

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE PESQUISAS HIDRÁULICAS

**PREVISÃO DE LONGO PRAZO DE NÍVEIS NO SISTEMA
HIDROLÓGICO DO TAIM**

CARLOS HENRIQUE PEREIRA ASSUNÇÃO GALDINO

Dissertação de Mestrado

Orientador: Prof. Dr. Juan Martín Bravo

Banca Examinadora

Prof ^ª . Dr ^ª . Rita de Cássia Marques Alves	CEPSRM/UFRGS
Prof. Dr. David Manuel Lelinho da Motta Marques	IPH/UFRGS
Prof. Dr. Olavo Correa Pedrollo	IPH/UFRGS

Porto Alegre, 08 de julho de 2015

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE PESQUISAS HIDRÁULICAS

**PREVISÃO DE LONGO PRAZO DE NÍVEIS NO SISTEMA
HIDROLÓGICO DO TAIM**

CARLOS HENRIQUE PEREIRA ASSUNÇÃO GALDINO

Dissertação submetida ao Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do título de Mestre em Engenharia.

Porto Alegre
2015

Resumo

O crescimento populacional e a degradação dos corpos d'água vêm exercendo pressão à agricultura moderna, a proporcionar respostas mais eficientes quanto ao uso racional da água. Para uma melhor utilização dos recursos hídricos, faz-se necessário compreender o movimento da água na natureza, onde o conhecimento prévio dos fenômenos atmosféricos constitui uma importante ferramenta no planejamento de atividades que utilizam os recursos hídricos como fonte primária de abastecimento. Nesse trabalho foram realizadas previsões de longo prazo com antecedência de sete meses e intervalo de tempo mensal de níveis no Sistema Hidrológico do Taim, utilizando previsões de precipitação geradas por um modelo de circulação global. Para realizar as previsões foi elaborado um modelo hidrológico empírico de regressão, onde foram utilizadas técnicas estatísticas de análise e manipulação de séries históricas para correlacionar os dados disponíveis aos níveis (volumes) de água no banhado. Partindo do pressuposto que as previsões meteorológicas são a maior fonte de incerteza na previsão hidrológica, foi utilizada a técnica de previsão por conjunto (*ensemble*) e dados do modelo COLA, com 30 membros, para quantificar as incertezas envolvidas. Foi elaborado um algoritmo para gerar todas as possibilidades de regressão linear múltipla com os dados disponíveis, onde oito equações candidatas foram selecionadas para realizar as previsões. Numa análise preliminar dos dados de entrada de precipitações previstas foi observado que o modelo de circulação global não representou os extremos observados de forma satisfatória, sendo executado um processo de remoção do viés. O modelo de empírico de simulação foi posteriormente executado em modo contínuo, gerando previsões de longo prazo de níveis para os próximos sete meses, para cada mês no período de junho/2004 a dezembro/2011. Os resultados obtidos mostraram que a metodologia utilizada obteve bons resultados, com desempenho satisfatórios até o terceiro mês, decaindo seu desempenho nos meses posteriores, mas configurando-se em uma ferramenta para auxílio à gestão dos recursos hídricos do local de estudo.

Palavras-chave: Previsão hidrológica; previsão de precipitação; modelagem empírica; regressão múltipla.

Abstract

Population growth and degradation of water bodies have been pressuring modern agriculture, to provide more efficient responses about the rational use of water. For a better use of water resources, it is necessary to understand the movement of water in nature, where prior knowledge of atmospheric phenomena is an important tool in planning activities that use water as the primary source of supply. In this study were performed long-term forecasts of water levels (seven months of horizon, monthly time-step) in the Hydrological System Taim, using rainfall forecasts generated by a global circulation model as input. To perform predictions was developed an empirical hydrological regression model. This model was developed based on statistical techniques of analysis and manipulation of historical data to correlate the input data available to the levels (volume) of water in a wetland. Assuming that weather forecasts are a major source of uncertainty in hydrological forecasting, we used an ensemble forecast from COLA 2.2 with 30 members to quantify the uncertainties involved. An algorithm was developed to generate all the multiple linear regression models with the available data, where eight candidates equations were selected for hydrological forecasting. In a preliminary analysis of the precipitation forecast was observed that the global circulation model did not achieve a good representation of extremes values, thus a process of bias removal was carried out. Then the empirical model was used to generate water levels forecast for the next seven months, in each month of the period june/2004 to december/2011. The results showed that the methodology used has a satisfactory performance until the lead time three (third month in the future) where the performance starts to show lower values. Beside the sharply lost of performance in the last lead times, the model is a support tool that can help the decision making in the management of water resources for the study case.

Keywords: hydrological forecast; precipitation forecast; empirical modeling; Multiple regression.

Dedico esse trabalho primeiramente a Deus, e a
minha família, que são minha razão de viver.

AGRADECIMENTOS

A Deus pelo dom da vida, e por me permitir estudar e ampliar meus conhecimentos.

A meus pais, Rui e Josefa, por dedicarem suas vidas a criação de seus filhos, e por direcionarem a mim o mais puro e sincero amor. A minha irmã Karina e meu cunhado Sidnei, por estarem sempre ao meu lado, proporcionando-me um lar saudável e feliz, em um lugar tão distante de minha casa. Aos meus sobrinhos Pedro e Ana, pela alegria de poder compartilhar de tantos momentos felizes. A todos os meus familiares, por serem presentes em minha vida, em especial meu primo Bruno, pelo companheirismo, principalmente nos momentos mais difíceis.

A minha esposa Betânia pelo carinho e compreensão, nessa árdua jornada. Ao longo desses dois anos, você me ajudou a suportar tantos momentos difíceis. Sem seu apoio teria sido mais difícil minha jornada.

Aos meus amigos, que mesmo distante se fizeram presentes. A Senhorinha e Aloísio Carolino, por serem a família que Deus me presenteou. Aos amigos que conheci ao longo dessa jornada, em especial o Mestre Domingos por todo conhecimento que tem compartilhado comigo.

Aos amigos Daniela, Bibiana, Juan, Virgílio, Siqueira, Vinícius, Ada, Mariane, Regis, Daniel, Paulo, Rosane, Diogo, Nilson, Maurício, Gabriel, Fernando, Junior, Fernando Fan, Mino, Tirzah pelo companheirismo nos momentos bons e maus compartilhados ao longo do mestrado.

Ao professor Ruberto, pelo seu incentivo ao estudo na área de recursos hídricos.

A todos os professores que contribuíram para minha formação do mestrado, em especial os professores Walter, Beatriz, Olavo e David por toda dedicação, e por sempre estarem disponíveis quando eu precisei.

Ao meu orientador Juan Martín, por todo conhecimento e dedicação que a mim direcionou. Seus comentários e instruções foram de grande importância para a construção desse trabalho, onde além de orientador, encontrei um amigo.

Ao Instituto de Pesquisas Hidráulica (IPH/UFRGS) pela oportunidade. Ao CNPq pela concessão da bolsa.

“O coração do homem propõe o seu caminho;
mas o Senhor lhe dirige os passos”.

Provérbios 16:9

Sumário

1	INTRODUÇÃO	16
1.1	Justificativa e relevância do estudo	16
1.2	Objetivo	18
2	PREVISÕES METEOROLÓGICAS	19
2.1	Aspectos gerais	19
2.2	Previsão meteorológica de longo prazo	20
2.3	Previsão de precipitação	21
2.4	Previsões climáticas	22
2.5	Previsão por conjunto (“ensemble”)	24
2.6	Modelos de previsão de precipitação	25
2.6.1	Modelos de circulação global (GCM)	26
2.6.2	Modelos regionais de previsão (<i>downscaling</i>)	27
2.6.3	Modelo de circulação global COLA	28
3	PREVISÕES HIDROLÓGICAS	30
3.1	Aspectos gerais	30
3.2	Previsões hidrológicas de longo prazo	31
3.3	Modelos hidrológicos	32
3.4	Modelos empíricos	33
3.4.1	Correlação	34
3.4.2	Regressão	36
3.5	Variáveis de entrada	42
3.5.1	Série temporal	43
3.6	Medidas de desempenho	49
4	CORREÇÃO DA PRECIPITAÇÃO PREVISTA	55
4.1	Detecção de Viés	55
4.2	Remoção do viés	57
4.2.1	Função cumulativa de probabilidade	57
4.2.2	Teste de aderência	59
5	ESTUDO DE CASO	60
5.1	Características gerais	60
5.2	Clima e precipitação	62
5.3	Uso do solo, geologia e relevo	62
5.4	Biodiversidade	62
5.5	Conflitos pelo uso da água	63
5.6	Dados hidrológicos	64
5.6.1	Dados de precipitação observada	64
5.6.2	Dados de previsão de precipitação e climatologia do modelo de circulação global	65
5.6.3	Dados de níveis de água, evaporação e retiradas para irrigação	66

5.7	Estudos antecedentes utilizando modelos matemáticos	66
6	METODOLOGIA	70
6.1	Análise preliminar dos dados hidrológicos	72
6.1.1	Divisão da amostra de dados.....	76
6.1.2	Correlações entre as variáveis.....	78
6.1.3	Retiradas para irrigação	81
6.1.4	Testes não paramétricos.....	84
6.2	Estimativa de modelos empíricos de simulação	85
6.2.1	Modelo de regressão linear simples	86
6.2.2	Modelo de regressão não linear simples	87
6.2.3	Modelo de regressão linear múltipla	87
6.3	Verificação de viés nos dados do modelo COLA	92
6.3.1	Detecção do viés	92
6.3.2	Remoção do viés.....	92
6.4	Considerações finais sobre o desenvolvimento do modelo	93
6.5	Previsão de longo prazo de níveis.....	94
7	RESULTADOS E DISCUSSÕES	97
7.1	Estatística das séries	97
7.1.1	Resultados estatísticos e ranking dos meses com maior disponibilidade de água.....	97
7.1.2	Correlação entre as variáveis	99
7.1.3	Testes não paramétricos.....	113
7.2	Estimativa de equações de regressão	115
7.2.1	Modelo de regressão linear simples	115
7.2.2	Modelo de regressão não linear simples	118
7.2.3	Modelo de regressão linear múltipla	118
7.3	Detecção e remoção de viés nos dados do modelo COLA	124
7.3.1	Detecção do viés	124
7.3.2	Remoção do viés.....	129
7.4	Análises dos modelos empíricos	144
7.4.1	Análise dos resíduos das equações de regressão linear múltipla.....	145
7.4.2	Validação dos modelos	146
7.5	Previsões de longo prazo de níveis	148
7.5.1	Previsões utilizando chuva perfeita.....	148
7.5.2	Previsões utilizando chuva prevista	150
8	CONCLUSÕES E RECOMENDAÇÕES	167
	REFERÊNCIAS.....	171
	ANEXO A.....	180
	APÊNDICE A.....	183

Lista de Figuras

Figura 1 – Temperatura da superfície do mar. INPE (2015).	23
Figura 2 - Ilustração da técnica da previsão hidrológica por conjunto: a) previsão determinística; b) previsão probabilística. Adaptado de Meller (2012).	25
Figura 3 - Discretização espacial dos modelos de circulação global. Adaptado de Cardoso (2010).	26
Figura 4 - Detalhamento da resolução dos modelos climáticos através da técnica de "downscaling". Adaptado de Silva (2005).	28
Figura 5 - Exemplo de correlação: a) correlação positiva; b) correlação negativa.	35
Figura 6 - Exemplos do rank histograma para $N_s=8$.	53
Figura 7 – Exemplos de possíveis resultados da aplicação do SBD. Adaptado de Meller (2012).	54
Figura 8 - Mapeamento das probabilidades para correção do viés. Adaptado de Fernandes (2012).	57
Figura 9 - Delimitação do banhado do Taim. Adaptado de Cardoso (2010).	61
Figura 10 – Delimitação da Lagoa Mangueira. Adaptado de Cavalcante (2013).	61
Figura 11 - Fluxograma da metodologia para elaboração das equações de regressão.	71
Figura 12 - Fluxograma para avaliação e correção do Viés nos dados da precipitação prevista pelo modelo de circulação global.	71
Figura 13 – Fluxograma da realização das previsões.	72
Figura 14 – Série temporal dos níveis no banhado do Taim (1960 - 2011).	72
Figura 15 - Sazonalidade da série de precipitação (1960 - 2011).	73
Figura 16 - Histograma da precipitação observada (1960 - 2011).	73
Figura 17 - Histogramas da precipitação para os meses de janeiro a dezembro (a - l).	74
Figura 18 - Sazonalidade da série de evaporação (1960 - 2011).	75
Figura 19 - Discretização mensal da série de níveis.	76
Figura 20 - Gráfico dos níveis com os resultados estatísticos plotados.	77
Figura 21 - Precipitação no banhado do Taim (1960 a 2011).	79
Figura 22 - Análise de correlação: a) mês a mês; b) mês atual para os níveis com o mês anterior das médias móveis de dois meses da precipitação.	81
Figura 23 - Gráfico da curva Cota-Volume da Lagoa Mangueira.	82
Figura 24 - Perfis topográficos longitudinais da interface entre o banhado do Taim e a lagoa Mangueira, adaptado de Paz (2003).	83
Figura 25 – Níveis do Sistema Hidrológico do Taim para o intervalo entre 2001 - 2007: a) níveis observados e simulados (adaptado de Fragoso, 2009); b) níveis no banhado do Taim em metros.	83
Figura 26 - Comparação entre o regime hidrológico real no banhado do Taim e cenário hipotético sem irrigação. Adaptado de Tassi (2008).	84
Figura 27 – Fluxograma simplificado do algoritmo elaborado para realizar regressões lineares múltiplas.	88
Figura 28 – Fluxograma das previsões de níveis.	96
Figura 29 - Series de níveis anuais somadas com as retiradas em cada ano no banhado do Taim (1960 - 2011).	98
Figura 30 - Gráfico dos dados normalizados (1960 e 2011).	99
Figura 31- Diagrama de dispersão dos dados do banhado do Taim entre 1960 a 1975. Onde os níveis estão em metros (m), a precipitação e evaporação estão em milímetros (mm).	100
Figura 32 - Diagrama de dispersão dos dados do banhado do Taim entre 1976 a 2003. Onde os níveis estão em metros (m), a precipitação e evaporação estão em milímetros (mm).	100
Figura 33 - Diagrama de dispersão dos dados do banhado do Taim entre 2004 a 2011. Onde os níveis estão em metros (m), a precipitação e evaporação estão em milímetros (mm).	101
Figura 34 - Tendência-ciclo da série de níveis do banhado do Taim (1960 a 2011).	102
Figura 35 - Tendência-ciclo da série de precipitação do banhado do Taim (1960 a 2011).	102
Figura 36 - Tendência-ciclo da série de evaporação do banhado do Taim (1960 a 2011).	102
Figura 37 - Linhas de tendências dos níveis para cada intervalo da elaboração do modelo.	103
Figura 38 - Linhas de tendências da precipitação para cada intervalo da elaboração do modelo.	103
Figura 39 - Linhas de tendências da evaporação para cada intervalo da elaboração do modelo.	104
Figura 40 - Ciclo das séries de níveis, precipitação e evaporação (1960 a 2011).	104
Figura 41 - Ciclo dos níveis e das precipitações no banhado.	105
Figura 42 - Diagramas de dispersão entre os níveis (eixo horizontal em m) e o logaritmo natural da precipitação (eixo vertical em mm) para diferentes intervalos: a) série completa (1960 - 2011); b) intervalo de formulação (1976 - 2003); c) intervalo de verificação (1960 - 1975) e; d) intervalo de simulação (2004 - 2011).	105
Figura 43 - Alisamento exponencial das séries: a) precipitação e b) evaporação.	113

Figura 44 - Diagrama de dispersão entre os valores observados e simulados utilizando os dados de precipitação suavizada exponencialmente para os períodos de: a) formulação; b) validação.	117
Figura 45 - Rank histogram calculado para os dados de precipitação prevista para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.	126
Figura 46 - Gráfico da precipitação observada sobreposta aos membros previstos para o horizonte de um mês de antecedência.	127
Figura 47 - Spread-bias diagram calculado para os dados de precipitação para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.	128
Figura 48 - Função cumulativa de probabilidades da precipitação observada e a: a) média da climatologia; b) média organizada de forma crescente da climatologia.	129
Figura 49 - Distribuições acumulada de probabilidades dos membros da climatologia.	130
Figura 50 - Rank histogram calculado para os dados de precipitação prevista com remoção do viés utilizando a média dos membros da climatologia para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.	131
Figura 51 - Membros previstos da precipitação para o horizonte de um mês de antecedência com remoção do viés utilizando a média dos membros da climatologia.	131
Figura 52 - Média dos membros previstos da precipitação para o horizonte de um mês de antecedência sem e com remoção do viés utilizando a média dos membros da climatologia.	131
Figura 53 - Spread-bias diagram calculado para os dados de precipitação prevista com remoção do viés utilizando a média dos membros da climatologia para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.	132
Figura 54 - Rank histogram calculado para os dados de precipitação prevista com remoção do viés utilizando a média organizada de forma crescente dos membros da climatologia para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.	134
Figura 55 - Membros previstos da precipitação para o horizonte de um mês de antecedência com remoção do viés utilizando a média organizada de forma crescente dos membros da climatologia.	134
Figura 56 - Média dos membros previstos da precipitação para o horizonte de um mês de antecedência sem e com remoção do viés utilizando a média organizada de forma crescente dos membros da climatologia.	134
Figura 57 - Spread-bias diagram calculado para os dados de precipitação prevista com remoção do viés utilizando a média organizada de forma crescente dos membros da climatologia para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.	136
Figura 58 - Dispersão entre os valores dos membros da climatologia do MCG e os dados observados (1960 a 2002).	137
Figura 59 - Função cumulativa de probabilidades da média organizada de forma crescente dos membros (b) e o membro dez da climatologia (1960 a 2002).	137
Figura 60 - Função cumulativa de probabilidades mensal da climatologia do MCG (membro um) e dos dados observados da precipitação.	139
Figura 61 - Rank histogram calculado para os dados de precipitação prevista com remoção do viés mensal realizado para todos os meses para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.	139
Figura 62 - Spread-bias diagram calculado para os dados de precipitação prevista com remoção do viés mensal realizado para todos os meses para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.	140
Figura 63 - Rank histogram calculado para os dados de precipitação prevista com remoção do viés mensal realizado para os meses de março, abril, maio e dezembro para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.	142
Figura 64 - Spread-bias diagram calculado para os dados de precipitação prevista com remoção do viés mensal realizado para os meses de março, abril, maio e dezembro para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.	143
Figura 65 - Fluxograma da análise dos resultados.	144
Figura 66 - Gráfico dos resíduos vs. valores ajustados, para as equações de n°: a) 33; b) 35; c) 87; d) 90; e) 91; f) 162; g) 195; h) 197.	146
Figura 67 - Diagrama de dispersão entre os dados observados e simulados para o intervalo de verificação das equações: a) 33; b) 35; c) 87; d) 90; e) 91; f) 162; g) 195; h) 197.	148
Figura 68 - Diagrama de dispersão entre os níveis observados e previstos pela equação n° 197 utilizando a precipitação perfeita para os horizontes de: a) um mês; b) dois meses; c) três meses; d) quatro meses; e) cinco meses; f) seis meses; g) sete meses.	150

<i>Figura 69 – Diagrama de dispersão entre os níveis observados e previstos pela equação de nº 197 para os horizontes de: a) um mês; b) dois meses; c) três meses; d) quatro meses; e) cinco meses; f) seis meses; g) sete meses.</i>	<i>152</i>
<i>Figura 70 – Gráficos dos níveis previstos pela equação de nº 197 para os horizontes de: a) um mês; b) dois meses; c) três meses; d) quatro meses; e) cinco meses; f) seis meses; g) sete meses.</i>	<i>153</i>
<i>Figura 71 – Rank histogram calculado para membros previstos dos níveis para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.</i>	<i>154</i>
<i>Figura 72 – Spread-bias diagram calculado para membros previstos dos níveis para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.</i>	<i>155</i>
<i>Figura 73 – Horizontes de previsão.</i>	<i>156</i>
<i>Figura 74 – Previsões iniciando em maio de 2004 a março de 2005.</i>	<i>157</i>
<i>Figura 75 – Previsões iniciando em maio de 2005 a março de 2006.</i>	<i>158</i>
<i>Figura 76 – Previsões iniciando em maio de 2006 a março de 2007.</i>	<i>159</i>
<i>Figura 77 – Previsões iniciando em maio de 2007 a março de 2008.</i>	<i>160</i>
<i>Figura 78 – Previsões iniciando em maio de 2008 a março de 2009.</i>	<i>161</i>
<i>Figura 79 – Previsões iniciando em maio de 2009 a março de 2010.</i>	<i>162</i>
<i>Figura 80 – Previsões iniciando em maio de 2010 a março de 2011.</i>	<i>163</i>
<i>Figura 81 – Previsões iniciando em maio de 2011 a dezembro de 2011.</i>	<i>164</i>
<i>Figura 82 – Decaimento do desempenho mensurado pelo coeficiente de Nash-Sutcliffe calculados para os resultados das previsões realizadas com precipitação perfeita e prevista.</i>	<i>166</i>

Lista de Tabelas

Tabela 1 - Estações pluviométricas.....	65
Tabela 2 - Estatísticas das séries de dados do banhado do Taim (1960 - 2011).	97
Tabela 3 - Estatística das amostras utilizadas para a modelagem empírica.	97
Tabela 4 – Ranking dos meses com níveis mais elevados.	99
Tabela 5 - Coeficientes de correlação.....	101
Tabela 6 - Correlação entre as séries de dados e os níveis no banhado.....	107
Tabela 7 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das precipitações (1960 à 2011).	108
Tabela 8 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das precipitações (1976 à 2003).	108
Tabela 9 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das precipitações (1960 à 1975).	108
Tabela 10 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das precipitações (2004 à 2011).	109
Tabela 11 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das evaporações (1960 à 2011).	110
Tabela 12 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das evaporações (1976 à 2003).	110
Tabela 13 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das evaporações (1960 à 1975).	110
Tabela 14 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das evaporações (2004 à 2011).	111
Tabela 15 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas do balanço entre precipitação e evaporações (1960 à 2011).	111
Tabela 16 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas do balanço entre precipitação e evaporações (1976 à 2003).	111
Tabela 17 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas do balanço entre precipitação e evaporações (1960 à 1975).	112
Tabela 18 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas do balanço entre precipitação e evaporações (2004 à 2011).	112
Tabela 19 - Testes de hipóteses não paramétricos (1960 a 2011).	114
Tabela 20 - Testes de hipóteses não paramétricos realizado para o intervalo de formulação (1976 a 2003)...	114
Tabela 21 - Testes de hipóteses não paramétricos realizado para o intervalo de formulação (1960 a 1975)...	115
Tabela 22 - Testes de hipóteses não paramétricos realizado para o intervalo de formulação (2004 a 2011)...	115
Tabela 23 - Parâmetros estimados para as equações lineares simples.	115
Tabela 24 - Medidas de desempenho para as regressões lineares simples.	116
Tabela 25 - Medidas de desempenho calculadas para a equação obtida utilizando os dados de precipitação suavizada exponencialmente (intervalo de validação).	117
Tabela 26 - Medidas de desempenho calculadas para equação obtida utilizando os dados transformados de precipitação.....	118
Tabela 27 – Medidas de desempenho calculadas para as 47 equações que não foram rejeitadas pelos critérios pré-estabelecidos.....	118
Tabela 28 - Erro absoluto e relativo percentual para as medidas de desempenho calculadas para as 47 equações que não foram rejeitadas pelos critérios pré-estabelecidos.	120
Tabela 29 – Equações de regressão linear múltipla selecionadas.	121
Tabela 30 - Tabela ANOVA da equação de regressão múltipla nº 33.....	122
Tabela 31 - Tabela ANOVA da equação de regressão múltipla nº 35.....	122
Tabela 32 - Tabela ANOVA da equação de regressão múltipla nº 87.....	123
Tabela 33 - Tabela ANOVA da equação de regressão múltipla nº 90.....	123
Tabela 34 - Tabela ANOVA da equação de regressão múltipla nº 91.....	123
Tabela 35 - Tabela ANOVA da equação de regressão múltipla nº 162.....	123
Tabela 36 - Tabela ANOVA da equação de regressão múltipla nº 195.....	123
Tabela 37 - Tabela ANOVA da equação de regressão múltipla nº 197.....	124
Tabela 38 – Teste de significância (F total).	124
Tabela 39 - Resultados dos coeficientes de mensuração do viés.....	125
Tabela 40 – Resíduos entre os valores observados e os dados da climatologia mensal.	141
Tabela 41 – Estatística dos Resíduos das equações de regressão.....	145
Tabela 42 – Medidas de desempenho calculadas para as equações de regressão.....	146

<i>Tabela 43 - Medidas de desempenho calculadas para as previsões de níveis utilizando a precipitação perfeita.</i>	148
<i>Tabela 44 - Medidas de desempenho para equação de nº 197.</i>	150
<i>Tabela 45 – Diferença entre as previsões realizadas com precipitação perfeita e precipitação prevista para os sete horizontes.</i>	165

Lista de Símbolos

ANA – Agência Nacional de Águas

ARMA – *Auto Regressive Moving Average*

CPTEC – Centro de Previsão de Tempo e Estudos Climáticos

COLA – *Center for Ocean-Land-Atmosphere Studies*

ENSO – *El Niño Southern Oscillation*

EPR – Erro Padrão da Regressão

FCP – Função Cumulativa de Probabilidade

INPE – Instituto Nacional de Pesquisas Espaciais

IRI – *International Research Institute for Climate and Society*

MCG – Modelo de Circulação Global

NOAA – *National Oceanic and Atmospheric Administration*

ODP – Oscilação Decadal do Pacífico

PELD – Programa de Pesquisa Ecológica de Longa Duração

PNT – Previsão Numérica do Tempo

REQM – Raiz Quadrada Do Erro Quadrático Médio

RLM – Regressão Linear Múltipla

SBD – *Spread-Bias Diagram*

SHT – Sistema Hidrológico do Taim

SST – *Sea Surface Temperature*

TSM – Temperatura da Superfície do Mar

1 INTRODUÇÃO

1.1 Justificativa e relevância do estudo

Vivemos um momento de pressão mundial pelo combate à fome e uma conseqüente busca pelo aumento da produção de alimentos, impulsionado a agricultura moderna a proporcionar respostas mais eficientes nesse sentido. Neste contexto, a produção agrícola transforma-se em uma das atividades que mais utilizam os recursos naturais e uma das que mais sofrem com as variações da disponibilidade hídrica. A finalidade básica da irrigação é proporcionar água as culturas, visando atender as exigências hídricas durante todo seu ciclo de produção, possibilitando altas produtividades e produtos de boa qualidade. Assim, a quantidade de água necessária às culturas, é função da espécie cultivada, produtividade desejada, e outras variáveis.

Para uma melhor utilização dos recursos hídricos, faz-se necessário compreender o movimento da água na natureza e, nesse contexto, o estudo dos fenômenos atmosféricos constitui um importante mecanismo para análise dos fatores que influenciam no ciclo hidrológico, permitindo avaliar os desvios do comportamento médio do sistema de forma antecipada, contribuindo para previsão de variáveis hidrológicas. Logo, o conhecimento prévio dos fenômenos atmosféricos apresenta um papel determinante no planejamento de atividades que utilizam os recursos hídricos como fonte primária de abastecimento.

A modelagem de sistemas hídricos constitui uma importante ferramenta na análise e previsão de variáveis hidrológicas, auxiliando no entendimento dos processos que envolvem os recursos hídricos. O uso desses modelos vem ganhando terreno ao longo dos anos e o grande número de trabalhos produzidos neste tema são indícios que permitem concluir que o desenvolvimento de modelos, tornou-se alvo de pesquisas em todo o mundo.

Muitas técnicas são usadas para modelagem e simulação de recursos hídricos e dentre elas, uma que vem ganhando força ao longo dos últimos anos é a previsão por conjunto (ensemble). Esse método consiste em uma técnica que procura melhorar e ampliar o horizonte das previsões numéricas de tempo (precipitação), partindo do pressuposto que as previsões meteorológicas são a maior fonte de incerteza na previsão hidrológica, desconsiderando-se as incertezas nos modelos hidrológicos. A consideração de mais de uma fonte de incerteza, gera conjuntos de previsão com um grande número de membros, o que pode trazer um benefício em termos de confiabilidade da previsão (Meller, 2012).

Um exemplo claro de aplicação de simulação utilizando modelos hidrológicos consiste na previsão de níveis de água em reservatórios naturais (lagos, lagoas, banhados, etc.). As flutuações dos níveis d'água em reservatórios naturais são importantes no planejamento e projeto de estruturas hidráulicas ao longo das suas margens e também na gestão dos reservatórios de água doce para fins de abastecimento de água. A fim de controlar as mudanças de nível futuros em um reservatório é necessário o desenvolvimento de modelos de simulação, os quais avaliam as variações extremas ou anormais de níveis (Sen et al., 2000).

A modelagem das entradas e saídas de água nos reservatórios naturais é fundamentada pelos processos hidrológicos e isso dá origem a mudanças dinâmicas no nível de água, que podem ser explicados por equações simples. As mudanças de níveis também podem ser movidas por ventos de superfície que levam à configuração de seiches como ondas estacionárias. Há outros processos que ocorrem dentro de reservatórios naturais que são motivados por correntes térmicas e processos de mistura, criando movimentos físicos no corpo d'água, mas a magnitude da alteração no nível é muitas vezes considerada pequena (Khatibi et al., 2014).

Atualmente existem muitos modelos usados para previsão de níveis, cada um possuindo peculiaridades quanto a sua utilização. Dentre as metodologias empregadas podem-se destacar as empíricas, que são modelos formulados em equações que não possuem nenhuma relação com os processos físicos envolvidos. Os parâmetros desses modelos são obtidos através de ajustes baseados na série histórica das variáveis e, por não fornecerem explicações sobre o funcionamento dos fenômenos, mas apenas os resultados, são chamados de modelos do tipo empírico (Pedrollo, 2000).

Existem diversas técnicas de modelagem empírica e uma bastante difundida entre os estudos que envolvem recursos hídricos são as equações de regressão, que representam uma técnica estatística utilizada para estimar a relação entre variáveis, investigando o valor esperado de uma variável Y , dados os valores das variáveis X (x_1, x_2, \dots, x_n), considerando que exista um relacionamento funcional entre as variáveis Y e X . Assim, a regressão fornece um valor médio de uma das variáveis em função de outra ou outras, onde os pontos observados apresentam uma variação nas proximidades da curva que representa a função de regressão. Uma regressão é chamada de linear se a relação da resposta às variáveis for uma função linear de alguns parâmetros, caso contrário é chamada de não linear.

Motivado pela falta de trabalhos utilizando uma abordagem empírica para as previsões de níveis de água no Sistema Hidrológico do Taim, esse trabalho elaborou um modelo, utilizando apenas uma abordagem empírica, para responder a pergunta: Um modelo de

equações simples, utilizando uma abordagem empírica é capaz de obter resultados tão bons como um modelo complexo de base física?

1.2 Objetivo

Este trabalho tem por objetivo avaliar a aplicabilidade de um modelo empírico na previsão de longo prazo de níveis no Banhado do Taim, utilizando previsões meteorológicas por conjunto, num contexto de disputa pelo uso da água.

2 PREVISÕES METEOROLÓGICAS

2.1 Aspectos gerais

Os fenômenos atmosféricos são os processos naturais que ocorrem na camada da atmosfera terrestre, que influenciam no ciclo hidrológico, causando uma variabilidade sazonal no regime de precipitação. A complexidade do ambiente atmosférico pode gerar eventos extremos (tornados, furacões, inundações, secas, etc.), que correspondem a grandes desvios do comportamento médio do sistema (Yu et al., 2013). Esses fenômenos podem ser mensurados pelas variáveis meteorológicas (temperatura, pressão, umidade do ar, etc.), como também em escala que levam em consideração a magnitude e a duração do fenômeno.

A previsão meteorológica pode ser considerada como a aplicação da ciência e da tecnologia para fazer uma descrição detalhada de ocorrências futuras esperadas na atmosfera num certo local. As previsões incluem o uso de modelos objetivos baseados em certos parâmetros atmosféricos que são alimentados por dados obtidos por radar Doppler, balões atmosféricos, boias marítimas, estações meteorológicas, etc.

Para um melhor entendimento dos processos que envolvem o ciclo hidrológico é necessário investigar a presença de água na atmosfera, em suas diferentes fases. O ramo da ciência que une a meteorologia e a hidrologia é chamado de hidrometeorologia e é responsável por estudar a transferência de água e de energia entre a superfície e a atmosfera. Embora estudos envolvendo hidrometeorologia sejam considerados novos, existem diversos trabalhos envolvendo essa temática.

Embora os fenômenos atmosféricos não sejam consequência da ação humana, a interferência antrópica pode intensificar ou alterar a dinâmica de tais fenômenos. Nos dias atuais, o buraco na camada de ozônio tem sido algo de diversos estudos, sendo correlacionado com alguns fenômenos atmosféricos como o aquecimento global. Entretanto, existem pessoas que acreditam que “o clima da Terra tem variado ao longo das eras, forçado por fenômenos de escalas de tempo decadal até milenar” (Molion, 2007).

Nessa seção é apresentada uma breve descrição dos processos meteorológicos que envolvem as previsões de variáveis hidrológicas, onde nesse estudo será dado enfoque para as previsões de precipitações de longo prazo.

2.2 Previsão meteorológica de longo prazo

Um tema bastante atual e que tem sido alvo de um crescente número de estudos está relacionado à capacidade de prever o tempo, em particular, prever as flutuações do clima em escalas sazonais e correlaciona-los aos impactos causados às variáveis hidrológicas. Um dos primeiros relatos escritos sobre esse tema é encontrado em Gênesis 41, onde José interpreta o sonho do Faraó prevendo sete anos de fartura seguidos de sete anos de seca. A previsão sugere uma gestão na oferta de alimentos, onde se fazendo reservas para a nação do Egito constitui uma forma organizada para lidar com o período de baixa produtividade. Nesse exemplo, pode-se verificar os potenciais benefícios práticos da previsão a longo prazo. Mais tentativas sistemáticas de prever o tempo no futuro tiveram de aguardar o desenvolvimento científico, evoluir e agregar o conhecimento de técnicas de medição.

A precipitação pode ser considerada uma das variáveis meteorológicas mais difíceis de prever, devido à natureza multi-escalar (regional e global) envolvendo os processos atmosféricos que levam ao seu desenvolvimento e pela sua variabilidade no espaço e no tempo. Para tentar contornar essa dificuldade na previsão de precipitação, tem-se utilizado modelos de previsão numérica do tempo (PNT), o qual tem sido um importante aliado na busca por previsões mais precisas e significativas (Antolik, 2000).

As previsões hidrometeorológicas de longo prazo são importantes em estudos de alterações climáticas e de gestão de recursos hídricos. Geralmente, séries temporais hidrometeorológicas são previstas com modelos baseados em considerações físicas ou teorias numéricas, como o método de análise linear regressiva (Salas et al., 1980), baseado na teoria do caos (Jayawardena e Lai, 1994; Islam e Sivakumar, 2002), na similaridade local de sinais (Jang et al., 1997) e em séries temporais não-lineares com modelos de redes neurais artificiais (Raman e Sunilkumar, 1995). Embora exista uma variedade de métodos para previsão hidrometeorológica, todos esbarram na mesma dificuldade de previsão, pois os sinais hidrometeorológicos são sistemas não lineares de alta complexidade e possuem uma acentuada variação no tempo e no espaço, que tornam difícil sua previsão exata (Yu et al., 2013).

As condições atuais do clima podem ser explicadas investigando-se as mudanças que ocorreram no passado. A observação das mudanças nas variáveis hidrometeorológicas pode ser feita através da investigação de registros históricos, que é considerada um importante meio para determinar as características do clima passado e recente. Esse tipo de investigação pode revelar padrões estatísticos e conduzir a uma explicação para mudanças observadas, atualmente. Uma mudança na série pode ocorrer de forma abrupta (mudança de passo),

gradualmente (tendência), pode assumir uma forma cíclica ou uma combinação de tendências. Essas mudanças podem afetar as propriedades estatísticas globais da série histórica, tais como a média, mediana e a variância.

Periodicidades em séries temporais naturais são, geralmente, devidas aos ciclos astronômicos, como a rotação da Terra em torno do Sol ou outros fenômenos climáticos cíclicos como o ENSO (El Niño / La Niña-Oscilação Sul). O exemplo mais óbvio de uma tendência cíclica é a sazonalidade. Compreender as mudanças na série temporal é muito importante para o planejamento de recursos hídricos e outras aplicações práticas que necessitem da compreensão do aspecto futuro do clima (Nyeko-Ogiramoi et al., 2013).

A compreensão sobre a variabilidade dos parâmetros climáticos locais em bacias hidrográficas é uma importante ferramenta para verificar o comportamento desses sistemas hídricos ao longo dos anos. O entendimento sobre os parâmetros climáticos desses sistemas ambientais auxiliam a realização de previsões visando o planejamento dos recursos hídricos (Marengo, 2008). O atraso no início da estação chuvosa pode afetar a agricultura e a geração de energia. A ocorrência de enchentes e secas de grande escala tem produzido fortes impactos na economia e na população. As mudanças nos padrões e nos regimes de precipitação poderão também afetar as vazões dos rios. Assim, as incertezas dos cenários sempre devem ser consideradas, para o estabelecimento de ações para adaptação e mitigação, bem como planos de monitoramento dos recursos hídricos para avaliar os riscos climáticos (Marengo et al, 2010).

2.3 Previsão de precipitação

Entende-se por precipitação como sendo toda água proveniente do meio atmosférico que atinge a superfície. Existem várias formas de precipitação (neblina, chuva, granizo, saraiva, orvalho, geada e neve), onde a forma mais comum é a chuva (Tucci, 1993). As previsões de precipitação são tipicamente resultado de simulações de modelos de previsão numérica do tempo (PNT), advindas da integração numérica das equações de mecânica dos fluídos, que preveem as mudanças na circulação da atmosfera (Silva et al., 2006).

As bases para as previsões meteorológicas de longo prazo surgiram no início da década de 1960 quando um meteorologista americano chamado Edward Lorenz elaborou um programa de computador que simulava o movimento de massas de ar. Realizando testes em seu modelo Lorenz percebeu que pequenas mudanças ou pequenos erros em um par de variáveis produziam efeitos desproporcionais, chamando sua descoberta de “efeito borboleta”,

onde em 1979 publicou com o tema: “Previsibilidade: O bater de asas de uma borboleta no Brasil desencadeou um Tornado no Texas”. A Teoria do caos aborda sistemas complexos e dinâmicos deterministas, mas que apresentam um fenômeno fundamental de instabilidade chamado sensibilidade às condições iniciais (GHYS, 2012).

A teoria do caos diz que as componentes das forças internas do sistema atmosférico podem ser previstas com antecedência de aproximadamente até 14 dias, ou seja, não é possível fazer previsões de tempo para meses ou anos no futuro, obtendo uma sequência de eventos atmosféricos diários (Lorenz, 1973). Assim, a previsão de precipitação está dividida em previsão do tempo e do clima: a) a previsão de tempo, se refere à previsão feita para até 14 dias de antecedência; b) a previsão de clima, se refere à previsão para antecedências maiores, normalmente meses no futuro.

Sivakumar (2004) revisou os resultados da aplicação da teoria do caos em diversos estudos envolvendo os fenômenos geofísicos (por exemplo, as chuvas, o fluxo dos rios, o transporte de sedimentos, a temperatura, a pressão, etc.), apresentando técnicas de identificação e previsão do caos. Esse autor concluiu que a compreensão adequada de um fenômeno geofísico é importante para avaliar a possibilidade de previsão da sua evolução a curto e longo prazo.

2.4 Previsões climáticas

O tempo meteorológico é considerado o tempo atual ou tempo a ser previsto pelos meteorologistas, o qual se estende por no máximo por 14 dias. Já o clima é o conjunto de estados do tempo meteorológico que caracterizam o meio ambiente atmosférico de uma determinada região ao longo do horizonte previsto (período superior a 14 dias). Para se definir o clima, é considerado um subconjunto dos possíveis estados atmosféricos, o qual requer uma análise de uma extensa série de dados. A organização mundial de Meteorologia (WMO) recomenda uma série de pelo menos 30 anos para a análise climática (INPE, 2015).

O conceito pelo qual não é possível prever o tempo para meses ou anos individuais para o futuro surgiu com o advento da teoria do caos. Embora o caos imponha limites rigorosos para a previsibilidade determinista da atmosfera, este acaba caindo em uma sucessão limitada de padrões. Assim, uma tentativa de prever padrões futuros do clima pode ser entendida como prever mudanças e transformações na distribuição de probabilidades do clima (Palmer, 1993). Isso se justifica, pois a atmosfera tem grande interação com a superfície, realizando assim trocas constantes de fluxos de calor, momento e água. Dessa

forma mudanças na condição da superfície causarão consequentes alterações no estado da atmosfera.

Sob uma perspectiva global, a variabilidade da temperatura da superfície do mar (TSM), é considerada a principal responsável pelas mudanças climáticas de um ano para outro. As temperaturas do oceano são mais estáveis do que as temperaturas da superfície terrestre, isso se justifica devido à grande capacidade de calor da água e a eficiência na mistura contínua entre a água da superfície e a água das camadas inferiores. As anomalias da TSM possuem um ciclo sazonal médio de magnitude da ordem de 1° C, com escalas espaciais que podem chegar a milhares de quilômetros e duração entre um a vários meses (Stockdale, 2000).

No Brasil os dados climáticos podem ser adquiridos juntos ao Instituto Nacional de Pesquisas Espaciais (INPE), o qual fornece dados de variáveis hidrológicas, meteorológicas, imagens de satélite e outros produtos. O INPE foi criado em 1961 com o objetivo de capacitar o país nas pesquisas científicas e nas tecnologias espaciais. O Instituto é considerado um centro de excelência e referência internacional em pesquisas de ciências espaciais e atmosféricas, engenharia espacial, meteorologia, observação da Terra por imagens de satélite e estudos de mudanças climáticas (INPE, 2015). A Figura 1 é exemplo de um dos produtos distribuídos pelo INPE.

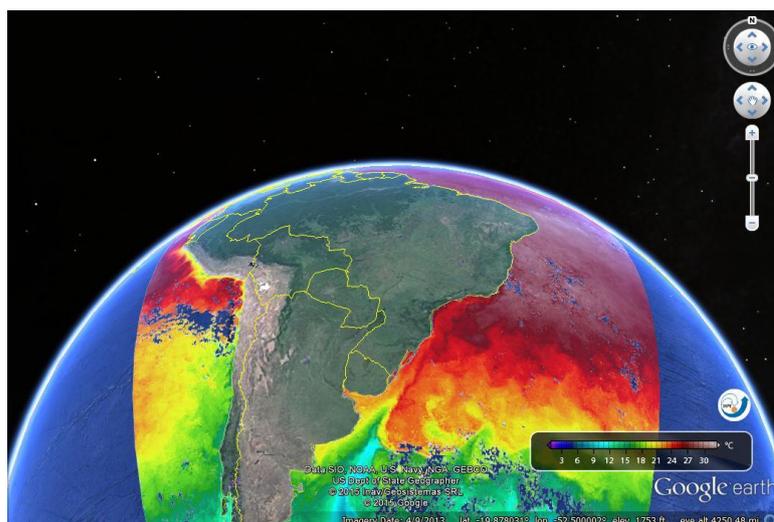


Figura 1 – Temperatura da superfície do mar. INPE (2015).

Existem diversos centros de previsões climáticas, dentre eles pode-se citar o *International Research Institute for Climate and Society (IRI)*, que foi fundado a partir de um acordo de cooperação entre o *National Oceanic and Atmospheric Administration (NOAA)* e *The Earth Institute, Columbia University* da cidade de New York. O IRI disponibiliza diversos dados climáticos simulados com diferentes modelos, onde cada um desses modelos

possui suas singularidades como resolução, número de membros, cobertura, período simulado, dentre outras características (IRI, 2015). Nessa pesquisa foi utilizado dados de precipitação prevista, disponibilizados pelo NOAA.

2.5 Previsão por conjunto (“ensemble”)

Dentre as técnicas usadas para modelagem e simulação de recursos hídricos, uma que vem ganhando força ao longo dos últimos anos é a previsão por conjunto (ensemble). Com o advento da teoria do Caos, descobriu-se que as equações que representam os movimentos da atmosfera possuem grande sensibilidade às condições iniciais, onde pequenos erros nas condições iniciais podem levar a grandes erros na previsão futura (Lorenz, 1973). Por esta razão, a previsão climática pode ser entendida em termos da evolução temporal de uma função densidade de probabilidade do estado da atmosfera (Stockdale, 2000).

A geração de uma previsão por conjunto pode ser realizada das seguintes formas: a) supondo-se que os modelos sejam perfeitos, ou seja, considerando que as incertezas ocorrem apenas nas condições iniciais. Dessa forma, o conjunto de previsões é gerado pelo modelo a partir de cenários perturbados nas condições iniciais; b) a partir de dois ou mais modelos, assim o conjunto de previsões é o resultado da integração desses diferentes modelos.

A técnica de previsão de tempo por ensemble surgiu como uma tentativa de aumentar o prazo de previsão e a previsibilidade dos modelos. Na meteorologia, seria necessário um número ilimitado de condições iniciais para representar todos os estados atmosféricos possíveis, o que seria impraticável, pois, para cada condição inicial, corresponderia uma rodada do modelo. Contudo, sabendo que a taxa de crescimento dos erros depende do regime de circulação da atmosfera, da estação do ano e do domínio geográfico, é possível construir um número limitado de perturbações “ótimas”, que representem os modos de crescimento mais rápido associados aos padrões atmosféricos da condição inicial. Assim, podem-se eliminar condições que geram previsões semelhantes entre si (CPTEC, 2015).

De uma forma geral, a previsão por conjunto procura fornecer informações sobre as incertezas das previsões hidrológicas, propondo para cada intervalo de tempo um conjunto de soluções, onde se pode estimar uma distribuição de probabilidades do preditor. Esta metodologia surgiu em contraste com as previsões determinísticas, que apresentam uma estimativa única (pontual), para cada intervalo de tempo (Meller, 2012). A Figura 2 ilustra a representação esquemática de uma previsão determinística e por conjunto.

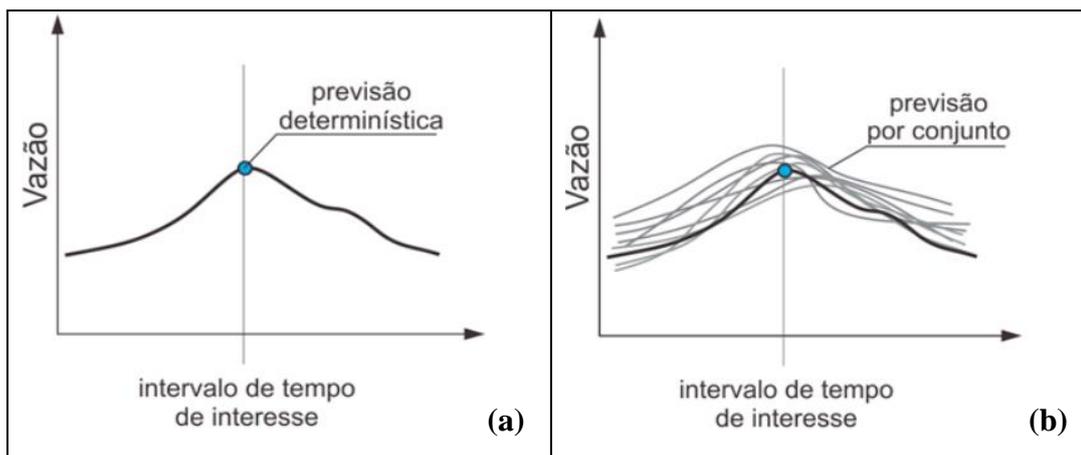


Figura 2 - Ilustração da técnica da previsão hidrológica por conjunto: a) previsão determinística; b) previsão probabilística. Adaptado de Meller (2012).

2.6 Modelos de previsão de precipitação

A precipitação é uma das variáveis de entrada em modelos hidrológicos que, por sua vez, é a variável de saída nos modelos atmosféricos. Dessa forma, é fácil deduzir que o desempenho de modelos hidrológicos está intimamente relacionado à qualidade da estimativa da precipitação. A caracterização adequada da quantidade de chuva é primordial para o sucesso da modelagem, fundamentada na teoria física ou empírica, podendo produzir bons resultados com dados de precipitação de baixa resolução espacial (McMillan et al, 2011).

Os modelos climáticos são ferramentas numéricas utilizadas para analisar o comportamento do clima em escala regional ou global. Basicamente, são elaborados para solucionar sistemas de equações diferenciais parciais que descrevem os fenômenos físicos envolvidos na previsão do clima.

Os recentes avanços na medição e previsão meteorológica criaram oportunidades para melhorar as previsões de variáveis hidrológicas. A precisão das previsões meteorológicas tem melhorado de forma constante ao longo dos últimos anos. Os modelos PNT usam as condições climáticas atuais como dados de entrada para prever a evolução dos sistemas meteorológicos. Esses modelos representam a atmosfera como um fluido dinâmico, resolvendo as equações que simulam o comportamento da atmosfera através do uso de equações de mecânica dos fluidos e termodinâmica. Esses modelos melhoraram desde a década de 1940 por causa das vantagens computacionais e melhorias na tecnologia de medição digital, incluindo satélites meteorológicos e redes de radar (Cuo et al., 2011).

Os modelos de previsão de precipitação, conhecidos como modelos atmosféricos, são divididos segundo sua área de abrangência em: modelos de circulação global e modelos regionais.

2.6.1 Modelos de circulação global (GCM)

O modelo de circulação global, GCM (*Global Circulation Model*), refere-se aos modelos matemáticos tridimensionais que simulam os principais processos que governam os movimentos da água na atmosfera, inserindo suas interações com os oceanos e a superfície dos continentes. Geralmente discretizam o globo em uma malha de resolução em cerca de 2 a 4 graus de latitude-longitude (ou elementos de 50 a 1000 km) na horizontal, e na vertical, a atmosfera e os oceanos são divididos em várias camadas, normalmente entre 10 e 30 camadas (1 a 5 km) (Figura 3). O modelo de circulação global é usado para prever o comportamento médio da atmosfera, ou seja, o seu clima, e não para fazer uma previsão determinística do tempo num momento específico (Tucci, 1998; Cardoso 2010).

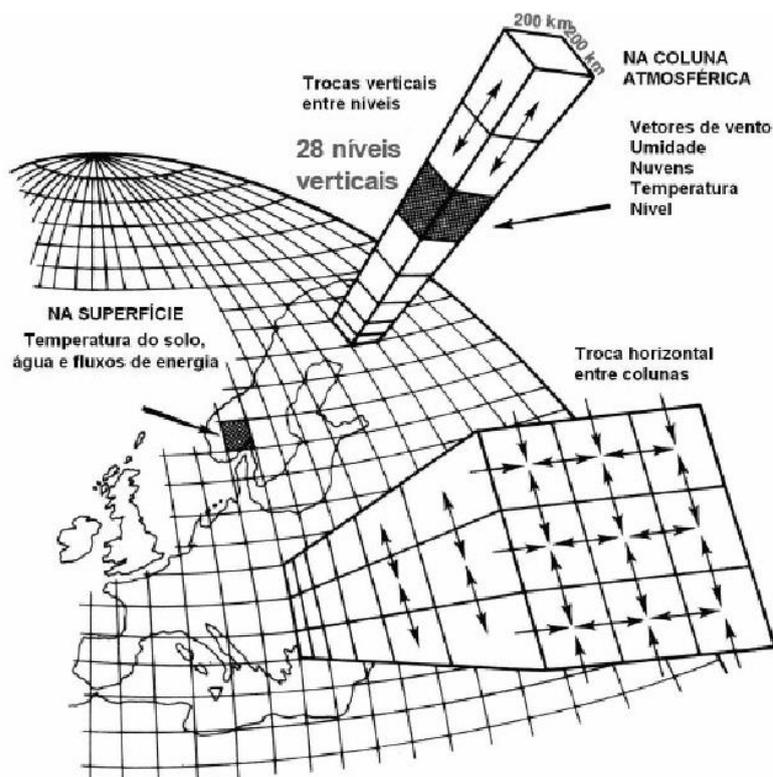


Figura 3 - Discretização espacial dos modelos de circulação global. Adaptado de Cardoso (2010).

Os GCMs são atualmente considerados como as ferramentas disponíveis mais sofisticadas para simulação do clima atual global e projeções de futuros cenários climáticos. Sua formulação, geralmente, considera o comportamento e as interações entre a biosfera,

hidrosfera, criosfera, atmosfera e geosfera. Os GCMs são capazes de simular temperaturas extremas razoavelmente bem, no entanto, a simulação de precipitações extremas é dependente da resolução, parametrização e dos limiares escolhidos. Os GCMs não preveem com precisão o clima local, mas as equações que compõem esses modelos climáticos, que são de base física, fornecem estimativas aceitáveis, a partir de relações das condições atuais do clima global, para gerar cenários previstos de variáveis climáticas, como a precipitação e temperatura (Green et al., 2011).

De acordo com Silva et al. (2006), entre as limitações destes modelos pode-se destacar: (i) A discretização retrata somente os processos atmosféricos de macroescala na superfície da terra; (ii) Muitos processos são representados com fortes limitações no modelo, tais como os processos hidrológicos, por exemplo, que variam na microescala; (iii) O custo e o tempo de processamento para uma discretização mais detalhada destes modelos é extremamente elevado.

2.6.2 Modelos regionais de previsão (*downscaling*)

Os modelos globais de previsão climática possuem baixa resolução espacial, da ordem de 100 a 200 km, no entanto, os fatores que controlam o clima regional ocorrem tipicamente em escalas bem menores. As características do uso e ocupação do solo (topografia, tipo de vegetação, etc.) podem influenciar de forma significativa no clima regional. Para realizar previsões utilizando um modelo global de alta resolução, implicaria em um alto custo computacional. Uma alternativa para contornar esse problema é a utilização de modelos regionais com resolução de algumas dezenas de quilômetros. Os modelos regionais de previsão, chamados de *downscaling* dinâmico, são alimentados nas fronteiras pelas condições produzidas pelos modelos globais, ou até mesmo por dados observados (Allasia, 2007).

A Figura 4 exemplifica como o modelo regional usa as condições de fronteira na célula B do modelo global, com resolução de 200x200 km, para gerar uma simulação com resolução de 40x40 km internamente a B de forma que os processos de troca de energia entre a atmosfera e a superfície tornam-se bem mais detalhadas do que no modelo global.

Por refinarem a malha de resolução espacial, acredita-se que os modelos regionais representem com maior detalhe os processos que agem regionalmente, dessa forma, produzindo previsões de tempo e clima com maior precisão. Esses modelos são importantes porque a sua escala de resolução espacial é considerada equivalente à escala em que são realizados os estudos hidrológicos. Assim, a aplicação dos modelos atmosféricos regionais na

área de recursos hídricos, obtém melhores resultados do que quando utilizados os modelos globais (Silva et al, 2006).

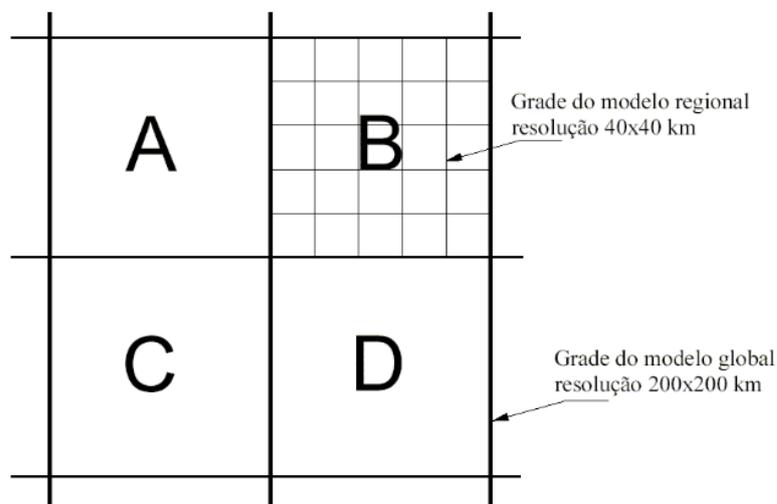


Figura 4 - Detalhamento da resolução dos modelos climáticos através da técnica de "downscaling". Adaptado de Silva (2005).

2.6.3 Modelo de circulação global COLA

No final dos anos 1960, o *Geophysical Fluid Dynamics Laboratory* do NOAA em Princeton-New Jersey, desenvolveu o seu primeiro modelo climático de circulação global, que combinava os processos oceânicos e atmosféricos. Esse modelo proporcionou aos cientistas uma melhor compreensão sobre como o oceano e a atmosfera interagem para influenciar o clima. O modelo também era capaz de prever como as mudanças nos fatores naturais que controlam o clima, como o mar, as correntes atmosféricas e a temperatura influenciavam nas mudanças climáticas (NOAA, 2015). Os modelos foram evoluindo ao longo dos anos e diversos centros de estudos climáticos elaboraram seus próprios modelos para simularem o clima em escala global.

O centro de estudos *Ocean-Land-Ambiente* (COLA), localizado na *George Mason University* em Fairfax-Virginia, é uma instituição que realiza pesquisas interdisciplinares relacionadas à variabilidade e à previsibilidade do clima sazonal do planeta, em escalas de tempo decadal. O COLA investiga elementos previsíveis do clima atual que tornam possíveis prever com precisão as variações climáticas globais. As pesquisas desenvolvidas pelo COLA partem do pressuposto que há previsibilidade no meio do caos, dando enfoque às previsões climáticas a longos prazos (Iges, 2015). Essa pesquisa utilizou previsões de precipitação realizadas pelo modelo de circulação global COLA 2.2, por esse motivo, abaixo se encontram

alguns comentários sobre as características do modelo e alguns resultados obtidos por trabalhos que utilizaram o COLA.

O modelo COLA 2.2 é um modelo de circulação global com 18 níveis verticais desigualmente espaçados e truncamento espectral T63 (Delsole & Shukla, 2006; Kirtman et al., 2003). Atualmente, os resultados das simulações do modelo COLA são disponibilizados pelo IRI, onde podem ser encontrados resultados com dez membros utilizando a temperatura da superfície do mar do Atlântico (ASST), e com 30 membros utilizando a TSM do Pacífico (SSST), ambas com frequência mensal.

Delsole & Shukla (2006), analisaram as características do inverno norte americano utilizando quatro modelos de circulação global (GCMs) e um modelo de regressão linear múltipla com preditor derivado de observações. O modelo COLA 2.2 estava entre os modelos utilizados para análise, onde foram utilizados os dados de previsões de precipitação baseadas na temperatura da superfície do mar. Das técnicas utilizadas, nenhuma foi capaz de melhorar as previsões em todos os modelos avaliados. Os resultados mostraram que as anomalias da TSM (El Niño e La Niña) influenciaram nas previsões realizadas com o modelo COLA.

Cash et al. (2007) analisaram as ligações entre a temperatura da superfície do mar (*Sea Surface Temperature* - SST) do Pacífico tropical e o clima regional de Bangladesh, utilizando o modelo COLA para quantificar a magnitude e a extensão geográfica da ligação. Os autores constataram que as chuvas são intensificadas ao longo de Bangladesh durante os eventos *El Niño*, onde as alterações na circulação global levam ao aumento da convergência sobre a cidade.

Zheng et al. (8394), avaliaram o desempenho do modelo COLA em simular a variabilidade interanual do inverno global, usando um conjunto de 50 anos de simulações, realizadas para TSM observada e comparados com os resultados de dados de reanálise do modelo NCEP-NCAR. Os resultados mostraram uma significativa correlação entre as previsões realizadas para o inverno e os fenômenos ENSO, concluindo que as simulações do modelo obtiveram bons resultados.

Misra (2003) avaliou a previsibilidade da precipitação sazonal do verão da América do Sul. Foram utilizados os membros do modelo de circulação global COLA, para realização das análises por conjunto (*ensemble*). O modelo foi calibrado utilizando as TSM previstas. Os resultados mostraram que as variações sazonais superaram as simuladas pelo modelo, onde a climatologia avaliada nos meses de janeiro, fevereiro e março era superiores aos valores observados, exceto pela região nordeste, onde as simulações mostraram melhores resultados.

3 PREVISÕES HIDROLÓGICAS

3.1 Aspectos gerais

Previsão hidrológica pode ser entendida como a estimativa de valores futuros das variáveis hidrológicas, com definida antecedência temporal (Meller, 2012). Logo, é correto afirmar que a motivação maior para a previsão é a incerteza sobre os eventos futuros. Contudo, de um modo geral, a previsão não elimina a incerteza, ela apenas a reduz, ou a quantifica. A incerteza das variáveis hidrológicas evolui com o tempo e para prazos mais longos, podendo ser quantificada em termos de uma função de distribuição de probabilidades, estimada a partir de dados climáticos (Krzysztofowicz, 2001).

O conhecimento das variáveis meteorológicas e hidrológicas de longo prazo é importante no planejamento de recursos hídricos, pois as previsões consideram as variações sazonais e interanuais da disponibilidade dessas variáveis. Especialmente na ocorrência de secas ou cheias excepcionais, onde dispor destas informações pode levar à preparação antecipada de medidas preventivas contra as calamidades previstas (Galvão, 1999). Por esta razão, a previsão é entendida como uma das técnicas utilizadas para minimizar o impacto das incertezas do clima na gestão dos recursos hídricos (Tucci & Collischonn, 2003).

O intervalo de tempo no futuro, o qual define a antecedência da previsão, recebe o nome de horizonte de previsão e pode ser classificado em curto e longo prazo: a) a previsão de curto prazo é realizada com antecedência de horas ou de alguns dias; b) a previsão de longo prazo é realizada com antecedência de semanas ou meses, a qual recebe muitas vezes o nome de previsão sazonal (Collischonn & Tucci, 2005). A frequência da previsão define o período de tempo no qual são emitidos novos dados de previsão. Logo, uma previsão de nível que se estende por cinco dias e que é repetida a cada dia, possui uma frequência diária e um horizonte de cinco dias (Bravo et al, 2009).

Outra classificação dada às previsões é dita como determinística ou probabilística: a) a previsão determinista especifica uma estimativa pontual da variável prevista; b) a previsão probabilística apresenta uma função de distribuição de probabilidade da variável prevista, com a finalidade de informar aos usuários (as agências governamentais, o público em geral) sobre a incerteza associada as previsões, fornecendo uma base para tomada decisões.

Para realização das previsões são necessários muitos anos de antecedência de dados ou apenas alguns minutos de antecedência. A quantidade de dados disponíveis torna-se, algumas

vezes, um fator determinante para o método a ser utilizado para previsão. Pois existem métodos que exigem muitos dados para realizarem boas previsões e métodos que necessitam de poucos dados para efetuarem previsões com bons resultados. Qualquer que sejam as circunstâncias ou horizontes de tempo envolvido, a previsão é uma ajuda importante para um planejamento eficiente e eficaz.

Pode-se dizer que a previsão de variáveis hidrológicas de longo prazo é uma área em constante desenvolvimento, apresentando grandes avanços ao longo dos últimos anos. Este fato é justificado graças ao grande número de dados coletados e ao uso de computadores de grande capacidade de processamento (Bravo, 2010).

Nos próximos tópicos deste capítulo serão abordadas as bases teóricas referentes à modelagem e previsão de longo prazo, utilizando modelos empíricos para realização das simulações, no qual está inserido o objetivo deste trabalho.

3.2 Previsões hidrológicas de longo prazo

Os níveis futuros no banhado, que constituem o objeto principal desse estudo, são desconhecidos e assim pode-se pensar neles como uma variável aleatória. Existe uma diversidade de valores possíveis para os níveis, mas até que o evento ocorra, não se sabe qual valor exato será. Contudo, quanto mais próximo é o mês previsto, tem-se mais confiança em prever os prováveis valores que os níveis podem assumir. À medida que o horizonte aumenta, as incertezas vão aumentando, tornando os valores futuros mais variáveis. Logo, quanto maior o horizonte previsto, mais incertas são as previsões.

A previsão de variáveis hidrológicas pode ser realizada para diversas finalidades e o horizonte de previsão está intimamente ligado ao interesse da investigação. Assim, a previsão de longo prazo, que comumente é realizada para um horizonte de até 9 meses, frequentemente é utilizada para propósitos de planejamento e manejo dos recursos hídricos (Silva et al. 2006, Bravo, 2010). Pode-se dizer que as previsões de futuros cenários de variáveis hidrológicas é uma ferramenta indispensável para o gerenciamento dos recursos hídricos.

Existem diversas atividades que possuem uma estreita dependência a variáveis hidrológicas e o conhecimento antecipado dessas variáveis pode ajudar no planejamento e operação dessas atividades. A agricultura é um típico exemplo de atividade que possui uma grande vulnerabilidade climática interanual, pois o tempo de semeadura, o preparo do solo, o manejo da irrigação, as estimativas de produtividade e colheita, são exemplos de práticas agrícolas que dependem da variabilidade do clima (Bravo, 2010).

3.3 Modelos hidrológicos

Os modelos hidrológicos são representações matemáticas simplificadas do comportamento dos sistemas hídricos de modo a produzir resultados próximos aos encontrados na natureza, utilizando séries de equações com dados de entrada diversos, gerando informações hidrológicas como vazão em rios, balanço de água no solo e níveis da água em lagos. A representação dos fenômenos naturais é uma tarefa bastante complexa, por essa razão, a modelagem hidrológica trabalha com simplificações desses fenômenos.

Os modelos de simulação são os principais meios utilizados para previsão dos futuros prováveis valores de séries temporais de variáveis hidrológicas, portanto, oferecem uma oportunidade para o gestor prever variáveis de decisão, tais como os níveis da água máximos de um lago, e sua persistência ao longo do tempo. A análise de séries temporais é uma técnica usada na construção de modelos de simulação, identificando, a partir da série histórica, tendências, saltos, sazonalidades e componentes estocásticas que constituem uma base para os estudos de simulação hidrológica (Sen et al., 2000).

A forma como os modelos são elaborados os classificam em conceituais (de base física) ou empíricos (caixa-preta): a) os modelos ditos conceituais buscam representar os processos hidrológicos através de equações, que são baseadas nos processos físicos que envolvem os fenômenos hidrológicos. Sua formulação possui a vantagem de simular cenários diferentes daqueles usados na calibração, ou seja, possibilidade de extrapolação de cenários com maior consistência metodológica (Tucci, 1998). Um problema na utilização desta metodologia consiste na variabilidade espacial dos parâmetros e do grande número de variáveis que envolvem a simulação, tornando os modelos pouco parcimoniosos; b) os modelos empíricos (ou caixa-preta) são baseados em relações empíricas entre as variáveis de entrada e as variáveis de saída, onde não é considerado nenhum conceito físico dos processos hidrológicos (Tucci, 2005). Dessa forma, procuram ajustar os dados calculados aos dados observados, sem considerar explicitamente a relação entre as funções e os processos físicos.

Os modelos matemáticos conceituais, na teoria, apresentam vantagens com respeito a representação dos fenômenos hidrológicos, em contrapartida, possuem dificuldades quanto a atualização contínua dos parâmetros. Segundo Tucci (2005), a definição de modelo conceitual é artificial, uma vez que são utilizadas funções empíricas, como as equações de Darcy e Horton, na elaboração desses modelos. Por este motivo, alguns autores utilizam uma subdivisão entre estes modelos: semi-conceituais e físicos.

As principais desvantagens dos modelos empíricos consistem na sua falta de representação física dos fenômenos relacionados à simulação e na sua dependência aos dados históricos disponibilizados para o ajuste. Entretanto, possuem algumas vantagens importantes, como a facilidade de elaboração e a simplicidade de cálculo.

Cabe ressaltar que, independentemente da metodologia empregada para formulação do modelo hidrológico (conceitual ou empírico), o fato do modelo estar devidamente calibrado e possuir bons indicadores de desempenho, não implica dizer que o modelo irá obter bons resultados na previsão.

Nos próximos tópicos serão abordadas algumas metodologias utilizadas para formulação e análise de um modelo empírico, com enfoque dado a previsão e simulação de variáveis hidrológicas.

3.4 Modelos empíricos

Uma das traduções da palavra “empírica” na língua portuguesa é que a metodologia esta pautada somente em experiências vividas, ou seja, na observação de fatos e não em teorias e métodos científicos. Em outras palavras, o empírico é aquele conhecimento adquirido durante toda a vida, no dia-a-dia que não possui comprovação científica. Contextualizando para hidrologia, é razoável supor que alguns aspectos dos padrões do passado possam continuar no futuro. Assim, a utilização de uma metodologia empírica pode ser usada desde que existam informações numéricas sobre o passado das variáveis que se deseja prever.

De um modo geral, um modelo empírico de previsão consiste em uma equação de recorrência, onde os valores previstos são obtidos a partir de séries históricas das variáveis intervenientes no processo. Não existe uma preocupação em representar o conhecimento do fenômeno através de leis físicas, os parâmetros são obtidos através de ajustes baseados na série histórica das variáveis. Por não fornecerem explicações sobre o funcionamento dos fenômenos, mas apenas os resultados, são chamados de modelos do tipo "não conceituais". Os modelos empíricos fazem uso de dados anteriores e atuais da própria variável e outras variáveis exógenas, na realização das previsões para um alcance determinado (Pedrollo, 2000).

A introdução de modelos baseados em inteligência artificial no campo dos recursos hídricos, como os modelos empíricos, deu início a novas abordagens na modelagem hidrológica. Esses modelos têm sido amplamente aplicados a uma variedade de problemas,

incluindo a simulação e a previsão hidrometeorológica. A metodologia utilizada por esses modelos provam ser uma alternativa eficiente aos métodos tradicionais, como a aplicação de regressão linear múltipla, ARMA, ARIMA, suavização exponencial e as redes neurais artificiais para a modelagem de variáveis hidrológicas, tanto em aspectos qualitativos como quantitativos (Karamouz et al., 2008).

A utilização de diferentes arquiteturas baseadas em modelos empíricos para modelagem de recursos hídricos tem-se mostrado eficiente na capacidade de representar os sistemas hidrológicos. Esses modelos comprovaram ser capazes de extrair informações relevantes a partir de índices climáticos complexos, melhorando assim a compreensão de previsões em escalas regionais (Coulibaly et al., 2000).

As previsões empíricas são baseadas em séries temporais e são úteis quando a variável que se deseja prever está variando ao longo do tempo (por exemplo: precipitação mensal). O objetivo de prever dados de uma série temporal consiste em estimar como a sequência de observações vai continuar no futuro. Assim, as previsões podem ser realizadas apenas com a série temporal da variável a ser prevista (autorregressivos), ou com a incorporação de outras variáveis.

Contudo, a relação entre os dados nem sempre é exata, como será visto nos próximos capítulos deste trabalho, podendo haver mudanças nos dados que não podem ser explicados pelas variáveis independentes. Tornando difícil medir as relações que governam o seu comportamento.

3.4.1 Correlação

Em estudos que envolvem simulações (ou previsões) de variáveis hidrológicas, que utilizam uma metodologia empírica, é comum o interesse em conhecer o relacionamento entre as variáveis disponíveis. A mensuração do grau de relacionamento entre as variáveis é conhecida como medida de associação ou coeficiente de correlação. Existem diversas formas de correlação entre variáveis, aonde para esse estudo será dado enfoque à discretização em linear e não linear.

A teoria da probabilidade e estatística aborda esse tema, correlação ou co-relação referindo-se à medida da relação entre duas variáveis, embora correlação não implique causalidade. Neste sentido, existem vários coeficientes medindo o grau de correlação, adaptados à natureza dos dados. Esses coeficientes são utilizados em diferentes situações, que de um modo geral, não são garantias de relação entre os dados, podendo ocorrer falhas em

capturar dependência em algumas circunstâncias, sendo recomendada a utilização de outras medidas de associação.

3.4.1.1 Coeficiente de correlação linear de Pearson

O termo “co-relação” foi proposto por Galton em 1888. Karl Pearson, que era seu aluno, desenvolveu a fórmula matemática que é usada atualmente e que tem seu nome por homenagem. O símbolo do coeficiente de correlação amostral r vem da primeira letra da palavra regressão, em reconhecimento a Galton (SCHULTZ e SCHULTZ, 1992).

Admite-se que duas variáveis apresentam uma correlação linear quando os pontos, analisados em um diagrama de dispersão, aproximam-se de uma reta. Essas correlações são ditas: a) positivas, quando para os valores crescentes de uma das variáveis, exista uma tendência a valores crescentes na outra; b) negativa, quando para valores crescentes de uma das variáveis, observa-se uma tendência de valores decrescente na outra (Figura 5).

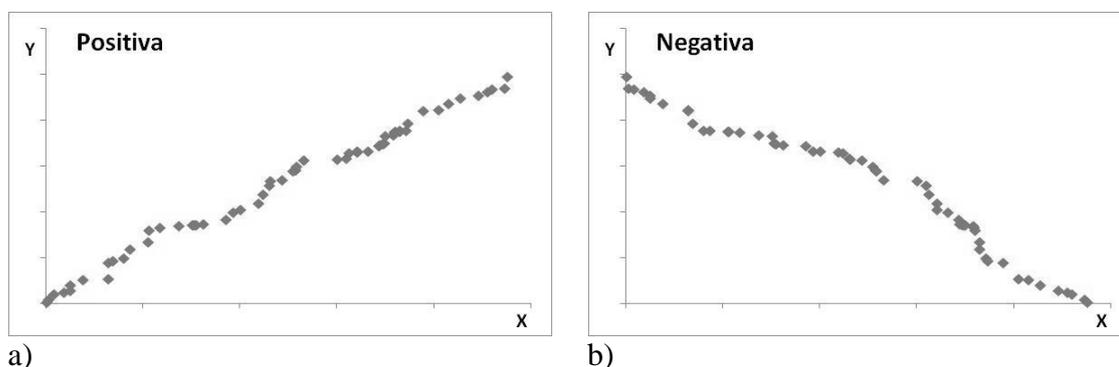


Figura 5 - Exemplo de correlação: a) correlação positiva; b) correlação negativa.

O coeficiente de correlação linear, também conhecido como coeficiente de Pearson, mede o grau de correlação e a direção dessa correlação (positiva ou negativa), entre duas variáveis aleatórias (Naghetini & Pinto, 2007). Quanto mais forte a relação linear, os pontos tendem a ficarem mais perto da linha de regressão. Esse coeficiente assume apenas valores entre -1 e 1, podendo ser expresso pela equação 1:

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \quad (1)$$

onde $\sigma_{X,Y}$ é a covariância entre as variáveis X e Y ; σ_X e σ_Y são o desvio-padrão da variável X e Y , respectivamente.

Na ocorrência das variáveis X e Y serem estatisticamente independentes, o coeficiente de correlação linear é $\rho = 0$. No entanto, se o coeficiente de correlação linear for igual à zero, $\rho = 0$, não significa que as variáveis sejam independentes, podendo existir dependência não

linear. Quando $\rho = 1$, existe uma correlação perfeita positiva entre as duas variáveis, e quando $\rho = -1$, existe uma correlação negativa perfeita, ou seja, se uma aumenta a outra sempre diminui.

Para o cálculo do coeficiente de correlação linear é necessário calcular a covariância entre duas variáveis. A covariância consiste em uma medida do grau de interdependência (ou inter-relação) numérica entre duas variáveis aleatórias. Ou seja, é uma medida estatística que quantifica o quanto as duas variáveis se alteram. Podendo ser interpretado como momento de primeira ordem de duas variáveis aleatórias, centrado nas respectivas médias, expressa pela equação 2:

$$\sigma_{X,Y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1} \quad (2)$$

onde $\sigma_{X,Y}$ é a covariância amostral entre as variáveis X e Y ; \bar{x} e \bar{y} são as médias aritméticas de cada uma das variáveis; n é o tamanho da amostra; x_i e y_i são as observações simultâneas das variáveis.

O desvio-padrão das amostras são calculados pelas equações 3 e 4:

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (3)$$

$$\sigma_Y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}} \quad (4)$$

3.4.2 Regressão

Dentre os métodos estatísticos existentes para analisar associações entre duas ou mais variáveis aleatórias, o método mais difundido é a teoria da regressão. Podem-se citar como exemplos de associações entre variáveis aleatórias hidrológicas, as relações: a) entre as intensidades, durações e frequências das precipitações intensas; b) entre as vazões médias anuais e as áreas de drenagem; c) entre as alturas de precipitação e as altitudes dos postos pluviométricos.

Os modelos de regressão são modelos matemáticos que relacionam o comportamento de uma variável Y com uma função $f(x_1, x_2, \dots, x_n)$. Quando a função f possui apenas uma variável independente, temos o modelo de regressão simples. Um modelo de regressão é

múltiplo quando envolve uma relação causal com duas ou mais variáveis independentes. Isso ocorre quando o comportamento de y pode ser explicado por mais de uma variável.

Uma primeira abordagem desse método consiste na análise do comportamento simultâneo das variáveis tomadas duas a duas, por meio do diagrama de dispersão. Os diagramas de dispersão são gráficos que permitem a identificação entre causa e efeito, avaliando assim relacionamentos entre as variáveis e o grau de intensidade dessas relações. Verificando se a variação de uma delas está associada à variação da outra, ou se não existe nenhuma forma de dependência entre elas. Cabe ressaltar que o fato de duas variáveis estarem altamente correlacionadas não resulta necessariamente qualquer relação de causa e efeito, o contrário também é verdadeiro. As evidências devem ser obtidas com base no conhecimento dos processos envolvidos na geração dessas variáveis (Naghetini & Pinto, 2007).

Existem muitos benefícios quanto à utilização de diagramas de dispersão para análise de dados, pois deles podemos inferir relações casuais entre as variáveis, tornando visível a identificação de possíveis problemas em determinado período de tempo. Podendo ser utilizado como ferramenta de análise de qualidade. Mas a usual utilização é como um método gráfico de análise que permite verificar a existência ou não de relação entre duas variáveis de natureza quantitativa.

Definidos os dados de entrada e suas correlações, precisamos estabelecer a equação de regressão e os parâmetros que a compõe. Assim, dependendo da quantidade de dados que serão utilizados na regressão, a equação pode ser simples ou múltipla, linear ou não-linear. Então os tipos de relação, ao qual estão correlacionados os dados e a regressão encontram-se fortemente ligadas. A vantagem da regressão é que ela afirma uma relação preditiva entre as variáveis, quantificando isso de uma forma que seja útil para a previsão.

3.4.2.1 Regressão linear

A regressão linear é uma técnica estatística utilizada para estimar a relação entre variáveis, investigando o valor esperado de uma variável Y , dados os valores das variáveis X , considerando que exista um relacionamento funcional entre as variáveis Y e X . Uma regressão é denominada linear, quando a relação da resposta as variáveis for uma função linear dos parâmetros. No entanto, as regressões que não apresentam uma característica linear são chamadas de não-lineares (Naghetini & Pinto, 2007).

Quando apenas uma variável independente relaciona-se com a variável dependente (de previsão), dizemos que é um modelo linear simples. O qual fornece um valor médio de uma

das variáveis em função da outra, onde os pontos observados apresentam uma variação nas proximidades da linha que representa a função de regressão. A linha de regressão pode ser estimada pela equação 5:

$$Y_i = \alpha + \beta \cdot X_i + e_i \quad (5)$$

onde Y_i é a variável dependente (é o valor que se deseja conhecer), X_i é a variável independente (representa o fator explicativo na equação), α é uma constante que representa o intercepto com o eixo vertical, β é uma constante que representa o coeficiente angular da reta e e_i constitui os erros ou resíduos da regressão. Os parâmetros α e β são estimados com base nos dados utilizados para realizar a regressão.

Os parâmetros dos modelos lineares podem ser estimados pelas equações 6 e 7:

$$\beta = \frac{\sum_{i=1}^n x_i \cdot y_i - n \cdot \bar{y} \cdot \bar{x}}{\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2} \quad (6)$$

$$\alpha = \bar{y} - \beta \cdot \bar{x} \quad (7)$$

Assim determinamos a linha de regressão, que é utilizada para realizar as previsões. Para cada valor de x_i podemos prever um valor correspondente de y_i . Contudo, é importante notar que a interpretação do intercepto da linha de regressão com o eixo y , requer que o valor de $x = 0$ faça sentido. Portanto, deve-se tomar cuidado quanto a análise dos dados que serão simulados.

A regressão linear múltipla (RLM) procura estudar o comportamento de uma variável dependente Y_i em função de duas ou mais variáveis independentes $X_{j,i}$. No caso da variável Y_i possuir um comportamento linear com as variáveis $X_{j,i}$, pode-se adotar um modelo geral seguindo a forma da equação 8:

$$Y_i = \beta_1 \cdot X_{1,i} + \beta_2 \cdot X_{2,i} + \dots + \beta_p \cdot X_{p,i} \quad (8)$$

onde Y_i é a variável dependente ou prevista; $X_{j,i}$ são as variáveis independentes ou explicativas e β_i são os coeficientes de regressão.

Os modelos de regressão linear são consideradas importantes ferramentas para previsão de variáveis hidrológicas a longo prazo, tendo sido usado para prever a distribuição de neve acumulada (Moreno et al, 2010), analisar os efeitos das mudanças climáticas sobre as concentrações de nitrato em grandes rios (Zweimuller et al., 2008), encontrar relações chuva-vazão sazonais em planícies arborizada (Torres et al., 2011), avaliação das tendências de escoamento futuras sob vários cenários de mudanças climáticas (Jung & Chang, 2011), dentre outras previsões de longo prazo.

3.4.2.2 Regressão não-linear

Existem situações nas quais não é desejável, ou mesmo possível, descrever um fenômeno através de um modelo de regressão linear, dessa forma ao invés de se fazer uma descrição puramente empírica do fenômeno em estudo, pode-se trabalhar no sentido de obter uma relação teórica entre as variáveis observadas.

A regressão não-linear é uma forma de analisar dados observados que não se comportam de forma linear, em que os dados são modelados por uma função que é uma combinação não-linear de parâmetros do modelo, dependendo de uma ou mais variáveis independentes. Os dados são ajustados utilizando métodos de aproximações sucessivas, tais como a regressão linear.

Segundo Mattos (2013), o modelo de regressão não linear pode ser escrito como a equação 9:

$$Y_i = f(X_j, \theta) + \varepsilon_i \quad (9)$$

em que Y_i é a variável resposta, f é uma função não linear contínua com forma conhecida do vetor de variáveis explicativas X_j , e dos parâmetros desconhecidos θ , ε_i são os erros que possuem média zero, variância constante, e não são correlacionados entre si.

A estimativa dos parâmetros pode ser realizada utilizando métodos matemáticos como os mínimos quadrados e verossimilhança.

3.4.2.3 Mínimos quadrados

O método de estimação por mínimos quadrados é uma técnica de otimização matemática que, minimiza a soma dos quadrados das diferenças entre os valores observados de uma amostra e seus respectivos valores esperados, onde o objetivo é minimizar a equação 10:

$$f(\theta) = \sum_{i=1}^n (Y_i - \theta X_i)^2 \quad (10)$$

O valor mínimo é obtido calculando-se as derivadas parciais em função dos parâmetros, pela equação 11:

$$\frac{\partial f(\theta)}{\partial \theta} = \sum_{i=1}^n (Y_i - \theta X_i)(-2X_i) = 0 \quad (11)$$

Resolvendo a equação acima obtêm-se a equação 12:

$$\theta = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} \quad (12)$$

3.4.2.4 Resíduos

Os valores resultantes das previsões de y_i , obtidas a partir das variáveis independentes $X_{j,i}$, são chamadas de “valores ajustados”. Nesse caso, escrevemos \hat{y}_i como sendo o valor simulado pela equação de regressão para $i = 1, \dots, n$ onde cada valor de \hat{y}_i representa um ponto da linha (ou curva) de regressão correspondente as variáveis $X_{j,i}$.

A diferença entre os valores observados de y_i e os valores ajustados \hat{y}_i correspondentes são os “resíduos” (equação 13).

$$e_i = y_i - \hat{y}_i \quad (13)$$

Os resíduos possuem algumas propriedades úteis (equação 14), dentre elas podemos destacar:

$$\sum_{i=1}^n e_i = 0 \quad \text{e} \quad \sum_{i=1}^n X_i \cdot e_i = 0 \quad (14)$$

Das propriedades acima podemos concluir que a média dos resíduos é zero e, que a correlação entre os resíduos e as observações das variáveis de previsão é zero.

Como os resíduos resultam das subtrações entre os valores observados e os valores ajustados é de se esperar que os resíduos possuam distribuição aleatória, sem demonstrar qualquer tipo de padrão sistemático. Uma maneira simples e rápida para uma primeira verificação desse resultado consiste em analisar o gráfico de dispersão dos resíduos em relação à variável prevista. Constatado um padrão não aleatório, para os resíduos de um modelo linear, isso pode ser um indicativo de uma relação não linear entre os dados, sugerindo que um modelo linear não seja adequado para realizar previsões com esses dados (Hyndman & Athanasopoulos, 2013).

O termo “erro” muitas vezes é designado para os resíduos, porém os resíduos não devem ser interpretados como um erro e sim como um desvio do modelo de regressão para os valores observados. Os resíduos são boas ferramentas no rastreamento de viés, pois se sua média for diferente de zero, implica dizer que as previsões serão sistematicamente tendenciosas.

3.4.2.5 Colinearidade e multicolinearidade

Um problema comum em regressões múltiplas ocorre quando as variáveis independentes estão correlacionadas umas com as outras. Isso porque variáveis colineares não

forneem novas informações, o que acaba dificultando a interpretação dos coeficientes obtidos na regressão. As consequências percebidas são erros padrão elevados, sinal do coeficiente de regressão oposto do esperado, resultando em estimativas pouco confiáveis. Uma alternativa para contornar esse problema consiste na eliminação de uma, entre cada conjunto de duas variáveis explicativas, que apresentarem coeficientes de correlação superiores a 0,85. Dessa forma, espera-se que as variáveis mantidas no modelo de regressão contribuam significativamente para explicar a variabilidade da variável resposta (Cruz & Carneiro, 2006; Naghettini & Pinto, 2007). Entretanto, se as variáveis não estão correlacionadas entre si, são chamadas de ortogonais. Na prática é muito difícil que as variáveis de entrada sejam ortogonais, contudo a falta de ortogonalidade não é prejudicial.

Embora muitas vezes os termos colinearidade e multicolinearidade serem usados como sinônimos, alguns autores fazem distinção entre os mesmos. Assim, colinearidade é a existência de relação linear entre duas variáveis explicativas (matriz de correlação) e multicolinearidade é a existência de relação linear entre uma variável explicativa e as demais (Gujarati, 2006). Algumas características no ajuste do modelo podem indicar problema de colinearidade ou multicolinearidade:

- β_i com sinal oposto ao esperado;
- Grandes mudanças em β_i quando adicionamos ou excluimos variáveis ou observações;
- β_i não significativa para um X_i teoricamente importante.

onde β_i é o coeficiente de regressão para a variável X_i .

A ocorrência de colinearidade e multicolinearidade podem ser diagnosticadas verificando se a matriz de correlação das variáveis explicativas apresenta correlações. Se a correlação de duas ou mais variáveis for próxima de 1, indica problema.

3.4.2.6 Teste de significância da regressão linear múltipla

Quando se trabalha com regressão múltipla é comum a utilização do teste de significância para avaliar a existência de uma relação significativa entre a variável dependente e as variáveis explicativas (Naghettini & Pinto, 2007). O teste é realizado para duas hipóteses:

- $H_0: \beta_1 = \beta_2 = \dots = \beta_n = 0$ (a relação entre as variáveis não é linear);
- H_1 : pelo menos um $\beta_i \neq 0$.

Esse teste também conhecido como “teste do F total” e é utilizada para testar a razão entre duas variâncias, sendo empregado para verificar a hipótese nula. O valor de F é calculado pela equação 15:

$$F = \frac{SQM\ Reg}{SQM\ Res} \quad (15)$$

Onde $SQM\ Reg$ é o somatório dos quadrados médios da regressão e $SQM\ Res$ é o somatório dos quadrados médios dos resíduos.

A hipótese nula será aceita se $F < F(\alpha, P, n - P - 1)$, onde alfa é o nível de significância, p e $n-p-1$ são os graus de liberdade da distribuição de Snedecor, sendo p o número de variáveis explicativas.

3.5 Variáveis de entrada

As variações temporais e espaciais dos fenômenos que envolvem o ciclo hidrológico na atmosfera podem ser descritas pelas variáveis hidrológicas, através de medições realizadas em locais distintos. Estas amostras podem ser usadas para caracterizar ou inferir informações sobre estas variáveis. De um modo geral, as variáveis hidrológicas são registradas por meio das chamadas séries históricas, que constituem as observações de sua ocorrência no tempo. A partir de análises realizadas nas séries históricas das variáveis hidrológicas ao longo dos anos, verificou-se um comportamento aleatório dos eventos, em que poderiam ser descritos por distribuições de probabilidade. Uma forma de simular as variáveis hidrológicas são os modelos matemáticos que são usados para prever futuros valores para essas variáveis.

Os modelos de simulação (ou previsão) estão sujeitos a diversas incertezas. Essas imprecisões podem ser devido a cobertura incompleta de dados, erros de medição, ou mesmo imperfeições da própria estrutura do modelo devido sua parametrização de processos físicos, como também de problemas referentes as escalas que envolvem os processos simulados. Mesmo partindo da suposição de um modelo com estrutura perfeita, as estimativas dos parâmetros podem ser incertas, dadas as incertezas observacionais que afetam a calibração do modelo (Liu et al, 2012).

Dentre as diversas variáveis climáticas, a precipitação é considerada a mais difícil de prever. Isso ocorre devido ao comportamento caótico da atmosfera, onde a estimativa de precipitação é limitada pela baixa representatividade espacial da maioria das redes de estações (Silva et al, 2006; Renard et al, 2010). Dessa forma, é fácil concluir que o impacto dos erros

nas estimativas da precipitação acarreta em incertezas nas respostas dos modelos de previsão de variáveis hidrológicas.

Existe uma gama de variáveis de entrada possíveis para previsões hidrológicas usando modelos empíricos. Assim, a escolha das variáveis de entrada requer um estudo das condições físicas que mais influenciam na variável hidrológica que será prevista. Logo é importante compreender os processos hidrológicos que envolvem o estudo. Os modelos empíricos estão intimamente ligados aos dados de entrada, pois deles é que são estabelecidos os parâmetros do modelo e sua forma (linear, exponencial, logarítmica, etc.). Logo é fácil deduzir que a validade desses modelos está inserida nos limites (amplitude) dos dados que o originaram. Assim, para o período de formulação da equação do modelo, devem-se analisar cuidadosamente os dados que serão usados nesta etapa. Pois, para o modelo ser eficaz, essas amostras necessitam serem representativas do comportamento hidrológico que se deseja simular. O processo de elaboração de um modelo hidrológico normalmente envolve 3 etapas: a) formulação (ou elaboração); b) verificação (ou validação) e; c) simulação. Para cada uma dessas etapas são necessárias amostras dos dados, onde se faz necessário repartir a série temporal de dados disponíveis para realizar cada etapa.

3.5.1 Série temporal

Uma série temporal é uma sequência de pontos de dados, tipicamente medidos em instantes de tempo sucessivamente espaçados em intervalos equidistantes. Normalmente, os dados de séries temporais possuem uma ordenação temporal natural, o que torna sua análise distinta de outros problemas de dados comuns, onde não existe ordem natural das observações (Kattan et al, 2015). Assim, pode-se supor que, em uma série temporal, existem fenômenos (fatores) relacionados com o tempo, que influenciaram no passado e pode continuar no futuro. O objetivo principal da análise de séries temporais consiste em identificar padrões não aleatórios e a detecção desse comportamento passado pode permitir fazer previsões sobre os futuros valores da série.

3.5.1.1 Testes de hipótese não paramétricos

Uma série de dados hidrológicos são amostras de observações, ou inferência das amostras, utilizadas na realização de estudos para uma melhor compreensão dos fenômenos envolvidos, auxiliando a tomada de decisões. Entretanto, da simples visualização dos dados não é possível realizar inferências sobre o comportamento estatístico da amostra. Uma

importante ferramenta para o entendimento concernente a forma da uma distribuição probabilística são os testes de hipóteses. Essas deduções partem do pressuposto de rejeitar ou não rejeitar a conjectura sugerida, confrontadas com as observações amostrais. A rejeição afirma que a hipótese proposta não corresponde à realidade física da amostra. Contudo, não rejeição afirma que não existem argumentos suficientes para descartar a proposta recomendada. Cabe ressaltar que a não rejeição da hipótese não implica em aceitá-la. Em sua maioria, a série de dados hidrológicos são considerados aleatórios e a decisão quanto rejeitar ou não rejeitar uma hipótese, deve ser baseada em certo nível de significância, para reduzir o grau de subjetividade associada a decisão de um teste de hipótese (Naghetini & Pinto, 2007).

Os testes de hipótese constituem uma importante ferramenta na detecção de comportamentos na série histórica. Admitindo um conjunto de observações de uma variável hidrológica, pode-se avaliar as hipóteses de aleatoriedade, independência, homogeneidade e estacionariedade dessa série, avaliando as características distributivas das variáveis e pelo número de elementos, podem ser realizados pelos testes não-paramétricos. Os testes não paramétricos abordam técnicas que não dependem de dados pertencentes a nenhuma distribuição particular.

O termo aleatório é utilizado para exprimir quebra de ordem ou imprevisibilidade. Um processo aleatório é o seguimento repetitivo cujo resultado não descreve um padrão determinístico, mas segue uma distribuição de probabilidade. Contextualizando para hidrologia, pode-se afirmar que uma variável transcorre de causas naturais se suas oscilações possuem um progresso aleatório. A detecção de aleatoriedade nos dados hidrológicos é de difícil demonstração, mas pode ser rejeitada a hipótese por evidências de natureza não aleatórias nos dados. NERC (1975) sugere que a hipótese de aleatoriedade de uma série hidrológica possa ser rejeitada ou não-rejeitada por meio do teste não-paramétrico baseado nos pontos de virada (*turning point*). O teste pode ser realizado da seguinte forma: 1) contabiliza-se o número de inflexões na série; 2) calcula-se a média (equação 16); 3) calcula-se a variância da série (equação 17); 4) realiza o teste por meio da estatística não paramétrica (equação 18).

$$E[p] = \frac{2(N - 2)}{3} \quad (16)$$

$$Var[p] = \frac{16.N - 29}{90} \quad (17)$$

$$T = \frac{p - E[p]}{\sqrt{Var[p]}} \quad (18)$$

onde p é o número de pontos de virada, N é o número total de dados, $E[p]$ é a média, $Var[p]$ é a variância e T é a estatística de teste. O teste é bilateral realizado para $N > 30$ a um nível de significância α , a decisão deve ser a de rejeitar a hipótese nula se $|T| > Z_{1-\alpha/2}$.

Na língua portuguesa, o termo independência significa a desassociação de um ser em relação a outro. Em estatística, essa denominação pode ser interpretada de forma que nenhuma observação da série analisada pode influenciar na ocorrência, ou não ocorrência, de outra observação presente na amostra.

A hipótese de independência de uma série hidrológica pode ser rejeitada ou não-rejeitada por meio do teste não-paramétrico proposto por Wald e Wolfowitz (1943 apud Naghettini & Pinto, 2007). O teste avalia a independência temporal da sequência de dados, ou seja, se um intervalo é independente do intervalo anterior. Dada uma amostra $\{X_1, X_2, \dots, X_N\}$, com tamanho N , onde as diferenças entre as observações X_i e a média amostral \bar{X} são $\{X'_1, X'_2, \dots, X'_N\}$. A estatística do teste é dada pela equação 19:

$$R = \sum_{i=1}^{N-1} X'_i \cdot X'_{i+1} + X'_1 \cdot X'_N \quad (19)$$

Admitindo a hipótese de que as observações são independentes, a média e a variância pode ser calculada pelas equações 20 e 22. A estatística não paramétrica é dada pela equação 23:

$$E[R] = -\frac{s_2}{N-1} \quad (20)$$

$$Var[R] = \frac{s_2^2 - s_4}{N-1} + \frac{s_2^2 - 2 \cdot s_4}{(N-1) \cdot (N-2)} - \frac{s_2^2}{(N-1)^2} \quad (21)$$

$$s_r = \sum_{i=1}^N (X'_i)^r \quad (22)$$

$$T = \frac{R - E[R]}{\sqrt{Var[R]}} \quad (23)$$

onde s_r é o momento amostral em relação a origem. O teste é bilateral realizado a um nível de significância α , a decisão deve ser a de rejeitar a hipótese nula se $|T| > Z_{1-\alpha/2}$.

A homogeneidade é uma característica que um corpo apresenta em sua composição ou aparência completamente igual, em qualquer parte analisada. Em estatística, pode ser interpretada quando todos os elementos de uma certa amostra provêm de uma única e idêntica população. A rejeição da hipótese de homogeneidade implica dizer que a série é heterogênea, porém a heterogeneidade de amostras muito pequenas é de difícil detecção. O teste de

homogeneidade pode ser realizado pelo método proposto por Mann e Whitney (1947 apud Naghettini & Pinto, 2007).

Dada uma amostra $\{X_1, X_2, \dots, X_N\}$, com tamanho N , separa-se em duas sub-amostras $\{X_1, X_2, \dots, X_{N_1}\}$, com tamanho N_1 e $\{X_{N_1+1}, X_{N_1+2}, \dots, X_N\}$, com tamanho N_2 , de maneira que $N_1 + N_2 = N$ sendo que N_1 e N_2 sejam aproximadamente iguais com $N_1 \leq N_2$. O próximo passo consiste em classificar em ordem crescente a amostra completa indicando a ordem de classificação m de cada observação, especificando se ela adveio da primeira ou da segunda sub-amostra. A estatística do teste é dada pelo menor valor entre as quantidades de V_1 e V_2 (equações 24 e 25):

$$V_1 = N_1 \cdot N_2 + \frac{N_1 \cdot (N_1 + 1)}{2} - R_1 \quad (24)$$

$$V_2 = N_1 \cdot N_2 - V_1 \quad (25)$$

onde R_1 é a soma das ordens de classificação dos elementos da primeira sub-amostra.

Admitindo a hipótese nula (H_0) de que a amostra é homogênea e que $N_1, N_2 > 20$, a média e a variância pode ser calculada pelas equações 26 e 27. A estatística do teste não paramétrico é dada pela equação 28:

$$E[V] = \frac{N_1 \cdot N_2}{2} \quad (26)$$

$$Var[V] = \frac{N_1 \cdot N_2 \cdot (N_1 + N_2 + 1)}{12} \quad (27)$$

$$T = \frac{V - E[V]}{\sqrt{Var[V]}} \quad (28)$$

Segundo explicado por Naghettini & Pinto (2007), se as duas sub-amostras não forem homogêneas, os elementos da primeira devem apresentar ordens de classificação mais baixas (ou mais elevadas), em relação às ordens de classificação da segunda. O teste é bilateral realizado a um nível de significância α , e a decisão deve ser a de rejeitar a hipótese nula se $|T| > Z_{1-\alpha/2}$.

Em análises de séries temporais, é comum em muitas técnicas a suposição de estacionariedade dos dados. Um processo estacionário possui a propriedade de que a média, a variância e a estrutura de autocorrelação não mudam no decorrer do tempo. Uma série temporal é dita estacionária quando ela se desenvolve no tempo de forma aleatoria ao redor de uma média constante. O teste de hipótese de estacionariedade pode ser realizado pelo teste não paramétrico de Spearman (Naghettini & Pinto, 2007), que está baseado no coeficiente de correlação entre as ordens de classificação m_i da sequência X_i e os índices de tempo T_i que são iguais a 1,2,3, ..., N . a estatística do teste é dada pela equação 29. Admitindo hipótese nula

(H_0) de que não há correlação entre m_t e T_t e $N > 10$, a média e a variância pode ser calculada pelas equações 30 e 31. A estatística do teste não paramétrico pode ser calculada pela equação 32:

$$r_s = 1 - \frac{6 \cdot \sum_{t=1}^N (m_t - T_t)^2}{N^3 - N} \quad (29)$$

$$E[r_s] = 0 \quad (30)$$

$$Var[r_s] = \frac{1}{N - 1} \quad (31)$$

$$T = \frac{r_s}{\sqrt{Var[r_s]}} \quad (32)$$

O teste é bilateral, realizado a um nível de significância α , a decisão deve ser a de rejeitar a hipótese nula se $|T| > Z_{1-\alpha/2}$.

3.5.1.2 Modelo clássico das séries temporais

Uma forma de analisar as séries temporais consiste na utilização de um modelo univariado, que utiliza formulações matemáticas simples, para separar a série em quatro componentes principais. Essa metodologia é conhecida por decomposição clássica e comumente é utilizada para realizar análises e previsões em séries temporais (Bouzada, 2012). Os quatro componentes são:

- Tendência (**T**), que se refere à direção geral segundo a qual a série temporal se desenvolve ao longo do tempo. Pode ser causado por mudanças que afetem a variável a longo prazo como urbanização, aquecimento global, etc.;
- Ciclo (**C**), que corresponde às oscilações a longo prazo ou aos desvios em torno da linha de tendência. As flutuações nos valores da variável comumente possuem duração superiores a um ano e se repetem com certa periodicidade, podem ocorrer por fenômenos climáticos como El Niño;
- Sazonalidade (**S**), são os padrões que uma série temporal está submetida. As flutuações geralmente possuem duração inferior a um ano e que se repetem todos os anos, geralmente em função das estações do ano.
- Termo aleatório (**A**), que aparece com flutuações de curto prazo e deslocamento inexplicável. Esses fatos fortuitos e inesperados podem ser resultados de catástrofes naturais.

A decomposição clássica permite que a série seja escrita como uma soma ou multiplicação de componentes não observáveis (Morettin & Tolo, 2004). Assumindo que as

componentes da série são somadas para formar os dados, podemos escrevê-la na forma da equação 33. Supondo que as componentes são multiplicadas para formar os dados, podemos escrevê-lo na forma da equação 34:

$$y_i = S_i + T_i + C_i + A_i \quad (33)$$

$$y_i = S_i \times T_i \times C_i \times A_i \quad (34)$$

onde y_i é a série temporal, T_i é a componente tendência, S_i é a componente sazonal, C_i é a componente ciclo e A_i é a componente aleatória (ou resto).

Cabe ressaltar que nem sempre uma série temporal irá apresentar todos os componentes citados. A decomposição permitirá identificar quais componentes estão atuando naquele conjunto particular, além de possibilitar obter os índices e as equações para realizar previsões. Dentre os componentes apenas a tendência possui a mesma unidade da variável, os demais assumem comumente valores próximos a unidade.

A média móvel é um importante indicador para rastreamento de tendências. Ela suaviza os movimentos dos dados analisados, reduzindo os ruídos causados pelas oscilações existentes. Assim, quanto mais próximas no tempo estão as observações, seus valores estão mais susceptíveis de estarem perto do mesmo valor. A média móvel elimina alguma aleatoriedade nos dados, suavizando a tendência. Com isso torna-se mais simples entender o comportamento global da série analisada.

A Média móvel Simples (MMS) ou aritmética (equação 35) é calculada através do somatório de um conjunto de valores numéricos dividido pela quantidade de elementos somados. A cada novo dado inserido, o primeiro dado da série é descartado, fazendo com que a média se movimente ao longo do tempo.

$$MMS_{t+1} = \frac{\sum_{i=t+1-n}^t x_i}{n} \quad (35)$$

onde MMS_{t+1} é a média móvel simples no período de tempo $t+1$, x_i valores da série a serem somados e n é o número de períodos incluídos em cada média.

A Média Móvel Exponencial (MME) é caracterizada por dar maior peso ao valor mais recente no cálculo da média. Assim, ficam mais evidentes as mudanças nos dados analisados. O cálculo da média móvel exponencial pode ser realizado pela equação 36:

$$MME_{t+1} = (x_i - MME_t) \cdot \left(\frac{2}{n+1} \right) + MME_t \quad (36)$$

onde MME_{t+1} é a média móvel exponencial no período de tempo $t+1$, x_i é o valor atual da série, n é o número de períodos incluídos em cada média e MME_t é a média móvel exponencial no período de tempo t .

Apesar das peculiaridades entre as médias móveis simples e exponenciais, não existe um consenso entre os especialistas em qualificar qual a melhor. A escolha sobre qual usar deve partir das necessidades ou preferência de cada usuário.

3.5.1.3 Normalização dos dados

Uma forma de visualizar padrões em série temporal consiste em análise de gráficos. Contudo, algumas séries não possuem as mesmas unidades. Assim, uma das formas de contornar esse problema está em normalizar os dados, tornando todos os valores das séries compreendidos entre 0 e 1. Dessa forma, podemos visualizar o comportamento geral das séries disponível, sem levar em consideração as unidades, verificando quais séries possuem um comportamento semelhante. As séries podem ser normalizadas pela equação 37:

$$Z_i^N = \frac{Z_i - Z_{\min}}{Z_{\max} - Z_{\min}} \quad (37)$$

onde Z_i^N é o valor normalizado, Z_i é o valor que será normalizado, Z_{\min} é o menor valor da série e Z_{\max} é o maior valor da série.

3.5.1.4 Suavização exponencial

Algumas séries possuem valores com oscilações bruscas que podem ser oriundas de erros de medição ou ruídos que podem prejudicar a interpretação dos dados, sendo desejável corrigir essas deficiências baseadas nos valores da série. Uma das técnicas utilizada para suavizar os movimentos dos dados de uma série histórica é a suavização exponencial, que consiste em um método de análise ou previsão de séries históricas, sendo adequado para os dados de previsão (variáveis dependentes) que não possuam uma tendência ou padrão sazonal bem definidos. As suavizações dos valores da série são calculadas utilizando médias ponderadas por pesos que diminuem exponencialmente. A suavização exponencial simples pode ser estimada pela equação 38:

$$S_i = \alpha \cdot y_{i-1} + (1 - \alpha) \cdot S_{i-1} \quad (38)$$

onde S_i é o valor suavizado exponencialmente e α é a constante ou parâmetro de suavização. Onde o valor de α possui amplitude entre 0 e 1.

3.6 Medidas de desempenho

A falta de conhecimento sobre a eficiência do processo de previsão pode comprometer a utilidade dos resultados. Por essa razão, torna-se imprescindível quantificar o grau de

confiabilidade dos resultados obtidos pelas previsões. Diversos autores tentaram responder essa pergunta, avaliando que elementos devem ser considerados para definir se as previsões são boas ou ruins.

Murphy (1993) definiu esses elementos, identificando três maneiras distintas em que as previsões podem ser boas (más). Estes três tipos podem ser descritos resumidamente da seguinte forma: a) consistência, a previsão é boa se corresponde ao melhor julgamento de quem faz as previsões, baseado em seus conhecimentos, ou seja, leva em conta a correspondência entre as previsões e o julgamento dos resultados; b) qualidade, a previsão é boa se os resultados das previsões são muito semelhantes às condições observadas (ou durante) ao tempo de validade das previsões; e c) utilidade, a previsão é considerada boa quando utilizada por um ou mais usuários como insumo em seus processos de tomada de decisão, resulta em benefícios econômicos incrementais.

Nesse trabalho será dado enfoque a qualidade dos resultados do modelo, por meio de medidas de desempenho, que quantificam o erro entre os valores previstos e observados. Dessa forma, comparam uma série de pares de valores, que para esse trabalho são os pares previstos e observados, e pondera as diferenças. A seguir são apresentadas algumas medidas de desempenho utilizadas na avaliação de modelos de previsão.

O Erro Quadrático Médio (EQM) é o índice que quantifica quadraticamente os erros, fazendo com que as diferenças de maior magnitude influenciem de forma mais significativa. Obter um EQM com valor zero significa que o estimador prevê observações com precisão perfeita, mas na maioria das situações nunca é possível. O EQM pode ser estimado pela equação 39:

$$EQM = \sum_{i=1}^n \frac{(N_{i,calc} - N_{i,obs})^2}{n} \quad (39)$$

O Coeficiente de Persistência (CP) está embasado no Coeficiente de Eficiência Relativa, o qual compara o modelo proposto com o modelo que sistematicamente previria a variável assumindo o valor observado no presente. Diz-se que esta é a previsão sempre disponível, tendo por base a suposição de que a situação presente persistirá (Pedrollo, 2000). O CP varia de $-\infty$ a 1, cujos valores negativos indicam que o modelo proposto é pior que a precipitação persistida (normal climatológica); para $CP = 0$, os modelos são equivalentes e; para valores positivos, o modelo proposto é tanto melhor quanto mais próximo de 1 (Silva, 2005). O coeficiente de Persistência pode ser calculado pelas equações 40, 41 e 42:

$$S_{Persistência} = \sqrt{\frac{\sum(P_{obs} - P_{Pers})^2}{n - 1}} \quad (40)$$

onde P_{obs} é o valor da precipitação observada mensal, P_{Pers} é a precipitação persistida (normal climatológica) e n é o número de meses da série.

$$\sigma = \left(\frac{S_{Persistência} - S_{Modelo}}{S_{Persistência}} \right) * 100 \quad (41) \quad \rightarrow \quad CP = 1 - \frac{\sum(P_{Prev} - P_{Obs})^2}{\sum(P_{Pers} - \bar{P}_{Obs})^2} \quad (42)$$

onde P_{Prev} é a precipitação prevista e \bar{P}_{Obs} é a precipitação média observada.

Uma técnica bastante difundida para verificar quão bem um modelo de regressão linear se ajusta aos dados é através do coeficiente de determinação (r^2). Ele pode ser calculado como o quadrado da correlação entre o valor observado de y_{obs_i} e o valor predito y_{prev_i} (Hyndman & Athanasopoulos, 2013). Que também pode ser calculado pela equação 43:

$$r^2 = \frac{\sum_{i=1}^n (y_{prev_i} - \bar{y}_{obs})^2}{\sum_{i=1}^n (y_{obs_i} - \bar{y}_{obs})^2} \quad (43)$$

Os valores de r^2 situam-se entre 0 e 1. Quanto mais as previsões aproximam-se dos valores reais, mais r^2 aproxima-se de 1.

O Erro Padrão da Regressão (EPR) é uma medida de desempenho bastante eficiente quanto a verificação do desempenho de um modelo de regressão. Também conhecido como “desvio padrão dos resíduos”, mas um pouco diferente do desvio padrão usual. Onde dividimos por $N-j$, porque estimamos j parâmetros no cálculo dos resíduos. Seu cálculo pode ser estimado pela equação 44:

$$EPR = \sqrt{\frac{1}{N - j} \cdot \sum_{i=1}^n e_i^2} \quad (44)$$

O coeficiente de Nash e Sutcliffe (NS) (Nash & Sutcliffe, 1970) que compara a redução do desvio quadrático do erro do modelo, com o desvio quadrático do modelo alternativo de prever sempre a média dos valores. Os valores de NS situam-se entre $-\infty$ e 1. Assim, quanto mais próximo da unidade for o valor de NS (Equação 45), melhor o ajuste do modelo.

$$NS = 1 - \frac{\sum (y_{obs_i} - y_{prev_i})^2}{\sum (y_{obs_i} - \bar{y}_{obs})^2} \quad (45)$$

onde y_{obs_i} é o nível observado, y_{prev_i} é o nível previsto e \bar{y}_{obs} é o nível médio previsto.

As anomalias são desvios em relação à média e sua análise é de grande importância, pois as previsões climáticas são mais confiáveis em valores relativos do que absolutos (Cardoso, 2010). Esse índice de anomalia pode ser estimado com as equações 46 e 47:

$$AO = \frac{NO - NMO_j}{NMO_j} \quad (46)$$

$$AP = \frac{NC - NMC_j}{NMC_j} \quad (47)$$

onde AO é a anomalia observada, AP a anomalia prevista, NC é o nível calculado, NO é o nível observado, NMO_j é a média dos níveis observados no mês j e NMC_j é a média dos níveis previstos no mês j .

As medidas de desempenho que foram citadas anteriormente, embora sejam voltadas para avaliação determinísticas, podem ser usadas na avaliação de previsões probabilísticas. Assim, o conjunto de previsões deve ser representado por uma previsão única, na qual pode ser representada por meio de medidas de tendência central do conjunto como a média, moda ou mediana (Meller, 2012). Contudo, para explorar de forma mais eficiente a metodologia por conjunto, faz-se necessário utilizar medidas de desempenho que avaliem todos os membros do conjunto. Logo, as medidas de desempenho que seguem são comumente utilizadas para medir a eficiência da modelagem por *ensemble*.

O *rank histogram* é utilizado para avaliar se o conjunto de previsões, aparentemente, inclui as observações como sendo um membro equiprovavelmente previsto. Assim, para cada um dos n conjuntos previstos, se os membros e o valor observado foram retirados de uma mesma distribuição, então a classificação do valor observado encontra-se inserido nos N_s+1 valores igualmente prováveis, podendo assumir qualquer um dos i ($i=1, 2, \dots, N_s+1$) valores (Wilks, 2006). A classificação é realizada da seguinte forma:

- Se valor observado for inferior a todos os membros do conjunto N_s , então sua classificação é $i=1$;
- Se o valor observado for superior a todos os membros do conjunto N_s , então sua classificação é $i=N_s+1$;
- Se o valor observado encontrar-se inserido nos intervalos intermediários os membros, então sua classificação encontra-se inserida nesses $1 \leq i \leq N_s+1$ (distribuição de classificação dos membros).

As n classes de verificação são conjuntamente plotadas sob a forma de um histograma. Se a condição de consistência foi cumprida, o histograma será uniforme (Figura 6), refletindo a equiprobabilidade das observações sobre o conjunto.

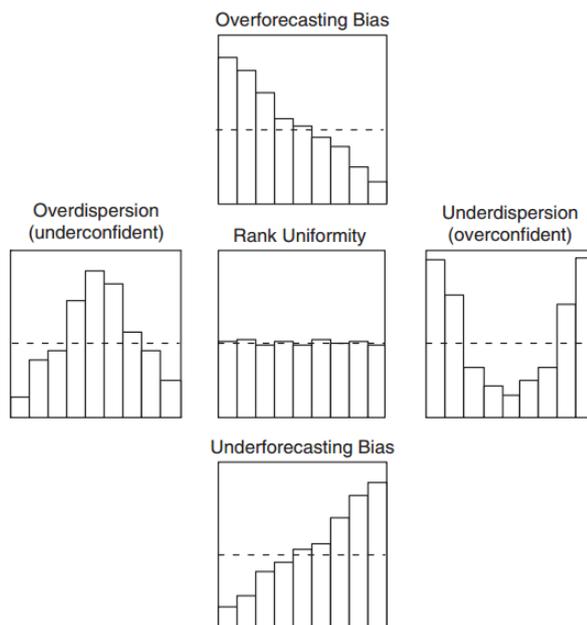


Figura 6 - Exemplos do *rank* histograma para $N_s=8$.

O *Spread-bias diagram* (SBD) descreve a proporção dos dados observados que se enquadram dentro em um determinado segmento da distribuição do conjunto previsto. O eixo Y representa a percentagem de observações que caem dentro ou abaixo desse segmento (ou janela) de previsão. O eixo X representa o tamanho da janela em unidades de probabilidade (a partir de menor previsão). Quanto mais próximo da diagonal, representa que as previsões são confiáveis.

Para as variáveis aleatórias contínuas, o *Spread-bias diagram* (SBD) fornece uma medida simples da confiabilidade das previsões. Trata-se da contagem da fração de observações, SBD (I) (equação 48), que caem dentro de um intervalo de probabilidade de largura fixa sobre o suporte da previsão, $I = (c, d | c, d \in [0,1])$ (EVS, 2015).

$$SBD(I) = \frac{1}{n} \sum_{i=1}^n 1\{\hat{F}_{m_i}(x_i^o) \in I\} \quad (48)$$

onde x_i^o é a observação associada a previsão i , \hat{F}_{m_i} é o quantil da distribuição de probabilidades da previsão por conjunto correspondente à observação e $1\{\cdot\}$ é uma função degrau que assume valor 1 se a condição for atendida e zero caso contrário. A previsão por conjunto é considerada confiável, se no intervalo I da distribuição, a proporção de observações captada for igual a $(d-c)$.

Uma previsão por conjunto é considerada confiável quando é capaz de captar as observações na mesma proporção da amplitude do intervalo da previsão. Isso significa dizer que a observação tem a mesma chance de ocorrer em qualquer intervalo da distribuição

estimada do conjunto (Meller, 2012; Brown et al., 2010). A Figura 7 ilustra os possíveis resultados para o SBD.

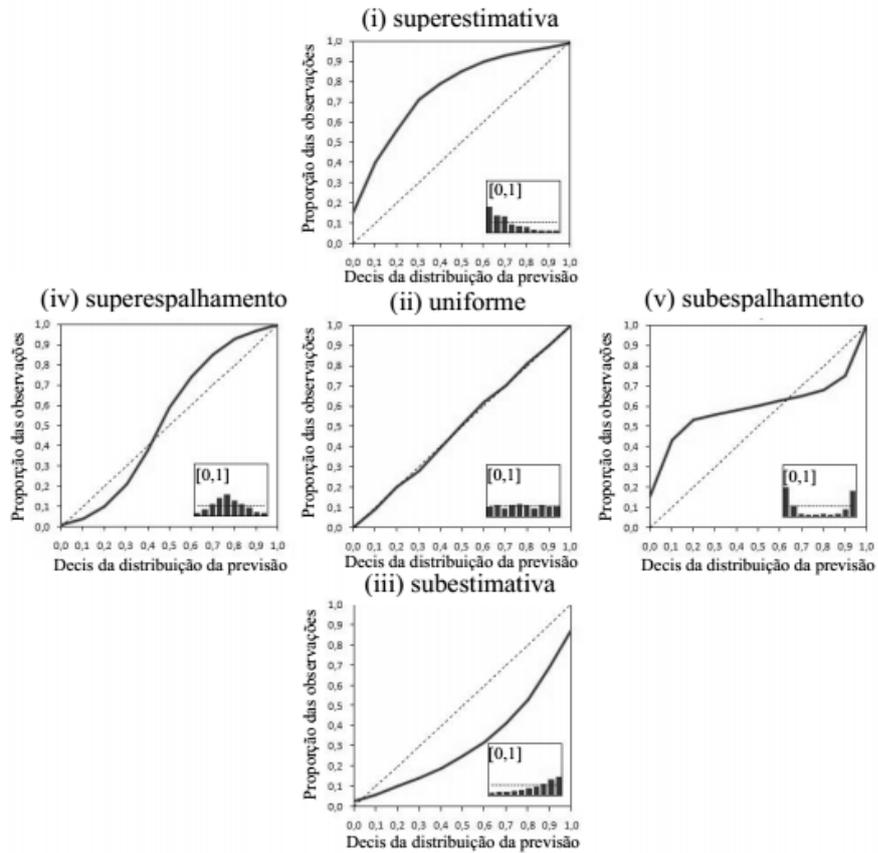


Figura 7 – Exemplos de possíveis resultados da aplicação do SBD. Adaptado de Meller (2012).

4 CORREÇÃO DA PRECIPITAÇÃO PREVISTA

Uma das principais dificuldades encontradas nos estudos hidrológicos consiste em avaliar a propagação dos erros e incertezas da precipitação prevista, na simulação ou previsão de variáveis hidrológicas como vazões em rios ou níveis em reservatórios. Essas dificuldades estão diretamente relacionadas à inconsistência entre a modelagem meteorológica e as respostas hidrológicas (Ferraris et al., 2002). Ao utilizar-se de uma abordagem probabilística, essa mensuração torna-se mais complexa, pois o tamanho do conjunto membros de previsão torna-se um grau de liberdade para a cadeia de previsões (Silvestro & Reborá, 2014).

Ao utilizar-se de simulações advindas de MCG, presume-se que os futuros desvios em relação à série histórica da climatologia do modelo, representam futuros cenários de eventos extremos (períodos de secas ou cheias). Por outro lado, sabe-se que as simulações de MCG afetam diretamente na estimativa de variáveis hidrológicas e na existência de uma tendenciosidade, faz-se necessário uma correção desse viés. Essa retificação procura reduzir a incerteza estrutural do modelo que poderiam levar a avaliações incorretas, em termos qualitativos das respostas do modelo de previsão (Johnson & Sharma, 2015).

Na existência de uma proposta de análise dos dados pautada na metodologia probabilística (*ensemble*), torna-se imprescindível a correção dos dados, haja vista que a presença de erros pode acarretar em falha na interpretação dos dados, tornando os resultados pouco confiáveis. Assim, a probabilidade de ocorrência de um determinado valor da série poderia ser tendenciosa, perdendo todo o valor científico da análise.

Geralmente, os erros das previsões numéricas de tempo (PNT) estão associados, principalmente, a dois fatores: as incertezas relacionadas às condições iniciais e as deficiências na formulação física e dinâmica dos modelos (Palmer, 1993). Por esse motivo, a correção das precipitações previstas a partir de técnicas estatísticas pode compensar, ao menos parcialmente, os erros devido à representação incompleta da dinâmica e dos processos físicos da atmosfera pelo modelo (Silva et al, 2008).

4.1 Detecção de Viés

O Viés (bias) é a média da diferença entre a variável prevista e a observada, apontando se o modelo superestima (viés úmido) ou subestima (viés seco) os valores médios das variáveis analisadas (Neves, 2012; Oliveira, 2014). De uma forma geral, pode ser interpretada

como uma tendência a apresentar ou possuir uma perspectiva parcial em detrimento de outras alternativas. Na estatística, esse termo é comumente usado para expressar erros sistemáticos ou de tendenciosidade. O viés pode ser detectado pela equação 49:

$$Viés = \frac{1}{N} \cdot \sum_{i=1}^N (P_i - P_i^*) \quad (49)$$

onde N é a quantidade de amostras, P é a precipitação calculada pelo modelo de previsão de precipitação e P^* é a precipitação observada.

A correção de viés é algo bastante subjetivo, onde sua correção depende da precisão desejada, porém Ávila et al (2009), afirma que a aplicação da correção do viés é indicada nos casos em que o coeficiente RV (equação 50), que corresponde a relação entre o valor absoluto do viés e o valor do desvio absoluto médio do erro em relação ao viés (equação 51), apresentar valor maior do que 1.

$$DViés = \frac{1}{N} \cdot \sum_{i=1}^N |P_i - P_i^* - Viés| \quad (50)$$

$$RV = \frac{|Viés|}{|DViés|} \quad (51)$$

A raiz quadrada do erro quadrático médio (REQM) fornece uma medida de precisão (acurácia) do modelo, permitindo apurar o quanto o valor simulado é próximo do valor verdadeiro. Como o erro médio é elevado ao quadrado, os maiores erros são realçados de modo que quanto maior REQM, menor a precisão do valor simulado (Mendonça, 1999; SANTOS, 2014; Silva et al, 2008; Oliveira, 2014; Batista, 2012). Esse coeficiente também é utilizado para avaliar a necessidade de correção do viés, contudo, não existe um consenso quanto a um valor ideal para indicar os casos onde a correção faz-se necessário, ficando por conta do usuário fixar um valor aceitável que atenda suas exigências. A REQM pode ser estimada pela equação 52:

$$REQM = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (P_i - P_i^*)^2} \quad (52)$$

Cabe ressaltar que os métodos utilizados pelos trabalhos citados nesse tópico utilizaram simulações e não previsões. Contudo, a metodologia será testada como uma primeira estimativa para avaliar a necessidade de correção nos dados.

4.2 Remoção do viés

Existem diversas técnicas de remoção do viés e uma bastante utilizada entre os hidrólogos é a técnica denominada Mapeamento quantil-quantil, que é baseada na comparação das funções cumulativas de probabilidade da variável observada e da variável estimada por um modelo climático (Fernandes, 2012; Silveira et al, 2014; Silveira et al, 2014).

A aplicação do método segue o procedimento a seguir:

1. Ajusta-se uma função cumulativa de probabilidade para os dados observados;
2. Ajusta-se a mesma função cumulativa de probabilidade do passo 2 para a variável estimada por um modelo climático;
3. Para cada valor mensal da série da variável analisada, é identificado o valor da probabilidade acumulada na série do modelo climático. Posteriormente, o valor mensal corrigido é igual ao valor observado que apresenta o mesmo valor de probabilidade acumulada, a Figura 8 ilustra o procedimento para remoção.

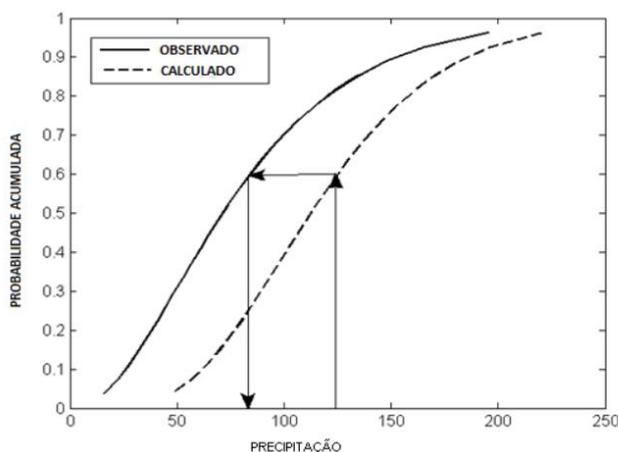


Figura 8 - Mapeamento das probabilidades para correção do viés. Adaptado de Fernandes (2012).

O ajuste da função cumulativa de probabilidade pode ser realizado mediante uma função conhecida ou realizando um teste de aderência para verificar a adequação de uma função cumulativa de probabilidade.

4.2.1 Função cumulativa de probabilidade

A função cumulativa de probabilidade (FCP) é uma forma de descrever como as probabilidades são associadas aos valores ou aos intervalos de valores de uma variável aleatória. Uma FCP de uma variável aleatória X é uma função que fornece a probabilidade de

não superação do argumento x , ou seja, mostram a probabilidade da variável ser menor ou igual a um determinado valor (equação 53).

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_x(x) dx \quad (53)$$

O conhecimento da FCP é suficiente para fazer inferências sobre o comportamento de uma variável aleatória. Existem diversas FCP e abaixo seguem algumas das mais utilizadas em estudos hidrológicos, segundo a descrição de Naghettini & Pinto (2007).

A distribuição Normal é utilizada para descrever o comportamento de uma variável aleatória que esboça uma forma simétrica em torno de um valor central. Constitui um modelo de dois parâmetros, cuja função cumulativa de probabilidade é expressa pela equação 54:

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\theta_2^2}} \exp \left[-\frac{1}{2} \left(\frac{x - \theta_1}{\theta_2} \right)^2 \right] dx \quad (54)$$

onde θ_1 e θ_2 são parâmetros da função.

A FCP da distribuição Normal, dada pela equação 54, não possui solução analítica. Assim, para contornar esse problema, pode-se realizar uma transformação linear pela equação 55, da variável Normal X de parâmetros μ e σ .

$$Z = \frac{X - \mu}{\sigma} \quad (55)$$

Onde Z é a variável Normal central reduzida, μ é a média e σ é desvio padrão.

A FCP de Z é dada pela equação 56:

$$F_Z(z) = \Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{z^2}{2} \right] dz \quad (56)$$

A FCP da distribuição normal padrão $\Phi(z)$ pode ser obtida por integração numérica, onde os resultados da integração encontram-se na tabela A1 do Anexo A.

O cálculo da FCP GAMA pode ser aproximado pela *distribuição Normal padrão* $\Phi(u)$, calculada no ponto u , definido pela equação 57:

$$u = 3\sqrt{\eta} \left(\sqrt[3]{\frac{\xi}{\eta}} - 1 + \frac{1}{9\eta} \right) \quad (57)$$

onde ξ (equação 58) é a variável Gama padrão, η e θ são parâmetros de forma e de escala, que podem ser estimados pelas equações 59 e 60.

$$\xi = \frac{x}{\theta} \quad (58)$$

$$E[x] = \eta\theta \quad (59)$$

$$Var[x] = \eta\theta^2 \quad (60)$$

A FCP de excedência de Weibull pode ser calculada pela equação 61.

$$q_i = \frac{i}{n+1} \quad (61)$$

onde q é a probabilidade de excedência, i é a posição de plotagem (ou ordem) e n é o número total de dados.

4.2.2 Teste de aderência

Quando um teste de hipótese refere-se à forma de uma função cumulativa de probabilidade, ou seja, do modelo distributivo da população de onde a amostra foi extraída, são chamados testes de aderência. Esses testes constituem em uma importante ferramenta para verificar uma eventual adequação entre as probabilidades ou frequências de um modelo hipotético, a determinados valores de uma amostra de dados observados. Um teste de aderência não paramétrico bastante difundido entre os hidrólogos, utilizado para séries contínuas, é o teste de Kolmogorov-Smirnov (KS). Este método está baseado na diferença máxima entre as funções cumulativas de probabilidade empíricas e teóricas, de variáveis aleatórias contínuas (Naghetini & Pinto, 2007).

Classifica-se uma amostra $\{X_1, X_2, \dots, X_N\}$ em ordem crescente e calcula-se a distribuição empírica para cada elemento pela equação 62. O próximo passo consiste em calcular as probabilidades teóricas $F_X(X_i)$ com a distribuição que se deseja testar a adequação a série. A estatística do teste KS é dada pela equação 63:

$$F_N(X_i) = \frac{i}{N} \quad (62)$$

$$D_N = \sup_{-\infty < X < \infty} |F_N(X_i) - F_X(X_i)| \quad (63)$$

O valor de D_N corresponde a maior diferença entre as probabilidades empíricas e teóricas. Para amostras com tamanho entre $10 < N < 40$ deve-se utilizar os valores críticos da tabela A2 que encontra-se no Anexo A. Para valores superiores pode-se encontrar na mesma tabela equações para estiver o valor crítico para um determinado nível de significância. Para $D_N^{calc} < D_{N,\alpha}$ a decisão é a de não rejeitar a hipótese (H_0) de que a série pode ser modelada pela distribuição testada.

5 ESTUDO DE CASO

5.1 Características gerais

A metodologia proposta foi aplicada no Sistema Hidrológico do Taim (SHT), que constitui um complexo composto de banhado e lagoas interconectadas. Segundo Villanueva (1997), o SHT é formado por diferentes subsistemas: (a) Subsistema Norte (387 km²), composto pela Lagoa Caiubá, pela Lagoa das Flores e pelo Banhado do Maçarico, onde este último sofre afluência para a Lagoa das Flores; (b) Subsistema do Banhado (270 km²) é o Banhado propriamente dito, também composto pelas lagoas Nicola e Jacaré. O escoamento é caracterizado pelas baixas velocidades superficiais, devido à ocorrência de macrófitas aquáticas, e a saída d'água acontece pelas comportas no extremo norte, em direção à Lagoa Mirim; (c) Subsistema Sul (1.597 km²) é formado pela Lagoa Mangueira e por sua bacia de contribuição. A união do Subsistema Sul com o Banhado é realizada através do canal junto a BR-471 por escoamento concentrado e a interface lagoa-banhado por escoamento difuso.

O Sistema Hidrológico do Taim (SHT) possui área total com aproximadamente 2254 km², localizada no Sul do Rio Grande do Sul, compreendendo partes dos Municípios de Santa Vitória do Palmar e de Rio Grande entre as latitudes 32°20'S e 33°00' S, e pela Lagoa Mirim e o Oceano Atlântico sul entre as longitudes 52°20'W e 52°45'W (Cavalcanti, 2013).

Devido ao crescente processo de expansão das lavouras de arroz, percebeu-se que a fauna e a flora do Banhado do Taim encontravam-se sob ameaça, e no dia 21 de julho de 1986, o Banhado do Taim foi promovido a Unidade de Conservação Federal pelo Decreto de nº 92.963, onde foi criada a Estação Ecológica do Taim (ESEC-Taim) (Tassi, 2008). A área delimitada pelo ESEC-Taim (Figura 9) não abrange a totalidade da sua bacia de contribuição, dessa forma, as atividades praticadas fora dos limites da Estação (como por exemplo, a retirada de água para a irrigação do arroz) influenciam diretamente na disponibilidade de água e nas condições ambientais no seu interior (Marques et al., 2001). As saídas de água que ocorrem no sistema são dadas basicamente por evaporação, demanda para irrigação e, quando o nível está alto, escoamento para a Lagoa Mirim por meio de uma comporta (Paz, 2003; Villanueva, 1997).

A lagoa Mangueira (Figura 10), que constitui a principal fonte de água para o banhado, possui uma área superficial de 820 km² com profundidade média de 2,6 metros, e

máxima de 6,5 metros. A lagoa é cercada por uma variedade de habitats, como praias, dunas, florestas e banhados.

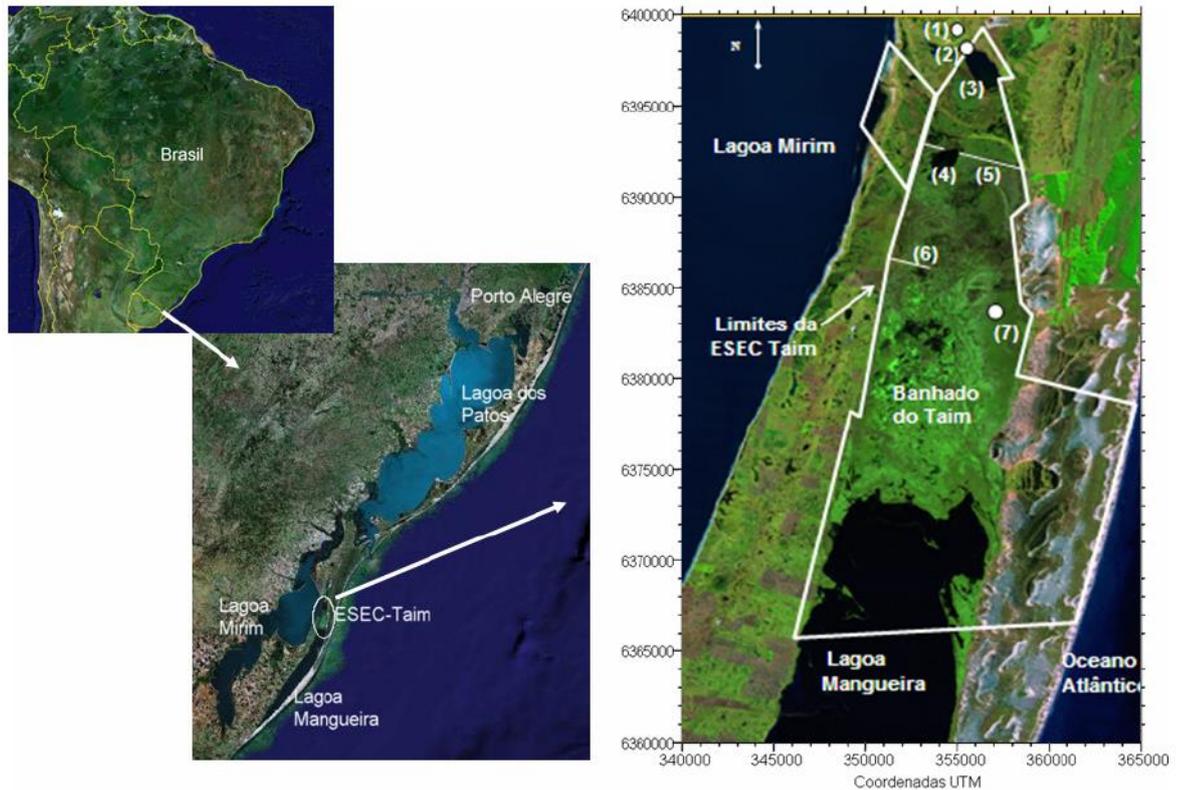


Figura 9 - Delimitação do banhado do Taim. Adaptado de Cardoso (2010).

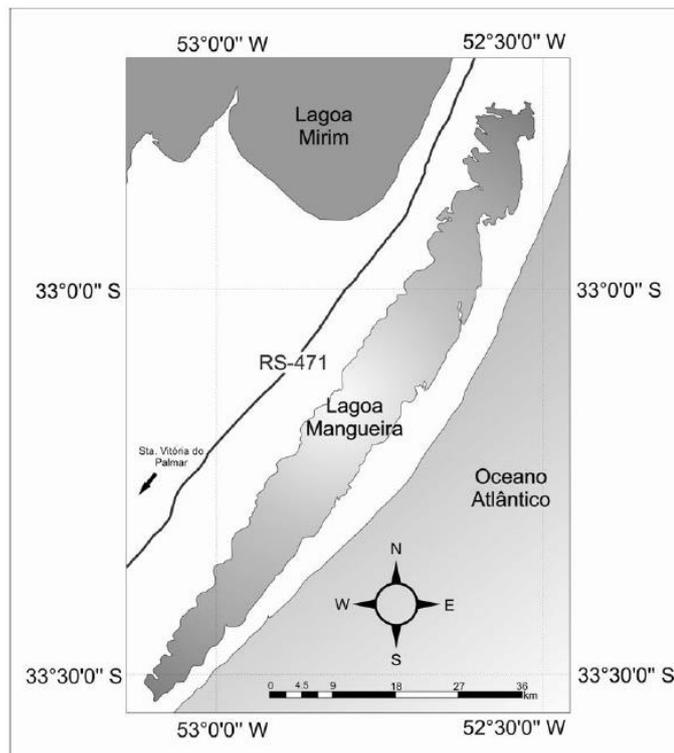


Figura 10 – Delimitação da Lagoa Mangueira. Adaptado de Cavalcante (2013).

5.2 Clima e precipitação

O SHT encontra-se inserido na região sul do Brasil, onde o clima é do tipo subtropical, e os mecanismos responsáveis pela precipitação na região estão relacionados com os sistemas dinâmicos transientes e as configurações orográficas. A região possui uma temperatura média anual de 18° C, com precipitação média anual de 1100 mm, apresentando um inverno frio e chuvoso e um verão quente e seco, com ocorrência de ventos bastante intensos (Paz, 2003).

A precipitação na região sul do Brasil, em geral, é bem distribuída ao longo do ano, fazendo com que a região não possua uma estação chuvosa ou seca bem definida. Cardoso (2010) afirma que o clima exerce uma forte influência na disponibilidade de água e no hidroperíodo do SHT, existindo uma relação entre a precipitação e os fenômenos El Niño e La Niña, e a Oscilação Decadal do Pacífico (ODP).

5.3 Uso do solo, geologia e relevo

Ao longo do sistema e em seu entorno, existem diferentes tipos de coberturas de solo, onde pode ser identificado dentre eles: solo agrícola, pastagens, reflorestamento, dunas e áreas baixas, alagadas permanente ou temporariamente.

A formação geológica onde se encontra o SHT se deu em função dos movimentos de recuo do mar, no qual originou o litoral do Rio Grande do Sul e que resultaram no maior complexo lagunar da América do Sul. O que hoje constitui o banhado, campos e matas palustres dentro do SHT, constituía um antigo canal ligando a Lagoa Mirim ao Oceano Atlântico (Paz, 2003). A região é caracterizada por terras baixas, onde os relevos mais acentuados formados por dunas litorâneas que margeiam o oceano Atlântico e por falésias marginais próximas a lagoa Mirim, cuja estrutura geológica é gerada pela deposição de sedimentos de origem marinha, lacustre e eólica (Tassi, 2008).

5.4 Biodiversidade

A diversidade biológica dentro do SHT é considerada bastante ampla, onde vários estudos visaram quantificar e detalhar as flutuações temporais e espaciais de diversas comunidades (plâncton, macrófitas aquáticas, invertebrados, peixes, répteis, anfíbios, aves, mamíferos, etc.) que fazem desse ecossistema seu habitat (Fragoso, 2009).

O regime hidrológico pode modificar as propriedades físicas e químicas, tais como disponibilidade de nutrientes, grau de anóxia do substrato, salinidade do solo, propriedades do sedimento e pH. Assim, as condições hidrológicas em banhados mudam, mesmo que ligeiramente, podendo a biota responder com mudanças acentuadas, tanto na riqueza de espécies como na produtividade do ecossistema. O banhado do Taim possui, em sua maioria, a área coberta por macrófitas emergentes e flutuantes, o que pressupõe uma grande quantidade de biomassa funcionando como um gerador de matéria orgânica (PELD, 2013).

As oscilações dos níveis e a disponibilidade de água influenciam de forma direta no habitat de banhados, pois algumas espécies existem (ou habitam) em uma determinada região em função do volume d'água, da quantidade e da natureza do sedimento que são deslocados (em processos naturais, tais como escoamento) e da natureza e composição dos materiais que formam as margens do reservatório. A diminuição no transporte de sedimentos e nutrientes causa impacto nas propriedades dos reservatórios, causando perda de habitat para diversas espécies. No trabalho de Tassi (2008), foram delimitados alguns níveis de referência para um gerenciamento, baseado em indicadores biológicos, no banhado do Taim.

5.5 Conflitos pelo uso da água

Segundo Villanueva et al. (2000), o principal vetor de pressão sobre a biodiversidade da SHT está associado à orizicultura, considerada a principal atividade econômica da região, a qual exige a manutenção de lâminas d'água sobre os cultivos. A água usada para a irrigação da orizicultura é extraída diretamente da Lagoa Mangueira, afetando diretamente as trocas de água entre a Lagoa e o Banhado do Taim. A referida captação coincide com os picos de evapotranspiração, que são observados nos meses de Outubro a Março, onde a combinação das duas fontes de retirada de água pode ocasionar a diminuição no volume de água necessário à manutenção dos ecossistemas palustres (Villanueva, 1997).

Observam-se interesses conflitantes na definição quanto aos níveis de funcionamento do sistema, onde, para a irrigação, o fechamento da estrutura de saída, garantiria altos níveis de água durante todo o período, assim retendo volumes que escoariam para a lagoa Mirim. No entanto, faz-se necessário que ocorram flutuações dos níveis, para a conservação do sistema. As mudanças nos níveis d'água, sejam causadas por eventos naturais ou pelo uso não planejado dos recursos hídricos do sistema, podem alterar de forma significativa as funções desempenhadas pelo banhado, como os padrões de comunidades, as riquezas de espécies, a produtividade e o habitat (Marques et al., 2002).

Outra atividade implantada na região nos anos de pousio do plantio de arroz é a pecuária, gerando uma fonte difusa de poluentes para o sistema. Outras alterações civis no SHT influenciaram em um padrão diferente para a circulação das águas do sistema, como a construção da BR 407, obras de drenagem e a construção da barragem eclusa do Canal de São Gonçalo (Fragoso, 2009).

Para este estudo o interesse hídrico concentrou-se no banhado do Taim, analisando a utilização de previsões hidrometeorológicas de longo prazo na estimativa dos níveis futuros no banhado. Contudo, a metodologia proposta, pode ser aplicada a qualquer ecossistema de terras úmidas onde se pretende realizar estudos hidrológicos.

5.6 Dados hidrológicos

Uma vantagem na utilização do Banhado do Taim como área de estudo, está no fato dele estar sendo monitorado desde 1998, quando inserido pelo CNPq no programa de PELD – Sítio 7. O Programa de Pesquisa Ecológica de Longa Duração (PELD) consiste em uma rede de sítios de referência para a pesquisa científica no tema de Ecologia de Ecossistemas. É executado pelo CNPq, e sua estrutura de gestão é estabelecida pela resolução normativa nº 23/2011, que define os comitês gestor e científico do programa e as suas atribuições (CNPq, 2014).

Nesse contexto o projeto PELD disponibiliza diferentes variáveis hidrológicas, meteorológicas, físicas, químicas e biológicas que vêm sendo monitoradas ao longo dos últimos anos. Dentre estas variáveis incluem: níveis da Lagoa Mangueira e no Banhado, precipitação, evaporação, temperatura da água, radiação solar incidente na superfície da água, velocidade e direção do vento.

Nos sub-tópicos que se seguem, encontram-se uma descrição dos dados disponíveis para essa pesquisa: fonte e tratamento realizado aos dados.

5.6.1 Dados de precipitação observada

Os dados de precipitação utilizados nessa pesquisa foram cedidos pelo Instituto de Pesquisas Hidráulicas da Universidade Federal do Rio Grande do Sul (IPH-UFRGS) e compreendem os anos de 1960 a 2011, onde as séries de dados foram obtidas pelo Sistema de Informações Hidrológicas da Agência Nacional de Águas, o 8º Distrito de Meteorologia (INMET) e a ESEC-TAIM (Tabela 1).

Tabela 1 - Estações pluviométricas.

Estação	Orgão	Código	Coordenadas
Santa Vitória do Palmar	8º Distrito de Meteorologia		33°31'S e 53°21'W
Rio Grande	Furg		32°04'S e 52°10'W
ESEC-TAIM	Ibama		
Granja Santa Maria	ANA	(3252008)	32°36'S e 52°36'W
Rio Grande	ANA	(3252020)	32°02'S e 52°06'W
Granja Osório	ANA	(3253003)	32°57'S e 53°09'W

Para a determinação da média mensal foi utilizado o método empregado por Villanueva (1996), onde os pesos adotados foram de 50% para o posto correspondente a porção média, 25% para os postos correspondentes a região norte e 25% a região sul. A determinação da porção média foi realizada utilizando 3 postos identificados pelos códigos 3252008 (correspondente à região média da área do Taim), 3253003 (região sul) e 3252020 (região norte).

5.6.2 Dados de previsão de precipitação e climatologia do modelo de circulação global

Os dados de previsão de precipitação foram disponibilizados pelo *International Research Institute for Climate and Society* (IRI). Para essa pesquisa foram utilizados dados simulados pelo modelo de circulação global COLA C2.2, o qual possui um conjunto de 30 membros para a forçante do Pacífico tropical SSST (cenário de anomalia da TSM), com resolução de 1,875 x 1,864678 graus e horizonte de 7 meses com frequência mensal.

A série adquirida possui uma disponibilidade temporal de junho de 2004 a outubro de 2015, porém a série de níveis possui disponibilidade de 1960 a 2011. Assim, o período utilizado para realizar as simulações e análises com os dados de precipitação prevista ficou compreendido entre junho de 2004 a dezembro de 2011. Como a frequência é mensal, a primeira rodada do modelo foi realizada em maio de 2004, sendo este mês o ponto de partida para as análises.

Foi adquirida a série histórica da climatologia do modelo, que possui uma frequência mensal e 10 membros, onde os dados disponíveis possuem disponibilidade entre 1950 a 2002. Devido ao período disponível da série dados observados de precipitação (1960 a 2011), foi

utilizado o período compreendido entre 1960 a 2002, pois é o intervalo em comum entre os dados.

5.6.3 Dados de níveis de água, evaporação e retiradas para irrigação

Conforme mencionado por Paz (2003), apesar de existência de algumas régua limnimétricas e linígrafos, poucos dados de níveis d'água estavam disponíveis e em condições de serem aproveitados nesta pesquisa, por vários fatores como: incerteza quanto à cota de referência; interrupção na operação e leitura dos dados. Tassi (2008), explica que, em estudos anteriores, houve a necessidade de obtenção de uma série de longo período para tentar caracterizar o regime hidrológico local. Na ocasião, foi utilizado o modelo hidrodinâmico de células elaborado por Villanueva (1997), onde a calibração foi realizada utilizando o monitoramento do período de setembro de 1995 a agosto de 1996. O modelo encontrou resultados satisfatórios, onde após a calibração a série de níveis no banhado foi estendida desde janeiro de 1960 até dezembro de 1996. Atualmente, essa série foi estendida utilizando o mesmo modelo até a data de dezembro de 2011.

Para esse trabalho foi adotada a série de níveis cedida pelo IPH-UFRGS e compreendem os anos de 1960 a 2011, advinda do modelo de Villanueva (1997), pois está constitui a única fonte disponível de dados, considerada confiável, para realização desse estudo.

Os dados de evaporação e retirada para irrigação foram fornecidos pelo IPH-UFRGS e projeto PELD, com disponibilidade temporal entre 1960 a 2011.

5.7 Estudos antecedentes utilizando modelos matemáticos

O modelo proposto por Villanueva (1997) simula a lagoa mangueira (que constitui a principal fonte de água do banhado) e o banhado propriamente dito até sua conexão com a lagoa Mirim nas comportas de saída do banhado. O modelo possui dois módulos de processamento: a) módulo da Mangueira, composto por um modelo hidrológico que simula a transformação chuva-vazão e uma equação de balanço da lagoa; b) módulo do banhado, que simula o escoamento no banhado e no canal lateral, recebendo influência da lagoa Mirim através das comportas de saída do banhado.

O módulo Mangueira utiliza um modelo de balanço de volumes representado pela equação 64:

$$S(t + \Delta t) = S(t) + A_b C_e P + A_M P - K_t E A_M - I_r - Q_T \Delta t \quad (64)$$

onde S é o volume na lagoa Mangueira (m^3), A_b é a área de contribuição à Mangueira (m^2), C_e é o coeficiente de escoamento, P é a precipitação no intervalo de tempo considerado (m), A_M é a área da lagoa Mangueira (m^2), K_t é o coeficiente da evaporação de tanque, E é a evaporação no intervalo de tempo, I_r é o volume retirado para irrigação (m^3), Q_T é a vazão que sai da Mangueira para o Taim (m^3/s).

O modelo do Taim (Villanueva, 1997) utiliza um modelo hidrodinâmico de células para representar o banhado, usando 13 seções no canal e 28 células no banhado. O comprimento dos trechos é da ordem de 3 km, e a área das células de aproximadamente 9 km^2 .

Esse modelo foi usado por Cardoso (2010) para previsão hidroclimática do sistema hidrológico do Taim, para calcular cenários de níveis baseados na previsão de precipitação. Dessa forma, foram realizadas quatro tipos de previsão: a) previsão baseada em chuva nula ou chuva zero; b) previsão baseada na chuva observada; c) previsão baseada na chuva prevista; d) previsão baseada na normal climatológica.

Cardoso (2010) realizou inicialmente previsões de nível baseada na chuva observada, que serviram de base para comparação da qualidade das outras previsões, uma vez que, operacionalmente, não poderiam ser obtidas em tempo real. Depois, realizou previsões para um horizonte de até seis meses, onde a partir da análise dos resíduos obtidos das simulações e dos índices obtidos das medidas de desempenho, pode concluir que os resultados foram promissores e que a utilização das previsões meteorológicas pode ser uma ferramenta importante para gestão dos recursos hídricos. Para as previsões, Cardoso (2010) utilizou três modelos de previsão de precipitação: a) modelo de circulação global CPTEC/COLA; b) modelo regional MRERS; c) modelo de regressão linear múltipla baseado em índices climáticos.

O Modelo de circulação global MCGA CPTEC/COLA está baseado no código do modelo utilizado pelo *Center for Ocean-Land-Atmosphere Studies* (COLA) dos EUA. Os dados do modelo são disponibilizados em uma grade de pontos distanciados de 1,875 graus em longitude e 1,875 graus em latitude. O modelo possui cinco membros, com frequência mensal e horizonte de seis meses. O modelo regional MRERS pertencente ao Centro de Pesquisas e Previsões Meteorológicas (CPPMet) da Universidade Federal de Pelotas (UFPEL), com resolução de $1^\circ \times 1^\circ$. O modelo de simulações de previsões do modelo de regressão linear múltipla está baseado em índices climáticos (ModReg) (Cardoso, 2010).

As previsões de níveis foram realizadas para um horizonte de até seis meses de antecedência, para o período de outubro de 1997 a dezembro de 2003, sendo realizadas sempre no mês de setembro.

Tassi (2008) utilizou o mesmo modelo (Villanueva, 1997) para um horizonte de 8 meses: iniciando em Agosto e finalizando em Março, onde o ponto de partida para a previsão residia no fato do mesmo marcar a época de início de preparo do solo para o plantio de arroz irrigado. Esta previsão possuía um enfoque na identificação de uma condição crítica para manutenção do ecossistema, onde seria possível alertar aos orizicultores que poderia haver restrições com relação à disponibilidade de água. Para o referido trabalho, foi realizada uma previsão hidrológica com o uso da técnica de modelagem por conjunto (*ensemble forecast*), utilizando dados de previsão de precipitação advindos de modelos climáticos.

Com os resultados obtidos, Tassi (2008) observou que, embora a média da previsão por conjunto para o período simulado não tenha sido satisfatória, alguns membros do conjunto da previsão mostraram um cenário de seca extrema, onde as cotas do nível da água sempre foram menores que as previstas para o cenário sem irrigação.

Tassi (2008) menciona que, para esse estudo, utilizou previsões de níveis no banhado do Taim com os modelos autorregressivos (AR), modelos de média móvel (*Moving Average - MA*), e autoregressivos com média móvel (*Auto Regressive Moving Average - ARMA*). Para servir como condição de contorno de jusante, durante a simulação, foram realizadas previsões na lagoa Mirim utilizando modelos estatísticos do tipo AR, MA, ARMA e SARIMA, sendo que o último modelo forneceu os melhores resultados. Contudo, similarmente ao ocorrido nas previsões realizadas para o banhado, o modelo não conseguiu representar bem algumas ocasiões de eventos extremos de cheia e estiagem.

Scuderi (2010) aplicou um modelo simplificado de balanço de volume, que utiliza a taxa de evapotranspiração (estimada em seu trabalho). Esse modelo representou, de forma aproximada, o comportamento mensal do nível de água no banhado. As simulações foram realizadas utilizando programação em linguagem fortran, com um período de análise desde janeiro de 1960 até julho de 2006. Para a realização das simulações o modelo utilizou as curvas cota-área e cota-volume, que foram obtidas a partir de levantamento batimétrico e das equações utilizadas por Villanueva (1997).

O modelo utilizado por Scuderi (2010) é constituído por dois módulos, semelhante ao modelo utilizado por Villanueva (1997): o módulo da lagoa Mangueira e outro módulo para o banhado. Quando os resultados do modelo foram comparados ao longo do período de simulação, verificou-se que o modelo de balanço de volume proposto não podia reproduzir os

valores mínimos. O autor concluiu que os resultados do modelo simplificado apresentam um comportamento semelhante ao modelo hidrodinâmico de Villanueva (1997), no seguimento da tendência simulada. O módulo Mangueira utilizado no modelo simplificado de balanço de volumes foi simulado pela equação 65:

$$\Delta V_{MANG} = (P - E) \cdot A_{SMANG} + P \cdot C_e \cdot A_{BACIA} - I \pm Qt_{MAT} \quad (65)$$

onde ΔV_{MANG} é a variação de volume na lagoa no período de tempo analisado em m³, P é a precipitação ocorrida sobre a bacia em m, E é a taxa de evaporação da lagoa em m, A_{SMANG} é a área superficial da lagoa, que é função do nível d'água em m², C_e é o coeficiente de escoamento da bacia de contribuição à lagoa (adimensional), A_{BACIA} é a área da bacia que contribui à lagoa em m², I é o volume da água para irrigação de arroz extraído da lagoa em m³ e Qt_{MAT} é a vazão de transferência entre a lagoa e o banhado no período de tempo analisado em m³. O autor afirma que a direção do fluxo é da lagoa para o banhado, na maioria dos casos.

A transferência de vazão entre a lagoa Mangueira e o Banhado do Taim foi calculada considerando o problema de desnível, o qual havia sido explicado no trabalho de Villanueva (1997), onde o autor relata um desnível da ordem de 50 cm que não deveria existir pela continuidade física dos corpos d'água, concluindo que o erro pode ser devido a problemas na determinação dos zeros das régua.

As simulações realizadas no banhado obtiveram bons resultados, onde o modelo elaborado por Villanueva (1997), obteve resultados bastante satisfatório, porém as previsões de níveis não obtiveram os mesmos desempenhos. Entretanto, os resultados das previsões revelaram ser uma promissora ferramenta de gestão do banhado. Os trabalhos que realizaram previsões no banhado, utilizaram um modelo hidrodinâmico de base física, composto por complexas equações que simulam o banhado e a lagoa Mangueira. Apenas Tassi (2008), realizou tentativas utilizando modelos empíricos autoregressivos, onde a autora relata que as previsões realizadas não conseguiu representar bem algumas ocasiões de eventos extremos de cheia e estiagem. Motivado pela falta de trabalho utilizando uma abordagem empírica para as previsões no banhado, esse trabalho propõe a elaboração de um modelo utilizando apenas uma abordagem empírica para responder a pergunta: Um modelo de equações simples, utilizando uma abordagem empírica é capaz de obter resultados tão bons como um modelo complexo de base física?

6 METODOLOGIA

Este capítulo descreve a aplicação da metodologia utilizada para obtenção do modelo hidrológico elaborado para simular o Sistema Hidrológico do Taim. A estrutura do modelo descrito aqui foi baseada em formulações matemáticas, utilizando a metodologia empírica de simulação de dados. Nesse item, são abordadas as etapas de análises, simplificações e critérios adotados em cada processo da elaboração das regressões.

Estudos anteriores já realizaram previsões de níveis no banhado utilizando uma abordagem física. Contudo, esse estudo parte de uma metodologia completamente empírica, utilizando o princípio da parcimônia.

As equações propostas nesse estudo são recomendadas apenas para utilização no banhado do Taim, pois, como a equação não utiliza conceitos físicos dos fenômenos envolvidos, seu uso fica restrito a área em que foi elaborada. Contudo a metodologia proposta pode ser usada para estimar equações para outras áreas.

A metodologia está dividida em três vertentes:

1. Análise preliminar das séries temporais;
2. Estimativa de equações de regressão;
3. Verificação de correção do viés de precipitação;
4. Previsão.

Para a estimativa das equações foram testadas as metodologias de regressão simples e múltipla. De uma forma geral, as etapas realizadas para elaboração das regressões foram as seguintes:

1. Seleção dos dados - Seleciona os dados que comporão as equações de regressão;
2. Estimativa dos parâmetros - São estimados os parâmetros das equações;
3. Avaliação do desempenho - As equações são utilizadas para estimar a variável resposta no modo simulação (intervalo de formulação) e com base nos resultados são calculadas as medidas de desempenho, onde as que obtiverem baixo desempenho serão descartadas;
4. Análise de sensibilidade - É avaliada a influência das variáveis que compõem cada equação para a variável resposta. As equações que não atenderem aos critérios pré-definidos serão rejeitadas;

Nos passos 3 e 4 foram adotados alguns critérios para rejeição das equações conforme apresentado no tópico 6.2. A Figura 11 ilustra os passos utilizados para a elaboração das equações de regressão.

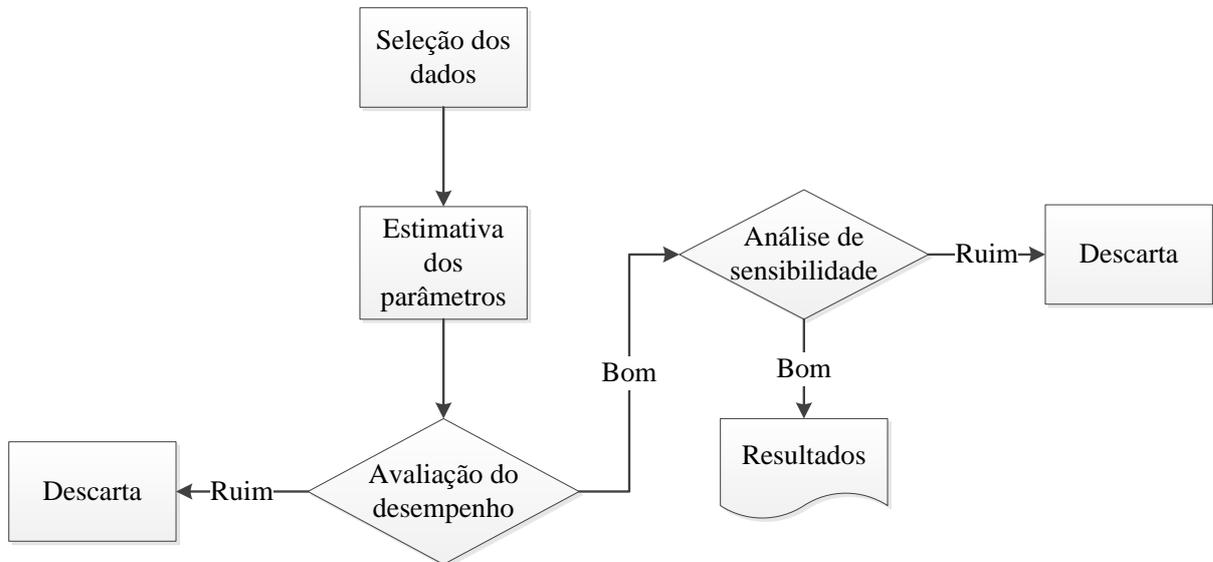


Figura 11 - Fluxograma da metodologia para elaboração das equações de regressão.

Para reduzir as incertezas que envolvem as previsões, nesse estudo foi utilizada a técnica de previsão por conjunto (ensemble), para assim agregar as informações probabilísticas aos resultados. Foram utilizadas previsões de precipitação simuladas por um modelo de circulação global. Devida à importância da precipitação para esse estudo, foi avaliada a necessidade de correção aos dados de precipitação prevista. Os comentários e critérios utilizados na avaliação e correção aos dados encontram-se no tópico 6.3. A Figura 12 ilustra os passos seguidos para as análises e correção nos dados de precipitação prevista.

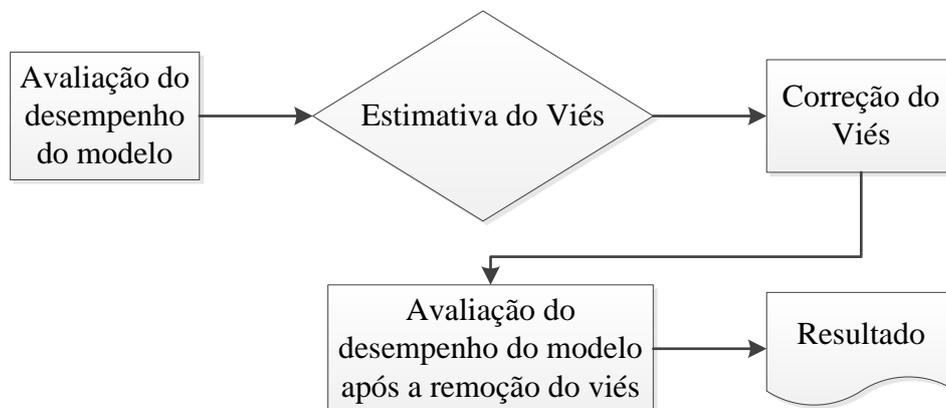


Figura 12 - Fluxograma para avaliação e correção do Viés nos dados da precipitação prevista pelo modelo de circulação global.

Uma vez com o modelo ajustado, e feita a verificação da remoção de viés, o modelo é rodado em modo previsão seguindo o procedimento apresentado na Figura 13. As previsões

foram realizadas em duas etapas: Utilizando chuva perfeita e chuva prevista. No item 6.5 apresentam-se maiores detalhes sobre o procedimento utilizado na realização das previsões.

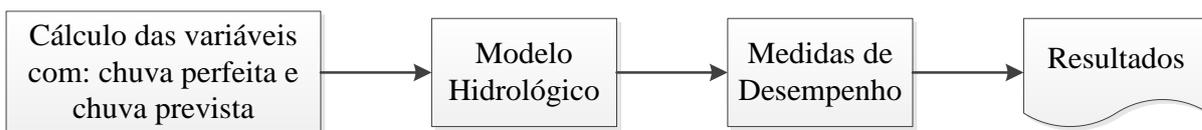


Figura 13 – Fluxograma da realização das previsões.

Para a utilização da proposta metodológica, foram realizadas análises preliminares nas séries históricas, onde foram utilizadas algumas ferramentas de cálculo estatísticos, que permitiram transformar e combinar as séries históricas dos dados hidrológicos, disponibilizados para esse estudo.

6.1 Análise preliminar dos dados hidrológicos

Segundo Naghettini & Pinto (2007), a análise preliminar dos dados hidrológicos visa extrair as características empíricas essenciais do padrão de distribuição de uma variável hidrológica. Essa averiguação compreende um conjunto de métodos e técnicas estatísticas, na busca de evidências e padrões empíricos de variabilidade. Por essa razão, dado que o objetivo da estatística é a produção da melhor informação possível a partir dos dados disponíveis, foi utilizada esta ferramenta para obter as primeiras conclusões sobre as características empíricas do banhado.

A primeira tarefa realizada na investigação estatística foi a análise visual das séries, onde foram traçados gráficos dos dados disponíveis. Essa avaliação permite visualizar características dos dados como padrões, mudanças ao longo do tempo e as relações entre as variáveis. Na Figura 14, encontra-se a série mensal de níveis do banhado para o período compreendido entre 1960 a 2011. A análise visual da série não demonstra um padrão bem definido, todavia pode-se perceber a ocorrência de uma oscilação bastante acentuada em alguns períodos ao longo da série.

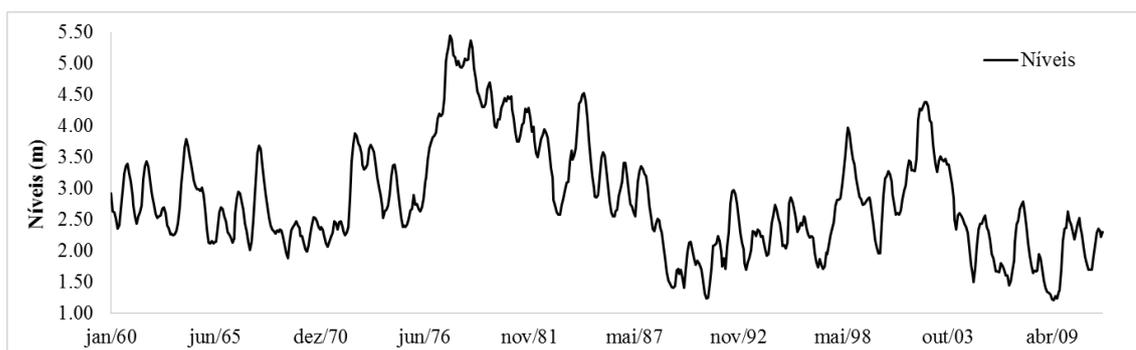


Figura 14 – Série temporal dos níveis no banhado do Taim (1960 - 2011).

A Figura 15 apresenta a sazonalidade mensal dos dados de precipitação, onde pode-se notar que as médias mensais estão uniformemente distribuídas ao longo do ano, com um sensível aumento para o período de inverno. Na Figura 16 encontra-se o comportamento da precipitação mensal e suas respectivas frequências. Observa-se uma assimetria acentuada o qual é justificada pela sazonalidade da distribuição da precipitação. Na Figura 17 encontra-se o comportamento da precipitação mensal discretizado mensalmente, onde pode-se observar de forma mais clara as variações ao longo dos anos para cada mês.

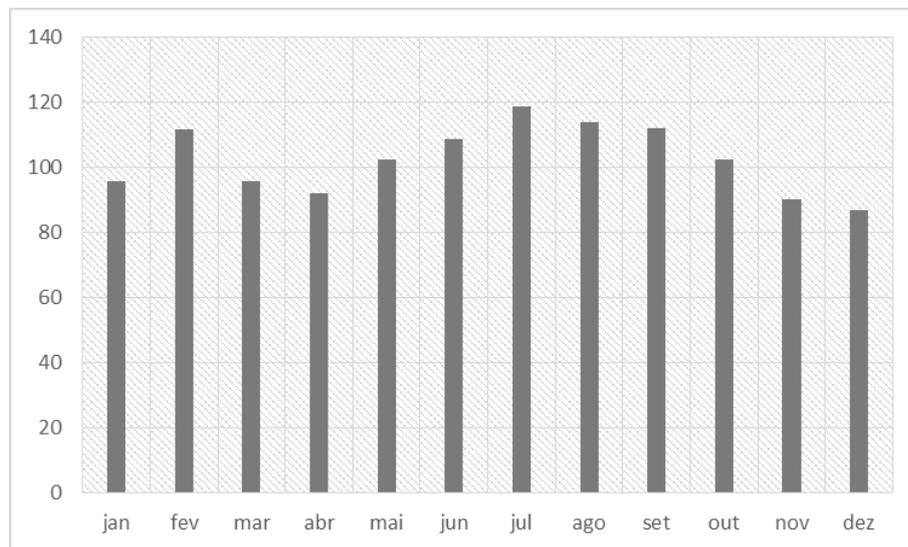


Figura 15 - Sazonalidade da série de precipitação (1960 - 2011).

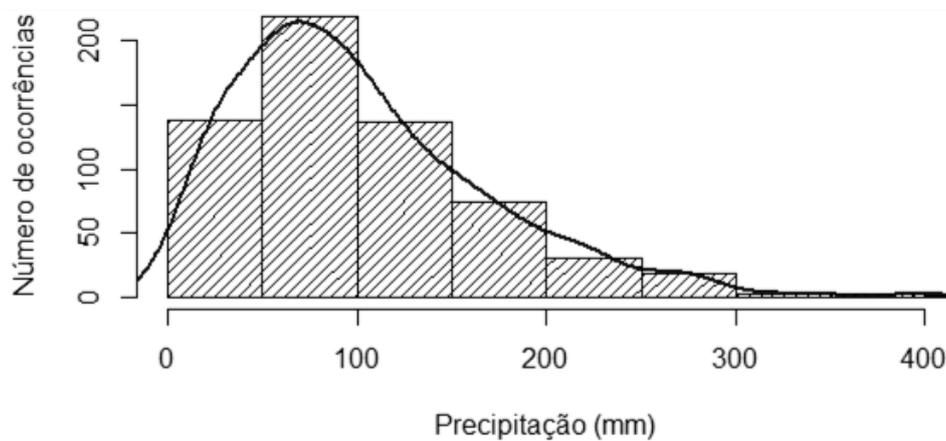


Figura 16 - Histograma da precipitação observada (1960 - 2011).

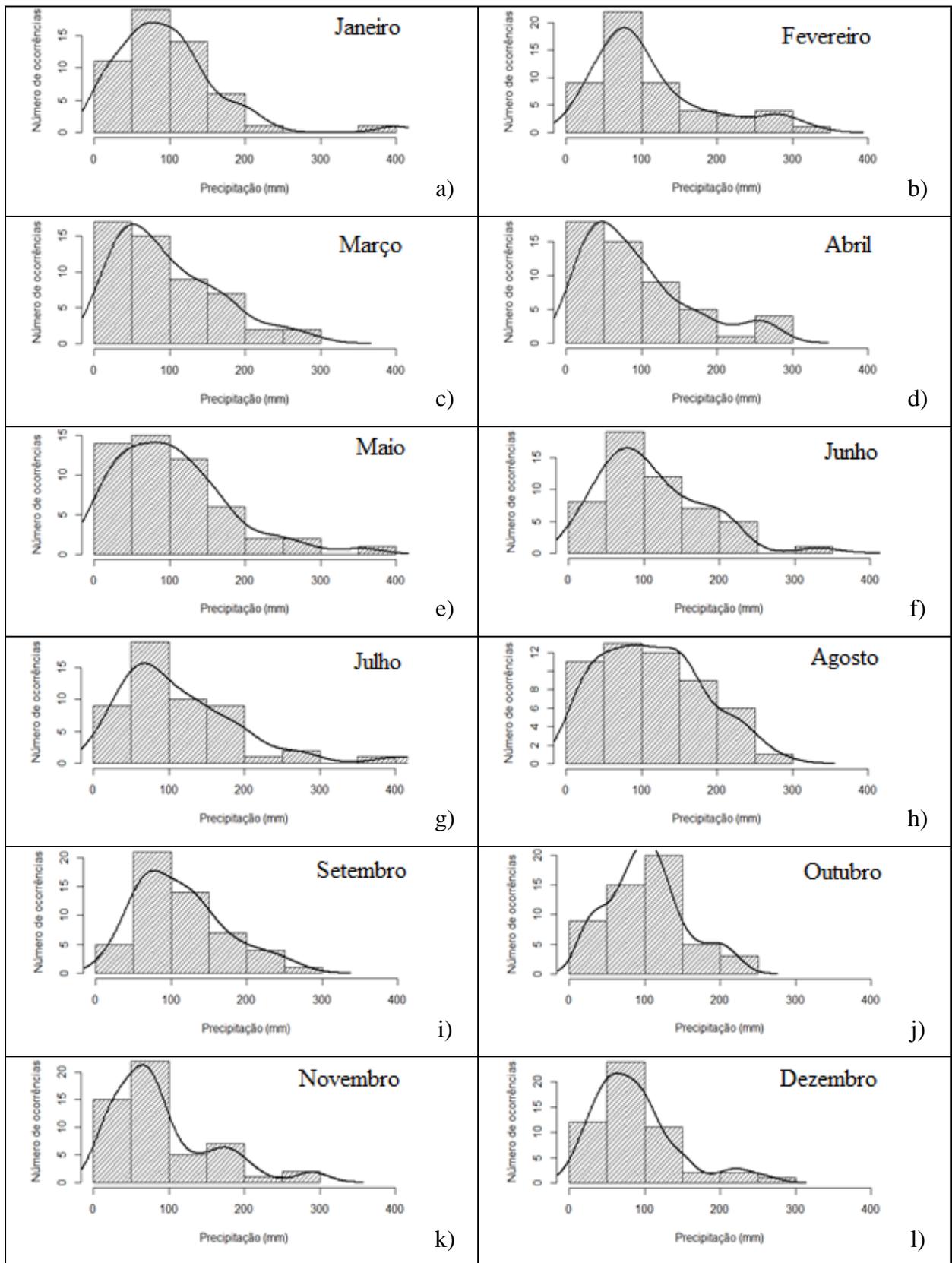


Figura 17 - Histogramas da precipitação para os meses de janeiro a dezembro (a - l).

Nas Figura 17.h e Figura 17.j que representam respectivamente os meses de agosto e outubro, pode-se notar uma forma quase simétrica, revelando uma variação bastante

acentuada em relação aos outros meses. Nas demais figuras existe um comportamento similar ao da série anual, com uma assimetria bem definida.

A Figura 18 apresenta a sazonalidade mensal dos dados de evaporação, onde pode-se notar que as médias mensais possuem uma variação mais acentuada, com maiores picos nos meses de janeiro e dezembro. Tassi (2008) explica que esse efeito é devido “há uma forte componente sazonal onde ocorrem maiores taxas de evaporação durante os períodos mais quentes do ano”.

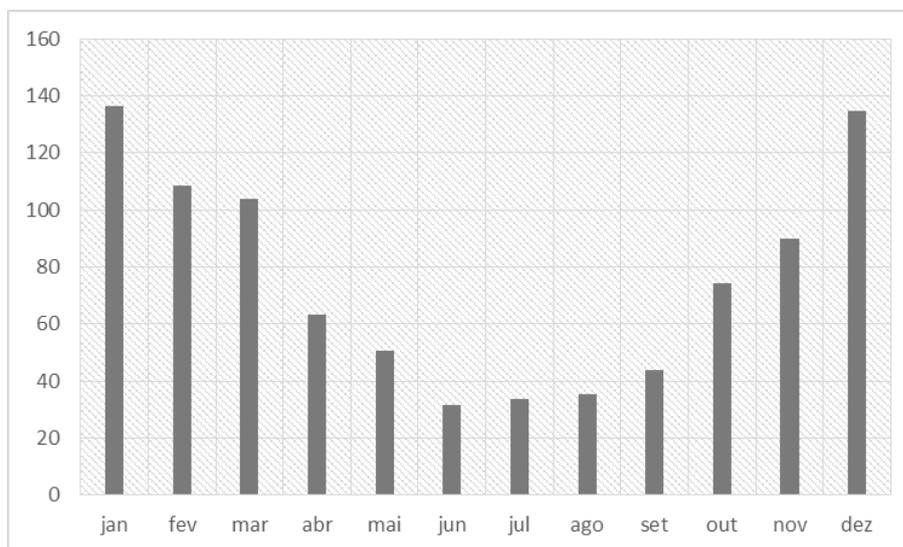


Figura 18 - Sazonalidade da série de evaporação (1960 - 2011).

A Figura 19 ilustra a série de níveis discretizada anualmente, onde cada linha representa um ano entre 1960 a 2011. Nela pode-se verificar de modo mais claro as oscilações durante os anos, onde fica mais evidente o comportamento mensal da série. Visualmente é fácil perceber a existência de anos completamente acima da média, indicando boa disponibilidade d'água, como também a existência de anos completamente abaixo da média, o qual necessita-se de uma atenção maior quanto a utilização racional da água.

Da análise da série inferiu-se que a maior variação registrada no mesmo ano ocorreu em 1967, onde o nível mínimo de 2,01 m ocorreu no mês de abril e o máximo de 3,68 no mês de outubro, acarretando em uma variação de 1,67 m. O ano de 1978 foi o ano em que os níveis no banhado estiveram mais elevados, estando acima de 4,93 m durante todo o ano, possuindo uma variação máxima de 0,43 m. Em 2009 o banhado obteve seus níveis mais baixos, com a cota mais elevada de 2,35 m no mês de dezembro e valor mínimo de 1,21 no mês de maio.

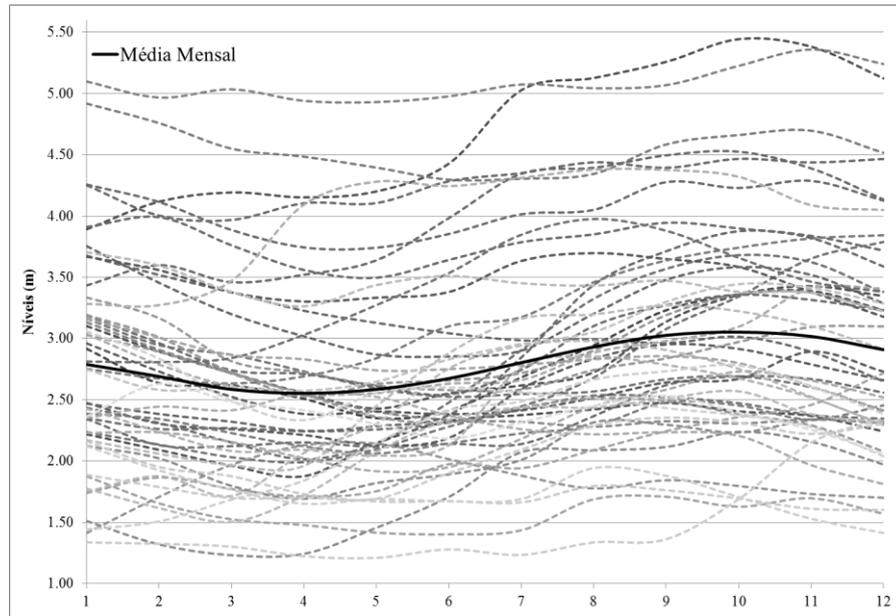


Figura 19 - Discretização mensal da série de níveis.

Para esse estudo existem 4 meses ao qual será dado maior importância, são os meses que são efetuadas as retiradas para irrigação. Uma primeira análise é identificar quais os meses que existem uma maior disponibilidade de água e verificar se coincidem com os meses em que são realizadas as retiradas. Para identificar os meses de maior disponibilidade hídrica foi realizado um ranking entre os meses, da seguinte forma:

1. Atribui a cada mês do respectivo ano uma classificação variando entre 1 a 12 (essa classificação é calculada com base no nível, ou seja, se um mês obteve o maior nível em relação aos outros meses do ano ele recebe o valor 1 e o mês com nível mais baixo recebe o valor 12);
2. Somam-se todos os valores atribuídos para cada respectivo mês;
3. Os meses que obtiverem os menores valores serão os meses que possuem os níveis mais elevados ao longo dos anos.

Esse ranking serve para identificar quais os melhores meses para se realizar as retiradas.

6.1.1 Divisão da amostra de dados

Quando se deseja verificar a eficiência de um modelo precisa-se de dados observados para a mensuração dos erros. Por esse motivo, para elaboração de um modelo e realização de testes quanto à sua eficiência, faz-se necessário possuir dados para cada etapa de construção do modelo, o que implica em possuir amostras de dados suficientes para cada etapa. Assim, os dados foram divididos em três partes para cada etapa de elaboração do modelo: a) formulação

e calibração; b) verificação (ou validação) e; c) simulação. A série de níveis possui dados entre os anos de 1960 a 2011, que corresponde a um total de 51 anos (624 meses). Esse intervalo é comum aos dados de precipitação, evaporação e retiradas para irrigação.

Conforme foi discutido no capítulo 3, a validade dos modelos empíricos fica limitada à amplitude dos dados que os originaram. Assim, para o período de formulação do modelo, faz-se necessário utilizar uma amostra dos dados que seja representativa de todo o comportamento geral da série. Dessa forma, analisando-se visualmente a série disponível de níveis, ficou constatado que o melhor período para a formulação e a calibração do modelo está compreendido entre o intervalo de 1976 a 2003. Nessa amostra, ficam evidentes os períodos mais úmidos e os secos de toda a série. Assim, além de conter os extremos amostrais, ela também possui trechos com um comportamento similar aos períodos remanescentes.

Da análise das estatísticas da série de dados pode-se verificar que a escolha da divisão da amostra possui intervalos razoavelmente bons, onde o intervalo reservado para formulação do modelo possui o valor máximo e o valor aproximado do valor mínimo da série temporal (diferença de 0,023 que corresponde a 23 mm, valor que pode ser considerado pequeno devido aos erros de medição convencionais). Na Figura 20 pode-se verificar de forma mais clara os resultados da estatística e como eles exercem influência na escolha dos períodos.

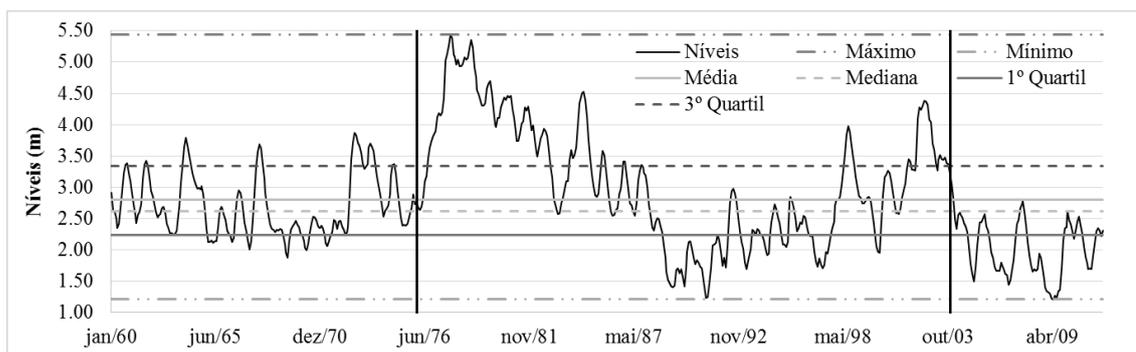


Figura 20 - Gráfico dos níveis com os resultados estatísticos plotados.

Da linha que indica o 3º quartil pode-se constatar que a maior parte dos 25% dos valores que estão acima de 3,34 m esta compreendida entre o intervalo de formulação e calibração do modelo (1976 a 2003). Entre as linhas que indicam o valor mínimo e o 1º quartil pode-se verificar que parte significativa dos 25% dos valores da amostra que estão abaixo de 2,23 m encontram-se compreendidos no intervalo de formulação e calibração do modelo. Assim, fica comprovado que a divisão da amostra está razoavelmente boa, compreendendo boa parte das flutuações dos níveis para o período de formulação do modelo.

Conforme anteriormente mencionado, são necessários dados observados para verificar a eficiência do modelo, assim, para os períodos de validação e simulação, usou-se os dados do passado como se fossem dados atuais, para mensurar quão bom o modelo consegue simular (validação) e prever (simulação) os níveis em condições reais.

Os dados de precipitação prevista que serão utilizados para este estudo, possuem uma disponibilidade temporal compreendida entre anos de 2004 a 2015. Mas, os dados de níveis estão limitados até o ano de 2011. Logo, o período de simulação deve estar compreendido nos anos em que existem dados de previsão de precipitação. Assim, o período de simulação ficou compreendido entre os anos de 2004 a 2011, pois são anos em comum para os níveis e a precipitação prevista. Restando o período entre 1960 a 1975 para a etapa de verificação.

6.1.2 Correlações entre as variáveis

Existem diferentes formas pela qual os dados podem se relacionar e para detectar esse encadeamento foram utilizadas algumas ferramentas estatísticas para auxiliar nessa investigação, onde foi usada a inspeção por meio de gráficos, diagramas de dispersão, coeficiente de correlação linear de Pearson e o coeficiente de correlação de Spearman, onde os cálculos desses coeficientes foram realizados utilizando os programas RStudio (versão 0.98.1028) e o R (versão 3.1.2), na versão para Windows.

6.1.2.1 Decomposição das séries temporais

Variáveis que apresentam um comportamento irregular ao longo do tempo, como é o caso da precipitação, são difíceis de identificar um padrão. Da análise do gráfico das precipitações (Figura 21), pode-se verificar a ocorrência de um comportamento irregular, aparentemente desprovido de qualquer tipo de padrão. Por esse motivo, faz-se necessário uma investigação fragmentada desse comportamento. Portanto, para efetuar uma melhor análise dos dados, realizou-se a fragmentação das séries utilizando o método conhecido por Decomposição Clássica, que tem por objetivo decompor a série temporal, para auxiliar na exploração e previsão de futuros valores da variável em estudo.

As séries foram decompostas em quatro partes: 1) tendência; 2) ciclo; 3) sazonalidade e; 4) termo aleatório. As séries de dados estão em escala de tempo mensal. Contudo, como não é conhecido em que escala de tempo as componentes decompostas se manifestam, foi investigado o comportamento das séries na mesma grandeza dos dados originais, ou seja, em escala mensal.

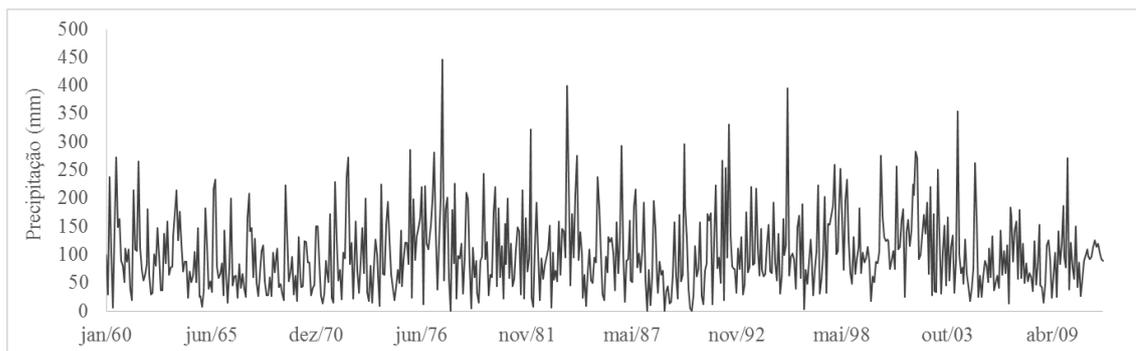


Figura 21 - Precipitação no banhado do Taim (1960 a 2011).

Primeiramente as séries foram fragmentadas na componente tendência-ciclo, que constituem o comportamento de longo prazo das séries, permitindo a exploração dos dados sem o efeito da sazonalidade e dos componentes aleatórios. Essa separação foi realizada utilizando o método das médias móveis e considerando-se que a componente ciclo comumente possui flutuações com duração superiores a um ano, foi adotado o período de um ano para as médias móveis. Para a realização do ajuste da tendência-ciclo faz-se necessário que o intervalo das médias móveis seja um número ímpar, assim foi efetuado as médias móveis centradas, que consiste na média móvel entre as duas médias móveis consecutivas, situando-se no ponto médio entre as duas.

6.1.2.2 Teste de correlação não linear

Conforme mencionado nos tópicos anteriores, não foi encontrada boa correlação entre as séries de dados. Por essa razão foram testadas algumas formas de correlação conhecidas. Essa pratica é bastante utilizada em cálculo de regressão, conhecidas como regressão não-linear e regressão com linearização de funções não lineares. Esta metodologia parte do pressuposto de que alguns dados se correlacionam de forma não-linear, assim utiliza transformações nos dados permitindo a aplicação da regressão não linear ou a linearização de funções não lineares.

O primeiro teste realizado consistiu em calcular o logaritmo natural das séries. Esta opção calcula o logaritmo natural (neperiano) de todas as observações das séries selecionadas. Esta metodologia é bastante útil quando deseja-se modificar as séries de comportamento exponencial acelerado ou para o cálculo de regressões. Posteriormente foram realizados testes com funções exponenciais, polinomial e raiz.

6.1.2.3 Defasagem entre as séries

Com o objetivo de avaliar como as séries se correlacionam, foi realizado o teste de correlação com as séries defasada no tempo. A análise das séries foi realizada da seguinte forma:

1. Sem defasagem;
2. Com defasagem.

A análise sem defasagem, ou seja, o mesmo mês para os níveis e as séries analisadas, foram realizadas para avaliar como as variáveis se relacionam no mesmo intervalo de tempo.

A análise com defasagem, ou seja, mês atual para a série de níveis correlacionadas com as séries analisadas no tempo passado, foram realizadas para avaliar como as séries de dados no tempo passado explicam o comportamento da série de níveis no tempo atual. As análises foram realizadas para o tempo t para os níveis e para as séries avaliadas foram variam de $t-1$ a $t-12$.

6.1.2.4 Teste com médias móveis

Conforme foi demonstrado no tópico 7.1.2.1, as médias móveis causam um amortecimento na variação dos dados, ocasionando uma redução dos ruídos existentes nas amostras. O objetivo da utilização das médias móveis nesse trabalho consiste em examinar a influência da ocorrência dos valores passados das variáveis explanatórias (precipitação e evaporação) em relação à variável resposta (níveis no banhado). A vantagem na utilização desse método está no fato de se analisar o comportamento médio dessas variáveis hidrológicas, reduzindo a aleatoriedade dos dados para tentar compreender a influência do comportamento geral da série.

Foram analisadas as médias móveis aritméticas e exponenciais para um período máximo de 12 meses, onde realizou-se estudos das correlações por meio dos coeficientes de correlação e dos diagramas de dispersão. Conforme será visto nos próximos tópicos, foi testada a possibilidade de uma única equação para simular diferentes horizontes de previsão, ou uma equação para cada horizonte. Portanto, as análises de correlação foram realizadas da seguinte forma: a) mês a mês (o mesmo mês para os níveis e as variáveis resultantes das médias móveis) (Figura 22.a) e; b) mês atual para os níveis com 1º ao 7º mês no passado para as médias móveis (que corresponde aos horizontes de previsão proposto) (Figura 22.b).

Datas	Níveis	Precipitação	Média Móvel de dois meses
jan/60	2.915915	99.0	
fev/60	2.635750	29.0	64
mar/60	2.617624	238.0	133.5
abr/60	2.507614	103.0	170.5
mai/60	2.350440	6.0	54.5
jun/60	2.412825	146.0	76
jul/60	2.683105	272.0	209
ago/60	2.945150	149.0	210.5
set/60	3.230636	164.0	156.5
out/60	3.358797	89.0	126.5

a)

Datas	Níveis	Precipitação	Média Móvel de dois meses
jan/60	2.915915	99.0	
fev/60	2.635750	29.0	64
mar/60	2.617624	238.0	133.5
abr/60	2.507614	103.0	170.5
mai/60	2.350440	6.0	54.5
jun/60	2.412825	146.0	76
jul/60	2.683105	272.0	209
ago/60	2.945150	149.0	210.5
set/60	3.230636	164.0	156.5
out/60	3.358797	89.0	126.5

b)

Figura 22 - Análise de correlação: a) mês a mês; b) mês atual para os níveis com o mês anterior das médias móveis de dois meses da precipitação.

Foi determinado um período máximo de 12 meses para as médias móveis, com base no horizonte máximo de previsão (sete meses), pois, à medida que aumenta-se a quantidade de dados para as médias móveis, as oscilações oriundas dos membros de previsão serão absorvidas pelos valores observados, tornando inviável a utilização do método proposto para essa pesquisa.

6.1.2.5 Teste com suavização exponencial

A escolha de testar esta técnica consiste na possibilidade de amortecimento dos valores da série baseados nos argumentos passados. Assim, um determinado mês propagaria sua influência aos meses subsequentes. Pode-se pensar numa explicação física quanto a sua utilização, como forma de avaliar quanto às quantidades precipitadas (ou evaporadas) no mês anterior, poderiam influenciar no período atual.

O método pode ser utilizado para um ou mais meses no passado e para este trabalho foi utilizado apenas o mês anterior como referência. Essa escolha foi motivada pela análise das séries, como pode-se citar as precipitações, na qual oscilam muitos entre seus picos variando entre valores máximos e mínimos. A escolha do parâmetro de suavização α foi realizada por meio de tentativas começando pela unidade, variando até encontrar um valor ótimo para o coeficiente de determinação.

6.1.3 Retiradas para irrigação

Sabendo-se que a água usada para a irrigação da orizicultura é extraída diretamente da Lagoa Mangueira e que, em estudos anteriores de simulação de níveis no banhado, percebeu-se que a principal fonte de água do banhado advém da lagoa Mangueira. Logo, é plausível pensar no banhado como uma extensão da lagoa Mangueira. Assim, as simulações de retirada

de água para irrigação foram realizadas com base na curva cota-volume da lagoa Mangueira e ajustada para o banhado do Taim. A curva cota-volume da Lagoa Mangueira (Figura 23) foi obtida através de medições batimétricas, cedidas pelo instituto de pesquisas hidráulicas (IPH-UFRGS).

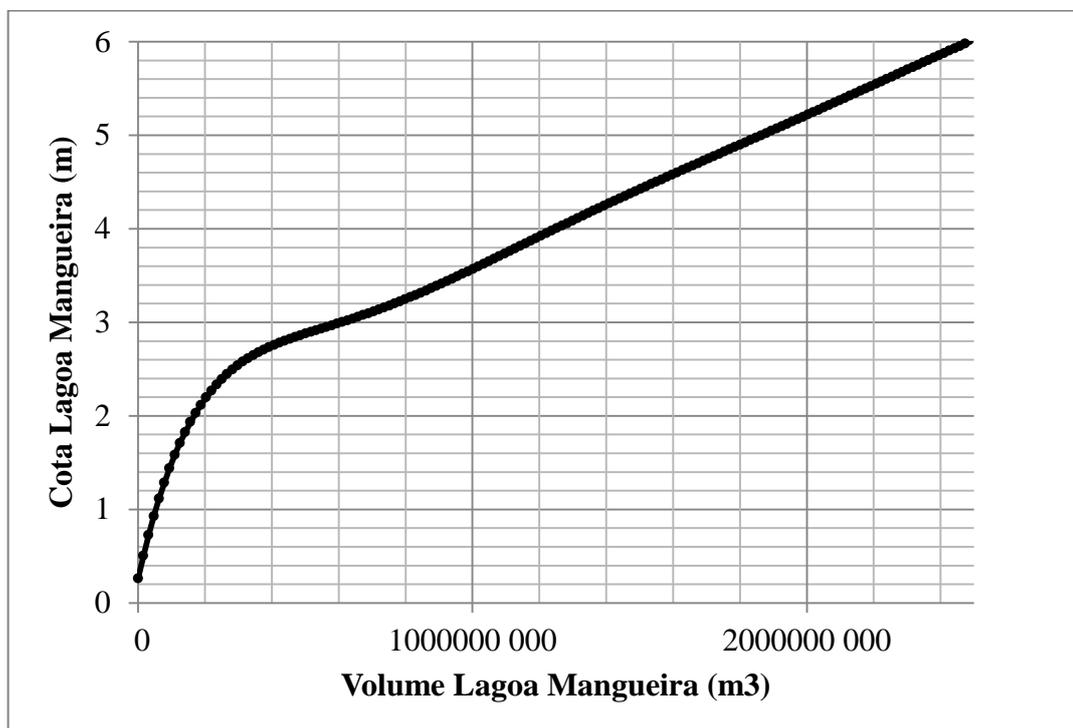


Figura 23 - Gráfico da curva Cota-Volume da Lagoa Mangueira.

Para fazer uso da curva cota-volume da lagoa Mangueira, faz-se necessário compreender a interface entre o banhado do Taim e a lagoa Mangueira (Figura 24). Paz (2003) descreve a essa interface da seguinte maneira: “A interface entre o banhado do Taim e a Lagoa Mangueira constitui uma região de pouquíssimas informações, apesar da importância no que diz respeito a interação entre os dois sistemas”. Villanueva (1997) relata um desnível da ordem de 50 cm que não deveria existir, pela continuidade física dos corpos d’água. Devido à falta de informação referente a essa interação, foi necessário analisar os resultados obtidos em trabalhos que realizaram simulações entre a lagoa e o banhado.

Fragoso (2009) realizou simulações na lagoa Mangueira utilizando um modelo que divide a lagoa em células, onde cada célula possui uma cota para um determinado referencial. Ele comparou os resultados de sua simulação com os valores observados na lagoa (Figura 25.a), obtendo bons resultados. Logo, sabendo que os picos na lagoa correspondem aos picos no banhado, comparou-se os valores dos picos onde concluiu-se uma diferença média de 1,50 m entre os valores das cotas no banhado e a lagoa Mangueira. Assim, adicionou-se esse

incremento de 1,50 m aos valores do banhado para fazer uso da curva cota-volume da lagoa Mangueira quando analisada a retirada de água.

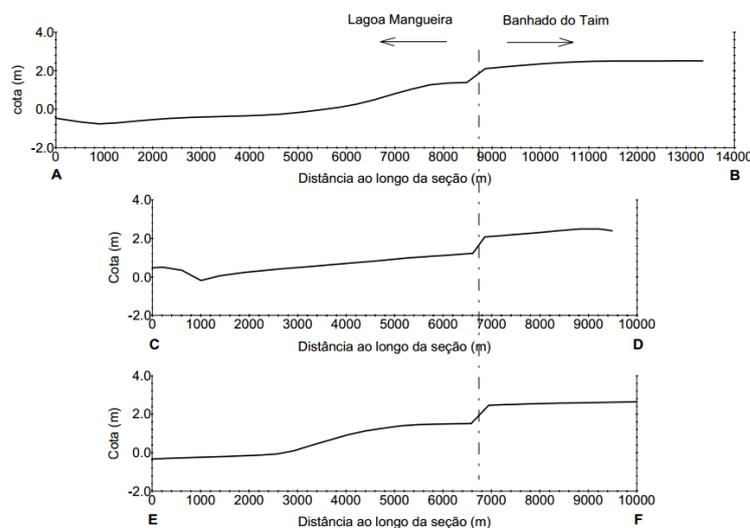


Figura 24 - Perfis topográficos longitudinais da interface entre o banhado do Taim e a lagoa Mangueira, adaptado de Paz (2003).

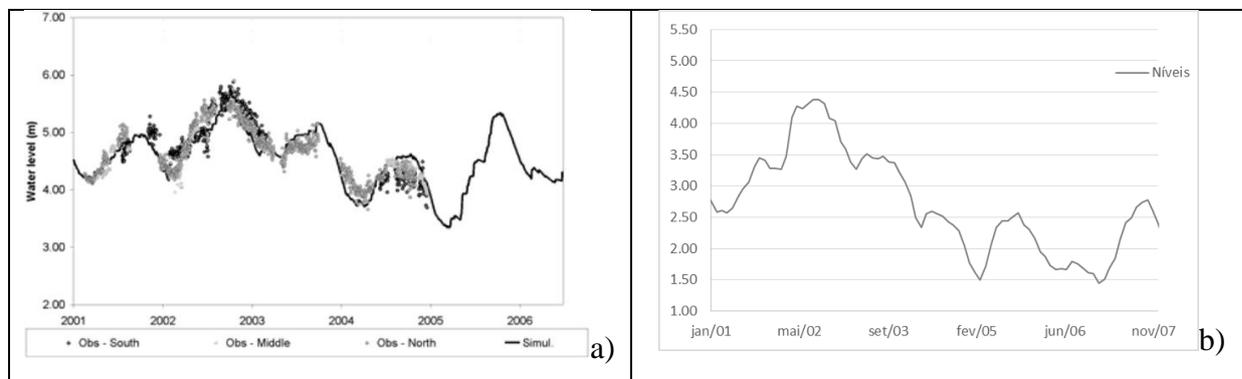


Figura 25 – Níveis do Sistema Hidrológico do Taim para o intervalo entre 2001 - 2007: a) níveis observados e simulados (adaptado de Fragoso, 2009); b) níveis no banhado do Taim em metros.

A retirada de água da Lagoa Mangueira para irrigação foi considerada da seguinte forma:

- a) estipula-se o volume a ser retirado para irrigação;
- b) o volume a ser retirado será captado em 4 etapas (16,7% em dezembro, 33,3% em janeiro, 33,3% em fevereiro e 16,7% em março).

Na determinação do volume a ser retirado é considerada uma demanda de água para irrigação. Assim, o cálculo é realizado considerando a área plantada e o consumo de água por área plantada para irrigação (no caso do banhado é o arroz). Um estudo realizado pelo IPH (1996), estimou esse consumo por área plantada de arroz em 2 L/s/ha. De posse da área que se deseja plantar, multiplica-se pelo consumo por área plantada e obtêm o volume estimado para irrigação. O exemplo abaixo demonstra esse cálculo:

Ex.: Suponha uma estimativa de área plantada de 25430,81 há,

$$\left[\left(\frac{25430,81 \times 2 \times 3600 \times 24 \times 90}{1000000000} \right) \times 0,4 \right] \times 1000000 = 158200000 \text{ m}^3$$

onde 0,4 é uma constante de proporcionalidade que representa o retorno de água esperado.

Para utilizar essa captação de água no modelo proposto, precisamos transformar o volume em uma diferença de nível, para isso faz-se necessário realizar algumas simplificações. Como a retirada é realizada em 4 etapas, dividiu-se o volume estimado pelas suas porcentagens. Para transformar os volumes em cotas, precisamos de uma referência, mas os meses em que ocorrem as retiradas ainda não foram previstos. Logo, foi utilizado o mês de novembro (mês anterior ao primeiro mês com captação de água) como referência. Com os volumes definidos para cada retirada, realiza-se o cálculo como se as retiradas fossem efetuadas no mês de referência, faz-se uma regra de três com os valores da curva cota-volume. Assim, obtém-se uma aproximação para o impacto das retiradas para os meses com captação.

Em Tassi (2008) encontra-se um comparativo entre o regime hidrológico real no banhado do Taim e um cenário hipotético sem irrigação (Figura 26), onde o modelo utilizado para gerar as simulações considera os processos físicos como infiltração, evaporação, hidrodinâmica, etc. Assim, pode-se ter uma melhor dimensão do impacto causado pelas retiradas para irrigação.

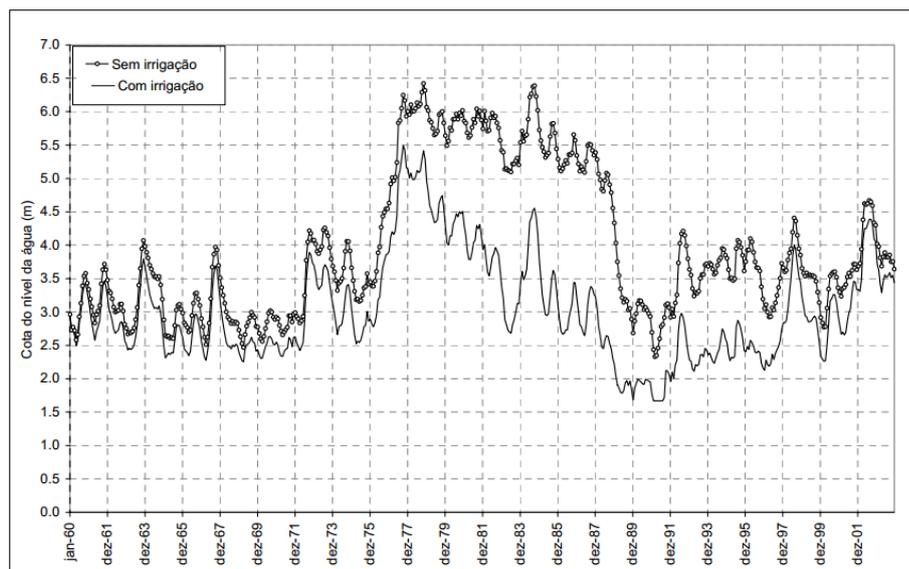


Figura 26 - Comparação entre o regime hidrológico real no banhado do Taim e cenário hipotético sem irrigação. Adaptado de Tassi (2008).

6.1.4 Testes não paramétricos

Selecionados os dados que serão utilizados para estimar as equações que simularão o banhado, faz-se necessário conhecer a representatividade estatística dos dados. Em estudos

que utilizam séries dados hidrológicos é de fundamental importância a verificação da premissa de base, na qual se trata de uma amostra aleatória simples extraída de uma população única, cuja função de probabilidades é desconhecida a priori. Inserida nessa premissa estão implícitas as hipóteses de aleatoriedade, independência, homogeneidade e estacionariedade.

Foram realizados testes de hipóteses não paramétricos, onde as séries foram avaliadas para os intervalos de elaboração do modelo (formulação, validação e simulação). Assim, pode-se verificar se as características dos dados se mantêm ao longo dos períodos estabelecidos.

6.2 Estimativa de modelos empíricos de simulação

Os parâmetros do modelo empírico estão intimamente ligados com a forma da equação, ou seja, com as características da equação, seja ela linear, polinomial, exponencial, etc. O objetivo é ajustar a equação composta pelas variáveis explanatórias à variável resposta, onde os parâmetros da equação ponderam (atribuem pesos) aos valores dos dados de entrada.

Existem diversos métodos para estimar parâmetros, seja para equações lineares ou não-lineares. Contudo, esse trabalho abordará este tema fazendo uso de metodologias empíricas que possam ajustar uma equação para simular níveis no banhado do Taim, utilizando apenas os dados das séries de precipitação e evaporação. Os tratamentos e inferências, realizadas nas séries, já foram discutidos nos tópicos anteriores. Assim, este tópico abordará os passos que foram seguidos na determinação das equações encontradas.

Uma importante análise para previsão de futuros cenários no banhado consiste em analisar o impacto das retiradas para irrigação. Entretanto, a série que se deseja prever é fruto de um balanço hídrico entre precipitação, evaporação, trocas subterrâneas, retiradas, etc. Ou seja, ela já possui as retiradas para irrigação em sua amostragem, sendo necessário desagregar esse valor das previsões. Dessa forma, o usuário do modelo poderá ir variando o valor das retiradas, podendo fazer julgamentos sobre um valor ótimo a ser retirado nos meses futuros, servindo assim como importante ferramenta para gestão.

As metodologias utilizadas nesse trabalho visam ajustar parâmetros às variáveis explicativas, formando uma equação que obtenha como resultado os níveis no banhado. Assim, a variável responsável pelas retiradas para irrigação obterá um peso ponderado na simulação dos níveis, o que não seria desejável, pois reduziria sua influência no produto final (os níveis). Esse problema foi contornado da seguinte forma:

- a) Transformam-se os volumes retirados em alturas de águas retiradas (esse procedimento encontra-se detalhado no tópico 6.1.3);
- b) Adicionam-se os valores das retiradas no passo “a” aos níveis do banhado;
- c) Estima-se uma equação para simular os níveis somados as retiradas (que seria os níveis se não tivessem ocorrido as retiradas);
- d) Adiciona-se um termo de retirada à equação.

Conforme discutido no tópico 6.1.3, as retiradas para irrigação analisadas pela metodologia física diferem do modo como abordamos nesse trabalho. Assim, a afirmação no passo “c” constitui uma simplificação, podendo ocorrer uma propagação dessa influência nos meses subsequentes.

O processo de elaboração do modelo foi composto de três etapas:

1. Formulação do modelo, nessa etapa os dados são utilizados para gerarem a equação que simulará o banhado;
2. Verificação (ou validação), nesse estágio a equação elaborada no passo anterior será utilizada para simular os níveis;
3. Simulação, fase em que a equação será utilizada para realizar previsões.

6.2.1 Modelo de regressão linear simples

A metodologia dos modelos lineares simples possui apenas uma variável explicativa, e como o objetivo desse estudo consiste em realizar previsões de níveis utilizando dados de precipitação prevista, os dados utilizados para gerar essas equações necessitam conter a precipitação. Logo, existem cinco possibilidades para as variáveis selecionadas no capítulo anterior, que são: 1) precipitação (P); 2) precipitação-evaporação (P-E); 3) média móvel de cinco meses das precipitações (MMP_5); 4) média móvel de sete meses das precipitações- evaporações (MMP-E_7); 5) suavização exponencial das precipitações (Sav_P).

Os parâmetros das equações foram estimados por mínimos quadrados. Os critérios utilizados para analisar os resultados foram as medidas de desempenho e a avaliação visual por meio de gráficos. Todas as análises foram realizadas utilizando-se os dados observados comparados aos valores simulados pelas equações.

As equações lineares simples possuem a forma da equação 66:

$$Y(x) = a + b \cdot x \tag{66}$$

onde “a” é a constante de proporcionalidade, “b” é o parâmetro que pondera a variável explicativa e “Y” é a variável resposta.

6.2.2 Modelo de regressão não linear simples

As regressões não lineares foram realizadas mediante transformações nas variáveis independentes utilizando-se as funções testadas no tópico 6.1.2.2, o qual realizou o cálculo de correlação linear de Pearson entre os níveis e os dados calculados com algumas funções conhecidas. Assim, as regressões foram realizadas da seguinte forma:

1. Realiza-se a transformação das variáveis independentes utilizando o cálculo da função;
2. Estima-se os parâmetros utilizando o método dos mínimos quadrados para cada um dos resultados do passo 1.

As funções testadas para as regressões não lineares simples foram: a) polinomial, calculando-se a função $f(X)^i$, para $i=2,3,4$; b) logarítmica, calculando-se a função $f(LN(X))$, para os valores da variável independente inferiores a unidade, foi atribuído o valor 1 para evitar a obtenção de valores negativos; c) raiz, calculando-se a função $f(\sqrt[i]{X})$, para $i=2,3,4$. O cálculo das funções exponenciais obteve resultados ruins, por essa razão não foram realizados testes de regressão não linear usando as funções exponenciais.

Partindo do pressuposto do tópico anterior que as equações precisam conter os dados de precipitação, as regressões foram realizadas para os mesmos dados utilizados para as regressões lineares simples.

6.2.3 Modelo de regressão linear múltipla

Das análises realizadas no tópico 6.1, foram selecionadas dez variáveis para a realização da regressão linear múltipla. Seguindo o mesmo raciocínio dos tópicos anteriores, a equação da regressão precisa conter a precipitação. Existem diversas formas de combinar essas variáveis (contendo a precipitação), o que resultaria em um considerável número de equações a serem avaliadas. Assim, para auxiliar na seleção dos dados, elaboração e avaliação dessas equações de regressão, foi implementado um algoritmo, em linguagem Fortran, para realizar as regressões visando analisar todas as combinações possíveis, com os dados selecionados. O algoritmo é composto por sub-rotinas específicas para cada análise com o objetivo final de gerar equações de regressão linear múltipla para N variáveis independentes. Na Figura 27 encontra-se um fluxograma simplificado das etapas realizadas pelo algoritmo. No Apêndice A encontram-se o algoritmo elaborado.

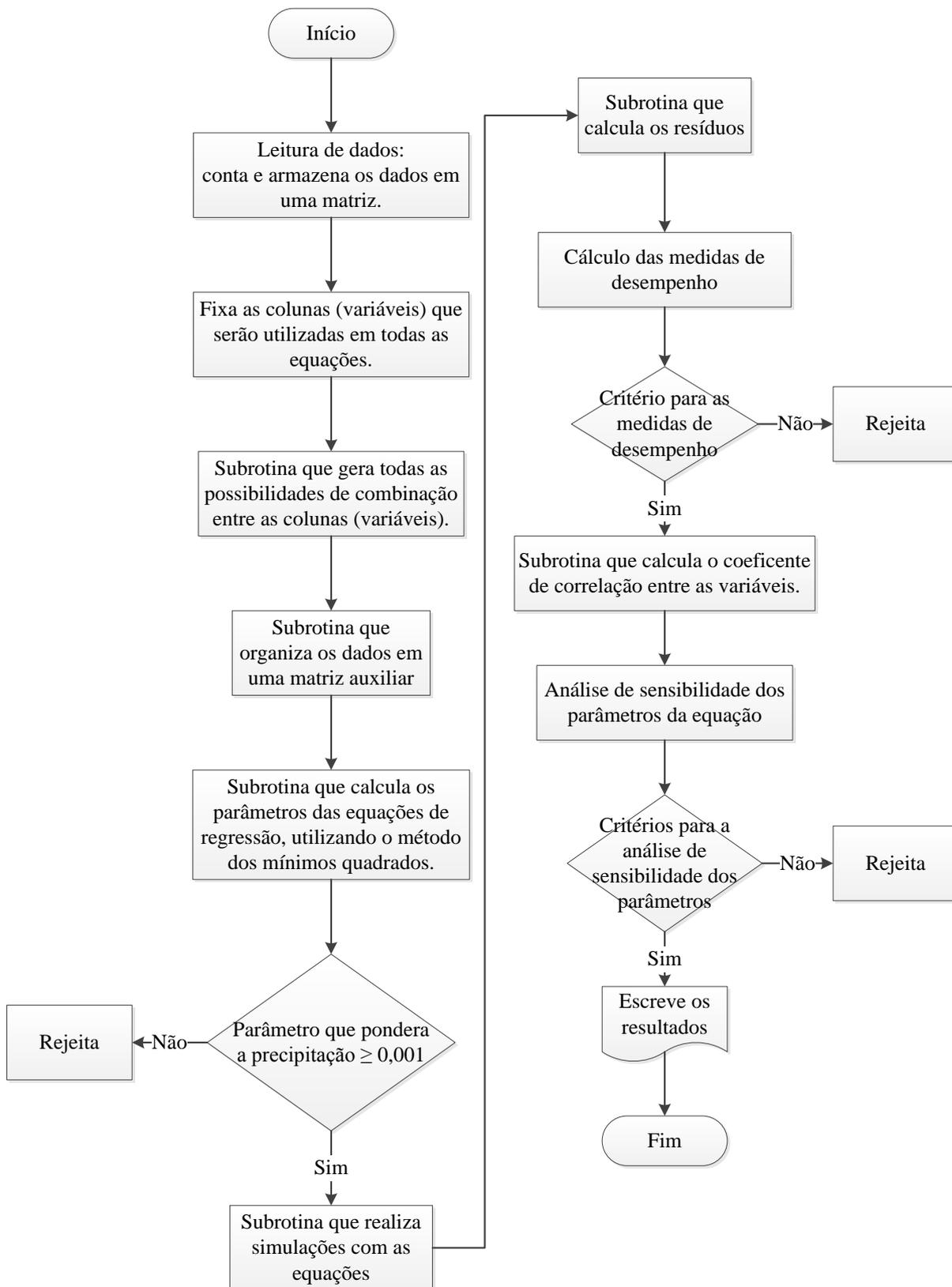


Figura 27 – Fluxograma simplificado do algoritmo elaborado para realizar regressões lineares múltiplas.

As equações de regressão linear múltipla possuem a forma semelhante à equação 67:

$$Y(X) = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_n \cdot x_n \quad (67)$$

onde $A(a_0, a_1, \dots, a_n)$ são os parâmetros da equação e $X(x_1, x_2, \dots, x_n)$ são as variáveis independentes.

Fixando uma variável independente (precipitação), o algoritmo calculou 255 possibilidades (arranjo simples) de equações candidatas. Os parâmetros das equações foram estimados pelo método dos mínimos quadrados e o critério utilizado para analisar as equações foram as medidas de desempenho previamente apresentadas.

Mesmo a equação contendo a precipitação, sua influência para geração dos níveis pode ser considerada muito baixa (irrelevante) dependendo do parâmetro que a pondera na equação, o que não justificaria a proposta metodológica desse estudo. Por essa razão, foi utilizado um critério inicial para a avaliação das equações, onde apenas as equações que obtivessem um determinado valor mínimo para o parâmetro que pondera a precipitação fossem escolhidas para serem avaliadas pelas medidas de desempenho, sendo descartadas as demais. Foi adotado o limite inferior para o parâmetro que pondera a variável precipitação nas equações de regressão como 10% do menor valor observado da série de níveis. Assim, visto que a média dos valores observados da precipitação é de 102,5 mm e que o valor mínimo observado dos níveis é de aproximadamente 1 m, tem-se:

$$10\% \text{ de } 1 \text{ m} = 0,1 \text{ m} \quad \text{e} \quad 0,1\% \text{ de } 102,5 \text{ mm} = 0,1025 \text{ m} \quad (0,1 \approx 0,1025)$$

Logo, o valor limite para o parâmetro que pondera a precipitação foi igual ou superior a 0,001. Cabe ressaltar que os modelos de regressão são metodologias empíricas que não levam em consideração as unidades, que são parâmetros físicos. As equações ajustam as variáveis por meio dos parâmetros aos dados que serão previstos (ou simulados).

6.2.3.1 Avaliação das equações de regressão linear múltipla

Devido ao grande número de equações candidatas (255) a serem avaliadas, foram estabelecidos alguns limites para as medidas de desempenho para servirem como critério de exclusão para as equações que não obtivessem tais desempenhos.

As diferenças entre os níveis observados e os simulados (resíduos) constitui uma análise delicada, pois, conforme discutido no tópico 6.1.3, a cota está diretamente relacionada com o volume de água. Para evitar a ocorrência de erros grosseiros, foi estabelecido um limite máximo para os resíduos em 0,3 m. É evidente que o valor estabelecido para o limite dos resíduos é bastante grosseiro se comparado com as diferentes oscilações possíveis para os

níveis, mas para um critério de exclusão, foi considerado plausível. Assim, 47 equações (aproximadamente 18,4% do total) não apresentaram falhas ao critério adotado. Cabe ressaltar que, para esse momento, as análises dos resíduos serviram apenas como um critério de exclusão, porém, no tópico dos resultados será abordada uma análise mais detalhada para os resíduos das previsões.

Tassi (2008), que utilizou o modelo elaborado por Villanueva (1997), relatou que o erro na determinação dos níveis excedeu 70 cm somente em um mês, onde ela relatou considerar que erros dessa ordem poderiam ser decorrentes da própria incerteza na modelagem. Com base nesse argumento, pode-se admitir que o critério adotado de 0,3 m constitui uma avaliação rigorosa, porém bastante eficiente para filtrar a quantidade de equações.

O coeficiente de determinação (R^2) foi calculado para as 47 equações selecionadas, onde foi estabelecido um valor limite de maior ou igual a 0,7 para exclusão das equações que não obtivessem esse limiar. Esse valor significa que 70% da variável dependente consegue ser explicada pelas variáveis independentes presentes no modelo. Assim, todas as 47 equações obtiveram esse valor mínimo, sendo selecionadas para as próximas análises. O erro padrão da regressão (EPR) constitui o desvio padrão dos valores simulados da variável dependente ao redor da linha de regressão estimada, podendo ser interpretado como uma primeira estimativa dos limites da regressão. Os resultados para as 47 equações foram bastante satisfatórios (inferiores a 0,1), o que revela boas estimativas dos valores simulados. Para o coeficiente de Nash-Sutcliffe, foi estabelecido um limite igual ou superior de 0,7 para exclusão das equações que não obtivessem esse limiar. Todas as 47 equações obtiveram valores superiores a 0,7, podendo-se concluir que as equações possuem boa eficiência em representar a variável resposta para o período avaliado.

O cálculo das anomalias foi realizado para os referidos meses do ano, ou seja, o mês de janeiro foi avaliado com a média dos meses observados de janeiro e assim sucessivamente. Foi calculada a anomalia observada e simulada. O menor valor encontrado foi de 0,8 m, valor em comum para as anomalias observadas e calculadas, o que se pode concluir que as anomalias simuladas obtiveram bons resultados, haja vista que os valores são comuns às anomalias observadas.

6.2.3.2 Análise de sensibilidade das equações de regressão linear múltipla

A análise de sensibilidade procura mensurar o efeito de uma variável sobre o comportamento total da equação, tornando-se um instrumento útil para determinar a importância de uma variável sobre o resultado final. Nesse trabalho, essa análise foi mensurada baseada nos coeficientes que foram utilizados para medir o desempenho das equações.

A análise de sensibilidade pode ser utilizada para avaliar a resposta do modelo no caso de alguma variável possuir alguma percentagem de erro, que podem ser oriundos de falhas de medição, erros de arredondamento, etc.

As equações que possuíam um parâmetro ponderador para a precipitação elevado, foram reprovadas pelos critérios utilizados para medir o desempenho dos modelos, sendo essas equações consideradas ineficientes para simular o sistema. As equações selecionadas possuem uma baixa influência da precipitação e uma elevada influência do nível anterior. Esse resultado é plausível, dado que os níveis anteriores possuem uma elevada correlação linear com o nível atual, assim era de se esperar que a metodologia utilizada pelos modelos autoregressivos desse um maior peso a essa variável.

Em decorrência do grande número de equações encontradas, a análise de sensibilidade serviu como critério adicional de exclusão das equações, de que a variação da precipitação não seja percebida pelas medidas de desempenho. A análise de sensibilidade foi realizada utilizando-se o erro relativo percentual e o erro absoluto, onde o cálculo foi realizado da seguinte forma:

1. Reduz uma variável da equação a certa percentagem;
2. Calcula-se o valor da equação (nível) com a variável reduzida;
3. Calcula-se o erro absoluto e o erro relativo percentual.

As variáveis foram reduzidas em 10%, 20%, 30%, 40% e 50% do valor observado. As análises ocorreram variando-se apenas uma variável por vez. O erro absoluto foi calculado pela equação 68 e o erro relativo percentual pela equação 69.

$$E_{abs} = Md_c - Md_s \quad (68)$$

$$E_{rel} = \left| \frac{Md_c - Md_s}{Md_s} \right| \times 100 \quad (69)$$

onde E_{abs} é o erro absoluto, E_{rel} é o erro relativo percentual, Md_s é a medida de desempenho sem redução dos coeficientes, Md_c é a medida de desempenho com redução dos valores das variáveis.

Foi avaliada ainda a significância das equações de regressão múltipla utilizando o teste conhecido como “teste do F total”. O cálculo do valor de F das equações foi realizado pela análise de variância, onde foram construídas tabelas ANOVA para auxiliar nos cálculos.

6.3 Verificação de viés nos dados do modelo COLA

6.3.1 Detecção do viés

Para avaliar a necessidade de correção nos dados de precipitação prevista foram calculadas medidas de desempenho determinísticas e probabilísticas. As medidas de desempenho determinísticas foram utilizadas para uma primeira estimativa na detecção do viés. As medidas de desempenho probabilísticas foram calculadas para avaliar o desempenho do modelo de circulação global.

Para uma avaliação determinística, foram calculados três índices para mensurar os resíduos entre os dados observados da precipitação e os simulados pelo modelo de circulação global. Os índices calculados para mensurar o viés foram: a) Erro médio; Um índice que mede a relação entre o valor absoluto do viés e o valor do desvio absoluto médio do erro em relação ao viés (RV); c) Raiz quadrada do erro quadrático médio (REMQ).

Para uma avaliação probabilística foram calculados dois índices para avaliar o desempenho do MCG. Os coeficientes calculados foram: a) *rank histogram*; b) *Spread-bias diagram*.

6.3.2 Remoção do viés

A remoção do viés foi realizada em três etapas:

1. Ajuste de uma função cumulativa de probabilidade;
2. Remoção do viés utilizando o método do Mapeamento Quantil-Quantil;
3. Avaliação do desempenho após a remoção.

Para a utilização do método de remoção do viés, faz-se necessário o ajuste de uma função cumulativa de probabilidade (FCP) aos dados de precipitação observada. Assim, foi realizado o teste de aderência de Kolmogorov-Smirnov para verificar a adequação de uma distribuição a série de dados. O teste foi realizado para as distribuições mais utilizadas em estudos hidrológicos a um nível de significância de 0,05 (nível usualmente utilizado na literatura). Dessa forma, as distribuições testadas foram: a) normal; b) gama; c) empírica de Weibull.

Conforme discutido no capítulo 5, a climatologia do modelo possui dez membros com frequência mensal, onde para essa análise foi realizada a média entre os membros para o ajuste da FCP. O cálculo da média dos membros foi efetuado de duas formas:

- a) Média entre os membros para cada data prevista correspondente;
- b) Os membros foram organizados de forma crescente e efetuada a média entre as ordens correspondentes.

O desempenho do modelo após a remoção do viés foi avaliado pelas medidas de desempenho probabilísticas, haja vista que os coeficientes utilizados para avaliação determinística não indicaram necessidade de correção aos dados.

Devido a fatores como a sazonalidade, estações do ano, limitações na formulação do MCG, etc., o modelo pode possuir dificuldade em representar alguns meses específicos. Assim, a remoção do viés também foi realizada da seguinte forma:

1. Seleccionam-se os dados da climatologia para cada mês do ano;
2. Seleccionam-se os dados da precipitação observada para cada mês do ano;
3. Ajusta-se uma FCP para os dados de precipitação observada em cada mês do ano;
4. Ajusta-se a mesma FCP do passo 3, para os dados da climatologia de cada mês do ano;
5. Remoção do viés mensal pelo método do Mapeamento Quantil-Quantil:
 - 5.1. Todos os meses;
 - 5.2. Meses específicos.

6.4 Considerações finais sobre o desenvolvimento do modelo

Para realização das regressões foram adotados alguns critérios para seleção das equações que foram elaboradas nesse estudo. Esses critérios foram adotados para limitar o número de equações, haja vista que existe um grande número de combinações possível para as variáveis selecionadas no capítulo anterior.

Como será visto nos resultados, todas as equações selecionadas possuem a variável nível anterior em sua composição. Isso se deu pelo fato do nível anterior possuir uma grande correlação linear com os níveis atuais, que são os produtos das previsões. As equações selecionadas obtiveram os melhores desempenhos em relação às outras que foram rejeitadas. Logo, as equações de regressão selecionadas para simular (intervalo de validação) e prever

(intervalo de simulação, ou previsão) os níveis no banhado, constituem as melhores equações possíveis para os dados selecionados para essa pesquisa (fixando-se a precipitação).

A correção do viés foi realizada utilizando algumas simplificações. No decorrer das análises, percebeu-se que o MCG possui dificuldade em representar os valores extremos observados e que a correção utilizando a climatologia do modelo possui uma tendência em superestimar os valores observados. Contudo, foram obtidos resultados satisfatórios para a remoção do viés mensal, realizado para os meses de março, abril, maio e dezembro. Sendo utilizados os dados sem remoção para os meses restantes.

6.5 Previsão de longo prazo de níveis

As previsões de níveis foram realizadas utilizando dados observados (evaporação, precipitação e níveis de água no banhado) e previstos (precipitação) para alimentar o modelo empírico. Assim, as previsões foram executadas de duas formas:

- Utilizando chuva perfeita;
- Utilizando chuva prevista.

As previsões utilizando chuva perfeita (observada) foi realizada em caráter determinístico, utilizando apenas dados observados, onde as variáveis explicativas estão dispostas em dois intervalos de tempos (passado e atual).

As previsões utilizando chuva prevista foi realizada em caráter probabilístico, utilizando os membros previstos de precipitação pelo modelo de circulação global COLA. As variáveis explicativas são compostas por dados observados (evaporação e níveis de água) e previstos (precipitação). Utilizando a média dos membros previstos de níveis de água, foram realizadas avaliações determinísticas para analisar o desempenho do modelo em paralelo as análises probabilísticas.

A metodologia utilizada para elaboração das regressões obteve como resultado equações que possuem dados em tempo atual (t) e dados no passado utilizados para explicar o futuro ($t-1$). As previsões foram realizadas para um horizonte máximo de sete meses de antecedência. Todas as equações que foram selecionadas possuem o nível anterior em sua composição. Isso significa dizer que o resultado da previsão do nível para o horizonte de um mês de antecedência, é a variável nível anterior para a previsão do segundo mês de antecedência e assim sucessivamente, como pode ser visto no exemplo abaixo realizado para os dados de precipitação prevista utilizando a equação de nº 33:

Previsão para o horizonte de um mês de antecedência (N_{t-1} , observado),

$$N_t = -0,04 + 0,002.P_t + 0,0007.MMPE_7_t + 0,97.N_{t-1}$$

Previsão para o horizonte de dois meses de antecedência (N_t , previsto),

$$N_{t+1} = -0,04 + 0,002.P_{t+1} + 0,0007.MMPE_7_t + 0,97.N_t$$

Previsão para o horizonte de três meses de antecedência (N_{t+1} , previsto),

$$N_{t+2} = -0,04 + 0,002.P_{t+2} + 0,0007.MMPE_7_{t+2} + 0,97.N_{t+1}$$

⋮

Previsão para o horizonte de sete meses de antecedência (N_{t+5} , previsto),

$$N_{t+6} = -0,04 + 0,002.P_{t+6} + 0,0007.MMPE_7_{t+6} + 0,97.N_{t+5}$$

A média móvel de sete meses do balanço entre a precipitação e evaporação, é composta por seis meses de dados no passado (observados) e um dado do mês atual (previsto), para o primeiro horizonte. Para o segundo horizonte de previsão são cinco dados no passado e dois previstos e, assim sucessivamente.

A metodologia utilizada para realização das médias móveis é a seguinte:

Seja “m” o horizonte de previsão e “n” a quantidade de dados das médias móveis, assim,

Para $n > m$, as médias móveis são realizadas pela equação 70:

$$MMS_{t+1} = \frac{\sum_{i=(t+1-n)}^{t-m} x_i + \sum_{j=1}^m x_{p_j}}{n} \quad (70)$$

Para $n \leq m$, as médias móveis são realizadas pela equação 71:

$$MMS_{t+1} = \frac{\sum_{j=1}^m x_{p_j}}{n} \quad (71)$$

onde x é a variável com dados observados no passado e x_p é a variável com dados previstos pelo modelo de circulação global ou precipitação perfeita (observado). Como exemplo, pode-se citar as médias móveis de sete meses do balanço entre a precipitação e evaporação:

Para o horizonte e um mês de antecedência ($m = 1$ e $n = 7$), assim,

$$x_i = P_{o_j} - E_i$$

$$x_{p_j} = P_{p_j} - E_j$$

$$MMS_{t+1} = \frac{\sum_{i=(t-6)}^{t-1} x_i + \sum_{j=1}^1 x_{p_j}}{7} \quad (72)$$

Para a suavização exponencial da precipitação foram utilizados apenas dados previstos. Para as médias móveis e suavização exponencial sem dados de precipitação, foram utilizados apenas dados observados.

A Figura 28, ilustra os passos seguidos para as previsões de níveis.

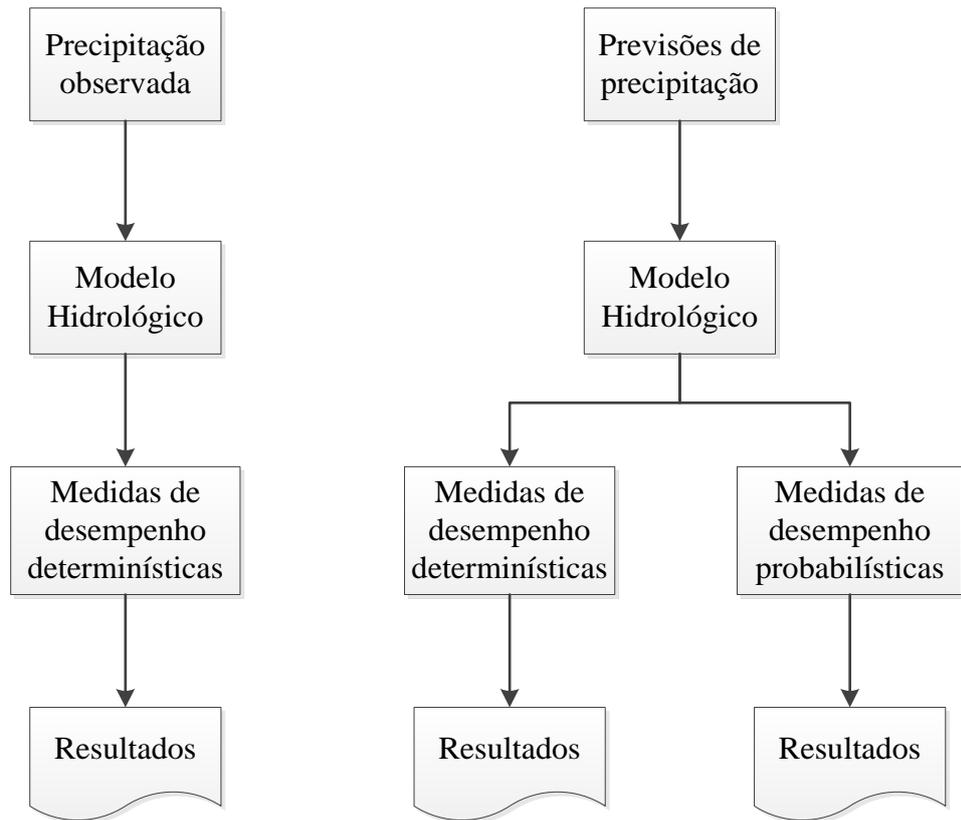


Figura 28 – Fluxograma das previsões de níveis.

7 RESULTADOS E DISCUSSÕES

7.1 Estatística das séries

7.1.1 Resultados estatísticos e ranking dos meses com maior disponibilidade de água

Utilizando os programas RStudio (versão 0.98.1028) e o R (versão 3.1.2), na versão para Windows, calculou-se algumas medidas de dispersão da série de dados, para examinar se a divisão proposta pela análise visual se justifica. Na Tabela 2 encontram-se as estatísticas dos dados disponíveis para a realização dessa pesquisa, com disponibilidade entre os anos de 1960 a 2011.

Tabela 2 - Estatísticas das séries de dados do banhado do Taim (1960 - 2011).

Estatística	Níveis	Precipitação	Evaporação	Função no R
Máximo	5,44	446,30	201,20	max() ou range()
Mínimo	1,21	0	11,60	min() ou range()
Média	2,80	102,49	75,59	mean()
Mediana	2,62	88,65	68,35	median()
Variância	0,75	4682,98	1913,28	var()
Desvio Padrão	0,87	68,43	43,74	sd()
1º Quartil	2,23	53,70	39,88	quantile()
2º Quartil	2,62	88,65	68,35	quantile()
3º Quartil	3,34	138,58	103,43	quantile()

Na Tabela 3 encontram-se as estatísticas das amostras, da série de níveis, propostas para a elaboração do modelo, podendo -se perceber que o intervalo selecionado para realização das previsões constitui um intervalo atípico, onde todos os dados encontram-se abaixo da média observada.

Tabela 3 - Estatística das amostras utilizadas para a modelagem empírica.

Estatística	Formulação	Validação	Simulação
Quantidade	336	192	96
Máximo	5,44	3,88	3,06
Mínimo	1,23	1,88	1,21
Média	3,08	2,71	2,03
Mediana	2,91	2,58	2,07
Variância	0,95	0,23	0,19
Desvio	0,98	0,48	0,44

Padrão			
1º Quartil	2,30	2,34	1,69
2º Quartil	2,91	2,58	2,07
3º Quartil	3,82	3,02	2,39

O resultado do ranking realizado para avaliar os meses com maior disponibilidade de água, mostrou que os meses de agosto a novembro obtiveram as melhores pontuações, ou seja, são os meses que possuem os níveis mais elevados durante o ano. Contudo, as retiradas para irrigação, como foi visto no tópico 6.1.3, são realizadas nos meses de dezembro a março. Isso seria problemático, pois não se aproveitaria o potencial hídrico do banhado de forma racional. No entanto, a série observada é fruto de medições, que é o resultado do balanço entre as entradas e saídas. Assim, para essa avaliação, faz-se necessário adicionar a série de retiradas para irrigação (o cálculo para as retiradas no banhado estão demonstrados no tópico 6.1.3) aos níveis observados. Refazendo o ranking com os níveis, somados às retiradas para irrigação, verifica-se um cenário diferente (Figura 29), onde os meses de janeiro e fevereiro adquirem as primeiras posições. Cabe ressaltar que esse procedimento constitui uma simplificação grosseira para uma primeira estimativa, pois como foi visto no tópico 6.1.3, o somatório da série de retirada para irrigação aos níveis, propagaria sua influência para os meses subsequentes.

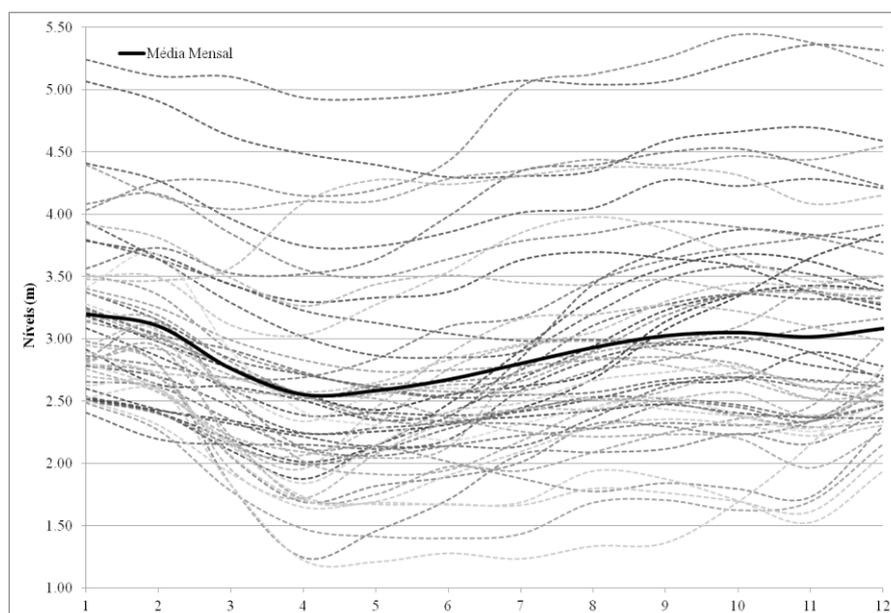


Figura 29 - Series de níveis anuais somadas com as retiradas em cada ano no banhado do Taim (1960 - 2011).

Na Tabela 4 podem-se verificar os resultados obtidos com o método adotado para o cálculo do ranking e o método de soma total dos níveis mensais, onde se pode observar que

alguns meses obtiveram posições diferentes para os métodos empregados. Contudo, os valores não diferem muito, em termos quantitativos. Os resultados mostram que os picos de retirada de água para irrigação ocorrem nos meses em que existe a maior disponibilidade hídrica, justificando as retiradas nesses respectivos meses.

Tabela 4 – Ranking dos meses com níveis mais elevados.

Meses	Soma das classificações	Ranking	Soma dos Níveis	Ranking
janeiro	176	1°	166,45	1°
fevereiro	230	2°	161,33	2°
março	379	8°	143,55	9°
abril	518	11°	132,73	12°
maio	520	12°	134,42	11°
junho	474	10°	139,04	10°
julho	393	9°	145,76	8°
agosto	303	7°	152,42	7°
setembro	249	4°	157,36	5°
outubro	243	3°	158,72	4°
novembro	303	7°	156,89	6°
dezembro	268	5°	160,39	3°

A análise por meio do ranking mostrou-se mais eficiente, pois a soma total dos níveis para cada mês é influenciada pelos anos de cheias, como também pelos anos de escassez.

7.1.2 Correlação entre as variáveis

Para uma primeira avaliação visual, os dados foram plotados em gráficos de linhas (Figura 30), com o objetivo de identificação de uma correlação ou padrão entre as séries de dados. Porém, as séries possuem unidades diferentes e para contornar esse problema os dados foram normalizados para os valores entre 0 e 1 (adimensionalização). Assim, pode-se visualizar o comportamento das séries sem levar em consideração as unidades, verificando quais séries possuem um comportamento semelhante.

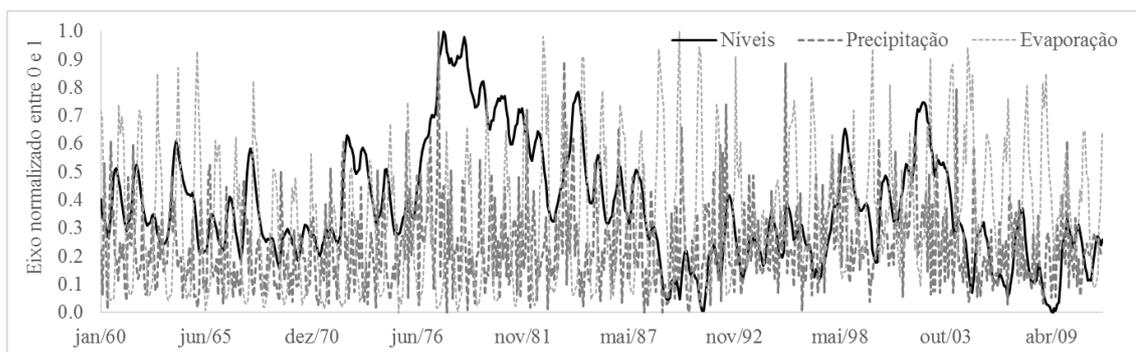


Figura 30 - Gráfico dos dados normalizados (1960 e 2011).

Conforme era esperado, a série de precipitação possui um comportamento aleatório, o que torna difícil a identificação de uma relação com as outras séries. A série de evaporação possui um comportamento mais homogêneo onde suas variações são similares ao longo dos anos.

Nas Figura 31 a Figura 33, encontram-se os diagramas de dispersão dos dados observados de níveis, precipitação e evaporação, para análise de correlação entre os intervalos de elaboração do modelo. Conforme se pode verificar nas referidas figuras, não existe uma aparente relação de causa e efeito entre os dados, onde os ajustes de linhas de tendência lineares, exponenciais, logarítmicas e polinomiais não se mostraram eficientes (R^2 inferiores a 0,035).

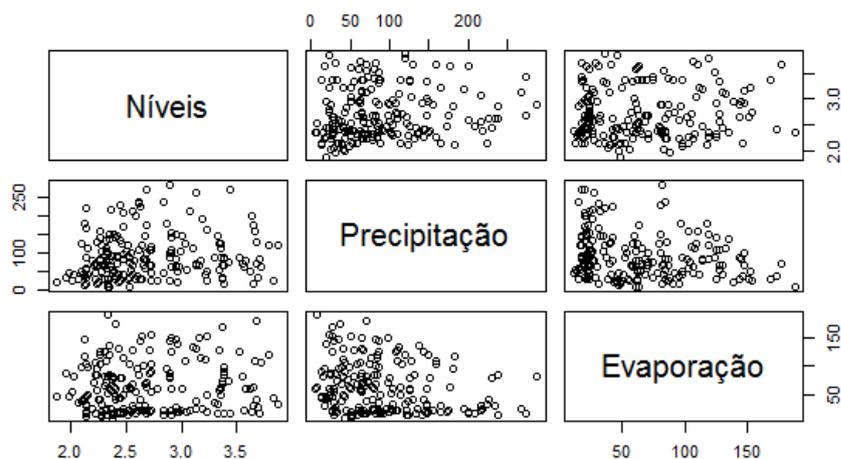


Figura 31- Diagrama de dispersão dos dados do banhado do Taim entre 1960 a 1975. Onde os níveis estão em metros (m), a precipitação e evaporação estão em milímetros (mm).

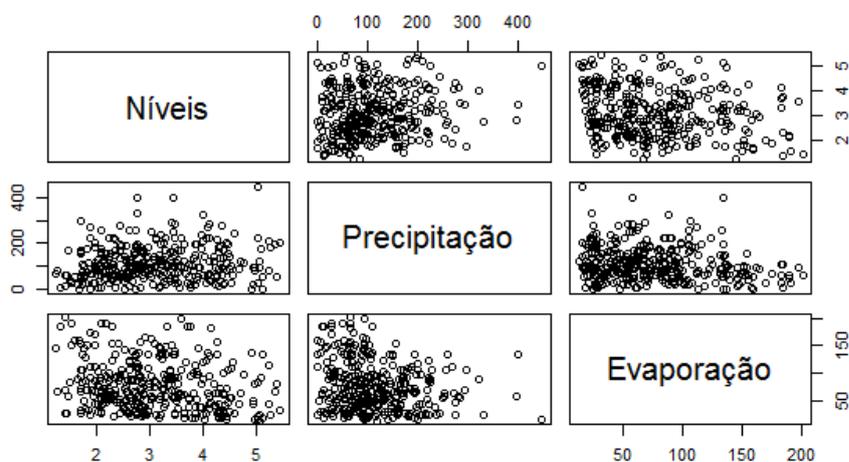


Figura 32 - Diagrama de dispersão dos dados do banhado do Taim entre 1976 a 2003. Onde os níveis estão em metros (m), a precipitação e evaporação estão em milímetros (mm).

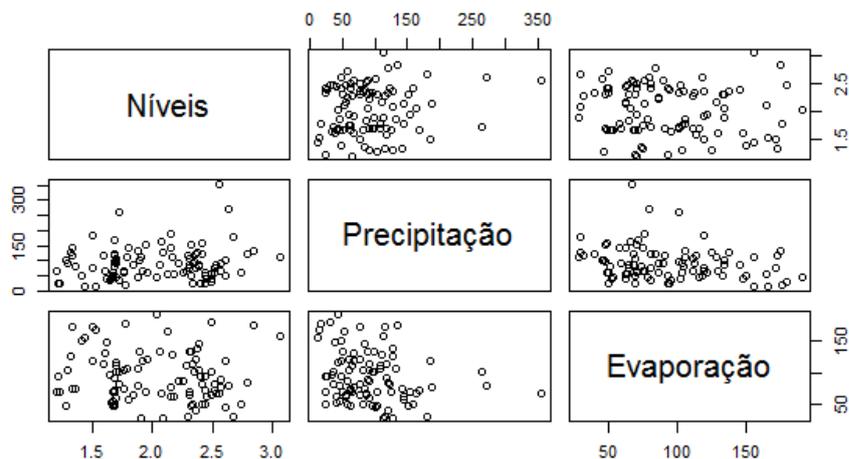


Figura 33 - Diagrama de dispersão dos dados do banhado do Taim entre 2004 a 2011. Onde os níveis estão em metros (m), a precipitação e evaporação estão em milímetros (mm).

Na Tabela 5 encontram-se os resultados dos cálculos dos coeficientes de correlação linear de Pearson e o coeficiente de correlação de Spearman para os referidos diagramas, onde se pode constatar valores relativamente baixos, indicando uma baixa correlação (quase nula). Entretanto, a ocorrência de uma baixa correlação, ou a inexistência de correlação, não significa que os dados não possuam uma relação de causa e efeito.

Tabela 5 - Coeficientes de correlação.

Método	Níveis x Precipitação			Níveis x Evaporação		
	1960-1975	1976-2003	2004-2011	1960-1975	1976-2003	2004-2011
Spearman	0,19	0,15	0,11	0,07	-0,18	-0,12
Pearson	0,14	0,15	0,15	0,08	-0,19	-0,10

A principal fonte de água para o sistema composto por lagoas, no qual está inserido o banhado do Taim, é a precipitação. As baixas correlações podem ser explicadas pelo alto potencial de retenção de água do complexo de lagoas e as trocas de água entre as mesmas. Isso se justifica pelas grandes dimensões do Sistema Hidrológico do Taim que possui uma área aproximada de 2254 km², onde a lagoa Mangueira (no qual encontra inserido o banhado) possui aproximadamente 820 km² de área superficial.

7.1.2.1 Decomposição das séries temporais

Nas Figura 34 a Figura 36 pode-se observar pela componente tendência-ciclo que as flutuações da série de níveis diminuíram, tornando mais fácil a visualização do comportamento geral da série, que é fruto da tendência e do ciclo percebidos em longo prazo.

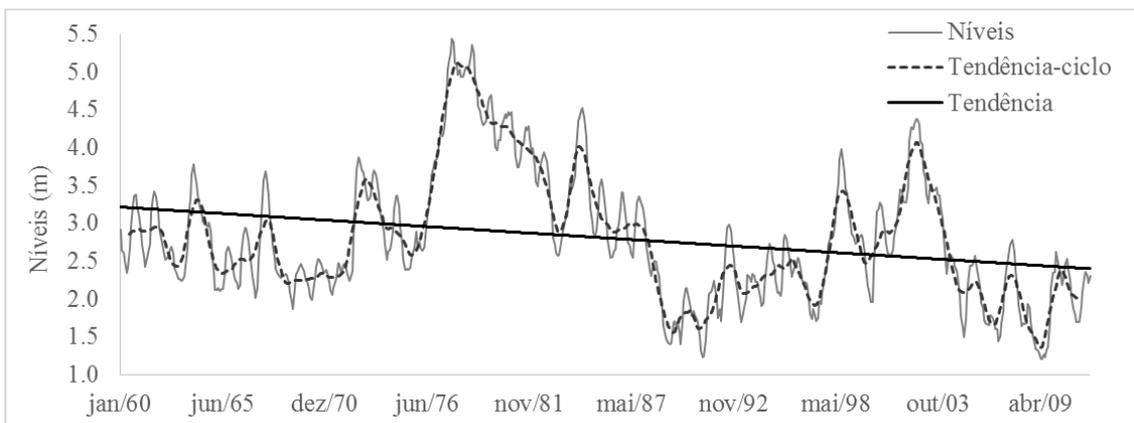


Figura 34 - Tendência-ciclo da série de níveis do banhado do Taim (1960 a 2011).

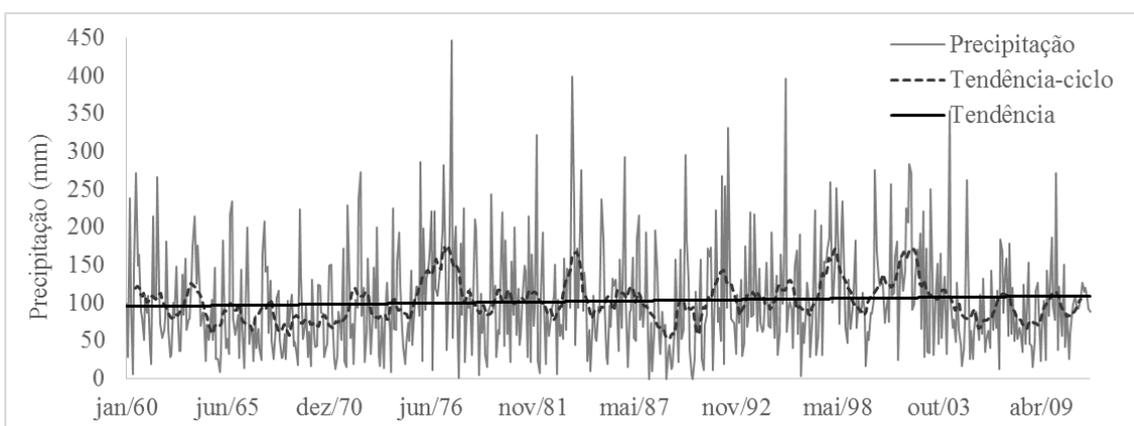


Figura 35 - Tendência-ciclo da série de precipitação do banhado do Taim (1960 a 2011).

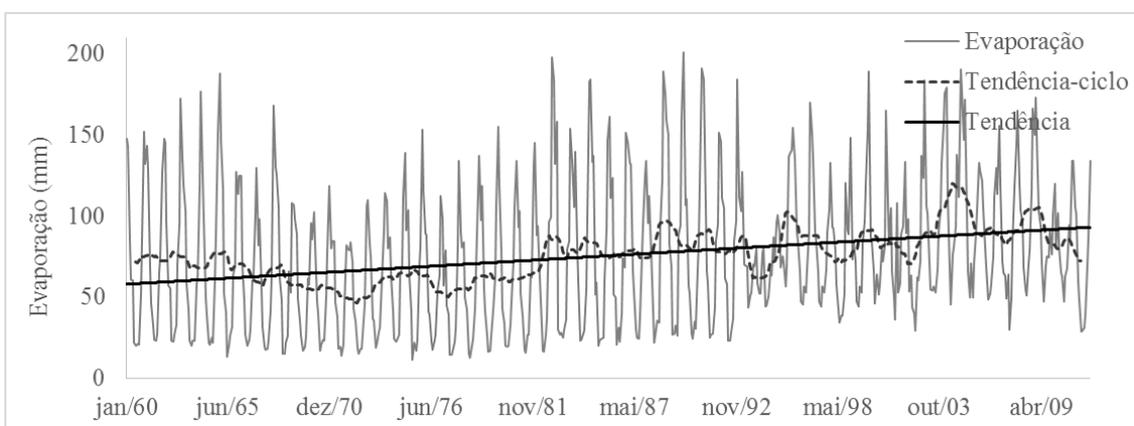


Figura 36 - Tendência-ciclo da série de evaporação do banhado do Taim (1960 a 2011).

As linhas de tendência plotada nos gráficos dos níveis e de precipitação (Figura 34 e Figura 35), acima referidos, não revelam muita informação sobre a tendência real das séries. Essa conclusão é motivada pela considerada perturbação dos valores da curva tendência-ciclo no entorno das linhas de tendência e pelo baixo valor dos coeficientes de determinação

(0,0826 e 0,0277, respectivamente). No entanto, a linha de tendência do gráfico da evaporação (Figura 36) demonstra uma possível adequação ao comportamento da série a longo prazo, no qual o coeficiente de determinação calculado (0,4539) manifesta uma adequação da linha de tendência à curva tendência-ciclo, revelando uma explicação aproximada de 45%.

Para refinar um pouco a busca pela tendência, as séries foram decompostas para cada intervalo de elaboração do modelo proposto. Na Figura 37, pode-se notar que houve uma quebra na linha de tendência, o que leva à conclusão que não existe uma tendência para a série de níveis (ao menos para o intervalo disponível). A mesma conclusão foi obtida para as séries de precipitação (Figura 38) e de evaporação (Figura 39), contrariando a conclusão anterior para a tendência da série de evaporação. Assim, não foi detectada uma tendência em longo prazo para as séries disponíveis.

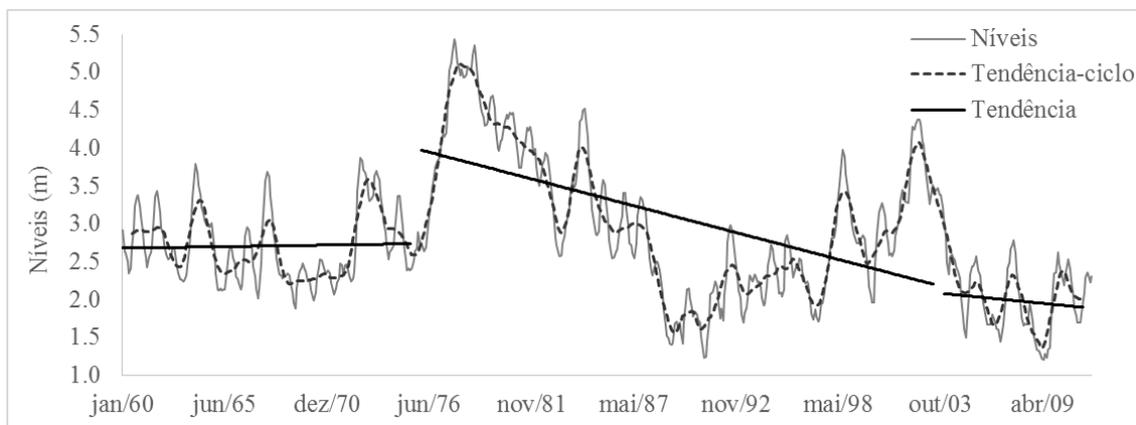


Figura 37 - Linhas de tendências dos níveis para cada intervalo da elaboração do modelo.

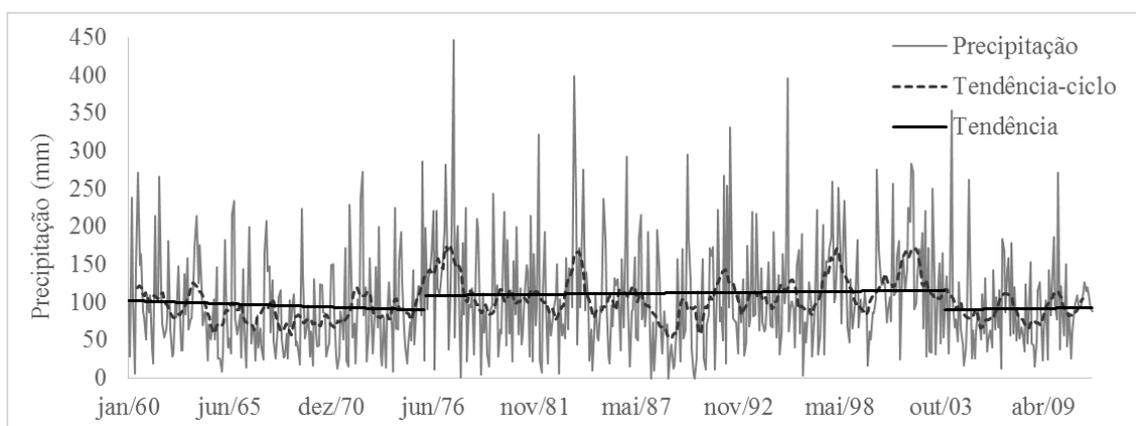


Figura 38 - Linhas de tendências da precipitação para cada intervalo da elaboração do modelo.

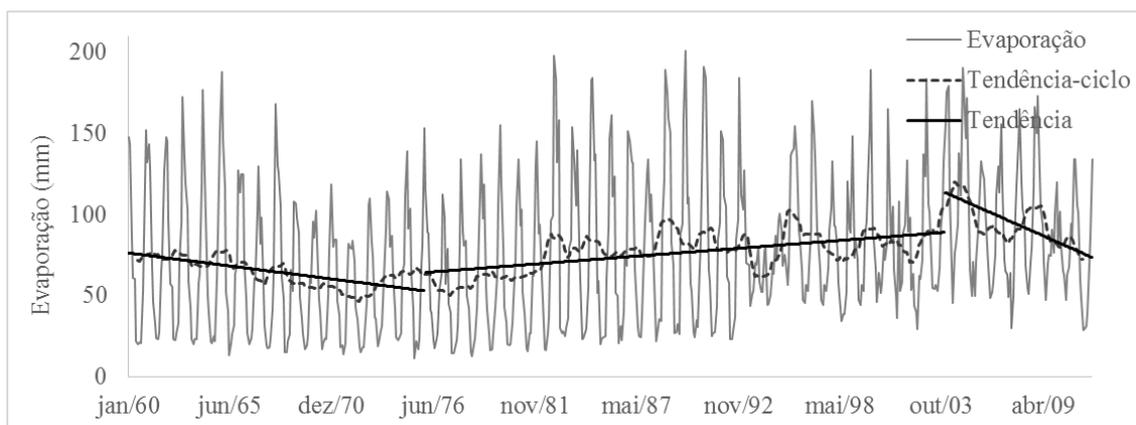


Figura 39 - Linhas de tendências da evaporação para cada intervalo da elaboração do modelo.

Como o objetivo desse estudo está pautado em previsões de longo prazo, a decomposição na componente tendência-ciclo é suficiente para as análises. Contudo, devido à baixa informação sobre as tendências e as relações de causa e efeito entre os dados, efetuou-se a fragmentação entre a tendência e o ciclo das séries de dados.

Da análise dos ciclos (Figura 40) não foi possível a identificação de padrões que pudessem sugerir relações de causa e efeito entre a série de evaporação com as demais séries analisadas. Entretanto, das séries de níveis e de precipitação percebeu-se uma semelhança entre os gráficos, na qual se pode perceber uma defasagem entre os picos de valores máximos variando em média entre 8 a 5 meses (Figura 41), evidenciando uma relação de causa e efeito entre as duas séries. O mesmo foi constatado em Tassi (2008) onde evidenciou-se que “não há coincidência entre os valores extremos de precipitação e de nível, o período de cheia ocorre aproximadamente três meses após o pico da precipitação”. No mesmo trabalho ficou constatado que o banhado do Taim tende a esvaziar-se mais rapidamente que encher, no entanto pode-se notar que as variações da precipitação são mais acentuadas que as ocorridas nos níveis.

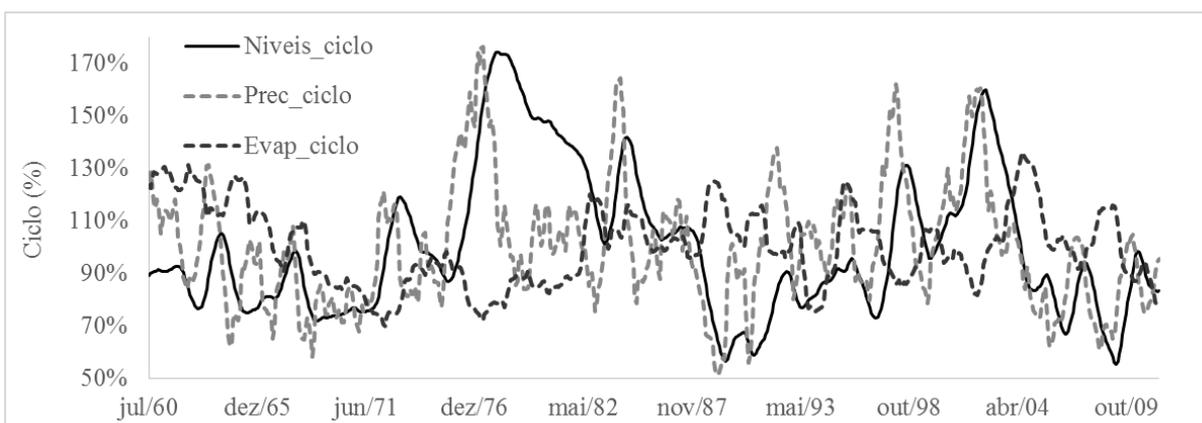


Figura 40 - Ciclo das séries de níveis, precipitação e evaporação (1960 a 2011).

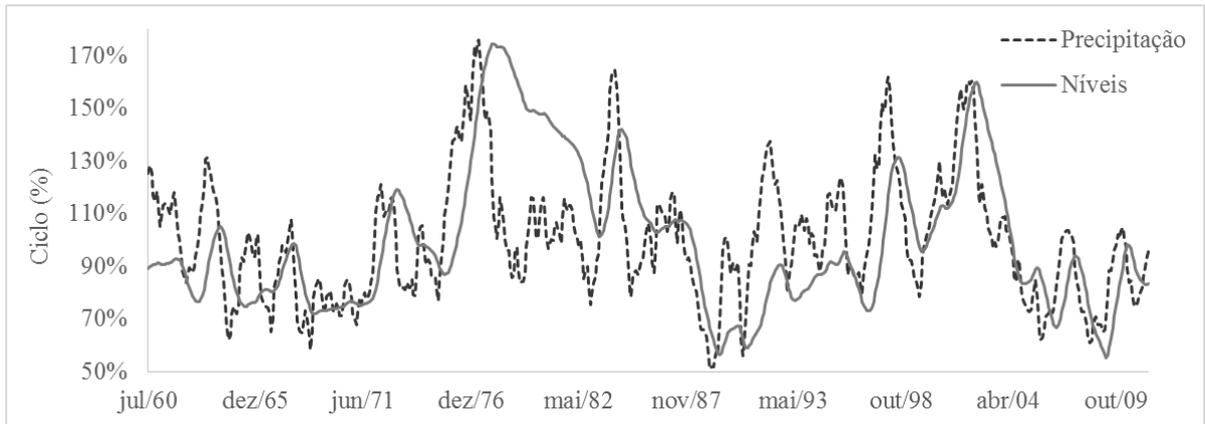


Figura 41 - Ciclo dos níveis e das precipitações no banhado.

7.1.2.2 Teste de correlação não linear

Na Figura 42 pode-se visualizar o resultado das correlações entre os níveis no banhado e o logaritmo neperiano da série de precipitação para diferentes intervalos, onde a linha no gráfico representa o ajuste logarítmico. Dessa análise pode-se concluir a inexistência de relação entre causa e efeito por meio de uma correlação logarítmica.

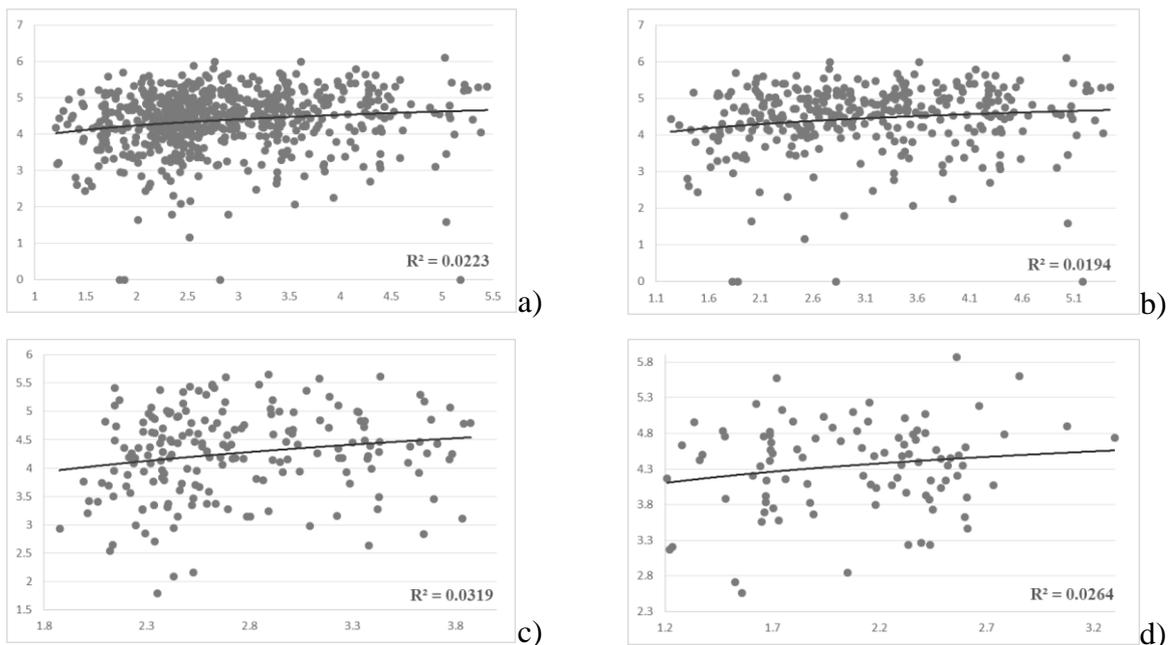


Figura 42 - Diagramas de dispersão entre os níveis (eixo horizontal em m) e o logaritmo natural da precipitação (eixo vertical em mm) para diferentes intervalos: a) série completa (1960 - 2011); b) intervalo de formulação (1976 - 2003); c) intervalo de verificação (1960 - 1975) e; d) intervalo de simulação (2004 - 2011).

A mesma conclusão foi obtida do cálculo do logaritmo natural da série de evaporação, onde foram obtidos resultados inferiores. Cabe ressaltar que foi realizado um ajuste nos dados

para o cálculo dos logaritmos, pois o logaritmo de zero não está definido, da mesma forma, não existe precipitação ou evaporação negativa. Assim todos os dados da série que possuíam valor inferior à unidade foram substituídos por um e calculados os logaritmos, obtendo assim o valor zero.

O teste de correlação para a função exponencial constitui uma operação inversa do logaritmo, onde esta opção calcula o exponencial de todas as observações das séries. Os resultados foram pouco satisfatórios, obtendo resultados inferiores a 0,002 do coeficiente de determinação (R^2). Não sendo possível a visualização de uma correlação por inspeção visual dos gráficos, na qual a nuvem de pontos não se aproximava das linhas de ajustes testadas, nem ficou evidente a formação de padrões visuais.

O teste potencial foi realizado da seguinte forma: a) elevaram-se todos os valores das séries a uma potência de valor dois; b) plotou-se gráficos de dispersão entre os níveis no banhado e o resultado do passo “a”, onde foram ajustadas linhas (ajuste polinomial, logaritmo, exponencial, linear) para calcular o coeficiente de determinação com base em cada ajuste; c) adicionou-se um ao valor da potência e refeitos os passos “a” e “b”. Os resultados obtidos não foram satisfatórios, onde os valores dos coeficientes calculados ficaram abaixo de 0,035. Ficou evidente que a medida que se aumentava o valor da potência, os resultados reduziam o valor do coeficiente de determinação, onde o efeito dessa redução era visível na inspeção dos gráficos.

O teste da raiz foi realizado de forma semelhante ao teste potencial, no qual os índices das raízes é que variavam (raiz quadrada, cúbica, etc.). Os resultados não foram satisfatórios, obtendo valores do coeficiente de determinação inferiores a 0,032 e a inspeção dos gráficos não revelaram padrões de correlação.

Cabe ressaltar que os testes realizados neste tópico não obtiveram valores satisfatórios para os dados utilizados nesta pesquisa. Contudo, esta metodologia é bastante eficiente e utilizada para encontrar padrões entre séries temporais. Conforme foi discutido no tópico anterior, existe uma relação de causa de efeito entre os dados, porém ocorrem fatores que influenciam de forma direta na determinação dos níveis no banhado que não podem ser explicados de forma empírica.

7.1.2.3 Defasagem entre as séries

No tópico 7.1.2.1, ficou constatada uma defasagem entre os picos dos valores máximos dos ciclos ao redor da tendência das séries de níveis e precipitação. Percebeu-se um

descompasso variando em média entre 5 a 8 meses. Assim, ficou clara uma correlação entre os dados, fazendo-se necessário compreender melhor qual a distância no tempo em que os fenômenos se correlacionam. Por esse motivo, avaliou-se as correlações dos dados em relação aos níveis, deslocando no tempo as séries temporais.

Na Tabela 6 encontram-se os valores obtidos do coeficiente de correlação linear de Pearson das séries analisadas com os níveis no banhado do Taim. Na referida tabela pode-se verificar que foram realizados alguns cálculos envolvendo as séries disponíveis, com o intuito de analisar o comportamento dessas interações. Pode-se ressaltar a coluna dos níveis ponderados pelas precipitações que obtiveram bons resultados. Esse dado pode ajudar a prever os meses futuros, ponderando-se o nível e a precipitação atual, para explicar o comportamento dos níveis nos meses subsequentes. Os resultados obtidos que envolvem a precipitação e a evaporação, sem a influência dos níveis, não foram satisfatórios. Contudo, as análises realizadas buscaram explicar as interações mensais, sem avaliar a propagação dessas influências ao longo do tempo.

Tabela 6 - Correlação entre as séries de dados e os níveis no banhado.

	P *	E *	P-E *	I *	P-E-I *	P+N *	P*N *	P-N *	P/N *	Níveis
mês a mês	0,15	-0,19	0,21	-0,13	0,22	0,16	0,49	0,13	-0,24	1,00
1 mês	0,19	-0,27	0,28	-0,18	0,30	0,20	0,52	0,18	-0,21	0,99
2 meses	0,21	-0,33	0,32	-0,20	0,34	0,22	0,52	0,20	-0,18	0,95
3 meses	0,22	-0,35	0,34	-0,19	0,35	0,23	0,52	0,21	-0,16	0,91
4 meses	0,22	-0,32	0,33	-0,14	0,32	0,23	0,50	0,21	-0,16	0,87
5 meses	0,21	-0,26	0,29	-0,07	0,26	0,22	0,48	0,20	-0,15	0,83
6 meses	0,21	-0,18	0,25	0,00	0,20	0,22	0,46	0,20	-0,15	0,80
7 meses	0,20	-0,11	0,21	0,06	0,15	0,21	0,45	0,19	-0,15	0,78
8 meses	0,19	-0,06	0,18	0,08	0,11	0,20	0,44	0,18	-0,15	0,77
9 meses	0,19	-0,04	0,17	0,07	0,11	0,20	0,44	0,18	-0,15	0,76
10 meses	0,18	-0,06	0,17	0,02	0,13	0,19	0,42	0,17	-0,16	0,75
11 meses	0,17	-0,12	0,19	-0,05	0,17	0,18	0,41	0,16	-0,16	0,74
12 meses	0,16	-0,19	0,22	-0,13	0,23	0,17	0,40	0,15	-0,15	0,72
Máximo	0,22	-0,04	0,34	0,08	0,35	0,23	0,52	0,21	-0,15	1,00
Mínimo	0,15	-0,35	0,17	-0,20	0,11	0,16	0,40	0,13	-0,24	0,72

*Abreviações: P (precipitação); E (evaporação); I (irrigação) e; N (níveis).

7.1.2.4 Teste com médias móveis

Para os testes realizados com as médias móveis, os resultados utilizando o coeficiente de correlação linear de Pearson e o coeficiente de Spearman obtiveram valores muito próximos (variando entre a segunda e a terceira casa decimal), assim, para evitar a duplicidade de informação, as tabelas disponibilizadas nesse tópico conterão apenas o

coeficiente de correlação linear de Pearson. Os resultados entre os coeficientes de correlação das médias móveis aritméticas e exponenciais obtiveram valores muito próximos (variando entre a segunda e a terceira casa decimal), não sendo possível a identificação de qual dos métodos adicionaria um maior benefício para as previsões. Assim, optou-se pelo método das médias móveis aritméticas, pela maior simplicidade na realização dos cálculos.

Nas Tabela 7 a Tabela 10 encontram-se os valores dos coeficientes de Pearson calculados para a correlação entre as séries de níveis e de precipitação.

Tabela 7 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das precipitações (1960 à 2011).

Nº de meses utilizado para as médias móveis											
Horizonte	2	3	4	5	6	7	8	9	10	11	12
mês a mês	0,277	0,359	0,427	0,483	0,525	0,555	0,578	0,594	0,605	0,614	0,621
1º mês	0,329	0,408	0,468	0,514	0,550	0,572	0,588	0,599	0,604	0,612	0,620
2º mês	0,356	0,428	0,478	0,520	0,550	0,567	0,579	0,584	0,590	0,599	0,608
3º mês	0,364	0,427	0,474	0,510	0,535	0,548	0,556	0,562	0,570	0,580	0,590
4º mês	0,352	0,414	0,455	0,487	0,509	0,519	0,528	0,537	0,546	0,558	0,567
5º mês	0,339	0,394	0,432	0,460	0,478	0,489	0,501	0,512	0,523	0,535	0,543
6º mês	0,322	0,374	0,407	0,430	0,450	0,464	0,479	0,491	0,502	0,513	0,519
7º mês	0,301	0,348	0,377	0,402	0,425	0,442	0,458	0,471	0,480	0,489	0,491

Tabela 8 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das precipitações (1976 à 2003).

Nº de meses utilizado para as médias móveis											
Horizonte	2	3	4	5	6	7	8	9	10	11	12
mês a mês	0,231	0,300	0,358	0,406	0,444	0,474	0,499	0,516	0,529	0,542	0,553
1º mês	0,273	0,341	0,394	0,433	0,467	0,493	0,513	0,527	0,536	0,546	0,553
2º mês	0,295	0,360	0,406	0,443	0,473	0,494	0,512	0,522	0,528	0,535	0,544
3º mês	0,306	0,364	0,408	0,441	0,467	0,487	0,501	0,509	0,513	0,521	0,530
4º mês	0,301	0,359	0,400	0,430	0,455	0,471	0,483	0,489	0,494	0,504	0,513
5º mês	0,294	0,348	0,387	0,416	0,438	0,451	0,462	0,469	0,476	0,485	0,493
6º mês	0,288	0,338	0,375	0,401	0,420	0,431	0,443	0,452	0,459	0,467	0,473
7º mês	0,276	0,326	0,358	0,381	0,398	0,411	0,425	0,434	0,441	0,447	0,449

Tabela 9 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das precipitações (1960 à 1975).

Nº de meses utilizado para as médias móveis											
Horizonte	2	3	4	5	6	7	8	9	10	11	12
mês a mês	0,286	0,348	0,419	0,489	0,557	0,623	0,672	0,709	0,736	0,754	0,767
1º mês	0,341	0,402	0,470	0,531	0,589	0,648	0,690	0,722	0,740	0,756	0,770
2º mês	0,375	0,435	0,493	0,547	0,598	0,651	0,688	0,712	0,728	0,745	0,766

3º mês	0,395	0,446	0,499	0,545	0,592	0,639	0,667	0,687	0,705	0,731	0,760
4º mês	0,388	0,435	0,480	0,522	0,564	0,603	0,630	0,654	0,681	0,716	0,753
5º mês	0,370	0,408	0,452	0,488	0,522	0,560	0,592	0,626	0,664	0,707	0,744
6º mês	0,342	0,378	0,416	0,445	0,478	0,520	0,564	0,610	0,657	0,701	0,732
7º mês	0,313	0,343	0,375	0,404	0,446	0,500	0,557	0,612	0,660	0,697	0,714

Tabela 10 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das precipitações (2004 à 2011).

Nº de meses utilizado para as médias móveis											
Horizonte	2	3	4	5	6	7	8	9	10	11	12
mês a mês	0,251	0,354	0,463	0,558	0,621	0,667	0,705	0,735	0,756	0,757	0,732
1º mês	0,345	0,458	0,540	0,632	0,686	0,710	0,730	0,743	0,736	0,723	0,700
2º mês	0,412	0,503	0,582	0,669	0,702	0,708	0,711	0,696	0,675	0,664	0,630
3º mês	0,424	0,513	0,588	0,652	0,668	0,658	0,634	0,606	0,588	0,565	0,535
4º mês	0,409	0,496	0,548	0,594	0,594	0,558	0,523	0,501	0,472	0,452	0,440
5º mês	0,405	0,464	0,497	0,521	0,493	0,449	0,421	0,389	0,363	0,362	0,357
6º mês	0,363	0,403	0,417	0,410	0,372	0,338	0,303	0,277	0,270	0,273	0,269
7º mês	0,296	0,317	0,302	0,280	0,253	0,215	0,188	0,184	0,184	0,186	0,170

Das Tabela 7 e Tabela 8, pode-se observar uma semelhança entre os valores calculados para os coeficientes de correlação, concluindo-se que as médias móveis da precipitação, calculadas para o intervalo de formulação, possuem uma correlação similar à correlação para toda a série. Na Tabela 9, consegue-se verificar que as correlações para o intervalo de validação obtiveram melhores resultados, comparados com os outros intervalos propostos. Na Tabela 10, pode-se observar que, para o sexto e o sétimo mês no passado, as correlações obtiveram valores relativamente baixo, se comparados com os outros horizontes de previsão.

Com o intuito de aproveitar ao máximo as oscilações fornecidas pela precipitação prevista, buscou-se utilizar o número mínimo de dados para as médias móveis. Assim, com base nas tabelas de correlação, percebeu-se que a partir da média móvel de cinco meses, o coeficiente era de aproximadamente 0,4. Logo, as médias móveis de cinco meses da precipitação foram selecionadas para compor os dados que serão utilizados nesse estudo.

Os resultados das correlações entre os níveis e a evaporação estão dispostos nas Tabela 11 a Tabela 14, onde os critérios para avaliação das correlações foram similares aos da precipitação. Contudo, devido ao fato não existirem dados de evaporação prevista, as médias móveis poderiam possuir um número maior de meses que não prejudicariam as conclusões.

Tabela 11 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das evaporações (1960 à 2011).

Nº de meses utilizado para as médias móveis											
Horizonte	2	3	4	5	6	7	8	9	10	11	12
mês a mês	-0,16	-0,22	-0,28	-0,34	-0,39	-0,43	-0,47	-0,50	-0,52	-0,52	-0,51
1º mês	-0,25	-0,31	-0,35	-0,39	-0,42	-0,44	-0,46	-0,48	-0,49	-0,51	-0,51
2º mês	-0,33	-0,36	-0,38	-0,40	-0,40	-0,41	-0,42	-0,43	-0,46	-0,48	-0,51
3º mês	-0,36	-0,37	-0,36	-0,36	-0,35	-0,35	-0,35	-0,38	-0,42	-0,46	-0,51
4º mês	-0,34	-0,32	-0,30	-0,29	-0,27	-0,27	-0,29	-0,32	-0,38	-0,45	-0,50
5º mês	-0,28	-0,25	-0,22	-0,20	-0,19	-0,20	-0,23	-0,29	-0,36	-0,44	-0,50
6º mês	-0,19	-0,16	-0,14	-0,13	-0,13	-0,16	-0,21	-0,28	-0,37	-0,45	-0,49
7º mês	-0,11	-0,09	-0,07	-0,08	-0,11	-0,15	-0,22	-0,30	-0,39	-0,46	-0,49

Tabela 12 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das evaporações (1976 à 2003).

Nº de meses utilizado para as médias móveis											
Horizonte	2	3	4	5	6	7	8	9	10	11	12
mês a mês	-0,20	-0,25	-0,31	-0,37	-0,42	-0,48	-0,53	-0,59	-0,65	-0,70	-0,71
1º mês	-0,28	-0,32	-0,37	-0,41	-0,44	-0,48	-0,52	-0,57	-0,63	-0,68	-0,71
2º mês	-0,34	-0,37	-0,39	-0,41	-0,43	-0,45	-0,48	-0,52	-0,58	-0,65	-0,70
3º mês	-0,36	-0,37	-0,37	-0,37	-0,38	-0,39	-0,42	-0,47	-0,54	-0,63	-0,70
4º mês	-0,34	-0,33	-0,32	-0,31	-0,31	-0,32	-0,35	-0,41	-0,50	-0,61	-0,69
5º mês	-0,28	-0,26	-0,24	-0,23	-0,23	-0,26	-0,30	-0,38	-0,48	-0,60	-0,68
6º mês	-0,21	-0,18	-0,17	-0,17	-0,18	-0,22	-0,28	-0,37	-0,49	-0,60	-0,67
7º mês	-0,14	-0,12	-0,12	-0,13	-0,16	-0,21	-0,29	-0,39	-0,51	-0,61	-0,65

Tabela 13 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das evaporações (1960 à 1975).

Nº de meses utilizado para as médias móveis											
Horizonte	2	3	4	5	6	7	8	9	10	11	12
mês a mês	0,03	-0,08	-0,19	-0,29	-0,36	-0,41	-0,43	-0,41	-0,35	-0,25	-0,15
1º mês	-0,19	-0,29	-0,37	-0,42	-0,44	-0,44	-0,41	-0,36	-0,28	-0,22	-0,18
2º mês	-0,36	-0,42	-0,45	-0,44	-0,42	-0,37	-0,30	-0,23	-0,17	-0,15	-0,19
3º mês	-0,44	-0,44	-0,41	-0,36	-0,29	-0,21	-0,14	-0,07	-0,05	-0,10	-0,21
4º mês	-0,41	-0,36	-0,28	-0,20	-0,11	-0,02	0,04	0,08	0,05	-0,05	-0,21
5º mês	-0,27	-0,18	-0,09	0,01	0,09	0,15	0,19	0,17	0,10	-0,05	-0,21
6º mês	-0,08	0,02	0,11	0,19	0,24	0,26	0,25	0,19	0,07	-0,08	-0,20
7º mês	0,13	0,21	0,27	0,30	0,31	0,28	0,22	0,12	-0,01	-0,13	-0,18

Tabela 14 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas das evaporações (2004 à 2011).

Nº de meses utilizado para as médias móveis											
Horizonte	2	3	4	5	6	7	8	9	10	11	12
mês a mês	-0,09	-0,21	-0,31	-0,36	-0,42	-0,45	-0,46	-0,45	-0,42	-0,37	-0,32
1º mês	-0,25	-0,33	-0,37	-0,42	-0,44	-0,44	-0,42	-0,39	-0,35	-0,31	-0,27
2º mês	-0,34	-0,36	-0,39	-0,40	-0,39	-0,36	-0,32	-0,28	-0,24	-0,22	-0,20
3º mês	-0,33	-0,34	-0,33	-0,31	-0,27	-0,22	-0,17	-0,13	-0,11	-0,10	-0,11
4º mês	-0,27	-0,25	-0,21	-0,17	-0,12	-0,06	-0,01	0,02	0,02	0,01	0,00
5º mês	-0,16	-0,11	-0,07	-0,01	0,05	0,09	0,12	0,13	0,12	0,11	0,09
6º mês	-0,02	0,04	0,09	0,14	0,17	0,20	0,21	0,21	0,20	0,17	0,14
7º mês	0,12	0,17	0,21	0,23	0,25	0,25	0,24	0,23	0,22	0,19	0,16

Os coeficientes de correlação calculados para os intervalos de validação (Tabela 13) e simulação (Tabela 14) obtiveram valores muito baixos (-0,09 para média móvel de cinco meses para o quinto mês do horizonte de previsão). Assim, foi escolhida a média móvel de dez meses para evaporação, onde o critério para escolha foi a análise das correlações das médias móveis para o período de formulação (Tabela 12).

Os dados de evaporação e as médias móveis da evaporação obtiveram baixa correlação com os níveis. E como consequência, na elaboração do modelo, sua influência pode ser muito baixa, ou irrelevante. Assim, visando explorar melhor as informações da evaporação, foram realizadas médias móveis do balanço entre a precipitação e a evaporação (P-E). Os resultados estão dispostos nas Tabela 15 a Tabela 18.

Tabela 15 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas do balanço entre precipitação e evaporações (1960 à 2011).

Nº de meses utilizado para as médias móveis											
Horizonte	2	3	4	5	6	7	8	9	10	11	12
mês a mês	0,302	0,385	0,462	0,528	0,587	0,638	0,678	0,706	0,722	0,732	0,736
1º mês	0,386	0,464	0,527	0,575	0,619	0,655	0,683	0,702	0,713	0,727	0,739
2º mês	0,436	0,501	0,547	0,582	0,612	0,638	0,657	0,671	0,687	0,709	0,732
3º mês	0,451	0,499	0,532	0,553	0,574	0,593	0,609	0,628	0,654	0,687	0,720
4º mês	0,431	0,466	0,486	0,499	0,514	0,532	0,554	0,583	0,620	0,664	0,703
5º mês	0,390	0,411	0,424	0,432	0,447	0,472	0,506	0,548	0,595	0,646	0,683
6º mês	0,335	0,347	0,356	0,366	0,391	0,429	0,476	0,530	0,584	0,632	0,661
7º mês	0,277	0,285	0,297	0,319	0,358	0,411	0,470	0,529	0,581	0,620	0,635

Tabela 16 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas do balanço entre precipitação e evaporações (1976 à 2003).

Nº de meses utilizado para as médias móveis											
Horizonte	2	3	4	5	6	7	8	9	10	11	12
mês a mês	0,277	0,352	0,421	0,482	0,539	0,589	0,626	0,651	0,667	0,677	0,685

1º mês	0,344	0,416	0,477	0,520	0,566	0,604	0,631	0,651	0,663	0,676	0,688
2º mês	0,381	0,446	0,493	0,527	0,562	0,592	0,614	0,631	0,645	0,662	0,682
3º mês	0,394	0,445	0,484	0,508	0,535	0,560	0,580	0,600	0,620	0,645	0,672
4º mês	0,378	0,421	0,451	0,468	0,490	0,515	0,538	0,564	0,593	0,626	0,657
5º mês	0,349	0,381	0,404	0,417	0,440	0,468	0,499	0,535	0,573	0,610	0,639
6º mês	0,308	0,333	0,353	0,367	0,395	0,432	0,474	0,520	0,563	0,598	0,619
7º mês	0,265	0,286	0,308	0,329	0,367	0,415	0,468	0,519	0,559	0,585	0,593

Tabela 17 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas do balanço entre precipitação e evaporações (1960 à 1975).

Nº de meses utilizado para as médias móveis											
Horizonte	2	3	4	5	6	7	8	9	10	11	12
mês a mês	0,191	0,317	0,447	0,547	0,629	0,697	0,751	0,780	0,784	0,765	0,731
1º mês	0,400	0,510	0,601	0,662	0,706	0,732	0,744	0,733	0,716	0,715	0,727
2º mês	0,529	0,602	0,648	0,671	0,675	0,667	0,640	0,607	0,599	0,635	0,690
3º mês	0,559	0,587	0,591	0,577	0,552	0,516	0,474	0,450	0,471	0,546	0,639
4º mês	0,497	0,486	0,455	0,419	0,373	0,330	0,303	0,306	0,356	0,463	0,574
5º mês	0,376	0,330	0,280	0,231	0,188	0,164	0,167	0,201	0,279	0,401	0,509
6º mês	0,219	0,154	0,097	0,060	0,044	0,053	0,091	0,154	0,248	0,366	0,444
7º mês	0,060	-0,01	-0,05	-0,05	-0,03	0,013	0,079	0,159	0,253	0,344	0,379

Tabela 18 - Análise de correlação linear de Pearson entre os níveis e as médias móveis aritméticas do balanço entre precipitação e evaporações (2004 à 2011).

Nº de meses utilizado para as médias móveis											
Horizonte	2	3	4	5	6	7	8	9	10	11	12
mês a mês	0,255	0,405	0,507	0,572	0,614	0,643	0,653	0,644	0,642	0,646	0,653
1º mês	0,397	0,524	0,569	0,619	0,648	0,642	0,642	0,620	0,605	0,602	0,595
2º mês	0,479	0,554	0,589	0,629	0,623	0,605	0,589	0,556	0,535	0,525	0,506
3º mês	0,475	0,542	0,573	0,575	0,553	0,515	0,479	0,447	0,427	0,409	0,388
4º mês	0,447	0,503	0,495	0,479	0,431	0,364	0,329	0,308	0,288	0,277	0,261
5º mês	0,410	0,410	0,386	0,343	0,259	0,208	0,185	0,169	0,159	0,155	0,138
6º mês	0,316	0,281	0,235	0,157	0,088	0,076	0,063	0,051	0,047	0,040	0,036
7º mês	0,200	0,128	0,054	-0,01	-0,03	-0,02	-0,02	-0,03	-0,04	-0,04	-0,05

Da Tabela 15 pode-se perceber que a partir da média móvel de sete meses os valores foram melhores que os encontrados para as outras séries. A mesma conclusão foi obtida da análise das correlações para o período de formulação (Tabela 16). Contudo, para os intervalos de validação (Tabela 17) e simulação (Tabela 18), percebe-se que a influência dos baixos valores da evaporação prejudicou as correlações a partir do horizonte do quarto mês de

previsão. Assim, foi escolhida a média móvel de sete meses para o balanço entre a precipitação e evaporação (P-E).

Foram calculadas as médias móveis das séries resultantes dos testes realizados no tópico 6.1.2.2, onde os resultados não foram satisfatórios, obtendo baixos valores de correlação e gráficos com linhas de ajustes (linear, exponencial, logaritmo, polinomial) que obtiveram coeficientes de determinação (R^2) com baixos valores (inferiores a 0,1). Um outro teste realizado foi o método de acumulação dos dados, contudo os valores de correlação e os gráficos foram os mesmos obtidos com os das médias móveis, sendo este último escolhido, pois a dimensão dos resultados era mais próxima aos da variável resposta.

7.1.2.5 Teste com suavização exponencial

Na Figura 43 pode-se verificar os resultados obtidos com o método da suavização exponencial. É fácil notar um ajuste linear entre os dados, melhorando de forma considerável as correlações entre os níveis e as variáveis independentes.

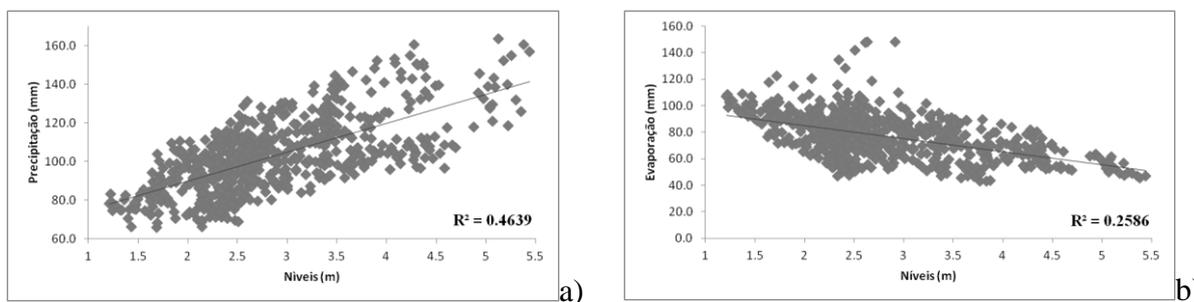


Figura 43 - Alisamento exponencial das séries: a) precipitação e b) evaporação.

A suavização para as séries resultantes do tópico 6.1.2.2 não mostraram melhoras significativas comparadas às séries originais. Para evitar a duplicidade da informação, optou-se por descartá-las. Assim, a suavização foi utilizada nas séries de precipitação e de evaporação.

7.1.3 Testes não paramétricos

Nas Tabela 19 a Tabela 22 encontram-se os resultados obtidos para os testes de hipóteses não paramétricos. A hipótese de aleatoriedade foi realizada para verificar se as flutuações ocorridas nas séries são decorrentes de causas naturais. Das séries analisadas apenas a precipitação não foi rejeitada a hipótese de aleatoriedade, homogeneidade e estacionariedade para todos os intervalos avaliados. Esse resultado revela que a série possui um comportamento semelhante para todos os intervalos avaliados. A hipótese de

independência foi aceita apenas para a série de precipitação nos intervalos de formulação e simulação, onde a rejeição ocorrida no intervalo de validação é um indício de que a ocorrência de algum evento pode ter sido influenciada pela ocorrência de outro.

A série de evaporação da região possui uma baixa variação (se comparada com a precipitação) e não uniforme, sendo influenciada pelas estações do ano, o que explicaria a rejeição de todos os testes realizados, onde apenas para o intervalo de simulação foram aceitas as hipóteses de homogeneidade e estacionariedade. A série de níveis é influenciada pelo conjunto de lagoas e trocas de águas subterrâneas, tornando o processo de formação dos níveis complexo, onde apenas as hipóteses de homogeneidade e estacionariedade foram aceitas para o intervalo de validação e estacionariedade para o intervalo de simulação. Com exceção das séries de níveis, precipitação e evaporação, as outras séries são frutos de cálculos realizados das séries observadas, onde os métodos utilizados para geração dos dados agem como um filtro, alterando as premissas que ocasionalmente possam existir nos dados observados.

Tabela 19 - Testes de hipóteses não paramétricos (1960 a 2011).

	Aleatoriedade	Independência	Homogeneidade	Estacionariedade
Níveis				
Precipitação	X		X	X
Evaporação				
P-E*				X
MMP_5*				
MME_10*				
MMP-E_7*				X
Precip_Sua*				
Evap_Sua*				
Nív_Ant*				
Irrigação				

*P-E (precipitação menos evaporação); MMP_5 (média móvel de cinco meses da precipitação); MME_10 (média móvel de dez meses da evaporação); MMP-E_7 (média móvel de sete meses da precipitação menos evaporação); Precip_Sua (precipitação suavizada exponencialmente); Evap_Sua (evaporação suavizada exponencialmente); Nív_ant (nível um mês anterior ao nível atual).

Tabela 20 - Testes de hipóteses não paramétricos realizado para o intervalo de formulação (1976 a 2003).

	Aleatoriedade	Independência	Homogeneidade	Estacionariedade
Níveis				
Precipitação	X	X	X	X
Evaporação				
P-E*			X	X
MMP_5*				X
MME_10*				
MMP-E_7*			X	X
Precip_Sua*				
Evap_Sua*				

Nív_Ant*
Irrigação

Tabela 21 - Testes de hipóteses não paramétricos realizado para o intervalo de formulação (1960 a 1975).

	Aleatoriedade	Independência	Homogeneidade	Estacionariedade
Níveis			X	X
Precipitação	X		X	X
Evaporação				
P-E*			X	X
MMP_5*			X	X
MME_10*				
MMP-E_7*			X	
Precip_Sua*				
Evap_Sua*				
Nív_Ant*				X
Irrigação				

Tabela 22 - Testes de hipóteses não paramétricos realizado para o intervalo de formulação (2004 a 2011).

	Aleatoriedade	Independência	Homogeneidade	Estacionariedade
Níveis				X
Precipitação	X	X	X	X
Evaporação			X	X
P-E*	X		X	
MMP_5*			X	X
MME_10*				
MMP-E_7*				
Precip_Sua*				
Evap_Sua*				
Nív_Ant*				X
Irrigação				

7.2 Estimativa de equações de regressão

7.2.1 Modelo de regressão linear simples

Na Tabela 23 encontram-se os coeficientes estimados para as cinco possibilidades mencionadas, calculados para o intervalo de formulação.

Tabela 23 - Parâmetros estimados para as equações lineares simples.

Variável Explicativa	a	b
Precipitação	2,9071	0,0019
P-E	3,0543	0,0019
MMP_5	1,9110	0,0108
MMP-E_7	2,4612	0,0159
Sav_P	-0,8167	0,0353

Na Tabela 24, pode-se visualizar os resultados obtidos (intervalo de formulação) para os índices de desempenho utilizados para mensurar a eficiência das regressões, onde os coeficientes utilizados foram: a) erro médio; b) erro médio absoluto; c) erro médio quadrático; d) raiz quadrada do erro médio absoluto; e) coeficiente de Nash-Sutcliffe; f) coeficiente de persistência; g) coeficiente de correlação de Pearson e; h) coeficiente de determinação.

Tabela 24 - Medidas de desempenho para as regressões lineares simples.

Medidas de Desempenho	Precipitação	P-E	MMP_5	MMP-E_7	Sav_P
Erro médio	1,02E-06	8,69E-07	3,18E-07	7,95E-07	-7,67E-07
Erro médio absoluto	0,80	0,80	0,73	0,63	0,59
Erro médio quadrático	0,92	0,91	0,79	0,62	0,52
Raiz quadrada do erro médio	0,96	0,95	0,89	0,78	0,72
Coeficiente de Nash-Sutcliffe	0,04	0,05	0,17	0,35	0,45
Coeficiente de persistência	-31,30	-30,89	-26,67	-20,66	-17,38
Coeficiente de correlação de Pearson	0,19	0,22	0,42	0,59	0,67
Coeficiente de determinação	0,04	0,05	0,17	0,35	0,45

Os coeficientes que mensuram a diferença entre o valor observado e simulado pelos modelos, obtiveram valores relativamente elevados, indicando a possibilidade de ocorrência de erros de grande magnitude. A inspeção visual dos gráficos identificou uma considerável diferença entre os gráficos simulados e observados.

O coeficiente de Nash-Sutcliffe obteve resultados próximos a zero, revelando que as equações não são capazes de simular o banhado de forma eficiente, porém a equação obtida com base na suavização exponencial da precipitação obteve valor próximo a 0,5, que pode ser considerado relativamente bom. O coeficiente de persistência obteve valores baixos, podendo-se concluir que as equações avaliadas não conseguem melhorar o desempenho de um modelo hipotético onde o valor atual fornecido é igual ao último valor observado.

Com relação aos ajustes entre os valores simulados e observados, o coeficiente de correlação linear de Pearson e o coeficiente de determinação, obtiveram valores baixos, o que indica que os valores simulados pelas equações possuem um comportamento diferente dos valores observados. Pode-se concluir que as limitações das equações, ocasionadas pelas simplificações do método utilizado para elaboração das mesmas, forneceram resultados que não possuem uma boa correlação linear entre os dados observados, tornando as equações ineficientes em simular o banhado.

A equação estimada com os dados de precipitação suavizada exponencialmente obteve os melhores resultados dos índices de desempenho utilizados para mensurar a eficiência das regressões. Assim, a equação foi utilizada para simular os níveis no banhado para o período

de validação e avaliada pelos mesmos índices utilizados para o período de formulação, conforme apresentado na Tabela 25.

Tabela 25 - Medidas de desempenho calculadas para a equação obtida utilizando os dados de precipitação suavizada exponencialmente (intervalo de validação).

Medidas de Desempenho	Sav_P
Erro médio	-0,360
Erro médio absoluto	0,423
Erro médio quadrático	0,262
Raiz quadrada do erro médio	0,512
Coefficiente de Nash-Sutcliffe	-0,142
Coefficiente de persistência	-9,016
Coefficiente de correlação de Pearson	0,698
Coefficiente de determinação	0,487

Os índices de desempenho que mensuram o erro calculado para o período de validação entre os dados observados e simulados, obtiveram resultados que podem ser considerados aceitáveis, onde se pode verificar uma significativa melhora em relação ao período formulação. Contudo, os índices que mensuram o desempenho dos modelos em simular os dados obtiveram resultados muito baixos, podendo-se concluir que a equação não é capaz de simular ou prever os níveis no banhado.

Na Figura 44, pode-se visualizar a dispersão entre os dados observados e simulados utilizando os dados de precipitação suavizada exponencialmente. Na Figura 44.a, pode-se verificar que, apesar da dispersão em torno da linha, os dados tendem a ajustar-se a uma reta. Enquanto na Figura 44.b, pode-se observar que os resultados encontram-se agrupados, não sendo possível identificar um ajuste.

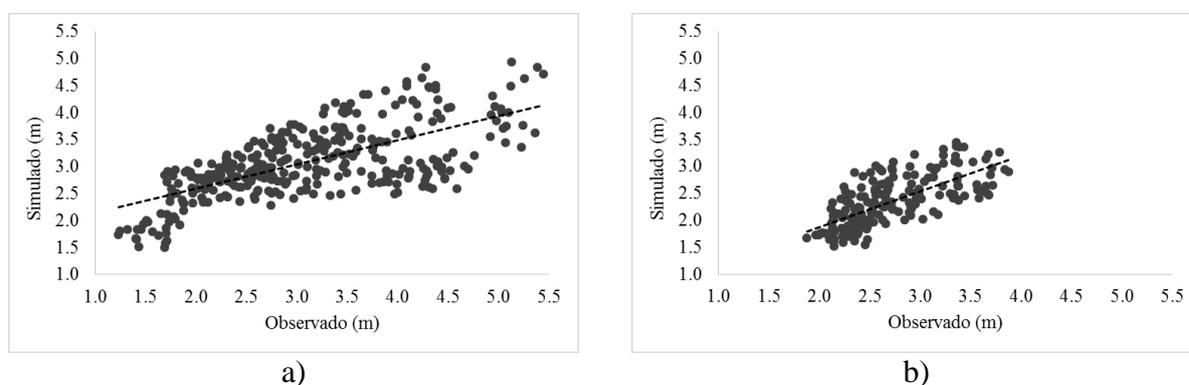


Figura 44 - Diagrama de dispersão entre os valores observados e simulados utilizando os dados de precipitação suavizada exponencialmente para os períodos de: a) formulação; b) validação.

7.2.2 Modelo de regressão não linear simples

Na Tabela 26, encontram-se os resultados dos índices de desempenho calculados para as regressões não lineares, utilizando-se os dados transformados de precipitação para as funções testadas. Os índices obtidos revelaram uma baixa eficiência, podendo-se concluir que a metodologia utilizada não encontrou resultados satisfatórios.

Tabela 26 - Medidas de desempenho calculadas para equação obtida utilizando os dados transformados de precipitação.

	X ²	X ³	X ⁴	LN(X)	$\sqrt[2]{X}$	$\sqrt[3]{X}$	$\sqrt[4]{X}$
Erro médio absoluto	0,80	0,80	0,80	0,80	0,80	0,80	0,80
Erro médio quadrático	0,92	0,92	0,93	0,93	0,92	0,92	0,92
Raiz quadrada do erro médio	0,96	0,96	0,96	0,96	0,96	0,96	0,96
Coefficiente de Nash-Sutcliffe	0,03	0,03	0,03	0,03	0,03	0,03	0,03
Coefficiente de persistência	-31,36	-31,49	-31,59	-31,59	-31,37	-31,40	-31,42
Coefficiente de correlação de Pearson	0,19	0,18	0,17	0,17	0,19	0,18	0,18
Coefficiente de determinação	0,04	0,03	0,03	0,03	0,04	0,03	0,03

Os testes realizados para os outros dados transformados também não encontraram bons resultados, onde os índices calculados obtiveram valores muito próximos aos da Tabela 26. Assim, pode-se concluir que a metodologia utilizada para estimar as equações não lineares simples não encontrou resultados satisfatórios.

7.2.3 Modelo de regressão linear múltipla

Devido à grande quantidade de equações que foram geradas, serão apresentadas apenas as equações que não foram rejeitadas pelos critérios estabelecidos. Na Tabela 27 encontram-se os resultados obtidos pelos coeficientes analisados para as 47 equações de regressões lineares múltiplas que não foram rejeitadas pelos critérios estabelecidos, onde a última coluna indica a quantidade de variáveis na equação com ocorrência de problemas de colinearidade.

Tabela 27 – Medidas de desempenho calculadas para as 47 equações que não foram rejeitadas pelos critérios pré-estabelecidos.

Nº da equação	R ²	EPR	Nash-Sut.	Colinearidade
26	0,9919	0,089	0,9918	
33	0,9915	0,091	0,9915	
35	0,9912	0,092	0,9912	
47	0,9929	0,083	0,9929	
76	0,9919	0,089	0,9919	

79	0,9919	0,089	0,9918	
81	0,9919	0,088	0,9919	
82	0,9919	0,088	0,9919	
85	0,9916	0,090	0,9916	
87	0,9912	0,092	0,9912	
90	0,9915	0,091	0,9915	
91	0,9915	0,091	0,9915	
111	0,9929	0,083	0,9929	
114	0,9929	0,083	0,9929	
116	0,9930	0,083	0,9929	
117	0,9929	0,083	0,9929	
120	0,9925	0,085	0,9925	
125	0,9925	0,085	0,9925	
150	0,9919	0,089	0,9919	
152	0,9920	0,088	0,9920	
153	0,9923	0,086	0,9923	2
155	0,9920	0,088	0,9919	
156	0,9920	0,088	0,9919	
157	0,9920	0,088	0,9920	
159	0,9916	0,090	0,9916	
162	0,9915	0,091	0,9915	
185	0,9929	0,083	0,9929	
187	0,9930	0,083	0,9929	
188	0,9930	0,083	0,9929	2
190	0,9930	0,083	0,9929	
191	0,9930	0,083	0,9929	
192	0,9930	0,083	0,9930	
194	0,9925	0,086	0,9925	
195	0,9925	0,086	0,9925	2
197	0,9925	0,086	0,9925	
214	0,9920	0,088	0,9920	
215	0,9923	0,087	0,9923	2
216	0,9923	0,086	0,9923	2
217	0,9920	0,088	0,9920	
218	0,9918	0,089	0,9918	2
235	0,9930	0,083	0,9929	
236	0,9930	0,083	0,9929	2
237	0,9930	0,083	0,9930	2
238	0,9930	0,083	0,9930	
239	0,9925	0,086	0,9925	2
246	0,9923	0,087	0,9923	2
253	0,9930	0,083	0,9930	2

7.2.3.1 Análise de sensibilidade das equações de regressão linear múltipla

As análises realizadas nas 47 equações obtiveram 1245 resultados gerados a partir das perturbações das variáveis de cada equação. Devido ao grande número de análises que seriam necessárias para avaliar cada resultado, esse estudo se restringiu em avaliar as variações ocorridas pelas perturbações fornecidas a variável precipitação, haja vista que essa constitui a variável independente de interesse.

Na Tabela 28 encontram-se os resultados das análises para a redução de 50% nos valores dos dados observados. As perturbações na variável precipitação foram sensivelmente percebidas pelos coeficientes de determinação e Nash-Sutcliffe, onde se pode notar uma alteração relativa próxima a 1% quando comparados os resultados com e sem a perturbação da variável de entrada. O erro padrão da regressão mostrou-se sensível às variações da precipitação. Os erros dos resíduos foram calculados pelo maior valor obtido de cada equação, comparando-se a situação sem a alteração da variável e com a alteração da variável, onde se pode perceber que o erro absoluto obteve o maior resultado, de 0,3 m, comparando o nível obtido com e sem a perturbação da variável de entrada, representando em torno de 0,6 m quando comparado com o valor observado.

Tabela 28 - Erro absoluto e relativo percentual para as medidas de desempenho calculadas para as 47 equações que não foram rejeitadas pelos critérios pré-estabelecidos.

Nº da equação	R ²		EPR		Nash-Sutcliffe		Resíduos	
	Absoluto	Relativo (%)	Absoluto	Relativo (%)	Absoluto	Relativo (%)	Absoluto (m)	Relativo (%)
26	-0,0098	0,99	0,043	48,63	-0,010	0,99	0,2	66,67
33	-0,0113	1,14	0,048	52,87	-0,011	1,15	0,3	100,00
35	-0,0126	1,27	0,052	56,75	-0,013	1,29	0,3	100,00
47	-0,0088	0,89	0,041	49,77	-0,009	0,89	0,2	66,67
76	-0,0098	0,99	0,043	48,95	-0,010	1,00	0,2	66,67
79	-0,0098	0,99	0,043	48,58	-0,010	0,99	0,2	66,67
81	-0,0091	0,92	0,041	46,21	-0,009	0,93	0,2	66,67
82	-0,0097	0,97	0,043	48,59	-0,010	0,98	0,2	66,67
85	-0,0112	1,13	0,048	53,01	-0,011	1,14	0,2	66,67
87	-0,0127	1,28	0,052	56,94	-0,013	1,29	0,3	100,00
90	-0,0113	1,14	0,048	53,04	-0,012	1,16	0,3	100,00
91	-0,0113	1,14	0,048	52,97	-0,011	1,15	0,3	100,00
111	-0,0087	0,88	0,041	49,55	-0,009	0,89	0,2	66,67
114	-0,0088	0,89	0,041	49,79	-0,009	0,89	0,2	66,67
116	-0,0082	0,83	0,039	47,35	-0,008	0,83	0,2	66,67
117	-0,0087	0,88	0,041	49,74	-0,009	0,89	0,2	66,67
120	-0,0103	1,04	0,047	54,35	-0,010	1,05	0,2	66,67
125	-0,0104	1,05	0,047	54,65	-0,011	1,06	0,2	66,67

150	-0,0098	0,99	0,043	48,96	-0,010	1,00	0,2	66,67
152	-0,0091	0,92	0,041	46,35	-0,009	0,92	0,2	66,67
153	-0,0098	0,99	0,044	50,95	-0,010	0,99	0,2	66,67
155	-0,0091	0,91	0,041	46,06	-0,009	0,92	0,2	66,67
156	-0,0097	0,98	0,043	48,69	-0,010	0,98	0,2	66,67
157	-0,0092	0,93	0,041	46,78	-0,009	0,93	0,2	66,67
159	-0,0111	1,12	0,048	52,75	-0,011	1,13	0,2	66,67
162	-0,0114	1,15	0,049	53,53	-0,012	1,17	0,3	100,00
185	-0,0087	0,88	0,041	49,55	-0,009	0,89	0,2	66,67
187	-0,0082	0,83	0,039	47,34	-0,008	0,83	0,2	66,67
188	-0,0088	0,89	0,042	50,39	-0,009	0,90	0,2	66,67
190	-0,0082	0,83	0,039	47,33	-0,008	0,83	0,2	66,67
191	-0,0088	0,88	0,041	49,90	-0,009	0,89	0,2	66,67
192	-0,0083	0,83	0,040	47,70	-0,008	0,84	0,2	66,67
194	-0,0104	1,04	0,047	54,59	-0,011	1,06	0,2	66,67
195	-0,0104	1,05	0,047	54,66	-0,011	1,06	0,3	100,00
197	-0,0104	1,05	0,047	54,83	-0,011	1,06	0,3	100,00
214	-0,0009	0,91	0,041	46,03	-0,009	0,91	0,2	66,67
215	-0,0098	0,99	0,044	51,03	-0,010	1,00	0,2	66,67
216	-0,0094	0,95	0,043	49,54	-0,010	0,96	0,2	66,67
217	-0,0092	0,93	0,041	46,92	-0,009	0,94	0,2	66,67
218	-0,0117	1,18	0,050	56,44	-0,012	1,20	0,2	66,67
235	-0,0082	0,83	0,039	47,34	-0,008	0,83	0,2	66,67
236	-0,0088	0,89	0,042	50,51	-0,009	0,90	0,2	66,67
237	-0,0084	0,84	0,040	48,36	-0,008	0,85	0,2	66,67
238	-0,0083	0,84	0,040	47,99	-0,008	0,84	0,2	66,67
239	-0,0105	1,06	0,048	55,39	-0,011	1,07	0,2	66,67
246	-0,0095	0,95	0,043	49,70	-0,010	0,96	0,2	66,67
253	-0,0084	0,85	0,040	48,65	-0,009	0,86	0,2	66,67

Dentre as equações de regressão linear múltipla estimadas foram selecionadas oito equações (Tabela 29) que não foram rejeitadas pelos critérios pré-estabelecidos e que se mostraram sensíveis às variações nos dados observados de precipitação. Cabe ressaltar que as equações que foram selecionadas obtiveram melhores desempenhos que as equações que foram rejeitadas.

Tabela 29 – Equações de regressão linear múltipla selecionadas.

nº da equação	Equações
33	$N = -0,04 + 0,002.P + 0,0007.MMPE_7 + 0,97.N_{ant}$
35	$N = -0,04 + 0,002.P + 0,0007.Sav_P + 0,97.N_{ant}$
87	$N = -0,14 + 0,002.P + 0,0002.MME_{10} + 0,0007.Sav_P + 0,97.N_{ant}$

90	$N = -0,06 + 0,002.P + 0,0006.MMPE_7 + 0,0001.Sav_P + 0,96.N_{ant}$
91	$N = -0,03 + 0,002.P + 0,0006.MMPE_7 - 0,0001.Sav_E + 0,96.N_{ant}$
162	$N = -0,03 + 0,002.P + 0,0006.MMPE_7 + 0,0003.Sav_P - 0,0004.Sav_E + 0,96.N_{ant}$
195	$N = -0,01 + 0,002.P - 0,0007.E + 0,0005.MME_{10} + 0,0006.MMPE_7 - 0,0004.Sav_E + 0,97.N_{ant}$
197	$N = -0,02 + 0,002.P - 0,0007.E + 0,0005.MMPE_7 + 0,0003.Sav_P - 0,0001.Sav_E + 0,96.N_{ant}$

*Obs.: P é a precipitação; E é a evaporação; MME₁₀ é a média móvel de dez meses da evaporação; MMP₅ é a média móvel de cinco meses da precipitação; MMPE₇ é a média móvel de sete meses da precipitação subtraída da evaporação; Sav_E é a evaporação suavizada exponencialmente; Sav_P é a precipitação suavizada exponencialmente; N_{ant} é o nível anterior ao simulado.

7.2.3.2 Teste da significância das equações de regressão linear múltipla

Para avaliar a existência de uma relação significativa entre a variável dependente e variáveis independentes (ou explicativas), foi realizado o teste de significância, também conhecido como “teste do F total”. O teste foi realizado para duas hipóteses:

- H₀: $a_1 = a_2 = \dots = a_n = 0$ (a relação entre as variáveis não é linear);
- H₁: pelo menos um $a_i \neq 0$.

O cálculo do F das equações foi realizado pela análise de variância das equações, onde os resultados encontram-se nas Tabela 30 a Tabela 37.

Tabela 30 - Tabela ANOVA da equação de regressão múltipla nº 33.

Fonte	Graus de liberdade	Somatório dos quadrados	Quadrado médio	Estatística F
Regressão	3	311,97	103,99	12651,04
Resíduos	332	2,73	0,0082	
Total	335	314,7		

Tabela 31 - Tabela ANOVA da equação de regressão múltipla nº 35.

Fonte	Graus de liberdade	Somatório dos quadrados	Quadrado médio	Estatística F
Regressão	3	311,89	103,96	12230,59
Resíduos	332	2,80	0,0085	

Total	335	314,7
--------------	-----	-------

Tabela 32 - Tabela ANOVA da equação de regressão múltipla n° 87.

Fonte	Graus de liberdade	Somatório dos quadrados	Quadrado médio	Estatística F
Regressão	4	311,89	77,97	9217,5
Resíduos	331	2,80	0,0085	
Total	335	314,7		

Tabela 33 - Tabela ANOVA da equação de regressão múltipla n° 90.

Fonte	Graus de liberdade	Somatório dos quadrados	Quadrado médio	Estatística F
Regressão	4	311,97	77,99	9462,1
Resíduos	331	2,73	0,0082	
Total	335	314,7		

Tabela 34 - Tabela ANOVA da equação de regressão múltipla n° 91.

Fonte	Graus de liberdade	Somatório dos quadrados	Quadrado médio	Estatística F
Regressão	4	311,97	77,99	9462,1
Resíduos	331	2,73	0,0082	
Total	335	314,7		

Tabela 35 - Tabela ANOVA da equação de regressão múltipla n° 162.

Fonte	Graus de liberdade	Somatório dos quadrados	Quadrado médio	Estatística F
Regressão	5	311,9	62,36	7552,1
Resíduos	330	2,73	0,0083	
Total	335	314,7		

Tabela 36 - Tabela ANOVA da equação de regressão múltipla n° 195.

Fonte	Graus de liberdade	Somatório dos quadrados	Quadrado médio	Estatística F
Regressão	6	312,29	52,05	7102,35
Resíduos	329	2,41	0,0073	

Total	335	314,7
--------------	-----	-------

Tabela 37 - Tabela ANOVA da equação de regressão múltipla nº 197.

Fonte	Graus de liberdade	Somatório dos quadrados	Quadrado médio	Estatística F
Regressão	6	312,29	52,05	7102,35
Resíduos	329	2,41	0,0073	
Total	335	314,7		

Os testes foram realizados a um nível de significância 0,05 (5%), onde os valores calculados para cada equação foram avaliados pelo F da distribuição de Snedecor. Assim, se o F calculado for maior que o $F(\alpha, P, n-P-1)$ de Snedecor, rejeita-se a hipótese nula (H_0). Na tabela A3 do Anexo A, encontram-se os valores da distribuição de Snedecor, calculados a um nível de significância de 5%. Na Tabela 38, encontram-se os resultados dos testes de significância realizados para as oito equações de regressão obtidas nesse estudo.

Tabela 38 – Teste de significância (F total).

nº da equação	F calculado	F Snedecor	Teste
33	12651,04	$[F(0,05; 3; 120) = 2,68] > F(0,05; 3; 332)$	Rejeita-se H_0
35	12230,59	$[F(0,05; 3; 120) = 2,68] > F(0,05; 3; 332)$	Rejeita-se H_0
87	9217,5	$[F(0,05; 4; 120) = 2,45] > F(0,05; 4; 331)$	Rejeita-se H_0
90	9462,1	$[F(0,05; 4; 120) = 2,68] > F(0,05; 4; 331)$	Rejeita-se H_0
91	9462,1	$[F(0,05; 4; 120) = 2,68] > F(0,05; 4; 331)$	Rejeita-se H_0
162	7552,1	$[F(0,05; 5; 120) = 2,29] > F(0,05; 5; 330)$	Rejeita-se H_0
195	7102,35	$[F(0,05; 6; 120) = 2,18] > F(0,05; 6; 329)$	Rejeita-se H_0
197	7102,35	$[F(0,05; 6; 120) = 2,18] > F(0,05; 6; 329)$	Rejeita-se H_0

A hipótese nula foi rejeitada em todos os testes realizados, revelando que existe uma relação significativa entre a variável dependente e as variáveis independentes em todas as oito equações avaliadas.

7.3 Detecção e remoção de viés nos dados do modelo COLA

7.3.1 Detecção do viés

Os resultados das análises realizadas com os coeficientes propostos para verificar a necessidade de correção do viés estão dispostos de forma resumida na Tabela 39, onde se

encontram os valores máximos e os mínimos calculados para os sete horizontes de previsão, no qual foram utilizados os 30 membros para a realização dos cálculos. Da investigação dos resultados, percebeu-se que os índices não possuíam relação, ou seja, alguns membros possuíam um valor relativamente baixo em um índice, e elevado em outro.

Tabela 39 - Resultados dos coeficientes de mensuração do viés.

Horizonte	Erro Médio		Desvio*		RV*		REMQ*	
	Maior	Menor	Maior	Menor	Maior	Menor	Maior	Menor
1 mês	24,06	4,74	52,14	42,28	0,55	0,10	67,75	58,16
2 meses	17,96	4,77	52,24	42,28	0,40	0,10	69,04	52,62
3 meses	18,20	-3,42	58,62	45,11	0,38	0,03	74,21	59,68
4 meses	13,11	1,34	56,50	45,89	0,25	0,03	73,34	60,28
5 meses	14,03	-1,90	59,30	40,01	0,31	0,03	74,99	57,24
6 meses	15,39	1,19	55,88	45,79	0,31	0,02	75,36	60,33
7 meses	15,01	-6,07	55,24	43,59	0,31	0,02	72,50	57,39

O Erro Médio (Viés) apresentou os valores mais elevados para o primeiro horizonte de antecedência, o que não era esperado, pois a incerteza das previsões tende a aumentar com o aumento do horizonte. A amplitude máxima entre os valores calculados para os 30 membros foi detectada para o horizonte de três meses de antecedência, o que indica que houve um maior espalhamento dos dados previstos pelo MCG. Da Tabela 39 pode-se perceber que alguns membros obtiveram resultados bastante satisfatórios, obtendo valores próximos a unidade.

O coeficiente RV obteve valores relativamente baixos (inferiores a unidade), o que pode ser um indício de que os membros não necessitem de correção, podendo ser utilizados os dados brutos nas previsões. Cabe ressaltar que os maiores valores encontrados foram para o horizonte de um mês de antecedência, o que não é uma surpresa, haja vista que o Erro Médio (que obteve os maiores erros para o primeiro horizonte) compõe a equação para o cálculo desse coeficiente.

Considerando que a quantidade de dados utilizados para essa análise varia entre 91 a 85 amostras para os horizontes de previsão, é plausível admitir que o REQM encontrasse valores relativamente elevados (entre 52,6 a 75,4). Esse fato pode ser justificado pela existência de erros pontuais de grande magnitude que podem acarretar em graves erros em alguns períodos extremos (cheias ou secas severas).

Os resultados obtidos com o método de verificação do *rank histogram* encontram-se na Figura 45, calculado para a série de precipitação prevista para os sete meses do horizonte de antecedência. Pode-se perceber que não existe uma uniformidade nos histogramas, em que

os gráficos aproximam-se da forma de U. Essa configuração indica que o modelo possui dificuldade em representar os valores extremos. É possível perceber uma maior frequência para a última classe, que pode ser interpretada como uma tendência do MCG em subestimar a precipitação observada.

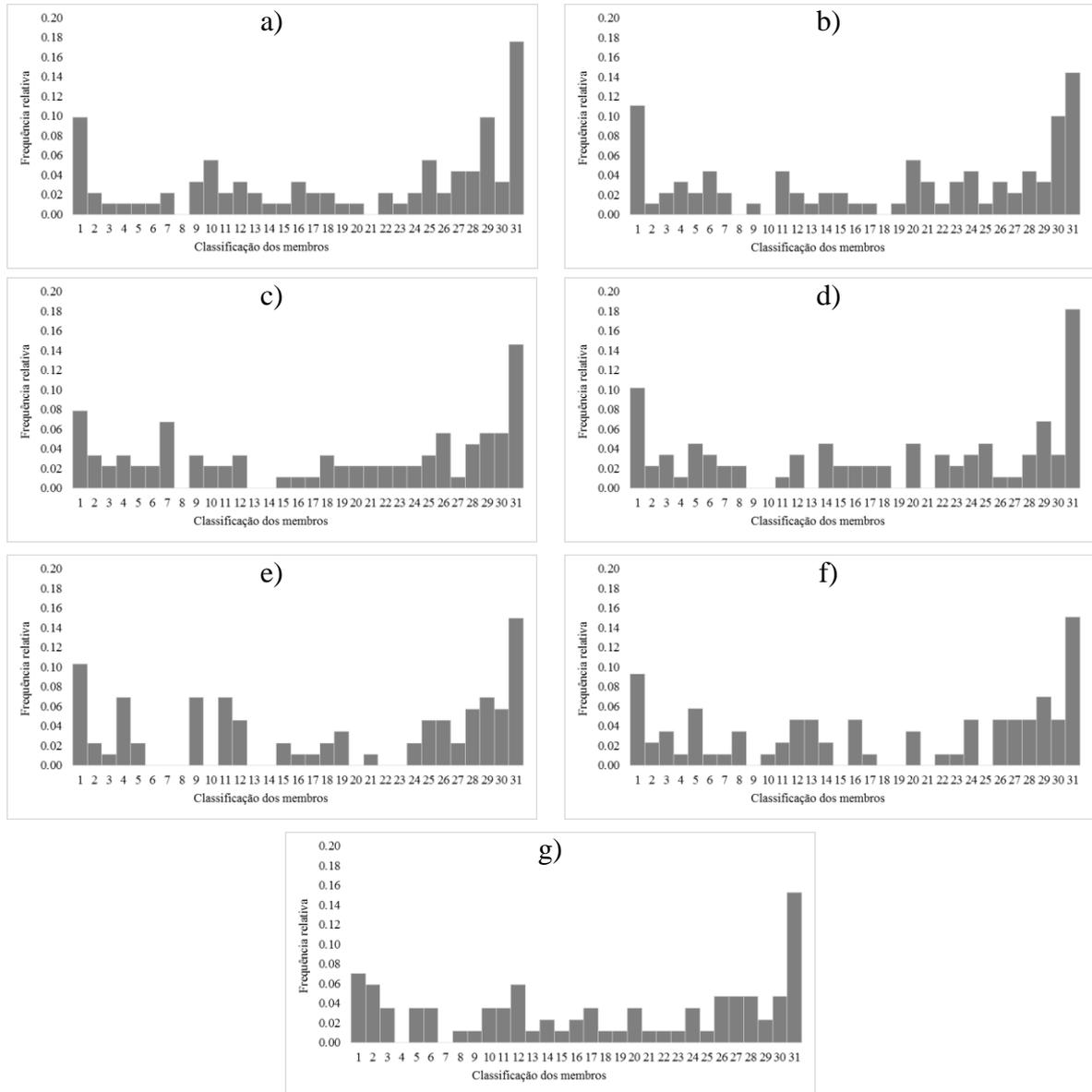


Figura 45 - Rank histogram calculado para os dados de precipitação prevista para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.

Na Figura 46 encontra-se o gráfico dos membros previstos da precipitação para o horizonte de um mês de antecedência, onde se pode verificar o comportamento da Figura 45.a sendo observado que quanto mais distante da média observada os membros obtiveram maiores dificuldades em representar os picos e vales.

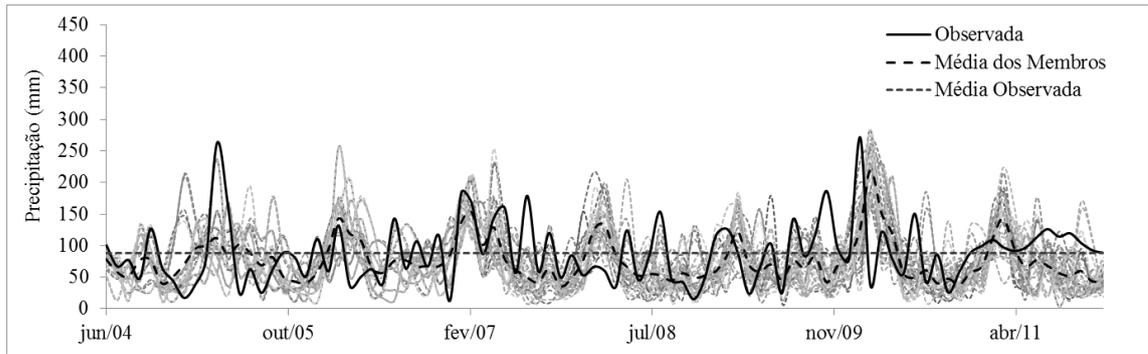
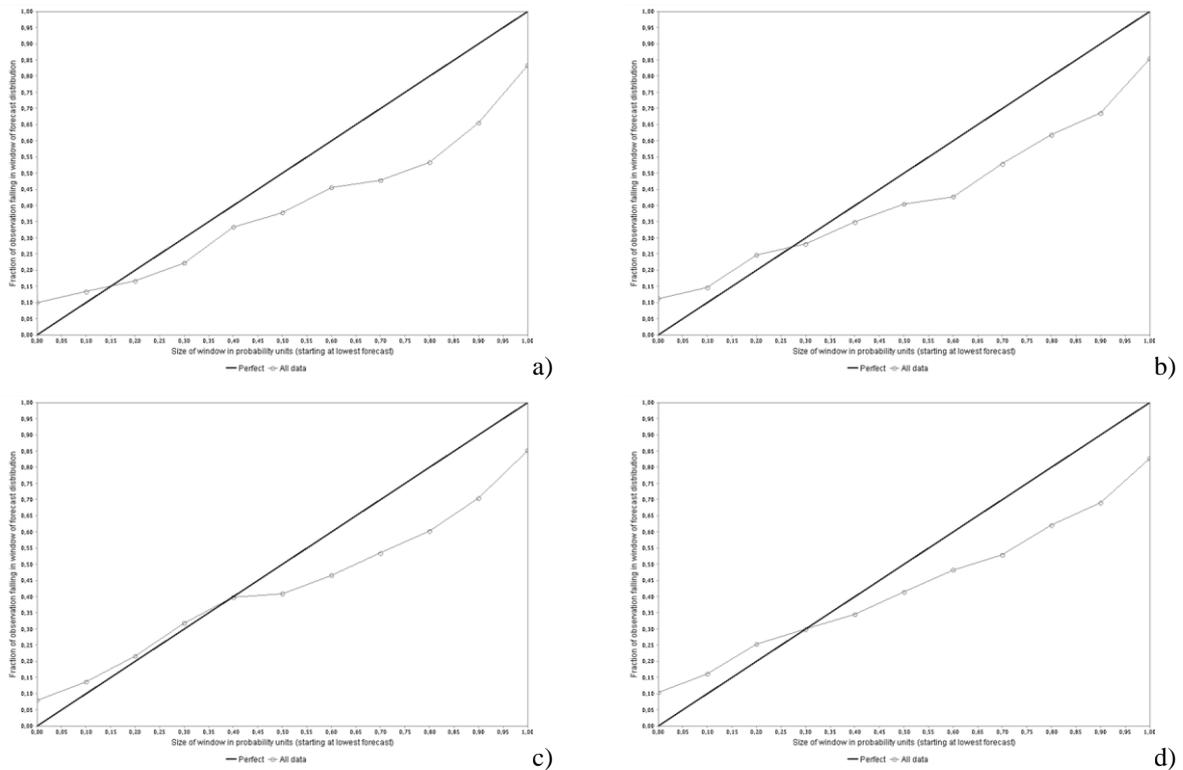


Figura 46 - Gráfico da precipitação observada sobreposta aos membros previstos para o horizonte de um mês de antecedência.

Os resultados obtidos pelo método de verificação *Spread-bias diagram* encontram-se na Figura 47, calculado para a série de precipitação prevista para os sete meses do horizonte de antecedência. Pode-se perceber que os horizontes de três, cinco e sete meses de antecedência obtiveram melhores resultados, para as probabilidades de ocorrência mais baixas (inferiores a 50%). Apesar do subespalhamento observado em todos os gráficos, é visível a tendência em subestimar os valores extremos para todos os horizontes previstos, conforme foi diagnosticado pelo *rank histogram*.



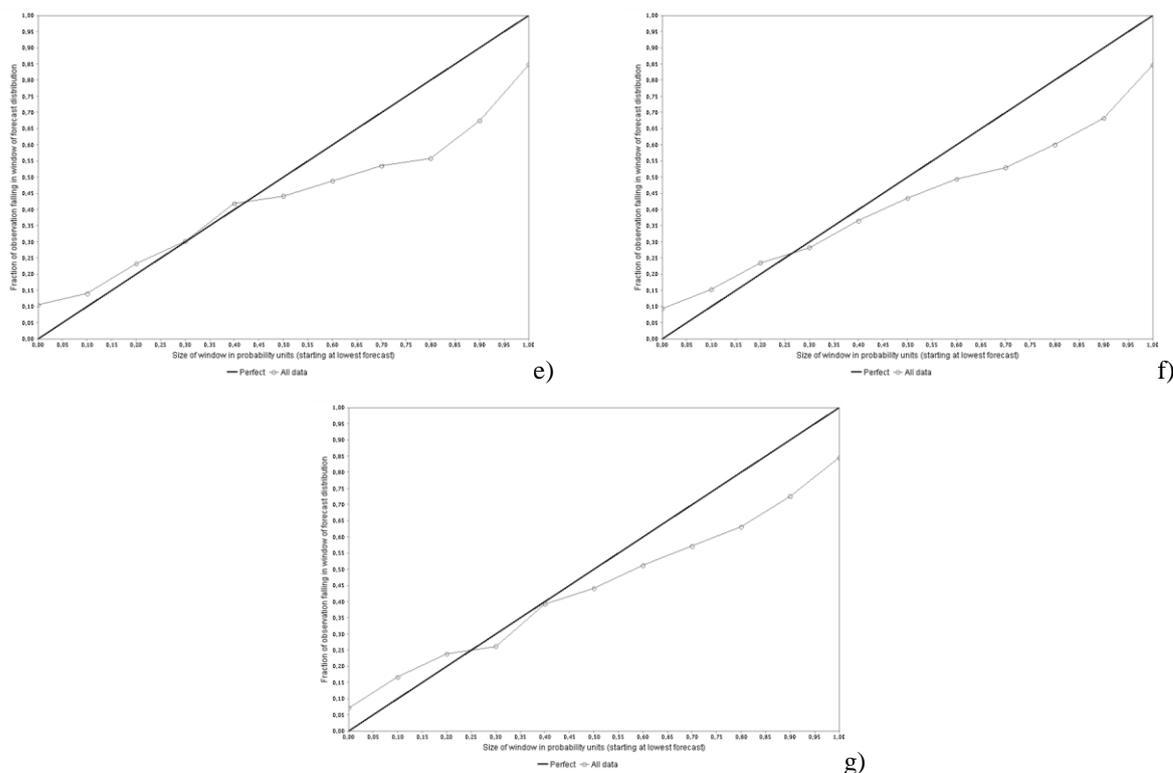


Figura 47 – *Spread-bias diagram* calculado para os dados de precipitação para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.

A quantidade de dados não permite uma discretização mensal, pois a análise seria bastante precária, com o total de dados variando entre 6 a 8 dados, que o seria um número inferior à quantidade de membros previstos.

Os resultados obtidos pelos coeficientes na avaliação determinística indicam que, apesar dos desvios mensurados, a série de precipitação prevista poderia ser utilizada sem grandes prejuízos para os resultados. Entretanto, os resultados obtidos pelos coeficientes na avaliação probabilística, indicam um subespalhamento dos dados, com uma tendência em subestimar os dados observados. O *rank histogram* revelou uma distribuição não uniforme, onde a forma de “U” indica que a série observada possui uma maior frequência nos intervalos não cobertos pelos membros previstos, o que pode ser prejudicial para períodos extremos como secas ou cheias. A mesma conclusão foi obtida pela análise do *Spread-bias diagram*. Cabe ressaltar que as previsões realizadas nesse estudo encontram-se inseridas em um contexto de disputa pelo uso racional da água (irrigação x biodiversidade), onde os períodos mais críticos são de imensa importância para gerar discussões quando da disponibilidade desse bem escasso que é a água. Assim, realizou-se a aplicação de uma metodologia para remover (ou reduzir) o viés presente nos dados de precipitação prevista.

7.3.2 Remoção do viés

Na Figura 48 encontram-se os gráficos das FCP's para as respectivas médias, onde o teste de aderência revelou que a série de precipitação pode ser simulada para as FCP's empírica (Weibull) e Gama. Na Figura 48.a, pode-se verificar que o modelo superestima a precipitação observada para valores inferiores a uma probabilidade próxima a 0,177 (ou precipitação próxima a 39,9 mm), subestimando para as demais observações. Na Figura 48.b, pode-se verificar que o modelo subestima todos os valores da precipitação observada, com exceção da distribuição Gama na qual superestima valores inferiores a uma precipitação de 5,5 mm.

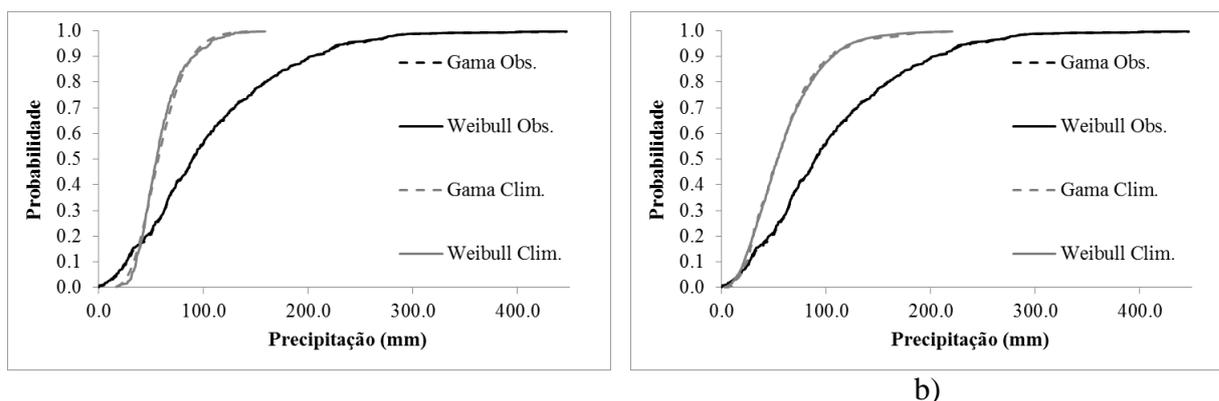


Figura 48 - Função cumulativa de probabilidades da precipitação observada e a: a) média da climatologia; b) média organizada de forma crescente da climatologia.

Da Figura 48 nota-se que as FCP's possuem um comportamento similar, sendo difícil identificar qual obteria resultados mais satisfatórios, podendo-se utilizar apenas uma delas para a correção do viés. Assim, escolheu-se a FCP de Weibull pela simplicidade na realização dos cálculos.

Na Figura 49, encontram-se os gráficos dos dez membros da climatologia, as médias dos membros (a e b) e uma FCP contendo todos os valores dos membros calculados pela FCP de Weibull. Com exceção da média dos membros (a), pode-se verificar que todas as FCP's possuem gráficos semelhantes. Assim, para a remoção do viés foram utilizadas a média dos membros da climatologia (a), a média organizada de forma crescente dos membros da climatologia (b) e o primeiro membro da climatologia (membro escolhido aleatoriamente).

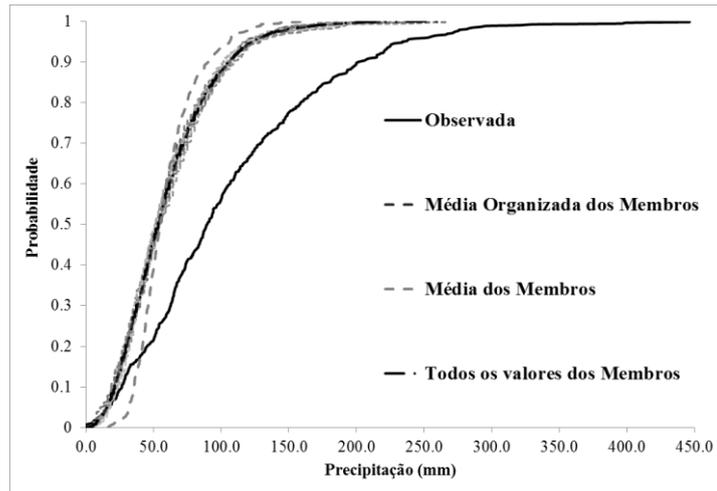
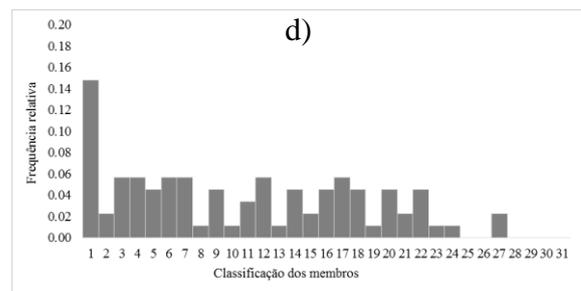
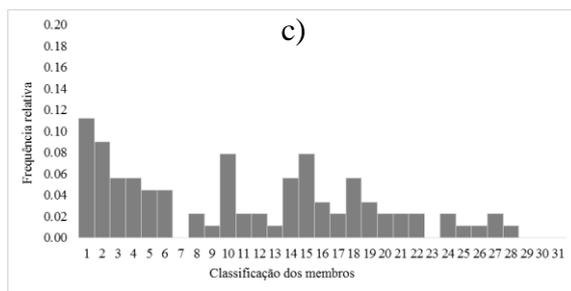
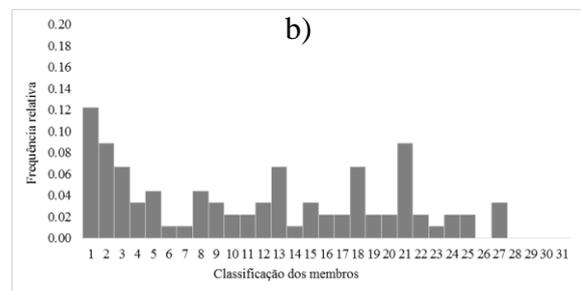
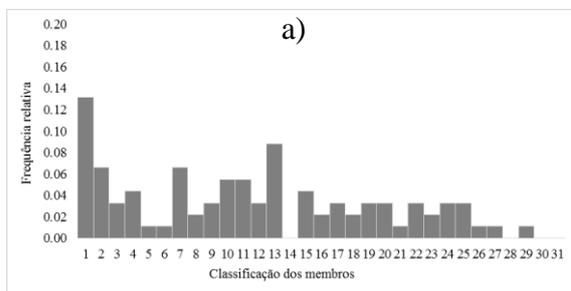


Figura 49 - Distribuições acumulada de probabilidades dos membros da climatologia.

A primeira remoção do viés foi realizada utilizando a média dos membros da climatologia (a). Os resultados obtidos com o método de verificação do *rank histogram* após a correção do viés encontram-se na Figura 50. Pode-se perceber que apesar do maior espalhamento dos histogramas, existe uma maior concentração de dados na primeira classe do histograma, indicando que os membros estão superestimando os valores observados. Na Figura 51, encontra-se o gráfico dos membros previstos para o horizonte de um mês corrigido, onde se pode observar que os membros estão superestimando a precipitação observada. Na Figura 52, pode-se verificar que as médias dos membros previstos para o horizonte de um mês antes e após a remoção possuem uma diferença acentuada, no qual a média antes da correção possui um comportamento mais próximo ao da precipitação observada, indicando que a remoção foi ineficiente.



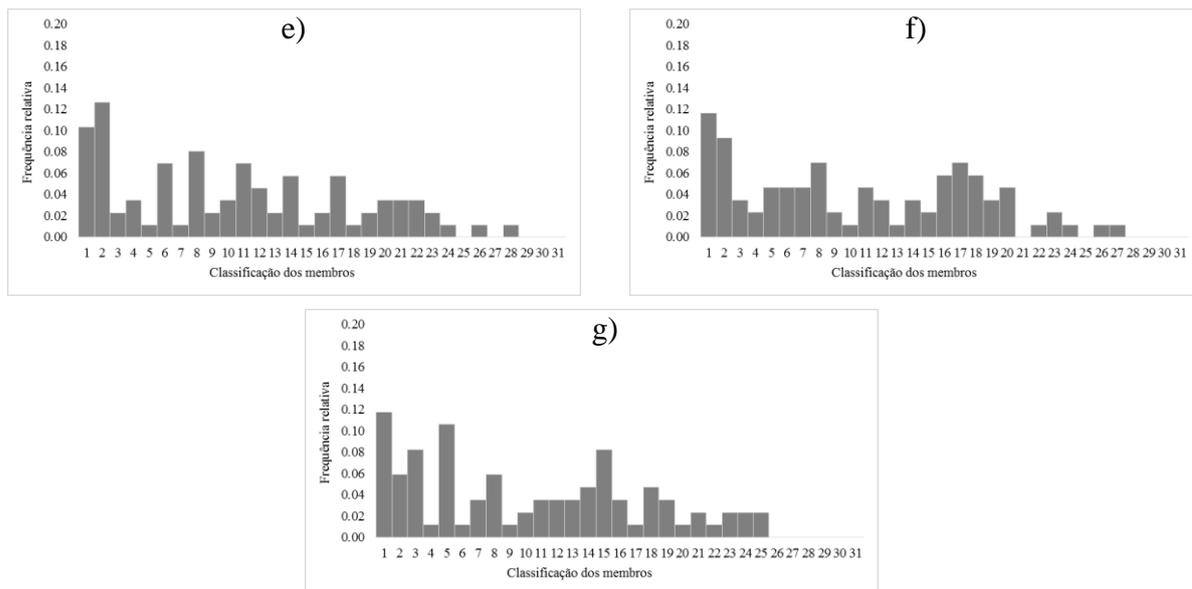


Figura 50 - Rank histogram calculado para os dados de precipitação prevista com remoção do viés utilizando a média dos membros da climatologia para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quatro meses; e) cinco meses; f) seis meses; g) sete meses.

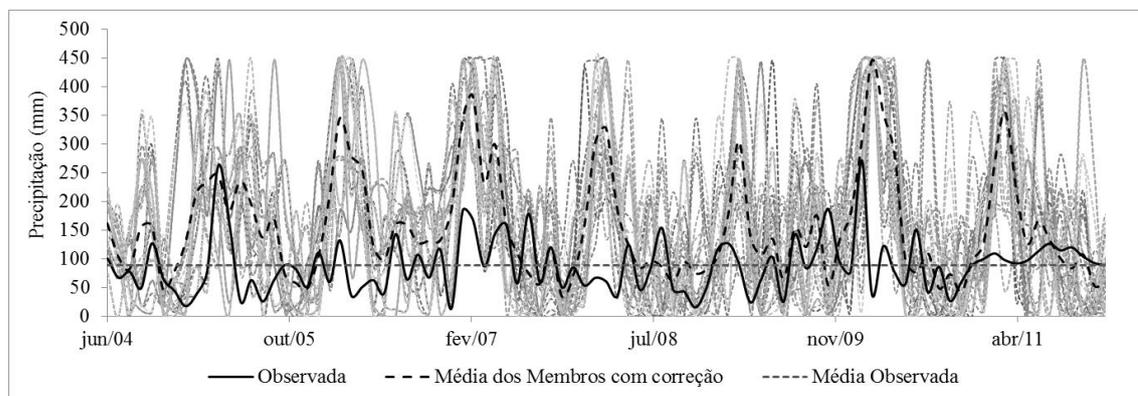


Figura 51 - Membros previstos da precipitação para o horizonte de um mês de antecedência com remoção do viés utilizando a média dos membros da climatologia.

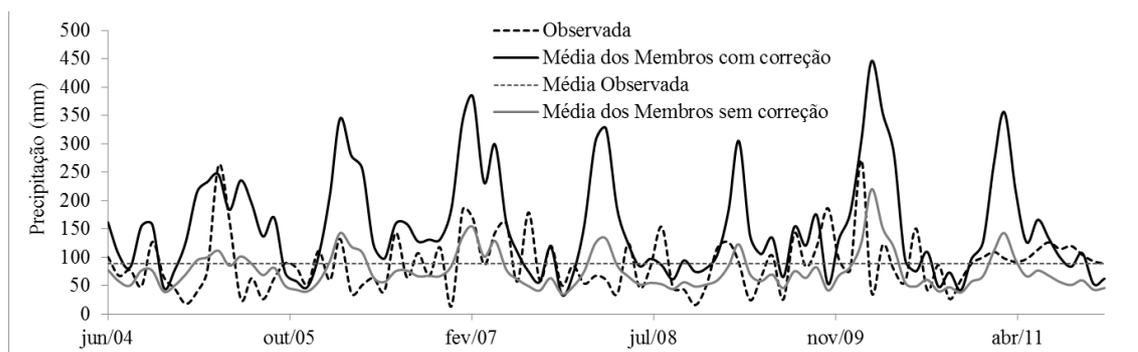


Figura 52 - Média dos membros previstos da precipitação para o horizonte de um mês de antecedência sem e com remoção do viés utilizando a média dos membros da climatologia.

Os resultados obtidos com o método de verificação do *Spread-bias diagram* após a correção do viés encontram-se na Figura 53. Pode-se perceber que para todos os horizontes previstos a linha que une os pontos está acima da linha de 45° do diagrama, indicando uma superestimativa dos valores da precipitação. Podendo-se concluir que a remoção do viés foi ineficiente.

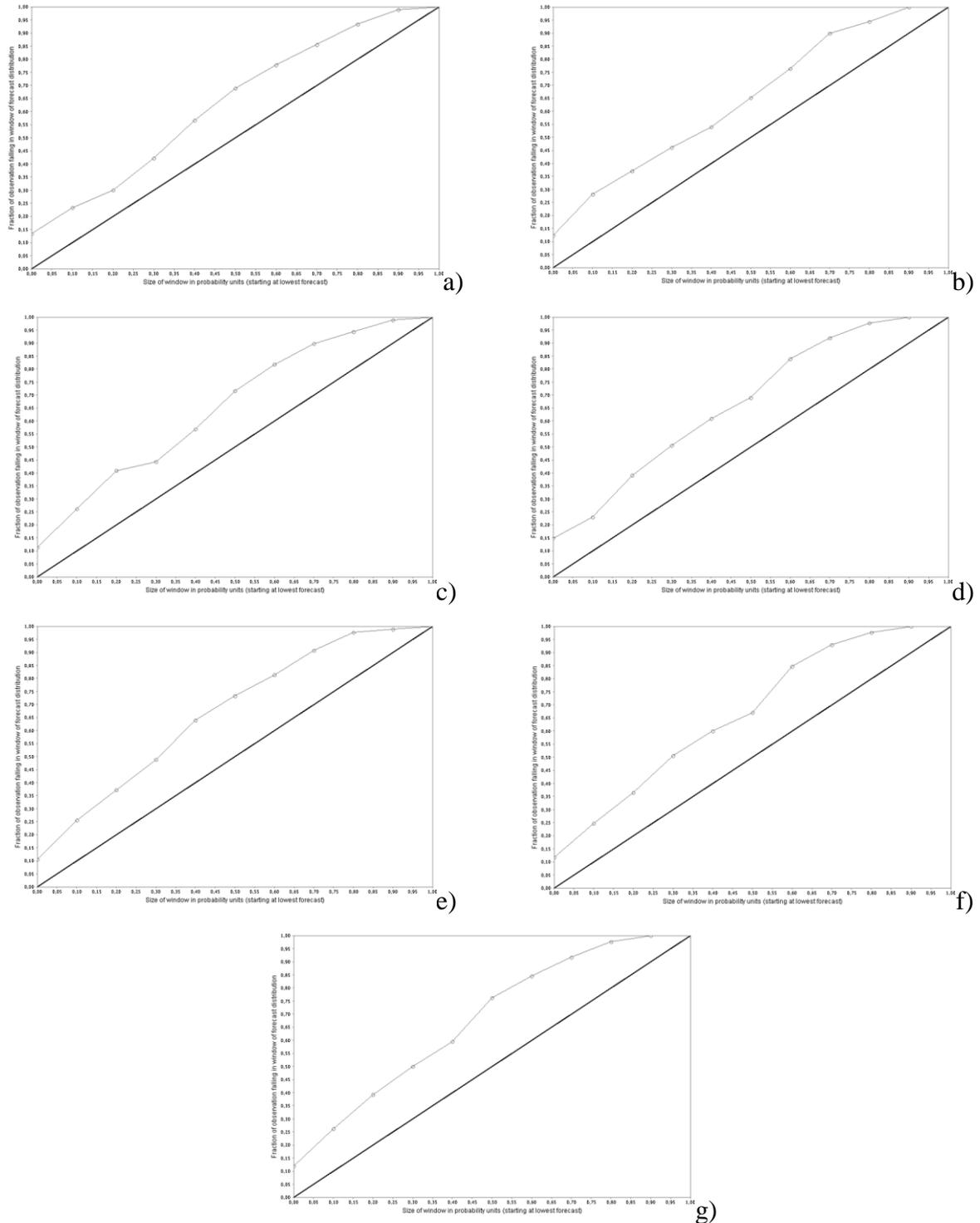
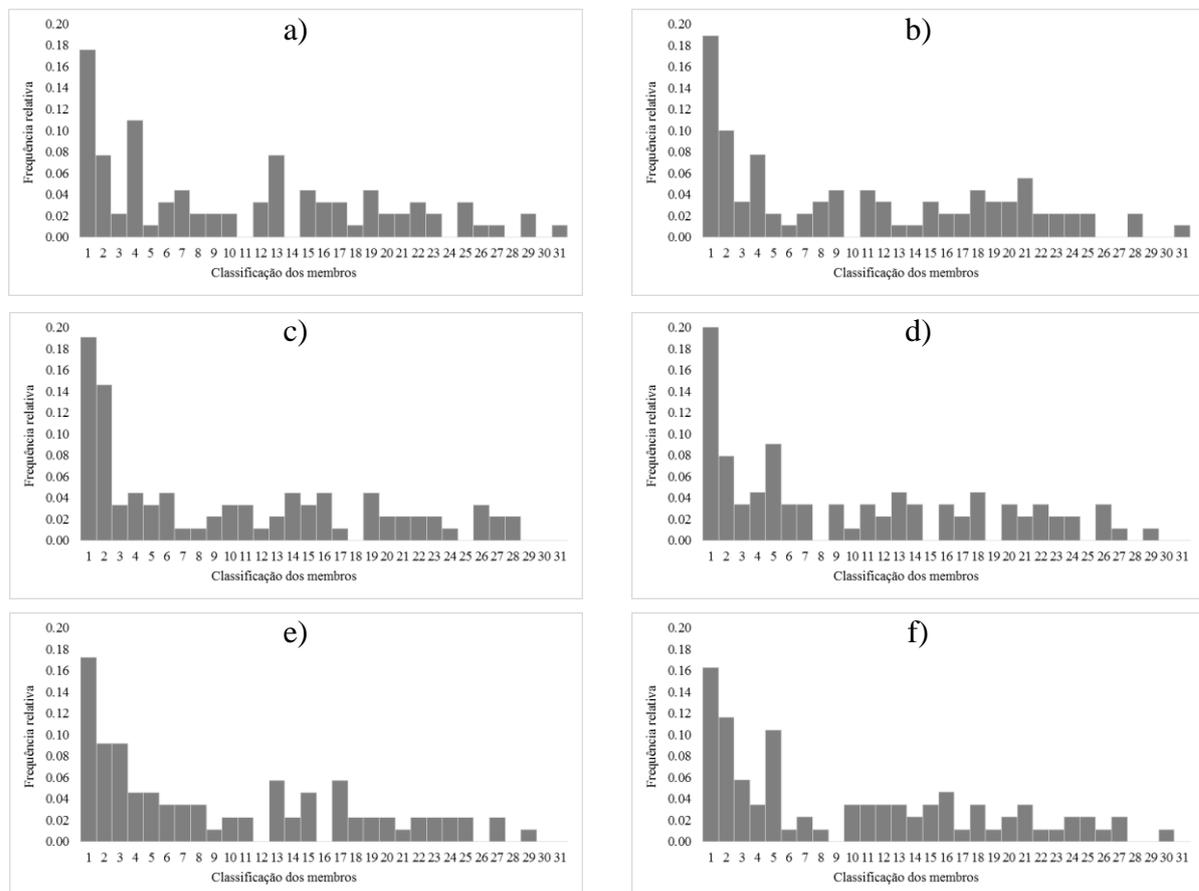


Figura 53 - *Spread-bias diagram* calculado para os dados de precipitação prevista com remoção do viés utilizando a média dos membros da climatologia para os horizontes de

antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.

Os coeficientes calculados para avaliar o desempenho do MCG após a remoção do viés, utilizando a média dos membros da climatologia (a), indicaram que o método utilizado não obteve resultados satisfatórios. Os resultados após a remoção obtiveram desempenho inferior aos obtidos antes da remoção. Portanto, pode-se concluir que a remoção do viés, utilizando a média dos membros da climatologia (a), foi ineficiente.

Foi realizada a remoção do viés utilizando a média organizada de forma crescente dos membros da climatologia (b). Os resultados obtidos com o método de verificação do *rank histogram* após a correção do viés encontram-se na Figura 54. Pode-se verificar que existe uma maior frequência para a primeira classe, para todos os horizontes de previsão propostos, indicando que uma parcela significativa dos dados observados encontra-se abaixo dos membros previsto. Essa configuração indica que a distribuição do conjunto superestimou os dados observados.



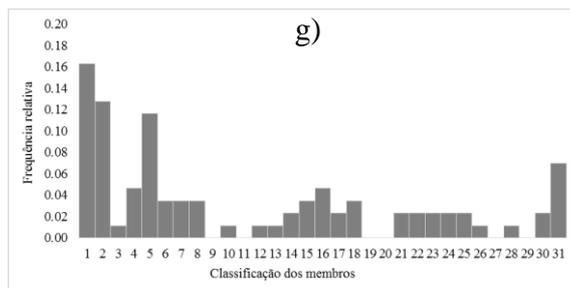


Figura 54 - Rank histogram calculado para os dados de precipitação prevista com remoção do viés utilizando a média organizada de forma crescente dos membros da climatologia para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.

Na Figura 55 encontra-se o gráfico dos membros previstos da precipitação para o horizonte de um mês de antecedência após a remoção do viés, utilizando a média organizada de forma crescente dos membros da climatologia (b), onde se pode verificar o comportamento da Figura 54.a, em forma de gráfico de linhas. Na Figura 56, pode-se perceber um incremento significativo entre a média dos membros antes e após a correção. É possível perceber que a média dos membros após a remoção do viés está acima do gráfico dos dados observados.

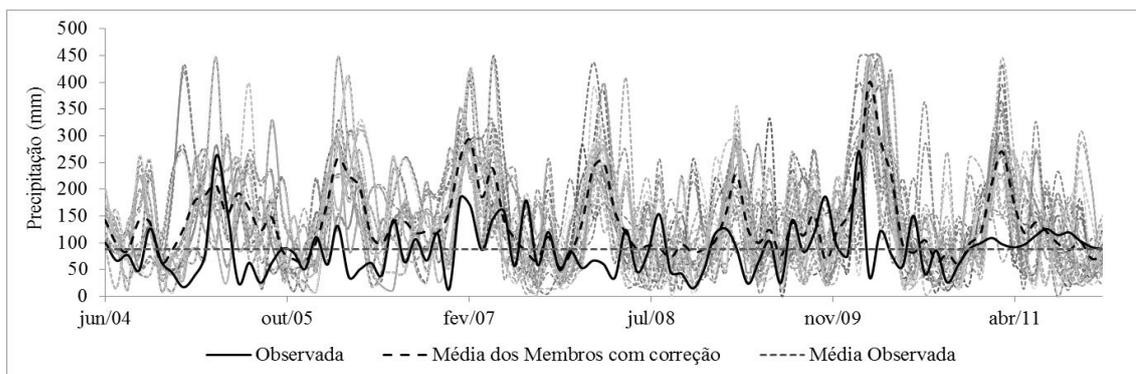


Figura 55 - Membros previstos da precipitação para o horizonte de um mês de antecedência com remoção do viés utilizando a média organizada de forma crescente dos membros da climatologia.

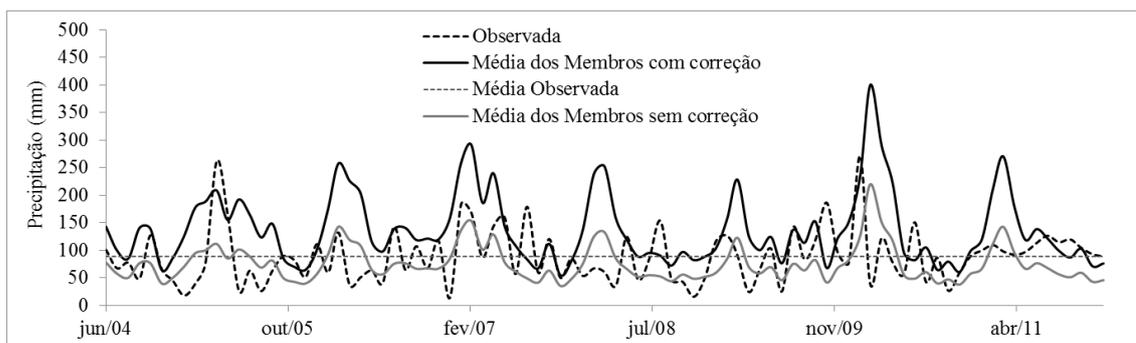
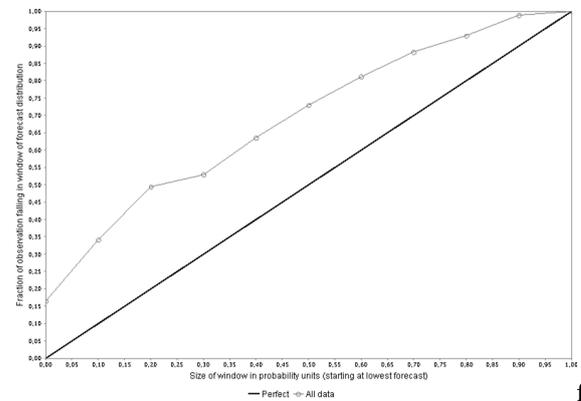
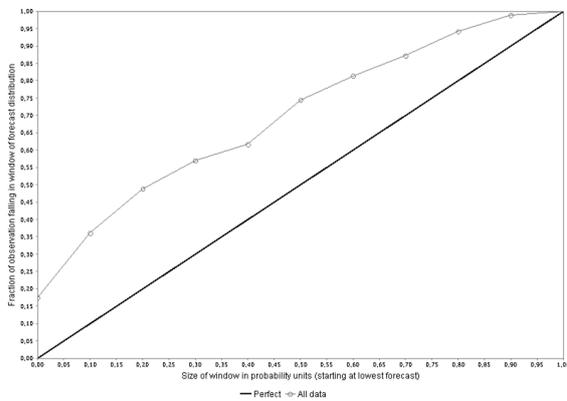
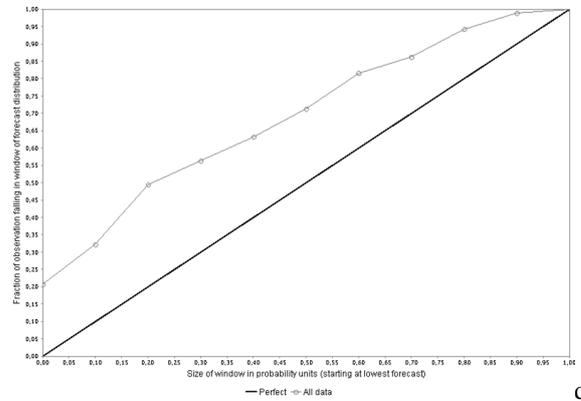
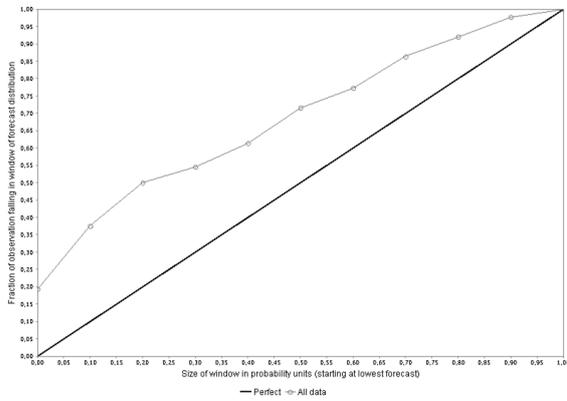
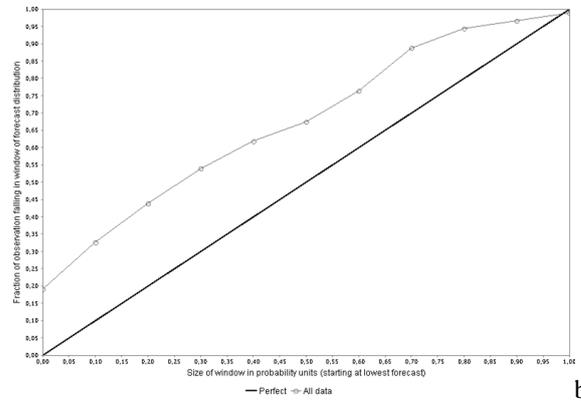
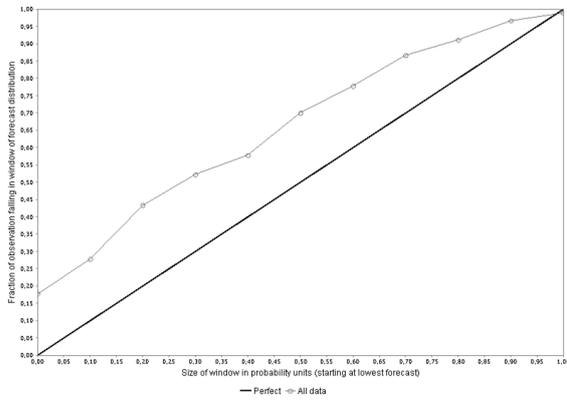
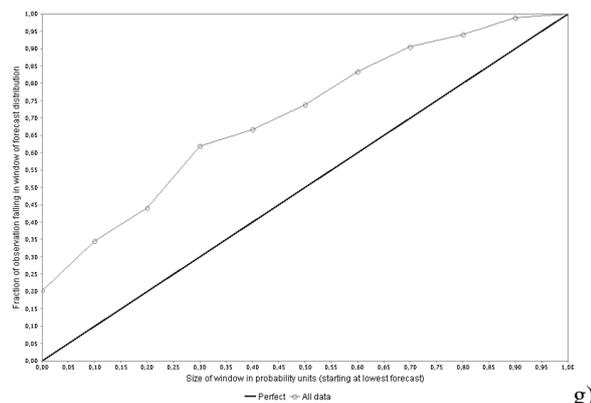


Figura 56 - Média dos membros previstos da precipitação para o horizonte de um mês de antecedência sem e com remoção do viés utilizando a média organizada de forma crescente dos membros da climatologia.

Os resultados obtidos com o método de verificação do *Spread-bias diagram* após a remoção do viés, utilizando a média organizada de forma crescente dos membros da climatologia (b), encontram-se na Figura 57. Pode-se perceber que, para todos os sete horizontes previstos, a linha que une os pontos está acima da linha de 45° do diagrama, indicando uma superestimativa da precipitação. Com base nessa configuração, pode-se concluir que a remoção do viés foi ineficiente.





g)

Figura 57 - *Spread-bias diagram* calculado para os dados de precipitação prevista com remoção do viés utilizando a média organizada de forma crescente dos membros da climatologia para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.

Os coeficientes calculados para avaliar o desempenho do MCG após a remoção do viés, utilizando a média organizada de forma crescente dos membros da climatologia (b), indicaram que o método utilizado não obteve resultados satisfatórios. Os resultados após a remoção obtiveram desempenho inferior aos obtidos antes da remoção. Portanto, pode-se concluir que a remoção do viés, utilizando a média organizada de forma crescente dos membros da climatologia (b), foi ineficiente.

Os resultados obtidos da remoção do viés, utilizando o membro um da climatologia (escolhido aleatoriamente entre os membros, pois apresentaram praticamente o mesmo comportamento), obtiveram resultados aproximadamente iguais aos obtidos pela remoção do viés utilizando a média organizada de forma crescente dos membros da climatologia (b). Esse resultado já era esperado, haja vista que as FCP's entre o membro um e a média organizada de forma crescente dos membros da climatologia (b) possuem um comportamento semelhante.

As tentativas de remoção do viés não foram bem sucedidas, obtendo resultados inferiores aos dados brutos. A metodologia de remoção do viés, utilizando a climatologia do MCG, para os dados dessa pesquisa, tende a superestimar os dados observados. Na Figura 58, encontra-se o diagrama de dispersão entre os membros previstos da climatologia e a precipitação observada. Pode-se perceber que existe uma maior concentração de pontos abaixo da linha de 45° do gráfico, indicando que os membros da climatologia do MCG subestimam os valores observados.

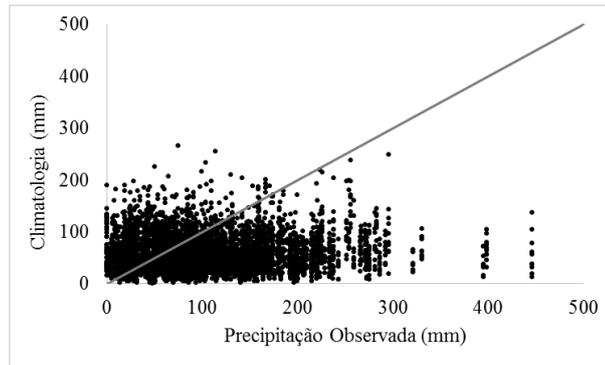


Figura 58 – Dispersão entre os valores dos membros da climatologia do MCG e os dados observados (1960 a 2002).

Quanto à escolha da climatologia utilizada para remoção do viés, a média organizada de forma crescente dos membros (b) e o membro um da climatologia obtiveram melhores resultados do que a média dos membros (a) da climatologia. Porém, para a discretização mensal, a média organizada de forma crescente dos membros (b) da climatologia perde relação entre os meses simulados. Assim, foi adotado o membro um como climatologia para remoção do viés mensal. Na Figura 59, pode-se perceber que as FCP's da média organizada de forma crescente dos membros (b) e o membro dez da climatologia são semelhantes, podendo-se concluir que os resultados obtidos serão similares.

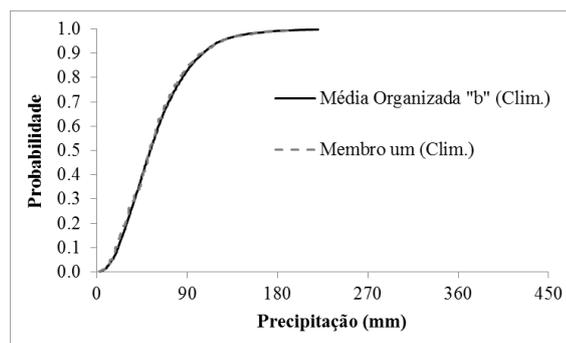
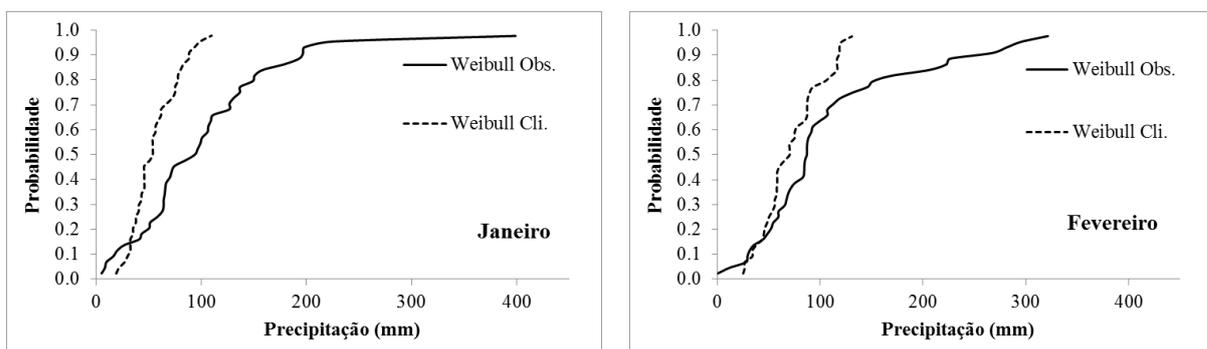


Figura 59 – Função cumulativa de probabilidades da média organizada de forma crescente dos membros (b) e o membro dez da climatologia (1960 a 2002).

Na Figura 60, encontra-se as FCP's da climatologia do MCG (membro um) e dos dados observados da precipitação discretizadas mensalmente.



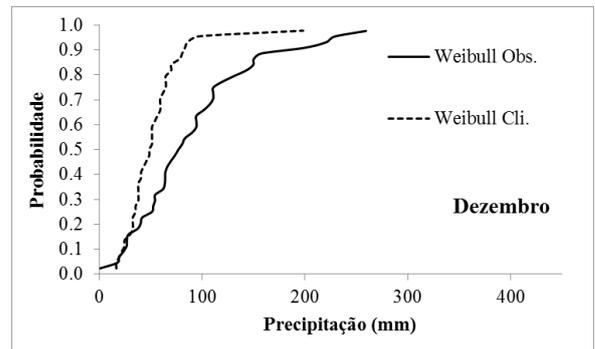
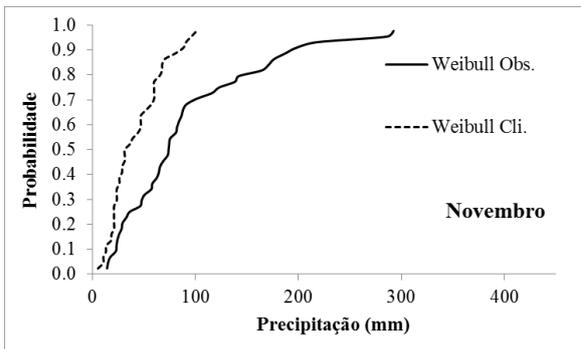
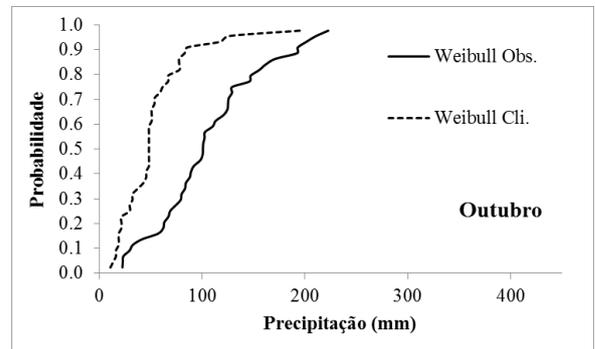
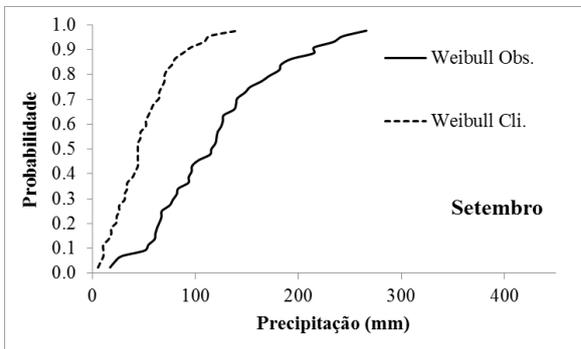
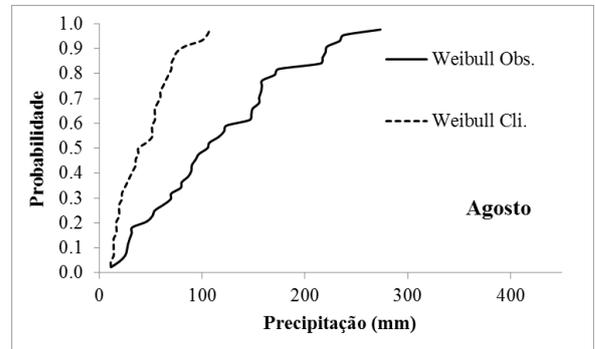
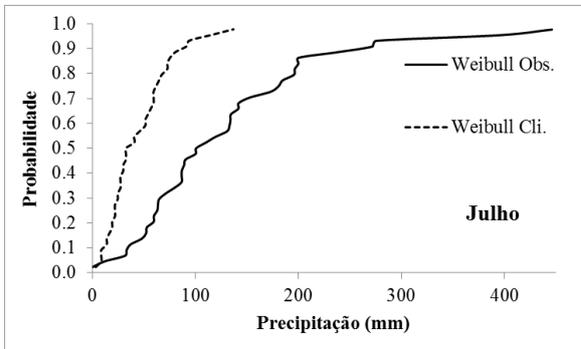
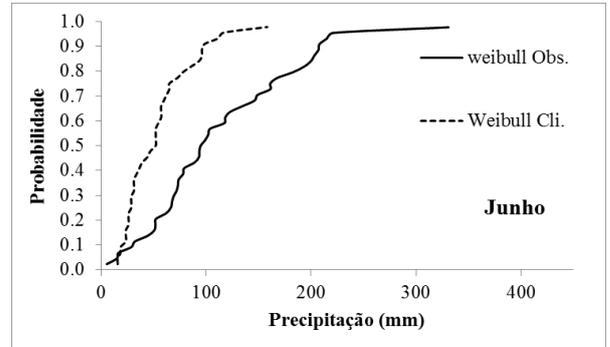
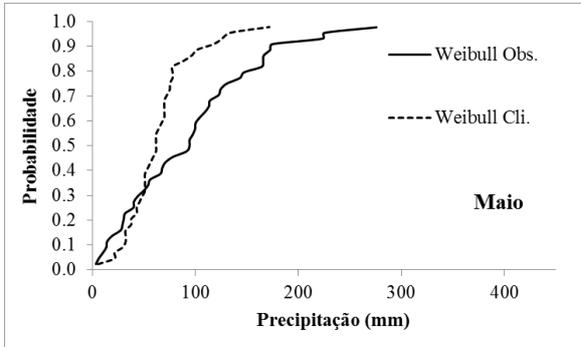
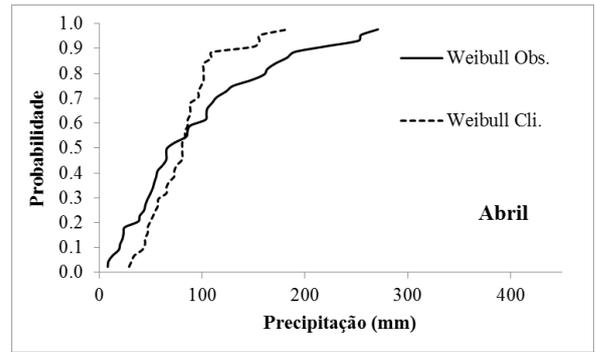
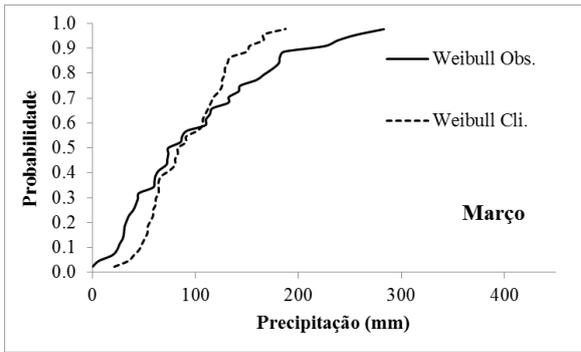


Figura 60 – Função cumulativa de probabilidades mensal da climatologia do MCG (membro um) e dos dados observados da precipitação.

Pode-se observar que alguns meses obtiveram FCP's com características semelhantes, com destaque para os meses de março e abril. Porém alguns meses revelaram uma subestimativa mais marcada, como se pode perceber nos meses de julho, agosto e setembro.

Os resultados da remoção do viés mensal, realizado para todos os meses do ano, encontram-se nas Figura 61 e Figura 62. Onde se pode perceber que os resultados obtiveram um desempenho inferior ao desempenho do modelo sem remoção do viés. Ocorrendo uma superestimativa mais acentuada que os outros métodos analisados até o momento.

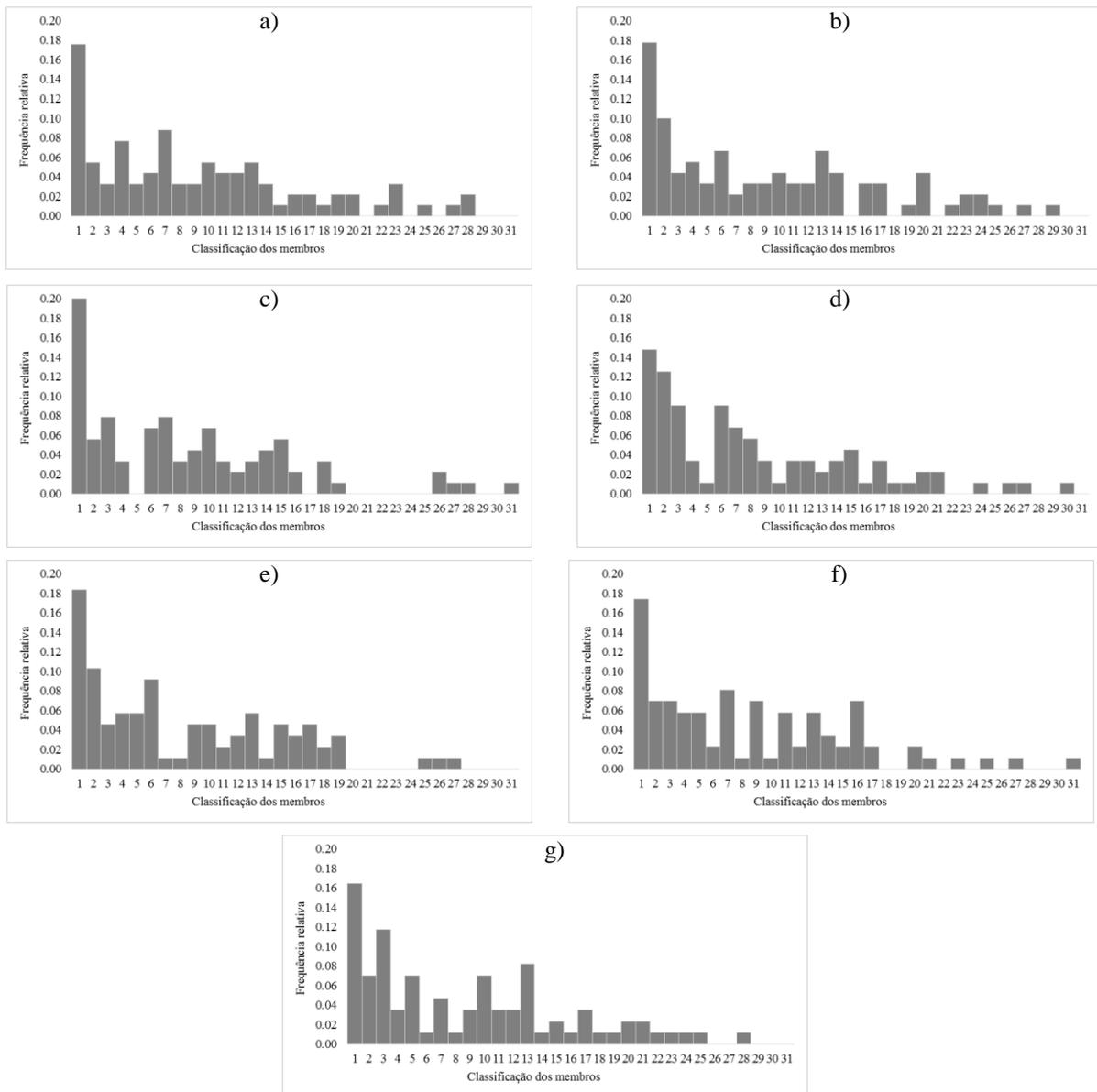


Figura 61 – Rank histogram calculado para os dados de precipitação prevista com remoção do viés mensal realizado para todos os meses para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.

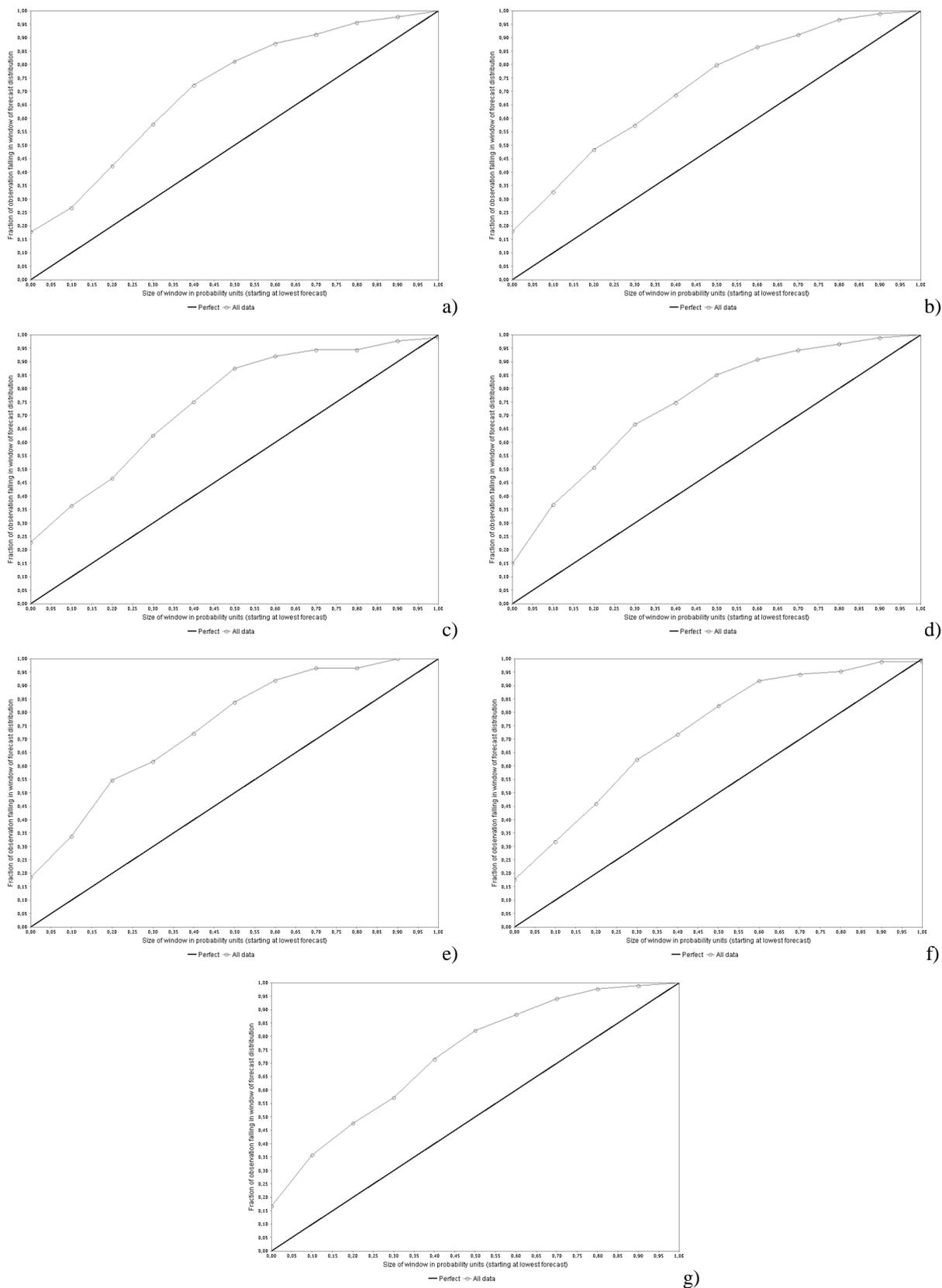


Figura 62 – *Spread-bias diagram* calculado para os dados de precipitação prevista com remoção do viés mensal realizado para todos os meses para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.

Não foi possível avaliar mensalmente as remoções do viés para cada mês de forma isolada, pois existem 30 membros previstos pelo MCG e apenas 43 dados da discretização mensal da climatologia do MCG. Logo, não seria possível avaliar o desempenho das remoções do viés mensal pelos coeficientes utilizados nesse estudo. Assim, foram adotados critérios para avaliação do desempenho da remoção do viés mensal para intervalos específicos.

O critério utilizado para a correção do viés mensal para meses definidos foi a análise dos resíduos entre os valores observados da precipitação e os dados da climatologia do MCG, para os respectivos meses do ano. Assim, a remoção do viés foi realizada apenas para os meses que obtivessem os menores resíduos, ou seja, a climatologia mensal fosse a mais próxima aos valores observados. Na Tabela 40, encontram-se os resultados da análise dos resíduos, onde se pode verificar a maior valor do resíduo, o menor valor, a média, o desvio padrão e o somatório dos valores absolutos dos resíduos.

Tabela 40 – Resíduos entre os valores observados e os dados da climatologia mensal.

	Maior Valor	Menor Valor	Média	Desvio Padrão	Somatório
Jan	289,19	-17,28	45,74	51,98	2118,69
Fev	190,86	-25,06	35,65	51,54	1633,73
Mar	95,51	-28,82	6,41	32,86	1050,05
Abr	98,88	-24,06	12,21	38,46	1279,95
Mai	104,38	-18,04	29,16	35,09	1499,13
Jun	172,89	-10,35	59,59	38,70	2587,95
Jul	309,70	-2,68	84,26	64,11	3628,46
Ago	166,26	0,69	70,83	43,98	3045,50
Set	129,05	11,82	69,76	30,84	2999,74
Out	109,17	7,13	54,04	25,17	2323,70
Nov	191,41	4,63	49,15	45,22	2113,53
Dez	134,89	-15,47	37,38	34,30	1640,80

Portanto, a remoção do viés mensal foi realizada para os meses de março, abril, maio e dezembro. Pois, apresentaram os menores valores das médias, desvio padrão e somatório dos valores absolutos respectivamente.

Na Figura 63, podem-se visualizar os resultados do desempenho do MCG após a remoção do viés mensal, realizado apenas para os meses de março, abril, maio e dezembro, avaliado pelo *Rank histogram*.

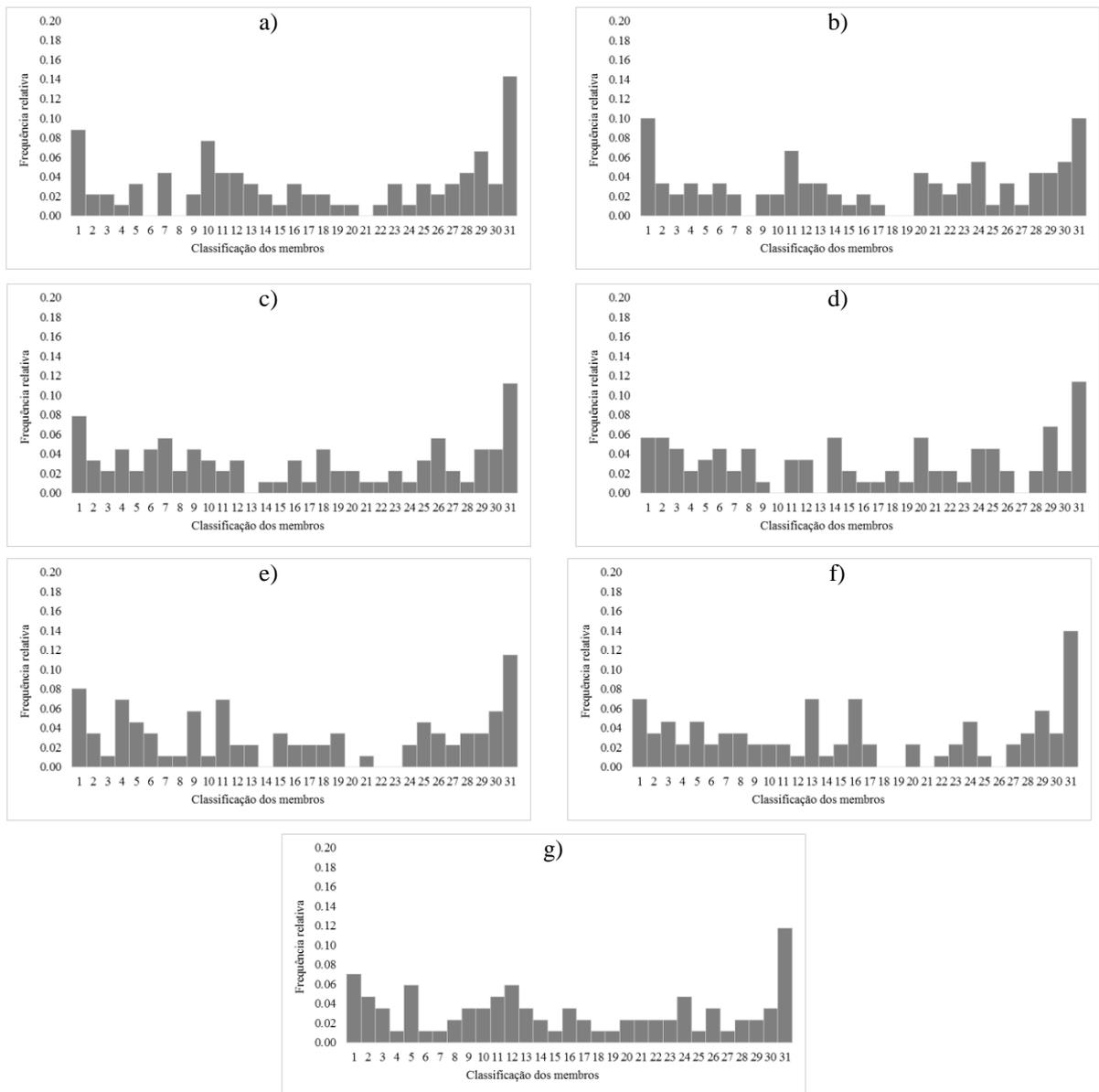


Figura 63 – *Rank histogram* calculado para os dados de precipitação prevista com remoção do viés mensal realizado para os meses de março, abril, maio e dezembro para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.

Pode-se perceber que houve uma melhor distribuição dos histogramas, onde os resultados possuem uma configuração próxima à configuração ideal (uniforme).

Na Figura 64, encontram-se os resultados do desempenho do MCG após a remoção do viés mensal, realizado para os meses de março, abril, maio e dezembro, avaliado pelo *Spread-bias diagram*.

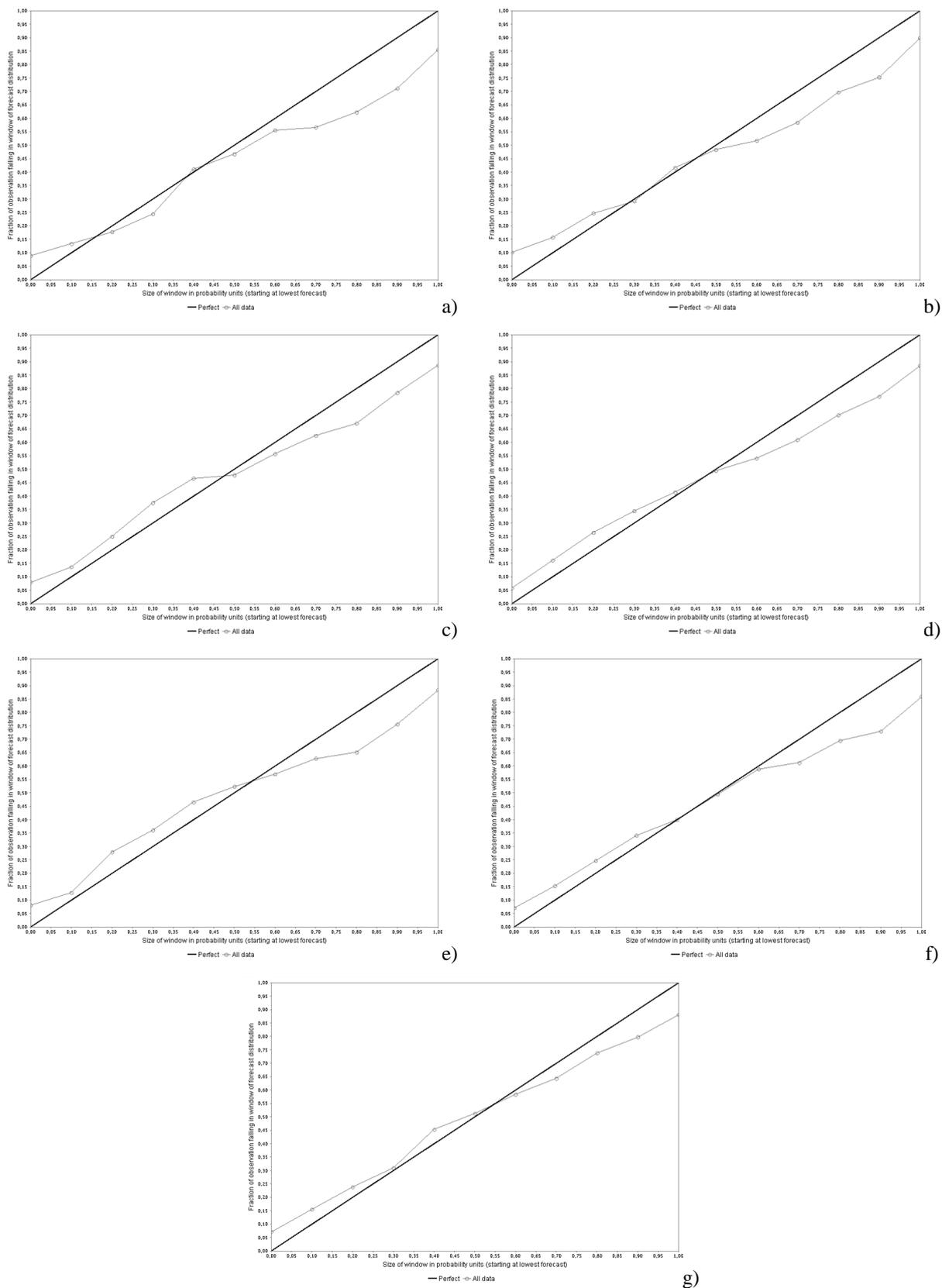


Figura 64 – *Spread-bias diagram* calculado para os dados de precipitação prevista com remoção do viés mensal realizado para os meses de março, abril, maio e dezembro para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.

Pode-se perceber que as linhas que unem os pontos encontram-se próximas à linha de 45° do gráfico, onde apenas para o horizonte de antecedência de um mês (Figura 64.a) observa-se uma defasagem acentuada na parte direita do gráfico. Essa defasagem consiste na dificuldade do modelo em representar os valores máximos para o referido horizonte.

Portanto, a remoção mensal realizada para os meses de março, abril, maio e dezembro, mostrou-se mais eficiente em termos qualitativos. Sendo selecionados para serem utilizados na realização das previsões utilizando as equações de regressão obtidas nessa pesquisa. Assim, os dados de precipitação prevista possuem os dados dos meses com remoção do viés mensal e os meses restantes sem remoção do viés.

7.4 Análises dos modelos empíricos

A metodologia proposta obteve como resultado oito equações de regressão linear múltipla. Nesse tópico, foram avaliados os resultados obtidos pelas oito equações utilizando os dados para os intervalos de verificação e simulação (previsão). A análise dos resultados foi realizada em três etapas:

- Análise das equações: nessa etapa encontram-se os resultados das análises das equações de regressão e as análises dos resíduos;
- Validação do modelo: nessa etapa encontram-se os resultados das simulações realizadas pelas oito equações de regressão elaboradas nesse estudo no período de validação (1960 a 1976);

A Figura 65 ilustra a divisão da análise dos resultados.

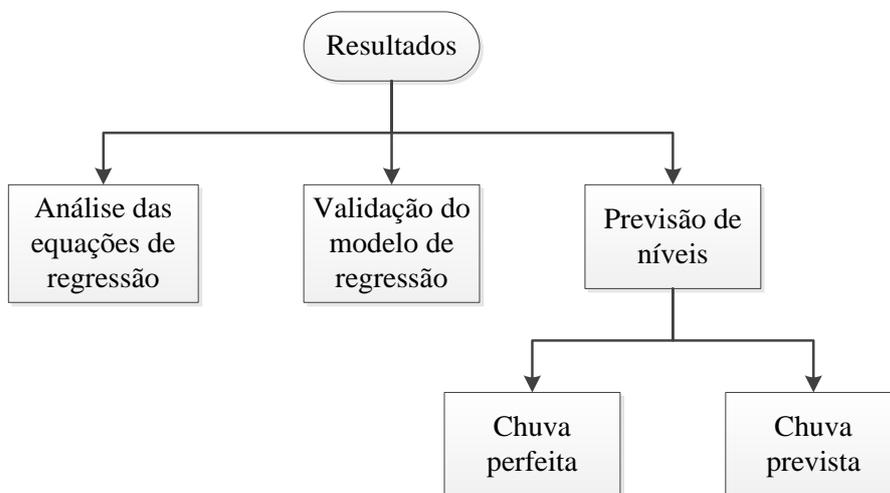


Figura 65 – Fluxograma da análise dos resultados.

7.4.1 Análise dos resíduos das equações de regressão linear múltipla

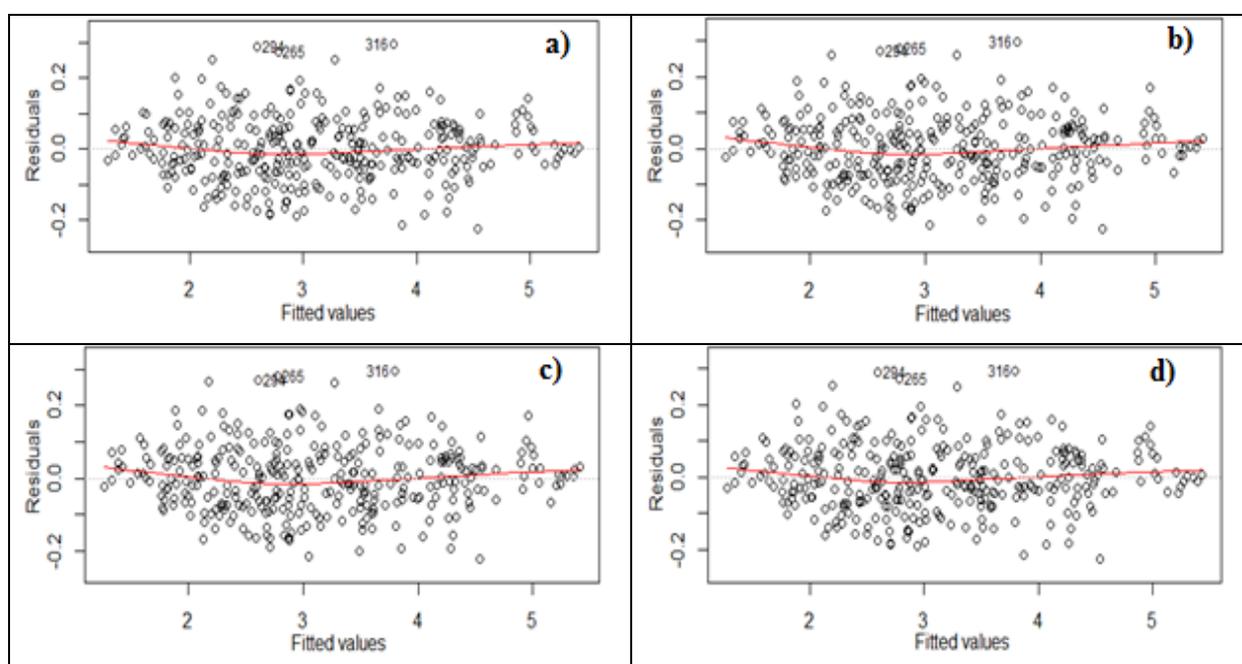
Para avaliar a adequabilidade dos modelos de regressão, foram analisados os resíduos entre os valores observados e os calculados por cada equação. Na Tabela 41, encontram-se as estatísticas dos resíduos das equações de regressão obtidas nesse estudo.

Tabela 41 – Estatística dos Resíduos das equações de regressão.

nº da equação	Mínimo	1Q	Média	Mediana	3Q	Máximo	Desvio Padrão	R ²
33	-0,22	-0,06	-0,01	-1,79E-18	0,06	0,30	0,09	0,99
35	-0,22	-0,06	-0,001	-1,56E-18	0,05	0,30	0,09	0,99
87	-0,22	-0,06	0,001	-1,47E-18	0,06	0,29	0,09	0,99
90	-0,22	-0,06	-0,01	-1,25E-18	0,06	0,29	0,09	0,99
91	-0,22	-0,06	-0,01	-4,89E-18	0,06	0,30	0,09	0,99
162	-0,22	-0,06	-0,01	1,56E-18	0,06	0,29	0,09	0,99
195	-0,21	-0,06	0,001	1,69E-18	0,05	0,30	0,09	0,99
197	-0,22	-0,06	-0,0003	-1,64E-18	0,05	0,30	0,09	0,99

Pode-se perceber que a média e a mediana das equações são aproximadamente zero. O coeficiente de determinação mostra um bom ajuste dos resíduos, e assim, um bom ajuste do modelo linear.

Na Figura 66, encontra-se o gráfico dos resíduos vs. valores ajustados pelas equações de regressão.



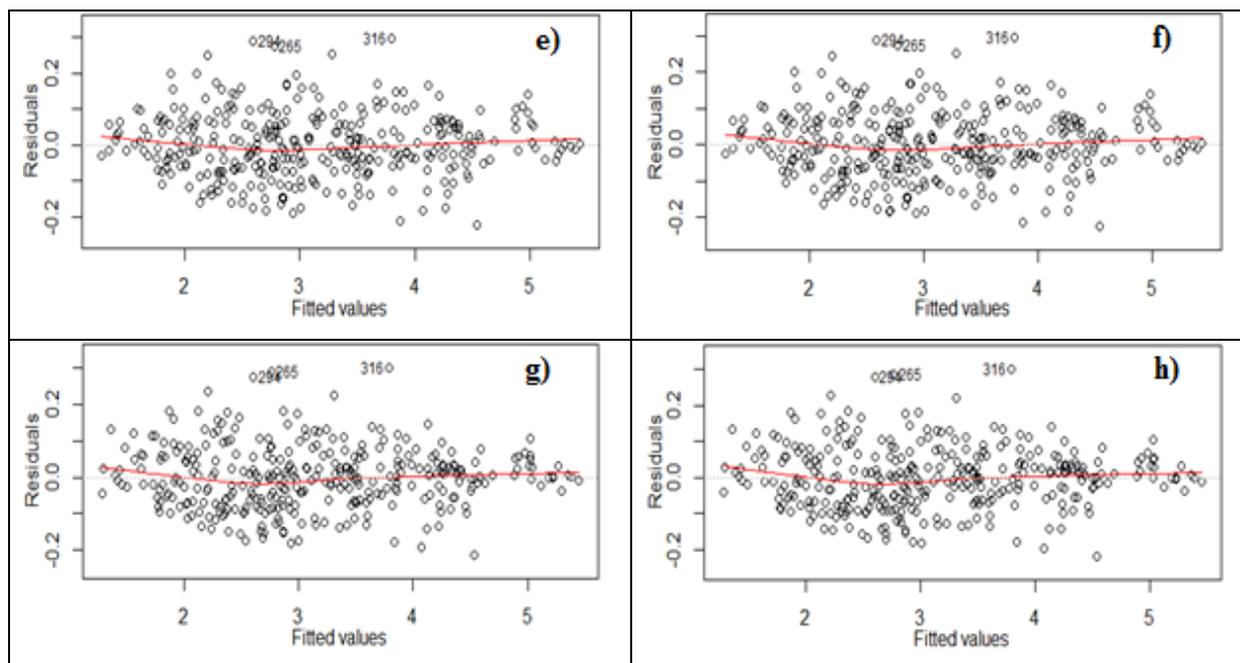


Figura 66 – Gráfico dos resíduos vs. valores ajustados, para as equações de nº: a) 33; b) 35; c) 87; d) 90; e) 91; f) 162; g) 195; h) 197.

Pode-se perceber uma leve curvatura ao ajuste horizontal nos gráficos, porém é correto afirmar que um modelo linear é capaz de representar os dados.

7.4.2 Validação dos modelos

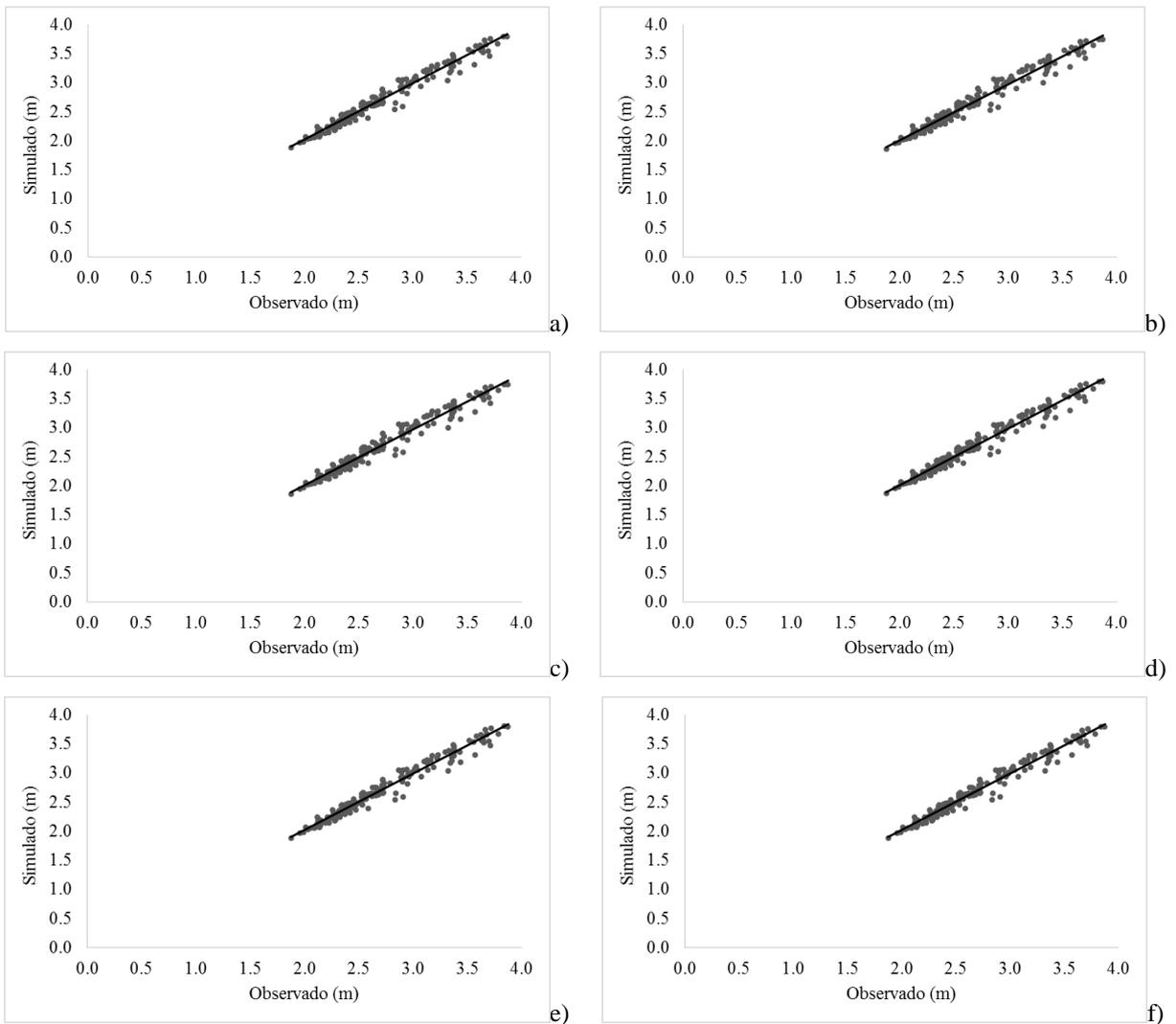
Essa é a etapa de validação do modelo, onde as equações de regressão foram utilizadas para simularem os níveis no banhado para o intervalo de verificação. As equações foram testadas para um intervalo diferente do que foram formuladas, utilizando os dados observados. Na Tabela 42, encontram-se os resultados das medidas de desempenho calculadas para mensurar a eficiência das equações, avaliadas em caráter determinístico.

Tabela 42 – Medidas de desempenho calculadas para as equações de regressão.

	33	35	87	90	91	162	195	197
Erro médio	-0,007	-0,019	-0,022	-0,009	-0,005	-0,007	0,002	0,002
Erro médio absoluto	0,06	0,07	0,07	0,06	0,06	0,06	0,05	0,05
Erro médio quadrático	0,007	0,009	0,009	0,008	0,007	0,007	0,006	0,006
Raiz quadrada do erro médio	0,09	0,09	0,10	0,09	0,09	0,09	0,07	0,07
Coefficiente de Nash-Sutcliffe	0,97	0,96	0,96	0,97	0,97	0,97	0,98	0,98
Coefficiente de persistência	0,71	0,66	0,65	0,71	0,71	0,72	0,79	0,78
Correlação de Pearson	0,98	0,98	0,98	0,98	0,98	0,98	0,99	0,99
Coefficiente de determinação	0,97	0,96	0,96	0,97	0,97	0,97	0,98	0,98

Os quatro primeiros índices, que mensuram os erros associados as simulações, obtiveram valores baixos, indicando que não ocorreram erros de grande magnitude. Os coeficientes de Nash-Sutcliffe e Persistência obtiveram resultados próximos a unidade, indicando uma boa eficiência das equações em simulares os níveis para o intervalo avaliado. Os coeficientes de Pearson e Determinação obtiveram resultados próximos a unidade, indicando que existe um bom ajuste (linear) entre os valores simulados e observados.

Na Figura 67, encontram-se os diagramas de dispersão entre os dados observados e simulados pelas equações.



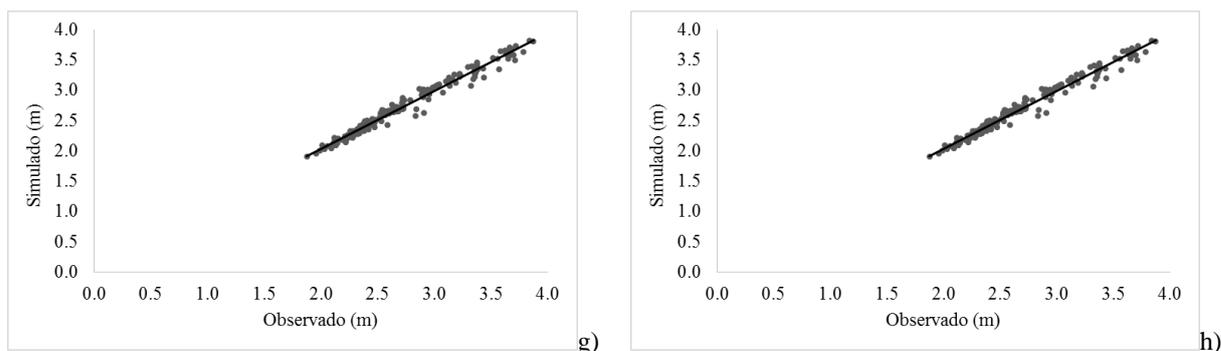


Figura 67 – Diagrama de dispersão entre os dados observados e simulados para o intervalo de verificação das equações: a) 33; b) 35; c) 87; d) 90; e) 91; f) 162; g) 195; h) 197.

Dos resultados obtidos, pode-se perceber que as oito equações possuem desempenho bastante similares, com desempenho sensivelmente melhor para as duas últimas equações.

7.5 Previsões de longo prazo de níveis

Conforme foi demonstrado no tópico anterior, as equações possuem desempenho similares e resultados variando a partir da segunda casa decimal, após a vírgula. Por esse motivo, serão demonstrados os resultados para uma única equação. Assim, foi escolhida a equação nº 197 (equação 73) para serem analisados os resultados das previsões. Contudo, cabe ressaltar que todas as equações foram avaliadas e os resultados para todos os horizontes propostos foram similares. Nessa etapa encontram-se os resultados das previsões de níveis no intervalo entre 2004 a 2011.

$$N_t = -0,02 + 0,002.P_t - 0,0007.E_t + 0,0005.MMPE7_t + 0,0003.SavP_t - 0,0001.SavE_t + 0,96.N_{t-1} - Irri \quad (73)$$

7.5.1 Previsões utilizando chuva perfeita

Na Tabela 43, encontram-se os resultados das medidas de desempenho calculadas para os resultados obtidos das previsões realizadas pela equação nº 197, utilizando a precipitação perfeita, avaliadas em caráter determinístico.

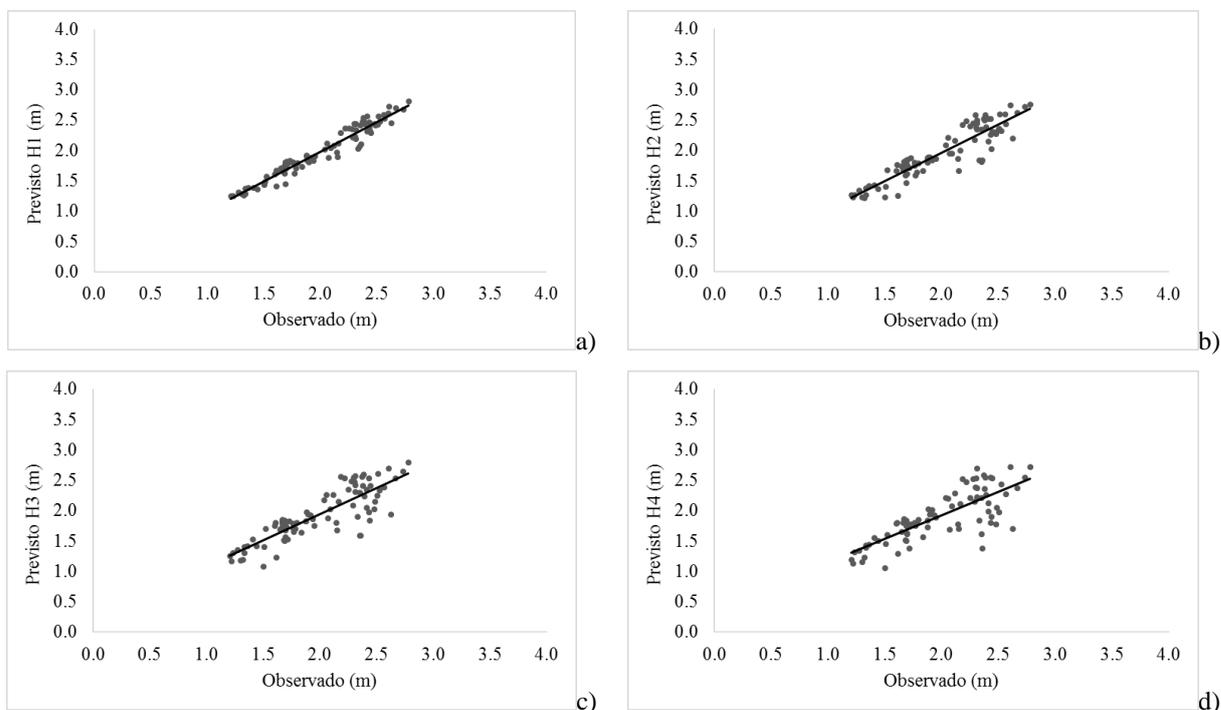
Tabela 43 - Medidas de desempenho calculadas para as previsões de níveis utilizando a precipitação perfeita.

	Horizontes de previsão (mês)						
	1	2	3	4	5	6	7
Erro médio	-0,02	-0,04	-0,06	-0,08	-0,10	-0,12	-0,14
Erro médio absoluto	0,08	0,13	0,17	0,20	0,21	0,23	0,25
Erro médio	0,01	0,03	0,06	0,08	0,10	0,12	0,14

quadrático							
Raiz quadrada do erro médio	0,10	0,18	0,24	0,28	0,32	0,34	0,37
Coefficiente de Nash-Sutcliffe	0,94	0,81	0,67	0,54	0,42	0,31	0,21
Coefficiente de persistência	0,54	-0,39	-1,40	-2,30	-3,11	-3,85	-4,69
Correlação de Pearson	0,97	0,91	0,85	0,78	0,72	0,67	0,61
Coefficiente de determinação	0,94	0,83	0,72	0,62	0,52	0,44	0,38

Pode-se perceber que o desempenho do modelo vai decaindo, à medida que o horizonte de previsão vai aumentando. A partir do horizonte de três meses de antecedência, o modelo reduz significativamente seu desempenho, porém, cabe ressaltar que ao se trabalhar com previsões de longo prazo, é de se esperar que o desempenho reduza com o aumento do horizonte, logo, pode-se considerar que os resultados são relativamente bons, possuindo erros aceitáveis.

Na Figura 68, encontram-se os diagramas de dispersão entre os níveis observados e previstos pela equação nº 197 utilizando a precipitação perfeita para os sete horizontes propostos. Pode-se perceber que, à medida que o horizonte vai aumentando, os pontos vão se distanciando da reta de ajuste.



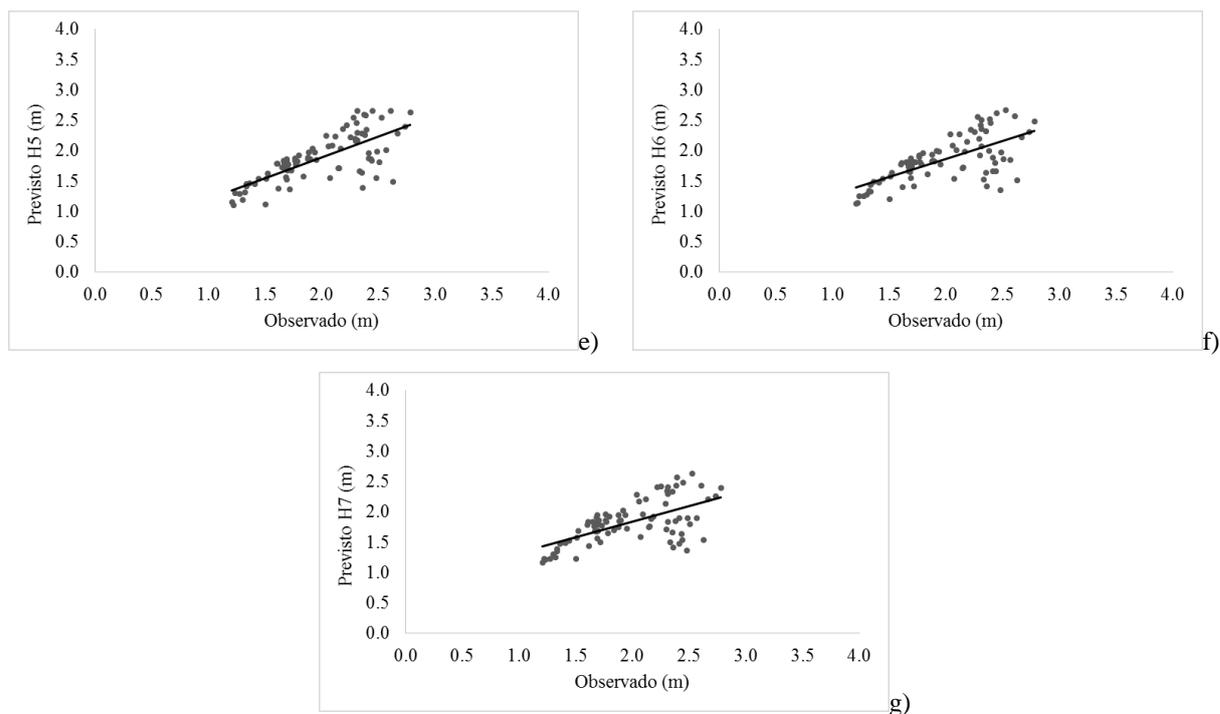


Figura 68 – Diagrama de dispersão entre os níveis observados e previstos pela equação n° 197 utilizando a precipitação perfeita para os horizontes de: a) um mês; b) dois meses; c) três meses; d) quatro meses; e) cinco meses; f) seis meses; g) sete meses.

7.5.2 Previsões utilizando chuva prevista

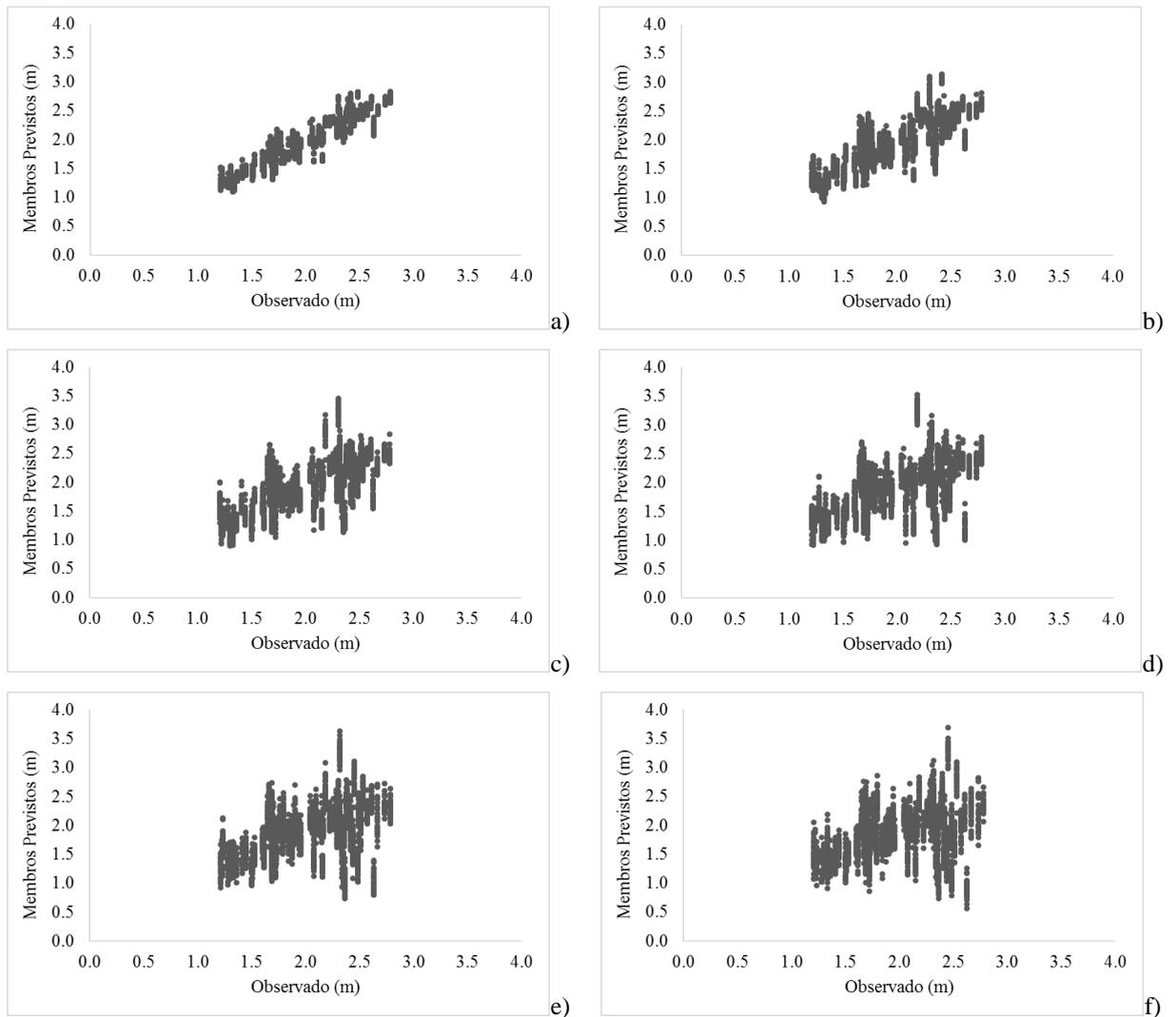
Utilizando a média dos membros previstos pelo modelo hidrológico, calculou-se as medidas de desempenho para uma avaliação determinística. Na Tabela 44, encontram-se os resultados das medidas de desempenho calculadas para os resultados obtidos das previsões realizadas pela equação n° 197, utilizando a precipitação prevista pelo MCG.

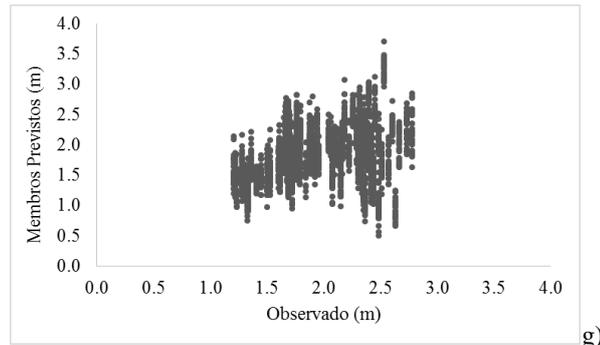
Tabela 44 - Medidas de desempenho para equação de n° 197.

	Horizontes de previsão (mês)						
	1	2	3	4	5	6	7
Erro médio	-0,02	-0,05	-0,06	-0,08	-0,10	-0,11	-0,13
Erro médio absoluto	0,11	0,18	0,24	0,28	0,30	0,33	0,36
Erro médio quadrático	0,02	0,06	0,11	0,15	0,19	0,22	0,24
Raiz quadrada do erro médio	0,14	0,25	0,33	0,39	0,43	0,47	0,49
Coefficiente de Nash-Sutcliffe	0,88	0,64	0,36	0,10	-0,10	-0,28	-0,42
Coefficiente de persistência	0,11	-1,68	-3,67	-5,44	-6,82	-7,98	-9,20
Correlação de Pearson	0,94	0,83	0,70	0,58	0,48	0,40	0,33
Coefficiente de determinação	0,89	0,69	0,49	0,34	0,23	0,16	0,11

Pode-se perceber que, a partir do horizonte de três meses de antecedência, as medidas de desempenho indicaram uma redução significativa da eficiência do modelo. Para um horizonte de sete meses no futuro, pode-se considerar que as medidas de desempenho que mensuram o erro associado as previsões obtiveram resultados satisfatórios. Os coeficientes de Nash-Sutcliffe e Persistência, demonstraram que, a partir do horizonte de três meses de antecedência, o modelo reduziu de forma significativa sua capacidade de previsão.

Na Figura 69, pode-se visualizar os diagramas de dispersão entre os dados observados e previstos pela equação nº 197, utilizando a precipitação prevista pelo MCG.



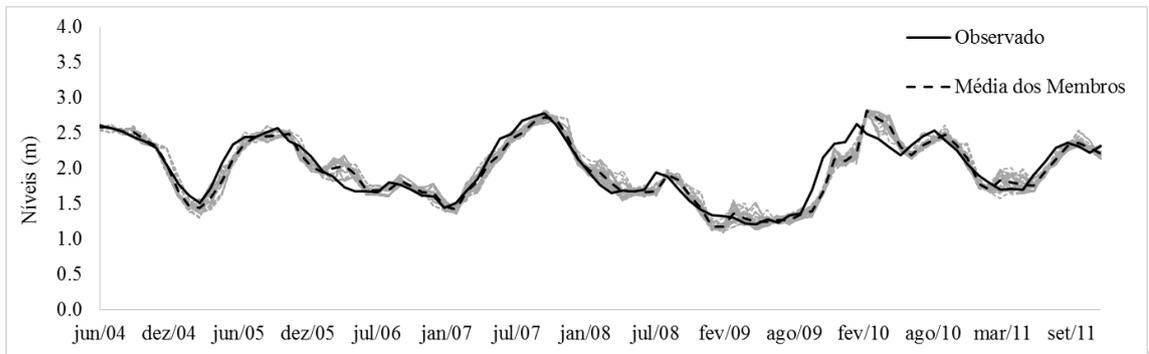


g)

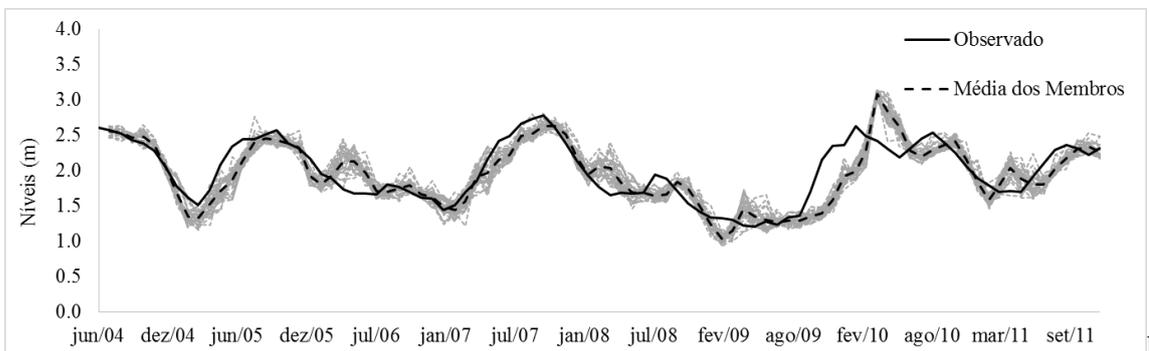
Figura 69 – Diagrama de dispersão entre os níveis observados e previstos pela equação de nº 197 para os horizontes de: a) um mês; b) dois meses; c) três meses; d) quatro meses; e) cinco meses; f) seis meses; g) sete meses.

Pode-se perceber que, à medida que o horizonte de previsão aumenta, a propagação dos erros vai aumentando de forma acentuada.

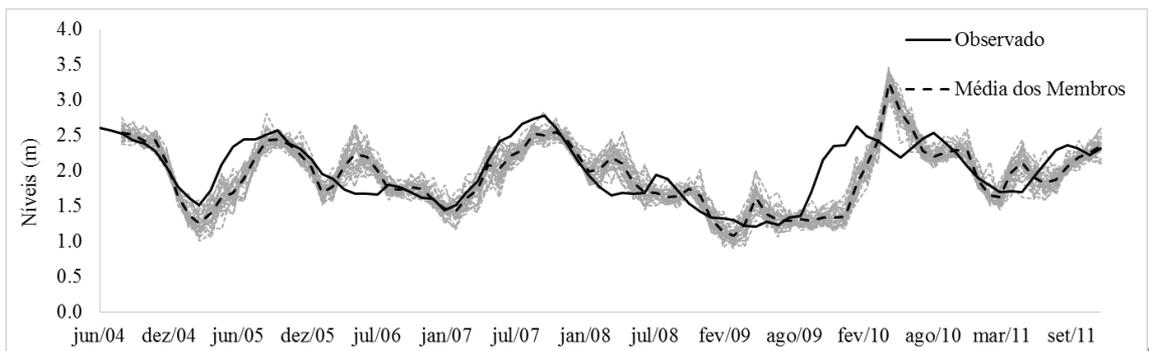
Na Figura 70, encontram-se os gráficos de linha para cada horizonte previsto.



a)



b)



c)

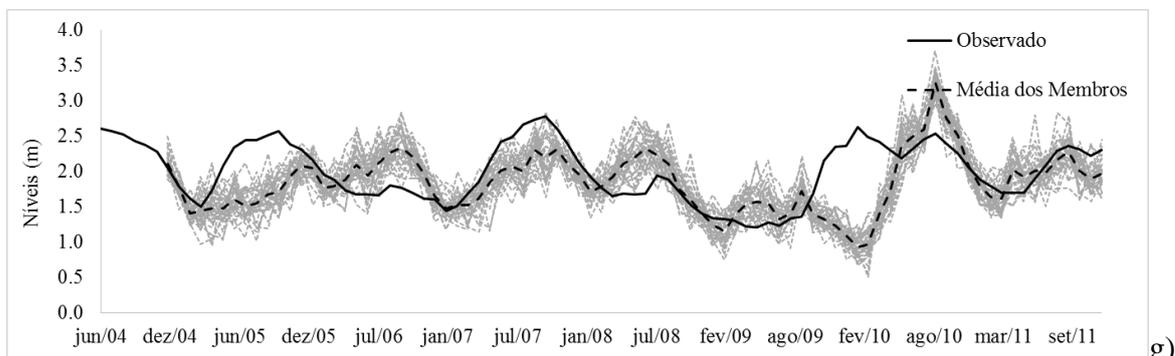
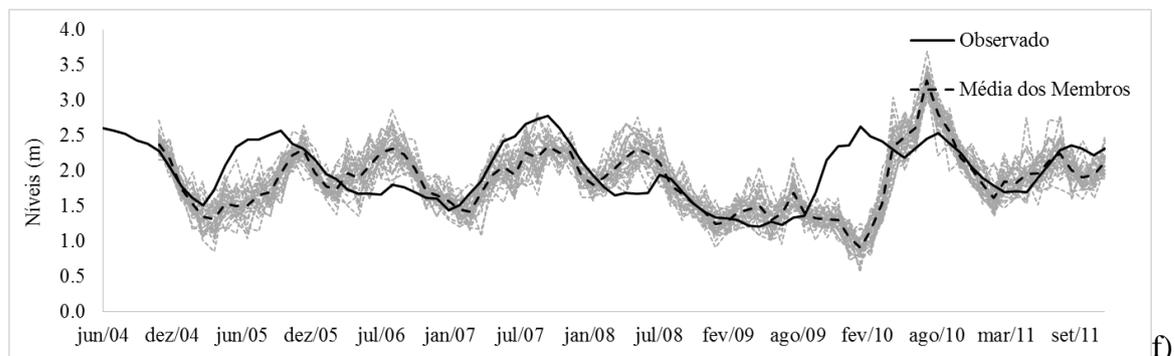
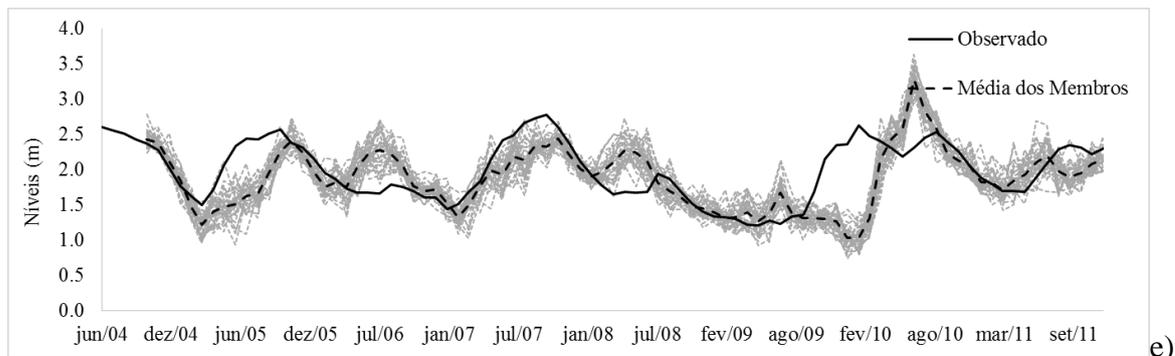
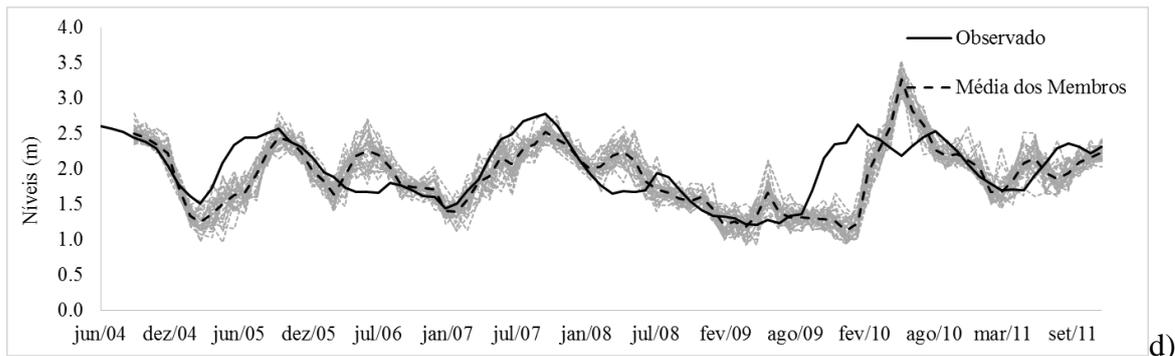


Figura 70 – Gráficos dos níveis previstos pela equação de nº 197 para os horizontes de: a) um mês; b) dois meses; c) três meses; d) quatro meses; e) cinco meses; f) seis meses; g) sete meses.

Pode-se perceber que, à medida que o horizonte de previsão vai aumentando, os erros advindos da previsão anterior vão se propagando acentuadamente.

Na Figura 71, encontram-se os resultados obtidos utilizando o *rank histogram* para avaliar o desempenho do modelo para os sete horizontes previstos.

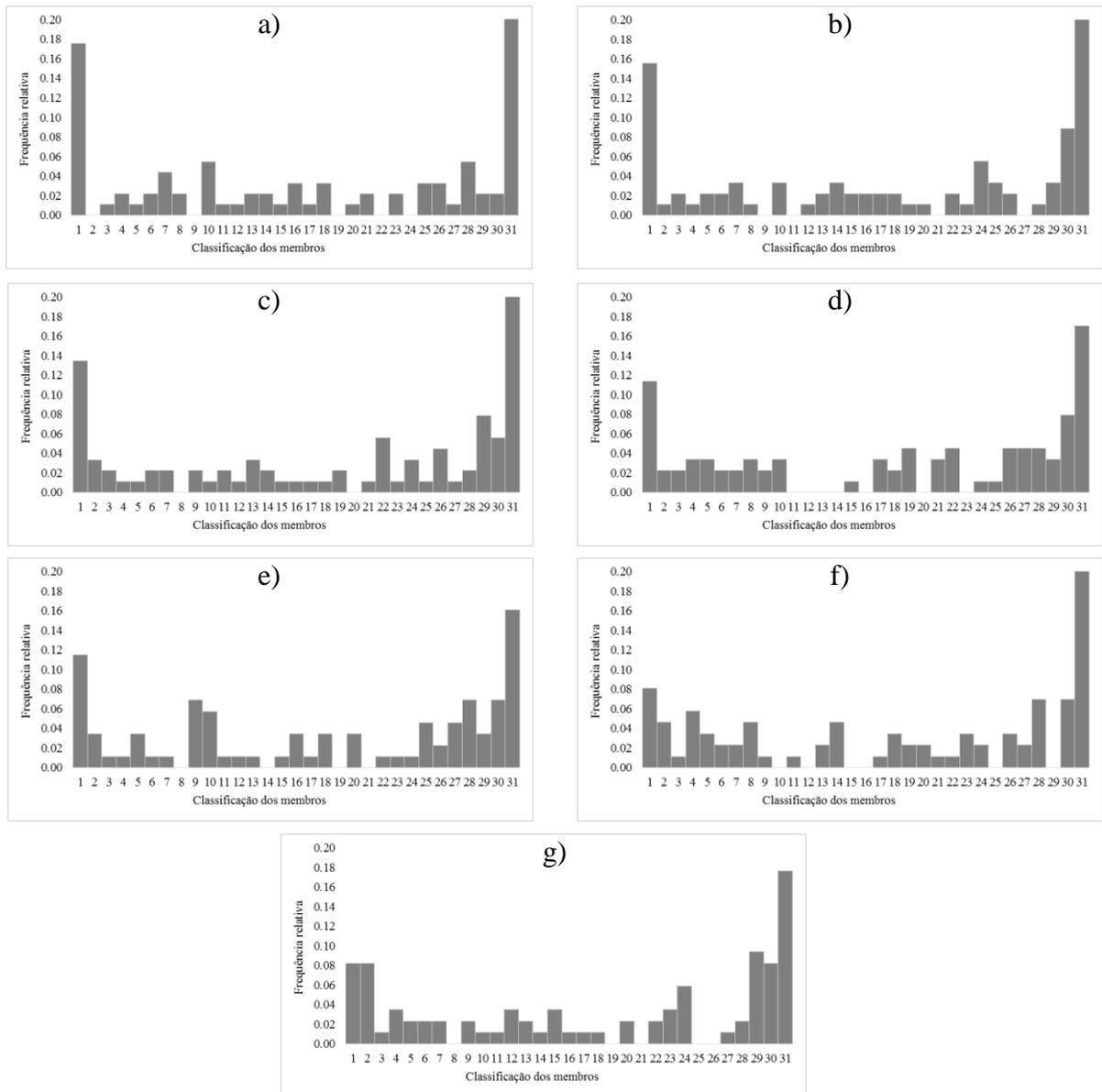


Figura 71 – *Rank histogram* calculado para membros previstos dos níveis para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.

Pode-se perceber que existe uma maior concentração de dados entre os extremos dos histogramas, com uma configuração em forma de “U”. Essa configuração indica um subespalhamento dos dados, onde o modelo não foi capaz de representar de forma eficiente os extremos observados.

Na Figura 72, encontram-se os resultados obtidos utilizando o *Spread-bias diagram* para avaliar o desempenho do modelo para os sete horizontes previstos.

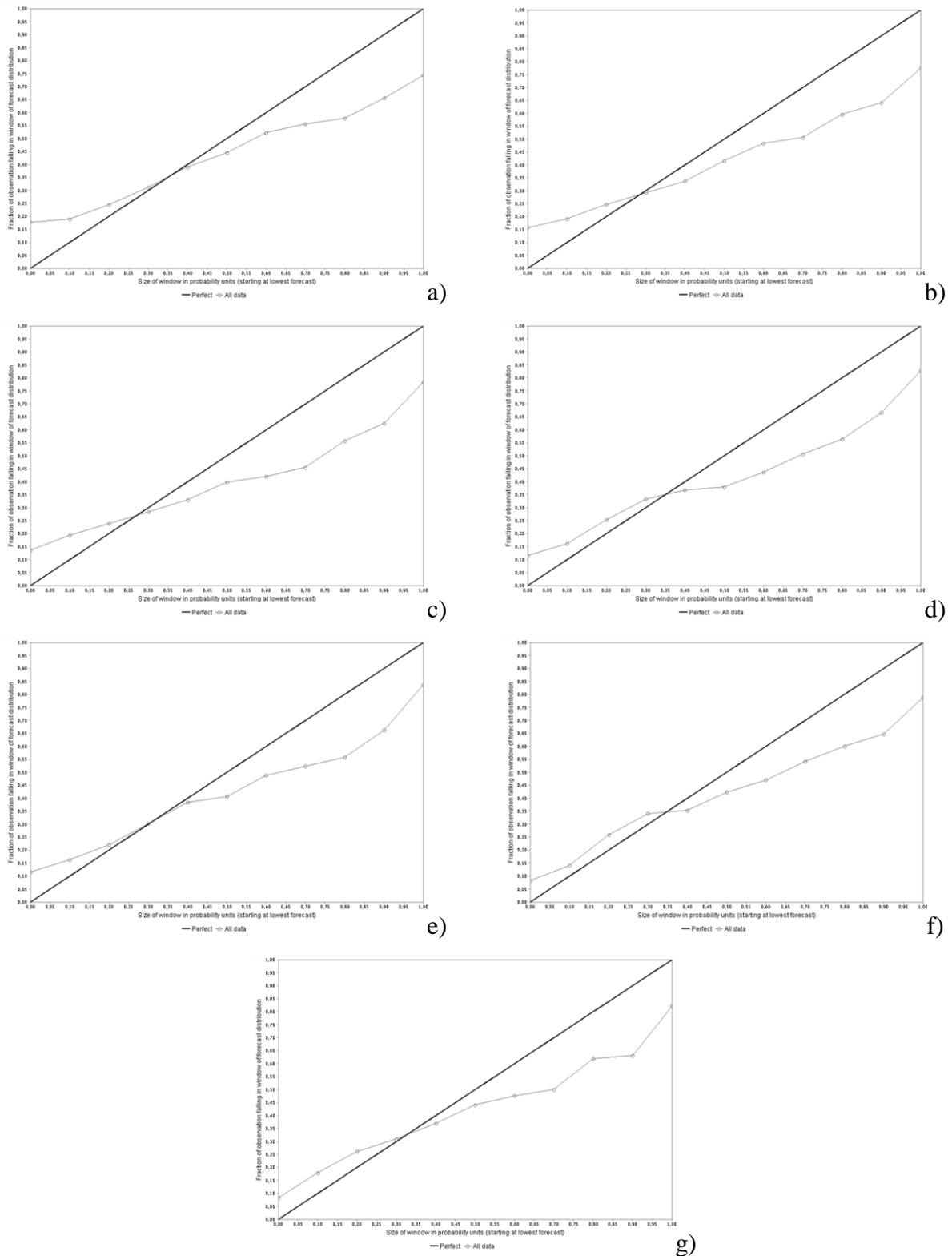


Figura 72 – *Spread-bias diagram* calculado para membros previstos dos níveis para os horizontes de antecedência de: a) um mês; b) dois meses; c) três meses; d) quarto meses; e) cinco meses; f) seis meses; g) sete meses.

Pode-se perceber que em todos os gráficos existe um subespalhamento da distribuição, onde a linha que une os pontos encontra-se em sua maior parte, abaixo da linha de 45° do gráfico, indicando uma tendência em subestimar os dados observados.

Até o presente momento, as avaliações foram realizadas para todo o intervalo de previsão (2004 a 2011), porém, conforme discutido nos trabalhos de Tassi (2008) e Cardoso (2010), os meses de interesse para a gestão, são os meses de dezembro, janeiro, fevereiro e março, que são os meses em que são realizadas as retiradas para irrigação. Por esse motivo, foram avaliadas as previsões obtidas nesse estudo para os meses em que são realizadas as retiradas para irrigação.

Devido a quantidade limitada de dados, os cálculos das medidas de desempenho foram tendenciosos, e obtiveram resultados acima do esperado para previsões de longo prazo. Assim, optou-se pela avaliação visual dos gráficos para analisar a precisão do modelo. A Figura 73, ilustra como estão dispostos os horizontes de previsão, onde cada ponto da linha prevista, significa um horizonte de previsão.

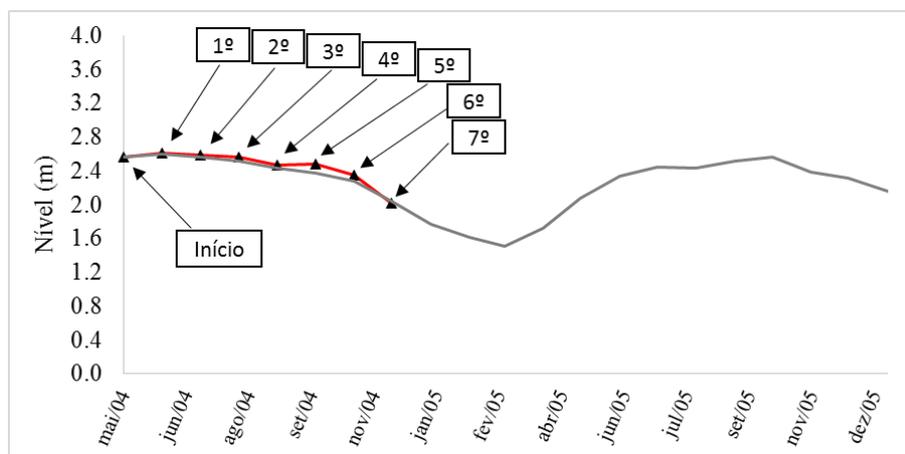


Figura 73 – Horizontes de previsão.

Nas Figura 74 a Figura 81 apresentam a sequência de previsões com sete meses de horizonte executadas a partir de maio até março, considerando as previsões com precipitação perfeita e as previsões de precipitação do modelo COLA. As quatro primeiras figuras, constituem as previsões para os horizontes máximos (sete meses) para os meses em que são realizadas as retiradas de água. As demais figuras, são a evolução das previsões ao longo do tempo, até o horizonte de um mês de antecedência para o último mês em que ocorrem as retiradas, ou seja, até março. É importante salientar que a linha vermelha constitui as previsões utilizando chuva perfeita, isso significa dizer que essa linha representa a melhor previsão que pode-se alcançar, melhorando-se (ou corrigir) as previsões de precipitação.

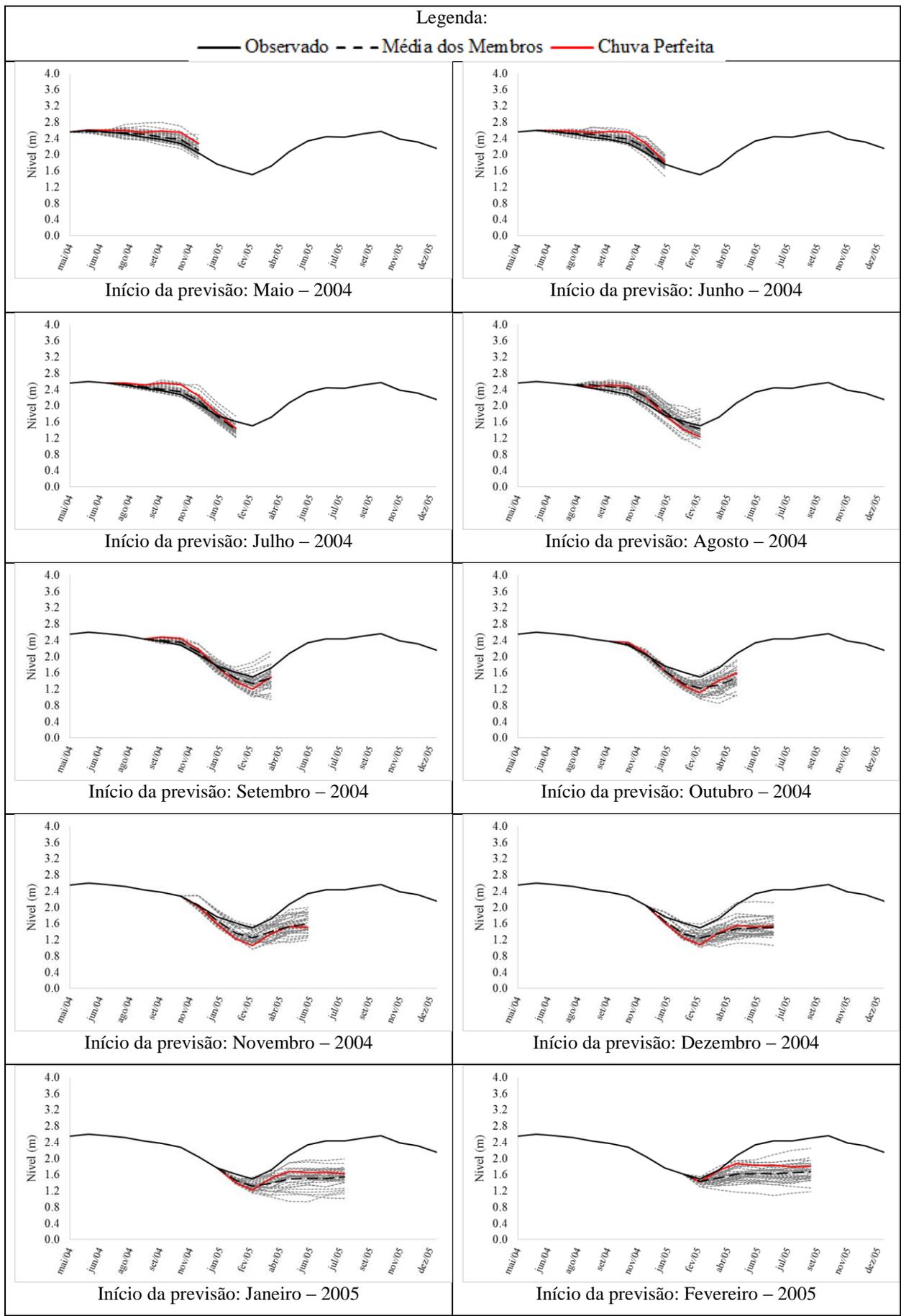


Figura 74 – Previsões iniciando em maio de 2004 a março de 2005.

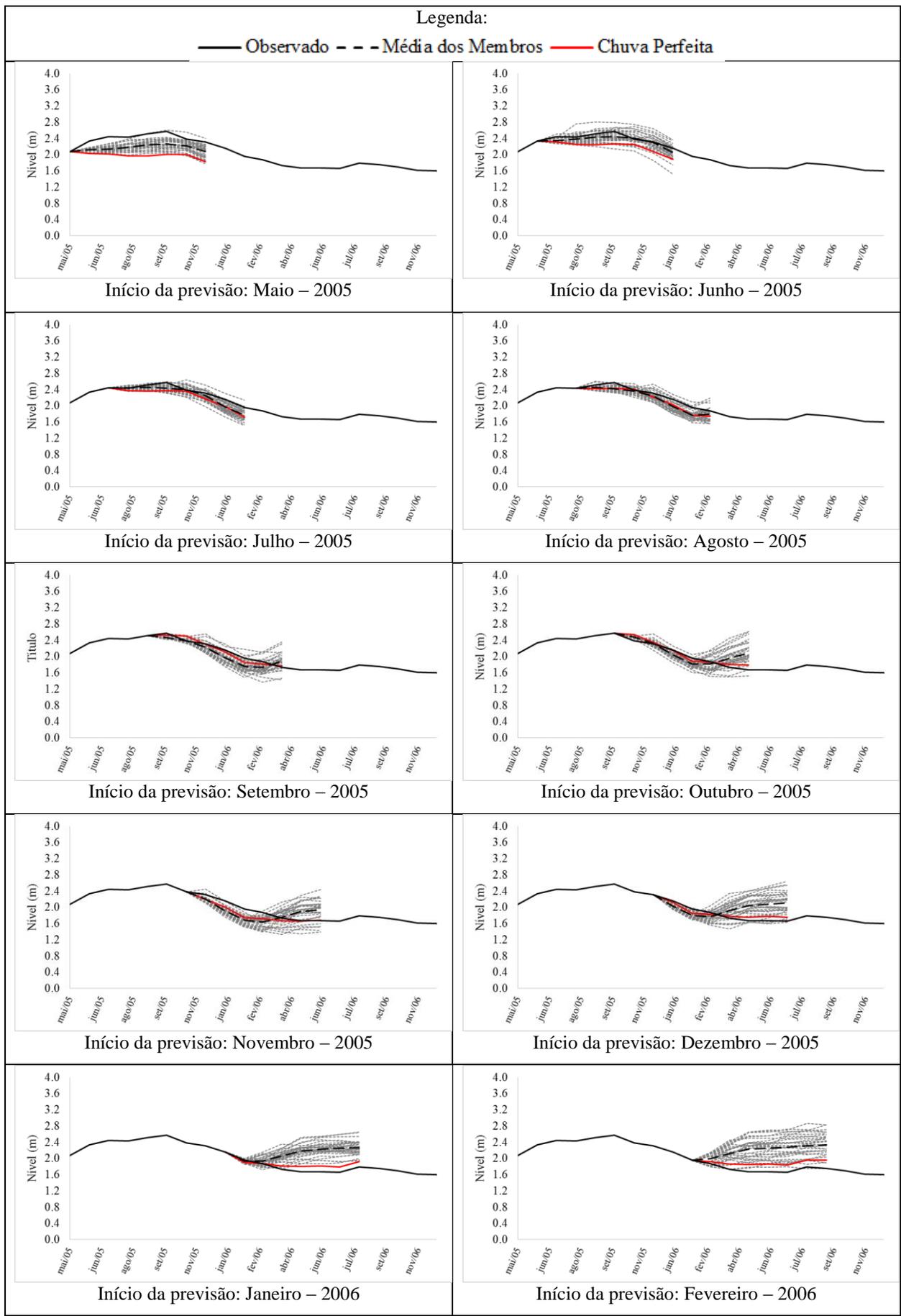


Figura 75 – Previsões iniciando em maio de 2005 a março de 2006.

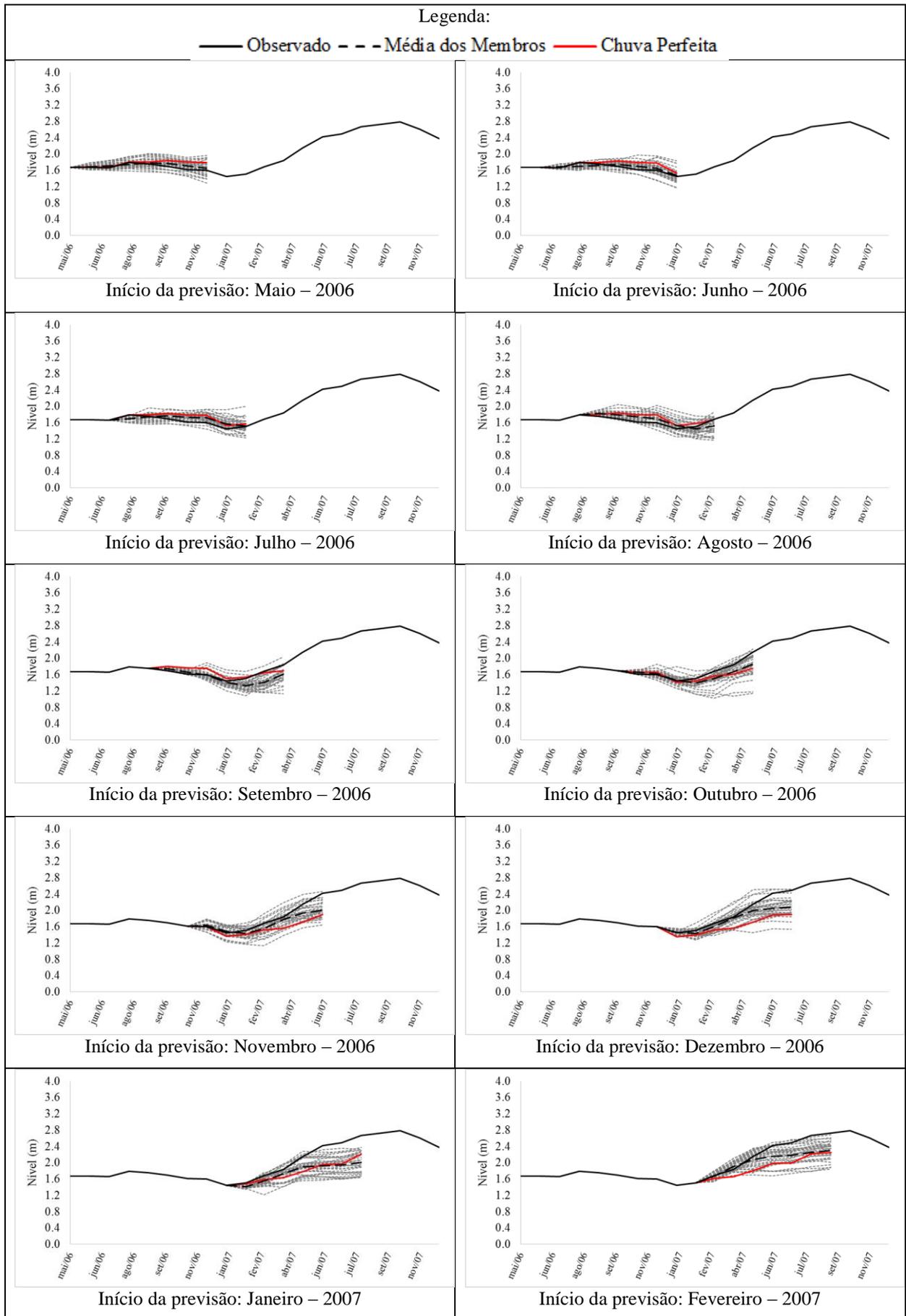


Figura 76 – Previsões iniciando em maio de 2006 a março de 2007.

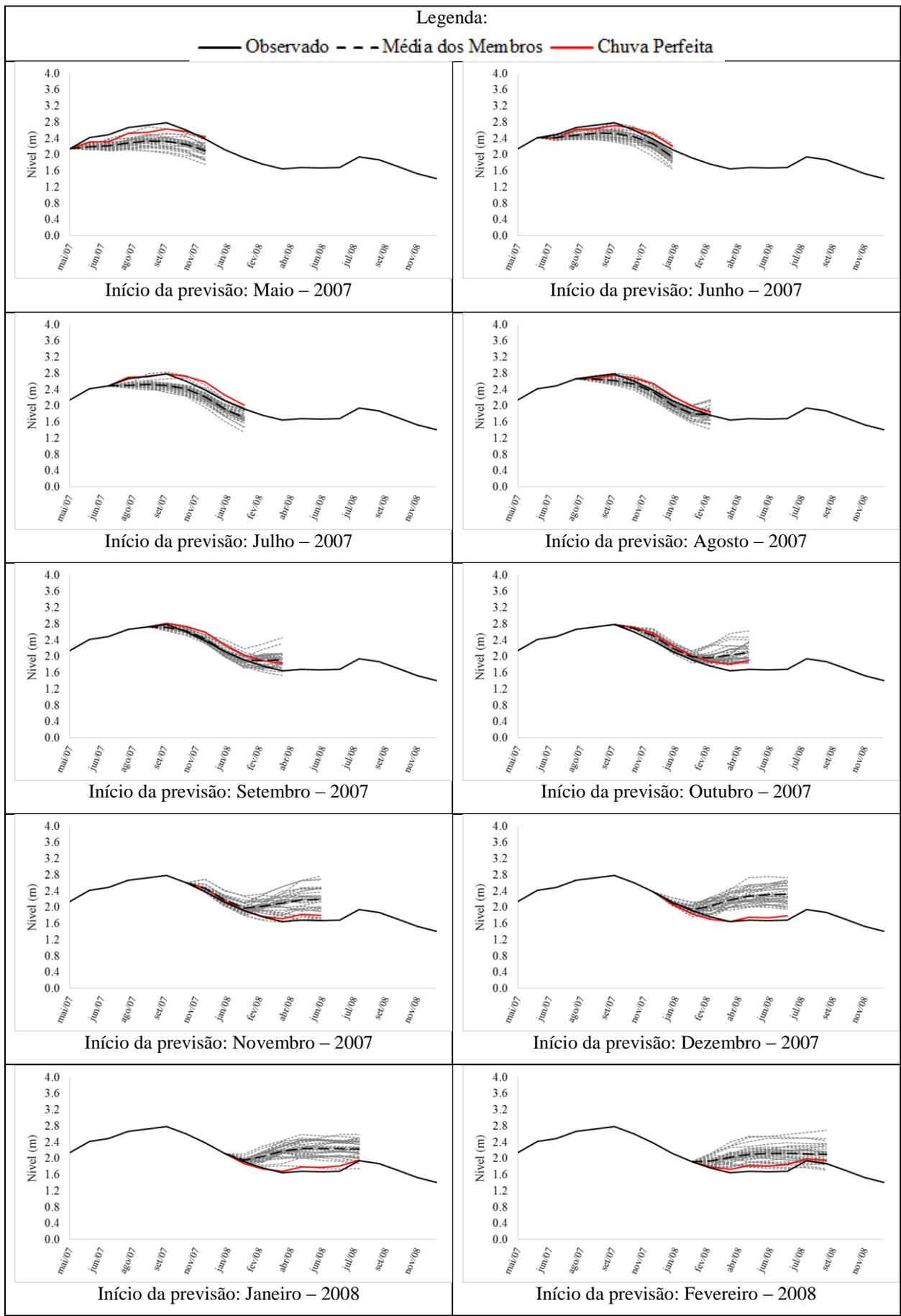


Figura 77 – Previsões iniciando em maio de 2007 a março de 2008.

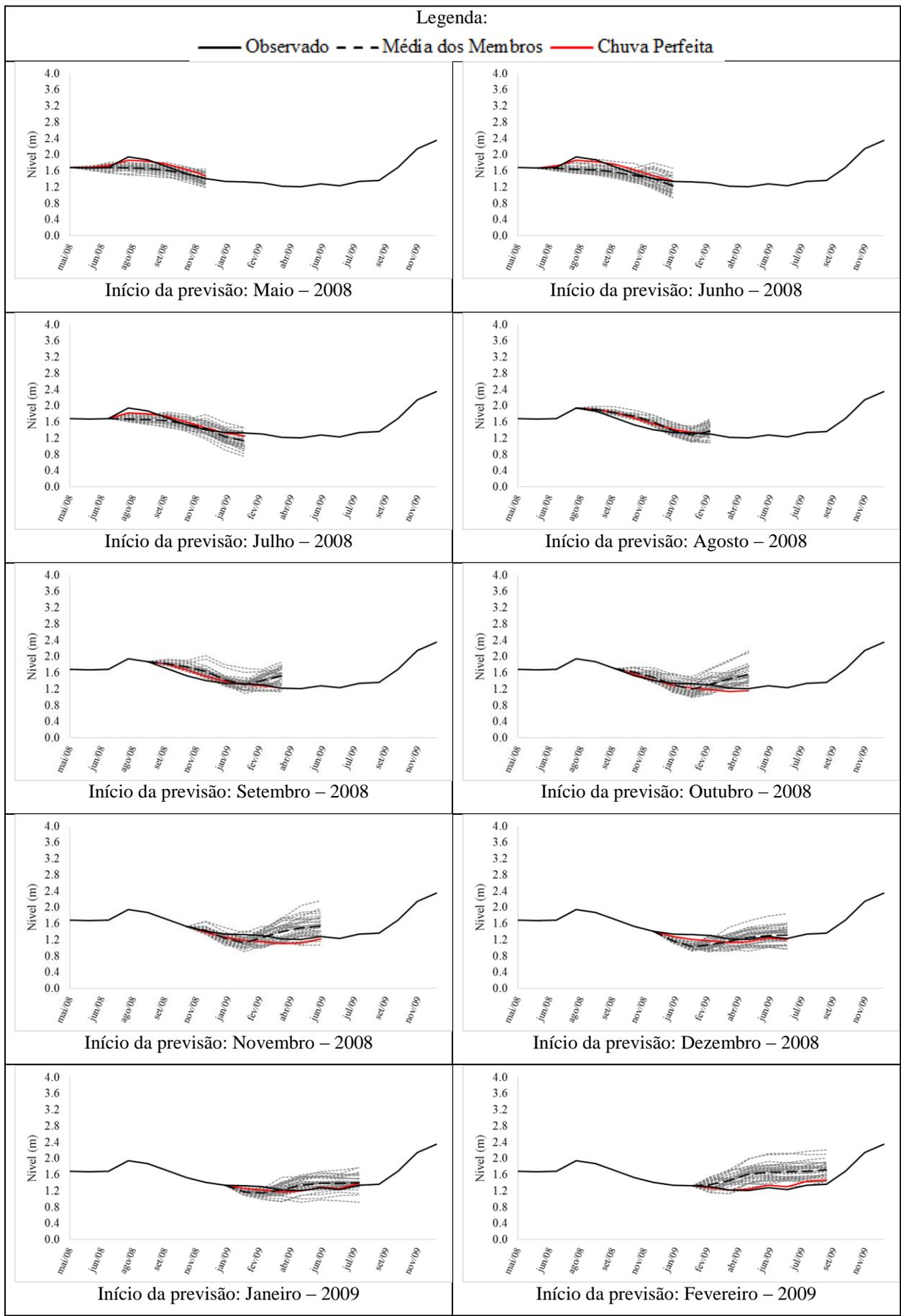


Figura 78 – Previsões iniciando em maio de 2008 a março de 2009.

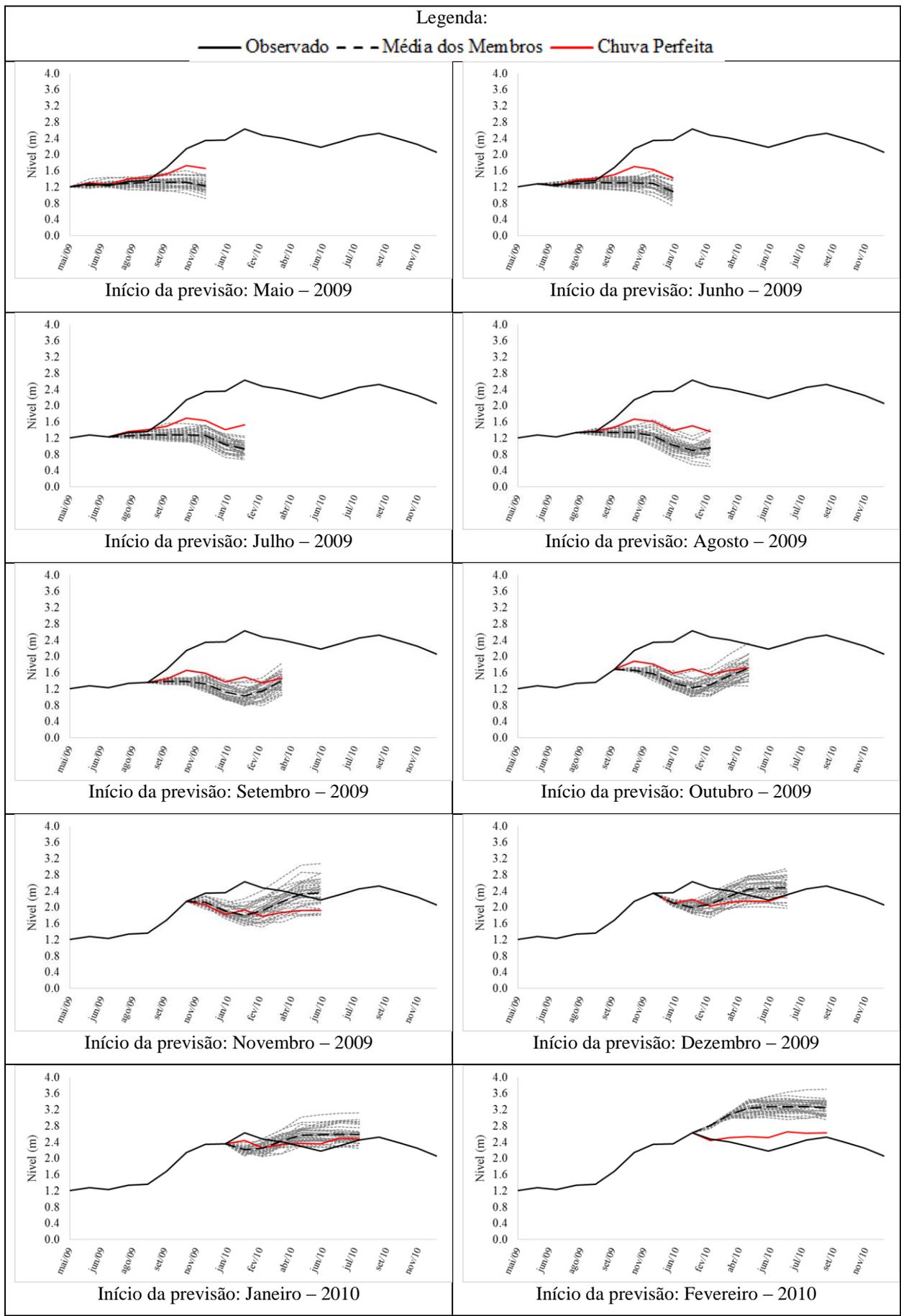


Figura 79 – Previsões iniciando em maio de 2009 a março de 2010.

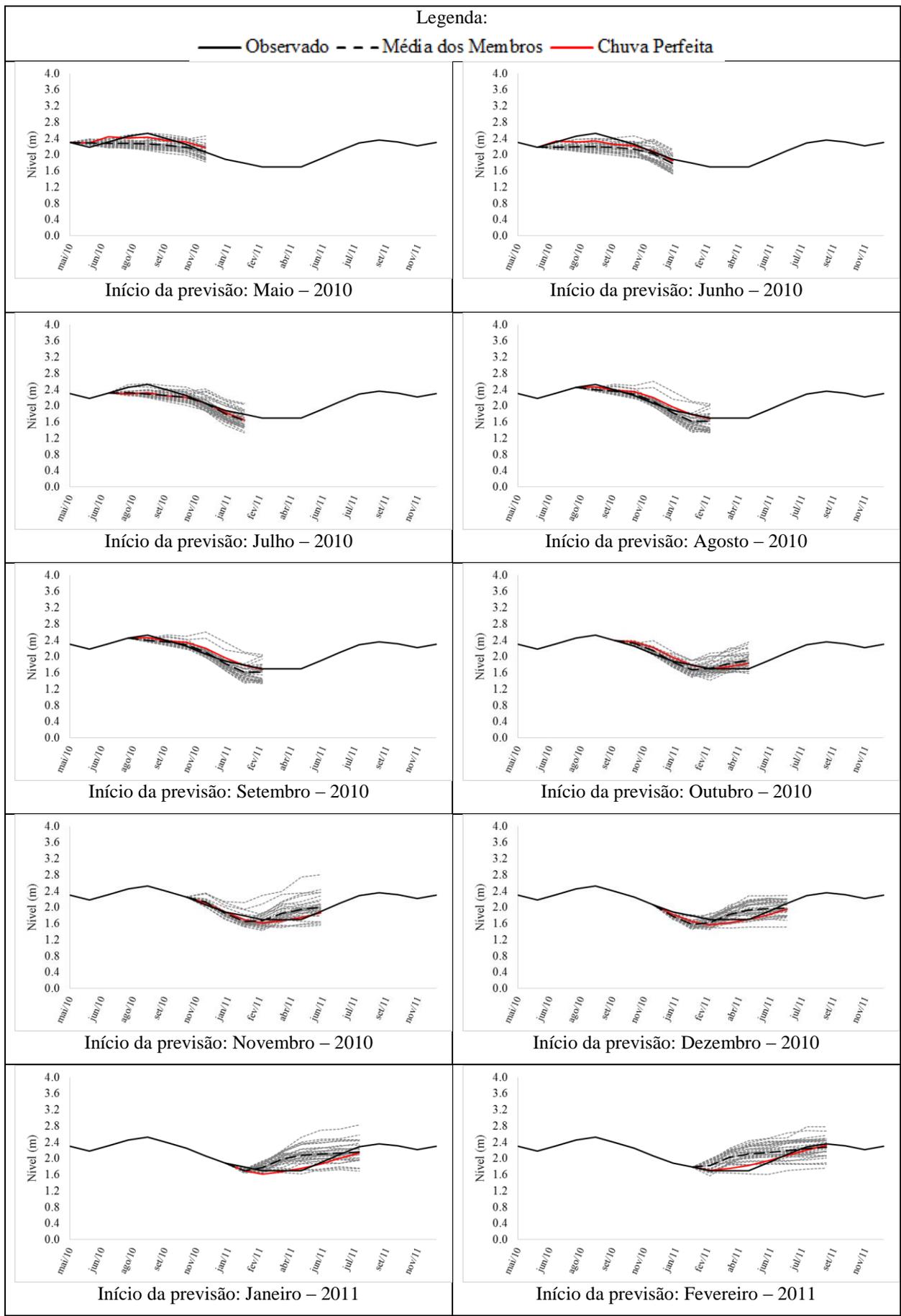


Figura 80 – Previsões iniciando em maio de 2010 a março de 2011.

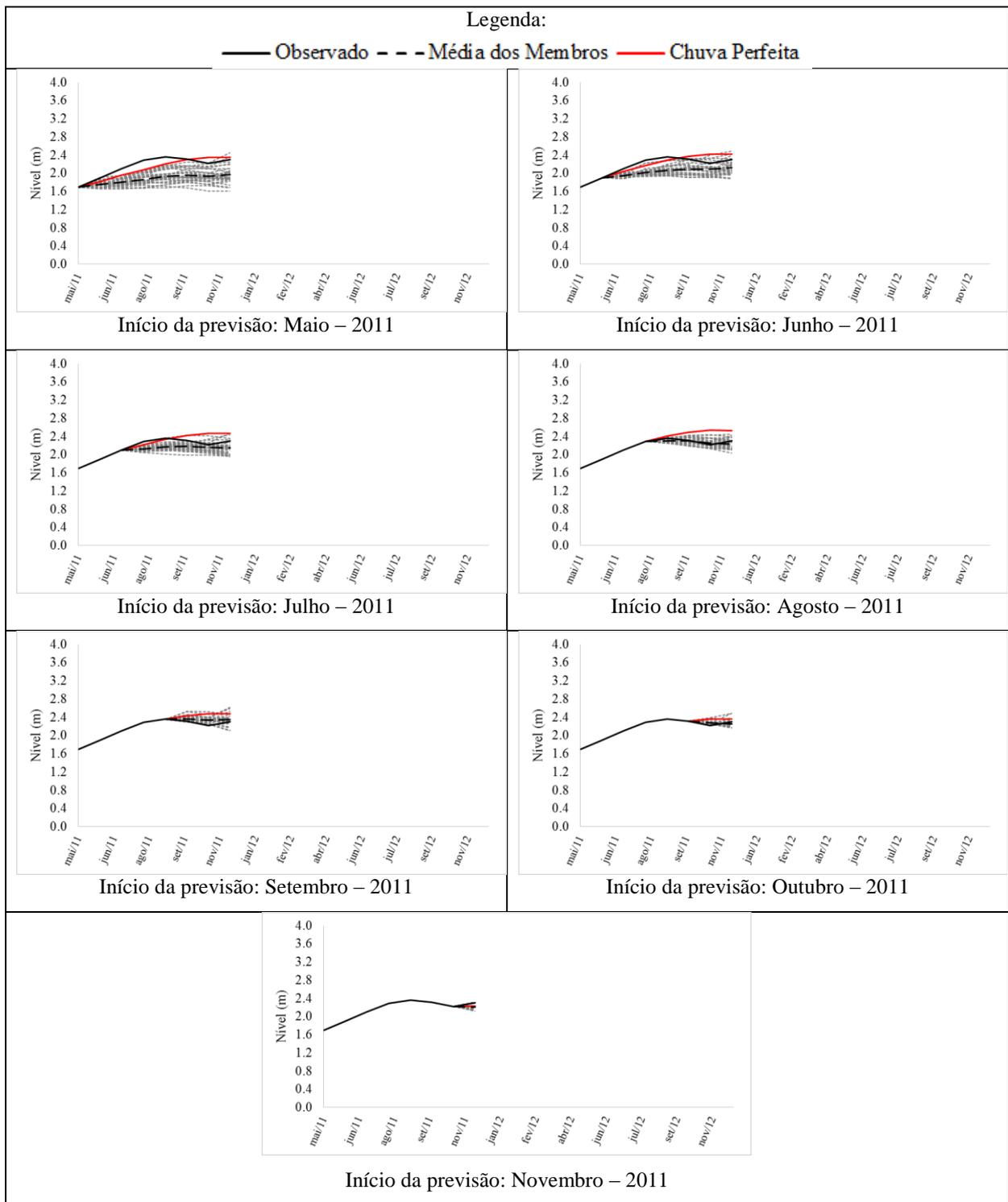


Figura 81 – Previsões iniciando em maio de 2011 a dezembro de 2011.

As quatro primeiras previsões (iniciadas em maio, junho, julho e agosto), que representam o horizonte máximo de sete meses, para os meses em que são realizadas as retiradas para irrigação, obtiveram bons resultados. Do ponto de vista da gestão, o horizonte máximo é de suma importância, pois quanto maior o horizonte de previsão, maior é o tempo

para análise dos possíveis impactos causados pelas previsões. Assim, pode-se concluir que o modelo obteve resultados satisfatórios.

Tassi (2008), determinou intervalos de níveis de água para gestão ambiental do banhado do Taim. No presente estudo, foi identificado que o ano de 2009 foi um ano crítico, com níveis de água baixo, onde seguindo os intervalos determinados por Tassi (2008), não eram para serem realizadas retiradas, ou serem realizadas em menor proporção. Como os modelos empíricos estão limitados as amplitudes dos dados para que foram calibrados, pode-se verificar que no ano de 2009 (Figura 79) o modelo teve dificuldade em simular os níveis.

A análise visual dos gráficos permitiu verificar que algumas previsões obtiveram um maior espalhamento dos membros, porém, pode-se verificar que as previsões nas quais o espalhamento foi menor, houve um maior acerto das previsões. Também ficou constatado que as previsões iniciadas nos meses de novembro a fevereiro obtiveram um maior espalhamento dos membros. Isso, está relacionado com as oscilações advindas das previsões de precipitação realizadas pelo MCG, onde as condições iniciais para esses meses conduzem a um maior espalhamento dos membros.

Na Tabela 45 encontra-se os resultados da diferença, entre as medidas de desempenho calculadas para as previsões realizadas com precipitação perfeita e precipitação prevista para os sete horizontes.

Tabela 45 – Diferença entre as previsões realizadas com precipitação perfeita e precipitação prevista para os sete horizontes.

	Horizontes de previsão (mês)						
	1	2	3	4	5	6	7
Δ Erro médio	0,01	0,01	0,00	0,00	0,00	-0,01	-0,02
Δ Erro médio absoluto	-0,03	-0,05	-0,06	-0,08	-0,09	-0,10	-0,11
Δ Erro médio quadrático	-0,01	-0,03	-0,05	-0,07	-0,09	-0,10	-0,11
Δ Raiz quadrada do erro médio	-0,04	-0,07	-0,09	-0,11	-0,12	-0,12	-0,12
Δ Coeficiente de Nash-Sutcliffe	0,06	0,17	0,31	0,44	0,52	0,58	0,62
Δ Coeficiente de persistência	0,43	1,29	2,27	3,14	3,71	4,13	4,51
Δ Correlação de Pearson	0,03	0,08	0,15	0,20	0,24	0,26	0,29
Δ Coeficiente de determinação	0,05	0,14	0,23	0,28	0,29	0,28	0,27

O coeficiente de Nash-Sutcliffe mostrou-se mais sensível ao aumento do horizonte de previsão, onde pode-se perceber que as previsões podem melhorar significativamente com as correções das precipitações previstas.

Na Figura 82, pode-se verificar o decaimento do desempenho mensurado pelo coeficiente de Nash-Sutcliffe calculados para os resultados das previsões realizadas com precipitação perfeita e prevista. Apesar das figuras não estarem na mesma escala, pode-se perceber que, nos dois casos, o decaimento obteve um bom ajuste linear.

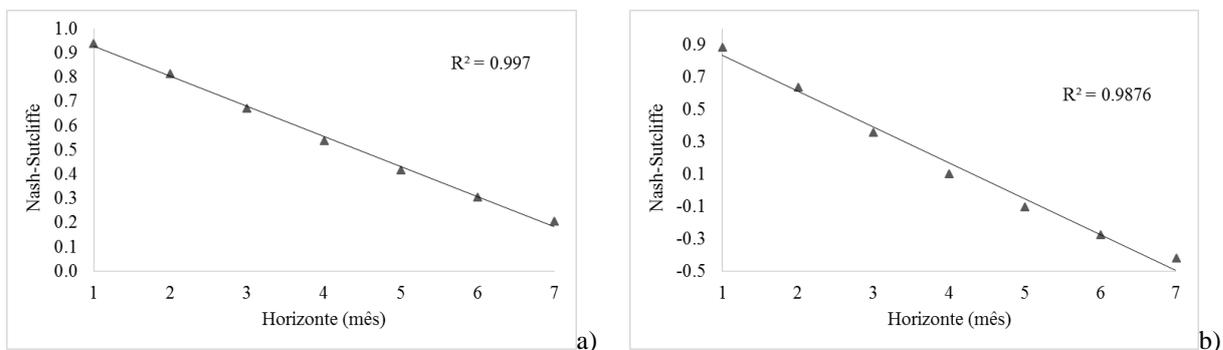


Figura 82 – Decaimento do desempenho mensurado pelo coeficiente de Nash-Sutcliffe calculados para os resultados das previsões realizadas com precipitação perfeita e prevista.

8 CONCLUSÕES E RECOMENDAÇÕES

Esse estudo realizou previsões de níveis no Sistema Hidrológico do Taim utilizando uma metodologia empírica, por meio de regressões múltiplas, utilizando dados observados de precipitação, evaporação, nível de água e precipitação prevista pelo modelo de circulação global COLA 2.2. As previsões foram realizadas para o horizonte de sete meses, em intervalo de tempo mensal e com frequência mensal. Foi utilizada a técnica de previsão por conjunto (*ensemble*), para melhorar as previsões e reduzir as incertezas produzidas pelas condições iniciais das previsões de precipitação.

Os dados brutos disponíveis não mostraram boa correlação com os níveis de água no banhado, por esse motivo, foram realizadas análises e tratamentos nos dados para verificar se melhores correlações eram obtidas. A decomposição clássica das séries proporcionou uma melhor compreensão sobre o comportamento dos dados, onde ficou constatado que algumas técnicas como média móvel, defasagem no tempo e suavização exponencial melhoram significativamente as correlações entre os dados disponíveis e a série de níveis de água no banhado. Os volumes de água retirados por irrigação foram considerados a partir dos valores de área irrigada, que posteriormente foram transformados em diferenças de níveis a partir do uso da curva cota-volume da lagoa Mangueira. Das análises realizadas, foram selecionadas dez variáveis para serem utilizadas na estimativa das equações de regressão.

As séries foram divididas em três intervalos: a) formulação e calibração (1976 a 2003); b) validação (1960 a 1975); c) simulação (2004 a 2011). Foram realizados testes não paramétricos para avaliar o comportamento das séries, e avaliado se suas características permanecem nos três intervalos. Os testes revelaram que os intervalos de formulação e validação possuem características semelhantes, porém, o intervalo de simulação possui características atípicas, se comparada com o restante da série, com níveis abaixo da média observada. Porém este período foi escolhido para a simulação em modo previsão, pois representa o período com dados de precipitação prevista.

Com o objetivo de melhorar a previsão de precipitação, foram testados dois procedimentos de remoção do viés das precipitações previstas (viés total e mensal). Foram utilizadas medidas de desempenho para avaliar a eficiência do modelo de circulação global, onde ficou constatada a necessidade de correção nos dados previstos pelo modelo. Foi utilizada a climatologia do modelo para remoção do viés, onde a mesma possui dez membros

com frequência mensal. A remoção do viés foi realizada, utilizando a técnica de mapeamento das probabilidades Quantil-quantil. Foi realizado o teste de aderência de Kolmogorov-Smirnov para ajustar uma função cumulativa de probabilidades aos dados observados e a climatologia do modelo. Das análises realizadas, percebeu-se que entre os anos de 1976 a 2002, ocorreram alguns anos bastante úmidos, com eventos de picos de precipitação, que não foram bem representados pelo modelo de circulação global. Conforme discutido anteriormente, o período de simulação constitui um intervalo atípico, com baixos valores precipitados, o que ocasionou problemas na metodologia de correção do viés com a climatologia do modelo, fazendo com que as correções superestimassem os valores observados. Assim, foram calculadas funções cumulativa de probabilidades para cada mês do ano, onde foi realizada a correção do viés mensal.

Posteriormente foram desenvolvidos modelos empíricos com base em regressão linear e não linear simples com as variáveis selecionadas. Estes modelos não apresentaram bom desempenho, com ocorrência de erros de grandes magnitudes nas simulações de níveis de água. Assim, foi elaborado um algoritmo para avaliar todas as possibilidades de regressão linear múltipla com os dados selecionados. Para avaliar as equações de regressão, foram estabelecidos limites para as medidas de desempenho pré-definidas, sendo excluídas as equações que não obtiveram o desempenho mínimo pré-definido. O algoritmo obteve 255 possibilidades de combinações entre os dados disponíveis, onde 47 equações não foram rejeitadas pelos critérios de exclusão. Devido ao grande número de equações obtidas foi adicionada uma sub-rotina para análise de sensibilidade das equações de regressão, que serviu como critério de exclusão para as equações que fossem pouco sensíveis às variações da variável precipitação, que constitui a variável independente mais importante desse estudo.

Das 47 equações elaboradas, foram selecionadas oito equações que obtiveram melhores desempenhos, onde percebeu-se que possuíam características semelhantes como parâmetro ponderador da precipitação e do nível anterior de aproximadamente 0,002 e 0,96, respectivamente. A influência das outras variáveis que compõem cada equação, forneceu sensíveis oscilações entre os resultados obtidos por cada equação, podendo-se concluir que as equações são semelhantes. Foi realizado o teste de significância “T” para avaliar a relação entre as variáveis que compõem cada equação e a variável resposta (níveis previstos), onde a hipótese nula foi rejeitada em todos os testes, revelando que existe uma relação significativa entre a variável dependente e as variáveis independentes em todas as oito equações. Foi realizada análise dos resíduos entre dados observados e simulados pelas equações de regressão, onde ficou constatado que um modelo linear era capaz de representar os dados.

As equações foram avaliadas para o intervalo de validação (1960 a 1975) e simulação (2004 a 2011), onde foram utilizadas medidas de desempenho e gráficos para analisar a eficiência das equações. Como o intervalo de validação não possui dados de precipitação prevista foi avaliado utilizando uma abordagem determinística. Os resultados obtidos foram bons, onde as medidas de desempenho revelaram a inexistência de erros de grande magnitude, sendo confirmada pela inspeção dos gráficos. Dos resultados obtidos dos intervalos de formulação e validação, pode-se concluir que as equações produzem resultados semelhantes e foi escolhida uma delas para representar o modelo empírico nas análises posteriores.

O modelo empírico mostrou-se sensível às oscilações da precipitação prevista, onde se pode perceber uma semelhança entre os resultados obtidos pelas medidas de desempenho *Rank histogram* e *Spread-bias diagram* para os dados de precipitação prevista e os níveis previstos pelo modelo de regressão. As configurações dos resultados são bastante semelhantes, podendo-se concluir que a eficiência do modelo de previsão de precipitação influencia diretamente na eficiência do modelo hidrológico de regressão obtido. Avaliando os resultados das previsões de níveis para os meses críticos, onde são efetuadas as retiradas de água para irrigação, ficou constatado que o modelo empírico obteve bons resultados, com as previsões utilizando chuva perfeita, muito próximas das previsões utilizando chuva prevista. Podendo-se concluir, que o modelo empírico pode ser utilizado para realizar previsões de níveis no banhado.

Da análise dos resultados das medidas de desempenho e da inspeção dos gráficos, pôde-se concluir que a medida que o horizonte aumenta, aumentam-se as incertezas, e assim, os erros na previsão do modelo, conforme era esperado. Da comparação dos resultados obtidos para as previsões utilizando chuva perfeita e prevista, pode-se concluir que se melhorando as previsões de precipitação, obtém-se uma melhora significativa das previsões de níveis.

As medidas de desempenho utilizadas neste estudo foram bastante eficientes em detectar os erros e avaliar a eficiência da previsão, onde foram confirmados os resultados pela inspeção dos gráficos. A técnica de previsão por conjunto proporcionou uma maior discussão sobre as previsões, onde os resultados demonstraram uma melhor confiabilidade para a tomada de decisão.

Os resultados das previsões foram promissões, onde as previsões para os meses de maior interesse para gestão obtiveram bons resultados. Da análise dos resíduos, percebeu-se que existe uma relação linear entre os resíduos das previsões do primeiro mês de antecedência, com os demais horizontes. Utilizando o algoritmo evolutivo SCE-UA,

estimaram-se algumas equações polinomiais que eram capazes de prever os erros, porém, o ponto de partida para as previsões era o primeiro mês de antecedência, o qual está no tempo futuro onde não existem dados observados. Assim, recomenda-se a análise de correções entre o ponto de partida para as previsões e os resíduos produzidos por cada horizonte. Pois, se os erros pudessem ser estimados, melhorariam consideravelmente as previsões.

REFERÊNCIAS

- ANTOLIK, M. S. (2000). An overview of the National Weather Service's centralized statistical quantitative precipitation forecasts. *Journal of Hydrology* 239 (2000) 306–337
- ÁVILA, A. M. H. de; CARDOSO, A. de O.; PINTO, H. S. (2009). Aplicação da Correção Estatística na Previsão de Tempo Estendida, para Três Localidades da Região Sul. XVI Congresso Brasileiro de Agrometeorologia. Belo Horizonte - MG.
- BATISTA, R. J. R. (2012). Climatologia e Previsão Climatológica, Via Modelo RegCM3, de Índices de Conforto Térmico para a Região Metropolitana de São Paulo. Dissertação de Mestrado. Departamento de Ciências Atmosféricas do Instituto de Astronomia, Geofísica e Ciências Atmosféricas da Universidade de São Paulo – USP. 115p.
- BELAYNEH, A.; ADAMOWSKI, J.; KHALIL, B.; OZGA-ZIELINSKI, B. (2014). Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *Journal of Hydrology* 508: 418–429
- BOUZADA, M. A. C. (2012). Aprendendo Decomposição Clássica: Tutorial para um Método de Análise de Séries Temporais. *TAC*, v.2, pp. 1-18.
- BRAVO, J. M.; COLLISCHONN, W.; UVO, C. B.; TASSI, R. PICCILLI, G. A. (2009). Avaliação da Eficácia de Modelos de Previsão de Vazão de Afluência a Reservatórios com Base na Utilidade da Previsão. *Simpósio Brasileiro de Recursos Hídricos*. 19p.
- BRAVO, J. M. (2010). Subsídios à operação de reservatórios baseada na previsão de variáveis hidrológicas. Tese de Doutorado. Universidade Federal do Rio Grande do Sul/Instituto de Pesquisas Hidráulicas. 205p.
- BROWN, J. D; DEMARGNE, J.; DONG-JUN SEO & LIU, Y. (2001). The Ensemble Verification System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. *Environmental Modelling & Software*, v.25, issue 7, 854-872p.
- CARDOSO, M. A. G. (2010). Previsão Hidroclimática para o Gerenciamento do Banhado do Taim. Tese de Doutorado. Universidade Federal do Rio Grande do Sul/Instituto de Pesquisas Hidráulicas. 154p.

- CASH, B. A.; RODÓ, X.; KINTER III, J. L. (2008). Links between Tropical Pacific SST and Cholera Incidence in Bangladesh: Role of the Eastern and Central Tropical Pacific. *J. Climate*, 21, 4647–4663.
- CAVALCANTI, J. R. de A. (2013). Influência da Hidrodinâmica no Metabolismo de Lagos Rasos. Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul/Instituto de Pesquisas Hidráulicas. 92p.
- COLLISCHONN, W. & TUCCI, C. E. M. (2005). Previsão Sazonal de Vazão na Bacia do Rio Uruguai 1: Ajuste e Verificação do Modelo Hidrológico Distribuído. *Revista Brasileira de Recursos Hídricos - RBRH*, v. 10, n. 4, p. 43-59.
- CNPQ, Conselho Nacional de Desenvolvimento Científico e Tecnológico. Disponível em: <<http://www.cnpq.br/web/guest/apresentacao7>>, Acesso em 07 agosto 2014.
- COULIBALY, P.; ANCTIL, F.; RASMUSSEN, P.; BOBÉE, B. (2000). A recurrent neural networks approach using indices of low-frequency climatic variability to forecast regional annual runoff. *Hydrol. Process.*14, 2755-2777
- CPTEC, Centro de Previsão de Tempo e Estudos Climáticos. Disponível em: <cptec.inpe.br>, acesso em 10 de abril de 2015.
- CUO, L.; PAGANO, T. C.; WANG, Q. J. (2011). A Review of Quantitative Precipitation Forecasts and Their Use in Short- to Medium-Range Streamflow Forecasting. *JOURNAL OF HYDROMETEOROLOGY*. october V 12. 10.1175/2011JHM1347.1
- DELSOLE, T. & SHUKLA, J. (2006). Specification of Wintertime North American Surface Temperature. *J. Climate*, 19, 2691–2716.
- DEMERRITT, D.; NOBERT, S.; CLOKE, H. L.; PAPPENBERGER, F. (2013). The European Flood Alert System and the communication, perception, and use of ensemble predictions for operational flood risk management. *HYDROLOGICAL PROCESSES*. 27, 147–157
- DORNELLES, F. (2007). Previsão Contínua de Níveis Fluviais com Redes Neurais Utilizando Previsão de Precipitação: Investigação Metodológica da Técnica. Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul/Instituto de Pesquisas Hidráulicas. 99p.
- FRAGOSO JÚNIOR, C. R. (2009). Modelagem Tridimensional da Estrutura Trófica em Ecossistemas Aquáticos Continentais Rasos. Tese de Doutorado. Universidade Federal do Rio Grande do Sul/Instituto de Pesquisas Hidráulicas. 299p.

FERNANDES, W. de S. (2012). Avaliação do Impacto das Mudanças Climáticas na Oferta Hídrica da Bacia Hidrográfica do Reservatório Óros Usando os Modelos de Mudanças Climáticas do IPCC-AR4, Levando em Consideração as Diversas Incertezas Associadas. Dissertação de Mestrado. Departamento de Engenharia Hidráulica e Ambiental da Universidade Federal do Ceará. 128p.

FERRARIS, L.; RUDARI, R.; SICCARDI, F. (2002). The Uncertainty in the Prediction of Flash Floods in the Northern Mediterranean Environment. *Journal Hydrometeor.* v3, 714-727.

GALVÃO, C. O. (1999). Previsão de Precipitação de Longo Prazo no Nordeste do Brasil. Tese de Doutorado. Universidade Federal do Rio Grande do Sul/Instituto de Pesquisas Hidráulicas. 153p.

GHYS, É. (2012). 12º International Congress on Mathematical Education. 8 July – 15 July, 2012, COEX, Seoul, Korea

GREEN, T. T.; TANIGUCHI, M.; KOOI, H.; GURDAK, J. J.; ALLEN, D. A.; HISCOCK, K. M.; TREIDEL, H.; AURELI, A. (2011). Beneath the Surface of Global Change: Impacts of Climate Change on Groundwater. *Journal of Hydrology.* 405: 532-560.

GUJARATI, D. N. (2006). *Econometria Básica.* 4 ed. Rio de Janeiro: Campus Elsevier. 812 p.

HUANG, W.; XU, B.; CHAN-HILTON, A. (2004). Forecasting flows in Apalachicola River using neural. *HYDROLOGICAL PROCESSES.* 18, 2545– 2564

HYNDMAN, R. J.; ATHANASOPOULOS, G. (2013). *Forecasting: principles and practice.*

IGES, Center for Ocean-Land-Atmosphere Studies. Disponível em: <<http://www.iges.org/cola.html>> Acesso em 25 maio de 2015.

INPE, Instituto Nacional de Pesquisas Espaciais. Disponível em: <www.inpe.br>, acesso em 10 de abril de 2015.

IRI, International Research Institute for Climate and Society. Disponível em: <<http://iri.columbia.edu/>>, Acesso em 24 setembro de 2014.

ISLAM, M. N.; SIVAKUMAR, B. (2002). Characterization and prediction of runoff dynamics: a nonlinear dynamical view. *Adv. Water Resour.* 25, 179–190

- JANG, J. S. R.; SUN, C. T.; MIZUTANI, E. (1997). Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence. Prentice Hall, Upper Saddle River, New Jersey.
- JAYAWARDENA, A. W.; LAI, F. (1994). Analysis and prediction of chaos in rainfall and streamflow time series. *J. Hydrol. (Amst)* 153, 23–52.
- JOHNSON, F. & SHARMA, A. (2015). What are the impacts of bias correction on future drought projections? *Journal of Hydrology*. V. 525, 472-485 p.
- JUNG, I-W; CHANG, H. (2011). Assessment of future runoff trends under multiple climate change scenarios in the Willamette River Basin, Oregon, USA. *HYDROLOGICAL PROCESSES*. 25, 258– 277
- KARAMOUZ, M.; RAZAVI, S.; ARAGHINEJAD, S. (2008). Long-lead seasonal rainfall forecasting using time-delay recurrent neural networks: a case study. *HYDROLOGICAL PROCESSES*. 22, 229– 241
- KATTAN, A.; FATIMA, S.; ARIF, M. (2015). Time-series event-based prediction: An unsupervised learning framework based on genetic programming. *Information Sciences*. 301: 99-123.
- KHATIBI, R.; GHORBANI, M. A.; NAGHIPOUR, L.; JOTHIPRAKASH, V.; FATHIMA, T. A.; FAZELIFARD, M. H. (2014). Inter-comparison of Time Series Models of Lake Levels Predicted by Several Modeling Strategies. *Journal of Hydrology*. 511: 530-545.
- KIRTMAN, B. P.; SCHOPF, D. M. P. S.; SCHNEIDER, E. K. (2003). A New Approach for Coupled GCM Sensitivity Studies. October 2003, 50 pp.
- KRZYSZTOFOWICZ, R. (2001). The Case for Probabilistic Forecasting in Hydrology. *Journal of Hydrology*. 249: 2-9.
- LIU, Y.; WEERTS, A. H.; CLARK, M.; FRANSSEN, H. J. H.; KUMAR, S.; MORADKHANI, H.; SEO, D. J.; SCHWANENBERG, D.; SMITH, P.; van DIJK, A. I. J. M.; van VELZEN, N.; HE, M.; LEE, H.; NOH, S. J.; RAKOVEC, O.; RESTREPO, P. (2012). Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences*. 16: 3863–3887
- LORENZ, E.N. On the Existence of Extended Range Predictability. *Journal of Applied Meteorology*. V.12, apr., p.543-546. 1973.

- MARENGO, J. A. (2008). Água e Mudanças Climáticas. Estudos Avançados. v.22, n. 63, p.83-96.
- MARENGO, J. A.; TOMASELLA, J.; NOBRE, C. A. (2010). Mudanças Climáticas e Recursos Hídricos. In: BICUDO, C. E. de M.; TUNDISI, J. G.; SCHEUENSTUHL, M. C. B. (eds.). Águas do Brasil: Análises Estratégicas. Capítulo 12, 224p. <abc.org.br/rubrique.php3?id_rubrique=150>. Acessado em: 22/02/2015.
- MARQUES, D. M.; Villanueva, A. O. N. (2001). Regime Hidrológico do Banhado do Taim e sua Conservação. ABRH Notícias, Revista Informativa da ABRH, n. 4, p. 22-61.
- MARQUES, D. M.; TUCCI, C.; CALAZANS, D.; CALLEGARO, V. L. M. & VILLANUEVA, A. (2002). O Sistema Hidrológico do Taim, Site 7. Os Sites e o Programa Brasileiro de Pesquisas Ecológicas de Longa Duração. Belo Horizonte: FURG/UFMG. p. 125-144.
- MCMILLAN, H.; JACKSON, B.; CLARK, M.; KAVETSKI, D. & WOODS, R. (2011). Rainfall Uncertainty in Hydrological Modelling: An Evaluation of Multiplicative Error Models. Journal of Hydrology. 400: 83–94
- MEKANIK, F.; IMTEAZ, M. A.; GATO-TRINIDAD, S.; ELMAHDI, A. (2013). Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. Journal of Hydrology. 503: 11–21
- MELLER, A. (2012). Previsão de Cheias por Conjunto em Curto Prazo. Tese de Doutorado. Universidade Federal do Rio Grande do Sul/Instituto de Pesquisas Hidráulicas. 204p.
- MENDONÇA, A. M. (1999). Desempenho do Modelo Global CPTEC/COLA Durante Episódios de ZCAS, Utilizando os Esquemas de Convecção Profunda Tipo Kuo e Arakawa-Schubert Relaxada. Dissertação de Mestrado. Instituto Nacional de Pesquisas Espaciais. São José dos Campos. 166 p.
- MISRA, V. (2003). An Evaluation of the Predictability of Austral Summer Season Precipitation over South America. 35 pp.
- MOLION, L. C. B. (2007). Desmistificando o Aquecimento Global. Intergeo, v. 5, p. 13-20.
- MORENO, J. I. L.; LATRON, J.; LEHMANN, A. (2010). Effects of sample and grid size on the accuracy and stability of regression-based snow interpolation methods. HYDROLOGICAL PROCESSES. 24, 1914– 1928

- MORETTIN, P. & TOLOI, C. (2004). *Análise de Series Temporais*. São Paulo: Edgard Blücher.
- MURPHY, A. H. (1993). What is a good forecast? An essay on the nature of goodness in weather forecasting. *Weather and Forecasting*. v. 8.
- NAGHETTINI, M.; PINTO, É. J. DE A. (2007). *Hidrologia estatística*. Belo Horizonte: CPRM. 552 p.
- NASH, J. E. & SUTCLIFFE, J. V. (1970). River flow forecasting through conceptual models. Part I – a discussion of principles *J. Hydrol.*, 10, pp. 282–290
- NEVES, D. G. das. (2012). *Influência da Vegetação na Precipitação Pluviométrica Sazonal do Estado do Amapá: Um Estudo de Sensibilidade Climática*. Tese de doutorado em Biodiversidade Tropical. Universidade Federal do Amapá - UNIFAP. 129 p.
- NOAA, National Oceanic and Atmospheric Administration. Disponível em: <<http://www.noaa.gov/>>, Acesso em 06 junho de 2015.
- NYEKO-OGIRAMOI, P.; WILLEMS, P.; NGIRANE-KATASHAYA, G. (2013). Trend and variability in observed hydrometeorological extremes in the Lake Victoria basin. *Journal of Hydrology* 489 (2013) 56–73
- OLIVEIRA, C. de. (2014). *Identificação e Correção da Banda Brilhante em Dados de Radar Meteorológico*. Dissertação de Mestrado. Departamento de Matemática da Universidade Federal do Paraná. 73p.
- PALMER, T. N. (1993) Extended range atmospheric prediction and the Lorenz model. *Bull. Am. Meteorol. Soc*, 74: 49-65
- PARTAL T. (2009). Modelling evapotranspiration using discrete wavelet transform and neural networks. *HYDROLOGICAL PROCESSES*. 23, 3545– 3555
- PAZ, A. (2003). *Análise dos Principais Fatores Intervenientes no Comportamento Hidrodinâmico do Banhado do Taim (RS)*. Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul/Instituto de Pesquisas Hidráulicas. 179 p.
- PEDROLLO, O. C. (2000). *Previsão em Tempo Atual de Cheias com Uso de Sistema Especialista Difuso*. Tese de Doutorado. Universidade Federal do Rio Grande do Sul/Instituto de Pesquisas Hidráulicas. 123p.

PELD - CNPq dez anos do Programa de Pesquisas Ecológicas de Longa Duração do Brasil : achados, lições e perspectivas. Recife : Ed. Universidade da UFPE, 2013.

PICCILLI, D. G. A. (2007). Avaliação da previsão hidroclimática no alto paraguai. Tese de Doutorado. Universidade Federal do Rio Grande do Sul/Instituto de Pesquisas Hidráulicas. 324p.

RAMAN, H.; SUNILKUMAR, N. (1995). Multivariate modeling of water resource time series using artificial neural network. *Hydrol. Sci. J.* 40 (21), 145–163.

RENARD, B., KAVETSKI, D., KUCZERA, G., THYER, M., AND FRANKS, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resources. Res.*, 46, W05521, doi:10.1029/2009WR008328

SALAS, J. D.; DELLEUR, J. W.; YEVJEVICK V.; LANE, W. L. (1980). Applied modeling of hydrologic time series. Water Resouces Publications. P. O. Box 2841. Littleton, Colorado 80161, USA

SANTOS, T. N. dos. (2014). Redes Neurais Artificiais e Relação ZR Aplicadas à Estimativa de Chuva. Dissertação (Mestrado) — Programa de Pós-Graduação em Métodos Numéricos para Engenharia, Universidade Federal do Paraná, 2014.

SCHULTZ, D. P.; SCHULTZ, S. E. (1992). História da psicologia moderna. São Paulo: Cultrix. ed. 16. 439 p.

SCUDERI, C. M. (2010). Estimativa da taxa de evapotranspiração da macrófita aquática "Zizaniopsis Bonariensis". Aplicação à simulação do nível de água do longo período do Banhado do Taim (RS). Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul/Instituto de Pesquisas Hidráulicas. 175p.

SEN, Z.; KADIOGLU, M.; BATUR, E. (2000). Stochastic Modelling of the Van Lake Monthly Level Fluctuations in Turkey. *Theor. Appl. Climatol.* 65: 99-110.

SILVA, B. C. da; TUCCI, C. E. M.; COLLISCHONN, W. (2006). Previsão de Vazão com Modelos Hidroclimáticos. *RBRH – Revista Brasileira de Recursos Hídricos* Volume 11 n.3 Jul/Set 2006, 15-29

SILVA, M. C. L da.; MENDONÇA, A. M.; BONATTI, J. P.; (2008). Determinação das Previsões de Temperaturas Mínimas e Máximas a Partir do Histórico das Previsões de Tempo por Conjunto do CPTEC. *Revista Brasileira de Meteorologia.* v. 23, n. 4. p: 431 - 449.

- SILVEIRA, C. da S.; OLIVEIRA, J. L.; SILVA, O. F. V.; NÓBREGA, M. T.; SOUZA, S. A. de. (2014). Variabilidade e Mudanças Climáticas na Bacia do Rio São Francisco - Avaliação dos Modelos Globais Quanto à Representação da Precipitação. XII Simpósio de Recursos Hídricos do Nordeste.
- SILVEIRA, C. da S.; FILHO, F. de A. de S.; LOPES, J. E. G.; BARBOSA, P. S. F.; TIEZZI, R. O. (2014). Análise das Projeções de Vazões nas Bacias do Setor Elétrico Brasileiro Usando Dados do IPCC-AR4 para o Século XXI. Revista Brasileira de Recursos Hídricos. v. 19, n. 4, 59-71p.
- SILVESTRO, F. & REBORA, N. (2014). Impact of Precipitation Forecast Uncertainties and Initial Soil Moisture Conditions on a Probabilistic Flood Forecasting Chain. Journal of Hydrology. V. 519, Part A, 1052-1067 p.
- SIVAKUMAR, B. (2004). Chaos theory in geophysics: past, present and future. Chaos, Solitons and Fractals 19 (2004) 441–462
- TASSI, R. (2008). Gerenciamento Hidroambiental de Terras Úmidas. Tese de Doutorado. Universidade Federal do Rio Grande do Sul/Instituto de Pesquisas Hidráulicas. 240p.
- TORRES, I. B. La; AMATYA, D. M.; SUN, G.; CALLAHAN, T. J. (2011). Seasonal rainfall–runoff relationships in a lowland forested watershed in the southeastern USA. HYDROLOGICAL PROCESSES. 25, 2032– 2045
- TUCCI, C.E.M. (1993). Hidrologia: ciência e aplicação. Quarta Edição. ABRH: Porto Alegre. 943 pg.
- TUCCI, C. E. M. & COLLISCHONN, W. (2003). Previsão de Vazão. In: Clima e Recursos Hídricos no Brasil. Porto Alegre: ABRH, p.281-348.
- TUCCI, C.E.M. (1998). Modelos Hidrológicos. Primeira Edição. ABRH: Porto Alegre. 669pg.
- TUCCI, C.E.M. (2005). Modelos Hidrológicos. Segunda Edição. ABRH: Porto Alegre. 678pg.
- VILLANUEVA, A.O. N. (1997). Simulação de áreas de inundação dinâmicas: canais compostos e wetlands. Universidade Federal do Rio Grande do Sul/Instituto de Pesquisas Hidráulicas. 125 p.

- VILLANUEVA, A.O.N., Marques, D. M. & TUCCI, C. E. M. (2000). The Taim Wetland Conflict: a Compromise Between Environment Conservation and Irrigation. *Water International*. 25: 610-616.
- YOUNG, R. M. B. (2010). Decomposition of the Brier score for weighted forecast-verification pairs. *QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY*. 136: 1364–1370
- YU, S-P; YANG, J-S; LIU, G-M. (2013). A novel discussion on two long-term forecast mechanisms for hydro-meteorological signals using hybrid wavelet-NN model *Journal of Hydrology* 497: 189–197
- ZHANG, G.; PATUWO, B. E.; HU, M. Y. (1998). Forecasting with Artificial Neural Networks: The State of the Art. *International Journal of Forecasting*. 14: 35–62.
- ZHENG, X.; FREDERIKSEN, D. S. C. S.; GRAINGER, S. (2008). Potentially Predictable Patterns of Seasonal Mean Geopotential Heights in an Ensemble of Climate Simulations with the COLA AGCM. 42pp.
- ZWEIMÜLLER, I.; ZESSNER, M.; HEIN, T. (2008). Effects of climate change on nitrate loads in a large river: the Austrian Danube as example. *HYDROLOGICAL PROCESSES*. 22, 1022– 1036.

ANEXO A

$$F_Z(z) = \Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right] dz$$

Tabela A1– Função cumulativa de probabilidades da distribuição Normal Padrão.

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5606	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8585	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9137	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

Tabela A2 – Valores críticos da estatística $D_{N,\alpha}$ do teste de aderência KS.

N	$D_{N,0,10}$	$D_{N,0,05}$	$D_{N,0,02}$	$D_{N,0,01}$	N	$D_{N,0,10}$	$D_{N,0,05}$	$D_{N,0,02}$	$D_{N,0,01}$
10	0,369	0,409	0,457	0,489	26	0,233	0,259	0,290	0,311
11	0,352	0,391	0,437	0,468	27	0,229	0,254	0,284	0,305
12	0,338	0,375	0,419	0,449	28	0,225	0,250	0,279	0,300
13	0,325	0,361	0,404	0,432	29	0,221	0,246	0,275	0,295
14	0,314	0,349	0,390	0,418	30	0,218	0,242	0,270	0,290
15	0,304	0,338	0,377	0,404	31	0,214	0,238	0,266	0,285
16	0,295	0,327	0,366	0,392	32	0,211	0,234	0,262	0,281
17	0,286	0,318	0,355	0,381	33	0,208	0,231	0,258	0,277
18	0,279	0,309	0,346	0,371	34	0,205	0,227	0,254	0,273
19	0,271	0,301	0,337	0,361	35	0,202	0,224	0,251	0,269
20	0,265	0,294	0,329	0,352	36	0,199	0,221	0,247	0,265
21	0,259	0,287	0,321	0,344	37	0,196	0,218	0,244	0,262
22	0,253	0,281	0,314	0,337	38	0,194	0,215	0,241	0,258
23	0,247	0,275	0,307	0,330	39	0,191	0,213	0,238	0,255
24	0,242	0,269	0,301	0,323	40	0,189	0,210	0,235	0,252
25	0,238	0,264	0,295	0,317	>40	$1,22/\sqrt{N}$	$1,36/\sqrt{N}$	$1,52/\sqrt{N}$	$1,63/\sqrt{N}$

Tabela A3 – Distribuição F de Snedecor a 5% ($p=0,05$)

	1	2	3	4	5	6	7	8	9	10	12	14	15	16	18	20	30	40	60	120
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,42	19,43	19,43	19,44	19,45	19,46	19,47	19,48	19,49
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,71	8,70	8,69	8,67	8,66	8,62	8,59	8,57	8,55
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,87	5,86	5,84	5,82	5,80	5,75	5,72	5,69	5,66
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,64	4,62	4,60	4,58	4,56	4,50	4,46	4,43	4,40
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,96	3,94	3,92	3,90	3,87	3,81	3,77	3,74	3,70
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,53	3,51	3,49	3,47	3,44	3,38	3,34	3,30	3,27
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,24	3,22	3,20	3,17	3,15	3,08	3,04	3,01	2,97
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,03	3,01	2,99	2,96	2,94	2,86	2,83	2,79	2,75
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,86	2,85	2,83	2,80	2,77	2,70	2,66	2,62	2,58
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,74	2,72	2,70	2,67	2,65	2,57	2,53	2,49	2,45
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,64	2,62	2,60	2,57	2,54	2,47	2,43	2,38	2,34
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,55	2,53	2,51	2,48	2,46	2,38	2,34	2,30	2,25
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,48	2,46	2,44	2,41	2,39	2,31	2,27	2,22	2,18
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,42	2,40	2,38	2,35	2,33	2,25	2,20	2,16	2,11
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,37	2,35	2,33	2,30	2,28	2,19	2,15	2,11	2,06
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,33	2,31	2,29	2,26	2,23	2,15	2,10	2,06	2,01
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,29	2,27	2,25	2,22	2,19	2,11	2,06	2,02	1,97
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,26	2,23	2,21	2,18	2,16	2,07	2,03	1,98	1,93
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,22	2,20	2,18	2,15	2,12	2,04	1,99	1,95	1,90
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,20	2,18	2,16	2,12	2,10	2,01	1,96	1,92	1,87
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,17	2,15	2,13	2,10	2,07	1,98	1,94	1,89	1,84
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,15	2,13	2,11	2,08	2,05	1,96	1,91	1,86	1,81
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,13	2,11	2,09	2,05	2,03	1,94	1,89	1,84	1,79
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,11	2,09	2,07	2,04	2,01	1,92	1,87	1,82	1,77
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,09	2,07	2,05	2,02	1,99	1,90	1,85	1,80	1,75
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,08	2,06	2,04	2,00	1,97	1,88	1,84	1,79	1,73
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,06	2,04	2,02	1,99	1,96	1,87	1,82	1,77	1,71
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,05	2,03	2,01	1,97	1,94	1,85	1,81	1,75	1,70
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,04	2,01	1,99	1,96	1,93	1,84	1,79	1,74	1,68
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,95	1,92	1,90	1,87	1,84	1,74	1,69	1,64	1,58
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,86	1,84	1,82	1,78	1,75	1,65	1,59	1,53	1,47
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,83	1,78	1,75	1,73	1,69	1,66	1,55	1,50	1,43	1,35

APÊNDICE A

Algoritmo em Fortran para elaboração de equações de regressão linear múltipla.

```
program Regressao
!*****
!*****
!*****
!***** Método para regressão linear múltipla *****
!*****
!*****
!*****
!*****
!obs: essa rotina realiza regressão linear múltipla e avaliaos resultados
!por meio de medidas de desempenho.
use modulo !uso do módulo para declarar as variáveis
implicit none !retira todo tipo de declaração
!*****
*****
! 0. Declaração das variáveis utilizadas
integer i,j,N,NS,NS2,h1,h2,h3,h4,h5,h6,NSi,NSf,NSc
integer NP1,NP2,NP3,NP4,NP5,NP6,NP7,Nc1
integer di !resposta das diferenças em cm (resíduos)
real ano_err !resposta da anomalia em cm
real,allocatable::A3(:,:)
!*****
*****
! 1. Rotina para ler os dados de entrada
!Rotina para ler os dados de entrada e armazena em N
open(unit=1,file='matriz.txt',action='read',status='old',IOSTAT=IO)
call conta(1) !subrotina para contar a quantidade de dados
close(1)
N=Nc !quantidade de dados
!*****
*****
!Rotina para gerar as combinações possíveis para os índices
open(unit=2,file='indices.txt',action='write',status='unknown')
NSc=8 !quantidade de colunas com dados para combinação
NSi=3 !ponto inicial da combinação (fixando as duas primeiras colunas)
NSf=10 !ponto final da combinação
do i=1,NSc
call combinacao(NSi,NSf,i)
enddo
close(2)
open(unit=2,file='indices.txt',action='read',status='old',IOSTAT=IO)
call conta(2) !subrotina para contar a quantidade de dados
close(2)
NS2=Nc !quantidade de combinações de dados
!*****
*****
! 2. abrir o arquivo de dados e gerar arquivos de saída
open(unit=1,file='matriz.txt',action='read',status='old')
open(unit=2,file='indices.txt',action='read',status='old')
open(unit=3,file='residuos.txt',action='write',status='unknown')
open(unit=4,file='R_2.txt',action='write',status='unknown')
open(unit=5,file='EPR.txt',action='write',status='unknown')
open(unit=6,file='Nash.txt',action='write',status='unknown')
open(unit=7,file='Anomalia.txt',action='write',status='unknown')
open(unit=8,file='datas.txt',action='read',status='old')
open(unit=9,file='correlacao.txt',action='write',status='unknown')
open(unit=10,file='resultados.txt',action='write',status='unknown') !testes
```

```

open(unit=10,file='Sen_R2.txt',action='write',status='unknown')
open(unit=11,file='Sen_EPR.txt',action='write',status='unknown')
open(unit=12,file='Sen_Nash.txt',action='write',status='unknown')
open(unit=13,file='Sen_Residuos.txt',action='write',status='unknown')
open(unit=14,file='Sen_Anomalias.txt',action='write',status='unknown')
!*****
*****
! 3. Alocando o tamanho dos vetores
NS=11 !<---quantidade de colunas da matriz de dados----verificar antes de iniciar
allocate(A(N,NS+1))
allocate(A2(N,NS+1))
allocate(A3(N,NS))
allocate(IND(NS2,NS+1))
allocate(Niv(N))
allocate(Niv_c(N))
allocate(irr(N))
allocate(Dat(N,3))
allocate(E(N))
allocate(E1(N))

!*****
*****
! 4. Armazenar os dados na matriz utilizada pela rotina
do i=1,N
  read(1,*)(A3(i,j),j=1,NS) !armazenando todos os dados uma única coluna
enddo
do j=2,NS-1 !sem a coluna dos níveis e irrigação
  do i=1,N
    A2(i,j)=A3(i,j) !armazenando os dados na matriz que será usada para
  enddo !calcular os coeficientes da equação
enddo
do i=1,NS2
  read(2,*)(IND(i,j),j=1,NS) !armazena os dados na matriz de índices
enddo
do i=1,N
  Niv(i)=A3(i,1) !armazena os dados no vetor de níveis
enddo
do i=1,N
  irr(i)=A3(i,NS) !armazena os dados no vetor irrigação
enddo
do i=1,N
  A2(i,1)=Niv(i)+irr(i) !adicionando a coluna dos níveis somada com a irrigação
enddo
do i=1,N
  read(8,*)(Dat(i,j),j=1,3) !armazena os dados no vetor data
enddo

!*****
*****
! 5. Rotina principal
do h1=1,NS2 !todas as combinações de dados
  h3=IND(h1,1) !número de coeficiente
  call Organiza(h1,h3,N) !organiza os dados para a rotina miniqua
  call Miniqua(N,h3) !estima os parâmetros pelo método dos mínimos quadrados
  if(E(2)<0.001)then !limite mínimo para o coeficiente da precipitação
    goto 10
  endif
  E1=E !adicionado pois na análise de sensibilidade E vai variar
  call Simul(N,h3,h1) !calcula os níveis com a equação estimada
  call Dife(h1,N,di) !calcula as diferenças entre obs e calc
  if(di>3)then
    goto 10
  endif
enddo

```

```

endif
call R_2(N,h3,h1)      !coeficiente R²
if(R_Q_t<0.7)then
  goto 10
endif
call epr(h1,N,h3)      !erro padrão da regressão
call nash_s(N,h3,h1)  !nash-sutcliffe
if(F_t<0.7)then
  goto 10
endif
call anomalia(N,h1,ano_err) !anomalia
if(ano_err>0.9)then
  goto 10
endif
call correlacao(N,h3,h1) !colinearidade e multicolinearidade
!<----- escrever os resultados -----
write(100, '(11i9)')(IND(h1,i),i=2,11)
write(3, '(2i9)')h1,di
write(4,*)h1,R_Q_t
write(5,*)h1,epr_t
write(6,*)h1,F_t
write(7,*)h1,ano_err
!<----- Análise de sensibilidade dos modelos -----
call armazena(di,ano_err) !armazena os valores anteriores para análise de
sensibilidade
do h5=2,h3 !loop das variáveis com exclusão da primeira ()
  h5=2
  do h2=1,5 !loop para adicionar % de erros nas análises
    h2=5
    call porc_param(h2,h5)
    call Simul(N,h3,h1) !calcula os níveis com a equação estimada
    call Dife(h1,N,di) !calcula as diferenças entre obs e calc
    call R_2(N,h3,h1) !coeficiente R²
    call epr(h1,N,h3) !erro padrão da regressão
    call nash_s(N,h3,h1) !nash-sutcliffe
    call anomalia(N,h1,ano_err) !anomalia
    call analise_sen(h5,h2,h1,di,ano_err) !análise de sensibilidade
  enddo
enddo
!<-----
10 continue
enddo

endprogram Regressao

!<-----

module Modulo

real,allocatable::A(:, :) !matriz de variáveis organizada
real,allocatable::A2(:, :) !matriz de variáveis secundária
integer,allocatable::IND(:, :) !matriz de índices
real,allocatable::E(:) !parâmetros da equação
real,allocatable::E1(:) !parâmetros auxiliares
real,allocatable::Niv(:) !níveis observados
real,allocatable::Niv_c(:) !níveis calculados
real,allocatable::irr(:) !irrigação
real,allocatable::Dat(:, :) !vetor data
real,allocatable::vt_2(:, :) !vetor auxiliar subrotina "interp"
integer Nc, IO !sub-rotina conta !,MMPP
real R_Q_t !resposta do R²
real epr_t !resposta do EPR

```

```

real F_t      !Nash-S
real R_Q_t_a  !R² anterior
real epre_t_a !EPR anterior
real F_t_a    !Nash-S anterior
integer di_a  !resíduo anterior
real ano_err_a !anomalia anterior

end module

!<-----
subroutine analise_sen(x2,x1,x3,di,ano_err)
!#####
!##### Esta subrotina realiza as análises de #####
!##### sensibilidade das equações selecionadas #####
!#####
!obs.:x1=seleção da porcentagem;x2=parâmetro analisado
!x3=equação avaliada
use Modulo
implicit none
integer x1 !porcentagem reduzida
integer x2 !variável analisada
integer x3 !linha da equação
integer di !resposta das diferenças
real ano_err !resposta da anomalia
real E_A_R2 !erro absoluto do R2
real E_R_R2 !erro relativo do R2
real E_A_EPR !erro absoluto do EPR
real E_R_EPR !erro relativo do EPR
real E_A_F !erro absoluto do Nash
real E_R_F !erro relativo do Nash
real E_A_di !erro absoluto das diferenças
real E_R_di !erro relativo das diferenças
real E_A_ano !erro absoluto das anomalias
real E_R_ano !erro relativo das anomalias
!<----- R² -----
E_A_R2=R_Q_t-R_Q_t_a
E_R_R2=(abs(R_Q_t-R_Q_t_a)/R_Q_t_a)*100
write(10, '(7f9.4)')real(x3),real(x2),real(x1),E_A_R2,E_R_R2
!<----- EPR -----
E_A_EPR=epre_t-epre_t_a
E_R_EPR=(abs(epre_t-epre_t_a)/epre_t_a)*100
write(11, '(7f9.4)')real(x3),real(x2),real(x1),E_A_EPR,E_R_EPR
!<----- Nash -----
E_A_F=F_t-F_t_a
E_R_F=(abs(F_t-F_t_a)/F_t_a)*100
write(12, '(7f9.4)')real(x3),real(x2),real(x1),E_A_F,E_R_F
!<----- Resíduos -----
E_A_di=real(di)-real(di_a)
E_R_di=(abs(real(di)-real(di_a))/real(di_a))*100
write(13, '(7f9.4)')real(x3),real(x2),real(x1),E_A_di,E_R_di
!<----- Anomalia -----
E_A_ano=ano_err-ano_err_a
E_R_ano=(abs(ano_err-ano_err_a)/ano_err_a)*100
write(14, '(7f9.4)')real(x3),real(x2),real(x1),E_A_ano,E_R_ano
return
end

!<-----
subroutine anomalia(x2,x1,ano_err)
!#####
!##### Esta subrotina calcula as anomalias #####

```

```

#####
!obs.:x2=quantidade de dados;x1=linha da equação
use Modulo
implicit none
integer x1 !linha da equação
integer x2 !quantidade de dados
integer x3 !auxiliar dos loops
integer IDMES !auxiliar para o cálculo mensal
integer cont !contador (soma)
integer erro_tc !erro calculado
integer erro_to !erro observado
integer erro_toc !erro calculado com med observada
real M_Niv !média dos níveis
real sig !limite para o erro
real ano_err !resposta da anomalia
real,allocatable::Media_N(:) !média mensal observada
real,allocatable::Media_Nc(:) !média mensal calculada
real,allocatable::AP(:) !anomalia calculada
real,allocatable::AO(:) !anomalia observada
real,allocatable::AOc(:) !anomalia calculada com a média observada
allocate(Media_N(12))
allocate(Media_Nc(12))
allocate(AP(x2))
allocate(AO(x2))
allocate(AOc(x2))
!*****
*****
##### Série Total #####
! . Cálculo das médias mensais, ou seja, a média de todos os valores de janeiro,
fevereiro, ...
!médias observadas
do IDMES=1,12
  Media_N(IDMES)=0.0
  cont=0
  M_Niv=0.0
  do x3=1,x2
    if(IDMES==Dat(x3,2))then
      M_Niv=M_Niv+Niv(x3)
      cont=cont+1
    endif
  enddo
  Media_N(IDMES)=M_Niv/real(cont)
enddo
!médias calculadas
do IDMES=1,12
  Media_Nc(IDMES)=0.0
  cont=0
  M_Niv=0.0
  do x3=1,x2
    if(IDMES==Dat(x3,2))then
      M_Niv=M_Niv+Niv_c(x3)
      cont=cont+1
    endif
  enddo
  Media_Nc(IDMES)=M_Niv/real(cont)
enddo
!*****
*****
! . Cálculo das anomalias
!anomalias observadas
do x3=1,x2
  if(Dat(x3,2)==1)then

```

```

        AO(x3)=(Niv(x3)-Media_N(1))/Media_N(1)
    endif
    if(Dat(x3,2)==2)then
        AO(x3)=(Niv(x3)-Media_N(2))/Media_N(2)
    endif
    if(Dat(x3,2)==3)then
        AO(x3)=(Niv(x3)-Media_N(3))/Media_N(3)
    endif
    if(Dat(x3,2)==4)then
        AO(x3)=(Niv(x3)-Media_N(4))/Media_N(4)
    endif
    if(Dat(x3,2)==5)then
        AO(x3)=(Niv(x3)-Media_N(5))/Media_N(5)
    endif
    if(Dat(x3,2)==6)then
        AO(x3)=(Niv(x3)-Media_N(6))/Media_N(6)
    endif
    if(Dat(x3,2)==7)then
        AO(x3)=(Niv(x3)-Media_N(7))/Media_N(7)
    endif
    if(Dat(x3,2)==8)then
        AO(x3)=(Niv(x3)-Media_N(8))/Media_N(8)
    endif
    if(Dat(x3,2)==9)then
        AO(x3)=(Niv(x3)-Media_N(9))/Media_N(9)
    endif
    if(Dat(x3,2)==10)then
        AO(x3)=(Niv(x3)-Media_N(10))/Media_N(10)
    endif
    if(Dat(x3,2)==11)then
        AO(x3)=(Niv(x3)-Media_N(11))/Media_N(11)
    endif
    if(Dat(x3,2)==12)then
        AO(x3)=(Niv(x3)-Media_N(12))/Media_N(12)
    endif
enddo
!*****
*****
! . Cálculo das anomalias
!anomalias calculadas com as média calculadas
do x3=1,x2
    if(Dat(x3,2)==1)then
        AP(x3)=(Niv_c(x3)-Media_Nc(1))/Media_Nc(1)
    endif
    if(Dat(x3,2)==2)then
        AP(x3)=(Niv_c(x3)-Media_Nc(2))/Media_Nc(2)
    endif
    if(Dat(x3,2)==3)then
        AP(x3)=(Niv_c(x3)-Media_Nc(3))/Media_Nc(3)
    endif
    if(Dat(x3,2)==4)then
        AP(x3)=(Niv_c(x3)-Media_Nc(4))/Media_Nc(4)
    endif
    if(Dat(x3,2)==5)then
        AP(x3)=(Niv_c(x3)-Media_Nc(5))/Media_Nc(5)
    endif
    if(Dat(x3,2)==6)then
        AP(x3)=(Niv_c(x3)-Media_Nc(6))/Media_Nc(6)
    endif
    if(Dat(x3,2)==7)then
        AP(x3)=(Niv_c(x3)-Media_Nc(7))/Media_Nc(7)
    endif
endif

```

```

if(Dat(x3,2)==8)then
  AP(x3)=(Niv_c(x3)-Media_Nc(8))/Media_Nc(8)
endif
if(Dat(x3,2)==9)then
  AP(x3)=(Niv_c(x3)-Media_Nc(9))/Media_Nc(9)
endif
if(Dat(x3,2)==10)then
  AP(x3)=(Niv_c(x3)-Media_Nc(10))/Media_Nc(10)
endif
if(Dat(x3,2)==11)then
  AP(x3)=(Niv_c(x3)-Media_Nc(11))/Media_Nc(11)
endif
if(Dat(x3,2)==12)then
  AP(x3)=(Niv_c(x3)-Media_Nc(12))/Media_Nc(12)
endif
enddo
!*****
*****
! . Cálculo das anomalias calculadas com a média observada
do x3=1,x2
  if(Dat(x3,2)==1)then
    AOc(x3)=(Niv_c(x3)-Media_N(1))/Media_N(1)
  endif
  if(Dat(x3,2)==2)then
    AOc(x3)=(Niv_c(x3)-Media_N(2))/Media_N(2)
  endif
  if(Dat(x3,2)==3)then
    AOc(x3)=(Niv_c(x3)-Media_N(3))/Media_N(3)
  endif
  if(Dat(x3,2)==4)then
    AOc(x3)=(Niv_c(x3)-Media_N(4))/Media_N(4)
  endif
  if(Dat(x3,2)==5)then
    AOc(x3)=(Niv_c(x3)-Media_N(5))/Media_N(5)
  endif
  if(Dat(x3,2)==6)then
    AOc(x3)=(Niv_c(x3)-Media_N(6))/Media_N(6)
  endif
  if(Dat(x3,2)==7)then
    AOc(x3)=(Niv_c(x3)-Media_N(7))/Media_N(7)
  endif
  if(Dat(x3,2)==8)then
    AOc(x3)=(Niv_c(x3)-Media_N(8))/Media_N(8)
  endif
  if(Dat(x3,2)==9)then
    AOc(x3)=(Niv_c(x3)-Media_N(9))/Media_N(9)
  endif
  if(Dat(x3,2)==10)then
    AOc(x3)=(Niv_c(x3)-Media_N(10))/Media_N(10)
  endif
  if(Dat(x3,2)==11)then
    AOc(x3)=(Niv_c(x3)-Media_N(11))/Media_N(11)
  endif
  if(Dat(x3,2)==12)then
    AOc(x3)=(Niv_c(x3)-Media_N(12))/Media_N(12)
  endif
enddo
!*****
*****
! . escrevendo os resultados
do sig=0.1,5,0.1
  erro_tc=0

```

```

do x3=1,x2
  if(AP(x3)>=sig)then !<----- modificar
    erro_tc=erro_tc+1
  endif
enddo
erro_to=0
do x3=1,x2
  if(A0(x3)>=sig)then !<----- modificar
    erro_to=erro_to+1
  endif
enddo
erro_toc=0
do x3=1,x2
  if(A0c(x3)>=sig)then !<----- modificar
    erro_toc=erro_toc+1
  endif
enddo
if(erro_tc==0.and.erro_to==0.and.erro_toc==0)then
  ano_err=sig
  goto 3
endif
enddo
!<-----
3 continue
deallocate(Media_N,Media_Nc,AP)
return
end

!<-----

subroutine armazena(di,ano_err)
!#####
!##### Esta subrotina armazena os resultados obtidos #####
!#####
!##### lobs.:esta subrotina armazena os resultados obtidos com as equações
!##### selecionadas para futuras análises de sensibilidade
use Modulo
implicit none
integer di !resposta das diferenças em cm (resíduos)
real ano_err !resposta da anomalia em cm
!<----- R² -----
R_Q_t_a=R_Q_t
!<----- EPR -----
epre_t_a=epre_t
!<----- Nash -----
F_t_a=F_t
!<----- resíduo -----
di_a=di
!<----- anomalia -----
ano_err_a=ano_err
return
end

!<-----

subroutine combinacao(x1,x2,x3)
!#####
!##### Esta subrotina realiza todas as combinações #####
!##### possíveis entre os números fixandos #####
!#####
implicit none
integer::x1 !valor inicial
integer::x2 !valor final

```

```

integer::x3 !número de colunas
integer x8,x9,x10 !variáveis auxiliares
integer,allocatable::zeros(:) !vetor zeros
integer,allocatable::comb(:) !vetor resposta
allocate(comb(x3)) !alocação do vetor resposta
x8=(8-x3)!ajuste para o vetor zeros
allocate(zeros(x8))
!<-----
zeros=(/(0,x9=1,x8)/) !organizando o vetor zeros
x10=2+x3
!<-----
call iterate(x1,x2-x3+1,1) !subrotina que gera combinações
!<-----
contains
!<-----
recursive subroutine iterate(x4,x5,x6)
implicit none
integer,intent(IN)::x4,x5,x6
integer x7

do x7=x4,x5
comb(x6)=x7
if(x6.LT.x3)then
call iterate(x7+1,x5+1,x6+1)
endif
if(x6.EQ.x3)then
!write(*,*)comb
write(2,'(11i12)')x10,1,2,comb,zeros
endif
end do

end subroutine iterate
!<-----

end
!<-----

subroutine Conta(x1)
!#####
!##### Subrotina que conta a quantidade de dados #####
!##### contida em um arquivo .txt #####
!#####
!obs.:x1=número da unidade (arquivo)
use modulo
implicit none !remove todas as declarações padrões
integer x1
Nc=0
IO=0 !está rotina ler a quantidade de dados e
armazena
do while(IO>=0) !na variável N
read(x1,*,IOSTAT=IO)
if(IO==0)then
Nc=Nc+1
else
endif
enddo
return
end
!<-----

```

```

subroutine correlacao(x2,x3,x1)
!#####
!##### Esta subrotina realiza testes de correlação #####
!##### entre as variáveis #####
!#####
!obs.:x2=quantidade de dados;x3=quantidade de parâmetros
!x1=linha da equação
use modulo

integer x1 !linha da equação
integer x2 !quantidade de dados
integer x3 !quantidade de variáveis
integer x4,x5,x6 !auxiliares dos loops
real SUM1,SUM2 !auxiliares de soma
real Med_Dad1,Med_Dad2 !médias das variáveis em análise
real desv_Dad1,desv_Dad2 !desvio padrão das variáveis em análise
real C_1 !covariância
real r !coeficiente de correlação linear de Pearson
real,allocatable::desv_D1(:)
real,allocatable::desv_D2(:)
real,allocatable::qua_desv_D1(:)
real,allocatable::qua_desv_D2(:)
real,allocatable::w_t(:)
real,allocatable::cor(:)
real,allocatable::cor1(:)
real,allocatable::cor2(:)
allocate(desv_D1(x2))
allocate(desv_D2(x2))
allocate(qua_desv_D1(x2))
allocate(qua_desv_D2(x2))
allocate(w_t(x2))
allocate(cor1(x2))
allocate(cor2(x2))

!##### série total #####
do x4=2,(x3-1) !começa na variável 2 pois os níveis tem correlação com todos os
dados
do x5=x4+1,x3
! 1. média entre as variáveis
SUM1=0.0
SUM2=0.0
do x6=1,x2 !<-----
SUM1=SUM1+A(x6,x4)
SUM2=SUM2+A(x6,x5)
enddo
Med_Dad1=SUM1/real(x2) !<-----
Med_Dad2=SUM2/real(x2) !<-----
!2. desvio em relação a média
SUM1=0.0
SUM2=0.0
do x6=1,x2 !<-----
desv_D1(x6)=A(x6,x4)-Med_Dad1
desv_D2(x6)=A(x6,x5)-Med_Dad2
qua_desv_D1(x6)=(desv_D1(x6))**2
qua_desv_D2(x6)=(desv_D2(x6))**2
SUM1=SUM1+qua_desv_D1(x6)
SUM2=SUM2+qua_desv_D2(x6)
enddo
desv_Dad1=sqrt(SUM1/real(x2-1))
desv_Dad2=sqrt(SUM2/real(x2-1))
! 3. cálculo da covariância
SUM1=0.0

```

```

do x6=1,x2                                !<-----
  w_t(x6)=desv_D1(x6)*desv_D2(x6)
  SUM1=SUM1+w_t(x6)
enddo
C_1=SUM1/real(x2-1)                        !<-----
! 4. cálculo do coeficiente de correlação linear
r=0.0 !<-----
r=C_1/(desv_Dad1*desv_Dad2) !<-----
erro=0
if(r>=0.85)then !<-----
  erro=erro+1 !<-----
  write(9,*)x1,x4,x5,r
endif
enddo
enddo
!<-----

deallocate(desv_D1,desv_D2,qua_desv_D1,qua_desv_D2,w_t)

return
end

!<-----

subroutine dife(x1,x2,di)
!#####
!##### Esta subrotina calcula a diferença entre #####
!##### o valor observado e o calculado dos níveis #####
!#####
!obs.:x2=quantidade de dados;x1=linha da equação;
use Modulo
implicit none
! 0. Declarando variáveis, vetores e matrizes
integer x1 !linha da equação
integer x2 !quantidade de dados
integer x4 !variável auxiliar do loop
integer falha_dt !quantidade de falhas
integer di !resposta das diferenças <-----
real soma !auxiliar para soma
real x7 !limite para as diferenças
integer,allocatable::num(:)
real,allocatable::auxi_(:)
real,allocatable::dif(:)
allocate(num(x2))
allocate(auxi_(x2))
allocate(dif(x2))
!x7=0.3
!*****
*****
! . cálculo dos resíduos
do x4=1,x2
  dif(x4)=Niv(x4)-Niv_c(x4)
enddo
!<----- média e desvio padrão -----
soma=0.0
do x4=1,x2
  soma=soma+dif(x4)
enddo
!<-----
do x7=0.1,5,0.1
  falha_dt=0
  do x4=1,x2 !<-----limite total

```

```

        auxi_(x4)=abs(dif(x4))
        if(auxi_(x4)>=x7)then !<-----alterar
            falha_dt=falha_dt+1
        endif
    enddo
    if(falha_dt==0)then
        di=int(x7*10)
        !write(3,'(2i9)')x1,di
        goto 2
    endif
enddo
2 continue
return
end

!<-----

subroutine epr(x1,x2,x3)
!#####
!##### Esta subrotina calcula o erro padrão das #####
!##### regressões #####
!#####
!obs.:x1=linha da equação;x2=quantidade de dados;
!x3=quantidade de parâmetros
use Modulo

implicit none
integer x1 !linha da equação
integer x2 !quantidade de dados
integer x3 !quantidade de variáveis
integer x4 !auxiliar do loop
real auxi_1,auxi_2 !auxiliares para soma
real,allocatable::dif(:) !resíduos
allocate(dif(x2))
!*****
! .
do x4=1,x2
    dif(x4)=Niv(x4)-Niv_c(x4)
enddo
!total
auxi_2=0.0
do x4=1,x2
    auxi_1=dif(x4)**2.
    auxi_2=auxi_2+auxi_1
enddo
epre_t=0.0
epre_t=sqrt(auxi_2/real(x2-x3))
!<-----
!write(5,*)x1,epre_t

return
end

!<-----

subroutine miniqua(x2,x3)
!#####
!##### Esta subrotina estima o valor dos parâmetrosrealiza troca das colunas
!##### da regressão utilizando o métodos dos mínimos quadrados
!#####
!#####

```

```

#####
#####
!obs.:x2=quantidade de dados;x3=quantidade de parâmetros
use Modulo
implicit none
! 0. Declarando variáveis, vetores e matrizes
integer x1,x4,x5,x6,x10,x11 !variáveis auxiliares para os loops
integer x3 !quantidade de variáveis em cada matriz
integer x9 !limite da matriz e vetor auxiliares
integer x2 !quantidade de dados
real x7 !auxiliar do escalonamento
real x8 !soma na substituição retroativa
real,allocatable::soma1(:) !vetor auxiliar para soma
real,allocatable::soma2(:) !vetor auxiliar para soma
real,allocatable::AM(:,) !Matriz de dados auxiliar para realizar as regressões
real,allocatable::B(:,) !Matriz de dados auxiliar para realizar as regressões
real,allocatable::C(:,) !Matriz de dados auxiliar para realizar as regressões
real,allocatable::auxi(:,) !Matriz de dados auxiliar para realizar as regressões
!alocando vetores e matrizes
allocate(soma1(x3))
allocate(B(x3-1,x3))
allocate(C(x2,x3+1))
x9=x3*(x3-1) !limite da matriz e vetor auxiliares
allocate(soma2(x9))
allocate(auxi(x2,x9))
allocate(AM(x2,x3+1)) !alocando os dados na matriz da subrotina
!*****
*****
! 1. Armazenando os dados e calculando as componentes da matriz de regressão
!armazenando os dados na matriz da subrotina
do x10=1,x3
do x11=1,x2
AM(x11,x10)=A(x11,x10)
enddo
enddo
!calculando as componentes da matriz de regressão
do x4=1,x3
soma1(x4)=0.0 !zerando o valor inicial
do x1=1,x2
soma1(x4)=soma1(x4)+AM(x1,x4) !somando todas as colunas (variáveis)
enddo
enddo
x5=0
do x6=2,x3
do x4=1,x3
x5=x5+1
auxi(1,x5)=0.0
soma2(x5)=0.0
do x1=1,x2
auxi(x1,x5)=AM(x1,x6)*AM(x1,x4) !multiplicando cada coluna com as
outras colunas
soma2(x5)=soma2(x5)+auxi(x1,x5) !soma as colunas multiplicadas
enddo
enddo
enddo
!*****
*****
! 2. Montando as equações do sistema linear
C(1,1)=x2 !o primeiro valor constitui ao número de observações
C(1,x3+1)=soma1(1) !a primeira linha e ultima coluna é o somatório de Y
do x4=2,x3
C(1,x4)=soma1(x4) !as colunas da primeira linha são o somatório das colunas

```

```

    enddo
do x1=2,x3
    C(x1,1)=soma1(x1) !a primeira coluna é o somatório de todas as colunas
enddo
do x1=1,x3-1
    do x4=1,x3
        B(x1,x4)=0.0 !Y1.x1 x1.x1 x1.x2
        B(x1,x4)=soma2(x4+x3*(x1-1)) !Y1.x2 x1.x2 x2.x2
    enddo
enddo
do x1=1,x3-1
    do x4=2,x3 !x1.x1 x1.x2
        C(x1+1,x4)=B(x1,x4) !x1.x2 x2.x2
    enddo
enddo
do x1=1,x3-1 !Y1.x1
    C(x1+1,x3+1)=B(x1,1) !Y1.x2
enddo
!*****
*****
! 3. Escalonando o sistema
do x1=1,x3-1
    do x4=x1+1,x3
        x7=C(x4,x1)/C(x1,x1)
        do x5=x1,x3+1
            C(x4,x5)=C(x4,x5)-x7*C(x1,x5)
        enddo
    enddo
enddo
!*****
*****
! 4. Substituição retroativa
E(x3)=C(x3,x3+1)/C(x3,x3)
do x1=x3-1,1,-1
    x8=0.0
    do x4=x1+1,x3
        x8=x8+C(x1,x4)*E(x4)
    enddo
    E(x1)=(C(x1,x3+1)-x8)/C(x1,x1)
enddo
deallocate(AM)
return
end

!<-----

subroutine nash_s(x2,x3,x1)

!#####
!##### Esta subrotina calcula o coeficiente de #####
!##### Nash-Sutcliffe #####
!#####
!obs.:x2=quantidade de dados;x3=quantidade de parâmetros
!x1=linha da equação
use Modulo
implicit none
integer x1 !linha da equação
integer x2 !quantidade de dados
integer x3 !quantidade de variáveis
integer x4 !auxiliar dos loops
real aux1,soma_o,soma_c !auxiliares para soma
real Med_Niv_t !média dos níveis

```

```

!*****
*****
! . Coeficiente de NASH-SUTCLIFFE
!total
!cálculo da média dos níveis
auxil=0.0
do x4=1,x2
    auxil=auxil+Niv(x4)
enddo
Med_Niv_t=auxil/real(x2)
!cálculo das somas utilizadas pelo coeficiente
soma_o=0.0
soma_c=0.0
do x4=1,x2
    soma_c=soma_c+(Niv_c(x4)-Niv(x4))**2.
    soma_o=soma_o+(Niv(x4)-Med_Niv_t)**2.
enddo
!cálculo do coeficiente
F_t=1.-(soma_c/soma_o)
!-----
return
end

!-----

subroutine Organiza(x4,x2,x5)
!#####
!#####      Subrotina que organiza os dados para as      #####
!#####      matrizes de regressão                          #####
!#####
!obs.:x4=linha onde esta a equação;x2=quantidade de coeficientes;
!x5=quantidade de dados
use modulo
implicit none !remove todas as declarações padrões
integer x1,x2,x3,x4,x5,x6
do x1=2,x2+1 !
    x3=0
    x3=IND(x4,x1)
    !MMPP=0
    !if(x3==5)then !para ser utilizada a média móvel da precipitação prevista
    !    MMPP=x1-1
    !endif
    do x6=1,x5
        A(x6,x1-1)=A2(x6,x3) !matriz de variáveis organizada
    enddo
enddo
return
end

!-----

subroutine porc_param(x1,x2)
!#####
!#####      Esta subrotina reduz os parâmtros para      #####
!#####      análise de sensibilidade                      #####
!#####
!obs.:x1=seleção da porcentagem;x2=parâmetro analisado
use Modulo
implicit none
integer x1 !seleção da porcentagem a ser reduzida
integer x2 !parâmetro a ser reduzido
!<-----

```

```

select case(x1)
  case(1)
    E1(x2)=(E1(x2)*90.)/100. !redução de 10%
    !<-----
  case(2)
    E1(x2)=(E1(x2)*80.)/100. !redução de 20%
    !<-----
  case(3)
    E1(x2)=(E1(x2)*70.)/100. !redução de 30%
    !<-----
  case(4)
    E1(x2)=(E1(x2)*60.)/100. !redução de 40%
    !<-----
  case(5)
    E1(x2)=(E1(x2)*50.)/100. !redução de 50%
    !<-----
end select
return
end

!<-----

subroutine R_2(x2,x3,x1)
!#####
!##### Esta subrotina calcula o coeficiente R² #####
!#####
!obs.:x2=quantidade de dados;x3=quantidade de parâmetros
use Modulo

implicit none
integer x2 !quantidade de dados
integer x3 !quantidade de variáveis
integer x1 !linha da equação
integer x4 !auxiliar para os loops
real aux1,soma_1,soma_2,soma_3 !auxiliares para somas
real Med_Niv_t !média dos níveis
!*****
*****
! . Coeficiente de R²
!cálculo da média dos níveis
!total
aux1=0.0
soma_1=0.0
soma_2=0.0
do x4=1,x2
  aux1=aux1+Niv(x4)
enddo
Med_Niv_t=aux1/real(x2)
!cálculo das somas utilizadas pelo coeficiente
do x4=1,x2
  soma_1=soma_1+(Niv_c(x4)-Niv(x4))**2.
  soma_2=soma_2+(Niv_c(x4)-Med_Niv_t)**2.
enddo
soma_3=soma_1+soma_2
!cálculo do coeficiente
R_Q_t=1.-(soma_1/soma_3)
!-----
!write(4,*)x1,R_Q_t

return
end

```

```

!-----
subroutine simul(x2,x3,x4)
!#####
!##### Esta subrotina realiza simulações com as #####
!##### equações de regressão #####
!#####
!obs.:x2=quantidade de dados;x3=quantidade de parâmetros;
!x4=linha onde esta a equação
use Modulo
implicit none
integer x1,x2,x3,x4,x5,x7
real soma
do x7=1,x2
Niv_c(x7)=0.0
soma=0.0
do x1=2,x3
Niv_c(x7)=(A(x7,x1)*E1(x1))
soma=soma+Niv_c(x7)
enddo
Niv_c(x7)=E1(1)+soma-irr(x7)
enddo
return
end

```