PREDICTING WATER-FLOODING PERFORMANCE IN LOW-PERMEABILITY RESERVOIRS WITH LINEAR DYNAMICAL SYSTEM

by

Dashun Wang

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Computer Engineering

Department of Electrical and Computer Engineering
University of Alberta

Abstract

Several interwell connectivity models such as multiple linear regression (MLR) and the capacitance model (CM) have been proposed to model waterflooding performance in high-permeability reservoirs based on observed production data. However, the existing methods are not effective at characterizing the behavior of transient flows which are prevalent in low-permeability reservoirs.

This thesis presents a novel dynamic waterflooding model based on linear dynamical systems (LDS) to characterize the injection-production relationships in an oilfield during both stationary and nonstationary production phases. We leverage a state space model, in which the changing rates of control volumes between injector-producer pairs in the reservoir of interest serve as time-varying hidden states depending on the reservoir condition, and can thus predict production rates more accurately for low-permeability reservoirs and many dynamic scenarios. We propose a self-learning procedure for the model to train its parameters as well as the evolution of the hidden states only based on past observations of injection and production rates.

We tested LDS in comparison with the state-of-the-art CM method in a wide range of synthetic reservoir models including both high-permeability and low per-

meability reservoirs, as well as various dynamic scenarios involving varying bottom-hole injection pressure (BHIP), injector shut-ins and reservoirs of a larger scale. We also tested LDS on the real production data collected from Changqing oilfield containing low-permeability formations. Testing results demonstrate that LDS significantly outperforms CM in terms of modeling and predicting waterflooding performance in low-permeability reservoirs and various dynamic scenarios.

Acknowledgments

I would like to thank all the people who contributed in some way to the work described in this thesis. First and foremost, I am deeply indebted to my academic advisors, Dr. Di Niu and Dr. Huazhou Li. Their guidance and supervision help me to learn how to be a good researcher and pursue an academic career. Their willingness to discussion helped me through two important years of my life.

Additionally, I would like to thank my committee members Dr. Jie Chen and Dr. Zhehui Jin for their interest in my work and their time in my defence.

Finally, I would like to acknowledge friends and family who supported me during my time here. First and foremost I would like to thank Mom, Dad and other family members for their constant love and support. Also, I am lucky to have met lots of good friends here and have a happy life. Thanks for their friendship and unyielding support!

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Acronyms	Definition
BHIP	Bottom-Hole Injection Pressure
CM	Capacitance Model
CVCR	Control Volume Changing Rate
EKF	Extended Kalman Filter
EM	Expectation Maximization
EnKF	Ensemble Kalman Filter
HPRM	High-Permeability Reservoir Model
LDS	Linear Dynamical System
LLPRM	Larger-scale Low-Permeability Reservoir Model
LPRM	Low-Permeability Reservoir Model
MLE	Maximum Likelihood Estimator
SSM	State Space Model

Chapter 1

Introduction

1.1 Thesis Outline

This thesis is organized as follows. Chapter 1 introduces the problem of reservoir waterflooding prediction. Chapter 2 introduces our proposed model and the corresponding algorithms for learning the parameters of the model. We validate our model in chapter 3 and summarize the conclusion in chapter 4.

1.2 Previous Literature

The production of oil and gas resources plays an important role in fulfilling the ever-increasing demand for energy supply worldwide. As a secondary recovery method, waterflooding injects water into an oilfield through injection wells to maintain the pressure in the reservoir, driving oil towards production wells. The data of injection and production rates can be readily collected during the production process, based on which important insights into the characteristics of the reservoir can be derived.

Field operators of waterflooding need to control various operational parameters on a regular basis, e.g., the injection rates and the bottom-hole pressure (BHP). To make optimal decisions regarding water injection, it is important to predict the responses of oil production under different water injection schedules. In addition to conventional numerical simulation, several lightweight correlative methods have

been developed, among which some methods specifically focus on characterizing the interwell connectivity. Heffner *et al.* [1] used Spearman rank correlations to relate injection/production well pairs and associated these relationships with geomechanical properties of the reservoir. Panda and Chopra [2] evaluated the relationships between injection and production rates with classical artificial neural networks. Albertoni and Lake [3] proposed an interwell connectivity model based on multiple linear regression (MLR). Their work was extended to the well-known capacitance model (CM) by Yousef *et al.* [4]. CM takes into account both the compressibility and transmissibility effects so as to better capture the true attenuation and time lag between injection rates and production rates. CM has been later modified to accommodate different application scenarios. For instance, Kaviani *et al.* [5] made CM capable of modeling the cases with unmeasured fluctuating BHP. Soroush *et al.* [6] improved CM by taking into account variations in skin factors and frequent production interruptions.

However, weaknesses still exist in these improved capacitance models. First, by assuming a constant productivity index, CM is not suitable for modeling transient flows [7], which are prevalent and can last for an extended period of time in low-permeability reservoirs. Second, parts of the high-frequency injection signals will be lost due to large dissipations in injection rates [4]. Considering these limitations, some time-dependent models have been proposed to capture the injection-production dynamics during waterflooding. Liu *et al.* [8] proposed a nonstationary waterflooding model based on Extended Kalman filter (EKF), followed by an Ensemble Kalman Filter (EnKF) model proposed by Zhang *et al.* [9]. Liu *et al.* [8] assumed that the impulse response between injectors and producers could be modeled by the continuous-time unimodal function. But such an assumption is empirical in nature. Zhang *et al.* [9] extended CM to model waterflooding in multi-layered reservoirs. Although the parameters are dynamically inferred in the models proposed by Liu *et al.* [8] and Zhang *et al.* [9], they incorporate the states of the reservoir as a part of the model parameters and suffer from one drawback: since their model parameters are not learned based on data, the auto-regressive coefficients can only

be set to fixed values in advance, which may undermine the prediction accuracy especially in low-permeability reservoirs.

1.3 Linear Dynamical System

In this study, we develop a novel dynamic waterflooding model based on linear dynamical systems (LDS) [10] to characterize the injection-production relationships in an oilfield during both stationary and nonstationary production phases. Especially, our new model can yield more accurate production predictions for low-permeability reservoirs than CM. The proposed LDS-based method models the dynamic changes in the reservoir condition over time through a hidden state variable called *control volume* between each injector-producer pair, which may change over time and affect the injection-production responses in the reservoir. Furthermore, unlike the EKF by Liu *et al.* [8] and EnKF by Zhang *et al.* [9], we do not artificially assign the coefficients in LDS. Instead, we take advantage of an expectation-maximization (EM) algorithm [11] to train the model parameters and coefficients, allowing the LDS to choose the most suitable coefficients according to observation data.

We evaluated the performance of the proposed LDS method in a wide range of settings based on both synthetic data and real production data collected from an oilfield. We observe that in high-permeability reservoirs, LDS shows performance similar to CM. However, in low-permeability reservoirs, LDS significantly outperforms CM in terms of the prediction accuracy in production rates, under a wide range of scenarios, including the scenarios of varying bottom-hole injection pressure (BHIP), injector shut-ins, larger-scale reservoirs and the real field.

Chapter 2

Methodology

We propose a new procedure to model the relationship between injection and production rates in waterflooding reservoirs based on linear dynamical system (LDS). Due to its capability of characterizing transient and non-steady-state flow behavior, the proposed model can better predict waterflooding performance in low-permeability reservoirs, which was not well studied in prior work.

2.1 Background

In a waterflooding project, the oil reservoir can be treated as a system where the water injection rates serve as continuous-time impulse and production rates are the response [8]. Let $i(t)$ and $q(t)$ denote the aggregate injection rate and production rate at time t, respectively. The governing material balance equation for the whole reservoir is given by

$$c_t V_p \frac{d\bar{p}}{dt} = i(t) - q(t) \tag{2.1}$$

where c_t is the total compressibility, V_p is the pore volume of the reservoir, and $\frac{d\bar{p}}{dt}$ is the average changing rate in reservoir pressure.

Traditionally, CM assumes [4]

$$\bar{p}(t) = \frac{q(t)}{J} + p_{wf}(t), \tag{2.2}$$

where the productivity index J is a constant, and $p_{wf}(t)$ represents the bottom-hole pressure (BHP). The basic idea of CM is to combine the material balance Eq. 2.1 and the assumption $\bar{p}$ in Eq. 2.2, and use optimization methods to approximately solve the resulted differential equations to model the relationships between injection rates and production rates.

However, this assumption is mainly applicable to high-permeability reservoirs where the periods of transient flows can be short [4][12][6]. In contrast, in a low-permeability reservoir, transient flows can be dominant for a long period of time. In this case, the assumption made by CM that the productivity index J is constant cannot hold any more [7].

In our model, we do not assume that the productivity index J is a constant or that the reservoir pressure $\bar{p}(t)$ is linearly related to $q(t)$. Instead, the reservoir pressure is a general function of time and position, without assuming its specific form. And the changing rate of $\bar{p}(t)$ [13] is given by

$$\frac{d\bar{p}}{dt} = f(\vec{x}, t) \tag{2.3}$$

where $\vec{x}$ is a vector representing a specific geological position in the reservoir. In the following, we will combine the material balance equation in Eq. 2.1 with Eq. 2.3 to derive an LDS to model the injection-production dynamics in an oilfield.

2.2 Mathematical Formulation of LDS

For brevity, let $V = c_t V_p \bar{p}$ denote the reservoir volume and we use $V'(t)$ to denote its derivative $\frac{dV}{dt}$. Then, the material balance Eq. 2.1 is converted to

$$V'(t) = i(t) - q(t) \tag{2.4}$$

where $V'(t)$ is the changing rate of the reservoir volume. Due to Eq. 2.3, $V'(t)$ is also a function of both time and position.

We now formulate the material balance equations for a reservoir with I injectors

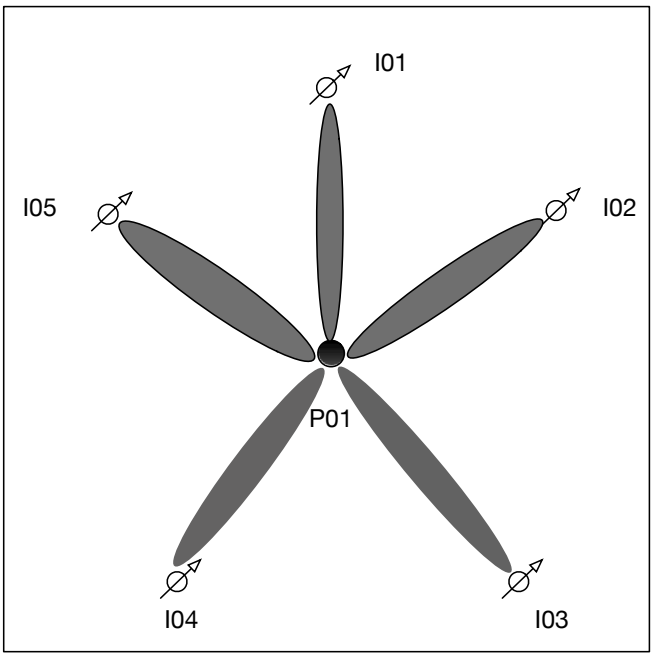

Fig. 2.1. The illustration of control volume from five injectors to a producer P01.

and J producers. In this case, let $\vec{i}(t) := [i_1(t), \cdots i_i(t), \cdots i_I(t)]^\top$, where $i_i(t)$ denotes the injection rate of the ith injector. Let $\vec{q}(t) := [q_1(t), \cdots q_j(t), \cdots q_J(t)]^\top$ where $q_j(t)$ denotes the production rate of the jth producer.

Our model will rely on the notion of the *control volume*, which refers to the volume of a reservoir space between a specific injector and a specific producer [14]. Through this reservoir space, water and/or oil flows from the injector to the producer. Let V_{ij} denote the control volume from injector i to producer j, as illustrated in Fig. 2.1 for a 5-spot reservoir example.

We denote the control volume changing rate (CVCR) from injector i to producer j by $V'_{ij}(t)$. Thus, similar to Eq. 2.4, we can write the material balance equation for each particular injector-producer pair (i, j) as

$$V'_{ij}(t) = i_i(t) - q_j(t) \tag{2.5}$$

For any given injector-producer pair, its CVCR V'_{ij} varies as a function of both time and position. Since V'_{ij} with different i and j indices correspond to wells at

different positions, V'_{ij} can reflect the variation of the reservoir volume dynamics across different geological locations.

For a particular producer j, we assume its production rate q_j is influenced by all the injectors. And let

$$\vec{V}'_j := [V'_{1j}, V'_{2j}, \cdots, V'_{ij}, \cdots, V'_{Ij}]^\top$$

represent the vector of CVCRs from all I injectors to producer j. For each given producer j, considering the weighted contributions from different injectors, Eq. 2.5 is converted to

$$\alpha^\top \vec{V}'_j(t) = \beta^\top \vec{i}(t) - q_j(t), \tag{2.6}$$

where α and β are coefficient column vectors.

Incorporating the injection rates as inputs, the production rates as outputs, and the CVCRs as hidden state variables, we can model the reservoir of interest as an LDS, which is essentially a state space model (SSM). Specifically, the state update of CVCRs should depend on both the previous CVCRs and the system inputs, i.e., the injection rates. Let $\vec{V}'_j(t+\Delta t)$ denote the CVCRs of producer j at time $t+\Delta t$. In other words, the auto-regressive exogenous update for CVCRs $\vec{V}'_j$ associated with producer j is

$$\vec{V}'_j(t+\Delta t) = f(\vec{V}'_j(t), \vec{i}(t)) + \varepsilon(t) \tag{2.7}$$

where $\varepsilon(t)$ is a Gaussian noise for updating $\vec{V}'_j(t)$. The discretized form of Eq. 2.7 is

$$\vec{V}'_j(n+1) = f(\vec{V}'_j(n), \vec{i}(n)) + \varepsilon(n) \tag{2.8}$$

where n is the discretized time step.

For computational efficiency, we model the CVCRs associated with each producer j as a linear recursive function of their previous values and the injection rates, leading to

$$\vec{V}'_j(n+1) = \mathbf{A}\,\vec{V}'_j(n) + \mathbf{B}\,\vec{i}(n) + \varepsilon(n) \tag{2.9}$$

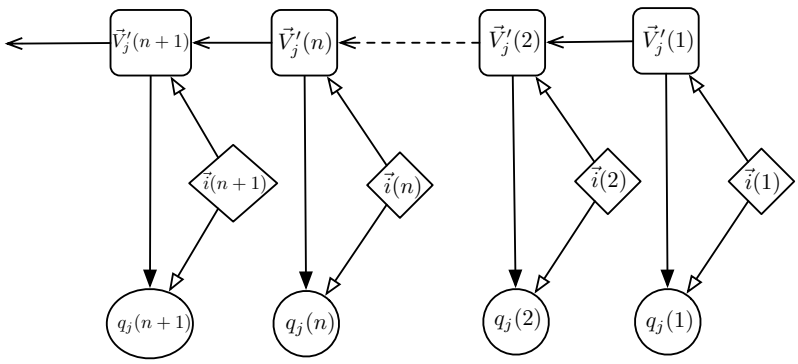

Fig. 2.2. An LDS to model waterflooding in a reservoir. CVCRs are denoted by squares. The injection rates are denoted by diamonds. The production rates are denoted by circles.

where $\mathbf{A}$ is the coefficient matrix of $\vec{V}'_j(n)$ and $\mathbf{B}$ is the coefficient matrix of $\vec{i}(n)$.

Similarly, discretizing the material balance equations in Eq. 2.6, we can obtain the observation equation for each producer j as

$$q_j(n) = -\alpha^\top \vec{V}'_j(n) + \beta^\top \vec{i}(n) + \gamma(n) \tag{2.10}$$

where $\gamma(n)$ is a Gaussian noise for the observed production rate.

Combining Eq. 2.9 and Eq. 2.10, we can obtain an LDS for each producer j as follows:

$$\vec{V}'_j(n+1) = \mathbf{A}\,\vec{V}'_j(n) + \mathbf{B}\,\vec{i}(n) + \varepsilon(n) \tag{2.11a}$$

$$q_j(n) = -\alpha^\top \vec{V}'_j(n) + \beta^\top \vec{i}(n) + \gamma(n) \tag{2.11b}$$

2.3 Physical Meanings of the LDS Model

Fig. 2.2 illustrates the dynamics between injection and production rates conditioned on the hidden reservoir states, i.e., the CVCRs. At each time step, the injection rates $\vec{i}(n)$ serve as the model input, and the production rate $q_j(n)$ serves as the observed output at producer j. The update of hidden CVCRs depends on both their previous values and the injection rates, based on Eq. 2.11a. On the other hand, the

output $q_j(n)$ is jointly determined by the CVCRs associated with producer j and the injection rates at time step n, according to Eq. 2.11b.

Let us now explain the physical meanings of model parameters. In Eq. 2.11b, the vector β represents the direct influence of all the injectors on the production rate at each producer. Thus, the meaning of each element in β is similar to that of *inter-well connectivity* in CM [4]. However, what is different from CM is that in LDS, there is an additional vector α that quantifies the impact of the introduced hidden reservoir states, represented by the CVCRs $\vec{V}_j'$, on each producer j. Therefore, the proposed LDS model can better characterize the transient behavior in waterflooding, leading to more accurate prediction especially in low-permeability reservoirs and various dynamic scenarios involving varying BHIP or injector shut-ins.

Note that the basic idea of CM is to express productivity index J as a constant and then substitute $\frac{d\bar{p}}{dt}$ in the material balance equation, i.e., Eq. 2.1, using Eq. 2.2. The same idea is also adopted in other multiple linear regression (MLR) models. In contrast, our model treats CVCRs $\vec{V}_j'$ as dynamic variables and trains such hidden variables together with all the model parameters $\alpha, \beta, \mathbf{A}, \mathbf{B}$ based on the past observations of injection and production data. Therefore, in this case, the productivity index is no longer assumed constant.

2.4 Training and Prediction

We now describe how to train the LDS model in Eq. 2.11 given historical observations of injection and production rates up to a certain time period N. Since in Eq. 2.11, each producer can be handled individually, without loss of generality, we can focus on a specific producer j. To simplify notations, we use a vector x_n to represent the hidden state variable at time step n, i.e., the CVCRs $\vec{V}_j'(n)$, with x_0 denoting the initial value of x_n, i.e., $\vec{V}_j'(0)$. We use a vector u_n to denote the system inputs at time step n, i.e., the injection rates $\vec{i}(n)$, and use a scalar y_n to denote the system output, i.e., the production rate $q_j(n)$ of producer j, at time step n. Then, we can rewrite the LDS in Eq. 2.11 in the nomenclature of *system identification* as

follows:

$$x_{n+1} = A\,x_n + B\,u_n + \varepsilon_n \tag{2.12a}$$

$$y_n = C\,x_n + D\,u_n + \gamma_n \tag{2.12b}$$

$$x_0 = \mu_0 + v \tag{2.12c}$$

where v is the noise of the initial state variable, ε_n is the system noise, and γ_n is the observation noise. All three noise terms follow Gaussian distributions, i.e., $v \sim \mathcal{N}(0, V_0)$, $\varepsilon_n \sim \mathcal{N}(0, \Sigma)$, $\gamma_n \sim \mathcal{N}(0, \Gamma)$, where V_0, Σ and Γ are the covariance matrices of v, ε_n and γ_n.

The system parameters $\theta = \{A, B, C, D, \Sigma, \Gamma, \mu_0, V_0\}$, can be learned based on historical inputs $u_1, \ldots, u_N$ and outputs $y_1, \ldots, y_N$ using a maximum-likelihood estimator (MLE) [15]. The expectation-maximization (EM) algorithm is a well-known iterative method to find the maximum likelihood of the parameters [16]. The general methodology of the EM algorithm is described in Section 2.5. Therefore, we apply the EM algorithm to derive the posterior estimation of the hidden state variables, $\{\tilde{x}_1, \ldots, \tilde{x}_N\}$, based on which the system parameters θ can be estimated. Such training process is applied iteratively until the convergence condition is met, and future production rates can be predicted given a certain sequence of planned injection rates.

To estimate the parameters $\theta = \{A, B, C, D, \Sigma, \Gamma, \mu_0, V_0\}$ of LDS in Eq. 2.12, we apply EM algorithm to the hidden sequence [10].

E-Step. Kalman Filter uses the past and current observations to predict the current state [17], which works well in estimating the current state. However, it is not sufficient to provide estimation on sequential data, which necessitates Kalman Smoothing [18][19]. The first phase of Kalman Smoothing is to use Kalman Filter to find the posterior marginal for a state x_n given observations $y_1, \ldots, y_n$. Assuming $x_n \sim \mathcal{N}(\mu_n, V_n)$, we have the forward updating phase as follows:

$$\hat{x}_n^- = A\,\hat{x}_{n-1} + B\,u_{n-1} \tag{2.13a}$$

$$\hat{V}_n^- = A\,\hat{V}_{n-1}\,A^\top + \Sigma \tag{2.13b}$$

$$K_n = \hat{V}_n^- C^\top (C\hat{V}_n^- C^\top + \Gamma)^{-1} \tag{2.13c}$$

$$\hat{x}_n = \hat{x}_n^- + K_n(y_n - C\,\hat{x}_n^- - D\,u_n) \tag{2.13d}$$

$$\hat{V}_n = (I - K_n\,C)\,\hat{V}_n^- \tag{2.13e}$$

where K_n is *Kalman Gain Matrix* at time step n, and I is the identity matrix. We let $\hat{x}_n^-$ represent the *prior* estimate at time step n given knowledge of the process prior to time step n, and let $\hat{x}_n$ represent the *posterior* state estimate at time n given the observation y_n. Similarly, $\hat{V}_n^-$ and $\hat{V}_n$ represent the prior and posterior estimates of V_n.

With the initial guess of parameters θ^{old}, Kalman Filter (the forward phase) can generate $\{\hat{x}_1, \ldots, \hat{x}_N\}$ and $\{\hat{V}_1, \ldots, \hat{V}_N\}$. The subscript N represents the finish time of the sequential data. Then, the forward phase can initialize the *backward phase*, which is given by

$$J_n = \hat{V}_n\,A^\top\,(\hat{V}_{n+1}^-)^{-1} \tag{2.14a}$$

$$\tilde{x}_n = \hat{x}_n + J_n(\tilde{x}_{n+1} - \hat{x}_{n+1}^-) \tag{2.14b}$$

$$\tilde{V}_n = \hat{V}_n + J_n(\tilde{V}_{n+1} - \hat{V}_{n+1}^-)J_n^\top \tag{2.14c}$$

where we can find J_n from $A\,\hat{V}_n = \hat{V}_{n+1}^-\,J_n^{-1}$. Moreover, $\hat{V}_{n+1}^-$ and $\hat{x}_{n+1}^-$ can be calculated using

$$\begin{aligned} \hat{x}_{n+1}^- &= A\,\hat{x}_n + B\,u_n \\ \hat{V}_{n+1}^- &= A\,\hat{V}_n\,A^\top + \Sigma \end{aligned} \tag{2.15}$$

Since the posterior estimate has been calculated in the forward phase, the final estimation of the state variable $\tilde{x}_n$ and its covariance $\tilde{V}_n$ can be obtained. Then, the covariance between state x_n and x_{n+1} is

$$\tilde{V}_{n,n+1} = \tilde{V}_{n+1}\,J_n^\top \tag{2.16}$$

M-Step. In the *E-Step*, we have inferred the LDS system states, assuming the model parameters $\theta = \{A, B, C, D, \Sigma, \Gamma, \mu_0, V_0\}$ are known. Here, we wil determine θ with MLE.

Since LDS is based on sequential data, which is recursively determined by the previous state, as shown in Fig. 2.2, we have a key conditional independence property that x_{n-1} and x_{n+1} are independent given x_n. Therefore, the joint distribution of the state variable and observation is given by

$$p(Y, X) = p(x_1)\left[\prod_{n=2}^{N} p(x_n|x_{n-1})\right]\prod_{n=1}^{N} p(y_n|x_n) \tag{2.17}$$

Taking the logarithm of Eq. 2.17, the complete-data log-likelihood function is

$$\begin{aligned}\log p(Y, X|\theta) =& \log p(x_1|\mu_0, V_0) + \sum_{n=2}^{N} \log p(x_n|x_{n-1}, A, B, \Sigma) \\ &+ \sum_{n=1}^{N} \log p(y_n|x_n, C, D, \Gamma)\end{aligned} \tag{2.18}$$

where $\theta = \{A, B, C, D, \Sigma, \Gamma, \mu_0, V_0\}$ is treated as the argument.

We now take the expectation of Eq. 2.27 with respect to the posterior distribution $p(X|Y, \theta^{\text{old}})$, and obtain

$$\begin{aligned}\Theta(\theta, \theta^{\text{old}}) &= \sum_{X} Q(X) \log p(Y, X|\theta) \\ &= \mathbb{E}_{x\in Q}[\log p(Y, X|\theta)]\end{aligned} \tag{2.19}$$

which also defines a lower bound of Eq. 2.27. Then, this function is maximized with respect to θ.

According to the *multivariate Gaussian distribution* [10], the following expectations will be required to calculate the log-likelihood:

$$\mathbb{E}\left[x_n\right] = \tilde{x}_n \tag{2.20a}$$

$$\mathbb{E}\left[x_n x_{n-1}^{\top}\right] = J_{n-1}\tilde{V}_n + \tilde{x}_n\tilde{x}_{n-1}^{\top} \tag{2.20b}$$

$$\mathbb{E}\left[x_n x_n^\top\right] = \tilde{V}_n + \tilde{x}_n \tilde{x}_n^\top \tag{2.20c}$$

As an example, let us consider the parameters μ_0 and V_0 first. If we take expectation with respect to $X \propto Q(X)$, we can obtain

$$\begin{aligned}\Theta(\mu_0, V_0) = &-\frac{1}{2}\log|V_0| - \mathbb{E}\Big[\frac{1}{2}(x_1-\mu_0)^\top V_0^{-1}(x_1-\mu_0)\Big] \\ &+ \text{const}\end{aligned} \tag{2.21}$$

where all the terms that are not dependent on μ_0 and V_0 are incorporated into the const term. Then, it is straightforward to find the optimal μ_0 and V_0 based on the maximum-likelihood solution for Gaussian distribution:

$$\mu_0^{\text{new}} = \mathbb{E}\left[x_1\right] \tag{2.22a}$$

$$V_0^{\text{new}} = \mathbb{E}\left[x_1 x_1^\top\right] - \mathbb{E}\left[x_1\right]\mathbb{E}\left[x_1^\top\right] \tag{2.22b}$$

We can find the optimal values of the other parameters A, B, C, D in a similar fashion and obtain

$$\begin{aligned}[A\ B] = &\begin{pmatrix}\sum \mathbb{E}\left[x_n x_{n-1}^\top\right] & \sum \mathbb{E}\left[x_{n-1}\right] u_{n-1}^\top\end{pmatrix} \cdot \\ &\begin{pmatrix}\sum \mathbb{E}\left[x_{n-1} x_{n-1}^\top\right] & \sum \mathbb{E}\left[x_{n-1}\right] u_{n-1}^\top \\ \sum u_{n-1}\mathbb{E}\left[x_{n-1}^\top\right] & \sum u_{n-1} u_{n-1}^\top\end{pmatrix}^{-1}\end{aligned} \tag{2.23}$$

$$\begin{aligned}[C\ D] = &\begin{pmatrix}\sum y_n \mathbb{E}\left[x_n^\top\right] & \sum y_n u_n^\top\end{pmatrix} \cdot \\ &\begin{pmatrix}\sum \mathbb{E}\left[x_n x_n^\top\right] & \sum \mathbb{E}\left[x_n\right] u_n^\top \\ \sum u_n \mathbb{E}\left[x_n^\top\right] & \sum u_n u_n^\top\end{pmatrix}^{-1}\end{aligned} \tag{2.24}$$

based on which, Σ, Γ can be found with the newly calculated parameters A, B, C, D. That is,

$$\Sigma = \frac{1}{N-1}\sum_{n=2}^{N} \mathbb{E}[(x_n - Ax_{n-1} - Bu_{n-1})(x_n - Ax_{n-1} - Bu_{n-1})^\top] \tag{2.25}$$

$$\Gamma = \frac{1}{N}\sum_{n=1}^{N} \mathbb{E}\left[(y_n - Cx_n - Du_n)(y_n - Cx_n - Du_n)^\top\right] \tag{2.26}$$

We can optimize the system parameters by iteratively applying the E-step and the M-step mentioned above.

Once we have obtained the posterior estimation of the hidden state variables, $\{\tilde{x}_1, \ldots, \tilde{x}_N\}$, and their covariance matrices, $\{\tilde{V}_1, \ldots, \tilde{V}_N\}$, as well as the system parameters from the EM algorithm, the prior estimate [20][10] of Kalman Filter could be applied to predict the system outputs $y_{N+1}, y_{N+2}, \ldots$ and their variances $W_{N+1}, W_{N+2}, \ldots$ for the future, as described in Algorithm 2.1.

Algorithm 2.1 Reconstructed and Predicted Production Rates

Input parameters $\{A, B, C, D\}$ and $\{\tilde{x}_1, \cdots \tilde{x}_N\}$
for $n = 1$ to N **do**
    $\tilde{y}_n = C\,\tilde{x}_n + D\,u_n$
Set $x_N^- = \tilde{x}_N$
for $n = N+1, N+2, \ldots$ **do**
    $x_n^- = A\,x_{n-1}^- + B\,u_{n-1}$
    $y_n = C\,x_n^- + D\,u_n$
    $V_n = A\,V_{n-1}\,A^\top + \Sigma$
    $W_n = C\,V_n\,C^\top + \Gamma$
Output the historical output reconstruction $\{\tilde{y}_1, \ldots \tilde{y}_N\}$
and the prediction $\{y_{N+1}, y_{N+2}, \ldots\}$

2.5 The EM Algorithm

In order to present the specific EM algorithm applied to our sequential data and the LDS in Eq. 2.12, we first describe the general EM algorithm [11][21], which is widely used to calculate the maximum likelihood of observed data $p(y|\theta)$ in the presence of a hidden variable x. It alternates between an *E-Step* and an *M-Step*.

In the *E-Step*, we first generate the expectation of the log-likelihood evaluated at the current estimated parameters, while in the *M-Step*, we update the parameters by maximizing the expected log-likelihood obtained in the *E-Step*. Then, the parameters obtained in the *M-Step* are used to create the new expectation in the next *E-Step*.

In the following, we briefly describe how to determine the expectation of the log-likelihood function. For simplicity, we denote $Y = \{y_1, \dots y_n, \dots\}$, $X = \{x_1, \dots x_n, \dots\}$. The log-likelihood function is given by

$$\log p(Y|\theta) = \log \int_X p(Y, X|\theta) \mathrm{d}X \tag{2.27}$$

In a discrete form, assigning any probability distribution $Q(X)$ over hidden variables X and applying Jensen's inequality due to the concavity of $\log$ function, we can obtain a lower bound of the log-likelihood function:

$$\begin{aligned}
\log p(Y|\theta) &= \log \left\{ \sum_X p(Y, X|\theta) \right\} \\
&= \log \left\{ \sum_X Q(X) \frac{p(Y, X|\theta)}{Q(X)} \right\} \\
&= \log \left\{ \mathbb{E}_{X \in Q} \frac{p(Y, X|\theta)}{Q(X)} \right\} \\
&\geqslant \mathbb{E}_{X \in Q} \left\{ \log \frac{p(Y, X|\theta)}{Q(X)} \right\} \\
&= \sum_X Q(X) \log \frac{p(Y, X|\theta)}{Q(X)}
\end{aligned} \tag{2.28}$$

To maximize the lower bound, $p(Y, X|\theta)/Q(X)$ should be a constant value independent of θ. Therefore, we have

$$Q(X) \propto p(Y, X|\theta) \tag{2.29}$$

Considering $\sum_X Q(X) = 1$, we have

$$\begin{aligned}
Q(X) &= \frac{p(Y, X|\theta)}{\sum_X p(Y, X|\theta)} = \frac{p(Y, X|\theta)}{p(Y|\theta)} \\
&= p(X|Y, \theta)
\end{aligned} \tag{2.30}$$

To summarize, in the *E-Step*, we calculate the conditional distribution of X given Y under the current estimate of the parameters θ (the lower bound of the log-likelihood). Then, in the *M-Step*, we aim to find the parameters θ that maximize the

lower bound

$$\begin{aligned} &\sum_X Q(X) \log \frac{p(Y, X|\theta)}{Q(X)} \\ =&\sum_X Q(X) \log p(Y, X|\theta) - \sum_X Q(X) \log Q(X) \end{aligned} \tag{2.31}$$

where we only need to maximize the first term in the RHS of Eq. 2.31, since the second term does not depend on θ. The general EM algorithm is shown in Algorithm 2.2 [10].

Algorithm 2.2 EM Algorithm

1: Initialize parameters θ^{old}
2: **for** $k = 1$ to maxIter **do**
3: *E-Step* Evaluate $Q(X) := p(X|Y, \theta^{\text{old}})$
4: *M-Step* Evaluate θ^{new}, given by
$\theta^{\text{new}} := \arg\max_\theta \sum_X Q(X) \log p(Y, X|\theta^{\text{old}})$
5: Check convergence of either log-likelihood or θ
6: Output parameters θ

Chapter 3

Performance Evaluation

We validate the newly developed model by proving that the production rates predicted by the new model are close to the real production rates. We compare this new model against CM which is widely used in industry [4] in terms of prediction accuracy for different types of reservoirs under a wide range of parameter settings.

3.1 Experimental Setup

We will conduct performance evaluation on three synthetic reservoir models built with a commercial reservoir simulator, CMG, as shown in Fig. 3.1, 3.2 and 3.3, respectively. The first synthetic model represents a heterogeneous high-permeability reservoir, while the second and the third models represent homogeneous low-permeability reservoirs of smaller and larger scales, respectively. The regular five-spot well pattern is adopted in all the three synthetic models. Both the first and second reservoirs are 5×4, employing 5 injectors and 4 producers, while the third reservoir is 10×10, employing 10 injectors and 10 producers. We also evaluate our proposed scheme based on the production data collected from a real oilfield, as shown in Fig. 3.4. The detailed setup of the these test cases are described as follows:

- **High-Permeability Reservoir Model (HPRM)**: the model (5×4) has a size of $50\,\text{m} \times 50\,\text{m} \times 1\,\text{m}$ and contains two high-permeability channels, as shown

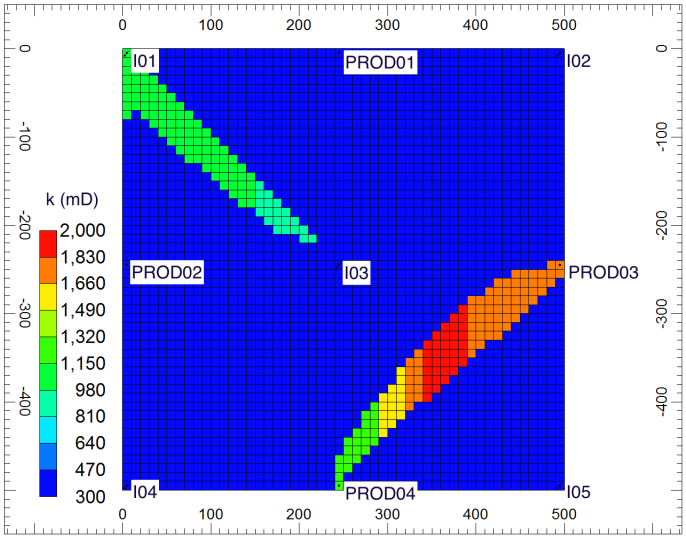

Fig. 3.1. The synthetic reservoir model, HPRM, comprised of 5 injectors and 4 producers (5×4), representing a heterogeneous high-permeability reservoir with two channels. The heat map quantifies the permeability distribution (mD).

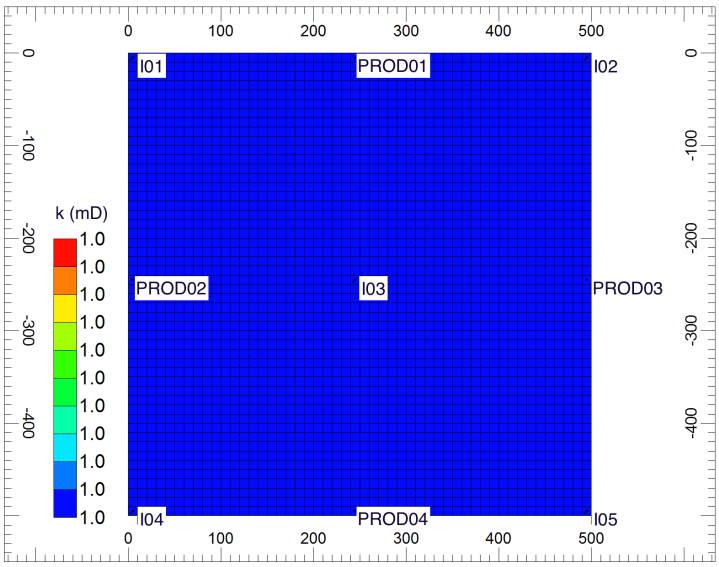

Fig. 3.2. The synthetic reservoir model, LPRM, comprised of 5 injectors and 4 producers (5×4), representing a homogeneous low-permeability reservoir. The permeability of this reservoir is 1mD.

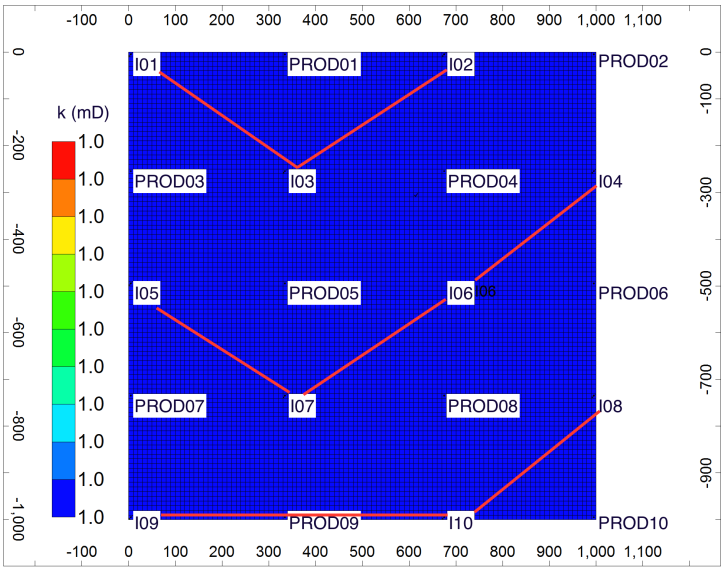

Fig. 3.3. The synthetic reservoir model, LLPRM, comprised of 10 injectors and 10 producers (10×10), representing a homogeneous low-permeability reservoir. The permeability of this reservoir is 1mD. Each red line represents a set of injectors at a certain distance from PROD01.

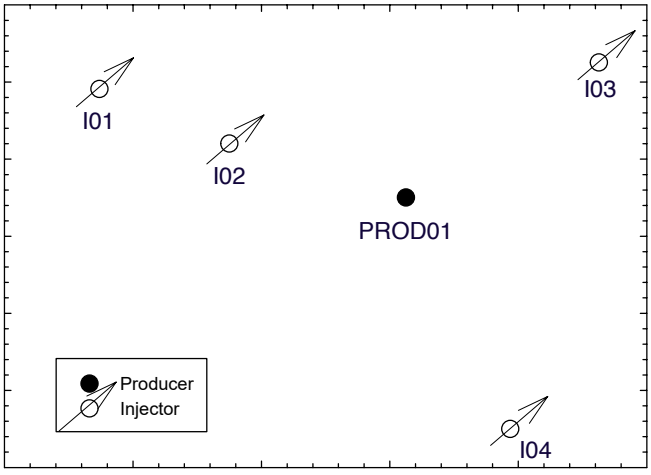

Fig. 3.4. The well patterns and relative locations of one producer and its most adjacent injectors in a waterflooding project in Changqing oilfield.

in Fig. 3.1. We set the BHP of all the producers to 689.4757 kPa. The permeability of the areas excluding the channel is set to 300 mD, while the permeability of the two channels is set to 800-2000 mD.

- **Low-Permeability Reservoir Model (LPRM)**: this homogeneous model (5×4) has a size of $50\,\text{m} \times 50\,\text{m} \times 1\,\text{m}$, as shown in Fig. 3.2. The permeability is set to 1 mD, and the BHP of all the producers is set to 689.4757 kPa. For this model, we consider three scenarios: 1) the scenario without varying BHP and well shut-in; 2) the scenario with varying BHIP; 3) the scenario with injector shut-ins.

- **Larger-scale Low-Permeability Reservoir Model (LLPRM)**: this homogeneous model (10×10) has a size of $100\,\text{m} \times 100\,\text{m} \times 1\,\text{m}$, as shown in Fig. 3.3. The permeability is set to 1 mD, and the BHP of producers is set to 689.4757 kPa.

- **Real Field Data Test**: we also evaluate the performance of LDS on the real data collected from the Changqing oilfield, which is located in northwest China and contains typical low-permeability production formations. Waterflooding is applied to a small well pattern comprised of 4 injectors and 1 producer, as shown in Fig. 3.4. The target reservoir depth is between 2018–2793 m and the permeability is distributed in the range of 0.7–58.7 mD.

3.2 Results

3.2.1 HPRM and LPRM

We first evaluate the performance of LDS in comparison to CM when applied to HPRM and LPRM, in terms of prediction accuracy. All the numerical simulations are executed over a period of 96 months, with a time interval $\Delta t = 1$ month. We use the first 3/4 data (first 72 months) as the training data, while the remaining data serve as the test data to validate the prediction accuracy of each method. We treat

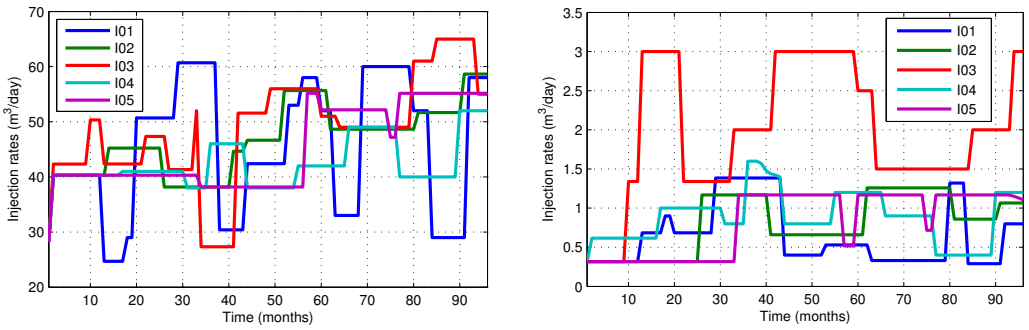

Fig. 3.5. The injection rates used for HPRM, as shown in Fig. 3.1 (left sub-figure), and for LPRM, as shown in Fig. 3.2 (right sub-figure). The first 72 months' injection rates data are used as input of training data, and the rest 24 months' injection rates data are used as input to validate the prediction results.

TABLE 3.1
RELATIVE ERRORS OF THE PREDICTED PRODUCTION RATES FOR HPRM AS SHOWN IN FIG. 3.1.

Method	Type	Relative Error (%)			
		PROD01	PROD02	PROD03	PROD04
LDS	Mean	1.7453	0.9899	1.2744	1.1870
CM	Mean	0.3830	0.4948	0.5154	0.8805
LDS	Max	3.1542	2.1641	2.1928	1.9677
CM	Max	1.1928	1.2933	0.8721	1.6175

Note: the relative error is the absolute difference between the predicted and real production rates divided by the real production rates.

each production well as being independent of other production wells, while being influenced by all the injection wells. We terminate LDS algorithm if the difference in the log-likelihood between two iterations is smaller than 10^{-4}. In order to avoid overfitting, we set the maximum number of iterations to 200.

Fig. 3.5 shows all the injection rates for HPRM and LPRM. The production rates predicted by LDS and CM are plotted in Fig. 3.6 and Fig. 3.7, respectively, for HPRM and LPRM, in comparison to the true production rates given by CMG. Comparing to CM, we observe that LDS can significantly improve the prediction accuracy for the low-permeability reservoir, whereas it exhibits a prediction accuracy similar to CM for the high-permeability reservoir.

The mean and maximum relative prediction errors are shown in Table 3.1 and

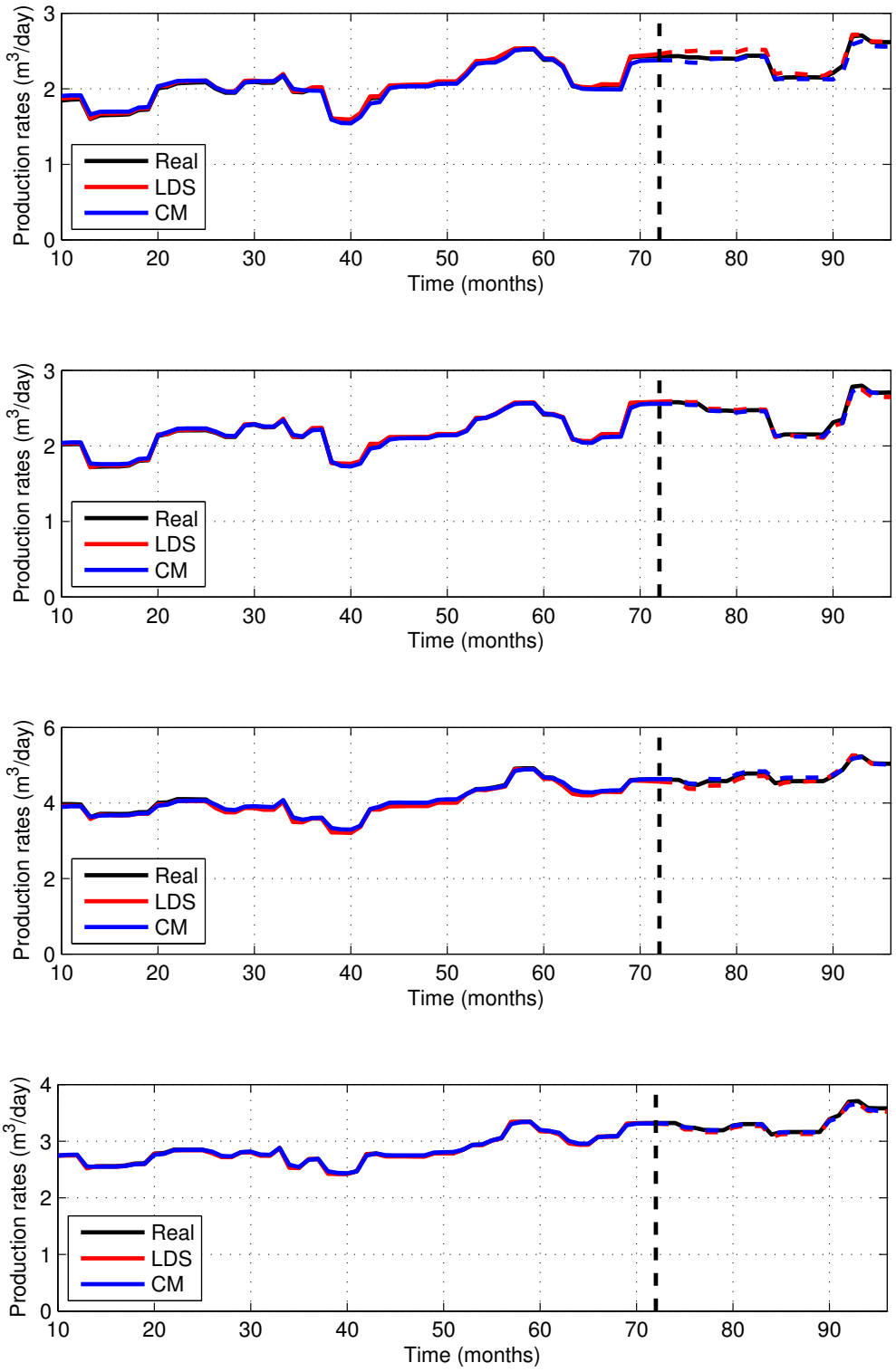

Fig. 3.6. Performance comparison of LDS and CM when they are applied to HPRM as shown in Fig. 3.1. Each sub-plot compares the actual production rates for each producer and calculated ones by using LDS and CM, for PROD01, PROD02, PROD03 and PROD04, respectively. The period from 1st to 72th months corresponds to the reconstruction stage, while the period from 73th to 96th months corresponds to the prediction stage.

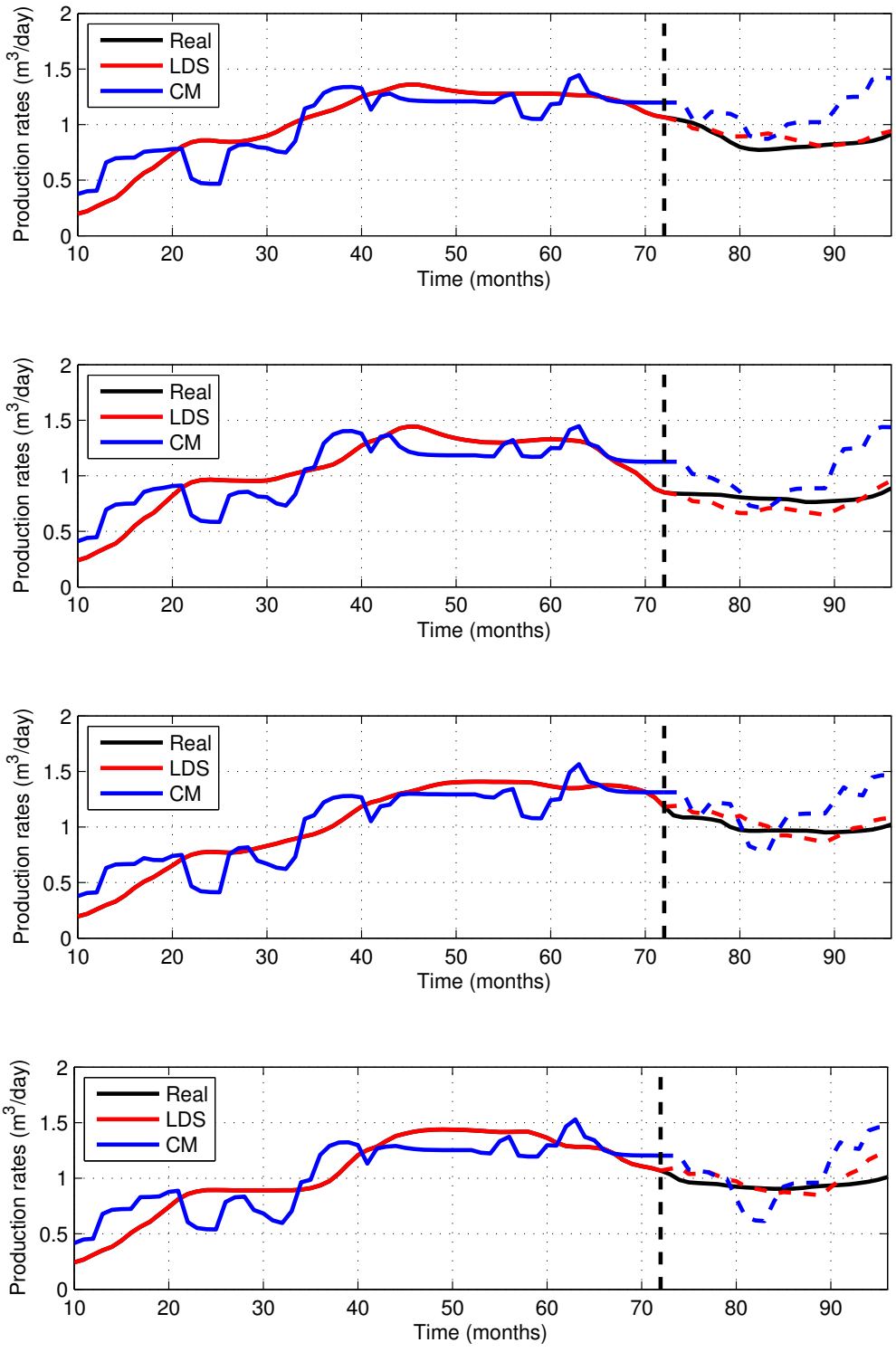

Fig. 3.7. Performance comparison of LDS and CM when they are applied to LPRM as shown in Fig. 3.2. Each sub-plot compares the actual production rates for each producer and calculated ones by using LDS and CM, for PROD01, PROD02, PROD03 and PROD04, respectively. The period from 1st to 72th months corresponds to the reconstruction stage, while the period from 73th to 96th months corresponds to the prediction stage.

TABLE 3.2
RELATIVE ERRORS OF THE PREDICTED PRODUCTION RATES FOR LPRM AS SHOWN IN FIG. 3.2.

Method	Type	Relative Error (%)			
		PROD01	PROD02	PROD03	PROD04
LDS	Mean	7.3159	8.1447	9.2772	8.3129
CM	Mean	29.2854	28.6264	21.3266	20.6699
LDS	Max	13.7871	17.3040	18.9399	23.0701
CM	Max	64.6709	73.5282	48.0139	47.7024

Note: the relative error is the absolute difference between the predicted and real production rates divided by the real production rates.

TABLE 3.3
β MATRIX TRAINED BY LDS FOR HPRM AS SHOWN IN FIG.3.1.

Injector	Producer			
	PROD01	PROD02	PROD03	PROD04
I1	0.3559	0.3467	0.1523	0.1298
I2	0.3581	0.1520	0.3778	0.1114
I3	0.2633	0.2467	0.2458	0.2182
I4	0.2122	0.3148	0.1688	0.2770
I5	0.1533	0.1441	0.2573	0.4687

Table 3.2, for HPRM and LPRM, respectively. It can be observed that as compared to CM, LDS substantially reduces both the mean and maximum relative prediction errors for LPRM.

Furthermore, we explore why LDS could provide better prediction than CM in LPRM. Recall that β matrix implies the direct connection between injection rates and production rates, which resembles the interwell connectivity in CM. We show the learned β matrices in LDS for HPRM and LPRM in Table 3.3 and Table 3.4, respectively. For HPRM, the geographically closer injector-producer pairs will have larger values of β, conforming to what we have expected. On the other hand, for LPRM, the elements in β matrix have relatively small values, as compared to HPRM. The underlying reason is that the fluid flow from each injector to a producer in tight formations is much more difficult than that in high-permeability formations.

TABLE 3.4
β MATRIX TRAINED BY LDS FOR LPRM AS SHOWN IN FIG.3.2.

Injector	Producer			
	PROD01	PROD02	PROD03	PROD04
I1	0.0292	0.0126	0.0506	0.0210
I2	0.0197	0.0292	0.0370	0.0292
I3	0.0281	0.0404	0.0474	0.0568
I4	0.0469	0.0821	0.0759	0.0969
I5	0.0842	0.1118	0.1224	0.1355

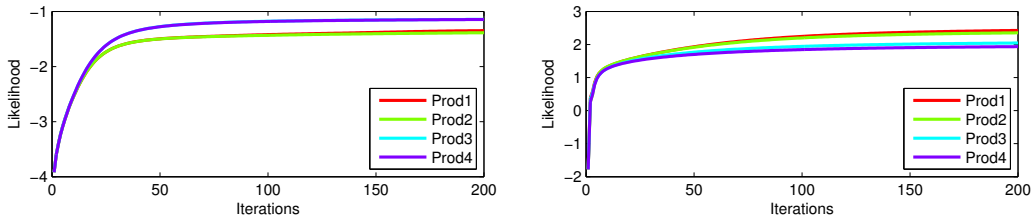

Fig. 3.8. Evolution of the log-likelihood for the 4 producers over the iterations: (a) HPRM; (b) LPRM.

In EM training process, the log-likelihood function is maximized to find out the best-fit parameters θ. Fig. 3.8 shows the evolution of the log-likelihood function, i.e., the objective function in the EM algorithm, as the iteration proceeds. From Fig. 3.8, we can observe that the convergence is faster for HPRM than for LPRM.

Prediction intervals can be used to quantify the uncertainty surrounding the prediction of each individual production rate at a future point. Since we have the variances of state variables, we can predict the variance of production rates based on Algorithm 2.1. Let σ_n denote the standard deviation of the predicted production rate at time n. With Gaussian noise, the 95% prediction interval is $[y_n - z\sigma_n, y_n + z\sigma_n]$, where $z = 1.96$ [22]. Even with non-Gaussian error terms and production rate distribution, the 95% prediction interval can be well approximated by the same value above if we have observed a large number of training samples from the history. We plot the 95% prediction intervals for HPRM and LPRM in Fig. 3.9 and Fig. 3.10, respectively.

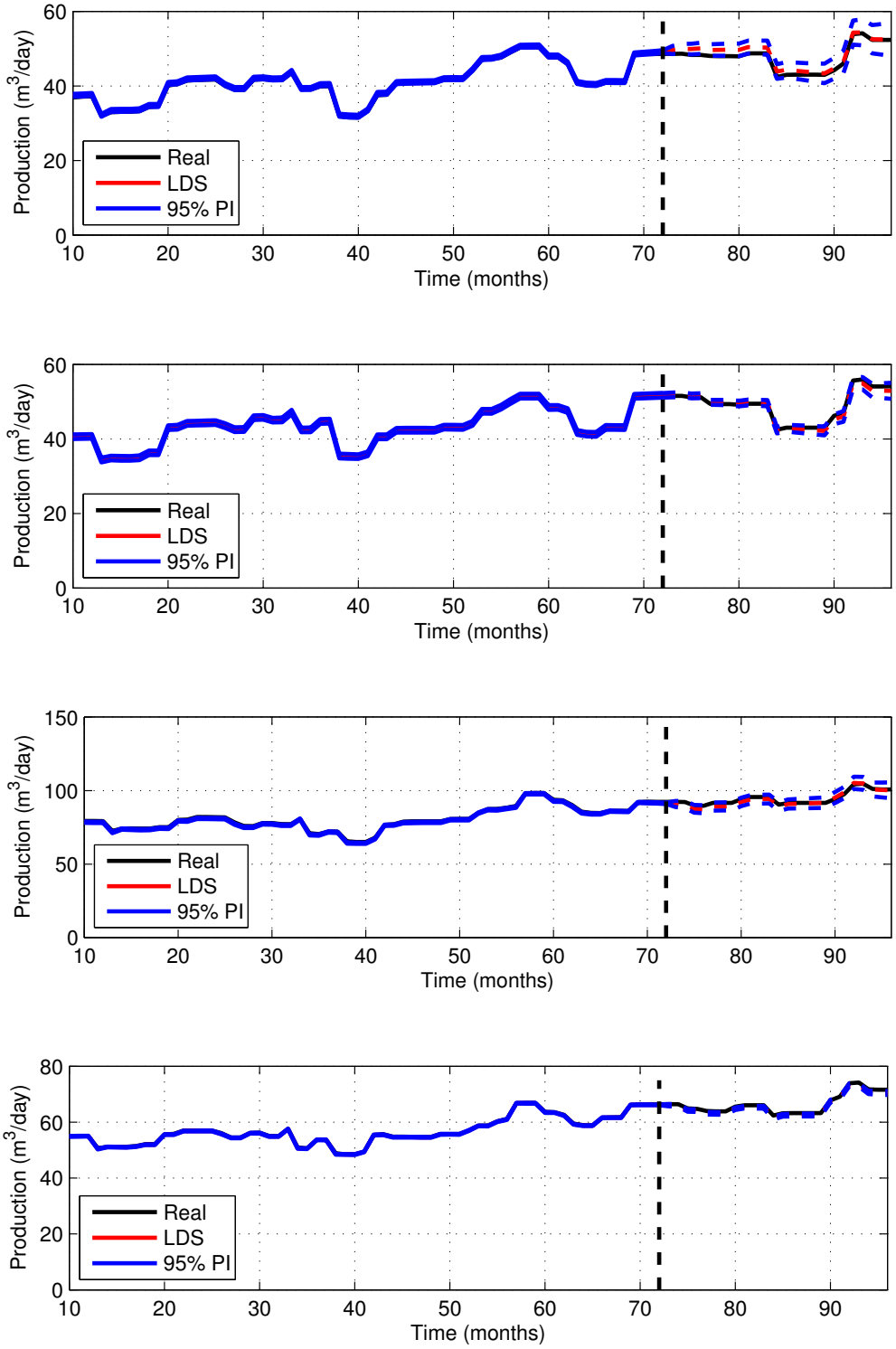

Fig. 3.9. Performance of LDS and its 95% prediction interval when they are applied to HPRM as shown in Fig. 3.1. Each sub-plot compares the actual production rates for each producer and predicted ones by using LDS, for PROD01, PROD02, PROD03 and PROD04, respectively.

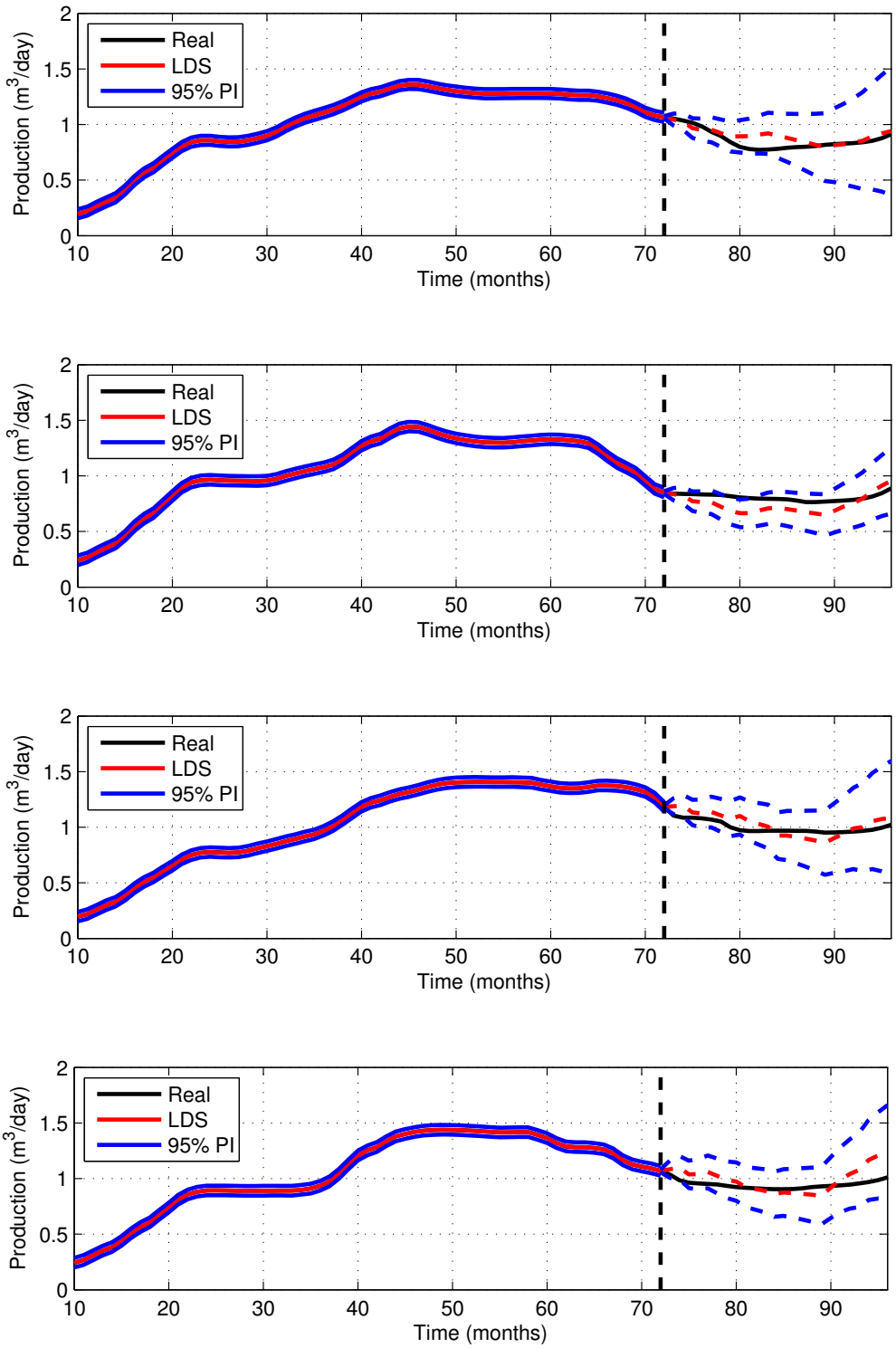

Fig. 3.10. Performance of LDS and its 95% prediction interval when they are applied to LPRM as shown in Fig. 3.2. Each sub-plot compares the actual production rates for each producer and predicted ones by using LDS, for PROD01, PROD02, PROD03 and PROD04, respectively.

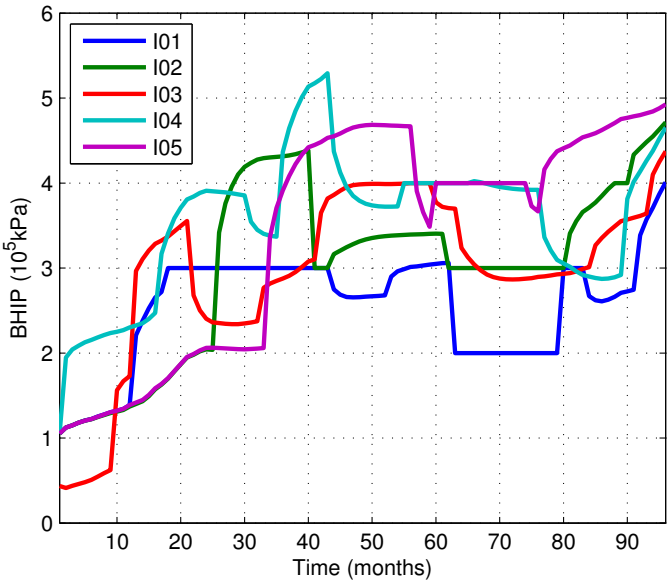

Fig. 3.11. The variation pattern of BHIP in LPRM as shown in Fig. 3.2.

3.2.2 LPRM with Varying BHIP

BHIP is the downhole pressure at which we inject treatment fluid into the injectors. In the following scenarios, we apply LDS to LPRM in comparison to CM, when BHIP varies. We design the variation pattern of BHIP for different injectors, as shown in Fig. 3.11. For this scenario, we use the same injection rates as the previous LPRM, as shown in Fig. 3.5 (right).

Fig. 3.12 shows the reconstructed production rates as well as the predicted ones for LPRM under varying BHIP. First, we compare LDS with CM in the reconstruction stage, i.e., 1st to 72th month, as shown in Fig. 3.7 and Fig. 3.12. We can observe that the reconstructed production rates in LDS can almost exactly match the real production rates, with no fluctuations. In contrast, the reconstructed production rates of CM cannot well approximate the real production rates, yielding larger fluctuations. This is attributed to the fact that CM cannot well capture the larger variation of BHIP [4].

For the prediction stage, due to varying BHIP, we can see from Fig. 3.12 that CM leads to even larger deviation from the real production rates in its prediction.

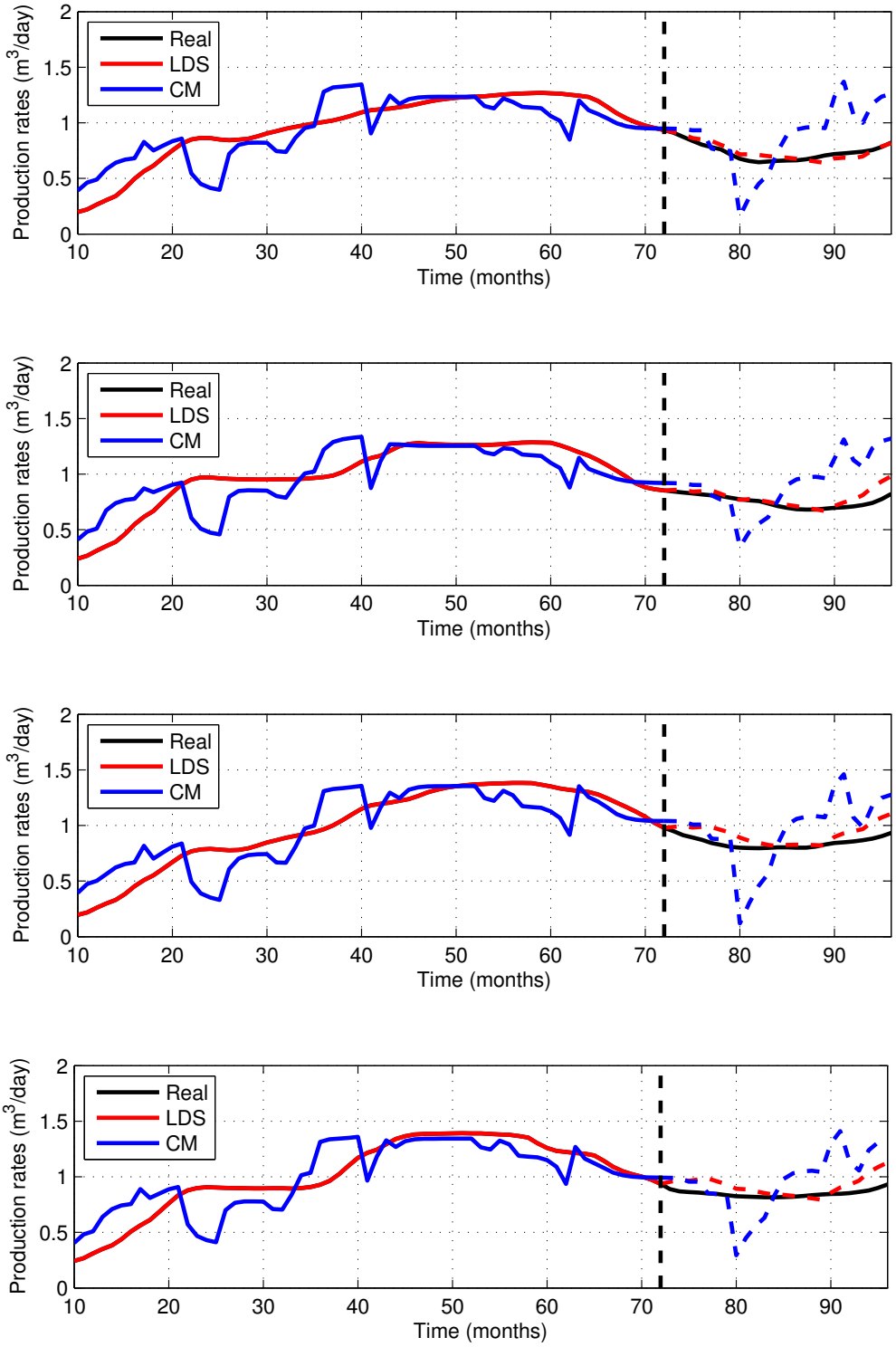

Fig. 3.12. Performance comparison of LDS and CM when they are applied LPRM as shown in Fig. 3.2. In this example, BHIP vary according to Fig. 3.11. Each sub-plot compares the actual production rates for each producer and calculated ones by using LDS and CM, for PROD01, PROD02, PROD03 and PROD04, respectively. The period from 1st to 72th months corresponds to the reconstruction stage, while the period from 73th to 96th months corresponds to the prediction stage.

TABLE 3.5
RELATIVE ERRORS OF THE PREDICTED PRODUCTION RATES FOR LPRM WITH VARYING BHIP, AS SHOWN IN FIG. 3.11.

Method	Type	Relative Error (%)			
		PROD01	PROD02	PROD03	PROD04
LDS	Mean	5.0186	5.4894	9.4581	9.2786
CM	Mean	35.2317	35.1791	30.8553	27.5190
LDS	Max	10.3932	20.9146	19.1693	21.5035
CM	Max	87.3713	89.0316	86.4592	66.3168

Note: the relative error is the absolute difference between the predicted and real production rates divided by the real production rates.

However, LDS can still provide prediction performance similar to that in LPRM without the varying BHIP, which has been shown in Fig. 3.7 and Fig. 3.12. Moreover, Table 3.5 summarizes the relative prediction errors of LDS and CM for the varying BHIP scenario. The maximum prediction error across all the months considered is 89.03% for CM, while it is only 21.50% for LDS, indicating the obvious superiority of LDS in terms of modeling waterflooding in low-permeability reservoirs with varying BHIP.

3.2.3 LPRM with Injector Shut-ins

Considering the low production rates of low-permeability reservoirs, it is possible for operators to suspend water injection at certain wells from time to time, which is often referred to as "well shut-in" in practice. In our model, since the injection rates serve as the system inputs, well shut-in can be readily handled with LDS by setting the injection rates at the corresponding shut-in injector to zero.

We consider two scenarios: 1) injectors I02 and I04 are shut in at the 44th month, which is in the middle of all 96 months considered; 2) all injectors except for I03 are shut in at the 44th month. Fig. 3.13 shows the corresponding injection rates in these two scenarios. The first 72 months will serve as the training stage, based on which predictions on production rates are made for the following 24 months.

Fig. 3.14 and Fig. 3.15 show the reconstructed and predicted production rates for the two scenarios with injector shut-ins, where the benefit of LDS over CM is

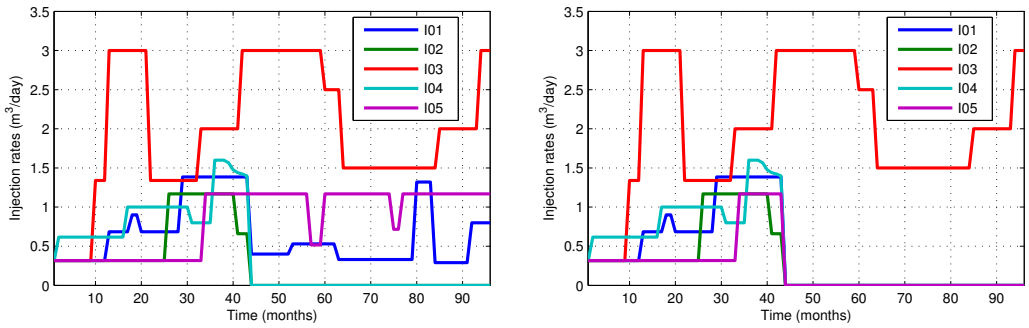

Fig. 3.13. The injection rates used for LPRM as shown in Fig. 3.2. Some injectors are shut in from the 44th month to test the prediction of production rates under the injectors shut-in scenario. Left sub-figure is the scenario where injectors I02, I04 are shut in, and the right sub-figure is the scenario where all injectors but I03 are shut in.

TABLE 3.6
RELATIVE ERRORS OF THE PREDICTED PRODUCTION RATES FOR LPRM, WHERE INJECTOR I02 AND I04 ARE SHUT IN, AS SHOWN IN FIG. 3.13 (LEFT).

Method	Type	Relative Error (%)			
		PROD01	PROD02	PROD03	PROD04
LDS	Mean	3.2975	5.5955	10.8349	6.3597
CM	Mean	69.8319	62.0303	40.8823	57.7797
LDS	Max	7.9852	16.7926	25.6571	14.2984
CM	Max	127.9042	127.3063	65.6108	101.0536

Note: the relative error is the absolute difference between the predicted and real production rates divided by the real production rates.

obviously observed. In both reconstruction and prediction stages, CM cannot well approximate the real production rates, while LDS can match the production rates much more closely.

Table 3.6 and Table 3.7 show the prediction errors of LDS and CM. We can see that with well shut-ins, the relative prediction errors of CM are much larger than those of LDS. Furthermore, comparing the two tables, we can find that when more injectors are shut in, the relative prediction errors of CM increase significantly, while the prediction errors of LDS only increase slightly. Thus, LDS can still maintain an acceptable prediction accuracy in the cases of well shut-ins.

The reason underlying the superiority of LDS to CM is that LDS can capture the changing dynamics of control volumes in the reservoir after the well shut-ins, while

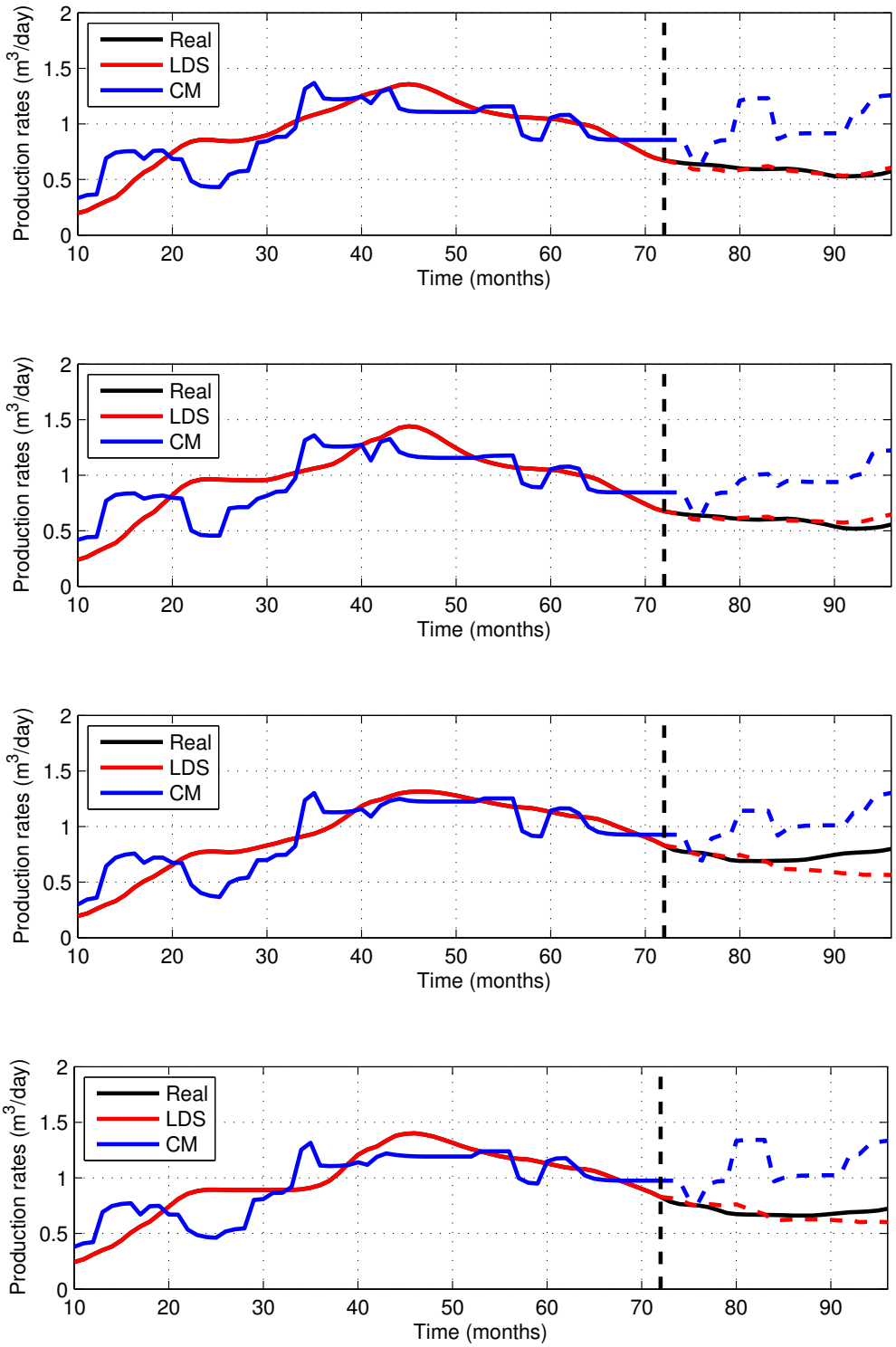

Fig. 3.14. Performance comparison of LDS and CM when they are applied to LPRM as shown in Fig. 3.2. In this example, the injectors $I02$ and $I04$ are both shut in at the 44th month. Each sub-plot compares the actual production rates for each producer and calculated ones by using LDS and CM, for PROD01, PROD02, PROD03 and PROD04, respectively. The period from 1st to 72th months corresponds to the reconstruction stage, while the period from 73th to 96th months corresponds to the prediction stage.

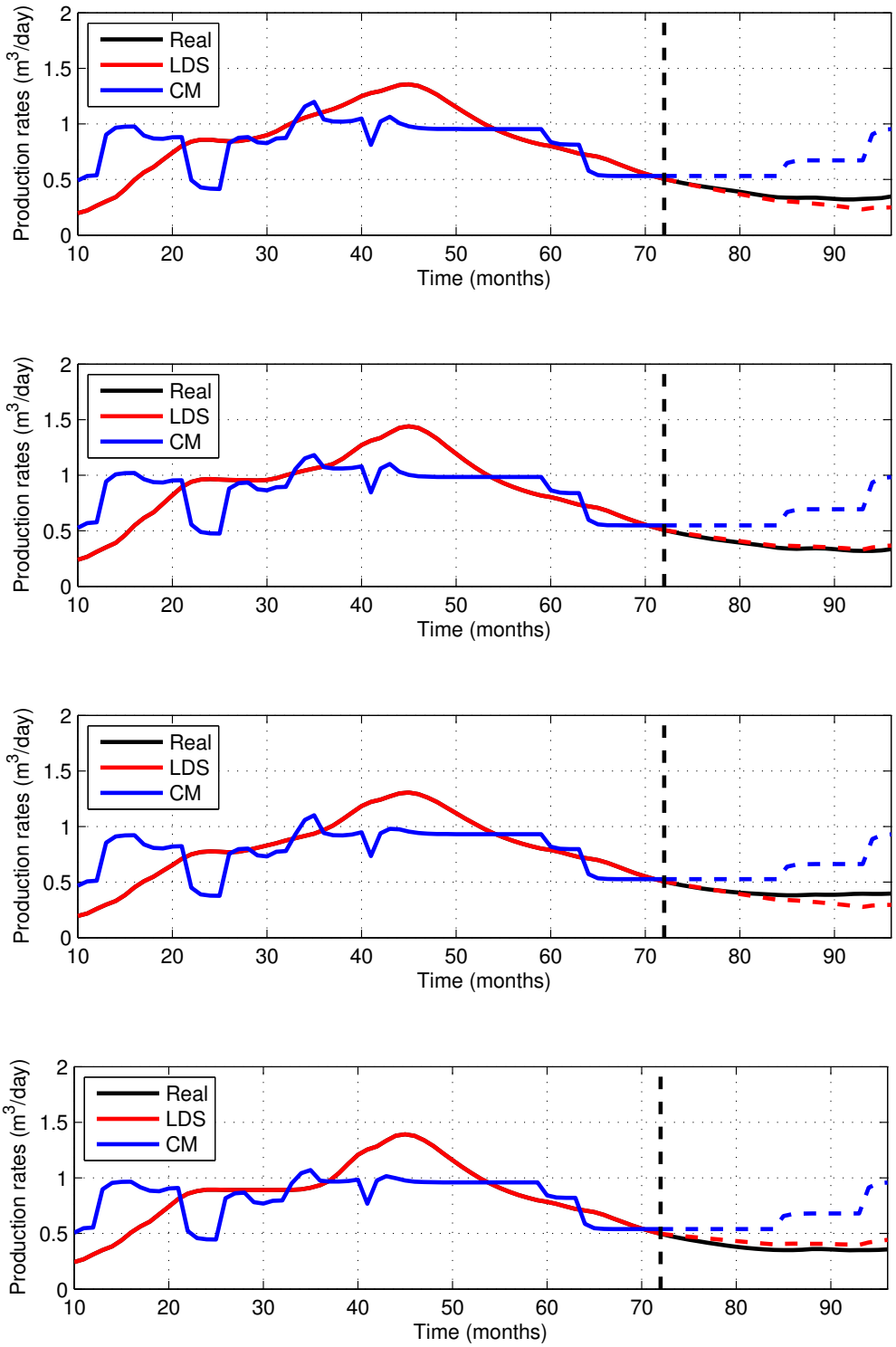

Fig. 3.15. Performance comparison of LDS and CM when they are applied to LPRM as shown in Fig. 3.2. In this example, the injectors $I01$, $I02$, $I04$ and $I05$ are all shut in at the 44th month. Each sub-plot compares the actual production rates for each producer and calculated ones by using LDS and CM, for PROD01, PROD02, PROD03 and PROD04, respectively. The period from 1st to 72th months corresponds to the reconstruction stage, while the period from 73th to 96th months corresponds to the prediction stage.

TABLE 3.7
RELATIVE ERRORS OF THE PREDICTED PRODUCTION RATES FOR LPRM, WHERE ALL INJECTORS EXCEPTION FOR I03 ARE SHUT IN, AS SHOWN IN FIG. 3.13 (RIGHT).

Method	Type	Relative Error (%)			
		PROD01	PROD02	PROD03	PROD04
LDS	Mean	11.9900	4.7310	12.2951	13.2868
CM	Mean	76.8022	81.7842	55.1571	72.5148
LDS	Max	28.6283	11.4820	29.9701	23.7883
CM	Max	183.6089	201.5525	134.7882	170.2932

Note: the relative error is the absolute difference between the predicted and real production rates divided by the real production rates.

CM cannot reflect such dynamics by assuming constant interwell connectivities.

3.2.4 LLPRM

We further test our method on a larger 10×10 LLPRM field as shown in Fig. 3.3. As the field is larger, the validation is done under a two-step procedure: 1) selecting the most relevant injectors for each producer, and 2) predicting the production rates of each producer only based on the most relevant injectors.

The rationale for selecting the most relevant injectors is the following. When there are more wells involved in a low-permeability reservoir, the injectors of different distances from a given producer will have significantly different impact to the producer: the farther the injector is away, the less influence it will have on the producer. In this case, if we train LDS with random initialization of β, which essentially assumes similar contributions from different injectors towards each producer, extra noise will be introduced into the model training, since injectors far away actually have little influence on the producer being considered.

To verify this idea, we compare the prediction accuracy for a particular producer PROD01. Fig. 3.3 depicts the injectors with different distances from PROD01: each distance range is represented by a different red line. The validation is under three strategies: 1) predict the production rates for PROD01 based on its immediately adjacent injectors: I01, I02 and I03; 2) predict the production rates for PROD01 based on I04, I05, I06 and I07 in addition to I01, I02 and I03; 3) predict the production

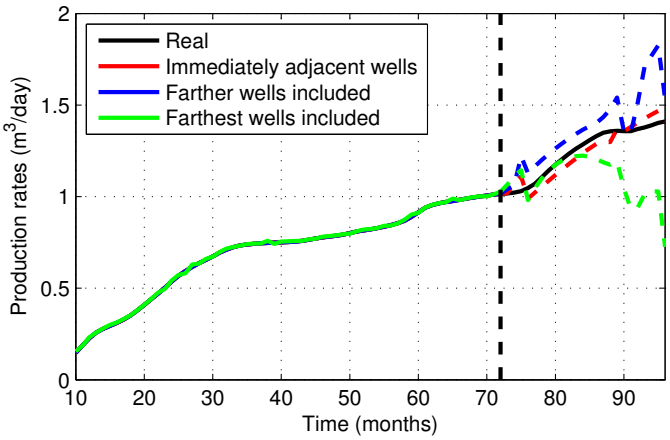

Fig. 3.16. Comparison between the actual production rates for producer PROD01 and calculated ones by using LLPRM under three different scenarios. These three scenarios are developed based on the injection data collected for the three groups of injectors, respectively: injectors located at the 1st red line, injectors located at the 1st and 2nd red lines, and injectors located at the 1st, 2nd and 3rd red lines (See Fig. 3.3). The period from 1st to 72th months corresponds to the reconstruction stage, while the period from 73th to 96th months corresponds to the prediction stage.

rates for PROD01 based on all 10 injectors.

The results are shown in Fig. 3.16, which conforms to our observation mentioned above: the prediction accuracy decreases as farther injectors are included, since the farther injectors do not have noticeable contributions to the producer, and may only introduce noise to the training process. Similar tests have been performed on other producers, from which we have observed a similar phenomenon. Therefore, in the following, we predict the production rates of each producer in the reservoir only based on its immediately adjacent injectors.

The total time intervals are also 96 months, and the first 72 months are utilized for training and the rest are used to test the prediction performance. Fig. 3.17 plots the reconstructed and predicted production rates for all 10 producers using both LDS and CM. We observe that the prediction error of LDS can be controlled within a small range by our injector selection strategy based on immediately adjacent injectors.

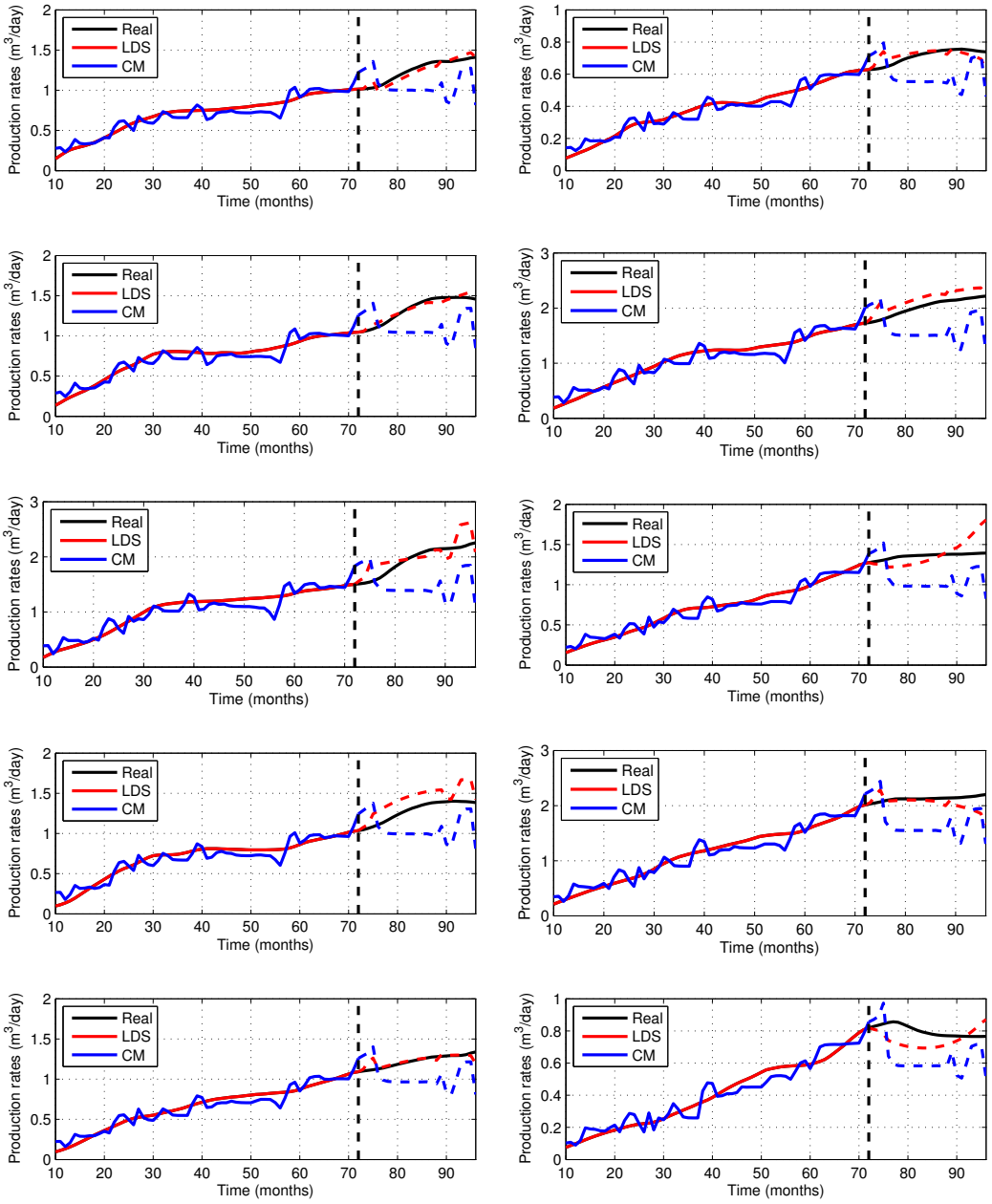

Fig. 3.17. Performance comparison of LDS and CM when they are applied to LLPRM as shown in Fig. 3.3. In this example, only the immediately adjacent injectors are considered for each producer. Each sub-plot compares the actual production rates for each producer and calculated ones by using LDS and CM, representing from PROD01 to PROD10 in order, respectively. The period from 1st to 72th months corresponds to the reconstruction stage, while the period from 73th to 96th months corresponds to the prediction stage.

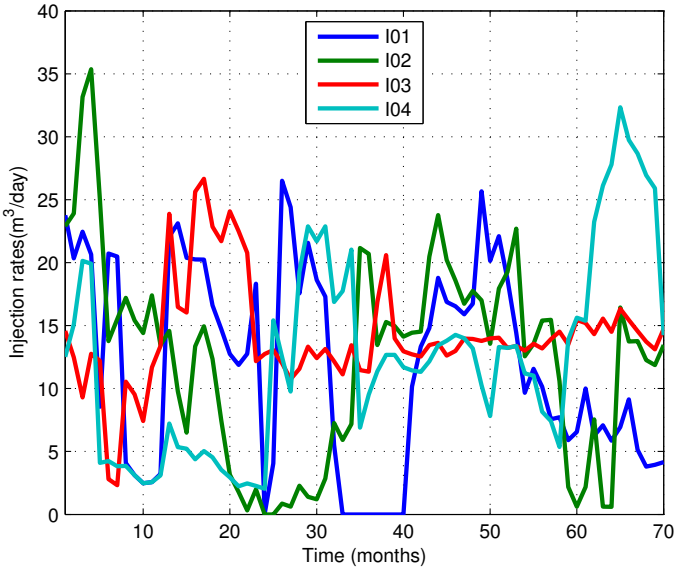

Fig. 3.18. The injection rates collected from the real data in Changqing oilfield.

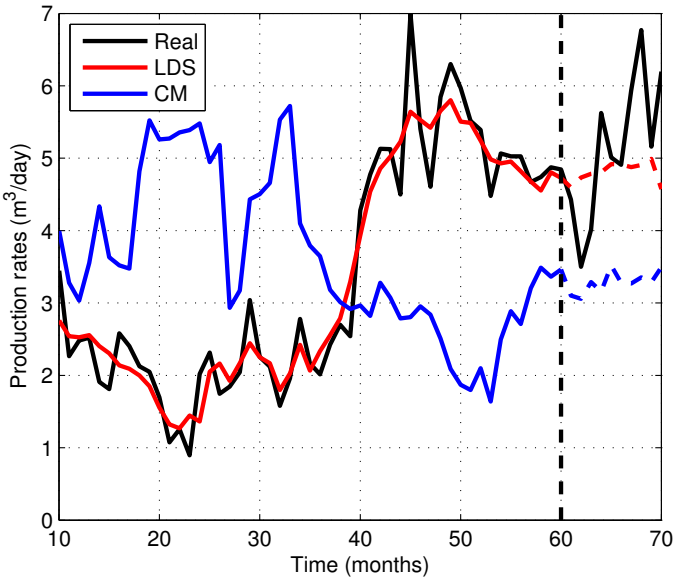

Fig. 3.19. Performance comparison of LDS and CM when applied to real data collected from Changqing oilfield. The period from the 1st to 60th month corresponds to the reconstruction stage, while the period from the 61st to 70th month corresponds to the prediction stage.

3.2.5 The Real Field

For real-field validation, we use the injection and production data collected from Changqing oilfield, containing low-permeability production formations, over a 70 month period, from June 2004 to March 2010. In this waterflooding project, 4 water injection wells are used to drive oil towards one producer, as shown in Fig. 3.4. We use the data of the first 60 months to train the model, based on which the production rates over the next 10 months are predicted.

Fig. 3.18 shows the injection rates and Fig. 3.19 shows the reconstructed and predicted production rates. First, in the reconstruction stage, although LDS cannot exactly match real production rates due to the volatile nature of real production data, it can still closely match them in trend attributed to the use of CVCRs as dynamic hidden states. In contrast, CM does not even generate a correct trend in the reconstruction stage, implying that CM is not a suitable model for real low-permeability reservoirs.

Furthermore, in the prediction stage, the prediction made by LDS is close to the mean value of production rates despite the drastic time-varying nature of the real production rates, while CM cannot provide meaningful results. Therefore, these results indicate that as compared to CM, LDS is more capable of handling the noisy injection and production rates observed in real low-permeability fields.

Chapter 4

Conclusions and Future Work

In this thesis, we propose a new method based on linear dynamical systems (LDS) to model waterflooding performance in oilfields. By leveraging control volumes as hidden state variables of the reservoir of interest, LDS is especially effective at characterizing the behavior of transient and non-stationary flows in low-permeability reservoirs. Compared to the well-known capacitance model (CM), LDS does not only explain the interwell connectivity among injectors and producers, but also quantifies the influence of the underlying dynamics of control volumes on the production rates.

We present an EM algorithm for the model to learn its parameters as well as the hidden reservoir states only based on the past observations of injection and production rates in the oilfield. To evaluate the effectiveness of the proposed model, we make predictions about future production rates given the planned injection schedules and the learned model in a wide range of scenarios, including a synthetic high-permeability reservoir and several synthetic low-permeability reservoirs of different scales with varying Bottom-Hole Injection Pressure (BHIP) and injector shut-ins. We have also tested the model performance on the real production data collected from Changqing oilfield, which contains low-permeability formations.

The following conclusions can be made. First, when applying LDS in high-permeability reservoirs, the mean prediction errors range from 0.9% to 1.7%, which are close to those given by CM, ranging from 0.3% to 0.9%. However, LDS signifi-

cantly outperforms CM in low-permeability reservoirs, with mean prediction errors ranging from 7% to 10%. In comparison, CM yields mean prediction errors from 20% to 30% in this case. Second, when applied to low-permeability reservoirs with varying BHIP and injector shut-in scenarios, LDS can successfully maintain a mean prediction error of approximately 10% and a maximum prediction error of about 20%. In contrast, for the same scenarios, the mean prediction error of CM ranges from 30% to 60%, with a maximum error ranging from 60% to 130%. Finally, when applied to larger-scale low-permeability reservoirs and the highly time-varying production data collected from a real oilfield containing low-permeability formations, LDS can still achieve correct trend prediction for mean production rates, while the CM fails to do so.

Some work can be done in the future. The prediction of the future production relies on previous trained parameters of the reservoir, which requires time-invariant condition [15]. When confronted with more difficult cases, such as extremely heterogeneous low-permeability reservoirs, sophisticated interaction cannot be quantified by the linear assumption in LDS, and more accurate model is required [23]. Considering that LDS can only provide the prediction of the total liquid rate, how to enable the model to capture the liquid rate and water cut is still open to explore.

References

[1] K. J. Heffer, R. J. Fox, C. A. McGill, N. C. Koutsabeloulis *et al.*, "Novel techniques show links between reservoir flow directionality, earth stress, fault structure and geomechanical changes in mature waterfloods," *SPE Journal*, vol. 2, no. 02, pp. 91–98, 1997.

[2] M. Panda, A. Chopra *et al.*, "An integrated approach to estimate well interactions," in *SPE India Oil and Gas Conference and Exhibition*. Society of Petroleum Engineers, 1998.

[3] A. Albertoni, L. W. Lake *et al.*, "Inferring interwell connectivity only from well-rate fluctuations in waterfloods," *SPE Reservoir Evaluation & Engineering*, vol. 6, no. 01, pp. 6–16, 2003.

[4] A. A. Yousef, P. H. Gentil, J. L. Jensen, L. W. Lake *et al.*, "A capacitance model to infer interwell connectivity from production and injection rate fluctuations," in *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2005.

[5] D. Kaviani, J. L. Jensen, and L. W. Lake, "Estimation of interwell connectivity in the case of unmeasured fluctuating bottomhole pressures," *Journal of Petroleum Science and Engineering*, vol. 90, pp. 79–95, 2012.

[6] M. Soroush, D. Kaviani, and J. L. Jensen, "Interwell connectivity evaluation in cases of changing skin and frequent production interruptions," *Journal of Petroleum Science and Engineering*, vol. 122, pp. 616–630, 2014.

[7] M. Soroush, "Interwell connectivity evaluation using injection and production fluctuation data," Ph.D. dissertation, University of Calgary, 2014.

[8] F. Liu, J. M. Mendel, A. M. Nejad *et al.*, "Forecasting injector/producer relationships from production and injection rates using an extended kalman filter," *SPE Journal*, vol. 14, no. 04, pp. 653–664, 2009.

[9] Z. Zhang, H. Li, and D. Zhang, "Water flooding performance prediction by multi-layer capacitance-resistive models combined with the ensemble kalman filter," *Journal of Petroleum Science and Engineering*, vol. 127, pp. 1–19, 2015.

[10] C. Bishop, "Pattern recognition and machine learning (information science and statistics), 1st edn. 2006. corr. 2nd printing edn," 2007.

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.

[12] M. Sayarpour, C. S. Kabir, L. W. Lake *et al.*, "Field applications of capacitance-resistance models in waterfloods," *SPE reservoir evaluation & engineering*, vol. 12, no. 06, pp. 853–864, 2009.

[13] T. Ahmed, P. McKinney *et al.*, *Advanced reservoir engineering*. Gulf Professional Publishing, 2011.

[14] H. Zhao, Y. Li, S. Cui, G. Shang, A. C. Reynolds, Z. Guo, and H. A. Li, "History matching and production optimization of water flooding based on a data-driven interwell numerical simulation model," *Journal of Natural Gas Science and Engineering*, vol. 31, pp. 48–66, 2016.

[15] L. Ljung, "System identification," in *Signal Analysis and Prediction*. Springer, 1998, pp. 163–173.

[16] Z. Ghahramani and G. E. Hinton, "Parameter estimation for linear dynamical systems," Technical Report CRG-TR-96-2, University of Totronto, Dept. of Computer Science, Tech. Rep., 1996.

[17] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[18] B. D. Anderson and J. B. Moore, "Optimal filtering. 1979," 1979.

[19] R. G. Brown and P. Y. Hwang, "Introduction to random signals and applied kalman filtering: with matlab exercises and solutions," *Introduction to random signals and applied Kalman filtering: with MATLAB exercises and solutions, by Brown, Robert Grover.; Hwang, Patrick YC New York: Wiley, c1997.*, vol. 1, 1997.

[20] G. Welch and G. Bishop, "An introduction to the kalman filter. department of computer science, university of north carolina," 2006.

[21] C. J. Wu, "On the convergence properties of the em algorithm," *The Annals of statistics*, pp. 95–103, 1983.

[22] R. K. Betty and A. Jonathan, "Essential medical statistics," *Kirkwood and Jonathan AC Sterne: Blackwell Science Ltd*, vol. 414, p. 425, 2003.

[23] B. Boots and G. J. Gordon, "An online spectral learning algorithm for partially observable nonlinear dynamical systems." in *AAAI*, 2011.