

TRABALHO DE GRADUAÇÃO

**IDENTIFICAÇÃO DE SISTEMAS LINEARES APLICADO  
A SISTEMAS DE PRODUÇÃO DE PETRÓLEO**

Gabriel de Castro Arão

Brasília, Novembro de 2021



**ENGENHARIA  
MECATRÔNICA**  
UNIVERSIDADE DE BRASÍLIA

UNIVERSIDADE DE BRASÍLIA  
Faculdade de Tecnologia  
Curso de Graduação em Engenharia de Controle e Automação

TRABALHO DE GRADUAÇÃO

IDENTIFICAÇÃO DE SISTEMAS LINEARES APLICADO  
A SISTEMAS DE PRODUÇÃO DE PETRÓLEO

Gabriel de Castro Arão

*Relatório submetido como requisito parcial de obtenção  
de grau de Engenheiro de Controle e Automação*

Banca Examinadora

Prof. Eugênio L. F. Fortaleza, ENM/UnB

*Orientador*

\_\_\_\_\_

Prof. Carlos Humberto Llanos Quintero,

ENM/UnB

*Examinador interno*

\_\_\_\_\_

Prof. José Oniram de Aquino Limaverde Filho,

ENE/UnB

*Examinador interno*

\_\_\_\_\_

**Brasília, Novembro de 2021**

## FICHA CATALOGRÁFICA

GABRIEL C., ARÃO

Análise dos efeitos dos parâmetros de identificação de sistemas de exploração de reservatórios usando modelos lineares,

[Distrito Federal] 2021.

viii, 62p., 297 mm (FT/UnB, Engenheiro, Controle e Automação, 2021). Trabalho de Graduação – Universidade de Brasília. Faculdade de Tecnologia.

1. Identificação de Sistemas

2. Engenharia de Reservatórios

3. Sistemas de Produção de Petróleo

I. Mecatrônica/FT/UnB

II. Título (Série)

## REFERÊNCIA BIBLIOGRÁFICA

ARÃO, G. C., (2021). Análise dos efeitos dos parâmetros de identificação de sistemas de exploração de reservatórios usando modelos lineares. Trabalho de Graduação em Engenharia de Controle e Automação, Publicação FT.TG-*n*º 08, Faculdade de Tecnologia, Universidade de Brasília, Brasília, DF, 62p.

## CESSÃO DE DIREITOS

AUTOR: Gabriel de Castro Arão

TÍTULO DO TRABALHO DE GRADUAÇÃO: Análise dos efeitos dos parâmetros de identificação de sistemas de exploração de reservatórios usando modelos lineares.

GRAU: Engenheiro

ANO: 2021

É concedida à Universidade de Brasília permissão para reproduzir cópias deste Trabalho de Graduação e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desse Trabalho de Graduação pode ser reproduzida sem autorização por escrito do autor.

---

Gabriel de Castro Arão

Guará I, QE 14, Conjunto D, Lote 5.

71015-041 Brasília – DF – Brasil.

## **Dedicatória**

*Dedico este trabalho a Deus, por me reconfortar, a meus pais, que sempre me apoiaram, à Emília, minha companheira e a todos os amigos que fiz ao longo dessa jornada.*

*Gabriel de Castro Arão*

## Agradecimentos

*Agradeço à Deus, pois sem ele esse trabalho não seria possível. Agradeço à minha família por todo o apoio, mesmo a distância. Por fim, sou grato a todos os amigos e pessoas incríveis que me fazem melhorar a cada dia.*

*Gabriel de Castro Arão*

---

## RESUMO

Sem objetivar a busca por bons modelos, o presente trabalho busca comparar diferentes métodos de identificação de sistemas e seus parâmetros no contexto de identificação de poço produtor de petróleo. Ao manter o modelo fixo e variar apenas os parâmetros dos métodos de identificação, é possível observar diferenças entre os resultados e na qualidade da previsão de cada modelo.

Neste trabalho é usado o modelo ARX (autorregressivo com entradas exógenas) em conjunto com métodos de identificação por batelada, com janela deslizante e identificação recursiva para produzir variados modelos de poço. Os modelos produzidos usam uma estrutura simplificada do ARX, onde todos os parâmetros possuem grau 1.

Foi constatada, com este trabalho, a superioridade que os métodos recursivos tem sobre os métodos por batelada, dentro do contexto do trabalho. A adaptabilidade do parâmetros e o foco em obter a melhor previsão do próximo passo conferem ao método recursivo resultados muito superiores e uma aplicabilidade bem mais promissora, em contraste com os métodos por batelada, que possuem parâmetros fixos e um foco em obter boas previsões observando o todo, ao invés de passo a passo.

Palavras Chave: Identificação de Sistemas, Engenharia de Reservatórios, Sistemas de Produção de Petróleo

---

## ABSTRACT

Without aiming to search for good models, the work seeks to compare the methods and their parameters in the context of identifying an oil-producing well. By keeping the model fixed and varying only the parameters of the identification methods, it is possible to observe differences between the results and the quality of prediction of each model.

In this work, the ARX model (autoregressive with exogenous inputs) is used together with batch identification methods, with sliding window and recursive identification to produce various well models. The models produced use a simplified structure of the ARX, where all parameters have grade 1.

It was verified, with this work, the superiority that the recursive methods have over the batch methods, within the context of the work. The adaptability of parameters and the focus on obtaining the best prediction of the next step give the recursive method much higher results and a much more promising applicability, in contrast to the batch methods, which have fixed parameters and a focus on obtaining good predictions by observing the whole, rather than step by step.

Keywords: System Identification, Reservoirs Engineering, Oil Production Systems

# SUMÁRIO

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	CONTEXTUALIZAÇÃO	1
1.2	DEFINIÇÃO DO PROBLEMA	1
1.3	OBJETIVOS DO TRABALHO	3
1.4	RESULTADOS OBTIDOS	4
1.5	APRESENTAÇÃO DO MANUSCRITO	4
<b>2</b>	<b>Fundamentação Teórica</b>	<b>6</b>
2.1	EXPLORAÇÃO DE RESERVATÓRIOS DE PETRÓLEO	6
2.1.1	MÉTODOS DE PRODUÇÃO E RECUPERAÇÃO	6
2.1.2	FLUIDOS PRODUZIDOS	9
2.1.3	CORTE DE ÁGUA	10
2.1.4	SIMULADORES DE RESERVATÓRIOS	10
2.1.5	EGG MODEL	10
2.2	MODELAGEM MATEMÁTICA	11
2.2.1	METODOLOGIA DE IDENTIFICAÇÃO CAIXA PRETA	12
2.3	REPRESENTAÇÕES LINEARES EM TEMPO DISCRETO	14
2.3.1	MODELO AUTO-REGRESSIVO COM ENTRADAS EXÓGENAS	15
2.3.2	ESTIMAÇÃO DE UMA CONSTANTE NO MODELO	17
2.4	ESTIMADOR DE MÍNIMOS QUADRADOS	17
2.4.1	ESTIMANDO MODELO ARX USANDO MQ	19
2.4.2	MÍNIMOS QUADRADOS RECURSIVO COM FATOR DE ESQUECIMENTO	20
2.4.3	INICIALIZAÇÃO DO ALGORITMO RECURSIVO	22
<b>3</b>	<b>Metodologia</b>	<b>23</b>
3.1	DADOS UTILIZADOS NA IDENTIFICAÇÃO	23
3.1.1	EXEMPLO DE SIMULAÇÃO 1	25
3.1.2	EXEMPLO DE SIMULAÇÃO 2	27
3.2	ESTRUTURA E ESTIMADORES UTILIZADOS	30
3.3	PARÂMETROS DE IDENTIFICAÇÃO E GERAÇÃO DE RESULTADOS	31
3.3.1	PARÂMETROS DO MODELO	31
3.3.2	PARÂMETROS DOS ESTIMADORES	31
3.3.3	MÉTODO DE IDENTIFICAÇÃO	32

3.4	VALIDAÇÃO DO ALGORITMO DE IDENTIFICAÇÃO .....	34
3.4.1	IDENTIFICAÇÃO COM ESTRUTURA PERFEITA .....	34
3.4.2	IDENTIFICAÇÃO DE CONSTANTE NO MODELO.....	36
3.4.3	IDENTIFICAÇÃO COM ESTRUTURA DE GRAU MENOR QUE O IDEAL .....	38
<b>4</b>	<b>Resultados.....</b>	<b>41</b>
4.1	RESULTADOS GRÁFICOS .....	41
4.1.1	RESULTADOS MQ POR BATELADA.....	44
4.1.2	RESULTADOS MQ POR BATELADA COM JANELA .....	46
4.1.3	RESULTADOS MQ RECURSIVO COM FATOR DE ESQUECIMENTO .....	50
4.2	RESULTADOS NUMÉRICOS .....	54
4.2.1	BONS FITS OBTIDOS.....	55
<b>5</b>	<b>Conclusões.....</b>	<b>58</b>
5.1	USO DE MODELOS LINEARES.....	58
5.2	ESTIMADORES UTILIZADOS .....	59
5.3	DADOS DO SISTEMA.....	59
5.4	PERSPECTIVAS FUTURAS.....	60
	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>61</b>

# LISTA DE FIGURAS

2.1	Reservatório com mecanismo de gás em solução [1].....	7
2.2	Reservatório com mecanismo de capa de gás [1].....	8
2.3	Reservatório com mecanismo de influxo de água [1].....	8
2.4	Representação do Egg Model, destacando os canais de alta permeabilidade e os poços injetores (em azul) e produtores (em vermelho).....	11
2.5	Diagrama do modelo geral.....	15
2.6	Diagramas do modelo ARX.....	16
3.1	Taxa de produção de óleo para os poços produtores na simulação 1. ....	26
3.2	Taxa de produção de água para os poços produtores na simulação 1. ....	26
3.3	Taxa de injeção de água para os poços injetores na simulação 1. ....	27
3.4	Pressão de fundo do poço para os poços produtores na simulação 1. ....	27
3.5	Taxa de produção de óleo para os poços produtores na simulação 2. ....	28
3.6	Taxa de produção de água para os poços produtores na simulação 2. ....	29
3.7	Taxa de injeção de água para os poços injetores na simulação 2. ....	29
3.8	Pressão de fundo do poço para os poços produtores na simulação 2. ....	30
3.9	Resultado da identificação por batelada. ....	35
3.10	Resultado da identificação por janela.....	35
3.11	Resultado da identificação recursiva. ....	36
3.12	Resultado da identificação por batelada. ....	37
3.13	Resultado da identificação por janela.....	37
3.14	Resultado da identificação recursiva. ....	38
3.15	Resultado da identificação por batelada. ....	39
3.16	Resultado da identificação por janela.....	39
3.17	Resultado da identificação recursiva. ....	40
4.1	Exemplo de resultado obtido. ....	41
4.2	Resultado do estimador em vários pontos. ....	42
4.3	Resultado do estimador em vários pontos. ....	43
4.4	WPR.....	43
4.5	WIR.....	44
4.6	BHP.....	44
4.7	Exemplo de resultado 1 do estimador MQ por batelada ....	45
4.8	Exemplo de resultado 2 do estimador MQ por batelada ....	46

4.9	Exemplo de resultado 1 do estimador MQ por batelada com janela $J = J_{min}$ .	47
4.10	Exemplo de resultado 1 do estimador MQ por batelada com janela $J = J_{min} + 6$ .	47
4.11	Exemplo de resultado 1 MQ por batelada com janela $J = J_{min} + 12$ .	48
4.12	Exemplo de resultado 2 do estimador MQ por batelada com janela $J = J_{min}$ .	49
4.13	Exemplo de resultado 2 do estimador MQ por batelada com janela $J = J_{min} + 6$ .	49
4.14	Exemplo de resultado 2 do estimador MQ por batelada com janela $J = J_{min} + 12$ .	50
4.15	Exemplo de resultado 1 do estimador MQ Recursivo com fator de esquecimento $\lambda = 0.95$ .	51
4.16	Exemplo de resultado 1 do estimador MQ Recursivo com fator de esquecimento $\lambda = 0.97$ .	51
4.17	Exemplo de resultado 1 do estimador MQ Recursivo com fator de esquecimento $\lambda = 0.99$ .	52
4.18	Exemplo de resultado 2 do estimador MQ Recursivo com fator de esquecimento $\lambda = 0.95$ .	53
4.19	Exemplo de resultado 2 do estimador MQ Recursivo com fator de esquecimento $\lambda = 0.97$ .	53
4.20	Exemplo de resultado 2 do estimador MQ Recursivo com fator de esquecimento $\lambda = 0.99$ .	54

# LISTA DE TABELAS

4.1	Proporção de bons modelos em treino e previsão considerando o total de dados .....	55
4.2	Proporção de bons modelos por estimador.....	55
4.3	Proporção de bons modelos para cada valor de previsão.....	56



## Siglas

ARX	Autorregressivo com entradas exógenas
ARMAX	Autorregressivo com entradas exógenas e média móvel
NRMSE	Raiz do erro quadrático médio normalizado
OPR	Taxa de produção de óleo
WPR	Taxa de produção de água
WIR	Taxa de injeção de água
BHP	Pressão de fundo de poço
EnKF	<i>Ensemble Kalman Filter</i>
MRST	<i>MATLAB Reservoir Simulation Toolbox</i>
OPM	<i>Open Porous Media</i>
RAO	Razão Água Óleo
RGO	Razão Gás Óleo
BSW	Água e sedimentos básicos

# Capítulo 1

## Introdução

### 1.1 Contextualização

A exploração de reservatórios de petróleo é algo que vem sendo feito há anos e que evoluiu muito ao longo do tempo. No entanto, mesmo com todos os avanços já alcançados, a maior parte do petróleo encontrado no mundo é considerada irrecuperável devido à baixa taxa de recuperação dos métodos de recuperação convencionais [1].

Completamente reconhecido pela indústria petrolífera, esse fato incentiva o desenvolvimento de pesquisas envolvendo a exploração de reservatórios dado o seu potencial enorme de retorno financeiro, tanto que diversos estudos são financiados pela indústria todos os anos, no Brasil e no mundo.

Aliado a isso, a modelagem de sistemas é uma disciplina tradicional e que já foi objeto de estudos no contexto de exploração de reservatórios [2, 3, 4]. O estudo de métodos de identificação aplicados ao problema de modelagem das taxas de produção de poços de petróleo é algo relevante.

### 1.2 Definição do problema

O processo de exploração de reservatórios, em sua operação normal, tem algumas características que tornam extremamente difícil a estimação de bons modelos, representando assim um grande desafio a busca por melhores estratégias que se aproveitem dos conceitos relacionados a teoria de identificação de sistemas. Tais características são:

- **Não linearidade do processo:** Reservatórios de petróleo normalmente envolvem grandes áreas subterrâneas, e o fluxo do óleo no subsolo passa por rochas de diferentes composições, porosidades e permeabilidades [1]. Além disso, ao se iniciar a produção, ocorrem alterações na pressão do reservatório, o que incorre em alterações no volume de fluidos, além da mudança de fase de compostos químicos, como a vaporização ou liquefação de fluidos. Ressalta-se também que, se algum tipo de fluido for introduzido no reservatório de

forma a manipular seu estado, os efeitos dessa entrada na saída do sistema apresentará, muito provavelmente, uma natureza não linear.

- **Impossibilidade de aplicação de entradas próprias para identificação:** Em se tratando de identificação de sistemas, é comum o uso de sinais de entradas próprios, que excitam o sistema a ser identificado em diferentes frequências em torno do ponto de operação. Alguns trabalhos, como [2], discutem o projeto de sinais de entrada para identificação de modelos de poços produtores de petróleo. Em uma situação real, entretanto, é praticamente impossível ter a liberdade para se realizar tais experimentos. Além disso, a própria aplicação da entrada altera o estado do reservatório, de forma que os resultados obtidos após os testes deixam de ser válidos logo na sequência, devido à realização do teste.
- **Poucas amostras para identificação:** É muito comum que o tempo de amostragem seja muito alto, com vários dias entre uma amostra e outra. Isso acontece pois o processo tem dinâmicas extremamente lentas, não sendo necessário nem interessante amostragens mais frequentes. No entanto, isso faz com que o número de amostras disponíveis para a estimação de parâmetros de modelos seja pequena, o que compromete bastante a qualidade dos modelos obtidos, além de tornar o modelo mais suscetível a ruídos e perturbações.

Devido às dificuldades apresentadas, entre outros motivadores, encontram-se vários trabalhos na literatura relacionada que propõem métodos que utilizam as informações disponíveis do reservatório de forma a se obter modelos que representam bem o comportamento dinâmico do sistema.

Usando um modelo bidimensional muito simples com um poço injetor, um poço produtor e um formato quadrangular, [3] testou 26 entradas diferentes, variando a frequência com que a água era injetada no reservatório, e avaliou, para cada entrada, 5 estruturas de modelagem distintos. Os resultados indicaram que todos os modelos obtiveram fits relativamente baixos no momento do corte de água se comparados com os fits obtidos nos outros momentos. Outra conclusão é que os modelos têm melhor desempenho com entradas que têm baixa frequência e maior amplitude e que os modelos lineares caixa preta podem oferecer boas previsões, embora a falta de interpretação física dos parâmetros obtidos no modelo atrapalhe na tomada de decisão. Apesar de conseguirem boas previsões em um curto prazo, o autor não considera que os modelos caixa preta sejam confiáveis o suficiente para serem usados em *loops* de otimização autônomos.

Ainda no campo teórico, mas com um modelo de reservatório mais complexo, [2] também cria modelos lineares caixa preta para os poços produtores, e investiga o efeito que as entradas têm sobre os resultados obtidos, buscando quais entradas diminuem o erro de predição dos modelos. No trabalho, o autor analisa e categoriza as não linearidades presentes no processo e conclui que, visando mitigar os efeitos das não linearidades sobre o resultado dos modelos lineares, deve-se minimizar o tempo de amostragem e o tamanho dos blocos da *grid* do modelo do reservatório. Além disso, é observado que os efeitos das não linearidades são bem mais presentes no início da exploração, sendo que o ideal é realizar a identificação do reservatório já maturado, mais próximo de um ponto de operação.

Ressalta-se também que o autor utiliza o sinal PRBS, que é uma espécie de ruído branco

bastante conhecido no contexto de identificação de sistemas, em cada entrada do sistema, tomando o cuidado para que não haja correlação entre os sinais de entrada para que o efeito de cada entrada possa ser avaliado separadamente. No trabalho é discutido a definição dos parâmetros do PRBS, como frequência, amplitude e tempo do experimento.

São obtidos ótimos modelos, porém são resultados que dependem do uso das entradas especificamente feitas para identificação do sistema, em torno de um ponto de operação do poço já maturado. Na prática, sacrificar um período de produção do poço para realizar um experimento de identificação é inviável economicamente.

Outro trabalho que também utiliza PRBS para construir modelos lineares caixa preta é encontrado em [4]. Este trabalho utiliza um modelo ARX, identificado por batelada, e o foco é apresentar uma metodologia para definir a melhor estrutura do modelo, que resulte nos menores erros de predição.

Determinar a estrutura ideal de um modelo para identificação não é uma tarefa trivial e, frequentemente, se resume à tentativa e erro. A metodologia apresentada se resume a dividir os dados conhecidos em conjuntos de treino e validação, escolher um ponto de partida nos parâmetros e refinar os parâmetros de acordo com o desempenho do modelo treinado, testando no conjunto de validação. Em resumo, o refinamento dos parâmetros se assemelha bastante com uma busca em profundidade [5].

### 1.3 Objetivos do trabalho

Os trabalhos comentados anteriormente inspiraram e serviram de motivação para o atual trabalho, que irá seguir no caminho de tentar modelar os poços de reservatórios através de modelos lineares caixa preta. O foco do trabalho será em observar as diferenças encontradas entre diferentes métodos de identificação. Também serão analisados os efeitos dos parâmetros de identificação sobre a qualidade de previsões dos modelos.

O objetivo não é, entretanto, desenvolver uma metodologia para otimizar as previsões dos modelos. Os trabalhos anteriores tratam bastante do tema, discutindo diversos aspectos e metodologias que podem ser aplicadas para esse fim. No trabalho atual, o objetivo é evidenciar as diferenças obtidas ao variar, principalmente, o método de identificação utilizado.

No estudo proposto, será utilizado o modelo ARX devido aos resultados positivos observados em trabalhos anteriores, como [4], e será avaliada a qualidade da modelagem de poços usando uma função linear de 1<sup>a</sup> ordem a partir de 3 métodos distintos. As diferentes simulações são selecionadas e classificadas de forma a explicitarem alguma característica específica, visando avaliar os métodos de identificação utilizando essa estrutura linear de grau 1 em diferentes situações.

Os métodos de identificação utilizados serão detalhados nas seções seguintes, mas se resumem ao estimador de Mínimos Quadrados (MQ) por batelada, com janela e recursivo. Além disso, realizaram-se pequenas variações nos parâmetros de cada estimador de forma a se obter uma ideia dos melhores e piores parâmetros.

O principal motivador para a realização deste trabalho foi a esperança de que os métodos de identificação recursivos pudessem obter melhores resultados que os métodos por batelada, os quais foram utilizados nos trabalhos citados anteriormente. Essa esperança se justifica no fato de que métodos recursivos de identificação resultam em modelos mais responsivos e adaptáveis, focados em diminuir o erro de predição do próximo passo, ao invés de uma redução global.

Para se observar as diferenças entre os métodos de identificação, não é necessário construir modelos que se adaptem bem aos dados em questão. Ao contrário, preferiu-se utilizar uma estrutura linear simples, de primeiro grau, mantendo todos os parâmetros da estrutura iguais, variando-se somente o método de identificação e seus parâmetros.

Pela falta de preocupação em se otimizar a estrutura do modelo aos dados observados, os resultados são sub-ótimos em termos de qualidade da predição. No entanto, foi possível observar, bem claramente, a diferença entre os métodos de identificação empregados. Acredita-se que, unindo essas informações com o que já se é conhecido sobre o tema, deve ser possível elevar o nível de qualidade da modelagem de reservatórios usando modelos lineares caixa preta.

Para se realizar esse estudo foram utilizadas simulações de um mesmo reservatório excitado de diferentes formas, de maneira que diferentes características surgiram quanto à produção de óleo. Após um processo de filtragem para selecionar o conjunto de simulações consideradas mais interessantes, tal conjunto foi submetido aos mesmos processos de identificação de sistema de forma que os diferentes resultados possam ser atribuídos às diferenças entre as simulações em si.

## 1.4 Resultados obtidos

Ao final do trabalho, foram obtidas informações acerca do desempenho dos estimadores aqui estudados, comparando uns com os outros e também os relacionando com algumas variáveis do processo.

Em resumo, observou-se um melhor desempenho ao se utilizar estimadores recursivos ou com janela quando comparado com o método por batelada clássico. Além disso, identificaram-se diferenças entre a performance dos estimadores baseado no momento em que se encontra a produção do poço.

Com relação a quais estimadores obtiveram melhores resultados para previsões de curto ou longo prazo, não foi observada uma diferença muito grande com variações dos parâmetros de identificação, porém, mais uma vez, se mostrou uma superioridade para os métodos recursivos.

## 1.5 Apresentação do manuscrito

O manuscrito está dividido em 5 capítulos. O primeiro capítulo se dedica a contextualizar as motivações por trás do trabalho, além de apresentar, de maneira sucinta, os principais objetivos e resultados.

O capítulo 2 apresenta a fundamentação teórica necessária para se compreender os principais pontos do trabalho, como aspectos da exploração de reservatórios de petróleo, além de uma introdução dos modelos e estimadores utilizados, que estão englobados pela disciplina de identificação de sistemas.

No capítulo 3 é apresentada a metodologia do trabalho. Mais precisamente, explicam-se algumas das características dos dados que estão sendo utilizados, o pré-processamento pelo qual eles são submetidos, além de apresentar propriamente os estimadores e configurações de identificação que serão estudadas.

O capítulo 4 lista os resultados obtidos, apresentando, primeiramente, alguns resultados gráficos e, na sequência, um panorama numérico dos resultados obtidos. Por fim, o capítulo 5 conclui alguns dos principais resultados observados a partir dos resultados obtidos.

## Capítulo 2

# Fundamentação Teórica

### 2.1 Exploração de Reservatórios de Petróleo

Segundo [6]:“Petróleo (do latim *petra* = rocha e *oleum* = óleo) é o nome dado às misturas naturais de hidrocarbonetos que podem ser encontradas no estado sólido, líquido ou gasoso, a depender das condições de pressão e temperatura a que estejam submetidas.”

Hidrocarbonetos são compostos orgânicos formados por carbono e hidrogênio. Para a formação do petróleo, é necessário que matéria orgânica esteja em um ambiente livre de processos de oxidação, além condições adequadas de temperatura e pressão e, por se tratar de uma longa cadeia de processos lentos, leva muito tempo.

Devido às condições necessárias, grande parte do petróleo do mundo se origina de microorganismos marinhos e algas, presentes em ambientes de deposição de sedimentos de baixa permeabilidade [1], que manteve a matéria orgânica isolada o suficiente.

A formação de grandes reservatórios de petróleo se dá em áreas que possuam características geológicas tais que o petróleo formado na região migre para um mesmo lugar e fique preso em algum tipo de armadilha geológica. Nesse processo o petróleo fica armazenado no interior de rochas porosas e permeáveis, que facilitam a migração, chamadas de rochas-reservatório, mas com a sua movimentação impedida por rochas de baixa permeabilidade, conhecidas como rochas-selantes [1].

#### 2.1.1 Métodos de Produção e Recuperação

Para se extrair o petróleo da rocha-reservatório, faz-se necessário perfurá-la, abrindo assim um canal para que ela possa subir à superfície. Consequentemente, o petróleo armazenado nela naturalmente irá subir e jorrar na superfície, devido à diferença de pressão entre a superfície e o reservatório. Esse método, entretanto, resulta em um baixo *fator de recuperação* de petróleo, definido como a razão entre o *volume recuperável* sobre o *volume original* de petróleo no reservatório [1].

Uma forma de se aumentar o *fator de recuperação* é através da manutenção da pressão do

reservatório. A pressão do reservatório é medida através da pressão do fundo do poço (BHP, do inglês *Bottom Hole Pressure*). A manutenção do BHP pode ser feita aproveitando as características naturais do reservatório, o que caracteriza os métodos de recuperação primária, ou através de métodos de recuperação secundária.

### 2.1.1.1 Mecanismos de Recuperação Primária

Os reservatórios possuem, inicialmente, uma quantidade de energia armazenada, chamada de *energia primária* [6]. Antes de se iniciar a produção de um reservatório deve ser feito um trabalho de prospecção e estudo do reservatório, de forma a se identificar os mecanismos de recuperação primária disponíveis no reservatório.

De acordo com [1], têm-se, por exemplo, os seguintes dos mecanismos de recuperação primária:

1. **Mecanismo de gás em solução:** Nesse caso, temos um reservatório majoritariamente formado por hidrocarbonetos líquidos bem isolado, como mostra a figura 2.1. Não existem corpos hídricos nem gases acumulados, de forma que toda a energia disponível está armazenada na própria zona do óleo.

O mecanismo de produção consiste em retirar o petróleo do reservatório, o que reduz a pressão do reservatório e, como consequência, os fluídos lá presentes se expandem e dão continuidade à produção. Esse mecanismo é marcado por baixos fatores de recuperação, obtendo somente cerca de 20% [1] do volume original de petróleo do reservatório.

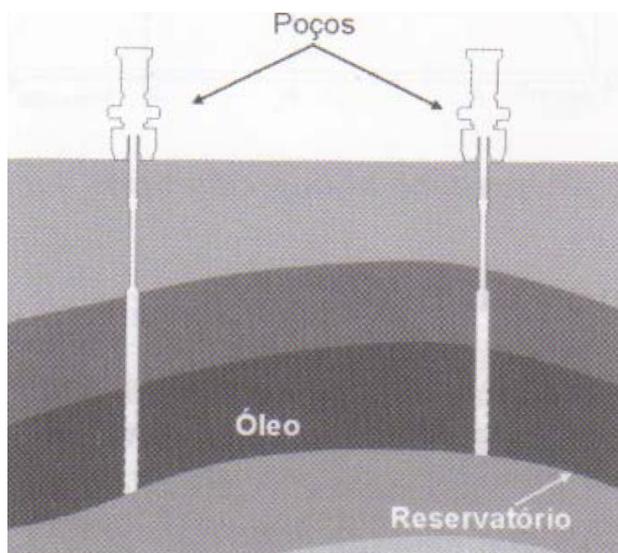


Figura 2.1: Reservatório com mecanismo de gás em solução [1].

2. **Mecanismo de capa de gás:** Nesse caso, o reservatório apresenta hidrocarbonetos em fase líquida e gasosa. Por ser menos denso, é formada uma capa de gás na parte superior do reservatório (Figura 2.2). É possível posicionar os poços de produção de forma a ficarem abaixo da capa de gás e, quando for iniciada a produção, a queda de pressão no reservatório levará à uma expansão da capa de gás, que força o óleo na direção dos poços. Esse método faz

com que a pressão do reservatório caia mais lentamente e resulta em fatores de recuperação médios, tipicamente entre 20% e 30% [1].

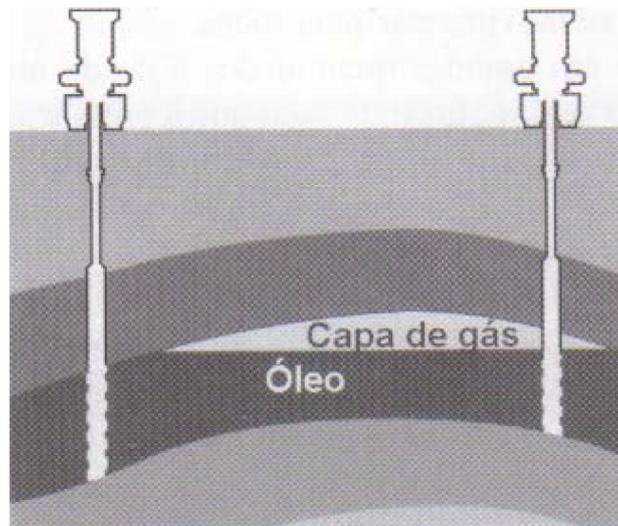


Figura 2.2: Reservatório com mecanismo de capa de gás [1].

- Mecanismo de influxo de água:** Este mecanismo funciona em reservatórios que estão em contato direto com grandes corpos hídricos. Devido à densidade maior, a água fica abaixo do óleo, de forma que o poço de produção deve estar acima do nível da água (Figura 2.3). Ao se iniciar a produção, a queda de pressão faz com que a água se expanda e force o óleo na direção do poço produtor. Esse mecanismo de produção necessita de grandes corpos hídricos para funcionar, pois a compressibilidade da água não é grande. No geral, se obtém fatores de recuperação altos com esse mecanismo, girando em torno de 30% a 40%, em alguns casos chegando a 75% [1].

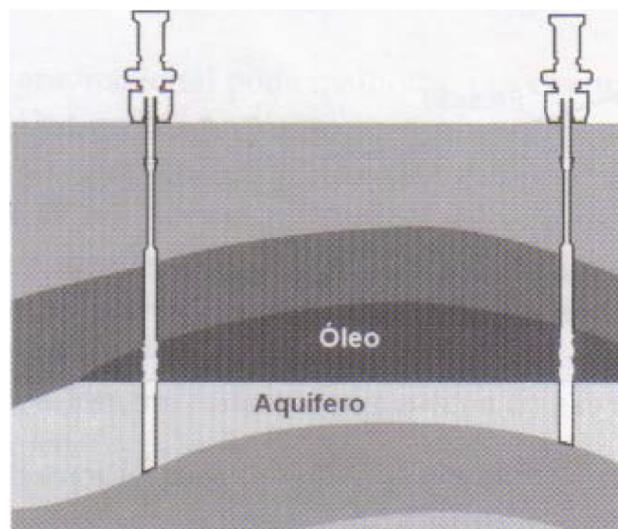


Figura 2.3: Reservatório com mecanismo de influxo de água [1].

Como é possível observar pelos exemplos acima, o mecanismo usado depende do reservatório.

Existem outros mecanismos de produção possíveis e, como é frequente, um mesmo reservatório pode conter diversos mecanismos de produção primários.

### 2.1.1.2 Métodos Convencionais de Recuperação Secundária

Métodos de recuperação secundária são métodos aplicados com o objetivo de suplementar a energia do reservatório, sendo aplicados logo após a fase de recuperação primária, mas também podem ser usados para aumentar o fator de recuperação e/ou acelerar a produção em relação à produção primária [6].

Alguns dos métodos convencionais de recuperação secundária se resumem à injeção de fluidos no reservatório, através de poços chamados de *poços injetores*. Os fluidos injetados podem ser, por exemplo, água, gás, compostos químicos diversos e um misto de injeções. De qualquer forma, os poços injetores normalmente são dispostos de forma a produzir mecanismos que levem o petróleo na direção dos poços produtores.

Com o uso de métodos de recuperação secundárias é possível realizar a manutenção do BHP de um poço produtor por um período bem extenso, através da injeção de mais energia no reservatório. Por ser possível manter e controlar a energia do reservatório, algumas métricas devem ser usadas para se definir o tempo de vida de um poço produtor, como a razão água/óleo que é produzida, e é normal que métricas econômicas predominem neste contexto.

Por fim, o uso de tais métodos eleva significativamente os fatores de recuperação, se comparados com o uso de mecanismos de produção primária, sendo amplamente usados. Além disso, esses métodos dão um controle maior no processo de produção, sendo possível estudar técnicas de controle que otimizem os parâmetros econômicos da operação.

### 2.1.2 Fluidos Produzidos

Além de hidrocarbonetos, que é o principal objetivo exploratório, é comum que outros fluidos venham a tona, como água e gases. A produção de tais fluidos pode ocorrer naturalmente, quando eles já estão presentes no reservatório, ou podem ser resultado da injeção dos mesmos via poços injetores. Além disso, sedimentos sólidos também são produzidos, trazidos pelos fluidos.

No contexto de exploração de reservatórios de petróleo, é comum monitorar o que está sendo produzido através de algumas razões de produção largamente conhecidas na literatura, tais como [1]:

1. **Razão Gás Óleo (RGO):** É a relação entre a vazão de gás e a vazão de óleo produzido por um poço produtor em um dado instante de tempo, medidas na superfície. Uma alta RGO pode significar que o reservatório está depletado, ou que uma grande fração dos componentes do reservatório é volátil.
2. **Razão Água Óleo (RAO):** É a relação entre a vazão de água e a vazão de óleo produzido por um poço produtor em um dado instante de tempo, medidas na superfície. Uma RAO

baixa significa que, na época da coleta desses fluidos, a zona na qual se encontra o poço produtor tem uma quantidade de água irrisória.

3. **Água e Sedimentos Básicos (BSW)**: É o quociente entre a vazão de água mais os sedimentos que estão sendo produzidos e a vazão total de líquidos e sedimentos.

Essas e outras variáveis do processo, medidas em todos os poços, formam a base de dados que constitui o histórico de produção do reservatório. Este histórico de dados, juntamente com o métodos de recuperação discutidos anteriormente, pode ser usados para se propor modelos do processo, projetar malhas de controle, entre outras possibilidades.

### 2.1.3 Corte de água

Ao se utilizar poços injetores para injeção de água nos reservatórios como um método de recuperação secundária, tem-se como desvantagem o aumento da RAO no reservatório. Além disso, a injeção contínua de água forma algo como uma parede de água que empurra o óleo em direção aos poços produtores, mas que, ao alcançá-los, os torna muito improdutivos. O aumento brusco na RAO observada em um poço produtor é chamado de corte de água e é uma importante variável para o controle de injeção de água dos poços injetores, além de ser um indicador do estado atual do reservatório.

Por ser tão significativo, há muito interesse em modelos que possam prever o corte de água [7] e extrair informações relevantes nesse momento [8]. Se trata, entretanto, de um evento complexo e muitas vezes sem precedentes para o poço produtor, sendo difícil de ser modelado.

### 2.1.4 Simuladores de Reservatórios

Por mais que seja desejável utilizar dados de um processo real, tais dados normalmente não são públicos e permanecem sob domínio das empresas petrolíferas. Além disso, para os propósitos do trabalho, é necessário se explorar o mesmo reservatório de diferentes maneiras. Por isso, os dados utilizados provêm de um simulador de reservatórios, porém as simulações são feitas de forma que se busque estar o mais próximo possível de uma exploração real.

Alguns dos simuladores de reservatórios mais conhecidos, atualmente, são o simulador MRST (*MATLAB Reservoir Simulation Toolbox*) [9], um simulador gratuito e de código aberto, o Flow [10], o qual foi desenvolvido pela OPM (*Open Porous Media*), e o Storm [11], para citar alguns. Os dados utilizados neste trabalho foram gerados no simulador Flow.

### 2.1.5 Egg Model

O *Egg Model* é um modelo de reservatório sintético amplamente usado como objeto de estudo em pesquisas relacionadas à reservatórios de petróleo. Alguns exemplos de artigos que o usaram para simular a produção de petróleo são [12], [13] e [14].

Segundo [15], que descreve alguns aspectos do *Egg Model*, a primeira publicação que se refere ao modelo aparece em [16], contendo apenas uma realização determinística do mesmo. Posteriormente, [17] apresentou uma versão do *Egg Model* usando um conjunto de 100 realizações de um reservatório, sendo essa a primeira versão "estocástica" do modelo.

O *Egg Model*, desde então, consiste em um modelo formado por um conjunto de realizações de reservatórios canalizados na forma de campos de permeabilidade discreta modelados com  $60 \times 60 \times 7 = 25.200$  células, dentre as quais somente 18.553 são células ativas. As células não ativas estão todas na parte externa do modelo, dando ao conjunto de células ativas um formato oval, que nomeia o modelo.

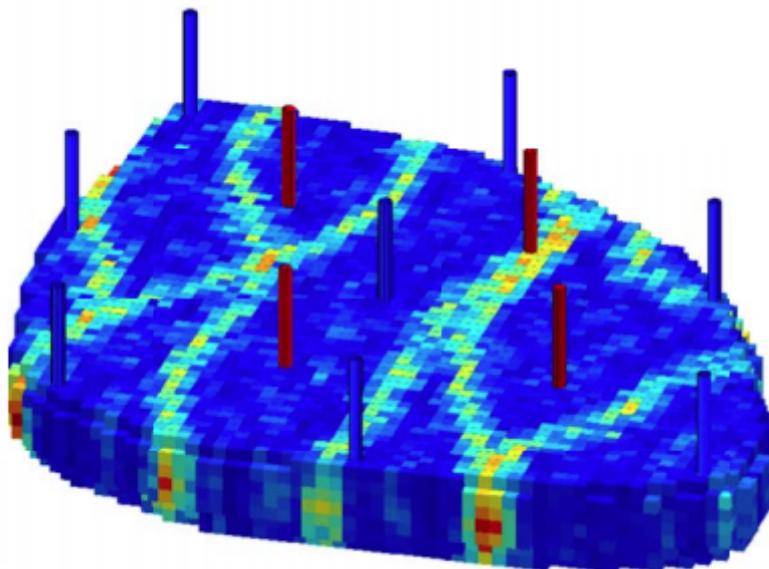


Figura 2.4: Representação do Egg Model, destacando os canais de alta permeabilidade e os poços injetores (em azul) e produtores (em vermelho)

O *Egg Model* é considerado um modelo de reservatório simples, pois não conta com aquíferos ou capas de gás, o que torna a produção primária desprezível [15]. Dessa forma, o principal mecanismo de produção é a injeção de água, feita através de 8 poços injetores que cercam os 4 poços produtores, conforme pode ser visto na Figura 2.4.

## 2.2 Modelagem Matemática

A modelagem matemática é uma área do conhecimento que estuda maneiras de representar sistemas quaisquer utilizando modelos matemáticos. Assim como qualquer modelo, o modelo matemático é uma representação do sistema em questão, simulando algumas de suas características, mas possuindo suas limitações.

Da mesma forma que uma maquete é o modelo de um espaço físico, podendo simular bem a aparência do local, mas sem simular, por exemplo, a resistência física dos materiais presentes na realidade, da mesma forma os modelos matemáticos também procuram simular bem alguns

aspectos do sistema, enquanto ignoram completamente outros aspectos.

Apesar de não serem modelos perfeitos, os modelos matemáticos, de maneira geral, conseguem simular mais realisticamente um sistema que é bem conhecido, onde vários parâmetros reais podem ser informados ao modelo. A modelagem matemática de um sistema bem conhecido é chamada de *modelagem* (ou *identificação*) *caixa branca*. Esse tipo de modelagem também é chamado de *modelagem pela física* ou *natureza do processo* ou ainda *modelagem fenomenológica* ou *conceitual* [18].

Um exemplo de modelagem caixa branca é um circuito elétrico com todos os parâmetros (como o valor das resistências, capacitâncias e indutâncias) conhecidos. Dessa forma, é possível utilizar as equações próprias para sistemas elétricos, inserir os parâmetros adequados, e obter diversas informações do sistema real a partir do seu modelo matemático.

Na prática, porém, nem sempre temos conhecimento de importantes parâmetros do sistema real. Obter o valor de tais parâmetros pode ser custoso ou até impossível, de forma que a modelagem caixa branca nem sempre é possível. Reconhecendo o quão comum essa situação é, outros métodos de identificação de sistemas foram desenvolvidos, sendo conhecidos como *modelagem caixa preta* e *modelagem caixa cinza*.

A *modelagem* (ou *identificação*) *caixa preta* é um conjunto de métodos de identificação de sistemas que pressupõe pouco ou nenhum conhecimento prévio do sistema a ser modelado. Nesse caso, utilizaram-se modelos gerais que são *ajustados* para se encaixarem no sistema sendo identificado.

No caso da modelagem caixa preta, o pouco conhecimento que se tem do sistema pode ser usado para guiar o modelista a escolher o modelo e método mais apropriado para tentar modelar o sistema em questão, mas baseado em características muito gerais do sistema. Por exemplo, o conhecimento do sistema pode guiar o modelista a escolher entre uma representação em tempo contínuo ou em tempo discreto.

Por fim, é chamada de *modelagem* (ou *identificação*) *caixa cinza* a modificação de métodos tradicionalmente conhecidos como caixa preta, buscando incorporar algumas das informações conhecidas do sistema, como seria típico da identificação caixa branca. Isso se reflete como o uso de uma estrutura, modelo ou metodologia de identificação especificamente pensado para aquela situação, tendo como base os conhecimentos da natureza do sistema a ser modelado.

### 2.2.1 Metodologia de Identificação Caixa Preta

As técnicas de identificação caixa preta serão o foco do presente trabalho. Portanto, para melhor compreender o processo de modelagem a ser empregado, serão apresentadas as principais etapas da metodologia, de acordo com [18]:

1. **Coleta de dados:** Independentemente do método de identificação empregado, é sempre interessante buscar o máximo de dados possível do sistema a ser modelado. O mais comum é que se tenha somente os dados de operação 'normal' do sistema. Entretanto, é desejável, se possível, excitar o sistema de forma a maximizar a quantidade de informação sobre a

dinâmica do sistema.

Nessa etapa, se for possível excitar o sistema dessa maneira, existem diversos parâmetros a serem definidos, como a amplitude e a frequência dos sinais de excitação, a forma desses sinais, o tempo de amostragem, entre outros detalhes.

2. **Determinação da representação matemática:** Existem diversas representações matemáticas que podem ser usadas para modelar qualquer sistema. Em se tratando de uma modelagem caixa preta, é necessário decidir, por exemplo, se será usado um modelo linear ou não linear, discreto ou contínuo, e então escolher dentre as várias representações disponíveis em cada uma dessas categorias.

A escolha de um desses modelos nem sempre reflete a verdadeira natureza do sistema. Por exemplo, pode ser escolhido um modelo linear para se modelar um sistema que é sabidamente não linear, porém que está restrito a certos valores de operação, o que permite o uso de um modelo linear identificado em torno daquela região de operação.

Se possível, o uso de um modelo linear é preferível sobre o uso de modelos não lineares devido à complexidade reduzida, o que torna mais simples e menos custoso o processo de identificação, podendo até produzir resultados melhores do que um modelo não linear.

Exemplos de representações matemáticas incluem funções de transferência, matrizes de estados e modelos ARX, ARMAX, entre outros. Apesar de ser possível guiar a escolha de uma representação específica baseado no sistema real, essa escolha depende bastante do que o modelista deseja, sendo, muitas vezes, um processo de tentativa e erro.

O estudo e comparação do desempenho de diferentes representações matemáticas para um determinado problema é um assunto que rende muitos artigos e trabalhos científicos, incluindo o presente trabalho.

3. **Determinação da estrutura do modelo:** Após escolhido um modelo matemático, deve ser escolhido uma estrutura para o modelo. Por exemplo, se for escolhido uma função de transferência, devem ser escolhidos quantos pólos e zeros serão usados para modelar o sistema.

A escolha da estrutura de um modelo é, assim como a escolha do modelo em si, motivo de diversos estudos e comparações. Nesse caso, porém, existem diversas técnicas para auxiliar na escolha da estrutura apropriada.

Um exemplo de metodologia que pode ser seguido para se determinar a melhor estrutura de um modelo ARX no contexto de identificação de poços de petróleo foi proposto em [4], que pode ser visto como um algoritmo de busca em profundidade.

4. **Estimação dos parâmetros:** A etapa atual começa com a escolha de um algoritmo de estimação dos parâmetros. Um dos algoritmos clássicos mais usados é o método dos Mínimos Quadrados (MQ) e suas variantes. A escolha de um algoritmo de estimação de parâmetros, porém, fica restrita à representação matemática escolhida na etapa 2. O método dos Mínimos Quadrados, por exemplo, só pode ser usado se o modelo em questão for *linear nos parâmetros* [18].

A escolha do método de estimação dos parâmetros também é influenciada pelo processo em questão e a forma como se deseja fazer a estimação. Por exemplo, se a identificação for feita em tempo real ou se é desejável que os parâmetros variem no tempo, o uso de estimadores recursivos é preferível.

Também é importante notar que os dados coletados na etapa 1 irão ter um grande impacto na determinação dos parâmetros. Quanto mais informação estiver contida nos dados coletados, melhor será o modelo obtido.

5. **Validação do modelo:** Após a identificação dos parâmetros, é obtido um modelo matemático para o sistema em questão. Como foi possível observar nas etapas anteriores, não existe um único modelo possível, pelo contrário, existem diversas opções de modelos para o mesmo sistema.

Tendo isso em mente, é interessante possuir métricas para avaliar o desempenho de cada um dos modelos, estruturas e estimadores usados para modelar um determinado sistema. Dessa forma é possível verificar se uma determinada combinação de escolhas é significativamente superior às outras.

Existem diversos indicadores possíveis de serem usados, assim como diversos métodos de validação. A escolha de um ou outro depende do que se deseja alcançar com a identificação. Por exemplo, o objetivo pode ser encontrar o modelo mais simples e com maior poder de previsão, de forma que pode ser construída uma função custo que leva em conta esses dois parâmetros (simplicidade e taxa de acertos na previsão) para se definir o melhor modelo.

É possível observar que, em geral, o espaço de soluções possíveis para o problema de se determinar o melhor modelo, estrutura e estimador para um sistema é bem grande. Isso motiva o uso de algoritmos de otimização, algoritmos genéticos, algoritmos de inteligência artificial, entre outros, para se buscar um modelo apropriado.

## 2.3 Representações Lineares em Tempo Discreto

Como discutido na Seção 2.2, métodos de identificação caixa preta partem do pressuposto que se possui pouco ou nenhum conhecimento prévio do sistema a ser identificado. Os métodos de modelagem caixa preta lançam mão, portanto, de representações gerais, que podem descrever uma ampla gama de sistemas, juntamente com métodos de estimação de parâmetros, para ajustar os parâmetros de uma representação generalista para um sistema particular.

Na atual seção iremos discutir algumas representações para sistemas lineares discretizados no tempo. Considere um sistema qualquer, onde estejam disponíveis os sinais de entrada e de saída, que são coletados a uma taxa de amostragem  $T_s$  e formam, respectivamente, as séries temporais  $u(k)$  e  $y(k)$ . Considere também o seguinte modelo geral:

$$A(q)y(k) = \frac{B(q)}{F(q)}u(k) + \frac{C(q)}{D(q)}\nu(k)$$

$$y(k) = \frac{B(q)}{F(q)A(q)}u(k) + \frac{C(q)}{D(q)A(q)}\nu(k) \quad (2.1)$$

$$y(k) = G(q)u(k) + H(q)\nu(k)$$

Sendo que  $q^{-1}$  é o operador de atraso, de modo que  $u(k)q^{-i} = u(k-i)$ , e  $\nu$  é ruído branco. Os demais componentes do modelo são polinômios definidos como segue:

$$\begin{aligned} A(q) &= 1 - a_1q^{-1} - \dots - a_{n_y}q^{-n_y}; \\ B(q) &= b_1q^{-1} - \dots - b_{n_u}q^{-n_u}; \\ C(q) &= 1 - c_1q^{-1} - \dots - c_{n_\nu}q^{-n_\nu}; \\ D(q) &= 1 - d_1q^{-1} - \dots - d_{n_d}q^{-n_d}; \\ E(q) &= 1 - e_1q^{-1} - \dots - e_{n_e}q^{-n_e}. \end{aligned} \quad (2.2)$$

Os polinômios  $H(q)$  e  $G(q)$  são chamados, respectivamente, de funções de transferência do ruído e do processo. Esse modelo geral pode ser representado pelo diagrama de blocos da Figura 2.5

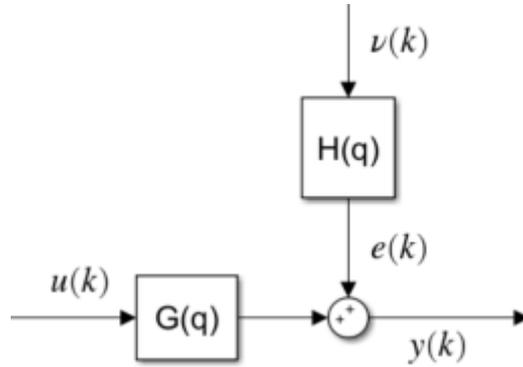


Figura 2.5: Diagrama do modelo geral.

### 2.3.1 Modelo Auto-Regressivo com Entradas Exógenas

O modelo Autorregressivo com Entradas Exógenas ou Autorregressivo com Entradas Externas (ARX) pode ser obtido a partir do modelo geral da equação (2.1) tornando  $C(q) = D(q) = F(q) = 1$ , obtendo o modelo representado na equação (2.3).

$$\begin{aligned} A(q)y(k) &= B(q)u(k) + \nu(k) \\ y(k) &= \frac{B(q)}{A(q)}u(k) + \frac{1}{A(q)}\nu(k) \end{aligned} \quad (2.3)$$

Comparando (2.3) com o modelo geral (2.1), pode-se definir  $G(q) = \frac{B(q)}{A(q)}$  e  $H(q) = \frac{1}{A(q)}$ . A partir das equações do modelo (2.3), é possível também definir os possíveis diagramas de blocos do modelo ARX:

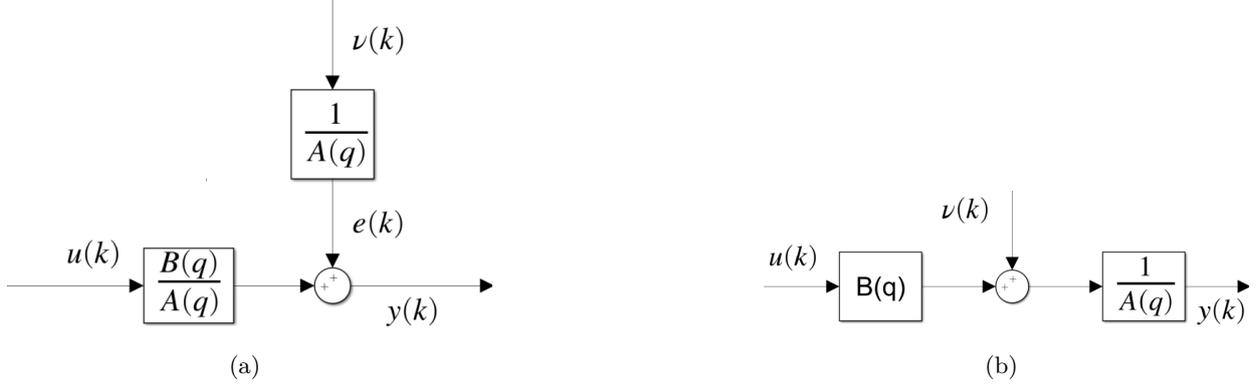


Figura 2.6: Diagramas do modelo ARX

Observando as definições de  $A(q)$  e  $B(q)$  em (2.2), é possível observar que existem 2 parâmetros a serem definidos ao se decidir pelo uso do modelo ARX:

1.  $n_y$ : Define quantos termos passados da saída do sistema serão considerados para descrever a próxima saída.
2.  $n_u$ : Define quantos termos passados da entrada do sistema serão considerados para descrever a próxima saída.

### 2.3.1.1 ARX Multivariável

O modelo ARX também pode ser usado para sistemas com múltiplas entradas e/ou múltiplas saídas. Considere um sistema de  $r$  entradas e  $s$  saídas, a representação ARX multivariável é dada como: [18, 19].

$$y = A_1 \mathbf{y}(k-1) \dots + A_{n_y} \mathbf{y}(k-n_y) + B_1 \mathbf{u}(k-1) \dots + B_{n_u} \mathbf{u}(k-n_u) + \mathbf{e}(k) \quad (2.4)$$

onde  $A_i \in \mathbb{R}^{s \times s}$ ,  $B_i \in \mathbb{R}^{s \times r}$ ,  $\mathbf{y}(k) = [y_1(k) \ y_2(k) \ \dots \ y_s(k)]^T$ ,  $\mathbf{u}(k) = [u_1(k) \ u_2(k) \ \dots \ u_r(k)]^T$  e  $\mathbf{e}(k) = [e_1(k) \ e_2(k) \ \dots \ e_s(k)]^T$ .

### 2.3.1.2 Tempo Morto

É comum que o efeito de uma entrada não tenha efeito sobre o valor da próxima amostra da saída. O tempo morto da entrada  $u$  é definido como o tempo entre o instante da variação da entrada  $u$  e o instante em que o efeito dessa variação é percebido na saída do sistema  $y$ .

Pode-se contabilizar o tempo morto no modelo ARX, como se segue:

$$B(q) = b_1q^{-n_k} - \dots - b_{n_u}q^{-(n_k+n_u)} \quad (2.5)$$

Que também pode ser estendido para o caso multivariável. Dessa forma, os parâmetros para definição de uma estrutura ARX passam a ser:  $n_y$ ,  $n_u$  e  $n_k$ .

### 2.3.2 Estimação de uma constante no modelo

Apesar de ser um modelo linear, é possível realizar a identificação de uma constante no modelo ARX de maneira simples. A identificação dessa constante amplia a usabilidade do modelo e será útil para identificação do modelo de poço, já que busca modelar ao menos uma não linearidade que pode estar presente.

A identificação da constante  $C$  se dá simplesmente pela adição de uma entrada  $u_c$  no modelo, que é uma entrada constante com valor 1. Definindo  $n_b = n_k = 1$  para essa entrada, temos que, em um sistema de múltiplas entradas e uma saída (MISO):

$$\begin{aligned} y = & C u_c(k-1) \\ & + A_1 y(k-1) \dots + A_{n_y} y(k-n_y) \\ & + B_1 \mathbf{u}(k-1) \dots + B_{n_u} \mathbf{u}(k-n_u) + \mathbf{e}(k) \end{aligned} \quad (2.6)$$

que se torna

$$\begin{aligned} y = & C + A_1 y(k-1) \dots + A_{n_y} y(k-n_y) \\ & + B_1 \mathbf{u}(k-1) \dots + B_{n_u} \mathbf{u}(k-n_u) + \mathbf{e}(k) \end{aligned} \quad (2.7)$$

## 2.4 Estimador de Mínimos Quadrados

O estimador de Mínimos Quadrados (MQ) é um dos estimadores de parâmetros mais populares e utilizados [20], possuindo, inclusive, algumas variações. O princípio básico do estimador MQ consiste em encontrar os parâmetros que minimizem o Erro Quadrático Médio (EQM) das observações.

É possível categorizar os estimadores MQ em duas categorias:

1. **Estimadores em batelada:** os parâmetros são estimados usando todos os dados disponíveis de uma só vez, ou seja, todos os dados são utilizados para se calcular uma estimativa somente.
2. **Estimadores recursivos:** os parâmetros estimados são atualizados a cada nova informação. Dessa forma, se existem  $N$  amostras do sinal, os parâmetros serão atualizados  $N$  vezes. Portanto, existirá uma estimativa para cada ponto.

Estimadores em batelada são mais eficientes em processar uma grande quantidade de dados de uma só vez, gerando o modelo final em um só passo. O estimador MQ por batelada também pode

ser usado em sua forma ponderada, o que faz com que algumas informações sejam mais relevantes que outras.

No caso deste trabalho, será usada, além da forma não ponderada, um método de identificação que considera somente uma janela de tempo no momento de se construir o modelo linear a partir do método MQ por batelada.

Quanto aos estimadores recursivos, entre as vantagens que esse tipo de estimador apresenta está a possibilidade do procedimento poder ser realizado em tempo real de maneira muito menos custosa do que o uso de estimadores em batelada, o que é especialmente útil em contextos de controle de sistemas em tempo real [18]. É importante ressaltar que, porém, que a estimação recursiva também pode ser aplicada também de maneira *offline*, ou seja, após a obtenção de todos os dados.

#### 2.4.0.1 Estimador de Mínimos Quadrados por Batelada

Considere o sistema descrito pela equação abaixo [18]:

$$y = \mathbf{x}^T \boldsymbol{\theta} \quad (2.8)$$

em que  $\mathbf{x}$  e  $y$  são valores conhecidos e mensuráveis e  $\boldsymbol{\theta}$  é um vetor de parâmetros que se deseja estimar. Considerando que  $\boldsymbol{\theta} \in \mathbb{R}^n$ , para determinarmos os  $n$  parâmetros, são necessários, no mínimo,  $n$  observações (ou restrições) de (2.8). Dessa forma, pode-se reescrever a equação 2.8 na sua forma matricial:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1^1 & x_1^2 & \dots & x_1^n \\ x_2^1 & x_2^2 & \dots & x_2^n \\ \vdots & \vdots & \vdots & \vdots \\ x_n^1 & x_n^2 & \dots & x_n^n \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \quad (2.9)$$

$$\mathbf{y} = X\boldsymbol{\theta}$$

Portanto, tem-se que:

$$\boldsymbol{\theta} = X^{-1}\mathbf{y} \quad (2.10)$$

Agora considere que, apesar de  $y$  ser mensurável, há sempre um erro de medição associado, de forma que a equação (2.8) passa a ser reescrita como:

$$y = \mathbf{x}^T \hat{\boldsymbol{\theta}} + \xi \quad (2.11)$$

onde  $\hat{\boldsymbol{\theta}}$  é o vetor de parâmetros estimável e  $\xi$  é o erro de estimação, considerado aleatório. Nesse caso, devido à presença de ruído de estimação, é necessário haver mais do que  $n$  observações de  $y$ , de forma a mitigar os efeitos do ruído.

A forma matricial da equação (2.11), considerando um sistema sobredeterminado, ou seja, com  $N > n$  observações, é dado por:

$$\mathbf{y} = X\hat{\boldsymbol{\theta}} + \boldsymbol{\xi} \quad (2.12)$$

onde  $\boldsymbol{\xi} \in \mathbb{R}^n$  é chamado de vetor de resíduos, representando o erro de medição em cada amostra.

O estimador MQ busca estimar  $\hat{\boldsymbol{\theta}}$  de forma que, além de satisfazer (2.12), o erro de medição seja o menor possível. Para isso, define-se a função custo a ser minimizada como o somatório do quadrado dos erros:

$$J_{MQ}(\hat{\boldsymbol{\theta}}, Z^N) = \sum_{i=1}^N \xi(i)^2 = \boldsymbol{\xi}^T \boldsymbol{\xi} = \|\boldsymbol{\xi}\|^2 \quad (2.13)$$

sendo importante notar que o vetor  $\hat{\boldsymbol{\theta}}$  que minimiza a função custo depende do conjunto de dados  $Z^N$  utilizado.

Substituindo  $\boldsymbol{\xi}$  de (2.12) em (2.13), tem-se:

$$\begin{aligned} J_{MQ} &= (\mathbf{y} - X\hat{\boldsymbol{\theta}})^T (\mathbf{y} - X\hat{\boldsymbol{\theta}}) \\ &= \mathbf{y}^T \mathbf{y} - \mathbf{y}^T X\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}^T X^T \mathbf{y} + \hat{\boldsymbol{\theta}}^T X^T X \hat{\boldsymbol{\theta}} \end{aligned} \quad (2.14)$$

Para minimizar a função custo  $J_{MQ}$  com relação a  $\hat{\boldsymbol{\theta}}$ , é necessário fazer  $\frac{\partial J_{MQ}}{\partial \hat{\boldsymbol{\theta}}} = 0$  e  $\frac{\partial^2 J_{MQ}}{\partial \hat{\boldsymbol{\theta}}^2} > 0$ . Portanto, temos que:

$$\begin{aligned} \frac{\partial J_{MQ}}{\partial \hat{\boldsymbol{\theta}}} &= -(\mathbf{y}^T X)^T - X^T \mathbf{y} + (X^T X + X^T X) \hat{\boldsymbol{\theta}} \\ &= -X^T \mathbf{y} - X^T \mathbf{y} + 2X^T X \hat{\boldsymbol{\theta}} \end{aligned} \quad (2.15)$$

e

$$\frac{\partial^2 J_{MQ}}{\partial \hat{\boldsymbol{\theta}}^2} = 2X^T X. \quad (2.16)$$

Observa-se que (2.16) é positiva por definição, portanto a condição  $\frac{\partial^2 J_{MQ}}{\partial \hat{\boldsymbol{\theta}}^2} > 0$  está satisfeita. Igualando (2.15) a zero, se obtém:

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= [X^T X]^{-1} X^T \mathbf{y} \\ &= X^\dagger \mathbf{y} \end{aligned} \quad (2.17)$$

onde  $X^\dagger$  representa a pseudoinversa de  $X$ . A equação (2.17) representa, portanto, o vetor de parâmetros  $\hat{\boldsymbol{\theta}}$  tal que satisfaça a equação (2.12) garantindo o menor erro de medição possível.

#### 2.4.1 Estimando modelo ARX usando MQ

O estimador MQ pode ser usado para estimar os parâmetros de um modelo ARX (discutido na seção 2.3.1). Para isso, consideramos o modelo da saída sendo dada por:

$$y = \psi \hat{\theta} + \xi, \quad (2.18)$$

onde  $\psi$  é chamado de *vetor de regressores*. No modelo ARX, ele é definido como

$$\psi(k-1) = [y(k-1) \ y(k-2) \ \dots \ y(k-n_y) \ u(k-1) \ u(k-2) \ \dots \ u(k-n_u)]^T \quad (2.19)$$

que também pode ser estendido para um sistema MISO (múltiplas entradas e uma saída) com  $r$  entradas:

$$\psi(k-1) = [y(k-1) \ \dots \ y(k-n_y) \ u_1(k-1) \ \dots \ u_1(k-n_{u_1}) \ \dots \ u_r(k-1) \ \dots \ u_r(k-n_{u_r})]^T. \quad (2.20)$$

Portanto, aplicando (2.17) em (2.18):

$$\hat{\theta}_{MQ} = [\psi^T \psi]^{-1} \psi^T \mathbf{y}. \quad (2.21)$$

que representa o estimador Mínimos Quadrados.

#### 2.4.2 Mínimos Quadrados Recursivo com Fator de Esquecimento

O estimador de Mínimos Quadrados Recursivo busca cumprir os mesmos objetivos do método por batelada, porém processando um passo de cada vez, criando um modelo que se adapta a cada passo para se moldar aos dados que surgem.

Neste trabalho será usado o estimador MQ Recursivo com Fator de Esquecimento, que nada mais é que uma forma ponderada do estimador MQ Recursivo padrão. É importante adotar o fator de esquecimento para que se tire proveito da adaptabilidade do estimador recursivo neste caso em que se busca simular o comportamento de um sistema não linear e de ordem maior que a estrutura adotada.

O modelo a ser adotado é o mesmo apresentado anteriormente, na equação 2.18, porém na sua forma escalar. Portanto, para cada instante de tempo  $k$ , tem-se que:

$$y(k) = \psi_k^T \theta_{k-1} + \xi_k \quad (2.22)$$

que pode ser estimado por

$$\hat{y}(k) = \psi_k^T \hat{\theta}_{k-1} \quad (2.23)$$

Aqui,  $\psi_k$  é o vetor de regressores do instante  $k$ , ou seja, os sinais do processo que se tem acesso até o instante  $k-1$ . Considerando um processo com  $n$  entradas, temos:

$$\psi_k = \begin{bmatrix} y(k-1) \\ y(k-2) \\ \vdots \\ y(k-n_y) \\ u_1(k-1) \\ u_1(k-2) \\ \vdots \\ u_1(k-n_u) \\ \vdots \\ u_n(k-1) \\ \vdots \\ u_n(k-n_u) \end{bmatrix} \quad (2.24)$$

onde  $n_y$  e  $n_u$  são os parâmetros usados no modelo ARX. Para cada regressor existe um parâmetro no vetor  $\hat{\theta}_{k-1}$ , que resultam na saída estimada da equação 2.23.

Em contraste com o método de MQ por batelada, no estimador de MQ recursivo o vetor de parâmetros é corrigido a cada passo, considerando o erro de estimativa  $\epsilon = y(k) - \hat{y}(k)$  e um ganho  $K_k$ , chamado de *ganho de Kalman*, que também é atualizado a cada instante de tempo k. Portanto, no instante de tempo k,  $\hat{\theta}_k$  é calculado como:

$$\hat{\theta}_k = \hat{\theta}_{k-1} + K_k \cdot \epsilon \quad (2.25)$$

É possível observar que, se o erro  $\epsilon$  for nulo,  $\hat{\theta}_{k-1}$  é considerado um modelo perfeito e não é alterado. Se o erro não for nulo, a magnitude da correção depende da magnitude do erro e também da matriz de ganho  $K_k$ . Ressalta-se também que o valor de  $K_k$  depende da confiabilidade que se tem no parâmetros de  $\hat{\theta}$ . Essa confiabilidade é dada pela matriz de covariância  $P_{k-1}$ , onde está quantificada a covariância dos parâmetros.

Em resumo, o método de identificação recursiva pode ser dividido em duas partes: a previsão da próxima saída (ver 2.26), que usa os dados disponíveis até o instante anterior, e a atualização dos parâmetros (ver 2.27), assim que os dados do instante atual estão disponíveis.

$$K_k = \frac{P_{k-1}\psi_k}{\psi_k^T P_{k-1} \psi_k + \lambda}$$

$$\hat{y}(k) = \psi_k^T \hat{\theta}_{k-1} \quad (2.26)$$

Parte 1: previsão

$$\hat{\theta}_k = \hat{\theta}_{k-1} + K_k[y(k) - \hat{y}(k)]$$

$$P_k = \frac{1}{\lambda} \left( P_{k-1} - \frac{P_{k-1}\psi_k\psi_k^T P_{k-1}}{\psi_k^T P_{k-1}\psi_k + \lambda} \right) \quad (2.27)$$

Parte 2: atualização

O algoritmo de identificação recursiva com fator de esquecimento completo é apresentado na equação 2.28:

$$K_k = \frac{P_{k-1}\psi_k}{\psi_k^T P_{k-1}\psi_k + \lambda}$$

$$\hat{y}(k) = \psi_k^T \hat{\theta}_{k-1} \quad (2.28)$$

$$\hat{\theta}_k = \hat{\theta}_{k-1} + K_k[y(k) - \hat{y}(k)]$$

$$P_k = \frac{1}{\lambda} \left( P_{k-1} - \frac{P_{k-1}\psi_k\psi_k^T P_{k-1}}{\psi_k^T P_{k-1}\psi_k + \lambda} \right)$$

Para um dado instante de tempo, a assimilação de um novo conjunto de dados forma um novo vetor de regressores  $\psi_k$  que atualiza os parâmetros  $\hat{\theta}_k$  levando em conta o erro da estimativa  $\hat{y}(k)$  ponderado por um ganho  $K_k$ . Se o modelo passa a ter uma boa taxa de acertos a matriz de covariância  $P_k$  reflete uma maior confiança nos parâmetros atuais com valores pequenos em sua diagonal principal.

### 2.4.3 Inicialização do algoritmo recursivo

A inicialização dos parâmetros parte do princípio que, inicialmente, não temos nenhuma certeza dos parâmetros. Portanto, é possível utilizar quaisquer valores em  $\hat{\theta}(0)$  e inicializar a matriz de covariâncias  $\mathbf{P}(0)$  com valores altos na diagonal principal, na ordem de  $10^6$ , de forma a refletir a incerteza nos parâmetros. Dessa forma o algoritmo entende que a incerteza sobre os parâmetros atuais é alta, o que leva a altos valores de correção e acelera a convergência de  $\hat{\theta}$  para os valores reais.

# Capítulo 3

## Metodologia

O atual capítulo tem por objetivo apresentar a metodologia proposta nesse manuscrito. A seguir serão apresentados tanto os dados que foram utilizados como os estimadores e parâmetros que serão considerados nos testes, além de algumas considerações acerca dessas escolhas.

### 3.1 Dados Utilizados na Identificação

Para se atingir o objetivo deste trabalho optou-se por utilizar apenas um modelo de reservatório, porém excitado de formas distintas. Para melhor simular uma exploração real, não foram usadas entradas especificamente desenhadas para identificação de sistemas, pelo contrário, as entradas usadas seguiram uma lógica de otimização de produção, como é comum em explorações reais.

Usando um único modelo de reservatório é possível garantir uma consistência entre os dados e os resultados obtidos, uma vez que todos eles seguem, na realidade, o mesmo modelo. A diferença é que, em cada simulação, excitou-se o reservatório de maneira distinta, revelando dinâmicas diferentes em cada uma.

Os dados utilizados no trabalho vêm de simulações do *Egg Model* utilizando o simulador *Flow* (ver seções 2.1.4 e 2.1.5). Sejam as entradas do sistema as taxas de injeção de água pelos poços injetores (WIR) e os valores de pressão no fundo dos poços produtores (BHP), realizaram-se mais de 100 simulações distintas a partir de variações dessas entradas.

As informações coletadas em cada simulação foram:

- Taxa de produção de óleo (OPR, do inglês *Oil Production Rate*), aferida nos poços produtores;
- Taxa de produção de água (WPR, do inglês *Water Production Rate*), aferida nos poços produtores;
- Pressão no fundo do poço (BHP, do inglês *Bottom Hole Pressure*), aferida nos poços produtores;

- Taxa de injeção de água (WIR, do inglês *Water Injection Rate*), aferida nos poços injetores.

A taxa de amostragem desses sinais é de 30 dias. Conforme discutido anteriormente, o *Egg Model* possui 4 poços produtores (PROD1, PROD2, PROD3 e PROD4) e 8 poços injetores de água (INJECT1 a INJECT8).

Apesar de terem sido feitas mais de 100 simulações, essas simulações foram filtradas e, no fim, apenas 16 simulações foram usadas para gerar os resultados finais. A filtragem foi feita de forma a diminuir o tempo de processamento computacional que envolveria identificar, ponto a ponto, mais de 100 simulações. Os pontos de interesse que foram buscados para selecionar os dados foram os seguintes:

- **Tempo de vida do poço:** O tempo de vida de um poço é definido por quanto tempo o poço passa produzindo. O fim da vida do poço é decretado quando o custo de se manter o poço em produção é mais caro do que os custos de manutenção da produção daquele poço.

Foram selecionadas as simulações com os poços que possuíram o maior e o menor tempo de vida, dentre todas as simulações. Respectivamente, esses tempos de vida foram de 3600 e 450 dias.

- **Instante de *Water Cut* no poço:** O instante de *Water Cut* (WC) de um poço é definido pelo instante em que o poço passa a produzir água, além de óleo. Esse momento é uma marca significativa na produção do poço e muda significativamente a dinâmica de produção.

Foram selecionadas as simulações com os poços que apresentaram o WC mais precoce, com 150 dias, e mais tardio, com 810 dias. Também foi analisada a razão entre o instante de WC e o tempo de vida dos poços, para selecionar as simulações com WC mais precoce e tardio com relação ao tempo de vida, mas as simulações que se destacaram foram as mesmas das selecionadas anteriormente. A simulação com WC mais precoce ocorre com 6% da vida do poço completa, enquanto que a mais tardia ocorre com 54% da vida do poço completa.

- **Variações nos parâmetros de entrada:** São consideradas variáveis de entrada do sistema as taxas de injeção de água dos poços injetores, além do BHP dos poços produtores. Foram selecionadas simulações buscando diferentes casos para as variações nesses parâmetros.

Portanto, entre as simulações usadas, estão simulações onde só houve variação na injeção de água, ou que só registrou variação no BHP dos poços, ou que houve variação em todos os parâmetros. Dessa forma, as simulações selecionadas fornecem uma ampla gama de situações onde o reservatório foi excitado de maneiras distintas.

Quanto à escolha da entrada para o modelo, após alguns testes, definiu-se que a melhor maneira de se modelar era usando apenas o sinal de BHP do próprio poço produtor. Essa escolha foi feita após observar que o uso de outras entradas pioravam bastante os resultados da modelagem. Suspeita-se que isso é causado por dois motivos, o primeiro é que o BHP do poço produtor é a única variável que está, de fato, diretamente correlacionada com a produção que está sendo modelada. Outras variáveis de entrada, como o BHP de outros

poços ou a WIR dos poços injetores, apesar de influenciarem a produção do poço, não se correlacionam diretamente.

O segundo motivo é que, com mais variáveis de entrada, o tamanho do modelo cresce e o número de parâmetros a serem estimados aumenta, tornando mais difícil a determinação de um modelo ótimo, ainda mais considerando a baixa quantidade de dados disponíveis para convergência dos valores dos parâmetros. Portanto, o uso de mais variáveis de entrada resultam em modelos grandes, complexos e com entradas que não possuem uma alta correlação com a saída em vários instantes de tempo.

Para se ter uma ideia melhor acerca da natureza dos dados coletados, serão apresentados 2 exemplos que foram selecionados das 16 simulações usadas.

### **3.1.1 Exemplo de Simulação 1**

Um primeiro exemplo de simulação é mostrado nas Figuras 3.1-3.4. Nessa simulação, a taxa de injeção de água nos 8 poços injetores foi mantida constante, como pode ser constatado na Figura 3.3. As taxas de produção de óleo e água podem ser observadas nas Figuras 3.1 e 3.2, respectivamente.

Com uma taxa de injeção de água mantida constante, as taxas de produção de água crescem rapidamente, com destaque para o poço PROD2 que registrou a produção de grandes volumes de água e óleo.

Em contrapartida com as taxas de injeção de água constantes, a pressão de fundo de poço dos poços produtores não foi mantida constante, sofrendo um declínio lento e gradual, que pode ser observado na Figura 3.4. O fim da vida do poço, que chega no momento em que o poço se torna economicamente inviável, é marcado por uma queda abrupta no BHP do poço, que simboliza o seu fechamento e, conseqüentemente, o fim da produção. Mais uma vez o destaque para o poço PROD2, que teve o maior tempo de vida dentre os poços produtores.

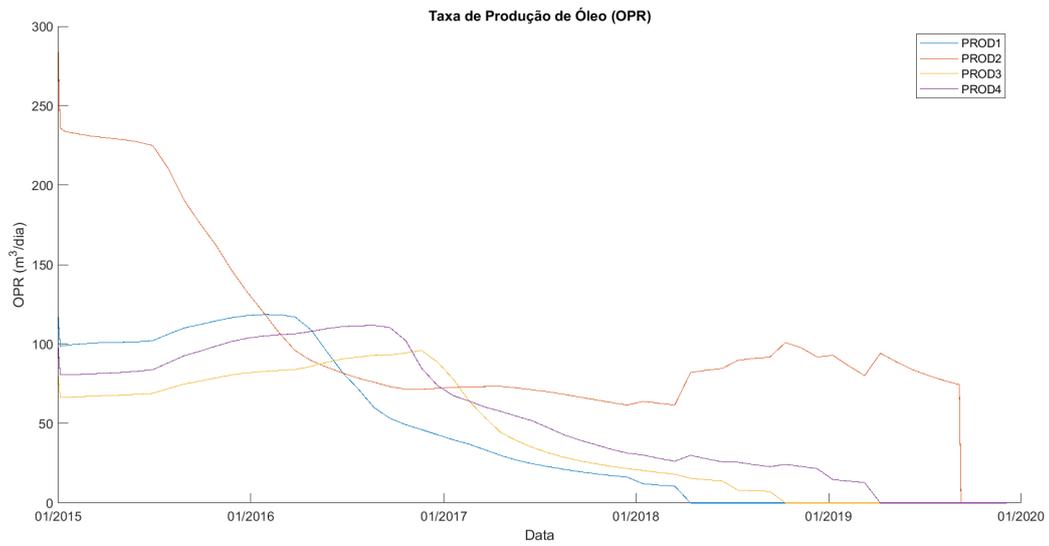


Figura 3.1: Taxa de produção de óleo para os poços produtores na simulação 1.

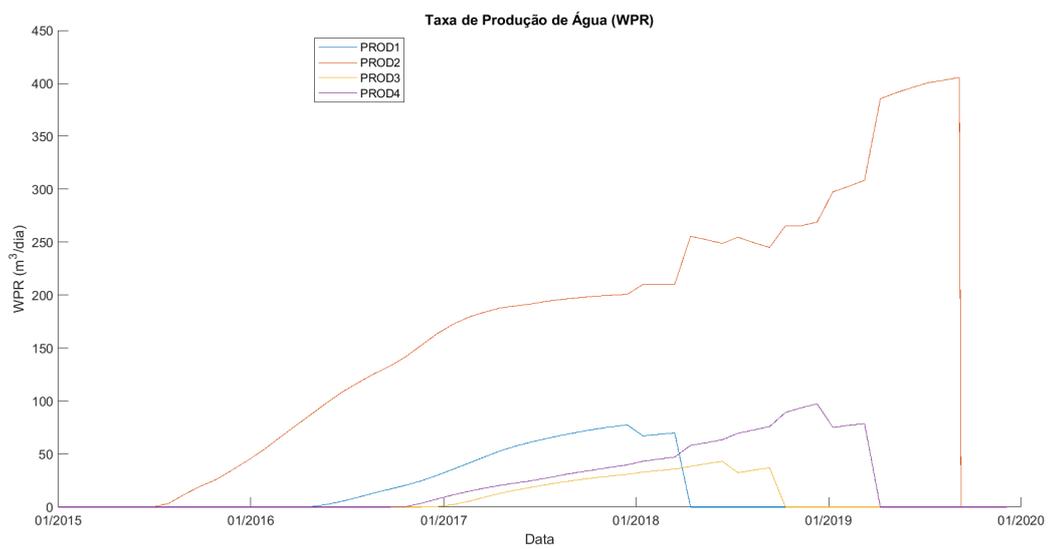


Figura 3.2: Taxa de produção de água para os poços produtores na simulação 1.

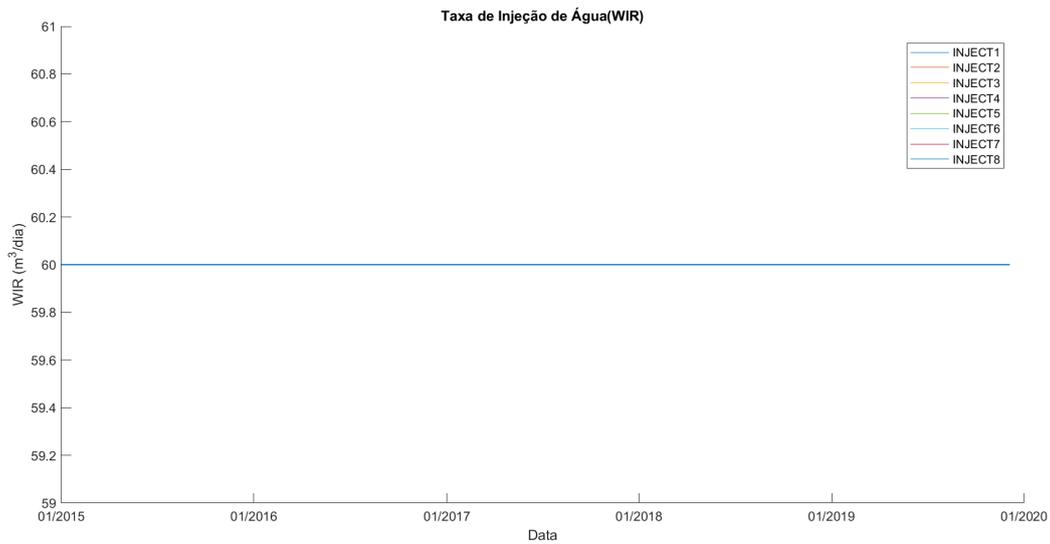


Figura 3.3: Taxa de injeção de água para os poços injetores na simulação 1.

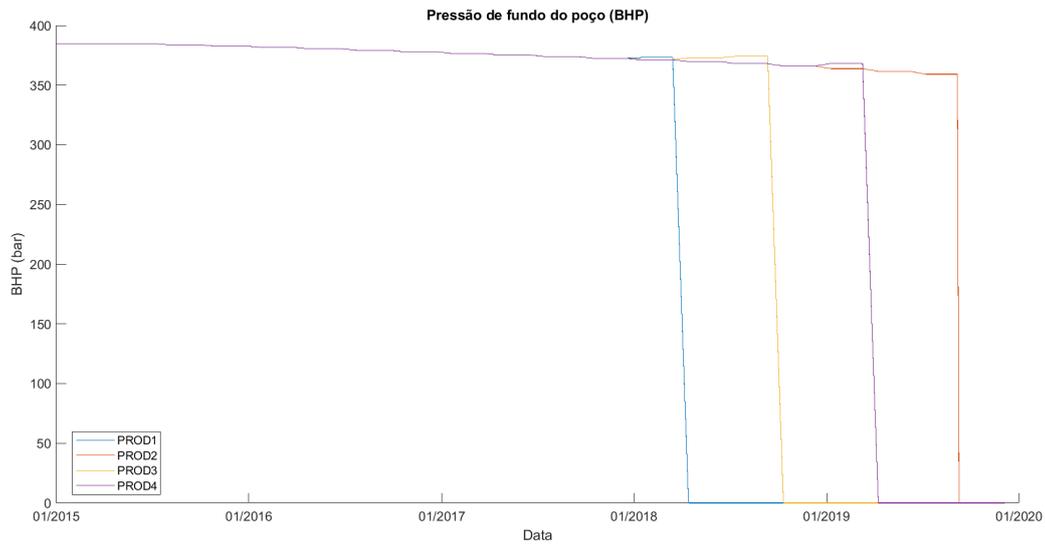


Figura 3.4: Pressão de fundo do poço para os poços produtores na simulação 1.

### 3.1.2 Exemplo de Simulação 2

O segundo exemplo de simulação, exibido nas Figuras 3.5-3.8, mostra uma exploração completamente diferente do exemplo anterior, uma vez que as taxas de injeção de água nos poços injetores variou, seguindo um padrão descendente em escada, como é possível visualizar na Figura 3.7. Além disso, a Figura 3.8 mostra um BHP estável até o fim da vida do poço, que representa a única variação dessa variável nos poços produtores.

As curvas de produção de óleo e água, nas Figuras 3.5 e 3.6 respectivamente, continuam trazendo um destaque para o poço PROD2, assim como foi no exemplo anterior, o que faz sentido,

já que tratam do mesmo modelo de reservatório. Porém, apesar das semelhanças, é possível perceber o quanto o método de exploração impacta nas curvas de produção, pois são bem diferentes das curvas de produção do exemplo anterior.

Uma das diferenças mais notáveis é o tempo de produção dos poços produtores, que é bem menor no primeiro caso. O poço PROD2, por exemplo, ganha em torno de 4 anos do primeiro para o segundo exemplo. Esse tempo a mais pode ser explicado pelas menores taxas de injeção de água, o que torna a exploração mais lenta e duradoura.

Uma exploração mais longeva não significa, necessariamente, uma exploração maior, melhor ou mais lucrativa. Em termos de produção de óleo, por exemplo, é possível observar taxas de produção muito maiores do poço PROD2 no exemplo 1 do que no 2. Isso também pode ser explicado pelas taxas de injeção de água diferentes.

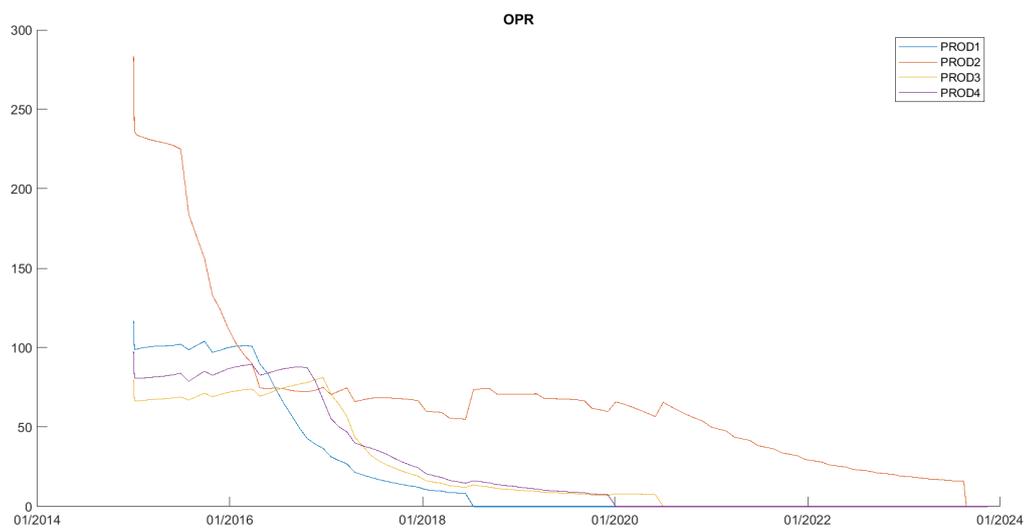


Figura 3.5: Taxa de produção de óleo para os poços produtores na simulação 2.

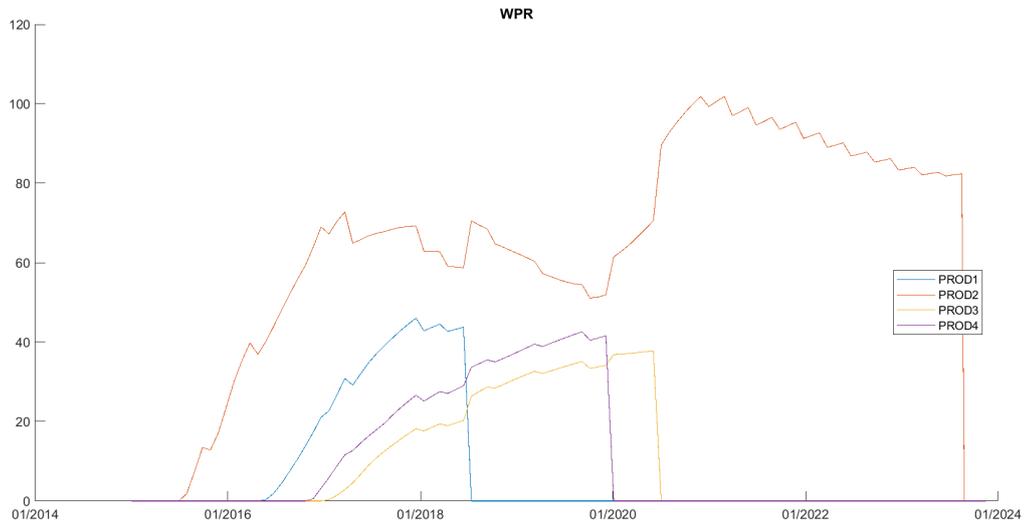


Figura 3.6: Taxa de produção de água para os poços produtores na simulação 2.

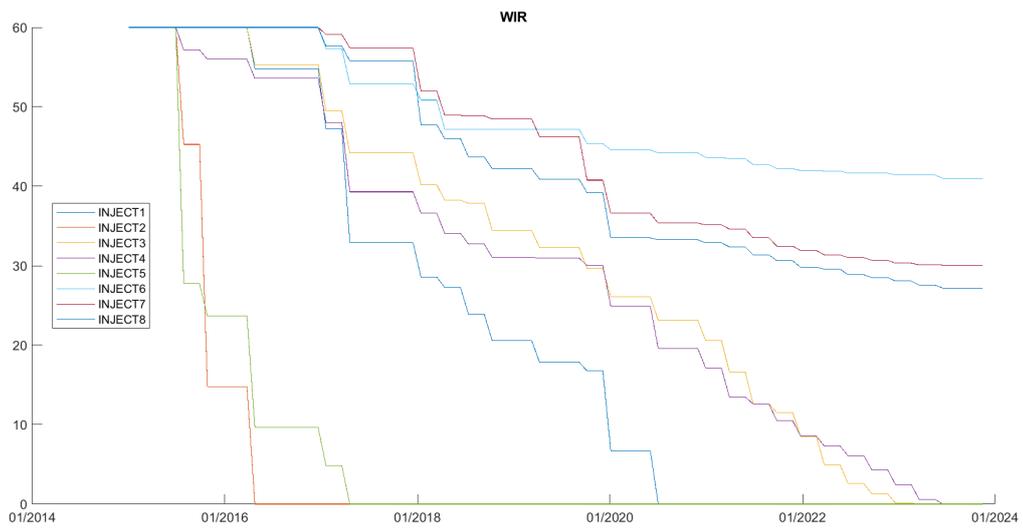


Figura 3.7: Taxa de injeção de água para os poços injetores na simulação 2.

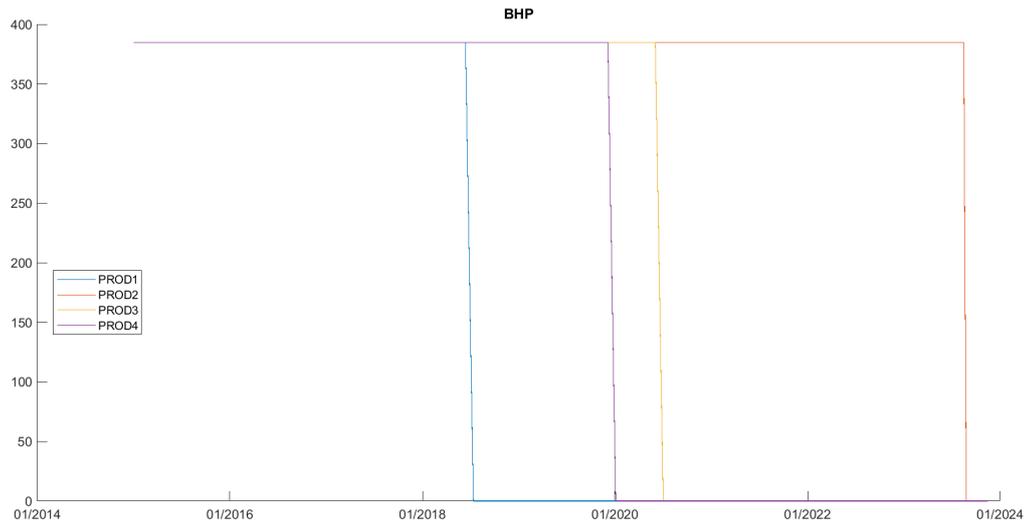


Figura 3.8: Pressão de fundo do poço para os poços produtores na simulação 2.

É possível observar, pelos exemplos apresentados, que existem várias características que os distinguem, tendo como similaridade somente o fato de serem feitos usando o mesmo modelo, no mesmo simulador. Com as simulações usadas, se conta com uma variedade de curvas de produção de óleo e água, diferentes comportamentos de WIR e BHP, além de poços com diferentes tempos de vida e momentos de *water cut*. Dessa forma, pode-se comparar o comportamento de uma estrutura de identificação em diferentes casos e contextos de produção.

## 3.2 Estrutura e Estimadores Utilizados

A estrutura usada foi um modelo ARX (ver Seção 2.3.1) MISO. Além disso, acrescentou-se, em todas as identificações, uma entrada constante, cujo valor é 1, para identificação de uma constante no processo, conforme explicado na Seção 2.3.2. Quanto aos métodos de estimação dos parâmetros do modelo, foram utilizados 3 métodos distintos:

1. **Mínimos Quadrados (MQ) por batelada:** Usa-se o estimador MQ para estimar o modelo ARX (ver Seção 2.4.1), por batelada, ou seja, de uma só vez. Nesse caso, considera-se todos os dados conhecidos para se estimar os parâmetros. Essa estratégia foi escolhida por ser uma estratégia de identificação clássica e simples, fornecendo resultados confiáveis para comparação com os demais métodos.
2. **MQ por batelada com janela:** Usa o mesmo esquema de estimação de parâmetros usando MQ por batelada, porém o estimador utiliza somente os dados disponíveis dentro de uma certa janela de tempo. Por exemplo, caso seja definida uma janela de tamanho 48, o algoritmo de MQ por batelada se limitará a utilizar as últimas 48 amostras disponíveis, ignorando todo o histórico passado fora dessa janela de tempo.

Esse método de estimação foi escolhido pois, devido a não linearidade bem conhecida do sistema, não se espera que a identificação por batelada clássica tenha bons resultados ao considerar todo o histórico passado. O uso de uma janela de tempo é uma forma de simular um fator de esquecimento dentro do método por batelada e se espera que esse método tenha um desempenho geral melhor que o método de MQ por batelada clássico.

3. **MQ recursivo com fator de esquecimento:** O método de Mínimos Quadrados Recursivo com fator de esquecimento possui uma maior capacidade de adaptação dos seus parâmetros. O uso deste estimador parte da hipótese de que um sistema não linear, como é o caso do processo de exploração de reservatórios, pode ser modelado de forma satisfatória com um sistema linear de parâmetros variantes no tempo.

Independente do método, força-se que o sinal de previsão tenha erro nulo no primeiro passo da previsão para que o estimador se inicie de um ponto correto na linha do tempo.

### 3.3 Parâmetros de Identificação e Geração de Resultados

#### 3.3.1 Parâmetros do Modelo

O modelo considerado foi o modelo ARX MISO com  $r = 2$  entradas, cuja saída  $y$  é representada pela taxa de produção de óleo de 1 dos 4 poços produtores e as entradas  $u$  são representadas pela entrada constante e a pela pressão no fundo do poço do poço produtor de onde se mede a saída.

Quanto aos demais parâmetros que estão presentes em (2.20), têm-se que:

- $n_y = 1$
- $n_{u_i} = 1, i = 1, 2 \dots r$
- $n_{k_i}, i = 1, 2 \dots r$ : este parâmetro foi estimado caso a caso, e é um valor entre 1 e 3. A estimativa é feita da seguinte forma: considerando uma estrutura ARX com  $n_y$  e  $n_u$  definidos, são testadas todas as possibilidades de  $n_k$ , usando o estimador MQ por batelada clássico, e é escolhida a que apresenta melhor fit aos dados. A exceção é a entrada constante, que recebeu  $n_k = 0$  sempre.

Portanto, tem-se um total de 3 parâmetros para estimação.

#### 3.3.2 Parâmetros dos Estimadores

Com relação aos estimadores utilizados discutidos na Seção 3.2, dois deles tem parâmetros a serem definidos. Em relação ao estimador de MQ por batelada com janela, pode-se variar o tamanho da janela  $J$ , enquanto que o estimador de MQ recursivo com fator de esquecimento, pode-se variar o fator de esquecimento  $\lambda$ . Dentro deste contexto, utilizaram-se os seguintes parâmetros:

- **Tamanho da janela  $J$ :** O tamanho mínimo da janela foi definido como

$$J_{min} = n_y + \sum_{i=1}^r (n_{u_i} + n_{k_i}) = 1 + r \sum_{i=1}^r n_{k_i} \quad (3.1)$$

Foram feitas identificações utilizando os seguintes valores de  $J$ :

1.  $J = J_{min}$ ;
2.  $J = J_{min} + 6$ ;
3.  $J = J_{min} + 12$ .

Dessa forma, considerando um tempo de amostragem  $T_s$  de 30 dias que é, aproximadamente, um mês, foram usadas e avaliadas janelas com o tamanho mínimo, com tamanho intermediário e uma janela maior, sendo que este último considera um ano a mais do que o mínimo necessário para estimar um modelo.

- **Fator de esquecimento  $\lambda$ :** O fator de esquecimento tem valores que variam, tipicamente, entre 0.95 e 0.99 [18].

Valores menores que 0.95 representa um esquecimento muito rápido do histórico de dados, o que faz com que o modelo seja muito reativo aos dados mais recentes, sendo mais suscetível à ruídos e falhando em identificar as dinâmicas do processo. Usar  $\lambda = 1$  faz com que o estimador deixe de ser considerado MQ recursivo com fator de esquecimento e se torne, simplesmente, MQ recursivo.

Portanto, os seguintes valores de  $\lambda$  foram considerados durante a geração de resultados:

1.  $\lambda = 0.95$ ;
2.  $\lambda = 0.97$ ;
3.  $\lambda = 0.99$ .

### 3.3.3 Método de Identificação

Considerando uma simulação com  $N$  amostras, saída  $y$  e entradas  $u$ , para cada amostra  $k$  são realizadas identificações usando o modelo e os estimadores já apresentados. Os dados considerados disponíveis para estimação do modelo no instante  $k$  são os dados disponíveis desde a primeira amostra até o instante  $k - 1$ . A partir dos modelos estimados, devem ser previstas as  $p$  amostras seguintes.

Portanto, no instante  $k$ , os parâmetros do modelo  $M$  são estimados usando as informações de saída disponíveis de  $y(1)$  até  $y(k - 1)$  e as informações das entradas disponíveis de  $u(1)$  até  $u(k - 1)$ . Estes são chamados de dados de treino de saída  $y_{treino}$  e entrada  $u_{treino}$ , respectivamente. A saída que o modelo  $M$  apresenta durante essa fase de treino é denominada  $\hat{y}_{treino}$ .

Após a estimação, o modelo  $M$ , juntamente com as entradas do sistema, são usados para calcular os valores de  $\hat{y}(k + 1), \hat{y}(k + 2), \dots, \hat{y}(k + p)$ . Para melhor guiar o modelo, é feito com que

$\hat{y}(k) = y(k)$ , forçando erro nulo no primeiro valor estimado. A saída que o modelo  $M$  apresenta durante a fase de previsão é denominada  $\hat{y}_{prev}$ , e pode ser comparada com os valores reais da saída do processo  $y_{prev}$ .

É necessário haver um mínimo de valores disponíveis para que seja possível estimar um modelo. É impossível, por exemplo, estimar um modelo a partir de uma amostra somente. Por isso, os valores de  $k$  são definidos como:

$$J_{min} < k \leq N - p \quad (3.2)$$

onde  $J_{min}$  é o tamanho mínimo da janela do estimador de MQ por batelada com janela, definido em (3.1). Para  $p$ , foram considerados os seguintes valores:

1.  $p = 3$ ;
2.  $p = 6$ .

Logo, considerando o tempo de amostragem  $T_s$  de 30 dias, temos estimativas de curto e médio prazo sendo consideradas e testadas.

Portanto, para cada instante de tempo  $k$ , são estimados modelos ARX usando cada um dos estimadores, usando cada uma das configurações dos estimadores, e todos estes modelos estimados realizam previsões de curto e médio prazo, de acordo com o valores de  $p$ . Deste processo, são calculados 2 resultados principais:

1. **Fit de treino:** Representa o desempenho do modelo em aderir aos dados fornecidos para estimar o modelo.
2. **Fit de previsão:** Representa o desempenho do modelo em prever as próximas  $p$  amostras do modelo.

Esses resultados são coletados de todas as configurações dos estimadores, em todos os instantes  $k$  de todos os poços produtores, repetido para cada uma das 90 simulações disponíveis. Dessa forma, obtém-se uma quantidade significativa de dados para que sejam feitas análises estatísticas com relação ao comportamento destes estimadores em diferentes contextos. Além das análises numéricas que serão realizadas, alguns gráficos serão gerados para complementar a análise dos resultados obtidos.

### 3.3.3.1 Cálculo do fit

O fit  $fit_x$  de duas curvas  $x$  e  $x_{ref}$  mostra o quanto  $x$  se aproxima de  $x_{ref}$ , podendo ter um valor que vai de 1, significando que os sinais são idênticos, até  $-\infty$ , sendo que o valor de  $fit_x = 0$  significa que uma simples linha reta (como a média do sinal, por exemplo), tem um fit tão bom quanto o sinal analisado.

O valor de  $fit_x$  é  $1 - \text{NRMSE}(x, x_{ref})$ , onde NRMSE é a raiz do erro quadrático médio normalizado (*Normalized Root Mean Squared Error*). Portanto, temos que:

$$fit_x = 1 - \frac{\|x_{ref} - x\|}{\|x_{ref} - \bar{x}_{ref}\|} \quad (3.3)$$

É importante notar que o valor do fit é apresentado como um percentual nos resultados a seguir.

## 3.4 Validação do Algoritmo de Identificação

Para se confirmar o correto funcionamento do código escrito, bem como a validação do algoritmo de identificação, realizaram-se alguns testes de identificação de diferentes sinais sintéticos. A seguir estão alguns dos resultados obtidos nos testes realizados.

### 3.4.1 Identificação com Estrutura Perfeita

Primeiramente, utilizou-se a função de transferência mostrada em (3.4). Ela pode ser perfeitamente descrita usando uma estrutura ARX com parâmetros  $n_a = n_b = 1$ , a qual foi usada na identificação.

$$G(s) = \frac{2}{s + 2} \quad (3.4)$$

Nas figuras 3.9-3.11 são mostrados, respectivamente, os resultados obtidos na identificação deste sinal usando os métodos por batelada clássica, por janela e identificação recursiva. Como esperado, o fit encontrado é perfeito, sendo que a única falha é vista na identificação recursiva, pois o modelo identificado demora pelo menos uma amostra para identificar os parâmetros e se ajustar à curva.

Nas figuras que exibem os resultados, a reta vertical cinza simboliza a separação entre os dados usados para treino e previsão. Na identificação por janela, a janela usada para treino do modelo está contida entre a primeira e a segunda reta vertical.

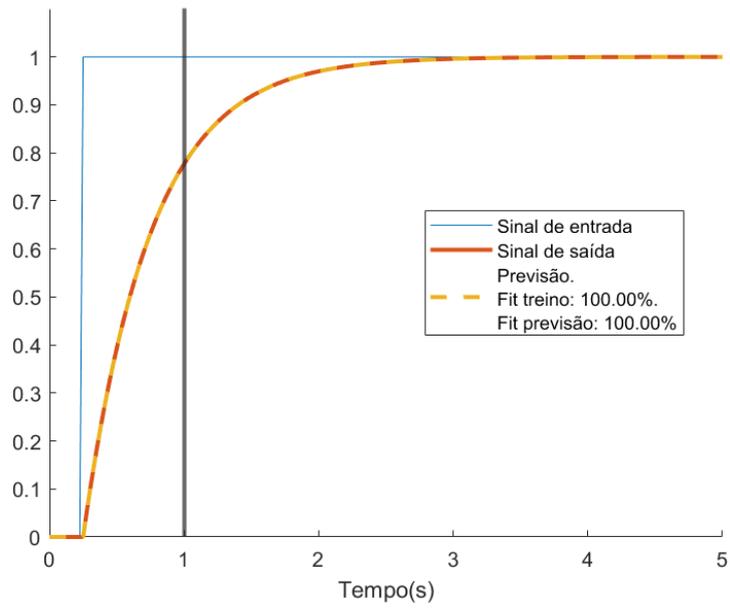


Figura 3.9: Resultado da identificação por batelada.

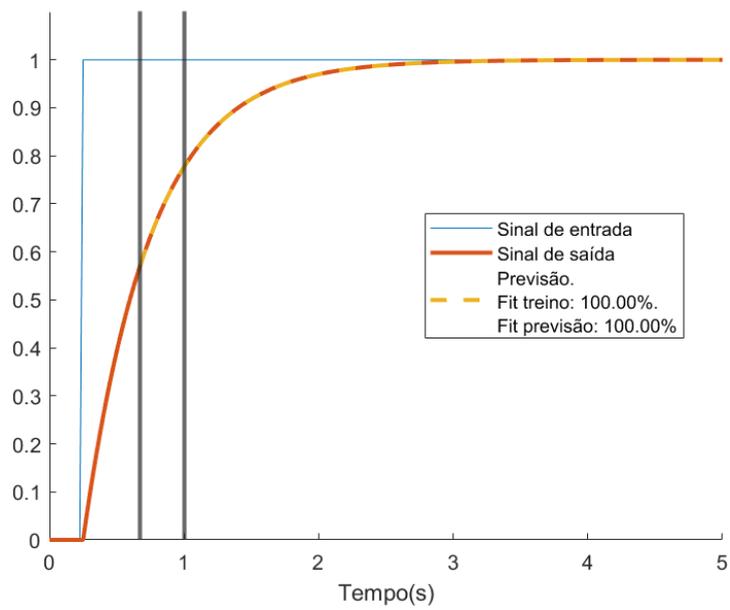


Figura 3.10: Resultado da identificação por janela.

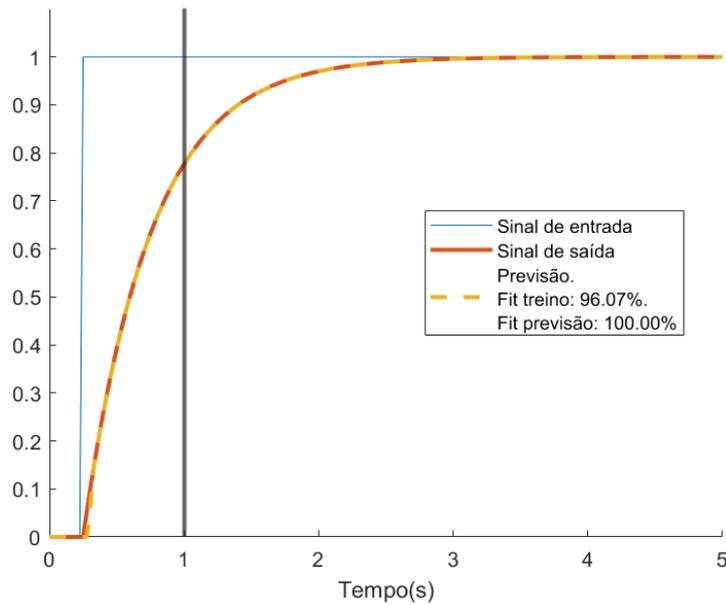


Figura 3.11: Resultado da identificação recursiva.

### 3.4.2 Identificação de Constante no Modelo

Está sendo usada uma adaptação ao modelo ARX, que o permite identificar constantes somadas ao processo. O modelo em si não deixa de ser linear, pois essa constante é vista pelo modelo como o resultado de uma entrada unitária constante, mas que, na prática, permite a identificação de modelos não lineares.

Para o teste, foi identificado o seguinte modelo:

$$y(k) = -0.1 + y(k - 1) + 0.3u(k - 1) \quad (3.5)$$

que pode ser reescrito da seguinte forma:

$$y(k) = y(k - 1) - 0.1u_1(k) + 0.3u_2(k - 1) \quad (3.6)$$

onde  $u_1$  é uma entrada constante igual a 1 e  $u_2$  é a entrada original do sistema. Essa segunda estrutura apresentada foi a forma usada para realizar a identificação. Além disso, utilizou-se uma onda quadrada variando entre 0 e 1 para excitar o sistema.

Assim como no teste anterior, os parâmetros estimados são os parâmetros reais, o que resulta em uma identificação perfeita, como pode ser visto nas figuras 3.12-3.14, que exhibe os resultados para os 3 métodos de identificação em estudo.

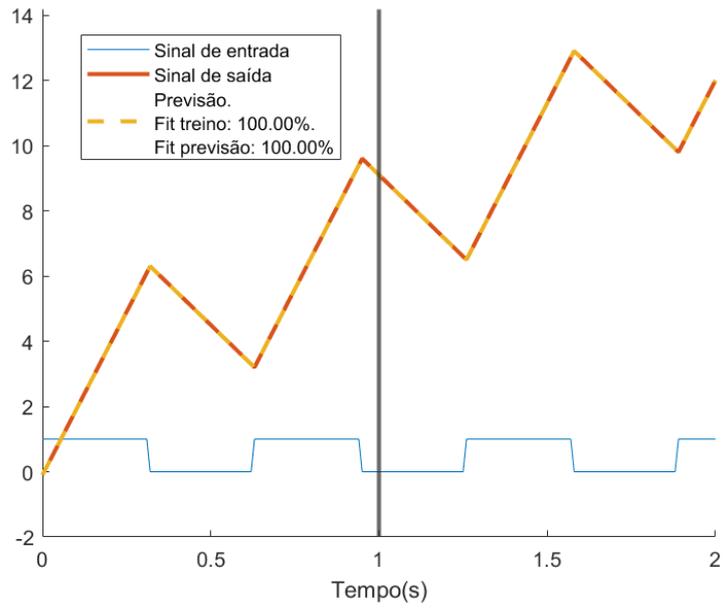


Figura 3.12: Resultado da identificação por batelada.

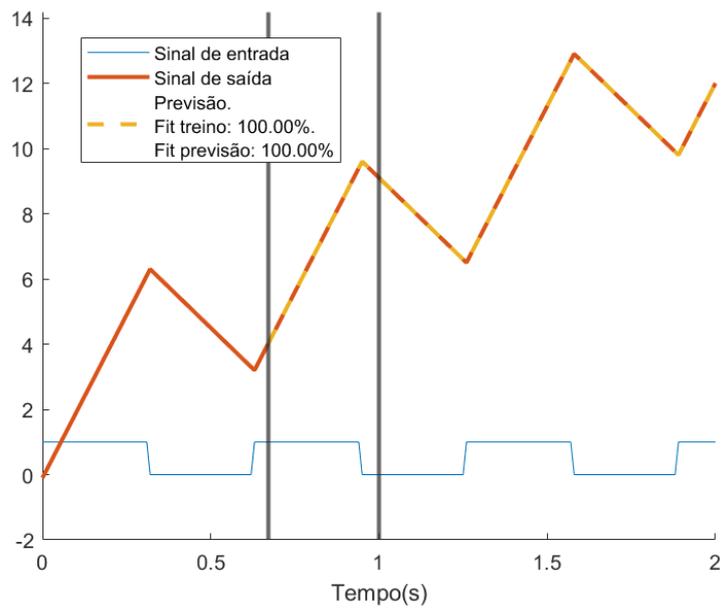


Figura 3.13: Resultado da identificação por janela.

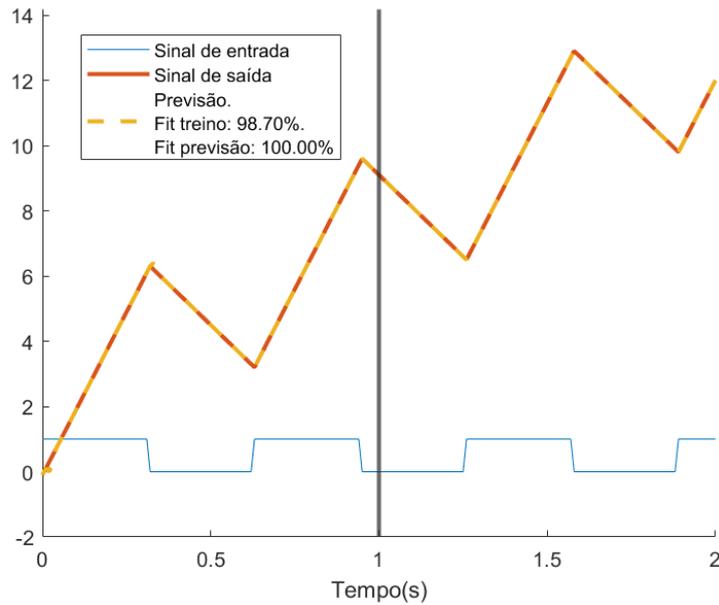


Figura 3.14: Resultado da identificação recursiva.

### 3.4.3 Identificação com estrutura de grau menor que o ideal

Para o segundo teste, foi mantida a mesma estrutura do modelo ARX usada anteriormente, com  $n_a = n_b = 1$ , porém foi usado um sinal gerado pela função de transferência exibida em (3.7), que pode ser descrita perfeitamente por uma estrutura ARX com  $n_a = n_b = 2$ . Ou seja, a estrutura ARX usada para identificação não será capaz de determinar um modelo exato para o sinal apresentado.

$$G(s) = \frac{3}{s^2 + 2s + 3} \quad (3.7)$$

As Figuras 3.15-3.17 mostram os resultados obtidos pelos 3 métodos discutidos até aqui. Como esperado, nenhum deles se adequou perfeitamente ao sinal apresentado, porém é possível distinguir o comportamento de cada método.

No resultado obtido pela identificação usando batelada clássica, que pode ser visto na Figura 3.15, o modelo de 1ª ordem identificado se ajusta o melhor possível aos dados de treino, e apresenta o mesmo comportamento na etapa de previsão. Além disso, é possível ver, claramente, que o primeiro ponto da previsão é forçado para ter erro nulo.

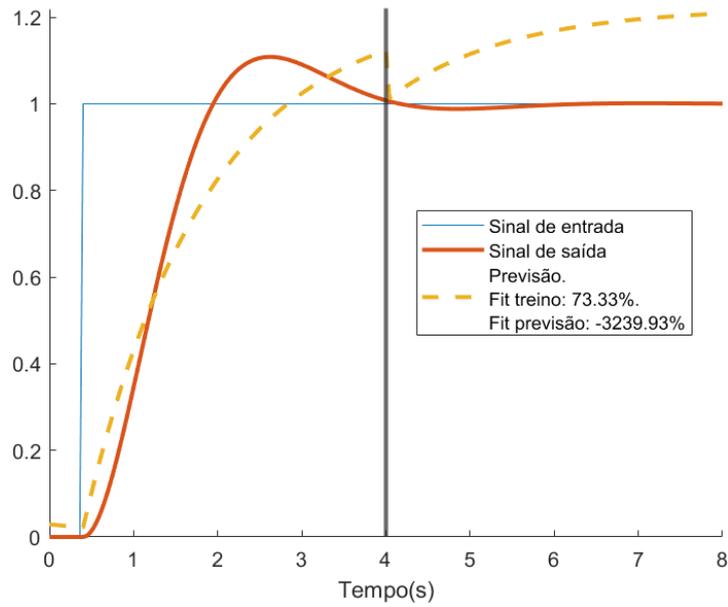


Figura 3.15: Resultado da identificação por batelada.

Na Figura 3.16, que mostra o resultado após a identificação com janela, é possível ver que o modelo identificado ignora todo o histórico anterior e mantém, na previsão, a mesma dinâmica que foi identificada dentro da janela.

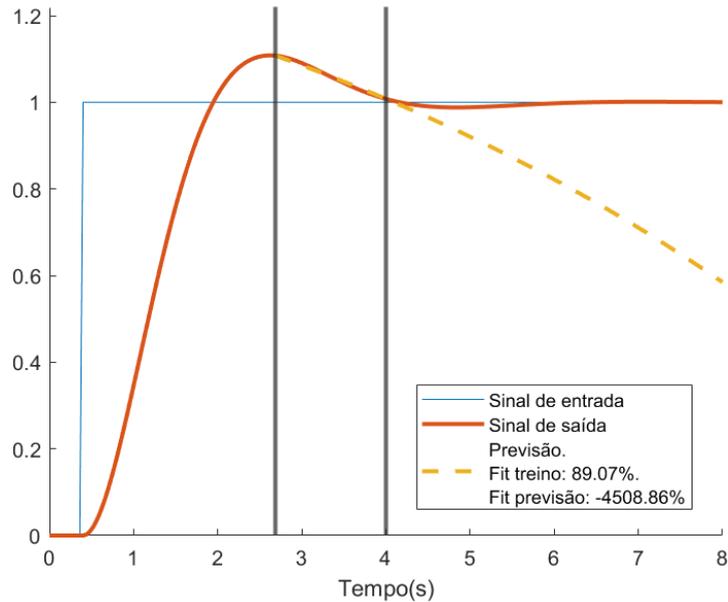


Figura 3.16: Resultado da identificação por janela.

Por fim, na identificação recursiva, cujo resultado é mostrado na Figura 3.17, o modelo consegue adaptar seus parâmetros e ter uma ótima aderência à curva. Isso ocorre pois o modelo se comporta como um sistema de 1ª ordem com parâmetros variantes no tempo, o que o permite simular o

comportamento do sistema de 2ª ordem. Por fim, apesar de ele ter um fit de previsão muito ruim, ele tem um resultado melhor que os métodos anteriores, neste caso.

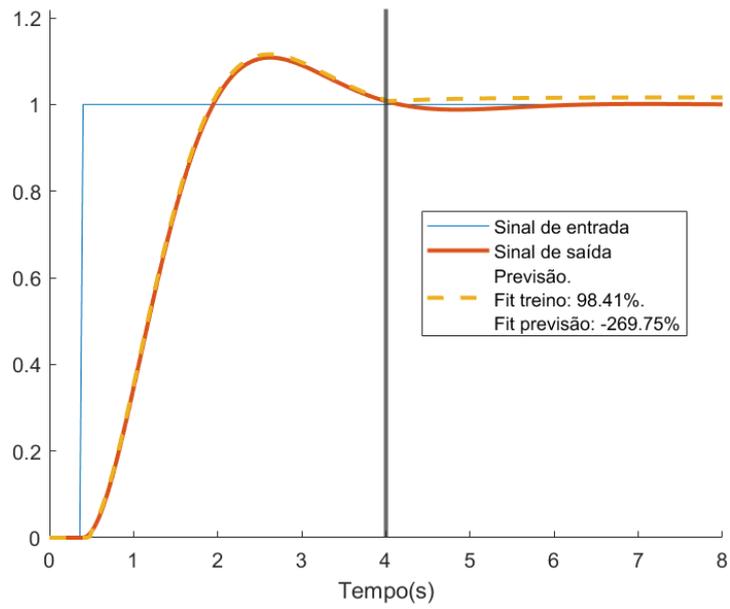


Figura 3.17: Resultado da identificação recursiva.

# Capítulo 4

## Resultados

### 4.1 Resultados Gráficos

A Figura 4.1 mostra o resultado da identificação de WPR do poço PROD2 usando o modelo ARX identificado com o estimador de MQ recursivos, considerando as primeiras 40 amostras do sinal e realizando uma previsão de  $p = 6$  amostras.

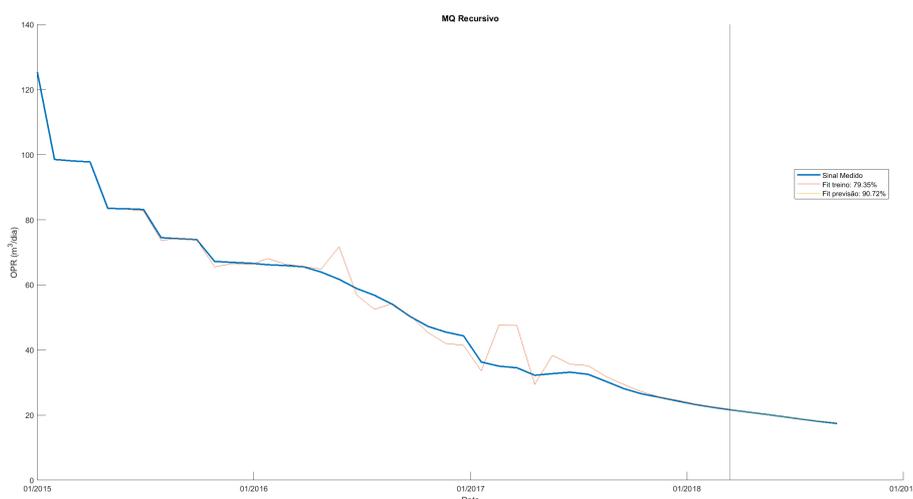


Figura 4.1: Exemplo de resultado obtido.

É possível ver, pelo gráfico, que durante a fase de treino, que está à esquerda da barra vertical, o modelo varia em torno do sinal medido, obtendo um fit de treino de 79,35%. No fim da fase de treino, o sinal medido se estabiliza, e é possível ver que o modelo rapidamente converge para os valores estabilizado. Na fase de previsão, o modelo assumiu, corretamente, que a tendência de estabilidade continuaria, alcançando um fit de previsão de 90,72%.

É, possível ver, também na Figura 4.1, que os valores iniciais do modelo de treino são idênticos aos valores do sinal medido. Isso acontece pois os valores do modelo são inicializados dessa forma,

para facilitar a convergência do modelo. O número de pontos inicializados no modelo recursivo é igual a  $J_{min}$ , conforme a equação (3.1). Esses pontos não são contabilizados no cálculo do fit de treino.

Embora seja útil e interessante visualizar o comportamento do modelo no momento do treino, é ainda mais interessante visualizar as previsões de um dado estimador em vários pontos distintos. Por isso, as análises gráficas, de agora em diante, irão omitir o comportamento do modelo durante o treino para focar somente em exibir os resultados dos modelos durante a fase de previsão. A Figura 4.2 mostra a simulação completa da Figura 4.1, porém com as previsões dos modelos estimados em vários momentos distintos. Os parâmetros são os mesmos informados anteriormente.

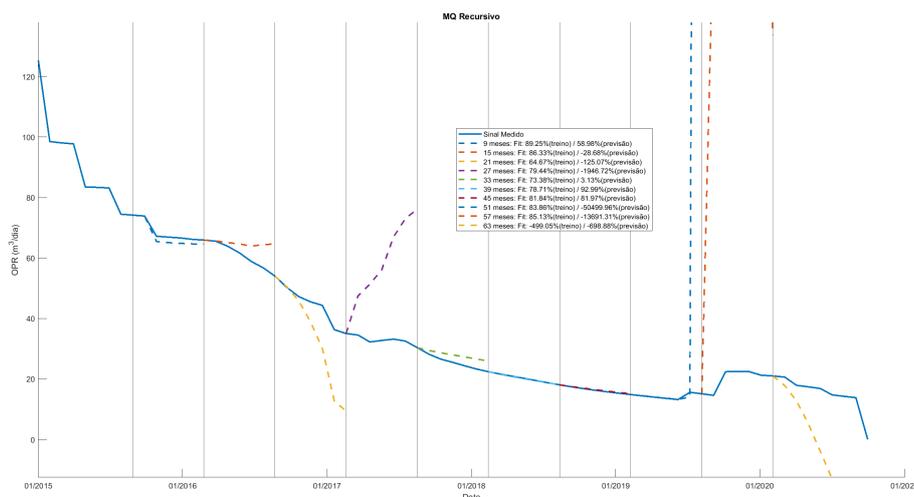


Figura 4.2: Resultado do estimador em vários pontos.

Na Figura 4.2 é possível destacar os seguintes pontos sobre o comportamento do estimador:

1. Logo no início, aos 9 meses, o modelo prevê com certa precisão uma mudança de direção do sinal, o que é compreensível pois parece seguir a tendência que o sinal apresentava anteriormente. É possível, inclusive, confirmar que o modelo teve um bom aprendizado pelo fit de treino de 89,25%.
2. Na sequência do bom resultado, aos 15 meses, nota-se que uma mudança no comportamento do sinal. É possível, aparentemente, associar o bom fit de treino (86,36%) com o mau desempenho da previsão, pois o modelo claramente esperava que a tendência anterior continuasse.
3. Nos meses 21 e 27, o modelo apresenta dificuldades para se adaptar aos dados, pois o modelo prevê uma descida e subida que estão bem fora do padrão do sinal medido. O mesmo ocorre no fim da simulação, no mês 63.
4. Nos meses 39 e 45, o modelo toma proveito de uma estabilidade do sinal medido e prevê, corretamente, a continuidade da estabilidade ao longo de mais de 1 ano. Além disso, é

possível observar, no mês 51, que o modelo prevê corretamente o momento que se quebra a estabilidade, ou seja após uma longa sequência de dados com comportamento de queda a uma taxa quase constante, houve um momento de subida no sinal medido.

Apesar de estimar corretamente o momento da subida, o modelo superestimou, e muito, a magnitude da mesma. No entanto, é possível observar que o sinal medido tem, majoritariamente, uma tendência de declínio, portanto é de se esperar que um movimento ascendente não seja bem modelado.

Para contextualizar e tornar possível uma análise mais aprofundada, os demais dados da simulação acima são mostrados nas figuras que seguem:

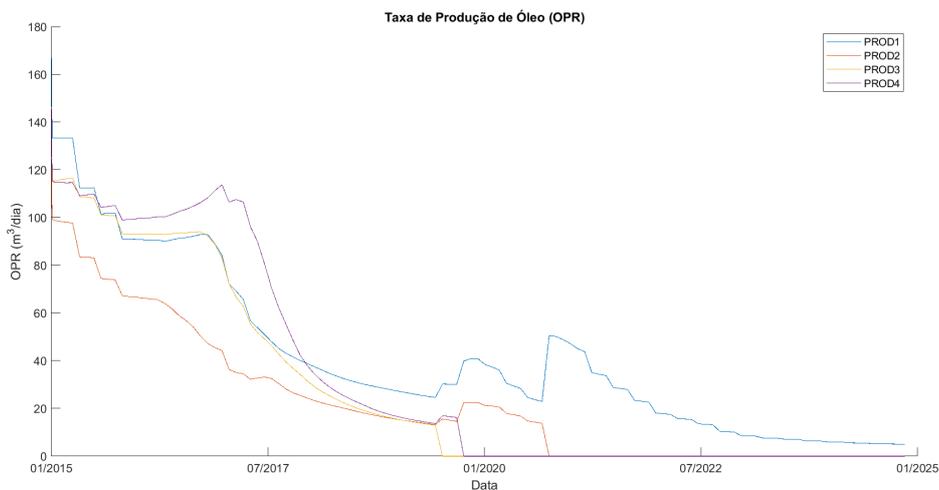


Figura 4.3: Resultado do estimador em vários pontos.

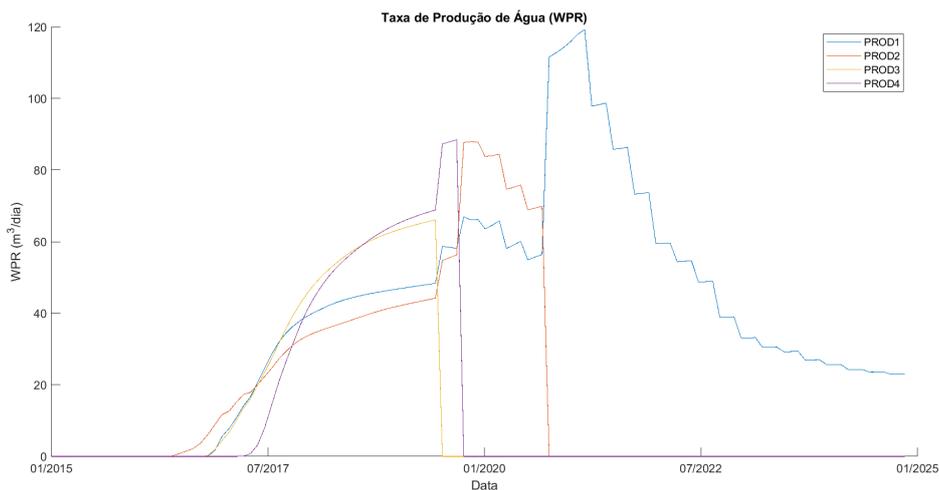


Figura 4.4: WPR

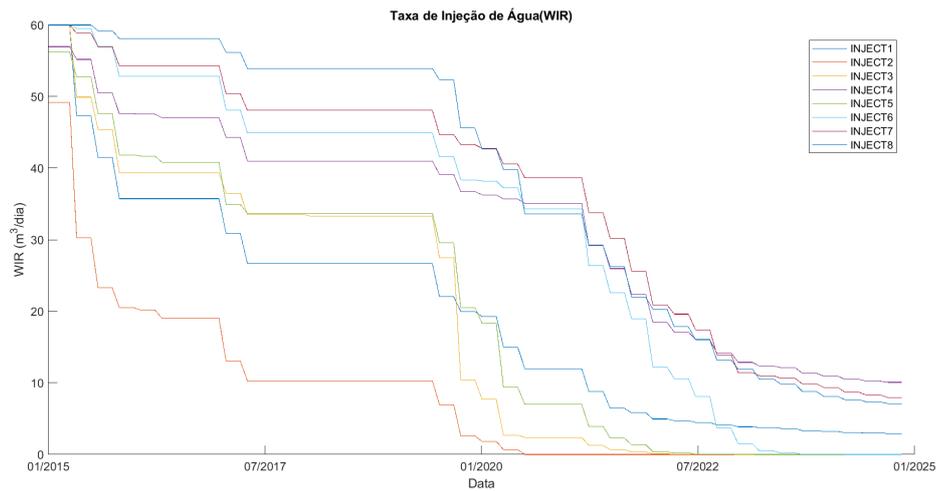


Figura 4.5: WIR

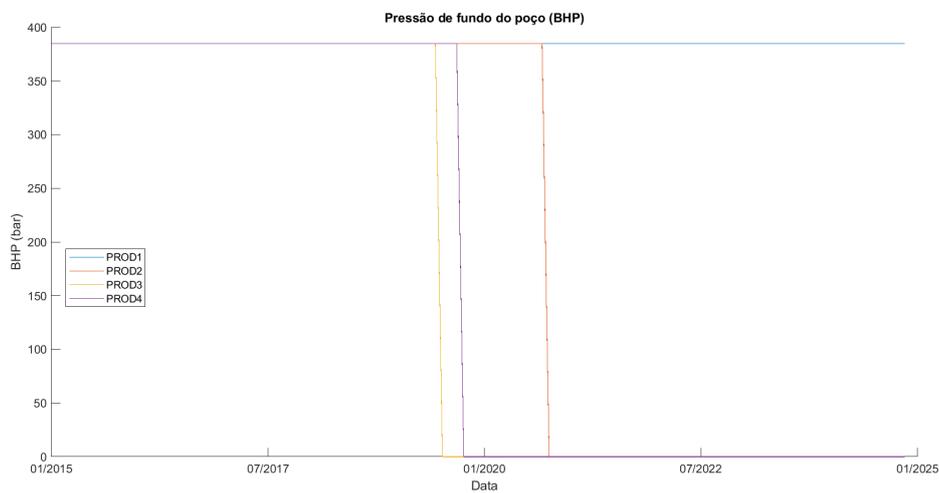


Figura 4.6: BHP

A seguir, serão apresentados alguns resultados gráficos, como mostrado na figura 4.2, para cada um dos estimadores e parâmetros utilizados. Serão utilizado os mesmos dados da simulação anteriormente exibidos, porém considerando a saída do sistema como sendo a produção de óleo do poço PROD1.

#### 4.1.1 Resultados MQ por Batelada

A seguir estão dois resultados obtidos usando o método de Mínimos Quadrados por Batelada, o método clássico de identificação. Os resultados foram selecionados de duas simulações distintas e representam o sinal de poços diferentes.

#### 4.1.1.1 Exemplo de resultado 1

A Figura 4.7 mostra o desempenho do estimador de Mínimos Quadrados por batelada no sinal de OPR do poço PROD1. Nota-se, pela figura, que o estimador tem um desempenho muito baixo no fit de previsão, e um desempenho razoável no fit de treino.

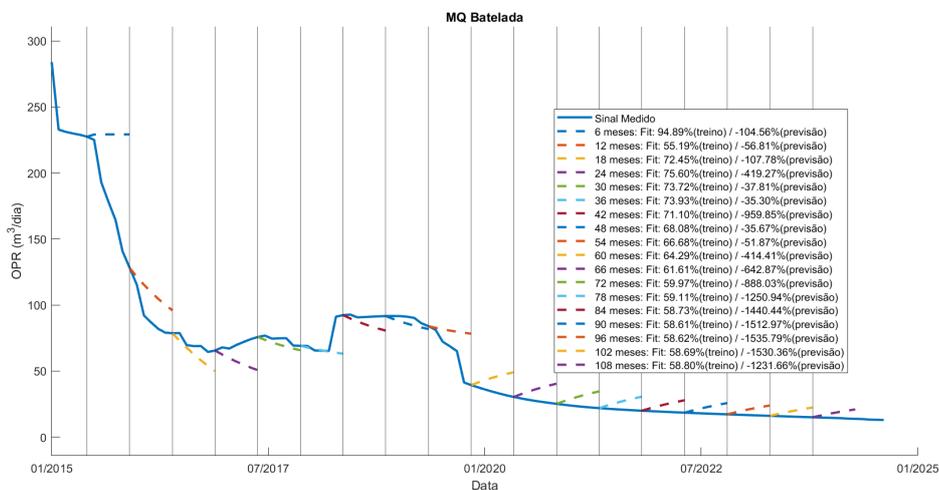


Figura 4.7: Exemplo de resultado 1 do estimador MQ por batelada

Acerca da figura 4.7, no que se refere ao desempenho do estimador em questão, pode-se pontuar:

1. O estimador por batelada não é capaz de acompanhar bem o sinal, mesmo em seus momentos de estabilidade.
2. O estimador têm momentos de grande instabilidade, o que geram previsões que se distanciam muito do sinal medido.

#### 4.1.1.2 Exemplo de resultado 2

Como segundo exemplo de resultado, a Figura 4.8 mostra o desempenho do estimador de Mínimos Quadrados por batelada no sinal de OPR do poço PROD2, em outra simulação. É possível notar o mesmo padrão observado no resultado apresentado anteriormente, com altos fits de treino porém fits de previsão pobres.

A partir do comportamento observado, é possível concluir que o método não é apropriado para a tarefa de identificação de produção de óleo. Esse resultado é realmente frustrante, uma vez que existem diversos trabalhos que exibem ótimos resultados utilizando essa mesma estrutura e esse mesmo método de identificação, porém em ambientes artificiais, com entradas, tempo de amostragem e demais variáveis escolhidas e manipuladas para se obter resultados ótimos. Em uma situação mais próxima da real, porém, o método claramente passa longe do ideal.

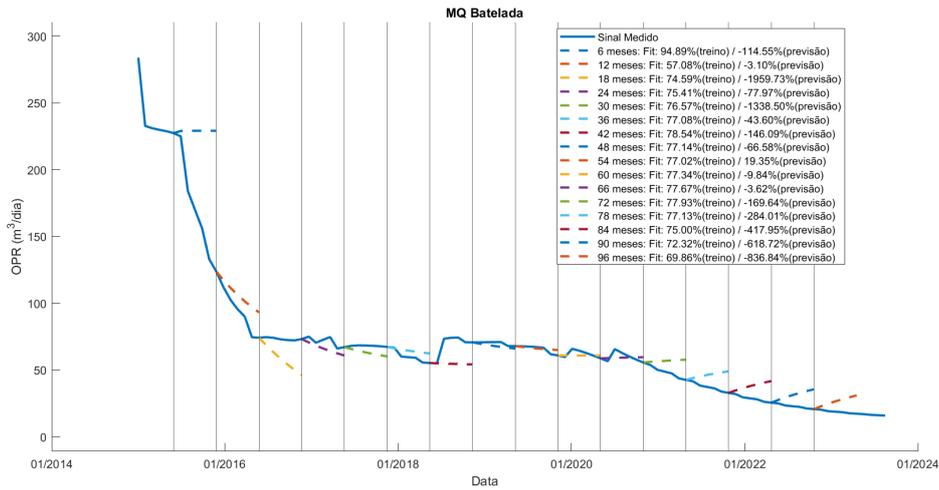


Figura 4.8: Exemplo de resultado 2 do estimador MQ por batelada

## 4.1.2 Resultados MQ por Batelada com Janela

Abaixo estão dois exemplos de resultados obtidos usando o método de Mínimos Quadrados por Batelada com Janela. O método visa melhorar os resultados do método clássico por batelada, uma vez que valores fora da janela são completamente ignorados, anulando o efeito que dinâmicas muito antigas têm sobre as previsões futuras.

Os resultados foram selecionados de simulações diferentes e são dados de poços produtores distintos.

### 4.1.2.1 Exemplo de resultado 1

As figuras 4.9, 4.10 e 4.11 mostram o desempenho do estimador de Mínimos Quadrados por batelada com janelas de tamanho  $J_{min} = 6$ ,  $J_{min} + 6 = 12$  e  $J_{min} + 12 = 18$ .

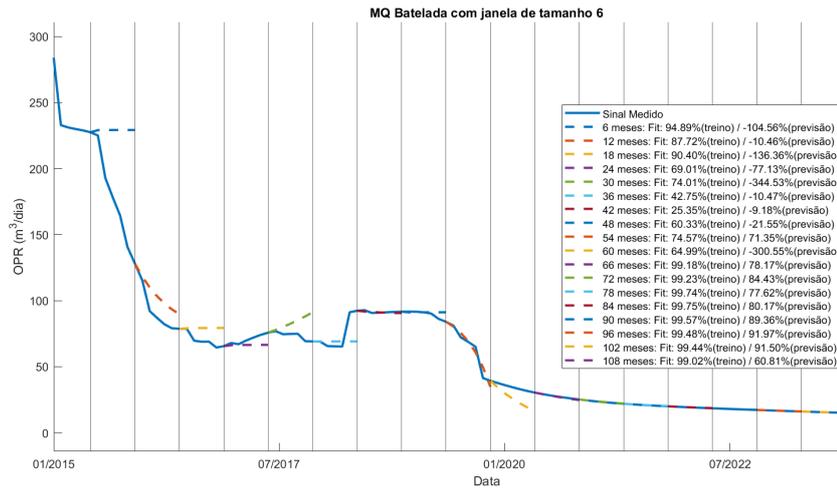


Figura 4.9: Exemplo de resultado 1 do estimador MQ por batelada com janela  $J = J_{min}$ .

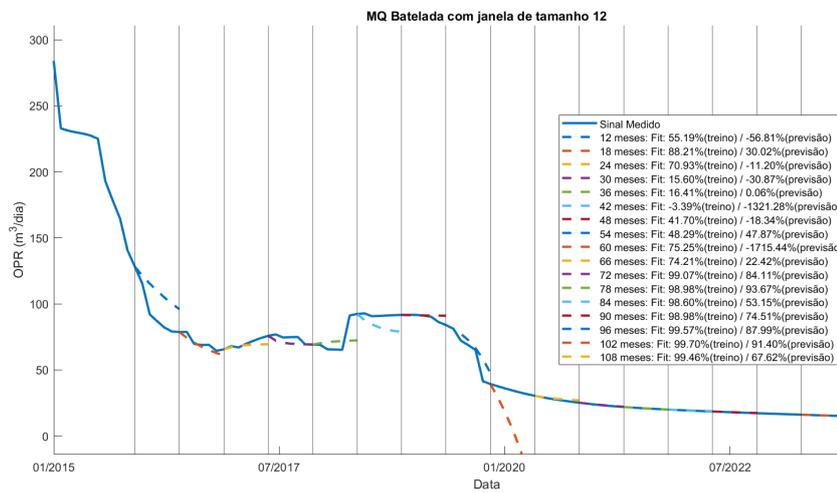


Figura 4.10: Exemplo de resultado 1 do estimador MQ por batelada com janela  $J = J_{min} + 6$ .

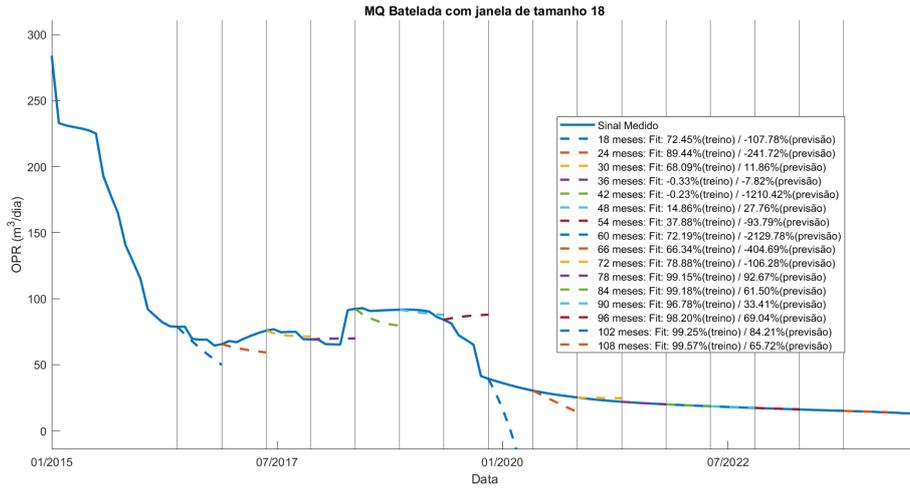


Figura 4.11: Exemplo de resultado 1 MQ por batelada com janela  $J = J_{min} + 12$ .

Sobre os resultados apresentados, pode-se afirmar:

1. Em comparação com o método por batelada sem janela, é visível a melhora nos fits de treino e previsão.
2. É possível observar que a janela menor obteve um desempenho melhor. Tanto para acompanhar o sinal em momento de instabilidade do sinal quanto para o momento de estabilidade do sinal, ao fim da simulação. Observa-se, por exemplo, que a previsão feita aos 66 meses, que marca o início de um momento de estabilidade do poço, se deteriora conforme se aumenta o tamanho da janela. O mesmo acontece com a previsão aos 54 meses, que representa uma variação bem grande no sinal.

#### 4.1.2.2 Exemplo de resultado 2

As figuras 4.12, 4.13 e 4.14 mostram o desempenho do estimador de Mínimos Quadrados por batelada com janelas de tamanho  $J_{min} = 6$ ,  $J_{min} + 6 = 12$  e  $J_{min} + 12 = 18$ .

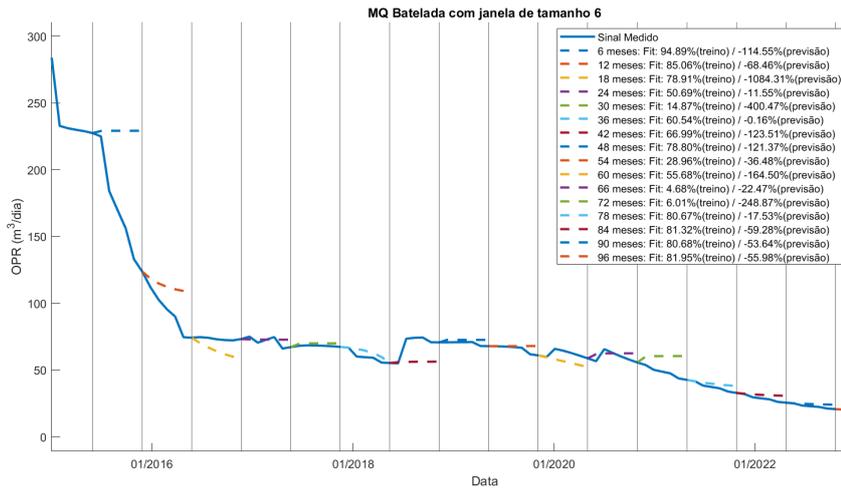


Figura 4.12: Exemplo de resultado 2 do estimador MQ por batelada com janela  $J = J_{min}$ .

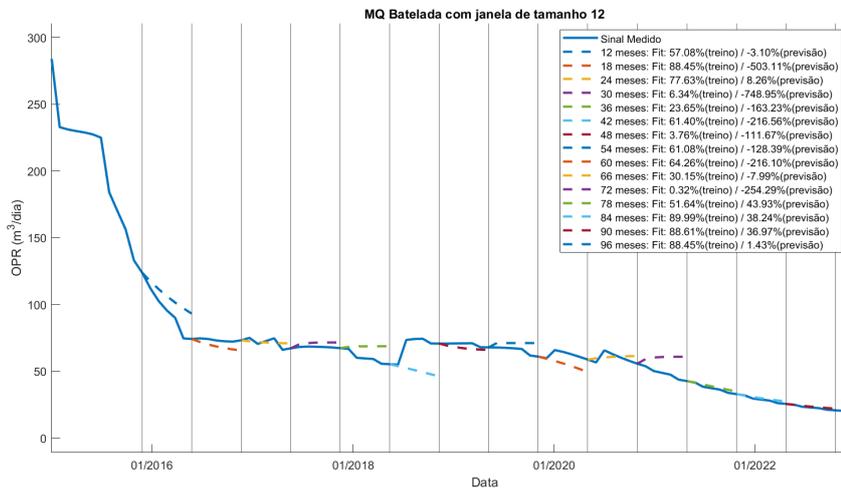


Figura 4.13: Exemplo de resultado 2 do estimador MQ por batelada com janela  $J = J_{min} + 6$ .

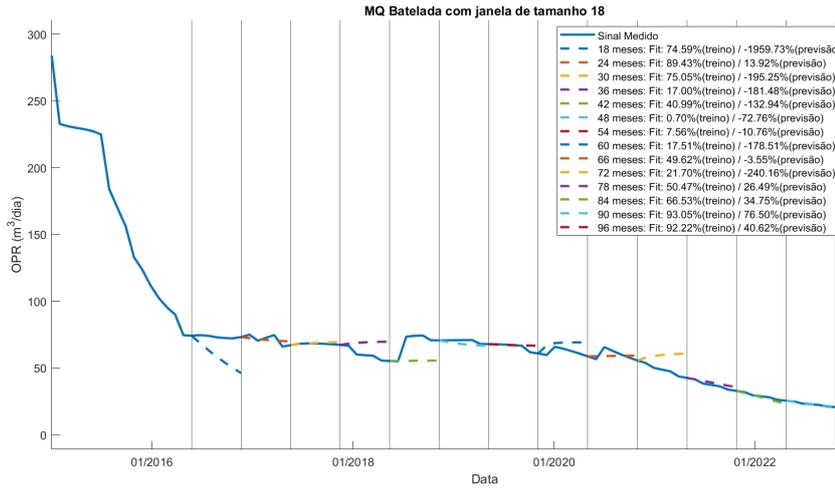


Figura 4.14: Exemplo de resultado 2 do estimador MQ por batelada com janela  $J = J_{min} + 12$ .

O segundo exemplo de resultados é um ótimo contraponto ao primeiro, pois mostra uma situação onde os fits de treino e previsão se tornam melhores com o uso de uma janela maior, ao contrário do exemplo anterior. Isso reforça o fato que, para se obter resultados ótimos, é necessário adequar todos os parâmetros para os dados que se tem em mãos.

Apesar dos resultados melhorarem com o uso de janelas maiores, o uso de uma janela grande demais pode fazer com que os resultados convirjam para os resultados obtidos no método por batelada sem janela, exibido na Figura 4.8. Isso pode ser desejável em alguns momentos, mas em outros não, especialmente no fim da simulação. Levando isso em consideração, o tamanho da janela deve ser variado ao longo da simulação para se obter os melhores resultados.

### 4.1.3 Resultados MQ Recursivo com Fator de Esquecimento

Por fim, dois exemplos de resultados obtidos usando o método de Mínimos Quadrados Recursivo com Fator de Esquecimento. Esse método difere completamente dos métodos por batelada, pois os parâmetros do modelo identificado são atualizados ponto a ponto, buscando trazer a melhor previsão possível ao invés de melhorar o fit de treino global.

Os resultados foram selecionados de simulações diferentes e são dados de poços produtores distintos.

#### 4.1.3.1 Exemplo de resultado 1

As figuras 4.15, 4.16 e 4.17 mostram, respectivamente, o desempenho do estimador de Mínimos Quadrados Recursivo com fator de esquecimento  $\lambda = 0.95$ ,  $\lambda = 0.97$  e  $\lambda = 0.99$ .

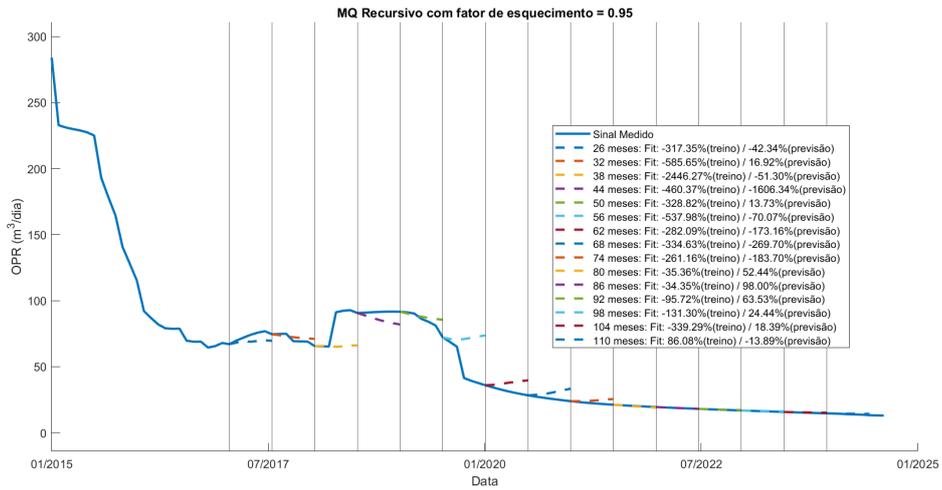


Figura 4.15: Exemplo de resultado 1 do estimador MQ Recursivo com fator de esquecimento  $\lambda = 0.95$ .

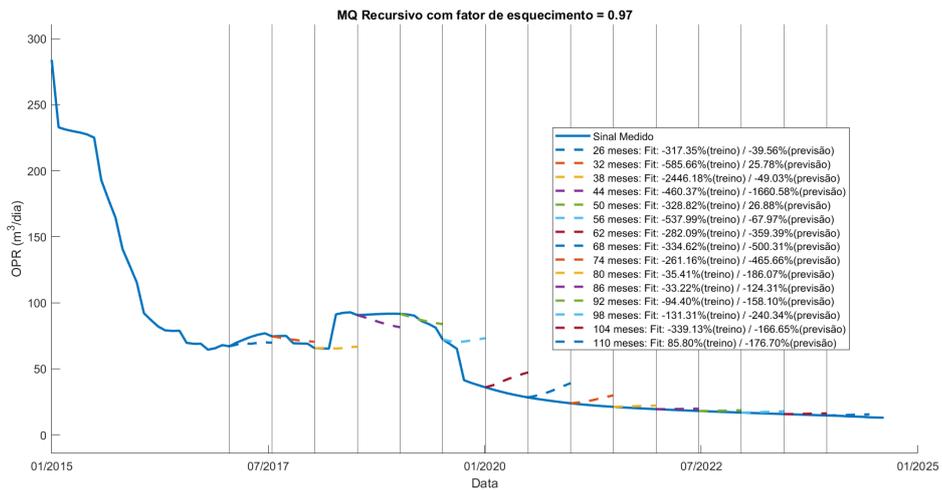


Figura 4.16: Exemplo de resultado 1 do estimador MQ Recursivo com fator de esquecimento  $\lambda = 0.97$ .

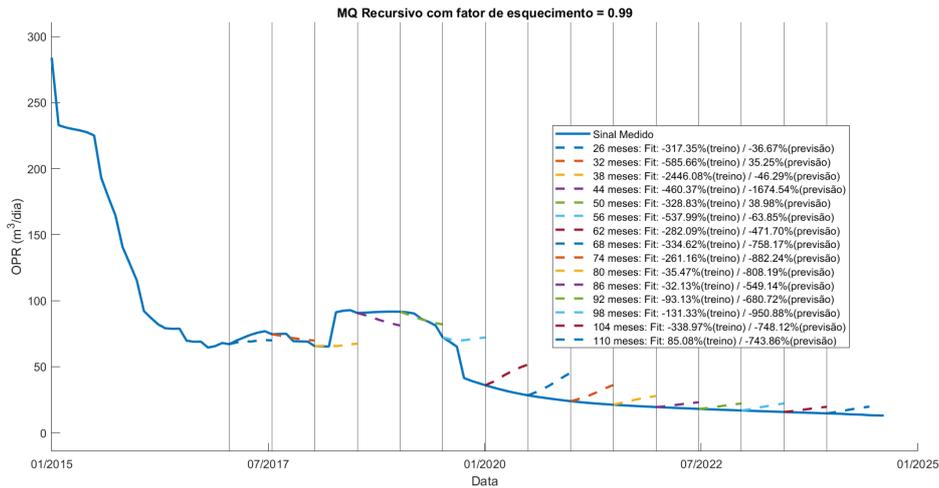


Figura 4.17: Exemplo de resultado 1 do estimador MQ Recursivo com fator de esquecimento  $\lambda = 0.99$ .

Observando os resultados obtidos com diferentes fatores de esquecimento, pode-se dizer que:

1. A diferença do resultado entre os valores distintos de  $\lambda$  é perceptível, sendo que os valores menores obtiveram resultados melhores, neste caso.
2. Ao fim da simulação, nos momentos de estabilidade, a previsão parece estar correta, mas em alguns casos ela possui um pequeno erro sistemático, o que reduz grandemente o valor do fit de previsão alcançado.
3. Ao contrário dos outros métodos, o fit de treino do algoritmo foi muito ruim durante a maior parte da simulação. Além disso, ele não reflete o desempenho obtido no momento da previsão.

#### 4.1.3.2 Exemplo de resultado 2

As figuras 4.18, 4.19 e 4.20 mostram, respectivamente, o desempenho do estimador de Mínimos Quadrados Recursivo com fator de esquecimento  $\lambda = 0.95$ ,  $\lambda = 0.97$  e  $\lambda = 0.99$ .

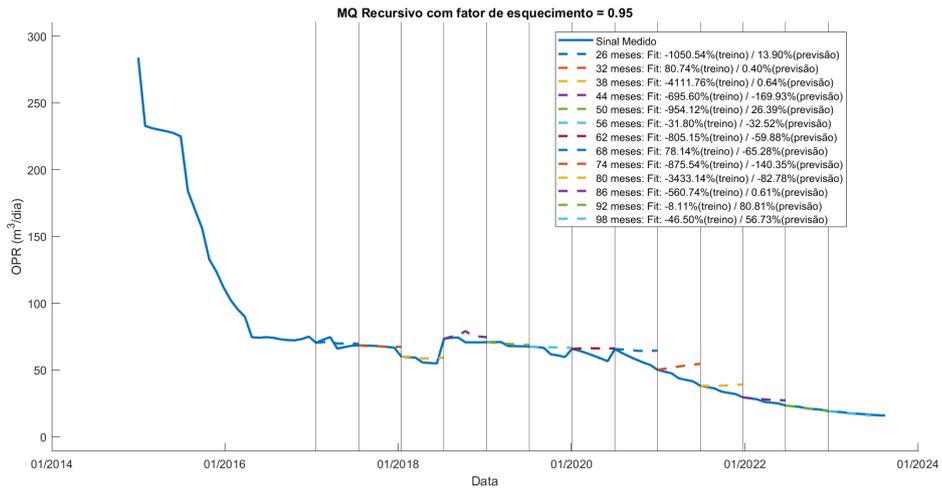


Figura 4.18: Exemplo de resultado 2 do estimador MQ Recursivo com fator de esquecimento  $\lambda = 0.95$

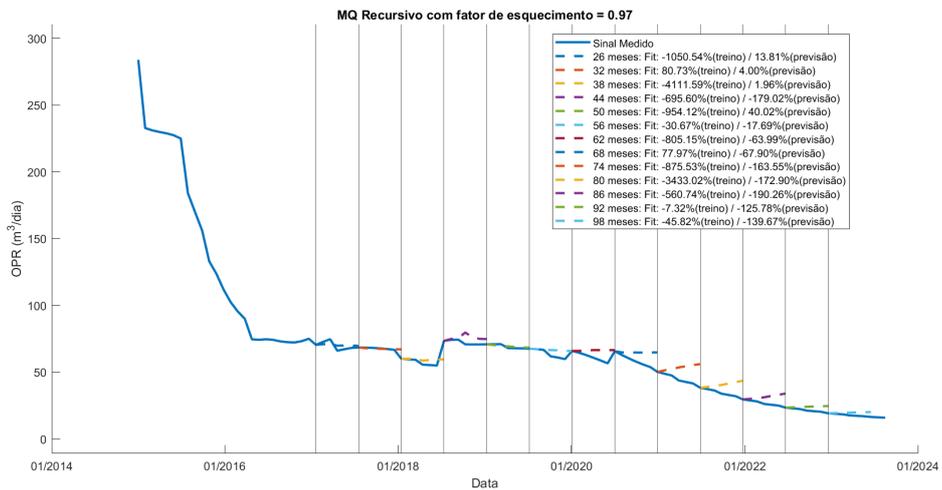


Figura 4.19: Exemplo de resultado 2 do estimador MQ Recursivo com fator de esquecimento  $\lambda = 0.97$

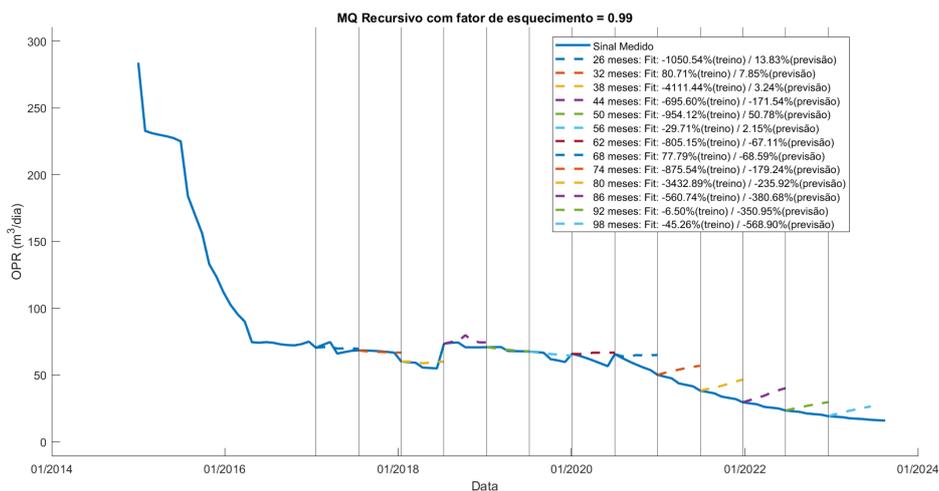


Figura 4.20: Exemplo de resultado 2 do estimador MQ Recursivo com fator de esquecimento  $\lambda = 0.99$

No segundo exemplo se confirma um padrão observado anteriormente, que o uso de um  $\lambda$  menor garante uma melhor aderência aos dados em momentos de relativa estabilidade do sinal. Também se observa como os fits de treino diferem dos resultados de previsão, assim como no exemplo anterior.

Em resumo, o segundo exemplo parece confirmar as impressões que se tira do primeiro. Além disso, em comparação com os outros métodos, o método recursivo tem um bom desempenho, porém possui um ponto fraco em momentos de estabilidade do sinal, que normalmente surgem no fim da produção do poço. Nesse momento, parece mais aconselhável o uso do método por batelada com janela.

## 4.2 Resultados Numéricos

Resultados como os apresentados na Seção 4.1 representam uma pequena parte do total de resultados obtidos. É possível observar que, apesar de uma previsão ter um fit baixo, ela pode conter informações importantes, sendo que uma análise gráfica é o mais indicado.

Devido ao baixo número de pontos de validação da previsão, algo que é inerente ao processo sob análise, o fit de treino nem sempre representa exatamente o quanto a previsão acompanhou o sinal medido, pois qualquer pequena divergência é suficiente para reduzir drasticamente o fit.

Além disso, o mesmo modelo de primeira ordem está sendo aplicado em todos os dados, em todos os pontos, o que também contribui para um baixo desempenho de forma geral. Em uma situação real, os dados do processo seriam passados por um processo de análise mais aprofundado, e toda a estrutura poderia ser estimada novamente para cada ponto.

Por isso, é interessante considerar que fits de previsão baixos também podem representar bons modelos que, após alguns ajustes feitos caso a caso, obteriam previsões bem melhores, e não

significa simplesmente uma incapacidade do método de identificação.

Um dos objetivos de se gerar uma grande massa de dados é permitir uma análise estatística que mostre a tendência geral do estimador com relação à dados típicos de um processo de exploração de reservatórios, independente do ajuste da estrutura para os dados em questão.

Por esses motivos, foi escolhido que um fit de previsão  $> 50\%$  é considerado um bom fit, e a análise dos resultados obtidos será baseada nesse valor de corte.

#### 4.2.1 Bons Fits Obtidos

A seguir é analisada a proporção de modelos considerando bons, ou seja, com fit maior que 50%. É importante notar que, para se chegar a essas proporções, foi coletada a mesma quantidade absoluta de dados de todos os métodos.

Com relação à proporção de fits com mais de 50%, considerando todos os dados coletados, temos os seguintes resultados gerais apresentados na Tabela 4.1:

Tabela 4.1: Proporção de bons modelos em treino e previsão considerando o total de dados

Fit de treino $>50\%$	Fit de previsão $>50\%$
77,21%	7,68%

Percebe-se que um treino bem sucedido é muito mais provável do que uma previsão bem sucedida, como já era de se esperar. Ainda assim, a taxa de treinos que atingiram um valor satisfatório no fit é relativamente alta. Por fim, a taxa de modelos com um bom fit de previsão é baixa, mesmo considerando um fit de somente 50%.

##### 4.2.1.1 Bons Fits por Estimador

A Tabela 4.2 mostra a proporção de bons fits de treino e validação para cada um dos estimadores, considerando as variações nos seus parâmetros.

Tabela 4.2: Proporção de bons modelos por estimador

Estimador	Fit de treino $>50\%$	Fit de previsão $>50\%$
MQ batelada	99,58%	0,24%
MQ batelada com janela $J = J_{min}$	96,54%	2,36%
MQ batelada com janela $J = J_{min} + 6$	95,97%	1,54%
MQ batelada com janela $J = J_{min} + 12$	96,11%	1,77%
MQ recursivo com $\lambda = 0,95$	40,13%	10,23%
MQ recursivo com $\lambda = 0,97$	41,89%	7,97%
MQ recursivo com $\lambda = 0,99$	45,04%	6,22%

É, possível ver, nesses resultados, uma clara separação entre os métodos por batelada e recur-

sivo. Os métodos por batelada priorizam o fit de treino, obtendo o fit mais alto possível aos dados de treinamento, enquanto que o método recursivo se preocupa em atualizar seu modelo para errar menos suas previsões.

Essa diferença nas mecânicas básicas dos métodos de identificação reflete nos resultados obtidos, onde os métodos por batelada alcançam bons resultados em treinamento com muito mais frequência que o método recursivo, mas deixam a desejar no momento de realizar a previsão do sinal, que é onde o método recursivo se mostra muito superior.

Vendo os parâmetros utilizados nos métodos, não é possível perceber grandes diferenças nos resultados ao se variar o valor dos parâmetros usados, mas existe uma variação. Com relação ao método recursivo, os resultados de previsão são melhores quanto menor o valor de  $\lambda$ . De forma semelhante, os resultados em previsão foram melhores para janelas menores.

#### 4.2.1.2 Bons Fits por Tamanho da Previsão

Na Tabela 4.3 está mostrado como os bons fits de cada método se dividiram com relação ao tamanho da previsão considerada.

Tabela 4.3: Proporção de bons modelos para cada valor de previsão

Estimador	Fit de previsão >50% com previsão de 3 passos	Fit de previsão >50% com previsão de 6 passos
MQ batelada	48,33%	51,67%
MQ batelada com janela $J = J_{min}$	74,00%	26,00%
MQ batelada com janela $J = J_{min} + 6$	67,63%	32,37%
MQ batelada com janela $J = J_{min} + 12$	74,70%	25,30%
MQ recursivo com $\lambda = 0,95$	64,00%	36,00%
MQ recursivo com $\lambda = 0,97$	61,33%	48,67%
MQ recursivo com $\lambda = 0,99$	68,70%	31,30%

Com exceção do estimador por batelada padrão, os modelos têm melhores previsões quando essas previsões são mais curtas, como era de se esperar, uma vez que, com menos pontos a serem previstos, é maior a chance da dinâmica predominante identificada recentemente não mudar e o modelo conseguir um bom resultado.

Observando os resultados obtidos, o estimador por batelada padrão se destaca pelos resultados perfeitamente divididos entre as previsões de 3 e 6 meses. Outro ponto interessante de ser observado nos resultados obtidos é que alguns estimadores conseguiram obter mais fits bons do que outros na previsão de 6 passos.

Era de se esperar que estimadores com parâmetros que ignoram dinâmicas de longo prazo, como a janela  $J = J_{min}$  e  $\lambda = 0,95$  fossem os piores para estimativas longas, de 6 meses. Entretanto, de forma surpreendente, os estimadores com parâmetros que consideraram mais as dinâmicas de

longo prazo, sendo eles a janela  $J = J_{min} + 12$  e  $\lambda = 0,99$ , tiveram os piores desempenhos.

A diferença nos resultados entre os métodos e entre os parâmetros considerados não variou tanto, porém é necessário destacar o desempenho sistematicamente melhor do método recursivo para previsões de longo prazo se comparado com os métodos de batelada com janela.

#### **4.2.1.3 Correlação entre Fit de Treino e Fit de Previsão**

Procurou-se alguma correlação entre os valores de fit de treino e fit de previsão. Ao se calcular a correlação entre o fit de treino e o fit de previsão, foi encontrada uma correlação de -2,51%. Porém, ao se considerar somente os casos em que o fit de treino foi maior que 50 %, foi encontrada uma correlação de 11,21%.

Essa não é considerada uma grande correlação, porém é uma correlação positiva e que confirma algo que já era esperado: bons fits de treino tendem a gerar bons fits de previsão. Entretanto, como foi visto na análise gráfica, o modelo muitas vezes se frustra ao prever a continuidade de uma certa dinâmica e obtém baixos fits de previsão, mesmo quando obtém bons fits de treino, o que condiz com a baixa correlação encontrada.

# Capítulo 5

## Conclusões

O processo de exploração de reservatórios é, sabidamente, um processo complexo e não linear. Modelar tal sistema é, portanto, um desafio que motiva muitas empresas e pesquisadores todos os anos. Este trabalho se propôs, até aqui, a estudar o comportamento de diferentes estimadores, de forma a identificar padrões e tendências na modelagem do sistema de exploração de reservatórios através de modelos lineares.

Foram gerados diversos resultados, a partir das simulações disponíveis, identificando e testando modelos ponto a ponto de todas as curvas de produção de petróleo, usando diferentes estimadores e parâmetros de identificação. A partir dos resultados obtidos, foi possível realizar uma análise estatística do desempenho destes estimadores.

A qualidade dos dados parte do princípio de que, estimando os parâmetros do mesmo modelo em uma diversidade de dados, variando os parâmetros de identificação de forma controlada, é possível se obter dados e resultados consistentes, que tem aplicabilidade para outros casos que se aproximem da situação apresentada.

### 5.1 Uso de Modelos Lineares

Neste trabalho foi usado o modelo Auto-Regressivo com Entradas Exógenas (ARX) de primeira ordem para se modelar o processo de produção de petróleo. Foi observado que, apesar de limitado, o modelo ARX consegue modelar certos aspectos do processo, como já foi, inclusive, verificado em outros trabalhos, como em [4].

Conforme pode ser verificado na tabela 4.2, o uso de estimadores por batelada ou recursivos teve bastante impacto nos resultados obtidos. Devido à não linearidade do processo, é natural se esperar que modelos recursivos, principalmente envolvendo fator de esquecimento, consigam adaptar seus parâmetros mais rapidamente para manter a aderência aos dados, como foi registrado, de fato.

A grande vantagem do uso de modelos lineares é a simplicidade com que é possível se estimar seus parâmetros. Dessa forma, se garante mais agilidade no teste de estruturas e na tomada de

decisão. Além disso, dadas certas condições do reservatório, a curva de declínio da produção de óleo pode ser bem representada por um modelo linear simples, como foi mostrado na figura 4.1.

A simplicidade de se estimar os parâmetros de modelos lineares também se prova útil quando se considera a escassez de amostras disponíveis para identificação. Por ser um processo muito lento, é comum o uso de períodos de amostragem mensais, o que dificulta o trabalho de estimadores mais complexos, que, normalmente, necessitam de mais informação.

É importante pontuar que, ao longo de todo o trabalho, os sinais estavam livres de ruído ou perturbações. Se fossem se considerar um sinal com ruído, provavelmente o modelo ARX não seria mais tão indicado, e seria mais interessante experimentar com modelos como o Auto-Regressivo com Média Móvel e Entradas Exógenas (ARMAX, do inglês *autoregressive moving average with exogenous inputs*, que busca modelar o ruído usando um processo de média móvel, que tende a suavizá-lo.

## 5.2 Estimadores Utilizados

Foi utilizado o estimador de Mínimos Quadrados em batelada e recursivo, com variações que buscaram reduzir o peso dado a informações do passado, como forma de se verificar se as dinâmicas atuais do sistema podem ser modelados mais rapidamente e com mais precisão.

Como conclusões acerca dos estimadores em batelada, foi verificado que eles tiveram, de forma geral, desempenhos parecidos, e que a capacidade preditiva dos mesmos é bastante inferior se comparados aos estimadores recursivos. Como é possível visualizar na seção 4.1, os modelos identificados com estimadores em batelada não se aderiram bem aos dados.

Já os estimadores recursivos pareceram bastante promissores, apresentando boas taxas de estimação de modelos considerados bons e, como foi possível ver graficamente, garantiu uma boa aderência aos dados em dados momentos de maior estabilidade na dinâmica do sistema.

## 5.3 Dados do sistema

Os dados usados para realizar os testes foram coletados a partir de simulações de um mesmo modelo de reservatório, que é um modelo simples de duas fases (água e óleo), que conta com a injeção de água como principal mecanismo de recuperação.

As alterações feitas em cada simulação se referem, portanto, somente com relação à condições iniciais, parâmetros do reservatório e aplicação de diferentes esquemas de injeção de água. Com isso, foi possível obter uma miscelânea de dados, simulando os mais diversos cenários.

Nas simulações, foram considerados como entrada do sistema o BHP do poço produtor. Essa escolha foi motivada pois se verificou, após alguns testes, que os resultados obtidos ao se considerar as demais variáveis de entrada resultavam em um modelo não ideal e propenso a instabilidades.

Isso pode ser explicado quando se considera que, com exceção do BHP do próprio poço pro-

dutor, todas as outras variáveis de entrada não estão diretamente relacionadas com a produção daquele poço produtor. Além disso, não é viável usar entradas próprias para se determinar qual a correlação que cada uma dessas outras variáveis têm com o poço produtor.

Aliado à impossibilidade de aplicação de entradas próprias para identificação está o fato de que existem poucas amostras disponíveis para determinação da dinâmica, uma vez que a dinâmica de produção do poço é muito lenta. Além disso, existe um grande período entre a aplicação de uma entrada e a mensuração de seus efeitos no poço produtor. Quando finalmente se torna mensurável o efeito que certa entrada teve no poço produtor, a dinâmica do reservatório já se alterou e, portanto, os parâmetros determinados deixam de ser válidos.

Por fim, reduzir o número de entradas tem a vantagem de simplificar o modelo, reduzindo o número de parâmetros que devem ser estimados. Isso permite que a complexidade do modelo cresça em outros sentidos, como, por exemplo, o uso de modelos de grau maior que um.

## 5.4 Perspectivas Futuras

Alguns pontos podem ser explorados para dar continuidade às investigações iniciadas no presente trabalho. Com relação aos dados utilizados, podem ser usado um novo modelo de reservatório, de forma a verificar se os resultados obtidos com o *Egg Model* se mantêm com um modelo diferente.

Ainda com relação ao uso de um modelo de reservatório diferente, seria interessante trazer para discussão modelos mais complexos, como por exemplo modelos com 3 fases (óleo, água e gás) ou ainda modelos que contam com outros métodos de recuperação primária e/ou secundária.

Além disso, podem ser feitas investigações no sentido de se estudar os efeitos de ruído e perturbação nos dados do sistema. Dessa forma, se torna interessante o uso de outros modelos além do ARX, como por exemplo o modelo ARMAX. Pode ser estudado, então, os efeitos de tais perturbações sobre as estimativas obtidas.

Com relação aos estimadores utilizados, podem ser feitos testes usando o estimador de Mínimos Quadrados Ponderados, que define pesos distintos para os dados de estimação, como uma alternativa de estimador por batelada. Para adicionar uma opção de estimador recursivo, pode ser usado o estimador de Mínimos Quadrados Recursivo de Mínima Covariância. Ambos estimadores são apresentados em [18].

Também pode ser usado alguma versão de estimador recursivo com fator de esquecimento variável, como o apresentado em [21], e avaliar o uso de tais métodos no contexto de exploração de reservatórios.

# REFERÊNCIAS BIBLIOGRÁFICAS

- [1] THOMAS, J. E. *Fundamentos de engenharia de petróleo*. : Interciência, 2001.
- [2] TAFTI, T. A. et al. Injection scheduling design for reduced order waterflood modeling. In: SOCIETY OF PETROLEUM ENGINEERS. *SPE Western Regional & AAPG Pacific Section Meeting 2013 Joint Technical Conference*. 2013.
- [3] REZAPOUR, A. et al. Reservoir waterflooding system identification and model validation with injection/production rate fluctuations. In: SOCIETY OF PETROLEUM ENGINEERS. *SPE Western Regional Meeting*. 2015.
- [4] MIYATAKE, L. K. *Optimization And Control Of Reservoir Models Using System Identification And Machine Learning Tools*. Dissertação (Mestrado) — UFRJ/COPPE/Programa de Engenharia Elétrica, 2019.
- [5] EVEN, S. *Graph algorithms*. : Cambridge University Press, 2011.
- [6] ROSA, A. J.; CARVALHO, R. de S.; XAVIER, J. A. D. *Engenharia de reservatórios de petróleo*. : Interciência, 2006.
- [7] LIU, P. et al. A new combined solution model to predict water cut in water flooding hydrocarbon reservoirs. *international journal of hydrogen energy*, Elsevier, v. 42, n. 29, p. 18685–18690, 2017.
- [8] LIU, Z.; LIU, H. An effective method to predict oil recovery in high water cut stage. *Journal of Hydrodynamics*, Springer, v. 27, n. 6, p. 988–995, 2015.
- [9] LIE, K.-A. *An introduction to reservoir simulation using MATLAB/GNU Octave: User guide for the MATLAB Reservoir Simulation Toolbox (MRST)*. : Cambridge University Press, 2019.
- [10] RASMUSSEN, A. F. et al. The open porous media flow reservoir simulator. *Computers & Mathematics with Applications*, Elsevier, v. 81, p. 159–185, 2021.
- [11] BRATVOLD, R. B. et al. Storm: Integrated 3d stochastic reservoir modeling tool for geologists and reservoir engineers. *SPE Computer Applications*, Society of Petroleum Engineers, v. 7, n. 03, p. 58–67, 1995.
- [12] KIM, J.; YANG, H.; CHOE, J. Robust optimization of the locations and types of multiple wells using cnn based proxy models. *Journal of Petroleum Science and Engineering*, Elsevier, v. 193, p. 107424, 2020.

- [13] SIRAJ, M. M. et al. Model and economic uncertainties in balancing short-term and long-term objectives in water-flooding optimization. In: SOCIETY OF PETROLEUM ENGINEERS. *SPE Reservoir Simulation Symposium*. 2015.
- [14] FORTALEZA, E. L. F.; NETO, E. P. B.; MIRANDA, M. E. R. Production optimization using a modified net present value. *Computational Geosciences*, Springer, v. 24, n. 3, p. 1087–1100, 2020.
- [15] JANSEN, J.-D. et al. The egg model—a geological ensemble for reservoir simulation. *Geoscience Data Journal*, Wiley Online Library, v. 1, n. 2, p. 192–195, 2014.
- [16] ZANDVLIET, M. et al. Bang-bang control and singular arcs in reservoir flooding. *Journal of Petroleum Science and Engineering*, Elsevier, v. 58, n. 1-2, p. 186–200, 2007.
- [17] ESSEN, G. van et al. Robust waterflooding optimization of multiple geological scenarios. *Spe Journal*, OnePetro, v. 14, n. 01, p. 202–210, 2009.
- [18] AGUIRRE, L. A. *Introdução à identificação de sistemas—Técnicas lineares e não-lineares aplicadas a sistemas reais.* : Editora UFMG, 2004.
- [19] KAILATH, T. *Linear systems.* : Prentice-Hall Englewood Cliffs, NJ, 1980.
- [20] BJÖRCK, Å. Least squares methods. *Handbook of numerical analysis*, Elsevier, v. 1, p. 465–652, 1990.
- [21] PALEOLOGU, C.; BENESTY, J.; CIOCHINA, S. A robust variable forgetting factor recursive least-squares algorithm for system identification. *IEEE Signal Processing Letters*, IEEE, v. 15, p. 597–600, 2008.