



**Simpósio de Métodos
Numéricos em Engenharia**

25 a 27 de outubro, 2017

Simulação Monte Carlo via Cadeias de Markov Aplicada à Previsão Hidrológica Probabilística

Alana Renata Ribeiro, Maurício Felga Gobbi
PPGMNE
UFPR
Curitiba, Brasil
alanar89@gmail.com, gobbi@ufpr.br

Eduardo Alvim Leite
SIMEPAR
Curitiba, Brasil
alvim@simepar.br

Resumo—O objetivo deste trabalho é desenvolver um método probabilístico de previsão de vazão com o uso dos algoritmos de amostragem de Monte Carlo via Cadeias de Markov e dos conceitos da inferência bayesiana. Este método avalia o comportamento dos erros resultantes de um modelo de previsão determinístico aplicado à sub-bacia de São Bento, no estado do Paraná. Comparações entre os valores das previsões dos modelos probabilístico e determinístico são realizadas a fim de quantificar as melhorias na qualidade das previsões. Além disso, apresenta-se como resultado um sistema dinâmico de visualização destas previsões.

Palavras-chave—vazão; previsões determinísticas; previsões probabilísticas

I. INTRODUÇÃO

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, nem sempre tem sido suficiente para suprir tais necessidades, com isso abordagens probabilísticas tem angariado cada vez mais espaço na área hidrológica.

Este estudo tem como objetivo propor um modelo probabilístico de previsão de vazão construído com algoritmos de amostragem MCMC (Monte Carlo via Cadeias de Markov). Informações a priori sobre os parâmetros de distribuições probabilísticas são combinadas às informações extraídas dos erros resultantes de um

modelo determinístico, e por meio das distribuições posteriores dos parâmetros, para os quais desejam-se extrair inferências, são feitas as previsões probabilísticas de vazão a curto prazo.

Um sistema de visualização dos resultados das previsões probabilísticas, com suas bandas de incerteza, é apresentado a fim de facilitar a análise dos resultados obtidos.

II. INFERÊNCIA BAYESIANA

A inferência bayesiana [7], considera informações prévias à coleta dos dados (informações à priori) sobre os parâmetros (θ), além da informação contida nos dados (Q). Ambas as informações, distribuição de probabilidade à priori ($p(\theta)$) e a função de verossimilhança que resume o modelo para os dados ($P(Q|\theta)$), são combinadas usando o teorema de Bayes (1) gerando a distribuição de probabilidade à posteriori ($P(\theta|Q)$) sobre os parâmetros, com $P(Q)$ constante de proporcionalidade.

$$P(\theta|Q) = \frac{p(\theta)P(Q|\theta)}{P(Q)} \quad (1)$$

Segundo [9] os métodos bayesianos são muito atraentes para uso em problemas hidrológicos pois incorporam conhecimento prévio de um especialista da área. Porém, o esforço computacional necessário para implementar os procedimentos dificulta sua aplicação. Este esforço é devido à dificuldade em avaliar momentos e outras

estatísticas de interesse da distribuição à posteriori, que é uma função multidimensional expressa em termos dos parâmetros (θ) desconhecidos do modelo, e pode assumir uma forma que nem sempre é fácil de integrar.

O método de amostragem de Monte Carlo via Cadeia de Markov (*Markov chain Monte Carlo* - MCMC) soluciona as dificuldades computacionais. O objetivo é gerar amostras dos valores dos parâmetros da distribuição à posteriori simulando um processo aleatório que tem a distribuição à posteriori como sua distribuição estacionária.

Existem diversos tipos de algoritmos de amostragem MCMC, dos quais Metropolis-Hastings [8] e Gibbs [6] são os mais populares. Outro algoritmo conhecido como HMC (Hamiltonian Monte Carlo ou Hybrid Monte Carlo) [4] 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 [1], 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. Para o desenvolvimento deste estudo optou-se por aplicar este algoritmo por meio da linguagem de modelagem probabilística Stan do software R.

III. MATERIAIS E MÉTODOS

Modelos hidrológicos determinísticos conceituais são baseados na representação dos processos físicos que envolvem o fenômeno estudado. Resultados, obtidos pelo Sistema Meteorológico do Paraná (SIMEPAR), de um modelo determinístico do tipo conceitual, chuva-vazão, chamado de *Rainfall-Runoff-Routing* (3R) [11], são utilizados como informações à priori para os modelos probabilísticos.

Para a aplicação do modelo determinístico e posteriormente do probabilístico, são utilizados dados de precipitação e vazão da sub-bacia de São Bento pertencente à bacia do rio Iguaçu, na cidade da Lapa no estado do Paraná, com área de drenagem de 2012km² e vazão média de 34,9m³/s, monitorada pelo SIMEPAR (Sistema Meteorológico do Paraná).

O modelo determinístico 3R foi previamente calibrado com 13 anos de dados entre 1998 e 2010, realizou previsões de até 72h, de 6h em 6h (totalizando 12 horizontes de previsão), em 3 anos de dados entre meados de 2013 e meados de 2016. Este período foi escolhido devido à disponibilidade de dados de precipitação previstos que foram utilizados em conjunto com os dados de precipitação e vazão observados.

Como entrada para o modelo probabilístico utilizou-se os erros ($Q_o - Q_p$, diferença entre vazão observada e vazão prevista) cometidos pelo modelo 3R entre meados de 2013 e meados de 2015 (período de calibração). E para aplicação e validação dos resultados utilizou-se os dados do período de meados de 2015 à meados de 2016.

A Fig. 1 apresenta as estimativas da distribuição de probabilidade (estimadas por meio do método *gaussian kernel density estimation* [12]) dos erros do modelo determinístico 3R para os diferentes horizontes de previsão. Nota-se a presença de deslocamentos, assimetrias, e distanciamento da distribuição normal. Pode-se observar também que a distribuição dos erros positivos difere-se da dos erros negativos.

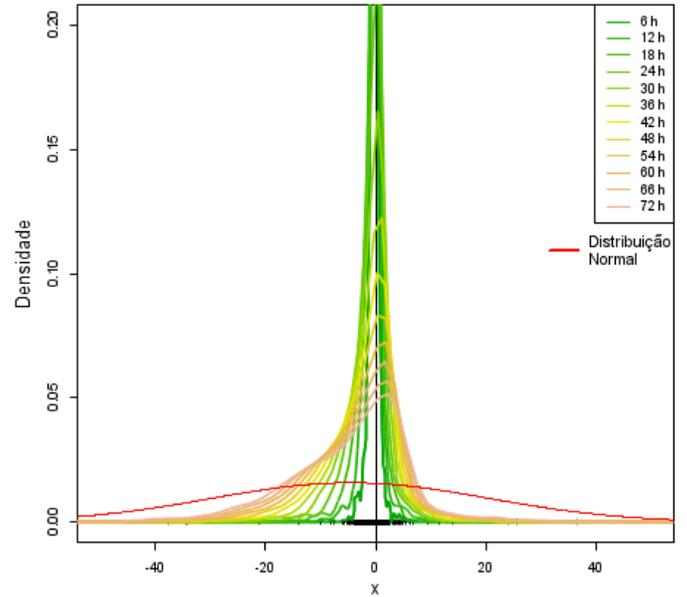


Figura 1: Estimativa das densidades dos erros do modelo determinístico 3R para todos os horizontes de previsão

As Fig. 2(a) e 2(b) apresentam as correlações, auto-correlações e correlações cruzadas entre os erros do modelo determinístico 3R para os diferentes horizontes de previsão, demonstrando uma forte estrutura no comportamento dos erros entre horizontes de uma mesma previsão [0, 24 a 0, 99], bem como a existência de correlação entre os erros de diferentes previsões [-0, 32 a 0, 78].

Devido ao comportamento assimétrico da densidade (Fig. 1), e das correlações existentes entre os erros nos diferentes horizontes de previsão do modelo determinístico (Fig. 2), a calibração do modelo probabilístico de previsão de vazão utiliza a distribuição de probabilidade multivariada *M-Skew-t* [2], cuja distribuição marginal em cada horizonte é a *Skew-t* univariada.

A função *M-Skew-t* permite uma grande flexibilidade no ajuste das probabilidades pois é uma generalização das distribuições do tipo Normal, Cauchy, *t-Student*, *Skew-Normal* e *Skew-Cauchy*. Sua utilização exige o ajuste dos parâmetros de localização (ξ), dispersão (Ω), inclinação (λ) e curtose (ν), ajustando a distribuição dos erros de maneira completa.

Um modelo probabilístico multivariado de previsão de vazão é construído de forma dinâmica, ou seja, com perturbações dos parâmetros da distribuição *M-Skew-t* utilizando um conjunto de 32 preditores que dependem do estado hidrometeorológico da sub-bacia, das previsões determinísticas de chuva e vazão, e de uma sequência recente de erros observados dos modelos determinísticos.

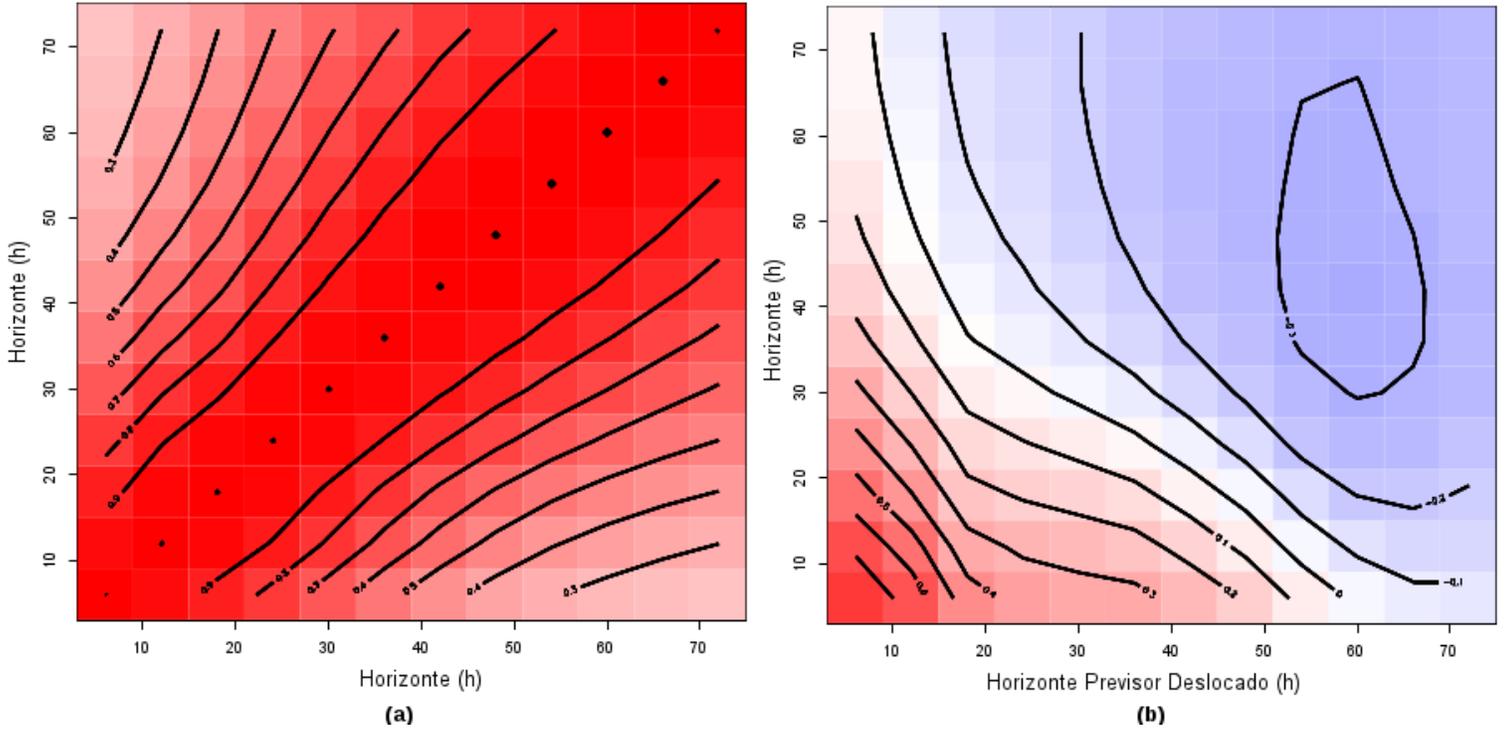


Figura 2: Correlações do modelo determinístico 3R. (a) Correlação entre os erros de acordo com o horizonte de previsão. (b) Correlação cruzada entre erros com deslocamentos temporais

Uma transformação *logit* [3] é aplicada aos dados de vazão observada e prevista a fim de que as previsões não ultrapassem os valores máximo e mínimo de vazão observados em todo o histórico de dados.

As etapas do desenvolvimento do modelo probabilístico são:

$x = l, d, i, c.$

$$\begin{aligned} \text{logit}(Q_o) \sim \\ M - \text{Skew} - t(fl(t) + \xi + \text{logit}(Q_p), fd(t) * \Omega, fi(t) + \lambda, \\ fc(t) * \nu) \end{aligned} \quad (3)$$

- Calibração de um modelo probabilístico estático multivariado que não depende do conjunto de preditores (2);

$$\text{logit}(Q_o) \sim M - \text{Skew} - t(\xi + \text{logit}(Q_p), \Omega, \lambda, \nu) \quad (2)$$

- Geração dos valores individuais de perturbação no momento de previsão t : $ppx(t)[ppl(t), ppd(t), ppi(t), ppc(t)]$, dependentes dos preditores, para o período de calibração otimizando o modelo para cada momento de previsão;
- Desenvolvimento de 4 modelos regressivos do tipo *Generalized Linear Models* (GLM) [10] para previsão das perturbações baseadas nos preditores apresentados (uso da técnica de *stepwise* [5] para eliminação de preditores não significativos);
- Geração das estimativas dos valores de perturbação a partir da aplicação dos modelos regressivos desenvolvidos;
- Calibração do modelo probabilístico dinâmico multivariado; como em (3), em que $f_x(t) = GLM(f_{x_1}, f_{x_2}, ppx(t))$, e

IV. RESULTADOS

A avaliação do desempenho do modelo probabilístico é realizada por meio da análise de sua verossimilhança como (4), que representa o somatório do logaritmo da probabilidade posterior total dos valores observados.

$$LPT = \sum_{i=1}^n \log(pdf(x_i)) \quad (4)$$

A Tabela I apresenta os valores de LPT para o modelo probabilístico aplicado ao período de validação do modelo 3R, comparando as utilizações da distribuição normal e *M-Skew-t* para o modelo estático, sem a utilização das perturbações por preditores, e da distribuição *M-Skew-t* para o modelo dinâmico, com a utilização das perturbações por preditores.

Tabela I: LPT PARA OS MODELOS PROBABILÍSTICOS ESTÁTICO E DINÂMICO NO PERÍODO DE VALIDAÇÃO DO MODELO 3R

Modelo	Modelos Probabilísticos			
	Determinístico	Normal Estático	<i>M-Skew-t</i> Estático	<i>M-Skew-t</i> Dinâmico
3R		-19508	-8729	-7419

Os resultados demonstram um aumento, e portanto uma melhoria significativa nos valores de LPT obtidos com a utilização da distribuição *M-Skew-t*, bem como com a dinâmica obtida com o uso do conjunto de preditores perturbadores dos parâmetros do modelo probabilístico. A aplicação do teste estatístico *Vuong-test* [13] para os modelos não aninhados (com diferentes distribuições de probabilidades) aponta que a utilização da distribuição *M-Skew-t* é preferível à distribuição Normal com mais de 95% de confiança. O teste da razão da verossimilhança [13] para modelos aninhados (com mesma distribuição de probabilidade *M-Skew-t*) demonstra que o modelo dinâmico é preferível ao estático também com mais de 95% de confiança.

A Fig. 3 apresenta os valores de $LPTH$ (LPT por horizonte de previsão) para o modelo probabilístico com distribuição *M-Skew-t*, destacando uma melhoria em média de 13%, em se utilizar uma abordagem dinâmica no lugar da estática, sendo que maiores melhorias são relacionadas aos menores horizontes de previsão.

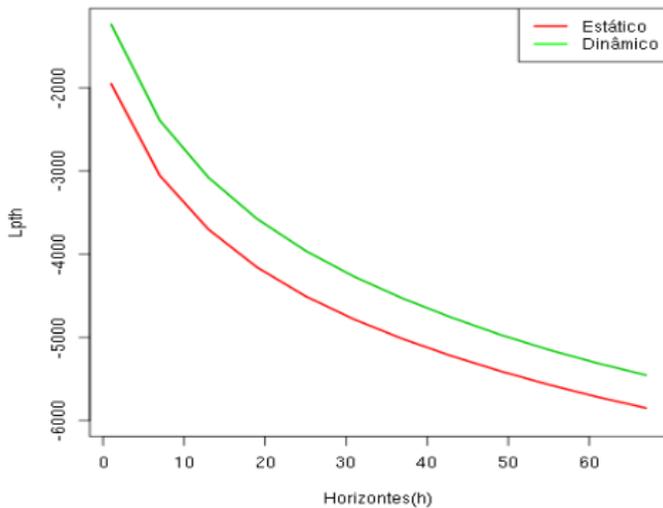


Figura 3: Comparações: $LPTH$ do modelo probabilístico estático e dinâmico com distribuições *M-Skew-t*

Outra avaliação do desempenho do modelo pode ser feita por meio da derivação da previsão determinística a partir da moda da distribuição posterior, no período de validação do modelo dinâmico multivariado com distribuição *M-Skew-t* aplicada ao modelo 3R. A Fig. 4 apresenta a comparação do erro médio quadrático por horizonte de previsão entre o modelo probabilístico e determinístico, destacando uma melhoria em média de 29%, em se utilizar um modelo probabilístico derivado de um determinístico.

Um sistema dinâmico de previsões probabilísticas foi construído para visualização dos resultados estáticos e dinâmicos. As Fig. 5(a,

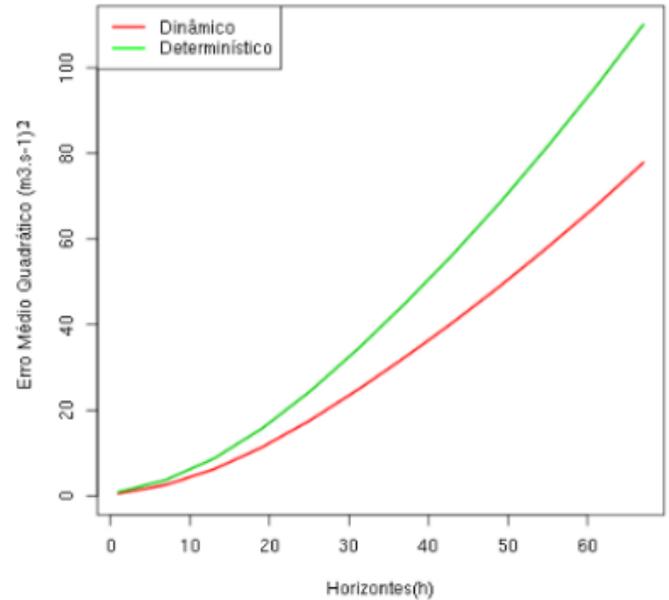


Figura 4: Comparações: Erro médio quadrático por horizonte de previsão dos modelos probabilístico dinâmico e determinístico 3R

b, c e d) apresentam quatro exemplos de diferentes momentos de previsões probabilísticas para o modelo 3R. Os valores de vazão observados, de vazão previstos pelo modelo determinístico, previsões pelos modelos probabilísticos estático e dinâmico com seus valores de LPT , são apresentados simultaneamente proporcionando ao usuário do sistema uma visão completa do cenário para todos os horizontes de previsão.

As cheias observadas nas Fig. 5(a, b, c e d) estão completamente previstas pela faixa de incerteza contruída com o modelo probabilístico, demonstrando sua eficiência de previsão em casos de possíveis enchentes. Os valores de LPT para estes quatro casos demonstram a superioridade do modelo dinâmico sobre o modelo estático, fato que ocorre na maioria dos momentos de previsão.

V. CONCLUSÕES

Neste estudo duas abordagens multivariadas (considerando 12 horizontes de previsão correlacionados) foram desenvolvidas e avaliadas através de comparações entre os valores máximos dos logaritmos das distribuições de probabilidade posteriores obtidas. A abordagem estática foi testada com a distribuição de probabilidade normal multivariada e com a distribuição *M-Skew-t*. Seus resultados demonstraram grande vantagem em se utilizar a segunda distribuição.

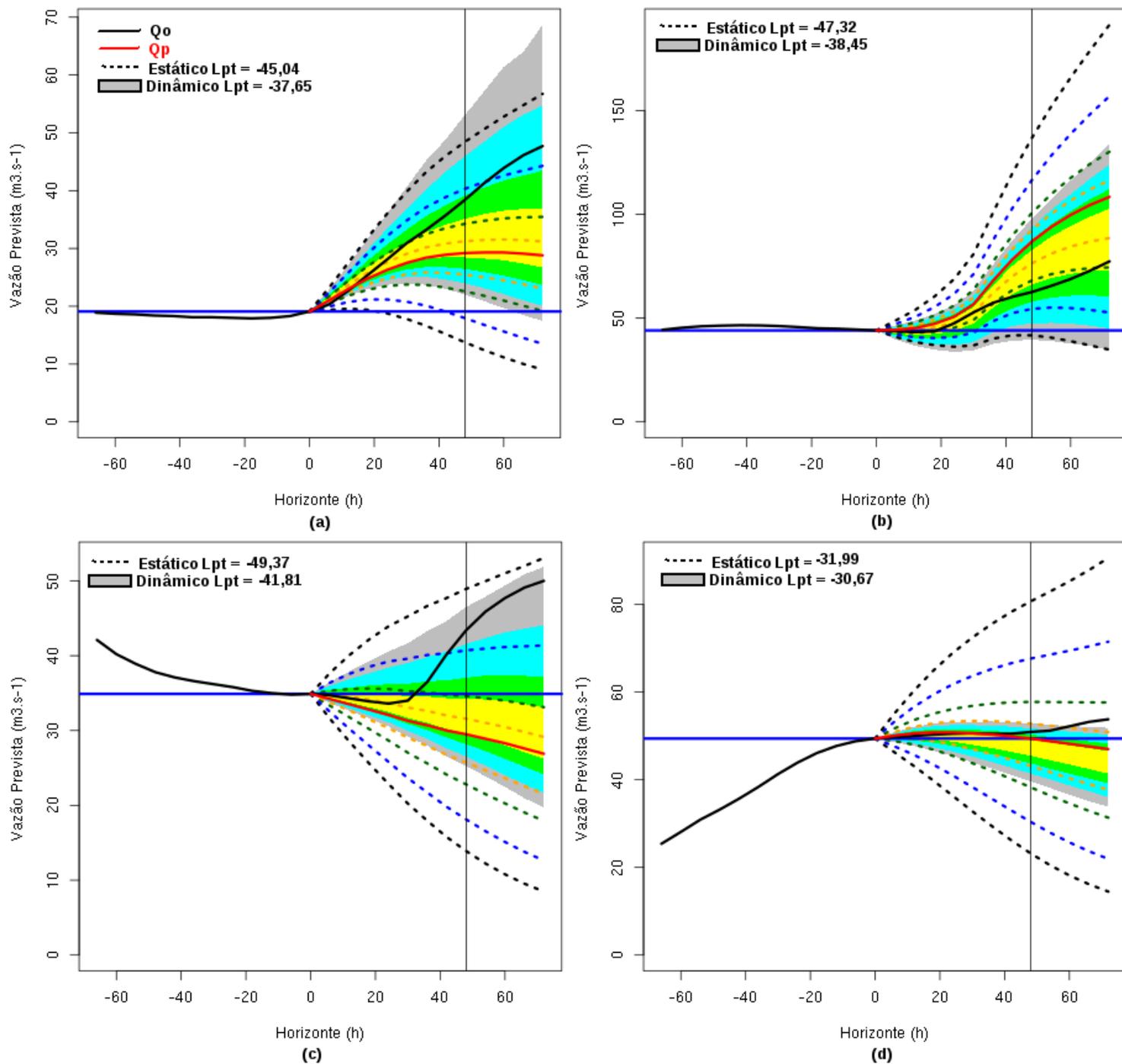


Figura 5: Visualização dinâmica para quatro momentos de previsão dos modelos probabilísticos estático e dinâmico aplicados ao modelo determinístico 3R

A abordagem dinâmica que utiliza perturbações dos parâmetros por preditores foi testada com a distribuição de probabilidade *M-Skew-t*.

O resultado da aplicação demonstrou uma melhoria de aproximadamente 62% com a utilização da distribuição *M-Skew-t* substituindo

a distribuição Normal, e em torno de 15% de melhoria com a abordagem dinâmica no lugar da estática. Além disso, possibilitou-se a construção de um sistema de visualização das previsões probabilísticas em torno das previsões determinísticas, e por meio das comparações dos resultados notou-se uma melhoria de até 29% no erro médio

quadrático principalmente para os maiores horizontes de previsão.

Pode-se observar ainda que mesmo em casos de difícil previsão como subidas ou descidas repentinas, e com os valores previstos pelo modelo determinístico muito diferentes dos valores de vazão observados, a região que representa a probabilidade de ocorrência das vazões foi, em muitos casos, capaz de envolver os valores futuros observados.

AGRADECIMENTOS

Os autores agradecem a CAPES - Coordenação para a Melhoria do Pessoal de Educação Superior - pelo apoio financeiro dado a essa pesquisa.

REFERÊNCIAS

- [1] M. Alfaki, Improving efficiency in parameter estimation using the Hamiltonian Monte Carlo algorithm, Dissertação de Mestrado, University of Bergen, 2008.
- [2] R. B. Arellano-Vale and M. G. Genton, Multivariate extended skew-t distributions and related families, *Metron*, 68:201–234, 2010. DOI:10.1007/BF03263536
- [3] J. S. Cramer, The origins and development of the logit model, *Disponível em: http://www.cambridge.org/resources/0521815886/1208_default.pdf*, 2017.
- [4] S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth, Hybrid monte carlo, *Physics Letters B*, 195:216–222, 1987. DOI: 10.1016/0370-2693(87)91197-X
- [5] M. A. Efronson, “Multiple Regression Analysis,” In: A. Ralston and H. S. Wilf, Eds., *Mathematical Methods for Digital Computers*, John Wiley, New York, 1960.
- [6] A. E. Gelfand and A. F. M. Smith, Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, 85:398–409, 1990. DOI: 10.2307/2289776
- [7] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall, New York, 2003.
- [8] W. K. Hastings, Monte carlo sampling methods using markov chains and their applications, *Biometrika*, 57:97–109, 1970.
- [9] L. Marshall, D. Nott, A. Sharma, A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, *Water Resources Research*, 40, 2004. DOI:10.1029/2003WR002378.
- [10] J. A. Nelder, R. W. M. Wedderburn, Generalized Linear Models, *Journal of the Royal Statistical Society*, 135:370–384, 1972.
- [11] R. K. Neto, A. K. Guetter, M. R. M. Mine, Modelo hidrológico com atualização de estado parte i: Modelagem hidrológica, *Revista Brasileira de Recursos Hídricos*, 12:93–103, 2007. DOI: 10.21168/rbrh.v12n3.p93-103
- [12] E. Parzen, On estimation of a probability density function and mode, *Annals of Mathematical Statistics*, 33:1065–1076, 1962.
- [13] Q. H. Vuong, Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses, *Econometrica*, 57:307–333, 1989.