

**EFEITO DO TAMANHO DA AMOSTRA DE TREINAMENTO NA PREDIÇÃO DE
PERFIS FALTANTES POR TÉCNICA DE ANÁLISE MULTIVARIADA**

ADALBERTO FRANCISCO MONTEIRO FILHO

**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
UNIDADE ACADÊMICA DE MINERAÇÃO E GEOLOGIA**

CAMPINA GRANDE (PB)

JULHO 2011

EFEITO DO TAMANHO DA AMOSTRA DE TREINAMENTO NA PREDIÇÃO DE
PERFIS FALTANTES POR TÉCNICA DE ANÁLISE MULTIVARIADA

ADALBERTO FRANCISCO MONTEIRO FILHO

Dissertação apresentada a Unidade Acadêmica de Mineração e Geologia da Universidade Federal de Campina Grande, como parte das exigências para obtenção do título de Mestre em Ciências.

Orientador: Prof. José Agnelo Soares, DSc.

CAMPINA GRANDE (PB)

JULHO 2011





FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL DA UFCC

M775e

Monteiro Filho, Adalberto Francisco

Efeito do Tamanho da Amostra de treinamento na Predição de Perfis Faltantes por técnica de Análise Multivariada / Adalberto Francisco Monteiro Filho. — Campina Grande, 2011.

112 f.: il. col.

Dissertação (Mestrado em Engenharia de Minas) – Universidade Federal de Campina Grande, Centro de Tecnologia e Recursos Naturais.

Referências.

Orientador: Prof. Dr. José Agnelo Soares.

1. Perfilagem de Poços. 2. Análise Multivariada. 3. KNN. 4. Petróleo. I. Título.

CDU 553.982(043)

EFEITO DO TAMANHO DA AMOSTRA DE TREINAMENTO NA PREDIÇÃO DE
PERFIS FALTANTES POR TÉCNICA DE ANÁLISE MULTIVARIADA

ADALBERTO FRANCISCO MONTEIRO FILHO

Dissertação apresentada a Unidade
Acadêmica de Mineração e Geologia
da Universidade Federal de Campina
Grande, como parte das exigências
para obtenção do título de Mestre em
Ciências.

Aprovada em 04/07/2011

Comissão Examinadora:


Arlindo José Bazante (DSc. - UAMG/UFCG)


Francisco César Nogueira (DSc. - DEM/UFCG)


José Agnelo Soares - Orientador (DSc. - UAMG/UFCG)

2011

AGRADECIMENTOS

Gostaria de expressar o mais profundo agradecimento a todas as pessoas que direta ou indiretamente contribuíram para o êxito deste trabalho, pela ajuda e apoio durante o período que trabalhei nesta dissertação. Seria impossível mencionar a todas.

A minha família, pela paciência e motivação que sempre manifestaram ao longo deste trabalho.

Ao meu orientador, Prof. Dr. José Agnelo Soares, pelo apoio permanente, suas sugestões valiosas, paciência e dedicação.

Ao Prof. Dr. Arlindo José Bazante, pelas valiosas sugestões e contribuições para o desfecho do trabalho.

Aos membros da banca examinadora pela leitura crítica e valiosas sugestões que muito contribuíram para a melhoria deste trabalho.

À Universidade Federal de Campina Grande (UFCG).

À Unidade Acadêmica de Mineração e Geologia, seus coordenadores e funcionários pelo apoio prestado ao longo do trabalho.

RESUMO

Perfis geofísicos de poços constituem uma importante fonte de dados para a avaliação do potencial de uma dada área, com fins de exploração de recursos naturais subterrâneos. Por diversas razões, dentre as quais destacamos as dificuldades operacionais, indisponibilidade de sondas ou por questões relativas à priorização no momento da contratação do serviço de aquisição dos dados, assim sendo dificilmente se dispõe do conjunto completo de perfis, de modo que pode haver perfis faltantes em um dado poço. No entanto, uma característica marcante dos perfis geofísicos de poços é que existe certo nível de redundância entre eles, de modo que se pode fazer uso desta característica para estimar um dado perfil faltante a partir dos demais perfis disponíveis. Neste trabalho as estimativas de perfis faltantes foram realizadas através da aplicação da técnica de estatística multivariada KNN (K-ésimos valores mais próximos), a qual se baseia na medição da similaridade, no espaço multidimensional, entre valores dos diversos perfis de uma amostra de treinamento. A estimativa do valor de um dado perfil faltante é obtida pela similaridade entre os perfis disponíveis e os perfis da amostra de treinamento. Em particular, neste trabalho se avaliou o efeito do tamanho da amostra de treinamento sobre a eficiência da predição pelo método KNN. Os resultados obtidos mostram que o aumento do tamanho da amostra de treinamento conduz à redução da diferença entre a curva real e a estimada. No entanto, amostras de treinamento com apenas 10% dos dados disponíveis já fornecem resultados aceitáveis de predição para os perfis cuja variação ocorre em escala linear, como os perfis de densidade, sônico, raios gama e porosidade de nêutrons. Para o perfil de resistividade elétrica não se obteve resultados aceitáveis, pois o mesmo não guarda redundância em relação aos demais perfis disponíveis.

Palavras-chave: Perfilagem de poços, análise multivariada, KNN, petróleo.

ABSTRACT

Geophysical well logs constitute an important data source for the evaluation of the potential of a given area for underground natural resources exploitation purposes. For several reasons, hardly the complete log set is available, so that there may be missing logs in a given well. However, a geophysical well logs characteristic is that there is some level of redundancy between them, so this feature can be used to estimate a given missing log from other available logs. This work estimate missing logs through the application of the multivariate statistical technique KNN (K-thnearest neighbor) which is based on the measurement of similarity between the values of several logs of a training sample set. The estimated value of a given missing log is obtained by the similarity between the available logs and the training set values. In particular this work evaluates the effect of the training set size on the efficiency of KNN prediction. The results obtained show that increasing the training set size leads to a reduction of the difference between the actual log and the estimated one. However, training set with only 10% of the available data already provide acceptable prediction for logs whose variation occurs in linear scale, as density, sonic, gamma ray and neutron porosity logs. For electrical resistivity log KNN not achieved acceptable results, because this log does not have redundancy with other available logs.

Key-words: Well logging, multivariate analysis, KNN, petroleum.

LISTA DE ILUSTRAÇÕES

Figura 1 - Movimento e direção das ondas P e S	8
Figura 2 - Classificação de um novo elemento	15
Figura 3 - Mapa de localização dos poços do Campo de Namorado.....	18
Figura 4 - Tela de dados do programa SAS <i>Enterprise Guide</i> ® com o plugin da Função Discriminante	20
Figura 5 - Tela do SAS <i>Enterprise Guide</i> ® para seleção do perfil a estimar a partir dos perfis de análise	21
Figura 6 - Tela do SAS <i>Enterprise Guide</i> ® para seleção do método estatístico de análise	22
Figura 7 - Tela do SAS <i>Enterprise Guide</i> ® para definição dos parâmetros da função discriminante	23
Figura 8 - Tela do SAS <i>Enterprise Guide</i> ® para incluir os dados de classificação e gerar arquivo com os dados de saída	24
Figura 9 - Tela do SAS <i>Enterprise Guide</i> ® com o arquivo de saída, o qual inclui a coluna com o perfil faltante estimado	25
Figura 10 - Fluxograma para obtenção dos conjuntos de dados de treinamento	27
Figura 11 - Fluxograma para aplicação das regras discriminantes e estimação dos perfis faltantes	27
Figura 12 - Perfis sônicos do poço NA-01a	29
Figura 13 - Perfis sônicos do poço NA-02	30
Figura 14 - Perfis sônicos do poço NA-04.....	31
Figura 15 - Perfis sônicos do poço NA-05a	32
Figura 16 - Perfis sônicos do poço NA-07.....	33
Figura 17 - Perfis sônicos do poço NA-11a	34
Figura 18 - Perfis sônicos do poço NA-12.....	35
Figura 19 - Perfis sônicos do poço NA-17a	36
Figura 20 - Perfis sônicos do poço NA-21b	37
Figura 21 - Perfis sônicos do poço RJS-19.....	38
Figura 22 - Perfis sônicos do poço RJS-42.....	39

Figura 23 - Perfis sônicos do poço RJS-234.....	40
Figura 24 - Perfis raios gama do poço NA-01a	41
Figura 25 - Perfis raios gama do poço NA-02	42
Figura 26 - Perfis raios gama do poço NA-04	43
Figura 27 - Perfis raios gama do poço NA-05a	44
Figura 28 - Perfis raios gama do poço NA-07	45
Figura 29 - Perfis raios gama do poço NA-11a	46
Figura 30 - Perfis raios gama do poço NA-12	47
Figura 31 - Perfis raios gama do poço NA-17a	48
Figura 32 - Perfis raios gama do poço NA-21b	49
Figura 33 - Perfis raios gama do poço RJS-19	50
Figura 34 - Perfis raios gama do poço RJS-42	51
Figura 35 - Perfis raios gama do poço RJS-234.....	52
Figura 36 - Perfis resistividade elétrica do poço NA-01a	53
Figura 37 - Perfis resistividade elétrica do poço NA-02	54
Figura 38 - Perfis resistividade elétrica do poço NA-04	55
Figura 39 - Perfis resistividade elétrica do poço NA-05a	56
Figura 40 - Perfis resistividade elétrica do poço NA-07	57
Figura 41 - Perfis resistividade elétrica do poço NA-11a	58
Figura 42 - Perfis resistividade elétrica do poço NA-12	59
Figura 43 - Perfis resistividade elétrica do poço NA-17a	60
Figura 44 - Perfis resistividade elétrica do poço NA-21b	61
Figura 45 - Perfis resistividade elétrica do poço RJS-19.....	62
Figura 46 - Perfis resistividade elétrica do poço RJS-42	63
Figura 47 - Perfis resistividade elétrica do poço RJS-234.....	64
Figura 48 - Perfis porosidade de nêutrons do poço NA-01a	65
Figura 49 - Perfis porosidade de nêutrons do poço NA-02	66

Figura 50 - Perfis porosidade de nêutrons do poço NA-04	67
Figura 51 - Perfis porosidade de nêutrons do poço NA-05a	68
Figura 52 - Perfis porosidade de nêutrons do poço NA-07	69
Figura 53 - Perfis porosidade de nêutrons do poço NA-11a	70
Figura 54 - Perfis porosidade de nêutrons do poço NA-12	71
Figura 55 - Perfis porosidade de nêutrons do poço NA-17a	72
Figura 56 - Perfis porosidade de nêutrons do poço NA-21b	73
Figura 57 - Perfis porosidade de nêutrons do poço RJS-19	74
Figura 58 - Perfis porosidade de nêutrons do poço RJS-42	75
Figura 59 - Perfis porosidade de nêutrons do poço RJS-234	76
Figura 60 - Perfis densidade do poço NA-01a	77
Figura 61 - Perfis densidade do poço NA-02	78
Figura 62 - Perfis densidade do poço NA-04	79
Figura 63 - Perfis densidade do poço NA-05a	80
Figura 64 - Perfis densidade do poço NA-07	81
Figura 65 - Perfis densidade do poço NA-11a	82
Figura 66 - Perfis densidade do poço NA-12	83
Figura 67 - Perfis densidade do poço NA-17a	84
Figura 68 - Perfis densidade do poço NA-21b	85
Figura 69 - Perfis densidade do poço RJS-19	86
Figura 70 - Perfis densidade do poço RJS-42	87
Figura 71 - Perfis densidade do poço RJS-234	88
Figura 72 - Diferença RMS para o perfil DT	89
Figura 73 - Diferença RMS para o perfil GR	90
Figura 74 - Diferença RMS para o perfil ILD	91
Figura 75 - Diferença RMS para o perfil NPHI	92
Figura 76 - Diferença RMS para o perfil RHOB	93

LISTA DE EQUAÇÕES

Resistência em função da resistividade.....	5
Arranjo matricial do tipo n-por-m	13
Distância Euclidiana entre dois elementos.....	14
Cálculo da diferença RMS (<i>Root Mean Square</i>).....	27

LISTA DE TABELAS

Tabela 1 - Resistividades elétricas de materiais	5
Tabela 2 - Velocidades da onda P típicas para diversos tipos de rocha	9
Tabela 3 - Tempo de trânsito característico de alguns materiais	12
Tabela 4 - Perfis geofísicos disponíveis e utilizados neste trabalho	19
Tabela 5 - Quadro-resumo das diferenças RMS máximas e mínimas, em relação ao valor médio, obtidas na predição de perfis faltantes	92

LISTA DE SÍMBOLOS E SIGLAS

DT.....	Perfil Sônico
GR	Perfil de Raios Gama
ILD.....	Perfil de Indução
KNN.....	K Nearest Neighbors
NPHI.....	Perfil Neutrônico
RHOB.....	Perfil de Densidade
SAS.....	Statistical Analysis System
SP.....	Perfil Potencial Espontâneo
SPSS	Statistical Package for the Social Sciences

SUMÁRIO

Resumo.....	i
Abstract.....	ii
Lista de ilustrações.....	iii
Lista de equações.....	vi
Lista de tabelas.....	vii
Lista de símbolos e siglas.....	viii

Capítulo I

1.1 Introdução.....	1
1.2 Objetivos.....	3
1.2.1 Propriedades físicas registradas nos perfis.....	3
1.2.1.1 Resistividade elétrica.....	5
1.2.1.2 Propriedades radioativas.....	6
1.2.1.3 Propriedades acústicas.....	8
1.2.2 Perfis utilizados neste trabalho.....	10
1.2.2.1 Raios gama (GR).....	10
1.2.2.2 Neutrônico (NPHI).....	10
1.2.2.3 Indução (ILD).....	11
1.2.2.4 Sônico (DT).....	11
1.2.2.5 Densidade (RHOB).....	12
1.3 Análise multivariada.....	12
1.3.1 Medidas de similaridade.....	15
1.3.1.1 Método da ligação completa.....	15
1.3.1.2 Método da ligação média.....	16
1.3.1.3 Método do centróide.....	16
1.3.1.4 Método de Ward.....	17

Capítulo II

2.1 Dados e programas utilizados.....	18
2.1.1 Dados utilizados.....	18
2.1.2 Programa <i>SAS Enterprise Guide</i> [®]	19
2.2 Metodologia.....	26

Capítulo III

3.1 Resultados	29
3.1.1 Perfil sônico (DT)	29
3.1.2 Perfil de raios gama (GR)	41
3.1.3 Perfil de resistividade (ILD).....	53
3.1.4 Perfil de porosidade de nêutrons (NPHI).....	65
3.1.5 Perfil de densidade (RHOB)	77
3.2 Análise da precisão dos resultados	89
3.2.1 Diferença RMS.....	89

Capítulo IV

4.1 Conclusões	94
-----------------------------	----

Capítulo V

5.1 Referências bibliográficas	96
---	----

CAPÍTULO I

1.1 - INTRODUÇÃO

Os fenômenos geológicos geralmente não apresentam natureza determinística, assim sendo, a complexidade na realização de estimativas é muito grande. Os modelos probabilísticos são os métodos mais viáveis para contornar esta dificuldade, a qual poderia ser minimizada pela obtenção de testemunhos de poços, porém este processo é muito oneroso e limitado em relação a quantidade de dados coletados. Por outro lado, se faz necessário o conhecimento da relação existente entre os diversos poços adjacentes e a possível correlação dos dados coletados e dos dados faltantes entre eles. Tal fato possibilita a extrapolação dos dados para os poços que por algum motivo não puderam ser testemunhados. Dentre as possíveis dificuldades, destaca-se: as operacionais, indisponibilidade de sondas, ou questões relativas à priorização no momento da contratação do serviço de aquisição dos dados ou impossibilidade de se conhecer algum dos perfis geofísicos. Genericamente os perfis são obtidos através dos parâmetros medidos em função das suas propriedades derivadas.

Do ponto de vista operacional, a rápida difusão destes métodos se deve, dentre outros fatores, às facilidades proporcionadas por novos programas e máquinas. Interfaces amigáveis de pacotes estatísticos, como SAS® e SPSS®, tornaram bastante simples a tarefa de construir modelos para a geração de curvas sintéticas [Buchebe et al., 1997 *apud* Ribeiro, 2008].

Diversos estudos mostram uma boa aplicabilidade das técnicas da análise estatística multivariada no processamento e interpretação de dados geofísicos [Couto, 1994 *apud* Ribeiro, 2008]. No caso da geofísica de poço, uma aplicação frequentemente apresentada trata da identificação e separação das eletrofácies e a sua associação aos diversos litotipos atravessados pelo poço. Outros estudos, onde também se têm aplicado a estatística multivariada com sucesso, se referem a problemas da engenharia de reservatório [Nitters et al., 1995 *apud* Ribeiro, 2008].

Wilbert [1986 *apud* Ribeiro, 2008] afirma que se considera rotina operacional no reestudo de campos de petróleo por perfis, a extrapolação, para os demais poços do

campo, de parâmetros e procedimentos calibrados com litologias distinguíveis em poços testemunhados (litofácies), dentro de eletrofácies rastreáveis através de perfis. Segundo este mesmo autor, a caracterização das eletrofácies, normalmente se faz considerando suas respostas típicas em distintos perfis ou em associações de perfis por grupo litológico. O padrão de resposta eletrofácies, assumido como representativo de uma determinada situação litológica definida nos testemunhos por petrografia e ou petrofísica, é rastreado pelo computador sobre dados digitalizados de perfis, ao longo de toda a área.

Sendo assim, é importante ter uma suíte de perfis completa para a determinação computacional mais precisa das eletrofácies. Como uma solução para aumentar a precisão desta caracterização, esta pesquisa visa obter curvas sintéticas faltantes de poços perfilados através de modelagem estatística.

Para contornar este problema, o primeiro passo é a obtenção de dados de boa qualidade e, com o auxílio da modelagem numérica, poder oferecer estimativas razoáveis e coerentes como resultado. Neste contexto, a modelagem de dados se torna uma importante ferramenta, pois permite a simulação dos dados de regiões desconhecidas, tornando a exploração mais segura.

1.2 - OBJETIVOS

O objetivo principal deste estudo é aplicar técnicas de análise multivariada para realizar estimativas de perfis geofísicos faltantes a partir dos demais perfis existentes nos poços do Campo de Namorado, na Bacia de Campos. A ideia central é aproveitar a redundância das informações existentes nos dados para estabelecer correlações, às vezes não lineares, entre os diversos perfis geofísicos de um poço. Desta forma se pode realizar a estimativa de um perfil faltante em um poço, ou de um perfil inexistente em um dado intervalo de profundidade. Outro objetivo central é avaliar a influência do tamanho da amostra de treinamento sobre a eficiência de predição.

A fim de atingir estes objetivos principais, os seguintes objetivos específicos devem ser perseguidos:

- Edição do arquivo contendo os dados de todos os poços a analisar;
- Extração dos arquivos com as amostras de treinamento selecionadas de forma aleatória;
- Executar a análise, gerando os perfis estimados;
- Preparar os gráficos para a comparação visual dos perfis estimados com os seus correspondentes perfis reais;
- Calcular as diferenças médias quadráticas entre os perfis estimados e os reais;
- Analisar os resultados obtidos.

1.2.1 – PROPRIEDADES FÍSICAS REGISTRADAS NOS PERFIS

A procura pela otimização da exploração dos reservatórios petrolíferos é fundamental para o crescimento da indústria do petróleo. Para que haja um cenário propício para a otimização de um reservatório é necessário que seja feita uma caracterização completa que esboça e identifica suas formas e suas propriedades petrofísicas (porosidade, permeabilidade, argilosidade e saturação). A partir daí técnicas de correlação entre atributos sísmicos e propriedades petrofísicas são obtidas e analisadas.

Quando da perfuração de poços de petróleo a perfilagem é uma ferramenta das mais úteis e mais importantes, possibilitando o mapeamento completo do poço, mensurando as diferentes propriedades físicas das formações, e a partir da interpretação dos perfis, identifica os diversos componentes formadores do reservatório (zonas de óleo, gás ou água), além de fazer a estimativa do volume de hidrocarbonetos que pode vir a ser produzido. A perfilagem geofísica apesar de fundamentar a avaliação do reservatório, é um método que representa um pequeno percentual do custo total do poço, entretanto fornece importantes informações sobre as propriedades da rocha nas condições do reservatório.

Esses perfis podem ser usados para realizar a definição da litologia, uma vez que respondem a uma grande quantidade de parâmetros geológicos e podem ser corridos em toda a extensão do poço obtendo-se medidas contínuas em todo o trajeto, além de revelar diferentes propriedades das formações. Porém, realizar a definição litológica, a partir dos perfis de poços, pode ser uma tarefa complexa, especialmente quando grandes quantidades de informações estão envolvidas. Para tornar essa tarefa eficaz, a estatística multivariada, que envolve etapas como a redução da dimensionalidade e o reconhecimento de padrões, é uma prática adotada com frequência [Talaat & Mostafa, 1989]. A análise multivariada é adotada neste trabalho com o objetivo de estimar perfis inexistentes em um poço.

O perfil de um poço é obtido a partir de ferramentas que são descidas no poço, onde os valores são registrados e em seguida são armazenados em arquivos digitais para posteriormente serem recuperados e analisados.

A efetivação do registro das propriedades físicas dos poços é baseada principalmente nas propriedades elétricas, radioativas e acústicas das rochas, sendo esta última também muito importante no estudo de propagação das ondas sísmicas.

1.2.1.1 - RESISTIVIDADE ELÉTRICA

A resistência elétrica é definida como sendo a medida da capacidade de um material qualquer de se opor à passagem da corrente elétrica. A resistência (R) é diretamente proporcional ao comprimento (L), a ser percorrido pela corrente elétrica, e à resistência específica ou resistividade do meio (ρ) e inversamente proporcional à seção transversal (A) atravessada. A resistência é dada por:

$$R = \rho \frac{L}{A} (\Omega) \quad (1)$$

A rocha sedimentar é formada, quase que na sua totalidade, por minerais de resistividade muito alta, conseqüentemente a corrente elétrica não tem propagação pelos grãos, mas sim através das soluções eletrolíticas que ocupam os espaços vazios interligados, que são muito mais condutivos.

A resistividade das rochas sedimentares varia entre 0,5 e 1.000 Ωm . A resistividade das soluções eletrolíticas intersticiais (que estão nos poros das rochas, e em vesículas ou fraturas), variam entre 0,05 e 100 Ωm (ver Tabela 1). Desta forma, uma rocha se torna mais ou menos condutora da corrente elétrica, a depender da maior ou menor interconexão entre seus poros (porosidade e permeabilidade) e da maior ou menor concentração iônica de sua solução (salinidade). A escassez de fluidos condutores ou então a presença de fluidos isolantes, como o petróleo, torna a rocha menos condutiva.

Tabela 1 - Resistividades elétricas de materiais (modificado de Tittman, 1986)

Material	Resistividade (Ωm a 18 -20°C)
Quartzo	10^{12} a 3×10^{14}
Petróleo	10^9 a 10^{16}
Água destilada a 2 kppm NaCl	3,4
Água destilada a 10 kppm NaCl	0,72
Água destilada a 20 kppm NaCl	0,38
Água destilada a 100 kppm NaCl	0,09
Água destilada a 200 kppm NaCl	0,06
Argila / Folhelho	2 – 10
Arenito com água salgada	0,5 – 10
Arenito com óleo	$5 - 10^3$
Calcário compactado	10^3 a 10^4
Dolomita	10^3 a 10^4

A gênese do registro elétrico é atribuída a Conrad Schlumberger, professor de física da École de Mines de Paris, França, que concebeu a ideia da prospecção de depósitos de minério de metal usando os princípios elétricos de condutividade para distingui-los de outros menos condutores. Um dos primeiros testes, de acordo com historiadores, Schlumberger realizou em sua banheira, que estava cheia de várias rochas para o experimento. Trabalhando com seu irmão Marcel Schlumberger, Conrad começou uma série de pesquisas na Europa, África e América do Norte durante um período de três anos.

Em 1926, os irmãos formaram a *Société de Prospection Electrique* e desenvolveram a teoria que adicionando informações de resistividade das formações mais profundas aumentaria a eficácia da prospecção. A experiência consistiu em baixar uma sonda elétrica a profundidade de 1.600 pés (~48 m) em um poço no campo de Pechelbronn, França, em 05 de setembro de 1927. Assim os irmãos Schlumberger criaram o primeiro perfil de poço. Este registro foi meticulosamente gravado ponto por ponto, metro a metro, utilizando um equipamento improvisado e as sucessivas leituras foram então plotadas juntas. A tecnologia era simples, mas funcional. Era composta por três eletrodos A, M e N, que eram descidos ao fundo do poço em três fios isolados. Um eletrodo de corrente A passa pela lama de perfuração e se espalha pela formação. Os potenciais medidos em M e N são transmitidos para a superfície onde eles são medidos. Ao medir a diferença de potencial entre M e N, e a intensidade da corrente de A, a resistividade aparente da formação é calculada.

Após o sucesso inicial com este primeiro registro a tecnologia de registrar permitiu a aquisição de diferentes conjuntos de dados nascidos a partir de ferramentas que forneciam dados sobre raios gama, SP, resistividade, sônico e caliper com uma simples coleta.

1.2.1.2- PROPRIEDADES RADIOATIVAS

As propriedades químicas são determinadas pelo número de prótons nos núcleos atômicos e caracterizam diferentes elementos químicos. Um mesmo elemento pode ter um número definido de prótons podendo, entretanto, variar o seu número de nêutrons (isótopos) cada isótopo tem suas próprias propriedades radioativas. Alguns

isótopos são estáveis, ou seja, não há mudança na estrutura atômica ou estado energético. Por outro lado, quando as forças nucleares se desestabilizam, os isótopos modificam sua estrutura, passando a emitir energia em forma de radiação, na tentativa de retornar à estabilidade, e se transformar em elementos diferentes. De quase 1400 isótopos conhecidos atualmente, 1130 deles são instáveis e apenas 65 ocorrem naturalmente, os demais são artificiais [Thomas, 2001].

A magnitude de radioatividade natural das rochas depende do teor de três elementos: urânio, tório e do isótopo radioativo do potássio (K^{40}). A importância destes elementos deve-se à relativa abundância litológica, em relação a outros elementos radioativos, e também às suas meias-vidas elevadas, se aproximando à idade da Terra.

De acordo com Thomas [2001], as rochas sedimentares podem ser divididas, de acordo com sua radioatividade natural, em três grandes grupos:

- Rochas altamente radioativas – folhelhos de águas profundas, folhelhos pretos betuminosos, evaporitos potássicos e algumas rochas ígneas e metamórficas;
- Rochas medianamente radioativas – folhelhos e arenitos argilosos de águas rasas, carbonatos e dolomitos argilosos;
- Rochas de baixa radioatividade – grande maioria de arenitos, carvões e evaporitos não potássicos (halita, anidrita, gipsita).

1.2.1.3 - PROPRIEDADES ACÚSTICAS

Dois tipos importantes de mecanismos de transporte de energia são suportados pelo meio elástico: ondas compressoriais (onda P) e cisalhantes (onda S). A Figura 1 apresenta uma explicação esquemática para estes tipos de onda.

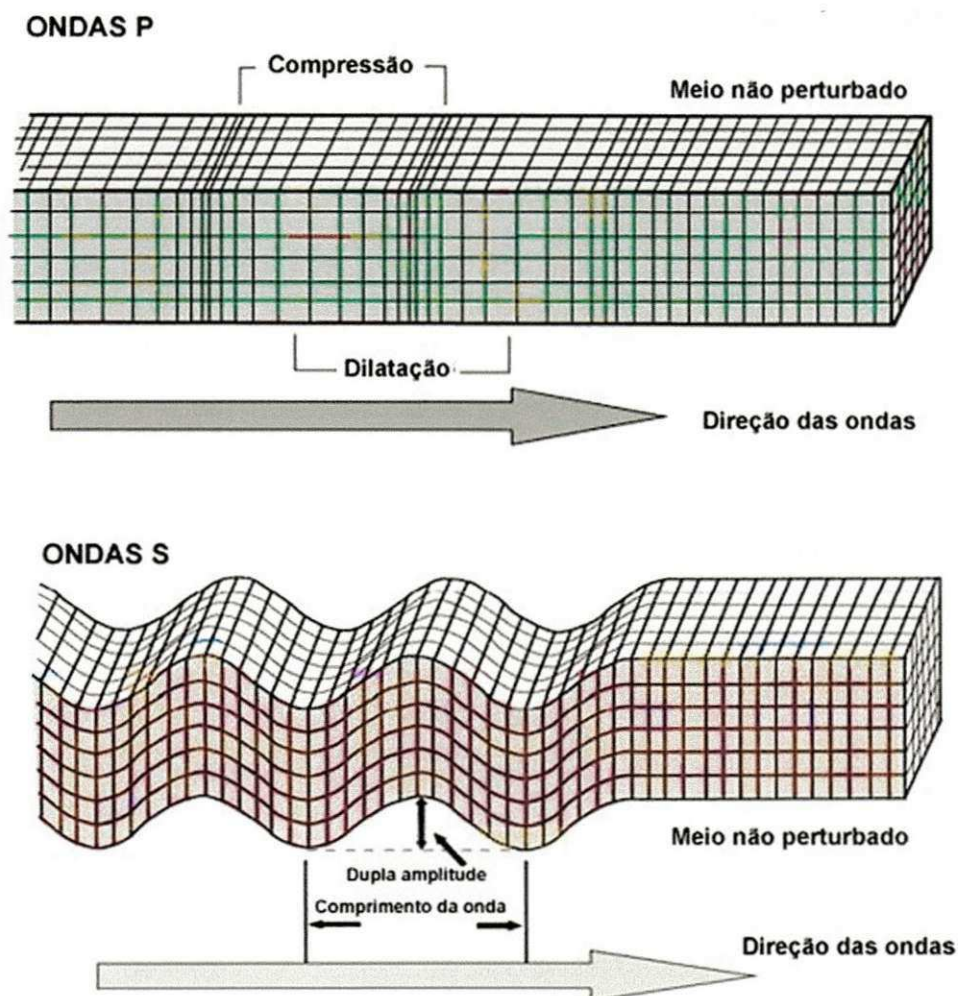


Figura 1 – Movimento e direção das ondas P e S (adaptado de http://www.apolo11.com/terremotos_wave.php).

A velocidade de propagação do som depende do meio em que viajam as ondas elásticas, sendo muito mais rápido nos sólidos que nos fluidos. Portanto, a velocidade de propagação da onda depende do material (mineralogia), da separação entre os componentes sólidos (geometria porosa) e da concentração dos

componentes fluidos (saturação). Desta forma, o tempo que o som leva para percorrer um determinado espaço de um material pode ser usado para determinação das constantes elásticas deste material e análise quantitativa da porosidade. A Tabela 2 apresenta as velocidades típicas da onda compressional para diversos materiais geológicos.

Tabela 2 – Velocidades da onda P típicas para diversos tipos de rocha.

Velocidade (m/s)	Provável tipo de rocha
< 400	Solos, depósitos superficiais de sedimentos não consolidados.
400 - 1400	Argilas e areias, não consolidados.
1400 - 1800	Areias saturadas, argilas compactadas, rochas bastante alteradas.
1800 - 2400	Sedimentos consolidados e provavelmente saturados, rochas metamórficas ou ígneas, altamente fraturadas e/ou alteradas, arenitos e folhelhos.
2400 - 3700	Folhelhos, arenitos, rochas ígneas e metamórficas alteradas e /ou fraturadas.
3700 - 4500	Rochas ígneas e metamórficas fracamente alteradas ou fraturadas, carbonatos.
4500 - 6000	Rochas ígneas e metamórficas sãs, não fraturadas.

Existem vários tipos de perfis utilizados para as mais diversas aplicações, todos com o objetivo de avaliar melhor as formações geológicas quanto à ocorrência de uma jazida comercial de hidrocarbonetos. Os perfis mais comuns são: Potencial Espontâneo, Raios Gama, Neutrônico, Indução, Sônico, Densidade e Caliper.

1.2.2 – PERFIS UTILIZADOS NESTE TRABALHO

1.2.2.1 - RAIOS GAMA (GR)

Este perfil permite detectar e avaliar a radioatividade natural da formação geológica. O principal parâmetro de medição é a amplitude de um pulso radioativo oriundo das rochas, sendo função da energia do fóton que é capturado no detector. A intensidade, ou quantidade da radiação, está relacionada com o número de fótons detectados por unidade de tempo. Este método começou a ser utilizado no final dos anos 30, sendo o primeiro perfil de poço não resistivo. Nessa época, foi utilizado para distinguir formações argilosas de formações com pouca argila (limpas).

Utilizado na identificação da litologia, identificação de minerais radioativos e para o cálculo do volume de argilas ou argilosidade. Também pode ser útil para interpretação de ambientes deposicionais e na investigação da subida do contato óleo-água em reservatórios fraturados.

1.2.2.2 - NEUTRÔNICO (NPHI)

Sendo os nêutrons partículas destituídas de carga elétrica, podem penetrar profundamente na matéria, atingindo os núcleos dos elementos que compõem a rocha, onde interagem através de choques elásticos e/ou inelásticos. Os perfis mais antigos medem a quantidade de raios gama de captura após excitação artificial através de bombardeio dirigido de nêutrons rápidos. Os mais modernos medem a quantidade de nêutrons epitermais e/ou termais da rocha após o bombardeio.

É utilizado para estimativa de porosidade, determinação do volume de argila, pode auxiliar na identificação da litologia e dos fluidos da formação e na detecção de hidrocarbonetos leves ou gás.

1.2.2.3 - INDUÇÃO (ILD)

Fornece leitura aproximada da resistividade, através da medição de campos elétricos e magnéticos induzidos nas rochas. A resistividade é a propriedade da rocha permitir ou não a passagem da corrente elétrica. A ferramenta de indução foi desenvolvida para ler medidas de resistividades profundas na formação com o mínimo de distorção provocada pela zona invadida.

A ferramenta possui uma ou mais bobinas transmissoras com corrente alternada que é usada para energizar as rochas adjacentes ao poço. A maior parte desta corrente é focalizada através do poço, o que a faz penetrar profundamente na formação. Desta forma, as voltagens induzidas nas bobinas receptoras estimam a condutividade da formação [Doveton, 1986].

É um perfil muito utilizado na indústria petrolífera por ser um método muito eficiente para indicar o contato água-óleo devido ao contraste muito elevado de valores destes fluidos [Thomas, 2001].

1.2.2.4 - SÔNICO (DT)

Conhecido também como perfil acústico mede a diferença nos tempos de trânsito de uma onda mecânica através das rochas. É utilizado para estimativa de porosidade, identificação de litologia, correlação poço a poço, estimativas do grau de compactação das rochas, estimativa das constantes elásticas, detecção de fraturas e apoio à sísmica. Diferentemente da medida de resistividade, que pode ser usada diretamente na detecção de petróleo, e das medidas radioativas que foram inicialmente voltadas para determinação de porosidade, o perfil acústico iniciou como um método complementar para a exploração sísmica [Thomas, 2001].

Esta ferramenta registra o tempo decorrido entre o momento em que um pulso sonoro compressional é emitido por um transmissor, até sua chegada a dois receptores distintos. A Tabela 3 apresenta os tempos de trânsito típicos para alguns materiais.

Tabela 3 - Tempo de trânsito característico de alguns materiais [Doveton, 1986].

Material	Tempo de trânsito ($\mu\text{s}/\text{m}$)
Quartzo	16,8
Calcita	14,5
Dolomita	13,3
Petróleo (média)	72,5
Água doce	56,7

1.2.2.5 - DENSIDADE (RHOB)

Este perfil tem a função de detectar os raios gama defletidos pelos elétrons orbitais dos elementos componentes das rochas, após terem sido emitidos por uma fonte colimada situada dentro do poço. Além da densidade das camadas, permite o cálculo da porosidade e a identificação das zonas de gás. É utilizado também como apoio à sísmica para o cálculo do sismograma sintético.

A ferramenta de densidade é formada por uma fonte radioativa de raios gama, geralmente césio-137, montada em um mandril com patins metálicos, os quais são pressionados contra a parede do poço. A fonte é responsável pela emissão de raios gama de alta energia, que interagem com os elétrons dos átomos da formação. Esta interação pode ser de três modos diferentes: efeito fotoelétrico, efeito Compton e produção do par elétron-pósitron; onde apenas os dois primeiros são de interesse para a geofísica de poço.

1.3 - ANÁLISE MULTIVARIADA

A análise multivariada visa entender e descrever a relação entre um número arbitrário de variáveis. Os geocientistas frequentemente lidam com conjuntos de dados multivariados, tais como propriedades físicas e químicas de minerais e rochas.

Se existem relações complexas entre os diferentes parâmetros, estatísticas univariadas ignoram o conteúdo da informação dos dados. Um conjunto de dados multivariados consiste em medidas de p variáveis em n objetos. Tais conjuntos de

dados são normalmente armazenados em arranjos matriciais do tipo n-por-m [Trauth, 2006].

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{1m} \\ x_{21} & x_{22} & x_{2m} \\ \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{nm} \end{bmatrix} \quad (2)$$

Onde as colunas da matriz representam as variáveis m, e as linhas representam os n objetos.

A análise multivariada é usada para descrever numerosas observações ou variáveis que são obtidas para cada indivíduo ou unidade estudada. Descobrir as inter-relações complexas entre diversas medidas no mesmo indivíduo ou unidade, e fazer a interpretação desses resultados é o papel predominante da análise multivariada. Quando algo é estimado, na realidade é realizado um modelamento numérico do que será medido, seguindo alguma regra para atribuir um número à característica particular que está sendo estimada.

Os algoritmos mais comumente usados para este modelamento são classificados em duas categorias gerais: (1) hierárquica e (2) não hierárquica

Os procedimentos hierárquicos envolvem a construção de uma hierarquia de uma estrutura do tipo árvore tendo como característica que os resultados de um estágio anterior são sempre aninhados com os resultados de um estágio posterior, criando ramificações parecidas com uma árvore (dendrograma).

Dentre os algoritmos hierárquicos os mais populares usados para desenvolver agregados são (1) ligação individual, (2) ligação média, (3) ligação completa, (4) método de Ward e (5) método do centróide. Esses algoritmos diferem na forma como a distância entre os agrupamentos é computada.

O procedimento de ligação individual é baseado em distância mínima. A técnica de agrupamento hierárquico interliga as amostras por suas similaridades, produzindo um dendrograma onde as amostras semelhantes, segundo as variáveis escolhidas, são agrupadas entre si. A suposição básica de sua interpretação é que quanto menor a distância entre os pontos que representam as amostras no espaço das

variáveis, maior a semelhança entre estas amostras. Desta maneira dois objetos separados pela menor distância são colocados no primeiro agrupamento. Em seguida a próxima distância mais curta é determinada, e um terceiro objeto se junta aos dois primeiros para formar um agregado, ou um novo agrupamento de dois membros é formado. O processo continua até que todos os objetos formem um só agregado. Esse procedimento também é chamado de abordagem ou classificação do vizinho mais próximo (KNN).

A classificação utilizando a regra do vizinho mais próximo (*K-Nearest Neighbors*) é baseada na comparação da distância entre amostras, para isto é calculada uma matriz n-dimensional entre todas as amostras no conjunto de dados, onde n é o número de variáveis nos dados medidos. Numa versão mais simples, KNN consiste na classificação de uma amostra junto com o vizinho mais próximo, dentro de um número de classes pré-definidas. A classe predita de uma amostra teste é então determinada baseada na identidade daquelas amostras mais próximas da amostra desconhecida.

O algoritmo KNN é uma técnica iterativa bastante simples que permite agrupar elementos por analogia. Em um conjunto de treino constituído por vetores n-dimensionais, o algoritmo KNN procura os k elementos mais próximos no grupo de treino. Destes k vizinhos, a classe de maior ocorrência é normalmente utilizada para classificar o novo elemento. Contudo, em algumas classificações específicas, poder-se-á escolher outras relações entre estes k vizinhos para decidir qual a classificação do novo elemento. Podem ser utilizadas diversas distâncias métricas como método de similaridade: euclidiana, Mahalanobis completa ou Mahalanobis diagonal. Neste trabalho é utilizada a distância Euclidiana que mede a extensão geométrica entre dois elementos em um espaço multidimensional.

$$d(h_1, h_2) = \sqrt{\sum_{i=1}^n (|h_1(i) - h_2(i)|)^2} \quad (3)$$

Onde:

h_1 é o valor do primeiro elemento;

h_2 é o valor do segundo elemento.

O maior problema se verifica na correta escolha do valor de quantos vizinhos devem ter similaridade para determinar o valor do próximo elemento (k). Uma regra geral

para esse valor será de considerar $k = (m)^{1/2}$, onde m é o número de elementos considerados na matriz de treinamento. Teoricamente k deverá ser tão grande quanto possível de modo a minimizar o erro, mas à medida que se aumenta o valor de k , aumenta-se o tempo computacional e perde-se resolução devido a suavização excessiva. Se for considerado um valor de k pequeno, apenas serão considerados os elementos muito próximos. [Almeida et al.,2008]

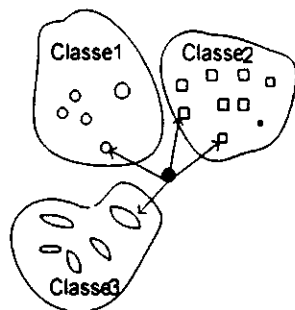


Figura 2 - Classificação de um novo elemento adaptado de Almeida et al.,2008.

1.3.1 – Medidas de similaridade

1.3.1.1 - Método da ligação completa

O procedimento de ligação completa é semelhante ao da ligação individual, exceto em que o critério de agrupamento se baseia em distância máxima, sendo uma das técnicas de hierarquização aglomerativa de maior aplicação na análise de agrupamento [Gama, 1980]. Como no método de ligação individual, aqui também não é exigida a fixação, *a priori*, do número de agrupamentos.

Conforme Bussab et al. [1990], agrupamentos são conectados um com o outro a distância máxima ou similaridade mínima no método da ligação completa, também denominado vizinho mais distante. A dissimilaridade entre dois grupos é definida como sendo aquela apresentada pelos indivíduos de cada grupo que mais se parecem, ou seja, formam-se todos os pares com um membro de cada grupo, e a dissimilaridade entre os grupos é definida pelo par que mais se parece. Este método, geralmente, leva a grupos compactos e discretos, tendo os seus valores de dissimilaridade relativamente grandes.

1.3.1.2 - Método da ligação média

O método de ligação média começa da maneira similar à ligação individual ou completa [Hair et al., 2005], diferenciando-se na maneira de formar os agrupamentos que são feitos a partir da distância média de todos os elementos em um agrupamento aos demais em outro. Essa técnica não depende de valores extremos, como ocorre com a ligação individual ou completa, e a seleção baseada em todos os elementos dos agregados, ao invés de um único par de membros extremos. Abordagens de ligação média tendem a combinar agregados com pequena variação interna. Elas também tendem a produzir agregados com aproximadamente a mesma variância.

1.3.1.3 - Método do centróide

O método do centróide foi proposto por Sokal & Michener [1958] e teve como origem a caracterização da matriz de dados como pontos do espaço Mahalanobis. Cada agrupamento é considerado um simples ponto, representado pelo seu centro de massa, chamado centróide. O presente método utiliza uma função de agrupamento para medir a distância entre os centros de massa dos dados. Esta técnica é de hierarquização aglomerativa.

Este algoritmo se caracteriza pela redefinição, a cada passo, da matriz de dados. Cada agrupamento é representado pelo vetor médio das p variáveis envolvidas. Na realidade, uma nova matriz de distâncias é determinada a cada interação.

No método do centróide, a distância entre dois grupos é definida como a distância entre os seus centróides, pontos definidos pelas médias das variáveis caracterizadoras dos indivíduos de cada grupo, isto é, o método do centróide calcula a distância entre dois grupos como a diferença entre as suas médias, para todas as variáveis. Uma desvantagem desse método é que se dois grupos forem muito diferentes em termos de dimensão, o centróide de um novo agrupamento estará, mais próximo daquele que for maior e as características do grupo menor tenderão a se perder. De fato, com esse método, o centróide de um novo grupo é uma combinação ponderada dos centróides dos dois grupos separados, sendo as ponderações proporcionais ao tamanho destes grupos [Reis, 1997].

1.3.1.4 - Método de Ward

Ward [1963] propõe um processo geral de classificação em que n elementos são progressivamente reunidos dentro de grupos através da minimização de uma função objetiva para cada $(n - 2)$ passos de fusão.

Inicialmente, este algoritmo admite que cada um dos elementos se constitui em um único agrupamento. Considerando a primeira reunião de elementos em um novo agrupamento, a soma dos desvios dos pontos representativos de seus elementos, em relação à média do agrupamento, é calculada, e dá uma indicação de homogeneidade do agrupamento formado. Esta medida fornece a “perda de informação” que se produz ao reunir os elementos em um agrupamento [Gama, 1980].

CAPÍTULO II

2.1 – DADOS E PROGRAMAS UTILIZADOS

2.1.1 - Dados utilizados

Neste trabalho foram utilizados dados de perfilagem geofísica de 12 poços de petróleo do Campo de Namorado, situado na parte central da Bacia de Campos, litoral do estado do Rio de Janeiro, localizado a 70 km da costa. Os poços cujos dados são utilizados neste trabalho estão localizados conforme apresentado no mapa da Figura 3.

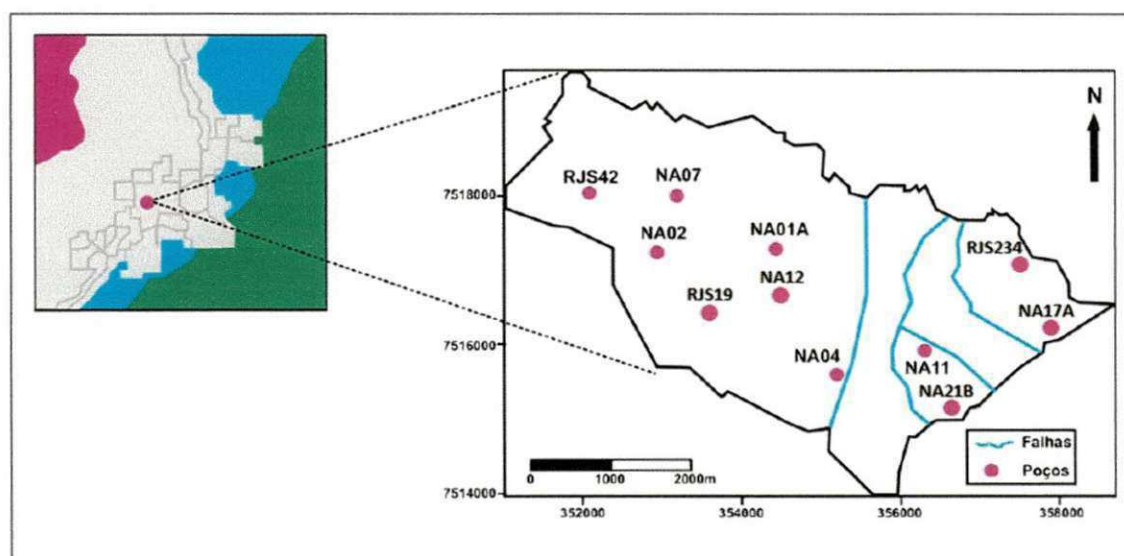


Figura 3 – Mapa de localização dos poços do Campo de Namorado, utilizados neste estudo (Adaptado de Rosa et al., 2008).

A Tabela 4 apresenta a suíte de perfis geofísicos disponíveis em cada poço analisado. Os dados foram fornecidos pela ANP (Agência Nacional de Petróleo) para uso acadêmico pelas instituições de pesquisa nacionais.

Tabela 4 – Perfis geofísicos disponíveis e utilizados neste trabalho.

PERFIL POÇO	DT	GR	ILD	NPHI	RHOB
NA01	✓	✓	✓	✓	✓
NA02	✓	✓	✓	✓	✓
NA04	✓	✓	✓	✓	✓
NA05A	✓	✓	✓	✓	✓
NA07	✓	✓	✓	✓	✓
NA11	✓	✓	✓	✓	✓
NA12	✗	✓	✓	✓	✓
NA17	✓	✓	✓	✓	✓
NA21	✓	✓	✓	✓	✓
RJS019	✓	✓	✓	✓	✓
RJS042	✓	✓	✓	✓	✓
RJS234	✓	✓	✓	✓	✓

2.1.2 – Programa SAS Enterprise Guide®

Este foi o programa utilizado para a análise estatística dos dados. Trata-se de uma interface amigável para utilização do pacote SAS® (*Statistical Analysis Software*) sem a necessidade de conhecimento da linguagem de programação SAS. Este programa permite a incorporação de *plugins* desenvolvidos pelo usuário. Neste trabalho foi utilizado o *plugin* “Discrim Petrobras”, anteriormente desenvolvido [Soares, 2005]. O *plugin* tem a finalidade de criar uma interface amigável para o usuário, dispensando-o de conhecer a linguagem de programação SAS®.

A Figura 4 apresenta a tela inicial com os dados do conjunto de treinamento para definição da regra discriminante pelo método KNN, através do programa SAS Enterprise Guide®. Nesta figura observa-se a incorporação dos *plugins* definidos pelo usuário na aba *Analysis / Custom / Função Discrim Petrobras*.

The screenshot shows the SAS Enterprise Guide interface. The main window displays a data table with the following columns: DT, GR, LD, tRPH, RHOB, and _BTO. The data rows are numbered from 12 to 44. A custom plugin menu is open, showing 'Função Discrim Petrobras' highlighted. The interface includes a menu bar (File, Edit, View, Insert, Format, Tools, Data, Analysis, Graph, Code, Window, Help) and a task pane on the left with categories like 'Add Items to Project', 'Tools', and 'Data'.

	DT	GR	LD	tRPH	RHOB	_BTO	
95	4766	74.375	1.1445	25.7249	2.4829	92.0547	
94	4845	78.6992	1.1443	25.7249	2.448	87.9492	
92	4727	80.2625	1.1631	25.7617	2.4205	87.9492	
91	4291	81.2391	1.2468	26.0015	2.4719	87.9492	
91	6703	81.8203	1.3105	25.3872	2.4176	87.9492	
90	6586	83.1836	1.3596	25.7031	2.4678	92.1445	
88	6662	83.6914	1.4053	26.9023	2.4412	100.5276	
88	1382	82.2656	1.4685	25.9531	2.4716	92.1445	
87	8492	79.2586	1.5227	24.662	2.4775	87.9492	
88	4765	78.4838	1.573	24.3438	2.5266	87.9492	
			1.663	24.8353	2.522	92.1445	
12	2954.6		1.6511	23.9453	2.5193	92.1445	
13	2954.8	84.1523	1.7231	23.0938	2.4861	81.8984	
14	2955	84.1436	1.782	22.2891	2.4873	81.8984	
15	2955.2	84.7383	1.8259	21.3594	2.4864	82.5625	
16	2955.4	85.9844	1.8714	21.1884	2.4836	89.9375	
17	2955.6	88.4102	1.915	21.3036	2.4863	92.1445	
18	2955.8	89.8047	1.958	21.8477	2.4818	92.1445	
19	2956	87.8802	1.9509	22.3123	2.4878	83.5625	
20	2956.2	85.797	1.9665	21.688	2.5017	83.5625	
21	2956.4	85.2283	1.9795	20.2481	2.4871	83.5625	
22	2956.6	86.9789	1.9715	19.8477	2.5271	83.5625	
23	2956.8	88.7461	1.9546	19.6641	2.5404	83.5625	
24	2957	92.5625	1.9245	19.9219	2.5148	83.5625	
25	2957.2	94.6172	1.9901	21.5864	2.5181	92.1445	
26	2957.4	91.2305	1.9549	23.0547	2.527	81.8984	
27	2957.6	90.6657	1.91797	1.9186	22.457	2.5518	83.5625
28	2957.8	89.7383	1.9509	1.9568	21.6802	2.5645	83.5625
29	2958	91.25	1.95391	1.4707	22.1914	2.5611	83.5625
30	2958.2	90.6211	1.82583	1.4749	22.4316	2.4834	83.5625
31	2958.4	86.375	1.9453	1.5803	23.4414	2.4745	81.8984
32	2958.6	81.8984	1.7661	24.5208	2.4653	81.8984	
33	2958.8	80.8436	1.8246	23.9102	2.4866	81.8984	
34	2959	83.1406	1.8692	2.8646	22.4543	2.5027	92.1445
35	2959.2	85.9141	1.8121	2.0234	22.707	2.4751	92.1445
36	2959.4	88.2891	1.8268	2.0547	22.7422	2.4395	92.1445
37	2959.6	86.9977	1.8789	2.1047	21.9727	2.4519	92.1445
38	2959.8	82.9531	1.8769	2.1211	19.9922	2.4742	83.5625
39	2960	82.1211	1.8121	1.8509	2.4891	83.5625	
40	2960.2	83.8477	1.8445	1.8847	2.4599	83.5625	
41	2960.4	85.5147	1.81367	1.8287	19.4786	2.4576	83.5625
42	2960.6	83.1875	1.8121	1.8121	20.6211	2.4717	83.5625
43	2960.8	81.8977	1.83058	2.334	21.1067	2.4879	89.9375
44	2961	80.7988	1.8125	2.5339	20.7305	2.4795	81.125

Figura 4 – Tela de dados do programa SAS Enterprise Guide® com o plugin da Função Discriminante em destaque.

A Figura 5 apresenta a tela do programa SAS Enterprise Guide® referente à seleção das variáveis de análise e classificação. No exemplo mostrado nesta figura, o perfil sônico (variável DT) foi utilizado como a variável a ser estimada a partir dos demais perfis geofísicos disponíveis.

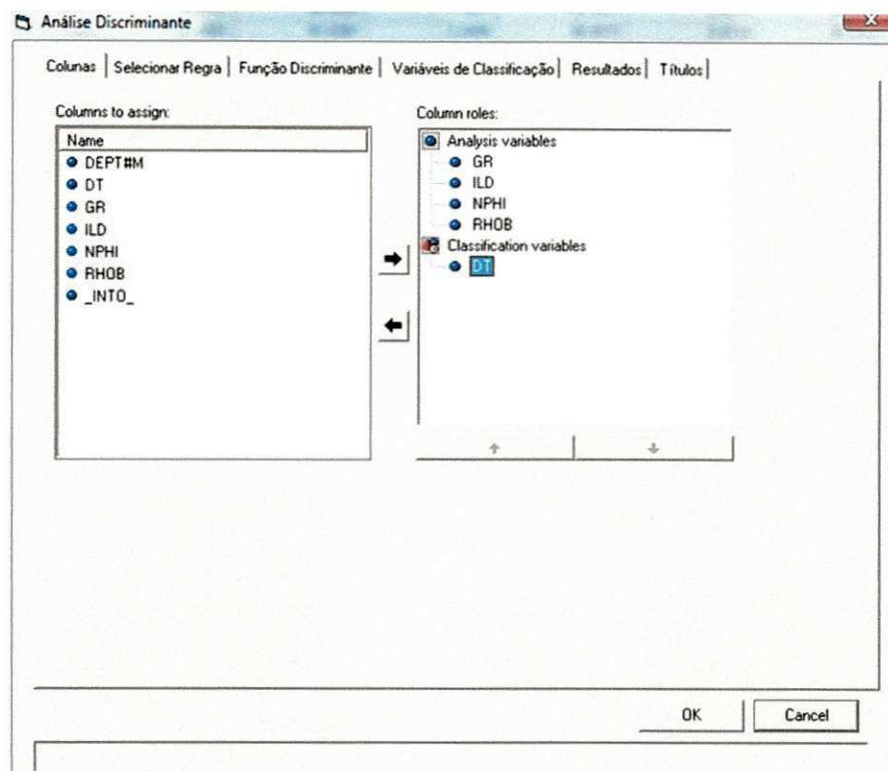


Figura 5- Tela do SAS Enterprise Guide® para seleção do perfil a estimar a partir dos perfis de análise.

A janela do programa SAS Enterprise Guide®, apresentada na Figura 6, permite realizar a escolha do método estatístico a ser empregado na estimativa do perfil faltante. No caso da escolha do método KNN, os parâmetros definidos pelo usuário são o valor de K (número de vizinhos no espaço n-dimensional) e o tipo de métrica utilizada para determinação das distâncias entre as n amostras no espaço investigado.

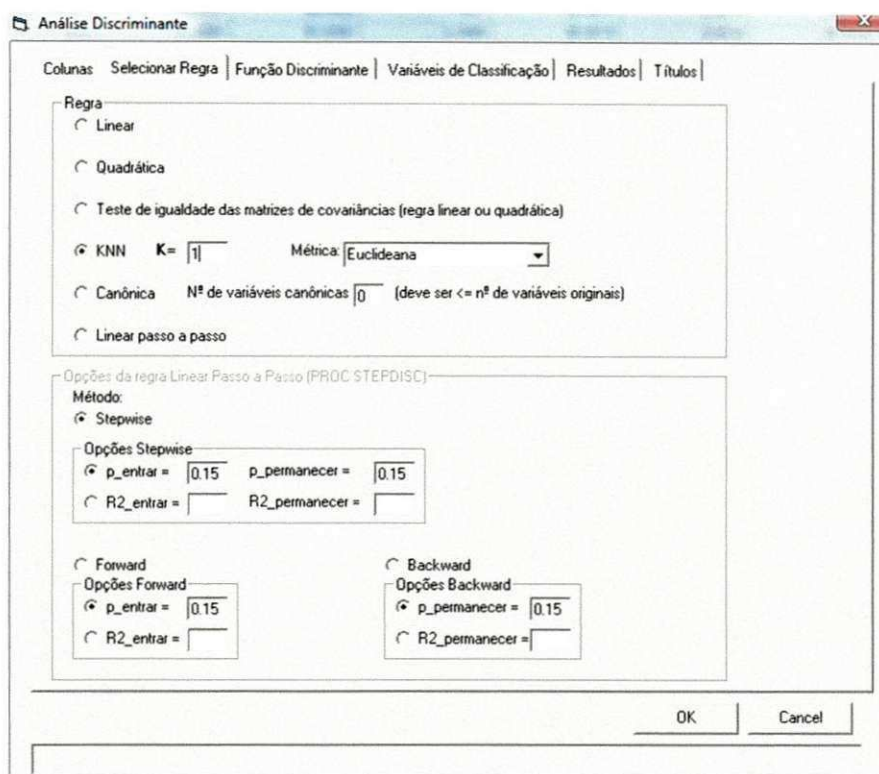


Figura 6 - Tela do SAS Enterprise Guide® para seleção do método estatístico de análise.

A Figura 7 apresenta a tela do programa SAS Enterprise Guide®, onde o usuário tem a possibilidade de adotar diversas opções relacionadas à implementação da função discriminante e do padrão de probabilidades *a priori*, para a ocorrência de uma dada classificação. A escolha da opção de probabilidades iguais significa que se considera o evento como aleatório, ou seja, a probabilidade de ocorrência de um dado valor é, *a priori*, a mesma que qualquer outro valor.

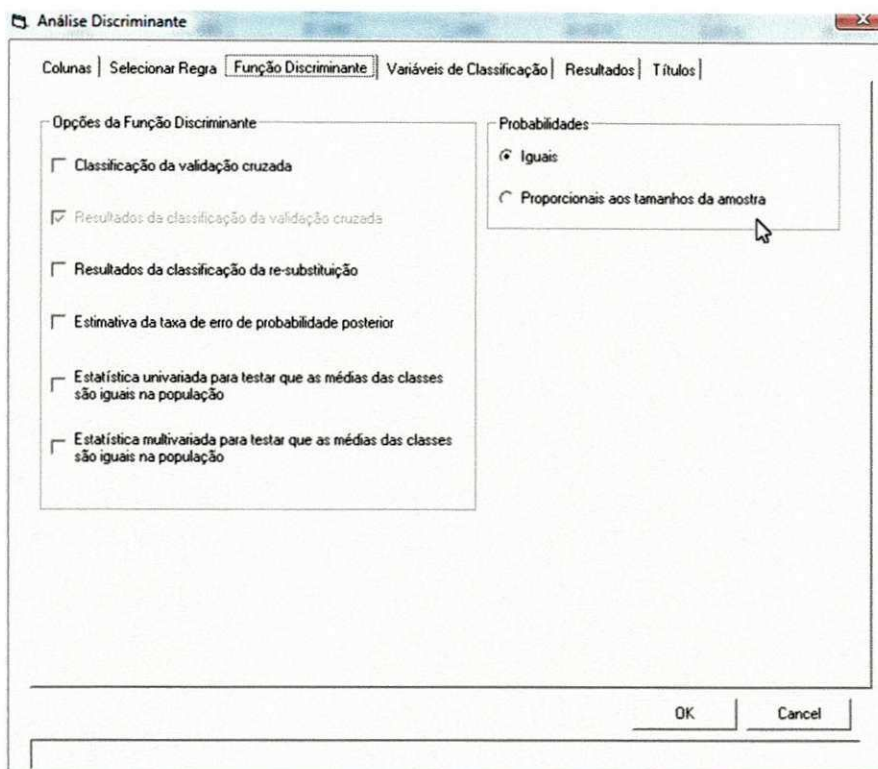


Figura 7 - Tela do *SAS Enterprise Guide*[®] para definição dos parâmetros da função discriminante.

Para gerar um dado perfil geofísico faltante, com base nas propriedades físicas registradas nos demais perfis de um poço, é preciso definir qual o arquivo que contém os dados para classificação (perfis disponíveis), como também o arquivo de saída, o qual deve conter uma coluna com os valores estimados para o perfil faltante. A Figura 8 apresenta a tela do programa *SAS Enterprise Guide*[®] que permite ao usuário a definição destes parâmetros.

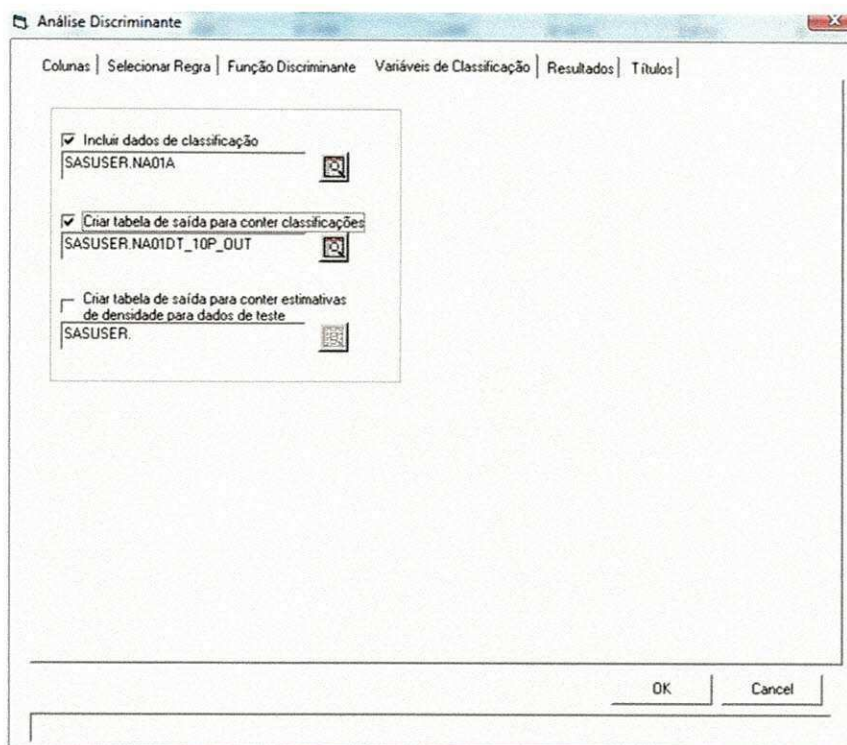


Figura 8 - Tela do SAS Enterprise Guide® para incluir os dados de classificação e gerar arquivo com os dados de saída (dados estimados).

A Figura 9 apresenta o arquivo com a coluna “_INTO_”, a qual contém os valores estimados para o perfil faltante. No exemplo apresentado nesta figura, a coluna “_INTO_” procura reproduzir os valores da curva DT. Observa-se nesta figura a proximidade dos valores entre a coluna de dados realmente medida (coluna DT) e a coluna estimada considerando-se a ausência do perfil DT (coluna _INTO_).

na01a_dt_10p100 (read-only)

	DEPT#M	DT	GR	ILD	NPHI	RHOB	_INTO_
1	2952.4	95.4766	74.375	1.1445	25.7249	2.4929	92.0547
2	2952.6	94.8945	78.6992	1.1443	25.7249	2.448	87.9492
3	2952.8	92.9727	80.0625	1.1831	25.7617	2.4355	87.9492
4	2953	91.8281	81.0391	1.2468	26.0015	2.4119	87.9492
5	2953.2	91.0703	81.8203	1.3105	25.3672	2.4176	87.9492
6	2953.4	90.0586	83.1836	1.3596	25.7031	2.4678	92.1445
7	2953.6	88.6602	83.6914	1.4053	26.9023	2.4412	100.5276
8	2953.8	88.1382	82.2656	1.4685	25.9531	2.4116	92.1445
9	2954	87.9492	79.0586	1.5327	24.082	2.4775	87.9492
10	2954.2	88.5703	78.6836	1.573	24.3438	2.5206	87.9492
11	2954.4	87.6211	81.9766	1.603	24.8853	2.522	92.1445
12	2954.6	85.3672	84.875	1.6511	23.9453	2.5193	92.1445
13	2954.8	84.1523	85.9375	1.7231	23.0938	2.4961	81.8984
14	2955	84.1436	87.3555	1.782	22.2891	2.4873	81.8984
15	2955.2	84.7383	87.832	1.8059	21.3594	2.4864	83.5625
16	2955.4	85.9844	85.3711	1.8014	21.1084	2.4836	89.9375
17	2955.6	88.4102	83.5508	1.7815	21.3086	2.4863	92.1445
18	2955.8	89.8047	84.9375	1.7536	21.8477	2.4818	92.1445
19	2956	87.6602	89.4258	1.7509	22.3123	2.4978	83.5625
20	2956.2	85.707	93.1797	1.7665	21.668	2.5017	83.5625
21	2956.4	85.2263	95.4766	1.7705	20.2461	2.4971	83.5625
22	2956.6	86.8789	96.0195	1.7415	19.8477	2.5271	83.5625
23	2956.8	88.7461	93.1875	1.6846	19.6641	2.5404	83.5625
24	2957	92.5625	87.5	1.6245	19.9219	2.5149	83.5625
25	2957.2	94.6172	84.4492	1.5901	21.5664	2.5181	92.1445
26	2957.4	91.2305	86.8125	1.5549	23.0547	2.527	81.8984
27	2957.6	90.0657	91.1797	1.5186	22.457	2.5518	83.5625

Figura 9 - Tela do SAS Enterprise Guide® com o arquivo de saída, o qual inclui a coluna com o perfil faltante estimado (coluna em destaque).

2.2 - METODOLOGIA

Para montar a base de dados deste trabalho foram escolhidos os 12 poços do Campo de Namorado citados anteriormente, e as variáveis utilizadas foram todos os perfis disponíveis (DT, GR, ILD, NPHI e RHOB). O critério adotado para a escolha dos poços foi o de incluir apenas os poços verticais, excluído os direcionais, os quais precisam ser interpretados de forma não convencional, e por isso foram excluídos da base de dados deste trabalho.

Inicialmente foi feita a concatenação dos perfis de todos os poços em um único arquivo, totalizando aproximadamente onze mil registros. Esta fusão foi realizada com o objetivo de permitir que o processo de amostragem aleatória, para fins de geração do conjunto de dados de treinamento, tenha a possibilidade de escolher amostras representativas de todas as formações geológicas que ocorrem em quaisquer dos poços analisados. A falta das características representativas de uma formação no conjunto de dados de treinamento pode induzir a regra discriminante a não ser capaz de reconhecer tal valor daquela propriedade durante a fase de estimação. Deste modo, a amostragem realizada sobre um arquivo que contém os perfis de todos os poços minimiza essa possibilidade, reduzindo o vício da amostra. Neste trabalho uma amostra é entendida com sendo o conjunto de valores dos vários perfis geofísicos numa dada profundidade de um poço.

Posteriormente, utilizando o programa *SAS Enterprise Guide*[®] foram feitas amostragens aleatórias, sobre o arquivo concatenado, nas proporções de 10%, 20%, 30% e 50%, com o objetivo de avaliar o efeito do tamanho da amostra de treinamento sobre a eficiência da estimativa de perfis faltantes. Na realidade, neste trabalho não há perfis faltantes (exceto a curva DT do poço NA-12), mas essa situação foi simulada para todos os perfis de todos os poços, excluindo a cada simulação um determinado perfil da base de dados e utilizando os demais perfis como dado de entrada para a estimação daquele perfil "faltante". Após cada simulação, a eficiência deste processo é medida através do cálculo da diferença RMS (*Root Mean Square*) entre o perfil real e o simulado. Foram também gerados gráficos para comparação visual entre os perfis reais e simulados para cada um dos poços analisados.

$$Dif_{RMS} = \sqrt{\frac{\sum_{i=1}^m (Real_i - Estimado_i)^2}{n-1}} \quad (4)$$

A Figura 10 apresenta o fluxograma que descreve o método utilizado para a obtenção dos conjuntos de dados de treinamento utilizados neste trabalho; enquanto que a Figura 11 exibe o fluxograma referente ao processo de geração e aplicação das regras discriminantes para a estimação dos perfis “faltantes”, análise da eficiência das estimativas e visualização gráfica dos resultados.

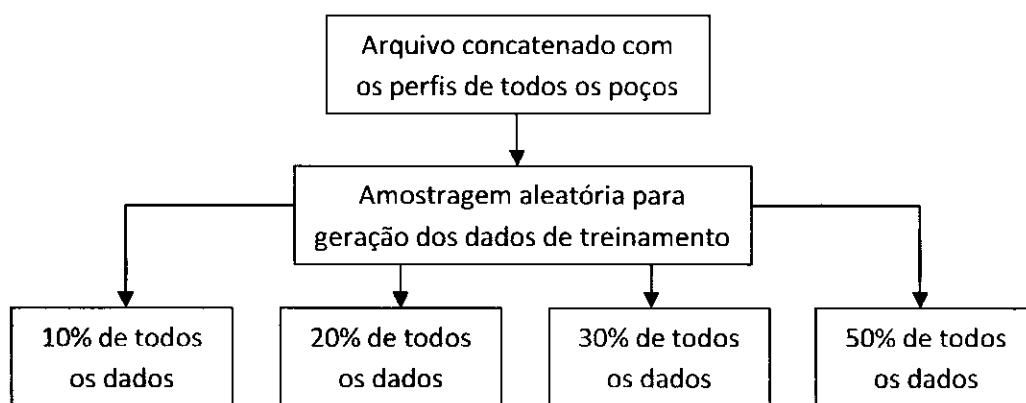


Figura 10 – Fluxograma para obtenção dos conjuntos de dados de treinamento.

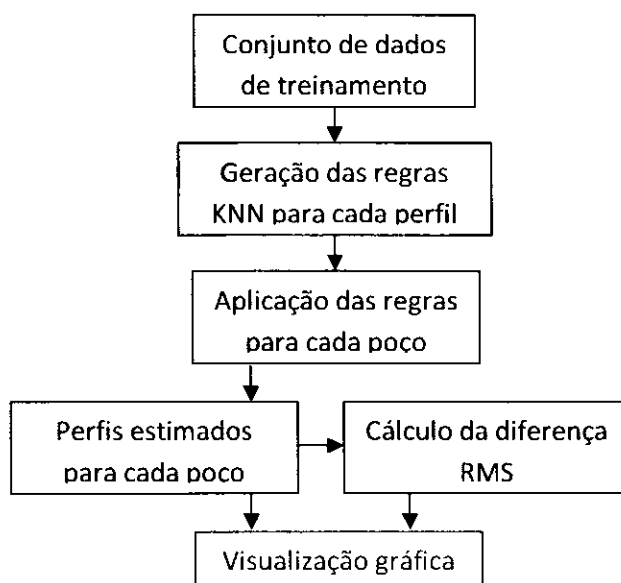


Figura 11 – Fluxograma para aplicação das regras discriminantes e estimação dos perfis faltantes.

Neste trabalho, para a implementação da regra KNN, foi utilizado $K=1$. Embora para valores de K muito pequenos essa técnica tenda a produzir resultados ruidosos, para a aplicação específica da técnica em perfis geofísicos de poços essa característica fica bastante atenuada pelo fato de se dispor de conjunto de dados com elevado número de amostras. Por outro lado, a adoção de um valor elevado para K induziria a suavização excessiva nos resultados, além de aumentar consideravelmente o custo computacional.

CAPÍTULO III

3.1 - RESULTADOS

Os resultados desta dissertação estão organizados por tipo de perfil e por poço.

3.1.1 - Perfil Sônico (DT)

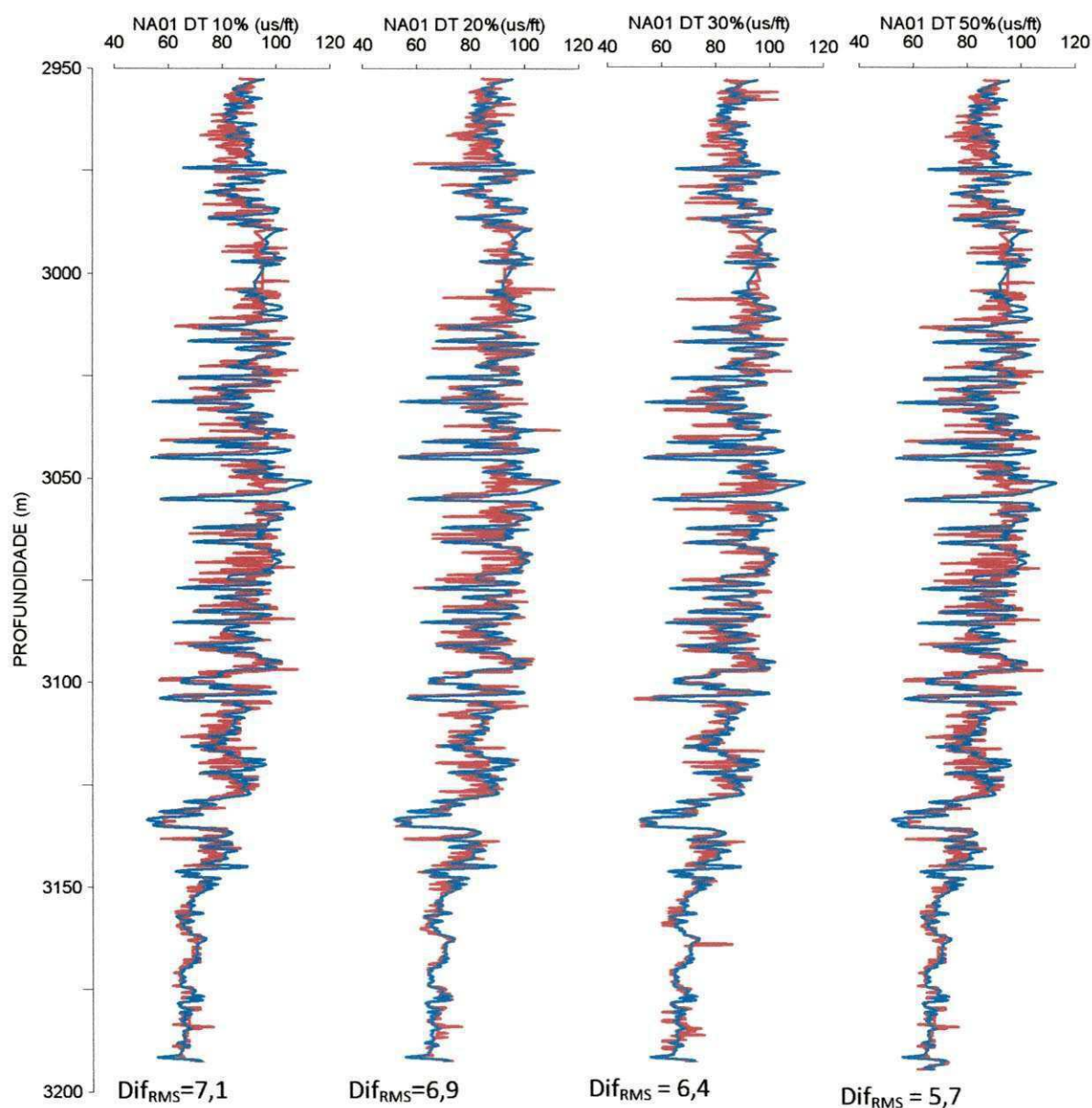


Figura 12 – Perfis sônicos do poço NA-01a: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

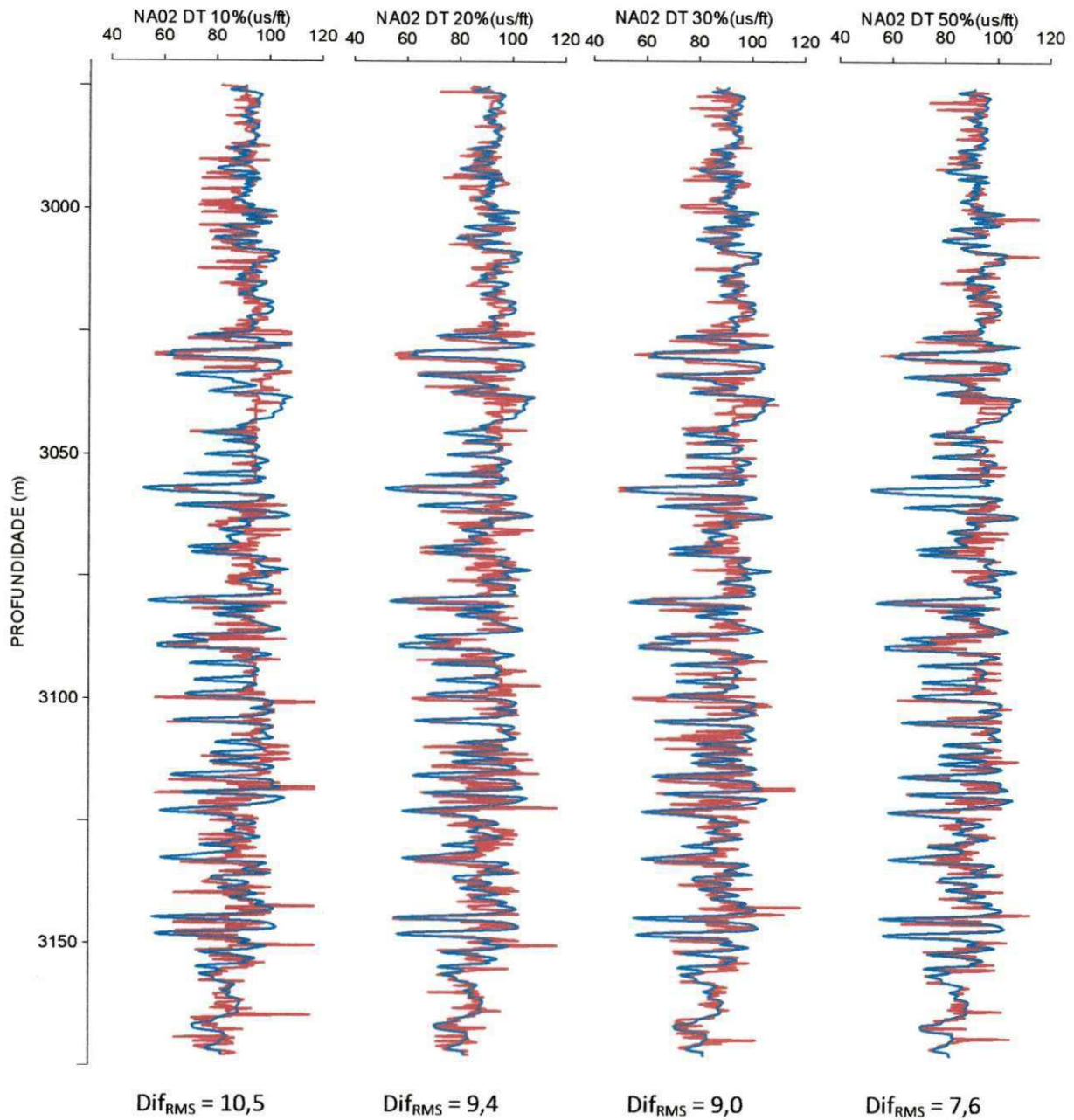


Figura 13 – Perfis sônicos do poço NA-02: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

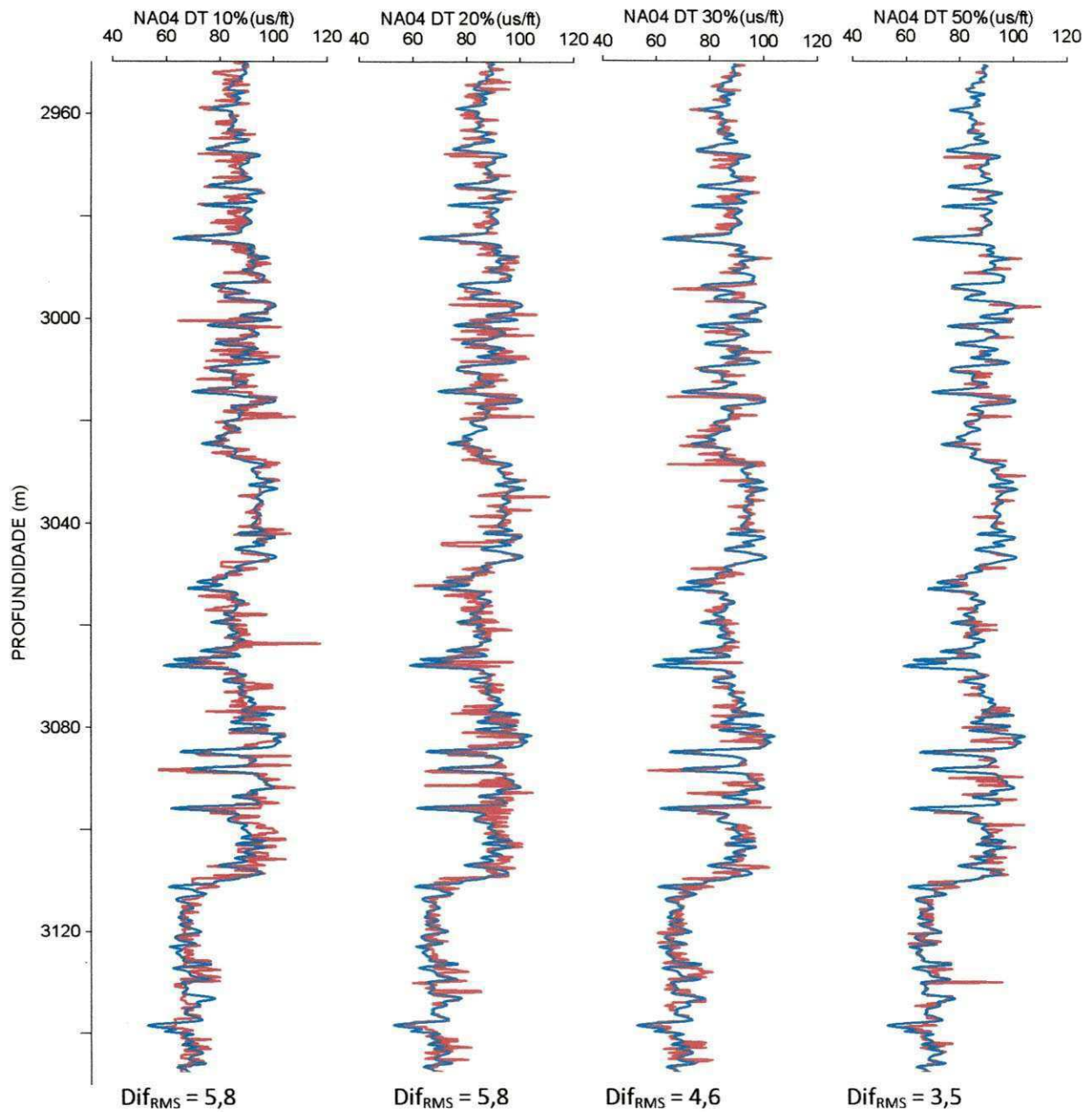


Figura 14 – Perfis sônicos do poço NA-04: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

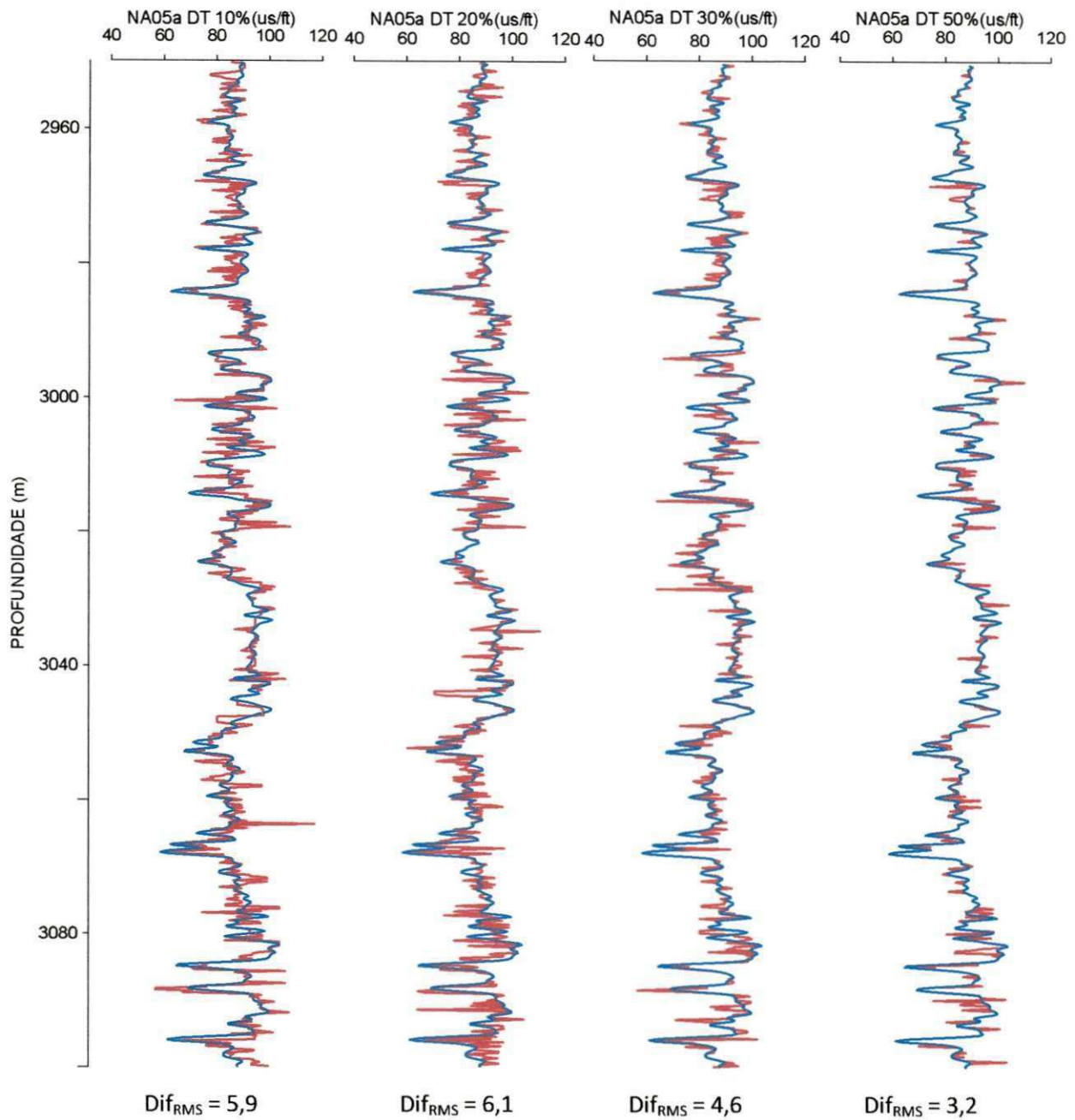


Figura 15 – Perfis sônicos do poço NA-05a: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

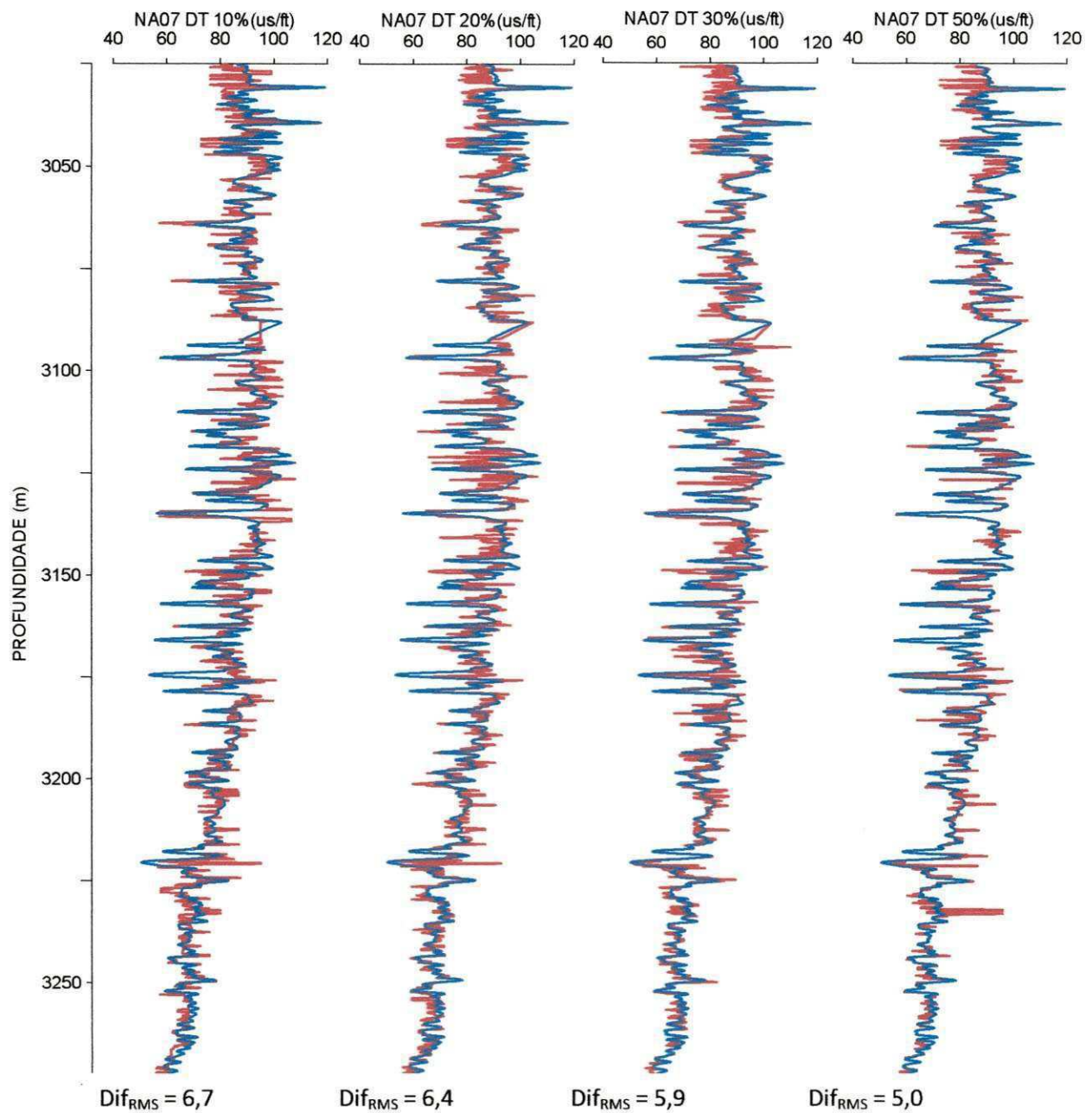


Figura 16 – Perfis sônicos do poço NA-07: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

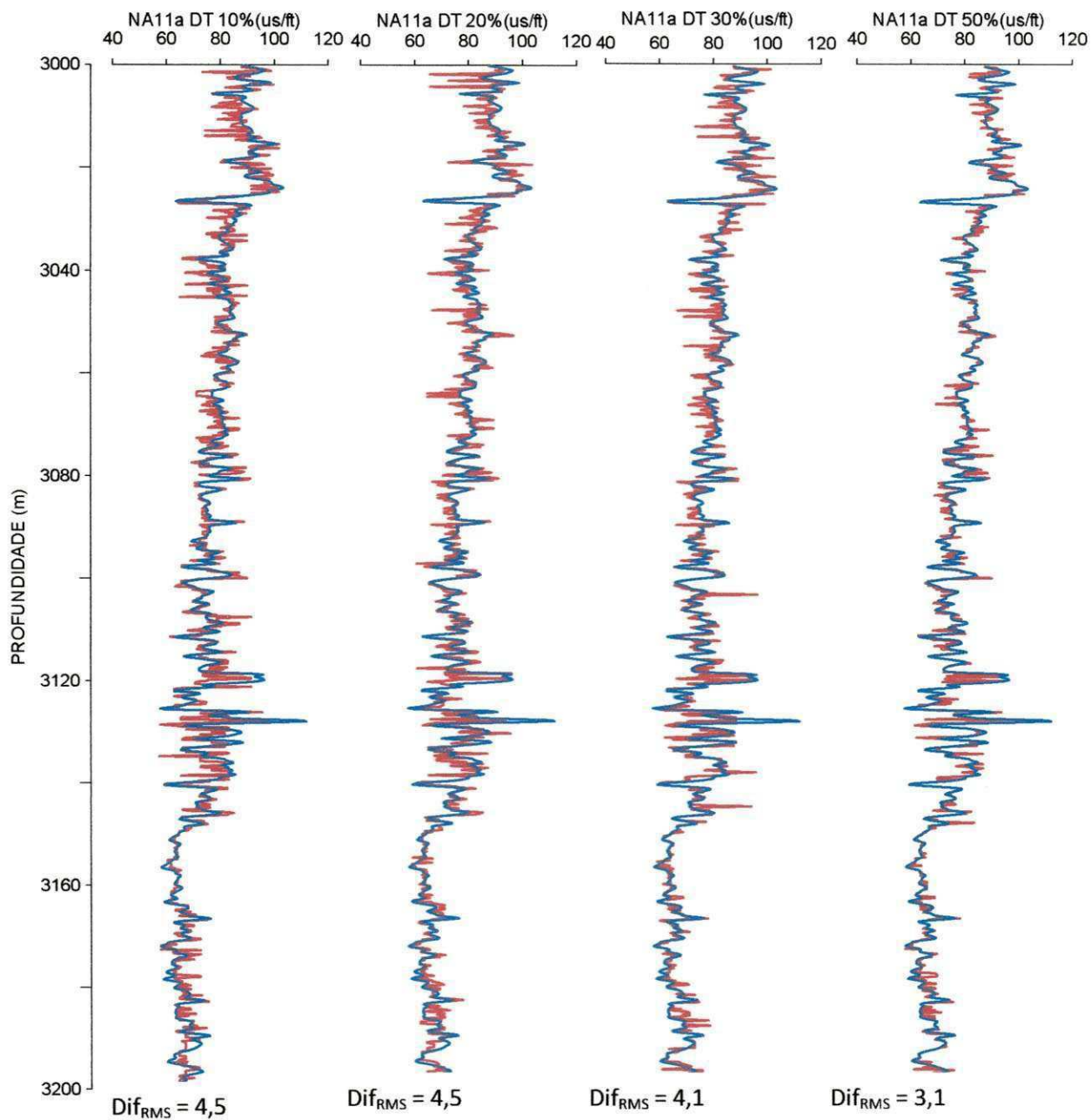


Figura 17 – Perfis sônicos do poço NA-11a: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

O perfil DT da Figura 18 apresenta o esboço esperado para o perfil sônico e que provavelmente descreve este perfil faltante com boa precisão, embora não se possa quantificar essa precisão, devido à ausência do perfil real. Como a precisão é satisfatória para os demais poços, supõe-se que também o seja para este.

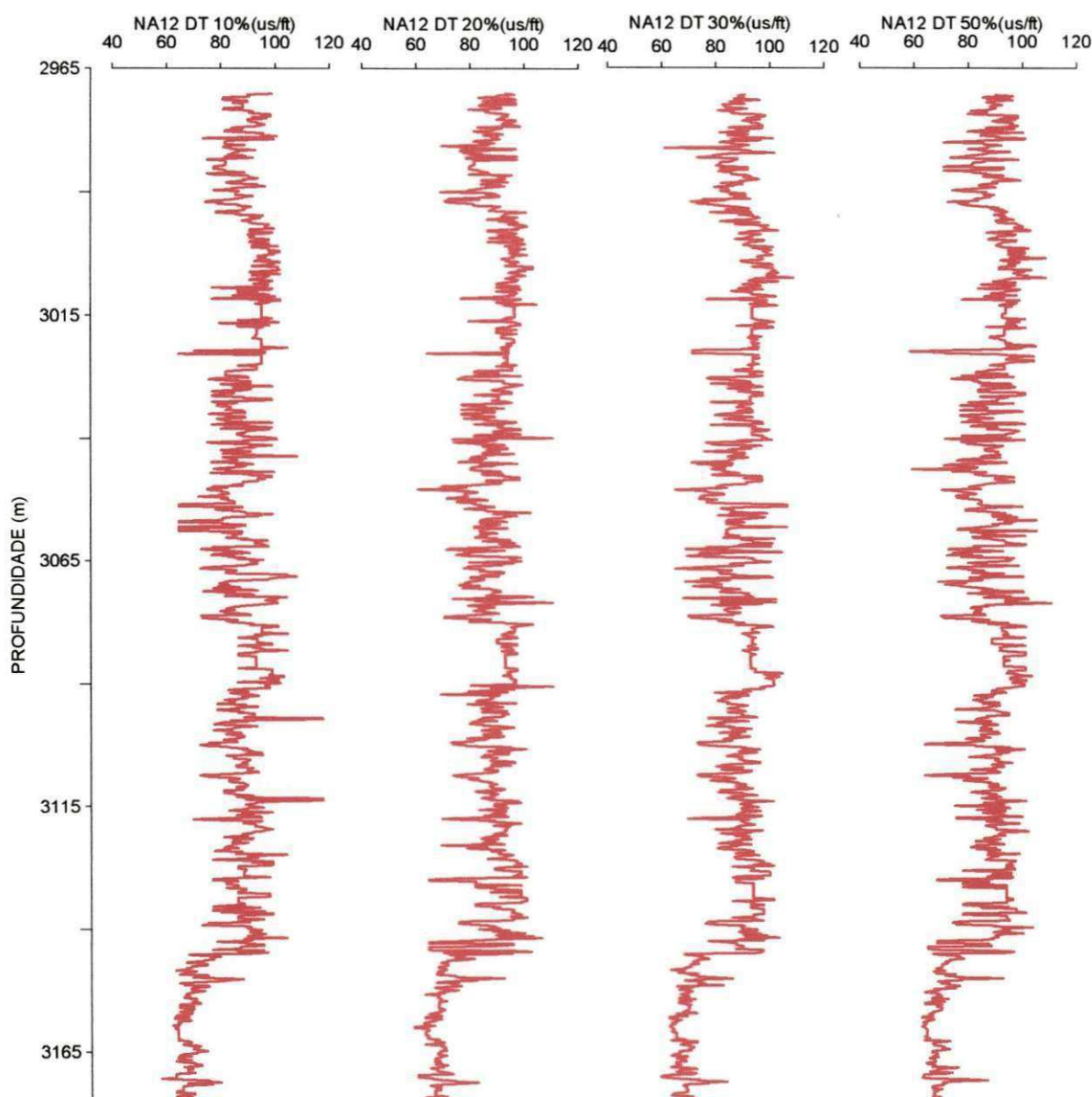


Figura 18 – Perfis sônicos do poço NA-12. Neste poço este perfil real não foi registrado, portanto não se pode estimar a precisão das curvas estimadas. Os perfis apresentados são os estimados para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

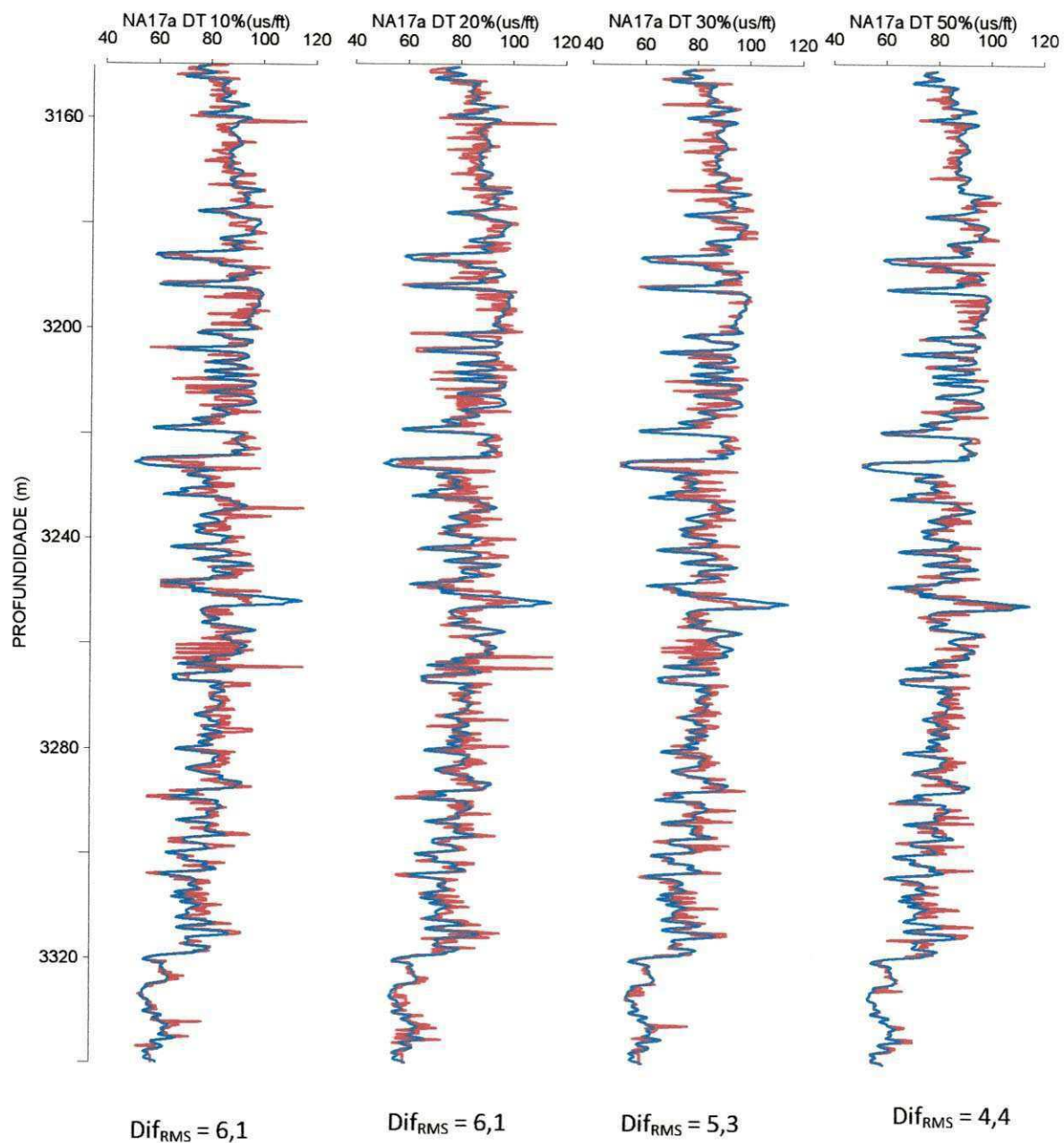


Figura 19 – Perfis sônicos do poço NA-17a: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

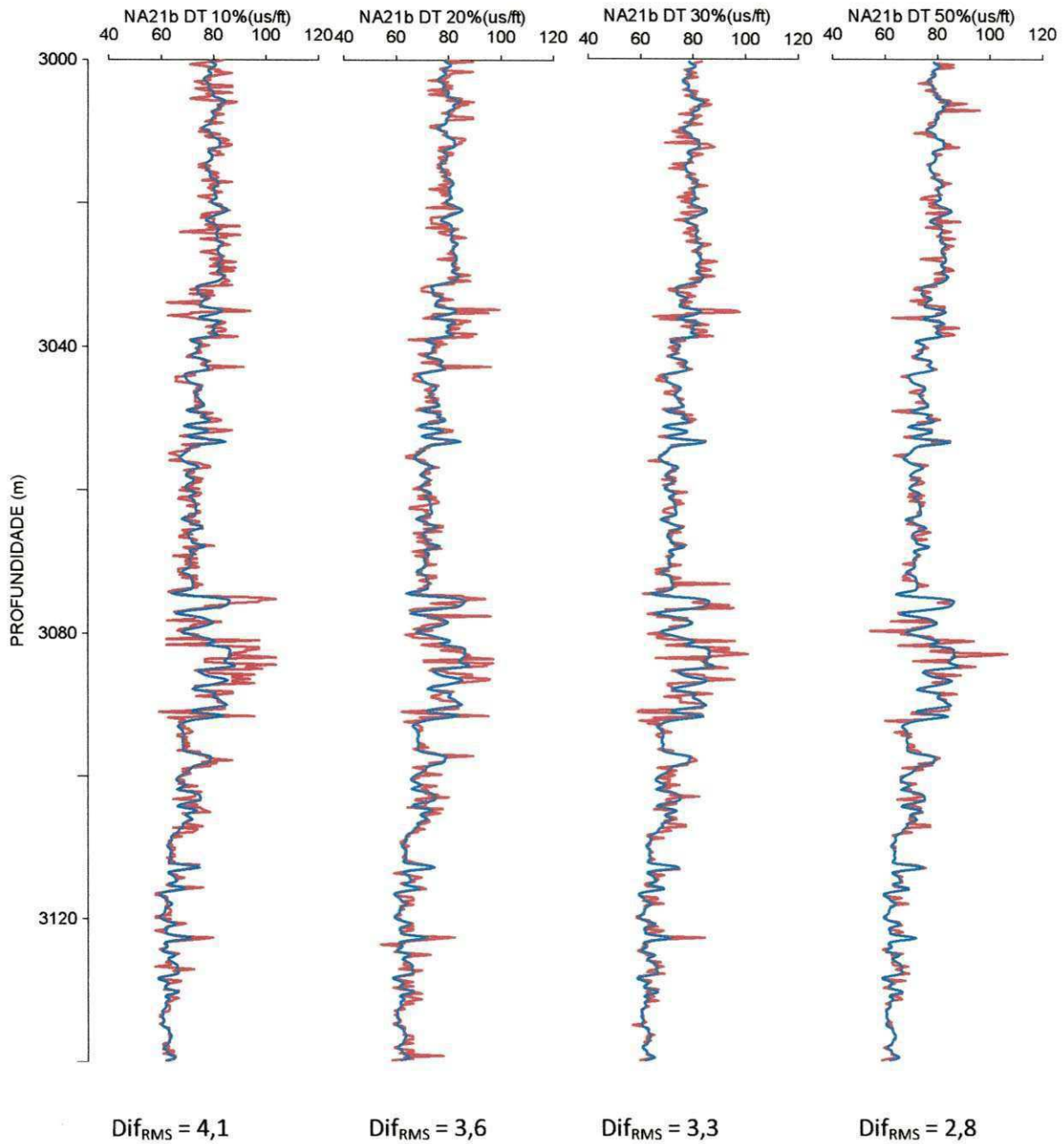


Figura 20 – Perfis sônicos do poço NA-21b: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

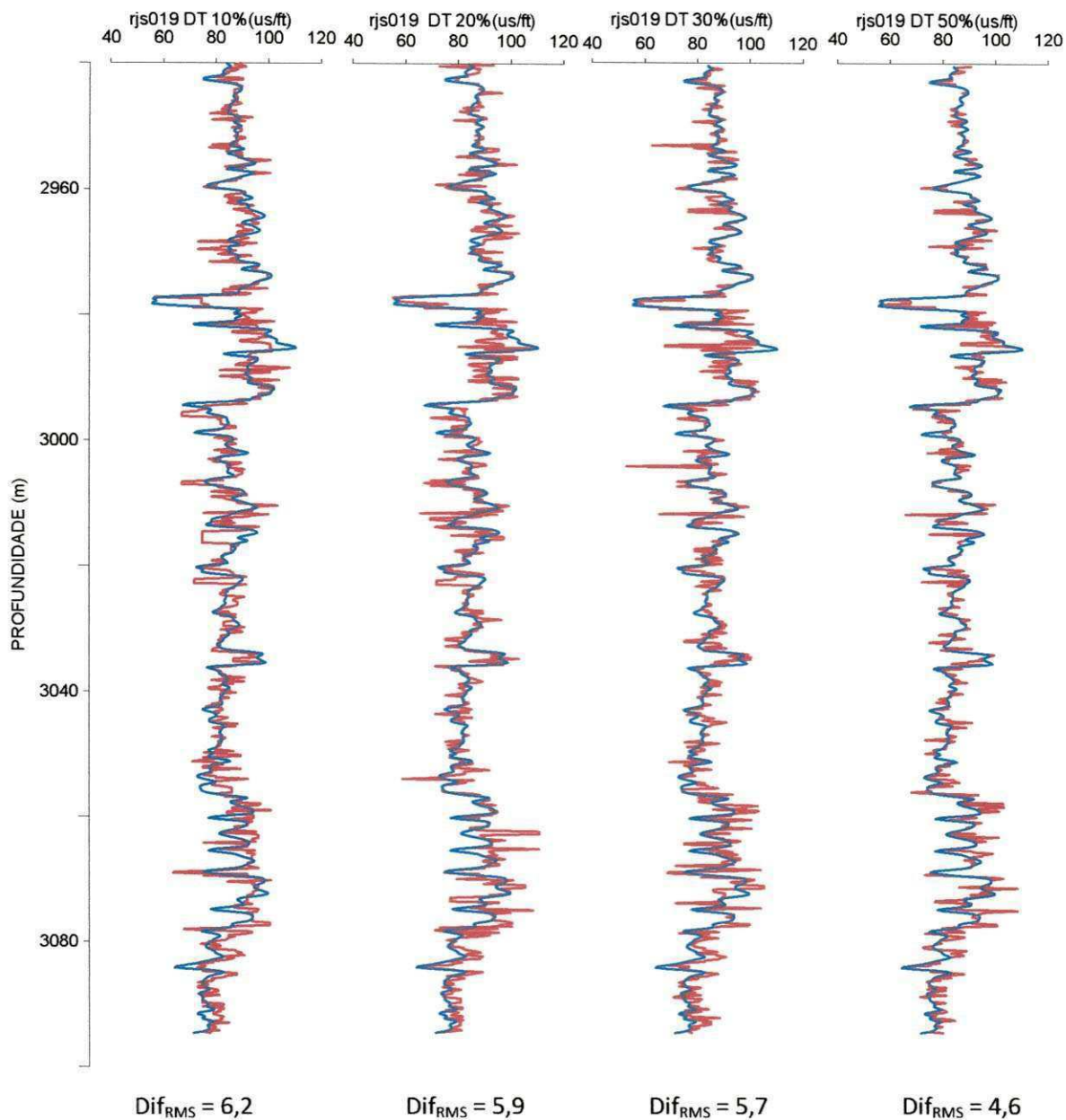


Figura 21 – Perfis sônicos do poço RJS-19: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

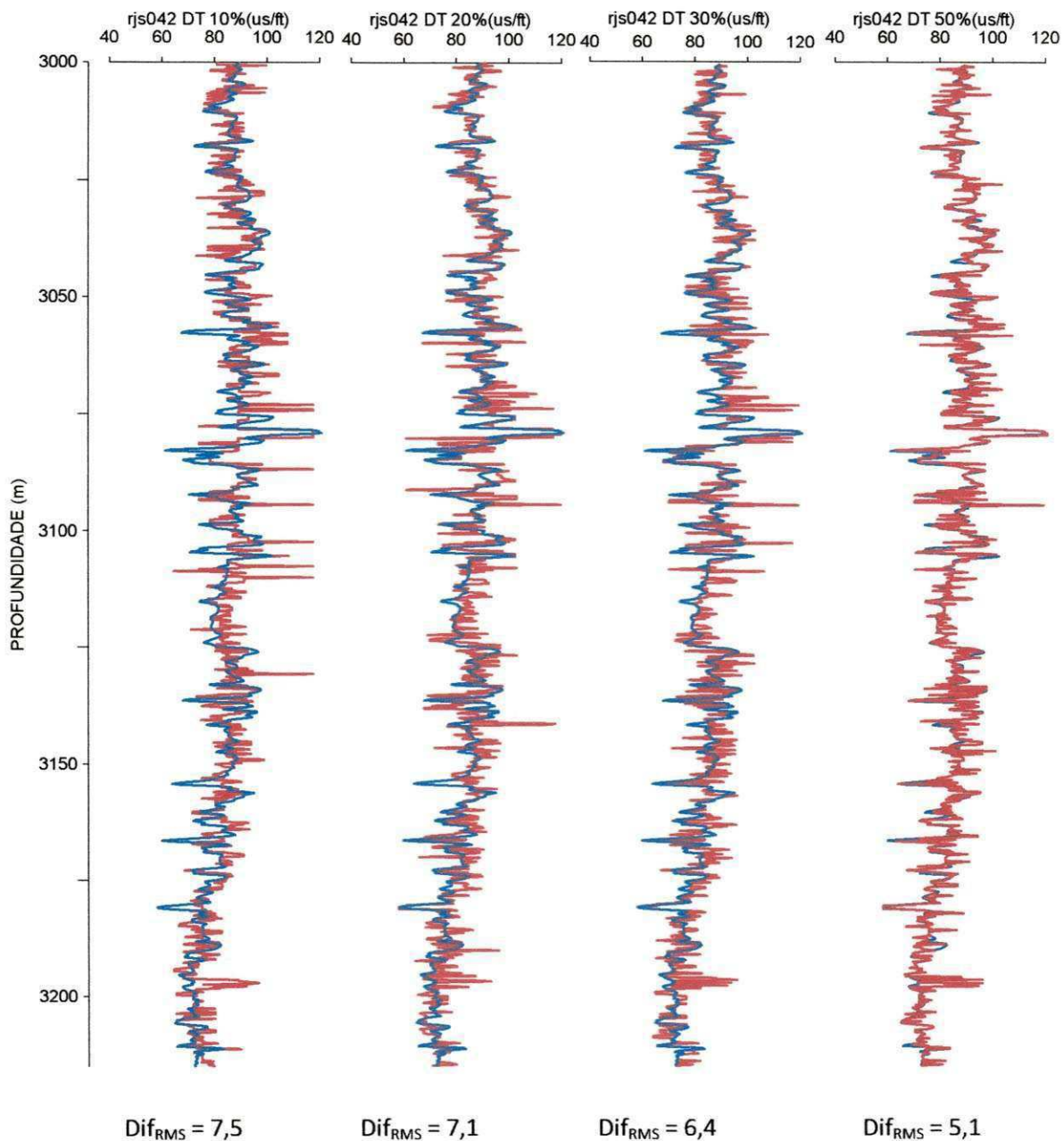


Figura 22 – Perfis sônicos do poço RJS-42: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

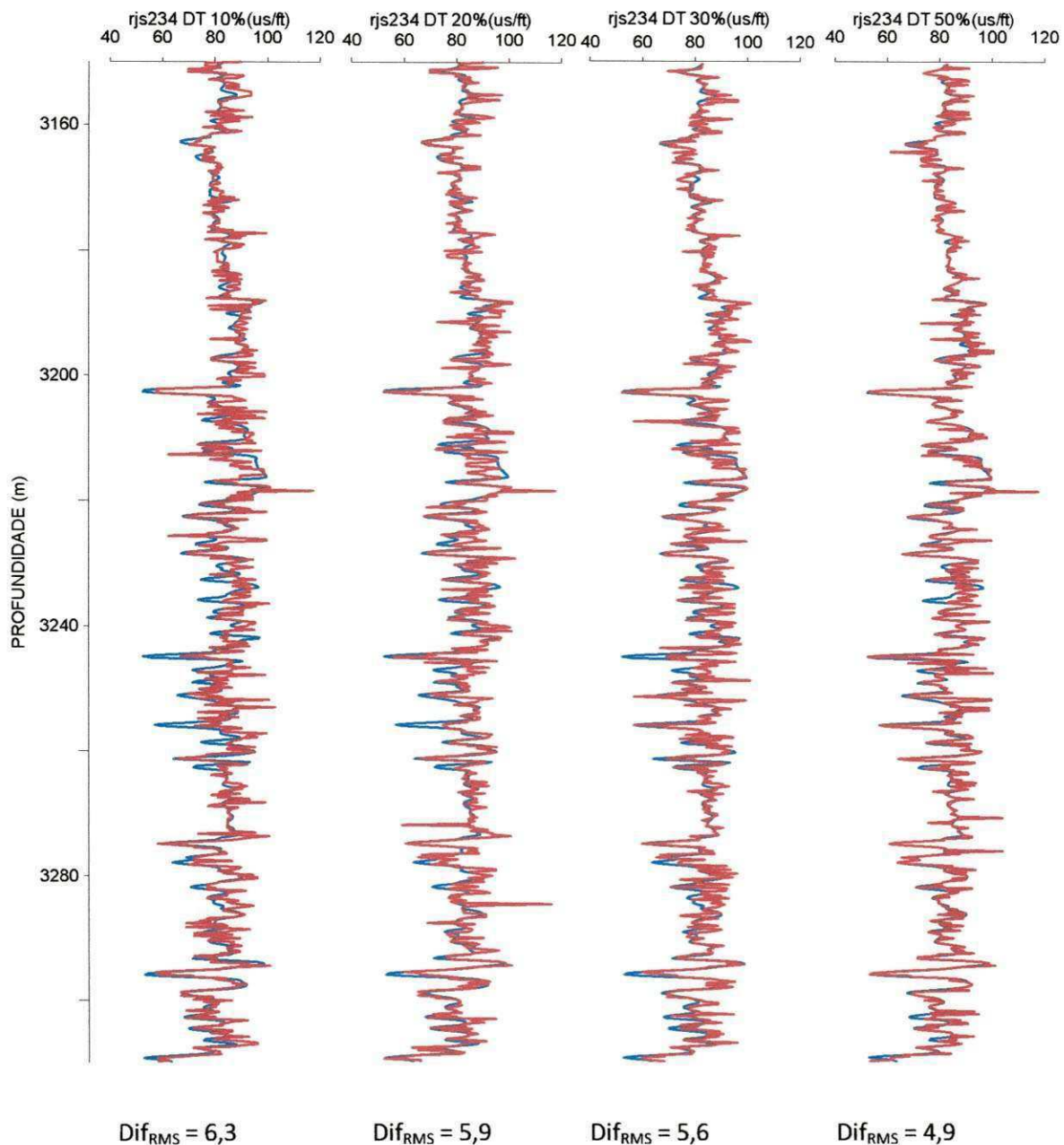


Figura 23 – Perfis sônicos do poço RJS-234: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

3.1.2 - Perfil de Raios Gama (GR)

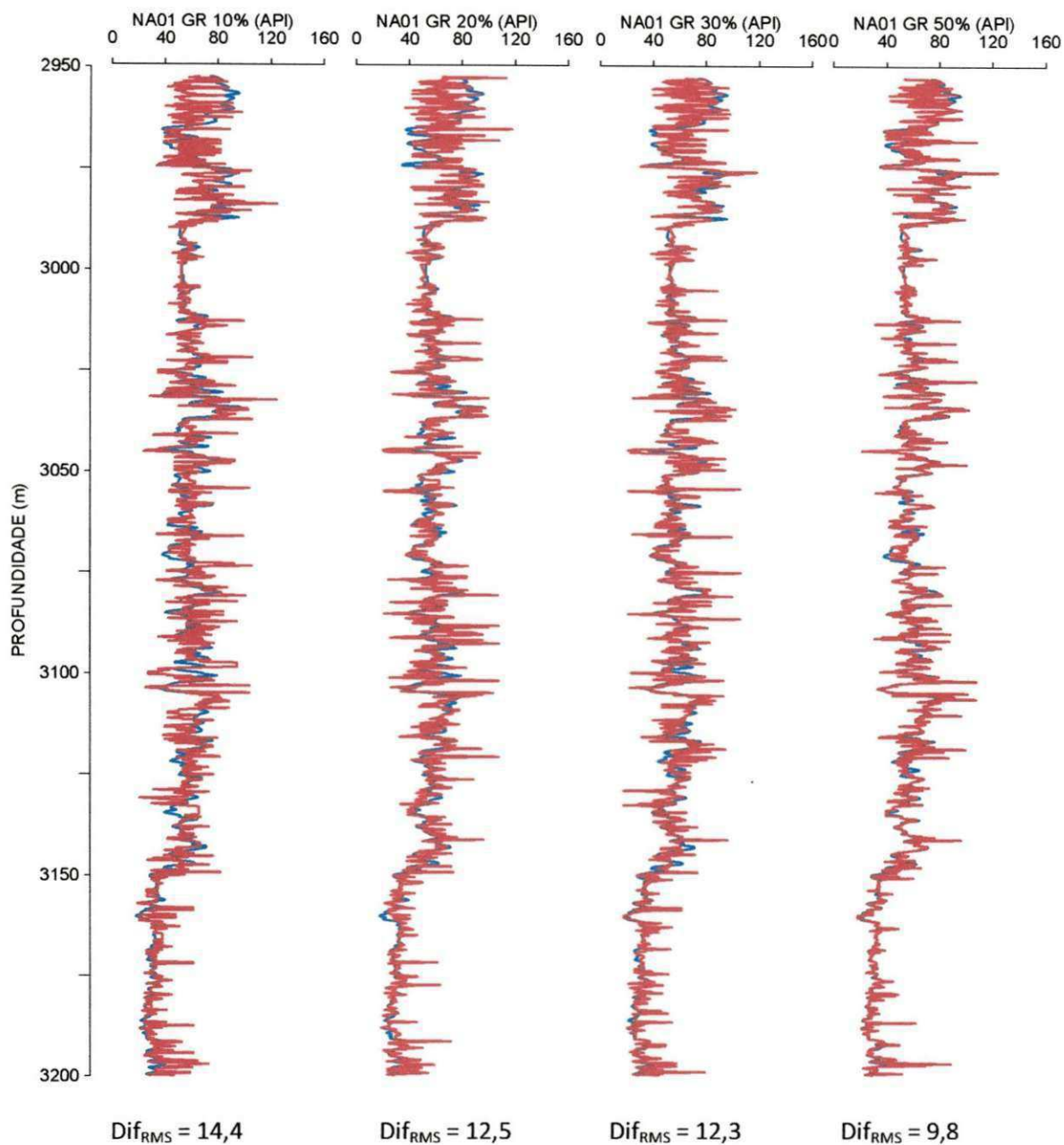


Figura 24 – Perfis de raios gama do poço NA-01: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

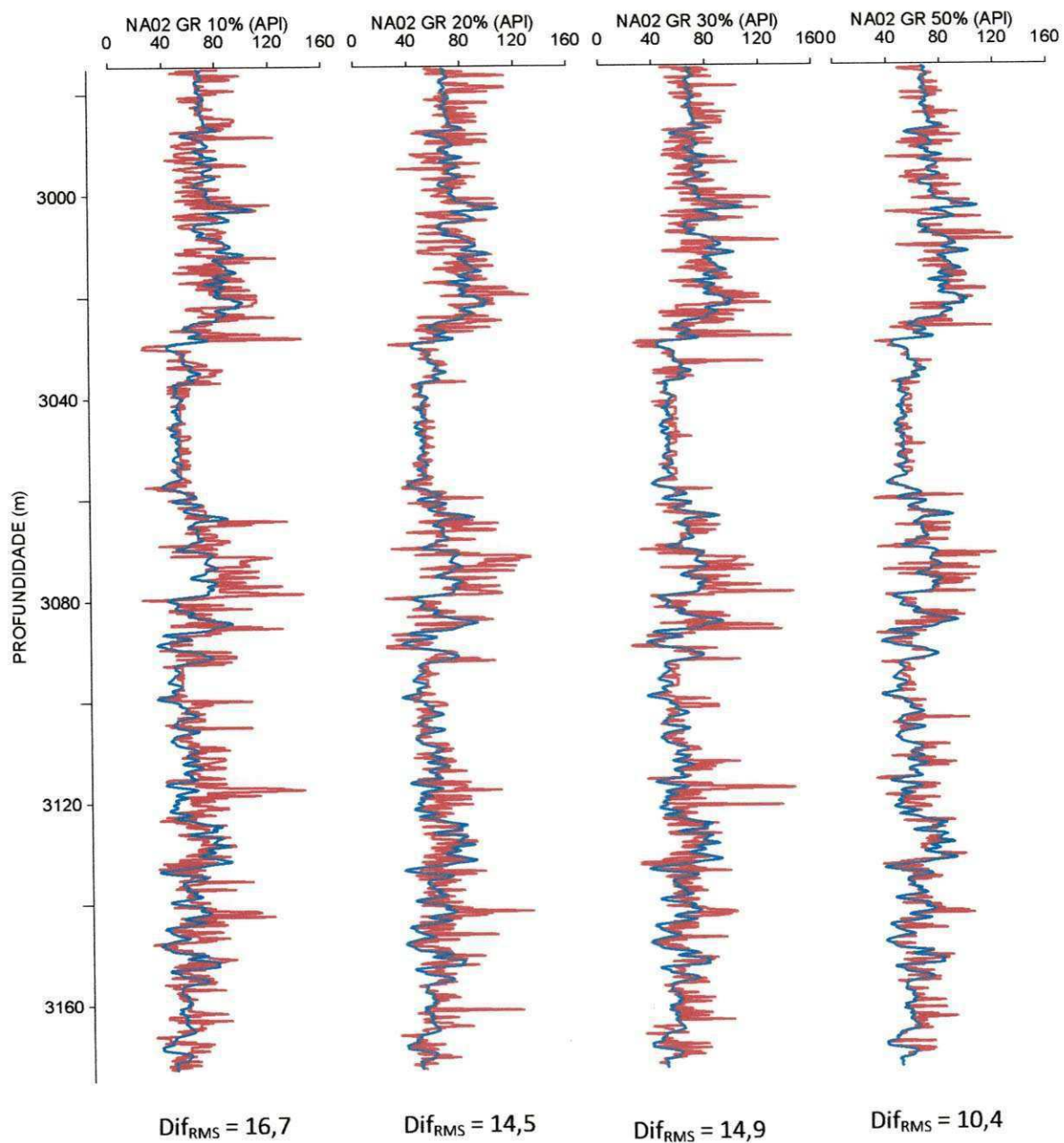


Figura 25 – Perfis de raios gama do poço NA-02: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

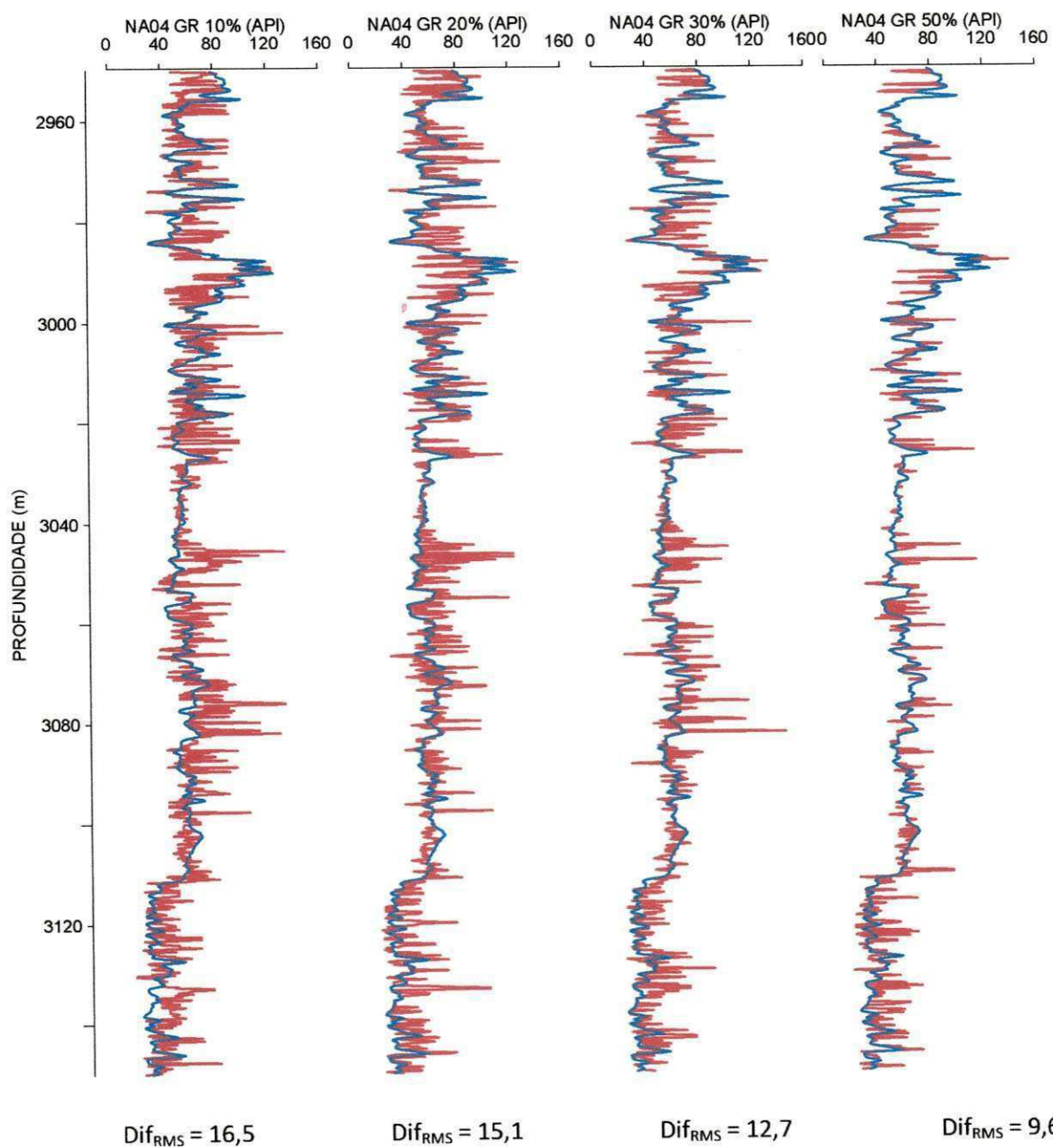


Figura 26 – Perfis de raios gama do poço NA-04: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

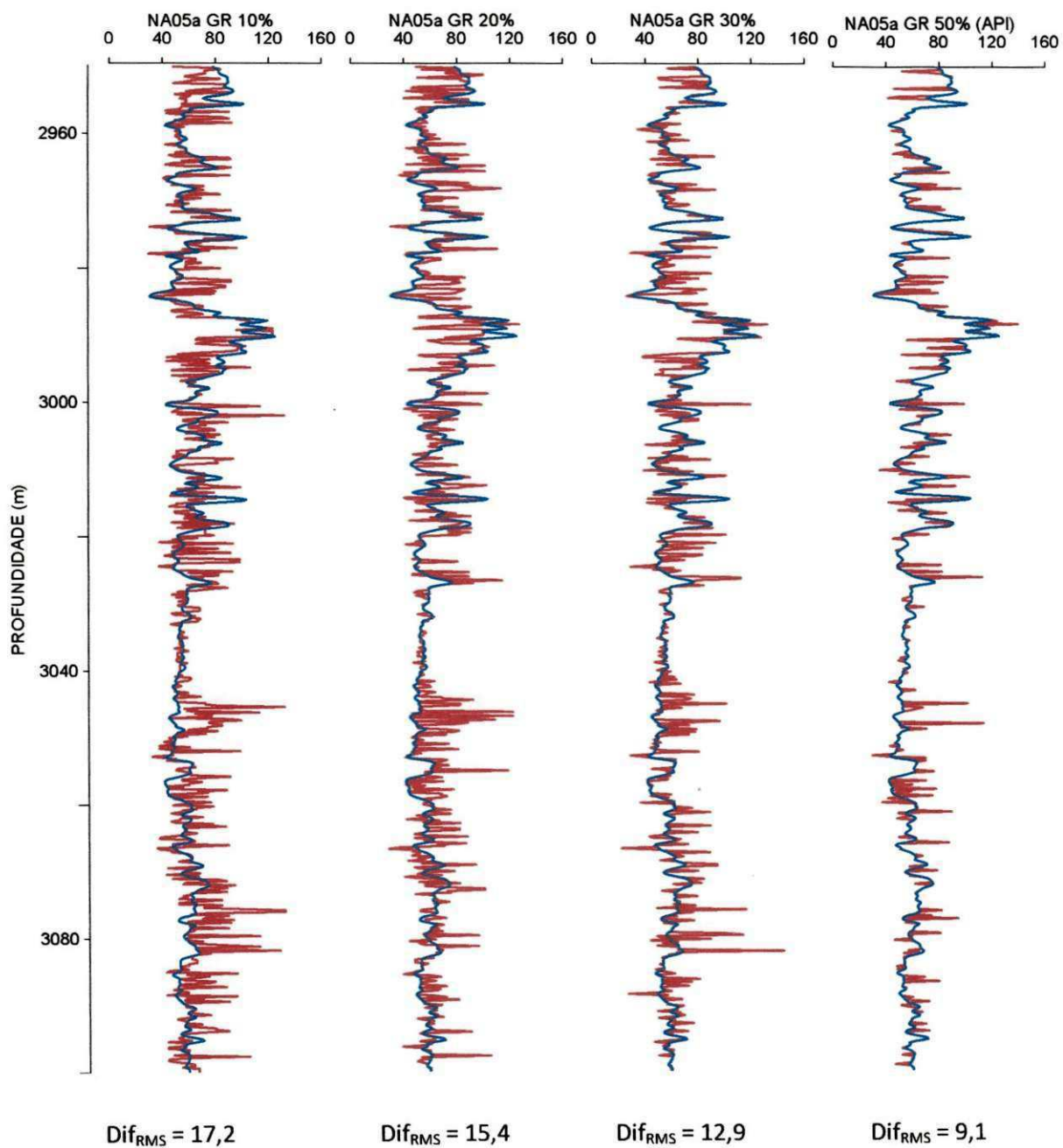


Figura 27 – Perfis de raios gama do poço NA-05a: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

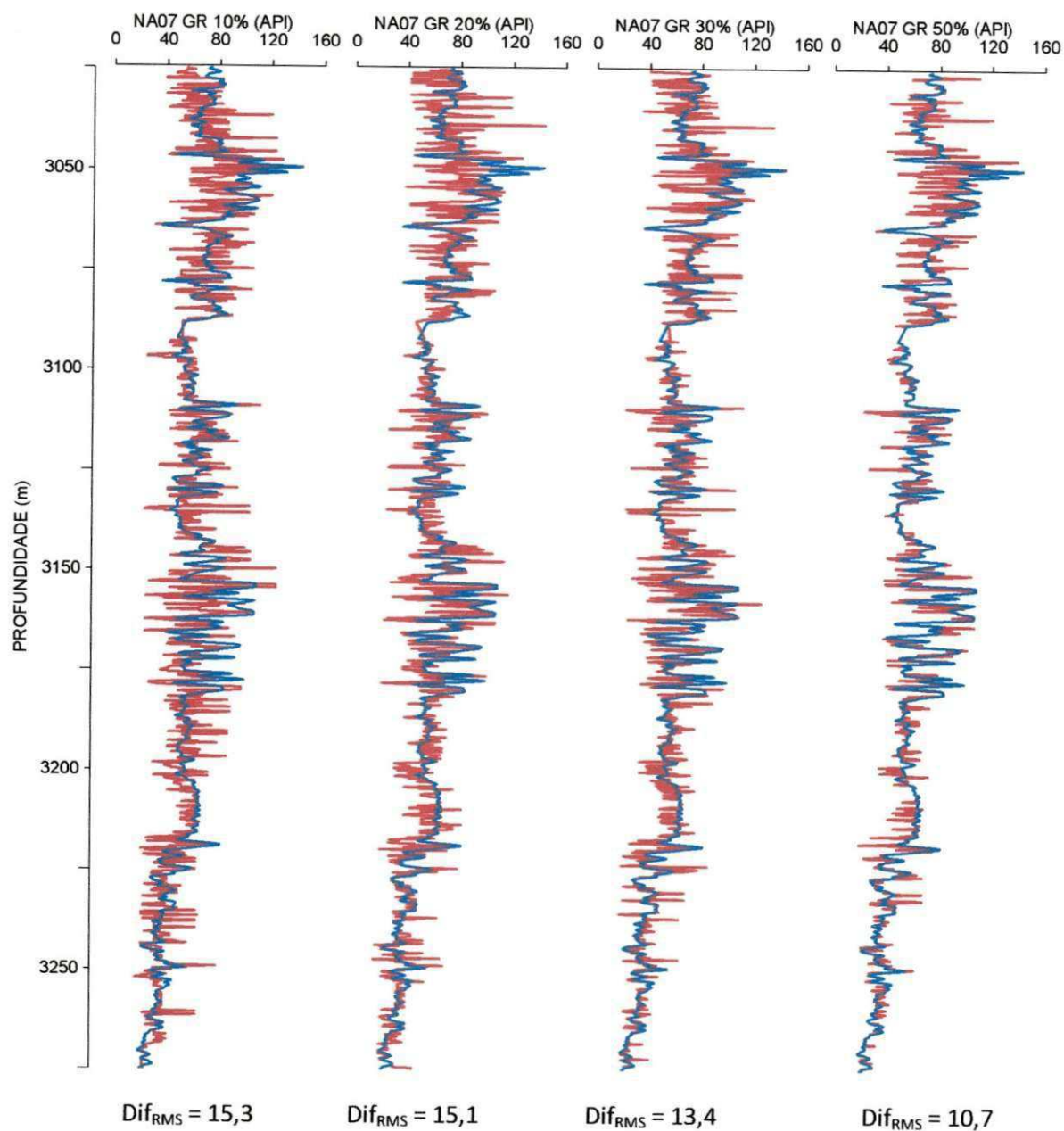


Figura 28 – Perfis de raios gama do poço NA-07: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

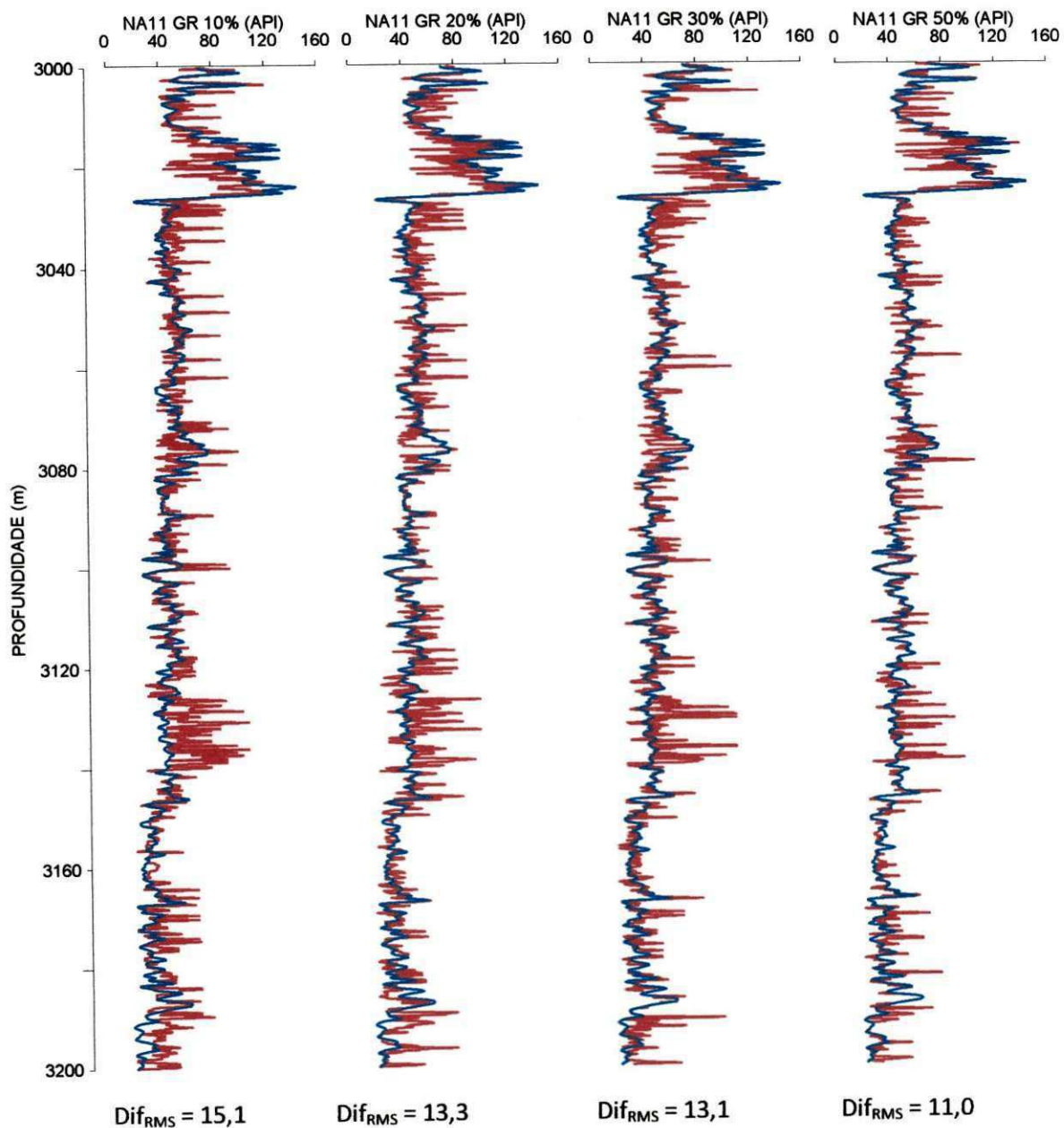


Figura 29 – Perfis de raios gama do poço NA-11: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

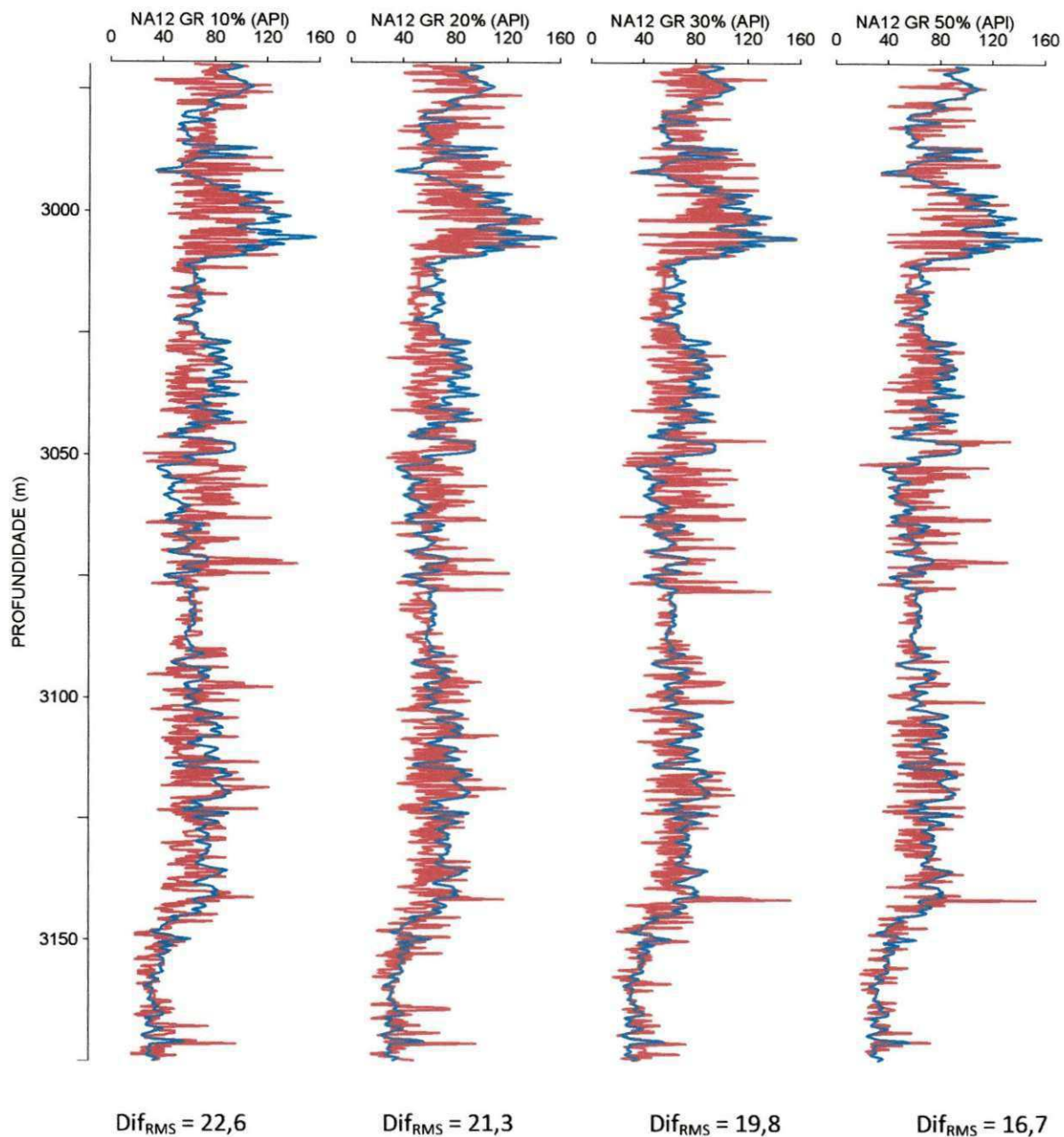


Figura 30 – Perfis de raios gama do poço NA-12: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

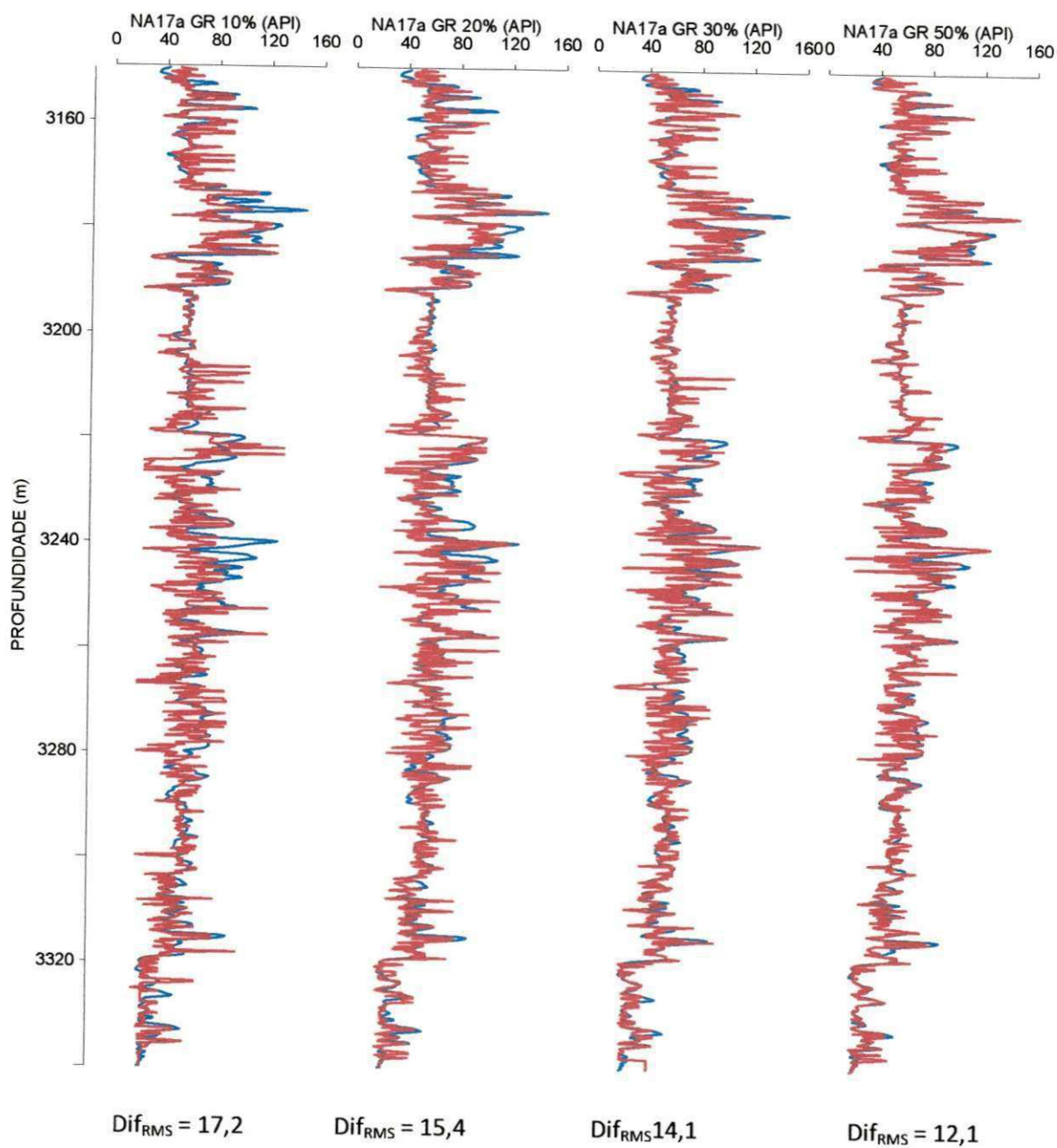


Figura 31 – Perfis de raios gama do poço NA-17a: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

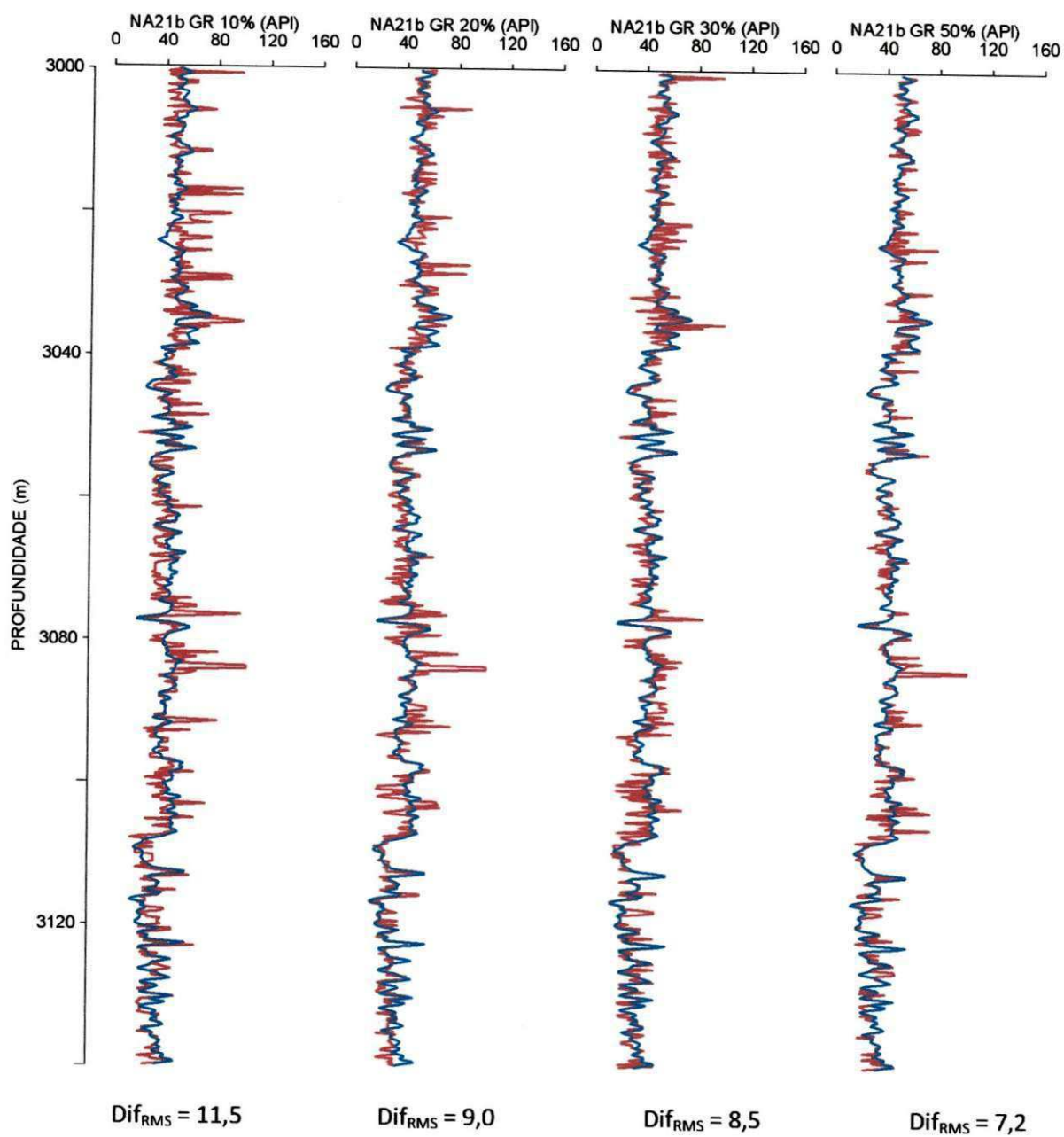


Figura 32 – Perfis de raios gama do poço NA-21b: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

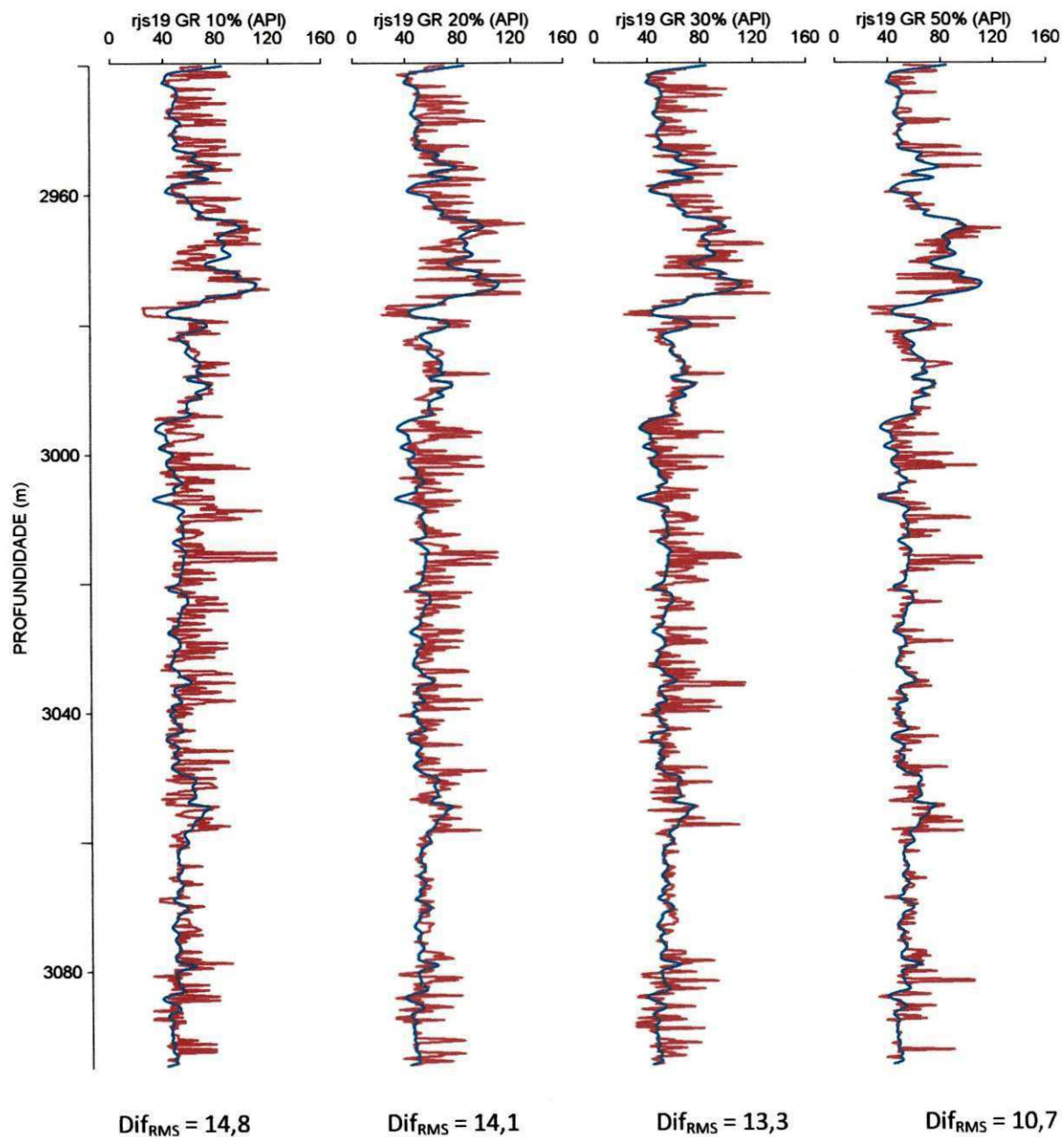


Figura 33 – Perfis de raios gama do poço RJS-19: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

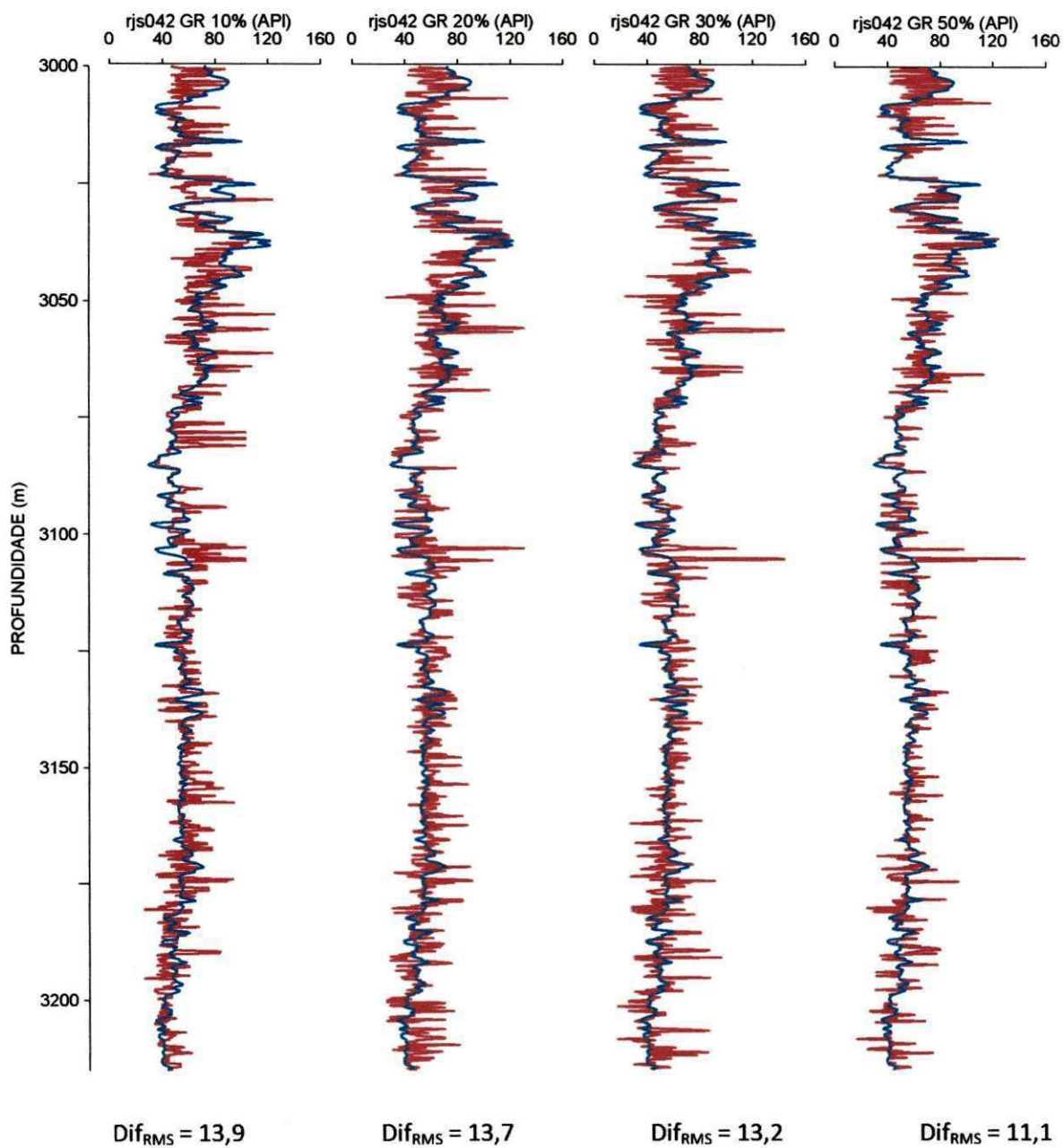


Figura 34 – Perfis de raios gama do poço RJS-42: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

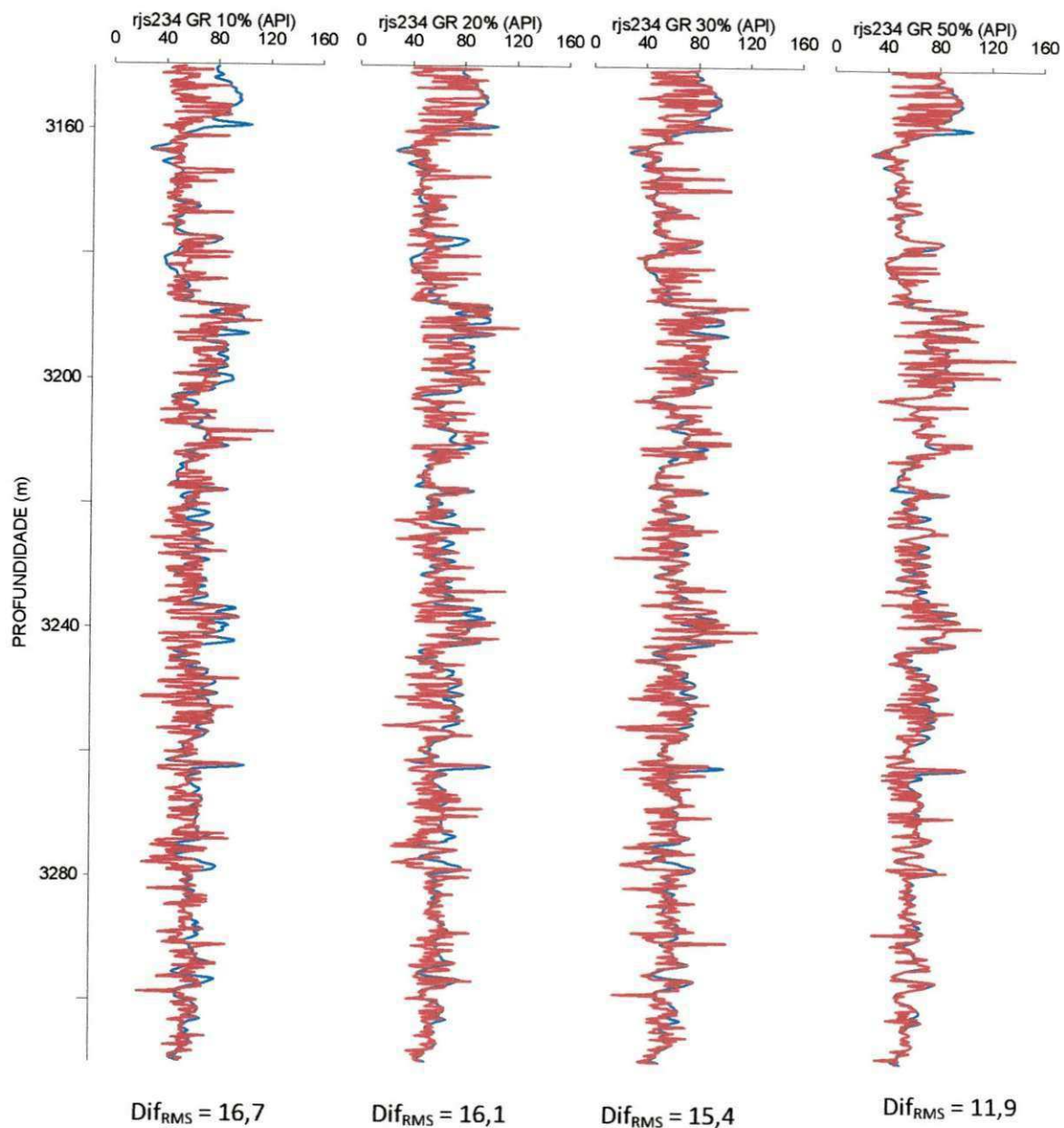


Figura 35 – Perfis de raios gama do poço RJS-234: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

3.1.3 - Perfil de resistividade (ILD)

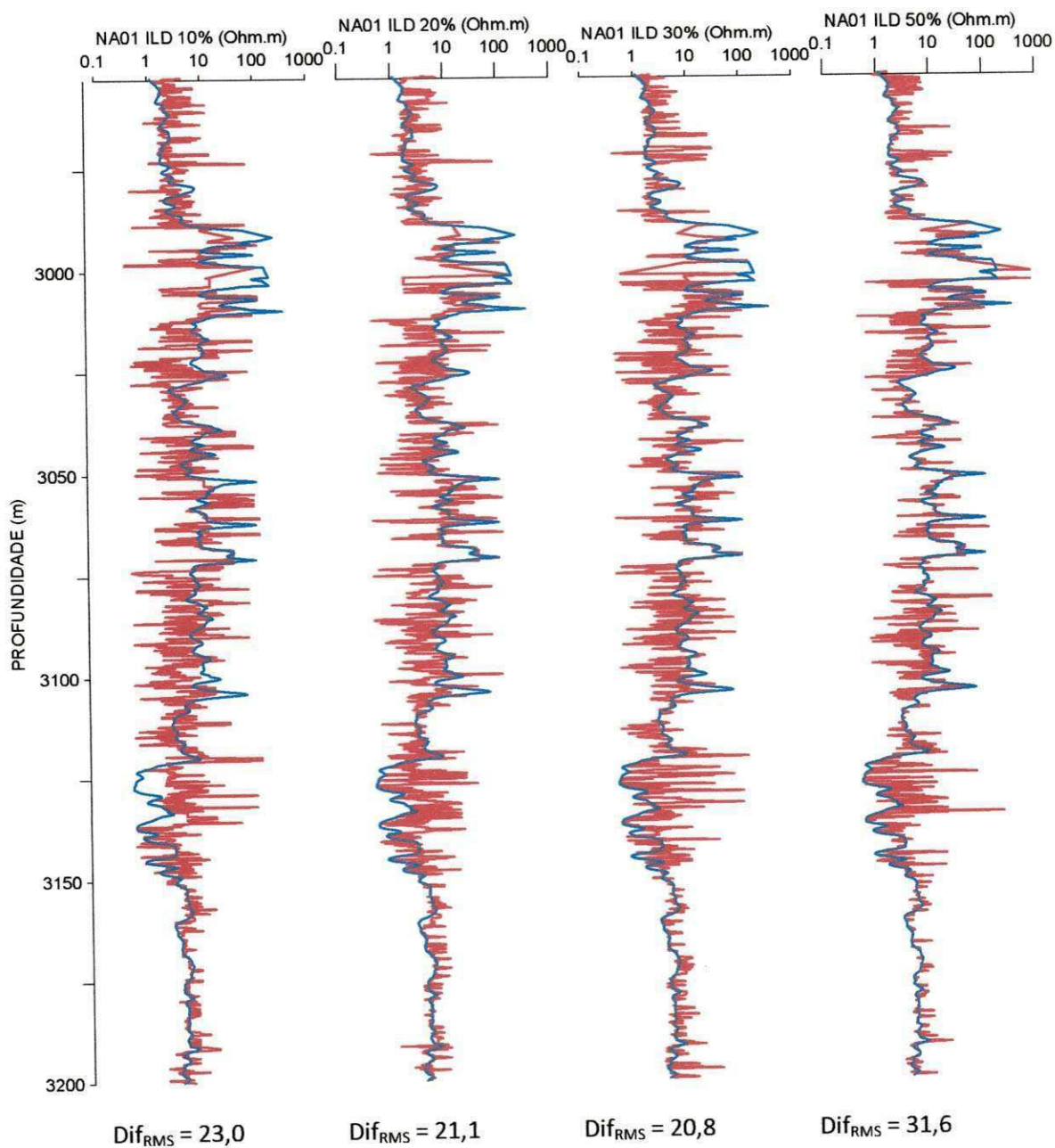


Figura 36 – Perfis de resistividade elétrica do poço NA-01: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

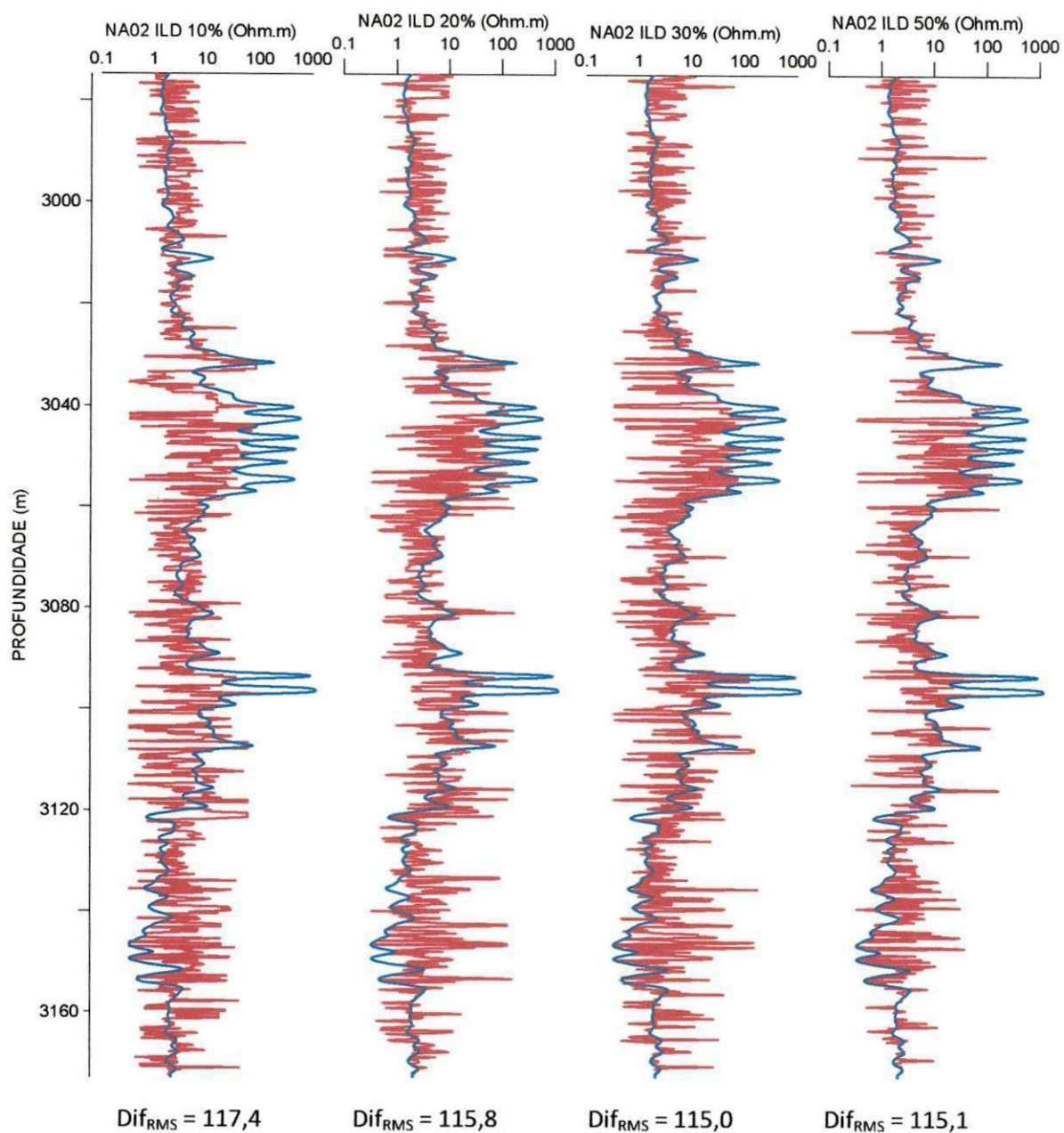


Figura 37 – Perfis de resistividade elétrica do poço NA-02: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

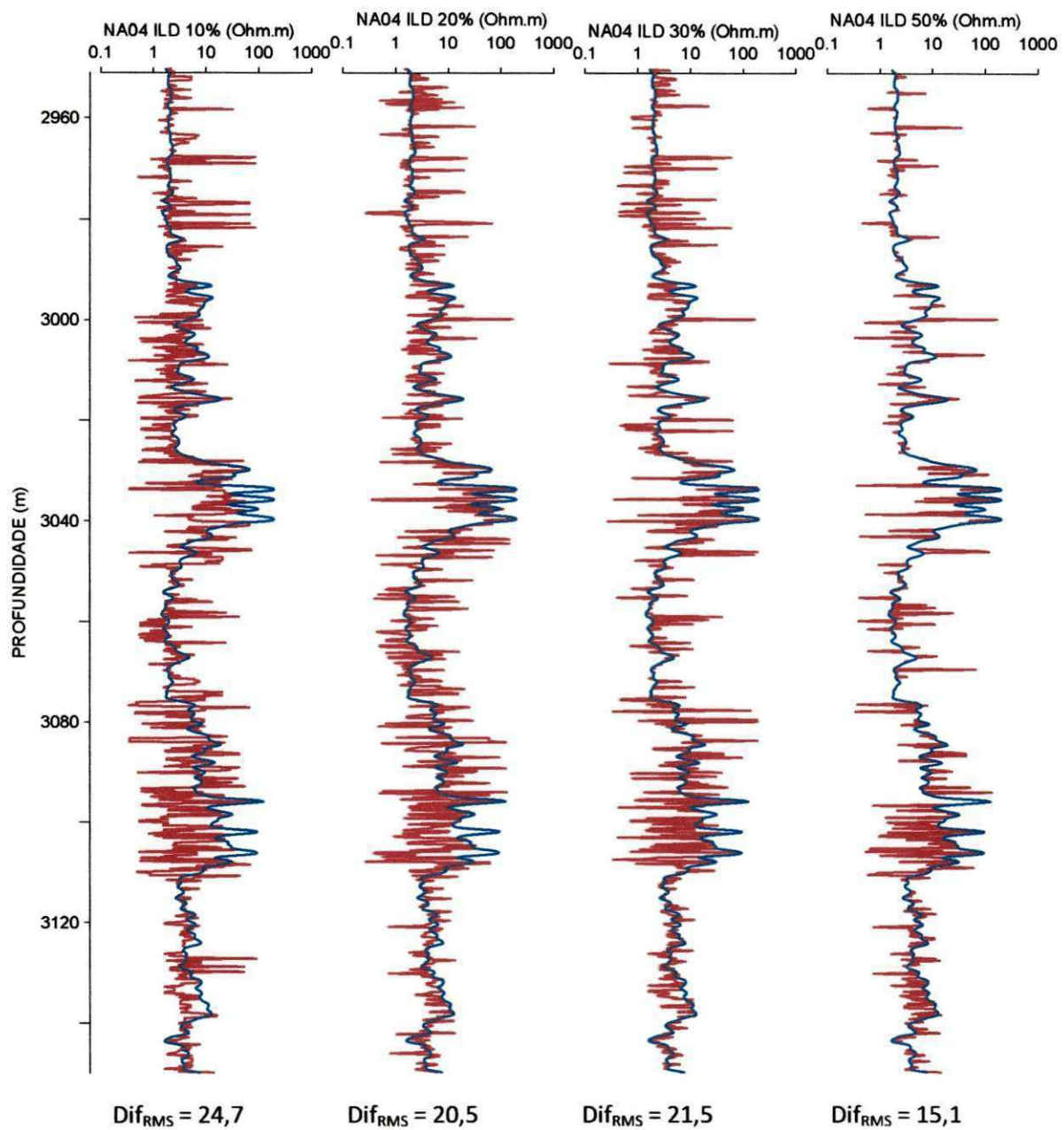


Figura 38 – Perfis de resistividade elétrica do poço NA-04: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

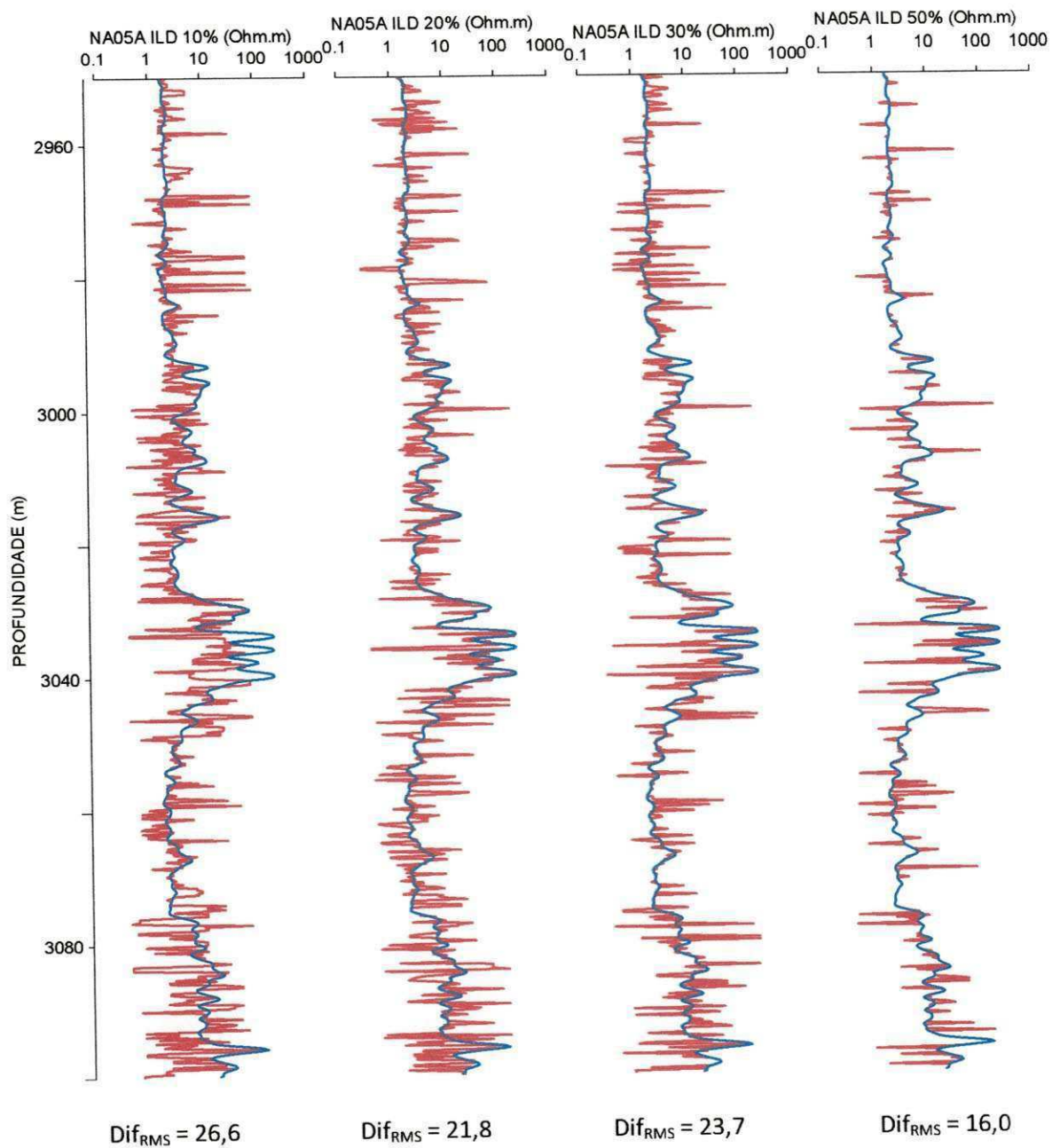


Figura 39 – Perfis de resistividade elétrica do poço NA-05a: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

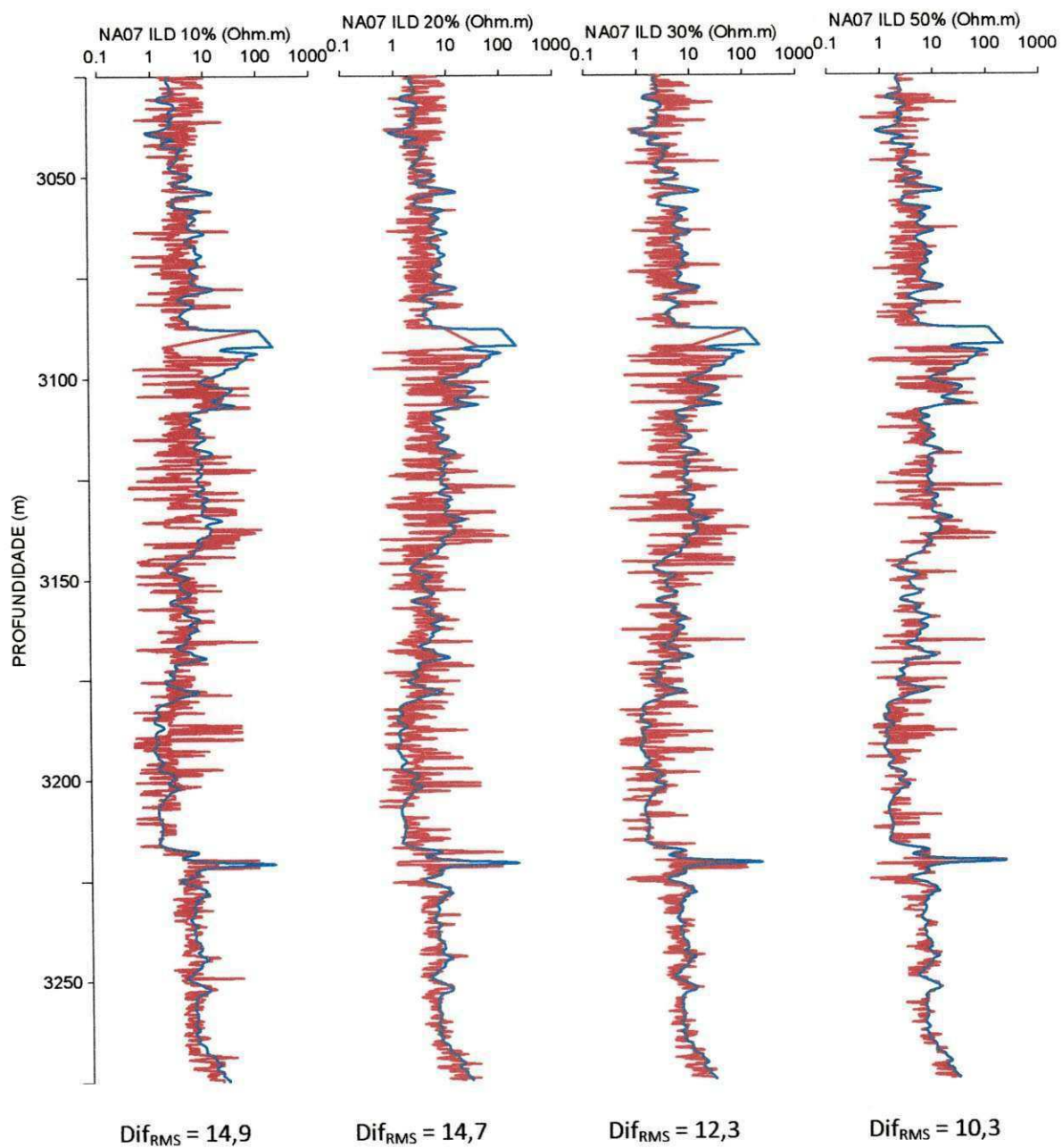


Figura 40 – Perfis de resistividade elétrica do poço NA-07: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

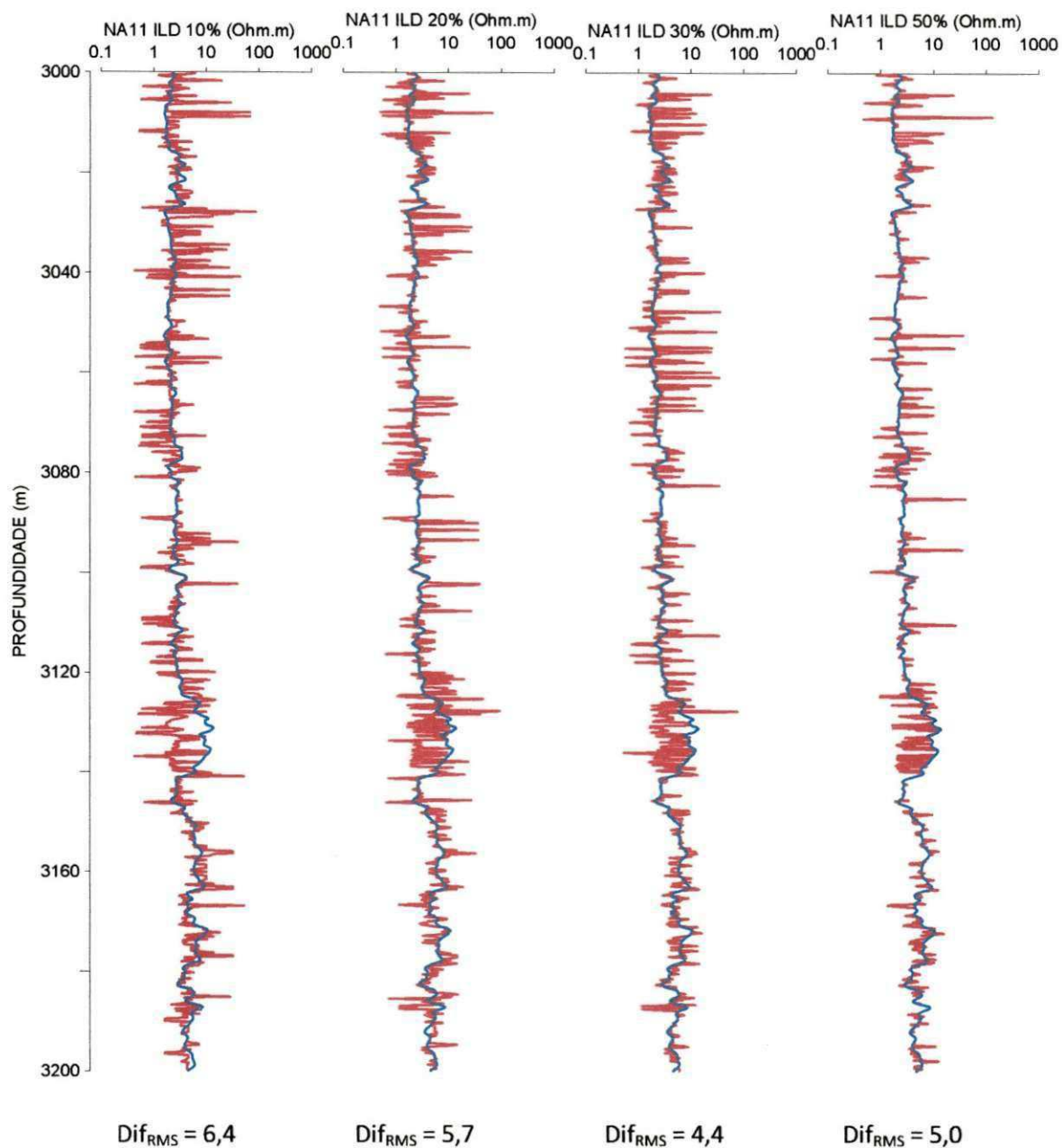


Figura 41 – Perfis de resistividade elétrica do poço NA-11: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

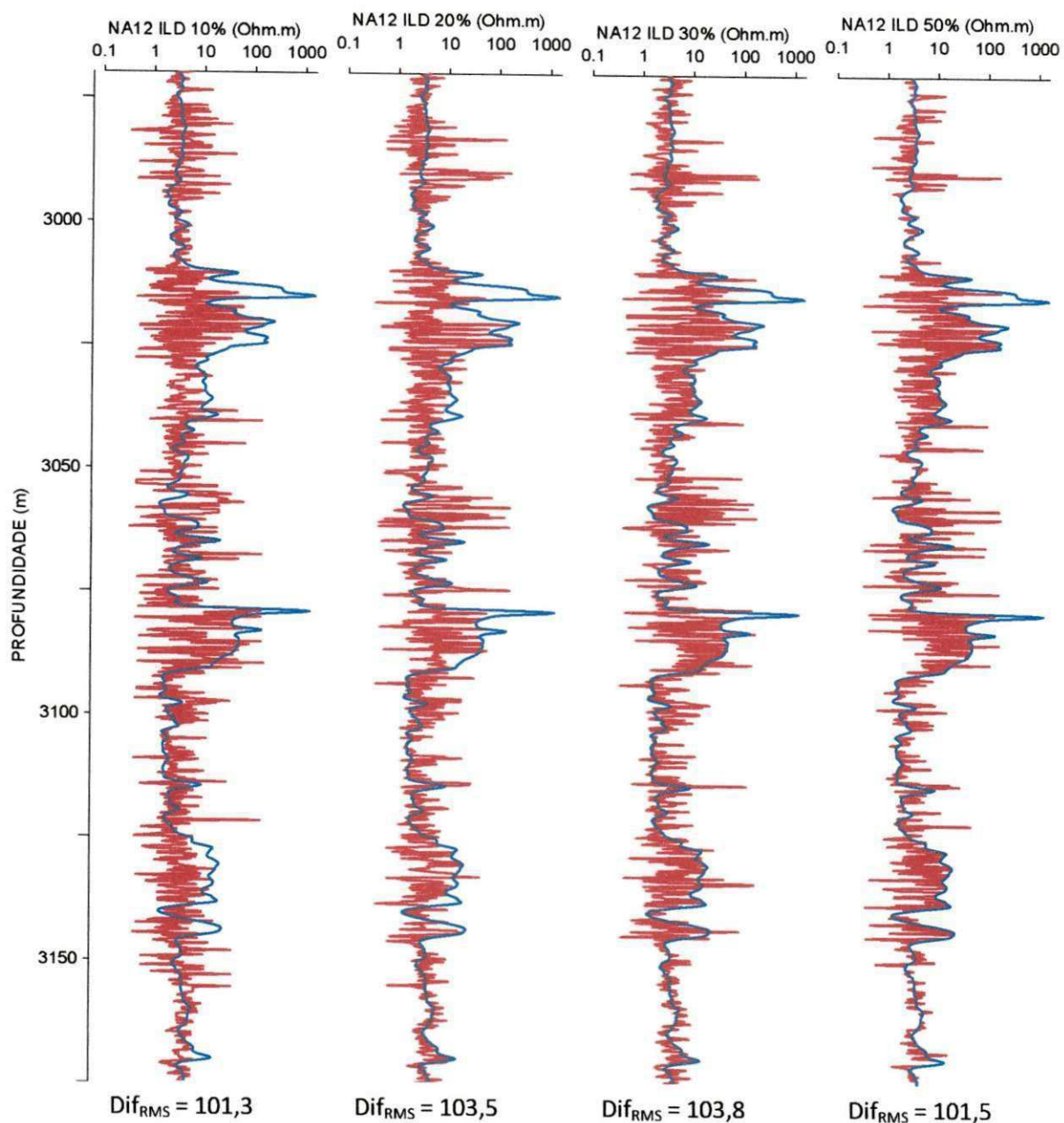


Figura 42 – Perfis de resistividade elétrica do poço NA-12: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

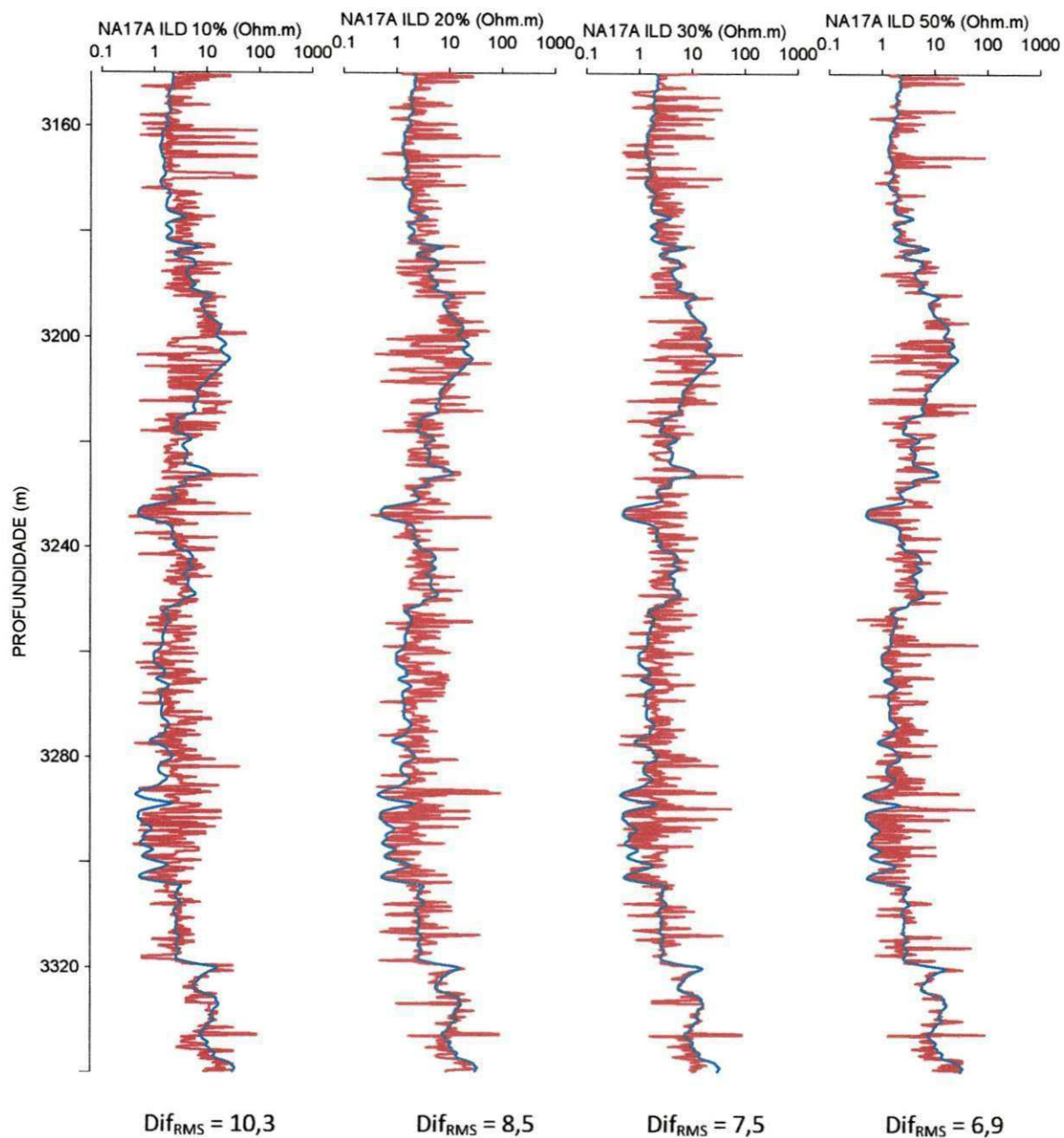


Figura 43 – Perfis de resistividade elétrica do poço NA-17a: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

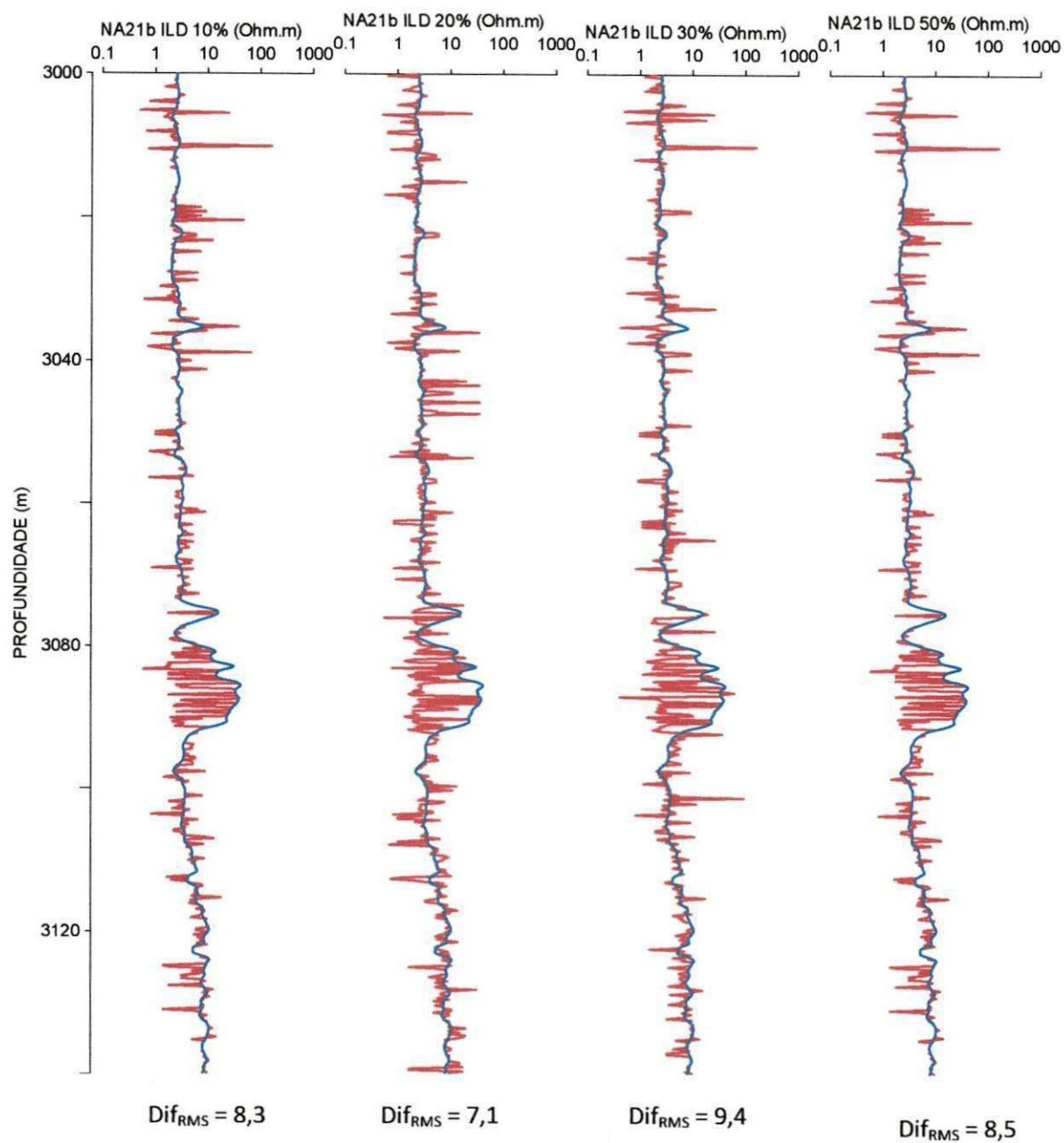


Figura 44 – Perfis de resistividade elétrica do poço NA-21b: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

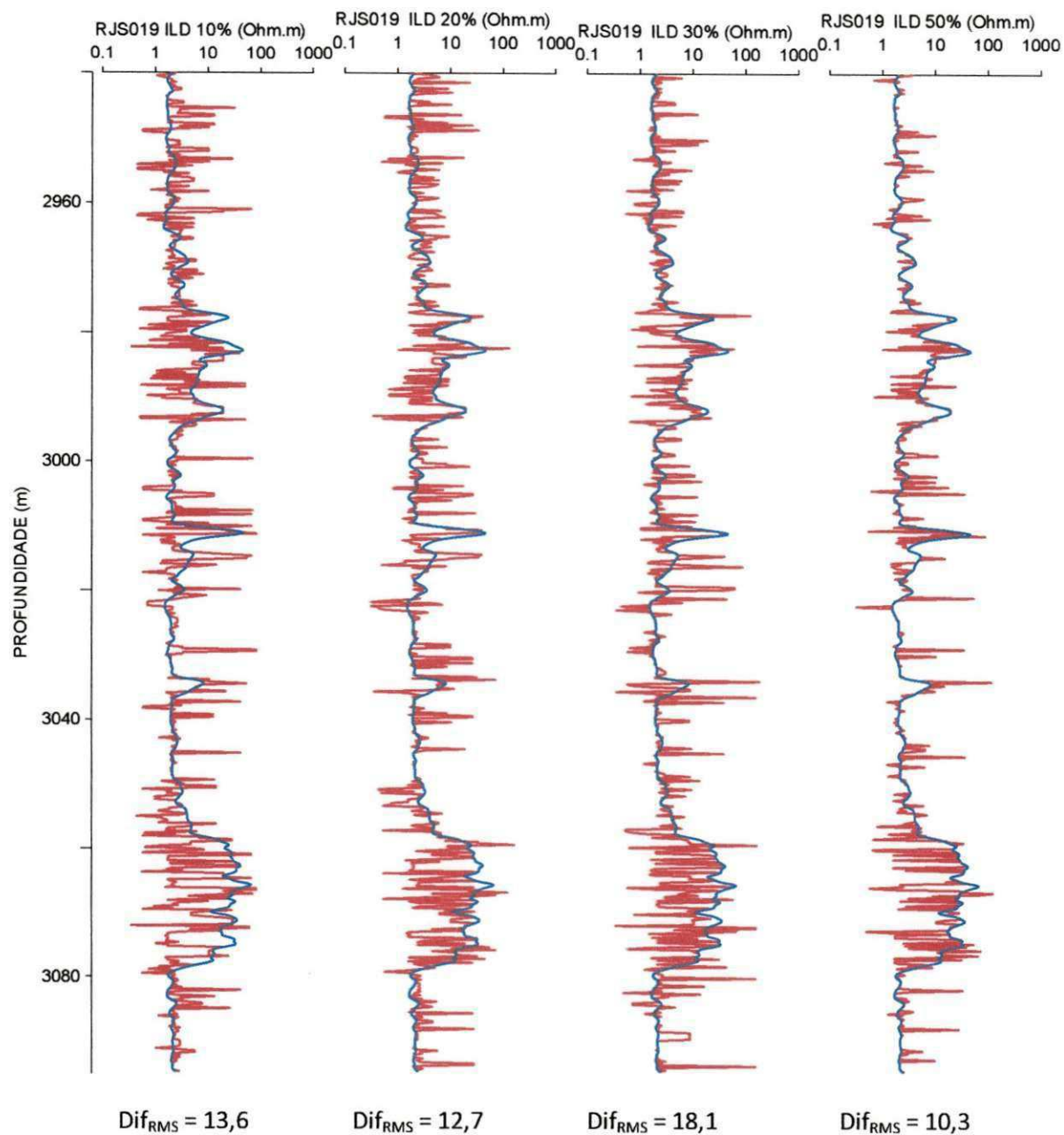


Figura 45 – Perfis de resistividade elétrica do poço RJS-19: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

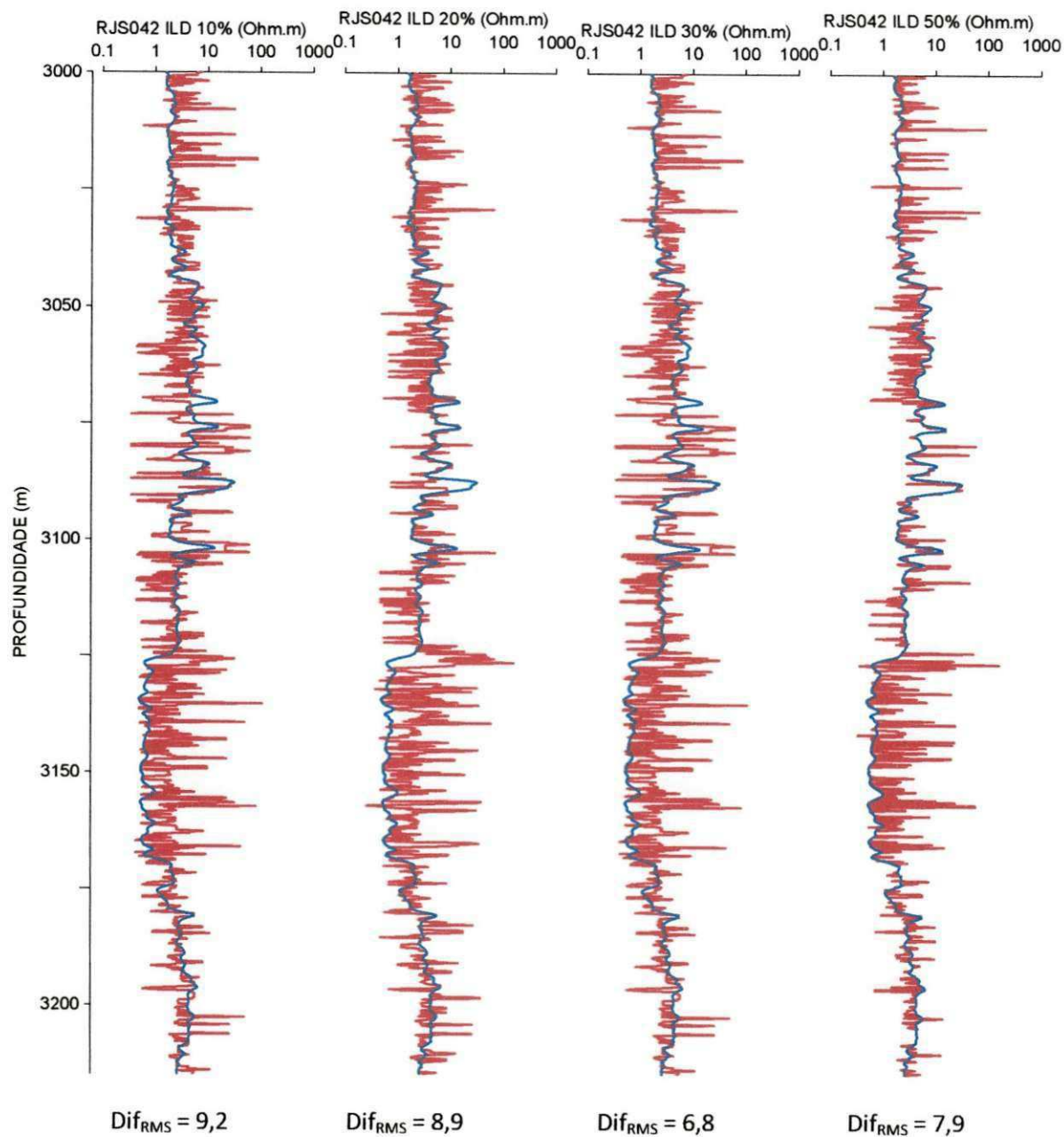


Figura 46 – Perfis de resistividade elétrica do poço RJS-42: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

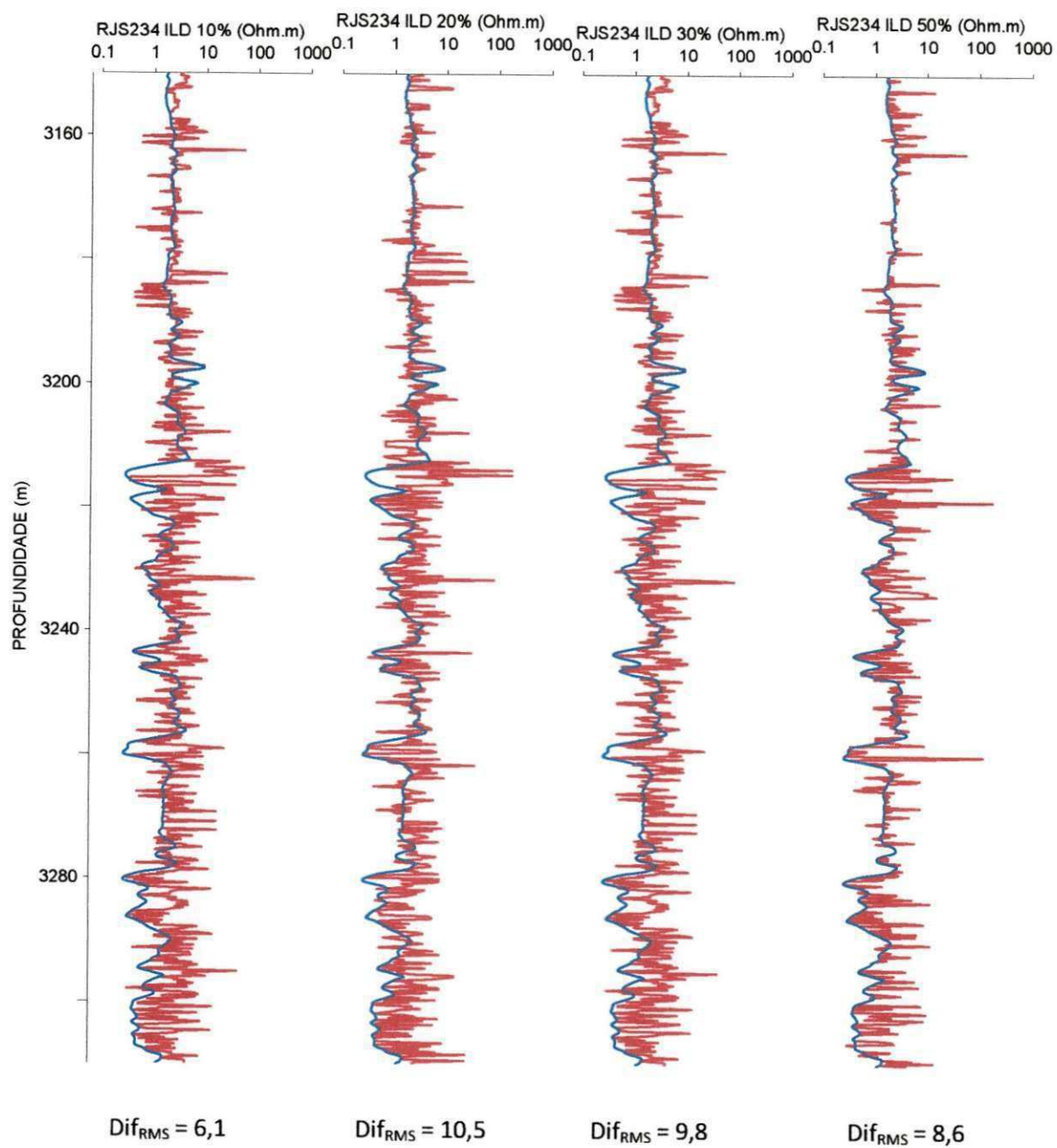


Figura 47 – Perfis de resistividade elétrica do poço RJS-234: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

3.1.4 - Perfil de porosidade de nêutrons (NPHI)

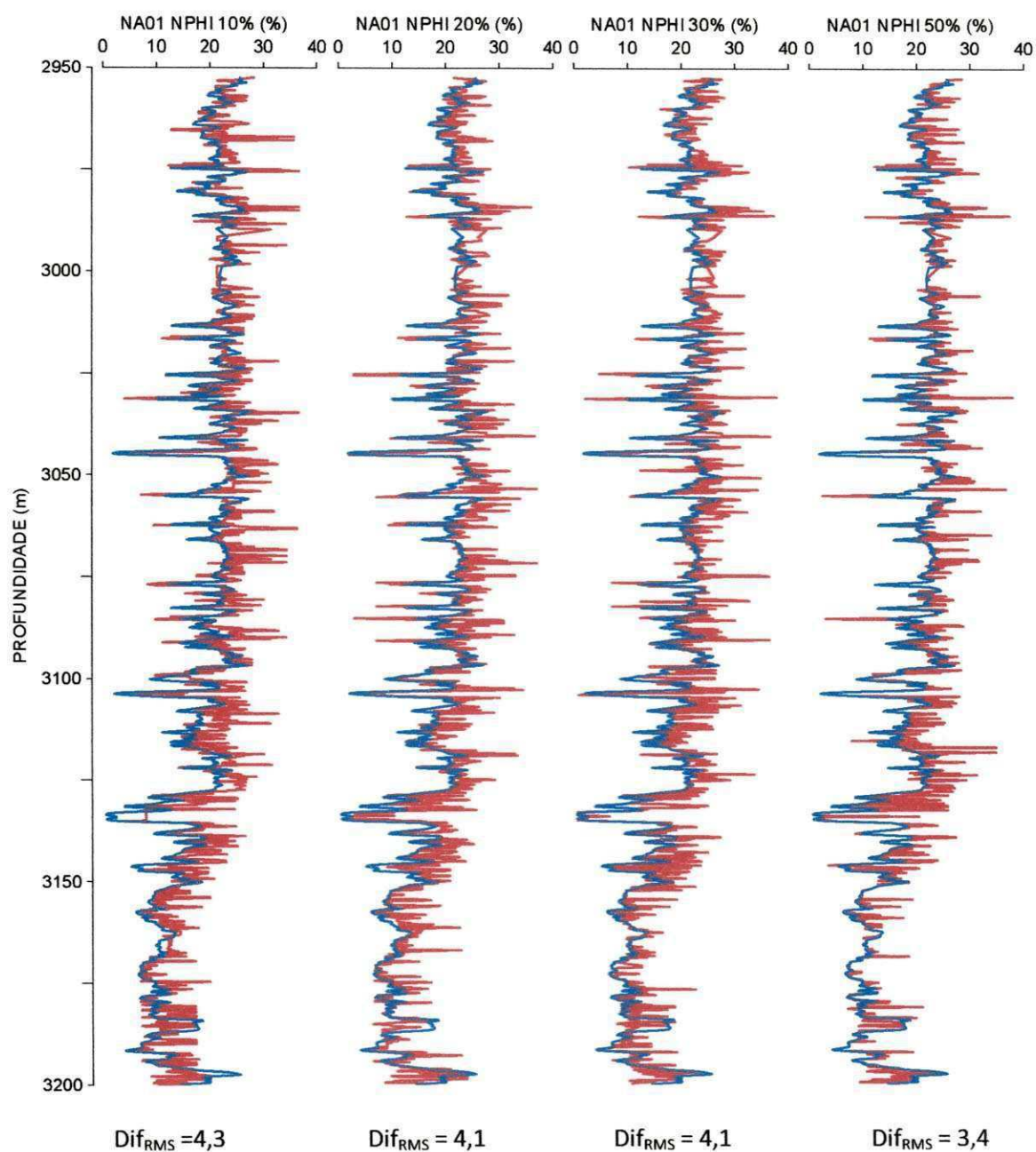


Figura 48 – Perfis de porosidade de nêutrons do poço NA-01: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

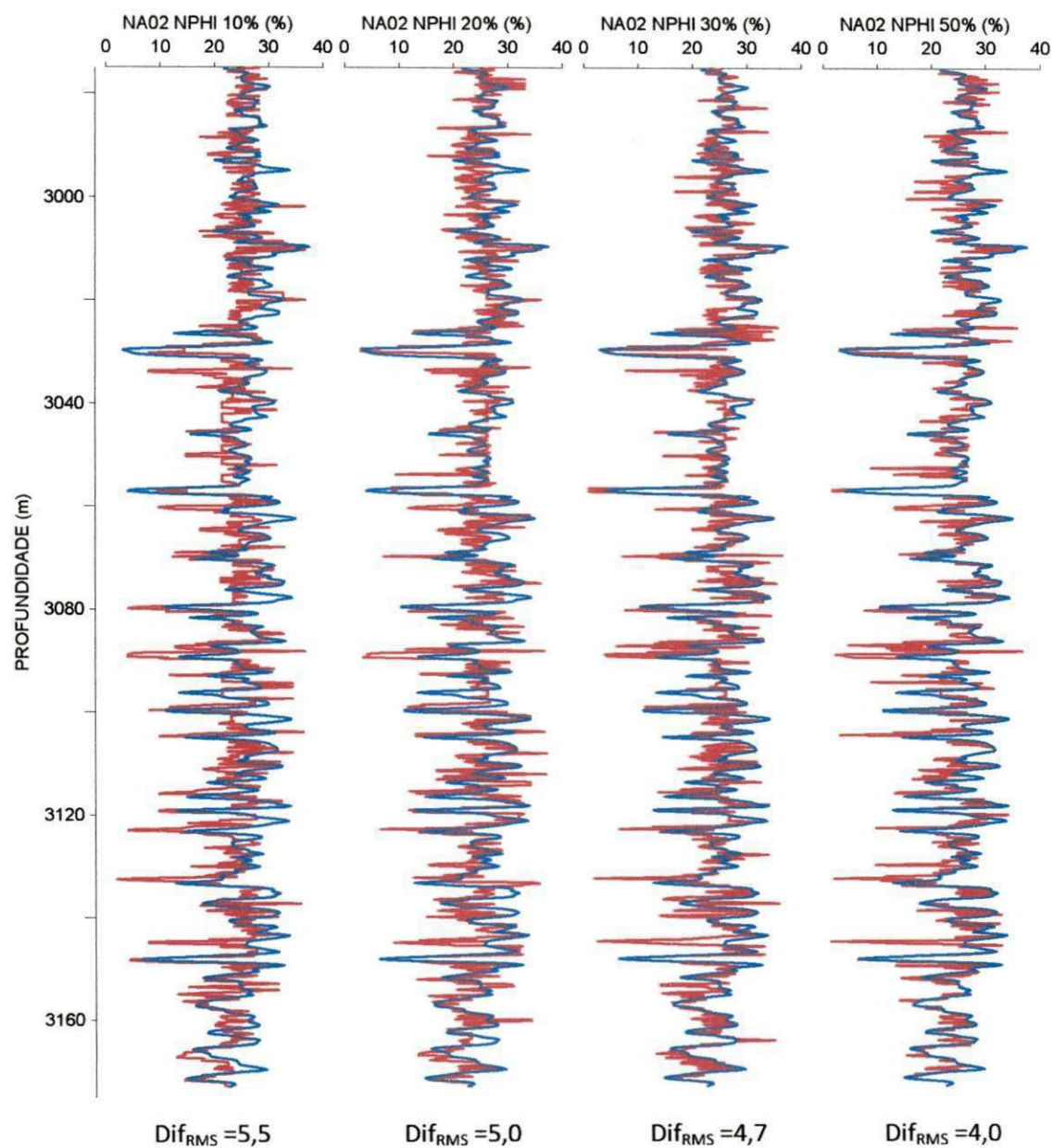


Figura 49 – Perfis de porosidade de nêutrons do poço NA-02: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

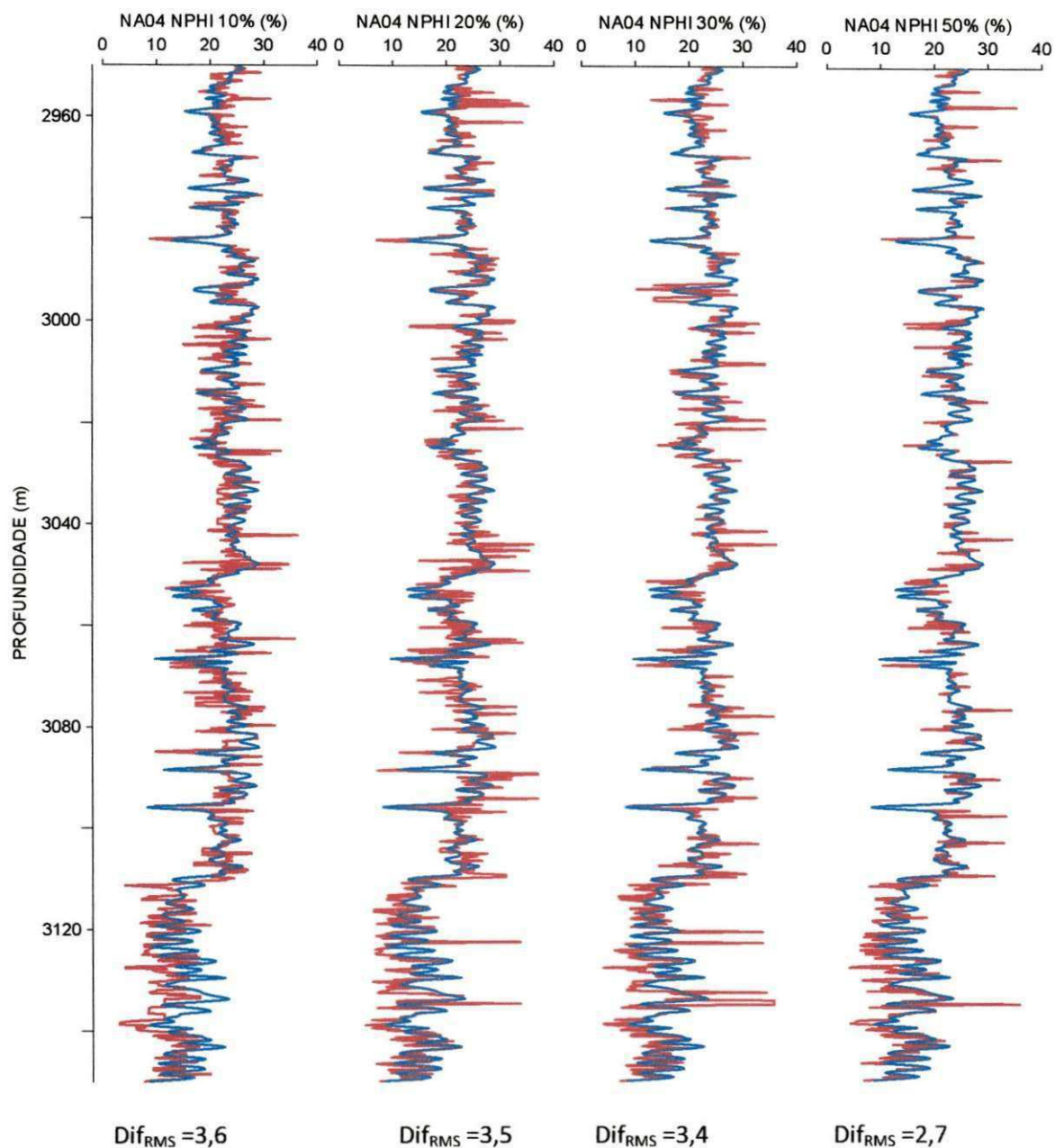


Figura 50 – Perfis de porosidade de nêutrons do poço NA-04: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

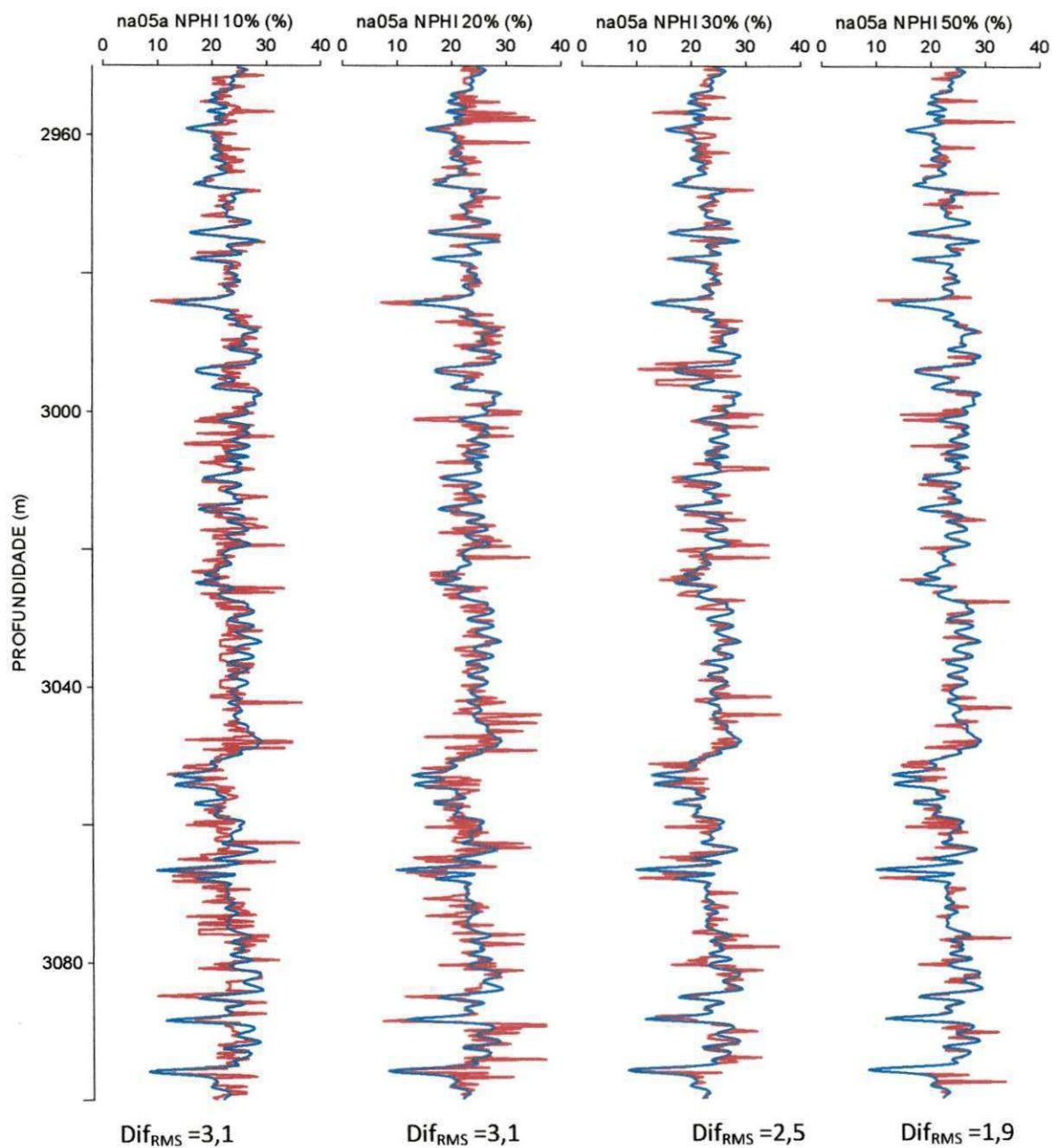


Figura 51 – Perfis de porosidade de nêutrons do poço NA-05a: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

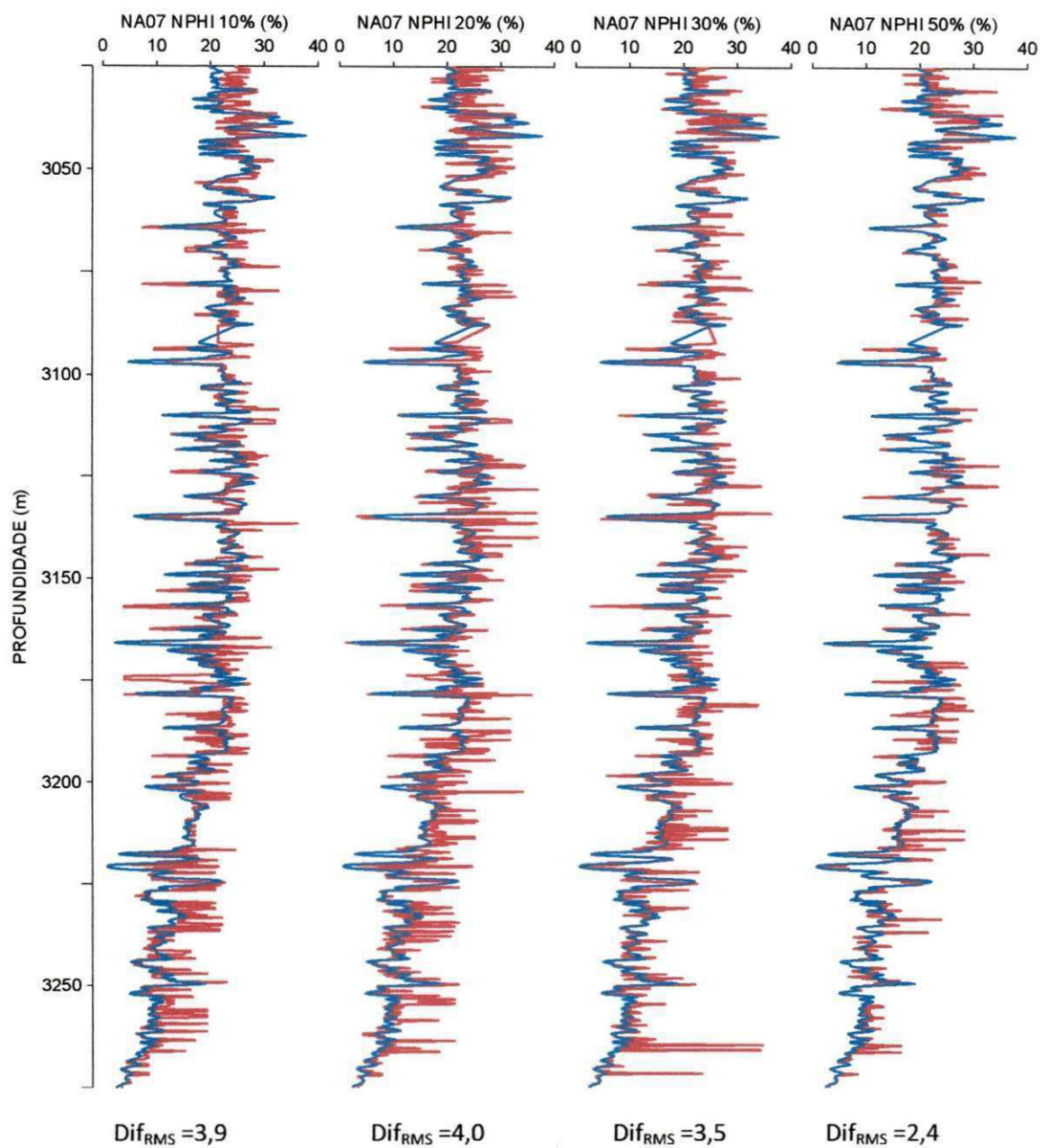


Figura 52 – Perfis de porosidade de nêutrons do poço NA-07: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

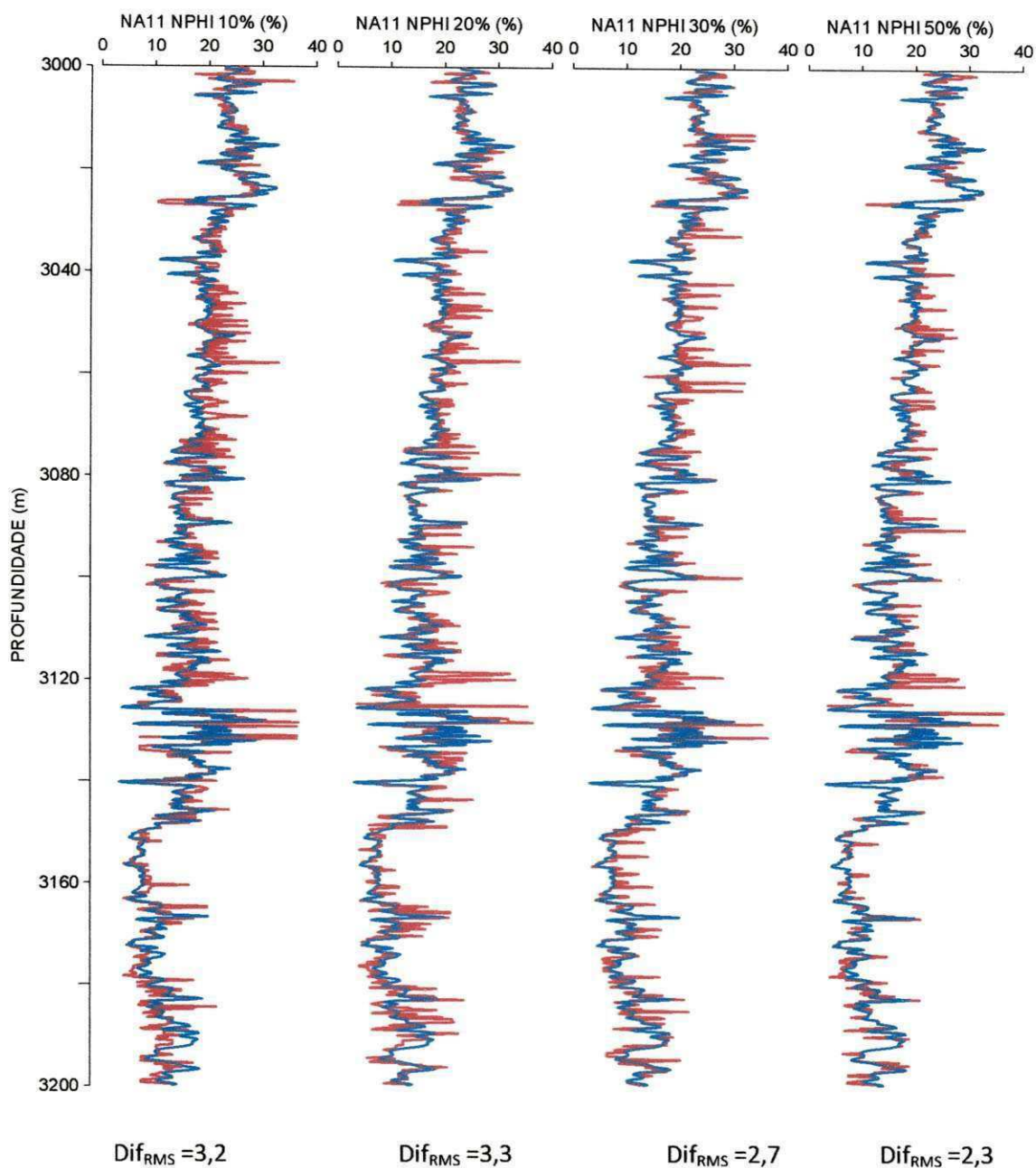


Figura 53 – Perfis de porosidade de nêutrons do poço NA-11: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

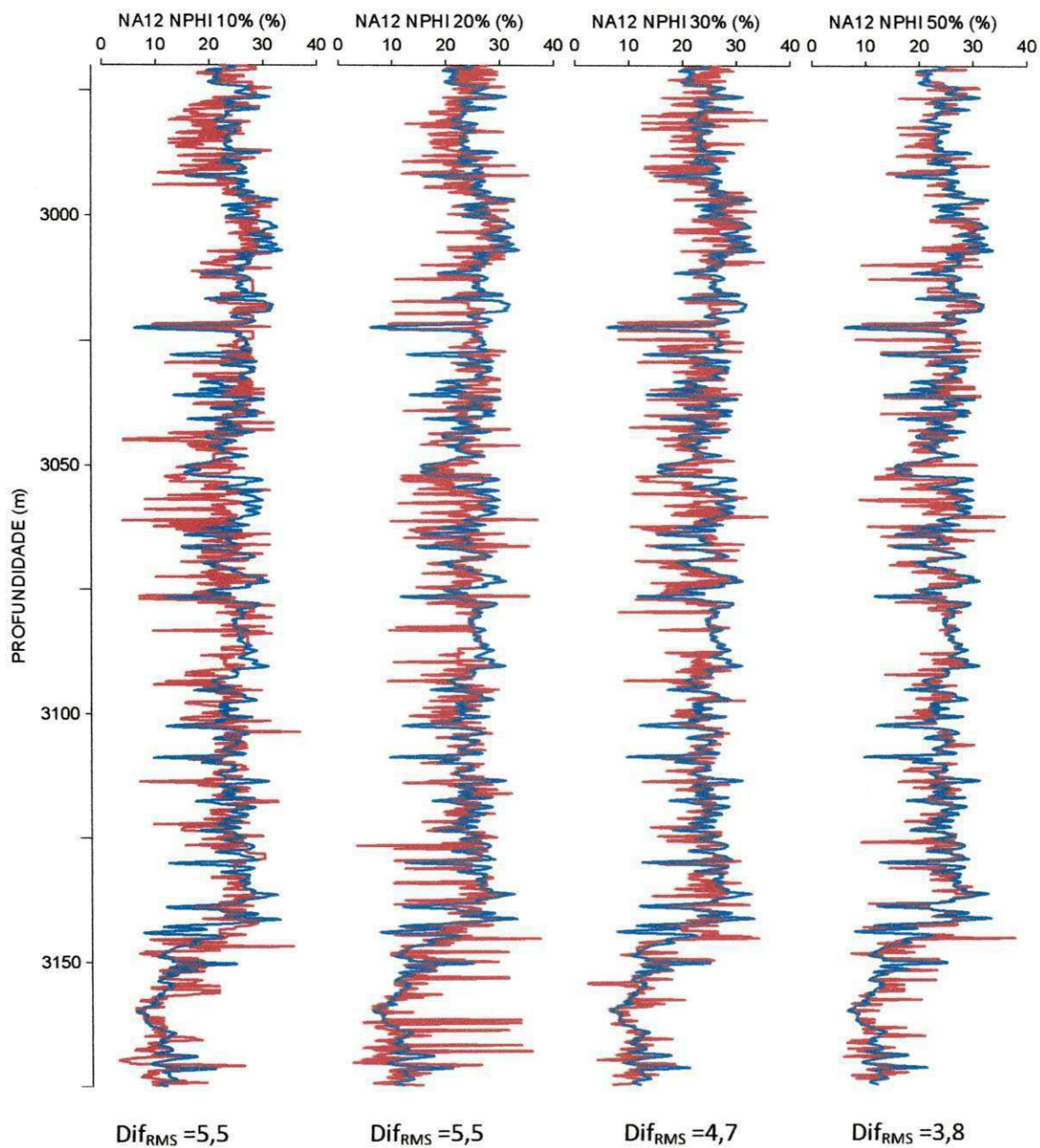


Figura 54 – Perfis de porosidade de nêutrons do poço NA-12: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

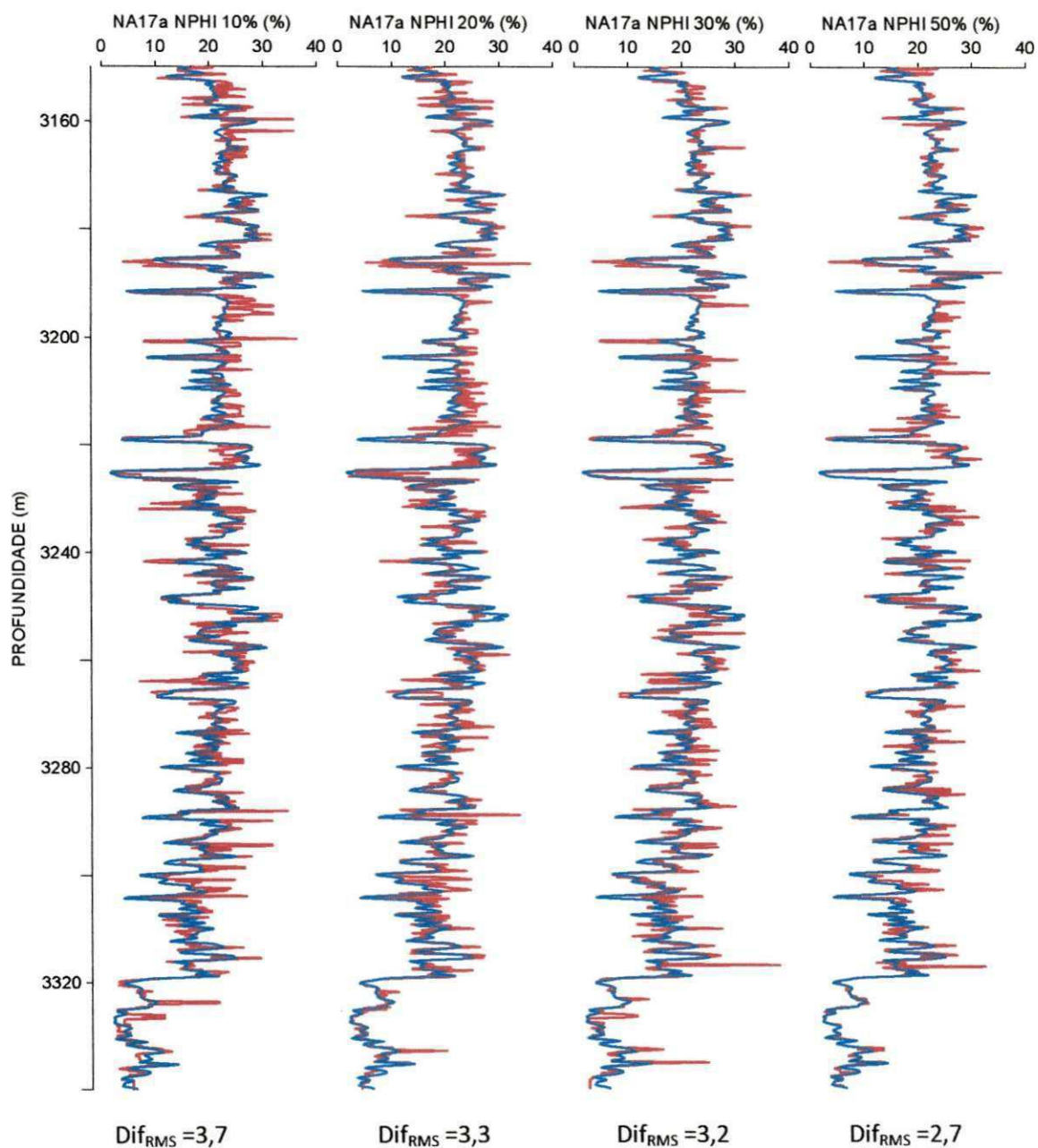


Figura 55 – Perfis de porosidade de nêutrons do poço NA-17a: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

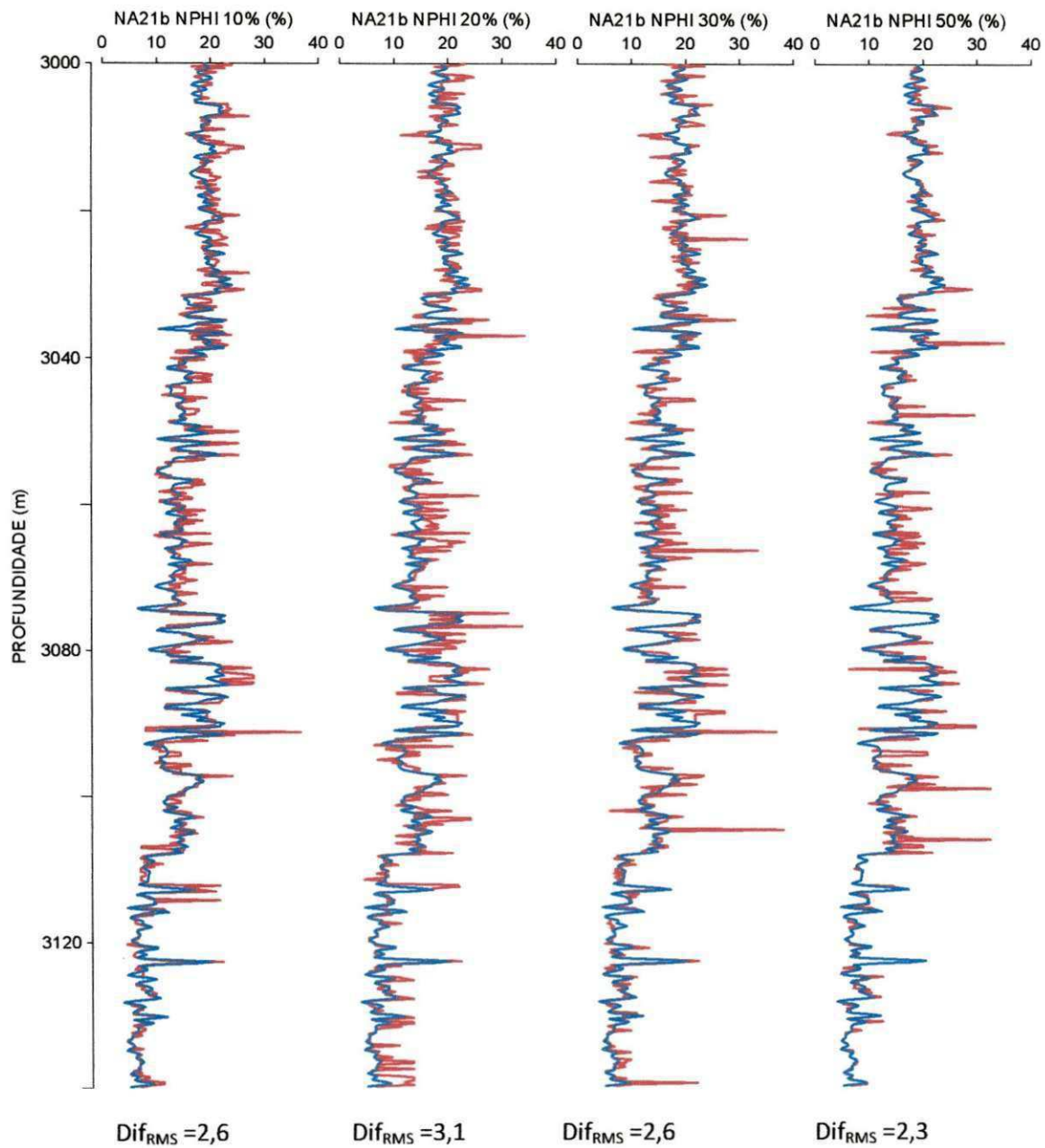


Figura 56 – Perfis de porosidade de nêutrons do poço NA-21b: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

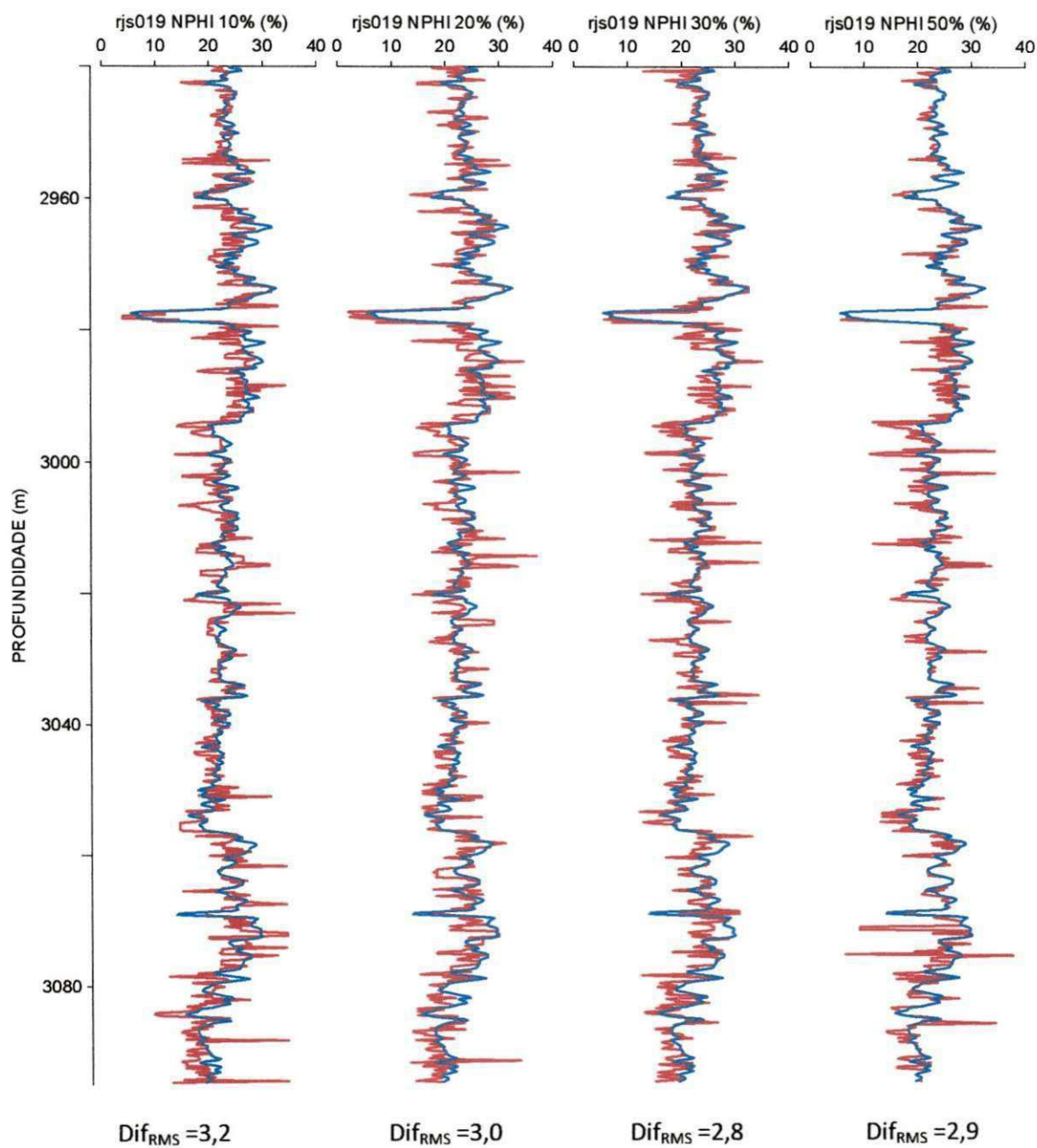


Figura 57 – Perfis de porosidade de nêutrons do poço RJS-19: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

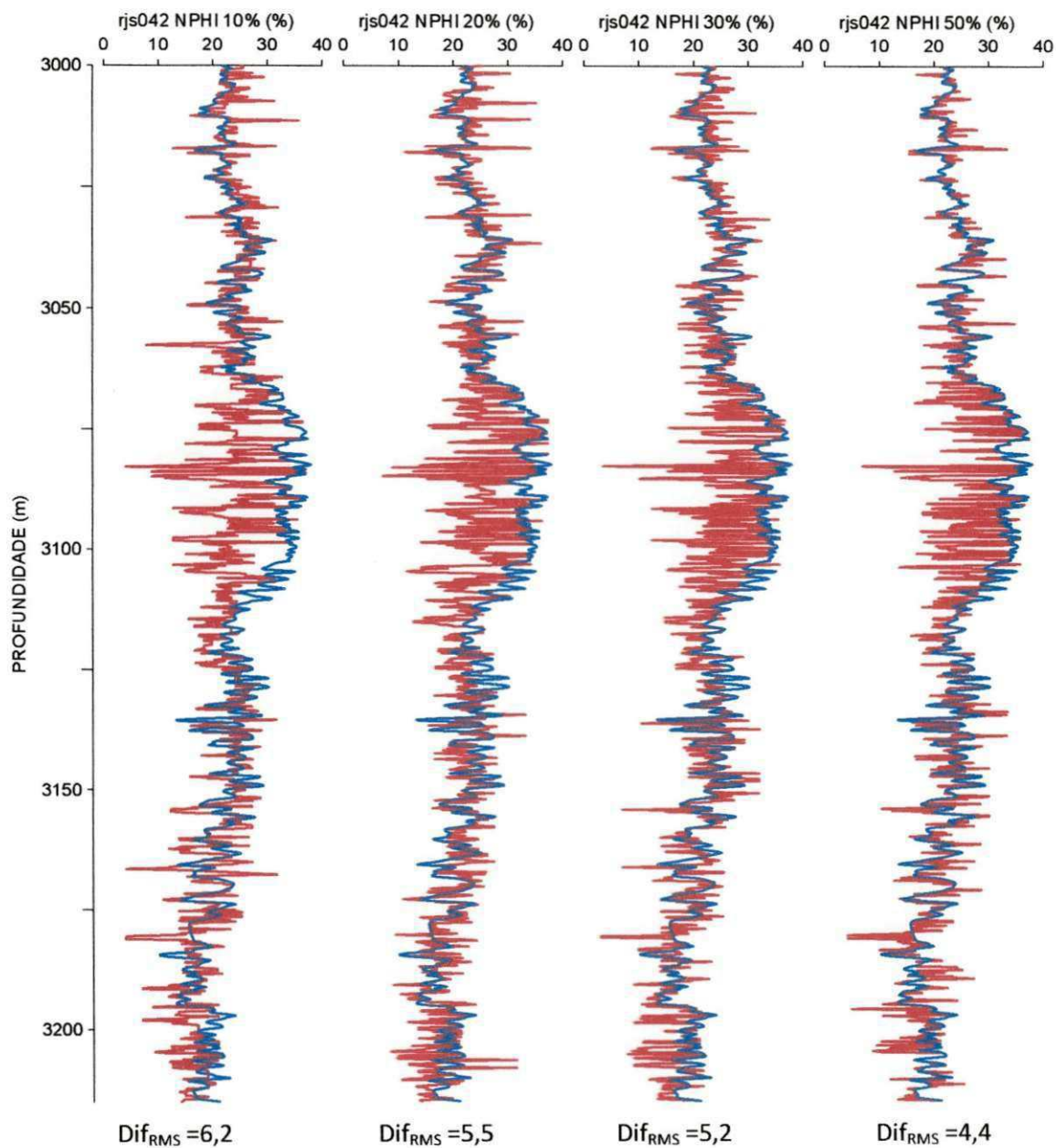


Figura 58 – Perfis de porosidade de nêutrons do poço RJS-42: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

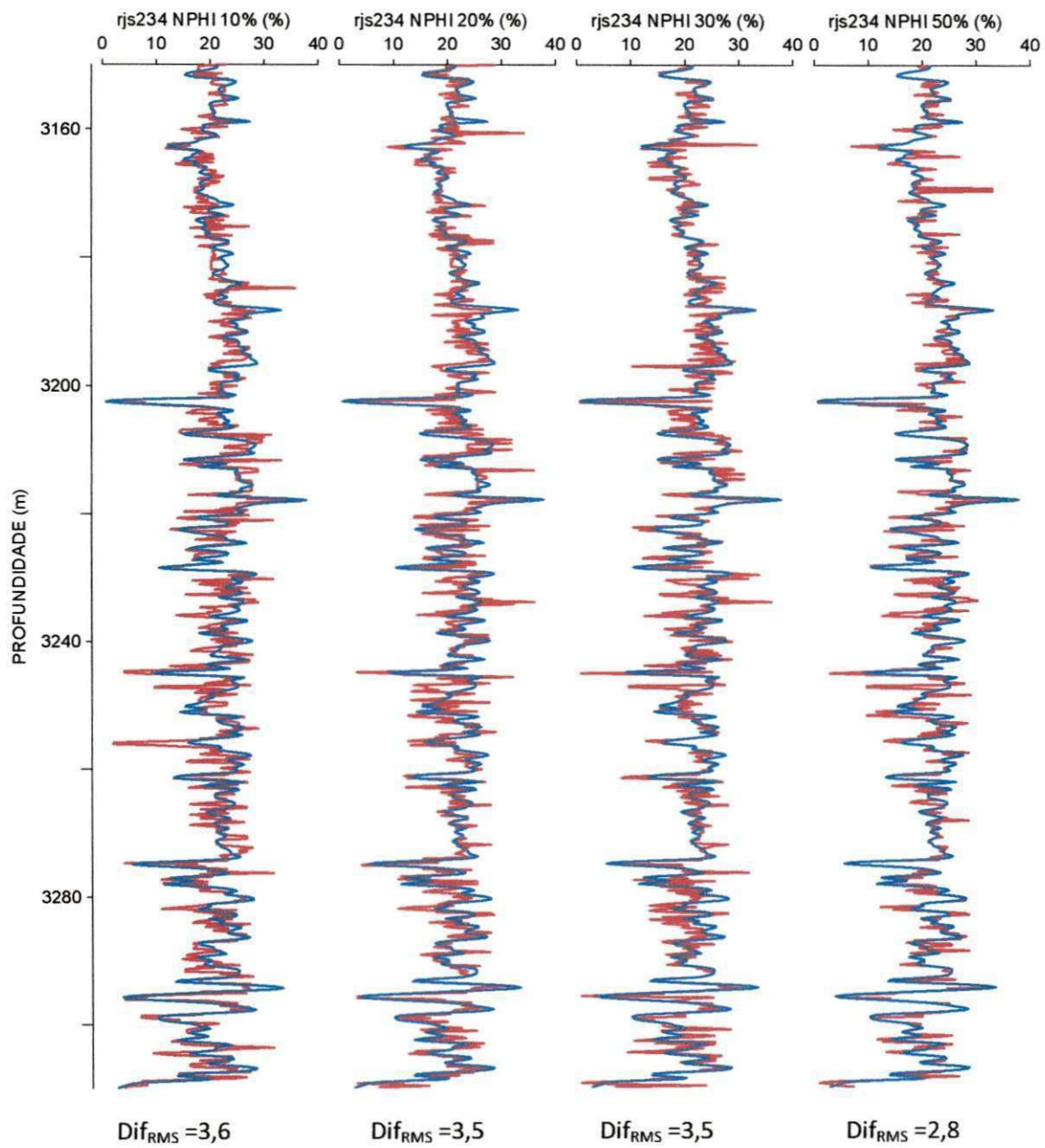


Figura 59 – Perfis de porosidade de nêutrons do poço RJS-234: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

3.1.5 - Perfil de densidade (RHOB)

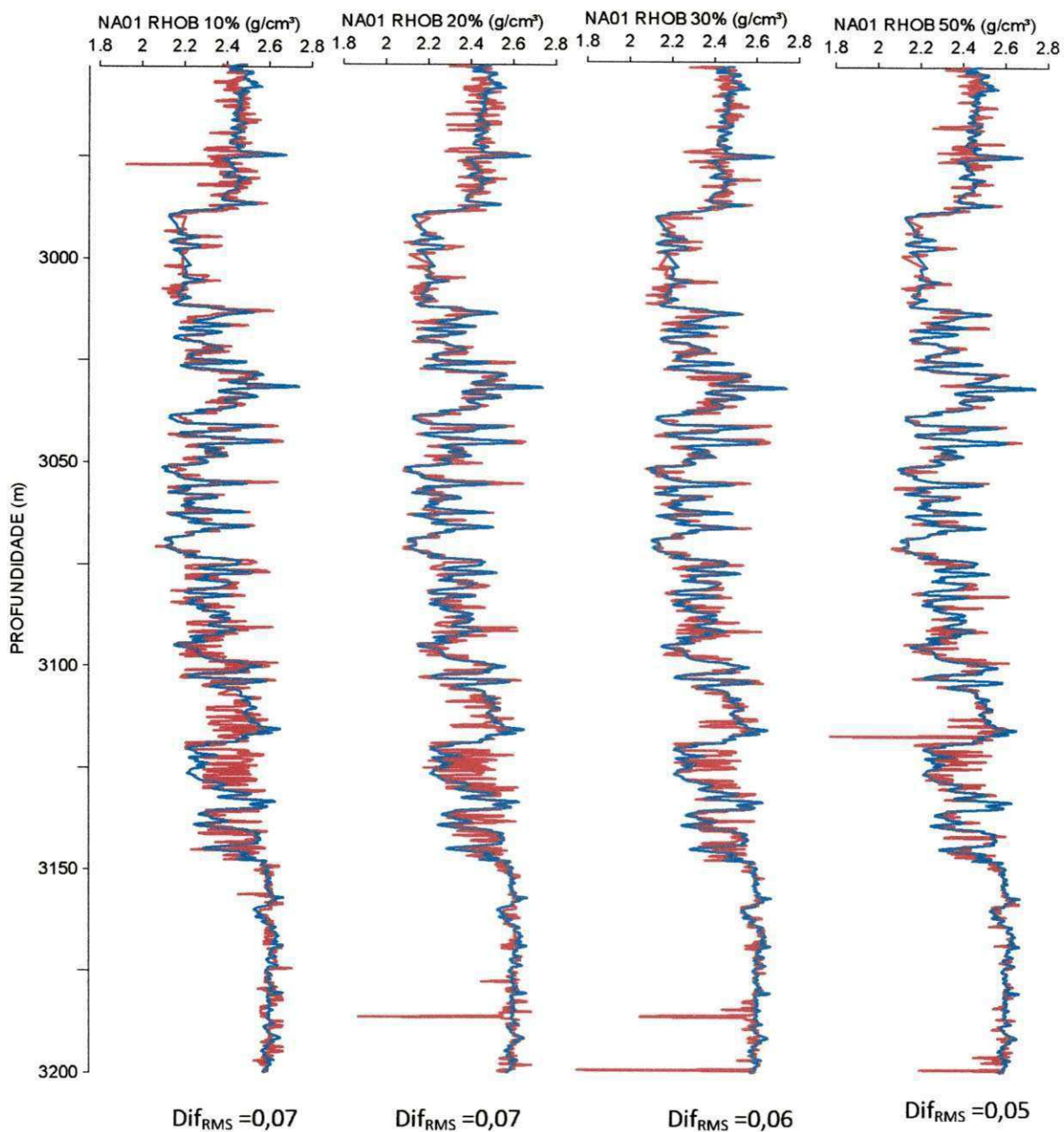


Figura 60 – Perfis de densidade do poço NA-01: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

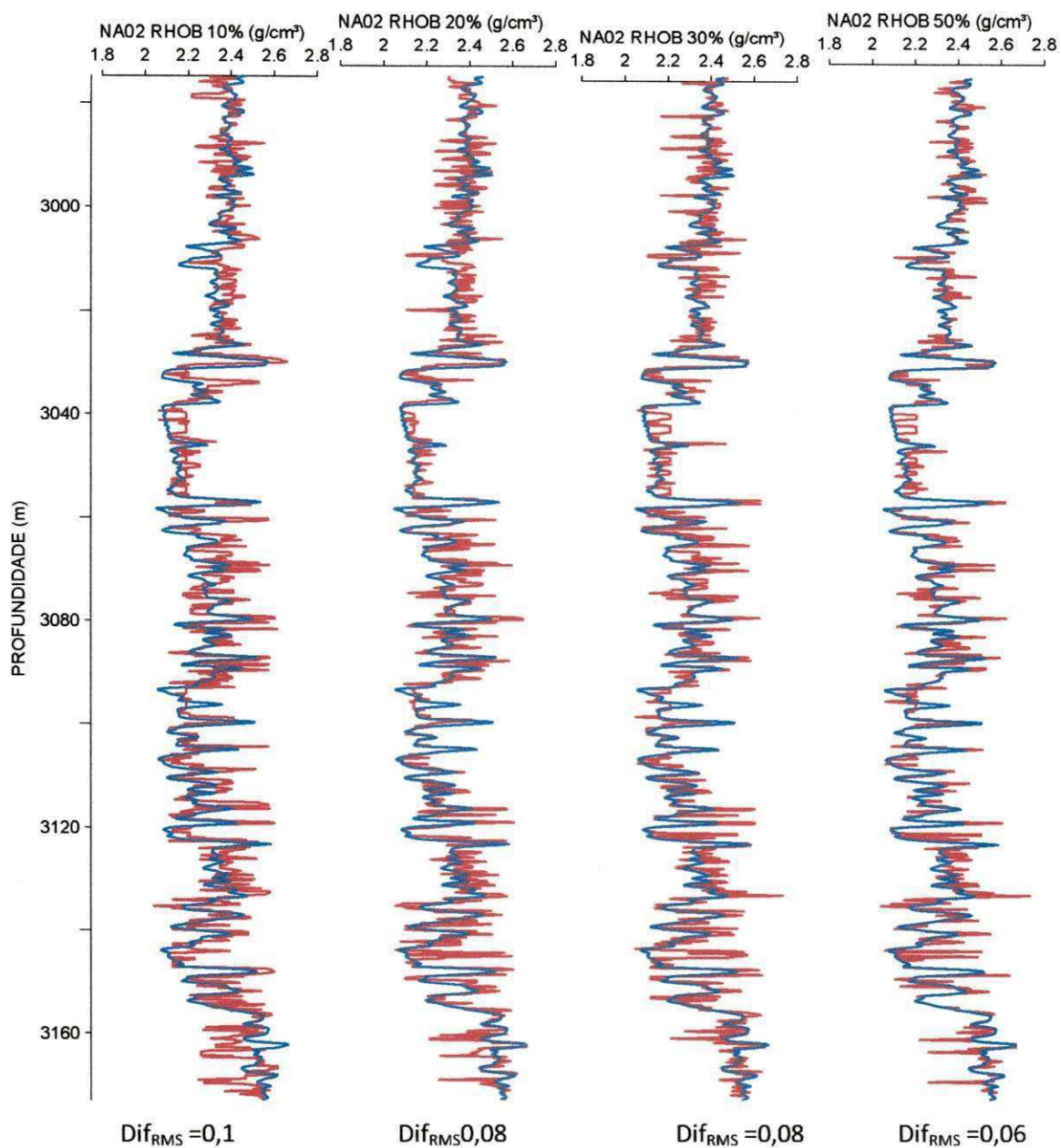


Figura 61 – Perfis de densidade do poço NA-02: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

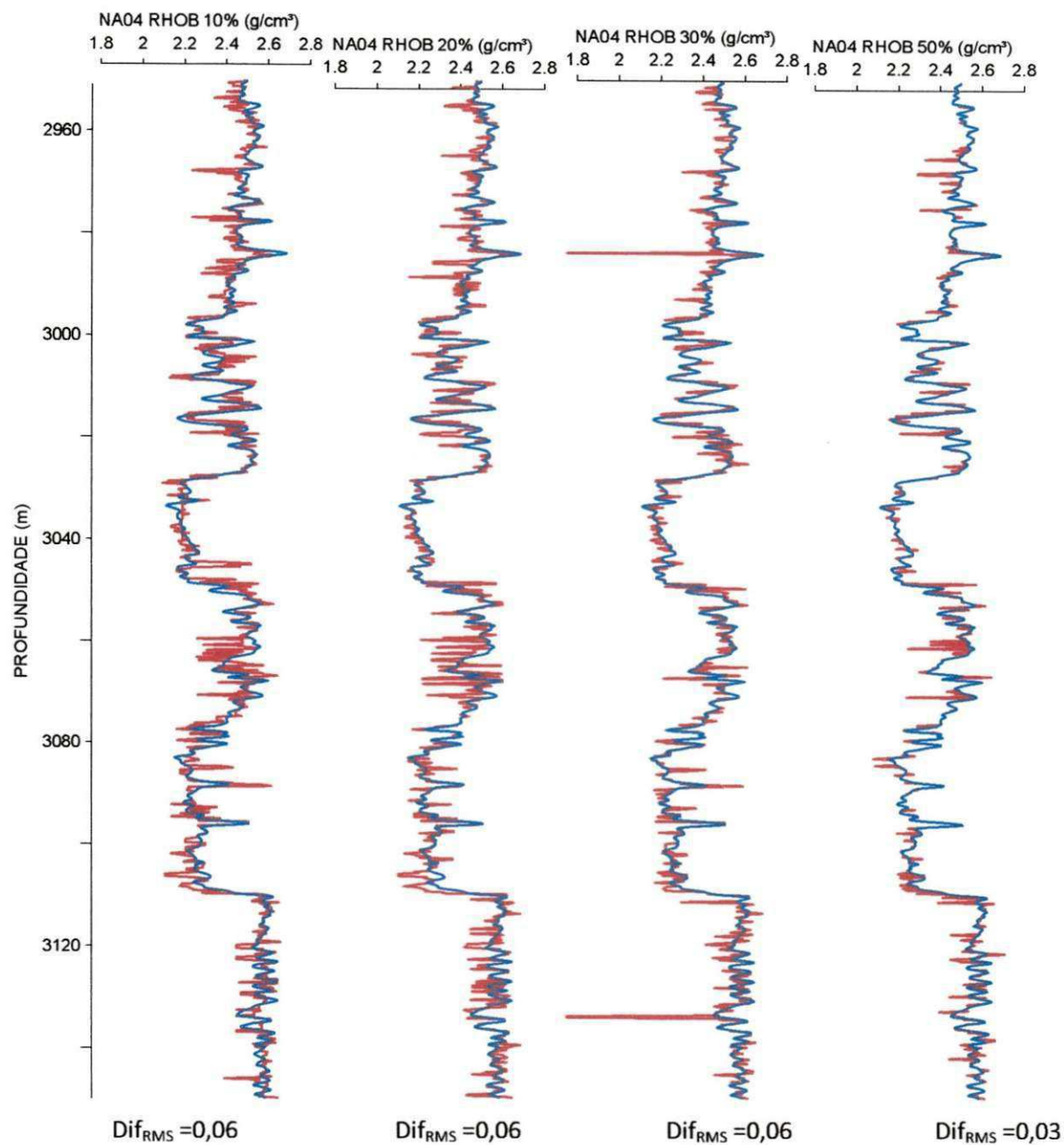


Figura 62 – Perfis de densidade do poço NA-04: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

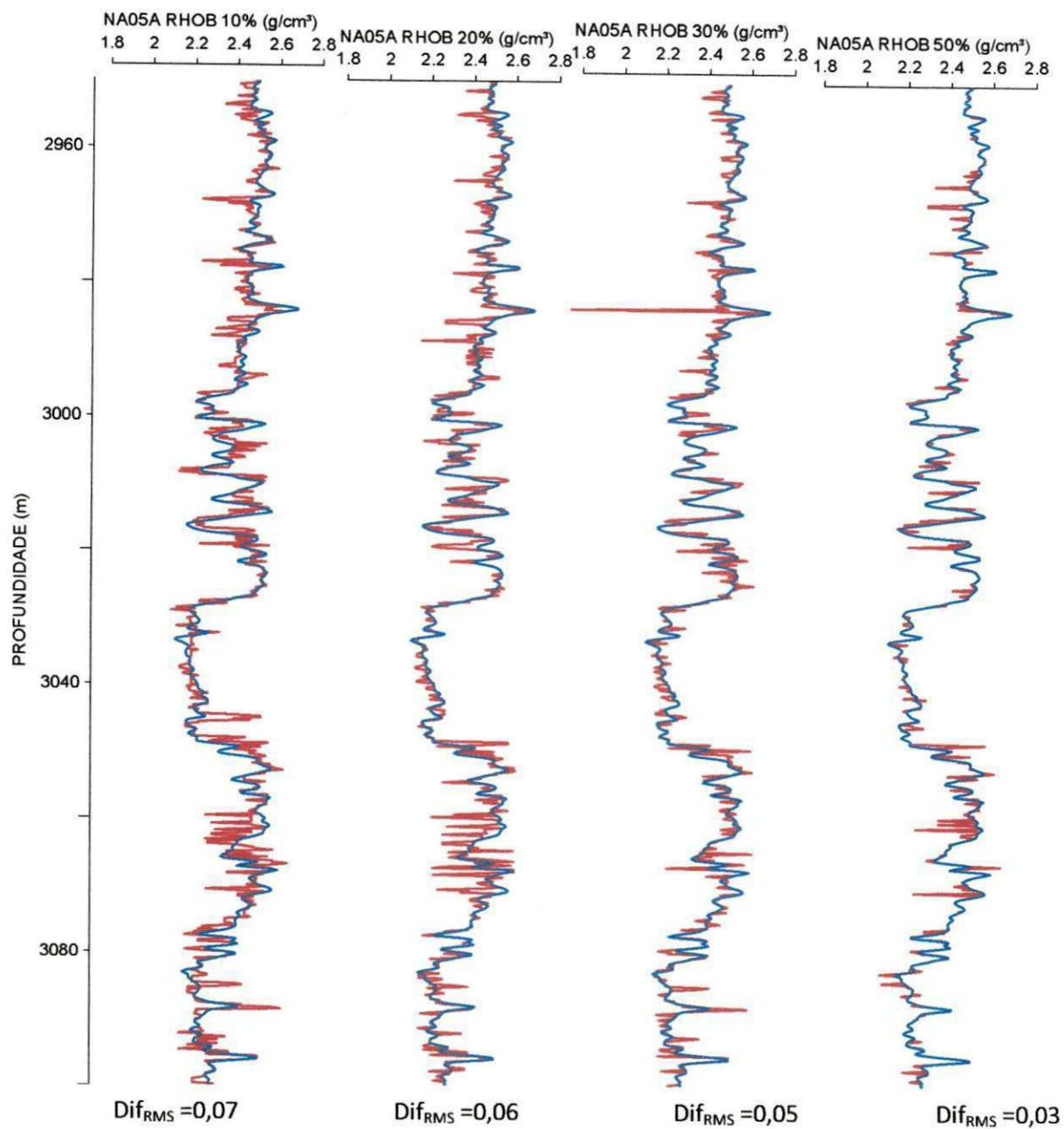


Figura 63 – Perfis de densidade do poço NA-05a: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

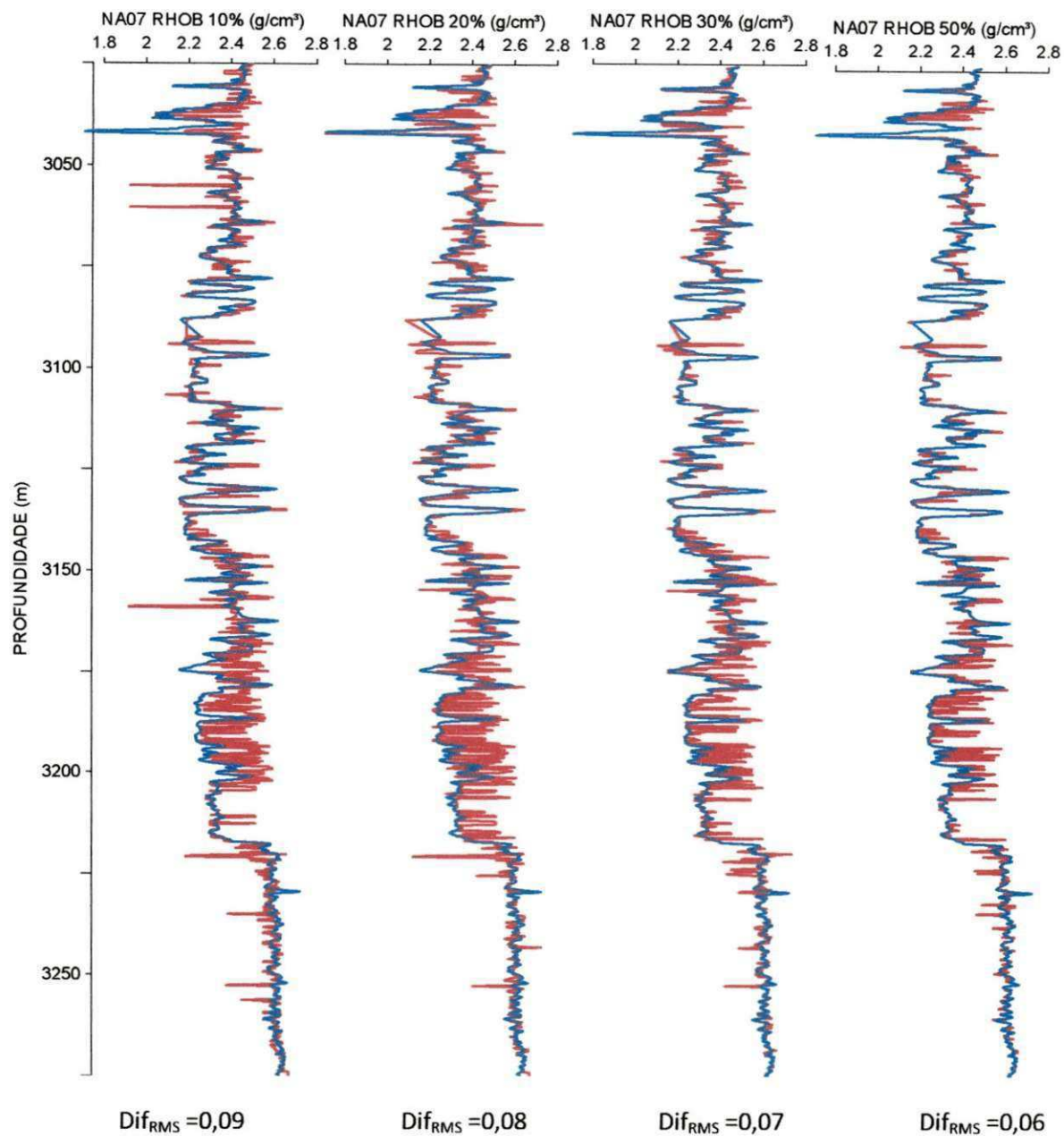


Figura 64 – Perfis de densidade do poço NA-07: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

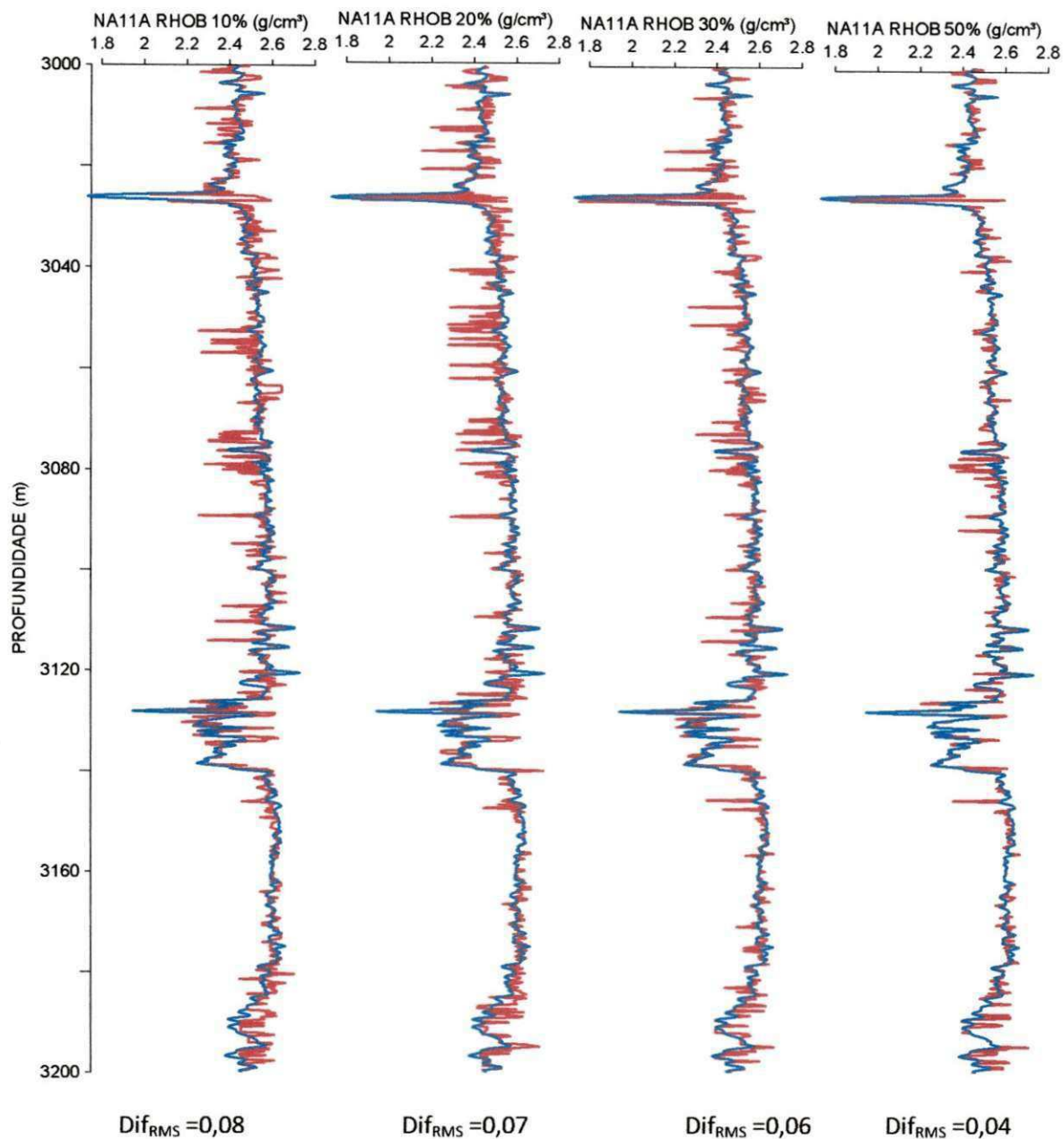


Figura 65 – Perfis de densidade do poço NA-11a: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

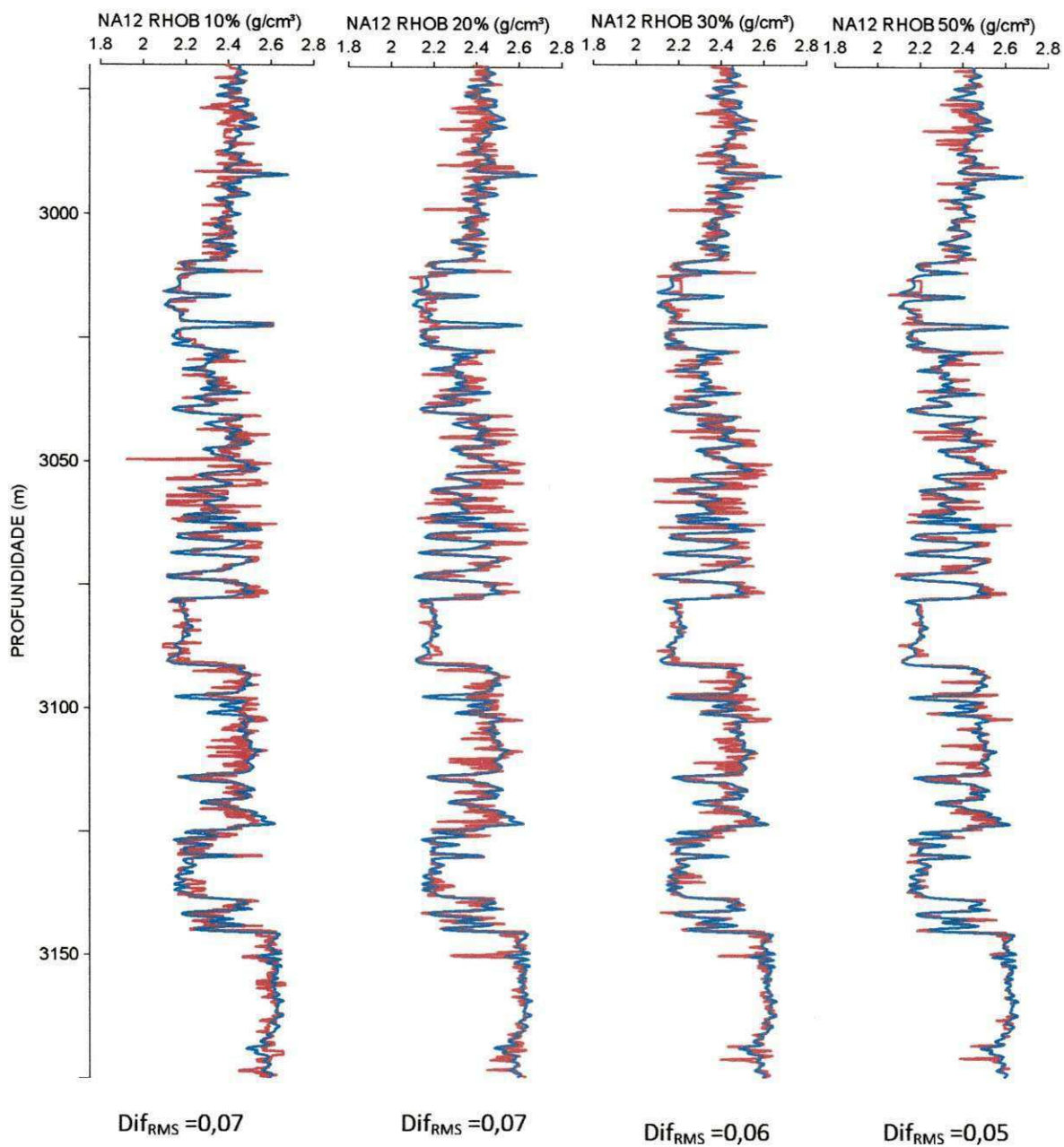


Figura 66 – Perfis de densidade do poço NA-12: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

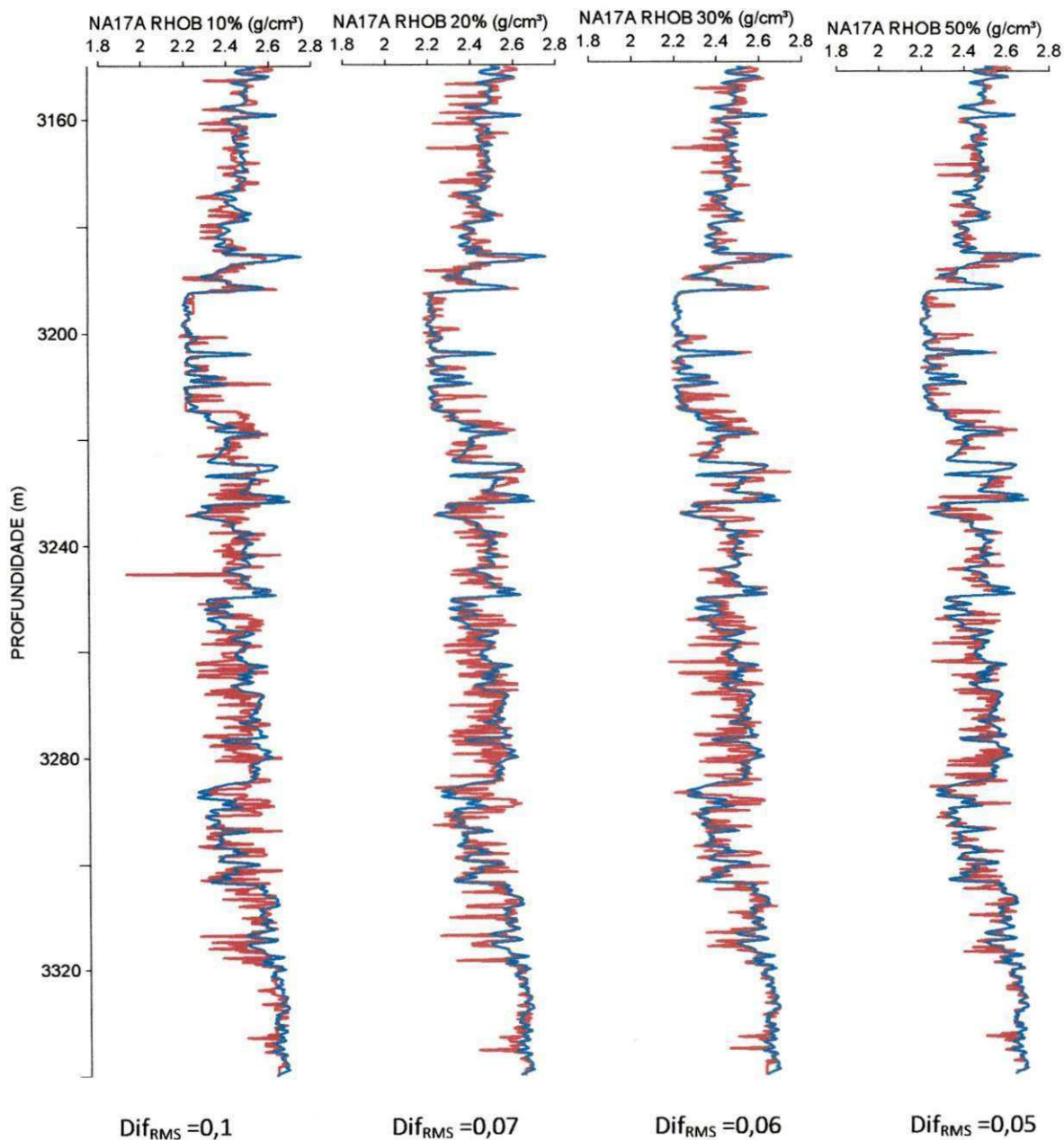


Figura 67 – Perfis de densidade do poço NA-17a: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

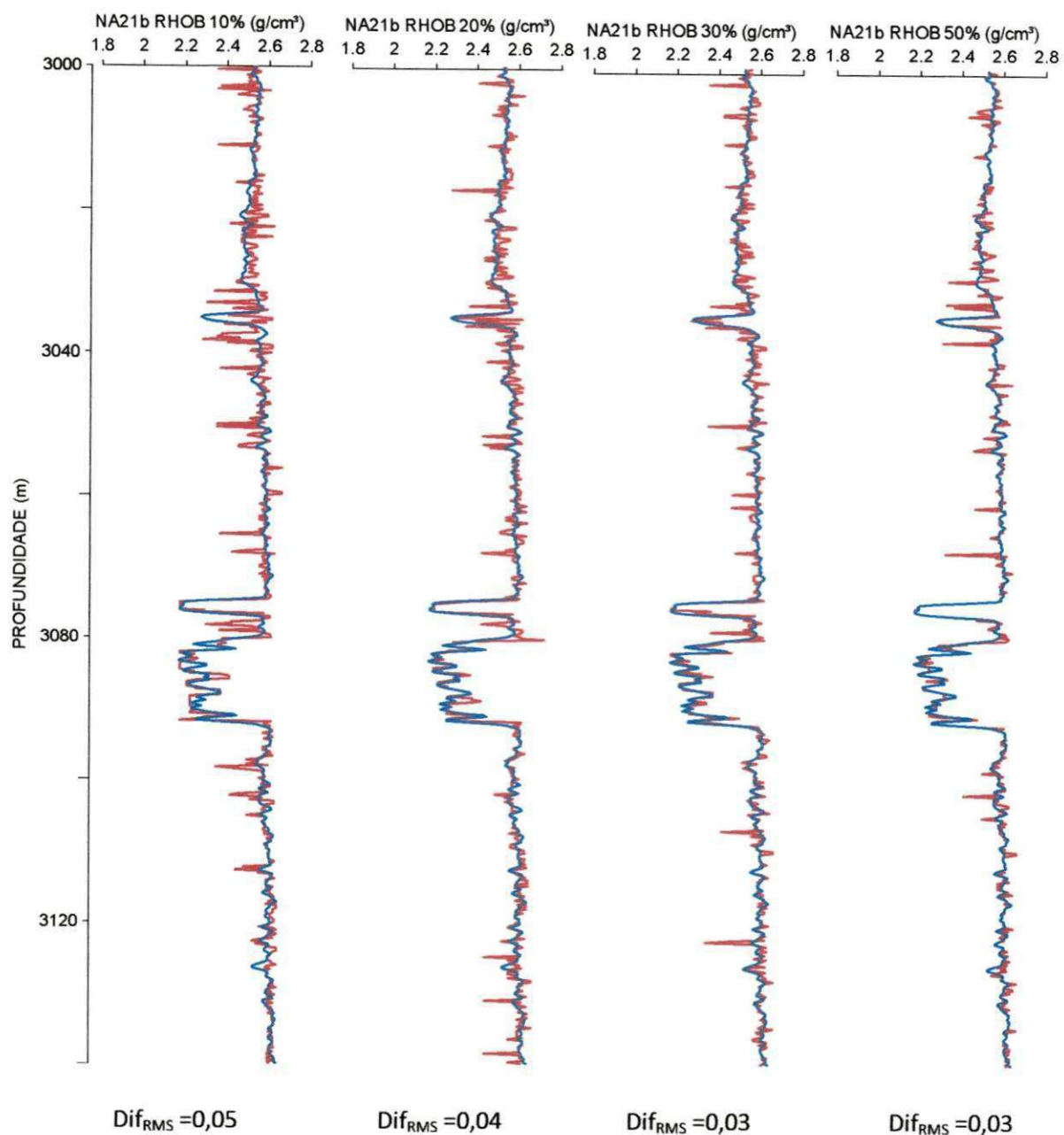


Figura 68 – Perfis de densidade do poço NA-21b: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

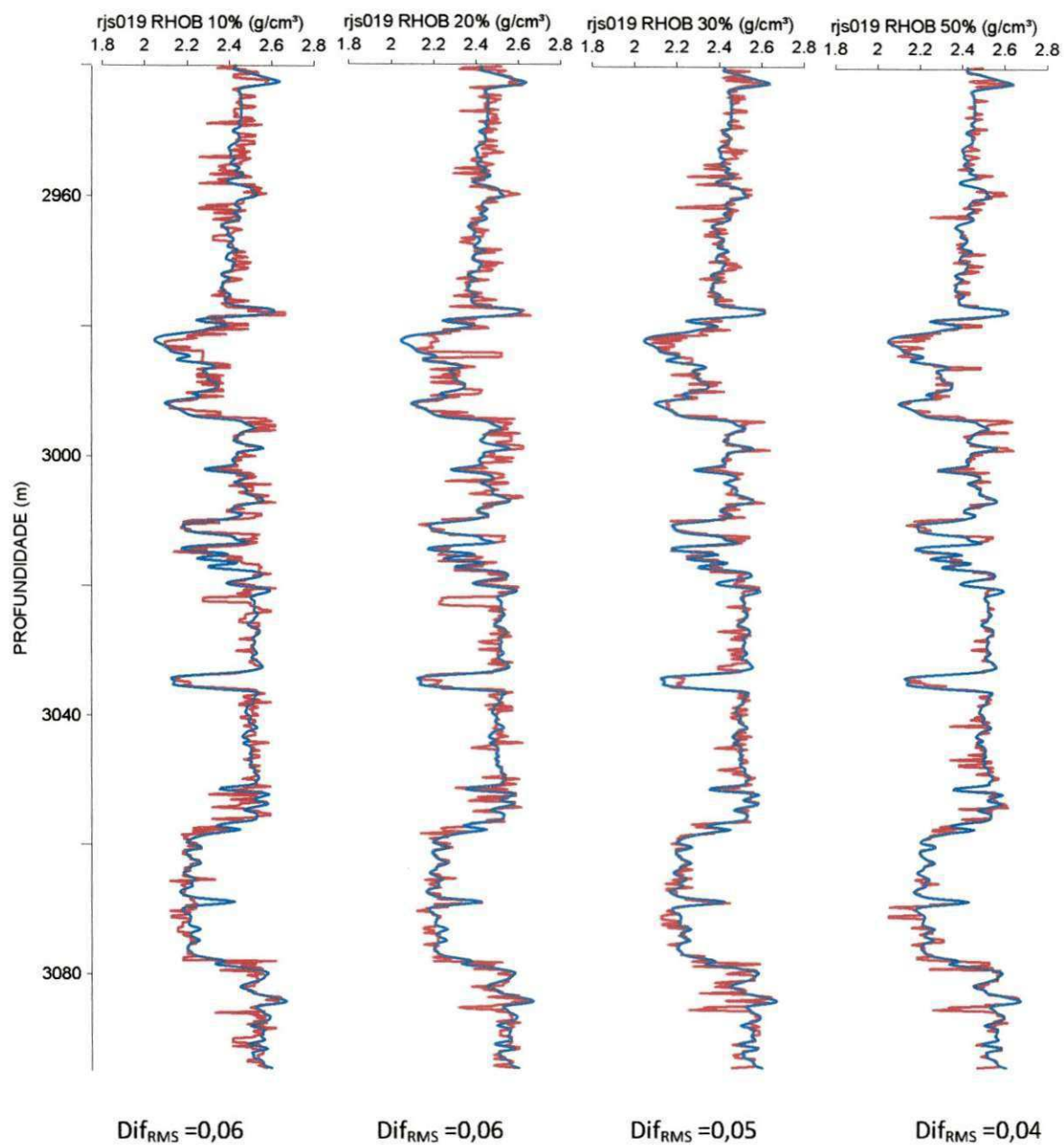


Figura 69 – Perfis de densidade do poço RJS-19: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

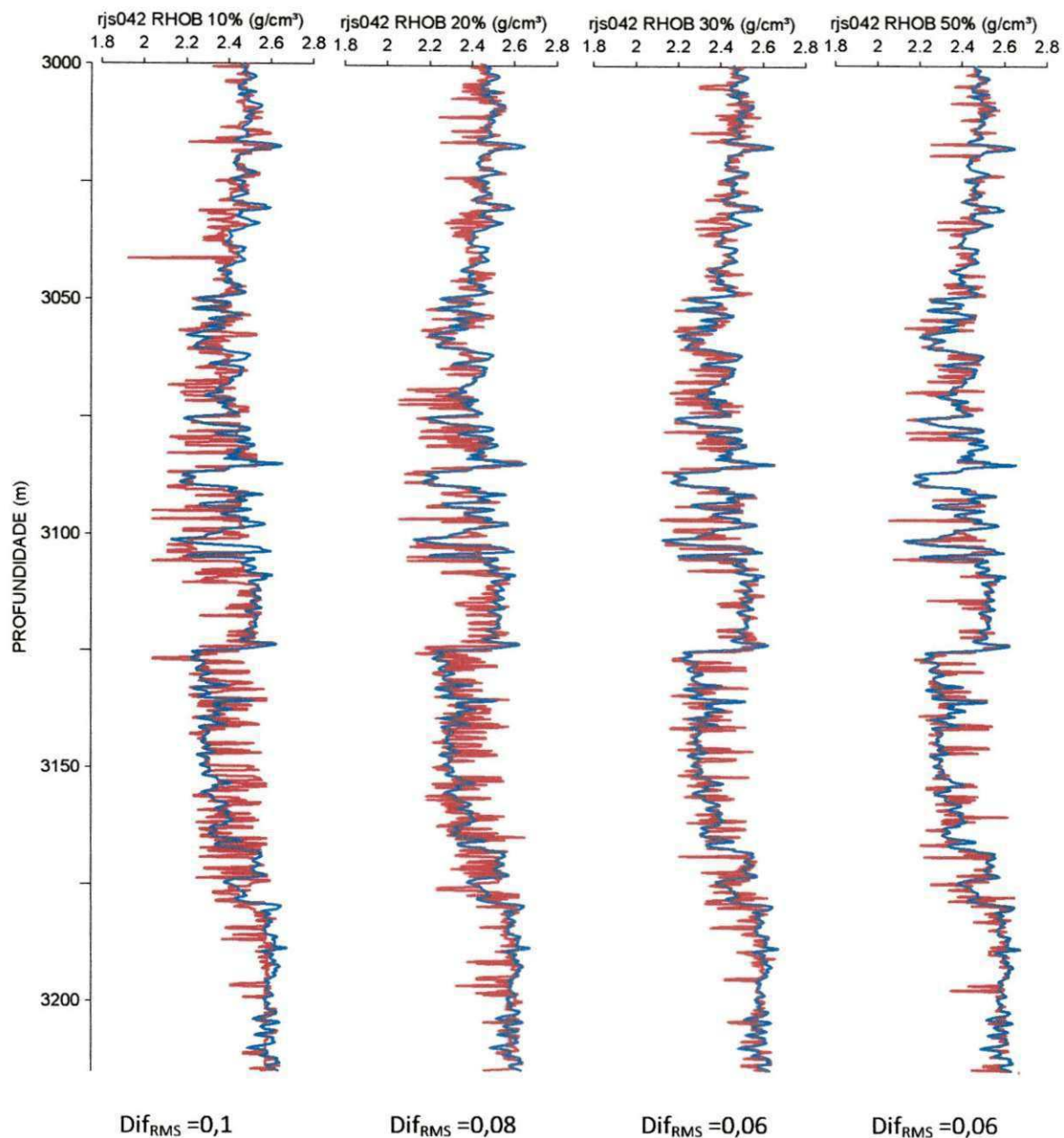


Figura 70 – Perfis de densidade do poço RJS-42: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

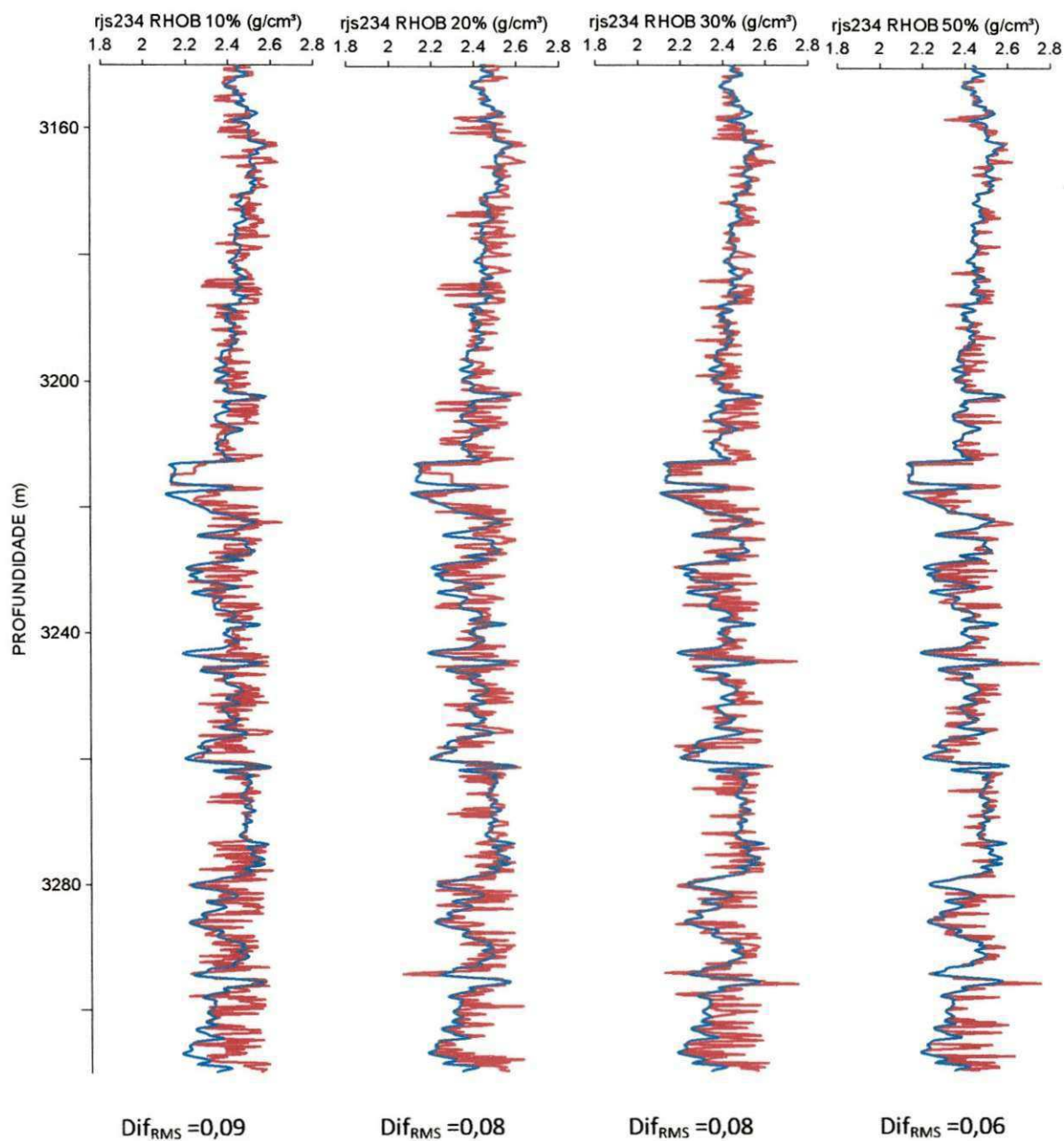


Figura 71 – Perfis de densidade do poço RJS-234: perfil registrado (curva azul) e perfis estimados (curvas vermelhas) para dados de treinamento com 10%, 20%, 30% e 50% dos dados totais, respectivamente.

3.2 - ANÁLISE DA PRECISÃO DOS RESULTADOS

3.2.1 - Diferença RMS

Com o objetivo de avaliar a precisão dos resultados alcançados nesta dissertação, a diferença média quadrática (RMS) entre os perfis estimados e os respectivos perfis registrados (equação 4) foi calculada para todos os perfis de todos os poços.

A Figura 72 apresenta a diferença RMS do perfil DT para todos os poços analisados. As diferenças máxima e mínima são de 10,5 $\mu\text{s}/\text{ft}$ e 2,8 $\mu\text{s}/\text{ft}$, respectivamente. Considerando que a escala do perfil sônico varia, em geral, entre 40 $\mu\text{s}/\text{ft}$ e 140 $\mu\text{s}/\text{ft}$, com valor médio de 90 $\mu\text{s}/\text{ft}$, estas diferenças RMS correspondem a 11,67% e 3,1% do valor médio, respectivamente.

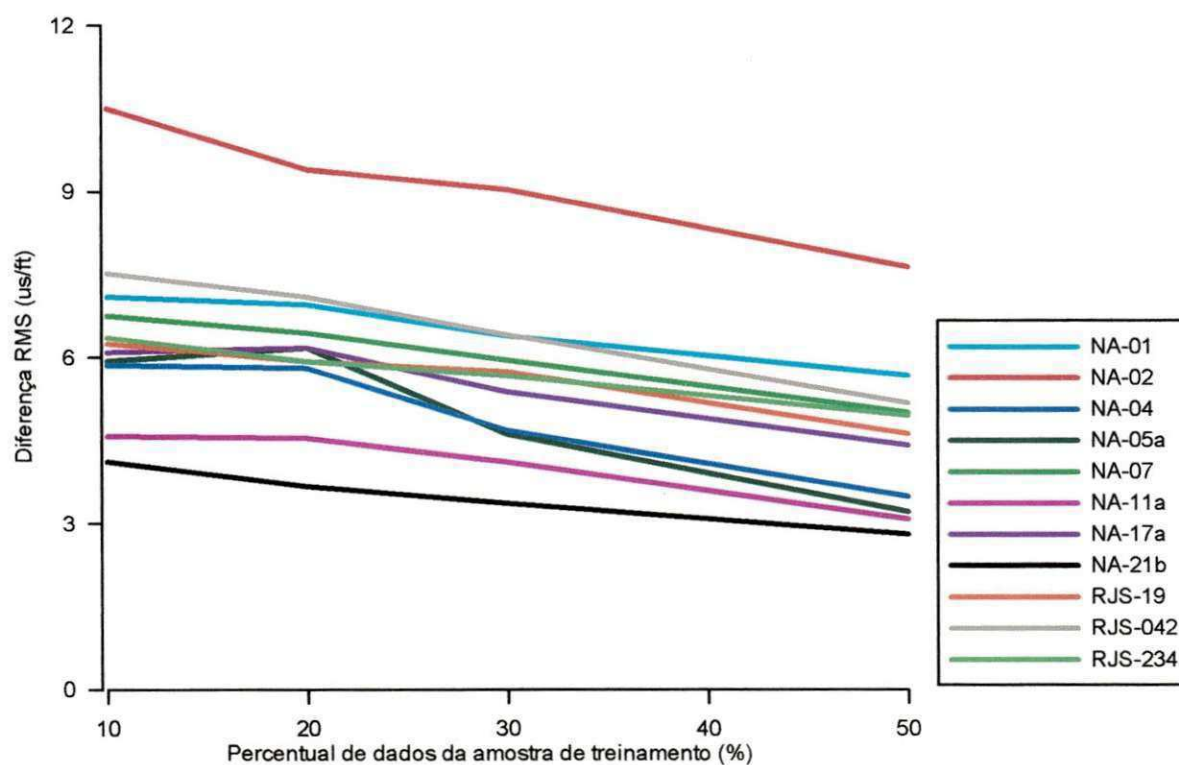


Figura 72 – Diferença RMS para o perfil DT de todos os poços analisados (exceto o poço NA-12 cujo perfil DT não foi registrado).

A Figura 73 apresenta a diferença RMS do perfil GR para todos os poços analisados. As diferenças RMS máxima e mínima são de 22,60 API e 7,29 API, respectivamente. Considerando que a escala do perfil de raios gama varia, em geral,

entre 0 e 150 API, com valor médio de 75 API, estas diferenças RMS correspondem a 30,1% e 9,7% do valor médio, respectivamente.

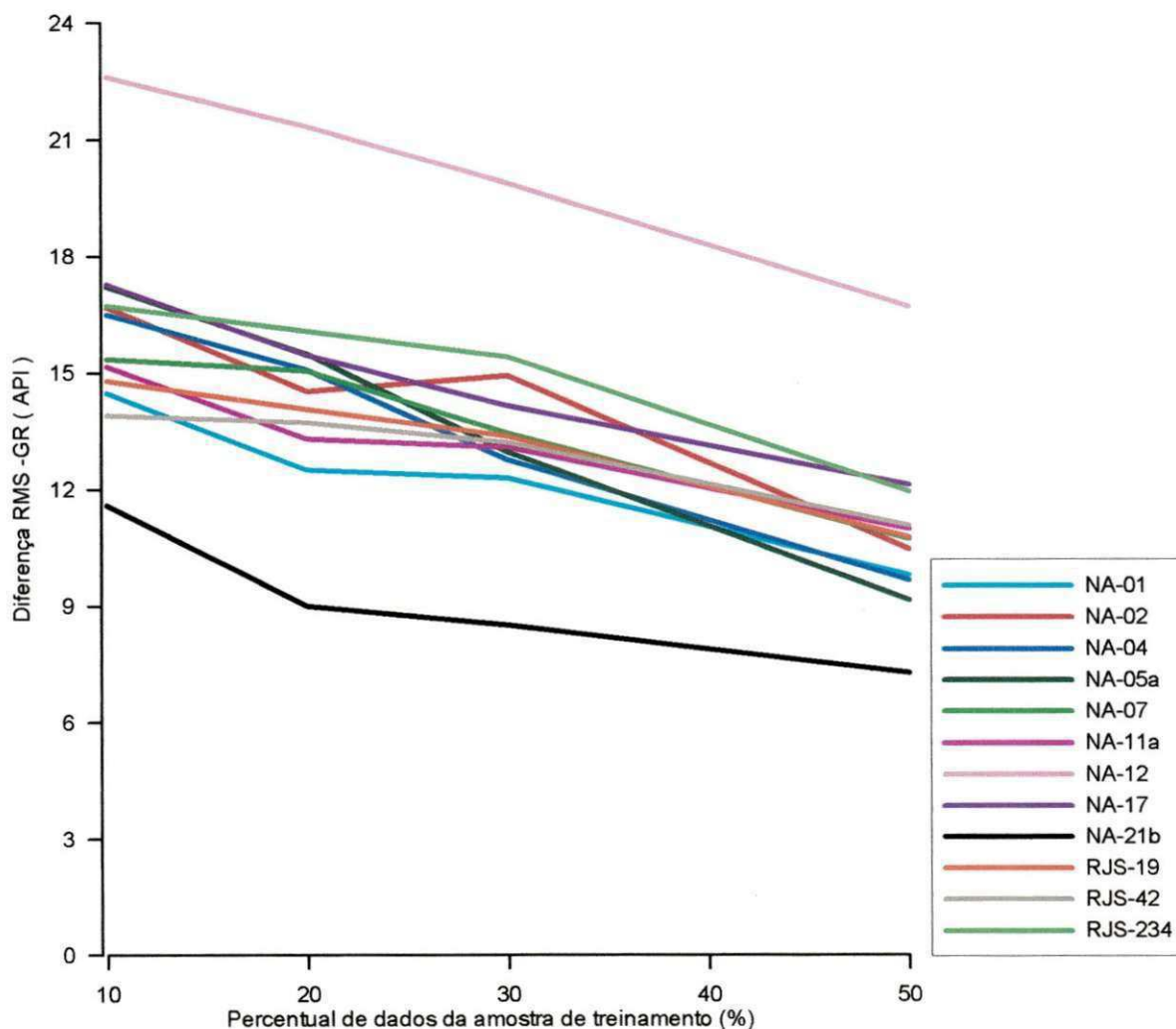


Figura 73 – Diferença RMS para o perfil GR de todos os poços analisados.

A Figura 74 apresenta a diferença RMS do perfil ILD para todos os poços analisados. As diferenças máxima e mínima são de 117,47 ohm.m e 4,4 ohm.m, respectivamente. Como o perfil de resistividade elétrica é apresentado em escala logarítmica, o cálculo da diferença RMS em percentual será realizado em termos dos logaritmos dos valores máximo, mínimo e médio. Considerando que os perfis de resistividade elétrica analisados apresentam valor médio da ordem de 3 ohm.m, estas diferenças RMS correspondem a 433,8% e 134,8% do valor médio, respectivamente.

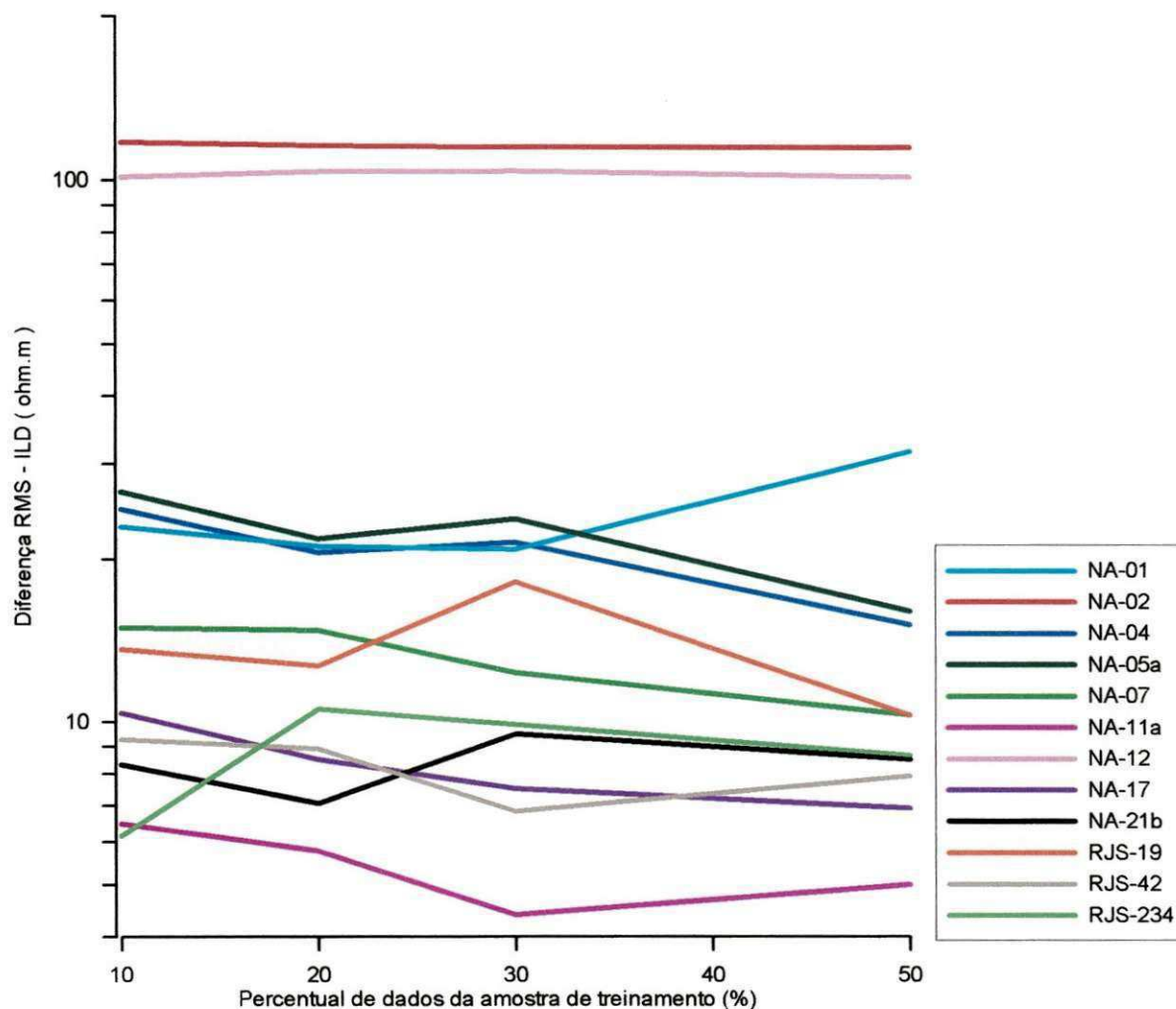


Figura 74 – Diferença RMS para o perfil ILD de todos os poços analisados.

A Figura 75 apresenta a diferença RMS do perfil NPHI para todos os poços analisados. As diferenças RMS máxima e mínima são de 6,2% e 1,95%, respectivamente. Considerando que a escala do perfil de porosidade de nêutrons varia, em geral, entre 0% e 45%, com valor médio de 22,5%, estas diferenças RMS correspondem a 26,7% e 8,7% do valor médio, respectivamente.

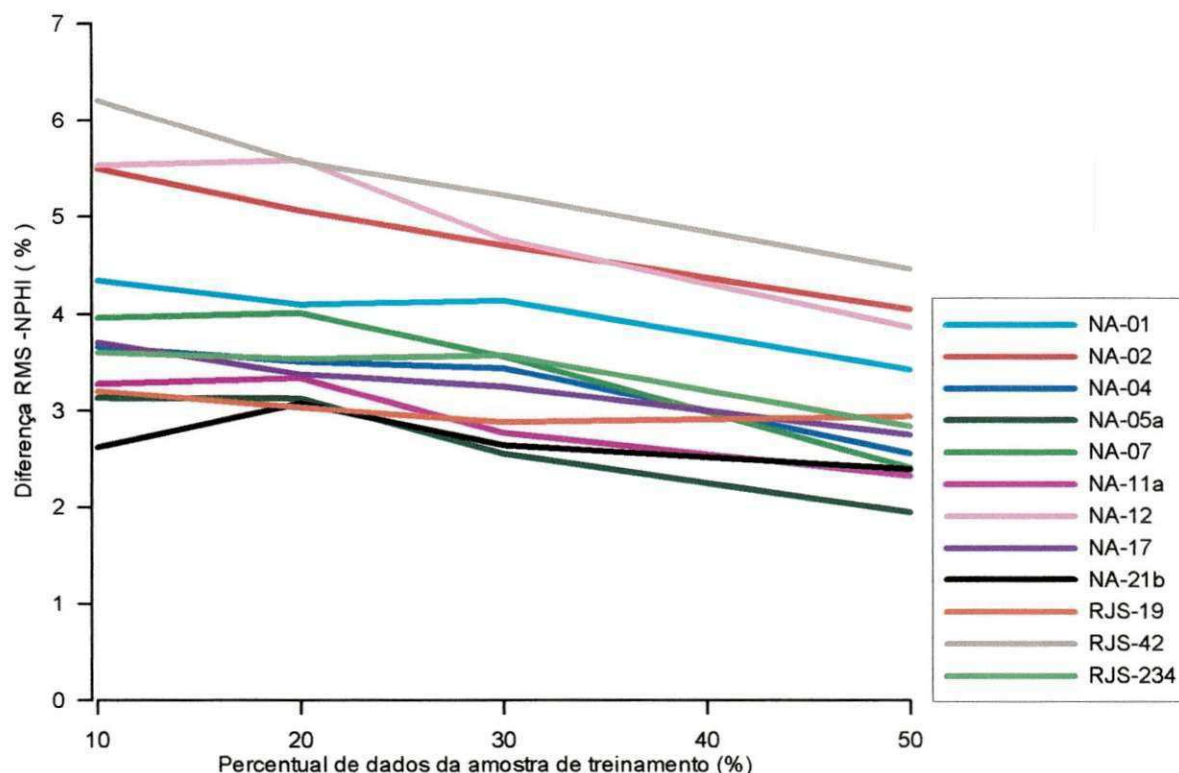


Figura 75 – Diferença RMS para o perfil NPHI de todos os poços analisados.

A Figura 76 apresenta a diferença RMS do perfil RHOB para todos os poços analisados. As diferenças RMS máxima e mínima são de 0,12 g/cm³ e 0,03 g/cm³, respectivamente. Considerando que a escala do perfil de densidade varia, em geral, entre 2 g/cm³ e 3 g/cm³, com valor médio de 2,5 g/cm³, estas diferenças RMS correspondem a 4,8% e 1,2% do valor médio, respectivamente.

A fim de consolidar a análise de precisão nas estimativas dos perfis faltantes, a Tabela 5 apresenta um quadro-resumo das diferenças RMS máximas e mínimas, em relação ao valor médio, obtidas para cada perfil.

Tabela 5 – Quadro-resumo das diferenças RMS máximas e mínimas, em relação ao valor médio, obtidas na predição de perfis faltantes.

Dif_RMS	DT	GR	ILD	NPHI	RHOB
Máxima (%)	11,7	30,1	433,8	26,7	4,8
Mínima (%)	3,1	9,7	134,8	8,7	1,2

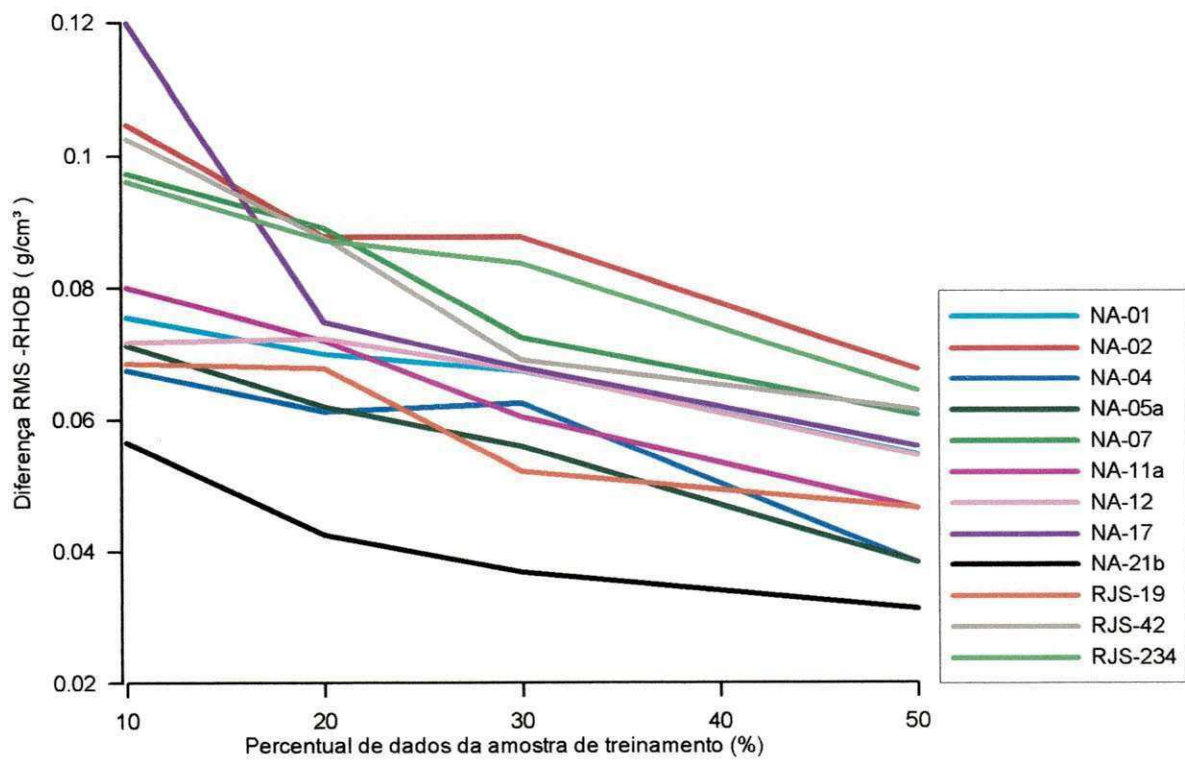


Figura 76 – Diferença RMS para o perfil RHOB de todos os poços analisados.

CAPÍTULO IV

4.1 - CONCLUSÕES

A aplicação do método KNN de análise multivariada se mostrou eficiente para a predição de perfis faltantes nos poços investigados neste trabalho. A inspeção visual das curvas estimadas, apresentadas no Capítulo de Resultados, mostra que na grande maioria dos perfis, a forma geral das curvas originais foi reproduzida nas curvas estimadas.

Dos resultados apresentados conclui-se que o aumento do tamanho da amostra de treinamento conduz à redução da diferença RMS entre a curva real e a estimada. No entanto, mesmo para amostras de treinamento tão pequenas quanto apenas 10% dos dados disponíveis, a diferença RMS máxima em relação à média foi de aproximadamente 30% (excluindo o perfil de resistividade elétrica, o qual possui escala logarítmica e não pode ser avaliado pelo mesmo critério). Considerando que as diferenças RMS em relação à média (para os perfis de escala linear) correspondem ao erro médio, os resultados da predição de todos eles apresentaram erros aceitáveis, ou seja, abaixo de 30% [Hair et al., 2005].

Para o perfil de resistividade elétrica (ILD), comparando as curvas estimadas com as curvas reais, conclui-se que não se obteve resultados aceitáveis para a predição deste perfil. A causa mais provável para o fato da predição do perfil ILD não ocorrer de forma satisfatória é que a técnica de classificação supervisionada KNN explora as similaridades e as redundâncias entre os vários perfis, os quais são afetados por características comuns da rocha. O perfil ILD, diferentemente dos demais, é afetado quase que exclusivamente pelas propriedades do fluido contido na rocha, de modo que pouca redundância existe entre este perfil e os demais.

Em termos de eficiência na predição de perfis faltantes, a técnica KNN mostrou resultados mais favoráveis para os perfis de densidade (RHOB) e sônico (DT), seguidos pelos perfis de porosidade de nêutrons (NPHI) e de raios gama (GR). É importante notar que estes perfis guardam certa redundância entre si, pois os três

primeiros são fortemente dependentes da porosidade da rocha, enquanto que os perfis de densidade e de raios gama são, em grande parte, indicativos do tipo litológico.

A diferença RMS calculada neste trabalho serve como um indicador quantitativo da precisão na estimativa de perfis faltantes. No entanto, não há uma correlação perfeita entre este indicador e o ajuste visual entre o perfil estimado e o perfil real, conforme pode ser observado nas figuras do capítulo de Resultados. Isto ocorre porque a diferença RMS representa um valor médio, de forma que, uma curva estimada que apresente um pequeno número de valores com grande diferença em relação à curva real pode resultar em uma diferença RMS maior do que aquela obtida para uma curva estimada que apresente um grande número de valores com pequena diferença em relação à curva real.

Há de se ressaltar que os perfis estimados apresentados neste trabalho não passaram por uma etapa de filtragem, de modo a suavizar as curvas através da minimização de possíveis ruídos de alta frequência. Estes ruídos podem existir como resultado da adoção do $K=1$ na técnica KNN. No entanto, adotar valores maiores que a unidade para o parâmetro K poderia suavizar demais as curvas estimadas, bem como não ser capaz de encontrar o número K de vizinhos mais próximos no espaço multidimensional, resultando na não determinação do valor estimado. De qualquer modo, graças ao grande número de amostras dos perfis geofísicos dos poços analisados, vê-se que mesmo adotando um K unitário, e sem filtragem dos dados, as diferenças RMS foram aceitáveis.

CAPITULO V

5.1 - REFERÊNCIAS BIBLIOGRÁFICAS

Albuquerque, M.A. Estabilidade em análise de agrupamento (*cluster analysis*). Dissertação de mestrado, UFRPE, Departamento de Física e Matemática. 2005.

Aldenderfer, M.S.; Blashfeld, R.K. Cluster analysis. Beverly Hills; Sage, 1984.

Almeida, G.; Melicio, F.; Pinheiro, A.M.G. Descritor do histograma para similaridade semântica de imagens. 5º Congresso Luso-Moçambicano de Engenharia; 2º Congresso de Engenharia de Moçambique, 2008.

Anderberg, M.R. Cluster analysis for applications. New York: Academic Press, 1973.

Anderson, T.W. An introduction to multivariate statistical analysis. John Wiley & Sons, New York, 1984.

Bucheb, J. A., Rodrigues, F. S., Aspectos Práticos Relativos ao Procedimento de Geração de Curvas Sintéticas de Perfis Por Meio De Regressão. *Boletim de Geociências da PETROBRAS*, v.11, n. 1/2, pp. 33-47, 1997.

Bussab, W.O.; Miazaki, E.S; Andrade, D. Introdução à análise de agrupamentos. São Paulo: Associação Brasileira de Estatística, 1990.

Couto, N.A.P., *Construção de Bancos de Dados Petrofísicos para a Determinação Automática de Litologias*. Dissertação de Mestrado., CG/UFGA, Belém, PA, Brasil, 1994.

Doveton, J.H. Log analysis of subsurface geology: concepts and computer methods. Wiley-Interscience. New York, 1986.

Everitt, B. Cluster analysis. London: Heinemann Educational Books, 1974.

Gama M.P. Bases da análise de agrupamentos ("Cluster Analysis"). Dissertação de Mestrado em Estatística e Métodos Quantitativos. Universidade de Brasília, 1980.

Gower, J.C.; Legendre, P. Metric and euclidean properties of dissimilarity coefficients. *Journal of Classification*, v. 3, p. 5-48, 1986.

Gunn, S.R. Support Vector Machines for Classification and Regression. Technical Report, University of Southampton, 1998.

Hair, J.F.Jr; Anderson, R.E.; Tatham, R.L.; Black, W.C. Análise Multivariada de Dados. Porto Alegre: Bookman, 2005.

http://www.apolo11.com/terremotos_wave.php. Acessado em 25/02/2011.

Mardia, A.K.V.; Kent. J.T.; Bibby, J.M. Multivariate Analysis. London: Academic Press, 1997.

Nitters G., Davies D.R., Epping W.J.M.. Discriminant Analysis and Neural Nets: Valuable Tools to Optimize Completion Practices. *In: SPE Drilling & Completions*, v. 10, pp. 127-133, 1995.

Reis, E. Estatística multivariada aplicada. Lisboa: Edições Silabo, 1997.

Ribeiro, F.S.A. Aplicação de análise multivariada na predição de perfis faltantes de poços do Campo de Namorado. Dissertação de Mestrado. Programa de Pós-graduação em Engenharia Civil, Universidade Federal do Rio de Janeiro, 2008.

Rosa, H.; Suslick, S.B.; Vidal, A.C.; Sakai, G.K. Caracterização de eletrofácies por meio de ferramentas estatísticas multivariadas. *Rev. Esc. Minas*, vol.61, no.4. Ouro Preto, 2008.

Soares, J.A. Um fluxo de trabalho para modelagem de eletrofácies com entrelaçamento de técnicas de classificação supervisionada e não-supervisionada. *Proceedings of the 9th International Congress of the Brazilian Geophysical Society*, 2005.

Sokal, R.R.; Michener, C.D. A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Buli*, 38(22) :1409-1438, 1958.

Serra, O. Fundamentals of well-log interpretation. Elsevier Science Publishing Company inc. New York, 1988.

Talaat, S.E.; Mostafa, M.S. An efficient technique for lithology classification. *IEEE Transactions on Geoscience and Remote Sensing*, 27(5):465-480, 1989.

Thomas, J.E. (Org.) Fundamentos de Engenharia de Petróleo - 2ª Ed.: Editora Interciência Ltda, 2001.

Tittman, J. Geophysical well logging. Academic Press, 1986.

Trauth M.H. MATLAB® Recipes for Earth Sciences. Springer-Verlag, Berlin Heidelberg, 2006.

Ward, J.H. Hierarchical grouping to optimize an objective function. Journal of American Statistical Association, v. 58, p. 236-244, 1963.

Wilbert, A., O Uso da Análise Discriminante na Caracterização Eletrofisiológica do Reservatório Carmópolis, no Campo de Riachuelo, Sergipe, Brasil. *In: Anais do 34º Congresso Brasileiro de Geologia*, v. 4, Goiânia, GO, 1986.