



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21c/2019/05.31.19.58-TDI

OBTENÇÃO DAS COMPONENTES DO TENSOR DE IMPEDÂNCIAS DO MÉTODO MAGNETOTELÚRICO USANDO A TRANSFORMADA DE KRAMERS-KRONIG

Marcos Banik de Pádua

Tese de Doutorado do Curso de Pós-Graduação em Geofísica Espacial/Ciências do Ambiente Solar-Terrestre, orientada pelo Dr. Antonio Lopes Padilha, aprovada em 26 de junho de 2019.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34R/3TD7FP8>>

INPE
São José dos Campos
2019

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GBDIR)

Serviço de Informação e Documentação (SESID)

CEP 12.227-010

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/7348

E-mail: pubtc@inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE - CEPPII (PORTARIA Nº 176/2018/SEI-INPE):

Presidente:

Dr. Marley Cavalcante de Lima Moscati - Centro de Previsão de Tempo e Estudos Climáticos (CGCPT)

Membros:

Dra. Carina Barros Mello - Coordenação de Laboratórios Associados (COCTE)

Dr. Alisson Dal Lago - Coordenação-Geral de Ciências Espaciais e Atmosféricas (CGCEA)

Dr. Evandro Albiach Branco - Centro de Ciência do Sistema Terrestre (COCST)

Dr. Evandro Marconi Rocco - Coordenação-Geral de Engenharia e Tecnologia Espacial (CGETE)

Dr. Hermann Johann Heinrich Kux - Coordenação-Geral de Observação da Terra (CGOBT)

Dra. Ieda Del Arco Sanches - Conselho de Pós-Graduação - (CPG)

Silvia Castro Marcelino - Serviço de Informação e Documentação (SESID)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon

Clayton Martins Pereira - Serviço de Informação e Documentação (SESID)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Simone Angélica Del Ducca Barbedo - Serviço de Informação e Documentação (SESID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SESID)

EDITORAÇÃO ELETRÔNICA:

Ivone Martins - Serviço de Informação e Documentação (SESID)

Cauê Silva Fróes - Serviço de Informação e Documentação (SESID)



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21c/2019/05.31.19.58-TDI

OBTENÇÃO DAS COMPONENTES DO TENSOR DE IMPEDÂNCIAS DO MÉTODO MAGNETOTELÚRICO USANDO A TRANSFORMADA DE KRAMERS-KRONIG

Marcos Banik de Pádua

Tese de Doutorado do Curso de Pós-Graduação em Geofísica Espacial/Ciências do Ambiente Solar-Terrestre, orientada pelo Dr. Antonio Lopes Padilha, aprovada em 26 de junho de 2019.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34R/3TD7FP8>>

INPE
São José dos Campos
2019

Dados Internacionais de Catalogação na Publicação (CIP)

Pádua, Marcos Banik de.

P136o Obtenção das componentes do tensor de impedâncias do método magnetotelúrico usando a transformada de Kramers-Kronig / Marcos Banik de Pádua. – São José dos Campos : INPE, 2019.

xx + 139 p. ; (sid.inpe.br/mtc-m21c/2019/05.31.19.58-TDI)

Tese (Doutorado em Geofísica Espacial/Ciências do Ambiente Solar-Terrestre) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2019.

Orientador : Dr. Antonio Lopes Padilha.

1. Sondagem magnetotelúrica. 2. Fórmula de Kramers-Kronig. 3. Análise de regressão. 4. Robustez (matemática). 5. Valores anômalos (estatística). I.Título.

CDU 550.837.2



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](https://creativecommons.org/licenses/by-nc/3.0/).

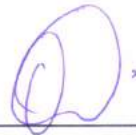
This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).

Aluno (a): **Marcos Banik de Pádua**

Título: "OBTENÇÃO DAS COMPONENTES DO TENSOR DE IMPEDÂNCIAS DO MÉTODO MAGNETOTELÚRICO USANDO A TRANSFORMADA DE KRAMERS-KRONIG"

Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de **Doutor(a)** em **Geofísica Espacial/Ciências do Ambiente Solar-Terrestre**

Dr. Cristiano Max Wrasse



Presidente / INPE / São José dos Campos - SP

() Participação por Vídeo - Conferência

Aprovado () Reprovado

Dr. Antonio Lopes Padilha

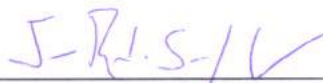


Orientador(a) / INPE / SJC Campos - SP

() Participação por Vídeo - Conferência

Aprovado () Reprovado

Dr. Jonas Rodrigues de Souza



Membro da Banca / INPE / SJC Campos - SP

() Participação por Vídeo - Conferência

Aprovado () Reprovado

Dr. Severino Luiz Guimarães Dutra



Membro da Banca / INPE / SJC Campos - SP

() Participação por Vídeo - Conferência

Aprovado () Reprovado

Este trabalho foi aprovado por:

() maioria simples

unanimidade

Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de **Doutor(a)** em

**Geofísica Espacial/Clências do Ambiente
Solar-Terrestre**

Dra. Andrea Cristina Lima Santos Matos

Membro da Banca / INPE / São José dos Campos - SP

Participação por Video - Conferência

Aprovado Reprovado

Dr. Sergio Luiz Fontes

Convidado(a) / ON / Rio de Janeiro - RJ

Participação por Video - Conferência

Aprovado Reprovado

Dr. Mauricio de Souza Bologna

Convidado(a) / IAG/USP / São Paulo - SP

Participação por Video - Conferência

Aprovado Reprovado

Este trabalho foi aprovado por:

maioria simples

unanimidade

São José dos Campos, 26 de junho de 2019

“Science is knowledge which we understand so well that we can teach it to a computer; and if we don’t fully understand something, it is an art to deal with it”.

DONALD KNUTH

*“Turing Award Lecture, Communications of the ACM 17 (12),
(December 1974), pp. 667–673”*,

AGRADECIMENTOS

Ao meu orientador, Antonio Lopes Padilha, pela orientação e por acreditar que os resultados viriam.

Ao meu irmão, Marcelo Banik de Pádua, pelo apoio. À Dra. Andréa Santos Matos por todo o estímulo, principalmente nas horas em que queria desistir.

Ao Instituto Nacional de Pesquisas Espaciais pelo suporte técnico e pelos dados e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pela bolsa concedida.

RESUMO

O método magnetotelúrico (MT) é utilizado para inferir a distribuição espacial da condutividade elétrica no interior da Terra, um parâmetro útil em diferentes estudos geofísicos. Os dados de campo MT são frequentemente contaminados por ruído eletromagnético (EM) e técnicas modernas de processamento e aquisição de dados têm sido propostas para eliminar os efeitos desses ruídos quando são coerentes e intermitentes. Esta tese propõe um método alternativo de processamento de dados baseado em processos de *reamostragem* e nas relações de dispersão de Kramers-Kronig para lidar com situações que envolvem esses ruídos EM extremos. O algoritmo desenvolvido foi inicialmente aplicado a dados sintéticos e mostrou eficiência para situações multidimensionais, embora apresentando limitações devido à função de extrapolação definida para períodos muito curtos e muito longos. Esses testes sintéticos também mostraram que a metodologia pode ser aplicada em situações tridimensionais (3D) para avaliar a confiabilidade da malha de discretização utilizada em algoritmos de inversão de dados. Por outro lado, a aplicação a dados experimentais contaminados por ruídos extremos (linhas de transmissão HVDC) não mostrou resultados adequados. Foi observado que o ruído coerente e contínuo contamina todo o espectro da amostra e não permite encontrar regiões onde o sinal natural predomine. Um procedimento alternativo, usando dados derivados do tradicional processamento robusto como entrada para a transformada Kramers-Kronig, foi bem sucedido na recuperação de funções de transferência de boa qualidade quando o ruído concentra-se exclusivamente na parte real ou imaginária da impedância. Esta última formulação pode ser usada para corrigir respostas MT, mesmo em situações 3D, quando o ruído EM estiver presente em apenas uma dessas componentes.

OBTAINING THE MAGNETOTELLURIC IMPEDANCE TENSOR COMPONENTS USING THE KRAMERS-KRONIG TRANSFORM

ABSTRACT

The magnetotelluric (MT) method is used to infer the electrical conductivity spatial distribution inside the Earth, a useful parameter for different geophysical studies. Field MT data are often contaminated by electromagnetic (EM) noise and modern data processing and acquisition techniques have been proposed to eliminate the effects of these noises when they are not coherent and intermittent. This thesis proposes an alternative data processing method based on resampling processes and in the Kramers-Kronig dispersion relations to deal with situations involving these extreme EM noises. The algorithm was initially applied to synthetic data and showed efficiency for multidimensional situations, although presenting limitations due to extrapolation function defined for very short and very long periods. These synthetic tests have also shown that the methodology can be applied in three-dimensional (3D) situations to evaluate the reliability of the discretization mesh used in algorithms of data inversion. On the other hand, application to experimental data contaminated by extreme noise (HVDC transmission lines) did not give adequate results. It was observed that coherent and continuous noise contaminates the entire sample spectra and does not allow finding regions where the natural signal prevails. An alternative procedure using data derived from the traditional robust processing as input to the Kramers-Kronig transform was successful in recovering good quality transfer functions when the noise is focused exclusively on the real or the imaginary part of the impedance. This last formulation can be used to correct MT responses even in 3D situations when the EM noise is present in only one of these components.

LISTA DE FIGURAS

	<u>Pág.</u>
3.1 Funções de transferência de estação MT coletada no estado de São Paulo (proximidades da cidade de Ourinhos) obtidas usando o processamento robusto de Egbert e Booker (1986). São mostradas as variações de resistividade aparente (ρ_{xy} e ρ_{yx}), em Ω m; fase (ϕ_{xy} e ϕ_{yx}), em graus; e as componentes reais e imaginárias de uma função derivada dos elementos do tensor de impedância ($Z_{xx}/\sqrt{\mu_0\omega}$, $Z_{xy}/\sqrt{\mu_0\omega}$, $Z_{yx}/\sqrt{\mu_0\omega}$ e $Z_{yy}/\sqrt{\mu_0\omega}$). As elipses vermelhas são discutidas no texto.	24
3.2 Funções de transferência de estação MT coletada no estado de São Paulo (proximidades da cidade de Sorocaba). São mostradas as variações de resistividade aparente (ρ_{xy} e ρ_{yx}), em Ω m, fase (ϕ_{xy} e ϕ_{yx}), em graus, componentes reais e imaginárias de uma função derivada dos elementos do tensor de impedância (Z_{xx} , Z_{xy} , Z_{yx} e Z_{yy}). O quadrado vermelho é discutido no texto.	26
3.3 Localização das estações MT analisadas. As linhas azuis representam as atuais linhas HDVC e as linhas vermelhas as futuras linhas HDVC.	30
4.1 Comparação da transformada <i>kk-ri</i> com resultados sintéticos para o modelo unidimensional. A discriminação das funções mostradas no gráfico é apresentada na Figura 3.1. Os pontos azuis foram obtidos por meio de modelagem unidimensional e os pontos em vermelho foram obtidos usando a Equação 4.3b. Para poder distinguir os casos em que as rotinas que calculam a transformada obtiveram resultados com melhor precisão daqueles em que as rotinas mostraram resultados mais imprecisos, os dados do modelo estão em tamanho maior.	37
4.2 Transformada <i>kk-ri</i> comparada com dados sintéticos para o modelo unidimensional usando períodos estendidos. As cores seguem o mesmo padrão da Figura 4.1.	38
4.3 Transformada <i>kk-ir</i> comparada com dados sintéticos para o modelo unidimensional. As cores seguem o mesmo padrão da Figura 4.1.	39
4.4 Transformada <i>kk-ir</i> comparada com dados sintéticos para o modelo unidimensional usando períodos estendidos. As cores seguem o mesmo padrão da Figura 4.1.	40

4.5	Visão superior para o modelo 3D sintético. O ponto representa a posição da estação ($x = -285.888676, y = -178.568076$) cujos dados sintéticos são utilizados e as linhas representam os cortes apresentados nas Figuras 4.6 e 4.7	41
4.6	Corte vertical do plano em $x = 260$ para o modelo 3D sintético. O triângulo na superfície marca a posição da estação.	42
4.7	Corte vertical do plano em $y = 160$ para o modelo 3D sintético. O triângulo na superfície marca a posição da estação.	42
4.8	Transformada $kk-ri$ comparada com dados sintéticos para o modelo tridimensional fornecido junto com o código <i>ModEM</i> . As cores seguem o mesmo padrão da Figura 4.3.	44
4.9	Transformada $kk-ir$ comparada com dados sintéticos para o modelo tridimensional fornecido junto com o código <i>ModEM</i> . As cores seguem o mesmo padrão da Figura 4.3.	45
4.10	Dados sintéticos para o modelo 3D original fornecido com o código <i>ModEM</i> , em vermelho, expandidos para períodos entre 0,231 s e 5 336 800 s, em azul.	47
4.11	Dados sintéticos para o modelo 3D original fornecido com o código <i>ModEM</i> , em vermelho, comparados com os gerados pelo modelo estendido verticalmente (inclusão de 4 novas camadas) para períodos entre 0,231 s e 5 336 800 s, em azul.	48
4.12	Transformada $kk-ri$ comparada com dados sintéticos expandidos para períodos mais curtos e mais longos no modelo tridimensional.	50
4.13	Transformada $kk-ir$ comparada com dados sintéticos expandidos para períodos mais curtos e mais longos no modelo tridimensional.	51
4.14	Transformada $kk-ri$ comparada com dados sintéticos expandidos para períodos mais curtos e mais longos no modelo 3D, com maior discretização da parte superficial do modelo.	52
4.15	Transformada $kk-ir$ comparada com dados sintéticos expandidos para períodos mais curtos e mais longos no modelo 3D, com maior discretização da parte superficial do modelo.	53
5.1	Funções de transferência de estação MT coletada no estado do Paraná. São mostradas as variações de resistividade aparente (ρ_{xy} e ρ_{yx} , em Ω m), fase (ϕ_{xy} e ϕ_{yx} , em graus), e as componentes reais e imaginárias de funções derivadas dos elementos do tensor de impedância (Z_{xx} , Z_{xy} , Z_{yx} e Z_{yy}).	56

5.2	Transformada de Kramers-Kronig para a estação <i>apg023a</i> usando a parte imaginária como entrada. As funções de transferência em azul foram obtidas por meio de regressão robusta com referência remota enquanto aquelas em vermelho foram obtidos usando a Equação 4.3a.	58
5.3	Transformada de Kramers-Kronig para a estação <i>apg023a</i> usando a parte real como entrada. As funções de transferência em azul foram obtidas por meio de regressão robusta com referência remota enquanto aquelas em vermelho foram obtidos usando a Equação 4.3b.	59
5.4	Transformada de Kramers-Kronig para a estação <i>apg023a</i> usando um procedimento iterativo. As funções de transferência em azul foram obtidas por meio de regressão robusta com referência remota enquanto aquelas em vermelho foram obtidas iterativamente usando as Equações 4.3a e 4.3b.	62
5.5	Distribuição das amostras da componente Z_{yx} para vários períodos. A linha superior contém o período de cada amostra e a coluna da direita contém o número de elementos de cada segmento utilizado para calcular o espectro. O eixo horizontal de cada caixa representa a parte real das impedâncias e o vertical a parte imaginária. A cor de cada hexágono é determinada pelo número de impedâncias contida no hexágono, conforme indicado pela contagem no lado direito da figura.	67
5.6	Distribuição das amostras da componente Z_{yx} em escala reduzida para o período 2,2857 s e $n = 128$ da Figura 5.5. O ponto azul corresponde ao valor obtido através da regressão robusta com referência remota (EGBERT; BOOKER, 1986), enquanto o ponto vermelho corresponde ao resultado do procedimento iterativo das transformadas de Kramers-Kronig mostrado na Figura 5.4.	68
5.7	Resumo do método <i>dbscan</i> . As cores dos pontos são definidas como: ● são os pontos originais do conjunto; ● é o ponto que esta sendo classificado; ● são os pontos dentro do alcance do ponto a ser classificado; ● são pontos classificados como núcleo de um aglomerado; ● são pontos classificados como borda de um aglomerado; ○ são pontos classificados como ruído.	70

5.8	Distribuição das distâncias dos primeiros vizinhos $kNN\{Z\}$ para vários comprimentos n de segmentos da série temporal dos dados EM. A distribuição foi calculada a partir do método <i>dbscan</i> sobre os conjuntos de Z_{yx} da Figura 5.5 com $T = 2,2857$ s. Os limites inferior e superior das caixas estabelecem os valores do primeiro ($Q1$) e terceiro ($Q3$) quartis da distribuição. A linha no interior das caixas representa o segundo quartil (mediana). As linhas que se estendem a partir da caixa são conhecidas como bigodes e podem ser interpretadas como limites para os <i>outliers</i> da distribuição. O valor do bigode superior é o maior ponto x da distribuição com $x \leq Q3 + 1.5(Q3 - Q1)$. O valor do bigode inferior é o menor ponto x da distribuição com $x \geq Q1 - 1.5(Q3 - Q1)$	72
5.9	Aglomerados obtidos a partir da utilização do método <i>dbscan</i> sobre as amostras da Figura 5.5. Cada aglomerado é representado por um polígono convexo de cor diferente.	73
5.10	Processo de amostragem da componente Z_{yx} na estação <i>apg023a</i> . Os gráficos no topo da figura contém os centroides para os sete períodos anômalos. Para cada um desses períodos é escolhido um dos centroides, o qual será utilizado para obter o vetor \mathbf{Z}_{yx}^a , exibido como pontos coloridos nos gráficos inferiores. Os pontos pretos nos gráficos central e inferior representam os pontos de períodos robustos $\mathbf{Z}_{yx}^{(r)}$. As curvas contínuas são obtidas por meio de <i>splines</i> do vetor de estimativas $\mathbf{Z}_{yx}^{(e)} = \mathbf{Z}_{yx}^{(r)} \cup \mathbf{Z}_{yx}^{(a)}$	75
5.11	Aplicação das transformadas <i>kk-ri</i> e <i>kk-ir</i> nos valores da impedância Z_{yx} da Figura 5.10. As funções $\Re(Z_{yx}^{(e)}(\omega))$ e $\Im(Z_{yx}^{(e)}(\omega))$ daquela figura estão representadas nos gráficos do lado esquerdo, enquanto o resultado das transformadas de Kramers-Kronig são apresentados no lado direito.	77
5.12	Cálculo dos resíduos para as componentes reais e imaginárias do elemento Z_{yx} da estação <i>apg023a</i> . Os gráficos a esquerda contém os resultados para as partes real e imaginária das amostras (em azul) e das transformadas de Kramers-Kronig (em vermelho), enquanto os gráficos a direita mostram os módulos dos resíduos entre esses valores.	79
5.13	Transformada de Kramers-Kronig para a estação <i>apg023a</i> usando a parte imaginária como entrada e a técnica mista de regressão robusta e reamostragem. As funções de transferência em azul foram obtidas por meio de regressão robusta enquanto aquelas em vermelho foram obtidos usando o método de reamostragem e a transformada <i>kk-ir</i> dada pela Equação 4.3a.	81

6.1	Saída do algoritmo Rho^+ comparada com dados sintéticos expandidos para períodos mais curtos e mais longos no modelo tridimensional. . .	84
6.2	Saída do algoritmo Rho^+ sobre os dados da estação <i>ser009a</i> da região Nordeste do Brasil processados pelo código robusto de Egbert e Booker (1986).	86
6.3	Transformada de Kramers-Kronig para a estação <i>ser009a</i> da região Nordeste do Brasil usando a parte real como entrada. As funções de transferência em azul foram obtidas por meio de regressão robusta enquanto aquelas em vermelho foram obtidos usando a Equação 4.3b.	87
6.4	Transformada de Kramers-Kronig para a estação <i>ser009a</i> da região Nordeste do Brasil usando a parte imaginária como entrada. As funções de transferência em azul foram obtidas por meio de regressão robusta enquanto aquelas em vermelho foram obtidos usando a Equação 4.3a.	88
6.5	Transformada <i>kk-ir</i> comparada com dados preditos pelo modelo inverso 3D para a estação condutora <i>bobc12a</i> na região Nordeste do Brasil.	94
6.6	Transformada <i>kk-ri</i> comparada com dados preditos pelo modelo inverso 3D para a estação condutora <i>bobc12a</i> na região Nordeste do Brasil.	95
6.7	Transformada <i>kk-ir</i> comparada com dados preditos pelo modelo inverso 3D para a estação resistiva <i>bobc18a</i> na região Nordeste do Brasil.	96
6.8	Transformada <i>kk-ri</i> comparada com dados preditos pelo modelo inverso 3D para a estação resistiva <i>bobc18a</i> na região Nordeste do Brasil.	97
6.9	Transformada de Kramers-Kronig para a estação <i>pru036a</i> da Bacia do Paraná, usando a parte imaginária como entrada. As funções de transferência em azul foram obtidas por meio da regressão robusta enquanto aquelas em vermelho foram obtidas usando a Equação 4.3a.	100
B.1	As quatro primeiras sequencias DPSS com $n = 32$ elementos e semi-largura de banda $nw = 4$	126
B.2	Auto-espectro das quatro primeiras sequencias DPSS da Figura B.1. . .	128
B.3	Auto-espectro das quatro primeiras sequencias DPSS da Figura B.2 interpolado utilizando 2^{15} zeros.	129

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO	1
2 O MÉTODO MAGNETOTELÚRICO	5
2.1 Fontes primárias das variações geomagnéticas	6
2.2 Difusão eletromagnética na Terra	7
2.3 Resistividade aparente, fase e profundidade pelicular	9
2.4 Impedância MT sobre uma Terra unidimensional	11
2.4.1 Limite de altas frequências	12
2.4.2 Limite de baixas frequências	12
2.4.3 Embasamento condutivo $\sigma_2 \gg \sigma_1$	12
2.4.4 Embasamento resistivo $\sigma_2 \ll \sigma_1$	13
2.5 Impedância MT de uma terra bidimensional	13
2.6 O tensor de impedâncias	14
2.6.1 Terra bidimensional	14
2.6.2 Terra tridimensional	15
2.7 Vetores de indução	15
2.8 Aquisição e processamento dos dados	16
2.9 Efeito de ruídos	19
3 O PROBLEMA PROPOSTO: RUÍDOS EM EXTREMOS NAS FUN- ÇÕES DE TRANSFERÊNCIA MT	23
4 A TRANSFORMADA DE KRAMERS-KRONIG	31
4.1 Implementação da transformada de Kramers-Kronig	32
4.2 Cálculo numérico da transformada de Kramers-Kronig	34
4.3 Validação do algoritmo desenvolvido - testes em dados sintéticos	35
4.3.1 Aplicação ao caso unidimensional	35
4.3.2 Aplicação ao caso tridimensional	41
5 APLICAÇÃO A DADOS EXPERIMENTAIS	55
5.1 Resultado da transformada de Kramers-Kronig usando como entrada a regressão robusta	57

5.2	Resultado da transformada de Kramers-Kronig usando uma entrada mista de regressão robusta e reamostragem	63
5.2.1	Análise espectral	64
5.2.2	Reamostragem	65
5.2.3	Redução de dados por meio de aglomeração	68
5.2.4	Construção de $Z_{ij}(\omega)$ a partir dos aglomerados.	74
5.2.5	Estimativa da transformada de Kramers-Kronig	76
5.2.6	Minimização dos resíduos	76
5.2.7	Resultado do uso misto de regressão robusta e reamostragem	80
6	DISCUSSÕES SOBRE O USO PRÁTICO DO ALGORITMO DESENVOLVIDO	83
6.1	Relação com o algoritmo Rho^+ e edição de respostas MT	83
6.2	Discretização do semiespaço infinito	90
6.3	Caso de ruídos extremos	98
7	CONCLUSÕES	101
	REFERÊNCIAS BIBLIOGRÁFICAS	105
	APÊNDICE A – BIBLIOTECAS EXTERNAS E PARÂMETROS UTILIZADOS PELAS ROTINAS	119
	APÊNDICE B – PESQUISA REPRODUZÍVEL	121
B.1	Implementação da análise reproduzível	123
B.2	Análise de dados remota	130
B.3	Dependências externas	136
B.4	Configuração do emacs	138

1 INTRODUÇÃO

O método magnetotelúrico (MT) é uma técnica de prospecção geofísica que busca obter a distribuição da condutividade elétrica no interior da Terra através da medição simultânea das variações temporais de campos elétricos e magnéticos na superfície da Terra. Sua principal característica é a versatilidade que permite investigar variações de condutividade desde profundidades rasas (alguns metros) até grandes profundidades (centenas de quilômetros), na dependência da frequência do sinal analisado e da condutividade elétrica do meio. Em altas frequências (1 Hz até centenas de milhares de hertz), as variações do campo são geradas por interações no guia de onda Terra-ionosfera, com a fonte de energia eletromagnética (EM) sendo a ocorrência de raios. As variações de baixa frequência (1 Hz até dezenas de milhares de segundos) devem-se a interações entre o vento solar e a magnetosfera da Terra. Essa capacidade de sondar diferentes profundidades com pequenas variações na instrumentação utilizada pode ser atribuída à maneira difusiva com que a energia EM se dissipa na Terra (SIMPSON; BAHR, 2005; CHAVE; JONES, 2012a).

O processamento inicial dos dados MT objetiva derivar funções de transferência entre os campos EM medidos, os quais são frequentemente contaminadas por ruídos artificiais ou naturais. Essas funções de transferência são geralmente apresentadas na forma de um tensor de impedâncias (ou tensor MT), o qual contém informação sobre a distribuição de condutividade do subsolo, pois se correlacionam por meio de processos de indução eletromagnética. Como as relações entre os campos elétrico e magnético no interior da Terra são lineares, os métodos utilizados para obter o tensor MT foram inicialmente baseados em regressão linear. Porém, a regressão linear clássica (método dos mínimos quadrados) (SIMS et al., 1971) mostrou-se ineficiente nesse processo devido à comum ocorrência de valores anômalos (*outliers*) nos espectros do campo EM. Para mitigar esse problema foram propostos uma série de métodos baseados em regressões lineares robustas, onde a influência dos valores anômalos é reduzida através de pesos adaptativos (CHAVE; THOMSON, 1989; SUTARNO; VOZOFF, 1989; SUTARNO, 2008).

Mesmo esses métodos robustos fornecem resultados pouco confiáveis em frequências afetadas por ruídos coerentes, nos quais os ruídos que afetam tanto o campo elétrico como o campo magnético estão correlacionados (FONTES et al., 1988). Sinais eletromagnéticos antrópicos constituem-se um exemplo típico de ruídos coerentes (ESCALAS et al., 2013) e difíceis de serem tratados (BANKS, 1998). Técnicas paliativas foram desenvolvidas para lidar com esse tipo de ruído baseadas em

procedimentos de seleção ou rejeição dos sinais contaminados a partir de algumas propriedades, tais como a coerência ou a polaridade do sinal (JONES; JÖDICKE, 1984; EGBERT; LIVELYBROOKS, 1996; GARCIA; JONES, 2008; ESCALAS et al., 2013). Nos casos em que é possível realizar duas medidas simultâneas dos campos geomagnéticos em localidades diferentes, a técnica de referência remota é utilizada para remover ruídos EM locais que ocorram apenas na vizinhança da área em estudo (GOUBAU et al., 1978; GAMBLE et al., 1979). Também é possível unir a técnica de referência remota à seleção do sinal EM, como proposto em (EGBERT; BOOKER, 1986; LARSEN et al., 1996; RITTER et al., 1998; OETTINGER et al., 2001).

O nível de proteção a valores anômalos fornecida por todos esses métodos baseados na regressão linear robusta é identificada pelo ponto de ruptura (*breakdown point*). O maior valor possível para o ponto de ruptura é de 50 % (MARONNA et al., 2006, p. 58; ROUSSEEUW; LEROY, 2003, p. 14). Isso quer dizer que, na presença de mais de 50 % de valores anômalos, a regressão linear passa a tratar os valores anômalos como se fossem o sinal. Como consequência, nos casos em que os ruídos coerentes são persistentes é impossível obter estimativas confiáveis dos elementos do tensor MT através desses métodos. Essa situação é historicamente crítica em diversos países devido à disseminação de linhas de transmissão de energia elétrica, ferrovias operadas por corrente contínua e cercas elétricas. No caso do Brasil, uma fonte comum de intenso ruído EM antrópico coerente são as linhas de transmissão de energia elétrica utilizando corrente contínua. Essa forma de transmissão de energia (*high voltage direct current* - HVDC) é uma das técnicas mais eficientes de transmissão de energia a grandes distâncias, especialmente no caso em que as áreas de demanda e de geração de energia são separadas por centenas a milhares de quilômetros. A perspectiva de expansão dessas linhas no país em futuro imediato torna urgente a necessidade de desenvolver métodos capazes de estimar o tensor MT na presença do tipo de ruído por elas gerado.

Este trabalho teve como objetivo principal substituir a técnica correntemente utilizada de processamento de dados MT através de regressão linear robusta por um método alternativo de reamostragem, a ser utilizado quando o ruído coerente for persistente. Essa técnica de reamostragem seria usada para obter os valores do tensor de impedâncias que melhor satisfizessem o critério de causalidade em sistemas dispersivos, representado pela transformada de Kramers-Kronig. Observe-se que uma proposição similar para obter as funções de transferência MT foi utilizada por (SUTARNO, 2008), mas que se baseou na média entre a transformada de Kramers-Kronig e o valor obtido por um método de regressão robusto. A tentativa

aqui feita de substituição completa da técnica de processamento robusto não se mostrou totalmente satisfatória. A disseminação e quantidade de amostras contaminadas geradas pelo ruído artificial de grande intensidade (acima de 50 % de valores anômalos) domina completamente o espectro dos elementos do tensor de impedâncias, não permitindo localizar a região onde estariam concentrados os sinais naturais que obedecem a limitação de ondas planas usada no método MT. Por outro lado, uma abordagem mista utilizando como partida os resultados gerados por programas de regressão linear robusta para o cálculo da transformada de Kramers-Kronig obteve sucesso na correção do ruído quando esse se concentra em apenas uma das componentes (real ou imaginária da impedância), individualmente para cada frequência. Essa alternativa pode ser utilizada para substituir o código Rho^+ atualmente empregado em situações unidimensionais ou bidimensionais para a correção de dados MT distorcidos por ruídos artificiais (PARKER; BOOKER, 1996). A formulação aqui proposta tem a vantagem de poder ser utilizada em qualquer situação, inclusive tridimensional.

2 O MÉTODO MAGNETOTELÚRICO

O método magnetotelúrico (MT) é uma técnica passiva de geofísica que registra a variação temporal dos campos elétricos e magnéticos naturais na superfície da Terra. A amplitude e a fase dessas componentes do campo eletromagnético (EM) na superfície dependem da estrutura elétrica no interior da Terra. De forma mais específica, as razões entre os valores medidos simultaneamente dos campos elétrico e magnético são usadas para visualizar variações espaciais da resistividade elétrica na Terra. A resistividade elétrica ρ (ou sua quantidade recíproca, a condutividade elétrica σ) é sensível a variações de diferentes conteúdos minoritários no interior da Terra, incluindo a presença de fluidos em rochas da crosta e do manto superior, além de outras fases condutoras interconectadas tais como grafita, sulfetos e óxidos de ferro. Esses conteúdos minoritários são porém extremamente relevantes para avaliar a evolução geotectônica das diferentes regiões nas quais o método é empregado.

O método foi proposto inicialmente na década de 1950 com uma breve descrição de como a relação entre as componentes ortogonais dos campos elétrico e magnético na superfície, dependentes da frequência, poderia ser usada para medir a estrutura de resistividade da Terra (TIKHONOV, 1950). De forma mais específica, esse autor mostrou que deveria existir uma proporcionalidade e um deslocamento de fase entre as componentes perpendiculares dos campos EM horizontais na superfície. De forma independente, Cagniard (1953) publicou a base teórica do método em mais detalhes. Um pressuposto subjacente crítico ao método MT é que os campos EM incidentes satisfaçam a uma aproximação de onda plana na superfície. A validade dessa suposição foi inicialmente contestada (WAIT, 1954; PRICE, 1962), colocando-se em dúvida a aplicação prática dessa técnica. Essa discussão prosseguiu até que Madden e Nelson (1964) demonstrassem de forma convincente que, para um modelo realista de resistividade da Terra, os campos da fonte externa geradora dos sinais EM usados em MT podem ser adequadamente caracterizados como espacialmente uniformes na superfície. Foi a partir desse resultado que o método MT se estabeleceu como uma técnica válida de exploração geofísica.

Em termos das propriedades físicas envolvidas, o comportamento eletromagnético de um meio é geralmente expresso por três parâmetros: condutividade elétrica σ , permissividade dielétrica ε e permeabilidade magnética μ , os quais descrevem respectivamente as propriedades condutivas, capacitivas e indutivas do meio. Para materiais isotrópicos lineares, essas quantidades são definidas pelas seguintes

relações constitutivas:

$$\mathbf{J} = \sigma \mathbf{E}, \quad (2.1)$$

$$\mathbf{D} = \epsilon \mathbf{E}, \quad (2.2)$$

$$\mathbf{B} = \mu \mathbf{H} \quad (2.3)$$

onde \mathbf{J} , \mathbf{E} , \mathbf{D} , \mathbf{B} e \mathbf{H} representam respectivamente a densidade da corrente elétrica, a intensidade do campo elétrico, o deslocamento elétrico, a indução magnética e a força do campo magnético. No entanto, variações em relação ao valor da permeabilidade observado no vácuo ($\mu_0 = 4\pi \times 10^{-7} \text{ N/A}^2$) são muito pequenas para a quase totalidade das rochas observadas na Terra (SIMPSON; BAHR, 2005). Além disso, nas baixas frequências medidas pelos instrumentos magnetotelúricos, as variações na permissividade dielétrica podem ser negligenciadas (a ser discutido nas Equações 2.14 e 2.15 que seguem). Por outro lado, a resistividade elétrica varia por mais de 4 ordens de grandeza nos materiais presentes no interior da Terra. Essa ampla variabilidade torna a resistividade elétrica uma propriedade útil para determinar a presença dos elementos condutores e as propriedades mecânicas da crosta e do manto.

Descrições teóricas sobre o método MT foram inicialmente limitadas a seções dentro de textos mais amplos sobre o uso de métodos eletromagnéticos em geofísica (por exemplo, em Telford et al. (1990) e Vozoff (1991)). Mais recentemente alguns livros dedicados exclusivamente à descrição tanto da teoria como de aspectos práticos dessa técnica geofísica foram publicados (SIMPSON; BAHR, 2005; CHAVE; JONES, 2012b). Esses últimos textos fornecem uma visão abrangente de toda a teoria e aplicações MT.

2.1 Fontes primárias das variações geomagnéticas

O método MT requer medições das variações temporais de campos elétricos e magnéticos na superfície da Terra, as quais são feitas na presença do campo magnético interno da Terra, considerado estático para efeito da teoria MT. A magnetização remanescente das rochas crustais e a ação do dínamo no núcleo da Terra compreendem essa parte interna. No entanto, variações temporais desses campos são pequenas para efeito das sondagens geofísicas, de tal forma que são os campos externos que fornecem a fonte de energia primária para o método. Os campos

externos são gerados por dois mecanismos que operam dentro de regimes distintos de frequências. Acima de $f = 1$ Hz, os sinais originam-se de atividades meteorológicas, particularmente raios que ocorrem em todo o mundo e que atravessam o globo dentro de um guia de ondas delimitado por extremidades relativamente condutoras, a superfície da Terra e a base da ionosfera. A energia para campos com $f < 1$ Hz tem origem no vento solar, que é um fluxo formado predominantemente por núcleos de hidrogênio e hélio ionizados e elétrons emitidos radialmente para fora da superfície do Sol. O campo magnético da Terra desvia essas partículas carregadas provenientes do Sol ao longo das linhas do campo geomagnético em direção às regiões polares, formando a magnetosfera (região no espaço em que o campo geomagnético permanece confinado). No entanto, mudanças temporais no vento solar, devido principalmente a atividade magnética mais intensa no Sol, podem incrementar e causar flutuações significativas no fluxo dessas partículas, causando impacto na magnetosfera. Esses fenômenos excitam e amplificam campos elétricos variáveis na magnetosfera, os quais induzem correntes elétricas de grande escala na ionosfera. Essas correntes induzidas na ionosfera geram mudanças no campo magnético medido na superfície da Terra e correspondem à fonte natural de sinais EM explorada pelo método MT em períodos mais longos.

2.2 Difusão eletromagnética na Terra

As equações que descrevem a propagação de campos EM na Terra podem ser derivadas das equações de Maxwell. Essas equações formam um conjunto de quatro equações diferenciais parciais fundamentais que descrevem completamente o comportamento dos campos EM:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad \text{Lei de Faraday,} \quad (2.4)$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J} \quad \text{Lei de Ampère,} \quad (2.5)$$

$$\nabla \cdot \mathbf{E} = \frac{\eta_e}{\epsilon} \quad \text{Lei de Gauss,} \quad (2.6)$$

$$\nabla \cdot \mathbf{B} = 0 \quad \text{Lei de Gauss,} \quad (2.7)$$

onde η_e é a densidade de cargas elétricas livres. Usando a identidade vetorial,

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} \quad (2.8)$$

e tomando o rotacional da Lei de Faraday (Equação 2.4) obtém-se,

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\frac{\partial \nabla \times \mathbf{B}}{\partial t}$$

Assumindo que não existem cargas elétricas livres na Terra (isto é, $\eta_e = 0$), tem-se pela lei de Gauss, que

$$\nabla \cdot \mathbf{E} = 0$$

Usando a lei de Ampère juntamente com as relações constitutivas chega-se a,

$$\nabla^2 \mathbf{E} = \mu\sigma \frac{\partial \mathbf{E}}{\partial t} + \mu\epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \quad (2.9)$$

Iniciando com o rotacional da lei de Ampère chega-se a uma equação análoga para \mathbf{B} ,

$$\nabla^2 \mathbf{B} = \mu\sigma \frac{\partial \mathbf{B}}{\partial t} + \mu\epsilon \frac{\partial^2 \mathbf{B}}{\partial t^2} \quad (2.10)$$

Assumindo que as variações temporais dos campos elétricos e magnéticos naturais na superfície têm uma dependência temporal harmônica¹ (isto é, $e^{i\omega t}$), as Equações 2.9 e 2.10 podem ser reescritas como funções da frequência angular ($\omega = 2\pi f$),

$$\nabla^2 \mathbf{E} + \mu\omega(\omega\epsilon - i\sigma) \mathbf{E} = 0 \quad (2.11)$$

$$\nabla^2 \mathbf{B} + \mu\omega(\omega\epsilon - i\sigma) \mathbf{B} = 0 \quad (2.12)$$

que assumem a forma de equações de Helmholtz, com a constante de propagação dada por:

$$k = \sqrt{\mu\omega(\omega\epsilon - i\sigma)} \quad (2.13)$$

Nas sondagens com o método MT, os dados são tipicamente adquiridos dentro

¹Aqui está sendo adotada a representação senoidal $\exp(i\omega t)$, como em [Chave e Jones \(2012a, p. 4\)](#) e [Zorin e Alekseev \(2018\)](#), enquanto que [Berdichevsky e Dmitriev \(2008, p. viii\)](#) e [Berdichevsky e Dmitriev \(2002, p. 27\)](#) adotam $\exp(-i\omega t)$.

da faixa de frequência de $10^{-5} \text{ Hz} < f < 10^3 \text{ Hz}$, ou equivalentemente dentro do intervalo de períodos de $10^{-3} \text{ s} < T < 10^5 \text{ s}$. Considerando que os valores de permissividade relativa de rochas são inferiores a 40, tem-se que o primeiro termo na Equação 2.13 é desprezível ($\omega\epsilon \ll \sigma$), mesmo que se considere o limite mínimo para a condutividade da crosta continental (por exemplo, $\sigma = 0,000 \text{ 1 S m}^{-1}$). Assim, pode-se negligenciar o primeiro termo da equação que representa a corrente de deslocamento e as equações de Helmholtz se simplificam para

$$\nabla^2 \mathbf{E} - i\omega\mu\sigma \mathbf{E} = 0 \quad (2.14)$$

$$\nabla^2 \mathbf{B} - i\omega\mu\sigma \mathbf{B} = 0. \quad (2.15)$$

Dessa forma, dentro da faixa espectral utilizada pelo método MT, os campos elétrico e magnético se propagam via difusão dentro da Terra. Em contraste, no ar isolante, os campos viajam via propagação de ondas. Observe-se que, como a permeabilidade magnética é considerada constante, \mathbf{B} e \mathbf{H} são intercambiáveis.

2.3 Resistividade aparente, fase e profundidade pelicular

A análise simplificada dos dados MT depende da suposição de incidência de campos EM uniformes. Portanto, supondo uma onda plana verticalmente incidente (direção z) é necessário que \mathbf{E} e \mathbf{H} sejam constantes em superfícies normais à incidência. Tomando arbitrariamente o campo \mathbf{E} como sendo polarizado na direção x , requer que o campo \mathbf{H} esteja polarizado na direção y para que se tenha a incidência normal. Na forma de componentes, a Equação 2.14 se simplifica para,

$$\frac{d^2 E_x}{dz^2} - i\omega\mu\sigma E_x = 0$$

que é uma equação diferencial homogênea de segunda ordem, com coeficientes constantes. Soluções gerais para essa equação satisfazem a forma,

$$E_x = E_1 e^{i\omega t} e^{-ikz} + E_2 e^{i\omega t} e^{+ikz}$$

onde E_1 e E_2 são constantes e $k = \sqrt{-i\omega\mu\sigma}$, é o número de onda em um meio condutor, que pode ser reescrita como:

$$k = \sqrt{\frac{\mu\omega\sigma}{2}} (1 - i) \quad (2.16)$$

No caso de um semi-espaço infinito, considerando que em profundidade infinita o campo incidente deve decair até uma amplitude zero e que na superfície a amplitude do campo elétrico é definida como E_0 , obtém-se a solução final

$$E_x = E_0 e^{i\omega t} e^{-(1+i)\sqrt{\omega\mu\sigma/2}z} \quad (2.17)$$

Expandindo a lei de Faraday para a forma de componentes e simplificando os termos de acordo com a suposição de ondas polarizadas no plano chega-se à expressão,

$$H_y = -\frac{1}{i\omega\mu} \frac{\partial E_x}{\partial z} \quad (2.18)$$

Substituindo a Equação 2.17 em 2.18 e isolando para a resistividade elétrica obtém-se

$$\rho_{xy} = \frac{1}{\omega\mu} \left| \frac{E_x}{H_y} \right|^2 = \frac{1}{\omega\mu} |Z_{xy}|^2 \quad (2.19)$$

A Equação 2.19 descreve matematicamente a base fundamental do método MT conforme originalmente proposta por Tikhonov (1950) e Cagniard (1953), relacionando a razão entre os campos EM medidos na superfície (definida como sendo a impedância eletromagnética, Z) com a resistividade elétrica em subsuperfície. Para um semi-espaço homogêneo, essa medida de resistividade é igual à resistividade real da subsuperfície. No entanto, sobre uma Terra não uniforme, que é a situação real, essa estimativa é mais corretamente referida como uma resistividade aparente ρ_a , definida como a resistividade média sobre um hemisfério de raio igual a uma profundidade pelicular (a ser definida na Equação 2.21). Assim, a Equação 2.19 é comumente expressa na forma mais geral,

$$\rho_a = \frac{1}{\omega\mu} |Z|^2 \quad (2.20)$$

A profundidade pelicular (“*skin depth*”) é a medida da distância que os campos EM se difundem em um meio condutor, tal como a Terra. E é definida como a distância em que a amplitude dos campos EM observados na superfície é reduzida por um fator $1/e$. A partir da Equação 2.17, a profundidade pelicular é expressa como,

$$\delta = \sqrt{\frac{2}{\omega\mu\sigma}} = \sqrt{\frac{\rho T}{\pi\mu}} \quad (2.21)$$

de modo que os sinais EM penetram mais profundamente para períodos mais longos em um meio resistivo. Por outro lado, tanto a redução do período do sinal como o aumento da condutividade diminuem a profundidade pelicular. A profundidade pelicular relaciona a resistividade à profundidade em subsuperfície através da frequência do sinal, ou equivalentemente, do período. É essa propriedade que permite ao método MT amostrar diferentes profundidades no interior da Terra, simplesmente utilizando diferentes frequências do sinal medido.

Variações na resistividade em subsuperfície também influenciam a diferença de fase entre os campos elétricos e magnéticos medidos. Observe-se que em um meio não isolante (isto é, $\sigma \neq 0$), a constante de propagação é complexa (Equação 2.13) e existe um deslocamento de fase entre as componentes elétrica e magnética de uma onda EM. Essa mudança de fase é definida por

$$\phi_{xy} = \tan^{-1} \left(\frac{\Im(Z_{xy})}{\Re(Z_{xy})} \right) \quad (2.22)$$

Para um semi-espço de resistividade constante, a fase se torna 45° . A fase é bastante sensível a mudanças na resistividade com profundidade: quando a resistividade aumenta, a fase se torna menor que 45° , e vice-versa.

Observe-se que a amplitude absoluta dos campos EM incidentes não é necessária nas Equações 2.19 e 2.20, visto que somente a razão entre as componentes elétricas e magnéticas é usada. Além disso, pode ser demonstrado que as ondas EM incidentes em todos os ângulos serão refratadas verticalmente no interior Terra, devido ao imenso contraste na resistividade elétrica entre o ar (isolante) e a superfície da Terra (meio condutor).

2.4 Impedância MT sobre uma Terra unidimensional

Um modelo típico de Terra unidimensional (1D) consiste em variações de condutividade apenas com a profundidade. Para esse modelo de múltiplas camadas homogêneas, Wait (1954) derivou uma fórmula recursiva para calcular analiticamente a impedância na superfície em um processo iterativo a partir da camada mais baixa, tomada como um semi-espço homogêneo. Esse modelo simplificado é útil para demonstrar os princípios do método MT, mas muitas vezes não é aplicável à interpretação de dados.

Considere-se um meio 1D composto de N camadas isotrópicas com eixo de co-

ordenadas z vertical positivo para o interior da Terra com origem na superfície, cujo topo de cada camada é z_n , com número de onda $k_n = (1 - i)/\delta_n$ e espessura $h_n = z_{n+1} - z_n$. Então a impedância no topo de cada camada pode ser calculada recursivamente da seguinte forma:

$$Z(z_n) = \frac{\omega\mu}{k_n} \left[\frac{k_n Z(z_{n+1}) + \omega\mu \tanh(ik_n h_n)}{\omega\mu + k_n Z(z_{n+1}) \tanh(ik_n h_n)} \right] \quad (2.23)$$

Um modelo com duas camadas é suficiente para ilustrar os conceitos chaves envolvidos no método MT, com a Equação 2.23 se simplificando para

$$Z(0) = \frac{\omega\mu}{k_1} \left[\frac{k_1 + k_2 \tanh(ik_1 h_1)}{k_2 + k_1 \tanh(ik_1 h_1)} \right] = \frac{\omega\mu}{k_1} \left[\frac{(k_1 - k_2)e^{-ik_1 h_1} + k_1 + k_2}{(k_2 - k_1)e^{-ik_1 h_1} + k_1 + k_2} \right] \quad (2.24)$$

2.4.1 Limite de altas frequências

Nesse caso, $(k_1 h_1) \rightarrow \infty$, e a Equação 2.24 se reduz a $Z(0) = Z_1$ e conseqüentemente $\rho_a = \rho_1$. Dessa forma, em altas frequências MT é sensível apenas à camada superior, pois a profundidade pelicular (Equação 2.21) é pequena.

2.4.2 Limite de baixas frequências

Nesse caso, $(k_1 h_1) \rightarrow 0$, e a Equação 2.24 se reduz a $Z(0) = Z_2$ e $\rho_a = \rho_2$. Em baixas frequências, a profundidade pelicular é grande e a resposta MT é dominada pela resistividade do semi-espaço profundo.

Para explorar variações nas propriedades da camada inferior é necessário evitar esses efeitos de frequência. Uma aproximação de camadas finas pode ser considerada, impondo-se que $\lambda_1 \gg h_1$. Dessa forma, $k_1 h_1 \ll 1$ e assim $\tanh(ik_1 h_1) \rightarrow ik_1 h_1$, dando

$$Z(0) = \omega\mu \left[\frac{1 + ik_2 h_1}{k_2 + ik_1^2 h_1} \right] \quad (2.25)$$

2.4.3 Embasamento condutivo $\sigma_2 \gg \sigma_1$

Nesse caso, $k_2 \gg k_1$, e a Equação 2.24 se reduz a $Z(0) = i\omega\mu h_1$ e $\rho_a = \omega\mu h_1^2$. A resistividade aparente depende apenas da espessura da camada resistiva, ilustrando que o método MT é bastante sensível à profundidade do topo de uma camada condutora.

2.4.4 Embasamento resistivo $\sigma_2 \ll \sigma_1$

Nesse caso, $k_2 \ll k_1$, e a Equação 2.24 se reduz a $Z(0) = (\omega\mu)/(ik_1^2 h_1) = 1/(\sigma_1 h_1)$ e portanto $\rho_a = [\omega\mu(\sigma_1 h_1)^2]^{-1}$. Dessa forma, a resistividade aparente depende do produto da condutividade e da espessura da camada superior, denominada condutância. Esse caso revela uma não-unicidade inerente ao método. MT não consegue resolver de forma independente a condutividade e a espessura de uma camada. Ao invés disso, as medições são sensíveis apenas ao produto dessas quantidades. É necessária informação independente para superar essa incerteza caso um dos parâmetros for desejado individualmente.

2.5 Impedância MT de uma terra bidimensional

Uma Terra 2D contém variações de resistividade em apenas uma direção horizontal (direção y), além da variação em profundidade. Nesse caso, expandindo as leis de Ampère e Faraday e eliminando as derivadas parciais em relação a x (aqui assumida como a direção invariante das principais estruturas geológicas, também chamada de *strike* geológico) possibilita que as equações subsequentes sejam acopladas em dois modos independentes contendo apenas componentes mutuamente ortogonais do campo:

$$\begin{aligned} \frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} &= \mu\sigma E_x \\ \frac{\partial E_x}{\partial z} &= -i\omega B_y \\ -\frac{\partial E_x}{\partial y} &= -i\omega B_z \end{aligned} \quad (2.26)$$

$$\begin{aligned} \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -i\omega B_x \\ \frac{\partial B_x}{\partial z} &= \mu\sigma E_y \\ -\frac{\partial B_x}{\partial y} &= \mu\sigma E_z \end{aligned} \quad (2.27)$$

As Expressões 2.26 definem o modo transversal elétrico (TE) de propagação do sinal EM no interior da Terra, enquanto as Expressões 2.27 definem o modo transversal magnético (TM). O modo TE contém apenas medidas dos campos elétricos paralelos ao *strike* e dos campos magnéticos perpendiculares ao *strike*. Em con-

traste, o modo TM é composto por campos magnéticos paralelos ao *strike* e campos elétricos perpendiculares ao *strike*.

2.6 O tensor de impedâncias

As relações de amplitude e fase entre as razões dos campos elétricos e magnéticos ortogonais horizontais na superfície da Terra são representadas pela impedância eletromagnética que é definida como sendo,

$$\mathbf{E} = \mathbf{ZH}$$

A impedância é conhecida como função de transferência, pois descreve a conexão, dependente da frequência, entre a entrada (campo magnético) e a saída (campo elétrico) de um sistema linear e invariante no tempo. Em geral, a impedância representa uma quantidade tensorial e é expressa como

$$\begin{bmatrix} E_x \\ E_y \end{bmatrix} = \begin{bmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{bmatrix} \begin{bmatrix} H_x \\ H_y \end{bmatrix} \quad (2.28)$$

O tensor de impedâncias contém informações sobre a direção geoeletrica, dimensionalidade e condutividade da Terra e se simplifica nos casos limites de uma ou duas dimensões (CAGNIARD, 1953; GROOM et al., 1993). Para o caso unidimensional, têm-se:

$$\mathbf{Z}_{1D} = \begin{bmatrix} 0 & Z_0 \\ -Z_0 & 0 \end{bmatrix} \quad (2.29)$$

Os elementos diagonais são zero, visto que a lei de Faraday descreve que as variações do campo magnético induzem apenas campos elétricos ortogonais. Dessa forma, as componentes paralelas não são correlacionadas. Os elementos anti-diagonais têm o mesmo módulo com sinal invertido, de modo que o tensor é invariante por rotação, o que é esperado para uma Terra unidimensional.

2.6.1 Terra bidimensional

No caso 2D, a impedância da superfície é dependente do azimute das medições do campo EM ortogonal. No caso especial de que a direção principal das estruturas geoeletricas (*strike*) está alinhada com uma das direções ortogonais dos campos

eletromagnéticos medidos, o tensor de impedância se reduz para,

$$\mathbf{Z}_{2D} = \begin{bmatrix} 0 & Z_{xy} \\ Z_{yx} & 0 \end{bmatrix} \quad (2.30)$$

Em geral, a direção do *strike* geoeletrico é desconhecida antes da aquisição dos dados e, conseqüentemente, o tensor medido conterá entradas diagonais diferentes de zero. Para situações perfeitamente 2D, uma rotação matemática reduz o tensor de impedância medido para a forma da Equação 2.30. No entanto, o ângulo do *strike* não pode ser determinado apenas a partir dos elementos do tensor de impedância. Como uma rotação de 90° simplesmente troca os modos TE e TM, duas direções ortogonais são possíveis. Vetores de indução (descritos abaixo) são frequentemente usados para superar essa ambigüidade.

2.6.2 Terra tridimensional

O caso mais geral é quando a resistividade varia em três dimensões. Na prática, os dados MT medidos nunca são inteiramente 1D ou 2D, e os elementos diagonais do tensor de impedância são sempre diferentes de zero. No entanto, em muitos casos, a resposta da Terra é dominada por estruturas locais 1D ou 2D, permitindo que a análise de dados seja feita dentro dessas dimensões. Por conseguinte, é necessária uma avaliação detalhada da dimensionalidade dos dados MT para determinar se os efeitos 3D presentes são significativos de modo a impedir uma análise mais simples em 1D ou 2D.

2.7 Vetores de indução

O tensor de impedâncias não captura a informação completa disponível nos dados MT medidos, uma vez que usa apenas as componentes do campo horizontal. Na prática, variações laterais na condutividade induzem campos magnéticos verticais não nulos. Analogamente à impedância EM, as razões entre os campos magnéticos verticais e os horizontais são correlacionadas pela função de transferência magnética, comumente conhecida como função “*tipper*” \mathbf{T} (PARKINSON, 1959),

$$\begin{bmatrix} H_z \end{bmatrix} = \begin{bmatrix} T_x & T_y \end{bmatrix} \begin{bmatrix} H_x \\ H_y \end{bmatrix} \quad (2.31)$$

A amplitude do tipper é dada por,

$$|T| = \sqrt{|T_x|^2 + |T_y|^2} \quad (2.32)$$

e é proporcional à intensidade das concentrações anômalas de corrente que dão origem aos campos verticais. O tipper é comumente visualizado através de representações vetoriais de suas funções, referidas como vetores de indução.

De forma similar à mudança de fase que existe entre os campos EM horizontais, o tipper também inclui componentes em fase (real) e fora de fase (imaginária). Esse atraso de fase se origina do tempo finito necessário para induzir campos verticais em limites condutivos em subsuperfície. Para condutores mais fortes, a resposta do tipper é predominantemente em fase e, quando plotados como vetores de indução, as setas apontam para áreas de concentração anômala de corrente (na convenção de Parkinson). Esses vetores de indução são, portanto, sensíveis às variações laterais de condutividade e, além disso, podem fornecer informações sobre a dimensionalidade da Terra. Em um modelo 2D, os vetores devem se alinhar perpendicularmente ao *strike* geoeletrico.

2.8 Aquisição e processamento dos dados

Em um levantamento de campo para estudos MT, as variações temporais das três componentes ortogonais do campo magnético (B_x , B_y e B_z) e das duas componentes horizontais do campo elétrico (E_x e E_y) são medidas simultaneamente em um determinado sítio sob o qual se deseja conhecer a distribuição de resistividade elétrica. Como esses sinais são geralmente de baixa amplitude é necessário que se disponha de sensores com excelente razão de sinal sobre ruído e alta capacidade de amplificação.

Em geral, o campo elétrico é medido de forma indireta a partir do registro da variação temporal na diferença de potencial entre pares de sensores elétricos separados por pelo menos uma centena de metros. Essa separação é necessária para que a tensão medida seja muito maior que o ruído gerado pelos próprios sensores, os quais são geralmente eletrodos porosos não polarizáveis. Esses eletrodos consistem de potes com base cerâmica, preenchidos com soluções, normalmente cloreto de chumbo, e mantidos em contato direto com o solo.

Para medida das variações temporais do campo magnético utilizam-se magnetômetros com alta sensibilidade e baixo ruído intrínseco nos intervalos de frequência

de interesse. Em baixas frequências (inferiores a cerca de 0,1 Hz; correspondente a períodos superiores a 10 s) utilizam-se magnetômetros do tipo *fluxgate*, adequados à relação sinal/ruído nesse intervalo. Em frequências mais altas (períodos inferiores a 10 s), porém, as variações naturais do campo geomagnético têm amplitudes muito baixas, inferiores ao nível do ruído intrínseco a esses magnetômetros. Nessas frequências, magnetômetros do tipo indução são empregados para medir as variações geomagnéticas.

Além do ruído EM dos próprios instrumentos, fontes ativas nas proximidades dos locais de medida adicionam energia eletromagnética nas medições e podem submergir completamente o sinal natural em algumas situações. Essas variações de caráter antrópico também são consideradas ruído EM, pois não obedecem as premissas de onda plana utilizadas para obtenção das funções de transferência entre campos elétricos e magnéticos do método MT.

O objetivo do processamento de dados MT é extrair desses sinais medidos no campo, geralmente ruidosos, um conjunto de funções contínuas e suaves que representem a resposta da Terra à propagação da onda EM. Essas características de suavidade e continuidade são consistentes com as diversas funções de transferência MT geradas por diferentes modelos da Terra, incluindo aqueles em que ocorre uma significativa variação lateral e em profundidade da resistividade elétrica. Isso se deve ao fato de que a propagação da onda EM, e conseqüentemente o processo de indução eletromagnética no interior da Terra, é difusivo e portanto sensível apenas à média espacial da condutividade (LARSEN et al., 1996). Por isso, se grandes variações nas funções de transferência são observadas entre frequências contíguas, pode-se inferir que essas funções foram erroneamente estimadas.

O processamento de dados envolve a estimativa das funções de transferência MT, no domínio da frequência, a partir das séries temporais dos campos EM medidos. Geralmente, esse processo envolve a estimativa de um pequeno número de funções de transferência quando comparado com o número bastante grande de dados medidos. Vários métodos matemáticos foram desenvolvidos e aplicados na estimativa dos elementos do tensor de impedâncias MT. Incluem métodos de mínimos quadrados, métodos de admitância e referência remota, métodos paramétricos com e sem restrições de dispersão de impedância, métodos no domínio do tempo, métodos de wavelets e métodos robustos. Jones et al. (1989) e Chave (2012) revisaram alguns desses métodos e compararam seu desempenho nas estimativas de impedância tanto em períodos geomagneticamente calmos como em períodos

geomagneticamente perturbados.

O procedimento utilizado no processamento deste trabalho é descrito detalhadamente em Pádua (2014), de forma que apenas uma descrição resumida será aqui apresentada. A primeira etapa na estimativa das funções de transferência é a transformação dos dados medidos no domínio de tempo para o domínio de frequência. Para isso, é necessário pré-condicionar os dados para reduzir o efeito das tendências, remover ruídos mais salientes (picos) e reduzir os efeitos devidos a discretização e limitação temporal dos dados. O pré-condicionamento divide as séries temporais em segmentos, cujo tamanho depende da faixa de período que está sendo calculada. Quanto mais segmentos houver, melhores serão os resultados estatísticos a serem obtidos na sequência do processamento, mas menor será o período máximo estimado. O pré-condicionamento em cada um dos segmentos é normalmente feito com uma função de janelas (por exemplo, utilizando uma janela Hanning, Hamming ou Parzen) (EGBERT; BOOKER, 1986).

A próxima etapa é a conversão dos dados segmentados do domínio do tempo para o domínio da frequência, normalmente usando uma Transformada de Fourier discreta, ou uma transformada *wavelet*. Cada componente do campo EM medido deve ser calibrado em relação à sensibilidade do instrumento usado na medida. Algumas calibrações de instrumento (por exemplo, sensor de bobina de indução) dependem da frequência e, portanto, as calibrações precisam ser aplicadas como tal, enquanto outras (por exemplo, o magnetômetro *fluxgate*) são independentes da frequência. Idealmente, são escolhidas de 6 a 10 frequências alvo por década, equiespaçadas em escala logarítmica. Não são necessárias mais frequências pois as relações de dispersão (WEIDELT, 1972) predizem resultados semelhantes para frequências vizinhas. Porém, um número menor de frequências pode resultar em *aliasing* no domínio da frequência (SIMPSON; BAHR, 2005).

O procedimento usado para cálculo das médias utiliza os auto-espectros e espectros cruzados calculados a partir dos espectros de potência, A_k , para cada uma das componentes do campo EM e em cada um dos segmentos em que os dados foram divididos. Vozoff (1991) indicou que a densidade do auto-espectro pode ser calculada para um segmento das séries temporais usando:

$$\langle A(f_j) \rangle = \sqrt{\frac{1}{2m+1} \sum_{k=j-m}^{j+m} A_k A_k^*} = \sqrt{A_j A_j^*} \quad (2.33)$$

onde f_j é a frequência alvo na qual o espectro será calculado, $2m + 1$ é o número de estimativas espectrais a serem mediadas e o asterisco indica o complexo conjugado. Tomando-se o quadrado do lado direito da Equação 2.33 obtém-se a densidade espectral da auto-potência (*autopower*) na frequência f_j . A densidade da potência cruzada (*crosspower*) de duas componentes é:

$$\langle A(f_j), B(f_j) \rangle = \frac{1}{2m + 1} \sum_{k=j-m}^{j+m} A_k B_k^* = \langle A_j B_j^* \rangle \quad (2.34)$$

As estimativas espectrais finais envolvem a média das auto-potências e das potências cruzadas sobre os múltiplos segmentos em que os dados foram subdivididos. Esse passo é seguido pelo cálculo dos elementos do tensor de impedâncias usando essas auto-potências e potências cruzadas armazenadas em uma matriz espectral. Conforme discutido por Sims et al. (1971), Vozoff (1991) e Simpson e Bahr (2005), esses elementos podem ser estimados por equações do tipo:

$$Z_{xx} = \frac{\langle E_x H_x^* \rangle \langle H_y H_y^* \rangle - \langle H_y H_x^* \rangle \langle E_x H_y^* \rangle}{\langle H_x H_x^* \rangle \langle H_y H_y^* \rangle - \langle H_y H_x^* \rangle \langle H_x H_y^* \rangle} \quad (2.35)$$

e outras expressões similares tanto para Z_{xx} como para os demais elementos do tensor de impedâncias (Z_{xy} , Z_{yx} e Z_{yy}).

2.9 Efeito de ruídos

Na obtenção das funções de transferência MT, um aspecto crucial é minimizar o efeito dos ruídos EM nos dados medidos, pois somente dados com uma boa relação sinal-ruído podem ser usados para obter modelos confiáveis da distribuição de condutividade do subsolo. Inicialmente, as estimativas numéricas das funções de transferência basearam-se em métodos de mínimos quadrados para minimização de ruído, os quais consideram que os erros estatísticos seguem uma distribuição gaussiana (SIMPSON; BAHR, 2005). No entanto, logo se reconheceu que uma distribuição de ruído gaussiana não é realista e que mesmo um número pequeno de eventos ruidosos (*outliers*) poderia ter efeito significativo nas estimativas das funções de transferência. Isso ocorre porque esses ruídos seguem uma distribuição não gaussiana de forma que estimativas por mínimos quadrados são consideravelmente afetadas (EGBERT; BOOKER, 1986). Os procedimentos iniciais usaram uma seleção manual nas séries temporais para eliminar esses eventos ruidosos, uma abordagem demorada, não-rigorosa e propensa a erros.

O maior efeito da presença dos ruídos ocorre nas componentes de auto-espectro utilizadas nas expressões usadas para obter as funções de transferência. Para remover esse efeito foi proposto o método de referência remota (GAMBLE et al., 1979) que envolve a substituição dos termos conjugados na Equação (2.35) por dados adicionais coletados simultaneamente em um local remoto. A ideia básica é que na localização remota o sinal natural usado pelo método MT será o mesmo, mas o ruído não será correlacionado de forma que seu efeito será removido (SIMPSON; BAHR, 2005). As referências remotas magnéticas horizontais são preferidas pois os campos magnéticos são geralmente menos contaminados pelo ruído que o campo elétrico e são mais homogêneos nas áreas em torno de inomogeneidades laterais. Usando esse método, o elemento de impedância apresentado na Equação (2.35) pode ser estimado como:

$$Z_{xx} = \frac{\langle E_x R_x^* \rangle \langle H_y R_y^* \rangle - \langle H_y R_x^* \rangle \langle E_x R_y^* \rangle}{\langle H_x R_x^* \rangle \langle H_y R_y^* \rangle - \langle H_y R_x^* \rangle \langle H_x R_y^* \rangle} \quad (2.36)$$

onde R_x^* e R_y^* são os complexos conjugados dos campos magnéticos horizontais medidos na estação remota.

Embora o desenvolvimento do método de processamento de referência remota tenha sido um marco na produção de sondagens MT de alta qualidade, esse método funciona bem apenas se o ruído for estacionário e não correlacionado. O método falha quando essa situação não é observada. Em seguida, diversos investigadores (EGBERT; BOOKER, 1986; CHAVE; THOMSON, 1989) desenvolveram esquemas de processamento estatisticamente robustos que iterativamente colocam pesos à contribuição de diferentes segmentos de dados para as estimativas das funções de transferência. Esses métodos mostraram-se adequados para produzir estimativas estáveis das funções de transferência com erros razoáveis na maioria das situações reais atualmente encontradas e constituem-se na principal ferramenta para o processamento de dados MT.

A situação mais crítica enfrentada pelo método MT é aquela da ocorrência de ruídos coerentes que se espalham por grandes distâncias. Nesse caso, nem a técnica de referência remota nem a metodologia de processamento robusto conseguem garantir que resultados razoáveis sejam obtidos. O exemplo mais comum de ruído que leva a essa limitação é aquele gerado por fontes de ruídos de corrente contínua (DC), tais como os gerados por linhas de trem operando com corrente contínua ou linhas de transmissão de energia de alta tensão em corrente contínua (HVDC - High-Voltage

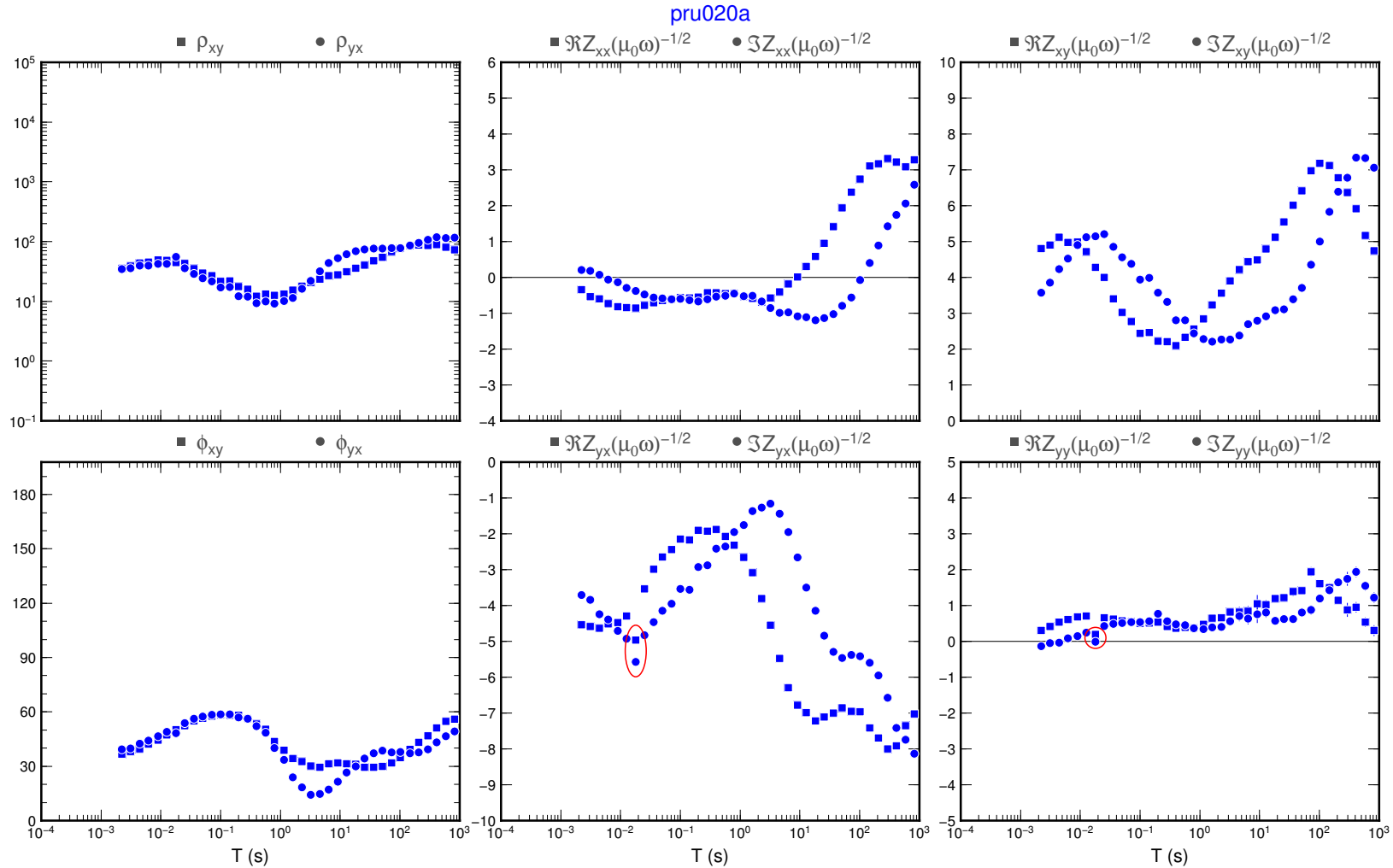
Direct Current). Egbert (1997) mostrou que, quando os dados MT são severamente contaminados por esse tipo de ruído coerente, apenas em algumas situações especiais é possível extrair resultados úteis dos dados: quando se encontra um local distante para alguma estação de referência não afetada pelo ruído coerente ou quando a fonte do sinal ruidoso coerente é intermitente. Quando essas situações não estão presentes, outros critérios/metodologias devem ser considerados para analisar os dados.

3 O PROBLEMA PROPOSTO: RUÍDOS EM EXTREMOS NAS FUNÇÕES DE TRANSFERÊNCIA MT

A apresentação das funções de transferência MT de um determinado local de medidas é geralmente feita por meio de gráficos cartesianos dessas funções, no eixo das ordenadas, em relação ao logaritmo do período, em segundos, no eixo das abscissas. Em geral, são mostrados apenas os gráficos das resistividades aparentes e fases para os dois elementos da diagonal secundária do tensor de impedâncias (ρ_{xy} , ρ_{yx} , ϕ_{xy} e ϕ_{yx}), pois estas são as funções geralmente utilizadas para interpretação das medidas. Para gráficos de resistividade aparente, o eixo das ordenadas é o logaritmo da resistividade aparente (em $\Omega \text{ m}$), enquanto para a fase o eixo das ordenadas é linear (em unidades de graus). Raramente são apresentados os valores da impedância por sua grande variação em função do período. Neste trabalho, para facilitar a visualização dessa função de transferência e a consequente discussão dos efeitos dos ruídos, vamos apresentar a variação de uma função da impedância (dividida pela raiz quadrada da frequência e da permeabilidade magnética).

A Figura 3.1 mostra um exemplo de um gráfico típico de respostas MT. Essa sondagem foi adquirida no segundo semestre de 2015, nas proximidades da cidade de Ourinhos, no estado de São Paulo, e fez parte de um perfil de medidas usado na Dissertação de Mestrado de Horvath (2017). Os dados foram processados usando o código computacional de regressão robusta desenvolvido por Egbert e Booker (1986) e são de boa qualidade a partir dos critérios de suavidade e continuidade na sequência das frequências alvo das várias funções de transferência. As componentes Z_{yx} e Z_{yy} apresentam apenas um ponto anômalo (*outlier*) cada, identificados pelas elipses vermelhas nas componentes. É associado a linhas de transmissão elétrica de corrente alternada (AC) operando em 60 Hz. O ruído é significativo para afetar o sinal na sua frequência original, mas não é elevado o suficiente para produzir harmônicos e sub-harmônicos em outras frequências. Seu efeito pode ser minimizado na interpretação excluindo essa frequência das análises.

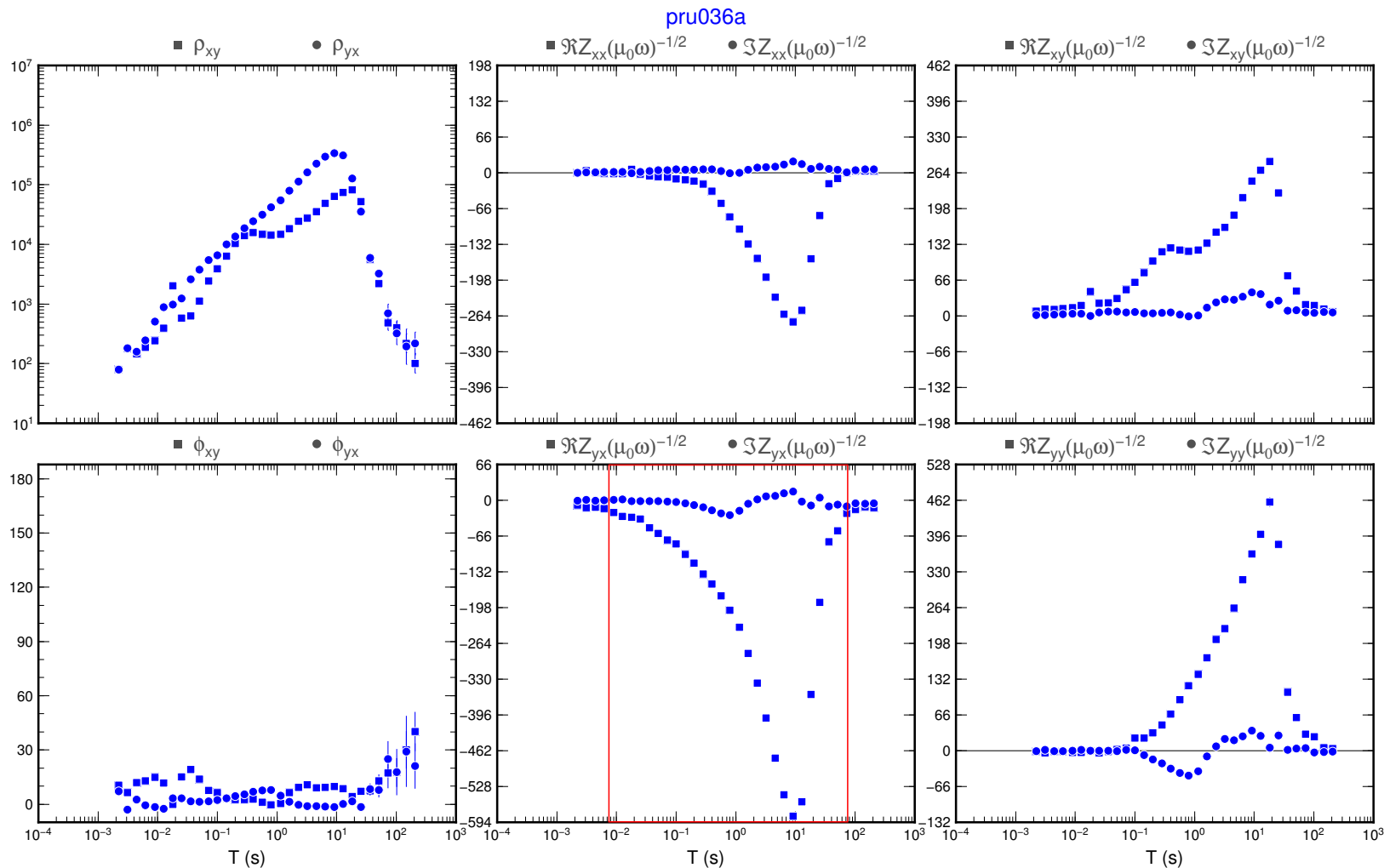
Figura 3.1 – Funções de transferência de estação MT coletada no estado de São Paulo (proximidades da cidade de Ourinhos) obtidas usando o processamento robusto de Egbert e Booker (1986). São mostradas as variações de resistividade aparente (ρ_{xy} e ρ_{yx}), em Ωm ; fase (ϕ_{xy} e ϕ_{yx}), em graus; e as componentes reais e imaginárias de uma função derivada dos elementos do tensor de impedância ($Z_{xx}/\sqrt{\mu_0\omega}$, $Z_{xy}/\sqrt{\mu_0\omega}$, $Z_{yx}/\sqrt{\mu_0\omega}$ e $Z_{yy}/\sqrt{\mu_0\omega}$). As elipses vermelhas são discutidas no texto.



Por outro lado, a Figura 3.2 mostra resultados de outra estação do mesmo perfil MT, também coletada no segundo semestre de 2015 e processada pelo mesmo algoritmo robusto da figura anterior. Essa estação situa-se bem mais a leste que a anterior, nas proximidades da cidade de Sorocaba, estado de São Paulo. Nesse caso, o efeito do ruído não aparece na forma de algum ponto anômalo localizado em frequência específica, mas em uma ampla faixa de frequências (identificada pelo quadrado vermelho na componente Z_{yx}). Esse ruído afeta completamente a parte real de todos os elementos do tensor de impedâncias fazendo com que sua amplitude seja centenas de vezes maior do que a observada na parte imaginária (comparar as amplitudes da impedância com aquelas da Figura 3.1). Como consequência desse ruído, as resistividades aparentes e fases derivadas da impedância MT são completamente distorcidas (resistividade aparente aumenta com aumento do período com ângulo de 45° na escala bilogarithmica e a fase se concentra próxima a 0°), mascarando completamente as características da variação geolétrica em subsuperfície. Esse é um resultado típico de sondagens MT que sofrem interferência industrial severa (QIAN; PEDERSEN, 1991).

Esse ruído intenso é majoritariamente associado a uma linha de transmissão de energia HVDC que liga a usina de Itaipú (Estado do Paraná) a Ibiúna (Estado de São Paulo). Esse sistema de transmissão é formado por duas linhas de ± 600 kV, com extensão de cerca de 820 km, e transporta energia para a região em torno da cidade de São Paulo (após conversão de DC para AC em Ibiúna). O uso desse tipo de transmissão de energia elétrica se justifica por ser mais econômico que a transmissão em modo AC em longas distâncias (perdas de transmissão no sistema HVDC são significativamente mais baixas). Conforme anteriormente discutido, em termos de levantamentos que utilizam sinais eletromagnéticos naturais, a presença dessas linhas HVDC representa uma imensa fonte de ruído eletromagnético coerente e contínuo que se espalha a grandes distâncias (pela alta voltagem utilizada) e contamina a totalidade do espectro de frequências (pelo nível DC). A estação medida nas proximidades de Ourinhos (*pru020a*) dista cerca de 200 km da linha HVDC e foi a última dentre várias estações do perfil, espaçadas em cerca de 30 km, em que não se observou o efeito desse tipo de ruído. Por outro lado, a estação de Sorocaba (*pru036a*) situa-se a pouco mais de 40 km da linha HVDC e é severamente contaminada pelo ruído DC.

Figura 3.2 – Funções de transferência de estação MT coletada no estado de São Paulo (proximidades da cidade de Sorocaba). São mostradas as variações de resistividade aparente (ρ_{xy} e ρ_{yx} , em Ω m, fase (ϕ_{xy} e ϕ_{yx}), em graus, componentes reais e imaginárias de uma função derivada dos elementos do tensor de impedância (Z_{xx} , Z_{xy} , Z_{yx} e Z_{yy}). O quadrado vermelho é discutido no texto.



Para descrever qualitativamente as características do ruído EM observado utiliza-se a definição básica da impedância elétrica. É um número complexo (Z) que possui uma parte real (resistência - R) e uma parte imaginária (reatância - X), todos dados em *ohms*. Assim, pode ser expressa como:

$$Z = R + iX \quad (3.1)$$

onde:

$$R = \frac{\Delta V}{I} \quad (3.2a)$$

$$X_L = \omega L \quad (3.2b)$$

sendo ΔV a diferença de potencial em *Volts*, I a corrente elétrica em *Amperes*, X_L é a parte indutiva da reatância (não se está considerando aqui a reatância capacitiva), L é a indutância em *Henrys* e ω a frequência angular. Como pode ser observado, uma diferença fundamental entre as duas partes da impedância é que apenas a parte imaginária depende da frequência.

Vamos analisar três situações:

- Sistema alimentado apenas por Corrente Contínua: Nesse caso, a impedância vai ser um valor real, relacionado com a amplitude da corrente contínua injetada, e independente da frequência.
- Sistema alimentado apenas por Corrente Alternada: A parte real da impedância vai estar relacionada com a amplitude da corrente alternada e a parte imaginária com a variação em frequência dessa corrente.
- Sistema alimentado simultaneamente por Corrente Contínua e Alternada: A parte real vai ser formada pela soma das amplitudes de ambas as correntes, enquanto a parte imaginária vai depender da variação da corrente alternada com a frequência.

A estação MT mostrada na Figura 3.1 é um exemplo típico de um sistema alimentado apenas por corrente alternada, na qual essa corrente está relacionada com o

sinal geomagnético natural cuja amplitude varia em função da frequência. Esse sinal natural é contaminado por outra fonte de sinal artificial em corrente alternada, mas que afeta um espectro limitado de frequências. Essa é a situação ideal para uma sondagem MT. Por outro lado, a estação MT da Figura 3.2 enquadra-se na situação do sistema alimentado pelas duas correntes (AC e DC), na qual a parte real é contaminada pela corrente contínua, enquanto a parte imaginária seria aparentemente dominada principalmente pela resposta do sinal natural (corrente alternada).

As características da parte real na Figura 3.2 podem ser usadas para ilustrar o efeito do ruído DC na sondagem MT, em função da frequência do sinal. Considerando a fonte de ruído fixa e posicionada a uma certa distância da estação MT, seu efeito sobre os elementos do tensor de impedância podem ser divididos em partes de campo próximo (*near field*) e distantes (*far field*), em função da relação entre a frequência do sinal considerado e o *skin depth* para essa frequência. Em frequências muito altas, a fonte de ruído DC está suficientemente distante (mais que 3 *skin depths*) para que os sinais possam ser considerados do tipo *far field* e, portanto, ser tratados como ondas planas pelo método MT (SANDBERG; HOHMANN, 1982). Nesse caso, o resultado da parte real vai se aproximar daquele da parte imaginária. Para frequências nas quais a distância da fonte é inferior a 3 *skin depths*, o campo magnético relacionado à fonte DC (ruído EM) vai ser muito maior que o campo magnético associado ao sinal natural. Esse é o caso de *near field* em que não se obtém respostas sobre a impedância do meio pelo qual as ondas EM se propagam, mas apenas sobre a geometria da fonte do sinal artificial. Nessa situação, a hipótese de ondas planas sobre a qual se assenta o método MT não se aplica. Eventualmente, para frequências ainda mais baixas, o sinal natural aumenta significativamente de intensidade e pode sobrepujar o sinal artificial. Nessa situação, a parte real da impedância volta a se aproximar do resultado esperado para ondas planas.

Presentemente, além da linha Itaipú-Ibiúna (que em sua maior extensão corre sobre o pacote vulcânico-sedimentar da Bacia do Paraná), há apenas outra linha HVDC em operação no país. Situa-se entre Rio Madeira (Porto Velho, Estado de Rondônia) e Araraquara (Estado de São Paulo). Essa linha com quase 2 400 km de extensão também é operada com ± 600 kV e atravessa em sua quase totalidade as bacias do Pantanal e do Paraná. Da mesma forma que a linha Itaipú-Ibiúna, deve também ser uma fonte significativa de ruído eletromagnético coerente na região centro-oeste do país. Em ambos os casos, os efeitos dessas linhas é restrito pela condutividade dos sedimentos das bacias sedimentares que atravessam, mas mesmo assim podem

chegar a distâncias da ordem de 200 km, conforme anteriormente discutido neste trabalho.

A Figura 3.3 mostra um mapa do Brasil com a localização das atuais e futuras linhas HVDC no país. Afora as duas linhas HVDC já citadas, estão em construção duas outras ligando Belo Monte (Estado do Pará) a Estreito (Minas Gerais) e ao Rio de Janeiro. Essas linhas vão operar em ± 800 kV e a previsão inicial é que estariam operacionais até 2020. Contrariamente às duas outras linhas HVDC, essas linhas vão se situar quase integralmente sobre terrenos de alta resistividade formados por rochas pré-cambrianas do Craton do São Francisco e seus cinturões marginais. Conseqüentemente, espera-se que o ruído por elas gerado deva ser ainda mais intenso e se espalhar por várias centenas de quilômetros cobrindo grande parte do território nacional. A viabilização de medidas de campo eletromagnético natural nessas regiões vai exigir, portanto, o desenvolvimento de novas técnicas de processamento que reduzam a influência desse tipo de ruído a fim de melhorar a precisão das estimativas das funções de resposta geofísicas.

A característica de aparente robustez da parte imaginária da impedância em relação aos intensos ruídos gerados por essas linhas de transmissão HVDC pode ser explorada como uma alternativa para tentar reconstruir a parte real da impedância e assim recuperar a resposta MT que seria gerada por um sinal natural não afetado pelo ruído. Saliente-se que essa alternativa só é válida para o caso de a parte imaginária não ser afetada (ou ser minimamente afetada) por ruídos do tipo AC. Não se aplica, portanto, às extremidades ou proximidades de sub-estações das linhas HVDC onde ocorra conversão de DC para AC e que linhas de transmissão de corrente alternada de alta tensão (HVAC) também estejam presentes. Na prática, porém, é virtualmente impossível encontrar locais em que apenas o ruído DC esteja presente. Por isso, iremos discutir neste trabalho possíveis aplicações da transformada de Kramers-Kronig para tentar minimizar efeitos de ruídos presentes tanto na parte real como na parte imaginária das componentes do tensor de impedâncias magnetotelúrico.

4 A TRANSFORMADA DE KRAMERS-KRONIG

As relações de dispersão de Kramers-Kronig fazem parte da teoria de dispersão da óptica e sua aplicação ao método MT foi inicialmente sugerida por Weidelt (1972). Ele mostrou que a impedância Z é uma função analítica no semi-plano complexo e, como resultado, as relações de dispersão podem ser aplicadas às partes reais e imaginárias das respostas MT para aproveitar informações redundantes que elas contêm. No método MT, relações de dispersão de dois tipos são consideradas: entre as partes reais e imaginárias da impedância Z e entre a resistividade aparente e a fase da impedância.

Zorin e Alekseev (2018) discutem que as relações do segundo tipo têm sido bastante usadas em sondagens MT, especialmente para controlar a qualidade dos dados e para aumentar a precisão na determinação do tensor de impedância. O exemplo mais claro é o algoritmo *Rho+*, comumente utilizado para validar a consistência entre as resistividades aparentes e as fases (PARKER; BOOKER, 1996). No entanto, essas relações entre resistividade aparente e fase são afetadas por distribuições multidimensionais de condutividade, de forma que a relação é estritamente válida apenas no modo 1D e no modo TM no caso de situações 2D. Seu uso é questionável para o modo TE e impraticável para muitas situações 3D. Por outro lado, a relação do primeiro tipo, entre as partes reais e imaginárias do tensor de impedâncias, é válida para um espectro muito mais amplo de situações. Não são, porém, utilizadas pois a maioria dos dados MT é tradicionalmente processada e interpretada na representação “*amplitude e fase versus frequência*” (ZORIN; ALEKSEEV, 2018).

Utilizando as derivações apresentadas em Berdichevsky (1999) e Berdichevsky e Dmitriev (2008, p. 38) a transformada de Kramers-Kronig pode ser definida como:

$$R(\omega_0) = \frac{2}{\pi} \text{pv} \int_0^{\infty} \frac{X(\omega)}{\omega^2 - \omega_0^2} \omega d\omega \quad (4.1a)$$

$$X(\omega_0) = -\frac{2\omega_0}{\pi} \text{pv} \int_0^{\infty} \frac{R(\omega)}{\omega^2 - \omega_0^2} d\omega \quad (4.1b)$$

onde *pv* significa que o valor principal de Cauchy da integral é utilizado no cálculo

e as funções $R(\omega)$ e $X(\omega)$ são descritas para o método MT na forma:

$$R = \Re\left(\frac{Z}{i\mu_0\omega}\right) \quad (4.2a)$$

$$X = \Im\left(\frac{Z}{i\mu_0\omega}\right) \quad (4.2b)$$

A partir das Equações (4.1) e (4.2) podem-se obter as partes reais e imaginárias das componentes do tensor MT (BERDICHEVSKY; DMITRIEV, 2002, p. 71) que vão ser dadas por:

$$\Re(Z(\omega_0)) = \frac{2\omega_0^2}{\pi} \text{Pv} \int_0^\infty \frac{\Im(Z(\omega))}{\omega(\omega^2 - \omega_0^2)} d\omega \quad (4.3a)$$

$$\Im(Z(\omega_0)) = -\frac{2\omega_0}{\pi} \text{Pv} \int_0^\infty \frac{\Re(Z(\omega))}{(\omega^2 - \omega_0^2)} d\omega \quad (4.3b)$$

A partir daqui vamos nos referir à Equação 4.3a como transformação *kk-ir* e a Equação 4.3b como transformação *kk-ri*, onde o prefixo *kk* é formado pelas iniciais de Kramers-Kronig, os sufixos *ir* e *ri* representam as iniciais da componente complexa das funções de entrada e saída. Essa definição é feita considerando que na Equação 4.3a a função de entrada (o integrando da transformada) é a componente imaginária da impedância, enquanto a saída (o resultado da transformada) é a componente real da impedância. Dessa forma, essa transformada é denominada transformada *kk-ir*. A situação recíproca é válida para a outra transformada.

4.1 Implementação da transformada de Kramers-Kronig

O foco deste trabalho é o uso da transformada de Kramers-Kronig para obter o tensor de impedâncias (Equação 2.28). Atualmente, os programas que estimam o tensor MT, tais como Egbert e Booker (1986) ou Pádua (2014), obtêm as componentes do tensor realizando algum tipo de regressão robusta sobre a Equação linear 2.28 para cada período selecionado. As curvas de resistividade e fase são em seguida obtidas utilizando o resultado da regressão para cada um desses períodos. Os resultados obtidos por esses métodos, apesar de robustos, estão limitados pelo ponto de ruptura (*breakdown point*) que é no máximo 50%. Isso quer dizer que

se mais de 50 % dos pontos estiverem contaminados, a regressão tentará ajustar a parte contaminada pelo ruído EM, e não o sinal natural da onda plana EM originária de descargas elétricas atmosféricas distantes ou dos fenômenos solares.

Para obter as curvas MT para os períodos em que os ruídos são persistentes é necessário abandonar os métodos de regressão linear tradicionais. Para isso é necessário utilizar outra relação para o tensor MT que não venha da Equação 2.28. A relação que será explorada neste trabalho é a transformada de Kramers-Kronig fornecida pelas Equações 4.1. Essa transformação estabelece uma relação causal entre a parte real e imaginária do tensor MT.

Dessa forma, passamos a ter um problema completamente distinto do original. No problema original, o sinal de entrada era o campo \mathbf{B} e o sinal de saída o campo \mathbf{E} . Na implementação aqui proposta, a variável de entrada é a parte imaginária de \mathbf{Z} e a de saída é a parte real de \mathbf{Z} , considerando nesse caso a transformada *kk-ir* fornecida pela Equação 4.3a. Ou, caso seja escolhida a transformada *kk-ri* fornecida pela Equação 4.3b, a entrada será obtida com valores de $\Re\mathbf{Z}$ e a saída com valores de $\Im\mathbf{Z}$. Uma segunda distinção é que os métodos robustos buscam parâmetros fixos de uma função linear pré-estabelecida, o hiperplano complexo dado pela Equação 2.28. Por outro lado, a transformada de Kramers-Kronig não fornece funções pré-estabelecidas que possam ser ajustadas, tampouco espera que as funções sejam lineares. A situação a ser equacionada não é mais de ajuste de uma função (uma relação que leva de um conjunto a outro) mas de ajuste de um funcional (uma relação que leva uma função a outra). A terceira distinção é que na regressão linear, \mathbf{Z} é um valor fixo para cada período, relacionado com o parâmetro que melhor ajusta o hiperplano complexo dada pela Função 2.28. Usando a transformada de Kramers-Kronig, para gerar as funções de entrada e saída serão gerados um conjunto de valores de \mathbf{Z} para cada período selecionado e as funções de entrada e saída são obtidas pelo produto cartesiano desses conjuntos. A última distinção entre os dois métodos é que o método tradicional procura eliminar os valores anômalos (*outliers*) do campo EM. Na transformada de Kramers-Kronig são procurados exatamente os valores anômalos de \mathbf{Z} para os períodos contaminados, pois já se sabe antecipadamente que os valores mais frequentes são suspeitos (contaminados por ruídos EM).

Pode-se utilizar a transformada de Kramers-Kronig de três maneiras distintas, de acordo com a fonte utilizada para obter as componentes do tensor de impedâncias:

entrada unicamente por meio de regressão robusta: Nesse caso são utilizadas como entrada as componentes das impedâncias obtidas pelos métodos de regressão robusta usualmente usados em processamento de dados MT, como aqueles mostrados nas Figuras 3.1 e 3.2. Essa alternativa também pode ser utilizada para verificar para quais períodos o método de regressão robusta apresenta resultados consistentes com os da transformada de Kramers-Kronig.

entrada mista de regressão robusta e reamostragem: Nesse caso, os valores de entrada provêm de duas fontes distintas. Para os períodos em que os resultados da regressão robusta foram considerados consistentes com a transformada, serão usadas as componentes das impedâncias obtidas por meio de regressão robusta. Por outro lado, para os períodos onde as impedâncias não foram consideradas consistentes com as obtidas pela transformada, será utilizado um método de reamostragem para tentar localizar entre os *outliers* dos métodos de regressão robusta aqueles valores que seriam consistentes com a transformada de Kramers-Kronig.

entrada puramente através de reamostragem: Nesse caso tenta-se encontrar estimativas das componentes do tensor de impedâncias consistentes com a transformada de Kramers-Kronig usando apenas o método de reamostragem, sem utilizar a regressão robusta.

4.2 Cálculo numérico da transformada de Kramers-Kronig

Para calcular as transformadas *kk-ir* e *kk-ri* é necessário obter a função $Z(\omega)$ para $\omega \in \mathbb{R}^+$, ou seja, ($\omega > 0$). No entanto, como a instrumentação de coleta de dados MT obtém as medidas de modo discreto no tempo e em um intervalo de tempo limitado, as impedâncias são calculadas apenas para um conjunto discreto e limitado de frequências angulares ($\omega_{\min} \leq \omega_i \leq \omega_{\max}$), com $i \in \mathbb{N}$. As funções de estimativas $\hat{Z}(\omega)$ são obtidas a partir de $Z(\omega_i)$ da seguinte maneira.

- Para ($\omega_{\min} \leq \omega_i \leq \omega_{\max}$), $\hat{Z}(\omega)$ será obtida usando uma função *spline* com os valores de $Z(\omega_i)$ como pontos de suporte como em (MARCUELLO et al., 2005).
- Para $\omega < \omega_{\min}$ e $\omega > \omega_{\max}$ vamos assumir que $\hat{Z}(\omega)$ se comporta como um

semiespaço infinito (Equação 2.19), de modo que,

$$\hat{Z}(\omega) = Z(\omega_0) \sqrt{\frac{\omega}{\omega_0}}, \text{ com } \omega_0 = \begin{cases} \omega_{\min} & \text{para } \omega < \omega_{\min} \\ \omega_{\max} & \text{para } \omega > \omega_{\max} \end{cases} \quad (4.4)$$

Maiores detalhes sobre a implementação computacional da transformada de Kramers-Kronig e das bibliotecas externas e parâmetros utilizados pelas rotinas de cálculo são fornecidos no apêndice A.

4.3 Validação do algoritmo desenvolvido - testes em dados sintéticos

Para validar o algoritmo desenvolvido para cálculo numérico da transformada de Kramers-Kronig são discutidos exemplos de aplicação em casos sintéticos. Funções de transferência MT derivadas para dois modelos de distribuição de resistividades, um unidimensional e outro tridimensional, são usadas para verificar o funcionamento e as limitações do código computacional desenvolvido.

4.3.1 Aplicação ao caso unidimensional

Um modelo unidimensional foi desenvolvido, similar àqueles descritos na Seção 2.4, e as respostas MT por ele geradas são usadas como parâmetros de entrada e saída para validar o algoritmo desenvolvido para calcular a transformada. O modelo unidimensional consiste de um terreno com 4 camadas, onde a camada superficial tem resistividade $\rho_1 = 800 \Omega \text{ m}$ e espessura $e_1 = 2000 \text{ m}$, seguida de camadas com resistividades e espessuras de $\rho_2 = 6000 \Omega \text{ m}$, $e_2 = 18000 \text{ m}$; $\rho_3 = 400 \Omega \text{ m}$, $e_3 = 25000 \text{ m}$; e $\rho_4 = 2000 \Omega \text{ m}$, $e_4 = \infty$.

Com esse modelo e a Fórmula 2.23 foram gerados dois conjuntos de impedâncias. O primeiro abrange o período que vai de $1 \times 10^{-3} \text{ s}$ a $1 \times 10^4 \text{ s}$ e o segundo se estende de $1 \times 10^{-3} \text{ s}$ até $1 \times 10^6 \text{ s}$. O uso desses dois intervalos de períodos vai nos permitir observar algumas limitações na implantação da transformada de Kramers-Kronig. Ambos os conjuntos contém 4 períodos por década (entende-se aqui como década o intervalo que se estende de 10^n a 10^{n+1} , com $n \in \mathbb{Z}$). Cada um desses conjuntos servirão de entrada para as rotinas que calculam as transformadas *kk-ir* (Equação 4.3a) e *kk-ri* (Equação 4.3b).

As Figuras 4.1 a 4.4 apresentam funções de transferência MT comparadas com os resultados das validações das funções que calculam as transformadas *kk-ri* e *kk-ir*. Nesse caso, os dados MT são compostos de seis gráficos, onde os da primeira

coluna à esquerda representam as curvas da resistividade aparente (Equação 2.20) e fase (Equação 2.22) das componentes xy e yx , e os quatro gráficos restantes representam a parte real e imaginária do tensor de impedâncias (Equação 2.28), escalonados por $\sqrt{\mu_0\omega}$. Esse escalonamento permite uma melhor visualização dos dados em uma escala linear (GROOM et al., 1993) e é semelhante à impedância normalizada $Y = Z/\sqrt{i\omega\mu_0}$ proposta por Basokur (1994), com o fator imaginário na raiz do denominador removido para que a fase da impedância não fosse alterada.

As Figuras 4.1 a 4.4 são facilmente identificáveis como 1D pois, conforme previsto pela Equação 2.29, os gráficos para Z_{xx} e Z_{yy} indicam que esses elementos são iguais a zero, enquanto os elementos Z_{xy} e Z_{yx} estão invertidos. A Figura 4.1 mostra o resultado das rotinas utilizadas para calcular a transformada $kk-ri$ para um intervalo de períodos de 1×10^{-3} s a 1×10^4 s. Nesse caso, percebe-se que as rotinas que calculam a transformada foram capazes de reproduzir o resultado esperado para o modelo. Por outro lado, a Figura 4.2 calcula a mesma transformada para um intervalo de períodos estendido até 1×10^6 s. Analogamente ao caso anterior, também nessa situação as rotinas foram capazes de reproduzir o resultado esperado para a recuperação da parte imaginária a partir da parte real do tensor.

Já na Figura 4.3 observa-se que para a transformada $kk-ir$ (4.3a), calculada com um intervalo de 1×10^{-3} s a 1×10^4 s, começa a haver um desvio sistemático mais acentuado, especialmente para períodos acima de 1×10^3 s, no cálculo da parte real a partir da parte imaginária. Esse desvio é consequência da formula de extrapolação (Equação 4.4), visto que a Figura 4.4 indica que esse efeito é bastante atenuado quando o período se estende a 1×10^6 s. O efeito da função de extrapolação pode ser identificado nos gráficos observando que a Função 4.4 implicaria em uma reta horizontal para os gráficos $\Re Z$ e $\Im Z$ das Figuras 4.1 a 4.4. Comparando as Figuras 4.1 e 4.2 vemos que a parte real já mostra o comportamento de semiespaço infinito (tendência para uma linha reta no valor da impedância) a partir de 1×10^4 s, mas a parte imaginária só alcança esse resultado de semiespaço infinito a partir de 1×10^5 s. Por essa razão, a transformada $kk-ri$ (que tem como entrada a parte real) apresenta um resultado satisfatório para períodos até 1×10^4 s, enquanto a transformada $kk-ir$ (cuja entrada é a parte imaginária) precisa de períodos maiores para se estabilizar no semi-espaço infinito. Cumpre observar que apesar da diferença ser facilmente verificável nos gráficos $\Re Z$ e $\Im Z$ ela é praticamente imperceptível nos gráficos de resistividade aparente (ρ) e fase (ϕ).

Figura 4.1 – Comparação da transformada kk - ri com resultados sintéticos para o modelo unidimensional. A discriminação das funções mostradas no gráfico é apresentada na Figura 3.1. Os pontos azuis foram obtidos por meio de modelagem unidimensional e os pontos em vermelho foram obtidos usando a Equação 4.3b. Para poder distinguir os casos em que as rotinas que calculam a transformada obtiveram resultados com melhor precisão daqueles em que as rotinas mostraram resultados mais imprecisos, os dados do modelo estão em tamanho maior.

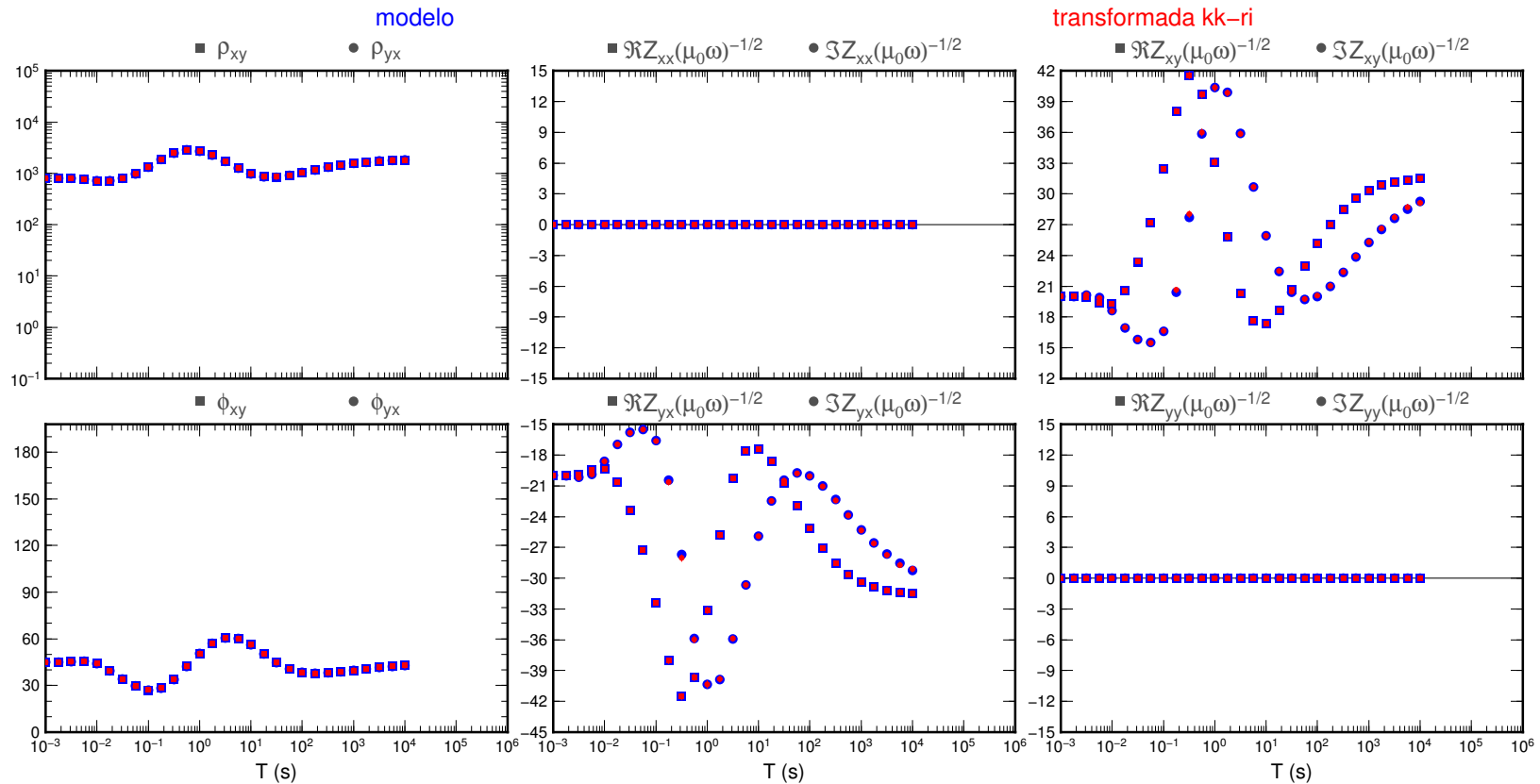


Figura 4.2 – Transformada $kk-ri$ comparada com dados sintéticos para o modelo unidimensional usando períodos estendidos. As cores seguem o mesmo padrão da Figura 4.1.

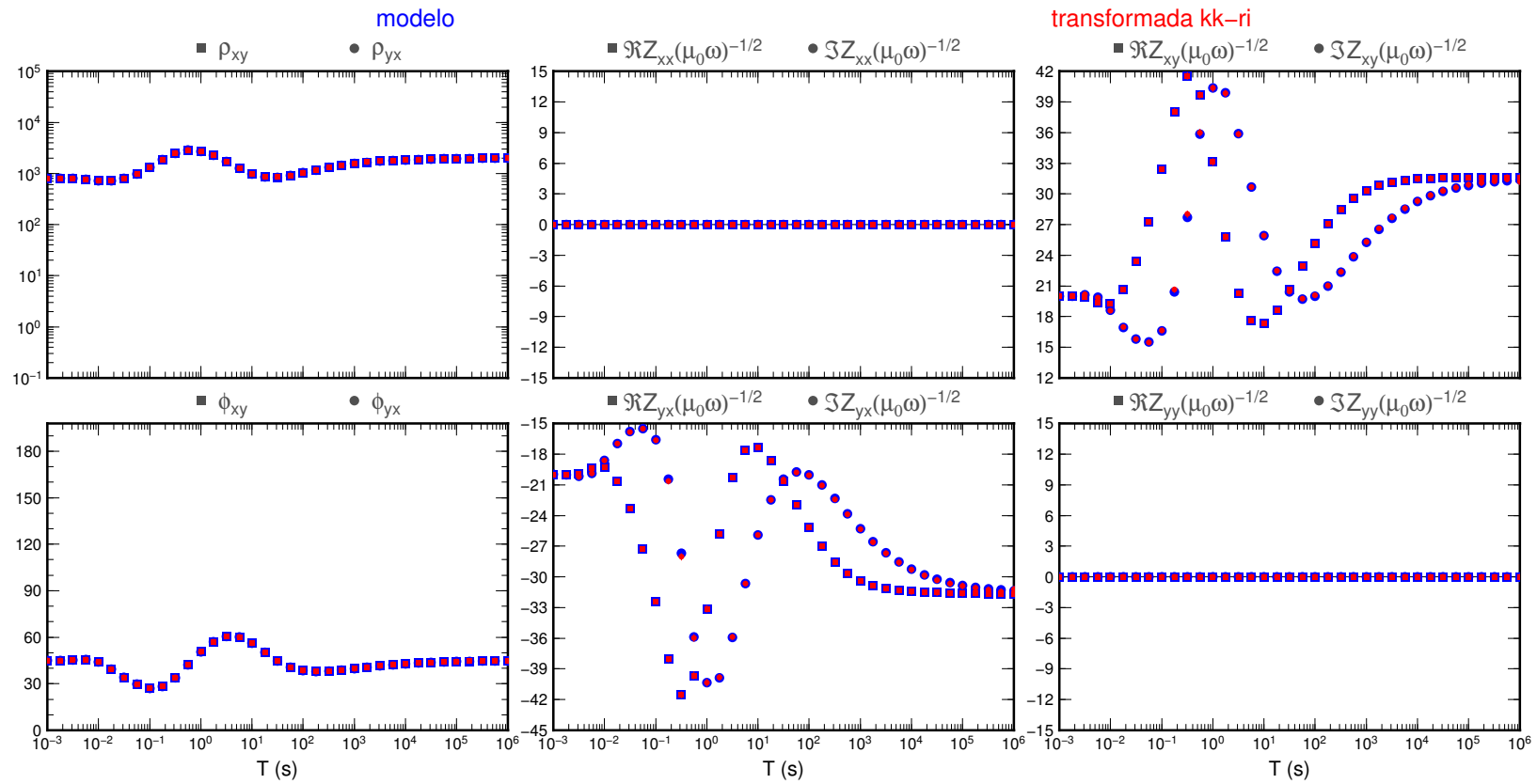


Figura 4.3 – Transformada kk - ir comparada com dados sintéticos para o modelo unidimensional. As cores seguem o mesmo padrão da Figura 4.1.

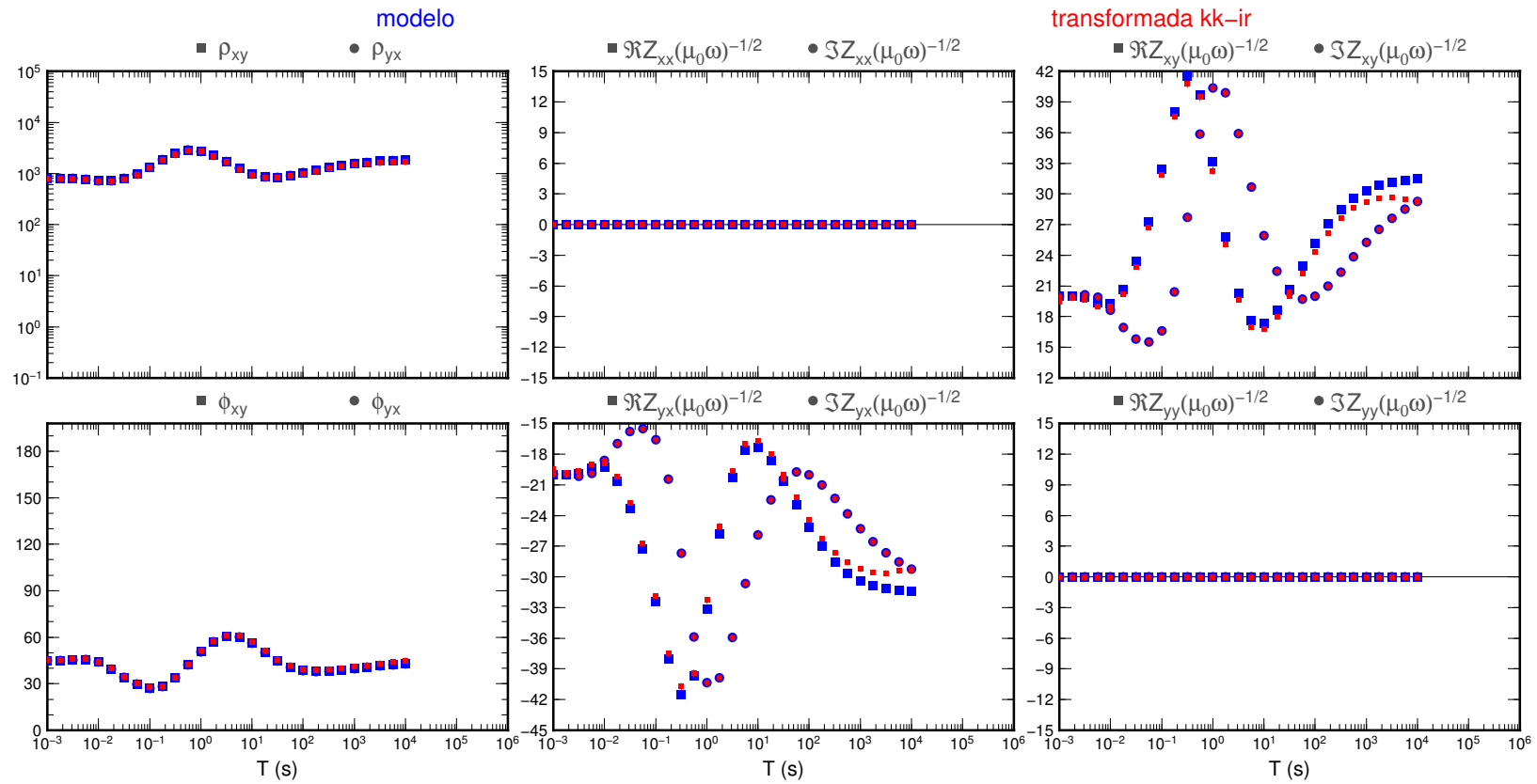
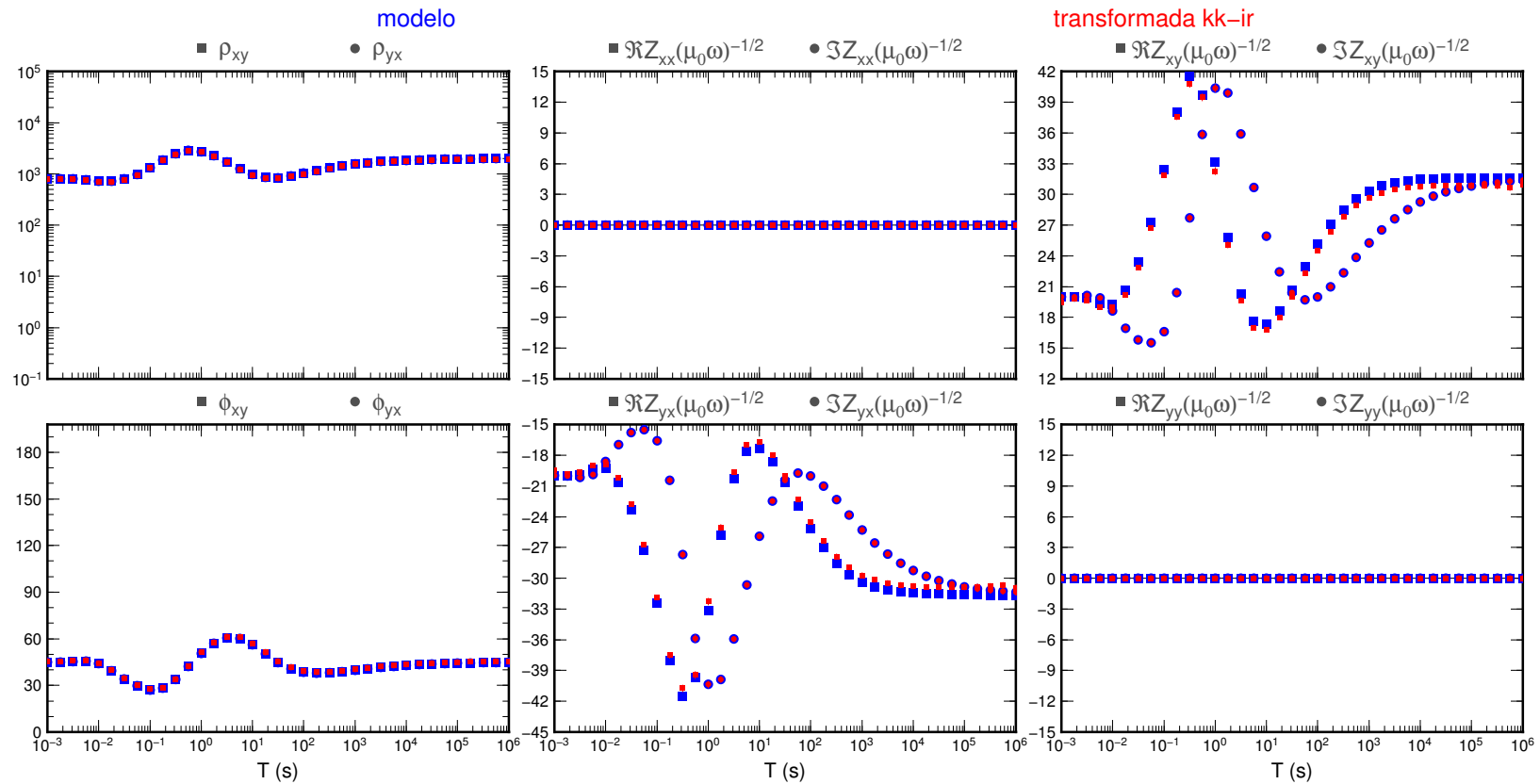


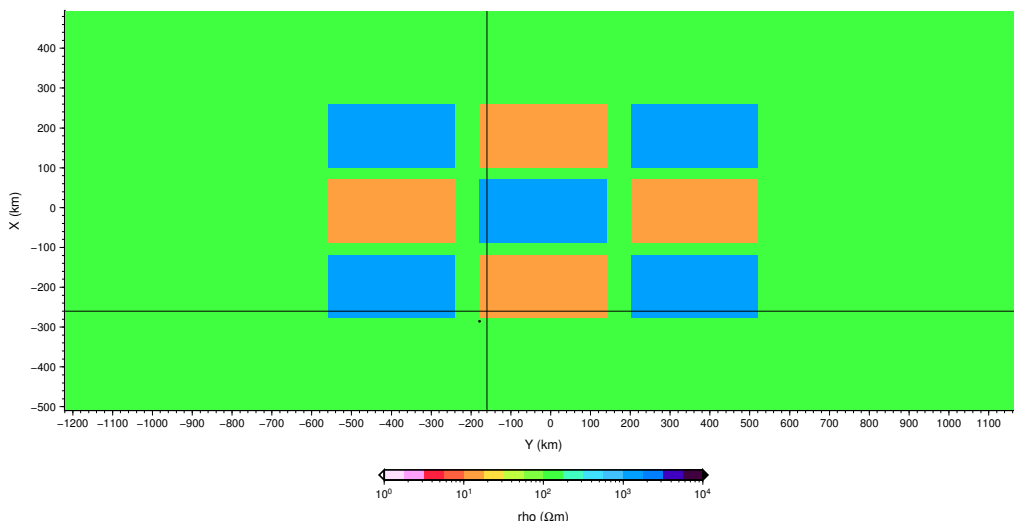
Figura 4.4 – Transformada kk - ir comparada com dados sintéticos para o modelo unidimensional usando períodos estendidos. As cores seguem o mesmo padrão da Figura 4.1.



4.3.2 Aplicação ao caso tridimensional

Um procedimento similar foi feito para dados gerados por um modelo sintético tridimensional, ilustrado nas Figuras 4.5, 4.6 e 4.7. Nesse caso, os dados sintéticos foram obtidos dos arquivos de teste fornecidos junto com o código *ModEM* de inversão tridimensional de dados MT (EGBERT; KELBERT, 2012). Os dados foram derivados a partir de um modelo 3D que consiste de blocos com resistividade de 10 e 1000 Ω m distribuídos em um semiespaço homogêneo de 100 Ω m. São fornecidas as funções de transferência dos elementos do tensor de impedâncias para 12 períodos, logaritmicamente equiespaçados entre 10 e 10 000 s.

Figura 4.5 – Visão superior para o modelo 3D sintético. O ponto representa a posição da estação ($x = -285.888676$, $y = -178.568076$) cujos dados sintéticos são utilizados e as linhas representam os cortes apresentados nas Figuras 4.6 e 4.7



As Figuras 4.8 e 4.9 mostram o resultado da transformada de Kramers-Kronig para esse modelo sintético tridimensional. Conforme esperado para um modelo 3D, as funções de transferência sintéticas não se anulam em nenhum dos quatro elementos do tensor e não mostram a simetria observada anteriormente no caso 1D. Na transformada *kk-ri* da Figura 4.8 observa-se que o código tem dificuldade para reproduzir os elementos imaginários dos quatro elementos do tensor de impedâncias para os períodos extremos (períodos mais curtos e o período mais longo). Usando os resultados anteriormente discutidos para o modelo 1D, pode-se argumentar que isso se deve ao comportamento da parte real dos elementos do

Figura 4.6 – Corte vertical do plano em $x = 260$ para o modelo 3D sintético. O triângulo na superfície marca a posição da estação.

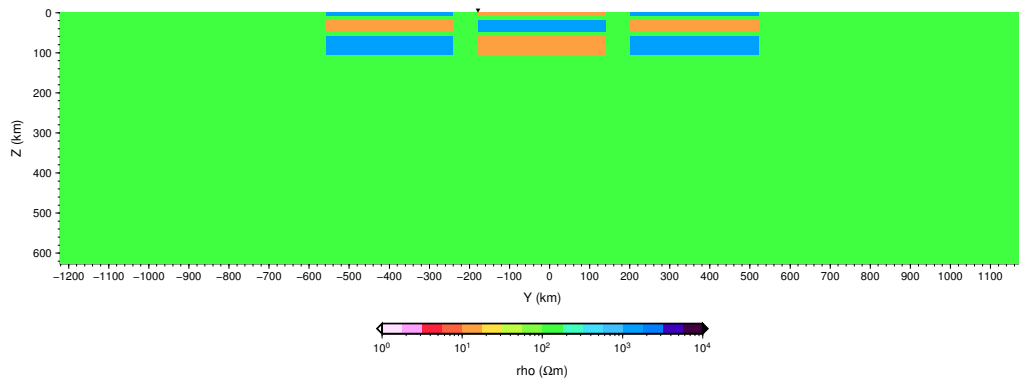
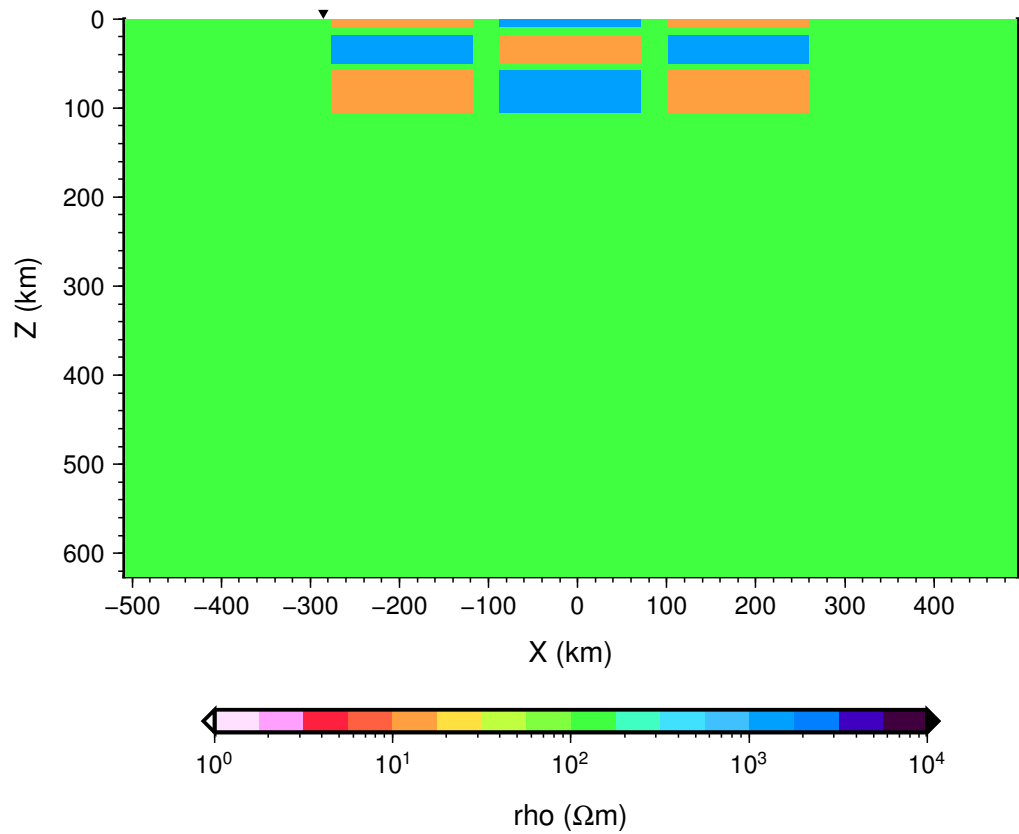


Figura 4.7 – Corte vertical do plano em $y = 160$ para o modelo 3D sintético. O triângulo na superfície marca a posição da estação.



tensor que em todos os casos ainda não alcançaram a estabilidade (valor aproximadamente constante) que seria esperada para um semiespaço homogêneo tanto em períodos curtos como em períodos longos (Equação 4.4). Em ambas as situações seria necessário incorporar dados de períodos muito mais curtos e muito mais longos do que os utilizados para estabilizar a transformada. Por outro lado, as resistividades aparentes dos elementos da diagonal secundária do tensor são mais bem comportadas, com os desvios aparecendo apenas no maior período utilizado. As fases são mais sensíveis a essa alteração nos resultados do tensor e mostram maior discrepância em relação aos resultados sintéticos gerados pelo modelo.

Na transformada *kk-ir* da Figura 4.9 observa-se que os períodos mais curtos da parte imaginária já tendem a um valor estável de um semiespaço homogêneo próximo da superfície. Com isso, a transformada reproduz de forma adequada a componente real nesses períodos. O mesmo não ocorre nos períodos mais longos, onde a parte imaginária ainda não alcançou a estabilidade do semiespaço infinito inferior. Como consequência, ocorrem discrepâncias entre os resultados gerados pela transformada e pelo modelo sintético. Novamente, a resistividade aparente é pouco afetada por esses resultados discrepantes dos elementos do tensor, enquanto a fase aqui também é pouco afetada.

Figura 4.8 – Transformada kk - ri comparada com dados sintéticos para o modelo tridimensional fornecido junto com o código *ModEM*. As cores seguem o mesmo padrão da Figura 4.3.

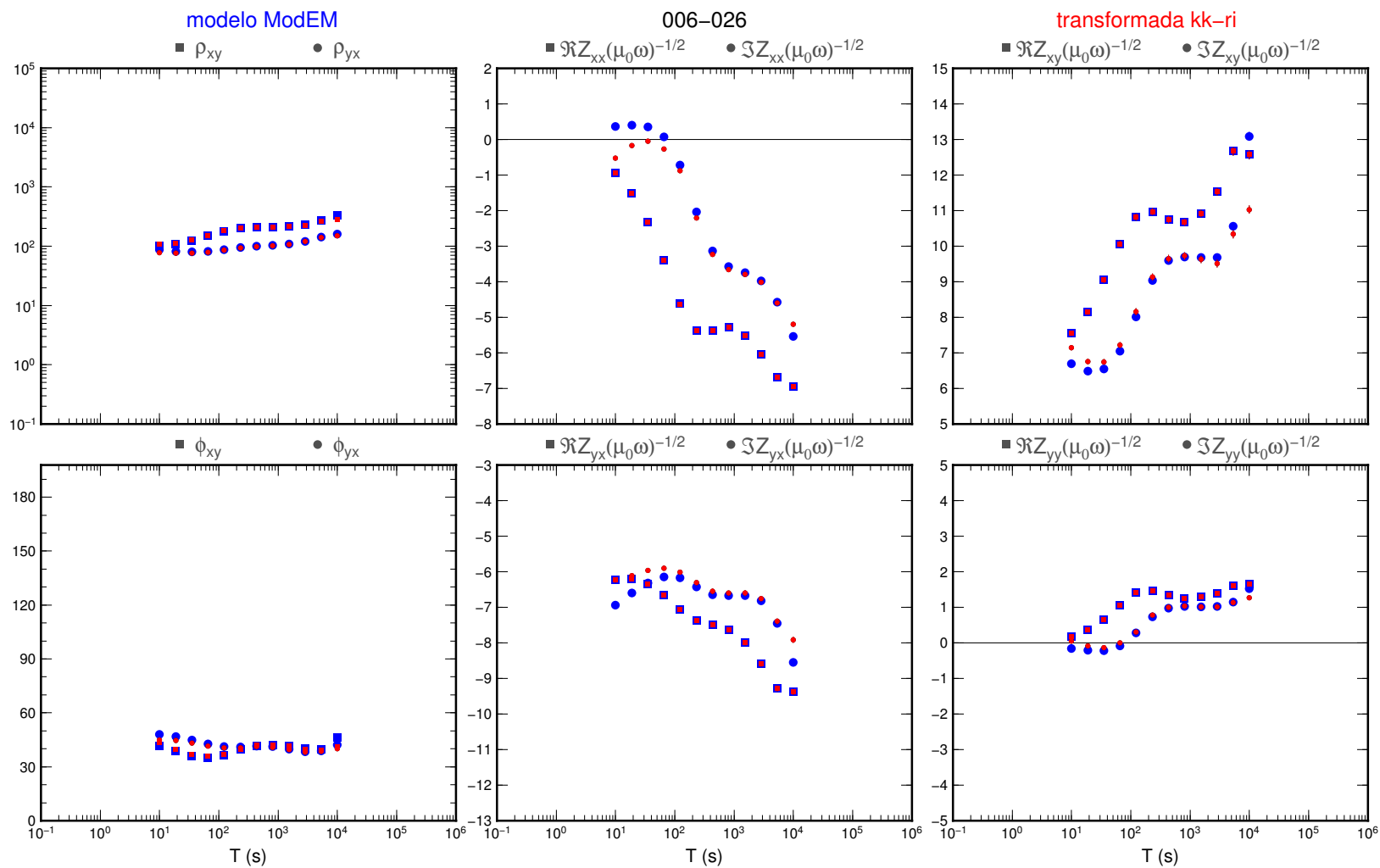
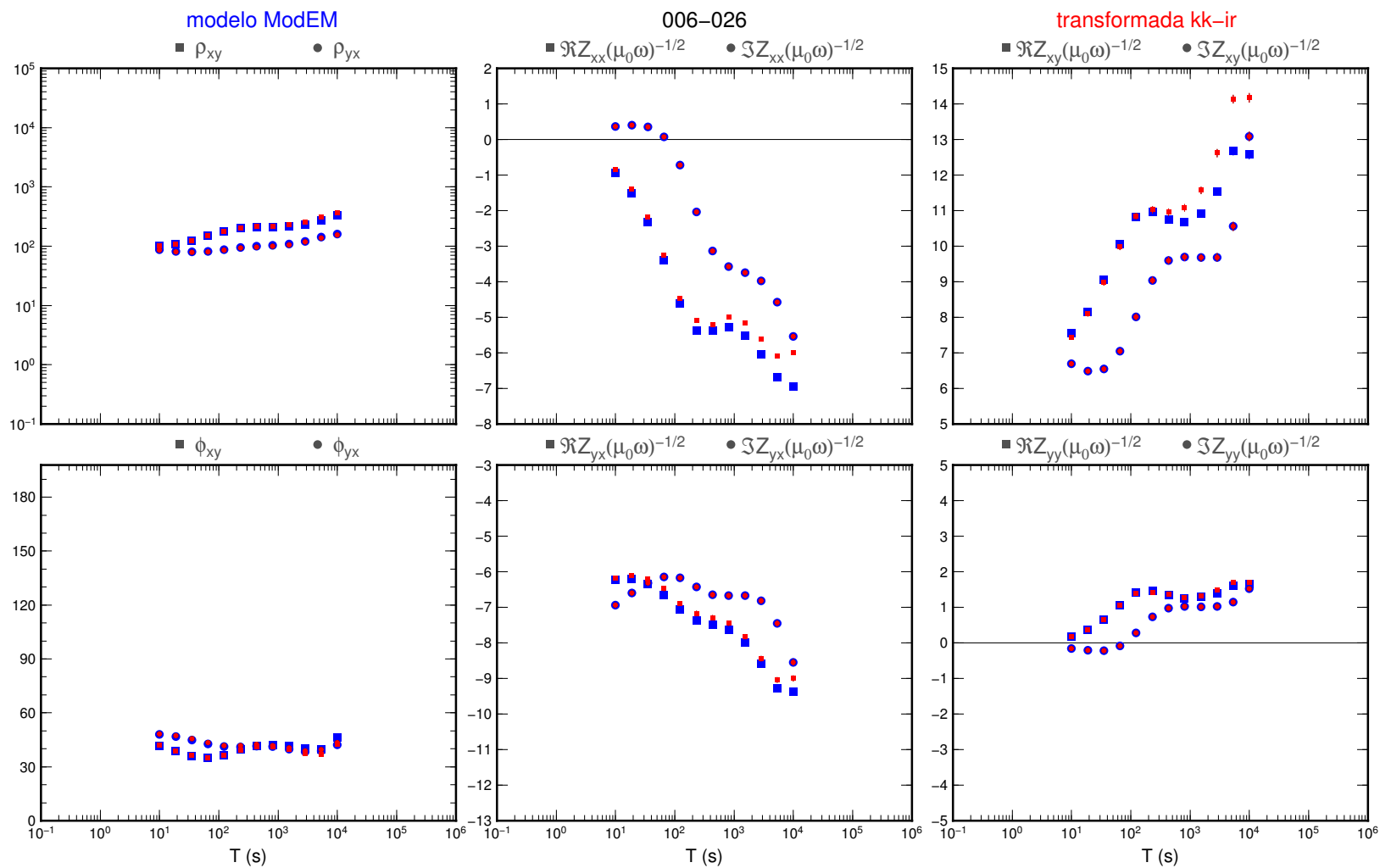


Figura 4.9 – Transformada kk - ir comparada com dados sintéticos para o modelo tridimensional fornecido junto com o código *ModEM*. As cores seguem o mesmo padrão da Figura 4.3.



De forma similar ao caso 1D, um estudo detalhado do efeito da expansão dos períodos analisados também foi realizado para o caso desse modelo 3D. O código *ModEM* de Egbert e Kelbert (2012) foi usado para obter as funções de transferência MT para o modelo sintético 3D apresentado nas Figuras 4.5, 4.6 e 4.7, em um intervalo de períodos estendido em relação àquele originalmente fornecido nos arquivos de teste do código. Usando cálculo direto, foram derivadas as funções de transferência para 30 períodos logaritmicamente equiespaçados entre 0, 231 e 5 336 800 s. A Figura 4.10 mostra as funções de transferência obtidas simplesmente expandindo o intervalo de períodos, sem qualquer alteração no modelo de entrada fornecido junto com o programa. Observa-se que os resultados são idênticos nos períodos originais disponibilizados junto com o código, indicando que o código fonte está adequadamente implementado. Porém, quando se analisa em detalhes os resultados de resistividade aparente e fase dos elementos da diagonal secundária observa-se que a resistividade cai abruptamente com o aumento do período e a fase tende para 90° nos períodos mais longos. Esse é um resultado inesperado considerando que a última camada do modelo 3D é um semiespaço condutor com resistividade de $100 \Omega \text{ m}$. Nessa situação seria esperado que a resistividade aparente nos períodos mais longos estivesse próxima a esse valor, enquanto a fase deveria se aproximar de 45° . O resultado anômalo da Figura 4.10 indica que ao aumentar significativamente o período máximo a ser analisado foram introduzidos problemas de condição de contorno na modelagem. Ou seja, para analisar esses períodos muito maiores é necessário aumentar também a profundidade máxima do semiespaço profundo.

O modelo original tem espessura total de 626 km na direção vertical. Para eliminar problemas numéricos relacionados com condições de contorno (limite inferior do modelo) foram acrescentadas 4 camadas à base do modelo (todas com a mesma resistividade de $100 \Omega \text{ m}$ do semiespaço condutor), estendendo a espessura vertical total a 5 426 km. A Figura 4.11 mostra uma comparação das funções de transferência fornecidas nos arquivos de teste do código com aquelas obtidas pelo modelo estendido verticalmente (inclusão de 4 novas camadas na parte inferior da grade) para os mesmos períodos expandidos. Note-se que os problemas nas condições de contorno foram agora eliminados, com os valores de resistividade aparente e fase apresentando o comportamento esperado. Também, observam-se diferenças nos períodos mais longos dos resultados fornecidos junto com o código *ModEM*, indicando que esses valores apresentam problemas de estabilidade (profundidade máxima deveria ser muito maior que a usada de 626 km para o cálculo direto).

Figura 4.10 – Dados sintéticos para o modelo 3D original fornecido com o código *ModEM*, em vermelho, expandidos para períodos entre 0,231 s e 5 336 800 s, em azul.

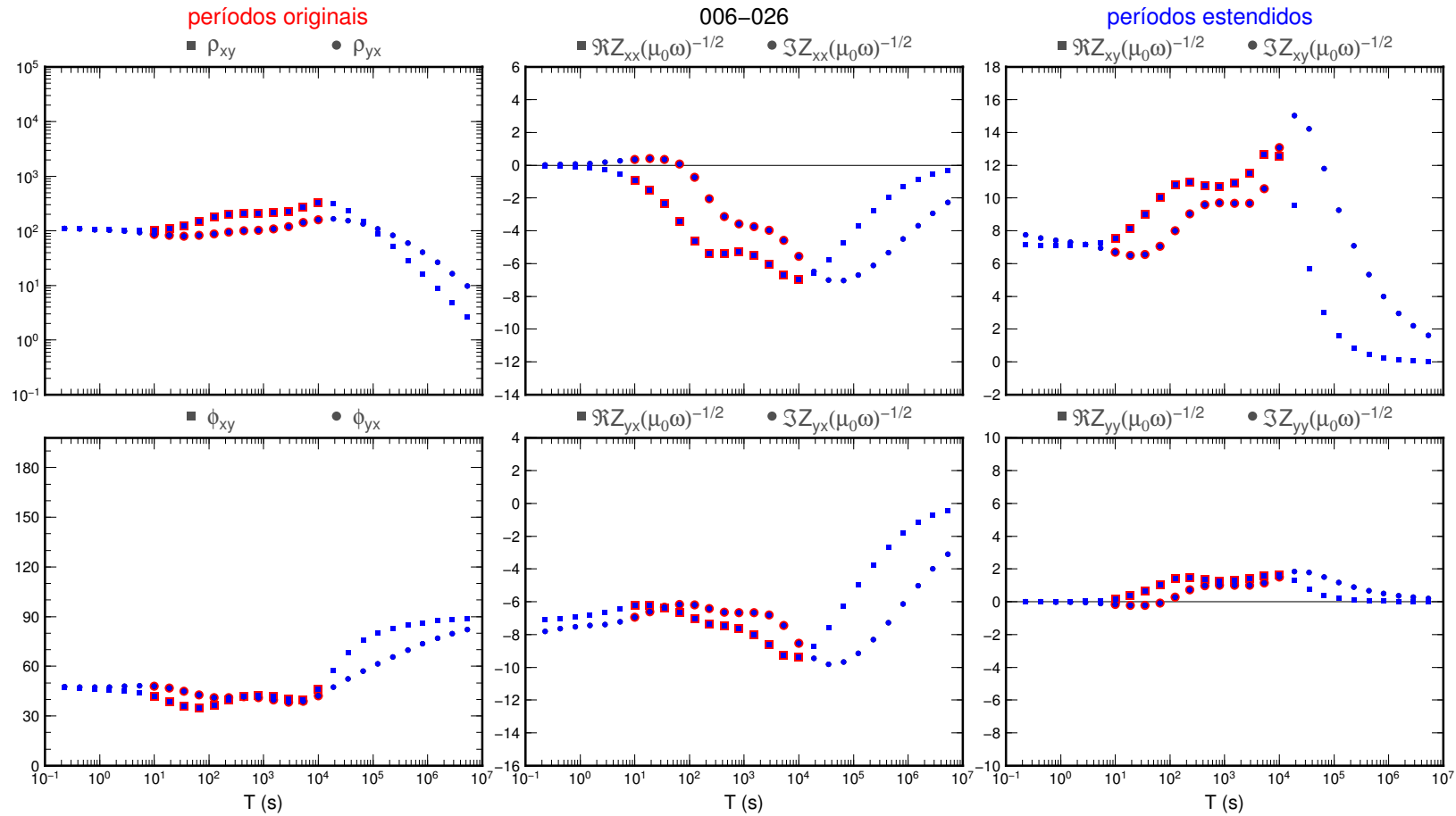
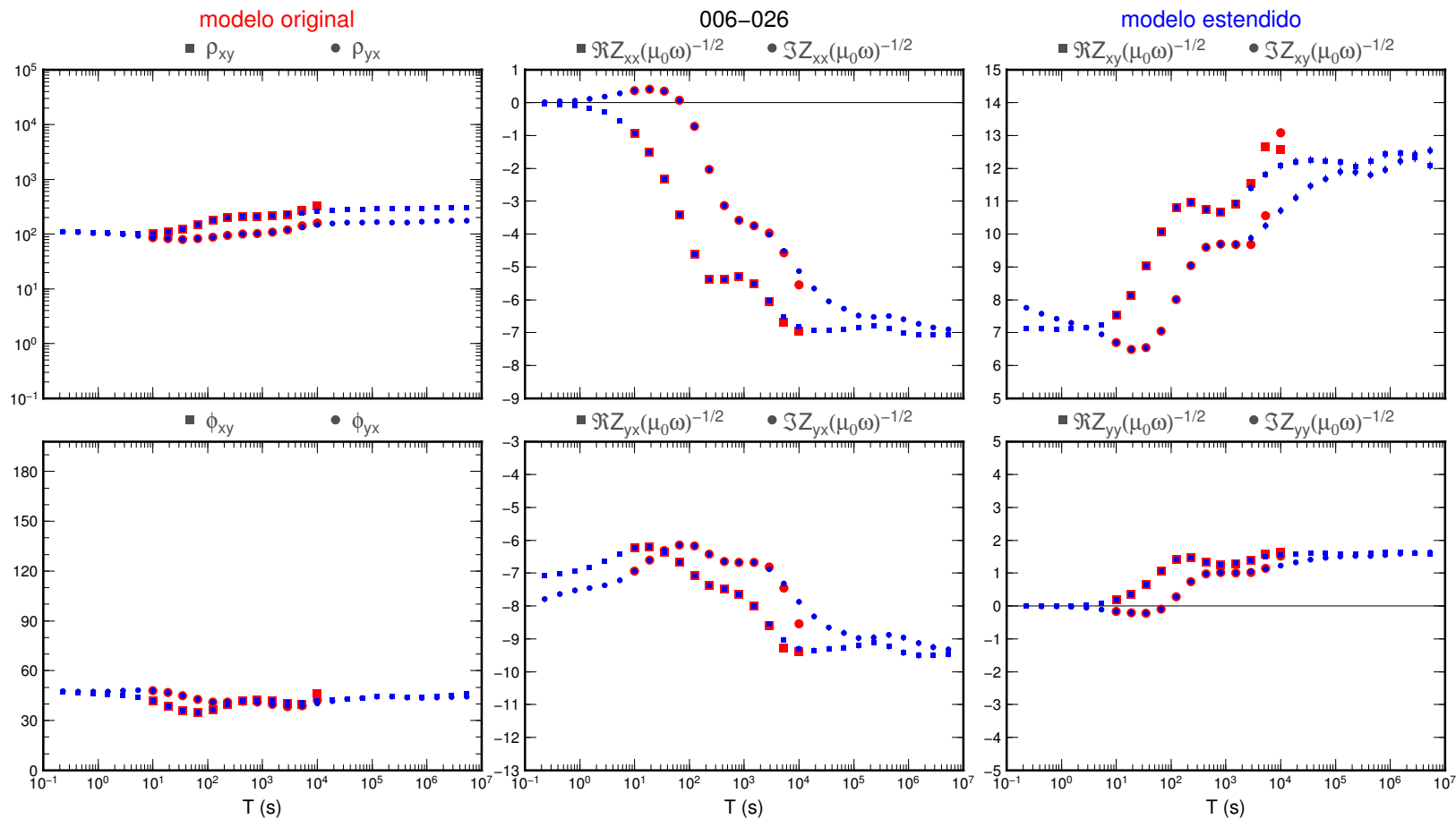


Figura 4.11 – Dados sintéticos para o modelo 3D original fornecido com o código *ModEM*, em vermelho, comparados com os gerados pelo modelo estendido verticalmente (inclusão de 4 novas camadas) para períodos entre 0,231 s e 5 336 800 s, em azul.



As funções de transferência obtidas com os períodos expandidos e a parte inferior do modelo estendida foram utilizadas para testar a transformada de Kramers-Kronig no caso 3D. As Figuras 4.12 e 4.13 mostram respectivamente o resultado para transformada $kk-ri$ e $kk-ir$. Os resultados mostram que o código agora reproduz mais adequadamente tanto a parte real como a parte imaginária dos elementos de impedância do que anteriormente observado nas Figuras 4.8 e 4.9. Por outro lado, analogamente ao observado no caso 1D (Figuras 4.2 e 4.4), a transformada $kk-ri$ é aparentemente mais estável e recupera os valores imaginários melhor do que a transformada $kk-ir$ recupera os valores reais.

A única discrepância entre os resultados esperados e aqueles gerados pela transformada $kk-ri$ ocorre nos períodos inferiores a 2 s dos elementos Z_{xy} e Z_{yx} (ver Figura 4.12). Analisando o modelo, verificou-se que a estação modelada encontra-se a quase 10 km de distância do bloco condutor mais próximo. Para uma resistividade de $100 \Omega \text{ m}$, isso equivale a mais que 3 *skin depths* de distância para o menor período modelado. Nesse caso, as partes reais e imaginárias deveriam ser as mesmas, como em um semiespaço infinito. A separação entre as partes reais e imaginárias nas componentes Z_{xy} e Z_{yx} só pode estar associada a uma grade grosseira para as primeiras camadas na direção vertical. Foi feito, então, um novo modelo, refinando os primeiros 2,5 km do modelo na direção vertical. As 5 camadas presentes nesse intervalo foram discretizadas para 15 camadas. Os resultados obtidos com esse novo modelo são mostrados nas Figuras 4.14 e 4.15. Pode se observar que as partes real e imaginária para os períodos mais curtos ficaram muito próximos e que a transformada $kk-ri$ ficou quase perfeita.

Como resultado desses testes, pode-se afirmar que, com uma devida extrapolação das curvas de resposta MT para períodos mais curtos e mais longos que permitam obter estabilidade (*enxergar* o semiespaço homogêneo na superfície e na base do modelo), o algoritmo desenvolvido funciona muito bem para este teste de um modelo 3D bastante intenso. Ainda mais importante é a oportunidade que o algoritmo oferece de ser utilizado para verificar a consistência de um cálculo direto 3D, auxiliando no diagnóstico de problemas nas condições de contorno e na resolução da grade do modelo.

Figura 4.12 – Transformada kk - ri comparada com dados sintéticos expandidos para períodos mais curtos e mais longos no modelo tridimensional.

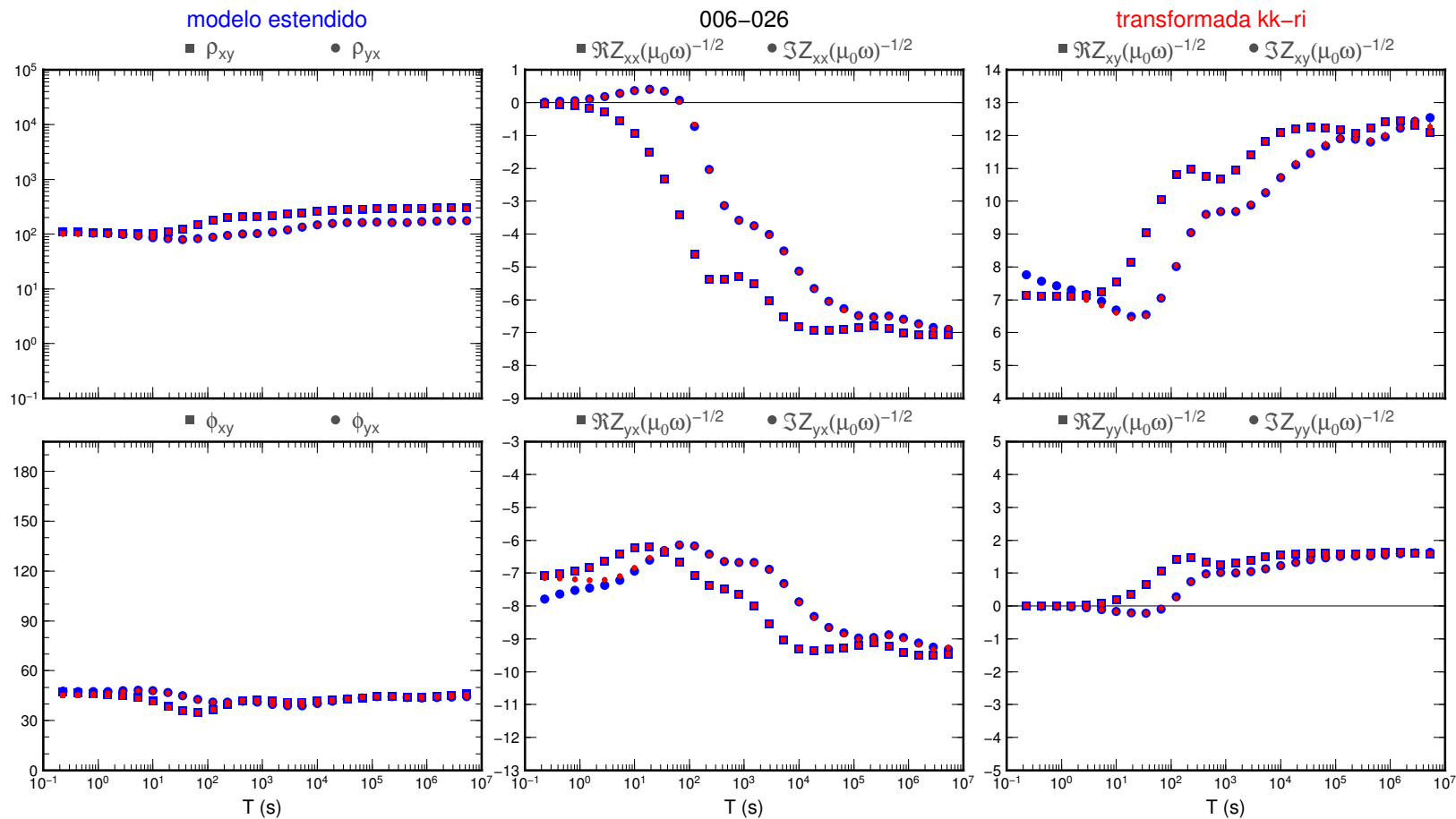


Figura 4.13 – Transformada kk - ir comparada com dados sintéticos expandidos para períodos mais curtos e mais longos no modelo tridimensional.

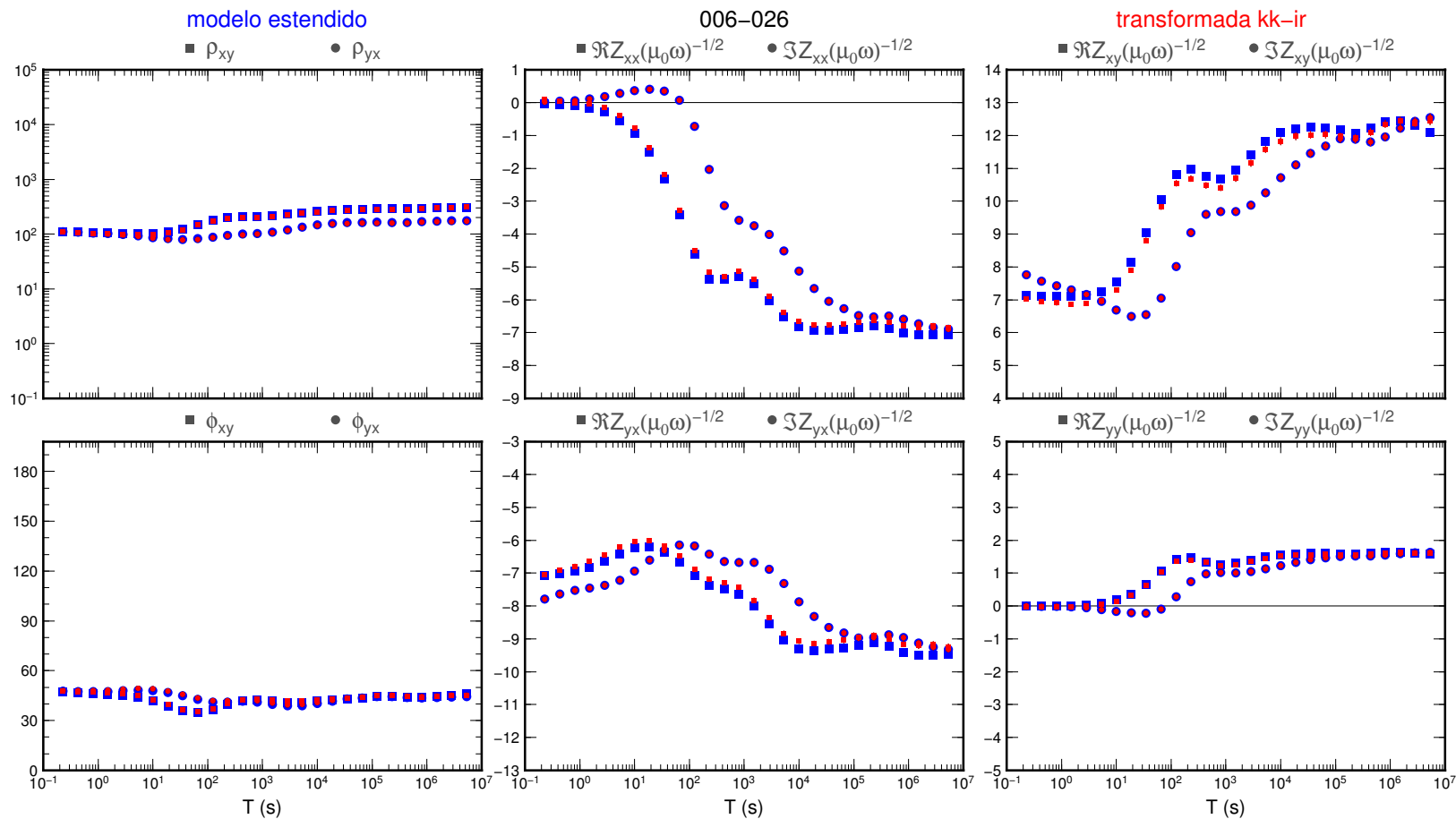


Figura 4.14 – Transformada kk - ri comparada com dados sintéticos expandidos para períodos mais curtos e mais longos no modelo 3D, com maior discretização da parte superficial do modelo.

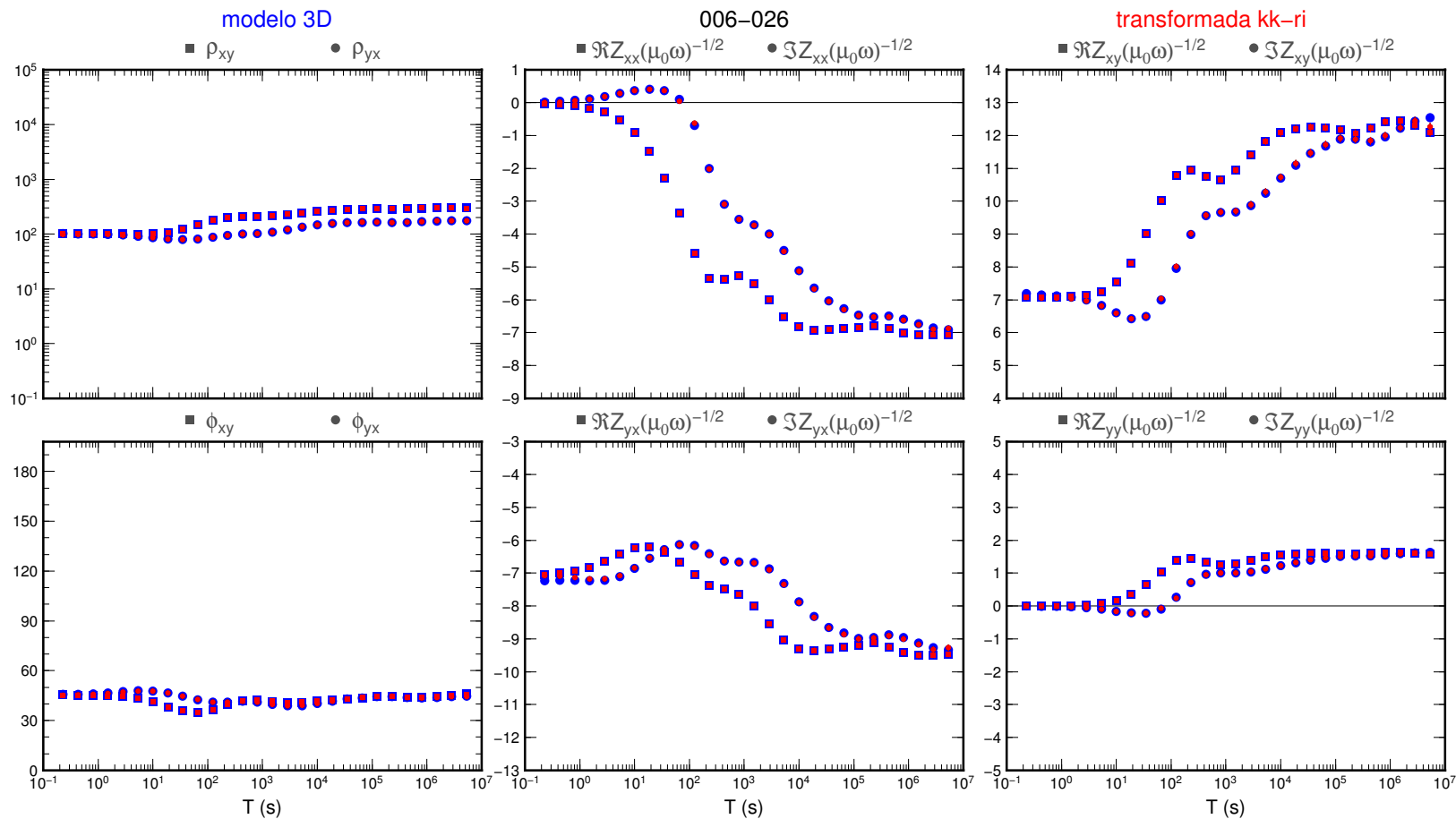
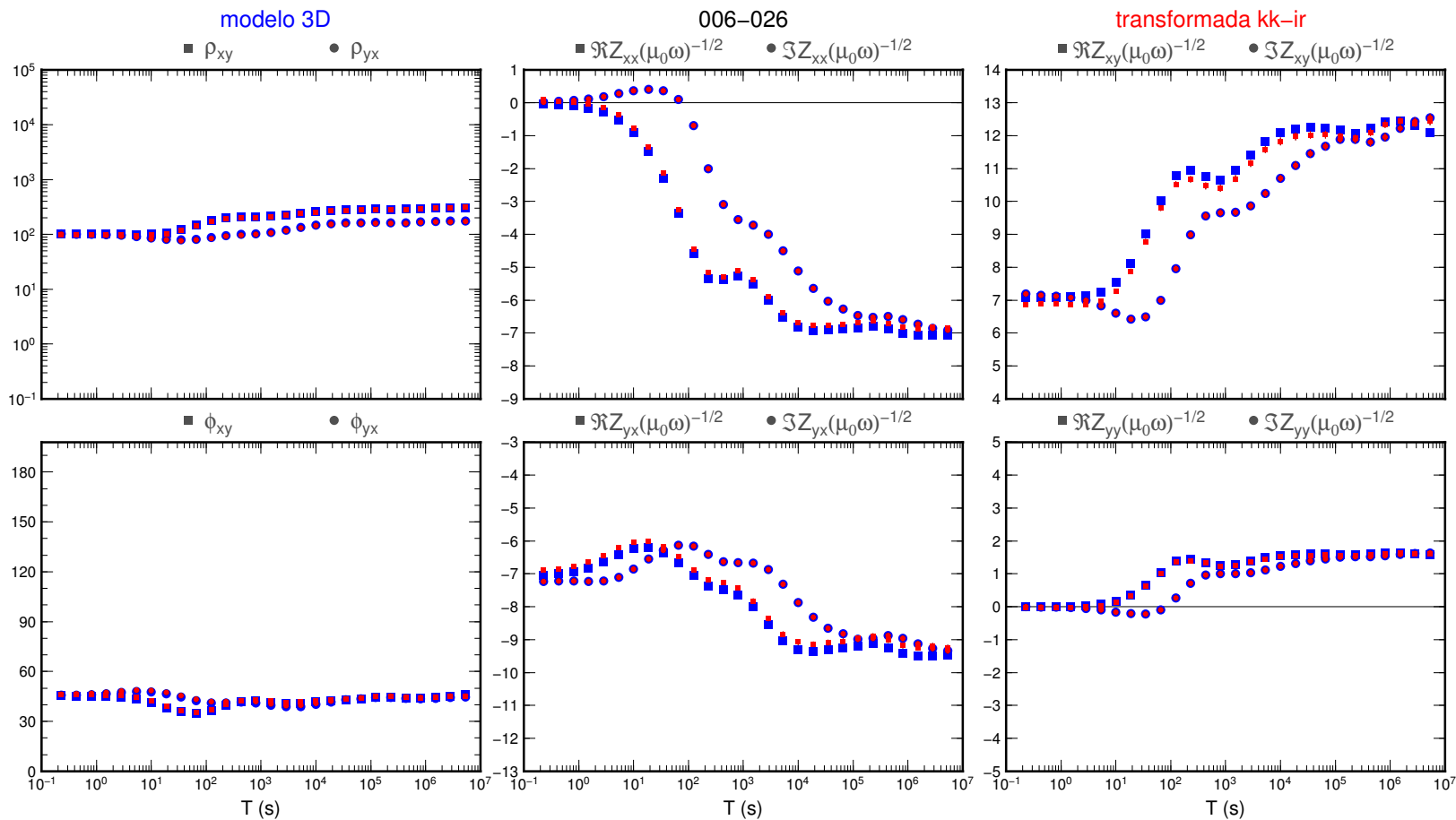


Figura 4.15 – Transformada $kk-ir$ comparada com dados sintéticos expandidos para períodos mais curtos e mais longos no modelo 3D, com maior discretização da parte superficial do modelo.



Como conclusão desses exercícios com modelos sintéticos pode-se afirmar que o algoritmo desenvolvido consegue recuperar as funções de transferência MT individualmente a partir das partes reais e imaginárias dos elementos do tensor. Esse resultado é válido inclusive para situações tridimensionais. Há limitações no uso da técnica quando se tem um intervalo de períodos que não alcança a estabilidade da resposta que seria gerada por semiespaços homogêneos em períodos curtos e em períodos longos (função de transferência tem que convergir para um valor constante nos extremos do espectro de períodos, de acordo com a simplificação utilizada na Equação 4.4). Também, a resistividade aparente é mais robusta que a fase a essas limitações da transformada.

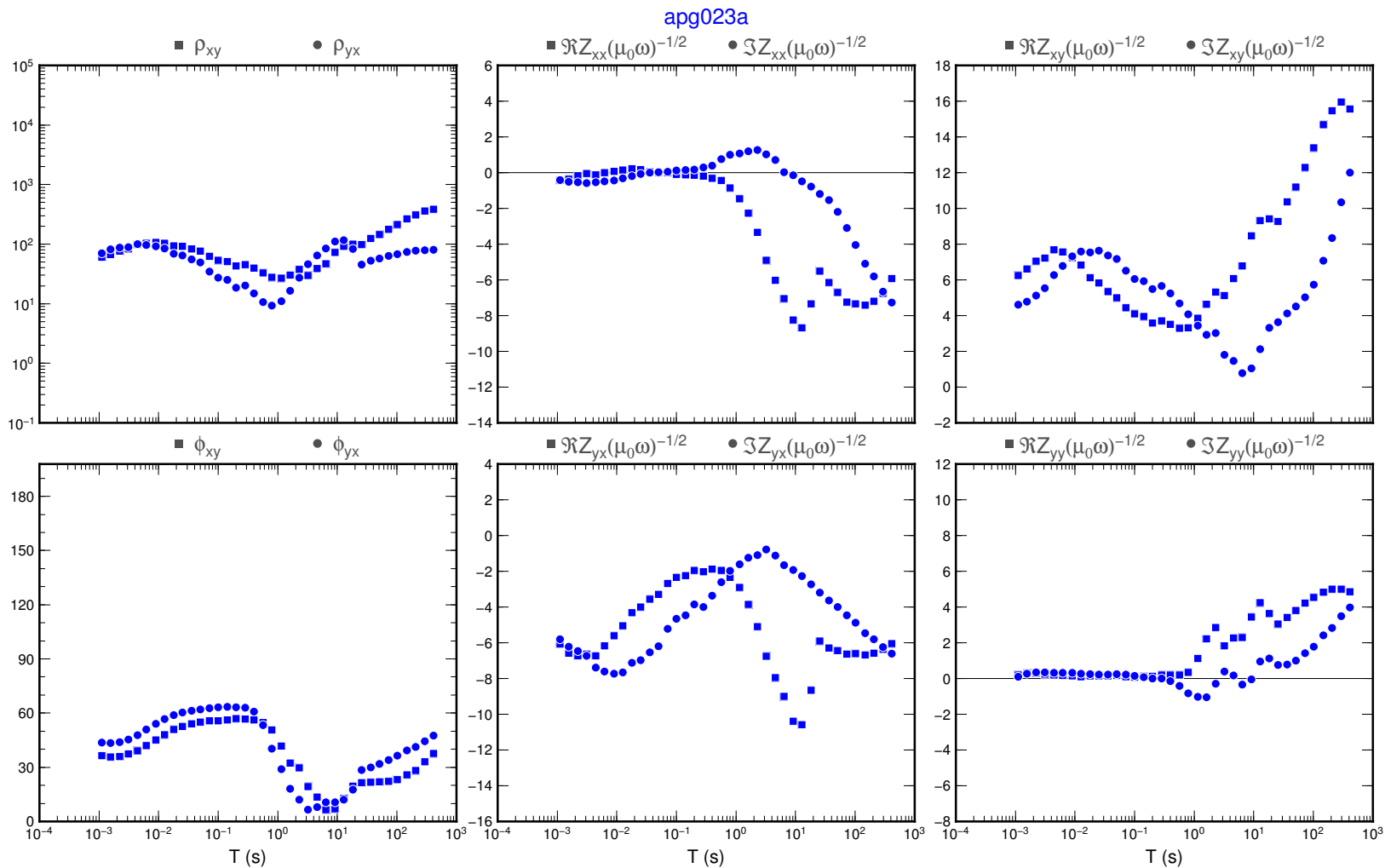
5 APLICAÇÃO A DADOS EXPERIMENTAIS

A metodologia desenvolvida no capítulo anterior foi aplicada a dados experimentais de sondagens MT coletadas no país. Para exemplificar os resultados obtidos foi escolhida uma estação (*apg023a*) adquirida sobre as rochas vulcano-sedimentares da Bacia do Paraná. Os dados dessa estação foram adquiridos e processados com as técnicas mais modernas atualmente disponíveis para o método MT, usando o algoritmo robusto de Egbert e Booker (1986) e também a técnica de referência remota (GAMBLE et al., 1979) para minimizar ruídos locais não correlacionados. A Figura 5.1 mostra o resultado desse processamento com as 4 componentes do tensor de impedância e as fases e resistividades aparentes da diagonal secundária.

Apesar das técnicas modernas de aquisição e processamento de dados, a estação é afetada pelos mesmos ruídos da linha HVDC entre Itaipú e Ibiúna, somente que em nível mais moderado do aquele exibido para a estação *pru036a*. A estação *apg023a* encontra-se a pouco mais de 100 km da linha HVDC e em uma região com pacote sedimentar bem mais espesso, o que contribui para amortecer o nível do ruído. Pode-se notar que o ruído principal é observado na parte real das impedâncias Z_{xx} e Z_{yx} , as quais apresentam valores perturbados em períodos entre aproximadamente 0,5 s e 20 s, com as funções de transferência caindo abruptamente (pico da distorção em torno de 10 s). Nas outras duas impedâncias, Z_{xy} e Z_{yy} , a distorção aparece também na parte imaginária. Em Z_{xy} , o pico descendente centrado em torno de 10 s na componente imaginária afeta as funções de transferência entre aproximadamente 2 s e 20 s. Na componente Z_{yy} , dois picos ascendentes são identificados tanto na parte real como na parte imaginária, no primeiro caso em torno de 2 s e 10 s e no segundo caso centrados em 3 s e 10 s. Há ainda oscilações na parte imaginária de Z_{xy} e Z_{yx} em torno de 0,1 s. Essas diferentes distorções, típicas de sondagens MT de campo, mostram a complexidade do ruído (composto por efeitos de corrente contínua e corrente alternada) afetando o sinal EM natural e tornando mais complexa a aplicação da transformada de Kramers-Kronig.

As resistividades aparentes e fases para os elementos da diagonal secundária mostradas nessa mesma figura refletem as principais distorções acima descritas. Observa-se claramente uma quebra no comportamento suave da resistividade aparente ρ_{yx} , entre aproximadamente 1 s e 20 s, enquanto a resistividade aparente ρ_{xy} aparenta ser melhor comportada em todo o intervalo de períodos analisados. A fase em ambos os elementos do tensor aparece perturbada em torno desse mesmo intervalo, aproximando-se de 0° entre 2 s e 20 s.

Figura 5.1 – Funções de transferência de estação MT coletada no estado do Paraná. São mostradas as variações de resistividade aparente (ρ_{xy} e ρ_{yx} , em Ω m), fase (ϕ_{xy} e ϕ_{yx} , em graus), e as componentes reais e imaginárias de funções derivadas dos elementos do tensor de impedância (Z_{xx} , Z_{xy} , Z_{yx} e Z_{yy}).



De acordo com a forma de utilização da transformada de Kramers-Kronig descrita na Seção 4.1, vão-se descrever os resultados obtidos usando como entrada as componentes das impedâncias geradas pelo método de regressão robusta e referência remota e o procedimento de entrada mista que utiliza os resultados da regressão robusta e aqueles provenientes do método de reamostragem.

5.1 Resultado da transformada de Kramers-Kronig usando como entrada a regressão robusta

As funções de transferência obtidas usando o processamento robusto de Egbert e Booker (1986) (incluindo a opção de referência remota) apresentadas na Figura 5.1 foram usadas como parâmetros de entrada para cálculo das transformadas. Elas são mostradas em azul nas Figuras 5.2 e 5.3, enquanto as respostas geradas pelas transformadas $kk-ir$ e $kk-ri$ são mostradas em vermelho.

A Figura 5.2 mostra a comparação dos resultados obtidos pelo código de regressão robusta com aqueles derivados da transformada $kk-ir$ (Equação 4.3a). Como a entrada dessa transformada é a parte imaginária das componentes do tensor de impedâncias, a parte imaginária da transformada e da regressão robusta coincidem em todos os casos. Observa-se que a transformada corrige a distorção principal observada na parte real da componente Z_{xx} (entre aproximadamente 0,5 s e 20 s), mas não ajusta adequadamente a parte real não distorcida nos períodos mais longos. Conforme observado nos testes sintéticos, esse resultado se deve ao fato de a parte imaginária ainda não ter se estabilizado para o resultado de um semiespaço homogêneo nesses longos períodos. Em Z_{xy} , a transformada mostra valores mais baixos que aqueles obtidos na parte real pelo processamento robusto em praticamente todo o intervalo de períodos. Novamente, o ajuste não é satisfatório nos períodos mais longos pela resposta da parte imaginária não se aproximar daquela de um semiespaço homogêneo nesses períodos. Em Z_{yx} , a transformada mostra valores mais altos que os da parte real no intervalo de períodos mais curtos. Nesse mesmo intervalo, em torno de 0,1 s, as oscilações localizadas na parte imaginária dessa componente geram uma oscilação na parte real da transformada que não é observada nos dados do processamento robusto. Por outro lado, a transformada corrige a distorção da parte real no intervalo perturbado entre 0,5 s e 20 s, mas não ajusta adequadamente os resultados robustos nos períodos mais longos. Por fim, em Z_{yy} , as ondulações localizadas na parte imaginária dos resultados do processamento robusto são refletidas no cálculo da parte real da transformada, apenas que atenuados em relação àqueles observados na parte real dos resultados robustos.

Figura 5.2 – Transformada de Kramers-Kronig para a estação *apg023a* usando a parte imaginária como entrada. As funções de transferência em azul foram obtidas por meio de regressão robusta com referência remota enquanto aquelas em vermelho foram obtidos usando a Equação 4.3a.

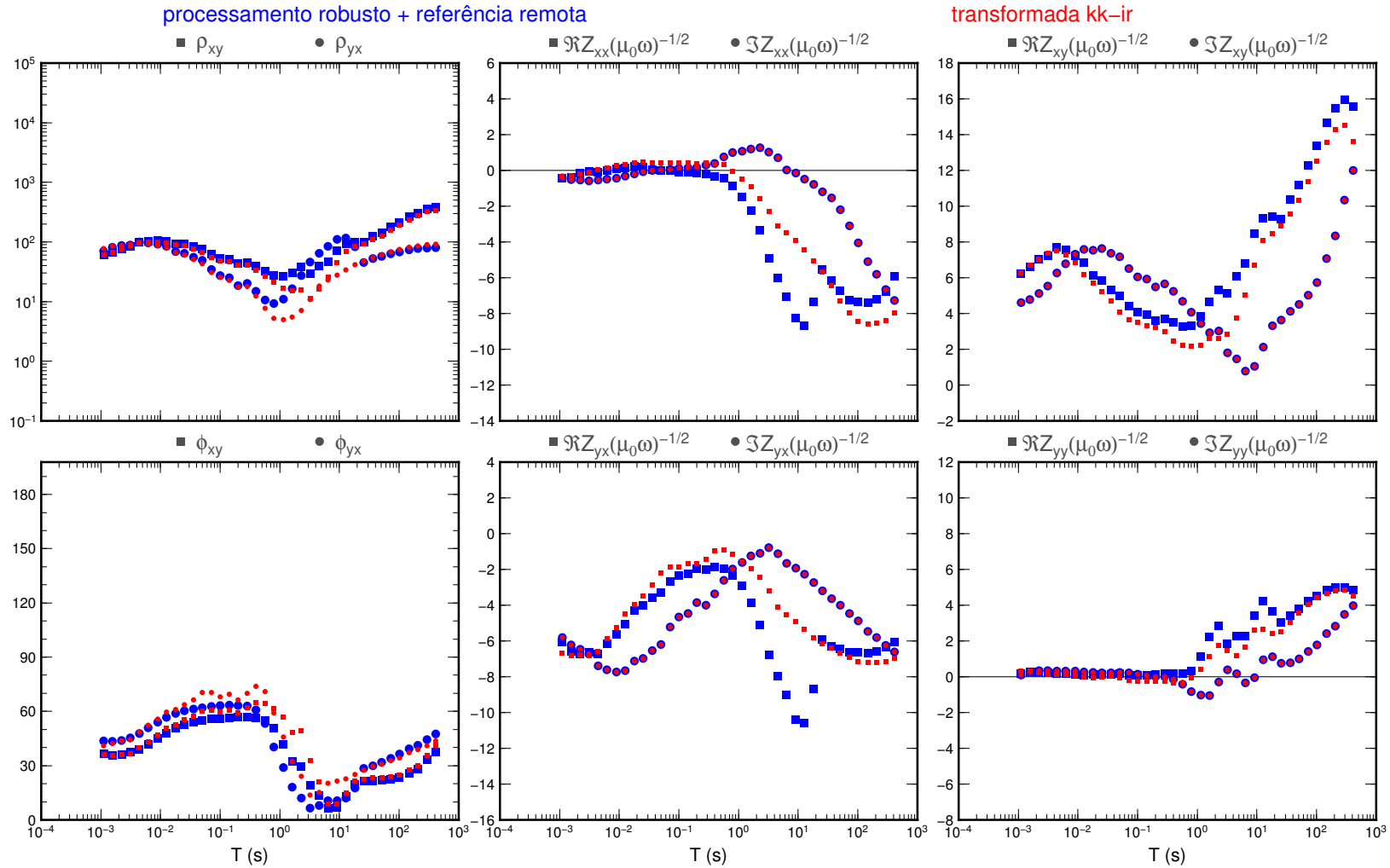
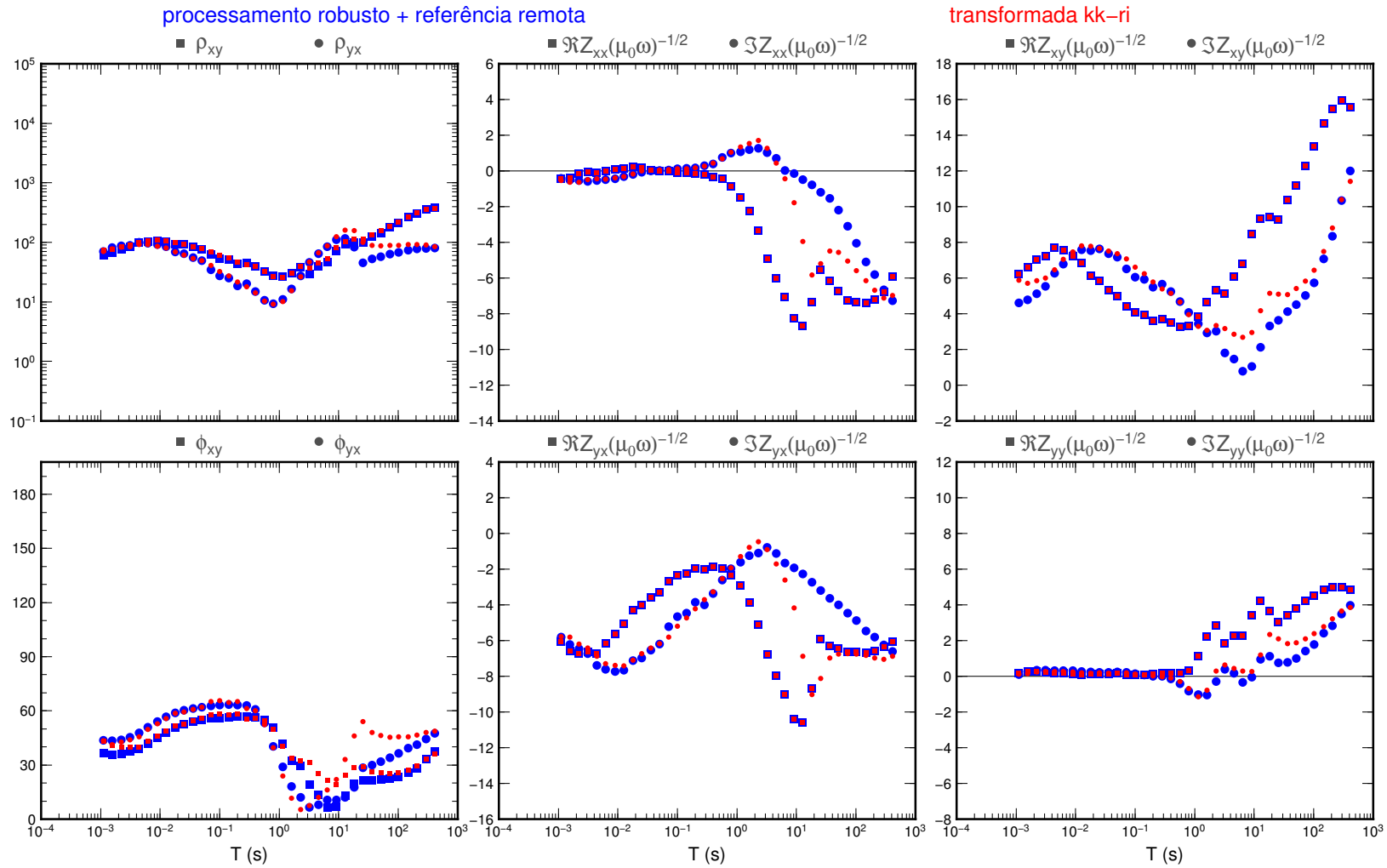


Figura 5.3 – Transformada de Kramers-Kronig para a estação *apg023a* usando a parte real como entrada. As funções de transferência em azul foram obtidas por meio de regressão robusta com referência remota enquanto aquelas em vermelho foram obtidos usando a Equação 4.3b.



Na análise dos resultados de resistividade aparente dessa mesma Figura 5.2 observa-se que a transformada *kk-ir* corrige a distorção da componente ρ_{yx} no intervalo entre 0,5 s e 20 s. Também, ainda em relação aos resultados gerados pelo processamento robusto, a transformada fornece valores mais baixos para essa componente nos períodos em torno de 1 s. Já para a resistividade aparente ρ_{xy} , a transformada fornece valores mais baixos de resistividade que os do processamento robusto em todo o intervalo entre 0,5 s e 10 s, mas é distorcida em períodos entre aproximadamente 3 s e 8 s.

As fases da Figura 5.2 mostram que a transformada *kk-ir* também é eficiente para corrigir a fase ϕ_{yx} no intervalo de períodos entre 2 s e 20 s. Há continuidade dessa fase com a obtida nos períodos mais longos e a fase mínima nesse intervalo fica em torno de 20° . Por outro lado, essa fase fica muito perturbada no intervalo de períodos entre aproximadamente 0,02 s e 1 s devido às oscilações anteriormente citadas, geradas na transformada da parte real da componente Z_{yx} . Quanto à fase ϕ_{xy} , a transformada *kk-ir* não consegue corrigir a distorção entre 3 s e 20 s devido ao efeito do ruído na parte imaginária da componente Z_{xy} , usada como entrada para a transformada. Em períodos mais curtos, essa fase apresenta valores superiores àqueles derivados do processamento robusto.

A Figura 5.3 mostra a comparação dos resultados do código de regressão robusta com os obtidos com a transformada *kk-ri* da Kramers-Kronig (Equação 4.3b). Em geral, essa transformada ajusta melhor os resultados da regressão robusta em períodos mais curtos do que a transformada *kk-ir*. Porém, no intervalo mais perturbado entre 0,2 s e 20 s, os resultados para as componentes Z_{xx} e Z_{yx} são muito diferentes daqueles do processamento robusto. Isso se deve ao forte ruído observado na parte real dessas componentes, a qual é usada como entrada pela transformada. Também, as oscilações observadas em períodos em torno de 10 s na parte real das componentes Z_{xy} e Z_{yy} causam grandes oscilações na parte imaginária gerada pela transformada. Como consequência, as resistividades aparentes e fases reproduzem de forma adequada os valores gerados pelo processamento robusto em períodos mais curtos, mas não é corrigida a quebra na resistividade aparente ρ_{yx} em períodos entre 2 s e 20 s e as distorções na fase são mais acentuadas nos períodos longos.

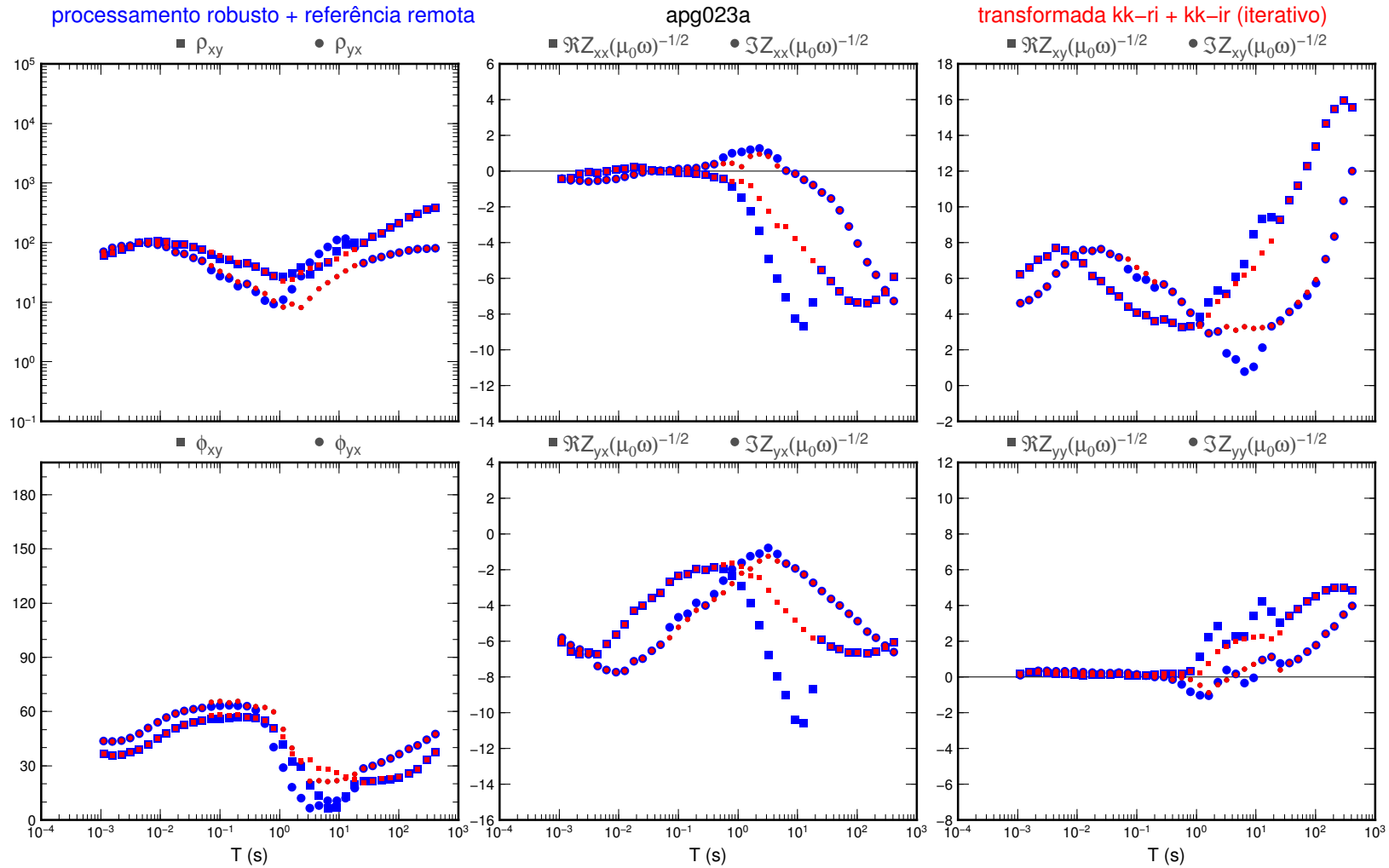
De forma sumarizada, esses resultados indicam que o procedimento proposto tem potencial para corrigir resultados anômalos nas funções de transferência MT quando esses se concentram em apenas uma das partes (real ou imaginária) das componentes do tensor de impedâncias. Esse é o caso das distorções na parte

real das componentes Z_{xx} e Z_{yx} no intervalo aproximado entre 2 s e 20 s, o que é confirmado pelos resultados suaves e contínuos de resistividade aparente e fase relacionados a Z_{yx} obtidos pela transformada *kk-ir* nesse intervalo de períodos (Figura 5.2). Por outro lado, os resultados mostram que a aplicação da transformada é desnecessária nos intervalos de períodos em que as funções de transferência mostram-se suaves (sem saltos abruptos) ou com pequenas oscilações em apenas uma das partes real ou imaginária. Essas pequenas oscilações nos elementos do tensor não produzem alterações significativas no cálculo de resistividades aparentes ou fases. Porém, deve-se tomar cuidado ao usar a transformada quando esses ruídos estão localizados na parte usada como entrada na transformada. Mesmo que de pequena monta, esses ruídos na componente de entrada acabam sendo transferidos também para a componente de saída da transformada. Com isso, vai se ter resultados anômalos, especialmente quando se calcula a fase do elemento tensorial, conforme patente na fase ϕ_{yx} da mesma Figura 5.2 em períodos em torno de 0,1 s.

Uma alternativa para minimizar efeitos de ruídos localizados em uma das transformadas (parte real ou imaginária) mas que não aparecem, ou são mais sutis, no mesmo intervalo de períodos da outra transformada é utilizar um procedimento iterativo. As transformadas podem ser aplicadas apenas nos períodos perturbados para corrigir o efeito de ruído localizado na outra parte do sinal e o resultado corrigido usado para corrigir o sinal inicial de entrada. Um aspecto desse procedimento iterativo é que ele depende do fator humano, pois quem vai escolher o intervalo a ser corrigido e o que está correto nas funções de transferência é o operador do programa. De qualquer forma, procedimento iterativo é comum em técnicas de processamento geofísico, sendo usado por exemplo no programa *Rho+* de validação entre resistividade aparente e fase (PARKER; BOOKER, 1996).

Um procedimento iterativo foi feito usando os resultados das transformadas *kk-ir* e *kk-ri* mostrados nas Figuras 5.2 e 5.3 e utilizando uma transformada para corrigir os resultados da outra, e vice-versa. O resultado final do procedimento é apresentado na Figura 5.4, a qual mostra uma significativa melhora nas funções de transferência, mais claramente observada nos resultados de resistividade aparente e fase. Esse resultado não foi obtido por qualquer processo de suavização, mas simplesmente aplicando de forma iterativa as transformadas *kk-ir* e *kk-ri* que têm base conceitual física sólida e bem definida. Novamente, os resultados evidenciam o potencial da transformada para minimizar ruídos nos dados MT.

Figura 5.4 – Transformada de Kramers-Kronig para a estação *apg023a* usando um procedimento iterativo. As funções de transferência em azul foram obtidas por meio de regressão robusta com referência remota enquanto aquelas em vermelho foram obtidas iterativamente usando as Equações 4.3a e 4.3b.



5.2 Resultado da transformada de Kramers-Kronig usando uma entrada mista de regressão robusta e reamostragem

Nesta seção se descreve um procedimento alternativo para buscar nos dados coletados em campo valores para a impedância que satisfaçam a transformada de Kramers-Kronig nos períodos em que a regressão robusta não apresenta um resultado satisfatório. Se a Seção 5.1 pode ser considerada como uma proposta de substituição do programa Rho^+ de Parker e Booker (1996) pela transformada $kk-ir$ ou $kk-ri$ aqui proposta, o primeiro válido apenas para situações específicas 1D e 2D e o segundo válido em situações 3D gerais, esta seção apresenta uma tentativa mais ampla de tentar substituir completamente programas de regressão robusta, tais como o de Egbert e Booker (1986) e que inclua a opção de referência remota de Gamble et al. (1979), nos períodos em que esses programas não forneçam respostas consistentes com avaliações prévias da transformada de Kramers-Kronig.

O procedimento adotado é realizado nas seguintes etapas:

Regressão Robusta Um conjunto inicial de impedâncias é obtido pelos métodos tradicionais de regressão robusta. Pode-se utilizar tanto a alternativa com apenas uma estação ou com a opção de referência remota (EGBERT; BOOKER, 1986; GAMBLE et al., 1979).

Verificação da Consistência das Funções de Transferência MT Verifica-se para quais períodos a regressão robusta apresenta resultados consistentes com a transformada de Kramers-Kronig (conforme discutido na Seção 5.1 e mostrado nas Figuras 5.2 e 5.3). Os períodos consistentes são considerados robustos e não necessitam de processamento adicional, os demais períodos serão considerados anômalos e selecionados para o procedimento de reamostragem.

Análise Espectral Envolve o procedimento padrão de transformar os dados coletados no domínio do tempo para o domínio da frequência (Seção 5.2.1), visto serem as impedâncias calculadas no domínio de frequência.

Reamostragem É a parte mais complexa do procedimento e envolve a tentativa de encontrar estimativas mais adequadas para as impedâncias a partir dos dados de campo. Nessa etapa nenhum valor é descartado, o que inclui os valores que seriam considerados *outliers* pelos algoritmos de regressão robusta. O procedimento utilizado é descrito em detalhes na Seção 5.2.2.

Redução de Dados A reamostragem fornece um número de amostras muito maior do que o que se pode tratar com os recursos computacionais atualmente disponíveis. Para reduzir o número de amostras utilizou-se um processo de redução de dados por aglomeração, descrito na Seção 5.2.3.

Construção de $Z(\omega)$ A Seção 5.2.4 descreve como construir as funções de transferência MT a partir dos aglomerados.

Aplicação da transformada de Kramers-Kronig A Seção 5.2.5 explica como a transformada de Kramers-Kronig é aplicada sobre $Z(\omega)$, de modo semelhante ao já feito no Capítulo 4.

Escolha da melhor estimativa de $Z(\omega)$ Finalmente, a Seção 5.2.6 mostra como montar uma função de resíduos para selecionar a função $Z(\omega)$ mais consistente com a transformada de Kramers-Kronig.

5.2.1 Análise espectral

Para obter as componentes do tensor de impedâncias Z , a relação $\mathbf{E} = \mathbf{Z}\mathbf{B}$ (Equação 2.28) é separada em duas equações lineares:

$$E_x = Z_{xx}B_x + Z_{xy}B_y \quad (5.1a)$$

$$E_y = Z_{yx}B_x + Z_{yy}B_y \quad (5.1b)$$

Para obter os valores das componentes utilizando algum método de regressão linear, seja por mínimos quadrados ou qualquer outro método mais robusto, é necessário obter um conjunto de valores para cada componente de \mathbf{E} e \mathbf{B} . Por essa razão as séries temporais coletadas em campo são divididas em segmentos.

Além disso, a relação $\mathbf{E} = \mathbf{Z}\mathbf{B}$ é definida no domínio da frequência, mas os valores de \mathbf{E} e \mathbf{B} são coletados no domínio do tempo. Por essa razão, depois de dividir a série temporal em segmentos, deve-se transformar as séries do domínio do tempo para o domínio da frequência. Com esses valores obtêm-se vários espectros das componentes do campo EM. Esses espectros são obtidos de uma maneira semelhante, mas não idêntica, ao procedimento adotado por Egbert e Booker (1986).

- Remoção de tendência linear.

- Aplicação de um *taper* para diminuir o vazamento das frequências com maior densidade espectral. O *taper* utilizado será a sequência DPSS (*Discrete Prolate Spheroidal Sequences*).
- Aplicação da transformada de Fourier discreta.

5.2.2 Reamostragem

Para obter funções de transferência MT que satisfaçam as transformadas de Kramers-Kronig (Equação 4.1) é necessário estabelecer um meio de obter respostas para as partes reais e imaginárias de \mathbf{Z} a partir dos espectros obtidos na sessão anterior. Em seguida, deve-se encontrar uma forma de selecionar qual dessas respostas é a mais consistente com a transformada de Kramers-Kronig.

Para obter as funções de transferência MT, este trabalho testou estimar os elementos do tensor magnetotelúrico \mathbf{Z} utilizando um método de reamostragem. Métodos de reamostragens são utilizados em estimadores estatísticos não paramétricos, tais como o *jackknife* e o *bootstrap* (EFRON, 1979), ou *reweighted least squares* (RLS) e *least median of squares regression* (LMS) (ROUSSEEUW, 1984).

Conforme já descrito anteriormente, durante a análise das séries temporais as medidas do sinal natural são divididas em N_s segmentos. Escolhem-se dois desses segmentos de índices ($1 \leq \varsigma_1 < \varsigma_2 \leq N_s$) a fim de obter uma estimativa para os elementos do tensor a partir dos espectros na seguinte forma:

$$Z_{xx}^{(\varsigma_1, \varsigma_2)} = \frac{[B_y B_y^*][B_x^* E_x] - [B_x^* B_y][B_y^* E_x]}{[B_x B_x^*][B_y B_y^*] - [B_x B_y^*][B_x^* B_y]}$$

$$Z_{xy}^{(\varsigma_1, \varsigma_2)} = \frac{[B_x B_x^*][B_y^* E_x] - [B_x B_y^*][B_x^* E_x]}{[B_x B_x^*][B_y B_y^*] - [B_x B_y^*][B_x^* B_y]}$$

$$Z_{yx}^{(\varsigma_1, \varsigma_2)} = \frac{[B_y B_y^*][B_x^* E_y] - [B_x^* B_y][B_y^* E_y]}{[B_x B_x^*][B_y B_y^*] - [B_x B_y^*][B_x^* B_y]}$$

$$Z_{yy}^{(\varsigma_1, \varsigma_2)} = \frac{[B_x B_x^*][B_y^* E_y] - [B_x B_y^*][B_x^* E_y]}{[B_x B_x^*][B_y B_y^*] - [B_x B_y^*][B_x^* B_y]}$$

O índice $(\varsigma_1, \varsigma_2)$ tem a função de indicar os segmentos utilizados para calcular a amostra, e os colchetes representam a soma dos espectros desses segmentos da série temporal. Por exemplo, para os segmentos ς_1 e ς_2 , temos $[B_x B_x^*] = (B_x B_x^*)_{\varsigma_1} + (B_x B_x^*)_{\varsigma_2}$. Essas soluções equivalem à solução das Equações lineares (5.1) e às

estimativas de mínimos quadrados (Equação 2.35), mas aqui usando apenas os segmentos ς_1 e ς_2 .

A Figura 5.5 mostra a distribuição das amostras da componente Z_{yx} para a mesma estação analisada na Figura 5.2. Neste caso, foram escolhidos apenas 7 períodos entre 2 s e 20 s por serem os que apresentaram maior discrepância entre os valores obtidos pela regressão robusta e pela transformada *kk-ir*. É mostrada na figura a dispersão dos valores da impedância (parte real no eixo das abscissas e parte imaginária no eixo das ordenadas) considerando diversas janelas de dados. Pode-se observar na figura que à medida em que diminui o número de elementos n de cada segmento considerado na série temporal diminui, aumenta o número de segmentos N_s . Consequentemente, também aumentam a quantidade de amostras geradas para Z e sua dispersão, podendo-se observar alguns *outliers* extremos. Esse comportamento é semelhante ao que ocorre nos métodos de regressão robusta, onde se perde resolução espectral com segmentos menores, mas ganha-se em números de pontos para a regressão, melhorando seus resultados.

A dispersão das amostras é difícil de ser visualizada na Figura 5.5, pois ela foi gerada para se mostrar também a presença dos *outliers* mais distantes. Para facilitar a visualização é mostrada na Figura 5.6 uma expansão da distribuição das amostras para o período 2,2857 s e série temporal com $n = 128$ elementos. Nesse caso, as escalas dos eixos real e imaginário foram reduzidas para facilitar a visualização dos hexágonos nos quais as amostras de Z_{yx} estão distribuídas. A figura mostra também a localização tanto do resultado de impedância (partes real e imaginária) obtido através do método tradicional de procedimento robusto com a opção de referência remota (EGBERT; BOOKER, 1986) como aquele derivado do procedimento iterativo da transformada de Kramers-Kronig descrito anteriormente e mostrado na Figura 5.4 (considerado o melhor resultado para esse período).

Figura 5.5 – Distribuição das amostras da componente Z_{yx} para vários períodos. A linha superior contém o período de cada amostra e a coluna da direita contém o número de elementos de cada segmento utilizado para calcular o espectro. O eixo horizontal de cada caixa representa a parte real das impedâncias e o vertical a parte imaginária. A cor de cada hexágono é determinada pelo número de impedâncias contida no hexágono, conforme indicado pela contagem no lado direito da figura.

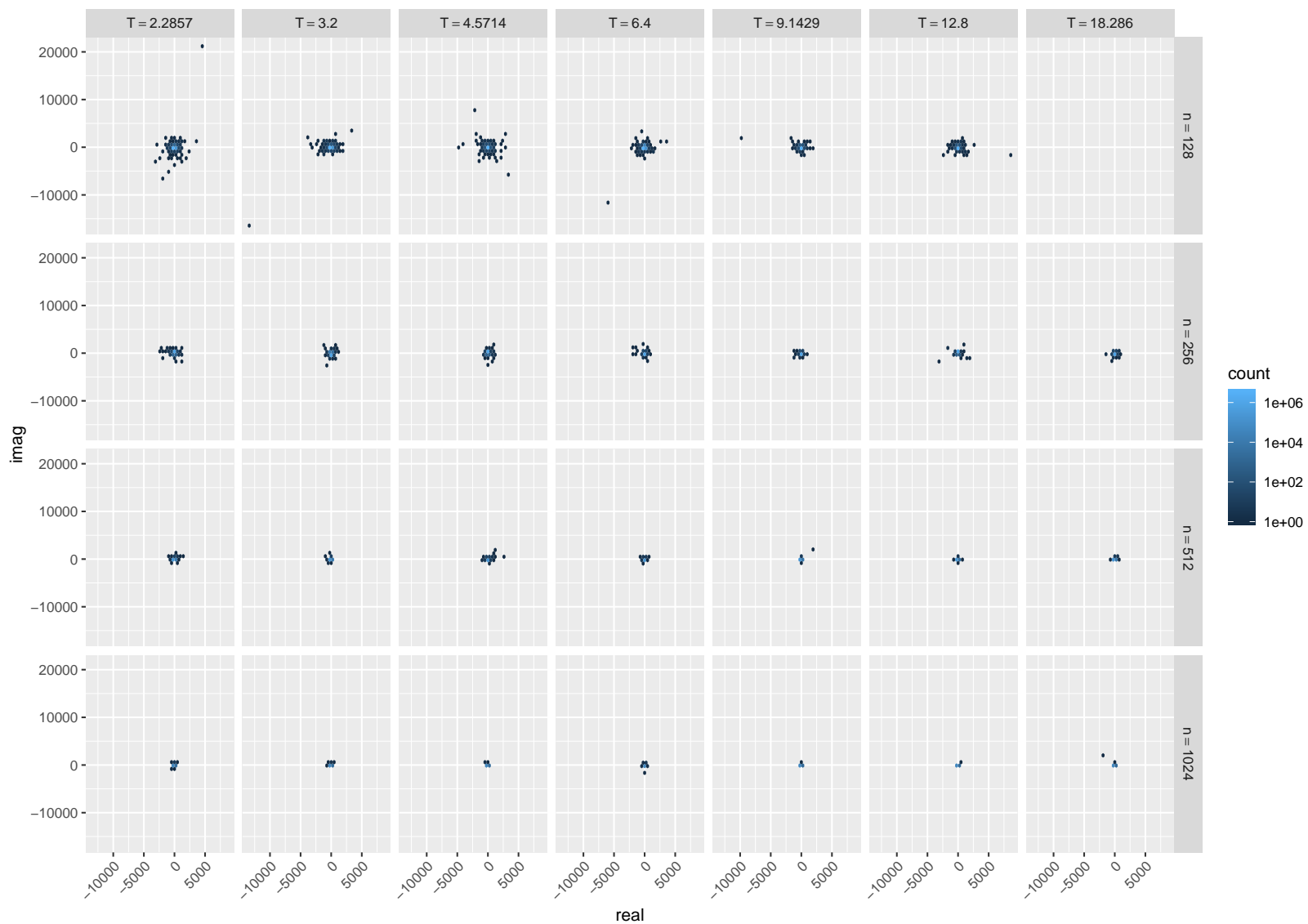
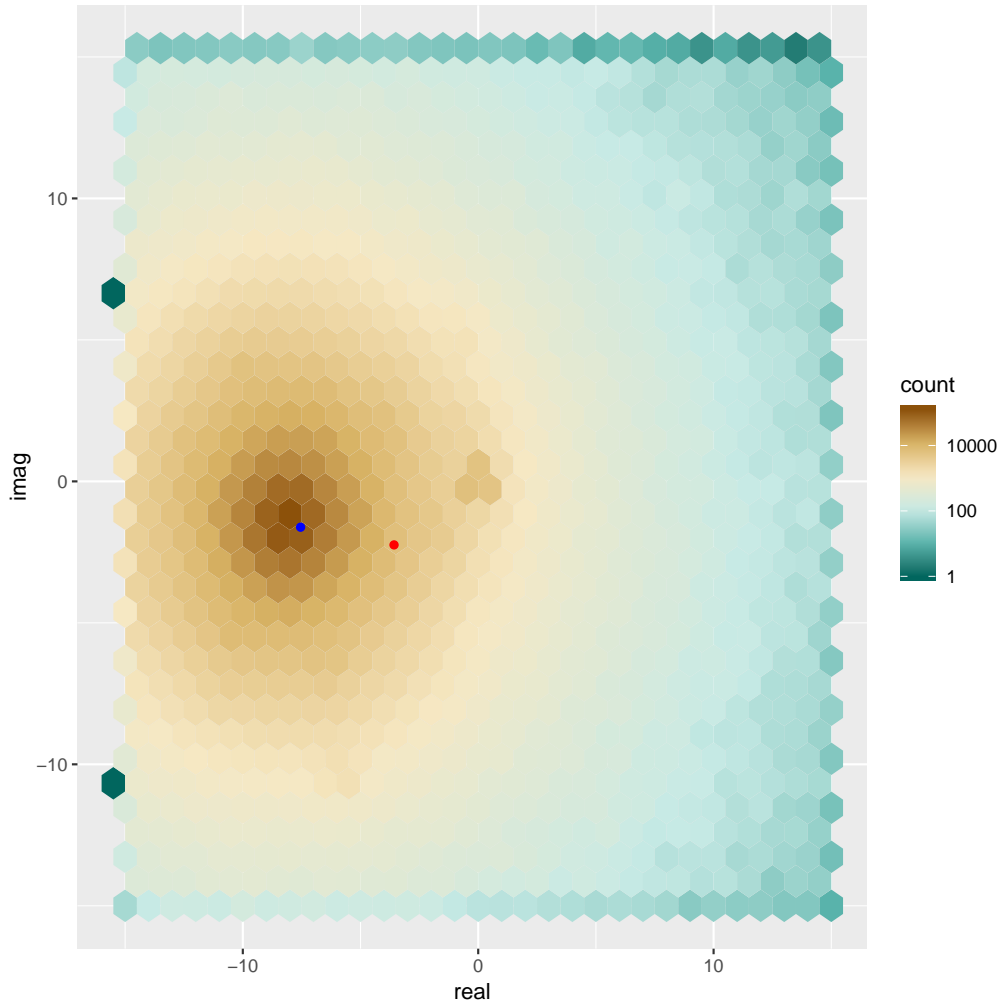


Figura 5.6 – Distribuição das amostras da componente Z_{yx} em escala reduzida para o período 2,2857 s e $n = 128$ da Figura 5.5. O ponto azul corresponde ao valor obtido através da regressão robusta com referência remota (EGBERT; BOOKER, 1986), enquanto o ponto vermelho corresponde ao resultado do procedimento iterativo das transformadas de Kramers-Kronig mostrado na Figura 5.4.



5.2.3 Redução de dados por meio de aglomeração

Idealmente, poder-se-ia utilizar os pontos da Figura 5.5 para construir várias funções para as componentes de $Z(\omega)$ e verificar qual delas seria a mais consistente com a transformada de Kramers-Kronig. No entanto, deve-se observar que, para alguns períodos, o número de amostras chega a 10^6 elementos. Considerando que no exemplo apresentado são 7 períodos amostrados em cada elemento do tensor de impedâncias, o número de amostras a serem testadas seria impraticável considerando os recursos computacionais hoje disponíveis.

De modo geral, verifica-se que se as séries temporais forem divididas em N_s segmentos, o número de amostras N_a possíveis para cada frequência alvo vai ser

$$N_a = \frac{N_s(N_s - 1)}{2}$$

Dessa forma, para N_s grande, vai se ter $N_a \sim N_s^2$. Para N_f frequências alvo escolhidas, o número de respostas possíveis vai ser $N = N_s^{2N_f}$. Como o número de frequências alvo é importante para a qualidade das curvas, é preferível diminuir o número de amostras escolhidas. Para reduzir o número de amostras vai ser aqui utilizada a análise de aglomerados.

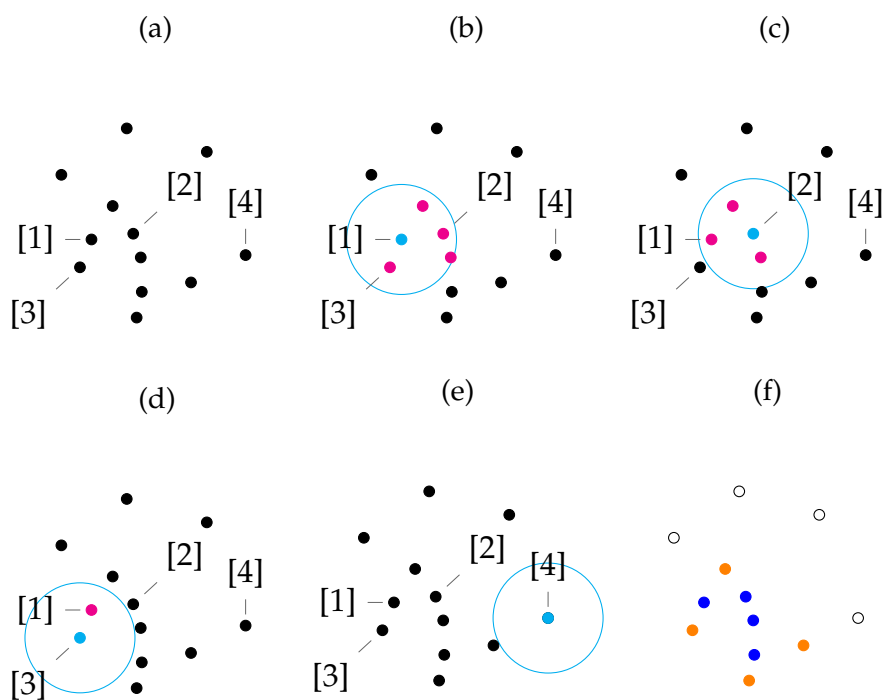
O objetivo da análise de aglomerados é descobrir os aglomerados naturais de um conjunto de padrões, pontos ou objetos (JAIN, 2010). A análise de aglomerados será aqui utilizada para substituir as amostras contidas em um aglomerado pelos centros desses aglomerados.

Para a análise de aglomerados foi utilizado o método *dbscan* descrito em (ESTER et al., 1996) e ilustrado na Figura 5.7. O algoritmo *dbscan* utiliza dois parâmetros: um raio ϵ e um número mínimo de pontos (*pts*). A partir desses parâmetros é calculado o número de pontos n contido no círculo de raio ϵ centrado em cada ponto do conjunto. Em seguida cada ponto é classificado da seguinte maneira:

- se $n \geq pts$, o ponto é considerado um núcleo do aglomerado.
- se um dos pontos contidos dentro do círculo de um núcleo também for um núcleo, esses dois núcleos pertencem ao mesmo aglomerado.
- se um dos pontos contidos dentro do círculo de um núcleo não for um núcleo, esse ponto é considerado borda do aglomerado daquele núcleo.
- pontos com $n < pts$ e que não são bordas de nenhum aglomerado são considerados ruído.

A Figura 5.7 ilustra o processo de classificação do algoritmo *dbscan*. Na Figura 5.7a é apresentado um conjunto inicial de pontos que devem ser agrupados em aglomerados com $\epsilon = 1$ e $pts = 4$. Na Figura 5.7b observa-se que dentro da circunferência de raio $\epsilon = 1$ do ponto [1] existem 5 pontos (indicados em cores). Como $5 > pts$, o ponto [1] é classificado como um dos núcleos do aglomerado. Da mesma forma, a Figura 5.7c mostra que o ponto [2] também é um núcleo, pois tem 4 pontos dentro

Figura 5.7 – Resumo do método *dbscan*. As cores dos pontos são definidas como: ● são os pontos originais do conjunto; ● é o ponto que está sendo classificado; ● são os pontos dentro do alcance do ponto a ser classificado; ● são pontos classificados como núcleo de um aglomerado; ● são pontos classificados como borda de um aglomerado; ○ são pontos classificados como ruído.



do círculo. Além disso, como o ponto [1] está dentro do círculo do ponto [2], e vice-versa, os pontos [1] e [2] são núcleos do mesmo aglomerado. Já a Figura 5.7d mostra que dentro do círculo do ponto [3] existem apenas 2 pontos. Como $2 < pts$, o ponto [3] não é um núcleo. Por outro lado, a Figura 5.7b mostra que o ponto [3] está dentro do raio do núcleo [1], e por essa razão ele é um ponto de borda do aglomerado a que pertence o núcleo [1]. Finalmente, a Figura 5.7e mostra que o ponto [4] tem apenas 1 ponto dentro do seu círculo e que esse ponto não está dentro do círculo de nenhum núcleo. Por essas razões, ele é considerado um ruído e não faz parte de qualquer aglomerado. Após realizar as classificações de todos os pontos, mostra-se o resultado final na Figura 5.7f, onde temos um aglomerado com 4 núcleos (indicados por pontos ●) e 4 pontos de borda (pontos ●).

O algoritmo *dbscan* utiliza o raio ϵ como um valor absoluto. Para o processa-

mento MT, esse aspecto se mostra um problema porque os valores do tensor de impedâncias não são espalhados uniformemente para os diferentes períodos ou para o tamanho do segmento, conforme mostrado na Figura 5.5. Isso tornou inviáveis nossas tentativas de encontrar um valor adequado para ϵ . Por essa razão, a opção adotada foi calcular o valor de ϵ a partir do conjunto de valores de Z , utilizando a seguinte fórmula:

$$\epsilon = P(p, \text{kNN}\{k, Z\{T\}\}) \quad (5.2)$$

onde $Z\{T\}$ representa a distribuição para uma das componentes do tensor de impedâncias para um determinado período. Essa distribuição é obtida a partir dos espectros das séries temporais do sinal EM com n elementos, como as mostradas em cada quadro da Figura 5.5 para a componente Z_{yx} .

Para cada ponto da distribuição $Z\{T\}$ foi calculada a distância desse ponto para o seu k -ésimo vizinho mais próximo $\text{kNN}(Z, k)$ (o nome da função vem do termo em inglês para *k-Nearest Neighbor*). O conjunto de todas essas distâncias forma a distribuição $\text{kNN}\{Z\{T\}, k\}$. A Figura 5.8 ilustra a distribuição de $\text{kNN}\{Z\{T\}\}$ por meio de gráficos “*box plot*” (MCGILL et al., 1978) para a componente Z_{yx} da Figura 5.5 com período $T = 2,2857$ s. A variação dessas distribuições mostra como uma escolha de ϵ para uma distribuição pode ser inadequada para uma distribuição diferente.

Para obter o valor de ϵ foi calculado o p -ésimo percentil da distribuição $\text{kNN}\{k, Z\{T\}\}$. Os aglomerados *dbscan* obtidos para os conjuntos da Figura 5.5 estão ilustrados na Figura 5.9. Os aglomerados foram obtidos com no mínimo 79 pontos dentro do círculo de raio ϵ . O raio ϵ para cada conjunto foi obtido usando a mediana (percentil $p = 50$) dos primeiros vizinhos (kNN com $k = 1$).

Os conglomerados obtidos para cada um dos períodos e nos diferentes valores de n utilizados são representados na Figura 5.9 por domínios com diferentes cores. Observe-se que o procedimento não foi eficiente e em todos os casos definiu domínios exclusivamente na região em torno de impedâncias real e imaginária próximas a 0, onde se concentra o maior número de amostras de Z na Figura 5.5. Dessa forma, o resultado não convergiu para um domínio mais distante onde se esperava que estariam concentrados as impedâncias não contaminadas por ruído EM. Ou seja, ele converge para uma solução similar àquela obtida pelo procedimento robusto tradicional e não para o resultado ótimo derivado anteriormente na Figura 5.4. Pode-se argumentar que esse resultado do procedimento iterativo não representa

um conglomerado secundário bem definido no espaço de amostras, como pode ser observado na distribuição mostrada na Figura 5.6.

Figura 5.8 – Distribuição das distâncias dos primeiros vizinhos $kNN\{Z\}$ para vários comprimentos n de segmentos da série temporal dos dados EM. A distribuição foi calculada a partir do método *dbscan* sobre os conjuntos de Z_{yx} da Figura 5.5 com $T = 2,2857$ s. Os limites inferior e superior das caixas estabelecem os valores do primeiro ($Q1$) e terceiro ($Q3$) quartis da distribuição. A linha no interior das caixas representa o segundo quartil (mediana). As linhas que se estendem a partir da caixa são conhecidas como bigodes e podem ser interpretadas como limites para os *outliers* da distribuição. O valor do bigode superior é o maior ponto x da distribuição com $x \leq Q3 + 1.5(Q3 - Q1)$. O valor do bigode inferior é o menor ponto x da distribuição com $x \geq Q1 - 1.5(Q3 - Q1)$.

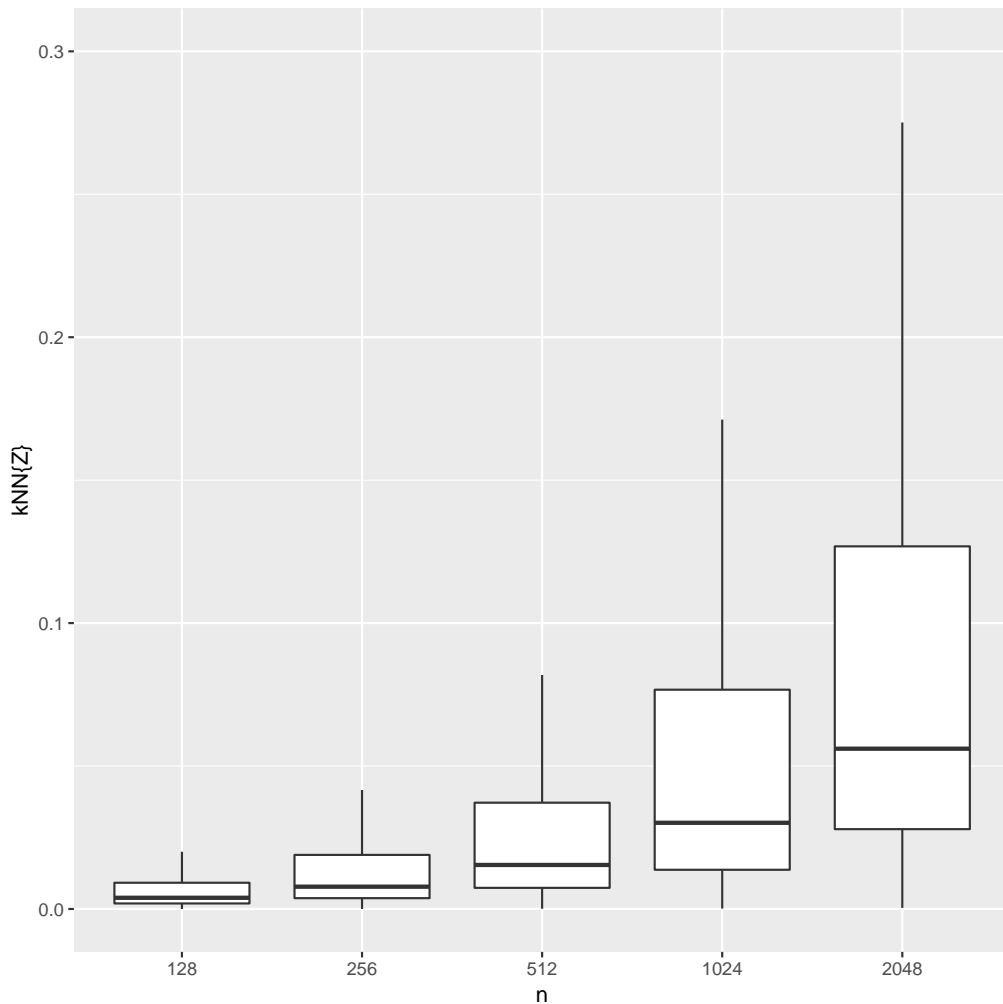
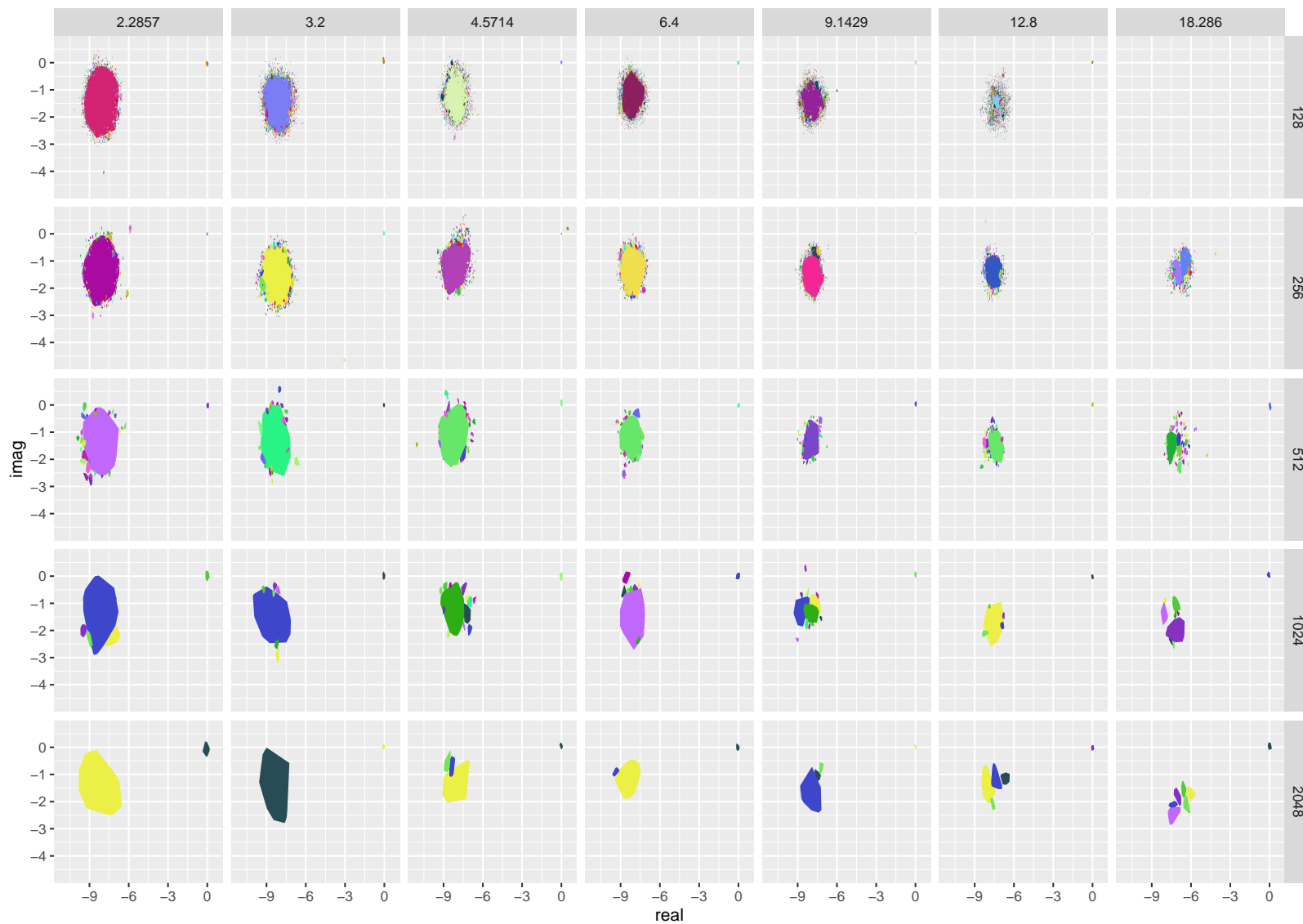


Figura 5.9 – Aglomerados obtidos a partir da utilização do método *dbscan* sobre as amostras da Figura 5.5. Cada aglomerado é representado por um polígono convexo de cor diferente.



5.2.4 Construção de $Z_{ij}(\omega)$ a partir dos aglomerados.

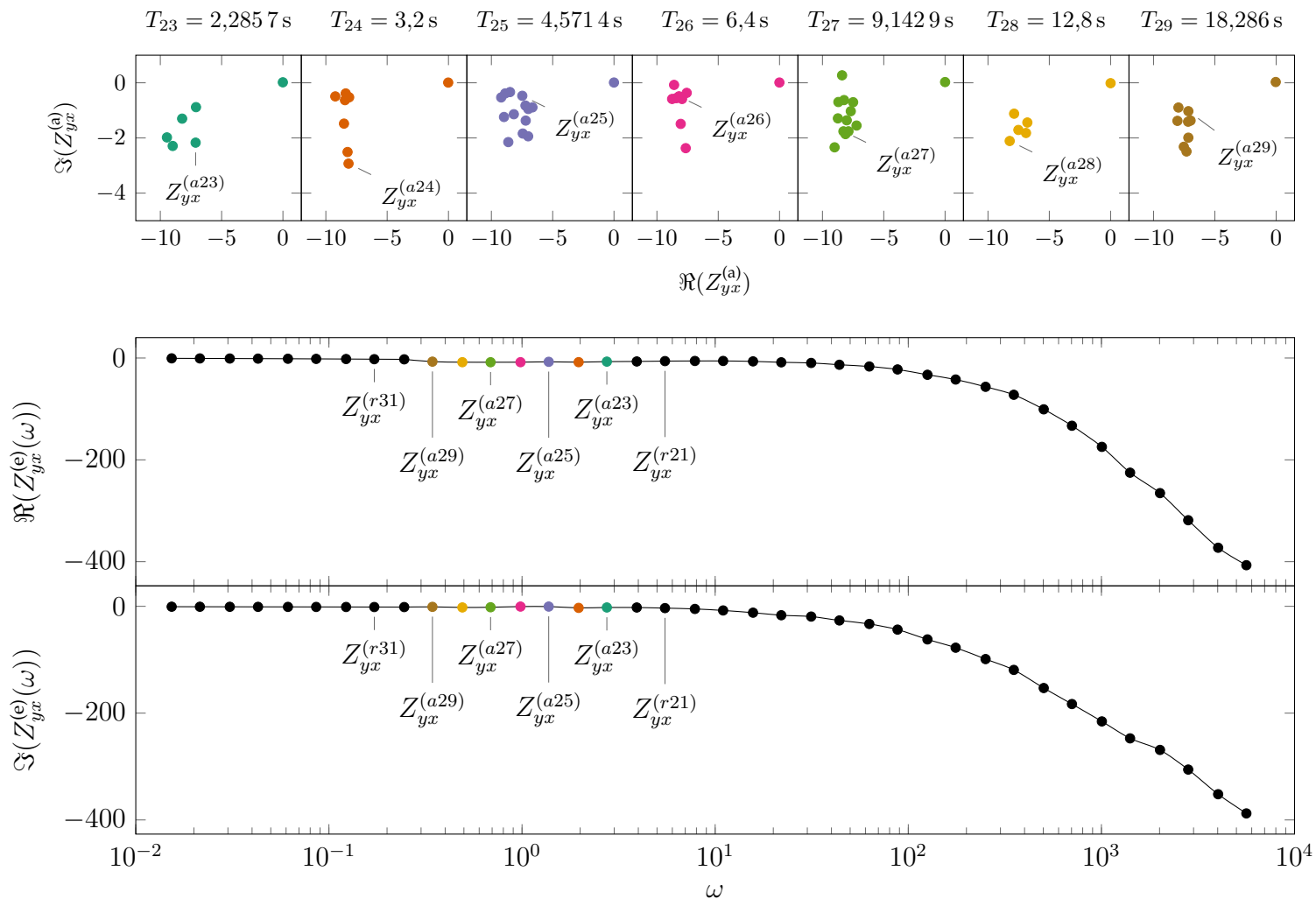
Conforme anteriormente discutido, para construir a função de transferência $Z_{ij}(\omega)$ da componente ij do tensor \mathbf{Z} , os diferentes períodos são divididos em duas categorias:

períodos robustos: aqueles nos quais a transformada de Kramers-Kronig forneceu resultado satisfatório com os próprios resultados da regressão robusta. Para esses períodos, a fonte de $Z_{ij}(\omega)$ são os valores obtidos por meio da regressão robusta e serão representados pelo vetor $\mathbf{Z}_{ij}^{(r)}$.

períodos anômalos: são aqueles em que a transformada não forneceu valores adequados. A fonte desses $Z_{ij}(\omega)$ para esses períodos são os aglomerados obtidos de acordo com o método descrito na Seção 5.2.3 e serão representados pelo vetor $\mathbf{Z}_{ij}^{(a)}$.

A Figura 5.10 ilustra de maneira esquemática como a amostragem é utilizada para montar uma curva de impedância $Z_{ij}(\omega)$. Para exemplificar o processo, escolheu-se a curva Z_{yx} da estação *apg023a* anteriormente discutida, na qual há 7 períodos anômalos no intervalo entre 2,2857 s – 18,286 s. Os 7 quadros no topo da figura representam os centroides dos aglomerados para esses períodos anômalos nos quais serão calculadas as funções de transferência (os indicadores 23 a 29 correspondem a cada um dos períodos a analisar dentro do espectro total de períodos amostrados). Para cada um desses períodos, foi escolhido aleatoriamente um dos centroides do vetor $\mathbf{Z}_{yx}^{(a)}$. A união dos vetores para os períodos robustos e anômalos selecionados fornece o vetor de estimativas $\mathbf{Z}_{yx}^{(e)} = \mathbf{Z}_{yx}^{(r)} \cup \mathbf{Z}_{yx}^{(a)}$ que será usado para montar a função $Z_{yx}(\omega)$. Considera-se aqui que $\mathbf{Z}_{yx}^{(e)}$ é o vetor de estimativas que representa os valores da impedância em todas as frequências ω , $\mathbf{Z}_{yx}^{(r)}$ são as impedâncias não ruidosas calculadas pelo programa robusto e $\mathbf{Z}_{yx}^{(a)}$ são as impedâncias anômalas que se pretende obter pela transformada de Kramers-Kronig. A partir do vetor de estimativas, são criados dois vetores contendo a parte real $\Re_{yx}^{(e)} = \Re(\mathbf{Z}_{yx}^{(e)})$ e a parte imaginária $\Im_{yx}^{(e)} = \Im(\mathbf{Z}_{yx}^{(e)})$, os quais estão representados como pontos nos gráficos central e inferior da figura. Nesses gráficos, os pontos coloridos correspondem aos valores anômalos, enquanto os pontos pretos correspondem àqueles obtidos pelo processamento robusto. Funções *splines* que interpolam essas amostras estão representados por meio das linhas contínuas.

Figura 5.10 – Processo de amostragem da componente Z_{yx} na estação *apg023a*. Os gráficos no topo da figura contém os centroides para os sete períodos anômalo. Para cada um desses períodos é escolhido um dos centroides, o qual será utilizado para obter o vetor Z_{yx}^a , exibido como pontos coloridos nos gráficos inferiores. Os pontos pretos nos gráficos central e inferior representam os pontos de períodos robustos $Z_{yx}^{(r)}$. As curvas contínuas são obtidas por meio de *splines* do vetor de estimativas $Z_{yx}^{(e)} = Z_{yx}^{(r)} \cup Z_{yx}^{(a)}$.



5.2.5 Estimativa da transformada de Kramers-Kronig

Uma vez estabelecidos os valores dos elementos do tensor de impedâncias em cada uma das frequências alvo, o passo seguinte consiste em estimar a transformada de Kramers-Kronig para verificar a correlação entre as partes reais e imaginárias desses elementos. A Figura 5.11 ilustra essa etapa com a utilização das transformadas *kk-ri* e *kk-ir* sobre as diferentes frequências do vetor $\mathbf{Z}_{ij}^{(e)}$. O procedimento é o mesmo daquele descrito na Seção 4.2.

5.2.6 Minimização dos resíduos

O procedimento discutido nas seções anteriores mostrou como obter os vetores da transformada *kk-ir* ($\mathfrak{R}_{ij}^{(kk)}$) e *kk-ri* ($\mathfrak{S}_{ij}^{(kk)}$) utilizando as estimativas obtidas da união das impedâncias de regressão robusta e aglomerados ($\mathbf{Z}_{ij}^{(e)} = \mathbf{Z}_{ij}^{(r)} \cup \mathbf{Z}_{ij}^{(a)}$). No caso dos aglomerados, essa estimativa é feita individualmente para cada um dos centroides da frequência alvo. O passo seguinte é determinar, entre todos os centroides possíveis (e consequentemente entre todos os vetores $\mathbf{Z}_{ij}^{(e)}$ possíveis), qual deles é o mais adequado para representar o elemento da impedância naquela frequência específica. O procedimento aqui sugerido é escolher o valor mais consistente com as transformadas de Kramers-Kronig.

Essa escolha pode ser feita a partir de uma função de resíduos entre as impedâncias estimadas e aquelas calculadas pela transformada, na forma $r(\mathbf{Z}_{ij}^{(kk)}, \mathbf{Z}_{ij}^{(e)})$. Diferentes funções podem ser adotadas para encontrar esses resíduos. Assim, se estivermos interessados apenas no resultado da transformada *kk-ri* pode-se adotar:

$$\hat{\mathbf{Z}}_{ij} = \min_{\mathbf{Z}_{ij}^{(e)}} \left| \mathfrak{S}_{ij}^{(kk)} - \mathfrak{S}_{ij}^{(e)} \right| \quad (5.3)$$

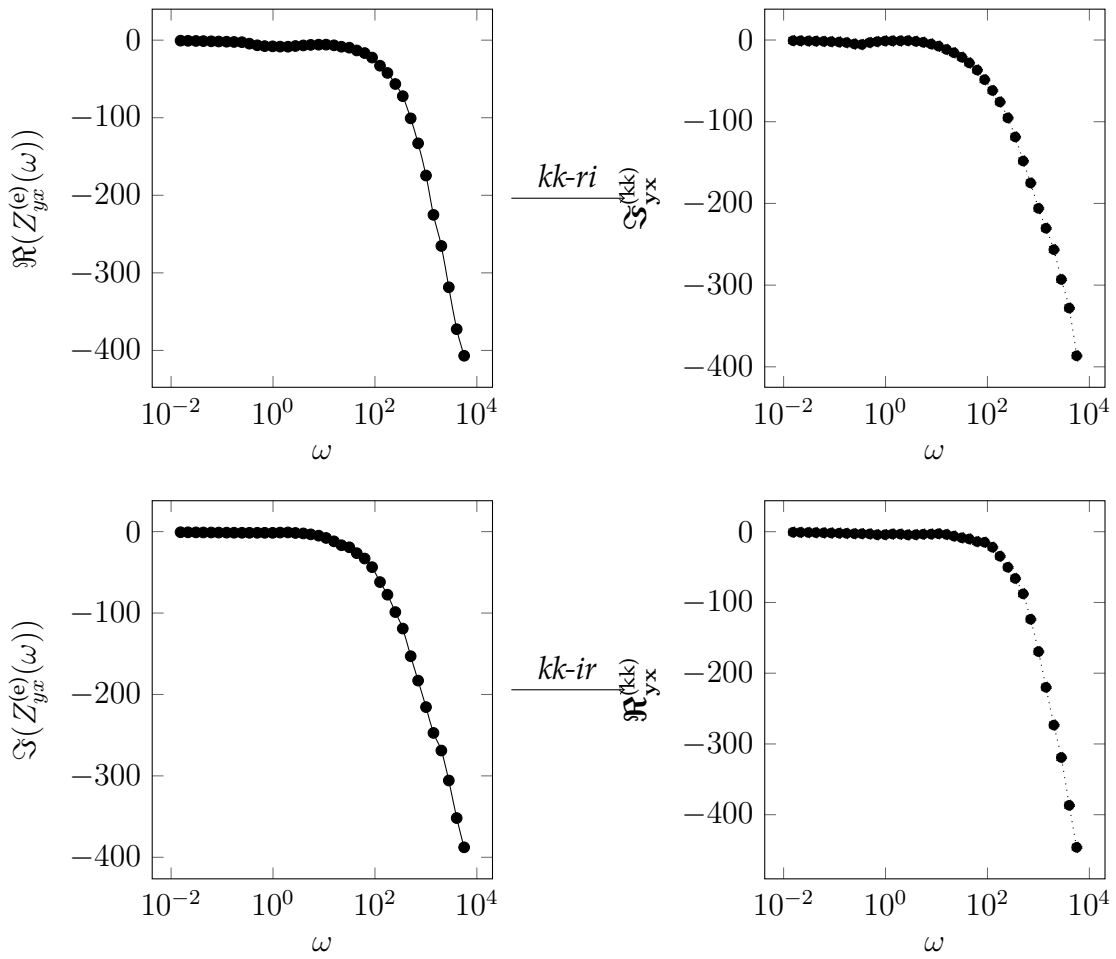
onde a estimativa $\hat{\mathbf{Z}}_{ij}$ dessa equação deve ser entendida como o valor de $\mathbf{Z}_{ij}^{(e)}$ que minimiza o módulo dos resíduos $r = \left| \mathfrak{S}_{ij}^{(kk)} - \mathfrak{S}_{ij}^{(e)} \right|$.

Por outro lado, se estivermos interessados apenas no resultado da transformada *kk-ir*, a expressão a adotar é:

$$\hat{\mathbf{Z}}_{ij} = \min_{\mathbf{Z}_{ij}^{(e)}} \left| \mathfrak{R}_{ij}^{(kk)} - \mathfrak{R}_{ij}^{(e)} \right| \quad (5.4)$$

Outras opções possíveis de minimização dos resíduos para escolha do valor representativo da impedância também podem ser utilizadas. Assim, por exemplo,

Figura 5.11 – Aplicação das transformadas $kk-ri$ e $kk-ir$ nos valores da impedância Z_{yx} da Figura 5.10. As funções $\Re(Z_{yx}^{(e)}(\omega))$ e $\Im(Z_{yx}^{(e)}(\omega))$ daquela figura estão representadas nos gráficos do lado esquerdo, enquanto o resultado das transformadas de Kramers-Kronig são apresentados no lado direito.



se escolhermos minimizar simultaneamente as partes real e imaginária de uma determinada frequência, pode-se utilizar:

$$\hat{Z}_{ij} = \min_{\Re_{ij}^{(e)}} \left| \Im_{ij}^{(kk)} - \Im_{ij}^{(e)} \right|^2 + i \min_{\Im_{ij}^{(e)}} \left| \Re_{ij}^{(kk)} - \Re_{ij}^{(e)} \right|^2 \quad (5.5)$$

Considerando-se simultaneamente apenas algumas das componentes do tensor (por exemplo, Z_{xx} e Z_{xy}), a minimização do resíduo pode ser dada por:

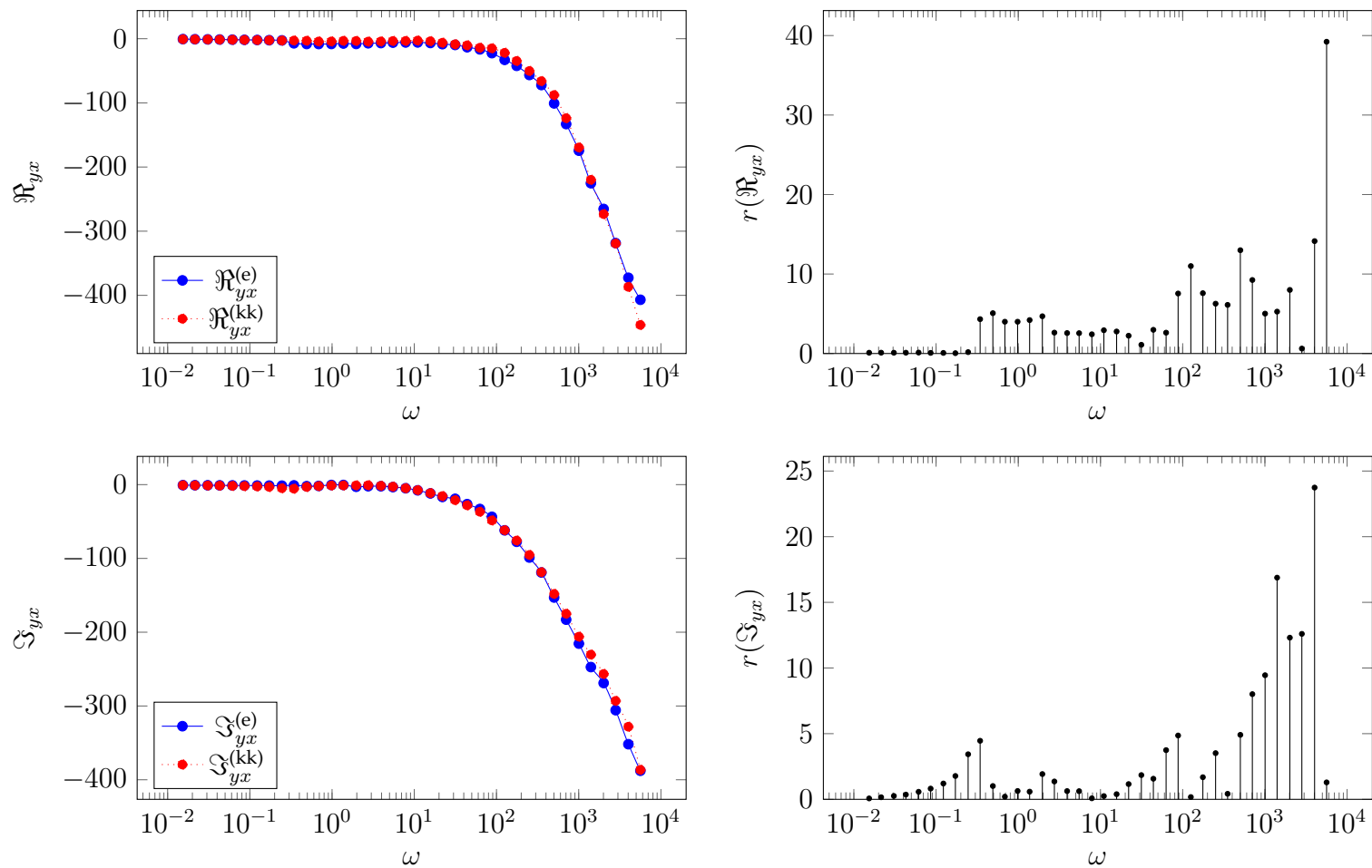
$$\left(\hat{Z}_{xx}, \hat{Z}_{xy} \right) = \min_{Z_{xx}^{(e)}, Z_{xy}^{(e)}} \left| Z_{xx}^{(kk)} - Z_{xx}^{(e)} \right|^2 + \left| Z_{xy}^{(kk)} - Z_{xy}^{(e)} \right|^2 \quad (5.6)$$

ou ainda

$$\begin{aligned} \left(\hat{Z}_{xx}, \hat{Z}_{xy} \right) = \min_{\Re_{xx}^{(e)}, \Re_{xy}^{(e)}} & \left(\left| \Im_{xx}^{(kk)} - \Im_{xx}^{(e)} \right|^2 + \left| \Im_{xy}^{(kk)} - \Im_{xy}^{(e)} \right|^2 \right) \\ & + i \min_{\Im_{xx}^{(e)}, \Im_{xy}^{(e)}} \left(\left| \Re_{xx}^{(kk)} - \Re_{xx}^{(e)} \right|^2 + \left| \Re_{xy}^{(kk)} - \Re_{xy}^{(e)} \right|^2 \right) \end{aligned} \quad (5.7)$$

Como exemplo, a Figura 5.12 mostra o resultado obtido para os resíduos utilizando as Equações 5.3 e 5.4 para os dados da impedância Z_{yx} mostrados na Figura 5.11.

Figura 5.12 – Cálculo dos resíduos para as componentes reais e imaginárias do elemento Z_{yx} da estação *apg023a*. Os gráficos a esquerda contém os resultados para as partes real e imaginária das amostras (em azul) e das transformadas de Kramers-Kronig (em vermelho), enquanto os gráficos a direita mostram os módulos dos resíduos entre esses valores.

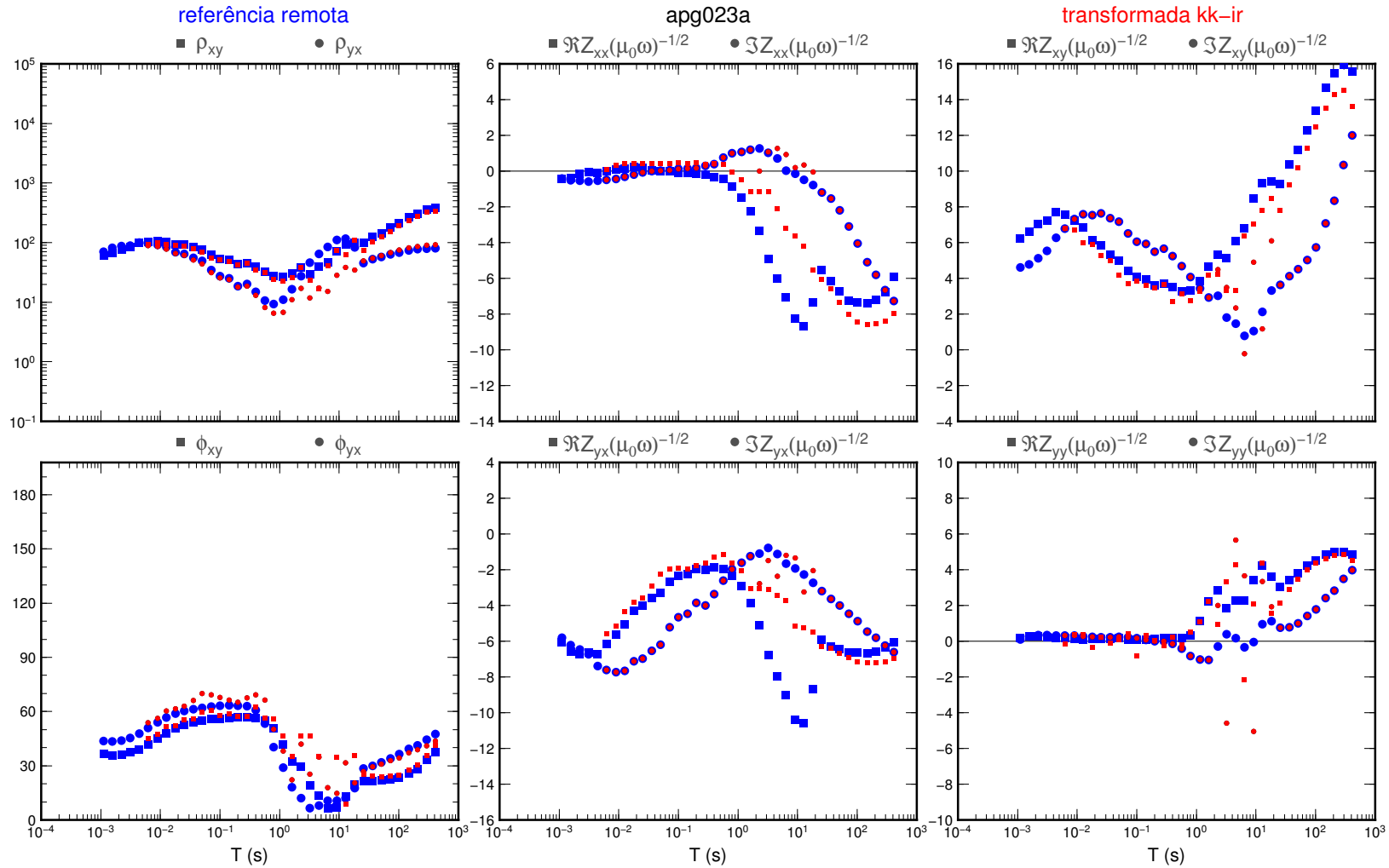


5.2.7 Resultado do uso misto de regressão robusta e reamostragem

A metodologia apresentada nas Seções 5.2.1 a 5.2.6 foi aplicada às séries temporais da estação *apg023a*, usando os resultados do processamento robusto nas frequências supostamente não contaminadas por ruído e a técnica de reamostragem nas 7 frequências perturbadas. O resultado da transformada *kk-ir* é apresentado na Figura 5.13. Esse resultado pode ser comparado diretamente com aquele obtido no procedimento em que se usam exclusivamente os resultados do processamento robusto como entrada para essa transformada, mostrado anteriormente na Figura 5.2.

Pode-se observar que o procedimento não foi bem sucedido para obter resultados adequados para as frequências contaminadas, mostrando valores dispersos para os elementos de impedância calculados nesse intervalo. Certamente, esse resultado deriva do fato de não se ter encontrado, no procedimento adotado de busca, a região onde se concentrariam as amostras não contaminadas pelo ruído. Conforme mostrado na Figura 5.6, o resultado ideal obtido pelo procedimento iterativo das transformadas de Kramers-Kronig (ponto vermelho nessa figura) não caracteriza uma concentração significativa de amostras. Isso seria esperado caso o resultado derivado do sinal natural (ondas planas) ficasse concentrado em torno de um resultado ótimo e gerasse um aglomerado bem definido de valores (distribuição gaussiana). Nesse caso, porém, qualquer possível concentração de pontos oriundos do sinal natural fica obscurecido pela distribuição mais espalhada e com grande número de amostras do ruído artificial.

Figura 5.13 – Transformada de Kramers-Kronig para a estação *apg023a* usando a parte imaginária como entrada e a técnica mista de regressão robusta e reamostragem. As funções de transferência em azul foram obtidas por meio de regressão robusta enquanto aquelas em vermelho foram obtidas usando o método de reamostragem e a transformada *kk-ir* dada pela Equação 4.3a.



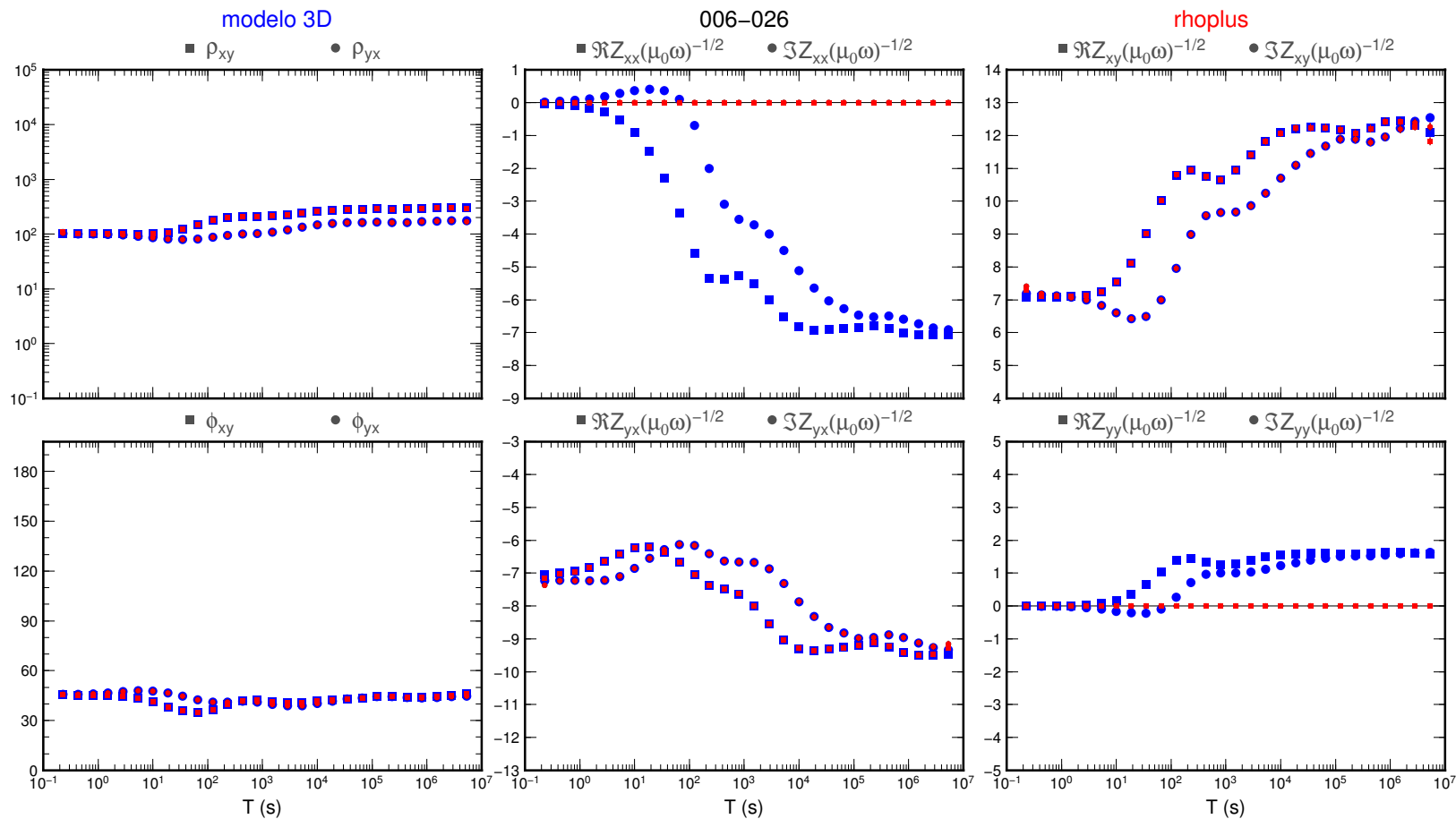
6 DISCUSSÕES SOBRE O USO PRÁTICO DO ALGORITMO DESENVOLVIDO

Os resultados apresentados no capítulo anterior evidenciam o potencial da transformada de Kramers-Kronig no apoio à obtenção de funções de transferência de melhor qualidade para uso pelo método MT em regiões sujeitas a intensos ruídos EM. Particularmente, a transformada pode ser bastante útil em situações em que falham as técnicas atualmente utilizadas de processamento de dados, envolvendo algoritmos robustos e referência remota. No estágio atual de implementação dessa transformada, conforme descrito nesta tese, a opção que usa como parâmetros de entrada a informação diretamente derivada de processamentos prévios de regressão robusta já se encontra operacional. A segunda opção, que consiste em um formalismo totalmente independente de obtenção das funções de transferência MT, utilizando o processo de *reamostragem* e que poderia ser aplicado a situações de ruídos extremos (ponto de ruptura superior a 50%) não obteve sucesso, valendo mais como conceito no presente estágio. Na sequência são discutidas situações em que o algoritmo desenvolvido para implantação da transformada poderia ser utilizado em substituição ou apoio aos procedimentos atualmente em utilização para auxílio no processamento de dados MT.

6.1 Relação com o algoritmo Rho^+ e edição de respostas MT

A abordagem que usa resultados prévios derivados pelo método robusto de regressão linear como parâmetro de entrada, pode ser utilizada em substituição ao algoritmo Rho^+ . Esse código computacional resolve o problema inverso de encontrar respostas MT em 1D que melhor ajustem um conjunto de dados de resistividade aparente e fase da impedância considerando a distribuição χ^2 (*qui-quadrado*). Como o algoritmo pode inverter tanto a resistividade aparente como a fase de forma independente, é possível utilizá-lo para prever um elemento a partir do outro de uma maneira fisicamente consistente. Embora originalmente desenvolvido para situações 1D, o algoritmo tem sido utilizado em diversas situações multidimensionais (2D e mesmo 3D) em que as relações entre resistividade aparente e fase podem ser aproximadas para 1D (SMITH, 1995). Um exemplo efetivo do uso desse algoritmo em situações 3D é o resultado que pode ser obtido com sua aplicação aos dados 3D sintéticos aqui usados para validar a transformada de Kramers-Kronig (Figuras 4.14 e 4.15). A Figura 6.1 apresenta o resultado da aplicação do Rho^+ nesses dados sintéticos e mostra que o algoritmo também funciona adequadamente nessa situação 3D.

Figura 6.1 – Saída do algoritmo Rho^+ comparada com dados sintéticos expandidos para períodos mais curtos e mais longos no modelo tridimensional.



Esse uso do Rho^+ em ambientes 3D não se aplica sempre, pois há situações em que as relações entre resistividade aparente e fase diferem substancialmente de situações 1D. Um caso típico é aquele de distorções 3D (ou efeitos de anisotropia no interior da Terra) que fazem com que a fase se estenda para fora do primeiro quadrante (intervalo entre 0° e 90°). Um exemplo experimental típico é observado na região Nordeste do Brasil, sobre a estrutura resistiva do bloco Serrinha do Cráton do São Francisco. Dados MT de uma das estações são mostrados na Figura 6.2 (processamento efetuado para a publicação de Bologna et al. (2017)) e comparados com os resultados da saída do Rho^+ .

Observe-se que os dados dessa estação são de excelente qualidade, não havendo necessidade de aplicar qualquer tipo de correção para minimizar efeitos de possíveis ruídos artificiais. São, porém, instrutivos para mostrar a inadequação do Rho^+ frente a esse tipo de distorção natural. O algoritmo claramente não ajusta as funções de transferência Z_{xy} e Z_{yx} (e as consequentes resistividades aparentes e fases) nos períodos mais longos, onde se concentram os efeitos 3D.

Por outro lado, as Figuras 6.3 e 6.4 mostram a aplicação da transformada de Kramers-Kronig nos dados dessa estação. A transformada $kk-ri$ (Figura 6.3) ajusta de forma adequada os dados de longo período, incluindo os efeitos 3D mais claramente observados na fase. O ajuste não é tão bom nos períodos mais curtos (tipicamente em períodos inferiores a 0,03 s). Por sua vez, a transformada $kk-ir$ (Figura 6.4) ajusta corretamente os dados de curto período, mas tem um ajuste pior nos dados de período mais longo (acima de cerca de 3 s). Nesse caso, um procedimento híbrido pode ser adotado, utilizando a transformada $kk-ir$ para os dados de curto período e a $kk-ri$ para os períodos mais longos.

Figura 6.2 – Saída do algoritmo Rho^+ sobre os dados da estação *ser009a* da região Nordeste do Brasil processados pelo código robusto de Egbert e Booker (1986).

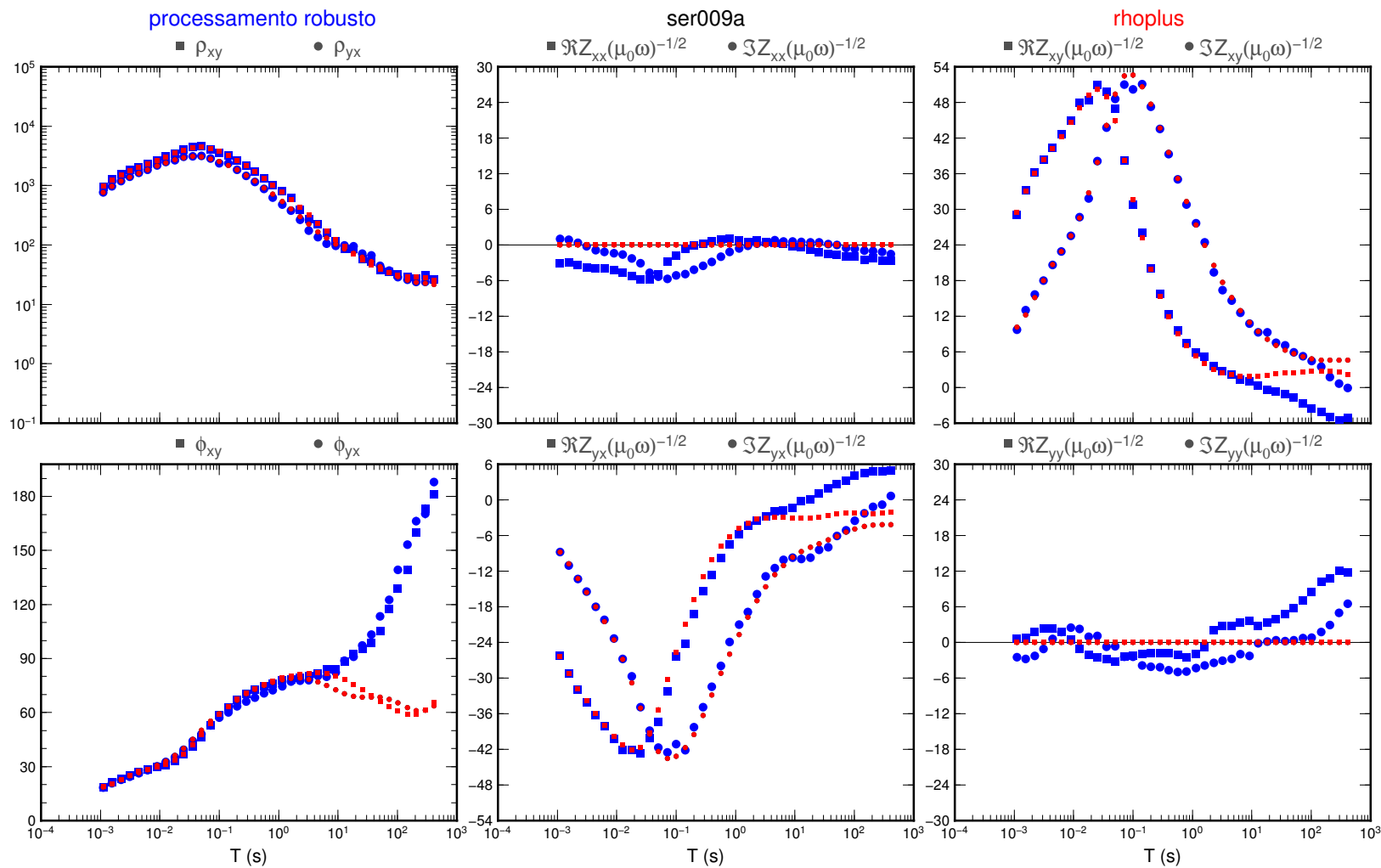


Figura 6.3 – Transformada de Kramers-Kronig para a estação *ser009a* da região Nordeste do Brasil usando a parte real como entrada. As funções de transferência em azul foram obtidas por meio de regressão robusta enquanto aquelas em vermelho foram obtidos usando a Equação 4.3b.

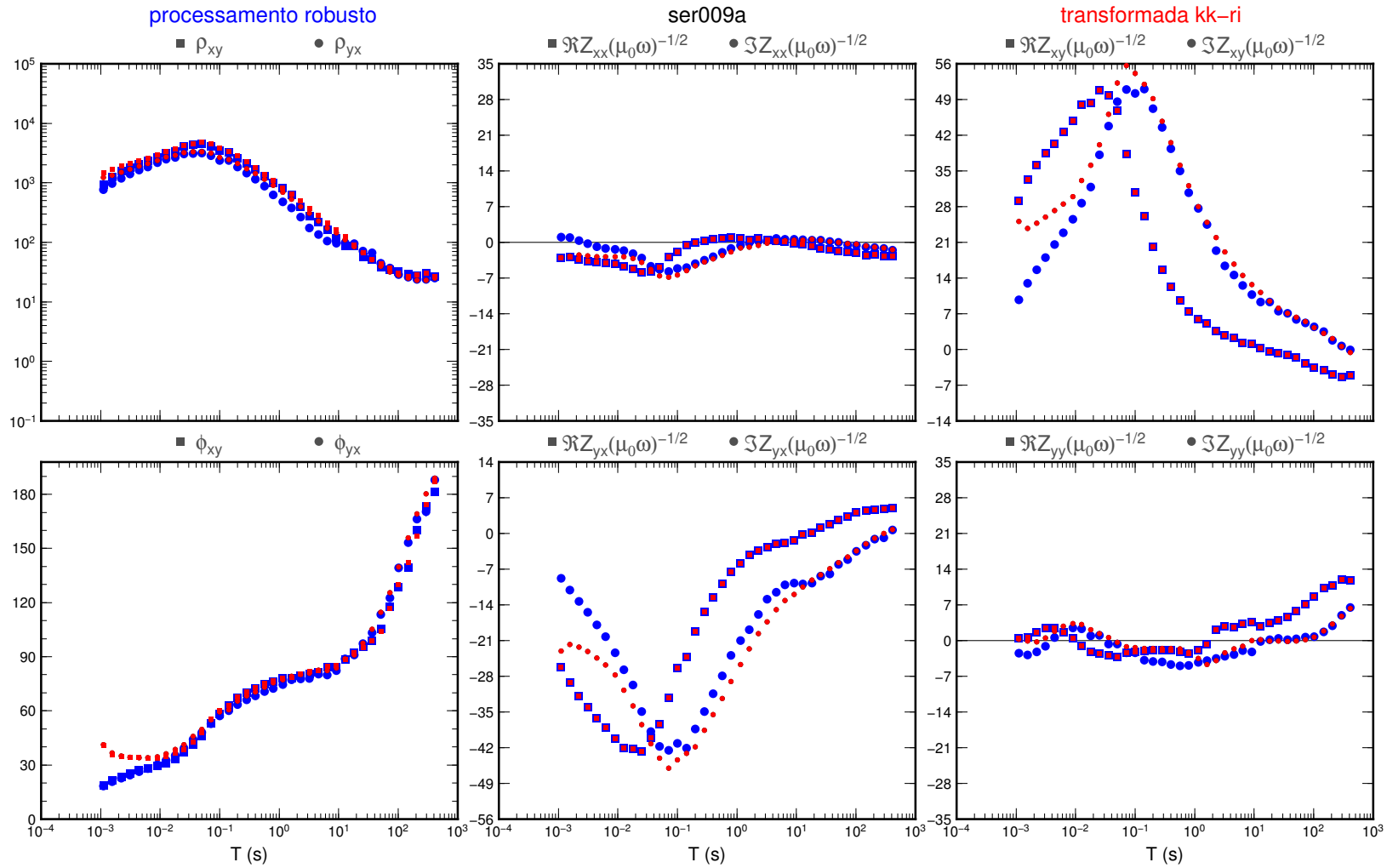
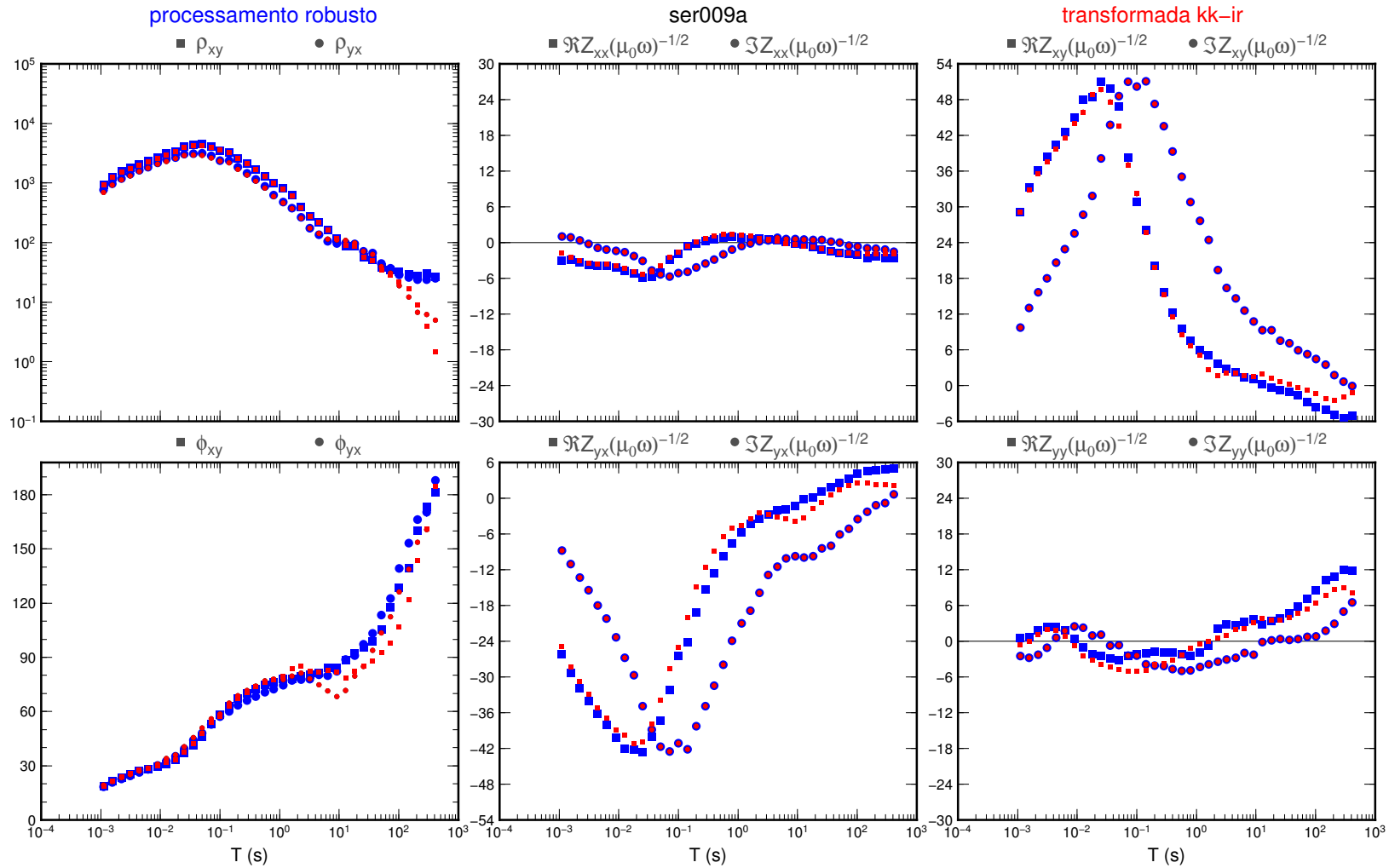


Figura 6.4 – Transformada de Kramers-Kronig para a estação *ser009a* da região Nordeste do Brasil usando a parte imaginária como entrada. As funções de transferência em azul foram obtidas por meio de regressão robusta enquanto aquelas em vermelho foram obtidos usando a Equação 4.3a.



Outro aspecto saliente nessa comparação entre os resultados do Rho^+ e da transformada de Kramers-Kronig refere-se aos elementos da diagonal principal do tensor de impedâncias (Z_{xx} e Z_{yy}). Obviamente, o Rho^+ não se aplica a esses elementos e os resultados nas Figuras 6.1 e 6.2 são apresentados como “zero”. Por outro lado, a transformada de Kramers-Kronig resolve perfeitamente esses dois elementos, conforme mostrado nos vários exemplos discutidos neste texto.

Historicamente, os dados MT foram inicialmente interpretados por modelos 1D ou 2D por serem mais simples de analisar e pela inexistência, até cerca de 15 anos atrás, de códigos computacionais livremente disponíveis para inversão de dados MT em 3D (SIRIPUNVARAPORN et al., 2005). Mesmo após a disponibilização desses programas de inversão, o uso dos elementos da diagonal principal dos elementos do tensor de impedâncias nos procedimentos de inversão 3D ainda são bastante discutidos (MIENSOPUST, 2017). Nesse artigo de revisão, ela argumenta que, embora não seja “zero”, a amplitude desses elementos é, em grande parte dos casos, ordens de grandeza menores que os da diagonal secundária (relacionados a Z_{xy} e Z_{yx}) e, portanto, é basicamente “zero” de forma que os elementos podem ser negligenciados na inversão 3D. Também, discute que o desempenho da inversão 3D é degradado pela inclusão dos elementos da diagonal principal devido às suas menores amplitudes e, conseqüentemente, menores relações sinal-ruído desses elementos. De qualquer forma, porém, os elementos da diagonal principal nos dados MT obtidos no campo, embora com menor amplitude que os da diagonal secundária, são sempre muito maiores do que os derivados nos casos de estudos sintéticos.

Os dados MT de campo geralmente são coletados em perfis únicos ou amplamente separados, ao invés de uma grade de estações bem distribuídas, como idealizado para os programas de inversão 3D. Além disso, mesmo que a distribuição das estações prevista para o campo seja direcionada para uma matriz regular e densa, ela muitas vezes não pode ser cumprida devido a restrições logísticas e circunstâncias externas no campo (por exemplo, terreno, infraestrutura, condições de ruído) ou por falta de tempo, fundos e/ou instrumentação que acabem por levar a uma decisão entre custo *versus* benefício. Em geral, para estações MT adquiridas ao longo de perfis, são feitas tentativas para modelar os dados através de inversões 2D. Possíveis efeitos 3D nos dados são geralmente associados a heterogeneidades 3D locais próximas da superfície e removidos por algoritmos especialmente desenhados para esse objetivo, conforme discutido em Groom et al. (1993), Jones (2012). No entanto, se a geologia regional for efetivamente 3D, essas simplificações podem

levar a artefatos nos modelos gerados pela inversão e a interpretação 3D deve ser usada, mesmo se os dados forem adquiridos ao longo de perfis (PATRO; EGBERT, 2011).

A inversão 3D para dados MT coletados ao longo de perfis é atualmente bastante disseminada (BOLOGNA et al., 2019; ROCHA et al., 2019; SINGH et al., 2019; TRIAHADINI et al., 2019). Um dos aspectos que tem emergido desses estudos com dados reais é que imagens mais realistas sob o perfil de medidas e melhor resolução de estruturas próximas fora do perfil são obtidas quando todos os elementos do tensor de impedâncias são incluídos na inversão. O uso dos elementos da diagonal principal é, porém, afetado pela relação sinal-ruído anteriormente citada. Além disso, Cumming e Mackie (2010) mostraram que as inversões MT em 3D, em geral, são menos tolerantes a dados ruidosos e à ausência de dados em alguns períodos do que as inversões 1-D e 2-D. Isso se deve ao número muito maior de elementos e graus de liberdade que devem ser ajustados pelos dados em um modelo 3-D (comparado a 1-D ou 2-D), o que gera mais possibilidades para uma inversão 3-D produzir resultados irreais se dados ruidosos forem incluídos.

A alternativa que tem sido utilizada para minimizar esses efeitos de ruído na inversão 3D é a realização de uma edição cuidadosa dos dados de entrada. As estratégias usadas incluem a remoção de respostas (funções de transferência) claramente ruidosas, suavização das respostas com uma função do tipo *spline*, interpolação ou inspeção visual e edição manual das respostas (HEISE et al., 2010; TIETZE; RITTER, 2013; MIENSOPUST, 2017). Nenhuma dessas alternativas tem uma base física sólida, como a aqui proposta para a transformada de Kramers-Kronig. Dessa forma, além de uma substituição viável ao Rho^+ , o algoritmo desenvolvido pode ser utilizado em alguns casos para editar os dados de entrada de todos os elementos do tensor de impedância para a inversão 3D.

6.2 Discretização do semiespaço infinito

A obtenção de um modelo de distribuição de resistividades para o interior da Terra (variando lateralmente e com a profundidade) a partir das funções de transferência MT obtidas em várias estações na superfície passa geralmente pelo processo de inversão de dados. A inversão nada mais é do que um procedimento automático de geração de modelos de resistividade para a subsuperfície que ajuste estatisticamente os dados observados. A inversão otimiza automaticamente o modelo gerado entre diferentes soluções possíveis, atualizando iterativamente os parâmetros do modelo de modo a obter o melhor ajuste possível para os dados dentro de um

limite previamente estabelecido.

A parte central de qualquer algoritmo de inversão é o processo de cálculo direto (*forward modeling*) que envolve o cálculo da resposta teórica gerada na superfície pelos modelos de variação de resistividades para o interior da Terra (suposto um semiespaço infinito). Genericamente, o objetivo do cálculo direto é resolver as equações de Maxwell em meios uniformes, sendo baseado em diferentes tipos de abordagens que incluem técnicas de equações integrais - *IE*, elementos finitos - *FE* ou diferenças finitas - *FD* (SIRIPUNVARAPORN, 2012). Para o caso 3D, os métodos de diferenças finitas têm a vantagem de serem mais adequados para modelos mais complicados e generalizados e é o usado pelo código *ModEM* utilizado neste trabalho. Esse método exige a discretização do domínio onde vai ser feita a inversão, incluindo não apenas a área de investigação mas também um domínio adicional em torno dessa área. A discretização é normalmente feita por planos horizontais e verticais que se intersectam gerando um grande número de células (cubos) no domínio da inversão. A cada um desses cubos é atribuído um valor de resistividade que é alterado no processo de inversão.

Com base nessa forma discreta do modelo, as equações fundamentais do método MT (incluindo condições de contorno apropriadas) são resolvidas. A formulação exata do sistema de equações depende da aproximação usada (*IE*, *FE* ou *FD*), mas geralmente o sistema de equações é resolvido para um dos campos (elétrico ou magnético) em toda a malha. Como os campos elétrico e magnético são relacionados entre si pelas Equações de Maxwell, o campo calculado pode ser usado para determinar o outro (também em toda a malha). Conhecendo os campos elétrico e magnético, pode-se estimar os valores de impedância (e , conseqüentemente, resistividade aparente e fase) nos locais de observação. Essa resposta sintética derivada do modelo é comparada com os dados medidos e sua diferença é geralmente representada pelo desvio quadrático médio (*RMS*). As alterações aplicadas aos modelos diretos nos programas de inversão são impulsionadas pela minimização do *RMS* entre os dados observados e aquele predito a partir do modelo.

Como o algoritmo de inversão é baseado nesse cálculo direto, qualquer imprecisão ou erro nesse processo de solução direta é propagado para a inversão. A confiabilidade desse cálculo deriva em grande parte da construção de uma malha 3D adequada. Esse processo de discretização do domínio onde vai se fazer a inversão vai ser sempre ditado por um compromisso entre a precisão numérica desejada na modelagem direta e os custos computacionais (ou seja, tempo e recursos ne-

cessários). Quanto mais fina for a grade, maior será o tempo de processamento computacional e o recurso de memória necessários.

A confiabilidade do cálculo direto das respostas MT obtidas a partir de uma determinada discretização do semiespaço em que é compreendido o modelo inverso é um dos temas ainda em aberto nos estudos experimentais. Em geral, toma-se bastante cuidado para digitalizar a parte superficial dos modelos para considerar, por exemplo, efeitos de topografia. Por outro lado, a parte inferior da malha é bastante negligenciada. O procedimento geralmente utilizado para permitir mais flexibilidade e melhor ajuste dos dados é estender o domínio da inversão até uma profundidade maior do que a profundidade pelicular na frequência mais baixa (MIENSOPUST, 2017). Esse cálculo, porém, é baseado no modelo de partida utilizado para dar início ao processo de inversão. O modelo final não é testado, sendo que qualquer problema com convergência dos resultados relacionada com a profundidade da parte inferior do modelo pode gerar artefatos que podem inclusive afetar partes distantes do modelo. Esses artefatos são, na maioria das vezes, automaticamente evitados em estudos sintéticos, já que geralmente para testes numéricos usa-se uma combinação correspondente de estrutura de modelo e malha, mas eles podem ser relevantes para a inversão de dados de campo.

Os resultados obtidos na Seção 4.3.2 de testes sintéticos da transformada de Kramers-Kronig em um caso tridimensional mostram que o algoritmo desenvolvido pode auxiliar na avaliação da conformidade da discretização do semiespaço do modelo 3D. Observe-se que as extrapolações feitas naquela seção, tanto para obter resultados em períodos tão longos (5 336 800 s) como na extensão do semiespaço até profundidades de 5 426 km mantendo a resistividade fixa em $100 \Omega \text{ m}$, não tem qualquer significado prático. Em geral, os dados MT de longo período coletados no campo chegam até períodos pouco superiores a $1 \times 10^4 \text{ s}$, raramente se aproximando de $1 \times 10^5 \text{ s}$. Períodos muito acima desses valores podem trazer problemas com a aproximação de ondas planas utilizada pelo método MT. Também, baseado em condições do manto e minerais e com o apoio de diferentes observações sobretudo de dados de observatórios geomagnéticos, sabe-se que há uma importante descontinuidade na resistividade elétrica em profundidades maiores que 660 km. Nas malhas projetadas para grande profundidade de penetração do sinal MT, as profundidades acima desse limite comumente aparecem como um semiespaço de baixa resistividade que é mantido fixo durante o processo de inversão. Valores usados para a resistividade desse condutor profundo podem chegar a $3 \Omega \text{ m}$ (BURD et al., 2014), enquanto outros estudos sugerem valores ainda mais baixos, da ordem

de $1 \Omega \text{ m}$ (OLSEN, 1999). A inclusão de valores de resistividade tão mais baixos na base do modelo 3D faria com que os resultados convergissem muito mais rapidamente, sem a necessidade de estender o modelo até profundidades tão grandes como nesse teste sintético. Independente dessas considerações, as comparações com a transformada de Kramers-Kronig mostraram que a discretização originalmente usada pelo modelo 3D de teste (profundidade máxima de 626 km) apresenta instabilidade no cálculo direto das funções de transferência nos períodos mais longos.

Esse último resultado oferece uma alternativa para utilizar o procedimento desenvolvido neste trabalho para avaliar a qualidade da malha utilizada na inversão 3D. Pode-se testar o resultado final da inversão através de cálculo direto para obter as respostas previstas pelo modelo inverso. Sobre esses resultados previstos pode-se aplicar a transformada de Kramers-Kronig na forma descrita anteriormente para obter os resultados mostrados nas Figuras 4.10 e 4.11. Dessa forma, pode-se verificar possíveis problemas nas condições de contorno na parte mais profunda do modelo. Em princípio, não é necessário testar integralmente o modelo inverso, mas apenas escolher as regiões de menor condutância integrada (somatória das razões entre espessura e resistividade para cada uma das camadas do modelo) pois espera-se que essa seja a região mais sensível a problemas na discretização do modelo.

Um exemplo desse uso do resultado previsto para testar a malha usada na inversão foi feita com a modelagem 3D de dados MT da região Nordeste do Brasil. Nesse caso, o modelo foi gerado por Padilha et al. (2017) para a região limítrofe entre a Província Borborema e a Bacia do Parnaíba. Para testar o uso da transformada de Kramers-Kronig foram escolhidas duas estações: uma sobre a região condutora da Bacia do Arararipe, que além do pacote sedimentar superficial apresenta também um condutor crustal mais profundo (estação *bobc12a*) e outra sobre a região resistiva do domínio Ceará-Central da Província Borborema (estação *bobc18a*). As Figuras 6.5 a 6.8 mostram o resultado desses testes.

Figura 6.5 – Transformada $kk-ir$ comparada com dados preditos pelo modelo inverso 3D para a estação condutora *bobc12a* na região Nordeste do Brasil.

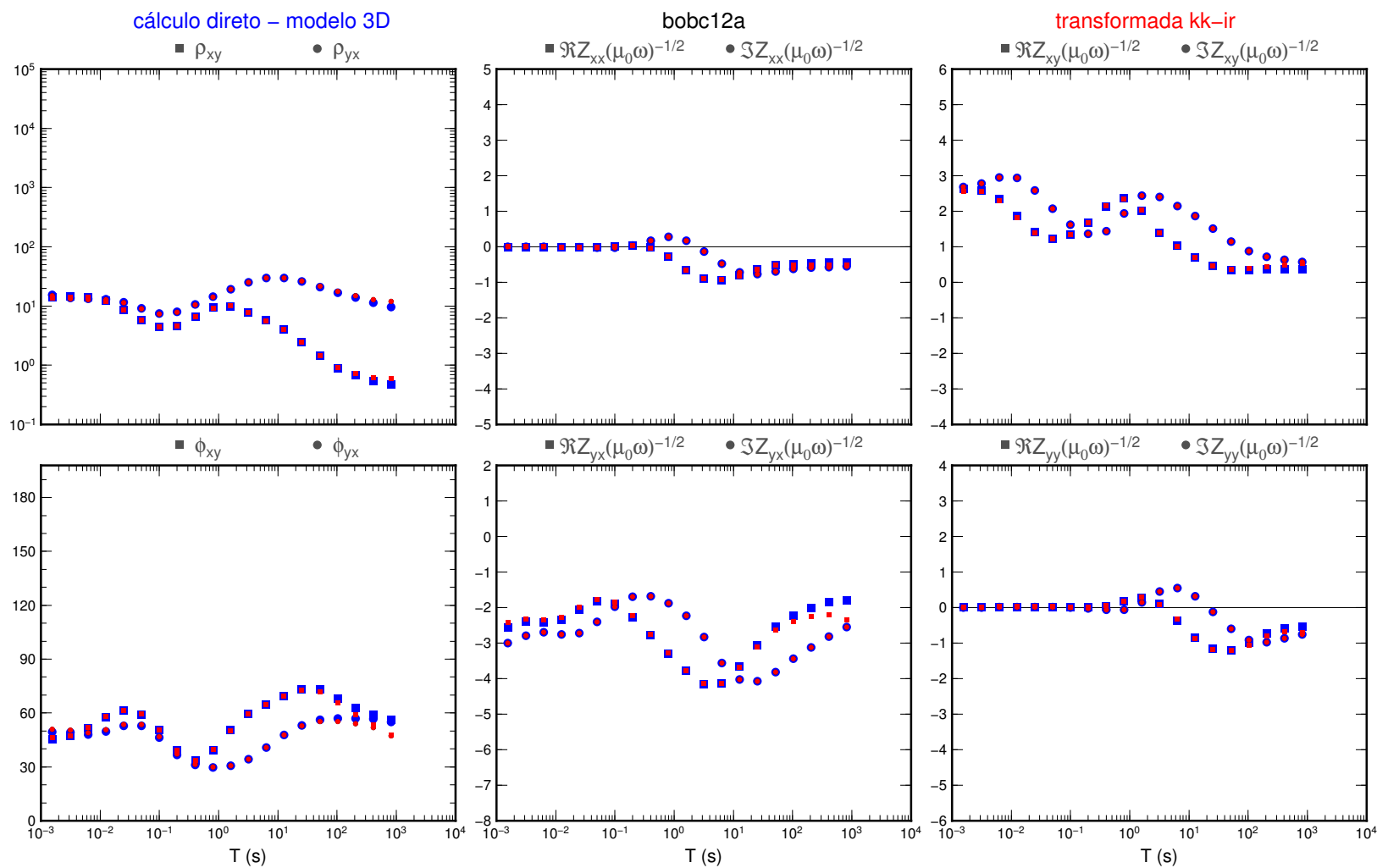


Figura 6.6 – Transformada kk - ri comparada com dados preditos pelo modelo inverso 3D para a estação condutora *bobc12a* na região Nordeste do Brasil.

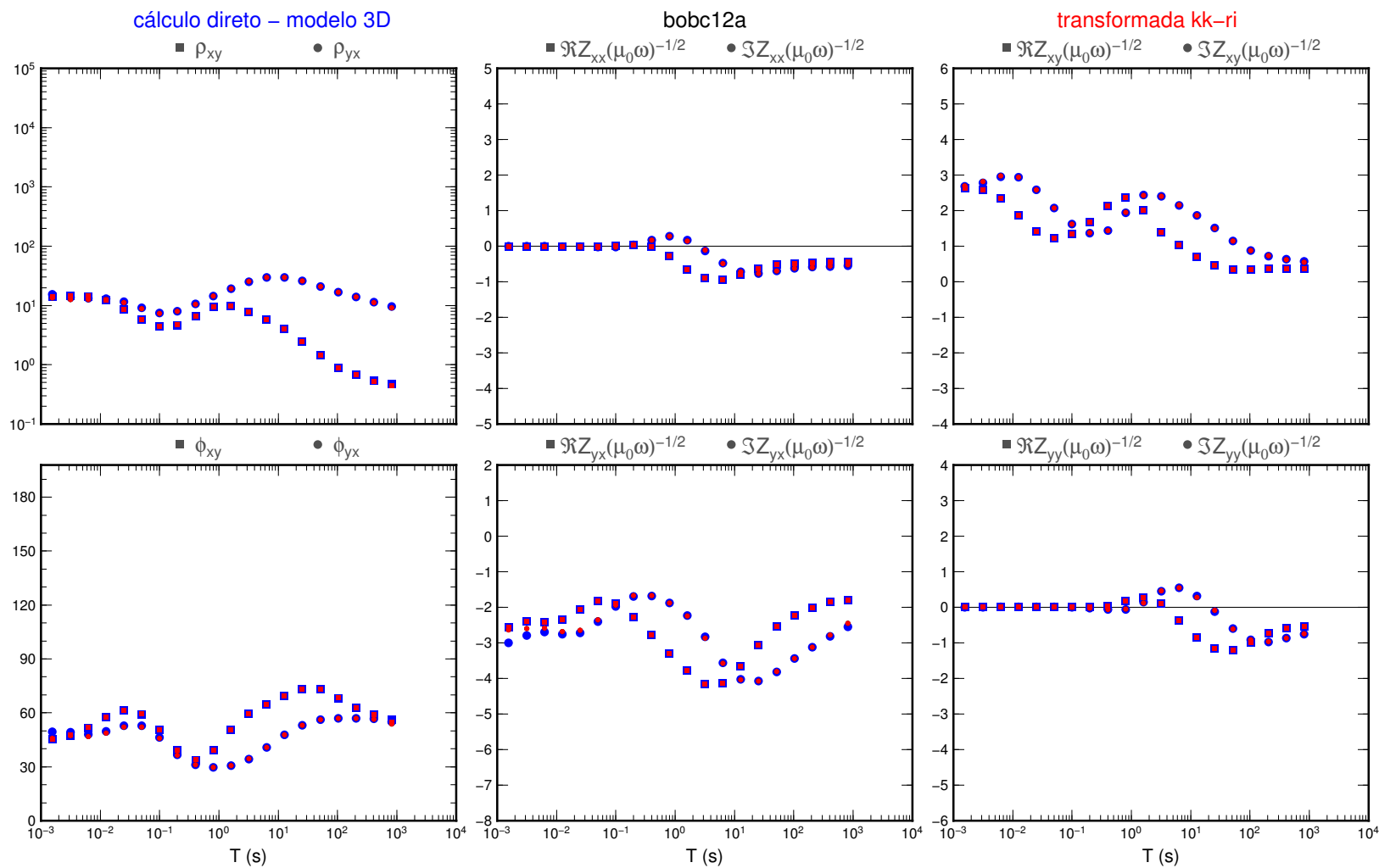


Figura 6.7 – Transformada kk - ir comparada com dados preditos pelo modelo inverso 3D para a estação resistiva *bobc18a* na região Nordeste do Brasil.

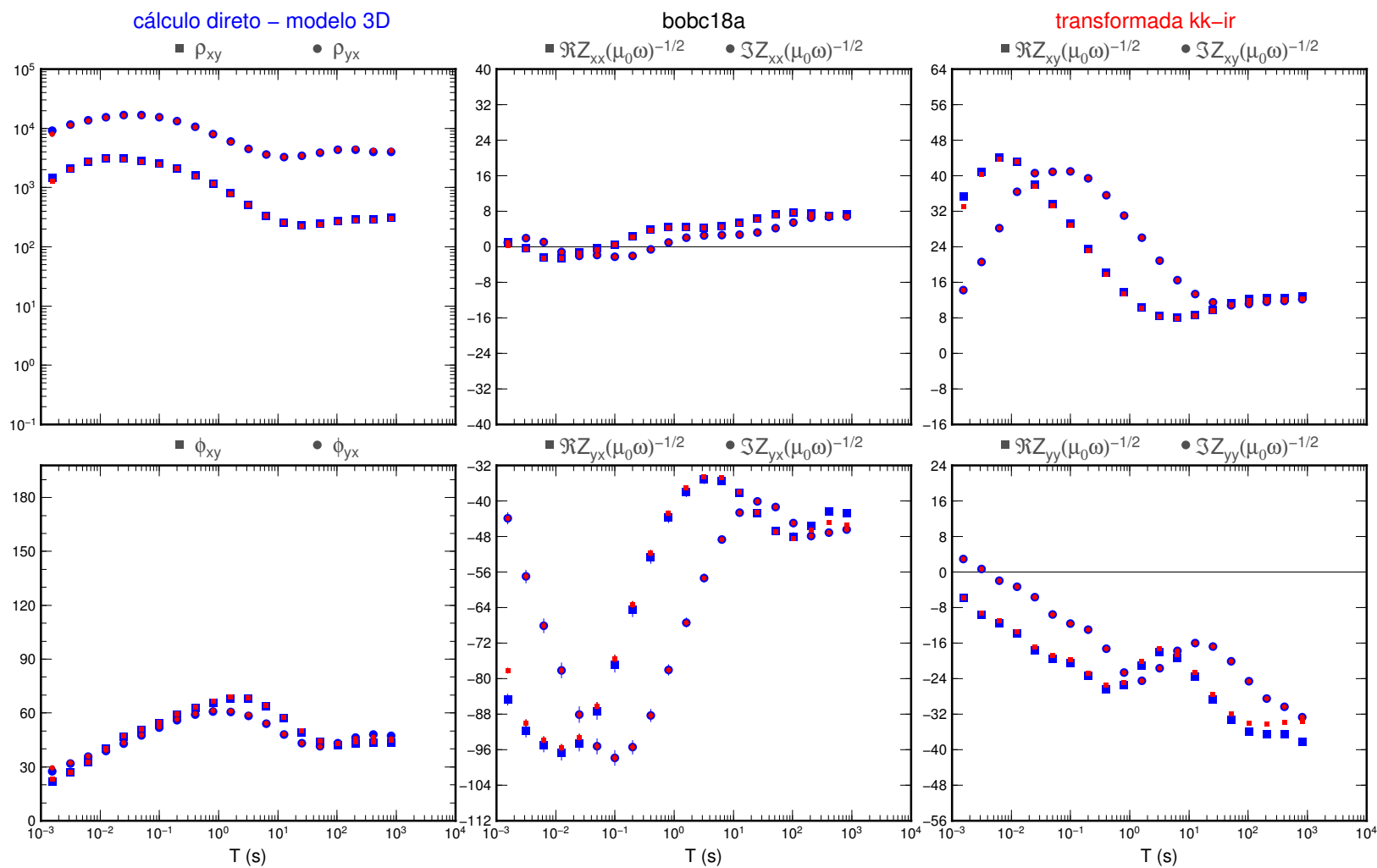
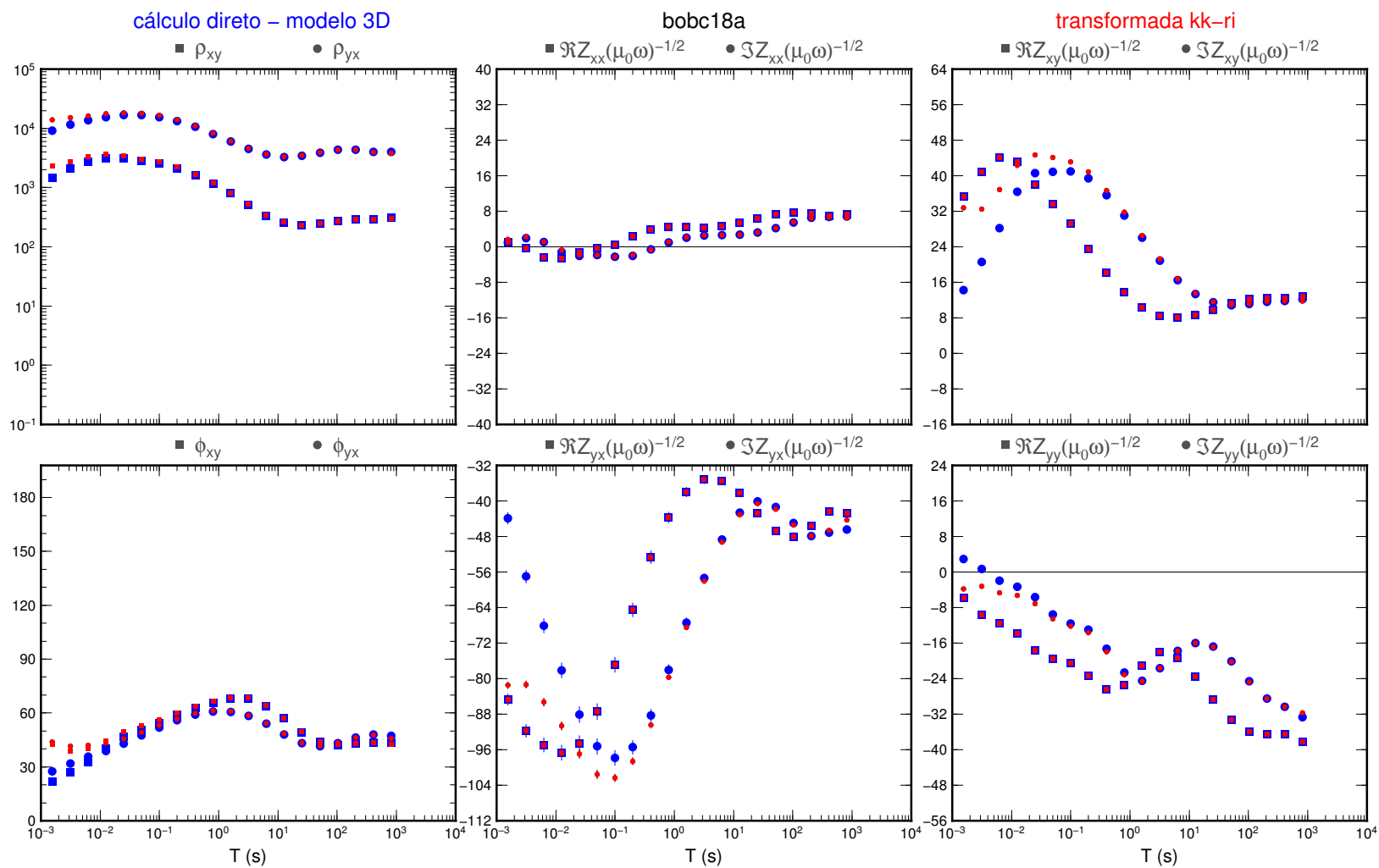


Figura 6.8 – Transformada kk - ri comparada com dados preditos pelo modelo inverso 3D para a estação resistiva *bobc18a* na região Nordeste do Brasil.



A Figura 6.5 mostra a transformada $kk-ir$ sobre os dados preditos da estação *bobc12a*. Pode-se notar que a transformada reproduz os valores da parte real, com pequena discrepância nos períodos mais longos. Como observado nos testes com dados sintéticos, essa diferença se deve ao fato de ainda não se ter alcançado a estabilidade do semiespaço condutor na parte imaginária dos dados gerados pelo modelo. Não há, porém, problema aparente com a malha utilizada na inversão. Já a transformada $kk-ri$ sobre os dados dessa mesma estação (Figura 6.6) mostra diferenças pequenas concentradas nos períodos mais curtos. A mesma explicação anterior pode ser usada, com a diferença nos resultados sendo gerada por não se ter atingido a estabilidade na parte real dos elementos do tensor requerido pelo semiespaço infinito na parte rasa do modelo. Novamente, esse não é um problema com a malha utilizada, pois as discrepâncias que seriam geradas pelas condições de contorno deveriam ser bem mais significativas (ver os testes sintéticos das Figuras 4.8 e 4.9).

A comparação dos dados preditos pelo modelo 3D com a transformada $kk-ir$ da estação *bobc18a* é mostrada na Figura 6.7. Analogamente ao caso da estação anterior, aparecem pequenas discrepâncias em períodos longos, mas que não podem ser atribuídos a problemas de digitalização do semiespaço. Por outro lado, a transformada $kk-ri$ dessa mesma estação (Figura 6.8) mostra significativas diferenças em períodos mais curtos. Nesse caso, a fase não converge para 45° e, pelos testes anteriores da Seção 4.3.2, o problema poderia ser corrigido subdividindo a parte mais rasa da malha. Como conclusão, as regiões mais resistivas do modelo 3D de Padilha et al. (2017) apresentam divergências entre as partes reais e imaginárias dos elementos do tensor de impedâncias que podem ser resolvidas com uma discretização mais fina das camadas superficiais do modelo (subdivisão das camadas).

6.3 Caso de ruídos extremos

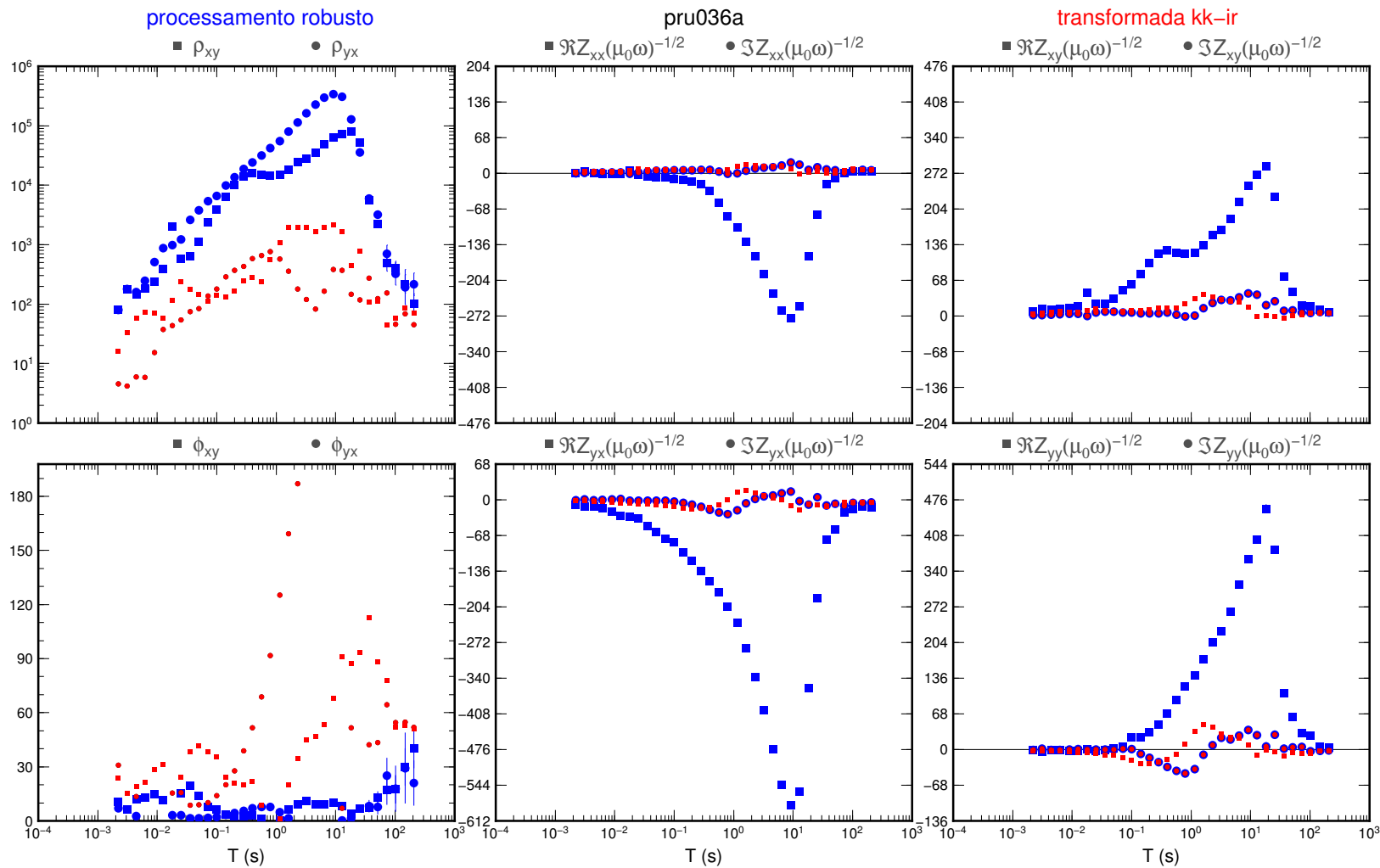
O objetivo principal do trabalho era buscar um processamento alternativo àqueles estabelecidos pela comunidade de indução eletromagnética para os dados MT que permitisse minimizar o efeito de ruídos estacionários e não correlacionados, especialmente considerando a expansão prevista de redes HVDC no Brasil. A estratégia antecipada era baseada na suposição de que amostras derivadas do sinal natural e que obedecessem o princípio de ondas planas formariam uma concentração identificável de valores no espaço real-imaginário dos elementos do tensor de impedâncias para cada frequência alvo. Infelizmente essa estratégia não funcionou no caso prático usado como teste, pois o ruído EM com ponto de ruptura superior a 50 % contamina todo esse espaço e com uma concentração de

amostras muito superior àquela que poderia estar associada ao sinal natural.

A alternativa disponível no momento é tentar usar a opção da transformada de Kramers-Kronig utilizando como dados de entrada os resultados do processamento robusto. Conforme discutido anteriormente, essa alternativa só se aplica no caso dos ruídos não serem tão significativos, não se concentrando simultaneamente nas componentes reais e imaginárias da impedância. Essa situação dificilmente vai ser encontrada nas proximidades de uma linha de transmissão de energia de alta voltagem (especialmente no caso das HVDC). Como exemplo, a Figura 6.9 mostra a aplicação da transformada *kk-ir* nos dados de contaminação extrema da estação *pru036a* na Bacia do Paraná. Nas discussões realizadas no Capítulo 3 desta tese, esse foi o exemplo de estação usada para mostrar o efeito do ruído afetando principalmente uma das componentes (no caso, a parte real). Na prática, porém, o ruído afeta também a parte imaginária, apenas que em menor intensidade que na parte real. Esse aspecto é salientado no resultado da transformada *kk-ir* nessa figura que mostra resultados extremamente perturbados para as resistividades aparentes e fases obtidas a partir das impedâncias geradas pela transformada.

Figura 6.9 – Transformada de Kramers-Kronig para a estação *pru036a* da Bacia do Paraná, usando a parte imaginária como entrada. As funções de transferência em azul foram obtidas por meio da regressão robusta enquanto aquelas em vermelho foram obtidas usando a Equação 4.3a.

100



7 CONCLUSÕES

Este trabalho mostra que a transformada de Kramers-Kronig pode ser utilizada para substituir o difundido método Rho^+ para verificar a coerência dos tensores de impedâncias estimados pelos métodos robustos tradicionais utilizados no processamento de dados MT. A maior vantagem do método aqui descrito é que ele é capaz de verificar todas as componentes do tensor, enquanto que o método Rho^+ verifica apenas as componentes da diagonal secundária. Essa vantagem fica mais evidente em situações com estruturas 3D onde a fase dos elementos da diagonal secundária ultrapassam 90° e em sondagens de longos períodos quando os valores das componentes da diagonal principal do tensor de impedâncias se afastam do zero.

Este trabalho também mostra como a transformada de Kramers-Kronig pode ser utilizada como método de validação dos resultados de inversões MT 3D. Nesse caso a transformada foi capaz de indicar corretamente se malha de discretização era compacta e profunda o suficiente para assegurar os resultados da inversão.

Apesar do sucesso em utilizar a transformada de Kramers-Kronig para validar os tensores de impedâncias obtidos por meio de métodos robustos, o método proposto para substituir a regressão linear por uma técnica de amostragem associada à transformada para obter os tensores não obteve sucesso. A quantidade de memória computacional utilizada e de tempo de processamento se mostrou insuficiente para obter bons resultados. No futuro esse método pode ser melhor utilizado com técnicas mais eficientes de seleção de amostras.

Sugestões para atividades futuras buscando incrementar a aplicação proposta para a transformada de Kramers-Kronig incluem:

alterar procedimentos para obtenção da transformada de Kramers-Kronig:

- buscar alternativas para as funções de extrapolações utilizadas (Equações 4.4): o processo de extrapolação não se mostrou adequado para várias das situações em períodos extremos (muito curtos e muito longos); é necessário buscar outras opções para melhorar o processo de validação e assim permitir o uso do código desenvolvido como uma espécie de $\text{Rho}^+ - 3D$;
- avaliar outros métodos para a análise espectral (seção 5.2.1): o ob-

jetivo seria verificar se se consegue uma melhor identificação do sinal natural no processo de busca, em um espectro dominado por ruídos artificiais; alternativas podem incluir cálculos de espectros auto-regressivos (*ar*), auto-regressivos com média móvel (*arma*) e *wavelets*.

tentar alternativas para o processo de busca:

- usar o resultado da validação para filtrar os segmentos da série temporal que contenham apenas os resultados consistentes com o da transformada de Kramers-Kronig; equivaleria a selecionar o hexágono que contém o ponto vermelho da Figura 5.6;
- com os elementos filtrados do item anterior, fazer a análise e classificação dos segmentos que são consistentes e inconsistentes com a transformada; pode-se aplicar um algoritmo de classificação em IA (inteligência artificial) para auxiliar na tarefa; caso seja encontrado um padrão nos espectros consistentes, buscar uma interpretação para esses espectros para tentar melhorar o processo de busca.

incrementar o código com outros resultados:

- implementar a transformada de Kramers-Kronig para o *tipper*;
- estimar um intervalo de confiança (barra de erros) para a transformada de Kramers-Kronig; não pode ser feito usando as fórmulas convencionais de propagação de erro e uma alternativa seria uma espécie de método de Monte Carlo, usando os erros da regressão robusta;
- implementar a alternativa de usar informação de referência remota junto com a transformada.

integrar com o código robusto desenvolvido no Mestrado (PÁDUA, 2014):

- com essa integração não vai ser mais necessário executar o programa robusto de Egbert e Booker (1986) antes de executar a transformada; um programa integrado pode fazer o cálculo robusto e, em seguida, desenvolver a validação com a transformada de Kramers-Kronig.

acelerar o processamento e automatizar o código para distribuição a terceiros:

- automatizar o processo iterativo das transformadas para desenvolvimento operacional de um novo código ($Rho^+ - 3D$), que permita

obter de forma automática os resultados mostrados na Figura 5.4; vários testes serão necessários incluindo a definição de por qual transformada se deve iniciar (*kk-ir* ou *kk-ri*); esse processo tem que ser feito depois de modificar a função de extrapolação, pois ela influencia significativamente o resultado da transformada em períodos curtos e longos;

- selecionar automaticamente os períodos anômalos (Figuras 5.4 e 5.10); nesse caso, basta sugerir um determinado desvio prévio (*epsilon*) e verificar em quais períodos a diferença entre o valor da transformada e da regressão robusta excedem esse *epsilon*;
- otimizar o código e paralelizar, para permitir que as diferentes frequências sejam analisadas simultaneamente;
- implementar uma forma de "container docker" (<https://www.docker.com/>) com os programas e scripts utilizados no processamento; esse processo facilita a distribuição e gerenciamento das dependências dos programas e das bibliotecas utilizadas e é bastante utilizado em trabalhos de pesquisa reproduzível.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDERSON, E.; BAI, Z.; BISCHOF, C.; BLACKFORD, S.; DEMMEL, J.; DONGARRA, J.; CROZ, J. D.; GREENBAUM, A.; HAMMARLING, S.; McKenney, A.; SORENSEN, D. **LAPACK users' guide**. 3. ed. Philadelphia, PA: Society for Industrial and Applied Mathematics, 1999. ISBN 0-89871-447-8 (paperback). Disponível em: <<http://www.netlib.org/lapack/lug/>>. 137

ARYA, S.; MOUNT, D.; KEMP, S. E.; JEFFERIS, G. **RANN: fast nearest neighbour search (wraps ANN library) using L2 metric**. [s.n.], 2019. R package version 2.6.1. Disponível em: <<https://github.com/jefferis/RANN>>. 137

BACHE, S. M.; WICKHAM, H. **magrittr: a forward-pipe operator for R**. [s.n.], 2014. R package version 1.5. Disponível em: <<https://CRAN.R-project.org/package=magrittr>>. 137

BANKS, R. J. The effects of non-stationary noise on electromagnetic response estimates. **Geophysical Journal International**, v. 135, n. 2, p. 553–563, 1998. Disponível em: <<http://dx.doi.org/10.1046/j.1365-246X.1998.00661.x>>. 1

BASOKUR, A. T. Definitions of apparent resistivity for the presentation of magnetotelluric sounding data. **Geophysical Prospecting**, v. 42, n. 2, p. 141–149, 1994. Disponível em: <<https://doi.org/10.1111/j.1365-2478.1994.tb00203.x>>. 36

BERDICHEVSKY, M. Marginal notes on magnetotellurics. **Surveys in Geophysics**, v. 20, n. 3/4, p. 341–375, 1999. Disponível em: <<https://doi.org/10.1023/a:1006645715819>>. 31

BERDICHEVSKY, M.; DMITRIEV, V. **Magnetotellurics in the context of the theory of ill-posed problems**. Society of Exploration Geophysicists, 2002. Disponível em: <<http://library.seg.org/doi/abs/10.1190/1.9781560802068>>. 8, 32

BERDICHEVSKY, M. N.; DMITRIEV, V. I. **Models and methods of magnetotellurics**. [S.l.]: Springer-Verlag, 2008. 8, 31

BLANCO, J. L.; RAI, P. K. **nanoflann: a C++ header-only fork of FLANN, a library for Nearest Neighbor (NN) with KD-trees**. 2014. Disponível em: <<https://github.com/jlblancoc/nanoflann>>. 136

BOLOGNA, M. S.; DRAGONE, G. N.; MUZIO, R.; PEEL, E.; NUÑEZ-DEMARCO, P.; USSAMI, N. Electrical structure of the lithosphere from rio de la plata craton to Paraná basin: amalgamation of cratonic and refertilized lithospheres in sw gondwanaland. **Tectonics**, v. 38, n. 1, p. 77–94, 2019. Disponível em:

<<https://doi.org/10.1029/2018tc005148>>. 90

BOLOGNA, M. S.; EGBERT, G. D.; PADILHA, A. L.; PÁDUA, M. B.; VITORELLO, Í. 3-D inversion of complex magnetotelluric data from an archean-proterozoic terrain in northeastern São Francisco craton, Brazil. **Geophysical Journal International**, v. 210, n. 3, p. 1545–1559, 2017. Disponível em:

<<https://doi.org/10.1093/gji/ggx261>>. 85

BOND-LAMBERTY, B. Data sharing and scientific impact in eddy covariance research. **Journal of Geophysical Research: Biogeosciences**, v. 123, n. 4, p. 1440–1443, 2018. Disponível em:

<<https://doi.org/10.1002/2018jg004502>>. 121

BURD, A. I.; BOOKER, J. R.; MACKIE, R.; FAVETTO, A.; POMPOSIELLO, M. C. Three-dimensional electrical conductivity in the mantle beneath the Payún Matrú volcanic field in the andean backarc of Argentina near 36.5°S: Evidence for decapitation of a mantle plume by resurgent upper mantle shear during slab steepening. **Geophysical Journal International**, v. 198, n. 2, p. 812–827, 2014.

Disponível em: <<https://doi.org/10.1093/gji/ggu145>>. 92

BURRELL, A. G.; HALFORD, A.; KLENZING, J.; STONEBACK, R. A.; MORLEY, S. K.; ANNEX, A. M.; LAUNDAL, K. M.; KELLERMAN, A. C.; STANSBY, D.; MA, J. Snakes on a spaceship - an overview of python in heliophysics. **Journal of Geophysical Research: Space Physics**, v. 123, n. 12, p. 10,384–10,402, 2018.

Disponível em: <<https://doi.org/10.1029/2018ja025877>>. 121

CAGNIARD, L. Basic theory of the magneto-telluric method of geophysical prospecting. **Geophysics**, v. 18, n. 3, p. 605–635, jul. 1953. Disponível em:

<<https://doi.org/10.1190/1.1437915>>. 5, 10, 14

CAMPOS-ENRIQUEZ, J.; CORBO-CAMARGO, F.; ARZATE-FLORES, J.; KEPPIE, J.; ARANGO-GALVÁN, C.; UNSWORTH, M.; BELMONTE-JIMÉNEZ, S. The buried southern continuation of the Oaxaca-Juarez terrane boundary and Oaxaca fault, southern Mexico: magnetotelluric constraints. **Journal of South American Earth Sciences**, v. 43, p. 62–73, abr. 2013. ISSN 0895-9811. Disponível em:

<<http://www.sciencedirect.com/science/article/pii/S0895981113000060>>. 123

CHANG, G.; ROTH, C. B.; REYES, C. L.; PORNILLOS, O.; CHEN, Y.-J.; CHEN, A. P. Retraction. **Science**, v. 314, n. 5807, p. 1875–1875, 2006. Disponível em: <<http://dx.doi.org/10.1126/science.314.5807.1875b>>. 122

CHAVE, A. D. Estimation of the magnetotelluric response function. In: CHAVE, A. D.; JONES, A. G. (Ed.). **The magnetotelluric method: theory and practice**. [S.l.]: Cambridge University Press, 2012. ISBN 9780521819275. 17

CHAVE, A. D.; JONES, A. G. Introduction to the magnetotelluric method. In: CHAVE, A. D.; JONES, A. G. (Ed.). **The magnetotelluric method: theory and practice**. [S.l.]: Cambridge University Press, 2012. ISBN 9780521819275. 1, 8

CHAVE, A. D.; JONES, A. G. (Ed.). **The magnetotelluric method: theory and practice**. [S.l.]: Cambridge University Press, 2012. ISBN 9780521819275. 6

CHAVE, A. D.; THOMSON, D. J. Some comments on magnetotelluric response function estimation. **Journal of Geophysical Research**, v. 94, n. B10, p. 14215–14225, 1989. ISSN 0148-0227. Disponível em: <<http://dx.doi.org/10.1029/JB094iB10p14215>>. 1, 20

CHEN, X.; DALLMEIER-TIESEN, S.; DASLER, R.; FEGER, S.; FOKIANOS, P.; GONZALEZ, J. B.; HIRVONSALO, H.; KOUSIDIS, D.; LAVASA, A.; MELE, S.; RODRIGUEZ, D. R.; ŠIMKO, T.; SMITH, T.; TRISOVIC, A.; TRZCINSKA, A.; TSANAKTSIDIS, I.; ZIMMERMANN, M.; CRANMER, K.; HEINRICH, L.; WATTS, G.; HILDRETH, M.; IGLESIAS, L. L.; LASSILA-PERINI, K.; NEUBERT, S. Open is not enough. **Nature Physics**, v. 15, n. 2, p. 113–119, 2018. Disponível em: <<https://doi.org/10.1038/s41567-018-0342-2>>. 136

CONSTANTINE, W. **ifultools: insightful research tools**. [s.n.], 2016. R package version 2.0-4. Disponível em: <<https://CRAN.R-project.org/package=ifultools>>. 137

CUMMING, W.; MACKIE, R. Resistivity imaging of geothermal resources using 1D, 2D and 3D MT inversion and TDEM static shift correction illustrated by a glass mountain case history. In: **WORLD GEOTHERMAL CONGRESS 2010. Proceedings...** [s.n.], 2010. p. 1–10. Disponível em: <<http://www.geothermal-energy.org/pdf/IGAstandard/WGC/2010/1396.pdf>>. 90

CURTIN, R. R.; EDEL, M.; LOZHNIKOV, M.; MENTEKIDIS, Y.; GHASAS, S.; ZHANG, S. mlpack 3: a fast, flexible machine learning library. **Journal of Open Source Software**, v. 3, n. 26, p. 726, 2018. Disponível em: <<https://doi.org/10.21105/joss.00726>>. 136

ECMA INTERNATIONAL. **ECMA-404: the JSON Data Interchange Format. 2.** ed. Geneva, Switzerland: ECMA (European Association for Standardizing Information and Communication Systems), 2017. Disponível em: <http://www.ecma-international.org/publications/standards/Ecma-404.htm>. 137

EFRON, B. Bootstrap methods: another look at the jackknife. **The Annals of Statistics**, v. 7, n. 1, p. 1–26, 1979. Disponível em: <https://doi.org/10.1214/aos/1176344552>. 65

EGBERT, G.; LIVELYBROOKS, D. Single station magnetotelluric impedance estimation: coherence weighting and the regression M-estimate. **Geophysics**, v. 61, n. 4, p. 964–970, jul. 1996. ISSN 0016-8033. Disponível em: <http://dx.doi.org/10.1190/1.1444045>. 2

EGBERT, G. D. Robust multiple-station magnetotelluric data processing. **Geophysical Journal International**, v. 130, n. 2, p. 475–496, 1997. Disponível em: <http://dx.doi.org/10.1111/j.1365-246X.1997.tb05663.x>. 21

EGBERT, G. D.; BOOKER, J. R. Robust estimation of geomagnetic transfer functions. **Geophysical Journal of the Royal Astronomical Society**, v. 87, n. 1, p. 173–194, out. 1986. ISSN 1365-246X. Disponível em: <http://gji.oxfordjournals.org/content/87/1/173.abstract>. xiii, xv, xvii, 2, 18, 19, 20, 23, 24, 32, 55, 57, 63, 64, 66, 68, 86, 102, 122, 131

EGBERT, G. D.; KELBERT, A. Computational recipes for electromagnetic inverse problems. **Geophysical Journal International**, v. 189, n. 1, p. 251–267, 2012. ISSN 1365-246X. Disponível em: <https://doi.org/10.1111/j.1365-246x.2011.05347.x>. 41, 46

ESCALAS, M.; QUERALT, P.; LEDO, J.; MARCUELLO, A. Polarisation analysis of magnetotelluric time series using a wavelet-based scheme: a method for detection and characterisation of cultural noise sources. **Physics of the Earth and Planetary Interiors**, v. 218, p. 31–50, maio 2013. ISSN 0031-9201. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0031920113000307>. 1, 2

ESTER, M.; KRIEGEL, H.-P.; SANDER, J.; XU, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: **INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 2., 1996. Proceedings...** [S.l.: s.n.], 1996. 69

FONTES, S.; HARINARAYANA, T.; DAWES, G.; HUTTON, V. Processing of noisy magnetotelluric data using digital filters and additional data selection criteria. **Physics of the Earth and Planetary Interiors**, v. 52, n. 1, p. 30–40, 1988. ISSN 0031-9201. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0031920188900556>>. 1

FRIGO, M.; JOHNSON, S. The design and implementation of fftw3. **Proceedings of the IEEE**, v. 93, n. 2, p. 216–231, 2005. ISSN 0018-9219. Disponível em: <<https://doi.org/10.1109/jproc.2004.840301>>. 137

GAMBLE, T. D.; GOUBAU, W. M.; CLARKE, J. Magnetotellurics with a remote magnetic reference. **Geophysics**, v. 44, n. 1, p. 53–68, jan. 1979. Disponível em: <<http://link.aip.org/link/?GPY/44/53/1>>. 2, 20, 55, 63

GARCIA, X.; JONES, A. G. Robust processing of magnetotelluric data in the amt dead band using the continuous wavelet transform. **Geophysics**, v. 73, n. 6, p. F223–F234, dez. 2008. Disponível em: <<http://geophysics.geoscienceworld.org/content/73/6/F223.abstract>>. 2

GIL, Y.; DAVID, C. H.; DEMIR, I.; ESSAWY, B. T.; FULWEILER, R. W.; GOODALL, J. L.; KARLSTROM, L.; LEE, H.; MILLS, H. J.; OH, J.-H.; PIERCE, S. A.; POPE, A.; TZENG, M. W.; VILLAMIZAR, S. R.; YU, X. Toward the geoscience paper of the future: best practices for documenting and sharing research from data to software to provenance. **Earth and Space Science**, v. 3, n. 10, p. 388–415, 2016. Disponível em: <<https://doi.org/10.1002/2015ea000136>>. 136

GOUBAU, W. M.; GAMBLE, T. D.; CLARKE, J. Magnetotelluric data analysis: removal of bias. **Geophysics**, v. 43, n. 6, p. 1157–1166, out. 1978. Disponível em: <<http://geophysics.geoscienceworld.org/content/43/6/1157.abstract>>. 2

GOUGH, B. **GNU scientific library reference manual**. 3. ed. [S.l.]: Network Theory, 2009. ISBN 0954612078, 9780954612078. 119

GROOM, R. W.; KURTZ, R. D.; JONES, A. G.; BOERNER, D. E. A quantitative methodology to extract regional magnetotelluric impedances and determine the dimension of the conductivity structure. **Geophysical Journal International**, v. 115, n. 3, p. 1095–1118, 1993. Disponível em: <<https://doi.org/10.1111/j.1365-246x.1993.tb01512.x>>. 14, 36, 89

GUENNEBAUD, G. et al. **Eigen v3**. 2010. Disponível em: <<http://eigen.tuxfamily.org>>. 137

HEISE, W.; CALDWELL, T. G.; BIBBY, H. M.; BENNIE, S. L. Three-dimensional electrical resistivity image of magma beneath an active continental rift, Taupo volcanic zone, New Zealand. **Geophysical Research Letters**, v. 37, n. 10, p. L10301, maio 2010. ISSN 0094-8276. Disponível em: <<https://doi.org/10.1029/2010gl043110>>. 90

HORVATH, K. A. **Estruturas geoeletricas do embasamento sob a porção central da bacia do Paraná a partir de um perfil magnetotelúrico**. Dissertação (Mestrado em Geofísica Espacial) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2017. 23

HÖSLER, C. **Rwantshue: R adaptor for the "I want hue" color palette generator**. [S.l.: s.n.], 2019. R package version 0.0.3. 137

INCE, D. C.; HATTON, L.; GRAHAM-CUMMING, J. The case for open computer programs. **Nature**, v. 482, n. 7386, p. 485–488, fev. 2012. ISSN 0028-0836. Disponível em: <<http://www.nature.com/nature/journal/v482/n7386/full/nature10836.html>>. 121

JAIN, A. K. Data clustering: 50 years beyond K-means. **Pattern Recognition Letters**, v. 31, n. 8, p. 651–666, 2010. ISSN 0167-8655. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167865509002323>>. 69

JONES, A. G. Distortion of magnetotelluric data: its identification and removal. In: CHAVE, A. D.; JONES, A. G. (Ed.). **The magnetotelluric method: theory and practice**. [S.l.]: Cambridge University Press, 2012. ISBN 9780521819275. 89

JONES, A. G.; CHAVE, A. D.; EGBERT, G. D.; AULD, D.; BAHR, K. A comparison of techniques for magnetotelluric response function estimation. **Journal of Geophysical Research: Solid Earth**, v. 94, p. 14,201–14,213, 1989. Disponível em: <<https://doi.org/10.1029/JB094iB10p14201>>. 17

JONES, A. G.; JÖDICKE, H. Magnetotelluric transfer function estimation improvement by a coherence-based rejection technique. In: **SEG ANNUAL MEETING, 1984, Proceedings...** Atlanta, Georgia: Society of Exploration Geophysicists, 1984. p. 51–55. Disponível em: <<http://homepages.dias.ie/~ajones/papers/publications.html>>. 2

JOPPA, L. N.; MCINERNY, G.; HARPER, R.; SALIDO, L.; TAKEDA, K.; O'HARA, K.; GAVAGHAN, D.; EMMOTT, S. Troubling trends in scientific software use.

- Science**, v. 340, n. 6134, p. 814–815, 2013. Disponível em:
<<http://dx.doi.org/10.1126/science.1231535>>. 122
- KNUTH, D. E. Literate programming. **The Computer Journal**, v. 27, n. 2, p. 97–111, 1984. Disponível em:
<<https://doi.org/10.1093/comjnl/27.2.97>>. 133
- LARSEN, J. C.; MACKIE, R. L.; MANZELLA, A.; FIORDELISI, A.; RIEVEN, S. Robust smooth magnetotelluric transfer functions. **Geophysical Journal International**, v. 124, n. 3, p. 801–819, 1996. Disponível em:
<<http://dx.doi.org/10.1111/j.1365-246X.1996.tb05639.x>>. 2, 17
- LEWIS, B. et al. **GNU emacs lisp reference manual**. [S.l.]: Free Software Foundation, 2019. 130
- MADDEN, T.; NELSON, P. **A defense of Cagniard's magnetotelluric method**. Cambridge, USA: [s.n.], 1964. 5
- MARCUELLO, A.; QUERALT, P.; LEDO, J. Applications of dispersion relations to the geomagnetic transfer function. **Physics of the Earth and Planetary Interiors**, v. 150, n. 1–3, p. 85–91, maio 2005. ISSN 0031-9201. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0031920104003644>>. 34
- MARONNA, R. A.; MARTIN, R. D.; YOHAI, V. J. **Robust statistics: theory and methods**. John Wiley & Sons, 2006. (Wiley Series in Probability and Statistics). ISBN 9780470010921, 9780470010945. Disponível em:
<<http://onlinelibrary.wiley.com/book/10.1002/0470010940>>. 2
- MCGILL, R.; TUKEY, J. W.; LARSEN, W. A. Variations of box plots. **The American Statistician**, v. 32, n. 1, p. 12, 1978. Disponível em:
<<https://doi.org/10.2307/2683468>>. 71
- MERALI, Z. Computational science: ...error. **Nature**, v. 467, n. 7317, p. 775–777, 2010. Disponível em: <<https://doi.org/10.1038/467775a>>. 121, 122, 123
- MIENSOPUST, M. P. Application of 3-D electromagnetic inversion in practice: challenges, pitfalls and solution approaches. **Surveys in Geophysics**, v. 38, n. 5, p. 869–933, 2017. Disponível em:
<<https://doi.org/10.1007/s10712-017-9435-1>>. 89, 90, 92
- MILLER, G. A scientist's nightmare: software problem leads to five retractions. **Science**, v. 314, n. 5807, p. 1856–1857, 2006. Disponível em:
<<http://dx.doi.org/10.1126/science.314.5807.1856>>. 122

NATURE. Reality check on reproducibility. **Nature**, v. 533, n. 7604, p. 437–437, 2016. Disponível em: <<http://dx.doi.org/10.1038/533437a>>. 121

OETTINGER, G.; HAAK, V.; LARSEN, J. C. Noise reduction in magnetotelluric time-series with a new signal-noise separation method and its application to a field experiment in the saxonian granulite massif. **Geophysical Journal International**, v. 146, n. 3, p. 659–669, 2001. ISSN 1365-246X. Disponível em: <<http://dx.doi.org/10.1046/j.1365-246X.2001.00473.x>>. 2

OLSEN, N. Long-period (30 days-1 year) electromagnetic sounding and the electrical conductivity of the lower mantle beneath europe. **Geophysical Journal International**, v. 138, n. 1, p. 179–187, 1999. Disponível em: <<https://doi.org/10.1046/j.1365-246x.1999.00854.x>>. 93

PADILHA, A. L.; VITORELLO, I.; PADUA, M. B.; FUCK, R. A. Cryptic signatures of neoproterozoic accretionary events in northeast Brazil imaged by magnetotellurics: implications for the assembly of west gondwana. **Tectonophysics**, v. 699, p. 164–177, 2017. Disponível em: <<http://dx.doi.org/10.1016/j.tecto.2017.01.022>>. 93, 98

PÁDUA, M. B. de. **Cálculo do tensor magnetotelúrico resistente a valores anômalos e pontos de influência**. Dissertação (Mestrado em Geofísica Espacial) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2014. 18, 32, 102

PARKER, R. L.; BOOKER, J. R. Optimal one-dimensional inversion and bounding of magnetotelluric apparent resistivity and phase measurements. **Physics of the Earth and Planetary Interiors**, v. 98, n. 3, p. 269–282, 1996. Disponível em: <[https://doi.org/10.1016/S0031-9201\(96\)03191-3](https://doi.org/10.1016/S0031-9201(96)03191-3)>. 3, 31, 61, 63

PARKINSON, W. D. Directions of rapid geomagnetic fluctuations. **Geophysical Journal International**, v. 2, n. 1, p. 1–14, 1959. Disponível em: <<https://doi.org/10.1111/j.1365-246x.1959.tb05776.x>>. 15

PATRO, P. K.; EGBERT, G. D. Application of 3D inversion to magnetotelluric profile data from the Deccan volcanic province of western India. **Physics of the Earth and Planetary Interiors**, v. 187, n. 1-2, p. 33–46, 2011. Disponível em: <<https://doi.org/10.1016/j.pepi.2011.04.005>>. 90

PENG, R. D. Reproducible research in computational science. **Science**, v. 334, n. 6060, p. 1226–1227, 2011. ISSN 0036-8075. Disponível em: <<http://science.sciencemag.org/content/334/6060/1226>>. 123

PERCIVAL, D. B.; WALDEN, A. T. **Spectral analysis for physical applications: multitaper and conventional univariate techniques**. [S.l.]: Cambridge University Press, 1993. ISBN 0 521 35532 x. 127

PICCOLO, S. R.; FRAMPTON, M. B. Tools and techniques for computational reproducibility. **GigaScience**, v. 5, n. 1, p. 30, 2016. Disponível em: <<https://doi.org/10.1186/s13742-016-0135-4>>. 136

PIESSENS, R.; DONCKER-KAPENGA, E. de; ÜBERHUBER, C. W.; KAHANER, D. K. **QUADPACK: a subroutine package for automatic integration**. Berlin: Springer-Verlag, 1983. ISBN 9783540125532. Disponível em: <<https://doi.org/10.1007/978-3-642-61786-7>>. 119

PRICE, A. T. The theory of magnetotelluric methods when the source field is considered. **Journal of Geophysical Research**, v. 67, n. 5, p. 1907–1918, 1962. Disponível em: <<https://doi.org/10.1029/jz067i005p01907>>. 5

QIAN, W.; PEDERSEN, L. B. Industrial interference magnetotellurics: an example from the Tangshan area, China. **Geophysics**, v. 56, n. 2, p. 265–273, 1991. Disponível em: <<http://dx.doi.org/10.1190/1.1443039>>. 25

R CORE TEAM. **R: a language and environment for statistical computing**. Vienna, Austria, 2019. Disponível em: <<https://www.R-project.org/>>. 125, 137

RAMEY, C.; FOX, B. **Bash reference manual**. [S.l.], 2013. Version 4.3.42(1). 124

RAMSEY, N. Literate programming simplified. **IEEE Software**, v. 11, n. 5, p. 97–105, 1994. Disponível em: <<https://doi.org/10.1109/52.311070>>. 133

RITTER, O.; JUNGE, A.; DAWES, G. New equipment and processing for magnetotelluric remote reference observations. **Geophysical Journal International**, v. 132, n. 3, p. 535–548, 1998. Disponível em: <<http://dx.doi.org/10.1046/j.1365-246X.1998.00440.x>>. 2

ROCHA, N. S.; FONTES, S. L.; LA-TERRA, E. F.; FUCK, R. A. Lithosphere structures of the Parnaíba basin and adjacent provinces revealed by deep magnetotelluric imaging. **Journal of South American Earth Sciences**, v. 92, p. 1–11, 2019. Disponível em: <<https://doi.org/10.1016/j.jsames.2019.02.020>>. 90

ROUSSEEUW, P. J. Least median of squares regression. **Journal of the American Statistical Association**, v. 79, n. 388, p. 871–880, dez. 1984. ISSN 01621459.

Disponível em:

<<https://doi.org/10.1080/01621459.1984.10477105>>. 65

ROUSSEEUW, P. J.; LEROY, A. M. **Robust regression and outlier detection**. New York: John Wiley & Sons, 2003. (Wiley series in probability and mathematical statistics). ISBN 0-471-48855-0. Disponível em:

<<http://onlinelibrary.wiley.com/book/10.1002/0471725382>>. 2

SANDBERG, S. K.; HOHMANN, G. W. Controlled-source audiomagnetotellurics in geothermal exploration. **Geophysics**, v. 47, n. 1, p. 100–116, 1982. Disponível em: <<http://dx.doi.org/10.1190/1.1441272>>. 28

SANDERSON, C.; CURTIN, R. Armadillo: a template-based c++ library for linear algebra. **The Journal of Open Source Software**, v. 1, n. 2, p. 26, 2016. Disponível em: <<https://doi.org/10.21105/joss.00026>>. 136

_____. A user-friendly hybrid sparse matrix class in c++. In: _____. **Mathematical software - ICMS 2018**. Berlin: Springer International Publishing, 2018.

(Mathematical Software - ICMS 2018), p. 422–430. Disponível em:

<https://doi.org/10.1007/978-3-319-96418-8_50>. 136

SANDVE, G. K.; NEKRUTENKO, A.; TAYLOR, J.; HOVIG, E. Ten simple rules for reproducible computational research. **PLoS Computational Biology**, v. 9, n. 10, p. e1003285, 2013. Disponível em:

<<http://dx.doi.org/10.1371/journal.pcbi.1003285>>. 122, 123

SCHULTE, E.; DAVISON, D. Active documents with org-mode. **Computer Science and Engineering**, v. 13, n. 3, p. 66–73, 2011. Disponível em:

<<http://dx.doi.org/10.1109/MCSE.2011.41>>. 124

SIMPSON, F.; BAHR, K. **Practical magnetotellurics**. Cambridge: Cambridge University Press, 2005. ISBN 0-521-81727-7. 1, 6, 18, 19, 20

SIMS, W. E.; BOSTICK, J. F. X.; SMITH, H. W. The estimation of magnetotelluric impedance tensor elements from measured data. **Geophysics**, v. 36, n. 5, p. 938–942, 1971. Disponível em:

<<http://link.aip.org/link/?GPY/36/938/1>>. 1, 19

SINGH, R. K.; MAURYA, V. P.; SHALIVAHAN; SINGH, S. Imaging regional geology and Au-sulphide mineralization over Dhanjori Greenstone belt:

implications from 3-D inversion of audio magnetotelluric data and petrophysical characterization. **Ore Geology Reviews**, v. 106, p. 369–386, 2019. Disponível em: <<https://doi.org/10.1016/j.oregeorev.2019.01.027>>. 90

SIRIPUNVARAPORN, W. Three-dimensional magnetotelluric inversion: an introductory guide for developers and users. **Surveys in Geophysics**, v. 33, n. 1, p. 5–27, 2012. ISSN 0169-3298. Disponível em: <<https://doi.org/10.1007/s10712-011-9122-6>>. 91

SIRIPUNVARAPORN, W.; EGBERT, G.; LENBURY, Y.; UYESHIMA, M. Three-dimensional magnetotelluric inversion: data-space method. **Physics of the Earth and Planetary Interiors**, v. 150, n. 1-3, p. 3–14, maio 2005. ISSN 0031-9201. Disponível em: <<https://doi.org/10.1016/j.pepi.2004.08.023>>. 89

SMITH, J. T. Understanding telluric distortion matrices. **Geophysical Journal International**, v. 122, n. 1, p. 219–226, 1995. Disponível em: <<http://dx.doi.org/10.1111/j.1365-246X.1995.tb03549.x>>. 83

SOERGEL, D. A. W. Rampant software errors may undermine scientific results. **F1000Research**, 2015. Disponível em: <<http://dx.doi.org/10.12688/f1000research.5930.2>>. 122

SUTARNO, D. Constrained robust estimation of magnetotelluric impedance functions based on a bounded-influence regression M-estimator and the Hilbert transform. **Nonlinear Processes in Geophysics**, v. 15, n. 2, p. 287–293, mar. 2008. ISSN 1023-5809. Disponível em: <<http://www.nonlin-processes-geophys.net/15/287/2008/>>. 1, 2

SUTARNO, D.; VOZOFF, K. Robust M-estimation of magnetotelluric impedance tensors. **Exploration Geophysics**, v. 20, n. 3, p. 383–398, jan. 1989. Disponível em: <<http://www.publish.csiro.au/paper/EG989383>>. 1

TELFORD, W. M.; GELDART, L. P.; SHERIFF, R. E. **Applied geophysics**. 2. ed. [S.l.]: Cambridge University Press, 1990. 6

TIETZE, K.; RITTER, O. Three-dimensional magnetotelluric inversion in practice — the electrical conductivity structure of the San Andreas fault in Central California. **Geophysical Journal International**, v. 195, n. 1, p. 130–147, 2013. Disponível em: <<https://doi.org/10.1093/gji/ggt234>>. 90

TIKHONOV, A. N. On determining electrical characteristics of the deep layers of the Earth's crust. **Doklady**, v. 73, p. 295–297, 1950. Disponível em: <<http://>>

[//mtnet.dias.ie/papers/ClassicPapers/classic_papers.html](http://mtnet.dias.ie/papers/ClassicPapers/classic_papers.html)>. 5, 10

TRIAHADINI, A.; AIZAWA, K.; TEGURI, Y.; KOYAMA, T.; TSUKAMOTO, K.; MURAMATSU, D.; CHIBA, K.; UYESHIMA, M. Magnetotelluric transect of Unzen Graben, Japan: conductors associated with normal faults. **Earth, Planets and Space**, v. 71, n. 1, p. 28, 2019. Disponível em: <<https://doi.org/10.1186/s40623-019-1004-z>>. 90

VOZOFF, K. The magnetotelluric method. In: NABIGHIAN, M. N. (Ed.). **Electromagnetic methods in applied geophysics**. Tulsa, OK: Society of Exploration Geophysicists, 1991. ISBN 9781560802686. Disponível em: <<http://dx.doi.org/10.1190/1.9781560802686.ch8>>. 6, 18, 19

WAIT, J. R. On the relation between telluric currents and the Earth's magnetic field. **Geophysics**, v. 19, n. 2, p. 281–289, 1954. Disponível em: <<https://doi.org/10.1190/1.1437994>>. 5, 11

WEIDELT, P. The inverse problem of geomagnetic induction. **Zeitschrift für Geophysik**, v. 38, p. 257–289, 1972. 18, 31

WICKHAM, H. **ggplot2: elegant graphics for data analysis**. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. Disponível em: <<https://ggplot2.tidyverse.org>>. 125

WICKHAM, H.; FRANÇOIS, R.; HENRY, L.; MÜLLER, K. **dplyr: a grammar of data manipulation**. [s.n.], 2019. Disponível em: <<http://dplyr.tidyverse.org>, <https://github.com/tidyverse/dplyr>>. 134

WICKHAM, H.; HENRY, L. **tidyr: easily tidy data with 'spread()' and 'gather()' functions**. [s.n.], 2019. Disponível em: <<http://tidyr.tidyverse.org>, <https://github.com/tidyverse/tidyr>>. 125

WICKHAM, H.; HESTER, J.; FRANÇOIS, R. **readr: read rectangular text data**. [s.n.], 2019. Disponível em: <<http://readr.tidyverse.org>, <https://github.com/tidyverse/readr>>. 134

WILSON, G.; ARULIAH, D. A.; BROWN, C. T.; HONG, N. P. C.; DAVIS, M.; GUY, R. T.; HADDOCK, S. H. D.; HUFF, K. D.; MITCHELL, I. M.; PLUMBLEY, M. D.; WAUGH, B.; WHITE, E. P.; WILSON, P. Best practices for scientific computing.

PLoS Biology, v. 12, n. 1, p. e1001745, 2014. Disponível em:

<<http://dx.doi.org/10.1371/journal.pbio.1001745>>. 122, 123

YLÖNEN, T. SSH–secure login connections over the internet. In: USENIX.

USENIX SECURITY SYMPOSIUM, 6., 1996. San Jose, California, 1996. 130

ZORIN, N.; ALEKSEEV, D. Causality and dispersion relations in electrical prospecting. **Russian Geology and Geophysics**, v. 59, n. 3, p. 313–323, 2018.

Disponível em: <<https://doi.org/10.1016/j.rgg.2018.03.009>>. 8, 31

APÊNDICE A – BIBLIOTECAS EXTERNAS E PARÂMETROS UTILIZADOS PELAS ROTINAS

A obtenção de uma estimativa válida para cálculo da transformada de Kramers-Kronig é parte essencial deste trabalho. Para obter esses resultados numericamente, foi utilizada a função `gsl_integration_qawc` em linguagem C da biblioteca GSL (GOUGH, 2009), que, por sua vez, reimplementa a função `QAWC` da biblioteca QUADPACK (PIESSENS et al., 1983) em linguagem Fortran. Essas funções calculam o valor principal da integral de uma função f no intervalo (a, b) com uma singularidade em c , definida da seguinte forma:

$$I = \int_a^b \frac{f(x)}{x - c} dx$$

Os parâmetros da função que calculam os limites a e b da integral são, eventualmente, fontes de instabilidade do programa. Quando eles não são bem escolhidos, a função `gsl_integration_qawc` retorna um erro e não resolve a integral. Idealmente, os parâmetros deveriam ser $a = 0$ e $b = \infty$. A solução mais simples seria adotar $a = 0$ e $b = \text{std::numeric_limits}<\text{double}>::\text{max}() \approx 1.8 \times 10^{+308}$ que é o maior valor possível para um tipo `double` no C++. No entanto, essa abordagem geralmente faz com que a função retorne um erro. O procedimento adotado foi atribuir $c = \omega_0$, $b = c + l$ e $a = \max(z, c - l)$. Dessa forma, $b - a = 2l$ para $c \geq l$. Tendo-se l na forma $l = 10^n$, onde $n = 22, \dots, 11$, a integral não é mais calculada até o infinito como esperado pelas Fórmulas 4.3a e 4.3b, mas limitada em b . Além disso, essas fórmulas admitem $z = 0$, mas esse valor eventualmente também retorna erro, principalmente no caso da Equação 4.3a. Quando isso ocorria foi utilizado $z \approx 2,225\,073\,858\,507\,201\,4 \times 10^{-308}$ ou $z \approx 2,225\,073\,858\,507\,201\,9 \times 10^{-308}$, que são os dois `doubles` positivos mais próximos de zero no C++. Ainda assim, em alguns casos (como na Figura 4.4) chegou-se a utilizar $z = 1 \times 10^{-250}$ e 1×10^{-212} .

APÊNDICE B – PESQUISA REPRODUZÍVEL

Para apresentar o resultado deste trabalho, optou-se por implementar o procedimento de pesquisa reproduzível. Esse tipo de procedimento objetiva melhorar a capacidade de reprodução da pesquisa científica. Em um recente editorial, a revista *Nature* descreve uma pesquisa de opinião onde 2/3 dos participantes descreveram o nível atual da reprodução de resultados acadêmicos como um grande problema (NATURE, 2016).

A *National Science Foundation* define reprodutibilidade como a capacidade de um pesquisador de duplicar os resultados de um estudo anterior utilizando os mesmos materiais e procedimentos utilizados pelo pesquisador original. Embora essa descrição faça parecer que a pesquisa reproduzível é uma atividade simples, a prática mostra que a quantidade de ferramentas computacionais utilizadas atualmente tornou a pesquisa reproduzível uma atividade complicada. A quantidade de *scripts* e programas necessários para coletar, preparar, limpar e processar os dados para gerar apenas uma figura exige uma familiaridade com ferramentas de automação além do conhecimento de boa parte dos pesquisadores (MERALI, 2010), e é motivo de preocupação com a *crise de reprodutibilidade* em ciências espaciais (BURRELL et al., 2018). Fatores humanos, como a mobilidade de pós-doutores e a aposentadoria de pesquisadores, também prejudicam a reprodutibilidade (BOND-LAMBERTY, 2018). Como os trabalhos normalmente descrevem de forma muito pobre (ou simplesmente não descrevem) o local onde os dados foram armazenados e o procedimento detalhado para montar os programas e obter os resultados, a informação necessária para se recuperar projetos já desenvolvidos são perdidos. Neste apêndice irá-se descrever o ambiente computacional que foi utilizado para que a tese desenvolvida estivesse dentro dos critérios de reprodutibilidade exigidos atualmente.

Existem vários guias de boas práticas para desenvolvimento de código e de reprodutibilidade. Um código bem escrito é como um artigo bem escrito, facilita a adoção e adaptação para uso em outros projetos (BURRELL et al., 2018).

- a) **Uso (quase) exclusivo de código aberto.** Para que um pesquisador independente (ou membro da banca) consiga reproduzir os resultados é necessário que ele tenha acesso a todos os códigos e dados necessários. Por essa razão, este trabalho utiliza apenas código e ferramenta de código aberto, como recomendado por Ince et al. (2012). Todo o código

desenvolvido pelo autor é feito em código aberto, mas neste trabalho também foi utilizado o código desenvolvido por [Egbert e Booker \(1986\)](#), com pequenas modificações que permitiram que o programa processasse séries temporais mais longas. Apesar do código original ter distribuição livre, a sua modificação foi permitida pelo autor (Gary Egbert) apenas para uso interno do INPE e não pode ser distribuída para terceiros.

- b) **Controle de versão.** Tanto o código do programa utilizado na tese quanto este texto estão armazenados sob controle de versão como recomendado por [Merali \(2010\)](#), [Wilson et al. \(2014\)](#), [Sandve et al. \(2013\)](#). O uso de controle de versão não só garante um modo padronizado de distribuição, mas também um modo de acompanhamento passo-a-passo tanto do programa quanto do relatório. O programa utilizado para o controle de versão é o `git` (<https://git-scm.com/about>).
- c) O processo de gerar os programas em código fonte em linguagem de máquina é uma tarefa tediosa. Por essa razão foi adotado um processo de montagem automática dos programas como recomendado por [Wilson et al. \(2014\)](#). A ferramenta de montagem utilizada para os programas em `c++` é o `cmake` (<https://cmake.org/>).
- d) **Teste automático do código.** Essa etapa tecnicamente não faz parte do processo de análise reproduzível. Mas, uma vez que o número de retratações em artigos científicos devido a erros básicos de programação tem aumentado ([SOERGEL, 2015](#); [WILSON et al., 2014](#); [JOPPA et al., 2013](#); [MERALI, 2010](#); [CHANG et al., 2006](#); [MILLER, 2006](#)), vale a pena utilizar essas ferramentas. Para o teste do código em `c++` é utilizado o `googletest` (<https://github.com/google/googletest>) em conjunto com o `cmake`.
- e) **Documentação do código.** Se um pesquisador independente ficar com dúvidas sobre algum resultado, uma das suas primeiras iniciativas será analisar o código. Um código bem documentado facilita a análise e ferramentas de documentação de código aberta estão disponíveis. ([MERALI, 2010](#); [WILSON et al., 2014](#)). A documentação do código é feita utilizando o programa `doxygen` (<http://www.doxygen.org>) em conjunto com o `cmake`.
- f) **Formatação e estilo consistente.** Mais uma vez, códigos bem formatados facilitam a leitura, o que diminui o tempo de encontro e correção de

falhas (WILSON et al., 2014). A ferramenta utilizada para a formatação automática do código é o `clang-format` (<http://clang.llvm.org/docs/ClangFormat.html>).

g) Registro dos passos de execução dos programas (MERALI, 2010; SANDVE et al., 2013). O modo de fazer isso automaticamente é descrito na seção **Implementação da análise reproduzível**

h) Substituição de processos manuais por registros em programas, *scripts* ou arquivos de configuração.

O Item h dificilmente é armazenado e é um grande entrave para a reprodução de resultados científicos. Essa etapa do processamento de dados normalmente é registrada em processos MT com frases semelhantes a “*As curvas de fase e resistividade aparente resultantes são editadas para eliminar dados ruidosos*”, como encontrado em (CAMPOS-ENRIQUEZ et al., 2013). Expressões como “*dados ruidosos*” são subjetivas e dependem da experiência e da interpretação pessoal de quem realiza o processamento. Como raramente duas pessoas terão os mesmos critérios, acaba-se por obter curvas diferentes para os mesmos dados de entrada. Nesse caso, um arquivo com os períodos e componentes dos pontos que são considerados ruidosos e um *script* para remover esses pontos do arquivo original bastam para tornar o processo reproduzível.

B.1 Implementação da análise reproduzível

Para uma pesquisa ser reproduzível é necessário que um pesquisador seja capaz de repetir todos os passos para obter o mesmo resultado que o autor original. Segundo Peng (2011), o padrão de reprodutibilidade consiste em reconhecer que todo experimento computacional tem, em teoria, um registro detalhado de toda ação tomada pelo computador. Tornando os códigos disponíveis para os outros permite um nível de detalhamento com relação à análise que é maior que as descrições não-computacionais impressas em revistas usando uma linguagem natural.

No entanto, vale a pena enfatizar que apenas enviar uma cópia de milhares de linhas de códigos fonte em várias linguagens de programação e *scripts* não bastam para que o pesquisador reproduza os resultados. É necessário que ele saiba a ordem em que todos os programas devem ser executados. Isso geralmente é feito enviando um arquivo de *log* que o pesquisador pode utilizar para copiar e colar em

um terminal de programas como *matlab*, *bash* ou *python* para gerar os resultados. Porém, tanto o processo de gerar o arquivo de *log* pelo autor original quando o de copiar esse arquivo para o terminal do pesquisador, envolvem atividades manuais e repetitivas, e como tais, tediosas e sujeitas a erro.

Para evitar esse processo, neste trabalho foi utilizada a linguagem `org-mode`. Essa linguagem consiste na inclusão de marcações em um arquivo texto que permitem a combinação de dados, código e prosa (SCHULTE; DAVISON, 2011).

Para exemplificar a linguagem, vamos mostrar dois blocos de códigos em `org-mode` utilizados para gerar o gráfico de um *taper dpss* (*discrete prolate spherical sequence*) que foi utilizado para gerar os espectros utilizados na tese. A descrição do *taper* não é importante agora, a intenção aqui é apenas mostrar como obtê-lo de maneira reproduzível e automática.

O bloco a seguir ilustra como configurar um bloco `org-mode`. A primeira linha fornece um nome para o bloco; as linhas marcadas com `#caption:` estabelecem a legenda do bloco e a linha iniciada por `#+header:` especifica que o conteúdo gerado pelo bloco deve ser redirecionado para o arquivo `dpss_time.dat`. Essas linhas são usadas para configurar parâmetros de execução e exibição dos resultados e apenas serão exibidas nos demais blocos deste trabalho quando introduzirem um conceito novo. O conteúdo entre `#+begin_src sh` e `#end_src` define que o *script* em `bash` (RAMEY; FOX, 2013) que deve ser executado para obter os resultados exibidos neste trabalho. No caso específico, o *script* executa o comando `dpss` que consiste em um programa que foi executado para gerar o *taper dpss*.

Bloco B.1 – Exemplo de bloco em `org-mode`

```
#+name: dpss-time
#+caption: linha de execução do código que gera os
#+caption: valores do /taper/ dpss com $n = 32$ pontos
#+caption: e semi-largura de banda \ (nw = 4\).
#+header: :file dpss_time.dat
#+begin_src sh :exports code :results silent file
, dpss 32 3 4
#+end_src
```

Quando o `org-mode` encontra o bloco acima ele executa duas funções. Primeiro, envia para o terminal `bash` o comando `dpss 32 3 4` e em seguida gera o código em `latex` que resulta no bloco a seguir.

Bloco B.2 – linha de execução do código que gera os valores do *taper* dpss com $n = 32$ pontos e semi-largura de banda $nw = 4$.

```
dpss 32 3 4
```

Uma vez gerados os dados do *taper*, é necessário gerar o gráfico. Para isso foi utilizado o seguinte bloco em `org-mode`.

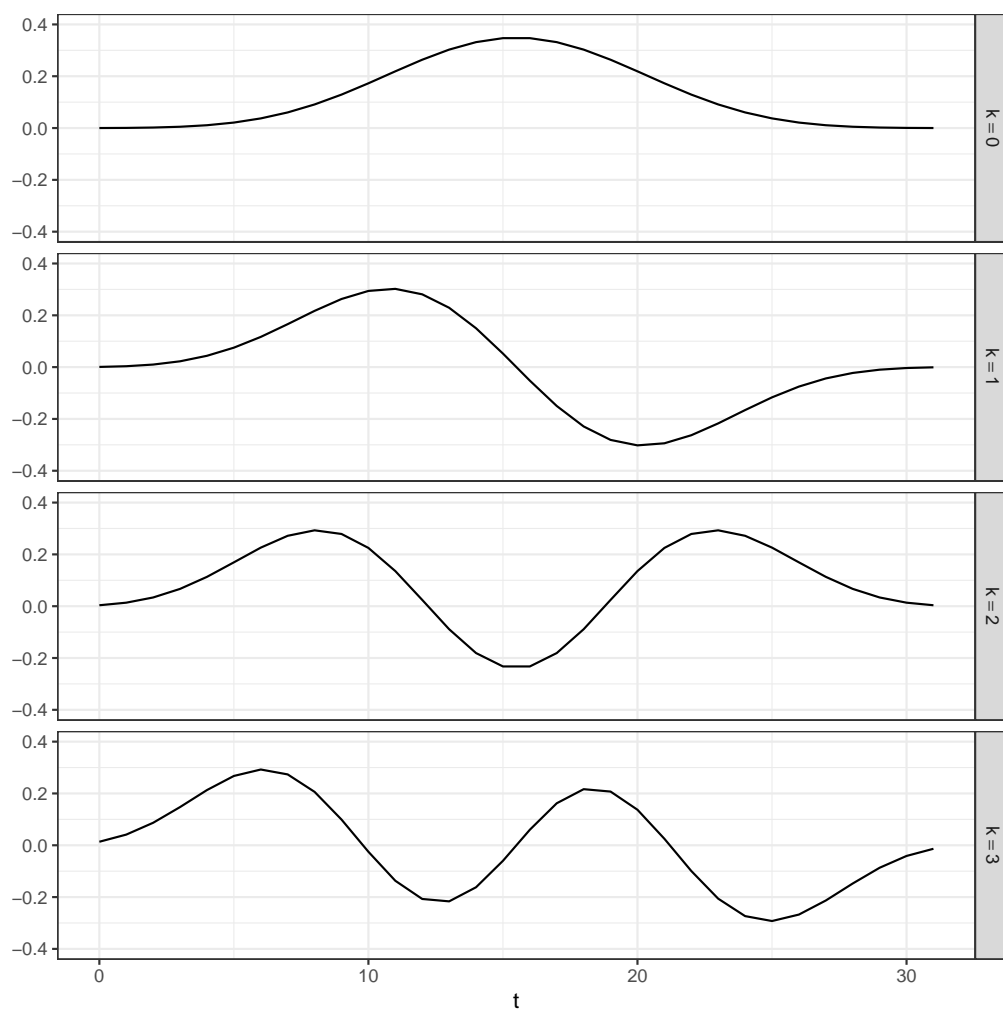
Bloco B.3 – Bloco em `org-mode` utilizado para gerar a Figura B.1.

```
#+name: dpss-time-graphic
#+headers: :results graphics
#+headers: :file dpss_time.png
#+begin_src R :session :exports results
raw <- read.table("dpss_time.dat", header=FALSE)
K_MAX <- ncol(raw) - 1
colnames(raw) <- 0:K_MAX
N <- nrow(raw); N <- nrow(raw);
library(tidyr)
long <- gather(data = raw, key = k)
t <- rep(0:(N - 1), K_MAX + 1)
long$t <- t
library(ggplot2)
ggplot() + geom_line(data = long,
                     aes(x = t, y = value)) +
  facet_grid(k ~ .,
            labeller=label_bquote(k == .(k))) +
  ylim(-0.4, 0.4) + theme_bw() +
  theme(axis.title.y = element_blank())
#+end_src
```

No Bloco B.3, o cabeçalho `:results graphics` estabelece que a saída do bloco é um arquivo gráfico, e o trecho entre `#+begin_src` e `#end_src` contém o código em linguagem R (R CORE TEAM, 2019) que irá gerar a Figura B.1. Nesta listagem foi utilizados os pacotes `tidyr` (WICKHAM; HENRY, 2019) para manipulação de dados e `ggplot2` (WICKHAM, 2016) para a geração de gráficos.

O responsável pela avaliação (um revisor de artigo ou membro da banca) poderia requisitar o registros das operações necessárias para reproduzir o resultado do trabalho. Tipicamente esse registro consiste de vários arquivos e *scripts* contendo os trechos entre as linhas `#+begin_src` e `#end_src` dos dois blocos apresentados. Utilizando análise reproduzível, todos o códigos e a ordem em que eles devem ser executados estão embebidos no texto.

Figura B.1 – As quatro primeiras sequencias DPSS com $n = 32$ elementos e semi-largura de banda $nw = 4$.



Uma vantagem do uso de análise reproduzível é a diminuição do risco de esquecer algum parâmetro utilizado para gerar uma figura ou um resultado. Para gerar a figura o autor não envia manualmente esses códigos para um ambiente `shell`, `R`, `Python` ou qualquer outro ambiente de processamento. A execução das rotinas de processamento de dados e geração de gráficos fica a cargo do `org-mode` durante o processo de produção do arquivo `pdf`. A chance de perder um trecho do arquivo de `log` é praticamente inexistente já que se os códigos apresentados não funcionarem, os gráficos não serão gerados. Se o leitor estiver vendo a Figura B.1 na Página 126 então o resultado deste trabalho é plenamente reproduzível. Caso contrário, os códigos apresentados não são capazes de gerar a figura.

Para ilustrar um exemplo em que a posse do código facilita a compreensão do resultado, vamos registrar o que ocorreu durante o desenvolvimento do código utilizado para calcular os espectros dos sinais eletromagnéticos. Para calcular esses espectros foi desenvolvido um pequeno programa `dpss_spectra` para gerar os auto-espectros de uma série temporal. Para validar os resultados desse programa foram calculados os auto-espectros do *taper dpss* da Figura B.1 para comparar com a imagem dos espectros contidos em (PERCIVAL; WALDEN, 1993, p. 107), o qual afirma apresentar uma ilustração desses auto-espectros. O resultado gerado pelo código desenvolvido está na Figura B.2.

A figura obtida não tem qualquer semelhança com o resultado esperado, o que gerou dúvida sobre a qualidade do código desenvolvido. Depois de muito tempo revisando o código do programa, o problema foi encontrado. Não no código, mas nos parâmetros de entrada. Percival e Walden (1993) esqueceram de informar que para obter a figura era necessário interpolar a sequência da Figura B.1 por meio de uma técnica conhecida como *zero padding*, a qual consiste em acrescentar uma sequência de zeros à série temporal depois de multiplicada pelo *taper*. Acrescentando um total de 2^{15} zeros, o resultado obtido pelo código desenvolvido está na Figura B.3.

A Figura B.3 apresenta os lóbulos esperados para o auto-espectro e está de acordo com a figura em (PERCIVAL; WALDEN, 1993). É evidente que o caso apresentado acima é trivial e acredito que a omissão do uso da interpolação foi porque para os autores era óbvio que *zero padding* devia ser utilizado para séries pequenas (apenas 32 pontos). Mas de qualquer forma, fica ilustrado como é comum a omissão de detalhes para a obtenção dos resultados e como a distribuição do código utilizado esclarece as técnicas utilizadas.

Figura B.2 – Auto-espectro das quatro primeiras sequencias DPSS da Figura B.1.

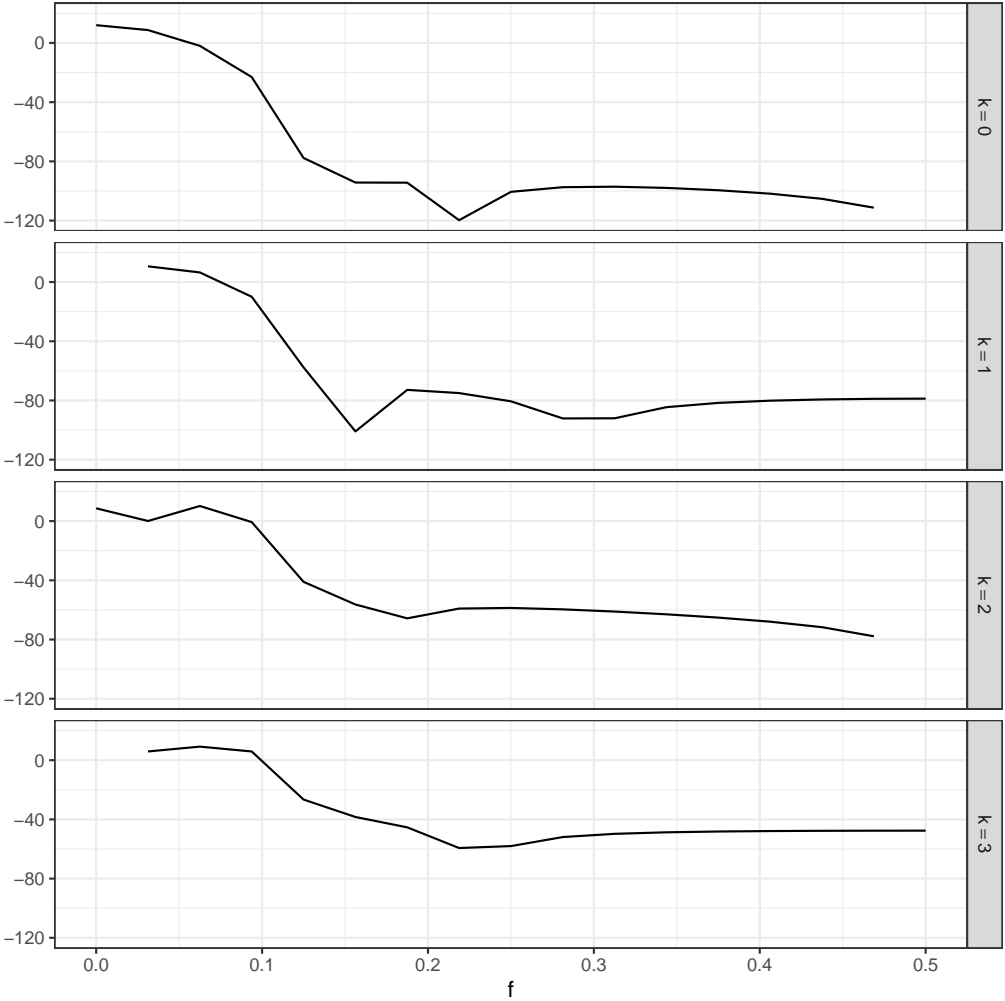
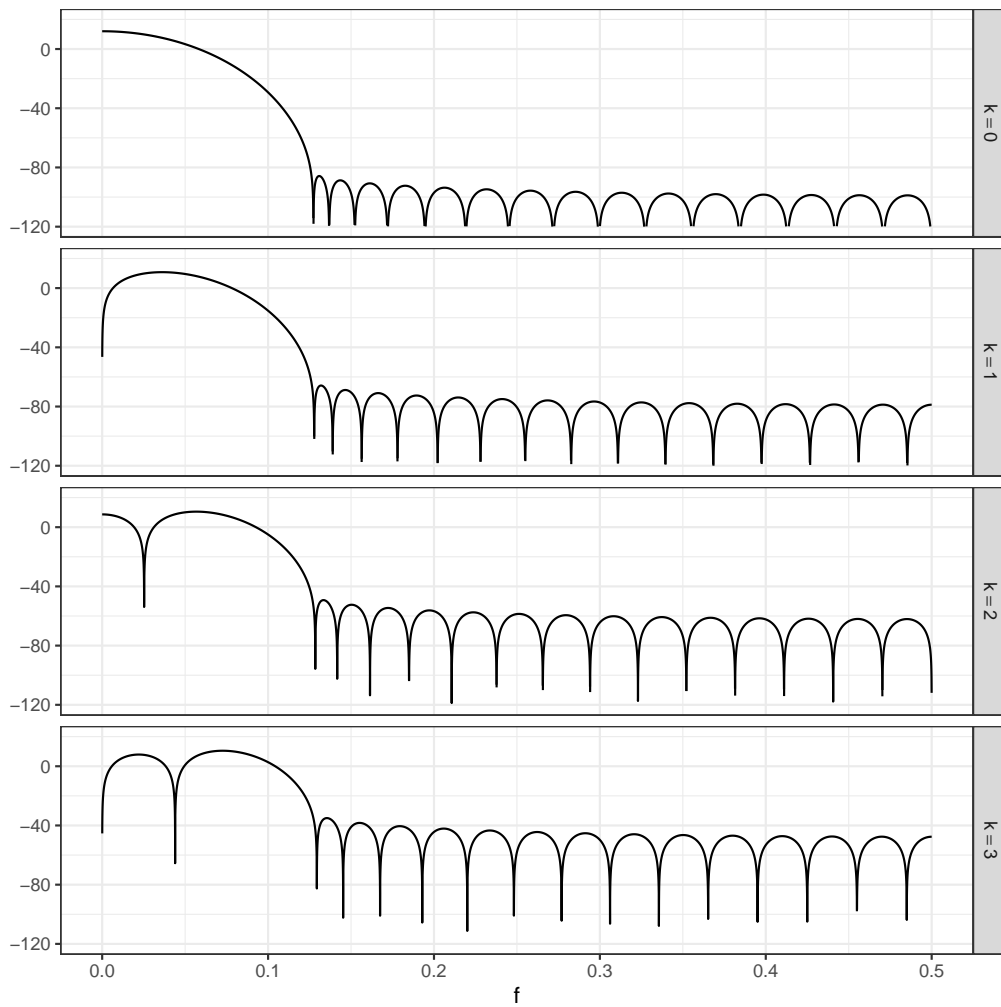


Figura B.3 – Auto-espectro das quatro primeiras sequencias DPSS da Figura B.2 interpolado utilizando 2^{15} zeros.



B.2 Análise de dados remota

Como discutido na Seção 5.2.3, o número de amostras geradas durante o processo de reamostragem atinge facilmente um número maior que os computadores de mesa atuais são capazes de armazenar na memória. Para contornar essa limitação, parte do processamento foi realizado em um servidor com 32 GB de memória RAM.

O uso de um servidor remoto para análise dos dados exige que os blocos em `org-mode`, que contenham o código responsável pelo processamento, sejam executados no servidor e não no computador onde o texto está sendo processado. Além disso, como não há garantias que o servidor esteja disponível no momento em que o texto está sendo processado, é necessário que haja mecanismos para configurar o acesso ao servidor.

O recurso que o `org-mode` fornece para executar código em um servidor remoto é o cabeçalho `#+header: :dir /method:user@host:path/`, onde `method` estabelece o método de acesso ao servidor; `user` é o nome de usuário utilizado para acessar o servidor (que pode ser omitido se for o mesmo do usuário processando o texto); `host` é o endereço ip do servidor; e `path` é o caminho do diretório no servidor em que o código deve ser executado. Dessa maneira, o cabeçalho `#+header: :dir /ssh:127.0.0.1:` faz com que o servidor no endereço `127.0.0.1` seja acessado com o mesmo nome de usuário do computador local, via `ssh` (YLÖNEN, 1996) e o código executado no diretório padrão do usuário.

O parâmetro `:dir` do cabeçalho permite executar o código em um servidor remoto, mas ainda é preciso que ele seja configurável. Pedir que o leitor troque o endereço de ip em todos os blocos é tedioso e sujeito a erros. Por essa razão, optou-se por fazer a configuração utilizando o recurso de propriedades do `org-mode`. Propriedades do `org-mode` são pares na forma chave-valor que podem ser configuradas no início do arquivo em uma linha na forma `#+PROPERTY: chave valor`. Neste trabalho, o nome da chave com o endereço de ip do servidor é `proc-server-ip`. Para recuperar esse valor é utilizado, nos cabeçalhos dos blocos que serão executados no servidor, um pequeno trecho de código em `lisp` (LEWIS et al., 2019): `:dir (format "/ssh:%s:" (org-entry-get nil "proc-server-ip" t))`. Assim o parâmetro `dir` do cabeçalho passa a ser configurável utilizando a chave `proc-server-ip`

Resta resolver ainda mais um problema. Se o leitor não tem acesso ao servidor e deixar a variável `proc-server-ip` vazia, o `org-mode` tentará executar o código

via `ssh` em um endereço de ip vazio, o que acarretará em erro, impedindo a produção do documento. Para evitar que isso ocorra, recorreu-se ao parâmetro `:eval` nos cabeçalhos dos blocos `org-mode`. Se parâmetro tiver o valor `yes`, ela executa o código, se o valor for `no`, ela não executa o código. Mais uma vez, para ajustar esse parâmetro automaticamente, foi utilizado o seguinte trecho de código em `lisp`: `eval (if (org-entry-get nil "proc-server-ip" t) "yes" "no")`. Desse modo o código é executado apenas se `proc-server-ip` estiver configurado.

A partir daqui, vamos explicitar os passos para obter as Figuras 5.5 e 5.6.

Para efetuar o processamento, o diretório `data_analysis/apg/apg023a` deve conter os arquivos `ats` com dados de campo. Os arquivos `ats` estão em formato binário, mas os programas desenvolvidos utilizam como entrada os arquivos textos em formato `asc` criados para executar o processamento robusto com os programas descritos em (EGBERT; BOOKER, 1986). O bloco a seguir converte os arquivos utilizando o programa `ats2asc` desenvolvido localmente.

Bloco B.4 – Converte os arquivos em formato binário para arquivos texto.

```
PROCDIR="${PROC_SERVER_DIR}/${CAMPANHA}"
if [[ ! -d "${PROCDIR}" ]]; then
    exit
fi
cd "${PROCDIR}"

# se já existir "DATA/${file}.asc" não é necessário
# executar o programa
if [[ ! -f "DATA/${file}.asc" ]]; then
    ats2asc --site-name "${ESTACAO}" "${ESTACAO}/gms06/"
fi
```

O código em `bash` do Bloco B.4 utiliza as variáveis `PROC_SERVER_DIR`, `CAMPANHA` e `file`. Essas variáveis foram configuradas no cabeçalho dessa seção, como mostra o Bloco B.5. A Linha 1 estabelece o título do cabeçalho e o número de `*` no começo dessa linha determina se o cabeçalho é um capítulo, seção ou subseção (no atual caso é uma seção desta tese). As Linhas 2 e 12 marcam o início e o fim de um bloco de propriedades dessa seção, entendendo-se que propriedades de seção são locais à seção e não são válidas nas demais seções. A propriedade `CUSTOM_ID` é utilizada para atribuir um *label* que é utilizado internamente para identificar essa seção. As propriedades `header-args` declaram os argumentos que serão utilizados em todos os blocos executados na seção.

Bloco B.5 – Cabeçalho em org-mode e configuração de propriedades de seção.

```
1 ** Análise de Dados Remota
2   :PROPERTIES:
3   :CUSTOM_ID: sec:analise.de.dados.remota
4   :header-args: :var CAMPANHA="apg" ESTACAO="apg023a"
5   :header-args+: :var PROC_SERVER_DIR="data_analysis"
6   :header-args+: :var max_asc_lines=314110
7   :header-args+: :var file = "apg023a059_06D"
8   :header-args+: :var subfile = "apg023a059_07D"
9   :header-args+: :var ts_lengths = "128 256 512 1024 2048"
10  :header-args+: :var quantity = "zyx"
11  :header-args+: :var calibration_archive="modelo.tar.xz"
12  :END:
```

Como foi afirmado no início deste Apêndice, os dados foram processados em um servidor externo com mais memória para permitir a criação de um conjunto de impedâncias maior. Ainda assim, o arquivo `apg023a059_06D` gerou um conjunto de impedâncias maior que o servidor é capaz de alocar na memória. Para resolver o problema foi criado o arquivo `apg023a059_07D` com as primeiras 314110 do arquivo original. As variáveis utilizadas nesse bloco foram configuradas como propriedades dessa seção como mostrado no Bloco B.5.

```
PROCDIR="${PROC_SERVER_DIR}/${CAMPANHA}"

head -n "${max_asc_lines}" "${PROCDIR}/DATA/${file}.asc" \
  > "${PROCDIR}/DATA/${subfile}.asc"

cp "${PROCDIR}/DATA/${file%.asc}.clk" \
  "${PROCDIR}/DATA/${subfile%.asc}.clk"
cp "${PROCDIR}/SP/${file%.asc}.sp" \
  "${PROCDIR}/SP/${subfile%.asc}.sp"

for s in $(grep TXT "${PROCDIR}/SP/${subfile%.asc}.sp" |\
  cut -f 1 -d ' '); do \
  cp -v "${PROCDIR}/sensors/$s" "${PROCDIR}";
done
```

Os espectros calculados precisam ser corrigidos pela curva de calibração do equipamento utilizado na coleta dos dados. A variável `PROC_SERVER_IP` não está entre as propriedades dessa seção como mostrado no Bloco B.5. Ela é definida no cabeçalho `:var PROC_SERVER_IP=(org-entry-get nil "proc-server-ip" t)` do próximo bloco em `bash`. Esse bloco copia o arquivo com os dados compactados das calibrações para o servidor.

```

PROCDIR="${PROC_SERVER_DIR}/"
rsync -a ${calibration_archive} \
      "${PROC_SERVER_IP}:${PROCDIR}"/

```

Depois de copiado, o arquivo com as calibrações deve ser expandido no servidor, o que é feito no próximo bloco.

```

if [[ ! -d "${PROC_SERVER_DIR}/${CAMPANHA}/sensors" ]]; then
    tar xvf "${PROC_SERVER_DIR}/modelo.tar.xz" \
        --transform "s,^modelo,${CAMPANHA}," \
        -C "${PROC_SERVER_DIR}"
fi

```

O bloco a seguir é usado para gerar o nome base para os arquivos com os resultados dos Blocos B.7, B.8 e B.9.

Bloco B.6 – Código em bash que gera o nome do arquivo (sem a extensão)

```

echo "${ESTACAO}_cloud_${quantity}"

```

Tabela B.1 – valores dos períodos utilizados no Bloco B.7

2.2857e+00
3.2000e+00
4.5714e+00
6.4000e+00
9.1429e+00
1.2800e+01
1.8286e+01

O Bloco B.7 em bash é o mais importante para os resultados apresentados na Seção 5.2.2 e merece ser discutido em mais detalhes. O código entre as Linhas 1 e 5 verifica se o diretório com os dados existe e muda para aquele diretório. A expressão «`results_filename()`» na Linha 7 é um recurso da programação literária (KNUTH, 1984), conhecido como `noweb` (RAMSEY, 1994). A sintaxe `noweb` é utilizada para injetar o resultado do Bloco B.6 no Bloco B.7. A Linha 9 verifica se o arquivo de saída existe antes de executar o processamento. Esse procedimento se mostrou importante nos blocos onde o processamento leva muito tempo, já que esperar os dados processarem antes de cada alteração de texto implicava um gasto de tempo significativo. O cuidado que se deve ter com esse tipo de checagem é que se houver atualizações nos programas que geram esse

arquivo, ela impede que o arquivo de saída seja atualizado. O procedimento atual é deixar a critério de quem está processando o texto apagar os arquivos temporários a cada atualização significativa dos programas de processamento de dados e geração de gráficos. A diferença entre o modo como as variáveis `ts_lengths` (Linha 10) e `periods` (Linha 11) são utilizadas se deve ao modo como elas são declaradas no arquivo `=org=`. A variável `ts_lengths` é declarada como um `string` no cabeçalho dessa seção, utilizando a propriedade do `org-mode` `:header-args+: :var ts_lengths = "128 256 512 1024 2048"`; já `periods` é declarado na Tabela B.1 o que faz com que o `org-mode` passe os valores para o `bash` como um `array`. O bloco encerra com a execução em um laço `for` do programa `kkmt` desenvolvido para calcular os resultados deste trabalho.

Bloco B.7 – código em `bash` que gera o arquivo `apg023a_cloud_zyx.csv`

```

1 PROCDIR="${PROC_SERVER_DIR}/${CAMPANHA}"
2 if [[ ! -d "${PROCDIR}" ]]; then
3     exit
4 fi
5 cd "${PROCDIR}"
6
7 OUTFILE="<<results_filename()>>.csv"
8
9 if [[ ! -f "${OUTFILE}" ]]; then
10     for L in ${ts_lengths}; do
11         for T in "${periods[@]}"; do
12             kkmt --cloud.quantities "${quantity}" \
13                 --cloud.file "${OUTFILE}" \
14                 --data.asc.files "${subfile}" \
15                 --spectral.periods "${T}" \
16                 --spectral.lengths.d "${L}" \
17                 --spectral.lengths.c "${L}"
18         done
19     done
20 fi

```

O Bloco B.8 exibe o código em linguagem `R` que gera o gráfico exibido na Figura 5.5. A técnica `noweb` descrita no Bloco B.7 é utilizada nas Linhas 4 e 11. A leitura dos arquivos é feita utilizando funções do pacote `readr` (WICKHAM et al., 2019) e o pacote `dplyr` (WICKHAM et al., 2019) é utilizado para manipulação dos dados. Para poupar tempo de processamento, a verificação da Linha 7 abandona o código caso o arquivo da figura já exista.

Bloco B.8 – código em R que gera a Figura 5.5.

```
1 procdir <- paste(PROC_SERVER_DIR, CAMPANHA, sep = "/")
2 setwd(procdir)
3
4 plotfile <- paste("<<results_filename()>>", ".pdf",
5                 sep = "")
6
7 if (file.exists(plotfile)) {
8   q()
9 }
10
11 cloudfile <- paste("<<results_filename()>>", ".csv",
12                  sep="")
13
14 if (! file.exists(cloudfile)) {
15   q()
16 }
17
18 library(readr)
19 cloud.zyx <-
20   read_tsv(file=cloudfile,
21            col_types = cols(
22              target_period = col_factor(NULL),
23              length = col_integer()))
24 nrow(cloud.zyx)
25
26 library(dplyr)
27 library(ggplot2)
28
29 p <- ggplot(data = cloud.zyx %>% filter(length < 2048)) +
30   aes(x = real, y = imag) + geom_hex(bins = 46) +
31   facet_grid(
32     facets = length~target_period,
33     labeller =
34       label_bquote(
35         cols = T == .(as.character(target_period)),
36         rows = n == .(length)) +
37     scale_fill_gradient(trans = "log10") +
38     theme(axis.text.x = element_text(angle=45, hjust=1,
39                                       vjust=1))
40
41 ggsave(plot = p, filename=plotfile,
42        width = 297, height = 210, units = "mm")
```

O Bloco B.8 gera o arquivo da Figura 5.5 no servidor. Para tornar essa figura disponível para a geração do arquivo pdf desta tese, é necessário copiar o arquivo de volta para o diretório onde o documento está sendo gerado. O Bloco B.9 é executado no computador local e realiza a cópia. A sintaxe `noweb` do Bloco B.6 foi utilizada na Linha 1.

Bloco B.9 – Copia o arquivo da Figura 5.5 do servidor

```
1 file="<<results_filename()>>.pdf"
2 PROCDIR="${PROC_SERVER_DIR}/${CAMPANHA}"
3
4 rsync -a "${proc_server_ip}:${PROC_DIR}/${file}" .
```

Ainda resta resolver uma situação não contemplada nesta apresentação. Caso o servidor não possa ser acessado no momento do processamento, o arquivo com o gráfico da Figura 5.5 não estará disponível e o documento não poderá ser produzido. Para evitar isso, esse arquivo foi acrescentado ao repositório `git` onde este documento é armazenado. Usualmente, os arquivos de figuras não precisam ser armazenados nos repositórios de pesquisas reproduzíveis, pois eles são produzidos no momento do processamento do texto. Mas, para garantir a produção do texto correto sempre que este trabalho fosse processado, os arquivos com figuras geradas no servidor têm uma cópia local no repositório `git@gitlab.com:marcos.banik/tese.git`.

B.3 Dependências externas

Para garantir a reprodutibilidade é necessário manter um registro das bibliotecas externas utilizadas no trabalho (CHEN et al., 2018; GIL et al., 2016; PICCOLO; FRAMPTON, 2016). No caso deste trabalho, além daquelas já citadas, também foram utilizadas as seguintes bibliotecas:

- `armadillo` (SANDERSON; CURTIN, 2016; SANDERSON; CURTIN, 2018): é uma biblioteca em linguagem C++ utilizada para cálculos de álgebra linear. Esta biblioteca é dependência da biblioteca `mlpack`.
- `mlpack` (CURTIN et al., 2018): biblioteca em linguagem C++ utilizada em operações de aprendizado de máquina. É esta biblioteca que fornece as rotinas para identificação dos aglomerados utilizados na Seção 5.2.3.
- `nanoflann` (BLANCO; RAI, 2014): biblioteca em C++ utilizada para obter as distancias dos primeiros vizinhos como mostrado na Figura 5.8.

- `boost`¹: é um conjunto de bibliotecas em linguagem C++ auditadas por um processo de revisão de pares. Neste trabalho, a biblioteca `boost` é utilizada para implementar a manipulação de arquivos e `strings`, cálculos em ponto flutuante com precisão múltipla, cálculos com números racionais, fornece valores numéricos para várias constantes físicas e matemáticas, interação com a linha de comando e com o sistema de arquivos do sistema operacional.
- `Eigen` (GUENNEBAUD et al., 2010): biblioteca em linguagem C++ usada para cálculos de álgebra linear.
- `FFTW` (FRIGO; JOHNSON, 2005): biblioteca em linguagem C utilizada para calcular as transformadas de Fourier.
- `LAPACK` (ANDERSON et al., 1999): biblioteca em `Fortran` utilizada para solução de problemas em álgebra linear. Neste trabalho, ela é usada para obter os `tapers` DPSS utilizados para calcular os espectros e exibidos nas Figuras B.1, B.2 e B.3.
- `range-v3`²: biblioteca em linguagem C++ utilizada para manipulação de vetores de dados.
- `nlohmann/json`³: biblioteca em C++ utilizada para manipular arquivos em formato `json` (ECMA INTERNATIONAL, 2017).
- `ifultools` (CONSTANTINE, 2016): biblioteca utilizada pelo R (R CORE TEAM, 2019) para calcular valores em decibels. Utilizada para obter a Figura B.3.
- `magrittr` (BACHE; WICKHAM, 2014): implementa os operadores de fluxo `%>%` e `%<>%` em linguagem R. Esses operadores facilitam a manipulação do fluxo de dados e foram usados para obter as Figuras 5.5, 5.6, 5.8 5.9, B.1, B.2 e B.3. Seu uso pode ser visto no Bloco B.8.
- `RANN` (ARYA et al., 2019): usado para calcular a distância dos primeiros vizinhos e utilizado para gerar a Figura 5.8.
- `rwantshue` (HÖSLER, 2019): biblioteca utilizada para gerar um conjunto com um grande número de cores como o utilizado na Figura 5.5.

¹<https://www.boost.org/>

²<https://github.com/ericniebler/range-v3>

³<https://github.com/nlohmann/json>

B.4 Configuração do emacs

Essa tese utilizou alguns pacotes que não fazem parte da instalação padrão do emacs. O bloco a seguir ilustra um exemplo mínimo de configuração do emacs utilizado para a reprodução deste trabalho.

Bloco B.10 – Configuração dos pacotes do emacs utilizados neste trabalho.

```
;;; package
(load "package")
(add-to-list 'package-archives
             '("melpa" . "http://melpa.org/packages/") t)
(add-to-list 'package-archives
             '("org" . "https://orgmode.org/elpa/") t)
(package-initialize)

;;; use-package
(when (not (package-installed-p 'use-package))
      (package-install 'use-package))

;; This is only needed once, near the top of the file
(eval-when-compile
  (require 'use-package)
  (setq use-package-verbose t))

;;; ess
(use-package ess
  :ensure t
  :defer t)

(use-package org
  :ensure org-plus-contrib
  :config
  (org-babel-do-load-languages
   'org-babel-load-languages
   '((R . t)
     (lisp . t)
     (shell . t))))

;;; org-ref
(use-package org-ref
  :ensure t
  :after org)
```

O leitor que desejar reproduzir este trabalho deve conferir se os arquivos de inicia-

lização do emacs (usualmente localizado em `~/ .emacs`, ou `~/ .emacs.el`, ou ainda `~/ .emacs.d/init.el`) são compatíveis com o Bloco [B.10](#).