Model Predictive Automatic Control of Sucker Rod Pump System with Simulation Case Study

Brigham Hansen\textsuperscript{a}, Brandon Tolbert\textsuperscript{b}, Cory Vernon\textsuperscript{a}, John D. Hedengren\textsuperscript{a,*}

\textsuperscript{a}Chemical Engineering Department, Brigham Young University, Provo, UT 84602, USA
\textsuperscript{b}University of Oklahoma, Norman, OK 73019, USA

Abstract

This work enables accelerated fluid recovery in oil and gas reservoirs by automatically controlling fluid height and bottomhole pressure in wells. Several literature studies show significant increase in recovered oil by determining a target bottomhole pressure but rarely consider how to control to that value. This work enables those benefits by maintaining bottomhole pressure or fluid height. Moving Horizon Estimation (MHE) determines uncertain well parameters using only common surface measurements. A Model Predictive Controller (MPC) adjusts the stroking speed of a sucker rod pump to maintain fluid height. Pump boundary conditions are simulated with Mathematical Programs with Complementarity Constraints (MPCCs) and a nonlinear programming solver finds a solution in near real-time. A combined rod string, well, and reservoir model simulate dynamic well conditions, and are formulated for simultaneous optimization by large-scale solvers. MPC increases cumulative oil production vs. conventional pump off control by maintaining an optimal fluid level height.

Keywords: sucker rod pump, model predictive control, optimization, reservoir modeling

1. Introduction

Sucker-rod pumping is a widely used artificial lift method for extraction of oil and gas resources. In sucker-rod pumping, a pump at the bottom of the

*Corresponding author. Tel.: +1 801 477 7341
Email address: john.hedengren@byu.edu (John D. Hedengren)
well is actuated by the linear up and down stroking of a surface unit via a rod string. Sucker-rod artificial lift systems are often over-sized after partially depleting the reservoir pressure. Pumping a well at a fluid production rate that exceeds the reservoir’s maximum fluid influx leads to a condition called pump-off. Pump-off leads to operating inefficiency and equipment damage. Equipment stress occurs when the pump plunger strikes liquid standing in the pump barrel during the down stroke. This condition is called fluid pound. To prevent both inefficient fluid production and equipment damage many methods have been developed to detect pump-off and control the surface unit. With few exceptions these methods are reactive controllers that either control motor speed or simply shutdown the unit for some predetermined time interval. This work proposes an advanced control system that utilizes a novel combination of sucker rod pump, well, and reservoir models. The system as a whole has both slow and fast dynamics, similar to traditional chemical processes. The mass balance of fluid in the annular space between the well casing and production tubing has relatively fast dynamics. Changes to reservoir conditions and reservoir fluid influx, typically, are much slower. Pump boundary conditions are formulated as Mathematical Programs with Complementarity Constraints (MPCCs) to allow for simulation and optimization using simultaneous methods with large-scale solvers. Moving horizon estimation estimates uncertain parameters for control. An advanced controller reduces equipment damage and maximizes fluid production.

1.1. Rod Pump Controller

Westerman [62] summarizes many traditional methods for detecting pump-off using a variety of sensors to measure fluid level, flow, vibration, motor current, and rod loading. Gibbs et al. [23] presents several methods for detecting pump-off conditions using pump dynamometer cards. It is recommended that several methods be used simultaneously to assure proper controller response. Pattern recognition is also used to diagnose pumping conditions for control [54, 58, 50, 64]. Pattern recognition schemes require training with data sets.

In addition to traditional pump-off controllers, recent work uses a variety of methods to control/diagnose pumping performance with many reported benefits. Ghareeb et al. [21] show significant production increases by using an automatic pump controller that utilizes a downhole pressure sensing device and automatic control algorithm. Ahmed and Nabil [1] show that smart rod-pump controllers can be used to accurately infer production rates and
calculate bottomhole pressures in close agreement with traditional measure-
ment techniques. Ehimeakhe [18] develops an algorithm comprised of four
methods for accurately determining pump fillage from pump dynamometer
cards to allow for control of pumping units. A case study shows a propri-
etary algorithm that controls stroking speed to minimize equipment damage
and maximize production in horizontal well applications where slugging is
common [19]. Other methods optimize the production from unconventional
oil and gas wells using sucker-rod lift [44, 41]. Sanchez et al. [52] show signif-
icant benefits of automatically controlling rod-pump systems. These benefits
include energy savings, optimized production, expense reduction, and better
manpower utilization.

Often, producing oil fields are equipped with supervisory control and
data acquisition systems (SCADA) that collect data from individual wells
to a central database. These systems are often referred to as expert sys-
tems because the computer program diagnoses, recommends, and in some
cases, automatically implements the course of action that an expert in the
field of artificial lift would recommend [13, 38, 58, 57, 24, 40]. The goal of
these systems is often to do some or all of the following: recommend optimal
system design (such as equipment selection and sizing), implement pump-off
controls, and finally, predict equipment failure and maintenance resource al-
location. Expert systems are frequently a rule-based chain of logic leading to
a recommended course of action [13]. Artificial neural networks (ANN) and
genetic algorithms (GA) are also used in expert systems to diagnose pumping
conditions and determine pump configurations, respectively [38]. SCADA
systems include closed loop adjustments of manipulated variables, such as
strokes per minute (SPM) and injection rates. They also compute downhole
dynamometer cards to diagnose pumping conditions and recommend system
design [58]. Vazquez and Fernandes [57] perform computer optimization of a
system model using the wave equation to determine optimal SPM. Optimiza-
tion of motor rotations per minute throughout a single pump stroke is also
performed with the objective to maximize production subject to equipment
maximum loading constraints [45]. Application of automatic stroking speed
controllers is shown to maximize production [48].

The literature demonstrates that implementing automatic control systems
for sucker rod pumps is beneficial to production. Despite the development
of many first principles models for sucker rod pumping systems, they are
rarely used directly in rod pump controllers. The two applications the au-
thors are aware of are the following: 1) Vazquez and Fernandes [57] use a
wave equation simulation to optimally set surface unit SPM, and 2) Pałaśka and Czyż [15] use a similar wave equation simulation to optimally determine motor RPM throughout a single stroke to maximize SPM subject to maximum loading constraints. Both of these works use shooting methods with sequential simulation. This work will expand the application of first principles modeling in rod pumping controllers by combining rod string, well, and reservoir models, and by posing them in a form suitable for simultaneous simulation and optimization by large-scale solvers.

1.2. Sucker Rod Pump and Well Modeling

To enable predictive analysis of rod pumping systems Gibbs [22], simulates the behavior of the rod string systems using the 1-D wave equation with viscous damping. Surface unit kinematics and proposed pump boundary conditions allow the simulation of various pumping conditions. The model is solved using a finite difference solution method. Doty and Schmidt [16] improve the simulation of sucker rod systems by including the effects of liquid inertia and viscosity which they show to have a significant effect on pump performance. Wang and Liu [61] show improved well modeling by including viscous friction, tubing friction, liquid inertia, and plunger barrel friction. This study will build on the model proposed by Gibbs [22]. Influx of reservoir fluids and height of fluid in the well annulus will be dynamically considered to simulate the changing production conditions over time. A summary of the dynamics considered in this study is shown in Figure 1. Kinematic equations describe the motion of the sucker rod pump at the surface as a function of motor torque input. A wave equation describes the force and position propagation through the sucker rod string. A material balance on the annular fluid determines the annular fluid height. Finally, a material balance reservoir model describes the influx of fluids into the wellbore from the reservoir.

1.3. Reservoir Modeling for Well Control

This section reviews reservoir modeling with emphasis on well control. Well control technologies are currently generating significant interest in the oil and gas industry primarily because estimations demonstrate that they can increase estimated ultimate recovery by 10-15% [31]. Further, as the speed of computers and data acquisition systems continues to improve, the industry is anticipating that control technologies will become a more integral part of the life cycle optimization process of a well as explained in Jansen et al.
To adopt an optimum production strategy, it is necessary to consider a closed loop control process. Most conventional reservoir simulators use a history-matching process to align model values with production data from a real system [31]. The history matching process typically involves the manual adjustment of model parameters over a period of years. Some drawbacks of this approach include (1) manual adjustment of model parameters, (2) violation of system constraints, and (3) overfitting of real-time data [10]. To help optimize reservoir performance under geologic uncertainties and incorporate real-time parameter updating of changing reservoir parameters, the oil and gas industry has placed an emphasis on exploring smart reservoir modeling. Like history matching, smart reservoir modeling is a closed loop process. However, it has the added benefit of (1) systematic updating using real-time data, (2) the ability to minimize the effect of bad data, and (3) reduce uncertainty [31]. Smart well modeling may also include additional constraints and data cleansing. Grbovic et al. [25] recommend developing system models that decrease dependence on historical data by including additional system constraints. Deshpande et al. [14] show that model based data cleansing methods significantly improves data quality. Udy et al. [56] reviewed prior applications of reservoir modeling for well control. The benefits of combining traditionally separate optimizations are highlighted. Brouwer and Jansen [11] applied optimal well control to adjust downhole valve settings of horizontal injector wells to optimize production. The results from his simulations suggested the cumulative recovery could be improved by 20% by implementing optimal control schemes. Sarma and Chen [53] implemented optimization control strategies to control bottomhole pressures of reservoir models depicting offshore wells. The optimal control algorithm increased NPV by 17% over an 8 year simulation period. Simulations confirm that smart reservoir modeling is superior to traditional methods. This study considers optimal control schemes to control the fluid level in the annulus of a rod pumped well while obeying certain constraints.

Literature studies show significant increases in ultimate oil recovery and project NPV as a result of optimally controlling producing well bottomhole pressure [12, 3, 27]. However, these researchers rarely consider how the well bottomhole pressure will be maintained at optimal values. In rod pumped wells, the bottomhole pressure is largely dictated by the height of the fluid column in the well annulus [51]. This research demonstrates a Moving Horizon Estimation (MHE) to determine fluid level from common measurements, and presents several controllers for controlling annular fluid height. This re-
search enables the benefits shown by others in simulation by demonstrating automatic control of bottomhole pressure.

Figure 1: Summary of the rod string, well, and reservoir dynamics

1.4. Nomenclature

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Value or Range</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>drainage area</td>
<td>2 acres</td>
<td>constant</td>
</tr>
<tr>
<td>$A_f$</td>
<td>cross sectional area of annular fluid</td>
<td>$ft^2$</td>
<td>constant</td>
</tr>
<tr>
<td>$A_c$</td>
<td>cross sectional rod area</td>
<td>$0.785 \text{ in}^2$</td>
<td>constant</td>
</tr>
<tr>
<td>$A_t$</td>
<td>cross sectional area of tubing</td>
<td>$\text{in}$</td>
<td>constant</td>
</tr>
<tr>
<td>API</td>
<td>oil API</td>
<td>45 API</td>
<td>constant</td>
</tr>
<tr>
<td>$B$</td>
<td>combined viscous friction coefficient</td>
<td>$\text{lb} \cdot \text{ft} \cdot \text{s}$</td>
<td>constant</td>
</tr>
<tr>
<td>$B_o$</td>
<td>formation volume factor</td>
<td>$1.2 \frac{\text{bbl}}{\text{STB}}$</td>
<td>constant</td>
</tr>
<tr>
<td>$c$</td>
<td>total compressibility factor</td>
<td>$0.000013 \text{ psi}^{-1}$</td>
<td>constant</td>
</tr>
</tbody>
</table>
\[ C_A \quad \text{drainage area shape factor} \quad 31.6 \quad \text{constant} \]
\[ C_{\text{initial}} \quad \text{initial cost of rod pump} \quad $1,000,000 \quad \text{constant} \]
\[ D_c \quad \text{diameter of well casing} \quad 8 \text{ in} \quad \text{constant} \]
\[ D_r \quad \text{diameter of sucker rod} \quad 1 \text{ in} \quad \text{constant} \]
\[ D_t \quad \text{diameter of production tubing} \quad 2.5 \text{ in} \quad \text{constant} \]
\[ E \quad \text{rod string modulus of elasticity} \quad 3.2 \times 10^6 \text{ psi} \quad \text{constant} \]
\[ \dot{\varepsilon}_{\text{mech,loss}} \quad \text{power to overcome frictional losses} \quad Kw \quad \text{state variable} \]
\[ E_{\text{cost}} \quad \text{cost of electricity} \quad \frac{\$}{\text{kwh}} \quad \text{constant} \]
\[ f \quad \text{rod load} \quad \frac{\text{lb}}{\text{f}} \quad \text{state variable} \]
\[ g \quad \text{gravity acceleration} \quad 32 \text{ ft/s}^2 \quad \text{constant} \]
\[ h \quad \text{reservoir thickness} \quad 8 \text{ ft} \quad \text{constant} \]
\[ h_f \quad \text{height of fluid in well annulus} \quad \text{ft} \quad \text{state variable or controlled variable} \]
\[ J_0 \quad \text{combined inertia of motor, load, and gear train} \quad \text{lb} \cdot \text{ft} \cdot \text{s}^2 \quad \text{constant} \]
\[ k \quad \text{permeability} \quad 15 \text{ md} \quad \text{constant} \]
\[ \text{NPV} \quad \text{net present value} \quad \$ \quad \text{state variable} \]
\[ \bar{P} \quad \text{average reservoir pressure} \quad \text{psi} \quad \text{state variable} \]
\[ P_{\text{bd}}(t) \quad \text{pump boundary condition parameter} \]
\[ P_i \quad \text{initial reservoir pressure} \quad 2078.4 \text{ psi} \quad \text{constant} \]
\[ P_o \quad \text{price of oil} \quad \frac{\$}{\text{bbl}} \quad \text{constant} \]
\[ p_{\text{pi}} \quad \text{pump intake pressure} \quad \text{psi} \quad \text{state variable} \]
\[ P_{\text{wf}} \quad \text{bottomhole flowing pressure} \quad \text{psi} \quad \text{state variable} \]
\[ P_{\text{wh}} \quad \text{well head pressure} \quad \text{psi} \quad \text{constant} \]
\[ p_r \quad \text{reservoir pressure} \quad \text{psi} \quad \text{parameter} \]
\[ PI \quad \text{well productivity index} \quad \frac{\text{bbl}}{\text{day} \cdot \text{psi}} \quad \text{parameter} \]
\[ r \quad \text{annual interest rate} \quad 12 \% \quad \text{constant} \]
\[ R \quad \text{revenue} \quad \$ \quad \text{state variable} \]
\[ r_w \quad \text{wellbore radius} \quad 0.328 \text{ ft} \quad \text{constant} \]
\[ S \quad \text{skin} \quad 0 \quad \text{constant} \]
\[ s_a \quad \text{slack variable} \quad \text{slack variable} \]
\[ s_b \quad \text{slack variable} \quad \text{slack variable} \]
\[ SPM \quad \text{stroking speed} \quad \frac{\text{strokes}}{\text{min}} \quad \text{manipulated variable} \]
2. Methods

This section describes the sucker rod, well, and reservoir models. Next, this section describes economic considerations, and poses general formulations of MHE and Model Predictive Control (MPC) problems.

2.1. Well and Rod String System

Figure 2a is a diagram of the well bottomhole assembly. Reservoir fluid flows into the well bore through perforations in the well casing. Fluid accumulates in the annular space between the well casing and the production tubing. Figure 2b shows a sucker rod pump surface unit adapted from Gibbs
A prime mover reciprocates the rod string via a four bar linkage. The rod string is connected to a positive displacement pump at the bottom of the well. The pump lifts reservoir fluid to the well surface. When the surface unit lifts the pump, reduced pressure causes fluid to flow into the bottom of the pump assembly through a stationary one way valve (standing valve), filling the pump barrel. When the surface unit lowers the pump, the fluid trapped in the pump barrel is forced past another one way valve (traveling valve) into the production tubing. The surface unit reciprocates the rod string and the pump produces fluid each upstroke.

Figure 2: Pump and well diagrams

2.2. Surface Unit Equations of Motion

Equations 1 and 2 describe the vertical position of a conventional four bar linkage sucker rod pump as a function of the prime mover angle, $\theta$. $L_1$ through $L_5$ are unit dimensions indicated in Figure 2b from Gibbs [22]. The same study simulates the polished rod motion assuming constant stroking speed (SPM).

To improve upon Gibbs [22], kinematic equations that account for non-constant SPM are developed for the crank arm. A simplified free body of the crank arm for non-constant angular velocity is shown in Figure 3. Figure 3 has three torques that act on the crank armature. They are the torque supplied by the motor ($T_M$), a loading torque ($T_L$), and a frictional torque..
(\(T_f\)). The torque speed characteristics of the crank arm with constant angular acceleration can be described by the kinematic equations for rotational motion, Equations 3 and 4. Using Figure 3 and assuming clockwise is positive, a torque balance taken at the center shaft of the crank arm (point c), reduces to Equation 5.

\[
u(0, \theta(t)) = L_3 \left[ \arcsin \left( \frac{L_1 \sin \theta(t)}{h} \right) + \arccos \left( \frac{h^2 + L_3^2 - L_4^2}{2L_3h} \right) \right]
\]

(1)

\[h = \sqrt{L_1^2 + L_2^2 + 2L_1L_2 \cos(\theta(t))}
\]

(2)

\[
\frac{d\theta}{dt} = \omega
\]

(3)

\[
\frac{d\omega}{dt} = \alpha
\]

(4)
\[
\sum M_c = J_0 \alpha = -T_f - T_L + T_M
\]  

(5)

The frictional torque within a motor system can be modeled with Equation 6 referenced in Virgal and Kelemen [59]:

\[
T_f = B \omega
\]  

(6)

Equation 6 is the viscous friction torque model that is commonly used as a damping term in the modeling of electric motors [63]. To simplify the analysis, the loading torque and motor torque in Equation 5 are lumped into a parameter we will define as net torque \( T_{\text{net}} \) where \( T_{\text{net}} = T_M - T_L \). Further, the rotational inertia is assumed to be the combined inertia of the motor, load, and gear train of the rod pump system. Combining Equations 3, 5, and 6 the torque balance equation reduces to Equation 7:

\[
J_0 \frac{d\omega}{dt} = -B \omega + T_{\text{net}}
\]  

(7)

Rearranging the equation so that it is in standard form, the equation becomes:

\[
J_0 \frac{d\omega}{B \, dt} = -\omega + \frac{1}{B} T_{\text{net}}
\]  

(8)

which reduces to a standard form for a first order system that is commonly encountered in process control as shown in Equation 9:

\[
\tau \frac{d\omega}{dt} = -\omega + k T_{\text{net}}
\]  

(9)

where \( \tau = \frac{J_0}{B} \) is the time constant and \( k = \frac{1}{B} \) is the gain of the system. For our analysis, it is convenient to express the dynamic equations in terms of SPM. The angular velocity relates to SPM by a simple relation given in Equation 10:

\[
\omega = \frac{2\pi}{60} SPM
\]  

(10)

Combining Equations 9 and 10 and solving explicitly for the time derivative results in the following relation:

\[
\frac{dSPM}{dt} = -\frac{1}{\tau} SPM + \frac{60 \, k}{2\pi \, \tau} T_{\text{net}}
\]  

(11)
Equation 11 expands to a second order system by adding an additional equation that relates \( SPM \) to \( \frac{d\theta}{dt} \). This can be achieved by combining Equations 3 and 10. The result is Equation 12:

\[
\frac{d\theta}{dt} = \frac{2\pi}{60} SPM \tag{12}
\]

Equations 11 and 12 describe the equations of motion for the crank arm in terms of \( SPM \). They can be synced with Equations 1 and 2 to simulate the surface position of the rod string at non-constant \( SPM \) values. The surface position is then translated to the dynamics of lower segments of the rod string through the wave equation given in Equation 13.

### 2.3. Rod String and Well Modeling

The one-dimensional wave equation with viscous damping models the rod string dynamics, shown in Equation 13 [22, 34, 20], with buoyant gravitational effects included, and describes the propagation of forces and motion in the rod string.

\[
\frac{\partial^2 u(x, t)}{\partial t^2} = a^2 \frac{\partial^2 u(x, t)}{\partial x^2} - \frac{\pi a v}{2L} \frac{\partial u(x, t)}{\partial t} - \left( 1 - \frac{\rho_w \gamma}{\rho_r} \right) g \tag{13}
\]

Simulating the rod string requires two boundary conditions. First, the position of the polished rod load is specified, as shown generally in Equation 14, where \( F(t) \) represents an arbitrary motion profile determined by the surface unit. This work uses kinematic equations for a conventional rod pump, given in Section 2.2. Second, the behavior of the downhole pump is modeled by Equation 15 where \( \alpha \), \( \beta \), and \( P_{bd}(t) \) depend on the pumping conditions [22]. This study assumes the produced fluid to be liquid and incompressible. In this case \( \alpha \) is equal to 0, \( \beta \) is equal to 1, and \( P_{bd}(t) \) is given by Equation 16. Equation 16 implies that the load at the pump is zero when the pump is descending (the fluid column is held up by production tubing) and equal to the buoyant weight of the fluid in the production tubing when the pump is ascending.

\[
u(0, t) = F(t) \tag{14}\]

\[P_{bd}(t) = \alpha_p + \beta \frac{\partial u(x_f, t)}{\partial x} \tag{15}\]
\[
P_{bd}(t) = \begin{cases} 
W_f - (A_t - A_e)P_w & \text{if: } \frac{\partial u(x_f,t)}{\partial t} > 0.0 \\
0.0 & \text{if: } \frac{\partial u(x_f,t)}{\partial t} \leq 0.0 
\end{cases} 
\] (16)

A differential form of Hook’s Law determines the load at each rod segment as shown in Equation 17.

\[
f(x,t) = E A_c \frac{\partial u(x,t)}{\partial x} 
\] (17)

This work also includes the dynamic effects of annulus fluid level and reservoir influx. A mass material balance on the downhole well assembly shown in Figure 2a yields Equation 18.

\[
\frac{dm}{dt} = \rho_w \gamma (q_{in} - q_{prod}) 
\] (18)

Where \( \frac{dm}{dt} \) is the change of mass in the well annulus, and \( q_{in} \) and \( q_{prod} \) are the influx of fluids from the reservoir and the fluid produced by the pump. \( q_{in} \) and \( q_{prod} \) are shown in Equations 19 and 23 for incompressible fluids. Equation 19 is piecewise because fluid is only removed from the control volume during the pump upstroke.

\[
q_{prod} = \begin{cases} 
A_{c_{pump}} \frac{\partial u(x_f,t)}{\partial t} & \text{if: } \frac{\partial u(x_f,t)}{\partial t} > 0.0 \\
0.0 & \text{if: } \frac{\partial u(x_f,t)}{\partial t} \leq 0.0 
\end{cases} 
\] (19)

Assuming the fluid is incompressible, Equation 18 expands to Equation 20. Simplifying leads to the final equation describing the change in height of fluid in the well annulus, Equation 21. Reformulating Equation 21 into oilfield units gives Equation 22. The simulation solves Equations 13, 14, 19, 21, and 23 simultaneously, to dynamically simulate the well.

\[
\rho_f \gamma (A_{casing} - A_{tubing}) \frac{dh}{dt} = \rho_w \gamma (q_{in} - q_{prod}) 
\] (20)

\[
\frac{dh}{dt} = \frac{(q_{in} - q_{prod})}{(A_{casing} - A_{tubing})} 
\] (21)

\[
\frac{dh}{dt} = 1617 \frac{(q_{in} - q_{prod})}{2 (A_{casing} - A_{tubing})} 
\] (22)
2.4. Reservoir Modeling

The simplified well model considers a solution gas drive reservoir in pseudo-steady state flow. To simplify the analysis further, we assume gas is held in the solution over the life of the well, i.e. the oil pressure never falls below bubble point pressure. Thus we do not have to consider more complicated dynamics such as two phase flow and relative permeability in the reservoir. Using the assumptions defined above the inflow performance relationship for the reservoir is defined by Equation 23 [17].

\[
q_{in} = \frac{kh(T - P_{wf})}{141.2B_0\mu \left( \frac{1}{2} \ln \frac{4A}{\gamma C A_{r_w}^2} + S \right)}
\] (23)

At pressures above the bubble point, the fluid recovery from an oil reservoir depends entirely on the fluid expansion as the reservoir pressure declines. This behavior can be described by the isothermal compressibility defined by Equation 24 [39].

\[
c = -\frac{1}{V} \frac{\partial V}{\partial P}
\] (24)

Integrating Equation 24 by the method of separation of variables from initial reservoir pressure to the current average reservoir pressure, the solution to the partial differential equation becomes the following:

\[
\frac{V}{V_i} = e^{c(P_i - \bar{P})}
\] (25)

It should be noted that \(c\) is assumed constant over the producing life of the well. The volume at the lower average reservoir pressure \(\bar{P}\) includes the volume left in the reservoir, \(V_i\), and the volume of the fluid that has been produced, \(V_p\), i.e.

\[
V = V_i + V_p
\] (26)

Combining Equations 25 and 26 an explicit solution for average reservoir pressure can be derived as a function of the cumulative volume extracted from the reservoir as shown in Equation 27.

\[
\bar{P} = P_i - \frac{1}{c} \ln \left( \frac{V_p}{V_i} + 1 \right)
\] (27)

The cumulative volume produced is given by Equation 28.
Using Equations 23, 27, and 28 an inflow performance relationship (IPR) model of a reservoir can be developed. The values of the constants given in the equations are shown in Table 1.

2.5. Well Vertical Lift Performance

The Vertical Lift Performance (VLP) of a rod pumped well is defined by Equation 29:

\[ P_{wf} = 0.433\gamma_h f \]  

To obtain Equation 29 we assumed that the pressure at the liquid surface in the annulus is at atmospheric pressure, i.e., the casing surface pressure is atmospheric and the hydrostatic pressure of the gas column above the fluid in the annulus is negligible. Combining Equation 23 from IPR and Equation 29 from VLP, nodal analysis can be performed to acquire the flow rate supplied by the reservoir as the fluid level changes in the annulus.

2.6. Economics

In the oil and gas industry, NPV is a common metric used to measure the economic viability of a project. NPV considers the time value of money and helps distinguish between multiple investment options. The continuous form of NPV is given by Equation 30:

\[ NPV = \int S(t)e^{-rt}dt - C_{initial} \]  

The profit rate, \( S(t) \), is given by the difference between the revenue and expense rates as shown in Equation 31:

\[ S(t) = \frac{dR(t)}{dt} - \frac{dX(t)}{dt} \]  

For the rod pump system, the revenue rate depends on the production rate and the price of oil as shown in Equation 32:

\[ \frac{dR(t)}{dt} = P_o q_{out}(t) \]
The expenses depend on the initial cost of the rod pump and the cost to operate the motor to pump the fluid. Applying an energy balance and mass balance between the bottom of the tubing string to the well head, the work required to produce the fluid with an electric motor can be derived. The initial energy balance is given by Equation 33. The mass balance is given by Equation 34:

\[ \dot{W}_{\text{motor}} + \dot{m} \left( \frac{P_{\text{wf}}}{p} + \frac{v_{\text{in}}^2}{2} + g z_{\text{in}} \right) = \dot{m} \left( \frac{P_{\text{wh}}}{p} + \frac{v_{\text{out}}^2}{2} + g z_{\text{out}} \right) + \dot{e}_{\text{mech,loss}} \] (33)

\[ \rho V_{\text{in}} A_t = \rho V_{\text{out}} A_t = \rho Q \] (34)

Combining Equations 33 and 34 and assuming the fluid is incompressible, an explicit expression for the rate of work of the rod pump motor can be formulated in terms of the pumping rate, \( Q \). The expression is given by Equation 35:

\[ \dot{W}_{\text{motor}} = Q(P_{\text{wh}} - P_{\text{wf}} + \rho g z_{\text{out}}) + \dot{e}_{\text{mech,loss}} \] (35)

The horsepower to overcome frictional losses for a rod pump can be empirically estimated by Equation 36 [17]:

\[ \dot{e}_{\text{mech,loss}} = 6.31 \times 10^{-7} W_r S_{\text{length}} (SPM) \] (36)

Combining Equations 35 and 36 and reformulating the equation so that it is compatible with oilfield units gives Equation 37. The rate of work of the rod pump motor has units of kilowatts.

\[ \dot{W}_{\text{motor}} = 1.2687 \times 10^{-5} q_{\text{out}} (P_{\text{wh}} - P_{\text{wf}} + 0.433 \gamma TVD) + 4.7053 \times 10^{-7} W_r S_{\text{length}} SPM \] (37)

The expense rate of operating the motor can then be determined. The result is given by Equation 38:

\[ \frac{dX(t)}{dt} = 24 \dot{W}_{\text{motor}} E_{\text{cost}} \] (38)

where \( E_{\text{cost}} \) is the cost of electricity per kWh. Substituting Equations 32 and 38 into Equation 30 gives the final expression to calculate the NPV of operating the rod pump over the life of the well:
\[ NPV = \int (P_{\text{fuel}}(t) - 24W_{\text{motor}}E_{\text{cost}})e^{-rt}dt - C_{\text{initial}} \]  

(39)

It is important to note that 100% motor efficiency and no mechanical downtime is assumed when deriving Equation 39.

2.7. Moving Horizon Estimation and Model Predictive Control

MHE is a dynamic optimization technique that looks back at a time horizon and fits model parameters to historical data. MHE approximates uncertain parameters or variables, and performs well on systems that include large amounts of noise in the data set [37]. MHE also allows process information to be directly considered during optimization [55]. The main disadvantage of MHE is computational time, but real time solutions are possible, even for large systems [65]. In this work, an optimizer uses a squared error objective, as described in Hedengren et al. [29], to minimize the error in the model fit by varying uncertain model parameters or states. Equation 40 shows the general MHE squared error. The model equations (Equations 1 - 39 and 55 - 59) are included as Equations 41 - 43 in nonlinear open-equation form with differential, algebraic, and inequality constraints, respectively. The objective function minimizes the sum of square error between measured data and model predictions. Terms used in Equations 40 - 43 are shown in Table 2.

\[
\min_{x,y,p} \Phi = (y_x - y)^TW_m(y_x - y) 
\]

(40)

\[
s.t. \quad 0 = f\left( \frac{dx}{dt}, x, y, p, u \right) 
\]

(41)

\[
0 = g(x, y, p, u) 
\]

(42)

\[
0 \leq h(x, y, p, u) 
\]

(43)

Estimation of uncertain parameters \((p)\) in wells and reservoirs is an important area of research [42, 56]. The process of estimating reservoir parameters from historical data \((y_x)\) is often called history matching. Prior work illustrates estimation of reservoir properties using simultaneous techniques [33, 66]. The ability and importance of estimating parameters in real-time is also shown [2, 26]. This work expands prior work by utilizing a combined rod string, well, and reservoir model to estimate parameters in real-time and utilize those parameters in automatic well control.
### Table 2: Objective Function Terms from Hedengren et al. [29] for MHE and MPC

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Phi$</td>
<td>objective function</td>
</tr>
<tr>
<td>$y_x$</td>
<td>measurements $(y_{x,0}, ..., y_{x,n})^T$</td>
</tr>
<tr>
<td>$y$</td>
<td>model values $(y_0, ..., y_n)^T$</td>
</tr>
<tr>
<td>$W_m$</td>
<td>measurement deviation penalty</td>
</tr>
<tr>
<td>$x, u, p$</td>
<td>states $(x)$, inputs $(u)$, and parameters $(p)$</td>
</tr>
<tr>
<td>$\Delta p$</td>
<td>change in parameters</td>
</tr>
<tr>
<td>$f, g, h$</td>
<td>equation residuals, output fraction, and inequality constraints</td>
</tr>
<tr>
<td>$y_t, y_{t,hi}, y_{t,lo}$</td>
<td>desired trajectory target or dead-band</td>
</tr>
<tr>
<td>$W_{hi}, W_{lo}$</td>
<td>penalty outside trajectory dead-band</td>
</tr>
<tr>
<td>$c_y, c_u, c_{\Delta u}$</td>
<td>cost of $y$, $u$, and $\Delta u$, respectively</td>
</tr>
<tr>
<td>$\tau_c$</td>
<td>time constant of desired controlled variable response</td>
</tr>
<tr>
<td>$e_{lo}, e_{hi}$</td>
<td>slack variable below or above the trajectory dead-band</td>
</tr>
<tr>
<td>$sp_{lo}, sp_{hi}$</td>
<td>lower and upper bounds to final set point dead-band</td>
</tr>
</tbody>
</table>

MPC determines future manipulated variables at control intervals. Industrial MPC began in the energy industry and has been applied to many systems [47]. The MPC application minimizes the $l_1$-norm error between the desired controlled variable trajectory and the model prediction by varying the variables that can be manipulated $(u)$. At each time step the next optimal control move $(u_0)$ is implemented, then the process repeats. A general formulation of the MPC problem is given in Equations 44–51 Hedengren et al. [29]. The objective function (Equation 44) is composed of four terms that minimize an upper trajectory error, a lower trajectory error, the cost of the controlled variable $(y)$, and the cost of manipulated variable movement $(\Delta u)$. Equations 45–47 are the nonlinear model equations after spatial discretization of the PDE with differential, algebraic, and inequality constraints, respectively. Equations 48–49 are first-order upper and lower reference trajectories that guide the speed of the controlled variable response with the parameter $\tau_c$. Equations 50–52 are error variables that are used in the objective function to penalize deviations outside the reference trajectory dead-band. The dead-band is set to $\pm0.1$ ft to account for the natural variation in fluid height with the discrete stroke cycles. This dead-band can be further increased to reduce the variation of the stroking speed manipulated variable adjustments due to process disturbances. Terms used in Equations 44–52
are shown in Table 2.

\[ \min_{x,y,u} \Phi = W_{hi}^T e_{hi} + W_{lo}^T e_{lo} + y^T c_y + \Delta u^T c_{\Delta u} \]  
\[ \text{s.t.} \quad 0 = f \left( \frac{dx}{dt}, x, y, p, d, u \right) \]  
\[ 0 = g(x, y, p, u) \]  
\[ 0 \leq h(x, y, p, u) \]  
\[ s_{phi} = \tau_c \frac{dy_{hi}}{dt} + y_{hi} \]  
\[ s_{plo} = \tau_c \frac{dy_{lo}}{dt} + y_{lo} \]  
\[ e_{hi} \geq y - y_{hi} \]  
\[ e_{lo} \geq y_{lo} - y \]  
\[ 0 \geq e_{hi}, e_{lo} \]  

3. Solution Methods

Two solution methods to the wave equation include a numerical approximation and an exact analytical solution. The following sections discuss a numerical and analytical approach to a simplified problem with the analytical solution verifying the numerical approximations that are used in optimization, and in the non-ideal case.

3.1. Numerical Methods

Equation 13 is a second order hyperbolic partial differential equation (PDE). Numerically simulating hyperbolic PDEs is difficult since they are unstable, or conditionally stable when solved with some numerical methods. Also, some numerical simulation methods suffer from numerical damping.

Gibbs [22] uses finite differencing to convert Equation 13 into an algebraic expression. First order forward differencing approximates \( \frac{\partial u(x,t)}{\partial t} \). First order central differencing approximates \( \frac{\partial^2 u(x,t)}{\partial x^2} \) and \( \frac{\partial^2 u(x,t)}{\partial x^2} \). Doty and Schmidt [16] use a similar finite difference method to simulate a sucker rod system.

For converting differential equations into algebraic equations this work utilizes both finite differencing and orthogonal collocation on a finite element. Partial derivatives with respect to space \((x)\) in Equation 13 are approximated.
using first order central differencing. First and second order derivatives with respect to time \((t)\) are approximated by orthogonal collocation on a finite element. \(\frac{\partial u(x,t)}{\partial x}\) in Equation 17 is approximated with a first order central difference at internal rod positions and by a second order forward difference at the surface, and a second order backward difference at the pump. The GEKKO optimization suite with the IPOPT solver simulates the system and solves subsequent MHE and MPC applications [8, 60, 7]. To verify the validity of this method a special case of the solution method is compared to a Fourier series analytical solution. A grid independence study determines by observation the stability and discretization requirements to achieve stable and accurate solutions to the well and reservoir model.

3.2. Simulating Pump Boundary Conditions for Optimization

Simulating the sucker rod system requires time varying boundary conditions at the pump. For example, in the case that the pump is liquid full and the pump valves are operating correctly, when the pump is traveling upwards it is loaded with the buoyant weight of the fluid in the production tubing, and when the pump is descending it is unloaded (the weight is borne by the production tubing). Previous work simulates these changing conditions with a series of conditional statements or tests [22, 16]. However, these conditional statements create discontinuities in the model and these discontinuities are inappropriate for dynamic optimization using simultaneous solvers [29]. This work formulates the pump boundary conditions with MPCCs to solve the well model in both dynamic simulation and optimization modes with large-scale solvers. MPCCs have been used in similar problems with discrete decisions or switching points [49, 6, 4, 5, 46]. The approach taken in this work is to include the bilinear switching terms as objective function terms and optionally as an inequality constraint. An alternative strategy is to include the bilinear switching terms as an equality constraint but it is found that the small slack value \(10^{-4}\) as shown in Equation 60 is needed to improve solver convergence speed. The additional constraints add computational expense and are not needed for simulation and MPC but are needed for MHE with model mismatch. When there is model mismatch the MPCC objective and MHE objective trade off producing non-physical results. Adding the inequality constraint ensures the MPCC objective approaches zero, and the MHE application can accurately estimate model parameters. The MPCC objective function contributions are verified to be zero at the solution when only the
objective function form is used. If the objective function is not zero, Equation 60 is added to the model. It is found that the inequality constraint is only needed for the one particular case (MHE with model mismatch) which would typically occur with real data. The additional slack variables and bilinear switching terms enable gradient-based solution methods to be used with the discontinuous pump boundary conditions. An MPCC formulation of the signum function, shown in Equation 53, determines when the pump is traveling up or down. When the velocity of the pump is positive it is traveling upward, and \( y \) is equal to 1. When the pump is descending the velocity is negative and \( y \) is equal to -1.

\[
y(t) = \text{sgn} \left( \frac{\partial u(x_f, t)}{\partial t} \right)
\] (53)

The solver solves Equations 54 through 58 and determines the sign of the velocity, \( y \). Equation 59 uses \( y \) to determine the boundary condition at the pump. \( P_{bd}(t) = 0 \) when \( y = -1 \) and \( P_{bd}(t) = \frac{W_f - (A_t - A_c)P_{wf}}{EA_c} \) when \( y = 1 \).

\[
\begin{align*}
\min & \quad s_a(1 - y) + s_b(y - 1) \\
\frac{\partial u(x_f, t)}{\partial t} & = s_b - s_a \\
-1 & \leq y \leq 1 \\
&s_a \geq 0 \\
&s_b \geq 0
\end{align*}
\] (54) (55) (56) (57) (58)

\[
P_{bd}(t) = \frac{W_f - (A_t - A_c)P_{wf}}{EA_c} \cdot \frac{(1 + y)}{2}
\] (59)

\[
10^{-4} \geq s_a(1 - y) + s_b(y - 1)
\] (60)

3.3. Comparison to Analytical Solution

A simplified version of Equation 13 is obtained when the viscous damping term is equal to zero and gravitational effects are neglected. The resulting wave equation is shown in Equation 61. When initial position and velocity of the rod string are given by Equations 63 and 64, a Fourier series provides an analytical solution. The analytical solution to this case is given in Equation 62. This solution is called the d’Alembert solution to the Cauchy problem for the wave equation 43. Where \( \phi \) and \( \psi \) are the initial position and velocity of \( u \).
\[
\frac{\partial^2 u(x, t)}{\partial t^2} = a^2 \frac{\partial^2 u(x, t)}{\partial x^2} 
\]

(61)

\[
u(x, t) = \frac{1}{2} (\phi(x + at) + \phi(x + at)) + \frac{1}{2a} \int_{x-at}^{x+at} \psi(s)ds
\]

(62)

\[
u(x, 0) = \phi(x)
\]

(63)

\[
\frac{\partial u(x, 0)}{\partial t} = \psi(x)
\]

(64)

A simulation of this special case compares the analytical solution to the proposed method. The initial position and velocity of the rod are given in Equations 65 and 66. The boundary condition in space is that the rod is of infinite length. Equation 62 finds the analytical solution to this initial value problem which is given in Equation 67.

\[
u(x, 0) = \cos(x)
\]

(65)

\[
\frac{\partial u(x, 0)}{\partial t} = \sin(2x) \quad \text{for} \quad -\infty < x < \infty
\]

(66)

\[
u(x, t) = \cos(x) \cos(at) + \frac{1}{2a} \sin(2x) \sin(2at)
\]

(67)

Figure 4 shows the analytical solution method versus the orthogonal collocation (in time) with the finite difference method (in space) over the time domain \(0.0 \leq t \leq 0.0005\) and the spatial domain \(0.0 \leq x \leq 2\pi\). The solutions are very similar. Figure 5 shows the difference in the proposed solution method versus the analytical solution. The difference is small for most of the domain. Error increases further out from the initial time, \(t = 0\), indicating that the proposed solution method, in this case, appears to have numerical damping. Despite moderate numerical damping, the results indicate that the proposed method appropriately simulates the simplified wave equation.

A grid independence study identifies criteria for discretization in time and space that allows for sufficiently accurate solutions to simulate the rod string behavior in the combined rod string, well, and reservoir model.

22
Figure 4: Comparison of analytical and GEKKO solutions

Figure 5: Error of proposed solution method vs. analytical solution
3.4. Grid Independence Study

A grid independence study evaluates the solution dependence on time and space grid size. Discretizing in time and space to simulate the system creates a large problem with thousands of variables and equations. This section determines minimum spatial and time resolution to capture system dynamics. Simulations for the combined rod string, well, and reservoir model are solved with increasingly fine discretization in time and space. The solver simulates the horizon iteratively in one second intervals.

The solver attempts to simulate all combinations with the sucker rod discretized in 3, 4, and 5 to 30 segments in increments of 5, and time discretized with 5 to 30 intervals in each second, in increments of 5. Figures 6 and 7 show representative results of the study. In these figures, the polished rod (surface) load is shown vs. time. Figure 6 shows results for 10 rod segments and varying time discretization. Figure 7 shows results for 20 time segments in each second and varying rod discretization. All simulations are initialized with $u(x, t) = 0$ and $\frac{du}{dt} = 0$. Cases that do not successfully solve for the first second of the horizon are omitted.

Figure 6 shows that when fewer time steps are used, the solution does not adequately capture the dynamics of the force and position propagation in the sucker rod. 20 segments in time appears to adequately captures the system dynamics. Figure 7 shows that increasing the number of rod segments changes the shape of the simulation results very little. However, the results do show that as the number of rod segments increase the entire curve shifts upward, although the amplitude of each case stay very similar. Few rod segments are required to capture system dynamics. These results indicate that 20 time segments per second and 10 rod segments capture a majority of the system dynamics, shown on Figures 6 and 7 in dashed red. This discretization is selected for simulation, estimation, and control in Section 6 as a compromise between computational speed and accuracy. Results shown in Section 4 are for 30 points in time per second and 30 rod segments.
Figure 6: Fixed rod length discretization of 10 sections

Figure 7: Fixed number of time points at 20
4. Modeling Results

Modeling results are discussed for reservoir and rod string dynamics. A unique contribution of this work is the combined models that are used in optimizing production and NPV. Both MHE and MPC require a system model. A more accurate system model leads to more accurate parameter estimation and improved MPC controller performance. This section verifies that the proposed model accurately describes the system, and is appropriate for MHE and MPC use.

4.1. Reservoir Pressure and Production Dynamics

The well and reservoir model described in Sections 2 and 3 simulate the well system over a 30 minute horizon using 30 rod string discretizations and 30 time points per second. It took 3.14 hours to run the entire 30 minute simulation. Figures 8 through 16 illustrate the results of the simulation. Figure 12 gives a modified version of the time horizon to demonstrate NPV. Figure 13 captures the transient dynamics within the first ten seconds after the simulation is initialized. Figures 14 to 16 capture the steady state dynamics of the rod string shown in the time interval 20 to 40 seconds after the simulation is initialized.
Figure 8 shows how the annular fluid level changes over the 30 minute simulation period with adjustments to net torque using a doublet test. $T_{\text{net}}$ is defined as the manipulated variable. $T_{\text{net}}$ is initialized at 10 ft-lbs and is adjusted between 5 and 15 ft-lbs over the 30 minute simulation. The height of the fluid level in the annulus is defined as the controlled variable. During the simulation, the fluid level in the annulus changes dependent on the torque input. In cases where the fluid level in the annulus rises, the inflow rate of the reservoir exceeds the motor pumping rate. In cases where the fluid level in the annulus declines, the pumping rate exceeds the inflow rate of the reservoir.
Figure 9 illustrates the reservoir pressure decline over the 30 minute simulation period for the solution gas drive reservoir. As fluid is produced and enters the annulus, the reservoir reserves are depleted. In this case, the cumulative volume produced is dependent on the bottomhole flowing pressure and pumping rate of the motor. As the cumulative volume produced increases, we observe a decline in pressure in accordance with Equation 27.
Figure 10: The system is a dynamic blend of inflow and outflow of fluids

Figure 10 shows the interacting dynamics between the reservoir and rod pump. Figure 10a shows the first ten seconds of the simulation. The rod pump is stroking with a $T_{\text{net}}$ of 10 ft-lbs and the fluid level is oscillating up and down. The fluid level in the annulus exhibits non-linear behavior because of the intermittent pumping rate of the rod string. The producing flow rate rises greatly with the upstroke of the rod string and is zero when the pump is not pulling fluid on the downstroke. It can also be observed that the fluid rate flowing into the annulus ($q_{\text{in}}$) is fairly constant for the first ten seconds. This is because the bottomhole pressure and reservoir pressure does not change significantly over the first ten seconds of the simulation. Figure 10b shows the interacting dynamics between the reservoir and rod pump over the full 30 minute simulation. Initially, the fluid level in the annulus rises because the reservoir inflow exceeds the pumping rate of the rod string. When the $T_{\text{net}}$ is increased to 15 ft-lbs the pumping rate increases, the fluid level in the annulus decreases. This is because the pumping rate exceeds the reservoir inflow. The "tug of war" between inflow and outflow is constantly observed through the changing liquid level height in the annulus and is difficult to control manually due to the inherent non-linear nature of the system. Using a control scheme like MPC, the fluid level can be controlled to a desired set point and maintained over the simulation period that otherwise may be difficult to control manually.
Figures 11 shows the interacting dynamics between power consumption, SPM, $T_{net}$, and Flowing Bottomhole Pressure (FBHP). When SPM is constant it is equal to $T_{net}$, and in the longer simulation there is a very short time constant following a first order response. Figure 11a shows the first ten seconds of the simulation. The rod pump is stroking at a rate of 10 SPM (which corresponds to $T_{net} = 10$) and the bottomhole pressure and power consumption is fluctuating up and down. The bottomhole pressure and power consumption exhibit non-linear behavior because of the intermittent pumping rate of the rod string. The power consumption rises with the upstroke of the well. At the same time, the bottomhole pressure declines because the fluid level in the annulus drops slightly. On the downstroke, the power consumption is zero and the bottomhole pressure increases. The bottomhole pressure increases on the downstroke because the liquid level height in the annulus increases due to reservoir inflow. Figure 11b shows the interacting dynamics between power consumption, SPM, $T_{net}$, and FBHP during the full 30 minute simulation. It is clear that the power consumption increases with increasing $T_{net}$. This makes sense because at a higher $T_{net}$ setting, more volume of fluid moves in a smaller amount of time. Further, power consumption increases with time at fixed $T_{net}$ inputs. This occurs because as FBHP declines, there is less assistance from the reservoir to help push the column of fluid above the pump to surface. Hence, power consumption increases with time at a fixed $T_{net}$ value.
Figure 12: Shows how NPV varies over an 8 year simulation period

Figure 12 illustrates how NPV changes over the life of the well. Because the resolution of a 30 minute simulation cannot capture the entire dynamics of NPV, the simulation is extended to 8 years. As is expected, the NPV is negative during the early development of the well due to the initial cost of the rod pump which is assumed to be $1 million. Initially, the rod pump produces at the maximum SPM and $T_{net}$ setting and NPV increases. As time progresses, the income is discounted and begins to deviate from a linear profile as shown between year 1 and year 3. Additional factors that cause the trajectory to deviate from a linear profile include: (1) the inflow rate from the well declines, forcing the pump to produce at a lower $T_{net}$ value and (2) the operating costs increase because the bottomhole pressure declines providing less assistance to raise the fluid to surface. As time progresses, the NPV attains a maximum value around year 3 as shown in Figure 12. Beyond this point, the operating costs begin to exceed the income produced from production for the reasons mentioned above. Therefore, a decline in
NPV is apparent. As illustrated by the figure above, the goal of a producer is to attain the maximum NPV. Using control methods combined with optimization, higher NPV can be realized that would be difficult to attain through manual control methods.

4.2. Rod String Dynamics

The rod string and pump dynamics are illustrated in this section. The first ten seconds and the 20 to 40 second interval of the 30 minute simulation (to avoid the noise from the transient start up of the system) show the quick dynamics of the rod string and pump system. The simulation takes into account the changes in the reservoir and the oscillations of the pump as it removes production fluid.

Figure 13: Dynamic rod string position in space and time (10 seconds)

Figure 13 displays the sucker rod position in time and space. Distance 0 ft represents the rod at the surface that is actuated by the surface unit.
Distance 4800 ft is the rod position at the pump. Figure 14 gives the position and velocity of the pump. The sign of the velocity is also shown, scaled by 10 to be visible. The solver accurately determines when the velocity is positive, and when it is negative, which determines the fluid load at the pump, simulating pumping conditions.

![Figure 14: Pump position, velocity, and velocity sign (20-40 seconds)](image)

Figures 15a, 15b, 16a and 16b refer to the load on the system at the surface, for the pump, and as the pump load changes with position and time. Offset lines on the surface card and pump card would be present from startup dynamics of the first 10 seconds so the segment of 20-40 seconds is shown. The rod load varies depending on stroking rate and the fluid load at the pump. The dynamic load is essentially zero when the pump rod rests buoyantly in the annular fluid. The polished rod load mimics the motion of the pump at the surface and the rod load versus the position of the pump draws out a dynamometer card in roughly the shape of a rectangle which
corresponds to actual data and previous research on rod pumps.

![Graph](image)

Figure 15: Rod load vs. position

![Graph](image)

Figure 16: Rod load vs. time

5. Moving Horizon Estimation Results

In the oil and gas industry, it is common not to have all the available data to carry out meaningful reservoir engineering analysis. The cost to acquire new information through well intervention techniques such as well logging, pressure build up tests, core analysis, and well surveys, is expensive. Further, those measurements are typically limited to a snapshot in time and require
the well to be shut-in. Other methods used to acquire continuous reservoir surveillance data through downhole electronic sensors always suffer in reliability due to the extreme pressures and temperatures encountered downhole \[9\]. Thus, there is a high degree of risk involved in implementing downhole measuring devices. Because operators are trained to minimize expenses, reduce downtime, and limit risks, this information is usually never realized, or acquired indirectly using stochastic estimates from other wells in the area. With MHE, reservoir properties and flow variables can be estimated at no cost, without shutting in the well, and at no risk. This section highlights the results of two simulation case studies where MHE is implemented to (1) infer the fluid level in the annulus of a rod pumped well and (2) estimate reservoir parameters. Both case studies use the simulation model from Section 4. Two measured parameters are passed to the MHE: (1) the surface polished rod load and (2) the net torque produced by the motor. This is convenient for operators because both measurements can be easily acquired from the surface through surface dynagraph cards and gauges on the motor. The case studies are shown in Table 3. In Case 1 the MHE estimates the annular fluid height given the rod load at the surface (polished rod load) and the net torque produced by the motor. The MHE estimate of annular fluid height is compared to the true value to validate that the fluid level in the annulus can be inferred from the surface load. In Case 2, MHE is used to estimate reservoir properties such as porosity and flow properties such as skin using the surface load and net torque produced by the motor. The true values are compared to the estimated values. All MHE figures use a discretization of 20 points in each one second interval and 10 rod sections.

Table 3: Estimation Case Studies

<table>
<thead>
<tr>
<th>Estimation Case</th>
<th>Estimated Variable(s)</th>
<th>Measured Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Annular fluid height</td>
<td>Polished rod load, $T_{net}$</td>
</tr>
<tr>
<td>2</td>
<td>Well Skin factor and Porosity</td>
<td>Polished rod load, $T_{net}$</td>
</tr>
</tbody>
</table>

5.1. Annular Fluid Height Estimation

In rod pumped wells, it is critical to maintain the annular fluid level above the pump to avoid pump-off and fluid pound conditions \[36\]. Fur-
ther, it is desirable to produce at the lowest level possible to maximize NPV over the lifetime of the well. Traditional methods used to measure the fluid level height in rod pumped wells include echo-meter surveys and electronic bottomhole pressure gauges. Although these tools have benefits, there are some drawbacks as described in the previous section. Figure 17 highlights the MHE estimate of the annular fluid height from the surface load values and the net torque produced by the motor excluding measurement noise. At the initial time, it is assumed that the well is not flowing and the hydrostatic head of the fluid is equal to the reservoir pressure. Therefore, for this scenario the initial condition assigned to the MHE is assumed to be equivalent to the true initial condition as demonstrated in Figure 17.

![Figure 17: Height is estimated with varying inputs of $T_{net}$ (which adjusts based off the load)](image)

From Figure 17 it is clear that the MHE estimate of annular fluid level height is identical to the true value. In practice, there may be measurement

36
noise in the load measurements taken at the surface. Further, there may be uncertainty in the starting fluid level height of the annulus. To test the robustness of the MHE estimator, an additional scenario is investigated which includes added measurement noise and an initial condition different from the actual fluid level height. In this scenario, the initial condition is set to 100 ft below the actual fluid level height (2400 ft) at the initial time. Normally distributed white noise is added to the measured force from the simulator with a standard deviation of 50 lbs. The results are shown in Figure 18.

![Figure 18: Height is estimated with varied $T_{net}$, noise, and an offset initial condition](image)

In this case an inequality constraint, Equation [60], is added to the MHE model to ensure the MPCC objective term approaches zero. If there is no inequality constraint, the optimizer finds non-physical solutions when there is process model mismatch as a result of measurement noise. Figure 18 suggests that the measurement noise adversely affects the estimator performance. However, the MHE is adequately successful in tracking the actual fluid level.
height. The estimator approaches the actual fluid height after about 60 seconds even when the initial condition is unknown. It is interesting to note that because the system has intermittent production, i.e. fluid is produced on the upstroke but not on the downstroke, the MHE estimator is only able to retrieve useful loading data on the upstroke. This is observed in Figure 18 where the estimated height is constant then steps during periods where the fluid level is decreasing. This is because on the downstroke, when no fluid is produced, the loading condition is constant and does not provide useful information for the MHE to react. Despite white measurement noise, the MHE application is able to accurately track the fluid level height over the course of the simulation. The results from the two scenarios validate that the MHE can use surface loading measurements to infer the annular fluid level height in the well. Thus downhole pressure gauge measurements and acoustic surveys are not necessarily required. By estimating annular fluid height the MHE can be integrated with MPC to lead to automated control as described in Section 6.

5.2. Reservoir Parameter Estimation

Parameters of interest to reservoir engineers include drainage area, porosity, pressure, permeability, skin factor, etc. This study estimates the porosity and skin using MHE.

In general, parameters such as porosity and skin are considered time-invariant in reservoir modeling. However, in practice, these parameters change during depletion of a well and are not necessarily time-invariant. For example, depleting reservoirs generally experience a change in total stress state in the reservoir rock as pressure reduction occurs over time [30]. Therefore, it is expected that pore volume reduction occurs as the reservoir is depleted. Changes in temperature and pressure in the reservoir can also cause formation skin damage to change over time. For example, as the temperature or pressure declines with time, compounds can precipitate from the wellbore fluids, this results in permeability impairment in the near wellbore region [17]. The precipitates can build up over time resulting in a constantly changing skin value. MHE has the capability to track such changes over the life cycle of a well. Further, MHE is an additional method to retrieve useful information of reservoir parameters. In practice there are several tools used to estimate the same reservoir properties (e.g. core analysis, logging, seismic, and stochastic methods can all be used to estimate porosity). Each method
has pros and cons. For example, core analysis has a high resolution, however it is expensive and limited to a snapshot in time. Stochastic methods are fast and inexpensive, however the results may not be representative of a particular well because it is relying on a distribution of a statistical data set from other wells of a given sample size. Seismic is extremely expensive and is also limited to snapshots in time. What makes MHE unique over the other methods, is it would be relatively inexpensive, does not require well intervention, and can dynamically estimate reservoir parameters with varying conditions which allows the reservoir engineer to monitor the health of a particular well in real time.

Porosity is a measure of the amount of void space in the reservoir and is important to petroleum engineers because it determines fluid storage capacity. Knowing porosity allows a reservoir engineer to calculate the amount of oil in a reservoir. This can subsequently be used to determine an overall recovery factor and can help engineers plan the depletion strategy of a well. Traditional methods used to infer porosity include logging tools such as acoustic, density, neutron, and NMR [15]. Direct measurement of porosity is determined through core analysis in a lab environment. It is not always convenient to run logging tools or retrieve core samples from a well. Figure 17 highlights the MHE estimate of the porosity from the surface load values and net torque supplied by the motor. All other reservoir parameters are assumed to be known.
An initial guess of porosity is provided to the MHE with a value of 12%. From Figure 19 it is clear that the MHE is able to closely approximate the true value of porosity after more data became available. This shows that the current MHE model is capable of estimating reservoir parameters from a rod pumped well without well intervention. Thus, well logging and core analysis are not necessarily required.

Skin factor is another parameter of interest to reservoir engineers. Skin is a measure of formation damage or production enhancement near the wellbore and can change over the producing life of the well. Examples of formation damage include mud filtrate invasion during drilling, plugging of pore space by cement solids during completion, and scale buildup during production. Examples of production enhancement include an acid wash, solvent wash, and various other chemical treatments. Traditional methods used to estimate skin include pressure drawdown or pressure buildup tests. Each method requires well intervention. With MHE, the skin can be estimated without shutting in...
the well. Figure 20 highlights the MHE estimate of skin from the surface load values and the net torque supplied by the motor. Note, all other reservoir parameters are assumed to be known.

An initial guess of skin is provided to the MHE with a value of 0. From Figure 20 it is clear that the MHE is able to find the true value of skin after more data became available. This shows that the current MHE model is capable of estimating reservoir flow parameters from a rod pumped well without well intervention. Thus pressure drawdown and pressure buildup tests are not necessarily required. Assuming skin and porosity are unknown, Figure 21 shows the case where both parameters are estimated simultaneously.
The initial guesses for skin and porosity provided to the MHE are the same values used previously. From Figure 21 it is clear that the MHE estimate of porosity and skin are almost identical to the true values. It is interesting to note that at different starting initial conditions, the MHE converged to different values of porosity and skin. Further, from Figure 21 it can be observed that in regions where the solution oscillates, the porosity and skin compensates by an analog negative slope. This behavior suggests that the system of equations and variables may have multiple solutions for the given simulation conditions when estimating more than one variable. More information on collinearity of solutions, improving initial conditions, and sensitivity analysis can be found in Lewis et al. [37]. As demonstrated from the analysis above, MHE is capable of estimating reservoir and flow parameters from a rod pumped well using information from the surface. Further, MHE does not require well intervention. For the reasons described above, MHE
has distinct advantages over traditional methods.

6. Dynamic Optimization & Model Predictive Control

This section shows how an optimum SPM input to the well is determined to maintain a constant fluid level in the annulus. The optimized SPM is implemented into the model. Results are shown and discussed. Three control strategies are demonstrated and compared. Table 4 describes each case. All cases utilize the model of a conventional four bar linkage sucker rod pump described in Section 2.2 with three different control strategies. Case 1 shows control results for pump-off control. Case 2 combines the MHE estimation of fluid height with a proportional integral (PI) feedback controller. Case 3 shows control when MHE estimates fluid levels and a MPC application controls the unit.

When a producing well is shut down, the fluid influx from the reservoir continues until the hydrostatic pressure of fluid accumulating in the well annulus equals the reservoir pressure. When fluid production restarts, the annular fluid level is pumped down and reservoir influx reoccurs. The well is pumped down until one of two conditions occurs. Either a minimum annular fluid level, or a predetermined fluid level or FBHP is obtained. The lowest possible FBHP occurs when the fluid level is drawn down to a minimum, this results in maximum fluid production. This is common in practice. In this case, the main purpose of the controller is to maintain adequate fluid over the pump to prevent equipment damage. However, this practice does not always result in the largest NPV or ultimate recovery for a reservoir system. Multiple studies show that determining and maintaining the optimal FBHPs for a set of wells in a reservoir system increases NPV and ultimate oil recovery [12, 3, 27]. In either case, the initial pump down period is not interesting for comparing controller performance since the pump will operate at a maximum rate until the fluid level approaches the set point. In view of this, the simulation case studies compare the quality of the control when the fluid level approaches and maintains the set point, which requires a relatively short simulation time.
Table 4: Simulated Control Cases

<table>
<thead>
<tr>
<th>Case</th>
<th>Description</th>
<th>Manipulated Variable(s)</th>
<th>Controlled Variable &amp; Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Conventional Unit Control</td>
<td>On/OFF</td>
<td>Annular fluid level</td>
</tr>
<tr>
<td></td>
<td>- PUMP-OFF/TIMER</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Conventional Unit Control</td>
<td>$T_{net}$ (SPM)</td>
<td>Annular fluid level</td>
</tr>
<tr>
<td></td>
<td>- PI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Conventional Unit Control</td>
<td>$T_{net}$ (SPM)</td>
<td>Annular fluid level</td>
</tr>
<tr>
<td></td>
<td>- MPC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6.1. **Case 1 - Pump-Off Control With Timer**

This case simulates a typical pump-off control system that is commonly used in industry. For these methods, equipment such as a load cell or vibration sensor detect when a pumped off condition has occurred and/or when the fluid level has fallen below a predetermined threshold. The controller then turns off the unit for a predetermined time period. The unit restarts after the timer period ends.

Figure 22 shows the results of a 3 minute pump-off control simulation. A major benefit of pump-off controllers is their simplicity. Additionally, if measurements to detect pump-off are not available the pumping unit can be run fully on timers where the percent of time the unit operates in a day is manually tuned to match nominal production rates. These controllers are a great option when the unit is not electrified since combustion units rarely have automatically adjustable stroking speed. Further, this type of control may be preferred when the minimum stroking speed of the unit is faster than the reservoir influx. However, a major downside to these systems is that sub-optimal production rates are achieved. The time when the unit is down and fluid level is increasing represents loss in production. In order to maintain maximum production, frequent start stop cycles are required. These start stop cycles lead to increased equipment degradation. Operators strive to balance these effects. In some wells where sand is present in the produced fluid, problems occur when the unit stops and dispersed sand settles.
6.2. Case 2 - Fluid Height Control With PI

In this case, it is assumed that the field is electrified and the simulated rod pump are equipped with a variable frequency drive so that the net motor torque, $T_{net}$, can be automatically adjusted. The $T_{net}$ is adjusted by a PI controller. This is a simplification of typical real systems, which is derived in Section 2.2. In practice an electric motor is equipped with a variable frequency drive that allows control of the motor rotational speed or torque. In practice the PI controller or MPC application act as a primary controller and send a command to the motor controller acting as the secondary controller. The response time of such motor controllers is rapid, and is neglected in this study [35]. The objective is to control the annular fluid height to a desired set point. In practice this would often be as low as possible, while still maintaining fluid over the pump. However, in some cases the optimal fluid height may be set at some other value to maximize ultimate reservoir recovery, or to prevent gas interference at the pump as the pressure falls.
below the fluid bubble point. Figure 23 shows a diagram of the simulated control system for cases 2 and 3. First, the model simulates a one second horizon. The system then passes the surface load to the MHE application. The MHE application estimates the annular fluid height, and passes it to the PI or MHE controller. The controller then determines the motor torque, $T_{net}$, for the next simulator cycle.

\[ T_{net} = K_e e(t) + \frac{K_e}{\tau_I} \int e(t) dt \]  

(68)

Figure 23: Block diagram of PI and MPC controllers

This application utilizes the MHE described in Section 5.1 because in practice the fluid level is rarely measured in real-time. The fluid height is usually measured infrequently with an acoustic device, but the manual measurement and infrequency make it impractical for automatic control. However, many current pump-off control systems measure the polished rod load and acceleration. From these, the polished rod load and position are obtained as a function of time. The MHE application illustrated in 5.1 converts the measured loads into predictions of the fluid height. The MHE passes the estimated height into the PI controller application as a “measurement”. The PI controller compares the “measured” height from the MHE application and compares it to the set point height. The resulting error, $e(t)$, determines the motor torque output via Equation 68.
A 180 second simulation shows the results of Case 2. These results are shown in Figure 24. The controller draws down the fluid level rapidly, but suffers from overshoot. These results are for an aggressive tuning $K_c = 10 \frac{ft \cdot lb}{ft}$ and $\tau_I = 5$ seconds. The controller holds the set point very well but there is large chatter in the net motor torque output, which leads to accelerated equipment failure. When tuning the PI controller an operator balances holding the set point accurately and smoothing the controller response.

Figure 24: Aggressive PI controller performance $K_c = 1 \frac{ft \cdot lb}{ft}$ and $\tau_I = 1s$

One benefit of the combined PI and MHE application is computation time. Figure 25 shows the controller solve time for each cycle in Case 2 on an Intel i7 processor. The average cycle time is 1.23 seconds, indicating that, the controller is nearly appropriate for real-time control. To be appropriate for real-time control a controller cycle time must be faster than real-time. In this case most of the computation time is solving the MHE application to estimate fluid height, so faster computation may be achieved with a faster
processor or by model simplification, such as, using fewer rod discretizations.

To verify that real time control is possible, Case 2 is re-simulated where the MHE model utilizes 10 discretizations per second and 5 rod discretizations (npt=10, npx=5). The resulting average controller solve time is 0.62, indicating that the method is suitable for real time control. The disadvantage of using a reduced resolution model in the MHE application is that the model-process mismatch is greater, resulting in less accurate parameter estimates. However, these results indicate that the proposed controller utilizing MHE and PI control is possible for field-testing with reduced resolution. Future work could explore model linearization around the nominal operating conditions for both the estimator and controller to further improve computational speed.
6.3. Case 3 - Fluid Height Control With MPC

This case is similar to Case 2, except that instead of PI control a MPC determines the pumping unit net motor torque. Again, the MHE application described in Section 5.1 estimates the fluid height, and passes it as a “measurement” to MPC application.

Figure 26 shows the results of the combined MHE and MPC controller. The MPC application determines a 10 second horizon to achieve and hold the set point, varying $T_{net}$. The MPC then implements the first second of net motor torque as the control move. The MPC controller achieves the set point with no overshoot and holds the set point adequately. It is interesting to note that the set point fluid height is constant, but the actual fluid level is inherently cyclical since fluid is only produced from the well during the pump upstroke. This is different from traditional chemical processes where cyclical changes to a controlled variable are uncommon. However, the results indicate that MPC is able to track the set point well despite the cycling fluid level, with a sufficiently long prediction horizon. A measurement dead band is used to allow the height to oscillate around the set point without penalty. The maximum change in $T_{net}$ per controller cycle is held at 1 $ft\cdot lb$ per second and a minor penalty is allotted to each change in $T_{net}$, to prevent controller chatter. Overall, the MPC application performs very well.
A major downside to the MPC application is long computation time. Figure 27 shows the combined MPC and MHE controller computation time per cycle. The average computation time for the combined MHE MPC application is 7.4 seconds per one second cycle on an Intel i7 processor. One unsuccessful MPC solve occurred during the horizon at approximately 65 seconds. A speedup in computation time of 7-8X is required for real-time control with this application. A faster processor or model simplification, such as, using fewer rod discretizations may decrease computation time. To verify this, Case 3 is resolved using 10 time discretizations per second and 5 rod discretizations (npt = 10, npx = 5). The average controller solve time with this discretization is 2.39 seconds, much closer to real time applicability. As noted in Section 6.2, the main disadvantage of reduced resolution is increased model-process mismatch, which may lead to suboptimal control. In addition to reduced model resolution, a shorter MPC prediction horizon can be used, although, the controller performance degrades. A future area of study is to
explore model simplification such as linear MPC or control parameterization methods such as explicit nonlinear MPC [28].

6.4. Comparison of Cases

The cases each have pros and cons. Case 1 has the benefit of simplicity, and is well suited for fields where electric power is not available, and functions well when reservoir influx is less than the minimum stroking rate. Case 2 is intermediate in its complexity, and is a common control algorithm. Combining MHE to determine annular fluid level with PI control enables good fluid level or bottomhole pressure control with intermediate computational requirements. Finally, Case 3 with MPC provides the best control of annular fluid level, but is the most complex. Case 3 also has the ability to consider multiple objectives as well as simultaneously consider well and reservoir interactions. These features enable optimal production and profitability.
Figure 28 illustrates the differences in incremental cumulative recovery between MPC, MPC with reduced model discretization resolution, PI, PI control with reduced MHE model discretization resolution, and pump-off control for a 180 second simulation. In Figure 28, pump-off control is the base case, and increases in production are shown relative to it. From the figure, it is clear that MPC, MPC with reduced discretization, PI, and PI with reduced MHE discretization each outperform pump-off control. Further, there is little difference between the the controller performance when the full or reduced discretization models are used. This result indicates that there is little controller performance loss due to reducing the model discretization to increase computational speed. Both MPC cases produce slightly more oil than the PI cases, but this is because the PI cases hold the average fluid level slightly above the set point and MPC cases maintain an average fluid level slightly below the set point. The increase in cumulative recovery is 0.029% over the simulation period for the MPC case, 0.031% for the MPC
case with reduced discretization, 0.026% for the PI case, and 0.028% for the reduced PI case. Although these appear to be small increases in incremental production for a single well, the cumulative effect on a large-scale field would imply significant gains in field wide production. This needs to be verified in future work with high fidelity reservoir simulators to quantify production and profitability gains from improved height control.

7. Conclusion and Future Work

This work illustrates the ability to use MPC for oil production systems with sucker rod pumps. The oil production system model includes both reservoir and well/pumping unit dynamics. The models are combined to optimize for closed loop control. This determines optimal operation by directly considering all major physical constraints of the system simultaneously. Increased ultimate recovery and NPV are enabled by automatically controlling annular fluid height and bottomhole pressure.

The ability to estimate uncertain reservoir parameters is illustrated with MHE. Specifically, reservoir skin estimation is illustrated where the load and position of the rod pumping unit are measured. When fluid height is also measured, additional reservoir parameters, such as the average reservoir pressure and porosity, can be estimated.

Predictively setting strokes per minute for a conventional rod pumping unit is shown where the annular fluid level is inferentially controlled. The fluid height remains controlled within the acceptable region with minimal stroking speed adjustments. MPC application results in a slight 0.03% increase in cumulative oil production vs. conventional pump off control over a period of 3 minutes. Future work will verify production gains over longer periods with high fidelity reservoir simulations.

The models and methods in this work represent a novel contribution in oil field production well control and reservoir optimization. Simultaneously considering many, or all, system degrees of freedom and constraints may lead to more optimal reservoir production. Future work should expand these methods to include: 1) Additional artificial lift methods such as gas lift and electronic submersible pumps 2) Additional system constraints such as gathering systems, 3) More rigorous well, pump unit, and reservoir models that include prime mover dynamics, multiple phase flow, and fluid compressibility 4) Investigate and improve solve time to determine applicability for real-time control 5) Tune the application for use in unconventional reservoir settings.
Acknowledgments

The authors acknowledge the financial assistance of Utah Science Technology and Research (USTAR) through a University Technology Acceleration Grant (UTAG). USTAR is a technology-based economic development initiative funded by the state of Utah.

Source Code

Python source code is available at the following repository: https://github.com/BYU-PRISM/USTAR-Artificial-Lift.

References


