USC-SIPI REPORT #393

Intelliegent Signal Processing for Oilfield Waterflood Magagement

by

Feilong Liu

May 2008

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Viterbi School of Engineering
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Suite 400
Los Angeles, CA 90089-2564 U.S.A.

Dedication

To Brycen, Chuping and my parents

Acknowledgments

I would like to express my deepest gratitude and thanks to my advisor, Professor Jerry M. Mendel, for his invaluable guidance, inspiration, support, encouragement and patience through all these years in all aspects of my research and career. He has taught me the way to do scientific research, the way to be an authority in researching fields, and the way to be the member or leader of a team. I would also like to express my great gratitude to Dr. Iraj Ershaghi for his consistent support and suggestions on this research; Drs. Iraj Ershaghi and B. Keith Jenkins for serving on my qualifying exam and dissertation committee; and Drs. Antonio Ortega and Shrikanth S. Narayanan for serving on my qualifying exam committee.

My gratitude also goes to Dave Tuk, Jim Brink, Dave Rossi, John Houghton, Tai Nguyen and Vega Sankur for their very helpful recommendations on Chevron job, and Steve Giffoni for providing the summer internship opportunity and job offer.

I also would like to thank my officemates, teammates (e.g., Amir Mohammad Nejad), basketball buddies, soccer buddies and many friends at USC. With them, the life in US becomes very interesting and memorable.

Finally, I would like to appreciate my beloved wife, chuping, my funny son, Brycen, and my beloved parents for their unconditional support. With their companion, understanding, encouragement and support, my life in US becomes much cozy, blissful and fruitful.

Table of Contents

List of Tables

List of Figures

Abstract

This thesis addresses two problems about water-flood management: (1) infer reservoir heterogeneity using measured injection and production rates, and (2) construct a decision support system to optimize oil production using computing with words (CWW) and the inferred reservoir heterogeneity.

To infer reservoir heterogeneity, we first present an adaptive method using an Extended Kalman Filter (EKF) for the case of multiple injectors and a single producer, and then present a pseudo-virtual reservoir method for the case of multiple injectors and multiple producers, respectively. In the EKF method, a very simple parametric model, one with two parameters per injector, is used so that if a producer depends upon N injectors our model contains exactly $2N$ parameters; and the EKF is used to adaptively estimate the $2N$ parameters, from which the injector-producer relationship (IPR) between each injector and a producer is then estimated. In the pseudo-virtual reservoir method, a virtual reservoir model is used to model the reservoir, and a pseudo- virtual reservoir model is used to estimate the regional impact from each injector, so that the problem for this case reduces to the problem for the case of multiple injectors and a single producer.

The study for decision support system using CWW focuses on the encoder component (called type-2 fuzzistics) of a *perceptual computer* (Per-C), a specific architecture for CWW using interval type-2 fuzzy sets, and our work so far has been to transform linguistic perceptions, words, into interval type-2 fuzzy sets (IT2 FS) that activate a CWW engine. We have proposed a new and simple approach, called the

Interval Approach (*IA*), to type-2 fuzzistics, one that captures the strong points of both the person membership function and interval end-points approaches. It collects interval end-point data from subjects, does not require subjects to be knowledgeable about fuzzy sets, has a straightforward mapping from data to footprint of uncertainty (FOU), does not require an a priori assumption about whether or not a FOU is symmetric or non-symmetric, and leads to an IT2 FS word model that reduces to a T1 FS word model automatically if all subjects provide the same intervals.

Chapter 1

Introduction

1.1 Motivation and Objective

As the oil demand increases globally, and the numbers of newly discovered significant oilfields decrease, the requirement to produce oil more efficiently and economically becomes more and more urgent. Primary recovery using natural producing mechanisms, i.e., liquid and rock expansion and solution gas drive, leave behind 80% or more of the original oil in place (OOIP). Consequently, the secondary recovery becomes more significant than ever to recover billions of barrels of additional oil from oil reservoirs. Various secondary recovery methods, e.g., water-flooding, steam-flooding, CO_2 flooding, hydrocarbon gas injection and chemical injection etc, have been proposed. Water-flooding is one of the most widely used methods because water is easy to obtain and is inexpensive [40]. Unfortunately, the combined primary and current secondary recovery still leave about two-thirds of the OOIP behind. This stimulates the demand to bring more technologies into water-flood management.

Water-flood management is a broad topic involving a range of activities, some of which are performed only once or twice during the life of the water-flood, e.g., water-flood design and facilities/well planning. Other water-flood activities are performed periodically based on analyses of recurring measurement data, e.g., setting water

injection flow rates in the injection wells. The success of water-flood management depends on an understanding of reservoir heterogeneity. The more we know about the reservoir, the greater the chance water-flooding can succeed. One way to infer the reservoir heterogeneity is to determine some expression (often either mathematically or empirically derived) of the inter-relationship between injectors and corresponding affected producers using measured production and injection rates, which are the most abundant data available in oilfield. In this thesis, this inter-relationship is referred as the "injector-producer relationship (IPR)" for the case of multiple injectors and a single producer, and the "regional impact" for the case of multiple injectors and multiple producers. By knowing the IPR and regional impact, we might be able to tell something about the directional sweep efficiency of a given pattern, obtain a better understanding of zonal conformance and reservoir heterogeneity, and make better decisions on waterflood management.

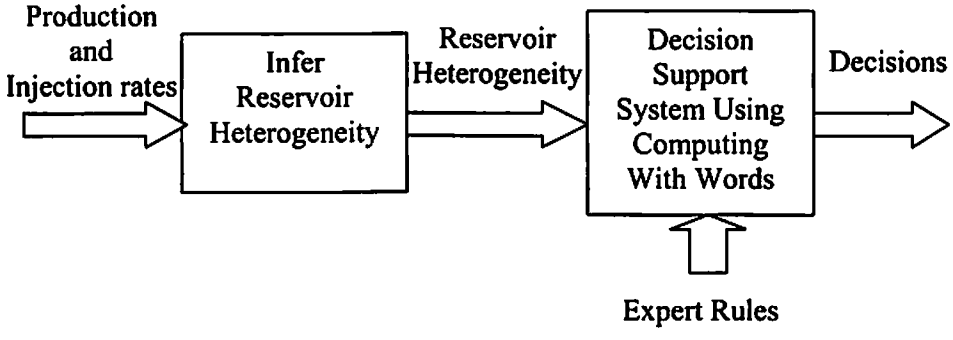


Figure 1.1: Block diagram of our research program.

As depicted in Fig. 1.1, the objectives of this thesis is to address two problems about water-flood management: (1) infer reservoir heterogeneity using measured injection and

production rates, and (2) construct a decision support system to optimize oil production using computing with words (CWW) and the inferred reservoir heterogeneity.

To infer reservoir heterogeneity, we first present an adaptive method using an Extended Kalman Filter (EKF) for the cases of multiple injectors and a single producer, and then present a pseudo-virtual reservoir approach for the case of multiple injectors and multiple producers, respectively. In the EKF method, a very simple parametric model, one with two parameters per injector, is used so that if a producer depends upon N injectors our model contains exactly $2N$ parameters; and the EKF is used to adaptively estimate the $2N$ parameters, from which the injector-producer relationship (IPR) between each injector and a producer is then estimated. In the pseudo-virtual reservoir method, a virtual reservoir model is used to model the reservoir, and a pseudo-virtual reservoir model is used to estimate the regional impact from each injector, so that the problem for the latter reduces to the problem for the former.

The study for decision support system using CWW focuses on the encoder component of a *perceptual computer* (Per-C), called type-2 fuzzistics. The Per-C is described in Chapter 5 (see Fig. 5.1) and is a specific architecture for CWW that uses interval type-2 fuzzy sets. Our work as far has been to transform linguistic perceptions, words, into interval type-2 fuzzy sets (IT2 FS) that activate a CWW engine. We have proposed a new and simple approach to type-2 fuzzistics, called the *Interval Approach* (*IA*), one that captures the strong points of both the person membership function and interval end-points approaches. It collects interval end-point data from subjects, does not require subjects to be knowledgeable about fuzzy sets, has a straightforward

mapping from data to footprint of uncertainty (FOU), does not require an a priori assumption about whether or not a FOU is symmetric or non-symmetric, and leads to an IT2 FS word model that reduces to a T1 FS word model automatically if all subjects provide the same intervals.

1.2 Thesis Outline

The rest of this thesis is organized as follows.

In Chapter 2, we begin with reservoir model, scale function and injector-producer model, define the injector-producer relation, construct the state variable model, and explain the EKF approach. Tests using synthetic data, CMG simulation data and real data are given to provide us some preliminary results on this EKF approach.

Chapter 3 begins with a virtual reservoir model, defines the regional impact, explains the similarity between this virtual reservoir model and the reservoir model in Chapter 2, and introduces a pseudo- virtual reservoir model to estimate regional impact. Tests using synthetic data confirm the feasibility of this virtual reservoir approach.

Chapter 4 reviews some required and concise background knowledge about type-1 and type-2 fuzzy logic systems that are needed in the subsequent chapters.

Chapter 5 briefly outlines the decision support system using computing with words, after which a new and simple approach, called the *Interval Approach* (*IA*), to type-2 fuzzistics is presented.

Finally, in Chapter 6, we summarize our work and propose some future work.

Chapter 2

Forecasting Injector-Producer Relationship from Production and Injection Rates Using Extended Kalman Filter

2.1 Introduction

Water-flood management is a broad topic involving a range of activities, some of which are performed only once or twice during the life of the water-flood, e.g., water-flood design and facilities/well planning. Other water-flood activities are performed periodically based on analyses of recurring measurement data, e.g., setting water injection flow rates in the injection wells. The success of water-flood management depends on our understanding of reservoir heterogeneity. The more we know about the reservoir, the greater chance the water-flooding can succeed. One way to infer the reservoir heterogeneity is to determine some expression (often either mathematically or empirically derived) of the inter-relationship between injectors and corresponding affected producers using the production and injection rates, the most abundant data available in the oilfield. Usually, this inter-relationship is referred to as the "injector-producer relationship (IPR)" for the case of multiple injectors and a single producer. By knowing the relative IPRs, we might be able to tell something about the directional sweep efficiency of a given pattern or even directional fracture, leading to a better

understanding of zonal conformance and reservoir heterogeneity, and maximizing the oil production by changing the injection rates from different injectors.

Many methods have been previously used to infer IPRs using only injection and production rates [1], [2], [6], [35]-[37], [39], [50], among which, the most recent works are from Albertoni and Lake [1], and Yousef *et al* [50]. In their works, the reservoir is considered to be a system, which can be represented as a continuous impulse response that converts the input signal (i.e., injection rates) into an output signal (i.e., production rates). Albertoni and Lake model the reservoir by a resistive model (RM), which is characterized by a collection of weights, whereas Yousef *et al.* model the reservoir by a capacitance model (CM), which is characterized by two parameters, λ and τ, that quantify the connectivity and the dissipation between wells, respectively. The IPRs (i.e., the interwell connectivities) are quantitatively evaluated by the model parameters, i.e., weights for RM, and λ for CM. The parameters in these models are estimated using the same approach, multiple linear regression (MLR); however, there are some weak points of both approaches: (1) In the process of estimating the IPRs, the parameters in the models and the IPRs are assumed stationary, i.e., unchanged over the window during which the data are analyzed so that when the IPRs change the analysis needs to be repeated for the new situation, and this may not be practical because the reservoir is dynamic and it may be very difficult to recognize when a change has occurred; (2) The CM model is very complex, and although it is characterized by two parameters λ and τ, to use it, the primary production and bottom hole pressure impact also need to be determined.

To overcome these difficulties, we propose a new approach to determine the IPRs in which [17]:

1) Learning from expert knowledge and the shape of a diffusivity filter, an exact two-parameter auto-regressive (AR) model is used to model the impulse response between a single injector and a single producer, and the area under the impulse response is then used as a measure of the IPR. If a producer depends on N injectors, our model will contain exactly $2N$ parameters;
2) Instead of assuming stationarity, the IPRs are considered to be non-stationary, i.e., IPRs may change constantly due to changing the bottom hole pressure, work-over or earthquake, but we have no idea about when the changes will happen, and how often they will occur;
3) The *Extended Kalman Filter* (EKF) is used to forecast and track the non-stationary IPR parameters because, as shown below, the *state variable model* (SVM) for our IPR model is nonlinear. Our forecasts of IPRs are nonlinear and adaptive functions of the measured data, where the nature of the nonlinearities is established by the mathematics of the EKF, and adaptation automatically permits the forecasters to track the changing conditions.

In this Chapter, we focus on the multiple injectors and single producer situation; the work for the case of multiple injectors and multiple producers will be discussed in Chapter 3.

2.2 Reservoir Model

Our "reservoir model" is a high-level system-oriented model and is not a petroleum engineering model. As such, it views a reservoir as an interconnection of injectors to a producer, and interprets this as a communication system in which the injectors are transmitting messages to the producer, but the messages are distorted by the channel between each injector and the producer. One could call this a synthetic reservoir model, but for short, we call it a reservoir model.

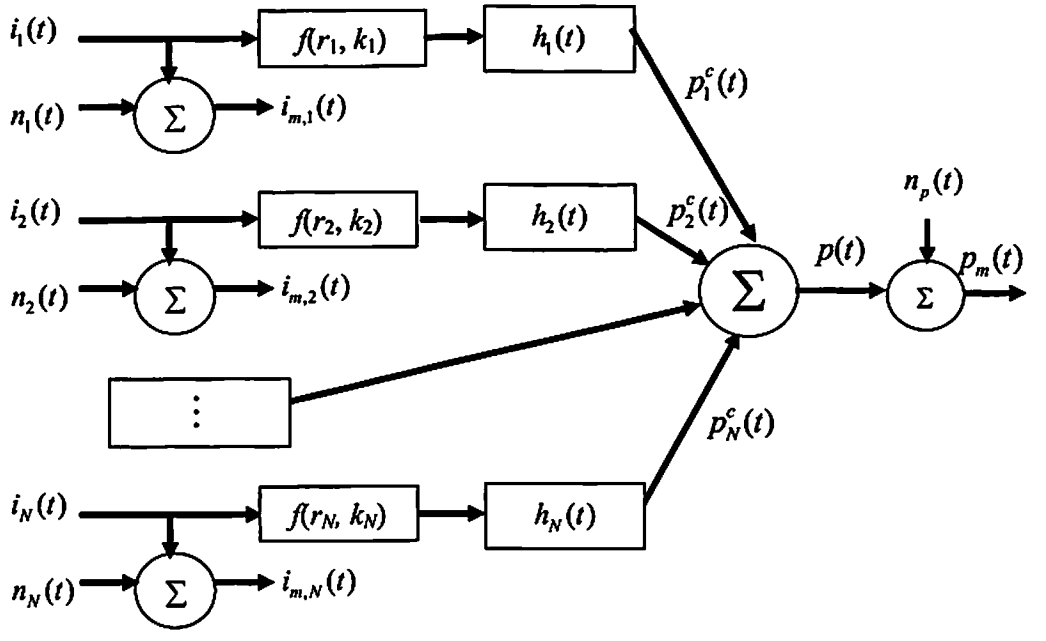


Figure 2.1: Reservoir model for N injectors and a single producer.

The reservoir model for N injectors and a single producer is depicted in Fig. 2.1, in which each injector-producer pair is considered as an independent subsystem; hence, for a reservoir with N injectors and a single producer, the reservoir is represented by N

independent subsystems, where each subsystem is modeled as a continuous-time communication channel. The notations in this reservoir model are:

1) $i_j(t)$, $n_j(t)$ and $i_{m,j}(t)$ ($j=1,\ldots,N$) are the actual injection rates that flow into the reservoir, the corresponding measurement noises for measuring injection rates, and the measured injection rates, respectively;
2) $f(r_j,k_j)$ ($j=1,\ldots,N$) are linear or nonlinear scalar functions of r_j and k_j, which relate to the connectivities between injectors and the producer, where r_j and k_j are distance and permeability between injector j and the producer, respectively. Heuristically, the greater the value of $f(r_j,k_j)$ the better is the connectivity between the injectors and producer;
3) $h_j(t)$ ($j=1,\ldots,N$) is the channel impulse response for subsystem j that is used to model the continuous-time communication channel between injector j and the producer. Each impulse response will be modeled as described in (2.4), and is characterized by two parameters;
4) $p(t)$, $n_p(t)$ and $p_m(t)$ are the actual production rate, the corresponding measurement noise for measuring production rate, and the measured production rate, respectively. (Note that production rate means the total production rate, i.e., the sum of oil, gas and water production rates.);
5) $p_j^c(t)$ is the channel production rate produced by injector j, which represents the amount of production rate in $p(t)$ caused by injector j;

6) $P(s)$, $P_j^c(s)$, $I_j(s)$ and $H_j(s)$ are Laplace transforms of $p(t)$, $p_j^c(t)$, $i_j(t)$, and $h_j(t)$, respectively; and

7) $P(z)$, $P_j^c(z)$, $I_j(z)$ and $H_j(z)$ are z-transforms of $p(n)$, $p_j^c(n)$, $i_j(n)$ and $h_j(n)$, respectively.

It will be seen later that:

1) The noise-free signals $i_j(t)$, $p_j^c(t)$ ($j=1,\ldots,N$) and $p(t)$ are used for constructing reservoir model, whereas the measured signals $i_{m,j}(t)$ ($j=1,\ldots,N$) and $p_m(t)$ are used for estimating the parameters of the reservoir model during our processing, because the noise-free signals $i_j(t)$, $p_j^c(t)$ ($j=1,\ldots,N$) and $p(t)$ are not directly available;

2) The Laplace transforms $P(s)$, $P_j^c(s)$, $I_j(s)$ and $H_j(s)$ are used to describe the reservoir model, whereas the z-transforms $P(z)$, $P_j^c(z)$, $I_j(z)$ and $H_j(z)$ are used for estimating the parameters of the reservoir model during our processing, because only discrete measurement data are available for such processing.

From Fig. 2.1, we see that

$$p(t)=\sum_{j=1}^{N} p_j^c(t) \tag{2.1}$$

and

$$p_j^c(t)=f(r_j,k_j)\left[i_j(t)*h_j(t)\right] \tag{2.2}$$

where * is the convolution operation. Taking the Laplace transform of both sides of (2.1) and (2.2), it follows that

$$P(s) = \sum_{j=1}^{N} P_j^c(s) = \sum_{j=1}^{N} f(r_j, k_j) H_j(s) I_j(s) \tag{2.3}$$

2.2.1 Scale function $f(r_j, k_j)$

Our view is that $f(r_j, k_j)$ is a linear or nonlinear scalar function of r_j and k_j. To date, we have no explicit formula for that. Even if we had an explicit formula for $f(r_j, k_j)$, it would still be impossible for us to use it because we don't know the permeability of the channel between the injector and producer; however, we do have some useful heuristic knowledge about $f(r_j, k_j)$, namely,

1) $f(r_j, k_j)$ is a monotonically decreasing function of r_j;
2) $f(r_j, k_j)$ is a monotonically increasing function of k_j; and,
3) The larger $f(r_j, k_j)$ is, the better is the connectivity from the injector to producer; the shorter is the delay time (i.e., the larger is a_j), and the larger is the impact from the injectors (i.e., the larger is IPR_j).

This knowledge will be used by us for our experimental designs when we have to choose the parameters for different reservoir models. Also, as we will see in Section 2.2.4, the $f(r_j, k_j)$ will be absorbed into γ_j when we estimate IPR_j.

2.2.2 Injector-Producer Subsystem Model

Each injector-producer pair is modeled as a continuous-time impulse response that converts the injection rates into production rate. Learning from expert knowledge (i.e., Chevron Engineers) and the shape of a diffusivity filter, this continuous-time impulse response has been assumed to be the continuous-time uni-modal function depicted in Fig. 2.2, which can be characterized as

$$h(t) = bte^{-at}. \tag{2.4}$$

The Laplace transform of $h(t)$, $H(s)$ has two poles at $-a$, i.e.,

$$H(s) = \frac{b}{(s+a)^2} \tag{2.5}$$

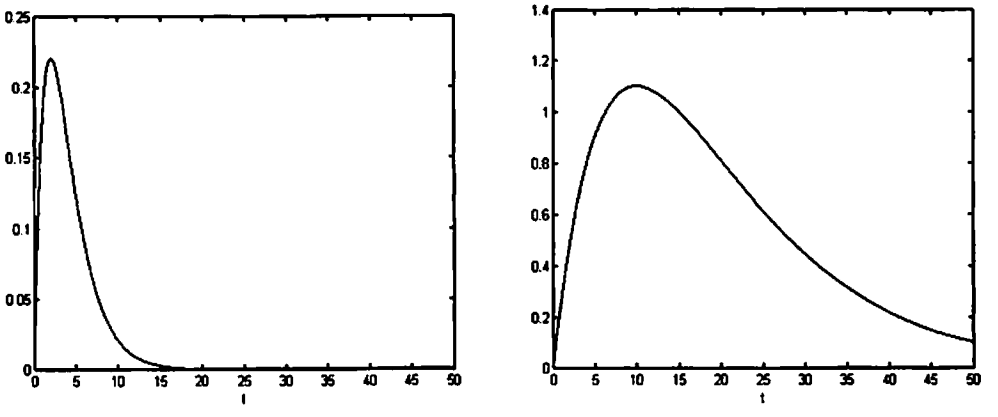


Figure 2.2: Two examples of the impulse response between an injector and a producer.

Because only sampled data (i.e., measurements of injection and production rates) are available for processing, the continuous-time impulse response is sampled, and a discrete impulse response is used to approximate it. Two ways to model a discrete impulse response that has a uni-modal shape are:

(1) Use an auto-regressive (AR) model that has two equal poles, e.g., for a single injector and single producer, such an AR model is

$$p(k+1)-2\alpha p(k)+\alpha^2 p(k-1)=\gamma i(k) \tag{2.6}$$

where $p(k)$ and $i(k)$ are the production and injection rates, respectively, and, $\alpha = e^{-aT}$ and $\gamma = b\alpha T$ are the two parameters that characterize the model.

(2) Use a moving-average (MA) model, e.g., for a single injector and single producer, the MA model is

$$p(k)=\sum_{j=1}^{l}\beta_j i(k-j) \tag{2.7}$$

where $p(k)$ and $i(k)$ are the production and injection rates, respectively, β_j, $(j=1,\ldots,l)$ are the weights, and l is the length of the impulse response. Usually, to better approximate the continuous-time impulse response, the value of l is much greater than 2.

The AR model has the following advantages over the MA model: (1) It needs fewer parameters to describe the discrete impulse response than the MA model, i.e., for N

injectors, the AR model only needs $2N$ parameters, whereas the MA model needs $\sum_{i=1}^{N} l_i$ parameters; and (2) It is difficult to determine the lengths of the MA model, i.e., the values of $l_1,\ldots,l_N$. Consequently, our preference for this project is to use the AR model, which has the following z-transform formulation:

$$H_j(z) = \frac{\gamma_j z^{-1}}{(1-\alpha_j z^{-1})^2} \tag{2.8}$$

where $\alpha_j = e^{-a_j T}$ and $\gamma_j = b_j \alpha_j T$ are the two parameters that characterize this discretized channel model. An example of such a $h_j(t)$ and $h_j(n)$ is depicted in Fig. 2.3, where the solid curve represents $h_j(t)$ and the dots represent the non-zero values of $h_j(n)$. Note that (2.8) is derived directly from (2.4).

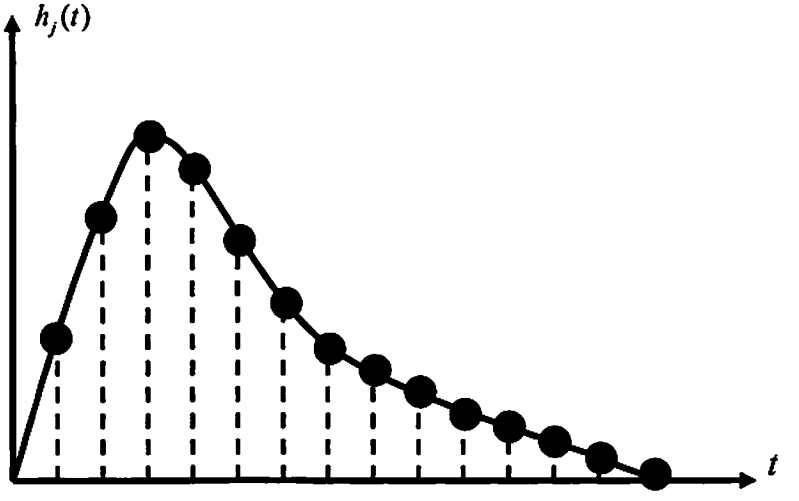


Figure 2.3: An example of channel impulse response $h_j(t)$ and $h_j(n)$ (the solid dots represent the non-zero samples).

2.2.3 Injector-Producer Relationship (IPR)

As shown in Fig. 2.4, when the j^{th} injection rate has a step change $\Delta I_j u(t-t_j)$ at time t_j, it produces a change in $p_j^c(t)$, $\Delta p_j^c(t)$. Because the Laplace transform of $\Delta I_j u(t-t_j)$ is $\Delta I_j \frac{1}{s} e^{-st_j}$, it follows from (2.3) that:

$$\Delta P(s) = \sum_{j=1}^{N} \Delta P_j^c(s) = \sum_{j=1}^{N} f(r_j, k_j) H_j(s) \Delta I_j(s) = \sum_{j=1}^{N} f(r_j, k_j) H_j(s) \Delta I_j \frac{1}{s} e^{-st_j} \qquad (2.9)$$

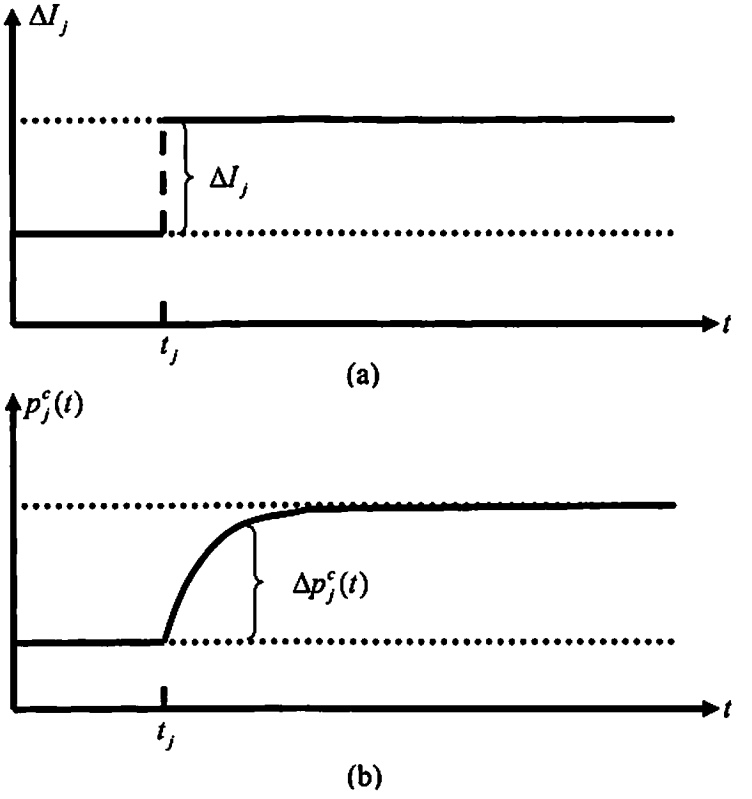


Figure 2.4: An example of (a) step change of injection rate; and (b) consequent change of production rate.

Taking the inverse Laplace transform of (2.9), the change of channel production rate produced by injector j, $\Delta p_j^c(t)$, can be expressed as:

$$\Delta p_j^c(t) = f(r_j, k_j)\left[\Delta I_j u(t-t_j) * h_j(t)\right] = \Delta I_j f(r_j, k_j)\int_0^t u(\tau - t_j)h_j(t-\tau)d\tau \quad (2.10)$$

When $t \to \infty$, the steady-state value of $\Delta p_j^c(t)$, $\Delta p_j^c(\infty)$, is

$$\begin{aligned}\Delta p_j^c(\infty) &= \Delta I_j f(r_j, k_j)\int_0^\infty u(\tau - t_j)h_j(\infty - \tau)d\tau \\ &= \Delta I_j f(r_j, k_j)\int_{t_j}^\infty h_j(\infty - \tau)d\tau \\ &= \Delta I_j f(r_j, k_j)\int_0^\infty h_j(t)dt\end{aligned} \quad (2.11)$$

The second line of (2.11) has made use of the fact that $u(\tau - t_j) = 0$ when $\tau < t_j$ and $u(\tau - t_j) = 1$ when $\tau > t_j$. The third line of (2.11) has made use of the transformation of variables $t = \infty - \tau$.

Definition 1: The *impact* from injector $i_j(t)$ to producer $p(t)$, IP_j, is defined as

$$IP_j \equiv \frac{\Delta p_j^c(\infty)}{\Delta I_j} = f(r_j, k_j)\int_0^\infty h_j(t)dt \quad (2.12)$$

This formula shows that a unit step change of the injection rate at an injector, i.e., $\Delta I_j = 1$, can cause a step change of steady-state production rate that equals $f(r_j,k_j)\int_0^\infty h_j(t)dt$. Obviously, $\int_0^\infty h_j(t)dt$ is the area under the continuous-time impulse response $h_j(t)$.

In order to compute IP_j, we need $h_j(t)$; however, instead of using $h_j(t)$, our method for computing IP_j is to use the sampled value of $h_j(t)$, $h_j(n)$, and to approximate IP_j in (2.12) as

$$\begin{aligned} IP_j &= f(r_j,k_j)\int_0^\infty h_j(t)dt \\ &\approx Tf(r_j,k_j)\sum_{n=0}^{\infty} h_j(n) \\ &= Tf(r_j,k_j)H_j(z)|_{z=1} \\ &= T\frac{\gamma_j f(r_j,k_j)}{(1-\alpha_j)^2} \end{aligned} \tag{2.13}$$

The fact that $\sum_{n=0}^{\infty} h_j(n) = H_j(z)|_{z=1}$ follows from the definition of $H_j(z) = \sum_{n=0}^{\infty} h_j(n)z^{-n}$.

Because the sample period T is normally the same for all injectors, we omit it and evaluate the impact using the relative *injector-producer relationship*.

Definition 2: The *injector-producer relationship* for the j^{th} injector, IPR_j, is defined as

$$IPR_j = \frac{IP_j}{T} \approx \frac{\gamma_j f(r_j,k_j)}{(1-\alpha_j)^2} = f(r_j,k_j)H_j(z)|_{z=1} \tag{2.14}$$

In order to estimate IPR_j from measured data, we need to estimate α_j, γ_j and $f(r_j,k_j)$; however, $f(r_j,k_j)$, which can be a linear or nonlinear scalar function of r_j and k_j, is unknown. Conceptually, since γ_j and $f(r_j,k_j)$ appear as a product, it is impossible to resolve each individually; hence, we re-express IPR_j in (2.14) as

$$IPR_j = \frac{IP_j}{T} \approx \frac{\gamma_j'}{(1-\alpha_j)^2} \tag{2.15}$$

where $\gamma_j' = \gamma_j f(r_j,k_j)$. (2.15) shows that each *IPR* is characterized by two parameters α_j and γ_j'. The process to estimate IPR_j becomes:

1) Estimate α_j and γ_j', and obtain $\hat{\alpha}_j$ and $\hat{\gamma}_j'$ $(j=1,\ldots,N)$; and,

2) Estimate IPR_j by $\hat{IPR}_j$, where

$$\hat{IPR}_j = \frac{\hat{\gamma}_j'}{(1-\hat{\alpha}_j)^2} \tag{2.16}$$

This is done by the EKF.

2.3 The Extended Kalman Filter

Using the available sampled measurements of injection and production rates, our goals are:

1) Estimate the channel impulse response parameters γ_j' and α_j $(j=1,\ldots,N)$;

2) Compute $\hat{IPR}_j$ using $\hat{\gamma}_j'$ and $\hat{\alpha}_j$ $(j=1,\ldots,N)$; and,

3) Forecast the future production rate.

2.3.1 State Variable Model (SVM)

The first step to using an EKF is to construct a *SVM* [19]. To illustrate how to do this, consider a simple case with three injectors and a single producer (it is easy to generalize the *SVM* from three injectors to *N* injectors.), for which the reservoir model in (2.3) is:

$$\begin{aligned} P(z) &= P_1^c(z) + P_2^c(z) + P_3^c(z) \\ &= f(r_1)H_1(z)I_1(z) + f(r_2)H_2(z)I_2(z) + f(r_3)H_3(z)I_3(z) \\ &= \sum_{j=1}^{3} \frac{\gamma_j' z^{-1}}{(1-\alpha_j z^{-1})^2} I_j(z) \end{aligned} \tag{2.17}$$

where $P_j^c(z) = f(r_j)H_j(z)I_j(z)$ ($j=1,2,3$) are z-transforms for the respective production rates from injectors 1, 2 and 3, respectively. Separate $P_j^c(z)$ from $P(z)$ to obtain ($j=1,2,3$):

$$\frac{P_j^c(z)}{I_j(z)} = \frac{\gamma_j' z^{-1}}{(1-\alpha_j z^{-1})^2} = \frac{\gamma_j' z^{-1}}{1-2\alpha_j z^{-1}+\alpha_j^2 z^{-2}} \tag{2.18}$$

(2.18) can be expressed as:

$$\left(1-2\alpha_j z^{-1}+\alpha_j^2 z^{-2}\right) P_j^c(z) = \gamma_j' z^{-1} I_j(z) \tag{2.19}$$

Multiply z on both sides of (2.19), and determine the inverse z-transform of the resulting equation, to see that (2.19) can be expressed in the time-domain as:

$$\begin{aligned} p_j^c(k+1)-2\alpha_j p_j^c(k)+\alpha_j^2 p_j^c(k-1) &= \gamma_j' i_j(k) \\ &= \gamma_j'\left(i_{m,j}(k)-n_j(k)\right) \\ &= \gamma_j' i_{m,j}(k)+n_{p_j^c}(k) \end{aligned} \tag{2.20}$$

where $i_j(k) = i_{m,j}(k) - n_j(k)$ can be seen from Fig. 2.1, and $n_{p_j^c}(k) = -\gamma_j' n_j(k)$. This is a second-order finite-difference equation that is described by two state variables $p_j^c(k-1)$ and $p_j^c(k)$. In addition, the unknown parameters, α_i and γ_i are also treated as state variables so that they can be estimated by an EKF, and their state equation models are

$$\left.\begin{aligned} \alpha_j(k+1) &= \alpha_j(k)+n_{\alpha_j}(k) \\ \gamma_j'(k+1) &= \gamma_j'(k)+n_{\gamma_j'}(k) \end{aligned}\right\} \tag{2.21}$$

where $n_{\alpha_j}(k)$ and $n_{\gamma_j}(k)$ are discrete-time zero mean white random noises. Eqs. (2.20) and (2.21) together are described by a 4×1 state vector $\boldsymbol{x}_j$, where

$$\boldsymbol{x}_j = \left[\alpha_j(k), \gamma_j'(k), p_j^c(k-1), p_j^c(k)\right]' \tag{2.22}$$

Using standard techniques [19], the *SVM* for p_j^c is ($j=1,2,3$):

$$\begin{cases} \boldsymbol{x}_j(k+1) = \begin{bmatrix} x_{j1}(k+1) \\ x_{j2}(k+1) \\ x_{j3}(k+1) \\ x_{j4}(k+1) \end{bmatrix} = \begin{bmatrix} \alpha_j(k+1) \\ \gamma_j(k+1) \\ p_j^c(k) \\ p_j^c(k+1) \end{bmatrix} = \begin{bmatrix} x_{j1}(k) \\ x_{j2}(k) \\ x_{j4}(k) \\ 2x_{j1}(k)x_{j4}(k) - x_{j1}^2(k)x_{j3}(k) \\ +x_{j2}(k)i_{m,j}(k) \end{bmatrix} + \boldsymbol{n}_{x_j}(k) \\ \\ p_j^c(k+1) = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} \boldsymbol{x}_j(k+1) \end{cases} \tag{2.23}$$

where $p_j^c(k+1)$ and $i_{m,j}(k)$ are the respective production rate and measured injection rate for injector j ($j=1,2,3$), and,

$$\boldsymbol{n}_{x_j}(k) = [n_{\alpha_j}(k),\ n_{\gamma_j}(k),\ 0,\ n_{p_j^c}(k)]' \tag{2.24}$$

is a 4×1 additive zero-mean white noise for the state equations.

Let $x(k+1) \equiv col\left[x_1(k+1),\, x_2(k+1), x_3(k+1)\right]$; then, the complete SVM for all three injector subsystems is:

$$\begin{cases} x(k+1) = \begin{bmatrix} x_{11}(k+1) \\ x_{12}(k+1) \\ x_{13}(k+1) \\ x_{14}(k+1) \\ x_{21}(k+1) \\ x_{22}(k+1) \\ x_{23}(k+1) \\ x_{24}(k+1) \\ x_{31}(k+1) \\ x_{32}(k+1) \\ x_{33}(k+1) \\ x_{34}(k+1) \end{bmatrix} = \begin{bmatrix} x_{11}(k) \\ x_{12}(k) \\ x_{14}(k) \\ 2x_{11}(k)x_{14}(k) - x_{11}^2(k)x_{13}(k) + x_{12}(k)i_{m,1}(k) \\ x_{21}(k) \\ x_{22}(k) \\ x_{24}(k) \\ 2x_{21}(k)x_{24}(k) - x_{21}^2(k)x_{23}(k) + x_{22}(k)i_{m,2}(k) \\ x_{31}(k) \\ x_{32}(k) \\ x_{34}(k) \\ 2x_{31}(k)x_{34}(k) - x_{31}^2(k)x_{33}(k) + x_{32}(k)i_{m,3}(k) \end{bmatrix} + \boldsymbol{n}_x(k) \\ p_m(k+1) = p_1^c(k+1) + p_2^c(k+1) + p_3^c(k+1) + n_p(k+1) \\ \qquad = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} x(k+1) + n_p(k+1) \end{cases} \tag{2.25}$$

where $p_m(k+1)$ is the measured production rate, $\boldsymbol{n}_x(k) = col[\boldsymbol{n}_{x_1}(k), \boldsymbol{n}_{x_2}(k), \boldsymbol{n}_{x_3}(k)]$ and $n_p(k+1)$ are additive zero-mean white noises for the state and measurement equations with covariance matrix $\boldsymbol{Q}_k$ and variance r_{k+1}, respectively, where

$$\boldsymbol{Q}_k = diag\left[r_{n_{\alpha_1}},\, r_{n_{f_1}},\, 0,\, r_{n_{p_1^c}}, r_{n_{\alpha_2}},\, r_{n_{f_2}},\, 0,\, r_{n_{p_2^c}}, r_{n_{\alpha_3}},\, r_{n_{f_3}},\, 0,\, r_{n_{p_3^c}} \right] \tag{2.26}$$

Because this SVM is nonlinear, we can't use a Kalman filter to estimate its twelve states; Instead, we use the EKF to estimate them. Note that after all twelve states are estimated, we can extract the estimates of the six injector-producer model parameters from them, namely $\hat{\alpha}_j(k+1|k+1)$, $\hat{\gamma}'_j(k+1|k+1)$ ($j=1,2,3$), after which we can compute $\hat{IPR}_j(k+1|k+1)$ ($j=1,2,3$) as

$$\hat{IPR}_j(k+1|k+1)=\frac{\hat{\gamma}'_j(k+1|k+1)}{(1-\hat{\alpha}_j(k+1|k+1))^2} \tag{2.27}$$

In this way, we see that the estimates of $\hat{IPR}_j$ are functions of time.

2.3.2 EKF approach

The EKF has been widely used [19], [5] as the standard technique for recursive nonlinear estimation. It provides a first-order approximation to optimal nonlinear mean-square state estimation for the following nonlinear discrete-time system:

$$\begin{cases} \boldsymbol{x}(k+1)=f\left[\boldsymbol{x}(k),k\right]+\boldsymbol{n}_x(k) \\ \\ y(k+1)=h\left[\boldsymbol{x}(k+1),k+1\right]+n_y(k+1), \quad k=1,2,\ldots \end{cases} \tag{2.28}$$

where $\boldsymbol{n}_x(k)$ and $n_y(k)$ are additive zero-mean white noises for the state and measurement equations with covariance matrices $\boldsymbol{Q}(k)$ and variance $r(k)$, respectively.

The EKF operates in two stages, predictor and corrector. The predictor computes the predicted value of $x(k+1)$, $\hat{x}(k+1|k)$, and the corrector computes the filtered value of $x(k+1)$, $\hat{x}(k+1|k+1)$.

In the EKF, the state equations are linearized about $\hat{x}(k|k)$, and the measurement equation is linearized about $\hat{x}(k+1|k)$, i.e.,

$$\begin{cases} x(k+1) \approx f\left[\hat{x}(k|k),k\right] + F_x\left[\hat{x}(k|k),k\right]\left[x(k) - \hat{x}(k|k)\right] + n_x(k) \\ y(k+1) \approx h\left[\hat{x}(k+1|k),k+1\right] + H_x\left[\hat{x}(k+1|k),k+1\right]\left[x(k+1) - \hat{x}(k+1|k)\right] + n_y(k+1) \end{cases} \tag{2.29}$$

where $\hat{x}(k+1|k) = f\left[\hat{x}(k|k),k\right]$, $F_x = \partial f\left[x(k),k\right]/\partial x(k)$ and $H_x = \partial h\left[x(k),k\right]/\partial x(k)$. Note that F_x is a $4N \times 4N$ Jacobian matrix, and H_x is a $1 \times 4N$ Jacobian matrix for the SVM of the reservoir model.

Example 1: For a single injector, using (2.23), it follows that:

$$F_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 2x_{14}(k) - 2x_{11}(k)x_{13}(k) & i_{m,1}(k) & -x_{11}^2(k) & 2x_{11}(k) \end{bmatrix} \tag{2.30}$$

and

$$H_x = [0 \quad 0 \quad 0 \quad 1]; \tag{2.31}$$

and, for three injectors, using (2.25), it follows that:

$$F_x = \begin{bmatrix} A_1 & \underline{0} & \underline{0} \\ \underline{0} & A_2 & \underline{0} \\ \underline{0} & \underline{0} & A_3 \end{bmatrix} \tag{2.32}$$

and

$$H_x = [0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 1]. \tag{2.33}$$

where

$$A_i = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 2x_{i4}(k) - 2x_{i1}(k)x_{i3}(k) & i_{m,i}(k) & -x_{i1}^2(k) & 2x_{i1}(k) \end{bmatrix} \tag{2.34}$$

for ($i = 1,2,3$), and $\underline{0}$ is a 4-by-4 zero matrix.

The EKF is summarized as follows:

1. Initialize the EKF with $\hat{x}(0|0)$, $P(0|0)$, Q_k and r_k.
2. Predictor ($k = 0,1,\ldots$):

$$\hat{x}(k+1|k) = f\left[\hat{x}(k|k), k\right] \tag{2.35}$$

$$P(k+1|k) = F_x(\hat{x}(k|k), k)P(k|k)F_x'(\hat{x}(k|k), k) + Q_k \tag{2.36}$$

3. Corrector ($k = 0,1,\ldots$):

$$\hat{x}(k+1|k+1) = \hat{x}(k+1|k) + K(k+1)\left\{y(k+1) - h\left[\hat{x}(k+1|k), k+1\right]\right\} \tag{2.37}$$

$$K(k+1) = \frac{P(k+1|k)H_x'(\hat{x}(k+1|k))}{H_x(\hat{x}(k+1|k))P(k+1|k)H_x'(\hat{x}(k+1|k)) + r_{k+1}} \tag{2.38}$$

$$P(k+1|k+1) = [I - K(k+1)H_x(\hat{x}(k+1|k))]P(k+1|k) \tag{2.39}$$

2.3.3 Production forecaster using SVM (Production Forecaster)

Using the SVM in (2.28), the future regional production rate can be forecasted using zero state and measurement noises and scheduled injection rates $i_{s,j}(k)$, i.e.,

$$\begin{cases} x(k+1) = f\left[x(k), k\right] \\ y(k+1) = h\left[x(k+1), k+1\right], \quad k = n, n+1, \ldots \end{cases} \tag{2.40}$$

Example 2: For a single injector, using (2.40) and (2.23), it follows that the future channel production rate $p_j^c(k+1)$ is forecasted as

$$\begin{cases} x_j(k+1) = \begin{bmatrix} x_{j1}(k) \\ x_{j2}(k) \\ x_{j4}(k) \\ 2x_{j1}(k)x_{j4}(k) - x_{j1}^2(k)x_{j3}(k) \\ +x_{j2}(k)i_{s,j}(k) \end{bmatrix} \\ p_j^c(k+1) = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} x_j(k+1) \end{cases} \tag{2.41}$$

and, for three injectors, using (2.40) and (2.25), it follows that the future regional production rate $\hat{p}_m(t)$ is forecasted as

$$\begin{cases} x(k+1) = \begin{bmatrix} x_{11}(k) \\ x_{12}(k) \\ x_{14}(k) \\ 2x_{11}(k)x_{14}(k) - x_{11}^2(k)x_{13}(k) + x_{12}(k)i_{s,1}(k) \\ x_{21}(k) \\ x_{22}(k) \\ x_{24}(k) \\ 2x_{21}(k)x_{24}(k) - x_{21}^2(k)x_{23}(k) + x_{22}(k)i_{s,2}(k) \\ x_{31}(k) \\ x_{32}(k) \\ x_{34}(k) \\ 2x_{31}(k)x_{34}(k) - x_{31}^2(k)x_{33}(k) + x_{32}(k)i_{s,3}(k) \end{bmatrix} \\ p_m(k+1) = p_1^c(k+1) + p_2^c(k+1) + p_3^c(k+1) \\ \qquad = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} x(k+1) \end{cases} \tag{2.42}$$

The forecasting is implemented as follows:

1. Initialize $x(k_1)$ in SVM using $\hat{x}(k_1 \mid k_1)$.

2. Set $n = k_1$, and let L_{fw} be the length of window for forecasting.

3. Forecast $\hat{p}_j^c(k+1)$ using (2.41) for single injector, and (2.42) for three injectors.

4. Set $n = n+1$; if $n < k_1 + L_{fw}$, go to Step 3; otherwise, go to Step 5.

5. Stop.

2.4 Simulations

Our approach was tested using synthetic data, CMG simulation data and real data. The tests using synthetic data used Monte-Carlo simulations, the tests using CMG simulation data used the data generated by petroleum commercial simulator, and the test using real data used data from Section 5 of Chevron's Lost Hills Field. The reason for using synthetic data is that we can validate our approach because we know what "truth" is. We can easily alter data sets to reflect plausible reservoir geology changes, so that we can know if our approach can adapt to such changes. Tests using reservoir simulation data are performed to confirm whether or not our estimated results match the models used for generating reservoir simulation data. Tests using real data are performed to see if our results match some expert knowledge.

2.4.1 Synthetic data

This section begins with the procedure for choosing the values of the parameters $f(r_j, k_j)$, a_j, b_j in reservoir model, introduces the procedure for synthetic data generation and Monte-Carlo simulation, and finally summarizes the tests.

A. Procedure for choosing the values of $f(r_j,k_j)$, a_j, b_j.

Recall from our third heuristic-knowledge for $f(r_j,k_j)$ in Section 2.2.1, the larger $f(r_j,k_j)$ is, the better is the connectivity from the injector to producer; the shorter is the delay time (i.e., the larger is a_j), and the larger is the impact from the injectors (i.e., the larger is IPR_j). Our procedure for choosing the values of $f(r_j,k_j)$, a_j, b_j is:

1) Specify the value of $f(r_j,k_j)$ in $[0, 1]$;
2) Choose the value of $a_j \in [1,1000]$ and $IPR_j \in [0, 1]$ using following rules:

   Rule 1: If $f(r_j,k_j)$ is *large*, then a_j and IPR_j are *large*.

   Rule 2: If $f(r_j,k_j)$ is *moderate*, then a_j and IPR_j are *moderate*.

   Rule 3: If $f(r_j,k_j)$ is *small*, then a_j and IPR_j are *small*.

3) Determine b_j using $\alpha_j = e^{-a_j T}$, $\gamma_j = b_j \alpha_j T$ and (2.8) as

$$b_j = \frac{\gamma_j}{\alpha_j T} = \frac{H_j(z)|_{z=1}}{(1-\alpha_j)^2 \alpha_j T} = \frac{H_j(z)|_{z=1}}{\left(1-e^{-a_j T}\right)^2 e^{-a_j T} T} = \frac{IPR_j}{f(r_j,k_j)\left(1-e^{-a_j T}\right)^2 e^{-a_j T} T} \tag{2.43}$$

where T is a small sample period, which corresponds to high sampling rate so that the generated discrete signal can be considered as a continuous signal.

B. Synthetic data generation

Synthetic data was generated as follows: (Note: the procedure is explained for the case of three injectors and two producers, and can be extended to any other situation.)

Step 1: Generate the noise-free injection rates $i_j^h(n)$ ($j=1,2,3$) using the injection rate model described in Appendix A.

Step 2: To generate the noise-free production rates $p^h(n)$, the *SVM* for multiple injectors and a single producer was implemented with zero-noise by specifying $f(r_j,k_j)$, a_j, b_j ($j=1,2,3$) using the procedure described in Section 2.4.1.A and a small $T=0.01$, which corresponds to a high sampling rate (recall that $\alpha_j=e^{-a_jT}$ and $\gamma_j=b_j\alpha_jT$).

Step 3: The noise-free injection rates $i_j(k)$ ($j=1,2,3$) and noise-free production rates $p(k)$ were obtained by down-sampling $i_j^h(n)$ ($j=1,2,3$) and $p^h(n)$ by 20. This corresponds to a low sampling rate.

Step 4: The measured injection and production rates were obtained by adding measurement noises $n_j(k)$ with $SNR=10\log_{10}\left[\frac{E\left(i_j(k)\right)^2}{E\left(n_j(k)\right)^2}\right]=20$ dB to $i_j(k)$ ($j=1,2,3$), and $n_p(k)$ with $SNR=10\log_{10}\left[\frac{E\left[\left(p(k)\right)^2\right]}{E\left[\left(n_p(k)\right)^2\right]}\right]=20$ dB to $p(k)$, i.e.,

$$\begin{cases} i_{m,j}(k) = i_j(k) + n_j(k) \\ p_m(k) = p(k) + n_p(k) \end{cases} \tag{2.44}$$

C. Procedure for implementing EKF estimator and production forecaster

Recall that our goals are to estimate the channel impulse response parameters γ_j' and $\hat{\alpha}_j$ $(j=1,\ldots,N)$; compute $IP\hat{R}_j$ using γ_j' and $\hat{\alpha}_j$ $(j=1,\ldots,N)$; and forecast the production rate $\hat{p}(t)$; and these works have been done either separately or simultaneously. An example with four time zones ($k_0 \le t \le k_1$, $k_1 \le t \le k_1 + L_{fw}$, $k_1 \le t \le k_1 + 2L_{fw}$ and $t \ge k_1 + 2L_{fw}$.) is depicted in Fig. 2.5, in which,

1) For $k_0 \le t \le k_1$, the measured injection and production rates, $i_{m,j}(t)$, $p_m(t)$, flow into the EKF estimator; and $\gamma_j'(t), \hat{\alpha}_j(t), IP\hat{R}_j(t)$ are obtained;

2) For $k_1 < t \le k_1 + L_{fw}$, where L_{fw} is the length of window for production forecasting, the estimated $\gamma_j'(k_1), \hat{\alpha}_j(k_1), IP\hat{R}_j(k_1)$ flow into both the EKF estimator and production forecaster as the initialization such that (a) $\gamma_j'(t), \hat{\alpha}_j(t), IP\hat{R}_j(t)$ are continuously updated using the new incoming injection and production rates, $i_{m,j}(t)$, $p_m(t)$, and (b) $\hat{p}(t)$ is forecasted from time k_1 to $k_1 + L_{fw}$ using the scheduled injection rates $i_{s,j}(t)$.

3) For $k_1 + L_{fw} < t \le k_1 + 2L_{fw}$, the estimated $\gamma_j'(k_1 + L_{fw}), \hat{\alpha}_j(k_1 + L_{fw})$,

$I\hat{P}\hat{R}_j(k_1 + L_{fw})$ flow into both the EKF estimator and production forecaster as the initialization such that (a) $\gamma_j'(t), \hat{\alpha}_j(t), I\hat{P}\hat{R}_j(t)$ are continuously updated using the new incoming injection and production rates, $i_{m,j}(t), p_m(t)$, and (b) the $\hat{p}(t)$ is forecasted after time k_2 using the scheduled injection rates $i_{s,j}(t)$.

4) For $t > k_1 + 2L_{fw}$, do the same as in step 3 except the initialization becomes $\gamma_j'(k_1 + 2L_{fw}), \hat{\alpha}_j(k_1 + 2L_{fw}), I\hat{P}\hat{R}_j(k_1 + 2L_{fw})$.

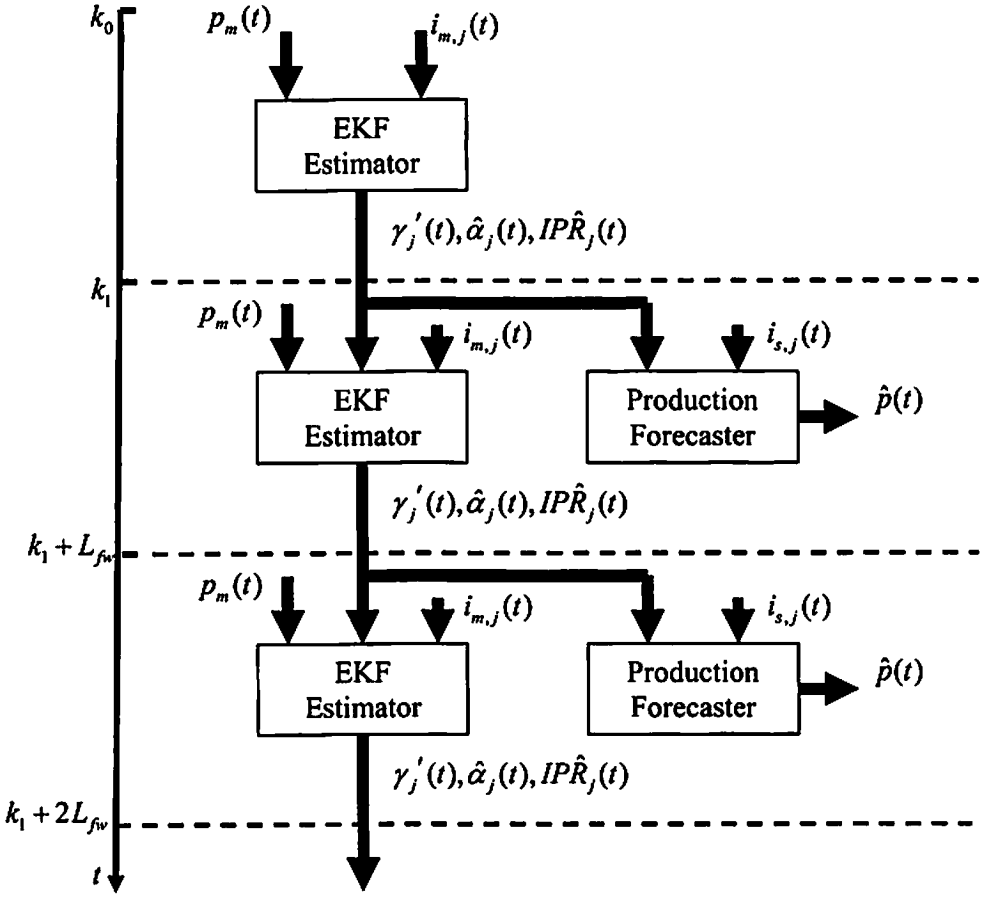


Figure 2.5: The diagram for implementing EKF estimator and production forecaster.

D. Monte-Carlo simulations

Simulations were implemented as follows:

Step 1: Set $i=1$, number of realizations $NR=100$, and the number of samples in each realization $M=500$ for Case 1 and $M=2000$ for Case 2.

Step 2: For the i^{th} realization ($j=1,2,3, k=1,\ldots,M$):

1) Generate $i_{m,j}(k)$ and $p_m(k)$, using the four-step procedure described in Section 2.4.1.B.
2) Implement the *EKF* using the just generated $i_{m,j}(k)$ and $p_m(k)$, and obtain $\hat{\gamma}'_j(k|k)$ and $\hat{\alpha}_j(k|k)$;
3) Compute the $\left[IP\hat{R}_j(k|k)\right]_i$ using (2.27).
4) Forecasting the production rate $\hat{p}(k)$, using the procedure described in Section 2.4.1.C.
5) Compute the forecasting error $\left[e(k)\right]_i$ using

$$\left[e(k)\right]_i = p_m(k) - \hat{p}(k) \tag{2.45}$$

Step 3: If $i \geq NR$, go to Step 4; otherwise, set $i=i+1$ and go to Step 2.

Step 4: Compute the mean and standard deviation of $\left[IP\hat{R}_j(k|k)\right]_i$ using

$$m_j(k) = \frac{1}{NR}\sum_{i=1}^{NR}\left[IP\hat{R}_j(k|k)\right]_i \tag{2.46}$$

and

$$\sigma_j(k) = \sqrt{\frac{1}{NR}\sum_{i=1}^{NR}\left(\left[\hat{IPR}_j(k\,|\,k)\right]_i - m_j(k)\right)^2}, \tag{2.47}$$

respectively.

Step 5: Compute the mean and standard deviation of $\left[e(k)\right]_i$ using

$$m_e(k) = \frac{1}{NR}\sum_{i=1}^{NR}\left[e(k)\right]_i \tag{2.48}$$

and

$$\sigma_e(k) = \sqrt{\frac{1}{NR}\sum_{i=1}^{NR}\left(\left[e(k)\right]_i - m_e(k)\right)^2}, \tag{2.49}$$

respectively.

Step 6: Stop.

E. Tests on synthetic data

Many tests were performed, and two of them are presented below. The first is for constant IPRs, and the second is for non-constant IPRs.

Case 1: Three injectors and a single producer with constant IPRs

For this case, the reservoir is static with constant IPRs. The parameters for this case are summarized in Table 1. Observe from the values of IPR_j that injector 2 has the

largest impact on the producer, injector 1 has about 75% less impact on the producer than injector 2, and injector 3 has the smallest impact on the producer, and this is about 95% less than injector 2.

Table 2.1: Parameters for Case 1.

j	$f(r_j,k_j)$	a_j	b_j	α_j	γ_j	IPR_j
1	0.5	200	120	0.1353	0.1624	0.1086
2	1	400	2400	0.0183	0.4396	0.4561
3	0.25	10	0.1	0.9048	0.0009	0.025

An example of measured production rates is depicted in Fig. 2.6. The test results are depicted in Figs. 2.7 and 2.8 for estimates of $IPR_j(k)$ ($j=1,2,3$), and in Figs. 2.9 and 2.10 for the production forecast error with $k_1=125$ and $L_{fw}=90$, respectively. Note that the original values of $IPR_j(k)$ are provided in the legend of Fig.2.7.

Observe that:

1) From Fig. 2.7, our approach can correctly estimate the relative IPRs, i.e., the largest $\hat{IPR}$ is obtained from the most influential injector, and the smallest $\hat{IPR}$ is obtained from least influential injector.
2) The $m_e(k)$ and $\sigma_e(k)$, which range mostly from -50 to 50, and 30 to 50, respectively, are much smaller than the generated measured production rates, which is around 1200. This confirms that our approach can successfully forecast future production rates.

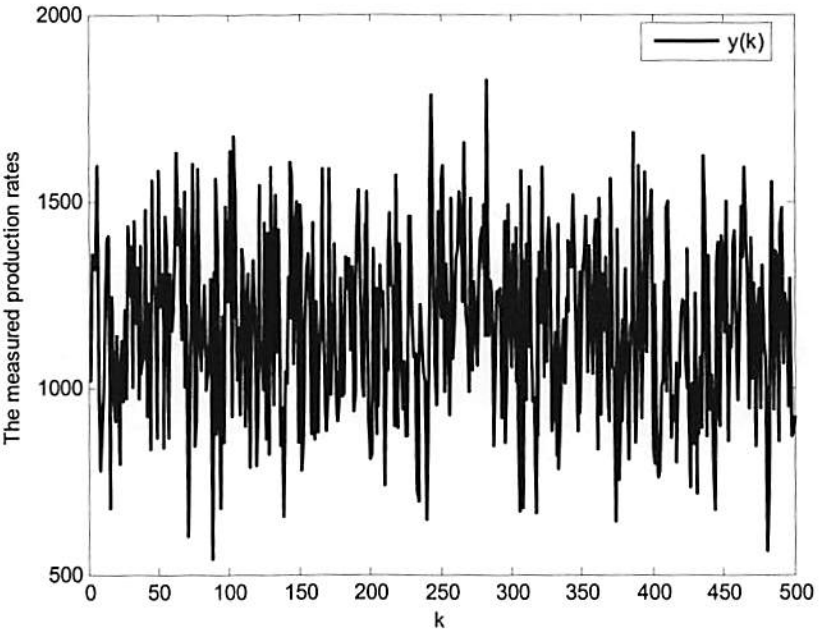


Figure 2.6: An example of generated measured production rates for Case 1.

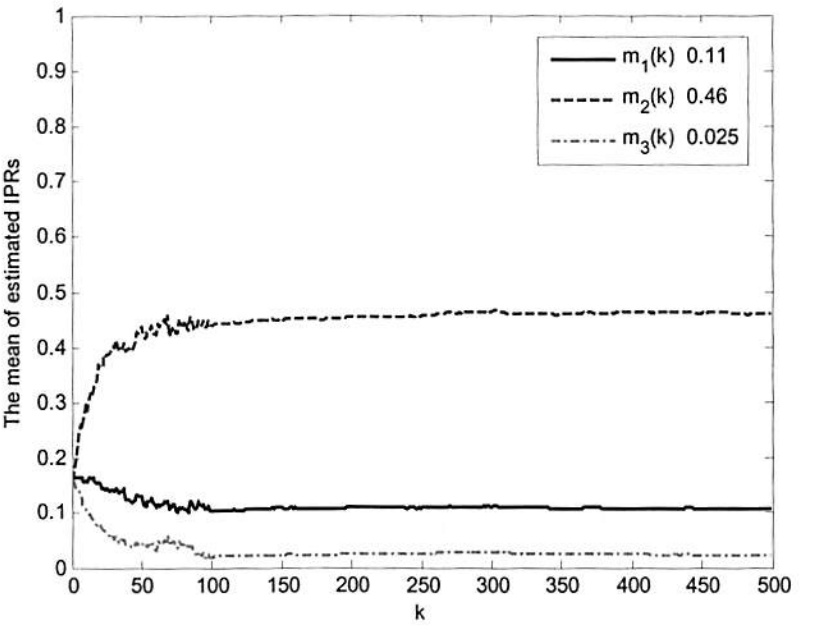


Figure 2.7: The mean of estimated $\hat{IPR}_j(k)$ for Case 1.

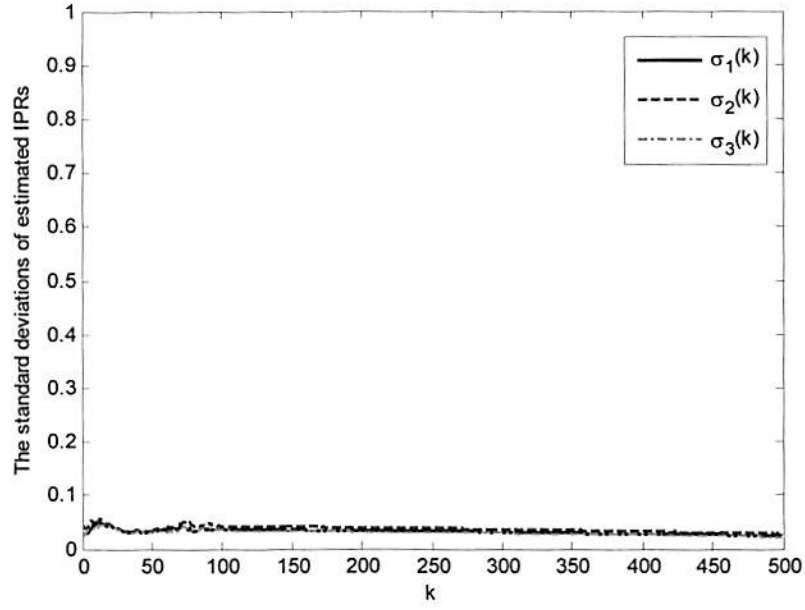


Figure 2.8: The standard deviations of estimated $\hat{IPR}_j(k)$ for Case 1.

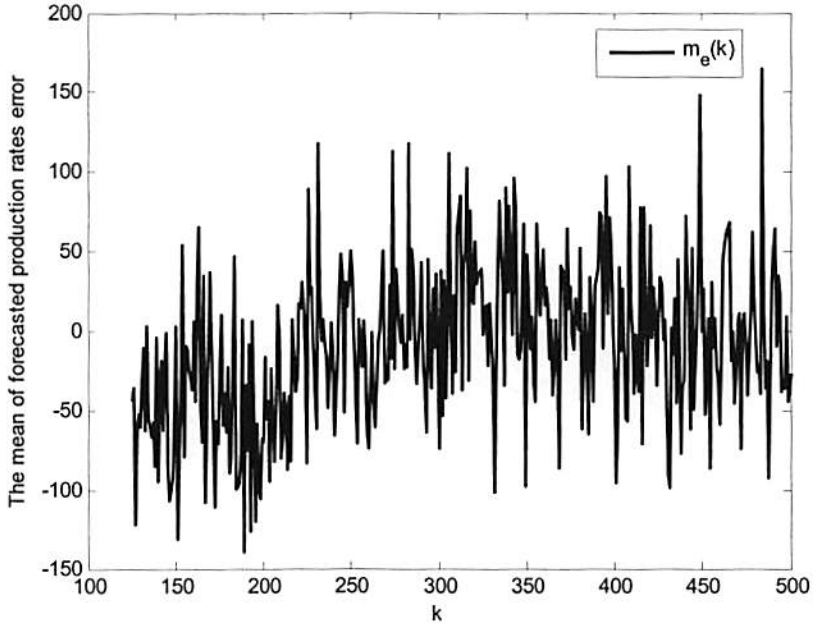


Figure 2.9: The mean $m_e(k)$ of forecasted production rates.

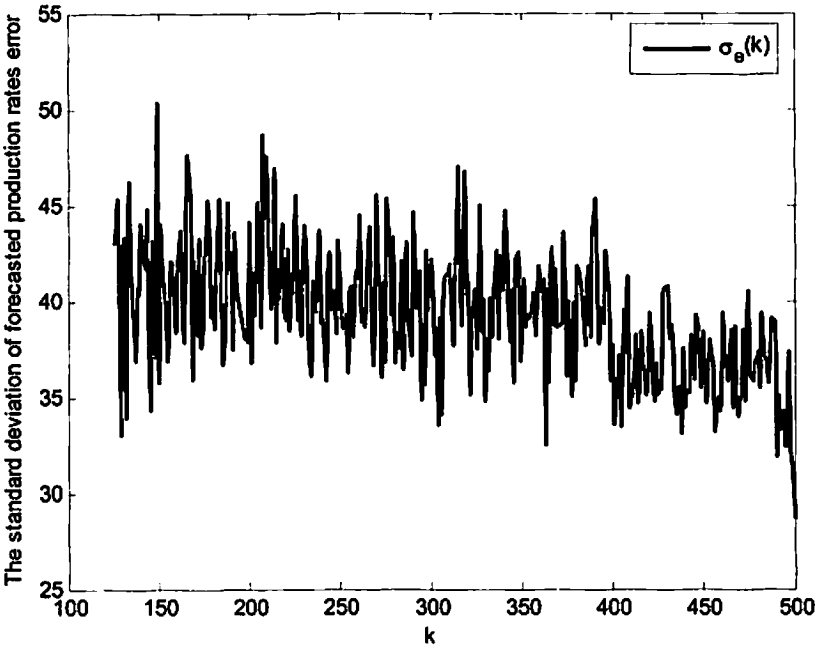


Figure 2.10: The standard deviations $\sigma_e(k)$ of forecasted production rates.

Case 2: Three injectors and a single producer with non-constant IPRs

Unlike Case 1, for which the IPRs were constant for all measurements, in this case, we assume that something such as a work-over has happened so that the reservoir model is different before and after $k = 1000$. The parameters for this case are summarized in Tables 2 and 3. Observe that, before $k = 1000$, injectors 1 and 2 have the largest and smallest impacts on the producer, respectively; and, after $k = 1000$, injector 2 becomes the most influential injector, whereas the influence from injector 1 reduces to around 45% of that from injector 2.

An example of measured production rates is depicted in Fig. 2.11. Test results are depicted in Figs. 2.12 and 2.13 for estimates of $IPR_j(k)$ ($j = 1,2,3$), and in Figs. 2.14

and 2.15 for the forecast error of production rates with $k_1 = 500$ and $L_{fw} = 90$, respectively. Note that the original values of $IPR_j(k)$ are provided in the legend of Fig. 2.12. Observe that:

1) Our approach can correctly estimate the relative IPRs, i.e., the largest $\hat{IPR}$ is obtained from the most influential injector, and the smallest $\hat{IPR}$ is obtained from least influential injector.
2) Around $k = 1000$, there is a change of the mean of the estimated $\hat{IPR}_j(k)$, which shows that our approach can correctly and quickly detect and track the changes of the IPRs.
3) The $m_e(k)$ and $\sigma_e(k)$ are much smaller than the generated measured production rates, which confirms that our approach can successfully forecast the future production rates.
4) Around $k = 1000$, the $m_e(k)$ and $\sigma_e(k)$ have a big change, which shows that the forecasting errors relate to the changes of the IPRs directly.

Table 2.2: Parameters for Case 2 before $k = 1000$.

j	$f(r_j, k_j)$	a_j	b_j	α_j	γ_j	IPR_j
1	1	200	300	0.1353	0.4060	0.5430
2	0.5	56	5.8	0.5703	0.0335	0.0907
3	0.8	100	41.8	0.3679	0.1539	0.3082

Table 2.3: Parameters for Case 2 after $k = 1000$.

j	$f(r_j, k_j)$	a_j	b_j	α_j	γ_j	IPR_j
1	0.7	200	120	0.1353	0.1624	0.1520
2	1	400	2400	0.0183	0.4396	0.4561
3	0.4	51	3.3	0.6	0.02	0.05

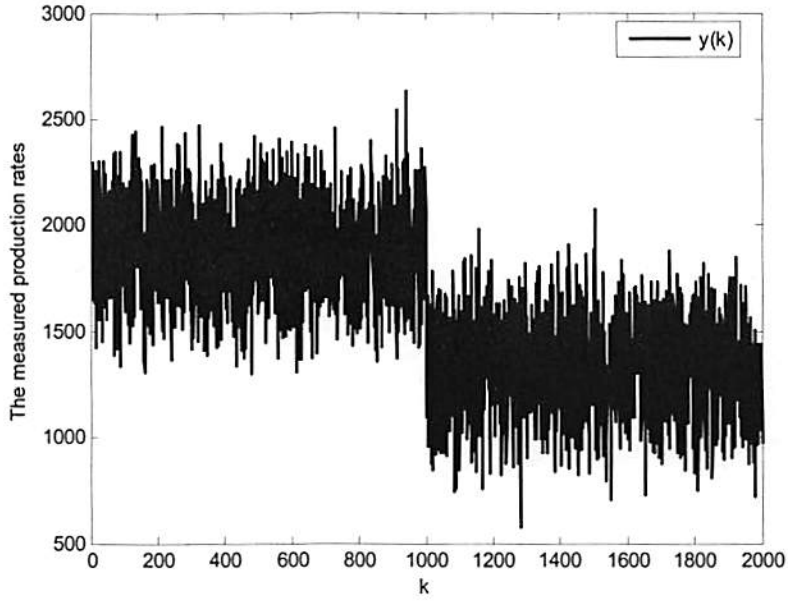


Figure 2.11: An example of generated measured production rates for Case 2.

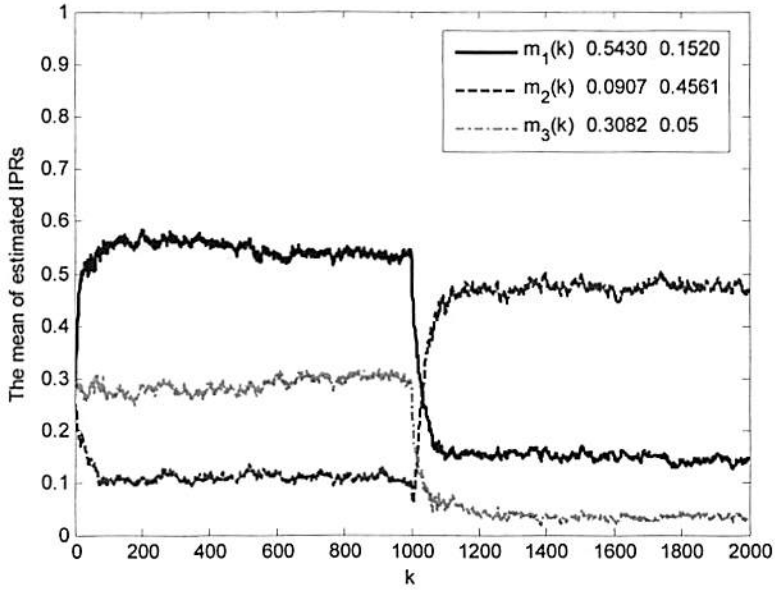


Figure 2.12: The mean of estimated $\hat{IPR}_j(k)$ for Case 2.

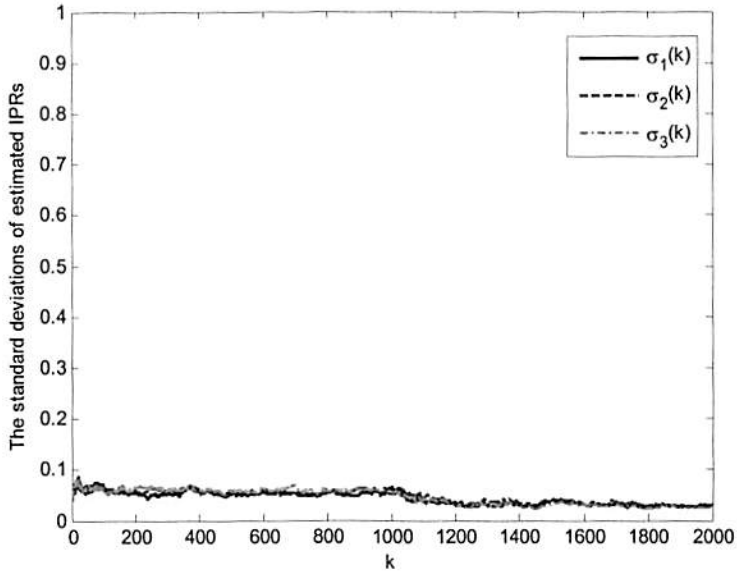


Figure 2.13: The standard deviations of the estimated $\hat{IPR}_j(k)$ for case 2.

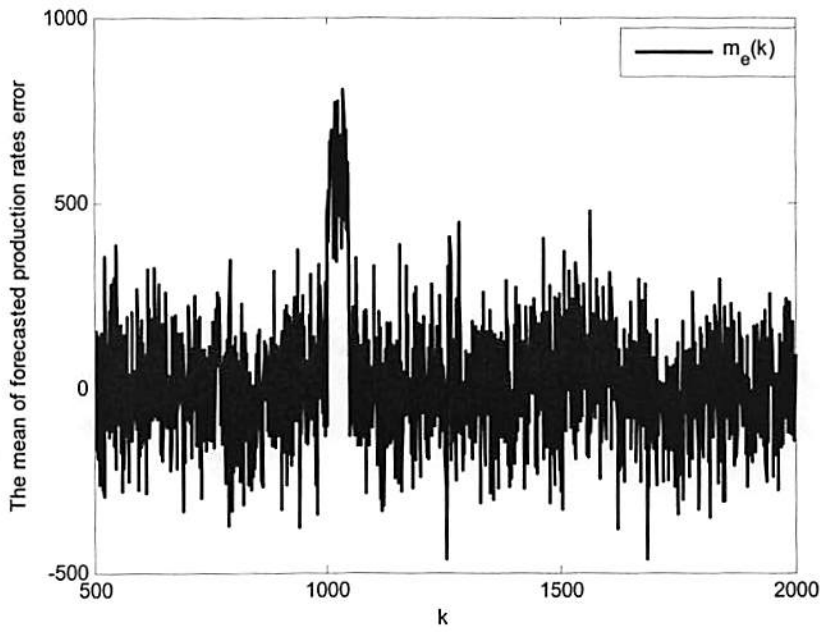


Figure 2.14: The mean $m_{error}(k)$ of forecasted production rates.

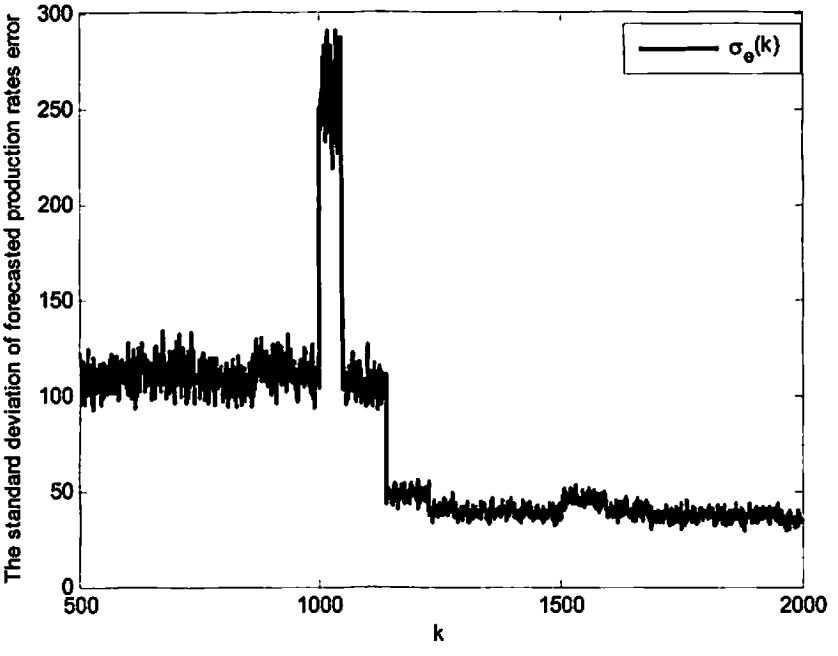


Figure 2.15: The standard deviations $\sigma_{error}(k)$ of forecasted production rates.

2.4.2 Simulation data

The simulation data were generated by Amir Mohammad Nejad using a commercial simulator. A simple five spot water injection pattern was adopted; and, the permeability, thickness and distance in a real oilfield were used to construct the models for generating simulation data. Different settings of permeability, thickness and distance were used, and data were generated for a broad range of models. Many tests were performed, and only one of them is presented herein.

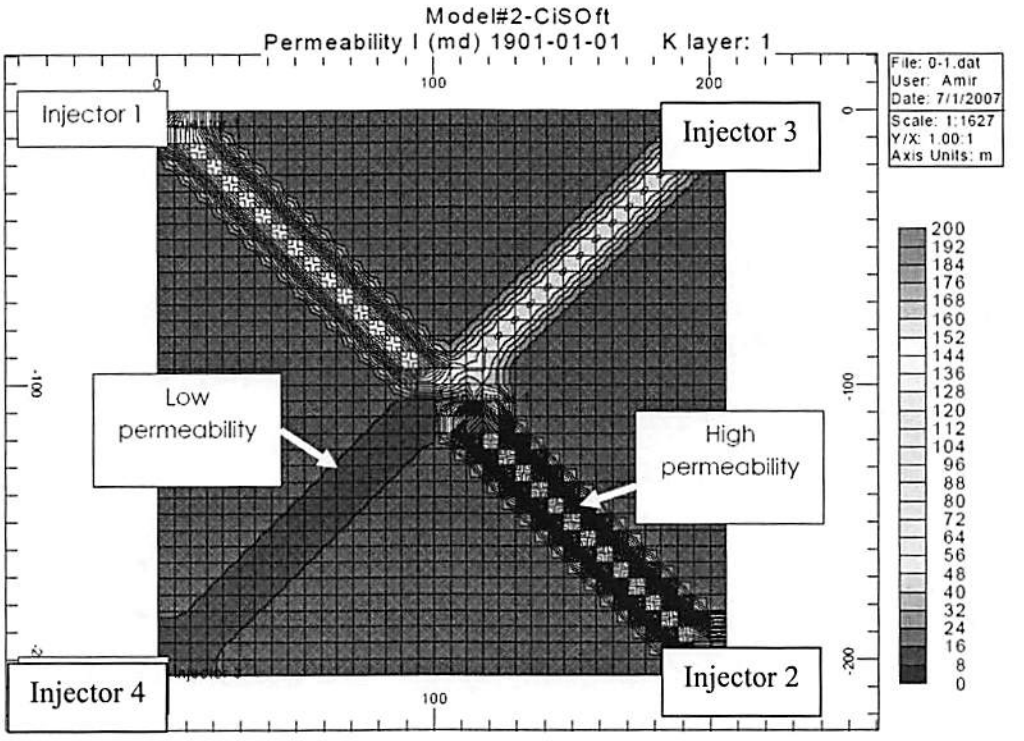


Figure 2.16: The reservoir model for simulation data generation.

Case 3: Five-Spot Static Reservoir

The reservoir model used for this case is depicted in Fig. 2.16. Observe that the permeability between injector 2 and producer has the highest value, then the permeability between injector 1 and producer, the permeability between injector 4 and producer, and the permeability between injector 3 and producer has the smallest value. Consequently, injector 2 has the largest impact on the producer; then, injectors 1, 4, 3 have smaller impacts on the producer in that order.

Test results are depicted in Fig. 2.17 for the estimated IPRs. Observe that the estimated $\hat{IPR}_j(k)$ match the model, i.e., the largest $\hat{IPR}_j(k)$ is obtained from the most

influential injector, and the smallest $\hat{IPR}_j(k)$ is obtained from the least influential injector.

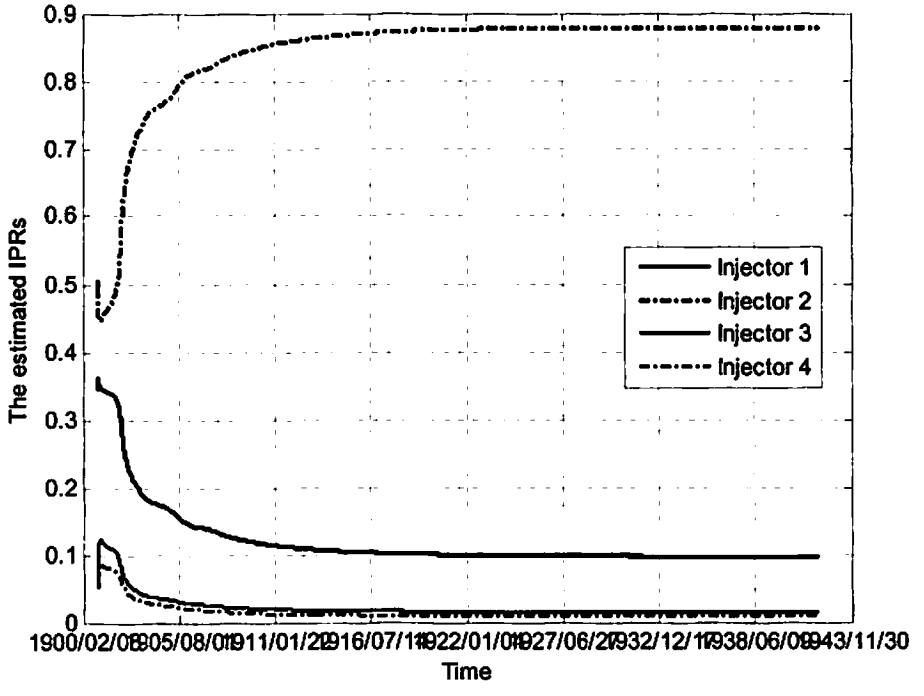


Figure 2.17: The estimated $\hat{IPR}_j(k)$ for Case 3.

2.4.2 Real data

A real dataset was provided by Chevron. It is weekly data from 30 injectors and 35 producers in Section 5 at Lost Hills. Recall that the objective of tests using real data is to see if our results match the expert knowledge. Two expert knowledges about the Lost Hills oilfield are: (1) Lost Hills is an oilfield with low permeability, so that injectors distant from the producer should have little impact upon it; and (2) there is a natural

fracture around North-East 55 degree at Chevron Lost Hill field. Many tests were performed, and two of them are presented herein.

Case 4: Reservoir model with two close injectors and three distant injectors

In this case, we use a two-step procedure to test our approach:

Step 1: We chose the producer *W5 12- 1A*, and its surrounding injectors *W5 12-1B* and *W5 11-1W*, for which $r_1 = 163.5532$ and $r_2 = 138.0518$, and obtained the estimated IPRs, as shown in Fig. 2.18; and,

Step 2: We add a distant injector (i.e., one that is far away from the producer such that one wouldn't anticipate it would have any influence on the producer.) at one time, and obtained the estimated IPRs depicted in Figs. 2.19 to 2.21, respectively, for three injectors *W5 11-2W, W5 11-1WA* and *W5 9-2W* with $r_3 = 639.6760, 556.7273$ and 1393.3.

From Figs. 2.18-2.21, observe that (1) the IPRs are dynamic, i.e., they change with time; (2) distant injectors impact the producer very little, which matches the first expert knowledge about Lost Hills that Lost Hills is an oilfield with low permeability, so that injectors distant from the producer should have little impact upon it; and, (3) the estimated IPRs of the surrounding injectors change very little when a distant injector is included.

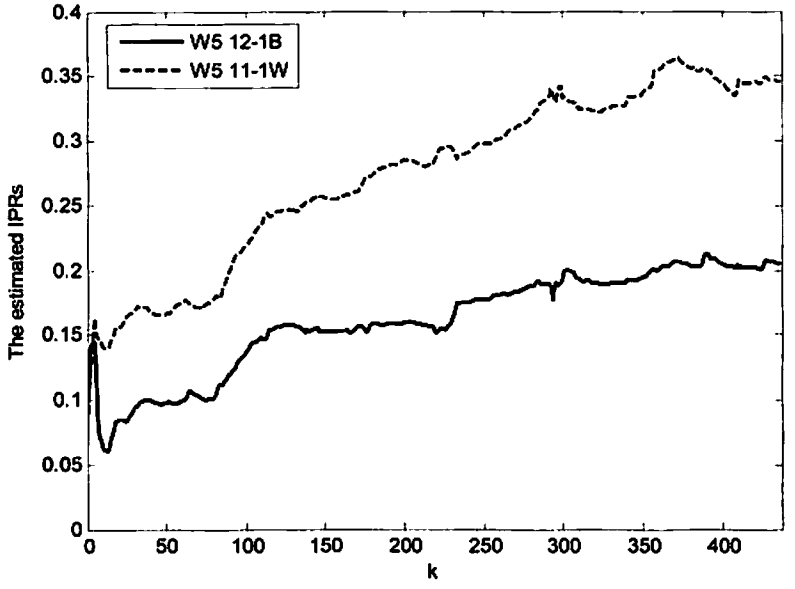


Figure 2.18: Estimated IPRs between producer *W5 12-1A* and its surrounding injectors *W5 12-1B* and *W5 11-1W*.

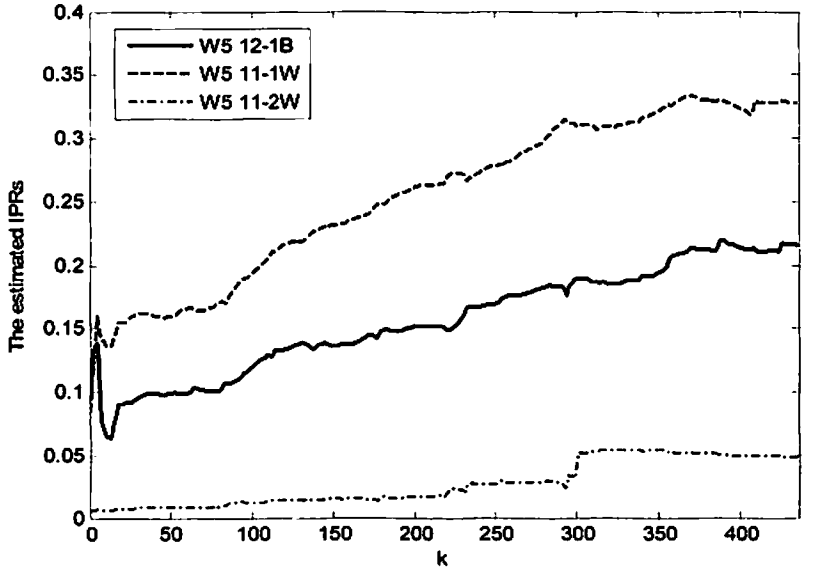


Figure 2.19: Estimated IPRs between producer *W5 12-1A* and its surrounding injectors *W5 12-1B* and *W5 11-1W* and a distant injector *W5 11-2W*.

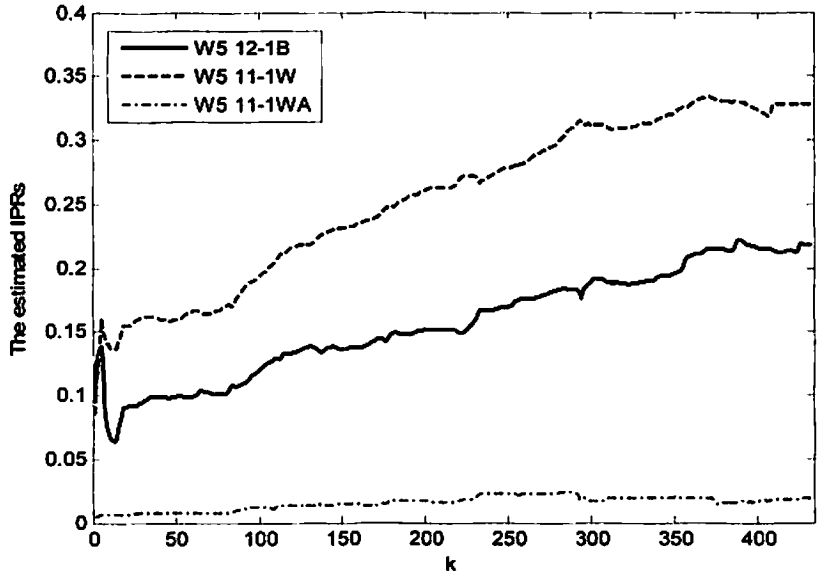


Figure 2.20: Estimated IPRs between producer *W5 12-1A* and its surrounding injectors *W5 12-1B* and *W5 11-1W* and a distant injector *W5 11-1WA*.

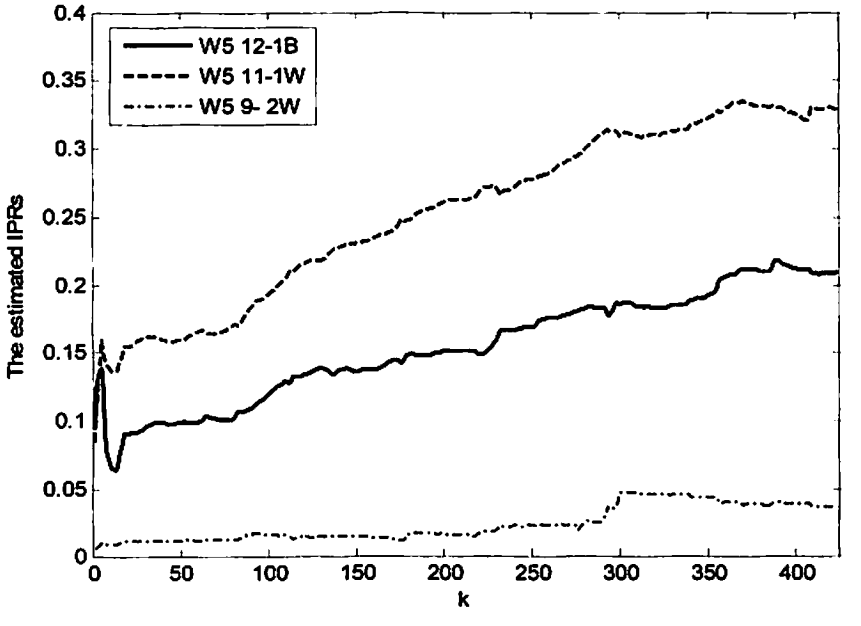


Figure 2.21: Estimated IPRs between producer *W5 12-1A* and its surrounding injectors *W5 12-1B* and *W5 11-1W* and a distant injector *W5 9-2W*.

Case 5: Reservoir with three close injectors.

For this case, we chose the producer *W5 10-3* and its surrounding injectors *W5 9-3W*, *W5 11-3WBS* and *W5 10-2WAS*, for which $r_1 = 333.5987$, $r_2 = 306.7820$ and $r_3 = 251.8346$, and obtain estimated IPRs depicted in Fig. 2.22.

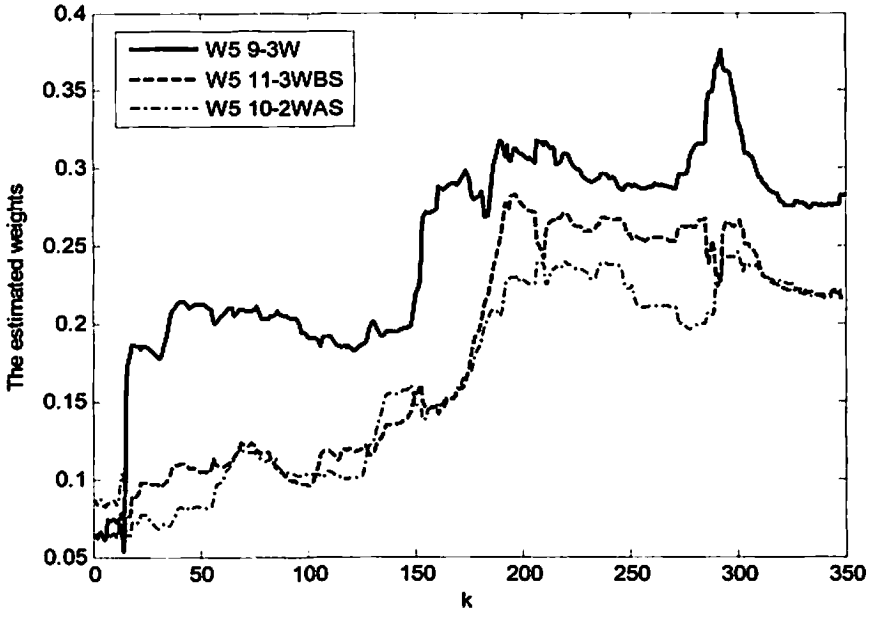


Figure 2.22: Estimated IPRs between producer *W5 10-3* and its surrounding injectors *W5 9-3W W5 11-3WBS* and *W5 10-2WAS*.

Observe that: (1) There is a change at about $k = 150$, which suggests something has happened, e.g., a work-over. This shows that our approach can quickly detect changes of the IPRs; (2) it appears that *W5 9-3W* has the most impact on the producer; however, from the distance, we know that this injector is the farthest one from the producer. An explanation for the second observation is there may be a fracture connection between

injector *W5 9-3W* and producer *W5 10-3,* which matches the second expert knowledge about Lost Hills, i.e., that there is a natural fracture around North-East 55 degree at Chevron Lost Hill field.

2.5 Conclusions

This chapter presents an adaptive method, using an extended Kalman filter (EKF), to forecast the injector-producer relationships (IPRs) between multiple injectors and a single producer based on measured production and injection rates.

Our approach has been tested on synthetic data, reservoir simulation data and real data. Test results on synthetic data and reservoir simulation data demonstrate the feasibility of the EKF method, and test results on the real data match expert knowledge about the IPRs between injectors and a producer. All results confirm that this EKF method can provide a good way to infer and track the IPRs so as to provide better insight about the IPRs.

Our approach may also be able to provide information about the directional sweep efficiency of a given pattern, a directional fracture, and may lead to a better understanding of zonal conformance and reservoir heterogeneity, all of which will assist Chevron engineers to maximize the oil production by changing the injection rates from different injectors.

Chapter 3

Inferring Regional Impact for Multiple Producers and Multiple Injectors, Using a Pseudo-virtual Reservoir Approach

3.1 Introduction

Current optimal decisions for water-flood management are made on a pattern basis, where a *pattern* generally consists of a group of injectors and producers. A *region* can be a single pattern or multiple patterns. The change of injection rate on a single injector may simultaneously have positive or negative impacts on multiple producers. Consequently, the impact of a single injector on a region may be more important than its impact on a single producer. If a regional impact is available, better decisions could be made to increase or decrease the injection rates for a particular injector in the region, such that the total oil production from the region is increased. In this thesis, our focus is on inferring the regional impact from each injector. By doing so, we will obtain a collection of tools that are able to provide flexibility for decision making under different scales, e.g., a region with a single producer, a collection of producers, a pattern, or a collection of patterns.

3.2 Pseudo-virtual reservoir approach

This section begins with the reservoir model, derives the virtual reservoir model and finally introduces the pseudo-virtual reservoir model.

Consider the simple example of a region depicted in Fig. 3.1, in which there are four injectors that surround two producers. Instead of determining the impact on a particular producer from each injector, which can be done by using a straight-forward extension of our results in Chapter 2 from one producer to more than one producer, our objective is to determine the regional impact from each injector on the region that encircles the two producers. The approach proposed in this report builds upon our previous work in Chapter 2 for multiple injectors and a single producer.

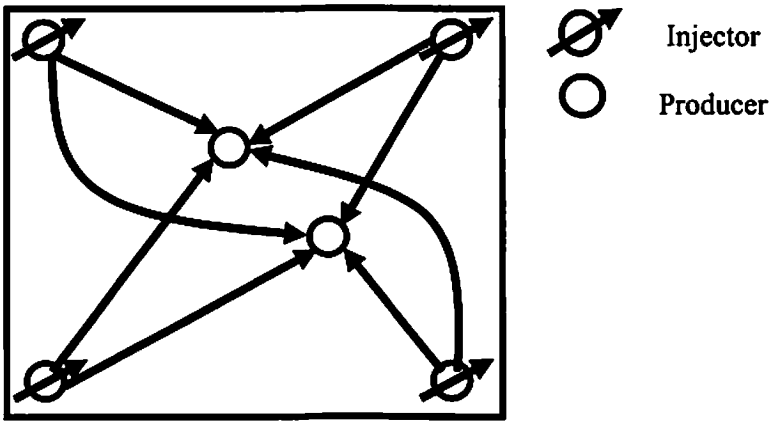


Figure 3.1: An example of a region with four injectors and two producers.

3.2.1 Reservoir model

For generality, we assume that there are N injectors and M producers in a region. The reservoir model for such a region is depicted in Fig. 3.2, in which the N injectors

and each producer pair is considered as an independent subsystem; hence, for a reservoir with N injectors and M producers, the reservoir is represented by M independent subsystems, where each subsystem is modeled as the reservoir model depicted in Fig. 2.1. The notations for this reservoir model are:

1) $i_j(t)$, $n_j(t)$ and $i_{m,j}(t)$ ($j=1,\ldots,N$) are the actual injection rates that flow into the reservoir, their corresponding measurement noises, and the measured injection rates, respectively;
2) $f(r_{j,l},k_{j,l})$ ($j=1,\ldots,N$; $l=1,\ldots,M$) are linear or nonlinear scalar functions of $r_{j,l}$ and $k_{j,l}$, which relate to the connectivities between injectors and producers, where $r_{j,l}$ and $k_{j,l}$ are the distance and permeability of the reservoir channel between injector j and producer l, respectively;
3) $h_{j,l}(t)$ ($j=1,\ldots,N$; $l=1,\ldots,M$) is the channel impulse response for the reservoir channel between injector j and producer l; it is modeled by a continuous-time uni-modal function, similar to (2.4), i.e., $h_{j,l}(t)=b_{j,l}te^{-a_{j,l}t}$;
4) $p_l(t)$, $n_{p_l}(t)$ and $p_{m,l}(t)$ are the actual production rate, their corresponding measurement noise, and the measured production rate for producer l, respectively;
5) $P_l(s)$, $I_j(s)$ and $H_{j,l}(s)$ are the Laplace transforms of $p_l(t)$, $i_j(t)$, and $h_{j,l}(t)$, respectively; and
6) $P_l(z)$, $I_j(z)$ and $H_{j,l}(z)$ are the z-transforms of $p_l(n)$, $i_j(n)$, and $h_{j,l}(n)$, respectively.

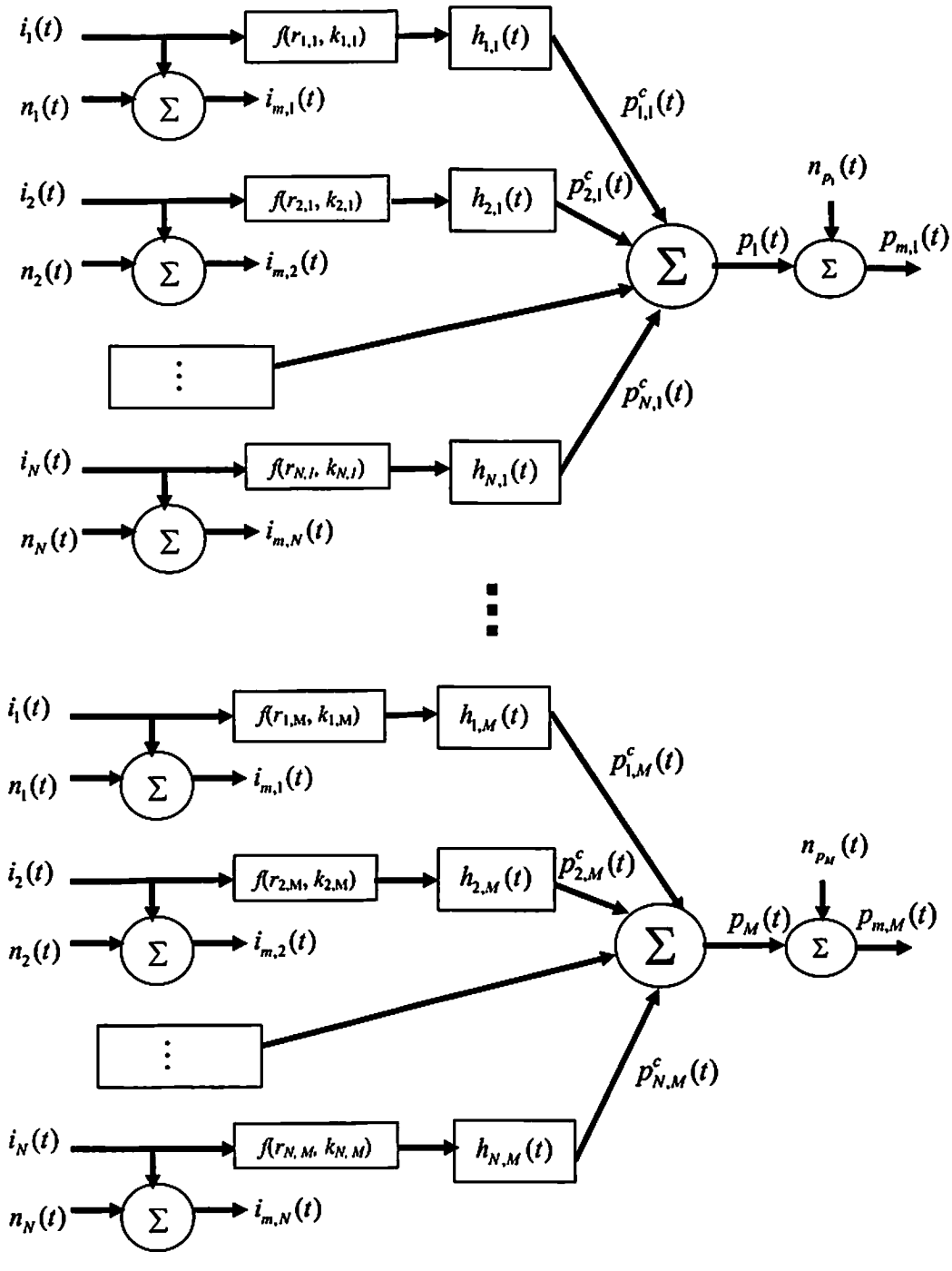


Figure 3.2: Reservoir model for N injectors and M producers.

3.2.2 Virtual reservoir model

To begin, we introduce some new definitions.

Definition 1: A *regional producer* is a pseudo-producer that comprises all producers in a region, with *actual regional production rate*, $p^R(t)$, and *measured regional production rates*, $p_m^R(t)$, given by:

$$p^R(t) = \sum_{l=1}^{M} p_l(t) \tag{3.1}$$

$$p_m^R(t) = \sum_{l=1}^{M} p_{m,l}(t) = \sum_{l=1}^{M} \left(p_l(t) + n_{p_l}(t) \right) = \sum_{l=1}^{M} p_l(t) + \sum_{l=1}^{M} n_{p_l}(t) = p^R(t) + n_{p^R}(t) \tag{3.2}$$

where $n_{p^R}(t) = \sum_{l=1}^{M} n_{p_l}(t)$ is the measurement noise for the actual regional production rate.

Definition 2: A *virtual channel* is a virtual communication channel between an injector and the regional producer, which converts the injection rates into regional production rates.

An example of regional producer and virtual channels for the reservoir depicted in Fig. 3.1 is shown in Fig. 3.3, in which the two producers are replaced by a single regional producer, and the dashed line from each injector to the regional producer denotes a virtual channel. We refer to this reservoir as a virtual reservoir.

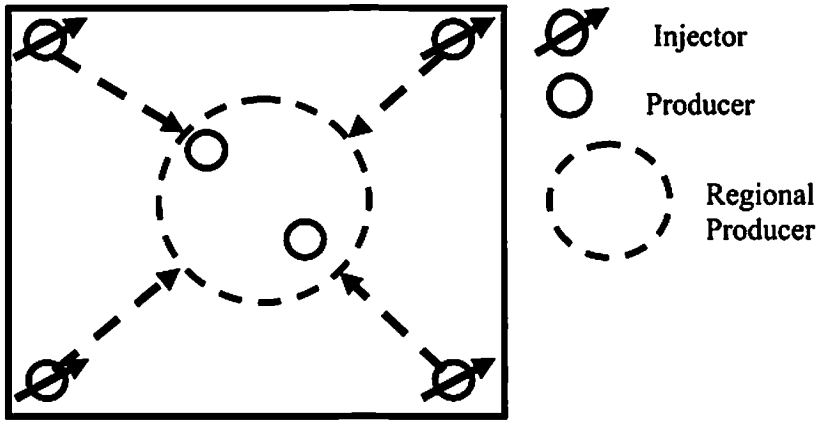


Figure 3.3: An example of virtual reservoir with four injectors and two producers.

Taking the Laplace- and z- transforms of both sides of (3.1), it follows that

$$P^R(s) = \sum_{l=1}^{M} P_l(s). \tag{3.3}$$

and

$$P^R(z) = \sum_{l=1}^{M} P_l(z). \tag{3.4}$$

It will be seen later that:

(1) $i_j(t)$ $\left(j=1,\ldots,N\right)$ and $p_l(t)$ $\left(l=1,\ldots,M\right)$ are used for constructing the virtual reservoir model; and $i_{m,j}(t)$ and $p_{m,l}(t)$ are used for estimating the parameters of the virtual reservoir model during our processing, because the noise-free signals $i_j(t)$ and $p_l(t)$ are not directly available;

(2) $P_l(s)$, $P^R(s)$, $I_j(s)$ and $H_{j,l}(s)$ are used for constructing the virtual reservoir model, and $P_l(z)$, $P^R(z)$, $I_j(z)$ and $H_{j,l}(z)$ are used for estimating the parameters of the virtual reservoir model during our processing, because only discrete measurement data are available for such processing.

Using (2.3), but for the l^{th} producer, $P_l(s)$ can be expressed as

$$P_l(s) = \sum_{j=1}^{N} f(r_{j,l}, k_{j,l}) H_{j,l}(s) I_j(s) \tag{3.5}$$

Substituting (3.5) into (3.3), it follows that

$$\begin{aligned} P^R(s) &= \sum_{l=1}^{M} P_l(s) = \sum_{l=1}^{M} \sum_{j=1}^{N} f(r_{j,l}, k_{j,l}) H_{j,l}(s) I_j(s) \\ &= \sum_{j=1}^{N} \left\{ \sum_{l=1}^{M} f(r_{j,l}, k_{j,l}) H_{j,l}(s) \right\} I_j(s) \\ &= \sum_{j=1}^{N} H_j^v(s) I_j(s) \end{aligned} \tag{3.6}$$

where $H_j^v(s)$ is the Laplace transform of the virtual channel between injector j and the regional producer, i.e.,

$$H_j^v(s) = \sum_{l=1}^{M} f(r_{j,l}, k_{j,l}) H_{j,l}(s) \tag{3.7}$$

Taking the inverse Laplace transform of both sides of (3.7), it follows that the virtual impulse response $h_j^v(t)$ can be expressed as

$$h_j^v(t) = \sum_{l=1}^{M} f(r_{j,l}, k_{j,l}) h_{j,l}(t). \tag{3.8}$$

which shows that the impulse response of a virtual channel is the weighted sum of M actual channel impulse responses. Although each $h_{j,l}(t)$ is characterized by a two parameter uni-modal function [see (2.4)], $h_j^v(t)$ will not be a two parameter uni-modal function. Instead, it is a sum of M AR model and has $2M$ parameters. An example of $h_j^v(t)$ for two producers (i.e., $M = 2$) is shown in Fig. 3.4, where the dotted and dashed curves represent the scaled impulse responses from injector j to producers 1 and 2, respectively, and the solid curve (which is the sum of these two curves) represents the virtual impulse response $h_j^v(t)$.

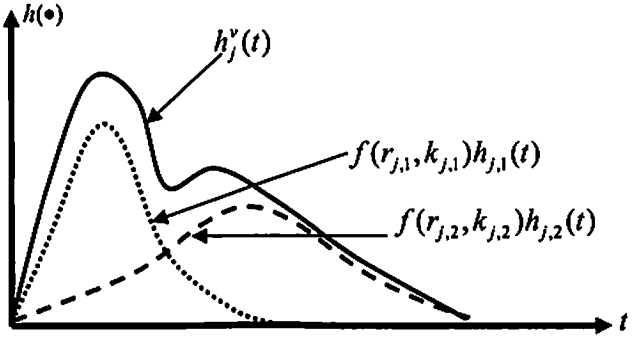


Figure 3.4: An example of virtual impulse response $h_j^v(t)$.

Let $p_j^v(t)$ be the virtual channel production rate due to injector j and the j^{th} virtual channel, its Laplace transform is given by

$$P_j^v(s) = H_j^v(s)I_j(s). \tag{3.9}$$

Substituting (3.9) into (3.6), it follows that the actual reservoir model can be expressed as:

$$P^R(s) = \sum_{j=1}^{N} P_j^v(s) = \sum_{j=1}^{N} H_j^v(s)I_j(s) \tag{3.10}$$

Observe from (3.10) that, by collecting multiple producers in a region as a regional producer, and by introducing virtual channels with virtual impulse responses, the actual reservoir model can be reformulated as a *virtual reservoir model*, as depicted in Fig. 3.5.

Comparing Figs. 3.5 and 2.1, observe that the virtual reservoir model is actually a reservoir model for multiple injectors and a single producer with $f(r_j,k_j)=1$ and $h_j(t) = h_j^v(t)$; hence, all of the results for the latter model that are in Section 2.2 can be used for the former model. Note, also, from (3.10), that the $f(r_{j,l},k_{j,l})$ have been absorbed into $h_j^v(t)$.

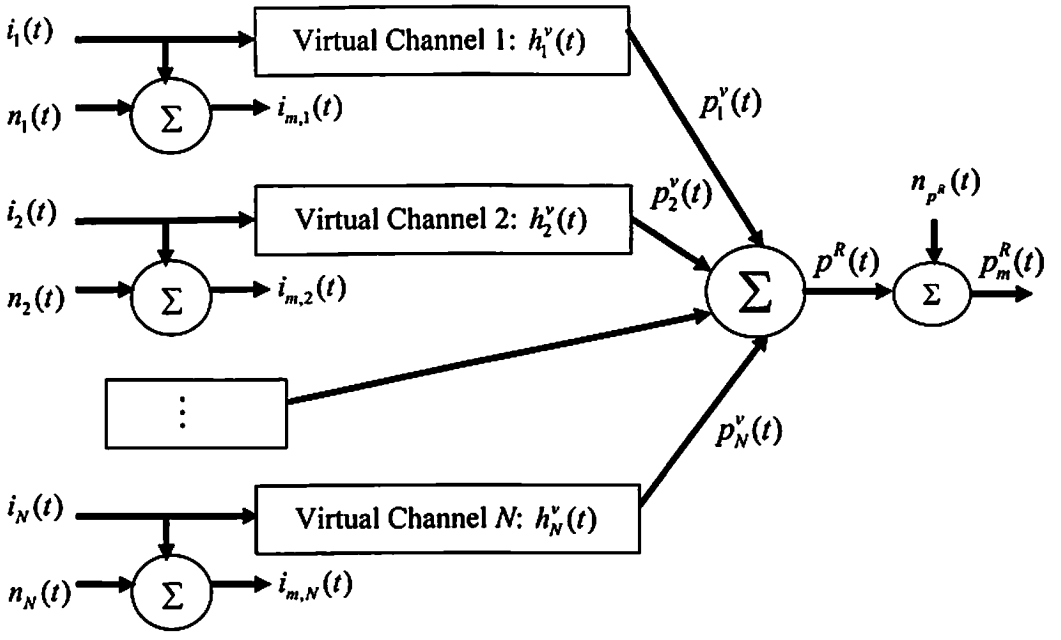


Figure 3.5: Virtual Reservoir Model for N injectors and M producers.

Definition 3: Let IP_j^R denotes the *regional impact*, then, analogous to IPR_j in (2.14),

$$\begin{aligned} IP_j^R = \frac{1}{T}\int_{t=0}^{\infty} h_j^v(t)dt \approx H_j^v(z)|_{z=1} &= \sum_{l=1}^{M} f(r_{j,l},k_{j,l})H_{j,l}(z)|_{z=1} \\ &= \sum_{l=1}^{M} IPR_{j,l} \end{aligned} \tag{3.11}$$

Note that the right-most side of the first row of (3.11) is obtained using (3.7), when (3.7) is expressed using z-transform, and $IPR_{j,l}$, which equals to $f(r_{j,l},k_{j,l})H_{j,l}(z)|_{z=1}$, is obtained using (2.14).

3.2.3 Pseudo-virtual reservoir model

As seen from (3.11), to estimate the regional impact IP_j^R one needs to estimate the virtual channel impulse response $h_j^v(n)$, which is the sum of M AR models having $2M$ unknown parameters. In order to simplify the estimation process, we reduce the number of unknown parameters by approximating the virtual channel $H_j^v(z)$ as a product of a pseudo-virtual channel $H_j^{pv}(z)$ and a pseudo-scale function $f(r_j^{pv}, k_j^{pv})$, i.e.,

$$H_j^v(z) \approx f(r_j^{pv}, k_j^{pv}) H_j^{pv}(z) \tag{3.12}$$

where

$$H_j^{pv}(z) = \frac{\gamma_j^{pv} z^{-1}}{(1-\alpha_j^{pv} z^{-1})^2} \tag{3.13}$$

and, r_j^{pv} and k_j^{pv} are the pseudo-distance and pseudo-permeability between the injector j and the regional producer. This approximation is made by equating the area under $h_j^v(t)$ to the area under $f(r_j^{pv}, k_j^{pv}) H_j^{pv}(z)$. An example is shown in Fig. 3.6, where the solid curve represents $h_j^v(t)$, the dash curve represents $f(r_j^{pv}, k_j^{pv}) H_j^{pv}(z)$.

Using $f(r_j^{pv}, k_j^{pv}) H_j^{pv}(s)$, $P^R(s)$ in (3.10) can be approximated, as

$$P^R(s) = \sum_{j=1}^{N} P_j^v(s) = \sum_{j=1}^{N} H_j^v(s) I_j(s) \approx \sum_{j=1}^{N} f(r_j^{pv}, k_j^{pv}) H_j^{pv}(s) I_j(s) \tag{3.14}$$

This demonstrates that the virtual reservoir model in Fig. 3.4 can be further reformulated as the pseudo-virtual reservoir model depicted in Fig. 3.7.

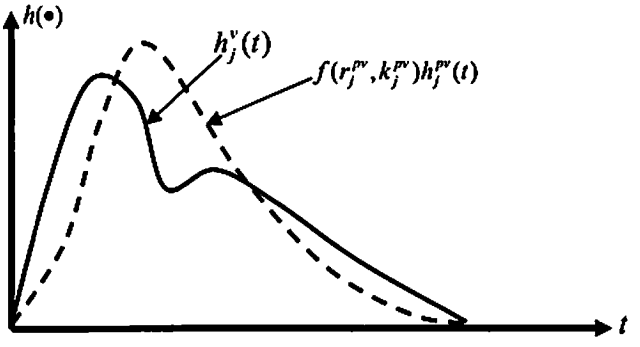


Figure 3.6: An example of the virtual impulse response $h_j^v(t)$ and the scaled pseudo-impulse response $f(r_j^{pv},k_j^{pv})h_j^{pv}(t)$.

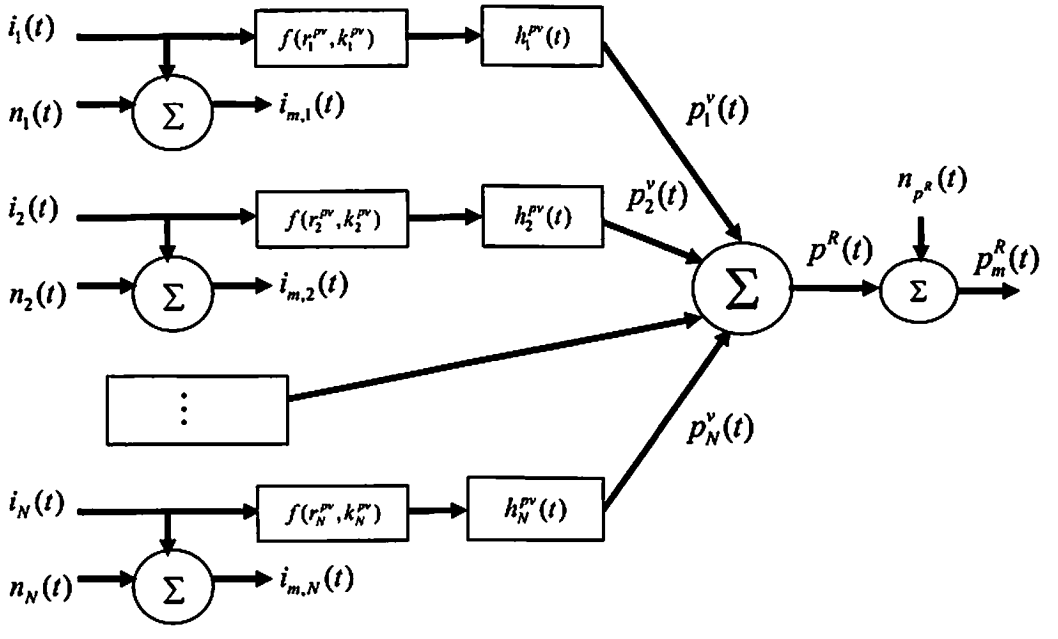


Figure 3.7: Pseudo- virtual reservoir model for N injectors and M producers.

Comparing Figs. 3.7 and 2.1, observe that the pseudo-virtual reservoir model is very similar to the reservoir model for multiple injectors and a single producer, as described in Section 2.2, except that:

1) The scale function $f(r_j,k_j)$ is replaced by the pseudo-scale function $f(r_j^{pv},k_j^{pv})$;
2) The channel $h_j(t)$ is replaced by the pseudo-virtual channel $h_j^{pv}(t)$; and,
3) The production rate $p(t)$ is replaced by the regional production rate $p^R(t)$.

Substituting the approximation of (3.12) into (3.12), it follows that the regional impact can be further approxiamted as

$$IP_j^R \approx H_j^v(z)|_{z=1} \approx f(r_j^{pv},k_j^{pv})H_j^{pv}(z)|_{z=1} \tag{3.15}$$

As in (2.14) and (2.15), (3.15) can be expressed as:

$$IP_j^R \approx f(r_j^{pv},k_j^{pv})H_j^{pv}(z)|_{z=1} = \frac{f(r_j^{pv},k_j^{pv})\gamma_j^{pv}}{(1-\alpha_j^{pv})^2} = \frac{\gamma_j^{pv'}}{(1-\alpha_j^{pv})^2} \tag{3.16}$$

where $\gamma_j^{pv'} = f(r_j^{pv},k_j^{pv})\gamma_j^{pv}$. Consequently, by replacing $h_j^v(t)$ with $f(r_j^{pv},k_j^{pv})h_j^{pv}(t)$, the estimation of IP_j^R has been tremendously simplified because the number of parameters that need to be estimated have been reduced from $2M$ to 2. In short, the

estimation of IP_j^R for a region becomes the same as the estimation of IPR_j for multiple injectors and a single producer.

Using the available sampled measurements of injection and production rates, our goals are:

1) Estimate the pseudo-virtual impulse response parameters $\gamma_j^{pv'}$ and α_j^{pv} $(j=1,\ldots,N)$; and,

2) Compute $\hat{IP}_j^R$ using $\hat{\gamma}_j^{pv'}$ and $\hat{\alpha}_j^{pv}$ $(j=1,\ldots,N)$;

These goals will be achieved by using the EKF estimator.

3.2.4 Estimator using Extended Kalman Filter (EKF estimator)

A. State variable model (SVM)

The derivation of the state variable model (SVM) for this pseudo-virtual reservoir model is similar to that in Section 2.3.1; however, for easy reference, we still explain the derivation process herein but using the notations for the pseudo-virtual channels. To illustrate how to do this, we consider the simple case of three injectors and multiple producers (it is easy to generalize the *SVM* from three injectors to *N* injectors.).

Observe from (3.9), (3.12) and (3.13) that for each pseudo-virtual channel:

$$\frac{P_j^v(z)}{I_j(z)} \approx \frac{f(r_j^{pv}, k_j^{pv})\gamma_j^{pv} z^{-1}}{(1-\alpha_j^{pv} z^{-1})^2} = \frac{\gamma_j^{pv'} z^{-1}}{1-2\alpha_j^{pv} + (\alpha_j^{pv})^2 z^{-2}} \tag{3.17}$$

which can be further expressed as:

$$\left(1-2\alpha_j^{pv}+(\alpha_j^{pv})^2 z^{-2}\right)P_j^v(z)\approx \gamma_j^{pv'} z^{-1} I_j(z) \tag{3.18}$$

Multiply z on both sides of (3.18) and determine the inverse z-transform of the resulting equation, to see that (3.18) can be expressed in the time-domain, as:

$$p_j^v(k+1)-2\alpha_j^{pv} p_j^v(k)+(\alpha_j^{pv})^2 p_j^v(k-1)\approx \gamma_j^{pv'} i_j(k) \tag{3.19}$$

Because only measured injection rates, $i_{m,j}(k)$, are available, where

$$i_{m,j}(k)=i_j(k)+n_j(k), \tag{3.20}$$

it follows that (3.19) becomes:

$$p_j^v(k+1)-2\alpha_j^{pv} p_j^v(k)+(\alpha_j^{pv})^2 p_j^v(k-1)\approx \gamma_j^{pv'} i_{m,j}(k)-\gamma_j^{pv'} n_j(k) \tag{3.21}$$

Letting $n_{m,j}(k)\equiv -\gamma_j^{pv'}{}_j n_j(k)$, (3.21) simplifies to

$$p_j^v(k+1)-2\alpha_j^{pv} p_j^v(k)+(\alpha_j^{pv})^2 p_j^v(k-1)\approx \gamma_j^{pv'} i_{m,j}(k)+n_{m,j}(k) \tag{3.22}$$

(3.22) is a second-order finite-difference equation that is described by two state variables $p_j^v(k-1)$ and $p_j^v(k)$. In addition, the unknown parameters, $\gamma_j^{pv'}$ and α_j^{pv} are also treated as state variables so that they can be estimated by an EKF, and their state equation models are

$$\left.\begin{aligned}\alpha_j^{pv}(k+1) &= \alpha_j^{pv}(k) + n_{\alpha_j^{pv}}(k) \\ \gamma_j^{pv'}(k+1) &= \gamma_j^{pv'}(k) + n_{\gamma_j^{pv'}}(k)\end{aligned}\right\} \tag{3.23}$$

where $n_{\alpha_j^{pv}}(k)$ and $n_{\gamma_j^{pv'}}(k)$ are discrete-time zero-mean white random noises. Eqs. (3.22) and (3.23) together are described by a 4×1 state vector x_j, where

$$x_j = \left[\alpha_j^{pv}(k), \gamma_j^{pv'}(k), p_j^v(k-1), p_j^v(k)\right]' \tag{3.24}$$

Using standard techniques [19], the following *SVM* is obtained for each $p_j^v(k)$ ($j = 1,2,3$):

$$\begin{cases} x_j(k+1) = \begin{bmatrix} x_{j1}(k+1) \\ x_{j2}(k+1) \\ x_{j3}(k+1) \\ x_{j4}(k+1) \end{bmatrix} = \begin{bmatrix} \alpha_j^{pv}(k+1) \\ \gamma_j^{pv'}(k+1) \\ p_j^v(k) \\ p_j^v(k+1) \end{bmatrix} = \begin{bmatrix} x_{j1}(k) \\ x_{j2}(k) \\ x_{j4}(k) \\ 2x_{j1}(k)x_{j4}(k) - x_{j1}^2(k)x_{j3}(k) \\ +x_{j2}(k)i_{m,j}(k) \end{bmatrix} + n_{x_j}(k) \\ p_j^v(k+1) = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} x_j(k+1) \end{cases} \tag{3.25}$$

where

$$\boldsymbol{n}_{x_i}(k) = [n_{\alpha_j^{p_r}}(k),\ n_{\gamma_j^{p_r}}(k),\ 0,\ n_{m,j}(k)]' \tag{3.26}$$

is a 4×1 zero-mean white noise vector. Letting $\boldsymbol{x}(k+1) \equiv col\left[\boldsymbol{x}_1(k+1),\ \boldsymbol{x}_2(k+1), \boldsymbol{x}_3(k+1)\right]$, the complete SVM for three injectors is:

$$\begin{cases} \boldsymbol{x}(k+1) = \begin{bmatrix} x_{11}(k+1) \\ x_{12}(k+1) \\ x_{13}(k+1) \\ x_{14}(k+1) \\ x_{21}(k+1) \\ x_{22}(k+1) \\ x_{23}(k+1) \\ x_{24}(k+1) \\ x_{31}(k+1) \\ x_{32}(k+1) \\ x_{33}(k+1) \\ x_{34}(k+1) \end{bmatrix} = \begin{bmatrix} x_{11}(k) \\ x_{12}(k) \\ x_{14}(k) \\ 2x_{11}(k)x_{14}(k) - x_{11}^2(k)x_{13}(k) + x_{12}(k)i_{m,1}(k) \\ x_{21}(k) \\ x_{22}(k) \\ x_{24}(k) \\ 2x_{21}(k)x_{24}(k) - x_{21}^2(k)x_{23}(k) + x_{22}(k)i_{m,2}(k) \\ x_{31}(k) \\ x_{32}(k) \\ x_{34}(k) \\ 2x_{31}(k)x_{34}(k) - x_{31}^2(k)x_{33}(k) + x_{32}(k)i_{m,3}(k) \end{bmatrix} + \boldsymbol{n}_x(k) \\ p_m^R(k+1) = p_1^v(k+1) + p_2^v(k+1) + p_3^v(k+1) + n_p(k+1) \\ \qquad = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \boldsymbol{x}(k+1) + n_{p^R}(k+1) \end{cases} \tag{3.27}$$

where $p_m^R(k+1)$ is the measured regional production rate, $\boldsymbol{n}_x(k) = col[\boldsymbol{n}_{x_1}(k), \boldsymbol{n}_{x_2}(k), \boldsymbol{n}_{x_3}(k)]$ and $n_{p^R}(k+1)$ are additive zero-mean white noises

for the state and measurement equations with covariance matrix $\boldsymbol{Q}_k$ and variance r_{k+1}, respectively, where

$$\boldsymbol{Q}_k = diag\left[r_{n_{\hat{\alpha}_1}},\ r_{n_{\hat{\gamma}_1}},\ 0,\ r_{n_{m,1}}, r_{n_{\hat{\alpha}_2}},\ r_{n_{\hat{\gamma}_2}},\ 0,\ r_{n_{m,2}}, r_{n_{\hat{\alpha}_3}},\ r_{n_{\hat{\gamma}_3}},\ 0,\ r_{n_{m,3}} \right] \tag{3.28}$$

Because this SVM is nonlinear, we can't use a Kalman filter to estimate its twelve states; Instead, we use an EKF to estimate them. After all 12 states are estimated, we extract the estimates of the six virtual reservoir model parameters from them, namely $\hat{\alpha}_j^{pv}(k+1|k+1)$ and $\hat{\gamma}_j^{pv'}(k+1|k+1)$ ($j=1,2,3$), after which we compute $I\hat{P}_j^R(k+1|k+1)$ ($j=1,2,3$) as

$$I\hat{P}_j^R(k+1|k+1) = \frac{\hat{\gamma}_j^{pv'}(k+1|k+1)}{(1-\hat{\alpha}_j^{pv}(k+1|k+1))^2} \tag{3.29}$$

In this way, we see that $I\hat{P}_j^R(k+1|k+1)$ ($j=1,2,3$) are functions of time, so that the EKF will provide dynamic update values for $I\hat{P}_j^R$, which can be very useful if there is a change to the reservoir.

B. EKF approach

The EKF approach herein is exactly the same as described in Section 2.3.2; hence, we omit it, and refer the reader to Section 2.3.2.

3.3 Simulations

The pseudo-virtual reservoir method will be tested using synthetic data, reservoir simulation data and real data. The tests using synthetic data, described below, evaluate the feasibility of our approach. The tests using reservoir simulation data, to be performed at a later date and not as part of this dissertation, will confirm whether or not our estimated results match the models used for the reservoir simulation data. The tests using real data, also to be performed at a later date and not as part of this dissertation, will confirm whether or not our results match some expert knowledge or actual field test result.

Table 3.1: Summarization of the settings for the test cases (For all nine cases, there are two producers in the region).

Case #	Name	Number of close injectors	Number of distant injectors	Constant or non-constant IPRs	Low or high impact from distant injectors
1	Static reservoir case 1	3	0	Constant	NA
2	Static reservoir case 2	3	0	Constant	NA
3	Dynamic reservoir case 1	3	0	Non-Constant	NA
4	Dynamic reservoir case 2	3	0	Non-Constant	NA
5	Static reservoir with low impact distant injector case 1	3	1	Constant	Low
6	Static reservoir with low impact distant injector case 2	3	1	Constant	Low
7	Dynamic reservoir with low impact distant injector	3	1	Non-Constant	Low
8	Static reservoir with high impact distant injector	3	1	Constant	High
9	Dynamic reservoir with high impact distant injector	3	1	Non-Constant	High

Our tests using synthetic data are summarized in Table 3.1, according to the number of close injectors, number of distant injectors, constant or non-constant IPRs and low or high impact from distant injectors, e.g., for Case 5, the region has three close injectors and one low impact distant injector with all constant IPRs.

3.3.1 Procedure for choosing the values of $f(r_{j,l},k_{j,l}), a_{j,l}$ and $b_{j,l}$

Recall from our third heuristic-knowledge for $f(r_{j,l},k_{j,l})$ in Section 2.1.B, the larger $f(r_{j,l},k_{j,l})$ is, the better is the connectivity from the injector to producer; the shorter is the delay time (i.e., the larger is $a_{j,l}$), and the larger is the impact from the injectors (i.e., the larger is $IPR_{j,l}$). Our procedure for choosing the values of $f(r_{j,l},k_{j,l}), a_{j,l}$, $b_{j,l}$ is:

1) Specify the value of $f(r_{j,l},k_{j,l})$ in $[0, 1]$;
2) Choose the value of $a_{j,l} \in [1, 1000]$ and $IPR_{j,l} \in [0, 1]$ using following rules:

   Rule 1: If $f(r_{j,l},k_{j,l})$ is *large*, then $a_{j,l}$ and $IPR_{j,l}$ are *large*.

   Rule 2: If $f(r_{j,l},k_{j,l})$ is *moderate*, then $a_{j,l}$ and $IPR_{j,l}$ are *moderate*.

   Rule 3: If $f(r_{j,l},k_{j,l})$ is *small*, then $a_{j,l}$ and $IPR_{j,l}$ are *small*.

3) Determine $b_{j,l}$ using $\alpha_{j,l} = e^{-a_{j,l}T}$, $\gamma_{j,l} = b_{j,l}\alpha_{j,l}T$ and (2.8) as

$$b_{j,l} = \frac{\gamma_{j,l}}{\alpha_{j,l}T} = \frac{H_{j,l}(z)|_{z=1}}{\left(1-\alpha_{j,l}\right)^2 \alpha_{j,l}T} = \frac{H_{j,l}(z)|_{z=1}}{\left(1-e^{-a_{j,l}T}\right)^2 e^{-a_{j,l}T}T}$$
$$= \frac{IPR_{j,l}}{f(r_{j,l},k_{j,l})\left(1-e^{-a_{j,l}T}\right)^2 e^{-a_{j,l}T}T} \tag{3.33}$$

where T is a small sample period, which corresponds to high sampling rate so that the generated discrete signal can be considered as a continuous signal.

The regional impact, IP_j^R, can then be computed using (3.11), i.e.,

$$IP_j^R = \sum_{l=1}^{M} IPR_{j,l} \tag{3.34}$$

3.3.2 Synthetic data generation

Synthetic data was generated as follows: (Note: the procedure is explained for the case of three injectors and two producers, and can be extended to any other situation.)

Step 1: Generate the noise-free injection rates $i_j^h(n)$ ($j=1,2,3$) using the injection rate model described in Appendix A.

Step 2: To generate the noise-free production rates $p_l^h(n)$ ($l=1,2$), the *SVM* for multiple injectors and a single producer was implemented with zero-noise by specifying $f(r_{i,j},k_{i,j})$, $a_{i,j}$, $b_{i,j}$ ($i=1,2,3; j=1,2.$) using the procedure described in Section 3.1 and a small $T=0.01$, which corresponds to a high sampling rate (recall that $\alpha_{i,j}=e^{-a_{i,j}T}$ and $\gamma_{i,j}=b_{i,j}\alpha_{i,j}T$).

Step 3: The noise-free injection rates $i_j(k)$ ($j=1,2,3$) and noise-free production rates $p_l(k)$ ($l=1,2$) were obtained by down-sampling $i_j^h(n)$ ($i=1,2,3$) and $p_l^h(n)$ by 20. This corresponds to a low sampling rate.

Step 4: The measured injection and production rates were obtained by adding measurement noises $n_j(k)$ with $SNR=10\log_{10}\left[\frac{E\left(i_j(k)\right)^2}{E\left(n_j(k)\right)^2}\right]=20$ dB to $i_j(k)$ ($j=1,2,3$), and $n_{p_l}(k)$ with $SNR=10\log_{10}\left[\frac{E\left[\left(p_l(k)\right)^2\right]}{E\left[\left(n_{p_l}(k)\right)^2\right]}\right]=20$ dB to $p_l(k)$ ($l=1,2$), i.e.,

$$\begin{cases} i_{m,j}(k)=i_j(k)+n_j(k) \\ p_{m,l}(k)=p_l(k)+n_{p_l}(k) \end{cases} \tag{3.35}$$

3.3.3 Monte-Carlo simulations

Simulations were implemented as follows:

Step 1: Set $i=1$, number of realizations $NR=100$, and the number of samples in each realization $M=500$.

Step 2: For the i^{th} realization ($j=1,2,3$, $l=1,2$, $k=1,\ldots,500$):

1) Generate $i_{m,j}(k)$ and $p_{m,l}(k)$, using the four-step procedure described in

Section 3.3.2, and compute the measured regional production rate $p_m^R(k)$ using (3.1);

2) Implement the *EKF* using the just generated $i_{m,j}(k)$ and $p_m^R(k)$, and obtain $\hat{\gamma}_j^{pv'}(k|k)$ and $\hat{\alpha}_j^{pv}(k|k)$;

3) Compute the estimated regional impact $\left[\hat{IP}_j^R(k|k)\right]_i$ using

$$\left[\hat{IP}_j^R(k|k)\right]_i = \frac{\hat{\gamma}_j^{pv'}(k|k)}{\left(1-\hat{\alpha}_j^{pv}(k|k)\right)^2}. \tag{3.36}$$

Step 3: If $i \geq NR$, go to Step 4; otherwise, set $i = i+1$ and go to Step 2.

Step 4: Compute the mean and standard deviation of $\left[\hat{IP}_j^R(k|k)\right]_i$ using

$$m_j(k) = \frac{1}{NR}\sum_{i=1}^{NR}\left[\hat{IP}_j^R(k|k)\right]_i \tag{3.37}$$

and

$$\sigma_j(k) = \sqrt{\frac{1}{NR}\sum_{i=1}^{NR}\left(\left[\hat{IP}_j^R(k|k)\right]_i - m_j(k)\right)^2}, \tag{3.38}$$

respectively.

Step 5: Stop.

Case 1: Static reservoir case 1

The parameters for Case 1 are summarized in Table 3.2. For this case,

1) Injector 1 impacts producer 2 the most while injector 2 impacts producer 1 the most because $f(r_{1,2},k_{1,2})=1$ and $f(r_{2,1},k_{2,1})=1$.

2) Injector 3 impacts producers 1 and 2 the least because $f(r_{3,1},k_{3,1})=0.5$ and $f(r_{3,2},k_{3,2})=0.5$.

3) Injector 2 has the largest regional impact, after which comes injector 1. Injector 3 has the smallest regional impact.

Table 3.2: The parameters for Case 1 (i refers to injector and j refers to producers).

i	j	$a_{i,j}$	$b_{i,j}$	$\alpha_{i,j}$	$\gamma_{i,j}$	$IPR_{i,j}$	$f(r_{i,j},k_{i,j})$	IP_j^R
1	1	200	120	0.1353	0.1624	0.22	0.7	0.384
	2	400	1200	0.0183	0.2198	0.23	1	
2	1	400	2400	0.0183	0.4396	0.456	1	0.576
	2	200	80	0.1353	0.1083	0.14	0.8	
3	1	10	0.1	0.9048	0.0009	0.1	0.5	0.075
	2	10	0.05	0.9048	0.0005	0.05	0.5	

The estimation results for regional impacts $\hat{IP}_j^R(k\,|\,k)$ ($j=1,2,3$) are depicted in Figs. 3.8 and 3.9. Observe that:

1) The means of $\hat{IP}_j^R(k\,|\,k)$ are very close to their actual regional impact as k increases.

2) The standard deviations of $\hat{IP}_j^R(k\,|\,k)$ are very small, which means that the estimation is very stable.

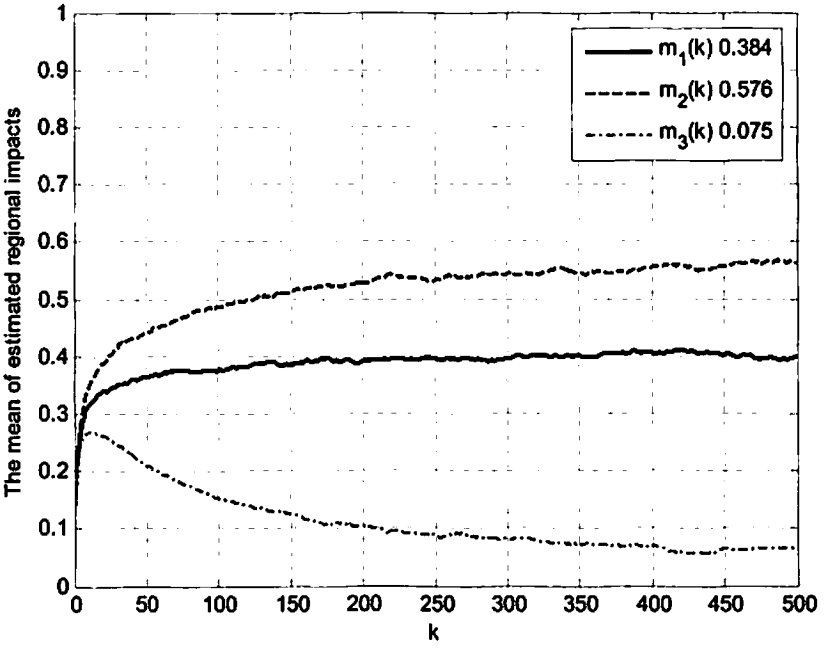


Figure 3.8: The mean of estimated $\hat{IP}_j^R(k \mid k)$ for Case 1.

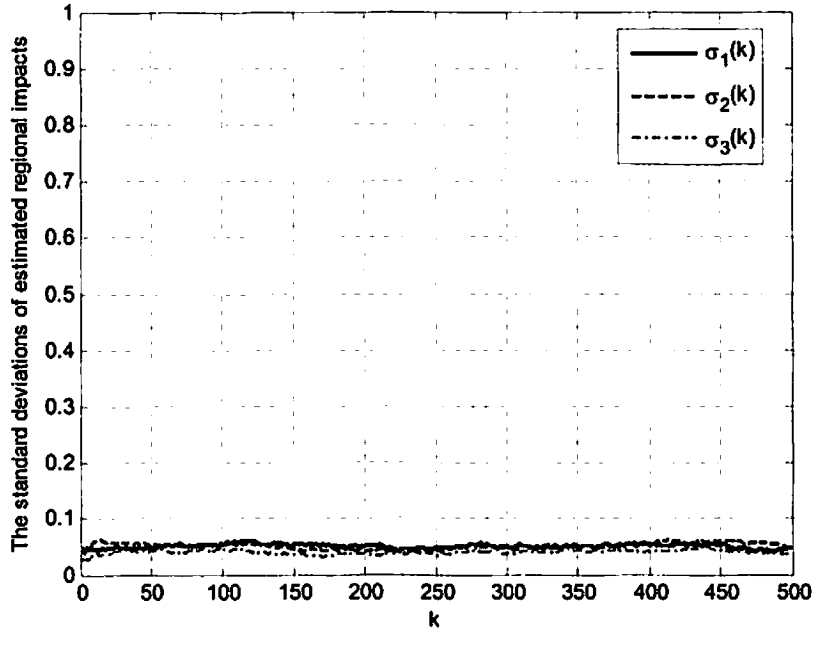


Figure 3.9: The standard deviation of estimated $\hat{IP}_j^R(k \mid k)$ for Case 1.

Case 2: Static reservoir case 2

The parameters for Case 2 are summarized in Table 3.3. For this case:

1) Injector 3, instead of injector 1, impacts producer 2 the most because $f(r_{3,2},k_{3,2})=1$ and $f(r_{1,2},k_{1,2})=1$.

2) Injector 1 has a regional impact that is very close to that of injector 3.

Table 3.3: The parameters for case 2 (i refers to injector and j refers to producers).

i	j	$a_{i,j}$	$b_{i,j}$	$\alpha_{i,j}$	$\gamma_{i,j}$	$IPR_{i,j}$	$f(r_{i,j},k_{i,j})$	IP_j^R
1	1	200	120	0.1353	0.1624	0.22	0.7	0.27
	2	200	80	0.1353	0.1083	0.14	0.8	
2	1	400	2400	0.0183	0.4396	0.456	1	0.485
	2	10	0.05	0.9048	0.0005	0.05	0.5	
3	1	10	0.1	0.9048	0.0009	0.1	0.5	0.28
	2	400	1200	0.0183	0.2198	0.23	1	

Estimation results for regional impacts $\hat{IP}_j^R(k\,|\,k)$ ($j=1,2,3$) are depicted in Figs. 3.10 and 3.11, which are similar to those for Case 1, i.e.,

1) The means of $\hat{IP}_j^R(k\,|\,k)$ are very close to their actual regional impact as k increases.

2) The standard deviations of $\hat{IP}_j^R(k\,|\,k)$ are very small, which means that the estimation is very stable.

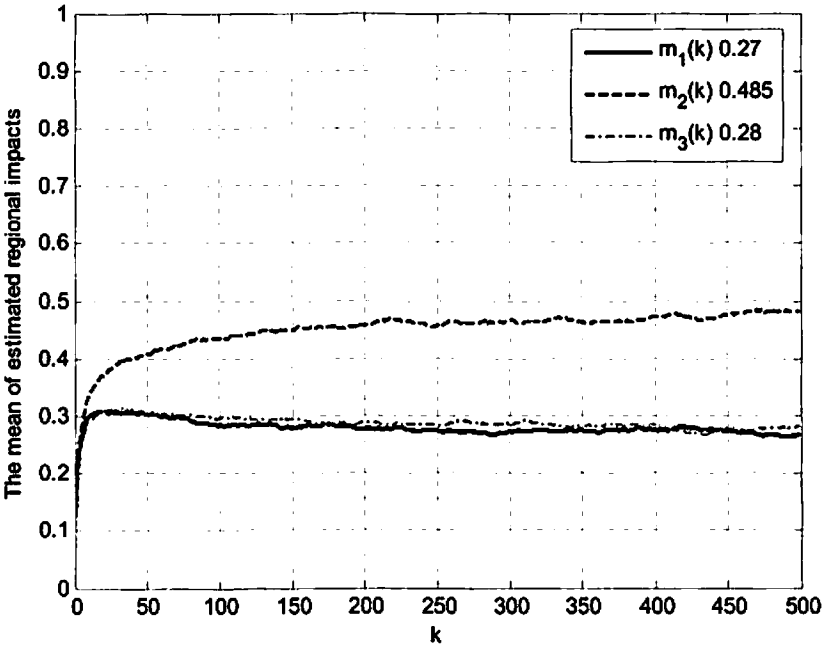


Figure 3.10: The mean of $\hat{IP}_j^R(k \mid k)$ for Case 2.

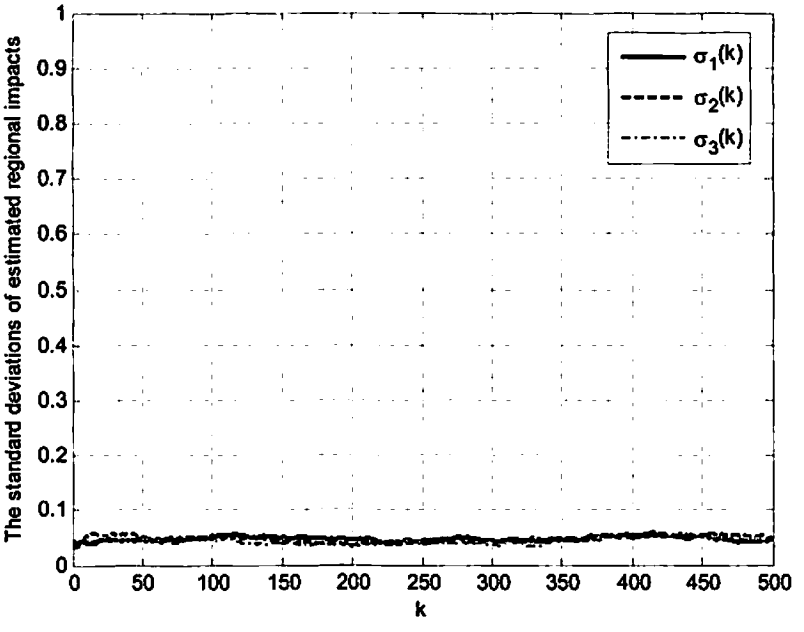


Figure 3.11: The standard deviation of $\hat{IP}_j^R(k \mid k)$ for Case 2.

Case 3: Dynamic reservoir case 1

Unlike Cases 1 and 2, for which the IPRs were constant for all measurements, in this case, we assume that something such as a work-over has happened so that the reservoir model is different before and after $k = 167$. The parameters for this case are summarized in Tables 3.4 and 3.5. Observe that:

1) Before $k = 167$, injectors 2 and 3 have the largest and smallest impacts on the region, respectively. After $k = 167$, injector 1 becomes most influential on producer 1 and least influential on producer 2, whereas injector 2 becomes least influential on producer 1.
2) Injector 2 has the largest regional impact before $k = 167$, whereas after $k = 167$, injector 1 has the largest regional impact.

Estimation results for regional impacts $\hat{IP}_j^R(k \mid k)$ ($j = 1,2,3$) are depicted in Figs. 3.12 and 3.13. Observe that:

1) $\hat{IP}_j^R(k \mid k)$ changes at time $k = 167$, which suggests that our method can successfully track a system change.
2) The means of $\hat{IP}_j^R(k \mid k)$ are very close to actual regional impacts, both before and after $k = 167$.
3) The standard deviations of $\hat{IP}_j^R(k \mid k)$ are very small, which means that the estimation is very stable.

Table 3.4: The parameters for Case 3 when $k=1,\ldots,167$ (i refers to injector and j refers to producers).

i	j	$a_{i,j}$	$b_{i,j}$	$\alpha_{i,j}$	$\gamma_{i,j}$	$IPR_{i,j}$	$f(r_{i,j},k_{i,j})$	IP_j^R
1	1	200	120	0.1353	0.1624	0.22	0.7	0.27
	2	200	80	0.1353	0.1083	0.14	0.8	
2	1	400	2400	0.0183	0.4396	0.456	1	0.686
	2	400	1200	0.0183	0.2198	0.23	1	
3	1	10	0.1	0.9048	0.0009	0.1	0.5	0.075
	2	10	0.05	0.9048	0.0005	0.05	0.5	

Table 3.5: The parameters for Case 3 when $k=168,\ldots,500$ (i refers to injector and j refers to producers).

i	j	$a_{i,j}$	$b_{i,j}$	$\alpha_{i,j}$	$\gamma_{i,j}$	$IPR_{i,j}$	$f(r_{i,j},k_{i,j})$	IP_j^R
1	1	400	2400	0.0183	0.4396	0.456	1	0.481
	2	10	0.05	0.9048	0.0005	0.05	0.5	
2	1	10	0.1	0.9048	0.0009	0.1	0.5	0.28
	2	400	1200	0.0183	0.2198	0.23	1	
3	1	200	120	0.1353	0.1624	0.22	0.7	0.27
	2	200	80	0.1353	0.1083	0.14	0.8	

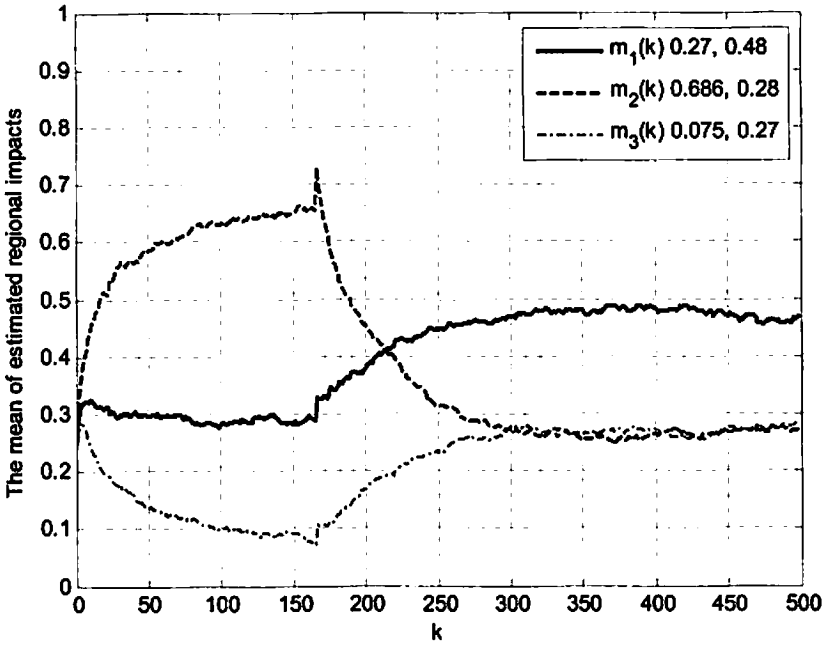


Figure 3.12: The mean of $\hat{IP}_j^R(k \mid k)$ for Case 3.

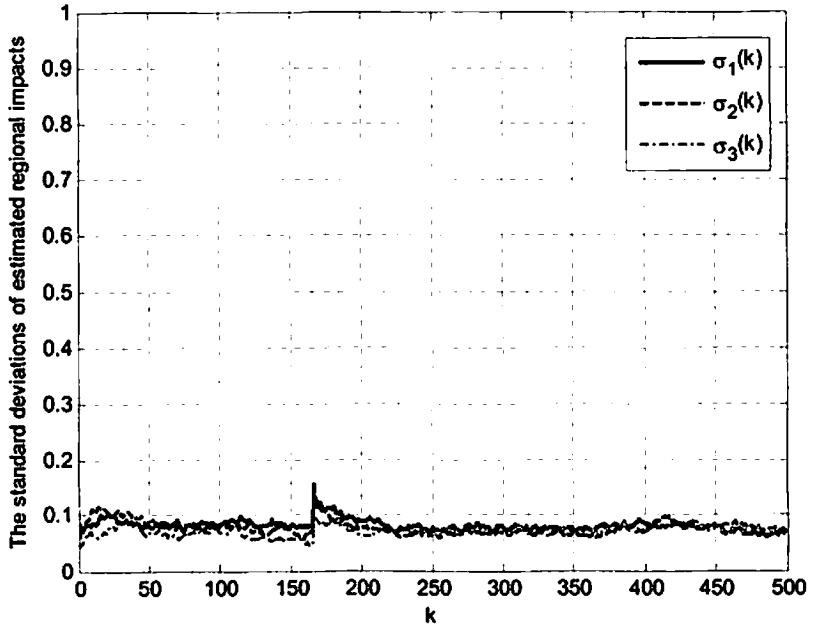


Figure 3.13: The standard deviation of $\hat{IP}_j^R(k \mid k)$ for Case 3.

Case 4: Dynamic reservoir case 2

As in Case 3, we assume that something such as a work-over has happened so that the reservoir model is different before and after $k = 167$. The parameters for this case are summarized in Table 3.6 and 3.7. Observe that:

1) Before $k = 167$, injector 1 impacts producer 1 the most, injector 2 impacts producer 2 the most, and injector 3 impacts producers 1 and 2 the least, respectively. After $k = 167$, injectors 1 and 2 impact producers 2 and 1 the least, respectively.
2) Before $k = 167$, injector 2 has moderate regional impact, which is much greater than that from injector 3. After $k = 167$, it is close to that of injector 3.

Table 3.6: The parameters for Case 4 when $k = 1,\ldots,167$ (i refers to injector and j refers to producers).

i	j	$a_{i,j}$	$b_{i,j}$	$\alpha_{i,j}$	$\gamma_{i,j}$	$IPR_{i,j}$	$f(r_{i,j},k_{i,j})$	IP_j^R
1	1	400	2400	0.0183	0.4396	0.456	1	0.576
	2	200	80	0.1353	0.1083	0.14	0.8	
2	1	200	120	0.1353	0.1624	0.22	0.7	0.384
	2	400	1200	0.0183	0.2198	0.23	1	
3	1	10	0.1	0.9048	0.0009	0.1	0.5	0.075
	2	10	0.05	0.9048	0.0005	0.05	0.5	

Table 3.7: The parameters for Case 4 when $k = 168,\ldots,500$ (i refers to injector and j refers to producers).

i	j	$a_{i,j}$	$b_{i,j}$	$\alpha_{i,j}$	$\gamma_{i,j}$	$IPR_{i,j}$	$f(r_{i,j},k_{i,j})$	IP_j^R
1	1	400	2400	0.0183	0.4396	0.456	1	0.481
	2	10	0.05	0.9048	0.0005	0.05	0.5	
2	1	10	0.1	0.9048	0.0009	0.1	0.5	0.28
	2	400	1200	0.0183	0.2198	0.23	1	
3	1	200	120	0.1353	0.1624	0.22	0.7	0.27
	2	200	80	0.1353	0.1083	0.14	0.8	

Estimation results for regional impacts $\hat{IP}_j^R(k|k)$ $(j=1,2,3)$ are depicted in Figs. 3.14 and 3.15, and are very similar to those for Case 3, i.e.,

1) $\hat{IP}_j^R(k|k)$ changes at time $k=167$, which suggests that our method can successfully track a system change.
2) The means of $\hat{IP}_j^R(k|k)$ are very close to actual regional impact, both before and after $k=167$.
3) The standard deviations of $\hat{IP}_j^R(k|k)$ are very small, which means that the estimation is very stable.

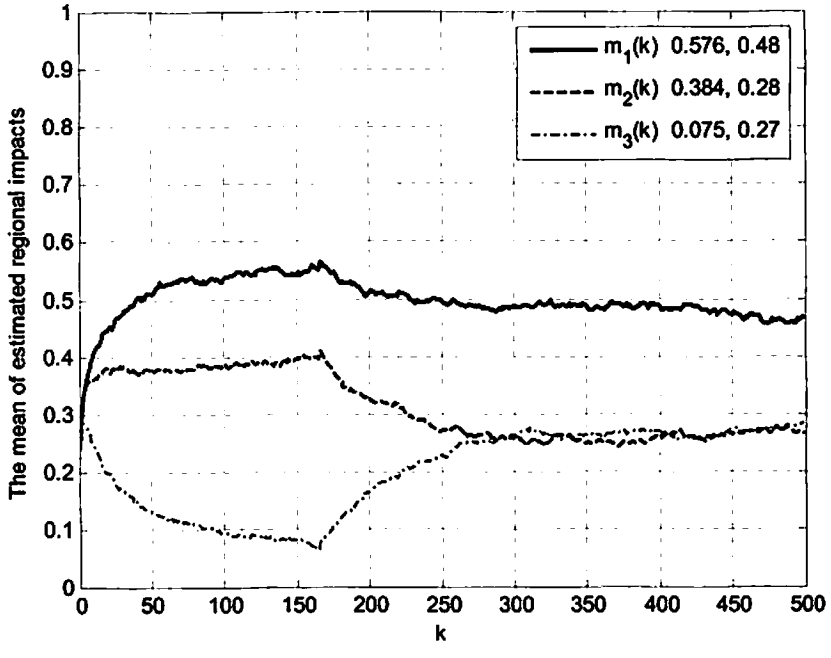


Figure 3.14: The mean of $\hat{IP}_j^R(k|k)$ for Case 4.

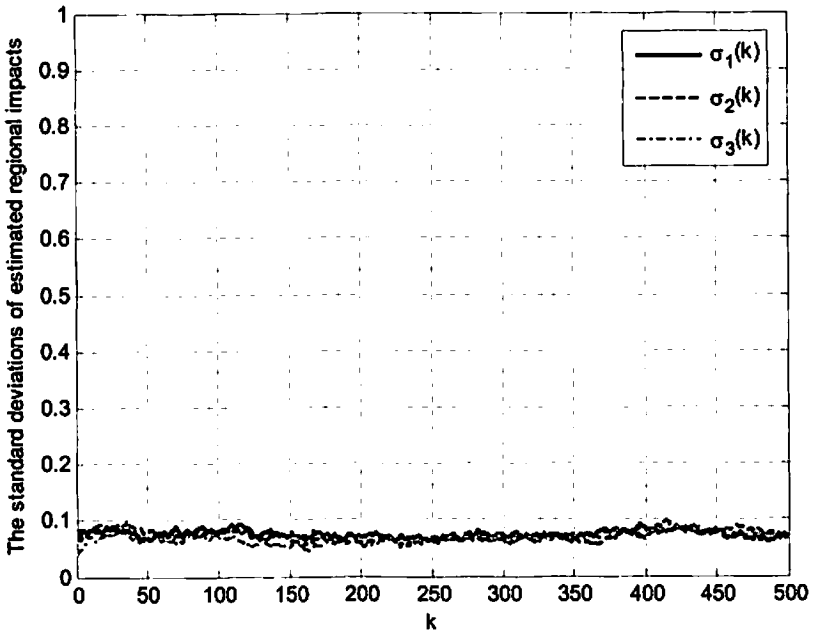


Figure 3.15: The standard deviation of $\hat{IP}_j^R(k \mid k)$ for Case 4.

Case 5: Static reservoir with low impact distant injector case 1

Unlike Cases 1 to 4, beginning with this case, we introduce a distant injector into the region. The purpose of this experiment is to see if omitting the distant injector from the EKF makes a difference on the estimated regional impact for close injectors. The corresponding parameters for this case are summarized in Table 3.8. Observe that:

1) The region includes three close injectors 1 to 3, one distant injector 4, and two producers, and the distant injector has very little impact on the producers.
2) Among the three close injectors, injector 2 has the largest regional impact whereas injector 3 has the smallest regional impact.

Estimation results for regional impacts $\hat{IP}_j^R(k\,|\,k)$ $(j=1,2,3)$ using injection rates from only the three close injectors are depicted in Figs. 3.16 and 3.17, and estimation results using injection rates from all four injectors are depicted in Figs. 3.18 and 3.19, respectively. Observe that:

1) The means and standard deviations of estimated regional impacts $\hat{IP}_j^R(k\,|\,k)$ for close injectors, without using the distant injector, are very close to those using all four injectors.
2) Regardless of using or not using the distant injector, the means of $\hat{IP}_j^R(k\,|\,k)$ for the three close injectors are very close to the actual regional impacts as k increases.
3) Regardless of using or not using the distant injector, the standard deviations of $\hat{IP}_j^R(k\,|\,k)$ for the three close injectors are very small, which means that the estimation is very stable.
4) $\hat{IP}_j^R(k\,|\,k)$ for the distant injector is very small, and this agrees with the reservoir model.

Table 3.8: The parameters for Case 5 (i refers to injector and j refers to producers).

i	j	$a_{i,j}$	$b_{i,j}$	$\alpha_{i,j}$	$\gamma_{i,j}$	$IPR_{i,j}$	$f(r_{i,j},k_{i,j})$	IP_j^R
1	1	200	120	0.1353	0.1624	0.22	0.7	0.27
	2	200	80	0.1353	0.1083	0.14	0.8	
2	1	400	2400	0.0183	0.4396	0.456	1	0.686
	2	400	1200	0.0183	0.2198	0.23	1	
3	1	10	0.1	0.9048	0.0009	0.1	0.5	0.065
	2	10	0.05	0.9048	0.0005	0.05	0.5	
4	1	1	0.0005	0.99	5×10^{-6}	0.05	0.3	0.03
	2	1	0.0005	0.99	5×10^{-6}	0.05	0.3	

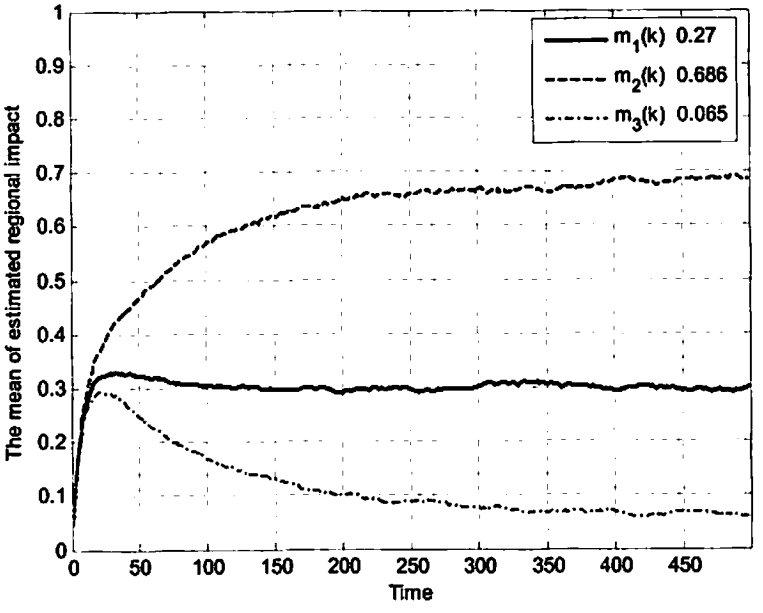


Figure 3.16: The mean of $I\hat{P}_j^R(k \mid k)$ using injection rates from only close injectors for Case 5.

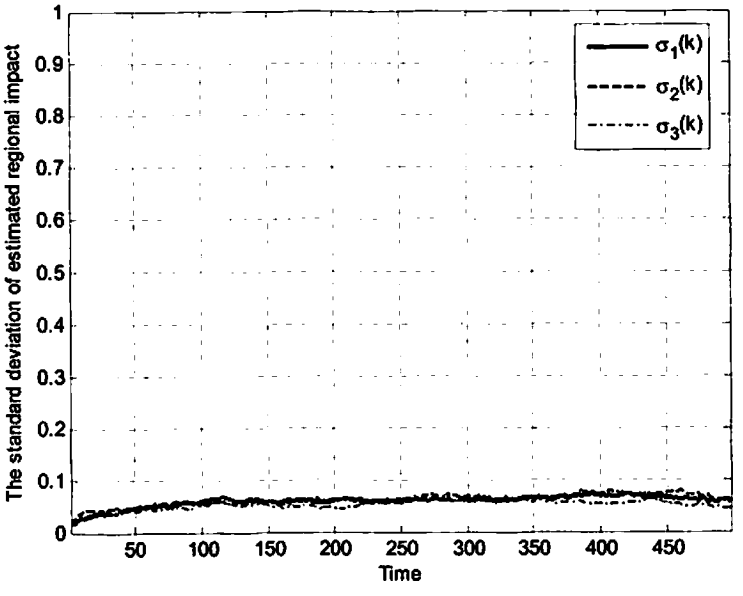


Figure 3.17: The standard deviation of $I\hat{P}_j^R(k \mid k)$ using injection rates from only close injectors for Case 5.

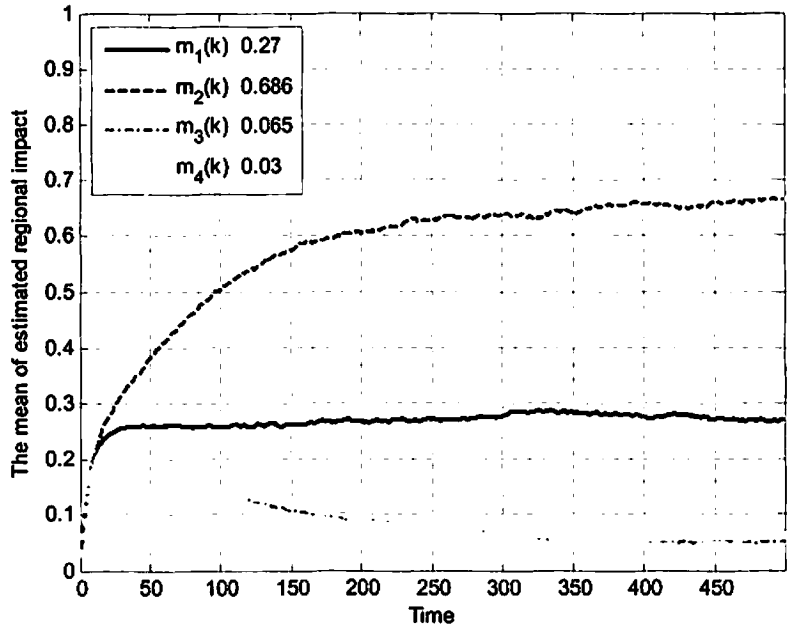


Figure 3.18: The mean of $\hat{IP}_j^R(k \mid k)$ using injection rates from all injectors for Case 5.

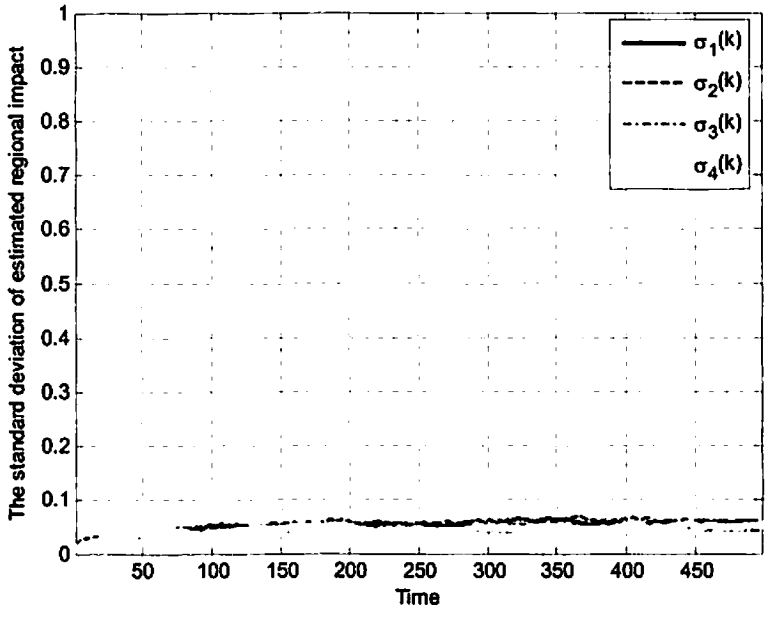


Figure 3.19: The standard deviation of $\hat{IP}_j^R(k \mid k)$ using injection rates from all injectors for Case 5.

Case 6: Static reservoir with low impact distant injector case 2

This Case is similar to Case 5 except that the impacts of close injectors for Case 6 are different from those for Case 5. The parameters for this case are summarized in Table 3.9.

Estimation results using injection rates from only the three close injectors are depicted in Figs. 3.20 and 3.21 for estimated regional impacts $\hat{IP}_j^R(k \mid k)$ ($j = 1,2,3$), and estimation results using injection rates from all four injectors are depicted in Figs. 3.22 and 3.23 for estimated regional impacts $\hat{IP}_j^R(k \mid k)$ ($j = 1,2,3,4$), respectively. All results are similar to those for Case 5, which further confirms the results obtained in Case 5.

Table 3.9: The parameters for Case 6 (i refers to injector and j refers to producers).

i	j	$a_{i,j}$	$b_{i,j}$	$\alpha_{i,j}$	$\gamma_{i,j}$	$IPR_{i,j}$	$f(r_{i,j},k_{i,j})$	IP_j^R
1	1	200	120	0.1353	0.1624	0.22	0.7	0.27
	2	200	80	0.1353	0.1083	0.14	0.8	
2	1	400	2400	0.0183	0.4396	0.456	1	0.485
	2	10	0.05	0.9048	0.0005	0.05	0.5	
3	1	10	0.1	0.9048	0.0009	0.1	0.5	0.28
	2	400	1200	0.0183	0.2198	0.23	1	
4	1	1	0.0005	0.99	5×10^{-6}	0.05	0.3	0.03
	2	1	0.0005	0.99	5×10^{-6}	0.05	0.3	

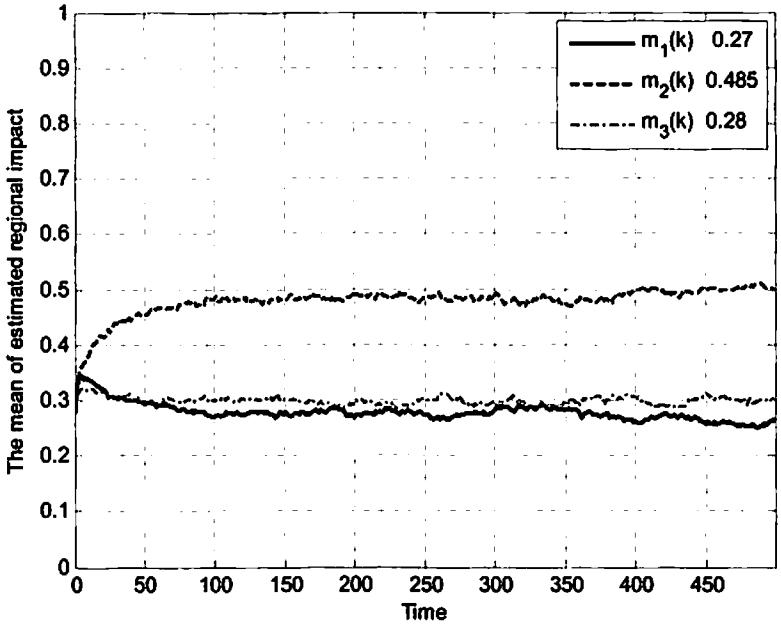


Figure 3.20: The mean of the estimated regional impact $I\hat{P}_j^R(k \mid k)$ using injection rates from only close injectors for Case 6.

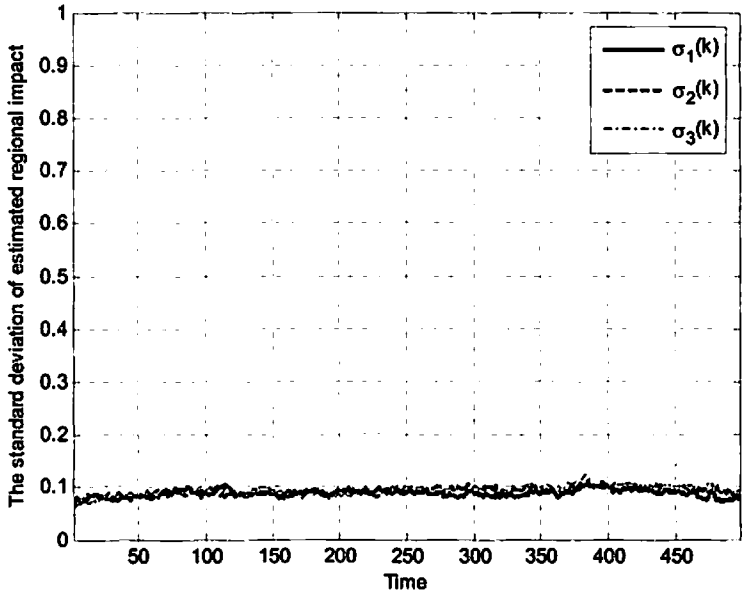


Figure 3.21: The standard deviation of the estimated regional impact $I\hat{P}_j^R(k \mid k)$ using injection rates from only close injectors for Case 6.

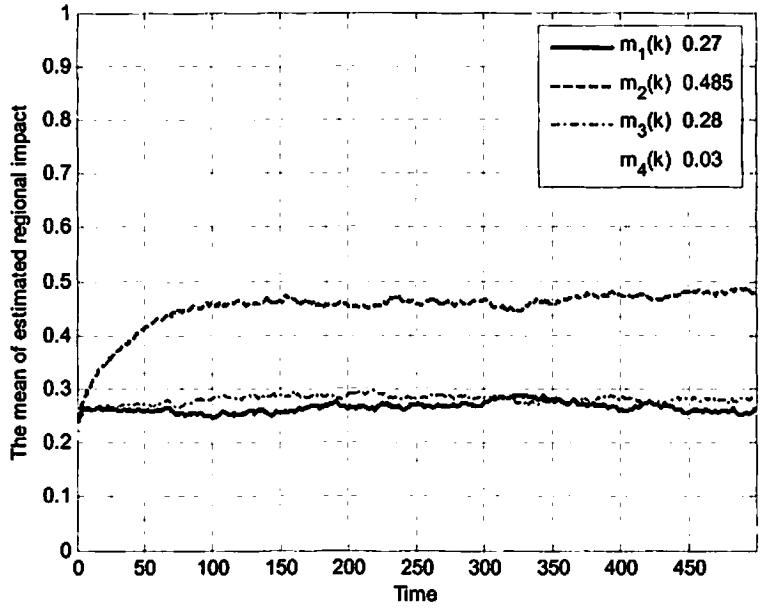


Figure 3.22: The mean of the estimated regional impact $I\hat{P}_j^R(k \mid k)$ using injection rates from all injectors for Case 6.

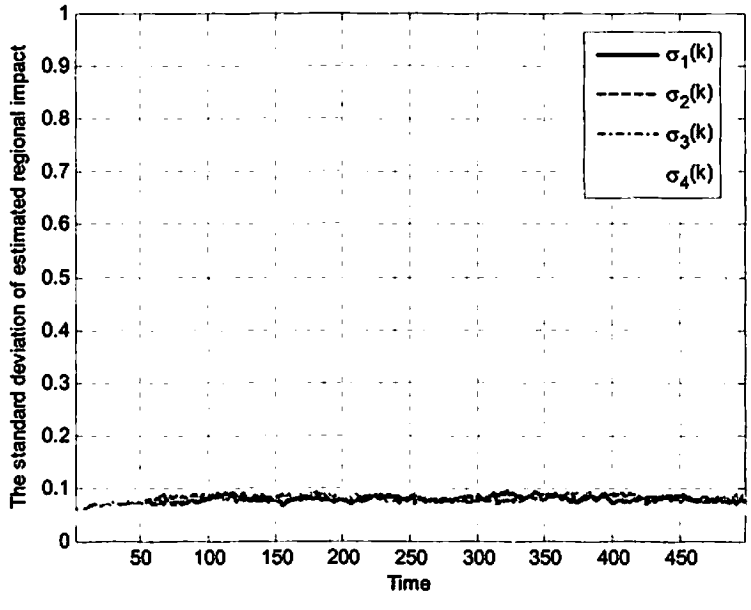


Figure 3.23: The standard deviation of the estimated regional impact $I\hat{P}_j^R(k \mid k)$ using injection rates from all injectors for Case 6.

Case 7: Dynamic reservoir with low impact distant injector

Unlike Case 6, in this case, we assume that something such as a work-over has happened so that the reservoir model is different before and after $k=167$. The parameters for this case are summarized in Tables 3.10 and 3.11. Observe that:

1) Before $k=167$, injector 2 impacts producers 1 and 2 the most, and injector 3 impacts producers 1 and 2 the least. After $k=167$, injector 1 impacts producer 1 the most and producer 2 the least, and injector 2 impacts producer 2 the most and producer 1 the least.
2) Before $k=167$, injector 2 has the largest regional impact, whereas after $k=167$, injector 1 has the largest regional impact.

Estimation results for regional impacts $I\hat{P}_j^R(k\,|\,k)$ $(j=1,2,3)$ using injection rates from only the three close injectors are depicted in Figs. 3.24 and 3.25, and estimation results using injection rates from all four injectors are depicted in Figs. 3.26 and 3.27. Observe that:

1) The estimated regional impacts $I\hat{P}_j^R(k\,|\,k)$ for the three close injectors, without using distant injector, are very close to those using all four injectors.
2) Regardless of using or not using the distant injector, the means of $I\hat{P}_j^R(k\,|\,k)$ for the three close injectors are very close to the actual regional impacts as k increases.
3) Regardless of using or not using distant injector, the $I\hat{P}_j^R(k\,|\,k)$ changes at time $k=167$, which verify that our method can adaptively track a system change.

4) Regardless of using or not using distant injector, the standard deviations of $I\hat{P}_j^R(k\,|\,k)$ for the three close injectors are very small, which means that the estimation is very stable.

5) $I\hat{P}_j^R(k\,|\,k)$ for the distant injector is very small, and this agrees with the reservoir model.

Table 3.10: The parameters for Case 7 when $k = 1,\ldots,167$ (i refers to injector and j refers to producers).

i	j	$a_{i,j}$	$b_{i,j}$	$\alpha_{i,j}$	$\gamma_{i,j}$	$IPR_{i,j}$	$f(r_{i,j},k_{i,j})$	IP_j^R
1	1	200	120	0.1353	0.1624	0.22	0.7	0.27
	2	200	80	0.1353	0.1083	0.14	0.8	
2	1	400	2400	0.0183	0.4396	0.456	1	0.686
	2	400	1200	0.0183	0.2198	0.23	1	
3	1	10	0.1	0.9048	0.0009	0.1	0.5	0.075
	2	10	0.05	0.9048	0.0005	0.05	0.5	
4	1	1	0.0005	0.99	5×10^{-6}	0.05	0.1	0.03
	2	1	0.0005	0.99	5×10^{-6}	0.05	0.5	

Table 3.11: The parameters for Case 7 when $k = 168,\ldots,500$ (i refers to injector and j refers to producers).

i	j	$a_{i,j}$	$b_{i,j}$	$\alpha_{i,j}$	$\gamma_{i,j}$	$IPR_{i,j}$	$f(r_{i,j},k_{i,j})$	IP_j^R
1	1	400	2400	0.0183	0.4396	0.456	1	0.481
	2	10	0.05	0.9048	0.0005	0.05	0.5	
2	1	10	0.1	0.9048	0.0009	0.1	0.5	0.28
	2	400	1200	0.0183	0.2198	0.23	1	
3	1	200	120	0.1353	0.1624	0.22	0.7	0.27
	2	200	80	0.1353	0.1083	0.14	0.8	
4	1	1	0.0005	0.99	5×10^{-6}	0.05	0.1	0.03
	2	1	0.0005	0.99	5×10^{-6}	0.05	0.5	

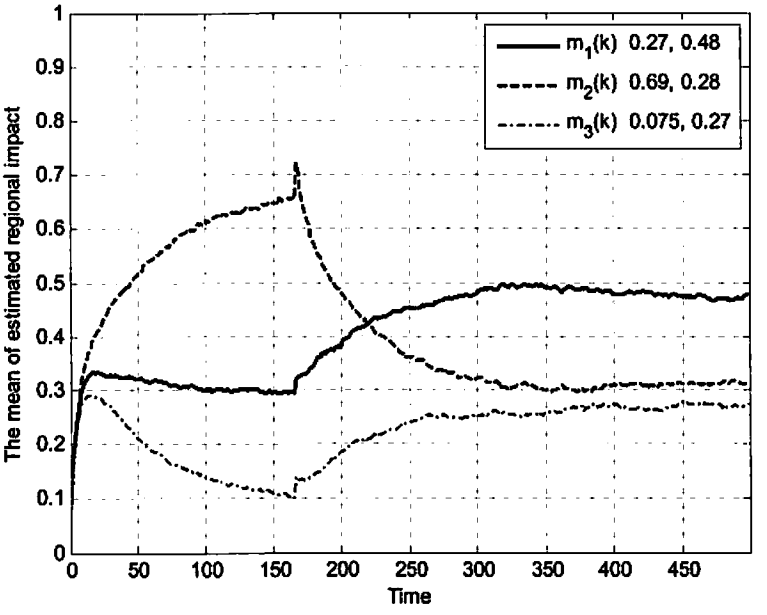


Figure 3.24: The mean of the estimated regional impact $I\hat{P}_j^R(k \mid k)$ using injection rates from only close injectors for Case 7.

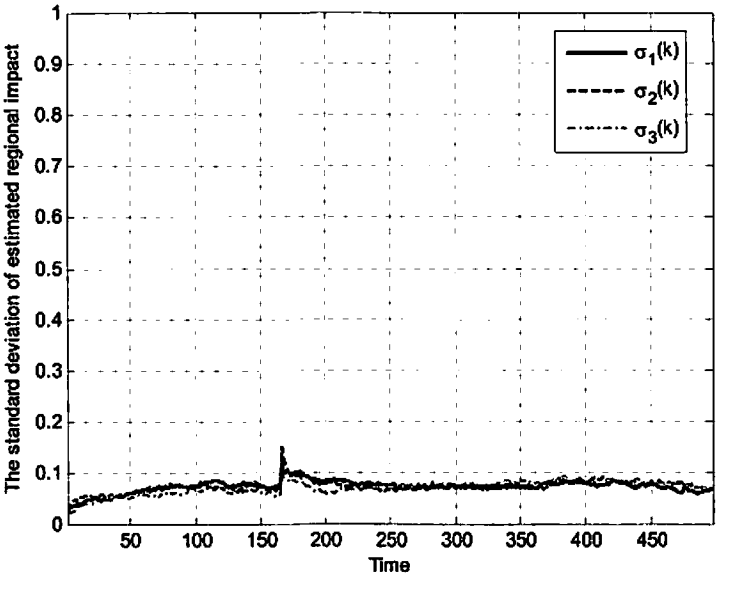


Figure 3.25: The standard deviation of the estimated regional impact $I\hat{P}_j^R(k \mid k)$ using injection rates from only close injectors for Case 7.

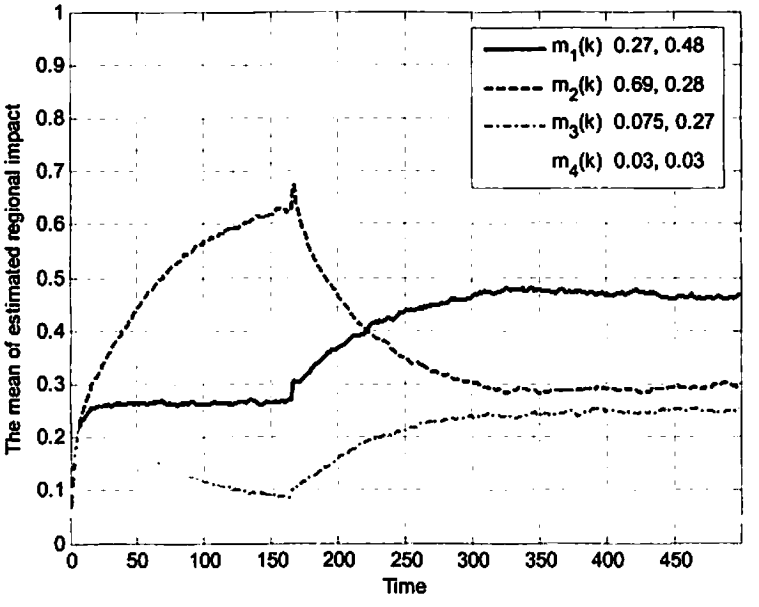


Figure 3.26: The mean of the estimated regional impact $\hat{IP}_j^R(k \mid k)$ using injection rates from all injectors for Case 7.

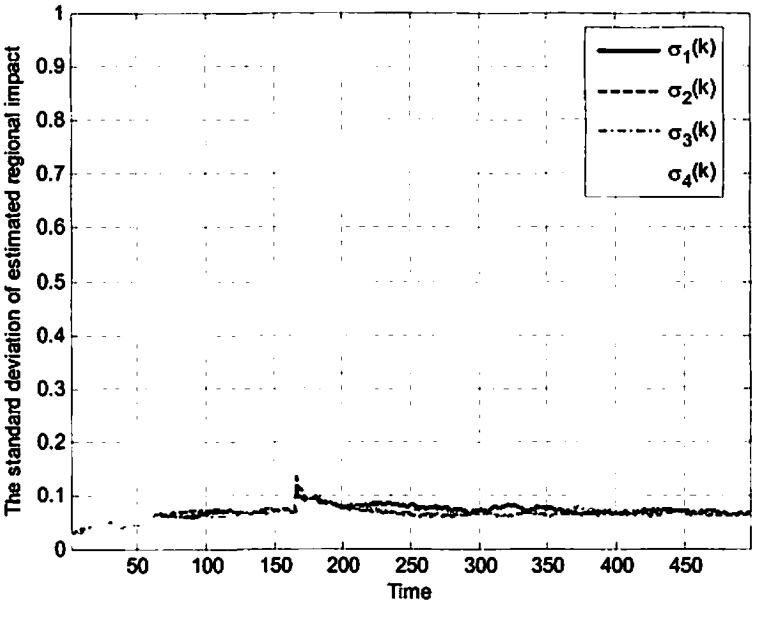


Figure 3.27: The standard deviation of the estimated regional impact $\hat{IP}_j^R(k \mid k)$ using injection rates from all injectors for Case 7.

Case 8: Static reservoir with high impact distant injector

Unlike Cases 5 and 6, for which the distant injector has the smallest regional impact, beginning with this case, the distant injector has the greatest regional impact. The parameters for this case are summarized in Table 3.12.

Table 3.12: The parameters for Case 8 (i refers to injector and j refers to producers).

i	j	$a_{i,j}$	$b_{i,j}$	$\alpha_{i,j}$	$\gamma_{i,j}$	$IPR_{i,j}$	$f(r_{i,j},k_{i,j})$	IP_j^R
1	1	200	120	0.1353	0.1624	0.22	0.7	0.27
	2	200	80	0.1353	0.1083	0.14	0.8	
2	1	10	0.1	0.9048	0.0009	0.1	0.5	0.39
	2	400	1800	0.0183	0.3297	0.34	1	
3	1	1	0.0005	0.99	5×10^{-6}	0.05	0.3	0.03
	2	10	0.05	0.9048	0.0005	0.05	0.3	
4	1	400	2400	0.0183	0.4396	0.456	1	0.471
	2	1	0.0005	0.99	5×10^{-6}	0.05	0.3	

Estimation results for regional impacts $I\hat{P}_j^R(k\,|\,k)$ ($j=1,2,3$) using injection rates from only the three close injectors are depicted in Figs. 3.28 and 3.29, and estimation results using injection rates from all four injectors are depicted in Figs. 3.30 and 3.31. Observe that:

1) The estimated regional impacts $I\hat{P}_j^R(k\,|\,k)$ for the three close injectors, without using the distant injector, are much different from those using the distant injector. Interestingly, even when the distant significant injector is omitted, the estimated $I\hat{P}_j^R(k\,|\,k)$ for close injectors still preserve their relative importance, i.e., the injector with the largest regional impact has the greatest value of $I\hat{P}_j^R(k\,|\,k)$, and the injector with the smallest regional impact $I\hat{P}_j^R(k\,|\,k)$ has the smallest value of $I\hat{P}_j^R(k\,|\,k)$.

2) Using the distant injector, the means of $I\hat{P}_j^R(k \mid k)$ are very close to actual regional impacts as k increases.

3) Using the distant injector, the standard deviations of $I\hat{P}_j^R(k \mid k)$ are very small, which means that the estimation is very stable.

4) Using the distant injector, $I\hat{P}_j^R(k \mid k)$ for the distant injector has the largest value, and this agrees with the reservoir model.

5) The standard deviations of $I\hat{P}_j^R(k \mid k)$ for the three close injectors using the distant injector are 50% smaller than those without using the distant injector.

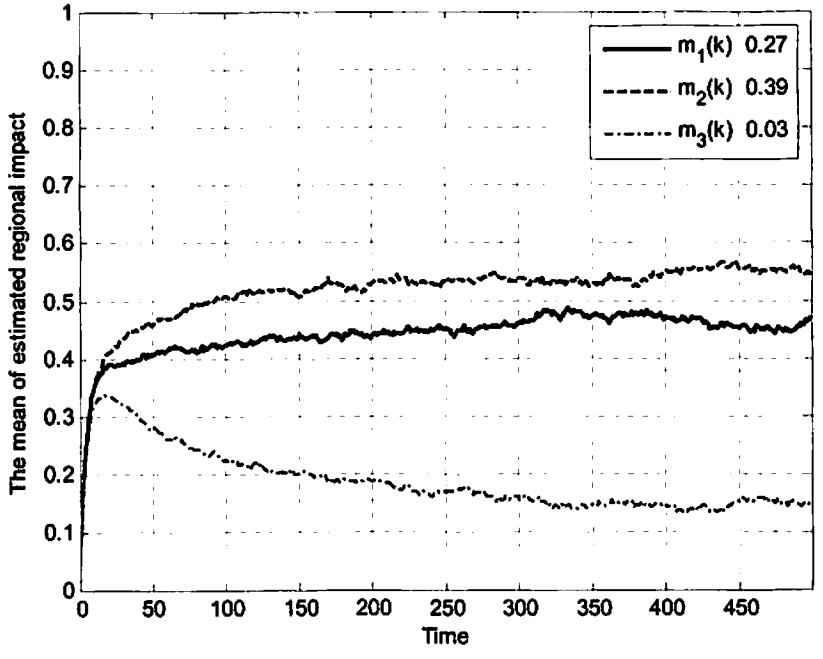


Figure 3.28: The mean of $I\hat{P}_j^R(k \mid k)$ using injection rates from only close injectors for Case 8.

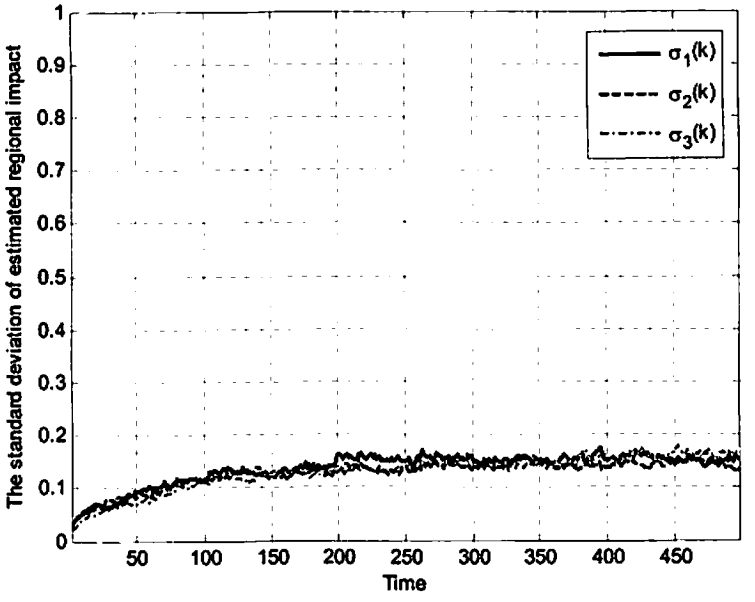


Figure 3.29: The standard deviation of $\hat{IP}_j^R(k \mid k)$ using injection rates from only close injectors for Case 8.

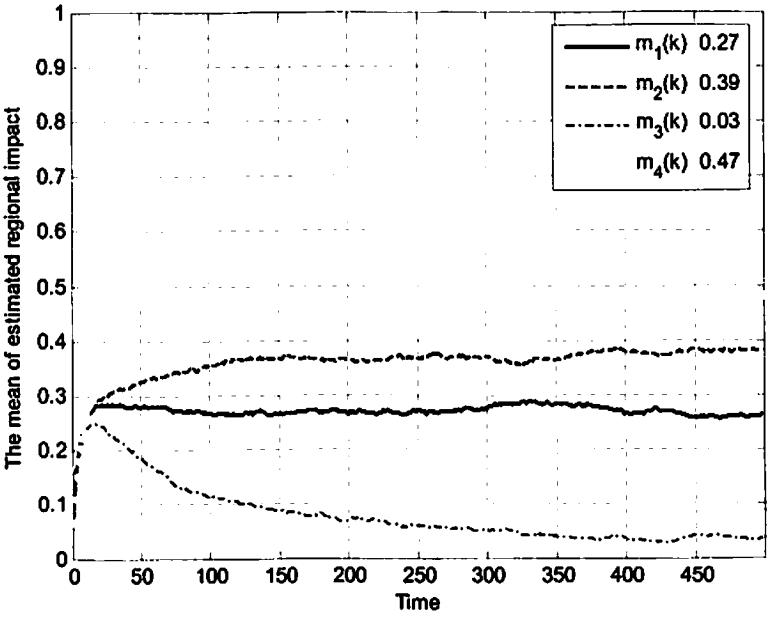


Figure 3.30: The mean of $\hat{IP}_j^R(k \mid k)$ using injection rates from all injectors for Case 8.

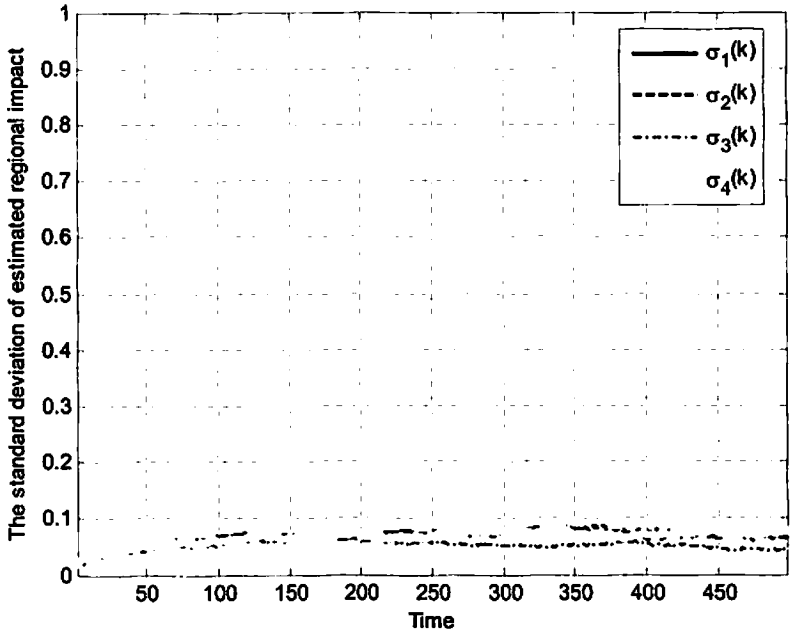


Figure 3.31: The standard deviation of $\hat{IP}_j^R(k \mid k)$ $\hat{ZI}_i$ using injection rates from all injectors for Case 8.

Case 9: Dynamic reservoir with high impact distant injector

Unlike Case 8, for which the IPRs are constant, in this case, we assume that something such as a work-over has happened so that the reservoir model is different before and after $k = 167$. The parameters for this case are summarized in Tables 3.13 and 3.14. Observe that, among the three close injectors, before $k = 167$, injectors 2 and 3 have the largest and smallest regional impacts, respectively, and after $k = 167$, the regional impact of injector 2 reduces to the medium.

Estimation results for regional impacts $\hat{IP}_j^R(k \mid k)$ ($j = 1,2,3$) using injection rates from only the three close injectors are depicted in Figs. 3.32 and 3.33, and estimation

results using injection rates from all four injectors are depicted in Figs. 3.34 and 3.35, respectively. All results are similar to those for Case 8 except that, regardless of using or not using distant injector, the $I\hat{P}_j^R(k\,|\,k)$ changes at time $k=167$, which further confirms that our method can adaptively track a system change.

Table 3.13: Parameters for Case 9 when $k=1,\ldots,167$ (i refers to injector and j refers to producers).

i	j	$a_{i,j}$	$b_{i,j}$	$\alpha_{i,j}$	$\gamma_{i,j}$	$IPR_{i,j}$	$f(r_{i,j},k_{i,j})$	IP_j^R
1	1	200	120	0.1353	0.1624	0.22	0.7	0.27
	2	200	80	0.1353	0.1083	0.14	0.8	
2	1	10	0.1	0.9048	0.0009	0.1	0.5	0.39
	2	400	1800	0.0183	0.3297	0.34	1	
3	1	1	0.0005	0.99	5×10^{-6}	0.05	0.3	0.03
	2	10	0.05	0.9048	0.0005	0.05	0.3	
4	1	400	2400	0.0183	0.4396	0.456	1	0.471
	2	1	0.0005	0.99	5×10^{-6}	0.05	0.3	

Table 3.14: The parameters for Case 9 when $k=168,\ldots,500$ (i refers to injector and j refers to producers).

i	j	$a_{i,j}$	$b_{i,j}$	$\alpha_{i,j}$	$\gamma_{i,j}$	$IPR_{i,j}$	$f(r_{i,j},k_{i,j})$	IP_j^R
1	1	10	0.1	0.9048	0.0009	0.1	0.5	0.39
	2	400	1800	0.0183	0.3297	0.34	1	
2	1	200	120	0.1353	0.1624	0.22	0.7	0.27
	2	200	80	0.1353	0.1083	0.14	0.8	
3	1	1	0.0005	0.99	5×10^{-6}	0.05	0.3	0.03
	2	10	0.05	0.9048	0.0005	0.05	0.3	
4	1	400	2400	0.0183	0.4396	0.456	1	0.471
	2	1	0.0005	0.99	5×10^{-6}	0.05	0.3	

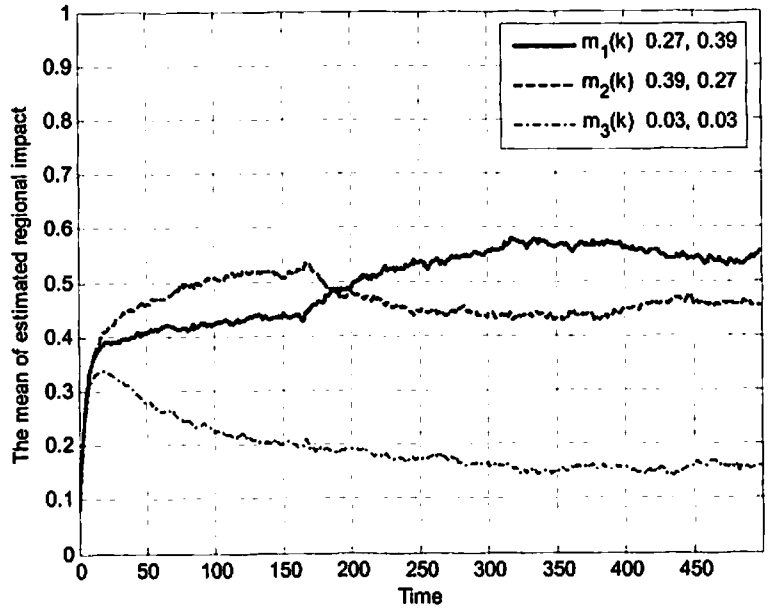


Figure 3.32: The mean of $I\hat{P}_j^R(k \mid k)$ using injection rates from only close injectors for Case 9.

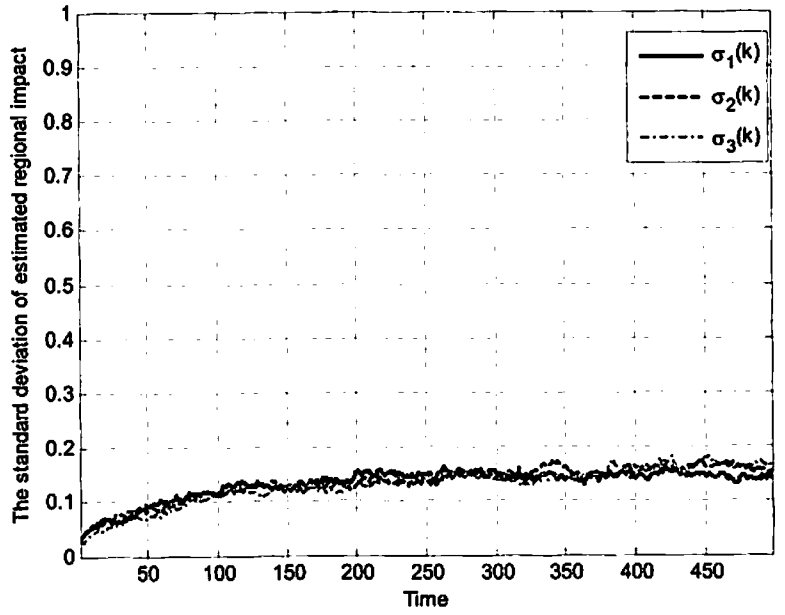


Figure 3.33: The standard deviation of $I\hat{P}_j^R(k \mid k)$ using injection rates from only close injectors for Case 9.

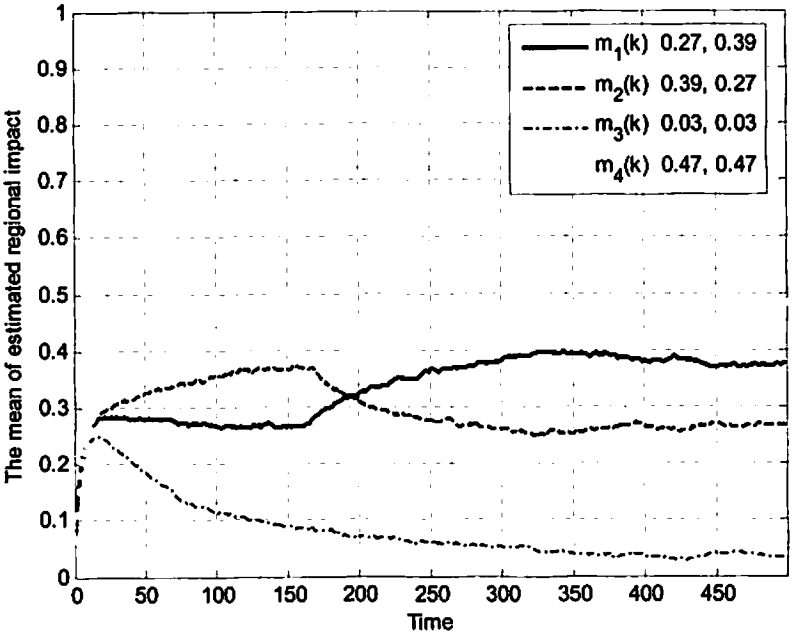


Figure 3.34: The mean of $\hat{IP}_j^R(k \mid k)$ using injection rates from all injectors for Case 9.

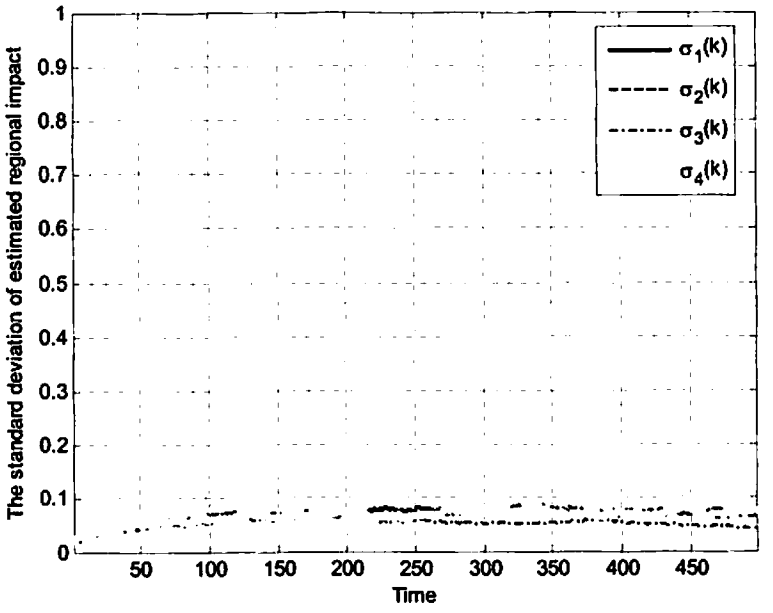


Figure 3.35: The standard deviation of $\hat{IP}_j^R(k \mid k)$ using injection rates from all injectors for Case 9.

3.4 Conclusions

This chapter presents a pseudo-virtual reservoir method to forecast the regional impacts between multiple injectors and multiple producers based on measured production and injection rates.

Our approach has been tested on synthetic data, and test results suggest that this pseudo-virtual reservoir method is feasible for evaluating regional impacts with multiple injectors and multiple producers.

Our approach could be used to determine regional impacts, and may lead to a better understanding of regional reservoir heterogeneity, all of which will assist Chevron engineers to make better decisions under different scales, e.g., a region with a single producer, a collection of producers, a pattern and a collection of patterns.

Chapter 4

Background Knowledge for Type-1 and Type-2 Fuzzy Sets

In this chapter, we briefly review some basic definitions and theorems for type-1 and type-2 fuzzy sets such that we can easily understand the materials in Chapter 5 [13]-[15], [21], [22], [24], [25], [28], [29].

4. 1 Type-1 Fuzzy Sets

Definition 4.1: A *type-1 fuzzy set*, denoted A, is characterized by a T1 membership function (MF) $\mu(x)$, where $x \in X$ and $\mu(x) \in [0,1]$, i.e.,

$$A=\{(x,\ \mu(x))\ |\ \forall x \in X, \mu(x) \in [0,1]\} \tag{4.1}$$

Definition 4.2: The $\alpha-cut$ of the T1 fuzzy set, A_α, is a crisp set that can be formulated as

$$A_\alpha = \{x \mid \mu(x) \geq \alpha\} \tag{4.2}$$

Obviously, A_α contains all elements whose grades are greater than or equal to a specific value of α. Consequently, A_α is a finite collection of elements for the discrete

case and an interval of values for the continuous case. An example is depicted in Fig. 4.1, where the type-1 membership function is triangular, and one of its $\alpha-cuts$ contains all the elements on x-axis between the two lower solid dots.

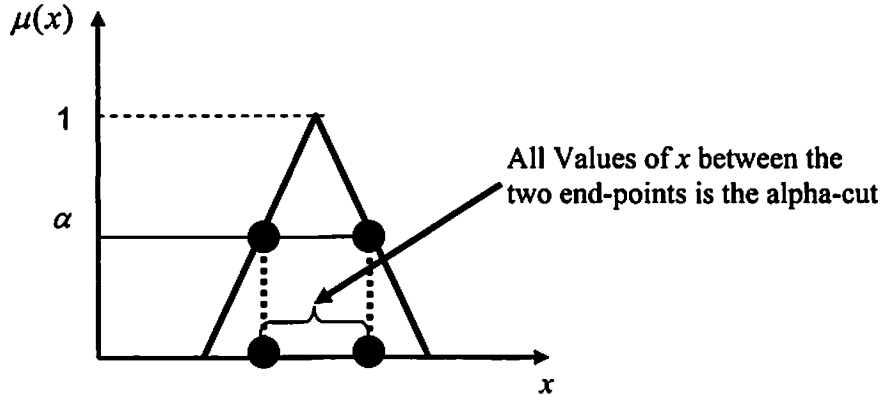


Figure 4.1: A type-1 triangular membership function and one of its alpha-cuts.

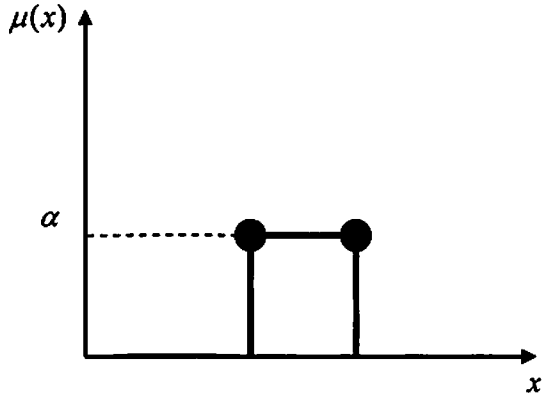


Figure 4.2: Membership function of $A(\alpha)$.

Definition 4.3: Let $I_{A_\alpha}(x)$ be an indicator function of the crisp set A_α, i.e.,

$$I_{A_\alpha}(x)=\begin{cases}1, & x\in A_\alpha\\ 0, & \text{otherwise}\end{cases} \tag{4.3}$$

The associated T1 fuzzy set for the $\alpha-cut\ A_\alpha$, $A(\alpha)$, is defined as

$$A(\alpha)=\left\{(x,\ \alpha I_{A_\alpha}(x))\ |\ \forall x\in X\right\} \tag{4.4}$$

An example is depicted in Fig. 4.2, where the $\alpha-cut\ A_\alpha$ has been raised to level α.

Although the T1 MF is the usual way to represent a T1 fuzzy set, $\alpha-cuts$ provide another way as in the following $\alpha-cut$ representation:

Theorem 4.1 ($\alpha-cut$ Representation Theorem for T1 FS): A T1 FS A is equal to the union of its associated T1 FSs $A(\alpha)$ for $\alpha\in[0,\ 1]$, i.e.,

$$A=\bigcup_{\alpha\in[0,1]}A(\alpha)=\sup_{\alpha\in[0,1]}A(\alpha) \tag{4.5}$$

Unlike the membership function representation, which is very useful for theoretical analyses, the $\alpha-cut$ representation usually not only provides us a good capability for computation, e.g., in computing a fuzzy weighted average, but it also is very useful for theoretical analyses.

4. 2 Type-2 Fuzzy Sets

In the shaded area of Fig. 4.3, at a specific value of x, say x', there no longer is a single value for the membership function. Instead, the membership function takes on

values wherever the dotted line intersects the shaded area. Different amplitude values can be assigned to these points $(x',\ u' = \mu(x'))$, which consequently leads to a three-dimensional membership function, called a type-2 membership function.

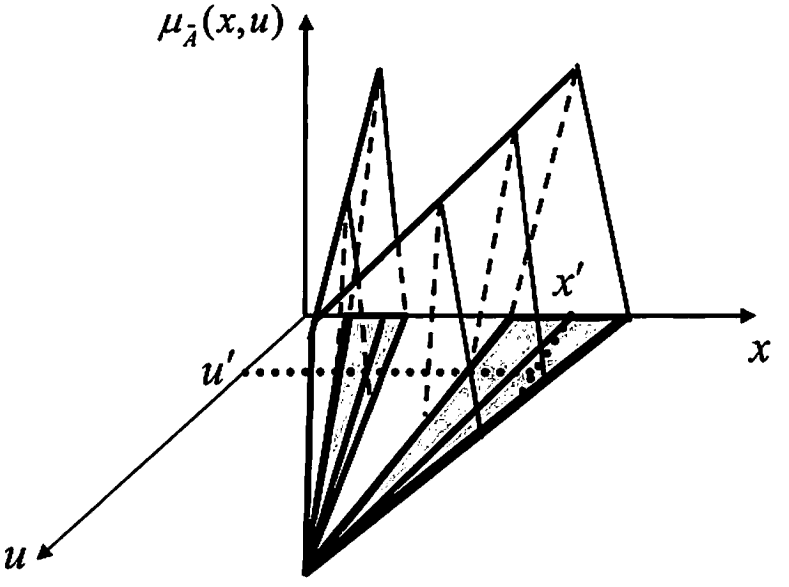


Figure 4.3: An example of a type-2 membership function.

Definition 4.4: At each value of *primary variable* x, there is an associated *secondary variable*, u, where $u \in [u_L(x), u_U(x)] \equiv J_x$, and $u_L(x)$ and $u_U(x)$ denote the lower and upper values for u at each x. J_x is a shorthand notation for $[u_L(x), u_U(x)]$, and is called the *primary membership* of x, and, $J_x \subseteq [0,1]$.

Definition 4.5: A *type-2 fuzzy set*, denoted $\tilde{A}$, is characterized by a type-2 membership function $\mu_{\tilde{A}}(x,u)$, where $x \in X, \forall u \in J_x \subseteq [0,1]$, i.e.,

$$\tilde{A} = \left\{(x,\ u), \mu_{\tilde{A}}(x,u) \mid \forall x \in X, \forall u \in J_x \subseteq [0,1]\right\} \tag{4.6}$$

in which $0 \le \mu_{\tilde{A}}(x,u) \le 1$, and J_x is the primary membership, which is the domain of the secondary membership function. $\tilde{A}$ can also be expressed as

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x} \mu_{\tilde{A}}(x,u)/(x,u) \quad J_x \subseteq [0,1] \tag{4.7}$$

where $\iint$ denotes union over all admissible x and u. For discrete universes of discourse $\int$ is replaced by $\sum$. In Definition 4.5, the first restriction that $u \in J_x = [u_L(x), u_U(x)] \subseteq [0,1]$ is consistent with the T1 constraint that $0 \le \mu_A(x) \le 1$, i.e., when uncertainties disappear a T2 MF must reduce to a T1 MF, in which case the variable u equals $\mu_A(x)$ and $0 \le \mu_A(x) \le 1$. The second restriction that $0 \le \mu_{\tilde{A}}(x,u) \le 1$ is consistent with the fact that the amplitudes of a MF should lie between or be equal to 0 and 1.

Definition 4.6: When $\mu_{\tilde{A}}(x,u) = 1$ for $\forall x \in X$ and $u \in J_x$, then $\tilde{A}$ is called an *interval T2 FS* (IT2 FS), in which case

$$\tilde{A} = \int_{x \in X} \int_{u \in J_x} 1/(x,u) \quad J_x = [u_L(x), u_U(x)] \subseteq [0,1] \tag{4.8}$$

Note that *the third dimension—the secondary grade—of the general T2 FS is no longer needed for an IT2 FS* because it conveys no new information, i.e. all secondary grades of an IT2 FS are the same, namely 1. We therefore have two choices: (1) abandon the third dimension entirely and introduce a different notation for an IT2 FS, since the T2 FS notation $\tilde{A}$ carries the connotation of a third dimension; or (2) continue to use the notation $\tilde{A}$ but in a simpler way than in (4.6). We choose to do the latter so that the reader will be able to easily connect with the existing literature on IT2 fuzzy systems.

Definition 4.7: As shown in Fig. 4.4, at each value of x, say $x = x'$, imagine that the 2D plane whose axes are u and $\mu_{\tilde{A}}(x',u)$ intersects the 3D type-2 MF $\tilde{A}$, and, consequently, obtain an intersection, which is called a *vertical slice*, denoted $\mu_{\tilde{A}}(x')$, of $\tilde{A}$, i.e.,

$$\mu_{\tilde{A}}(x') = \mu_{\tilde{A}}(x = x',\ u) = \int_{u \in J_{x'}} \mu_{\tilde{A}}(x = x', u)/u \quad J_{x'} \subseteq [0,1] \tag{4.9}$$

The *vertical slice* $\mu_{\tilde{A}}(x')$ is also called the *secondary membership function*, and its domain is the primary membership of x, J_x, where $J_x \subseteq [0, 1]$ for $\forall x \in X$. At each

value of x and u, the amplitude of a secondary membership function, $\mu_A(x,u)$, is called the *secondary grade.*

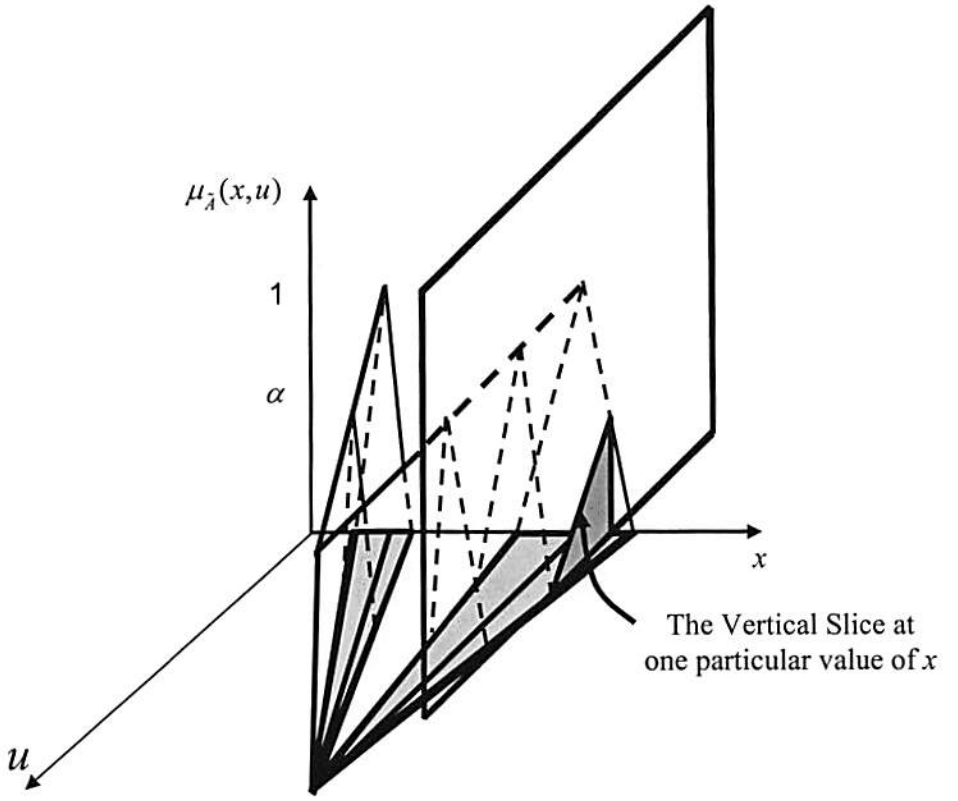


Figure 4.4: An example of the vertical slice for a particular type-2 membership function.

Definition 4.8: The union of all primary memberships, which represents the uncertainty in the primary memberships of a type-2 fuzzy set $\tilde{A}$, is called the *Footprint of uncertainty* (the shaded region in Fig. 4.3), denoted $\mathrm{FOU}(\tilde{A})$, i.e.,

$$\mathrm{FOU}(\tilde{A}) = \bigcup_{x \in X} J_x \tag{4.10}$$

An IT2 FS is completely described by its *footprint of uncertainty* (FOU), and an IT2 FS can therefore be represented in the following simpler way:

$$\tilde{A} = 1/FOU(\tilde{A}) = 1/\bigcup_{\forall x \in X} J_x \tag{4.11}$$

where a secondary grade of 1 has been put for at all elements in the $FOU(\tilde{A})$.

Definition 4.9: The *upper membership function* (UMF) and *lower membership function* (LMF) of $\tilde{A}$ are two T1 MFs that bound the FOU. The UMF is associated with the upper bound of $FOU(\tilde{A})$ and is denoted $\overline{\mu}_{\tilde{A}}(x)$, $\forall x \in X$, and the LMF is associated with the lower bound of $FOU(\tilde{A})$ and is denoted $\underline{\mu}_{\tilde{A}}(x)$, $\forall x \in X$, i. e.[1]

$$\overline{\mu}_{\tilde{A}}(x) \equiv \overline{FOU(\tilde{A})} \quad \forall x \in X \tag{4.12}$$

$$\underline{\mu}_{\tilde{A}}(x) \equiv \underline{FOU(\tilde{A})} \quad \forall x \in X \tag{4.13}$$

Using this notation, we can re-express the $FOU(\tilde{A})$ in (4.10), as

$$FOU(\tilde{A}) = \bigcup_{\forall x \in X} [\underline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{A}}(x)] \tag{4.14}$$

[1] In the T2 FS literature, notations $\underline{\mu}_{\tilde{A}}(x)$ and $\overline{\mu}_{\tilde{A}}(x)$ are used instead of $u_L(x)$ and $u_R(x)$, respectively. Obviously, these notations are interchangeable. We use the former notation, again so that readers can easily connect with other articles in the IT2 FS literature.

As an example, consider a Gaussian primary MF having a fixed mean, m, and an uncertain standard deviation that takes on values in $[\sigma_1, \sigma_2]$, i.e. $\mu_{\tilde{A}}(x) = \exp\{-\frac{1}{2}[(x-m)/\sigma]^2\}$, where $\sigma \in [\sigma_1, \sigma_2]$. Corresponding to each value of σ we get a different MF curve. The lower and upper MFs of this IT2 FS are $\underline{\mu}_{\tilde{A}}(x) = \exp\{-\frac{1}{2}[(x-m)/\sigma_1]^2\}$ and $\overline{\mu}_{\tilde{A}}(x) = \exp\{-\frac{1}{2}[(x-m)/\sigma_2]^2\}$, respectively. Many other examples of IT2 FSs and their lower and upper MFs can be found, e.g. in [21].

Definition 4.10: As shown in Fig. 4.5, for discrete universes of discourse X and J_x, an *embedded type-2 fuzzy set* contains exactly one element from $J_{x_1}, J_{x_2}, \ldots, J_{x_N}$, namely, $u_1, u_2, \ldots, u_N$, each with its associated secondary grade, namely $\mu_{\tilde{A}}(x_1, u_1), \mu_{\tilde{A}}(x_2, u_2), \ldots, \mu_{\tilde{A}}(x_N, u_N)$, i.e.,

$$\tilde{A}_e = \sum_{i=1}^{N} \left[\mu_{\tilde{A}}(x_i, u_i) / u_i \right] / x_i \tag{4.15}$$

An *embedded type-1 fuzzy set* A_e contains exactly one element from $J_{x_1}, J_{x_2}, \ldots, J_{x_N}$, namely, $u_1, u_2, \ldots, u_N$, i.e.,

$$A_e = \sum_{i=1}^{N} u_i / x_i \tag{4.16}$$

Note that A_e acts as the domain for $\tilde{A}_e$. When each u_i has been discretized into M_i levels, there are a total of $\prod_{i=1}^{N} M_i$ embedded T1 FSs, whereas if each u_i has been discretized into the same number of M levels, then there are a total of M^N embedded T1 FSs. An example of an embedded T1 FS is depicted in Fig. 4.5; it is the wavy curve[2] on the $x-u$ plane.

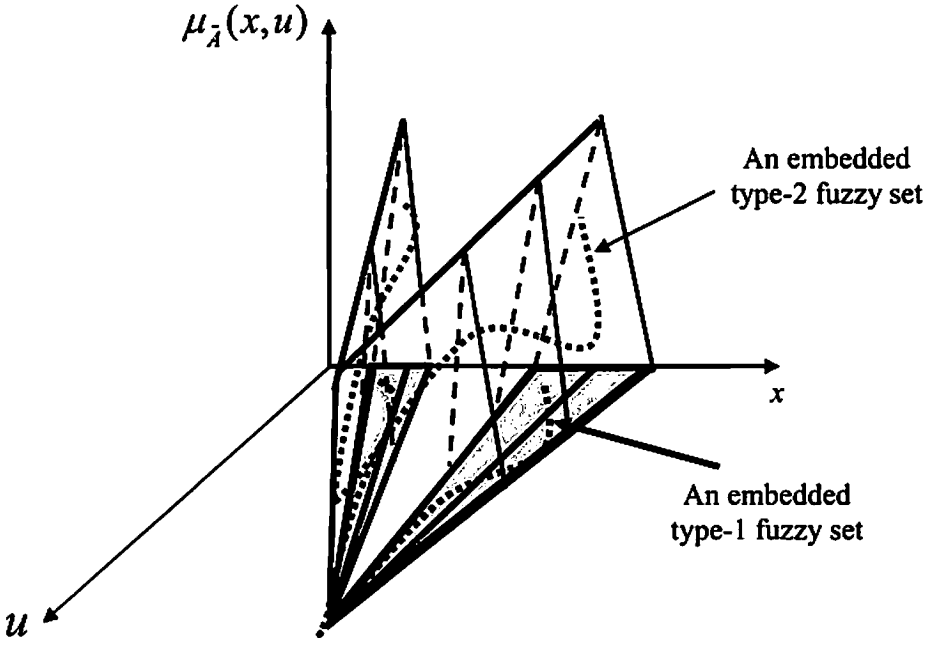


Figure 4.5: An example of embedded type-2 and type-1 fuzzy set.

Theorem 4.2 (Vertical Slice Representation Theorem for T2 FS): The first representation of a T2 FS in the T2 FS literature is the *vertical slice representation*, in

[2] We show the embedded T1 FS as a continuous curve because it is easy to draw such a curve; however, the actual embedded T1 FS, as given by (4.16), is obtained from this curve by sampling it at N values of the primary variable x.

which a type-2 fuzzy set $\tilde{A}$ is represented as the union of its vertical slices, i.e., for the continuous case,

$$\begin{aligned}\tilde{A} &= \int_{x\in X}\int_{u\in J_x} \mu_{\tilde{A}}(x,u)/(x,u) \\ &= \int_{x'\in X}\left[\int_{u\in J_x} \mu_{\tilde{A}}(x=x',u)/u\right]/x' = \int_{x'\in X} \mu_{\tilde{A}}(x')/x'\end{aligned} \tag{4.17}$$

and for the discrete case

$$\begin{aligned}\tilde{A} &= \sum_{x\in X}\left[\sum_{u\in J_x} \mu_{\tilde{A}}(x,u)/u\right]/x \\ &= \sum_{i=1}^{N}\left[\sum_{u\in J_{x_i}} \mu_{\tilde{A}}(x_i,u)/u\right]/x_i = \sum_{i=1}^{N} \mu_{\tilde{A}}(x_i)/x_i\end{aligned} \tag{4.18}$$

The vertical slice representation is very useful for developing practical computation algorithms.

Theorem 4.3 (Embedded Set Representation Theorem for T2 FS): The second representation of a T2 FS is the embedded set representation, in which a general type-2 fuzzy set $\tilde{A}$ is represented as the union of all of its embedded type-2 fuzzy sets, i.e.,

$$\tilde{A} = \sum_{j=1}^{n} \tilde{A}_e^j \tag{4.19}$$

This representation is very useful for theoretical analyses to develop the general structure of the solution to a problem. More important, this representation provides us great understanding about the relationship between type-1 fuzzy sets and type-2 fuzzy sets, as shown in [24].

(4.19) is also called a *wavy slice* representation of $\tilde{A}$. From (4.10), (4.15), (4.16) and (4.19) we see that $FOU(\tilde{A})$ can be obtained as the union of all of the T1 embedded sets, i.e.

$$FOU(\tilde{A}) = \sum_{j=1}^{n_A} A_e^j = \begin{cases} \left\{ \underline{\mu}_{\tilde{A}}(x), ..., \overline{\mu}_{\tilde{A}}(x) \right\} & \forall x \in X_d \\ \left[\underline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{A}}(x) \right] & \forall x \in X \end{cases} \tag{4.20}$$

The top line of (4.20) is for a discrete universe of discourse, X_d, and contains n_A elements (functions), and the bottom line is for a continuous universe of discourse and is an interval set of functions, meaning that it contains an uncountable number of functions that completely fills the space between $\underline{\mu}_{\tilde{A}}(x)$ and $\overline{\mu}_{\tilde{A}}(x)$, for $\forall x \in X$. Equation (4.20) is a *wavy-slice representation of* $FOU(\tilde{A})$, because all of its elements are functions, i.e. they are wavy-slices.

Chapter 5

Encoding Words into Interval Type-2 Fuzzy Sets Using an *Interval Approach*

5.1 Introduction

As mentioned in Chapter 1, the decisions about water-flood management involve a range of activities, some of which are performed only once or twice during the life of the water-flood, e.g., water-flood design and facilities/well planning. Other water-flood activities are performed periodically based on analyses of recurring measurement data, e.g., setting water injection flow rates in the injection wells. How to make decisions based on available information, e.g., the IPRs, is a very difficult problem, especially for large-scale complex system.

One possible way to deal with this problem is to construct a decision making system using computing with words (CWW), *a methodology,* proposed by Zadeh [51]-[52], *in which the objects of computation are words and propositions drawn from a natural language*. CWW is fundamentally different from the traditional expert systems that are tools to realize an intelligent system but are not able to process natural language because it is imprecise, uncertain and partially true.

A specific architecture for CWW using IT2 FSs, called a *Perceptual Computer* (Per-C), proposed in [23], is shown in Fig. 5.1. The Per-C consists of three components: encoder, CWW engine and decoder. The encoder transforms linguistic perceptions, i.e., words, into IT2 FSs that activate a CWW engine. A CWW engine maps its input IT2 FSs into output IT2 FSs, and this can be done in many different ways, including by rules [21], linguistic summarizations [9], [49], linguistic weighted average [46], etc. The decoder maps the IT2 FS outputs of the CWW engine into a specific word [47].

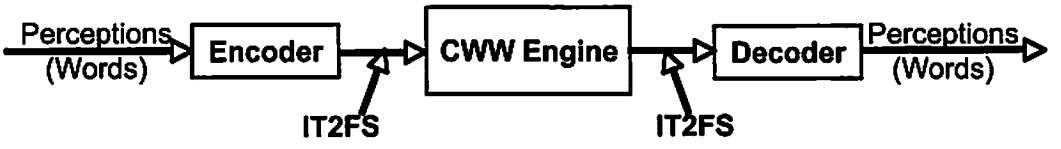


Figure 5.1: The Perceptual Computer (Per-C) for CWW using IT2 FSs.

In order to construct a rule based decision support system using the Per-C, three issues need to be addressed:

1) Design a way to extract rules. Usually, there are two ways to obtain the rules, learning from data and mining from experts using IF-THEN questionnaires. An example of the rules mined from experts may be:

R^l : IF the *injector - producer relation* is *small*,
and the *fracture* between producer and injector is *highly impossible*,
THEN this injector and producer are *highly not* within the same pattern.

The rule antecedents (the IF parts of a rule) injector-producer relation and fracture involve the forecasted injector-producer relationships using our newly developed Extended Kalman Filtering approach, and the geology information about the fracture; and each rule consequent (the THEN part of a rule), e.g., *highly not*, is a word to describe the pattern information. Each expert may provide some rules, and then for a group of experts we can obtain some amount of rules to support decision making.

2) Obtain the membership function for the words such as *small*, *highly impossible*, i.e., encoding problem.
3) Map the IT2 FS outputs of the CWW engine into a specific word [27], i.e., decoder problem.

This chapter is about the encoding problem, i.e., how to transform a word into an IT2 FS. A collection of application-specific words and their footprints of uncertainty (FOUs) (see Def. 4.8) is a *codebook* for the application. The *codebook* is also needed for the decoder.

How to collect data from a group of subjects, and how to then map that data into the parameters of a T1 MF has been reported on by a number of authors [e.g., Klir and Yuan (1995)] Names for the different T1 methods include: *polling*, *direct rating*, *reverse rating*, *interval estimation*, and *transition interval estimation*. Unfortunately, none of these methods transfers the uncertainties about collecting word data from a group of subjects into the MF of a T1 FS, because a T1 FS does not have enough degrees of freedom to do this; hence, they are not elaborated upon in this thesis.

Recently [30], two approaches have been described for collecting data about a word from a group of subjects and then mapping that data into an FOU for that word, the *person membership function (MF) approach* and the *interval end-points approach*. Both approaches have been referred to as *type-2 fuzzistics* [30].

In the *person MF approach* a subject provides their FOU for a word on a prescribed scale (e.g., 0-10), and this is done for a group of subjects. Each person FOU captures the intra-level of uncertainty about a word, i.e. the uncertainty that each subject has about the word. All of the person-FOUs are aggregated, which captures the inter-level of uncertainty about the word across the group of subjects. Finally, an IT2 FS model is fit to the aggregated data. Note that when the aggregation operation is the union then this approach is based on the T2 FS Representation Theorem (See Theorem 4.3) for an IT2 FS, which states that the FOU of an IT2 FS equals the union of all of its embedded T1 FSs. Each subject's person FOU can also be interpreted as a union of its embedded T1 FSs.

Strong points of this approach are: (1) The union of the person FOUs (the data) establishes the shape of the FOU directly; (2) No statistics about the data are used, i.e. all of the data (the person FOUs) are used so that no information is lost; (3) An a priori uncertainty measure of an FOU is not required in order to map the person FOUs into the IT2 FS that is fit to the aggregated data; and, (4) If all uncertainty disappears then the IT2 FS word model reduces to a T1 FS word model[3].

The weak point of this approach is it requires subjects to be knowledgeable about fuzzy sets. Unfortunately, this weakness may be so large that (in the opinion of the

[3] In this case, all subjects provide the same person T1 MF (not an FOU).

authors) it may obliterate the advantages of the approach; hence, the person FOU approach is very limited in applicability.

In the *interval end-points approach* each subject provides the end-points of an interval associated with a word on a prescribed scale. Subjects are asked:

> On a scale of 0-10, what are the end-points of an interval that you associate with the word *W*?

The mean and standard deviation are then computed for the two end-points using the data collected from all of the subjects. Doing this captures the inter-level uncertainties about the word. These end-point statistics are then mapped into an IT2 FS model for the word by bounding the end-points of the centroid of a pre-specified FOU. Mendel and Wu [26] have shown that the centroid of an IT2 FS is a measure of the uncertainty of such a FS, and that, although the centroid cannot be computed in closed form [10], centroid bounds[4] can be expressed explicitly in terms of the geometry (parameters) of an FOU. They provide formulas for these bounds for many different FOUs [26]. Design equations are then used to map the end-point statistics into FOU parameters. Explicit mappings have been provided only for symmetrical FOUs [27].

The strong point of this approach is that collecting interval end-point data is not limited to people who are knowledgeable about fuzzy sets; hence, it is presently a preferable way to collect data from people. Weak points of this approach are: (1) Closed-form mappings are only available for symmetrical FOUs that are associated with data intervals whose two end-point standard deviations are approximately equal, whereas actual interval end-point data show that most words do not have equal end-

[4] The centroid is an interval-valued set, and centroid bounds are lower and upper bound for both the left- and right- end points of that set.

point standard deviations [20], [21], [30]; (2) The shape of an FOU must be chosen ahead of time, independent of the end-point statistics; (3) The centroid has to be chosen as the measure of uncertainty of an FOU, and because no closed-form formulas are available for the centroid, uncertainty-bound surrogates must be used in their place; (4) Only data end-point mean and standard deviations are used to model word data-uncertainties; and, (5) If all uncertainty disappears (i.e., all subjects provide the same intervals) then the IT2 FS word model does not reduce to a T1 FS word model[5]. Regardless of all of the weaknesses of this approach, its strong point is so large that it has masked the weak points, and has been the recommended approach.

In this chapter, a new and simple approach, called the *Interval Approach* to type-2 fuzzistics is presented, one that captures the strong points (enumerated in Section 5.2) of both the person-MF and interval end-points approaches.

The rest of this chapter is organized as follows: Section 5.2 explains the Interval Approach, and provides some observations about the IA; Section 5.3 contains an extensive 32 words codebook example; and, Section 5.4 draws conclusions. Detailed proofs are provided in the appendixes.

5. 2 Interval approach

To begin, the *Interval Approach* (*IA*) to T2 fuzzistics is overviewed. The IA captures the strong points of both the person-MF and interval end-points approaches, i.e. it: (1) Collects interval end-point data from a group of subjects; (2) Does not require the

[5] This last point seems to have been missed when Mendel and Wu [27] set up their two design equations.

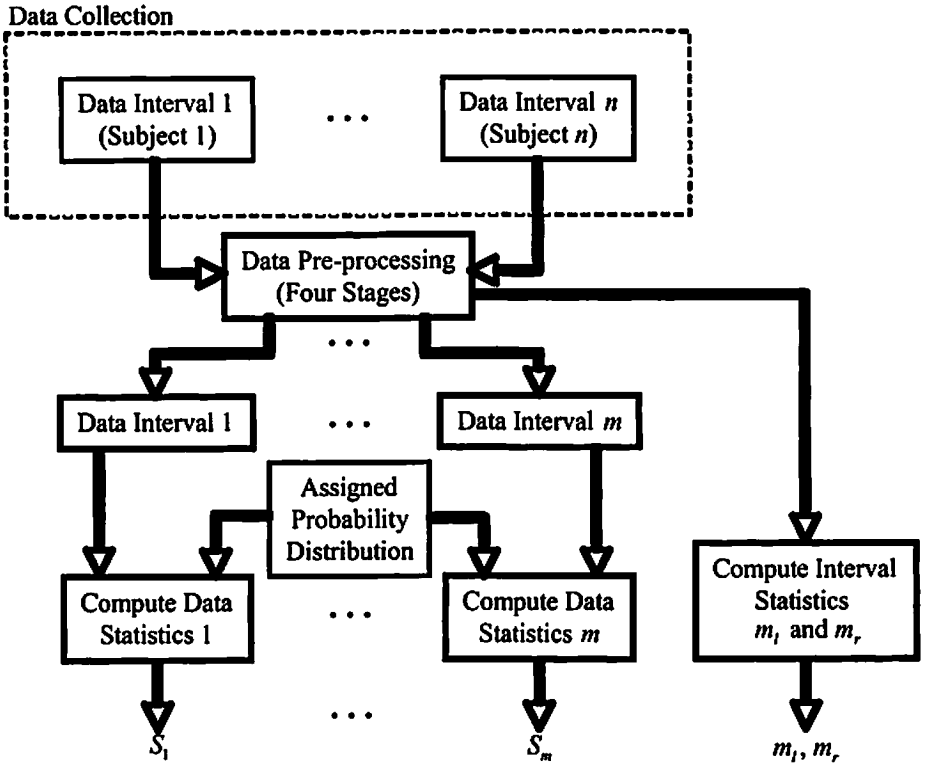


Figure 5.2: *Data Part* of the IA approach. Note that the data statistics, $S_1,...,S_m$, and interval statistics, m_l and m_r, feed into the *FS Part* of IA, in Fig. 5.5. The extra heavy lines and blocks denote the flow of processing once the data are collected.

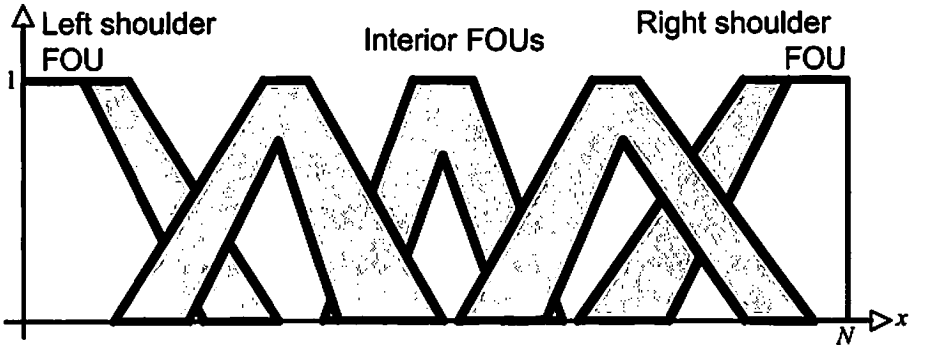


Figure 5.3: Left-shoulder, right-shoulder and interior FOUs, all of whose LMFs and UMFs are piecewise linear.

subjects to be knowledgeable about fuzzy sets; (3) Has a straightforward mapping from data to an FOU; (4) Does not require an a priori assumption about whether or not an FOU is symmetric or non-symmetric; and, (5) Leads to an IT2 FS word model that reduces to a T1 FS word model automatically if all subjects provide the same intervals.

The basic idea of the IA is to map each subject's data interval into a pre-specified T1 person MF, and to interpret the latter as an embedded T1 FS of an IT2 FS (this is motivated by the Representation Theorem for an IT2 FS. The IA consists of two parts, the *Data Part* (Fig. 5.2) and the *Fuzzy Set (FS) Part* (Fig. 5.5). In the Data Part, data that have been collected from a group of subjects are pre-processed, after which data statistics are computed for the surviving data intervals. In the FS Part, FS uncertainty measures are established for a pre-specified T1 MF [always beginning with the assumption that the FOU is an interior FOU (Fig. 5.3), and, if need be, later switching to a shoulder FOU (Fig. 5.3)]. Then the parameters of the T1 MF are determined using the data statistics, and the derived T1 MFs are aggregated using union leading to an FOU for a word, and finally to a mathematical model for the FOU.

5.2.1 Interval approach: data part

Once data intervals have been collected from a group of subjects the Data Part of the IA consists of two major steps: (1) Preprocessing the data intervals that have been collected from a group of subjects, and (2) Computing statistics for the data intervals that have survived the pre-processing step. The details of these steps are described in this section.

A. Data Preprocessing

Preprocessing (Fig. 5.2) the n interval end-point data $[a^{(i)},b^{(i)}]$ $(i=1,...,n)$ consists of *four stages*: (1) bad data processing, (2) outlier processing, (3) tolerance-limit processing, and (4) reasonable-interval processing. As a result of data preprocessing, some of the n interval data are discarded and the remaining m intervals are re-numbered, $1,2,...,m$. In the rest of this section details are provided for each of these four stages.

A.1 Stage 1–Bad data processing: Such processing removes nonsensical results (some subjects do not take a survey seriously and provide useless results). If interval end-points satisfy

$$\left.\begin{array}{c} 0 \le a^{(i)} \le 10 \\ 0 \le b^{(i)} \le 10 \\ b^{(i)} \ge a^{(i)} \end{array}\right\} \; i=1,...,n, \tag{5.1}$$

then an interval is accepted; otherwise, it is rejected. These conditions are obvious and do not need further explanations. After bad data processing there will be $n' \le n$ remaining data intervals.

A.2 Stage 2–Outlier processing: Such processing uses a Box and Whisker test [43] to eliminate outliers. Recall that outliers are points that are unusually large or small.

A Box and Whisker test is usually stated in terms of first and third quartiles and an inter-quartile range. Recall that the first and third quartiles, $Q(0.25)$ and $Q(0.75)$,

contain 25% and 75% of the data, respectively. The inter-quartile range, IQR, is the difference between the third and first quartiles; hence, IQR contains 50% of the data between the first and third quartiles.

Any point that is more than $1.5IQR$ above the third quartile or more than $1.5IQR$ below the first quartile is considered an outlier [43]. Consequently, if the subject interval end-points satisfy

$$\left.\begin{aligned} a^{(i)} &\in [Q_a(0.25) - 1.5IQR_a, Q_a(0.75) + 1.5IQR_a] \\ b^{(i)} &\in [Q_b(0.25) - 1.5IQR_b, Q_b(0.75) + 1.5IQR_b] \end{aligned}\right\} \; i = 1,\ldots,n' \qquad (5.2)$$

a data interval is accepted; otherwise, it is rejected. In these equations, Q_a (Q_b) and IQR_a (IQR_b) are the quartile and inter-quartile ranges for the left (right) end-points.

After outlier processing, there will be $m' \le n'$ remaining data intervals for which the following data statistics are then computed: m_l, s_l (sample mean and standard deviation of the m' left end-points) and m_r, s_r (sample mean and standard deviation of the m' right end-points).

A.3 Stage 3–Tolerance limit processing: If a data interval $[a^{(i)}, b^{(i)}]$ satisfies [43]

$$\left.\begin{aligned} a^{(i)} &\in [m_l - ks_l, m_l + ks_l] \\ b^{(i)} &\in [m_r - ks_r, m_r + ks_r] \end{aligned}\right\} \; i = 1,\ldots,m' \qquad (5.3)$$

it is accepted, otherwise it is rejected. In (5.3) k is determined as follows.

Recall that for a normal distribution [43] of measurements with unknown mean and standard deviation, *tolerance limits* are given by $m_l \pm ks_l$ (or $m_r \pm ks_r$), where *tolerance factor k* is determined so that one can assert with $100(1-\gamma)\%$ confidence that the given limits contain at least the proportion $1-\alpha$ of the measurements. Table 5.1 (adapted from Table A.7 in[6] [43]) gives k for eight values of m', two values of $1-\gamma$ and two values of $1-\alpha$. Knowing m' and choosing values for $1-\gamma$ and $1-\alpha$, one can obtain k. If, e.g., $k = 2.549$ (for which $m' = 30$) then one can assert with 95% confidence that the given limits contain at least 95% of the subject data intervals.

Table 5.1: Tolerance factor k for a number of collected data (m'), a proportion of the data ($1-\alpha$), and a confidence level $1-\gamma$ [43].

	$1-\gamma = 0.95$		$1-\gamma = 0.99$	
	$1-\alpha$		$1-\alpha$	
m'	0.90	0.95	0.90	0.95
10	2.839	3.379	3.582	4.265
15	2.480	2.954	2.945	3.507
20	2.310	2.752	2.659	3.168
30	2.140	2.549	2.358	2.841
50	1.996	2.379	2.162	2.576
100	1.874	2.233	1.977	2.355
1000	1.709	2.036	1.736	2.718
∞	1.645	1.960	1.645	1.960

Assumption: Data interval end-points are approximately normal, so that the tolerance limits that are given in Table 5.1 can be used.

Note that m' may be different for each word, because it is a result from pre-processing stages 1 and 2, and those stages are applied independently to each word.

[6] Their table is in turn adapted from [2], and contains entries for 47 values of n, beginning with $n = 2$.

After tolerance limit processing, there will be $m'' \leq m'$ remaining data intervals ($1 \leq m'' \leq n$), and the following data statistics are then recomputed: m_l, s_l (sample mean and standard deviation of the m'' left end-points) and m_r, s_r (sample mean and standard deviation of the m'' right end-points).

A.4 Stage 4–Reasonable-interval processing: In one of our first attempts at the IA [16], only the first three stages of data pre-processing were used. FOUs were obtained that did not look so good (this is subjective, but is demonstrated in Section VI), and many were filled in or almost filled in, i.e. $LMF(\tilde{A}) \approx 0$. Because the centroid of a filled-in FOU is completely independent of $UMF(\tilde{A})$ [32], such an FOU is not considered to be a good one. As a result, something else had to be done.

It dawned on us that, in addition to focusing on *words mean different things to different people* (which was our rationale for using IT2 FS models for words), one also needs to focus on *words mean similar things to different people.* In fact, if there is understanding about a word across a group of subjects, it is the latter that causes it. This led us to require only overlapping intervals be kept. Such intervals are called *reasonable*, i.e.:

Definition 5.1: A data interval is said to be *reasonable* if it overlaps with another data interval in the sense of Fig. 5.4.

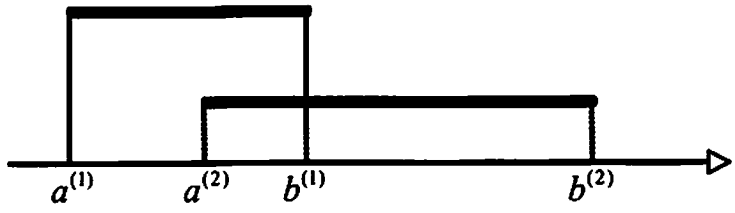


Figure 5.4: An example of two overlapping data intervals for the same word.

In the last step of data preprocessing only reasonable data intervals are kept. Appendix B provides a derivation of the following:

Reasonable-Interval Test: IF

$$\left.\begin{array}{l} a^{(i)} < \xi^* \\ b^{(i)} > \xi^* \end{array}\right\} \forall i = 1,\ldots,m'' \tag{5.4}$$

where ξ^* is one of the values

$$\xi^* = \frac{(m_r\sigma_l^2 - m_l\sigma_r^2) \pm \sigma_l\sigma_r\left[(m_l - m_r)^2 + 2(\sigma_r^2 - \sigma_l^2)\ln(\sigma_l / \sigma_r)\right]^{1/2}}{\sigma_l^2 - \sigma_r^2} \tag{5.5}$$

such that

$$m_l \le \xi^* \le m_r, \tag{5.6}$$

THEN the data interval is kept; OTHERWISE, it is deleted.

As a result of reasonable-interval processing, some of the m'' data intervals may be discarded and there will finally be m remaining data intervals ($1 \le m \le n$) that are re-numbered, $1,2,\ldots,m$.

B. Computing statistics

Two kinds of statistics are computed, one for all of the remaining m data intervals, and one for individual data intervals. Both kinds are used in the Fuzzy Part of the IA.

B.1 Compute interval statistics: The sample means of the left and right end-points, m_l and m_r, of the remaining m data intervals are computed. They are used as described in Section 5.2.2.E.

B.2 Compute data statistics for each interval: A probability distribution is assigned to *each* of the m surviving data intervals after which statistics are computed for each interval using the assumed probability model and the interval end-points. These statistics are used as described in Section 5.2.2.D.

Although many choices are possible for an assumed probability distribution for a subject's data interval, unless a subject provides more information about that interval (e.g., a greater belief in the center of the interval) then a *uniform distribution* is most sensible, and is the one chosen herein[7]. According to Dubois et al. [3], " ... a uniform probability distribution on a bounded interval ... is the most natural probabilistic

[7] Dubois et al. [3] explain how to map a collection of confidence intervals into a symmetrical triangle T1 MF, where the confidence intervals are associated with data that are collected from a group of subjects about a single point. More specifically, in their problem n measurements, $y_1, y_2, ..., y_n$, are collected, after which the sample mean, $\bar{m}_y$, is computed, as $\bar{m}_y = \sum_{i=1}^{n} y_i / n$. The α confidence intervals of $\bar{m}_y$, denoted $[\underline{CI}(\alpha), \overline{CI}(\alpha)]$, are then computed for a fixed value of α. When each confidence interval is assumed uniformly distributed, their method maps the confidence intervals into a symmetric triangular fuzzy number.

Note, however, their problem is different from ours, because in our problem we begin with a collection of n intervals rather than with collection of n numbers; so, their results have not been used by us.

representation of incomplete knowledge when only the support is known. It is non-committal in the sense of maximal entropy ... and it applies Laplace's indifference principle stating that what is equipossible is equiprobable."

In order to keep things as simple as possible, only two statistics are used for a uniform distribution, its mean and standard deviation. Recall, that if a random variable Y is uniformly distributed in $[a,b]$ [43], then

$$m_Y = \frac{a+b}{2} \tag{5.7}$$

$$\sigma_Y = \frac{b-a}{\sqrt{12}} \tag{5.8}$$

In the second stage of the Data Part, data statistics $S_1,...,S_m$ are computed for each interval, where

$$S_i = (m_Y^{(i)}, \sigma_Y^{(i)}) \;\; i = 1,...,m, \tag{5.9}$$

and these data statistics are then used in the FS part of the IA where they are mapped into the parameters of a T1 MF, as explained in Section 5.2.2.E.

This completes the Data Part of the IA.

5.2.2 Interval approach: fuzzy set (FS) part

The FS Part of the IA (Fig. 5.5) consists of nine steps, each of which is described in this section.

A. Choose a T1 FS Model

In the present IA, because the mapping from an interval of data to a T1 MF only uses the mean and variance of the (just) assumed uniform probability distribution, only T1 MFs with two degrees of freedom can be used. In this chapter only a symmetrical triangle interior T1 MF, or a left-shoulder T1 MF, or a right-shoulder T1 MF are used.

B. Establish FS Uncertainty Measures

Although many choices are possible for uncertainty measures of a T1 FS[8] [11] our approach is to focus on simplicity, and to therefore use the mean and standard deviation of a T1 FS.

Definition 5.2: The mean and standard deviation of a T1 FS A are:

$$m_A = \frac{\int_{a_{MF}}^{b_{MF}} x\mu_A(x)dx}{\int_{a_{MF}}^{b_{MF}} \mu_A(x)dx} \tag{5.10}$$

[8] Klir [11] states that there are in principle three uncertainty measures, namely generalized Hartley measure, generalized Shannon measure, and an aggregate measure; however, it has been found by us that these measures are too complicated to use in the IA.

$$\sigma_A = \left[\frac{\int_{a_{MF}}^{b_{MF}} (x - m_A)^2 \mu_A(x) dx}{\int_{a_{MF}}^{b_{MF}} \mu_A(x) dx} \right]^{1/2} \tag{5.11}$$

where a_{MF} and b_{MF} are the parameters of the MFs that are depicted in the figures of Table 5.2.

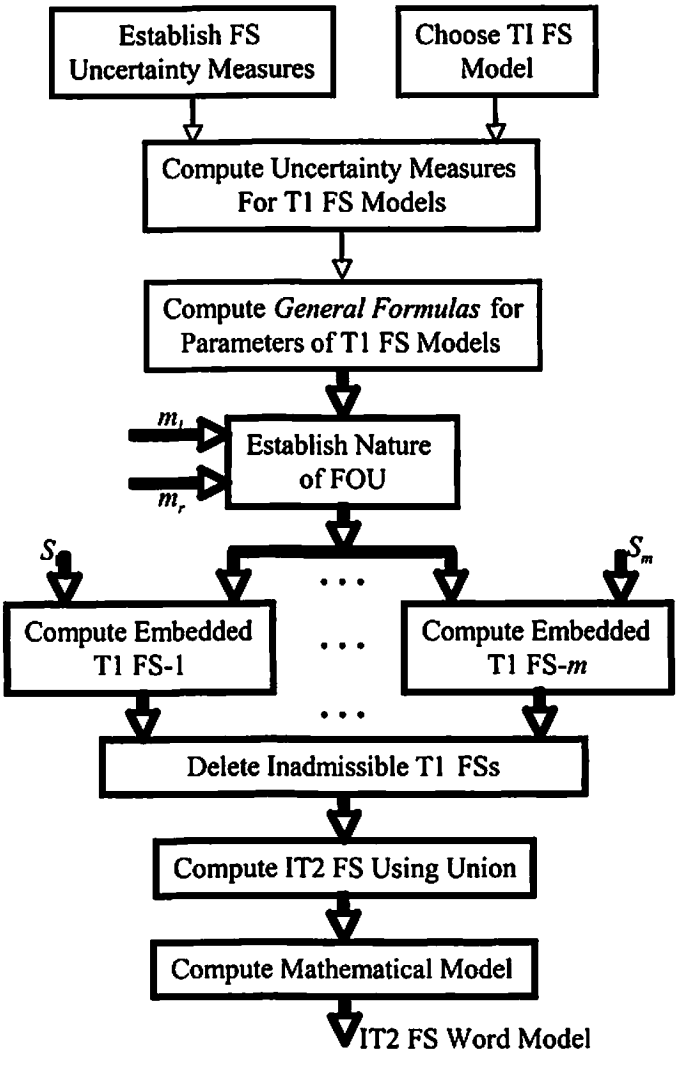
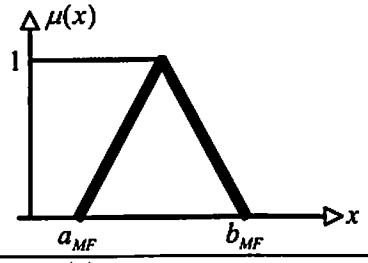
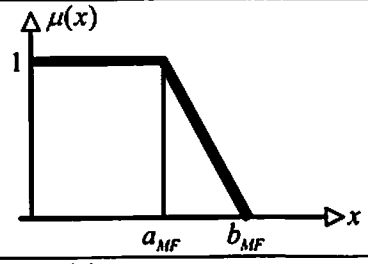
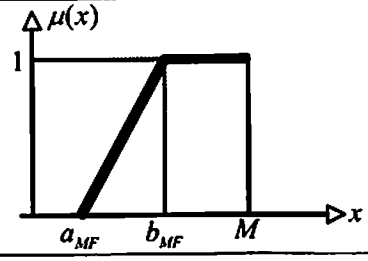


Figure 5.5: *FS Part* of the IA approach. The extra heavy lines and blocks denote the flow of processing once the data are collected.

Obviously, if $\mu_A(x) \Big/ \int_{a_{MF}}^{b_{MF}} \mu_A(x)dx$ is the probability of x, where $x \in [a_{MF}, b_{MF}]$, then (5.10) and (5.11) are the same as the mean and standard deviation used in probability.

Usually, a_{MF} and b_{MF} denote the left-end and right-end of the support of a T1 MF; however, shoulder T1 MFs pose a problem because, for a left shoulder T1 MF there is no uncertainty for $x \in [0, a_{MF}]$, whereas for a right shoulder T1 MF there is no uncertainty for $x \in [b_{MF}, M]$; hence, for shoulder MFs a_{MF} and b_{MF} do not cover the entire span of the MF, and are as shown in the second and third row figures of Table 5.2.

Table 5.2: Mean and standard deviation for interior and shoulder T1 MFs [16].

Name	MF	Mean (m_{MF}) Standard deviation (σ_{MF})
Symmetric triangle (interior MF)		$m_{MF} = (a_{MF} + b_{MF})/2$ $\sigma_{MF} = (b_{MF} - a_{MF})/2\sqrt{6}$
Left-shoulder		$m_{MF} = (2a_{MF} + b_{MF})/3$ $\sigma_{MF} = \left[\frac{1}{6}\left[(a_{MF} + b_{MF})^2 + 2a_{MF}^2\right] - m_{MF}^2\right]^{1/2}$
Right-shoulder		$m_{MF} = (2a_{MF} + b_{MF})/3$ $\sigma_{MF} = \left[\frac{1}{6}\left[(a'_{MF} + b'_{MF})^2 + 2a'^2_{MF}\right] - m'^2_{MF}\right]^{1/2}$ $a'_{MF} = M - b_{MF}$ $b'_{MF} = M - a_{MF}$ $m'_{MF} = M - m_{MF}$

C. Compute Uncertainty Measures for T1 FS Models

The mean and standard deviations for symmetric triangle (interior), left-shoulder and right-shoulder T1 MFs are easy to compute, and they are also summarized in Table II. Observe that, by using the primed parameters for the right-shoulder T1 MF, the equation for its σ_{MF} looks just like the comparable formula for the left-shoulder T1 MF.

D. Compute General Formulas for Parameters of T1 FS Models

The parameters of a T1 FS (triangle, left- or right-shoulder) are computed by equating the mean and standard deviation of a T1 FS to the mean and standard deviation, respectively, of a data interval, i.e., $m_{MF}^{(i)} = m_Y^{(i)}$ and $\sigma_{MF}^{(i)} = \sigma_Y^{(i)}$, where $m_{MF}^{(i)}$ and $\sigma_{MF}^{(i)}$ are in Table 5.2, and $m_Y^{(i)}$ and $\sigma_Y^{(i)}$ are in Table 5.3. This is done for each of the m remaining data intervals. The resulting T1 MF parameters, $a_{MF}^{(i)}$ and $b_{MF}^{(i)}$, are summarized in Table 5.3.

Table 5.3: Transformations of the uniformly distributed data interval $[a^{(i)}, b^{(i)}]$ into the parameters $a_{MF}^{(i)}$ and $b_{MF}^{(i)}$ of a T1 FS [16].

MF	Transformations
Symmetric triangle (interior MF)	$a_{MF}^{(i)} = \frac{1}{2}[(a^{(i)} + b^{(i)}) - \sqrt{2}(b^{(i)} - a^{(i)})]$ $b_{MF}^{(i)} = \frac{1}{2}[(a^{(i)} + b^{(i)}) + \sqrt{2}(b^{(i)} - a^{(i)})]$
Left-shoulder	$a_{MF}^{(i)} = \frac{(a^{(i)} + b^{(i)})}{2} - \frac{(b^{(i)} - a^{(i)})}{\sqrt{6}}$ $b_{MF}^{(i)} = \frac{(a^{(i)} + b^{(i)})}{2} + \frac{\sqrt{6}(b^{(i)} - a^{(i)})}{3}$
Right-shoulder	$a_{MF}^{(i)} = M - \frac{(a'^{(i)} + b'^{(i)})}{2} - \frac{\sqrt{6}(b'^{(i)} - a'^{(i)})}{3}$ $b_{MF}^{(i)} = M - \frac{(a'^{(i)} + b'^{(i)})}{2} + \frac{(b'^{(i)} - a'^{(i)})}{\sqrt{6}}$ $a'^{(i)} = M - b^{(i)}$ $b'^{(i)} = M - a^{(i)}$

E. Establish Nature of FOU

Given a set of m data intervals, they must be mapped into an interior FOU, left-shoulder FOU or a right-shoulder FOU. This is a *classification problem*. In this sub-section, a rationale for deciding which of these FOUs is chosen, and an FOU classification procedure is explained.

E.1 FOU Rationale: To begin, it is always assumed that the m data intervals can be mapped into an interior FOU, and if this cannot be done that the data can be mapped into a left-shoulder FOU, and if this cannot be done that the data can be mapped into a right-shoulder FOU. This rationale is the basis for the classification procedure that is given next.

E.2 FOU classification procedure: To begin, the following *Admissibility Requirement* for an interior FOU is defined.

Definition 5.3: For the scale [0, 10], an interior FOU is said to be *admissible* if and only if

$$\left.\begin{array}{l} a_{MF}^{(i)} \geq 0 \\ b_{MF}^{(i)} \leq 10 \end{array}\right\} \forall i = 1,\ldots,m \qquad (5.12)$$

By using the formulas for $a_{MF}^{(i)}$ and $b_{MF}^{(i)}$ that are given in the first row of Table 5.3, it is straightforward to show that (12) is equivalent to:

$$\left.\begin{array}{l} 1.207a^{(i)} - 0.207b^{(i)} \geq 0 \\ 1.207b^{(i)} - 0.207a^{(i)} \leq 10 \end{array}\right\} \forall i = 1,\ldots,m, \tag{5.13}$$

or, equivalently,

$$\left.\begin{array}{l} b^{(i)} \leq 5.831a^{(i)} \\ b^{(i)} \leq 0.171a^{(i)} + 8.29 \end{array}\right\} \forall i = 1,\ldots,m \tag{5.14}$$

Additionally, there is the obvious constraint that:

$$b^{(i)} \geq a^{(i)} \ \forall i = 1,\ldots,m \tag{5.15}$$

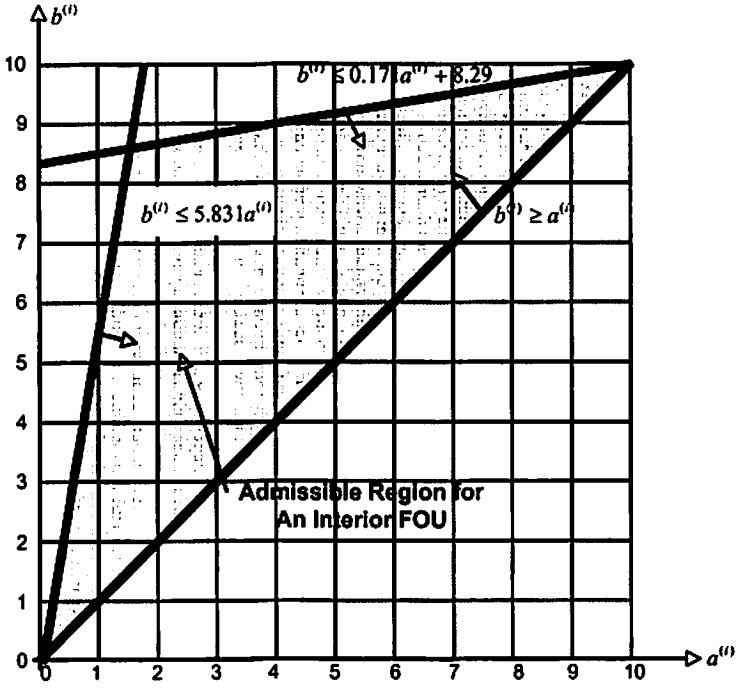


Figure 5.6: Admissible region for an interior FOU that is based on (5.14) and (5.15).

Fig. 5.6 depicts the three inequalities in (5.14) and (5.15) and shows the admissible region for an interior FOU. Unfortunately, requiring (5.14) and (5.15) to be satisfied for all m data intervals is too stringent, as explained next.

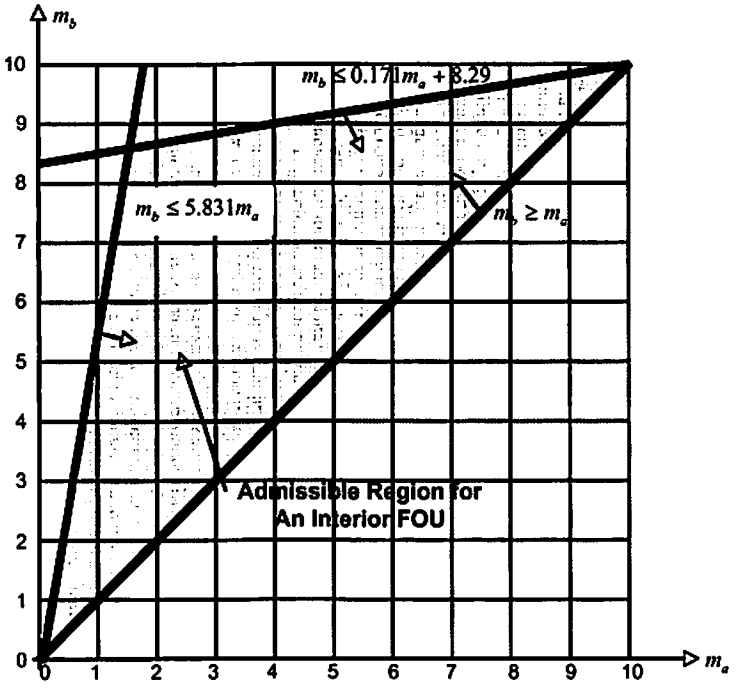


Figure 5.7: Admissible region for an interior FOU that is based on (5.16) and (5.17).

Consider two situations, where in the first situation only one of the m data pairs (barely) falls outside of the admissible region, whereas in the second situation more than half of the data pairs fall outside of that region. Using (5.14) and (5.15) an interior FOU would be rejected for both situations, which does not seem so reasonable; hence, requiring all $\{a^{(i)}, b^{(i)}\}_{i=1}^{m}$ to fall in the admissible region seems too stringent.

To that end, instead of using (5.14) and (5.15), their expected values are used, i.e.

$$\left.\begin{aligned} m_b &\le 5.831 m_a \\ m_b &\le 0.171 m_a + 8.29 \end{aligned}\right\} \tag{5.16}$$

$$m_b \ge m_a \tag{5.17}$$

The figure that is analogous to Fig. 5.6 is now depicted in Fig. 5.7. It looks just like the one in Fig. 5.6, except that the lines are for (5.16) and (5.17) instead of for (5.14) and (5.15).

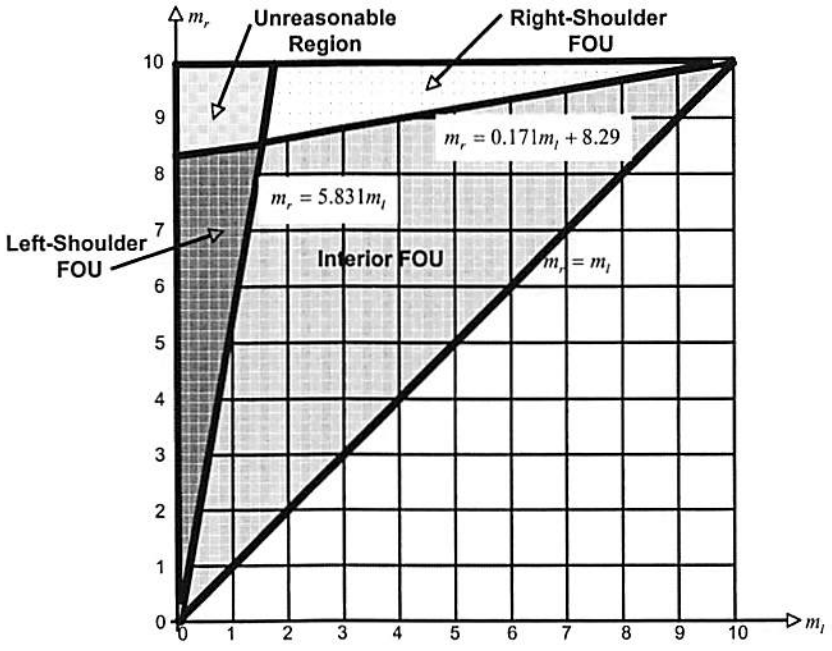


Figure 5.8: FOU decision regions.

In practice, the population-means, m_a and m_b, are not available. Instead, they are approximated using the interval end-point sample means, m_l and m_r, which are available from the Data Part of the IA, i.e.

$$\left.\begin{aligned} m_a &\approx m_l \\ m_b &\approx m_r \end{aligned}\right\} \tag{5.18}$$

Based on these discussions, our *FOU Classification Procedure* is: Using m_l and m_r,

$$\left.\begin{array}{ll} \text{IF } m_r \le 5.831 m_l,\ m_r \le 0.171 m_l + 8.29, \text{ and } m_r \ge m_l, \text{ THEN FOU is an} & \text{INTERIOR FOU} \\ \text{OTHERWISE, IF } m_r > 5.831 m_l, \text{ and } m_r < 0.171 m_l + 8.29, \text{ THEN FOU is a} & \text{LEFT SHOULDER FOU} \\ \text{OTHERWISE, IF } m_r > 0.171 m_l + 8.29, \text{ and } m_r < 5.831 m_l & \text{RIGHT SHOULDER FOU} \\ \text{OTHERWISE, IF } m_r > 0.171 m_l + 8.29, \text{ and } m_r > 5.831 m_l & \text{NO FOU} \end{array}\right\} \tag{5.19}$$

The FOU decision regions are depicted in Fig. 5.8. Observe that there is a small region for which no FOU is assigned. It is called the *Unreasonable Region*, because to assign a shoulder FOU for values in it leads to FOUs that extend unreasonably far to the left (for a right-shoulder FOU) or right (for a left-shoulder FOU). No interval data that we have collected have led to (m_l, m_r) that fall in the Unreasonable Region.

F. Compute Embedded T1 FSs

Once a decision has been made as to the kind of FOU for a specific word, each of the word's remaining m data intervals are mapped into their respective T1 FSs using the

equations that are given in Table III, i.e.

$$(a^{(i)},b^{(i)}) \rightarrow (a_{MF}^{(i)},b_{MF}^{(i)}),\ i=1,...,m \tag{5.20}$$

These T1 FSs, denoted $A^{(i)}$, are called *embedded T1 FSs*, because they will be used to obtain the FOU of the word, as described below in Section H.

G. Delete Inadmissible T1 FSs

It is possible that some of the m embedded T1 FSs are inadmissible, i.e. they violate (5.12). Those T1 FSs are deleted, so that there will be m^* remaining embedded T1 FSs, where $m^* \leq m$.

H. Compute an IT2 FS Using the Union

Using the Representation Theorem for an IT2 FS, a word's IT2 FS $\tilde{A}$ is computed as:

$$\tilde{A} = \bigcup_{i=1}^{m^*} A^{(i)} \tag{5.21}$$

where $A^{(i)}$ is the just-computed i^{th} embedded T1 FS.

I. Compute Mathematical Model for $FOU(\tilde{A})$

In order to compute a mathematical model for $FOU(\tilde{A})$, both $UMF(\tilde{A})$ and $LMF(\tilde{A})$ must be approximated. There are many ways in which this can be done. Our

approach is very simple and guarantees that all m^* embedded T1 FSs are contained within $FOU(\tilde{A})$. Regardless of the type of FOU, the following four numbers must first be computed:

$$\left.\begin{aligned} \underline{a}_{MF} &\equiv \min_{i=1,\ldots,m^*}\{a_{MF}^{(i)}\} \\ \overline{a}_{MF} &\equiv \max_{i=1,\ldots,m^*}\{a_{MF}^{(i)}\} \end{aligned}\right\} \tag{5.22}$$

$$\left.\begin{aligned} \underline{b}_{MF} &\equiv \min_{i=1,\ldots,m^*}\{b_{MF}^{(i)}\} \\ \overline{b}_{MF} &\equiv \max_{i=1,\ldots,m^*}\{b_{MF}^{(i)}\} \end{aligned}\right\} \tag{5.23}$$

I.1 Mathematical model for an interior FOU: Fig. 5.9 depicts this situation. The steps to approximate $UMF(\tilde{A})$ are:

a. Compute

$$C_{MF}^{(i)} = \frac{a_{MF}^{(i)} + b_{MF}^{(i)}}{2} \tag{5.24}$$

b. Compute

$$\underline{C}_{MF} = \min\{C_{MF}^{(i)}\} \tag{5.25}$$

$$\overline{C}_{MF} = \max\{C_{MF}^{(i)}\} \tag{5.26}$$

c. Connect the following points with straight lines: $(\underline{a}_{MF}, 0)$, $(\underline{C}_{MF}, 1)$, $(\overline{C}_{MF}, 1)$ and $(\overline{b}_{MF}, 0)$.

The result is a trapezoidal UMF.

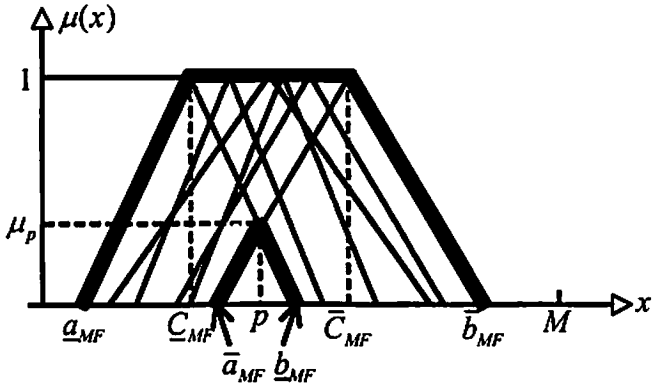


Figure 5.9: An example of the union of T1 triangle MFs. The heavy lines are the lower and upper MFs for the interior FOU.

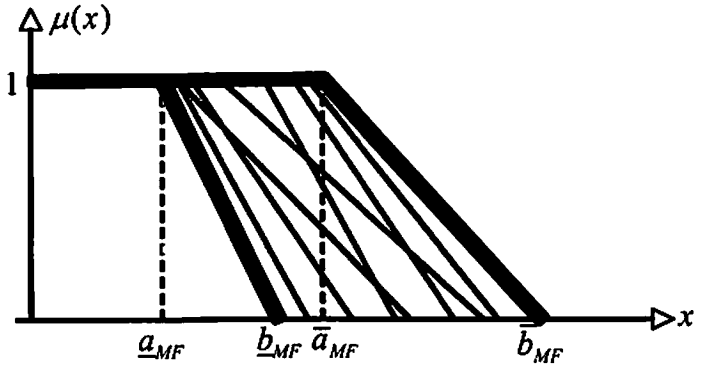


Figure 5.10: An example of the union of T1 left-shoulder MFs. The heavy lines are the lower and upper MFs for the left-shoulder FOU.

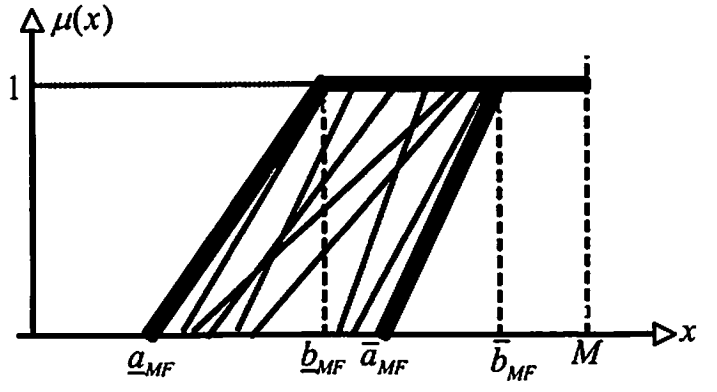


Figure 5.11: An example of the union of T1 right-shoulder MFs. The heavy lines are the lower and upper MFs for the right-shoulder FOU.

The steps to approximate $LMF(\tilde{A})$ are:

a. Compute the intersection point (p, μ_p) of the right leg and the left leg of the left and right most-extreme triangles (see Fig. 5.9), using:

$$p = \frac{\underline{b}_{MF}(\bar{C}_{MF} - \bar{a}_{MF}) + \bar{a}_{MF}(\underline{b}_{MF} - \underline{C}_{MF})}{(\bar{C}_{MF} - \bar{a}_{MF}) + (\underline{b}_{MF} - \underline{C}_{MF})} \tag{5.27}$$

$$\mu_p = \frac{\underline{b}_{MF} - p}{\underline{b}_{MF} - \underline{C}_{MF}} \tag{5.28}$$

b. Connect the following points with straight lines: $(\underline{a}_{MF}, 0)$, $(\bar{a}_{MF}, 0)$, (p, μ_p), $(\underline{b}_{MF}, 0)$ and $(\bar{b}_{MF}, 0)$.

The result is a triangle LMF.

I.2 Mathematical model for a left-shoulder FOU: Fig. 5.10 depicts this situation. To approximate $UMF(\tilde{A})$, connect the following points with straight lines: $(0,1)$, $(\bar{a}_{MF}, 1)$ and $(\bar{b}_{MF}, 0)$. The result is a left-shoulder UMF. To approximate $LMF(\tilde{A})$, connect the following points with straight lines: $(0,1)$, $(\underline{a}_{MF}, 1)$, $(\underline{b}_{MF}, 0)$ and $(\bar{b}_{MF}, 0)$. The result is a left-shoulder LMF.

I.3 Mathematical model for a right-shoulder FOU: Fig. 5.11 depicts this situation. To approximate $UMF(\tilde{A})$, connect the following points with straight lines: $(\underline{a}_{MF}, 0)$,

$(\underline{b}_{MF},1)$ and $(M,1)$. The result is a right-shoulder UMF. To approximate $LMF(\tilde{A})$, connect the following points with straight lines: $(\underline{a}_{MF},0)$, $(\overline{a}_{MF},0)$, $(\overline{b}_{MF},1)$ and $(M,1)$. The result is a right-shoulder LMF.

5.2.3 Observations

A. Canonical FOUs for Words

Figs. 5.9–5.11 are the only FOUs that can be obtained for a word using the IA, and so these FOUs are referred to herein as *canonical FOUs for a word.*

A word that is modeled by an interior FOU has an UMF that is a trapezoid and a LMF that is a triangle, but in general neither the trapezoid nor the triangle is symmetrical. A word that is modeled as a left- or right-shoulder FOU has trapezoidal upper and lower MFs; however, the legs of the respective two trapezoids are not necessarily parallel.

That there are only three canonical FOUs for a word is very different than in function approximation applications of IT2 FSs (e.g., as in fuzzy logic control, or forecasting of time-series) where one is free to choose the shapes of the FOUs ahead of time, and many different choices are possible.

B. No Completely Filled-in FOUs

In [32], it is explained that when $LMF(\tilde{A}) = 0$ then $FOU(\tilde{A})$ is completely filled-in. This is not considered to be a good FOU, because the centroid of such an FOU equals the span of $LMF(\tilde{A})$, and is therefore completely independent of $UMF(\tilde{A})$. The following theorem shows that the IA does not lead to completely filled-in FOUs.

Theorem 5.1: Using the IA, none of the obtained FOUs will be completely filled-in, i.e., (a) for an interior FOU, $\underline{b}_{MF} > \overline{a}_{MF}$ (see Fig. 5.9); (b) for a left-shoulder FOU, $\underline{b}_{MF} \geq \underline{a}_{MF} > 0$ (see Fig. 5.10); and, (c) for a right-shoulder FOU, $\overline{a}_{MF} < \overline{b}_{MF} < M$ (see Fig. 5.11).

A proof of this theorem is given in Appendix C.

C. Whose FOU?

In the field of probability elicitation, O'Hagan and Oakley [34] question how various individual expert probability distributions should be combined into a single distribution, and, regardless of the method used for combining, *whose distribution does this represent*? The latter question reflects the fact that regardless of how the distributions are combined, the final distribution has lost the uncertainties of the individual subjects, and, in fact, it may correspond to none of the subjects.

One can raise a similar question for the FOU of a word that is obtained from the IA, i.e. *whose FOU does it represent*? Unlike the probability elicitation field, where each expert is assumed to have a probability distribution[9], no assumption is ever made in our work that a subject has a personal FOU for a word. An FOU is a mathematical model

[9] While probability is an excellent model for unpredictable uncertainty, it is mathematics and not science; hence, to ascribe a personal probability distribution to an expert for a particular situation is fiction, regardless of how much data are elicited from that expert. A probability model may fit that data, and the fit may improve as more reliable data are extracted from the expert, but, that in no way proves that this is the expert's personal probability distribution.

that captures the uncertainties about a word, and is only used in later processing or analyses.

Note, however, that the union method for combining each subject's T1 FS preserves all of their uncertainties because each of their T1 FSs is contained within the FOU.

One can even provide a Bayesian-like interpretation to the construction of an FOU. Allow the analyst to be one of the subjects, so that her T1 FS is the first T1 FS of the remaining m^* T1 FSs. Then the resulting FOU not only uses her a priori FOU, which is a T1 FS, but modifies it by folding in the T1 FSs of the remaining m^*-1 data intervals.

5. 3 Codebook examples

A data set was collected from 28 subjects at the Jet Propulsion Laboratory (JPL)[10] for a vocabulary of 32 words. These words were randomized and for all words each subject was asked the question "On a scale of 0-10, what are the end-points of an interval that you associate with the word *W*?". All of the data was processed as described in Sections 5.2.

Table 5.4 (in which the 32 words have been ordered using a ranking method that is explained later in this section) summarizes how many data intervals survived each of the four pre-processing stages, and how many intervals, m^*, survived the Deletion of Inadmissible T1 FSs step in the FS Part of the IA. Observe that m^* is quite variable. Table 5.4 also gives the final left- and right end-point statistics that were used to

[10] This was done in 2002 when J. M. Mendel gave an in-house short course on fuzzy sets and systems at JPL.

Table 5.4: Remaining data intervals and their end-point statistics for m data intervals.

Word	Pre-processing				FS Part	Left-end statistic		Right-end statistic	
	Stage 1 n'	Stage 2 m'	Stage 3 m''	Stage 4 m	m^*	m_l	s_l	m_r	s_r
Teeny-weeny	28	27	25	12	12	0.0025	0.0043	0.6252	0.4478
A smidgen	26	22	20	9	9	0.0339	0.0662	0.9795	0.5869
None to very little	28	23	23	23	23	0	0	1.1371	0.3749
Tiny	28	25	25	12	12	0.0764	0.0825	1.1723	0.7466
Very little	28	23	21	13	8	0.2686	0.2205	1.2897	0.4294
Very small	28	21	21	8	6	0.3997	0.1749	1.3976	0.3137
A bit	28	25	23	12	12	1.0000	0	2.6254	0.4583
Low amount	28	26	25	17	14	0.7963	0.4278	2.6141	0.6348
Little	28	27	26	18	16	0.9612	0.2869	2.7625	0.5700
Small	28	26	25	22	20	1.2482	0.5230	3.2330	0.4653
Somewhat small	28	27	26	16	16	1.8493	0.5313	3.5449	0.5557
Some	28	26	25	17	17	2.7072	0.5834	5.2590	1.0153
Some to moderate	28	26	26	26	26	3.1236	0.6984	6.1169	0.6172
Moderate amount	28	24	23	21	21	3.8838	0.3352	6.0189	0.5770
Fair amount	28	24	22	22	22	3.9952	0.6697	6.2978	0.4932
Medium	28	24	23	20	20	4.3529	0.3853	5.8343	0.3048
Modest amount	28	21	20	19	19	4.3125	0.4369	6.0238	0.3113
Good amount	28	27	26	15	14	6.0692	0.9252	8.2282	0.4752
Sizeable	28	24	23	17	13	6.3915	0.7619	8.7298	0.7640
Quite a bit	28	27	26	14	12	6.0516	0.4770	8.3883	0.7110
Considerable amount	28	26	25	17	14	6.5049	0.7301	8.6187	0.7386
A lot	28	26	26	16	12	7.2669	0.5392	9.2706	0.4733
High amount	28	26	24	22	13	7.6442	0.7686	9.4406	0.4604
Substantial amount	28	26	25	15	10	7.6036	0.5896	9.3982	0.4499
Very sizeable	28	26	24	14	10	7.5715	0.7045	9.3577	0.4405
Large	28	17	16	16	16	7.3890	0.5366	9.0000	0
Very large	27	26	25	15	15	8.7070	0.3906	9.8906	0.1783
Humongous amount	28	23	23	22	22	9.0116	0.5658	10.0000	0
Huge amount	28	27	23	17	17	8.8236	0.3873	9.9573	0.0882
Very high amount	28	22	19	16	16	8.6886	0.5732	10.0000	0
Extreme amount	28	17	17	17	17	9.0000	0	10.0000	0
Maximum amount	27	25	23	15	15	9.2155	0.2601	10.0000	0

establish the nature of each word's FOU. These statistics are based on the m remaining data intervals after stage 4 of pre-processing.

Table 5.5 is the *codebook* for the 32 words. It provides the coordinates (code) for the LMF and UMF of each FOU. We warn the reader that *while it may be very tempting*

to use this codebook for your application(s), do not do this, because data collection is sensitive to scale and is application dependent; hence, although we advocate using this paper's methodologies we do not advocate using the paper's example word FOUs for specific applications.

Table 5.5: FOU data for all words (based on m^* data intervals)—the Codebook.

Word	LMF[a]	UMF[b]	Centroid	Mean of centroid
Teeny-weeny	(0.01, 0.13)	(0.14, 1.97)	[0.04, 1.06]	0.550
A smidgen	(0.05, 0.63)	(0.26, 2.63)	[0.20, 1.05]	0.625
None to very little	(0.03, 0.39)	(0.18, 2.63)	[0.13, 1.17]	0.650
Tiny	(0.05, 0.63)	(0.37, 3.92)	[0.21, 1.72]	0.965
Very little	(0.40, 0.75, 0.41, 0.92)	(0.03, 0.50, 1.25, 2.31)	[0.45, 1.49]	0.970
Very small	(0.40, 0.90, 0.59, 1.10)	(0.19, 0.75, 1.25, 2.31)	[0.66, 1.39]	1.025
A bit	(0.79, 1.68, 0.74, 2.21)	(0.59, 1.50, 2.00, 3.41)	[1.42, 2.09]	1.755
Low amount	(1.29, 1.80, 0.42, 2.21)	(0.09, 1.25, 2.50, 4.62)	[1.07, 3.13]	2.100
Little	(1.09, 1.83, 0.53, 2.21)	(0.38, 1.50, 2.50, 4.62)	[1.31, 2.95]	2.130
Small	(1.79, 2.28, 0.40, 2.81)	(0.09, 1.50, 3.00, 4.62)	[1.29, 3.34]	2.315
Somewhat small	(2.29, 2.75, 0.38, 3.21)	(0.59, 2.00, 3.50, 5.62)	[1.73, 4.16]	2.945
Some	(2.88, 3.61, 0.35, 4.21)	(0.38, 2.50, 5.00, 8.00)	[2.03, 5.90]	3.965
Some to moderate	(4.10, 4.60, 0.36, 5.20)	(0.96, 3.50, 5.50, 8.00)	[2.76, 6.31]	4.535
Moderate amount	(4.30, 4.75, 0.38, 5.20)	(2.17, 4.00, 5.00, 7.83)	[3.50, 6.28]	4.895
Fair amount	(4.79, 5.29, 0.41, 6.02)	(1.46, 4.25, 6.00, 8.50)	[3.47, 6.78]	5.125
Medium	(4.86, 5.00, 0.27, 5.14)	(3.59, 4.75, 5.50, 6.90)	[4.19, 6.19]	5.190
Modest amount	(4.79, 5.30, 0.42, 5.70)	(3.59, 4.75, 6.00, 7.41)	[4.57, 6.24]	5.405
Good amount	(6.79, 7.20, 0.34, 7.80)	(3.17, 6.00, 8.00, 9.62)	[5.04, 8.39]	6.715
Sizeable	(6.79, 7.38, 0.49, 8.21)	(4.38, 6.50, 8.00, 9.40)	[6.16, 8.15]	7.155
Quite a bit	(6.79, 7.38, 0.49, 8.21)	(4.38, 6.50, 8.00, 9.40)	[6.16, 8.15]	7.155
Considerable amount	(7.19, 7.58, 0.37, 8.21)	(4.38, 6.50, 8.25, 9.60)	[5.97, 8.52]	7.245
A lot	(7.69, 8.19, 0.47, 8.81)	(5.38, 7.50, 8.75, 9.83)	[6.99, 8.82]	7.905
High amount	(7.79, 8.30, 0.53, 9.20)	(5.38, 7.50, 8.75, 9.81)	[7.19, 8.82]	8.005
Substantial amount	(7.79, 8.30, 0.53, 9.20)	(5.38, 7.50, 8.75, 9.81)	[7.19, 8.82]	8.005
Very sizeable	(8.29, 8.56, 0.38, 9.20)	(5.38, 7.50, 9.00, 9.81)	[6.94, 9.10]	8.020
Large	(8.00, 8.36, 0.57, 9.17)	(5.98, 7.75, 8.60, 9.52)	[7.50, 8.75]	8.125
Very large	(8.70, 9.90)	(7.37, 9.40)	[9.03, 9.58]	9.305
Humongous amount	(9.74, 9.98)	(7.37, 9.80)	[8.71, 9.91]	9.310
Huge amount	(9.47, 9.96)	(7.37, 9.59)	[8.88, 9.83]	9.355
Very high amount	(9.34, 9.95)	(7.37, 9.82)	[8.97, 9.78]	9.375
Extreme amount	(8.68, 9.90)	(8.68, 9.90)	[9.56, 9.56]	9.560
Maximum amount	(9.60, 9.97)	(8.68, 9.)	[9.51, 9.87]	9.690

[a] For an interior LMF, the four numbers are $(\bar{a}_{MF}, p, \mu_p, \underline{b}_{MF})$; for a left-shoulder LMF, the two numbers are $(\underline{a}_{MF}, \underline{b}_{MF})$; and for right-shoulder LMF, the two numbers are $(\bar{a}_{MF}, \bar{b}_{MF})$.

[b] For an interior UMF, the four numbers are $\underline{a}_{MF}, \underline{c}_{MF}, \bar{c}_{MF}, \bar{b}_{MF})$; for a left-shoulder UMF, the two numbers are $(\bar{a}_{MF}, \bar{b}_{MF})$; and for right-shoulder UMF, the two numbers are $(\underline{a}_{MF}, \underline{b}_{MF})$.

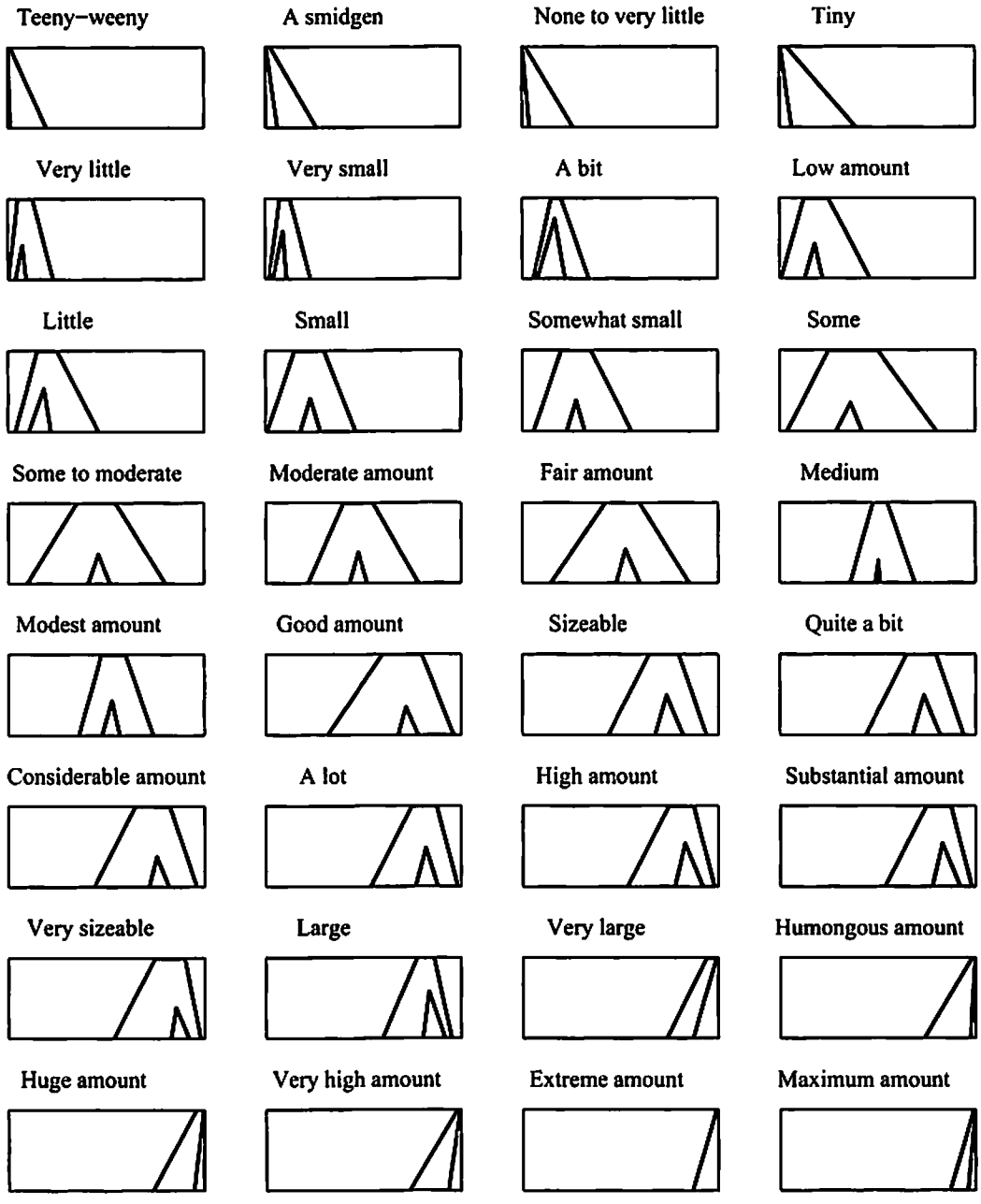


Figure 5.12: FOUs for all 32 words. Start at the top row and proceed downwards scanning from left-to-right.

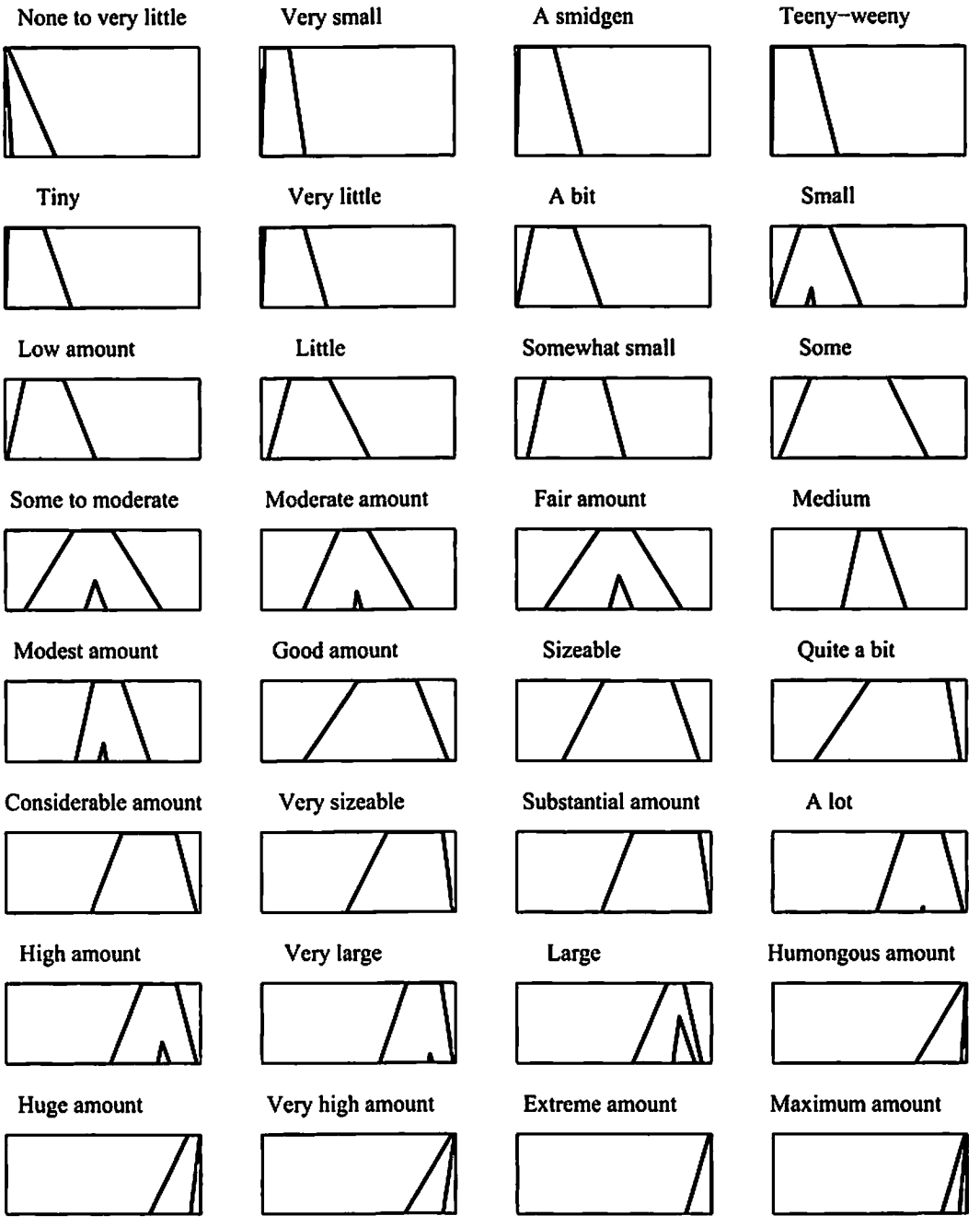


Figure 5.13: FOUs for all 32 words when Stage 4 of data pre-processing is omitted. Start at the top row and proceed downwards scanning from left-to-right.

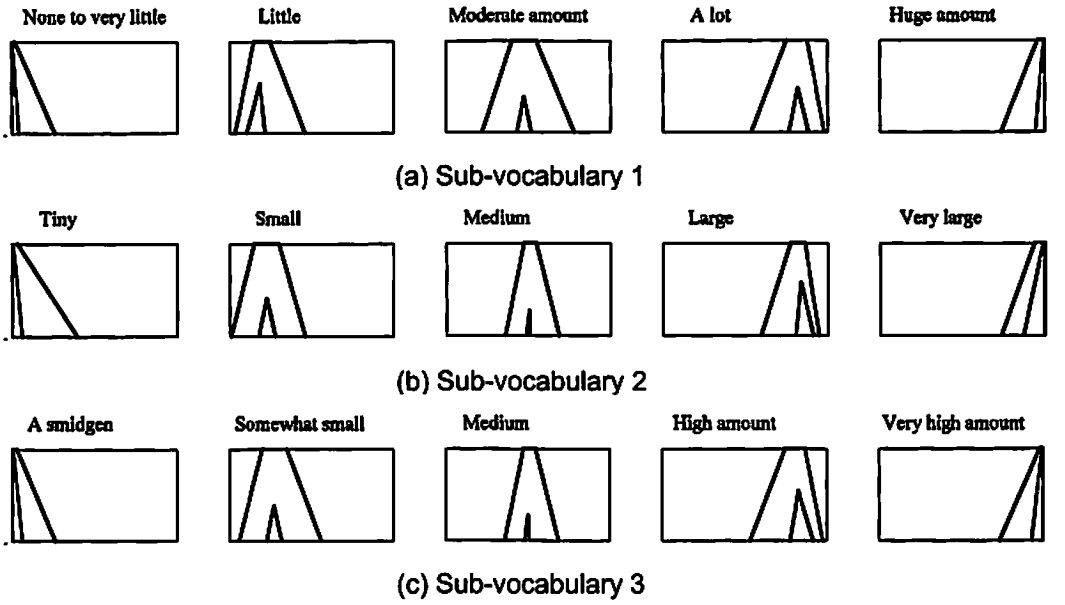


Figure 5.14: FOUs for three sub-vocabularies.

Fig. 5.12 depicts the FOUs for all 32 words. Observe that the words have been ordered so that there seems to be a very natural flow from left-shoulder FOUs to interior FOUs to right-shoulder FOUs. This flow was achieved by first computing the centroid of each FOU [20], [10], [32] and then the mean of each centroid. The results of these computations are given in the last two columns of Table 5.5. The words were then rank-ordered using the mean of the centroid. Ranking the words by this method seems to give visually acceptable results.

Interestingly, *sizeable* and *quite a bit* have the same FOUs, as do *high amount* and *substantial amount*. This can be seen in Fig. 5.12, and is also evident from the LMF and UMF columns in Table 5.5. It means that based on the data collected from the 28 JPL subjects, the words *sizeable* and *quite a bit* are 100% similar, and the words *high amount and substantial amount* are also 100% similar. Similarity of all of the words to

one another can be computed using the similarity measure for IT2 FSs that is described in [47]. Observe, also, that the word *extreme amount* is modeled as a T1 FS, demonstrating that the IA does provide a T1 FS model when all m^* data intervals are the same.

Section 5.4 argued for the inclusion of a fourth step of pre-processing, namely for reasonable-interval processing. Fig. 5.13 depicts FOUs for the 32 words when this stage was omitted. Observe that many of the FOUs are filled in or just about filled in, whereas none of the FOUs depicted in Fig. 5.12 are filled in, thereby confirming the results given in Theorem 5.1.

Herrera, Herrera-Viedma and Martinez [7] discuss *multi-granular linguistic term sets* and how in "... decision making problems with multiple sources of information, linguistic performance values that are given to the different sources can be represented as linguistic terms sets with different granularity and/ or semantics." Our interpretation of "linguistic term sets with different granularity" is as a *sub-vocabulary* from the codebook. Fig. 5.14 depicts three sub-vocabularies, where the FOUs in each sub-vocabulary cover the entire domain [0, 10]. Each sub-vocabulary was obtained from the results given in Table V and Fig. 5.12. When a codebook is established it contains within it many sub-vocabularies. One important use for a sub-vocabulary is in designing IF-THEN rules as a CWW Engine, where it is expedient to use a small (the smallest) sub-vocabulary that covers the entire domain in order to avoid rule-explosion.

5. 4 Conclusions

This chapter has presented a very practical type-2-fuzzistics methodology for obtaining IT2 FS models for words, one that is called an *Interval Approach* (IA). The basic idea of the IA is to collect interval end-point data for a word from a group of subjects, map each subject's data interval into a pre-specified T1 person MF interpret the latter as an embedded T1 FS of an IT2 FS, and obtain a mathematical model for the FOU for the word from these T1 FSs.

The IA consists of two parts, the *Data Part* and the *Fuzzy Set Part*. In the Data Part, the interval end-point data are pre-processed, after which data statistics are computed for the surviving data intervals. In the FS Part, the data are used to decide whether the word should be modeled as an interior, left-shoulder or right-shoulder FOU. Then the parameters of the respective embedded T1 MFs are determined using the data statistics and uncertainty measures for the T1 FS models. The derived T1 MFs are aggregated using union leading to an FOU for a word, and finally to a mathematical model for the FOU.

The IA has many strong points and, arguably, no weak points, i.e., it: (1) Collects interval end-point data from subjects, and this is easy to do; (2) Does not require subjects to be knowledgeable about fuzzy sets; (3) Has a straightforward mapping from data to an FOU; (4) Does not require an priori assumption about whether or not an FOU is symmetric or non-symmetric; and, (5) Leads to an IT2 FS word model that reduces to a T1 FS word model automatically if all subjects provide the same intervals.

Chapter 6

Conclusions and future works

6.1 Conclusions

This thesis addresses two problems about water-flood management: (1) infer reservoir heterogeneity using measured injection and production rates, and (2) construct a decision support system to optimize oil production using computing with words (CWW) and the inferred reservoir heterogeneity.

To infer reservoir heterogeneity, we first present an adaptive method using an Extended Kalman Filter (EKF) for the case of multiple injectors and a single producer, and then present a pseudo-virtual reservoir method for the case of multiple injectors and multiple producers, respectively.

In the EKF method, a very simple parametric model, one with two parameters per injector, is used so that if a producer depends upon N injectors our model contains exactly $2N$ parameters. The EKF is used to adaptively estimate the $2N$ parameters, from which the injector-producer relationship (IPR) between each injector and a producer is then estimated. Our approach has been tested on synthetic data, reservoir simulation data and real data. Test results on synthetic data and reservoir simulation data demonstrate the feasibility of the EKF method, and test results on the real data match expert knowledge about the IPRs between injectors and a producer. All results confirm

that this EKF method can provide a good way to infer and track the IPRs so as to provide better insight about the IPRs.

In the pseudo-virtual reservoir method, a virtual reservoir model is used to model the reservoir, and a pseudo- virtual reservoir model is used to estimate the regional impact from each injector, so that the problem for this case reduces to the problem for the case of multiple injectors and a single producer. Our approach has been tested on synthetic data, and test results suggest that this pseudo-virtual reservoir method is feasible for evaluating regional impacts with multiple injectors and multiple producers.

The study for decision support system using CWW focuses on the encoder component (called *type-2 fuzzistics*) of a *Perceptual Computer* (Per-C), a specific architecture for CWW using interval type-2 fuzzy sets. Our work so far has been to transform linguistic perceptions, words, into interval type-2 fuzzy sets (IT2 FS) that activate a CWW engine. A new and simple approach, called the *Interval Approach* (*IA*) to type-2 fuzzistics has been developed; it captures the strong points of both the person membership function and interval end-points approaches. It collects interval end-point data from subjects, does not require subjects to be knowledgeable about fuzzy sets, has a straightforward mapping from data to footprint of uncertainty (FOU), does not require an a priori assumption about whether or not a FOU is symmetric or non-symmetric, and leads to an IT2 FS word model that reduces to a T1 FS word model automatically if all subjects provide the same intervals. Experiments showed that this approach is easy to implement and the derived interval type-2 word models match our intuitions, i.e., the FOUs of the small-sounding words are located to the left, the FOUs of the medium-

sounding words are located in the middle, and the FOUs of the large-sounding words are located to the right.

6.2 Future Works

Water-flood management aims to increase secondary oil recovery. Generally, the system to fulfill such a goal operates in two stages. The first stage derives some information to characterize the reservoir using some available data, and the second stage makes decisions based on that information. This thesis adheres to this two-stage approach. Inferring reservoir heterogeneity is where we obtain the information about the reservoir heterogeneity, and constructing a decision support system is where we make better decisions based on the inferred reservoir heterogeneity and expert knowledge.

To construct a sound water-flood management system, further research is needed as follows. For inferring reservoir heterogeneity, we need to perform field tests to verify our obtained test results. For constructing a decision support system using computing with words, we need to construct a decision support system using computing with words for different applications such as pattern review process and water reallocation, so that we can use the inferred reservoir heterogeneity to help us make better decisions. Some details about this proposed research are given next.

6.2.1 Perform field tests

We are in the process of performing field tests on Chevron Lost Hills field data, and are ready to improve our current approach based on the feedback from the tests.

Based on our current research, we have found that a very important issue for current EKF and pseudo-virtual reservoir approach is validation on test results for real data. The validation for synthetic data is obvious because we know what "truth" is; however, for real data, we can't evaluate our results by comparing estimated IPRs or regional impact with the actual IPRs or regional impact because we don't know what the actual IPRs or regional impact are; hence, we need to use other ways to deal with this issue.

Currently, our validation is to see if our results can match some expert knowledge; however, there are some limitations for such a validation procedure, namely:

(1) Only a limited amount of expert knowledge is available, e.g., for Lost Hills, we only have the following two rules of expert knowledge: (a) Lost Hills is an oilfield with low permeability such that the injectors far away from the producer should impact the producer very little; and (b) there is a natural fracture around North-East 55 degree; and,

(2) A reservoir is a very complex system, such that expert knowledge may only be useful for a limited number of cases, and we have no idea about when our expert knowledge is or is not useful. Hence, even with expert knowledge, we still can't obtain convincible validation on test results because we can't tell if the validation is correct.

Consequently, we need to design a new approach to deal with this.

One approach is to forecast the production rates of producers using our EKF after a deliberate change has been made to an injector. By comparing the forecasted production rates with the actual measured production rates, we may gain more confidence on our

test results and be more confident of the validation. Two issues need to be examined:

1) Develop a strategy to change the injection rates. From the view of Chevron engineers, changing the injection rate may change the reservoir model, which directly relates to our estimated IPRs or regional impacts; hence, how to change the injection rate such that the reservoir model will not change much is an important issue that needs to be examined.
2) Develop a procedure for validation. Because the reservoir is a dynamic system, even if we don't intentionally change anything, the reservoir model may change constantly and naturally; hence, we also need to develop a procedure for validation such that it can correctly describe our test results, e.g., we may validate the test results in a small amount of time so that we can assume that the reservoir model does not change too much in that small amount of time.

6.2.2 Construct a Decision Support System Using Computing With Words

After inferring the reservoir heterogeneity, the next question we would like to address is what we can do for inferred reservoir heterogeneity. Pattern review process could be one such application that uses the inferred reservoir heterogeneity to make better decisions.

As an important part of water-flood management, the pattern review process often leads to the determination of remediation steps to optimize recovery and mitigate subsidence. Currently, the pattern review is performed by human experts using geology information (such as location, fault information, etc.) and production rates, etc. However, there are some limitations to this:

1) Because of the enormous amount of available data and the very complex reservoir environment, this process is very difficult and somewhat inefficient;
2) Neat patterns defined by geology information can't actually determine the correct relationships between injectors and producers, i.e., petroleum engineers know that the injectors that relate to a particular producer are frequently not always within the pattern of this producer. Sometimes, they are two or three patterns away. Sometimes, they are even farther away, bypassing some close producers and going to a more distant producer. This is more prevalent in highly fractured reservoirs.

Consequently, our long-term goal is to provide a decision support tool to supplement Chevron engineers carrying out asset activities and processes, such that a pattern can be identified automatically. The pattern would be identified based not only on geological information and other Chevron expert knowledge, but also on inferred reservoir heterogeneity, where the latter would be obtained using our newly developed EKF and pseudo-virtual reservoir approaches.

Future work should be conducted to develop a decision support CWW system for the pattern review process, because the CWW using interval type-2 FLS is very suitable when:

- Knowledge is mined from experts using IF-THEN questionnaires, but there is uncertainty about the meanings of the questionnaire words, because words mean different things to different people; and,
- Computation is done with words (words-in and words-out).

Some issues that should be studied are:

- Design a way to extract some expert knowledge about the pattern review process from Chevron Engineers. Usually, the expert knowledge is captured in the form of rules, which are mined from experts using IF-THEN questionnaires. An example of a rule is:

R^l : IF the *injector - producer relation* is *low*, and the *fracture* between producer and injector is *highly impossible*,
THEN this injector and producer are *very unlikely to be* within the same pattern.

(6.1)

In (6.1), rule antecedents (the IF parts of a rule) injector-producer relation and fracture involve the forecasted injector-producer relationships using our newly developed Extended Kalman Filtering approach, and the geology information about the fracture; and each rule consequent (the THEN part of a rule), e.g., *very unlikely*, is a term to describe the pattern information. Each expert may provide some rules, and then for a group of experts, a group of rules can be obtained to support decision making.

- Develop and implement a strategy to construct the CWW decision support system. When we construct the CWW decision support system, we need to decide if the system is distributed and hierarchical.
- Test this decision support system using real data and an interaction with Chevron engineers. This is very important for validating this decision system.

6.2.3 Some future works about interval approach for type-2 fuzzistics

Recall from Chapter 5 that, so far, it has been assumed that interval end-point data have been collected from a group of subjects. Although we strongly advocate doing this, so that the data will contain both intra- and inter-levels of uncertainty, we realize that there may be times when this is not possible, due, for example, to budget limitations, time-constraints, or unavailability of a subject pool. How to obtain an FOU from a single subject by an IA is an open research issue.

Finally, the following *validation problem* needs to be studied. Given an FOU whose parameters are specified, use it to generate N symmetrical triangle embedded T1 FSs. Such T1 FSs might be called the "realizations" of the FOU T1 FS generator. Using the equations that are given in Table 5.3, but in reverse, obtain $(a^{(i)},b^{(i)})$ for each of the N $(a_{MF}^{(i)},b_{MF}^{(i)})$. Prove that by using the IA, one can obtain the original FOU to within a quantified level of accuracy (that will depend upon N), i.e., prove some sort of convergence result for the IA. Because this validation problem is very different from the problem that is considered in Chapter 5, it will appear as the future work for type-2 fuzzistics.

APPENDIX A

THE MODEL OF INJECTION RATES

To test our proposed approach, we need to design the reasonable model, which matches practical situations, for injection rate data generation. Our understanding of the practical case is follows: when the engineer injects water at time t, he firstly has the scheduled set point of injection rate, called scheduled injection rate, which can be the same as or different from that at different time t'. However, because of all kind of issues, the actual injection rate has some deviation from the scheduled injection rate, e.g., 30%, 50% deviation. Therefore, the actual injection rate is different with the scheduled injection rate. Additionally, the environment may introduce some noises into the measurement of injection rate, such that the measured injection rate composes of the actual injection rate and measurement noises. Based on this understanding, the simulation model of injection rate generation for the i^{th} injector is:

$$I_i(n) = I_s\left(1 + V_i(n)\right) + N_i(n) \tag{A-1}$$

where $I_i(n)$, I_s are the measured injection rate and scheduled injection rate, respectively, $V_i(n)$ is the zero-mean random deviation about the set point, $I_s\left(1 + V_i(n)\right)$ is the actual injection rate, and $N_i(n)$ is the measurement noise of injection rate.

As some experts said, usually, the I_s is fixed, but there are infrequent changes; and the $N_i(n)$ is about 5%-7% of injection rate. Normally, we assume that the measurement

noise $N_i(n)$ is zero-mean Gaussian distributed, and independent on the actual injection rate $I_s\left(1+V_i(n)\right)$. Also, we assume that the $V_i(n)$ is zero-mean uniform distributed random variable in $[-a,\ a]$. Based on these assumptions, we can obtain the mean and variance of $I_i(n)$, $V_i(n)$ and $N_i(n)$ as:

$$\begin{cases} m_I = I_s \\ m_V = m_N = 0 \\ \delta_V^2 = \dfrac{1}{3}a^2 \\ \delta_I^2 = \dfrac{1}{3}I_s^2 a^2 + \delta_N^2 \end{cases} \tag{A-2}$$

where m_I, m_V, m_N are the mean of $I_i(n)$, $V_i(n)$ and $N_i(n)$, respectively, and δ_I^2, δ_V^2 and δ_N^2 are the variance of $I_i(n)$, $V_i(n)$ and $N_i(n)$, respectively.

The derivation of (A-2) is follows. As we have assumed, the $N_i(n)$ is zero-mean Gaussian distributed, and independent on the actual injection rate $I_s\left(1+V_i(n)\right)$, and the $V_i(n)$ is zero-mean uniform distributed random variable in $[-a,\ a]$, it follows that

$$m_V = m_N = 0 \tag{A-3}$$

Hence,

$$m_I = E\left\{I_s\left(1+V_i(n)\right)+N_i(n)\right\} = I_s \tag{A-4}$$

Let δ_N^2 be the variance of $N_i(n)$, and the variance of $V_i(n)$, which is uniformly distributed in $[-a,\ a]$, can be computed by:

$$\delta_V^2 = \frac{1}{2a}\int_{-a}^{a}(x-0)^2 dx = \frac{1}{3}a^2 \tag{A-5}$$

Because $N_i(n)$ is independent on the actual injection rate $I_s\left(1+V_i(n)\right)$, the variance of $I_i(n)$ is

$$\begin{aligned}\delta_I^2 &= \mathrm{var}\left\{I_s\left(1+V_i(n)\right)+N_i(n)\right\} = \mathrm{var}\left\{I_s\left(1+V_i(n)\right)\right\}+\mathrm{var}\left\{N_i(n)\right\} \\ &= I_s^2\delta_V^2+\delta_N^2 = \frac{1}{3}I_s^2a^2+\delta_N^2\end{aligned} \tag{A-6}$$

which completes the derivation of (A-2).

To generate the simulation data, we need to make choices on $I_i(n)$, $V_i(n)$ and $N_i(n)$. To make the choice more practically, we refer our choices to the mean and variance of a collection of injection rates from 10 injectors, which are summarized in Table A-1.

As we had discussed, let's assume $\delta_N^2 \approx 0.05\delta_I^2$, then $\frac{1}{3}I_s^2 a^2 \approx 0.95\delta_I^2$. Hence,

$$\begin{cases} a^2 \approx 2.85\left(\delta_I / I_s\right)^2 \\ a \approx 1.6882\left(\delta_I / I_s\right) \end{cases} \tag{A-7}$$

Table A-1: the mean and variance of the injection rates from 10 injectors.

Name of injector	Mean I_s	Variance δ_I^2	δ_I / I_s
I5 12- 1B	2136	508369	713/2136 = 0.3338
I5 11- 1W	2465	269361	519/2465 = 0.2105
I5 11- 1WA	2707	527076	727/2707 = 0.2686
I5 10- 1W	2392	373321	611/2392 = 0.2554
I5 10- 2WAS	1814	148225	385/1814 = 0.2122
I5 11-2WB (Before Aug. 2002)	2658	318096	564/2658 = 0.2122
I5 10- 2WB	1577	91204	302/1577 = 0.1915
I5 11- 2W	2341	320356	566/2341 = 0.2418
I5 11- 2WAL	2332	275625	525/2332 = 0.2251
I5 9- 2W	2522	625681	791/2522 = 0.3136

From Table A-1, we can see that the δ_I / I_s is about 0.19 – 0.33. Therefore,

$$a \approx 1.6882\left(\delta_I / I_s\right) \approx 0.3208 \sim 0.5571 \tag{A-8}$$

which means the deviation of the scheduled injection rate is about 32% to 56%.

In summary, my choices on these variables are:

(1) $I_s = 2000$.

(2) $a = 0.4$ such that $V_i(n)$ is the uniform random variable in domain $[-0.4,\ 0.4]$.

(3) $\delta_N = \sqrt{\frac{\frac{1}{3} I_s^2 a^2}{95\%} \times 5\%} = \sqrt{\frac{I_s^2 a^2}{3 \times 19}} = \sqrt{\frac{(2000 \times 0.4)^2}{3 \times 19}} = 106$.

APPENDIX B

DERIVATION OF REASONABLE INTERVAL TEST FOR (5.4)-(5.6)

In this Appendix, derivations of (5.4)-(5.6) are obtained. Examining Fig. 5.4, and using the requirement that reasonable data intervals must overlap, it must be true that:

$$\min_{\forall i=1,\ldots,m''} b^{(i)} > \max_{\forall i=1,\ldots,m''} a^{(i)} \tag{B-1}$$

A simple way to satisfy (B-1) is to require that:

$$\left.\begin{matrix} a^{(i)} < \xi \\ b^{(i)} > \xi \end{matrix}\right\}, \ \forall i = 1,\ldots,m'' \tag{B-2}$$

where threshold ξ has to be chosen, and there can be different ways to do this. In this paper, an optimal value of ξ, ξ^*, is chosen so that

$$\xi^* = \arg\min_{\xi}[P(a^{(i)} > \xi) + P(b^{(i)} < \xi)] \tag{B-3}$$

By choosing ξ^* in this way, data intervals that don't satisfy (B-2) will occur with the smallest probability[11].

[11] Another approach might be to choose ξ^* that maximizes $P[a^{(i)} < \xi, b^{(i)} > \xi]$. As of yet, we have not done this.

In order to compute ξ^*, it is assumed that each $a^{(i)}$ $(i = 1,...,m'')$ is Gaussian with mean m_a and standard deviation σ_a, and each $b^{(i)}$ $(i = 1,...,m'')$ is also Gaussian, but with mean m_b and standard deviation σ_b. It follows that:

$$P(a^{(i)} > \xi) + P(b^{(i)} < \xi) = \frac{1}{\sqrt{2\pi}\sigma_a} \int_{\xi}^{\infty} e^{-\frac{1}{2}\left[\frac{a^{(i)} - m_a}{\sigma_a}\right]^2} da^{(i)} + \frac{1}{\sqrt{2\pi}\sigma_b} \int_{-\infty}^{\xi} e^{-\frac{1}{2}\left[\frac{b^{(i)} - m_b}{\sigma_b}\right]^2} db^{(i)} \quad \text{(B-4)}$$

Setting the derivative of this function with respect to ξ equal to zero, ξ^* is found to be the solution of:

$$\frac{1}{\sqrt{2\pi}\sigma_a} e^{-\frac{1}{2}\left[\frac{\xi^* - m_a}{\sigma_a}\right]^2} = \frac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{1}{2}\left[\frac{\xi^* - m_b}{\sigma_b}\right]^2} \quad \text{(B-5)}$$

Observe that ξ^* occurs at the intersection of the two Gaussian distributions $p(a^{(i)})$ and $p(b^{(i)})$. Taking the natural logarithm of both sides of (B-5), one is led to the following quadratic equation:

$$(\sigma_a^2 - \sigma_b^2)\xi^{*2} + 2(m_a\sigma_b^2 - m_b\sigma_a^2)\xi^* + [m_b^2\sigma_a^2 - m_a^2\sigma_b^2 - 2\sigma_a^2\sigma_b^2 \ln(\sigma_a / \sigma_b)] = 0 \quad \text{(B-6)}$$

The two solutions of this equation are:

$$\xi^* = \frac{(m_b\sigma_a^2 - m_a\sigma_b^2) \pm \sigma_a\sigma_b\left[(m_a - m_b)^2 + 2(\sigma_a^2 - \sigma_b^2)\ln(\sigma_a / \sigma_b)\right]^{1/2}}{(\sigma_a^2 - \sigma_b^2)} \qquad \text{(B-7)}$$

The final solution is chosen as the one for which:

$$\xi^* \in [m_a, m_b] \qquad \text{(B-8)}$$

That this solution minimizes $P(a^{(i)} > \xi) + P(b^{(i)} < \xi)$, rather than maximizes it follows from showing that the derivative of (B-5) with respect to ξ, after which ξ is set equal to ξ^*, is positive. Because this is a very tedious calculation, an alternative is presented next.

Fact: $P(a^{(i)} > \xi) + P(b^{(i)} < \xi)$ is a concave function and its minimum value, ξ^*, occurs in the interval $[m_a, m_b]$.

A proof of this fact follows from (a) and (b) of Fig. B-1. From Fig. B-1 (a), observe that at $\xi = \xi^*$,

$$P(a^{(i)} > \xi^*) + P(b^{(i)} < \xi^*) = A_1 + (A_2 + A_3) \qquad \text{(B-9)}$$

and at $\xi = \xi'$

$$P(a^{(i)} > \xi') + P(b^{(i)} < \xi') = (A_1 + A_2 + A_4) + A_3; \qquad \text{(B-10)}$$

hence,

$$P(a^{(i)} > \xi') + P(b^{(i)} < \xi') > P(a^{(i)} > \xi^*) + P(b^{(i)} < \xi^*) \qquad \text{(B-11)}$$

Proceeding in a similar manner for Fig. B-1 (b), it follows that:

$$P(a^{(i)} > \xi'') + P(b^{(i)} < \xi'') > P(a^{(i)} > \xi^*) + P(b^{(i)} < \xi^*) \qquad \text{(B-12)}$$

(B-11) and (B-12) together prove that $P(a^{(i)} > \xi) + P(b^{(i)} < \xi)$ is a concave function about $\xi = \xi^*$. The concave shape of $P(a^{(i)} > \xi) + P(b^{(i)} < \xi)$ is depicted in Fig. B-1 (c).

Above, it has been proven that ξ^* occurs at the intersection of $p(a^{(i)})$ and $p(b^{(i)})$; but, it is clear from Fig. 5.B-1 (a) or (b) $\xi^* \in [m_a, m_b]$. Q. E. D.

Because access to the population means and standard deviations is unavailable, they must be estimated in order to compute ξ^* and to perform the test in (B-8). Our approach is to estimate those quantities as:

$$\left.\begin{aligned} \hat{m}_a &= m_l \\ \hat{m}_b &= m_r \\ \hat{\sigma}_a &= \sigma_l \\ \hat{\sigma}_b &= \sigma_r \end{aligned}\right\} \qquad \text{(B-13)}$$

Doing this, one obtains (5.4)-(5.6). Note that numerical values for m_l, m_r, σ_l and σ_r are available at the end of Tolerance Limit Processing, so that ξ^* can indeed be computed.

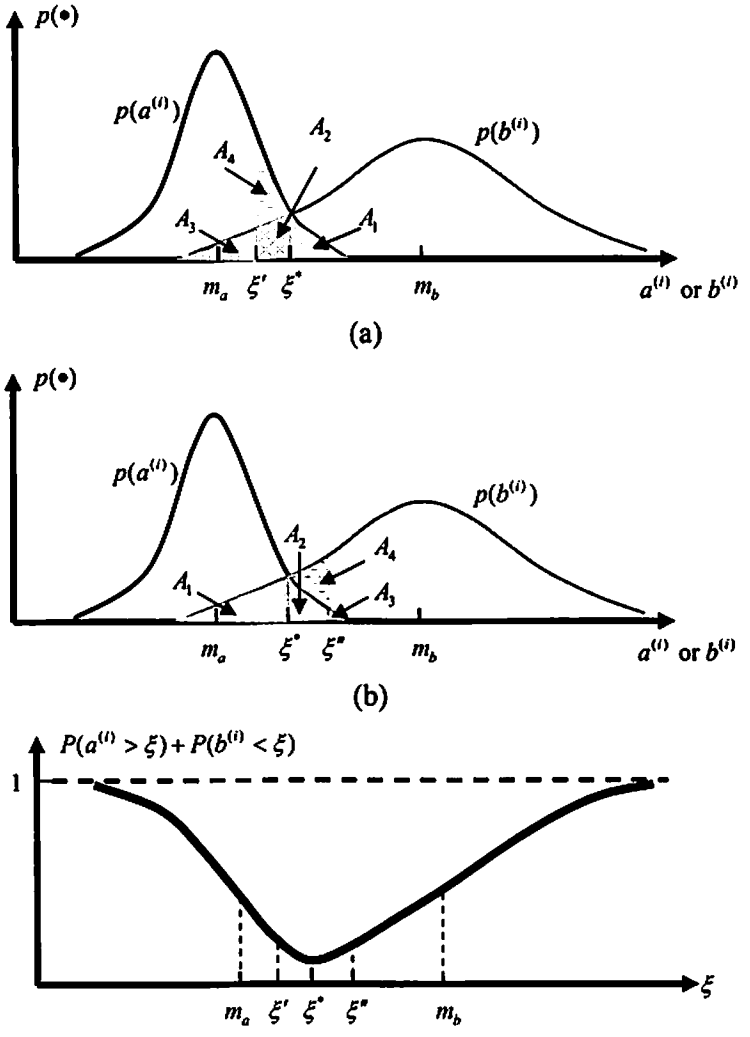


Figure B-1: $p(a^{(i)})$, $p(b^{(i)})$ and the four areas A_i $(i = 1,...,4)$ that can be used to compute $P(a^{(i)} > \xi^*) + P(b^{(i)} < \xi^*)$, and (a) $P(a^{(i)} > \xi') + P(b^{(i)} < \xi')$, or (b) $P(a^{(i)} > \xi'') + P(b^{(i)} < \xi'')$; and, (c) the concave shape of $P(a^{(i)} > \xi) + P(b^{(i)} < \xi)$.

APPENDIX C

PROOF OF THEOREM 5.1

This Appendix contains proofs for the three parts of Theorem 5.1. It uses $\underline{a}_{MF}$, $\overline{a}_{MF}$, $\underline{b}_{MF}$ and $\overline{b}_{MF}$ that are defined in (22) and (23).

(a) Using the equations for $a_{MF}^{(i)}$ and $b_{MF}^{(i)}$, that are given in the top row of Table 5.3, it follows that ($\forall i = 1,...,m^*$)

$$a_{MF}^{(i)} = \tfrac{1}{2}[(a^{(i)} + b^{(i)}) - \sqrt{2}(b^{(i)} - a^{(i)})] = a^{(i)} - \frac{\sqrt{2}-1}{2}(b^{(i)} - a^{(i)}) < a^{(i)} \qquad \text{(C-1)}$$

$$b_{MF}^{(i)} = \tfrac{1}{2}[(a^{(i)} + b^{(i)}) + \sqrt{2}(b^{(i)} - a^{(i)})] = b^{(i)} + \frac{\sqrt{2}-1}{2}(b^{(i)} - a^{(i)}) > b^{(i)} \qquad \text{(C-2)}$$

Let

$$\overline{a} \equiv \max_{i=1,...,m^*} \{a^{(i)}\} \qquad \text{(C-3)}$$

$$\underline{b} \equiv \min_{i=1,...,m^*} \{b^{(i)}\} \qquad \text{(C-4)}$$

Applying the definitions of $\overline{a}_{MF}$ and $\underline{b}_{MF}$ as well as (C-3) and (C-4) to (C-1) and (C-2), it follows that:

$$\bar{a}_{MF} < \bar{a} \tag{C-5}$$

$$\underline{b}_{MF} > \underline{b} \tag{C-6}$$

From (5.4), (C-3) and (C-4), it is true that:

$$\bar{a} < \xi^* < \underline{b} \tag{C-7}$$

Substituting (C-5) and (C-6) into (C-7), it follows that:

$$\bar{a}_{MF} < \xi^* < \underline{b}_{MF} \tag{C-8}$$

Consequently,

$$\underline{b}_{MF} > \bar{a}_{MF} \tag{C-9}$$

(b) Using the equation for $a_{MF}^{(i)}$, given in the second row of Table 5.3, it follows that:

$$a_{MF}^{(i)} = \frac{(a^{(i)} + b^{(i)})}{2} - \frac{(b^{(i)} - a^{(i)})}{\sqrt{6}} = \left(\frac{1}{2} + \frac{1}{\sqrt{6}}\right) a^{(i)} + \left(\frac{1}{2} - \frac{1}{\sqrt{6}}\right) b^{(i)} > 0 \tag{C-10}$$

Applying the definition of $\underline{a}_{MF}$ to (C-10), it follows that:

$$\underline{a}_{MF} > 0 \tag{C-11}$$

Additionally, it is always true that

$$b_{MF}^{(i)} \geq a_{MF}^{(i)} \quad \forall i = 1,...,m^{*}; \tag{C-12}$$

hence, applying the definitions of $\underline{b}_{MF}$ and $\underline{a}_{MF}$ to (C-12), it follows that:

$$\underline{b}_{MF} \geq \underline{a}_{MF} \tag{C-13}$$

Combining (C-13) and (C-11), it follows that $\underline{b}_{MF} \geq \underline{a}_{MF} > 0$; hence,

$$\underline{b}_{MF} > \underline{a}_{MF} \tag{C-14}$$

(c) Using the equation for $b_{MF}^{(i)}$, given in the third row of Table 5.3, it follows that:

$$b_{MF}^{(i)} = M - \frac{(a'^{(i)} + b'^{(i)})}{2} + \frac{(b'^{(i)} - a'^{(i)})}{\sqrt{6}} = \frac{(b^{(i)} + a^{(i)})}{2} + \frac{(b^{(i)} - a^{(i)})}{\sqrt{6}} \tag{C-15}$$

Consequently,

$$b_{MF}^{(i)} < \frac{(a^{(i)} + b^{(i)})}{2} + \frac{(b^{(i)} - a^{(i)})}{2} = b^{(i)} \qquad \text{(C-16)}$$

Applying the definitions of $\bar{b}_{MF}$ and $\bar{b}$ to (C-16), it follows that:

$$\bar{b}_{MF} < \bar{b}\,; \qquad \text{(C-17)}$$

but, it is also true that:

$$b^{(i)} \leq M \;\; \forall i = 1,\ldots,m\,* \qquad \text{(C-18)}$$

so that

$$\bar{b} \leq M \qquad \text{(C-19)}$$

Combining (C-17) and (C-19), observe that:

$$\bar{b}_{MF} < M \qquad \text{(C-20)}$$

Next, using the equation for $a_{MF}^{(i)}$, given in the third row of Table 5.3, it follows that:

$$a_{MF}^{(i)} = M - \frac{(a'^{(i)} + b'^{(i)})}{2} - \frac{\sqrt{6}(b'^{(i)} - a'^{(i)})}{3} = \frac{(b^{(i)} + a^{(i)})}{2} - \frac{\sqrt{6}(b^{(i)} - a^{(i)})}{3} \qquad \text{(C-21)}$$

Comparing (C-21) and (C-15), observe that:

$$b_{MF}^{(i)} > a_{MF}^{(i)} \ \forall i = 1,...,m^* \qquad \text{(C-22)}$$

Applying the definitions of $\bar{b}_{MF}$ and $\bar{a}_{MF}$ to (C-22), it follows that:

$$\bar{b}_{MF} > \bar{a}_{MF} \qquad \text{(C-23)}$$

Finally, combining (C-20) and (C-23), one obtains:

$$\bar{a}_{MF} < \bar{b}_{MF} < M \qquad \text{(C-24)}$$

References

[1] Alejandro Albertoni and Larry W. Lake, "Inferring Interwell Connectivity only from Well-Rate Fluctuations in Waterfloods," SPE 83381, 2003.

[2] A. N. Araqye-Martinez, "Estimation of autocorrelation and its use in sweep efficiency calculation," MS thesis, U. of Texas at Austin, Austin, Texas, 1993.

[3] D. Dubois, L. Foulloy, G. Mauris and H. Prade, "Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities," *Reliable Computing*, vol. 10, pp. 273-297, 2004.

[4] C. Eisenhart, M. W. Hastay and W. A. Wallis, *Techniques of Statistical Analysis,* Ch. 2, McGraw-Hill Book Co. NY, 1947.

[5] Simon Haykin, Kalman Filtering and Neural Networks. John Wiley & Sons, 2001.

[6] Heffer, K. J., Fox, R. J., and McGill, C.A., "Novel Techniques Show Links Between Reservior Flow Directionality, Earth Stree, Fault Structure and Geomechanical Changes in Mature Waterfloods," Paper SPE30711 Presented at the 1995 SPE Annual Technical Conference and Exhibition, Dallas, 22-25, Oct., 1995.

[7] F. Herrera, E. Herrera-Viedma and L. Martinez, "A fusion approach for managing multi-granularity linguistic term sets in decision making," *Fuzzy Sets and Systems*, vol. 114, pp. 43-58, 2000.

[8] F. Herrera and L. Martinez, "A type-2 fuzzy linguistic representation model for computing with words," IEEE Trans. on Fuzzy Systems, vol. 8, no. 6, pp. 746-752, 2000.

[9] J. Kacprzyk and R.R. Yager, "Linguistic summaries of data using fuzzy logic," Int'l. J. of General Systems, vol. 30, pp. 33-154, 2001.

[10] N. N. Karnik and J. M. Mendel, "Centroid of a type-2 fuzzy set," *Information Sciences*, vol. 132, pp. 195-220, 2001.

[11] George. J. Klir, "Uncertainty and Information: Foundations of Generalized Information Theory", John Wiley & Sons, Inc., New York, 2006.

[12] J. Lawry, "A methodology for computing with words," Int'l. J. of Approximate Reasoning, vol. 28, pp. 51-89, 2001.

[13] *Feilong Liu* and Jerry M. Mendel, "Aggregation Using Fuzzy Weighted Averages, as computed by the Karnik-Mendel Algorithm", Accepted by IEEE Transaction on Fuzzy Systems, 2006

[14] *Feilong Liu* and Jerry M. Mendel, "On Computing the Fuzzy Weighted Average Using the KM Algorithm", Proceeding of IPMU 2006, Paris, France, 2006.

[15] *Feilong Liu*, "An Efficient Centroid Type Reduction Strategy for General Type-2 Fuzzy Logic System", The website of IEEE Computational Intelligence Society (http://ieee-cis.org/edu/research/), 2006.

[16] *Feilong Liu* and Jerry M. Mendel, "An interval approach to Fuzzistics for interval type-2 fuzzy sets", Accepted by FUZZ-IEEE 2007, London, UK, 2007.

[17] *Feilong Liu* and Jerry M. Mendel, "Forecasting inter-well relation by using only injection rates and production rates", Accepted by ATCE2007, 2007.

[18] M. Margaliot and G. Langholz, "Fuzzy control of a benchmark problem: A computing with words approach," IEEE Trans. on Fuzzy Systems, vol.12, no. 2, pp. 230-235, 2004.

[19] Jerry M. Mendel, Lessons in Estimation Theory for Signal Processing, Communications, and Control. Prentice Hall PTR, 1995.

[20] J. M. Mendel, "Computing with words, when words can mean different things to different people," in Proc. 3rd International ICSC Symposium on Fuzzy Logic and Applications, Rochester Univ., Rochester, NY, June 1999, pp. 158-164.

[21] J. M. Mendel, Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions, Prentice-Hall, Upper Saddle River, NJ, 2001.

[22] J. M. Mendel and R. I. Bob John, "Type-2 fuzzy sets made simple," IEEE Trans. on Fuzzy Systems, vol. 10, pp. 117-127, April 2002.

[23] J. M. Mendel, "An architecture for making judgments using computing with words," Int'l. J. of Applied Mathematics and Computer Science, Vol. 12, no. 3, pp. 325-335, 2002.

[24] Jerry M. Mendel, Robert I. John and *Feilong Liu*, "On Using Type-2 Fuzzy Set Mathematics to Derive Interval Type-2 Fuzzy Logic System", NAFIPS'2005, Ann Arbor, Michigan, 2005.

[25] Jerry M. Mendel and *Feilong Liu*, "Super-Exponential Convergence of the KM Algorithms that are used for Type-Reduction in a Type-2 Fuzzy Logic System", Proceeding of IEEE WCCI 2006, Vancouver, BC, Canada, 2006.

[26] J. M. Mendel and H. Wu, "Type-2 fuzzistics for symmetric interval type-2 fuzzy sets: Part 1: forward problem," IEEE Trans. on Fuzzy Systems, Vol. 14, pp. 781-792, 2006.

[27] J. M. Mendel and H. Wu, "Type-2 fuzzistics for symmetric interval type-2 fuzzy sets: Part 2: Inverse problem," IEEE Trans. on Fuzzy Systems, 2007, in press.

[28] J. M. Mendel, "On eliciting membership function information from subjects," submitted for publication, 2006.

[29] Jerry M. Mendel, Robert I. John and *Feilong Liu*, "Interval Type-2 Fuzzy Logic Systems Made Simple", IEEE Transaction on Fuzzy Systems, vol. 14, pp. 808-821, No. 6, Dec., 2006.

[30] Jerry M. Mendel and *Feilong Liu*, "Super-Exponential Convergence of the Karnik-Mendel Algorithm for Computing the Centroid of an Interval Type-2 Fuzzy set", IEEE Transaction on Fuzzy Systems, vol. 15, No. 2, pp. 309-322, April, 2007.

[31] J. M. Mendel, "Computing with words and its relationship with fuzzistics," Information Sciences, Vol.177, pp. 988-1006, 2007.

[32] J. M. Mendel and H. Wu, "New results about the centroid of an interval type-2 fuzzy set, including the centroid of a fuzzy granule," Information Sciences, vol. 177, pp.360-377, 2007.

[33] J. M. Mendel, "Type-2 fuzzy sets and systems: an overview," *IEEE Computational Intelligence Magazine*, vol. 2, pp. 20-29, February 2007.

[34] A. O'Hagan and J. E. Oakley, "Probability is perfect, but we can't elicit it perfectly," *Reliability Engineering and System Safety*, vol. 85, pp. 239-248, 2004.

[35] M. N. Panda and A. K. Chopra, "An integrated approach to estimate well interaction," paper SPE 39563 presented at the 1998 SPE India oil and gas conference and exhibition, New Delhi, Feb., 17-19, 1998.

[36] De Sant'Anna Pizarro, J. O., "Estimating injectivity and lateral autocorrelation in heterogeneous media," Ph.D dissertation, U. of Texas at Austin, Austin, Texas, 1998.

[37] B. T. Refunjol, "Reservoir characterization of North Buck Draw field based on tracer response and production/injection analysis," MS thesis, U. of Texas at Austin, Austin, Texas, 1996.

[38] S. H. Rubin, "Computing with words," IEEE Trans. on Systems, Man and Cybernetics, Part B, vol. 29, no. 4, pp. 518-524, 1999.

[39] T. Soeriawinata and M. Kelkar, "Reservoir management using production data," paper SPE 52224 presented at the 1999 SPE Mid-Continent operation symposium, Oklahoma City, Oklahoma, March 28-31, 1999.

[40] Thakur, Ganesh C. and Abdus Satter, Integrated Waterflood Asset Management. PennWell Corporation, 1998.

[41] I. B. Turksen, "Type-2 representation and reasoning for CWW," Fuzzy Sets and Systems, vol. 127, pp. 17-36, 2002.

[42] Stephen B. Vardeman, "Statistics for Engineering Problem Solving", IEEE Press, New York, 1994.

[43] R. W. Walpole, R. H. Myers, A.L. Myers and K. Ye, *Probability & Statistics for Engineers and Scientists*," 8th Edition, Prentice-Hall, Upper Saddleback River, NJ, 2007.

[44] H. Wang and D. Qiu, "Computing with words via Turing machines: A formal approach," IEEE Trans. on Fuzzy Systems, vol.11, no. 6 , pp. 742-753, 2003.

[45] J. H. Wang and J. Hao, "A new version of 2-tuple fuzzy linguistic representation model for computing with words," IEEE Trans. on Fuzzy Systems, vol. 14, no. 3, pp. 435-445, 2006.

[46] D. Wu and J. M. Mendel, "The linguistic weighted average," Proc. of 2006 IEEE Int'l. Conf. on Fuzzy Systems, pp. 3030-3037, Vancouver, CA, July 2006.

[47] D. Wu and J. M. Mendel, "A vector similarity measure for interval type-2 fuzzy sets and type-1 fuzzy sets," submitted for publication, 2006.

[48] R. R. Yager, "On the retranslation process in Zadeh's paradigm of computing with words," IEEE Trans. on Systems, Man and Cybernetics, Part B, vol.34, no. 2, pp. 1184-1195, 2004.

[49] R. R. Yager, "A new approach to the summarization of data," Information Sciences, vol. 28, pp. 69-86, 1982.

[50] A. A. Yousef, P. Gentil, J. L. Jensen and L. W. Lake, "A Capacity Model to Infer Interwell Connectivity From Production and Injection Rate Fluctuations," SPE 95322, Presented at the SPE ATCE 2005, Dallas, Oct., 2005.

[51] L. A. Zadeh, "Fuzzy Logic = Computing with words," IEEE Trans. on Fuzzy Systems, vol. 4, pp. 103-111, 1996.

[52] L. A. Zadeh, "From computing with numbers to computing with words—from manipulation of measurements to manipulation of perceptions," IEEE Trans. on Circuits and Systems–I: Fundamental Theory and Applications, vol. 4, pp. 105-119, 1999.