

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA CIVIL E AMBIENTAL**

**ANÁLISE E PREVISÃO ESTATÍSTICA DO ÍNDICE DE
PRECIPITAÇÃO PADRONIZADA (SPI) PARA O NORDESTE
DO BRASIL**

MARIA VITÓRIA NAVA SILVA DO CARMO

ORIENTADOR: CARLOS HENRIQUE RIBEIRO LIMA

**DISSERTAÇÃO DE MESTRADO EM TECNOLOGIA AMBIENTAL E
RECURSOS HÍDRICOS**

BRASÍLIA/DF: FEVEREIRO – 2018

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA CIVIL**

**ANÁLISE E PREVISÃO ESTATÍSTICA DO ÍNDICE DE
PRECIPITAÇÃO PADRONIZADA (SPI) PARA O NORDESTE DO
BRASIL**

MARIA VITÓRIA NAVA SILVA DO CARMO

**DISSERTAÇÃO SUBMETIDA AO DEPARTAMENTO DE
ENGENHARIA CIVIL E AMBIENTAL DA FACULDADE DE
TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE
MESTRE EM TECNOLOGIA AMBIENTAL E RECURSOS HÍDRICOS.**

APROVADA POR:

**Prof. Carlos Henrique Ribeiro Lima, PhD (ENC-UnB)
(Orientador)**

**Prof. Dirceu Silveira Reis Júnior, PhD (ENC-UnB)
(Examinador Interno)**

**Prof. Eduardo Sávio Passos Rodrigues Martins, PhD (UFC/FUNCEME)
(Examinador Externo)**

BRASÍLIA/DF, 23 DE FEVEREIRO DE 2018

FICHA CATALOGRÁFICA

CARMO, MARIA VITÓRIA NAVA SILVA DO

Análise e previsão estatística do Índice de Precipitação Padronizada (SPI) para o Nordeste do Brasil [Distrito Federal] 2018.

xvi, 111p., 210 x 297 mm (ENC/FT/UnB, Mestre, Tecnologia Ambiental e Recursos Hídricos, 2018).

Dissertação de Mestrado – Universidade de Brasília. Faculdade de Tecnologia.

Departamento de Engenharia Civil e Ambiental.

1. Secas

2. Índices de seca

3. Nordeste

4. Variabilidade espaço-temporal

I. ENC/FT/UnB

II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

CARMO, M. V. N. S. (2018). Análise e previsão estatística do Índice de Precipitação Padronizada (SPI) para o Nordeste do Brasil. Dissertação de Mestrado em Tecnologia Ambiental e Recursos Hídricos, Publicação PTARH.DM-206/2018, Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília, DF, 111p.

CESSÃO DE DIREITOS

AUTOR: Maria Vitória Nava Silva do Carmo.

TÍTULO: Análise e previsão estatística do Índice de Precipitação Padronizada (SPI) para o Nordeste do Brasil.

GRAU: Mestre

ANO: 2018

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

Maria Vitória Nava Silva do Carmo

vitorianava@gmail.com

DEDICATÓRIA

Dedico este trabalho aos meus pais por todo o carinho, incentivo e apoio durante minha vida acadêmica.

AGRADECIMENTOS

Agradeço primeiramente ao meu orientador, Carlos Henrique Ribeiro Lima, pela oportunidade de expandir meus conhecimentos, pela paciência em responder aos meus questionamentos, pelos ensinamentos e pela confiança em mim depositada.

Agradeço também à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), pela bolsa de pesquisa concedida.

Agradeço principalmente a meus pais que sempre me deram amor, carinho e as ferramentas necessárias para que eu atingisse os meus objetivos.

RESUMO

ANÁLISE E PREVISÃO ESTATÍSTICA DO ÍNDICE DE PRECIPITAÇÃO PADRONIZADA (SPI) PARA O NORDESTE DO BRASIL

As secas são o principal desastre natural que afeta o Nordeste brasileiro e o entendimento da sua dinâmica é fundamental para a gestão do risco climático na região. Buscou-se neste trabalho caracterizar os padrões espaço-temporais das secas meteorológicas no Nordeste a partir da análise do Índice de Precipitação Padronizada (SPI, *Standardized Precipitation Index*) e desenvolver um modelo de previsão sazonal para esse índice, usando a temperatura na superfície do mar (TSM) como covariável climática preditora. A técnica análise das componentes principais foi utilizada para decompor o SPI em modos de variabilidade, visando identificar as variações no tempo e no espaço, tendo-se observado uma seca distribuída por toda a região na primeira componente principal (cerca de 42% da variabilidade dos dados), um padrão de variação do tipo dipolo norte-sul na segunda componente principal (11% da variabilidade dos dados) e um padrão zonal leste-oeste na terceira componente principal (6% da variância dos dados). Utilizando as duas primeiras componentes principais, identificou-se as secas mais severas ocorridas durante o período considerado e a porcentagem total da área do Nordeste atingida por cada evento, constatando-se uma coincidência entre os meses detectados como mais secos e os anos secos citados na literatura. A partir desse resultado, foi feita uma análise compósito, apontando que, nas regiões do Nordeste com seca mais severa durante os eventos extremos, as variáveis precipitação e temperatura encontravam-se abaixo e acima do normal, respectivamente. Identificou-se também, por meio de correlações cruzadas com índices climáticos, que eventos de *El Niño* (índice NINO3.4) e temperaturas mais altas no Atlântico Norte do que no Atlântico Sul (índice dipolo do Atlântico tropical) estavam relacionados com as secas na região. Foi feita ainda uma avaliação da correlação entre as componentes principais do SPI e a TSM, encontrando-se que a primeira componente principal seria mais influenciada pelo *El Niño* e a segunda, pelo dipolo do Atlântico. Para melhorar o desempenho dos preditores no modelo de previsão, foi usada a técnica Análise de Correlação Canônica Esparsa (SCCA, *sparse Canonical Correlation Analysis*). Os resultados do modelo indicaram que o desempenho do modelo completo ao prever a primeira componente principal, a terceira componente principal e a média espacial do SPI é maior do que do auto-regressivo, indicando que a informação climática melhora performance do modelo, e que as previsões na parte norte do Nordeste são melhores do que na parte sul para tempos de previsão de 1 e 2 meses.

Palavras-Chaves: secas; índices de seca; Nordeste; variabilidade espaço-temporal; previsão.

ABSTRACT

ANALYSIS AND STATISTICAL FORECAST OF STANDARDIZED PRECIPITATION INDEX (SPI) FOR NORTHEAST BRAZIL

Droughts are the main natural disaster that affects Northeast Brazil and understanding its dynamic is essential to advance in climate risk management in region. This work seeks to characterize spatial-temporal patterns of meteorological droughts in Northeast Brazil through analysis of Standardized Precipitation Index - SPI and to develop a seasonal forecast model for this index, using sea surface temperature (SST) as climate predictor. Principal component analysis technique was used to decompose SPI in modes of variability, in order to identify the index variations in time and space, observing a drought distributed throughout the region in the first principal component (about 42% of data variability), a north-south dipole variation pattern in the second principal component (11% of data variability) and an east-west zonal pattern in the third principal component (6% of data variance). Using the first two principal components, the most severe droughts that occurred during the considered period and the Northeast's percentage area reached in each event were identified, showing a coincidence between months detected as drier and drought years mentioned in literature. From this result, a composite analysis was performed, pointing out that, in Northeast regions with more severe droughts during extreme events, precipitation and temperature were below and above normal, respectively. It was also identified, through cross correlations with climate indexes, that *El Niño* events (NINO3.4 index) and higher temperatures in North Atlantic than in South Atlantic (tropical Atlantic dipole index) were related to droughts in region. It was made an evaluation of the correlation between SPI principal components and SST, finding that the first principal component would be more influenced by *El Niño* and the second principal component by Atlantic dipole. To improve predictor skill in forecast model, sparse Canonical Correlation Analysis (SCCA) technique was used. Results of the model indicated that the full model skill in forecasting SPI first principal component, third principal component and spatial average is greater than that of the auto-regressive model, indicating that climate information improves model performance, and that forecasts in northern part of Northeast are better than in southern part for lead times of 1 and 2 months.

Keywords: droughts; drought indexes; Northeast; spatial-temporal variability; forecasting.

SUMÁRIO

1 - INTRODUÇÃO	1
2 - OBJETIVOS	4
2.1 - OBJETIVO GERAL.....	4
2.2 - OBJETIVOS ESPECÍFICOS	4
3 - REFERENCIAL TEÓRICO	5
3.1 - CONCEITO DE SECAS	5
3.2 - CLASSIFICAÇÃO DAS SECAS.....	5
3.3 - CARACTERÍSTICAS DAS SECAS.....	7
3.4 - POSSÍVEIS CAUSAS.....	8
3.5 - IMPACTOS	9
4 - REVISÃO BIBLIOGRÁFICA	11
4.1 - ÍNDICES DE SECA.....	11
4.1.1 - ESCOLHA DE UM ÍNDICE	12
4.1.2 - EXEMPLOS DE ÍNDICES	13
4.1.3 - SPI	18
4.1.3.1 - Vantagens.....	22
4.1.3.2 - Limitações.....	23
4.2 - PREVISÃO DE SECAS.....	23
5 - ÁREA DE ESTUDO.....	27
5.1 - CARACTERIZAÇÃO DA REGIÃO NORDESTE.....	27
5.1.1 - Caracterização geral	27
5.1.2 - Padrões de precipitação	28
5.2 - SECAS NO NORDESTE.....	31
5.2.1 - Histórico e principais impactos.....	31
5.2.2 - Mecanismos de formação das secas no Nordeste	36
6 - METODOLOGIA.....	39
6.1 - DADOS.....	40
6.1.1 - SPI	40

6.1.2 - Dados Climáticos	40
6.1.3 - Índices Climáticos	40
6.1.4 - Temperatura na Superfície do Mar (TSM)	41
6.2 - PCA	42
6.3 - ANÁLISE DE EXTREMOS E DE ÁREAS.....	46
6.4 - ANÁLISE COMPÓSITO	46
6.5 - ANÁLISE DE CORRELAÇÃO CRUZADA	46
6.6 - CCA	47
6.7 - SCCA.....	50
6.8 - MODELO DE PREVISÃO	53
7 - RESULTADOS	57
7.1 - PCA	57
7.2 - ANÁLISE DE EXTREMOS E DE ÁREAS.....	60
7.2.1 - Análise de extremos.....	60
7.2.2 - Eventos de maior abrangência espacial	66
7.3 - ANÁLISE COMPÓSITO	69
7.4 - INFLUÊNCIA DE INDICADORES CLIMÁTICOS	74
7.5 - CORRELAÇÕES ENTRE AS COMPONENTES PRINCIPAIS DO SPI E A TSM	77
7.6 - SCCA.....	80
7.6.1 - Atraso de 0 meses	80
7.6.2 - Atraso de 3 meses	84
7.7 - PREVISÃO DE SECAS.....	88
8 - CONCLUSÕES E RECOMENDAÇÕES	97
REFERÊNCIAS BIBLIOGRÁFICAS	101
APÊNDICES.....	107
APÊNDICE A – COMPLEMENTO SOBRE CCA	108
APÊNDICE B – PORCENTAGENS DA ÁREA DO NORDESTE NOS EVENTOS EXTREMOS	111

LISTA DE TABELAS

Tabela 4.1 – Categorias de secas e impactos possíveis do Monitor de Secas.....	17
Tabela 4.2 – Categorias de seca do SPI	20
Tabela 5.1 – Principais registros de secas no Nordeste.....	32
Tabela 5.2 – Impactos de algumas secas no Nordeste brasileiro	32
Tabela B.1 - Porcentagens da área do Nordeste atingida por eventos severos e extremos de seca detectados pela primeira e segunda componentes principais do índice SPI.....	111

LISTA DE FIGURAS

Figura 3.1 – Relação entre os tipos de secas (Fonte: adaptado de WMO, 2006).....	7
Figura 3.2 – Parâmetros das secas (Fonte: adaptado de Keyantash e Dracup, 2002)	8
Figura 4.1 – Mapa do USDM para fevereiro de 2018 (Fonte: USDM)	15
Figura 4.2 – Mapa do Monitor de Secas para janeiro de 2018.....	16
Figura 4.3 – Mapa do IIS para o Nordeste em janeiro de 2018 (Fonte: CEMADEN).....	18
Figura 4.4 - Mapa da América do Sul para o SPI de 6 meses, em agosto/2017-janeiro/2018.	21
Figura 4.5 - Mapa do Brasil para o SPI de 3 meses, em janeiro de 2018	22
Figura 5.1 – Média anual do balanço entre a precipitação e a evaporação no Nordeste do Brasil (mm/dia) (Fonte: Projeto PROCLIMA, INPE/CPTEC <i>apud</i> CGEE e ANA, 2012).....	28
Figura 5.2 – Precipitação média anual para a América do Sul e o Atlântico Tropical Sul (em mm/ano) (Fonte: CGEE e ANA, 2012).....	29
Figura 5.3 – Anomalias de precipitação para o Nordeste brasileiro, de 1951 a 2015 (Fonte: Marengo <i>et al.</i> , 2016)	33
Figura 5.4 – Áreas de incidência de secas no Nordeste brasileiro, em porcentagem (Fonte: CGEE e ANA, 2012).....	36
Figura 6.1 – Exemplo em duas dimensões da rotação de eixos na PCA (Fonte: adaptado de Wilks, 2011).....	43
Figura 7.1 - Padrões espaciais dos pesos da primeira componente principal do SPI3.....	58
Figura 7.2 - Padrões espaciais dos pesos da segunda componente principal do SPI3	59
Figura 7.3 - Padrões espaciais dos pesos da terceira componente principal do SPI3	59
Figura 7.4 - Série temporal da primeira componente do SPI3, em que as linhas vermelhas representam os percentis de 6,7%	60
Figura 7.5 - Série temporal da segunda componente do SPI3, em que as linhas vermelhas representam os percentis de 6,7%	61
Figura 7.6 - Porcentagem da área do Nordeste atingida pelos eventos extremos de secas meteorológicas identificados pela primeira componente do SPI3	62
Figura 7.7 - Porcentagem da área do Nordeste atingida pelos eventos extremos de secas meteorológicas identificados pela segunda componente, parte inferior, do SPI3	63
Figura 7.8 - Porcentagem da área do Nordeste atingida pelos eventos extremos de secas meteorológicas identificados pela segunda componente, parte superior, do SPI3.....	64
Figura 7.9 - Abrangência espacial do SPI3 na seca de junho/1983 (40,18% da área atingida)67	

Figura 7.10 - Abrangência espacial do SPI3 na seca de março/1993 (56,43% da área atingida)	68
Figura 7.11 - Abrangência espacial do SPI3 na seca de abril/1993 (49,90% da área atingida)	68
Figura 7.12 - Abrangência espacial do SPI3 na seca de maio/2012 (51,08% da área atingida)	69
Figura 7.13 - Anomalias de precipitação nos extremos da primeira componente do SPI3.....	70
Figura 7.14 - Anomalias de temperatura nos extremos da primeira componente do SPI3	71
Figura 7.15 - Anomalias de precipitação nos extremos da segunda componente, parte inferior, do SPI3	72
Figura 7.16 - Anomalias de temperatura nos extremos da segunda componente, parte inferior, do SPI3	72
Figura 7.17 - Anomalias de precipitação nos extremos da segunda componente, parte superior, do SPI3	73
Figura 7.18 - Anomalias de temperatura nos extremos da segunda componente, parte superior, do SPI3	73
Figura 7.19 - Correlação cruzada entre o Dipolo e a primeira componente principal do SPI3, em que as linhas pretas tracejadas mostram as faixas a partir das quais as correlações são significantes ($\alpha = 5\%$)	75
Figura 7.20 - Correlação cruzada entre o Dipolo e a segunda componente principal do SPI3, em que as linhas pretas tracejadas mostram as faixas a partir das quais as correlações são significantes ($\alpha = 5\%$)	76
Figura 7.21 - Correlação cruzada entre o NINO3.4 e a primeira componente principal do SPI3, em que as linhas pretas tracejadas mostram as faixas a partir das quais as correlações são significantes ($\alpha = 5\%$)	76
Figura 7.22 - Correlação cruzada entre o NINO3.4 e a segunda componente principal do SPI3, em que as linhas pretas tracejadas mostram as faixas a partir das quais as correlações são significantes ($\alpha = 5\%$)	77
Figura 7.23 - Correlações entre a TSM e a primeira componente principal do SPI3	79
Figura 7.24 - Correlações entre a TSM e a segunda componente principal do SPI3.....	79
Figura 7.25 – Correlações entre a TSM e a terceira componente principal do SPI3	80
Figura 7.26 - Coeficiente canônico a_1 para valores de $\text{penalty}_x = 0,2$ (atraso de 0 meses).....	81
Figura 7.27 - Coeficiente canônico a_1 para valores de $\text{penalty}_x = 0,3$ (atraso de 0 meses).....	82
Figura 7.28 - Coeficiente canônico a_1 para valores de $\text{penalty}_x = 0,4$ (atraso de 0 meses).....	82
Figura 7.29 - Coeficiente canônico a_2 para valores de $\text{penalty}_x = 0,2$ (atraso de 0 meses).....	83

Figura 7.30 - Coeficiente canônico a_2 para valores de $\text{penalty}_x = 0,3$ (atraso de 0 meses).....	83
Figura 7.31 - Coeficiente canônico a_2 para valores de $\text{penalty}_x = 0,4$ (atraso de 0 meses).....	84
Figura 7.32 - Coeficiente canônico a_1 para valores de $\text{penalty}_x = 0,2$ (atraso de 3 meses).....	85
Figura 7.33 - Coeficiente canônico a_1 para valores de $\text{penalty}_x = 0,3$ (atraso de 3 meses).....	86
Figura 7.34 - Coeficiente canônico a_1 para valores de $\text{penalty}_x = 0,4$ (atraso de 3 meses).....	86
Figura 7.35 - Coeficiente canônico a_2 para valores de $\text{penalty}_x = 0,2$ (atraso de 3 meses).....	87
Figura 7.36 - Coeficiente canônico a_2 para valores de $\text{penalty}_x = 0,3$ (atraso de 3 meses).....	87
Figura 7.37 - Coeficiente canônico a_2 para valores de $\text{penalty}_x = 0,4$ (atraso de 3 meses).....	88
Figura 7.38 - SPI previsto pelo modelo completo (linhas azuis) e SPI observado (linhas pretas) para o tempo de previsão de 1 mês. A hachura cinza mostra o intervalo de previsão de 95%	89
Figura 7.39 - SPI previsto pelo modelo completo (linhas azuis) e SPI observado (linhas pretas) para o tempo de previsão de 2 meses. A hachura cinza mostra o intervalo de previsão de 95%	89
Figura 7.40 - SPI previsto pelo modelo completo (linhas azuis) e SPI observado (linhas pretas) para o tempo de previsão de 4 meses. A hachura cinza mostra o intervalo de previsão de 95%	90
Figura 7.41 - SPI previsto pelo modelo completo (linhas azuis) e SPI observado (linhas pretas) para o tempo de previsão de 6 meses. A hachura cinza mostra o intervalo de previsão de 95%	90
Figura 7.42 - Coeficientes de correlação em 1 a 6 meses de previsão do modelo completo (linhas cheias) e do modelo auto-regressivo (linhas pontilhadas) para a primeira (linhas pretas), para a segunda (linhas vermelhas) e para a terceira (linhas azuis) componentes principais do SPI e para a média espacial (linhas verdes) do SPI	92
Figura 7.43 - Valores do RMSE (<i>Root-Mean Squared Error</i>) em 1 a 6 meses de previsão do modelo completo (linhas cheias) e do modelo auto-regressivo (linhas pontilhadas) para a primeira (linhas pretas), para a segunda (linhas vermelhas) e para a terceira (linhas azuis) componentes principais do SPI e para a média espacial (linhas verdes) do SPI	93
Figura 7.44 - Valores do RMSE normalizado para o modelo completo (em relação ao modelo auto-regressivo), em tempos de previsão de 1 a 6 meses, para a primeira (linhas pretas), para a segunda (linhas vermelhas) e para a terceira (linhas azuis) componentes principais do SPI e para a média espacial (linhas verdes) do SPI	94
Figura 7.45 - Distribuição espacial do BSS (<i>Brier Skill Score</i>) para o tempo de previsão de 1 mês	95

Figura 7.46 - Distribuição espacial do BSS (<i>Brier Skill Score</i>) para o tempo de previsão de 3 meses	95
Figura 7.47 - Distribuição espacial do BSS (<i>Brier Skill Score</i>) para o tempo de previsão de 4 meses	96
Figura 7.48 - Distribuição espacial do BSS (<i>Brier Skill Score</i>) para o tempo de previsão de 6 meses	96

LISTA DE SÍMBOLOS, NOMECLATURA E ABREVIACÕES

α	Nível de significância
ANA	Agência Nacional de Águas
BS	<i>Brier Score</i>
BSS	<i>Brier Skill Score</i>
CCA	<i>Canonical Correlation Analysis</i>
CEMADEN	Centro Nacional de Monitoramento e Alertas de Desastres Naturais
CMI	<i>Crop Moisture Index</i>
Corr	Correlação
CO ₂	Dióxido de carbono
CPC	<i>Climate Prediction Center</i>
CPTEC	Centro de Previsão de Tempo e Estudos Climáticos
D	Duração
DAEE	Departamento de Águas e Energia Elétrica de São Paulo
ECMWF	<i>European Center for Medium-Range Weather Forecasts</i>
ENSO	<i>El Niño Oscilação Sul</i>
EOF	Empirical Orthogonal Function
FUNCEME	Fundação Cearense de Meteorologia e Recursos Hídricos
IBGE	Instituto Brasileiro de Geografia e Estatística
IIS	Índice Integrado de Seca
INMET	Instituto Nacional de Meteorologia
INPE	Instituto Nacional de Pesquisas Espaciais
IRI	<i>International Research Institute for Climate and Society</i>
GIDMaPS	<i>Global Integrated Drought Monitoring and Prediction System</i>
LDEO	<i>Lamont-Doherty Earth Observatory</i>
M	Magnitude
MSDI	<i>Multivariate Standardized Drought Index</i>
MSE	<i>Mean Squared Error</i>
NAO	<i>North Atlantic Oscillation</i>
NDMC	<i>National Drought Mitigation Center</i>
NOAA	<i>National Oceanic and Atmospheric Administration</i>
PCA	<i>Principal Component Analysis</i>
PDSI	<i>Palmer Drought Severity Index</i>
PHDI	<i>Palmer Hydrological Drought Severity Index</i>
PIB	Produto Interno Bruto
PMD	<i>Penalized Matrix Decomposition</i>
RMSE	<i>Root-Mean Squared Error</i>
S	Severidade
SCCA	<i>sparse Canonical Correlation Analysis</i>
SPEI	<i>Standardized Precipitation Evapotranspiration Index</i>
SPI	<i>Standardized Precipitation Index</i>
SSI	<i>Standardized Soil Moisture Index</i>
SWSI	<i>Surface Water Supply Index</i>
TNA	<i>Tropical Northern Atlantic Index</i>
TSA	<i>Tropical Southern Atlantic Index</i>
TSM	Temperatura da Superfície do Mar
USDM	<i>U.S. Drought Monitor</i>

Var	Variância
VCI	<i>Vegetation Condition Index</i>
VSWI	<i>Vegetation Supply Water Index</i>
WMO	<i>World Meteorological Organization</i>
<i>Z Index</i>	<i>Palmer Moisture Anomaly Index</i>
ZCAS	Zona de Convergência do Atlântico Sul
ZCIT	Zona de Convergência Intertropical

1 - INTRODUÇÃO

As secas configuram-se como uma catástrofe natural e socioeconômica, tendo causado vários problemas em muitos lugares do mundo ao longo da história. Trata-se do pior desastre natural em número de pessoas afetadas e de prejuízos financeiros (Rajsekhar *et al.*, 2015; Aghakouchak, 2015). O fato de não contar com uma definição universal, de se estender por grandes regiões e de haver uma dificuldade inerente em definir seu início e seu término (Wilhite e Buchanan-Smith, 2005) dificulta as ações políticas para lidar com esse problema. É normal então que as nações tenham ações reativas a esses fenômenos, quando o ideal seria a adoção de medidas preventivas, por serem mais eficientes (Banco Mundial, 2015). Essas medidas contribuem para reduzir a vulnerabilidade das regiões e uma das maneiras de se fazer isso é por meio da previsão desse evento extremo, o que demanda um conhecimento acerca da região atingida e do seu clima.

O Nordeste brasileiro é uma região conhecida pelos impactos causados por secas intensas, configurando-se entre os principais desastres climáticos dos trópicos em função das grandes perdas econômicas e problemas sociais que causam (Hastenrath *et al.*, 1984; Hastenrath, 1990). Apenas em 1958, por exemplo, aproximadamente 10 milhões de pessoas abandonaram suas residências em função da seca, enquanto a seca de 1979-1983 custou cerca de 7,8 bilhões de dólares para o governo (CGEE e ANA, 2012; Marengo *et al.*, 2016). As causas apontadas para as secas remontam ao clima predominantemente semiárido, com alta variabilidade interanual e espacial da precipitação, chegando a menos de 400 mm por ano em algumas áreas (Kousky, 1979), e à grande vulnerabilidade da região, em especial por parte da população mais carente. Apesar do investimento em medidas mitigatórias de infraestrutura, a população, especialmente os mais pobres, ainda sofre com as consequências desse fenômeno natural, lidando, por exemplo, com o risco à segurança alimentar, hídrica e energética (Anderson *et al.*, 2015).

A dinâmica espaço-temporal das chuvas no Nordeste é complexa e seu entendimento científico, assim como de mecanismos de teleconecção climática associados, remonta da década de 1970. Foram encontradas relações entre o desastre natural e alguns mecanismos climáticos, tais como a temperatura da superfície do mar (TSM) no Oceano Atlântico (Hastenrath, 1990; Hastenrath *et al.*, 1984; Hastenrath e Heller, 1977), a posição da Zona de Convergência Intertropical (ZCIT) (Hastenrath, 1990; Moura e Shukla, 1981; Rao *et al.*, 1993), a pressão ao nível do mar

nos Oceanos Atlântico e Pacífico tropical (Hastenrath e Heller, 1977; Hastenrath, 1978), os ventos alísios (Namias, 1972; Kousky, 1980; Moura e Shukla, 1981; Rao *et al.*, 1993), o fenômeno *El Niño* (Rao *et al.*, 1993; Marengo *et al.*, 2016) e as frentes frias (Kousky, 1979).

Em razão da dinâmica e dos impactos causados em diferentes escalas espaciais e temporais, torna-se fundamental desenvolver ferramentas que possam melhor caracterizar, monitorar e prever as secas. Nesse sentido, foram desenvolvidos diversos índices de seca, que buscam quantificar a severidade, a duração e a distribuição espacial do fenômeno a partir da evolução temporal da chuva e eventualmente de outras variáveis climáticas (Hao e Singh, 2015). O estudo de índices de seca possui uma extensa literatura, como, por exemplo, o trabalho de Hao e AghaKouchak (2013), que verificaram, também nos Estados Unidos, que o SPI identifica com antecedência o início de uma seca e mostra grande variabilidade. Santos *et al.* (2013) usaram uma análise de frequência regional para o cálculo do SPI, que foi usado para determinar períodos de seca em postos no Alto São Francisco. Dantas *et al.* (2014) adaptaram o SPI para classificar chuvas intensas juntamente com um estudo de regionalização, visando caracterizar os padrões de precipitação que geram inundações na bacia do rio Uno, em Pernambuco. Canamary (2015) verificou a capacidade do SPI em monitorar as secas no Nordeste do Brasil e avaliou a possibilidade de utiliza-se na previsão de secas, observando que esse índice consegue prever esses eventos com um desempenho melhor que a climatologia. Canamary *et al.* (2015) avaliaram a performance de previsões do SPI para bacias hidrográficas no estado do Ceará, observando resultados melhores do que a climatologia para tempos de previsão superiores a 2 meses. Costa *et al.* (2016) estudaram as relações entre diferentes índices climáticos, o SPI e os eventos de seca na estação seca no Leste do Nordeste do Brasil.

O monitoramento e a previsão das secas são muito importantes para subsidiar ações de gestão que contribuam para aumentar a resiliência da população atingida e evitar ou, pelo menos, para mitigar os impactos gerados. Quanto ao monitoramento, destaca-se a iniciativa do Monitor de Secas do Nordeste (disponível em <<http://monitordesecas.ana.gov.br/>>), que integra diferentes indicadores para fornecer mapas mensais da intensidade da seca na região. Em relação à previsão, ela tem sido, na maior parte dos casos, limitada no Brasil à variável precipitação, o que não é suficiente para caracterizar as secas, por não as medir diretamente. O mais adequado seria fazer uso dos índices de seca, mas as iniciativas de previsão de índices ainda são poucas. Nesse sentido, busca-se, neste trabalho, avançar no entendimento da dinâmica das secas meteorológicas no Nordeste a partir da análise dos padrões espaço-temporais de secas na região

indicados pelo índice SPI e utilizar então um modelo estatístico para previsão sazonal do índice a partir de informações climática de larga escala. Utiliza-se nesse caso a ferramenta de estatística multivariada Análise de Correlação Canônica Esparsa (SCCA, *sparse Canonical Correlation Analysis*) para obter um modelo sazonal (1-6 meses) de previsão do índice de seca SPI no Nordeste do Brasil, a partir da variável climática temperatura da superfície do mar (TSM).

Este trabalho está organizado em 8 capítulos, sendo o primeiro esta introdução. O segundo capítulo apresenta os objetivos gerais e específicos. O item 3 traz o marco teórico e conceitual acerca dos conceitos básicos sobre secas, incluindo suas classificações, causas e principais impactos. O capítulo 4 é uma revisão bibliográfica, em que são apresentados os conceitos, critérios de escolha, exemplos de aplicação e iniciativas de previsão dos principais índices de seca. O capítulo 5 apresenta uma descrição geral sobre o clima e os regimes de precipitação no Nordeste brasileiro, incluindo uma caracterização das secas na região. O capítulo 6 apresenta a metodologia adotada, descrevendo as técnicas estatísticas utilizadas, como a Análise de Correlação Canônica Esparsa (SCCA, *sparse Canonical Correlation Analysis*) e o modelo de previsão usado. O item 7 traz os resultados obtidos, bem como uma discussão acerca deles. Por fim, o capítulo 8 mostra as conclusões e as recomendações para trabalhos futuros.

2 - OBJETIVOS

2.1 - OBJETIVO GERAL

Construir um modelo de previsão sazonal (1-6 meses) para o índice Índice de Precipitação Padronizada (SPI) no Nordeste do Brasil.

2.2 - OBJETIVOS ESPECÍFICOS

- Avaliar o SPI como índice de seca capaz de caracterizar a seca meteorológica no Nordeste do Brasil;
- Avaliar e verificar variáveis climáticas relacionadas às secas meteorológicas caracterizadas pelo SPI;
- Identificar a relação entre a temperatura da superfície do mar e o SPI, de forma a subsidiar a obtenção do modelo de previsão.

3 - REFERENCIAL TEÓRICO

O referencial teórico traz as bases conceituais sobre as secas. É descrito o conceito (item 3.1), a classificação (item 3.2), as características (item 3.3), as possíveis causas (item 3.4) e, por último, os impactos (item 3.5) desse fenômeno.

3.1 - CONCEITO DE SECAS

As secas podem ser descritas como um evento natural causado por um déficit prolongado de precipitação em relação à quantidade necessária para atender às necessidades do ser humano e do meio ambiente (Wilhite e Buchanan-Smith, 2005; WMO, 2006). A Organização Meteorológica Mundial (WMO, *World Meteorological Organization*) e a Lei nº 13.153, de 30 de julho de 2015, que instituiu a “Política Nacional de Combate à Desertificação e Mitigação dos Efeitos da Seca” fornecem conceitos semelhantes. O fenômeno em si não constitui um desastre natural, podendo tornar-se a depender dos impactos causados nas esferas social e ambiental (Wilhite e Buchanan-Smith, 2005; WMO, 2006).

Não existe, entretanto, uma definição de seca precisa e que seja aceita globalmente, pois esse fenômeno é determinado de forma distinta para cada região, já que cada localidade possui um clima com características específicas (Wilhite e Buchanan-Smith, 2005). Além disso, cada setor econômico sofre com as secas de forma única, levando-os a conceituá-las de maneiras distintas, podendo haver diferenças inclusive dentro de um mesmo setor (Wilhite e Buchanan-Smith, 2005). Na agricultura, por exemplo, cada cultura reage ao déficit hídrico de uma maneira, já que a fase de crescimento e a sensibilidade à falta d’água e à mudança de temperatura são diferentes (Wilhite e Buchanan-Smith, 2005).

3.2 - CLASSIFICAÇÃO DAS SECAS

As secas são tradicionalmente classificadas na literatura em quatro tipos: meteorológica, hidrológica, agrícola e socioeconômica (Keyantash e Dracup, 2002; Brito *et al.*, 2018). A seca meteorológica é caracterizada por um período com escassez de precipitação em uma determinada região (Keyantash e Dracup, 2002; Mishra e Singh, 2010; CGEE e ANA, 2012;

Hao e Singh, 2015; Brito *et al.*, 2018), sendo a origem dos outros tipos de secas (Hao e Singh, 2015).

A seca agrícola ocorre quando a umidade do solo está abaixo da quantidade necessária para atender à demanda das plantas e para repor as perdas por evapotranspiração (Keyantash e Dracup, 2002), resultando na perda da cultura (Mishra e Singh, 2010). Geralmente, observa-se um atraso no início desse tipo de seca em relação à meteorológica, a depender do estado de umidade inicial do solo (Heim, 2002).

Esse déficit de precipitação, quando prolongado, tem como consequência uma redução das vazões e dos níveis dos lagos, dos reservatórios e das águas subterrâneas, caracterizando uma seca hidrológica, a qual persistente mesmo após o encerramento da seca meteorológica (CGEE e ANA, 2012; Rajsekhar *et al.*, 2015; Hao e Singh, 2015), em razão do lento tempo de recarga das águas superficiais e subterrâneas. Mishra e Singh (2010) classificam o déficit hídrico das águas subterrâneas como um quinto tipo de seca, a subterrânea, mas essa não é uma posição majoritária na literatura.

Por fim, tem-se a seca socioeconômica, que pode ser considerada uma consequência dos outros três tipos, os quais são considerados como tipos físicos de secas, em associação com fatores humanos (Wilhite e Buchanan-Smith, 2005). Ocorre quando a quantidade de água disponível não consegue atender às demandas da sociedade e dos setores econômicos (Mishra e Singh, 2010).

A Figura 3.1 traz a sequência mais usual de ocorrência dos diferentes tipos de secas. A seca agrícola, a hidrológica e a socioeconômica são menos frequentes do que a meteorológica, por estarem mais relacionadas à disponibilidade hídrica no sistema (WMO, 2006). Problemas nas culturas e nos pastos costumam demorar semanas para serem notados após o início do déficit de precipitação, enquanto a redução nos níveis dos reservatórios superficiais e subterrâneos pode ser percebida após alguns meses (WMO, 2006). Da esquerda para a direita da figura, os impactos vão tornando-se mais complexos, podendo as ações de gestão amenizá-los ou agravá-los.

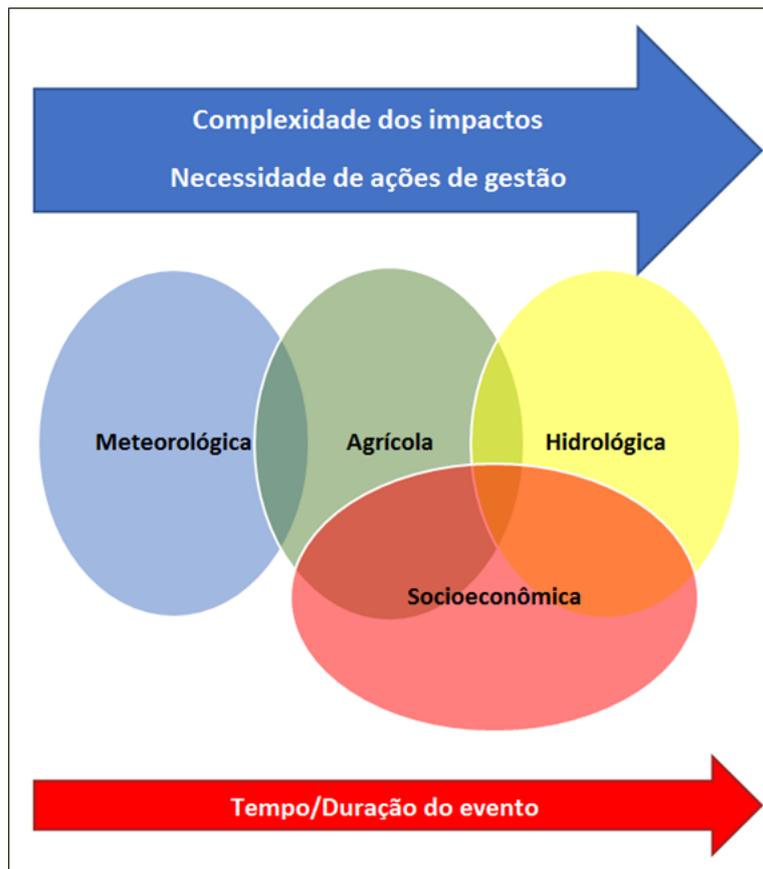


Figura 3.1 – Relação entre os tipos de secas (Fonte: adaptado de WMO, 2006)

3.3 - CARACTERÍSTICAS DAS SECAS

As secas podem ocorrer em qualquer clima, tanto em locais que apresentam alta pluviosidade quanto nos de baixa precipitação, o que faz com que cerca de metade da superfície terrestre seja suscetível à ocorrência do evento (Mishra e Singh, 2010). Elas não devem ser confundidas com aridez, já que esta é uma característica permanente do clima de uma região, enquanto a seca constitui um distúrbio temporário (Mishra e Singh, 2010).

As secas são fenômenos de início lento, o que faz com que seus efeitos se acumulem por um período de tempo considerável (Wilhite e Buchanan-Smith, 2005). Por esse motivo, a determinação do início e do final do evento apresenta uma dificuldade para cientistas e gestores, não existindo um consenso sobre qual fator considerar para determinar o término de uma seca (Wilhite e Buchanan-Smith, 2005; Mishra e Singh, 2010). O fato dos impactos serem não estruturais e se estenderem por grandes áreas dificulta a quantificação deles e, conseqüentemente, a adoção de medidas corretivas (Wilhite e Buchanan-Smith, 2005; Mishra e Singh, 2010). Além disso, as secas, diferentemente de outros desastres naturais, podem ter

participação humana em seu desencadeamento, a partir de comportamentos como desmatamento, irrigação excessiva e uso pouco consciente dos recursos hídricos (Wilhite e Buchanan-Smith, 2005).

De acordo com Wilhite e Buchanan-Smith (2005) e WMO (2006), as secas são tecnicamente diferenciadas com base em três parâmetros: intensidade, duração e alcance espacial. O primeiro refere-se à magnitude do déficit de precipitação ou dos seus impactos associados, sendo medido com base no desvio em relação às condições normais de alguma variável climática ou de um índice de seca. As secas costumam estabilizar-se após dois ou três meses, podendo durar de meses a anos, caracterizando o parâmetro duração. Ele, combinado com a intensidade ou magnitude, define a severidade dos impactos. A Figura 3.2 mostra as grandezas duração (D), intensidade ou magnitude (M) e severidade (S) com base na anomalia de uma variável ou índice qualquer. Além dessas três propriedades, a extensão espacial também é importante para o planejamento de ações mitigatórias.

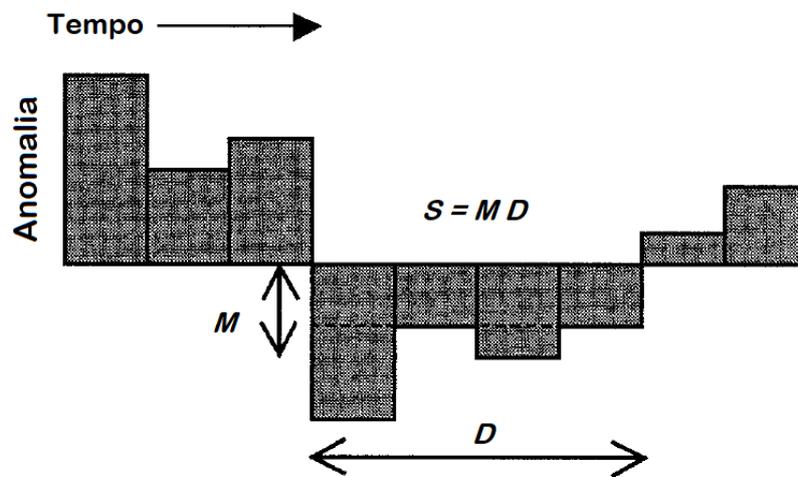


Figura 3.2 – Parâmetros das secas (Fonte: adaptado de Keyantash e Dracup, 2002)

3.4 - POSSÍVEIS CAUSAS

As secas são fenômenos complexos, já que estão conectadas à atmosfera e aos processos hidrológicos capazes de fornecer umidade para o ar. A seca meteorológica relaciona-se à variabilidade regional do ciclo global da água, o qual se associa aos padrões de circulação climática, sendo necessário estudar fenômenos globais para entender secas locais e regionais (AghaKouchak *et al.*, 2015). A ocorrência de secas relaciona-se não apenas às características

dos períodos chuvosos, mas também a outras variáveis, como alta temperatura, ventos fortes, umidade relativa do ar baixa, menor quantidade de água no solo e maior taxa de evapotranspiração (Mishra e Singh, 2010; Marengo *et al.*, 2013; Marengo e Bernasconi, 2015).

Mishra e Singh (2010) apontam um registro de acréscimo de eventos extremos, como secas e cheias, em alguns locais do mundo. Em relação à escassez hídrica, ela tem aumentado devido a um crescimento da demanda por água, em razão do desenvolvimento populacional, industrial, energético e agrícola das últimas décadas. Mudanças climáticas, que causam intensificação do ciclo hidrológico, e contaminação dos mananciais também podem contribuir para esse problema.

3.5 - IMPACTOS

A lista de impactos causados pelas secas é grande, devido às suas características de se difundir por grandes áreas e de possuir longa duração, consistindo em uma grande preocupação com a qualidade de vida das populações e com a segurança alimentar (WMO, 2006). Elas prejudicam a agricultura e os ecossistemas de maneira geral, implicando em muitos custos econômicos e sociais (Mishra e Singh, 2010). São responsáveis pela morte de animais, além de afetarem os recursos hídricos, tanto superficiais quanto subterrâneos, reduzindo a disponibilidade para os diversos usos, como abastecimento, geração de energia e atividades de recreação (Mishra e Singh, 2010). Também modificam a qualidade dos corpos hídricos, já que provocam uma alteração no regime hidrológico e, conseqüentemente, na química das águas superficiais (Webster *et al.*, 1996).

No caso da seca agrícola, o fenômeno pode gerar redução de produtividade ou mesmo perda da cultura (Mishra e Singh, 2010), causando, muitas vezes, a migração da população para outras áreas (WMO, 2006). Pode causar degradação dos solos e, algumas vezes, provocar o fenômeno da desertificação (WMO, 2006; Marengo e Bernasconi, 2015). AghaKouchak *et al.* (2015) apontam, ainda, o efeito das secas nos ciclos de carbono e de nitrogênio, já que a vegetação responde ao estresse hídrico de forma estrutural e fisiológica, como, por exemplo, reduzindo sua cobertura foliar, fazendo com que a captura geral de dióxido de carbono (CO₂) do ecossistema seja reduzida, aumentando assim a quantidade do gás presente na atmosfera.

É preciso destacar, entretanto, que os riscos de uma seca se associam à probabilidade de ocorrência do evento e também à vulnerabilidade da população local e essa vulnerabilidade, por sua vez, depende de aspectos políticos, sociais, econômicos, culturais (Wilhite e Buchanan-Smith, 2005). Uma sociedade menos favorecida economicamente, por exemplo, é mais vulnerável aos impactos de uma seca. É preciso considerar esses fatores na elaboração de planos de preparo, de mitigação e compensação para o desastre natural.

4 - REVISÃO BIBLIOGRÁFICA

Este capítulo traz uma revisão bibliográfica acerca de índices de seca (item 4.1) e das iniciativas existentes de previsão de secas (item 4.2). Dentro do tópico de índices de seca, além da conceituação dessa ferramenta, são fornecidos critérios para subsidiar a escolha de um índice para uma determinada aplicação (item 4.1.1), alguns exemplos de índices (item 4.1.2) e também é feita uma descrição do índice utilizado neste trabalho, o SPI (item 4.1.3), incluindo suas vantagens (item 4.1.3.1) e suas limitações (item 4.1.3.2).

4.1 - ÍNDICES DE SECA

As secas são fenômenos complexos e de grande impacto, fazendo com que seja necessário o desenvolvimento de métodos capazes de caracterizá-las e de prevê-las. Nessa questão, os índices de secas aparecem como uma importante ferramenta na integração de diferentes variáveis relacionadas ao fenômeno. Além de serem úteis na detecção e no monitoramento das condições de secas, eles também podem determinar o tempo e o nível das respostas ao desastre (Steinemann *et al.*, 2005). A partir desses índices, pode-se caracterizar e comparar diferentes eventos de secas, além de desenvolver planos e sistemas de monitoramento e de alerta de secas (Aghakouchak, 2015).

Índices de secas são variáveis capazes de representar numericamente as secas em relação aos seus efeitos e às suas propriedades físicas, tais como duração, severidade, extensão espacial e intensidade (Mishra e Singh, 2010; Hao e Singh, 2015; Steinemann *et al.*, 2005). Eles traduzem as informações de diversas variáveis e indicadores em um único número, sendo, portanto, úteis para caracterizar uma situação e para servir como suporte para a tomada de decisão (Hao e Singh, 2015). Podem ser compostos por diferentes variáveis relacionadas às secas, como precipitação, vazão, umidade do solo, armazenamento de reservatórios e nível das águas subterrâneas (Steinemann *et al.*, 2005; Hao *et al.*, 2016).

É comum classificar os índices de secas de acordo com o tipo de seca que eles caracterizam, ou seja, como índice de seca meteorológica, agrícola e hidrológica (Hao e Singh, 2015; Hao *et al.*, 2016). Os índices de seca meteorológica utilizam principalmente as variáveis precipitação, temperatura e evapotranspiração (Steinemann *et al.*, 2005). Dentre elas destaca-se a

precipitação, utilizada na maior parte dos índices sozinha ou em combinação com outros parâmetros, sendo útil por medir diretamente o fornecimento de água e por refletir os impactos das secas no tempo (Steinemann *et al.*, 2005). Os índices de seca hidrológica consideram variáveis que indicam como a seca hidrológica demora mais para se desenvolver e para encerrar do que a meteorológica, como nível das águas subterrâneas, vazões, umidade do solo, armazenamento dos reservatórios e quantidade de neve (Steinemann *et al.*, 2005). É comum, para os índices de seca agrícola, o uso da umidade do solo, por ser o fator de maior importância no desenvolvimento das culturas (Mishra e Singh, 2010).

Muito índices foram desenvolvidos nas últimas décadas, a partir de abordagens com diferentes graus de complexidade (Hao e Singh, 2015; Mishra e Singh, 2010). Alguns utilizam apenas uma variável, como precipitação, enquanto outros trabalham com uma combinação de algumas delas. Eles fazem uso de fontes diversas, como medições em campo, dados de sensoriamento remoto, simulações da superfície terrestre e até mesmo informações de especialistas (Hao e Singh, 2015).

Um aspecto importante quando se lida com secas é definir gatilhos ou patamares, que são limiares a partir dos quais se atinge um determinado nível de seca e determina-se quando ações de gestão devem ser iniciadas ou encerradas (Steinemann *et al.*, 2005). Eles são representados por um valor específico de um indicador ou índice e a definição de qual valor utilizar pode impactar muito os resultados de uma análise, fazendo com que essa escolha nem sempre seja simples (Steinemann *et al.*, 2005). É preciso considerar, por exemplo, as peculiaridades da região avaliada e os possíveis impactos (Steinemann *et al.*, 2005). A partir desses gatilhos, podem-se definir diferentes categorias para as secas, como “leve”, “moderada”, “severa”.

4.1.1 - ESCOLHA DE UM ÍNDICE

Escolher o índice a ser usado em uma determinada situação é fundamental para lidar adequadamente com o problema. O desempenho de um índice, porém, depende da localidade em que ele é aplicado, já que é função das variáveis meteorológicas e hidrológicas, que são específicas para cada região (Mishra e Singh, 2010) e, além disso, os índices também trabalham de forma distinta em períodos do ano diferentes (Steinemann *et al.*, 2005). Não existe, por isso, um critério universal para a escolha de um índice, tão pouco existe um que seja superior aos demais (Steinemann *et al.*, 2005). O contexto sempre precisa ser analisado para a definição do

índice a ser usado, entretanto, alguns aspectos podem ser listados para auxiliar na escolha para uma determinada aplicação.

Primeiramente, deve-se considerar a adequabilidade do índice ao tipo de seca que se deseja monitorar, pois, como as secas dependem de muitos fatores, não existe um único indicador que consiga caracterizar todos os tipos possíveis (Steinemann *et al.*, 2005; Rajsekhar *et al.*, 2015). Outra questão muito importante é a disponibilidade de dados e sua consistência, já que um índice precisa ser fácil de gerar e não pode ser muito custoso (Steinemann *et al.*, 2005) monetária e computacionalmente. Assim, ao selecionar um índice para uma determinada área, é preciso verificar, primeiramente, se os dados necessários para o seu cálculo estão disponíveis para aquela localidade.

É preciso também identificar se os índices são claros e válidos, sendo necessário que eles sejam fáceis de entender, para subsidiar adequadamente a tomada de decisão. Eles precisam ser testados tanto antes da ocorrência de uma seca quanto durante, para avaliar seu desempenho e a sensibilidade espacial e temporal também precisa ser reconhecida (Steinemann *et al.*, 2005). Precisam também estar associados a um determinado período de cálculo, de forma a definir uma escala de análise (Steinemann *et al.*, 2005). De acordo com Mishra e Singh (2010), o ideal é que um índice seja capaz de analisar uma seca por diferentes escalas temporais. Além disso, os gatilhos precisam ter consistência estatística. É também fundamental que os índices estejam diretamente ligados aos objetivos da gestão e às metas de redução de impactos (Steinemann *et al.*, 2005).

4.1.2 - EXEMPLOS DE ÍNDICES

A lista de índices de seca descritos na literatura é enorme. Apenas a título de exemplo, podem-se citar alguns dos mais conhecidos: Índice de Severidade de Seca de Palmer (PDSI, *Palmer Drought Severity Index*); Índice de Precipitação Padronizada (SPI, *Standardized Precipitation Index*); Índice de Umidade de Cultura (CMI, *Crop Moisture Index*); Índice de Anomalia de Umidade de Palmer (Z Index, *Palmer Moisture Anomaly Index*); Índice de Severidade de Seca Hidrológica de Palmer (PHDI, *Palmer Hydrological Drought Severity Index*); Índice Padronizado de Precipitação-Evapotranspiração (SPEI, *Standardized Precipitation Evapotranspiration Index*); Índice de Umidade do Solo Padronizada (SSI, *Standardized Soil*

Moisture Index); Índice de Condição da Vegetação (VCI, *Vegetation Condition Index*); Índice de Seca Padronizada Multivariada (MSDI, *Multivariate Standardized Drought Index*) .

Existem também exemplos de várias instituições que fazem uso de índices para monitorar espacialmente as condições de seca. O *Global Integrated Drought Monitoring and Prediction System* (GIDMaPS) (disponível em <<http://drought.eng.uci.edu/>>), vinculado à Universidade da Califórnia, em Irvine, Estados Unidos, fornecia, até fevereiro de 2016, informações sobre secas, constituindo um sistema de monitoramento e de previsão. Ele fazia uso de três índices (SPI, SSI e MSDI) e de outras fontes de dados para fornecer mapas da distribuição da magnitude dos índices na Terra. Também era possível obter mapas da previsão sazonal desses índices.

Nos Estados Unidos, existe o *National Drought Mitigation Center* (NDMC), fundado em 1995, com o intuito de auxiliar em medidas para reduzir a vulnerabilidade da sociedade às secas e na implementação de gestão de risco, ao invés de gestão de crise. O NDMC ajuda no monitoramento de secas, participando do *U.S. Drought Monitor* (USDm). O USDm (disponível em <<http://droughtmonitor.unl.edu/CurrentMap.aspx>>), criado em 1999, fornece mapas semanais das condições de seca em todo o território norte-americano, contando ainda com o apoio do *United States Department of Agriculture* e do *National Oceanic and Atmospheric Administration* (NOAA). O USDm baseia-se no uso de vários indicadores, como PDSI, SPI, vegetação, umidade do solo (WMO, 2006). A Figura 4.1 mostra um exemplo desses mapas, para fevereiro de 2018.

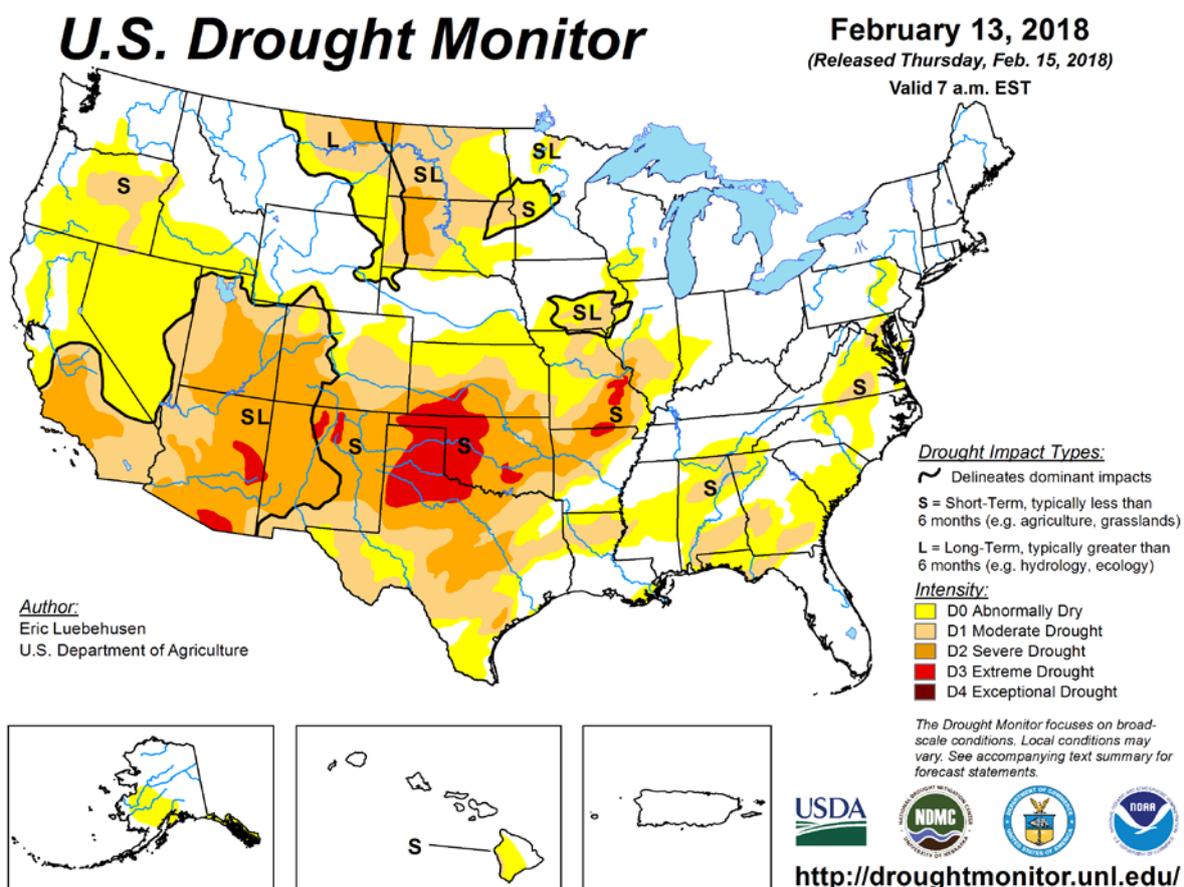


Figura 4.1 – Mapa do USDM para fevereiro de 2018 (Fonte: USDM)

O Brasil também possui um monitor de secas, desenvolvido especialmente para o Nordeste (disponível em <<http://monitordesecas.ana.gov.br/>>). Ele fornece, mensalmente, o chamado Mapa do Monitor de Secas, com a situação e a evolução da seca na região, por meio de indicadores de severidade, informando também se é uma seca de curto ou longo prazo. Esses indicadores integram indicadores meteorológicos, hidrológicos e agrícolas de diferentes fontes. O Monitor, que começou em 2014 em fase experimental, foi inspirado nas iniciativas dos Estados Unidos para lidar com esse fenômeno. Trata-se de um produto conjunto, produzido com a colaboração de instituições federais e estaduais, e que visa subsidiar políticas públicas, tomada de decisão e medidas para combater os impactos das secas, com o intuito de ser uma ferramenta de fácil interpretação tanto para especialistas quanto para a população em geral. O mapa é acompanhado por um texto com a descrição da situação.

A Figura 4.2 mostra o Mapa do Monitor de Secas para janeiro de 2018. Observa-se que grande parte do interior nordestino apresentava a maior intensidade do indicador, apresentando seca

excepcional (categoria S4). O sítio eletrônico do Monitor também fornece uma tabela (transcrita na Tabela 4.1) com as categorias para classificação da severidade da seca, bem como os impactos associados a elas.

Monitor de Secas Janeiro/2018

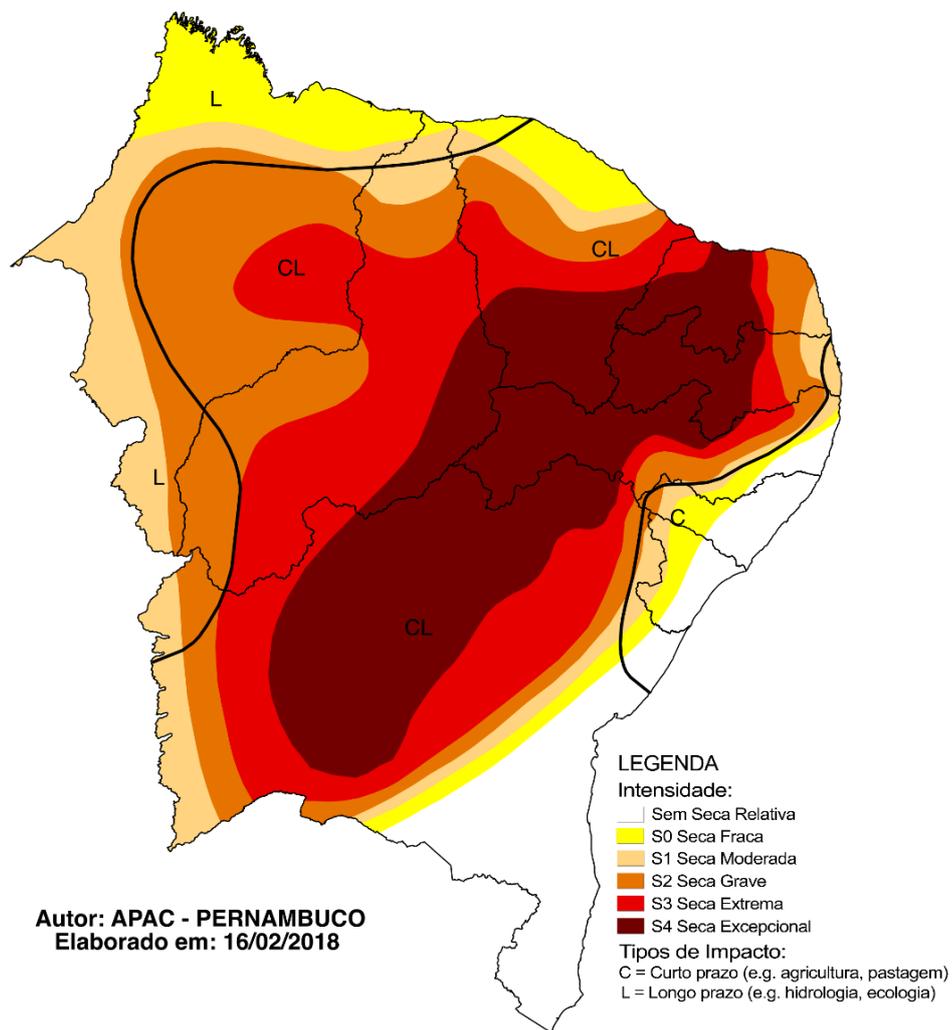


Figura 4.2 – Mapa do Monitor de Secas para janeiro de 2018
(Fonte: Monitor de Secas do Nordeste)

Tabela 4.1 – Categorias de secas e impactos possíveis do Monitor de Secas
(Fonte: adaptado de Monitor de Secas do Nordeste)

Categoria	Percentil	Descrição	Impactos Possíveis
S0	30 %til	Seca Fraca	Entrando em seca: veranico de curto prazo diminuindo plantio, crescimento de culturas ou pastagem. Saindo de seca: alguns déficits hídricos prolongados, pastagens ou culturas não completamente recuperadas.
S1	20 %til	Seca Moderada	Alguns danos às culturas, pastagens; córregos, reservatórios ou poços com níveis baixos, algumas faltas de água em desenvolvimento ou iminentes; restrições voluntárias de uso de água solicitadas.
S2	10 %til	Seca Grave	Perdas de cultura ou pastagens prováveis; escassez de água; restrições de água impostas.
S3	5 %til	Seca Extrema	Grandes perdas de culturas/pastagem; escassez de água generalizada ou restrições.
S4	2 %til	Seca Excepcional	Perdas de cultura/pastagem excepcionais e generalizadas; escassez de água nos reservatórios, córregos e poços de água, criando situações de emergência.

O Centro Nacional de Monitoramento e Alertas de Desastres Naturais – CEMADEN (disponível em <<http://www.cemaden.gov.br/categoria/monitoramento/secas-no-semiarido/>>) fornece diagnósticos mensais da situação das secas no semiárido nordestino, mostrando mapas baseados em diferentes índices de seca: Índice de Precipitação Padronizada (SPI, *Standardized Precipitation Index*), Índice Integrado de Seca (IIS) e o Índice de Suprimento de Água para a Vegetação (VSWI, *Vegetation Supply Water Index*). Eles mostram também a porcentagem da área impactada pelas secas. A Figura 4.3 mostra o mapa do IIS no Nordeste para janeiro de 2018.

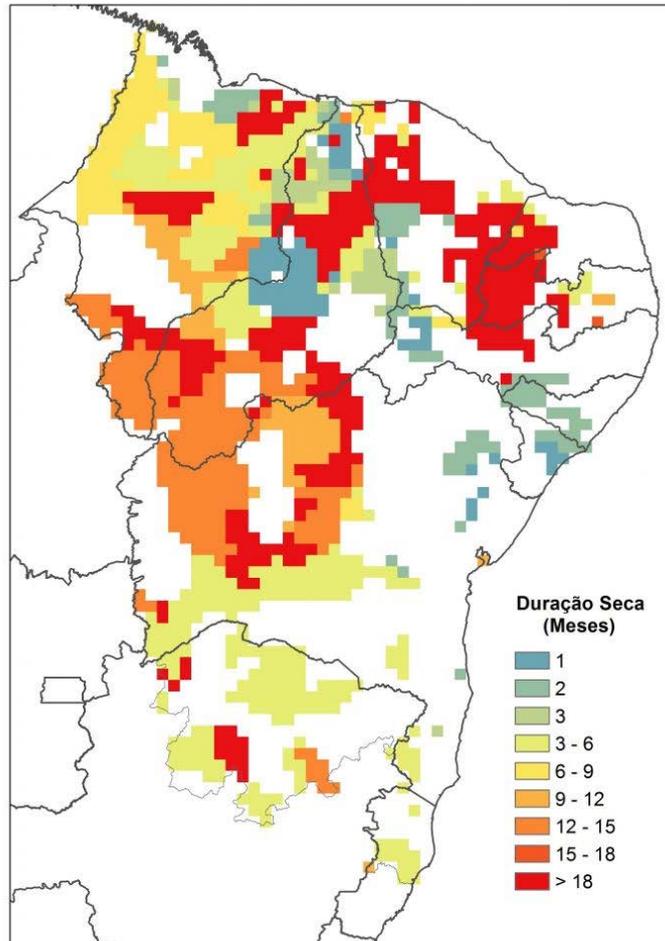


Figura 4.3 – Mapa do IIS para o Nordeste em janeiro de 2018 (Fonte: CEMADEN)

4.1.3 - SPI

Para avaliar as secas meteorológicas no Nordeste brasileiro, decidiu-se utilizar o Índice de Precipitação Padronizada (SPI, *Standardized Precipitation Index*), por ter grande aceitação no meio científico, sendo um dos índices mais utilizados no monitoramento. É um dos índices usado no *Global Integrated Drought Monitoring and Prediction System* (GIDMaPS), no Monitor de Secas do Nordeste e também está disponível no banco de dados do IRI/LDEO (*International Research Institute for Climate and Society/Lamont-Doherty Earth Observatory*).

A seleção desse índice foi corroborada pelo trabalho de Canamary (2015). Ela pesquisou a aplicabilidade de alguns índices de seca para o Nordeste setentrional, ou seja, para a região localizada no semiárido brasileiro, englobando nove estados: Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe, Bahia e Minas Gerais. Os índices foram

escolhidos com base na disponibilidade dos dados necessários à aplicação e também considerando aqueles que haviam sido mais usados por outros pesquisadores, em razão de serem muito conhecidos. Assim, foram selecionados PDSI, o PHDI, o Z index, o SPI e o SPEI, bem como um índice desenvolvido pela Fundação Cearense de Meteorologia e Recursos Hídricos (FUNCEME).

A autora verificou que todos os índices avaliados foram capazes de identificar as secas históricas ocorridas na região, embora tenham sido notados vários tipos de erros. O SPI e o SPEI apresentaram comportamentos semelhantes, tendo representado melhor os eventos no período de acumulação de 12 meses. Ela também analisou a capacidade de previsão do SPI mensal e trimestral, tendo ele mostrado-se mais eficiente para antecipar eventos de seca do que a climatologia.

O SPI também foi explorado no trabalho de Brito *et al.* (2018). Os autores usaram o SPI de 12 meses e o Índice de Saúde da Vegetação (VHI, *Vegetation Health Index*) para avaliar os eventos de seca ocorridos entre 1981 e 2016, com ênfase na seca atual (de 2011 a 2016), na região do semiárido nordestino. Eles fizeram uso do conceito de ano hidrológico (1º de outubro até 30 de setembro) e consideraram períodos de 5 anos e dados mensais. Para cada evento seco identificado, os autores determinaram o começo e o final, a duração, a severidade e a frequência (por período de 5 anos e pelos 36 anos considerados).

O *Standardized Precipitation Index* (SPI), ou, traduzindo, o Índice de Precipitação Padronizada, é um índice de seca meteorológica e caracteriza-se por incluir apenas dados de precipitação em seu cálculo (Mishra e Singh, 2010; Banco Mundial, 2015). Foi introduzido por McKee *et al.* (1993), que definiu precipitação padronizada como a diferença entre o valor de precipitação e a média histórica, dividida pelo desvio padrão. Trata-se de um dos índices mais utilizados, sendo recomendado pela Organização Meteorológica Mundial (WMO, *World Meteorological Organization*) para o monitoramento de secas meteorológicas (Aghakouchak, 2015; AghaKouchak *et al.*, 2015; Hao e Singh, 2015; Rajsekhar *et al.*, 2015). É usado também para previsão, para análise de frequência, para análise no espaço e no tempo e para avaliação de impactos climáticos (Mishra e Singh, 2010).

Ele utiliza, para o período e local de interesse, uma série temporal mensal longa de precipitação, preferencialmente de, pelo menos, 30 anos de dados (McKee *et al.*, 1993). O conjunto é ajustado

a uma distribuição de probabilidade, para obter-se a relação entre a probabilidade e a precipitação (McKee *et al.*, 1993), utilizado-se, geralmente, a distribuição Gama, sendo também comum a Pearson Tipo III, a Lognormal, a de Valores Extremos e Exponencial (Mishra e Singh, 2010). A distribuição é então transformada em uma distribuição normal padrão (McKee *et al.*, 1993; AghaKouchak *et al.*, 2015), para permitir a comparação entre diferentes localidades e o cálculo de probabilidades de excedência (Keyantash e Dracup, 2002). O valor do quantil Z na distribuição normal padrão, obtido a partir da probabilidade, corresponde ao valor do SPI e, a partir do valor do índice, pode-se obter o início, o encerramento, a intensidade e a magnitude de uma seca (McKee *et al.*, 1993).

Uma sequência de valores negativos do SPI indica uma condição de seca, enquanto os positivos representam um período úmido (Aghakouchak, 2015; AghaKouchak *et al.*, 2015; Banco Mundial, 2015; Brito *et al.*, 2018). Esses valores são usados para definir categorias de secas, conforme mostrado na Tabela 4.2.

Tabela 4.2 – Categorias de seca do SPI
(Fonte: McKee *et al.*, 1993; Steinemann *et al.*, 2005)

Valores do SPI	Categoria de Seca	Frequência Acumulada	Tempo na Categoria
0 a -0,99	Seca leve	16-50 %	~24%
-1,00 a -1,49	Seca moderada	6,8-15,9 %	9,2%
-1,50 a -1,99	Seca severa	2,3-6,7 %	4,4%
-2,00 ou menos	Seca extrema	<2,3%	2,3%
			~40%

Uma fonte de dados do SPI é o IRI/LDEO (disponível em <<http://iridl.ldeo.columbia.edu/maproom/Global/Precipitation/SPI.html>>), que permite a obtenção do índice para escalas temporais de 1, 3, 6, 9 e 12 meses, para diferentes locais. A Figura 4.4 mostra o mapa do SPI de 6 meses, entre agosto de 2017 e janeiro de 2018. Observa-se a indicação de condições secas no Brasil (valores negativos do índice), mas não na categoria mais extrema.

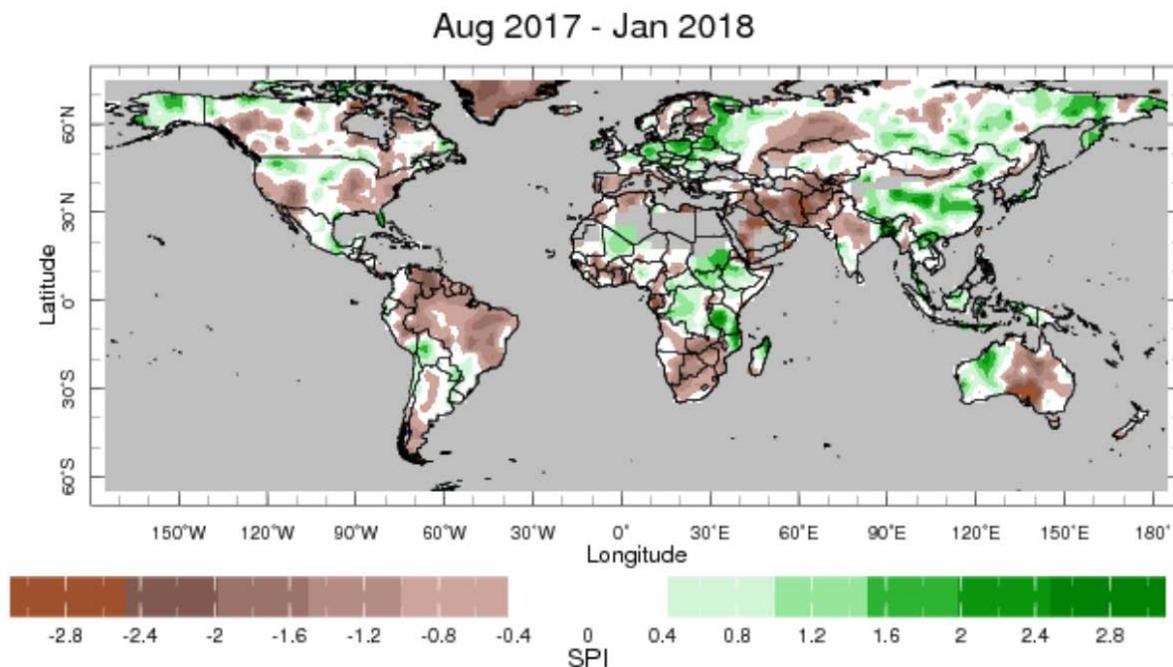


Figura 4.4 - Mapa da América do Sul para o SPI de 6 meses, em agosto/2017-janeiro/2018 (Fonte: IRI/LDEO)

O SPI é usado no monitoramento de secas pelo Centro de Previsão de Tempo e Estudos Climáticos do Instituto Nacional de Pesquisas Espaciais - CPTEC/INPE (disponível em <<http://clima1.cptec.inpe.br/spi/pt>>). A Figura 4.5 mostra o mapa mais recente, de janeiro de 2018, para o SPI na escala temporal de 3 meses. Nota-se a presença de todas as intensidades de seca no Nordeste. Além disso, o SPI é um dos índices usado para compor o Monitor de Secas do Nordeste (Banco Mundial, 2015) e para monitorar as secas pelo CEMADEN.

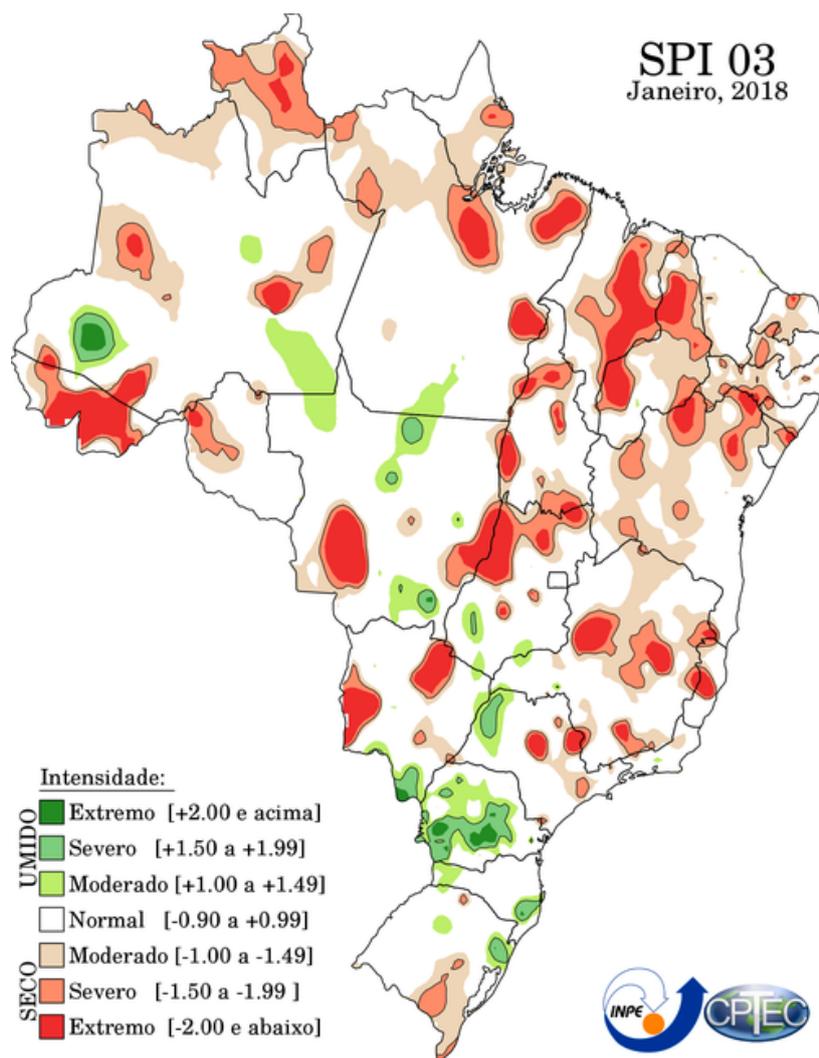


Figura 4.5 - Mapa do Brasil para o SPI de 3 meses, em janeiro de 2018
(Fonte: CPTEC/INPE)

4.1.3.1 - Vantagens

A principal vantagem do SPI é que ele pode ser calculado para várias escalas temporais: 3, 6, 12, 24 e 48 meses (McKee *et al.*, 1993; Steinemann *et al.*, 2005; Mishra e Singh, 2010; Banco Mundial, 2015; Rajsekhar *et al.*, 2015), por meio de uma média móvel. Assim, o índice é capaz de avaliar os impactos do déficit de precipitação em diferentes elementos do ciclo hidrológico (Steinemann *et al.*, 2005; Banco Mundial, 2015). Como a umidade do solo responde a anomalias de precipitação em uma escala curta, o índice consegue monitorar bem a seca agrícola. Já os níveis dos reservatórios, as águas subterrâneas e as vazões respondem mais lentamente, podendo ser avaliados pelo SPI também (Mishra e Singh, 2010; Steinemann *et al.*, 2005), caracterizando uma seca hidrológica.

O SPI também permite a comparação entre vários locais (Rajsekhar *et al.*, 2015; Banco Mundial, 2015), em razão do ajuste a uma distribuição normal (Keyantash e Dracup, 2002). Por ser um índice padronizado, ele reflete a mesma probabilidade de ocorrência em diferentes lugares, períodos e climas (Steinemann *et al.*, 2005). O fato de ser normalizado permite, ainda, por parte do índice, a indicação de períodos tanto secos quanto úmidos (McKee *et al.*, 1993). Além disso, Rajsekhar *et al.* (2015) fez um estudo comparando alguns índices e constatou a habilidade do SPI em prever o início de uma seca.

4.1.3.2 - Limitações

O grande problema apontado pelos autores em relação ao SPI é o fato dele considerar o déficit hídrico com base apenas na variável precipitação, o que pode não ser verdade em muitos casos (Tsakiris e Vangelis, 2005). Assim, ele pode não ser suficiente para uma gestão de secas proativa. É preciso também ter atenção ao tamanho da série temporal e à natureza da distribuição de probabilidade ao calcular o SPI. Quando o índice precisa ser ajustado para escalas temporais maiores, ou seja, acima de 24 meses, o ajuste pode apresentar viés por causa da falta de dados, o que é comum em muitos locais (Mishra e Singh, 2010). Além disso, em climas secos, que apresentam muitos valores nulos de precipitação, pode ser difícil ajustar a distribuição Gama e o SPI em escalas curtas pode não seguir uma distribuição normal, gerando erros (Mishra e Singh, 2010).

4.2 - PREVISÃO DE SECAS

A região Nordeste recebeu vários investimentos, nos Séculos XX e XXI, em obras de infraestrutura, como construção de reservatórios e de estruturas de irrigação, para combater os impactos das secas (Banco Mundial, 2015). Embora essas medidas tenham amenizado os problemas causados por esse desastre climático, o fato de até hoje ocorrerem impactos indica que outras medidas, relacionadas à gestão de eventos extremos, precisam ser tomadas (Banco Mundial, 2015). É preciso então alterar essa atuação reativa para uma proativa, modificando o foco para uma gestão de riscos, que visa trabalhar as vulnerabilidades locais e inclui investir no monitoramento e na antecipação desses eventos, de forma a criar sistemas de alerta, planos de prevenção e programas de resposta operacional (Banco Mundial, 2015). O foco deste trabalho

é tentar contribuir para a previsão de secas e este item traz algumas iniciativas já existentes nesse sentido.

As iniciativas de previsão de seca no Nordeste brasileiro acontecem desde o Século XX, mas o uso de índices de seca para esse fim ainda é pouco explorado. Dentre os trabalhos antigos, que não usam índices, destaca-se Hastenrath *et al.* (1984), que buscaram criar um método de previsão de chuvas para a parte norte do Nordeste, baseado em fatores intervenientes, como temperatura da superfície do mar do Oceano Atlântico, vento e pressão ao nível do mar. Já Marengo e Bernasconi (2015) procuraram prever, por meio de índices de aridez, o avanço da aridez no Nordeste até 2100, constatando uma provável redução na precipitação ao longo dos anos, resultando em uma piora das secas.

Além desses trabalhos, algumas instituições fornecem previsões de variáveis climáticas, disponíveis em seus sites. O Instituto Nacional de Meteorologia – INMET (disponível em <<http://www.inmet.gov.br/portal/>>) traz a previsão trimestral (para o mês atual e para os dois seguintes) de temperatura média e de precipitação. O Instituto Nacional de Pesquisas Espaciais – INPE, por meio do Centro de Previsão de Tempo e Estudos Climáticos – CPTEC (disponível em <<http://www.cptec.inpe.br/>>), fornece previsões globais também sazonais para precipitação e temperatura, além de previsões trimestrais da TSM, visando monitorar a ocorrência do *El Niño* e da *La Niña* e a posição da ZCIT (Zona de Convergência Intertropical). O *International Research Institute for Climate and Society* – IRI (disponível em <<http://iri.columbia.edu/our-expertise/climate/forecasts/>>) fornece previsões também trimestrais globais de precipitação e de temperatura, do ENSO (mostrando condições de *El Niño*, de *La Niña* e neutras) e de anomalias da TSM.

O Centro Nacional de Monitoramento e Alertas de Desastres Naturais – CEMADEN (disponível em <<http://www.cemaden.gov.br/>>) fornece previsões sazonais de precipitação, mostrando, em um mapa, as regiões com chuva acima e abaixo do normal. Esse mapa é acompanhado de um texto, contendo esclarecimentos e possíveis impactos dessa previsão. Essas previsões são feitas com base em uma análise das condições atmosféricas e oceânicas globais e em previsões de modelos dinâmicos e estatísticos. A Fundação Cearense de Meteorologia e Recursos Hídricos - FUNCEME (disponível em <<http://www3.funceme.br/previsao-climatica/>>) realiza previsões sazonais (para os três meses

seguintes) de precipitação, de temperatura e de evapotranspiração para todo o Nordeste. Além disso, a instituição fornece previsões também sazonais para o SPI e para o SPEI.

Nos trabalhos mais atuais, é possível encontrar contribuições para a previsão de índices de seca. Canamary *et al.* (2015) avaliou um sistema de previsão do SPI de 3 meses (SPI3) para os reservatórios Banabuiu, Castanhão e Orós, localizados no estado do Ceará, baseado em modelos atmosféricos globais e regionais: ECHAM4.5 e RSM (*Regional Spectral Model*). Foram utilizados dados observados e previstos para o período de 2007 a 2014, observando a previsão para os meses de janeiro e de fevereiro, por representarem o início da estação chuvosa, a qual se concentra entre fevereiro e maio. A conclusão do estudo foi a de que a previsão do SPI3 com um tempo de previsão de pelo menos 2 meses possui um desempenho superior à climatologia. Lima e AghaKouchak (2017) usaram um modelo de previsão sazonal para o Índice de Severidade de Seca de Palmer (PDSI, *Palmer Drought Severity Index*) na Amazônia. Os autores obtiveram correlações de até 0,60 e de 0,45, para os tempos de antecedência de 5 e de 9 meses, respectivamente, entre o PDSI médio observado e previsto.

Delgado *et al.* (2017) testaram modelos de previsão de secas para uma região no Semiárido nordestino. Foram usadas as previsões meteorológicas fornecidas pela Fundação Cearense de Meteorologia e Recursos Hídricos (FUNCEME) e pela *European Center for Medium-Range Weather Forecasts* (ECMWF), como modelos dinâmicos de previsão sazonal. Foram utilizados o SPI e o Índice Padronizado de Precipitação-Evapotranspiração (SPEI, *Standardized Precipitation Evapotranspiration Index*), em diferentes escalas temporais, para quantificar secas meteorológicas, além de algumas métricas para mensurar a seca hidrológica (*Surface Water Supply Index – SWSI*, volume do reservatório dividido pelo volume máximo, variação mensal do volume do reservatório). Os índices de seca hidrológica foram previstos por meio de um modelo linear generalizado, que usou os índices de seca meteorológica como preditores. O desempenho de períodos com baixa precipitação e das secas feitas com o SPI1 foram baixos, enquanto os do armazenamento relativo de reservatório e o SPEI1 foram bons. Eles concluíram que é viável prever secas com o modelo múltiplo em escalas temporais importantes para a gestão, mas que não é possível obter um bom desempenho em todos os elementos testados.

Focando agora em uma iniciativa internacional, os Estados Unidos mantêm um centro de previsão do clima (*Climate Prediction Center – CPC*), ligado ao *National Oceanic and Atmospheric Administration* (NOAA) (disponível em

<<http://www.cpc.ncep.noaa.gov/products/Drought/>>). Eles monitoram o sistema climático como um todo, incluindo oceano e atmosfera, e as secas (por meio do USDM, vide item 4.1.2) e fornecem previsões de precipitação, de temperatura e de secas. Eles fornecem mapas com as tendências das secas (se elas irão persistir ou não) para o mês seguinte e também a média para os três meses seguintes (previsão sazonal) para todo o País. O CPC fornece, ainda, previsões semanais, mensais e de três meses da precipitação e da temperatura para todas as partes do mundo.

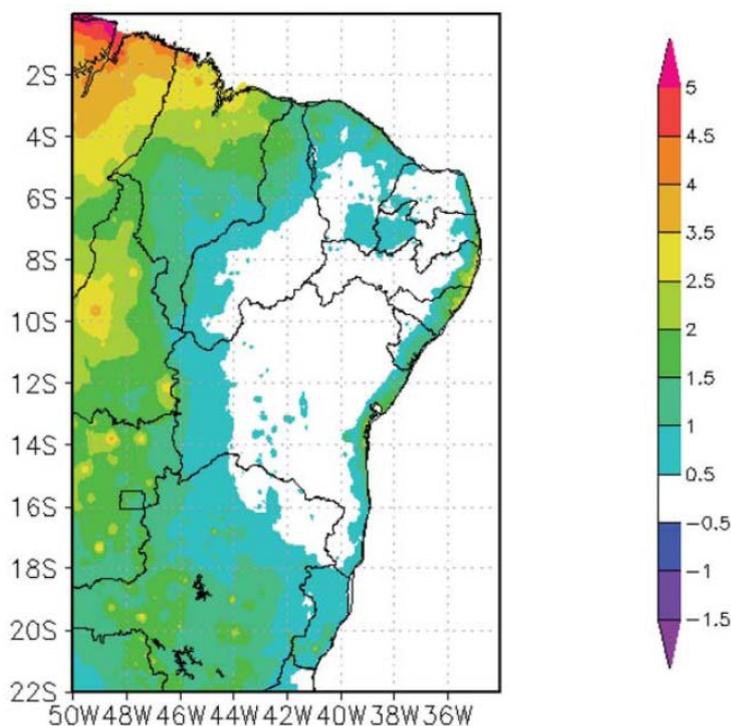
5 - ÁREA DE ESTUDO

5.1 - CARACTERIZAÇÃO DA REGIÃO NORDESTE

5.1.1 - Caracterização geral

A região Nordeste do Brasil está localizada aproximadamente entre 1°S e 18°S e 35°W e 47W° (Rao *et al.*, 1993; Uvo *et al.*, 1998; Kousky e Chu, 1978), constituindo uma área de aproximadamente um milhão e meio de quilômetros quadrados (CGEE e ANA, 2012; Marengo e Bernasconi, 2015). Trata-se de uma região densamente povoada, habitada por uma estimativa de 46.995.094 habitantes (IBGE). Podem-se distinguir três climas diferentes distribuídos ao longo da região. O clima tropical é predominante no litoral, no sul da Bahia e no centro do Maranhão, sendo caracterizado pela presença de uma estação chuvosa bem definida. O clima equatorial úmido encontra-se em uma área no oeste do Maranhão, apresentando altas temperaturas e grande pluviosidade. Por fim, tem-se o clima semiárido, característico da região central do Nordeste (sertão nordestino), no qual a precipitação é baixa e irregular, gerando períodos de estiagem (adaptado de CGEE e ANA, 2012).

Existe uma região no Nordeste que é mais regularmente afetada pelas secas, conhecida como Polígono das Secas. Ela possui ainda uma sub-região semiárida conhecida como Sertão, que abrange uma área de cerca de 912.200 km² (Marengo e Bernasconi, 2015), caracterizada pela presença de rios intermitentes e de um solo pedregoso e pobre (CGEE e ANA, 2012). A existência dessa região semiárida foi explicada por Moura e Shukla (1981) com base em alguns fatores. Primeiramente, está localizada na vizinhança de uma inversão de ventos alísios e, em razão de sua geografia, uma parte de sua área encontra-se na zona de influência permanente da alta subtropical do Atlântico Sul. O albedo diferencial e os efeitos locais da topografia também influenciam nos padrões climáticos. Outros elementos que podem contribuir são a existência na região de solos rasos, que possuem pouca capacidade de armazenamento hídrico, e de altas taxas de evaporação, em combinação com o fato das precipitações serem concentradas em uma época do ano (CGEE e ANA, 2012). A Figura 5.1 apresenta o balanço anual entre precipitação e evaporação no Nordeste, mostrando que, na média, essas duas variáveis igualam-se na parte central da região (CGEE e ANA, 2012), que é correspondente à área mais afetada pelos eventos de seca.



GrADS: OOLA/IGES

Figura 5.1 – Média anual do balanço entre a precipitação e a evaporação no Nordeste do Brasil (mm/dia) (Fonte: Projeto PROCLIMA, INPE/CPTEC *apud* CGEE e ANA, 2012)

5.1.2 - Padrões de precipitação

A precipitação anual é espacialmente e temporalmente muito diversa, variando, em média, entre 600 e 2000 mm (Rao *et al.*, 1993; Uvo *et al.*, 1998). A média da parte amazônica do Maranhão é de aproximadamente 1.800 mm e, no Semiárido, de 600 mm (CGEE e ANA, 2012), podendo ser inferior a 400 mm em algumas áreas (Kousky, 1979). A sequência das estações chuvosas é influenciada por vários sistemas climáticos e atmosféricos, capazes de favorecer ou inibir a precipitação, com destaque especial para as temperaturas na superfície do mar (TSM) nos Oceanos Atlântico e Pacífico (CGEE e ANA, 2012). O Nordeste pode ser dividido em diferentes regiões de acordo com o comportamento da estação chuvosa: parte leste ou costeira, região mais ao sul, região mais ao norte e parte central ou semiárida. A Figura 5.2 mostra a média anual da precipitação entre 1961 e 1990 na América do Sul e no Atlântico Tropical Sul. Pode-se observar que a média é maior no litoral do Nordeste e no Maranhão (1.200 a 1.500 mm/ano) do que na parte central (média de 900 mm/ano). Existem mecanismos que podem influenciar no Nordeste como um todo, como os Vórtices Ciclônicos de Altos Níveis e as oscilações de Madden e Julian (CGEE e ANA, 2012), mas, em geral, a estação chuvosa de cada uma dessas partes é regulada por mecanismos específicos.

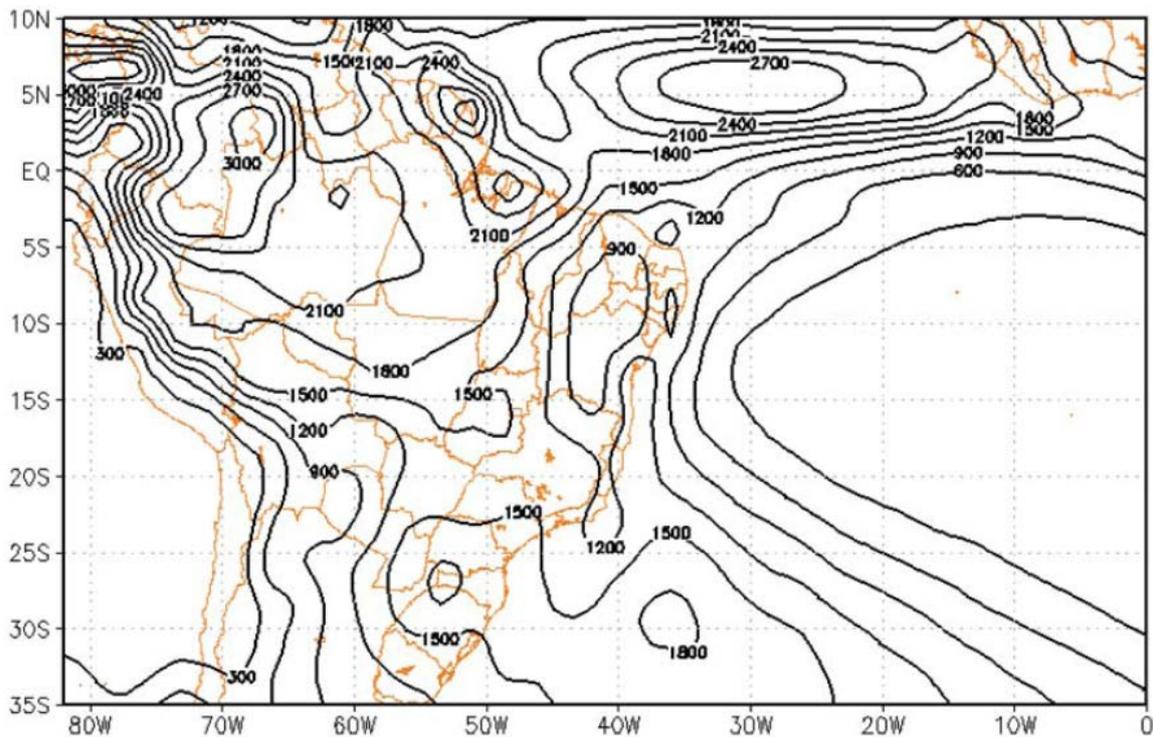


Figura 5.2 – Precipitação média anual para a América do Sul e o Atlântico Tropical Sul (em mm/ano) (Fonte: CGEE e ANA, 2012)

A parte leste do Nordeste possui estação chuvosa entre os meses de maio e julho (Rao *et al.*, 1993; Kousky, 1980), com pico em junho e julho (CGEE e ANA, 2012). Isso ocorre devido à propagação na direção oeste de sistemas de nuvens (Kousky, 1980; Rao *et al.*, 1993). Além disso, de acordo com Kousky (1979), frentes frias penetram a partir da parte sul e influenciam na precipitação ao longo da costa durante o outono e o inverno. Kousky (1980) investigou a influência dos sistemas de ventos na precipitação ao longo da costa. Ele constatou que haveria um vínculo direto com a intensidade da brisa terrestre, a qual se relaciona com a diferença entre a temperatura na superfície do mar (TSM) e a temperatura média do ar. Nos meses de outono e inverno, correspondentes à estação chuvosa, a TSM está mais quente do que o ar, o que favorece esse sistema. Quanto maior a diferença, mais favorável é a ocorrência de chuvas. O autor também verificou que a precipitação apresenta um máximo noturno ao longo da costa e um máximo diurno no interior até 300 km, em razão da influência do fluxo médio e da brisa terra-mar.

A variação dos ventos costeiros, relacionada com a intensidade e com o posicionamento da alta subtropical no Atlântico Sul (sistema de alta pressão), também parece ter influência, de acordo com Rao *et al.* (1993). Eles concluíram que as chuvas de outono e de inverno na região associam-se à existência de ventos alísios de sudeste soprando de forma perpendicular à costa, informação que também foi citada em CGEE e ANA (2012). Rao *et al.* (1993) acreditam também que uma TSM maior no Atlântico Sul aumenta a evaporação, podendo provocar anos mais úmidos. Mecanismos atmosféricos provenientes da África parecem também ter influência (CGEE e ANA, 2012). Talvez em razão da existência de tantos mecanismos, mesmo durante a estação seca, a parte leste recebe, geralmente, mais de 500 mm de precipitação (Costa *et al.*, 2016).

A porção mais ao sul, formada por parcelas dos estados da Bahia, do Maranhão e do Piauí, apresenta precipitação entre novembro a março, com pico em dezembro (CGEE e ANA, 2012; Moura e Shukla, 1981). Essas chuvas são causadas pelo movimento equatorial de frentes frias (Kousky, 1979; CGEE e ANA, 2012). Conforme já mencionado, de acordo com Kousky (1979), essas frentes penetram a partir da parte sul e influenciam também na precipitação ao longo da costa da Bahia até a do Rio Grande do Norte, durante o outono e o inverno. A Zona de Convergência do Atlântico Sul (ZCAS) também pode ser relacionada à precipitação nessa região (CGEE e ANA, 2012).

A região mais ao norte apresenta chuvas entre janeiro e abril, com máximo em março e abril (Kousky, 1979; Hastenrath *et al.*, 1984; CGEE e ANA, 2012), época em que a zona do cavado intertropical (Hastenrath e Heller, 1977), o eixo de confluência próximo ao Equador e a banda de convergência associada atingem suas posições mais ao sul (Hastenrath *et al.*, 1984; Hastenrath e Heller, 1977). A Zona de Convergência Intertropical (ZCIT) é também um dos principais mecanismos de formação de precipitação, estando ligada diretamente à TSM no Atlântico Equatorial (CGEE e ANA, 2012). Nesses meses, o contraste entre as águas quentes equatoriais do Norte e as frias do Sul fica mais fraco e, além disso, a TSM do Atlântico Sul equatorial está mais alta, “aumentando a umidade e o fluxo atmosférico” até o Nordeste (Hastenrath *et al.*, 1984). Durante os meses de outubro a novembro, a região não recebe precipitação, caracterizando a estação seca (Costa *et al.*, 2016).

Já a região semiárida, localizada no interior, possui chuvas máximas entre fevereiro e maio (Uvo *et al.*, 1998), especialmente em março e abril (Moura e Shukla, 1981; Hastenrath *et al.*,

1984), com uma variabilidade interanual alta, podendo chegar a 40% da média (Uvo *et al.*, 1998). A explicação é que essa é a época do ano em que a Zona de Convergência Intertropical (ZCIT), o cavado (de baixa pressão) próximo ao Equador e a confluência (dos ventos) associada ocupam posições mais ao sul. Além disso, de acordo com Hastenrath (1990), as águas do Atlântico Sul equatorial estão mais quentes, o que aumenta a umidade e a instabilidade no fluxo que penetra no Nordeste. O contraste entre a TSM do Atlântico Norte tropical (mais quente) com a TSM das águas mais ao sul, que induz uma circulação termal sobre o Nordeste, inibindo a precipitação, está mais fraca. A precipitação anual total varia entre 300 e 800 mm e é caracterizada por uma estação seca entre junho e setembro, iniciada a partir de maio e finalizada em outubro (Hastenrath, 1990).

5.2 - SECAS NO NORDESTE

5.2.1 - Histórico e principais impactos

Em razão da presença de solos rasos, de um embasamento cristalino (CGEE e ANA, 2012), de uma alta evaporação (Marengo *et al.*, 2013), do regime de precipitação concentrado em apenas alguns meses do ano (CGEE e ANA, 2012) e da variabilidade interanual alta da precipitação (Uvo *et al.*, 1998; Kousky, 1979; CGEE e ANA, 2012), a flutuação climática é muito alta, provocando os anos de seca, especialmente na região semiárida (CGEE e ANA, 2012), sendo comum o registro de ocorrência desse desastre desde o Século XVI. De 1559, quando foi registrada uma seca na Bahia, até 2011, haviam ocorrido cerca de 72 secas no Nordeste, das quais 40 foram anuais e 32, plurianuais, fazendo com que 112 dos 452 anos em questão fossem de seca (CGEE e ANA, 2012). A Tabela 5.1 mostra os anos de ocorrências das principais secas, enquanto a Tabela 5.2 cita impactos de alguns desses eventos. Observa-se que o número de secas por século aumenta com a maior ocupação das terras pelo ser humano (CGEE e ANA, 2012). A Figura 5.3 mostra as anomalias de precipitação no Nordeste para os meses de fevereiro a março, de 1951 a 2015. Os picos negativos indicam os anos de secas, em coerência com a Tabela 5.1.

Tabela 5.1 – Principais registros de secas no Nordeste
(Fonte: Moura e Shukla, 1981; CGEE e ANA, 2012; Marengo e Bernasconi, 2015; Banco Mundial, 2015)

1559	1603	1707-1711	1804	1900	2001-2002
1564	1614	1711	1808	1902-1903	2005
1583	1624	1721	1809	1907	2007
1587	1692	1723-1724	1810	1915	2010
		1730	1814	1919-1920	2012-2018
		1736-1737	1816-1817	1932-1933	
		1744-1746	1819-1820	1936	
		1754	1824-1825	1941-1944	
		1760	1827	1951-1953	
		1766-1767	1830-1833	1958	
		1772	1845-1846	1966	
		1776-1777	1877-1879	1970	
		1777-1780	1888-1889	1972	
		1784	1891	1976	
			1898	1979-1983	
				1987	
				1990-1993	
				1998	

Tabela 5.2 – Impactos de algumas secas no Nordeste brasileiro
(Fonte: Marengo *et al.*, 2013; Marengo *et al.*, 2016; Hastenrath e Heller, 1977)

Seca	Impacto
1777-1780	Morte de quase 85% do gado Morte de metade da população por fome
1877-1879	Morte de quase 200.000 pessoas por fome e doenças
1907	Afetou principalmente os estados do Ceará e do Rio Grande do Norte
1915	Morte de mais de 278.000 pessoas e emigração de cerca de 75.000 no estado do Ceará
1958	Emigração de cerca de 10 milhões de pessoas
1979-1981	Redução maior de 70% na produção de arroz, feijão e algodão Aumento em cerca de 100% do preço dessas culturas
1982-1983	Perda de 80% do gado
1998	Perda de 57% da produção agrícola Perdas econômicas de 5% do Produto Interno Bruto (PIB) da região Afetou 10 estados

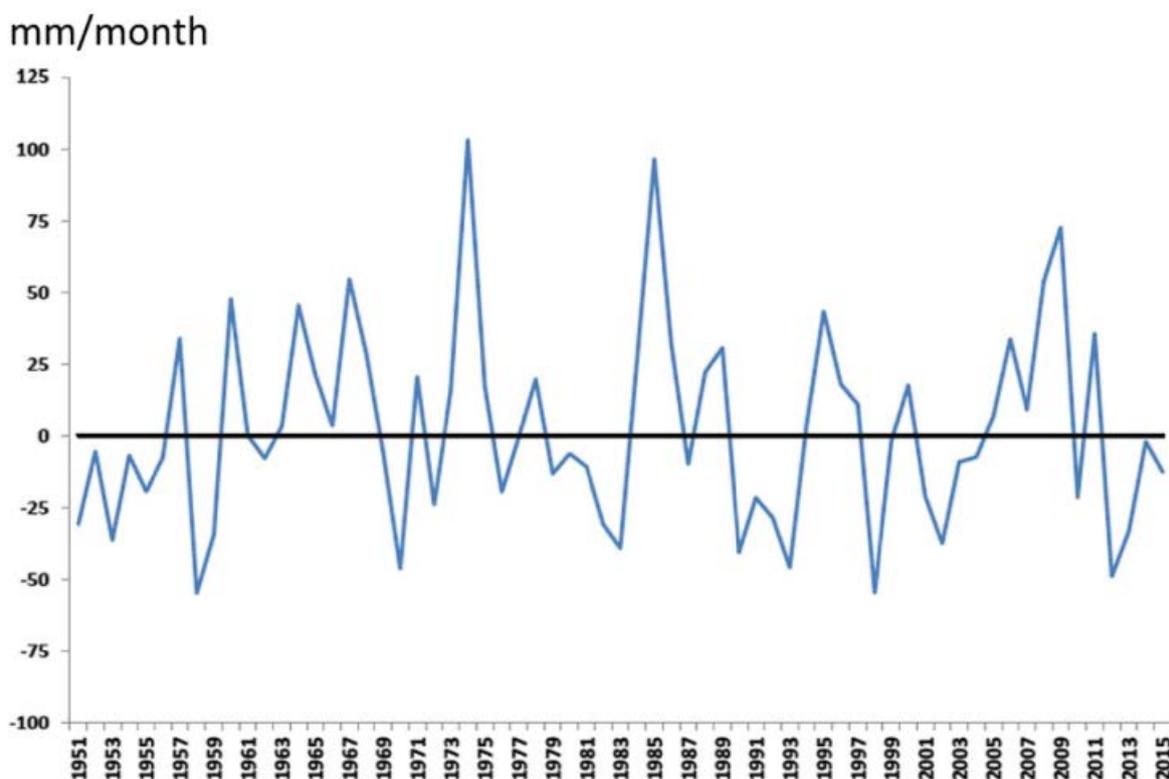


Figura 5.3 – Anomalias de precipitação para o Nordeste brasileiro, de 1951 a 2015 (Fonte: Marengo *et al.*, 2016)

Durante séculos, as secas no Nordeste causaram vários impactos ambientais, sociopolíticos e econômicos. Dentre os impactos ambientais, pode-se citar prejuízos às plantas, aos animais e aos peixes, alteração na qualidade da água e do ar, aumentando a quantidade de poeira e de poluentes, perda do solo (por erosão, salinização, alcalinização), além de alterações na paisagem (CGEE e ANA, 2012). Os impactos sociopolíticos podem incluir problemas de saúde (como, por exemplo, associados ao uso de águas provenientes de açudes ou poços, com qualidade inferior), impacto na vida da população afetada (CGEE e ANA, 2012), mortes humanas por fome, desnutrição e doenças e migração de pessoas (Marengo *et al.*, 2016). Os danos econômicos, por sua vez, podem ser representados por perdas na produção agrícola (CGEE e ANA, 2012; Marengo *et al.*, 2016), na produção de carne de laticínios, na produção de madeira, nos lucros com recreação, além de uma desorganização das atividades econômicas existentes (CGEE e ANA, 2012). As perdas monetárias também são altas: as secas de 1958, de 1970, de 1976 e de 1979-1983 custaram para o governo, respectivamente, US\$ 803 milhões, US\$ 430 milhões, US\$ 447 milhões e US\$ 7,8 bilhões, tendo sido a última possivelmente a mais cara (Marengo *et al.*, 2016).

Os impactos relacionados aos eventos ocorridos do Século XVI até a segunda metade do Século XVII foram concentrados mais na agricultura e na pecuária, mas, com o aumento da ocupação populacional, os impactos foram assumindo um caráter mais social e econômico (Banco Mundial, 2015). A seca de 1777-1778 gerou uma perda de 7/8 dos rebanhos no estado do Ceará (Banco Mundial, 2015). A seca de 1877-1879 causou impactos significantes na economia, além da morte de animais e de pessoas (CGEE e ANA, 2012). As secas de 1915, 1932, 1958 e 1970 também levaram prejuízos para a população residente do Polígono das Secas (CGEE e ANA, 2012). No caso da seca de 1970, o PIB das atividades agropecuárias caiu 29,62% em relação a 1969 (CGEE e ANA, 2012). Uma taxa de crescimento negativa do PIB da agropecuária foi observada em outros anos de seca, como, por exemplo: 1987-1990 (-9,48%), 1990-1991 (-10,01%), 1991-1992 (-0,74%), 1992-1993 (-12,30%), 1997-1998 (-13,30%) (CGEE e ANA, 2012). A seca de 1979-1983 foi a que afetou o maior número de pessoas na região Semiárida (CGEE e ANA, 2012).

Com o crescimento populacional elevado do Século XX, houve grande investimento em infraestrutura para lidar com os eventos de seca (CGEE e ANA, 2012). A seca de 2010, por exemplo, foi meteorológica e ocorreu principalmente no Ceará, com precipitações 47% abaixo da média, tendo sido causada principalmente pelo *El Niño* (CGEE e ANA, 2012). A safra foi 56,17% que a de 2009, tendo o Governo decretado situação de emergência em 79 municípios, mas essa seca foi pouco percebida pela população afetada, já que não houve falta de água, em razão dos investimentos feitos pelo Governo para minimizar os impactos, como o uso de caminhões-pipa e cisternas (CGEE e ANA, 2012).

De acordo com Marengo *et al.* (2013) e Marengo e Bernasconi (2015), a seca de 2012-2013, que se estende até 2018, está sendo a mais severa das últimas décadas, afetando, até 2014, cerca de 38% da população da região semiárida (aproximadamente 9 milhões de pessoas), destruindo algumas terras cultivadas e levando alguns estados a declarar calamidade pública. Ela começou quando houve um evento de *La Niña* no verão do hemisfério sul. Até então, a seca de 1998 tinha sido a pior da história mais recente, em que as precipitações ficaram cerca de 48% abaixo do normal (CGEE e ANA, 2012). Esses dois eventos, entretanto, não causaram mortes, diferentemente da pior seca da história nordestina, a de 1877, que matou milhares de pessoas.

Ainda hoje as secas no Nordeste têm como impacto a escassez de água, gerando risco à segurança hídrica e energética, afetando, principalmente, a parcela mais pobre da população

(CGEE e ANA, 2012). Ocorre também uma degradação do solo, que pode afetar a agricultura e, a longo prazo, pode transformar a região semiárida em árida, gerando um risco de desertificação (CGEE e ANA, 2012; Marengo e Bernasconi, 2015). Desde os anos 1950, o governo passou a desenvolver programas contra as secas, como a construção de reservatórios e de canais, além da instituição de programas sociais para ajudar as pessoas afetadas (CGEE e ANA, 2012; Marengo *et al.*, 2013; Marengo *et al.*, 2016). Assim, desde os anos 70 não se registram mais mortes causadas pelas secas, embora a migração de pessoas ainda ocorra (Marengo *et al.*, 2013; Marengo *et al.*, 2016), bem como a fome. Atualmente, a população e os governos lidam com esse desastre por meio de medidas adaptativas, como o armazenamento de água em cisternas, fundamental para garantir a dessedentação dos animais e a disponibilidade hídrica para os diversos usos (especialmente abastecimento público e irrigação), e o uso de caminhões pipa, para manter a agricultura de pequena escala e garantir a subsistência das populações, além de investimento em irrigação (CGEE e ANA, 2012). A importância da construção de reservatórios está em armazenar água em um ano chuvoso para ser usada durante os períodos de escassez.

Além dos impactos, também é importante entender a abrangência espacial das secas, visando conhecer as áreas mais susceptíveis à ocorrência desse desastre natural. A Figura 5.4 mostra as principais áreas de ocorrência de seca no Nordeste, em porcentagem. As regiões litorâneas, ao sul e a oeste são as menos afetadas, enquanto a parte central é mais atingida. Observa-se uma ocorrência mais frequente das secas no Ceará do que nos demais Estados (CGEE e ANA, 2012).

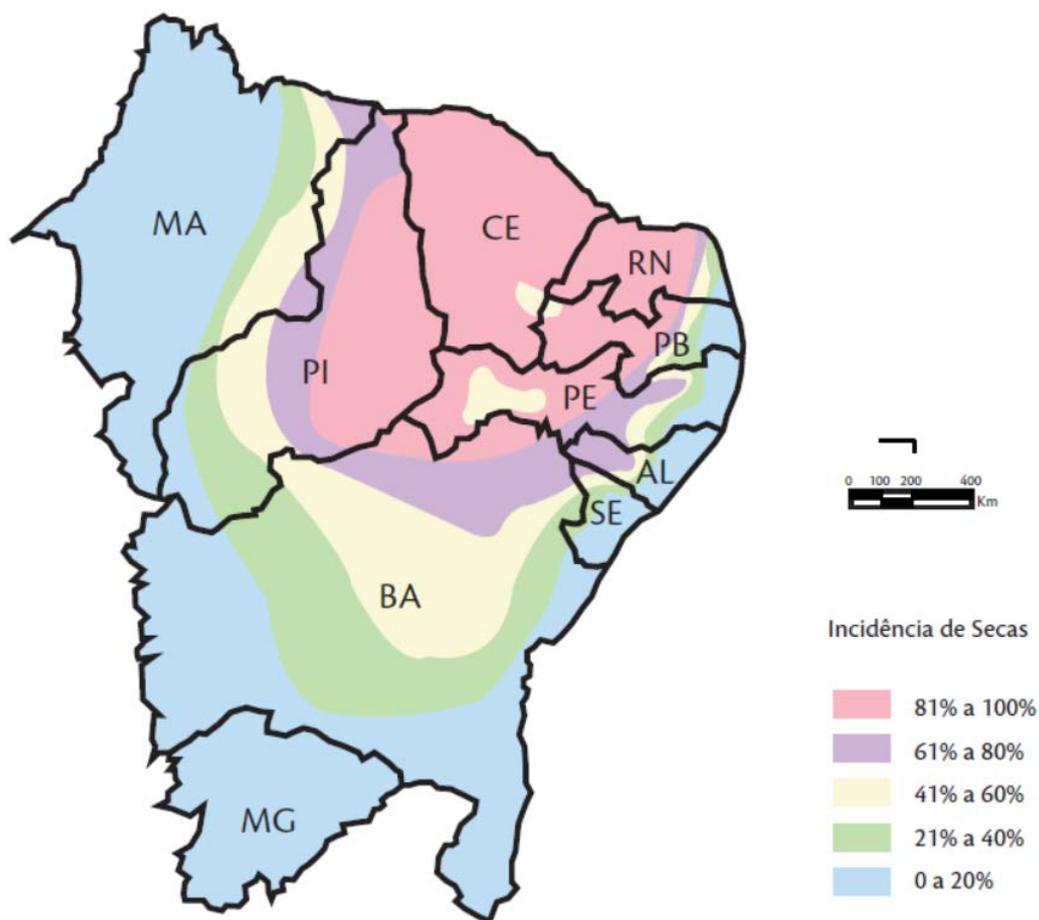


Figura 5.4 – Áreas de incidência de secas no Nordeste brasileiro, em porcentagem (Fonte: CGEE e ANA, 2012)

5.2.2 - Mecanismos de formação das secas no Nordeste

Vários autores investigaram, ao longo dos anos, mecanismos globais de circulação atmosférica e oceânica que poderiam causar as secas no Nordeste. A relação entre a Oscilação Sul e as chuvas na região foram estudadas por Walker (1928) *apud* Moura e Shukla (1981), estando as anomalias de precipitação associadas com a oscilação de baixa frequência, o que faz com que o cavado (de baixa pressão) próximo ao Equador seja deslocada para o Norte (Hastenrath, 1990). Hastenrath e Heller (1977) igualmente relacionaram, a posição latitudinal e a força da zona do cavado com o desastre natural, caracterizando a seca de 1958. Hastenrath *et al.* (1984) e Hastenrath e Heller (1977) também apontaram o deslocamento do cavado para o norte como uma causa de secas, em associação com uma expansão na direção do Equador da alta do Atlântico Sul e a uma retração na direção polar da alta do Atlântico Norte.

Hastenrath *et al.* (1984) apontaram como uma das causas de estações chuvosas mais secas do que o normal no norte do Nordeste o posicionamento mais ao norte, característico da estação mais seca, do eixo de confluência próximo ao Equador e da banda de convergência associada. Este último fenômeno relaciona-se a anomalias da pressão ao nível do mar, com pressões altas no Atlântico Sul e baixas no Atlântico Norte (Hastenrath, 1978).

Anos secos também são caracterizados por uma posição mais ao norte da Zona de Convergência Intertropical (ZCIT) (Hastenrath, 1990), reduzindo a convergência de umidade sobre a região. Durante esses anos, a TSM apresenta anomalias positivas no Atlântico Norte e negativas no Atlântico Sul, provocando uma célula de circulação meridional na atmosfera com subsidência sobre o Nordeste (Hastenrath, 1990; CGEE e ANA, 2012). Durante anos chuvosos, esse padrão é invertido (Hastenrath e Heller, 1977; Hastenrath *et al.*, 1984; Hastenrath, 1990; CGEE e ANA, 2012). Esses resultados também foram encontrados por Moura e Shukla (1981). De acordo com os autores, a presença desse padrão de temperatura provocaria um movimento ascendente e vorticidade ciclônica no norte e um momento descendente e vorticidade anticiclônica no sul, sobre o Nordeste. Essa relação poderia ser atribuída à posição da ZCIT, com sua intensificação no norte do Equador, causada por esse dipolo de anomalia de temperatura. Moura e Shukla (1981) também afirmaram que, quando a TSM no Atlântico Sul está mais fria, a evaporação no oceano é menor, fazendo com que os ventos alísios levem menos umidade para o Nordeste.

A ocorrência simultânea entre secas no Nordeste e eventos *El Niño* foi observada por Rao *et al.* (1993) e por Marengo *et al.* (2016). O *El Niño* é um fenômeno caracterizado por um aquecimento acima da média da TSM na região do Pacífico Equatorial Leste (Curtis, 2008), estando ligado à variabilidade no gradiente de pressão horizontal na região, também conhecido como Oscilação Sul (*El Niño* Oscilação Sul - ENOS) (Hastenrath e Heller, 1977). A anomalia positiva da TSM na região altera os padrões de pressão na superfície do mar, modificando a célula de *Walker* (Curtis, 2008), que se desloca e passa a ter uma subsidência sobre a Amazônia e parte do Nordeste (Kayano e Andreoli, 2004; CGEE e ANA, 2012), o que pode inibir a convecção atmosférica e causar (ou acentuar) secas. Os anos 1983, 1993, 1998, 2002 e 2015 foram de *El Niño* e a precipitação foi abaixo do normal no Nordeste (Marengo *et al.*, 2016; CGEE e ANA, 2012). Quando a TSM no pacífico está abaixo do normal, tem-se o fenômeno *La Niña*, que gera um efeito oposto, ou seja, de aumento da precipitação na Amazônia e no Nordeste (CGEE e ANA, 2012), no entanto, as secas de 1907 e de 2012-2013 ocorreram em

anos de *La Niña* (Marengo *et al.*, 2013), sugerindo que o ENSO não seja um mecanismo definitivo para a determinação das secas na região.

As secas no Nordeste também podem ter como causa fenômenos mais distantes. Namias (1972) relacionou a variação interanual da precipitação em Quixeramobim com a atividade ciclônica próxima à ilha de Terra Nova, pertencente ao Canadá, no inverno e primavera no Hemisfério Norte. Quando essa atividade é mais intensa do que o normal, ocorre um realce da circulação da célula de Hadley, aumentando os ventos alísios de nordeste e de sudeste, sendo os últimos responsáveis por um aumento da precipitação no Nordeste (Namias, 1972). De acordo com CGEE e ANA (2012), as anomalias de circulação atmosférica entre novembro e janeiro em Terra Nova e na Groelândia, que fazem parte da Oscilação do Atlântico Norte (*North Atlantic Oscillation* - NAO), influenciam nas precipitações entre março e maio no Nordeste.

Além dos mecanismos citados, pode-se mencionar outros que estariam ligados a uma variabilidade de precipitação em menores escalas espaciais e temporais, como as oscilações de Madden e Julian e distúrbios oriundos da África (CGEE e ANA, 2012). Os mecanismos mais associados à ocorrência de secas no Nordeste são, entretanto, o contraste da TSM no Atlântico Equatorial e o *El Niño*, sendo que este funcionaria como uma variável moduladora da ocorrência de períodos secos.

6 - METODOLOGIA

Para representar e caracterizar as secas meteorológicas no Nordeste do Brasil, foi escolhido o Índice de Precipitação Padronizada (SPI, *Standardized Precipitation Index*), por já ter sido explorado em trabalhos de outros autores e por já ser usado em sistemas de monitoramento. A primeira etapa metodológica consistiu em obter esse índice, além de variáveis climáticas associadas ao fenômeno das secas.

Realizou-se então uma avaliação inicial das dinâmicas espacial e temporal do SPI para o Nordeste. As variações desse índice no espaço e no tempo foram analisadas pela decomposição em modos de variabilidade, via análise das componentes principais. Utilizando as componentes principais obtidas, foi feita uma análise de extremos, com o objetivo de obter uma ideia das secas mais severas ocorridas durante o período considerado e a porcentagem total da área do Nordeste atingida por cada um desses eventos. Essa análise é importante para contribuir na avaliação da capacidade do SPI em detectar e descrever as secas históricas. A partir do resultado da análise de extremos, foi feita uma análise compósito, para estudar a influência das variáveis climáticas temperatura e precipitação nos padrões de secas. Por fim, foi realizada uma análise de correlação cruzada, visando identificar a relação entre a ocorrência de secas no Nordeste e indicadores climáticos representativos da temperatura da superfície do mar (TSM) global.

A partir desse conhecimento inicial, foi desenvolvido um modelo de previsão sazonal para as secas meteorológicas no Nordeste, caracterizadas por meio do SPI. A técnica SCCA, que é uma variação da CCA tradicional, foi utilizada para obter uma variável preditora e uma variável resposta altamente correlacionadas e, assim, construir um modelo sazonal de previsão. A TSM foi utilizada como covariável climática preditora do SPI para o Nordeste do Brasil, de forma a obter previsões com antecedências variando de 1 a 6 meses.

Esta metodologia utilizou os *softwares Matlab e R* para manipular os dados. A seguir, cada uma das etapas metodológicas descritas é detalhada: obtenção dos dados (item 6.1); Análise das Componentes Principais (item 6.2); análise de extremos e de áreas (item 6.3); análise compósito (item 6.4); análise de correlação cruzada (item 6.5); Análise de Correlação Canônica tradicional (item 6.6); Análise de Correlação Canônica esparsa (item 6.7); e modelo de previsão (item 6.8).

6.1 - DADOS

6.1.1 - SPI

O SPI foi calculado, na escala temporal de 3 meses (SPI3), a partir de dados de precipitação, para o período de 1980 a 2013, na escala mensal (totalizando 408 meses), sobre a área da região Nordeste do Brasil (entre 1,375°S e 18,125°S e 34,875°O e 48,125°O). Esses dados foram disponibilizados por Xavier *et al.* (2016), em uma grade de dados climáticos interpolados para todo o Brasil, na resolução espacial de 0,25° x 0,25°. Eles são obtidos a partir 3625 pluviômetros e 735 estações meteorológicas em todo o País, sendo fornecidos por diferentes fontes: Instituto Nacional de Meteorologia (INMET), Agência Nacional de Águas (ANA) e Departamento de Águas e Energia Elétrica de São Paulo (DAEE). A distribuição de probabilidades utilizada para o cálculo do SPI foi a empírica, realizada com base nos postos dos dados.

6.1.2 - Dados Climáticos

A fonte de dados (Xavier *et al.*, 2016) descrita no item anterior (6.1.1) foi utilizada para obter, além de dados de precipitação, dados também de temperatura, para o mesmo período em que o SPI foi calculado, também sobre a área da região Nordeste. Esses dados foram usados para obter anomalias mensais de precipitação e de temperatura, subtraindo, em cada mês, a média de cada valor observado. A partir desses valores, foram calculadas as anomalias médias dessas duas variáveis apenas nos meses detectados como mais extremos pela análise de extremos (item 6.3).

6.1.3 - Índices Climáticos

O principal fator responsável pela precipitação no Nordeste, como mencionado nos itens 5.1.2 e 5.2.2, é a Zona de Convergência Intertropical (ZCIT), cuja posição e intensidade estão ligadas a um gradiente ou dipolo de temperatura do Atlântico Tropical (Huang *et al.*, 2005). Esse dipolo é usualmente definido como a diferença entre a TSM no Atlântico Norte tropical (TNA, *Tropical Northern Atlantic Index*) e a TSM no Atlântico Sul tropical (TSA, *Tropical Southern Atlantic Index*). Quando esse gradiente é positivo (temperatura ao norte mais alta do que ao sul), a ZCIT é deslocada para o norte, reduzindo a precipitação sobre o Nordeste (Huang *et al.*, 2005). Hastenrath e Heller (1977), Hastenrath *et al.* (1984), Moura e Shukla (1981), Hastenrath

(1990) e Rao *et al.* (1993) relacionaram esse mecanismo climático à ocorrência de secas na região. Em geral, define-se o índice TNA como a média espacial da TSM na região 5,5°N a 23,5°N e 15°O e 57,5°O, enquanto o índice TSA é definido pela média espacial da TSM em 20°S e 0° e 30°O e 10°L (Hasanean e Almazroui, 2017). O índice dipolo pode então ser calculado como a diferença entre a TNA e a TSA.

Como mencionado no item 5.2.2, o *El Niño* (*El Niño* Oscilação Sul - ENOS) ocorre quando a TSM no Oceano Pacífico equatorial leste encontra-se acima do normal, alterando a circulação da célula de *Walker* (Curtis, 2008) e inibindo a ocorrência de chuvas no Nordeste, resultando, conseqüentemente, na ocorrência de secas. Rao *et al.* (1993) e Marengo *et al.* (2016) observaram a ocorrência, ao mesmo tempo, de secas na região e do fenômeno *El Niño*, que atua como uma variável moduladora da ocorrência desses períodos secos. Costa *et al.* (2016) também usaram o SPI e o ENOS para verificar a ocorrência de secas no leste do Nordeste. O ENOS é, em geral, representado pelo índice NINO3.4, definido como a média espacial da TSM na área entre 5°S e 5°N e 170°O e 120°O (Kaplan *et al.*, 1998), já que essa região concentra uma extensa parte das águas do Pacífico leste equatorial (Guo *et al.*, 2011).

Os índices climáticos dipolo do Atlântico Tropical e NINO3.4 foram obtidos para o mesmo período do SPI, tendo sido calculados a partir dos dados de TSM fornecidos por Kaplan *et al.* (1998) e por Reynolds *et al.* (2002). A influência desses índices sobre as secas no Nordeste foi avaliada por meio de uma análise de correlação cruzada entre o SPI e os índices climáticos, conforme descrito no item 6.5.

6.1.4 - Temperatura na Superfície do Mar (TSM)

Dados globais de TSM foram obtidos a partir dos dados de *reanalysis ERA Interim* na escala mensal, para o mesmo período do SPI, e estão disponíveis em <<http://apps.ecmwf.int/datasets/data/interim-full-moda/levtype=sfc/>>. Os dados são interpolados em uma grade de 2,5° x 2,5°. A TSM foi usada como preditor no modelo de previsão, a ser descrito no item 6.8.

6.2 - PCA

A Análise de Componentes Principais (PCA, *Principal Component Analysis*), também conhecida como *Empirical Orthogonal Function* (EOF), criada por Karl Pearson (1901) e fundamentada por Hotelling (1933) (Mingoti, 2005), é a técnica de estatística multivariada mais utilizada nas ciências atmosféricas, tendo sido introduzida nesse ramo de conhecimento por Obukhov (1947) (Wilks, 2011). Em razão das várias correlações observadas entre as variáveis atmosféricas, a PCA pode ser usada para representar os dados de maneira mais compacta e também para auxiliar na interpretação de seu comportamento temporal e espacial (Wilks, 2011), já que ajuda a explicar sua estrutura de variância (Mingoti, 2005). No estudo de secas, a PCA pode ser usada para extrair as informações essenciais quando há uma grande quantidade de variáveis e de índices representando o fenômeno (Hao e Singh, 2015).

Trata-se de uma ferramenta capaz de reduzir linearmente um conjunto de dados contendo um grande número de variáveis em um novo conjunto constituído de um menor número de variáveis, as quais são uma combinação linear das originais e contêm a maior parte da variabilidade dos dados iniciais (Wilks, 2011; Rajsekhar *et al.*, 2015; Mingoti, 2005), eliminando redundâncias geradas por correlações nos dados originais (Wilks, 2011). Os elementos desses novos dados são conhecidos como componentes principais (Wilks, 2011).

A técnica consiste inicialmente na decomposição da matriz de covariância $[S]$ dos dados iniciais (Mingoti, 2005). A partir dessa matriz pode-se obter a matriz ortogonal $[E]$, tal que $[E]^t[S][E]$ é uma matriz cuja diagonal é formada pelos chamados autovalores, apresentados em ordem decrescente (Mingoti, 2005). As colunas da matriz $[E]$ são representadas pelos chamados autovetores (Wilks, 2011; Mingoti, 2005). Na técnica PCA, os dados transformados preservam essa estrutura de covariância (Rajsekhar *et al.*, 2015).

Os dados originais, de dimensão n , são projetados em um subespaço que apresenta máxima variância (Wilks, 2011; Rajsekhar *et al.*, 2015; Zou *et al.*, 2006) e possui dimensão m , sendo $m \leq n$. Os autovetores, obtidos a partir da decomposição matriz de covariância dos dados iniciais (Mingoti, 2005; Rajsekhar *et al.*, 2015), rotacionam os eixos originais, criando um novo sistema de coordenadas, em que os eixos estão alinhados na direção da maior parte da variabilidade dos dados (Wilks, 2011). A PCA fornece as coordenadas dos dados de entrada nesse subespaço, em que os eixos principais possuem as direções especificadas pelos m autovetores (Wilks, 2011;

Rajsekhar *et al.*, 2015). Essa direção em que os dados estão mais representados corresponde ao autovetor associado ao maior autovalor, sendo que este corresponde à variância representada pelo respectivo eixo (Wilks, 2011). A segunda direção, que representa a segunda maior fração da variabilidade dos dados originais, é ortogonal à primeira e correspondente ao autovetor referente ao segundo maior autovalor, e assim por diante, sendo essa lógica é seguida até o m -ésimo autovetor (Rajsekhar *et al.*, 2015).

A Figura 6.1 traz um exemplo em duas dimensões da aplicação da PCA. Os dados originais correspondem a duas variáveis, x_1 (eixo das abscissas) e x_2 (eixo das ordenadas). Os autovetores e_1 e e_2 promovem uma rotação geométrica dos eixos, que deixam ser representados por x_1 e por x_2 e passam a retratar os dados em um novo sistema de coordenadas. Observa-se que o primeiro autovetor, e_1 , está alinhado na direção de maior variabilidade dos dados. O segundo autovetor, e_2 , precisa ser ortogonal ao primeiro e está voltado na direção da segunda maior variação dos dados.

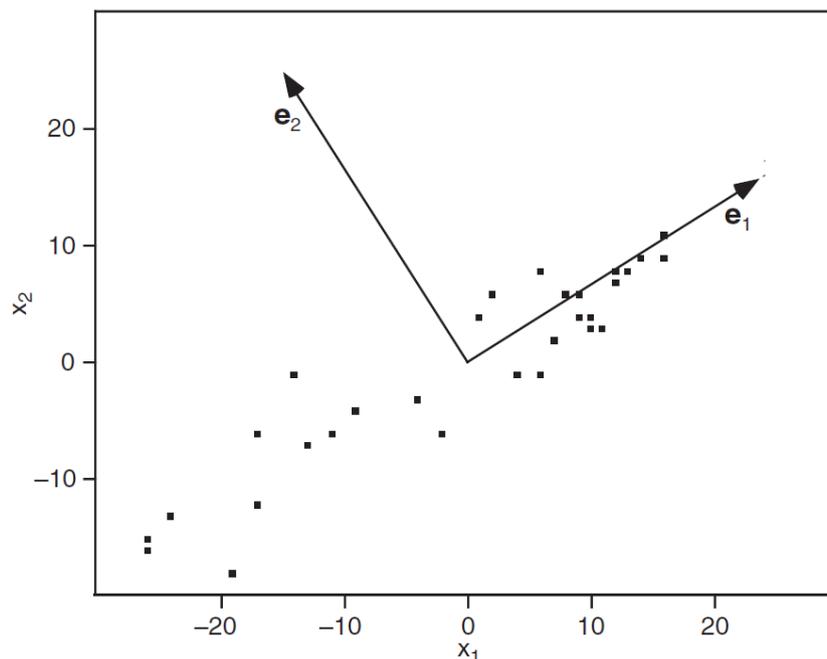


Figura 6.1 – Exemplo em duas dimensões da rotação de eixos na PCA (Fonte: adaptado de Wilks, 2011)

Cada componente principal é uma espécie de média ponderada dos elementos originais, em que os pesos são os autovetores correspondentes (Wilks, 2011). Sejam X a matriz de dados originais, com p observações e n variáveis x_k , u_i as componentes principais e e_i os autovetores

correspondentes. A primeira componente principal, u_1 , é uma combinação linear de X com a maior variância possível (Wilks, 2011). As componentes seguintes seguem a mesma lógica, diminuindo a variância correspondente (Wilks, 2011). Elas são definidas a partir dos autovetores (e_i) da matriz de covariância de X , $[S]$. Uma componente principal i é definida como a projeção dos dados originais no autovetor correspondente (Wilks, 2011):

$$u_i = e_i^t X = \sum_{k=1}^n e_{k,i} x_k, \quad i = 1, \dots, m \quad (6.1)$$

Em termos da matriz ortogonal $[E]$, que define a rotação, a Equação (6.2) pode ser reescrita como (Wilks, 2011):

$$u = [E]^t X \quad (6.2)$$

Tem-se como resultado m componentes principais que são não correlacionadas entre si (Wilks, 2011), já que a covariância entre cada par de componentes é zero (Wilks, 2011; Mingoti, 2005), podendo ser então serem modeladas separadamente. Como são p observações de dados originais (correspondendo a p linhas na matriz X), serão p observações para cada componente principal u_i . Os elementos dos autovetores são chamados de carregamentos (*loadings*) ou pesos e os das componentes principais, de escores (*scores*) (Wilks, 2011; Mingoti, 2005).

A PCA tem como objetivo representar as variáveis originais por um espaço de menor dimensão e uma forma de mensurar o quão boa é essa aproximação do espaço original pelo espaço das componentes principais é por meio do cálculo da proporção da variância total explicada por cada componente (Mingoti, 2005). A variância explicada (R_i^2) por cada componente principal i é fornecida pelo autovalor λ_i correspondente, como mostrado na Equação (6.3) (Wilks, 2011):

$$R_i^2 = \frac{\lambda_i}{\sum_{k=1}^n \lambda_k} \quad (6.3)$$

É possível reconstruir os dados originais a partir das componentes principais. Quando $m < n$, que é o caso mais comum, já que a PCA é usada para reduzir a dimensão dos dados, essa fórmula corresponde a uma aproximação (Equações (6.4), (6.5) e (6.6)) (Wilks, 2011):

$$X = [E]u \quad (6.4)$$

$$X \approx [E]u \quad (6.5)$$

$$x_k \approx \sum_{i=1}^m e_{k,i}u_i, \quad k = 1, \dots, n \quad (6.6)$$

A PCA também pode ser realizada utilizando-se a matriz de correlação $[R]$ ao invés da matriz de covariância $[S]$, equivalente à matriz de covariância da matriz de variáveis padronizadas Z (Wilks, 2011), a qual é obtida subtraindo de cada elemento de X a média μ e o desvio padrão σ (Mingoti, 2005). Quando se usa a matriz de covariância, a ênfase é em obter componentes principais com maior variância, já que a técnica busca maximizar a proporção da variância explicada (Wilks, 2011). Tem-se como resultado componentes principais que são mais influenciadas pelas variáveis originais de maior variabilidade, fazendo com que a técnica seja pouco útil quando há uma grande desigualdade nas variâncias iniciais (Mingoti, 2005). Isso acontece muitas vezes quando há diferença entre as unidades de medida das variáveis originais (Mingoti, 2005), fazendo com que, nesse caso, seja melhor usar a PCA com a matriz de correlação (Wilks, 2011), já que a padronização equilibra os valores de variância. Ao aplicar a PCA usando a matriz de correlação $[R]$, todas as variáveis padronizadas z_i recebem o mesmo peso (Wilks, 2011), pois a padronização equilibra os valores de variância.

A PCA é muito utilizada por apresentar algumas propriedades vantajosas. Primeiramente, ela é capaz de obter a variabilidade máxima dos dados originais, assegurando uma perda mínima de informação (Zou *et al.*, 2006; Wilks, 2011; Mingoti, 2005). Além disso, as componentes principais não são correlacionadas entre si, ou seja, a covariância entre elas é zero, fazendo com que elas possam ser tratadas de forma independente (Zou *et al.*, 2006; Wilks, 2011; Mingoti, 2005). Já uma das principais limitações da técnica é assumir a linearidade das transformações dos dados originais, o que pode não corresponder à realidade (Hao e Singh, 2015; Rajsekhar *et al.*, 2015; Zou *et al.*, 2006).

6.3 - ANÁLISE DE EXTREMOS E DE ÁREAS

Observa-se, pela Tabela 4.2, que valores do SPI abaixo de -1,5 indicam secas severas e extremas, correspondendo a 6,7% do tempo nas categorias. Assim, foi realizada uma análise de extremos procurando, nas componentes principais do SPI, os meses correspondentes aos valores 6,7% mais extremos, com o objetivo de identificar os padrões das maiores secas. Como o campo completo do SPI contém muitos pontos, realizar essa análise usando as componentes facilita o processamento os dados e o entendimento dos resultados.

Uma vez obtido os meses que as componentes principais do SPI identificam como mais extremos, pode-se avaliar as informações fornecidos pelo campo completo do índice de seca. Assim, foi feita uma busca, nesses meses extremos, pelos valores do SPI que estavam abaixo de -1,5 (ou seja, valores dos índices correspondentes a secas meteorológicas severas e extremas). Sabendo os pontos em que esses valores mais baixos estavam posicionados, foi possível determinar espacialmente a porcentagem da área do Nordeste afetada em cada mês de seca severa ou extrema.

6.4 - ANÁLISE COMPÓSITO

Como já mencionado, a temperatura e a precipitação são duas variáveis que desempenham um papel importante no desenvolvimento de uma seca (Mishra e Singh, 2010; Marengo *et al.*, 2013; Marengo e Bernasconi, 2015). Para avaliar esse comportamento e verificar essa influência, foi realizada uma análise compósito, que consistiu no cálculo das médias das anomalias dessas duas variáveis nos meses indicados como extremos por cada componente principal do SPI. Como resultado, foram gerados mapas contendo os padrões espaciais dessas anomalias médias, com o objetivo de verificar como esses padrões estavam relacionados com os padrões espaciais de secas indicadas pelo SPI.

6.5 - ANÁLISE DE CORRELAÇÃO CRUZADA

Essa análise consiste em estimar as correlações entre duas variáveis em diferentes tempos de adiantamento (ou atraso) de uma em relação à outra. Neste trabalho, essa análise foi utilizada para verificar a associação entre o índice de seca SPI e os índices climáticos descritos anteriormente (NINO3.4 e Dipolo), visando avaliar a influência da TSM nos Oceanos Atlântico

e Pacífico nas secas meteorológicas ocorridas no Nordeste, identificadas pelo SPI de 3 meses, e em qual atraso essa relação é máxima. Seja então V_1 a variável resposta, que corresponde a uma dada componente principal do SPI, V_2 o índice climático que a precede e τ o tempo de adiantamento. A correlação cruzada Cor , em um tempo t , é dada então pela Equação 6.7. Foram considerados valores de τ variando de 0 a 12 meses.

$$Cor(V_1(t + \tau), V_2(t)), \tau \geq 0 \quad (6.7)$$

6.6 - CCA

A Análise de Correlação Canônica (CCA, *Canonical Correlation Analysis*) é uma técnica de estatística multivariada, introduzida por Harold Hotelling (1935, 1936) (Wiesel *et al.*, 2008; Mingoti, 2005), capaz de identificar pares de padrões em dois conjuntos de dados e projetar os dados originais nesses padrões, obtendo assim um novo conjunto de variáveis transformadas (Wilks, 2011), a partir de combinações lineares (Mingoti, 2005). A escolha desses padrões é feita de forma que as relações entre os dois conjuntos de dados seja maximizada, ou seja, as novas variáveis devem possuir máxima correlação (Wilks, 2011; Mingoti, 2005). A matriz de correlação encontrada é diagonal, com as correlações presentes na diagonal sendo as maiores possíveis. O método CCA é reproduzido a seguir, conforme descrito por Wilks (2011).

Seja x o vetor de variáveis preditoras e y a variável a ser prevista. A CCA busca pares de conjuntos de pesos de maneira que as novas variáveis definidas pelos respectivos produtos escalares com x e y tenham as correlações entre si maximizadas. Na área de conhecimento geofísico, é comum que o vetor x contenha as observações de uma variável em diferentes localizações, enquanto o vetor y possua as séries temporais de outra variável em locais que podem não ser os mesmos de x . A CCA pode ser então usada para prever o vetor y utilizando observações de x (Wilks, 2011), que é a aplicação utilizada neste trabalho.

A CCA utiliza a matriz de covariância conjunta para obter as relações entre os dois conjuntos de dados. Sejam $x' = x - \bar{x}$ e $y' = y - \bar{y}$ os vetores de dados centrados (\bar{x} e \bar{y} são as respectivas médias), com I e J elementos, respectivamente. Eles são primeiramente associados em um vetor $c'^T = [x'^T, y'^T]$, que contém $I + J$ elementos. A forma matricial de x e de y é representada por X e por Y , considerando que cada conjunto de dados possui n observações, fazendo com que o

vetor c' seja representado por C' . A matriz de covariância $[Sc]$ de c' possui dimensão $((I + J) \times (I + J))$ e é repartida em quatro blocos, conforme a Equação (6.8):

$$[Sc] = \frac{1}{n-1} [C']^t [C'] = \begin{bmatrix} [S_{x,x}] & [S_{x,y}] \\ [S_{y,x}] & [S_{y,y}] \end{bmatrix} \quad (6.8)$$

A matriz $[C']$ possui dimensão $(n \times (I + J))$ e cada uma das n linhas possuem uma observação do vetor x' e uma do y' . A matriz $[S_{x,x}]$ é a matriz variância-covariância das I variáveis de x , apresentando dimensão $I \times I$, enquanto $[S_{y,y}]$ é a matriz variância-covariância das J variáveis de y , possuindo dimensão $J \times J$. Essas duas matrizes fornecem as associações existentes entre os elementos de x e de y , respectivamente (Mingoti, 2005). As matrizes $[S_{x,y}]$ (dimensão $I \times J$) e $[S_{y,x}]$ apresentam as covariâncias de todas as combinações dos elementos de x e de y , sendo que $[S_{x,y}] = [S_{y,x}]^t$. Quando as dimensões I e J são grandes, a interpretação de todos os elementos de $[S_{x,y}]$ é dificultada, sendo a CCA uma alternativa de resolução desse problema, já que ela permite que as relações entre x e y sejam avaliadas a partir de combinações lineares desses dois vetores (Mingoti, 2005).

Essa transformação dos vetores centrados dos dados originais, x' e y' , é feita utilizando-se os *vetores canônicos* (também conhecidos como coeficientes canônicos ou como pesos) a_m e b_m , respectivamente, resultando nas novas variáveis v_m e w_m , respectivamente, chamadas de *variáveis canônicas*. Isso é feito por meio dos produtos escalares mostrados nas Equações (6.9) e (6.10). Um vetor de dados e seu respectivo vetor canônico devem possuir a mesma dimensão, ou seja, x' e a_m possuem I elementos, enquanto y' e b_m , J elementos. As variáveis canônicas, entretanto, precisam apresentar a mesma dimensão (M), a qual é definida como a menor dimensão de x e y , ou seja, $M = \text{mínimo}(I, J)$.

$$v_m = a_m^t x' = \sum_{i=1}^I a_{m,i} x'_i, \quad m = 1, \dots, M \quad (6.9)$$

$$w_m = b_m^t y' = \sum_{i=1}^J b_{m,i} y'_i, \quad m = 1, \dots, M \quad (6.10)$$

As variáveis canônicas v_m e w_m são, portanto, uma combinação linear, uma espécie de média ponderada, dos respectivos vetores x' e y' . Os vetores canônicos a_m e b_m são escolhidos de forma que as variáveis canônicas apresentem algumas propriedades. A primeira delas, representada pela Equação (6.11), estabelece que cada par de variáveis canônicas apresente correlação menor ou igual ao par anterior. Essas correlações são nomeadas *correlações canônicas* (r_C).

$$\text{Cor}(v_1, w_1) \geq \text{Cor}(v_2, w_2) \geq \dots \geq \text{Cor}(v_M, w_M) \geq 0 \quad (6.11)$$

As próximas duas restrições (Equações (6.12) e (6.13)) estabelecem que as variáveis canônicas sejam correlacionadas apenas com seu respectivo par, sendo não correlacionadas com qualquer outra variável:

$$\text{Cor}(v_k, w_m) = \begin{cases} r_{C_m}, & k = m \\ 0, & k \neq m \end{cases} \quad (6.12)$$

$$\text{Cor}(v_k, v_m) = \text{Cor}(w_k, w_m) = 0, \quad k \neq m \quad (6.13)$$

Por fim, a Equação (6.14) determina que cada variável canônica possua variância unitária:

$$\text{Var}(v_m) = a_m^T [S_{x,x}] a_m = \text{Var}(w_m) = b_m^T [S_{y,y}] b_m = 1, \quad m = 1, \dots, M \quad (6.14)$$

É preciso estabelecer algumas restrições nos comprimentos dos vetores canônicos (a_m e b_m), o que pode ser feito de forma a garantir a variância unitária, sendo útil em algumas aplicações. Neste caso, a matriz de covariância para as variáveis canônicas assume a forma simples da Equação (6.15), em que R_C é a matriz diagonal das correlações canônicas (Equação (6.16)).

$$\text{Var} \left(\begin{bmatrix} v \\ w \end{bmatrix} \right) = \begin{bmatrix} [S_{v,v}] & [S_{v,w}] \\ [S_{w,v}] & [S_{w,w}] \end{bmatrix} = \begin{bmatrix} [I] & [R_C] \\ [R_C] & [I] \end{bmatrix} \quad (6.15)$$

$$[R_C] = \begin{bmatrix} r_{C_1} & 0 & 0 & \dots & 0 \\ 0 & r_{C_2} & 0 & \dots & 0 \\ 0 & 0 & r_{C_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & r_{C_M} \end{bmatrix} \quad (6.16)$$

As matrizes contendo as variáveis canônicas de x' e y' , denotadas por v e w , respectivamente, podem ser obtidas pelas Equações (6.17) e (6.18), em que as matrizes A e B contêm os vetores canônicos a_m e b_m , respectivamente, e, portanto, possuem dimensões $M \times I$ e $M \times J$, respectivamente. Os vetores v e w possuem dimensão $M \times I$ e podem ainda serem escritos na respectiva forma matricial, V e W , respectivamente.

$$v = [A]^T x' \quad (6.17)$$

$$w = [B]^T y' \quad (6.18)$$

Informações complementares acerca da técnica CCA, incluindo propriedades, variância explicada pelas variáveis canônicas e fórmulas de cálculos dos vetores canônicos, são mostradas no Apêndice A.

6.7 - SCCA

A CCA, como descrita no item anterior, é empregada utilizando as matrizes completas ou cheias (traduzindo do inglês, *full matrix*), ou seja, com valores diferentes de zero (Lima e AghaKouchak, 2017). Isso significa que os coeficientes canônicos/vetores canônicos/ pesos são todos diferentes de zero, fazendo com que, para os pontos que contribuem pouco, os valores sejam muito pequenos, mas nunca iguais a zero. Os resultados alcançados dessa forma apresentam ruídos e os estimadores obtidos são pouco robustos (Lima e AghaKouchak, 2017), o que significa que eles apresentam baixo desempenho em validação cruzada e em previsão em tempo real.

Uma alternativa para resolver esse problema consiste em utilizar o método Análise de Correlação Canônica Esparsa (SCCA, *sparse Canonical Correlation Analysis*) ou Análise de Correlação Canônica Robusta (*robust CCA, robust Canonical Correlation Analysis*), que atribui peso zero para coeficientes de baixo valor, que não contribuem efetivamente para a correlação após a projeção, ou seja, para a correlação entre as variáveis canônicas. Visa-se então obter preditores climáticos esparsos (covariáveis climáticas esparsas), com poucos pontos diferentes de zero e, portanto, mais robustos na previsão dos índices de seca, já que são obtidas combinações lineares usando poucas variáveis, mas que possuem máxima correlação (Wiesel

et al., 2008). Esse método, além de melhorar as previsões, gera resultados de mais fácil interpretação e visualização, por utilizar apenas os pontos-chave, além de melhorar a eficiência computacional e trazer maior estabilidade ao processo, já que a CCA é sensível a poucas observações (Wiesel *et al.*, 2008). A seguir é descrito o procedimento de cálculo adaptado daquele que foi proposto por Witten *et al.* (2009), que é uma aplicação de Decomposição de Matriz Penalizada (PMD, *Penalized Matrix Decomposition*) para obtenção de CCA penalizada.

Seja X a matriz contendo o campo observado da variável climática, de dimensão $n \times I$, em que n é o número de observações durante o período de tempo considerado e I é o número de pontos. Seja Y a matriz dos valores observados do índice de seca para o Nordeste brasileiro, de dimensão $n \times J$, em que J é o número de pontos para o índice. Assume-se que X e Y são matrizes centradas e escalonadas. A técnica CCA, como já explicado, busca os vetores canônicos A e B que maximizam a correlação das projeções de X e de Y nesses vetores ($\text{Cor}(Xa_m, Yb_m)$), resolvendo o problema de otimização mostrado na Equação (6.19). O vetor $A = [a_1 \dots a_m]$ possui dimensão I , o vetor $B = [b_1 \dots b_m]$ tem dimensão J e $M = \text{mínimo}(I, J)$.

$$\text{maximizar}_{B,A} B^t Y^t X A \quad \text{sujeito a } B^t Y^t Y B \leq 1, A^t X^t X A \leq 1 \quad (6.19)$$

O primeiro termo da Equação (6.19) indica o objetivo de maximizar a correlação entre a projeção dos dois campos. Se os pesos A e B , entretanto, pudessem ser aumentados indefinidamente, a correlação poderia ser maximizada de uma forma arbitrariamente alta, já que os pesos poderiam apresentar qualquer valor, o que não é desejável. É preciso então limitar o crescimento desses coeficientes, o que é feito pelo segundo termo da equação, que se trata de uma restrição que faz com que os pesos permaneçam dentro do círculo unitário. A Equação (6.19) mostra a formulação da CCA tradicional, que pode ser resolvida por autovalores e autovetores, como mostrado nas Equações (A.13) e (A.14).

A técnica, tradicionalmente, fornece vetores A e B que não são esparsos, o que é não desejável para algumas aplicações, principalmente quando I ou J é grande e maior do que n . Para obter vetores canônicos que sejam penalizados (vetores canônicos esparsos), ou seja, que possuam um certo número de elementos exatamente iguais a zero (Lima e AghaKouchak, 2017), pode-se incluir penalidades (P_1 e P_2) na Equação (6.19), formando assim a Equação (6.20). Essas penalidades restringem ainda mais o crescimento dos pesos, para constantes c_1 e c_2 , o que obriga que alguns pesos apresentem valores exatamente iguais a zero.

$$\text{maximizar}_{B,A} B^t Y^t X A \quad \text{sujeito a } B^t Y^t Y B \leq 1, A^t X^t X A \leq 1, P_1(B) \leq c_1, P_2(A) \leq c_2 \quad (6.20)$$

As penalidades P_1 e P_2 são funções de penalidade convexas, que podem assumir muitas formas. Neste trabalho, escolheu-se a penalidade Lasso, representada pelas Equações (6.21) e (6.22):

$$P_1(B) = \sum_{i=1}^M \sum_{j=1}^J |b_{ji}| \quad (6.21)$$

$$P_2(A) = \sum_{i=1}^M \sum_{j=1}^I |a_{ji}| \quad (6.22)$$

Apenas algumas faixas de valores de c_1 e de c_2 fornecem soluções possíveis, mostrados na Equação (6.23). Quanto maior os valores de c_1 e de c_2 , menor a penalização ou esparsidade, ou seja, o número de coeficientes diferentes de zero é menor.

$$1 \leq c_1 \leq \sqrt{J} \quad \text{e} \quad 1 \leq c_2 \leq \sqrt{I} \quad (6.23)$$

Os valores de c_1 e de c_2 podem ser obtidos por validação cruzada, em que A e B são o conjunto de treinamento e X e Y , o conjunto de teste independente, de forma a maximizar $\text{cor}(Xa, Yb)$. Esses valores também podem ser escolhidos de forma a obter a esparsidade desejada para a e para b .

A aplicação do método SCCA utilizando os campos completos do índice de seca (Y) e da covariável climática (X) poderia resultar em coeficientes canônicos e explicações físicas de difícil interpretação (Lima e AghaKouchak, 2017). Uma solução para isso é substituir o campo inteiro do índice por suas primeiras componentes principais (Lima e AghaKouchak, 2017). Assim, aqui em diante, a matriz Y passa a denotar as primeiras componentes do índice de seca, representativas de uma parcela considerável da variabilidade dos dados.

Para melhor avaliar os resultados da técnica SCCA, foi escolhido um valor para a constante c_1 tal que B seja formada por coeficientes canônicos b_m que sejam vetores unitários ortogonais, fazendo com que B seja uma matriz ortogonal de permutação dos eixos coordenados. Isso faz

com que a projeção YB seja igual a Y , ou seja, faz com que as variáveis canônicas sejam iguais às componentes principais do índice de seca (Lima e AghaKouchak, 2017). Para isso, utilizou-se uma penalidade $penaltyy = 0,3$ (Equação (24)). Feito isso, o valor de c_2 pode ser determinado pela Equação (6.25):

$$c_1 = \text{mínimo}(1, penaltyy \cdot \sqrt{J}) \quad (6.24)$$

$$c_2 = \text{mínimo}(1, penaltyx \cdot \sqrt{I}) \quad (6.25)$$

A penalidade, denotada por $penaltyy$ ou $penaltyx$, é um número que pode variar de 0 a 1. Quanto mais próximo de 1, menor é a penalização ocorrida, ou seja, o número de coeficientes canônicos diferente de zero diminui. Neste trabalho, foram testados diferentes valores de $penaltyx$ nas correlações entre o SPI e a TSM, para determinar o mais adequado para ser incorporado no modelo de previsão.

6.8 - MODELO DE PREVISÃO

Para obter previsões do índice de seca a partir de uma variável climática, foi selecionado um modelo de previsão linear multivariado. Inicialmente, são previstas as variáveis canônicas da CCA, a partir das quais é obtido o campo completo do índice. Em razão da escolha feita para o valor da constante c_1 , como explicado no item 6.7, as variáveis canônicas são iguais às componentes principais do índice de seca, podendo ser transformadas para o campo completo usando os autovetores (Equação (6.4) a (6.6)). Assim, o resultado da previsão é um vetor $Y_i(t)$ para cada uma das i componentes em um mês t ($i = 1, 2, 3$). A previsão é realizada para diversos tempos de antecedência τ , que em geral poderão variar de 1 a 6 meses. O vetor $Y_i(t)$, para um mês t , considerando uma previsão com antecedência de τ meses, é obtido pelo modelo de previsão representado pelas Equações (6.26) a (6.28) a seguir, para cada uma das três primeiras componentes:

$$Y_1(t) = \alpha_1 + \beta_{1,1}(X_1(t - \tau)) + \beta_{1,2}(X_2(t - \tau)) + \beta_{1,3}(X_3(t - \tau)) + \gamma_1(Y_1(t - \tau))_1 + \varepsilon_1 \quad (6.26)$$

$$Y_2(t) = \alpha_2 + \beta_{2,1}(X_1(t - \tau)) + \beta_{2,2}(X_2(t - \tau)) + \beta_{2,3}(X_3(t - \tau)) + \gamma_2(Y_2(t - \tau))_2 + \varepsilon_2 \quad (6.27)$$

$$Y_3(t) = \alpha_3 + \beta_{3,1}(X_1(t - \tau)) + \beta_{3,2}(X_2(t - \tau)) + \beta_{3,3}(X_3(t - \tau)) + \gamma_3(Y_3(t - \tau))_3 + \varepsilon_3 \quad (6.28)$$

Os coeficientes α , β e γ são os coeficientes de uma regressão linear. As três primeiras componentes principais da TSM, que é uma covariável climática, são representadas pelos termos $X_j(t - \tau)$, para cada uma das j ($j = 1, 2, 3$) componentes em um mês t e em um atraso τ . O $Y_i(t - \tau)$ e seu respectivo coeficiente γ representa o termo de persistência, ou seja, a parte da equação que relaciona o Y no tempo t com o Y em um tempo anterior $t - \tau$. Por fim, ε é o erro, que segue uma distribuição normal multivariada, com média zero e desvio padrão igual à covariância entre os erros. Esse erro precisa ter média zero, caso contrário haveria um viés (bias) na previsão.

O modelo usado para comparar (*benchmarking*), chamado de modelo auto-regressivo ou modelo de referência, é mostrado a seguir (Equações (6.29) a (6.31)). Nele, retira-se o campo da variável climática ($\beta = 0$) e utiliza-se apenas o termo de persistência para realizar as previsões. Deseja-se, portanto, verificar o ganho que o uso da informação climática agrega ao modelo.

$$Y_1(t) = \alpha_1 + \gamma_1(Y_1(t - \tau)) + \varepsilon_1 \quad (6.29)$$

$$Y_2(t) = \alpha_2 + \gamma_2(Y_2(t - \tau)) + \varepsilon_2 \quad (6.30)$$

$$Y_3(t) = \alpha_3 + \gamma_3(Y_3(t - \tau)) + \varepsilon_3 \quad (6.31)$$

Foi realizado uma validação cruzada com retirada de um ano. A técnica é executada retirando-se das Equações (6.26) a (6.28) as observações de Y e de X relativas ao primeiro ano dos dados e usando os dados restantes para estimar os parâmetros das Equações (6.26) a (6.28). Uma vez calibrado, o modelo é usado para calcular a previsão de Y para o primeiro ano. Em seguida, o método é repetido, retirando-se dessa vez os dados do segundo ano. Esse procedimento é repetido até o ano final dos dados. Essa metodologia é útil para avaliar o desempenho do modelo de previsão.

A etapa seguinte da previsão foi a verificação, visando determinar o quão boas foram as previsões obtidas. O desempenho (performance) do modelo de previsão linear multivariado foi

avaliado por meio de três métricas: correlação, raiz do erro médio quadrático (RMSE, *Root-Mean Squared Error*) e BSS (*Brier Skill Score*).

O coeficiente de correlação, também conhecido como coeficiente de correlação de *Pearson*, mede o grau de associação linear entre duas variáveis. Sendo essas variáveis os valores obtidos pelo modelo de previsão (y) e os valores observados (o), o coeficiente de correlação (r_{yo}) pode ser calculado como a razão entre a covariância de y e de o e a multiplicação de seus respectivos desvios-padrão (s_y e s_o , respectivamente), conforme mostrado na Equação (6.32) (Wilks, 2011). O coeficiente de correlação pode assumir valores de -1 a 1 ($-1 < r_{yo} < 1$) e, quanto mais próximo desses extremos, melhor é o ajuste. Essa métrica, no entanto, não é capaz de fornecer uma explicação física ou causal para a relação entre as variáveis envolvidas, tornando necessário o uso de outras medidas complementares (Wilks, 2011).

$$r_{yo} = \frac{Cov(y, o)}{s_y s_o} \quad (6.32)$$

A raiz do erro médio quadrático (RMSE, *Root-Mean Squared Error*) é uma medida de acurácia, sendo uma das métricas mais usadas para mensurar a precisão de uma previsão (Wilks, 2011). O RMSE é a raiz quadrada do erro médio quadrático (MSE, *Mean Squared Error*), apresentando as vantagens de ser expresso na mesma unidade das variáveis estudadas e de possuir uma magnitude de erro facilmente interpretável (Wilks, 2011). O MSE calcula as diferenças individuais ao quadrado dos valores previstos (y_k) e dos valores observados (o_k) para cada um dos n pares de valores (Wilks, 2011). A formulação do RMSE é apresentada na Equação (6.33). Quanto mais próximo de zero é o valor do RMSE, melhor são as previsões feitas.

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (y_k - o_k)^2} \quad (6.33)$$

Uma das métricas mais usadas para a verificação de previsões probabilísticas é o *Brier Score* (BS), definido pela Equação (6.34), em que y_k denota a probabilidade da variável estar em uma determinada faixa de valores e o_k a probabilidade para os valores observados, para cada um dos n pares de dados. Trata-se de uma métrica semelhante ao erro médio quadrático, em que a

observação é 0 se o evento não ocorrer e igual a 1, caso ele ocorra, fazendo com que a faixa de valores possíveis de BS seja entre 0 e 1 ($0 \leq BS \leq 1$) (Wilks, 2011).

$$BS = \frac{1}{n} \sum_{k=1}^n (y_k - o_k)^2 \quad (6.34)$$

A métrica calculada neste trabalho, *Brier Skill Score* (BSS), é calculada pela Equação (6.36), em que BS é a medida para os dados medidos pelo modelo de previsão e BS_{ref} (*Brier Score* para o modelo de referência) representa as previsões de referência (climatologia, modelo auto-regressivo), formadas pelas frequências relativas de referência. Para o cálculo do BS_{ref} , foram consideradas as secas 20% mais extremas (com probabilidade acumulada de 20%), ou seja, em cada ponto, o BS_{ref} é estimado a partir da probabilidade do quantil de 20% do SPI observado (Equação (6.35)). O valor de 20% já havia sido utilizado por Lima e AghaKouchak (2017), pois probabilidades menores, como 10% ou 6,7% (porcentagem que define uma seca como severa ou extrema), não forneceriam resultados adequados. Previsões perfeitas resultam em $BSS = 0$, então, quando maior for o valor de BSS , mais precisa é a precisão (Wilks, 2011). Assim, quando o valor de BSS é positivo (BS menor que o BS_{ref}), tem-se que as previsões são melhores do que a climatologia (Wilks, 2011).

$$BS_{ref} = \frac{1}{n} \sum_{k=1}^n (0,20 - o_k)^2 + ((1 - 0,20) - o_k)^2 \quad (6.35)$$

$$BSS = 1 - \frac{BS}{BS_{ref}} \quad (6.36)$$

7 - RESULTADOS

Este capítulo traz os resultados obtidos por cada uma das técnicas descritas no capítulo 6. O item 7.1 mostra os resultados da Análise das Componentes Principais; o item 7.2 expõe as informações obtidas a partir da análise de extremos (item 7.2.1), destacando alguns eventos de maior abrangência espacial (item 7.2.2); o item 7.3 exhibe os mapas obtidos a partir da análise compósito; e o item 7.4 mostra qual é a influência de indicadores climáticos (NINO3.4 e dipolo do Atlântico) sobre o SPI. Em relação à obtenção das previsões, o item 7.5 traz uma análise inicial das correlações entre o SPI e a TSM; o item 7.6 mostra a aplicação da técnica SCCA; e, por fim, o item 7.7 exhibe os resultados finais do trabalho, obtidos a partir do modelo de previsão sazonal.

7.1 - PCA

As três primeiras componentes do SPI explicam 41,53%, 11,52% e 5,70% da variância dos dados, respectivamente. Em geral, componentes que explicam menos do que 10% da variabilidade dos dados estão associadas com fortes ruídos, o que pode dificultar a interpretação. Assim, as análises mostradas nos próximos itens focam apenas nas duas primeiras componentes, que juntas representam 52,94% da variância do SPI. As Figuras 7.1 a 7.3 mostram os padrões espaciais dos pesos (*loadings*) das três primeiras componentes principais do SPI sobre o Nordeste brasileiro.

Observa-se que os pesos associados à primeira componente do SPI (Figura 7.1) apresentam apenas valores positivos, indicando um padrão homogêneo de seca meteorológica em toda região, sendo uma espécie de média para o Nordeste. Os valores apresentam uma maior magnitude ao longo de uma faixa central da região, indicando, portanto, a existência de condições mais secas nessa área.

Os pesos da segunda componente (Figura 7.2) mostram um comportamento meridional do tipo dipolo, com valores positivos ao sul e negativos ao norte, sugerindo que as condições ao sul são mais secas e ao norte são úmidas (ou próximas de neutras) e vice-versa, a depender do sinal da componente principal (ver Equações (6.4) a (6.6)). Isso significa que, quando a componente for positiva, o valor do SPI será positivo (menos seco) ao sul e negativo (mais seco) ao norte. Já

quando a componente for negativa, observa-se o inverso, com o valor do SPI negativo (mais seco) ao sul e positivo (menos seco) ao norte.

Já na terceira componente (Figura 7.3), os pesos apresentam um padrão zonal, com valores positivos a leste e negativos a oeste. Assim, tem-se um padrão de secas mais intensas a leste e do que a oeste e vice-versa, a depender do sinal da componente principal. Quando o sinal da componente principal é positivo, as condições são mais secas a oeste, quando a componente é negativa, é mais seco a leste.

É preciso destacar, entretanto, que as três componentes atuam simultaneamente. Assim, pode-se ter uma seca meteorológica determinada pela primeira componente atingindo o Nordeste como um todo, mas mais intensa a norte do que a sul, por exemplo, pelo efeito modulador da segunda componente.

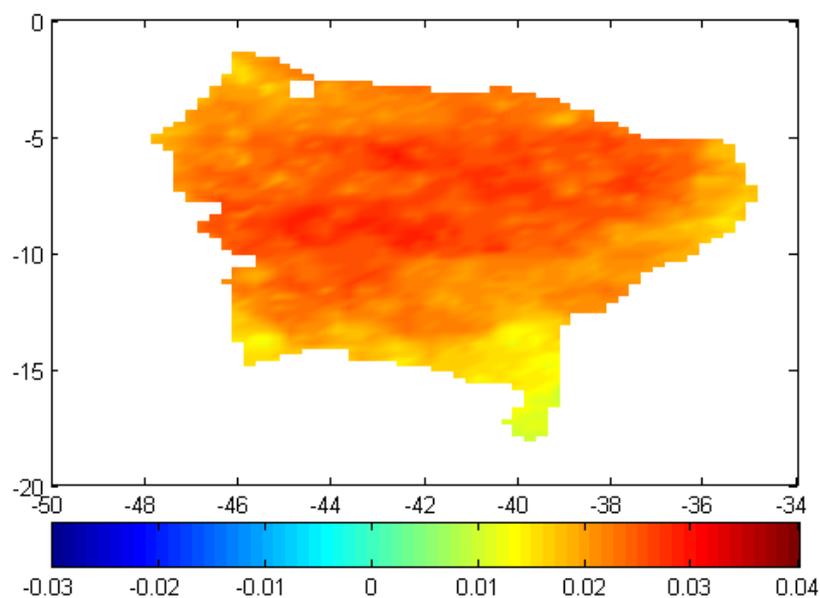


Figura 7.1 - Padrões espaciais dos pesos da primeira componente principal do SPI3

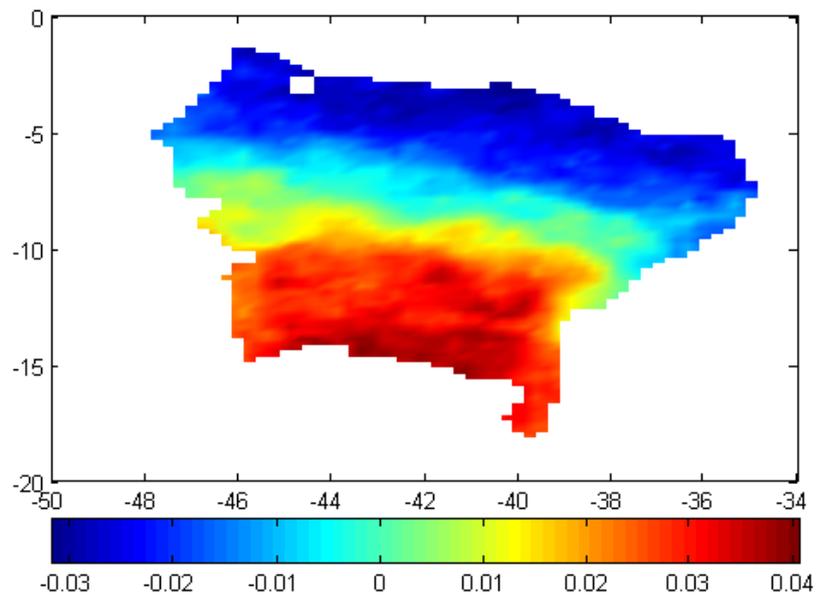


Figura 7.2 - Padrões espaciais dos pesos da segunda componente principal do SPI3

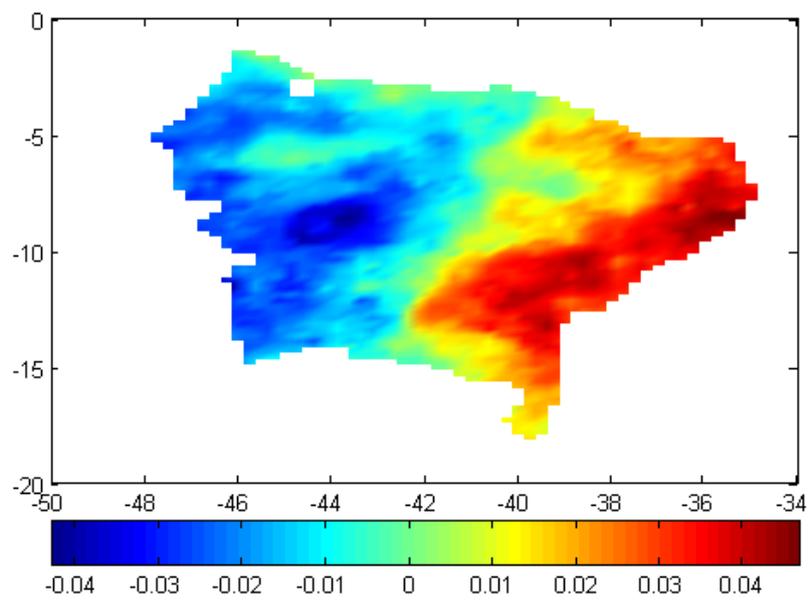


Figura 7.3 - Padrões espaciais dos pesos da terceira componente principal do SPI3

7.2 - ANÁLISE DE EXTREMOS E DE ÁREAS

A seguir são mostrados os resultados da análise de extremos (item 7.2.1) e o delimitamento de alguns dos eventos que apresentaram maior abrangência espacial (item 7.2.2).

7.2.1 - Análise de extremos

Para a primeira componente do SPI, foram detectados os valores 6,7% mais baixos, por ela representar um padrão homogêneo de secas no Nordeste. Já para a segunda componente, em razão do comportamento dipolo (com as secas alternando-se entre o norte e o sul, a depender do sinal da componente principal), identificou-se os valores 6,7% menores (SPI - 2ª componente inferior, valores negativos) e maiores (SPI - 2ª componente superior, valores positivos). As Figuras 7.4 e 7.5 mostram as séries temporais das duas primeiras componentes do SPI, em que as linhas horizontais representam os percentis de 6,7% considerados. É importante lembrar que a escolha desse percentil leva a classificação de secas severas ou extremas (vide Tabela 4.2).

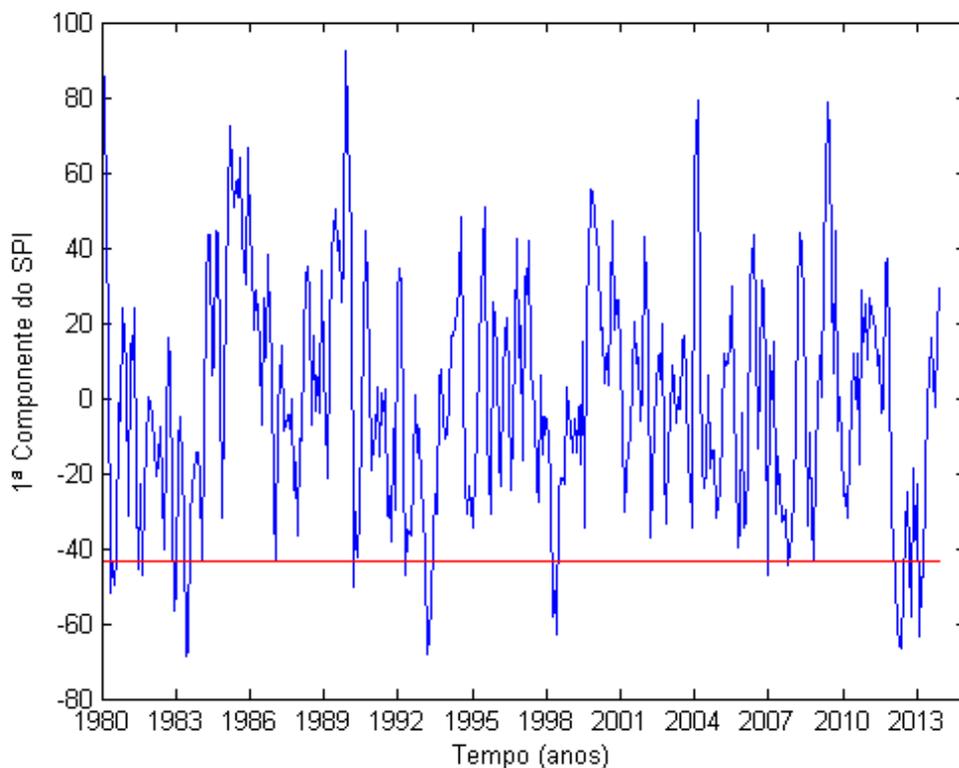


Figura 7.4 - Série temporal da primeira componente do SPI3, em que as linhas vermelhas representam os percentis de 6,7%

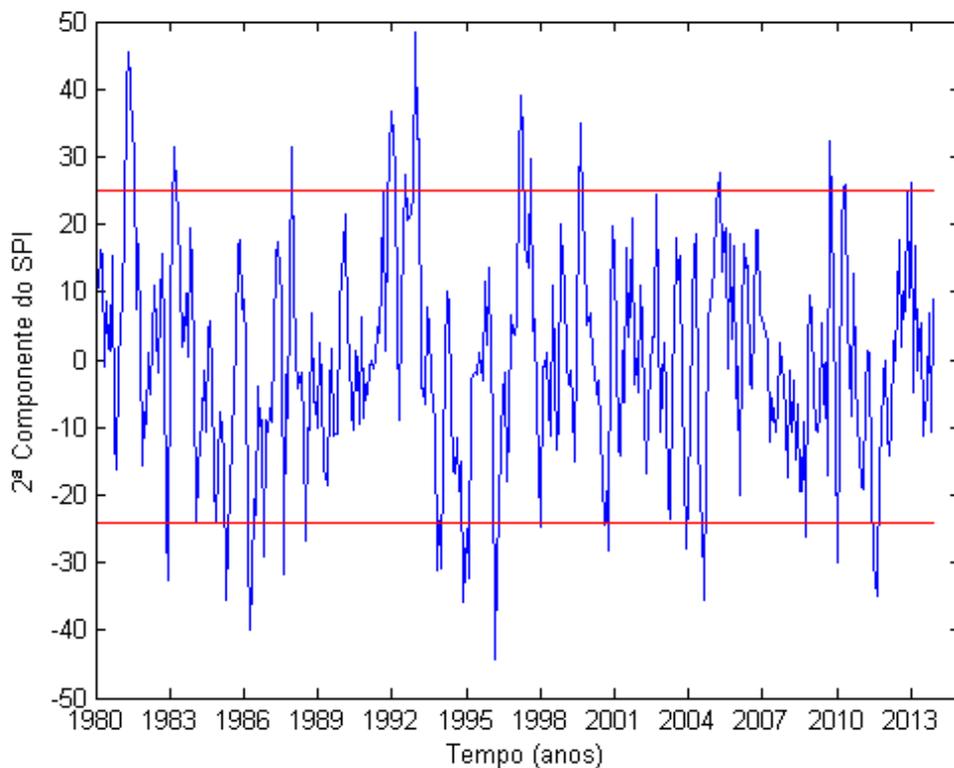


Figura 7.5 - Série temporal da segunda componente do SPI3, em que as linhas vermelhas representam os percentis de 6,7%

As Figuras 7.6 a 7.8 mostram a porcentagem da área do Nordeste atingida nos extremos indicados por cada componente principal (27 eventos extremos cada). Uma tabela detalhada contendo os valores exatos dessas áreas e os meses correspondentes é apresentada no Apêndice deste trabalho (Tabela B.1). A título ilustrativo, as Figuras 7.9 a 7.12 mostram o comportamento espacial do SPI nos meses correspondentes às secas que atingiram as maiores áreas ao longo do período considerado. A seca de maior abrangência espacial foi a de março de 1993 (56,43%) (Figura 7.10), seguida pela seca de maio de 2012 (51,08%) (Figura 7.12) e pela seca de abril de 1993 (49,90%) (Figura 7.11). Para avaliar outro período de seca que ocupou uma área também expressiva, de 40,18%, escolheu-se a seca de junho de 1983 (Figura 7.9), que foi o 6º evento de maior abrangência. Todos esses eventos foram sinalizados pela primeira componente do SPI, observando-se, então, que os eventos de maior abrangência espacial foram os sinalizados pela primeira componente, em função dela representar a maior parte da variabilidade dos dados.

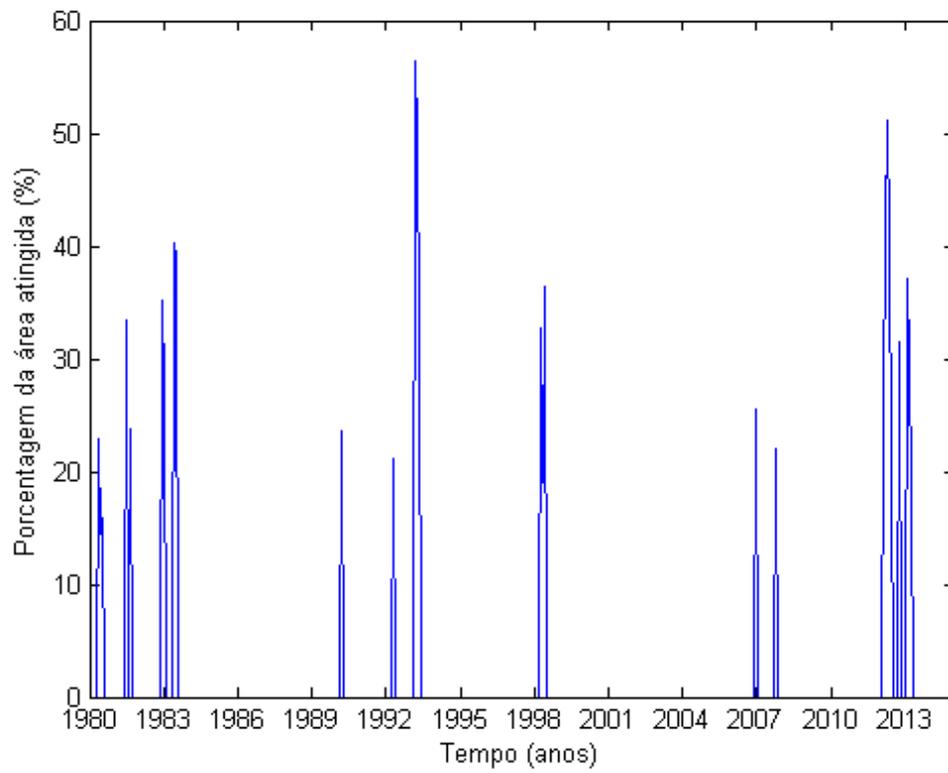


Figura 7.6 - Porcentagem da área do Nordeste atingida pelos eventos extremos de secas meteorológicas identificados pela primeira componente do SPI3

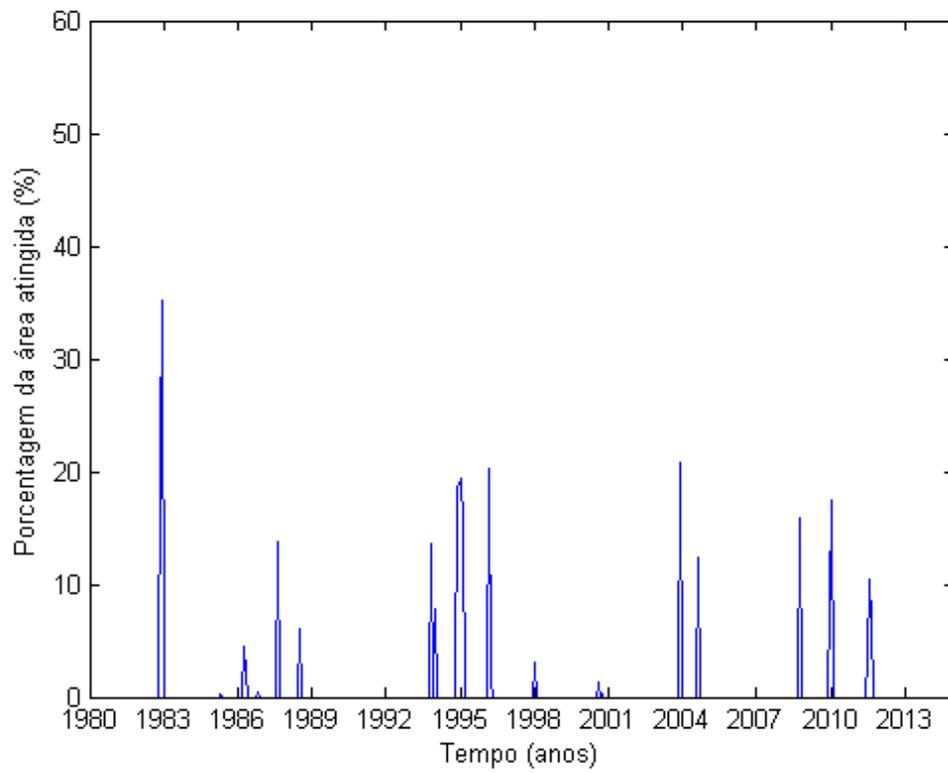


Figura 7.7 - Porcentagem da área do Nordeste atingida pelos eventos extremos de secas meteorológicas identificados pela pela segunda componente, parte inferior, do SPI3

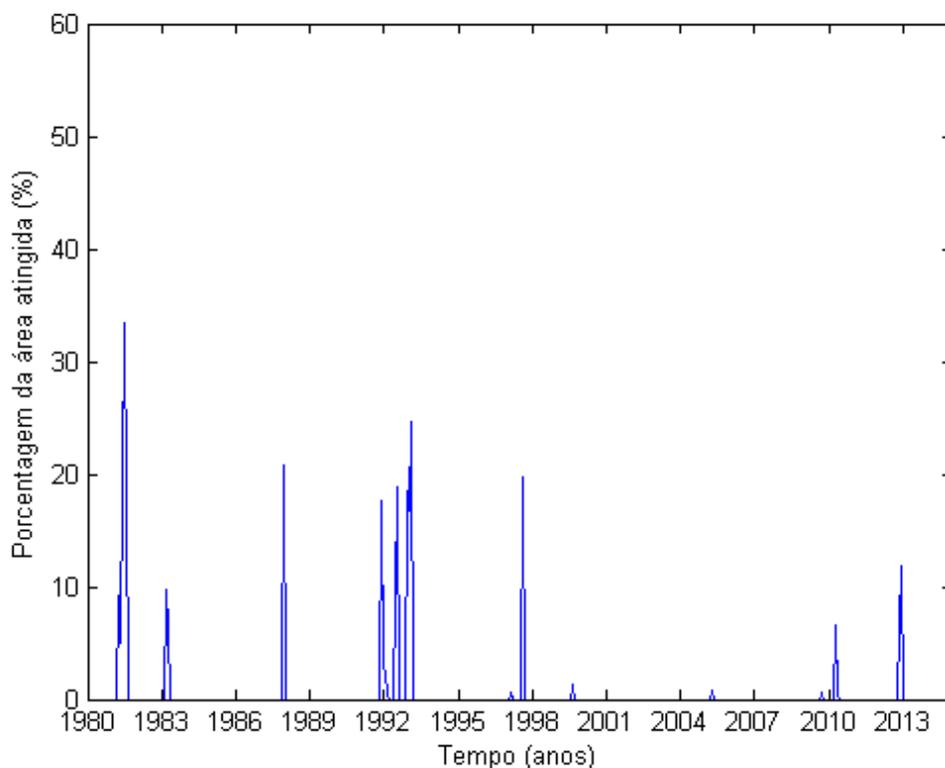


Figura 7.8 - Porcentagem da área do Nordeste atingida pelos eventos extremos de secas meteorológicas identificados pela segunda componente, parte superior, do SPI3

O Nordeste sofreu uma seca plurianual entre os anos de 1979 (não analisado neste trabalho) e de 1983, que afetou grande número de pessoas no Semiárido e produziu vários impactos na agricultura e na pecuária (CGEE, 2012; Marengo *et al.*, 2016). As duas componentes do SPI detectaram secas extremas em alguns meses entre os anos de 1980 a 1983. Na primeira componente, as secas corresponderam a um total de 9 meses dentre os 27 meses analisados, englobando todos os quatro anos. A seca de junho de 1983, detalhada na Figura 7.9, atingiu cerca de 40,18% da região nordestina. O mês seguinte, julho/1983, ainda apresentou uma grande área (38,95%) de seca severa ou extrema. A segunda componente, parte inferior e que atinge mais severamente a região sul do Nordeste (vide sinais dos pesos, Figura 7.2), sinalizou apenas 2 meses no ano de 1982, mas que corresponderam às maiores áreas (21,76% e 35,12%) atingidas de todos os eventos identificados pela componente. Já a segunda componente, parte superior, detectou 5 meses do ano de 1981 como extremos, sendo que julho de 1981 foi o que apresentou a maior área identificada pela componente (33,35%).

Nos anos de 1986 e 1987 houve uma seca na região, mas com poucos impactos (Marengo *et al.*, 2016). Ela foi detectada apenas pela segunda componente do SPI, representando uma área

considerável (20,78%), apenas em dezembro de 1987, para a parte superior da componente (secas mais severas na região norte do Nordeste, vide Figura 7.2). Brito *et al.* (2018) identificaram que o período 1986-1991 apresentou secas mais suaves que os demais, estando de acordo com os resultados aqui obtidos.

O período de 1990 a 1995 foi de uma seca longa, mais intensa em 1992-1993. Especificamente no ano de 1993, o fenômeno foi ligado ao *El Niño* (Marengo *et al.*, 2016). A primeira componente sinalizou meses dos anos de 1990, 1992 e 1993 como extremos, sendo que março e abril de 1993 foram, respectivamente, os eventos com a maior e a terceira maior abrangência espacial dos 27 considerados, tendo sido detalhados nas Figuras 7.10 e 7.11. A segunda componente, especialmente parte superior, indicou alguns meses dessa seca, com áreas expressivas em 1992 e 1993 (como 18,86% em julho/1992 e 24,66% em fevereiro/1993), no norte do Nordeste.

A seca de 1998 espalhou-se por vários estados e afetou a agricultura e a economia, sendo considerada a segunda seca mais severa do período de análise (Marengo *et al.*, 2016). A primeira componente do SPI detectou 3 meses, de abril a junho de 1998, como de seca severa ou extrema, período correspondente a uma área afetada de mais de 32% nos dois primeiros meses. A segunda componente sinalizou o mês de janeiro de 1998, com uma área afetada pequena, de 3,14%, localizada na parte sul do Nordeste.

As secas de 2005 e de 2007, citadas em Marengo *et al.* (2016), foram identificadas pela segunda componente, parte superior, e pela primeira componente, respectivamente. A seca de 2010 foi meteorológica e apresentou poucos impactos para a população, tendo sido localizada principalmente no Ceará (CGEE, 2012). Ela foi detectada apenas pela segunda componente e não representou uma grande área (17,49%, 6,53% e 0,44% para janeiro, abril e maio, respectivamente).

A seca do Nordeste que começou em 2012-2013 e tem persistido até 2018 é considerada a mais severa das últimas décadas e afetou, nos anos iniciais, cerca de 38% da população da região (Marengo e Bernasconi, 2015). Dos 27 eventos apontados pela primeira componente, 8 correspondem à seca atual, com 5 meses do ano de 2012 e 3 de 2013, o que mostra a capacidade do índice SPI em detectar essa seca. O mês de maio de 2012 apresentou a segunda maior porcentagem (51,08%) de área do Nordeste atingida por uma seca severa/extrema (Figura 7.12).

A segunda componente, parte superior (secas no norte do Nordeste), também sinalizou 2 meses da seca: novembro de 2012 (6,68%) e janeiro de 2013 (11,79%).

Das secas reportadas pela literatura (Marengo *et al.*, 2016; Banco Mundial, 2015; CGEE, 2012) como as principais ao longo do período de análise (1980 a 2013), apenas a seca de 2001-2002 não foi identificada pelo SPI. Por outro lado, a segunda componente do índice SPI identificou meses de seca severa/extrema nos anos de 1994-1996, 1999, 2008 e 2009, que não são reportados na literatura.

A análise de extremos foi realizada para obter uma ideia inicial da capacidade do SPI em retratar secas históricas, porém existem fontes de erro envolvidas. Inicialmente, pode-se citar o fato de que os registros históricos são anuais, enquanto os dados utilizados são mensais. Uma possível resolução para esse problema seria transformar os dados do SPI para uma escala anual e avaliar apenas a ocorrência ou não de seca, como foi feito por Canamary (2015). A autora analisou o acontecimento de uma seca com base nos valores dos índices durante a estação chuvosa, pois é esse período que determina se um ano será ou não seco. Além disso, o SPI de 3 meses é capaz de identificar secas meteorológicas, mas não existe clareza acerca de qual foi o tipo de seca observado em cada evento citado na literatura, já que não há um detalhamento grande dos registros.

Seria importante avaliar, ainda, a ocorrência de erros de falso alarme (quando o índice detecta uma seca que não ocorreu) ou de *missed* (quando o índice não detecta uma seca que aconteceu) (Canamary, 2015). Por fim, a análise foi feita considerando apenas a ocorrência ou não de uma seca, com base na severidade, enquanto outras propriedades são também importantes na gestão de risco, como a capacidade de um índice em detectar o início e o término de um período seco. A avaliação dessa transição é fundamental e pode mostrar diferentes situações: início correto, início falso, término correto e término falso (Canamary, 2015).

7.2.2 - Eventos de maior abrangência espacial

A seca de junho de 1983 (Figura 7.9) atingiu especialmente o estado do Piauí, com valores extremos de SPI (cerca de -4), além de partes dos estados do Maranhão e do Ceará. Os demais estados do Nordeste foram menos afetados. Brito *et al.* (2018) também encontraram uma maior severidade na parte central do semiárido do Nordeste durante o evento de seca de 1982-1983.

A seca de março de 1993 (Figura 7.10) apresentou valores de SPI em torno de -2, indicando uma seca de menor magnitude do que a seca de junho de 1983, mas com uma maior abrangência espacial, afetando os estados do Rio Grande do Norte, da Paraíba, de Pernambuco, de Alagoas, de Sergipe, do Ceará e da Bahia. A seca de abril de 1993 (Figura 7.11) apresentou comportamento semelhante, mas com uma menor área atingida. Observa-se um predomínio de condições mais severas na parte central e leste, estando de acordo com os resultados encontrados por Brito *et al.* (2018). A seca de maio de 2012 (Figura 7.12) apresenta-se bem espalhada em todos os estados da região, fato coerente com o encontrado por Brito *et al.* (2018), ocupando parte considerável dos estados do Rio Grande do Norte e da Paraíba. Os valores do SPI apresentaram magnitude semelhante às das secas de 1993. Nota-se que o Ceará foi um dos estados atingidos pela seca em maio de 2012, fato que também foi citado em Banco Mundial (2015) e em Brito *et al.* (2018). Brito *et al.* (2018) afirmaram que os outros dois estados mais afetados foram Bahia e Pernambuco, mas não foi possível chegar a essa conclusão neste trabalho. CGEE (2012) trouxe o dado de que as secas seriam mais comuns no Estado no Ceará e de que todas as secas do período estudado não ficaram restritas a apenas um estado, afirmativas que foram confirmadas nesta análise.

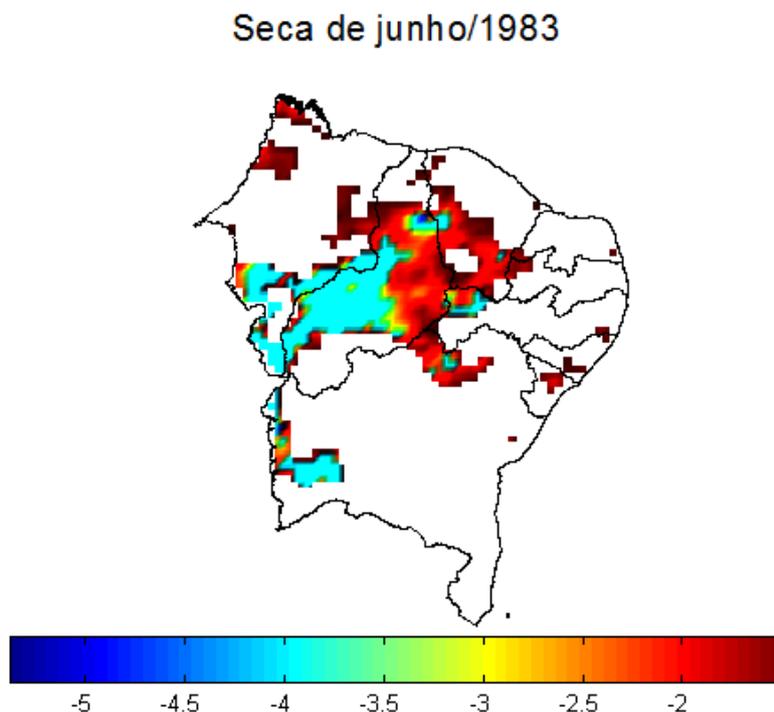


Figura 7.9 - Abrangência espacial do SPI3 na seca de junho/1983 (40,18% da área atingida)

Seca de março/1993

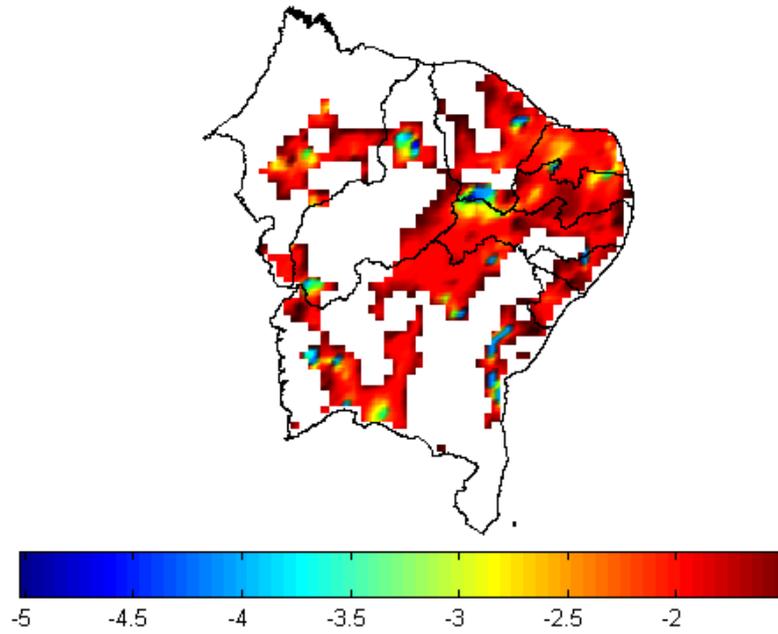


Figura 7.10 - Abrangência espacial do SPI3 na seca de março/1993 (56,43% da área atingida)

Seca de abril/1993

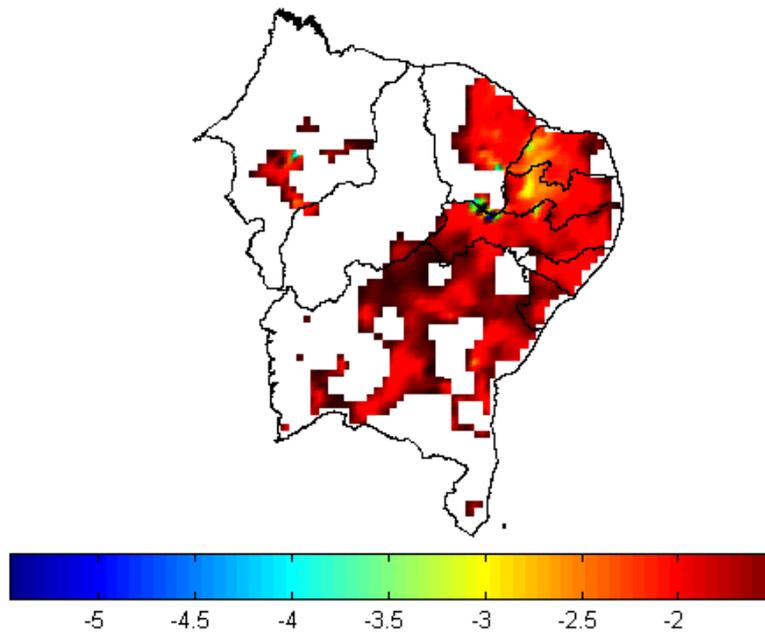


Figura 7.11 - Abrangência espacial do SPI3 na seca de abril/1993 (49,90% da área atingida)

Seca de maio/2012

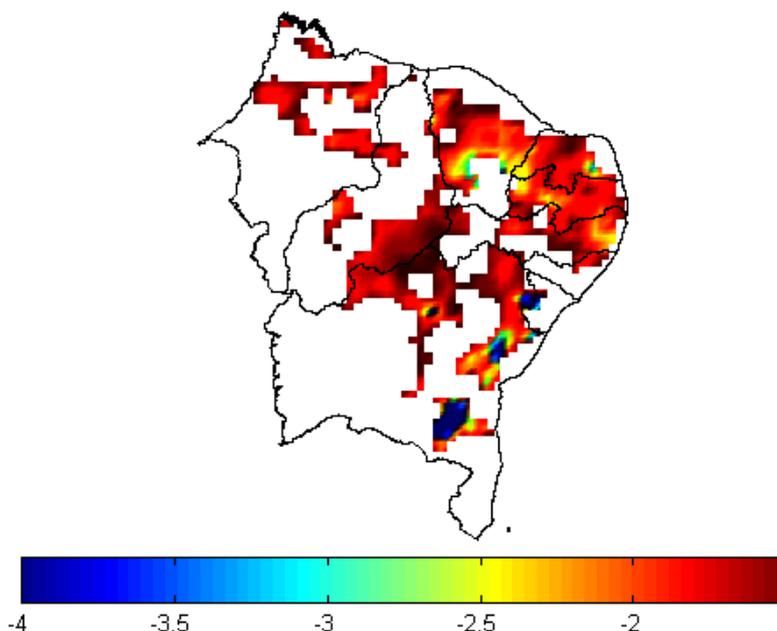


Figura 7.12 - Abrangência espacial do SPI3 na seca de maio/2012 (51,08% da área atingida)

7.3 - ANÁLISE COMPÓSITO

As Figuras 7.13, 7.15 e 7.17 mostram o resultado da análise compósito para a precipitação sobre o Nordeste em relação aos eventos extremos sinalizados pelas duas primeiras componentes do SPI (primeira componente, segunda componente - parte inferior e segunda componente - parte superior, respectivamente). Já as Figuras 7.14, 7.16 e 7.18 mostram a mesma análise para a variável temperatura.

Para a primeira componente (Figura 7.13), observa-se um comportamento homogêneo, com anomalias predominantemente negativas de precipitação (déficit de precipitação) em praticamente todo o Nordeste. Esse comportamento está coerente com os padrões espaciais dos pesos da primeira componente do SPI (Figura 7.1), que apresenta uma seca meteorológica homogênea sobre todo o Nordeste. Espera-se, de fato, que uma seca meteorológica coincida com precipitação abaixo da média. Apenas uma pequena parcela da região, localizada no Maranhão, apresenta anomalias positivas de precipitação.

Olhando para a análise compósito feita com a temperatura (Figura 7.14), nota-se um predomínio de anomalias positivas de temperatura por toda a região, o que também é coerente com os

padrões espaciais dos pesos (Figura 7.1) e com o descrito pela literatura (Mishra e Singh, 2010; Marengo *et al.*, 2013; Marengo e Bernasconi, 2015), que associa temperaturas mais altas com a ocorrência de secas. As anomalias mais altas estão concentradas na parte central, coincidindo com a região semiárida do Nordeste, o que está de acordo com o fato da maior parte das secas ocorrer nessa área (CGEE, 2012). As poucas faixas de anomalias negativas de temperatura coincidem com as de anomalias positivas de precipitação.

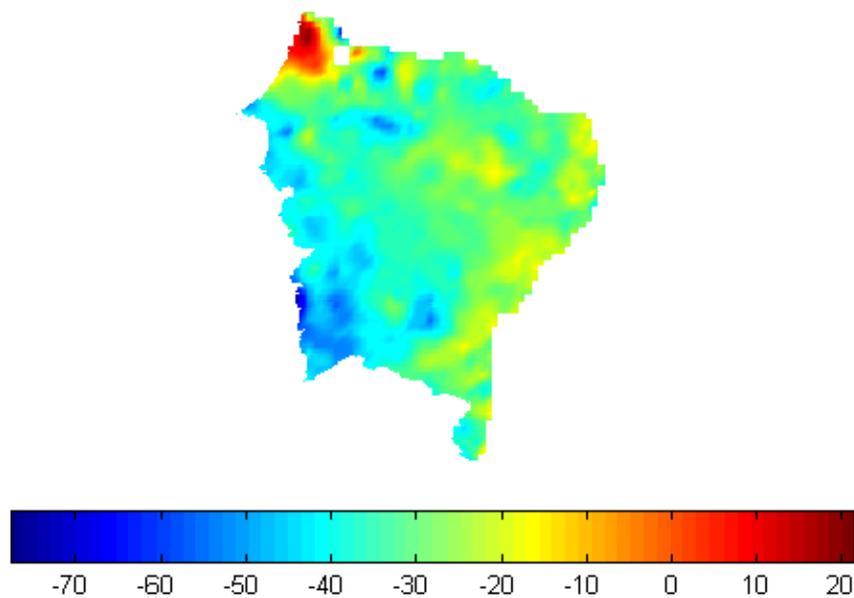


Figura 7.13 - Anomalias de precipitação nos extremos da primeira componente do SPI3

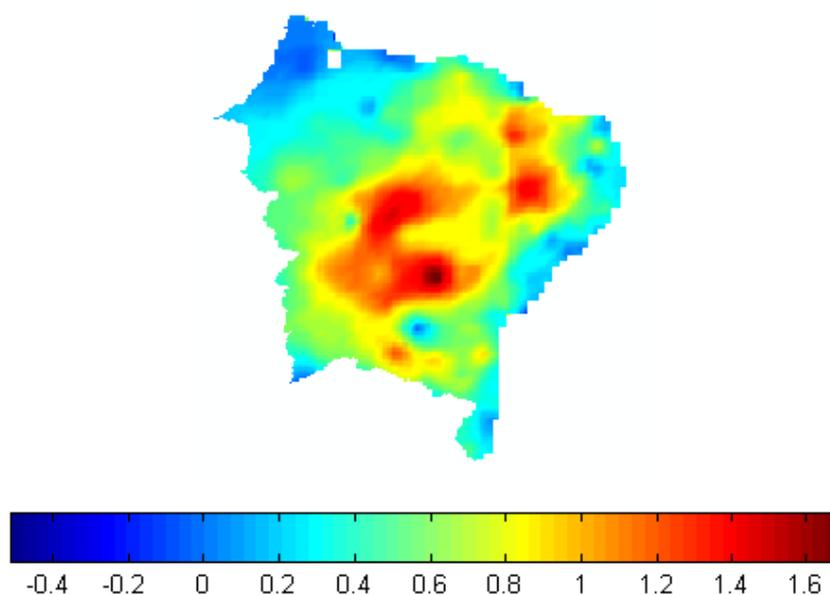


Figura 7.14 - Anomalias de temperatura nos extremos da primeira componente do SPI3

Para a segunda componente principal, parte inferior, observa-se um comportamento do tipo dipolo, com precipitação abaixo da média (anomalias negativas) (Figura 7.15) e temperaturas mais altas (anomalias positivas) (Figura 7.16) ao sul e precipitação acima da média e temperaturas mais baixas ao norte, indicando condições mais secas ao sul do que ao norte. Na análise de extremos, a segunda componente principal, parte inferior, corresponde aos valores menores e, portanto, mais negativos da componente. Como os pesos para a segunda componente principal (Figura 7.2) são negativos ao norte e positivos ao sul, a parte inferior da componente corresponde a secas no sul do Nordeste, o que está coerente com os resultados da análise compósito.

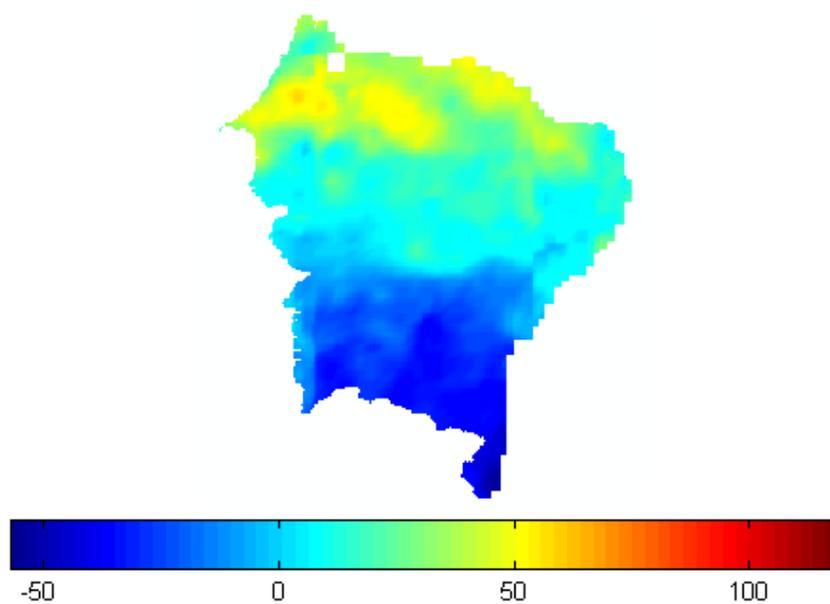


Figura 7.15 - Anomalias de precipitação nos extremos da segunda componente, parte inferior, do SPI3

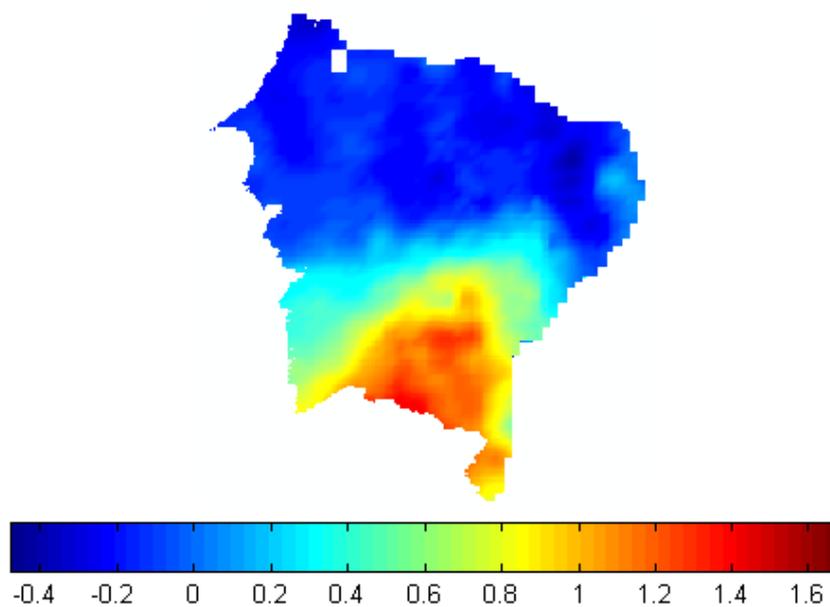


Figura 7.16 - Anomalias de temperatura nos extremos da segunda componente, parte inferior, do SPI3

Na análise compósito da segunda componente principal, parte superior, o padrão dipolo está invertido: precipitação abaixo da média (Figura 7.17) e temperaturas mais altas (Figura 7.18)

no norte. Os valores extremos mais altos correspondem aos valores positivos da componente principal e, como os pesos são negativos ao norte (Figura 7.2), eles representam secas ao norte, estando coerente com as anomalias de precipitação e de temperatura.

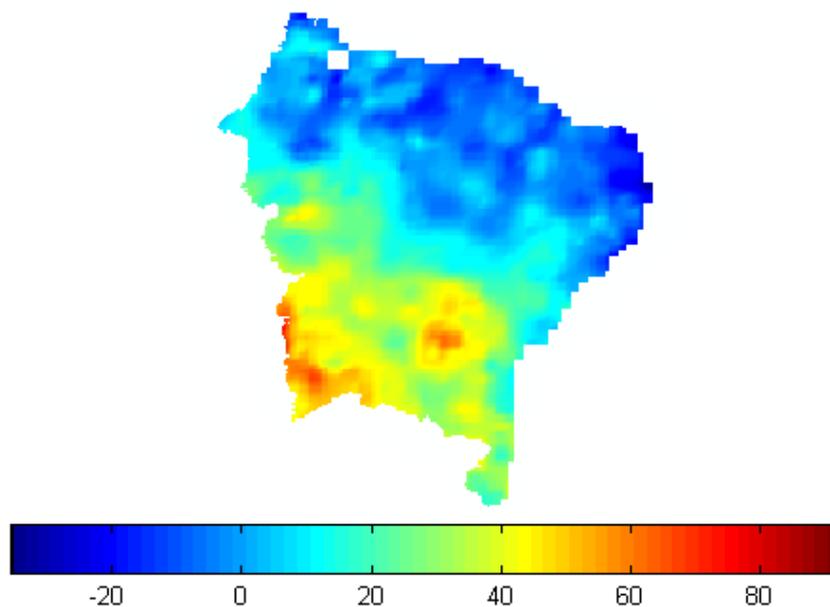


Figura 7.17 - Anomalias de precipitação nos extremos da segunda componente, parte superior, do SPI3

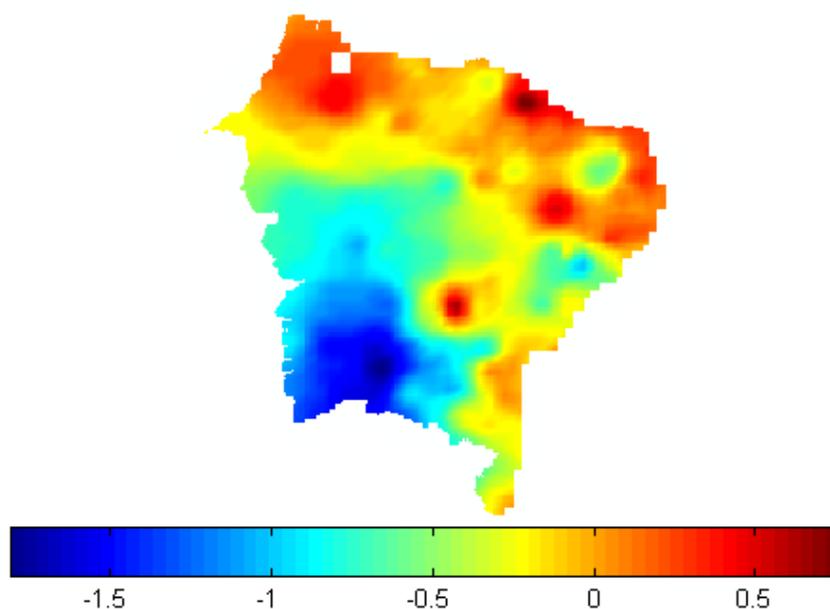


Figura 7.18 - Anomalias de temperatura nos extremos da segunda componente, parte superior, do SPI3

7.4 - INFLUÊNCIA DE INDICADORES CLIMÁTICOS

As Figuras 7.19 a 7.22 mostram as correlações cruzadas entre os índices Dipolo (Figuras 7.19 e 7.20) e NINO3.4 (Figuras 7.21 e 7.22), respectivamente, e as duas primeiras componentes do SPI. O índice que é citado em segundo lugar nos títulos é o que está adiantado (no caso, as componentes principais do SPI estão adiantadas) e as linhas pretas pontilhadas mostram as faixas a partir das quais as correlações são consideradas estatisticamente significantes ao nível de significância $\alpha = 5\%$. Observa-se que a maior parte dos valores de correlação encontra-se na faixa de significância estatística.

A Figura 7.19 apresenta todos os valores de correlação negativos, sugerindo que valores positivos do dipolo (Atlântico Norte Tropical mais quente do que o Atlântico Sul Tropical) estão associados com valores negativos da primeira componente do SPI, ou seja, relacionam-se a secas (como os pesos associados à primeira componente são positivos, valores negativos da componente levam a valores também negativos do índice de seca, vide Figura 7.1). Esse padrão de TSM provoca o deslocamento da ZCIT para o norte, reduzindo a precipitação sobre o Nordeste, como indicado na literatura (Hastenrath e Heller, 1977; Hastenrath *et al.*, 1984; Moura e Shukla, 1981; Hastenrath, 1990; Rao *et al.*, 1993; Huang *et al.*, 2005; CGEE, 2012). O pico de correlação acontece no atraso $\tau = 4$ meses, sugerindo tal período para a organização entre o aquecimento da TSM, efeito na posição da ZCIT e os impactos na seca representada pelo índice SPI.

Já na segunda componente (Figura 7.20), os valores positivos da correlação indicam sinais iguais entre o índice climático e o índice de seca. Quando o Dipolo é positivo, a segunda componente do SPI também é e, como os pesos são negativos na parte norte (Figura 7.2), tem-se secas ao norte. Nessa componente, o efeito é mais rápido, com o pico ocorrendo com um mês de atraso ($\tau = 1$), sugerindo um impacto espacial de rápido efeito, que se inicia na região norte do Nordeste e que se propaga em direção a porção sul da região à medida que o tempo avança.

O efeito do índice NINO3.4 (Figuras 7.19 e 7.20) assemelha-se ao observado para o dipolo (Figuras 7.21 e 7.22), apresentando os mesmos padrões de sinais e relações com as secas na região de estudo. O comportamento está de acordo com o descrito na literatura (Rao *et al.*, 1993; Curtis, 2008; CGEE, 2012; Marengo *et al.*, 2016), em que valores positivos do índice são

associados ao fenômeno *El Niño* e à ocorrência de secas no Nordeste do Brasil, com início na parte norte da região. O pico de correlação da primeira componente ocorreu no mesmo atraso que o Dipolo ($\tau = 4$), mas, na segunda componente, aconteceu em $\tau = 5$, indicando que, para esse índice, o efeito da segunda componente é mais lento.

Ao olhar para os sinais dos índices climáticos nos cinco meses que antecederam as quatro maiores secas detalhadas no item 7.2.2, observa-se um predomínio de valores positivos do NINO3.4 (ocorrência do *El Niño*). Apenas a seca de maio de 2012 mostrou valores negativos desse índice. Com relação ao dipolo do Atlântico, observa-se que um número grande de valores negativos do índice climático em alguns dos cinco meses de atraso, principalmente nas secas de março/93 e de abril/93. Isso indica que a seca do ano em questão deve ter sido mais influenciada pelo fenômeno do *El Niño* do que pelo Dipolo, fato que já havia sido observado por Marengo *et al.* (2016).

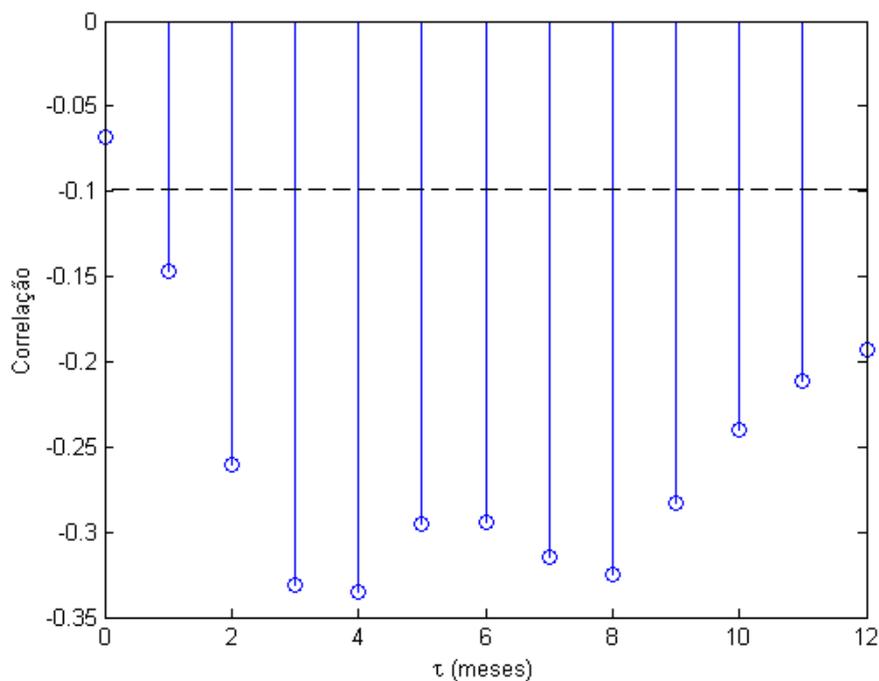


Figura 7.19 - Correlação cruzada entre o Dipolo e a primeira componente principal do SPI3, em que as linhas pretas tracejadas mostram as faixas a partir das quais as correlações são significantes ($\alpha = 5\%$)

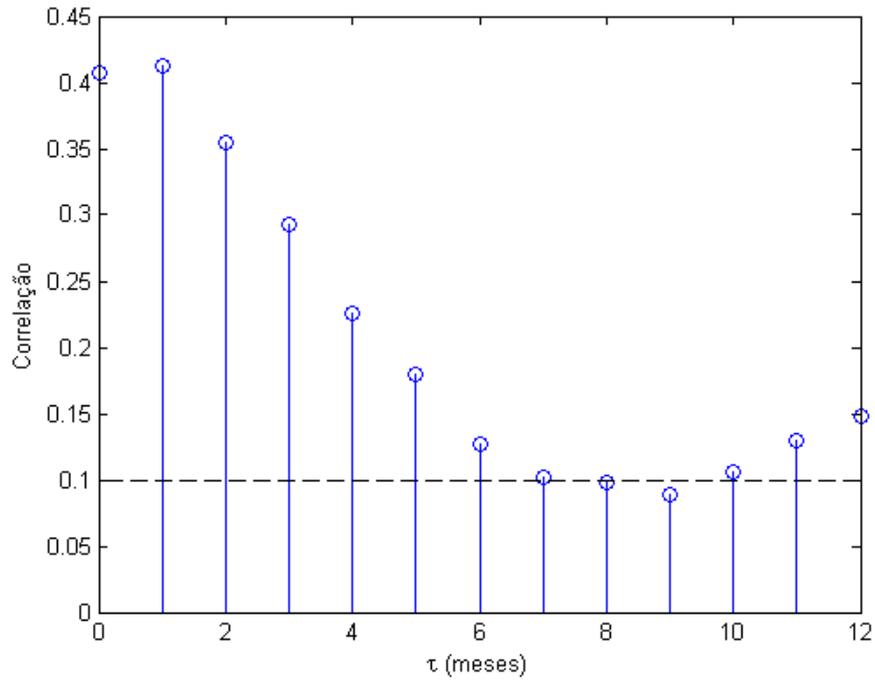


Figura 7.20 - Correlação cruzada entre o Dipolo e a segunda componente principal do SPI3, em que as linhas pretas tracejadas mostram as faixas a partir das quais as correlações são significantes ($\alpha = 5\%$)

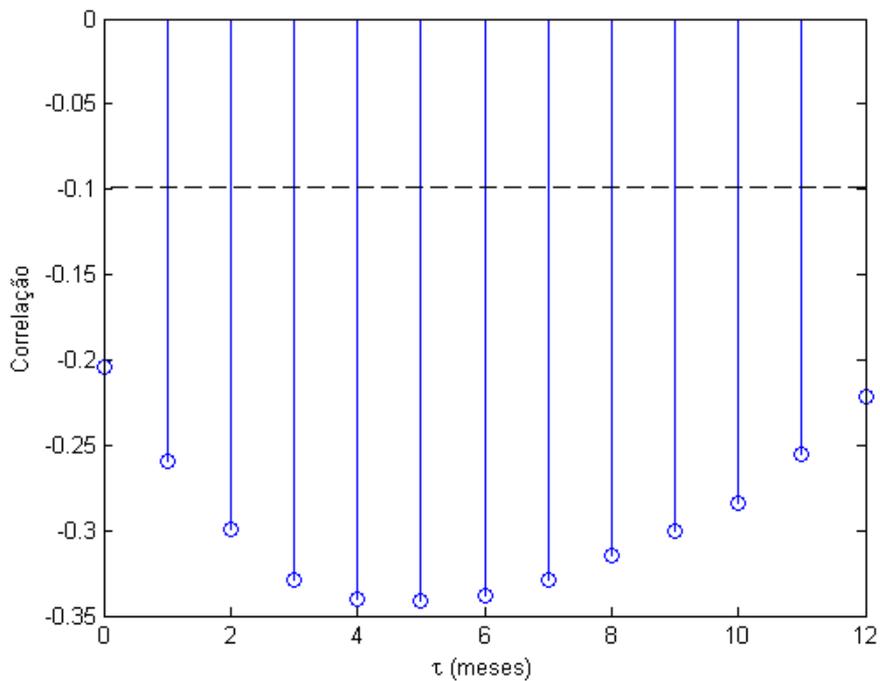


Figura 7.21 - Correlação cruzada entre o NINO3.4 e a primeira componente principal do SPI3, em que as linhas pretas tracejadas mostram as faixas a partir das quais as correlações são significantes ($\alpha = 5\%$)

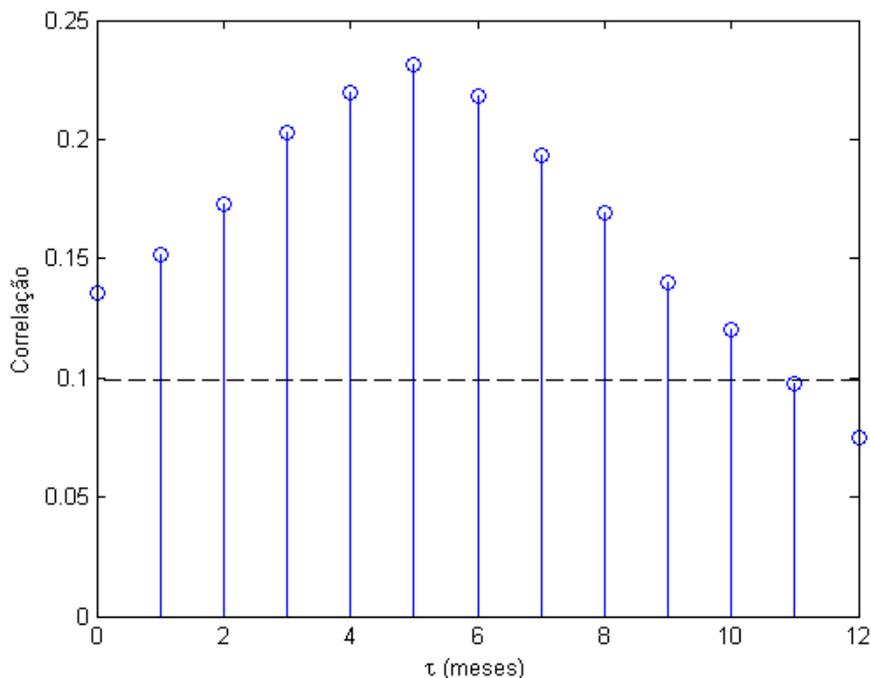


Figura 7.22 - Correlação cruzada entre o NINO3.4 e a segunda componente principal do SPI3, em que as linhas pretas tracejadas mostram as faixas a partir das quais as correlações são significantes ($\alpha = 5\%$)

7.5 - CORRELAÇÕES ENTRE AS COMPONENTES PRINCIPAIS DO SPI E A TSM

Esta análise consistiu em uma avaliação inicial da relação entre o SPI e a TSM, para fornecer as informações necessária para a construção do modelo de previsão. As Figuras 7.23 a 7.25 mostram as correlações entre a TSM mensal e as três primeiras componentes do SPI, respectivamente. A primeira componente principal (Figura 7.23) apresenta correlações negativas com a região do ENSO (Oceano Pacífico Equatorial). Como todos os pesos relacionados à primeira componente são positivos (Figura 7.1), valores negativos do SPI representam secas. Isso indica que temperaturas acima do normal nessa região do Oceano Pacífico, o que caracteriza o fenômeno do *El Niño*, estão associadas a valores negativos do SPI, ou seja, a secas. Isso está coerente com o trabalho de outros autores (Rao *et al.*, 1993; Curtis, 2008; CGEE, 2012; Marengo *et al.*, 2016) e com os resultados da análise de correlação cruzada (item 7.4), que mostraram a associação entre a ocorrência do *El Niño* e de alguns eventos de seca no Nordeste do Brasil. Observa-se, ainda, a presença de correlações positivas no Atlântico Sul Equatorial. Isso indica que, quando o oceano está mais quente, o SPI está positivo e ocorrem chuvas no Nordeste (ou, pelo menos, uma atenuação na magnitude das secas). Isso faz sentido, já que uma TSM maior estaria associada a uma maior evaporação e, conseqüentemente, a uma

maior quantidade de vapor d'água disponível para a ocorrência de precipitação na região (Rao *et al.*, 1993; Moura e Shukla, 1981).

Na segunda componente principal (Figura 7.24), pode-se observar a presença de correlações de sinais opostos na área do dipolo do Atlântico. Para analisar esse comportamento, é preciso lembrar que, na segunda componente principal, tem-se pesos negativos ao norte e positivos ao sul do Nordeste (Figura 7.2). A temperatura do Atlântico Sul apresentou valores de correlação negativos. Então, no norte do Nordeste, como os pesos são negativos, a relação da temperatura com o SPI é positiva, o que significa que, quando a TSM é maior, o SPI é positivo e tem-se chuvas na região, mostrando um comportamento semelhante ao da primeira componente. Já no sul do Nordeste, que possui pesos positivos, como a correlação é negativa, a relação é negativa. Então, quando a temperatura do Atlântico Sul é maior, o SPI é negativo e observam-se secas na parte sul da região.

A influência do Atlântico Norte, por ter apresentado correlações positivas, é inversa. No norte do Nordeste, como os pesos são negativos e a correlação é positiva, a relação é negativa, o que significa que um aumento da temperatura do Atlântico Norte gera um SPI negativo e, portanto, secas na parte norte da região. Esse comportamento é coerente com o descrito na literatura (Hastenrath e Heller, 1977; Hastenrath *et al.*, 1984; Moura e Shukla, 1981; Hastenrath, 1990; Rao *et al.*, 1993; Huang *et al.*, 2005; CGEE, 2012). De acordo com esses autores, quando o Atlântico Norte está mais quente, ocorre um deslocamento para o norte da Zona de Convergência Intertropical (ZCIT), reduzindo as chuvas sobre o Nordeste, configurando-se em um dos principais mecanismos responsáveis pela ocorrência de secas na região. Já na parte sul, os pesos são positivos e, com a correlação positiva, a relação também é positiva, indicando que um aumento da TSM corresponde a uma SPI positivo e, conseqüentemente, a chuvas no sul do Nordeste.

Ainda na segunda componente, é possível observar correlações positivas no Oceano Pacífico, havendo uma relação na região do índice NINO1+2 entre 0-10°S e 90°W-80°W, mas essa associação não é reportada na literatura. Na terceira componente (Figura 7.25), as correlações nas regiões do ENSO e do dipolo não são mais significantes. Aparece uma correlação negativa no litoral oeste da África e CGEE e ANA (2012) associam a ocorrência de secas no Nordeste com mecanismos atmosféricos advindos da região. Moura e Shukla (1981) concluíram que temperaturas mais altas no Atlântico Norte têm mais influência sobre a redução da precipitação

sobre o Nordeste do que anomalias mais quentes da TSM próxima à costa africana e isso está coerente com resultados aqui encontrados, já que a influência da TSM próxima à África aparece apenas na terceira componente principal do SPI.

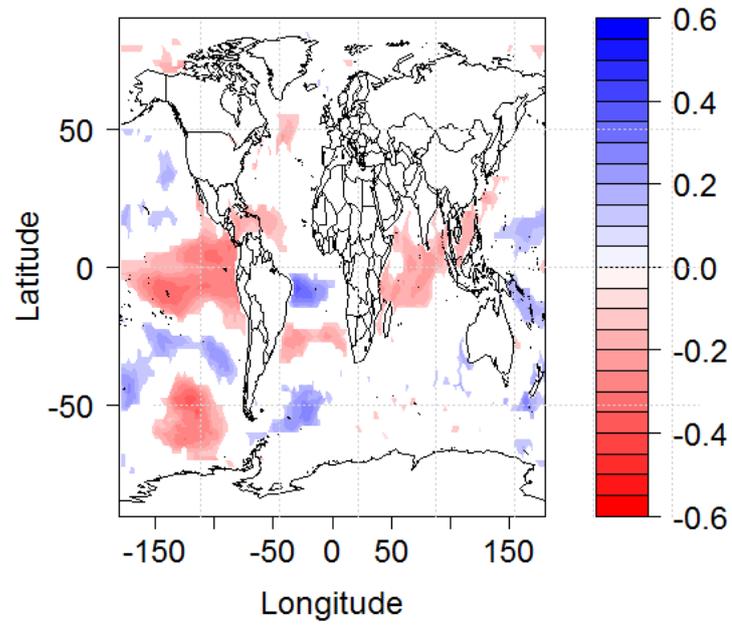


Figura 7.23 - Correlações entre a TSM e a primeira componente principal do SPI3

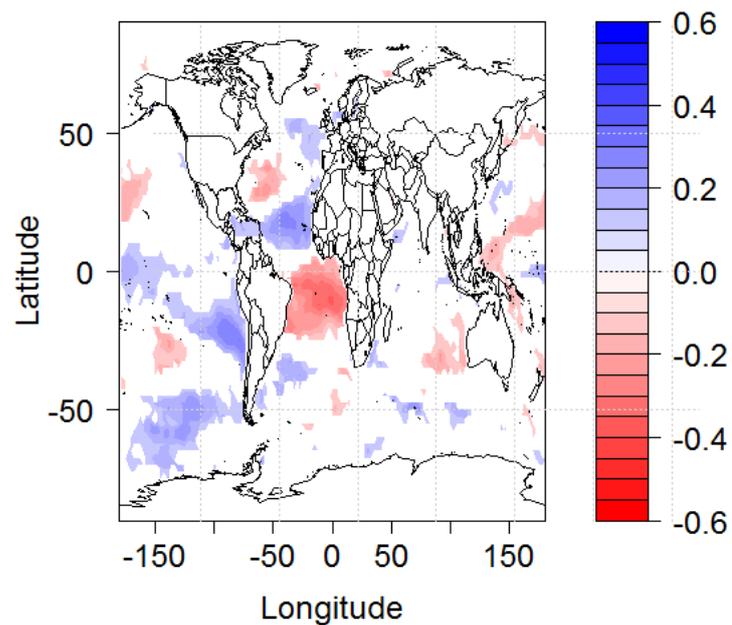


Figura 7.24 - Correlações entre a TSM e a segunda componente principal do SPI3

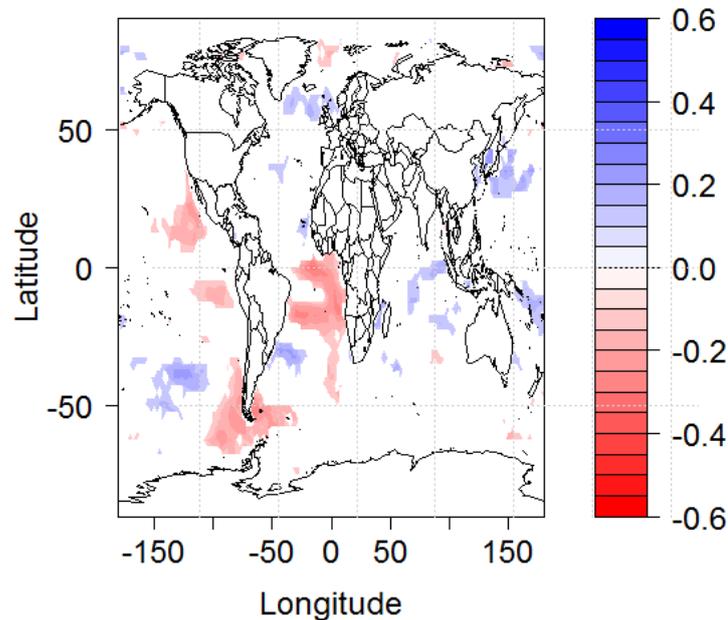


Figura 7.25 – Correlações entre a TSM e a terceira componente principal do SPI3

7.6 - SCCA

Esta seção mostra os resultados da Análise de Correlação Canônica Esparsa (SCCA, *Sparse Canonical Correlation Analysis*) para diferentes valores de *penaltyx* nos atrasos iguais a 0 e a 3 meses. O objetivo foi o de escolher o valor mais adequado de penalidade a ser usado no modelo de previsão (item 7.7).

7.6.1 - Atraso de 0 meses

Estuda-se primeiramente o atraso nulo. Esse atraso de 0 meses mostra a variação dentro de um mês, revelando mais a persistência da TSM, que demora para se alterar (variável de resposta lenta), do que o efeito em si. As Figuras 7.26 a 7.28 e 7.29 a 7.31 mostram os vetores (ou coeficientes) canônicos a_1 e a_2 , respectivamente, para três diferentes valores de *penaltyx*: 0,2 (Figuras 7.26 e 7.29), 0,3 (Figuras 7.27 e 7.30) e 0,4 (Figuras 7.28 e 7.31). Tratam-se dos coeficientes canônicos associados, respectivamente, com a primeira e com a segunda componente do SPI (vide Equações (6.9) e (6.10)). O coeficiente de correlação de Pearson ρ

das variáveis canônicas a_1X e a_2X com a primeira e a segunda componente principal do SPI, respectivamente, é indicado acima de cada figura. O *penalty* é mantido constante e igual a 0,3, por ser esse o valor que garante que a matriz B seja ortogonal (vide item 6.7 e Lima e AghaKouchak, 2017).

A medida que o valor de *penalty* aumenta, menos penalização ocorre, o número de coeficientes diferentes de zero aumenta e os padrões espaciais de correlação das Figuras 7.26 a 7.28 e 7.29 a 7.31 vão ficando mais parecido com os das Figuras 7.23 e 7.24, respectivamente. Ou seja, quanto menor a penalização, as correlações são mais próximas daquelas obtidas usando as matrizes cheias (completas), ou seja, a técnica SCCA vai se aproximando da CCA tradicional.

Olhando para as correlações de Pearson ρ obtidas, nota-se um aumento muito pequeno entre as correlações obtidas com o aumento da *penalty*. Observa-se ainda que a correlação é maior com a segunda componente do que com a primeira. Esse resultado, embora não esperado, pode indicar que a segunda componente principal está mais relacionada com a TSM do que a primeira componente.

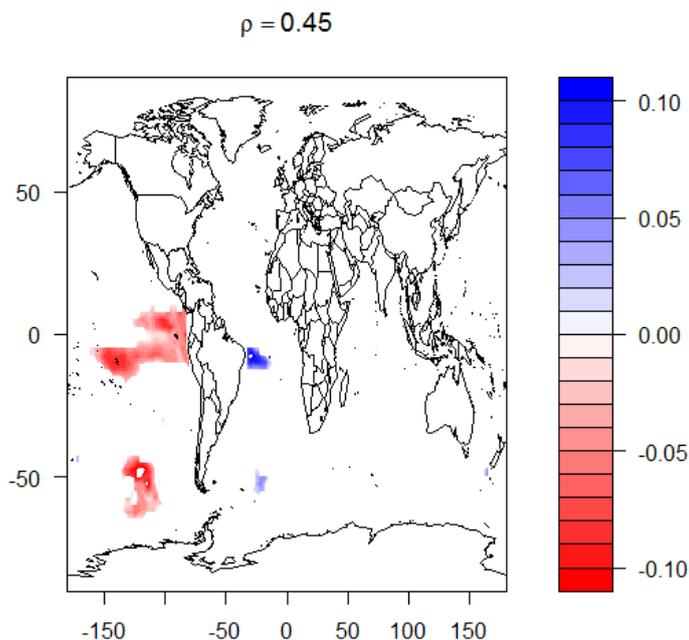


Figura 7.26 - Coeficiente canônico a_1 para valores de *penalty* = 0,2 (atraso de 0 meses)

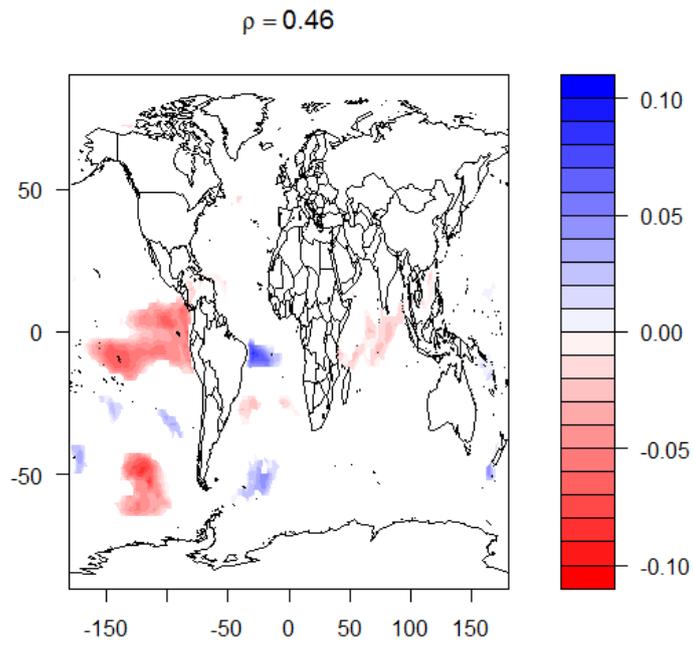


Figura 7.27 - Coeficiente canônico a_1 para valores de penaltyx = 0,3 (atraso de 0 meses)

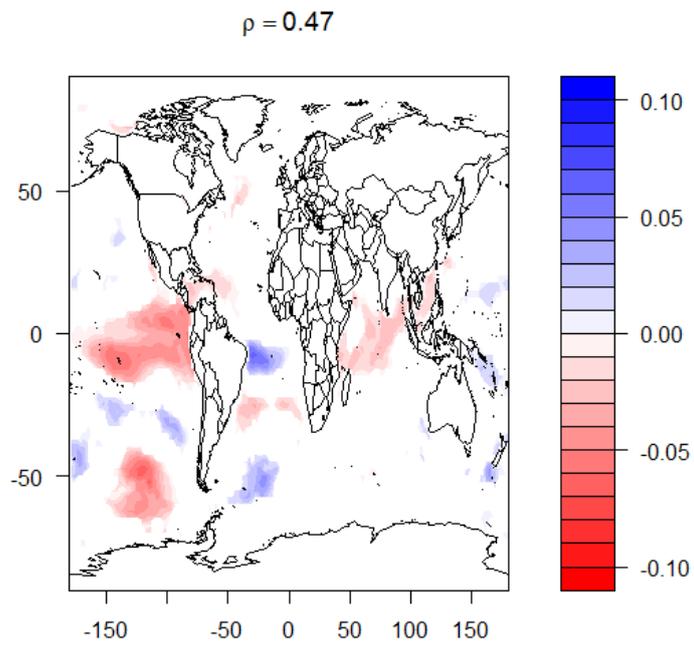


Figura 7.28 - Coeficiente canônico a_1 para valores de penaltyx = 0,4 (atraso de 0 meses)

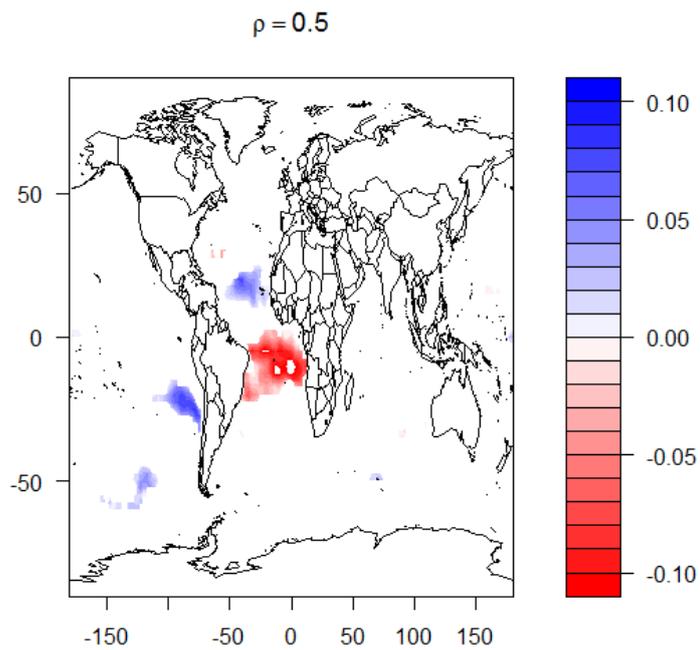


Figura 7.29 - Coeficiente canônico a_2 para valores de penaltix = 0,2 (atraso de 0 meses)

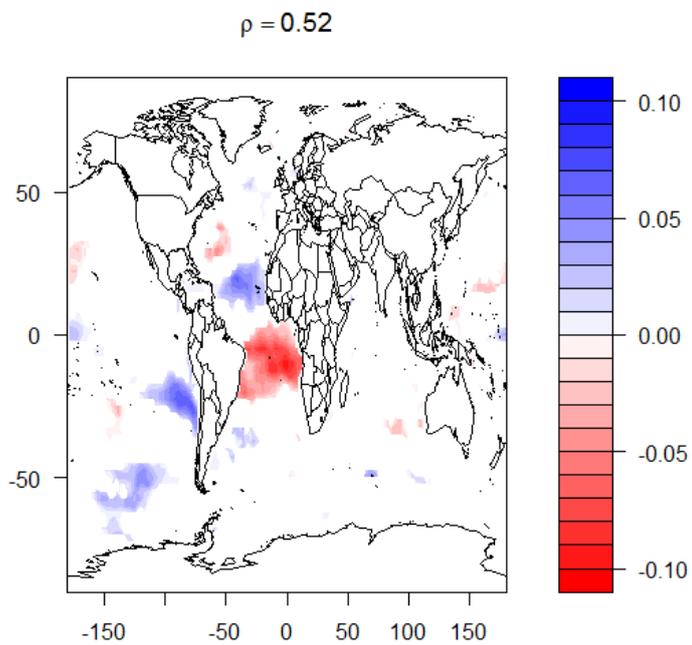


Figura 7.30 - Coeficiente canônico a_2 para valores de penaltix = 0,3 (atraso de 0 meses)

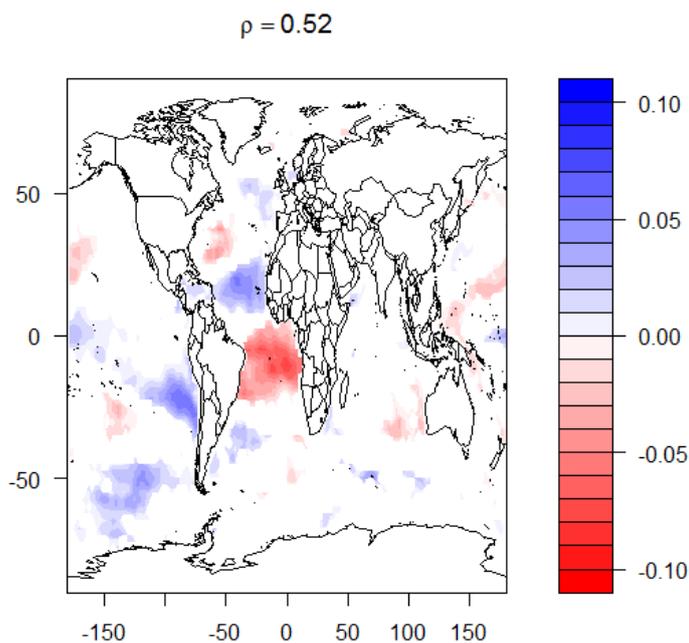


Figura 7.31 - Coeficiente canônico a_2 para valores de $penaltyx = 0,4$ (atraso de 0 meses)

7.6.2 - Atraso de 3 meses

Estuda-se agora o atraso de três meses. As Figuras 7.32 a 7.34 e 7.35 a 7.37 mostram os coeficientes canônicos a_1 e a_2 , respectivamente, para valores de $penaltyx$ de 0,2 (Figuras 7.32 e 7.35), 0,3 (Figuras 7.33 e 7.36) e 0,4 (Figuras 7.34 e 7.37). São os coeficientes canônicos associados, respectivamente, com a primeira e com a segunda componente principal do SPI. O coeficiente de correlação de Pearson ρ das variáveis canônicas a_1X e a_2X com a primeira e a segunda componente principal do SPI, respectivamente, é indicado acima de cada figura. O $penaltyy$ é mantido constante, com o valor de 0,3, pela razão explicada anteriormente (no atraso de 0 meses).

Para a primeira componente, observa-se que a correlação é maior no atraso de 3 meses do que no atraso de 0 meses, o que faz sentido, já que 3 meses pode ser o tempo necessário para a TSM ter efeito nas condições de secas dessa componente. Na segunda componente, entretanto, o resultado é o oposto: a correlação no atraso de 3 meses é inferior à do atraso nulo, pois a segunda componente está captando o efeito da TSM mais rápido. Observado as correlações para as diferentes penalizações, não é possível notar algum padrão.

Observa-se que, no atraso de 0 meses, para a $penaltyx = 0,2$, as principais correlações das Figuras 7.23 e 7.24, correspondentes à área do ENSO e da temperatura no Atlântico Sul na primeira componente e do dipolo do Atlântico na segunda, respectivamente, estão presentes, tendo havido a remoção das correlações associadas com ruídos. No atraso de 3 meses, a $penaltyx = 0,2$ mostra o mesmo padrão para a primeira componente. Na segunda componente, mesmo na $penaltyx = 0,4$, não foi possível observar o dipolo, aparecendo apenas a correlação no Atlântico Sul. Isso é uma evidência de que a temperatura no Atlântico Sul contribui mais para a ocorrência de secas do que a TSM do Atlântico Norte, então, a $penaltyx = 0,2$ é ainda a mais adequada. É interessante usar o menor valor possível de $penaltyx$, já que, quanto maior a penalização, menor é a quantidade de pontos usada no modelo de previsão, tornando os estimadores mais robustos e facilitando a interpretação dos resultados. A penalidade de 0,2, portanto, foi escolhida para ser usada no modelo de previsão (item 7.7).

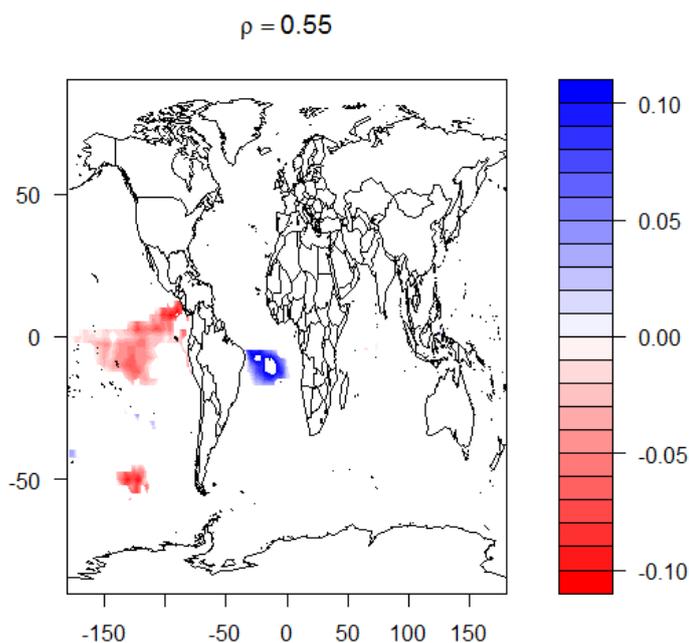


Figura 7.32 - Coeficiente canônico a_1 para valores de $penaltyx = 0,2$ (atraso de 3 meses)

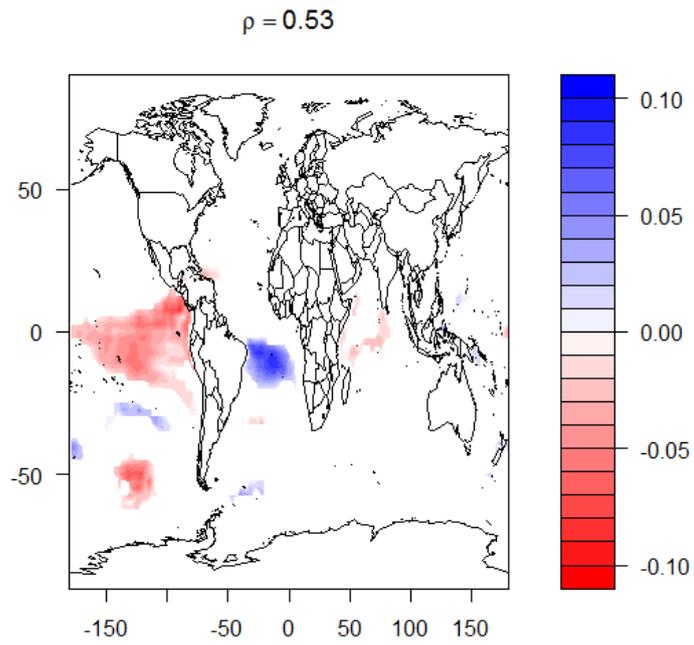


Figura 7.33 - Coeficiente canônico a_1 para valores de penaltix = 0,3 (atraso de 3 meses)

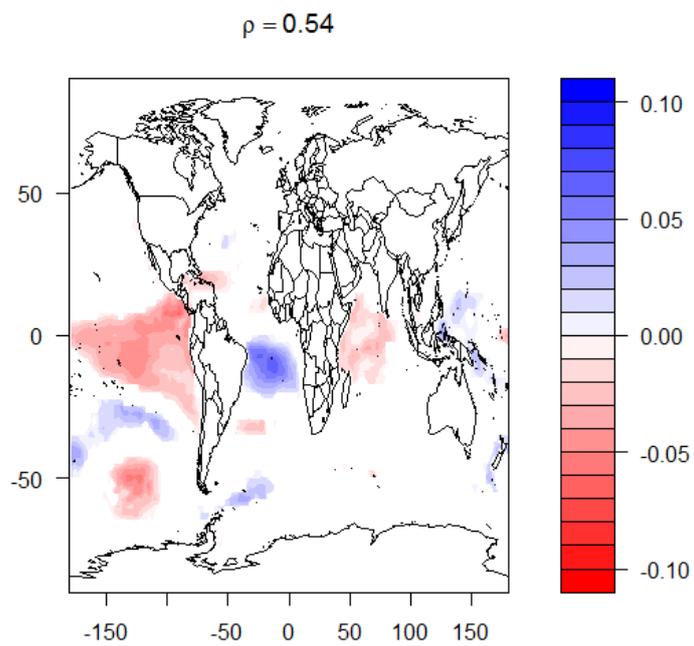


Figura 7.34 - Coeficiente canônico a_1 para valores de penaltix = 0,4 (atraso de 3 meses)

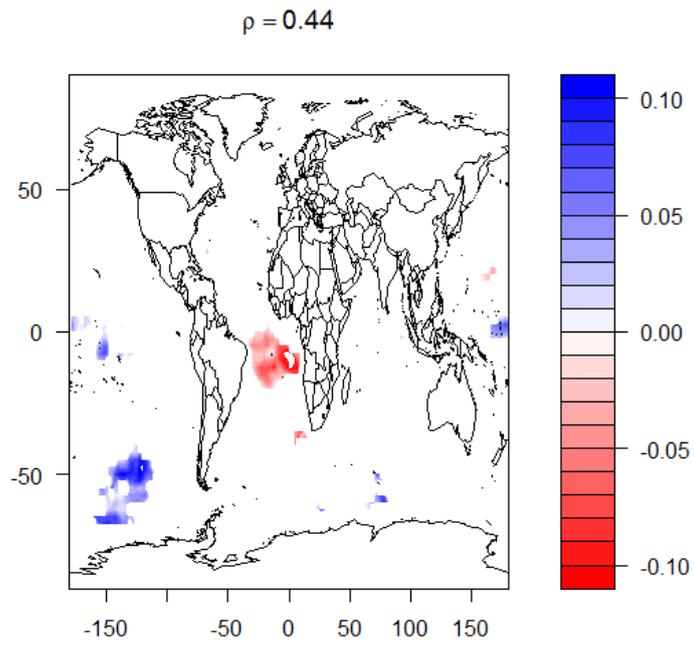


Figura 7.35 - Coeficiente canônico a_2 para valores de penaltix = 0,2 (atraso de 3 meses)

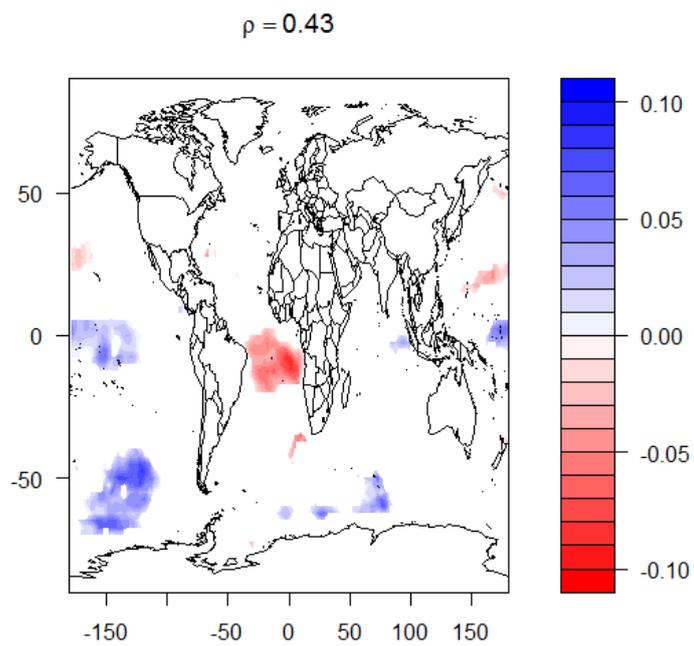


Figura 7.36 - Coeficiente canônico a_2 para valores de penaltix = 0,3 (atraso de 3 meses)

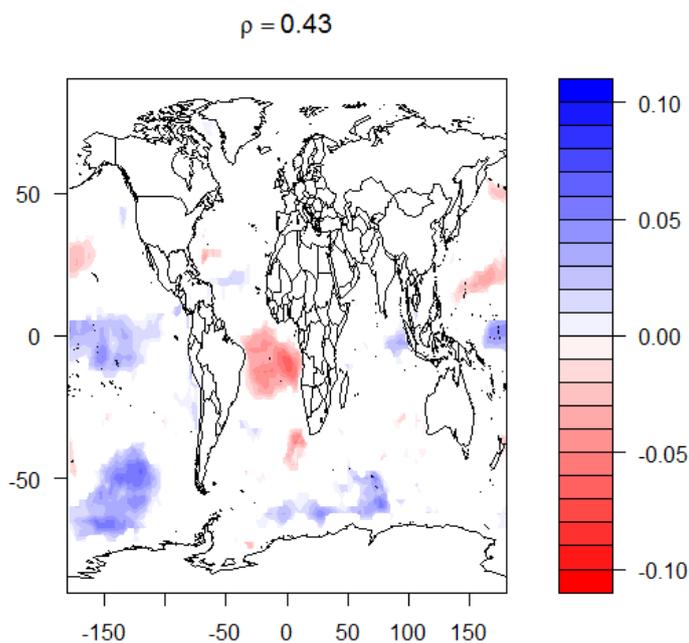


Figura 7.37 - Coeficiente canônico a_2 para valores de $penaltyx = 0,4$ (atraso de 3 meses)

7.7 - PREVISÃO DE SECAS

As Figuras 7.38, 7.39, 7.40 e 7.41 mostram as previsões para os tempos de previsão de 1, 2, 4 e 6 meses, respectivamente, obtidas usando a penalidade $penaltyx = 0,2$. As linhas em preto e em vermelho mostram o SPI observado e o previsto, respectivamente, para cada tempo de atraso. A hachura cinza indica o intervalo de previsão de 95%. Observa-se que a maior parte dos valores observados se encontra dentro dos limites do intervalo de previsão.

Observa-se que a série prevista com um mês de antecedência segue a mesma tendência geral dos valores observados, mas sem conseguir atingir os picos das séries originais, subestimando os valores mais extremos do SPI. Com o aumento do tempo de atraso, observa-se uma redução na qualidade da previsão. No gráfico para 2 meses de antecedência a forma é semelhante, mas mais distante dos valores de picos. A seca de 1998 foi detectada com um pico deslocado para o ano anterior. Em 4 meses de antecedência, a distância entre os picos previstos e observados é maior. Nos períodos de 2007-2008 e de 2012-2013, a série prevista não conseguiu captar a queda do SPI para valores negativos, indicativos de secas. O mesmo se observa para 6 meses de antecedência, em que os picos estão ainda mais distantes do observado.

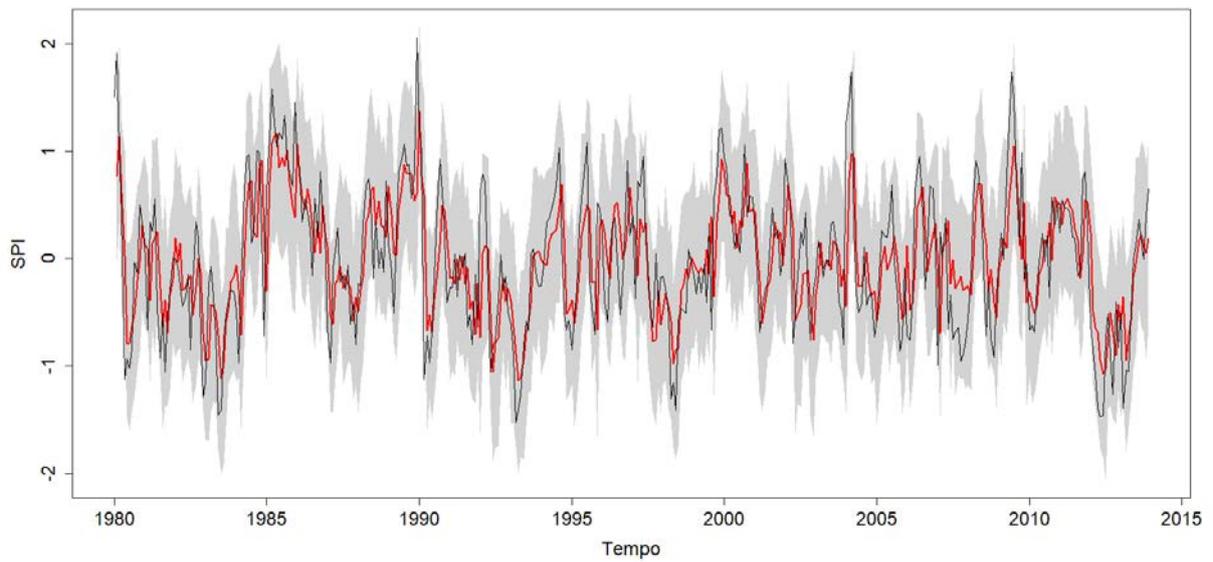


Figura 7.38 - SPI previsto pelo modelo completo (linhas azuis) e SPI observado (linhas pretas) para o tempo de previsão de 1 mês. A hachura cinza mostra o intervalo de previsão de 95%

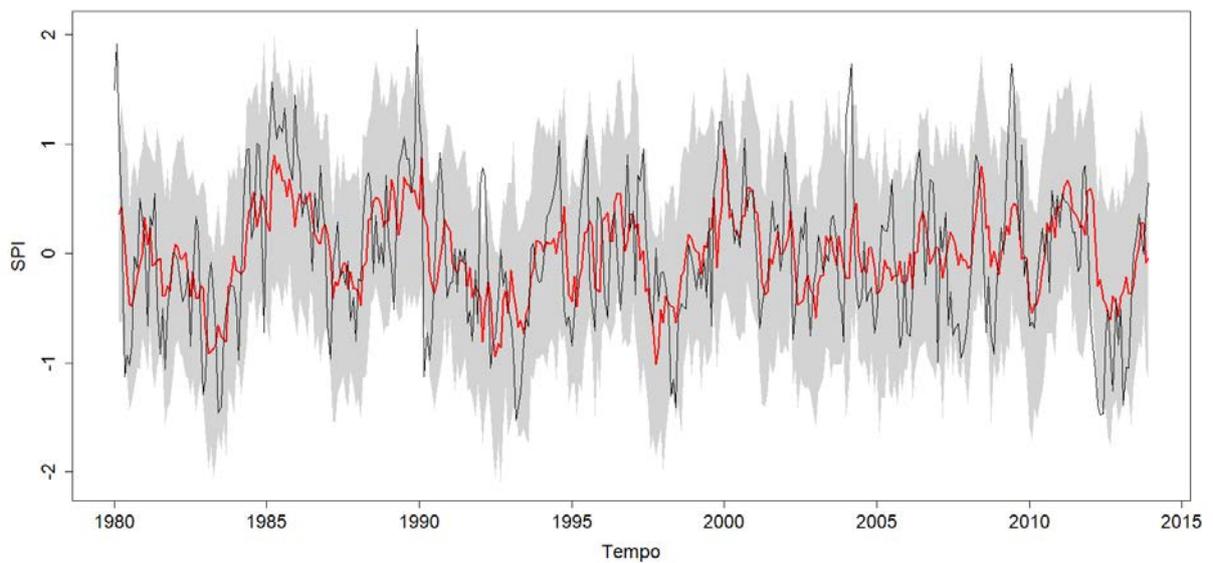


Figura 7.39 - SPI previsto pelo modelo completo (linhas azuis) e SPI observado (linhas pretas) para o tempo de previsão de 2 meses. A hachura cinza mostra o intervalo de previsão de 95%

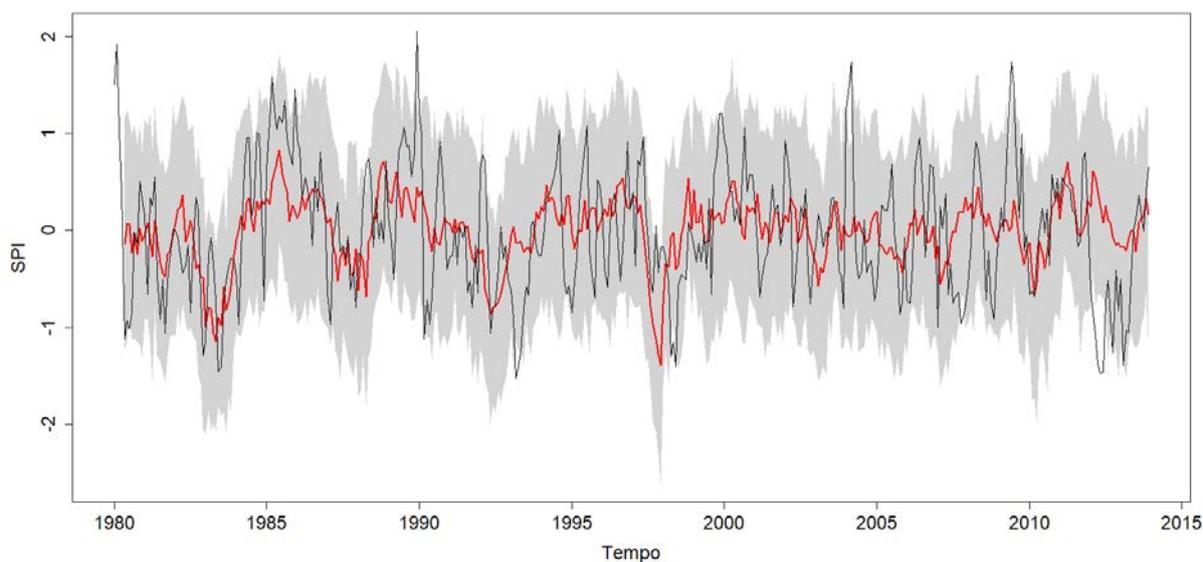


Figura 7.40 - SPI previsto pelo modelo completo (linhas azuis) e SPI observado (linhas pretas) para o tempo de previsão de 4 meses. A hachura cinza mostra o intervalo de previsão de 95%

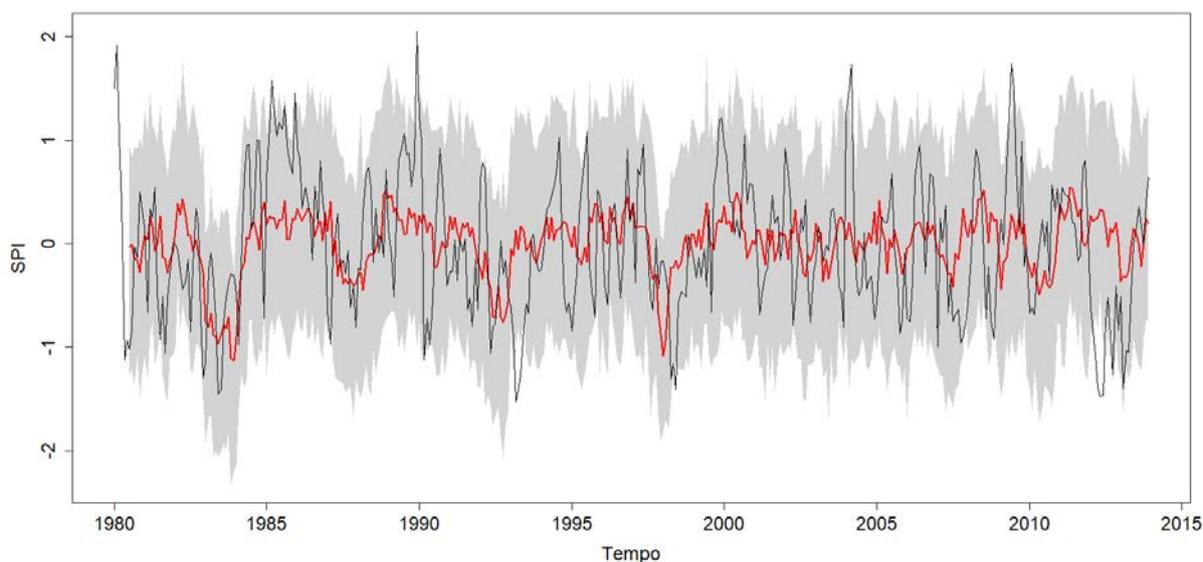


Figura 7.41 - SPI previsto pelo modelo completo (linhas azuis) e SPI observado (linhas pretas) para o tempo de previsão de 6 meses. A hachura cinza mostra o intervalo de previsão de 95%

As Figuras 7.42 e 7.43 mostram o desempenho do coeficiente de correlação e do raiz do erro médio quadrático (RMSE, *Root-Mean Squared Error*), respectivamente, para os tempos de previsão de 1 a 6 meses. As linhas pretas, vermelhas e azuis representam, respectivamente, a primeira, a segunda e a terceira componente principal do SPI, enquanto a linha verde mostra a média espacial do SPI. As linhas cheias representam os resultados para o modelo completo,

enquanto as linhas pontilhadas, para o modelo auto-regressivo. O comportamento dessas duas métricas foi de acordo com o esperado, ou seja, o coeficiente de correlação diminuiu e o RMSE cresceu com o aumento do tempo de previsão, com exceção da segunda componente, modelo completo, que apresentou um acréscimo de correlação do tempo de previsão de 3 para o de 4 meses.

Para o coeficiente de correlação, os maiores valores foram obtidos para a primeira componente principal do SPI (linhas pretas) e para o SPI médio (linhas verdes), modelo completo, os quais apresentaram valores de correlação bem semelhantes e máximos em 0,7349 e 0,7301, respectivamente, no atraso de 1 mês, caindo para 0,2894 e para 0,2777 no atraso de 6 meses. As correlações para os respectivos modelos auto-regressivos (da primeira componente e do SPI médio) são as próximas correlações em ordem decrescente, com valores máximos de 0,7203 e de 0,7166. O fato do modelo completo mostrar correlações maiores do que o auto-regressivo indica que a presença da informação climática melhora o desempenho das previsões, especialmente para tempos de atraso a partir de 2 meses. Tanto para o modelo completo quanto para o auto-regressivo, as correlações possuem uma queda menos acentuada após o tempo de previsão de 3 meses. Ou seja, observa-se que a primeira componente do SPI e o SPI médio apresentam comportamentos semelhantes, o que era esperado, já que essa componente apresenta a maior parte da variabilidade do sistema.

Ainda sobre o desempenho do coeficiente de correlação, a segunda componente principal (linhas vermelhas) apresenta-se com valores inferiores à primeira em todos os tempos de previsão, havendo uma queda mais acentuada, já que atinge correlações próximas a zero no tempo de previsão de 6 meses. Os modelos completo e auto-regressivo apresentam resultados semelhantes em todos os tempos de previsão, com o pico em aproximadamente 0,67 no tempo de previsão de 1 mês, indicando que a covariável climática não influencia no desempenho da previsão para essa componente. A terceira componente principal (linhas azuis) apresenta, inicialmente, um desempenho inferior ao da segunda e o modelo auto-regressivo apresenta-se ligeiramente melhor, mas, a partir do atraso de 3 meses, as correlações tornam-se mais altas do que para a segunda componente e o modelo completo passa a apresentar um desempenho superior ao auto-regressivo. Ou seja, para essa componente, a informação climática parece acrescentar ao modelo.

Quanto ao desempenho do RMSE, não se observa distinção entre o modelo completo e o auto-regressivo para o SPI médio (linhas verdes), estando todos os valores próximos a zero. Quanto mais próximo de zero for o RMSE, melhor é o ajuste feito, indicando que, na primeira componente principal, o modelo completo apresentou um RMSE menor do que o auto-regressivo, corroborando o que foi observado pelo coeficiente de correlação: o acréscimo da informação climática melhora o desempenho das previsões da primeira componente do SPI. Para a segunda e a terceira componente principal, a diferença entre os dois modelos é menor, com o auto-regressivo ligeiramente melhor (RMSE menor), confirmando que a covariável climática não contribui para as previsões, como já havia sido notado ao avaliar os coeficientes de correlação. Apenas para o tempo de previsão maior do que 3 meses, a informação climática parece contribuir de alguma forma na terceira componente (Figura 7.42). Esse resultado pode ser um indício de que essas duas componentes estejam relacionadas a algum outro mecanismo climático diferente da TSM.

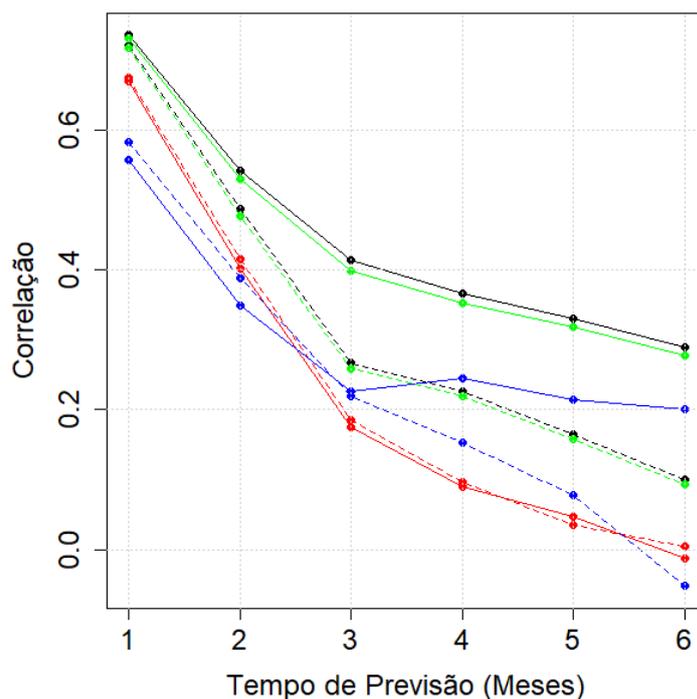


Figura 7.42 - Coeficientes de correlação em 1 a 6 meses de previsão do modelo completo (linhas cheias) e do modelo auto-regressivo (linhas pontilhadas) para a primeira (linhas pretas), para a segunda (linhas vermelhas) e para a terceira (linhas azuis) componentes principais do SPI e para a média espacial (linhas verdes) do SPI

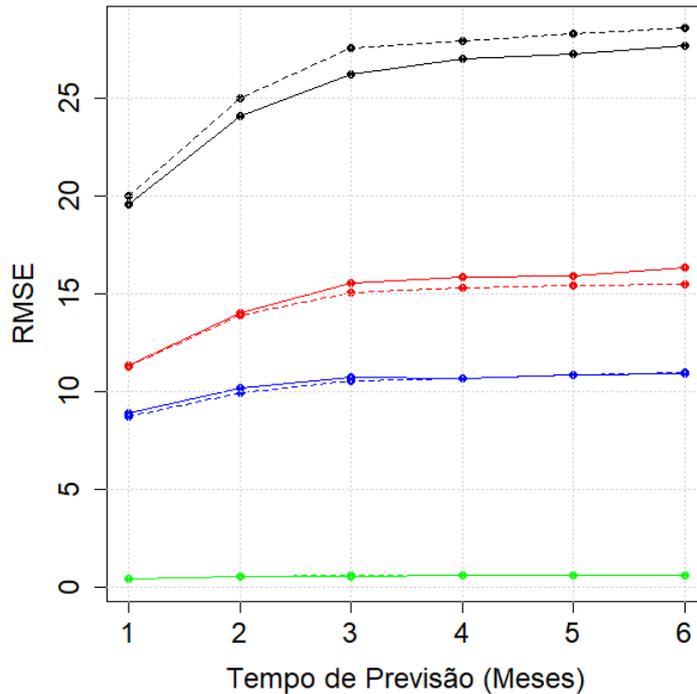


Figura 7.43 - Valores do RMSE (*Root-Mean Squared Error*) em 1 a 6 meses de previsão do modelo completo (linhas cheias) e do modelo auto-regressivo (linhas pontilhadas) para a primeira (linhas pretas), para a segunda (linhas vermelhas) e para a terceira (linhas azuis) componentes principais do SPI e para a média espacial (linhas verdes) do SPI

Para visualizar melhor os ganhos relativos do modelo completo, foi plotado um gráfico contendo o RMSE normalizado em relação ao modelo auto-regressivo (dividindo-se o RMSE do modelo completo pelo RMSE do modelo auto-regressivo), conforme mostrado na Figura 7.44. Como deseja-se que o RMSE seja o menor possível, valores abaixo de 1,00 significam uma melhora no modelo de previsão em razão do acréscimo da informação climática. O maior ganho é observado na previsão da média espacial (linha verde) e da primeira componente do SPI (linha preta), no tempo de previsão de 3 meses. A melhora, entretanto, foi semelhante para todos os tempos de previsão, tendo-se valores do RMSE normalizado entre 0,95 e 0,98 para todos os tempos de previsão da média espacial e da primeira componente.

Para a segunda componente principal prevista do SPI (linha vermelha), como todos os valores do RMSE normalizado permaneceram acima de 1,00, a informação climática não foi capaz de melhorar o desempenho das previsões em relação ao modelo auto-regressivo. Na terceira componente principal (linha azul), observa-se o mesmo, mas, a partir do tempo de previsão de 4 meses, o RMSE normalizado passa a apresentar valores menores do que 1,00. Como esses

valores, no entanto, são próximo de 1,00, não pode afirmar que houve um ganho, fazendo com que a conclusão para essa componente seja igual à da segunda componente principal.

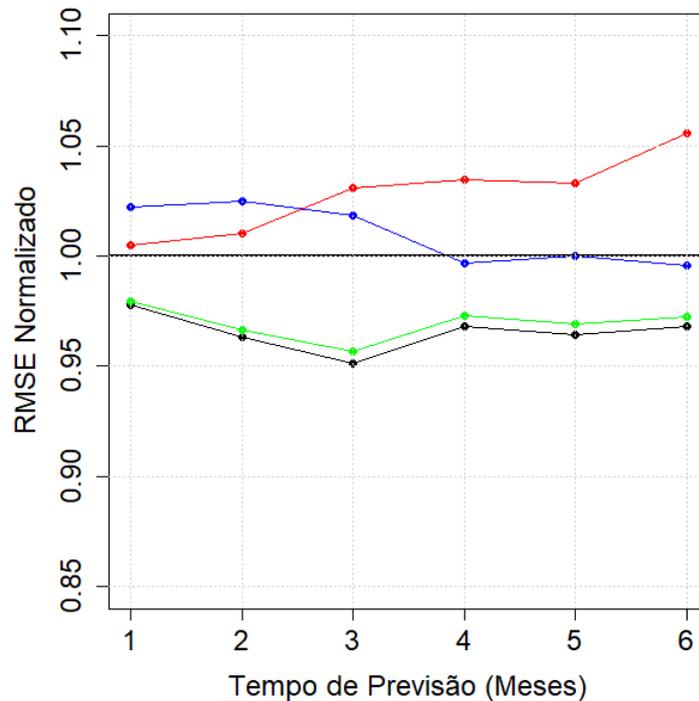


Figura 7.44 - Valores do RMSE normalizado para o modelo completo (em relação ao modelo auto-regressivo), em tempos de previsão de 1 a 6 meses, para a primeira (linhas pretas), para a segunda (linhas vermelhas) e para a terceira (linhas azuis) componentes principais do SPI e para a média espacial (linhas verdes) do SPI

As Figuras 7.45, 7.46, 7.47 e 7.48 mostram os mapas com as distribuições espaciais do BSS (*Brier Skill Score*) para tempos de previsão de 1, 2, 4 e 6 meses, respectivamente. Valores positivos do BSS indicam que as previsões são melhores do que a climatologia. Observa-se que os valores positivos se concentraram na faixa norte e um pouco no leste do Nordeste para os tempos de previsão de 1 e de 2 meses, indicando que, nessas regiões, as previsões foram melhores. Ou seja, o fato da região ao norte exibir valores positivos do BSS indica que há algum desempenho de previsão para até 2 meses de antecedência. Já a parte sul, por mostrar valores negativos do BSS, apresenta pior desempenho nas previsões. Nota-se que, com o aumento do tempo de previsão, essa faixa de valores positivos foi sendo substituída por valores negativos, indicando uma piora nos resultados das previsões. Os valores médios espaciais do BSS para esses tempos de previsão foram, respectivamente, 0,1220, 0,0345, -0,0353 e -0,0484. Isso confirma que, na média, os ajustes foram melhores do que a climatologia apenas para os tempos de previsão de 1 e de 2 meses.

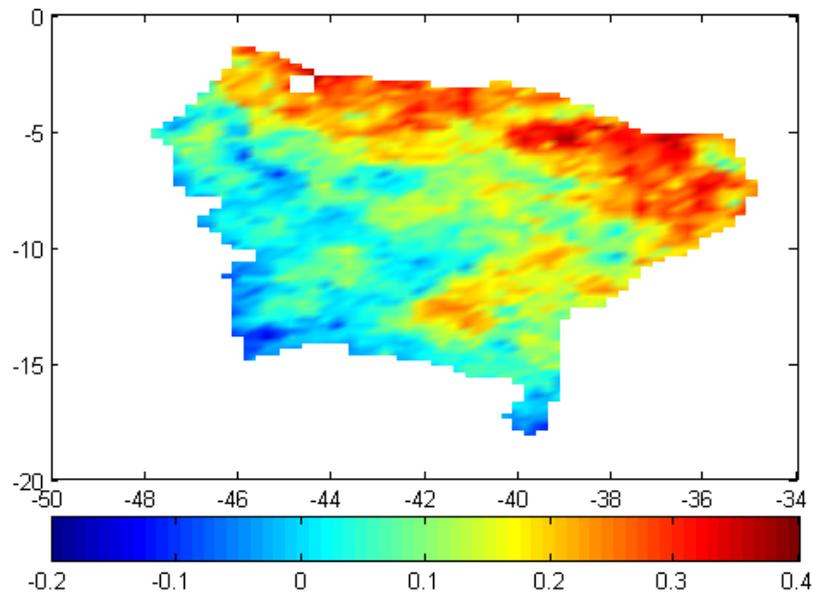


Figura 7.45 - Distribuição espacial do BSS (*Brier Skill Score*) para o tempo de previsão de 1 mês

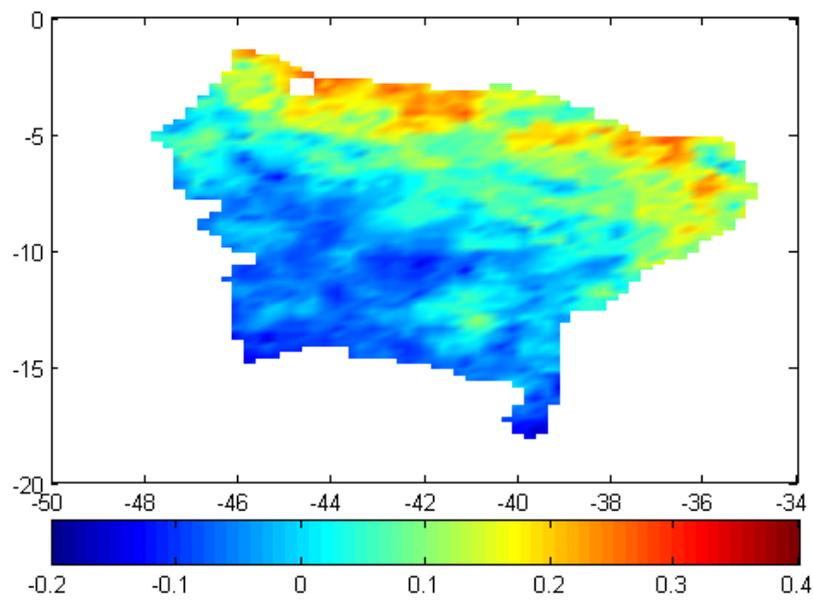


Figura 7.46 - Distribuição espacial do BSS (*Brier Skill Score*) para o tempo de previsão de 2 meses

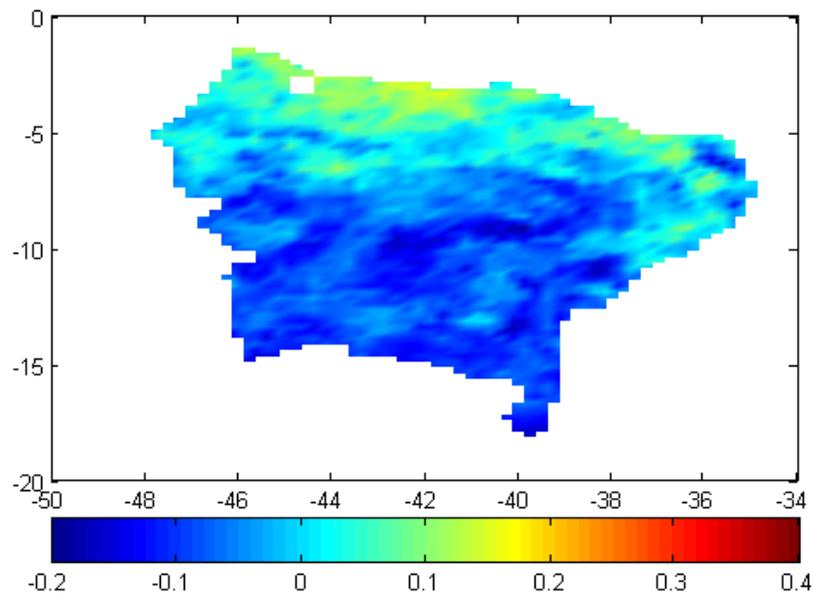


Figura 7.47 - Distribuição espacial do BSS (*Brier Skill Score*) para o tempo de previsão de 4 meses

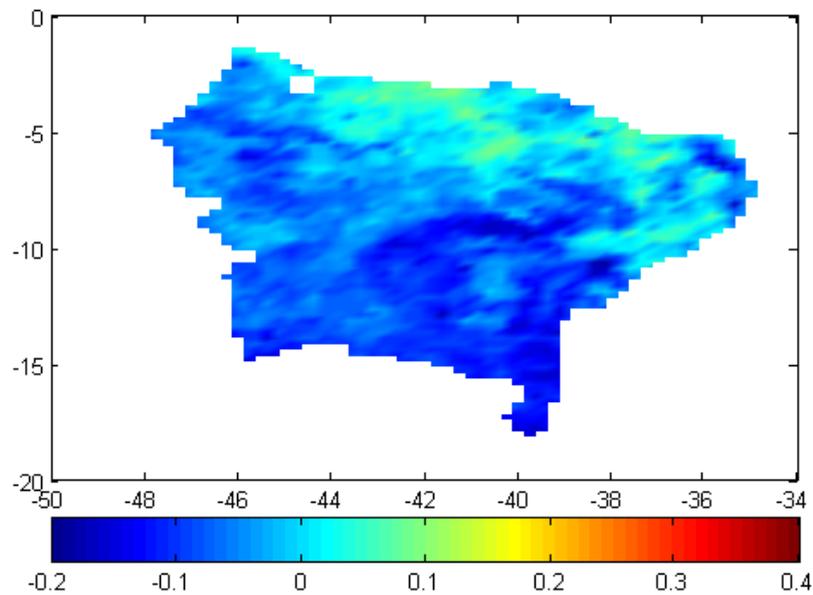


Figura 7.48 - Distribuição espacial do BSS (*Brier Skill Score*) para o tempo de previsão de 6 meses

8 - CONCLUSÕES E RECOMENDAÇÕES

O índice de seca Índice de Precipitação Padronizada (SPI, *Standardized Precipitation Index*) foi utilizado para avaliar a dinâmica espaço-temporal de secas meteorológicas no Nordeste do Brasil e a influência de forçantes climáticas de larga escala relacionadas à TSM (temperatura na superfície do mar) nos Oceanos Pacífico e Atlântico. O objetivo foi o de, a partir desse conhecimento, criar um modelo para prever o SPI e, conseqüentemente, as secas meteorológicas na região, utilizando as informações climáticas da TSM como preditores e usando a técnica Análise de Correlação Canônica Esparsa (SCCA, *sparse Canonical Correlation Analysis*).

A primeira avaliação realizada foi por meio da técnica Análise das Componentes Principais (PCA, *Principal Component Analysis*). Analisando os padrões espaciais dos pesos das três primeiras componentes principais, observou-se que a primeira componente principal do índice SPI representa as secas meteorológicas no Nordeste brasileiro seguindo um modelo de condições secas homogêneas em toda a região, explicando em torno de 41,61% da variabilidade espacial dos dados. Já a segunda componente principal apresenta um padrão do tipo dipolo norte-sul, que responde por 11,33% da variância dos dados. Por fim, a terceira componente principal mostrou um padrão zonal leste-oeste, correspondendo a 5,70% da variabilidade dos dados. Análises mais detalhadas foram feitas apenas para as duas primeiras componentes principais, por representarem juntas 52,94% da variância do índice de seca.

Em relação à análise de extremos, realizada considerando os percentis de 6,7% (correspondente a secas severas e extremas), o SPI mostrou-se adequado em identificar secas extremas em relação à severidade, visto que a maior parte dos resultados coincidiu com os registros na literatura (Moura e Shukla, 1981; CGEE e ANA, 2012; Marengo e Bernasconi, 2015; Banco Mundial, 2015). A única exceção foi a seca de 2001-2002, mas, por outro lado, o índice de seca conseguiu identificar as secas mais severas do período considerado (1998 e a atual, iniciada em 2012). Os eventos de secas com maior abrangência espacial ocorreram em março/1993, maio/2012, abril/1993 e junho/2012. Nesses eventos, assim como na seca de junho/1983, todos os estados do Nordeste foram atingidos, mesmo que em pequenas áreas.

Apesar da análise de extremos ter fornecido resultados favoráveis, ela não é definitiva para concluir que o SPI é adequado para representar as secas na região Nordeste, em razão das várias fontes de erro presentes. O SPI na escala temporal de 3 meses sinaliza especialmente a seca meteorológica e não se sabe qual foi o tipo de seca constante dos registros históricos, além do fato de que as informações da literatura são anuais, enquanto os dados utilizados neste trabalho são mensais. Além disso, não foi estudada a capacidade do SPI em detectar o início e o término de um evento de seca, o que é informação primordial, por exemplo, para que os Governos iniciem ou encerrem medidas de mitigação dos impactos das secas.

A análise compósito, realizada a partir dos dados obtidos na análise de extremos, forneceu resultados coerentes com os padrões espaciais dos pesos das componentes principais. As regiões do Nordeste que apresentaram indícios de secas mais severas coincidiram com os locais de anomalias negativas de precipitação (precipitação abaixo da média) e de anomalias positivas de temperatura (temperatura acima do normal). Como, segundo a literatura (Mishra e Singh, 2010; Marengo *et al.*, 2013; Marengo e Bernasconi, 2015), secas meteorológicas podem estar associadas não apenas a um déficit de precipitação como também a temperaturas altas, esses resultados estão de acordo com o esperado.

As correlações cruzadas entre o SPI e os índices climáticos dipolo do Atlântico e NINO3.4 mostraram que esses dois indicadores influenciam na formação das secas meteorológicas no Nordeste, especialmente com 3-4 meses de atraso da primeira componente principal do SPI. Os sinais negativos dessas correlações indicaram que, para a primeira componente, valores positivos desses dois índices estão associados a secas. Esse resultado está coerente com o indicado pela literatura (Hastenrath e Heller, 1977; Moura e Shukla, 1981; Hastenrath *et al.*, 1984; Hastenrath, 1990; Rao *et al.*, 1993; CGEE e ANA, 2012; Marengo *et al.*, 2016), já que o sinal positivo desses índices climáticos está associado à ocorrência de *El Niño* (temperatura no Oceano Pacífico acima do normal) e à presença de uma TSM no Atlântico Norte maior do que no Atlântico Sul e esses dois fenômenos são relacionados à ocorrência de secas no Nordeste do Brasil. Para a segunda componente do SPI, o dipolo apresenta correlações positivas e um efeito mais rápido, com pico em 1 mês de antecedência. O índice climático possui um efeito que se propaga da região norte para a sul do Nordeste. Em relação ao NINO3.4, o efeito é o mesmo, mas o pico de correlação ocorre apenas em 5 meses de atraso, indicando que a influência é mais lenta.

A partir do estudo das correlações entre as componentes principais do SPI e a TSM, foi possível concluir que a primeira componente principal está associada à ocorrência de *El Niño*, a segunda está relacionada às temperaturas no Atlântico Norte e no Atlântico Sul e a terceira está ligada a mecanismos climáticos provenientes do litoral africano. Esses resultados foram condizentes com estudos realizados por diversos autores (Moura e Shukla, 1981; Hastenrath, 1990; Rao *et al.*, 1993; CGEE e ANA, 2012; Marengo *et al.*, 2016).

Ao estudar diferentes valores de penalidade e os tempos de antecedência de 0 e de 3 meses na SCCA, optou-se por utilizar uma penalidade $penaltyx = 0,2$ no modelo de previsão, por ela manter os padrões espaciais de correlação entre as componentes principais do SPI e a TSM utilizando um número reduzido de pontos do preditor, o que é útil para melhorar o desempenho das previsões e a visualização dos resultados. Os valores previstos foram próximos ao SPI observado para os tempos de previsão de 1 e de 2 meses. Para os tempos de 4 e de 6 meses, o modelo não foi capaz de prever o índice de seca na magnitude adequada, subestimando os valores do SPI.

Ao analisar a qualidade das previsões, observou-se que o desempenho da primeira componente foi semelhante à média espacial do SPI, por essa componente responder pela maior parte da variabilidade dos dados, e superior às demais componentes na análise do coeficiente de correlação. Comparando o modelo completo com o modelo auto-regressivo, o primeiro apresentou um desempenho melhor do que o segundo na primeira componente e na média, tanto em termos de correlação quanto de RMSE. O mesmo não foi observado para a segunda, em todos os tempos de previsão, e para a terceira componente, até o tempo de previsão de 3 meses. Conclui-se então que o acréscimo da informação climática referente à TSM melhorou as previsões da primeira componente principal, da média espacial do SPI e da terceira componente principal (a partir do tempo de previsão de 3 meses), mas que esse efeito não observado na segunda componente principal, sendo um indício de que essa componente esteja ligada a algum outro mecanismo climático. Como esperado, o desempenho das previsões diminuiu com o crescimento do tempo de previsão, já que as correlações diminuíram e os valores do RMSE aumentaram com o crescimento do tempo de previsão.

Observando a métrica RMSE normalizado, confirmou-se que a informação climática forneceu ganho no desempenho do modelo para a primeira componente principal e para a média espacial do SPI, mas que essa melhora não aconteceu para a segunda e para a terceira componentes

principais. O maior ganho relativo foi observado no tempo de previsão de 3 meses. Em relação ao BSS, foi possível concluir que as melhores previsões, as quais são superiores à climatologia, encontraram-se na parte norte e na região mais ao norte do litoral leste do Nordeste, especialmente para os tempos de previsão de 1 e de 2 meses.

As técnicas estatísticas utilizadas mostraram-se adequadas para os fins propostos, de representação mais compacta dos dados originais e de avaliação entre diferentes variáveis. Os resultados do modelo de previsão ilustraram a dificuldade que em prever secas em diferentes tempos de antecedência. Observa-se também que a variável TSM não contribuiu tanto para a melhora das previsões, indicando a necessidade de se estudar outros mecanismos associados à formação de uma seca.

Uma sugestão para trabalhos futuros consiste na ampliação do estudo para outros índices de seca, que incluam outras variáveis responsáveis pelo desenvolvimento de uma seca além da precipitação. Seria interessante testar o modelo de previsão para índices representativos dos três tipos físicos de secas: meteorológica, agrícola e hidrológica. Pode-se também avançar no entendimento das secas históricas registradas aqui e em como a interação de índices climáticos, assim como de outras forçantes climáticas, poderiam determinar não só a intensidade e a abrangência espacial, mas também a duração das secas no Nordeste. É igualmente importante entender as secas de outras maneiras, como com a medição da distribuição de dias sem chuva. Outra possibilidade seria medir o desempenho do modelo focando em previsões apenas para a estação chuvosa, pois, como o regime de precipitação é muito concentrado, é essa época do ano que define se um ano será seco ou não, sendo uma maneira mais operacional de abordar o problema.

Para melhorar o modelo sazonal de previsão, além do uso de outros índices de seca, pode-se testar a inclusão de outras variáveis climáticas que possam capturar a evolução do sistema climático em uma menor escala temporal, como a pressão ao nível do mar, a altura geopotencial e previsões realizadas por Modelos de Circulação Geral (*General Circulation Models*, GCMs). É preciso testar e entender a variação de outras forçantes, já que diferentes eventos de secas estão relacionados a mecanismos distintos, tais como dipolo do vento, alta pressão no Atlântico Norte, padrões de aquecimento no Equador, frentes, vórtices, oscilações decadais e outros índices relacionados com o ENSO. É preciso avaliar como essas variáveis atuam na formação das secas e como fatores de curto e de longo prazo se combinam para isso.

REFERÊNCIAS BIBLIOGRÁFICAS

- Aghakouchak, A. (2015). A multivariate approach for persistence-based drought prediction: Application to the 2010–2011 East Africa drought. In: *Journal of Hydrology*, 526, 127-135.
- AghaKouchak, A.; Farahmand, A.; Melton, F. S.; Teixeira, J.; Anderson, M. C.; Wardlow, B. D.; Hain, C. R. (2015). Remote sensing of drought: Progress, challenges and opportunities. In: *Reviews of Geophysics Eighth*, 53, 452-480.
- Anderson, M. C.; Zolin, C. A.; Hain, C. R.; Semmens, K.; Yilmaz, M. T.; Gao, F. (2015). Comparison of satellite-derived LAI and precipitation anomalies over Brazil with a thermal infrared-based Evaporative Stress Index for 2003–2013. In: *Journal of Hydrology*, 526, 287-302.
- Banco Mundial (2015). A necessidade de um novo conceito na gestão de secas. In: Banco Mundial (2015). Monitor de Secas do Nordeste, em busca de um novo paradigma para a gestão de secas, 1ª edição, Brasília, 2015.
- Banco Mundial (2015). O histórico de secas da Região Nordeste. In: Banco Mundial (2015). Monitor de Secas do Nordeste, em busca de um novo paradigma para a gestão de secas, 1ª edição, Brasília, 2015.
- Banco Mundial (2015). Os indicadores de seca e produtos de apoio utilizados. In: Banco Mundial (2015). Monitor de Secas do Nordeste, em busca de um novo paradigma para a gestão de secas, 1ª edição, Brasília, 2015.
- Banco Mundial (2015). Retratos da variabilidade climática recente na região. In: Banco Mundial (2015). Monitor de Secas do Nordeste, em busca de um novo paradigma para a gestão de secas, 1ª edição, Brasília, 2015.
- Brasil. Centro de Gestão e Estudos Estratégicos – CGEE e Agência Nacional de Águas - ANA (2012). *A Questão da Água no Nordeste*. Brasília, DF.
- Brasil, Lei nº 13.153, de 30 de Julho de 2015. Institui a Política Nacional de Combate e Prevenção à Desertificação e Mitigação dos Efeitos da Seca e seus instrumentos; prevê a criação da Comissão Nacional de Combate à Desertificação; e dá outras providências. Diário Oficial [da República Federativa do Brasil], Brasília, DF, 30 jul. 2015.
- Brito, S. S. B.; Cunha, A. P. M. A.; Cunningham, C. C.; Alvalá, R. C.; Marengo, J. A.; Carvalho, M. A. (2018). Frequency, duration and severity of drought in the Semiarid Northeast Brazil region. In: *Int. J. Climatol.*, 38, 517-529.

- Canamary, E. A. (2015). *Avaliação de índices para fins de monitoramento e previsão de secas no Nordeste setentrional*. Dissertação de Mestrado em Tecnologia Ambiental e Recursos Hídricos, Publicação PTARH.DM-174/15, Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília, DF, 133p.
- Canamary, E. A. ; Reis junior, D. S. ; Martins, E. S. P. R. Evaluation of Ensemble SPI Forecasts for Ceará, Northeastern Brazil. In: *International Conference on Drought: Research and Science-Policy Interfacing*, 2015, Valencia. Drought: Research and Science-Policy Interfacing. Londres: CRC Press, 2015. 245-250.
- Centro Nacional de Monitoramento e Alertas de Desastres Naturais - CEMADEN. Disponível em: <<http://www.cemaden.gov.br/categoria/monitoramento/secas-no-semiarido/>>. Acesso em: 17 de janeiro de 2018.
- Climate Prediction Center* – CPC. Disponível em: <<http://www.cpc.ncep.noaa.gov/products/Drought/>>. Acesso em: 17 de janeiro de 2018.
- Costa D. D.; Pereira; T. A. S.; Fragoso Jr.; C. R.; Madani K.; Uvo, C. B. (2016). Understanding Drought Dynamics during Dry Season in Eastern Northeast Brazil. In: *Frontiers in Earth Science*, 4:69.
- Curtis, S. (2008). The El Niño–Southern Oscillation and Global Precipitation. In: *Geography Compass*, 2/3, 600-619.
- Dantas, C. E. O.; Cirilo, J. A.; Neto, A. R.; Silva, E. R. (2014). Caracterização da Formação de Cheias na Bacia do Rio Una em Pernambuco: Análise Estatística Regional. In: *Revista Brasileira de Recursos Hídricos*, 19 (4), 239-248.
- Delgado, J. M.; Voss, S.; Bürger, G.; Vormoor, K.; Murawski, A.; Rodrigues, M.; Martins, E.; Vasconcelos Júnior, F.; Francke, T. (2017). Seasonal Drought Prediction for Semiarid Northeast Brazil: Verification of Six Hydro-Meteorological Forecast Products. In: *Hydrol. Earth Syst. Sci. Discuss.*
- Fundação Cearense de Meteorologia e Recursos Hídricos – FUNCEME. Disponível em <<http://www3.funceme.br/previsao-climatica/>>. Acesso em: 17 de janeiro de 2018.
- Global Integrated Drought Monitoring and Prediction System* - GIDMaPS. Disponível em: <<http://drought.eng.uci.edu/>>. Acesso em: 17 de janeiro de 2018.
- Guo, J.; Sun, J.; Chang, X.; Guo, S.; Liu, X. (2011). Correlation Analysis of NINO3.4 SST and Inland Lake Level Variations Monitored with Satellite Altimetry: Case Studies of Lakes Hongze, Khanka, La-ang, Ulungur, Issyk-kul and Baikal. In: *Terr. Atmos. Ocean. Sci.*, 22, 203-213.

- Hao, Z.; Aghakouchak, A. (2013). Multivariate Standardized Drought Index: A parametric multi-index model. In: *Advances in Water Resources*, 57, 12-18.
- Hao, Z.; Hao, F.; Singh, V. P. (2016). A general framework for multivariate multi-index drought prediction based on Multivariate Ensemble Streamflow Prediction (MESP). In: *Journal of Hydrology*, 539, 1-10.
- Hao, Z.; Singh, V. P. (2015). Drought characterization from a multivariate perspective: A review. In: *Journal of Hydrology*, 527, 668-678.
- Hasanean, H. M.; Almazroui, M. (2017). Teleconnections of the tropical sea surface temperatures to the surface air temperature over Saudi Arabia in summer season. In: *International Journal of Climatology*, 37, 1040-1049.
- Hastenrath, S. (1978). On Modes of Tropical Circulation and Climate Anomalies. In: *Journal of the Atmospheric Sciences*, 35, 2222-2231.
- Hastenrath, S. (1990). Prediction of Northeast Brazil Rainfall Anomalies. In: *Journal of Climate*, 3, 893-904.
- Hastenrath, S.; Heller, L. (1977). Dynamics of climatic hazards in northeast Brazil. In: *Quart. J. R. Met. Soc.*, 103, 77-92.
- Hastenrath, S.; Wu, M-C.; Chu, P-S. (1984). Towards the monitoring and prediction of northeast Brazil droughts. In: *Quart. J. R. Met. Soc.*, 110, 411-425.
- Hayes, M. J.; Svoboda, M. D.; Wilhite, D. A.; Vanyarkho, O. V. (1999). Monitoring the 1996 Drought Using the Standardized Precipitation Index. In: *Bulletin of the American Meteorological Society*, 80, 429-438.
- Heim, R. R. (2002). A Review of Twentieth-Century Drought Indices Used in the United States. In: *American Meteorological Society*, 33, 1149-1165.
- Huang, H.; Robertson, A. W.; Kushnir, Y. (2005). Atlantic SST gradient and the influence of ENSO. In: *Geophysical Research Letters*, 32, 1-4.
- Instituto Brasileiro de Geografia e Estatística – IBGE. Indicadores Sociais Mínimos. Disponível em: <http://www.ibge.gov.br/home/estatistica/populacao/condicaodevida/indicadoresminimos/tabela1.shtm>>. Acesso em: 21 de março de 2017.
- Instituto Nacional de Meteorologia – INMET. Disponível em <http://www.inmet.gov.br/portal/>>. Acesso em: 18 de janeiro de 2018.
- Instituto Nacional de Pesquisas Espaciais – INPE. Centro de Previsão de Tempo e Estudos Climáticos – CPTEC. Disponível em <http://www.cptec.inpe.br/>>. Acesso em: 18 de janeiro de 2018.

- International Research Institute for Climate and Society/Lamont-Doherty Earth Observatory - IRI/LDEO.* Disponível em: <http://iridl.ldeo.columbia.edu/maproom/Global/Precipitation/SPI.html>. Acesso em: 18 de janeiro de 2018.
- Kaplan, A.; Cane, M. A.; Kushnir, Y.; Clement, A. C.; Blumenthal, M. B.; Rajagopalan, B. (1998). Analyses of global sea surface temperature 1856-1991. In: *Journal of Geophysical Research*, 103, 567-589.
- Kayano, M. T; Andreoli, R. V. (2004). Decadal variability of northern northeast Brazil rainfall and its relation to tropical sea surface temperature and global sea level pressure anomalies. In: *Journal of Geophysical Research*, 109, 1-8.
- Keyantash, J.; Dracup, J. A. (2002). The Quantification of Drought: An Evaluation of Drought Indexes. In: *American Meteorological Society*, 83, 1167-1180.
- Kousky, V. E. (1980). Diurnal Rainfall Variation in Northeast Brazil. In: *Monthly Weather Review*, 108, 488-498.
- Kousky, V. E. (1979). Frontal Influences on Northeast Brazil. In: *Monthly Weather Review*, 107, 1140-1153.
- Kousky, V. E.; Chu, P. S. (1978). Fluctuations in Annual Rainfall for Northeast Brazil. In: *Journal of the Meteorological Society of Japan*, 56, 457-465.
- Lima, C. H. R.; AghaKouchak, A. (2017). Droughts in Amazonia: Spatiotemporal variability, teleconnections, and seasonal predictions. In: *Water Resources Research*, 53, 10824-10840.
- Marengo J. A.; Alves L. M.; Soares W. R.; Rodriguez D. A.; Camargo H.; Riveros M. P.; Pabló A. D. (2013). Two Contrasting Severe Seasonal Extremes in Tropical South America in 2012: Flood in Amazonia and Drought in Northeast Brazil. In: *Journal of Climate*, 26, 9137-9154.
- Marengo, J. A.; Bernasconi, M. (2015). Regional differences in aridity/drought conditions over Northeast Brazil: present state and future projections. In: *Climatic Change*, 129, 103-115.
- Marengo, J. A.; Torres, R. R.; Alves, L. M. (2016). Drought in Northeast Brazil—past, present, and future. In: *Theor Appl Climatol*.
- McKee T. B.; Doesken N. J.; Kleist J. (1993). The Relationship of Drought Frequency and Duration to Time Scales. In: *Eighth Conference on Applied Climatology*. Anahaim, California.

- Mingoti, S. A. (2005). Análise de Componentes Principais. In: Mingoti, S. A. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*, editora UFMG, Belo Horizonte, Brasil.
- Mishra, A. K.; Singh, V. P. (2010). A review of drought concepts. In: *Journal of Hydrology*, 391, 202-216.
- Monitor de Secas do Nordeste do Brasil. Disponível em: <<http://monitordesecas.ana.gov.br/>>. Acesso em: 18 de fevereiro de 2018.
- Moura, A. D.; Shukla, J. (1981). On the Dynamics of Droughts in Northeast Brazil: Observations, Theory and Numerical Experiments with a General Circulation Model. In: *Journal of the Atmospheric Sciences*, 38, 2653-2675.
- Namias, J. (1972). Influence of northern hemisphere general circulation on drought in northeast Brazil. *Tellus*, 24, 336-343.
- Rajsekhar, D.; Singh, V. P.; Mishra, A. K. (2015). Multivariate drought index: An information theory based approach for integrated drought assessment. In: *Journal of Hydrology*, 526, 164-182.
- Rao, V. B.; Lima, M. C.; Franchito, S. H. (1993). Seasonal and Interannual Variations of Rainfall over Eastern Northeast Brazil. In: *Journal of Climate*, 6, 1754-1763.
- Reynolds, R. W.; Rayner, N. A.; Smith, T. M.; Stokes, D. C.; Wang, W. (2002). An Improved In Situ and Satellite SST Analysis for Climate. In: *Journal of Climate*, 15, 1609-1625.
- Santos, M. S.; Pinto, E. J. A.; Martins, L. K. L. A.; Pinto, J. A. O. (2013). Definição de Liminares de Secas e Cálculo do Índice de Precipitação Padronizada por Meio de Análise Regional de Frequências na Bacia do Alto São Francisco. In: *Revista Brasileira de Recursos Hídricos*, 18 (2), 95-105.
- Steinemann, A. C.; Hayes, M. J.; Cavalcanti, L. F. N. (2005). Drought Indicators and Triggers. In: Wilhite, D. A., *Drought and Water Crises Science, Technology, and Management Issues*. Taylor & Francis Group. Nova York.
- Tsakiris, G.; Vangelis, H. (2005). Establishing a Drought Index Incorporating Evapotranspiration. In: *European Water*, 9 (10), 3-11.
- U.S. Drought Monitor - USDM. Disponível em: <<http://droughtmonitor.unl.edu/Home.aspx>>. Acesso em: 18 de fevereiro de 2018.
- Uvo, C. B.; Repelli, C. A.; Zebiak, S. E.; Kushnir, Y. (1998). The Relationships between Tropical Pacific and Atlantic SST and Northeast Brazil Monthly Precipitation. In: *Journal of Climate*, 11, 551-562.

- Webster, K. E.; Kratz, T. K.; Bowser, C. J.; Magnuson, J. J.; Rose, W. J. (1996). The Influence of Landscape Position on Lake Chemical Responses to Drought in Northern Wisconsin. In: *Limnology and Oceanography*, 41 (5), 977–984.
- Wiesel, A.; Kliger, M.; Hero III, A. O. (2008). A greedy approach to sparse canonical correlation analysis. *arXiv:0801.2748v1 [stat.CO]*.
- Wilhite, D. A. e Buchanan-Smith, M. (2005). Drought as Hazard: Understanding the Natural and Social Context. In: Wilhite, D. A., *Drought and Water Crises Science, Technology, and Management Issues*. Taylor & Francis Group. Nova York.
- Wilks, D. S. (2011). Canonical Correlation Analysis (CCA). In: Wilks, D. S. *Statistical methods in the atmospheric sciences*. 3rd ed. p. cm. International geophysics series, v. 100, Estados Unidos.
- Wilks, D. S. (2011). Principal Component (EOF) Analysis. In: Wilks, D. S. *Statistical methods in the atmospheric sciences*. 3rd ed. p. cm. International geophysics series, v. 100, Estados Unidos.
- Witten, D. M.; Tibshirani, R.; Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation. In: *Biostatistics*, 10 (3), 515–534.
- World Meteorological Organization* (2006). Drought monitoring and early warning: concepts, progress and future challenges. WMO – No. 1006.
- Xavier, A. C.; King, C. W.; Scanlon, B. R. (2016). Daily gridded meteorological variables in Brazil (1980–2013). In: *Int. J. Climatol.*, 36, 2644–2659.
- Zou, H.; Hastie, T.; Tibshirani, R. (2006). Sparse Principal Component Analysis. In: *Journal of Computational and Graphical Statistics*, 15:2, 265-286, doi: 10.1198/106186006X113430.

APÊNDICES

APÊNDICE A – COMPLEMENTO SOBRE CCA

As correlações entre as variáveis originais e as canônicas, chamadas de *correlações homogêneas*, podem ser calculadas pelas Equações (A.1) e (A.2). Nelas, $[D_x]$ e $[D_y]$ são as matrizes diagonais que contêm os desvios padrões das variáveis x e y , respectivamente, possuindo dimensão $(I \times I)$ e $(J \times J)$, respectivamente.

$$\text{Cor}(v_m, x^T) = a_m^T [S_{x,x}] [D_x]^{-1} \quad (\text{A.1})$$

$$\text{Cor}(w_m, y^T) = b_m^T [S_{y,y}] [D_y]^{-1} \quad (\text{A.2})$$

Os vetores das correlações entre as variáveis canônicas e as outras variáveis originais, chamados de *correlações heterogêneas*, também podem ser calculados (Equações (A.3) e (A.4)):

$$\text{Cor}(v_m, y^T) = a_m^T [S_{x,y}] [D_y]^{-1} \quad (\text{A.3})$$

$$\text{Cor}(w_m, x^T) = b_m^T [S_{y,x}] [D_x]^{-1} \quad (\text{A.4})$$

A CCA pode ou não ser capaz de resumir a variabilidade das variáveis originais, como faz a PCA. Para calcular a proporção da variância (R^2_m) capturada por cada variável canônica, ou seja, quanto as variáveis canônicas representam a variabilidade original, é preciso detectar o quão bem as variáveis originais podem ser sintetizadas a partir das variáveis canônicas. Com base nas Equações (6.17) e (6.18), tem-se as Equações (A.5) e (A.6) (para as observações individuais) e as Equações (A.7) e (A.8) (para a forma matricial):

$$x' = [\tilde{A}]^{-1} v \quad (\text{A.5})$$

$$y' = [\tilde{B}]^{-1} w \quad (\text{A.6})$$

$$[X']^T = [\tilde{A}]^{-1} [V]^T \quad (\text{A.7})$$

$$[Y']^T = [\tilde{B}]^{-1} [W]^T \quad (\text{A.8})$$

O sinal gráfico til indica as colunas das inversas das matrizes correspondentes. Caso os vetores x e y possuam a mesma dimensão ($I = J$), as matrizes $[A]$ e $[B]$ são quadradas e $[\tilde{A}] = [A]^t$ e $[\tilde{B}] = [B]^t$, sendo possível calcular as respectivas inversas. Se, no entanto, as dimensões forem diferentes ($I \neq J$), as matrizes $[A]$ e $[B]$ não são quadradas e, portanto, não são invertíveis. Nesse caso, é preciso preencher as últimas colunas da matriz menor com vetores canônicos “fantasmas”.

Sendo as matrizes de covariância das variáveis canônicas iguais a $(n - 1)^{-1}[V]^T[V] = [I]$ e $(n - 1)^{-1}[W]^T[W] = [I]$, tem-se:

$$[S_{x,x}] = \frac{1}{n - 1} [X']^T [X'] = [\tilde{A}]^{-1} ([\tilde{A}]^{-1})^T = \sum_{m=1}^I \tilde{a}_m \tilde{a}_m^T \quad (\text{A. 9})$$

$$[S_{y,y}] = \frac{1}{n - 1} [Y']^T [Y'] = [\tilde{B}]^{-1} ([\tilde{B}]^{-1})^T = \sum_{m=1}^I \tilde{b}_m \tilde{b}_m^T \quad (\text{A. 10})$$

Assim, as proporções das variâncias de x e de y representadas pelas variáveis canônicas são dadas pelas Equações (A.11) e (A.12), respectivamente:

$$R_m^2(x) = \frac{\text{tr}(\tilde{a}_m \tilde{a}_m^T)}{\text{tr}([S_{x,x}])} \quad (\text{A. 11})$$

$$R_m^2(y) = \frac{\text{tr}(\tilde{b}_m \tilde{b}_m^T)}{\text{tr}([S_{y,y}])} \quad (\text{A. 12})$$

Uma maneira de calcular os vetores canônicos a_m e b_m , descrita por Mingoti (2005), é resolvendo o sistema de equações mostrado nas Equações (A.7) e (A.8), em que λ_m é o m -ésimo autovalor da matriz $([S_{x,x}]^{-1}[S_{x,y}][S_{y,y}]^{-1}[S_{y,x}]) = ([S_{y,y}]^{-1}[S_{y,x}][S_{x,x}]^{-1}[S_{x,y}])$ e deve satisfazer as equações características mostradas na Equação (A.13) e (A.14).

$$\begin{cases} ([S_{x,y}][S_{y,y}]^{-1}[S_{y,x}] - \lambda_m[S_{x,x}]) a_m = 0 \\ ([S_{y,x}][S_{x,x}]^{-1}[S_{x,y}] - \lambda_m[S_{y,y}]) b_m = 0 \end{cases} \quad (\text{A.13})$$

$$\begin{cases} |[S_{x,y}][S_{y,y}]^{-1}[S_{y,x}] - \lambda_m[S_{x,x}]| = 0 \\ |[S_{y,x}][S_{x,x}]^{-1}[S_{x,y}] - \lambda_m[S_{y,y}]| = 0 \end{cases} \quad (\text{A.14})$$

APÊNDICE B – PORCENTAGENS DA ÁREA DO NORDESTE NOS EVENTOS EXTREMOS

Tabela B.1 - Porcentagens da área do Nordeste atingida por eventos severos e extremos de seca meteorológica detectados pela primeira e segunda componentes principais do índice SPI

1ª Componente (secas em todo Nordeste)		2ª Componente – inferior (secas na região sul do Nordeste)		2ª Componente – superior (secas na região norte do Nordeste)	
Área (%)	Mês	Área (%)	Mês	Área (%)	Mês
56,43	mar/93	35,12	dez/82	33,35	jul/81
51,08	mai/12	21,76	nov/82	24,66	fev/93
49,90	abr/93	20,78	dez/03	20,78	dez/87
41,45	abr/12	20,28	mar/96	19,70	ago/97
40,72	jun/12	19,35	jan/95	19,55	jun/81
40,18	jun/83	18,76	dez/94	18,86	jul/92
38,95	jul/83	17,49	jan/10	18,47	dez/92
37,18	fev/13	15,82	out/08	18,37	ago/81
36,44	jun/98	15,32	fev/95	17,63	dez/91
35,12	dez/82	13,80	ago/87	16,75	jan/93
33,35	jul/81	13,65	nov/93	11,79	jan/13
32,76	abr/98	12,38	set/04	9,68	mar/83
32,47	mai/93	10,46	ago/11	9,23	abr/81
31,43	out/12	7,76	jan/94	6,68	nov/12
29,91	mar/13	6,83	set/11	6,53	abr/10
27,65	jan/83	6,14	jul/88	6,43	abr/83
25,54	mar/12	4,52	abr/86	5,06	mai/81
25,44	jan/07	3,78	jul/11	2,75	jan/92
23,77	set/81	3,14	jan/98	2,60	fev/92
23,53	mar/90	2,41	mai/86	1,28	set/99
22,94	mai/80	1,52	abr/96	0,74	abr/05
21,95	out/07	1,38	ago/00	0,69	out/09
21,22	mai/92	0,44	nov/86	0,54	mar/97
18,96	mai/98	0,29	abr/85	0,49	mar/92
18,17	abr/13	0,25	mar/86	0,44	mai/10
15,86	jul/80	0,25	out/00	0,10	abr/97
14,49	jun/80	0,10	mai/85	0,00	mai/97