



TÉCNICO
LISBOA

**Substituição de simuladores de fluidos em
metodologias de ajuste de histórico automático com o
uso da regressão simbólica**

ERIK EIDY SHIMABUKURO

Dissertação para obtenção do Grau de Mestre em Engenharia de Petróleos

Engenharia de Petróleos

Orientador: Prof. Dr. Leonardo Azevedo Guerra Raposo Pereira

Júri

Presidente: Prof. Dra. Maria João Correia Colunas Pereira

Orientador: Prof. Dr. Leonardo Azevedo Guerra Raposo Pereira

Vogal: Prof. Dr. Amílcar de Oliveira Soares

Outubro 2017

Resumo

Um modelo de reservatório geologicamente consistente e que seja capaz de reproduzir corretamente o comportamento da produção de um determinado campo só pode ser obtido através do processo de ajuste de histórico. Este processo ajusta os parâmetros do reservatório até que as produções e pressões do modelo se assemelhem ao histórico existente de produção dos poços. Neste domínio as metodologias assentes em modelos geoestatísticos têm ganho importância devido à sua capacidade de reproduzir os dados de produção ao mesmo tempo que permitem avaliar a incerteza dos modelos ajustados. No entanto, é necessário correr um simulador numérico de fluidos para cada conjunto de modelos petrofísicos o que aumenta consideravelmente o tempo de computação deste tipo de metodologias. Devido ao longo tempo que este procedimento leva, somado ao grande número de simulações necessárias para um bom ajuste de histórico, surge a necessidade de se encontrar um novo método mais eficiente de se obter as curvas de produção destas simulações. Para este trabalho, utiliza-se um algoritmo de regressão simbólica – uma técnica de *machine learning* – como uma alternativa para o uso de simuladores de fluidos, com o objetivo de se reduzir o tempo gasto em simulações como essas. A regressão simbólica é uma ferramenta capaz de gerar modelos que representam equações matemáticas baseadas no comportamento observado pelos reservatórios a serem analisados. Com o uso destes modelos é possível se prever, até certo ponto, o comportamento do reservatório para diferentes configurações de propriedades petrofísicas, excluindo-se a necessidade pelos lentos simuladores de fluidos. O trabalho se inicia com o estudo da ferramenta de regressão simbólica utilizada no âmbito deste trabalho, especialmente sua capacidade para a aplicação desejada, e passa para a substituição parcial do uso dos simuladores de fluidos na simulação dos modelos de porosidades e permeabilidades de reservatórios, através de um ciclo de ajuste de histórico geoestatístico. O reservatório no qual a metodologia foi testada é o Wattfield. Os resultados encontrados para este reservatório foram satisfatórios e indicam uma equivalência na qualidade entre a metodologia original e a metodologia proposta com a regressão simbólica. No entanto, a possibilidade de em menos tempo se realizar um número maior de simulações por iteração na metodologia proposta, possibilita uma convergência ainda mais eficiente devido à ampliação no espaço de exploração dos parâmetros do modelo.

Abstract

A geologically consistent reservoir model that is able to properly reproduce the behavior of the simulated reservoir may be attained by geostatistical history matching. This process will adjust the parameters of the reservoir until the productions and pressures of the model resemble the history of productions from the existing well data. In this domain, the existing methodologies for geostatistical models have become important due to their capacity of reproducing production data while at the same time allowing to evaluate the uncertainty on the adjusted models. However, it is necessary to run a fluid simulator for each set of petrophysical models, increasing considerably the computational time wasted in those methodologies. Due to the long time wasted with this process, and the countless number of simulations necessary for a good history matching, there is the need of finding a more efficient method to obtain the production curves of these simulations. This work studies the use of an algorithm of symbolic regression – a machine learning technique – as an alternative for the use of fluid flow simulators, in order to reduce the time spent with such simulations. Symbolic regression is a tool able of generating models representing mathematical equations based on the behavior observed by the analyzed reservoirs. With the use of these models it is possible to foresee, to a certain extent, the behavior of a reservoir for different configurations of the petrophysical properties, excluding the necessity for the time-consuming fluid flow simulators. This work starts with the study of the symbolic regression tool used in it, especially its capacity for the desired application, and then the partial replacement the fluid flow simulator on the simulation of porosity and permeability models of reservoirs on a geostatistical history matching loop. Watt-field is the reservoir in which the methodology was tested. The results achieved for the studied reservoir show an equivalence on the quality between the original methodology and the proposed methodology with symbolic regression. However, the possibility of, in less time, increasing the number of simulations per iteration on the proposed methodology allows an even more efficient convergence due to the ampliation of the exploration space of the model's parameters.

Sumário

Resumo	1
Abstract	2
1. Introdução	6
1.1. Motivação	6
1.2. Objetivos.....	7
1.3. Estrutura da tese	7
2. Estado da Arte	8
2.1. Simulação de reservatórios	8
2.2. Ajuste de histórico	9
2.3. Programação Genética	11
2.4. Regressão Simbólica.....	12
3. Metodologia	13
3.1. Geração dos modelos	13
3.2. Preparação dos dados	14
3.3. Integração de regressão simbólica no ajuste de histórico geoestatístico	14
4. Exemplos de aplicação.....	17
4.1. Reservatório sintético simples.....	17
4.1.1. Caso 1: WOPR Well P3 – 10 genes/5 gene depth	17
4.1.2. Caso 2: WOPR Well P3 – 3 genes/3 genes de profundidade.....	19
4.2. Watt-field	21
4.2.1. Ajuste de histórico geoestatístico tradicional (com simulador de fluidos).....	22
4.2.2. Ajuste de histórico com regressão simbólica	25
4.2.3. Análise estatística dos modelos petrofísicos gerados no final do processo iterativo	
32	
5. Conclusões	38
Referências	40

Lista de figuras

Figura 1 – Exemplo de um bom ajuste de histórico	9
Figura 2 – Algoritmo básico da programação genética (Poli et al., 2012)	12
Figura 3 – Diagrama simplificado da metodologia proposta de ajuste de histórico de produção geoestatístico com integração de regressão simbólica.	16
Figura 4 – WOPR Well P3: Predições para os dados de treinamento do caso 1 (modelo mais complexo)	18
Figura 5 – WOPR Well P3: Predições para os dados de teste (validação) do caso 1 (modelo mais complexo)	18
Figura 6 – WOPR Well P3: Predições para os dados de treinamento no caso 2 (modelo mais simples)	19
Figura 7 – WOPR Well P3: Predições para os dados de validação no caso 2 (modelo mais simples)	20
Figura 8 – Localização dos 16 poços do reservatório Watt-Field e polígonos de Voronoi associado a cada poço	22
Figura 9 – Watt-field WWPR Well 4a: Iterações 1 e 6 do ajuste de histórico sem regressão simbólica que ilustram boa convergência dos modelos petrofísicos.	23
Figura 10 – Watt-field WWPR Well 3: Iterações 1 e 6 do ajuste de histórico sem regressão simbólica que ilustram má convergência dos modelos petrofísicos.	24
Figura 11 – Watt-field: Menor valor da função objetivo global por iteração (ajuste de histórico geoestatístico com simulador de fluidos)	25
Figura 12 – Watt-field WOPR Well 8: Ocorre boa convergência das curvas de produção simuladas pelo algoritmo tradicional (esquerda) e com a regressão (direita) após as 6 iterações.	26
Figura 13 – Watt-Field WOPR Well 3: Iterações 1 e 6 do ajuste de histórico com regressão simbólica que ilustram má convergência dos modelos petrofísicos.	28
Figura 14 – Watt-field WBHP Well 1B: iterações 1 a 6 que ilustram um caso no qual as curvas simuladas já se encontram dentro dos limites de erro desde a primeira iteração	29
Figura 15 – Watt-field WWPR Well 7: iterações 1 a 6 que ilustram um caso no qual a regressão simbólica não é capaz de sempre reproduzir corretamente o comportamento do poço para todas as iterações	30
Figura 16 – Watt-field: Menor valor da função objetivo global por iteração (ajuste de histórico geoestatístico com regressão simbólica)	31
Figura 17 – Corte horizontal dos modelos reais (em cima) de porosidades/permeabilidades e dos modelos finais do ajuste tradicional (meio) e com a regressão (embaixo). Valores da escala de permeabilidade representam Log10(perm) em mD.	34
Figura 18 – Histogramas de porosidade e permeabilidade com o algoritmo original (esquerda) e com a regressão (direita)	36

Figura 19 – Gráfico Q-Q entre os dados experimentais dos poços e o modelo matemático obtido com a regressão simbólica (permeabilidade à esquerda; porosidade à direita) 37

Figura 20 – Gráfico Q-Q entre os modelos reais e os modelos obtidos pela regressão simbólica. (Porosidade à esquerda e Permeabilidade à direita) 37

1. Introdução

Este capítulo introduz de forma breve e objetiva os conceitos abordados neste trabalho e como estes se comunicam entre si, gerando a motivação necessária para o estudo deste tema.

1.1. Motivação

Conforme a produção de petróleo se aventura em zonas mais profundas e/ou de difícil exploração, nota-se também um aumento considerável no seu custo, especialmente para produções offshore. Desta forma, cada vez mais procura-se obter uma maior quantidade e uma melhor qualidade de informações confiáveis para se realizar um estudo e uma análise aprofundada acerca dos reservatórios em questão.

Após a descoberta de um reservatório de hidrocarbonetos e durante a fase de produção do mesmo, um dos principais estudos realizados, senão o principal, é a simulação do escoamento de fluidos de um dado reservatório. Tendo por base um modelo estático com a distribuição espacial das propriedades em estudo, é possível realizar tais simulações com o uso de simuladores de fluidos numéricos que são capazes de fornecer boas previsões para as produções totais de água, óleo e gases bem como a pressão no campo. Tendo-se essas estimativas, é possível realizar uma avaliação econômica do reservatório e definir a viabilidade de um projeto em cima dele.

No entanto, para se obter um bom modelo é necessário realizar um processo iterativo de ajuste de histórico. Este procedimento consiste no ajuste dos parâmetros de um modelo (i.e. a distribuição espacial de propriedades petrofísicas na subsuperfície) até que suas produções, obtidas pelos simuladores de fluidos, sejam semelhantes aos valores observados no histórico do reservatório analisado. Pelo fato destes simuladores serem bastante lentos, um bom ajuste de histórico pode levar muitos dias para ser completado, mesmo para reservatórios relativamente simples. Por este motivo, há uma busca por alternativas ao uso desses *softwares* que sejam mais eficientes mas ainda assim confiáveis. Um possível candidato é o uso de algoritmos de *machine learning* que são capazes de reduzir consideravelmente esse tempo.

Machine learning é um campo da inteligência artificial em ciência da computação que tem como objetivo a construção de modelos que podem aprender e fazer previsões em cima de dados. No âmbito da engenharia de petróleo esse campo ainda é bastante novo e pouco explorado, mas já há estudos sobre a aplicação de certas técnicas (eg. *Artificial Neural Network* e *Support Vector Machines*) em engenharia de reservatórios (Rojas et al., 2012).

Um dos algoritmos que utilizam esse conceito de *machine learning* é a regressão simbólica, um tipo de regressão que gera modelos simbólicos que descrevem as relações existentes em uma base de dados. Este algoritmo é inspirado nos princípios da seleção natural,

evoluindo-se artificialmente os modelos e selecionando-se os melhores, os quais serão utilizados para a formação da próxima geração de modelos (Vladislavleva et al., 2009).

O presente trabalho tem exatamente como objetivo a utilização da regressão simbólica como uma alternativa para a utilização de simuladores de fluidos numa metodologia de ajuste histórico geoestatístico, com o objetivo de reduzir o tempo gasto em simulações como essas. Para o problema apresentado, o algoritmo receberia alguns modelos de porosidade e permeabilidade, bem como suas estimativas de produção obtidas pelo simulador de fluidos, e então geraria um modelo matemático capaz de reproduzir essas previsões em cima de novos modelos petrofísicos do dado reservatório.

1.2. Objetivos

Este trabalho trata do estudo experimental do uso da programação genética, um algoritmo de *machine learning*, na engenharia de reservatórios e foi desenvolvido a partir dos seguintes objetivos específicos:

1. Estudo da regressão simbólica aplicada através da plataforma de software GPTIPS2 e avaliação de seus resultados.
2. Substituição parcial do uso do simulador de fluidos, Eclipse® (Schlumberger), em um ciclo de ajuste de histórico por modelos de equações treinadas a partir da regressão simbólica, a fim de se reduzir consideravelmente o tempo nas previsões de produções e pressões.

1.3. Estrutura da tese

Este trabalho se inicia com o estudo dos conceitos básicos necessários para o seu entendimento, como a simulação de reservatórios, ajuste de histórico e regressão simbólica. Depois, é explicada a metodologia utilizada, desde a configuração dos parâmetros para a geração dos modelos da regressão simbólica até o ciclo do ajuste de histórico com o uso da metodologia proposta. Finalmente, são exibidos e comentados os resultados obtidos para 2 diferentes reservatórios sintéticos.

2. Estado da Arte

2.1. Simulação de reservatórios

Segundo Zhangxin (2007), métodos tradicionais para a predição da performance do reservatório podem ser divididos em 3 categorias: métodos análogos, experimentais ou matemáticos. Métodos análogos usam as propriedades de reservatórios maduros que são geograficamente ou petrofisicamente similares ao reservatório alvo para que se tente estimar sua performance. Métodos experimentais medem as propriedades físicas (como taxas de produção, pressões e saturações) em modelos de laboratório e escalam esses resultados para o tamanho real das acumulações de hidrocarboneto. Finalmente, métodos matemáticos utilizam equações matemáticas para se predizer essa performance.

Uma outra abordagem vem do uso de simuladores de reservatórios. A simulação de reservatório combina a física, matemática, engenharia de reservatório e a programação computacional para desenvolver uma ferramenta capaz de predizer a performance de reservatórios de hidrocarbonetos sob diversas condições de operação (Taniar, 2010). Ela se refere a construção e operação de um modelo cujo comportamento assume a aparência de um verdadeiro reservatório .

Um conjunto de equações algébricas desenvolvidas a partir de um conjunto de equações parciais diferenciais reproduzem o comportamento do reservatório na abordagem de simulação de reservatório. Essas equações incorporam os mais importantes processos físicos que ocorrem no sistema do reservatório, incluindo, entre outros, o fluxo de fluidos (óleo, água e gás), e a transferência de massa entre as diversas fases. Os efeitos da viscosidades, capilaridade, e forças da gravidade no escoamento dos fluidos são levados em consideração através do uso de uma forma generalizada da lei de Darcy. Uma grande vantagem dessa abordagem está no fato de que o menor número de suposições simplificadas são usadas para a heterogeneidade do reservatório, transferência de massa entre as fases, e as forças/mecanismos responsáveis pelo fluxo. Além disso, variações espaciais das propriedades das rochas descritas por modelos estáticos existentes, propriedades dos fluidos e características da permeabilidade relativa podem ser precisamente representadas em simuladores de reservatórios.

A necessidade pela simulação de reservatórios é consequência da necessidade dos engenheiros de petróleo para se obter previsões precisas da performance de um reservatório de hidrocarbonetos sob diferentes condições de operação. Essa necessidade surge do fato de que em um projeto de recuperação de hidrocarbonetos, o qual pode envolver investimentos de centenas de milhões de dólares, o risco associado com o plano de desenvolvimento selecionado deve ser avaliado e minimizado.

Fatores que contribuem para o risco existente incluem a complexidade do reservatório devido a propriedades heterogêneas e anisotrópicas das rochas; variações regionais das

propriedades dos fluidos e características da permeabilidade relativa; a complexidade dos mecanismos de recuperação de hidrocarbonetos; e a aplicabilidade de outros métodos preditivos cujas limitações podem torná-los impróprios. Os primeiros três fatores estão além do controle do engenheiro, mas são levados em consideração em simulações de reservatórios através da generalidade dos dados de entrada presentes nos modelos de simulação de reservatórios e pela disponibilidade dos simuladores para as várias técnicas de recuperação avançada de petróleo (EOR).

2.2. Ajuste de histórico

Ajuste de histórico é o processo no qual se ajusta um modelo de reservatório que caracteriza a distribuição espacial das propriedades petrofísicas de interesse até que a simulação de sua produção através de um simulador de fluidos seja próxima do comportamento observado do referido reservatório. Como consequência, os modelos que passam por este processo fornecem previsões de produção com um grau de confiabilidade muito maior, especialmente se os ajustes realizados seguirem limitações de propriedades geológicas e de parâmetros de engenharia conhecidas do reservatório. Baseado no comportamento esperado do reservatório, os cenários de produção podem ser otimizados. A precisão do ajuste de histórico dependerá da qualidade do modelo de reservatório inicial e da qualidade e quantidade de dados existentes da produção e pressão (TNO Geo Energy, 2009). O exemplo da Figura 1 mostra um ajuste de histórico bastante preciso.

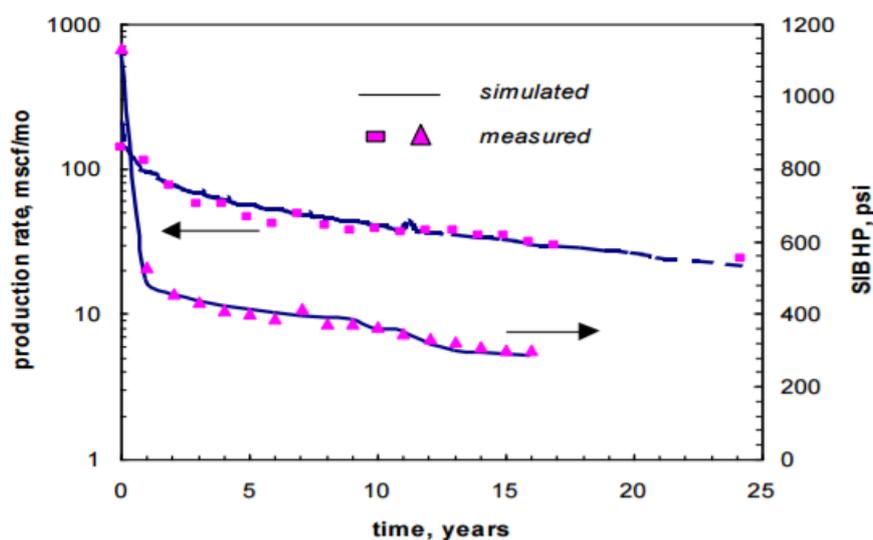


Figura 1 – Exemplo de um bom ajuste de histórico

A construção de um modelo de reservatório é uma tarefa incrivelmente difícil devido à falta de informações disponíveis para esse fim. Nossas mais confiáveis fontes de informação são provenientes de *logs* e *core data*, que apenas conseguem cobrir uma fração minúscula do

reservatório. Desta forma, gerar um modelo de reservatório utilizando apenas esses dados será na maioria das vezes insuficiente para representar de forma segura as produções e os dados de pressão do campo e, portanto, é essencial que o modelo de reservatório seja calibrado pelo processo de ajuste de histórico. (Baker et al, 2006)

É bastante provável que haja diversos modelos que se encaixem no perfil do history match. No entanto, embora não seja possível garantir que o modelo correto seja encontrado através do ajuste de histórico, é certo que sem esse processo o modelo não estará correto. Apesar desta etapa para a geração do modelo ser extremamente lenta e consumir um enorme tempo, ela é igualmente necessária.

Um caso onde o ajuste de histórico é essencial ocorre em poços de gás em reservatórios de baixa permeabilidade. Muitas vezes nesses reservatórios a permeabilidade encontrada pelo *core data* é muito maior que a determinada por *buildup tests*. Essas diferenças são causadas por *shales* internos às zonas do reservatório que resultam em um caminho tortuoso para o fluxo de fluidos. Se estes *shales* forem pequenos, muitas vezes eles não são mapeados pela descrição geológica inicial e resultam em falsas elevadas permeabilidades. Com o ajuste de histórico, esta incerteza pode ser evitada.

O ajuste de histórico é normalmente realizado através de dois métodos: manual ou automático. No processo manual, o simulador gera os dados de produção do modelo para um determinado período. Após a obtenção dos resultados, manualmente compara-se estes dados com os dados existentes do reservatório. Finalmente, realiza-se os ajustes necessários nos modelos estáticos do reservatório (dados de entrada) a fim de se aproximar do comportamento observado. Estes ajustes são realizados baseados na experiência e no conhecimento do reservatório. Já no processo automático, o simulador aperfeiçoa o modelo minimizando uma função objetivo construída através da diferença entre os dados de produção observados e os resultados obtidos através da simulação para um determinado modelo estático.

O ajuste manual dos parâmetros do reservatório se baseia em tentativa e erro. No entanto, na maioria dos casos é impossível utilizar-se toda a informação disponível. Este método não se adequa a modelos complexos por consumir muito tempo e fornecer resultados subótimos. Nestes casos, o método automático é mais aconselhável, apesar dele excluir a experiência e o conhecimento humano.

Atualmente, técnicas mais avançadas como o ajuste de histórico geoestatístico são aplicadas para a geração de melhores resultados. No ajuste de histórico convencional, a porosidade, permeabilidade e outras propriedades petrofísicas são modificadas utilizando-se multiplicadores. Esse procedimento pode gerar modelos geologicamente inconsistentes. Isso pode ser ainda pior quando as mudanças são realizadas regionalmente pois elas não respeitam as continuidades do reservatório. Para se manter a consistência geológica e respeitar a correlação espacial (variograma) de propriedades petrofísicas, o procedimento recomendado é

que se realize o processo de ajuste de histórico integrado ao modelamento geoestatístico (Maschio et al, 2005).

Azevedo, L. e Soares, A. (2017) resumiram o processo do ajuste de histórico geoestatístico nos seguintes passos:

1. A partir dos dados de poço disponíveis, simula-se um conjunto de modelos petrofísicos utilizando-se um modelo secundário derivado de um processo de inversão sísmica;
2. Obtém-se as produções destes modelos com o uso dos simuladores de fluidos;
3. Compara-se estas produções com os dados do verdadeiro histórico de produção;
4. Divide-se o reservatório em áreas de influência de cada poço considerado no ajuste;
5. Para cada área formada no passo anterior, seleciona-se o modelo petrofísico cujas produções mais se assemelham à produção desta determinada região;
6. Gera-se um novo conjunto de modelos petrofísicos a partir da co-simulação estocástica sequencial do modelo composto no item (5);
7. Repete-se os passos (2) a (6) até que todos os poços tenham uma produção simulada igual (ou ao menos semelhante) à produção observada pelos dados.

2.3. Programação Genética

A programação genética, ou *genetic programming* (GP), é um modelo de programação que usa a ideia de evolução biológica para resolver um problema complexo. Na GP, evolui-se uma população de programas de computação que, geração por geração, é estocasticamente transformada em uma nova, melhor população de programas. Sua natureza aleatória livra os resultados de pré concepções humanas ou enviesamentos, e a natureza adaptativa é responsável por gerar soluções comparáveis, senão melhores, que as encontradas pelo melhor esforço humano (<http://geneticprogramming.com/>). As etapas básicas do algoritmo do GP pode ser visto na Figura 2.

A GP inicia-se com a geração de centenas a milhares de programas criados aleatoriamente dentro de um conjunto de parâmetros definidos a priori. Essa população é evoluída progressivamente a partir de uma série de gerações, com o objetivo de se reduzir o valor de uma determinada função objetivo. Esta função objetivo é uma medida da distância entre as curvas de produção simulada e observada. Através do princípio Darwiniano de seleção natural, os programas de melhor desempenho, ou seja, com menor valor da função objetivo, são selecionados para reproduzir novos programas para a próxima geração. Segundo Poli et al. (2008), as principais operações que são utilizadas para a criação dos novos programas a partir dos já existentes são:

- *Crossover*: A criação do programa filho é realizada a partir de uma combinação aleatória entre partes escolhidas dos dois programas pais.

- *Mutation*: A criação do programa filho é realizada a partir de uma alteração aleatória em uma parte qualquer do programa pai.

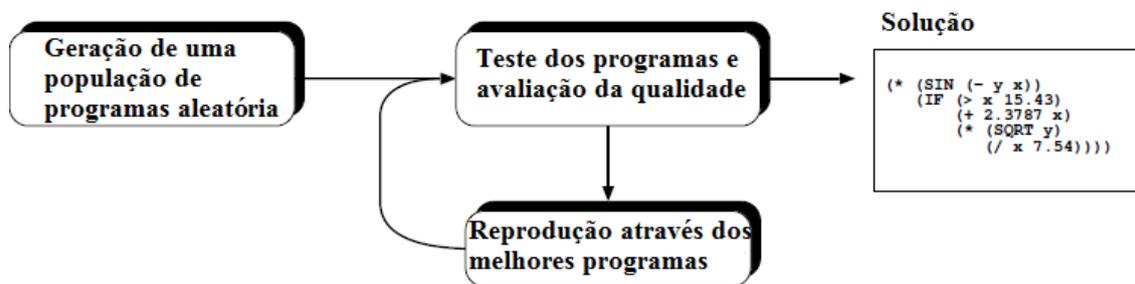


Figura 2 – Algoritmo básico da programação genética (Poli et al., 2012)

2.4. Regressão Simbólica

A regressão simbólica é um caso específico da programação genética e é um tipo regressão analítica de geração automática de modelos simbólicos que descrevem as relações funcionais em uma base de dados a partir de expressões (ou modelos) matemáticas. Utilizando o conceito da programação genética, a regressão simbólica irá promover a evolução artificial dos modelos, otimizando alguns deles. Através de uma análise que determina quão bem esses modelos se adequam à base de dados, avalia-se quais destes serão utilizados na geração dos seguintes modelos. As equações (modelos) são então formadas recombinação equações anteriores, utilizando a programação genética. Se nenhuma informação *a priori* é fornecida acerca da relação entre os dados de entrada e saída, o espaço de procura é o conjunto de todos os modelos simbólicos possíveis para as operações no conjunto de dados de entrada.

A regressão simbólica apresenta inúmeras vantagens sobre outras técnicas de *machine learning* (e.g. *support vector machines* e redes neurais) para problemas nos quais os modelos têm de ser simples expressões analíticas interpretáveis e possuir um comportamento generalizado. Além destas, esse método conta com diversos outros benefícios (Searson, 2015):

- Não necessita de suposições prévias sobre a estrutura dos modelos.
- Não requer o uso de programas especializados em modelamentos para se testar os modelos simbólicos.
- O modelo de predições final é escolhido através de um conjunto variado de modelos empíricos não lineares que são gerados automaticamente.
- A examinação das equações muitas vezes leva a *insights* sobre os processos físicos e a dinâmica envolvida.
- A simplicidade do modelo torna-o muito mais confiável do que os típicos modelos “caixa preta”.
- Fornece uma população de modelos ao invés de um único.

3. Metodologia

A implementação da regressão simbólica foi realizada através do GPTIPS2 (*Genetic Programming Toolbox for the Identification of Physical Systems*), uma plataforma de *software open source* baseada no matlab e livre de custos. Esta toolbox foi escolhida por ser simples e altamente customizável, pois permite que pesquisadores adicionem suas próprias implementações de código. A regressão simbólica foi selecionada como substituto de simuladores numéricos de fluidos numa metodologia de *history matching* geoestatístico devido à sua flexibilidade e por apresentar excelentes resultados, muitos destes superando outros métodos de *machine learning* como *Artificial Neural Networks* e *Support Vector Machines* noutras áreas não relacionadas com as geociências (Searson, 2015).

Neste capítulo são descritos detalhadamente todos os procedimentos realizados e valores adotados para se chegar nos resultados atingidos deste trabalho que serão descritos no Capítulo 4.

3.1. Geração dos modelos

A geração dos modelos matemáticos resultantes da regressão simbólica é realizada através de dois banco de dados diferentes: dados de treinamento e dados de validação. Os dados de treinamento serão os dados utilizados para a geração dos modelos – estes serão os dados que a regressão simbólica tentará reproduzir a partir das variáveis selecionadas. Neste trabalho, as produções e pressões deverão ser estimadas através de modelos tridimensionais de porosidades, permeabilidades e do tempo de produção. Já os dados de validação serão aqueles que confirmarão a qualidade do modelo matemático encontrado pela regressão simbólica e para outros dados que não sejam os dados de treinamento. Para que um modelo seja validado, é necessário que ele forneça resultados aceitáveis independente do conjunto de dados que seja testado. Para este trabalho, independente do modelo petrofísico (i.e., os pares de modelos de porosidade e permeabilidade), o modelo matemático da regressão simbólica deve sempre fornecer valores realistas para as produções e pressões, ou seja, não podemos ter valores negativos ou curvas completamente fora do esperado.

Como exemplo do uso destes dois bancos de dados, treinamento e validação, ao se simular 10 conjuntos de modelos petrofísicos no simulador de fluidos, pode-se utilizar 7-8 destes para o treinamento do modelo matemático e os outros 2 a 3 para a validação do modelo encontrado. É importante que fique claro que a análise da validação é feita pelo usuário posteriormente e que esta não influencia a evolução do modelo de regressão simbólica. Portanto no ciclo de ajuste de histórico proposto no âmbito desta tese, que é completamente automático, esta validação é dispensada, assim todos os 10 modelos petrofísicos são utilizados para o treinamento.

O treinamento dos modelos matemáticos na plataforma GPTIPS2 é realizado através de um arquivo principal onde se seleciona as funções desejadas, e outro que configura os parâmetros da regressão simbólica e seleciona as bases de dados que serão utilizadas. O ajuste de cada um destes parâmetros tem como objetivo reduzir o valor do *misfit* entre as produções simuladas no Eclipse® (Schlumberger) e os resultados obtidos pela regressão simbólica. A descrição dos principais parâmetros, assim como a configuração utilizada pode ser vista no Anexo A.

3.2. Preparação dos dados

Os dados adquiridos após a simulação numérica dos modelos petrofísicos em simuladores de fluidos, que resultam no conjunto de dados de produção simulados, necessitam passar por um pré-condicionamento antes que possam ser utilizados para o treinamento dos modelos de regressão simbólica. Este pré-processamento é consequência da necessidade do *upscaling* dos modelos petrofísicos, da formatação dos arquivos de treinamento e validação, e a solução do problema com as extensões dos arquivos. Estes processos são realizados através do uso de um programa escrito em matlab e foram totalmente automatizados para sua aplicação no ciclo de ajuste de histórico automático desenvolvido e implementado no âmbito desta tese.

Os modelos petrofísicos de porosidade e de permeabilidade possuem um valor para cada célula do modelo 3D que representa o reservatório. Como essas representações possuem milhares de células, e como os valores de cada célula representam uma variável para o modelo de equações a ser aproximado pela regressão simbólica, é necessário que se realize um *upscaling* destes modelos a fim de se reduzir a quantidade de variáveis independentes durante a simulação. Este passo de pré-processamento é necessário para reduzir o custo computacional da metodologia proposta tendo em conta que não deve existir um compromisso entre esta redução e o detalhe perdido na caracterização das propriedades petrofísicas de interesse. Todo o procedimento de *upscaling* está descrito em detalhe no Anexo B.

3.3. Integração de regressão simbólica no ajuste de histórico geoestatístico

O processo comum do ajuste de histórico geoestatístico, já descrito anteriormente, requer o uso contínuo dos simuladores de fluidos para se obter as produções e pressões para cada conjunto de modelos de porosidade e permeabilidade gerados através de simulação sequencial estocástica. As produções e pressões resultantes são então comparadas com o histórico de produção real dos poços e a diferença resultante conduz o processo iterativo à convergência. A principal limitação destas metodologias é o tempo gasto no passo de simulação numérica de fluidos. Com a integração da regressão simbólica no ajuste de histórico geoestatístico, espera-se reduzir o tempo gasto com as inúmeras simulações numéricas de fluidos sem no entanto perder a capacidade de ajustar a produção simulada e os modelos

petrofísicos convergirem para a realidade geológica. Esta seção descreve os passos adotados para a metodologia proposta no âmbito desta tese:

1. Inicialmente, através da simulação sequencial direta (Soares 2001), gera-se um determinado número de modelos de porosidade os quais são utilizados com variável secundária na co-simulação directa com distribuições conjuntas do mesmo número de modelos de permeabilidade. Durante a primeira iteração, utiliza-se apenas os dados dos poços como dados experimentais. Para as seguintes iterações, o modelo composto da última iteração também condiciona a geração dos modelos;
2. Os 10 primeiros pares de modelos de porosidade e permeabilidade são utilizados num simulador numérico de fluidos e são obtidas as respectivas produções simuladas;
3. Após a décima simulação, ocorre automaticamente o pré-processamento dos dados mencionados na seção 3.2 da metodologia e descritos em detalhe no Anexo B. Como resultado temos o upscaling os modelos petrofísicos;
4. O arquivo pronto, composto pelas 10 simulações, é utilizado para o treinamento dos modelos da regressão simbólica. Após isso, o modelo (expressão matemática) resultante é salvo para ser utilizado na etapa seguinte. Esta etapa é repetida para cada variável de produção de cada poço que será utilizado no ajuste de histórico. No caso deste trabalho, esta etapa foi repetida 48 vezes para o reservatório Watt-field (3 variáveis e 16 poços; Secção 4.2);
5. Para as seguintes simulações (a partir da 11^a), são aplicados os modelos matemáticos obtidos na etapa anterior para assim se obter as produções e pressões desejadas. Esta é a etapa que substitui o uso dos simuladores de fluidos. A utilização dos modelos matemáticos em substituição do simulador numérico de fluidos permite uma redução do tempo computacional;
6. A qualidade de cada modelo de porosidade e permeabilidade é avaliada por uma função objetivo que compara as produções geradas quer pelo simulador de fluidos quer pelo modelo de regressão simbólica.
7. Os modelos que petrofísicos que geram produções mais próximas das observadas para cada área de influência dos poços existentes são selecionados para condicionar a geração de novos modelos na próxima iteração pela co-simulação sequencial direta (Co-DSS). Os valores da função objectivo são transformados em coeficientes de correlação e então vão controlar a convergência do processo iterativo.

Ao final de todas as iterações, os gráficos das simulações de cada iteração, os valores obtidos pela função objetivo, e o melhor modelo petrofísico alcançado são analisados e comparados com os resultados obtidos pelo ciclo tradicional de ajuste de histórico geoestatístico. Um esquema simplificado deste algoritmo pode ser visualizado na Figura 3.

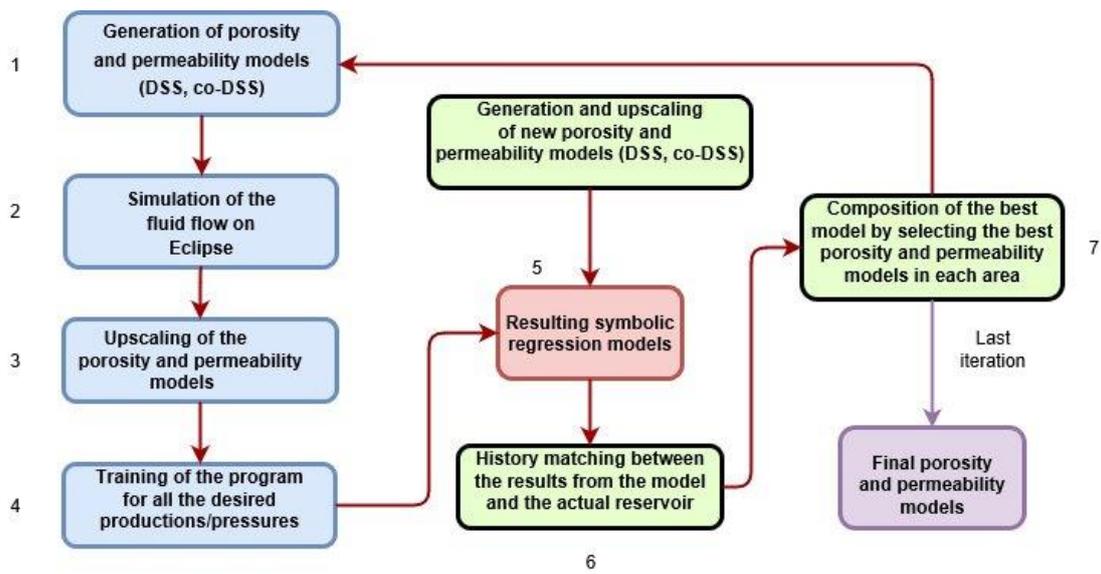


Figura 3 – Diagrama simplificado da metodologia proposta de ajuste de histórico de produção geoestatístico com integração de regressão simbólica.

4. Exemplos de aplicação

Para a obtenção dos resultados finais apresentados neste capítulo foram efectuadas diferentes parametrizações da regressão simbólica a fim de se testar a capacidade da técnica ou melhorar os resultados alcançados. Nesta seção será apresentado a evolução do trabalho através dos resultados dos principais exemplos de aplicação realizados.

4.1. Reservatório sintético simples

Para verificar a validade da utilização de regressão simbólica na previsão de produções de reservatórios de hidrocarbonetos através de modelos tridimensionais da subsuperfície, realizou-se um teste em um reservatório sintético mais simples. Este reservatório é composto por 6400 blocos com dimensão 20x20x16 (cada um com seu respectivo valor para porosidade e permeabilidade) e apresenta 5 poços no total: 4 produtores e 1 de injeção. Trabalhou-se com o resultado de 1000 simulações geradas ao longo de 50 iterações (20 simulações por iteração) pelo algoritmo apresentado anteriormente.

Após o passo de *upscaling*, o reservatório passou a ter apenas 100 células, cada uma com valores para porosidade e para permeabilidade. Juntamente com o tempo de produção (dividido em 66 *timesteps*), há um total de 201 variáveis para se encontrar o valor das produções desejadas.

Para o treino da regressão simbólica, utilizou-se os seguintes parâmetros no regressão simbólica do GPTIPS 2:

- Tamanho da população – 2000
- Número de gerações – 100
- Funções – *Times*, *Minus*, *Plus*, *Square*, *Exponential*, *Cube*, *Negative*, *Absolute*, *Mult3*, *Add3*. As últimas 2 funções são somas e multiplicações de 3 valores, em vez de apenas 2.
- Número de genes e profundidade – Variáveis conforme a necessidade por um modelo mais complexo.

Para os outros parâmetros desta *toolbox* de regressão simbólica, utilizou-se os valores padrões propostos (“*default values*”).

4.1.1. Caso 1: WOPR Well P3 – 10 genes/5 gene depth

Neste primeiro teste para o *Well Oil Production Rate* (WOPR), utilizou-se 10 genes com uma profundidade de 3. O objetivo inicial era de treinar o modelo matemático da regressão com metade das simulações (500) para tentar prever a produção da outra metade. O modelo matemático alcançado foi capaz de reproduzir as produções reais com um erro desprezável. De modo a avaliar o comportamento do método numa situação mais difícil, decidiu-se utilizar apenas

10 modelos de porosidade e permeabilidade e suas correspondentes produções para o treinamento do da regressão simbólica. O resultado alcançado é bom quer para o caso de treino (Figura 4) quer para o caso de teste/validação (Figura 5). O modelo matemático resultante da regressão simbólica, representado na Equação (1), consegue reproduzir os dados de taxa de produção de óleo corretamente. Na equação, x_1 é a variável do tempo de produção, x_2 a x_{101} são os valores para a permeabilidade e x_{102} a x_{201} são os valores para a porosidade. A Tabela 1 sumariza de forma quantitativa a performance da previsão de WOPR para este caso em estudo.

$$\begin{aligned}
 WOPR_{caso\ 1}^{P3} = & 3290 + 1,88 * (x_{177} + x_{101} - x_{172} - x_1) + 1,01 * (|x_{64} - x_1|) - \\
 & 0,541 * (x_{66} - x_{63} - x_{177} - x_{142}) - 4,6 * 10^{-4} * (x_{48} - x_1) - (|x_{48} - x_1|) + \\
 & 0,804 * (x_{198} * x_{x1} + x_{177} * x_{42} - x_{59} - x_{82}) - 7,3 * 10^{-4} * (x_{101}^2 * x_{98} + 2x_{80}^2) + \\
 & 1,41 * (x_1^2 * x_{157} * x_{177} * x_{178}) - 1,27 * (|x_{97} - x_1|) + 1,2 * 10^{-3} * (x_{109} + x_1)^2
 \end{aligned} \quad (1)$$

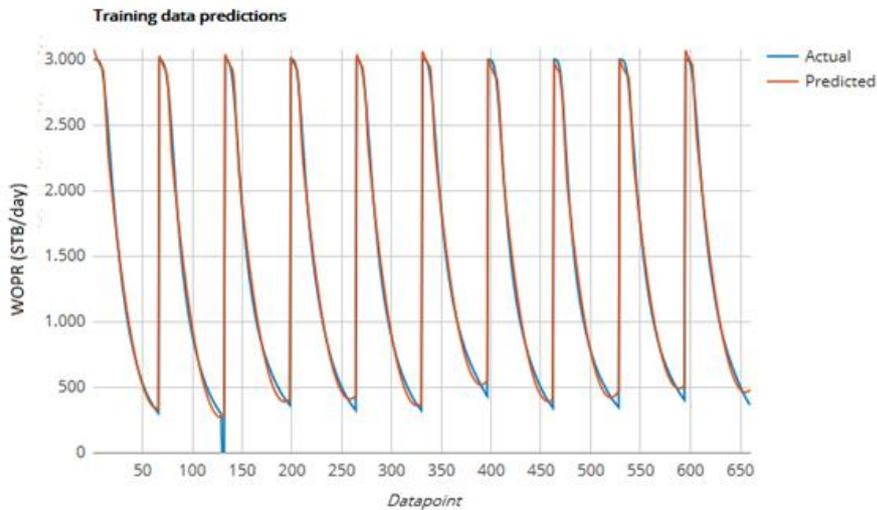


Figura 4 – WOPR Well P3: Predições para os dados de treinamento do caso 1 (modelo mais complexo)

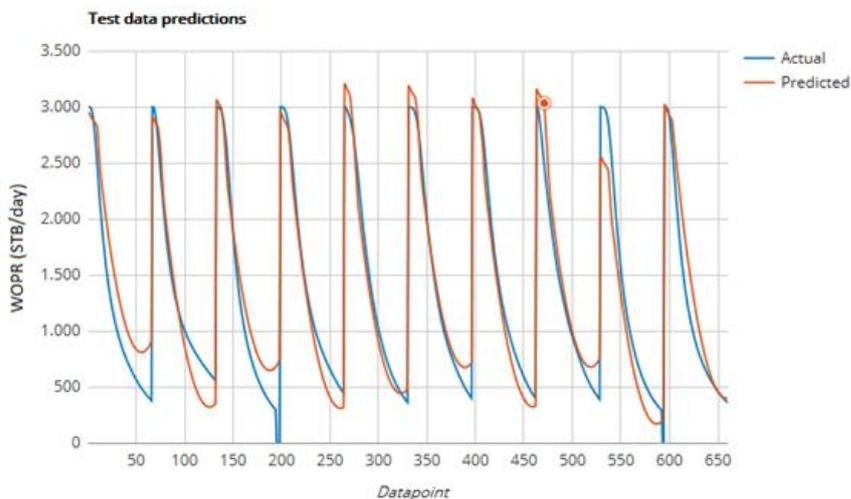


Figura 5 – WOPR Well P3: Predições para os dados de teste (validação) do caso 1 (modelo mais complexo)

Tabela 1 – WOPR Well P3: R² e erros dos dados de treinamento e validação do caso 1 (modelo mais complexo)

	Dados de Treinamento	Dados de Teste
R ²	0,996	0,934
Erro médio quadrático	57,4	224,7
Erro Absoluto Médio	46,0	184,9
Erro absoluto Máximo	275,0	734,7

4.1.2. Caso 2: WOPR Well P3 – 3 genes/3 genes de profundidade

A fim de se testar a capacidade da ferramenta com uma parametrização menos adequada que a utilizada no caso de estudo anterior, decidiu-se utilizar um modelo matemático a ser previsto pela regressão simbólica muito mais simples, com apenas 3 genes de profundidade 3 para equacionar a taxa de produção de óleo do poço P3. Os resultados podem ser visualizados pelas Figura 6 e Figura 7 e pela Tabela 2. Os resultados apresentaram uma correspondência muito boa tanto para os dados de treinamento (Figura 6) quanto para os dados de validação (Figura 7). De uma maneira geral, o modelo matemático obtido (que pode ser observado na equação 2), mais simples que no caso de estudo anterior, permite prever convenientemente as taxas de produção de óleo.

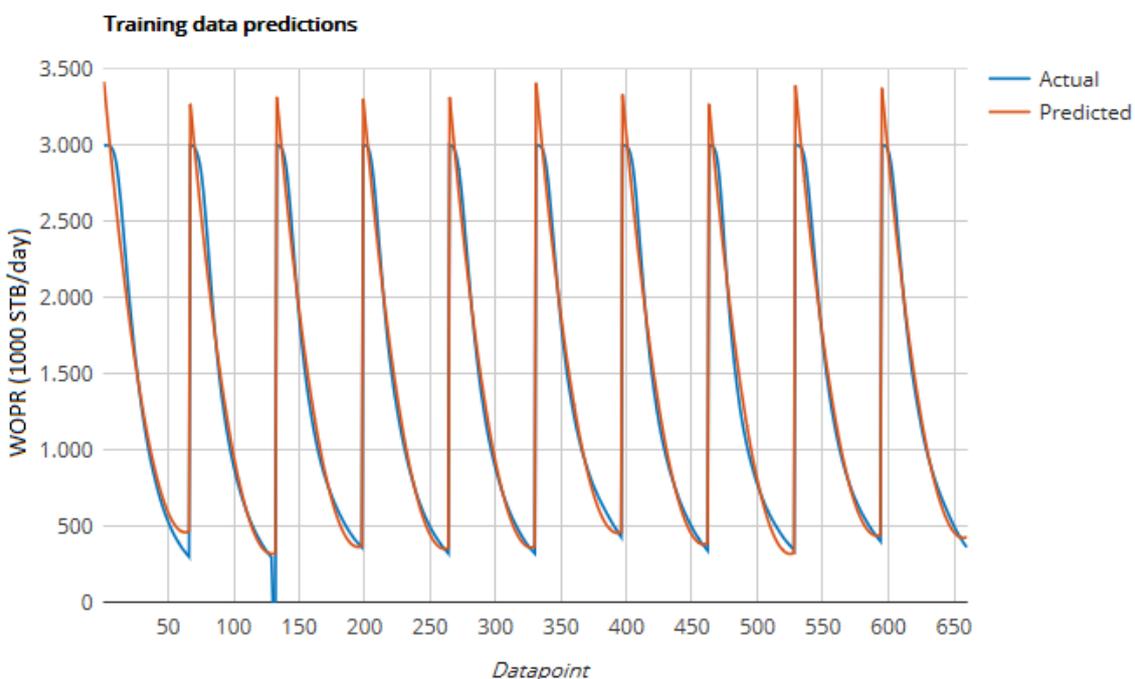


Figura 6 – WOPR Well P3: Predições para os dados de treinamento no caso 2 (modelo mais simples)

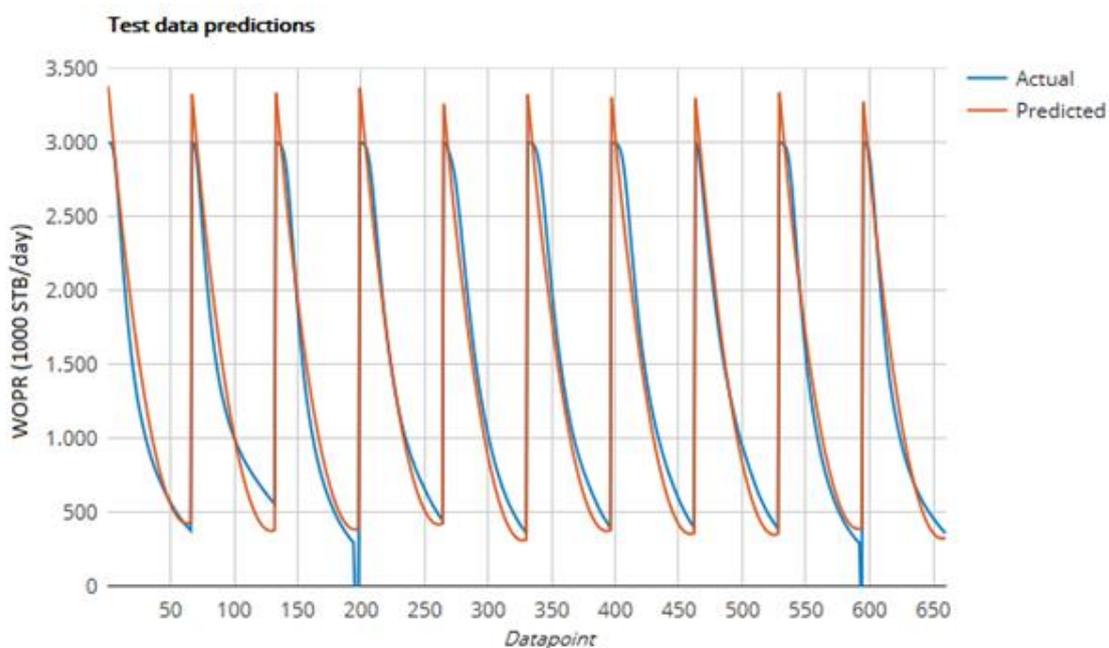


Figura 7 – WOPR Well P3: Predições para os dados de validação no caso 2 (modelo mais simples)

Pela Tabela 2 é possível interpretar, em comparação como caso anterior, que os erros para os dados de treinamento são maiores ao se usar modelos de regressão mais simples (e.g. Erro médio quadrático: 57,4 para o modelo anterior contra 108,0 deste modelo mais simples). No entanto, este caso se inverte para a validação, no qual podemos observar uma diminuição destes erros (e.g. Erro médio quadrático de 224,7 para 166,4). Além de provarmos a capacidade e simplicidade desta ferramenta, podemos verificar também o risco de um possível *overfitting* ao utilizarmos um excesso de genes nos modelos.

Tabela 2 - R² e erros dos dados de treinamento e validação do caso 2 (modelo mais simples)

	Dados de Treinamento	Dados de Teste
R ²	0,986	0,964
Erro Médio Quadrático	108,0	166,4
Erro Absoluto Médio	79,8	141,9
Erro Absoluto Máximo	415,7	394,9

O modelo matemático da regressão encontrado pode ser visualizado pela Equação (2~~Error! Reference source not found.~~). Observe a aparição da variável x1 em 2 dos 3 genes. Modelos bastante simplificados são obrigados a fazer as previsões apenas com as variáveis mais importantes. Neste caso, o tempo (x₁) é a variável com maior influência sobre as produções. Além do tempo, valores de porosidade (x₂ a x₁₀₁) e permeabilidade (x₁₀₂ a x₂₀₁) de diferentes regiões do poço foram escolhidas.

Como este reservatório sintético era bastante simples, mesmo um modelo pouco complexo foi facilmente capaz de reproduzir bem as curvas de produção e pressão. Esta facilidade não foi sempre possível para todos os casos de um reservatório mais complexo, como será visto na seção seguinte.

$$WOPR_{caso\ 2}^{P3} = 3350 - 3,18 * (x_{154}^2 * x_{42} + x_{42} * x_{118} * x_{113} + 2x_{180} + x_1) + 0,142 * (x_{46} * x_{97} * x_{150} * x_{182}) + 8,3 * 10^{-4} (x_1^2 - x_{64} * x_{52}) \quad (2)$$

4.2. Watt-field

O reservatório Watt-field é um reservatório semi-sintético baseado no estudo de caso de um ambiente deposicional de rio entrelaçado (Barrela, 2017). O modelo do reservatório original utilizado apresenta as dimensões de 226x59x40 células nas direções *i*-, *j*- e *k*- respectivamente. Para o treino da regressão simbólica foi efetuado o *upscaling* do reservatório para 45x14x20 células nas direções *i*-, *j*- e *k*- respectivamente. Para o ciclo de ajuste de histórico deste reservatório foram realizadas 20 simulações em cada uma do total de 6 iterações para a metodologia com regressão simbólica e 18 simulações para as 6 iterações da metodologia tradicional. Para ambas as metodologias, com ou sem a regressão simbólica, foram utilizados os mesmos dados de entrada.

O reservatório Watt-field apresenta 6 poços injetores que circundam os 16 poços produtores que são utilizados para o ajuste de histórico (Figura 8). O ajuste é realizado através de 3 variáveis: Pressão no fundo do poço (WBHP), Taxa de produção de óleo (WOPR) e taxa de produção de água (WWPR). A Figura 8 Mostra em detalhes a disposição dos 16 poços produtores deste reservatório. Cada zona representada nesta figura corresponde a uma área de influência para cada poço produtor calculada através de polígonos de Voronoi. Estas regiões são utilizadas para localmente gerar os melhores modelos de cada iteração que irão condicionar a iteração seguinte.

Em relação ao exemplo anterior, e mesmo com o *upscaling* dos modelos gerados através de simulação e co-simulação, o treinamento da regressão simbólica neste reservatório ainda apresentou um grande problema em relação ao tempo de computação: como foi descrito na Seção 3.3 da Metodologia, realiza-se um treinamento para cada produção ou pressão a qual se deseja prever (e.g. WOPR – Poço 7); devido ao grande número de poços, de nós da malha do reservatório a etapa de regressão simbólica é ainda computacionalmente bastante pesada. Outro fator que aumentou consideravelmente o tempo computacional, quando comparado com o exemplo anterior, foi a grande quantidade de *timesteps* em cada simulação. Como a simulação do reservatório Watt-field fornecia os dados de produção para 97 datas diferentes, o processo de regressão simbólica considerou um total de 970 (97 datas * 10 simulações) dados para o treinamento dos modelos de regressão.

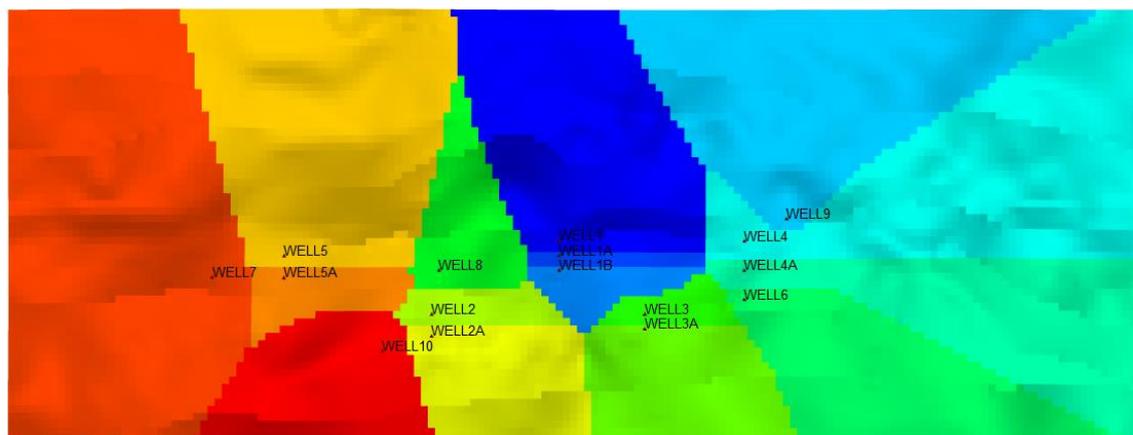


Figura 8 – Localização dos 16 poços do reservatório Watt-Field e polígonos de Voronoi associado a cada poço

4.2.1. Ajuste de histórico geoestatístico tradicional (com simulador de fluidos)

De forma a comparar a qualidade dos resultados obtidos através da metodologia proposta, foi primeiro realizado um ajuste histórico geoestatístico recorrendo à metodologia proposta por Barrela (2016). Os resultados foram satisfatórios quer a nível global quer a nível local. Os resultados podem ser analisados através da evolução do comportamento das curvas de produção para as variáveis em estudo ao longo das iterações (Figura 9 e Figura 10).

É possível observar que a cada iteração as curvas de produção correspondentes aos modelos petrofísicos simulados ao longo do processo iterativo (em azul) se aproximam mais da produção real (em preto). As linhas vermelhas representam os limites aceitáveis de erro (i.e., curvas que estejam dentro deste intervalo são consideradas como tendo um bom ajuste aos dados de produção reais). Na primeira iteração, a maioria das curvas apresentava um comportamento bastante diferente do observado pelos dados de produção. No entanto, com o passar das iterações, pouco a pouco é possível interpretar que as curvas se convergem para a curva real como podemos ver pela iteração 6, no qual as curvas azuis apresentam uma proximidade muito maior da curva preta, e pela iteração 6, na qual muitas das curvas já estão dentro dos limites aceitáveis do erro.

Para alguns poços, no entanto, não se observa um bom ajuste, como mostra a Figura 10. Pela figura observa-se novamente o comportamento bastante variado na primeira iteração, mas neste caso as iterações seguintes não se aproximam da curva real. Todas passam a ter um comportamento semelhante, uma vez que os modelos de porosidade e permeabilidade que lhes dão origem serão mais próximos do real, mas isto não é suficiente para se garantir a convergência correta das curvas. É possível que a função objetivo calculada para este poço tenha se prendido a um mínimo local devido às outras 2 variáveis (WBHP e WOPR). As curvas de produção destas 2 variáveis podem ser visualizadas pelo Anexo C. Este problema ocorreu para apenas alguns poços. É importante ressaltar que a não convergência resulta da complexidade do reservatório em estudo e do comportamento contraditório de alguns poços.

Em ambas as figuras podemos interpretar que a primeira iteração é sempre aquela cujos modelos petrofísicos apresentam maior variação, uma vez que são apenas condicionados aos dados de poço. Isto é importante para que se amplie o espaço de busca por um modelo adequado. Após a primeira iteração, os modelos já foram condicionados pelo melhor conjunto de modelos da iteração anterior através da co-simulação sequencial direta tendo em conta o *misfit* das curvas de produção geradas para cada modelo simulado e as curvas de produção observadas para as três variáveis de produção simultaneamente.

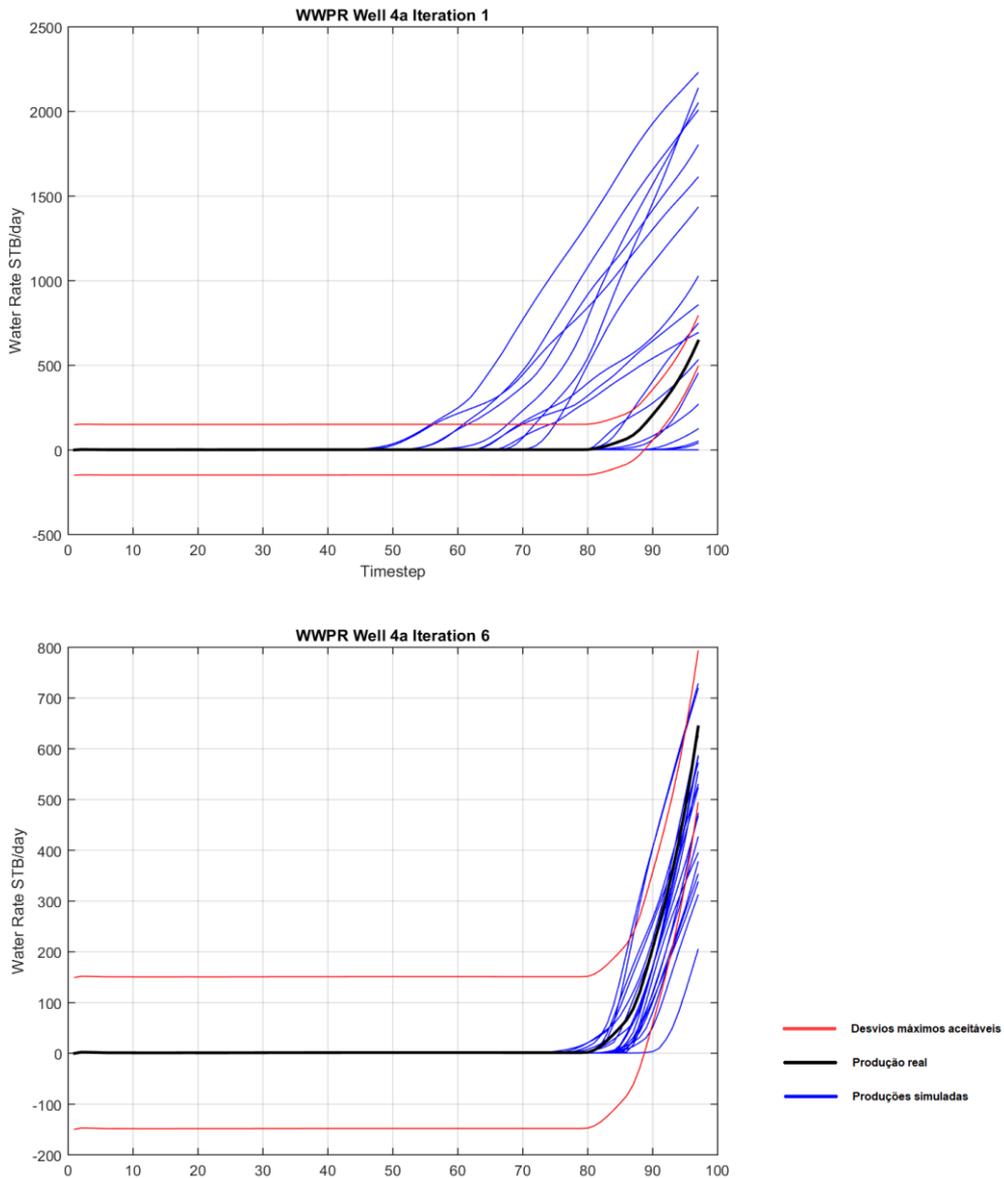


Figura 9 – Watt-field WWPR Well 4a: Iterações 1 e 6 do ajuste de histórico sem regressão simbólica que ilustram boa convergência dos modelos petrofísicos.

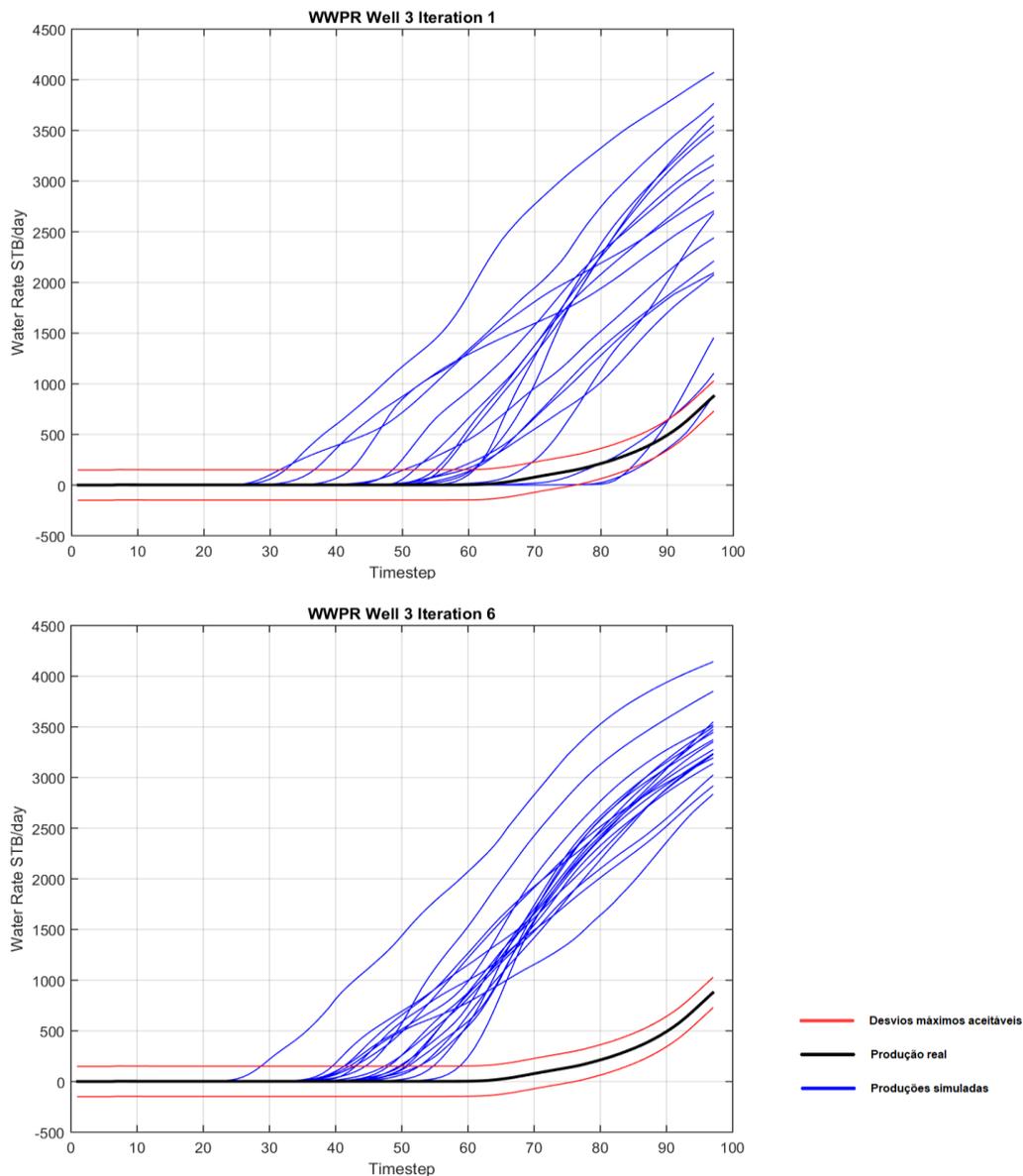


Figura 10 – Watt-field WWPR Well 3: Iterações 1 e 6 do ajuste de histórico sem regressão simbólica que ilustram má convergência dos modelos petrofísicos.

Por fim, pela Figura 11 é possível se analisar a evolução do menor valor atingido pela função objetivo global para cada uma das seis iterações realizadas. A figura ilustra a tendência de diminuição de seu valor conforme se passam as iterações e o valor mínimo sendo atingido na iteração 4 ($1,41 \cdot 10^6$). Este é o resultado esperado uma vez que indica uma redução das diferenças entre as curvas de produção e pressão para todos os dezasseis poços produtores do reservatório. Observa-se também que o valor para cada iteração não necessariamente irá cair em relação à iteração anterior. Isto ocorre tanto para a iteração 3 como para a iteração 5, onde é possível se observar um pequeno aumento em relação à iteração antecedente. Este fenômeno não se põe como um problema, uma vez que o ponto principal desta análise é garantir que o menor valor da função objetivo tem um comportamento decrescente em relação ao número da

iteração. Caso fosse possível realizar um grande número de iterações, essa tendência ficaria ainda mais evidente.

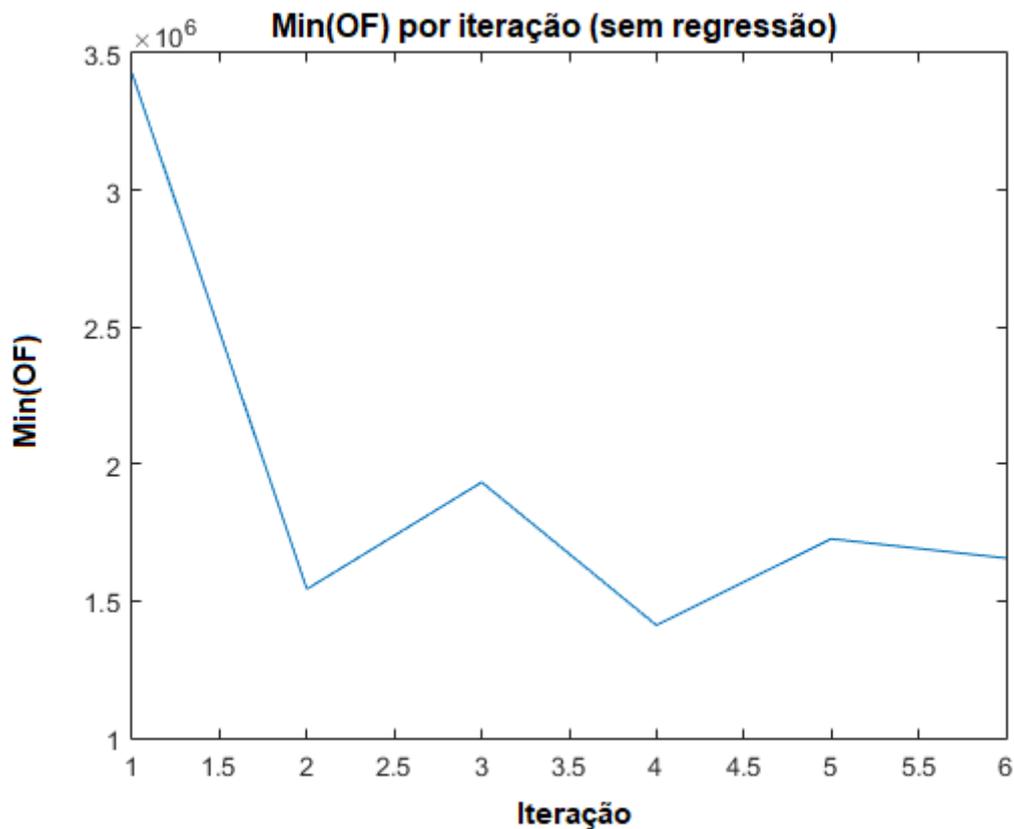


Figura 11 – Watt-field: Menor valor da função objetivo global por iteração (ajuste de histórico geostatístico com simulador de fluidos)

4.2.2. Ajuste de histórico com regressão simbólica

A geração dos modelos de regressão simbólica foi condicionada a parâmetros que variam conforme a dificuldade de se encontrar um bom modelo matemático que seja capaz de reproduzir corretamente a produção ou pressão desejada. Estes parâmetros variaram da seguinte forma:

- Tamanho da população: 800-5000
- Número de gerações: 30-50
- Número de execuções: 1-10
- Funções: “times”, “minus”, “plus”, “square”, “mult3”, “add3”, “cube”, “negative”, “absolute” e funções do tipo “if-then-else”.
- Número de genes: 3-10
- Profundidade dos genes: 3-6

Nesta seção são comparados os conjuntos de simulações entre o ciclo de ajuste de histórico sem a regressão simbólica (gráficos à esquerda) e com a regressão simbólica (gráficos à direita). Todos os resultados obtidos podem ser resumidos basicamente nos quatro casos a seguir, ilustrados pelas Figura 12, Figura 13, Figura 14 e Figura 15

Para os casos ilustrados pela Figura 12, poço 8, a convergência ocorre independente da metodologia de ajuste histórico incluir ou não regressão simbólica. Podemos observar que na metodologia com a regressão simbólica todas as 20 produções simuladas são bastante semelhantes, independente do uso do simulador de fluidos ou do modelo matemático proveniente da regressão, o que indica que o modelo encontrado pela regressão simbólica foi capaz de simular bem o comportamento dos poços.

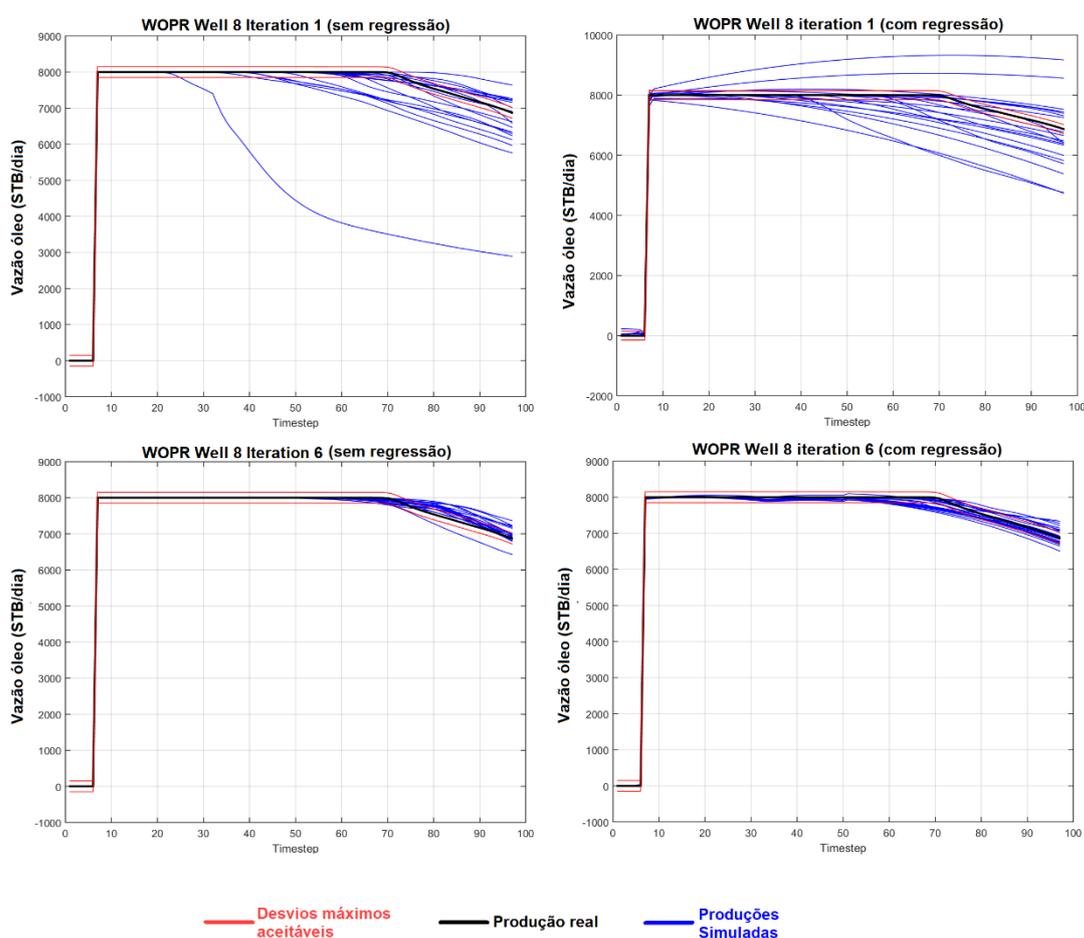


Figura 12 – Watt-field WOPR Well 8: Ocorre boa convergência das curvas de produção simuladas pelo algoritmo tradicional (esquerda) e com a regressão (direita) após as 6 iterações.

É interessante notar que neste caso em específico a convergência dos modelos petrofísicos do ciclo com regressão simbólica foi melhor que o ciclo tradicional. Comparando-se os valores da função objetivo para este poço e variável, o menor valor obtido para o algoritmo original foi de $1,13 \cdot 10^3$ STB/dia enquanto a média foi de $6,27 \cdot 10^3$ STB/dia, já para o algoritmo com a regressão simbólica, estes valores foram de $0,66 \cdot 10^3$ STB/dia e $5,89 \cdot 10^3$ STB/dia, respectivamente. Enquanto diversos modelos do novo algoritmo se mantiveram, em geral, dentro

dos desvios máximos, apenas alguns modelos do algoritmo tradicional tiveram resultados semelhantes. Este é um excelente indicativo do trabalho que mostra que a economia de tempo não reflete necessariamente em uma má qualidade dos resultados.

Para se analisar um dos modelos matemáticos gerados para este reservatório, escolheu-se justamente o caso do WOPR do poço 8, uma vez que esta variável para este poço apresentou boa convergência na última iteração. A Equação (3) apresenta o modelo matemático selecionado. Novamente, a variável x_1 indica o tempo de produção, x_2 a x_{12601} indicam porosidade e x_{12602} a x_{25201} indicam permeabilidade. Este é um modelo bastante complexo, apresentando 8 genes com profundidade máxima de 6. Percebe-se que diferentes valores de porosidade e permeabilidade são escolhidos para a geração do modelo, mas na maioria dos genes a variável “x1” está presente. Esta é a variável que representa o tempo, sendo essencial para se reproduzir corretamente o comportamento do poço. Por ser um modelo mais complexo, podemos perceber também a presença da função do tipo “if-then-else” que segue: “If $A \geq B$, then $X=1$, else $X=0$ ”.

Para os casos da Figura 13, não há convergência em nenhuma das metodologias testadas após 6 iterações. Assim como na Figura 10, estes modelos podem ter se prendidos em algum mínimo local que não permitiu uma boa convergência das curvas. Mesmo que em nenhum dos casos as curvas tenham convergido bem, percebe-se que o algoritmo com a regressão simbólica teve um desempenho ligeiramente melhor. A função objetivo para este poço e variável tiveram como valor mínimo $0,99 \cdot 10^5$ STB/dia no algoritmo com a regressão e $3,96 \cdot 10^5$ STB/dia sem ela. Além disso as médias foram de $6,40 \cdot 10^5$ STB/dia e $8,19 \cdot 10^5$ STB/dia respectivamente. Note que, como os modelos da regressão foram treinados a partir de maus resultados da simulação no reservatório, suas curvas também apresentaram um comportamento bastante diferente da curva real.

A Figura 14 apresenta situações em que uma variável de um determinado poço é ajustada logo na primeira iteração. Estas situações ocorreram unicamente para as pressões de fundo do poço (WBHP) de alguns poços, independente do uso da regressão ou não. Este fato acontece devido à pouca sensibilidade do reservatório a esta variável de produção e ao fato de se estar a ignorar os valores de taxas de injeção dos poços injetores existentes. Nestes casos as curvas simuladas já se encontram dentro dos desvios limites, mas é importante observar que com o passar das iterações todos os modelos petrofísicos foram aperfeiçoados e suas curvas de produção para esta variável chegaram próximos à uma convergência ideal.

$$\begin{aligned}
WOPR_{Watt-field}^{Well\ 8} = & 0,0075 + 6,9 * 10^{-3} * Y * (x_{9910}^3 * x_{10082} * (x_1 - x_{21345})) + \\
& 7,53 * (x_{16643} - x_{23753}) + 15,9 * Z * (x_{9627} + |x_{21219}| - |x_{19228}|) + \\
& 0,14 * (x_{14253} * x_1 - x_1 * |x_{15727}| + x_{3703} + x_{14253}) - \\
& 3,5 * 10^{-4} * ((W + (x_{1059} * x_{16418})^2 - x_{17001} + x_{8041}) * (x_{21345} - x_1^2) + x_{19742}) - \\
& 1,2 * 10^{-4} * (x_{15752} + |x_{21219}| - |x_{15413}|) * x_1^2 + 0,48 * (|x_1 * x_{17001}|) + \\
& 0,16 * (|(x_1 - |x_{23753}|) * x_{16643}|)
\end{aligned}
\tag{3}$$

if $-x_1 - x_{7415} \geq x_{22090} + x_{23753} - x_{16643}, Y = 1$
else $Y = 0$
If $x_{16643} - x_{23753} \geq 6x_1 + x_{3653}^2, Z = 1$
else $Z = 0$
If $x_{12026}^2 \geq 75,2, W = 1$
else $W = 0$

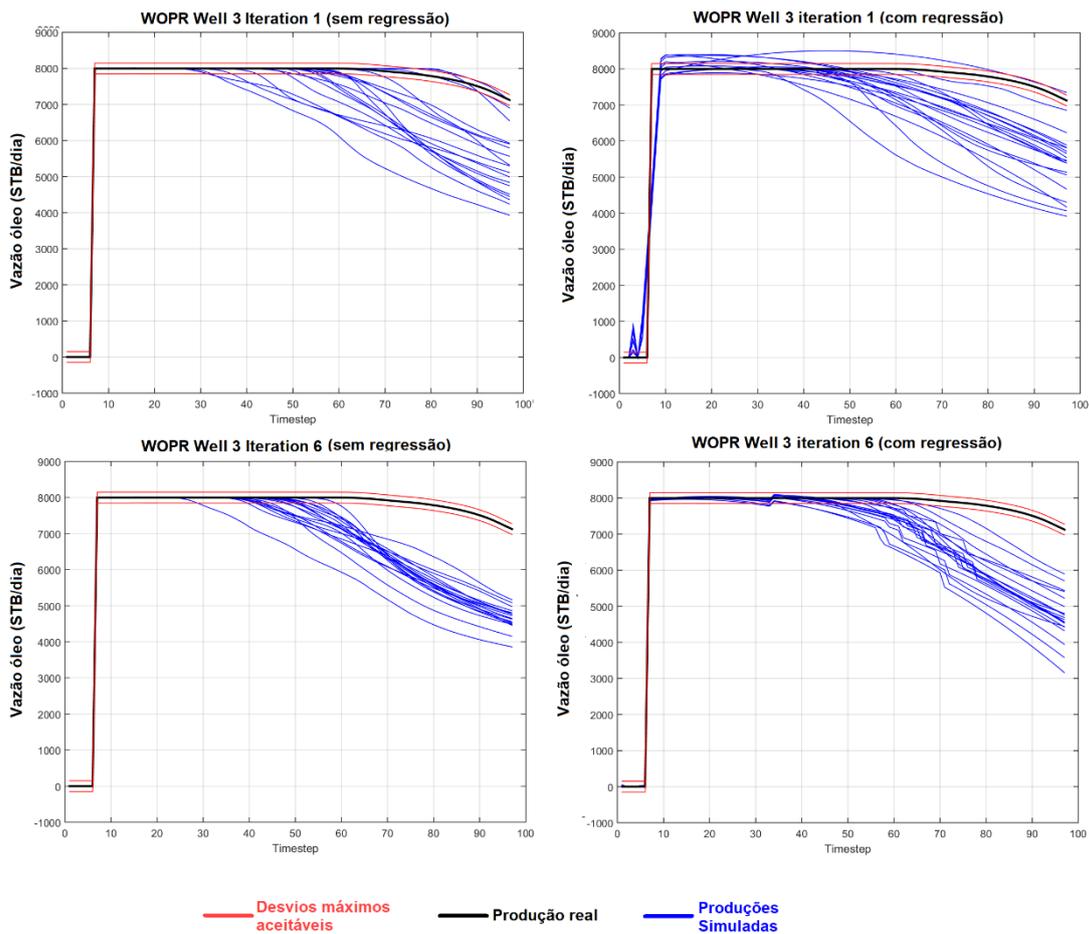


Figura 13 – Watt-Field WOPR Well 3: Iterações 1 e 6 do ajuste de histórico com regressão simbólica que ilustram má convergência dos modelos petrofísicos.

A interpretação da Figura 14 permite ainda identificar que os modelos simulados pela regressão simbólica falham ao reproduzir o comportamento do poço logo ao se iniciar a produção. Neste período o poço tem uma elevada queda de pressão e um curto período de estabilização até que decaia constantemente. O algoritmo genético, por não ter informações prévias sobre os dados inseridos, apresenta dificuldades em encontrar um modelo matemático que consiga ajustar corretamente à esta região. Além disso, como se trata de uma região pequena, o algoritmo dará prioridade em tentar equacionar corretamente a região mais extensa (comportamento médio) das curvas.

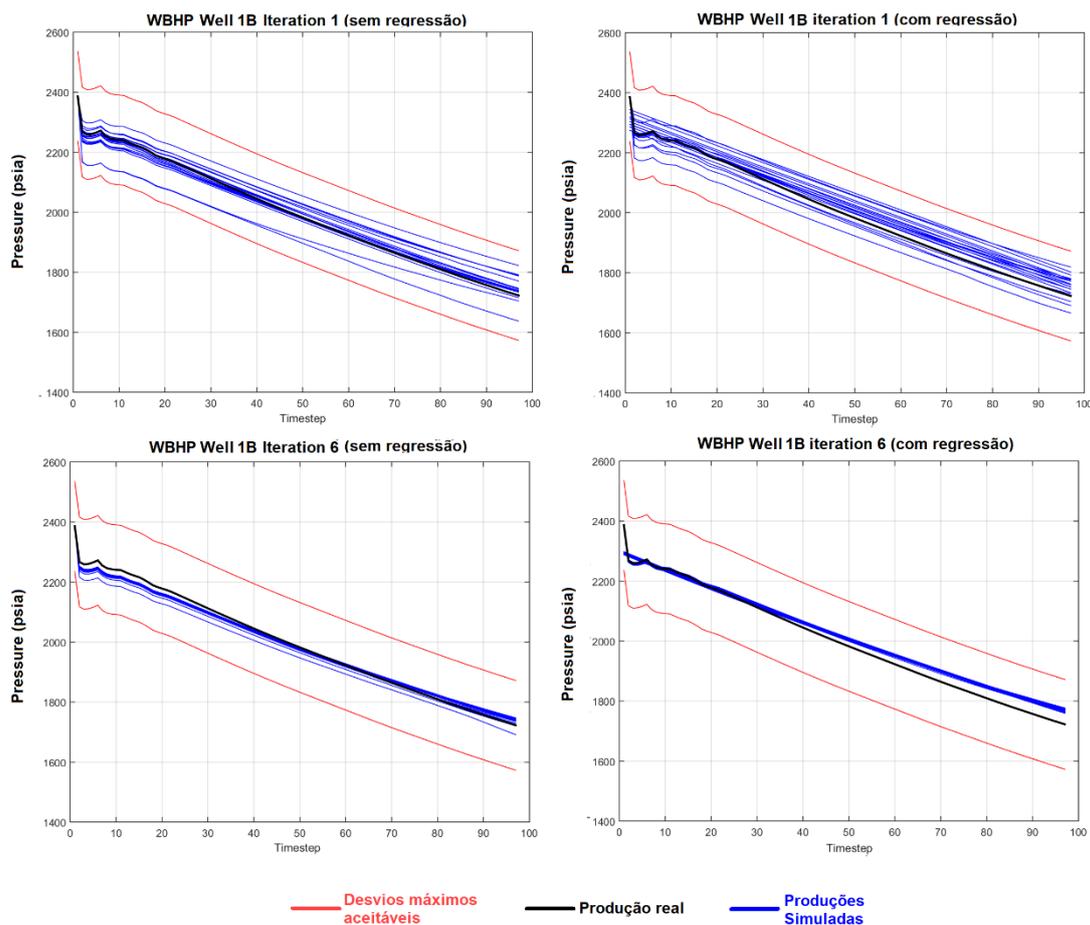


Figura 14 – Watt-field WBHP Well 1B: iterações 1 a 6 que ilustram um caso no qual as curvas simuladas já se encontram dentro dos limites de erro desde a primeira iteração

A Figura 15 é um exemplo onde a regressão simbólica não é sempre capaz de reproduzir o comportamento do reservatório. Como o treinamento ocorre a cada iteração, mesmo que o treinamento do modelo matemático falhe em uma das iterações, ele não necessariamente irá falhar em todas. No entanto, se uma produção ou pressão específica apresenta constantemente modelos de regressão que falham ao simular as produções e pressões, isso é um sinal de que é necessário se alterar os parâmetros na configuração da regressão, especialmente o número de execuções (*runs*).

Neste exemplo, apenas na sexta iteração os modelos da regressão conseguiram reproduzir corretamente o comportamento das curvas. Todas as outras iterações (que podem ser visualizadas no Anexo D) apresentaram algum desvio no comportamento da curva. Isto pode ser um dos motivos que levaram as produções do ciclo tradicional serem mais próximas da curva ideal. Como as curvas simuladas pela regressão simbólica eram todas bastante diferentes, a definição do melhor modelo petrofísico nesta determinada região foi baseada apenas pelos 10 modelos que passaram pelo simulador de fluidos, e não por todos os 20 modelos gerados na co-simulação ao final da iteração anterior. Esta redução em modelos por simulação é responsável por uma convergência mais lenta das curvas (enquanto no ciclo tradicional as curvas já convergiram, o ciclo com a regressão ainda não apresenta curvas dentro dos limites de desvio). Este é um caso em que se necessitaria um gasto de tempo maior na geração do modelo matemático proveniente da regressão simbólica.

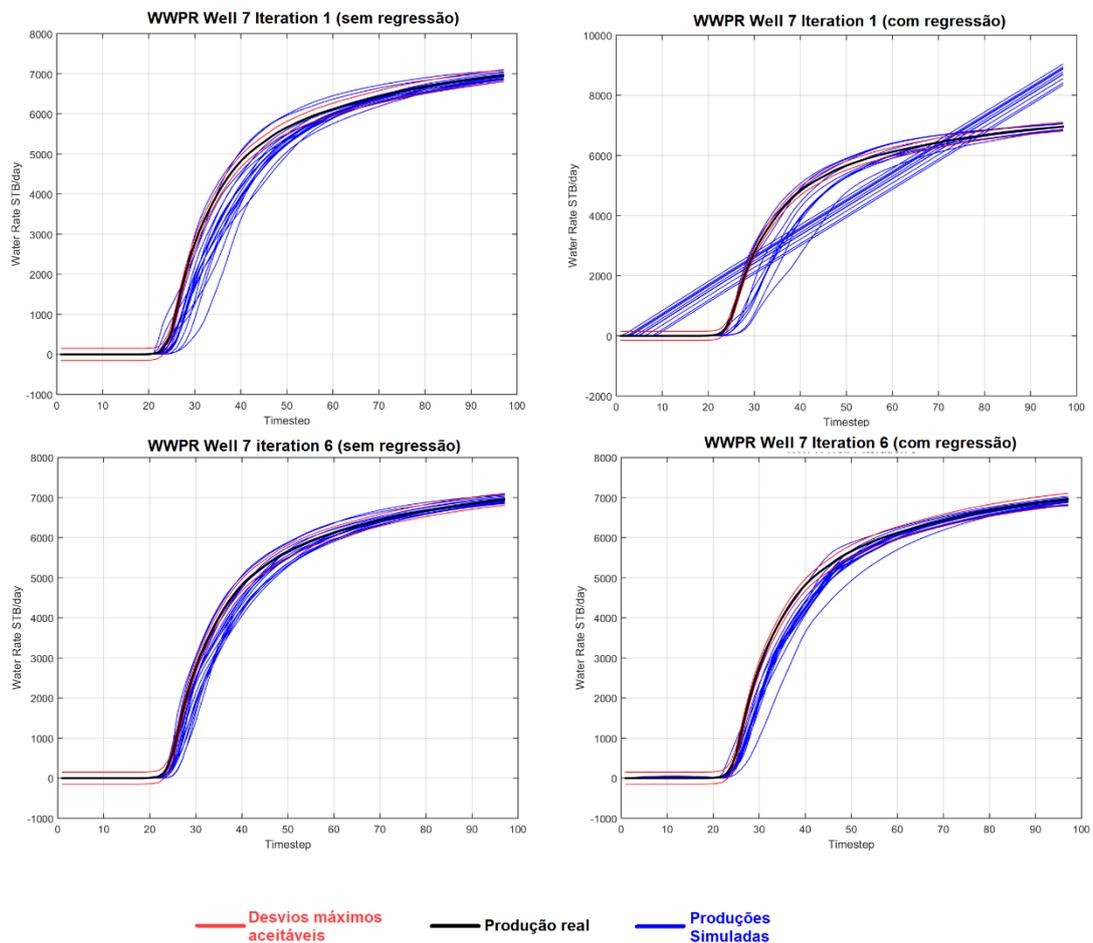


Figura 15 – Watt-field WWPR Well 7: iterações 1 a 6 que ilustram um caso no qual a regressão simbólica não é capaz de sempre reproduzir corretamente o comportamento do poço para todas as iterações

Pelo gráfico da Figura 16 realiza-se uma análise do menor valor obtido pela função objetivo global para as 6 iterações realizadas no ciclo de ajuste de histórico com a regressão

simbólica. Assim como no caso original com o simulador de fluidos, o valor mínimo da função objetivo global tem uma tendência negativa com o passar das iterações. Há um declínio constante até a terceira iteração, na qual se observa o menor valor atingido entre as 6 iterações ($1,68 \times 10^6$). Assim como explicado anteriormente, o aumento nos valores em relação à iteração anterior que pode ser observado pelas iterações 4 e 6 é normal, contanto que a tendência da curva seja em geral decrescente.

De toda maneira, ao se comparar as curvas e os valores mínimos obtidos pelos gráficos da Figura 11 e da Figura 16 (valores mínimos de $1,41 \times 10^6$ e $1,68 \times 10^6$, respectivamente), nota-se que, ao menos nestes ciclos de ajuste de histórico, o algoritmo original com simuladores de fluidos apresentou uma convergência ligeiramente melhor em relação ao algoritmo com a regressão simbólica. Isto não significa que este resultado não poderia ser invertido caso tivessem sido realizadas mais iterações. No entanto, é de salientar que esta diferença não é significativa e que a diminuição do tempo de computação obtido com a metodologia proposta permite por exemplo gerar mais pares de modelos simulados ou mais iterações num tempo menor que a abordagem convencional. Este aumento do número de modelos espera-se que diminua o valor da função objectivo e que permita uma melhor exploração do espaço dos parâmetros do modelo.

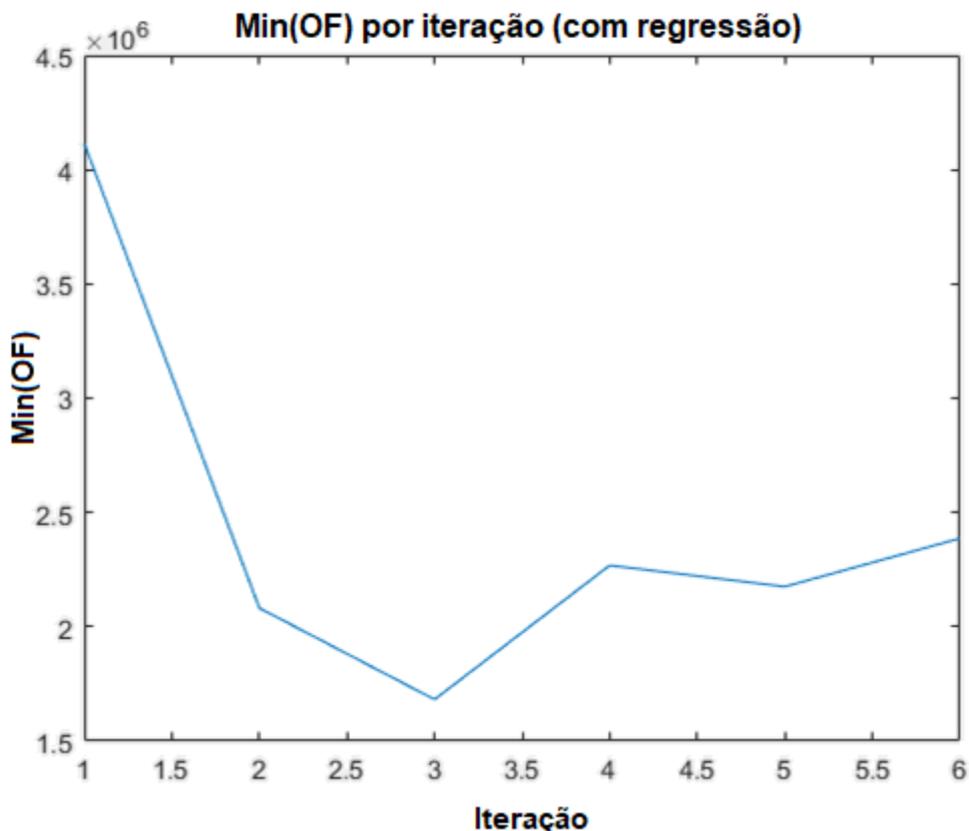


Figura 16 – Watt-field: Menor valor da função objetivo global por iteração (ajuste de histórico geostatístico com regressão simbólica)

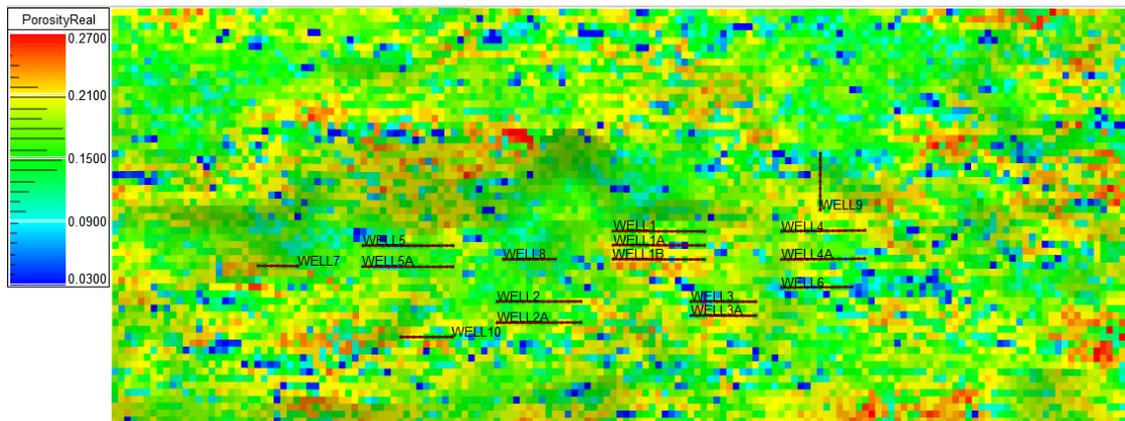
Por fim, o número de variáveis por poço bem ajustadas por cada método é comparado através da Tabela 3. Considerando-se apenas as 6 iterações realizadas, o ajuste é considerado bom para todas as variáveis em 9 dos 16 poços utilizados no ajuste de histórico para ambas as metodologias. O critério adotado foi a comparação das curvas da primeira e última iteração (caso houve uma melhora significativa entre as iterações), contanto que as curvas na última iteração se encontrem, em geral, dentro dos limites máximos de desvio da curva real de produção. Podemos verificar que a metodologia com a regressão simbólica pode ser tão eficiente quanto a metodologia original. Também é possível interpretar que o ajuste é mais fácil em alguns poços do que outros, uma vez que a maioria dos poços que tiveram um bom ou um mau ajuste na metodologia tradicional também tiveram o mesmo resultado na metodologia proposta. Caso fossem realizadas mais iterações, é possível que os ajustes tivessem resultados ainda melhores

Tabela 3 – Número de variáveis com bom ajuste por poço da metodologia tradicional e com a regressão simbólica.

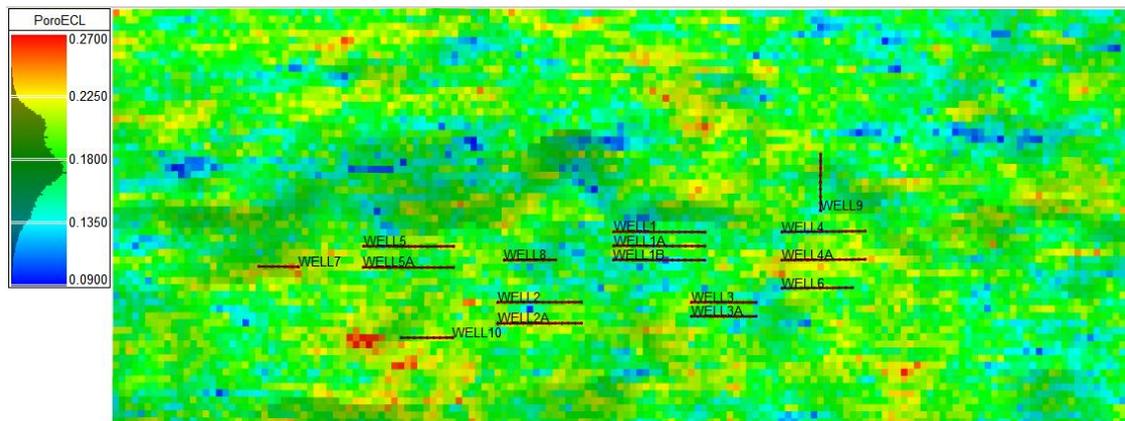
Poços	Metodologia Tradicional	Metodologia Proposta	Poços	Metodologia Tradicional	Metodologia Proposta
1	3	1	4a	3	3
1a	1	3	5	3	3
1b	1	3	5a	3	1
2	1	1	6	3	3
2a	3	3	7	3	3
3	1	1	8	3	3
3a	1	1	9	1	1
4	3	3	10	1	1

4.2.3. Análise estatística dos modelos petrofísicos gerados no final do processo iterativo

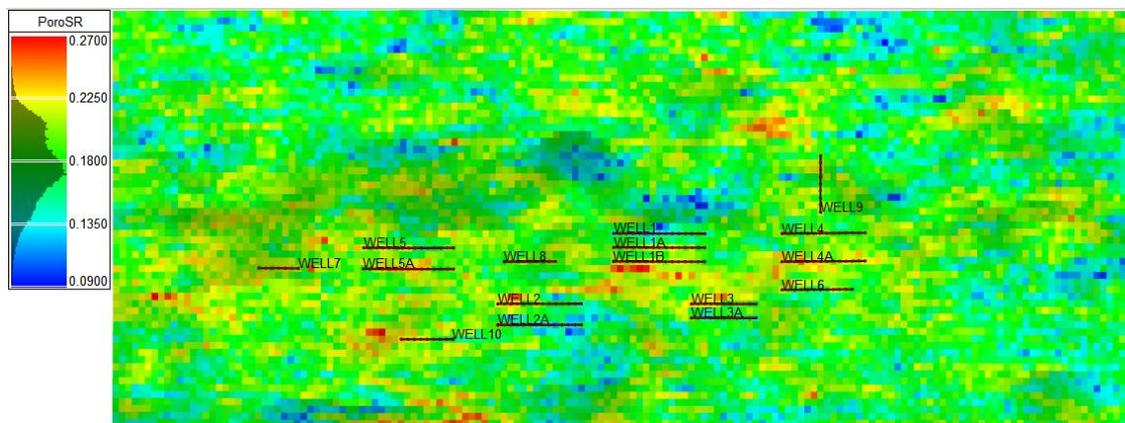
Por fim, para se avaliar o método adotado, realiza-se uma comparação entre os modelos reais e os modelos finais (i.e., o conjunto de modelos petrofísicos que gerou o menor valor de função objectivo) obtidos pelo ajuste de histórico tradicional e com a regressão simbólica. A Figura 17 ilustra cortes horizontais dos modelos de porosidade e permeabilidade próximas ao centro do reservatório. Ao se comparar as três imagens, observa-se que não existe uma grande correlação entre as três imagens obtidas. A variância local dos modelos simulados é muito menor que a imagem real, como pode ser muito bem observado pelas imagens das duas propriedades. É possível notar a presença de valores de alta e baixa porosidade/permeabilidade muito próximos um dos outros na imagem real.



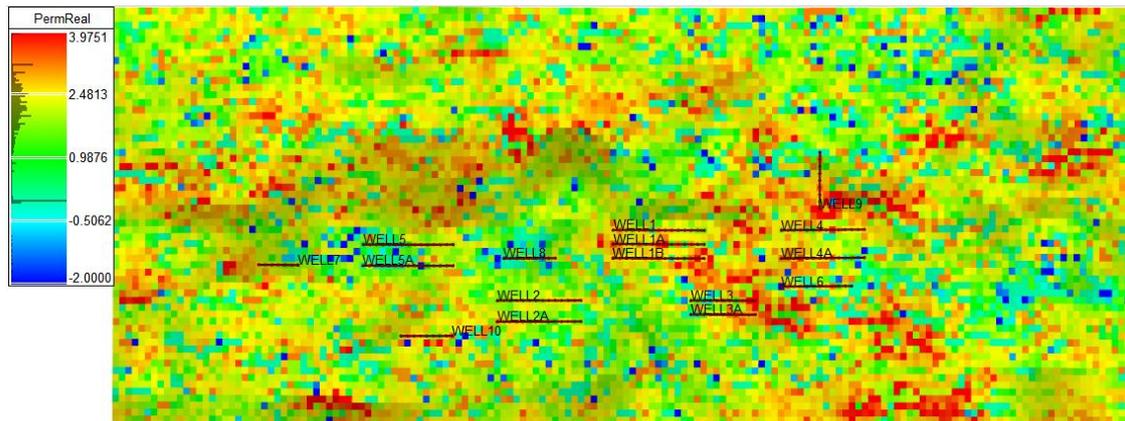
(a) Seção horizontal extraída da porosidade real reservatório Watt-Field



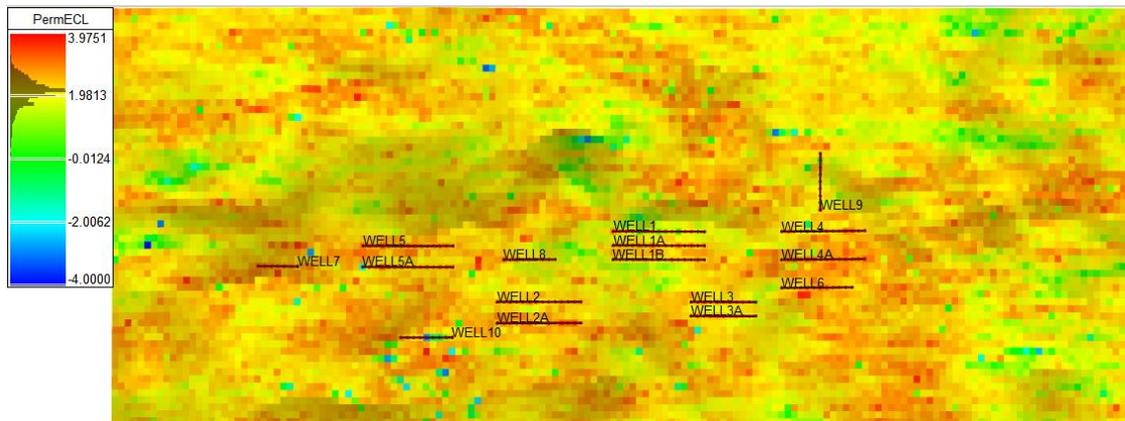
(b) Seção horizontal extraída do melhor modelo de porosidade obtido pelo ajuste de histórico sem regressão simbólica



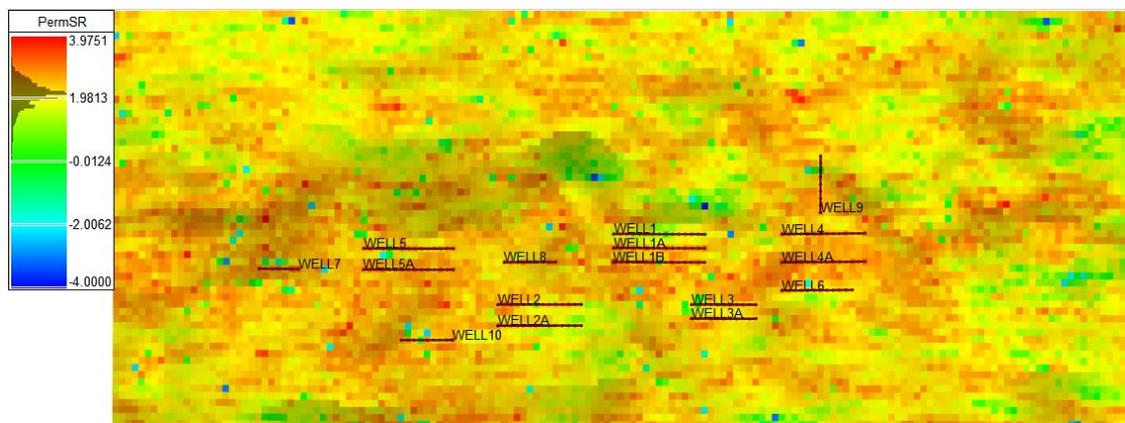
(c) Seção Horizontal extraída do melhor modelo de porosidade obtido pelo ajuste de histórico com regressão simbólica



(d) Seção horizontal extraída da permeabilidade real reservatório Watt-Field



(e) Seção horizontal extraída do melhor modelo de permeabilidade obtido pelo ajuste de histórico sem regressão simbólica



(f) Seção horizontal extraída do melhor modelo de permeabilidade obtido pelo ajuste de histórico com regressão simbólica

Figura 17 – Corte horizontal dos modelos reais (em cima) de porosidades/permeabilidades e dos modelos finais do ajuste tradicional (meio) e com a regressão (embaixo). Valores da escala de permeabilidade representam $\text{Log}_{10}(\text{perm})$ em mD.

Ainda na Figura 17, esta apresenta seções horizontais extraídas dos melhores modelos de porosidade e permeabilidade obtidos pelo ajuste de histórico com e sem a regressão simbólica. É possível observar que a semelhança dos modelos inferidos é variável dependendo da zona do modelo em consideração para ambos os métodos aplicados no âmbito desta tese. Estas diferenças são esperadas devido à natureza do próprio caso de estudo, altamente não

estacionário e heterogêneo. Esta diferença é consequência da natureza do ajuste de histórico que nos fornece múltiplos modelos bem ajustados, uma vez que este não é um problema de solução única. A existência de diferentes modelos permite um estudo de probabilidade, e consequentemente a redução do risco de erro, o que é uma vantagem para a tomada de decisões.

Apesar da diferença visual entre os modelos inferidos e reais, isto não significa que o ajuste de histórico em termos de curvas de produção não foi correto. O ajuste tem como objetivo apenas tentar reproduzir o comportamento do poço, mas dificilmente será capaz de encontrar o modelo exato das propriedades petrofísicas, especialmente devido à pouquíssima quantidade de informação que temos acerca do reservatório. De toda forma, isto não significa que os modelos possam ser completamente diferentes do reservatório real. Assim, uma análise estatística é necessária para se validar os modelos.

Uma característica dos modelos da simulação sequencial direta é que eles devem reproduzir os histogramas dos dados experimentais. No entanto, antes de se realizar uma comparação entre os modelos petrofísicos e os dados reais, realiza-se uma comparação entre os dois métodos aplicados. Podemos notar pela Figura 18 que os modelos simulados pelo algoritmo tradicional e com a regressão nos fornecem histogramas praticamente idênticos. A média, variância e mediana tanto para a porosidade quanto para a permeabilidade também apresentam uma diferença ínfima. Desta forma, podemos utilizar qualquer um destes dois para se comparar com as estatísticas com os dados reais.

O gráfico Q-Q (Quantil-Quantil) é um método gráfico para se comparar duas distribuições de probabilidades. Caso elas sejam semelhantes, os pontos no gráfico Q-Q formarão aproximadamente uma reta de 45° ($y=x$). Caso as distribuições sejam linearmente relacionadas, os pontos formarão uma reta, mas não necessariamente de 45° . O uso deste gráfico facilita muito a comparação entre dois histogramas, uma vez que ao se plotar um histograma, a escolha dos *bins* afeta muito a sua forma, podendo levar a conclusões precipitadas. Por este motivo utilizou-se esta forma de análise na Figura 19 para se comparar a estatística dos dados dos poços com a dos modelos encontrados no ajuste de histórico com a regressão simbólica. Para o gráfico da permeabilidade, podemos notar que os pontos formam quase uma reta de 45° , apesar de suas imperfeições. O mesmo vale para a porosidade. Apesar da escala do gráfico passar a impressão de que a reta não é de 45° , ao se observar os valores podemos notar que ela é, de fato, bastante próxima da reta $y=x$. Observe que as médias são bastante parecidas e a variância é maior para os dados dos poços, assim como se era esperado.

Tendo validado os modelos com os dados dos poços utilizados durante as simulações, finalmente realiza-se a comparação estatística com os resultados dos modelos reais do reservatório. Os resultados foram muito bons e podem ser observados pela Figura 20. Mesmo que os dados dos poços sejam dados reais, sua média e variância podem ser diferentes do reservatório em si. Portanto em uma comparação entre os modelos da regressão simbólica e os

modelos reais do reservatório, espera-se encontrar valores próximos, mas não necessariamente idênticos. Isto pode ser verificado através da Figura 20, na qual os valores da média e da mediana são bastante próximos. Os valores da variância dos modelos reais são consideravelmente maiores, como foi observado pela Figura 17 e já explicado anteriormente. Analisando-se os gráficos Q-Q, é possível verificar que os pontos formam uma reta tanto para a porosidade quanto para a permeabilidade. Apesar dessas retas não serem de 45°, elas indicam que há uma correlação linear das distribuições destas propriedades entre os modelos reais e os modelos petrofísicos obtidos com a regressão simbólica.

Tendo em consideração todos os resultados estatísticos analisados, conclui-se que a análise estatística realizada nesta seção valida os bons resultados gerados pelas simulações e co-simulações sequenciais diretas realizadas ao longo do ciclo de ajuste de histórico com regressão simbólica.

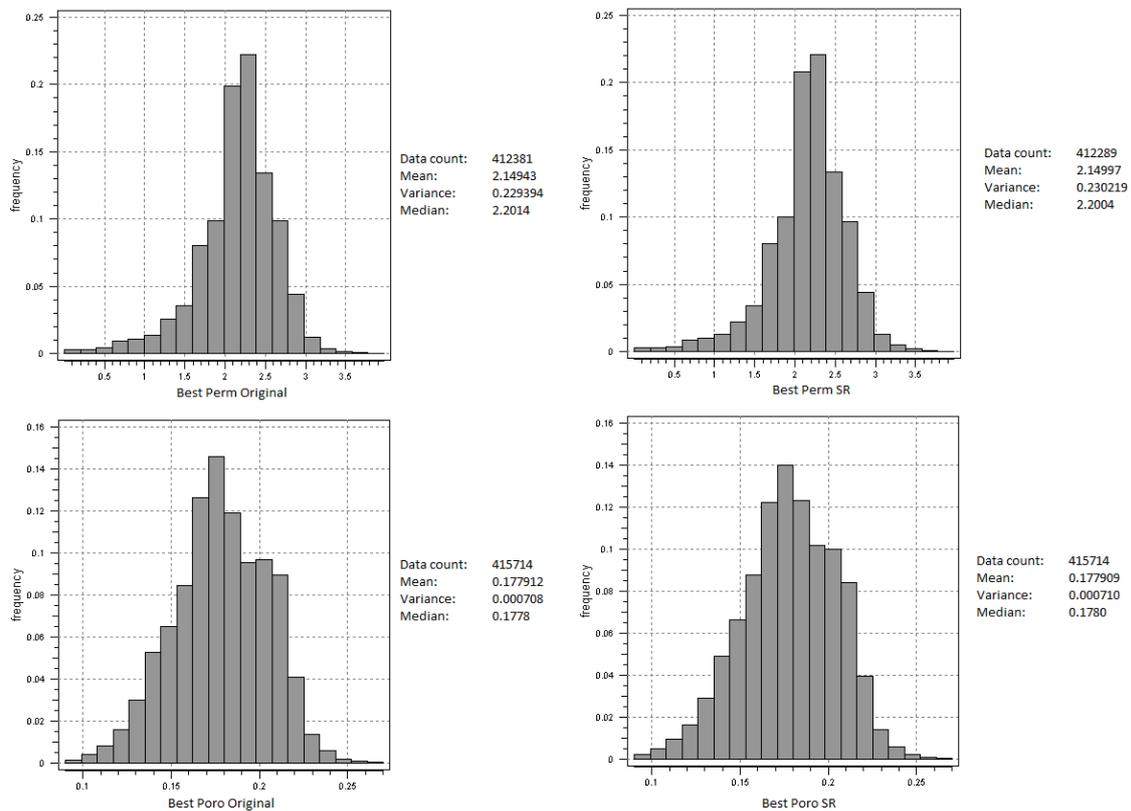


Figura 18 – Histogramas de porosidade e permeabilidade com o algoritmo original (esquerda) e com a regressão (direita)

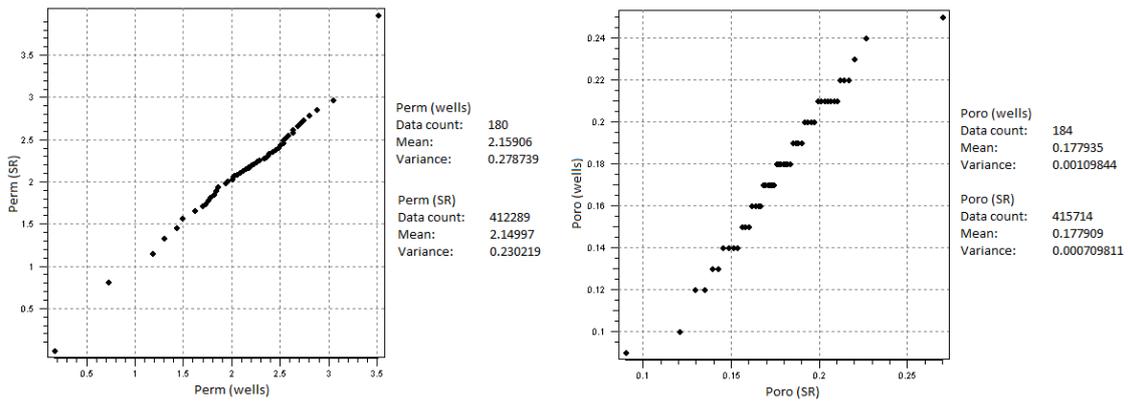


Figura 19 – Gráfico Q-Q entre os dados experimentais dos poços e o modelo matemático obtido com a regressão simbólica (permeabilidade à esquerda; porosidade à direita)

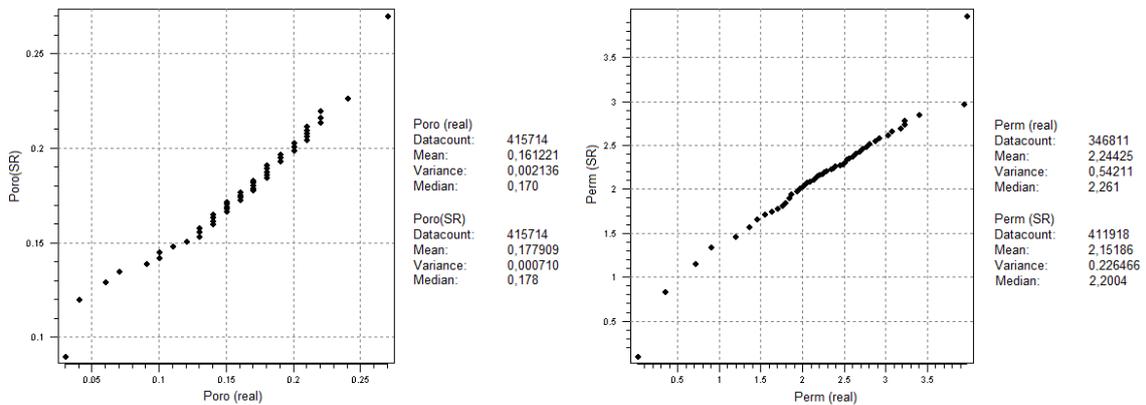


Figura 20 – Gráfico Q-Q entre os modelos reais e os modelos obtidos pela regressão simbólica. (Porosidade à esquerda e Permeabilidade à direita)

5. Conclusões

Este trabalho demonstrou a possibilidade de se substituir parcialmente o uso de um simulador de fluidos, neste caso de aplicação o Eclipse® (Schlumberger) em um ciclo de ajuste de histórico por um algoritmo de regressão simbólica. Esta foi uma proposta inovadora cujos resultados atingidos estão em linha quando comparados com os resultados do algoritmo tradicional. É de notar que o uso de regressão simbólica em geociências está ainda num estágio inicial o que aumentam o potencial dos resultados atingidos no âmbito desta tese. A regressão simbólica provou ser uma ferramenta poderosíssima e forneceu excelentes resultados na predição de muitas das produções e pressões, porém, não foi possível garantir bons modelos matemáticos para todas estas variáveis do ajuste de histórico. Nos casos no qual ela falha, a convergência dos modelos petrofísicos é mais demorada, mas ainda assim ocorre. Para se contornar o problema é necessário que se gaste mais tempo no treinamento do modelo matemático desta produção em específico.

Para se discutir sobre a redução no tempo do ajuste pela metodologia proposta, é necessário se comentar sobre dois factores do ajuste dos reservatórios apresentados neste trabalho. A primeira é o tempo da simulação num simulador de fluidos convencional como é o caso do Eclipse® (Schlumberger): o tempo de simulação dos reservatórios é considerável e aumenta com o aumento da dimensão do campo e a complexidade do modelo. A segunda diferença crucial é o tempo gasto no treinamento dos modelos da regressão simbólica: o número de produções e pressões que serão ajustadas irá depender do número de poços e das variáveis escolhidas. Para cada uma delas será necessário se treinar um modelo matemático. Além disso, cada simulação terá um número diferente de *timesteps*, o que significa mais dados para se realizar os treinamentos. Todos esses factores resultam em um tempo muito maior dos treinamentos, podendo superar inclusive o tempo das próprias simulações no Eclipse® (Schlumberger). É importante destacar que, por iteração, o tempo gasto com os treinamentos e com as 10 simulações no Eclipse® (Schlumberger) é fixo e independe do número de simulações por iteração. Desta forma, ajustes de histórico com diversas simulações por iteração serão beneficiados pela metodologia proposta (o treinamento só ocorre uma vez por iteração e, depois disso, todo modelo petrofísico simulado pela regressão terá uma resposta quase que instantânea).

Um ponto essencial que deve ficar claro é que é impossível de se treinar um conjunto de modelos matemáticos de regressão que seria capaz de prever com precisão as produções e pressões para qualquer modelo petrofísico selecionado de qualquer iteração futura. Um teste foi realizado no qual utilizou-se apenas a regressão simbólica na tentativa de se obter um bom ajuste de histórico. Como nenhuma simulação no Eclipse® (Schlumberger) foi feita, os modelos de regressão utilizados necessitaram ser previamente gerados. Os modelos selecionados foram os obtidos na sexta iteração do ajuste de histórico do Watt-field. Com eles tentou-se realizar o ajuste de histórico do reservatório, mas os resultados não foram tão bons, sendo que apenas cerca de

35% das produções/pressões convergiram pra uma solução correta e muitas das outras tiveram previsões bastante diferentes. Este experimento demonstrou a necessidade que existe de se realizar o treinamento para cada iteração no processo do ajuste de histórico.

De toda maneira este trabalho produziu resultados de elevado potencial que puderam ser observados por um relativamente bom ajuste de histórico em apenas 6 iterações. O estudo das técnicas de *machine learning* em engenharia de reservatórios, especialmente na simulação de reservatórios ainda é bastante novo. Os bons resultados obtidos neste trabalho se mostram apenas como um pequeno passo em direção a uma área com um enorme potencial, e que ainda será responsável por enormes avanços nas tecnologias atuais, inclusive para o tema abordado.

Referências

- Baker et al., "History Matching Standards; Quality Control and Risk analysis for Simulation", Paper 2006-129 (2006 Canadian International Petroleum Conference)
- Barrela, et al. (2017) "Geostatistical history matching – A Zonation-based Approach Using Direct Sequential Simulation". 79th EAGE Conference & Exhibition, Paris, França, 12-15 Junho 2017.
- Vladislavleva et al., (2009). "Order of nonlinearity as a complexity measure for models generated by symbolic regression via pareto genetic programming" IEEE Transactions on Evolutionary Computation. IEEE.*
- Genetic Programming, disponível em "<http://geneticprogramming.com>". Acedido a 13 de Janeiro de 2017.
- Azevedo, L., Soares, A. (2017) "Geostatistical Methods for Reservoir Geophysics". Springer International Publishing AG
- Maschio et al., (2015). "A New Framework For Geostatistics-Based History Matching Using Genetic Algorithm with Adaptive Bounds", Journal of Petroleum Science and Engineering.
- Poli et al., (2012) "A Field Guide to Genetic Programming". Publicado via <http://lulu.com> e disponível em <http://www.gp-field-guide.org.uk>, 2008.
- Rojas et al. (2012) "Reducing Uncertainty in Modelling Fluvial Reservoirs by Using Intelligent Geological Priors", Nono Congresso Internacional de Geoestatística, Oslo, Noruega, 0 Junho 11 – 15, 2012.
- Searson, D. P. (2015) "GPTIPS 2: an open source software platform for symbolic data mining". Capítulo 22 em "Handbook of Genetic Programming Applications", A.H. Gandomi et al, (Eds.), Springer, New York, NY, 2015.
- Soares, A. (2001). Direct sequential simulation and cosimulation. *Mathematical Geology*, 33 (8), 911– 92
- Taniar et al., (2010) "Computation Science and Its Applications" International Conference, Fukuoka, Japão, Março 2010.
- TNO Geo Energy, "Assisted History Matching", 2009. Disponível em <https://www.energydelta.org>.
- Zhangxin, C., (2007) "Reservoir Simulation: Mathematical Techniques in Oil Recovery" .,

Anexo A

Parâmetros da regressão simbólica

Este anexo irá discutir a função de cada parâmetro e quais valores foram adotados no trabalho.

- *Timeout*: define o tempo (em segundos) máximo de execução para cada realização da regressão. Caso o tempo exceda a esse limite, a execução será finalizada. A fim de se obter os melhores resultados, esse parâmetro foi desconsiderado e o tempo de cada simulação foi controlado ajustando-se outros parâmetros.
- *Runs*: número de execuções independentes que são realizadas e então combinadas. Este parâmetro pode ser importante para obter resultados com maior relevância estatística. Para produções mais simples realizou-se apenas 1 execução, mas para os casos difíceis de se obter um bom modelo matemático, várias execuções foram necessárias.
- *Elite Fraction*: fração dos melhores indivíduos que são copiados sem serem alterados para a próxima geração. Inicialmente utilizou-se um valor relativamente alto para este parâmetro (70%) pois este era o valor utilizado por configurações de alguns exemplos disponíveis. No entanto, notou-se que desta forma havia pouca variabilidade entre os modelos finais mesmo de simulações distintas. Para se evitar essa repetição, modificou-se o valor para 5%. Como o algoritmo da regressão busca sempre a redução do *misfit* dos modelos, essa alteração forçou a procura por novos genes em uma área mais ampla no espaço de busca dos parâmetros selecionados e assim evitar que a solução ficasse retida num mínimo local.
- *Pareto Tournament*: Indica a probabilidade de um *Pareto Tournament* ser utilizado para os eventos de seleção. *Pareto Tournaments* são torneios nos quais o *misfit* não é a única característica a ser analisada de forma a se determinar a qualidade de um modelo. É também analisado sua complexidade a fim de se selecionar os melhores que serão mantidos para a próxima geração. Desta forma, mesmo modelos com baixo *misfit*, serão descartados caso possuam uma elevada complexidade. No caso de estudo, como modelos bastante simples obtiveram resultados excelentes, não foi realizado nenhum *Pareto Tournament*.
- “*Number of genes*” e “*Max gene depth*”: estes valores alteram a complexidade do modelo matemático a ser gerado. A *Figure* ilustra dois genes com diferentes profundidades. Entre todos os parâmetros alteráveis, estes são os que mais diminuem o *misfit* ao serem aumentados. No entanto, aumentando-se muito estes valores torna as simulações bastante lentas devido ao aumento considerável no número de cálculos necessários realizados pelo computador, e também ocorre o risco de haver *overfitting* do modelo. Estes foram os principais parâmetros alterados a cada simulação conforme o desejado

por cada teste. Observe pela Figure A.1 que uma pequena alteração neste valor pode criar modelos muito mais complexos, mas que na maioria das vezes melhor se adequam à base de dados

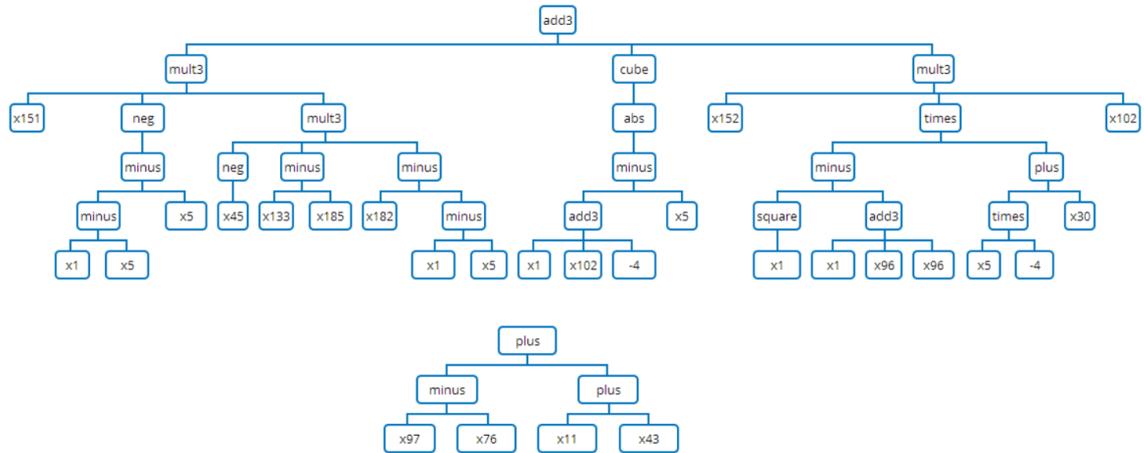


Figure A.1 – Comparação entre dois genes com diferentes profundidades (6 e 3).

- Population size: É o tamanho da população de modelos testes (aleatoriamente gerados de forma a maximizar sua diversidade) que são treinados pela base de dados até atingirem o menor *misfit* possível ao final de todas as gerações. Um valor grande para esse parâmetro é necessário para que se expanda ao máximo possível a busca por um modelo adequado. Para o Watt-field utilizou-se entre 800 e 4000.
- Number of generations: Número de gerações por cada realização da regressão.. Em cada geração, os modelos são selecionados probabilisticamente com base apenas em seu *misfit* ou então em seu *misfit* e sua complexidade (no caso de pareto tournaments) e, então, servem para gerar novos modelos por crossovers e mutações. Em todos modelos gerados utilizou-se de 30 a 40 gerações pois, acima disso, o *misfit* da maioria dos testes não se alterava.
- Functions: Este parâmetro define quais funções serão utilizadas para se gerar e treinar os modelos. As funções disponíveis são: *plus*, *minus*, *multiply*, *divide*, *add3* (adição de 3 números), *mult3* (multiplicação de 3 números), *tanh*, *cos*, *sin*, *exp*, \log_{10} , *square*, *power*, *abs* (módulo), *cube*, *sqrt*, *negative exp*, *if-then-else*, *-x*, *greater than* ($>$), *less than* ($<$), Gaussian ($\exp(x^2)$). Inicialmente utilizou-se todas as funções com exceção a “*if-then-else*”, “*greater than* ($>$)” e “*less than* ($<$)”, no entanto, desta forma notou-se frequentemente um problema nos resultados de validação dos modelos. Muitas vezes em que o modelo de equações era aplicado nos dados utilizados para sua validação, os resultados apresentaram valores completamente errados (geralmente com ordens de magnitude muito maiores que o esperado) ou até inexistentes. A Figure A.2 mostra um exemplo deste problema. Após investigação, descobriu-se que o motivo estava relacionado com algumas restrições de certas funções. Funções como \log_{10} ou *sqrt*, por exemplo, não permitem valores negativos

e, por isso, erros como esses são possíveis nos resultados de validação (uma vez que os valores utilizados são diferentes da base de dados pro treinamento). No exemplo da Figure A.2, os elevados valores eram consequência de uma divisão por um número inesperadamente pequeno, o que provocava um grave erro nos resultados. Em casos nos quais a conta era simplesmente impossível (e.g. raiz quadrada de um número negativo), uma mensagem informava do erro e os resultados da validação não eram exibidos. Desta forma, as funções utilizadas para a geração dos modelos foram: “times”, “minus”, “plus”, “square”, “mult3”, “add3”, “cube”, “neg (-x)”, “abs” e funções do tipo “if-then-else”.

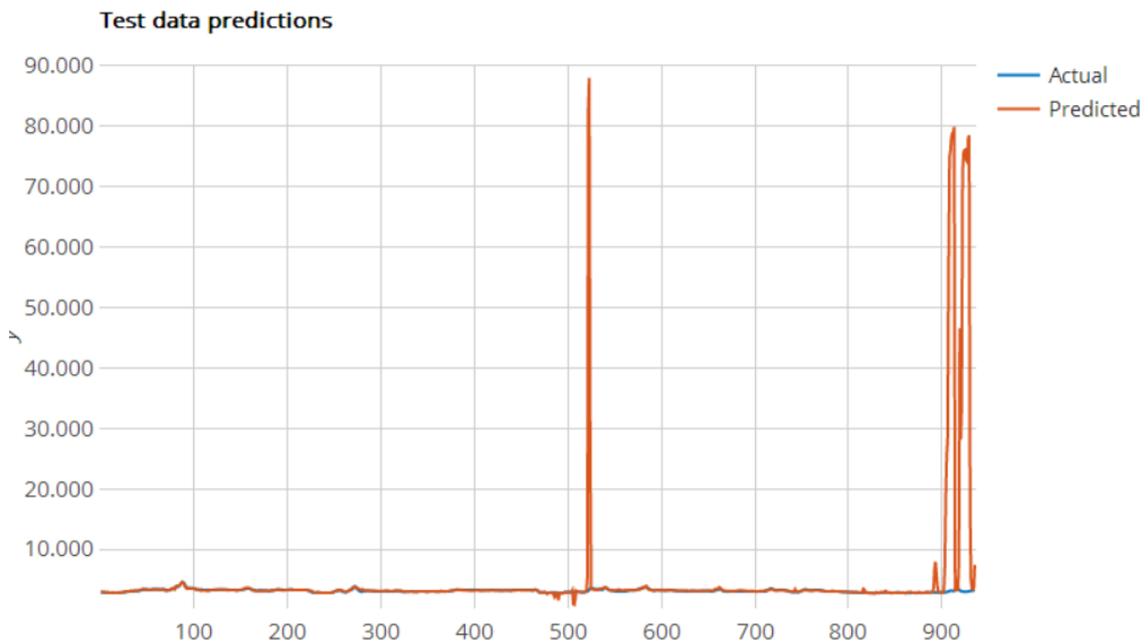


Figure A.2 – Problema causado por determinadas funções.

Para as funções existentes no GPTIPS2, apenas a “gpmodelexport” foi utilizada. Esta função gera um relatório detalhado da regressão desejada e foi apenas utilizada para se verificar os resultados dos modelos; na integração da regressão simbólica no ciclo de ajuste de histórico, esta função foi dispensada. As outras funções do GPTIPS2 também exigem algum processo manual, e portanto não foram utilizadas especialmente pois o objetivo deste trabalho é a substituição dos simuladores de fluidos por um método automático

Anexo B

Preparação da planilha

A preparação dos dados se inicia com o *upscaling* dos modelos petrofísicos. Este procedimento calcula a média aritmética dos valores de porosidade e permeabilidade das células próximas (número de células em cada direção é definida pelo usuário), e atribui o resultado a uma única célula de um vetor das propriedades petrofísicas. Após o *upscaling*, alguns passos são executados para que se obtenha a matriz (planilha) necessária para os treinamentos. Um exemplo desta matriz pode ser observado pela Figura 21.

		K1	...	Kn	ϕ_1	...	ϕ_n
1º modelo	T1	200		700	0,1		0,08
	...	200		700	0,1		0,08
	Tt	200		700	0,1		0,08
...							
Nº modelo	T1	150		800	0,15		0,07
	...	150		800	0,15		0,07
	Tt	150		800	0,15		0,07

Figura 21 – Formato do arquivo xls escrito pelo programa. N indica o número total de modelos, n indica o número de células após *upscaling* e t indica o número de *timesteps*.

1. Inicialmente, o usuário define o número de *timesteps* (T) que os dados de produção apresentam, o tamanho do modelo pelo número de células em X, Y e Z, assim como o tamanho desejado nessas mesmas direções após o upscale (note que esses valores têm de ser um múltiplo do outro);
2. O programa, então, lê o arquivo .txt que contém os valores da porosidade ou permeabilidade de um determinado modelo, e em seguida os valores carregados do arquivo são reorganizados em um arranjo tridimensional no qual cada dimensão representa as direções X, Y e Z;
3. Utilizando-se esse arranjo, o programa calcula as médias entre as células que serão combinadas e guarda-as todas em um único vetor
4. O vetor criado na etapa 3 é repetido T vezes, uma embaixo da outra, para o número de *timesteps* determinado pelo usuário.
5. Finalmente, o programa reinicia o processo com o próximo modelo, mas mantendo a mesma matriz.

Como se observa pela Figura 21, a matriz final terá um número total de linhas igual ao número de *timesteps* multiplicado pela quantidade de modelos a serem utilizados, enquanto o

número total de colunas é igual à quantidade de células existentes após o *upscaling* mais os *timesteps*.

Um problema encontrado foi o formato dos arquivos de porosidade e permeabilidade que se encontravam com a extensão .out e, portanto, precisam sofrer uma conversão antes de serem lidos pelo programa em Matlab. Para se evitar uma conversão manual de todas as centenas de modelos, o seguinte código pode ser aplicado para computadores com sistema operacional Windows:

1. No bloco de notas, escreve-se o código “copy ‘path’*.out ‘path’*.txt”, onde ‘path’ representa a pasta onde se localizam os arquivos a serem convertidos;
2. Em seguida, salva-se o arquivo com a extensão .bat, incluindo a extensão no nome do arquivo.
3. Por fim, executa-se o arquivo. O *prompt* de comando do windows irá abrir e fará uma cópia dos arquivos .out, salvando-os em formato .txt

Com o problema das extensões e do *upscaling* resolvido, a última etapa é a inclusão dos dados de produção na tabela que será utilizada para se treinar o programa. Essa inclusão é bastante simples e pode ser feita com um código simples de matlab que lê todas os arquivos de produção e os compila em uma única tabela.

Unindo-se os dados das duas tabelas (dados de produção e porosidades/permeabilidades após *upscaling*), finalmente obtém-se o arquivo utilizado para o treinamento e a geração dos modelos pelo GPTIPS2, que pode ser observado pela Figura 22. Desta forma, tendo-se os parâmetros configurados, é possível dar início às simulações.

	WOPR (STB Tempo (dias))	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}	
Model 1	8497,332	120	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	8498,728	273	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	8499,231	425	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	8499,462	578	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	8499,58	731	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	8496,811	851	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	8494,432	1004	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	8493,474	1096	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	8492,955	1155	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	8491,786	1308	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	8490,784	1461	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	8488,399	1581	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	8439,35	1734	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	7899,557	1886	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	7519,34	2039	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	7232,626	2192	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	7139,028	2312	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
	6955,312	2465	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634
6771,347	2616	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634	
6446,333	2769	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634	
6013,887	2922	0,187511	0,11852	0,095667	0,134963	0,106308	0,154539	0,085179	0,104021	0,090267	0,092634	
Model 2	8497,294	120	0,191178	0,117223	0,096476	0,135192	0,11135	0,160386	0,087152	0,105857	0,089366	0,093166
	8498,583	273	0,191178	0,117223	0,096476	0,135192	0,11135	0,160386	0,087152	0,105857	0,089366	0,093166
	8499,132	425	0,191178	0,117223	0,096476	0,135192	0,11135	0,160386	0,087152	0,105857	0,089366	0,093166
	8499,403	578	0,191178	0,117223	0,096476	0,135192	0,11135	0,160386	0,087152	0,105857	0,089366	0,093166
	8499,546	731	0,191178	0,117223	0,096476	0,135192	0,11135	0,160386	0,087152	0,105857	0,089366	0,093166
	8496,997	851	0,191178	0,117223	0,096476	0,135192	0,11135	0,160386	0,087152	0,105857	0,089366	0,093166
	8494,41	1004	0,191178	0,117223	0,096476	0,135192	0,11135	0,160386	0,087152	0,105857	0,089366	0,093166
	8493,297	1096	0,191178	0,117223	0,096476	0,135192	0,11135	0,160386	0,087152	0,105857	0,089366	0,093166
	8492,69	1155	0,191178	0,117223	0,096476	0,135192	0,11135	0,160386	0,087152	0,105857	0,089366	0,093166

Figura 22 – Exemplo real de parte de uma planilha utilizada para o treinamento do programa.

Anexo C

Ciclo sem regressão simbólica: WBHP e WOPR do poço 3

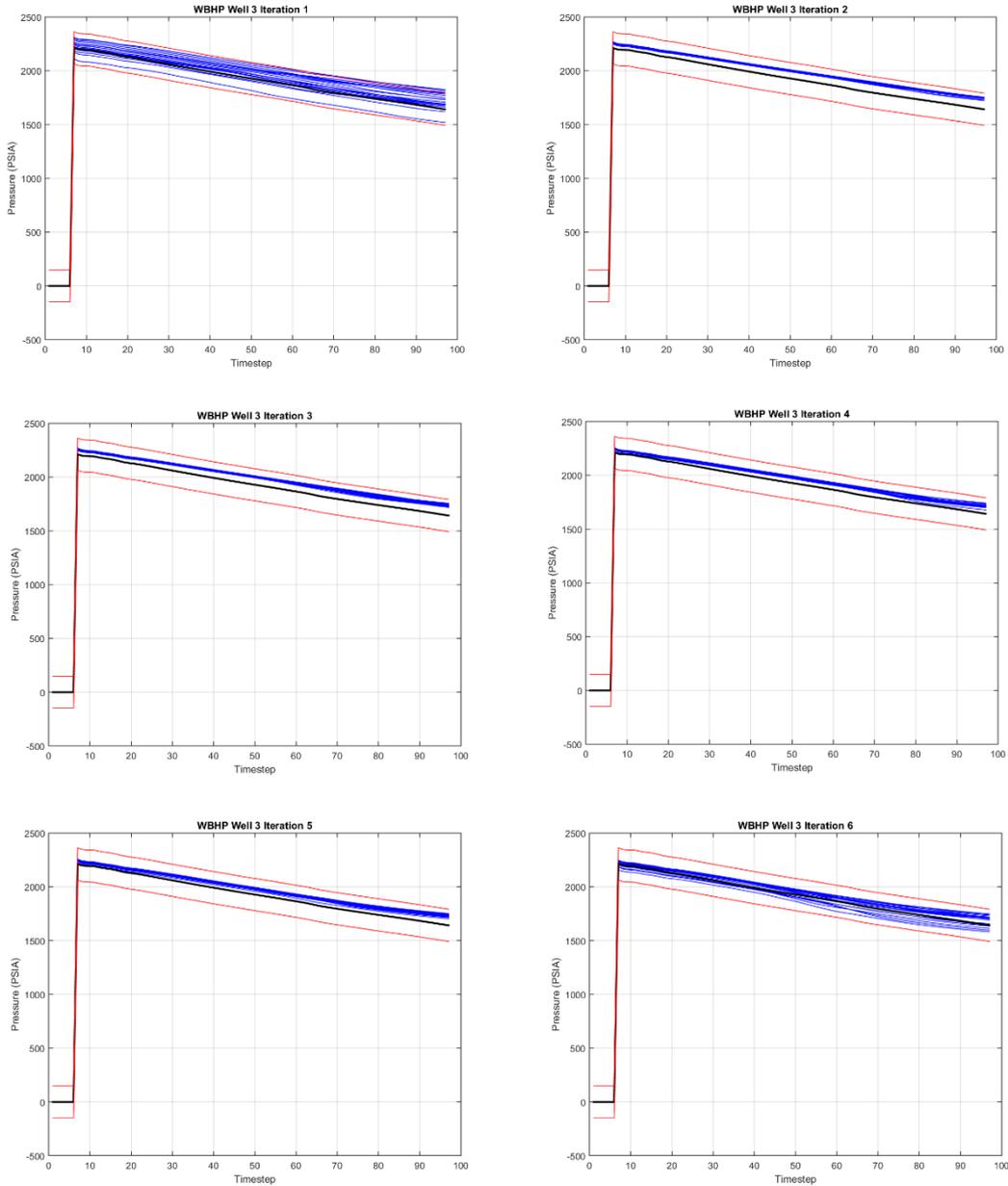


Figura 23 – WBHP poço 3: Ciclo sem regressão simbólica (iterações 1-6)

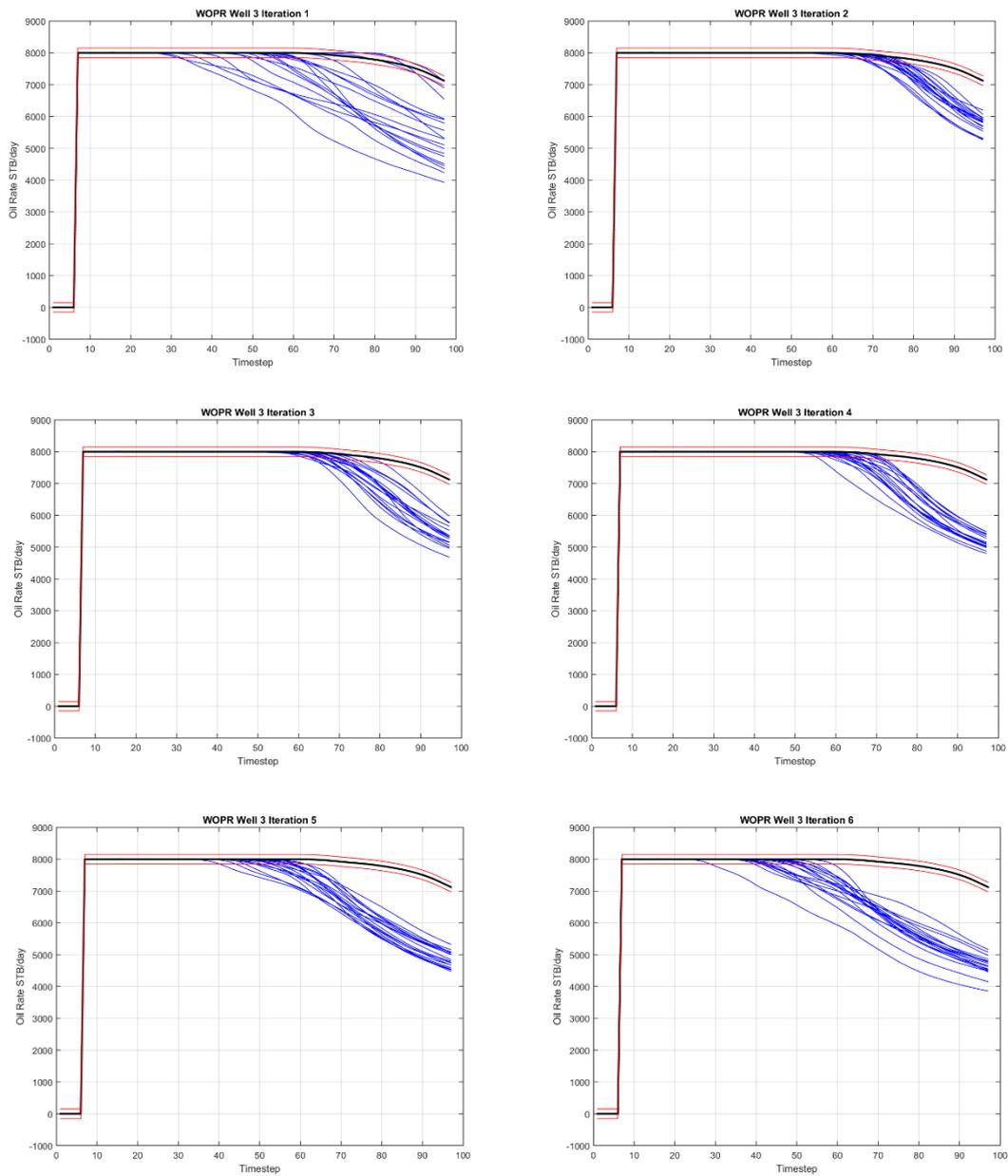


Figura 24 – WWPR poço 3: Ciclo sem regressão simbólica (iterações 1-6)

Anexo D

Ciclo com regressão simbólica: WWPR poço 7

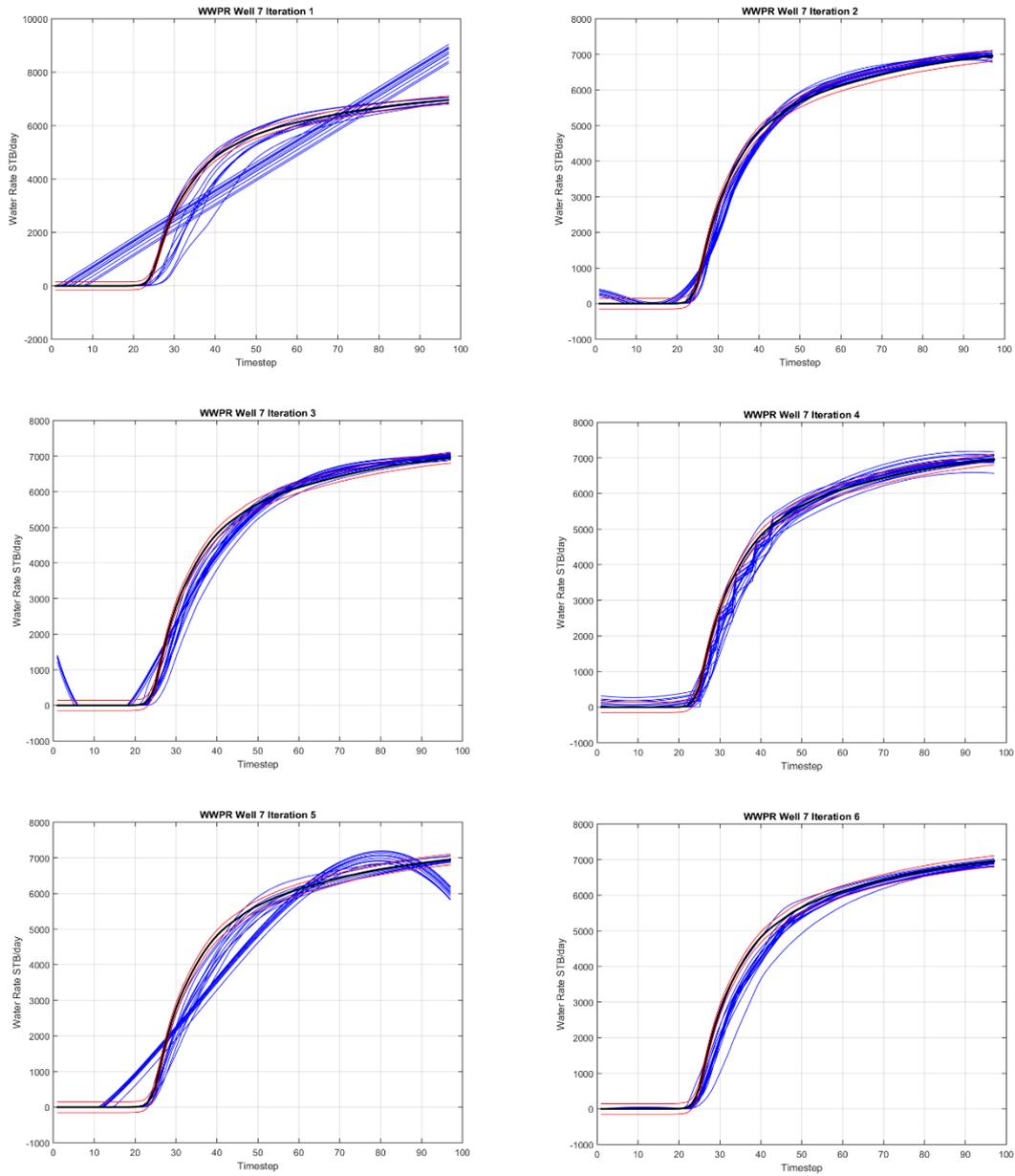


Figura 25 – WWPR Well 7: Exemplo de maus ajustes dos modelos de regressão simbólica (6 iterações)

Anexo E

Comparação dos métodos de ajuste de histórico

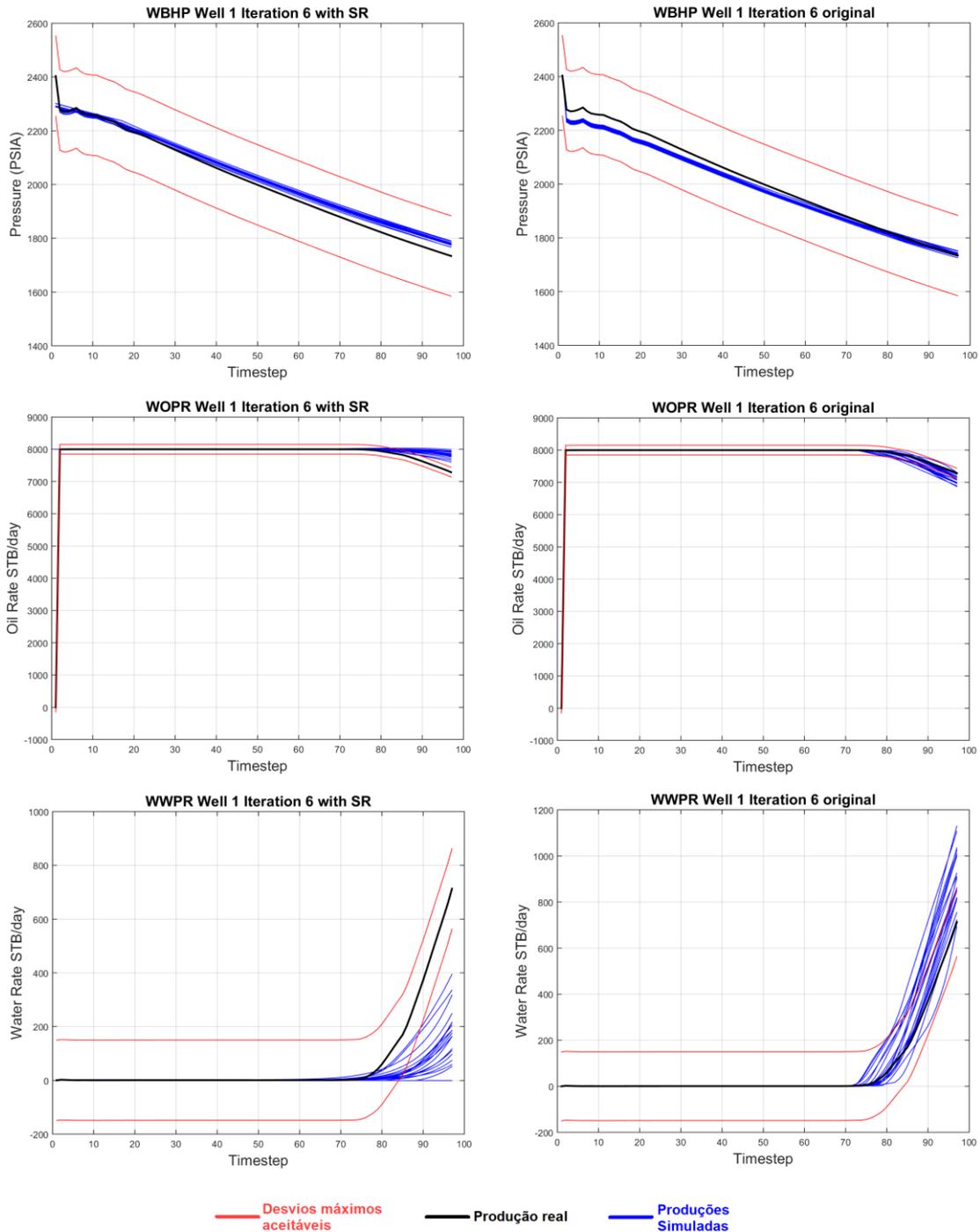


Figura 26E.1 – Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 1.

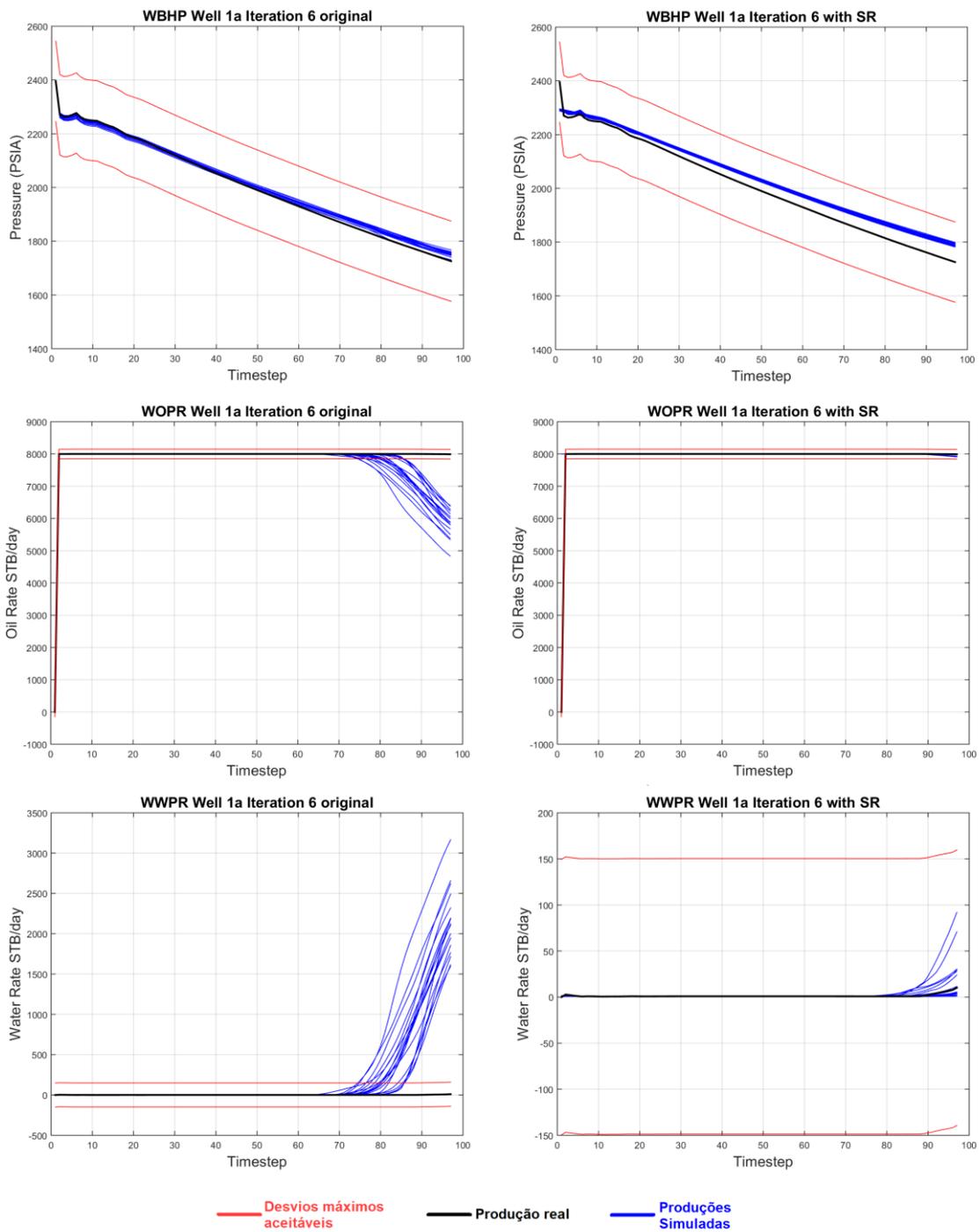


Figura E.2 – Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 1ª.

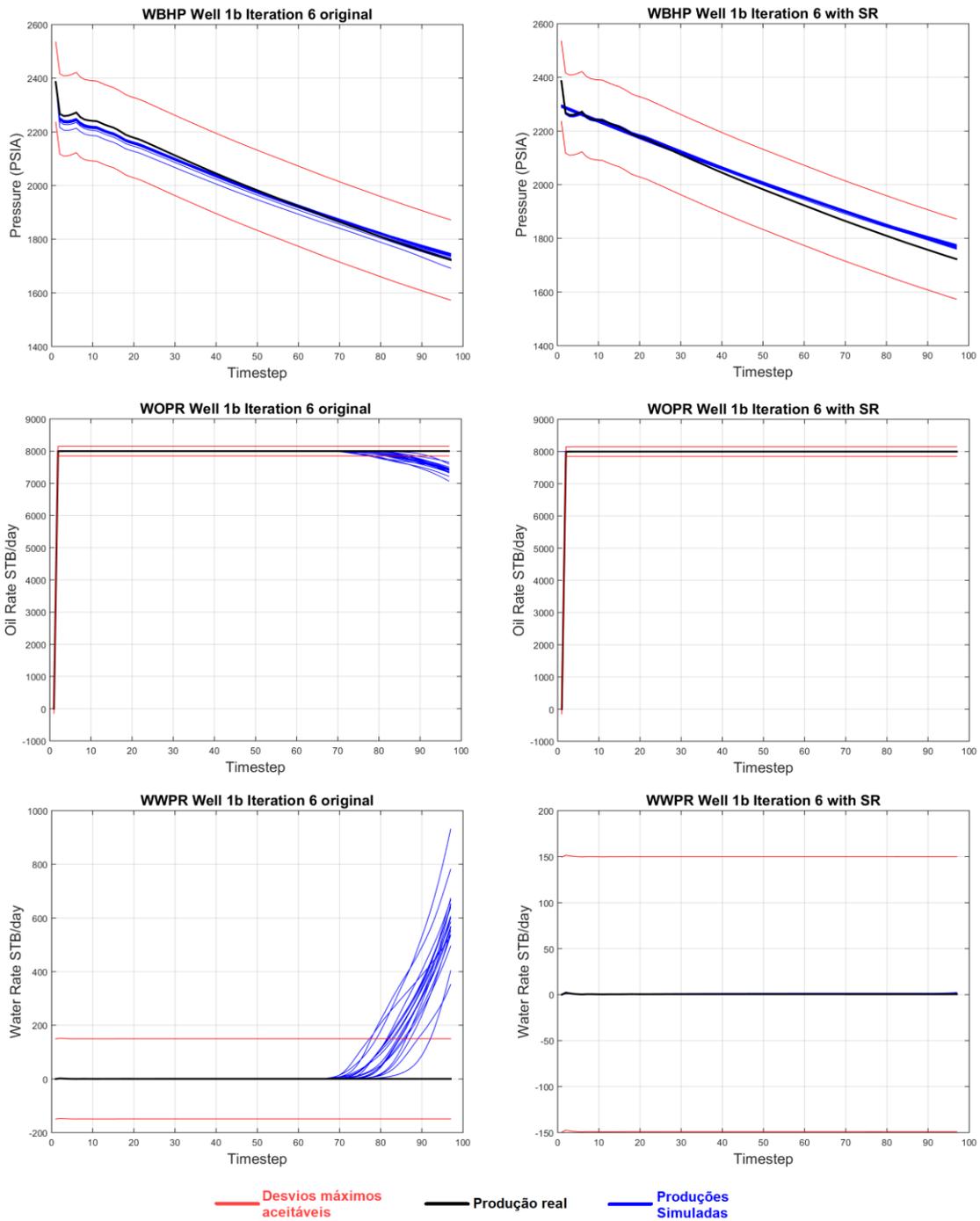


Figura E.3 – Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 1b.

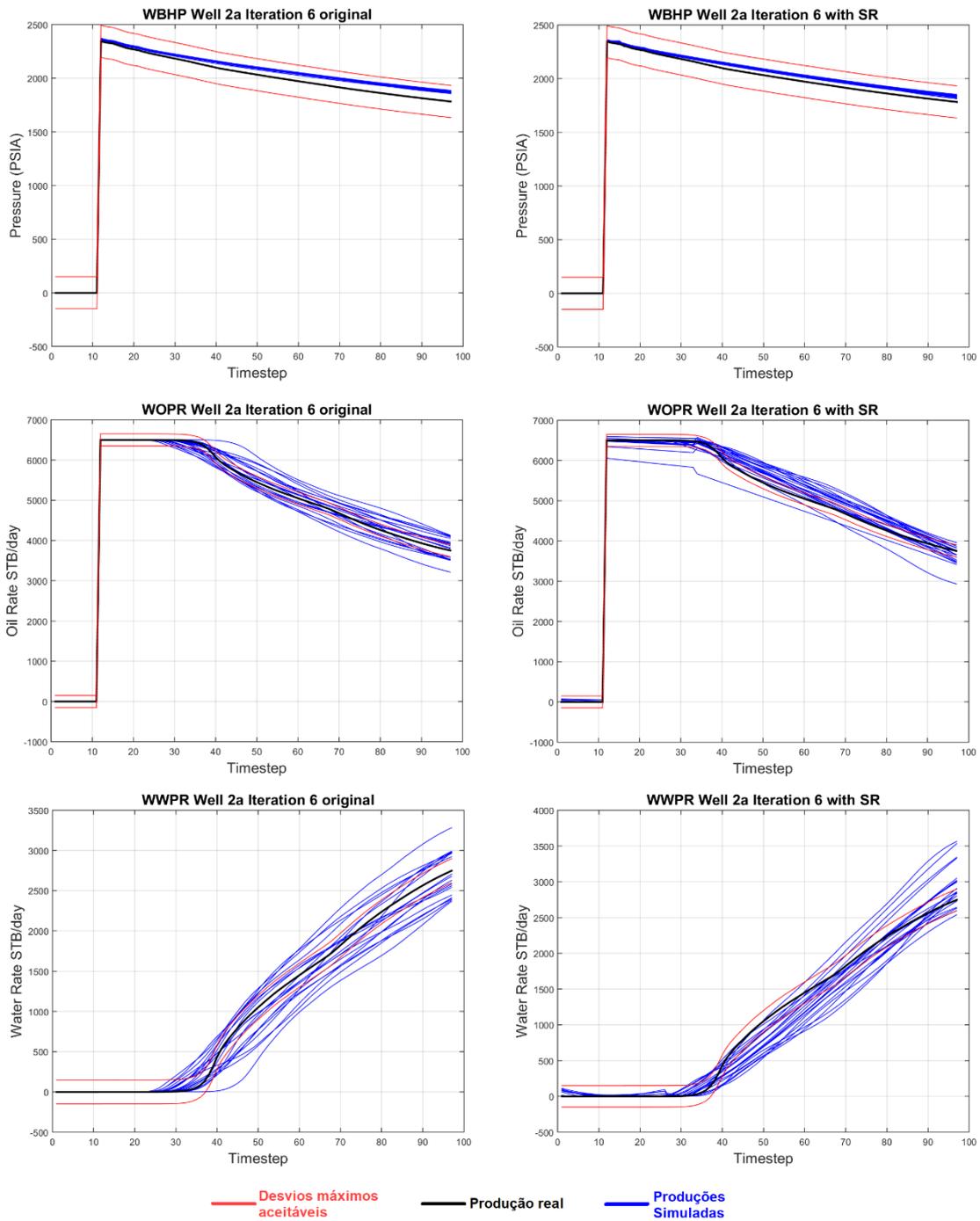


Figura E.4 – Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 2a.

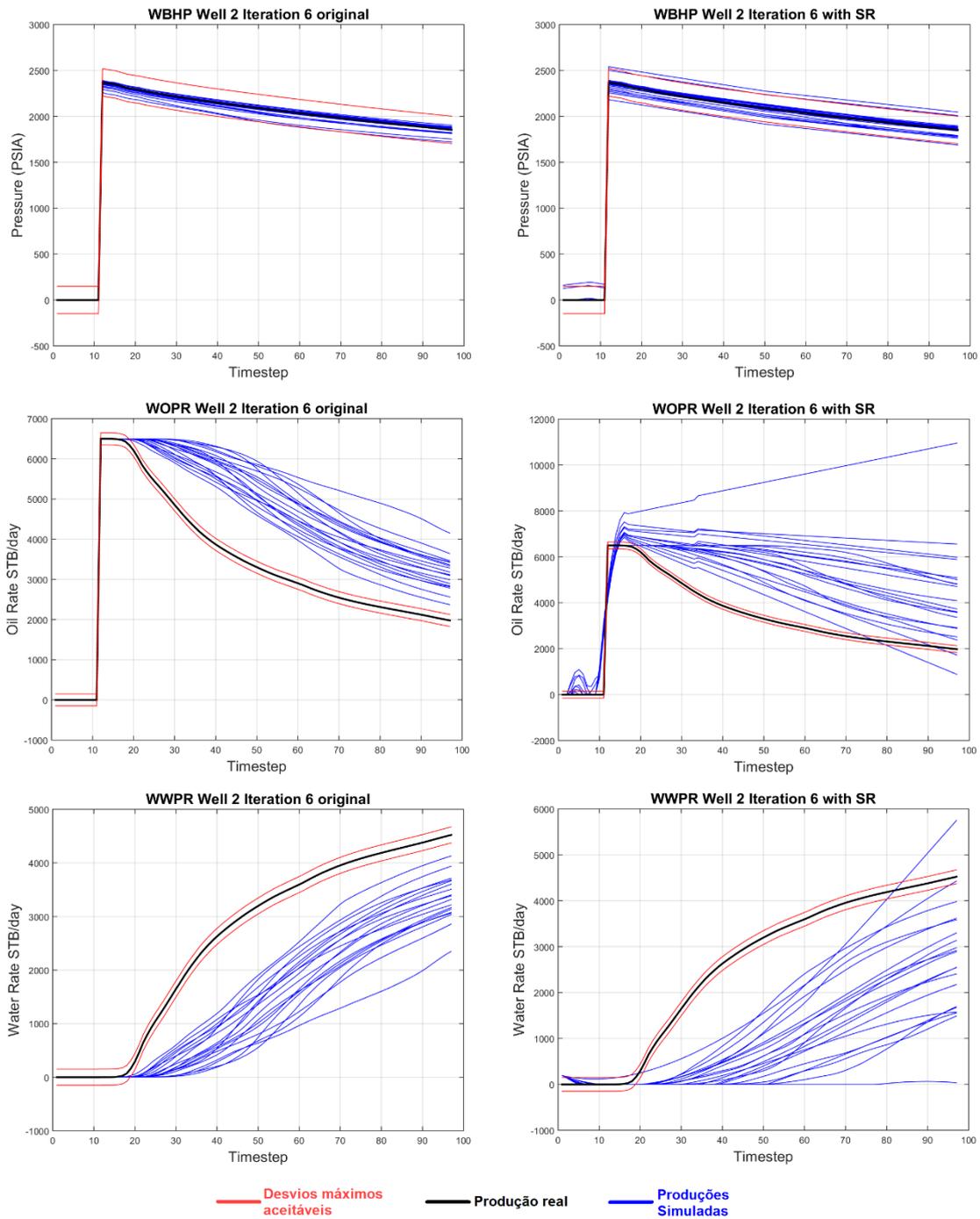


Figura 27E.5 – Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 2.

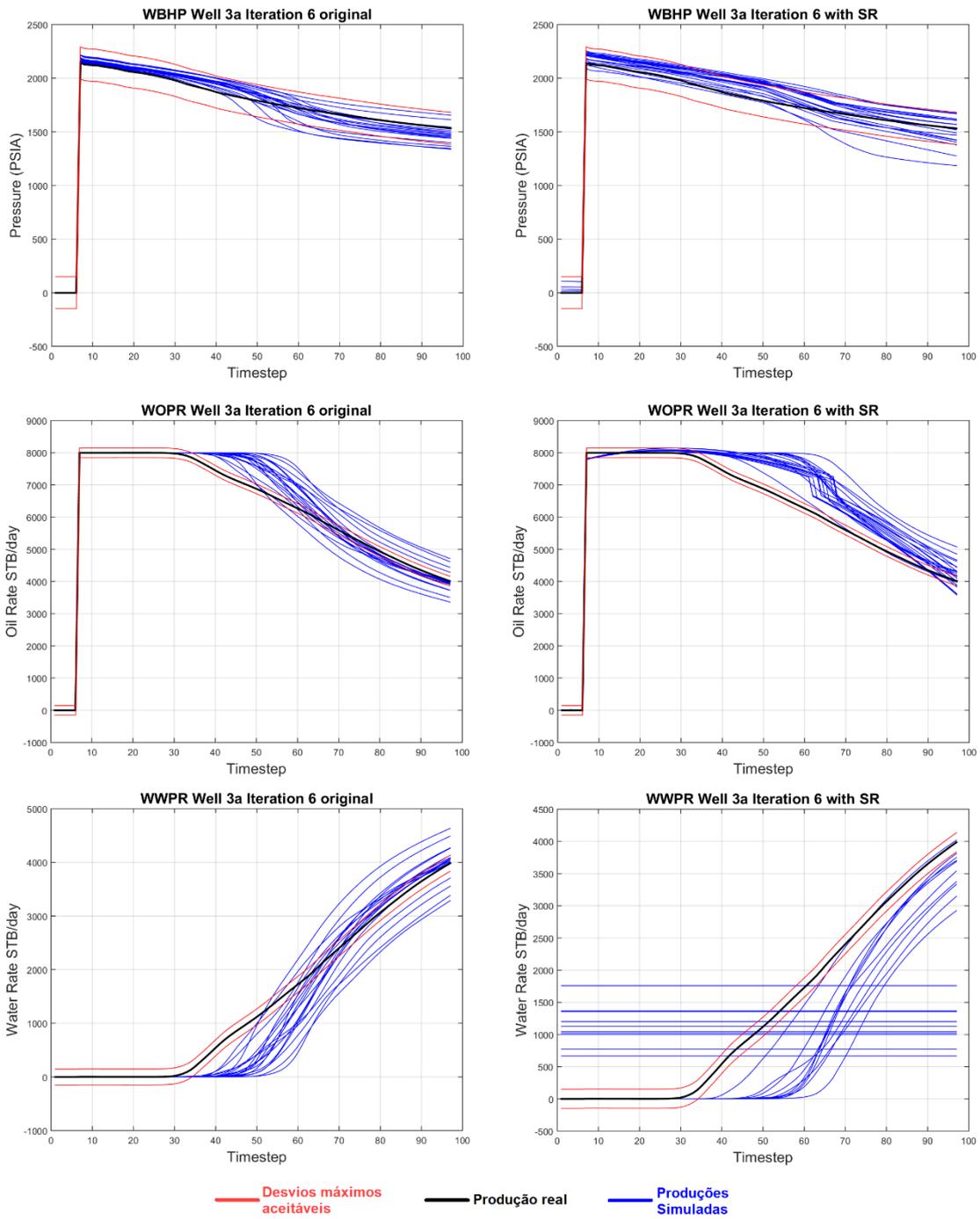


Figura E.6 – Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 3a.

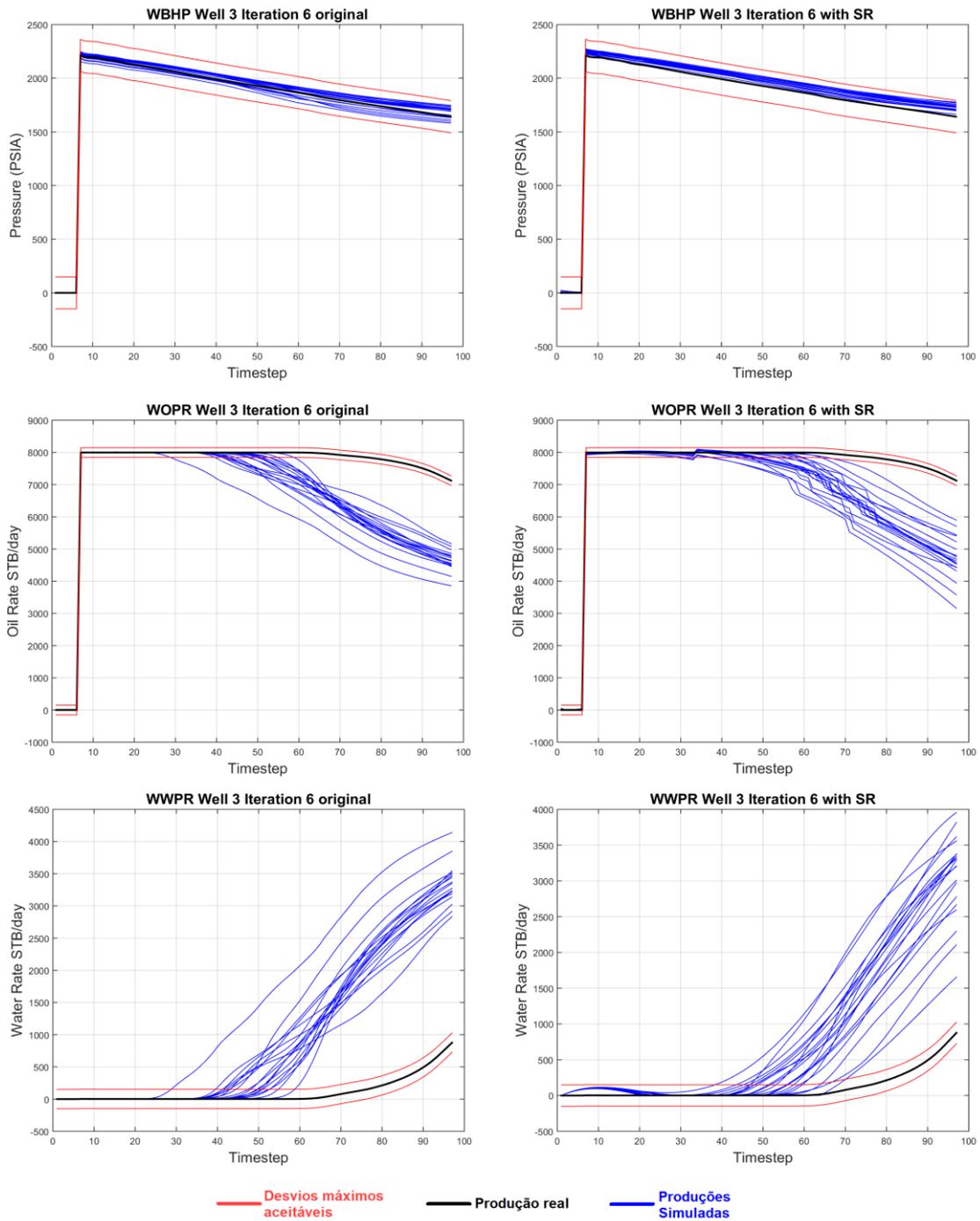


Figura 28E.7 – Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 3.

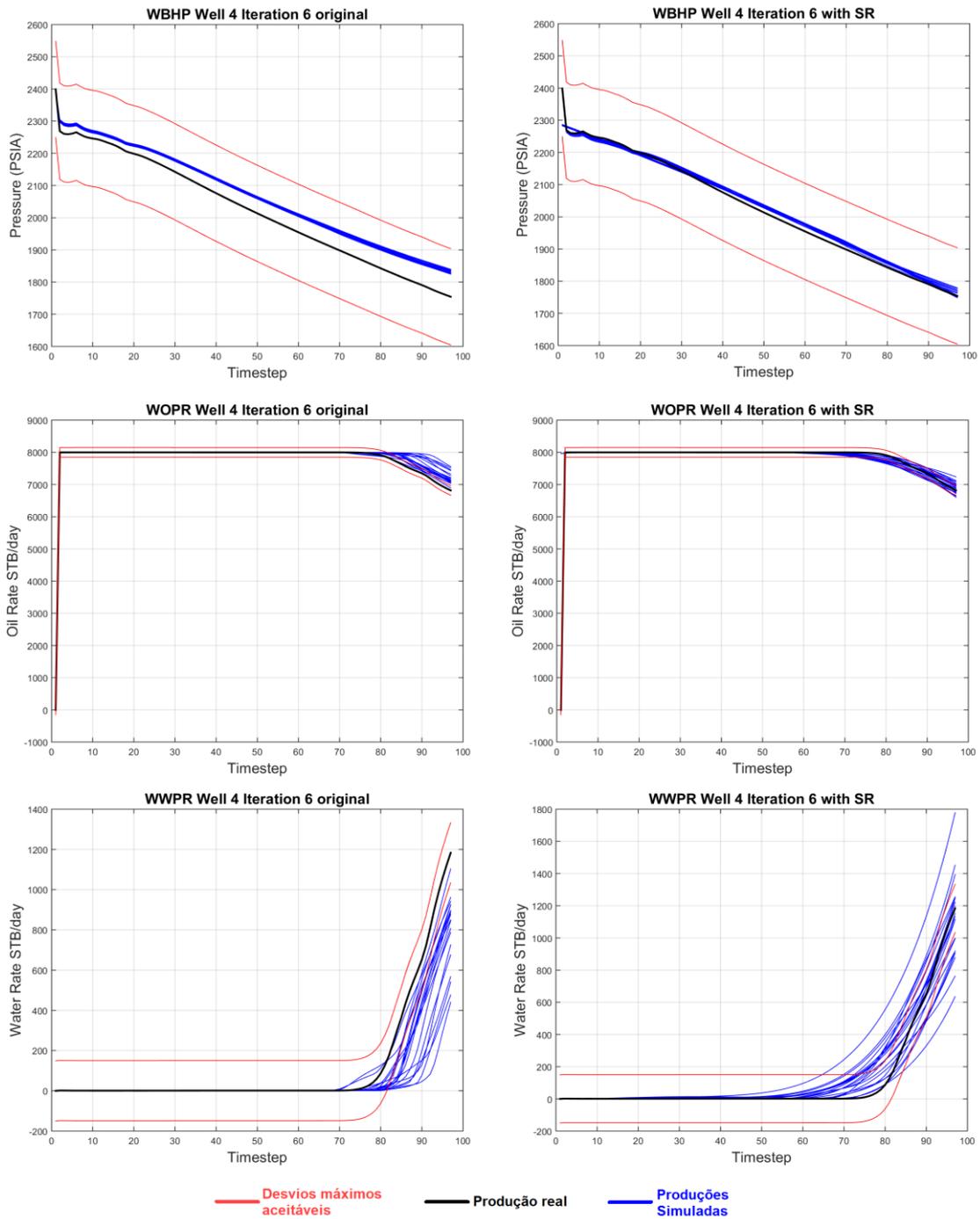


Figura E.8– Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 4.

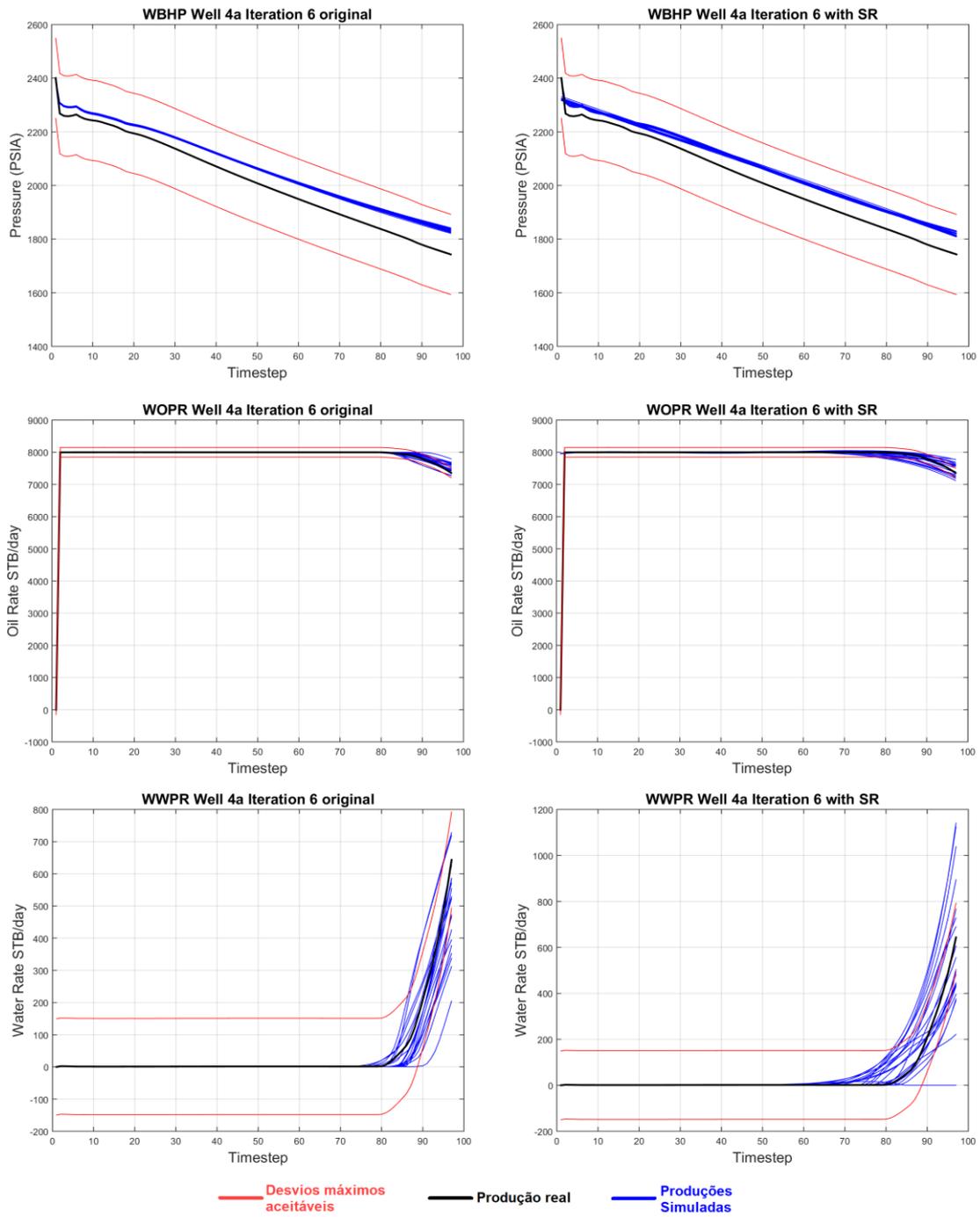


Figura E.9 – Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 4a.

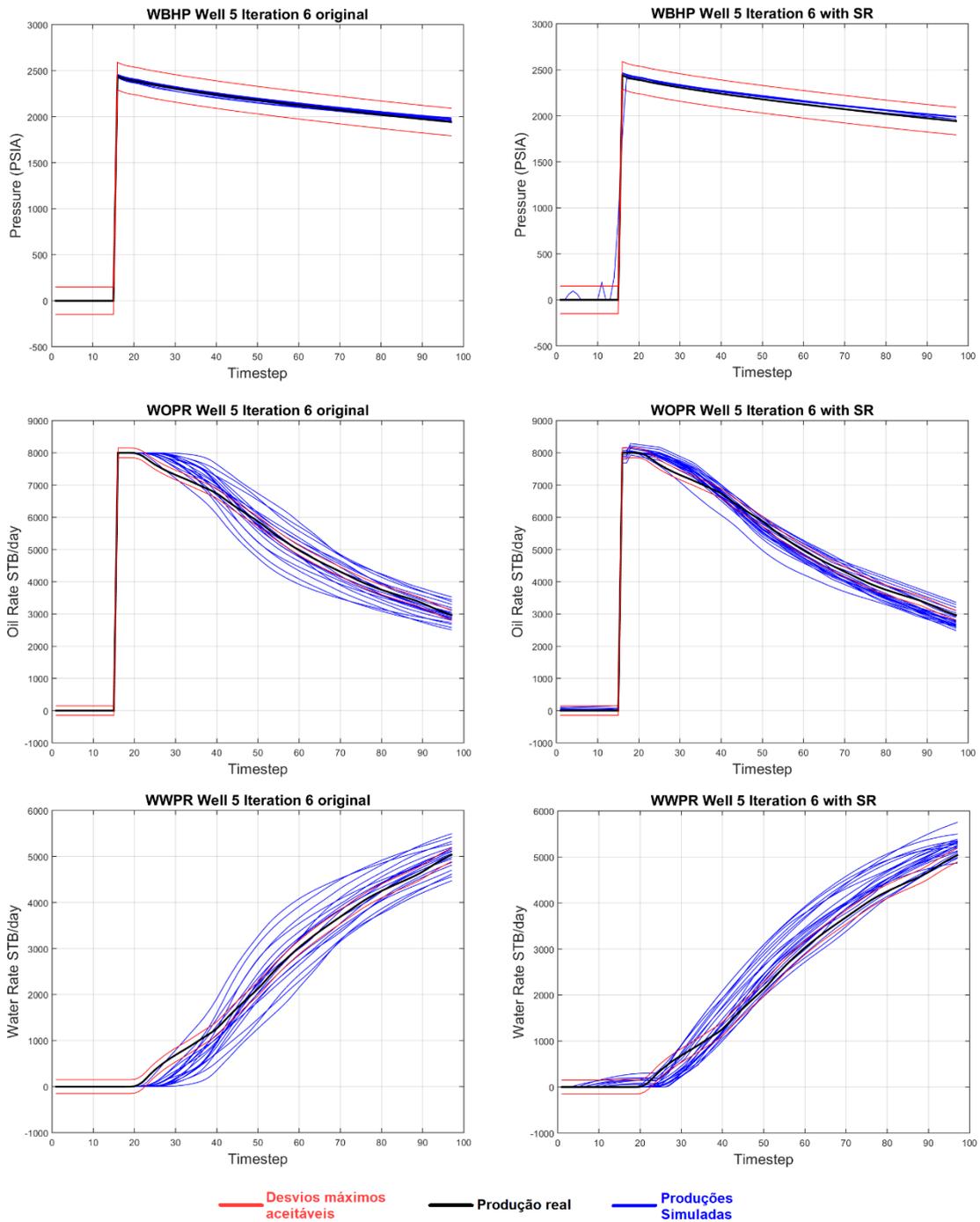


Figura E.10 – Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 5.

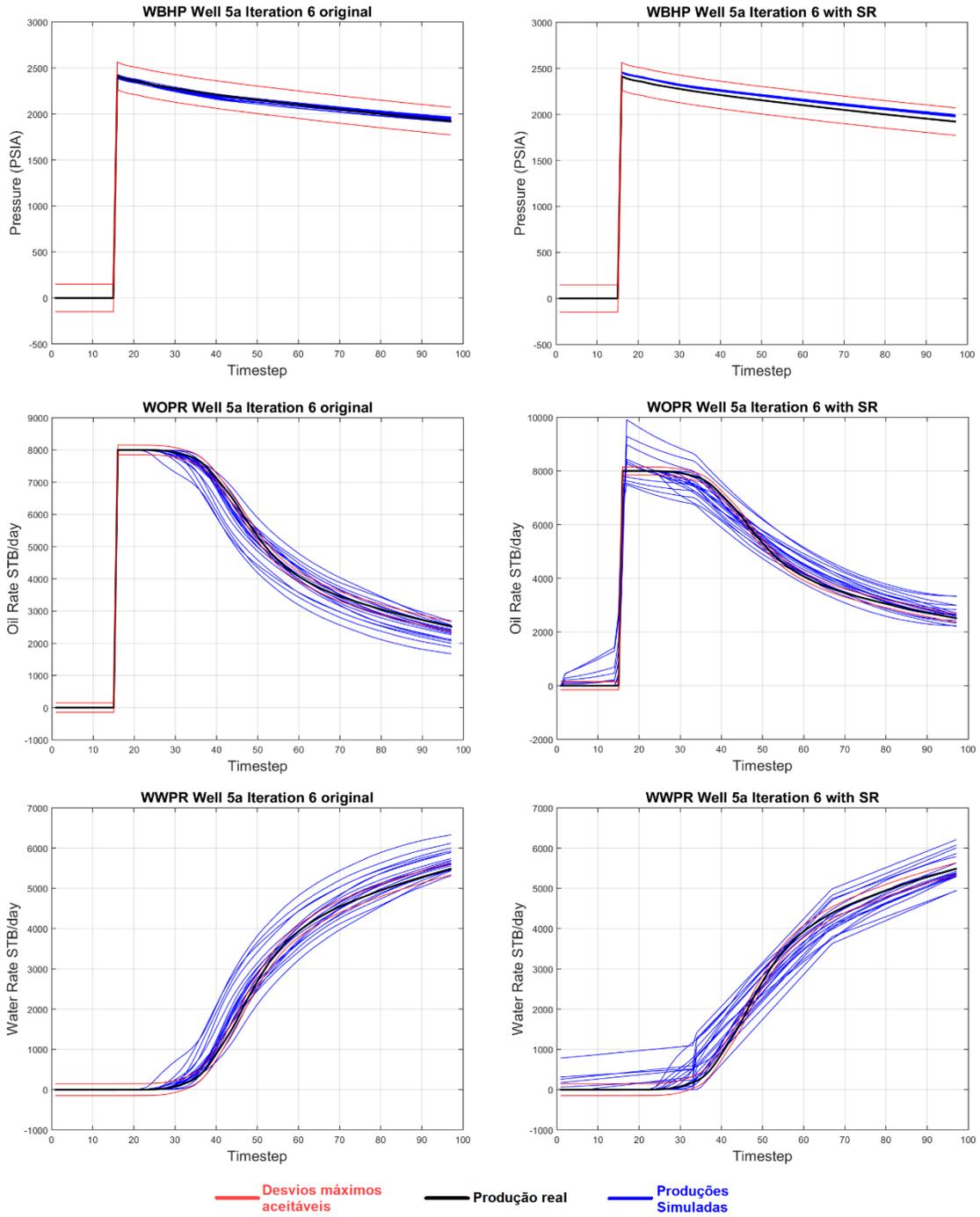


Figura E.11 – Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 5ª.

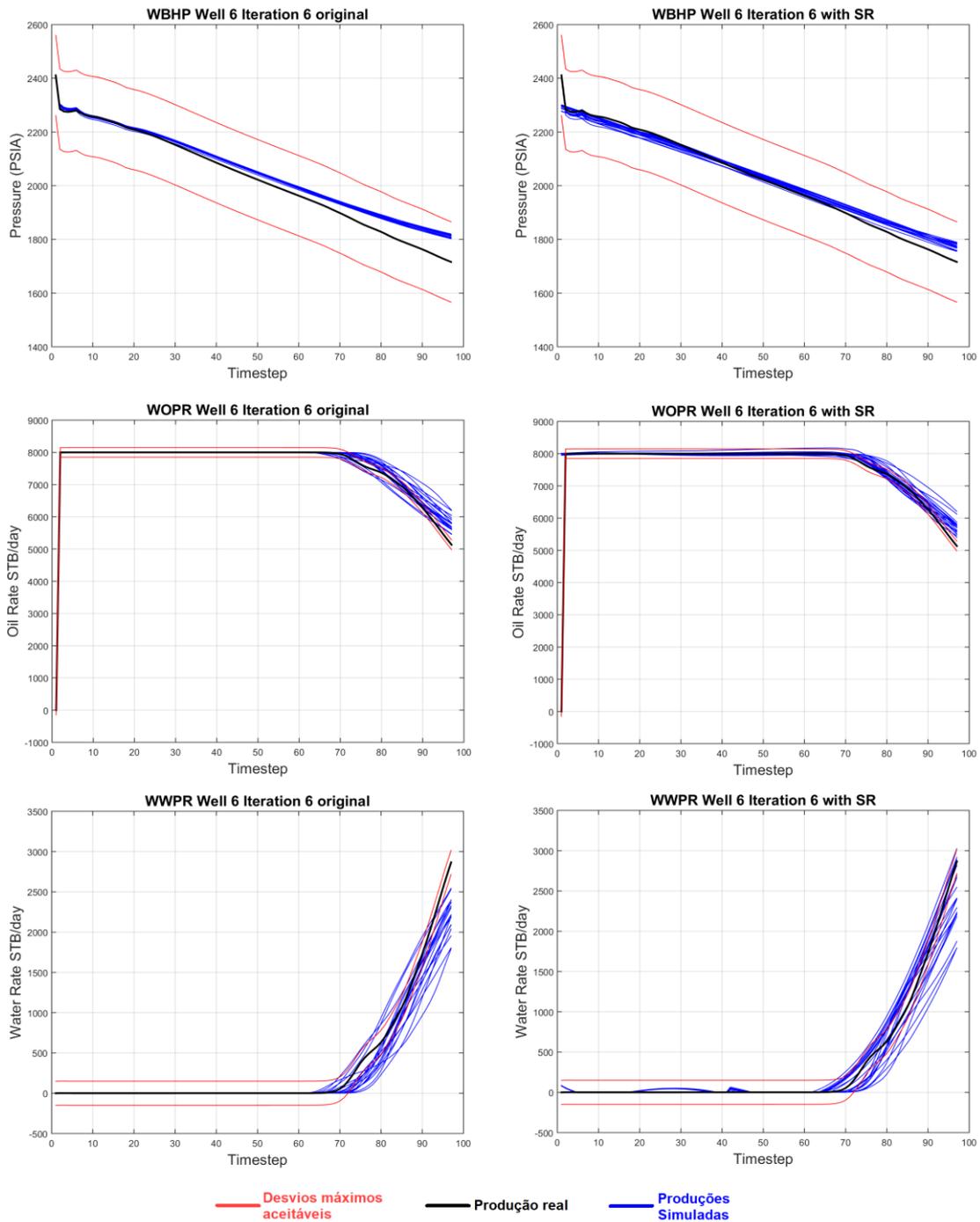


Figura 29E.12 – Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 6.

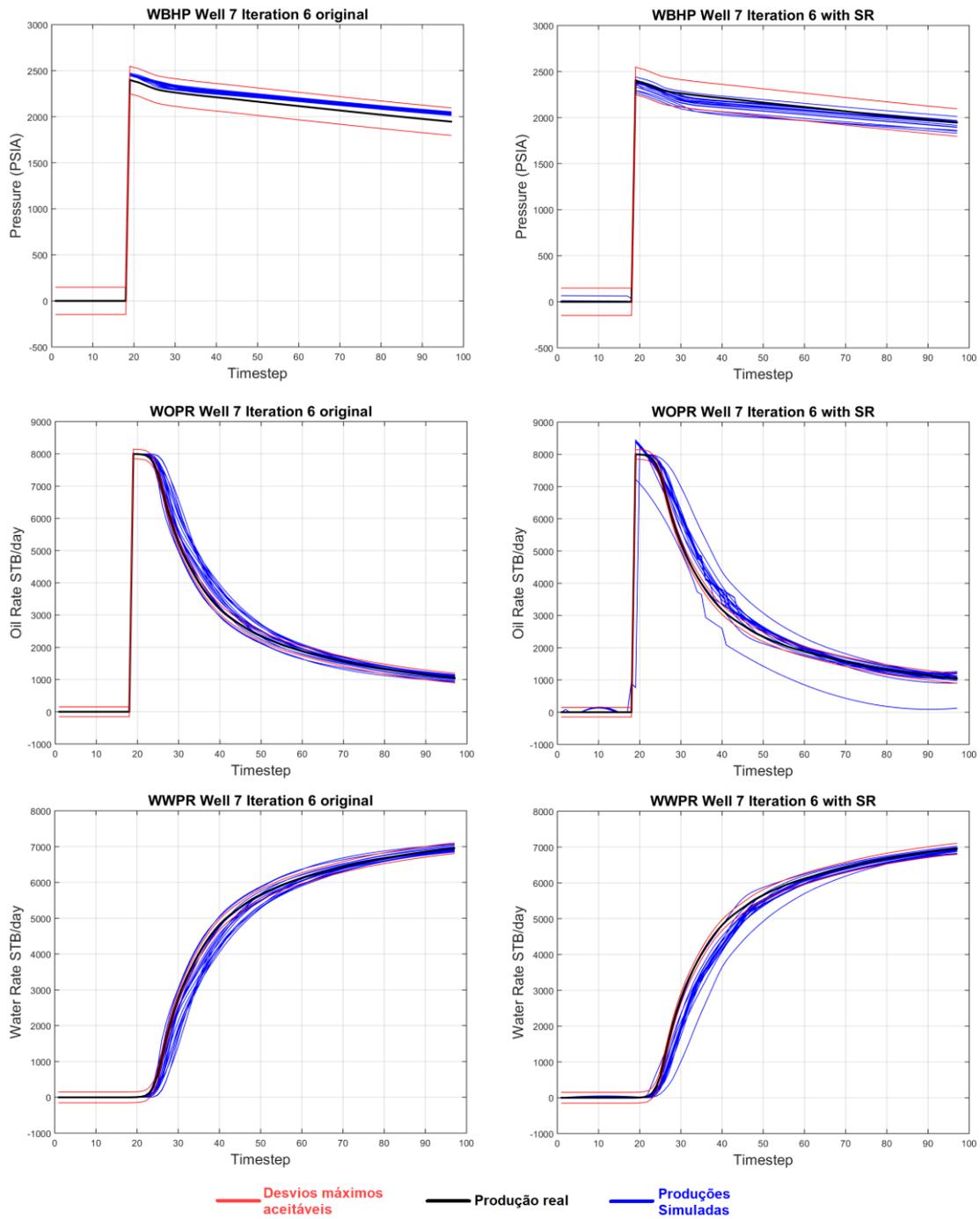


Figura E.13 – Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 7.

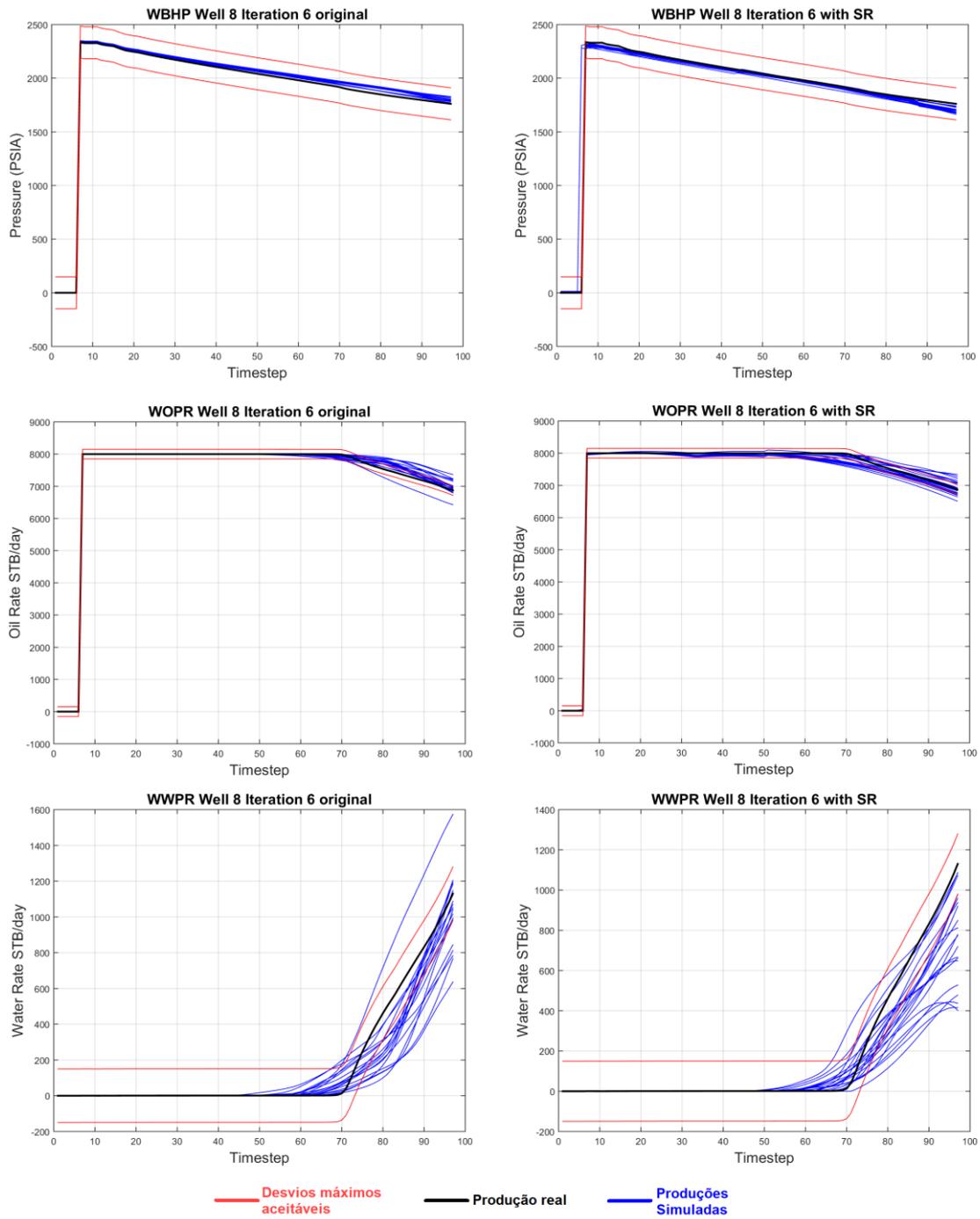


Figura E.14 – Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 8.

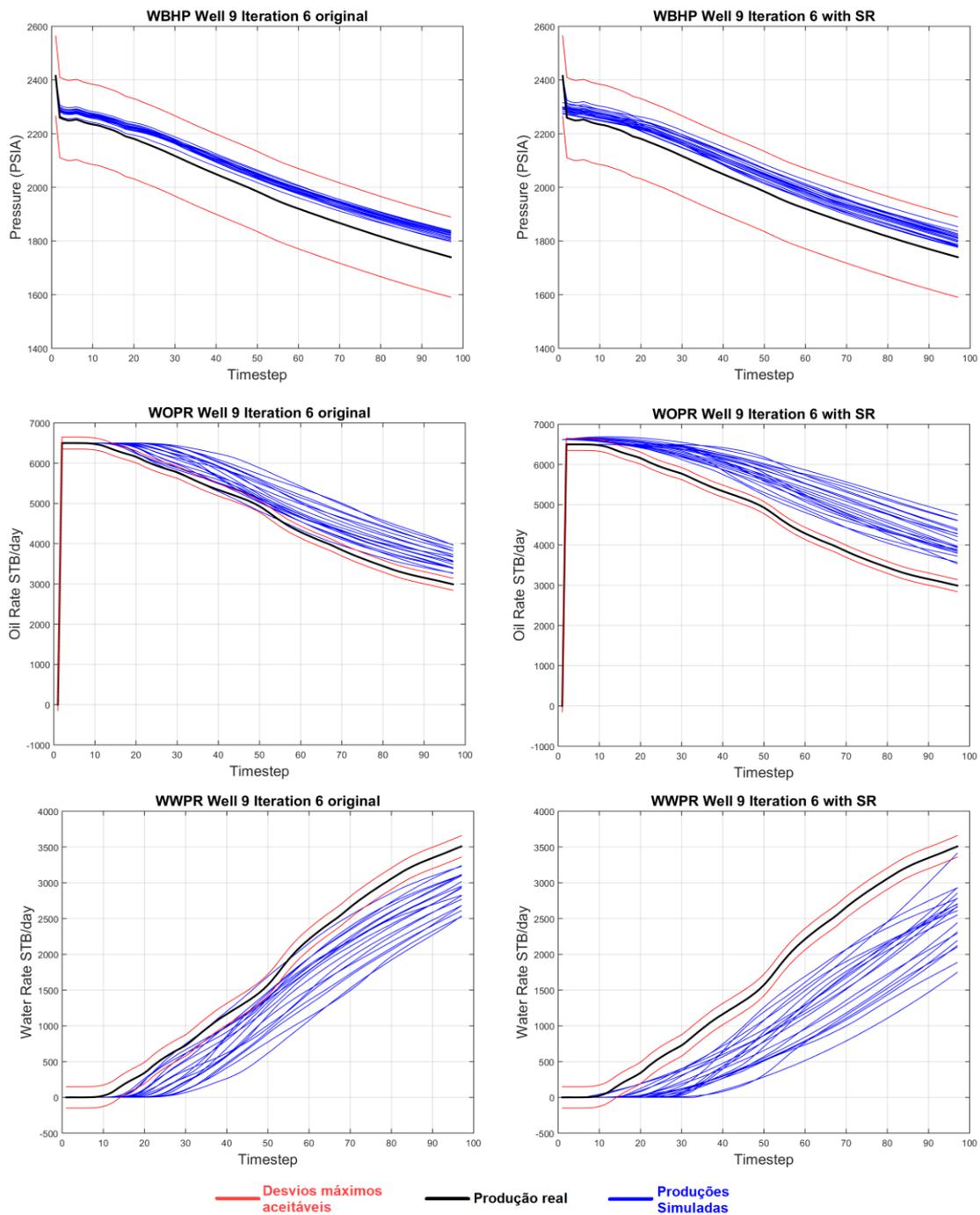


Figura E.15 – Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 9.

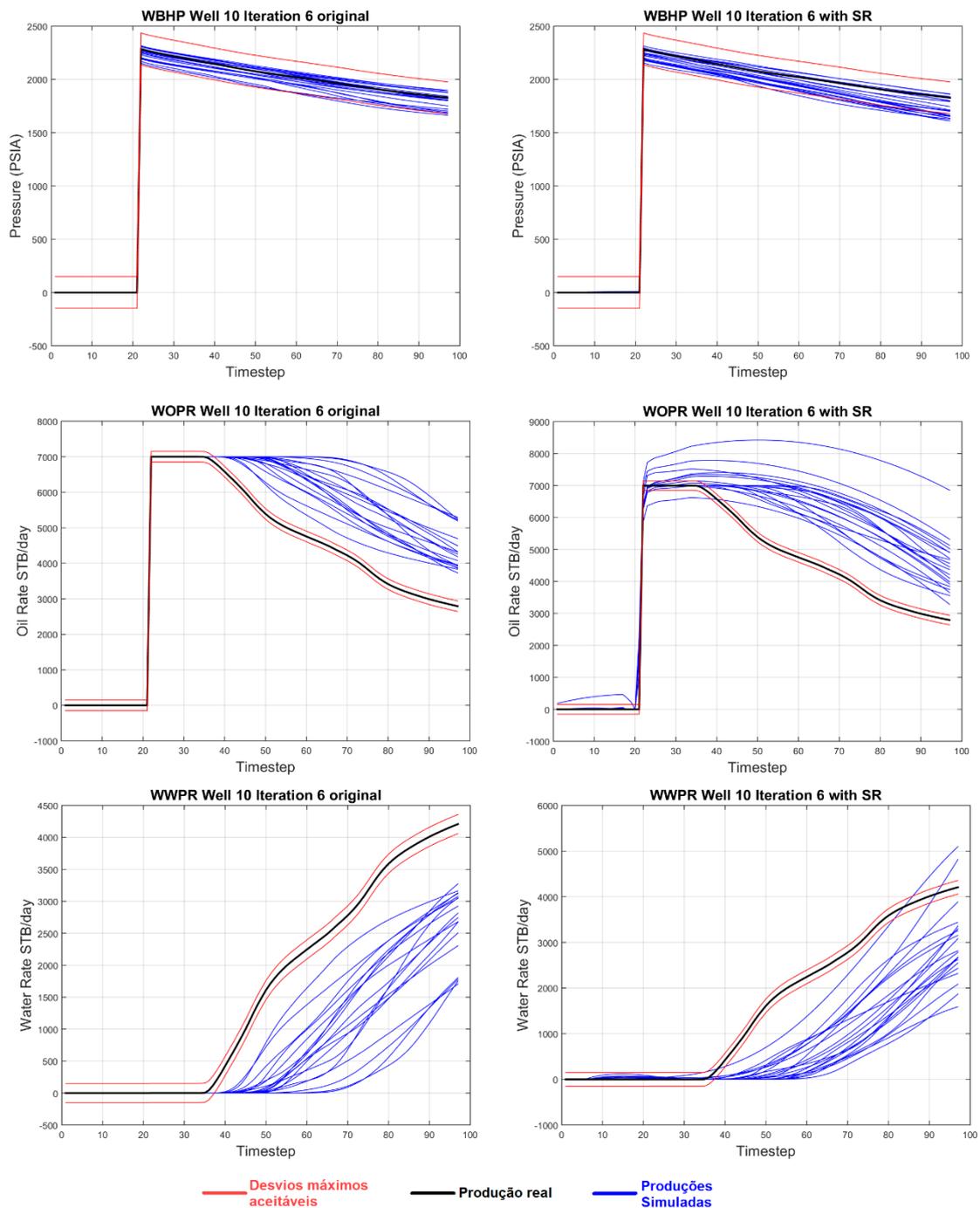


Figura E.16 – Comparação das convergências entre o método com e sem a regressão simbólica para o Poço 10.