

UNIVERSIDADE FEDERAL DO PARANÁ

ALANA RENATA RIBEIRO

PREVISÃO HIDROLÓGICA PROBABILÍSTICA: INFERÊNCIA BAYESIANA
APLICADA AOS ERROS DE PREVISÕES DETERMINÍSTICAS

CURITIBA
2017

ALANA RENATA RIBEIRO

PREVISÃO HIDROLÓGICA PROBABILÍSTICA: INFERÊNCIA BAYESIANA
APLICADA AOS ERROS DE PREVISÕES DETERMINÍSTICAS

Tese apresentada ao curso de Pós-Graduação em Métodos Numéricos em Engenharia, Área de Concentração em Programação Matemática, do Departamento de Matemática, Setor de Ciências Exatas e do Departamento de Construção Civil, Setor de Tecnologia, Universidade Federal do Paraná, como parte das exigências para a obtenção do título de Doutor em Ciências.

Orientador: Prof. Dr. Maurício Felga Gobbi

CURITIBA
2017

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

R484p

Ribeiro, Alana Renata

Previsão hidrológica probabilística: inferência bayesiana aplicada aos
erros de previsões determinísticas / Alana Renata Ribeiro. – Curitiba, 2017.
115 f. : il. color. ; 30 cm.

Tese - Universidade Federal do Paraná, Setor de Tecnologia, Programa
de Pós-Graduação em Métodos Numéricos em Engenharia, 2017.

Orientador: Maurício Felga Gobbi .

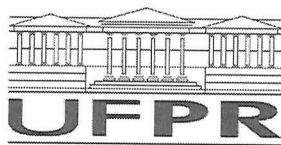
Bibliografia: p. 105-109.

Acesso dados abertos: http://dx.doi.org/10.11137/2017_2_266_277

1. Previsão hidrológica. 2. Vazão – Medidores. 3. Monte Carlo, Método de.
4. Markov, Processos de. 5. Teoria dos conjuntos . I. Universidade Federal
do Paraná. II. Gobbi, Maurício Felga. III. Título.

CDD: 551.489

Bibliotecário: Elias Barbosa da Silva CRB-9/1894



MINISTÉRIO DA EDUCAÇÃO
SETOR CIÊNCIAS EXATAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO MÉTODOS NUMÉRICOS
EM ENGENHARIA

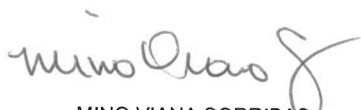
TERMO DE APROVAÇÃO

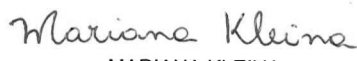
Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em MÉTODOS NUMÉRICOS EM ENGENHARIA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **ALANA RENATA RIBEIRO** intitulada: **PREVISÃO HIDROLÓGICA PROBABILÍSTICA: INFERÊNCIA BAYESIANA APLICADA AOS ERROS DE PREVISÕES DETERMINÍSTICAS**, após terem inquirido a aluna e realizado a avaliação do trabalho, são de parecer pela sua **APROVAÇÃO** no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

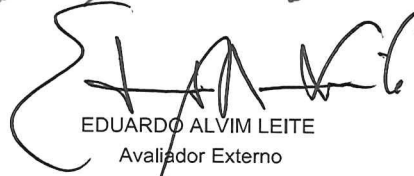
Curitiba, 15 de Dezembro de 2017.


MAURÍCIO FELGA GOBBI
Presidente da Banca Examinadora


MINO VIANA SORRIBAS
Avaliador Externo


MARIANA KLEINA
Avaliador Externo


PAULO JUSTINIANO RIBEIRO JÚNIOR
Avaliador Interno


EDUARDO ALVIM LEITE
Avaliador Externo

Dedico esta tese a minha avó (*in memoriam*),
Alair do Carmo Ribeiro (★ 1948 - † 2017)
Lala, te amo eternamente!

AGRADECIMENTOS

A Deus criador de todas as coisas, pela vida.

Ao meu amor Cleiton, pelo companheirismo e compreensão.

À minha mãe Iraci, ao meu pai Rogério, minha irmã Fabiane, meus sobrinhos Pedro e Isabelle, e a toda a minha família, por todo o apoio e dedicação.

Ao Professor Dr. Maurício Felga Gobbi, pela orientação.

Ao Dr. Eduardo Alvim Leite, por todos os ensinamentos.

A todos os meus amigos. Principalmente os que acompanharam a construção desse trabalho diariamente: Mariana, Diandra, Camila, Tiago, Renan, Jorge e Danilo.

Aos professores e funcionários do PPGMNE.

Aos colegas e funcionários do SIMEPAR.

A todos que contribuíram para a realização deste trabalho: o meu muito obrigada!

“O Senhor é meu pastor e nada me faltará.”

— Salmo 23.1

RESUMO

Neste trabalho modelos probabilísticos para previsão de vazão a curto prazo são desenvolvidos com base na inferência bayesiana aplicada em erros resultantes de previsões determinísticas. A inferência bayesiana permite a combinação entre informações prévias, sobre parâmetros de distribuições probabilísticas, com os erros cometidos por modelos determinísticos de previsão de vazão. Desta forma, os algoritmos de amostragem Monte Carlo via Cadeias de Markov facilitam a estimativa dos parâmetros de distribuições posteriores para as quais deseja-se extrair inferências. Os erros utilizados como dados de entrada para os modelos probabilísticos provêm de quatro modelos hidrológicos determinísticos diferentes: dois do tipo conceituais chuva-vazão, Sacramento e 3R; e outros dois empíricos, Regressivo Linear Múltiplo e Rede Neural do tipo *Multilayer Perceptron*. Apesar das diferenças existentes entre eles, os resultados obtidos demonstram que a abordagem probabilística adotada é capaz de considerar erros de diferentes tipos de modelos determinísticos. Os dados das séries horárias de vazão e precipitação utilizados, foram obtidos de uma estação automática de monitoramento pertencente ao Sistema Meteorológico do Paraná. Esta estação está localizada na sub-bacia de São Bento, pertencente à bacia do rio Iguaçu, no estado do Paraná, Brasil. Duas abordagens probabilísticas chamadas de estática e dinâmica são desenvolvidas. Na abordagem estática, os parâmetros da distribuição de probabilidade são mantidos fixos para qualquer momento de previsão. A abordagem dinâmica utiliza a abordagem estática como informação a priori e um conjunto de preditores do estado hidrometeorológico da sub-bacia em estudo, no momento da previsão, para variar os parâmetros da distribuição de probabilidade. Os modelos probabilísticos, estático e dinâmico, são aplicados aos quatro modelos determinísticos e avaliados por meio de testes estatísticos. O critério de informação de Akaike, o critério de informação bayesiano, o teste da razão da verossimilhança, e o Vuong teste, são os testes utilizados e permitem as comparações entre os valores máximos dos logaritmos das distribuições de probabilidade posteriores geradas. Os resultados demonstram a superioridade dos modelos probabilísticos com a abordagem dinâmica. Um novo modelo determinístico chamado *Naif* é construído a fim de incluir a climatologia de trinta anos de dados de vazão da região de estudo. Uma combinação entre os resultados dos modelos hidrológicos probabilísticos dinâmicos aplicados aos determinísticos e ao novo modelo *Naif*, permite a obtenção de um novo resultado determinístico chamado de *ensemble*. Os modelos probabilísticos, estático e dinâmico, são aplicados ao resultado do *ensemble*, promovendo uma melhora significativa na representação das incertezas inerentes ao processo de previsão de vazão. Os resultados da aplicação da previsão probabilística aplicada na previsão por conjunto demonstraram benefícios principalmente em casos de alertas de grandes cheias, preocupação constante dos órgãos de monitoramento hidrológico.

Palavras-chave: Previsão de vazão, Monte Carlo via Cadeias de Markov, *Ensemble*

ABSTRACT

In this work, short-term probabilistic streamflow forecast models are developed using bayesian inference applied to errors from deterministic forecasts. Bayesian inference allows a combination of previously obtained information about the parameters of the probabilistic distribution and the errors made by the deterministic streamflow forecast models. This way, the sampling algorithm Markov Chain Monte Carlo was used for the parameter estimation of posterior distributions that we inferred on. The errors used as input for the probabilistic models came from four different deterministic hydrological models. Two of them were rainfall-runoff conceptual models, Sacramento and 3R; and the remaining were empirical models, the Multiple Linear Regressive and the Multilayer Perceptron Neural Network. Despite the differences between them, the results obtained showed that the approach efficiently handled errors of different deterministic models. The hourly streamflow and precipitation time series used was obtained from an automatic monitoring station belonging to the Meteorological System of Paraná. This station is located on the São Bento subbasin, belonging to the Iguaçu river basin, on the state of Parana, Brazil. Two probabilistic approaches were built for the inference: static and dynamic. On the static approach, the probabilistic distribution parameters were fixed for any forecast time. On the other hand, the dynamic approach used the static one as prior information along with a set of hydrometeorological predictors from the subbasin state at the forecast time. The probabilistic models were applied to the four deterministic models and evaluated through statistical tests. The Akaike information criterion, the bayesian information criterion, the likelihood ratio test and the Vuong test were used for, they allow comparisons between the logarithms maximum values of the posterior probability distributions generated by each of them. The results showed the superiority of probabilistic models with the dynamic approach. A new deterministic model called *Naiif* was built, in order to have a model that includes thirty years of streamflow climatology in the studied region. A combination of the dynamic hydrological probabilistic models applied to the deterministic model's results and the *Naiif* model was made, allowing the obtention of a new deterministic result called ensemble. The probabilistic models, static and dynamic, are applied to the ensemble's results, yielding significant improvement on the representation of the uncertainties inherent to the streamflow forecast process. The results of the probabilistic forecast applied on the ensemble showed a better performance even in severe cases of floods, a constant preoccupation of hydrologic monitoring organs.

Key-words: Streamflow forecast, Markov Chain Monte Carlo, Ensemble

LISTA DE FIGURAS

FIGURA 1 – ESQUEMA DOS MODELOS PROBABILÍSTICOS	23
FIGURA 2 – ESQUEMA DOS MODELOS DETERMINÍSTICOS	23
FIGURA 3 – ESQUEMA DO PROCESSO DE MODELAGEM PROBABILÍSTICA DA ESTRUTURA DOS ERROS COMETIDOS POR MODELOS DE PREVISÃO DETERMINÍSTICOS	24
FIGURA 4 – ESQUEMA DO PROCESSO DE FORMAÇÃO DO <i>ENSEMBLE</i> ...	25
FIGURA 5 – CINCO SEQUÊNCIAS INDEPENDENTES DE UM MCMC PARA UMA DISTRIBUIÇÃO NORMAL PADRÃO, COM DIFERENTES PON- TOS DE PARTIDA REPRESENTADOS PELOS QUADRADOS SÓ- LIDOS. (A) APÓS 50 ITERAÇÕES AS SEQUÊNCIAS AINDA ES- TÃO LONGE DA CONVERGÊNCIA. (B) APÓS 1000 ITERAÇÕES AS SEQUÊNCIAS ESTÃO MAIS PRÓXIMAS DA CONVERGÊNCIA. (C) CONJUNTO DE SORTEIOS (CORRELACIONADOS) DA DISTRIBUI- ÇÃO A POSTERIORI ALVO	38
FIGURA 6 – APRESENTAÇÃO ESQUEMÁTICA DO MODELO SACRAMENTO	49
FIGURA 7 – ESTRUTURA DO MODELO 3R	51
FIGURA 8 – ARQUITETURA DE UMA REDE NEURAL MLP	55
FIGURA 9 – SUB-BACIA DE SÃO BENTO PERTENCENTE À BACIA DO RIO IGUAÇU	57
FIGURA 10– PERÍODOS DE TREINAMENTO E VERIFICAÇÃO DOS MODE- LOS DETERMINÍSTICOS E PROBABILÍSTICOS	58
FIGURA 11– VALORES DE REQM PARA PREVISÕES DOS QUATRO MODE- LOS DETERMINÍSTICOS QUE PODEM UTILIZAR PRECIPITAÇÃO OBSERVADA OU PREVISTA	60
FIGURA 12– VALORES DO COEFICIENTE DE <i>NASH-SUTCLIFFE</i> PARA OS QUATRO MODELOS DETERMINÍSTICOS UTILIZANDO PRECIP- ITAÇÃO PREVISTA	61

FIGURA 13– DISTRIBUIÇÕES DE PROBABILIDADE DOS ERROS POR HORIZONTE DE PREVISÃO DO MODELO DETERMINÍSTICO REGRESSIVO LINEAR	62
FIGURA 14– DISTRIBUIÇÃO DE PROBABILIDADE DOS ERROS PARA O HORIZONTE DE 72H DO MODELO DETERMINÍSTICO MLP	63
FIGURA 15– CORRELAÇÕES DO MODELO DETERMINÍSTICO SACRAMENTO. (A) CORRELAÇÃO ENTRE OS ERROS DE ACORDO COM O HORIZONTE DE PREVISÃO. (B) CORRELAÇÃO CRUZADA ENTRE ERROS COM DESLOCAMENTOS TEMPORAIS	63
FIGURA 16– COMPARAÇÃO ENTRE OS VALORES DO LOGARITMOS DA PROBABILIDADE POSTERIORI POR HORIZONTES DOS MODELOS PROBABILÍSTICOS ESTÁTICO E DINÂMICO, NO PERÍODO DE VERIFICAÇÃO	77
FIGURA 17– RESULTADOS DOS MODELOS PROBABILÍSTICOS APLICADOS AO MODELO DETERMINÍSTICO 3R NA DATA 13/06/2015 ÀS 16:00H (A) PREVISÕES DOS MODELOS ESTÁTICO COM AS DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS LINHAS PONTILHADAS COLORIDAS E PREVISÕES DO MODELO DINÂMICO COM DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS REGIÕES COLORIDAS. A DISPOSIÇÃO DAS CORES DO MODELO ESTÁTICO SE REFEREM ÀS CORES DO MODELO DINÂMICO CUJAS PROBABILIDADES SÃO APRESENTADAS (B) DISTRIBUIÇÕES DE PROBABILIDADE <i>SKEW-T</i> RESULTANTES DOS MODELOS ESTÁTICO E DINÂMICO OBTIDAS PARA O HORIZONTE DE 48H	79
FIGURA 18– RESULTADOS DOS MODELOS PROBABILÍSTICOS APLICADOS AO MODELO DETERMINÍSTICO SACRAMENTO NA DATA 13/06/2015 ÀS 16:00H (A) PREVISÕES DOS MODELOS ESTÁTICO COM AS DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS LINHAS PONTILHADAS COLORIDAS E PREVISÕES DO MODELO DINÂMICO COM DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS REGIÕES COLORIDAS. A DISPOSIÇÃO DAS CORES DO MODELO ESTÁTICO SE REFEREM ÀS CORES DO MODELO DINÂMICO CUJAS PROBABILIDADES SÃO APRESENTADAS (B) DISTRIBUIÇÕES DE PROBABILIDADE <i>SKEW-T</i> RESULTANTES DOS MODELOS ESTÁ-	

TICO E DINÂMICO OBTIDAS PARA O HORIZONTE DE 48H 81

FIGURA 19– RESULTADOS DOS MODELOS PROBABILÍSTICOS APLICADOS AO MODELO DETERMINÍSTICO SACRAMENTO NA DATA 04/11/2015 ÀS 04:00H (A) PREVISÕES DOS MODELOS ESTÁTICO COM AS DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS LINHAS PONTILHADAS COLORIDAS E PREVISÕES DO MODELO DINÂMICO COM DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS REGIÕES COLORIDAS. A DISPOSIÇÃO DAS CORES DO MODELO ESTÁTICO SE REFEREM ÀS CORES DO MODELO DINÂMICO CUJAS PROBABILIDADES SÃO APRESENTADAS (B) DISTRIBUIÇÕES DE PROBABILIDADE *SKEW-T* RESULTANTES DOS MODELOS ESTÁTICO E DINÂMICO OBTIDAS PARA O HORIZONTE DE 48H 82

FIGURA 20– RESULTADOS DOS MODELOS PROBABILÍSTICOS APLICADOS AO MODELO DETERMINÍSTICO REGRESSIVO NA DATA 18/05/2016 ÀS 16:00H (A) PREVISÕES DOS MODELOS ESTÁTICO COM AS DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS LINHAS PONTILHADAS COLORIDAS E PREVISÕES DO MODELO DINÂMICO COM DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS REGIÕES COLORIDAS. A DISPOSIÇÃO DAS CORES DO MODELO ESTÁTICO SE REFEREM ÀS CORES DO MODELO DINÂMICO CUJAS PROBABILIDADES SÃO APRESENTADAS (B) DISTRIBUIÇÕES DE PROBABILIDADE *SKEW-T* RESULTANTES DOS MODELOS ESTÁTICO E DINÂMICO OBTIDAS PARA O HORIZONTE DE 48H 83

FIGURA 21– RESULTADOS DOS MODELOS PROBABILÍSTICOS APLICADOS AO MODELO DETERMINÍSTICO REGRESSIVO NA DATA 18/05/2016 ÀS 22:00H (A) PREVISÕES DOS MODELOS ESTÁTICO COM AS DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS LINHAS PONTILHADAS COLORIDAS E PREVISÕES DO MODELO DINÂMICO COM DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS REGIÕES COLORIDAS. A DISPOSIÇÃO DAS CORES DO MODELO ESTÁTICO SE REFEREM ÀS CORES DO MODELO DINÂMICO CUJAS PROBABILIDADES SÃO APRESENTADAS (B) DISTRIBUIÇÕES DE PROBABILIDADE *SKEW-T* RESULTANTES DOS MODELOS ESTÁTICO E DINÂMICO OBTIDAS PARA O HORIZONTE DE 48H 84

FIGURA 22– EQM DOS MODELOS DETERMINÍSTICOS E DAS MODAS DOS MODELOS PROBABILÍSTICOS DINÂMICOS NO PERÍODO DE VE- RIFICAÇÃO	85
FIGURA 23– AVALIAÇÃO DOS MODELOS DETERMINÍSTICOS E DAS MODAS DOS MODELOS PROBABILÍSTICOS DINÂMICOS NO PERÍODO DE VERIFICAÇÃO SEGUNDO O COEFICIENTE DE <i>NASH-SUTCLIFFE</i>	86
FIGURA 24– VALORES DE REQM PARA PREVISÕES DOS QUATRO MODE- LOS DETERMINÍSTICOS E PARA O NOVO MODELO CLIMATOLÓ- GICO <i>NAIF</i>	89
FIGURA 25– COMBINAÇÃO DE DISTRIBUIÇÕES DE PROBABILIDADE <i>SKEW- T</i> PARA OBTENÇÃO DA MEDIANA DA DISTRIBUIÇÃO COMBINADA RESULTANTE PARA DETERMINADO HORIZONTE E MOMENTO DE PREVISÃO	91
FIGURA 26– VALORES DE <i>LPTH</i> DOS MODELOS DINÂMICOS <i>ENSEMBLE</i> E INDIVIDUAIS	91
FIGURA 27– EQM DOS MODELOS DETERMINÍSTICOS, E DAS MODAS DOS MODELOS PROBABILÍSTICOS DINÂMICOS	92
FIGURA 28– VALORES DOS COEFICIENTES DE <i>NASH-SUTCLIFFE</i> DOS MO- DELOS DETERMINÍSTICOS, E DAS MODAS DOS MODELOS PRO- BABILÍSTICOS DINÂMICOS	93
FIGURA 29– RESULTADOS DOS MODELOS PROBABILÍSTICOS APLICADOS AO <i>ENSEMBLE</i> NA DATA 02/10/2015 ÀS 22:00H (A) PREVISÕES DO MODELO ESTÁTICO COM AS DISTRIBUIÇÕES DE PROBA- BILIDADE ENTRE AS LINHAS PONTILHADAS COLORIDAS E PREVI- SÕES DO MODELO DINÂMICO COM DISTRIBUIÇÕES DE PROBA- BILIDADE ENTRE AS REGIÕES COLORIDAS. A DISPOSIÇÃO DAS CORES DO MODELO ESTÁTICO SE REFEREM ÀS CORES DO MODELO DINÂMICO CUJAS PROBABILIDADES SÃO APRESEN- TADAS (B) DISTRIBUIÇÕES DE PROBABILIDADE DOS MODELOS ESTÁTICO E DINÂMICO PARA O HORIZONTE DE 24H	94
FIGURA 30– RESULTADO DO MODELO PROBABILÍSTICO DINÂMICO, COM SEUS VALORES DE PROBABILIDADE DEFINIDOS ENTRE AS RE- GIÕES COLORIDAS, APLICADOS AO <i>ENSEMBLE</i> NA DATA DE 04/11/2015	

ÀS 04:00H	95
FIGURA 31– RESULTADO DO MODELO PROBABILÍSTICO DINÂMICO, COM SEUS VALORES DE PROBABILIDADE DEFINIDOS ENTRE AS REGIÕES COLORIDAS, APLICADOS AO <i>ENSEMBLE</i> NA DATA DE 20/09/2015	
ÀS 22:00H	96
FIGURA 32– RESULTADOS DOS MODELOS PROBABILÍSTICOS APLICADOS AO <i>ENSEMBLE</i> . PREVISÕES DO MODELO ESTÁTICO COM AS DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS LINHAS PONTILHADAS COLORIDAS E PREVISÕES DO MODELO DINÂMICO COM DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS REGIÕES COLORIDAS. A DISPOSIÇÃO DAS CORES DO MODELO ESTÁTICO SE REFERE ÀS CORES DO MODELO DINÂMICO CUJAS PROBABILIDADES SÃO APRESENTADAS (A) 04H DO DIA 18/05/2016 (B) 10H DO DIA 18/05/2016 (C) 16H DO DIA 18/05/2016 (D) 22H DO DIA 18/05/2016	
	97
FIGURA 33– RESULTADOS DOS MODELOS PROBABILÍSTICOS APLICADOS AO <i>ENSEMBLE</i> . PREVISÕES DO MODELO ESTÁTICO COM AS DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS LINHAS PONTILHADAS COLORIDAS E PREVISÕES DO MODELO DINÂMICO COM DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS REGIÕES COLORIDAS. A DISPOSIÇÃO DAS CORES DO MODELO ESTÁTICO SE REFERE ÀS CORES DO MODELO DINÂMICO CUJAS PROBABILIDADES SÃO APRESENTADAS (A) 10H DO DIA 23/09/2015 (B) 16H DO DIA 23/09/2015 (C) 22H DO DIA 23/09/2015 (D) 04H DO DIA 24/09/2015	
	98
FIGURA 34– RESULTADOS DOS MODELOS PROBABILÍSTICOS APLICADOS AO <i>ENSEMBLE</i> NA DATA 22/05/2016 ÀS 04:00H (A) PREVISÕES DO MODELO ESTÁTICO COM AS DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS LINHAS PONTILHADAS COLORIDAS E PREVISÕES DO MODELO DINÂMICO COM DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS REGIÕES COLORIDAS. A DISPOSIÇÃO DAS CORES DO MODELO ESTÁTICO SE REFEREM ÀS CORES DO MODELO DINÂMICO CUJAS PROBABILIDADES SÃO APRESENTADAS (B) DISTRIBUIÇÕES DE PROBABILIDADE DOS MODELOS	

ESTÁTICO E DINÂMICO PARA O HORIZONTE DE 24H	99
FIGURA 35– RESULTADOS DOS MODELOS PROBABILÍSTICOS APLICADOS AO <i>ENSEMBLE</i> NA DATA 25/05/2016 ÀS 04:00H (A) PREVISÕES DO MODELO ESTÁTICO COM AS DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS LINHAS PONTILHADAS COLORIDAS E PREVISÕES DO MODELO DINÂMICO COM DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS REGIÕES COLORIDAS. A DISPOSIÇÃO DAS CORES DO MODELO ESTÁTICO SE REFEREM ÀS CORES DO MODELO DINÂMICO CUJAS PROBABILIDADES SÃO APRESENTADAS (B) DISTRIBUIÇÕES DE PROBABILIDADE DOS MODELOS ESTÁTICO E DINÂMICO PARA O HORIZONTE DE 24H	100
FIGURA 36– RESULTADOS DOS MODELOS PROBABILÍSTICOS APLICADOS AO <i>ENSEMBLE</i> NA DATA 08/06/2014 ÀS 16:00H (A) PREVISÕES DO MODELO ESTÁTICO COM AS DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS LINHAS PONTILHADAS COLORIDAS E PREVISÕES DO MODELO DINÂMICO COM DISTRIBUIÇÕES DE PROBABILIDADE ENTRE AS REGIÕES COLORIDAS. A DISPOSIÇÃO DAS CORES DO MODELO ESTÁTICO SE REFEREM ÀS CORES DO MODELO DINÂMICO CUJAS PROBABILIDADES SÃO APRESENTADAS (B) DISTRIBUIÇÕES DE PROBABILIDADE DOS MODELOS ESTÁTICO E DINÂMICO PARA O HORIZONTE DE 24H	101
FIGURA 37– CÓDIGO DO MODELO ESTÁTICO MULTIVARIADO, COM POSTERIORI NORMAL MULTIVARIADA, PROGRAMADO EM LINGUAGEM STAN	110
FIGURA 38– CÓDIGO DO MODELO ESTÁTICO MULTIVARIADO, COM POSTERIORI <i>SKEW-T</i> MULTIVARIADA, PROGRAMADO EM LINGUAGEM STAN	111
FIGURA 39– CÓDIGO DA FUNÇÃO QUE REPRESENTA A DISTRIBUIÇÃO <i>SKEW-T</i> MULTIVARIADA, PROGRAMADO EM LINGUAGEM STAN	112
FIGURA 40– CÓDIGO DA FUNÇÃO QUE CONTROLA OS LIMITES DAS PERTURBAÇÕES DOS PARÂMETROS DO MODELO PROBABILÍSTICO DINÂMICO MULTIVARIADO, PROGRAMADO EM LINGUAGEM STAN	112
FIGURA 41– PRIMEIRA PARTE DO CÓDIGO DO MODELO DINÂMICO MULTI-	

VARIADO, COM POSTERIORI <i>SKEW-T</i> MULTIVARIADA, PROGRAMADO EM LINGUAGEM STAN	113
FIGURA 42– SEGUNDA PARTE DO CÓDIGO DO MODELO DINÂMICO MULTIVARIADO, COM POSTERIORI <i>SKEW-T</i> MULTIVARIADA, PROGRAMADO EM LINGUAGEM STAN	114
FIGURA 43– TERCEIRA PARTE DO CÓDIGO DO MODELO DINÂMICO MULTIVARIADO, COM POSTERIORI <i>SKEW-T</i> MULTIVARIADA, PROGRAMADO EM LINGUAGEM STAN	114
FIGURA 44– CÓDIGO DA FUNÇÃO <i>SKEW-T</i> UNIVARIADA UTILIZADA PARA O CÁLCULO DAS DISTRIBUIÇÕES MARGINAIS DO MODELO PROBABILÍSTICO DINÂMICO MULTIVARIADO, PROGRAMADO EM LINGUAGEM STAN	115
FIGURA 45– CÓDIGO DA FUNÇÃO <i>MARGINAL</i> UTILIZADA PARA O CÁLCULO DAS DISTRIBUIÇÕES MARGINAIS DO MODELO PROBABILÍSTICO DINÂMICO MULTIVARIADO, PROGRAMADO EM LINGUAGEM STAN	115

LISTA DE TABELAS

TABELA 1	– VALOR DO LOGARITMO DA PROBABILIDADE POSTERIORI TOTAL, OBTIDO COM A OTIMIZAÇÃO DO MODELO ESTÁTICO MULTIVARIADO COM POSTERIORI NORMAL MULTIVARIADA	68
TABELA 2	– VALOR DO LOGARITMO DA PROBABILIDADE POSTERIORI TOTAL OBTIDO COM A OTIMIZAÇÃO DO MODELO ESTÁTICO MULTIVARIADO COM POSTERIORI <i>SKEW-T</i> MULTIVARIADA	71
TABELA 3	– PERCENTUAL DE MELHORIA DO VALOR DO LOGARITMO DA PROBABILIDADE POSTERIORI TOTAL OBTIDO COM A OTIMIZAÇÃO DO MODELO ESTÁTICO MULTIVARIADO COM POSTERIORI <i>SKEW-T</i> MULTIVARIADA SUBSTITUINDO A POSTERIORI NORMAL MULTIVARIADA	71
TABELA 4	– VALORES DE AIC, BIC E VUONG DOS MODELOS ESTÁTICOS MULTIVARIADOS, COM POSTERIORI NORMAL OU <i>SKEW-T</i> MULTIVARIADAS, APLICADOS AO PERÍODO DE VERIFICAÇÃO	72
TABELA 5	– VALORES DO LOGARITMO DA PROBABILIDADE POSTERIORI TOTAL OBTIDO COM A OTIMIZAÇÃO DO MODELO DINÂMICO MULTIVARIADO	76
TABELA 6	– PERCENTUAL DE MELHORIA DOS VALORES DO LOGARITMO DA PROBABILIDADE POSTERIORI TOTAL OBTIDOS COM A OTIMIZAÇÃO DOS MODELOS DINÂMICOS MULTIVARIADOS SUBSTITUINDO OS MODELOS ESTÁTICOS	76
TABELA 7	– VALORES PARA O TESTE DA RAZÃO DA VEROSSIMILHANÇA COMPARANDO OS MODELOS ESTÁTICOS MULTIVARIADOS COM OS MODELOS DINÂMICOS, APLICADOS AO PERÍODO DE VERIFICAÇÃO	77
TABELA 8	– AVALIAÇÃO DE DESEMPENHO DO MODELO PROBABILÍSTICO BASEADO NA COMBINAÇÃO (<i>ENSEMBLE</i>) DOS MODELOS HIDROLÓGICOS PROBABILÍSTICOS DINÂMICOS	91

LISTA DE QUADROS

QUADRO 1 - DESCRIÇÃO DOS PARÂMETROS DA FASE BACIA DO MODELO
SACRAMENTO 50

QUADRO 2 - CONJUNTOS DE PREDITORES DOS ESTADOS HIDROLÓGICOS
UTILIZADOS NOS MODELOS PROBABILÍSTICOS DINÂMICOS 74

LISTA DE SIGLAS E ABREVIações

3R: Rainfall-Runoff-Routing
AIC: Critério de Informação de Akaike
ANA: Agência Nacional de Águas
AWBM: *Australian Water Balance Model*
BATEA: *Bayesian Total Error Analysis*
BFS: *Bayesian Forecasting System*
BIC: *Bayesian Information Criterion*
CEMIG: Companhia Energética de Minas Gerais
COPEL: Companhia Paranaense de Energia Elétrica
EQM: Erro Quadrático Médio
HMC: *Hamiltonian Monte Carlo*
Lpt: Logaritmo da probabilidade posterior total
Lpth: Logaritmo da probabilidade posterior total por horizonte
MC: Monte Carlo
MCMC: Monte Carlo via Cadeias de Markov
MCS: *Monte Carlo Simulation*
MFV: Máximo da Função de Verossimilhança
MLP: *Multilayer Perceptron*
NS: Coeficiente de *Nash-Sutcliffe*
REQM: Raiz do Erro Quadrático Médio
SIMEPAR: Sistema Meteorológico do Paraná
SNNS: *Stuttgart Neural Network Simulator*
TRV: Teste da Razão de Verossimilhança
VIC: *Variable Infiltration Capacity*

SUMÁRIO

1	INTRODUÇÃO	21
1.1	JUSTIFICATIVA	25
1.2	OBJETIVOS	28
1.2.1	Objetivo Geral	28
1.2.2	Objetivos Específicos	28
1.3	INOVAÇÕES PROPOSTAS	29
1.4	ESTRUTURA DO TRABALHO	29
2	INFERÊNCIA BAYESIANA	31
2.1	CONCEITOS	35
2.2	LINGUAGEM COMPUTACIONAL	40
2.2.1	Amostragens - <i>Sampling</i>	41
2.2.2	Otimizações - <i>Optmizing</i>	43
2.2.3	Logaritmo da probabilidade posterior total - <i>Lpt</i>	44
2.3	TESTES ESTATÍSTICOS PARA COMPARAÇÃO DE MODELOS	45
3	MODELOS HIDROLÓGICOS DETERMINÍSTICOS	47
3.1	MODELOS CONCEITUAIS CHUVA-VAZÃO	48
3.1.1	Modelo Sacramento	48
3.1.2	Modelo 3R	49
3.2	MODELOS EMPÍRICOS	50
3.2.1	Modelo Regressivo Linear Múltiplo	51
3.2.2	Modelo Rede Neural - MLP	54
3.3	DADOS E REGIÃO DE ESTUDO	57
3.4	COMPARAÇÃO ENTRE OS MODELOS DETERMINÍSTICOS	59

4	MODELOS HIDROLÓGICOS PROBABILÍSTICOS	65
4.1	MODELOS PROBABILÍSTICOS ESTÁTICOS MULTIVARIADOS	65
4.1.1	Ajuste de uma Distribuição Normal Multivariada	66
4.1.2	Ajuste de uma Distribuição Skew-t Multivariada	69
4.2	MODELOS PROBABILÍSTICOS DINÂMICOS MULTIVARIADOS	72
4.3	VISUALIZAÇÃO DOS RESULTADOS DOS MODELOS PROBABILÍSTICOS	78
4.4	PREVISÕES DETERMINÍSTICAS RESULTANTES DOS MODELOS PROBABILÍSTICOS	84
5	COMBINAÇÃO DAS PREVISÕES PROBABILÍSTICAS	87
5.1	CLIMATOLOGIA	88
5.2	<i>ENSEMBLE</i>	89
5.3	VISUALIZAÇÃO DOS RESULTADOS DO <i>ENSEMBLE</i>	93
6	CONCLUSÕES	102
6.1	TRABALHOS FUTUROS	104
	REFERÊNCIAS	105
	ANEXO A	110

1 INTRODUÇÃO

Estações hidrometeorológicas reúnem dados sobre níveis de rios (cotas), vazões, chuvas, evaporação, entre outros. A coleta destes dados é considerada uma importante ferramenta para a sociedade, pois os dados são utilizados para produzir estudos, definir políticas públicas e avaliar a disponibilidade hídrica (ANA, 2016). Por meio destas informações, a ANA (Agência Nacional de águas), o SIMEPAR (Sistema Meteorológico do Paraná), a COPEL (Companhia Paranaense de Energia Elétrica), entre outras entidades, monitoram eventos considerados críticos, como cheias e estiagens, disponibilizam informações para a execução de projetos, identificam o potencial energético, de navegação ou de lazer em um determinado ponto ou ao longo da calha do manancial, levantam as condições dos corpos de água para atender a projetos de irrigação ou de abastecimento público, entre outras aplicações (ANA, 2016).

Além disso, segundo Leite (2008), as informações hidrometeorológicas compostas pelo conjunto de informações qualitativas e quantitativas sobre o passado, presente e futuro da atmosfera e dos recursos hídricos, requerem uma análise do processo de produção de suas informações. Portanto, é necessário que se realize a previsão sistemática das condições do tempo, clima e recursos hídricos paralelamente à emissão de alertas de condições extremas e adversas.

As informações hidrometeorológicas em geral, e as previsões em particular, não são capazes de expressar, com toda a extensão e fidelidade, o estado atual e o comportamento futuro da atmosfera e dos recursos hídricos (WILKS, 1995, YEVJEVICH, 1974, WMO, 2007b ¹ apud LEITE, 2008). Nesse sentido, as informações climatológicas e de monitoramento incluem as incertezas provenientes dos processos imperfei-

¹WILKS, D. S., Statistical methods in the atmospheric sciences. **Academic Press**, San Diego, 1995; YEVJEVICH, V., Determinism and stochasticity in hydrology **Journal of Hydrology**, v. 22, p. 225-237, 1974; WMO, 2007B, WMO statement on the scientific basis for, and limitations of weather and climate forecasting. **WMO, Elements for life**, **Annexe 7**, Geneva, 2007.

tos de mensuração e estimação, enquanto que as previsões, baseadas em modelos preditivos de natureza objetiva ou subjetiva, incorporam as incertezas de propagação de erros, de compreensão e de representação dos fenômenos naturais. Portanto, pode-se afirmar que as informações hidrometeorológicas exibem limites de qualidade e previsibilidade, e que esses limites são impostos por questões de natureza teórica, tecnológica e econômica (KRZYSZTOFOWICZ, 2001², apud LEITE, 2008).

Incetezas com relação aos valores de vazão fluvial reduzem o conteúdo de suas informações para posteriores análises hidrológicas e afetam a confiabilidade do conhecimento que se pode inferir a partir destas análises (WESTERBERG *et al.*, 2011). Ao longo dos últimos anos, muitos estudos têm sido desenvolvidos com o intuito de classificar, quantificar, e minimizar tais incertezas (DIBALDASSARRE; MONTANARI, 2009; MCMILLAN *et al.*, 2010; DOMENEGHETTI; CASTELLARIN; BRATH, 2012), além de analisar seus possíveis efeitos sobre problemas de modelagem hidrológica (MONTANARI; BRATH, 2004; MONTANARI; GROSSI, 2008). Essas incertezas podem ser identificadas e comunicadas em um esforço específico próprio, de atualização periódica conforme o caso, ou ser incorporadas no próprio processo de geração e disseminação de informações, dando origem às chamadas previsões probabilísticas (LEITE, 2008).

Neste trabalho serão desenvolvidos dois tipos de modelos probabilísticos de previsão de vazão: um modelo probabilístico estático multivariado e um modelo probabilístico dinâmico multivariado. Os nomes dados aos modelos fazem referência à distribuição de probabilidade multivariada utilizada e às configurações dos parâmetros desta distribuição.

No modelo probabilístico do tipo estático, os parâmetros da distribuição de probabilidade são mantidos fixos para qualquer momento de previsão. No modelo probabilístico dinâmico, um conjunto de preditores do estado hidrometeorológico da sub-bacia em estudo é utilizado para condicionar os parâmetros da distribuição de probabilidade

²KRZYSZTOFOWICZ, R., The case for probabilistic forecasting in hydrology, **Journal of Hydrology**, v.249, n.1-4(Aug), pp. 2-9,2001.

de acordo com o momento de previsão (FIGURA 1).

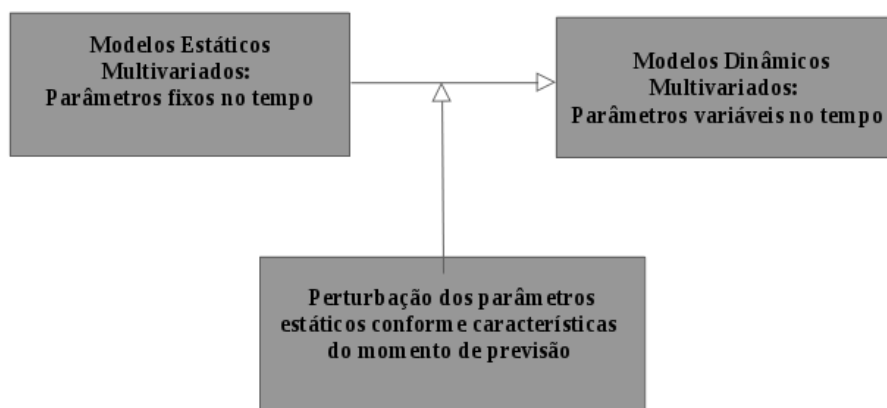


FIGURA 1: Esquema dos modelos probabilísticos

FONTE: A autora (2017)

Os modelos probabilísticos consideram a estrutura dos erros resultantes de diferentes modelos determinísticos de previsão de vazão (FIGURA 2), com o principal intuito de incorporar os resultados dessas previsões às informações obtidas a priori sobre os parâmetros das distribuições de probabilidade utilizadas, por meio da inferência bayesiana .

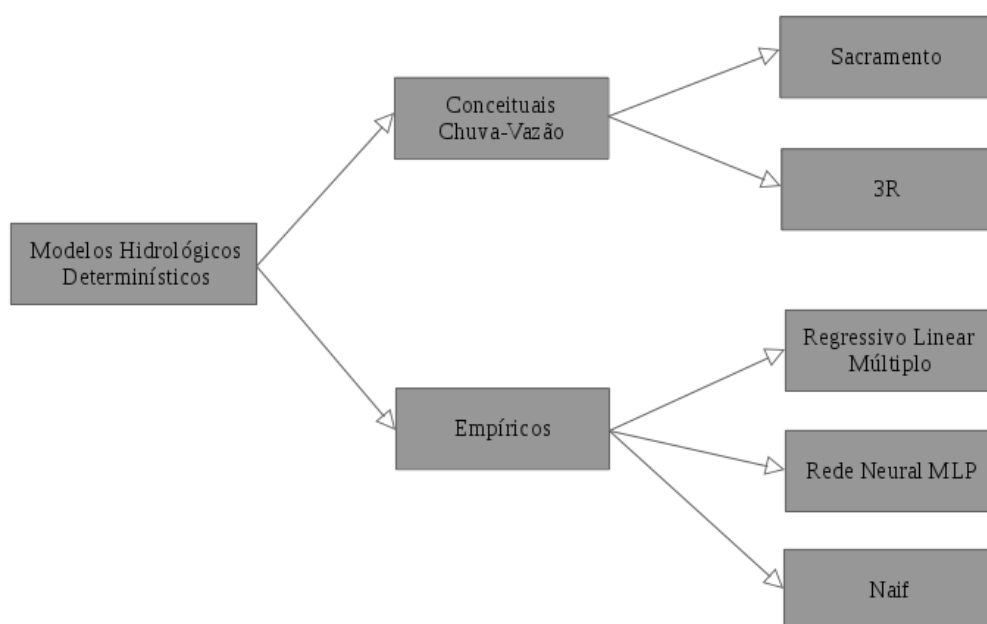


FIGURA 2: Esquema dos modelos determinísticos

FONTE: A autora (2017)

A Figura 3 esquematiza o processo da modelagem probabilística proposta. Resumidamente, os erros resultantes dos modelos determinísticos, de características distintas, são utilizados como entrada para modelos probabilísticos construídos a partir de informações a priori sobre os parâmetros de suas distribuições. Distribuições de probabilidade que representam a estrutura dos erros, valores de vazão previstos e incertezas globais em torno do processo de previsão desses valores, são determinadas.

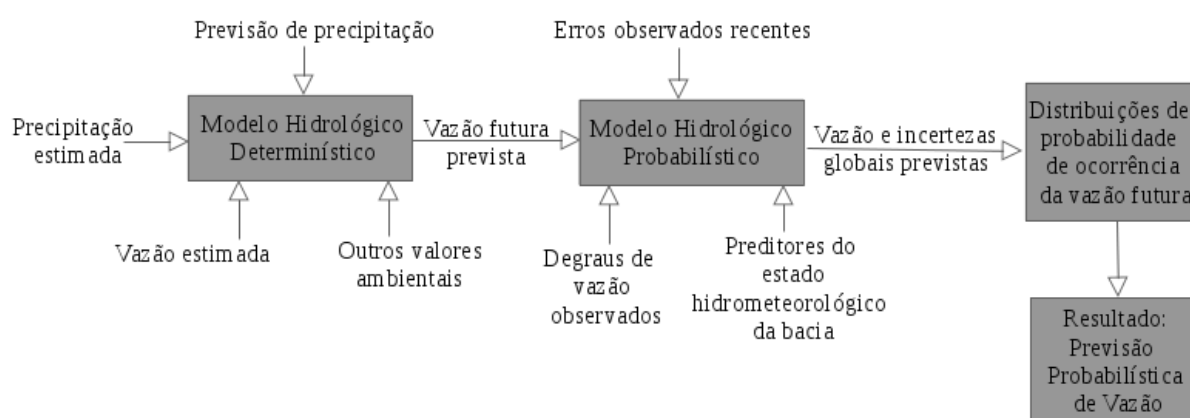


FIGURA 3: Esquema do processo de modelagem probabilística da estrutura dos erros cometidos por modelos de previsão determinísticos

FONTE: A autora (2017)

Desta forma, são obtidos os resultados de previsões probabilísticas estáticas. Os resultados dos modelos estáticos são utilizados como informação a priori para os modelos de previsões probabilísticas dinâmicas.

Testes estatísticos são utilizados para análise e comparação dos resultados obtidos com as abordagens construídas e suas diferentes aplicações. Um sistema de visualização dos resultados também é desenvolvido com o intuito de facilitar as análises.

Uma combinação (*ensemble*) dos resultados dos modelos probabilísticos dinâmicos, aplicados aos erros dos modelos determinísticos, é desenvolvida (FIGURA 4). Os objetivos da combinação dos modelos são: resumir as informações e resultados obtidos, e construir apenas uma previsão probabilística para a sub-bacia estudada neste trabalho.

A fim de completar o *ensemble*, proporcionando uma maior robustez ao modelo, os modelos probabilísticos são aplicados aos erros de um quinto modelo determinístico chamado de *Naif*. O modelo *Naif* faz uso de 30 anos de dados de vazão observados na região e foi desenvolvido para considerar a climatologia da região de estudo.

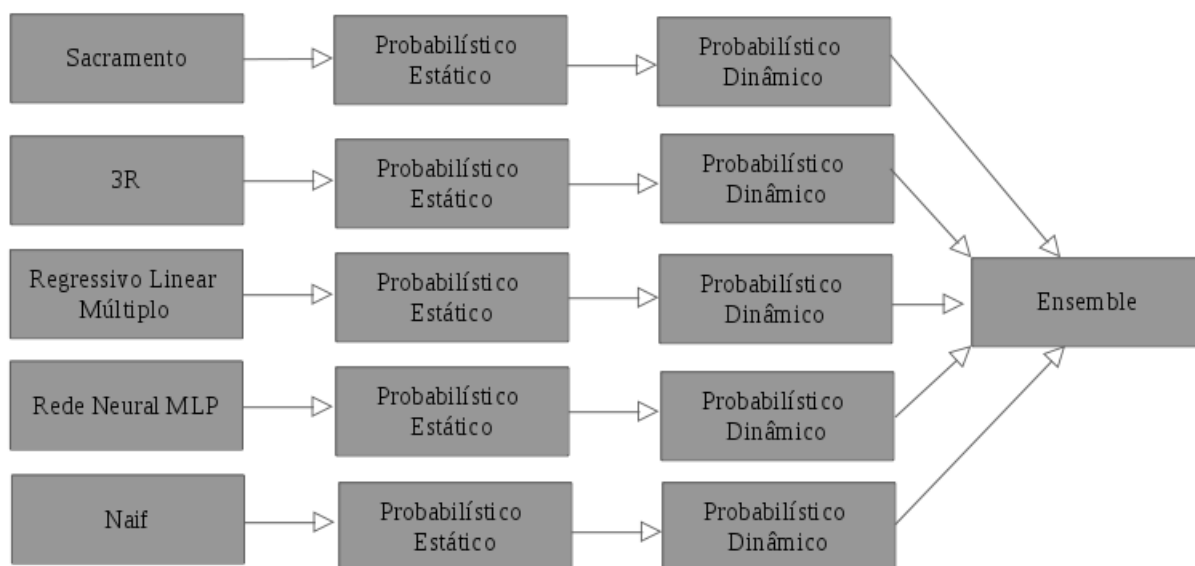


FIGURA 4: Esquema do processo de formação do *ensemble*

FONTE: A autora (2017)

Desta forma, o resultado do *ensemble* é utilizado para a construção do principal modelo probabilístico de previsão de vazão desenvolvido neste trabalho. Os resultados da previsão probabilística de vazão proposta, são considerados mais robustos, pois consideram toda a informação disponível da região de estudo. Espera-se portanto, que as previsões possam ser utilizadas, no futuro, como auxílio aos tomadores de decisões dos órgãos de monitoramento hidrológico e de gestão de recursos hídricos.

1.1 JUSTIFICATIVA

Eventos hidrológicos extremos como cheias e estiagens, e seus impactos são apenas alguns dos motivos que levam pesquisadores e órgãos de monitoramento e gestão de recursos hídricos do mundo inteiro a se preocuparem com a validade dos mais diversos tipos de dados hidrometeorológicos.

O monitoramento de dados hidrometeorológicos é realizado em postos, automáticos ou convencionais, onde são registrados os valores observados e armazenados em um banco de dados.

Além dos dados coletados, existem dados calculados para a operação de reservatórios, tal como vazão afluente, turbinada e vertida. Todavia, este monitoramento está sujeito à ocorrência de falhas, sejam erros provocados pelo equipamento de medição ou despropositadamente cometido pelo leitorista, implicando na inclusão de dados inconsistentes às séries de dados.

Além da presença de dados inconsistentes, outro fator que reduz a utilidade das séries é a ausência de informações (falhas) em alguns intervalos. Estes dados ausentes podem ocorrer por fatores tais como falha de transmissão e depredação dos equipamentos.

Outro problema bastante característico dos dados utilizados para a operação de reservatórios é a incerteza dos valores calculados. Ocorre com bastante frequência a geração de dados espúrios na contabilidade do balanço hídrico do reservatório.

Entretanto, com uma abordagem probabilística, os dados observados que são utilizados no balanço hídrico são expressos como uma distribuição de valores mais, ou menos, prováveis de serem os valores reais. Como consequência, o resultado do balanço também fornece uma informação na forma da distribuição para as variáveis estimadas. A vantagem disto é obter valores mais consistentes, considerando-se as incertezas intrínsecas das medições.

A análise probabilística de dados, estimativas e previsões, é uma abordagem recente que traz aos pesquisadores e usuários de sistemas de informações dados secundários, relevantes principalmente no processo de tomada de decisões.

Leite (2008) avaliou o valor da informação hidrometeorológica para o alerta de cheias no município paranaense de União da Vitória. Suas análises podem ser utilizadas para estabelecer níveis de atenção, alerta e urgência sobre cheias do Rio Iguaçu que venham a causar prejuízos no município.

Em um estudo posterior, Leite, Breda e Saboia (2011) apresentaram um sistema de previsão probabilística para União da Vitória mostrando como, de fato, a abordagem probabilística é importante na eficácia da tomada de decisão sobre a ocorrência de cheias e remoção de bens e pessoas dos locais de risco.

As previsões probabilísticas possuem a vantagem de serem mais completas, pois expressam as incertezas e permitem a derivação de melhores informações sob o ponto de vista do usuário.

Além disso, as previsões probabilísticas são apontadas como cientificamente mais honestas, pois: admitem e expressam as incertezas; permitem a emissão de avisos e alertas baseados em riscos; favorecem a tomada de decisão racional do usuário; agregam potencial para benefícios econômicos adicionais para toda a sociedade; e evitam mal entendidos nas atribuições de responsabilidade entre previsor e tomador de decisão, desacoplando os processos de previsão e uso das informações (KRZYSZTOFOWICZ, 1998; KRZYSZTOFOWICZ, 2001³, apud LEITE, 2008).

Finalmente, é grande a quantidade de modelos hidrológicos de previsão de vazão existentes, sejam eles determinísticos ou probabilísticos. Os usuários dos modelos de previsão, sejam eles agências de monitoramento e gestão de recursos hídricos, distribuidores de energia elétrica, entre outros tipos de empresas, calibram e utilizam esses modelos por muitos anos, aumentando a confiança em seus resultados e diminuindo a procura pelo desenvolvimento de novas abordagens. Desta forma, observa-se a necessidade de estimar incertezas de modelos que já estão sendo utilizados em larga escala, construir bandas de incerteza em torno de seus resultados, e considerar todas as incertezas intrínsecas ao processo de previsão, facilitando uma possível tomada de decisão resultante dos modelos hidrológicos.

³KRZYSZTOFOWICZ, R., Probabilistic hydrometeorological forecasts: toward a new era in operational forecasting, **Bulletin of the American Meteorological Society**, v.79, n. 2 (Feb), pp. 243-251, 1998
KRZYSZTOFOWICZ, R., The case for probabilistic forecasting in hydrology, **Journal of Hydrology**, v.249, n.1-4(Aug), pp. 2-9, 2001

1.2 OBJETIVOS

1.2.1 Objetivo Geral

Elaborar, por meio do uso de inferência bayesiana, uma previsão probabilística de vazão que considere a estrutura dos erros obtidos com a aplicação de diferentes modelos de previsão determinísticos. O sistema proposto contemplará um módulo de análise estatística das incertezas das previsões, com o intuito de explicitar mais informações sobre a vazão e disponibilizar bandas de incerteza de ocorrência das vazões previstas futuras.

1.2.2 Objetivos Específicos

Os objetivos específicos são:

- utilizar os erros cometidos por dois modelos determinísticos conceituais de chuva-vazão (Sacramento e 3R, extensivamente utilizados pelo SIMEPAR) como dados de entrada para os modelos probabilísticos que serão desenvolvidos;
- desenvolver e aplicar dois modelos determinísticos empíricos (Regressivo Linear e MLP) e utilizar seus erros como dados de entrada para os modelos probabilísticos que serão desenvolvidos;
- desenvolver, aplicar, avaliar e comparar duas abordagens probabilísticas derivadas da estrutura dos erros cometidos pelos modelos determinísticos, chamadas de: estática multivariada e dinâmica multivariada;
- desenvolver uma nova previsão determinística de previsão de vazão, chamada de *Naif*, para incorporar a climatologia da região de estudo;
- combinar os resultados do modelo probabilístico desenvolvido, de melhor desempenho, aplicado ao modelo *Naif* e aos modelos determinísticos, por meio de um *ensemble*, formando uma única previsão determinística que considere a climatologia e toda a informação disponível da sub-bacia estudada;

- aplicar a abordagem probabilística, de melhor desempenho, desenvolvida aos erros cometidos pelo *ensemble*, para formar um conjunto robusto de previsões probabilísticas para a sub-bacia estudada;
- apresentar os resultados da nova previsão probabilística por meio de um sistema de visualização que visa facilitar possíveis tomadas de decisão de um operador do sistema hidrológico.

1.3 INOVAÇÕES PROPOSTAS

Utilizar inferência bayesiana para criar bandas de incerteza em torno dos resultados de modelos determinísticos de previsão, desenvolvidos, calibrados e utilizados extensivamente por órgãos de gestão de recursos hídricos. Combinar seus resultados criando uma única previsão probabilística para a região de estudo para, futuramente, auxiliar o processo de tomada de decisão em um sistema robusto de alerta de enchentes e inundações.

1.4 ESTRUTURA DO TRABALHO

Este trabalho é composto de seis capítulos e um anexo. Sendo este primeiro o capítulo que introduz o tema a ser abordado justificando sua escolha e levantando os objetivos a serem atingidos.

No segundo capítulo serão discutidos os conceitos teóricos básicos utilizados para o desenvolvimento da pesquisa. Será apresentada uma visão geral sobre inferência bayesiana, sua utilização e as diferentes áreas de aplicação. Serão também apresentados algoritmos de amostragem específicos do tipo MCMC (Monte Carlo via Cadeias de Markov). Além disso, será feita uma breve descrição da linguagem computacional utilizada para implementação dos modelos probabilísticos.

O capítulo 3 apresenta um resumo do funcionamento de dois modelos hidrológicos determinísticos já calibrados e aplicados extensivamente pelo SIMEPAR, cha-

mados de conceituais chuva-vazão (Sacramento e 3R), e a construção e aplicação de dois modelos hidrológicos determinísticos empíricos (Regressivo Linear Múltiplo e rede neural MLP). Estes quatro modelos determinísticos serão utilizados para fornecer dados de entrada para os modelos probabilísticos que serão apresentados no capítulo 4. Ainda no terceiro capítulo os dados utilizados e a região de estudo serão apresentados, bem como uma comparação entre os resultados das previsões realizadas pelos modelos determinísticos.

O capítulo 4 apresenta a construção das duas abordagens probabilísticas de análise da estrutura de erros de previsão de vazão, os modelos probabilísticos estáticos multivariados e dinâmicos multivariados, e suas particularidades. É também apresentada a comparação e visualização dos resultados obtidos com a aplicação das duas abordagens probabilísticas aos quatro modelos determinísticos.

No capítulo 5 é apresentado o desenvolvimento de um novo modelo determinístico que considera a climatologia da região de estudo, chamado de *Naif*. Os modelos probabilísticos serão aplicados aos erros do *Naif* a fim de combinar este resultado com os resultados da previsão probabilística aplicada aos quatro modelos determinísticos, resultando em uma previsão determinística por *ensemble*. A abordagem probabilística desenvolvida anteriormente é aplicada aos erros do *ensemble*, resultando na previsão de vazão probabilística (principal objetivo deste trabalho), com suas bandas de incerteza.

O capítulo 6 apresenta as conclusões e as intenções de trabalhos futuros.

Por fim, no Anexo A os códigos desenvolvidos em linguagem Stan são apresentados.

2 INFERÊNCIA BAYESIANA

A melhoria de sistemas de previsão de vazão já existentes tem sido muito estudada nos últimos tempos, seja pela necessidade de emissão de alertas visando reduzir os danos causados por inundações, ou pelo desenvolvimento de novas técnicas de pesquisa e avanço computacional.

Muitos destes estudos consideram que utilizar uma abordagem determinística apenas, não tem sido suficiente para suprir tais necessidades. Com isso abordagens probabilísticas tem angariado cada vez mais espaço na área hidrológica.

Krzysztofowicz (1999) apresenta os fundamentos de um sistema de previsão bayesiano (*bayesian forecasting system* - BFS) para produzir uma previsão probabilística de um preditor hidrológico por meio de qualquer modelo determinístico. Segundo o autor, o BSF decompõe a incerteza total, intrínseca a um processo de previsão, em incerteza de entrada (associada às entradas aleatórias para o modelo como dados de precipitação previstos) e incerteza hidrológica (erros de modelo, parâmetro, estimação e medição), que são quantificadas de forma independente e posteriormente integradas em uma distribuição preditiva probabilística mediante o uso da inferência bayesiana. Além disso, o BSF possui a propriedade da auto-calibração (desde que o sistema de previsão que fornece a previsão probabilística de entradas aleatórias esteja, por sua vez, bem calibrado), protegendo o usuário, tomador de decisão, contra previsões notoriamente precárias. Em particular, se o modelo hidrológico não tem capacidade preditiva ou a densidade de entrada não é informativa para a previsão de saída, então a densidade preditiva converge automaticamente para a densidade prévia.

Krzysztofowicz (2002) apresentou uma síntese do sistema de previsão bayesiano (BFS) com suas expressões operacionais, seus procedimentos de estimativa, seus algoritmos numéricos, um exemplo completo e todos os requisitos de projeto, suposi-

ções de modelagem e atributos operacionais. Neste artigo, o autor desenvolve uma previsão probabilística a curto prazo do nível de rio, com base em uma previsão probabilística de precipitação quantitativa como entrada e um modelo hidrológico determinístico (de qualquer complexidade) como meio de simular a resposta de uma bacia de cabeceira, ressaltando que o uso do BSF pode ser adaptado para produzir uma previsão probabilística de vazão do rio, ou ainda uma previsão probabilística do volume de escoamento.

Montanari e Grossi (2008) propõem um método estatístico que se baseia no uso de um modelo estatístico meta-Gaussiano para quantificar a incerteza total das previsões hidrológicas. Os autores inferem sobre a distribuição de probabilidade do erro de previsão por meio de uma regressão múltipla com o objetivo de relacionar indiretamente os erros com as fontes de incerteza no processo de previsão. Os autores ajustam, com uma distribuição Gaussiana multivariada, variáveis aleatórias com distribuições marginais arbitrárias, incorporando a transformada *Normal Quantile Transformation* de cada variável na lei Gaussiana. O método proposto é operacionalmente simples e rápido e depende de suposições leves que são frequentemente atendidas em aplicações práticas. Em vista dessas características, esta técnica pode fornecer perspectivas úteis para estimar a incerteza das previsões hidrológicas em tempo real dentro de um quadro robusto. Os resultados indicam que as estimativas de incerteza representam com acurácia as estatísticas dos erros de previsão reais para os eventos analisados no estudo.

Nos últimos anos, a aplicação de inferência bayesiana e simulações MCMC também permitiram novas soluções para os problemas hidrológicos.

Marshall, Nott e Sharma (2004) comparam quatro algoritmos de amostragem do tipo MCMC aplicados à modelagem chuva-vazão por meio do modelo australiano de balanço hídrico (*Australian Water Balance Model* - AWBM), no sentido de comparar a facilidade, simplicidade, eficiência estatística na exploração do espaço paramétrico e velocidade de implementação, utilizando 11 anos de dados diários de precipitação

e vazão da bacia do rio Bass na Austrália. Os resultados obtidos demonstraram que a amostragem adaptada de Metropolis proposta por Haario, Saksman e Tamminen (2001), que exige menos especificações por parte do modelador, com relação às demais abordagens testadas, apresentou melhor desempenho em muitos aspectos e pode oferecer uma base relativamente simples para avaliar a incerteza de parâmetros em estudos de modelagem hidrológica.

Marshall, Nott e Sharma (2005) utilizaram a inferência bayesiana e os algoritmos MCMC, como meios para se avaliar e selecionar modelos hidrológicos, baseando-se no cálculo do fator de Bayes e estimando a distribuição de probabilidade marginal de cada modelo. A comparação de modelos por meio do fator de Bayes também fornece uma comparação formal do desempenho de modelos que assumem diferentes estruturas de erro.

Kavetski, Kuczera e Franks (2006a) desenvolveram uma metodologia bayesiana de análise de erros totais (BATEA - *Bayesian Total Error Analysis*) para modelos hidrológicos do tipo chuva-vazão que permite ao modelador incorporar, testar e refinar de forma direta a existência de todas as fontes de incerteza de dados em uma aplicação específica, incluindo incertezas de precipitação e de vazão. BATEA introduz variáveis latentes (não observáveis diretamente pelo pesquisador, ou que não podem ser medidas) adicionais à medida que a quantidade de dados incluídos na calibração aumenta, com isso os autores ressaltam a importância da utilização de métodos Monte Carlo (MC) e métodos numéricos para análise e otimização de funções, em combinação com o aumento do poder da computação, para obter soluções práticas e perspicazes para as equações de inferência e predição.

Kavetski, Kuczera e Franks (2006b) aplicam a metodologia BATEA em duas bacias hidrográficas norte-americanas: (1) French Broad River e (2) Potomac, e avaliam o desempenho do modelo conceitual *Variable Infiltration Capacity* (VIC) com e sem contabilização da incerteza de entrada (precipitação). Os resultados mostram os efeitos consideráveis de erros de precipitação nos hidrogramas (gráficos representativos

da variação, no tempo, de diversas observações hidrológicas, como cotas ou vazões) previstos (especialmente os limites de predição) e nos parâmetros calibrados.

DREAM (*Differential Evolution Adaptive Metropolis*), um novo modelo MCMC, foi desenvolvido por Vrugt *et al.* (2008) para estimar eficientemente a função densidade de probabilidade posterior de parâmetros do modelo hidrológico em problemas complexos de amostragem de alta dimensão. Além disso, o DREAM fornece novas maneiras de estimar a precipitação média das bacias hidrográficas, informação de extrema importância para testar a teoria hidrológica, diagnosticar erros estruturais em modelos e avaliar dispositivos de medição de precipitação adequadamente.

MCMC também foi utilizado por Benke, Lowell e Hamilton (2008), por Smith e Marshall (2008) e Vrugt *et al.* (2013), com o intuito de estimar as incertezas em torno dos parâmetros de diferentes modelos hidrológicos do tipo chuva-vazão.

McMillan *et al.* (2010) realizam simulações MCMC sobre a relação curva-chave (função que relaciona o nível do rio em uma seção com sua respectiva vazão) em um rio na Nova Zelândia, avaliando a incerteza e os erros das medidas de vazão que são utilizados para calibrar modelos chuva-vazão.

Ribeiro *et al.* (2017) testaram a inferência bayesiana para a construção de regiões de confiabilidade probabilística em torno de curvas-chave de diferentes postos fluvio-métricos monitorados pela CEMIG (Companhia Energética de Minas Gerais). Os resultados obtidos demonstraram que os algoritmos de amostragem MCMC permitiram a realização de uma avaliação das incertezas associadas à utilização das curvas-chave. O trabalho de Ribeiro *et al.* (2017) deu origem à esta pesquisa de doutorado.

Técnicas de amostragem baseadas em simulações de Monte Carlo, no contexto da análise de incerteza de modelagem hidrológica, são avaliadas com a finalidade de reduzir os custos computacionais e fornecer orientações sobre a sua utilização e avaliação de desempenho (ASSUMPÇÃO, 2015). Ainda neste estudo, é apresentado o desenvolvimento e aperfeiçoamento dos métodos de avaliação da qualidade de um método de amostragem, chamado de *Monte Carlo Simulation* (MCS), utilizado na mo-

delagem para avaliar a incerteza dos resultados. O estudo de caso utilizado foi um modelo hidrológico conceitual aplicado à bacia de Bagmati, no Nepal.

Enfim, é extensa a quantidade de pesquisas hidrológicas que utilizam a inferência bayesiana e suas técnicas de amostragem.

2.1 CONCEITOS

Segundo Bonat *et al.* (2012), a inferência estatística, de maneira geral, tem como objetivo fornecer subsídios, de natureza quantitativa, para a tomada de decisões a partir de dados coletados em estudos amostrais, experimentais ou observacionais.

Por sua vez, a inferência bayesiana considera informações prévias à coleta dos dados (informação a priori) sobre os parâmetros do modelo probabilístico, além da informação contida nos dados. Ambas as informações, a priori e dos dados, são combinadas usando o teorema de Bayes, gerando a informação a posteriori sobre os parâmetros do modelo (GELMAN *et al.*, 2003).

Segundo Gelman *et al.* (2013), a característica essencial dos métodos bayesianos é a utilização explícita da teoria de probabilidade para quantificar a incerteza em inferências com base em análise estatística. Os métodos consideram a utilização de parâmetros fixos para os modelos, porém desconhecidos, fazendo uso das distribuições a priori e posteriori sobre os parâmetros do modelo. Essas probabilidades a priori e a posteriori e as probabilidades preditivas posteriores caracterizam o conhecimento sobre os parâmetros do modelo e sobre observações futuras. As distribuições posteriores formam a base da inferência bayesiana.

A modelagem bayesiana pode ser resumida em três etapas:

1. construção de um modelo de probabilidade total para todas as quantidades observáveis e não observáveis. Este modelo deve ser consistente com o conhecimento dos dados existentes e com a maneira com a qual eles foram coletados;
2. cálculo da probabilidade posterior de quantidades desconhecidas condicionadas

às quantidades observadas. As incógnitas podem incluir quantidades não observadas, como parâmetros do modelo e quantidades potencialmente observáveis, tais como previsões para futuras observações;

3. avaliação dos modelos de ajuste aos dados, isto inclui a avaliação das consequências da distribuição posterior.

Este ciclo é repetido até que um ajuste aceitável seja obtido na terceira etapa.

A construção de modelos baseados em inferência bayesiana utiliza o teorema de Bayes combinando dados e informação a priori sobre os parâmetros do modelo.

O teorema de Bayes fornece uma expressão para a probabilidade condicional de A dado B :

$$P(A|B) = P(B|A)P(A)/P(B) \quad (1)$$

substituindo B por um vetor de observações y , A por um conjunto de parâmetros θ , e $P(\cdot)$ por funções de probabilidade ou densidades de probabilidade representadas por $p(\cdot)$, tem-se:

$$p(\theta|y) = p(y|\theta)p(\theta)/p(y) \quad (2)$$

em que $p(\theta)$ é a distribuição conjunta a priori para o vetor de parâmetros θ independentemente de y , $p(y|\theta)$ representa a verossimilhança de y , $p(\theta|y)$ é a distribuição a posteriori conjunta para o vetor de parâmetros θ . O denominador dado por:

$$p(y) = \int p(y|\theta)p(\theta)d\theta \quad (3)$$

define a verossimilhança marginal y .

Porém, a quantidade representada pela Equação (3) não precisa ser calculada para fins de inferência sobre os parâmetros do modelo, pois não depende dos parâmetros, podendo ser denotada como uma constante de proporcionalidade c . Então, a posteriori torna-se:

$$p(\theta|y) \propto p(y|\theta)p(\theta) \quad (4)$$

Nas aplicações propostas neste trabalho, y representa os erros obtidos com as aplicações de diferentes modelos hidrológicos do tipo determinísticos (apresentados no capítulo 3), θ são os parâmetros das distribuições de probabilidade, a priori e a posteriori ($p(\cdot)$), que serão testadas e aplicadas aos modelos hidrológicos probabilísticos (apresentados no capítulo 4).

O produto entre a verossimilhança $p(y|\theta)$ e a priori $p(\theta)$ na Equação (4), representa a distribuição conjunta não normalizada dos parâmetros. Esta posteriori só possui expressão definida para um conjunto limitado de problemas. Assim, utilizam-se aproximações analíticas, ou numéricas, ou simulação estocástica para obter uma aproximação desta distribuição, e infere-se sobre ela a partir de simulações de seus valores, geralmente por meio de algoritmos conhecidos como métodos MCMC.

MCMC são métodos gerais que possuem como base a determinação de valores dos parâmetros (aqui representados por θ) de distribuições de probabilidade aproximadas e que, após uma correção visam aproximar a distribuição a posteriori $p(\theta|y)$. As amostras são obtidas sequencialmente, com a distribuição da amostragem determinada em função do último valor obtido, formando uma cadeia de Markov.

Segundo Gelman *et al.* (2003), uma cadeia de Markov é uma sequência de variáveis aleatórias $\theta^{(1)}, \theta^{(2)}, \dots$, para a qual, para qualquer t , a distribuição de $\theta^{(t)}$ dados todos os θ' s anteriores depende somente do valor mais recente, $\theta^{(t-1)}$. Este fato é devido à propriedade de Markov que garante que a distribuição de probabilidade condicional de estados futuros do processo (condicional tanto em estados passados, como presentes), depende apenas do estado presente, não da sequência de eventos que o precedeu.

A propriedade de Markov é útil para provar a convergência dos métodos MCMC, que garante que as distribuições de probabilidade sejam melhor aproximadas em cada passo na simulação, no sentido de convergir para a distribuição posteriori alvo.

A Figura 5 ilustra um exemplo do algoritmo Metropolis (METROPOLIS *et al.*, 1953) do tipo MCMC, em que θ é um vetor de apenas duas componentes (θ_x e θ_y), com uma

distribuição a posteriori do tipo normal padrão ($\theta \sim N(0, 1)$).

Na Figura 5 (a), fases iniciais da simulação são retratadas, ou seja, depois de 50 iterações as sequências ainda estão longe de convergir para uma distribuição. Essa figura representa a gama de possíveis valores do parâmetro multivariado θ , e cada um dos cinco percursos irregulares representa o caminho no início de um passeio aleatório começando perto do centro ou dos extremos da distribuição alvo e salta por meio da distribuição de acordo com uma sequência apropriada de iterações aleatórias.

A Figura 5 (b) representa a fase de maturidade do mesmo MCMC (depois de 100 iterações as sequências estão perto de convergir), na qual os caminhos aleatórios simulados têm um traço para cada caminho ao longo do espaço de θ , com uma distribuição estacionária comum que é igual à distribuição a posteriori alvo.

A partir de uma simulação como essa, inferências sobre θ podem ser realizadas utilizando os pontos da segunda metade das cadeias de Markov simuladas, conforme apresentado na Figura 5 (c).

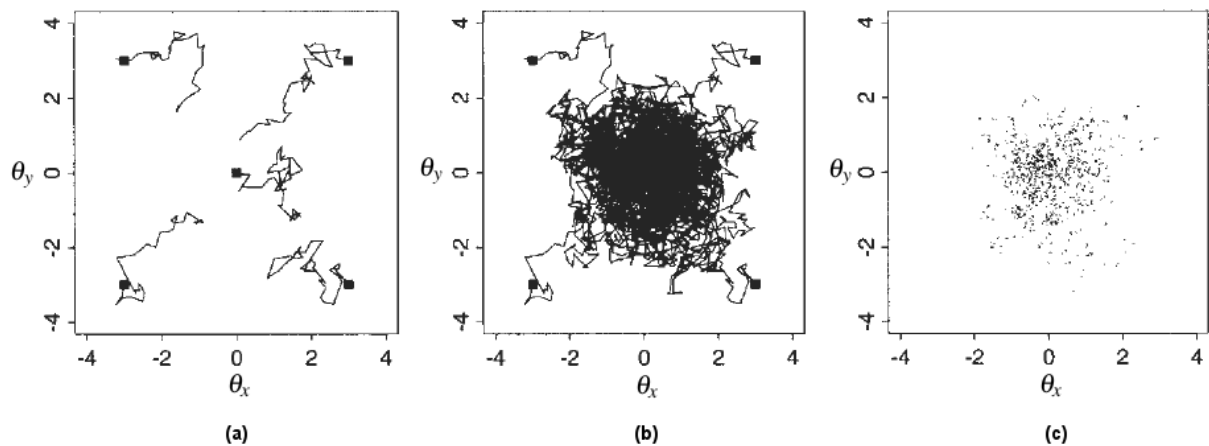


FIGURA 5: Cinco sequências independentes de um MCMC para uma distribuição normal padrão, com diferentes pontos de partida representados pelos quadrados sólidos. (a) Após 50 iterações as sequências ainda estão longe da convergência. (b) Após 1000 iterações as sequências estão mais próximas da convergência. (c) Conjunto de sorteios (correlacionados) da distribuição a posteriori alvo

FONTE: Gelman *et al.* (2003)

Nas aplicações de MCMC, várias sequências independentes são criadas; cada sequência $\theta^{(t)}$, $t = 1, 2, 3, \dots$, é produzida iniciando em algum ponto $\theta^{(0)}$ e, em seguida,

para cada t determinando $\theta^{(t)}$ a partir de uma distribuição de transição, $T_t(\theta^{(t)}|\theta^{(t-1)})$ que depende da determinação anterior $\theta^{(t-1)}$.

O MCMC é utilizado quando não é possível (ou não é computacionalmente eficiente) obter amostras de θ diretamente de $p(\theta|y)$. Desta forma, as amostras são obtidas de forma iterativa de tal maneira que, em cada passo do processo, espera-se determinar uma distribuição que se torna cada vez mais próxima de $p(\theta|y)$.

Para uma ampla classe de problemas (incluindo distribuições posteriores para muitos modelos hierárquicos), esta abordagem parece ser a maneira mais fácil de obter resultados confiáveis, pelo menos quando usada com cuidado (GELMAN *et al.*, 2003). Além disso, a cadeia de Markov e outros métodos de simulação iterativos tem muitas aplicações fora da estatística bayesiana, como a otimização.

Uma vez que o algoritmo de simulação é implementado e as simulações determinadas, faz-se necessário verificar a convergência das sequências simuladas. Por exemplo, as simulações da Figura 5 (a) estão longe da convergência e não estão perto de uma distribuição alvo. Se a convergência for muito lenta, o algoritmo deve ser alterado.

Existem diversos tipos de algoritmos chamados de MCMC. Entre eles, Metropolis-Hastings (METROPOLIS *et al.*, 1953; HASTINGS, 1970) e amostrador de Gibbs (GELMAN; GEMAN, 1984; GELFAND; SMITH, 1990) são os mais populares. Eles pertencem a uma classe de algoritmos de simulação sequencial gerando uma cadeia que, após atingir convergência, passa a gerar amostras da distribuição a posteriori (BONAT *et al.*, 2012).

Outro algoritmo conhecido como HMC (*Hamiltonian Monte Carlo*, ou *Hybrid Monte Carlo*) (DUANE *et al.*, 1987) também é um método MCMC que visa obter uma sequência de amostras aleatórias de uma distribuição probabilística para a qual a amostragem direta é muito complicada, esta sequência pode ser utilizada para aproximar a distribuição, ou para calcular sua integral.

Segundo Alfaki (2008), o HMC combina a atualização de amostragem de Gibbs

com a regra de aceitação-rejeição de Metropolis, obtendo melhores propriedades de convergência, e rapidez em sua aplicação. Portanto, o algoritmo HMC é aplicado aos erros dos modelos determinísticos de previsão de vazão para o desenvolvimento deste estudo, por meio da linguagem de modelagem probabilística Stan (Stan Development Team, 2014), que possui como característica o uso do HMC.

2.2 LINGUAGEM COMPUTACIONAL

Métodos de inferência bayesianos são, em geral, implementados por meio de métodos computacionalmente intensivos, exigindo a utilização de *softwares* específicos para este fim.

Para processamento dos dados previamente utilizados na construção dos modelos probabilísticos, e para processamento das inferências obtidas após sua aplicação, neste estudo, faz-se uso do *software* estatístico R (R Core Team, 2012).

R é um ambiente de análise estatística, com código fonte aberto, podendo ser modificado ou implementado com novos procedimentos desenvolvidos pelo próprio usuário. É uma linguagem de programação de utilização fácil e completa, amplamente utilizada nas mais diversas áreas científicas, especialmente no campo estatístico (KLEINA, 2015).

Além disso, para a construção dos modelos probabilísticos, fez-se uso da linguagem de alto nível Stan (GELMAN *et al.*, 2013) em que o usuário especifica um modelo, fornece valores iniciais, e implementa simulações da cadeia de Markov automaticamente para as distribuições posteriores resultantes.

Um programa construído no Stan define um modelo estatístico por meio de uma função de probabilidade condicional $p(\theta|y,x)$, onde θ é uma sequência de valores modelados desconhecidos (os parâmetros do modelo, variáveis latentes, dados em falta ou previsões futuras), y é uma sequência de valores modelados conhecidos, e x é uma sequência de preditores não modelados e constantes (parâmetros ou hiperparâmetros

das distribuições de probabilidade).

Os modelos construídos em linguagem Stan são compostos por blocos: blocos de novas funções que ainda não foram implementadas; de declarações de variáveis; de dados transformados; parâmetros; parâmetros transformados; quantidades geradas e um bloco especial do modelo que consiste em declarações que definem o logaritmo da probabilidade posterior total (L_{pt}) para o modelo.

Quando se executa um modelo definido no Stan primeiro leem-se e validam-se os valores conhecidos y e x , em seguida, gera-se uma sequência de amostras não independentes identicamente distribuídas $\theta^{(1)}, \theta^{(2)}, \dots$, cada uma das quais seguindo distribuições marginais $p(\theta|y, x)$.

O modelo pode ser executado utilizando-se amostragens (função *sampling*), ou otimizações (função *optimizing*). E, para que os resultados dos modelos construídos sejam avaliados e comparados, calcula-se a quantidade L_{pt} . O melhor modelo desenvolvido é aquele com maior valor de L_{pt} .

2.2.1 Amostragens - *Sampling*

No Stan, as amostragens são baseadas na simulação do HMC visando acelerar tanto a convergência para uma distribuição estacionária quanto a exploração do espaço paramétrico, utilizando o gradiente do logaritmo da função de probabilidade.

No processo de amostragem do HMC, o vetor paramétrico desconhecido θ é interpretado como a posição de uma partícula fictícia. Cada iteração gera um impulso aleatório e estimula o caminho da partícula com energia potencial determinando o logaritmo da função de probabilidade.

A decomposição de Hamilton mostra que o gradiente deste potencial determina a mudança no impulso e o impulso determina a mudança de posição. Estas alterações contínuas ao longo do tempo são aproximadas utilizando um algoritmo de salto, que quebra o tempo em passos discretos que são facilmente simulados.

A rejeição de Metropolis é então aplicada para corrigir qualquer erro de simulação e assegurar o equilíbrio detalhado das transições resultantes da cadeia de Markov (METROPOLIS *et al.*, 1953; HASTINGS, 1970).

O HMC padrão envolve três parâmetros de ajuste tornando seu comportamento mais sensível. Os amostradores do Stan permitem que esses parâmetros sejam ajustados manualmente ou se definam automaticamente sem a intervenção do usuário.

Os dois primeiros parâmetros de ajuste definem o tamanho do passo temporal da discretização do HMC e o número total de passos dados por iteração (o produto dos dois determina o tempo total de simulação). O Stan pode ser configurado com um tamanho de passo especificado pelo usuário ou pode estimar um tamanho ideal de passo durante a fase de aquecimento.

O terceiro parâmetro de ajuste é uma matriz de massa para a partícula fictícia. O Stan pode ser configurado para estimar uma matriz diagonal de massa ou uma matriz de massa completa durante o processo de aquecimento. Estimando uma matriz diagonal de massa, normaliza-se a escala de cada elemento θ_k da sequência variável desconhecida θ . Enquanto que estimando a matriz de massa completa, tanto a escala quanto a rotação são consideradas, porém exige-se uma maior memória e computação intensiva devido às operações de matrizes subjacentes.

As amostras em uma cadeia de Markov só são determinadas, com a distribuição marginal $p(\theta|y,x)$, depois que a cadeia convergir para a sua distribuição de equilíbrio.

Existem diversos métodos para testar se um MCMC falhou em sua convergência, mas infelizmente, eles não garantem convergência.

O método recomendado para o Stan, assim como em todas as demais maneiras de execução dos algoritmos MCMC, é executar várias cadeias de Markov cada uma com diferentes valores de parâmetros iniciais difusos, descartar as amostras do período de aquecimento ou adaptação, em seguida, dividir o restante de cada cadeia pela metade e calcular a redução de escala potencial estatística $\hat{R}$ (GELMAN *et al.*, 2013). Se $\hat{R}$ for alto, então deve-se prosseguir com outras simulações para melhorar a inferência

sobre a distribuição alvo, se $\hat{R} \rightarrow 1$ supõe-se que a cadeia convergiu para distribuição alvo.

Outro valor que também é retornado pelo Stan, e importante de ser observado para cada resultado das amostragens aplicadas, é n_{eff} que representa uma medida grosseira da dimensão eficaz das amostras, permitindo que se tenha noção da precisão obtida a partir das diversas simulações. Os cálculos das quantidades $\hat{R}$ e n_{eff} estão detalhados em Gelman *et al.* (2013).

2.2.2 Otimizações - *Optmizing*

As otimizações funcionam tão bem quanto as amostragens, além de serem mais rápidas. Iniciam-se com a leitura dos dados e, em seguida, inicializam-se os parâmetros.

Ao contrário da amostragem, a otimização produz uma saída determinística que não requer uma análise tão profunda para verificar que o otimizador converge para um valor alvo.

Dada uma função posterior $p(\theta|y)$, o Stan pode encontrar o valor alvo θ^* o qual é definido por:

$$\theta^* = \operatorname{argmax}_{\theta} p(\theta|y) \quad (5)$$

A notação $\operatorname{argmax}_u f(v)$ é usada para determinar o valor de v em que $f(v)$ é maximizada.

Se a distribuição priori utilizada for uniforme, o valor alvo corresponde à estimativa de máxima verossimilhança dos parâmetros. Se ela não for uniforme, o valor alvo é chamado de estimativa de máxima posterior. Se os parâmetros (tipicamente hierárquicos) forem marginalizados, o valor alvo é chamado de estimativa de probabilidade máxima marginal.

A estimativa θ^* é chamada de "estimativa pontual" e resume a distribuição a posteriori por um único ponto, em vez de utilizar uma distribuição. É claro que uma estima-

tiva pontual não consegue, por si só, levar em conta a variância estimada. Inferências preditivas posteriores $p(\tilde{y}|y)$ podem ser feitas usando o valor alvo dado os dados y como $p(\tilde{y}|\theta^*)$, mas eles não representam uma inferência bayesiana, mesmo que o modelo considere informações a priori ele não leva a incerteza posterior em consideração. Se a variação posterior é baixa e a média posterior está perto do valor alvo, inferências com estimativas pontuais podem ser muito semelhantes à uma inferência bayesiana completa. A verificação desta situação é necessária sempre que desejar utilizar a otimização como substituição da amostragem.

2.2.3 Logaritmo da probabilidade posterior total - L_{pt}

Segundo Stan Development Team (2014), a finalidade básica de um programa implementado na linguagem Stan é calcular a função logarítmica da probabilidade e seus derivados.

Um acumulador do logaritmo da probabilidade é inicializado em zero e depois incrementado de várias maneiras por um programa Stan. Primeiramente as variáveis são transformadas de não restritas para restritas e o determinante do logaritmo adicionado ao acumulador do logaritmo de probabilidade. Em seguida, o bloco principal do modelo é executado sobre os parâmetros restritos, com cada declaração de amostragem e com a declaração do incremento do logaritmo da probabilidade (*increment_log_prob*, ver os exemplos de modelos que utilizam esta função do Stan no Anexo A) adicionado ao acumulador. No final da execução do bloco de modelo, o valor do registro acumulado do logaritmo de probabilidade é o valor do logaritmo de probabilidade total (L_{pt}) retornado pelo programa Stan.

Avaliações de modelos construídos e aplicados em linguagem Stan, podem ser realizadas por meio da comparação entre os L_{pt} obtidos. Espera-se com a aplicação dos modelos que os valores de L_{pt} sejam maximizados.

Nas abordagens probabilísticas desenvolvidas neste trabalho, são utilizadas distribuições de probabilidade multivariadas, a fim de representar os diferentes horizontes

de previsão de vazão. Desta forma, um cálculo similar ao L_{pt} do Stan é feito para representar o logaritmo da probabilidade posterior total por horizonte de previsão (L_{pth}). A comparação dos valores de L_{pth} obtidos para os modelos também é utilizada como forma de avaliação das abordagens probabilísticas desenvolvidas.

2.3 TESTES ESTATÍSTICOS PARA COMPARAÇÃO DE MODELOS

A fim de avaliar as abordagens probabilísticas propostas neste trabalho diferentes testes estatísticos para análise dos resultados são aplicados. O intuito é construir modelos parcimoniosos, ou seja, que envolvam um número mínimo de parâmetros a serem estimados e que, ao mesmo tempo, expliquem comportamento da variável resposta de forma otimizada.

Diversos critérios para seleção de modelos são apresentados na literatura, dentre eles, os critérios baseados no máximo da função de verossimilhança (MFV) são os mais utilizados, com maior ênfase o Teste da Razão de Verossimilhança (TRV) (CASSELLA; BERGER, 2001), o Critério de Informação de Akaike (AIC) (BOZDONGAN, 1987) e o *Bayesian Information Criterion* (BIC) (SCHWARZ, 1978).

O teste da razão de verossimilhança é apropriado para comparar dois modelos, desde que um dos modelos seja um caso especial do outro (modelos aninhados), e é descrito por:

$$G = -2\ln(L_s) + 2\ln(L_c) \quad (6)$$

em que L_s é a máxima verossimilhança do modelo com menor número de parâmetros e L_c é a máxima verossimilhança do modelo com maior número de parâmetros. Testa-se a hipótese $H_0 : \beta_1 = 0$ contra a hipótese $H_1 : \beta_1 \neq 0$, e sob a hipótese nula, a estatística G tem distribuição chi-quadrado com graus de liberdade igual a diferença entre os números de parâmetros dos modelos.

O Critério de Informação de Akaike (AIC):

$$AIC = 2k - 2\ln(L) \quad (7)$$

em que k é o número de parâmetros do modelo e L sua máxima verossimilhança, estima a qualidade de cada modelo, em relação a cada um dos outros modelos, e não garante que os modelos comparados entre si sejam bons. O menor valor de AIC determina o melhor modelo na comparação.

Equivalente ao AIC, o *Bayesian Information Criterion* (BIC):

$$BIC = -2\ln(L) + k\ln(n) \quad (8)$$

considera n como tamanho da amostra. BIC é um critério para seleção de modelos entre um conjunto finito de modelos. O modelo com o menor BIC é preferido.

Para comparação entre os modelos com diferentes distribuições de probabilidade, ou seja, em que um dos modelos não é caso especial do outro (não aninhados), utiliza-se o teste de Vuong (VUONG, 1989). Na comparação entre dois modelos A e B , prefere-se o modelo A com nível de significância α , se a estatística Q_v dada por:

$$Q_v = \frac{2(Lpt_B - Lpt_A) - (p_B - p_A)\log(n)}{2\sqrt{\text{Var}(2(Lpt_{Bi} - Lpt_{Ai}))n}}, \quad i = 1, \dots, n \quad (9)$$

excede o quantil $(1 - \alpha)$ da distribuição normal padrão. Na Equação (9), Lpt_A e Lpt_B são os logaritmos das probabilidades posteriores dos modelos A e B respectivamente, p_B e p_A representam a quantidade de parâmetros de cada um dos modelos, com $p_B > p_A$, Lpt_{Ai} e Lpt_{Bi} são os logaritmos das probabilidades posteriores pontuais, e n tamanho da amostra.

Estes testes são utilizados para comparação dos resultados das diferentes abordagens criadas para os modelos probabilísticos aplicados aos erros dos modelos determinísticos.

3 MODELOS HIDROLÓGICOS DETERMINÍSTICOS

Modelos hidrológicos são ferramentas extremamente úteis que permitem, por meio do equacionamento dos processos, representar, entender e simular de forma aproximada o comportamento de bacias hidrográficas (TUCCI, 1998).

Segundo Breda (2008), gestores e planejadores de recursos hídricos, que utilizam estimativas de vazões futuras para diversos fins como a operação de reservatórios e prevenção das consequências de eventos extremos, são os principais empregadores de modelos hidrológicos dos mais diversos tipos, justificando assim sua grande utilidade.

Este mesmo autor ressalta que uma das maiores dificuldades no desenvolvimento destes modelos está na complexidade das características físicas envolvidas nos processos hidrológicos. Assim, faz-se necessário que simplificações sejam consideradas, implicando em uma generalização de parâmetros, ou de constantes, nas equações dos modelos.

Normalmente, as características dos processos variam de uma bacia hidrológica para outra devido aos diferentes fatores físicos envolvidos, sendo que ainda é possível que uma mesma bacia contenha regiões com características diferentes entre si. Deste modo, é preciso realizar um ajuste dos valores dos parâmetros do modelo hidrológico para cada bacia hidrográfica. Este ajuste pode ser efetuado com a escolha por tentativa e erro ou então por intermédio do emprego de métodos de calibração automática.

Muitas pesquisas, dentre elas, as que foram apresentadas no Capítulo 2, utilizam o MCMC com o intuito de estimar as incertezas em torno dos parâmetros de diferentes modelos hidrológicos determinísticos. Porém neste trabalho a proposta é utilizar

os resultados de modelos determinísticos que já estão sendo extensivamente utilizados, ou seja, que possuem parâmetros devidamente calibrados, para derivar os erros resultantes de suas previsões, e utilizá-los como entrada para os modelos de previsão probabilísticos.

Desta forma, para a aplicação dos modelos probabilísticos de previsão de vazão, desenvolvidos no decorrer deste trabalho, os resultados de quatro modelos determinísticos do tipo chuva-vazão são utilizados. Estes modelos determinísticos estão subdivididos entre modelos conceituais e modelos empíricos.

Os modelos conceituais são baseados na representação dos processos físicos envolvidos no fenômeno estudado. Estes modelos podem relacionar as características físicas do processo mantendo razoável empirismo na definição de seus parâmetros ou podem se aproximar muito da física do sistema.

Os modelos empíricos são, basicamente, definidos por equações matemáticas. A principal característica deste tipo de modelo consiste em estabelecer uma relação estável entre variáveis de entrada e saída, sem considerar neste processo quais leis físicas que governam o processo natural de transformação da chuva em vazão.

3.1 MODELOS CONCEITUAIS CHUVA-VAZÃO

Segundo Breda (2008), modelos conceituais chuva-vazão têm sido desenvolvidos para representar, de maneira realista, os sub-processos e relações envolvidas no processo chuva-vazão. Embora esses modelos variem significativamente quanto ao grau de complexidade, precisão e quantidade de parâmetros utilizados, eles possuem características essencialmente comuns.

3.1.1 Modelo Sacramento

O modelo *Sacramento Soil Moisture Accounting* (BURNASH, 1995) é um dos modelos conceituais chuva-vazão mais utilizados no mundo para a simulação de vazão

em curta escala de tempo (CUNHA, 2004). Sua base teórica envolve vários processos hidrológicos, além de apresentar certa hierarquia destes processos, e disto decorre o grande número de parâmetros necessários para ele (BREDA, 2008).

A Figura 6 apresenta um esquema resumido do modelo Sacramento que possui 16 parâmetros ao todo descritos na Tabela 1. Mais informações sobre o funcionamento deste modelo podem ser encontradas em Burnash (1995), Cunha (2004) e Breda (2008).

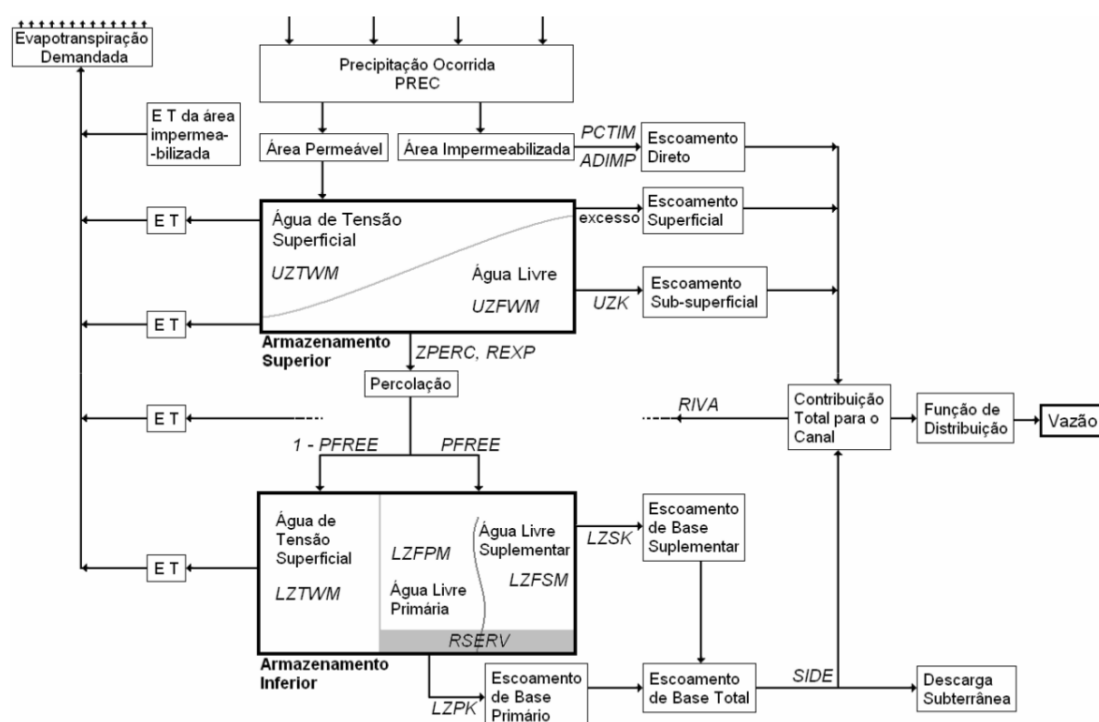


FIGURA 6: Apresentação esquemática do modelo Sacramento

FONTE: Breda (2008)

3.1.2 Modelo 3R

Segundo Neto, Guetter e Mine (2007), o modelo 3R, cujo nome tem origem nas iniciais das palavras *rainfall-runoff-routing* (chuva- vazão-propagação), é uma adaptação, para a aplicação em grandes bacias, do modelo Sacramento reduzindo a complexidade do modelo mas também diminuindo a capacidade de representar os detalhes dos processos de infiltração e percolação em pequenas bacias.

QUADRO 1: Descrição dos parâmetros da fase bacia do modelo Sacramento

FONTE: Breda (2008)

Parâmetro	Descrição	Unidade
UZW	Capacidade máxima do armazenamento superior de água de tensão superficial	mm
UZF	Capacidade máxima do armazenamento superior de água livre	mm
LZW	Capacidade máxima do armazenamento inferior de água de tensão superficial	mm
LZF	Capacidade máxima do armazenamento inferior primário de água livre	mm
LZFS	Capacidade máxima do armazenamento inferior suplementar de água livre	mm
UZK	Taxa de depleção lateral da zona superior de água livre	dia^{-1}
LZPK	Expoente da equação da percolação	dia^{-1}
LZSK	Expoente da equação da evapotranspiração ocorrida no armazenamento inferior	dia^{-1}
PCTIM	Fração de área constantemente impermeável na bacia	fração
ADIMP	Área impermeável adicional	fração
PFREE	Fração da água percolada da zona superior que vai direto para a zona de água livre inferior	fração
ZPERC	Taxa máxima de percolação	-
REXP	Expoente da equação de percolação	-
RSERV	Fração da água livre da zona inferior que não é transferível para a zona de água de tensão superficial	fração
RIVA	Fração de área coberta por mata ciliar na bacia	fração
SIDE	Fração do escoamento de base que compõe a perda de base	-

Assim, o modelo 3R é indicado para uso em grandes bacias e para a resolução temporal de algumas horas, ou preferencialmente, para a resolução diária. A estrutura do modelo 3R está ilustrada na Figura 7, sendo a sua formulação e seus parâmetros detalhados em Neto (2005).

3.2 MODELOS EMPÍRICOS

Os modelos empíricos utilizam relações baseadas em observações. Em geral, estes modelos são bastante simples e úteis no dia a dia. No entanto, são modelos pouco robustos uma vez que são específicos para aquela região para a qual as relações foram estimadas. Além disso, os modelos empíricos não possibilitam fazer simulações de mudanças em condições para as quais não previu, tais como, chuvas em grande

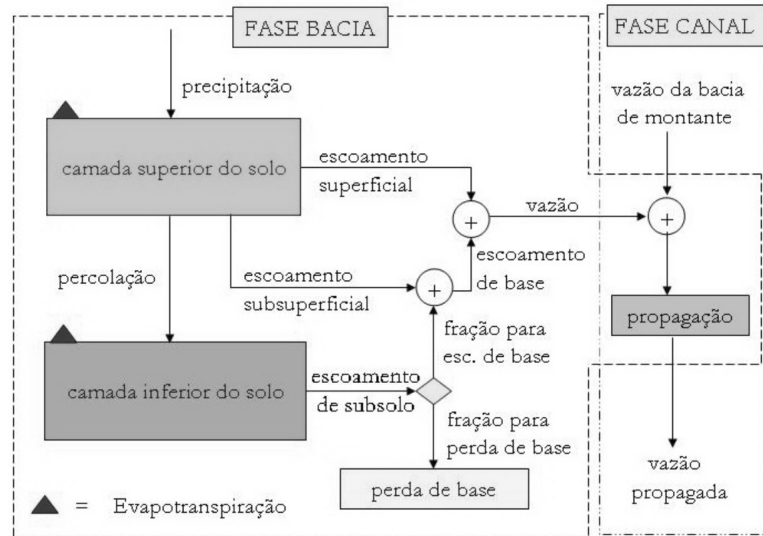


FIGURA 7: Estrutura do modelo 3R

FONTE: Neto, Guetter e Mine (2007)

quantidade, mudança de uso do solo, etc. (RENNO; SOARES, 2000).

3.2.1 Modelo Regressivo Linear Múltiplo

Segundo Davison (2003), modelos regressivos são utilizados para descrever como uma ou mais variáveis de resposta dependem de outras variáveis explicativas. Muitas vezes faz-se necessário prever ou controlar as respostas futuras alterando outras variáveis, ou compreendendo a relação entre elas. Se y representar a variável resposta e x as variáveis explicativas, a pergunta que se deseja responder é “o que acontece com y quando x varia?” ou ainda, “como as mudanças de x afetam y ?”. Para responder essas perguntas, constroem-se modelos de regressão linear simples, quando x representa apenas uma variável, ou múltiplos, quando x representa várias variáveis explicativas de um determinado processo:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = x_i^T \beta + \varepsilon_i, \quad i = 1, \dots, n. \quad (10)$$

Ou seja, dado um conjunto de dados $\{y_i, x_{i1}, \dots, x_{ip}\}_{i=1}^n$ de n unidades estatísticas, um modelo de regressão linear múltiplo supõe que a relação entre a variável dependente y_i e um p -vetor de regressores x_i é linear. Essa relação é modelada por meio de

um termo de perturbação ou variável de erro ε_i (uma variável aleatória que acrescenta ruído entre a relação linear entre a variável dependente e os regressores). Além disso, na Equação (10), o expoente T representa a operação transposta, de modo que $x_i^T \beta$ representa o produto interno entre os vetores x_i e β .

A solução de sistemas deste tipo (EQUAÇÃO (10)) pode ser determinada pelo Método dos Quadrados Mínimos que possui como objetivo minimizar a função:

$$L = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2, \quad (11)$$

derivando L em função dos β'_s , obtêm-se:

$$\begin{aligned} \frac{\partial L}{\partial \beta_0} &= -2 \sum_{i=1}^n [y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}], \\ &\vdots \\ \frac{\partial L}{\partial \beta_j} &= -2 \sum_{i=1}^n [y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}] x_{ji}, \quad j = 1, 2, \dots, p, \end{aligned} \quad (12)$$

igualando as derivadas parciais a zero e substituindo $\beta_0, \beta_1, \dots, \beta_p$ por $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$, tem-se o sistema:

$$\begin{cases} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{i1} + \cdots + \hat{\beta}_p \sum_{i=1}^n x_{ip} = \sum_{i=1}^n y_i \\ \hat{\beta}_0 \sum_{i=1}^n x_{i1} + \hat{\beta}_1 \sum_{i=1}^n x_{i1}^2 + \cdots + \hat{\beta}_p \sum_{i=1}^n x_{i1} x_{ip} = \sum_{i=1}^n x_{i1} y_i \\ \vdots \\ \hat{\beta}_0 \sum_{i=1}^n x_{ip} + \hat{\beta}_1 \sum_{i=1}^n x_{ip} x_{i1} + \cdots + \hat{\beta}_p \sum_{i=1}^n x_{ip}^2 = \sum_{i=1}^n x_{ip} y_i \end{cases} \quad (13)$$

Resolvendo este sistema (EQUAÇÃO (13)), obtêm-se os estimadores de mínimos quadrados $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ dos parâmetros do modelos em questão.

O modelo de Regressão Linear Múltiplo também pode ser escrito na forma matricial:

$$Y = X\beta + \varepsilon, \quad (14)$$

em que Y é um vetor $n \times 1$ cujos componentes correspondem às n respostas; X é uma matriz de dimensão $n \times (p+1)$ denominada matriz do modelo; ε é um vetor de dimensão $n \times 1$ cujos componentes são os erros; e β é um vetor $(p+1) \times 1$ cujos elementos são os coeficientes de regressão.

O método dos Quadrados Mínimos tem como objetivo encontrar o vetor $\hat{\beta}$ que minimiza:

$$\begin{aligned} L &= \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^T \varepsilon = (Y - X\beta)^T (Y - X\beta) = \\ &= Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta, \end{aligned} \quad (15)$$

sendo que $Y^T X\beta = \beta^T X^T Y$ pois o produto resulta em um escalar. Derivando em termos matriciais:

$$\frac{\partial L}{\partial \beta} = -2X^T Y + 2X^T X\beta. \quad (16)$$

Igualando a zero e substituindo o vetor β pelo vetor $\hat{\beta}$, tem-se:

$$(X^T X)\hat{\beta} = X^T Y. \quad (17)$$

Portanto, o modelo de regressão linear ajustado e o vetor de resíduos se tornam, respectivamente:

$$\hat{Y} = X\hat{\beta} \quad \text{e} \quad \varepsilon = Y - \hat{Y}. \quad (18)$$

Para a construção do modelo determinístico regressivo linear de previsão de vazão, degraus de valores de vazão são utilizados como regressores.

Os valores de vazão observados no momento em que a previsão será realizada são representados por q_0 ; um degrau previsto, por exemplo, em um horizonte adiante do momento da previsão é representado por $(q_1 - q_0)$, e degraus de momentos anteriores à previsão são representados por $(q_{-1} - q_{-2}), (q_{-2} - q_{-3}), \dots$.

Previsões de vazão com horizonte de até 72h são realizadas, considerando os valores de vazão de 6h em 6h, totalizando 12 horizontes de previsão. Assim, o resultado de um modelo regressivo de 6h será utilizado como entrada para o modelo regressivo de 12h, o de 12h para o de 18h e assim por diante, até que o resultado de 66h seja utilizado como entrada do modelo regressivo de 72h.

As componentes dos modelos regressivos lineares de 6h para a previsão de degraus de vazão foram definidas como:

- A variável resposta, ou variável dependente, y_i , que representa um degrau de vazão a ser previsto para o momento de previsão do tipo $(q_1 - q_0)$;
- As variáveis $x_{i1}, x_{i2}, \dots, x_{ip}$ são os regressores, as variáveis preditoras independentes, determinadas por meio das soluções do sistema de equações lineares a partir do método de quadrados mínimos implementado na função *lm* do *software* R (R Core Team, 2012). Os regressores são representados na forma de matrizes formadas por 12 degraus de valores de vazão passados $((q_{-11} - q_{-12}), \dots, (q_{-1} - q_{-2}), (q_0 - q_{-1}))$, com 12 valores de precipitação passados $(p_{-11}, \dots, p_{-1}, p_0)$, e 1 valor de precipitação previsto p_1 referente ao próprio momento da previsão de vazão, todos acumulados de 6h, para um período de treinamento (descrito na seção 3.3 deste capítulo);
- β um vetor paramétrico p-dimensional cujos elementos são chamados de coeficientes de regressão. Os coeficientes são determinados por meio da solução do sistema de equações lineares por meio da aplicação do método de quadrados mínimos, mas desta vez para um período de validação (descrito na seção 3.3).

Os resultados do modelo determinístico Regressivo Linear Múltiplo são obtidos na forma de degraus de vazão. Para a utilização destes resultados como entrada dos modelos probabilísticos de previsão de vazão, ou seja, para a obtenção dos erros cometidos por este método, transformam-se estes resultados em valores de vazão, somando o valor de vazão observada à diferença resultante.

3.2.2 Modelo Rede Neural - MLP

MLP (*Multilayer perceptron* (HAYKIN, 2001)) é uma rede neural artificial que mapeia conjuntos de dados de entrada para um conjunto de saídas apropriadas. Essa rede é composta por várias camadas de nós em um grafo direcionado, normalmente com cada camada completamente conectada à próxima. Exceto para os nós da camada de entrada, os demais nós podem ser chamados de neurônios ou elemento de processamento, admitindo funções de ativação não lineares.

A MLP é uma rede do tipo *feedforward* em que as conexões entre os nós que compõe a rede não formam um ciclo, ou seja, a informação é transmitida somente em uma direção, para frente, a partir dos nós de entrada passando pelos neurônios das camadas intermediárias até chegarem à camada de saída.

Esta rede utiliza uma técnica de aprendizagem supervisionada chamada de retropropagação para o seu treinamento.

A Figura 8 apresenta um exemplo da arquitetura de uma MLP, com uma camada de entrada formada por um nó, uma camada intermediária formada por vários neurônios, e uma camada de saída formada por um neurônio.

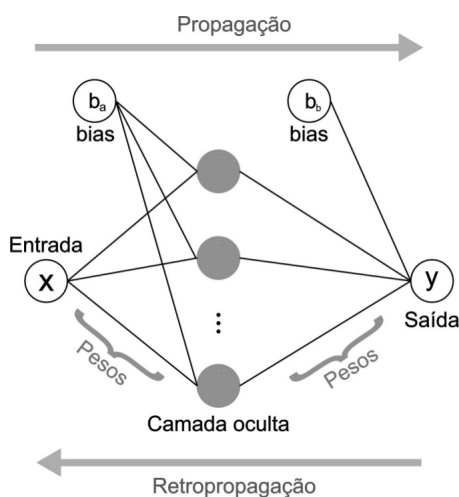


FIGURA 8: Arquitetura de uma rede neural MLP

FONTE: Santos (2014)

Algoritmo resumido da MLP:

1. inicialize os pesos das conexões (v e w) e dos bias (b_a , b_b) com valores aleatórios; inicialize a taxa de aprendizagem (α). Para cada padrão de entrada (x), execute os passos de 2 a 4:
2. calcule as entradas na camada escondida $z = f(z^*) = f(v^T x + b_a)$, e a saída da rede $y = f(y^*) = f(w^T z + b_b)$, sendo f uma função de ativação;
3. calcule as correções das conexões da camada de saída, erro do vetor y , $w = w + \Delta w$ e $b_b = b_b + \Delta b_b$;

4. calcule as correções das conexões da camada escondida, erro do vetor z , $v = v + \Delta v$ e $b_a = b_a + \Delta b_a$;
5. atualize a taxa de aprendizagem (α), verifique os erros para todos os padrões de entrada, e teste o critério de parada determinado.

Informações detalhadas sobre esse tipo de rede podem ser encontradas em Haykin (2001) e Santos (2014).

A construção do modelo determinístico MLP de previsão de vazão segue a mesma ideia do modelo regressivo linear utilizando valores acumulados de 6h de precipitação em conjunto com os degraus de vazão.

A rede construída contou com 25 nós na camada de entrada correspondentes a uma composição formada por 12 valores de degraus de vazão passados $((q_{-11} - q_{-12}), \dots, (q_{-1} - q_{-2}), (q_0 - q_{-1}))$, com 12 valores de precipitação passados e acumulados de 6h $(p_{-11}, \dots, p_{-1}, p_0)$, e 1 valor de precipitação previsto p_1 referente ao próprio momento da previsão de vazão, acumulado de 6h.

Sua camada intermediária é composta de 25 neurônios ativados pela função tangente hiperbólica, porém com o valor do neurônio dividido por 2. A camada de saída é composta por 1 neurônio que representa um degrau de vazão a ser previsto para o momento de previsão do tipo $(q_1 - q_0)$ ativado pela função de ativação logística.

Os pesos das conexões entre as camadas são inicializados aleatoriamente com valores no intervalo de $[1, -1]$.

A função de aprendizagem utilizada na rede é a do tipo *Backpropagation* considerando um termo de momento (HAYKIN, 2001).

O treinamento da rede neural é dividido em duas etapas, chamadas aqui de calibração e verificação, realizadas com 70% e 30%, respectivamente, dos dados que correspondem a um período chamado de treinamento descrito na seção 3.3 deste capítulo. A validação desta rede é aplicada aos dados correspondentes a um segundo período específico também descrito na próxima seção.

Definida a arquitetura da rede, e as funções a serem aplicadas, utilizou-se o SNNS (*Stuttgart Neural Network Simulator* (ZELL *et al.*, 1990)), um simulador para redes neurais, e o pacote RSNNS do *software* R (R Core Team, 2012) para avaliação dos resultados das previsões de degraus de vazão.

Os resultados obtidos são, novamente, transformados em valores de vazão somando o valor de vazão observada à diferença resultante, para a obtenção dos erros cometidos por este método que serão utilizados como entrada dos modelos probabilísticos descritos no próximo capítulo.

3.3 DADOS E REGIÃO DE ESTUDO

Para a aplicação dos modelos determinísticos, e posteriormente dos probabilísticos, são utilizados dados de precipitação e vazão de uma estação hidrometeorológica. Esta estação é monitorada pelo SIMEPAR e pela COPEL, possui coordenadas geográficas de latitude igual a -25,9333, longitude -49,7833, altura 799m e corresponde à exutória da sub-bacia de São Bento pertencente à bacia do rio Iguaçu, na cidade da Lapa, no estado do Paraná (FIGURA 9), com área de drenagem de 2012km^2 e vazão média de $34,9\text{m}^3/\text{s}$.

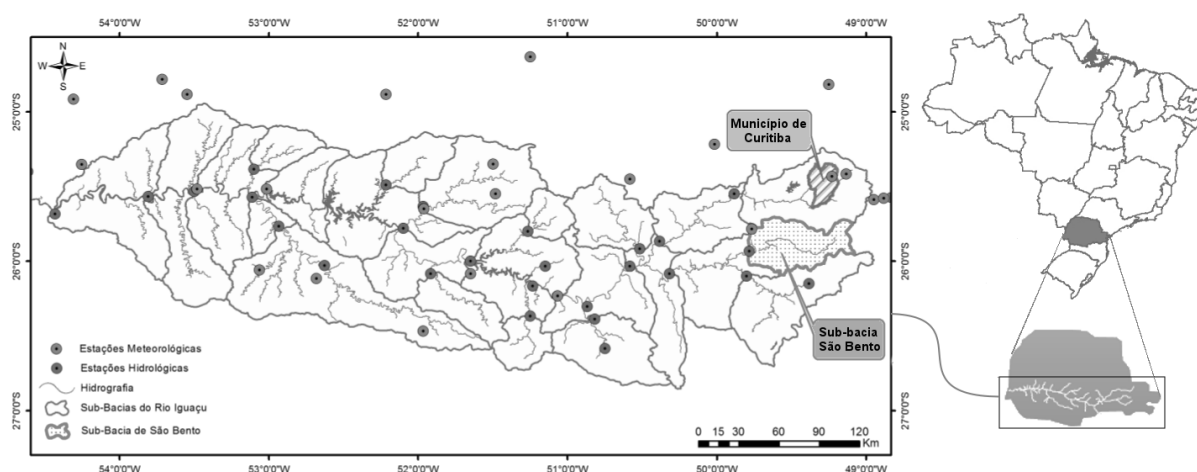


FIGURA 9: Sub-bacia de São Bento pertencente à bacia do rio Iguaçu

FONTE: A autora (2017)

Os modelos determinísticos conceituais, Sacramento e 3R, calibrados previamente

pelo SIMEPAR com dados de 05/01/1998 até 25/12/2010, foram aplicados e realizaram previsões de 6h em 6h correspondentes às datas de 24/05/2013 até 25/06/2016 (FIGURA 10), totalizando 4515 casos. Este período foi escolhido devido à disponibilidade de dados de precipitação previstos que foram utilizados em conjunto com os dados de precipitação e também de vazão observados.

Os modelos empíricos, Regressivo Linear e MLP, utilizam dados que são subdivididos em duas etapas: o treinamento com dados de vazão e precipitação observados, relativos ao período de 05/01/1998 até 25/12/2010 (totalizando 113707 casos horários); e a verificação, com dados de vazão e precipitação observados em conjunto com dados de precipitação previstos, correspondente ao período de 24/05/2013 até 25/06/2016 (totalizando 4515 casos de 6h em 6h, Figura 10).

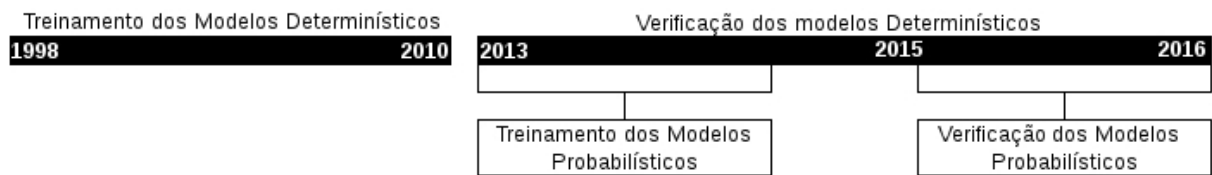


FIGURA 10: Períodos de treinamento e verificação dos modelos determinísticos e probabilísticos

FONTE: A autora (2017)

Como entrada para os modelos probabilísticos, descritos no capítulo a seguir, utilizou-se o erro cometido pelos modelos determinísticos no período de verificação. Este erro (ϵ) é um vetor que contém a diferença entre valor observado de vazão (q_{obs}) e valor previsto pelo modelo determinístico (q_{prev}):

$$\epsilon = q_{obs} - q_{prev} \quad (19)$$

A utilização, pelos modelos determinísticos, de dados de precipitação previstos, permite que os modelos probabilísticos tornem-se capazes de refletir as incertezas hidrometeorológicas em sua totalidade.

Assim como nos modelos determinísticos empíricos, para os modelos probabilís-

ticos, os dados de erros resultantes dos modelos determinísticos também foram subdivididos em períodos de treinamento (24/05/2013 até 02/06/2015, totalizando 2959 casos de 6h em 6h) e verificação (03/06/2015 até 25/06/2016, totalizando 1556 casos de 6h em 6h), como pode ser observado na Figura 10.

3.4 COMPARAÇÃO ENTRE OS MODELOS DETERMINÍSTICOS

A avaliação do desempenho dos modelos determinísticos pode ser realizada de duas maneiras com objetivos distintos.

A primeira maneira de avaliação utiliza a combinação entre dados de vazão observados, dados de precipitação observados passados e dados de precipitação observados futuros (de 6h em 6h). Esta maneira tem o intuito de avaliar os resultados dos modelos hidrológicos que realizam a previsão de dados de vazão considerando dados de precipitação observados para o futuro, permitindo uma comparação entre previsão e observação.

A segunda maneira de avaliação utiliza a combinação de dados de vazão observados, dados de precipitação observados passados e dados de precipitação previstos por modelos meteorológicos conhecidos (de 6h em 6h). Esta maneira possui o intuito de avaliar o processo de previsão, em que os modelos hidrológicos realizam a previsão de dados de vazão considerando dados de precipitação previstos para o futuro. Neste processo todas as incertezas inerentes à realização da previsão, inclusive a incerteza sobre os valores de precipitação, estão inclusas.

A Figura 11 apresenta a comparação do desempenho dos quatro modelos determinísticos, que podem utilizar precipitação observada ou prevista em seu desenvolvimento. As curvas são formadas pelos valores da raiz do erro quadrático médio:

$$REQM = \sqrt{\frac{1}{n} \sum_{i=1}^n (q_{obs_i} - q_{prev_i})^2} \quad (20)$$

para os 12 horizontes de previsão, do período de verificação dos modelos determinísticos. Na Equação (20) q_{obs_i} são os valores de vazão observados no momento i

de previsão, q_{prev_i} são os valores de vazão previstos pelo modelo determinístico no momento i , e n a quantidade de dados considerada.

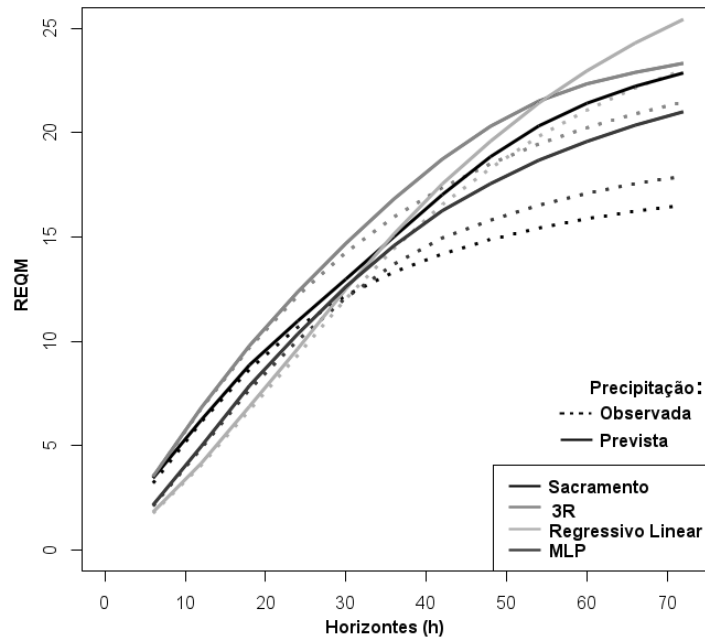


FIGURA 11: Valores de REQM para previsões dos quatro modelos determinísticos que podem utilizar precipitação observada ou prevista

FONTE: A autora(2017)

Na Figura 11, as curvas de desempenho demonstram que as incertezas quanto à previsão de precipitação, bem como o passar dos horizontes, contribuem para o aumento do erro de previsão de vazão. Com isso, neste trabalho, a análise dos erros dos modelos determinísticos será sempre realizada sobre as previsões que utilizam precipitação prevista em seu desenvolvimento. Desta forma, toda a incerteza hidrometeorológica envolvida no processo de previsão determinística será considerada.

O coeficiente de eficiência de *Nash-Sutcliffe* (NASH; SUTCLIFFE, 1970) também é muito utilizado para avaliar o poder preditivo de modelos hidrológicos. Definido por:

$$NS = 1 - \frac{\sum_{i=1}^n (q_{obs_i} - q_{prev_i})^2}{\sum_{i=1}^n (q_{obs_i} - \overline{q_{obs}})^2} \quad (21)$$

em que $\overline{q_{obs}}$ é a média dos valores de vazão observados, a eficiência pode variar de $-\infty$ a 1. Um coeficiente igual a 1 corresponde a um modelo de previsão perfeito, igual a zero indica que as previsões do modelo são tão precisas quanto a média dos dados

observados, e menores do que zero que a média observada é melhor preditor do que o modelo.

A Figura 12 apresenta a comparação dos valores dos coeficientes de *Nash-Sutcliffe* para os quatro modelos determinísticos, utilizando precipitação prevista, no período de verificação para os 12 horizontes de previsão.

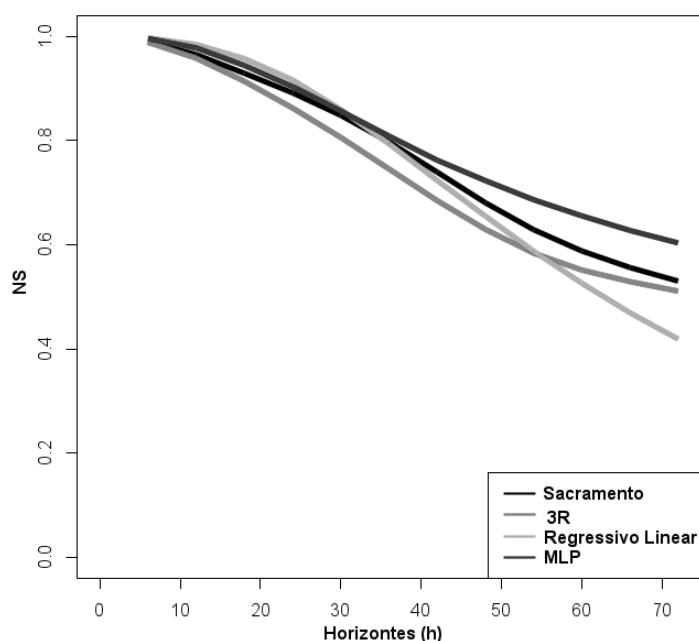


FIGURA 12: Valores do coeficiente de *Nash-Sutcliffe* para os quatro modelos determinísticos utilizando precipitação prevista

FONTE: A autora(2017)

Em geral, as Figuras 11 e 12 evidenciam as diferenças entre os resultados dos modelos determinísticos para diferentes horizontes de previsão, indicando a necessidade de um método robusto que seja capaz de considerar essas variações. Assim, o objetivo da construção dos modelos probabilísticos de erros, apresentados no capítulo a seguir, é justamente considerar as diferenças existentes entre os resultados dos diferentes modelos determinísticos a fim de construir uma previsão probabilística que suporte todas essas incertezas.

Neste sentido, um estudo aprofundado da distribuição dos erros dos modelos determinísticos torna-se relevante.

A Figura 13 apresenta as estimativas da distribuição de probabilidade (estimadas

por meio do método *gaussian kernel density estimation* (PARZEN, 1962)) dos erros do modelo determinístico Regressivo Linear para os diferentes horizontes de previsão.

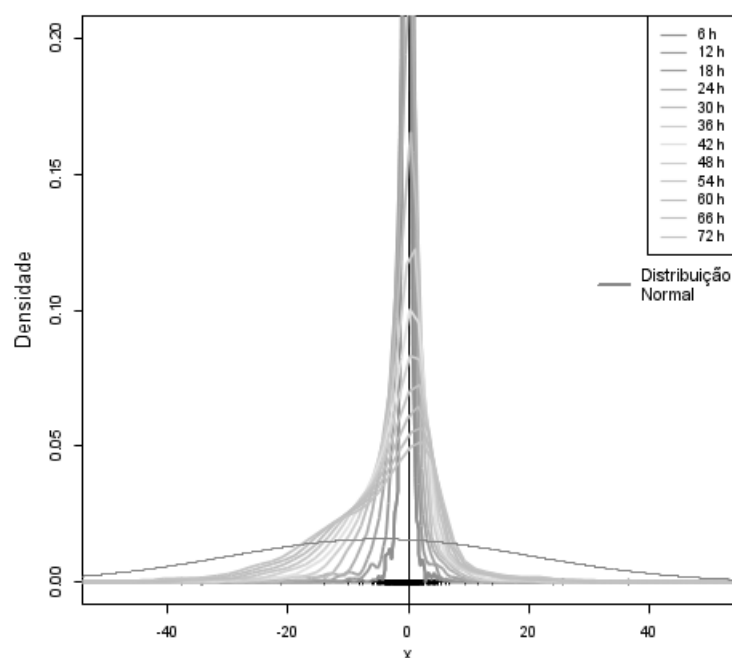


FIGURA 13: Distribuições de probabilidade dos erros por horizonte de previsão do modelo determinístico Regressivo Linear

FONTE: A autora (2017)

Na Figura 14, é apresentada uma estimativa da densidade dos erros do modelo determinístico MLP, para o horizonte de 72h.

Os mesmos estudos foram realizados sobre os erros dos modelos determinísticos conceituais Sacramento e 3R. Em todos os casos, nota-se a presença de deslocamentos, assimetrias, e distanciamento da distribuição normal. Pode-se observar que a distribuição dos erros positivos difere-se da dos erros negativos.

As Figuras 15 (a) e 15 (b) apresentam as correlações, auto-correlações e correlações cruzadas entre os erros do modelo determinístico Sacramento para os diferentes horizontes de previsão.

A Figura 15 (a) apresenta a correlação entre os erros do modelo Sacramento de acordo com cada um dos horizontes de previsão. Os tons de vermelho mais escuros representam altos valores de correlação (por exemplo, quando um horizonte está sendo comparado com ele mesmo), estes tons tornam-se mais claros conforme a cor-

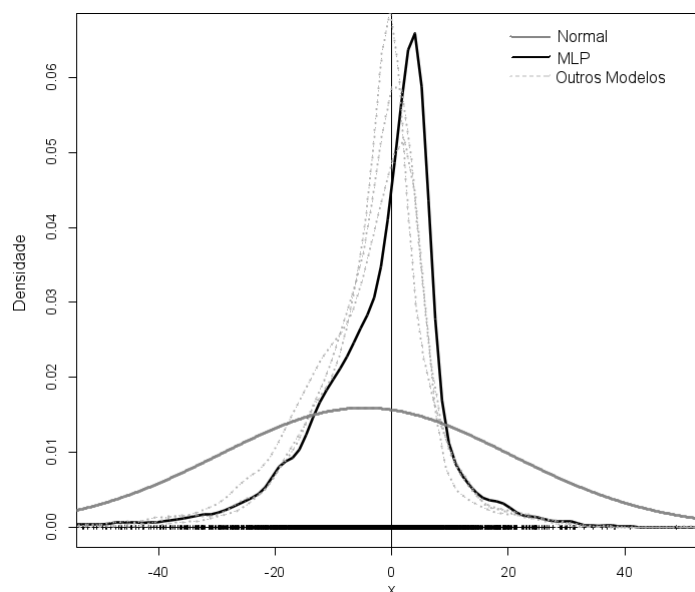


FIGURA 14: Distribuição de probabilidade dos erros para o horizonte de 72h do modelo determinístico MLP

FONTE: A autora (2017)

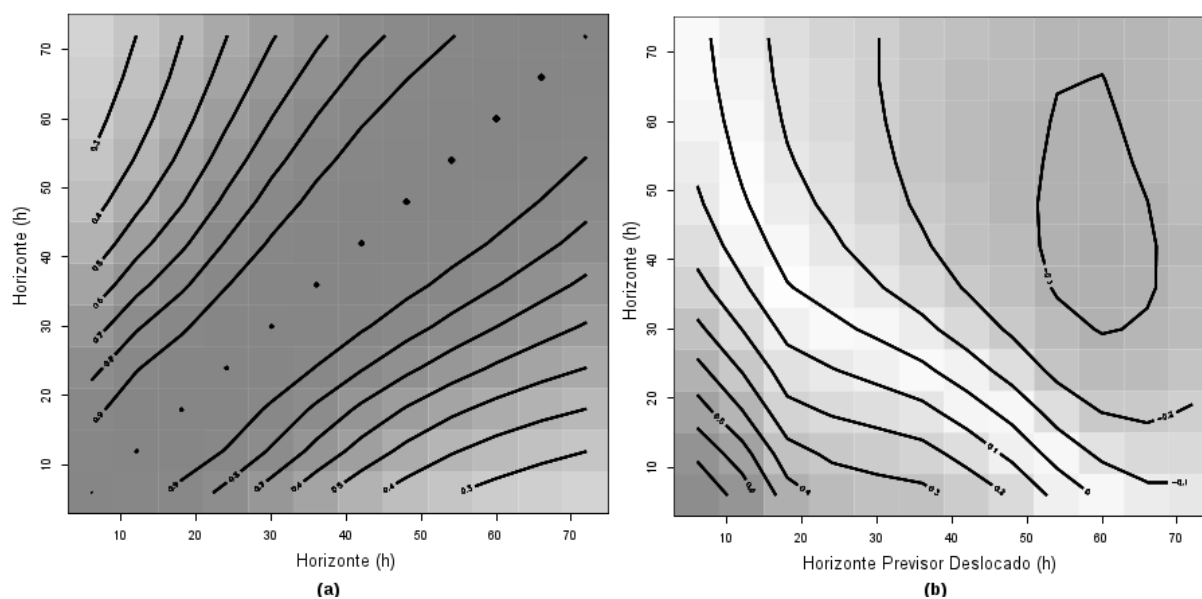


FIGURA 15: Correlações do modelo determinístico Sacramento. (a) Correlação entre os erros de acordo com o horizonte de previsão. (b) Correlação cruzada entre erros com deslocamentos temporais

FONTE: A autora (2017)

relação diminui (por exemplo, quando um horizonte menor se correlaciona com um horizonte maior).

A Figura 15 (b) apresenta a correlação cruzada entre os erros deslocados tempo-

ralmente. Os tons em vermelho escuro representam correlações altas positivas que quando diminuem e se aproximam de zero tornam os tons mais claros, transformando-se em tons de azul mais escuros conforme a correlação torna-se mais negativa.

Ambas as representações das Figuras 15 (a) e 15 (b) demonstram forte estrutura no comportamento dos erros entre horizontes de uma mesma previsão (0,22 a 0,99), bem como a existência de correlação entre os erros de diferentes previsões (-0,32 a 0,78).

Os erros obtidos pela aplicação dos quatro modelos determinísticos para horizontes de 6h são altamente correlacionados. Além disso, quanto menor o horizonte melhor o resultado do modelo, e quanto maior o horizonte, mais complexa se torna a realização da previsão.

Devido ao comportamento assimétrico das densidades (FIGURA 13 e FIGURA 14), e das correlações existentes entre os erros nos diferentes horizontes de previsão dos modelos determinísticos (FIGURA 15 (a) e FIGURA 15 (b)), a calibração do modelo probabilístico de previsão de vazão também necessita de um estudo prévio sobre a distribuição de probabilidade ideal para ser utilizada.

4 MODELOS HIDROLÓGICOS PROBABILÍSTICOS

A construção dos modelos probabilísticos é estruturada de acordo com a variação de seus parâmetros. Dessa forma, dois tipos de modelos probabilísticos são desenvolvidos.

No modelo probabilístico do tipo estático, os parâmetros são considerados fixos para qualquer momento de previsão, ou seja, não variam ao longo do tempo.

Para o modelo probabilístico do tipo dinâmico, um conjunto de preditores faz com que os parâmetros se tornem variáveis ao longo do tempo.

O modelo estático é utilizado como informação à priori no desenvolvimento do modelo dinâmico, dessa forma o modelo dinâmico é considerado uma melhoria do estático.

Os modelos probabilísticos, estáticos ou dinâmicos, são estruturados de forma que 12 horizontes de previsões de 6h em 6h (totalizando 72h) sejam correlacionados por meio de uma análise estatística multivariada.

4.1 MODELOS PROBABILÍSTICOS ESTÁTICOS MULTIVARIADOS

Parte-se da ideia inicial de que os erros obtidos pela aplicação dos quatro modelos determinísticos podem ser representados por uma distribuição normal multivariada (JOHNSON; WICHERN, 2007), ou seja, que a posteriori do modelo probabilístico seja deste tipo.

Porém, com a observação das densidades apresentadas pelas Figuras 13 e 14, acredita-se que uma distribuição de probabilidade multivariada assimétrica seria melhor ajustada aos erros dos modelos determinísticos. Com isso testa-se, além da distribuição normal, a distribuição de probabilidade multivariada *skew-t* (ARELLANO-

VALLE; GENTON, 2010).

Na seção 4.1.2 são apresentadas comparações entre as utilizações das distribuições normal multivariada e *skew-t* multivariada nos modelos probabilísticos do tipo estático, bem como as avaliações dessas aplicações por meio dos testes estatísticos AIC, BIC e Vuong.

4.1.1 Ajuste de uma Distribuição Normal Multivariada

A densidade normal multivariada é uma generalização da densidade normal univariada para $p \geq 2$ dimensões. A distribuição normal univariada, com média μ e variância σ^2 possui a seguinte função densidade de probabilidade:

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma \in \mathbb{R} \quad (22)$$

A função densidade de probabilidade conjunta de p variáveis independentes normais $x_1, x_2, \dots, x_p$ é:

$$f(x_1, \dots, x_p) = \frac{1}{(2\pi)^{\frac{p}{2}} \sqrt{\sigma_1^2 \sigma_2^2 \dots \sigma_p^2}} e^{-\frac{1}{2} \sum_{i=1}^p \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2} \quad (23)$$

Fazendo $\underline{x}^T = [x_1, x_2, \dots, x_p]$, $\underline{\mu}^T = [\mu_1, \mu_2, \dots, \mu_p]$ e $\Sigma = \begin{pmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_p^2 \end{pmatrix}$, pode-se escrever a densidade conjunta como:

$$f(\underline{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2} (\underline{x} - \underline{\mu}) \Sigma^{-1} (\underline{x} - \underline{\mu})} \quad (24)$$

Desta maneira, o modelo probabilístico estático multivariado que possui como posteriori a distribuição normal multivariada $NM(\underline{\mu}, \Sigma)$, também é composto de dois parâmetros chamados de localização e escala.

A localização corresponde à média $\underline{\mu}$ (vetor de 12 dimensões) da distribuição nor-

mal multivariada e a escala corresponde à variância Σ (matriz de covariância de dimensão 12×12) da distribuição normal multivariada ajustada aos erros de previsões dos modelos determinísticos.

Uma transformação *logit* (COLLINS *et al.*, 1992) é aplicada aos valores de vazão observados q_{obs} normalizados e de vazão previstos q_{prev} pelos quatro modelos determinísticos e normalizados da forma:

$$\begin{aligned} lq_{prev} &= \text{logit} \left(\frac{q_{prev} - q_{min}}{q_{max} - q_{min}} \right) = \log \left(\frac{q_{prev} - q_{min}}{q_{max} - q_{prev}} \right) \\ lq_{obs} &= \text{logit} \left(\frac{q_{obs} - q_{min}}{q_{max} - q_{min}} \right) = \log \left(\frac{q_{obs} - q_{min}}{q_{max} - q_{obs}} \right) \end{aligned} \quad (25)$$

A aplicação da transformação *logit* visa limitar os valores de vazão previstos aos q_{max} (valores máximos de vazão) e q_{min} (valores mínimos de vazão) já observados em toda a série histórica de dados da sub-bacia estudada, evitando assim que valores infactíveis (negativos ou extremamente altos) de vazão sejam previstos pelos modelos probabilísticos.

O vetor de erros de cada um dos modelos determinísticos torna-se então a variável aleatória $\underline{x}$ para a qual será aplicado o modelo probabilístico com distribuição de probabilidade normal multivariada:

$$\underline{x} = lq_{obs} - lq_{prev} \quad (26)$$

Otimizações do Stan (descritas na seção 2.2.2), de quatro cadeias de Markov em paralelo, são aplicadas para os 12 horizontes de previsão simultaneamente com o intuito de determinar representantes da distribuição posteriori dos parâmetros do modelo probabilístico estático.

O código, desenvolvido em linguagem Stan, para aplicação do MCMC para o modelo estático multivariado com distribuição de probabilidade posteriori normal multivariada, é apresentado na Figura 37 do Anexo A.

O método de amostragem ou *Sampling* (descrito na seção 2.2.1) proposto pelo

Stan, também é aplicado com a mesma implementação, apresentando resultados muito semelhantes às otimizações, devido à grande quantidade de dados de vazão disponíveis para a aplicação, porém com um custo computacional muito mais elevado.

A transformação *logit* afeta o cálculo do logaritmo da probabilidade posteriori total (L_{pt}), assim torna-se necessário ajustar os valores obtidos para que o desempenho do modelo probabilístico possa ser analisado no campo das vazões.

Para a análise dos resultados das amostras da distribuição normal multivariada adicionou-se um fator de transformação dos valores em escala *logit* para a escala original de vazão, segundo:

$$q_{obs} = inv.logit(lq_{obs}) = \log \left(\frac{q_{max} - q_{min}}{(q_{obs} - q_{min})(q_{max} - q_{obs})} \right) \quad (27)$$

A Tabela 1 apresenta os valores de L_{pt} , obtidos com a aplicação do modelo probabilístico estático multivariado, com posteriori normal multivariada, para o período de treinamento (24/05/2013 até 02/06/2015) e para verificação (03/06/2015 até 25/06/2016), aos quatro modelos determinísticos.

TABELA 1: Valor do logaritmo da probabilidade posteriori total, obtido com a otimização do modelo estático multivariado com posteriori normal multivariada

	L_{pt}			
	Sacramento	3R	Regressivo	MLP
Treinamento	-64548,79	-58849,00	-75424,54	-57755,81
Verificação	-22315,92	-19508,71	-25126,63	-19777,71

Pode-se observar na Tabela 1 a diferença existente entre os valores de L_{pt} obtidos para os diferentes modelos hidrológicos determinísticos. O melhor resultado foi encontrado para o modelo 3R, seguido da rede neural MLP, do modelo Sacramento e por fim do Regressivo.

Como se suspeitava, os baixos valores de L_{pt} obtidos levantam a hipótese de que a distribuição normal multivariada pode não ter se ajustado adequadamente aos erros dos modelos determinísticos, com isso optou-se por testar a distribuição multivariada *skew-t* a fim de maximizar os valores de L_{pt} e obter melhores resultados com as

aplicações dos modelos probabilísticos.

4.1.2 Ajuste de uma Distribuição Skew-t Multivariada

As distribuições normais servem de referência quanto à curtose das demais distribuições consideradas unimodais. Muito embora seja comum explicar a curtose como o “grau de achatamento” de uma distribuição, o que as medidas de curtose buscam indicar realmente é o grau de concentração de valores da distribuição em torno do centro desta distribuição. Com isso, a pouca quantidade de valores considerados extremos em dados de vazão (cheias ou estiagens) tornam-se menos importantes, prejudicando as análises dos resultados.

Além disso, a simetria de uma distribuição normal associa a ela uma curva com “caudas” simétricas, enquanto que em uma distribuição assimétrica esta “cauda” pode ser mais longa para a direita (assimetria positiva) ou para a esquerda (assimetria negativa). Nas distribuições assimétricas, os valores da moda, da mediana e da média divergem sendo que a média sempre estará do mesmo lado que a cauda mais longa.

Estes fatores contribuíram para a escolha da distribuição *skew-t* multivariada, ou *M-skew-t* (ARELLANO-VALLE; GENTON, 2010).

Um vetor aleatório p -dimensional Y possui uma distribuição multivariada *skew-t* estendida (MEST), denotada por $Y \sim EST_p(\xi, \Omega, \lambda, \nu, \tau)$, se sua função densidade em $y \in \mathbb{R}^p$ é:

$$EST_p(\xi, \Omega, \lambda, \nu, \tau) = \frac{1}{T_1\left(\tau/\sqrt{1+\lambda^T\Omega\lambda}; \nu\right)} t_p(y; \xi, \Omega, \nu) T_1\left\{(\lambda^T z + \tau) \left(\frac{\nu+p}{\nu+Q(z)}\right)^{1/2}; \nu+p\right\} \quad (28)$$

em que $z = \omega^{-1}(y - \xi)$, $Q(z) = z^T \Omega^{-1} z$, $\lambda \in \mathbb{R}^p$ é o parâmetro de inclinação, $\tau \in \mathbb{R}$ é o parâmetro de extensão. t_p é a função densidade da distribuição *t de Student* p -dimensional:

$$t_p(y; \xi, \Omega, \nu) = \frac{\Gamma((\nu+p)/2)}{|\Omega|^{1/2} (\nu\pi)^{p/2} \Gamma(\nu/2)} \left(1 + \frac{Q(z)}{\nu}\right)^{-(\nu+p)/2} \quad (29)$$

t_p possui parâmetro de localização $\xi \in \mathbb{R}^p$, matriz de dispersão $p \times p$ positiva definida Ω , com $\omega = \text{diag}(\Omega)^{1/2}$, matriz $p \times p$ de correlação $\bar{\Omega} = \omega^{-1} \Omega \omega^{-1}$, $\nu > 0$ graus de liberdade, e $T_1(y; \nu)$ representando a função distribuição acumulada *t de Student* padrão com $\nu > 0$ graus de liberdade.

A distribuição skew-*t* estendida fornece uma classe muito flexível de modelos estatísticos. Para $\lambda = \tau = 0$, tem-se a distribuição $t_p(\xi, \Omega, \nu)$, *t de Student* multivariada simétrica, que se reduz às distribuições multivariadas normal $N_p(\xi, \Omega)$ e Cauchy $C_p(\xi, \Omega)$ quando $\nu \rightarrow \infty$ e $\nu = 1$, respectivamente, o mesmo ocorre quando $\tau \rightarrow +\infty$. Para $\tau \rightarrow -\infty$ a densidade se degenera para zero. Para $\tau = 0$, tem-se a distribuição skew-*t* multivariada $ST_p(\xi, \Omega, \lambda, \nu)$ na forma proposta por Azzalini (2014) e esta distribuição é adotada neste trabalho.

O modelo probabilístico estático multivariado, que possui como posteriori a distribuição skew-*t* multivariada $ST_p(\xi, \Omega, \lambda, \nu)$, também é composto de quatro parâmetros. O parâmetro de localização ξ (vetor de 12 dimensões), escala Ω (matriz de covariância de dimensão 12×12), inclinação λ (vetor de 12 dimensões) e graus de liberdade ν (número real). Esta distribuição pode ser aplicada à variável aleatória que corresponde ao vetor de erros (transformados pela função *logit*) de cada um dos modelos determinísticos.

Otimizações do Stan, de quatro cadeias de Markov em paralelo, são aplicadas para os 12 horizontes de previsão de vazão simultaneamente com o intuito de determinar representantes da distribuição posteriori dos parâmetros do modelo.

O código, desenvolvido em linguagem Stan, para aplicação do MCMC para o modelo estático multivariado com distribuição de probabilidade posteriori skew-*t* multivariada, é apresentado na Figura 38 do Anexo A.

A função *multi_skew_t_log* utilizada para o cálculo da distribuição skew-*t* multivariada, deve ser implementada pois não se encontra dentre as funções básicas pertencentes à biblioteca do Stan. Um módulo *functions* é adicionado ao programa desenvolvido e apresentado na Figura 39 do Anexo A.

Para a análise dos resultados das amostras da distribuição *skew-t* multivariada, adicionou-se o fator de transformação (EQUAÇÃO (27)) dos valores em escala *logit* para a escala original de vazão.

A Tabela 2 apresenta os valores do logaritmo da probabilidade posteriori total (L_{pt}), representando todos os horizontes, obtidos com a aplicação do novo modelo probabilístico estático multivariado com distribuição *skew-t* multivariada para o período de treinamento (24/05/2013 até 02/06/2015) e para verificação (03/06/2015 até 25/06/2016).

TABELA 2: Valor do logaritmo da probabilidade posteriori total obtido com a otimização do modelo estático multivariado com posteriori *skew-t* multivariada

	L_{pt}			
	Sacramento	3R	Regressivo	MLP
Treinamento	-17848,69	-15336,78	-13694,86	-14460,95
Verificação	-11341,15	-8728,44	-8053,07	-8615,36

A melhoria obtida com a utilização da distribuição *skew-t* multivariada em substituição da distribuição normal multivariada pode ser observada na Tabela 3. A comparação é feita sob o modo de modelos não aninhados, pois duas distribuições diferentes, com diferentes quantidades de parâmetros, estão sendo comparadas.

TABELA 3: Percentual de melhoria do valor do logaritmo da probabilidade posteriori total obtido com a otimização do modelo estático multivariado com posteriori *skew-t* multivariada substituindo a posteriori normal multivariada

	L_{pt}			
	Sacramento	3R	Regressivo	MLP
Treinamento	72%	74%	82%	75%
Verificação	49%	55%	67%	56%

O ajuste da distribuição *skew-t* multivariada aos erros dos modelos determinísticos possibilitou um aumento significativo nos valores de L_{pt} se comparados aos da distribuição normal multivariada, confirmando a hipótese de que “caudas” menos achatadas da distribuição e sua assimetria representariam com maior confiabilidade os diferentes tipos de erros associados a cada modelo determinístico.

O aumento no número de parâmetros do modelo probabilístico com posteriori normal multivariada para o modelo com posteriori *skew-t* multivariada não prejudicou a

avaliação da melhoria dos valores de L_{pt} , como pode ser observado com os valores de AIC (EQUAÇÃO (7)) e BIC (EQUAÇÃO (8)), bem como pelo resultado do teste Vuong (EQUAÇÃO (9)), todos aplicados ao período de verificação e apresentados na Tabela 4.

TABELA 4: Valores de AIC, BIC e Vuong dos modelos estáticos multivariados, com posteriori normal ou *skew-t* multivariadas, aplicados ao período de verificação

	Sacramento	3R	Regressivo	MLP
	AIC			
Normal	44811,84	39197,42	50433,26	39735,42
Skew-t	22888,30	17662,88	16312,14	17436,72
	BIC			
Normal	45293,33	39678,91	50914,75	40216,91
Skew-t	23439,34	18213,92	16863,18	17987,76
	Vuong			
Q_v (EQUAÇÃO (9))	23,66	26,20	34,38	23,39

Na Tabela 4, observa-se que os valores de AIC e BIC foram menores para os modelos probabilísticos que utilizaram como posteriori a distribuição *skew-t* multivariada. Além disso, todos os valores de Q_v obtidos com a Equação (9), em que o modelo *A* utiliza distribuição *skew-t* e o modelo *B* distribuição normal, são maiores que o quantil de $(1 - \alpha)$ da distribuição normal padrão, com nível de significância $\alpha = 5\%$, portanto segundo o teste de Vuong, prefere-se o modelo com posteriori *skew-t* multivariada.

Mesmo com a melhoria obtida com a utilização da distribuição *skew-t* multivariada, nota-se a necessidade de desenvolver uma nova maneira de se incorporar os dados hidrológicos existentes às informações obtidas com a aplicação dos modelos estáticos.

Dessa forma, modelos dinâmicos são desenvolvidos utilizando os resultados dos modelos estáticos como priori, e utilizando as informações hidrológicas disponíveis em uma combinação proposta pela inferência bayesiana.

4.2 MODELOS PROBABILÍSTICOS DINÂMICOS MULTIVARIADOS

Os modelos dinâmicos multivariados são um aperfeiçoamento dos modelos estáticos. Variáveis preditoras condicionam os parâmetros do modelo de acordo com

o estado hidrometeorológico do momento da previsão e os resultados dos modelos estáticos tornam-se *prioris* dos dinâmicos.

A quantidade de parâmetros do modelo probabilístico dinâmico é a quantidade de parâmetros do modelo probabilístico estático acrescida de mais oito parâmetros, chamados de parâmetros de translação ($\xi_{t1}, \xi_{t2}, \lambda_{t1}, \lambda_{t2}$) e parâmetros de rotação do modelo dinâmico ($\omega_{r1}, \omega_{r2}, v_{r1}, v_{r2}$).

A distribuição multivariada *skew-t* $ST_p(\xi_{new}, \Omega_{new}, \lambda_{new}, v_{new})$ é composta por quatro parâmetros: localização (ξ_{new}), escala (Ω_{new}), inclinação (λ_{new}) e graus de liberdade (v_{new}), que passam a ser funções dos novos parâmetros dinâmicos:

$$\xi_{new} = \xi + (\xi_{t1} + \xi_{t2}p_{\xi}), \xi_{new} \in [\xi_{L0}, \xi_{L1}] \quad (30)$$

$$\Omega_{new} = \Omega(\omega_{r1} + \omega_{r2}p_{\omega}), (\omega_{r1} + \omega_{r2}p_{\omega}) \in [\omega_{L0}, \omega_{L1}] \quad (31)$$

$$\lambda_{new} = \lambda + (\lambda_{t1} + \lambda_{t2}p_{\lambda}), \lambda_{new} \in [\lambda_{L0}, \lambda_{L1}] \quad (32)$$

$$v_{new} = v(v_{r1} + v_{r2}p_v), (v_{r1} + v_{r2}p_v) \in [v_{L0}, v_{L1}] \quad (33)$$

Desta maneira a distribuição multivariada *skew-t* combina os parâmetros de translação e de rotação com informações de preditores do estado hidrológico ($p_{\xi}, p_{\omega}, p_{\lambda}, p_v$) às informações a priori dos modelos estáticos (ξ, Ω, λ, v), provocando uma perturbação no modelo estático controlada pelos limites $\xi_{L0}, \xi_{L1}, \omega_{L0}, \omega_{L1}, \lambda_{L0}, \lambda_{L1}, v_{L0}, v_{L1}$.

Os limites da perturbação são responsáveis pelo controle dos parâmetros da distribuição de probabilidade escolhida, eles podem ser escolhidos empiricamente, sempre observando o comportamento da distribuição posteriori obtida. A função *lim_vetorial* foi desenvolvida em linguagem Stan para controlar os limites das perturbações (código apresentado na Figura 40 do Anexo A)

O Quadro 2 apresenta os preditores representantes dos estados hidrometeorológicos separados em três grupos: o estado hidrometeorológico da bacia estudada no momento da previsão ou em um passado recente, as informações das previsões hidrometeorológicas determinísticas, e os erros recentes cometidos pelo sistema predictor.

QUADRO 2: Conjuntos de preditores dos estados hidrológicos utilizados nos modelos probabilísticos dinâmicos

Estado hidrometeorológico da bacia estudada no momento da previsão ou em um passado recente	Informações das previsões hidrometeorológicas determinísticas	Erros recentes cometidos pelo sistema predictor
P1: Vazão observada; P2: Vazão observada deslocada em um horizonte passado; P3: Vazão observada deslocada em dois horizontes passados; P4(P5): Degrau positivo (negativo) de vazão observada; P6(P7): Degrau positivo (negativo) de vazão observada deslocado em um horizonte passado; P8(P9): Degrau positivo (negativo) de vazão observada deslocado em dois horizontes passados; P10: Precipitação observada; P11: Precipitação observada deslocada em um horizonte passado; P12: Precipitação observada deslocada em dois horizontes passados;	P13(P14): Coeficientes positivos (negativos) da regressão linear das previsões; P15(P16): Coeficientes positivos (negativos) da regressão linear de degraus de previsão de vazão; P17: Amplitude da previsão de vazão (máx - min); P18: Precipitação prevista; P19: Precipitação prevista em um horizonte à frente; P20: Precipitação prevista em dois horizontes à frente; P21: Precipitação prevista acumulada em quatro horizontes à frente; P22: Precipitação prevista acumulada em doze horizontes à frente;	P23(P24): Erros positivos (negativos) do modelo determinístico; P25(P26): Erros positivos (negativos) do modelo determinístico deslocados em um horizonte passado; P27(P28): Erros positivos (negativos) do modelo determinístico deslocados em dois horizontes passados; P29(P30): Erros positivos (negativos) do modelo determinístico acumulados em quatro horizontes passados; P31(P32): Erros positivos (negativos) do modelo determinístico acumulados em doze horizontes passados;

Para a seleção dos preditores mais relevantes à formação de cada uma das perturbações $(p_{\xi}, p_{\omega}, p_{\lambda}, p_{\nu})$, uma otimização individual de cada um dos parâmetros é realizada a fim de encontrar valores ótimos $(\xi^*, \omega^*, \lambda^*, \nu^*)$ para os quais o valor de L_{pt} no período de treinamento seja maximizado.

Modelos lineares de regressão com os valores ótimos $(\xi^*, \omega^*, \lambda^*, \nu^*)$ dependentes e os 32 preditores, pertencentes aos 3 grupos do Quadro 2, como variáveis regres-

soras foram construídos individualmente. O método *stepwise* (CHAMBERS; HASTIE, 1992) é aplicado a esses modelos para que se faça uma seleção dos preditores mais significativos para cada perturbação dos parâmetros. Por fim, previsões das perturbações $(p_{\xi}, p_{\omega}, p_{\lambda}, p_{\nu})$ para o período de verificação são realizadas. Com estas previsões os parâmetros de translação e rotação são combinados para formarem as perturbações dos parâmetros da distribuição *skew-t* tornando o modelo dinâmico.

Novamente as entradas para o modelo dinâmico passaram por uma transformação *logit* a fim de que os resultados não ultrapassassem os valores de vazão máximos e mínimos.

O código, desenvolvido em linguagem Stan, para aplicação do MCMC para o modelo dinâmico multivariado com distribuição de probabilidade posteriori *skew-t* multivariada, é apresentado em três partes nas Figuras 41, 42 e 43 do Anexo A. O módulo *functions* deve ser adicionado ao código do modelo dinâmico multivariado, contendo as funções *lim_vetorial* (FIGURA 40), *multi_skew_t_log* (FIGURA 39), e as funções *skew_t_log* (FIGURA 44) e *marginal* (FIGURA 45) desenvolvidas para o cálculo das distribuições marginais, utilizadas nos cálculos dos valores de L_{pth} .

Otimizações do Stan, de quatro cadeias em paralelo, são aplicadas com o intuito de determinar amostras representativas da distribuição posteriori dos parâmetros do modelo. E para a análise dos resultados, adicionou-se o fator de transformação (EQUAÇÃO (27)) dos valores em escala *logit* para a escala original de vazão.

A Tabela 5 apresenta os valores do logaritmo da probabilidade posteriori total (L_{pt}), representando todos os horizontes, obtidos com a aplicação do novo modelo probabilístico dinâmico multivariado com distribuição *skew-t* para o período de treinamento e para verificação.

A melhoria obtida com a utilização de um modelo dinâmico multivariado em substituição do modelo estático pode ser observada na Tabela 6.

O gráfico da Figura 16 apresenta a comparação entre os valores do logaritmo da probabilidade posteriori para cada um dos horizontes de previsão separadamente

TABELA 5: Valores do logaritmo da probabilidade posteriori total obtido com a otimização do modelo dinâmico multivariado

	<i>Lpt</i>			
	Sacramento	3R	Regressivo	MLP
Treinamento	-15448,00	-13060,00	-11530,00	-12241,00
Verificação	-10145,28	-7419,88	-6531,30	-7192,65

TABELA 6: Percentual de melhoria dos valores do logaritmo da probabilidade posteriori total obtidos com a otimização dos modelos dinâmicos multivariados substituindo os modelos estáticos

	<i>Lpt</i>			
	Sacramento	3R	Regressivo	MLP
Treinamento	13,45%	17,84%	15,8%	15,35%
Verificação	10,54%	15%	19%	16,51%

(L_{pth}), no período de verificação, dos modelos estáticos e dinâmicos aplicados aos quatro modelos determinísticos. Pode-se observar que:

1. os valores de verossimilhança do modelo dinâmico são superiores aos do estático em todos os horizontes para todos os modelos determinísticos;
2. a aplicação do modelo probabilístico dinâmico aos determinísticos empíricos (Regressivo Linear e MLP) resultou em maiores valores de verossimilhança para horizontes de até 42 horas;
3. valores de verossimilhança se tornaram muito similares entre todos os modelos determinísticos nos horizontes de 48 a 72 horas; melhorias médias de 11% para a aplicação do modelo dinâmico ao modelo Sacramento, 13% para o 3R, 16% para o Regressivo Linear e 15% para a MLP.

A dinâmica obtida com a inserção das perturbações limitadas possibilitou aumento nos valores de L_{pt} se comparados aos dos modelos estáticos.

O acréscimo no número de parâmetros do modelo estático para o modelo dinâmico não prejudicou a avaliação da melhoria dos valores de L_{pt} , como pode ser observado com os resultados do teste da razão da verossimilhança (EQUAÇÃO (6)) apresentados na Tabela 7.

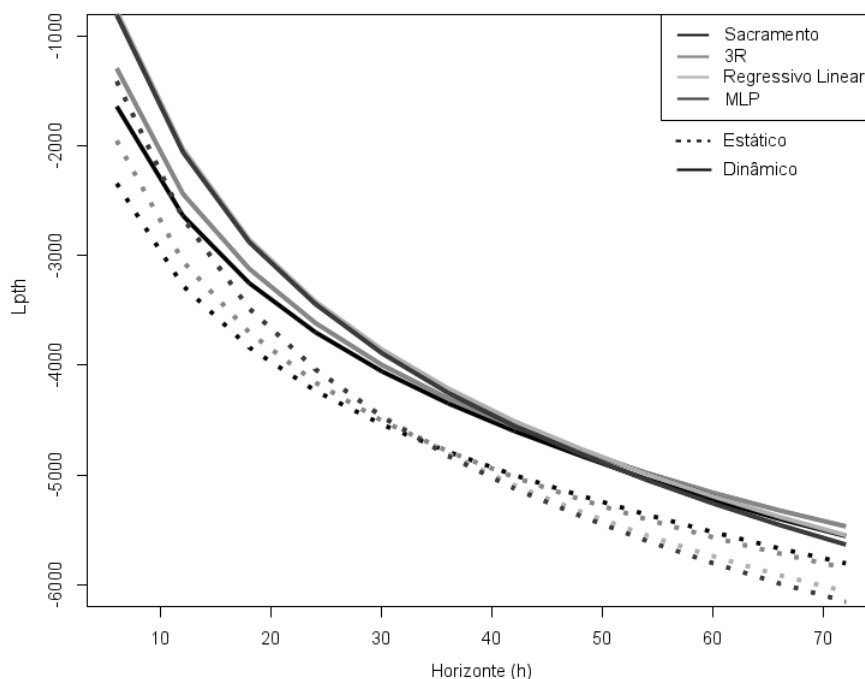


FIGURA 16: Comparação entre os valores do logaritmos da probabilidade posteriori por horizontes dos modelos probabilísticos estático e dinâmico, no período de verificação

FONTE: A autora (2017)

TABELA 7: Valores para o teste da razão da verossimilhança comparando os modelos estáticos multivariados com os modelos dinâmicos, aplicados ao período de verificação

	Razão da verossimilhança			
	Sacramento	3R	Regressivo	MLP
G (EQUAÇÃO(6))	2391,74	2617,12	3043,54	2845,42

Da Tabela 7 para todos os valores de G (EQUAÇÃO (6)) com distribuição chi-quadrado com 8 graus de liberdade (para o qual o modelo estático é o modelo com menor quantidade de parâmetros e o dinâmico com maior, totalizando uma diferença de 8 parâmetros), obtêm-se um p-valor sempre menor que 0,0001, rejeitando a hipótese nula. Assim, o teste da razão da verossimilhança demonstra que os parâmetros de translação e rotação são significativos para o modelo probabilístico dinâmico construído, sem alterar sua parcimônia.

4.3 VISUALIZAÇÃO DOS RESULTADOS DOS MODELOS PROBABILÍSTICOS

Por meio da análise das comparações entre os resultados das abordagens estáticas e dinâmicas, pode-se concluir que devido aos altos valores de L_{pt} e L_{pth} , o modelo dinâmico, com distribuição de probabilidade posterior *skew-t* multivariada, é considerado o modelo probabilístico de maior eficácia. Com isso, desenvolveu-se um sistema de visualização de seus resultados.

Os modelos probabilísticos foram aplicados aos erros cometidos por modelos hidrológicos determinísticos, dessa maneira seus resultados também são representados na forma de erros, isto é, a diferença entre vazão observada e vazão prevista. No entanto, para a construção do sistema de visualização, os resultados na forma de distribuição de probabilidades de erro são transformados novamente em valores de vazão.

A coletânea de figuras a seguir apresenta exemplos dessa visualização.

No quadro esquerdo (a) das Figuras 17 a 21, o intervalo entre as linhas pontilhadas representa as regiões de probabilidade de ocorrência das vazões construídas por meio dos resultados dos modelos probabilísticos estáticos multivariados. As regiões coloridas representam as probabilidades de ocorrência das vazões construídas por meio dos resultados de modelos probabilísticos dinâmicos multivariados. A disposição das cores do modelo estático fazem referência às cores do modelo dinâmico que estão relacionadas aos valores de probabilidades apresentados nas imagens. A linha contínua de cor preta representa os valores de vazão observados, e a linha contínua vermelha representa os valores de vazão previstos pelo modelo determinístico em questão. A linha horizontal azul representa o valor de vazão observado no momento exato da previsão.

No quadro do lado direito (b) das Figuras 17 a 21 são apresentadas as distribuições *skew-t* multivariadas resultantes da aplicação dos modelos estático e dinâmico ao modelo determinístico em questão, para um mesmo momento de previsão, em um

determinado horizonte escolhido. Nestas imagens, o horizonte de 48h foi escolhido, e pode ser alterado conforme necessidade do usuário. Além disso, são apresentados os valores dos erros de um determinado momento (neste caso da média) da distribuição de probabilidade condicionada pela previsão dos modelos probabilísticos estático e dinâmico com relação ao valor de vazão observada, e também os valores de vazão observada, e de vazão prevista pelo modelo determinístico.

As Figuras 17 (a) e (b) representam o resultado das previsões probabilísticas de vazão realizadas na data de 13/06/2015 às 16:00h pelos modelos estático e dinâmico multivariados aplicados ao modelo determinístico 3R, para todos os horizontes até 72h.

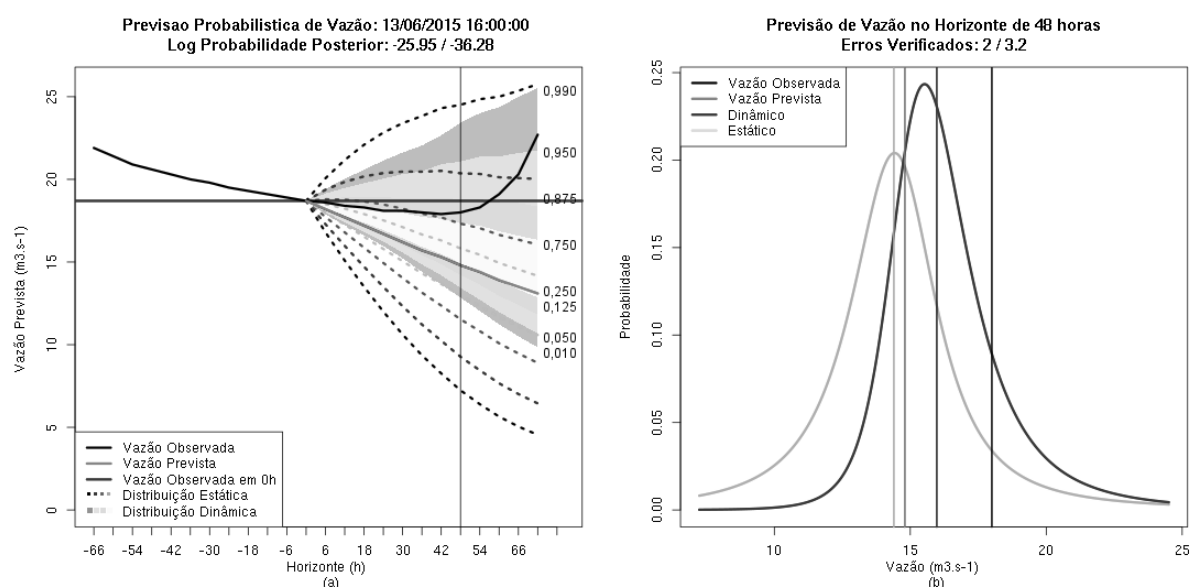


FIGURA 17: Resultados dos modelos probabilísticos aplicados ao modelo determinístico 3R na data 13/06/2015 às 16:00h (a) Previsões dos modelos estático com as distribuições de probabilidade entre as linhas pontilhadas coloridas e previsões do modelo dinâmico com distribuições de probabilidade entre as regiões coloridas. A disposição das cores do modelo estático se referem às cores do modelo dinâmico cujas probabilidades são apresentadas (b) Distribuições de probabilidade *skew-t* resultantes dos modelos estático e dinâmico obtidas para o horizonte de 48h

FONTE: A autora (2017)

O gráfico da Figura 17 (a) permite que se compare o resultado da previsão do modelo determinístico 3R com o valor real da vazão observada e com os resultados dos modelos probabilísticos estáticos e dinâmicos multivariados. Nesse caso, é espe-

rado um aumento nos valores de vazão para horizontes futuros (linha preta), porém o modelo 3R previu valores que diminuem ao longo dos horizontes. Os modelos probabilísticos estático e dinâmico foram capazes de englobar a subida em suas previsões. Além disso, podem-se observar os valores específicos de L_{pt} obtidos para os modelos probabilísticos. O valor de L_{pt} do modelo estático neste momento de previsão é $-36,28$ e do modelo dinâmico é $-25,95$, demonstrando a superioridade do modelo dinâmico neste caso.

O gráfico da Figura 17 (b) apresenta a distribuição *skew-t* multivariada do modelo estático e do modelo dinâmico, para o mesmo momento de previsão, aplicados ao modelo determinístico 3R, no horizonte específico de 48h. Também, pode-se observar o valor do erro de um momento (neste caso a média) da distribuição de probabilidade condicionada pela previsão do modelo probabilístico dinâmico ($2 \text{ m}^3/\text{s}$) que neste caso é menor do que o modelo probabilístico estático ($3,2 \text{ m}^3/\text{s}$).

No mesmo momento de previsão, da data de 13/06/2015 às 16:00h, pode-se visualizar os resultados dos modelos probabilísticos aplicados aos demais modelos determinísticos. Para o modelo Sacramento (FIGURAS 18 (a) e (b)), por exemplo, existe uma pequena diferença nos resultados das previsões, determinísticas e probabilísticas, bem como nos valores de L_{pt} e dos erros. Analisando os resultados para os quatro modelos determinísticos, uma informação completa e robusta das previsões pode ser obtida para futuras análises e estudos de alertas.

Para a data de 04/11/2015 às 04:00h, o valor de L_{pt} do modelo estático é $-70,04$, consideravelmente menor que $-54,09$ do modelo dinâmico, aplicados ao modelo determinístico Sacramento. Estes resultados são representados na Figura 19 (a). As distribuições de probabilidade *skew-t* obtidas para previsão desta data com um horizonte de 48h do modelo Sacramento são apresentadas na Figura 19 (b), o valor do erro de um momento da distribuição de probabilidade condicionada pela previsão do modelo probabilístico dinâmico foi de $19 \text{ m}^3/\text{s}$, menor que do modelo estático que foi de $30,7 \text{ m}^3/\text{s}$.

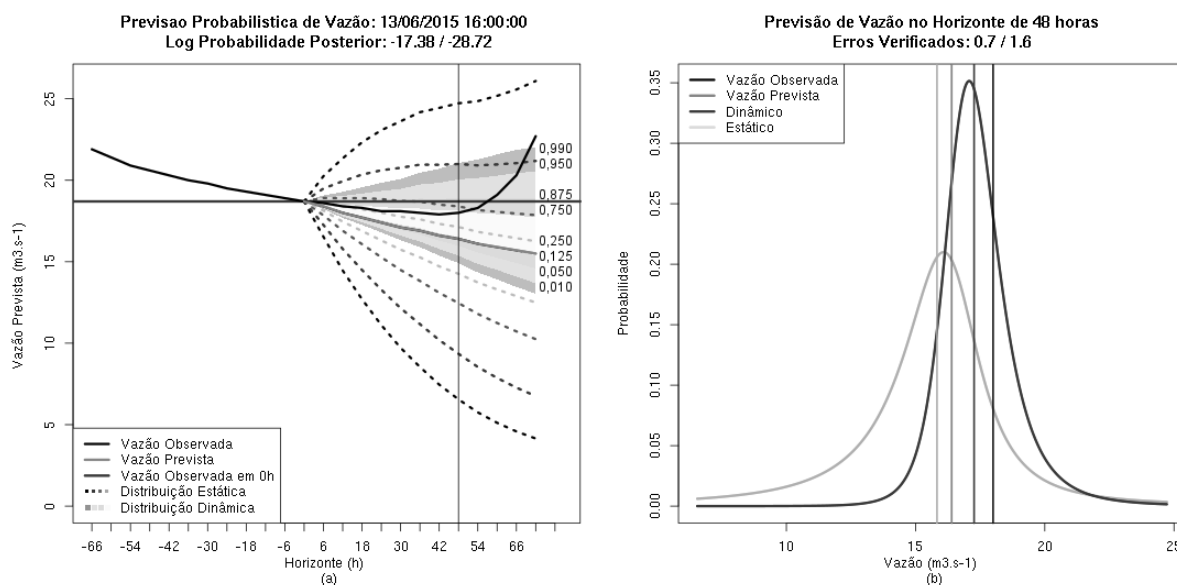


FIGURA 18: Resultados dos modelos probabilísticos aplicados ao modelo determinístico Sacramento na data 13/06/2015 às 16:00h (a) Previsões dos modelos estático com as distribuições de probabilidade entre as linhas pontilhadas coloridas e previsões do modelo dinâmico com distribuições de probabilidade entre as regiões coloridas. A disposição das cores do modelo estático se referem às cores do modelo dinâmico cujas probabilidades são apresentadas (b) Distribuições de probabilidade *skew-t* resultantes dos modelos estático e dinâmico obtidas para o horizonte de 48h

FONTE: A autora (2017)

No momento de previsão da Figura 19, os valores de vazão observados estão diminuindo ao longo dos horizontes, porém o modelo determinístico Sacramento previu um aumento desses valores, e novamente as previsões probabilísticas foram capazes de englobar o comportamento hidrológico futuro dentro de suas probabilidades. Novamente, de acordo com a necessidade do usuário, a visualização deste mesmo momento de previsão pode ser realizada para os demais modelos determinísticos.

Um exemplo em que a previsão probabilística dinâmica não foi capaz de prever os valores de vazão que foram observados nas 72h à frente da data de 18/05/2016 às 16:00h, pode ser observado na Figura 20. Neste caso, os modelos probabilísticos foram aplicados ao modelo determinístico Regressivo Linear Múltiplo, porém esse comportamento também ocorre com a aplicação nos demais modelos determinísticos.

Na data de 18/05/2016 às 22:00h, 6h após a previsão representada na Figura 20, o comportamento dos modelos probabilísticos aplicados ao determinístico Regressivo

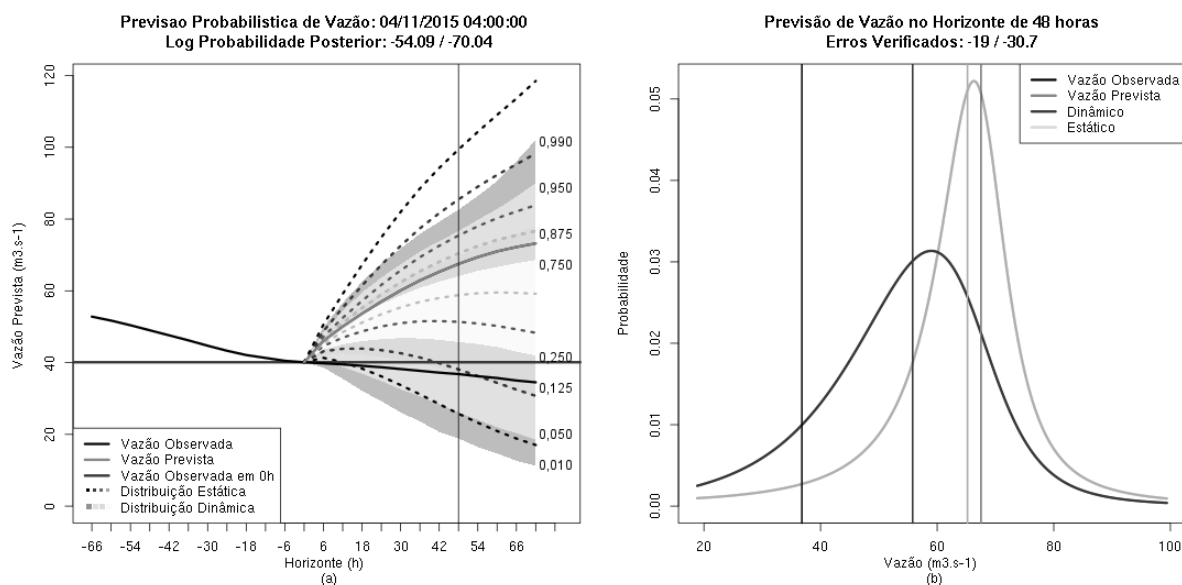


FIGURA 19: Resultados dos modelos probabilísticos aplicados ao modelo determinístico Sacramento na data 04/11/2015 às 04:00h (a) Previsões dos modelos estático com as distribuições de probabilidade entre as linhas pontilhadas coloridas e previsões do modelo dinâmico com distribuições de probabilidade entre as regiões coloridas. A disposição das cores do modelo estático se referem às cores do modelo dinâmico cujas probabilidades são apresentadas (b) Distribuições de probabilidade *skew-t* resultantes dos modelos estático e dinâmico obtidas para o horizonte de 48h

FONTE: A autora (2017)

Linear Múltiplo melhora consideravelmente, como pode ser observado na Figura 21. E este comportamento se repete com a aplicação aos demais modelos determinísticos.

Esta coletânea de imagens permite importantes visualizações dos momentos de previsão dos modelos probabilísticos estático e dinâmico aplicados aos diferentes modelos determinísticos, bem como as seguintes informações:

- variações dos valores de L_{pt} e dos erros de um momento da distribuição resultante do modelos probabilísticos estático e dinâmico, aplicados aos diversos modelos determinísticos e representados por essas figuras;
- variações das análises por horizonte, representado pelo horizonte de 48h escolhido, e a análise de todos os horizontes ao mesmo tempo, para um mesmo momento de previsão;
- importância de se utilizar distribuições multivariadas e assimétricas para cons-

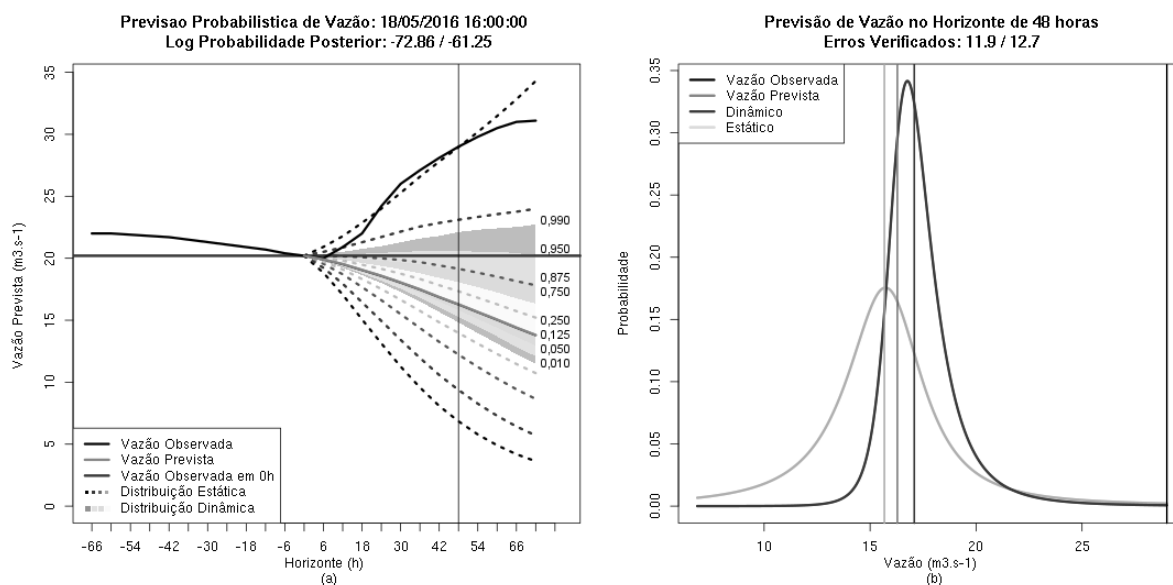


FIGURA 20: Resultados dos modelos probabilísticos aplicados ao modelo determinístico Regressivo na data 18/05/2016 às 16:00h (a) Previsões dos modelos estático com as distribuições de probabilidade entre as linhas pontilhadas coloridas e previsões do modelo dinâmico com distribuições de probabilidade entre as regiões coloridas. A disposição das cores do modelo estático se referem às cores do modelo dinâmico cujas probabilidades são apresentadas (b) Distribuições de probabilidade *skew-t* resultantes dos modelos estático e dinâmico obtidas para o horizonte de 48h

FONTE: A autora (2017)

trução dos modelos probabilísticos estático ou dinâmico;

- comparação do uso dos modelos estático e dinâmico.

No entanto, com a utilização dos quatro modelos determinísticos, são construídos quatro modelos probabilísticos estáticos e quatro modelos probabilísticos dinâmicos. A visualização e análise dos resultados, dessa forma, deve ser realizada de maneira individual para cada modelo determinístico, dificultando o trabalho de um usuário analista de hidrologia.

Com isso, para aprimoramento da previsão probabilística de vazão, principal objetivo deste trabalho, propõe-se uma combinação dos resultados dos modelos probabilísticos dinâmicos aplicados aos quatro determinísticos, e a um novo modelo determinístico, construído para a inclusão da climatologia da sub-bacia em estudo, em forma de *ensemble*. Esta combinação (assunto do capítulo 5) resulta em uma única previsão

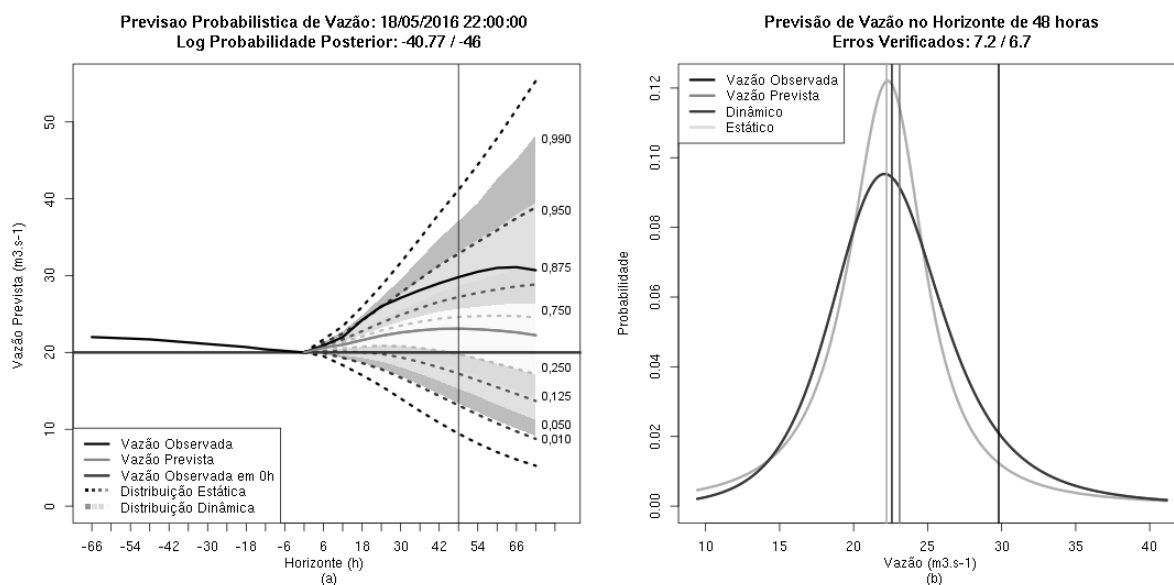


FIGURA 21: Resultados dos modelos probabilísticos aplicados ao modelo determinístico Regressivo na data 18/05/2016 às 22:00h (a) Previsões dos modelos estático com as distribuições de probabilidade entre as linhas pontilhadas coloridas e previsões do modelo dinâmico com distribuições de probabilidade entre as regiões coloridas. A disposição das cores do modelo estático se referem às cores do modelo dinâmico cujas probabilidades são apresentadas (b) Distribuições de probabilidade *skew-t* resultantes dos modelos estático e dinâmico obtidas para o horizonte de 48h

FONTE: A autora (2017)

determinística, para a qual os modelos probabilísticos são aplicados, incorporando todas as vantagens das abordagens apresentadas até o momento.

4.4 PREVISÕES DETERMINÍSTICAS RESULTANTES DOS MODELOS PROBABILÍSTICOS

Existe, ainda, uma dificuldade em se utilizar os resultados das previsões probabilísticas. Segundo Fan, Ramos e Collischonn (2015), esta dificuldade pode estar no fato de que os previsores ou utilizadores dessas previsões, os tomadores de decisão, não estarem certos sobre como a informação probabilística pode realmente ser útil no processo de avaliação e tomada de decisões.

"Como a decisão é sempre vista no seu resultado final, como uma escolha específica, um sim ou não, o uso de uma informação que se apresenta sob a forma de probabilidades (e.g., existe uma probabilidade de 30% de serem observadas vazões acima

de um determinado nível crítico) não é imediato.” (FAN; RAMOS; COLLISCHONN, 2015)

Dessa forma, pode ser necessário calcular previsões determinísticas a partir dos resultados das previsões probabilísticas. Para isso, neste trabalho, pode-se calcular a média, a mediana, ou a moda da distribuição posteriori do modelo probabilístico dinâmico.

Neste sentido, testes foram realizados e conclui-se que a moda da distribuição posteriori é a melhor estimativa de uma previsão determinística, para esta aplicação. Com isso, uma comparação entre essa nova previsão determinística e as previsões dos modelos determinísticos (do tipo chuva-vazão) pode ser realizada.

A Figura 22 apresenta a comparação entre os erros quadráticos médios (EQM) dos modelos determinísticos e da moda obtida dos modelos dinâmicos probabilísticos, para o período de verificação.

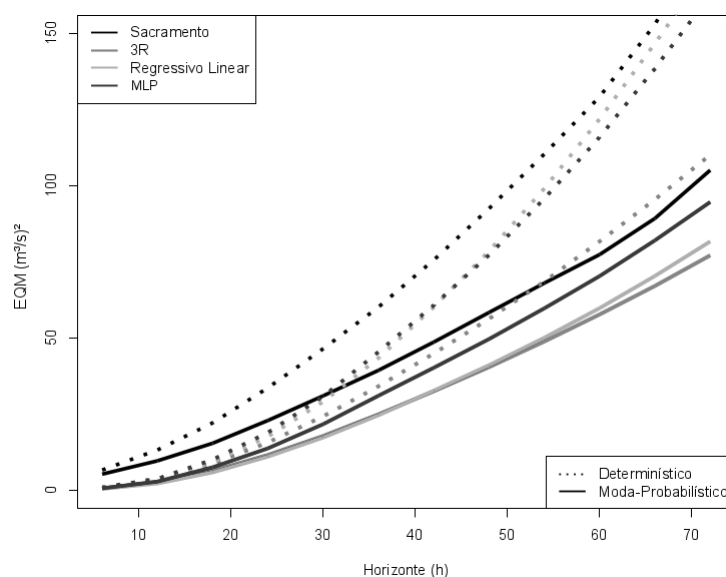


FIGURA 22: EQM dos modelos determinísticos e das modas dos modelos probabilísticos dinâmicos no período de verificação

FONTE: A autora (2017)

Observa-se que o uso da moda introduz maiores benefícios em maiores horizontes de previsão. As melhorias médias (em termos de EQM) obtidas pela comparação entre os modelos determinísticos e as modas são de 34% para o modelo Sacramento, de

29% para o 3R, 43% para o Regressivo Linear Múltiplo e 33% para a MLP .

O coeficiente de *Nash-Sutcliffe* para as modas dos modelos probabilísticos também é calculado, e apresentado na Figura 23. Os valores médios de *Nash-Sutcliffe* para a moda variam de 0,99 a 0,84 entre os horizontes de 6h à 72h.

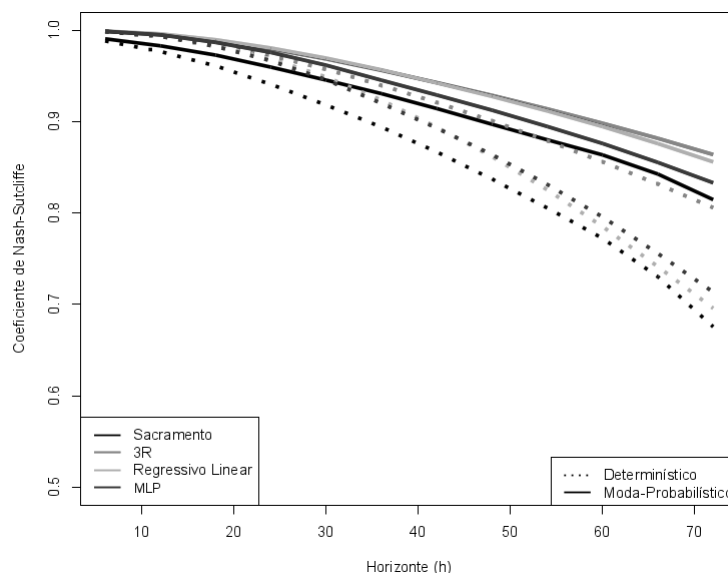


FIGURA 23: Avaliação dos modelos determinísticos e das modas dos modelos probabilísticos dinâmicos no período de verificação segundo o coeficiente de *Nash-Sutcliffe*

FONTE: A autora (2017)

Estes quatro conjuntos de previsões determinísticas, derivadas das modas das distribuições posteriores dos modelos probabilísticos dinâmicos, podem ser utilizados pelos usuários do sistema de previsão como uma opção facilitadora do processo de tomada de decisão.

5 COMBINAÇÃO DAS PREVISÕES PROBABILÍSTICAS

Um meio muito utilizado nos últimos anos para se considerar os diversos tipos de incerteza no processo de previsão hidrometeorológica, de uma maneira geral, é a técnica de previsão por conjunto, ou *ensemble* de previsões.

Segundo Fan *et al.* (2014), a comunidade meteorológica começou a reconhecer a incerteza existente nas previsões de modelos de previsão do tempo, e desenvolveu maneiras de lidar e expressar essa incerteza para uso em modelos de previsão de vazão. Isto foi realizado por meio da adoção de *ensemble* (BUIZZA *et al.*, 1999; KRZYSZTOFOWICZ, 2001; BOWLER *et al.*, 2008; HAMILL; HAGEDORN; HITAKER, 2008). Este fato implicou na adoção das previsões por conjunto como forma de considerar a incerteza na previsão hidrológica (CHEN; YU, 2007; DEMERITT *et al.*, 2007; DAVOLIO *et al.*, 2008; PAPPENBERGER *et al.*, 2008; REGGIANI; WEERTS, 2008; CLOKE; PAPPENBERGER, 2009; THIEMIG *et al.*, 2010; BAO *et al.*, 2011; BOGNER; PAPPENBERGER, 2011; CALVETTI, 2011; CUO; PAGANO; WANG, 2011; SCHELLEKENS *et al.*, 2011; ALFIERI *et al.*, 2012).

De maneira geral, previsões por *ensemble* podem ser desenvolvidas, por exemplo, com perturbações nas condições iniciais de modelos determinísticos, gerando uma saída do modelo para cada uma dessas condições; ou, com perturbações físicas dos modelos, e diferentes parametrizações, mantendo-se as condições iniciais; ou ainda, com a combinação de resultados de diferentes modelos, resultando em previsões hidrológicas probabilísticas.

Neste trabalho, será desenvolvido um *ensemble* que considera todas as informa-

ções obtidas com os quatro modelos determinísticos e com a climatologia da região de estudo, por meio de um novo modelo chamado de *Naif*. O principal objetivo da construção do *ensemble* é resumir as previsões de vazão da sub-bacia de São Bento em apenas uma previsão probabilística.

5.1 CLIMATOLOGIA

O novo modelo determinístico desenvolvido chamado de *Naif*¹, visa promover a incorporação da climatologia da sub-bacia em estudo ao *ensemble*, tornando a previsão de vazão mais completa.

A previsão *Naif* é realizada com base em uma longa série com 30 anos de dados de vazão observados, considerando a vazão atual e o último degrau de vazão (diferença entre o valor de vazão atual e o valor anterior). O último degrau de vazão é reproduzido de acordo com o horizonte que se deseja prever. Se a previsão for realizada para 12 horizontes à frente, então o degrau de vazão será reproduzido 12 vezes. Dessa forma, o comportamento da vazão deve-se manter ao longo do tempo.

Os dados considerados na construção do modelo *Naif* também passaram pela transformação *logit* a fim de que seus resultados não ultrapassassem os valores de vazão máximo e mínimo, observados em toda a série histórica.

A Figura 24 apresenta a comparação entre os valores de REQM (EQUAÇÃO (20)) calculados para os quatro modelos determinísticos, que já haviam sido apresentados na Figura 11, e também para o novo modelo *Naif*. Os resultados do *Naif* demonstraram previsões melhores que os demais modelos determinísticos, para horizontes de até 36h.

Portanto, um modelo probabilístico dinâmico aplicado aos dados desta nova previsão *Naif*, que representa a climatologia da sub-bacia, pode ser combinado às pre-

¹Naif, nome masculino, que tem como significado: aquele que se caracteriza pela simplicidade, ou ingenuidade, segundo o dicionário da língua portuguesa, também dá nome a um tipo de arte, principalmente da pintura, que se caracteriza por linhas e formas simples e resulta em composições sem artifícios (INFOPÉDIA, 2017).

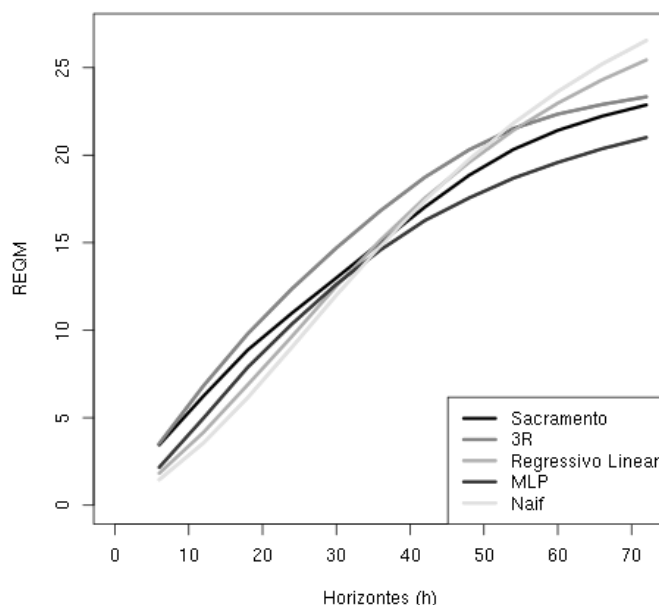


FIGURA 24: Valores de REQM para previsões dos quatro modelos determinísticos e para o novo modelo climatológico *Naif*

FONTE: A autora(2017)

visões dos demais modelos probabilísticos dinâmicos para o desenvolvimento do *ensemble*.

5.2 ENSEMBLE

Após a aplicação dos modelos probabilísticos dinâmicos, que utilizaram os dados dos 5 modelos determinísticos (Sacramento, 3R, Regressivo Linear, MLP e *Naif*), são obtidas estimativas dos parâmetros das distribuições posteriores, por meio da aplicação do MCMC.

Com os parâmetros das distribuições posteriores estimados, torna-se possível a reprodução de distribuições de probabilidade *skew-t* multivariadas para cada momento de previsão (2959 momentos, do período de treinamento dos modelos probabilísticos), de todos os 5 modelos, totalizando 14795 distribuições *skew-t* multivariadas diferentes.

Estas distribuições de probabilidade são combinadas de acordo com a multiplicação delas por uma matriz β formada pelas verossimilhanças apresentadas por horizonte de cada modelo, sendo:

$$\beta_{h,m} = \frac{e^{Lpth(h,m)}}{\sum_{k=1}^n e^{Lpth(h,k)}}, \quad \sum_{m=1}^n \beta_{m,h} = 1 \quad (34)$$

em que m representa cada um dos cinco modelos, h cada um dos doze horizontes, $Lpth$ a verossimilhança para cada modelo em cada um dos horizontes e k a quantidade de dados (os momentos de previsão). A combinação é dada por:

$$g_h(q) = \beta_{h,m} f_{m,h}(q) \quad (35)$$

com f representando as funções de distribuição obtidas a posteriori por meio dos m modelos probabilísticos dinâmicos, para h horizontes de previsão.

Da combinação das distribuições, pode-se obter a mediana resultante da distribuição combinada para cada momento e para cada horizonte de previsão. A mediana da distribuição representa o resultado determinístico da previsão de vazão por *ensemble*, dada por:

$$p_{gh} = \text{mediana}(g_h(q)) \quad (36)$$

Uma representação dessa combinação pode ser vista na Figura 25.

Com o novo resultado determinístico do *ensemble*, os modelos probabilísticos estático e dinâmico são aplicados, resultando nas previsões probabilísticas de vazão por *ensemble*.

A Tabela 8 apresenta a comparação entre os desempenhos do modelo probabilístico dinâmico aplicado ao *ensemble*, e aos demais modelos determinísticos individuais. O modelo combinado pela mediana apresenta maior valor de Lpt em relação a todos os modelos dinâmicos, com uma melhoria média de 38%.

O modelo probabilístico dinâmico aplicado ao *ensemble* também apresenta maior verossimilhança ($Lpth$) em todos os horizontes de previsão, com uma melhoria média de 10%. Estes resultados estão apresentados na Figura 26.

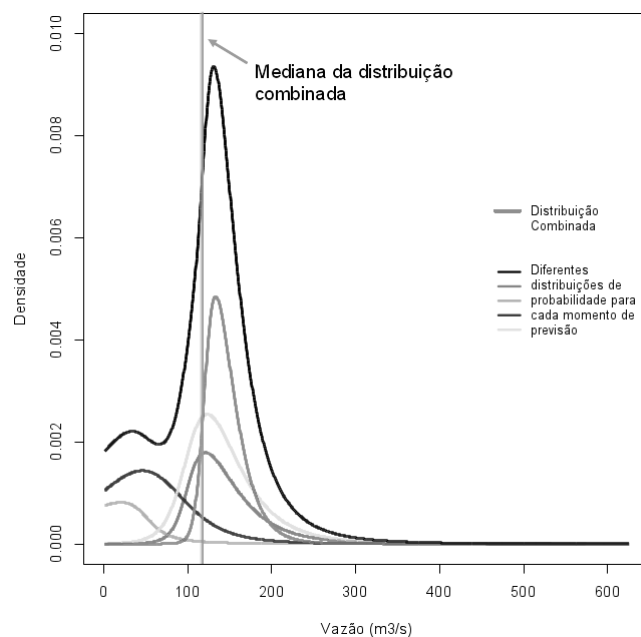


FIGURA 25: Combinação de distribuições de probabilidade *skew-t* para obtenção da mediana da distribuição combinada resultante para determinado horizonte e momento de previsão

FONTE: A autora (2017)

TABELA 8: Avaliação de desempenho do modelo probabilístico baseado na combinação (*ensemble*) dos modelos hidrológicos probabilísticos dinâmicos

	Sacramento	3R	Regressivo	MLP	<i>Ensemble</i>
<i>Lpt</i>	-10145,28	-7419,88	-6531,30	-7192,65	-4830
Melhoria	52%	35%	26%	33%	

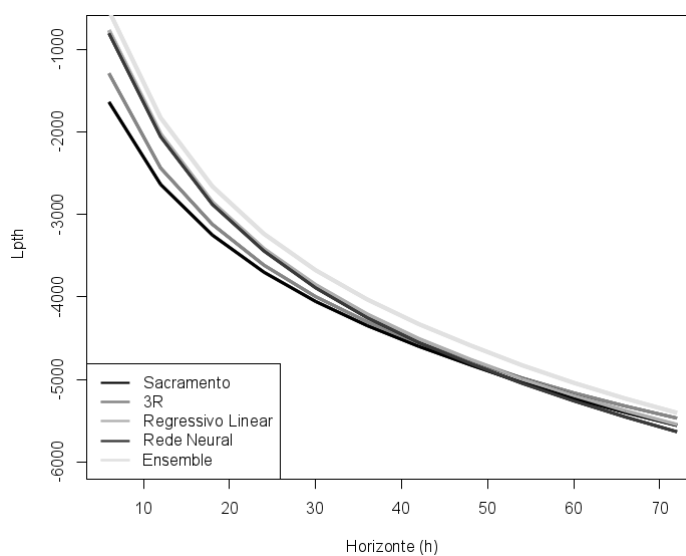


FIGURA 26: Valores de *Lpth* dos modelos dinâmicos *ensemble* e individuais

FONTE: A autora (2017)

A avaliação da previsão determinística oriunda da combinação dos modelos probabilísticos pode ser realizada por meio do cálculo do erro quadrático médio, apresentado na Figura 27, ou do coeficiente de *Nash-Sutcliffe*, conforme Figura 28.

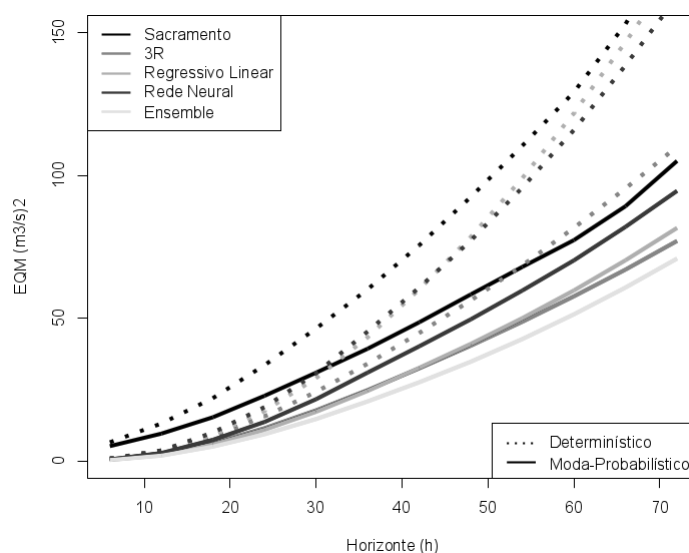


FIGURA 27: EQM dos modelos determinísticos, e das modas dos modelos probabilísticos dinâmicos

FONTE: A autora (2017)

O EQM da moda do modelo probabilístico dinâmico aplicado ao *ensemble* é inferior a todos os EQM das modas dos modelos probabilísticos dinâmicos aplicados aos demais modelos determinísticos, que estão representados na Figura 27.

As melhorias médias (em termos de EQM) obtidas pela comparação entre o resultado do modelo probabilístico dinâmico aplicado ao *ensemble* e aos demais modelos determinísticos, são de 51% para o modelo Sacramento, de 16% para o 3R, 13% para o Regressivo Linear e 30% para a MLP.

O coeficiente de *Nash-Sutcliffe* do *ensemble* variou apenas de 0,99 a 0,88, ao longo de todos os horizontes de previsão (FIGURA 28).

A melhoria obtida com a aplicação do *ensemble* permite uma atualização no sistema de visualização dos resultados da previsão de vazão.

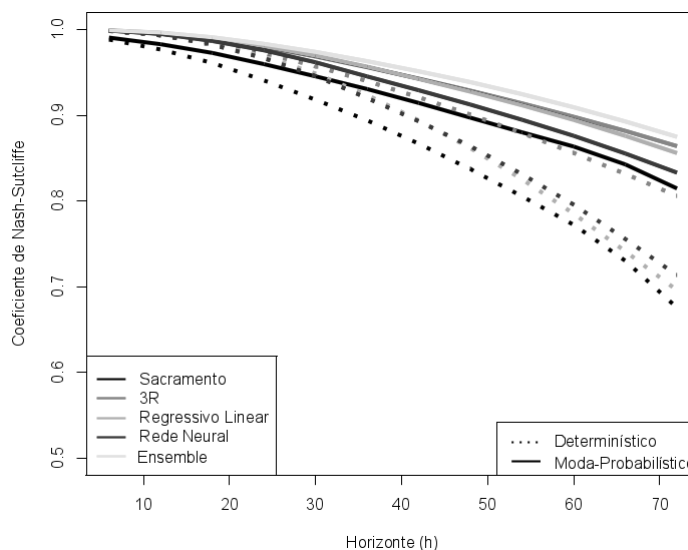


FIGURA 28: Valores dos coeficientes de *Nash-Sutcliffe* dos modelos determinísticos, e das modas dos modelos probabilísticos dinâmicos

FONTE: A autora (2017)

5.3 VISUALIZAÇÃO DOS RESULTADOS DO *ENSEMBLE*

A coletânea de figuras a seguir apresenta exemplos do sistema de visualização com os resultados dos modelos probabilísticos estáticos e dinâmicos aplicados ao *ensemble*. Nas Figuras 29 a 36, foram adicionadas quatro linhas, de cor magenta, contínuas que representam os resultados dos quatro modelos determinísticos, e uma linha pontilhada, de cor magenta, que representa o resultado do novo determinístico *Naif*.

A Figura 29 apresenta de maneira completa a visualização dos resultados dos modelos probabilísticos estático e dinâmico aplicados ao *ensemble* na data de 02/10/2015 às 22:00h, destacando o horizonte de 24h. Neste momento de previsão (FIGURA 29 (a)), os quatro modelos determinísticos indicam aumento nos valores de vazão. O modelo *Naif* indica estabilidade nos valores ao longo dos horizontes. O *ensemble* determinístico resulta em um pequeno aumento nos valores de vazão, e os probabilísticos estático e dinâmico englobam os valores de vazão observados que, ao contrário dos resultados das previsões determinísticas, diminuem ao longo dos horizontes. Neste momento de previsão, para o horizonte de 24h, a diferença entre o valor observado e

o valor previsto pelo modelo probabilístico dinâmico (dado pela média da distribuição de probabilidade obtida) é de apenas $-0,5 \text{ m}^3/\text{s}$, e entre o valor observado e o previsto pelo estático de $-0,8 \text{ m}^3/\text{s}$, as distribuições obtidas para este momento e horizonte de previsão podem ser observadas na Figura 29 (b).

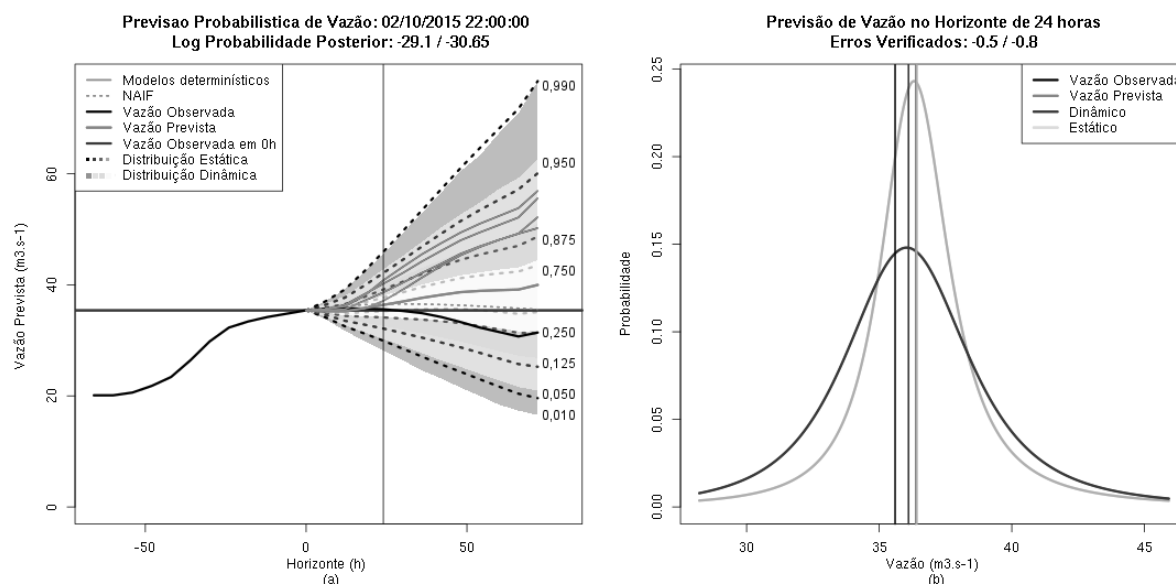


FIGURA 29: Resultados dos modelos probabilísticos aplicados ao *ensemble* na data 02/10/2015 às 22:00h (a) Previsões do modelo estático com as distribuições de probabilidade entre as linhas pontilhadas coloridas e previsões do modelo dinâmico com distribuições de probabilidade entre as regiões coloridas. A disposição das cores do modelo estático se referem às cores do modelo dinâmico cujas probabilidades são apresentadas (b) Distribuições de probabilidade dos modelos estático e dinâmico para o horizonte de 24h

FONTE: A autora (2017)

Para simplificar a visualização dos resultados das previsões para o dia 04/11/2015 às 04:00h, na Figura 30, apenas o modelo probabilístico dinâmico é apresentado. Neste momento pode-se destacar a importância da previsão do modelo *Naif* para a construção do modelo probabilístico. Os resultados do modelo dinâmico só são capazes de englobar a vazão observada pois o *ensemble* considera tanto os modelos determinísticos, que estão prevendo grande elevação nos valores de vazão, quanto o modelo *Naif* que apresenta uma continuidade nos valores de vazão que já haviam sido observados, e que são condizentes com a previsão.

A variação no comportamento da vazão observada no dia 20/09/2015 às 22h, apresentada na Figura 31, ao longo dos 12 horizontes de 6h, dificulta a realização da

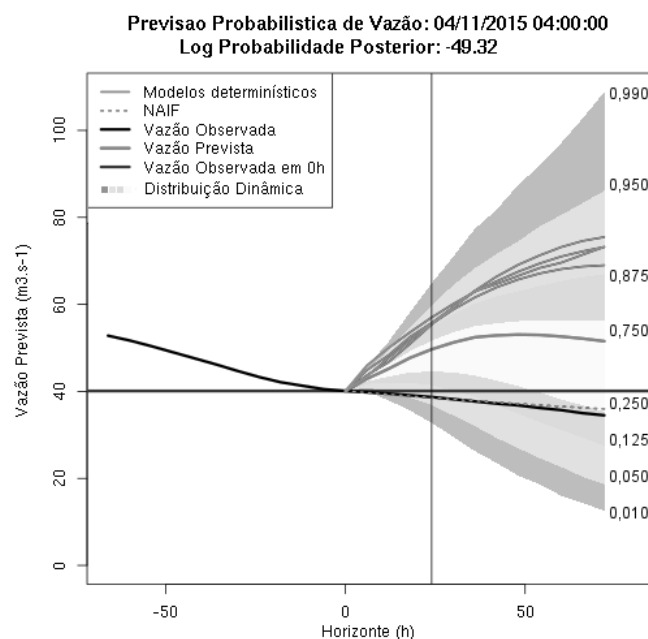


FIGURA 30: Resultado do modelo probabilístico dinâmico, com seus valores de probabilidade definidos entre as regiões coloridas, aplicados ao *ensemble* na data de 04/11/2015 às 04:00h

FONTE: A autora (2017)

previsão. No entanto, o modelo probabilístico dinâmico aplicado ao *ensemble* foi capaz de englobar em sua distribuição de probabilidade toda essa variação, demonstrando a vantagem de sua utilização e da utilização da abordagem multivariada que considera a correlação entre os horizontes de previsão.

A Figura 32 apresenta uma sequência de momentos de previsão dos modelos probabilísticos, estático e dinâmico aplicados ao *ensemble*, na data de 18/05/2016 às 04h (FIGURA 32 (a)), 10h (FIGURA 32 (b)), 16h (FIGURA 32 (c)) e 22h (FIGURA 32 (d)), que apresentam casos de cheia.

No primeiro momento às 04h (FIGURA 32 (a)) o aumento dos valores de vazão foram previstos pelos modelos determinísticos, pelo modelo *Naïf*, pelo *ensemble* e pelos probabilísticos estático e dinâmico. Porém, às 10h (FIGURA 32 (b)) e 16h (FIGURA 32 (c)) os valores continuaram aumentando e as previsões não acompanharam este comportamento. Às 22h (FIGURA 32 (d)), os modelos probabilísticos voltaram a englobar os valores de vazão observados. Este comportamento dos modelos de previsão se repete em muitos casos, demonstrando a importância de se acompanhar o

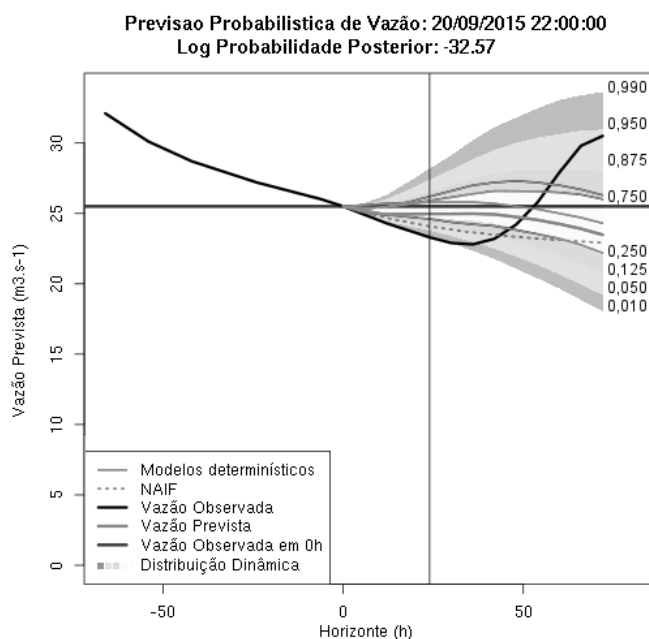


FIGURA 31: Resultado do modelo probabilístico dinâmico, com seus valores de probabilidade definidos entre as regiões coloridas, aplicados ao *ensemble* na data de 20/09/2015 às 22:00h

FORTE: A autora (2017)

desenvolvimento das previsões ao longo dos horizontes.

A Figura 33 apresenta uma sequência de momentos de previsão dos modelos probabilísticos, estático e dinâmico aplicados ao *ensemble*, na data de 23/09/2015 às 10h (FIGURA 33 (a)), 16h (FIGURA 33 (b)), 22h (FIGURA 33 (c)) e 04h do dia 24/09/2015 (FIGURA 33 (d)), que apresentam casos de estiagem. No início desta sequência, apenas o modelo probabilístico estático é capaz de englobar na previsão os valores de vazão, porém ao longo dos horizontes a previsão se torna mais adequada.

As Figuras 34 e 35 apresentam exemplos dos bons resultados do modelo probabilístico dinâmico em casos de cheias observadas no período de verificação dos modelos probabilísticos, que não foram corretamente previstas pelos modelos determinísticos.

A cheia do dia 22/05/2016 (FIGURA 34 (a)) foi subestimada pelos modelos determinísticos, não foi prevista pelo modelo *Naif*, porém ficou dentro das faixas de probabilidade dos modelos probabilísticos estático e dinâmico. Às 04h deste, dia o erro dado pela média da distribuição obtida pelo modelo dinâmico foi de apenas $0,6 \text{ m}^3/\text{s}$

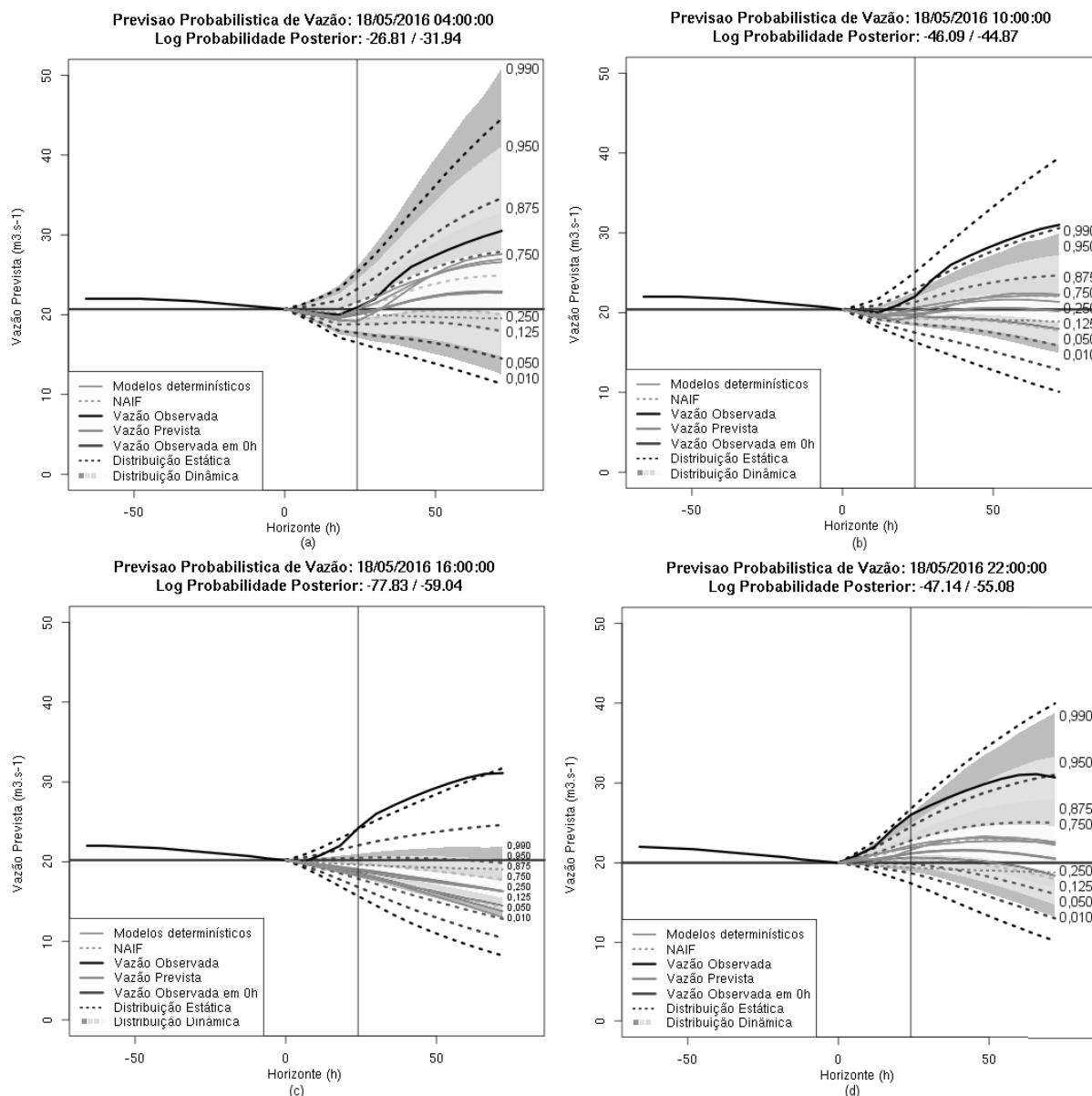


FIGURA 32: Resultados dos modelos probabilísticos aplicados ao *ensemble*. Previsões do modelo estático com as distribuições de probabilidade entre as linhas pontilhadas coloridas e previsões do modelo dinâmico com distribuições de probabilidade entre as regiões coloridas.

A disposição das cores do modelo estático se refere às cores do modelo dinâmico cujas probabilidades são apresentadas (a) 04h do dia 18/05/2016 (b) 10h do dia 18/05/2016 (c) 16h do dia 18/05/2016 (d) 22h do dia 18/05/2016

FONTE: A autora (2017)

e aproximadamente $0 \text{ m}^3/\text{s}$ para o modelo estático, no horizonte de 24h (FIGURA 34 (b)).

A cheia do dia 25/05/2016 (FIGURA 35 (a)) foi superestimada pelos modelos determinísticos, não foi prevista pelo modelo *Naif*, porém ficou dentro das faixas de pro-

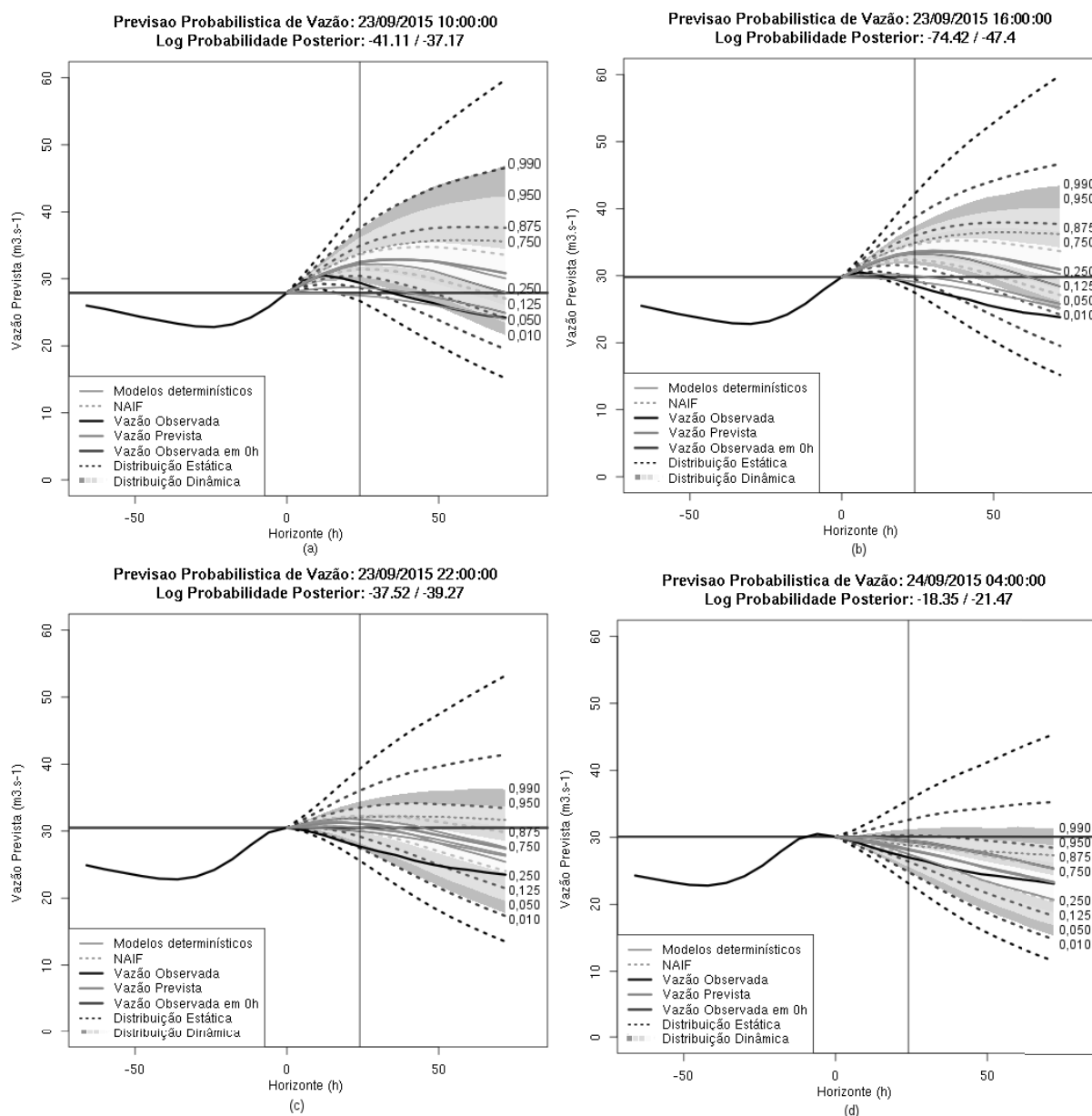


FIGURA 33: Resultados dos modelos probabilísticos aplicados ao *ensemble*. Previsões do modelo estático com as distribuições de probabilidade entre as linhas pontilhadas coloridas e previsões do modelo dinâmico com distribuições de probabilidade entre as regiões coloridas.

A disposição das cores do modelo estático se refere às cores do modelo dinâmico cujas probabilidades são apresentadas (a) 10h do dia 23/09/2015 (b) 16h do dia 23/09/2015 (c) 22h do dia 23/09/2015 (d) 04h do dia 24/09/2015

FONTE: A autora (2017)

bilidade dos modelos probabilísticos estático e dinâmico. Às 04h deste dia o erro dado pela média da distribuição obtida pelo modelo dinâmico foi de apenas $0,8 \text{ m}^3/\text{s}$ e aproximadamente $-2,2 \text{ m}^3/\text{s}$ para o modelo estático, no horizonte de 24h (FIGURA 35 (b)).

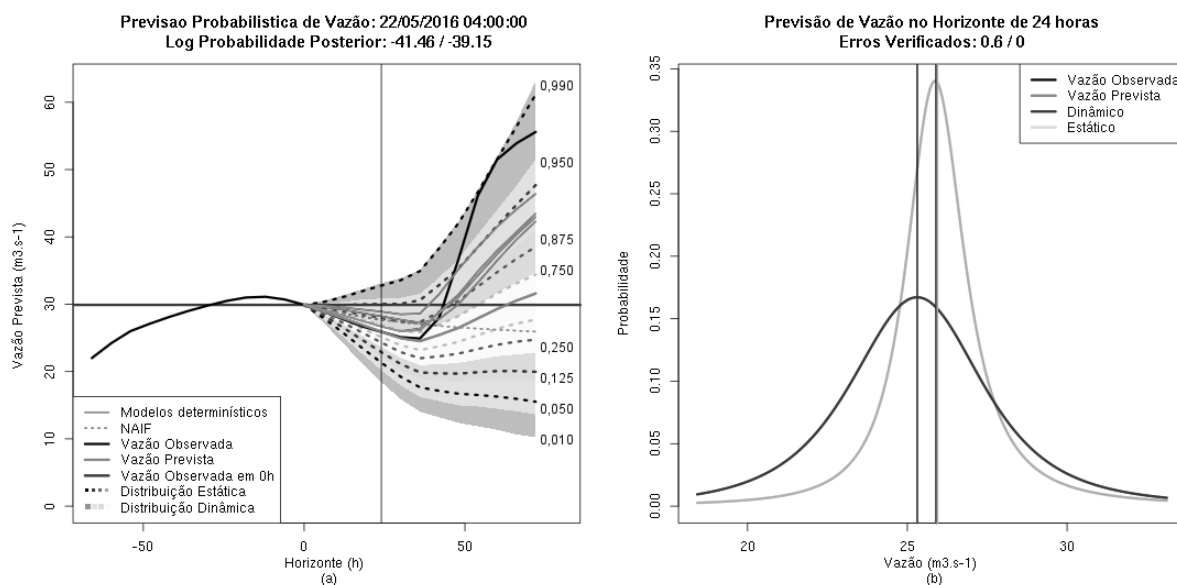


FIGURA 34: Resultados dos modelos probabilísticos aplicados ao *ensemble* na data 22/05/2016 às 04:00h (a) Previsões do modelo estático com as distribuições de probabilidade entre as linhas pontilhadas coloridas e previsões do modelo dinâmico com distribuições de probabilidade entre as regiões coloridas. A disposição das cores do modelo estático se referem às cores do modelo dinâmico cujas probabilidades são apresentadas (b) Distribuições de probabilidade dos modelos estático e dinâmico para o horizonte de 24h

FORTE: A autora (2017)

Todas as figuras apresentadas até este momento representam o período de verificação dos modelos probabilísticos. Porém, a maior cheia observada em todo o período de estudo, para esta sub-bacia, pertence ao período de treinamento dos modelos probabilísticos. A Figura 36 representa a visualização desta cheia.

O valores de vazão futuros à data de 08/06/2014 às 16:00h (FIGURA 36 (a)) aumentaram repentinamente e com grande intensidade. As previsões dos modelos determinísticos, do *Naif*, e do modelo probabilístico estático do *ensemble* subestimaram este aumento. Apenas o modelo probabilístico dinâmico foi capaz de acompanhar este crescimento.

Este sistema de visualização dos resultados das previsões dos modelos, pode se tornar extremamente útil para o acompanhamento, em tempo real, do desenvolvimento de cheias ou estiagens realizado por diversos órgãos de monitoramento. Além disso, a visualização permite o acompanhamento, por parte desses órgãos, do desempenho

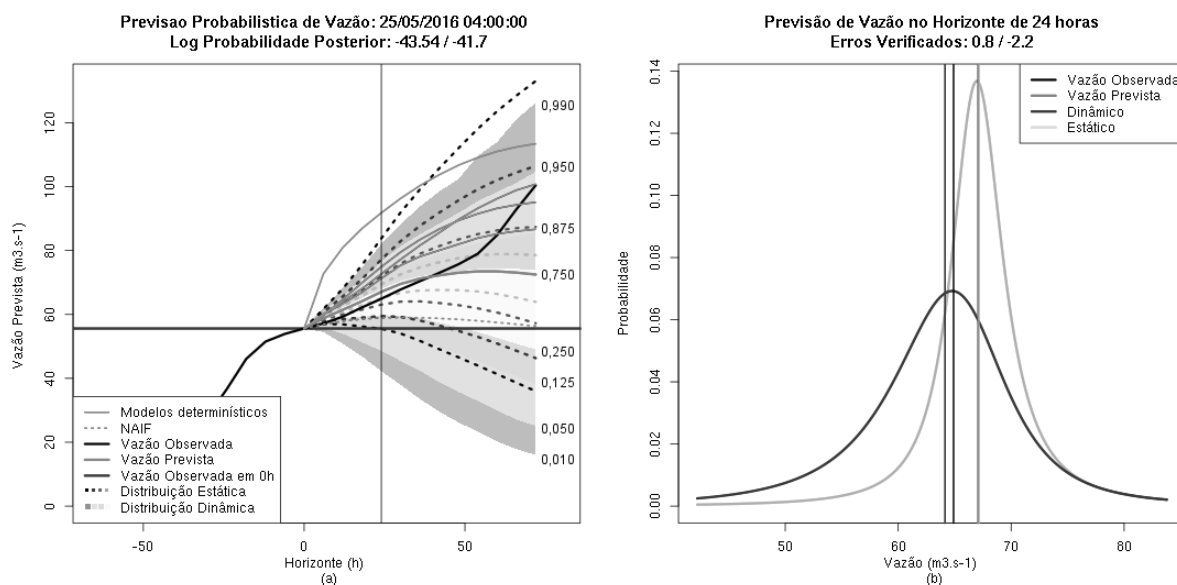


FIGURA 35: Resultados dos modelos probabilísticos aplicados ao *ensemble* na data 25/05/2016 às 04:00h (a) Previsões do modelo estático com as distribuições de probabilidade entre as linhas pontilhadas coloridas e previsões do modelo dinâmico com distribuições de probabilidade entre as regiões coloridas. A disposição das cores do modelo estático se referem às cores do modelo dinâmico cujas probabilidades são apresentadas (b) Distribuições de probabilidade dos modelos estático e dinâmico para o horizonte de 24h

FONTE: A autora (2017)

dos modelos e da necessidade de recalibrá-los com um período maior de dados que venham a ser disponibilizados.

Estudos futuros sobre valores limiares de probabilidade, valores limites superiores e inferiores de vazão, baseados no risco de cheias ou estiagens, podem ser úteis no processo de tomada de decisão, seja para a região de estudo, ou para qualquer outra região para a qual se pretenda aplicar este sistema de previsão probabilística.

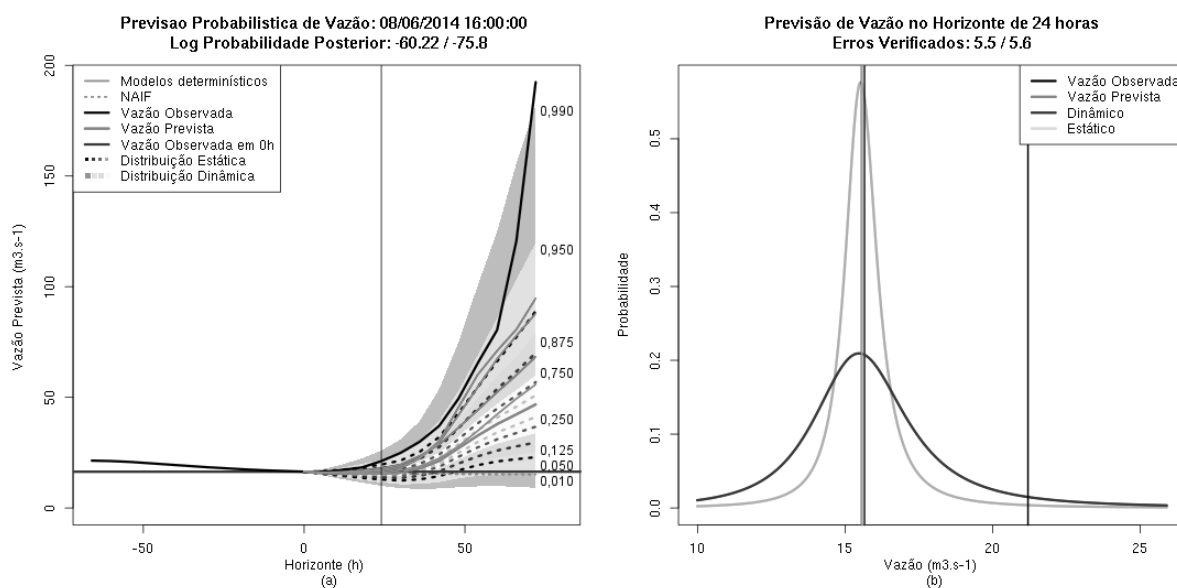


FIGURA 36: Resultados dos modelos probabilísticos aplicados ao *ensemble* na data 08/06/2014 às 16:00h (a) Previsões do modelo estático com as distribuições de probabilidade entre as linhas pontilhadas coloridas e previsões do modelo dinâmico com distribuições de probabilidade entre as regiões coloridas. A disposição das cores do modelo estático se referem às cores do modelo dinâmico cujas probabilidades são apresentadas (b) Distribuições de probabilidade dos modelos estático e dinâmico para o horizonte de 24h

FONTE: A autora (2017)

6 CONCLUSÕES

Este trabalho abordou o complexo problema de previsão probabilística de vazão por meio de inferência bayesiana. Modelos probabilísticos multivariados (estáticos e dinâmicos) foram desenvolvidos e aplicados de uma nova maneira, considerando os erros resultantes de modelos determinísticos do tipo chuva-vazão. Os modelos determinísticos utilizados podem ser subdivididos entre modelos conceituais (Sacramento e 3R) e modelos empíricos (Regressivo Linear e rede neural MLP).

A principal motivação para a escolha dos modelos hidrológicos determinísticos de características tão distintas foi verificar se os métodos de amostragem, como o MCMC da inferência bayesiana, são adequados para a calibração dos modelos probabilísticos desenvolvidos. Os valores do logaritmo da probabilidade posterior total (L_{pt}), bem como os valores do logaritmo da probabilidade posterior total por horizonte de previsão (L_{pth}), demonstraram o bom funcionamento dos modelos probabilísticos aplicados aos erros dos diferentes modelos determinísticos.

Ainda baseando-se nos resultados, ficou explícita a importância de se utilizar modelos multivariados que considerem todos os horizontes de previsão ao mesmo tempo, bem como a correlação existente entre horizontes vizinhos. É a abordagem multivariada que torna possível a geração de diferentes cenários de previsão de vazão, ou seja, é possível saber a probabilidade do valor da vazão ultrapassar um certo limite de risco nas próximas 24h, ou 48h, ou até mesmo 72h, entre outros horizontes. Por este motivo, a implementação (em linguagem Stan) e utilização da distribuição de probabilidade multivariada *Skew-t* também mostrou-se de extrema importância.

Além disso, a dinâmica obtida com o uso de preditores nos parâmetros do modelo probabilístico desenvolvido certamente contribuiu para a melhoria do desempenho das aplicações aos modelos determinísticos.

Por meio do desenvolvimento de um novo sistema de visualização, destacou-se o uso do modelo probabilístico dinâmico multivariado. Isto demonstra que, mesmo em casos de difícil previsão tais como cheias, estiagens, ou outros eventos com variações dos valores de vazão inesperados, o modelo probabilístico dinâmico superou as abordagens determinísticas e probabilísticas estáticas. Mesmo em casos que os valores previstos pelos modelos determinísticos foram muito diferentes dos valores de vazão observados, a abordagem proposta foi capaz de construir regiões que representaram a probabilidade de ocorrência das vazões.

Desta forma, para a sub-bacia de São Bento em estudo, são obtidos quatro resultados de previsões probabilísticas estáticas e quatro resultados de previsões probabilísticas dinâmicas, uma para cada modelo determinístico. Com isso, propôs-se um *ensemble* dos modelos probabilísticos para a criação de uma única previsão de vazão mais robusta e completa.

Com o intuito de completar a combinação dos modelos, desenvolveu-se um novo modelo determinístico, de maneira muito simplificada, que considera a climatologia da região de estudo. Este modelo combinou 30 anos de dados de vazão observados gerando uma nova previsão determinística chamada de *Naif*. O novo modelo *Naif* utiliza somente dados de vazão e não necessita de estimativas nem previsões de chuva.

Os modelos probabilísticos, estático e dinâmico, foram aplicados aos erros do novo modelo determinístico *Naif* e, finalmente, foram incorporados ao *ensemble*. Dessa forma, o *ensemble* combinou a climatologia e os resultados dos modelos probabilísticos dinâmicos aplicados aos quatro modelos determinísticos. Como resultado, obteve-se uma única previsão determinística proveniente de outras cinco. A inclusão da climatologia da região, por meio do modelo *Naif*, garantiu ao *ensemble* uma robustez na sua previsão determinística.

Os modelos probabilísticos, estático e dinâmico, foram aplicados novamente à previsão por *ensemble*, resultando maiores valores do logaritmo da probabilidade poste-

rior total (L_{pt}) e também do logaritmo da probabilidade posterior total por horizonte de previsão (L_{pth}). Seus resultados se demonstraram superiores aos testes que já haviam sido realizados.

O sistema de visualização dos resultados dos modelos probabilísticos estático e dinâmico aplicados à previsão por *ensemble* foi atualizado e apresentado, demonstrando que pode se tornar útil para o acompanhamento do desenvolvimento de cheias ou estiagens em tempo real, cumprindo o principal objetivo de aperfeiçoar as operações de reservatórios, facilitar a tomada de decisão e a prevenção das consequências de eventos extremos.

6.1 TRABALHOS FUTUROS

A inferência bayesiana utilizada neste trabalho mostrou-se adequada ao ajuste dos parâmetros dos modelos probabilísticos desenvolvidos. Desta forma, esta abordagem pode ser aplicada para modelos determinísticos dos mais diversos tipos, e espera-se para o futuro aprofundar os estudos sobre diferentes métodos de amostragem da inferência bayesiana. Diante disso, um aprimoramento dos conhecimentos em métodos de amostragem bayesianos mostra-se necessária, bem como de outros métodos utilizados e suas teorias.

Além disso, o sistema desenvolvido neste trabalho foi construído de forma a facilitar a aplicação da mesma metodologia em outras regiões de estudo. Como consequência, trabalhos futuros podem ser realizados abrangendo outras áreas para as quais o controle de cheias ou de estiagens sejam também relevantes.

O sistema atual considera casos passados, e portanto uma interface que permita a visualização em tempo real dos dados disponíveis mostra-se necessária e factível. Com esse novo módulo desenvolvido, o monitoramento das situações hidrológicas de risco seria ainda mais completo. Estas linhas serão investigadas e estudadas a fim de dar continuidade a este trabalho.

REFERÊNCIAS

- ALFAKI, M. **Improving efficiency in parameter estimation using the Hamiltonian Monte Carlo algorithm**. Dissertação (Mestrado) — Department of Informatics, University of Bergen, 2008.
- ALFIERI, L. *et al.* Operational early warning systems for water-related hazards in eu-rope. **Environmental Science Policy**, v. 21, p. 35 – 49, 2012.
- ANA. **Sistema de Informações Hidrológicas**. 2016. Disponível em: <http://www2.ana.gov.br/Paginas/servicos/informacoes/hidrologicas/redehidro.aspx>, acesso em: 12 Janeiro 2016.
- ARELLANO-VALLE, R. B.; GENTON, M. G. Multivariate extended skew-t distributions and related families. **METRON – International Journal of Statistics**, v. 68, p. 201 – 234, 2010.
- ASSUMPÇÃO, T. H. **Statistical assessment of the Monte Carlo method for uncertainty analysis of hydrological models**. Dissertação (Mestrado) — UNESCO-IHE Institute for Water Education, 2015.
- AZZALINI, A. **The Skew-normal and Related Families**. New York: Cambridge, 2014. With the collaboration of Antonela Capitanio.
- BAO, H.-J. *et al.* Coupling ensemble weather predictions based on tigre database with grid-xinjiang model for flood forecast. **Advances in Geosciences**, v. 29, p. 61 – 67, 2011.
- BENKE, K. K.; LOWELL, K. E.; HAMILTON, A. J. Parameter uncertainty, sensitivity analysis and prediction error in a water-balance hydrological model. **Mathematical and Computer Modelling**, v. 47, p. 1134 – 1149, 2008.
- BOGNER, K.; PAPPENBERGER, F. Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. **Water Resources Research**, v. 47, 2011.
- BONAT, W. H. *et al.* Métodos computacionais em inferência estatística. **20 SINAPE Simpósio Nacional de Probabilidade e Estatística**, 2012.
- BOWLER, N. E. *et al.* The mogreps short-range ensemble prediction system. **Quarterly Journal of the Royal Meteorological Society**, v. 134, p. 703–722, 2008.
- BOZDONGAN, H. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. **Psychometrika**, v. 52, p. 345 – 370, 1987.
- BREDA, A. **Avaliação de Melhorias para um Sistema de Previsão Hidrológica Horária**. Dissertação (Mestrado) — Departamento de Hidráulica e Saneamento, Universidade Federal do Paraná, 2008.

- BUIZZA, R. *et al.* Probabilistic predictions of precipitation using the ecmwf ensemble prediction system. **Weather and Forecasting**, v. 14, p. 168 – 189, 1999.
- BURNASH, R. J. C. The nws river forecast system - catchment model. **Computer Models of Watershed Hydrology**, v. 188, p. 311 – 366, 1995.
- CALVETTI, L. **Previsão hidrometeorológica probabilística na Bacia do Alto Iguaçu-PR com os modelos WRF e TopModel**. Tese (Doutorado) — Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, 2011.
- CASELLA, G.; BERGER, R. L. **Statistical Inference**. New York: Thomson Learning, 2001.
- CHAMBERS, J. M.; HASTIE, T. **Statistical Models in S**. Michigan: Wadsworth & Brooks/Cole Advanced Books & Software, 1992.
- CHEN, S.-T.; YU, P.-S. Real-time probabilistic forecasting of flood stages. **Journal of Hydrology**, v. 340, p. 63 – 77, 2007.
- CLOKE, H. L.; PAPPENBERGER, F. Ensemble flood forecasting: A review. **Journal of Hydrology**, v. 375, p. 613 – 626, 2009.
- COLLINS, J. *et al.* A random number generator based on the logit transform of the logistic variable. **Computers in Physics**, v. 6, p. 360 – 632, 1992.
- CUNHA, L. **Sistemas avançados de previsão hidrológica e operação hidráulica em tempo real**. Dissertação (Mestrado) — Departamento de Hidráulica e Saneamento, Universidade Federal do Paraná, 2004.
- CUO, L.; PAGANO, T. C.; WANG, Q. J. A review of quantitative precipitation forecasts and their use in short- to medium-range streamflow forecasting. **Journal of Hydrometeorology**, v. 12, p. 713 – 728, 2011.
- DAVISON, A. C. **Statistical models**. New York: Cambridge, 2003. Volume 11 de Cambridge Series in Statistical and Probabilistic Mathematics.
- DAVOLIO, S. *et al.* A meteo-hydrological prediction system based on a multi-model approach for precipitation forecasting. **Natural Hazards and Earth System Sciences**, v. 8, p. 143 – 159, 2008.
- DEMERITT, D. *et al.* Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting. **Environmental Hazards**, v. 7, p. 115 – 127, 2007.
- DIBALDASSARRE, G.; MONTANARI, A. **Hydrology and Earth System Sciences**, v. 13, p. 913 – 921, 2009.
- DOMENEGHETTI, A.; CASTELLARIN, A.; BRATH, A. Assessing rating-curve uncertainty and its effects on hydraulic calibration. **Hydrology and Earth System Sciences**, v. 16, p. 1191 – 1202, 2012.
- DUANE, S. *et al.* Hybrid monte carlo. **Physics Letters B**, v. 195, p. 216 – 222, 1987.
- FAN, F. M. *et al.* Ensemble streamflow forecasting experiments in a tropical basin: The são francisco river case study. **Journal of Hydrology**, v. 519, p. 2906 – 2919, 2014.

FAN, F. M.; RAMOS, M.-H.; COLLISCHONN, W. Sobre o uso de previsões hidrológicas probabilísticas para tomada de decisão. **Revista Brasileira de Recursos Hídricos**, v. 20, p. 914 – 926, 2015.

GELFAND, A. E.; SMITH, A. F. M. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, v. 85, p. 398 – 409, 1990.

GELMAN, A. *et al.* **Bayesian Data Analysis**. New York: Chapman and Hall, 2003.

GELMAN, A. *et al.* **Bayesian Data Analysis**. New York: Chapman and Hall, 2013.

GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. **IEEE Trans. on pattern analysis and machine intelligence**, v. 6, p. 721 – 741, 1984.

HAARIO, H.; SAKSMAN, E.; TAMMINEN, J. An adaptive metropolis algorithm. **Bernoulli**, v. 7, p. 223 – 242, 2001.

HAMILL, T. M.; HAGEDORN, R.; HITAKER, J. S. W. Probabilistic forecast calibration using ecmwf and gfs ensemble reforecasts. part ii: Precipitation. **Monthly Weather Review**, v. 136, p. 2620 – 2632, 2008.

HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. **Biometrika**, v. 57, p. 97 – 109, 1970.

HAYKIN, S. **Redes Neurais: princípios e prática**. Porto Alegre: Bookman, 2001.

INFOPÉDIA. **Dicionários Porto Editora**. 2017. Disponível em: <https://www.infopedia.pt/dicionarios/lingua-portuguesa/naif>, acesso em: 20 Junho 2017.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. New Jersey: Pearson, 2007.

KAVETSKI, D.; KUCZERA, G.; FRANKS, S. W. Bayesian analysis of input uncertainty in hydrological modeling: 1. theory. **Water Resources Research**, v. 42, 2006.

KAVETSKI, D.; KUCZERA, G.; FRANKS, S. W. Bayesian analysis of input uncertainty in hydrological modeling: 2. application. **Water Resources Research**, v. 42, 2006.

KLEINA, M. **Identificação, Monitoramento e Previsão de Tempestades Elétricas**. Tese (Doutorado) — PPGMNE, Universidade Federal do Paraná, 2015.

KRZYSZTOFOWICZ, R. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. **Water Resources Research**, v. 35, p. 2739 – 2750, 1999.

KRZYSZTOFOWICZ, R. The case for probabilistic forecasting in hydrology. **Journal of Hydrology**, v. 249, p. 2 – 9, 2001.

KRZYSZTOFOWICZ, R. Bayesian system for probabilistic river stage forecasting. **Journal of Hydrology**, v. 268, p. 16 – 40, 2002.

LEITE, E. A. **Gestão do valor da informação hidrometeorológica: O caso dos alertas de inundação para proteção de bens móveis em edificações residenciais de União da Vitória**. Tese (Doutorado) — COPPE, Universidade Federal do Rio de Janeiro, 2008.

LEITE, E. A.; BREDAS, A.; SABOIA, J. P. Sistema de previsão probabilística de variáveis hidrológicas e sua aplicação no alerta de inundações. In: **XXI Seminário Nacional de Produção e Transmissão de Energia Elétrica**. [S.l.: s.n.], 2011.

MARSHALL, L.; NOTT, D.; SHARMA, A. A comparative study of markov chain monte carlo methods for conceptual rainfall-runoff modeling. **Water Resources Research**, v. 40, 2004.

MARSHALL, L.; NOTT, D.; SHARMA, A. Hydrological model selection: a bayesian alternative. **Water Resources Research**, v. 41, 2005.

MCMILLAN, H. *et al.* Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. **Hydrological Process**, v. 24, p. 1270 – 1284, 2010.

METROPOLIS, N. *et al.* Equations of state calculations by fast computing machines. **The Journal of Chemical Physics**, v. 21, p. 1087 – 1091, 1953.

MONTANARI, A.; BRATH, A. A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. **Water Resources Research**, v. 40, 2004.

MONTANARI, A.; GROSSI, G. Estimating the uncertainty of hydrological forecasts: A statistical approach. **Water Resources Research**, v. 44, 2008.

NASH, J. E.; SUTCLIFFE, J. V. River flow forecasting through conceptual models part i — a discussion of principles*. **Journal of Hydrology**, v. 10, p. 282 – 290, 1970.

NETO, R. K. **Atualização de Modelos Chuva-Vazão-Propagação com Estimadores de Estado**. Dissertação (Mestrado) — Departamento de Hidráulica e Saneamento, Universidade Federal do Paraná, 2005.

NETO, R. K.; GUETTER, A. K.; MINE, M. R. M. Modelo hidrológico com atualização de estado parte i: Modelagem hidrológica. **Revista Brasileira de Recursos Hídricos**, v. 12, p. 93 – 103, 2007.

PAPPENBERGER, F. *et al.* New dimensions in early flood warning across the globe using grand-ensemble weather predictions. **Geophysical Research Letters**, v. 35, p. 7, 2008.

PARZEN, E. On estimation of a probability density function and mode. **Annals of Mathematical Statistics**, v. 33, p. 1065 – 1076, 1962.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2012. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org/>>.

REGGIANI, P.; WEERTS, A. H. Probabilistic quantitative precipitation forecast for flood prediction: An application. **Journal of Hydrometeorology**, v. 9, p. 76–95, 2008.

RENNO, C. D.; SOARES, J. V. **Modelos Hidrológicos para Gestão Ambiental**. [S.l.], 2000.

RIBEIRO, A. R. *et al.* Quantificação das incertezas em curvas de descarga de postos fluviométricos utilizando inferência bayesiana. **Anuário do Instituto de Geociências**, v. 40, p. 266–277, 2017.

SANTOS, T. N. dos. **Redes Neurais Artificiais e Relação ZR Aplicadas à Estimativa de Chuva**. Dissertação (Mestrado) — Departamento de Matemática e Departamento de Construção Civil, Universidade Federal do Paraná, 2014.

SCHELLEKENS, J. *et al.* The use of mogreps ensemble rainfall forecasts in operational flood forecasting systems across england and wales. **Advances in Geosciences**, v. 29, p. 77 – 84, 2011.

SCHWARZ, G. Estimating the dimension of a model. **Annals of Statistics**, v. 6, p. 461–464, 1978.

SMITH, T. J.; MARSHALL, L. A. Bayesian methods in hydrologic modeling: A study of recent advancements in markov chain monte carlo techniques. **WATER RESOURCES RESEARCH**, v. 44, 2008.

Stan Development Team. **Stan Modeling Language: User's Guide and Reference Manual**. [S.l.], 2014. Disponível em: <<http://mc-stan.org/>>.

THIEMIG, V. *et al.* Ensemble flood forecasting in africa: a feasibility study in the juba–shabelle river basin. **Atmospheric Science Letters**, 2010.

TUCCI, C. E. M. **Modelos Hidrológicos**. Porto Alegre: Editora da UFRGS, 1998.

VRUGT, J. A. *et al.* Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with markov chain monte carlo simulation. **Water Resources Research**, v. 44, 2008.

VRUGT, J. A. *et al.* Hydrologic data assimilation using particle markov chain monte carlo simulation: Theory, concepts and applications. **Advances in Water Resources**, v. 51, p. 457–478, 2013.

VUONG, Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. **Econometrica**, v. 57, p. 307 – 333, 1989.

WESTERBERG, I. *et al.* Stage-discharge uncertainty derived with a non-stationary rating curve in the choluteca river. **Hydrological Process**, v. 25, p. 603–6013, 2011.

ZELL, A. *et al.* **Stuttgart Neural Network Simulator**. Stuttgart, 1990. Disponível em: <<http://www.ra.cs.uni-tuebingen.de/SNNS/>>.

ANEXO A

```

data {
  int<lower=1> T;
  int<lower=1> T1;
  int<lower=1,upper=12> h;
  matrix[T,h] obs;
  matrix[T,h] prev;
  real q_min;
  real q_max;
}
transformed data {
  matrix[T,h] o_logit;
  matrix[T,h] p_logit;
  for(i in 1:T){
    o_logit[i]<-log((obs[i]-q_min) ./ (q_max-obs[i]));
    p_logit[i]<-log((prev[i]-q_min) ./ (q_max-prev[i]));
  }
}
parameters {
  vector[h] xi;
  cholesky_factor_corr[h] momega;
  vector<lower=0>[h] w;
}
transformed parameters{
  cov_matrix[h] omega;
  omega<-multiply_lower_tri_self_transpose(momega).*(to_matrix(w)*to_matrix(to_row_vector(w)));
}
model {
  real logp;
  for(i in 1:T){
    logp<-multi_normal_log(to_vector(o_logit[i]),xi+to_vector(p_logit[i]),omega)
      + sum(log((q_max-q_min) ./ ((obs[i]-q_min) .* (q_max-obs[i]))));
    increment_log_prob(logp);
  }
}
generated quantities {
  real logp[T-T1];
  for(i in (T1+1):T) {
    logp[i-T1]<-multi_normal_log(to_vector(o_logit[i]),xi+to_vector(p_logit[i]),omega)
      + sum(log((q_max-q_min) ./ ((obs[i]-q_min) .* (q_max-obs[i]))));
  }
}

```

FIGURA 37: Código do modelo estático multivariado, com posteriori normal multivariada, programado em linguagem Stan

FONTE: A autora (2017)

```

data {
  int<lower=1> T;
  int<lower=1> T1;
  int<lower=1,upper=12> h;
  matrix[T,h] obs;
  matrix[T,h] prev;
  real q_max;
  real q_min;
}
transformed data {
  matrix[T,h] o_logit;
  matrix[T,h] p_logit;
  for(i in 1:T) {
    o_logit[i]<-log((obs[i]-q_min) ./ (q_max-obs[i]));
    p_logit[i]<-log((prev[i]-q_min) ./ (q_max-prev[i]));
  }
}
parameters {
  vector[h] xi;
  cholesky_factor_corr[h] momega;
  vector<lower=0>[h] w;
  vector[h] alpha;
  real<lower=0> nu;
}
transformed parameters {
  cov_matrix[h] omega;
  omega<-multiply_lower_tri_self_transpose(momega).*(to_matrix(w)*to_matrix(to_row_vector(w)));
}
model {
  real logp;
  logp<-0.0;
  for(i in 1:T1) {
    logp<-logp + multi_skew_t_log(to_matrix(o_logit[i]),xi+to_vector(p_logit[i]),
                                omega,alpha,nu)+sum(log((q_max-q_min)/((obs[i]-q_min).*(q_max-obs[i]))));
  }
  increment_log_prob(logp);
}
generated quantities {
  real logp[T-T1];
  for(i in (T1+1):T) {
    logp[i-T1]<-multi_skew_t_log(to_matrix(o_logit[i]),xi+to_vector(p_logit[i]),
                                omega,alpha,nu)+sum(log((q_max-q_min)/((obs[i]-q_min).*(q_max-obs[i]))));
  }
}

```

FIGURA 38: Código do modelo estático multivariado, com posteriori *skew-t* multivariada, programado em linguagem Stan

FONTE: A autora (2017)


```

functions {
  real multi_skew_t_log(matrix x, vector xi, matrix omega, vector alpha, real nu) {
    int d;
    int t;
    matrix[dims(alpha)[1],dims(alpha)[1]] inv_omega;
    real log_det;
    real log_const;
    real log_pt;
    real log_PDF;
    matrix[dims(alpha)[1],dims(x)[1]] X;
    matrix[dims(alpha)[1],dims(x)[1]] Y;
    matrix[dims(x)[1],dims(alpha)[1]] tY;
    vector[dims(x)[1]] Q;
    vector[dims(x)[1]] L;
    vector[dims(x)[1]] log_lQ;
    vector[dims(x)[1]] log_dmt;
    d<-dims(alpha)[1];
    t<-dims(x)[1];
    inv_omega<-inverse_spd(omega);
    log_det<-log_determinant(omega);
    for(i in 1:t) tY[i]<-to_row_vector(xi);
    X<-(x - tY)';
    Y<-(inv_omega * X) .* X;
    for(i in 1:t) Q[i]<-sum(col(Y,i));
    for(i in 1:t) tY[i]<- to_row_vector( col(X,i) ./ exp(0.5*log(diagonal(omega))) );
    L<-to_vector(tY * alpha);
    log_const<-lgamma((nu+d)/2.0) - lgamma(nu/2.0) - 0.5*d*log(nu);
    log_lQ<-log(1.0+Q/nu);
    log_dmt<-log_const-0.5*(d*log(pi())+log_det+(nu+d)*log_lQ);
    log_pt<-student_t_cdf_log(to_array_ld(L .* exp(0.5*log((nu+d) ./ (Q+nu)))) ,nu+d,0,1);
    log_PDF<-sum(log(2) + log_dmt) + log_pt;
    return(log_PDF);
  }
}

```

FIGURA 39: Código da função que representa a distribuição *skew-t* multivariada, programado em linguagem Stan

FONTE: A autora (2017)

```

functions{
  vector lim_vetorial(vector lim, vector ref, real t, real r) {
    vector[dims(ref)[1]] ref_u;
    vector[dims(ref)[1]] y;
    int h;
    h<-dims(ref)[1];
    for(i in 1:h) {
      ref_u[i]<-logit((ref[i]-lim[1])/(lim[2]-lim[1]));
      y[i]<-lim[1]+(lim[2]-lim[1])*inv_logit(t + r*ref_u[i]);
    }
    return(y);
  }
}

```

FIGURA 40: Código da função que controla os limites das perturbações dos parâmetros do modelo probabilístico dinâmico multivariado, programado em linguagem Stan

FONTE: A autora (2017)

```

data {
  int<lower=1> T;
  int<lower=1,upper=12> h;
  int<lower=1> G;
  int<lower=0> T2;
  matrix[T,h] obs;
  matrix[T,h] prev;
  real q_min;
  real q_max;
  matrix[T,G] m_p;
  vector[h] i_alpha;
  cholesky_factor_corr[h] i_momega;
  vector[h] i_w;
  vector[h] i_gamma;
  vector[2] l_etha;
  vector[2] l_gamma;
  vector[2] l_nu;
  vector[2] l_alpha;
  real<lower=0> nu;
}
transformed data {
  int<lower=1> T1;
  matrix[T,G+1] mp;
  real ref_etha;
  real ref_nu;
  cov_matrix[h] i_beta;
  matrix[T-T2,G+1] m_p1;
  matrix[T2,G+1] m_p2;
  matrix[T,h] obs_l;
  matrix[T,h] prev_l;
  T1<-T-T2;
  i_beta<-multiply_lower_tri_self_transpose(i_momega).* (to_matrix(i_w)*to_matrix(to_row_vector(i_w)));
  ref_etha<-logit((1.0-l_etha[1])/(l_etha[2]-l_etha[1]));
  ref_nu<-logit((1.0-l_nu[1])/(l_nu[2]-l_nu[1]));
  mp<-append_col(to_vector(rep_array(1.,T)), m_p);
  m_p1<-block(mp,1,1,T1,G+1);
  m_p2<-block(mp,T1+1,1,T2,G+1);
  for(i in 1:T) {
    obs_l[i]<-log((obs[i]-q_min)/(q_max-obs[i]));
    prev_l[i]<-log((prev[i]-q_min)/(q_max-prev[i]));
  }
}

```

FIGURA 41: Primeira parte do código do modelo dinâmico multivariado, com posteriori *skew-t* multivariada, programado em linguagem Stan

FONTE: A autora (2017)

```

parameters {
  vector[2] alpha_t;
  vector[2] gamma_t;
  vector[2] etha;
  vector[2] nu_r;
}
transformed parameters {
  vector[T1] lpt;
  vector[h] lph;
  vector[h] lampda;
  vector[h] alpha;
  real r_etha;
  real r_nu;
  real r_obs0[1];
  lph<-rep_vector(0,h);
  for(i in 1:T1) {
    r_etha<-l_etha[1]+(l_etha[2]-l_etha[1])*inv_logit(ref_etha + etha[1]+m_pl[i,2]*etha[2]);
    r_nu<-l_nu[1]+(l_nu[2]-l_nu[1])*inv_logit(ref_nu + nu_r[1]+m_pl[i,4]*nu_r[2]);
    lampda<-marginal(sqrt(r_etha)*i_w,r_etha*i_beta,lim_vetorial(l_gamma,i_gamma,
      gamma_t[1]+m_pl[i,3]*gamma_t[2],1.0));
    alpha<-lim_vetorial(l_alpha,i_alpha,alpha_t[1]+m_pl[i,5]*alpha_t[2],1.0);
    lpt[i]<-multi_skew_t_log(to_matrix(obs_l[i]),
      alpha+to_vector(prev_l[i]),
      r_etha*i_beta,
      lim_vetorial(l_gamma,i_gamma,gamma_t[1]+m_pl[i,3]*gamma_t[2],1.0),
      r_nu*nu)+sum(log((q_max-q_min)/((obs[i]-q_min)*(q_max-obs[i]))));
    for(j in 1:h) {
      r_obs0[1]<-obs_l[i,j];
      lph[j]<-lph[j]+skew_t_log(r_obs0,alpha[j]+prev_l[i,j],sqrt(r_etha)*i_w[j],
        lampda[j],r_nu*nu)+log((q_max-q_min)/((obs[i,j]-q_min)*(q_max-obs[i,j]))));
    }
  }
}
model {
  increment_log_prob(sum(lpt));
}

```

FIGURA 42: Segunda parte do código do modelo dinâmico multivariado, com posteriori *skew-t* multivariada, programado em linguagem Stan

FONTE: A autora (2017)

```

generated quantities {
  vector[T2] lpt2;
  vector[h] lph2;
  vector[T] v_etha2;
  vector[T] v_nu2;
  matrix[T,h] lampda2;
  matrix[T,h] alpha2;
  real r_obs[1];
  lph2<-rep_vector(0,h);
  for(i in 1:T) {
    v_etha2[i]<-l_etha[1]+(l_etha[2]-l_etha[1])*inv_logit(ref_etha + etha[1]+mp[i,2]*etha[2]);
    v_nu2[i]<-l_nu[1]+(l_nu[2]-l_nu[1])*inv_logit(ref_nu + nu_r[1]+mp[i,4]*nu_r[2]);
    alpha2[i]<-to_row_vector(lim_vetorial(l_alpha,i_alpha,alpha_t[1]+mp[i,5]*alpha_t[2],1.0));
    lampda2[i]<-to_row_vector(marginal(sqrt(v_etha2[i])*i_w,v_etha2[i]*i_beta,
      lim_vetorial(l_gamma,i_gamma,gamma_t[1]+mp[i,3]*gamma_t[2],1.0)));
    if(i>T1) {
      lpt2[i-T1]<-multi_skew_t_log(to_matrix(obs_l[i]),
        to_vector(alpha2[i])+to_vector(prev_l[i]),
        v_etha2[i]*i_beta,
        lim_vetorial(l_gamma,i_gamma,gamma_t[1]+mp[i,3]*gamma_t[2],1.0),
        v_nu2[i]*nu)+sum(log((q_max-q_min)/((obs[i]-q_min)*(q_max-obs[i]))));
    }
    for(j in 1:h) {
      r_obs[1]<-obs_l[i,j];
      lph2[j]<-lph2[j]+skew_t_log(r_obs,alpha2[i,j]+prev_l[i,j],sqrt(v_etha2[i])*i_w[j],
        lampda2[i,j],v_nu2[i]*nu)+log((q_max-q_min)/((obs[i,j]-q_min)*(q_max-obs[i,j]))));
    }
  }
}

```

FIGURA 43: Terceira parte do código do modelo dinâmico multivariado, com posteriori *skew-t* multivariada, programado em linguagem Stan

FONTE: A autora (2017)

```

functions {
  real skew_t_log(real[] x, real xi, real omega, real alpha, real nu) {
    vector[dims(x)[1]] z;
    vector[dims(x)[1]] w;
    int n;
    n<-dims(x)[1];
    z<-(to_vector(x)-xi)/omega;
    w<-alpha*z .* exp(0.5*log(nu+1)-0.5*log((z .* z) + nu));
    return (n*(log(2)-log(omega))+student_t_log(z,nu,0,1)+student_t_cdf_log(w,(nu+1),0,1));
  }
}

```

FIGURA 44: Código da função *skew-t* univariada utilizada para o cálculo das distribuições marginais do modelo probabilístico dinâmico multivariado, programado em linguagem Stan

FONTE: A autora (2017)

```

functions{
vector marginal(vector w, matrix beta, vector gamma) {
  matrix[dims(gamma)[1],dims(gamma)[1]] omega_hat;
  vector[dims(gamma)[1]] lampda;
  matrix[dims(gamma)[1]-1,dims(gamma)[1]-1] o_jj;
  matrix[dims(gamma)[1]-1,1] o_ji;
  matrix[1,dims(gamma)[1]-1] o_ij;
  matrix[dims(gamma)[1]-1,1] l_j;
  int h;
  h<-dims(gamma)[1];
  omega_hat<-diag_matrix(1.0 ./ w) * beta * diag_matrix(1.0 ./ w);
  for(i in 1:h) {
    if(i==1) {
      o_jj<-block(omega_hat,2,2,h-1,h-1);
      o_ji<-block(omega_hat,2,1,h-1,1);
      o_ij<-block(omega_hat,1,2,1,h-1);
      l_j<-block(to_matrix(gamma),2,1,h-1,1);
    }
    if(i==h) {
      o_jj<-block(omega_hat,1,1,h-1,h-1);
      o_ji<-block(omega_hat,1,h,h-1,1);
      o_ij<-block(omega_hat,h,1,1,h-1);
      l_j<-block(to_matrix(gamma),1,1,h-1,1);
    }
    if(i>1 && i<h) {
      o_jj<-append_row(append_col(block(omega_hat,1,1,i-1,i-1),
        block(omega_hat,1,i+1,i-1,h-i)),
        append_col(block(omega_hat,i+1,1,h-i,i-1),
        block(omega_hat,i+1,i+1,h-i,h-i)));
      o_ji<-append_row(block(omega_hat,1,i,i-1,1),
        block(omega_hat,i+1,i,h-i,1));
      o_ij<-append_col(block(omega_hat,i,1,1,i-1),
        block(omega_hat,i,i+1,1,h-i));
      l_j<-append_row(block(to_matrix(gamma),1,1,i-1,1),
        block(to_matrix(gamma),i+1,1,h-i,1));
    }
    o_jj<-o_jj - (o_ji * (1/omega_hat[i,i])) * o_ij;
    lampda[i]<-(gamma[i]+(1/omega_hat[i,i]))*(o_ij*l_j)[1,1])/sqrt(1+(l_j'*o_jj*l_j)[1,1]);
  }
  return(lampda);
}
}

```

FIGURA 45: Código da função *marginal* utilizada para o cálculo das distribuições marginais do modelo probabilístico dinâmico multivariado, programado em linguagem Stan

FONTE: A autora (2017)