



Ministério da
Ciência e Tecnologia



sid.inpe.br/mtc-m19/2010/08.25.12.26-TDI

**MODELOS BASEADOS EM REDES NEURAIIS PARA O
ESTUDO DE PADRÕES CLIMÁTICOS SAZONAIS A
PARTIR DE DADOS TRATADOS COM A TEORIA DOS
CONJUNTOS APROXIMATIVOS**

Juliana Aparecida Anochi

Dissertação de Mestrado do Curso de Pós-graduação em Computação Aplicada,
orientada pelo Dr. José Demisio Simões da Silva, aprovada em 31 de agosto de
2010.

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/385N4PP>>

INPE
São José dos Campos
2010

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE (RE/DIR-204):**Presidente:**

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Membros:

Dr^a Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr^a Regina Célia dos Santos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Dr. Ralf Gielow - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr. Wilson Yamaguti - Coordenação Engenharia e Tecnologia Espacial (ETE)

Dr. Horácio Hideki Yanasse - Centro de Tecnologias Especiais (CTE)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Deicy Farabello - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Vivéca Sant´Ana Lemos - Serviço de Informação e Documentação (SID)



Ministério da
Ciência e Tecnologia



sid.inpe.br/mtc-m19/2010/08.25.12.26-TDI

**MODELOS BASEADOS EM REDES NEURAIIS PARA O
ESTUDO DE PADRÕES CLIMÁTICOS SAZONAIS A
PARTIR DE DADOS TRATADOS COM A TEORIA DOS
CONJUNTOS APROXIMATIVOS**

Juliana Aparecida Anochi

Dissertação de Mestrado do Curso de Pós-graduação em Computação Aplicada,
orientada pelo Dr. José Demisio Simões da Silva, aprovada em 31 de agosto de
2010.

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/385N4PP>>

INPE
São José dos Campos
2010

A72m Anochi, Juliana Aparecida.
Modelos baseados em redes neurais para o estudo de padrões climáticos sazonais a partir de dados tratados com a teoria dos conjuntos aproximativos / Juliana Aparecida Anochi. – São José dos Campos : INPE, 2010.
xxvi + 188 p. ; (sid.inpe.br/mtc-m19/2010/08.25.12.26-TDI)

Dissertação (Mestrado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2010.
Orientador : Dr. José Demisio Simões da Silva.

1. Previsão climática. 2. Inteligência artificial. 3. Descoberta de conhecimento em banco de dados. 4. Teoria dos conjuntos aproximativos. 5. Redes neurais artificiais. I.Título.

CDU 004.8

Copyright © 2010 do MCT/INPE. Nenhuma parte desta publicação pode ser reproduzida, armazenada em um sistema de recuperação, ou transmitida sob qualquer forma ou por qualquer meio, eletrônico, mecânico, fotográfico, reprográfico, de microfilmagem ou outros, sem a permissão escrita do INPE, com exceção de qualquer material fornecido especificamente com o propósito de ser entrado e executado num sistema computacional, para o uso exclusivo do leitor da obra.

Copyright © 2010 by MCT/INPE. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, microfilming, or otherwise, without written permission from INPE, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use of the reader of the work.

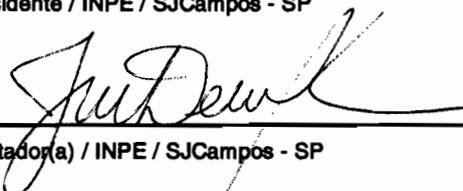
Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de Mestre em
Computação Aplicada

Dr. Stephan Stephany



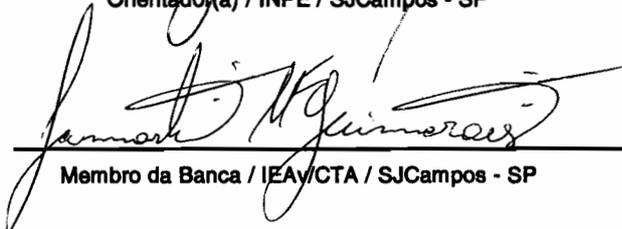
Presidente / INPE / SJC Campos - SP

Dr. José Demisio Simões da Silva



Orientador(a) / INPE / SJC Campos - SP

Dr. Lamartine Nogueira Frutuoso
Guimarães



Membro da Banca / IEAvCTA / SJC Campos - SP

Dr. João Onofre Pereira Pinto



Convidado(a) / UFMS / Campo Grande - MS



Aluno (a): Juliana Aparecida Anochi

São José dos Campos, 31 de agosto de 2010

A meus pais Olímpio e Izabel

AGRADECIMENTOS

Agradeço primeiramente a Deus, criador e razão de tudo, pela minha existência e por todas as bênçãos concebidas.

Ao meu orientador, Dr. José Demisio Simões da Silva, pela grandiosa orientação de forma segura com seus vastos conhecimento e sabedoria, pela confiança depositada, pelo incentivo, a minha entrada na área de pesquisa, e principalmente pela oportunidade de concretizar esta dissertação.

A meu querido Pai, Olimpio e minha Mãe, Izabel, a minha eterna graditão, pelo apoio e amor, sempre me incentivando em todos os momentos da minha vida.

Às secretárias Maria Cristina e Claudia Carraro pelo auxílio ao longo do curso.

A meu noivo, Ricardo Miranda, pela paciência e compreensão durante todo o tempo.

Às minhas irmãs, Andreia e Simone, pelo companheirismo e amor.

Às minhas amigas Juliana Bueno e Érica Ferreira, pelo apoio, amizade e ajuda nos momentos mais difíceis.

À Daniella Azevedo, irmã dada pela vida, por acreditar e torcer pelo meu sucesso.

Aos professores do INPE pelo conhecimento compartilhado.

Aos membros da banca examinadora pela disposição em analisar esta dissertação.

À CAPES, pelo suporte financeiro.

Finalmente, agradeço a todos aqueles que direta ou indiretamente contribuíram, de alguma maneira, na realização deste trabalho.

RESUMO

Esta dissertação trata do uso de redes neurais artificiais na derivação de modelos de previsão climatológica da variável precipitação, a partir de dados de reanálise processados por um método de mineração de dados baseado na Teoria dos Conjuntos Aproximativos, para gerar reduções de dados, com o objetivo de sintonizar os modelos de redes neurais artificiais para a previsão sazonal de precipitação, a partir de um conjunto de dados reduzidos, considerando a hipótese de que algumas variáveis são mais relevantes para o processo de previsão climática e que uma técnica de mineração de dados pode identificar estas variáveis. Os resultados obtidos com os processamentos propostos mostraram a eficácia da metodologia, apresentando estimativas de previsão similares às situações climatológicas consideradas como observações disponíveis na base de dados.

**NEURAL NETWORKS BASED MODELS IN THE STUDY OF
CLIMATIC PATTERNS DERIVED FROM DATA TREATED WITH
A ROUGH SET THEORY PROCEDURE**

ABSTRACT

This dissertation deals with the use of artificial neural networks in the derivation of climatological predictive models of the variable precipitation from reanalysis data processed by a data mining method based on Rough Sets Theory to generate reductions of data, with the goal to tune the models of artificial neural networks to forecast seasonal rainfall, from a reduced data set, considering the hypothesis that some variables are drivers for the process of weather forecasting and that a data mining technique may identify these variables. The results obtained in the conducted experiments show the effectiveness of the proposed methodology, presenting estimates similar to climatological situations considered as available observations in the database.

LISTA DE FIGURAS

	<u>Pág.</u>
3.1 Etapas do processo KDD	14
4.1 Interface da ferramenta ROSETTA.	33
5.1 Caricatura da estrutura de um neurônio biológico	42
5.2 Estruturas dos neurônios artificiais	44
5.3 Comparação entre os neurônios biológico e artificial	45
5.4 Função Sinal	46
5.5 Função Logística Sigmóide	46
5.6 Função Tangente Hiperbólica	47
5.7 Rede Camada Única	48
5.8 Rede alimentada adiante totalmente conectada com uma camada oculta .	49
5.9 Rede recorrente	49
5.10 Aprendizagem supervisionada	50
5.11 Aprendizagem não-supervisionado	51
5.12 Arquitetura da Rede Perceptron de Múltiplas Camadas	56
5.13 Rede recorrente com neurônios ocultos	62
5.14 Rede recorrente de Elman	64
5.15 Rede recorrente de Jordan	65
6.1 Área de estudo	67
6.2 Análise dependências dos dados.	71
6.3 Esquema de previsão sazonal.	73
7.1 Precipitação para o outono de 1999. (a) Precipitação Real; (b) Estimativa com RNA usando todos os dados; (c) Estimativa com dados processados por TCA.	77
7.2 Precipitação para o inverno de 1999. (a) Precipitação Real; (b) Estimativa com RNA usando todos os dados; (c) Estimativa com dados processados por TCA.	78
7.3 Precipitação para a primavera de 1999. (a) Precipitação Real; (b) Estimativa com RNA usando todos os dados; (c) Estimativa com dados processados por TCA.	78

7.4	Precipitação para o verão de 1999. (a) Precipitação Real; (b) Estimativa com RNA usando todos os dados; (c) Estimativa com dados processados por TCA.	78
7.5	Subáreas de estudo	80
7.6	Junção das 4 subáreas analisadas	81
7.7	Previsão para estação Outono de 1998	82
7.8	Previsão para estação Inverno de 1998	82
7.9	Previsão para estação Verão de 1998	83
7.10	Previsão para estação Primavera de 1998	83
7.11	Estimativa de precipitação. Verão de 1999. (a) Precipitação Real; (b) Estimativa com dados reduzidos por TCA, (c) Estimativa com todos os dados.	85
7.12	Estimativa de precipitação. Outono de 1999. (a) Precipitação Real; (b) Estimativa com dados reduzidos por TCA, (c) Estimativa com todos os dados.	85
7.13	Estimativa de precipitação. Inverno de 1999. (a) Precipitação Real; (b) Estimativa com dados reduzidos por TCA, (c) Estimativa com todos os dados.	86
7.14	Estimativa de precipitação. Primavera de 1999. (a) Precipitação Real; (b) Estimativa com dados reduzidos por TCA, (c) Estimativa com todos os dados.	86
7.15	Estimativa de precipitação no Verão de 2007 na região Nordeste.	91
7.16	Estimativa de precipitação no Verão de 2008 na região Nordeste.	91
7.17	Estimativa de precipitação no Inverno de 2007 na região Nordeste.	92
7.18	Estimativa de precipitação no Inverno de 2008 na região Nordeste.	92
7.19	Estimativa de precipitação no Outono de 2007 na região Nordeste.	93
7.20	Estimativa de precipitação no Outono de 2008 na região Nordeste.	93
7.21	Estimativa de precipitação na Primavera de 2007 na região Nordeste.	94
7.22	Estimativa de precipitação na Primavera de 2008 na região Nordeste.	94
7.23	Erro em cada ponto de grade: Rede de Elman com todos os dados.	95
7.24	Erro em cada ponto de grade: Rede de Elman usando TCA.	96
7.25	Estimativa de precipitação no Verão de 2007 na região Norte.	99
7.26	Estimativa de precipitação no Verão de 2008 na região Norte.	100
7.27	Estimativa de precipitação no Inverno de 2007 na região Norte.	101
7.28	Estimativa de precipitação no Inverno de 2008 na região Norte.	102
7.29	Estimativa de precipitação no Outono de 2007 na região Norte.	103
7.30	Estimativa de precipitação no Outono de 2008 na região Norte.	104

7.31	Estimativa de precipitação na Primavera de 2007 na região Norte.	105
7.32	Estimativa de precipitação na Primavera de 2008 na região Norte.	106
7.33	Erro em cada ponto de grade: Rede de Elman usando todos os dados.	107
7.34	Erro em cada ponto de grade: Rede de Elman usando a TCA.	108
7.35	Estimativa de precipitação no Verão de 2007 na região Centro Oeste.	112
7.36	Estimativa de precipitação no Verão de 2008 na região Centro Oeste.	113
7.37	Estimativa de precipitação no Inverno de 2007 na região Centro Oeste.	114
7.38	Estimativa de precipitação no Inverno de 2008 na região Centro Oeste.	115
7.39	Estimativa de precipitação no Outono de 2007 na região Centro Oeste.	116
7.40	Estimativa de precipitação no Outono de 2008 na região Centro Oeste.	117
7.41	Estimativa de precipitação na Primavera de 2007 na região Centro Oeste.	118
7.42	Estimativa de precipitação na Primavera de 2008 na região Centro Oeste.	119
7.43	Erro em cada ponto de grade: Rede de Elman usando todos os dados.	120
7.44	Erro em cada ponto de grade: Rede de Elman usando os dados reduzidos.	120
7.45	Estimativa de precipitação no Verão de 2007 na região Sudeste.	124
7.46	Estimativa de precipitação no Verão de 2008 na região Sudeste.	125
7.47	Estimativa de precipitação no Inverno de 2007 na região Sudeste.	126
7.48	Estimativa de precipitação no Inverno de 2008 na região Sudeste.	127
7.49	Estimativa de precipitação no Outono de 2007 na região Sudeste.	128
7.50	Estimativa de precipitação no Outono de 2008 na região Sudeste.	129
7.51	Estimativa de precipitação na Primavera de 2007 na região Sudeste.	130
7.52	Estimativa de precipitação na Primavera de 2008 na região Sudeste.	131
7.53	Erro em cada ponto de grade: Rede de Elman usando todos os dados.	132
7.54	Erro em cada ponto de grade: Rede de Elman usando a TCA.	132
7.55	Estimativa de precipitação no Verão de 2007 na região Sul.	136
7.56	Estimativa de precipitação no Verão de 2008 na região Sul.	137
7.57	Estimativa de precipitação no Inverno de 2007 na região Sul.	138
7.58	Estimativa de precipitação no Inverno de 2008 na região Sul.	139
7.59	Estimativa de precipitação no Outono de 2007 na região Sul.	140
7.60	Estimativa de precipitação no Outono de 2008 na região Sul.	141
7.61	Estimativa de precipitação na Primavera de 2007 na região Sul.	142
7.62	Estimativa de precipitação na Primavera de 2008 na região Sul.	143
7.63	Erro em cada ponto de grade: Rede de Jordan usando todos os dados.	144
7.64	Erro em cada ponto de grade: Rede de Jordan usando os dados reduzidos.	144
A.1	Variáveis Meteorológicas.	155
A.2	Experimento faltando variáveis no outono de 1998.	157
A.3	Experimento faltando variáveis no inverno de 1998.	157

A.4	Experimento faltando variáveis na primavera de 1998.	158
A.5	Experimento faltando variáveis no verão de 1998.	158
A.6	Experimento faltando variáveis indicadas pela TCA, no outono de 1998.	160
A.7	Experimento faltando variáveis indicadas pela TCA, no inverno de 1998.	160
A.8	Experimento faltando variáveis indicadas pela TCA, na primavera de 1998.	161
A.9	Experimento faltando variáveis indicadas pela TCA, no verão de 1998.	161
A.10	Experimentos acrescentando 1% de ruído.	164
A.11	Experimentos acrescentando 2% de ruído.	165
A.12	Experimentos acrescentando 5% de ruído.	166
B.1	Erro em cada ponto de grade: Rede de Elman com todos os dados.	167
B.2	Erro em cada ponto de grade: Rede de Jordan com todos os dados.	168
B.3	Erro em cada ponto de grade: Rede MLP com todos os dados.	168
B.4	Erro em cada ponto de grade: Rede de Elman usando a TCA.	169
B.5	Erro em cada ponto de grade: Rede de Jordan usando a TCA.	169
B.6	Erro em cada ponto de grade: Rede MLP usando a TCA.	170
B.7	Erro em cada ponto de grade: Rede de Elman usando todos os dados (CO).	171
B.8	Erro em cada ponto de grade: Rede de Jordan usando todos os dados (CO).	172
B.9	Erro em cada ponto de grade: Rede MLP usando todos os dados (CO).	172
B.10	Erro em cada ponto de grade: Rede de Elman usando a TCA(CO).	173
B.11	Erro em cada ponto de grade: Rede de Jordan usando a TCA(CO).	173
B.12	Erro em cada ponto de grade: Rede MLP usando a TCA(CO).	174
B.13	Erro em cada ponto de grade: Rede de Elman usando todos os dados(N).	175
B.14	Erro em cada ponto de grade: Rede de Jordan usando todos os dados(N).	176
B.15	Erro em cada ponto de grade: Rede MLP usando todos os dados(N).	176
B.16	Erro em cada ponto de grade: Rede de Elman usando a TCA(N).	177
B.17	Erro em cada ponto de grade: Rede de Jordan usando a TCA(N).	177
B.18	Erro em cada ponto de grade: Rede MLP usando a TCA(N).	178
B.19	Erro em cada ponto de grade: Rede de Elman usando todos os dados(SE).	179
B.20	Erro em cada ponto de grade: Rede de Jordan usando todos os dados(SE).	180
B.21	Erro em cada ponto de grade: Rede MLP usando todos os dados(SE).	180
B.22	Erro em cada ponto de grade: Rede de Elman usando a TCA(SE).	181
B.23	Erro em cada ponto de grade: Rede de Jordan usando a TCA(SE).	181
B.24	Erro em cada ponto de grade: Rede MLP usando a TCA(SE).	182
B.25	Erro em cada ponto de grade: Rede de Elman usando todos os dados(S).	183
B.26	Erro em cada ponto de grade: Rede de Jordan usando todos os dados(S).	184
B.27	Erro em cada ponto de grade: Rede MLP usando todos os dados(S).	184
B.28	Erro em cada ponto de grade: Rede de Elman usando a TCA(S).	185

B.29 Erro em cada ponto de grade: Rede de Jordan usando a TCA(S).	185
B.30 Erro em cada ponto de grade: Rede MLP usando a TCA(S).	186

LISTA DE TABELAS

	<u>Pág.</u>
4.1 Sistema de Informação.	24
4.2 Sistema de Decisão.	25
4.3 Relação de Indiscernibilidade.	27
4.4 Classe para IND(B).	27
4.5 Aproximações do Sistema de Decisão.	29
4.6 Matriz de Discernibilidade.	30
6.1 Variáveis para o conjunto de dados de 1980-2000.	69
6.2 Variáveis para o conjunto de dados de 2000-2009.	69
7.1 Variáveis reduzidas.	76
7.2 Erro quadrático médio.	76
7.3 Coordenadas das subáreas de estudo.	79
7.4 Redutos extraídos pela TCA.	80
7.5 Erro Quadrático Médio.	81
7.6 Variáveis extraídas pela TCA.	84
7.7 Erro Quadrático Médio.	85
7.8 Variáveis extraídas pela TCA - NE.	88
7.9 Erros quadráticos médios para modelos de redes neurais com todos os dados.	88
7.10 Erros quadráticos médios para modelos de redes neurais com dados reduzidos pela TCA.	89
7.11 Variáveis extraídas pela TCA na região Norte.	97
7.12 Erro quadrático médio para as quatro estações na região Norte.	97
7.13 Erro quadrático médio para as quatro estações na região Norte usando a TCA.	97
7.14 Variáveis extraídas pela TCA - CO.	109
7.15 Erro quadrático médio para as quatro estações da região Centro Oeste.	109
7.16 Erro quadrático médio para as estações na região Centro Oeste usando TCA.	110
7.17 Variáveis extraídas pela TCA para região Sudeste.	121
7.18 Erro quadrático médio para as quatro estações na região Sudeste.	121
7.19 Erro quadrático médio nas quatro estações na região Sudeste usando TCA.	122
7.20 Variáveis extraídas pela TCA - SUL.	133

7.21	Erro quadrático médio para as quatro estações da região Sul.	133
7.22	Erro quadrático médio das previsões feitas pelas redes neurais nas quatro estações da região Sul usando dados reduzidos pela TCA.	134
A.1	Erro quadrático médio.	156
A.2	Erro quadrático médio para a rede neural usando a TCA.	159
A.3	Erro quadrático médio - 1% de ruído.	163
A.4	Erro quadrático médio - 2% de ruído.	164
A.5	Erro quadrático médio - 5% de ruído.	166
C.1	Dados reanálise América do Sul.	187

LISTA DE ABREVIATURAS E SIGLAS

CO -- Centro Oeste
CPTEC -- Centro de Previsão de Tempo e Estudos Climáticos
CRS -- Centro Regional Sul de Pesquisas Espaciais
GrADS -- Grid Analysis and Display System
GUI -- Graphical User Interface
IA -- Inteligência Artificial
IND -- Relação de Indiscernibilidade
INPE -- Instituto Nacional de Pesquisas Espaciais
KDD -- Knowledge Discovery Databases
LHS -- Left Hand Side
MCGA -- Modelo de Circulação Geral Atmosférico
MLP -- Perceptron de Múltiplas Camadas
N -- Norte
NCAR -- The National Center for Atmospheric Research
NCEP -- National Centers for Environmental Prediction
NE -- Nordeste
NTNU -- Universidade Norueguesa de Ciência e Tecnologia
ODBC -- Open Database Connectivity
PSD -- Physical Sciences Division
RED -- Redutos
RHS -- Right Hand Side
RNAs -- Redes Neurais Artificiais
RNA -- Rede Neural Artificial
ROSETTA -- Rough Set Toolkit for Analysis of Data
S -- Sul
SD -- Sistema de Decisão
SE -- Sudeste
SI -- Sistema de Informação
TCA -- Teoria dos Conjuntos Aproximativos
TCN -- Teoria dos Conjuntos Nebulosos

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO	1
2 METEOROLOGIA E CLIMATOLOGIA	5
2.1 Subdivisões da Climatologia	6
2.2 Previsão Meteorológica e Climatológica	7
2.3 Aspectos climáticos sobre as regiões	9
3 DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS	13
3.1 Descoberta de Conhecimento em Banco de Dados	13
3.1.1 Conhecimento do domínio da aplicação	15
3.1.2 Seleção de Dados	15
3.1.3 Pré-processamento	15
3.1.4 Transformação de dados	16
3.1.5 Mineração de dados	17
3.1.6 Interpretação	17
3.1.7 Uso do conhecimento descoberto	17
3.2 Mineração de dados	17
3.3 Tarefas e Funcionalidades da mineração de dados	18
3.3.1 Caracterização	19
3.3.2 Classificação	19
3.3.3 Regras de Associação	19
3.3.4 Regressão	20
3.3.5 Agrupamento ou clusterização	20
3.4 Técnicas de mineração de dados	21
4 TEORIA DOS CONJUNTOS APROXIMATIVOS	23
4.1 Sistemas de Informação	24
4.2 Sistemas de Decisão	24
4.3 Relação de Indiscernibilidade	25
4.4 Aproximação de Conjuntos	27
4.5 Matriz de Discernibilidade	30
4.6 Função de Discernibilidade	30

4.7	Redução	31
4.8	Ferramenta ROSETTA	32
4.8.1	Importar e exportar dados	32
4.8.2	Pré-processamento	33
4.8.3	Discretização	34
4.8.4	Redução de atributos	36
4.8.5	Interpretação e a Avaliação	38
4.8.6	Implementação	38
5	REDES NEURAIS ARTIFICIAIS	39
5.1	Neurônios Biológicos	42
5.2	Neurônio Artificial	43
5.3	Função de Ativação	45
5.4	Arquiteturas de Rede	47
5.4.1	Redes Alimentadas adiante com camada única	47
5.4.2	Redes alimentadas com Múltiplas Camadas	48
5.4.3	Redes Recorrentes	48
5.5	Processos de Aprendizagem	50
5.5.1	Aprendizagem Supervisionada	50
5.5.2	Aprendizagem Não-Supervisionada	51
5.6	Regras de Aprendizagem	51
5.6.1	Aprendizagem por Correção do Erro	51
5.6.2	Aprendizagem baseada em memória	52
5.6.3	Aprendizagem Hebbiana	52
5.6.4	Aprendizagem Competitiva	53
5.6.5	Aprendizagem de Boltzmann	54
5.7	Perceptron	54
5.8	Perceptron de Múltiplas Camadas	55
5.8.1	Arquitetura da rede	55
5.8.2	Algoritmo de Retropropagação	57
5.9	Redes Recorrentes	61
5.9.1	Rede de Elman	62
5.9.2	Rede de Jordan	63
6	METODOLOGIA	67
7	RESULTADOS	75

7.1	Resultados na região Nordeste em 1999	75
7.2	Resultados para na região Nordeste em 1998	79
7.3	Resultados na região Sul em 1999	83
7.4	Resultados de precipitação para as cinco regiões do Brasil	86
7.4.1	Resultados na região Nordeste	88
7.4.2	Resultados na região Norte	96
7.4.3	Resultados na Região Centro Oeste	109
7.4.4	Resultados na região Sudeste	121
7.4.5	Resultados na região Sul	133
8	CONSIDERAÇÕES FINAIS	145
	REFERÊNCIAS BIBLIOGRÁFICAS	151
	APÊNDICE A - ANÁLISE DE ROBUSTEZ	155
A.1	Simulando a ausência de variáveis	155
A.2	Análise com ruídos	163
	APÊNDICE B - MAPAS DE ERRO EM CADA PONTO DE	
	GRADE	167
B.1	Mapas de erros para os modelos de previsão na região Nordeste	167
B.2	Mapas de erros para os modelos de previsão na região Centro Oeste	171
B.3	Mapas de erros para os modelos de previsão na região Norte	175
B.4	Mapas de erros para os modelos de previsão na região Sudeste	179
B.5	Mapas de erros para os modelos de previsão na região Sul	183
	APÊNDICE C - DADOS DE REANÁLISE	187

1 INTRODUÇÃO

Esta dissertação propõe o desenvolvimento de modelos empíricos de previsão climática, a partir de dados de reanálise, utilizando redes neurais artificiais (RNA) supervisionadas, treinadas com dados processados pela Teoria de Conjuntos Aproximativos (TCA), em uma abordagem de descoberta de conhecimento e de redução de dados.

O desenvolvimento de modelos de previsão climática a partir dados considera a hipótese de que é possível extrair dos dados históricos informações sobre o comportamento das condições climáticas. Neste sentido, a metodologia de desenvolvimento necessita dispor de um grande número de dados para garantir que o modelo considera uma vasta faixa de situações. Entretanto, apesar de poder garantir, a princípio, maior robustez aos modelos derivados, a manipulação de grandes volumes de dados pode exigir muito poder computacional, inviabilizando o uso da metodologia de desenvolvimento para a calibração de modelos locais.

A disponibilidade de dados tem crescido com o desenvolvimento da tecnologia computacional e da capacidade de processamento e armazenamento de informações em banco de dados.

Desta maneira, a análise e extração de conhecimento necessitam de novas abordagens que levem a resultados em tempo hábil para uso deste conhecimento em processos de tomada de decisão, principalmente naqueles que envolvem situações críticas para o ser humano, como nos processos de análise de clima e de tempo em Meteorologia.

Para este cenário, a Descoberta de Conhecimento em Banco de Dados (Knowledge Discovery in Database - KDD) é uma solução que pode ser usada para garantia da qualidade dos dados a serem utilizados durante a extração de informações úteis associadas a padrões escondidos entre os dados.

Os modelos empíricos desenvolvidos neste trabalho são meios de estudar as transformações ambientais para o entendimento do comportamento climático, que tem se tornado uma tarefa essencial para aos seres humanos em função de suas atividades diárias associadas, por exemplo, à agricultura, ao setor industrial, à aviação, entre outras. Os modelos físico-computacionais para previsão meteorológica tentam simular a física dos processos atmosféricos e necessitam de todas as informações disponíveis para reproduzir o estado do comportamento atmosférico. Desta maneira,

a construção de um modelo de previsão se torna complexo devido à necessidade de modelagem computacional dos fenômenos físicos e químicos, envolvendo um grande número de variáveis associadas aos fenômenos naturais e requerendo grande poder computacional para a execução. Esse fato é um limitante para a disseminação da tarefa de avaliação da atmosfera já que pode exigir altos investimentos. O Centro de Previsão de Tempo e Estudos Climáticos do Instituto Nacional de Pesquisas Espaciais (CPTEC-INPE) (www.cptec.inpe.br), por exemplo, é um dos órgãos responsáveis pelos estudos climáticos no Brasil que emprega um supercomputador para realização de estudos das condições do tempo e clima, através da elaboração de cenários futuros com alta resolução. Estes cenários apóiam os estudos de impactos e vulnerabilidade e ainda podem permitir a elaboração de projeções dos extremos climáticos do estado atmosférico.

Este trabalho de dissertação aborda o problema propondo a derivação de modelos de previsão, a partir de dados históricos de reanálise, usando as técnicas de redes neurais que podem ser executadas em qualquer tipo de máquina, mantendo um excelente desempenho, em termos de tempo de processamento. Para isso, o trabalho também aborda a questão da existência de grandes volumes de dados, através de uma técnica de redução de dados.

Para a derivação dos modelos, considera-se a hipótese de que a eliminação de redundâncias nos dados mantém o conhecimento necessário para o entendimento dos fenômenos relacionados e para realização da climatologia, cujos objetivos consistem em prognosticar, descobrir e explorar o comportamento atmosférico, visando os benefícios para os seres humanos.

Na metodologia apresentada nesta dissertação, a redução de dados é utilizada como método de adequação do volume de dados disponível para o processamento por RNAs.

Os modelos de previsão de clima com RNAs propostos são alternativas para a aquisição de prognósticos de variáveis climáticas, com o objetivo de auxiliar a tomada de decisão nos processos de previsão climática.

Pessoa (2004) desenvolveu um modelo computacional para realização de previsão de clima, utilizando o modelo de rede Perceptron de Múltiplas Camadas (MLP) e a TCA para redução de variáveis, a partir de dados do período entre janeiro de

1980 a dezembro de 2000, com propósito de realizar a previsão em cinco subáreas da América do Sul, cobrindo apenas uma parte das cinco regiões brasileiras. Em continuidade ao trabalho de (PESSOA, 2004), esta dissertação além de usar a rede MLP, também faz uso das redes recorrentes Elman e Jordan, para derivar novos modelos de previsão, utilizando dados de um período entre janeiro de 2000 a dezembro de 2009.

Portanto, o desenvolvimento dos modelos de previsão climática descritos nesta dissertação usa as redes: MLP, Elman e Jordan, com treinamento supervisionado, realizado com dados históricos, pré-processados pela TCA. Os modelos então utilizam os padrões com o objetivo de realizar previsões climáticas sazonais para um intervalo de tempo, sobre as cinco regiões do Brasil.

No pré-processamento, a TCA identifica os atributos mais relevantes para o processo de previsão, permitindo assim uma redução da complexidade do problema e a possibilidade de criar modelos de previsão climática a partir de um conjunto simplificado de dados e de realizar prognósticos do comportamento climático, utilizando modelos simplificados construídos diretamente dos dados.

Esta dissertação está organizada da seguinte forma:

- O Capítulo 1 faz uma introdução ao documento apresentando diretamente o tema do presente trabalho;
- O Capítulo 2 apresenta os principais conceitos relacionados ao problema de previsão climática;
- No Capítulo 3 são abordados alguns tópicos referentes aos conceitos utilizados em mineração de dados e descoberta de conhecimento, incluindo suas funcionalidades e objetivos;
- No Capítulo 4 são apresentados e exemplificados os conceitos fundamentais da teoria dos conjuntos aproximativos;
- O Capítulo 5 apresenta algumas definições sobre os paradigmas das redes neurais artificiais;
- O Capítulo 6 descreve um detalhamento do trabalho e a metodologia utilizada para realização do modelo de previsão climática;

- O Capítulo 7 apresenta os resultados experimentais e comparativos a respeito da teoria dos conjuntos aproximativos;
- O Capítulo 8 traz algumas considerações finais e futuras aplicações da metodologia desenvolvida.

2 METEOROLOGIA E CLIMATOLOGIA

A meteorologia é a ciência que estuda os fenômenos que ocorrem na atmosfera terrestre, com foco nos estados físicos, dinâmicos, químico-atmosférico e as interações entre eles. Neste campo são observadas as atividades humanas e os efeitos dinâmicos sobre a superfície terrestre e o oceano, sendo considerado um dos campos mais complexos da ciência natural e aplicada, devido à extensão da atmosfera e a grande variedade de fenômenos (VIANELLO; ALVES, 2006).

A pesquisa ciência atmosférica é usada para obter informações dos processos que causam a evolução das condições do tempo, a fim de, empregar esse conhecimento na previsão destas condições. Deste modo, além de realizar a previsão do tempo, a ciência atmosférica é usada para caracterizar a diversidade das áreas de estudo e as aplicações que decorrem nessa ciência, a qual possui o propósito de estudar fenômenos atmosféricos que ocorrem diariamente, tais como, poluição, mudanças na camada de ozônio, raios, entre outros.

A previsão de tempo e clima são conceitos distintos utilizados em meteorologia, os quais são usados para entender o comportamento da atmosfera em diferentes intervalos de tempo e são os aspectos considerados como os mais tradicionais e abordados em meteorologia.

O estudo do tempo consiste nas condições da atmosfera em uma determinada região do planeta em um intervalo de tempo muito curto, a partir do estado considerado. Essa é considerada uma tarefa complexa devido ao dinamismo inerente do estado da atmosfera em relação aos seus efeitos sobre a vida e as atividades dos seres humanos. Em um curto prazo, a previsão do tempo condiciona a segurança para os meios de transportes aéreo, naval e rodoviário, os quais podem contar com informações que possam interferir na viagem. Na agricultura a previsão do tempo traz informações direcionadas e precisas sobre as variáveis que interferem em culturas como café, soja, entre outras.

O estudo do clima é uma integração das condições do tempo em certo período representando uma caracterização mais abstrata. O clima é definido como um conjunto de condições naturais que dominam uma região, obtidas pelo comportamento médio da atmosfera em um intervalo de tempo maior. Desta maneira, as variações e as condições extremas do tempo são importantes para caracterizar uma região. Por

exemplo, os agricultores estão interessados em conhecer a precipitação média em um determinado mês, e ainda, os valores extremos de seca e sua probabilidade de ocorrência.

Vianello e Alves (2006) definem a climatologia como o estudo científico do clima, em que utiliza os mesmos dados básicos usados na meteorologia, e os meteorologistas utilizam seus resultados, nas previsões do tempo, nas aplicações em indústrias, na agricultura entre outros. O objetivo da climatologia consiste em prognosticar, descobrir e explorar o comportamento dos fenômenos atmosférico, visando ao benefício para o ser humano.

2.1 Subdivisões da Climatologia

A climatologia é uma ciência aplicada, a qual procura entender o funcionamento do clima cujos métodos são essencialmente meteorológicos, porém os objetivos e resultados são geográficos. O climatologista necessita conhecer as técnicas abordadas em meteorologia e essencialmente um conhecimento geográfico (VIANELLO; ALVES, 2006).

Na climatologia existem diferentes definições, todas válidas, porém variáveis de acordo com a finalidade do estudo. Koppen propõe o clima como o somatório das condições atmosféricas que fazem um lugar da superfície terrestre ser mais ou menos habitável para os seres humanos, essa classificação é desenvolvida com base em dados de precipitação, temperatura e sua distribuição dentro das estações. Hann define o clima como um conjunto de fenômenos meteorológicos que caracterizam o estado médio da atmosfera em um ponto da superfície terrestre. Marchi sugere o clima como um complexo das condições atmosféricas de uma região que a tornam favorável às necessidades do homem. Entretanto o meteorologista Poncelet definiu o clima como um conjunto habitual flutuante de elementos físicos, químicos e biológicos, os quais caracterizam a atmosfera de um determinado local e influencia na vida dos seres que nele se encontra. Esta definição foi examinada e estudada pela Comissão de Climatologia da Organização Meteorológica Mundial (VIANELLO; ALVES, 2006).

Vianello e Alves (2006) apresentam as seguintes subdivisões da climatologia:

- Climatologia Física: envolve a investigação do comportamento dos elementos do tempo ou processos atmosféricos em termos de princípios físicos e

químicos. Neste, estuda assuntos relativos à radiação eletromagnética, a composição e estrutura da atmosfera, e aos processos de trocas verticais entre a superfície da Terra e da atmosfera.

- **Climatologia Dinâmica:** consiste nos movimentos atmosféricos em termos hidrodinâmicos, especificamente na circulação geral da atmosfera, a qual utiliza dados de estações climatológicas, modelos matemáticos e análises baseadas em modelos teóricos. Os meteorologistas se preocupam com a explanação científica dos fenômenos que ocorrem em grande escala e prever eventos importantes dentro da meteorologia.
- **Climatologia Sinótica:** compreende o estudo do tempo e clima em uma área com relação às condições climáticas locais com circulação atmosférica. A climatologia sinótica se interessa pelos aspectos climatológicos, os quais podem ser essenciais no processo de previsão do tempo, utilizando dados de várias estações meteorológicas, tais como: temperatura, vento, umidade e tipo de nuvens são coletados.
- **Climatologia Aplicada:** consiste na análise científica dos dados climáticos, com intuito de aplicá-los na solução de projetos específicos e de problemas operacionais nos campos de indústria, agricultura, florestas entre outros.

2.2 Previsão Meteorológica e Climatológica

A previsão do tempo consiste na descrição detalhada de ocorrências futuras, baseadas nas médias de diversas medições, sobre como o tempo estará em determinado local e instante. A previsão do tempo depende de medições corretas das variáveis meteorológicas provenientes de estações meteorológicas, imagens de radar, balões atmosféricos, bóias marítimas, entre outras. Após os dados serem coletados, esses são processados em supercomputadores, os quais analisam as possibilidades da evolução do tempo, gerando previsões baseadas em probabilidades.

A previsão climática consiste na estimativa do comportamento médio da atmosfera com alguns meses de antecedência. Por exemplo, em escala de tempo sazonal, se está interessado em prever se o próximo inverno será mais frio que a média ou se haverá mais ou menos chuva. Cabe ainda à previsão climática analisar a friagem no inverno e as ondas de calor, visando prever as propriedades estatísticas do estado climático,

entretanto, tal estimativa não pode dizer exatamente a quantidade de chuvas ou se a temperatura estará mais ou menos elevada (VIANELLO; ALVES, 2006).

A previsão de tempo e clima realizada por centros de previsão como o Centro de Previsão de Tempo e Clima do Instituto Nacional de Pesquisas Espaciais (CPTEC-INPE), utiliza modelos que procuram descrever o comportamento das condições físico-químicas da atmosfera. Estes são executados em sistemas computacionais de grande porte, capazes de realizar estudos das condições do tempo e clima para a elaboração de cenários futuros de mudanças do clima com alta resolução. Estas atividades apóiam estudos de impactos e vulnerabilidade e permitem que se façam projeções dos extremos climáticos do estado atmosférico.

Atualmente no Centro de Previsão de Tempo e Clima (CPTEC) utiliza dois modelos numéricos para realização de previsões do tempo e clima. Dentre esses modelos existe o chamado Modelo de Circulação Geral Atmosférico (MCGA) o qual cobre todo o globo, e o modelo ETA, que cobre a América do Sul e de oceanos próximos.

Modelo de Circulação Geral Atmosférico

O modelo de circulação geral atmosférico (MCGA) teve sua origem no modelo usado para previsão de tempo de médio prazo pelo National Centers for Environmental Prediction (NCEP) em 1985 (BONATTI, 1996).

O MCGA tem sido utilizado tanto para previsão de tempo operacional, quanto para previsão climática de forma experimental, com o propósito de estudar a variabilidade e as mudanças climáticas. As variáveis prognósticas do modelo MCGA são: o logaritmo da pressão à superfície, a vorticidade, a divergência, a temperatura virtual e a umidade específica (BONATTI, 1996).

Modelo ETA

O modelo ETA é um modelo que utiliza o método de ponto de grade. Neste método, um conjunto de pontos (grade) é introduzido na região de interesse e as variáveis dependentes são inicialmente definidas e em seguida computadas nesses pontos. A resolução horizontal atual é de 40 km e a vertical de 38 camadas. A condição inicial é proveniente da análise do NCEP e as condições de contorno lateral são provenientes das previsões do MCGA do CPTEC e atualizadas a cada 6 horas. O prazo de integração é de 72 horas (CHOU, 1996).

Ensemble

A técnica de previsão de tempo por ensemble procura melhorar a qualidade da climatologia e simular o estado da atmosfera. [Mendonça e Bonatti \(2002\)](#) define a técnica de previsão de tempo por ensemble como uma tentativa de aumentar o prazo de previsão e a previsibilidade dos modelos dinâmicos.

Para a determinação da climatologia de um modelo, em uma determinada região, existem três parâmetros que precisam ser definidos, tais como: data de início da integração, comprimento da integração e número de membros.

Em geral, a previsão por ensemble faz o tratamento considerando apenas as incertezas na condição inicial, buscando através de alguma técnica específica, a estimativa dos erros relacionados às observações para criar um conjunto de condições iniciais, onde consiste em utilizar o mesmo modelo em termos dos mesmos parâmetros físicos atmosféricos e forçantes, porém executando com diversas datas diferentes ([MENDONÇA; BONATTI, 2002](#)).

A previsão de tempo por ensemble caracteriza uma grande quantidade de informações e frequentemente tem sido um desafio para os centros meteorológicos que implantaram este sistema, condensar essas informações e extrair delas as características mais importantes para o previsor ou para o usuário final ([MENDONÇA; BONATTI, 2002](#)).

Modelo numérico regional

Outro modelo numérico é o modelo regional, que é uma solução para aumentar a resolução do modelo sem aumentar o custo computacional. Este modelo prevê fenômenos de pequena escala, tais como, tempestades, brisa marítimas entre outros ([INPE.CPTEC, 2009](#)).

2.3 Aspectos climáticos sobre as regiões

O Brasil possui um extenso território o que possibilita uma diversidade de climas, que influencia diretamente as temperaturas e os índices pluviométricos nas regiões do país. De maneira geral, as estações no Brasil são bem definidas e o índice pluviométrico é regular. A temperatura no inverno é baixa e chove pouco, enquanto que no verão a temperatura é mais alta e chove mais ([INPE.CPTEC, 2009](#)).

Na região Norte do país consiste em um clima equatorial, com altas temperaturas e um alto índice de chuva, quase sem estação seca. O período chuvoso da região ocorre nas estações do verão e no outono, a exceção de Roraima e parte do norte do Amazonas, onde o máximo pluviométrico se dá no inverno, por influência do regime do hemisfério Norte.

O Nordeste do país é caracterizado como semi-árido e está associado a temperaturas elevadas o ano todo, sendo que o índice de chuva na região não possui uma boa distribuição no tempo e no espaço. A região possui uma onda de calor intensa e permanente, provocando uma intensa evaporação das águas de sua bacia hidrográfica fluvial. Na região do sertão o período chuvoso ocorre normalmente em apenas dois meses no ano, podendo, em alguns anos até não existir, ocasionando as denominadas secas regionais (VIANELLO; ALVES, 2006).

A climatologia da região Sudeste é bastante diversificada, com elevadas precipitações ao longo de todo o ano. A temperatura no verão, principalmente no mês de janeiro, podem atingir médias entre 30°C e 32°C. No inverno, a temperatura média varia entre 6°C e 20°C, com temperaturas mínimas absolutas entre -4° e 8°C, registradas nas áreas mais elevadas.

Com relação ao regime de chuvas o máximo pluviométrico da região Sudeste normalmente ocorre em janeiro e o mínimo em julho, enquanto o período seco ocorre normalmente no inverno com duração de seis meses.

A região Sul é uma região climaticamente mais regular. As chuvas são normalmente bem distribuídas no ano todo, sendo a única região brasileira com ocorrência de neve. Quanto ao regime térmico, o inverno é frio e o verão é quente. A temperatura média anual situa-se entre 14°C e 22°C. No verão, principalmente em janeiro, a temperatura média é superior a 24°C. A média das temperaturas máximas mantém-se em torno de 24°C a 27°C nas superfícies mais elevadas do planalto e, nas áreas mais baixas, entre 30°C e 32°C.

No inverno, principalmente em julho, a temperatura média se mantém relativamente baixa, oscilando entre 10°C e 15°C. A média das temperaturas máximas mantém-se baixa, em torno de 20°C a 24°C. A média das mínimas varia de 6°C a 12°C, sendo comum atingir temperaturas próximas de 0°C, ou mesmo alcançar índices negativos, acompanhados de geada e neve, quando da invasão das massas polares. O máximo

pluviométrico acontece no inverno e o mínimo no verão em quase toda a região (INPE.CPTEC, 2009).

Na região Centro-Oeste o clima é tropical semi-úmido, com sazonalidade tipicamente tropical. Mais de 70% do total de chuvas acumuladas durante o ano se precipitam de novembro a março. O inverno é excessivamente seco, pois as chuvas são muito raras. A temperatura média anual é de 22°C. Na primavera e no verão, são comuns temperaturas elevadas, quando a média do mês mais quente varia de 24°C a 26°C. A média das máximas de setembro (mês mais quente) oscila entre 30°C e 36°C. O inverno é uma estação com temperaturas amenas, mas ocorrem, com frequência, temperaturas baixas, em razão da invasão de ventos polares. A temperatura média do mês mais frio oscila entre 15°C e 24°C, e a média das temperaturas mínimas varia de 8°C a 18°C, não sendo raro a ocorrência de mínimas absolutas negativas (INPE.CPTEC, 2009).

3 DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS

Neste Capítulo são abordados alguns tópicos referentes ao processo de descoberta de conhecimento em banco de dados (*Knowledge Discovery in Database KDD*), tais como as funcionalidades e objetivos que podem ser alcançados, as etapas do processo de KDD e suas áreas de aplicação.

3.1 Descoberta de Conhecimento em Banco de Dados

Fayyad et al. (1996) definem o processo de descoberta de conhecimento em banco de dados, como o processo que busca extrair padrões válidos, novos, potencialmente úteis e compreensíveis presentes no conjunto de dados, mas que não são visíveis diretamente. O processo de KDD é conhecido também como extração do conhecimento, arqueologia de dados ou colheita de informações.

Com o desenvolvimento da tecnologia computacional e a possibilidade de armazenamento de informações em grandes bases de dados provenientes de equipamentos como radares, sensores, satélites entre diversos outros, a análise e extração de conhecimento resultou em uma crescente necessidade de obter novas abordagens, a fim de, alcançar resultados em tempo hábil para uso deste conhecimento em processos de tomada de decisão e habilidade de procurar informações potencialmente úteis, as quais dificilmente podem ser identificadas e utilizadas diante de um grande volume de dados. Esse crescente número de dados, armazenados em grandes bases de dados, existem em muitos domínios, tais como, financeiro, medicina, científico entre outros.

Diante da necessidade de uma nova geração de técnicas computacionais e ferramentas, exigidas para suportar a extração de conhecimento útil presentes em crescentes volumes de dados, surgiu a técnica de descoberta de conhecimento em banco de dados, com o propósito de entender o comportamento dos mesmos.

Fayyad et al. (1996) apresentam a descoberta de conhecimento em banco de dados como um processo interativo e iterativo, envolvendo diversas tomadas de decisão, sendo centrado na interação entre usuários e especialistas. O processo iterativo consiste em uma sequencias de operações, permitindo retornar a etapas anteriores, a fim de, ajustar o conjunto de parâmetros com propósito de obter resultados mais adequados em relação aos objetivos preestabelecidos.

No processo interativo, o usuário diante das descobertas realizadas, obtém novas

hipóteses, e neste caso, pode necessitar intervir e decidir pela retomada dos processos de mineração, ou uma nova seleção de atributos, com intuito de validar hipóteses que surgiram durante o processo.

As pesquisas em KDD evoluíram e atraíram esforços nos campos de pesquisas, tais como: bancos de dados, aprendizado de máquinas, reconhecimento de padrões, estatísticas, inteligência artificial, aquisição de conhecimento para sistemas especialistas, visualização de dados, recuperação de informação e computação de alto desempenho. Os sistemas de KDD incorporam algoritmos e métodos de todos estes campos de pesquisa, com objetivo de encontrar padrões característicos, os quais possam ser interpretados como conhecimento útil e atraente.

No processo de KDD, existem diferentes abordagens para as divisões das etapas de seus processos. [Fayyad et al. \(1996\)](#) dividem o processo em nove etapas, compreendendo o conhecimento do domínio da aplicação, obtenção conjunto de dados, limpeza e pré-processamento dos dados, redução do volume de dados, escolha da função de mineração, escolha do algoritmo, a mineração dos dados, interpretação dos resultados e a utilização do conhecimento descoberto. A [Figura 3.1](#) apresenta uma visão geral das etapas que compõem o processo de KDD segundo [Fayyad et al. \(1996\)](#)

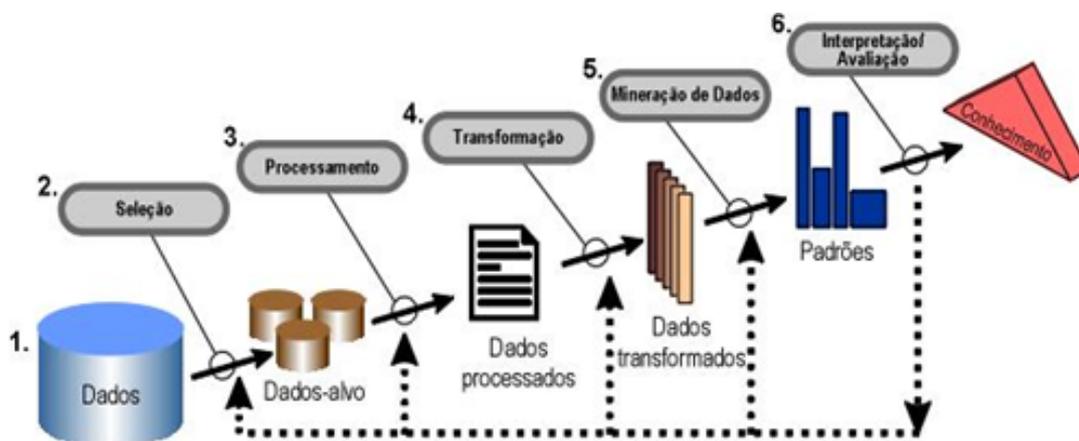


Figura 3.1 - Etapas do processo KDD
Fonte: (FAYYAD et al., 1996)

As principais etapas do processo de descoberta de conhecimento em banco de dados

são descritas a seguir:

3.1.1 Conhecimento do domínio da aplicação

Nesta etapa inclui o conhecimento inicial e relevante do domínio da aplicação, compreendendo na análise e identificação dos principais objetivos e restrições sobre o domínio.

Uma das dificuldades dessa etapa inicial está no fato de não existir padronização, uma vez que cada situação mantém suas características. Dessa forma, algumas considerações relevantes podem ser úteis nessa fase, na tentativa de definir o campo de aplicação. Dentre elas, podem-se destacar: definição dos objetivos do projeto de KDD, levantamento das possíveis bases de dados com as quais se pretende trabalhar, pesquisa sobre o conhecimento prévio existente, modo de distribuição do conhecimento extraído, definição das relações custo e benefício (OLIVEIRA, 2000).

3.1.2 Seleção de Dados

Esta etapa consiste na seleção do conjunto de dados, tabelas e atributos que estejam associados aos objetivos do usuário, sobre os quais serão realizadas as descobertas e a geração de um novo conjunto de dados, com propósito de serem submetidos às demais etapas de pré-processamento. Ainda nesta etapa envolve a identificação de possíveis problemas, tais como a ausência de dados de determinadas variáveis, e a determinação de aspectos que podem ser relevantes para a área de aplicação.

Os dados podem ser classificados em bases internas e externas. As bases internas pertencem diretamente ao domínio de aplicação e normalmente são constituídos de dados operacionais e históricos, os quais já estão armazenados em um repositório estruturado e requerem menos processamento. A base externa é composta por outros tipos de domínio público, como por exemplo, documentos pela internet. Diferente da base externa, esses tipos de dados, necessitam ser formatados, de forma que possam ser preparados para a análise (OLIVEIRA, 2000).

3.1.3 Pré-processamento

Nesta etapa os dados são organizados para a execução do algoritmo de extração de conhecimento, onde normalmente estes não estão em um formato adequado para o processamento, devido alguns problemas relacionados com a presença de ruídos,

dados corrompidos, valores desconhecidos e ausência de dados (REZENDE, 2003).

Diante desses problemas, a limpeza dos dados é realizada para garantir um resultado satisfatório, a qual consiste em remover registros com valores inválidos presentes em algum atributo, efetuando a limpeza do domínio, através da remoção de ruídos e erros, e o tratamento de atributos incompletos.

3.1.4 Transformação de dados

Após os dados serem pré-processados é possível que seja necessário a transformação do formato dos dados, pois os algoritmos que são usados durante o processo de KDD, requerem que esses estejam disponíveis em um formato apropriado, com intuito de superar quaisquer limitações existentes nos algoritmos de extração de padrões. Dentre as possíveis transformações, são usadas: a normalização, a discretização e a transformação de atributos qualitativos em quantitativos, os quais são descritos a seguir.

- A normalização compreende em transformar os valores dos dados originais em outro tipo, com intuito de utilizá-los mais adequadamente, por exemplo, converter (normalizar) os valores de atributos contínuos, modificando para um intervalo entre 0 e 1 ou de -1 e 1.
- A discretização envolve a transformação de um atributo contínuo para um atributo discreto. Por exemplo, é possível transformar um atributo contínuo, como a temperatura $^{\circ}\text{C}$, em um atributo categórico como Temperatura Alta e Temperatura Baixa, ou ainda, discretizar a faixa de valores de uma variável em um número de valores que podem ser caracterizadas como classes.
- A transformação de atributos qualitativos em quantitativos compreende em converter os atributos qualitativos em atributos quantitativos. Por exemplo, os atributos qualitativos sem ordem própria, tais como azul, amarelo e roxo, podem ser mapeados para números, portanto, esses acabam por criar uma ordem de valores que não é real, entretanto os atributos qualitativos com ordem tais como baixo, médio e alto, podem ser mapeados para valores numéricos de forma a manter a ordem dos valores, como baixo = 1, médio = 2 e alto = 3 (BATISTA, 2003).

3.1.5 Mineração de dados

Essa é a etapa essencial do processo de KDD, envolve a seleção de algoritmos, técnicas e modelos sobre o conjunto de dados, os quais serão aplicados para realizar a extração de padrões ou regras para construir uma base de dados consistente e confiável para a descoberta das informações desejadas. Uma grande variedade de técnicas provenientes de diversas áreas de conhecimento, como aprendizado de máquina, redes neurais artificiais, estatísticas, banco de dados entre outros, estão empregados na mineração de dados e podem ser combinadas para se obter resultados melhores.

3.1.6 Interpretação

Na etapa de interpretação envolve a análise e compreensão dos padrões obtidos, que podem ser utilizados como suporte a tomada de decisão por parte do usuário, a quem cabe a avaliação para se decidir pela necessidade de repetir algum passo do processo de KDD. Vale ressaltar que após a obtenção do conhecimento e caso os resultados não sejam satisfatórios, existe a possibilidade de retorno para qualquer das etapas anteriores, a partir do ponto que seja mais oportuno para tal análise.

3.1.7 Uso do conhecimento descoberto

A última etapa do processo inclui a incorporação do conhecimento na concepção de um sistema aplicado, o qual realiza ações baseadas nos padrões descobertos e a checagem para resolução de conflitos potenciais e a verificação do conhecimento extraído.

3.2 Mineração de dados

A etapa de mineração de dados é considerada como uma sequência de etapas dentro do processo de KDD, a qual envolve repetidas iterações, utilizando um conjunto de técnicas que envolvem métodos matemáticos, algoritmos inspirados em diferentes paradigmas para extração de padrões válidos e representativos para área de aplicação, além de descobrir regularidades em conjuntos de dados. A mineração de dados é caracterizada por selecionar, explorar e modelar grandes volumes de dados, cujo objetivo é detectar padrões de comportamento a partir de dados observados.

Mitra et al. (2002) apresentam que a mineração de dados pode ser aplicada em praticamente todas as áreas do conhecimento, tais como:

- Investimento financeiro: podem obter vantagens competitivas no mercado por meio da descoberta, por exemplo, obter informações dos perfis de clientes através do uso de cartão de crédito, o qual pode ser útil para combater fraudes, analisar o crédito e usar em futuras promoções;
- Medicina: o diagnóstico das informações armazenadas pelos sistemas de gestão hospitalar e os padrões de pacientes que desenvolveram algum tipo de doença, podem ser usadas na tentativa de prevenir ou antecipar o tratamento de doenças;
- Internet: consiste em melhorar a pesquisa e a extração de informações na Internet. O domínio de aplicações da mineração de dados na Internet apresentou um crescimento acentuado devido um aumento significativo de informações processadas. A área de segurança de redes é um exemplo de aplicação, podendo ser utilizada para detectar padrões de invasão;
- Mineração de dados científica: compreende em detectar padrões e construir modelos que sejam capazes de simular o comportamento de um determinado fenômeno, os quais demandam análises, tais como, físico, químico ou biológico, com propósito de utilizá-los na construção de sistemas especialistas, de maneira que auxiliem nos processos de diagnósticos em diversas áreas.

3.3 Tarefas e Funcionalidades da mineração de dados

Na mineração de dados, os tipos de padrões que podem ser descobertos dependem de suas funcionalidades (ou tarefas), as quais possuem dois tipos principais de tarefas que são descritivas e preditivas. A descritiva os sistemas extraem padrões que descrevem as características e comportamentos existentes nos dados envolvendo a descoberta de padrões. A preditiva utiliza algumas variáveis de um conjunto de dados para prever valores desconhecidos ou futuros de outras variáveis de interesse, buscando identificar a classe de uma nova amostra de dados, a partir do conhecimento adquirido de um conjunto de amostras (OLSON; DURSUN, 2008).

As tarefas da mineração de dados estão relacionadas com o domínio da aplicação e o interesse do usuário, as quais indicam o tipo de conhecimento que deve ser extraído. Funcionalmente, as tarefas de mineração de dados podem ser categorizadas como:

caracterização, classificação, regras de associação, regressão e agrupamento (OLSON; DURSUN, 2008)

3.3.1 Caracterização

A caracterização consiste em generalizar e resumir as características dos objetos em uma classe e produzir regras. Por exemplo, é possível caracterizar quais tipos de consumidores são compradores de imóveis em uma cidade X ou ainda caracterizar os consumidores de um supermercado que regularmente compraram mais de 30 alimentos por mês. As informações relevantes para a classe especificadas pelo usuário são retornadas por uma consulta ao banco de dados e passam sobre um módulo de resumo para extrair as características dos dados em diferentes níveis de abstração.

3.3.2 Classificação

Classificação é uma tarefa preditiva que consiste no mapeamento de um conjunto de dados em um número finito de categorias, onde cada exemplo pertence a um conjunto de classes categóricas predefinidas, as quais possuem exemplos de um conjunto de atributos e um atributo-meta discreto, cujo objetivo é encontrar alguma relação entre os atributos e uma classe, de forma que o processo possa realizar a previsão da classe de um novo exemplo. Desta forma, a classificação compreende em obter um modelo baseado em um conjunto de exemplos, o qual descreve uma função desconhecida com objetivo de utilizar esse modelo para prever o valor do atributo-meta de novos exemplos (REZENDE, 2003).

3.3.3 Regras de Associação

As regras associativas consistem na descoberta de relações entre vários atributos, de forma que, a relação de um atributo particular em uma transação de dados com outros atributos na mesma transação é usada para prever padrões. Desta forma, o objetivo é encontrar tendências que possam ser utilizadas para exploração dos padrões. Por exemplo, se um cliente compra um computador (X), então ele também pode comprar um mouse (Y). A partir de associações é possível criar regras do tipo (X) implica (Y) (OLSON; DURSUN, 2008).

O formato de uma regra de associação pode ser representado como uma implicação $LHS \Rightarrow RHS$, em que LHS e RHS são respectivamente, o lado esquerdo (*Left Hand Side - LHS*) e o lado direito (*Right Hand Side - RHS*) da regra, definidos por

um conjunto de atributos e seus respectivos valores.

Os autores (AGRAWAL; SRIKANT, 1994) definem uma regra de associação como: Seja D um banco de dados composto por um conjunto de itens ordenados $A = \{a_1, a_2, \dots, a_m\}$, em que, o conjunto total de transações é chamado de conjunto $T = \{t_1, t_2, \dots, t_m\}$, a qual é composta por um conjunto de atributos de A , ou seja, $t_i \subseteq A$.

Cada regra de associação possui dois parâmetros de avaliação. O primeiro parâmetro é o suporte, o qual identifica o percentual de vezes que o conjunto de atributos A aparece no conjunto de transações T . O segundo parâmetro é a confiança que consiste no percentual de ocorrência em que um item aparece em uma transação quando outro item surge. Os parâmetros confiança e suporte são essenciais para o funcionamento da regra de associação, pois esses determinam diretamente a quantidade e a qualidade das regras geradas.

3.3.4 Regressão

Regressão está relacionada com a tarefa preditiva, cujo objetivo é mapear um item de dado para a predição de outro. Além disso, a regressão permite a descoberta de relações entre variáveis dependentes e independentes (OLSON; DURSUN, 2008).

Os modelos de regressão obtidos durante o processo de mineração de dados podem expressar o conhecimento em diferentes formatos de representação. As regras e as árvores de decisão são exemplos de modelos de regressão. Outro exemplo de modelo são as redes neurais artificiais, as quais podem realizar a predição de mais de um atributo-meta, onde normalmente a maioria dos problemas relacionados a regressão realizam a predição de apenas um atributo-meta (REZENDE, 2003).

3.3.5 Agrupamento ou clusterização

Essa tarefa é conceitualmente similar à classificação, consiste em formar grupos ou classes de dados homogêneos, entretanto, nessa tarefa as classes são desconhecidas e o algoritmo deve descobrir classes aceitáveis. Para realização dessa tarefa são utilizados algoritmos que buscam semelhanças entre os dados, a fim de agrupá-los em uma mesma classe, baseando-se no princípio de maximizar a similaridade entre objetos da mesma classe e minimizar a similaridade entre objetos de classes diferentes (OLSON; DURSUN, 2008).

3.4 Técnicas de mineração de dados

Mineração de dados é uma área multidisciplinar que incorpora diversas técnicas em diferentes áreas tais como, na Inteligência Artificial, Teoria dos Conjuntos Aproximativos, Teoria dos Conjuntos Nebulosos, Redes Neurais Artificiais, Algoritmos Genéticos, Indução de regras, Redes Bayesianas, Árvores de decisão entre outras.

Tais técnicas descrevem um modelo de extração de conhecimento, de maneira que vários algoritmos possam seguir esse protótipo. Por exemplo, dentre as técnicas mais utilizadas em mineração de dados destacam-se: as regras de decisão, os algoritmos genéticos que fazem parte da computação evolutiva, as redes neurais artificiais que são usadas em diversas etapas de mineração de dados, podem ser transformadas em um conjunto de regras de classificação, ou seja, a rede neural será treinada para classificar corretamente novos dados (REZENDE, 2003).

Dentre as técnicas citadas acima, para o desenvolvimento desta dissertação foram utilizadas duas delas: a teoria dos conjuntos aproximativos para realização do pré-processamento dos dados, em uma tentativa de aprimorar os resultados, buscando o uso de variáveis de maior importância e o uso da técnica de redes neurais artificiais para o desenvolvimento de um modelo de previsão de climática. Essas técnicas são descritas nos próximos capítulos.

4 TEORIA DOS CONJUNTOS APROXIMATIVOS

A Teoria dos Conjuntos Aproximativos (TCA), proposta no início da década de 80 pelo matemático polonês, Zdzislaw Pawlak, como um formalismo matemático, para o tratamento de informações incertas e imprecisas em aplicações de Inteligência Artificial, por meio de aproximações de um conjunto de dados (PAWLAK, 1982).

A TCA baseia-se na similaridade entre objetos através de uma relação de indiscernibilidade, a qual essencialmente considera que se tal relação existe entre dois ou mais objetos, significa que esses possuem os mesmos valores para todos os atributos que os caracterizam, portanto não podem ser discernidos entre si. Esta relação permite o uso da TCA na construção de subconjuntos de atributos a partir de uma base de dados, os quais são capazes de representar o conhecimento da base de dados com seus atributos iniciais, através de um processo de eliminação de atributos irrelevantes proporcionando a redução dos dados, através dos chamados redutos (PAWLAK, 1982).

As principais vantagens da teoria dos conjuntos aproximativos em relação a outras abordagens de tratamento da incerteza são:

- possibilidade de gerar regras de decisão em função do processamento;
- tratamento quantitativo e qualitativo da incerteza;
- o fato de não necessitar de informações preliminares sobre os dados, tais como à distribuição de probabilidade ou grau de pertinência;
- métricas estatísticas para avaliação das regras.

A TCA é uma extensão da teoria dos conjuntos e frequentemente é comparada à Teoria dos Conjuntos Nebulosos (TCN), no sentido de ambas tratarem sobre o conhecimento impreciso. As diferenças estão associadas ao fato da TCA não necessitar de informação adicional sobre os dados, relacionadas com o grau de pertinência (como na TCN) ou a distribuição de probabilidade no tratamento da imprecisão dos dados (UCHÔA, 1998).

No desenvolvimento desta dissertação foi utilizado o sistema ROSETTA (*Rough Set Toolkit for Analysis of Data* - <http://www.lcb.uu.se/tools/rosetta>), o qual possui

um conjunto de algoritmos para análise tabular de dados sob a abordagem da TCA, construído como ferramenta de apoio a processos de mineração e descoberta de conhecimento em dados, cobrindo as diferentes etapas do processo. Uma característica importante do ROSETTA é a possibilidade de validação dos conjuntos reduzidos de dados e a análise das regras (se-então) geradas durante o processo (ØHRN, 1999a).

4.1 Sistemas de Informação

A representação dos dados na abordagem TCA é feita através de um Sistema de Informação (SI), em que os dados são representados no formato de tabela, onde as linhas correspondem aos objetos e as colunas representam os valores dos atributos (KOMOROWSKI; ØHRN, 1999).

A definição formal de um SI é um par ordenado $SI = (U, A)$ em que U , é um conjunto finito e não vazio de objetos, chamado de universo e A é um conjunto finito e não vazio de elementos chamados de atributos, tal que, $a : U \rightarrow V_a$ para todo $a \in A$. O conjunto V_a é o conjunto de valores de a , chamado de domínio de a . Os objetos que formam a tabela do SI representam as ocorrências observadas dos atributos. Na Tabela 4.1 observa-se um exemplo de SI.

Tabela 4.1 - Sistema de Informação.

	Atributos Condicionais		
U	Estação do Ano	Temperatura	Vento
<i>o1</i>	Outono	média	forte
<i>o2</i>	Inverno	baixa	moderado
<i>o3</i>	Primavera	alta	moderado
<i>o4</i>	Verão	alta	forte
<i>o5</i>	Outono	média	forte

4.2 Sistemas de Decisão

Um Sistema de Decisão (SD) é um SI considerando um atributo de decisão d que não pertence ao conjunto de atributos A . Formalmente, $SD = (U, Ad)$, em que $d \notin A$ é o atributo de decisão. (SKOWRON; POLKOWSKI, 1998).

Um SD pode ser determinístico ou não-determinístico. Um SD determinístico descreve unicamente as decisões e ações a serem efetuadas quando são satisfeitas algu-

mas condições. É não-determinístico, se as decisões não são determinadas excepcionalmente pelas condições.

Acrescentando o atributo de decisão Chuva, referindo-se à ocorrência de chuva, no sistema de informação da Tabela 4.1, tem-se o SD na Tabela 4.2. Aplicando a relação de indiscernibilidade para todos os atributos, os objetos do SD agrupam-se em classes de equivalência.

Tabela 4.2 - Sistema de Decisão.

U	Atributos Condicionais			Atributo de decisão
	Estação do Ano	Temperatura	Vento	Chuva
$o1$	Outono	média	forte	Sim
$o2$	Inverno	baixa	moderado	Não
$o3$	Primavera	alta	moderado	Sim
$o4$	Verão	alta	forte	Sim
$o5$	Outono	média	forte	Não

Para o SD apresentado na Tabela 4.2, tem-se cinco objetos no conjunto de objetos $U = \{o_1, o_2, o_3, o_4, o_5\}$, três atributos condicionais $A = \{\text{Estação do ano, Temperatura e Vento}\}$ e um atributo de decisão: $d = \{\text{Chuva}\}$.

4.3 Relação de Indiscernibilidade

A relação de indiscernibilidade (IND) é considerada como um dos principais conceitos envolvido em TCA, usada para medir a similaridade entre dois ou mais objetos. Se tal relação existe entre os objetos, isso implica que os valores de seus atributos são idênticos e não podem ser discernidos entre si, desta forma são definidos os atributos redundantes ou dispensáveis.

Um SI pode ser desnecessariamente grande devido à representação redundante de atributos idênticos ou à existência de atributos supérfluos. Portanto, uma relação de equivalência tem a capacidade de tratar esses problemas, de modo que, apenas um objeto represente toda uma classe. Por exemplo, considerando o conjunto de atributos estação do ano e temperatura mencionados na Tabela 4.2, observa-se que os atributos o_1 e o_5 são caracterizados pelos mesmos valores, desta forma, utilizando o conceito de relação de indiscernibilidade, esses podem ser reduzidos de modo que, apenas um objeto represente uma classe.

Dado um sistema de informação $SI = (U, A)$, então com qualquer conjunto $B \subseteq A$ existe uma relação de equivalência $IND_A(B)$ definida por Pawlak (1994):

$$IND_A(B) = \{(x, x') \in U^2 | \forall a \in B, a(x) = a(x')\} \quad (4.1)$$

Em que $IND_A(B)$ é a relação de indiscernibilidade B . Se $(x, x') \in IND_A(B)$, então os objetos x e x' são indiscerníveis relativamente a qualquer atributo no conjunto B .

Uma relação de equivalência IND é uma relação que é reflexiva, simétrica e transitiva. Para cada relação de equivalência existe uma maneira natural para dividir o conjunto em que são definidos subconjuntos exclusivos que são chamados de classes de equivalência (STEPANIUK, 2008).

Se X é um conjunto não vazio e R é uma relação em X , tem-se as seguintes situações:

- Reflexividade: R é dita reflexiva se xRx para todo $x \in X$, ou seja, se $(x, x) \in R$ para todo $x \in X$, todos os elementos se relacionam com si próprios.
- Simetria: Uma relação binária é simétrica se qualquer xRy implica yRx .
- Transitividade: A transitividade de uma relação binária vale quando xRy e yRz implicam que xRz .

Considerando o SD apresentado na Tabela 4.2, são mostrados alguns possíveis subconjuntos dos atributos condicionais que podem ser usados na criação de relações de indiscernibilidade:

- {Estação do Ano}.
- {Temperatura}.
- {Vento},
- {Estação do Ano, Temperatura},
- {Estação do Ano, Vento},
- {Temperatura, Vento},

- {Estação do Ano, Temperatura, Vento}.

Considerando o subconjunto $B = \{\text{Estação do ano, Temperatura e Vento}\}$ os objetos o_1 e o_5 apresentam os mesmos valores, portanto existe uma relação de indiscernibilidade, ou seja, não podem ser discernidos entre si. Para o subconjunto $B = \{\text{Temperatura}\}$, os objetos o_3 e o_4 estão na mesma classe de equivalência e são indiscerníveis, assim como os objetos o_1 e o_5 .

Desta forma, considerando os subconjuntos dos atributos condicionais, tem-se as seguintes relações de indiscernibilidade apresentados na [Tabela 4.3](#):

Tabela 4.3 - Relação de Indiscernibilidade.

IND (Estação do Ano)	$[\{o_1, o_5\}, \{o_2\}, \{o_3\}, \{o_4\}]$
IND (Temperatura)	$[\{o_1, o_5\}, \{o_2\}, \{o_3, o_4\}]$
IND (Vento)	$[\{o_1, o_4, o_5\}, \{o_2, o_3\}]$
IND (Estação do Ano, Temperatura)	$[\{o_1, o_5\}, \{o_2\}, \{o_3\}, \{o_4\}]$
IND (Estação do Ano, Vento)	$[\{o_1, o_5\}, \{o_2\}, \{o_3\}, \{o_4\}]$
IND (Temperatura, Vento)	$[\{o_1, o_5\}, \{o_2\}, \{o_3\}, \{o_4\}]$
IND (Estação do Ano, Temperatura, Vento)	$[\{o_1, o_5\}, \{o_2\}, \{o_3\}, \{o_4\}]$

Para o subconjunto $IND_A(B) = \{\text{Estação do Ano, Temperatura, Vento}\}$, os objetos o_1 e o_5 são indiscerníveis, dessa forma, é possível reduzi-los, formando assim a classe $C1$ apresentada na [Tabela 4.4](#).

Tabela 4.4 - Classe para IND(B).

	Atributos Condicionais		
U	Estação do Ano	Temperatura	Vento
C1	Outono	média	forte
C2	Inverno	baixa	moderado
C3	Primavera	alta	moderado
C4	Verão	alta	forte

4.4 Aproximação de Conjuntos

Uma relação de equivalência induz um particionamento do universo. Estas partições podem ser usadas para construir novos subconjuntos do universo. Considerando

um SD , os subconjuntos de maior interesse têm o mesmo valor do resultado do atributo de decisão, ou seja, as características dos atributos condicionais que definem o atributo de decisão chuva se vai chover ou não no caso da [Tabela 4.2](#).

Os subconjuntos do universo considerado podem ser descritos em termos dos conjuntos elementares, através de duas aproximações, a aproximação inferior e a aproximação superior.

Pode-se observar na [Tabela 4.2](#) que não há uma resposta única para o atributo de decisão para as classes o_1 e o_5 , elas possuem uma relação de indiscernibilidade nos atributos condicionais e valores discerníveis para o atributo de decisão, neste caso, existe uma contradição tornando essas classes não-determinísticas ([KOMOROWSKI et al., 1998](#)).

Os elementos de um SD podem ser particionados em regiões chamadas de aproximações. Existem quatro regiões formadas por este particionamento ([KOMOROWSKI; ØHRN, 1999](#)):

- Aproximação inferior;
- Aproximação superior;
- Região de Fronteira ou Borda;
- Região Externa.

Para o sistema de informação $SI = (U; A)$ tem-se:

- $B \subseteq A$, onde A são os atributos condicionais;
- $X \subseteq U$, onde U é o conjunto de elementos;
- X é o conjunto de classe alvo e $x \in X$.

A aproximação inferior é o conjunto de objetos que podem ser classificados, com total certeza, como membros do conjunto X . A aproximação inferior de X em relação a B é denotada por $\underline{B}X$, e a equação é definida pela seguinte forma:

$$\underline{B}X = U \{E \in U / IND(B) | E \subseteq X\} \quad (4.2)$$

em que E é um subconjunto de X .

A aproximação superior são os objetos que podem ser classificados como possíveis membros do conjunto X , ou seja, são tratados pelos conjuntos que apresentam padrões idênticos nos atributos condicionais e valores diferentes no atributo de decisão, gerando decisões contraditórias. A aproximação superior de X em relação a B é denotada por $\overline{B}X$, e a equação é definida pela seguinte forma (KOMOROWSKI et al., 1995):

$$\overline{B}X = U \{E \in U / IND(B) | E \cap X \neq \emptyset\} \quad (4.3)$$

A região de fronteira ou borda de X obtém-se pela diferença de $\underline{B}X$ para $\overline{B}X$, são os objetos que são impossíveis de serem classificados como membros do conjunto X . A equação é definida por:

$$BNB(X) = \underline{B}X - \overline{B}X \quad (4.4)$$

A região externa são os objetos que não pertencem a X , a equação é da seguinte forma:

$$EXT(X) = U - \overline{B}X \quad (4.5)$$

Para as aproximações referentes ao sistema de decisão apresentados na Tabela 4.2, para a classe de decisão do atributo "chuva", $X_s = Sim = \{o_1, o_3, o_4\}$ e $X_n = Não = \{o_2, o_5\}$ as aproximações são apresentadas na Tabela 4.5:

Tabela 4.5 - Aproximações do Sistema de Decisão.

$\underline{B}(X_s) = \{o_3\}, \{o_4\}$	$\underline{B}(X_n) = \{o_2\}$
$\overline{B}(X_s) = \{o_3\}, \{o_4\}, \{o_1, o_5\}$	$\overline{B}(X_n) = [\{o_2\}, \{o_1, o_5\}]$
$BNB(X_s) = \overline{B}(X_s) - \underline{B}(X_s) = \{o_1, o_5\}$	$BNB(X_n) = \overline{B}(X_n) - \underline{B}(X_n) = \{o_1, o_5\}$
$EXT(X_s) = U - \overline{B}(X_s) = \{o_2\}$	$EXT(X_n) = U - \overline{B}(X_n) = \{o_1\}, \{o_3\}, \{o_4\}$

Considerando as classes de decisão do atributo "chuva" $X_s = Sim$, para a aproxi-

mação inferior as classes o_3 e o_4 , apresentam valores exclusivos nos atributos condicionais e não geram decisões inconsistentes. Analisando a aproximação superior, observa-se que as classes o_1 e o_5 apresentam valores idênticos para os atributos condicionais e valores diferentes para o atributo de decisão, gerando decisões inconsistentes.

4.5 Matriz de Discernibilidade

A matriz de discernibilidade é feita por meio de uma comparação entre os valores dos atributos. Desta forma, os atributos que apresentam valores discerníveis são inseridos na matriz, onde cada entrada consiste em um conjunto de atributos que difere os objetos x_i e x_j (ØHRN, 1999a).

A matriz de discernibilidade do sistema é simétrica $n \times n$ com entradas $M_{i,j}$ que representa o conjunto de atributos que discerne os objetos (KOMOROWSKI; ØHRN, 1999).

$$M_{i,j} = \{a \in B | a(x_i) \neq a(x_j)\} \quad (4.6)$$

para $i, j = 1, \dots, n$

Fazendo as seguintes abreviações para efeito de simplificação Estação do Ano = EA , Temperatura = T e Vento = V , a Tabela 4.6 apresenta a matriz de discernibilidade do SD na Tabela 4.6.

Tabela 4.6 - Matriz de Discernibilidade.

	C_1	C_2	C_3	C_4
C_1	\emptyset	EA, T, V	EA, T, V	EA, T
C_2	EA, T, V	\emptyset	EA, T	EA, T, V
C_3	EA, T, V	EA, T	\emptyset	EA, V
C_4	EA, T	EA, T, V	EA, V	\emptyset

4.6 Função de Discernibilidade

Com a matriz de discernibilidade construída, o passo seguinte é a construção da função de discernibilidade que são as uniões de cada entrada a partir da Tabela 4.6. Em

cada célula, as variáveis booleanas são conectadas por meio do operador disjuntivo.

A função de discernibilidade f_A para um sistema de decisão é uma função booleana de m variáveis booleanas a_1^*, \dots, a_m^* (correspondente aos atributos a_1, \dots, a_m), onde $m_{i,j} = \{a^* | a \in m_{i,j}\}$ definida como (KOMOROWSKI; ØHRN, 1999):

$$f_A(a_1^*, \dots, a_m^*) = \bigwedge \left\{ \bigvee m_{i,j}^* | 1 \leq j \leq i \leq n, m_{i,j} \neq \emptyset \right\} \quad (4.7)$$

A função de discernibilidade para a Tabela 4.6 pode ser visualizada a seguir:

$$\begin{aligned} f_A(a_1^*, \dots, a_m^*) = & (EA \vee T \vee V) \wedge (EA \vee T \vee V) \wedge (EA \vee T) \wedge \\ & (EA \vee T \vee V) \wedge (EA \vee T) \wedge (EA \vee T \vee V) \wedge \\ & (EA \vee T \vee V) \wedge (EA \vee T) \wedge (EA \vee V) \wedge \\ & (EA \vee T) \wedge (EA \vee T \vee V) \wedge (EA \vee V) \wedge \\ & (EA \vee T \vee V) \end{aligned}$$

4.7 Redução

Os dados contidos em um sistema de informação podem ser utilizados para discernir atributos somente até um determinado grau. Contudo, nem todos os atributos podem ser necessários para desempenhar essa tarefa. Assim, a função de discernibilidade é utilizada para realizar o processo de redução de atributos sem alterar a relação de indiscernibilidade.

O processo de redução é feito através dos chamados redutos (RED), que são subconjuntos de atributos com capacidade de representar o conhecimento da base de dados, ou seja, agrupar em classes os objetos que são indiscerníveis a partir da matriz de discernimento, formada pelos atributos discerníveis (BAZAN et al., 1994).

Somente os atributos que preservam a relação de indiscernibilidade são mantidos na redução, ou seja, existe um subconjunto de atributos condicionais que mantêm a mesma partição, levando em consideração as classes formadas pelo atributo de decisão. Os atributos restantes são redundantes, ou supérfluos, desde que suas remoções mantenham a mesma classificação. Normalmente existem vários subconjuntos de atributos e esses que são mínimos são usualmente chamados de redutos (KOMOROWSKI; ØHRN, 1999).

4.8 Ferramenta ROSETTA

O software ROSETTA (*Rough Set Toolkit for Analysis of Data*) é voltado para a utilização da técnica da TCA, para realização dos processos de KDD, de maneira geral, os dados são analisados em buscas de regras que descrevam o seu comportamento (ØHRN, 1999a). Essa ferramenta foi desenvolvida pelo grupo de Lógica do Instituto de Matemática da Universidade de Varsóvia e o de Sistemas de Conhecimento da Universidade Norueguesa de Ciência e Tecnologia (NTNU). O ROSETTA possui duas partes distintas o kernel é uma biblioteca de classes criadas em C++ e foi desenvolvido na Universidade de Varsóvia e o front-end é a interface gráfica (*Graphical User Interface - GUI*) foi criado na NTNU (ØHRN, 1999a).

O ROSETTA possui diversos algoritmos para realização das diferentes etapas de KDD, o que possibilita diferentes abordagens e modelagens. O software é considerado uma ferramenta flexível devido a comunicação com diversos tipos de bancos de dados, a exportação dos resultados obtidos para vários formatos de arquivo além de possuir uma interface de interação com o usuário bastante intuitiva, facilitando a manipulação de informações e aplicação de algoritmos (ØHRN; KOMOROWSKI, 1997).

As informações em relação aos dados e etapas são apresentadas ao usuário em forma de árvore, o qual favorece a leitura e interpretação das informações contidos em um projeto no sistema. Na [Figura 4.1](#) apresenta a interface do software ROSETTA com todos os seus algoritmos:

O ROSETTA é um sistema que cobre todas as etapas de KDD, como descritas no [Capítulo 3](#), podendo deste modo ser empregado para realização da análise de dados utilizando como base a TCA. Diversos tipos de algoritmos para determinada etapa do KDD, como a mineração de dados, discretização, algoritmos para eliminação de ruídos, por exemplo. A seguir serão mostrados alguns algoritmos utilizados no ROSETTA, com suas devidas características.

4.8.1 Importar e exportar dados

O sistema ROSETTA possui diversos algoritmos de importação de banco de dados pertinentes à interface ODBC (*Open Database Connectivity*). Desta maneira permite que o ROSETTA importe dados tabulares diretamente de uma grande variedade de fontes. Os tipos mais comuns de dados importados são: Planilhas do Microsoft Excel, Arquivos de textos ASCII, Oracle, dBase, MSAccess, SAS. Além de exportar regras

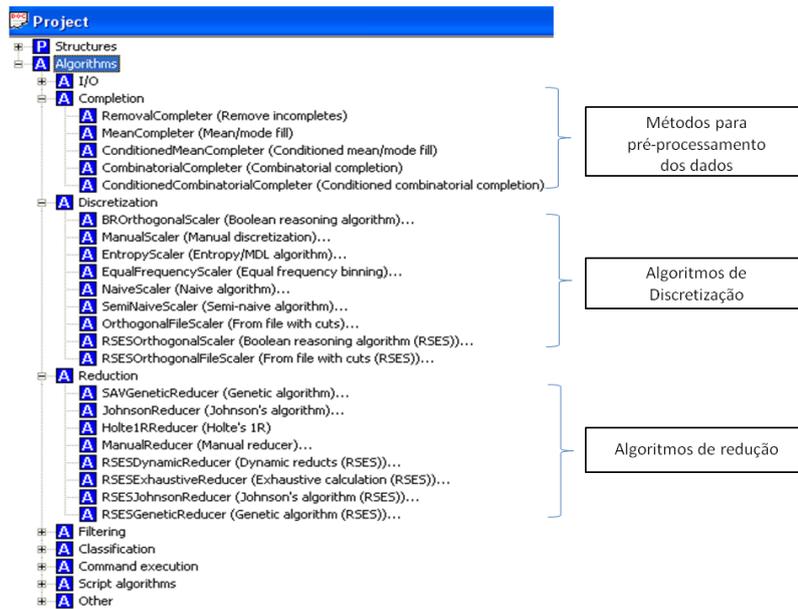


Figura 4.1 - Interface da ferramenta ROSETTA.

e redutos para diversos formatos, incluindo Prolog, C++, HTML e Matlab.

Outra característica durante a importação dos dados, é a capacidade de exportar ou importar um dicionário de dados, contendo informações sobre os atributos tais com nome, tipos e unidades, e seus respectivos mapeamentos.

4.8.2 Pré-processamento

Esta etapa compreende na limpeza dos dados, ou seja, a remoção de registros de dados que estejam incompletos ou ausentes contidos na base de dados, ainda nesta etapa consiste em completar valores de atributos com valores determinados e substituir valores de atributos. Os algoritmos disponíveis no sistema são:

- *Remove Incompletes* (Remoção dos incompletos), este faz a remoção de todos os atributos que possuem valores ausentes. A remoção de algum atributo, por sua vez pode afetar a descrição de seus valores;
- *Mean/Mode fill* (Substituição pela média), este algoritmo trata os valores ausentes substituindo pelo valor da média dos valores observados no atributo. Ao substituir os valores ausentes pela média um bias é introduzido na tabela de decisão (ØHRN, 1999b);

- *Conditioned mean/mode fill* (Substituição pela média condicionada à decisão), este tipo de algoritmo possui as mesmas funções do algoritmo anterior, porém os valores da média para atributos contínuos e os valores de maior frequência para atributos nominais são calculados considerando a classe de decisão (ØHRN, 1999b);
- *Combinatorial completion* (Substituição combinatorial), o tratamento de valores incompletos expande em vários exemplos de valores que o atributo pode assumir. Devido a essa expansão dos valores ausentes, a representação dos atributos é fortemente afetada, pois implica na multiplicação da representação dos valores nos demais atributos (ØHRN, 1999b);
- *Conditioned combinatorial completion* (Substituição combinatorial condicionada), este algoritmo é semelhante ao anterior, porém o tratamento dos valores ausentes, a representação dos atributos é menos afetada, devido a substituição ser condicionada à decisão.

Neste trabalho utilizou-se o algoritmo *Remove Incompletes* (Remoção dos incompletos), para realização do pré-processamento, o qual basicamente realiza a remoção de todos os valores ausente presentes nos atributos (ØHRN, 1999b).

4.8.3 Discretização

O procedimento mais comum de transformação em KDD baseado em discernibilidade é a discretização, em que basicamente envolve a transformação de atributos não-catóricos em atributos catóricos, mudar uma variável contínua para discreta.

Para atributos numéricos, podem-se introduzir intervalos que possibilitem a substituição linguística (exemplo: alto, médio e baixo) determinados pelos pontos de corte, podendo ser tratados de forma qualitativa ao invés de quantitativas. Porém quando os valores passam por esse processo de discretização ocorre perda de informação, pois os valores ficam reduzidos nos intervalos, entretanto existem muitos algoritmos que conseguem minimizar a perda de informação (ØHRN, 1999b).

Existem vários algoritmos de discretização disponíveis no sistema ROSETTA, os quais são citados a seguir:

- *Boolean reasoning algorithm* (Algoritmo de raciocínio booleano), baseado

na combinação dos cortes encontrados no algoritmo de Naive, utiliza o raciocínio booleano para descartar os pontos de corte excessivos e o conjunto de pontos de corte restante preservam a relação de indiscernibilidade (ØHRN, 1999b);

- *Manual discretization* (Discretização manual), neste algoritmo o usuário define os pontos de corte para discretizar os dados;
- *Entropy/MDL algorithm* (Algoritmo de Entropia/MDL), utiliza a recursividade para particionar os conjuntos de valores dos atributos, com o propósito de aperfeiçoar a medida da entropia (ØHRN, 1999b);
- *Equal frequency binning* (Frequências iguais), este algoritmo é não supervisionado e univariada, realiza a discretização dos dados procurando por n-1 pontos de corte, dividindo os conjuntos em n partes iguais. Por exemplo, uma variável que possui 30 registros e é discretizada em 3 classes, possui os valores limites da discretização ajustados para que cada classe fique com 10 registros (ØHRN, 1999b);
- *Naive algorithm* (Algoritmo Naive) o resultado de discretização deste algoritmo produz muitos cortes, na maior parte dos casos resulta em mais pontos de corte do que o necessário;
- *Semi-naive algorithm* (Algoritmo Semi-Naive), este tipo de discretização é semelhante ao Naive algorithm, a diferença está no mecanismo existente para manipular valores de exemplos que estão em uma mesma vizinhança, porém pertencentes a classes diferentes;
- *From file with cuts/ From file with cuts* (RSES) - (A partir do arquivo com cortes), estes algoritmos produzem a discretização usando arquivo de corte, a diferença entre eles é que no algoritmo de From file with cuts, deixa os atributos sem cortes não transformados (inalterados), enquanto que From file with cuts (RSES) os atributos sem cortes são definidos como 0, em vez de não transformados;
- *Boolean reasoning algorithm* (RSES) - (Algoritmo de raciocínio booleano (RSES)), é semelhante ao algoritmo de Boolean reasoning algorithm, porém os resultados são obtidos mais rápido.

Neste trabalho, o algoritmo utilizado para realização da discretização foi *Boolean reasoning algorithm (RSES)*, o qual utiliza o raciocínio booleano para descartar os pontos de corte excessivos. O conjunto restante é um conjunto de pontos mínimos de cortes que preservam a relação de indiscernibilidade. Øhrn (1999a) afirma ser o algoritmo que produz melhores resultados e eficiência, os valores ausentes são tratados como valores positivos. Caso não forem encontrados cortes nos atributos, todas as entradas para estes serão definidas para 0 (ØHRN, 1999b).

4.8.4 Redução de atributos

O sistema ROSETTA suporta em seus algoritmos o uso de informações sobre os dados, estas podem ser, por exemplo, informação de custos, cujo conteúdo é formado de dados do custo ou pesos para determinados atributos. Este tipo de informação pode ser usado para guiar algoritmos para soluções que sejam práticas, permitindo que os algoritmos de computação de reduções tenham habilidade para selecionar subconjuntos de atributo com baixos custos, preservando a relação de indiscernibilidade (ØHRN, 1999b).

Os tipos de algoritmos disponíveis para realização do processo de redução dos atributos no ROSETTA são:

- *Genetic algorithm* (Algoritmo Genético), implementa um algoritmo genético para computação dos conjuntos mínimos (MC), o qual suporta informações de custo e soluções aproximadas (ØHRN, 1999b). Os redutos são recombinados segundo uma função de custo em que os avaliam e os mantêm em uma lista e somente aqueles que satisfazem a função de custo são mantidos para serem recombinados (ØHRN, 1999b). A função de aptidão é mostrada abaixo:

$$f(B) = (1 - \alpha) \times \frac{\text{custo}(A) - \text{custo}(B)}{\text{custo}(A)} + \alpha \times \min \left\{ \epsilon, \frac{|\{s \text{ em } S \mid s \cap B \neq \emptyset\}|}{|S|} \right\} \quad (4.8)$$

Em que, ϵ é a fração dos candidatos (*hitting fraction*) para soluções aproximadas, α é o parâmetro que define o peso entre o conjunto de custo e a fração dos candidatos, B é um subconjunto de A (conjunto de atributos) e S é o conjunto de conjuntos correspondente à função de discernibilidade.

Outro parâmetro relativo as soluções aproximadas é o "extra keep list", denotado por k . *Keep list* são listas contendo subconjuntos da função de discernibilidade. Por exemplo, se $k = 0$ somente uma *keep list* referente à ϵ é retornado. Se $k > 0$ então $k + 1$ grupos de MC são retornados, cada grupo tendo uma "fração dos candidatos" aproximada entre ϵ e 1. É relevante observar que se $k = 1$ então o próprio MC é retornado (ØHRN, 1999b). Por exemplo: para soluções aproximadas com os parâmetros $\epsilon = 0.7$ e $k = 3$, então serão computados 4 grupos de MC, com as "frações de acerto" 0.7, 0.8, 0.9 e 1.

No final do processo de redução, cada reduto possui um valor de suporte associado como medida de qualidade do mesmo na representação do conhecimento.

- *Johnson's algorithm* (Algoritmo de Johnson), este algoritmo possui como principal característica retornar uma única redução. Possui um bias na busca de um simples primo implicante de tamanho mínimo;
- *Holte's 1R* (Algoritmo de Holte 1R), retorna todos os conjuntos de atributos únicos (reduções dinâmicas);
- *Manual reducer* (Cálculo manual), este algoritmo de redução, o usuário escolhe os atributos que serão reduzidos;
- *Dynamic reducts* (RSES) (Redutos dinâmicos (RSES)), um número de sub-tabelas é amostrado, os redutos são calculados utilizando um outro algoritmo e o reduto que aparecer com maior frequência é tratado como o reduto principal (ØHRN, 1999b);
- *Exhaustive calculation* (RSES) (Cálculo Exaustivo (RSES)), este algoritmo calcula todos os redutos através do cálculo exaustivo. Este algoritmo não é fornecido suporte para aproximar soluções e apenas é adequado para tabelas de decisões com tamanho moderado, devido ao cálculo exaustivo (ØHRN, 1999b);

Neste trabalho, alguns experimentos realizados mostraram que o algoritmo de genético foi o que produziu os melhores resultados, sendo, portanto este utilizado nessa dissertação.

4.8.5 Interpretação e a Avaliação

Um conjunto de regras de decisão pode ser avaliado confrontando as regras com objetos que não foram minerados, ou objetos não vistos. Alguns procedimentos de classificação oferecidos por ROSETTA são, por exemplo: Votação padrão e Classificação por rastreamento de objetos.

4.8.6 Implementação

Os classificadores, obtidos no processo de KDD, podem ser exportados pelo ROSETTA como código fonte, para uma subsequente implementação. Contado com esse recurso, as regras de decisão são exportadas como código C++, Matlab e Prolog dispondo assim o modelo para ser implementado em aplicações externas.

5 REDES NEURAIS ARTIFICIAIS

Redes Neurais Artificiais (RNAs) são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes, e que adquirem conhecimento através de experiência.

As redes neurais artificiais são sistemas paralelos distribuídos, compostos por neurônios ou unidades de processamento, que calculam determinadas funções matemáticas, normalmente não-lineares. Esses neurônios podem ser distribuídos em uma ou mais camadas interligadas por conexões (pesos sinápticos), os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede. O comportamento inteligente de uma rede neural surge das interações entre as unidades de processamento, a partir de seu ambiente através de um processo de aprendizagem, cuja função é alterar os pesos sinápticos da rede, a fim de, alcançar a proposta do projeto desejado (HAYKIN, 2001).

Em RNAs, o procedimento para a solução de problemas, passa por uma fase de aprendizagem, na qual consiste em um processo iterativo de ajuste de parâmetros de modo a computar uma função desejada ou realizar uma tarefa, com capacidade de extrair as informações necessárias para a aquisição das respostas dos problemas (CARVALHO et al., 2000).

As redes neurais se assemelham na estrutura do cérebro, e apresentam características similares ao comportamento humano, tais como:

- **Aprendizado:** envolve a capacidade de aprender através de exemplos, onde os pesos sinápticos são modificados, com objetivo de alcançar a resposta desejada, ou seja, à medida que se apresenta um exemplo à rede, esta é ajustada para relacioná-lo com uma saída (HAYKIN, 2001);
- **Associação:** a rede é capaz de fazer associações entre diferentes padrões;
- **Generalização:** consiste na capacidade da rede adaptar-se às novas situações, ou seja, produzir saídas adequadas para entradas que não estavam presentes durante o treinamento (HAYKIN, 2001);
- **Abstração:** alguns modelos de redes são capazes de abstrair a essência do conjunto de dados a elas apresentados, ou seja, a partir de padrões com ruídos, a rede pode extrair a informação do padrão sem ruído.

As primeiras pesquisas para o desenvolvimento de computadores baseados no comportamento no cérebro humano datam em 1943, quando McCulloch, um psiquiatra e neuroanatomista e Walter Pitts um matemático, publicaram um trabalho descrevendo sobre cálculo lógico das redes neurais, unificando os estudos de neurofisiologia e lógica matemática. Os autores propuseram um modelo matemático baseado nos neurônios biológicos para explicar a inteligência e mostraram que uma rede neural realizaria a computação de qualquer função computável (HAYKIN, 2001).

Em 1949 o livro *The Organization of Behavior* de Donald Hebb, um psicólogo, é publicado, no qual postula uma formulação de uma regra de aprendizagem para modificação da sinapse dos neurônios biológicos. Hebb determina que quanto mais correlacionadas estiverem as saídas de dois neurônios, maior será a intensidade da ligação entre eles e afirma que a aprendizagem de redes neurais artificiais é obtida pela variação dos pesos dos neurônios de entrada, de forma que, esse aprendizado é baseado no reforço das ligações sinápticas entre neurônios (HAYKIN, 2001).

No período entre 1951 e 1958, pouco se evoluiu nas pesquisas de redes neurais artificiais. Entretanto em 1959, Frank Rosenblatt, propõe o *Perceptron* como uma nova abordagem para o modelo proposto por McCulloch e Pitts, o qual é usado para classificar padrões linearmente separáveis, usando aprendizagem supervisionada. A topologia de uma rede *Perceptron*, consiste em estruturas de ligações entre os neurônios e um algoritmo de treinamento para execução de determinados tipos de funções.

Em 1960 Widrow e Hoff introduziram o algoritmo do mínimo quadrado médio (LMS, *least mean-square*) o qual foi utilizado para criar um elemento adaptativo chamado *Adaline* (*Adaptive Linear Element*). Esses modelos têm como base o aprendizado supervisionado por correção do erro. Widrow também introduziu a rede *Madaline* (*multiple-adaline*) foi uma das pioneiras em treinamento de múltiplos elementos adaptativos (HAYKIN, 2001).

Em 1969 no livro dos pesquisadores Marvin Minsk e Seymour Papert, apresentaram críticas ao *perceptron* introduzido por Rosenblatt, eles afirmam que o *perceptron* apresentava sérias restrições sendo capaz de solucionar apenas problemas linearmente separáveis e não conseguia distinguir padrões não-linearmente separáveis, como por exemplo, o problema do OU-Exclusivo. Em repercussão ao trabalho e as críticas de Minsky e Papert, as pesquisas na área de redes neurais artificiais foram descontinuadas.

Em meados dos anos 70 surgiram os mapas auto-organizáveis utilizando aprendizagem competitiva. Em 1976, Willshaw e Von Der Malsburg publicaram o primeiro artigo sobre a formação de mapas auto-organizáveis. Em 1980 Grossberg, inspirado na aprendizagem competitiva, estabeleceu um novo princípio de auto-organização, conhecido como Teoria da Ressonância *Adaptativa* (ART, *Adaptative Resonance Theory*) (HAYKIN, 2001).

O reaquecimento das pesquisas se deu em 1982 quando John Hopfield ressuscita as redes neurais artificiais com um modelo não linear conhecido como rede de Hopfield. Hopfield enfatiza as propriedades associativas e introduz o conceito de função de energia às redes, com aprendizagem não supervisionada.

No mesmo ano em que a rede de Hopfield foi introduzida, Kohonen propõe em seu artigo o aprendizado competitivo com auto-organização usando uma estrutura de rede unidimensional ou bidimensional, tornando-se conhecido como mapas de atributos auto-organizáveis SOFM (*self-organization feature maps*) (HAYKIN, 2001).

Na década de 80, ressurgiram as redes baseadas em *perceptron* com o modelo de rede *perceptron de múltiplas camadas*, usando um algoritmo chamado *backpropagation* ou retropropagação, com aprendizagem supervisionada, que resolveu em grande parte os problemas levantados por Minsky e Papert em 1962.

O algoritmo de *backpropagation* foi criado na mesma época por diferentes pesquisadores, tais como Werbos em 1974, que descreve em sua tese de doutorado, sobre a computação eficiente do gradiente em modo reverso aplicada em modelos de redes neurais, Parker em 1985, LeCun em 1985, os pesquisadores Rumelhart, Hinton e Williams em 1986, que conseguiram resolver em grandes partes dos problemas de aprendizagem existentes, como a utilização do algoritmo para aprendizagem de máquina e demonstraram como funcionava. Entretanto, foi com a publicação do livro "*Parallel Distributed Processing: Explorations in the Microstructures of Cognition*" de Rumelhart e McClelland, responsável pelo ressurgimento do interesse em redes neurais nos anos 80, e que exerceu uma grande influência na utilização de aprendizagem por retropropagação, que emergiu como o algoritmo de aprendizagem mais popular para o treinamento de perceptrons de múltiplas camadas. O algoritmo *backpropagation* foi aplicado com sucesso em uma grande variedade de problemas, em que estimulou o desenvolvimento de muitas pesquisas em redes neurais artificiais.

5.1 Neurônios Biológicos

O cérebro humano possui cerca de 100 bilhões de neurônios biológicos, constituindo-se de uma unidade completa e independente de processamento que podem se comunicar com milhares de outros neurônios continuamente e em paralelo (CARVALHO et al., 2000).

O neurônio biológico é caracterizado como um dispositivo computacional elementar do sistema nervoso, em que possui muitas entradas e uma saída. Os neurônios possuem um papel essencial na determinação do funcionamento e comportamento do corpo humano, os quais são formados pelos dendritos, que são um conjunto de terminais de entrada, pelo corpo central, e pelos axônios que são longos terminais de saída.

A semelhança entre o cérebro humano em relação à rede neural encontra-se em dois pontos: o conhecimento é obtido através de etapas de aprendizagem e os pesos sinápticos são usados para armazenar o conhecimento adquirido (HAYKIN, 2001).

A Figura 5.1 apresenta uma representação da estrutura de um neurônio biológico.

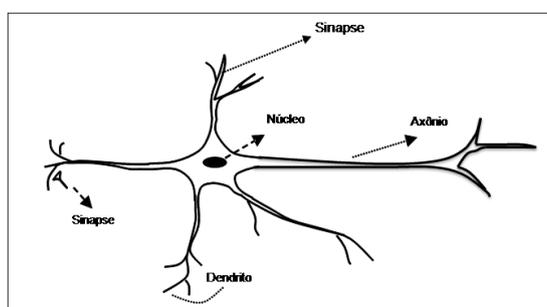


Figura 5.1 - Caricatura da estrutura de um neurônio biológico

O trabalho de Ramón y Cajal em 1911 tornou mais fácil a compreensão do cérebro humano, em que identificou a idéia dos neurônios como estrutura do cérebro. O cérebro tem a capacidade de organizar seus componentes estruturais de forma a desempenhar diversas operações. Cajal apresenta a célula nervosa (neurônio), sendo constituída pelo corpo da célula, os dendritos e o axônio, como o modelo básico de um neurônio. O corpo da célula é constituído por um núcleo, em que oferece suporte metabólico para toda célula. O axônio é a principal unidade condutora do neurônio,

capaz de conduzir sinais elétricos a longas e curtas distâncias. Os axônios possuem nas extremidades os botões sinápticos, responsáveis pelas conexões com os dendritos de outros neurônios, formando as sinapses. Os dendritos têm por função receber os impulsos nervosos provenientes dos axônios de outros neurônios, e são transmitidos ao corpo celular, e então de volta em um novo ciclo. A sinapse é o ponto de término entre o axônio e o dendrito, é a unidade funcional básica que une os neurônios e controlam a transmissão de impulsos (CARVALHO et al., 2000).

Alguns tipos de neurônios biológicos são: os neurônios bipolares, os quais conectam-se a vários neurônios dentro do cérebro e da medula espinhal; o neurônio unipolar o qual transporta sinais das extremidades do corpo para o sistema nervoso central e os neurônios multipolares que transportam sinais do sistema nervoso central para as extremidades do corpo (CARVALHO et al., 2000).

5.2 Neurônio Artificial

O modelo de neurônio artificial proposto por McCulloch e Pitts em 1943, interpreta o funcionamento do neurônio como sendo um circuito binário simples que combina várias entradas e apenas um sinal de saída. Sua descrição matemática resultou em um modelo com n terminais de entrada representando os dendritos, e apenas uma saída simulando o axônio. Para emular o comportamento das sinapses, os terminais de entrada do neurônio artificial possuem pesos, cujos valores podem ser positivos ou negativos (CARVALHO et al., 2000).

Um neurônio artificial é a unidade de processamento fundamental para a operação de uma rede neural. Seu funcionamento pode ser descrito da seguinte maneira:

- Cada entrada x é multiplicada por um peso sináptico w ;
- O resultado da multiplicação é somado com todos os sinais de entrada;
- O resultado é passado por uma função de ativação que tem a função de limitar a saída e introduzir não-linearidade ao modelo.

A Figura 5.2 mostra a representação esquemática de um neurônio artificial.

Cada entrada n é ponderada por um peso sináptico w , representado por w_n para $n = 1, 2, \dots, n$ e n é o número total de entradas. A soma ponderada do produto dos

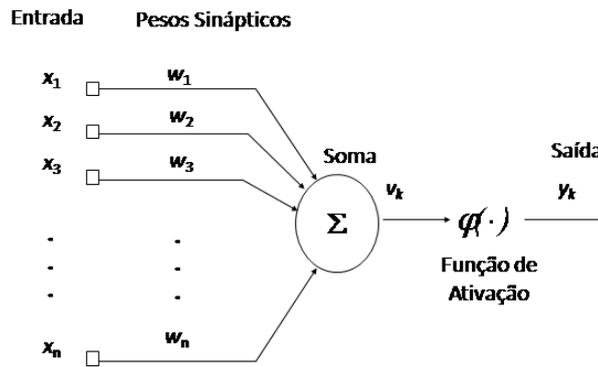


Figura 5.2 - Estruturas dos neurônios artificiais
 Fonte: (HAYKIN, 2001)

pesos de entrada e uma função de transferência são aplicadas sobre a função de ativação para gerar um sinal de saída. Entretanto a função básica de um neurônio artificial é adicionar entradas e reproduzir uma saída. Por exemplo, se o somatório das entradas for maior que um valor limite, então haverá uma saída.

Formalmente um neurônio pode ser descrito da seguinte forma:

$$V_k = \sum_{j=1}^n w_{kj} x_j \quad (5.1)$$

Os componentes do vetor de entrada são x_j para $j = 1, 2, \dots, n$ e o peso sináptico correspondente é representado por w_{kj} , que conecta a entrada j ao neurônio k . A atividade do neurônio k é dada por: $y_k = \phi$ onde ϕ é a função de ativação e y_k é o sinal de saída do neurônio.

O neurônio artificial possui um bias b que é somado com o produto das entradas e dos pesos, tendo o efeito de aumentar ou diminuir a entrada líquida da unidade, dependendo se o seu valor é positivo ou negativo, respectivamente.

O modelo de neurônio artificial proposto por McCulloch e Pitts procura simular a forma, o comportamento e as funções de um neurônio biológico. Sua estrutura será como nos neurônios biológicos, os dendritos são substituídos por *entradas* x , cujas ligações com o corpo celular artificial são realizadas através de elementos chamados *peso* w simulando as sinapses. Os estímulos captados pelas entradas são processados

pela função de soma, se o resultado do somatório for maior que um determinado valor limite, haverá uma *saída y*, caso contrário não haverá uma saída. E o limiar de disparo do neurônio biológico foi substituído pela função de ativação, a qual fornece o valor de saída *y*. Na Figura [Figura 5.3](#) é apresentada a comparação entre o neurônio biológico e o neurônio artificial proposto por McCulloch e Pitts:

A [Figura 5.3](#) apresenta uma comparação entre o neurônio biológico e o neurônio artificial proposto por McCulloch e Pitts.

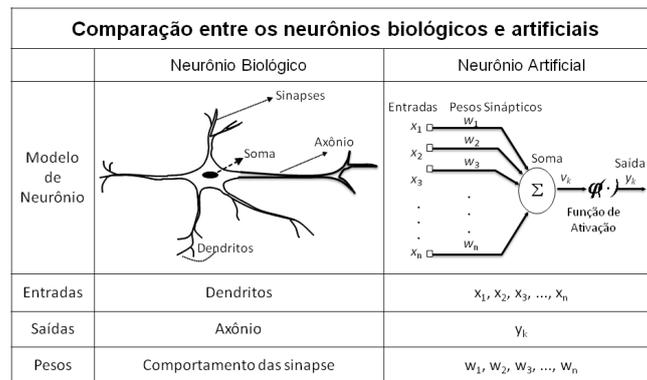


Figura 5.3 - Comparação entre os neurônios biológico e artificial

5.3 Função de Ativação

Uma função de ativação restringe a amplitude da saída de um neurônio e impõe um intervalo de amplitude do sinal de saída a um valor finito. A partir do modelo de McCulloch e Pitts, foram derivados vários outros modelos que permitem a produção de uma saída qualquer com diferentes funções de ativação. Existem diversas funções de ativação que podem ser aplicadas para gerar neurônios distintos (CARVALHO et al., 2000).

A função sinal assume valores entre 1 e -1 , representada por:

$$f_v = v \geq 0, f_v = 1v < 0, f_v = -1 \quad (5.2)$$

A [Figura 5.4](#) apresenta um exemplo de uma função sinal.

A função sigmóide pode assumir todos os valores entre 0 e 1. A representação mais

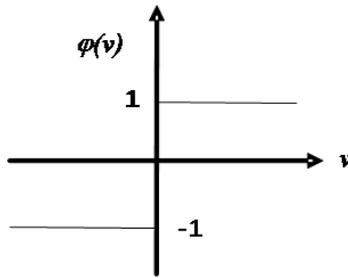


Figura 5.4 - Função Sinal
 Fonte: Adaptada de Carvalho et al. (2000).

utilizada para esta função é a função logística, cujo gráfico possui a forma de s, sendo a forma mais comum de função de ativação usada em redes neurais, definida por:

$$f_v = \frac{1}{1 + e^{-av}} \quad (5.3)$$

em que a é o parâmetro de inclinação da função sigmóide e v é o valor de ativação do neurônio. A Figura 5.5 apresenta uma ilustração da função logística sigmóide.

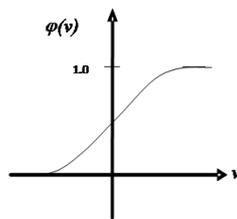


Figura 5.5 - Função Logística Sigmóide
 Fonte: Adaptada de Carvalho et al. (2000).

A função tangente hiperbólica, assim como a função logística, possui a forma de s, assumindo valores entre 1 e -1 , sendo representada por:

$$f_v = \frac{1 - e^{(-av)}}{1 + e^{(av)}} \quad (5.4)$$

A Figura 5.6 apresenta o gráfico da função tangente hiperbólica.

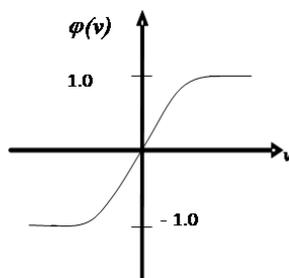


Figura 5.6 - Função Tangente Hiperbólica
Fonte: Adaptada de Carvalho et al. (2000).

5.4 Arquiteturas de Rede

As arquiteturas de redes neurais podem ser classificadas em diversos grupos, de acordo com suas características e propriedades. Haykin (2001) divide as redes neurais em três classes de arquiteturas diferentes, sendo compostas por: número de camadas, números de neurônios em cada camada, tipo de conexão entre os neurônios e topologia da rede. As classes de arquiteturas de redes neurais são descritas a seguir.

5.4.1 Redes Alimentadas adiante com camada única

A primeira classe de uma rede neural é alimentada adiante com camada única, sendo o modelo mais antigo de redes neurais. Neste tipo de rede, os neurônios estão organizados na forma de camadas, onde possui uma camada de entrada que se projeta sobre uma camada de saída de neurônios, os quais estão conectados por pesos sinápticos às entradas. Podem ser treinados por um algoritmo de aprendizagem simples, o qual irá calcular os erros entre a saída dos dados obtidos durante o treinamento da rede e a saída desejada. Este tipo de rede neural é utilizado para classificação de padrões linearmente separáveis. Na Figura 5.7 é ilustrado um modelo de rede neural de camada única, com três nós na camada de entrada e na de saída:

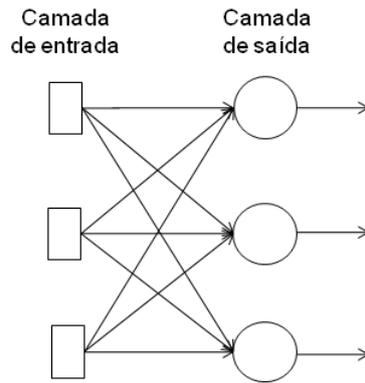


Figura 5.7 - Rede Camada Única
 Fonte: (HAYKIN, 2001)

5.4.2 Redes alimentadas com Múltiplas Camadas

A segunda classe de uma rede neural é alimentada adiante com múltiplas camadas. Este tipo de rede é uma extensão do modelo da rede com camada única, que se distingue pela presença de uma ou mais camadas intermediárias, de modo que, cada neurônio em uma camada possui conexões diretas a neurônios da próxima camada, com a função de intervir entre a entrada externa e a saída da rede. Este tipo de rede neural é utilizado para solução de problemas não linearmente separáveis.

Na Figura 5.8 mostra um exemplo de uma rede neural com múltiplas camadas, sendo a primeira camada referente às entradas, a segunda camada que não têm contato com o meio externo, é chamada de camada intermediária ou camada oculta e a última camada são as saídas. É uma rede totalmente conectada, devido a cada neurônio de uma camada da rede está conectado a todos os neurônios da camada adjacente seguinte (HAYKIN, 2001).

5.4.3 Redes Recorrentes

Uma rede neural recorrente pode consistir de uma camada de neurônios, em que cada neurônio alimenta seu sinal de saída de volta para as entradas de todos os outros neurônios. A presença de realimentação de informação permite a criação de representação interna e um dispositivo de memória capaz de processar e armazenar informações temporais e sinais sequenciais (HAYKIN, 2001).

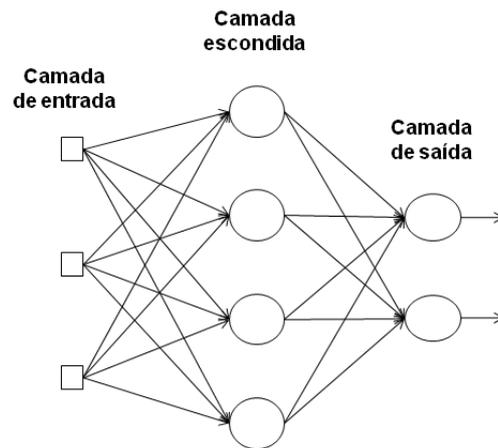


Figura 5.8 - Rede alimentada adiante totalmente conectada com uma camada oculta
 Fonte: (HAYKIN, 2001)

A Figura 5.9 apresenta uma arquitetura de uma rede recorrente, a qual consiste em uma única camada de neurônios com cada neurônio alimentando seu sinal de saída de volta para as entradas de todos os outros neurônios. Na Figura 5.9 não apresenta os laços de auto-realimentação na rede, a auto-realimentação se refere a uma situação onde a saída de um neurônio é realimentada para sua própria entrada (HAYKIN, 2001):

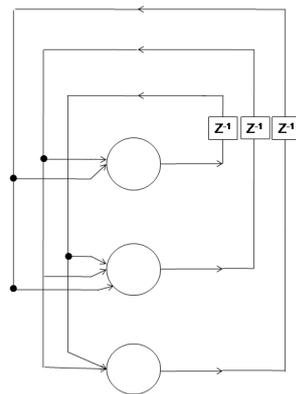


Figura 5.9 - Rede recorrente
 Fonte: (HAYKIN, 2001)

5.5 Processos de Aprendizagem

A principal propriedade de uma rede neural está na habilidade de incorporar conhecimento a partir de seu ambiente por meio de um processo iterativo de ajuste dos pesos sinápticos das conexões e níveis de bias.

No contexto de redes neurais, a aprendizagem está relacionada com a melhoria do desempenho do sistema. Isto pode ser alcançado pelo processo de ajuste dos pesos das conexões entre as unidades de processamento que guardam ao final do processo o conhecimento que a rede adquiriu do ambiente em que está operando. Existem dois paradigmas de aprendizagem, que dependem do ambiente externo onde a rede opera: o aprendizado supervisionado e o aprendizado não-supervisionado (HAYKIN, 2001).

5.5.1 Aprendizagem Supervisionada

No aprendizado supervisionado, também conhecido de aprendizagem com um professor, a entrada e a saída desejada são fornecidas por um supervisor externo. O objetivo é ajustar os parâmetros da rede, de forma a encontrar uma ligação entre os pares de entrada e saída fornecidos. Para este tipo de treinamento os padrões de entradas são apresentados à rede, a saída é comparada com a resposta desejada e os pesos das conexões são ajustados de modo que o erro seja minimizado. É comum a utilização da soma dos erros quadráticos de todas as saídas, como medida de desempenho das redes neurais, além como função de custo a ser minimizada pelo algoritmo de treinamento. A Figura 5.10 ilustra o processo de aprendizagem supervisionada.

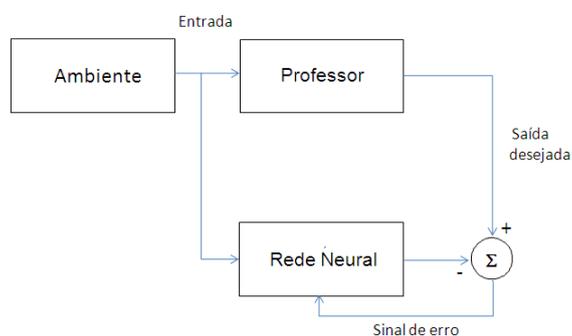


Figura 5.10 - Aprendizagem supervisionada
Fonte: Adaptada de Haykin (2001).

5.5.2 Aprendizagem Não-Supervisionada

No aprendizado não-supervisionado, não existe um supervisor para acompanhar o processo de aprendizado. Neste processo somente os padrões de entrada estão disponíveis na rede. No treinamento os pesos são ajustados de forma a distribuírem-se nos dados em função de características estatísticas que grupos de dados apresentam. A Figura 5.11 apresenta um exemplo do processo de aprendizagem não-supervisionada.

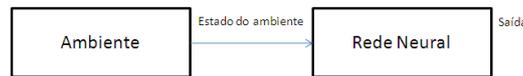


Figura 5.11 - Aprendizagem não-supervisionada
Fonte: Adaptada de Carvalho et al. (2000).

5.6 Regras de Aprendizagem

Existem várias maneiras para a implementação de aprendizagem em redes neurais artificiais. Um conjunto bem definido de regras é denominado um algoritmo de aprendizagem para a solução de um problema de aprendizagem. Haykin (2001) classifica as regras de aprendizagem em: aprendizagem por correção do erro, aprendizagem baseada em memória, aprendizagem Hebbiana, aprendizagem competitiva e aprendizagem de Boltzmann.

5.6.1 Aprendizagem por Correção do Erro

No processo de aprendizagem por correção de erros, consiste basicamente em minimizar a diferença entre as entradas ponderadas e a saída esperada. O sinal de saída gerado pela rede neural é comparado com a resposta esperada, e conseqüentemente, é produzido um sinal de erro, representado por (HAYKIN, 2001):

$$e_k = d_k - y_k \quad (5.5)$$

Em que para cada neurônio k : e_k é o erro; d_k é a saída desejada e y_k é a saída real da rede. Durante o aprendizado os pesos da rede são atualizados para garantir o menor valor de erro possível sobre os dados apresentados.

5.6.2 Aprendizagem baseada em memória

A aprendizagem baseada em memória, as experiências passadas são armazenadas em uma estrutura de memória contendo N exemplos de pares entrada-saída classificados corretamente em (HAYKIN, 2001):

$$(x_i, d_i)_{i=1}^N \quad (5.6)$$

onde x_i representa o vetor de entrada e d_i representa a resposta desejada.

Na aprendizagem baseada em memória existe um classificador pela busca dos dados dentro de uma vizinhança ou vizinhos mais próximos. Para isto, é necessário identificar os padrões classificados que se encontram mais próximos do vetor de teste X_{teste} para um número inteiro e atribuir o vetor de teste X_{teste} a classe que está mais frequentemente representada nos vizinhos mais próximos do vetor.

A regra de aprendizagem baseada em memória constitui de dois elementos essenciais: o critério utilizado para definir a vizinhança local do vetor de teste X_{teste} e a regra de aprendizagem aplicada aos exemplos de treinamento em uma vizinhança local de X_{teste} .

5.6.3 Aprendizagem Hebbiana

A aprendizagem Hebbiana é baseada na teoria de Hebb, sendo a aprendizagem mais antiga e mais conhecida de todas as regras de aprendizagem. Hebb propõe que o peso de uma conexão sináptica seja ajustado se houver sincronismo entre os níveis de atividade das entradas e saídas. Quando dois neurônios, em lados distintos da sinapse, são ativados simultaneamente, então deve haver reforço da sinapse. Entretanto, se os neurônios forem ativados assincronamente, então a sinapse será enfraquecida ou mesmo eliminada. Existem quatro mecanismos ou propriedades fundamentais que caracterizam uma sinapse (HAYKIN, 2001):

- Mecanismo Dependente do Tempo: Este se refere ao fato de que as modificações em uma sinapse Hebbiana estão sujeitas ao tempo exato de ocorrência entre os sinais pré-sinápticos e pós-sinápticos;
- Mecanismo Local: Uma sinapse é um local de transmissão onde sinais por-

tadores de informação estão em contiguidade espaço-temporal. Esta informação localmente disponível é utilizada por uma sinapse Hebbiana para produzir uma modificação sináptica local a qual é específica para a entrada;

- Mecanismo Interativo: no aprendizado Hebbiano, não se pode analisar atividades pré-sinápticas de forma isolada, ou seja, qualquer modificação na sinapse Hebbiana depende da interação em ambos os lados da sinapse (pré e pós- sinápticas) (HAYKIN, 2001);
- Mecanismo Conjuncional ou Correlacional: a sinapse Hebbiana pode ser chamada de sinapse conjuncional devido à ocorrência conjunta de atividades pré e pós-sinápticas ser suficiente para que haja uma modificação. Além disso, pode também ser chamada de sinapse correlacional visto que uma correlação entre estas mesmas atividades também é suficiente para gerar mudanças (HAYKIN, 2001).

5.6.4 Aprendizagem Competitiva

Neste tipo de aprendizagem os neurônios de saída competem entre si para serem ativados. Um único neurônio de saída é ativado em um determinado instante e seus pesos atualizados no treinamento. Tal característica torna a aprendizagem competitiva apropriada para descobrir características marcantes que podem ser utilizadas para classificar um conjunto de padrões de entrada. As unidades de entrada são diretamente conectadas às unidades de saída. A unidade de saída com maior ativação inicial terá maior chance de vencer a disputas em relação as outras unidades (CARVALHO et al., 2000).

Existem três elementos básicos em uma regra de aprendizagem competitiva (HAYKIN, 2001):

- Um conjunto de neurônios iguais, com exceção de alguns pesos sinápticos distribuídos aleatoriamente, que respondam diferentemente a um dado conjunto de entrada;
- Um limite imposto sobre a robustez de cada neurônio; e
- Um mecanismo que permita que os neurônios possam competir pelo direito de responder a um subconjunto de entradas, de forma que somente

um neurônio de saída seja ativado em um determinado instante (HAYKIN, 2001).

5.6.5 Aprendizagem de Boltzmann

O Aprendizado de Boltzmann é um algoritmo de aprendizagem estocástico que realiza o ajuste dos pesos baseando-se na mecânica estatística. A rede neural que utiliza este aprendizado é denominada máquina de Boltzmann.

Neste aprendizado, os neurônios constituem uma estrutura recorrente e operam de uma maneira binária, em que podem estar em um estado ligado, representado por 1, ou em um estado desligado, representado por -1 . Os neurônios se dividem em dois grupos: os visíveis e os ocultos. Os visíveis oferecem uma interface entre a rede e o ambiente em que ela opera, e os neurônios ocultos sempre operam livremente (HAYKIN, 2001).

5.7 Perceptron

O Perceptron foi proposto por Rosenblatt em 1958, como um modelo de rede neural com aprendizagem supervisionada. Os neurônios estão dispostos em uma única camada com pesos e bias ajustáveis. Rosenblatt provou que se os padrões usados para treinar o perceptron são retirados de duas classes linearmente separáveis, então o algoritmo do perceptron converge e posiciona a superfície de decisão na forma de um hiperplano entre as duas classes. Um único neurônio perceptron limita-se a realizar classificação de padrões de apenas duas classes (HAYKIN, 2001).

No processo de aprendizado, espera-se obter o valor do incremento Δw a ser aplicado ao vetor de pesos w sendo que o seu valor atualizado $w(n+1) = w(n) + \Delta w(n)$ esteja mais próximo da solução desejada do que $w(n)$ (CARVALHO et al., 2000).

A Equação 5.7 apresenta o erro entre a resposta esperada e a resposta produzida pela rede:

$$e_k(n) = d_k(n) - y_k(n) \quad (5.7)$$

de modo que $d_k(n)$ e $y_k(n)$ são os sinais de saída desejado e obtido no neurônio k no instante n .

Os cálculos para os novos pesos e bias são obtidos por:

$$w(n + 1) = w(n) + \eta x(n) \quad (5.8)$$

$$b(n + 1) = b(n) + \eta e(n)(-1) \quad (5.9)$$

em que, $w(n)$ representa os pesos, $x(n)$ são as entradas, $e(n)$ é o erro, $b(n)$ são os bias e η a taxa de aprendizagem.

Este tipo de rede neural é utilizado apenas para solução de problemas linearmente separáveis.

5.8 Perceptron de Múltiplas Camadas

As redes Perceptron de Múltiplas Camadas (MLP) surgiram como alternativa para solução de problemas não linearmente separáveis e têm sido aplicadas com sucesso para solução de diversos problemas complexos, através do seu treinamento supervisionado usando o algoritmo de retropropagação de erro (*error backpropagation*) baseado na regra de aprendizagem por correção do erro (HAYKIN, 2001).

As redes MLP são compostas basicamente por: uma camada de entrada, onde os padrões são apresentados à rede; uma ou mais camadas intermediárias, as quais trabalham como um reconhecedor de características que ficam armazenadas nos pesos sinápticos e são responsáveis pela maior parte do processamento e uma camada de saída, onde os resultados são apresentados. Os sinais de entrada são propagados camada por camada, até que o vetor de saída seja obtido na última camada. A rede exibe um alto grau de conectividade, determinadas pelas sinapses da rede. O processo de aprendizado termina quando se atinge um erro desejado ou quando se atinge um número máximo de épocas de aprendizado (HAYKIN, 2001).

5.8.1 Arquitetura da rede

A arquitetura de uma rede MLP consiste no arranjo topológico das unidades de processamento dos neurônios com os respectivos valores de pesos associados às conexões. Os pesos sinápticos são ajustados para fazer com que a resposta real da rede se aproxima da resposta desejada, em um sentido estatístico (HAYKIN, 2001).

A Figura 5.12 representa a arquitetura de uma rede MLP, com uma camada de entrada, uma escondida e uma camada de saída.

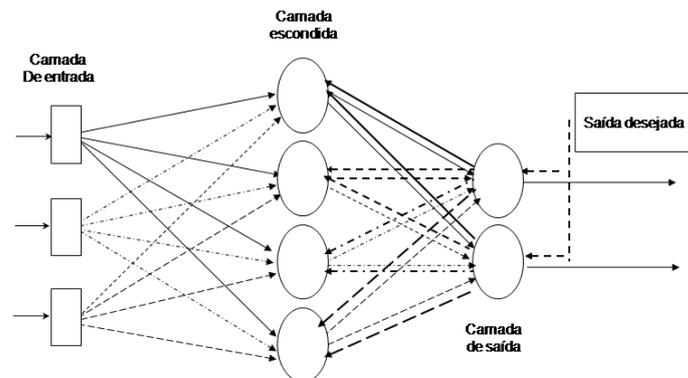


Figura 5.12 - Arquitetura da Rede Perceptron de Múltiplas Camadas

O algoritmo de retropropagação de erro consiste em dois passos:

- A propagação: consiste na ativação da rede, os pesos não são alterados, os sinais na camada são propagados camada por camada, até que o vetor de saída seja obtido na última camada. Durante o passo de propagação os pesos sinápticos são todos fixos (HAYKIN, 2001).
- Retropropagação: passo para trás, iniciando na camada de saída, calcula-se o gradiente local do neurônio, os pesos são ajustados de acordo com a regra delta. Onde a resposta real da rede é subtraída de uma resposta desejada para produzir um sinal de erro, que é então propagado para trás, através da rede. No passo de retropropagação os pesos sinápticos são ajustados de acordo com a regra de correção de erro para produzir um sinal de erro (HAYKIN, 2001).

O funcionamento do neurônio da rede MLP é semelhante ao neurônio de McCulloch e Pitts. Porém a ativação de cada neurônio é dada por uma função não-linear que deve ser uma função diferenciável. Neste tipo de rede normalmente são usadas as funções sigmóide e a tangente hiperbólica.

O número de camadas e o número de neurônios em cada camada são definidos empiricamente. Este número depende de uma verificação prévia nos dados e na

complexidade do problema. Ao utilizar muitos neurônios na camada escondida, pode levar a rede neural a memorizar os padrões de treinamento, ao invés de extrair informações na qual permitirão a generalização ou o reconhecimento de padrões, este problema é chamado de *overfitting*. Uma forma de evitar o *overfitting* é selecionar um subconjunto dos dados formando um conjunto de testes para o critério de parada antecipada no treinamento das redes. Por outro lado, uma rede com poucos neurônios na camada escondida pode não ser capaz de realizar o mapeamento desejado, o que é chamado de *underfitting* (CARVALHO et al., 2000).

5.8.2 Algoritmo de Retropropagação

O algoritmo de retropropagação é um algoritmo de aprendizado supervisionado, baseado na correção de erros. Sua função é encontrar na superfície de erro um mínimo global, para minimização do erro da rede. O algoritmo de retropropagação é considerado popular no contexto de redes neurais, devido sua capacidade de armazenamento de informações nos pesos sinápticos da rede.

As principais notações usadas no algoritmo de retropropagação são apresentadas a seguir (HAYKIN, 2001):

- Os índices i, j , e k , representam os neurônios da rede neural; o neurônio j se encontra em uma camada à direita do neurônio i , e o neurônio k se encontra em uma camada à direita do neurônio j , quando o neurônio j é uma unidade oculta;
- O símbolo $\varepsilon(n)$ representa a soma dos erros quadráticos ou a energia do erro na iteração n ;
- O símbolo $e_j(n)$ se refere ao sinal de erro na saída do neurônio j , para a iteração n ;
- A notação $d_j(n)$ representa a saída desejada para o neurônio j e é usada para calcular $e_j(n)$;
- A notação $y_j(n)$ se refere a saída do neurônio j , na iteração n ;
- O símbolo $w_{ij}(n)$, representa o peso sináptico da saída do neurônio i para a entrada do neurônio j , na iteração n ;

- A notação $v_j(n)$ constitui no sinal aplicado à função de ativação associado com o neurônio j ;
- O símbolo $\varphi'(\cdot)$ representa a função de ativação, que descreve a relação funcional de entrada-saída da não-linearidade associada ao neurônio j ;
- A notação η , se refere ao parâmetro da taxa de aprendizagem;
- O símbolo δ , consiste no gradiente local.

O sinal de erro na saída do neurônio j , na iteração n é definido por:

$$e_j(n) = d_j(n) - y_j(n) \quad (5.10)$$

em que j é o neurônio em um nó de saída.

O valor instantâneo da energia total do erro para o neurônio j é obtido somando os termos $\frac{1}{2}e_j^2(n)$ de todos os neurônios da camada de saída, os quais são os únicos neurônios visíveis em que os sinais de erro podem ser calculados diretamente, como descrito em:

$$\varepsilon(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \quad (5.11)$$

Onde C é um conjunto que inclui todos os neurônios da camada de saída da rede. Considerando que N representa o número total de padrões contidos no conjunto de treinamento. A energia média do erro quadrático é obtida somando-se os ε_{med} para todos os n e então normalizando em relação ao tamanho do conjunto N , como apresentado por:

$$\epsilon_{med} = \frac{1}{N} \sum_{n=1}^N \varepsilon(n) \quad (5.12)$$

A energia média do erro quadrático ε_{med} é uma função de todos os parâmetros livres (por exemplo, os pesos e bias) da rede. Dado um conjunto de treinamento, ε_{med} representa a função de custo como uma medida do desempenho de aprendizagem.

O objetivo do processo de aprendizagem é ajustar os pesos sinápticos e bias para minimizar o erro. Os pesos são ajustados usando a regra delta pela equação (HAYKIN, 2001):

Os pesos são ajustados usando a regra delta pela equação:

$$w_j(n+1) = w_j(n) + \Delta w_j(n) \quad (5.13)$$

Para minimização do erro, é aplicada uma correção $\Delta w_j(n)$ ao peso sináptico de acordo com a Regra Delta (HAYKIN, 2001):

$$\Delta w_{ji}(n) = -\eta \frac{\partial \epsilon(n)}{\partial w_{ji}(n)} \quad (5.14)$$

Onde η é o parâmetro da taxa de aprendizagem do algoritmo de retropropagação do erro. O uso do sinal negativo na Equação 5.14 indica a descida do gradiente no espaço de pesos, como por exemplo, busca uma direção para a mudança de peso que o reduza o valor de $\epsilon(n)$.

$$\Delta w_{ji}(n) = -\eta \delta_j(n) y_i(n) \quad (5.15)$$

Onde o gradiente local $\delta_j(n)$ é definido por:

$$\delta_j(n) = -\frac{\partial \epsilon(n)}{\partial v_j(n)} = \frac{\partial \epsilon(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} = -e_j(n) \varphi'_j(v_j(n)) \quad (5.16)$$

O gradiente na camada escondida é dado pela equação:

$$\delta_j(n) = \varphi'_j(v_j(n)) \sum_k \delta_k(n) w_{kj}(n) \quad (5.17)$$

O gradiente local aponta para as modificações necessárias nos pesos sinápticos. A partir da Equação 5.16, o gradiente local $\delta_j(n)$ para o neurônio de saída j é igual ao produto do sinal de erro $e_j(n)$ correspondente para aquele neurônio pela derivada

$\varphi'_j(v_j(n))$ da função de ativação associada (HAYKIN, 2001).

O cálculo do ajuste de peso $\Delta w_{ji}(n)$ é o sinal de erro $e_j(n)$ na saída do neurônio j . Diante disso, é possível identificar dois casos distintos. No caso 1, o neurônio j é um nó de saída, o qual é suprido com uma resposta desejada. E no caso 2, o neurônio j é um nó oculto, onde compartilham a responsabilidade por qualquer erro cometido na saída da rede (HAYKIN, 2001).

No caso 1 o neurônio j é um nó de saída, ou seja, quando o neurônio j está localizado na camada de saída da rede, este é suprido com uma resposta desejada. Desta maneira, utiliza-se a Equação 5.10 para calcular o sinal de erro $e_j(n)$ associado com este neurônio. Calculado o $e_j(n)$, determina-se diretamente o gradiente local, usando a Equação 5.16.

No caso 2, onde o neurônio j é um nó oculto, não existe uma resposta especificada para aquele neurônio. Portanto, o sinal de erro para um neurônio oculto deve ser recursivo, em termos dos sinais de erro de todos os neurônios aos quais os neurônios ocultos estão conectados (HAYKIN, 2001). De acordo com a Equação 5.16, é possível redefinir o gradiente local para o neurônio oculto j como:

De acordo com a Equação 5.16, é possível redefinir o gradiente local para o neurônio oculto j como:

$$\delta_j(n) = -\frac{\partial \varepsilon(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} = \frac{\partial \varepsilon(n)}{\partial y_j(n)} \varphi'_j(v_j(n)) \quad (5.18)$$

onde o neurônio j é oculto.

A Equação 5.19 o neurônio k é um nó de saída:

$$\varepsilon(n) = \frac{1}{2} \sum_{k \in C} e_k^2(n) \quad (5.19)$$

O neurônio k é um nó de saída, e este índice k é utilizado no lugar do índice j , para evitar a confusão com o uso do índice j , o qual se refere ao neurônio oculto no caso 2 (HAYKIN, 2001).

Diferenciando a Equação 5.19 em relação ao sinal funcional $y_i(n)$ tem-se:

$$\frac{\partial \epsilon(n)}{\partial y_j(n)} = \sum_k e_k \frac{\partial e_k(n)}{\partial y_j(n)} = \sum_k e_k \frac{\partial e_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial y_j(n)} \quad (5.20)$$

Na Equação 5.21, têm-se a derivada parcial desejada:

$$\frac{\partial \epsilon(n)}{\partial y_j(n)} = - \sum_k e_k(n) \varphi'_k(v_k(n)) w_{kj}(n) = - \sum_k \delta_k(n)_k w_{kj}(n) \quad (5.21)$$

Onde, na segunda linha, é utilizada a definição do gradiente local $\delta_k(n)$, com índice k substituído por j . Finalmente, a partir das Equação 5.21 e Equação 5.18, tem-se a fórmula de retropropagação para o gradiente local $\delta_j(n)$ como descrito (HAYKIN, 2001):

$$\delta_j(n) = \varphi'_k(v_k(n)) \sum_k \delta_k(n)_k w_{kj}(n) \quad (5.22)$$

em que o neurônio j é oculto.

5.9 Redes Recorrentes

As redes recorrentes são definidas como aquelas que possuem um ou mais laços de realimentação e que proporcionam um comportamento dinâmico. Existem dois modos de aplicar a realimentação: a primeira realimentação local ao nível de um único neurônio dentro da rede, e a realimentação global, a qual abrange toda a rede (HAYKIN, 2001).

Existem dois tipos de redes recorrentes: aquelas em que o padrão de entrada é fixo e a saída caminha, dinamicamente, para um estado estável e aquelas em que ambas, entrada e saída, variam com o tempo, sendo estas últimas mais gerais e utilizadas com maior frequência (CARVALHO et al., 2000).

As redes recorrentes ampliam a potencialidade de modelagem em dados temporais, e são utilizadas com sucesso em processamento de dados temporais, tanto para reconhecimento de padrões, quanto em previsões de séries temporais.

Uma rede neural é considerada dinâmica quando possui memória. Existem duas maneiras de fornecer memória a uma RNA, a primeira forma é utilizar atraso no tempo,

tais como as técnicas de *Time Delay Neural Network* (TDNN) e *FIR Multilayer Perceptron*. A segunda abordagem consiste na utilização de redes neurais recorrentes, tais como *Backpropagation Through Time*, *Real-Time Recurrent Learning*, *Cascade Correlation recorrente*, redes de Elman e Jordan (CARVALHO et al., 2000).

Os modelos de redes recorrentes mais utilizados para predição de séries temporais são as redes de Elman e Jordan. Estas foram utilizadas neste trabalho, para realização do processo de previsão climática. Estas redes propõem o uso de uma nova camada, chamada de camada de contexto, a qual é responsável em armazenar as informações temporais.

A Figura 5.13 apresenta uma arquitetura de uma rede recorrente, com os laços de realimentação envolvendo o uso de ramos particulares compostos de operadores de atraso unitário, resultando em um comportamento dinâmico não-linear (HAYKIN, 2001):

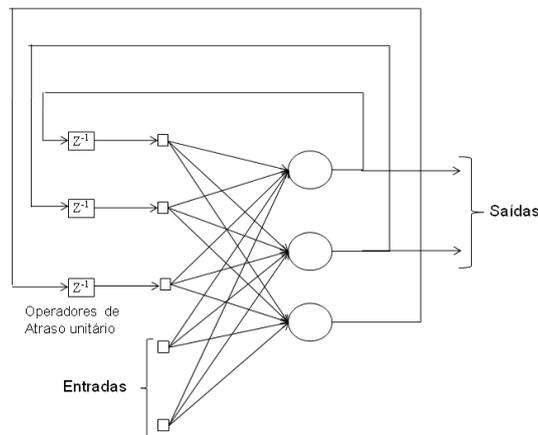


Figura 5.13 - Rede recorrente com neurônios ocultos
Fonte: (HAYKIN, 2001)

5.9.1 Rede de Elman

A rede de Elman contém conexões recorrentes dos neurônios escondidos para uma camada de contexto que consiste em atrasos unitários. Estas unidades de contexto são a memória da rede usada para armazenar o valor de saída das camadas intermediárias por um passo de tempo, e as realimenta de volta para a camada de entrada.

Diante disso, esta rede tem sido utilizada com sucesso para aprender as características temporais de um sinal que represente a uma série histórica, com realimentação global (HAYKIN, 2001).

A arquitetura da rede de Elman incorpora um perceptron de múltiplas camadas ou parte dele, e explora a capacidade de mapeamento não-linear da MLP, de tal forma que esta arquitetura pode ser treinada pelo algoritmo backpropagation convencional (HAYKIN, 2001).

A rede de Elman é considerada parcialmente recorrente, devido os elos de realimentação ocorrem entre a saída e a entrada da primeira camada oculta. Essa recorrência ocorre através da unidade de contexto, onde possui um atraso do tipo $n - 1$, o qual armazena a saída da primeira camada oculta para um passo de tempo.

As redes de Elman são compostas por:

- Camada de entrada: composta por neurônios de armazenamento que recebem um sinal externo e o propagam sem modificá-los;
- Camadas intermediárias: podem ter funções de ativação lineares e não lineares;
- Camada de contexto: são utilizadas para memorizar as ativações dos neurônios da camada intermediária e podem ser usadas com atraso no tempo;
- Camada de saída: consiste de neurônios lineares cujas saídas são somas de seus respectivos sinais de entrada.

Na Figura 5.14 é apresentada uma arquitetura da rede recorrente de Elman, a qual possui quatro camadas: uma camada de entrada, uma camada intermediária, a camada de contexto e a camada saída. A realimentação ocorre da saída de cada neurônio da camada escondida para as unidades de contexto. As unidades de contexto são totalmente conectadas com todos os neurônios da camada escondida.

5.9.2 Rede de Jordan

A rede recorrente de Jordan, assim como a rede de Elman, possui realimentação global, com treinamento de forma aproximada ao algoritmo de retropropagação, sendo utilizada em aplicações envolvendo informações temporais.

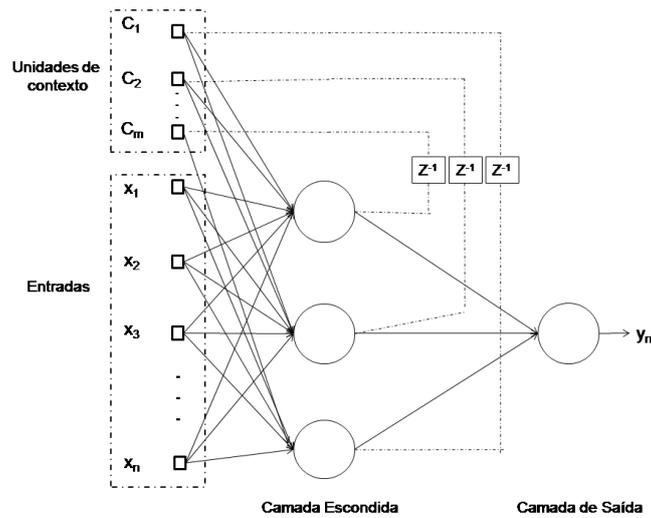


Figura 5.14 - Rede recorrente de Elman
 Fonte: (CARVALHO et al., 2000)

Enquanto a rede de Elman possui recorrência a partir da camada oculta, na rede de Jordan a recorrência é feita a partir da camada de saída para a camada de contexto, em que pode associar a mesma entrada a várias sequencias de saída em função da unidade de contexto.

Na Figura 5.15 é mostrada uma arquitetura da rede recorrente de Jordan, com apenas uma camada intermediária. A realimentação ocorre da camada de saída para as unidades de contexto.

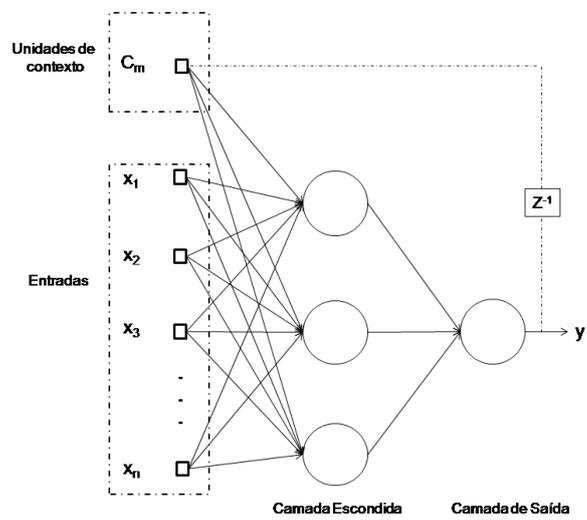


Figura 5.15 - Rede recorrente de Jordan
 Fonte: (CARVALHO et al., 2000)

6 METODOLOGIA

A metodologia adotada nessa dissertação consiste em duas abordagens para realização do processo de previsão climática: na primeira os modelos de redes neurais são treinados com todas as variáveis disponíveis; e na segunda abordagem, os dados disponíveis para treinamento são processados para gerar reduções (em uma abordagem de mineração de dados) que são usadas como dados de treinamento para as redes neurais. A fase de mineração utiliza a TCA, que identifica os atributos mais significativos para o processo de previsão climática, como aqueles de maior ocorrência, segundo a relação de indiscernibilidade. Os resultados de previsão obtidos com as redes neurais treinadas com os dados completos e com os dados reduzidos serão comparados com a resposta esperada que está disponível nos dados obtidos como descrito na sequência.

A área de estudo para realização dos experimentos abrange as cinco regiões do Brasil, Nordeste (NE), Sudeste (SE), Norte (N), Centro Oeste (CO) e Sul (S), como mostrado na [Figura 6.1](#). As coordenadas geográficas estão compreendidas entre as latitudes $[5^{\circ}\text{N}, 35^{\circ}\text{S}]$ e longitudes $[70^{\circ}\text{W}, 35^{\circ}\text{W}]$.

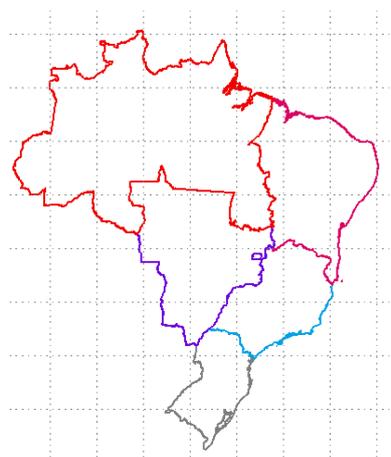


Figura 6.1 - Área de estudo

Os dados utilizados nos experimentos foram coletados da base de dados de re-análise do NCEP/NCAR (National Centers for Environmental Prediction / The National Center for Atmospheric Research) [<http://ww.ncep.noaa.gov>]. NOAA é a principal organização do National Weather Service (Serviço Nacional de Me-

teorologia dos Estados Unidos), que é responsável em realizar pesquisas meteorológicas e climáticas, para observar e entender o ambiente físico da Terra, a fim de, melhorar as previsões meteorológicas e climáticas em escala global a local. O NOAA mantém grandes conjuntos de dados de reanálise, para o uso em diagnósticos do meio ambiente mundial. Esses dados são criados a partir de assimilações de observações climáticas, provenientes de diferentes fontes, tais como: navios, aviões, satélites, estações terrestres, radar entre outras. Atualmente, o NOAA disponibiliza estes conjuntos de dados para o público através do site [<http://www.esrl.noaa.gov/psd/data/gridded/tables/monthly.html>] em formato netCDF. O conjunto de dados é mantido e atualizado utilizando observações em tempo real (NOAA, 2010).

Os dados coletados compreende uma janela de tempo de 31 anos entre janeiro de 1980 e dezembro de 2009, em uma área contida entre as latitudes [10°N, 35°S] e longitudes [80°W, 30°W], referente à América do Sul. A resolução espacial, em ambas as dimensões da grade é de 2.5° e resolução temporal (t) de 1 mês.

Inicialmente foram realizados estudos e experimentos sobre as regiões Nordeste e Sul do Brasil, utilizando um período de tempo de 21 anos, entre janeiro de 1980 a dezembro de 2000, dos quais foram selecionados 18 anos (janeiro de 1980 a dezembro de 1997) para o treinamento dos modelos de redes neurais e como entrada para o processamento pela TCA. Os outros 3 anos (janeiro de 1998 a dezembro de 2000) foram usados para a validação dos modelos. As variáveis contidas nesta base de dados estão descritas na [Tabela 6.1](#):

Posteriormente foram realizados estudos e experimentos sobre as cinco regiões do Brasil, com dados correspondentes ao período entre janeiro de 2000 e dezembro de 2009, dos quais foram selecionados 7 anos (janeiro de 2000 a dezembro de 2006) para realização dos treinamentos das redes neurais e como entrada para o processamento pela TCA e para a validação dos modelos foram utilizados os 3 anos restantes (janeiro de 2007 a dezembro de 2009). As variáveis contidas nesta base de dados são descritas na [Tabela 6.2](#):

Tabela 6.1 - Variáveis para o conjunto de dados de 1980-2000.

	Variável	Descrição (nível)	Unidade
1	airt	Temperatura do ar (superfície)	°C
2	div	divergência (1 s)	1/s
3	prec	Precipitação (superfície)	mm/dia
4	shum	Umidade Específica (2 m)	kg/kg
5	pres	Pressão (Superfície)	mb
6	temp	Temperatura (0.1- 2 m)	°C
7	u300	Vento Zonal (300 hPa)	m/s
8	u500	Vento Zonal (500 hPa)	m/s
9	u850	Vento Zonal (850 hPa)	m/s
10	v300	Vento Meridional (300 hPa)	m/s
11	v500	Vento Meridional (500 hPa)	m/s
12	v850	Vento Meridional (850 hPa)	m/s

Tabela 6.2 - Variáveis para o conjunto de dados de 2000-2009.

	Variável	Descrição (nível)	Unidade
1	temp	Temperatura do ar (superfície)	°C
2	prec	Precipitação (superfície)	mm/dia
3	shum	Umidade Específica (2 m)	kg/kg
4	pres	Pressão (Superfície)	mb
5	u300	Vento Zonal (300 hPa)	m/s
6	u500	Vento Zonal (500 hPa)	m/s
7	u850	Vento Zonal (850 hPa)	m/s
8	v300	Vento Meridional (300 hPa)	m/s
9	v500	Vento Meridional (500 hPa)	m/s
10	v850	Vento Meridional (850 hPa)	m/s

Os significados das variáveis meteorológicas são (INMET, 2010):

- Temperatura do ar em superfície: é a subtração entre o campo de temperatura do ar em superfície em um dado horário e o campo 24 horas antes. Os valores positivos indicam aumento de temperatura do ar em 24 horas. Por exemplo, diminuição considerável de temperatura do ar sugere que o modelo está prevendo a entrada de uma massa de ar frio sobre a região, enquanto que elevado aumento de temperatura do ar indica incursão de uma massa de ar quente sobre a região;
- Umidade: quantidade de vapor de água no ar. Tipos de umidade incluem: umidade absoluta, umidade e umidade específica;
- Vento: os campos a serem visualizados correspondem à magnitude (m/s) e direção (graus) dos ventos (nos níveis de pressão de 1000, 950, 850, 700, 500, 400, 300, 250, 200, 150, 100 e 50 hPa);
- Precipitação: todas as formas de água, líquida ou sólida, que caem das nuvens. Podem ser na forma de aguaceiros, chuva, chuveiro, granizo;
- Divergência: é a medida de expansão de um campo vetorial (qualquer quantidade física que varia no espaço tri-dimensional). Em Meteorologia, devido à predominância de movimentos horizontais, a divergência usualmente refere-se à divergência horizontal bi-dimensional do campo velocidade (unidade indicada 10^{-5} s^{-1}). A divergência horizontal do campo velocidade é relacionado às variações de movimento vertical e pressão, através das equações da continuidade e equações do movimento. Convergência é o negativo de divergência, sendo a contração do campo vetorial;
- Pressão: é a força por unidade de área, exercida pelo peso da atmosfera, sobre um ponto localizado na superfície da Terra ou acima da mesma.

O grau de relação entre as variáveis meteorológicas é mostrado na Figura 6.2. Observa-se na Figura 6.2(a) que as variáveis vento zonal u300 hPa, u500 hPa e u850 hPa estão bem correlacionadas. Assim como a variável vento meridional na Figura 6.2(b), possui uma forte associação entre os níveis 300 hPa e 500 hPa. E a Figura 6.2(c) mostra que as variáveis temperatura do ar e temperatura da superfície possuem alta correlação.

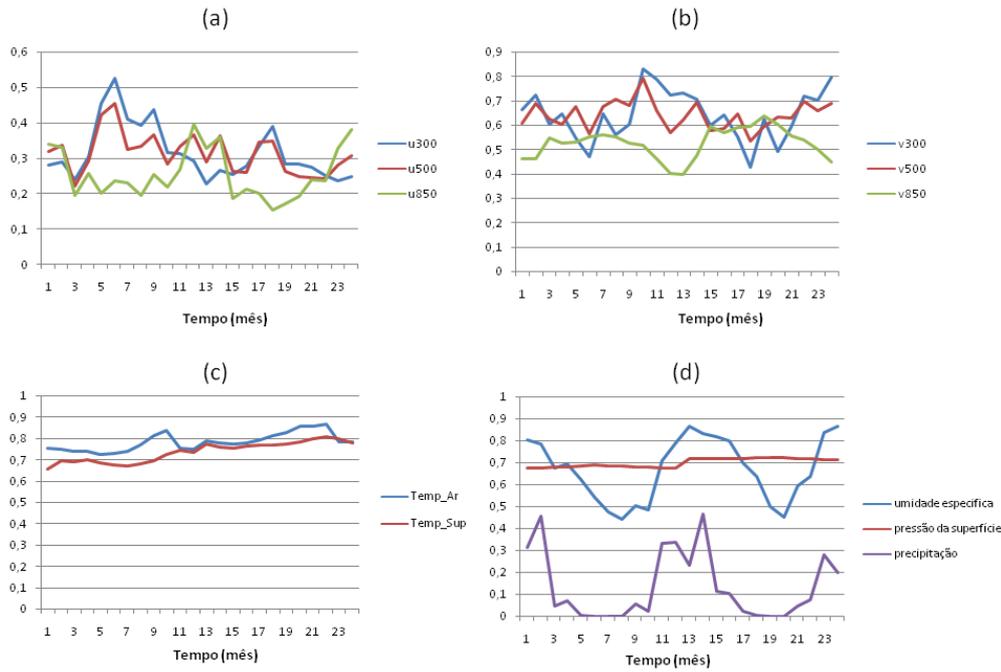


Figura 6.2 - Análise dependências dos dados.

Os modelos de redes neurais utilizados para desempenhar a previsão de precipitação foram: a MLP e os modelos recorrentes de Elman e Jordan, os quais utilizam o algoritmo de retropropagação do erro.

As arquiteturas dos modelos de redes usadas neste trabalho, foram configuradas durante testes preliminares, variando-se o número de neurônios na camada escondida e o número máximo de épocas de treinamento. Os testes conduzidos proporcionaram uma arquitetura constituída por apenas uma camada escondida com 16 neurônios, uma camada de saída, submetida ao máximo de 10000 épocas, onde cada neurônio foi configurado com a função de ativação do tipo logística sigmoidal.

A métrica para quantificar o desempenho da previsão foi o erro quadrático médio E dado por:

$$E = \frac{1}{N} \sum_{k=1}^N (y_k - \hat{y}_k)^2 \quad (6.1)$$

em que N é o número de pontos da grade, y_k é o valor real no ponto de grade e \hat{y}_k

é a estimativa produzida pela rede neural.

Para a realização do processo de redução dos atributos, utilizou-se a ferramenta ROSETTA (ØHRN, 1999a) para calcular os atributos relevantes para o processo de previsão climática. Primeiramente os dados são discretizados, em seguida são submetidos ao algoritmo de redução para o cálculo dos conjuntos candidatos aproximados, como descrito na [Subseção 4.8.4](#). O principal parâmetro é a fração dos candidatos, cujo valor neste trabalho foi de 0.7, ou seja, todos os atributos que possuem no mínimo 70% de presença na função de indiscernimento. Esses são apresentados como os redutos desejados para o treinamento das redes neurais.

Para a visualização dos resultados utilizou-se a ferramenta GrADS (*Grid Analysis and Display System*) que é um software utilizado para visualização e análise de dados em pontos de grade, que fornece um ambiente integrado para acesso, manipulação e exibição de dados. Atualmente o GrADS é a ferramenta mais utilizada na área de pesquisa nos grandes centros de meteorologia (DOTY, 2009).

O GrADS utiliza dois arquivos para realizar a visualização, um arquivo descritor ou *ctl* e arquivo de dados no formato binário. O arquivo descritor encontra-se no formato ASCII, sendo responsável por descrever as características do arquivo de dados, que descreve informações como: as coordenadas dos dados (latitude e longitude), tempo, o número e o nome das variáveis. A grade é interpolada para uma grade mais fina usando interpolação linear antes de traçar os contornos.

Neste trabalho as saídas obtidas através dos modelos de redes neurais, encontram-se no formato de texto ASCII e requerem um pré-processamento, para ser integradas no ambiente GrADS, desta forma, os dados são convertidos para o formato binário, a partir de *scripts* desenvolvidos usando a linguagem FORTRAN.

O modelo de previsão climática sazonal desenvolvido nesta dissertação tem como função desempenhar a previsão das propriedades estatísticas de algum estado climático futuro, o qual pode ser para um mês ou uma estação. Os experimentos realizados foram feitos para estimar a variável de precipitação em uma estação adiante, ou seja, é possível criar um conjunto de possíveis estados iniciais para os modelos de redes neurais, produzindo um conjunto de previsões para cada condição inicial. Supondo que o modelo receba como condições iniciais dados referentes à estação outono, como saída terá a previsão para 1 estação adiante, que neste caso será a estimativa

de precipitação para a estação inverno, conforme segue um exemplo na Figura 6.3:

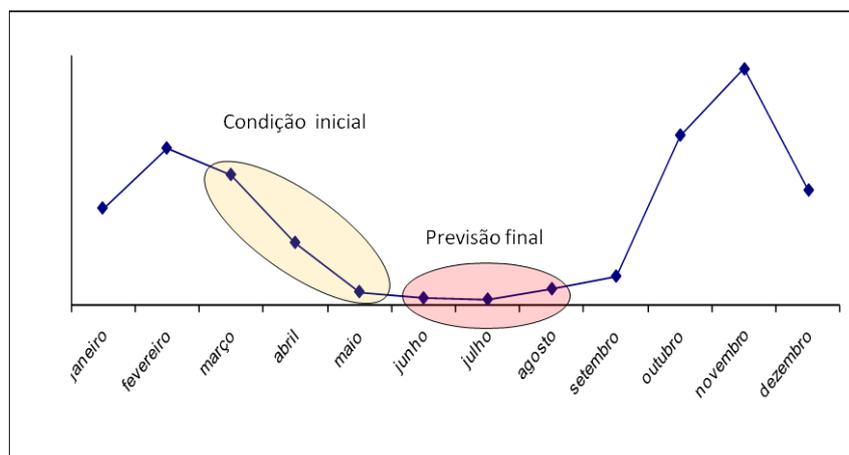


Figura 6.3 - Esquema de previsão sazonal.

7 RESULTADOS

Nesta seção são apresentados os resultados obtidos com os experimentos realizados usando dados obtidos do processo de redução pela Teoria dos Conjuntos Aproximativos - TCA. O principal objetivo desta dissertação é conceber um método para o desenvolvimento de um modelo de previsão da variável de precipitação a partir de um conjunto de dados reduzidos, com base na hipótese de que algumas variáveis são mais adequadas (ou seja, são vetores) para estimar o estado futuro da variável precipitação.

No treinamento das redes neurais a forma de apresentação dos dados para os modelos de previsão, considera duas abordagens: na primeira todos os dados disponíveis na base de dados são usados no treinamento das redes e na segunda, por meio da teoria dos conjuntos aproximativos, são inseridos os dados reduzidos, ou seja, aqueles que mais contribuem no processo de previsão climática.

Seis modelos de redes neurais foram construídos para cada uma das regiões de estudo: três dos modelos são gerados a partir do treinamento das redes MLP, Elman e Jordan, usando todas as variáveis disponíveis na base de dados, e para os outros três modelos (MLP, Elman e Jordan), foram usados os dados pré-processados pela TCA, ou seja, os dados reduzidos.

7.1 Resultados na região Nordeste em 1999

Nesta seção são mostrados os experimentos realizados sobre a região Nordeste do Brasil, entre as longitudes $[47^{\circ}\text{W}, 35^{\circ}\text{W}]$ e entre as latitudes $[0^{\circ}, -17.5^{\circ}\text{S}]$, compreendendo 48 pontos de grade, utilizando dados do período entre janeiro de 1980 e dezembro de 2000, para previsões climáticas sazonais da variável de precipitação. As variáveis usadas neste experimento são aquelas citadas na [Tabela 6.1](#) na [Capítulo 6](#).

A arquitetura da rede neural utilizada neste experimento foi escolhida em testes preliminares variando o número de neurônios na camada escondida e o número de épocas de treinamento. Após a realização de vários testes, verificou-se que o número de neurônios entre 18 e 22, com 10000 épocas de treinamento, era suficiente para a maioria dos casos testados, desta forma, escolheu-se configurações condizentes com estes números definidos empiricamente.

As variáveis reduzidas por meio do uso da TCA podem ser observadas na [Tabela 7.1](#).

Elas foram selecionadas em função da ocorrência igual ou superior a 70% de presença na função de discernibilidade e foram escolhidas como entradas para o treinamento das redes neurais. Na [Tabela 7.1](#), observa-se que dos 11 atributos mencionados na [Tabela 6.1](#), a dimensão do problema foi reduzida para 5 atributos.

Tabela 7.1 - Variáveis reduzidas.

Variável	%
airt	75%
u850	82%
u500	71%
v500	73%
v300	85%

Após a obtenção dos dois modelos de previsão, utilizando todos os dados e as reduções, estes foram generalizados (ou validados) para realização das previsões sobre os dados não utilizados no treinamento. A [Tabela 7.2](#) apresenta os erros quadráticos médios obtidos durante o processo de previsão climática.

Tabela 7.2 - Erro quadrático médio.

Erro Quadrático Médio			
Ano	Estação	RNA completo	RNA com TCA
1998	Outono	0,79293	2,71292
1998	Inverno	0,52551	0,49527
1998	Primavera	1,41907	1,0316
1998	Verão	1,54688	2,24434
1999	Outono	1,41584	0,58194
1999	Inverno	0,60117	0,6363
1999	Primavera	1,64285	1,82802
1999	Verão	1,42834	0,82173

Uma análise qualitativa dos resultados pode ser visualizada nas Figuras 7.1 a 7.4, as quais apresentam simultaneamente os resultados obtidos no processo de previsão climática, usando todos os dados disponíveis e os dados processados através da TCA. Os resultados são mostrados em um mapa criado pela ferramenta GrADS, para as quatro estações do ano de 1999.

Na Figura 7.1 são mostrados os resultados de precipitação obtidos no processo de previsão climática pelas redes neurais, para a estação outono de 1999. Observa-se que a previsão realizada com os dados processados pela TCA, mostra padrões visuais mais semelhantes àqueles presentes na Figura 7.1(a) que representa o observado (dado real esperado).

Na Figura 7.2 são apresentados os resultados de precipitação obtidos pelas redes neurais para a estação inverno de 1999. O resultado de previsão obtida com o modelo que utiliza os dados pré-processados pela TCA apresenta padrões mais semelhantes aos presentes na Figura 7.2(a) que representa os dados reais.

Na Figura 7.3 são mostrados os resultados de previsão de precipitação para a estação primavera de 1999. O resultado obtido com a rede usando os dados reduzidos através da TCA apresenta uma previsão mais compatível com o que foi observado como mostrado na Figura 7.3(a).

Na Figura 7.4 são apresentados os resultados de precipitação obtidos para a estação verão de 1999. Observa-se que ambas as estimativas por redes neurais apresentaram padrões muito semelhantes àqueles presentes na Figura 7.4(a).

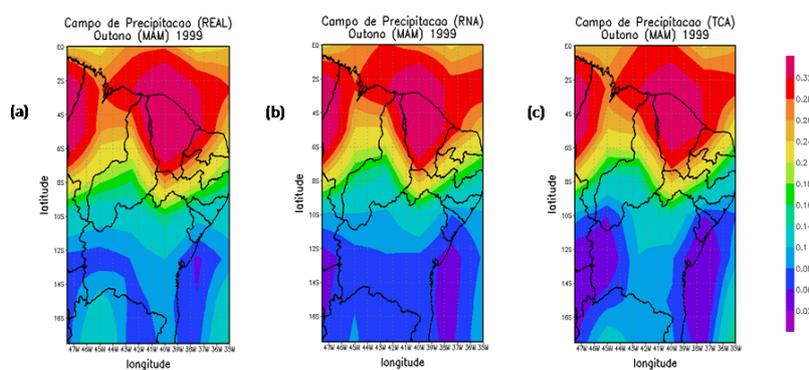


Figura 7.1 - Precipitação para o outono de 1999. (a) Precipitação Real; (b) Estimativa com RNA usando todos os dados; (c) Estimativa com dados processados por TCA.

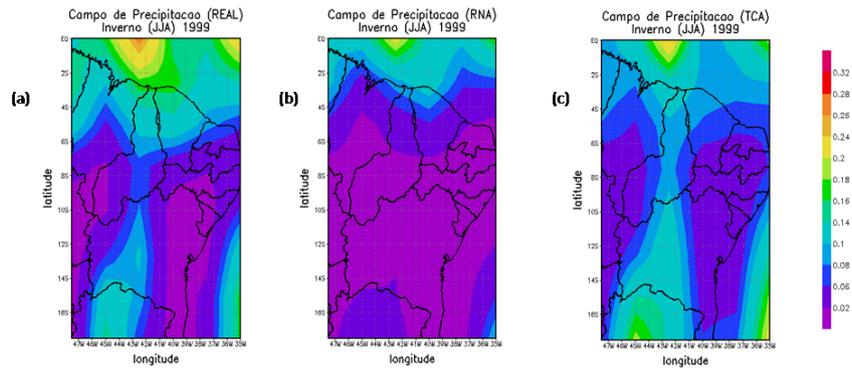


Figura 7.2 - Precipitação para o inverno de 1999. (a) Precipitação Real; (b) Estimativa com RNA usando todos os dados; (c) Estimativa com dados processados por TCA.

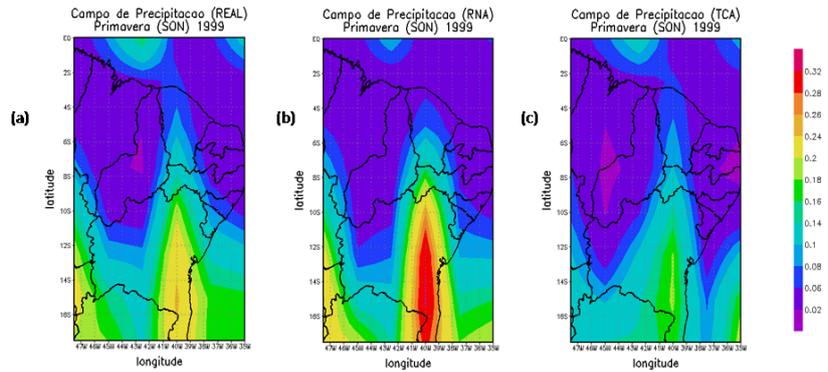


Figura 7.3 - Precipitação para a primavera de 1999. (a) Precipitação Real; (b) Estimativa com RNA usando todos os dados; (c) Estimativa com dados processados por TCA.

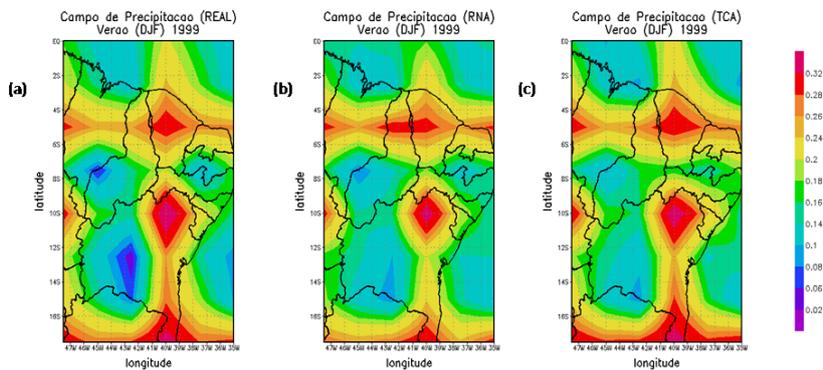


Figura 7.4 - Precipitação para o verão de 1999. (a) Precipitação Real; (b) Estimativa com RNA usando todos os dados; (c) Estimativa com dados processados por TCA.

Da análise de todos resultados apresentados, observa-se que ambos os modelos de redes neurais fizeram estimativas muito próximas quando comparadas aos dados reais. Entretanto, o uso da TCA reduziu o número de atributos de onze para cinco, implicando na redução do custo computacional durante o treinamento das redes neurais e no processo de ativação para previsão.

7.2 Resultados para na região Nordeste em 1998

Foram realizados experimentos com dados da Região Nordeste, para realização de previsão de precipitação. Neste experimento, buscou-se estudar a melhorias na eficiência da rede neural MLP, quando tratava a previsão de forma mais local. Assim, área de estudo foi dividida em quatro subáreas definidas pelas coordenadas geográficas relacionadas na [Tabela 7.3](#), compreendendo 20 pontos de grade para a subárea 1, 15 para a subárea 2, 16 para a subárea 3 e 12 para a subárea 4. Os dados utilizados para este experimento são aqueles do período de 1980 a 2000. As variáveis contidas na base de dados são: temperatura do ar (airt), divergência (div), precipitação (prec), umidade específica (shum), pressão da superfície (spres), componentes vento zonal hPa (u300), (u500) e (u850) e meridional hPa (v300), (v500) e (v850).

Tabela 7.3 - Coordenadas das subáreas de estudo.

Subárea	Longitude	Latitude
1	47°W, 40°W	-7.5°S, 0°
2	40°W, 35°W	-10°S, 0°
3	47°W, 40°W	-17.5°S, -10°S
4	40°W, 35°W	-10°S, -0°

Em cada subárea os dados são utilizados para construir uma base de dados reduzida, através da aplicação da TCA. Em seguida, a base de dados (ou a redução) obtida forma o conjunto de treinamento da rede neural MLP, considerando para as subáreas tratadas. Esse processo totaliza quatro redes MLP, para a obtenção do modelo de previsão da respectiva subárea. Na [Figura 7.5](#) são mostradas as quatro subáreas.

Cada rede neural teve sua topologia configurada variando-se, de maneira experimental, o número de neurônios nas camadas escondidas e o número máximo de épocas de treinamento, sobre o conjunto de treinamento. Os testes conduzidos indicaram uma arquitetura com apenas uma camada escondida contendo 10 neurônios, submetida

ao máximo de 10000 épocas de treinamento; cada neurônio foi configurado com a função de ativação do tipo logística sigmoidal. Durante o treinamento de cada rede, foi utilizada a estratégia de parada antecipada com erro calculado sobre o conjunto de teste e comparado com o erro de treinamento durante cada época de treinamento.

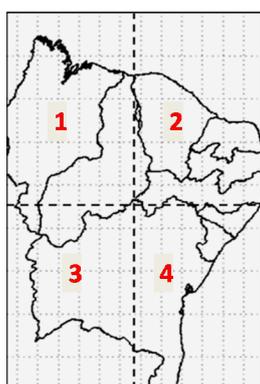


Figura 7.5 - Subáreas de estudo

Os resultados das reduções de variáveis obtidas neste experimento são apresentados na Tabela 7.4 para as quatro subáreas de estudo. Observe-se que das onze variáveis citadas na Tabela 6.1, oito são necessárias nas subáreas 1 e 2, sete na subárea 3 e apenas seis na subárea 4. Estas reduções são utilizadas então para o treinamento das redes neurais de cada subárea, na busca pelo modelo de previsão.

Tabela 7.4 - Redutos extraídos pela TCA.

Redutos			
Subárea 1	Subárea 2	Subárea 3	Subárea 4
airt	airt	airt	airt
u850	u850	u500	u500
u500	u500	v850	v850
v850	u300	v500	v300
v500	v500	v300	shum
v300	v300	spres	div
shum	shum	div	
div	spres		

A Tabela 7.5 apresenta o desempenho dos modelos de previsão, em termos de erro

quadrático médio, para as quatro estações do ano de 1998, considerando as quatro subáreas de estudo.

Tabela 7.5 - Erro Quadrático Médio.

Erro quadrático médio								
	Subárea 1		Subárea 2		Subárea 3		Subárea 4	
Est.	RNA	TCA	RNA	TCA	RNA	TCA	RNA	TCA
Out.	0.009	0.001	0.421	0.078	0.054	0.585	0.343	0.142
Inv.	0.021	0.002	0.315	0.141	0.348	1.012	1.176	0.176
Pri.	0.043	0.013	0.133	0.180	0.499	2.311	0.547	0.126
Ver.	0.005	0.050	0.441	0.382	0.259	1.763	0.536	0.292

Os resultados obtidos nesse experimento podem ser visualizados nas Figuras 7.7 a 7.8, as quais apresentam simultaneamente os resultados obtidos, das quatro estações do ano 1998, sobre as quatro subáreas analisadas. Para visualização e análise dos resultados, os quatro modelos de redes neurais referentes às subáreas foram unidos, como observado Figura 7.6.

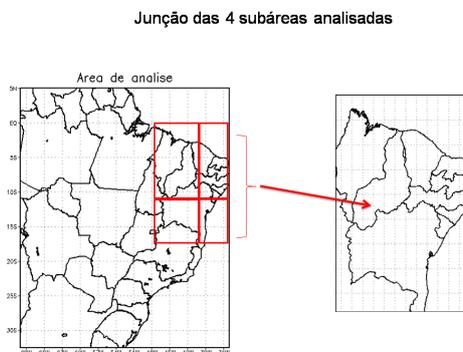


Figura 7.6 - Junção das 4 subáreas analisadas

As Figura 7.7(a), 7.7(b) e 7.7(c) representam respectivamente: as situações observadas (denominadas REAL), a previsão obtida com os dados processados pela TCA e o resultado de previsão com o conjunto de dados completos.

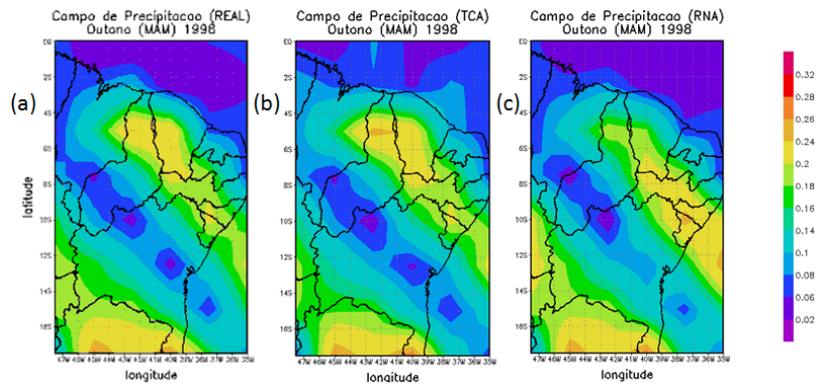


Figura 7.7 - Previsão para estação Outono de 1998

Na Figura 7.8 são apresentados os resultados da estimativa de precipitação para a estação inverno de 1998. Observa-se que a previsão realizada com os dados pré-processados pela TCA tem padrões mais semelhantes àqueles observados na Figura 7.8(a).

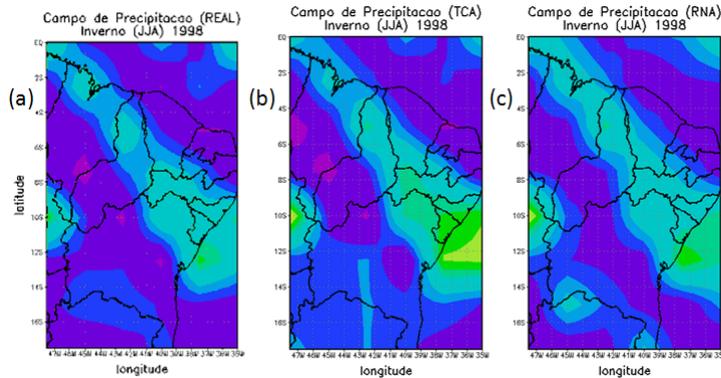


Figura 7.8 - Previsão para estação Inverno de 1998

Na Figura 7.9 são apresentados os resultados de precipitação obtidos para a estação verão de 1998. Observa-se que ambas as estimativas por redes neurais apresentaram padrões muito semelhantes àqueles presentes na Figura 7.9(a).

Na Figura 7.10 são mostrados os resultados de precipitação obtidos pelas redes neurais para a estação primavera de 1998. Também se observa uma grande semelhança entre as estimativas encontradas pelas redes neurais com a Figura 7.10(a) que representa a informação real desejada.

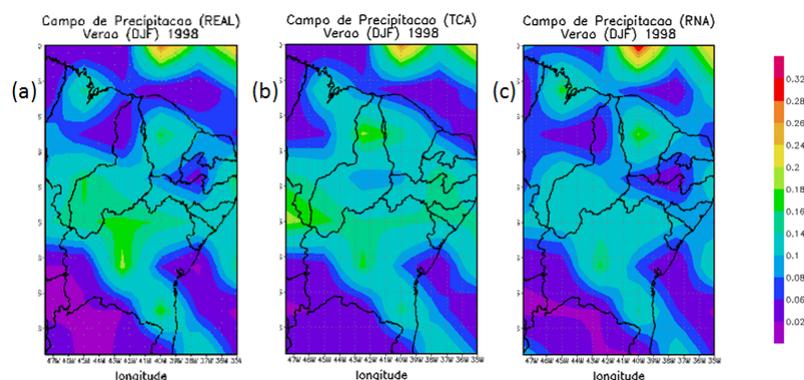


Figura 7.9 - Previsão para estação Verão de 1998

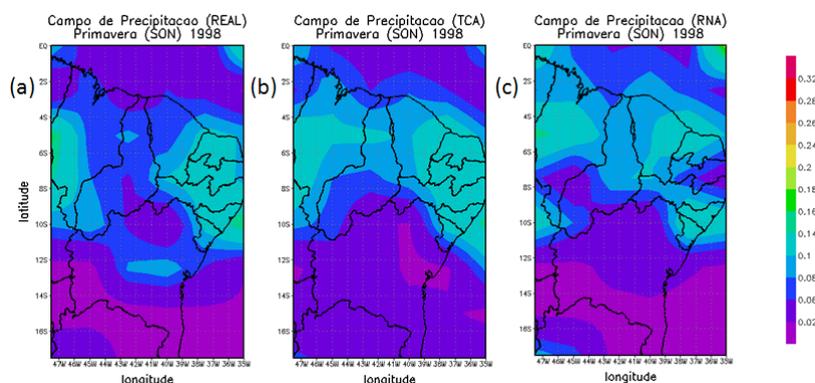


Figura 7.10 - Previsão para estação Primavera de 1998

Com os resultados obtidos foi possível notar que os conjuntos reduzidos de dados geraram previsões muito compatíveis com as previsões realizadas pelo modelo que utiliza todos os dados, mostrando que é possível realizar a previsão com base em um subconjunto de variáveis que, provavelmente, estão mais relacionadas com o comportamento físico da atmosfera, ou seja, subconjunto de variáveis que são vetores que indicam a atividade de precipitação.

7.3 Resultados na região Sul em 1999

Nesta seção são apresentados resultados de experimentos realizados sobre a região Sul do Brasil, para o comportamento sazonal da variável de precipitação. As coordenadas geográficas da área de estudo estão compreendidas entre as longitudes [60°W, 45°W] e entre as latitudes [35°S, 22.5°S], compreendendo 42 pontos de grade (6 latitudes x 7 longitudes). Os dados são mesmos utilizados nos experimentos já

relatados, compreendendo médias mensais do período de tempo de 21 anos entre janeiro de 1980 e dezembro de 2000. As variáveis contidas na base de dados são: temperatura do ar (airt), divergência (div), precipitação (prec), umidade específica (shum), pressão da superfície (spres), componentes vento zonal hPa (u300), (u500) e (u850) e meridional hPa (v300), (v500) e (v850).

A arquitetura da rede foi configurada durante testes preliminares, variando-se o número de neurônios nas camadas escondidas e o número máximo de épocas de treinamento. Os testes conduzidos levaram a uma arquitetura com uma camada escondida contendo 14 neurônios, submetida ao máximo de 10000 épocas, cada neurônio foi configurado com a função de ativação do tipo logística sigmoidal.

A redução dos atributos mais relevantes, com ocorrência igual ou superior a 70% de presença na função de discernimento são apresentados na [Tabela 7.6](#). Observa-se que a dimensão do problema foi reduzida de 11 atributos para 6 atributos.

Tabela 7.6 - Variáveis extraídas pela TCA.

Variáveis	%
u500	76%
u300	88%
v850	86%
v300	72%
spres	70%
div	71%

Na [Tabela 7.7](#) é apresentado o desempenho em termos de erro quadrático médio para os dois modelos de redes neurais, para as quatro estações dos anos de 1998 e 1999.

As Figuras 7.11(a), 7.11(b) e 7.11(c) representam respectivamente as situações observadas (real), a estimativa obtida pela rede com os dados processados pela TCA e a estimativa pela rede neural com todos os dados disponíveis, para a estação verão do ano de 1999.

Na [Figura 7.12](#) são mostrados os resultados de precipitação obtidos para a estação outono de 1999. A estimativa realizada com os dados reduzidos pela TCA apresenta padrões visuais mais semelhantes àqueles mostrados na [Figura 7.12\(a\)](#).

Tabela 7.7 - Erro Quadrático Médio.

Erro quadrático médio			
Ano	Estação	RNA	TCA
1998	Outono	0.0417	0.0138
1998	Inverno	0.0052	0.0701
1998	Primavera	0.0036	0.0489
1998	Verão	0.3930	0.6970
1999	Outono	0.0004	0.0554
1999	Inverno	0.0567	0.0199
1999	Primavera	0.0360	0.2767
1999	Verão	0.1751	0.1347

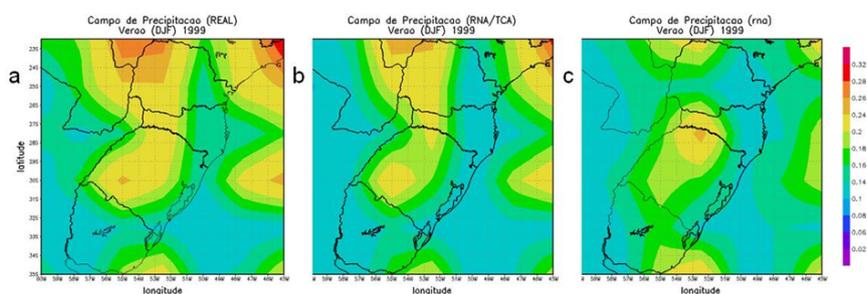


Figura 7.11 - Estimativa de precipitação. Verão de 1999. (a) Precipitação Real; (b) Estimativa com dados reduzidos por TCA, (c) Estimativa com todos os dados.

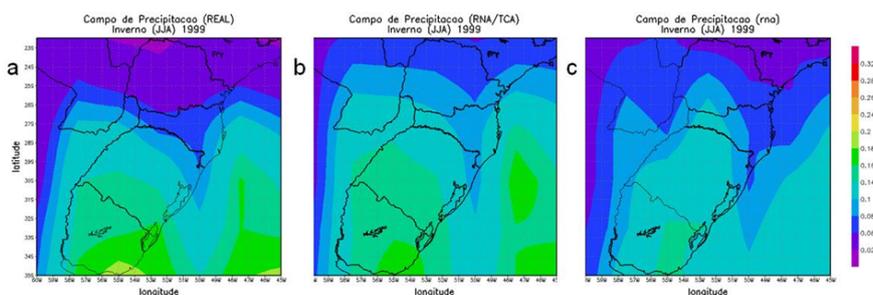


Figura 7.12 - Estimativa de precipitação. Outono de 1999. (a) Precipitação Real; (b) Estimativa com dados reduzidos por TCA, (c) Estimativa com todos os dados.

Na Figura 7.13 são mostradas as estimativas de precipitação obtidas pelos modelos de redes neurais para a estação inverno de 1999. O resultado de previsão obtido utilizando dados processados pela TCA apresenta padrões mais semelhantes aos presentes no resultado esperado mostrado na Figura 7.13(a).

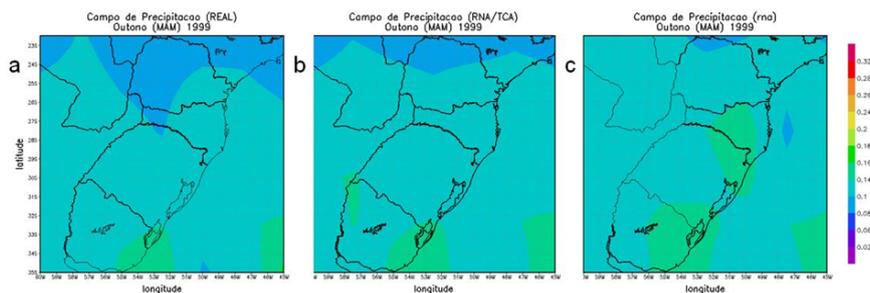


Figura 7.13 - Estimativa de precipitação. Inverno de 1999. (a) Precipitação Real; (b) Estimativa com dados reduzidos por TCA, (c) Estimativa com todos os dados.

Na Figura 7.14 são apresentados os resultados de previsão de precipitação para a estação primavera de 1999. Neste caso a previsão realizada com os dados reduzidos pela TCA, apesar de resultar em padrões semelhantes aos padrões presentes na Figura 7.14(a), tem valores que indicam uma subestimativa da precipitação, como pode ser observado pela escala de intensidades.

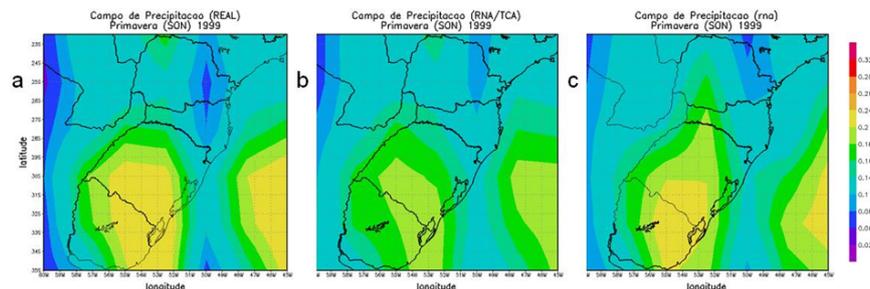


Figura 7.14 - Estimativa de precipitação. Primavera de 1999. (a) Precipitação Real; (b) Estimativa com dados reduzidos por TCA, (c) Estimativa com todos os dados.

Com base nos experimentos realizados sobre a região Sul do Brasil, observou-se que a abordagem por TCA proposta neste experimento, apesar de ser uma boa alternativa para a obtenção de modelos localizados para a previsão de clima, pode levar a resultados subestimados (ou superestimados) para a previsão.

7.4 Resultados de precipitação para as cinco regiões do Brasil

Nesta seção, são apresentados os resultados obtidos para o processo de previsão climática do comportamento sazonal da variável de precipitação utilizando dados

de um período de tempo modificado em relação aos resultados apresentados anteriormente, compreendendo uma janela de tempo de 10 anos, entre janeiro de 2000 e dezembro de 2009 com o intuito de mostrar a robustez dos modelos de previsão por redes neurais encontrados. Entretanto, apesar dos dados terem a mesma origem dos dados do período 1980 a 2000, o novo conjunto de dados não é semelhante ao anterior, por faltar várias variáveis.

Assim, os novos dados foram usados para derivar novos modelos de previsão por redes neurais. Do período disponível, os dados foram divididos em 7 anos (janeiro de 2000 a dezembro de 2006) para o treinamento das redes neurais e como entrada para o processamento das redes por meio da TCA, e os 3 anos (janeiro de 2007 a dezembro de 2009) restantes foram usados para a validação dos modelos.

Os modelos de redes neurais utilizados para realização do processo de previsão climática para os novos dados são: o Perceptron de Múltiplas Camadas (MLP) e as redes recorrentes de Elman e Jordan, os quais utilizam o algoritmo de retropropagação do erro. A métrica para quantificar o desempenho da previsão foi o erro quadrático médio.

As arquiteturas tomadas para os modelos de redes neurais empregadas nesta dissertação também foram configuradas a partir de testes preliminares, nos quais se modificavam de maneira *ad hoc*, o número de neurônios nas camadas escondidas e o número máximo de épocas de treinamento. Desta maneira, os testes conduzidos levaram a uma topologia com apenas uma camada escondida com 15 neurônios, submetida ao máximo de 10000 épocas. Cada neurônio foi configurado com a função de ativação do tipo logística sigmoideal. Durante o treinamento das redes, foi utilizada a estratégia de parada antecipada com erro calculado sobre o conjunto de teste e comparado com o erro de treinamento durante cada época de treinamento.

Em todos os experimentos, os atributos mais relevantes revelados pelo processo de redução dos dados pela TCA, para o processo de previsão climática, foram aqueles com ocorrência igual ou superior a 70% de presença na função de discernimento, os quais são usados para os treinamentos das redes neurais, com objetivo de alcançar a previsão sazonal de precipitação com redução da dimensão do problema.

Nestes experimentos, a área da América do Sul que compreende o Brasil foi dividida em cinco subregiões. As definições das áreas e os resultados de previsão são

apresentados nas seções seguintes.

7.4.1 Resultados na região Nordeste

Os resultados obtidos para o comportamento sazonal da variável de precipitação sobre a região Nordeste são apresentados nas Figuras 7.15 a 7.22. As coordenadas geográficas que definem a região estão compreendidas entre as longitudes [47°W, 35°W] e entre as latitudes [0°S, -17.5°N], compreendendo 48 pontos de grade (8 latitudes x 6 longitudes).

Os atributos mais significativos obtidos pela TCA para esta região são mostrados na Tabela 7.8. Observa-se que o conjunto de atributos foi reduzido de 10 atributos apresentados no Capítulo 6, na Tabela 6.2, para 5 atributos. A Tabela 7.8 mostra a variável e percentual de presença na função de discernibilidade.

Tabela 7.8 - Variáveis extraídas pela TCA - NE.

Redutos extraídos pela TCA	
Variáveis	%
temp	75%
u500	76%
v300	74%
pres	77%
shum	70%

A Tabela 7.9 apresenta o desempenho dos modelos de previsão com as redes MLP, Elman e Jordan, em termos de erro quadrático médio, sobre as quatro estações dos anos de 2007 e 2008, os quais utilizam os dados disponíveis nos dados de validação (generalização) com todas as variáveis.

Tabela 7.9 - Erros quadráticos médios para modelos de redes neurais com todos os dados.

Região NORDESTE								
	Outono		Inverno		Primavera		Verão	
Tipo RN	2007	2008	2007	2008	2007	2008	2007	2008
Jordan	0,0060	0,0491	0,0481	0	0,0367	0,2001	0,1164	0,0114
Elman	0,0731	0,0091	0,0142	0,0022	0,0257	0,0154	0,0096	0,0037
MLP	0,0102	0,0587	0,0725	0,0388	1,0257	0,5609	0,0954	0,0020

Na [Tabela 7.10](#) são mostrados os erros quadráticos médios, para os modelos MLP, Elman e Jordan, obtidos com os dados reduzidos através do uso da TCA.

Tabela 7.10 - Erros quadráticos médios para modelos de redes neurais com dados reduzidos pela TCA.

Região NORDESTE - TCA								
	Outono		Inverno		Primavera		Verão	
Tipo RN	2007	2008	2007	2008	2007	2008	2007	2008
Jordan	0,0027	0,0639	0,0777	0,0484	0,0035	0,0316	0,0236	0,0071
Elman	0,0478	0,0009	0,0709	0,0003	0	0,0011	0,0102	0,0004
MLP	0,0480	0,0001	0,2200	0,0006	0,0187	0,0003	0,2237	0,0037

Nas Figuras [7.15](#) a [7.22](#) são apresentadas análises qualitativas dos resultados. Em cada figura são apresentadas as observações reais (situações de previsão desejadas), as estimativas obtidas pelos modelos de redes neurais de Jordan, Elman, MLP, treinadas com os dados disponíveis com todas as variáveis e os resultados obtidos com os modelos de redes neurais de Jordan, Elman e MLP, treinados usando os dados reduzidos pelo uso da TCA, sobre as quatro estações dos anos de 2007 e 2008.

Na [Figura 7.15](#) são mostrados os resultados de precipitação obtidos para a estação verão de 2007. Neste caso a previsão realizada com a rede recorrente de Elman, usando os dados reduzidos por meio da TCA, apresenta padrões mais semelhantes aos observados na situação Real apresentada na mesma figura.

Na [Figura 7.16](#) mostra os resultados de previsão da precipitação para a estação verão de 2008. Também neste caso a previsão realizada com a rede recorrente de Elman, usando os dados reduzidos por meio da TCA, apresenta padrões presentes no Real mostrado na mesma figura.

Na [Figura 7.17](#) são mostradas as estimativas de precipitação obtidas pelos modelos de redes neurais para a estação inverno de 2007. Observa-se que os resultados obtidos usando os modelos de Jordan, com dados com todos os atributos disponíveis na base de dados e com os dados reduzidos, apresentam padrões semelhantes ao dado real. Os modelos de Elman em que utiliza os dados reduzidos por meio da TCA e MLP usando todos os dados, também apresentaram gradientes semelhantes aos presentes no Real, observado na mesma figura.

Na [Figura 7.18](#) são apresentados os resultados de previsão de precipitação para a estação inverno de 2008. Observa-se que o modelo recorrente de Elman usando os dados pré-processados pela TCA, apresenta um resultado com padrões semelhante aos presentes na situação Real na mesma figura. Analisando o resultado com a rede recorrente de Jordan usando todos os dados no treinamento, embora haja uma superestimação em determinados pontos da região, apresentou um resultado com padrões bem próximos ao Real apresentado na mesma figura.

Na [Figura 7.19](#) são mostrados os resultados de precipitação obtidos para a estação outono de 2007. A estimativa realizada com o modelo de Elman usando os dados reduzidos apresenta padrões visuais mais semelhantes em relação ao real. Os modelos Jordan e Elman, que utilizam todos os dados disponíveis, apresentam gradientes semelhantes ao observado (situação Real na mesma figura).

Na [Figura 7.20](#) são apresentadas as estimativas de precipitação para a estação outono de 2008. Observa-se que a estimativa realizada com o modelo de rede recorrente Elman usando os dados processados pela TCA apresenta resultados com padrões mais compatíveis aos presentes no Real (ver a mesma figura). Analisando o modelo MLP usando os dados reduzidos pela TCA, mostrou similaridade aos padrões presentes na situação Real, apesar da superestimação em alguns pontos da região.

As [Figuras 7.21](#) e [7.22](#) apresentam respectivamente os resultados de estimativas para a estação primavera nos anos de 2007 e 2008. Os resultados de maneira geral não apresentam padrões similares aos presentes na situação Real (ver as figuras). Entretanto, é possível notar que as estimativas obtidas com o modelo de Elman, para os dados processados por meio da TCA, apresentaram-se mais próximas da situação Real (ver as figuras).

Verão 2007

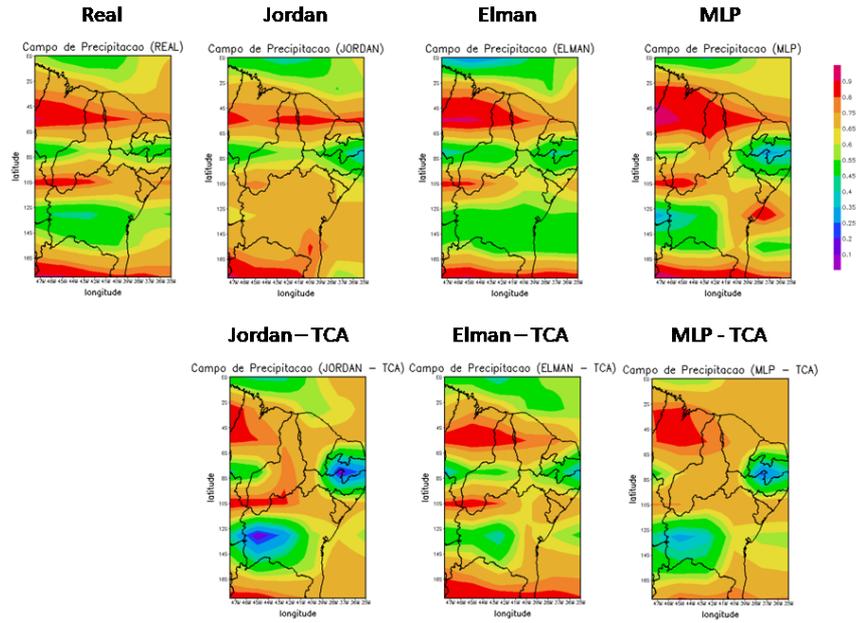


Figura 7.15 - Estimativa de precipitação no Verão de 2007 na região Nordeste.

Verão 2008

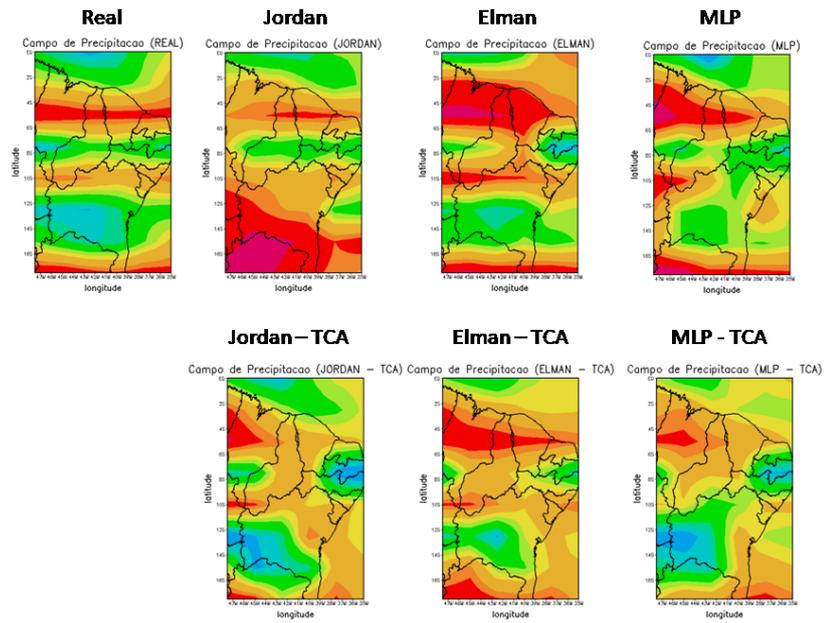


Figura 7.16 - Estimativa de precipitação no Verão de 2008 na região Nordeste.

Inverno 2007

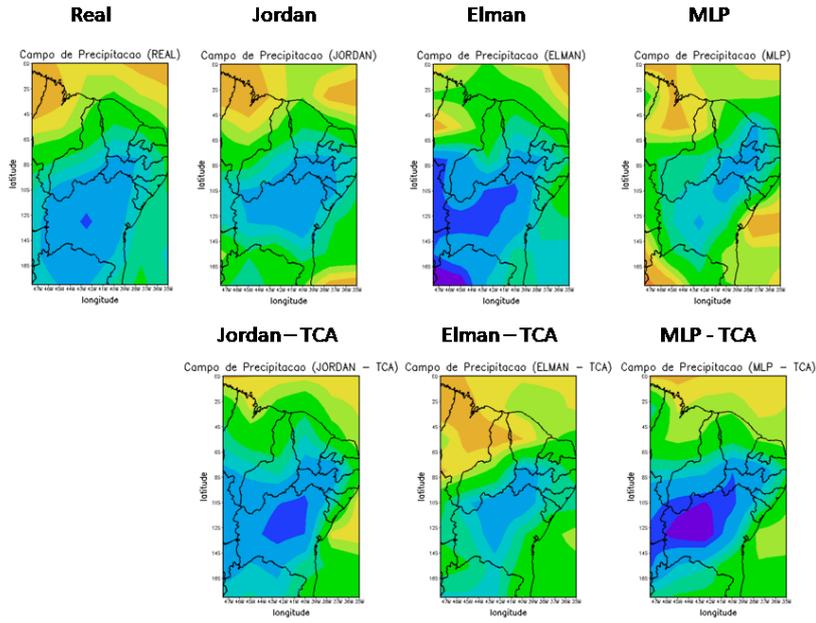


Figura 7.17 - Estimativa de precipitação no Inverno de 2007 na região Nordeste.

Inverno 2008

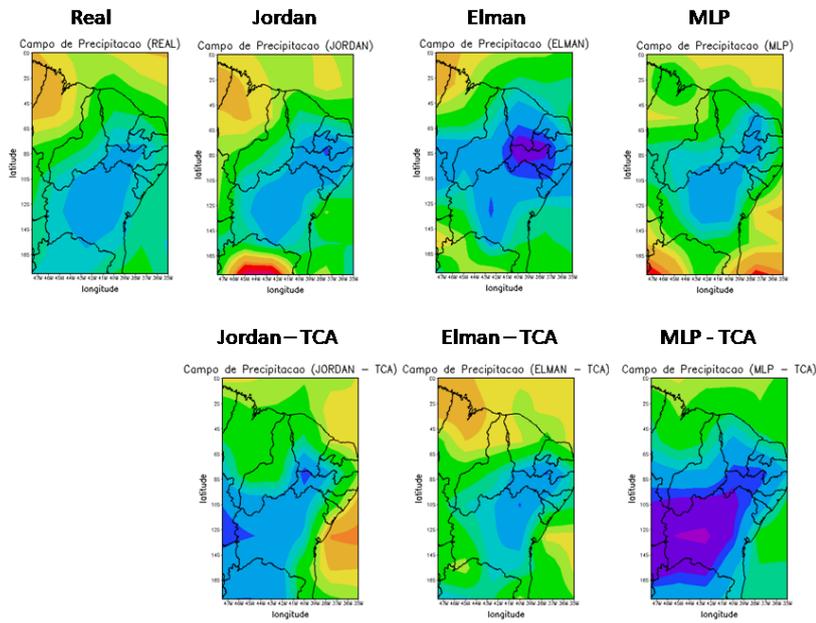


Figura 7.18 - Estimativa de precipitação no Inverno de 2008 na região Nordeste.

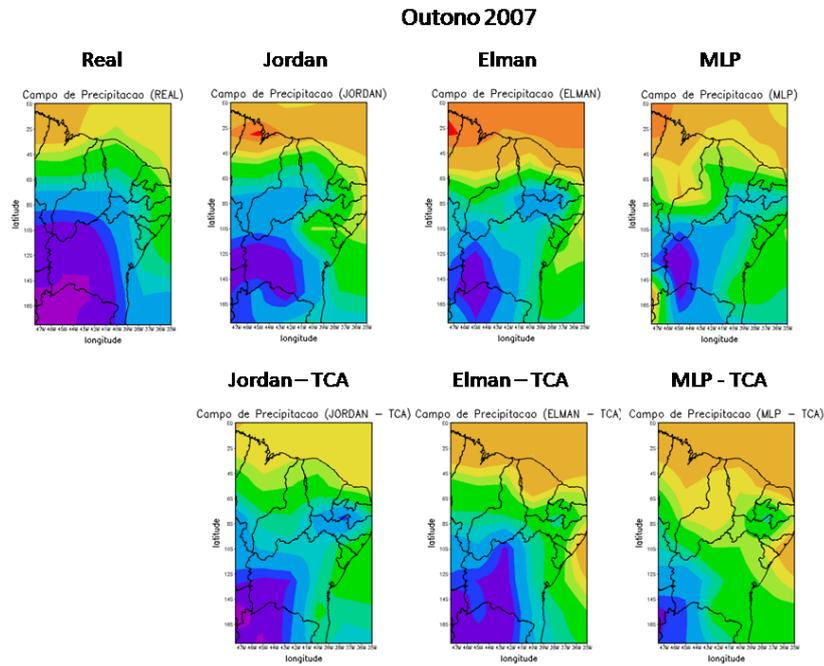


Figura 7.19 - Estimativa de precipitação no Outono de 2007 na região Nordeste.

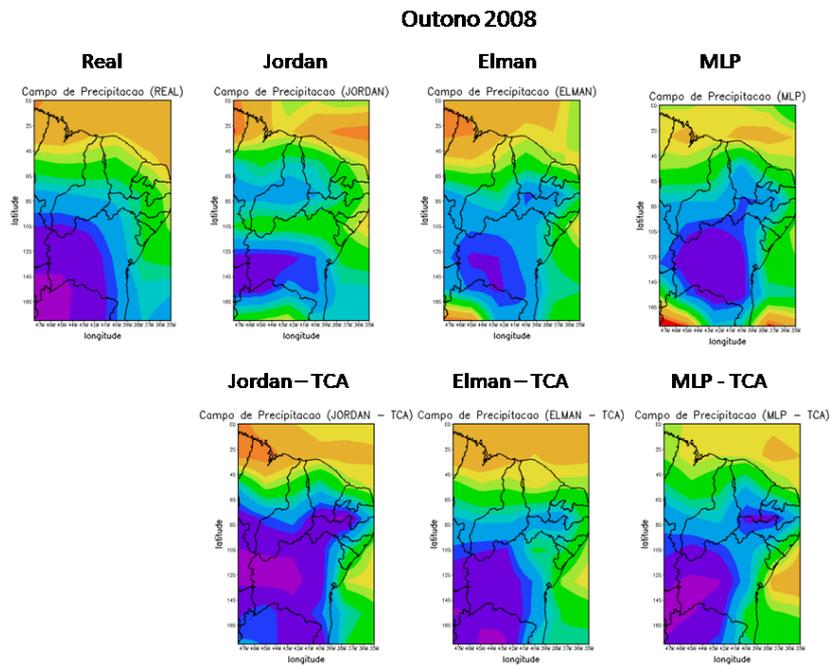


Figura 7.20 - Estimativa de precipitação no Outono de 2008 na região Nordeste.

Primavera 2007

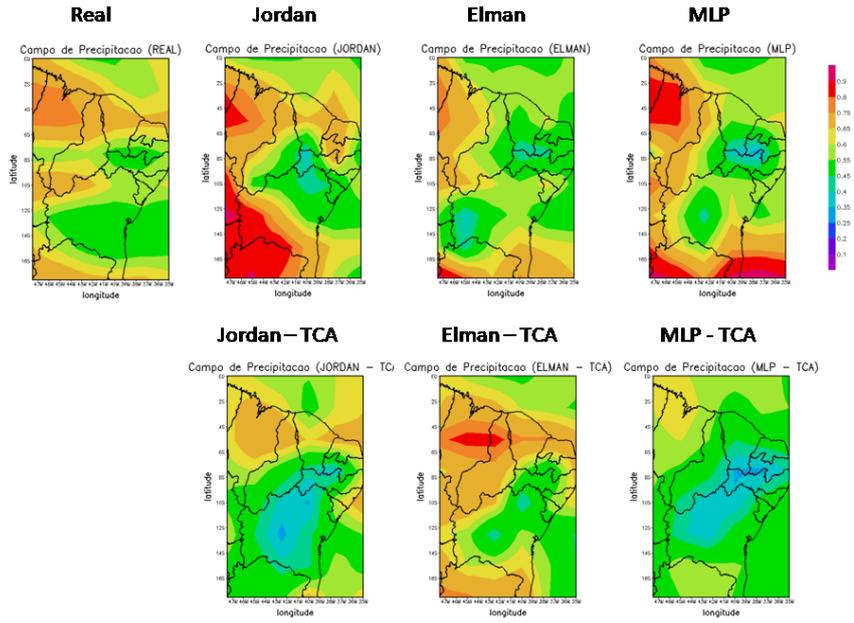


Figura 7.21 - Estimativa de precipitação na Primavera de 2007 na região Nordeste.

Primavera 2008

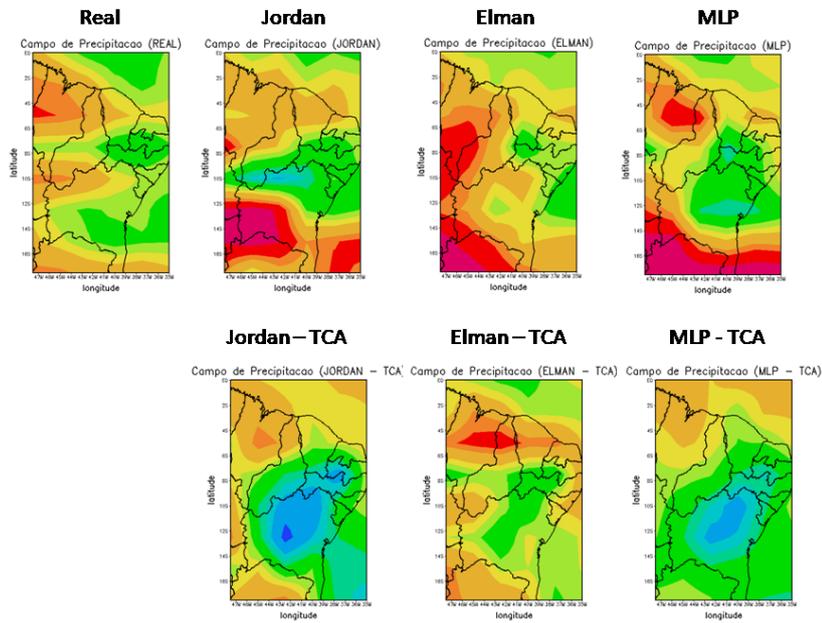


Figura 7.22 - Estimativa de precipitação na Primavera de 2008 na região Nordeste.

Analisando os resultados de previsão climática de forma qualitativa, observa-se que os padrões de gradientes obtidos nas estações inverno e outono, o modelo recorrente de Elman usando os dados reduzidos pela TCA, mostrou-se mais semelhante aos padrões presentes na situação real (REAL), entretanto o erro quadrático médio obtido com o modelo de Elman usando todos os dados, apresentou um menor erro. Diante deste problema, foram calculados os erros em cada ponto de grade, com objetivo de analisar os padrões de gradiente, identificando na grade uma possível ocorrência de subestimação ou superestimação.

As Figuras B.1 e B.4 apresentam os erros em cada ponto de grade. Observa-se em alguns pontos da grade, que os erros obtidos no modelo recorrente de Elman usando os dados reduzidos foram maiores em relação ao modelo de Elman usando todos os dados. As demais Figuras contendo os erros por ponto de grade podem ser visualizadas no Apêndice B.

Erro em cada ponto de grade – Rede de Elman usando todos os dados (NE)

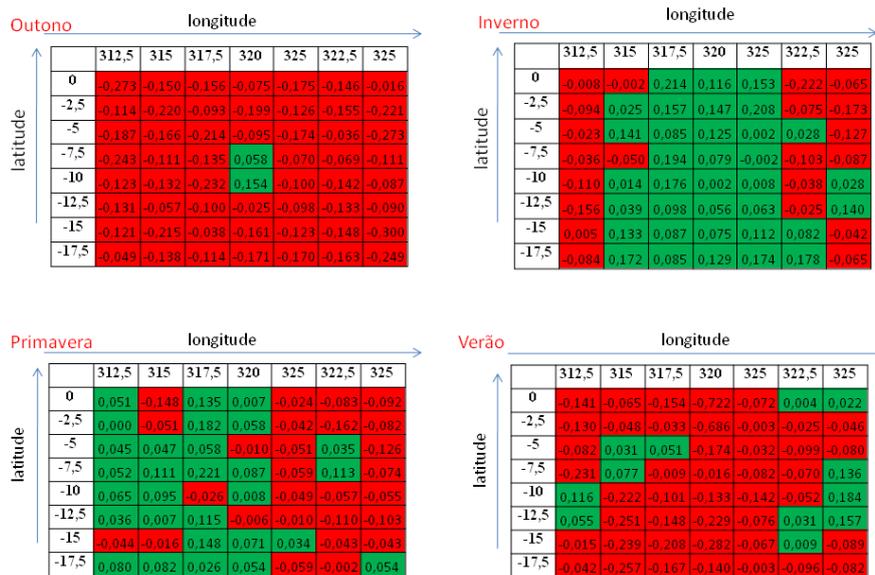


Figura 7.23 - Erro em cada ponto de grade: Rede de Elman com todos os dados.

Erro em cada ponto de grade – Rede de Elman usando TCA (NE)

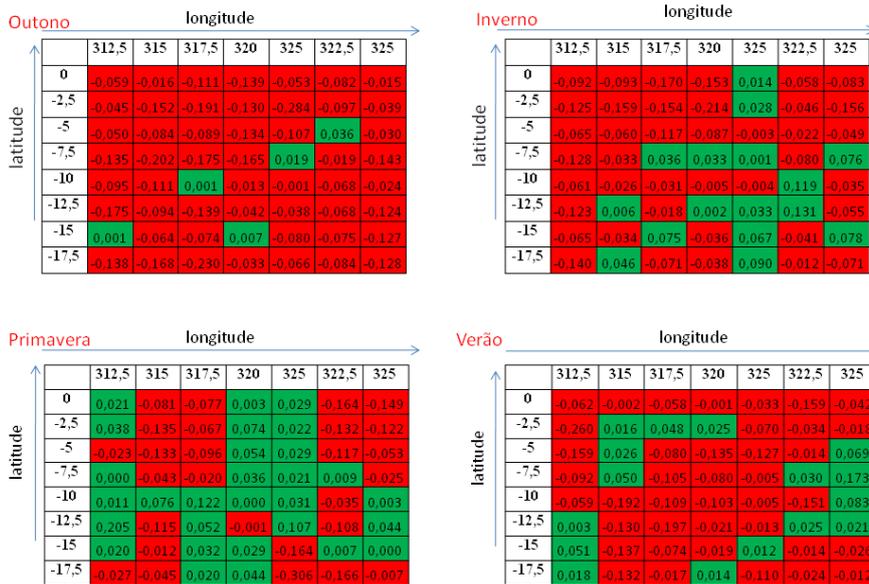


Figura 7.24 - Erro em cada ponto de grade: Rede de Elman usando TCA.

7.4.2 Resultados na região Norte

As Figuras 7.25 a 7.32 apresentam os resultados do comportamento sazonal da variável de precipitação, sobre a região Norte do Brasil. As coordenadas geográficas estão compreendidas entre as longitudes [75°W, 45°W] e entre as latitudes [15°S, 10°N], compreendendo 140 pontos de grade (14 latitudes x 10 longitudes).

Na Tabela 7.11 são apresentados os atributos mais relevantes obtidos por meio da TCA para essa região. Observa-se que das dez variáveis mencionadas no Capítulo 6, na Tabela 6.2, sete são necessários para realizar previsão sazonal de precipitação.

A Tabela 7.12 mostra os erros quadráticos médios dos modelos MLP, Elman e Jordan, para as previsões que utilizam os dados com todos os atributos disponíveis na base de dados, para as quatro estações dos anos de 2007 e 2008.

A Tabela 7.13 apresenta os erros quadráticos médios das previsões realizadas com os modelos MLP, Elman e Jordan, utilizando os dados pré-processados pela TCA, sobre as quatro estações dos anos de 2007 e 2008.

Nas Figuras 7.25 a 7.32 são apresentadas a precipitação observada ou real, os re-

Tabela 7.11 - Variáveis extraídas pela TCA na região Norte.

Redutos extraídos pela TCA	
Variáveis	%
temp	74%
u300	81%
u500	81%
v300	76%
v500	78%
pres	77%
shum	88%

Tabela 7.12 - Erro quadrático médio para as quatro estações na região Norte.

Região NORTE								
	Outono		Inverno		Primavera		Verão	
Tipo RN	2007	2008	2007	2008	2007	2008	2007	2008
Jordan	0,0134	0,0808	0,0118	0,0124	0,0106	0,0233	0,0045	0,0166
Elman	0,0007	0,0281	0,0021	0,0069	0,0037	0,0055	0,0002	0,0135
MLP	0,0689	0,0668	0,0320	0,0006	0,1237	0,0557	0,0203	0,0188

Tabela 7.13 - Erro quadrático médio para as quatro estações na região Norte usando a TCA.

Região NORTE - TCA								
	Outono		Inverno		Primavera		Verão	
Tipo RN	2007	2008	2007	2008	2007	2008	2007	2008
Jordan	0,0253	0,0011	0,0425	0,0080	0,0279	0,2689	0,0547	0,1359
Elman	0,0401	0,0094	0,0008	0,0002	0,0161	0,0274	0,0156	0,0148
MLP	0,0572	0,1499	0,0601	0,0585	0,0669	0,1299	0,0135	0,0602

sultados obtidos com os modelos de Jordan, Elman e MLP, utilizam os dados com as todas as variáveis disponíveis na base de dados, e os resultados obtidos para as mesmas redes utilizando os dados processados pela TCA, sobre as quatro estações dos anos de 2007 e 2008.

Na [Figura 7.25](#) são apresentados os resultados da estimativa de precipitação para verão de 2007. Observa-se que o resultado usando o modelo de rede Elman com dados completos, apresentou padrões similares aos observados no Real (ver figura).

Na [Figura 7.26](#) são apresentados os resultados das estimativas de precipitação para verão de 2008. Novamente observa-se que o resultado usando o modelo de rede Elman com os dados completos, apresentou padrões com certa similaridade àqueles presentes na situação Real (ver figura).

Na [Figura 7.27](#) mostra os resultados das estimativas de precipitação no inverno de 2007. Observa-se que os resultados obtidos com os modelos de Elman usando os dados com todas os atributos disponíveis na base de dados e usando os dados reduzidos através da TCA, apresentaram padrões semelhantes ao Real (ver figura).

Na [Figura 7.28](#) são mostrados os resultados das estimativas de precipitação no inverno de 2008. Os resultados obtidos com os modelos de Elman, tanto para os dados com todos os atributos, quanto para o modelo usando os dados processados pela TCA, embora apresentem superestimação em alguns pontos, apresentam resultados com padrões compatíveis com o Real (ver figura).

Na [Figura 7.29](#) são mostrados os resultados das estimativas de precipitação no outono de 2007. Observa-se que o modelo recorrente de Elman, que utiliza como entradas os dados com todas as variáveis disponíveis na base de dados, apresentou um resultado mais semelhante ao Real (ver figura).

Na [Figura 7.30](#) são mostrados os resultados das estimativas de precipitação no outono de 2008. Observa-se que o modelo de rede recorrente de Elman usando os dados processados pela TCA, mesmo que em alguns pontos da região tenha superestimado, apresenta padrões semelhantes àqueles observados na situação Real (ver figura).

Nas [Figuras 7.31 e 7.32](#), são mostrados os resultados das estimativas de precipitação na estação primavera nos anos de 2007 e 2008, respectivamente. Observa-se que os resultados obtidos pela rede recorrente Elman, utilizando os dados processados pela TCA, possuem padrões mais similares àqueles presentes nas respectivas situações Reais (ver figuras), apesar de apresentarem pontos de super e subestimação.

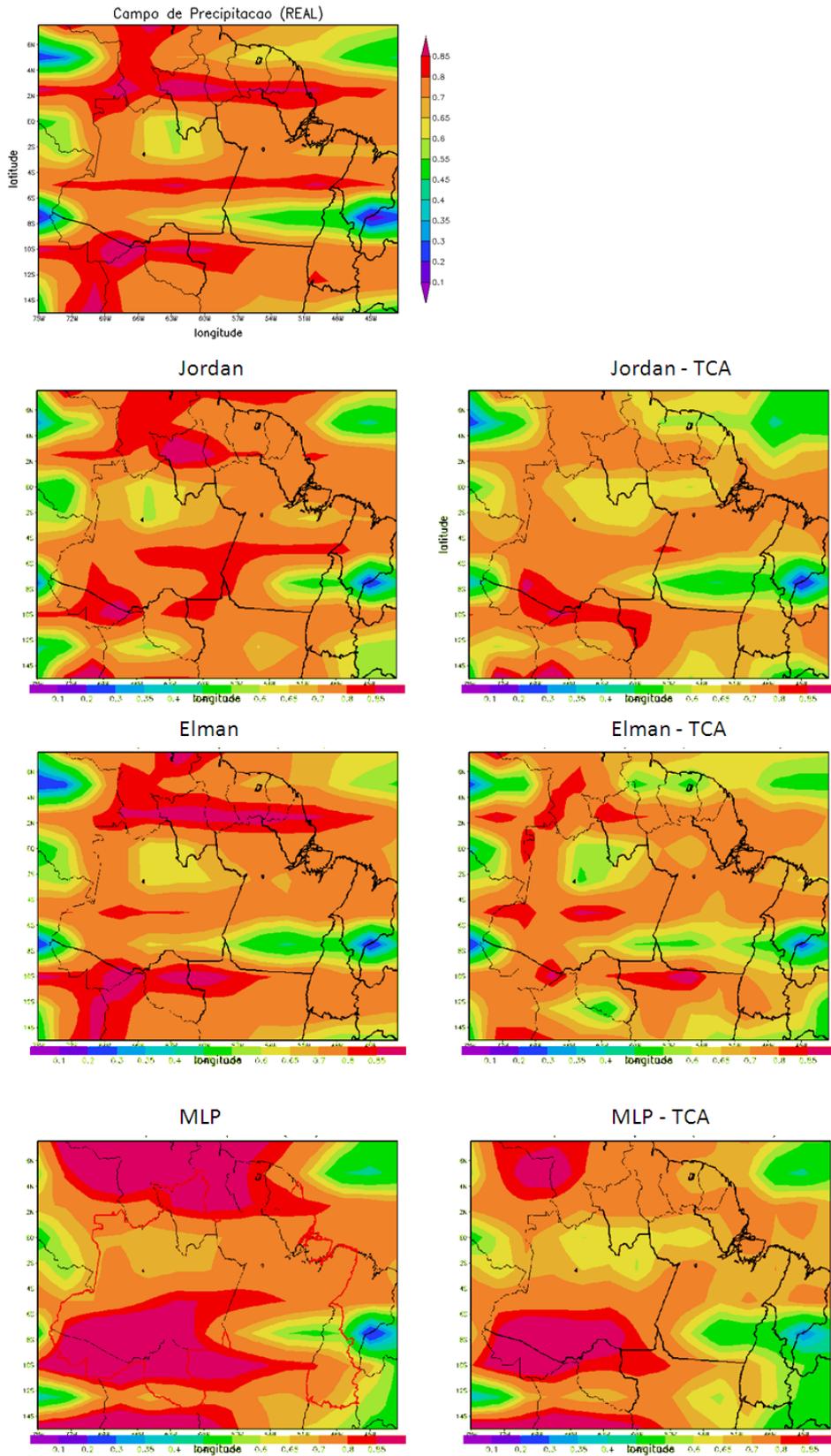


Figura 7.25 - Estimativa de precipitação no Verão de 2007 na região Norte.

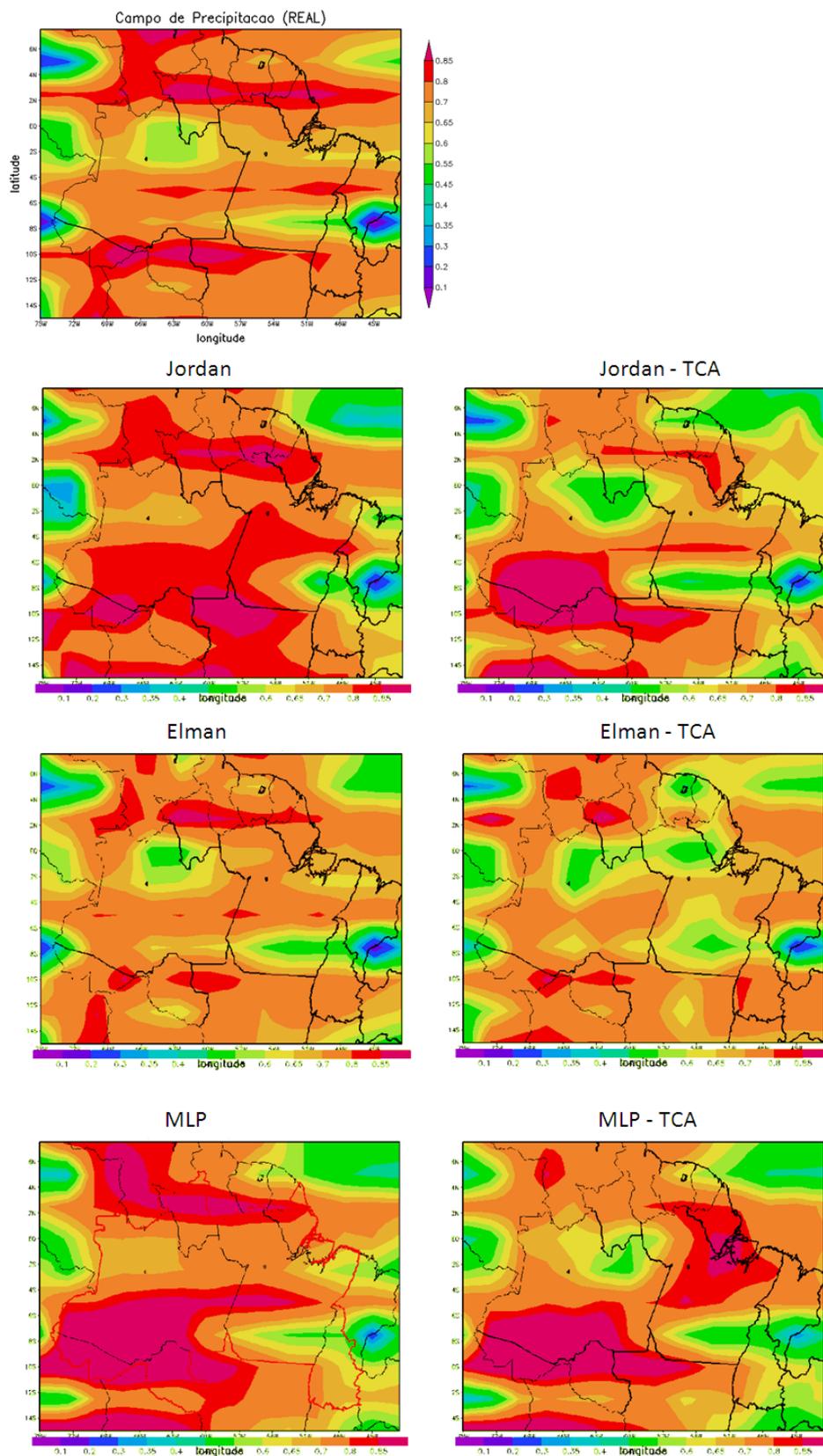


Figura 7.26 - Estimativa de precipitação no Verão de 2008 na região Norte.

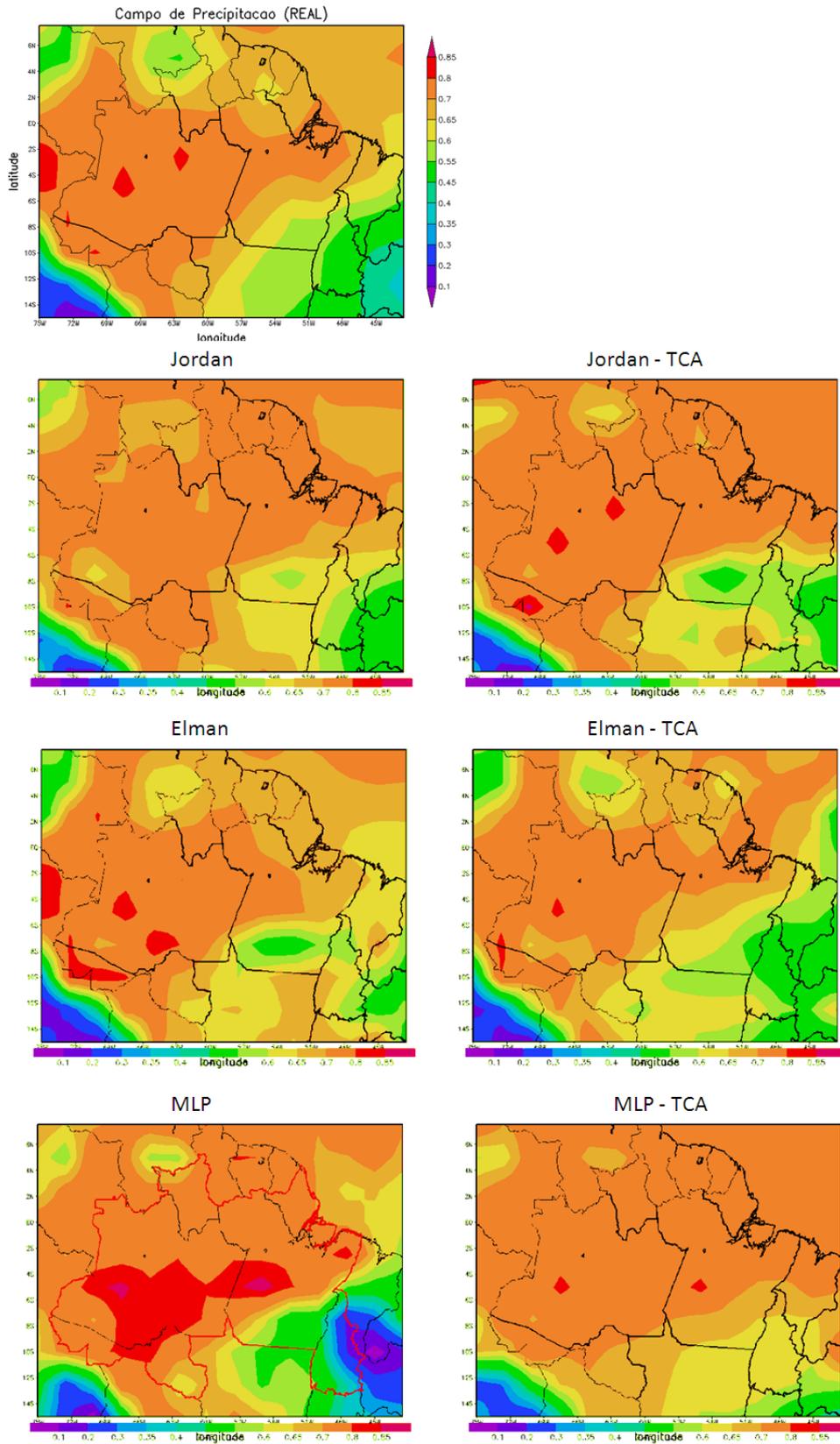


Figura 7.27 - Estimativa de precipitação no Inverno de 2007 na região Norte.

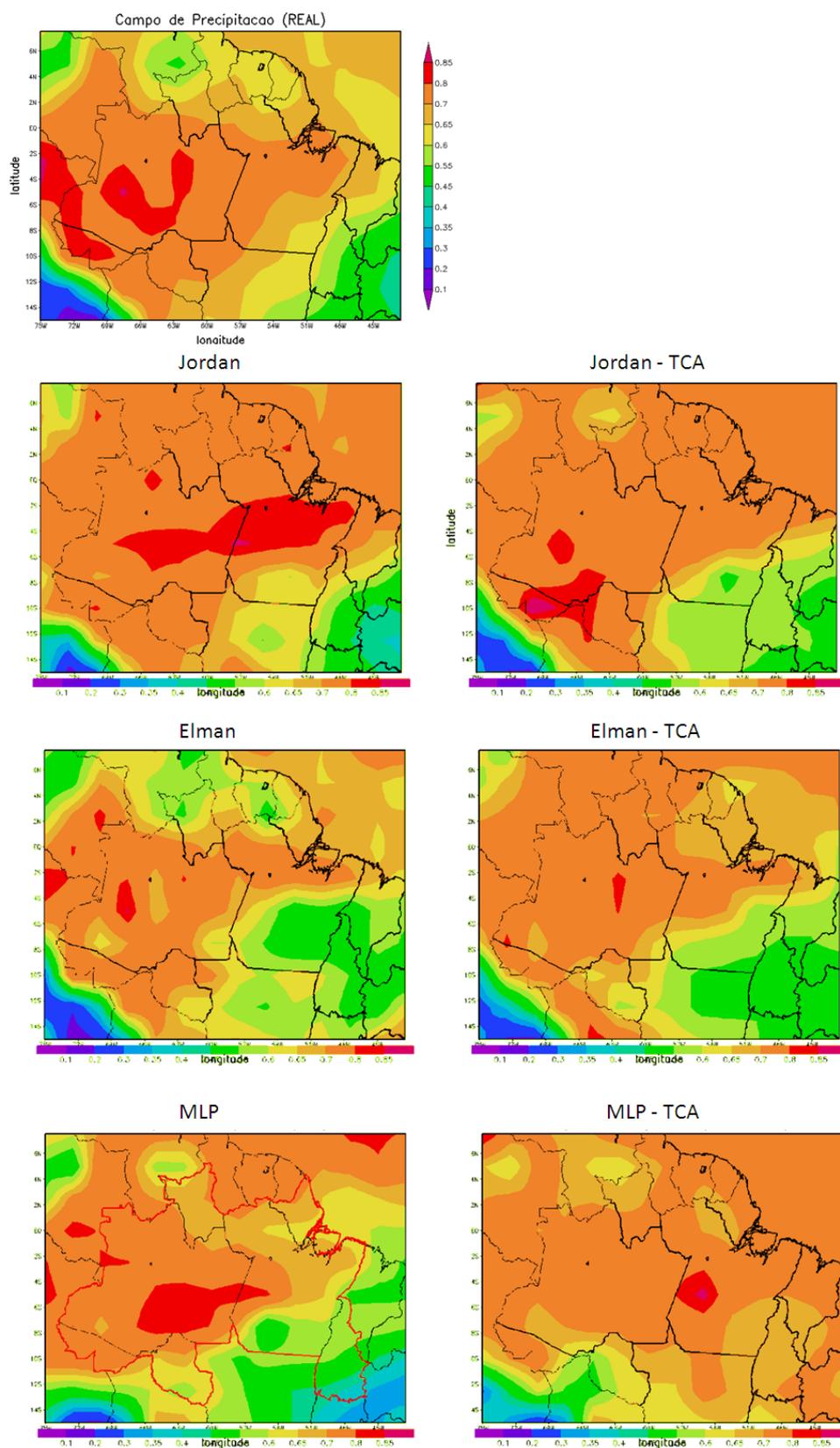


Figura 7.28 - Estimativa de precipitação no Inverno de 2008 na região Norte.

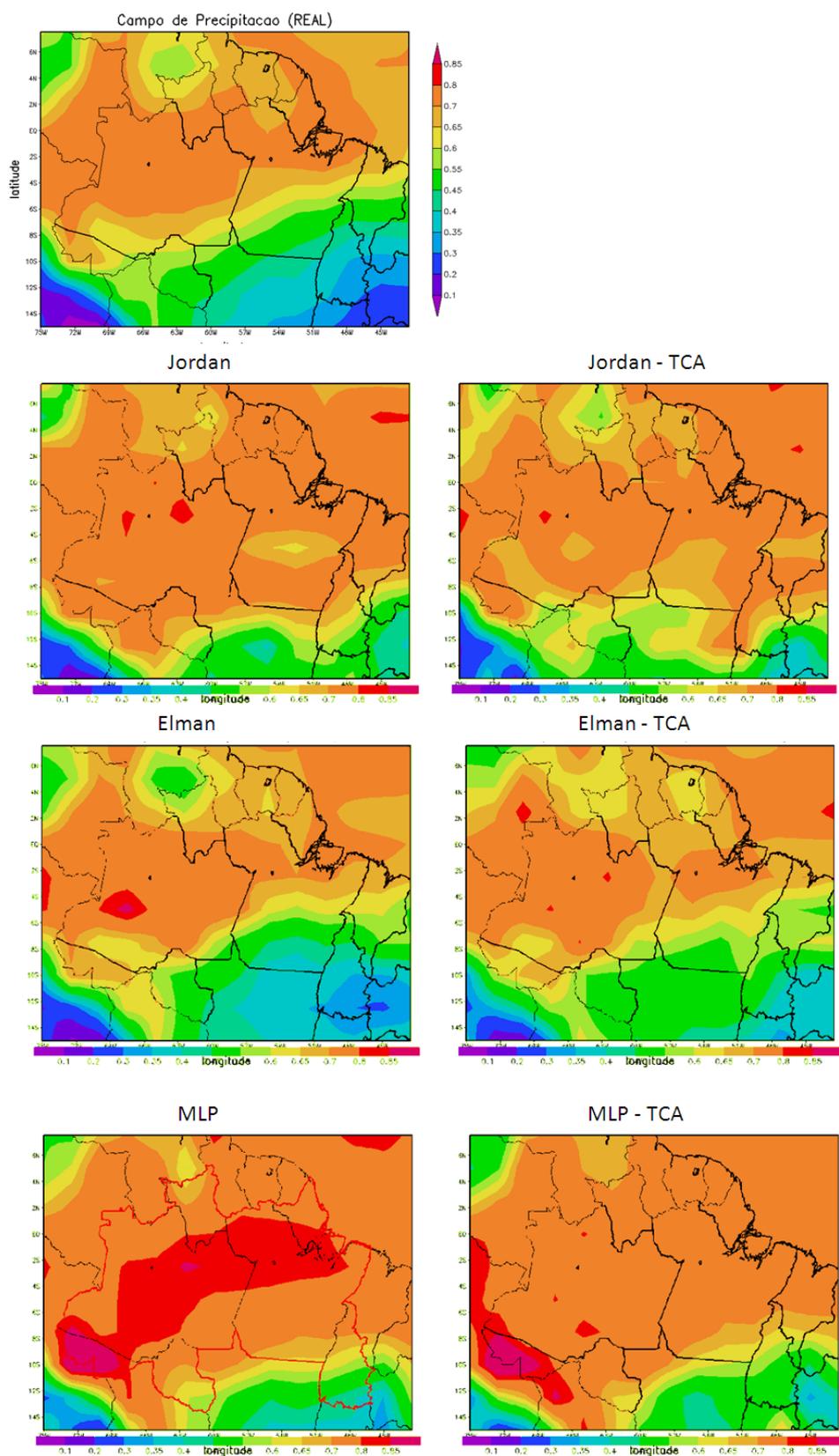


Figura 7.29 - Estimativa de precipitação no Outono de 2007 na região Norte.

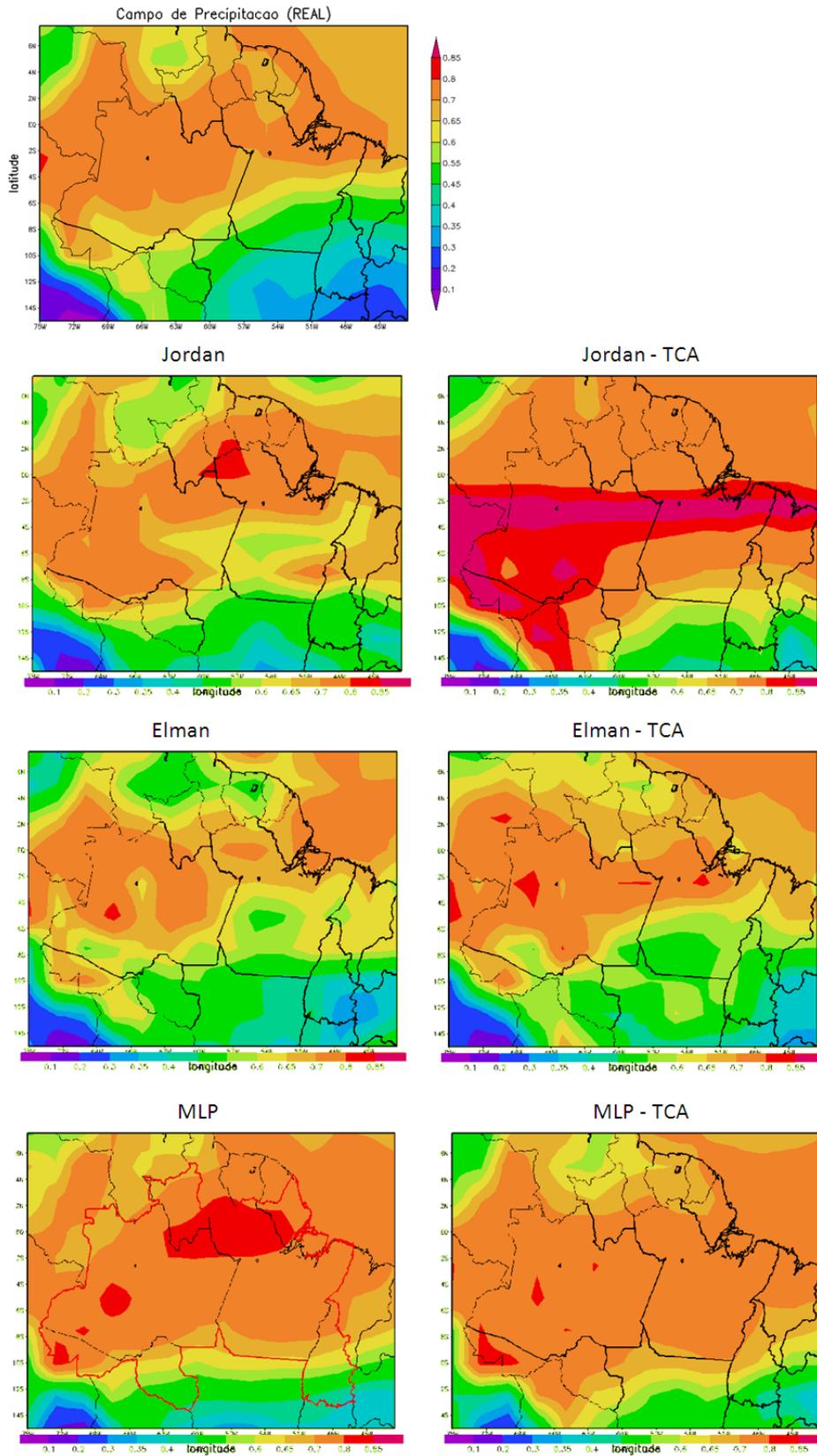


Figura 7.30 - Estimativa de precipitação no Outono de 2008 na região Norte.

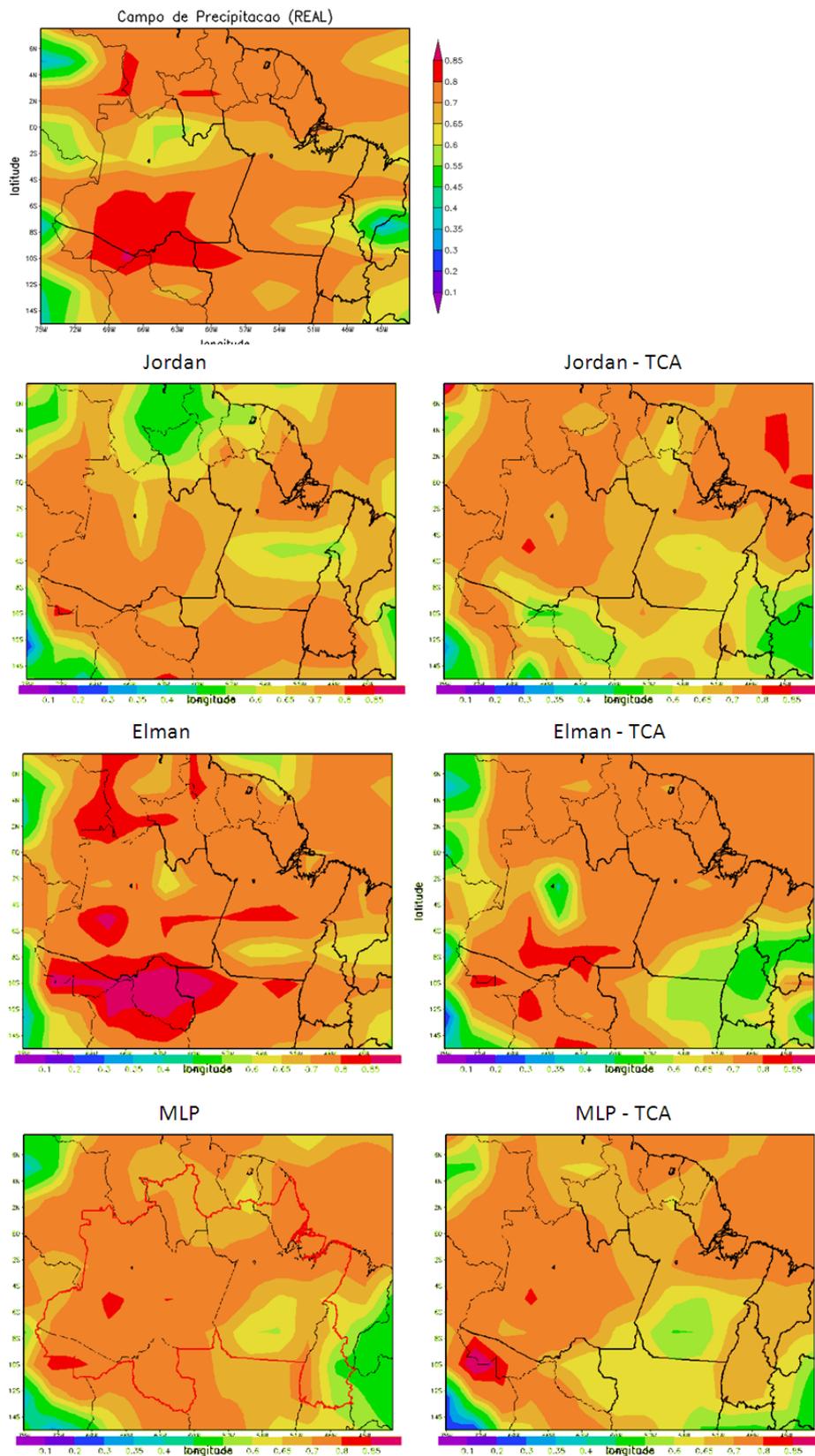


Figura 7.31 - Estimativa de precipitação na Primavera de 2007 na região Norte.

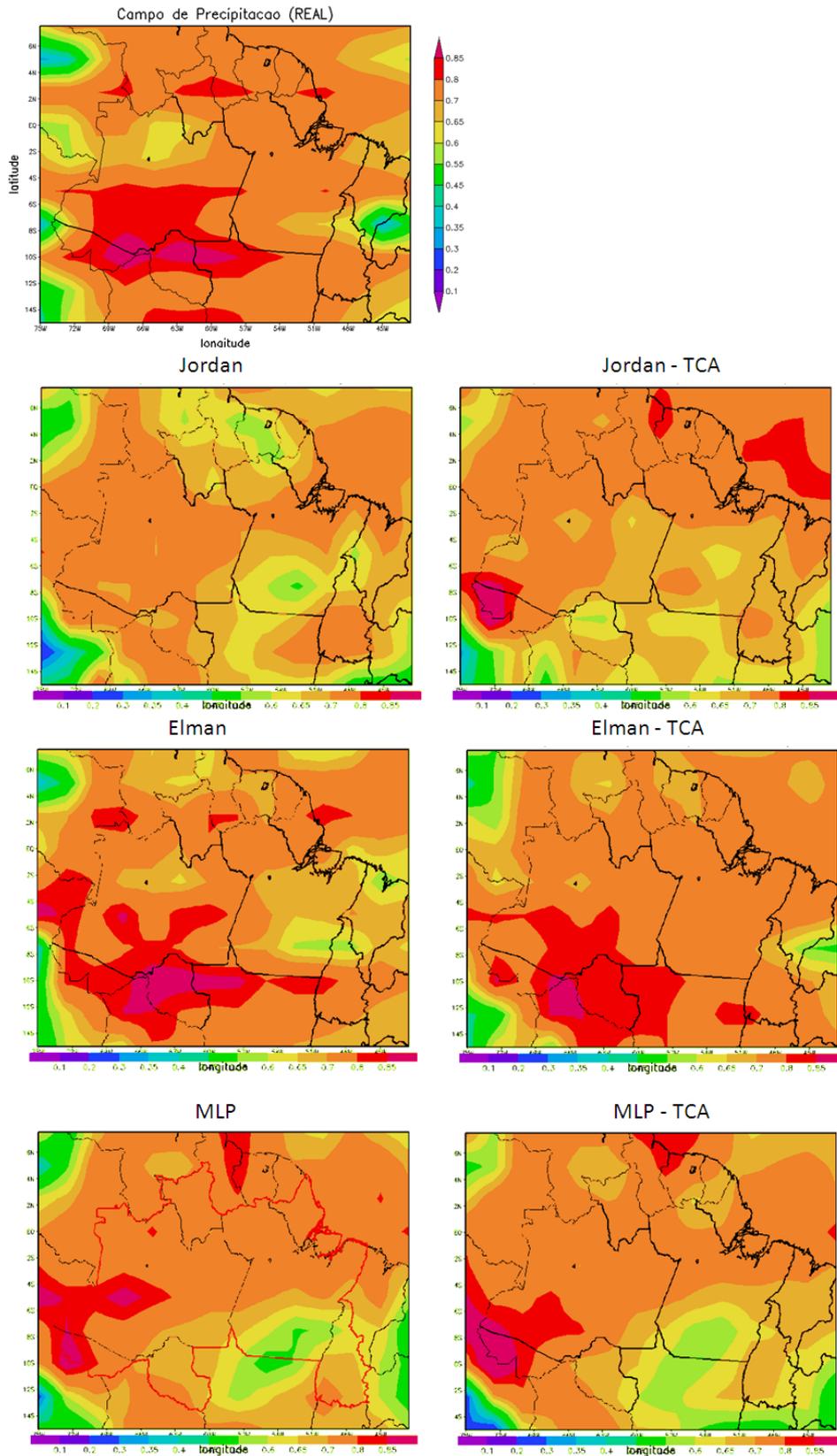


Figura 7.32 - Estimativa de precipitação na Primavera de 2008 na região Norte.

Com base nos experimentos realizados na região Norte do Brasil, observou-se que com o uso da abordagem por TCA, apresentou resultados mais similares aos padrões presentes na situação real, apesar dos erros quadráticos médios obtidos para as estações outono, primavera e verão, serem maiores em relação ao modelo de previsão usando todos os dados.

Diante disso, os erros foram calculados em cada ponto de grade, cujo objetivo é analisar os padrões de gradiente, identificando-se pontos em que houve ocorrência de subestimação ou superestimação. Este pode ser o principal motivo do erro quadrático médio dos modelos de previsão usando a TCA ter sido maior em relação aos modelos de previsão usando todos os dados.

As Figuras B.13 e B.16 apresentam os erros em cada ponto de grade. Observa-se em alguns pontos da grade, que os erros obtidos a partir dos modelos recorrentes de Elman e Jordan usando os dados reduzidos, foram maiores em relação aos modelos de Elman e Jordan usando todos os dados. As demais Figuras contendo os mapas de erros em cada ponto de grade podem ser visualizadas no Apêndice B.

Erro em cada ponto de grade – Rede de Elman usando todos os dados (N)

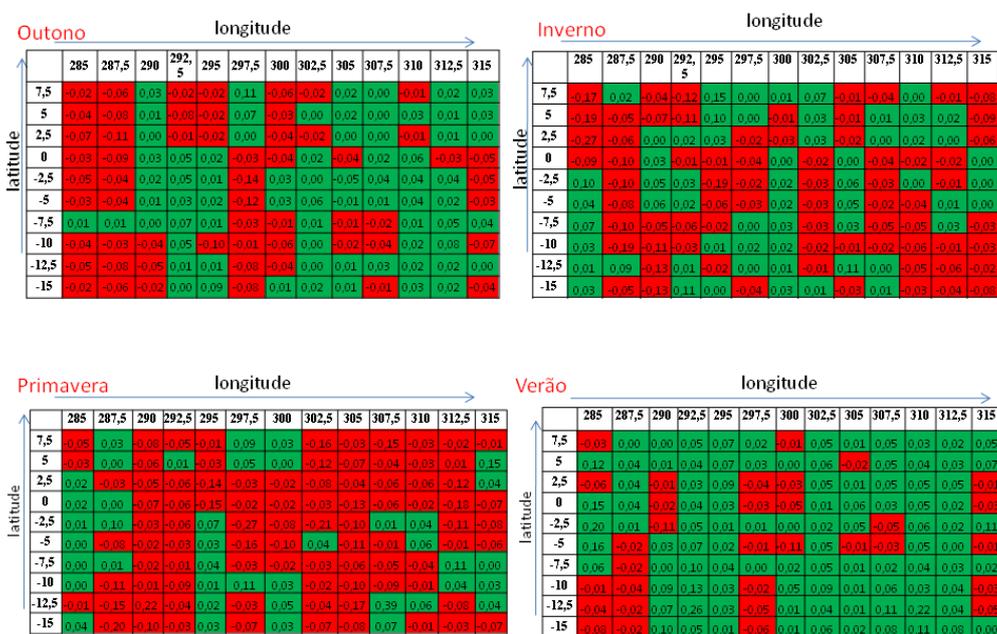


Figura 7.33 - Erro em cada ponto de grade: Rede de Elman usando todos os dados.

Erro em cada ponto de grade – Rede de Elman usando a TCA (N)

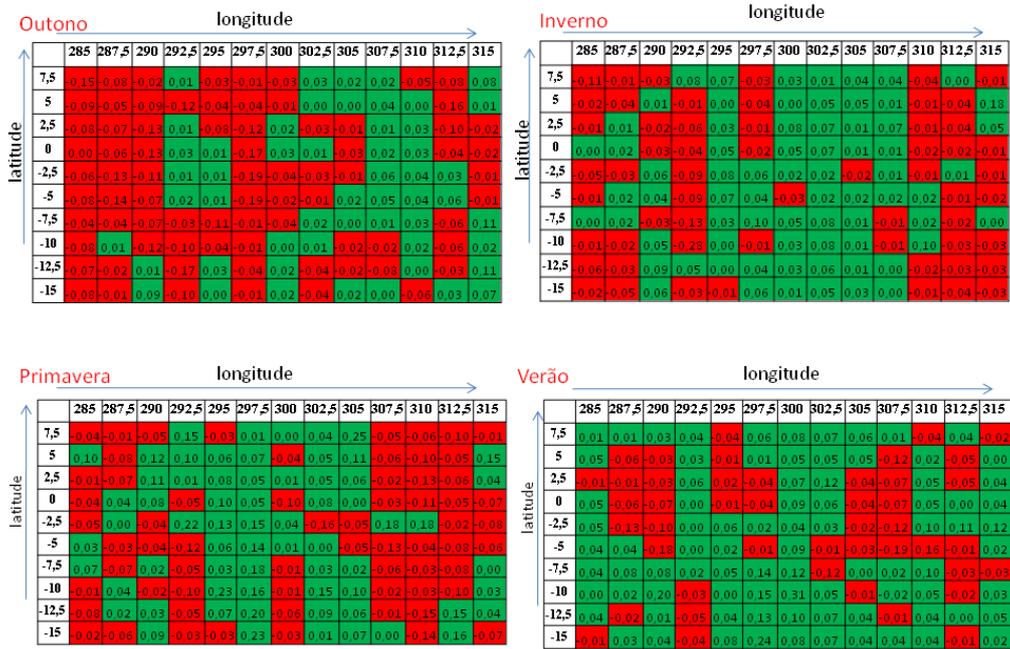


Figura 7.34 - Erro em cada ponto de grade: Rede de Elman usando a TCA.

7.4.3 Resultados na Região Centro Oeste

Os resultados de previsão sazonal da variável de precipitação sobre a região Centro Oeste (CO) do Brasil são apresentados nas Figuras 7.31 a 7.38. As coordenadas geográficas que definem a região estão compreendidas entre as longitudes [62°W, 48°W] e entre as latitudes [25°S, 7.5°S], compreendendo 56 pontos de grade (7 latitudes x 8 longitudes).

Na Tabela 7.14 são apresentados os atributos mais significativos encontrados através do uso da TCA para a região centro-oeste. A partir da Tabela 6.2, no Capítulo 6, pode-se observar que dos dez atributos contidos, apenas seis foram escolhidos como necessários para realizar o processo de previsão.

Tabela 7.14 - Variáveis extraídas pela TCA - CO.

Redutos extraídos pela TCA	
Variáveis	%
temp	72%
u300	74%
u500	79%
v300	73%
v500	75%
pres	77%

A Tabela 7.15 apresenta os erros quadráticos médios obtidos pelos modelos Jordan, Elman e MLP, que utilizaram os conjuntos de dados com todas as variáveis, para as quatro estações nos anos de 2007 e 2008.

Tabela 7.15 - Erro quadrático médio para as quatro estações da região Centro Oeste.

Região CENTRO OESTE								
	Outono		Inverno		Primavera		Verão	
Tipo RN	2007	2008	2007	2008	2007	2008	2007	2008
Jordan	0,1992	0,7434	0,0341	0,1488	0,0408	0,0971	0,1645	0,0159
Elman	0,0858	0,0082	0,0048	0,0108	0,0139	0,0002	0,0027	0,0335
MLP	0,6339	0,1535	0,0880	0,3281	0,5027	0,0722	0	0,0661

Na Tabela 7.16 são apresentados os erros quadráticos médios obtidos pelos modelos

Jordan, Elman e MLP, que utilizaram como entradas os dados reduzidos por meio do uso da TCA, para as quatro estações dos anos de 2007 e 2008.

Tabela 7.16 - Erro quadrático médio para as estações na região Centro Oeste usando TCA.

Região CENTRO OESTE - TCA								
Tipo RN	Outono		Inverno		Primavera		Verão	
	2007	2008	2007	2008	2007	2008	2007	2008
Jordan	0,0647	0,1056	0,0278	0,0555	0,0470	0,0337	0,0002	0,0661
Elman	0,4516	0,0038	0,0032	0,0294	0,0032	0,0123	0,0012	0,0096
MLP	0,3491	0,4167	0,0074	0,0675	0,0479	0,1139	0,0008	0,1373

Nas Figuras 7.35 a 7.42 são apresentadas a precipitação real, os resultados de previsão obtidos com os modelos de Jordan, Elman e MLP, que utilizaram os dados com todos os atributos disponíveis na base de dados e os resultados de previsão dos modelos Jordan, Elman e MLP que utilizaram os dados reduzidos através da TCA.

Na Figura 7.35 são mostradas as estimativas de precipitação para a estação verão de 2007. Observa-se que os resultados obtidos usando o modelo recorrente de Elman com os dados reduzidos pela TCA e usando todas as variáveis, apresentam padrões semelhantes àqueles presentes no Real (ver figura).

Na Figura 7.36 são apresentadas as estimativas de precipitação para a estação verão de 2008. O resultado da rede recorrente de Elman usando os dados com todos os atributos disponíveis apresenta padrões mais similares àqueles presentes na situação Real (ver figura).

Na Figura 7.37 são mostradas as estimativas de precipitação para a estação inverno de 2007. Observa-se que o resultado obtido com a rede de Elman utilizando os dados reduzidos pela TCA, tem padrões similares àqueles presentes na situação Real (ver figura).

Na Figura 7.38 são mostradas as estimativas de precipitação para a estação inverno de 2008. O resultado obtido com o modelo de Elman usando os dados com todas as variáveis disponíveis apresenta padrões similares aos presentes na situação Real (ver figura).

Nas Figuras 7.39 e 7.40 são mostradas as estimativas de precipitação para a estação

outono dos anos 2007 e 2008, respectivamente. Observa-se que os resultados obtidos com o modelo recorrente de Elman, mesmo com superestimação em determinados pontos da região, apresentaram padrões semelhantes aos que estão presentes na situação Real (ver figuras).

Nas Figuras 7.41 e 7.42 são apresentadas respectivamente as estimativas de precipitação para a estação primavera de 2007 e 2008. Observa-se que nesta estação os modelos de previsão não conseguiram produzir padrões visualmente similares àqueles presentes na situação Real (ver figura).

Com base nos resultados de previsão obtidos na região Centro Oeste do Brasil, observou-se que com o uso da abordagem por TCA, mostrou-se mais compatível quando comparada aos padrões presentes na situação real. Entretanto os erros quadráticos médios obtidos a partir do modelo recorrente de Elman usando os dados reduzidos pela TCA, apresentaram-se maiores em relação ao modelo de Elman usando todos os dados, para as quatro estações.

Desta forma, a partir das Figuras B.7 e B.10, é possível fazer analisar os erros em cada ponto de grade, verificando-se os pontos que apresentaram subestimação ou superestimação. As demais Figuras contendo os erros por ponto de grade podem ser visualizadas no Apêndice B.

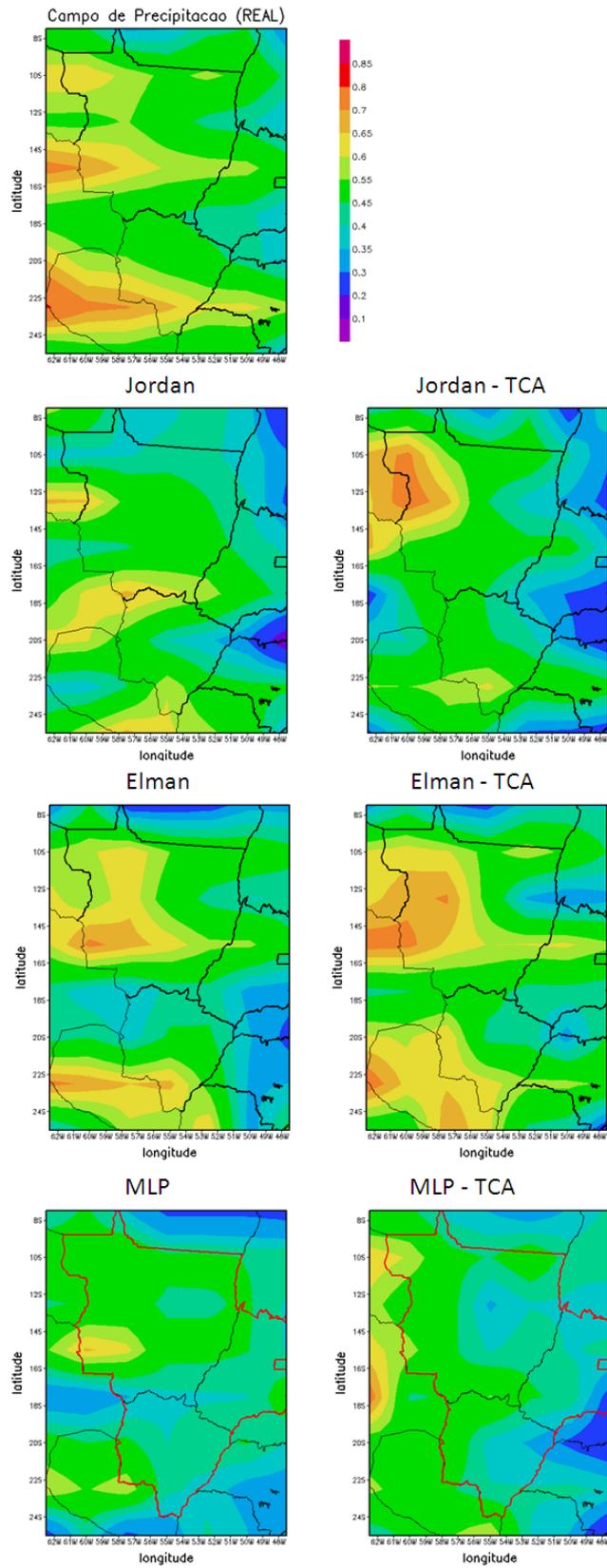


Figura 7.35 - Estimativa de precipitação no Verão de 2007 na região Centro Oeste.

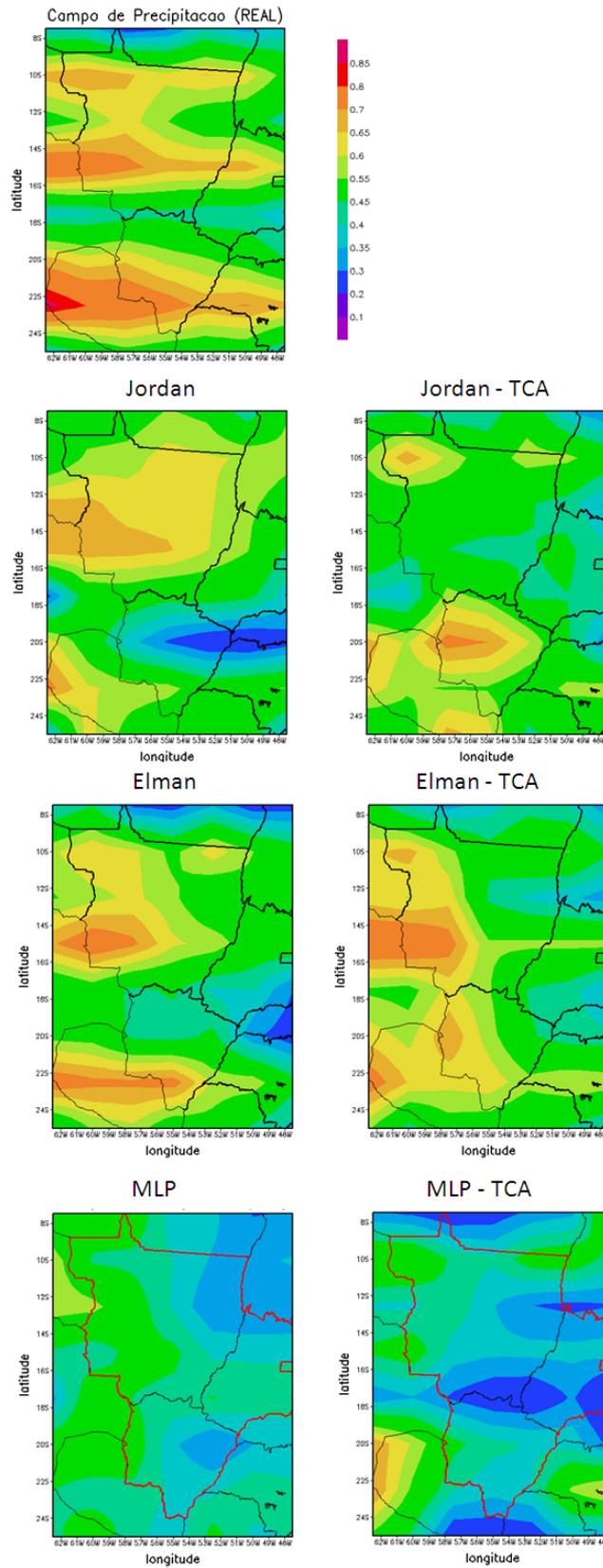


Figura 7.36 - Estimativa de precipitação no Verão de 2008 na região Centro Oeste.

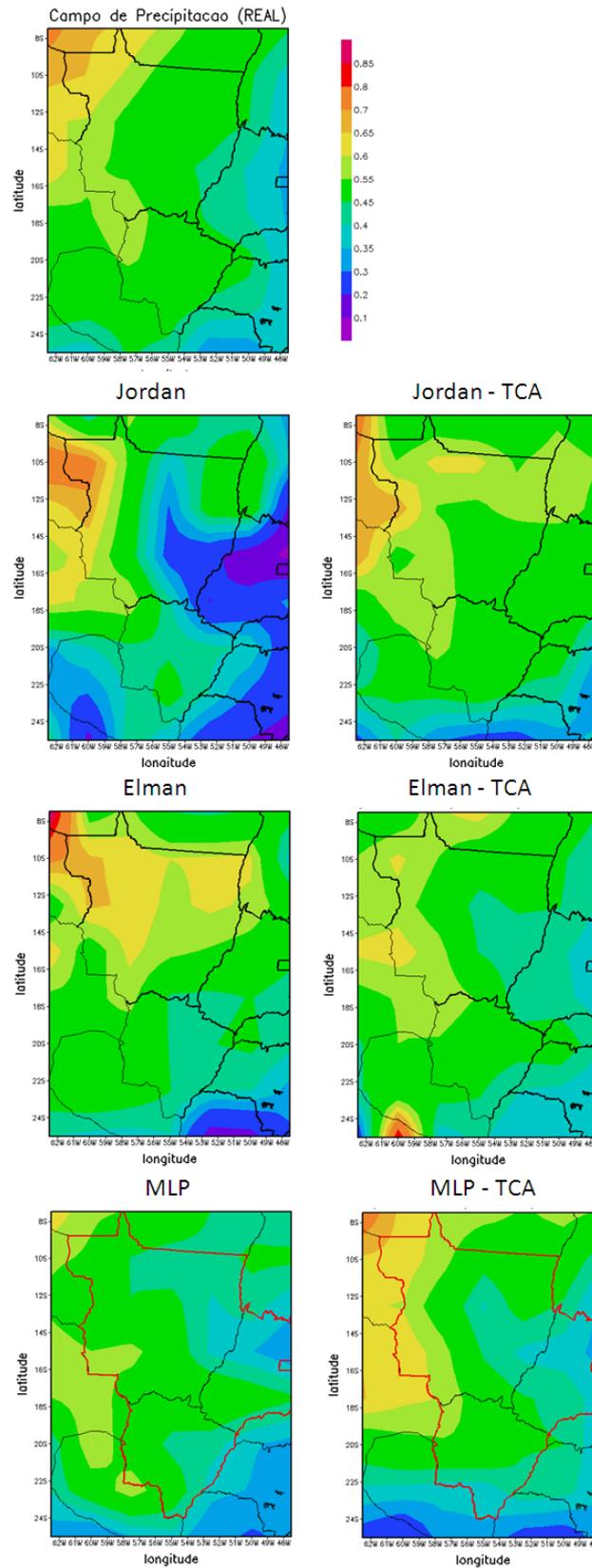


Figura 7.37 - Estimativa de precipitação no Inverno de 2007 na região Centro Oeste.

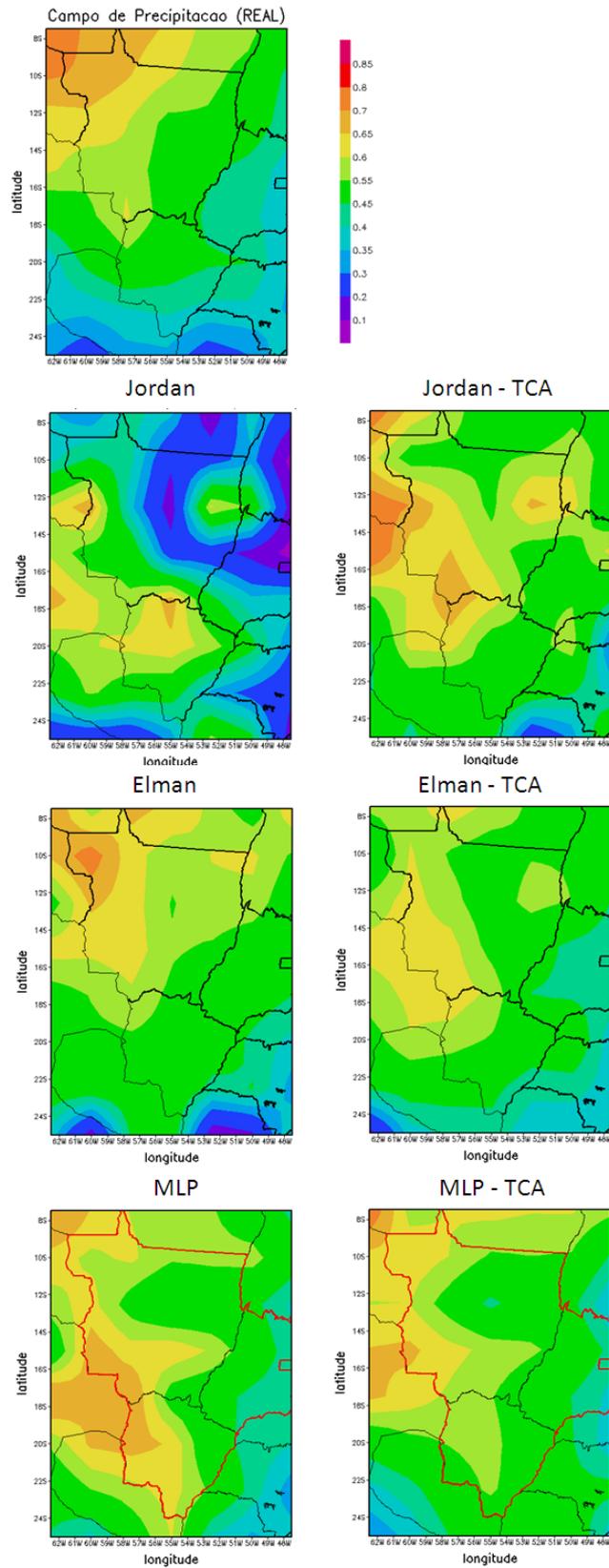


Figura 7.38 - Estimativa de precipitação no Inverno de 2008 na região Centro Oeste.

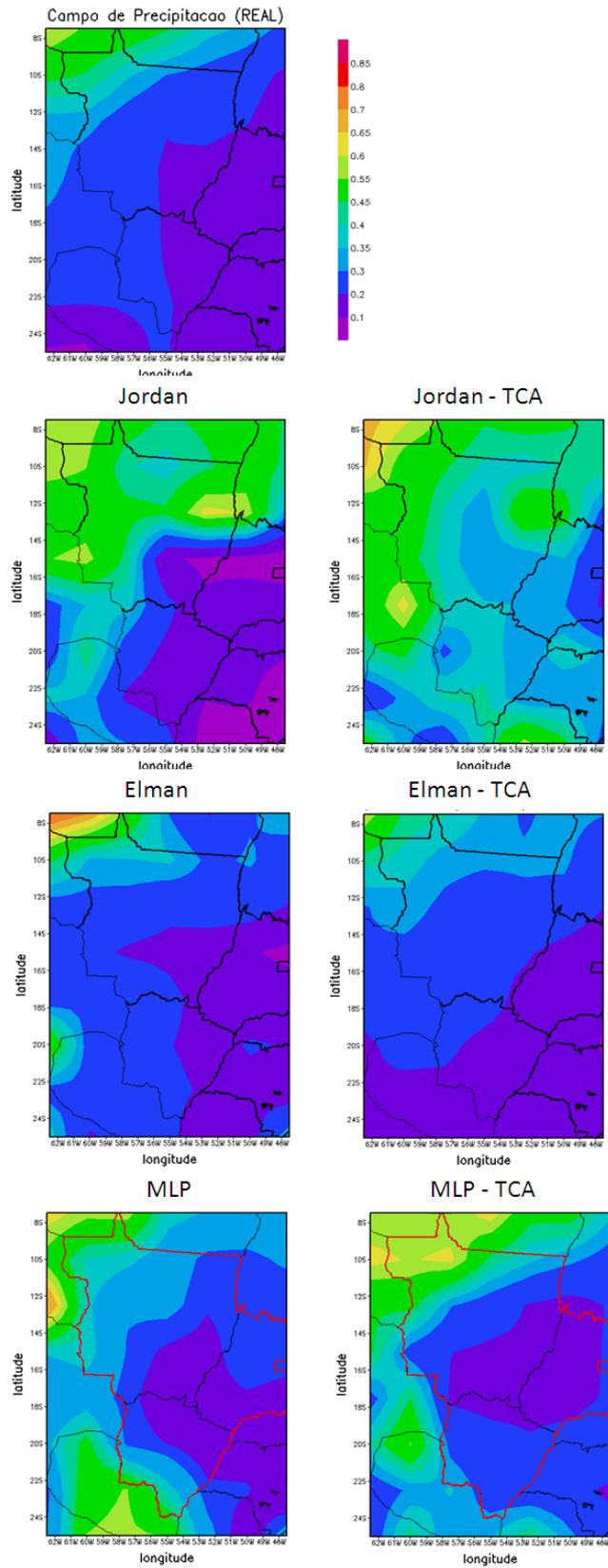


Figura 7.39 - Estimativa de precipitação no Outono de 2007 na região Centro Oeste.

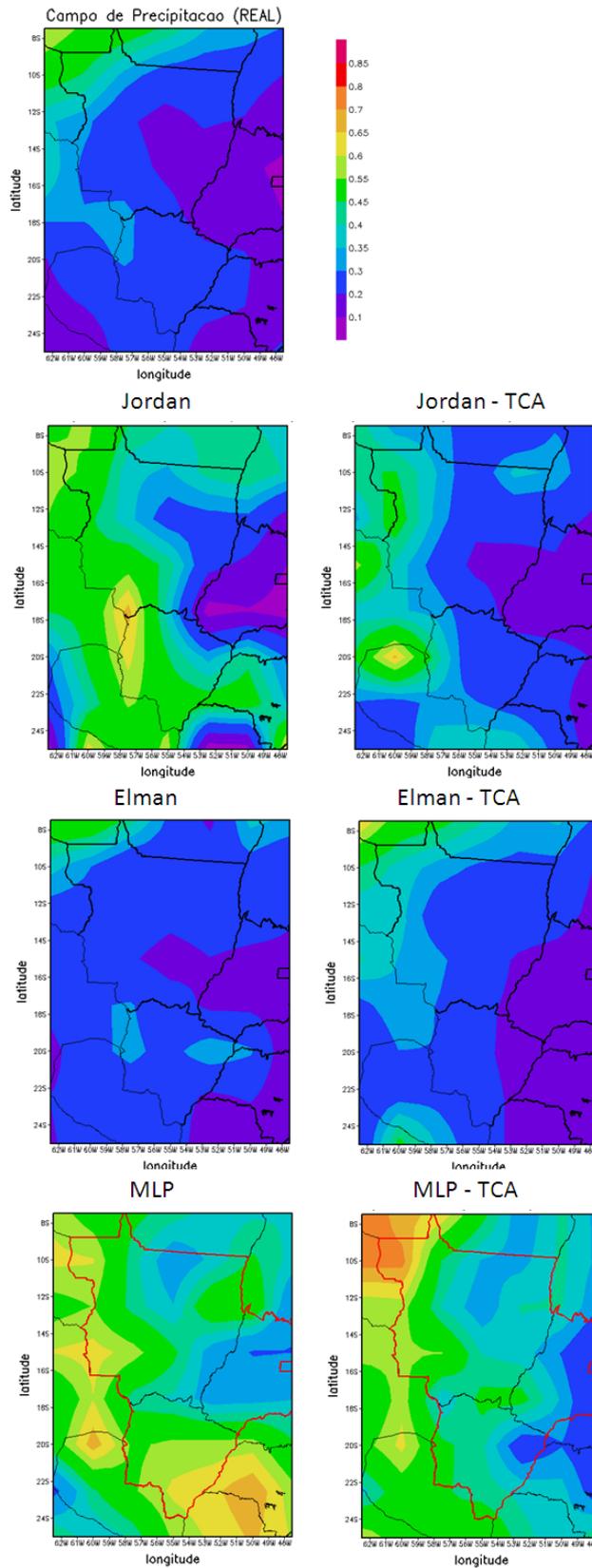


Figura 7.40 - Estimativa de precipitação no Outono de 2008 na região Centro Oeste.

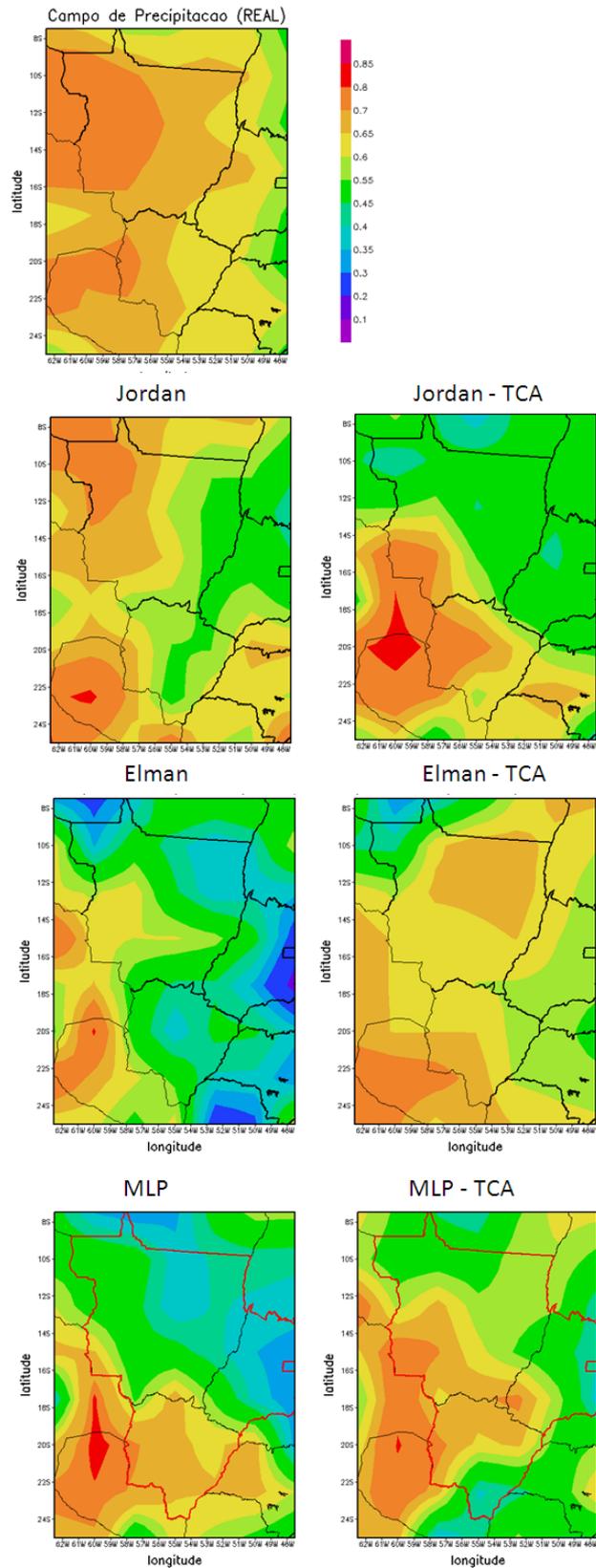


Figura 7.41 - Estimativa de precipitação na Primavera de 2007 na região Centro Oeste.

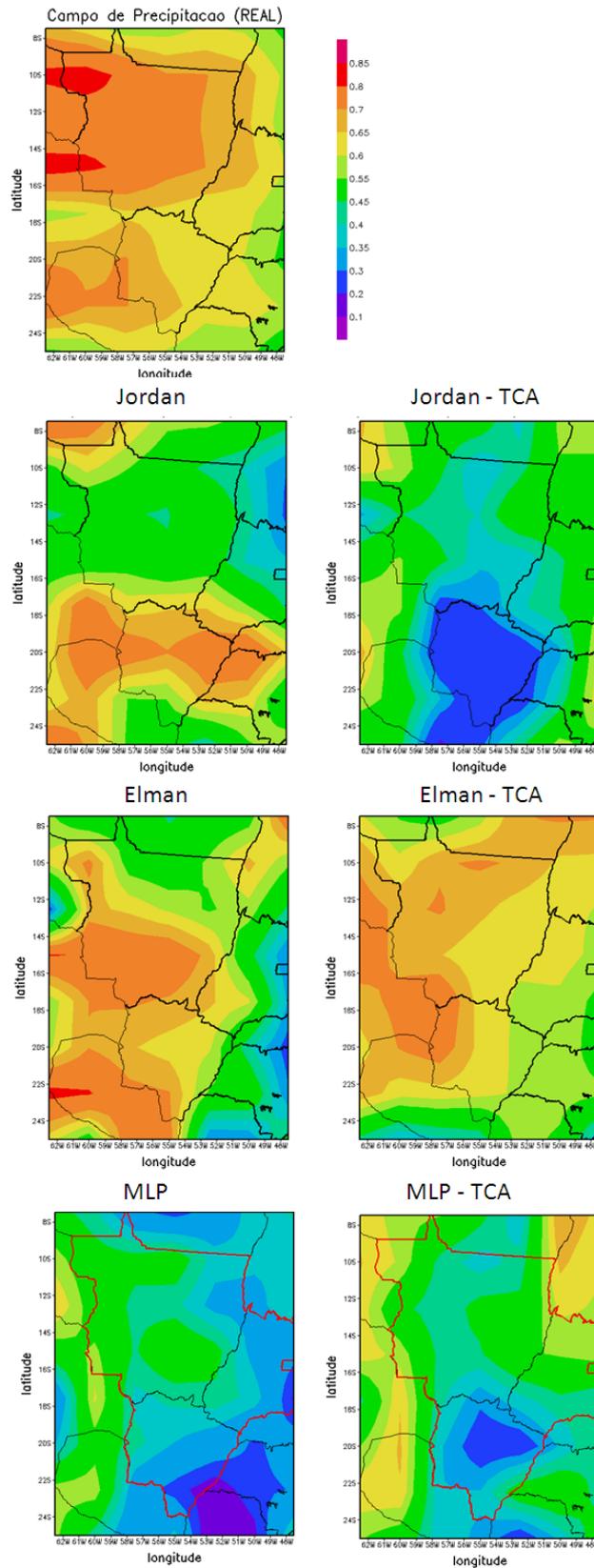


Figura 7.42 - Estimativa de precipitação na Primavera de 2008 na região Centro Oeste.

Erro em cada ponto de grade – Rede de Elman com todos os dados (CO)

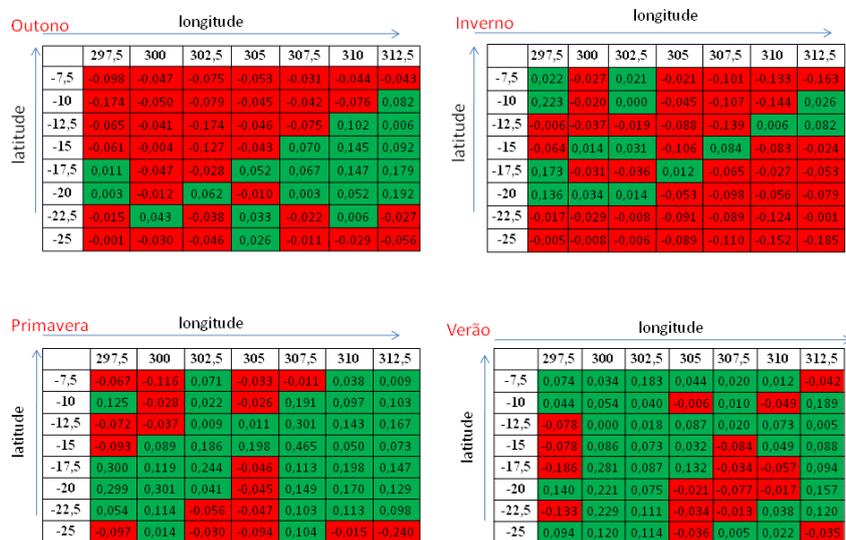


Figura 7.43 - Erro em cada ponto de grade: Rede de Elman usando todos os dados.

Erro em cada ponto de grade – Rede de Elman usando TCA (CO)

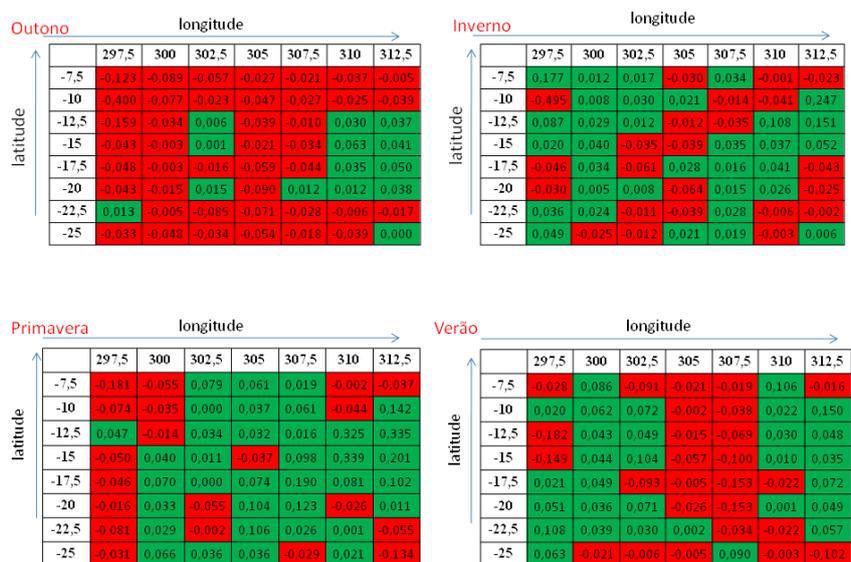


Figura 7.44 - Erro em cada ponto de grade: Rede de Elman usando os dados reduzidos.

7.4.4 Resultados na região Sudeste

As Figuras 7.45 a 7.52 apresentam os resultados de previsão do comportamento sazonal da variável de precipitação, sobre a região Sudeste do Brasil. As coordenadas geográficas que delimitam a área estão compreendidas entre as longitudes [52°W, 40°W] e entre as latitudes [25°S, 15°S], compreendendo 30 pontos de grade (5 latitudes x 6 longitudes).

A Tabela 7.17 apresenta os atributos mais relevantes para o processo de previsão na região Sudeste. Observa-se que dos 10 atributos mostrados na Tabela 6.2, apenas 5 foram considerados necessários pelo processamento da TCA, o que implica na redução da dimensão do problema.

Tabela 7.17 - Variáveis extraídas pela TCA para região Sudeste.

Redutos extraídos pela TCA	
Variáveis	%
temp	78%
u850	79%
v300	84%
v500	80%
pres	89%

A Tabela 7.18 apresenta o desempenho dos modelos de previsão em termos de erro quadrático médio, para as quatro estações dos anos de 2007 e 2008, para os modelos usando os dados com todas as variáveis disponíveis na base de dados.

Tabela 7.18 - Erro quadrático médio para as quatro estações na região Sudeste.

Região SUDESTE								
Tipo RN	Outono		Inverno		Primavera		Verão	
	2007	2008	2007	2008	2007	2008	2007	2008
Jordan	0,3822	0,8785	0,124	0,0511	0,5773	0,5697	0,1403	0,0679
Elman	2,4938	1,2269	0	0,0044	0,1298	0,1159	0,0127	0,0026
MLP	2,1273	0,3883	0,0539	1,7886	0	2,1507	2,4108	0,0048

A Tabela 7.19 apresenta os erros quadráticos médios obtidos pelas redes que utilizaram os dados reduzidos através do uso da TCA, para as quatro estações dos anos

de 2007 e 2008.

Tabela 7.19 - Erro quadrático médio nas quatro estações na região Sudeste usando TCA.

Região SUDESTE - TCA								
	Outono		Inverno		Primavera		Verão	
Tipo RN	2007	2008	2007	2008	2007	2008	2007	2008
Jordan	0,406	0,0912	0,0058	0,0465	0,8273	0,6432	0,1696	0,0103
Elman	0,7498	0,4793	0,01	0,0098	0,006	0,3711	0,0034	0,4833
MLP	2,0892	2,2069	0,3199	1,5653	0,9481	0,7176	0,4102	1,3295

As Figuras 7.45 a 7.52 apresentam a precipitação observada ou real, os resultados dos modelos Jordan, Elman e MLP, que utilizaram o conjunto de dados com todas as variáveis e os resultados dos modelos Jordan, Elman e MLP, usando os dados reduzidos pelo uso da TCA.

Na Figura 7.45 são apresentadas as estimativas de precipitação na região Sudeste do Brasil para a estação verão de 2007.

Na Figura 7.46 são apresentadas as estimativas de precipitação na estação verão para o ano de 2008.

Na Figura 7.47 são apresentadas as estimativas de precipitação para a estação inverno de 2007. Neste caso, observa-se que os resultados usando a rede de Jordan, tanto para o experimento que utiliza os dados com todas as variáveis, quanto para os dados reduzidos pela TCA, ainda que superestimadas em alguns pontos da região, apresentam padrões semelhantes aos presentes na situação Real (ver figura).

Na Figura 7.48 são apresentadas as estimativas de precipitação para a estação inverno de 2008. O resultado obtido com o modelo recorrente de Jordan, usando os dados com todas as variáveis apresenta padrões semelhantes aos presentes na situação Real (ver figura).

Na Figura 7.49 são apresentadas as estimativas de precipitação para a estação outono de 2007. O resultado obtido com o modelo de Elman, usando os dados reduzidos pela TCA apresenta os padrões mais próximos daqueles presentes na situação Real (ver figura).

Na [Figura 7.50](#) são apresentadas as estimativas de precipitação para a estação outono de 2008. Embora os resultados dos experimentos para estação outono de 2008 produzidos pelas diferentes redes neurais não apresentem padrões similares àqueles presentes na situação real, observa-se que os resultados da rede Elman que utilizou os dados processados pela TCA apresentam valores compatíveis com aqueles presentes na situação Real (ver figura).

Na [Figura 7.51](#) são apresentadas as estimativas de precipitação para a estação primavera de 2007. Neste caso, observa-se que o resultado obtido com o modelo de Jordan, o qual utiliza os dados com todas as variáveis, apresenta padrões compatíveis aos valores da situação Real (ver figura).

Na [Figura 7.52](#) são mostradas as estimativas de precipitação para a estação primavera de 2008. Observa-se que apenas o resultado obtido com a rede de Jordan utilizando os dados com todas as variáveis para estação primavera de 2008, apresenta padrões compatíveis com aqueles presentes na situação Real (ver figura).

Os resultados dos experimentos sobre a região Sudeste, referentes ao comportamento sazonal da variável de precipitação, não apresentaram padrões muito similares àqueles presentes na situação Real (ver figura).

As Tabelas [B.19](#) e [B.22](#) apresentam os erros calculados em cada ponto de grade, cujo objetivo é analisar os pontos que apresentaram subestimação ou superestimação. As demais Figuras podem ser visualizadas no [Apêndice B](#).

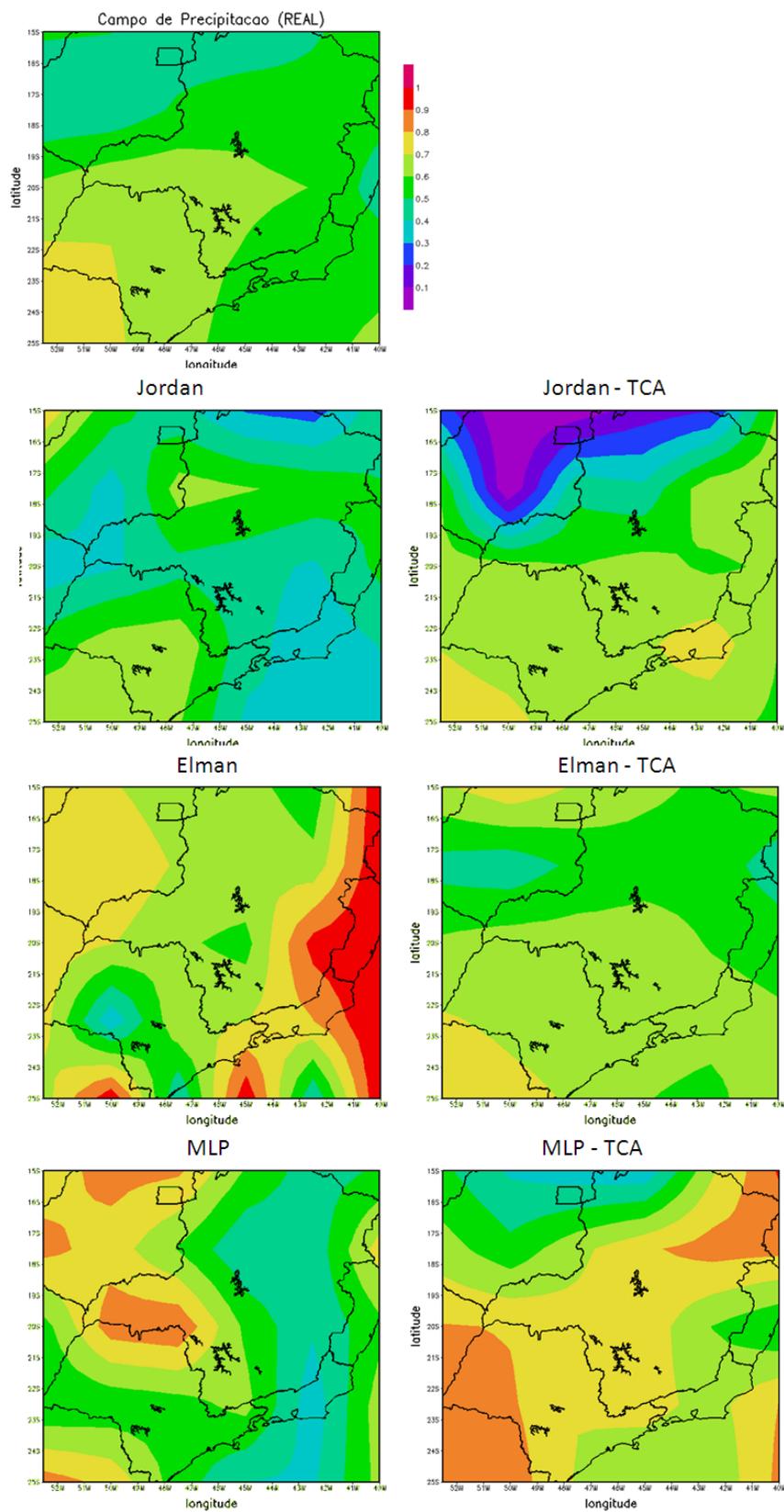


Figura 7.45 - Estimativa de precipitação no Verão de 2007 na região Sudeste.

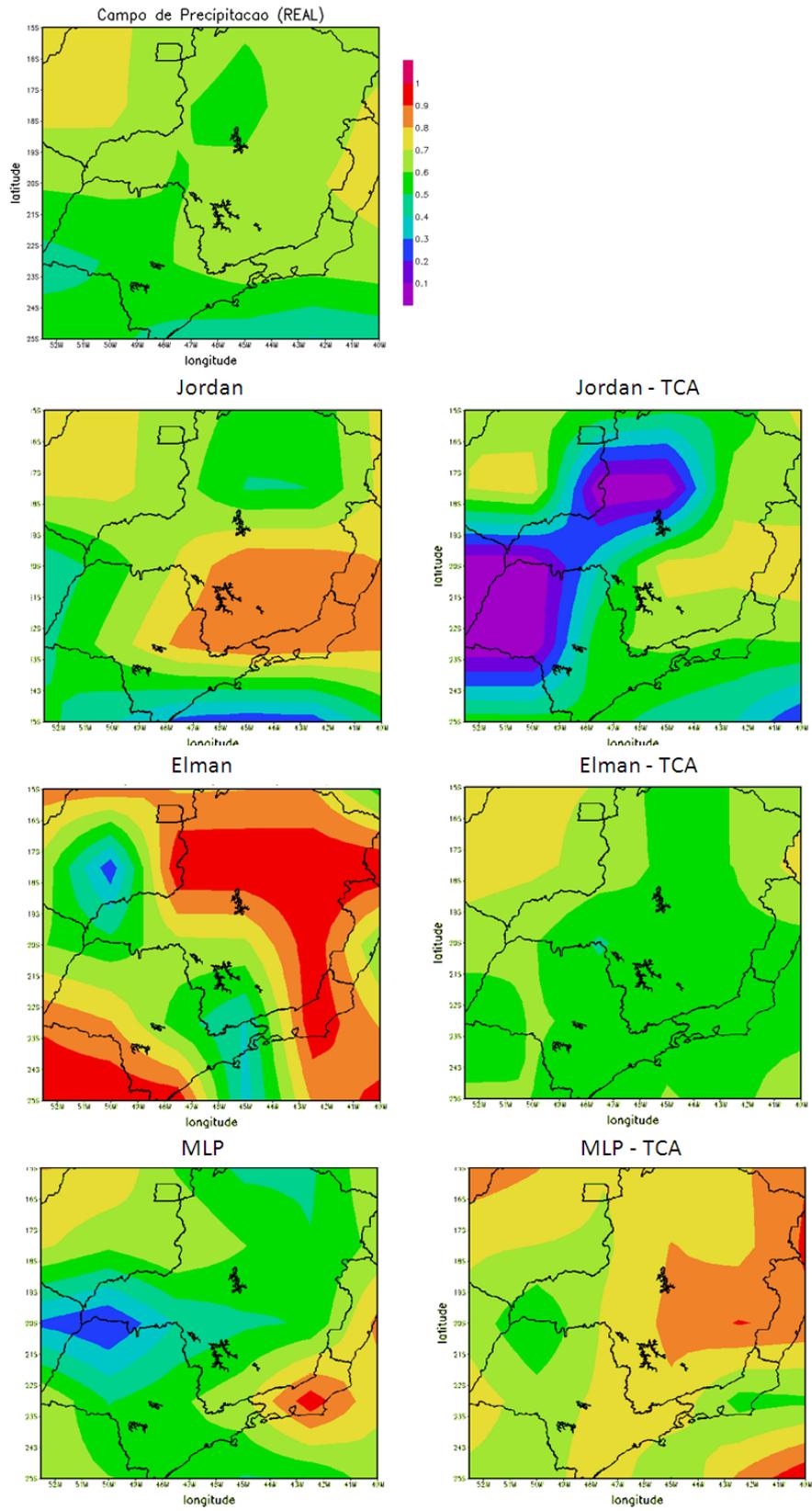


Figura 7.46 - Estimativa de precipitação no Verão de 2008 na região Sudeste.

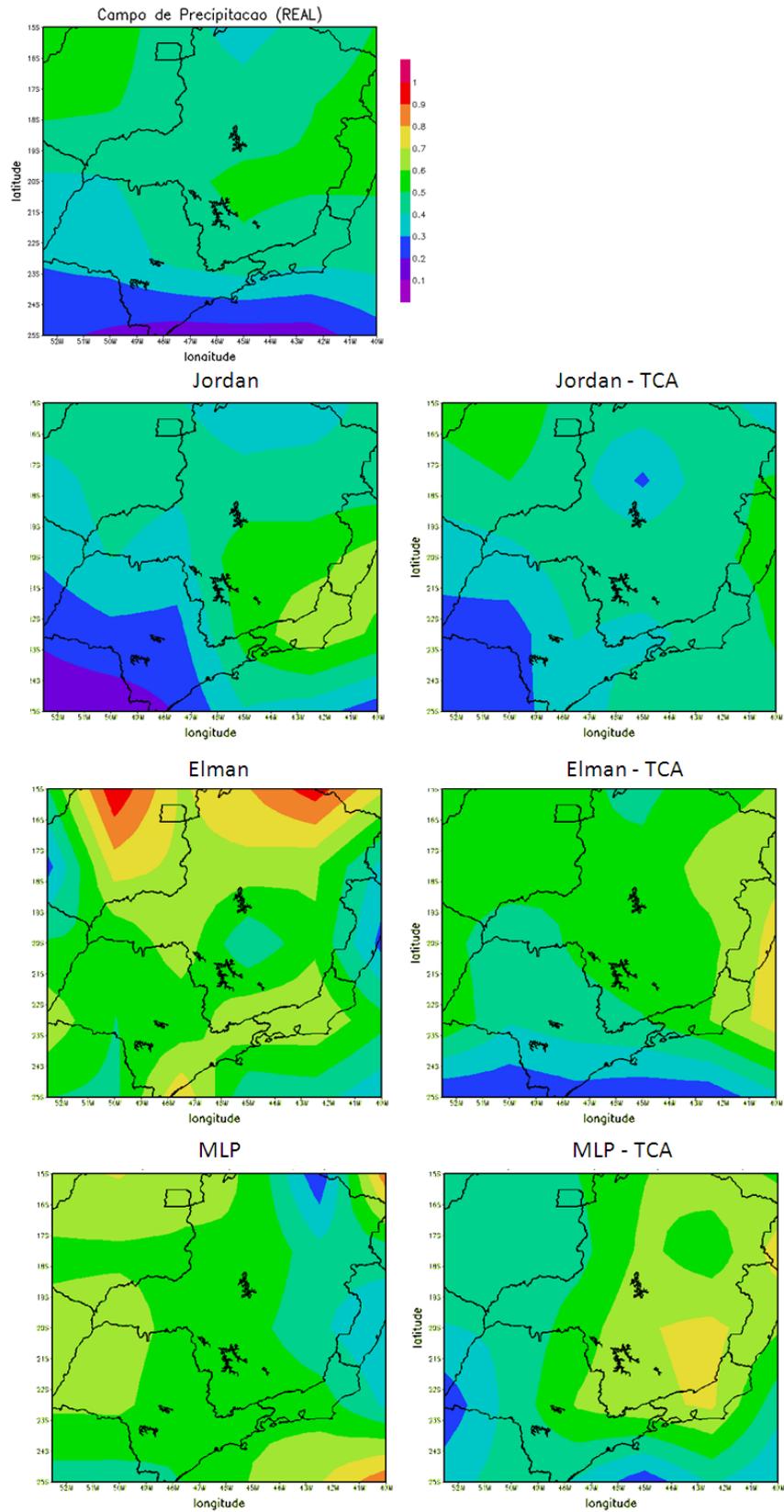


Figura 7.47 - Estimativa de precipitação no Inverno de 2007 na região Sudeste.

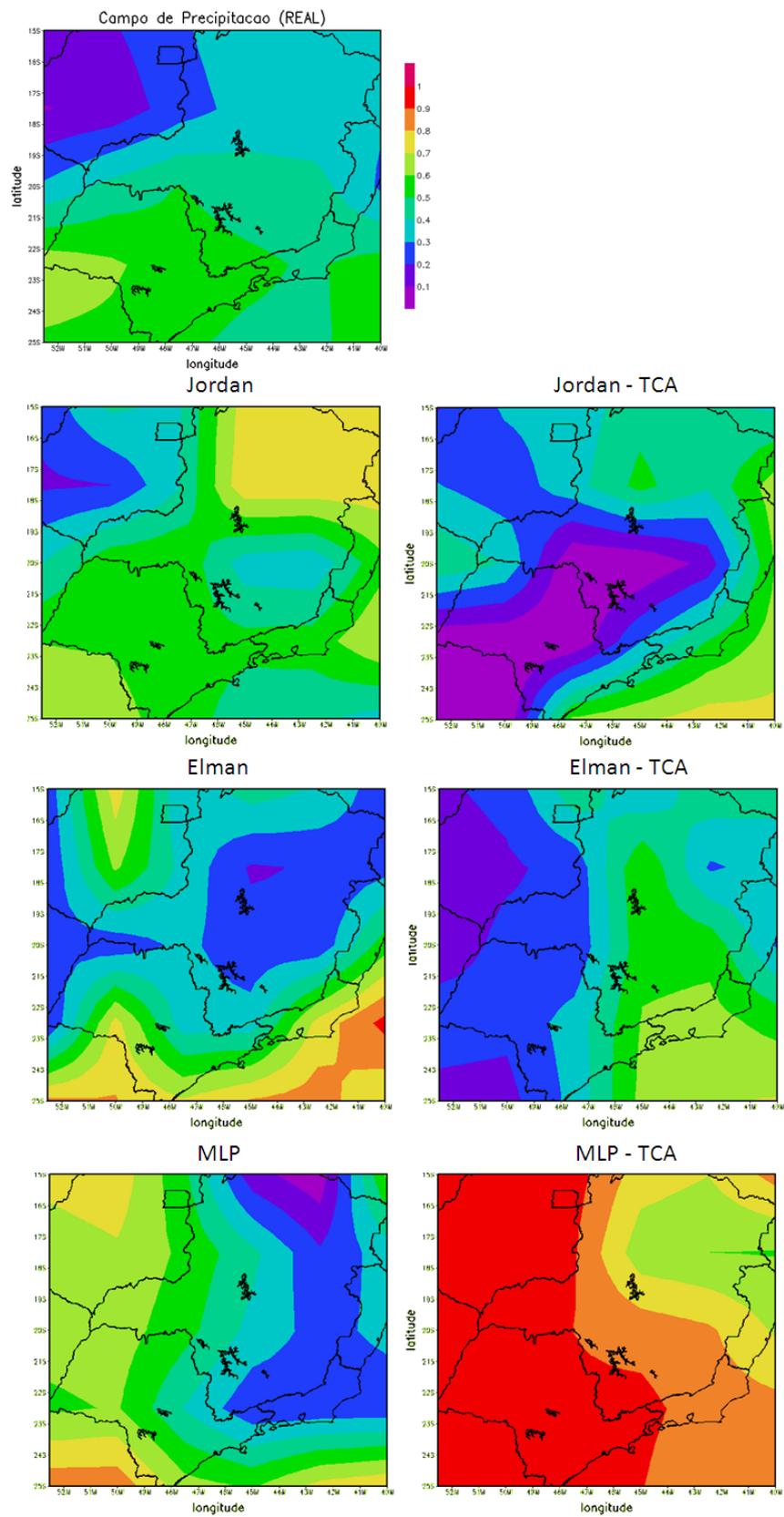


Figura 7.48 - Estimativa de precipitação no Inverno de 2008 na região Sudeste.

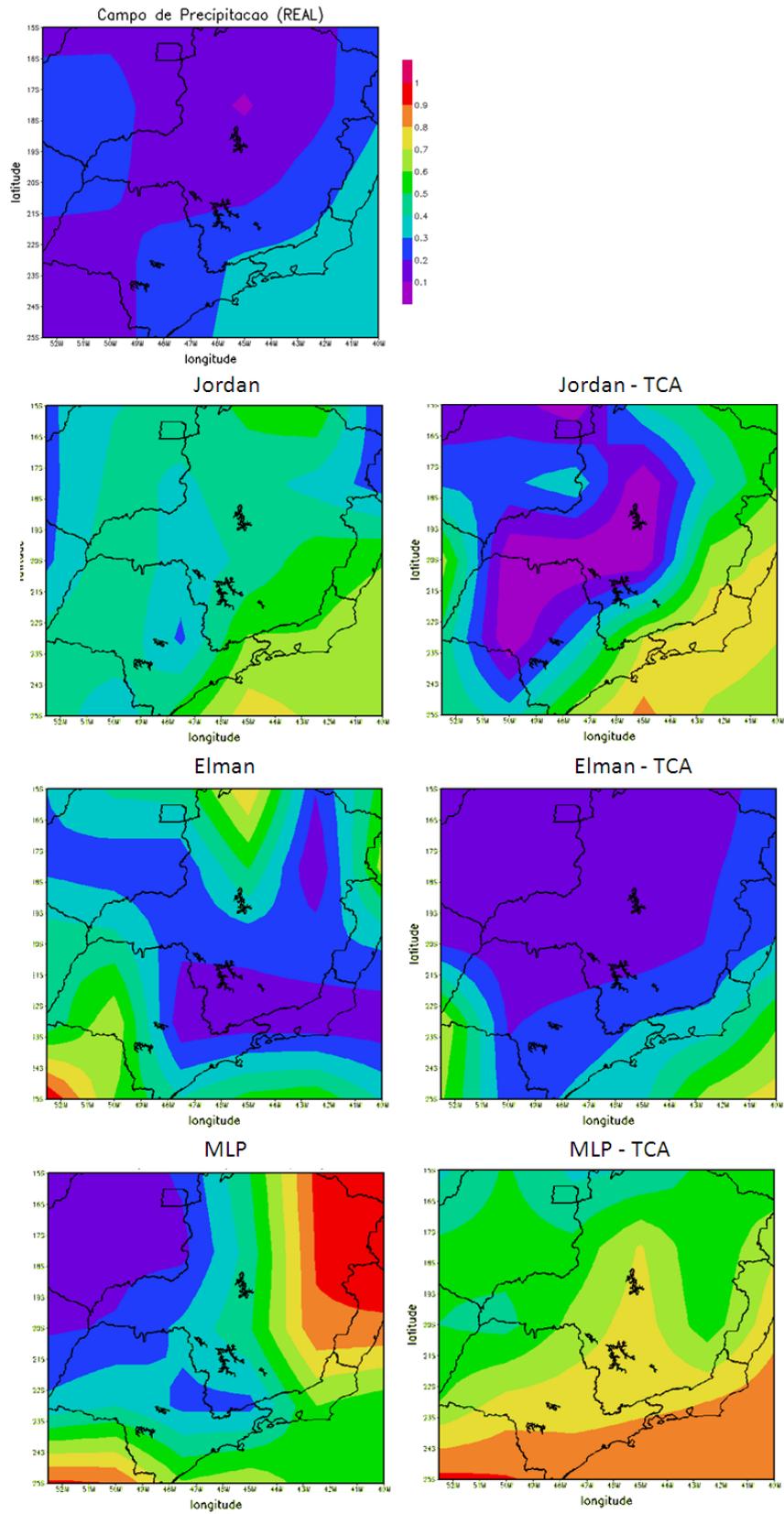


Figura 7.49 - Estimativa de precipitação no Outono de 2007 na região Sudeste.

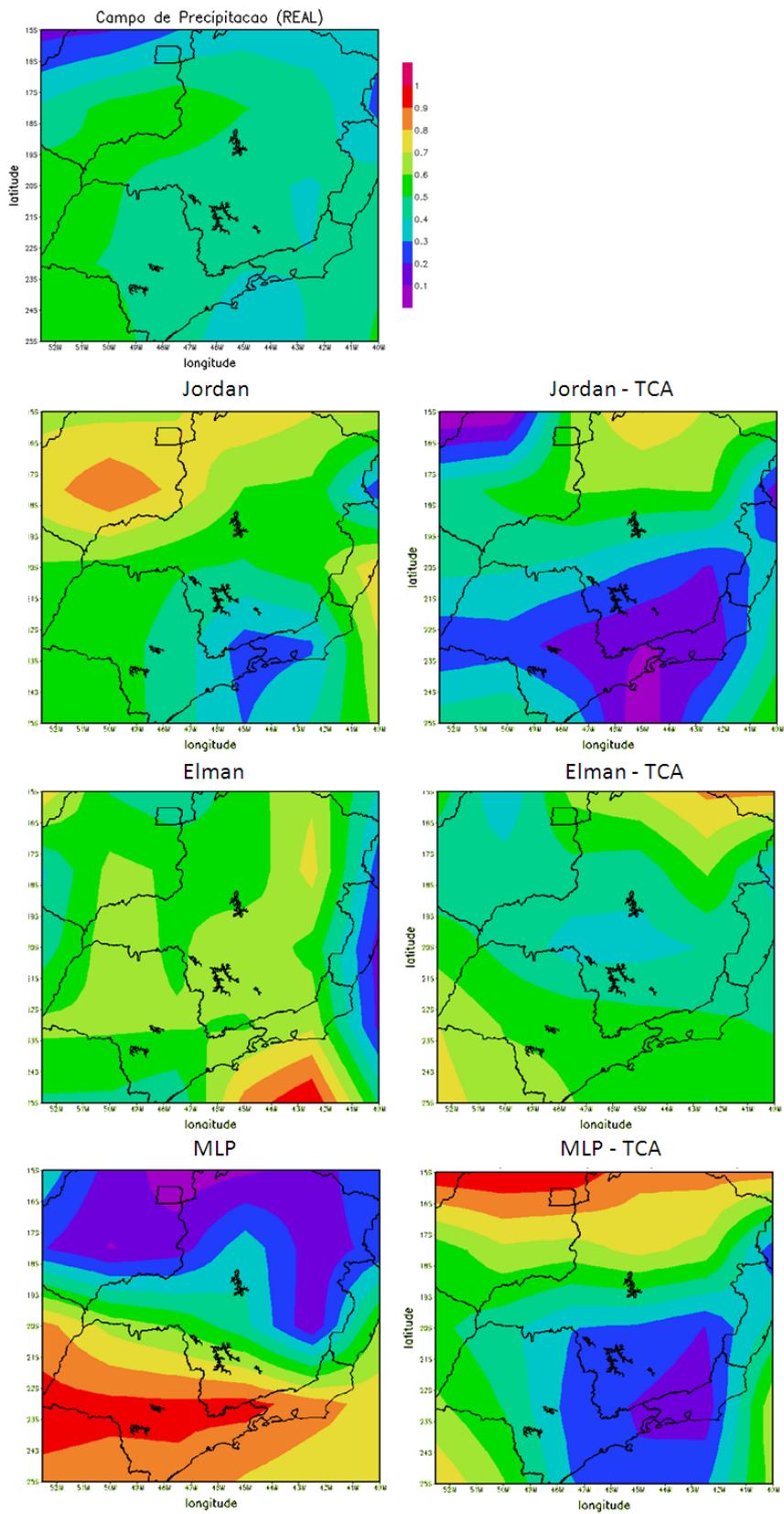


Figura 7.50 - Estimativa de precipitação no Outono de 2008 na região Sudeste.

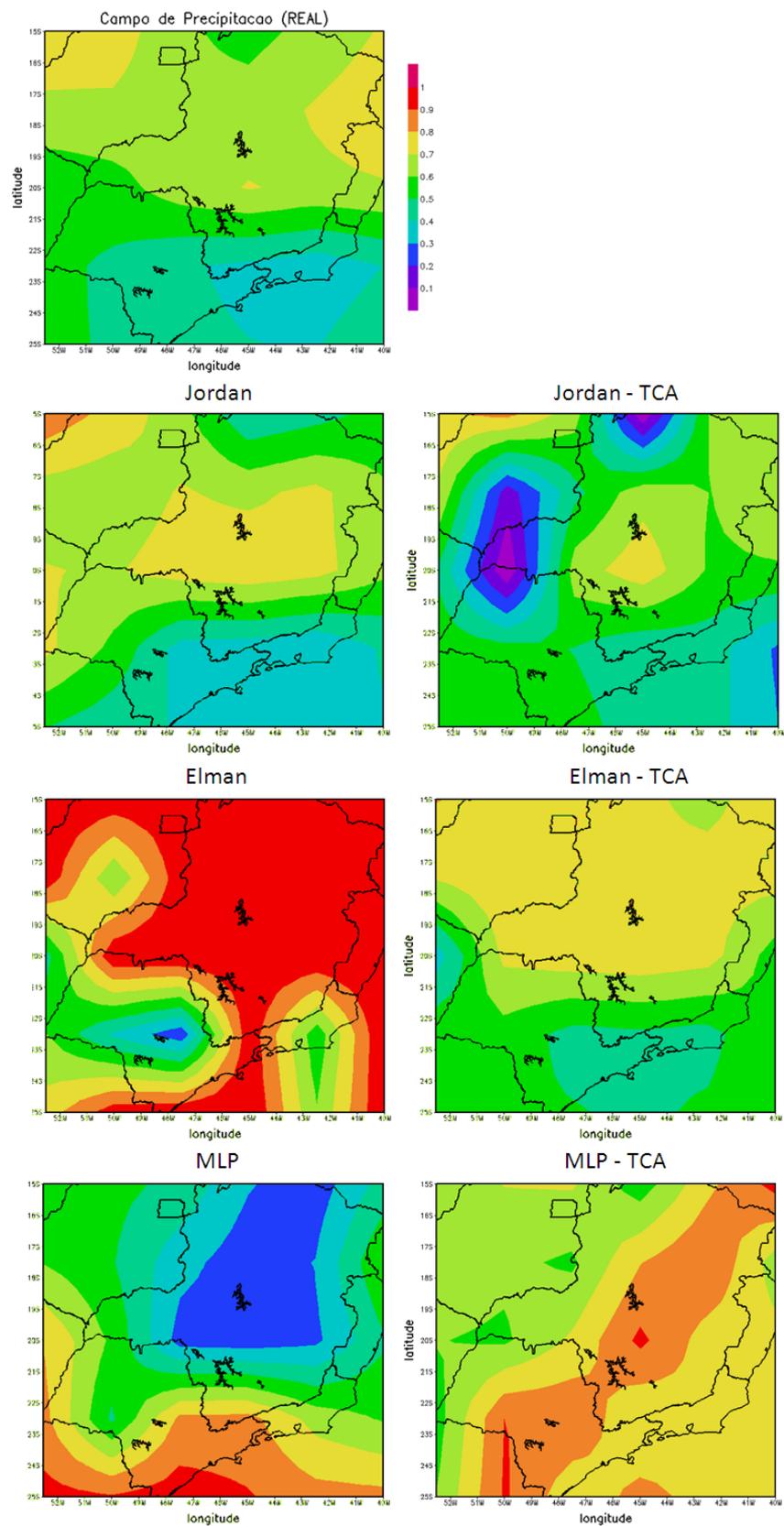


Figura 7.51 - Estimativa de precipitação na Primavera de 2007 na região Sudeste.

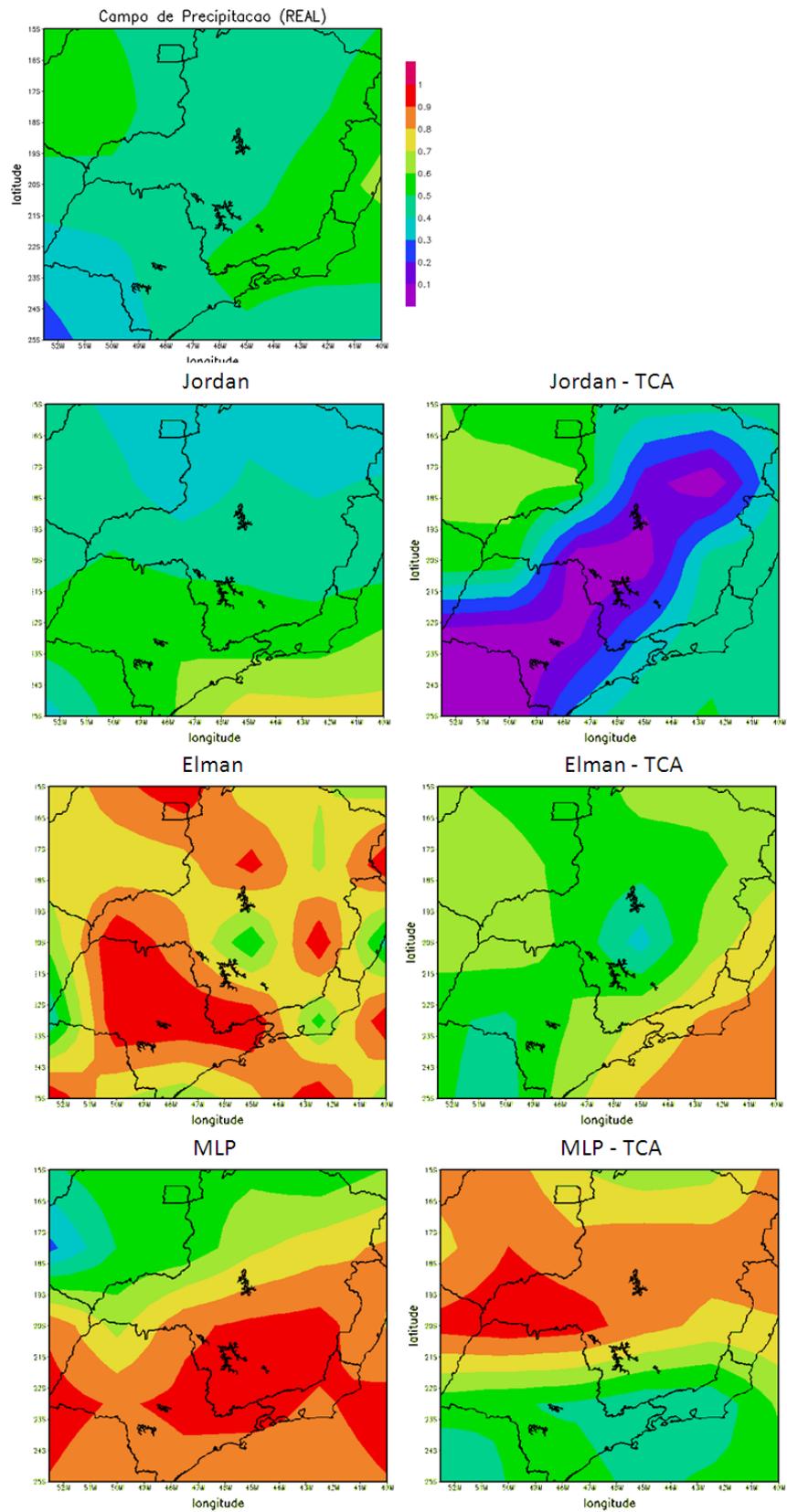


Figura 7.52 - Estimativa de precipitação na Primavera de 2008 na região Sudeste.

Erro em cada ponto de grade – Rede de Elman usando todos os dados (SE)

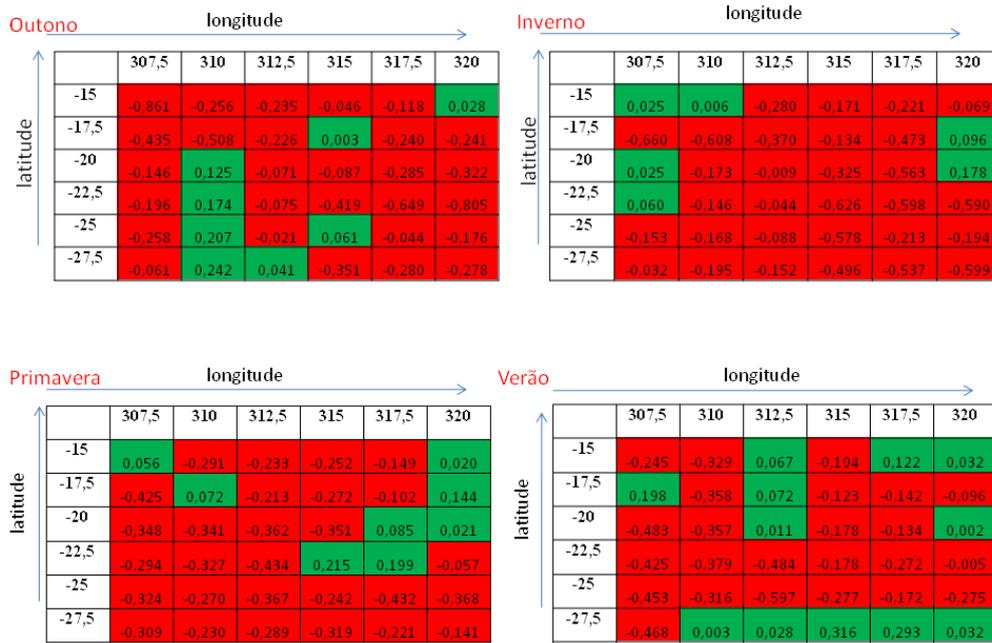


Figura 7.53 - Erro em cada ponto de grade: Rede de Elman usando todos os dados.

Erro em cada ponto de grade – Rede de Elman usando a TCA (SE)

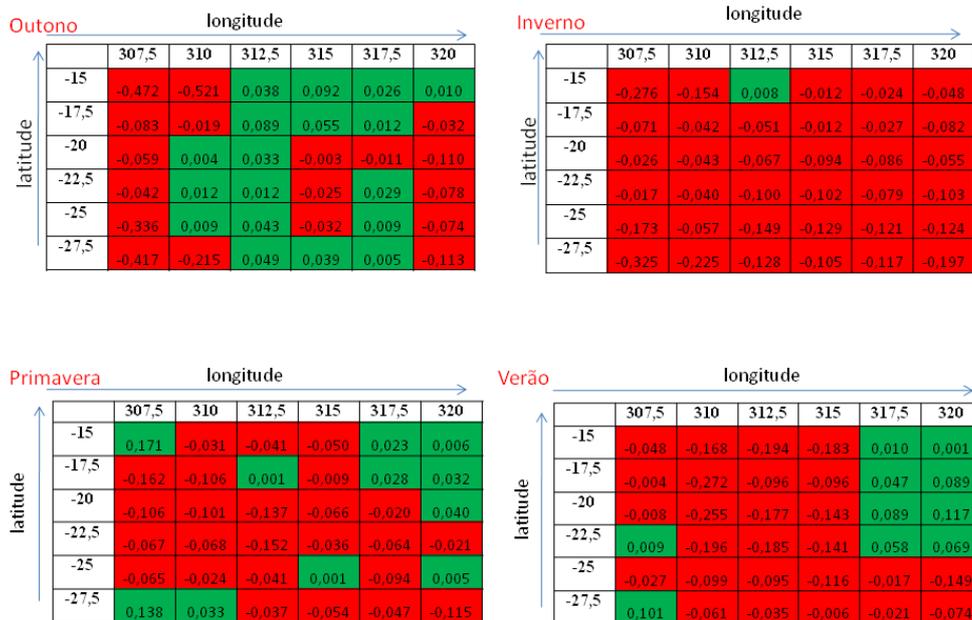


Figura 7.54 - Erro em cada ponto de grade: Rede de Elman usando a TCA.

7.4.5 Resultados na região Sul

Os resultados do comportamento sazonal da variável de precipitação na região Sul do Brasil são mostrados nas Figuras 7.55 a 7.62. As coordenadas geográficas que delimitam a área de estudo estão compreendidas entre as longitudes [60°W, 45°W] e entre as latitudes [35°S, 22.5°S], compreendendo 42 pontos de grade (7 latitudes x 6 longitudes).

Na Tabela 7.20 são apresentados os atributos mais significativos para o processo de previsão na região Sul identificados pelo processamento pela TCA. Observa-se que dos dez atributos mencionados na Tabela 6.2, sete foram considerados para o processo de previsão.

Tabela 7.20 - Variáveis extraídas pela TCA - SUL.

Redutos extraídos pela TCA	
Variáveis	%
Temp	77%
U300	79%
U850	80%
V300	77%
V500	79%
V850	70%
pres	78%

Na Tabela 7.21 é apresentado o desempenho dos modelos de redes neurais, em termos de erro quadrático médio, para as quatro estações dos anos de 2007 e 2008, que utilizaram os dados com todas as variáveis disponíveis na base de dados.

Tabela 7.21 - Erro quadrático médio para as quatro estações da região Sul.

Região SUL								
	Outono		Inverno		Primavera		Verão	
Tipo RN	2007	2008	2007	2008	2007	2008	2007	2008
Jordan	0,4354	0,0612	0,0249	0,0072	0,0455	0,0126	0,012	0,0197
Elman	1,0817	0,0192	0,0437	0,4866	0,0316	0,0211	0,0599	0,2608
MLP	0,0549	0,1099	0,0043	0,031	0,0001	0	0,0023	0,0017

A [Tabela 7.22](#) mostra o desempenho dos modelos de redes neurais, para as quatro estações dos anos de 2007 e 2008, que utilizaram os dados reduzidos pela TCA.

Tabela 7.22 - Erro quadrático médio das previsões feitas pelas redes neurais nas quatro estações da região Sul usando dados reduzidos pela TCA.

Região SUL - TCA								
Tipo RN	Outono		Inverno		Primavera		Verão	
	2007	2008	2007	2008	2007	2008	2007	2008
Jordan	0,1949	0,0201	0,1465	0,0282	0,0033	0,1489	0,005	0,0362
Elman	0,7391	0,0014	0,036	0,1476	0,0004	0,0003	0,002	0,0309
MLP	0,2333	0,0122	0,1135	0,0861	0,0274	0,004	0,0013	0,0051

Nas Figuras [7.55](#) a [7.62](#) são apresentados respectivamente a situação Real da precipitação, os resultados obtidos pelos modelos de Jordan, Elman e MLP usando os dados com todas as variáveis disponíveis e os resultados obtidos com os modelos Jordan, Elman e MLP, usando os dados processados pela TCA.

Na [Figura 7.56](#) é apresentado o resultado da estimativa de precipitação na estação verão 2008. Observando os resultados obtidos, o modelo MLP usando os dados reduzidos, apresenta padrões mais semelhantes àqueles presentes na situação Real (ver figura).

A [Figura 7.57](#) apresenta o resultado da estimativa de precipitação na estação inverno de 2007. O resultado obtido com o modelo de Elman, que utilizou os dados reduzidos, apresenta resultado semelhante à situação Real, em termos dos padrões presentes.

Na [Figura 7.58](#) é mostrado o resultado da estimativa de precipitação na estação inverno de 2008. Observa-se que o modelo de Elman usando os dados reduzidos através da TCA, produziu o resultado de previsão com padrões mais similares àqueles presentes na situação Real (ver figura).

A [Figura 7.59](#) apresenta o resultado da estimativa de precipitação na estação outono de 2007. O modelo Elman usando os dados reduzidos, produziu um resultado com padrões mais similares àqueles presentes na situação Real (ver figura).

Na [Figura 7.60](#) é apresentado o resultados da estimativa de precipitação na estação outono para o ano de 2008. Neste o caso, os modelos de previsão não produziram

resultados compatíveis em relação ao valor real.

Nas Figuras 7.61 e 7.62, são apresentados os resultados da estimativa de precipitação na estação primavera para os anos de 2007 e 2008.

Com base nos experimentos realizados na região Sul do Brasil, observou-se que com o uso da abordagem por TCA, apresentou resultados mais similares aos padrões presentes na situação real. Entretanto os erros quadráticos médios obtidos a partir do modelo recorrente de Jordan usando a TCA, apresentaram erros maiores em relação ao modelo de Jordan usando todos os dados, apesar de mostrarem um padrão mais semelhante à situação real.

As Tabelas B.26 e B.29 apresentam os erros calculados em cada ponto de grade, cujo objetivo é analisar os pontos que apresentaram subestimação ou superestimação. As demais tabelas contendo os erros por ponto de grade podem ser visualizadas no Apêndice B

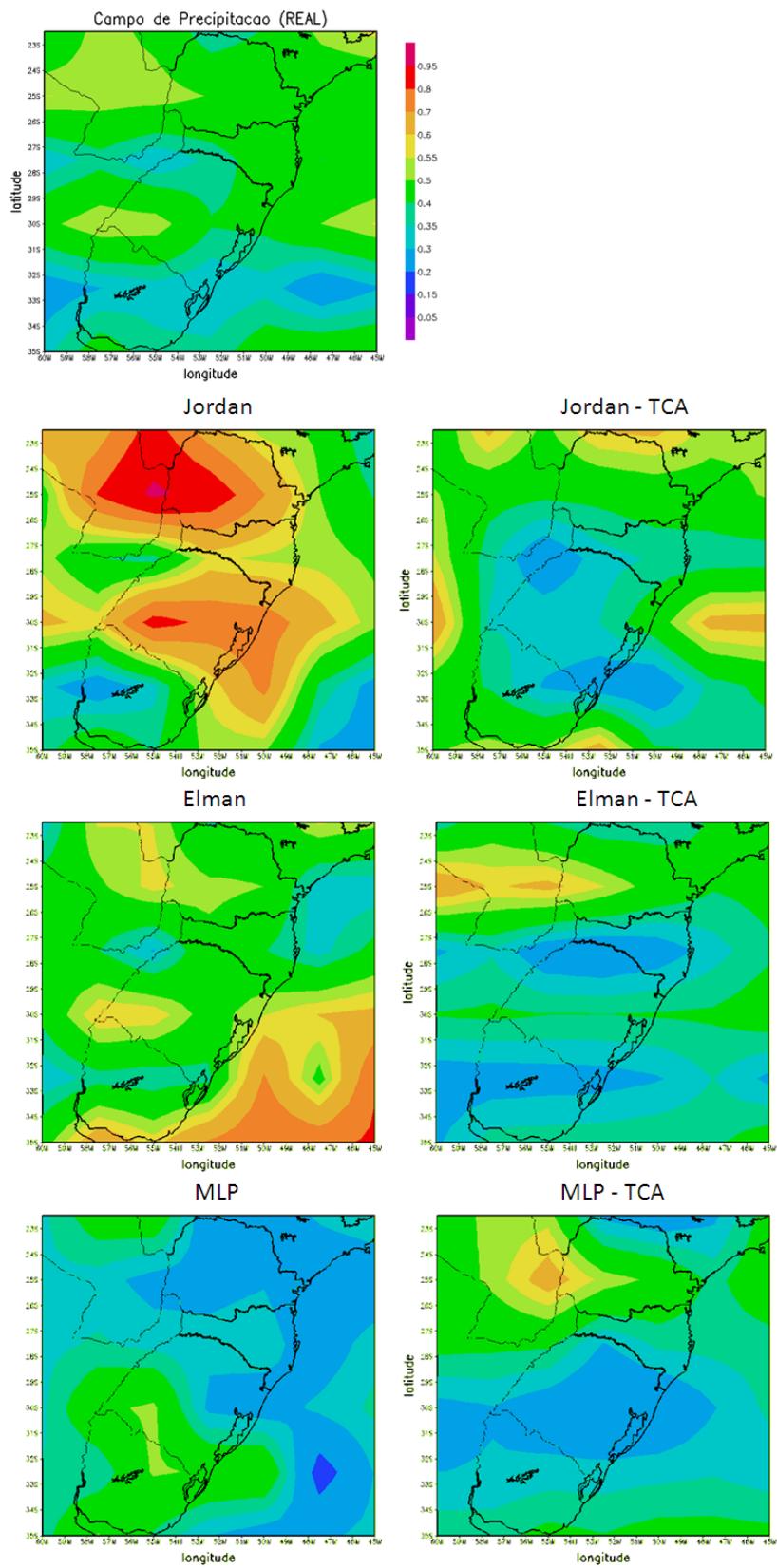


Figura 7.55 - Estimativa de precipitação no Verão de 2007 na região Sul.

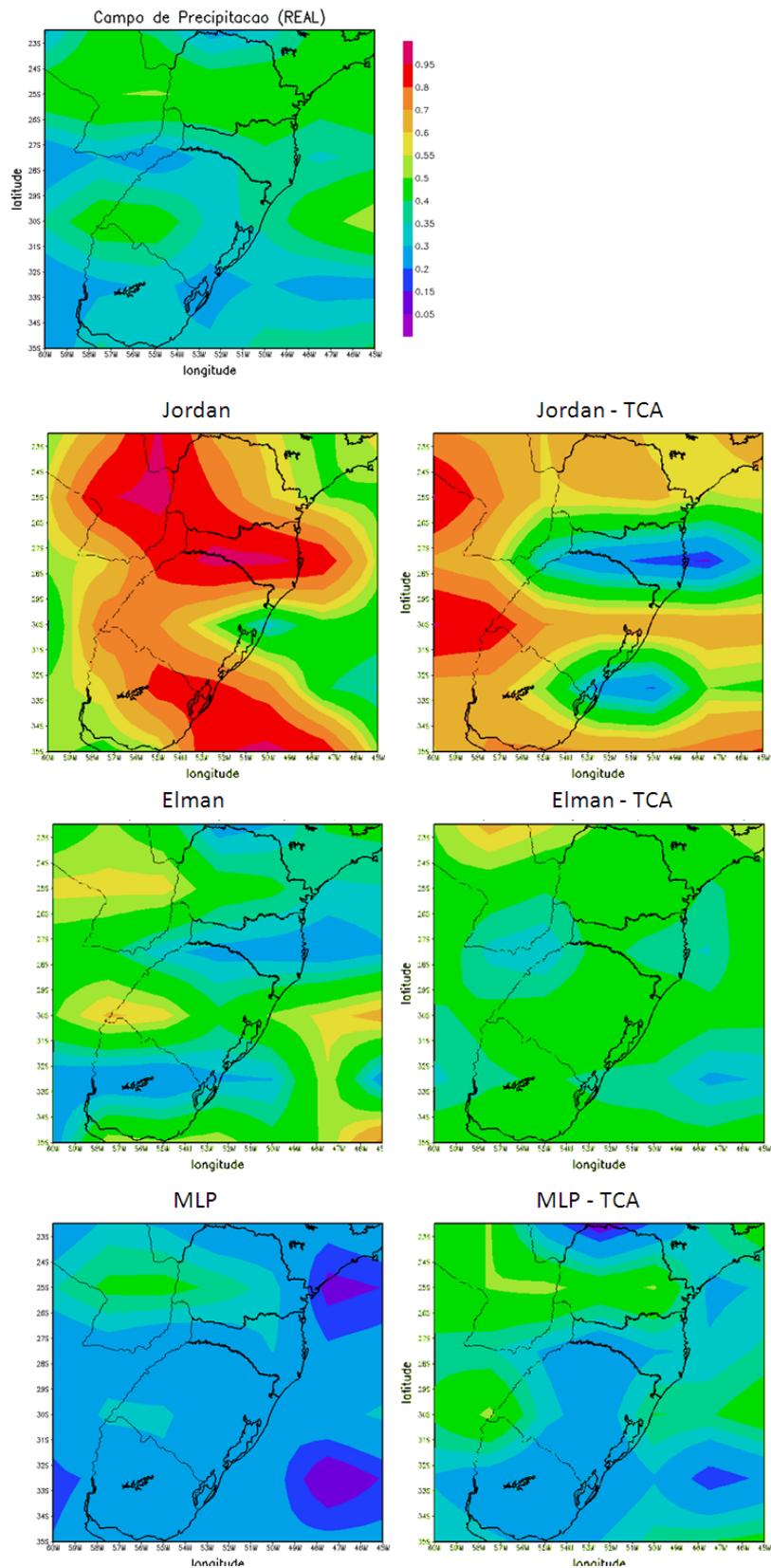


Figura 7.56 - Estimativa de precipitação no Verão de 2008 na região Sul.

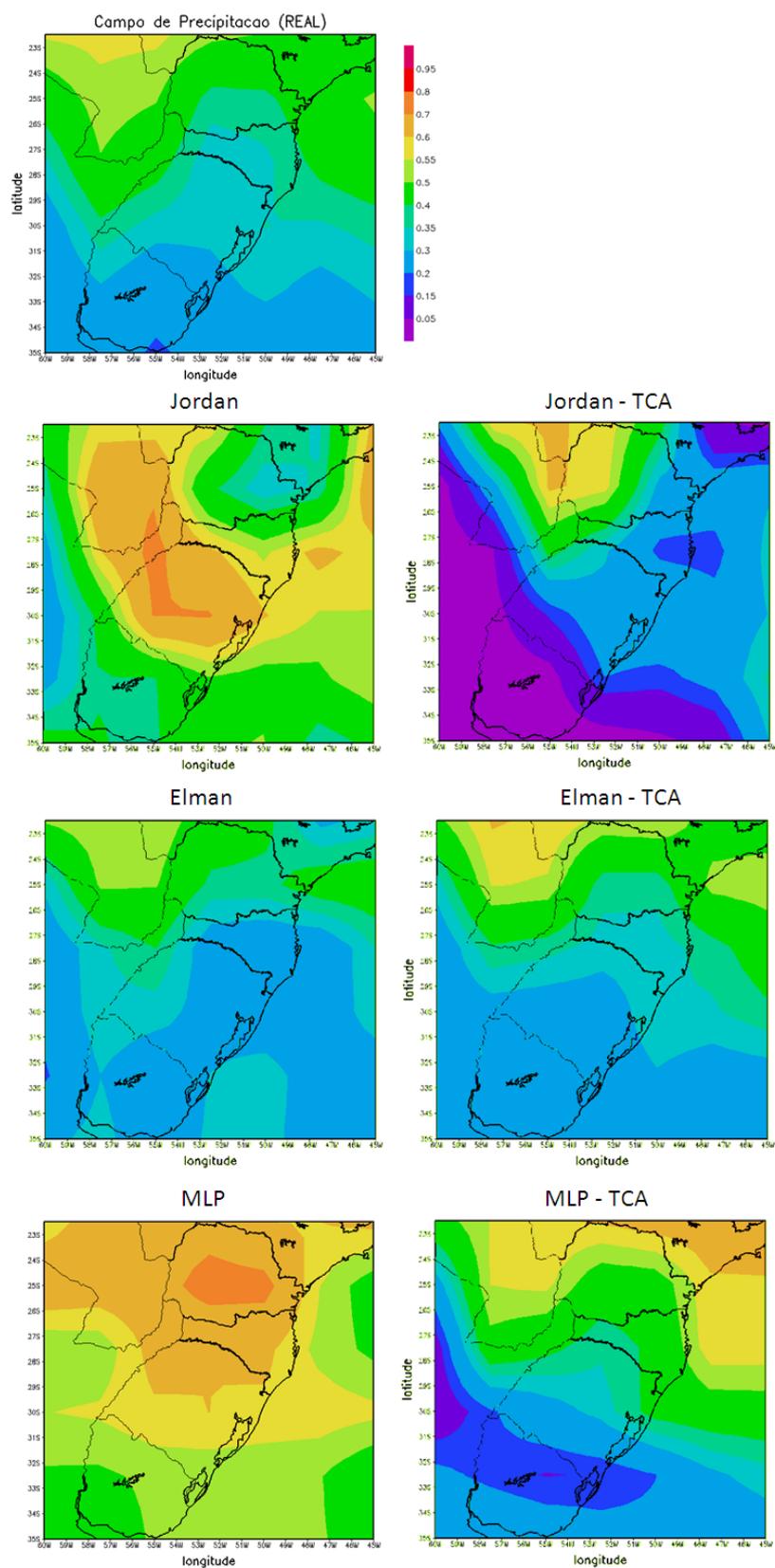


Figura 7.57 - Estimativa de precipitação no Inverno de 2007 na região Sul.

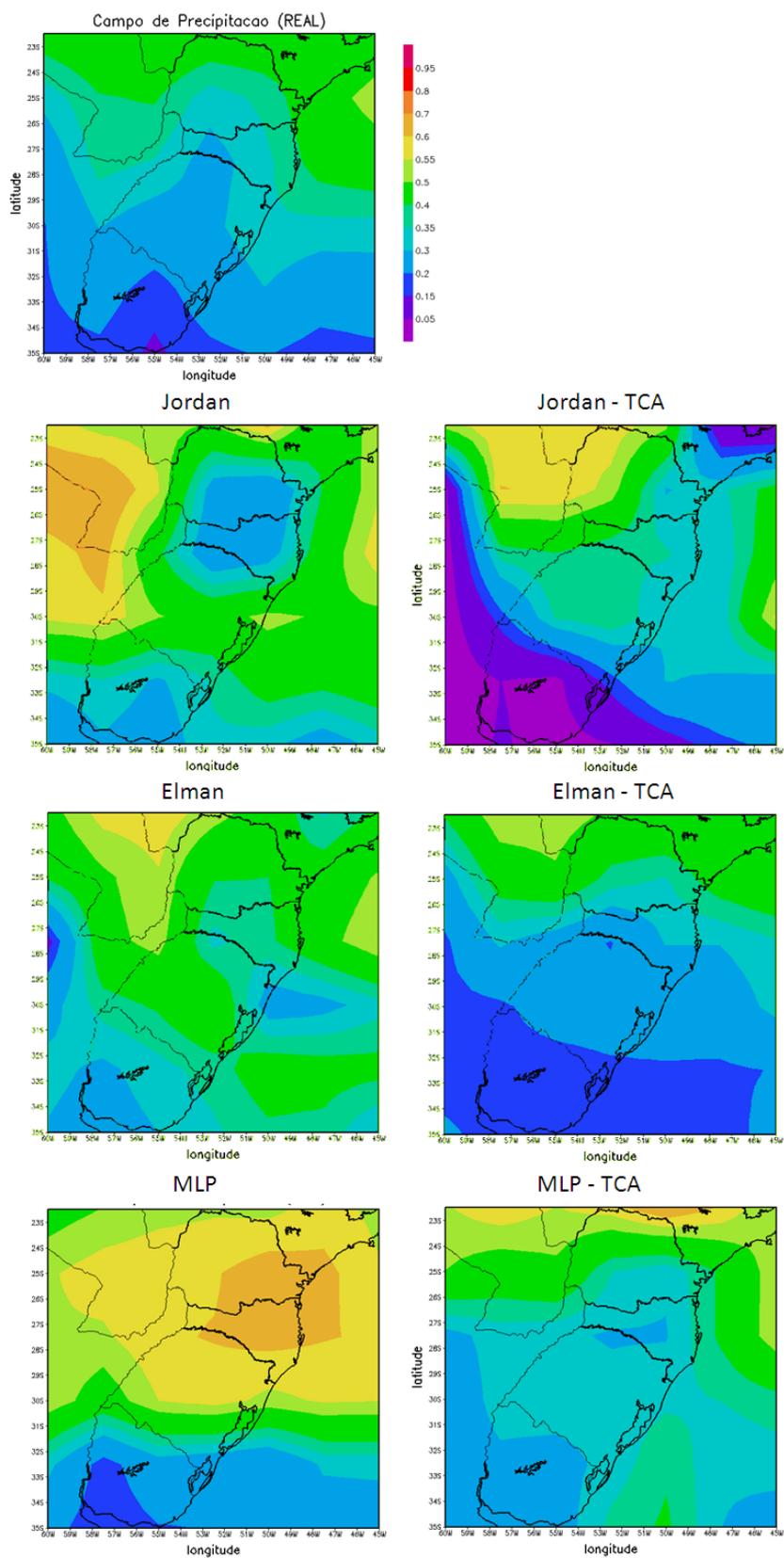


Figura 7.58 - Estimativa de precipitação no Inverno de 2008 na região Sul.

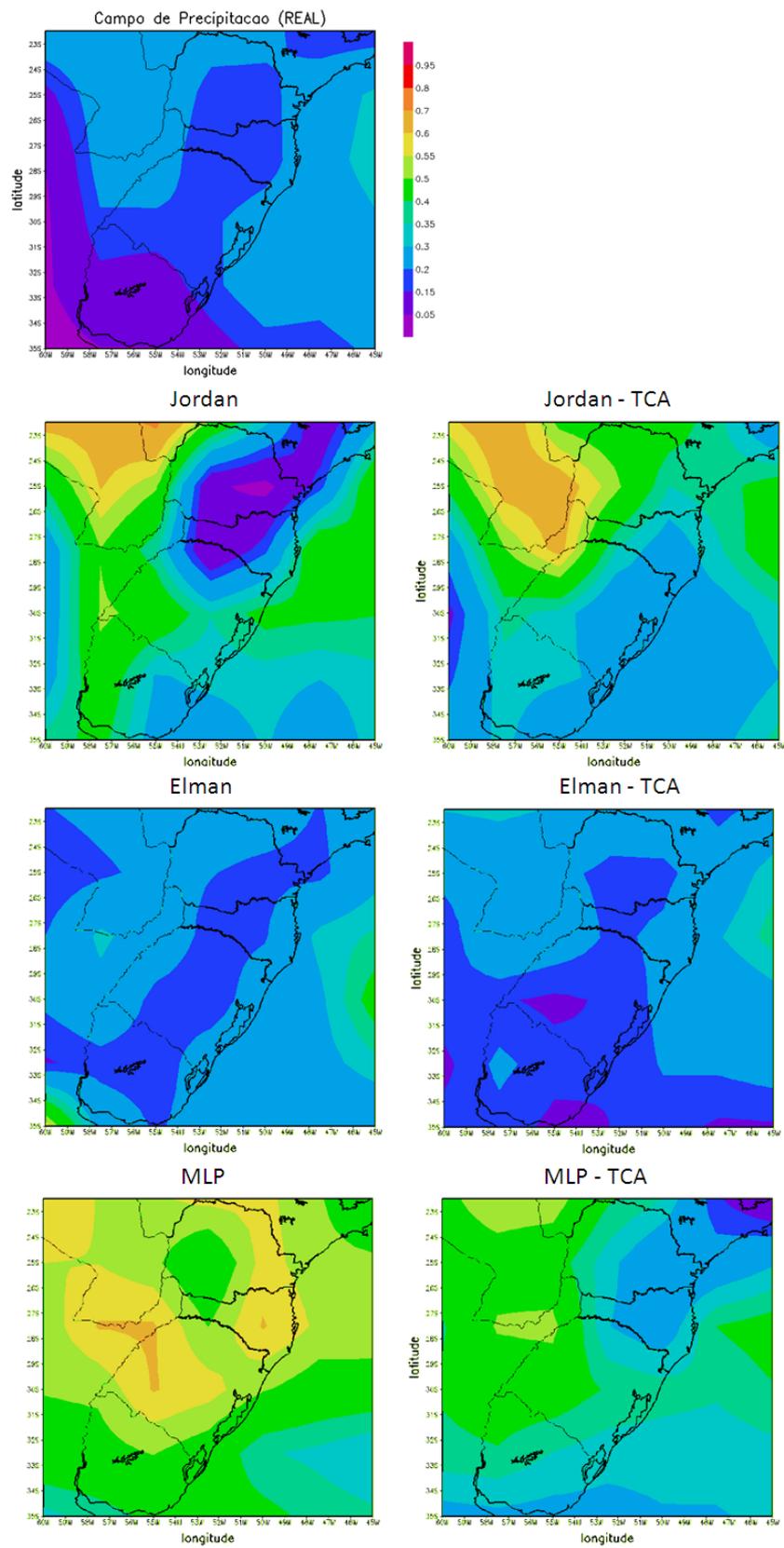


Figura 7.59 - Estimativa de precipitação no Outono de 2007 na região Sul.

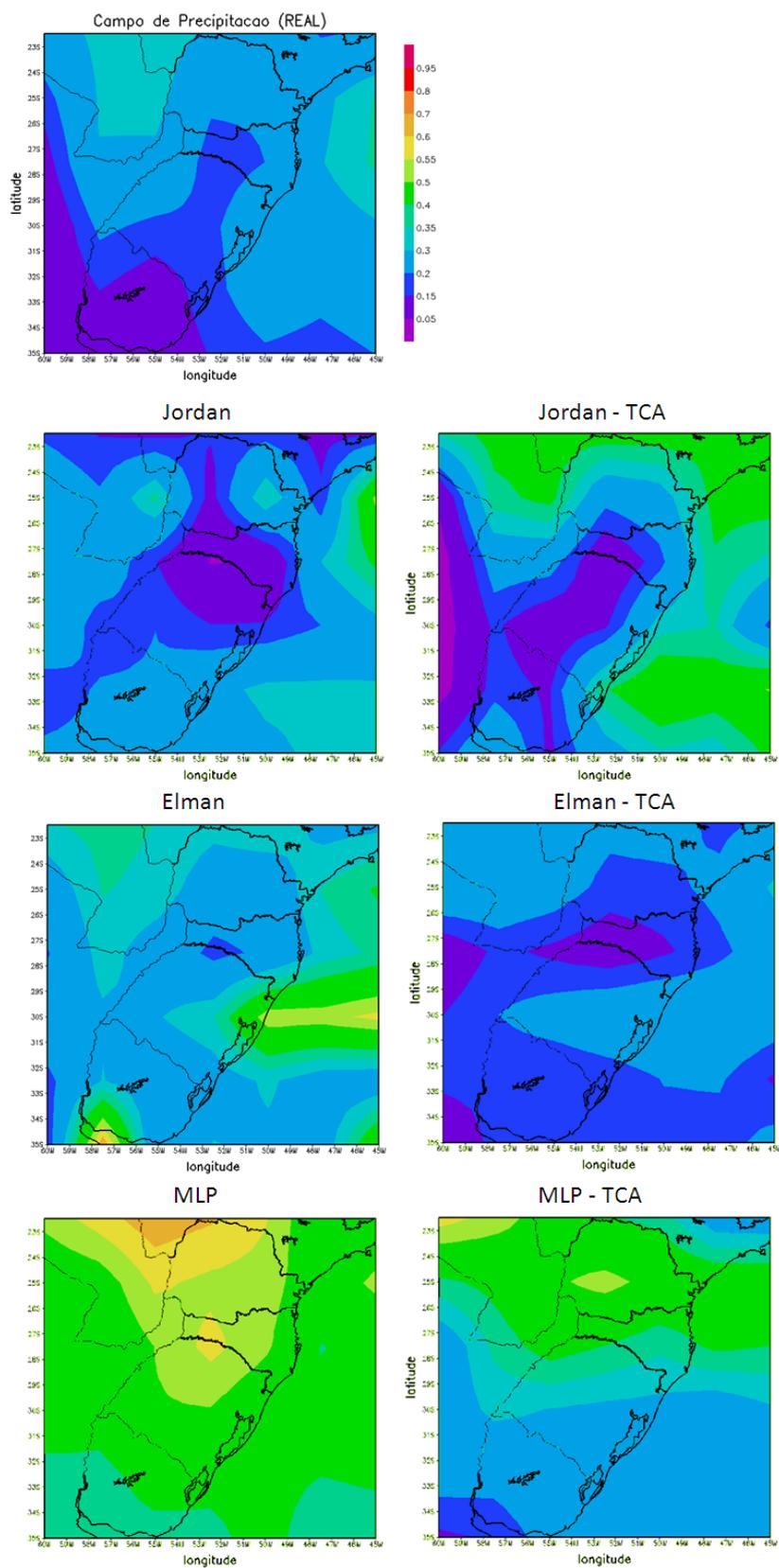


Figura 7.60 - Estimativa de precipitação no Outono de 2008 na região Sul.

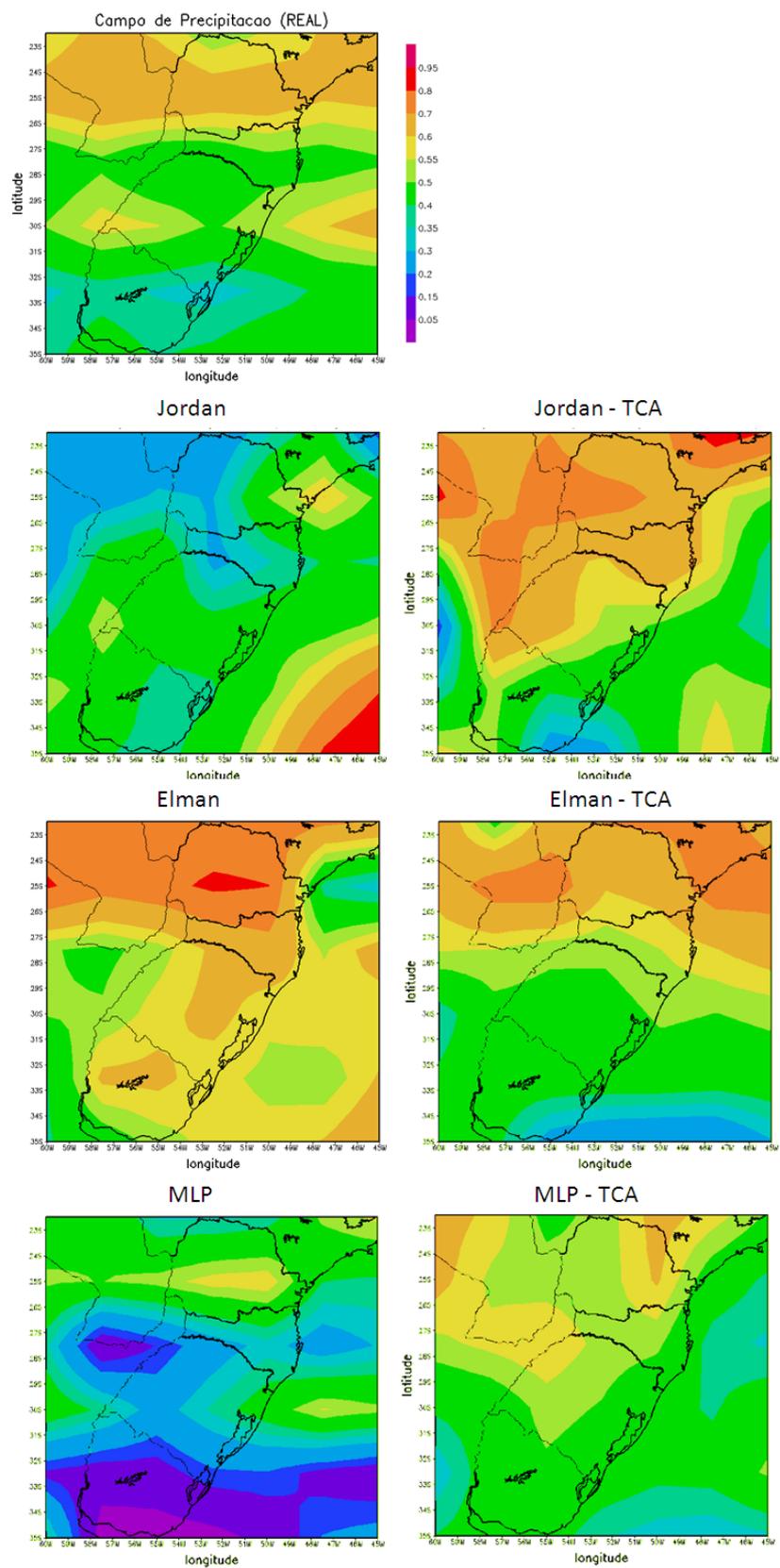


Figura 7.61 - Estimativa de precipitação na Primavera de 2007 na região Sul.

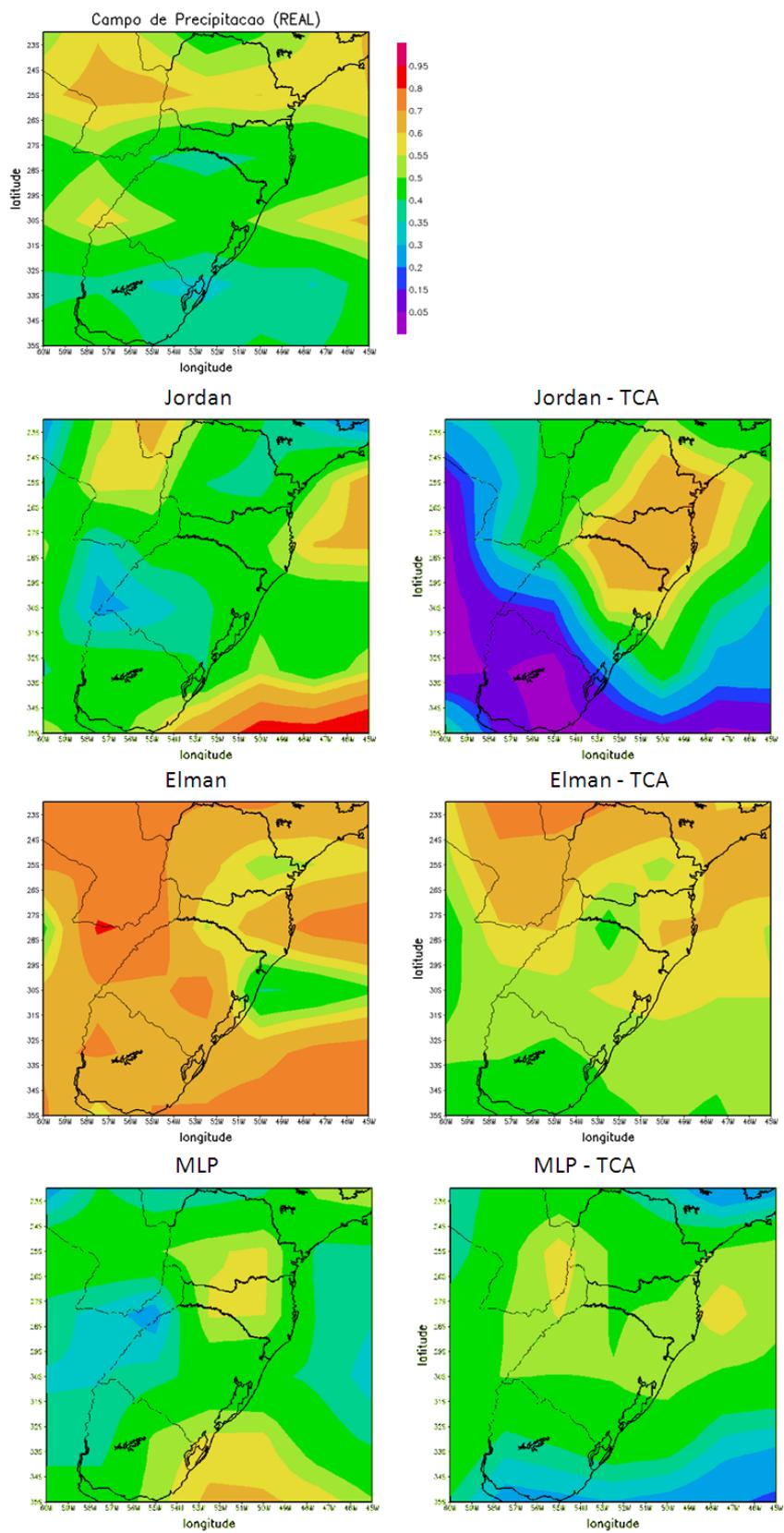


Figura 7.62 - Estimativa de precipitação na Primavera de 2008 na região Sul.

Erro em cada ponto de grade – Rede de Jordan usando todos os dados (S)

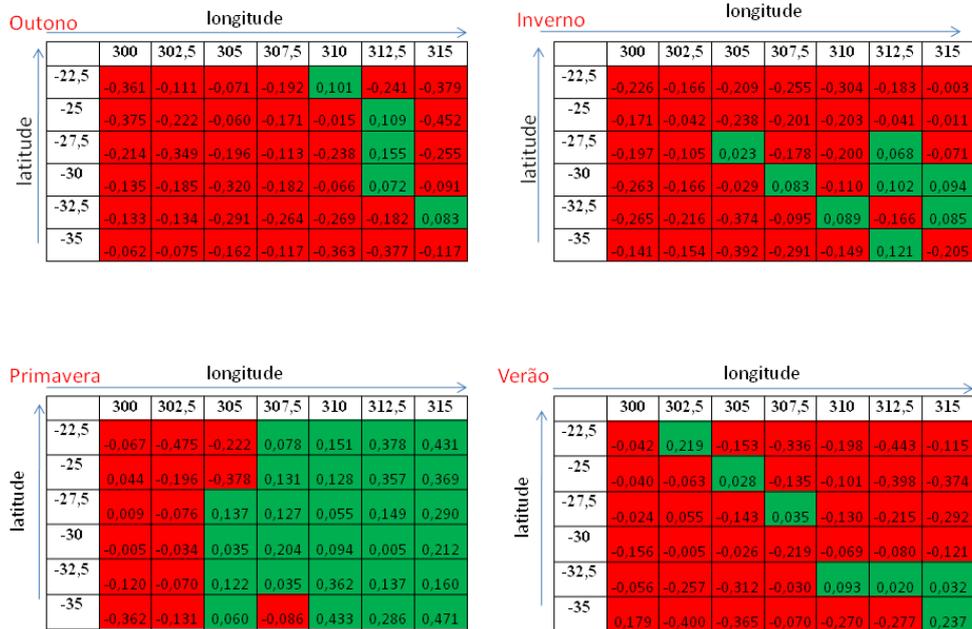


Figura 7.63 - Erro em cada ponto de grade: Rede de Jordan usando todos os dados.

Erro em cada ponto de grade – Rede de Jordan usando a TCA (S)

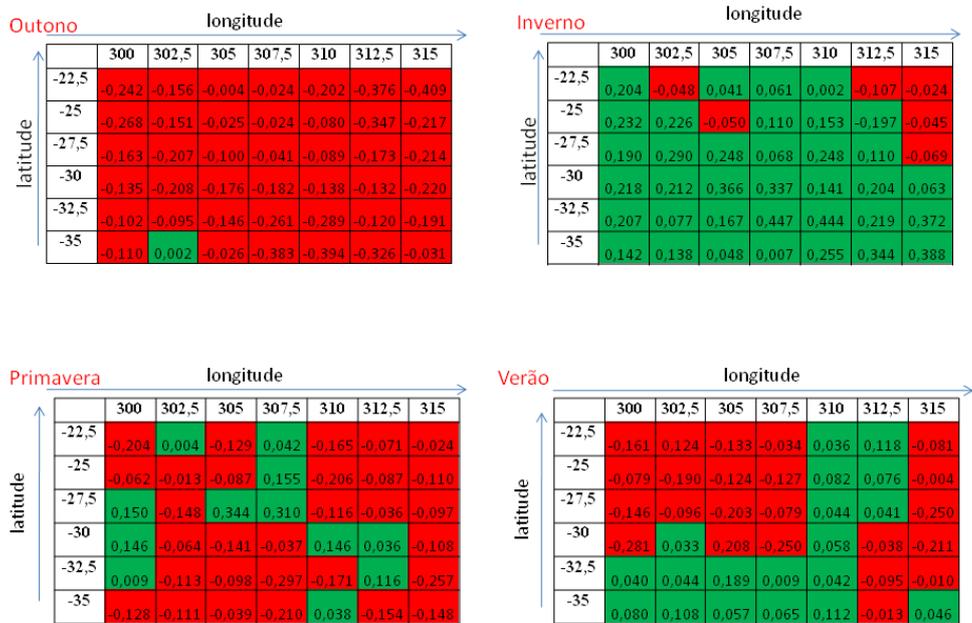


Figura 7.64 - Erro em cada ponto de grade: Rede de Jordan usando os dados reduzidos.

8 CONSIDERAÇÕES FINAIS

Nesta dissertação desenvolveu-se uma metodologia para redução de dados, aplicada no estudo de padrões climáticos sazonais, utilizando as abordagens Teoria dos Conjuntos Aproximativos e Redes Neurais Artificiais, com o propósito de estimar o estado futuro da variável de precipitação, em cinco áreas do Brasil, usando dados provenientes da análise do NCEP (reanálise). Para execução do processo de previsão, foram adotadas duas abordagens: a primeira os modelos de redes são treinados com todas as variáveis disponíveis na base de dados, e na segunda abordagem, os dados são processados pela TCA para selecionar os atributos mais relevantes, com base na hipótese de que alguns atributos exercem maior influência na compreensão do comportamento médio atmosférico.

Para o desenvolvimento do modelo de previsão, foram criados seis modelos de redes neurais para cada uma das regiões de estudo: três dos modelos são construídos a partir do treinamento das redes MLP, Elman e Jordan, usando todas as variáveis disponíveis na base de dados, e para os outros três modelos (MLP, Elman e Jordan), foram usados os dados pré-processados pela TCA, ou seja, os dados reduzidos em número de variáveis.

Os resultados obtidos nesta dissertação mostraram as previsões sazonais, utilizando os modelos de MLP, Elman e Jordan, cujos desempenhos foram avaliados em relação às observações de precipitação nas 5 regiões do território do Brasil, disponíveis na base de dados. Todos os modelos foram validados com a submissão de um conjunto de dados não utilizado durante o treinamento das redes neurais.

Com base nos experimentos iniciais realizados nas regiões Nordeste e Sul, descritos respectivamente nas Seções 7.1, 7.2 e 7.3, observaram-se que as estimativas produzidas utilizando o modelo de rede neural usando os dados pré-processados com a TCA, produziram estimativas mais semelhantes ao observado (real).

As previsões geradas pelos modelos descritos na Seção 7.4.1, mostraram que as previsões realizadas usando a rede recorrente de Elman, utilizando os dados reduzidos por meio da TCA, apresentaram padrões semelhantes àqueles presentes na situação real.

Entretanto, para os resultados de estimativas obtidas nas regiões Norte e Centro Oeste (descritos nas Seções 7.4.2 e 7.4.3), os modelos de Elman apresentaram padrões

mais semelhantes àqueles presentes na situação real, tanto para os modelos usando os dados com todos os atributos, quanto para o modelo usando os dados processados pela TCA.

Os resultados dos experimentos sobre a região Sudeste (descritos na Seção 7.4.4), referentes ao comportamento sazonal da variável de precipitação, não apresentaram padrões muito similares àqueles presentes na situação Real. Entretanto nas estações primavera, inverno e verão, as previsões produzidas pelos modelos de Jordan usando os dados reduzidos pela TCA, apresentam padrões compatíveis com aqueles presentes nas situações reais.

Para as previsões obtidas na região Sul (descrita na Seção 7.4.5) os modelos de Elman usando os dados processados pela TCA, mostraram maior semelhança àqueles presentes na situação real, apenas para as estações outono e inverno. E nas estimativas para a estação verão, o modelo MLP usando os dados reduzidos por meio do uso da TCA, apresentou padrões mais compatíveis com aqueles presentes na situação real.

Com os experimentos realizados, observou-se que as etapas de mineração de dados, usando a abordagem TCA, apresentaram-se adequadas para redução da dimensão dos dados. Foi possível utilizar um número reduzido de variáveis, considerados pela TCA como mais significativos para a atividade de previsão da precipitação, na sintonia dos modelos de redes neurais.

As previsões sazonais geradas pelos modelos mostraram subestimativas na estação primavera, enquanto que na estação verão, as redes neurais superestimaram. Entretanto os experimentos para as estações outono e inverno produziram previsões muito próximas da situação real. Particularmente sobre a região Sul e Sudeste, embora os modelos produzissem estimativas com padrões compatíveis com aqueles presentes nas situações reais, as previsões mostraram uma predominância de superestimação da variável precipitação. Neste caso, a realização de previsão seguindo critérios não geopolíticos, ou seja, voltado para regiões climáticas, poderia ser aplicado, com propósito de aprimorar os resultados de previsão.

De forma geral, a realização dos experimentos cujos resultados foram apresentados nesta dissertação, apresentam a adequação da metodologia para derivar modelos de previsão climatológica utilizando redes neurais, conforme já havia sido mostrado

por Pessoa (2004). Entretanto, neste trabalho foram apresentados novos resultados, utilizando dados de reanálise de um novo período, diferente daquele em (PESSOA, 2004), e novos modelos de redes neurais, as redes recorrentes de Jordan e Elman, foram utilizadas para conceber novos modelos.

Com os novos dados, utilizando a mesma metodologia de redução de variáveis, observou-se que os modelos recorrentes, no geral, apresentaram resultados com padrões de gradientes mais similares às situações reais observadas nos dados disponíveis.

A principal vantagem no uso da metodologia proposta para derivação de modelos de previsão, usando redes neurais, está na possibilidade que a técnica de mineração de dados oferece de reduzir a complexidade do problema e no fato de poderem-se usar métodos de previsão a partir de modelos derivados diretamente dos dados, sem necessidade de considerações sobre a física do problema, permitindo a captação do conhecimento da aplicação.

Por outro, a simplicidade de computação dos modelos de redes neurais, oferece a possibilidade de sintonia local do modelo, em função da disponibilidade de dados, garantindo que qualquer centro de previsão consiga sintonizar um modelo próprio.

Principais contribuições

As seguintes contribuições podem ser pontuadas:

- a) Metodologia para tratamento de dados, em uma abordagem de mineração de dados, utilizando a teoria dos conjuntos aproximativos, permitindo a simplificação do problema.
- b) Metodologia de derivação de modelos baseados em redes neurais para a previsão de precipitação a partir de dados de reanálise com complexidade do problema reduzido pela técnica de mineração com TCA.
- c) Utilização de modelos recorrentes para a previsão climatológica.
- d) Os experimentos realizados na região Sul do Brasil, possibilitou a incorporação da metodologia proposta nesta dissertação no Centro Regional Sul de Pesquisas Espaciais (CRS) de Santa Maria

(<http://www.inpe.br/crs/index.php>), visando auxiliar a previsão de clima na região e realizar operações meteorológicas.

e) Publicações realizadas:

- ANOCHI, J. A.; SILVA, J. D. S.. Uso de Teoria de Conjuntos Aproximativos e Redes Neurais Artificiais no Estudo de Padrões Climáticos Sazonais. *Learning and Nonlinear Models - Revista da Sociedade Brasileira de Redes Neurais*, v.7, No. 2p, pp. 83-91, 2009.
- ANOCHI, J. A.; SILVA, J. D. S.. Mineração de Dados Meteorológicos pela Teoria dos Conjuntos Aproximativos para Aplicação na Previsão de Precipitação Sazonal. In: XXXII Congresso Nacional de Matemática Aplicada e Computacional, 2009, Cuiabá - MT. XXXII CNMAC. São Carlos: SBMAC, pp. 204-210, 2009.
- ANOCHI, J. A.; SILVA, J. D. S.. Uso de Redes Neurais Artificiais e Teoria de Conjuntos Aproximativos no Estudo de Padrões Climáticos Sazonais. In: IX Congresso Brasileiro de Redes Neurais / Inteligência Computacional, 2009, Ouro Preto. IX Congresso Brasileiro de Redes Neurais/Inteligência Computacional. ISSN 2177-1200. 2009.
- ANOCHI, J. A.; SILVA, J. D. S.. Uso de Teoria de Conjuntos Aproximativos e Redes Neurais Artificiais no Estudo de Padrões Climáticos Sazonais. In: IX Workshop do Curso de Computação Aplicada do INPE, São José dos Campos. IX WORCAP, 2009.
- ANOCHI, J. A.; SILVA, J. D. S.; SANTOS, R. D. C.. Estudo de Séries Temporais Meteorológicas com Técnicas de Mineração de Dados e Redes Neurais Artificiais. In: V ERMAC - Encontro Regional de Matemática Aplicada e Computação, 2008, São José dos Campos. V ERMAC, 2008.

Sugestões para trabalhos futuros

Os modelos proposto nesta dissertação mostraram promissores, apresentando previsões similares às situações climatológicas consideradas como observações. No entanto, existem alguns pontos que podem ser explorados para a continuidade deste trabalho de pesquisa, tais como:

- Realizar novos experimentos utilizando bases de dados mais extensas, visando encontrar padrões que sejam desconhecidos para a climatologia;
- Utilizar outros modelos de redes neurais, com aprendizagem não- supervisionada para aprimorar a qualidade dos resultados de previsões;
- Incorporar a metodologia para efetuar previsões reais.
- Realizar testes de robustez dos modelos na geração de previsões com dados faltantes.
- Realizar a previsão climática considerando a divisão de regiões não geopolítica.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGRAWAL, R.; SRIKANT, R. Fast algorithms for mining association rules. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, 20., 1994. **Proceedings...** Santiago, Chile: Morgan Kaufmann, 1994. p. 487–499. 20
- BATISTA, G. E. de A. P. A. **Pré-processamento de dados em aprendizado de máquina supervisionado**. Tese (Doutorado) — Universidade de São Paulo - USP, São Carlos, mar 2003. 16
- BAZAN, J. G.; SKOWRON, A.; SYNAK, P. Dynamic reducts as a tool for extracting laws from decisions tables. **Methodologies for Intelligent Systems**, v. 869, p. 346–355, 1994. 31
- BONATTI, J. **Modelo de circulação geral atmosférico do CPTEC**. Cachoeira Paulista - SP, 1996. Disponível em: <<http://climanalise.cptec.inpe.br/~rcliman1/boletim/cliesp10a/bonatti.html>>. 8
- CARVALHO, A.; BRAGA, A.; LUDERMIR, T. **Redes neurais artificiais teorias e aplicações**. Rio de Janeiro: Rio de Janeiro: LTC, 2000. 262 p. 39, 42, 43, 45, 46, 47, 51, 53, 54, 57, 61, 62, 64, 65
- CHOU, S. C. **Modelo regional ETA**. Cachoeira Paulista - SP, Janeiro 1996. Disponível em: <<http://climanalise.cptec.inpe.br/~rcliman1/boletim>>. 8
- DOTY, B. **Grid Analysis and Display System (GrADS)**. março 2009. Center for Ocean - Land-Atmosphere Studies (COLA). Disponível em: <<http://grads.iges.org/grads/head.html>>. 72
- FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH., P. The kdd process for extracting useful knowledge from volumes of data. **Communications of the ACM**, v. 39, n. 11, p. 27–34, 1996. 13, 14
- HAYKIN, S. **Redes neurais princípios e práticas**. 2°. ed. Porto Alegre: Bookman, 2001. 900 p. 39, 40, 41, 42, 44, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 59, 60, 61, 62, 63
- ØHRN, A. **Discernibility and rough sets in medicine: tools and applications**. Tese (Doutorado) — Norwegian University of Science and

Technology, Department of Computer and Information Science, NTNU, Noruega, 1999. 24, 30, 32, 36, 72

ØHRN, A. **ROSETTA technical reference manual**. Warsaw, Novembro 1999. Disponível em: <<http://www.lcb.uu.se>>. 33, 34, 35, 36, 37

ØHRN, A.; KOMOROWSKI, J. Rosetta - a rough set toolkit for analysis of data. In: INTERNATIONAL JOINT CONFERENCE ON INFORMATION SCIENCES, 3., 1997. **Proceedings...** Durham, USA: Institute of Computer Science, 1997. p. 403–407. 32

INSTITUTO NACIONAL DE METEOROLOGIA (INMET). **Previsão de Consenso**. agosto 2010. Previsão Climática. Disponível em: <<http://www.inmet.gov.br/>>. Acesso em: 27 set. 2010. 70

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. CENTRO DE PREVISÃO DE TEMPO E ESTUDOS CLIMÁTICOS (INPE.CPTEC). **Portal de previsões numéricas**. 2009. Disponível em: <www.cptec.inpe.br>. Acesso em: 27 feb. 2009. 9, 11

KOMOROWSKI, J.; ØHRN, A. Modelling prognostic power of cardiac tests using rough sets. **Artificial Intelligence in Medicine**, v. 15, p. 167–191, 1999. 24, 28, 30, 31

KOMOROWSKI, J.; PAWLAK, Z.; POLKOWSKI, L.; SKOWRON, A. **Rough Set: a tutorial**. Polônia, 1998. 17 p. 28

KOMOROWSKI, J.; POLKOWSKI, L.; SKOWRON, A. Towards a rough mereology-based logic for approximate solution synthesis. **Studia Logica**, v. 58, p. 17–95, 1995. 29

MENDONÇA, A.; BONATTI, J. O sistema de previsão de tempo global por ensemble do cptec. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 12., 2002. **Anais...** Foz de Iguaçu, Paraná: CBMET Congresso Brasileiro de Meteorologia, 2002. p. 3341–3351. 9

MITRA, S.; PAL, S. K.; MITRA, P. Data mining in soft computing framework: A survey. **IEEE Transactions on Neural Networks**, v. 13, n. 1, p. 12, 2002. 17

NACIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (NOAA). **Reanalysis Datasets at PSD**. 2010. Reanalysis Datasets. Disponível em: <<http://www.esrl.noaa.gov/psd/>>. Acesso em: 17 set. 2010. 68

OLIVEIRA, R. B. T. de. **O processo de extração de conhecimento de base de dados apoiado por agentes de software**. Dissertação (Mestrado) — Universidade de São Carlos - USP, São Carlos, 2000. 15

OLSON, L.; DURSUN, D. **Advanced data mining techniques**. 1°. ed. Lincoln: Springer, 2008. 182 p. 18, 19, 20

PAWLAK, Z. Rough sets. **International Journal of Computer and Information Sciences**, v. 11, p. 341–356, 1982. 23

_____. Rough sets present state and further prospects. In: INTERNATIONAL WORKSHOP ON ROUGH SET, 3., 1994. **Proceedings...** Poland: Institute of Computer Science, 1994. p. 72–76. 26

PESSOA, A. S. A. **Mineração de dados meteorológicos pela teoria dos conjuntos aproximativos na previsão de clima por redes neurais artificiais**. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Dezembro 2004. 2, 3, 147

REZENDE, S. O. **Sistemas inteligentes fundamentos e aplicações**. 1°. ed. Barueri, SP: Manole, 2003. 550 p. 16, 19, 20, 21

SKOWRON, A.; POLKOWSKI, L. Rough mereological foundations for design, analysis, synthesis and control in distributed systems. **Information Sciences**, v. 104, p. 129–156, 1998. 24

STEPANIUK, J. **Rough granular computing in knowledge discovery and data mining**. 1°. ed. Warsaw, Poland: Springer, 2008. 157 p. 26

UCHÔA, J. **Representação e indução de conhecimento usando teoria de conjuntos aproximados**. Dissertação (Mestrado) — Universidade de São Carlos - USP, São Carlos, 1998. 23

VIANELLO, R. L.; ALVES, A. **Meteorologia básica e aplicações**. 4°. ed. Viçosa: UFV, 2006. 449 p. 5, 6, 8, 10

APÊNDICE A - ANÁLISE DE ROBUSTEZ

Nesta dissertação foi realizada uma análise de robustez, com propósito de avaliar o comportamento das variáveis e suas influências no processo de previsão climática e ainda verificar a adequação dos novos dados (janeiro de 2000 - dezembro de 2009), devido a indisponibilidade da variável divergência na base de dados do CDC/NOAA, como pode ser observado na Figura A.1.

Dados Janeiro de 2000 – Dezembro de 2009			Dados Janeiro de 1980 – Dezembro de 2000		
	Variável	Descrição		Variável	Descrição
1	temp	Temperatura	1	div	Divergência
2	shum	Umidade Específica	2	prec	Precipitação
3	u300	Vento Zonal 300 hPa	4	shum	Umidade Específica
4	u500	Vento Zonal 500 hPa	4	spres	Pressão da Superfície
5	u850	Vento Zonal 850 hPa	5	temp	Temperatura
6	v300	Vento Meridional 300 hPa	6	u300	Vento Zonal 300 hPa
7	v500	Vento Meridional 500 hPa	7	u500	Vento Zonal 500 hPa
8	v850	Vento Meridional 850 hPa	8	u850	Vento Zonal 850 hPa
9	pres	Pressão superfície	9	v300	Vento Meridional 300 hPa
10	prec	Precipitação	10	v500	Vento Meridional 500 hPa
			11	v850	Vento Meridional 850 hPa

Figura A.1 - Variáveis Meteorológicas.

Desta maneira, foram realizados experimentos, simulando a ausência de variáveis e foram acrescentados ruídos gaussiano de 1%, 2% e 5%, com propósito de analisar o desempenho de cada variável em relação aos dados de observação (precipitação real). Deste modo a solução que estiver mais próxima da resposta obtida pela rede neural com todas as variáveis, é uma solução robusta.

A.1 Simulando a ausência de variáveis

A Tabela A.1 apresenta os erros quadráticos médios obtidos através dos experimentos simulando a ausência de variáveis na região Nordeste, para as quatro estações dos anos de 2007 e 2008. As coordenadas geográficas que definem a região estão compreendidas entre as longitudes [47°W, 35°W] e entre as latitudes [0°S, -17.5°S]. Observa-se que os maiores erros encontram-se na ausência das variáveis: vento zonal (u500), vento zonal (u300), divergência (div) e pressão do nível do mar (pres).

Tabela A.1 - Erro quadrático médio.

Erro Quadrático Médio para rede faltando variáveis - Nordeste				
Variáveis	Outono	Inverno	Primavera	Verão
div	1,115	0,7429	0,9612	0,8202
u300	1,6171	0,6299	0,827	1,242
u500	1,5678	2,0298	1,3787	1,477
u850	0,1912	1,252	0,4434	0,3303
v300	0,6551	0,3839	0,6655	1,1316
v500	0,3783	0,2738	0,384	0,446
v850	0,2788	0,2689	0,3419	0,4155
pres	1,9422	0,5444	0,8003	1,4571
shum	0,1343	0,2219	1,2287	0,2273
prec	0,7085	0,3217	0,4949	0,954
airt	0,1611	0,1863	0,2488	0,2059
Todas variáveis	0,5819	0,6363	1,828	0,8217

O estudo da execução do modelo de previsão, simulando a ausência de variáveis, mostrou que o modelo treinado originalmente com todas as variáveis, é sensível a ausência individual de cada variável. Observa-se nos resultados que as redes produziram padrões de gradientes diferentes daqueles presentes na situação considerada Real (observado).

As Figuras A.2 a A.5 apresentam os resultados dos experimentos faltando variáveis, com propósito de verificar a variação e a representatividade das variáveis durante o treinamento das redes neurais.

A Figura A.3 apresenta os resultados faltando variáveis na estação inverno de 1998.

A Figura A.4 mostra o resultado obtido para a estação primavera de 1998. Neste caso, observa-se que todas as variáveis apresentam um alto índice de representatividade em relação ao estado de precipitação (real).

A Figura A.5 apresenta o resultado obtido para a estação verão de 1998.

A Tabela A.2 apresenta o desempenho dos experimentos realizados, em termos de erro quadrático médio, para o modelo de rede neural treinada com as variáveis indicadas pela TCA.

O estudo da execução do modelo de previsão, simulando a ausência de variáveis,

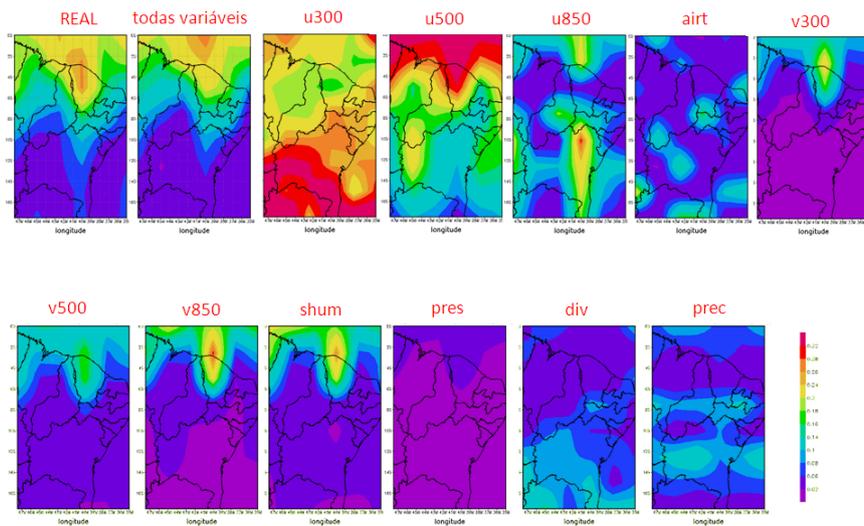


Figura A.2 - Experimento faltando variáveis no outono de 1998.

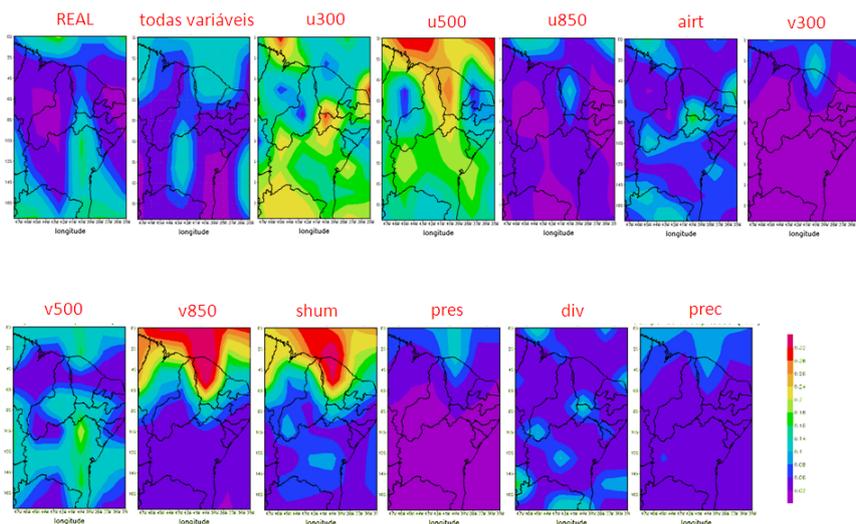


Figura A.3 - Experimento faltando variáveis no inverno de 1998.

mostrou que o modelo treinado com as variáveis selecionadas pela TCA é sensível a ausência individual de cada variável.

Observa-se nas Figuras A.6 a A.9 os resultados dos experimentos faltando variáveis, utilizando as variáveis que foram indicadas pela TCA. Os resultados que as redes produziram padrões de gradientes diferentes daqueles presentes na situação considerada Real (observado).

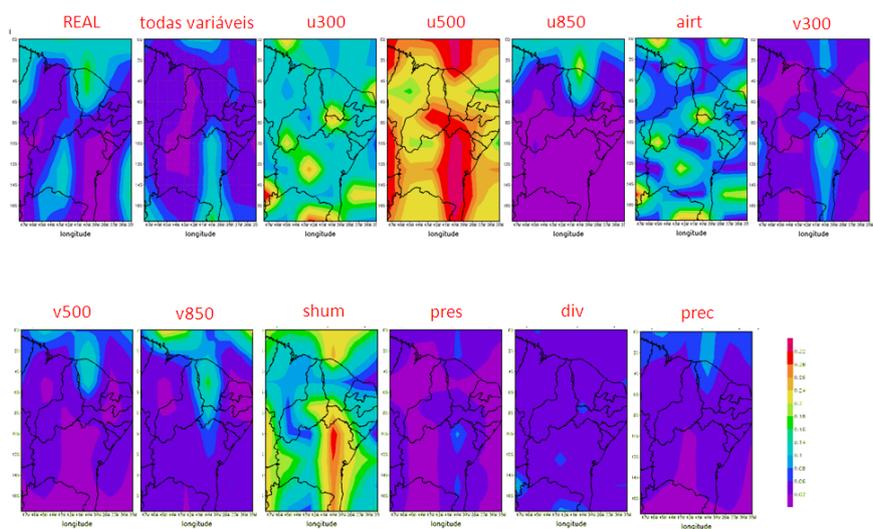


Figura A.4 - Experimento faltando variáveis na primavera de 1998.

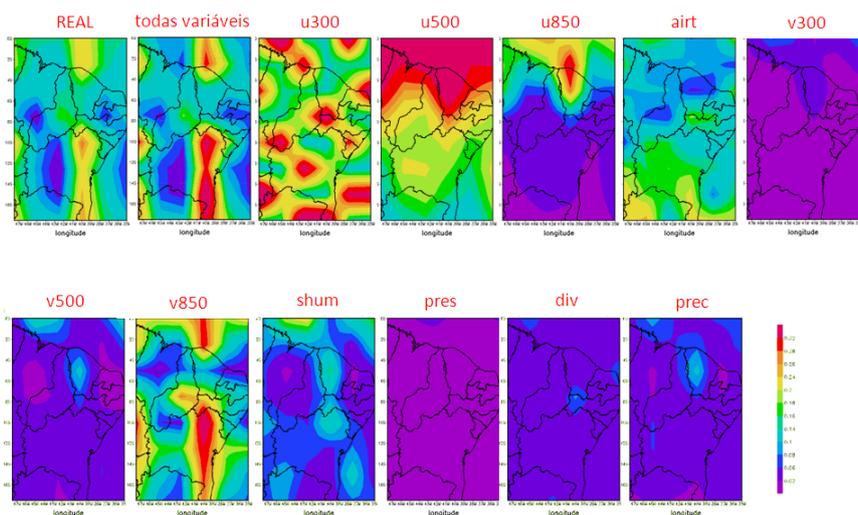


Figura A.5 - Experimento faltando variáveis no verão de 1998.

A Figura A.7 mostra o resultado para a estação inverno de 1998. Observa-se que todas as variáveis indicadas pela TCA na estação inverno de 1998 são indispensáveis.

A figura A.8 apresenta o resultado para a estação primavera de 1998. Neste caso, observa-se que a variável vento meridional (v300) possui uma pequena influência ao estado de precipitação, desta forma, poderia ser dispensada pelo fato de não interferir no resultado quando retiradas do treinamento.

Tabela A.2 - Erro quadrático médio para a rede neural usando a TCA.

Erro Quadrático Médio para rede faltando variáveis - Nordeste				
Variáveis	Outono	Inverno	Primavera	Verão
Airt	1,1007	0,5121	0,6109	1,4479
U850	0,9342	0,5358	0,5989	0,9111
U500	1,7983	0,7548	0,9264	2,2313
V500	1,7051	0,7412	0,9027	2,0906
V500	0,4261	0,2730	0,3392	0,5715
Todas variáveis TCA	0,5423	1,0670	0,7601	0,6722

A Figura A.9 mostra o resultado na estação verão de 1998. Observa-se que todas as variáveis são essenciais para obter um resultado semelhante ao estado de precipitação real.

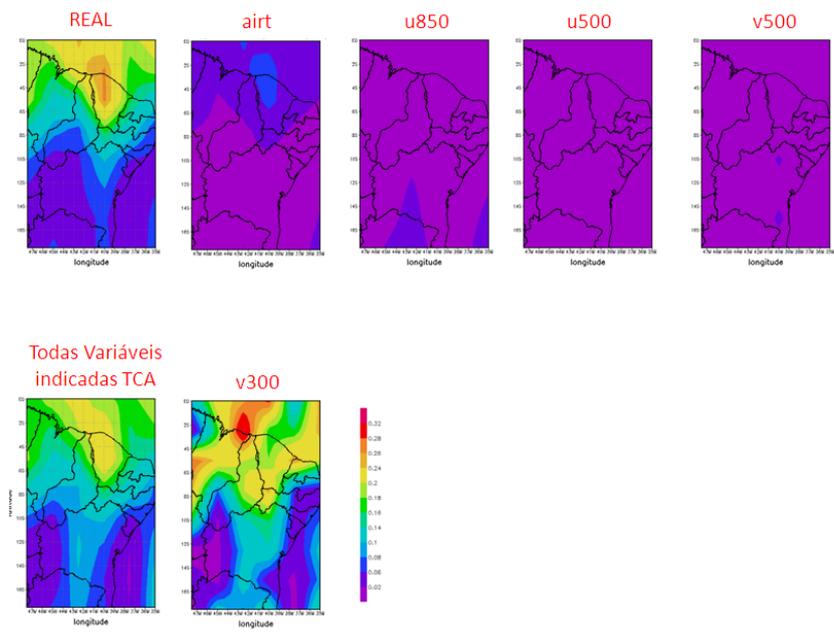


Figura A.6 - Experimento faltando variáveis indicadas pela TCA, no outono de 1998.

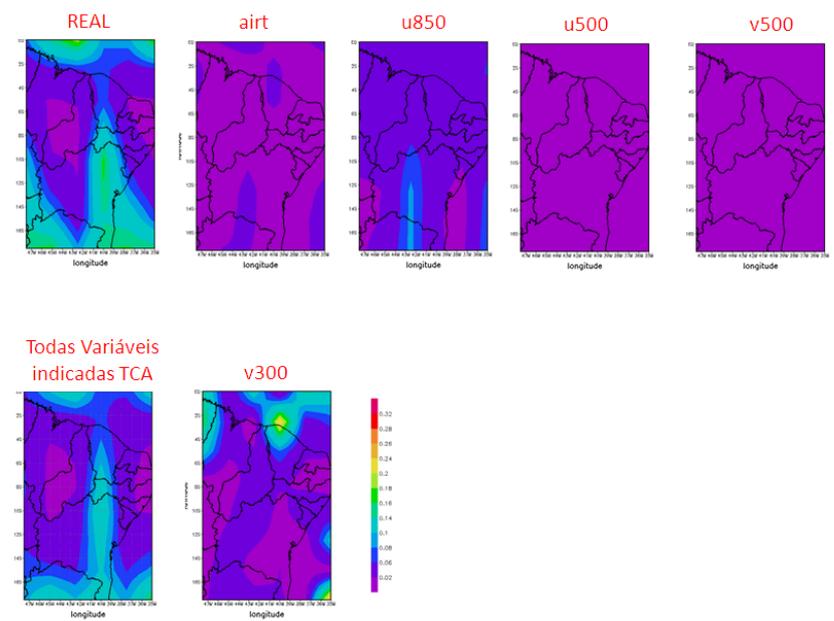


Figura A.7 - Experimento faltando variáveis indicadas pela TCA, no inverno de 1998.

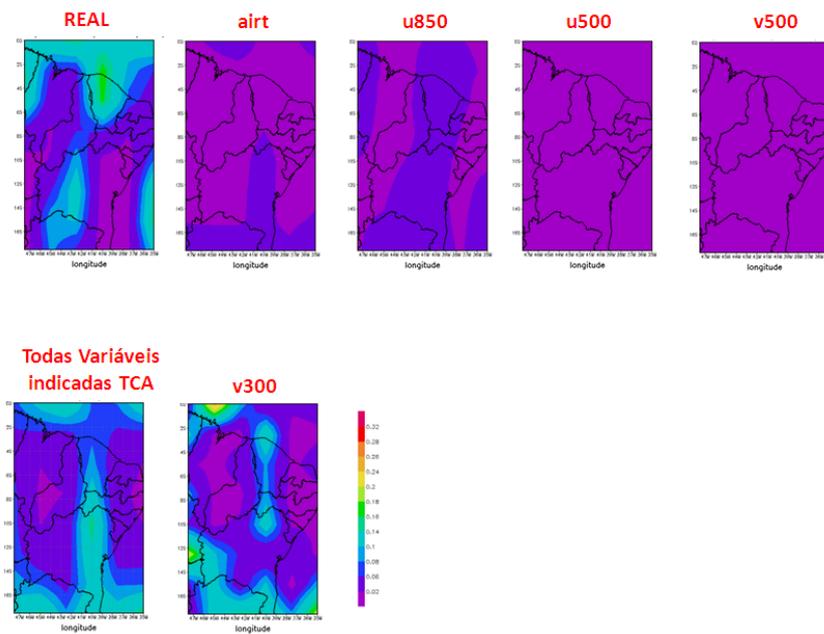


Figura A.8 - Experimento faltando variáveis indicadas pela TCA, na primavera de 1998.

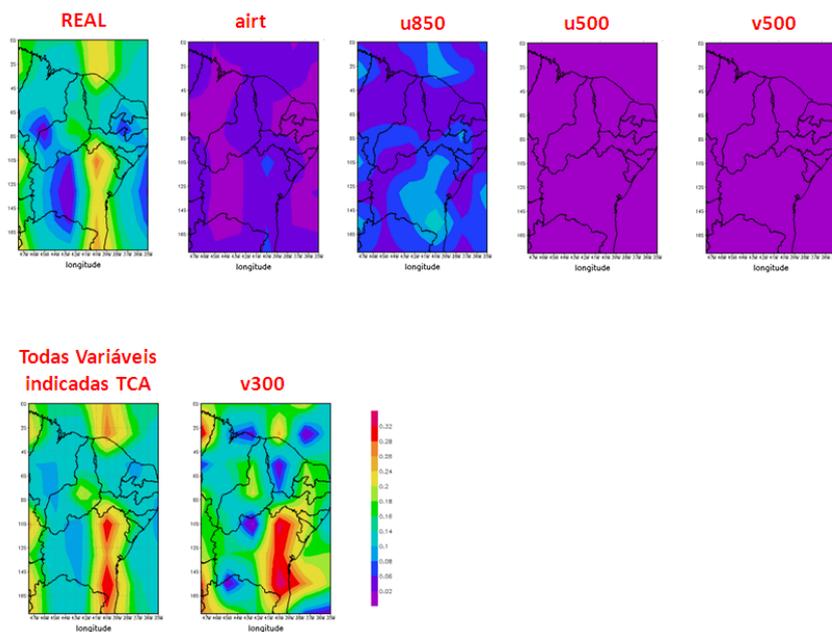


Figura A.9 - Experimento faltando variáveis indicadas pela TCA, no verão de 1998.

Os experimentos simulando ausência de variáveis mostram que os modelos são sensíveis a ausência de informação. Na simulação de ausência de variáveis para as redes treinadas com todas as variáveis, observou-se a sensibilidade do modelo a variáveis que não foram selecionadas pelo processo da aplicação da TCA, indicando que outras variáveis, além daquelas indicadas pela TCA também são importantes para o processo de previsão.

A explicação para este fato está relacionada com o processo de treinamento das redes neurais. Ou seja, as redes treinadas com todas as variáveis, aprenderam a realizar estimativas com base em todas as variáveis, estabelecendo associações entre todas as variáveis e a variável estimada.

Por outro lado, as redes treinadas com as variáveis selecionadas pela TCA, só estabeleceram associações da variável estimada com as variáveis escolhidas. Portanto, o estudo mostra que os modelos são sensíveis às variáveis que foram utilizadas para treiná-los.

A.2 Análise com ruídos

Neste trabalho, foi realizado um estudo nos dados, em que foram acrescentados ruídos gaussiano de 1%, 2% e 5%.

As Figuras A.10 a A.12 mostram os resultados das análises acrescentando ruídos gaussiano de 1%, 2% e 5%. Observa-se nos resultados que as redes produziram padrões de gradientes diferentes daqueles presentes na situação considerada Real (observado), mostrando que houve uma superestimativa nos valores máximos presentes na situação Real (observado) e tendem a subestimar os valores mínimos.

A Tabela A.3 mostra os erros quadráticos médios obtidos. Observa-se que os maiores índices de erros encontram-se nas estações primavera e verão.

Tabela A.3 - Erro quadrático médio - 1% de ruído.

1% de ruído				
Variáveis	Outono	Inverno	Primavera	Verão
airt	0,03	4,71	2,39	4,82
u850	0,00	4,72	2,25	4,36
u500	0,05	1,49	3,34	9,61
u300	0,00	0,16	9,25	7,76
v850	1,04	0,02	19,35	46,96
v500	0,13	0,68	26,79	60,36
v300	0,59	6,63	31,81	3,62
shum	0,02	6,66	26,54	3,47
pres	0,04	6,61	28,50	4,24
div	1,92	0,24	34,87	0,90
prec	1,58	0,11	36,43	1,27
Todas variáveis	0,00	25,17	1,06	6,87

A Figura A.10 apresenta o resultado obtido com o experimento acrescentando 1% de ruído nos dados.

A Tabela A.4 mostra os erros quadráticos médios obtidos, nos experimentos em que foram acrescentados 2% de ruído nos dados. Observa-se que os maiores índices de erros encontram-se nas estações inverno e verão, em particular, quando acrescentado ruído nas variáveis vento zonal (u300), vento meridional (v850), vento meridional (v500), vento meridional (v300) e umidade específica (shum).

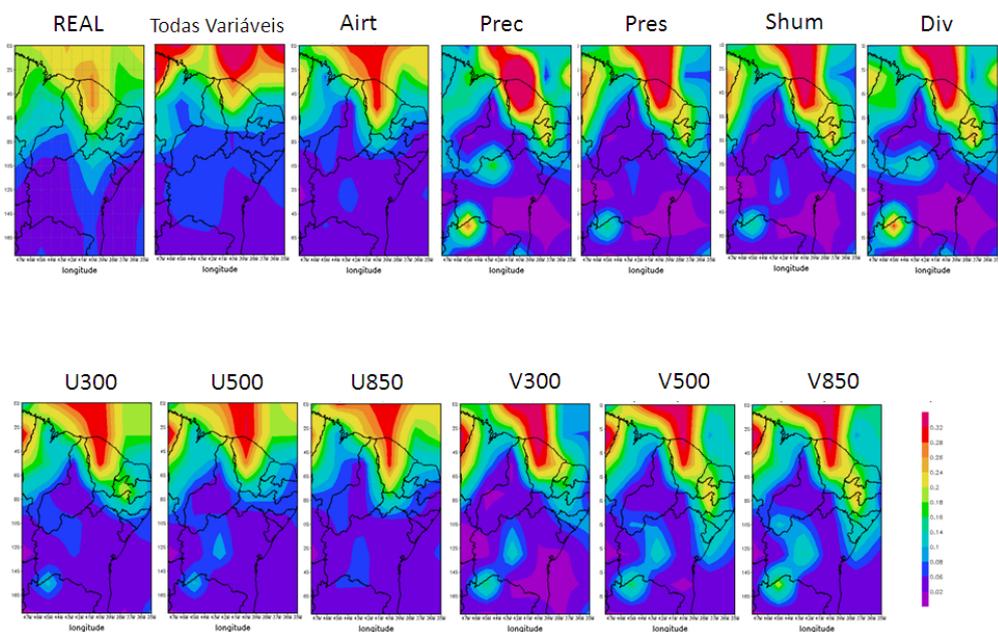


Figura A.10 - Experimentos acrescentando 1% de ruído.

Tabela A.4 - Erro quadrático médio - 2% de ruído.

2% de ruído				
Variáveis	Outono	Inverno	Primavera	Verão
airt	1,78	21,98	26,48	5,38
u850	1,42	21,93	25,23	5,43
u500	3,85	135,33	24,75	61,93
u300	3,59	205,84	24,56	95,91
v850	1,32	485,16	24,97	599,36
v500	6,11	404,42	24,99	827,79
v300	5,47	554,92	5,55	616,61
shum	51,01	1260,5	3,22	2988
pres	213,2	4641,9	0,54	2841
div	16,92	3974,9	160,34	2601,6
prec	22,97	3748,2	153,14	3045,5
Todas variáveis	0,04	18,04	0,27	1,08

A Figura A.11 apresenta o resultado dos experimentos, em que foram acrescentando 2% de ruído nos dados.

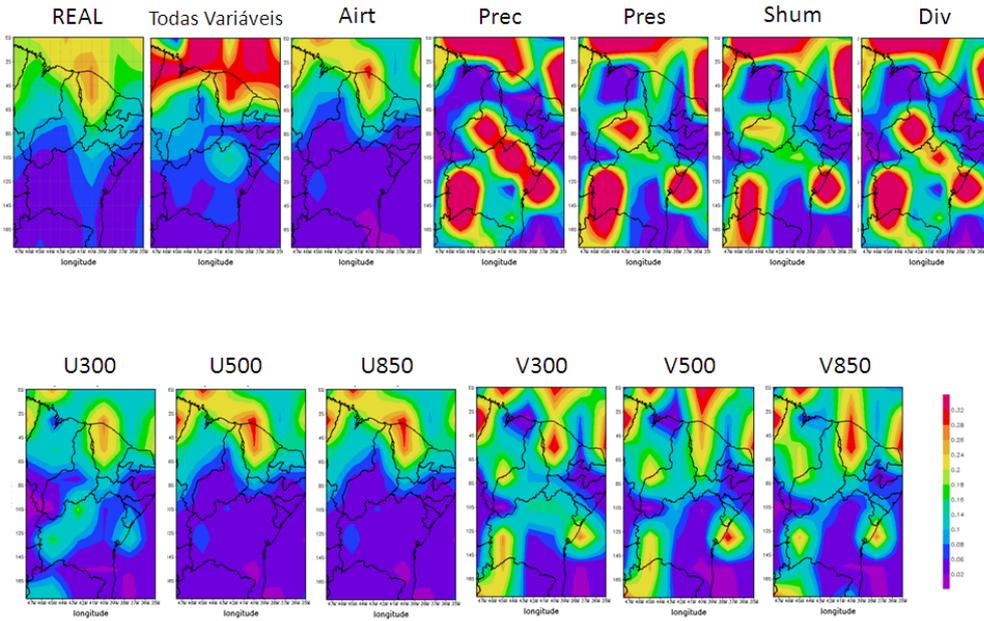


Figura A.11 - Experimentos acrescentando 2% de ruído.

A Tabela A.5 mostra os erros quadráticos médios obtidos com os experimentos acrescentando 5% de ruído nos dados. Neste caso, observa-se que os maiores índices de erros encontram-se nas estações inverno e verão, em particular, quando acrescentado ruído nas variáveis vento zonal (u300), vento meridional (v300), vento meridional (v500) e umidade específica (shum).

A Figura A.12 apresenta o resultado das análises acrescentando 5% de ruído nos dados. Observa-se que as redes não produziram padrões similares àqueles presentes na situação real (observação).

Tabela A.5 - Erro quadrático médio - 5% de ruído.

5% de ruído				
Variáveis	Outono	Inverno	Primavera	Verão
airt	0,3	0,25	0,88	0,53
u850	0,93	0,21	3,08	2,66
u500	10,75	1,88	8,74	0,53
u300	6,34	43,69	31,4	23,81
v850	7,94	0,01	18,71	9,15
v500	15,28	6,7	4,95	44,16
v300	16,17	102,25	9,59	6,36
shum	13,61	84,84	7,62	5,9
pres	12,84	0,03	32,17	128,47
div	5,76	7	32,14	0,14
prec	8,45	6,96	32,16	0,47
Todas variáveis	1,39	7,44	17,71	6,17

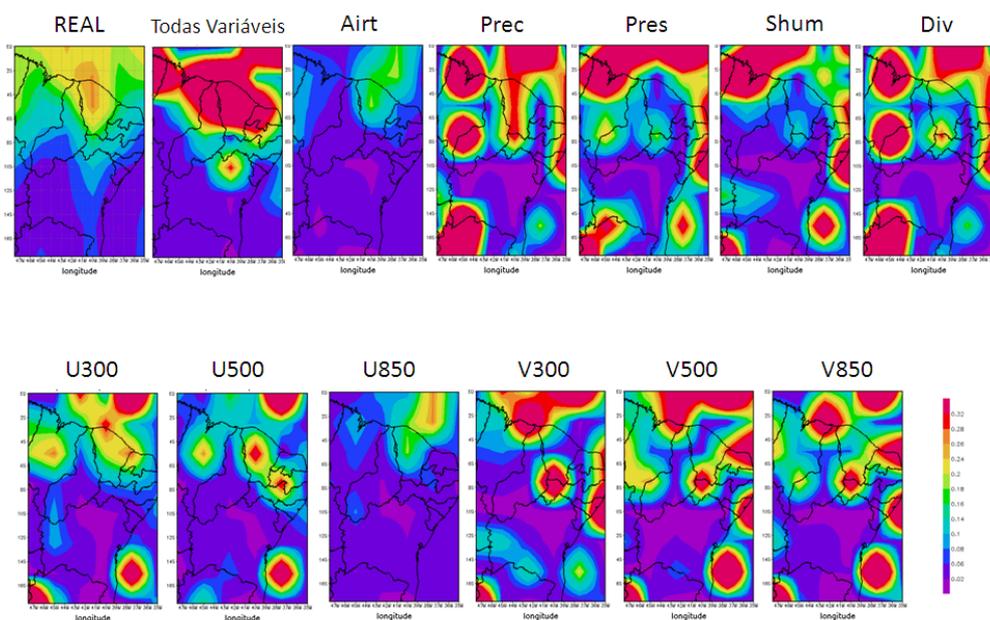


Figura A.12 - Experimentos acrescentando 5% de ruído.

APÊNDICE B - MAPAS DE ERRO EM CADA PONTO DE GRADE

B.1 Mapas de erros para os modelos de previsão na região Nordeste

Os erros foram calculados em cada ponto de grade da região. As coordenadas geográficas que definem a região Nordeste (NE) estão compreendidas entre as longitudes [47°W, 35°W] e entre as latitudes [0°S, -17.5°N], compreendendo 48 pontos de grade (8 latitudes e 6 longitudes).

As Figuras B.1 a B.3 apresentam os mapas de erros em cada ponto de grade, criados a partir dos modelos de RNAs: Elman, Jordan e MLP, em que utilizam todas as variáveis disponíveis na base de dados.

As Figuras B.4 a B.6 mostram os mapas de erros obtidos em cada ponto de grade, a partir dos modelos de RNAs: Elman, Jordan e MLP, utilizando os dados reduzidos através do uso da TCA.

Erro em cada ponto de grade – Rede de Elman usando todos os dados (NE)

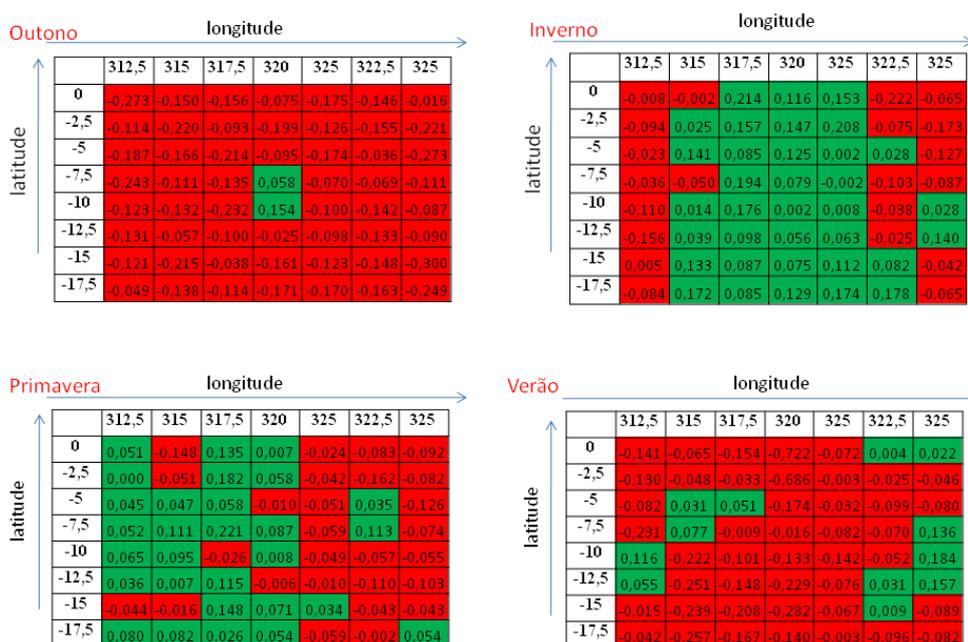


Figura B.1 - Erro em cada ponto de grade: Rede de Elman com todos os dados.

Erro em cada ponto de grade – Rede de Jordan usando todos os dados (NE)

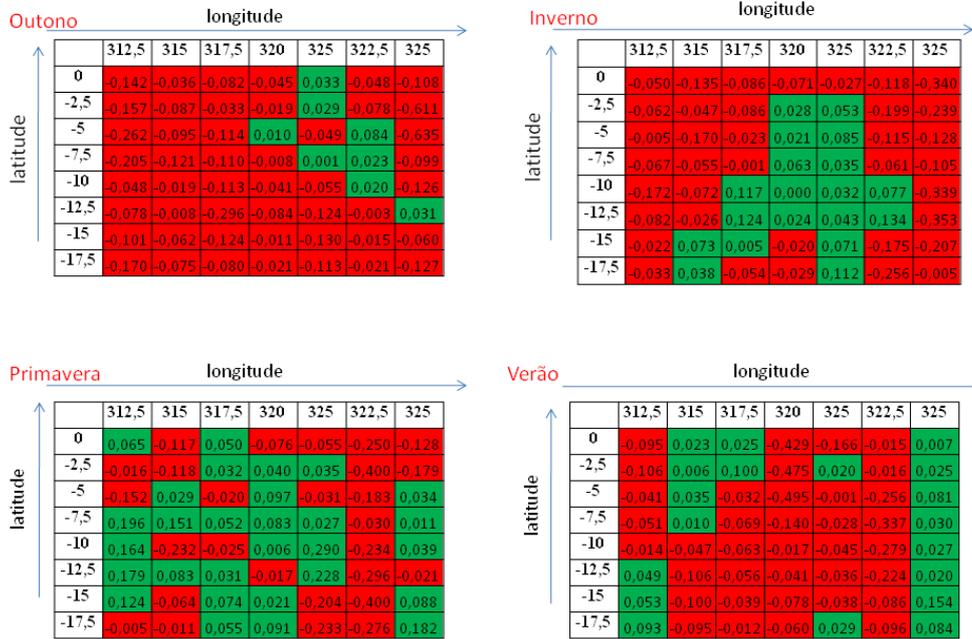


Figura B.2 - Erro em cada ponto de grade: Rede de Jordan com todos os dados.

Erro em cada ponto de grade – Rede MLP usando todos os dados (NE)

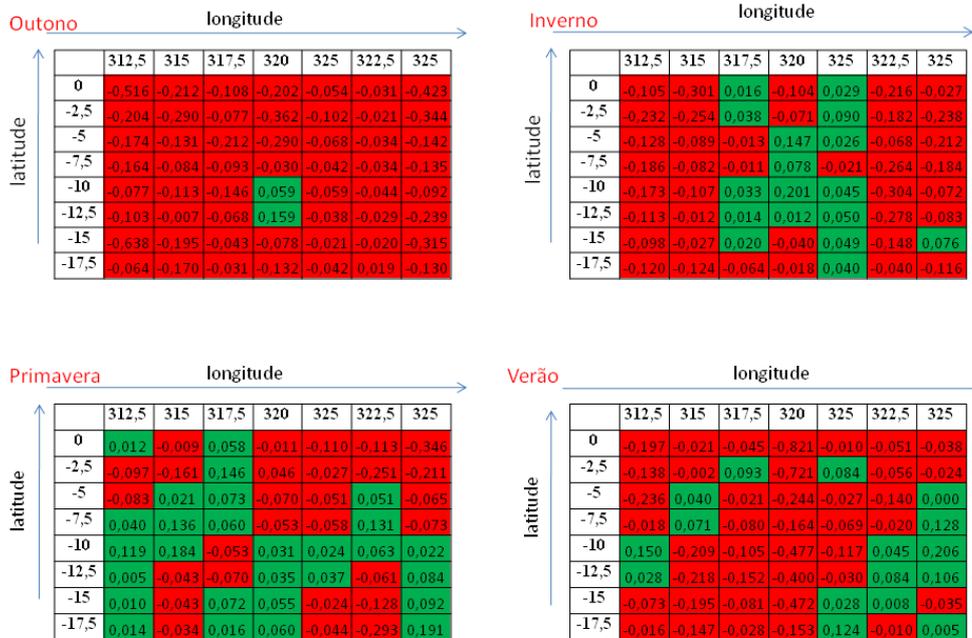


Figura B.3 - Erro em cada ponto de grade: Rede MLP com todos os dados.

Erro em cada ponto de grade – Rede de Elman usando TCA (NE)

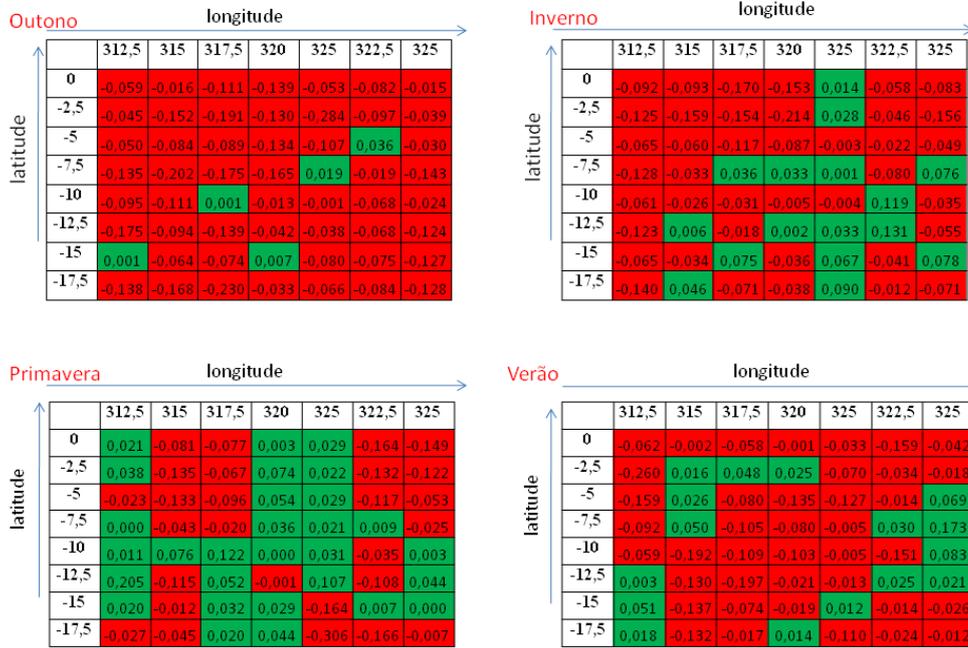


Figura B.4 - Erro em cada ponto de grade: Rede de Elman usando a TCA.

Erro em cada ponto de grade – Rede de Jordan usando TCA (NE)

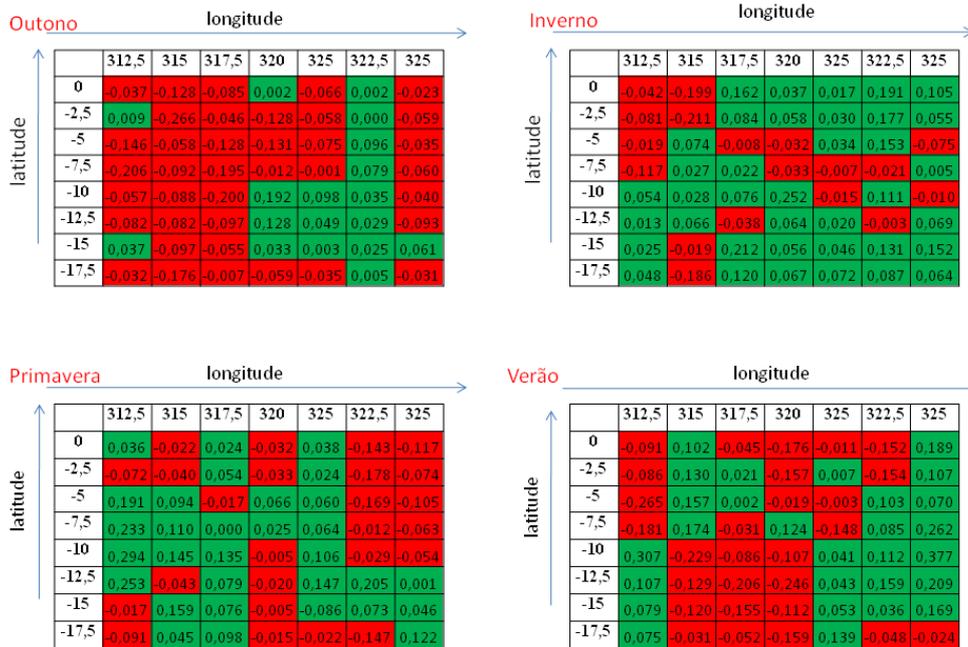


Figura B.5 - Erro em cada ponto de grade: Rede de Jordan usando a TCA.

Erro em cada ponto de grade – Rede MLP usando TCA (NE)

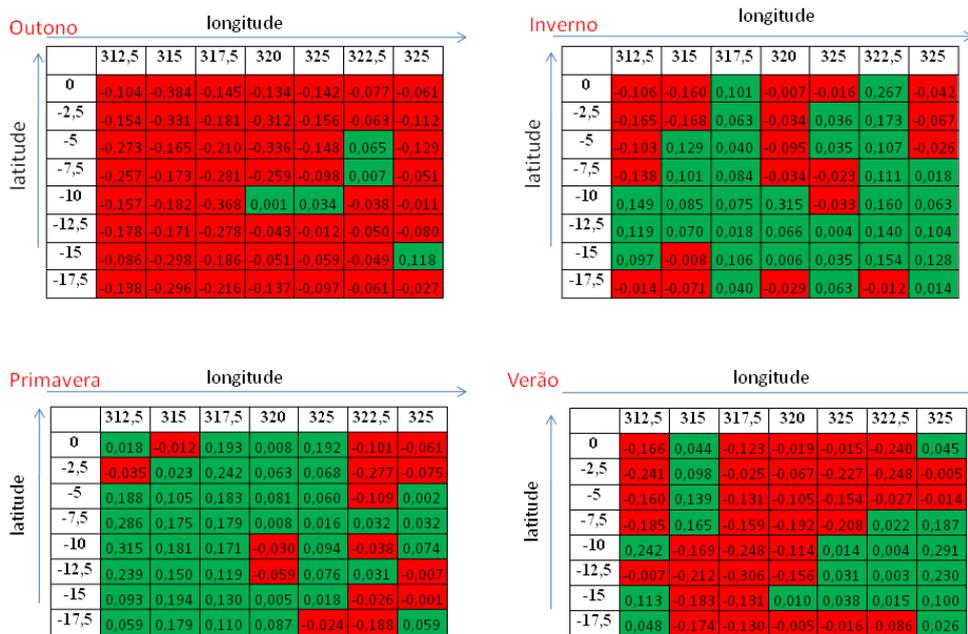


Figura B.6 - Erro em cada ponto de grade: Rede MLP usando a TCA.

B.2 Mapas de erros para os modelos de previsão na região Centro Oeste

Os erros foram calculados em cada ponto de grade da região. As coordenadas geográficas que definem a região Centro Oeste (CO) estão compreendidas entre as longitudes [62°W, 48°W] e entre as latitudes [25°S, 7.5°S], compreendendo 56 pontos de grade (7 latitudes e 8 longitudes).

As Figuras B.7 a B.9 mostram mapas de erros em cada ponto de grade, construídos a partir dos modelos de redes: Elman, Jordan e MLP, em que utilizam todas as variáveis disponíveis na base de dados.

As Figuras B.10 a B.11 apresentam os mapas de erros em cada ponto de grade obtidos a partir dos modelos de redes: Elman, Jordan e MLP, em que utilizam o conjunto de dados reduzidos através do uso da TCA.

Erro em cada ponto de grade – Rede de Elman com todos os dados (CO)

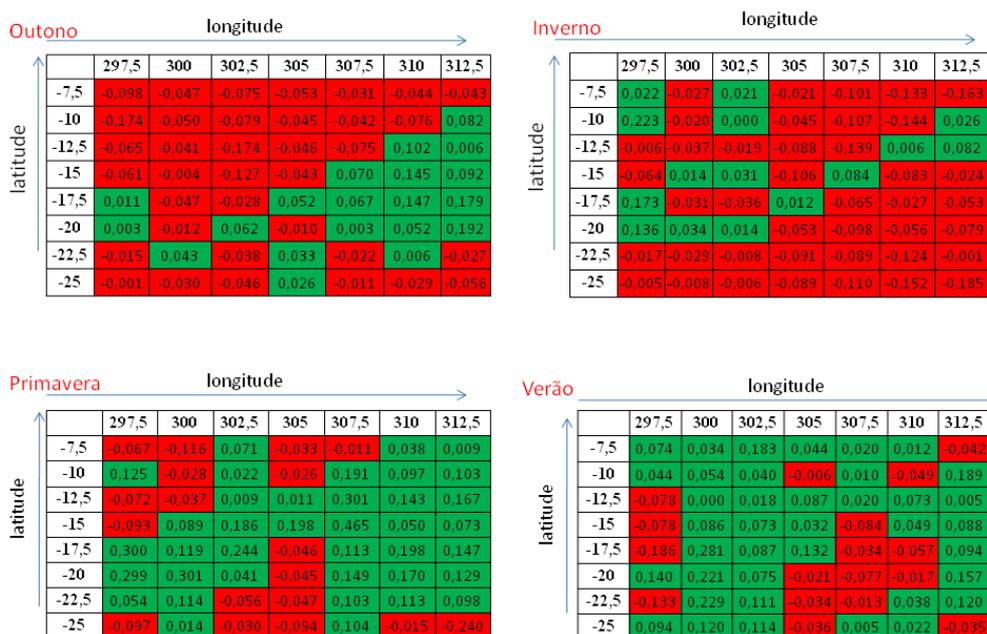


Figura B.7 - Erro em cada ponto de grade: Rede de Elman usando todos os dados (CO).

Erro em cada ponto de grade – Rede de Jordan com todos os dados (CO)

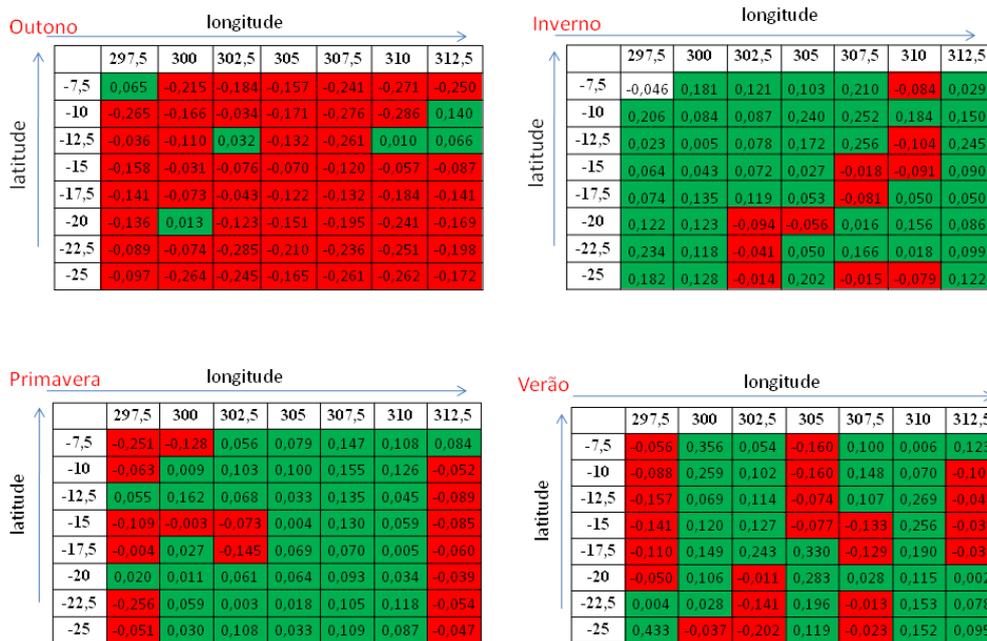


Figura B.8 - Erro em cada ponto de grade: Rede de Jordan usando todos os dados (CO).

Erro em cada ponto de grade – Rede MLP com todos os dados (CO)

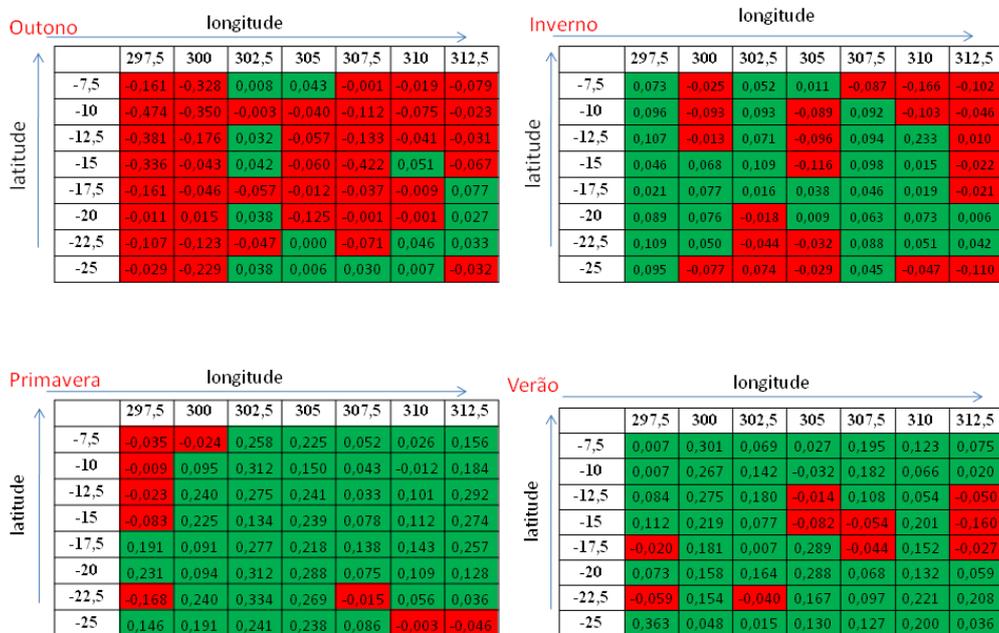


Figura B.9 - Erro em cada ponto de grade: Rede MLP usando todos os dados (CO).

Erro em cada ponto de grade – Rede de Elman usando TCA (CO)

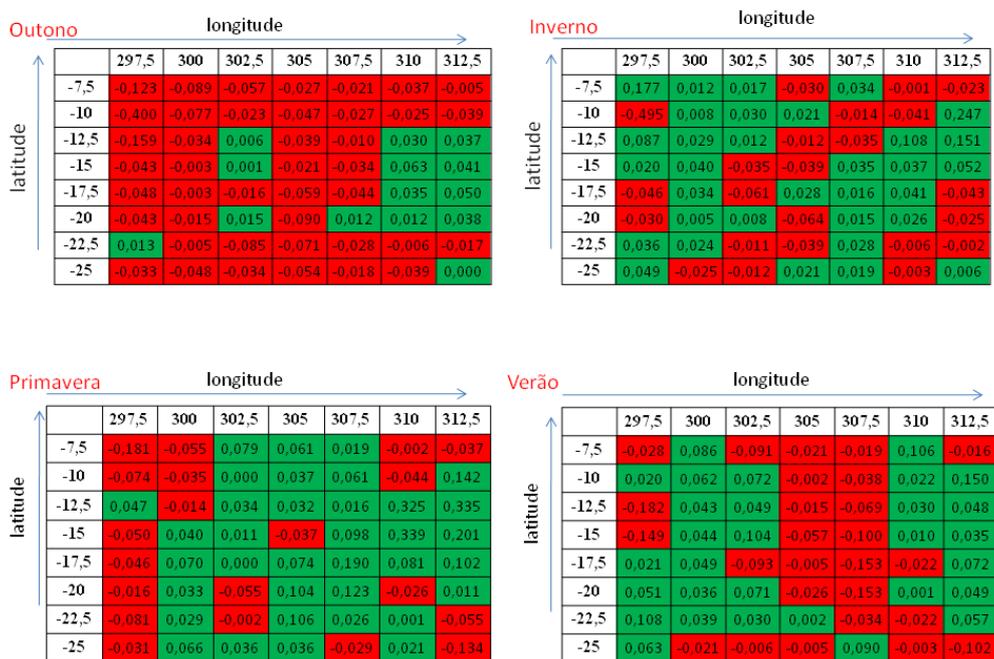


Figura B.10 - Erro em cada ponto de grade: Rede de Elman usando a TCA(CO).

Erro em cada ponto de grade – Rede de Jordan usando TCA (CO)

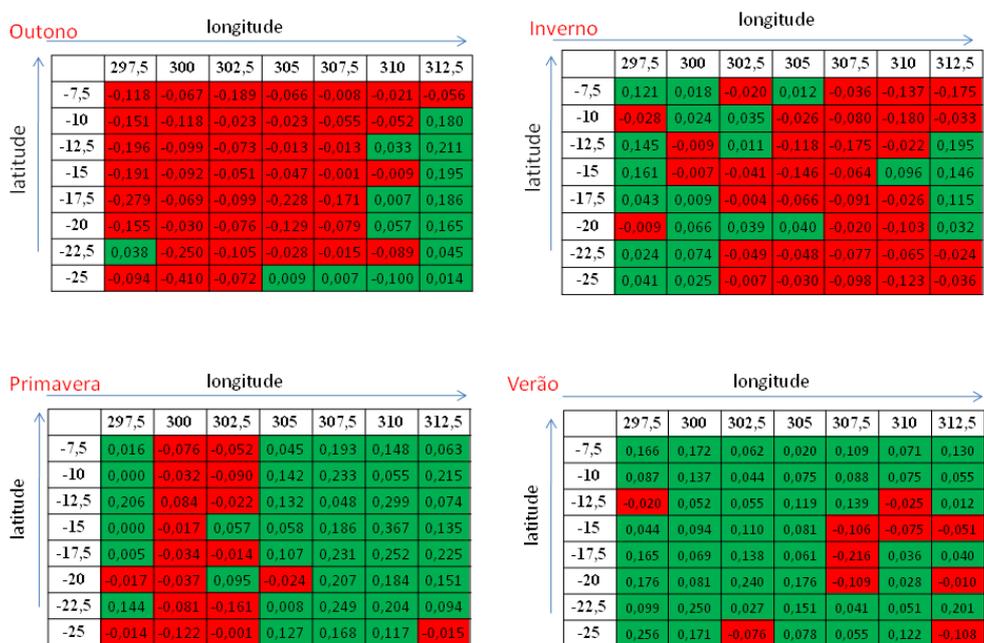


Figura B.11 - Erro em cada ponto de grade: Rede de Jordan usando a TCA(CO).

Erro em cada ponto de grade – Rede MLP usando TCA (CO)

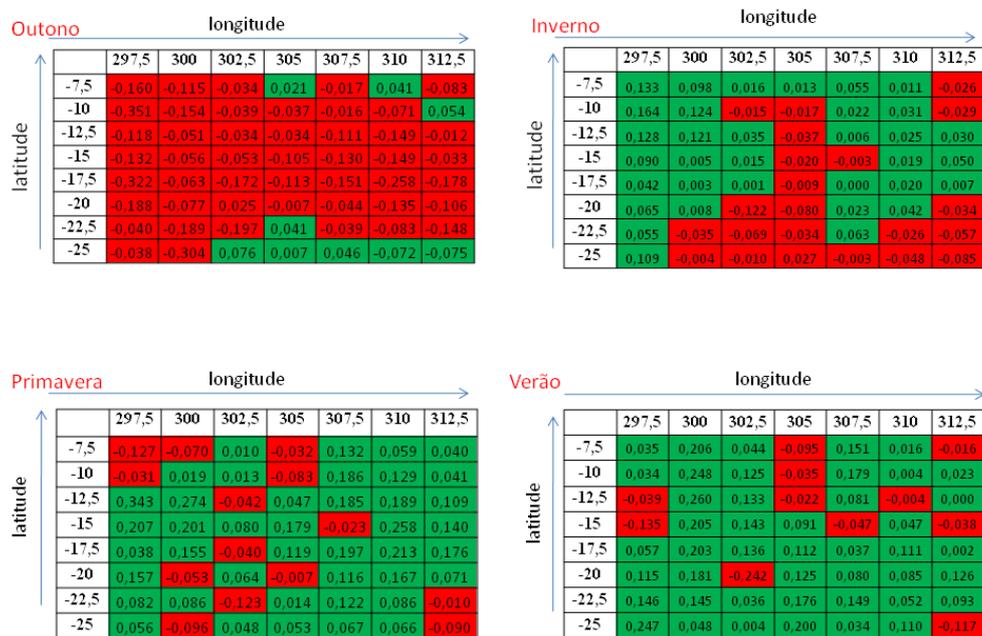


Figura B.12 - Erro em cada ponto de grade: Rede MLP usando a TCA(CO).

B.3 Mapas de erros para os modelos de previsão na região Norte

Os erros foram calculados em cada ponto de grade da região. As coordenadas geográficas que definem a região Norte (N) estão compreendidas entre as longitudes [75°W, 45°W] e entre as latitudes [15°S, 10°N], compreendendo 140 pontos de grade (14 latitudes e 10 longitudes).

As Figuras B.13 a B.15 apresentam os mapas de erros obtidos em cada ponto de grade, criados a partir dos modelos de Elman, Jordan e MLP usando todas as variáveis disponíveis na base de dados.

As Figuras B.16 a B.18 mostram os mapas de erros criados a partir dos modelos de previsão: Elman, Jordan e MLP, em que utilizam o conjunto de dados reduzidos pela TCA.

Erro em cada ponto de grade – Rede de Elman usando todos os dados (N)

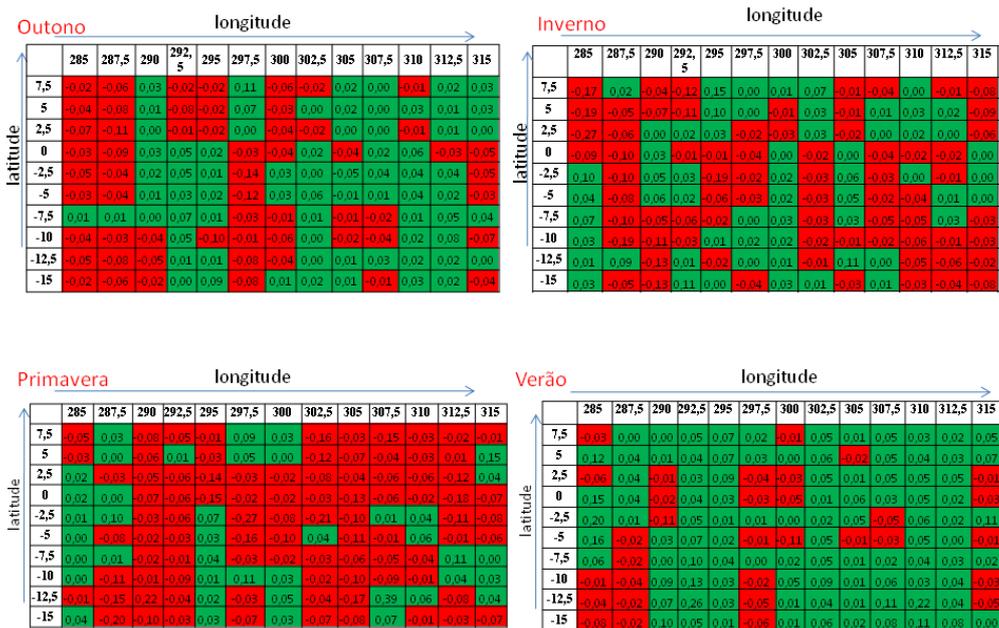


Figura B.13 - Erro em cada ponto de grade: Rede de Elman usando todos os dados(N).

Erro em cada ponto de grade – Rede de Jordan usando todos os dados (N)

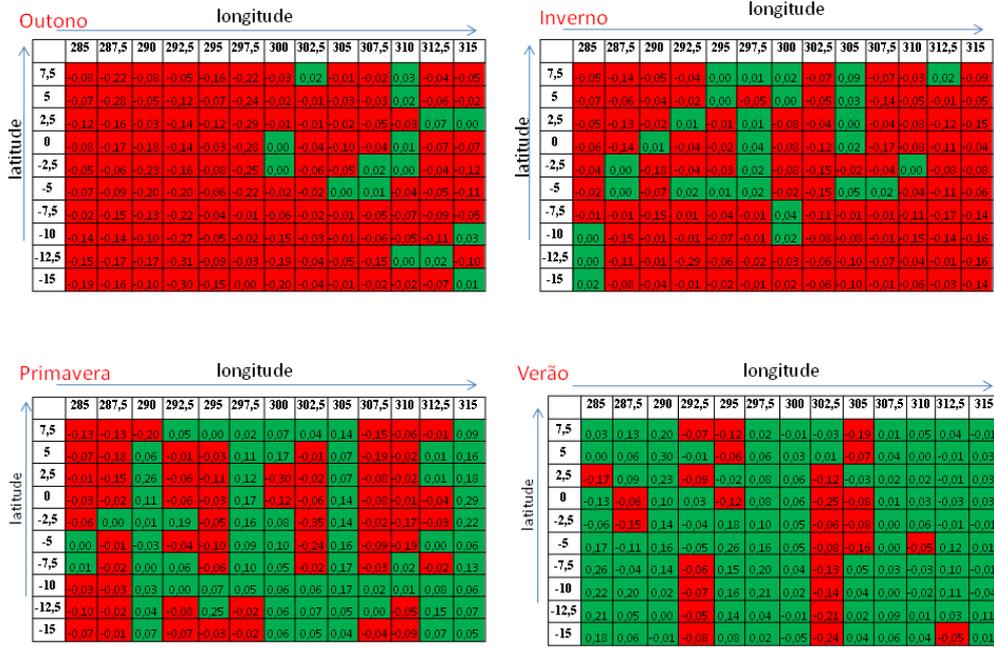


Figura B.14 - Erro em cada ponto de grade: Rede de Jordan usando todos os dados(N).

Erro em cada ponto de grade – Rede MLP usando todos os dados (N)

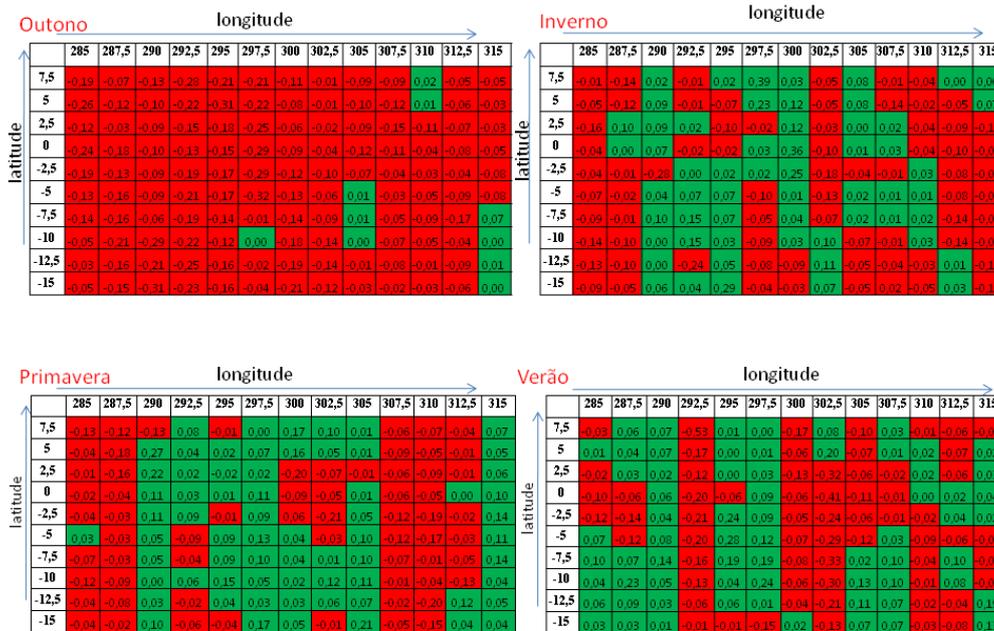


Figura B.15 - Erro em cada ponto de grade: Rede MLP usando todos os dados(N).

Erro em cada ponto de grade – Rede de Elman usando a TCA (N)

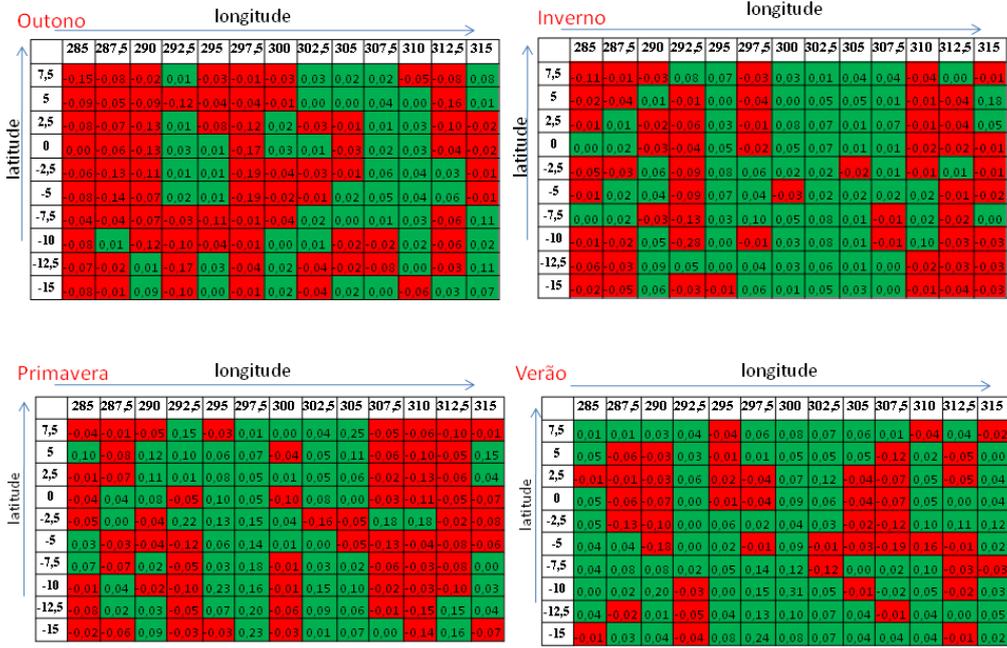


Figura B.16 - Erro em cada ponto de grade: Rede de Elman usando a TCA(N).

Erro em cada ponto de grade – Rede de Jordan usando a TCA (N)

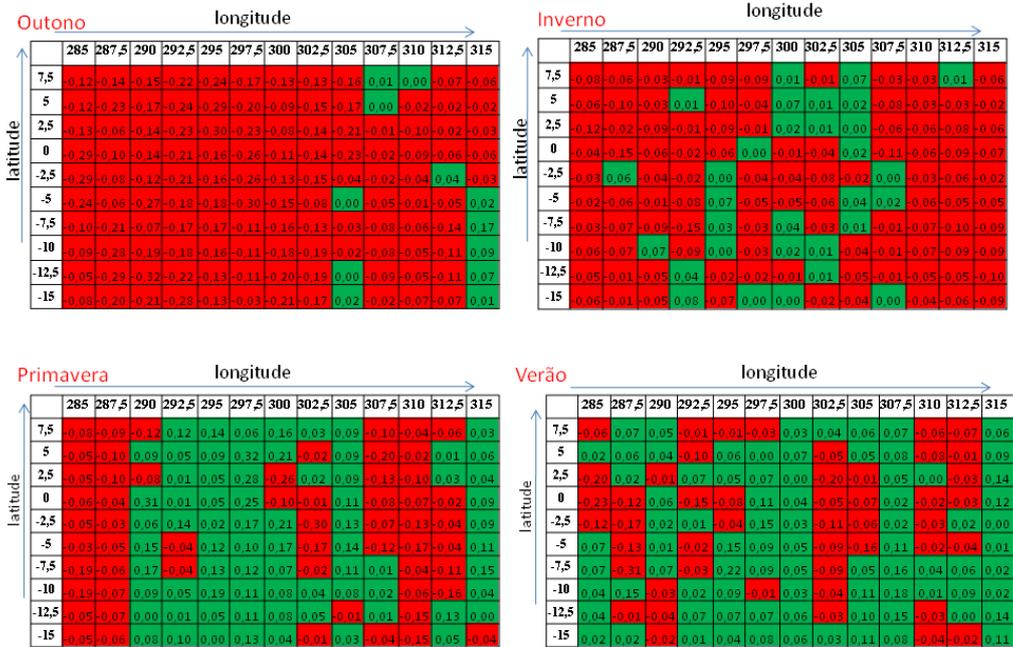


Figura B.17 - Erro em cada ponto de grade: Rede de Jordan usando a TCA(N).

Erro em cada ponto de grade – Rede MLP usando a TCA (N)

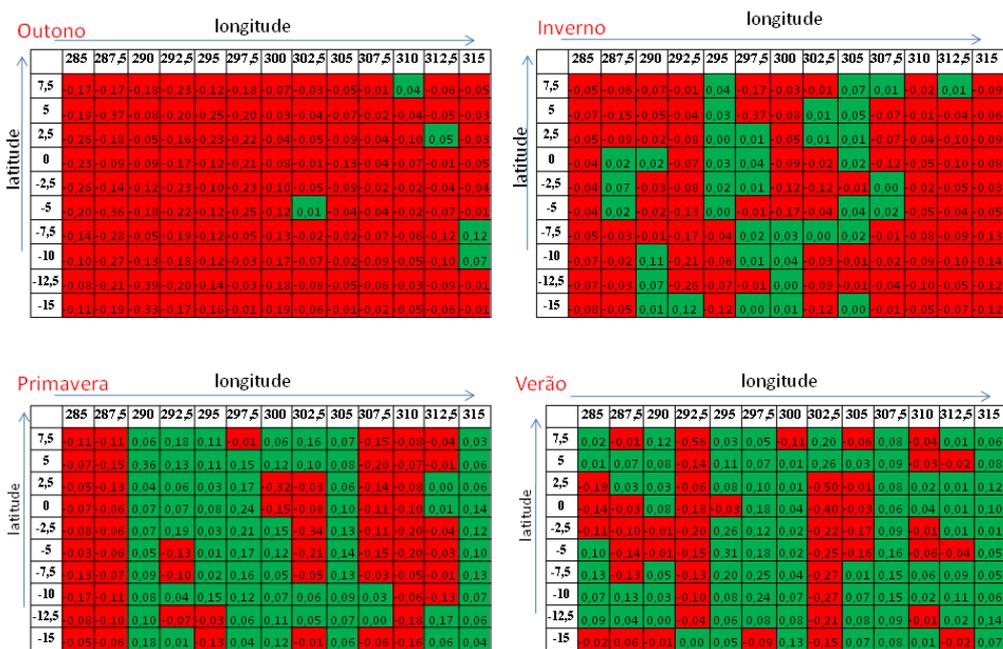


Figura B.18 - Erro em cada ponto de grade: Rede MLP usando a TCA(N).

B.4 Mapas de erros para os modelos de previsão na região Sudeste

Os erros foram calculados em cada ponto de grade da região Sudeste (SE). As coordenadas geográficas que definem a região estão compreendidas entre as longitudes [52°W, 40°W] e entre as latitudes [25°S, 15°S], compreendendo 30 pontos de grade (5 latitudes e 6 longitudes).

As Figuras B.19 a B.21 apresentam os mapas de erros em cada ponto de grade, criados a partir dos modelos de: Elman, Jordan e MLP usando todas as variáveis presentes na base de dados.

As Figuras B.22 a B.24 mostram os mapas de erros obtidos a partir dos modelos de RNAs: Elman, Jordan e MLP, em que utilizam os dados reduzidos através da TCA.

Erro em cada ponto de grade – Rede de Elman usando todos os dados (SE)

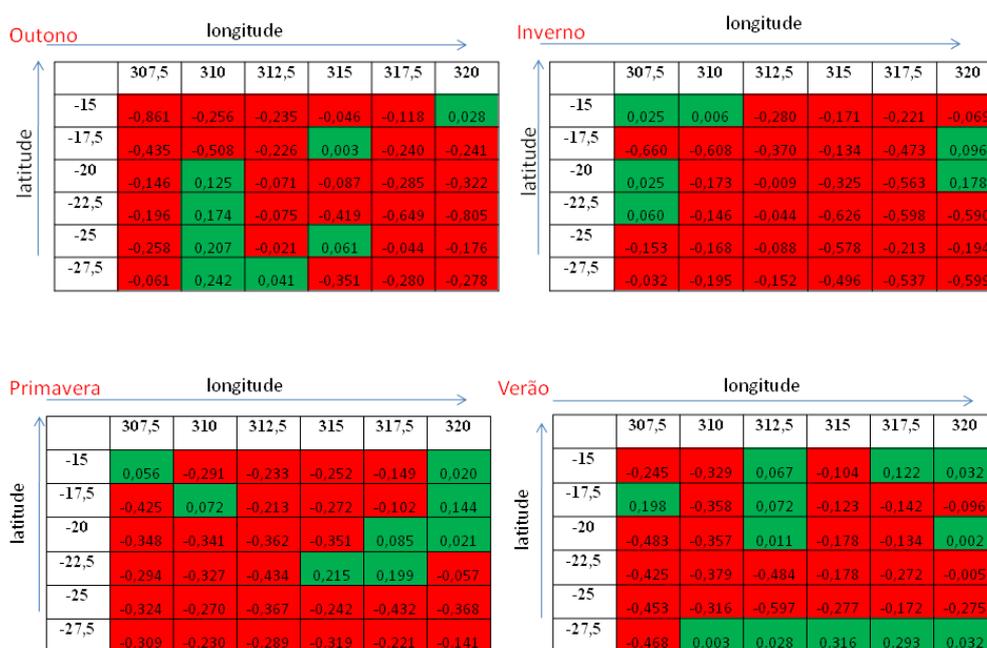


Figura B.19 - Erro em cada ponto de grade: Rede de Elman usando todos os dados(SE).

Erro em cada ponto de grade – Rede de Jordan usando todos os dados (SE)

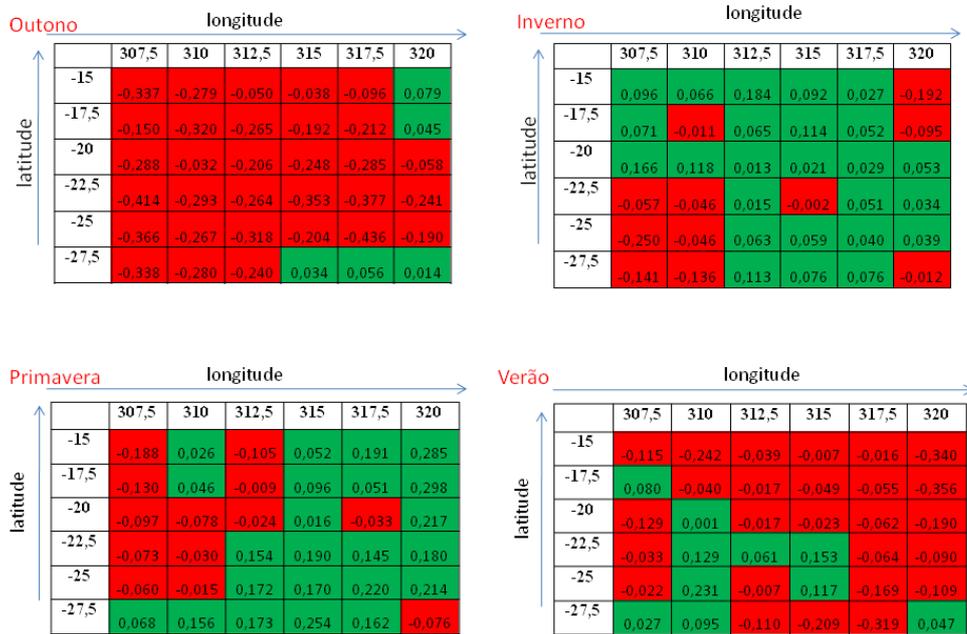


Figura B.20 - Erro em cada ponto de grade: Rede de Jordan usando todos os dados(SE).

Erro em cada ponto de grade – Rede MLP usando todos os dados (SE)

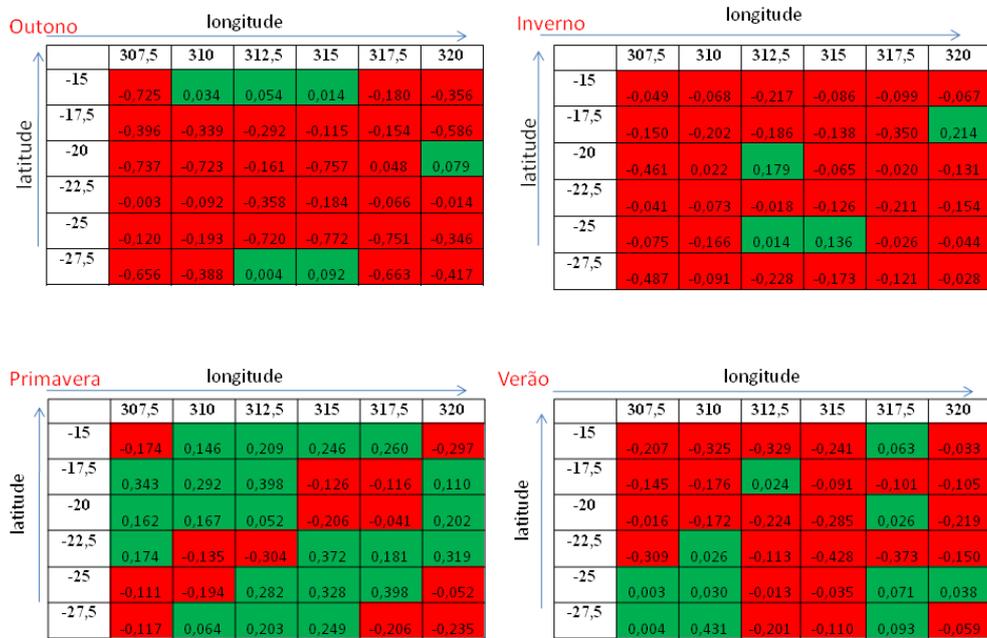


Figura B.21 - Erro em cada ponto de grade: Rede MLP usando todos os dados(SE).

Erro em cada ponto de grade – Rede de Elman usando a TCA (SE)

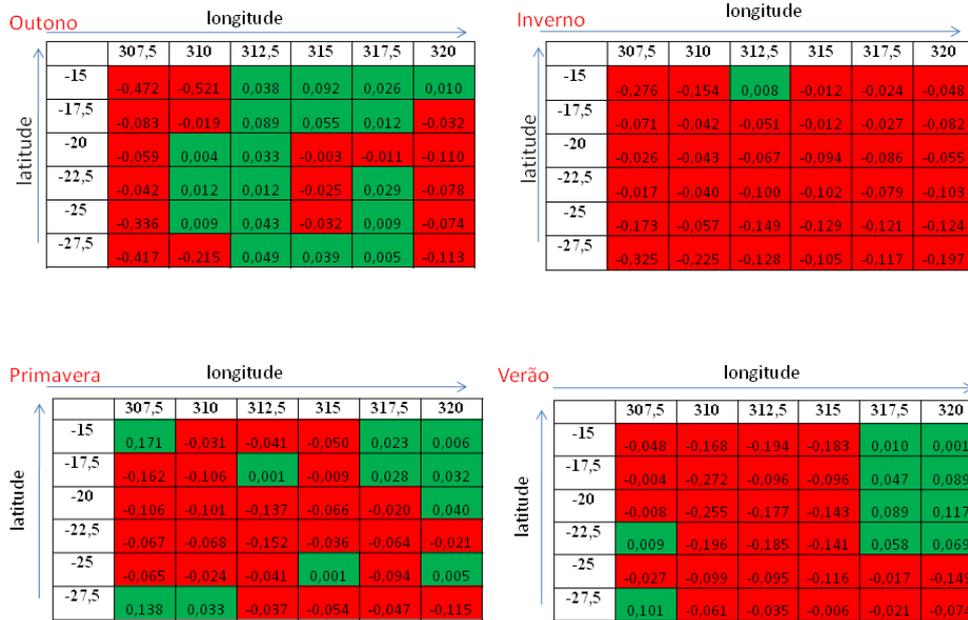


Figura B.22 - Erro em cada ponto de grade: Rede de Elman usando a TCA(SE).

Erro em cada ponto de grade – Rede de Jordan usando a TCA (SE)



Figura B.23 - Erro em cada ponto de grade: Rede de Jordan usando a TCA(SE).

Erro em cada ponto de grade – Rede MLP usando a TCA (SE)

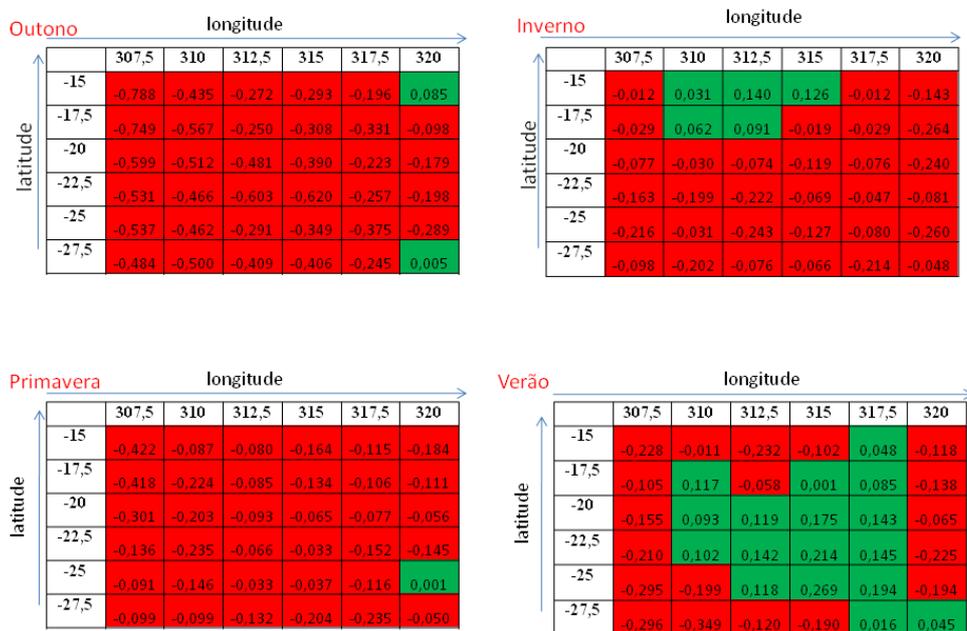


Figura B.24 - Erro em cada ponto de grade: Rede MLP usando a TCA(SE).

B.5 Mapas de erros para os modelos de previsão na região Sul

Os erros foram calculados em cada ponto de grade da região Sul(S). As coordenadas geográficas que definem a região estão compreendidas entre as longitudes [60°W, 45°W] e entre as latitudes [35°S, 22.5°S], compreendendo 42 pontos de grade (7 latitudes e 6 longitudes).

As Figuras B.25 a B.27 apresentam os mapas de erros em cada ponto de grade, construídos a partir dos modelos de redes: Elman, Jordan e MLP, em que utilizam todas as variáveis disponíveis na base de dados.

As Figuras B.28 a B.30 mostram os mapas de erros em cada ponto de grade, obtidos a partir dos modelos: Elman, Jordan e MLP, que utilizam o conjunto de variáveis reduzidas através do uso da TCA.

Erro em cada ponto de grade – Rede de Elman usando todos os dados (S)

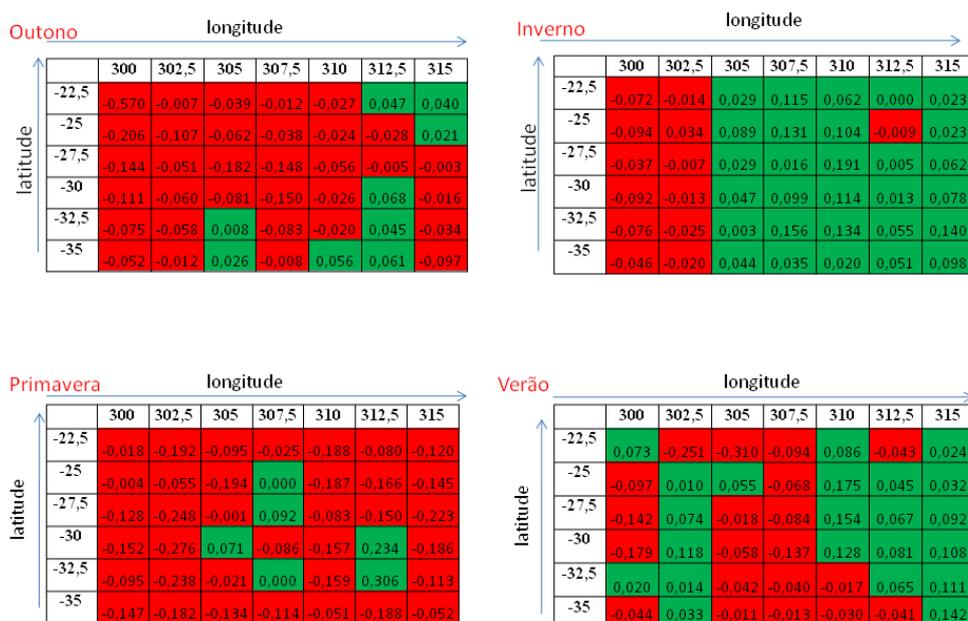


Figura B.25 - Erro em cada ponto de grade: Rede de Elman usando todos os dados(S).

Erro em cada ponto de grade – Rede de Jordan usando todos os dados (S)

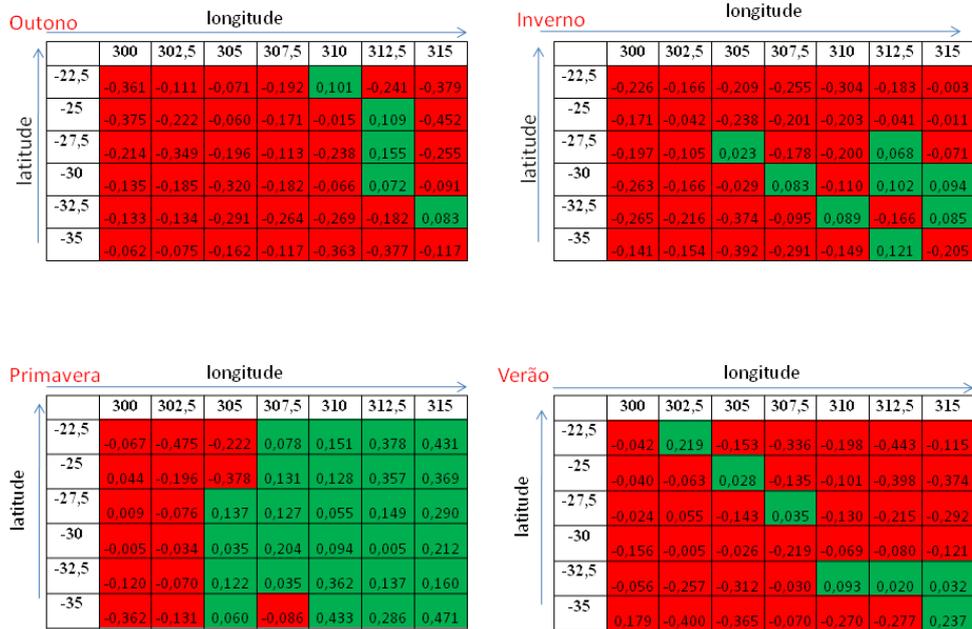


Figura B.26 - Erro em cada ponto de grade: Rede de Jordan usando todos os dados(S).

Erro em cada ponto de grade – Rede MLP usando todos os dados (S)

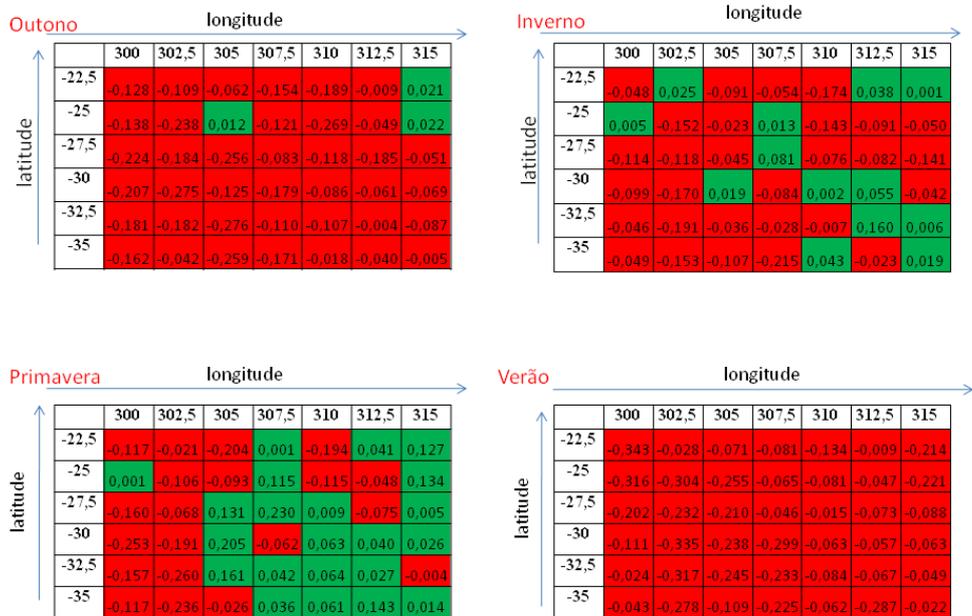


Figura B.27 - Erro em cada ponto de grade: Rede MLP usando todos os dados(S).

Erro em cada ponto de grade – Rede de Elman usando a TCA (S)

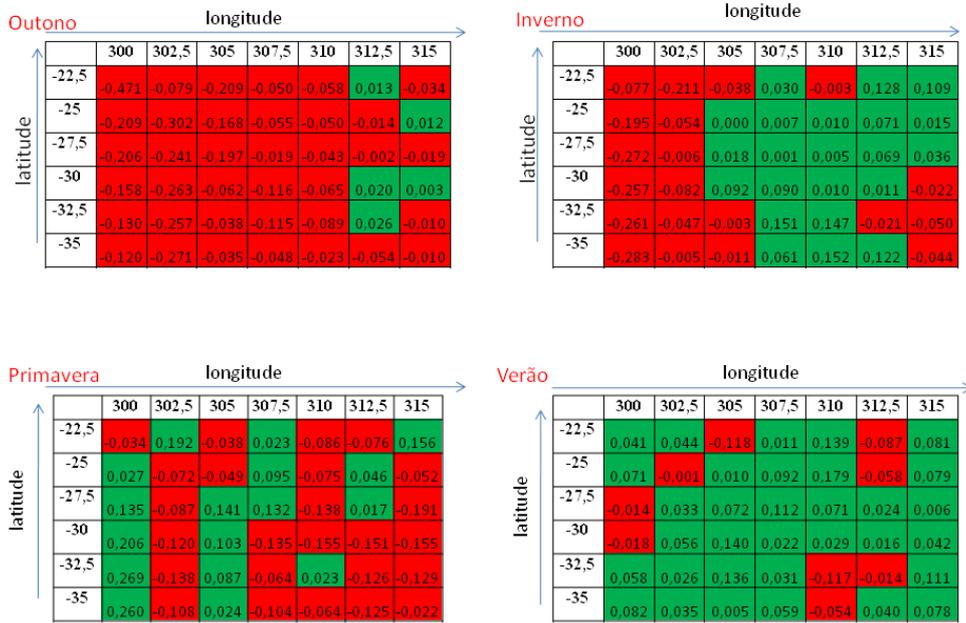


Figura B.28 - Erro em cada ponto de grade: Rede de Elman usando a TCA(S).

Erro em cada ponto de grade – Rede de Jordan usando a TCA (S)

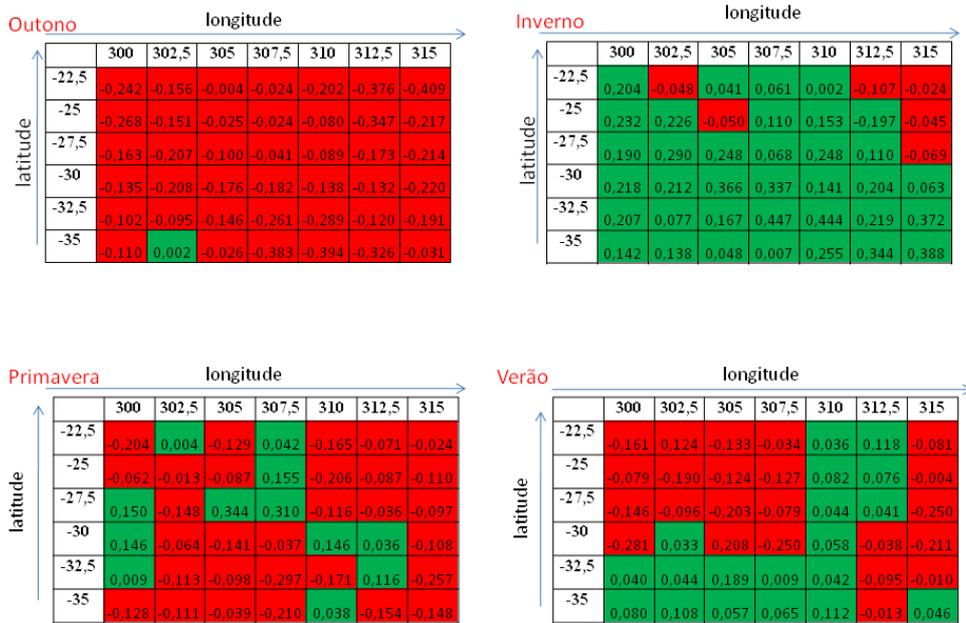


Figura B.29 - Erro em cada ponto de grade: Rede de Jordan usando a TCA(S).

Erro em cada ponto de grade – Rede MLP usando a TCA (S)

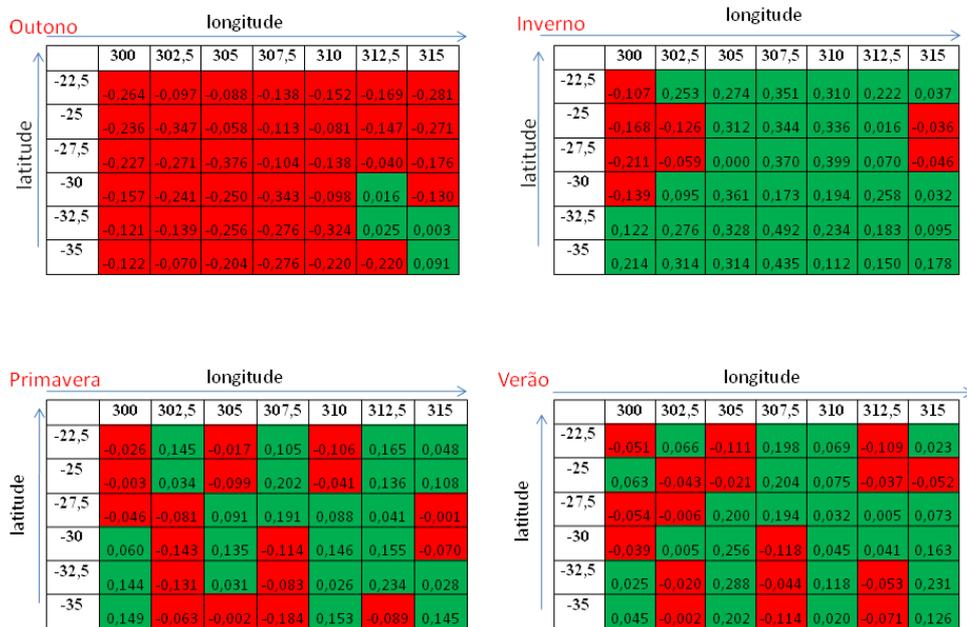


Figura B.30 - Erro em cada ponto de grade: Rede MLP usando a TCA(S).

APÊNDICE C - DADOS DE REANÁLISE

A base de dados utilizada nesta dissertação envolve um grande volume de dados meteorológicos, referente à América do Sul, contida entre as latitudes [10°N, 35°S] e longitudes [80°W, 30°W]. A resolução espacial em ambas as dimensões da grade é de 2.5° e resolução temporal (t) de 1 mês.

A área de estudo para realização dos experimentos abrange as cinco regiões do Brasil: Nordeste (NE), Sudeste (SE), Norte (N), Centro Oeste (CO) e Sul (S), desta forma, os dados foram divididos por regiões.

A Tabela C.1 mostra uma pequena amostra dos dados de reanálise utilizados nesta dissertação.

Tabela C.1 - Dados reanálise América do Sul.

ano	mês	lon	lat	temp	u300	u500	u850	v300	v500	v850	pres	shum	prec
2000	1	313	-18	24,875	-3,7	-1	0,2	3,6	-1,8	-3,5	653	92,282	35,779
2000	2	313	-18	19,134	-2,2	-1	-0,6	2,8	-0,2	-3,4	656	91,974	33,749
2000	3	313	-18	17,724	-0,6	-0,3	-3,3	2,6	-0,4	-4,1	655	92,089	34,632
2000	4	313	-18	17,382	6,8	0,7	-4,5	3,9	1,6	-2,2	657	90,133	29,346
2000	5	313	-18	23,948	18	6,7	-1,9	5,3	3	-1,2	658	68,941	22,129
2000	6	313	-18	24,52	17	7,2	-3,3	0,4	0	-1,6	660	59,641	18,918
2000	7	313	-18	25,695	24	8,7	-3	-3,2	0,1	-1,4	659	50,007	19,267
2000	8	313	-18	20,914	18	7,3	-2,7	2,2	1,4	-1,5	659	39	21,613
2000	9	313	-18	23,248	12	3,3	-4	3,5	1,4	-2	658	51,175	27,007
2000	10	313	-18	23,513	4,3	-0,9	-2,7	3,2	0,6	-0,8	657	50,927	27,605
2000	11	313	-18	23,27	7,6	3,5	-0,9	-0,9	-2,1	-3,8	655	92,658	35,823
2000	12	313	-18	23,475	2,3	2,9	2,4	3,6	0,9	-2,4	655	91,96	36,618
2001	1	313	-18	22,673	-2,8	-2,2	-1,9	3,2	-1,2	-3,4	657	91,258	29,131
2001	2	313	-18	25,053	-7,3	-4,3	-2,7	4,7	-1,5	-2,1	656	90,045	27,877
2001	3	313	-18	18,583	1,3	-1	-2,3	4,6	-0,5	-2,6	657	91,911	34,12
2001	4	313	-18	26,706	1,6	-1,4	-4,1	4,3	1,7	-2,9	658	87,758	29,295
2001	5	313	-18	26,918	20	7,9	-1,5	2,5	1,6	-1,4	658	78,353	26,706
2001	6	313	-18	26,525	15	4,5	-3,4	-3,4	-0,9	-1,8	661	62,149	21,021
2001	7	313	-18	26,207	17	5,4	-2,8	2,9	1,6	-1,7	660	50,016	19,446
2001	8	313	-18	25,548	16	1,6	-5,3	1,3	2,4	-1,9	662	44,573	18,679
2001	9	313	-18	26,051	12	4,7	-2,2	7,2	4,1	-1,1	657	45,375	27,14
2001	10	313	-18	22,817	13	5	-0,6	5,2	2,4	-2,4	657	74,717	33,063
2001	11	313	-18	18,063	7,7	3,3	1,1	2,4	-0,1	-2,6	655	92,017	37,284
2001	12	313	-18	17,704	7,1	2,9	0,3	4,9	0,4	-4,4	655	92,952	35,784
2002	1	313	-18	18,634	2,2	1	-1,6	2,1	-0,6	-3,1	655	91,726	34,946
2002	2	313	-18	24,976	5,2	1,9	0,7	3	0,1	-3,6	656	93	35,73
2002	3	313	-18	25,298	-2,2	-1,2	-2,4	2,3	-0,8	-3,5	657	91,556	32,603
2002	4	313	-18	25,913	-1,6	-3,1	-4,8	2,8	0,3	-3,2	657	87,083	26,506
2002	5	313	-18	20,563	15	3,3	-4,1	5,8	2,5	-2,6	658	75,774	25,423
2002	6	313	-18	23,072	6,2	1,4	-4,4	0,8	-1,3	-1,8	661	58,617	19,237

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.