A HIDDEN MARKOV APPROACH FOR SCREENOUT DETECTION IN UNCONVENTIONAL RESERVOIRS

by

Xiaodan Yu

A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Master of Science (Geophysics).

Golden, Colorado

Date _____

Signed: _____
Xiaodan Yu

Signed: _____
Dr. Whitney Trainor-Guitton
Thesis Advisor

Golden, Colorado

Date _____

Signed: _____
Dr. Paul Sava
Professor and Head
Department of Geophysics

ABSTRACT

Hydraulic fracturing has gained its popularity all over the world as more tight geologic formations are developed economically for hydrocarbon resources. While exploring for new unconventional resources such as shale plays, horizontal drilling and multi-stage hydraulic fracturing are required to stimulate the geologic units to increase well production. However, due to the stages' operating complexity, different kinds of disruptions in fracturing operations may occur and even result in great economic loss. Screenout is one of the issues caused by the blockage of proppant inside the fractures. In this project, a horizontal well landing in the Niobrara B shale, Denver-Julesburg (DJ) Basin, is simulated with multiple fracturing stages in a hydraulic fracturing software and various synthetic fracturing treatment data are forward modeled for both screenout and non-screenout scenarios. This thesis describes a screenout classification system based on Gaussian Hidden Markov Models, trained on simulated data, in order to predict screenouts and provide early warning by learning pre-screenout patterns in the simulated surface pressure signals. The classification system consists of two Gaussian Hidden Markov Models (screenout and non-screenout), each of which is fitted and optimized by its respective training set. Both Hidden Markov Models are assigned with two 1D Gaussian probability density functions to represent the distribution of their associated simulated surface pressure signals. During the classification process, once a new surface pressure sequence is observed, the log likelihood is calculated under both fitted models and the model with greater likelihood will be predicted as the class of this new observation. The classification system is validated and verified with a hold-out testing data set from the simulations and the statistics of the performance is presented in a confusion matrix. The results indicate the classification system achieves 86% accuracy for successfully predicting screenout events around 8.5 minutes prior to screenout occurring in the simulation. The described methodology is demonstrated to be a useful tool for early screenout detections

and shows its promising feasibility of other time-series analysis such as microseismic data.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

State Transition Probabilities . . . . . A

Emission Probabilities . . . . . B

Initial State Probabilities . . . . . Π

Set of Hidden States . . . . . S

Set of Observed Sequences . . . . . O

Training Set of Observed Screenout Sequences . . . . . O_s

Training Set of Observed Non-screenout Sequences . . . . . O_{ns}

Total Number of Hidden States . . . . . n

Total Number of Observations in a Sequence . . . . . m

Time . . . . . t

The Log Likelihood Function L

General Hidden Markov Model with Gaussian Emissions . . . . . λ

Hidden Markov Model with Gaussian Emissions for Screenout . . . . . λ_1

Hidden Markov Model with Gaussian Emissions for Non-screenout . . . . . λ_2

Conditional Probability of Event X Given Y has Occurred . . . . . $P(X|Y)$

Mean . . . . . μ

Variance . . . . . σ^2

LIST OF ABBREVIATIONS

Hidden Markov Model with Gaussian Emissions . . . . . GHMM

Hidden Markov Model . . . . . HMM

Probability Distribution Function . . . . . PDF

ACKNOWLEDGMENTS

First, I want to express my profound gratitude to my advisor, Dr. Whitney Trainor-Guitton, for her guidance, support, and encouragement throughout my two and a half years in the Department of Geophysics. Before entering my Master's program, I had limited research experience and GP background, Dr. Whitney always pulls me through and enlightens my road forward. It is my honor to meet such a mentor on my life's journey.

I am extremely grateful to Dr. Jennifer Miskimins for her generous help. I have known Dr. Miskimins since the summer of 2016 while I was in her summer externship program, her kindness and strict working attitude made a deep impression on me. During the time of my research, she spent a great amount of time to discuss the problems I encountered and provided invaluable advice.

I would also like to thank Dr. Jeffrey Shragge for his time and inputs to my research and thesis writing.

Huge thanks to all of my friends and teammates in W team, from whom I have learned so much in doing research.

In the end, I hope to send my heartful thanks to my family for always being supportive during my time at school. Most importantly, I want to thank my beloved husband, Rong, who provides endless inspirations in my life.

This work is dedicated to my parents, Dr. Yongwei Yu and Ms. Hongqing Lu.

To my grandparents, Ms. Xunlan Liu and Mr. Baoyu Lu.

CHAPTER 1

INTRODUCTION

With the world's growing demand for energy, the discovery of unconventional resources and their use are receiving more attention. However, many challenges still exist when exploring for unconventional hydrocarbons (shale gas/oil, tight gas/sandstone, etc.). Unlike conventional reservoirs, such as sandstone where fluids (oil and gas) can flow easily towards the wellbore, unconventional reservoirs have low-permeability and cannot produce economically without the aid of stimulation treatments or special recovery operations [1]. As illustrated in Figure 1.1, horizontal drilling technology is applied to create more reservoir contact area [2]. After a horizontal well is drilled and completed, hydraulic fracturing, as one of well stimulation treatments, is performed to enhance the well productivity by inducing high-permeable fractures near the wellbore. Currently, multiple fracturing stages are widely applied on the horizontal wellbores to achieve more economical benefits. However, the stages' operating complexity would sometimes result in several disruptions in fracturing operations and may even jeopardize the safety of field staff. Screenout is a problem that can happen during a fracturing stage due to the blockage of proppant inside the fractures. In the best case scenario, a screenout could cause a delay in the ensuring fracturing stages; in the worst case scenario, screenout can result in extra cost as a result of wellbore cleanout and loss of production days [3]. With no well-recognized and sophisticated methods to detect screenout in the development of unconventional reservoirs, it is clear that such a technology could have significant impact on the industry. This thesis explores a data-driven approach to analyze and identify the signatures of pre-screenout signals by applying a statistical learning model – the Hidden Markov Model on the fracturing time-series data.

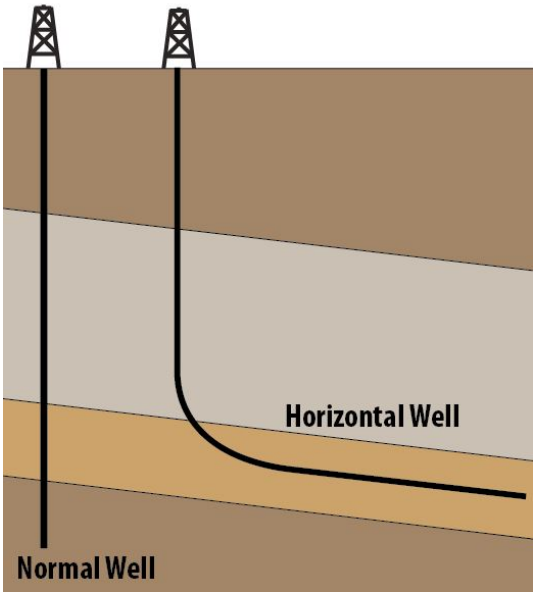


Figure 1.1: A normal/vertical well (left) and a horizontal well (right) drilled into an unconventional reservoir (the third layer from the top). As compared to the vertical well, the horizontal well has longer exposed section length into the producing formation, which improves the well productivity [4].

1.1 Hydraulic Fracturing in Unconventional Reservoirs

Hydraulic fracturing, as a practice of reservoir stimulation, has been broadly applied in the oil and gas industry since 1947 [5]. A great advancement of hydraulic fracturing is that it enables economic production in unconventional reservoirs that geologists used to believe were unable to produce. The hydraulic fracturing process mainly refers to a procedure of injecting fracturing fluid blended with special chemicals into the pay zone at high pumping rates and pressures. During the operation of hydraulic fracturing, a clean fluid called a "pad" is first pumped into the wellbore. After the fractures are initiated, they tend to close gradually. At this time, a slurry carrying a propping agent (called as a "proppant") is injected to the subsurface formation causing the fractures to be propped and extended. Once pumping is completed, the fracturing fluid chemically breaks down to a lower viscosity and flows back out of the wellbore. Before production activity starts, highly conductive fractures are expected to be around the wellbore [5]. As Figure 1.2 shows, fractures are created by

hydraulically stimulation near a horizontal wellbore. The proppants carried in the fracturing fluid stay inside of the fractures to maintain the fractures' opening, thereby facilitating the entry of a great amount of natural gas into the wellbore.

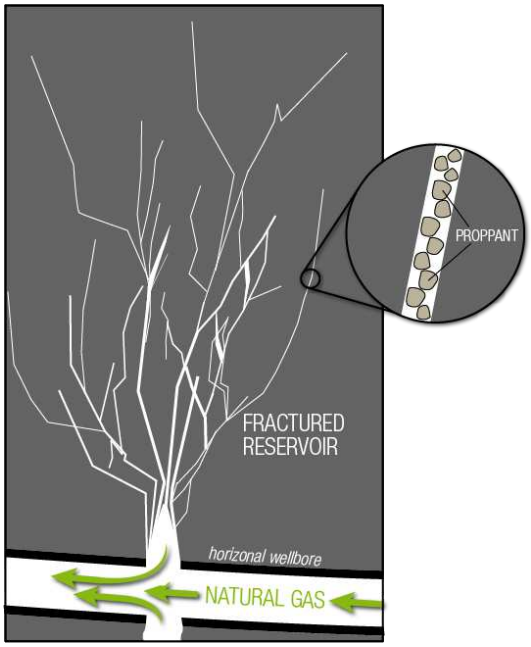


Figure 1.2: An illustration of fractures near a horizontal wellbore. The proppants carried in the fracturing fluid hold the fractures open so that the natural gas can easily enter to the production tubing [6].

Among the techniques of multi-stage well completion, "plug-and-perf" is commonly used as a completion approach, which is combined with hydraulic fracturing for the development of unconventional wells (see Figure 1.2) [7]. In this system, the perforation guns are first placed at the toe of the well (the far end of the horizontal wellbore) and multiple sets (clusters) of perforated holes are made successively (named as cluster perforations). Then, hydraulic fracturing takes place when fluids are simultaneously pumped into all clusters of perforations. Once the fracturing job at this stage is completed, a plug is set to isolate from the previous one. This process is then repeated until all stages have been fractured.

1.2 Screenouts during Hydraulic Fracturing Treatments

Proppants are one of the key factors that influence fracturing performance. Therefore, an assurance of good proppant transportation into the fracture network is pivotal, or a large economic loss would be caused. Among the various proppant transport issues, screenout has been considered by operators as one of the more serious problems. During the process of fracturing, surface treating pressure, known as wellhead pressure or surface injection pressure, is recorded by a gauge at or near the surface in a well where the fracturing fluids are pumping through to monitor the on-going fracturing job [8].

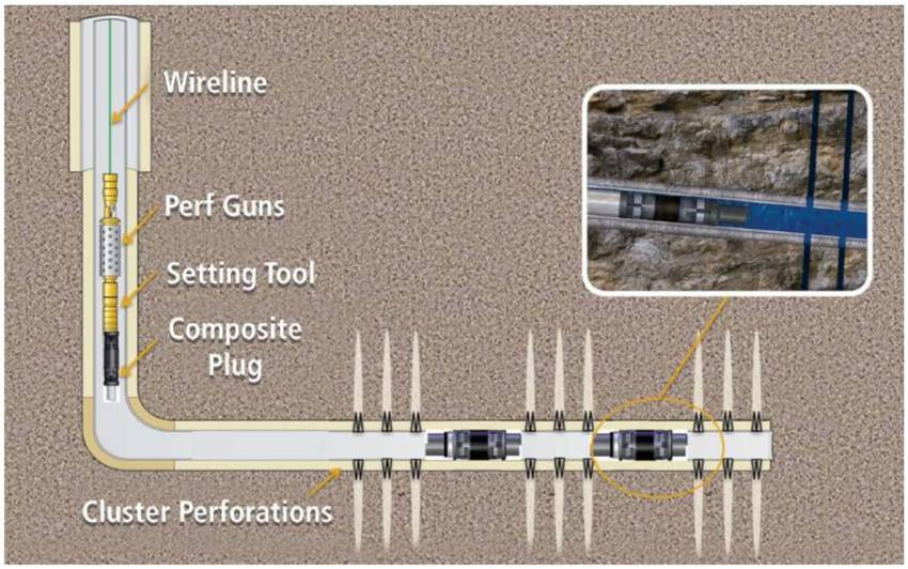


Figure 1.3: Wellbore diagram of a plug-and-perf system. After the fracturing job at the first (the rightmost) stage is finished, a wireline assembly is pumped down and a composite plug is placed at the bottom of the assembly. Once the assembly is taken to the intended depth, an electric signal will be sent through the wireline and activate the setting tool which is above the plug. Then, the plug will be set and released by the setting tool. When the assembly is moved to the place of the first cluster of perforations at the second stage, another electric signal will be sent to activate the wireline and fire the perforation guns. As this diagram shows, there are three clusters of perforations at each stage [7].

Under the majority of conditions, surface treating pressure is the most direct signal to detect screenouts through a drastic pressure rise within a short period of time (around 2.5 minutes). Figure 1.4 presents an example of screenout observed in the surface treating

pressure data (blue-colored line).

Screenout occurs when the accumulation of proppants near the wellbore reduces the fracture cross-section available for fluid flow, which results in higher frictional pressure losses inside the fracture. The existence of screenout would place the wellbore on a stagnant status for cleanout with extra cost, thus delaying the subsequent fracturing stages and losing production days [3]. More importantly, a sharp rise in pumping pressure caused by a screenout could damage the surface equipment, which even threatens field staff safety.

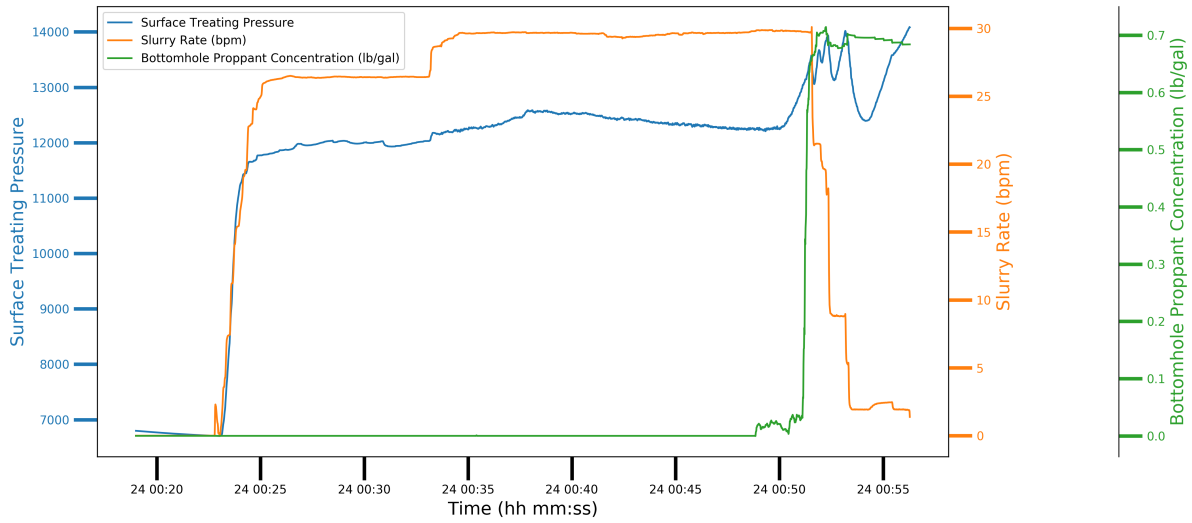


Figure 1.4: A field example of screenout in a surface treating pressure profile. The slurry rate and bottomhole proppant concentration is also included to illustrate the surface pressure behavior. The first rapid growth in the surface pressure (around 24 00:23) is due to the initiation of fluid pumping. The second growth (around 24 00:50) is caused by the arrival of proppant at the bottomhole. Then with a decreasing slurry rate and a flat injecting rate of proppant, the surface pressure suddenly rises (around 24:00:55) which indicates a screenout.

To effectively mitigate the influence of screenouts, [9] proposed the net-pressure plot as a real-time diagnostic tool and decision-control procedure to predict screenouts. The plot analyzes the net pressure behavior, which is the difference between the fracturing fluid pressure and the closure pressure based on the Perkins-Kern-Nordgren (PKN) theory [10]

during and after the fracturing treatment. By calculating the slope of net pressure versus time, different patterns of fracture growth can be determined (see Figure 1.5).

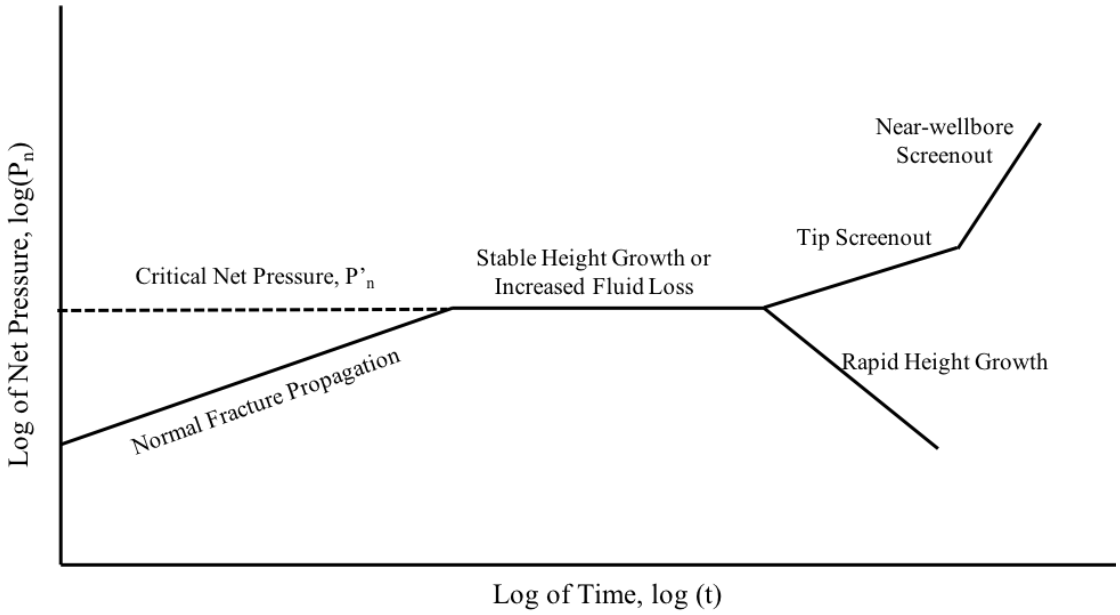


Figure 1.5: Diagram of a classic net-pressure plot [9]. Each slope of net pressure versus time represents a mode of fracture growths (five modes in total).

Although the net-pressure plot method has proven very effective in evaluating fracturing treatments during the conventional reservoir explorations, its qualitative assumption that the net pressure is constant becomes invalid in unconventional reservoirs. [11] explains the invalid assumption is due to a phenomenon called stress shadowing. When the fracturing treatments are carried out in horizontal wells, multiple fracturing stages are performed successively. Once a new nearby fracturing job is initiated, the interaction (including induced stress and strain fields around each existing fracture) cannot be ignored because it can affect fracture growth, geometry, and treating pressure. Such fracture-stress interference is described as stress shadowing, and it appears between fracturing stages, where multiple fracture sets are induced [11]. Figure 1.6 presents a view of stress shadowing between adjacent fractures. It can be seen that on the left side of the figure, if the vertical transverse fractures are far away from each other, the assumption of constant net pressure could still be effective. But if the

fractures are initiated closely, the net pressure inside the features is lower than outside [11].

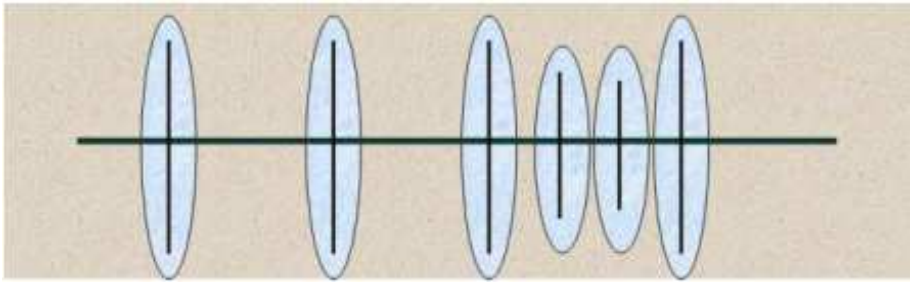


Figure 1.6: An illustration of stress shadowing. The black bold line stands for a horizontal wellbore and the vertical lines represent the hydraulic fractures initiated in multiple hydraulic fracturing stages. The light blue-shaded elliptical regions are the the stress shadowed regions [11].

1.3 Literature Review

Screenout prediction methods can be divided into two categories: post-minifrac and real-time diagnostic methods [3]. Minifrac refers to a small treatment that is performed before a main fracturing job, and its recorded data are useful to refine the parameters of the final fracturing job. Recently, various approaches have been proposed to specifically tackle screenout predictions in unconventional reservoirs under the category of post-minifrac. In [12], the pressure declining gradient is employed as a pressure declining feature of leak-off tests to predict the likeliness of screenout in a tight-gas formation. [3] describes a solution of using an indicator, named as screenout index (SOI), to predict screenouts in the Eagle Ford shale. The SOI is developed by analyzing rate step-down tests and used as a post-minifrac diagnostic tool to provide early warning of potential screenouts during the hydraulic fracturing treatments. However, there are not many new studies published on real-time screenout diagnosis in the previous decade. Besides the net-pressure plot [9] introduced in Section 1.2, [13] is the most recent one that finds a screenout is very likely to occur when the surface pressure slope deviates from the inverse slope (postive pressure slope), which could

deliver prompt and advanced warning to field staff.

1.4 Research Objective

Currently, there is no comprehensive real-time solution such as the net-pressure plot [10] for predicting screenout events in horizontal wells. Although people have investigated different methods for screenout diagnosis, they are all from a physics-based perspective that requires additional calculations such as pressure slope. Therefore, beginning from a statistics-based perspective, this work proposes a data-driven approach to study the unique pre-screenout patterns directly in the surface treating pressure data that could provide early warning before actual screenouts take place during the horizontal fracturing operations. The novelty of such approach is that it can be applied as a real-time screenout diagnostic tool that does not require a prior minifrac and additional pressure calculations. In this project, a statistical model, specifically a Hidden Markov Model, is implemented as a classifier to develop the screenout classification system.

1.5 Background of Hidden Markov Models

The Hidden Markov Model (HMM), named after the Russian mathematician Andrey Markov who is best known for the theory of stochastic Markov processes, was first introduced by Baum and Petrie in 1960s [14]. The HMM is a statistical model based on a Markov process with hidden states. In the HMM, an observed sequence is believed to be generated by an invisible Markov chain that contains a series of hidden states, and each hidden state generates one observation randomly out of this observation sequence. HMM has been widely used in many applications such as time-series analysis. For oil and gas applications, [15] develops an HMM-based classification method that achieves around 92% accuracy to distinguish different kinds of microseismic signals in a coal mine. [16] proposes an HMM-based method to detect a salt dome using migrated seismic data with an average classification accuracy of 98.7%. In [17], an HMM-based workflow is used to automatically interpret different stratigraphic zones using well-logging data. In this work, since surface treating pressures used to study

screenouts are time-series data, the research question can then be reformulated to a time-series analysis. Hence, HMM can be a good solution because it is reasonable to assume that pressure signals can be generated by some underlying stochastic processes.

1.6 Thesis Structure

This thesis consists of four chapters. The first chapter introduces the general information of hydraulic fracturing and the occurrence mechanisms of a screenout. Then, the difficulties of implementing the current existing screenout detection method in unconventional reservoirs are explained.

The second chapter discusses synthetic data generation and the methodology used to develop a screenout classification system. Three main aspects of the data simulation strategies: reservoir pressure design, well perforation design, and pumping schedule design. Regarding the screenout recognition system, an HMM with Gaussian emissions (GHMM)-based workflow is presented and two GHMMs are fitted and optimized by training sets of screenout (O_s) and non-screenout (O_{ns}) data, respectively.

The third chapter focuses on the discussion of the results. The synthetic surface pressure data are first compared with a screenout example acquired from the Wattenberg field. Then, a confusion matrix, a table with four different combinations of predicted and actual values, is evaluated to show the classification performance of the system. Finally, some typical misclassified examples are analyzed.

The last chapter summarizes the key points drawn from the work explained in the previous chapters and makes recommendations for future work.

CHAPTER 2

MATERIALS AND METHODOLOGIES

Due to the limited amount of field data available for screenout investigations, multiple synthetic hydraulic fracturing treatment data are developed for a horizontal well for both screenout and non-screenout scenarios. In this thesis, only one unconventional reservoir is examined to better understand screenout behavior. The reservoir conditions and geology are set to mimic the Niobrara shale formation, located in the Denver-Julesburg (DJ) Basin since this is the formation where great successes with horizontal drilling and multi-stage fracturing completions have been reported. The simulated treatment data are used to generate a screenout and non-screenout model (training state). These models are validated and tested from the respective hold-out data. Finally, a GHMM-based classification system is developed to help identify the pre-screenout behaviors using simulated surface treating pressure data.

2.1 Synthetic Fracturing Treatment Data Generation in the Niobrara-DJ Formation

Fracturing treatment data are generated using a hydraulic fracturing simulation software. A well-logging profile acquired from a vertical well located in the Denver-Julesburg (DJ) Basin is used as a well log reference in my fracturing simulations. The DJ Basin has been recognized as one of the largest producers in the United States [18]. Among the major producing sources of the DJ Basin, the Niobrara formation is selected in this study as the target zone into which a horizontal well is simulated landing. The Niobrara formation is an organic rich shale that is situated in Northeast Colorado, Northwest Kansas, Southwest Nebraska, and Southeast Wyoming [19]. As the best-known shale deposit in the DJ Basin, the Niobrara shale has been recognized as a primary petroleum production resources of the U.S. Rocky Mountain region [20]. The shale formation thickness ranges from 200 feet to 400 feet in northeastern Colorado, which requires horizontal drilling and multi-stage

fracturing [18]. As Figure 2.1 shows, the Niobrara shale is divided into three different reservoir zones, Niobrara A, Niobrara B, and Niobrara C followed by the underlying Fort Hays Limestone.

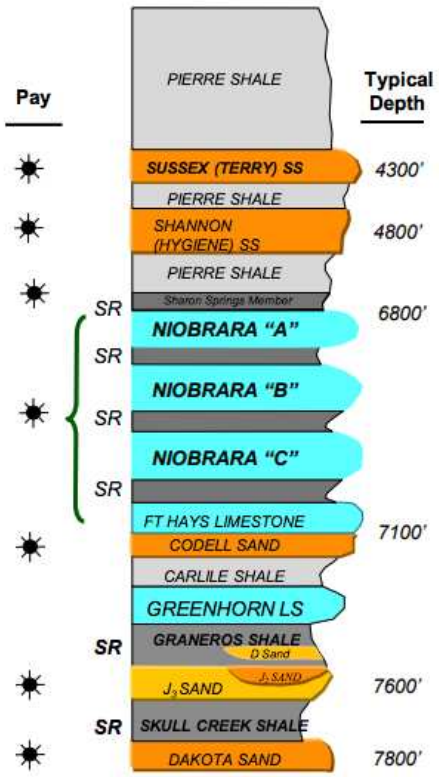
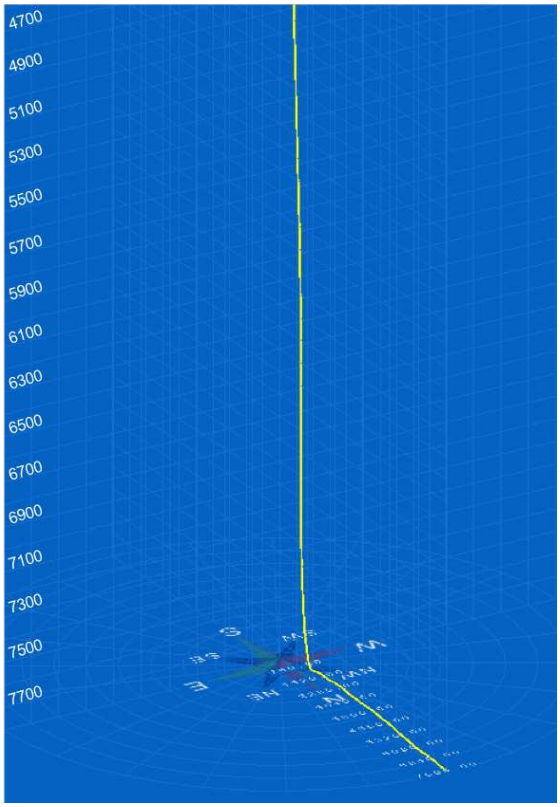


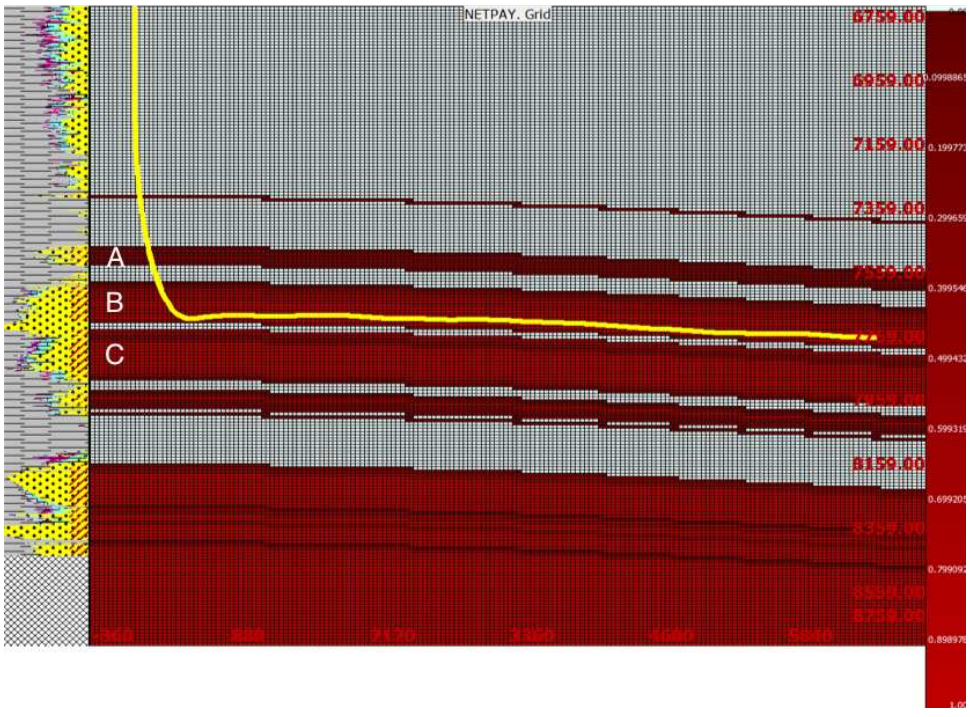
Figure 2.1: Detailed stratigraphic column of producing reservoir rocks in the Wattenberg Field (northeastern Colorado), a subset of the DJ Basin. The intervals labeled with SR are source beds for reservoir rocks. The key interval of interest in this work is indicated as Niobrara "B" [20].

In Figure 2.2(a), a well trajectory is presented for the fracturing treatment designs. One horizontal well was modeled landing specifically in the Niobrara B since it is the most actively producing zone of the Niobrara formation [21]. The horizontal well has a true vertical depth of 7800 feet with a lateral length of 6000 feet.

To better study different pre-screenout behaviors, diverse simulation results in both screenout and non-screenout scenarios are created by altering various fracturing design parameters. The fracturing design strategy is explained in the following three subsections: reservoir



(a)



(b)

Figure 2.2: Profile of the simulated horizontal well. (a) Screenshot of the horizontal well trajectory in 3D view. (b) View of the horizontal well trajectory in the pay zone where the horizontal wellbore is designed to land in the zone of Niobrara B.

pressure design, well perforation design, and pumping schedule design.

2.1.1 Reservoir Pressure Design

In general, reservoir properties such as pore pressure and permeability have significant influence on fracturing performance. In this thesis, overpressure and depletion (underpressure) scenarios are created in the Niobrara B through modifying pore pressure offset. By definition, pore pressure refers to the fluid pressure within the pores of the reservoir, which is often described as initial reservoir pressure. The pore pressure was increased and reduced from the original reservoir pore pressure by 50 psi, 100 psi, 150 psi, and 200 psi, respectively. By doing this, it allowed having nine different reservoir models, including the basecase, at the beginning of the treatment design.

2.1.2 Well Perforation Design

Three fracturing stages are assigned for the horizontal well where the fracturing treatment begins from Stage 1 to Stage 3. Figure 2.3 illustrates the perforation geometry for the lateral length of the well. Each stage is set to be 300 feet long and each perforation cluster is 100 feet apart. For each simulation, the number of the perforation shots in each cluster ranges from 55 to 65 shots.

2.1.3 Pumping Schedule Design

After determining the perforation design, a pumping schedule is required for each stage to initiate the simulation. In this study, all three stages share the same pumping schedule during one simulation, with each simulation outputting three unique treatment results, one for each stage. For the pumping schedule design, five different fracturing fluids are used for each reservoir pressure model, which covers water fracturing, linear gel fracturing, and hybrid fracturing. These fracturing fluids appear to have different abilities to transport proppants due to their different main composition. In addition, a screenout is usually caused by high proppant concentration, large proppant mesh size, and large masses of proppant

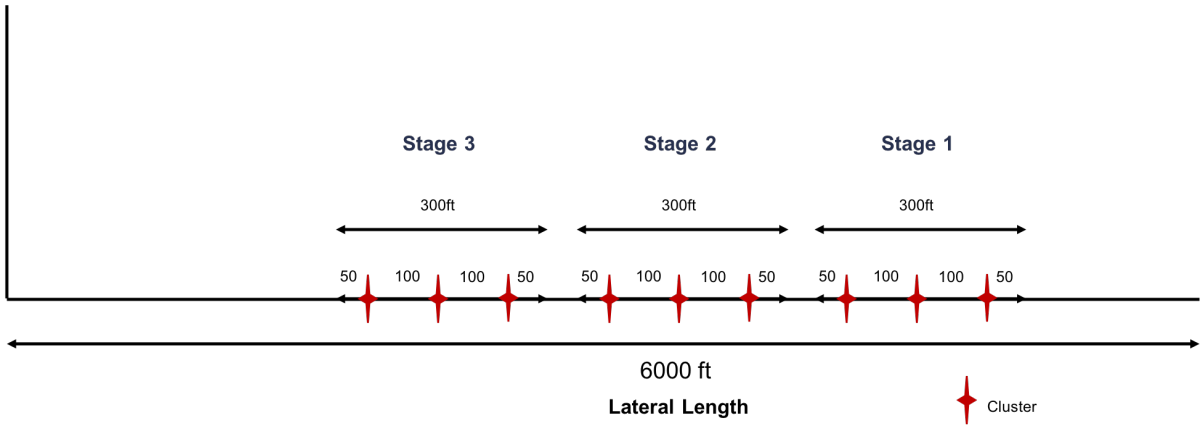


Figure 2.3: Diagram of the simple well perforation design for the well lateral length. The treatment process starts from Stage 1 (the toe of the well) to Stage 3 in sequence. Each cluster is a set of perforation shots (the number of shots varies between 55 and 65).

used in each stage of fracturing operation (cumulative proppant volume) [3]. Therefore, the proppant concentration and slurry/pumping rate (the proppant volume equals to proppant concentration multiplies by slurry/pumping rate) used in this study are mainly modified to simulate diverse screenout results. Although extremely large proppant concentration or proppant volume can induce screenout, it is essential to ensure my pumping schedules follow realistic fracturing schemes such that reasonable data are acquired for further analysis. Accordingly, the proppant concentration and pumping rate are not allowed to exceed certain limits based on their abilities to transport proppants. The linear and hybrid gel fluids will not exceed 8 pounds per gallon and slickwater will not exceed 6 pounds per gallon. Besides, the range of proppant concentration is adjusted as well depending on different fracturing fluid types. Table 2.1 lists all input five fracturing fluid names, corresponding fluid types, the range of proppant concentration, and the range of pumping rate, respectively.

Regarding the proppant type selection, the Raw Substrate sand with mesh size 20/40 is consistently used for all fracturing simulations since it is relatively large to induce screenout. As a result, 90 simulations (each having three stages of fracturing data) are modeled where 135 fracturing treatment data sets are labeled with screenout and 135 data sets are labeled

Table 2.1: Fracturing parameters experimented in the pumping schedules.

Fluid Name	Fluid Type	Proppant Concentration Range (pounds per gallon)	Slurry Rate (barrels per minute)
Guar	Linear Gel	0-8	0-65
HPG	Hybrid Gel	0-8	0-65
Slickwater_120F	Slickwater	0-6	0-65
Slickwater_150F	Slickwater	0-6	0-65
HEC	Linear Gel	0-8	0-65

as non-screenout. All simulation cases are summarized in Table 2.2. To begin to explore the data in terms of screenout detection, a Hidden Markov Model will be introduced as a tool in the next section.

2.2 The Definition of a Hidden Markov Model

An HMM has been recognized as a great analytic tool for time-series analysis. It is used to represent probability distributions for multiple observed sequential data [22]. According to the definition of an HMM: one observation at time t, o_t, is believed to be generated conditionally by the state s_t, then the Markov chain moves to the next hidden state, s_{t+1} and generates the next observation, o_{t+1}, and so on. In this thesis, a first-order HMM is selected which has two main assumptions:

1. **Markov Process**: According to the first-order Markov chain assumption, the current observation only depends on the previous observation, represents the simplest dynamic Bayesian network:

$$\mathrm{P(o)} \;=\; \prod_{\mathrm{i=1}}^{\mathrm{t}} \mathrm{p(o_t|o_{t-1})}. \tag{2.1}$$

Table 2.2: Summary of the fracturing simulation design.

<table>
<tr><td></td><td>Pore Pressure (psi)</td><td>Perforation Design</td><td>Pumping Schedule</td></tr>
<tr><td rowspan="9">Parameter Settings</td><td>200</td><td rowspan="3">Stage 1</td><td rowspan="9">As indicated in Table 2.2.</td></tr>
<tr><td>150</td></tr>
<tr><td>100</td></tr>
<tr><td>50</td><td rowspan="3">Stage 2</td></tr>
<tr><td>0</td></tr>
<tr><td>-50</td></tr>
<tr><td>-100</td><td rowspan="3">Stage 3</td></tr>
<tr><td>-150</td></tr>
<tr><td>-200</td></tr>
<tr><td>Total Scenarios:</td><td>9</td><td>3</td><td>10</td></tr>
<tr><td></td><td colspan="3">9×3×10=270 in total</td></tr>
</table>

2. **Observation Independence**: The observation o_t at time t, depends only on the current state s_t, which can be written as:

$$\mathrm{P(o_t|o_1...o_{t-1}, s_1...s_{t-1})} \;=\; \mathrm{P(o_t|s_t)}. \tag{2.2}$$

After clarifying these two assumptions, the next step is to determine the type of HMM. At present, there are many kinds of HMM can be applied, such as ergodic and left-right [23]. In my study, an ergodic model is selected since it performs better than the left-right model with the simulated surface treating pressure data. Figure 2.4 demonstrates the structure of an ergodic HMM with three hidden states and its required elements.

1. Number of hidden states, S=$\{s_1, s_2, \ldots, s_n\}$: The hidden states $s_1, s_2, \ldots, s_n$ come from a known and finite set, S. The sequence of hidden states is generated by the Markov process, each state can be connected to any other state and itself.

2. Observed sequence, O=$\{o_1, o_2, \ldots, o_m\}$: The observations $o_1, o_2, \ldots, o_m$ come from a known and finite set, O. In my case, each surface treating pressure data during the simulated fracturing treatments is treated as an observed sequence.

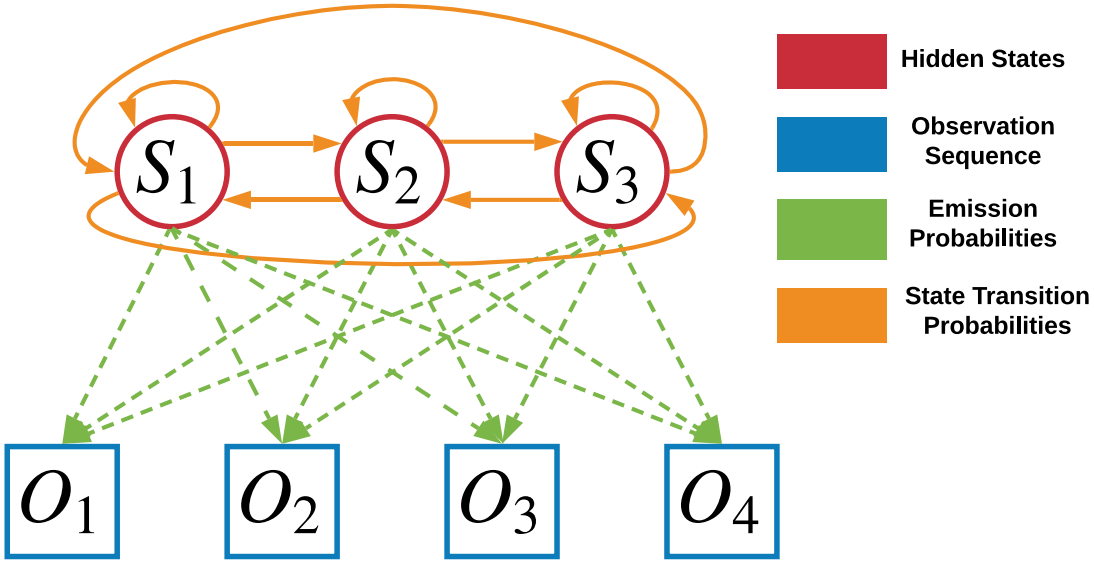


Figure 2.4: Structure of an ergodic HMM with three hidden states. For an ergodic model, the hidden states are fully connected: One hidden state could transit either to itself or any other states of the model. The observations o_1, o_2, o_3, o_4 are emitted by the hidden states. The state transition probabilities represent the probability of transitioning state s_i to state s_j $(1 \leq i, j \leq 3)$. The emission probabilities represent the probability of emitting observation k from state j $(1 \leq k \leq 4$ and $1 \leq i \leq 3)$.

3. State transition probabilities, A=$\{a_{ij}$=$P(s_j|s_i)\}$,where $1 \leq i, j \leq n$: It is the probability that the next state s_j is at time=$t+1$ given the current state s_i is at time=t. An ergodic HMM (Figure 2.4) allows each state to reach to any other state within one step since it is fully connected.

4. Emission probabilities, $B = \{b_{ik} = P(o_k|s_i)\}$, where $1 \leq i \leq n$ and $1 \leq k \leq m$: It is the probability that the observation is o_k given the current state is s_i. The form of these for the surface pressure data are described in Section 2.3.

5. Initial state probabilities, Π=$\{\pi_i$=$P(s_i$ at $t = 1)\}$, where $1 \leq i \leq n$: It is the probability set for all states at t=1. The initial state probabilities are not presented graphically in Figure 2.4. Numerical values are initially assigned.

For convenience, an HMM can then be denoted by the following set of parameters:

$$\lambda \quad = \quad (A, B, \Pi). \tag{2.3}$$

2.3 HMM for Continuous Surface Treating Pressure Data

During the fracturing treatment simulation, the pressure sequences are frequently sampled over calendar time. The typical sampling frequency for the simulated surface pressure is 1 Hz, which is consistent with the sampling rate recorded by a surface pressure gauge in the Niobrara-DJ Basin. Given that surface pressures are continuous variables, it is more suitable to use a continuous probability density function (PDF) as the emission probability function. In this thesis, a 1D Gaussian PDF is used for the continuous PDF associated with each state. Based on the current hidden state, the observation (o_k) is drawn from a 1D Gaussian PDF. Therefore, the total number of the hidden states, n, is equal to the number of the 1D Gaussian PDFs. The formula of a 1D Gaussian PDF is:

$$\mathrm{P(o_k|s_i)} \quad = \quad \mathrm{N(o_k|\mu_i, \sigma_i^2)} = \frac{1}{\sqrt{2\pi\sigma_\mathrm{i}^2}} \exp\left(\frac{-(\mathrm{o}-\mu_\mathrm{i})^2}{2\sigma_\mathrm{i}^2}\right), \tag{2.4}$$

where μ_i, σ_i^2 are the mean value and variance value associated with state s_i.

The 1D Gaussian PDF for state s_i is initialized with a μ_i=0 and σ_i^2=1 (Figure 2.5). Variables μ and σ will be updated and optimized during the model training process (called the learning problem) as explained in the following section.

2.4 Fundamental Problems of a GHMM in a Classification Setting

As described in Section 2.3, an HMM with Gaussian emissions (GHMM) is specified in this thesis. Among these various treatment outputs, the surface pressure data will be

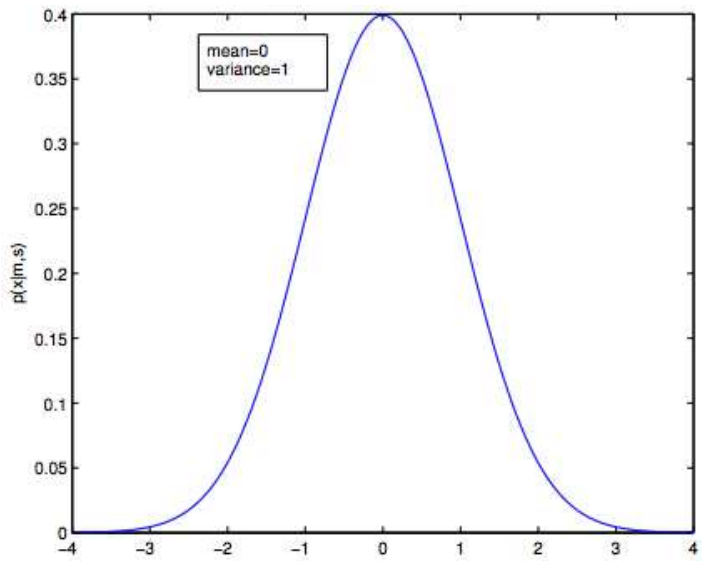


Figure 2.5: 1D Gaussian PDF with $\mu = 0$ and $\sigma^2 = 1$.

primarily studied since they are the most direct signal for screenout. It is worth noting that once the software detects screenout occuring, it will immediately stop the simulation. Therefore, the collected surface pressure signals for screenouts do not capture the subsequent increasing trends while the screenouts fully develop in the subsurface. With missing pressure signals during the whole screenout process, a GHMM-based classification system is more appropriate than a GHMM-based regression system because I want to explore and learn the pre-screenout patterns existing in my pressure data instead of the actual screenout behaviors. Furthermore, unlike a discriminative classifier such as Support Vector Machine, GHMM is fundamentally a generative model, which means that one GHMM can only model one class. Therefore, one GHMM, λ_1, is fitted by using screenout training data, and one GHMM, λ_2, is fitted by using non-screenout training data. An advantage of such classification system is that it can be further upgraded to a real-time alert system that evaluates whether the current operating well is very likely to screenout within an established period of time according to the incoming surface pressure time series. Moreover, the system allows the model to discover new unique patterns before a screenout takes place, which provides more time for field staff to adjust fracturing strategy.

Before applying each GHMM in practice, there are three main problems to be solved sequentially:

1. **The learning problem: How to optimize GHMM parameters $\lambda = (A, B, \Pi)$ to maximize the GHMM λ ability to predict observations, $P(O|\lambda)$?**

   This problem focuses on how to search the model parameters that best characterize the simulated surface pressure data. In this thesis, the model parameters are optimized in the training process of the classification system with an iterative Expectation-Maximization (EM) algorithm. I will discuss the training process and EM algorithm [24] in detail in Section 2.5.

2. **The evaluation problem: Given a GHMM λ=(A, B, Π) and a sequence of observations $\{O = o_1, o_2, \ldots, o_m\}$, what is the probability $P(O|\lambda)$ that observations are produced?**

   In my classification system, given the model parameters, λ_1 and λ_2 have been optimized first in the training process (after the learning problem is solved), this problem determines to which class (screenout or non-screenout) each new observed pressure sequence belongs. The solution applied here is maximum log likelihood, which will be illustrated in Section 2.5.

3. **The decoding problem: Given a GHMM λ=(A, B, Π) and a sequence of observations $\{O = o_1, o_2, \ldots, o_m\}$, what is optimal corresponding hidden state path in the model?**

   The solution to this problem would tell the hidden state sequence carried by the simulated surface pressure data. However, in a classification system, the solution to this problem is much less important, so it will not be addressed in this thesis.

2.5 GHMM-based Screenout Classification System

This section emphasizes on the development of a classification system that trains two GHMMs to classify whether the surface pressure sequence will be screenout or not. Figure 2.6 presents the overall workflow of the screenout classification system based on training GHMMs. Before initiating the classification system, the data set is split into training, validation, and testing sets based on 70%, 15%, and 15%, respectively.

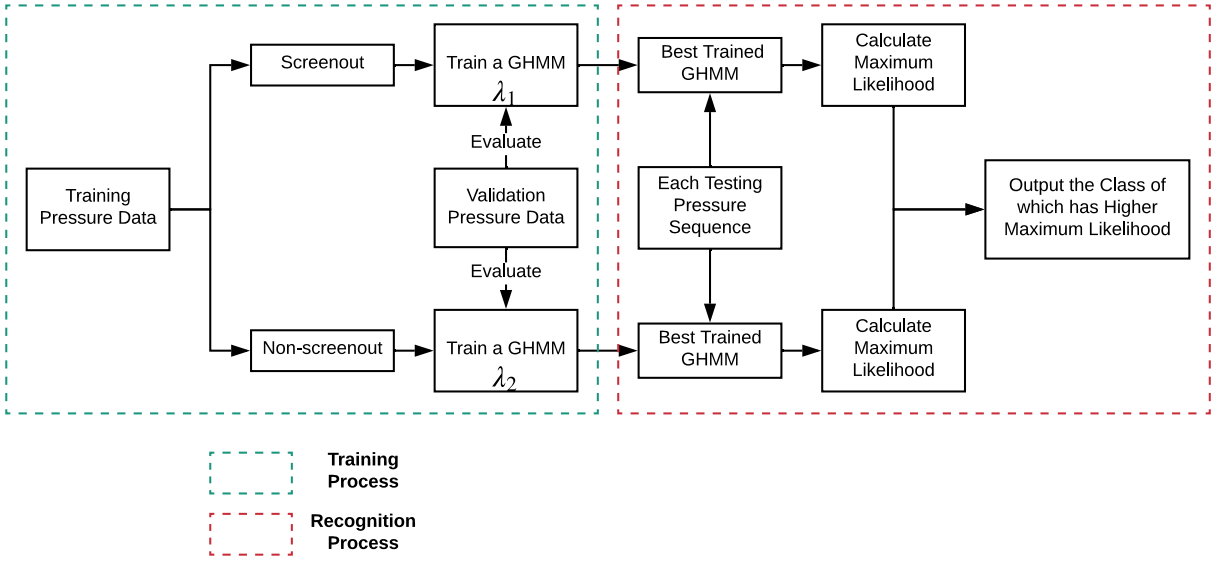


Figure 2.6: GHMM-based screenout event classification system. The system is divided into training and recognition process. The training process estimates the optimal model parameters including the number of hidden states for both screenout and non-screenout models. In the recognition process, the best-trained models are used to determine the class of each pressure sequence in the testing set.

In Figure 2.6, the classification system is divided into two processes: model training and model recognition. In the former, training data are used to train the model five times with different hidden state numbers (from 1 to 5), and the validation data are applied to determine which state number makes the model achieve optimal performance. The testing data are utilized in the recognition process. More importantly, to better explain how these two fundamental problems are solved during the training process (Figure 2.6), I include more

detailed procedures in the training process in Figure 2.7. The learning problem objective is to discover the optimal model parameters of both GHMMs such that the model optimally characterizes the features in the training pressure data for screenout and non-screenout events. As Figure 2.7 shows, I first initiate a reasonable guess for the parameters of the two GHMMs: λ_1, λ_2, and their number of hidden states (n) which starts at 1.

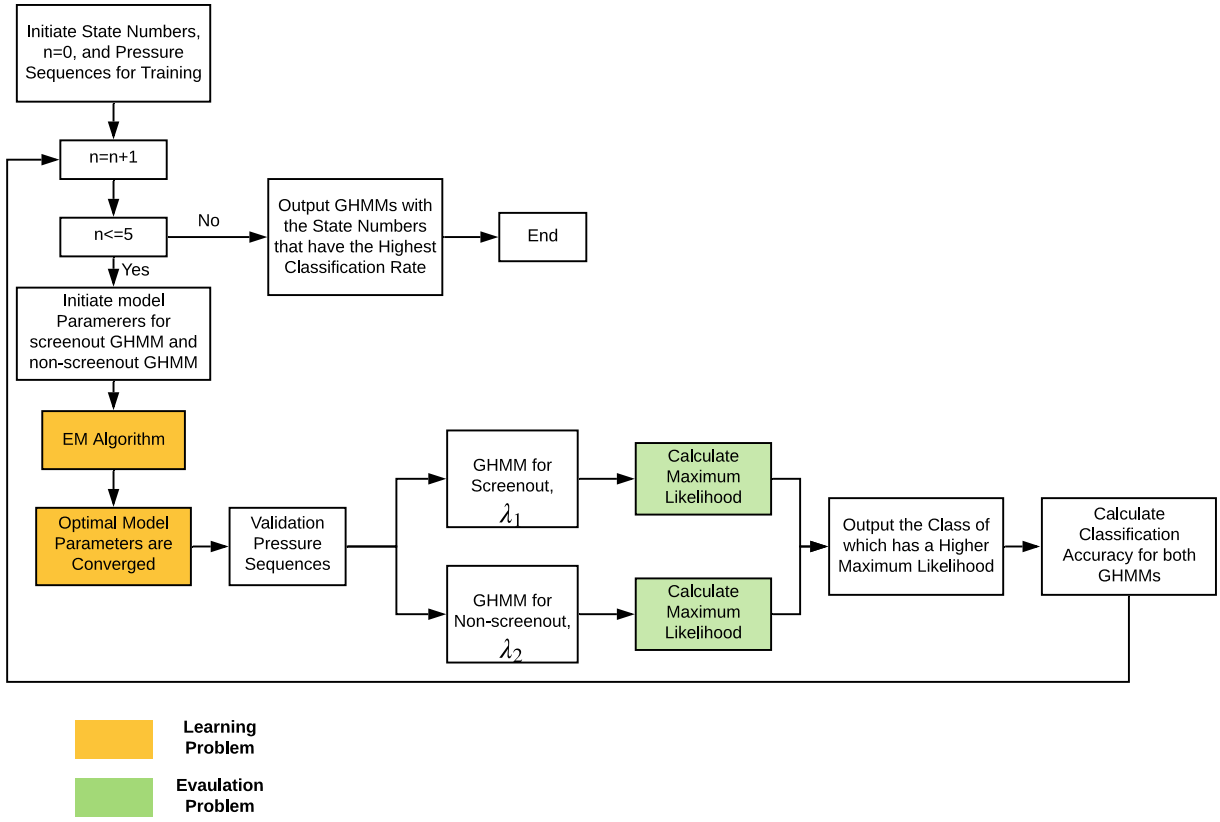


Figure 2.7: Detailed training process in the GHMM-based classification system. The model parameters λ=(A, B, Π) for both GHMMs are optimized by the EM algorithm [24]. The optimal number of hidden states in each GHMM, n, is determined by the highest classification accuracy using its respective validation set. All procedures fall into the range of training process in Figure 2.6.

In the next step, I tackle the learning problem discussed in Section 2.4. The learning problem can be mathematically reformulated as an optimization problem aimed at finding the model parameters that optimally describe the surface pressure data in the training set.

Since I have two GHMMs in the classification system, the whole training set is first separated into screenout and non-screenout events based on data labels. Hence, for each GHMM, I want to maximize the following formula:

$$\underset{\lambda_1}{\operatorname{argmax}} \mathrm{P}(\mathrm{O_s}|\lambda_1), \tag{2.5}$$

$$\underset{\lambda_2}{\operatorname{argmax}} \mathrm{P}(\mathrm{O_{ns}}|\lambda_2), \tag{2.6}$$

where O_s, O_{ns} stand for all collected training screenout and non-screenout observed sequences.

Then, an iterative Expectation-Maximization (EM) algorithm, known as the Baum-Welch algorithm [24], is used to solve Equations 2.5 and 2.6 by determining the model parameters that best explain the training observation sequences. Given the above definitions, I begin with an initial model λ_i=(A, B, Π) for each GHMM, and run its corresponding training data through the current model to estimate the mean value of each model parameter. Then, the model can be changed to maximize the values of the used paths. Through iterating, the optimal model parameters are finally converged [25].

After defining the optimal parameters for both GHMMs, each surface pressure sequence from the validation set is input to these two models to evaluate how likely each model would be to generate such a sequence. When an unseen surface sequence is observed, its maximum log likelihood under each GHMM (screenout and non-screenout) is calculated. The maximum log likelihood for the two GHMMs is:

$$\mathrm{L}(\mathrm{O}|\lambda_1) = \log\big(\mathrm{P}(\mathrm{O}|\lambda_1)\big), \tag{2.7}$$

$$\mathrm{L}(\mathrm{O}|\lambda_2) = \log\big(\mathrm{P}(\mathrm{O}|\lambda_2)\big). \tag{2.8}$$

The GHMM with a higher likelihood will be selected as the class of this validation sequence. When all validation sequences have been assigned with their predicted labels, the classification accuracy for λ_1 and λ_2 are calculated and recorded.

Besides the model parameter optimization, the influence of the number of states on the classification performance is investigated as well. The training process presented in Figure 2.7 is iterated with up to five hidden states, because increasing this number of hidden states may lead to model overfitting. In the last procedure, the model with the number of states resulting in the highest classification accuracy in each class is chosen as the optimally trained GHMM.

Once two GHMMs are finalized with their optimal number of states, the classification system then moves to the recognition process (Figure 2.6). During this stage, each pressure sequence in the testing set is imported into both best-trained GHMMs. Same as the likelihood evaluation part in the training process, the log likelihood under each GHMM is calculated and compared (Equations 2.7 and 2.8). The model with higher log likelihood will be predicted as the category of the testing surface sequence. At the end of the classification system, the true model classification performance is evaluated and will be discussed below.

CHAPTER 3

RESULTS AND DISCUSSION

This chapter has four main sections. The first section presents the results of surface treating pressure data from 270 simulations, and then compares the surface pressure data of screenout with a field screenout example acquired from the Wattenberg field. I also experiment with different time lengths to investigate the best cut-off time for all simulated data so that their distinct ending trends can be removed. The second section determines the optimal hidden states for both GHMMs from their respective validation set and analyses the performance of the GHMM-based screenout classification system using a confusion matrix approach. The third section presents and discusses the results of testing the screenout classification system based on seven screenout data from the Wattenberg field. The last section demonstrates a misclassified example and discusses the potential reasons for causing these incorrect classifications.

3.1 Surface Treating Pressure Data in 270 Simulations

In the process of synthetic data generation, five fracturing fluids and nine reservoir pressure models are used to model 135 screenout and 135 non-screenout events for a horizontal well with three fracturing stages. According to the simulations that end up with screenouts and non-screenouts, I summarize their respective number for each fracturing fluid in Table 3.1.

As Table 3.1 shows, the simulation case of HEC fracturing fluid results in the most screenouts while Slikcwater_120F generates the fewest screenouts. Although HEC is a linear gel fluid that exhibits a better ability to transport proppants than Slickwater, the results in Table 3.1 can be explained since HEC has the highest proppant concentration (the proppant concentration in all pumping schedule with HEC achieves 7.5 lb/gal in the end) while Slickwater_120F reaches only a max of 5.5 lb/gal for all scenarios.

Table 3.1: Summary of the number of screenout and non-screenout events for all five fracturing fluids.

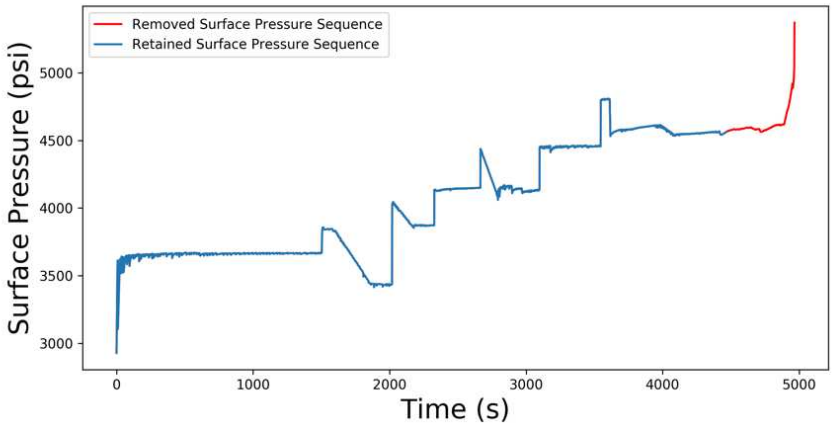
Fluid Name	Number of Screenouts	Number of Non-screenouts
HEC	34	20
Slickwater_150F	29	25
Guar	28	26
HPG	27	27
Slickwater_120F	17	37
Total	135	135

Figure 3.1 presents one representative surface pressure sequence for each class (screenout and non-screenout). I observe that the surface pressure of screenout (Figure 3.1(a)) stops with a rising trend while the non-screenout sequence (Figure 3.1(b)) appears to drop dramatically in the end since the fracturing job is completed. Therefore, it is necessary to remove these obvious trends so that the models can only focus on discovering the pre-screenout patterns. To find the optimal cut-off time, I experiment with different durations, each of which removes the discrepant ending patterns for all simulated data. Then, for each experiment, I iterate the data with the number of hidden states from 1 to 5 and the highest overall classification accuracy is obtained from the hold-out testing set. In Table 3.2, the highest overall classification accuracy is listed for each cut time length. As a result, the final procedure removes the final 500 s (red-colored time sequence in Figure 3.1) in all simulations data since this approach achieves the best overall classification performance.

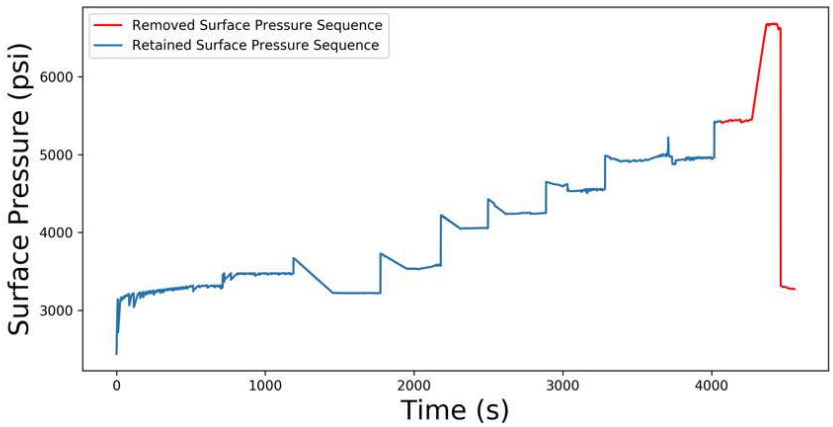
Comparing Figure 3.1(a) and Figure 3.2 illustrates that both field and synthetic data of screenout appear to capture similar pre-screenout patterns: the surface pressure signals hold steady for a period before screenout occurring (Figure 3.1(a) is 1100 s and Figure 3.2

Table 3.2: Classification results using different cut-off time durations.

Time Length (s)	450	500	550	600
Number of Hidden States	2	2	2	2
Overall Classification Accuracy	73%	81%	78%	73%



(a)



(b)

Figure 3.1: Examples of the synthetic surface pressure data from 270 hydraulic fracturing simulations for (a) screenout and (b) non-screenout scenarios. The red-colored data (last 500 s) are removed to eliminate distinct ending patterns for both classes while the blue-colored data are retained for use in training, validation, and testing for the classification system.

is around 850 s).

3.2 Results of the GHMM-based Screenout Classification System

The training-validation-testing procedure uses 188 training, 41 validation, and 41 testing sequences of the surface pressure data. Table 3.3 presents the overall classification accuracy for between 1 to 5 hidden states from the validation data. The optimal number of hidden states for both GHMMs is two, which results in an overall classification accuracy of 75%.

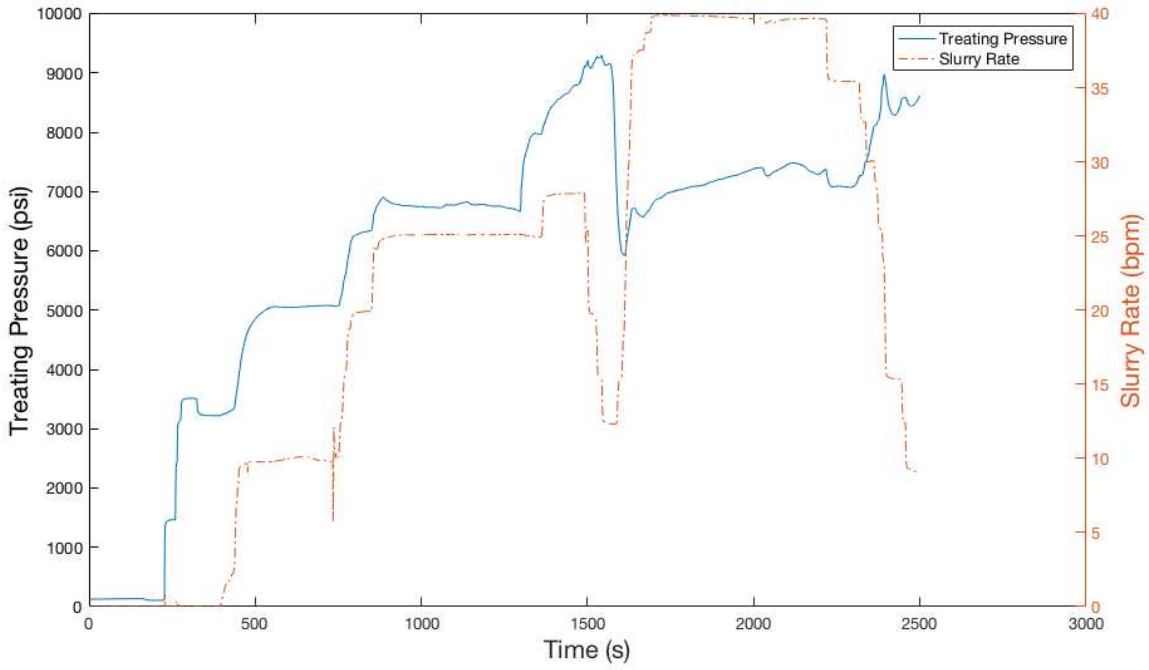


Figure 3.2: A screenout example acquired from the Niobrara formation, Wattenberg field. The surface treating pressure is indicated in blue, while the orange line represents the slurry rate. The screenout occurs at 2400 s.

After determining the optimal number of hidden states for both models, I apply the two GHMMs with two hidden states and initiate my classification workflow on the testing data (Figure 2.6). The final predictive performance of the GHMM-based classification system is evaluated and visualized in the confusion matrix defined in Table 3.4.

Table 3.3: Validation results from using different number of hidden states for both GHMMs.

Number of Hidden States	1	2	3	4	5
Overall Classification Accuracy	61%	75%	41%	50%	54%

Table 3.4: A sample of confusion matrix with detailed descriptions.

True Positive (TP) Correctly predicted screenout on screenout data	**False Positive (FP)** Incorrectly predicted screenout on non-screenout data	**Precision/Positive Prediction Value (PPV)** $=\frac{\text{TP}}{\text{TP+FP}} \times 100\%$
False Negative (FN) Incorrectly predicted non-creenout on screenout data	**True Negative (TN)** Correctly predicted non-screenout on non-screenout data	**Negative Prediction Value (NPV)** $=\frac{\text{TN}}{\text{TN+FN}} \times 100\%$
Overall Accuracy	$\frac{\text{TP+TN}}{\text{TN+FN+TP+TN}} \times 100\%$	

As Table 3.4 shows, each confusion matrix row represents the predicted class while each column represents the actual class. The green and orange shaded cells contain the number of sequences classified correctly and incorrectly by the trained GHMM, respectively. Furthermore, I continue with my evaluation analysis with additional metrics that are commonly used in a classification problem: positive prediction value (PPV) and negative prediction value (NPV). These two metrics are the proportions of the positive and negative results in the confusion matrix that are true positive and true negative results, respectively.

Table 3.5 shows results from my GHMM-based classification system that achieves an 81% overall accuracy with 7% misses and 12% false positive warnings. Also, based on the PPV and NPV calculations, I observe that my classification system can successfully identify 86% of the screenout events and 75% of the non-screenout events in the testing set. Since my training data set is made up equally of non-screenout and screenout events, the positive prediction rate of (86%) in screenout data reveals that the system can efficiently learn the

Table 3.5: Confusion matrix of the GHMM-based classification system.

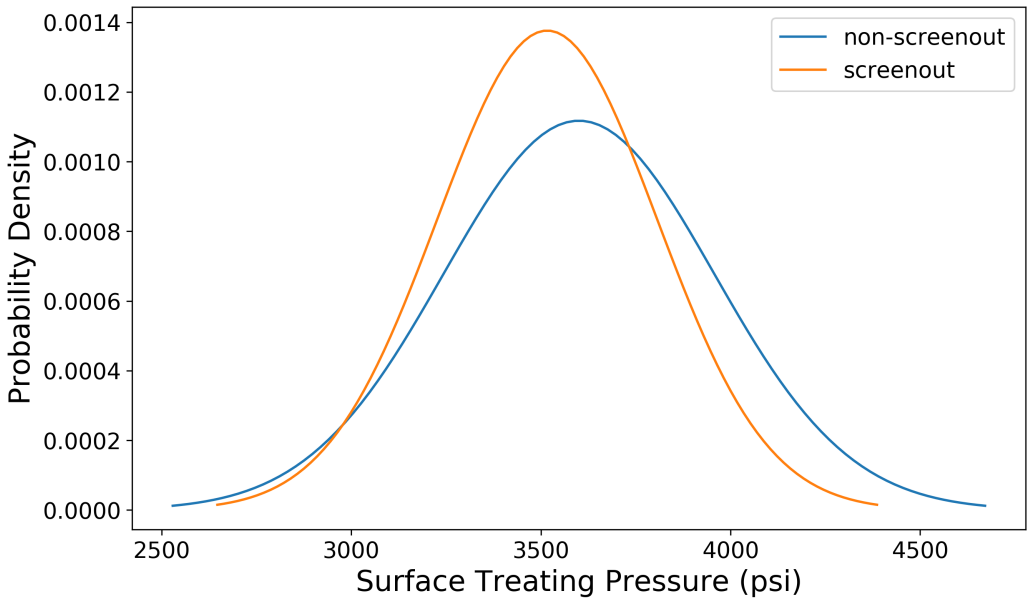
Testing data=41	Predicted Yes	Predicted No	
Actual Yes	18 44%	3 7%	PPV=86%
Actual No	5 12%	15 37%	NPV=75%
Overall Accuracy	81%		

pre-screenout patterns existing in my simulated surface pressure signals and send correct screenout warnings 500 s ahead. Besides, although the NPV states that the system performs less accurately when recognizing non-screenout events (75%), the false warnings could still be beneficial because these alarms could remind the field personnel to pay more attention on monitoring the subsurface status during the operation.

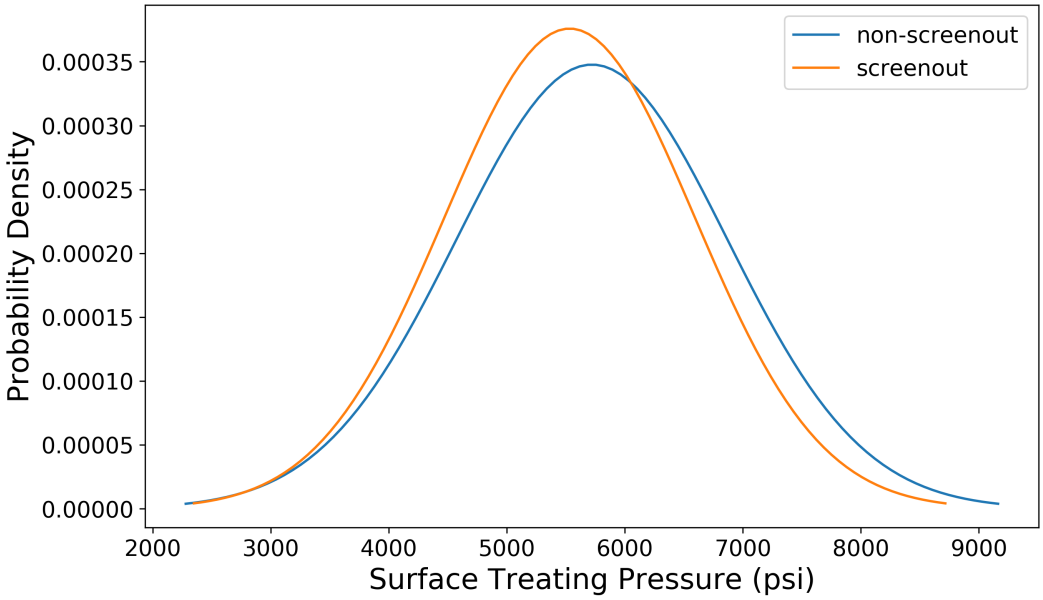
3.3 Discussion of the Misclassified Cases

Figure 3.3(a) and Figure 3.3(b) show the 1D Gaussian PDFs for both screenout and non-screenout models associated with States 1 and 2. As seen from the variance values for both states, both models have learned that the surface treating pressure data have two pressure regimes: low (State 1) and high (State 2) pressure. In addition, I find when both GHMMs estimate an observed sequence at State 2, their Guassian PDFs are very similar to each other.

Figure 3.4 presents a false negative for a true-labeled screenout data. Its hidden path under both models are demonstrated in Figure 3.5(a) and Figure 3.5(b). Before 1000 s in Figure 3.4, the surface pressure sequence (under 4000 psi) is estimated to be at the first state by the screenout model (Figure 3.5(a)) and the second state by the non-screenout



(a)



(b)

Figure 3.3: 1D Gaussian PDFs for screenout and non-screenout GHMM. (a) State 1 and (b) State 2.

model (Figure 3.5(b)), respectively. After 1000 s, the surface pressure sequence (from 4000 to 8000 psi) is estimated to be at the second state by the screenout model and the first state by the non-screenout model, respectively. Since the Gaussian PDFs at the first state have larger probability values than the second state, this leads to that the non-screenout model outputs a greater log likelihood (-37171.73) than the screenout model (-37841.29). Moreover, there is not a big difference between the log likelihood calculated by the screenout and non-screenout model and this could be due to the similar Gaussian PDF shapes for State 2.

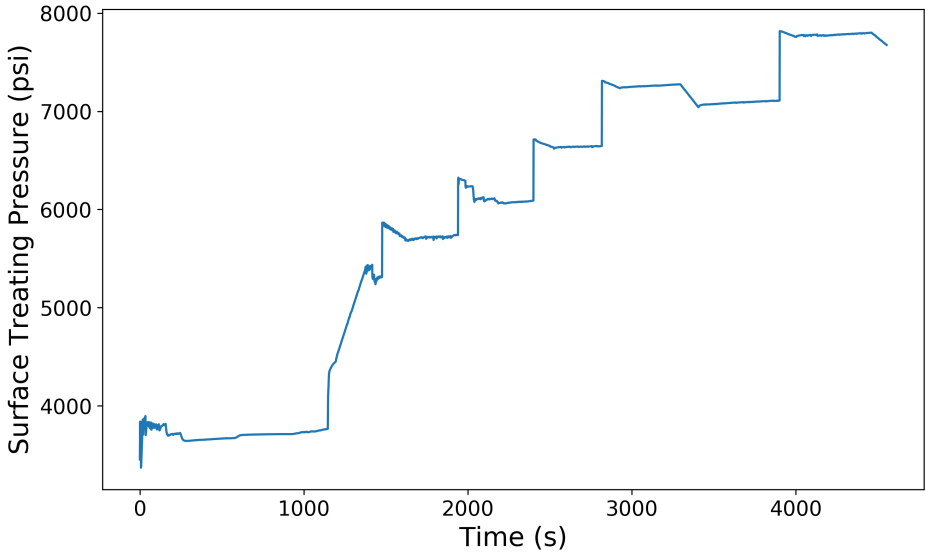
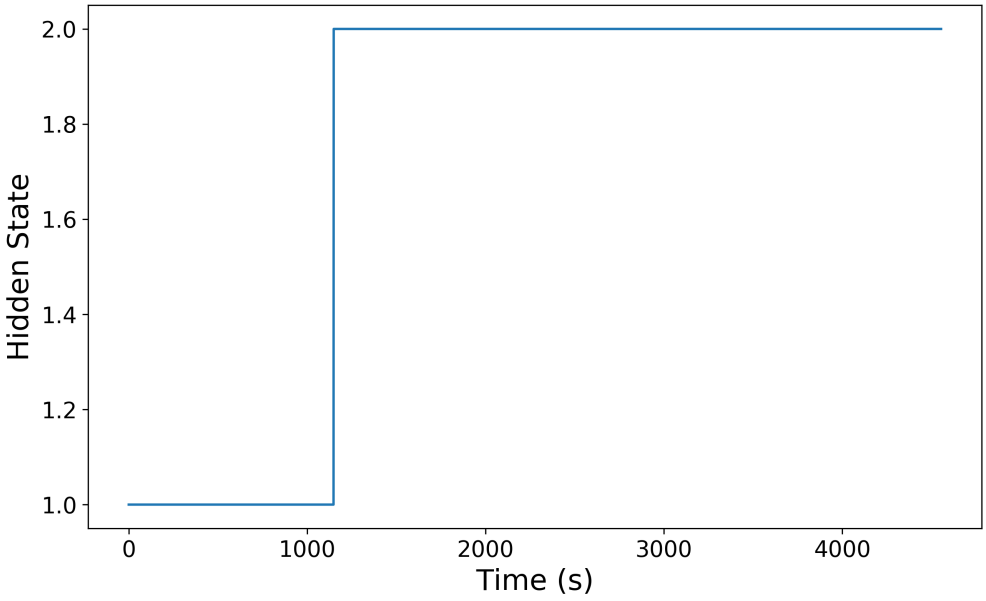
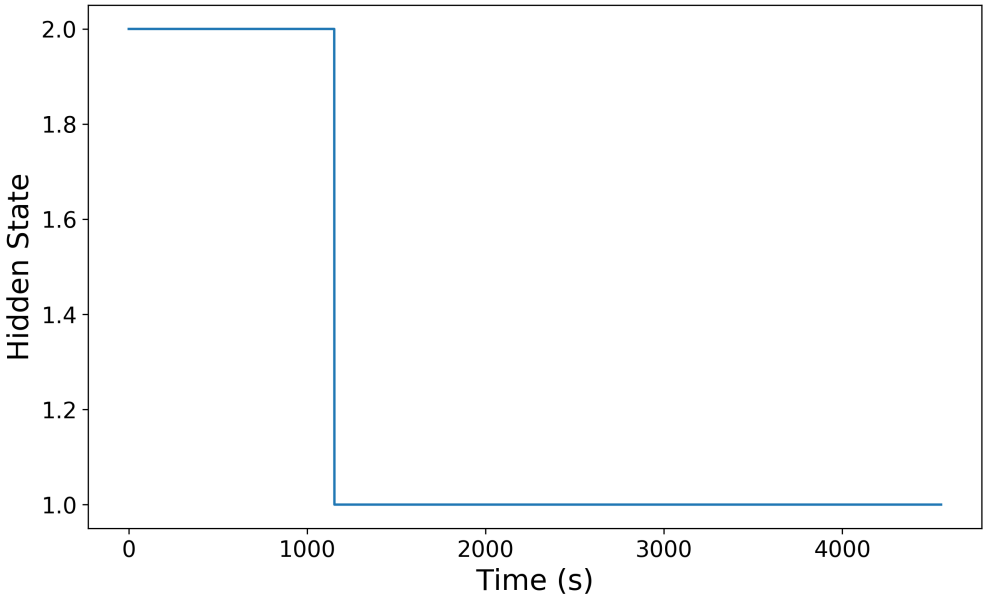


Figure 3.4: A false negative example. The screenout classification system incorrectly predict it as non-screenout.

In addition, since log likelihood is used as the criterion for determining the class for a new observation in the classification system, these false negatives and positives are made because the model of class they are actually labeled outputs smaller likelihood value. As described in Equation 2.7 and 2.8, the log likelihood is computed based on the two well-trained GHMMs (λ_1 and λ_2). In other words, the log likelihood implies how likely these two models (with defined parameters) are going to generate a new given observed sequence. In this project,



(a)



(b)

Figure 3.5: Hidden state path for both GHMMS. (a) screenout and (b) non-screenout.

the 1D Gaussian PDF is one of the model parameters that contribute to the likelihood calculation. Therefore, these misclassified cases may due to incorrect probabilities given by these Gaussian PDFs. For example, some observations from a true screenout sequence are assigned with much larger probabilities by the Gaussian PDFs of non-screenout, which leads to greater likelihood values under the non-screenout model (λ_2). One reason for resulting smaller probabilities by Gaussian PDFs of non-screenout could be that several observations are located at the tail regions of the screenout Gaussian PDFs. As seen in Figure 3.6, the tails of a Gaussian PDF drop off exponentially fast. Although Gaussian distributions are often assumed to represent real-valued random variables such as the surface treating pressure data studied in this project, whose distribution are not known, some observed sequences (misclassified surface treating pressure) could come from another distribution, for instance, t-distribution. Thus, these sequences could have larger probability values when they are drawn from a t-distribution since it has heavier tails than a Gaussian PDF as shown in Figure 3.7.

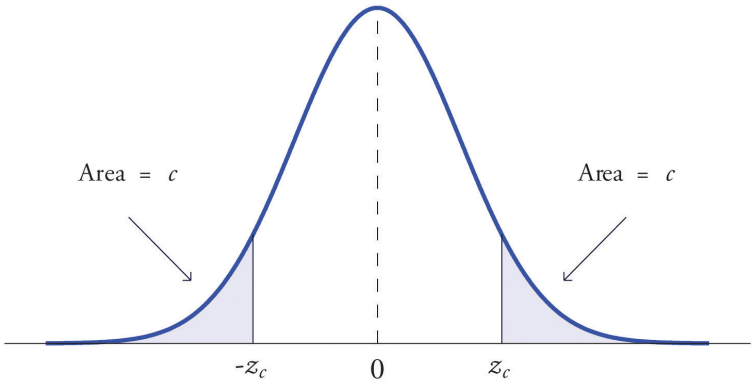


Figure 3.6: Diagram of a Gaussian PDF with a mean of zero. When the random variable is equal to z_c, it cuts off the right tail of the Gaussian PDF, which has an area of c. Similarly, the left tail region (area also equals to c) is formed when the random value equals to $-z_c$ [26].

Moreover, this project only investigates the process of calculating log likelihood under two GHMMs for the screenout detection. Another solution could be to evaluate the posterior probability, which would provide the probability of being screenout and non-screenout given

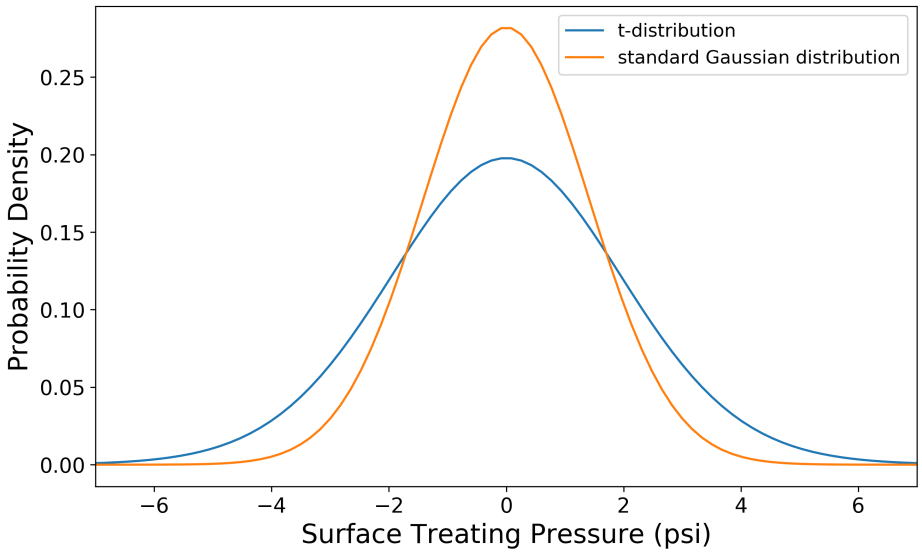


Figure 3.7: An illustration of a standard Gaussian (normal) distrbution with a t-distribution. The t-distribution has heavier tails, which lead to a more spread out distribution compared with the standard Gaussian distribution.

a new observed sequence, respectively. Then, the model with greater probability can be predicted as the class of the new observed sequence. However, with missing prior information, the posterior probability cannot be determined accurately in this study.

3.4 GHMM-based Screenout Classification System on Field Data

The GHMM-based screenout classification system is also tested on the seven field surface treating pressure data sets of screenout of which three are obtained from the Niobrara formation and four are collected from the Codell formation. As for the results, the classification system correctly identifies two Niobrara data sets and two Codell data sets. Figure 3.8 shows a field example of screenout in the Niobrara formation. Using the same pre-processing procedure for the simulated data, the final 500 s prior to screenout occurring are removed for all seven field data sets. Those misses could be due to multiple reasons. First, the classification system is only trained by the surface pressures modeled in the Niobrara formation. Thus, it is reasonable that the model might not recognize the pre-screenout signals presented in

Codell formation. Moreover, all the surface pressure data in this study are synthetic. Therefore, noise in the Niobrara data sets could mislead the model to incorrect classifications. The third reason could be the cut-off time length, which indicates the current 500 s may not be the best duration for the field data.

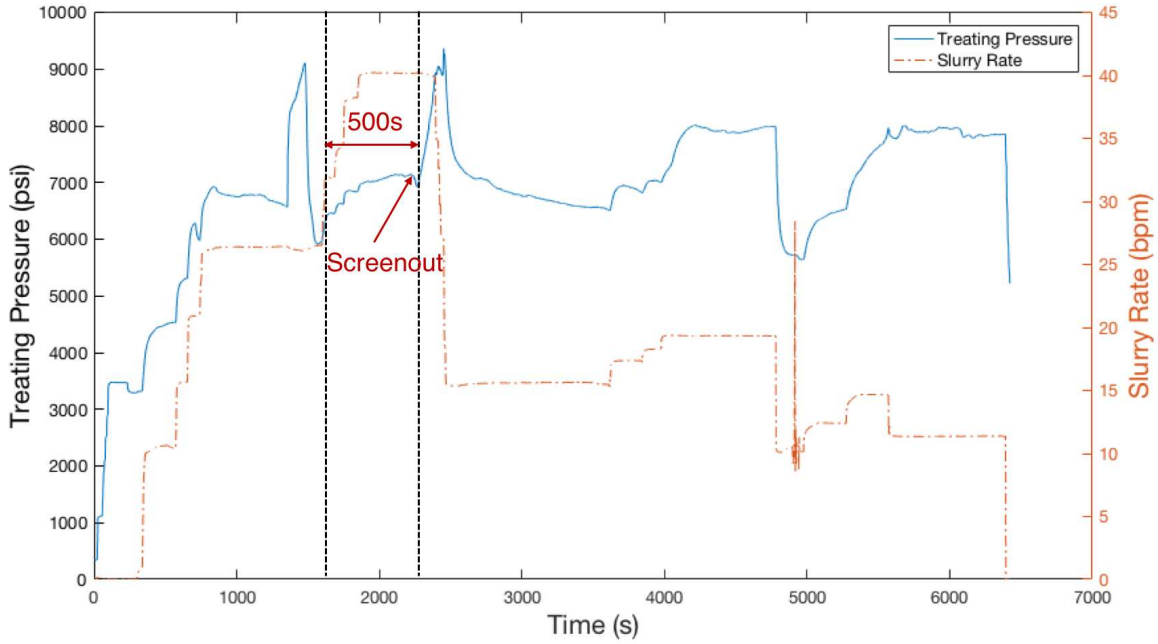


Figure 3.8: A screenout example acquired from the Niobrara formation, Wattenberg field. The screenout occurs when the pumping rate starts decreasing at 2500 s.

Overall, according to the testing statistics of the GHMM-based classification system, its performance shows it can be successfully employed as an effective tool for monitoring screenout events given the model assumptions used in this study.

CHAPTER 4

CONCLUSIONS AND FUTURE WORK

4.1 Conclusions

This thesis develops a screenout classification system based on Gaussian Hidden (GHMM) Markov Model through training two GHMMs. The novelty of such a recognition system is that it can be used as a real-time fracturing diagnostic tool for sreenout detections. More importantly, the system allows us to send 8.5 mins (500 s) of early warning of possible screenout events in surface treating pressure time-series data, which provides sufficient warning for field staff to take remedial actions without shutting down the well. Through an extensive set of experiments, the GHMM has demonstrated its strong learning ability in surface treating pressure time-series analysis with a particular example on specific model assumptions: The GHMM can differentiate non-screenouts and screenouts with an overall classification accuracy of 81% by learning the pre-screenout patterns. The following conclusions are summarized from this research:

- The simulated treating pressure data captured similar pressure behavior as obtained from the Niobrara formation, Wattenberg field.
- The cut-off time of 500 s removes the distinct ending patterns both for screenouts and non-screenouts in surface treating pressure data while keeping the largest information before screenouts, and thus could provide 8.5 mins of early warning.
- The 1D Gaussian probability density function (PDF) is effective as the emission probabilities for representing the randomness of continuous-valued simulated pressure treating data associated with each state.
- The statistics of model performance indicate that the proposed screenout classification system can successfully predict screenouts with a prediction accuracy of 86% from the

testing data.

- The current classification system still has limitations to predict screenout on field data since it is only trained by the synthetic data.

- The misclassified pressure sequences (false negatives and false positives) may due to the similar Gaussian PDFs for State 2 or some of the observations in the sequences are located at the tails of their belonging Gaussian PDFs.

4.2 Future Work

Some future work could be undertaken in order to improve model classification:

- Since the data studied in this thesis are synthetic, it is recommended to acquire more field data (both screenout and non-screenout) from the Niobrara formation and test the performance of the classification system.

- Experiment with another probability density function such as Gamma or t-distributions since they have heavier tails than the Gaussian PDF. Compare the model performance when using different PDFs in the GHMM-based classification system and see if they could achieve less incorrect classifications.

- If the number of the training data is great enough, employ another time-series model such as Long Short-Term Memory (LSTM) and compare the classification performance.

- Further analyze the impact of the cut-off time duration on the model performance with the surface treating pressure data. Correlate the cut-off time length with the pumping rate so that it could vary depending on each input surface pressure sequence.

- Add a procedure of feature extraction such as signal processing to obtain more representative characteristics of the surface treating pressure data. An early experiment of using discrete wavelet transform has been done to extract the features of the simulated surface pressure signals in four different frequency bands (see Appendix A). All four

signal features have been input as the substitute of the original simulated pressure data into the GHMM-based classification system. However, their classification performances are not as high as the original data.

- Develop a multivariate GHMM that includes additional attributes/variables, which are the natural responses from the reservoir into the model such as temperature data. By doing this, the model could correlate multiple inputs that represent subsurface conditions, and then better describe the differences between screenout and non-screenout.

REFERENCES CITED

[1] Charlie Smith, Lynda Ziane, et al. Effective fracture treatment determination in unconventional reservoirs. In *SPE Asia Pacific Oil and Gas Conference and Exhibition*. Society of Petroleum Engineers, 2012. doi: 10.2118/158349-MS.

[2] PJ Hudson, RP Matson, et al. Hydraulic fracturing in horizontal wellbores. In *Permian Basin Oil and Gas Recovery Conference*. Society of Petroleum Engineers, 1992. doi: 10.2118/23950-MS.

[3] Leon V Massaras, Timothy R McNealy, et al. Highly accurate prediction of screenouts in the eagle ford shale with the screenout index. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2012. doi: 10.2118/157613-MS.

[4] Introduction to well control for horizontal wells, 2014. https://www.drillingformulas.com/introduction-to-well-control-for-horizontal-wells.html. Accessed: 27 November 2019.

[5] John L. Gidley. *Recent advances in hydraulic fracturing*. Monograph ; v. 12. Henry L. Doherty series. Henry L. Doherty Memorial Fund of AIME, Society of Petroleum Engineers, Richardson, TX, 1989. ISBN 1555630200.

[6] Charles W. Zuppann and John C. Steinmetz. Hydraulic fracturing: an Indiana assessment, 2014. https://igws.indiana.edu/OilGas/HydraulicFracturing.html. Accessed: 17 October 2019.

[7] S Bagci, L Castro, J Flores, et al. Optimization of hydraulic fracturing and production enhancement: Case studies for us shale plays and tight sand reservoirs. In *SPE Europec featured at 79th EAGE Conference and Exhibition*. 2017. doi: 10.2118/185773-MS.

[8] Gun-Ho Kim. Interpretation of hydraulic fracturing pressure in lowpermeability gas reservoirs. Master's thesis, The Pennsylvania State University, 2010.

[9] Kenneth G Nolte, Michael B Smith, et al. Interpretation of fracturing pressures. *Journal of Petroleum Technology*, 33(09):1–767, 1981.

[10] RP Nordgren et al. Propagation of a vertical hydraulic fracture. *Society of Petroleum Engineers Journal*, 12(04):306–314, 1972.

[11] RD Barree. Stress shadowing and fracture interference in gohfer. *Barree & Associates*, 2015.

[12] Yuanhai Yang, Guonong Hu, Anne Kremer, et al. Using prefrac test information to predict and avoid screenout associated with slickwater frac in tight gas sands at the wattenberg field in the denver-julesburg basin. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2008.

[13] Leon V Massaras, Dimitri V Massaras, et al. Real-time screenout advanced warning with the inverse slope method. In *SPE International Symposium and Exhibition on Formation Damage Control*. Society of Petroleum Engineers, 2012. doi: 10.2118/150263-MS.

[14] Leonard E Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state markov chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563, 1966.

[15] QJ Zhu, FX Jiang, ZX Yu, ZL Han, SQ Li, and PX Zhang. Application of hidden markov model for mining microseimic signals classification. *Transit Development in Rock Mechanics: Recognition, Thinking and Innovation*, page 211, 2014.

[16] Asjad Amin, Mohamed Deriche, and Bo Liu. A novel approach for salt dome detection in seismic surveys using a hidden markov model. In *SEG Technical Program Expanded Abstracts 2016*, pages 1688–1692. Society of Exploration Geophysicists, 2016.

[17] Po-Yen Wu, Vikas Jain, Mandar S Kulkarni, and Aria Abubakar. Machine learning-based method for automated well-log processing and interpretation. In *SEG Technical Program Expanded Abstracts 2018*, pages 2041–2045. Society of Exploration Geophysicists, 2018.

[18] V Matthews. Colorados new oil boomthe niobrara. *CGS RockTalk*, 13(1), 2011.

[19] James G. Speight. Chapter three - gas and oil resources in tight formations. In James G. Speight, editor, *Deep Shale Oil and Gas*, pages 121 – 174. Gulf Professional Publishing, Boston, 2017. ISBN 978-0-12-803097-4. doi: https://doi.org/10.1016/B978-0-12-803097-4.00003-6.

[20] Stephen A Sonnenberg. The niobrara petroleum system, rocky mountain region. *Search and Discovery*, 80206:107–112, 2012.

[21] Jessica G Iriarte. Fluid-shale-proppant interactions and the degradation of hydraulic fracture conductivity in the niobrara formation. Master's thesis, Colorado School of Mines. Arthur Lakes Library, 2017.

[22] Zoubin Ghahramani. An introduction to hidden markov models and bayesian networks. In *Hidden Markov Models: Applications in Computer Vision*, pages 9–41. World Scientific, 2001.

[23] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[24] Leonard E Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical statistics*, 41(1):164–171, 1970.

[25] Yingjian Zhang. Prediction of financial time series with hidden markov models. Master's thesis, Simon Fraser University, 2004.

[26] Barbara Illowsky and Susan Dean. Introductory statistics. *OpenStax College, Houston*, 2015.

[27] Robi Polikar. The story of wavelets. *Physics and modern topics in mechanical and electrical engineering*, pages 192–197, 1999.

[28] Liu Chun-Lin. A tutorial of the wavelet transform. *NTUEE, Taiwan*, 2010.

[29] Rogelio Ramos, Benjamin Valdez-Salas, Roumen Zlatev, Michael Schorr Wiener, and Jose Mara Bastidas Rull. The discrete wavelet transform and its application for noise removal in localized corrosion measurements. *International Journal of Corrosion*, 2017: 1–7, 2017. ISSN 1687-9325.

[30] DE Rivera-Recillas, M Lozada-Zumaeta, JO Campos-Enriquez, and G Ronquillo-Jarillo. Calculation of seismic attributes with the discrete wavelet transform. In *SEG Technical Program Expanded Abstracts 2003*, pages 2028–2031. Society of Exploration Geophysicists, 2003.

[31] Jiang Li and Tsili Wang. Effective logging data compression using the discrete wavelet transform. In *SEG Technical Program Expanded Abstracts 2003*, pages 313–316. Society of Exploration Geophysicists, 2003.

[32] Werner Heisenberg. Multi-body problem and resonance in quantum mechanics ii. *Z Phys*, 41:239, 1927.

[33] E Unal, F Siddiqui, MY Soliman, et al. Wavelet analysis of fracturing pressure data. In *SPE Hydraulic Fracturing Technology Conference and Exhibition*. Society of Petroleum Engineers, 2018.

APPENDIX
DISCRETE WAVELET TRANSFORM

This appendix investigates an approach that implements a discrete wavelet transform to extract features of the simulated surface pressure signals.

A.1 Background of Wavelet Transform

The concept of the wavelet was first introduced in 1982 by Jean Morlet, a geophysicist, to solve a problem in processing seismic signals which had very high-frequency components with a short period, and low-frequency components with a long time period [27]. Afterward, Morlet formalized the wavelet transform with the physicist Alex Grossman in 1984, and then more researchers devoted to this field such as Yves Meyer, a Mathematician, who developed orthogonal wavelet basis functions that have better localization both in frequency and time [28]. Nowadays, the wavelet transform has been widely applied in the signal processing field for different kinds of stationary and non-stationary signal analysis with the purposes of electrical noises removal, detection of abrupt discontinuities, and compression of a great amount of data [29]. In oil and gas industry, [30] used a discrete wavelet transform to successfully obtain seismic attributes such as amplitude, phase, and instantaneous frequency from synthetic and real data. [31] described a compression methodology based on wavelet analysis for well-logging data transmission.

A.2 Definition of Wavelet Transform

Unlike the Fourier transform that decomposes the signals into a sum of an infinite series of sinusoidal basis functions, the wavelet transform approximates the signals by a sum of wavelets which have finite lengths (called compact support). Equation A.1 and Equation A.2 present the formulas of the Fourier transform and the wavelet transform. Similar to the Fourier Transform that multiplies a signal, $x(t)$ which is a function of time, by an

analyzing function into sines and cosines, the wavelet transform multiplies a signal by a wavelet analyzing function, which is localized in both the real and Fourier space.

$$\mathrm{X(F)} = \int_{-\infty}^{\infty} \mathrm{x(t)}\ \mathrm{e}^{-\mathrm{j}2\pi\mathrm{Ft}}\ \mathrm{dt}, \tag{A.1}$$

$$. \ \mathrm{X(a,b)} = \int_{-\infty}^{\infty} \mathrm{x(t)}\ \psi_{\mathrm{a,b}}(\mathrm{t})\ \mathrm{dt}. \tag{A.2}$$

In addition, as seen in Equation A.1 and A.2, the Fourier transform outputs coefficients as a function of frequency (F) and the wavelet transform outputs a two by two matrix of coefficients which is related to translation and scale (a and b). The term translation refers to shifting the wavelet forward in time as the entire signal is being analyzed. The term of scale simply means to what extent the wavelet is stretched. For example, a low scale wavelet is equivalent to a "compressed" wavelet which has better localization in time than a "stretched" (high scale) wavelet. The advantage of utilizing wavelet transform than the Fourier transform is that it can be used to solve problems of frequency and time domain resolution. According to the Heisenberg uncertainty principle, it is impossible to know simultaneously the precise location in time and frequency [32]. To better explain this concept, Figure A.1 shows a diagram of a Heisenberg box in a time-frequency plane based on the wavelet transform. When the high resolution in time is achieved, the tradeoff is low resolution in frequency space and vice versa. In a nutshell, the wavelet transform works as much the same as the Fourier transform, the only difference is that the wavelet transform tries to find the correlation between a signal and an analyzing function which is a wavelet so that the signal can be localized in time and frequency accurately.

A.3 Why Discrete Wavelet Transform?

In this project, all simulated surface treating pressure data are discrete, so it is more suitable to implement a discrete wavelet transform rather than a continuous wavelet transform.

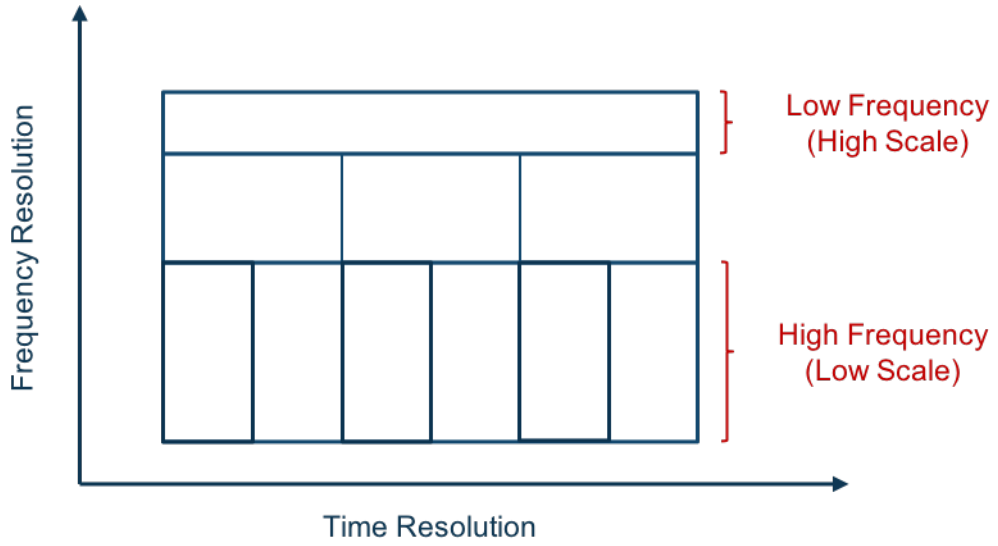


Figure A.1: An illustration of a Heisenberg box based on the wavelet transform in a time-frequency plane. Each box represents a wavelet coefficient. As approaching to high scale wavelets, more resolution in frequency and less resolution in time are obtained.

A.4 Definition of Discrete Wavelet Transform

In definition, the discrete wavelet transform decomposes a signal into a discrete set of wavelets which are orthogonal to their scales and translations [33]. Conceptually, given a signal, a wavelet (high pass) filter and a scaling (low pass) filter are implemented to have high-frequency and low-frequency components of the signal. Such filters have a small number of coefficients that result in good computation performance. In addition, the coefficients contained by high-frequency components are named as detail while the coefficients obtained by low-frequency components are called approximation. These two terms can be calculated in the following formulas:

$$C_{Dj} = \sum_{t=0}^{\infty} x(t)\ \psi_{j,k}(t), \tag{A.3}$$

$$C_{Aj} = \sum_{t=0}^{\infty} x(t)\ \phi_{j,k}(t), \tag{A.4}$$

where $\psi_{j,k}(t)$ is a wavelet function with a translation parameter k at decomposition level j and $\phi_{j,k}(t)$ is a scaling function with a translation parameter k at decomposition level j. During the discrete wavelet transform, the detail coefficients capture the discontinuities and singularities of the signals well and the approximation information presents very similar behaviors as the original signal [33]. To gain more information in the detail coefficients, it is desirable to decompose the original signal into different frequency bands (levels). As seen in Figure A.2, a three-level wavelet decomposition tree for a surface treating pressure signal is presented.

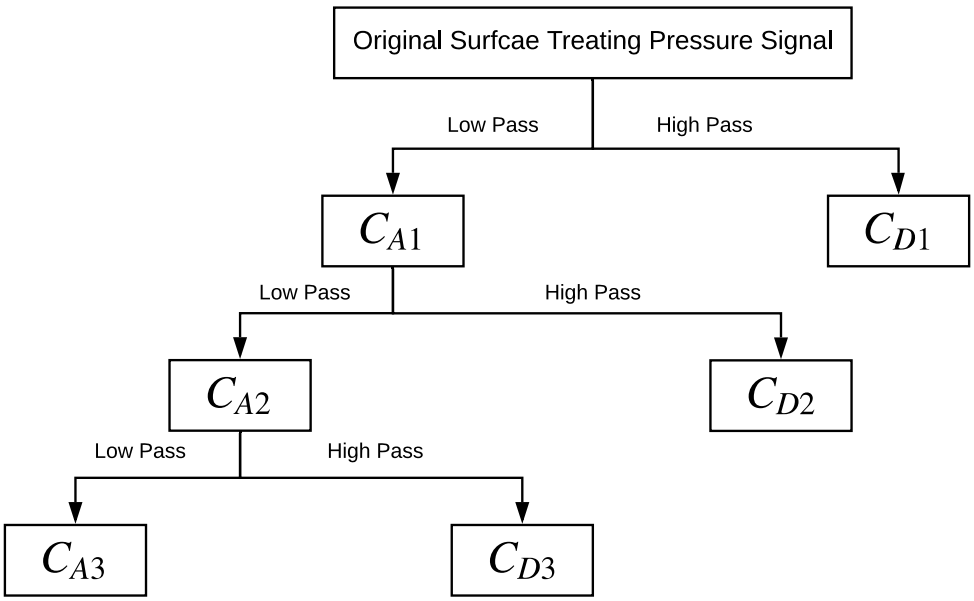


Figure A.2: A tree diagram of a three-level discrete wavelet decomposition. C_A is the approximation coefficient and C_D is the detail coefficient.

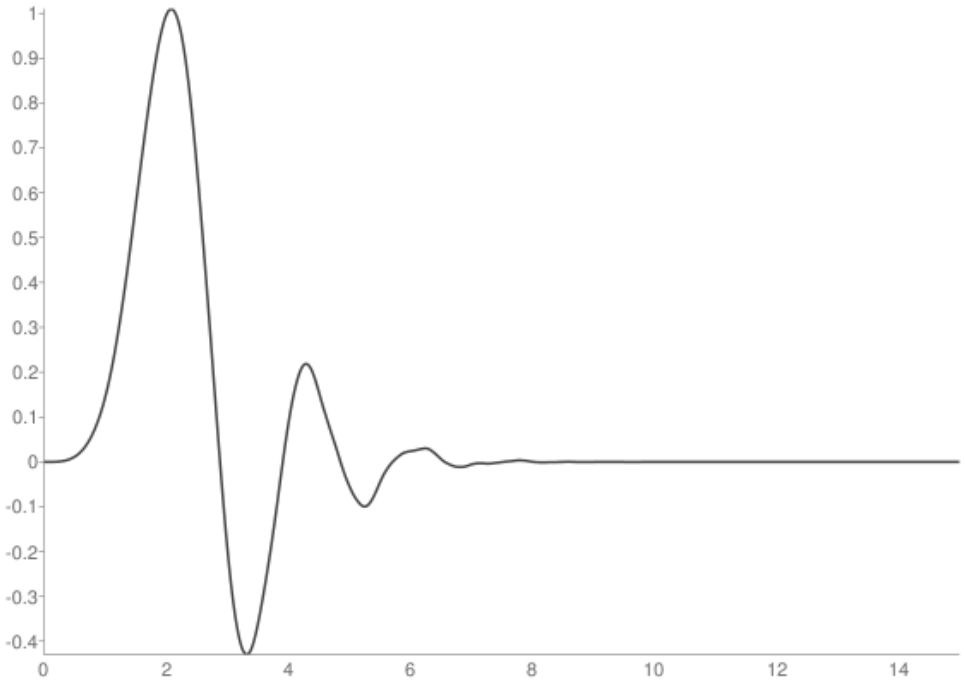
A.5 Discrete Wavelet Transform for Simulated Surface Treating Pressure Data

In this thesis, a three-level discrete wavelet transform (decomposition) is performed for all simulated data. This process is done before the two GHMMs are trained. The Daubechie (db) 8 wavelet, an orthogonal wavelet from Daubechies family, is used as the analyzing wavelet of

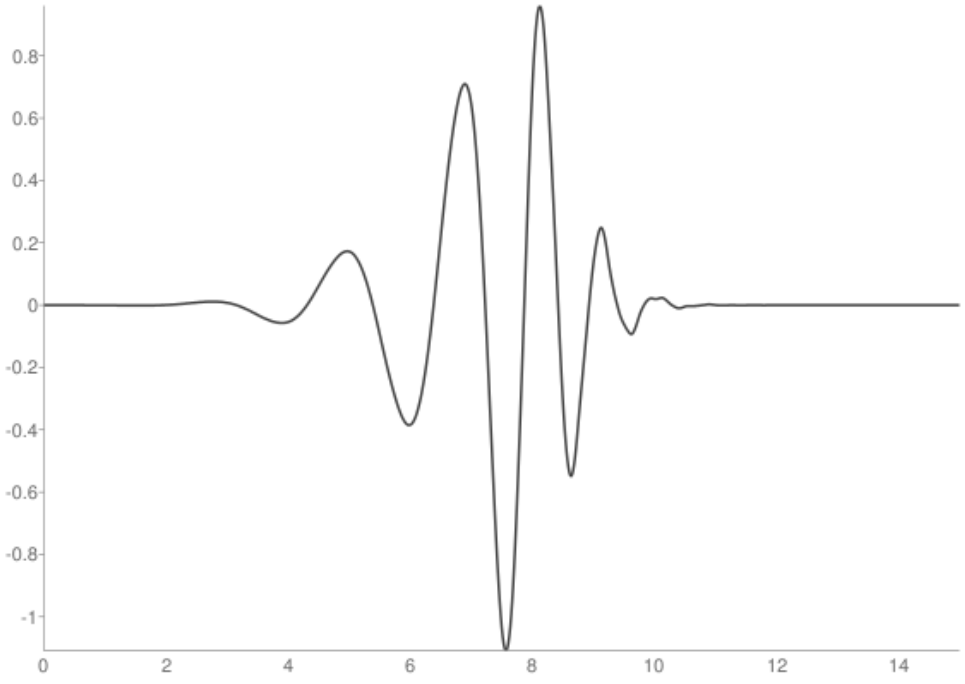
the discrete wavelet transform. Figure A.3(a) and Figure A.3(b) present the scaling function and wavelet function of a db8 wavelet. Figure A.4(a) and Figure A.3(b) show examples of a three-level wavelet decomposition for a screenout and a non-screenout surface treating pressure signal, respectively. From the figures, I observe that the screenout pressure signal obtains more peaks in its all three detail coefficients before screenout occurring, which could be a good feature to be included in the GHMM-based classification system.

A.6 Summary and Recommendations

In conclusion, the discrete wavelet transform could be utilized as a digital signal processing tool to extract features from the surface pressure signals. However, in this project, I only develop a univariate GHMM, which only allows 1D data as the input. Therefore, when I input each of these four decomposed signals into the GHMM-based classification system and their performances are far below as inputting the original data. It is due to one frequency band (level) that has limited ability to represent the information contained in the entire original surface pressure. In the future, it is recommended to propose a multivariate GHMM so that all of the decomposed levels can be input as features into the model.

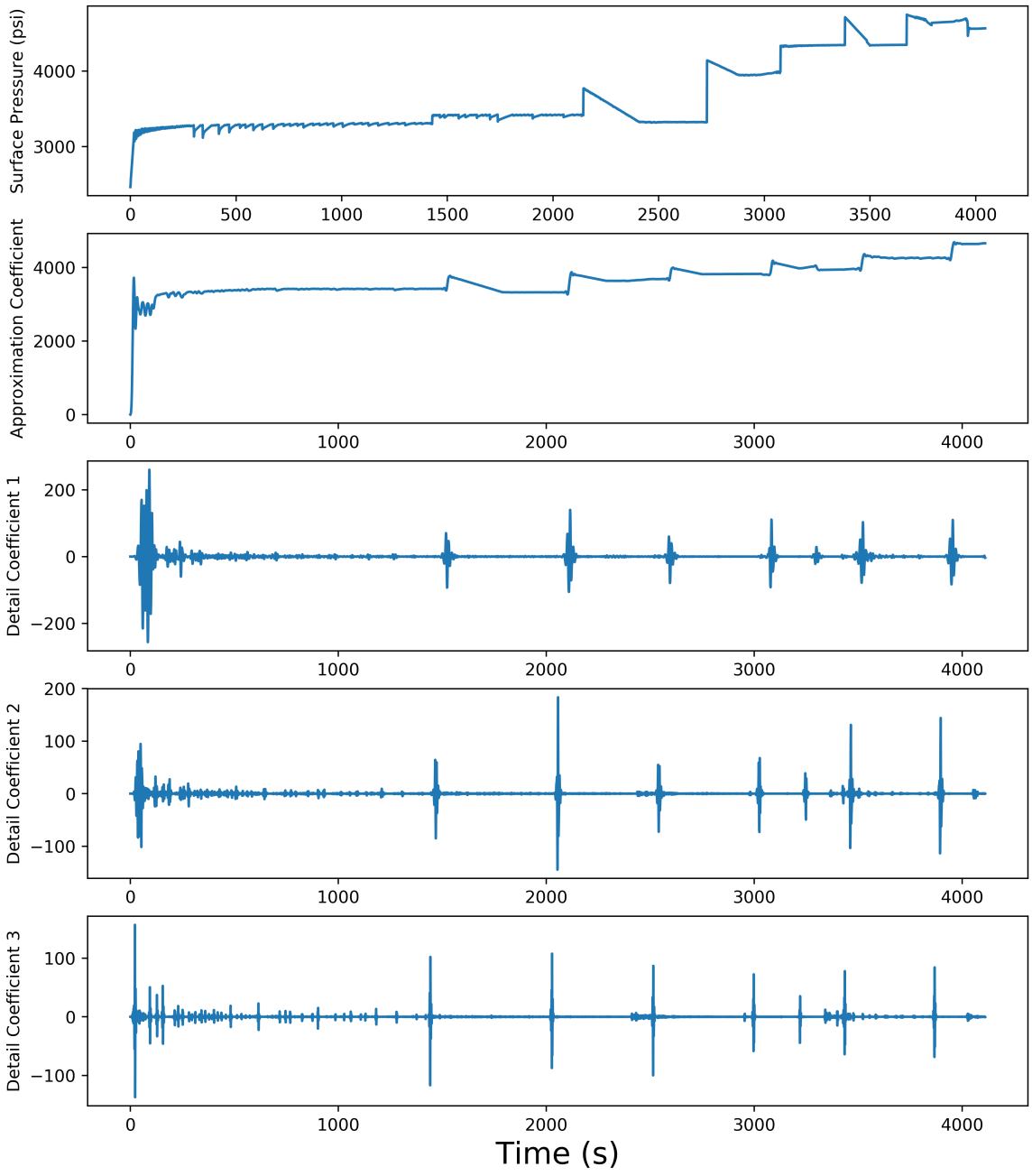


(a)

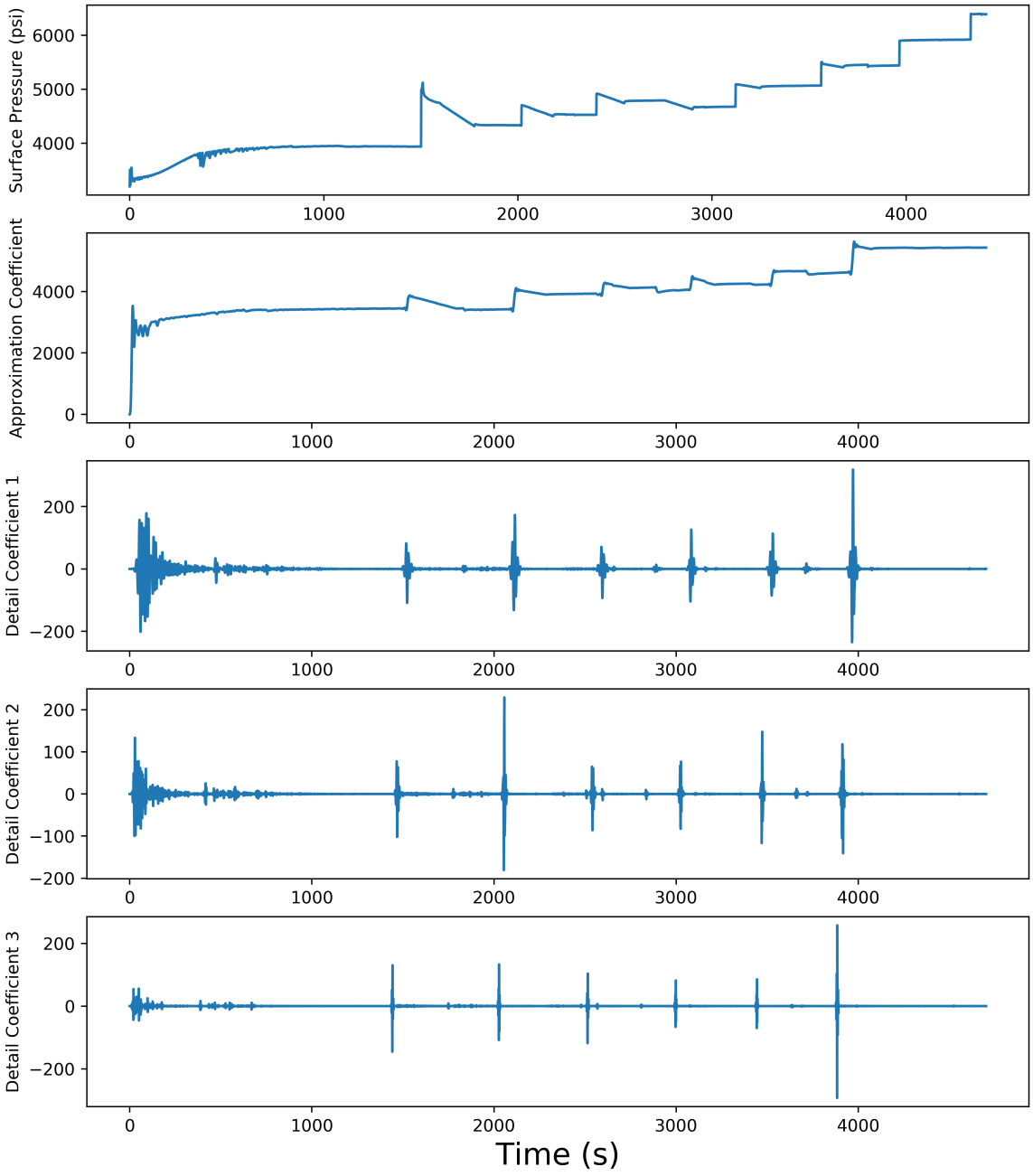


(b)

Figure A.3: Diagrams of (a) scaling function and (b) translating function of a db8 wavelet. They are applied in Equation A.3 and Equation A.4 to compute the detail coefficients and approximation coefficients at each decomposition level, respectively.


Surface Pressure (psi)
4000
3000
0
500
1000
1500
2000
2500
3000
3500
4000
Approximation Coefficient
4000
2000
0
0
1000
2000
3000
4000
Detail Coefficient 1
200
0
−200
0
1000
2000
3000
4000
Detail Coefficient 2
200
100
0
−100
0
1000
2000
3000
4000
Detail Coefficient 3
100
0
−100
0
1000
2000
3000
4000
Time (s)

(a)



(b)

Figure A.3: Decomposition of a surface treating pressure signal at level 3. (a) screenout and (b) non-screenout. The peaks represent the discontinuities within the original signal.