SPE 112186

Production Optimization: System Identification and Uncertainty Estimation

Steinar M. Elgsaeter, Olav Slupphaug, and Tor Arne Johansen, Norwegian University of Science & Technology

This paper was prepared for presentation at the 2008 SPE Intelligent Energy Conference and Exhibition held in Amsterdam, The Netherlands, 25–27 February 2008.

Production Optimization; System Identification and Uncertainty Estimation

Steinar M. Elgsaeter* Olav Slupphaug Tor Arne Johansen

1 Abstract

Real-time optimization of oil and gas production requires a production model, which must be fitted to data for accuracy. A certain amount of uncertainty must typically be expected in production models fitted to data due to the limited information content in data. It is usually not acceptable to introduce additional excitation at will to reduce this uncertainty due to the costs and risks involved.

The contribution of this paper is twofold. Firstly, this paper discusses estimation of uncertainty in production optimization resulting from fitting models to production data with low information content, a concept that has previously mainly been applied in reservoir management. Secondly, this paper illustrates how system identification can be used to find production models which can be solved with little computational effort and which are designed to be easily fitted to production data.

The method is demonstrated on a synthetic example before being applied to a case study of a North Sea oil and gas field. In offshore oil and gas production, the suggested method is expected to have applications in the development of structured approaches to uncertainty handling, for instance excitation planning and real-time optimization under uncertainty.

2 Introduction

Production in the context of offshore oil and gas fields, can be considered the total output of production wells, a mass flow of components including hydrocarbons, in addition to water, CO_2, H_2S, sand and possibly other components. Hydrocarbon production is for simplicity often lumped into oil and gas. Production travels as multiphase flow from wells through flow lines to a processing facility for separation, illustrated in Figure 1. Water and gas injection is used for optimizing hydrocarbon recovery of reservoirs. Gas lift can increase production to a certain extent by increasing the pressure difference between reservoir and well inlet.

Multiphase flow rates are hard to measure. Measurements of total produced single phase oil and gas rates are usually available, and estimates of total water rates can often be found by adding different measured water rates after separation. To determine the rates of oil, gas and water produced from individual wells, the production of a single well is usually routed to a dedicated test separator where the rate of each separated component is measured. In *single-rate well tests* rates are only measured for the current setpoint, while rates are measured for several setpoints in *multi-rate well tests*. The total amount of oil, gas and water which can be separated and processed is constrained by the capacity of facilities, these capacities are themselves uncertain. Normally production is at setpoints where some of these capacities are at their perceived constraints, therefore a multi-rate well test cannot be performed without simultaneously reducing production at some other well, which may cause lost production and a cost. There is also a risk that changes in setpoints during testing may cause some part of the facilities to exceed the limits of safe operation, which may force an expensive shutdown and re-start of production. Well tests are only performed when a need for tests has been identified due to the costs and risks involved. Well tests are a form of *planned excitation*, some planned variation in one or more decision variables designed to reveal information on production through measurements.

Production is constrained by several factors including, on the field level, the capacity of the facilities to separate components of production and the capacity of facilities to compress lift gas. The production of groups of wells may travel through shared flow lines or inlet separators which have a limited liquid handling capacity. The production of individual wells may be constrained due to slugging, other flow assurance issues or due to reservoir management constraints.

In the context of oil and gas producing systems, real-time optimization has been defined as a process of measure-calculate-control cycles at a frequency which maintains the system's optimal operating conditions within the time-constant constraints of the system [1]. It has been suggested that real-time optimization could be divided into subproblems on different time scales to limit complexity, and to consider separately reservoir management, optimization of injection and reservoir drainage on the time scales of months and years, and production optimization, maximization of value from the daily production of reservoir fluids [1]. Reservoir management typically specifies constraints on production optimization to link these problems.

The aim of production optimization is to determine setpoints for a set of chosen decision variables which are optimal by some criterion. These setpoints are implemented by altering the settings of production equipment. Decision variables may be any measured or computed variables associated with production which are influenced by changes in settings, but the number of decision variables is limited by the number of settings. We may for instance determine the settings of a gas lift choke by deciding a target lift gas rate, a target annulus pressure or a target gas lift choke opening. On short timescales the flow from individual wells can be manipulated by production choke settings, by gas lift choke settings and possibly by routing settings.

There are many reasons why a production model may not

*The authors would like to thank the Research Council of Norway, StatoilHydro and ABB for funding this work.

describe production accurately. One reason may be *structural uncertainty*, the model may have a structure which makes it impossible to describe production truly regardless of the choice of parameters. An important cause of structural uncertainty may be *un-modeled disturbances*, influences which are not accounted for in the model but which cause production to change with time. A second reason may be *measurement uncertainty*, measured production may differ from the actual production for some reason, for instance due to incorrect calibration of measurement equipment. A third reason may be *lack of informative data*, the data may have insufficient excitation to uniquely determine the parameters of the model. In practice all of these factors are usually present to some extent. Modeling uncertainty in the context of reservoir management has received some attention in recent years [2] [3], while uncertainty in production optimization has received less attention. One recent discussion of the topic is [4], which considered uncertainty in well tests for wells with rate-independent gas-oil ratio and water cut. In prior work we showed that the information content in production data may be low, and suggested investigating uncertainty in production optimization to be a topic for further research [5]. A recent technology survey has noted that few implementations of real-time optimization exist on offshore oil and gas production systems, which the authors attribute to the difficulty of fitting models to production data [6].

Production optimization requires a production model which is able to predict how changes in setpoints affect production. Models used in the production optimization of gas-lifted wells are normally based on commercial multiphase flow simulators, either by querying the simulator directly [7] or by building tables of simulator predictions, so called *proxy-models* [8]. Deriving production models using physics alone can be difficult, as physical equations describing multiphase flow may depend on a large number of parameters or variables which are not fully known. For instance, the relationship between flow rate and pressure drop in porous media such as reservoirs is very complex and depends on parameters such as rock properties, fluid properties, flow regime, fluid saturations in rock, compressibility of the flowing fluids, formation damage or stimulation, turbulence and drive mechanism [9]. Some physical phenomena in production are only modeled using empirical relationships, due to limited understanding of the physics involved. Multiphase flow in the reservoir and through flow lines, including gas-lift performance curves, flow through restrictions, and inflow from reservoir into well, require empirical closure relations (for recent discussion of such empirical relations for flow lines, see for instance [10], [11], for flow through restrictions see [12], for inflow relations see [9], and references therein). Empirical relationships can be fitted against laboratory experiments, but experiments can be costly and small deviations between laboratory model and field can produce large differences in observed flow [13]. Even the most carefully constructed production model will require some fitting against production data to reflect the influence of un-modeled disturbances and structural uncertainty, for instance skin effects near the well, erosion of chokes or the build up of wax or hydrates in flow lines.

Inferring relationships between past input-output data and present/future outputs of a system when very little a priori knowledge is available is known as *black-box modeling*, and the study of such methods is the topic of *system identification* [14]. System identification takes a pragmatic view of the choice of model structure, seeking model structures in a trial-and-error fashion which can be relatively easily fitted to data yet describe observations with sufficient accuracy. Emphasis is usually on keeping the number of parameters to be fitted low while introducing some physical knowledge to achieve required performance [15]. Experience has shown that for some applications black-box modeling may meet the requirements of industry as well as models derived using physical insight. For instance the majority model-predictive controllers are based on black-box models and these controllers are used extensively in refineries and petrochemical plants [16].

System identification has been applied to model offshore production of oil and gas earlier, notably for well monitoring in [17], and for production optimization of gas-coned wells [18]. Some authors have suggested that the main bottleneck in production optimization is the computational effort required to solve rigorous physical production models [8]. Models found through system identification tend to be solvable with little computational effort due to their simplicity, and to be easily maintainable, as they are designed to be updated against data with little human intervention.

2.1 Problem formulation

Cost-effective methods for the design and maintenance of production models is a significant hurdle for the proliferation of production optimization in oil and gas production. This paper makes two contributions toward reducing the cost of designing and implementing production optimization. Firstly, we investigate modeling production with the general methods of system identification, motivated by a desire for models which can be easily fitted to production data. Secondly, we study fitting production models to recent historical production data describing normal operations, as such data is available at little or no cost, and suggest how to quantify uncertainty that may result when such data has low information content.

3 Modeling for production optimization under uncertainty

In this section a system identification approach to modeling for production optimization is outlined, and we outline how to quantify parameter uncertainty when models are fitted to data with low information content.

3.1 Production optimization and parameter estimation

Throughout this paper, variables with a hat ($\hat{\ }$) denote estimates and variables with bars ($^-$) denote measurements. The intended application of the model suggested in this section is the production optimization problem on the form

$$\begin{bmatrix}\hat{u}(\theta) & \hat{x}(\theta)\end{bmatrix} = \arg\max_{u,x} M(x,u,d) \tag{1}$$

$$s.t. \quad 0 = f(x,u,d,\theta) \tag{2}$$

$$0 \leq c(x,u,d). \tag{3}$$

u are decision variables, for instance the opening of production valves or gas-lift rates. θ is a vector of model parameters which are fitted to a tuning set. d is a vector of measured and modeled disturbances which are independent of u. x is a vector which expresses the production of each modeled fluid for each modeled well. Which fluids to model in x is a design question which depends on the choice of $c(x, u, d)$ and $M(x, u, d)$. Production optimization determines decision variables $\hat{u}(\theta)$ and the associated optimal production $\hat{x}(\theta)$ which maximizes an objective function $M(x, u, d)$, while obeying the production model (2) and production constraints (3). (3) may express constraints in the capacity of downstream processing equipment to separate oil, gas and water, as well as reservoir constraints and other constraints. $M(x, u, d)$ is most often the total rate of produced oil.

Let $\bar{y}(t)$ be a vector of measurements, and let $\hat{y}(x(t), u(t), d(t), \theta)$ be an estimate of those measurements. The parameter estimation problem attempts to determine θ so that $\bar{y}(t)$ and $\hat{y}(x(t), u(t), d(t), \theta)$ match as closely as possible. Parameter estimation considers a set of historical production data called the *tuning set*

$$Z^N = \begin{bmatrix}\bar{y}(1) & \bar{d}(1) & \bar{u}(1) & \bar{y}(2) & \bar{d}(2) & \bar{u}(2) & \dots \\ & & & \bar{y}(N) & \bar{d}(N) & \bar{u}(N)\end{bmatrix}. \tag{4}$$

Let the residuals for a given model structure and estimate θ be given by

$$\epsilon(t, \theta) = \bar{y}(t) - \hat{y}(x(t), u(t), d(t), \theta), \quad \forall t \in \{1, 2, \dots, N\} \quad . \tag{5}$$

A parameter estimate $\hat{\theta}$ is found by minimizing the sum of squared residuals:

$$\hat{\theta} = \arg\min_{\theta} \sum_{t=1}^{N} w(t) \| D_y \epsilon(t, \theta) \|_2^2 + V_s(\theta), \tag{6}$$

$$s.t. \quad c_\theta(\theta) \leq 0, \tag{7}$$

where $w(t)$ is a user-specified weight. The components of y may have different ranges, yet parameter estimation should give similar weight to minimizing residuals of all measurements, which motivates normalizing residuals with a diagonal matrix D_y. To improve parameter estimates, physical knowledge can be included in (6)–(7) in terms of *soft constraints* $V_s(\theta)$ or *hard constraints* $c_\theta(\theta)$. From (1)–(3) it should be clear that $\hat{\theta}$ will influence $\hat{u}(\theta)$, while (4)–(6) illustrate that $\hat{\theta}$ is influenced by the information content in Z^N. These relationships are illustrated in Figure 2.

The tuning set should only consist of historical production data which is consistent with current production, which implies that the effects of un-modeled disturbances should be negligible over the time interval spanned by the tuning set. If the information content in the tuning set is low it may not be possible to determine a unique $\hat{\theta}$ from (6)–(7).

3.2 Local production models for optimization

Assume that the oil, gas and water rates measured at the most recent well test are q^l, measured at time t^l and at setpoint (u^l, d^l). Consider a model which is *locally valid* in the sense that it attempts to predict the rates $\hat{q}^i$ for values of (u, d) close to (u^l, d^l). The choice to consider locally valid production models is motivated by two observations. Firstly, a locally valid model may be sufficient as long as production optimization only attempts to suggest new setpoints close to u^l, and as long as d^l has not changed significantly since the last well test. Secondly, decision variables often vary within a narrow range in production data from normal operations [5].

To simplify modeling we assume that the effects of changes in (u, d, t) from (u^l, d^l, t^l) on q^i, the vector of modeled rates for well i, can be described by separate kernel functions f_u, f_d, f_t:

$$\hat{q}^i = q^{l,i} f_d^i(d^i, d^{l,i}, \theta) \cdot f_u^i(u^i, u^{l,i}, \theta) \cdot f_t^i(t, t^l, \theta) \tag{8}$$

Each kernel function may be further separated as necessary, f_t may for instance be divided into kernel functions describing depletion and transients, and f_u and f_d may be divided into kernel functions describing different components of u and d as necessary. In addition, terms describing measurement uncertainty may be added as appropriate for joint data reconciliation and parameter estimation [19]. Kernel functions can be found by different means, either simulators, physical knowledge, well tests or fitted to a tuning set in a black-box manner. The model structure (8) may need to be tailored to describe the characteristics of each particular field.

A balance is required between models which are too rigid to describe the observed tuning set and models which are too flexible. Models which are too flexible can suffer from a phenomenon known as *over-fitting*, where the fitted model describes the tuning data set well while the model describes other data poorly. Fitted models which are almost equally capable of describing a set of independent data, a "validation data set" and the tuning set should be preferred, as such models do not suffer from over-fitting [14].

3.3 Estimating parameter uncertainty

We wish to exploit production data from normal operations as much as possible as such data can be obtained at low costs, yet low information content can result in significant parameter uncertainty if models are fitted to such data. If $\hat{\theta}$ is erroneously assumed to describe production while parameter uncertainty is significant, production optimization may suggest setpoint changes $\hat{u}(\hat{\theta})$ that are infeasible, sub-optimal or may even reduce profit. Rather than abandon the use of production data from normal operations, we propose quantifying parameter uncertainty. Further work may focus on exploiting this quantification of uncertainty to devise strategies for production optimization under uncertainty.

A standard result of system identification is that the matrix

$$P_\theta = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \lambda_0 \left[E\left\{ \left(\frac{\partial \hat{y}(t, \hat{\theta})}{\partial \theta} \right) \left(\frac{\partial \hat{y}(t, \hat{\theta})}{\partial \theta} \right)^T \right\} \right]^{-1} \tag{9}$$

is the covariance matrix of the asymptotic distribution of prediction-error estimates [14]. Estimates $\hat{P}_\theta$ are an expression of parameter uncertainty which can be found from the tuning set and model, the diagonal elements of $\hat{P}_\theta$ are the variance of each component of θ.

The derivation of (9) requires $\epsilon(t,\hat{\theta})$ to be a sequence of zero mean independent variables with variance λ_0, as well as some technical conditions and invokes the central limit theorem. Approximations of P_θ can be obtained from a finite set of data of length N, but this approximation can introduce errors, especially when N is small. In nonlinear identification, numerical solvers may return estimates $\hat{\theta}$ which are one of several local optima rather than the global optima of (6)–(7). In such cases it is not clear that (9) will be a valid description of parameter uncertainty. The matrix product in (9) may be an ill-conditioned matrix when Z^N has low information content, in which case matrix inversion can be numerically inaccurate.

Parameter uncertainty can also be estimated numerically using *bootstrapping* [20]. Bootstrapping decomposes $\bar{y}(t)$ into a systematic, modeled component $y(u,d,\hat{\theta})$ and a stochastic process e: $\bar{y}(t) = y(u,d,t) + e(t)$, and assumes that the observed residuals ϵ are a representative distribution of $e(t)$. Bootstrapping uses this assumption to construct a large number of synthetic measurements with a set of re-sampled stochastic component, and re-estimates $\hat{\theta}$ for each of the synthetic measurements. Residuals for time-instances where measurement errors or large un-modeled disturbances cause gross errors should be detected in pre-analysis and excluded.

The advantage of bootstrapping over asymptotic analysis is that it makes assumptions about the model and data set which are possibly less stringent, but finding estimates of uncertainty requires significant computational effort as the parameter estimation problem is solved a large number of times numerically.

3.4 Summary

The suggested approach to modeling and parameter uncertainty estimation is summarized in Algorithm 1.

Algorithm 1 *Given a dataset Z^N (4) of historical data describing production, and let N_s be the desired number of re-samples.*

- *Choose a model structure (2) on the form (8) as applicable to the particular field, and*
- *estimate a nominal $\hat{\theta}$ from (6)–(7). Apply constraints to assist parameter estimation as applicable.*
- *For the number of re-samples N_s*
  - *generate a data set Z_r^N, by sampling observed residuals ϵ N times and adding them to the output estimated using the process model and the nominal $\hat{\theta}$, and*
  - *determine a re-sampled parameter estimate $\hat{\theta}_r$ for Z_r^N from (6)–(7).*
- *The distribution of $\hat{\theta}_{r,i} \quad \forall i = 1,\ldots,N_s$ is an estimate of the parameter uncertainty.*

4 Synthetic examples and a case study

In this section the suggested approach is validated on a synthetic example and a case study of actual field data is performed. All simulations in this paper are implemented and solved in MATLAB[1] using the TOMLAB[2] toolbox.

4.1 Production model

We consider a well *decoupled* from other wells when changes in its production do not influence the production of other wells. In this paper we will consider the case of a field with n_w decoupled, gas-lifted wells producing predominantly oil, gas and water. We will consider change in gaslift rates $\Delta q_{gl}^i \stackrel{\text{def}}{=} \frac{q_{gl}^i}{q_{gl}^{l,i}} - 1, i \in 1,2,\ldots n_w$ as the decision variable $u = \Delta q_{gl}$. We consider the relative production valve opening $z \in [0,1]$ as a modeled disturbance. Let the most recently measured rate of a given component of production be $q^{l,i}$, measured for gas lift rate $q_{gl}^{l,i}$ and relative valve opening $z^{l,i}$ at time t^l. As profit depends on total oil production and constraints are linked to total production of gas and water, we choose to model the production of oil, gas and water for each well, $(\hat{q}_o^i, \hat{q}_g^i, \hat{q}_w^i) \quad \forall i = 1,2,\ldots,n_w$. To simplify estimation of $\hat{\theta}$ we will assume that kernel functions $f_z^i(z^i, z^{l,i})$ based solely on physical knowledge will describe production sufficiently well. The error introduced by this assumption should be small, as the production chokes for most wells are either fully opened or fully closed most of the time. As gas lift rates are the decision variable, we choose to fit the parameters of gas-lift kernels $f_{gl}^i(q_{gl}^i, q_{gl}^{l,i}, \theta)$ to production data. It is feasible to model time-variant effects by interpolating between single-rate well tests, but we choose not to include such effects in our model as test-separator measurements of gas are unreliable and often exhibit large non-physical variations. We choose to not model transients as few transients are visible on the time-scales and sampling rate considered. We will assume that discrepancies between test separator and total rate measurements can be described sufficiently by identifying bias terms β_y for each total rate measurement, and will not attempt to compensate further for measurement uncertainty. We assume that wells are decoupled and elect to not model coupling between wells. The form of (8) we choose to consider in this paper is

$$\hat{q}_o^i = \max\{0, q_o^{l,i} \cdot f_z(z^i, z^{l,i})(1 + \alpha_o^i \Delta q_{gl}^i + \kappa_o^i (\Delta q_{gl}^i)^2)\} \quad (10)$$

$$\hat{q}_g^i = \max\{0, q_g^{l,i} \cdot f_z(z^i, z^{l,i})(1 + \alpha_g^i \Delta q_{gl}^i + \kappa_g^i (\Delta q_{gl}^i)^2)\} \quad (11)$$

$$\hat{q}_w^i = \max\{0, q_w^{l,i} \cdot f_z(z^i, z^{l,i})(1 + \alpha_w^i \Delta q_{gl}^i + \kappa_w^i (\Delta q_{gl}^i)^2)\} \quad (12)$$

Kernel functions $f_{gl}(q_{gl}^i, q_{gl}^{l,i}, \theta)$ are chosen as second order polynomials, where $\alpha_o^i, \alpha_g^i, \alpha_w^i$ expresses gradient information and $\kappa_o^i, \kappa_g^i, \kappa_w^i$ expresses curvature in oil, gas and water gas-lift curves for well i, respectively. $f_z(z^i, z^{l,i})$ is a nonlinear kernel which should express the nonlinear relationship between valve opening and production, which obeys $f_z(0, z^{l,i}) = 0$ and $f_z(z^i, z^{l,i}) = 1$. In this paper we choose

$$f_z(z^i, z^{l,i}) = \frac{1 - (1 - z^i)^k}{1 - (1 - z^{l,i})^k}. \quad (13)$$

We will consider the total rates of oil, gas and water $q_o^{tot} \stackrel{\text{def}}{=} \sum_{i=1}^{n_w} q_o^i, q_g^{tot} \stackrel{\text{def}}{=} \sum_{i=1}^{n_w} q_g^i, q_w^{tot} \stackrel{\text{def}}{=} \sum_{i=1}^{n_w} q_w^i$ as elements in the

[1] The Mathworks,Inc., version 7.0.4.365

[2] TOMLAB Optimization Inc., version 5.5

measurement vector $y \stackrel{\text{def}}{=} \begin{bmatrix} q_o^{tot} & q_g^{tot} & q_w^{tot} \end{bmatrix}^T$. Let R be a routing matrix of ones and zeros defined such that $\bar{y} \stackrel{\text{def}}{=} \bar{R}x$ when measurement uncertainties can be neglected. Let an estimate of production for a given parameter estimate $\hat{\theta}$ be

$$\hat{y}(u(t), d(t), \hat{\theta}, t) \stackrel{\text{def}}{=} Rx + \beta_y \tag{14}$$

where β_y is a time-invariant measurement bias to be estimated. The components of θ in this paper are $\alpha_o^i, \alpha_g^i, \alpha_w^i$ and $\kappa_o^i, \kappa_g^i, \kappa_w^i$ for all wells $i = 1, 2, \ldots, n_w$ and β_y. (10)-(12) does not assume rate-independent ratios, and may therefore be able to describe wells with oil-, gas- or water coning in steady state.

4.1.1 Numerical solution of the parameter estimation problem

Estimating the parameters of (8) is a nonlinear programming problem in general, while (10)–(12) has a linear-in-variables structure, and we estimate $\hat{\theta}$ with the linear-least squares solver *lssol* in TOMLAB.

The parameter estimation problem can be divided into three separate sub-problems, one for oil, one for gas and one for water, but we choose to solve these problems as a single parameter estimation problem and use constraints to link these problems, as we will discuss in the section below.

4.1.2 Physical knowledge in parameter estimation

It is impossible to give an exhaustive list of all conceivable physical constraints on θ, but a short review is given.

The ratios between phases, such as gas-oil ratio or water-oil ratio, are rate-independent for some wells, and this qualitative knowledge can be used to simplify the parameter estimation problem. The ratio r between phases m and n for well i is $r_{m,n}^i \stackrel{\text{def}}{=} \frac{q_m^i}{q_n^i}$. A rate-independent ratio $r_{m,n}^i$ obeys

$$r_{m,n}^i = \frac{q_m^{l,i}(1 + \alpha_m^i \Delta q_{gl}^i + \kappa_m^i \Delta (q_{gl}^i)^2)}{q_n^{l,i}(1 + \alpha_n^i \Delta q_{gl}^i + \kappa_n^i (\Delta q_{gl}^i)^2)} \quad \forall \Delta q_{gl}^i. \tag{15}$$

(15) could be enforced by hard constraints $\alpha_m^i = \alpha_n^i$ and $\kappa_m^i = \kappa_n^i$, or alternatively by a soft constraint

$$V_s(\theta) = w_r \left(\alpha_m^i - \alpha_n^i\right)^2 + w_r \left(\kappa_m^i - \kappa_n^i\right)^2. \tag{16}$$

On wells where the gas-oil or water-oil ratios are known to be rate-dependent we may still know that these ratios do not vary by more than a given percentage and this may be included as difference constraints. If we expect

$$\frac{q_m^{l,i}}{q_n^{l,i}} R_L < \frac{q_m^i(q_{gl}^i)}{q_n^i(q_{gl}^i)} < \frac{q_m^{l,i}}{q_n^{l,i}} R_U, \quad \forall q_{gl}^i, \tag{17}$$

where $0 < R^L < 1 < R_U$, we could enforce constraints $\alpha_o^i R_L \leq \alpha_g^i$,$\kappa_o^i R_L \leq \kappa_g^i$, $\alpha_o^i R_U \leq \alpha_g^i$ and $\kappa_o^i \leq R_U \kappa_g^i$.

For gas-lifted wells it is reasonable to expect the increasing friction at increased gas-lift rates to result in performance curves with negative curvature $\frac{\partial^2 m_{tot}^i}{\partial q_{gl}^2} < 0$, and this qualitative knowledge can be expressed in terms of θ. Let $\dot{m}^i(\Delta q_{gl}^i)$ be the mass rate of production from well i for a given gas lift rate Δq_{gl}^i, and let $\Delta \dot{m}^i \stackrel{\text{def}}{=} \dot{m}^i(\Delta q_{gl}^i) - \dot{m}^i(0)$. If the rate of produced mass from well i can be assumed to consist predominantly of oil, gas and water with densities ρ_o, ρ_g and ρ_w, respectively:

$$\begin{aligned} \Delta \dot{m}_{tot}^i = \rho_o q_o^{l,i} \left(\alpha_o^i \Delta q_{gl}^i + \kappa_o^i (\Delta q_{gl}^i)^2\right) + \\ \rho_g q_g^{l,i} \left(\alpha_g^i \Delta q_{gl}^i + \kappa_g^i (\Delta q_{gl}^i)^2\right) + \\ \rho_w q_w^{l,i} \left(\alpha_w^i \Delta q_{gl}^i + \kappa_w^i (\Delta q_{gl}^i)^2\right) + \\ \rho_g (q_{gl}^i - q_{gl}^{l,i}). \end{aligned} \tag{18}$$

with second derivative

$$\frac{\partial^2 \Delta \dot{m}_{tot}^i}{\partial (\Delta q_{gl}^i)^2} = \rho_o q_o^{l,i} \kappa_o^i + \rho_g q_g^{l,i} \kappa_g^i + \rho_w q_w^{l,i} \kappa_w^i < 0. \tag{19}$$

(19) could be added as a soft-constraint. One technique that can help reduce over-fitting is regularization [21]. Let α be a vector of $\alpha_o^i, \alpha_g^i, \alpha_w^i, \quad i = 1, \ldots, n_w$ and let κ be a vector of $\kappa_o^i, \kappa_g^i, \kappa_w^i, \quad i = 1, \ldots, n_w$. Regularization terms

$$\begin{aligned} V_s(\theta) = w_{reg}^\alpha (\alpha - \alpha_{reg})^T (\alpha - \alpha_{reg}) + \\ w_{reg}^\kappa (\kappa - \kappa_{reg})^T (\kappa - \kappa_{reg}) \end{aligned} \tag{20}$$

can be added as soft constrains, penalizing deviation from $\alpha = \alpha_{reg}$ and $\kappa = \kappa_{reg}$. w_{reg}^α and w_{reg}^κ are weighting parameters. α_{reg} and κ_{reg} could be estimates of gradients and curvature determined from multi-rate well tests. Upper and lower bounds on $\hat{\theta}$, based on knowledge of the expected shape of the gas lift performance curve, may improve the performance of some solvers. Such bounds should be *loose*, i.e. so wide that the solver is expected to find solutions within rather than on these bounds when there is any excitation of the decision variables associated with a given parameter. Loose bounds also ensure that unrealistic parameter values are not chosen in those cases when there is very little excitation of certain modes.

4.2 A synthetic example

In this subsection the properties and capabilities of the proposed modeling approach is illustrated on a synthetic example. Consider a field with 8 gas-lifted wells, each with production of oil q_o^i, gas q_g^i, and water q_w^i, at rates which vary with the gas lift rate q_{gl}^i. Wells are grouped into pairs with performance given by similar equations. Wells 1 and 2 have rate-independent gas-oil ratio and water-oil ratio:

$$q_o^i = 10 + 0.025 q_{gl} + \sqrt{q_{gl}^i} - 0.000013 (q_{gl}^i)^2 \tag{21}$$

$$q_g^i = 30 q_o^i \tag{22}$$

$$q_w^i = 0.5 q_o^i \tag{23}$$

Wells 3 and 4 have gas coning, and a rate-dependent gas-oil ratio is chosen, while the water-oil ratio is rate-independent:

$$q_o^i = 10 + 0.033 q_{gl} + \sqrt{q_{gl}^i} - 0.000017 (q_{gl}^i)^2 \tag{24}$$

$$q_g^i = 30 \left(\frac{q_{gl}^i}{1000} - 1\right) q_o^i \tag{25}$$

$$q_w^i = \frac{0.5}{1 - 0.5} q_o^i \tag{26}$$

Wells 5 and 6 have water coning, and have a rate-dependent water-oil ratio, while the gas-oil ratio is rate-independent:

$$q_o^i = 15 + 0.028q_{gl} + \sqrt{q_{gl}^i} - 0.000022(q_{gl}^i)^2 \tag{27}$$

$$q_g^i = 40q_o^i \tag{28}$$

$$q_w^i = \frac{0.4(1 + 0.3(q_{gl}/1000 - 1))}{1 - 0.4(1 + 0.3(q_{gl}/1000 - 1))} q_o^i \tag{29}$$

Wells 7 and 8 have oil coning, and have a rate-dependent water-oil ratio while the gas-oil rate is rate-independent:

$$q_o^i = 14 + 0.15q_{gl}^i - 0.00003(q_{gl}^i)^2 \tag{30}$$

$$q_g^i = 35q_o^i \tag{31}$$

$$q_w^i = \frac{0.6(1 - 0.20(q_{gl}^i/1000 - 1))}{1 - 0.6(1 - 0.20(q_{gl}^i/1000 - 1))} q_o^i \tag{32}$$

Suppose that we have a set of historical measured total production rates and accompanying gas lift rates of each well, against which to fit our model, shown in Figure 5. The model is fitted with Algorithm 1, firstly with only loose bounds on θ, secondly knowledge of rate-independent gas-oil and water-oil ratios were applied as soft constraints where appropriate. The resulting models and model uncertainties are shown in Figure 5. This example illustrates that the parameter estimation is able to gain information from measurements of total rates for a synthetic field with a small to medium number of wells with varying characteristics. As the synthetic example has pairs of similar wells, the effect of the level of excitation on the estimates of uncertainty is visible in Figure 5. The matrix P_θ was close to singular, with $\mathrm{cond}(P_\theta) \approx 10^{19}$, which would make it difficult to obtain meaningful estimates of parameter uncertainty using asymptotic analysis in this case.

Although the suggested methodology is able to produce uncertainty estimates which describe the true gaslift-curves well for most wells, the method can break down if the information on the gas lift performance of some wells is virtually nonexistent, as is the case for well 2 as shown in Figure 5. Such cases should be detected in post-analysis, and either further excitation of such wells should be performed, or these wells should be left out of production optimization.

4.3 Case study: Production data from a North Sea oil field

In this section, the proposed method is applied to the production data from an offshore North Sea oil field producing mainly oil, water, and gas from 20 gas-lifted wells. The operator of the field requested that data be kept anonymous and all results are therefore presented in terms of normalized variables. A tuning set spanning five months with sampling time of one hour was considered. To compare the significance on estimated uncertainty of including physical assumptions as soft constraints, we will fit the model to production data in two runs. In the first run, only loose bounds $0 < \alpha < 1$ and $-1 < \kappa < 0$ are implemented. In the second run soft constraints were on rate-independent ratios were enforced, difference constraints were applied to limit the rate-dependency of gas-oil ratios, chosen as $R_L = 0.7$,$R_U = 1.3$ for all wells, and the curvature constraint was enforced. Past well tests indicate that the watercut is rate-independent, and soft-constraints are included in the second simulation run to enforce this rate-independence. We choose $k = 5$, so that changes in z^i have a large influence on production when z^i is small and a small influence on production when z^i is large. To reduce the impact of reservoir depletion, and un-modeled disturbances, on estimates, measured oil rates were de-trended as is shown in Figure 5 and in addition residuals in (6) were weighted with a forgetting factor [14]

$$w(t) = \lambda^{N-t}, \tag{33}$$

chosen as $\lambda = 0.5^{\frac{1}{N-1}}$. As the production model is intended to be a local description around (q^l, u^l, d^l), we omitted shutdowns from the tuning set.

4.3.1 Results

The fit between the nominal production model and the tuning data set is shown in Figure 6. A set of $N_s = 100$ bootstrap replications were designed and the parameters were re-estimated, as outlined in Algorithm 1. The tuning sets used when bootstrapping are shown in Figure 7. Estimates of the bootstrap models are compared with the validation set in Figure 8. The resulting models are compared in Figure 9. Nominal parameter estimates result in models and estimates similar to the those shown in and Figure 9 and Figure 8.

4.3.2 Discussion

An advantage of the chosen modeling approach is that estimated models can be interpreted by the shape of performance curves, as in Figure 5, with which industry practitioners may already be familiar. In Figures 6 and 8 we observe several large changes in measured rates which cannot be explained with changes in z or q_{gl}. These deviations could be the result of measurement error or disturbances stemming from the reservoir or topside process facilities. As the model is intended for determining day-to-day operating setpoints rather than prediction for long time intervals, the bias between model estimates and measurements observed in Figure 8 is less significant than the ability of the model to predict the change in production resulting from changes in setpoints shown in Figure 8. From Figure 5 we see that the uncertainty varies greatly between different phases and for different wells. Comparing the relative degree of uncertainty for different wells may be useful for excitation planning. In this paper we have assumed that the only measurement uncertainty is a steady bias in total rate measurements β_y. Measurement uncertainty in rate measurements at the test separator may cause errors in the operating point, and we have not studied the significance of such errors on estimates or how to mitigate or estimate these uncertainties. The assumption of time-invariant production in the tuning set is clearly an approximation, as a falling trend in oil production due to depletion is visible and some wells have varying phase ratios during the course of the tuning set. By choosing a relatively long tuning set, we err on the side of caution with regards to estimated uncertainties, but risk introducing biases in parameter estimates. Reservoir dynamics and changed processing conditions are treated as un-modeled disturbances in this paper and only the most recent well test was used. A visible decline in measured oil production was treated by de-trending and introducing a forgetting factor. It may be possible to improve on this approach either through modeling disturbances and by exploiting older well tests. Although we have focused on gas-lifted wells in this paper, the system identification approach to modeling is general and has extensions to other types of field and wells, for instance wells where coupling is significant or wells where production choke settings are decision variables. The approach to estimating uncertainty resulting from a low information content is general and has extensions to estimating the uncertainty in proxy-models derived from commercial simulators as well. The model and uncertainty estimates described may be applied to estimate the significance of uncertainty on production profits, for formulating structured business cases for uncertainty mitigation or for designing structured approaches to decision making under uncertainty, as suggested in [5].

5 Conclusion

The contribution of this paper is twofold. Firstly, it is discussed how parameter uncertainty can be quantified when models are fitted to data with little information content, a concept that has been little explored in the context of modeling for production optimization. Secondly, it is suggested how system identification can be used to find models for production optimization with the aim of reducing costs of design and maintenance of production optimization. The method was demonstrated on a synthetic example before being applied to a case study of real field data from a North Sea oil and gas field. The method suggested in this paper has applications in real-time optimization of day-to-day production and in the development of structured approaches to handling uncertainty, such as excitation planning/well test planning or robust production optimization.

References

[1] L. Saputelli, S. Mochizuki, L. Hutchkins, R. Cramer, M. Anderson, J. Mueller, A. Escoricia, A. Harms, C. Sisk, S. Pennebaker, J. Han, C. Brown, A. Kabir, R. Reese, G. Nuñez, K. Landgren, C. McKie, and C. Airlie, "*Promoting real-time optimization of hydrocarbon producing systems*", in 2003 SPE Offshore Europe Aberdeen, 2003. SPE 839781.

[2] A. Diab, B. Griess, and R. Schulze-Riegert, "*Application of global optimization techniques for model validation and prediction scenarios of a north african oil field*", in SPE Europec/EAGE Annual Conference and Exhibition, 2006. SPE 100193.

[3] A. Little, H. Jutila, and A. Fincham, "*History matching with production uncertainty eases transition into prediction*", in SPE Europec/EAGE Annual Conference and Exhibition, 2006. SPE 100206.

[4] H. Bieker, O. Slupphaug, and T. Johansen, "*Optimal well-testing strategy for production optimization: A monte carlo simulation approach*", in SPE Eastern Regional Meeting, 2006. SPE 104535.

[5] S. Elgsaeter, O. Slupphaug, and T. Johansen, "*Challenges in paramter estimation of models for offshore oil and gas production*", in International Petroleum Technology Conference, 2007.

[6] H. Bieker, O. Slupphaug, and T. Johansen, "*Real-time production optimization of offshore oil and gas production systems: Technology survey*", in SPE Intelligent Energy Conference and Exhibition, 2006. SPE 99446.

[7] P. Wang, "*Development and application of production optimization for petroleum fields*". PhD thesis, 2003.

[8] G. Zangl, T. Graf, and A. Al-Kinani, "*Proxy modeling in production optimization*", in SPE Europec/EAGE Annual Conference and Exhibition, 2006. SPE 100131.

[9] H. Beggs, "*Production Optimization: using NODAL analysis*", Oil & Gas Consultants Intl., Tulsa, Oklahoma, 2nd ed., 2003.

[10] A. Ullmann and N. Brauer, "*Closure relations for two-fluid models of two-phase stratified smooth and stratified wavy flows*", International Journal of Multiphase Flow, 32 (2006), pp. 82–105.

[11] C. Tribbe and H. Müller-Steinhagen, "*An evaluation of the performance of phenomenological models for predicting pressure gradient during gas-liquid flow in horizontal pipelines*", Int. J. Multiphase Flow, 26 (2000), pp. 1019–1036.

[12] S. Rastoin, Z. Schmidt, and D. Doty, "*A review of multiphase flow through chokes*", Journal of Energy Resources Technology, 119 (1997).

[13] O. Utvik, T. Rinde, and A. Valle, "*An experimental comparison between a recombined hydrocarbon-water fluid and a model fluid system in three-phase pipe flow*", Journal of Energy Resources Technology, 123 (2001), pp. 253–259.

[14] L. Ljung, "*System identification: Theory for the user*", Prentice Hall, Cambridge, 2nd ed., 1999.

[15] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.-Y. Glorennec, H. Hjalmarsson, and A. Juditsky, "*Nonlinear black-box modelling in system identification: a unified overview*", Automatica, 31 (1995), pp. 1691–1724.

[16] S. Qin and T. Badgwell, "*A survey of industrial model predictive control technology*", Control Engineering Practice, 11 (2003), pp. 733–764.

[17] H. Poulisse, P. van Overschee, J. Briers, C. Moncur, and K.-C. Goh, "*Continous well production flow monitoring and surveillance*", in SPE Intelligent Energy Conference and Exhibition, 2006. SPE 99963.

[18] A. Mjaavatten, R. Aasheim, S. Saelid, and O. Gronning, "*A model for gas coning and rate-dependent gas/oil ratio in an oil-rim reservoir*", in SPE Russian Oil and Gas Technical Conference and Exhibition, 2006. SPE 102390.

[19] C. M. Crowe, "*Data reconciliation — progress and challenges*", Journal of process control, 6 (1996), p. 89–98.

[20] B. Efron and R. Tibshirani, "*An Introduction to the Bootstrap*", Chapman & Hall, 1993.

[21] T. Johansen, "*On tikhonov regularization, bias and variance in nonlinear system identification*", Automatica, 33 (1997), pp. 441–446.

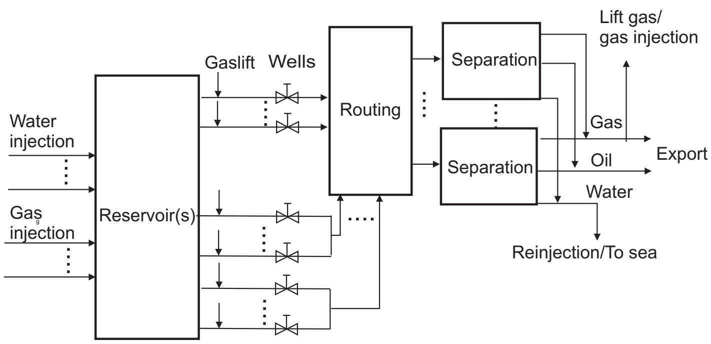

Figure 1: A schematic model of offshore oil and gas production.

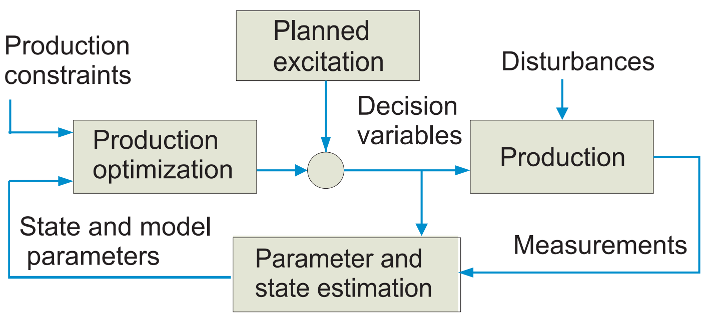

Figure 2: A model of the relationship between subproblems in production optimization.

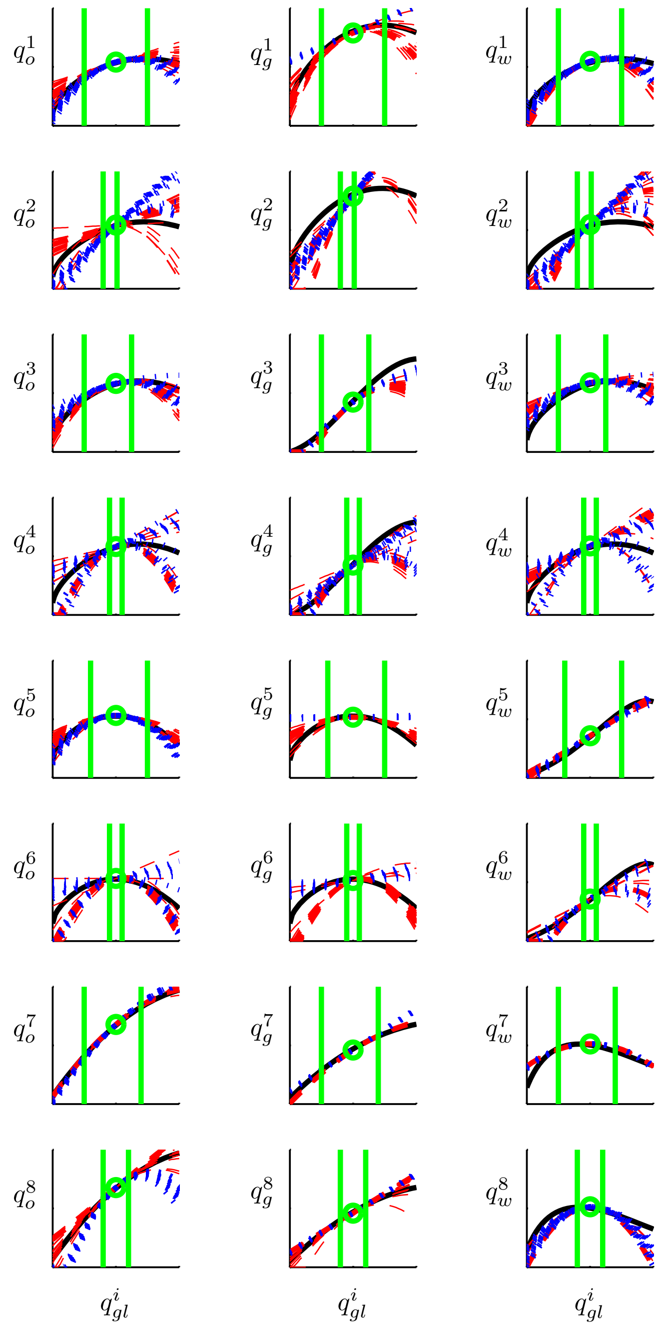

Figure 3: Synthetic example: True performance curve (solid), bootstrap performance curves without physical constraints (dashed), bootstrap performance curves with physical constraints (dotted), local operating point (circle) and the span of gas lift rates observed in tuning set (vertical solid).

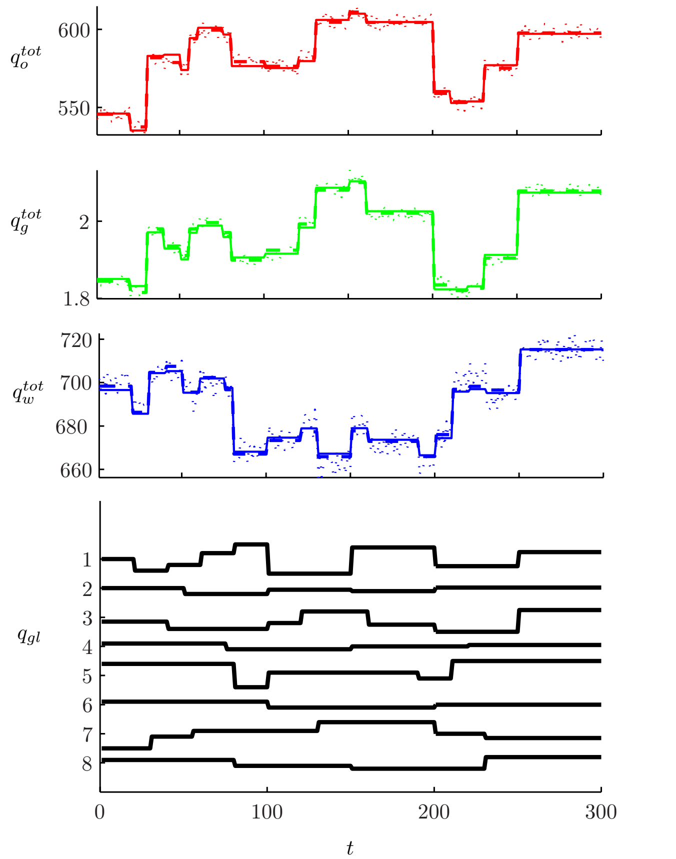

Figure 4: Synthetic example: Top three graphs: Measured total rates of oil, gas and water (dotted) compared with estimates of the fitted model without soft constraints (dashed) and with soft constraints (solid). Bottom graph: normalized gas lift rates of wells 1 through 8 plotted in ascending order.

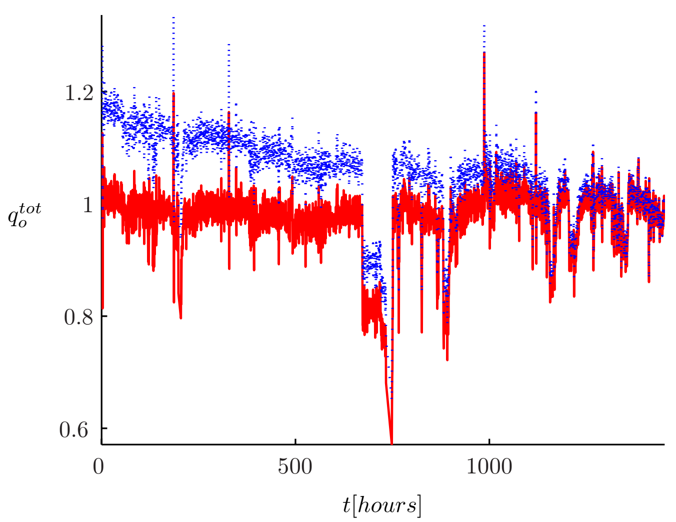

Figure 5: Case-study: Measured oil rates (dotted) and detrended oil rates used in tuning set (solid).

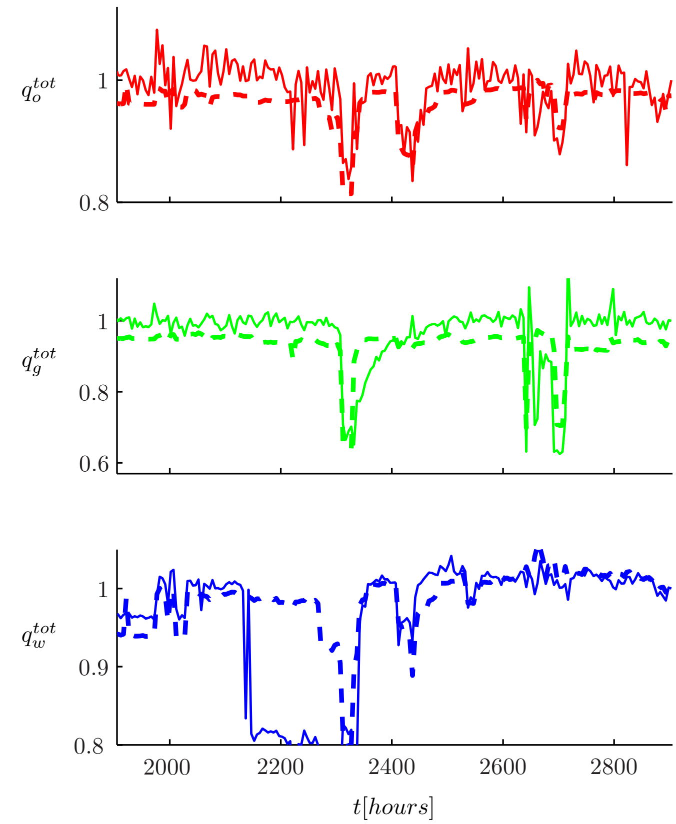

Figure 6: Case-study: Case study: Last portion of tuning set (solid) compared with estimates of the model including physical constraints (dashed). Sampling time reduced to five hours for clarity.

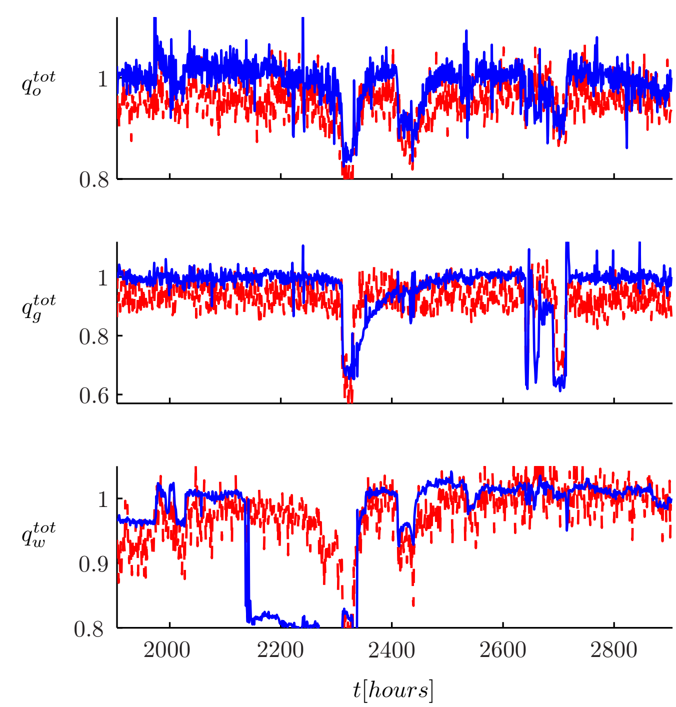

Figure 7: Case-study: Tuning set (solid) and one bootstrap replications (dashed), for model including physical constraints.

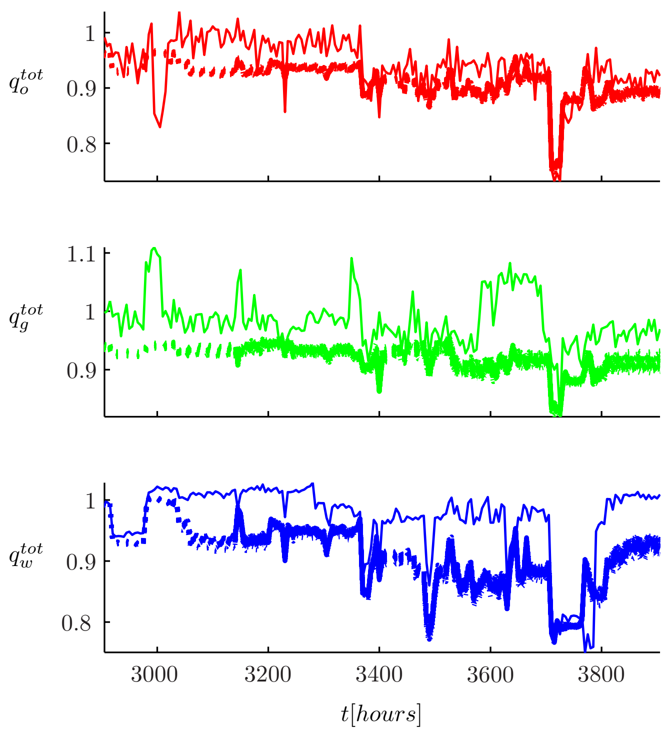

Figure 8: Case-study: Estimates of with bootstrap replications (dotted) compared with the validation data set (solid), for model including physical constraints.

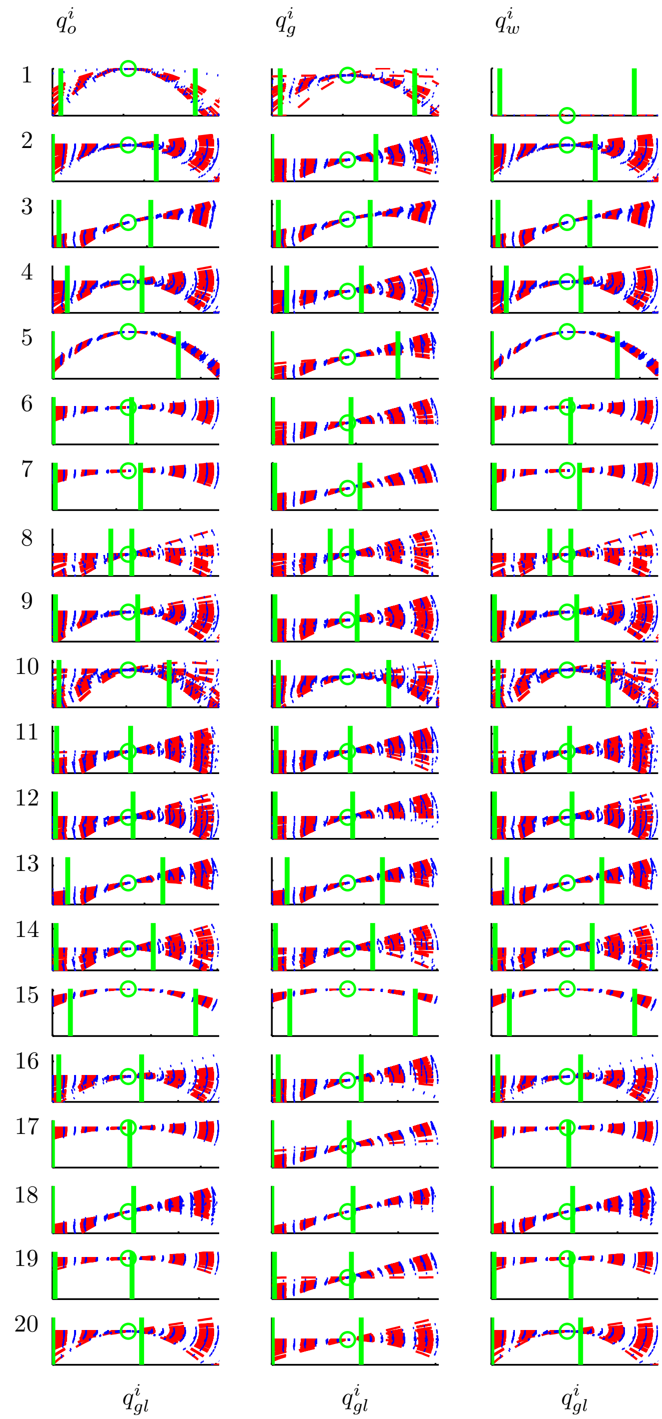

Figure 9: Case-study: Bootstrap performance curves without physical constraints (dashed), bootstrap performance curves with physical constraints (dotted), local operating point (circle) and the span of gas lift rates observed in tuning set (vertical solid). Lower left bound of all plots is (0,0). Well indices are along left margin.