



**Quantificação das Incertezas em Curvas de
Descarga de Postos Fluviométricos Utilizando Inferência Bayesiana**

Uncertainty Quantification in Discharge Curves of Fluviometric Stations Using Bayesian Inference

Alana Renata Ribeiro¹; Maurício Felga Gobbi¹; Eduardo Alvim Leite² & Mariana Kleina³

¹Universidade Federal do Paraná, Programa de Pós-graduação em Métodos Numéricos em Engenharia,
Av. Coronel Francisco Heráclito dos Santos, Centro Politécnico, Bloco LAME/CESEC, 19011, 81531-980, Curitiba, PR, Brasil

²Sistema Meteorológico do Paraná (SIMEPAR),

Av. Coronel Francisco Heráclito dos Santos, Centro Politécnico, 19100, 81531-980, Curitiba, PR, Brasil

³Universidade Federal do Paraná, Departamento de Engenharia de Produção,

Av. Coronel Francisco Heráclito dos Santos, Centro Politécnico, 19011, 81531-980, Curitiba, PR, Brasil

E-mails: alanar89@gmail.com; gobbi@ufpr.br; alvim@simepar.br; marianakleina11@gmail.com

Recebido em: 16/05/2017 Aprovado em: 03/07/2017

DOI: http://dx.doi.org/10.11137/2017_2_266_277

Resumo

O objetivo deste artigo é apresentar a construção de uma região de confiabilidade probabilística em torno de curvas de descarga previamente definidas para diferentes postos fluviométricos, com o intuito de considerar as incertezas na obtenção de valores de vazão fluviométrica calculados por meio da utilização destas curvas. Modelos Bayesianos com algoritmos de amostragem MCMC (Monte Carlo via Cadeias de Markov) são construídos e aplicados às distribuições de probabilidade dos erros obtidos com a comparação de valores de vazão fluviométrica previstas (resultantes da aplicação das curvas de descarga) e observadas (por meio de métodos convencionais). Para este estudo registros de três postos hidrológicos monitorados pela COPEL (Companhia Paranaense de Energia) e de dez postos hidrológicos monitorados pela CEMIG (Companhia Energética de Minas Gerais) foram analisados e utilizados. Os resultados demonstraram que a abordagem Bayesiana revelou-se adequada aos objetivos propostos, permitindo a construção da região de confiabilidade e, com ela, a avaliação das incertezas associadas à utilização das curvas de descarga.

Palavras-chave: Dados fluviométricos; confiabilidade probabilística; MCMC

Abstract

The aim of this paper is to present the construction of a probabilistic reliability region around rating curves previously defined for different fluviometric stations, in order to consider the uncertainty in obtaining fluviometric discharge values calculated by using these curves. Bayesian models with MCMC sampling algorithms (Monte Carlo Markov Chain) are constructed and applied to the error probability distributions obtained from the comparison of predicted fluviometric discharge values (resulting from the application of the rating curves) and observed (by conventional methods). For this study, records of three hydrological stations monitored by COPEL, and ten hydrological stations monitored by CEMIG, were analyzed and used. The results showed that the Bayesian approach has proved successful for the proposed objectives, allowing the construction of reliability region, and with it, the evaluation of the uncertainties associated with the use of rating curves.

Keywords: Fluviometric data; probabilistic reliability; MCMC

1 Introdução

Incertezas com relação aos valores de vazão fluviométrica 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 (Di Baldassarre & Montanari, 2009; McMillan *et al.*, 2010; Domeneghetti *et al.*, 2012), além de analisar seus possíveis efeitos sobre problemas de modelagem hidrológica (Montanari & Brath, 2004; Montanari & Grossi, 2008).

Estabelecer uma relação cota-vazão, conhecida como curva de descarga (Rantz, 1982; Schmidt, 2002), é a maneira mais comum e simples de se monitorar dados de vazão em uma estação hidrológica. Esta relação, independentemente pré-estabelecida para cada estação hidrológica de acordo com suas características individuais, permite que se convertam valores de cota (nível da água em um rio), continuamente registrados, em séries temporais de valores de vazão.

A precisão e a estabilidade de uma curva de descarga, também conhecida como curva-chave, dependem das condições hidráulicas prevalentes no local, do conhecimento dos processos físicos que relacionam cota e vazão no local determinado, e da disponibilidade e precisão das taxas individuais, ou seja, de um conjunto direto e determinado de observações de cota e de vazão (Le Coz *et al.*, 2014). Contudo, avaliar as incertezas associadas à curva-chave, durante um determinado período de tempo e para determinadas condições hidráulicas, tem sido tema de diversos estudos (Overleir, 2004; Moyeed & Clarke, 2005; McMillan *et al.*, 2010; Domeneghetti *et al.*, 2012), e é o objetivo deste artigo.

Segundo Le Coz *et al.* (2014) a aplicação de inferência Bayesiana e simulações de Monte Carlo via Cadeia de Markov (MCMC) permitiram que novas soluções para o problema da análise da classificação de incertezas nas curvas de descarga fossem encontradas nos últimos anos. Esta abordagem apresentou vantagens decisivas na resolução do problema, enquanto os custos computacionais permaneceram baixos.

O conhecimento prévio de séries históricas de dados de cota e vazão, das curvas-chave, e de características das estações hidrológicas estudadas, pode ser utilizado de forma explícita como informações para a construção de distribuições *a priori* dos parâmetros das funções assumidas pelo método MCMC. É possível, além disso, determinar uma função de verossimilhança que responde pela incerteza em classificações individuais. Com base nas incertezas correspondentes, a melhor informação a partir de observações (funções de verossimilhança) e modelo (informações *a priori*) será utilizada para produzir os resultados sob a forma de uma distribuição *a posteriori*. Além disso, como a técnica é diretamente baseada em funções de densidade de probabilidade, a análise de incerteza pode ser realizada de uma forma simples e clara utilizando percentis da distribuição *a posteriori*.

2 Materiais e Métodos

2.1 Curvas de Descarga

Para se determinar a vazão de um rio ao longo do tempo, estima-se para este rio uma curva-chave, ou curva de descarga (Figura 1), expressão que relaciona o nível do rio numa seção com sua respectiva vazão (Rantz, 1982).

A obtenção desta expressão necessita de medições de vazão para diversos níveis. No entanto, medir apenas o nível é muito mais fácil e barato que se utilizar qualquer um dos recursos existentes para medir a vazão. Contudo, medi-la ao menos uma vez para certo intervalo de níveis faz-se necessário para que se defina assim uma curva de descarga por meio de equações definidas, tabulações, ou de interpolações dos pares cota-vazão determinados (pontos na Figura 1).

Uma interpolação dos pontos dados pelos pares cota-vazão determina esta relação e, como esta operação não contempla todos os níveis possíveis, pode-se utilizar ainda extrapolações.

A relação biunívoca cota-vazão de um rio se mantém ao longo do tempo desde que as características geométricas do mesmo não sofram grandes variações.

Duas hipóteses hidrodinâmicas para que a curva-chave exista são: o escoamento deve ser

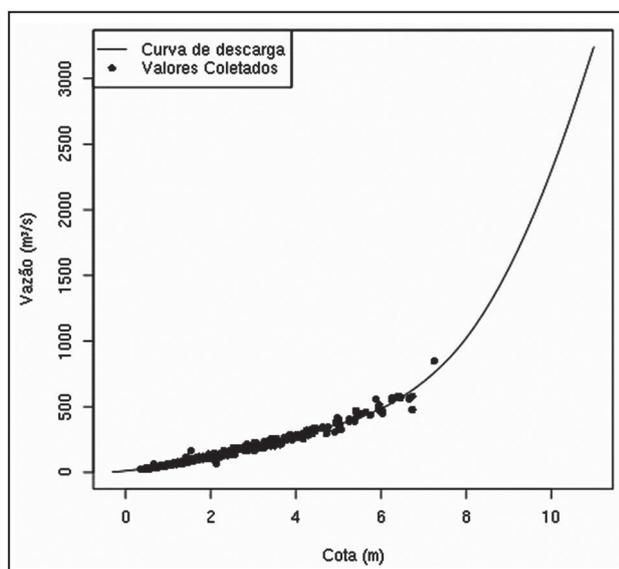


Figura 1 Curva de descarga: exemplo de uma relação cota-vazão.

uniforme, ou seja, a velocidade da vazão varia lentamente ao longo do rio e isso geralmente ocorre quando a seção transversal do rio é gradualmente variada; além disso, o escoamento deve ser aproximadamente permanente, ou seja, o nível e a velocidade da vazão variam lentamente com o passar do tempo, o que não ocorre, por exemplo, durante a passagem de uma onda de cheia.

Alterações na geometria da seção ou na declividade do rio geradas por erosões ou assoreamento ao longo do tempo causam mudanças na velocidade do escoamento e nas relações entre área, raio hidráulico e profundidade, afetando a relação cota-vazão.

As etapas para a obtenção de uma curva-chave podem ser resumidas como se segue:

- Instalação do mecanismo de medição de nível numa seção (régua, etc.);
- Medição da vazão por meio de diferentes recursos, para vários níveis observados pelo mecanismo instalado anteriormente;
- Determinação dos parâmetros da curva-chave;
- Monitoramento do nível.

Porém as aproximações, as violações das hipóteses, e até mesmo os erros envolvidos nas medições podem fazer com que o valor de vazão relacionado ao valor de cota contenha erros inerentes,

e por consequência há a inserção de incertezas em torno das curvas de descarga definidas para os diferentes postos de monitoramento hidrológico.

2.2 Inferência Bayesiana

Segundo Bonat *et al.* (2012) a inferência estatística, de maneira geral, tem como objetivo fornecer subsídios, de forma 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, 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 (Gelman *et al.*, 2003).

O teorema de Bayes (Equação 1) fornece uma expressão para a probabilidade condicional de A dado B.

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

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

Na equação (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, tem-se a Equação 2.

$$[\theta|y] = [y|\theta] [\theta]/[y] \quad (2)$$

Em que $[\theta]$ é a distribuição conjunta *a priori* para o vetor de parâmetros θ independentemente de **y**, $[y|\theta]$ representa a verossimilhança de **y**, $[\theta|y]$ é a distribuição *a posteriori* conjunta para o vetor de parâmetros θ , e o denominador (Equação 3) define a verossimilhança marginal **y**.

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

Porém, em algoritmos como os de MCMC 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 a Equação 4.

$$[\theta|y] \propto [y|\theta][\theta] \quad (4)$$

O produto entre a verossimilhança $[y|\theta]$ e a *a priori* $[\theta]$ (Equação 4) representa a distribuição conjunta não normalizada dos parâmetros. Esta *posteriori* não possui expressão definida para a maioria dos problemas, assim se utilizam aproximações analíticas, ou numéricas, ou simulação estocástica para se 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 de Monte Carlo via Cadeias de Markov (MCMC).

Existem diversos tipos de algoritmos chamados de MCMC, entre eles o Metropolis-Hastings (Metropolis *et al.*, 1953; Hastings, 1970), e o amostrador de Gibbs (Geman & 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 objetivo (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) 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, para o desenvolvimento deste estudo opta-se por aplicar este algoritmo às incertezas em torno das curvas-chave.

2.3 Linguagem Computacional

Métodos de inferência Bayesianos são, em geral, implementados através de métodos computacionalmente intensivos, exigindo a utilização de softwares específicos para este fim.

Neste estudo faz-se uso do software estatístico R (R Core Team, 2012) para processamento dos dados previamente utilizados para a construção dos modelos Bayesianos, e para processamento das inferências obtidas após a aplicação destes modelos. Além disso, para a construção dos modelos propriamente ditos, 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.

2.4 Regiões de Estudo

Para o desenvolvimento dos estudos, neste artigo, registros de campanhas realizadas em três postos hidrológicos monitorados pela COPEL (utilizados no desenvolvimento do período de calibração dos modelos), e de dez postos hidrológicos monitorados pela CEMIG (utilizados no desenvolvimento do período de aplicação dos modelos) foram analisados e utilizados. A Tabela 1 apresenta um resumo das informações relevantes de cada um destes postos de monitoramento hidrológico, e as Figuras 2 e 3 apresentam a localização destes postos nos estados do Paraná e de Minas Gerais, respectivamente.

Os nomes de cada um dos postos hidrológicos utilizados são apresentados na segunda coluna da Tabela 1, e para simplificar, no decorrer do texto serão substituídos pelas siglas correspondentes, presentes na primeira coluna da tabela. O código ANA, por sua vez, permite que cada posto seja identificado, entre os usuários de recursos hídricos, em um cadastro no Sistema de Informações Hidrológicas – HIDRO, da Agência Nacional de Águas, que disponibiliza informações relevantes relativas aos postos hidrológicos de todo o território nacional, bem como dados coletados por estes postos. As localizações geográficas referentes à Latitude e Longitude são apresentadas na quinta e sexta colunas respectivamente. E, por fim, a última coluna da Tabela 1 apresenta a quantidade de pares cota-vazão utilizados no estudo para cada um dos postos, esses pares são resultados de campanhas de medição, em que valores de cota e vazão fluviométricos foram coletados de maneira convencional, sem a utilização de curvas de descarga, ou qualquer outro método de cálculo automático.

Quantificação das Incertezas em Curvas de Descarga de Postos Fluviométricos Utilizando Inferência Bayesiana
 Alana Renata Ribeiro; Maurício Felga Gobbi; Eduardo Alvim Leite² & Mariana Kleina

N.	Nomes	Código ANA	Entidade Responsável	Lat.	Long.	Quantidade de Dados das Campanhas
C1	Divisa	26055019	COPEL	-26,0833	-50,3166	1120
C2	Fluviópolis	26025035	COPEL	-26,0333	-50,5833	590
C3	Porto Amazonas	25334953	COPEL	-25,55	-49,8833	617
P1	Fazenda Ouro Fino	56776500	CEMIG	-19,168	-42,834	104
P2	Poço Fundo – Piedade	61566000	CEMIG	-21,738	-45,958	18
P3	Ponte GO-206	60680004	CEMIG	-18,478	-49,616	204
P4	Porto Tiradentes	61107000	CEMIG	-21,121	-44,233	43
P5	Vila Terra Branca – Jusante	54010005	CEMIG	-17,315	-43,213	123
P6	Fazenda Boa Vista – MG	60228000	CEMIG	-19,693	-47,422	108
P7	Juiz de Fora – Jusante	58480500	CEMIG	-21,778	-43,326	101
P8	Ponte BR-040	90000002	CEMIG	-18,109	-45,461	58
P9	Porto das Andorinhas	40100001	CEMIG	-19,281	-45,289	39
P10	São Sebastião da Encruzilhada	56990000	CEMIG	-19,493	-41,162	8

Tabela 1 Informações sobre os postos de monitoramento hidrológico.

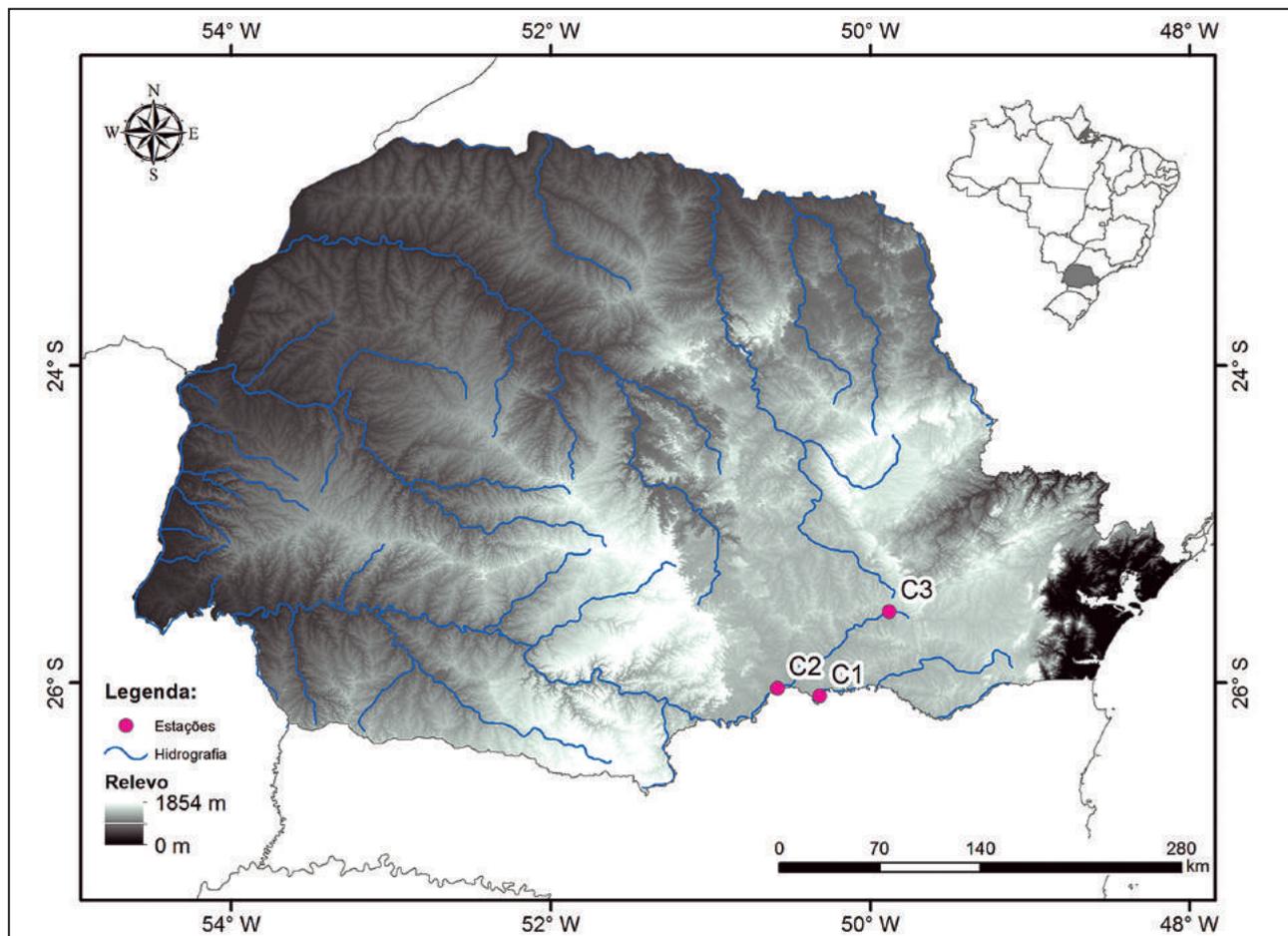


Figura 2 Localização dos postos de monitoramento hidrológicos utilizados no desenvolvimento do período de calibração dos modelos.

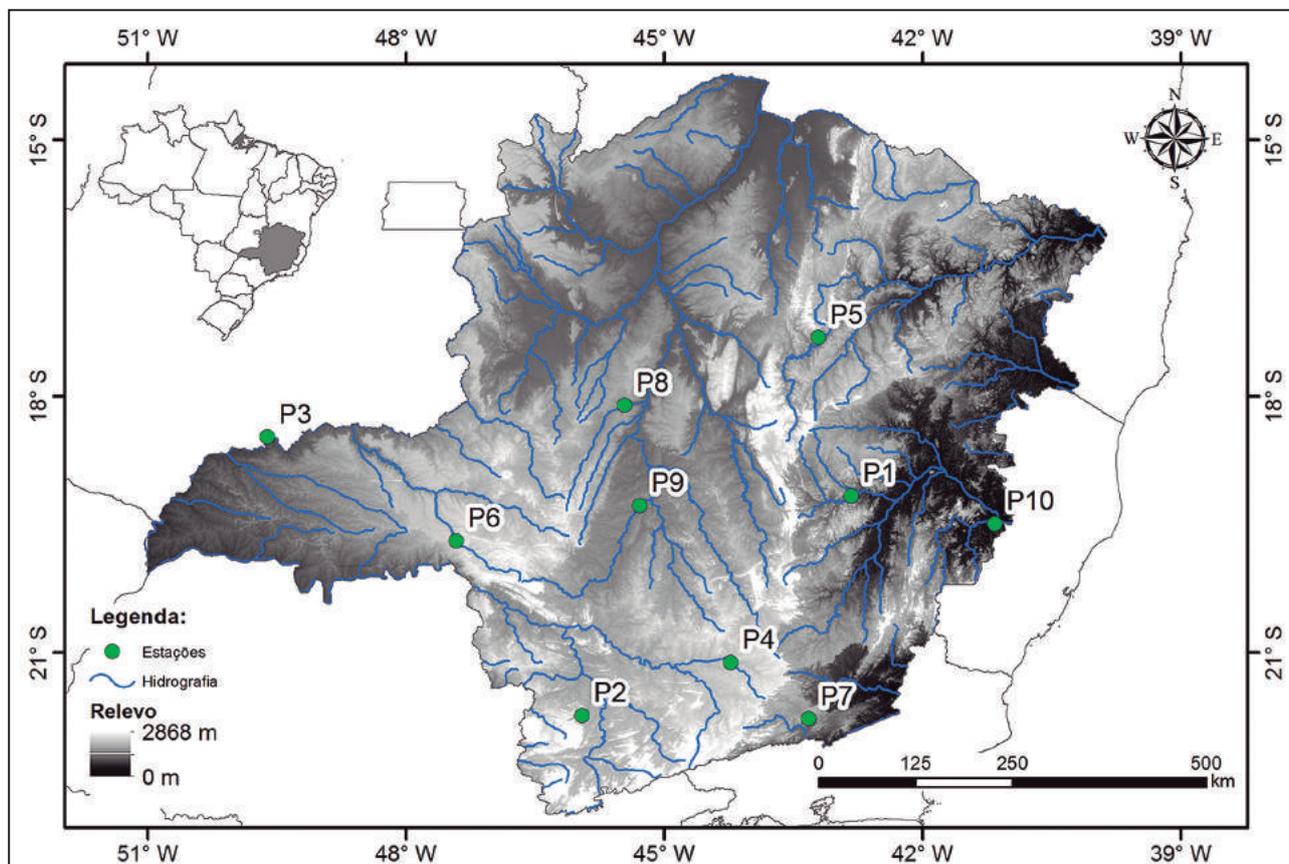


Figura 3 Localização dos postos de monitoramento hidrológicos utilizados no desenvolvimento do período de aplicação dos modelos.

3 Projeto de Aplicação

Sob um cenário comum ao de Silva *et al.* (2014), para que o modelo desenvolvido seja útil em situações hidrológicas práticas, ele deve ser capaz de reproduzir as vazões observadas em uma seção fluviométrica de interesse, com suficiente grau de acurácia.

Neste caso o modelo desenvolvido deve ser capaz de determinar regiões de confiabilidade probabilística em torno das curvas-chave previamente definidas para os dez postos fluviométricos monitorados pela CEMIG, levando em consideração as informações obtidas por meio dos pares cota-vazão previamente medidos.

Porém, devido à escassez de dados monitorados para os postos escolhidos, principalmente em situações de cheias, e em casos onde os valores de vazão são extremamente altos, faz-se necessária a calibração do modelo, permitindo com que ele seja capaz de identificar um conjunto de parâmetros

adequados que representem o sistema hidrológico em análise, e que possam ser utilizados como informações *a priori*.

Esta calibração é realizada por meio dos dados coletados dos outros três postos hidrológicos monitorados pela COPEL, os quais possuem uma alta disponibilidade de informações tanto em períodos de cheias quanto de estiagens, bem como uma grande quantidade de informações necessárias para a realização da calibração. Assim, o principal critério adotado para seleção dos postos foi a disponibilidade e consistência das informações hidrológicas empregadas, além da escolha de postos com comportamentos distintos uns dos outros, para que as informações pudessem ser utilizadas *a priori* sem perda de generalidade.

Após a calibração e levantamento dessas informações, aplica-se o modelo desenvolvido, considerando as informações *a priori*, aos postos monitorados pela CEMIG, a fim de estimar as funções densidade de probabilidade *a posteriori*

dos parâmetros, e a partir dos resultados obtidos calculam-se os percentis da distribuição *a posteriori*, que, por sua vez, determinarão a região de confiabilidade probabilística buscada.

As seções a seguir apresentam estas etapas, de calibração e aplicação, bem como as informações básicas e os métodos desenvolvidos para cada uma dessas etapas.

3.1 Calibração

Um gráfico do comportamento dos pares cota-vazão dos três postos hidrológicos escolhidos para a calibração do modelo, C1, C2, C3 (Tabela 1), é apresentado na Figura 4, pode-se perceber a diferença entre os postos com relação aos seus valores de cota e vazão, e à relação que existe entre eles, apresentada através de suas curvas de descarga teóricas definidas pela COPEL.

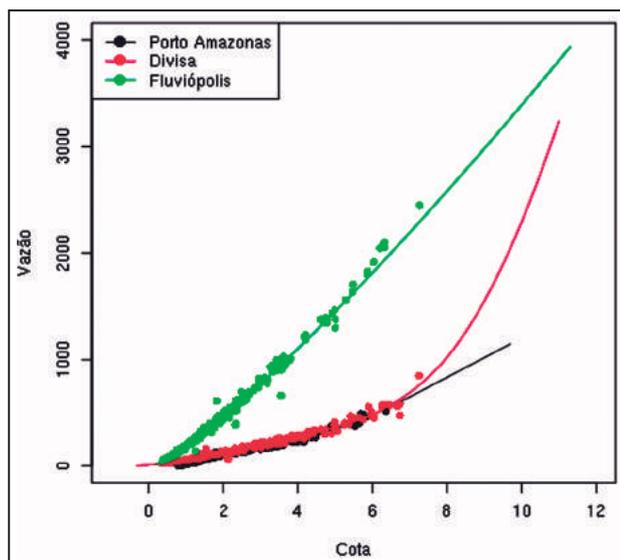


Figura 4 Pares cota-vazão e curvas-chave correspondentes aos postos fluviométricos de Porto Amazonas, Divisa e Fluviópolis, monitorados pela COPEL.

Por outro lado, a Figura 5 apresenta os pares cota-vazão normalizados (para este estudo, normaliza-se um dado x subtraindo-se a média populacional, e dividindo-se o resultado obtido pelo desvio padrão populacional), e por sua vez, a semelhança encontrada entre os três postos, quando seus valores são transformados. A partir disso, conclui-se que estas informações, sem perda de generalidade, podem ser utilizadas como informações *a priori* para os modelos de quaisquer

outros postos, com os mais diferentes tipos de comportamento, desde que os dados utilizados sejam devidamente normalizados.

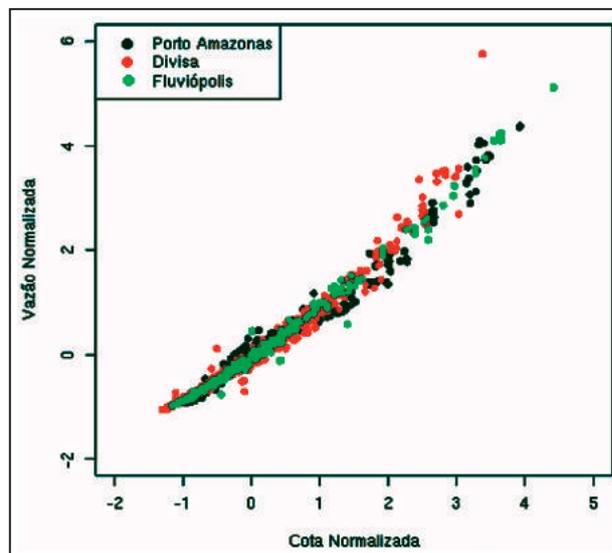


Figura 5 Pares cota-vazão normalizados e a semelhança entre três postos fluviométricos de Porto Amazonas, Divisa e Fluviópolis monitorados pela COPEL.

Com isso, dois modelos chamados de desinformados (que não possuem informações *a priori*) são aplicados aos dados de calibração com o intuito de aproximar, cada um, uma distribuição de probabilidade diferente aos erros obtidos por meio da diferença entre valores de vazão previstos por meio da utilização da curva-chave e valores de vazão observados por métodos convencionais.

O primeiro modelo aproxima uma distribuição de probabilidade Normal ($N(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0$), com parâmetros $\mu = 0$ e $\sigma^2 = \eta_1 + \eta_2 * cota_{norm}$, em que η_1 e η_2 são os parâmetros desinformativos modelados pelo MCMC, $cota_{norm}$ representa os valores de cota normalizados relativos aos valores de vazão, e $\mu = 0$ permite que o MCMC defina valores para os parâmetros η_1 e η_2 respeitando a função definida pela curva-chave previamente determinada para cada um dos três postos fluviométricos.

E, neste mesmo contexto, a segunda distribuição de probabilidade testada, para o segundo modelo, é a distribuição de Cauchy ($C(x_0, \gamma), x_0 \in \mathbb{R}, \gamma > 0$), com $x_0 = 0$ e $\gamma = \eta_1 + \eta_2 * cota_{norm}$, assim como na distribuição Normal.

Ambas as distribuições Normal e Cauchy são simétricas, justificando sua escolha, porém a principal diferença entre elas (Figura 6) é que na distribuição Cauchy as abas mais alongadas fazem com que os valores atípicos inerentes aos dados hidrológicos sejam mais bem representados.

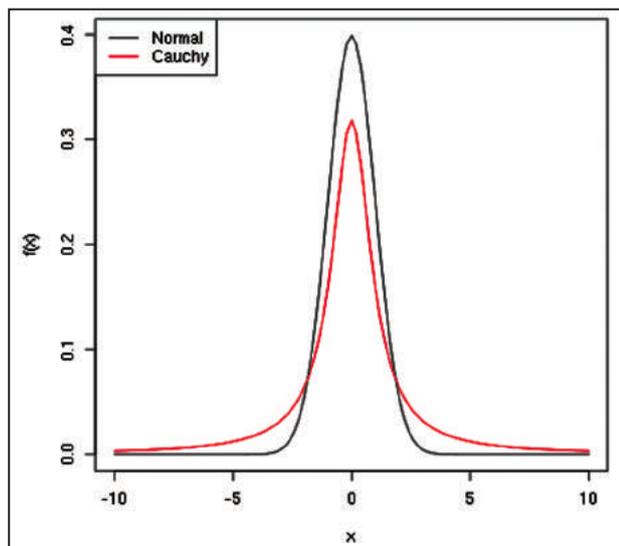


Figura 6 Comparação entre as distribuições: Normal e Cauchy.

A Tabela 2 descreve cada um dos modelos desinformados utilizados para a calibração, na qual q_p é o valor de vazão previsto obtido pela aplicação da função curva-chave e q_{obs} é o valor de vazão observado através de métodos convencionais.

Modelos	erro($q_p - q_{obs}$)
Normal	$erro \sim N(0, \eta_1 + \eta_2 * cota_{norm})$
Cauchy	$erro \sim C(0, \eta_1 + \eta_2 * cota_{norm})$

Tabela 2 Modelos desinformados para a calibração.

A partir da aplicação dos dois modelos aos três postos (C1, C2, C3), são definidos os valores de η_1 e η_2 para cada um deles, apresentados na Tabela 3, e as médias destes valores μ_1 e μ_2 são utilizadas como informações a priori nos modelos desenvolvidos para a etapa de aplicação.

Parâmetros		Div.	Fluv.	P. Am.	μ
Normal	η_1	0,04	0,03	0,06	0,04
	η_2	0,02	0,02	0,03	0,02
Cauchy	η_1	0,08	0,07	0,11	0,09
	η_2	0,06	0,05	0,06	0,06

Tabela 3 Resultados da calibração.

3.2 Aplicação

Utilizando as informações obtidas a priori μ_1 e μ_2 com os modelos que aproximam distribuições Normal e Cauchy aos erros de previsão de vazão, e assumindo $\sigma_1 = \sigma_2 = 0,01$ como medida de dispersão dos valores das distribuições em relação às suas médias, pode-se construir novos modelos, chamados de informativos, para a aplicação nos dez postos selecionados e monitorados pela CEMIG.

Assim como na calibração duas distribuições distintas, Normal e Cauchy são testadas como distribuições dos erros. Entretanto, nesta etapa, os parâmetros η_1 e η_2 deixam de ser desinformativos e passam a utilizar as informações obtidas a priori por meio da etapa de calibração. Estes parâmetros, combinados aos valores de cota normalizados, compõem os valores de σ^2 da distribuição Normal, e de γ da distribuição Cauchy.

Sabe-se por definição que σ^2 e γ devem ser ambos positivos, portanto os parâmetros η_1 e η_2 também devem ser. Desta maneira, utiliza-se a distribuição Gamma ($G(\alpha, \beta), \alpha > 0, \beta > 0$) como uma possível distribuição de probabilidade destes dois últimos parâmetros. Esta distribuição possui média e variância definidas como na Equação 5, que rearranjada resulta na Equação 6.

$$E[X] = \frac{\alpha}{\beta}, Var[X] = \frac{\alpha}{\beta^2} \quad (5)$$

$$\alpha = \frac{E[X]^2}{Var[X]}, \beta = \frac{E[X]}{Var[X]} \quad (6)$$

Assumindo que, $E_1[X] = \mu_1, E_2[X] = \mu_2$ e $Var_1[X] = Var_2[X] = \sigma_1 = \sigma_2$, para a Equação 6, obtêm-se os parâmetros apresentados na Tabela 4, e pode-se inferir sobre os parâmetros η_1 e η_2 como na Equação 7.

	α_1	α_2	β_1	β_2
Normal	81	36	900	600
Cauchy	16	4	400	200

Tabela 4 Valores dos parâmetros.

$$\eta_1 \sim G(\alpha_1, \beta_1), \eta_2 \sim G(\alpha_2, \beta_2) \quad (7)$$

Com as distribuições dos parâmetros definidas previamente constroem-se modelos similares aos utilizados na etapa de calibração, porém com a consideração das informações obtidas *a priori*. Os resultados da aplicação destes modelos aos postos da CEMIG são apresentados na seção seguinte.

4 Resultados e Discussões

Obtida a distribuição *a posteriori* dos parâmetros, tem-se todas as informações dos modelos disponíveis, (*a priori* e a obtida pelos dados) representadas por distribuições de probabilidades. Assim, podem-se calcular resumos descritivos ou probabilísticos sobre os parâmetros, ou simplesmente visualizar essas distribuições (Bonat *et al.*, 2012).

Os algoritmos de simulação de Monte Carlo via Cadeia de Markov (MCMC), são aplicados gerando uma cadeia que após atingir convergência passa a gerar amostras da distribuição objetivo. Esta classe de algoritmos se baseia em propor valores para os parâmetros que são aceitos por uma regra associada a valores *a posteriori*. Entretanto, os valores são simulados de forma simples sequencialmente em uma cadeia e tipicamente cada valor depende do anterior. Faz-se necessário descartar os primeiros valores simulados para evitar dependência do valor inicial.

Além disso, as amostras iniciais não são amostras da distribuição objetivo, a cadeia apenas atinge a distribuição estacionária, que é a de interesse, após um período de convergência. Essas amostras iniciais descartadas são as amostras simuladas durante o período de aquecimento da cadeia.

Devido à simulação sequencial, os valores sucessivos da amostra obtida apresentam dependência entre si. Portanto, para evitar armazenar informações (parcialmente) redundantes, pode-se optar por reter apenas valores a cada certo número de amostras simuladas.

Por meio da aplicação dos modelos construídos em linguagem Stan, todos estes passos são realizados automaticamente. Quando os cálculos são finalizados, faz-se necessário avaliar estes modelos. As tabelas 5 e 6 apresentam um resumo das inferências e convergências dos modelos aplicados a cada um dos dez postos fluviométricos em estudo.

Postos	$M\eta_1$	$M\eta_2$	lp_	Rhat
P1	0,08	0,06	-203,57	1
P2	0,09	0,05	-377,76	1
P3	0,08	0,04	-2,08	1
P4	0,07	0,03	-323,60	1
P5	0,07	0,08	-112,16	1
P6	0,08	0,05	-176,58	1
P7	0,10	0,05	-241,62	1
P8	0,10	0,04	-312,52	1
P9	0,07	0,04	-319,13	1
P10	0,16	0,04	-450,80	1

Tabela 5 Resultados da aplicação: Modelo com distribuição Normal para os erros.

Postos	$M\eta_1$	$M\eta_2$	lp_	Rhat
P1	0,04	0,02	131,30	1
P2	0,04	0,02	-46,10	1
P3	0,04	0,02	374,07	1
P4	0,03	0,01	25,59	1
P5	0,03	0,03	252,85	1
P6	0,03	0,02	185,89	1
P7	0,05	0,02	105,13	1
P8	0,03	0,01	51,88	1
P9	0,02	0,01	31,00	1
P10	0,05	0,01	-84,31	1

Tabela 6 Resultados da aplicação: Modelo com distribuição Cauchy para os erros.

As colunas $M\eta_1$ e $M\eta_2$ representam a média posterior estimada dos parâmetros η_1 e η_2 , a coluna lp_ representa a inferência da média do logaritmo da densidade posterior resultante em cada cadeia do HMC aplicado pelo Stan, e a coluna Rhat apresenta o valor do fator de redução da escala potencial.

O valor de Rhat próximo de 1 significa que as diferentes cadeias do MCMC convergem para um mesmo resultado, e isto acontece tanto nos modelos que aproximam uma distribuição Normal para os erros, quanto para os que aproximam uma distribuição Cauchy, com isso se conclui que ambos os modelos convergem.

O valor lp_ deve ser maximizado, e quanto maior o valor obtido, melhor calibrado está o referido modelo. Pode-se notar que nas dez aplicações lp_ possui maior valor quando a distribuição utilizada é a Cauchy.

Para o primeiro modelo que ajusta uma distribuição Normal para os erros (diferença entre

vazão prevista e vazão observada), os parâmetros $M\eta_1$ e $M\eta_2$ são utilizados para calcular o desvio-padrão $sd = M\eta_1 + M\eta_2 * cota_{norm}$ que determina uma amostra que segue uma distribuição Normal com média 0 e desvio-padrão sd . Desta amostra são calculados os percentis de 2,5%, 5%, 10%, 20%, 25%, 50%, 75%, 80%, 90% e 97,5%.

Analogamente para o segundo modelo que ajusta uma distribuição Cauchy para os erros, os mesmos valores de percentis são calculados a partir de uma amostra que segue distribuição Cauchy $C(x_0, \gamma)$, com $x_0 = 0$ e $\gamma = M\eta_1 + M\eta_2 * cota_{norm}$, assim como na distribuição Normal.

As Figuras 7, 8 e 9 (A e B) apresentam exemplos dos resultados (desnormalizados) obtidos com os cálculos dos percentis para os dois modelos e para três dos dez postos de monitoramento da CEMIG.

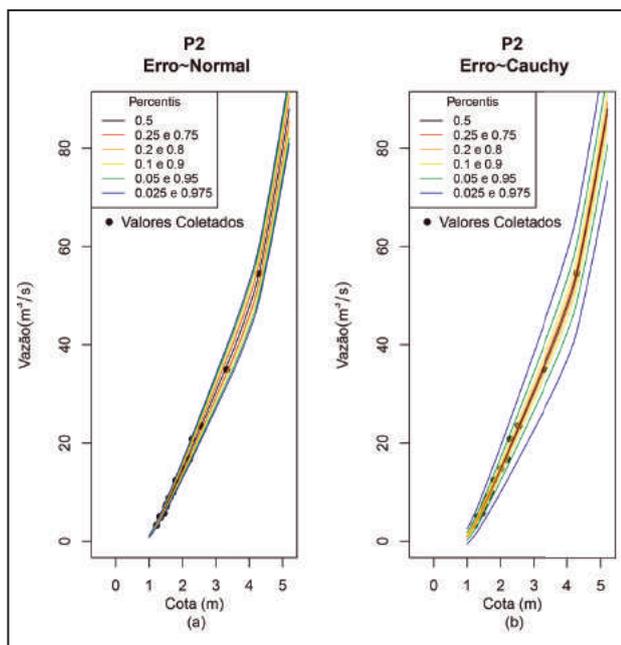


Figura 7 Resultados obtidos para o posto P2 da CEMIG; A. Percentis obtidos com os resultados da aplicação do modelo que aproxima uma distribuição Normal aos erros; B. Percentis obtidos com os resultados da aplicação do modelo que aproxima uma distribuição Cauchy aos erros.

Na Figura 7A os percentis obtidos através da aplicação do modelo que aproxima uma distribuição Normal aos erros do posto fluviométrico P2 (Posto de Poço Fundo Piedade) formam uma região de confiabilidade probabilística estreita, e muito fiel à curva-chave pré-definida, se comparada à

região obtida por meio dos percentis do modelo que aproxima uma distribuição Cauchy aos erros, apresentados na Figura 7B. Demonstrando que, apesar da quantidade reduzida de dados medidos de cota e vazão, o modelo que aproxima a distribuição Cauchy, mescla a importância dos dados com a importância da curva-chave de uma maneira melhor do que o modelo que aproxima a distribuição Normal.

O mesmo pode ser observado nas Figuras 8 (A e B), desta vez para o posto P6 (Posto de Fazenda Boa Vista) que possui uma maior quantidade de pares cota-vazão medidos de forma convencional. Pode-se observar que o formato da região de confiabilidade probabilística segue fielmente o formato da curva de descarga pré-estabelecida. Além disso, quanto maiores forem os valores de cota, e respectivamente os valores de vazão, maior é a abertura da região formada pelos percentis. Este fato ocorre em todos os postos analisados, e só foi possível graças às configurações dos modelos.

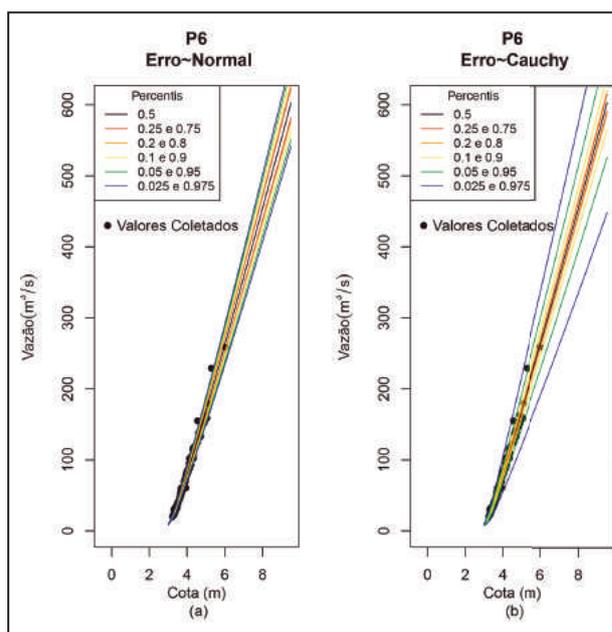


Figura 8 Resultados obtidos para o posto P6 da CEMIG; A. Percentis obtidos com os resultados da aplicação do modelo que aproxima uma distribuição Normal aos erros; B. Percentis obtidos com os resultados da aplicação do modelo que aproxima uma distribuição Cauchy aos erros.

Outros exemplos de resultados são apresentados nas Figuras 9 (A e B). O posto P8 (Posto de Ponte BR 040) possui poucos, e apenas baixos, valores para os pares cota-vazão medidos,

ou seja, pouca informação com relação aos dados, e muita informação com relação à curva-chave, com isso, os percentis resultantes seguem fielmente o formato da curva-chave, e formam uma região de confiabilidade probabilística pouco espaçada.

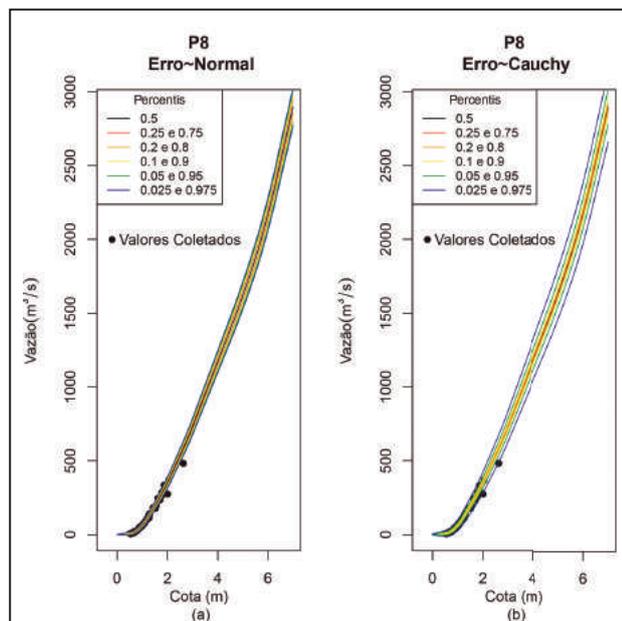


Figura 9 Resultados obtidos para o posto P8 da CEMIG; A. Percentis obtidos com os resultados da aplicação do modelo que aproxima uma distribuição Normal aos erros; B. Percentis obtidos com os resultados da aplicação do modelo que aproxima uma distribuição Cauchy aos erros.

Gráficos semelhantes aos das Figuras 7, 8 e 9 foram construídos para os demais postos de monitoramento em estudo, porém foram omitidos aqui, pois apresentam resultados semelhantes aos que foram mostrados anteriormente.

Os diferenciais entre os gráficos gerados para os dez postos de monitoramento são: o formato de suas curvas-chave, a quantidade de pares cota-vazão, e a diferença entre os valores destes pares. Deve ser destacado que ambos os modelos foram capazes de retornar resultados que possibilitassem a consideração tanto dos pares cota-vazão medidos de forma convencional, quanto da curva-chave pré-definida para os postos de monitoramento. Contudo, o modelo que ajustou a distribuição Cauchy aos erros destacou-se pela capacidade de abrangência em seus resultados.

5 Conclusões

Os resultados apresentados demonstram que a aplicação da inferência Bayesiana, por meio do algoritmo HMC, pertencente à classe dos algoritmos

MCMC, revelou-se adequada aos objetivos propostos, permitindo a construção de regiões de confiabilidade probabilística, bem construídas, em torno das curvas de descarga pré-determinadas para dez postos fluviométricos em estudo, monitorados pela CEMIG.

Dois modelos Bayesianos desinformados foram aplicados aos dados de calibração (pertencentes a três postos hidrológicos monitorados pela COPEL) com o objetivo de aproximar, cada um, uma distribuição de probabilidade aos erros obtidos com a aplicação das funções curva-chave (diferença entre valores de vazão previstos e valores de vazão observados). Foram utilizadas as distribuições Normal e Cauchy, para cada um destes modelos. E, após o período de calibração, com a utilização dos dados pertencentes ao período de aplicação (pertencentes a dez postos hidrológicos monitorados pela CEMIG) em conjunto com as informações obtidas *a priori*, os modelos informativos que ajustam cada um uma distribuição Normal e uma distribuição Cauchy foram testados e comparados. Por meio dos resultados obtidos pode-se concluir que, para os postos de monitoramento em estudo, o modelo que ajustou a distribuição Cauchy se destacou positivamente, se comparado com o modelo que ajustou a distribuição Normal.

As regiões de confiabilidade probabilística que consideram tanto as informações extraídas dos pares cota-vazão medidos convencionalmente, quanto às curvas de descarga, permitem que seja feita, posteriormente, uma avaliação das incertezas associadas à utilização das curvas-chave na determinação de valores de vazões previstos, com as mais diversas amplitudes. Facilitando, com isso, o trabalho de monitoramento, previsão, e diversos outros campos da hidrologia que se utilizam destes dados.

6 Agradecimentos

Os autores agradecem à Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG), à Companhia Energética de Minas Gerais (CEMIG) e à Companhia Paranaense de Energia (COPEL) pelos recursos providos através de projeto de pesquisa e desenvolvimento.

7 Referências

Alfaki, M. 2008. *Improving efficiency in parameter estimation using the Hamiltonian Monte Carlo algorithm.*

- University of Bergen, Norway, Dissertação de Mestrado, 85p.
- Bonat, W.H.; Krainski, E.T.; Ribeiro, J.R. & Zeviani, W.M. 2012. Métodos computacionais para inferência com aplicações em R. *In: SIMPOSIO BRASILEIRO DE PROBABILIDADE E ESTATÍSTICA, SINAPE, 20, CD-ROM.*
- Di Baldassarre, G. & Montanari, A. 2009. Uncertainty in river discharge observations: a quantitative analysis. *Hydrology and Earth System Sciences, 13*: 913-921.
- Domeneghetti, A.; Castellarin, A. & Brath, A. 2012. Assessing rating-curve uncertainty and its effects on hydraulic calibration. *Hydrology and Earth System Sciences, 16*: 1191-1202.
- Duane, S.; Kennedy, A.D.; Pedleton, B.J. & Roweth, D. 1987. Hybrid Monte Carlo. *Physics Letters B, 195*: 216-222.
- Gelfand, A.E. & Smith, A.F.M. 1990. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85*: 398-409.
- Gelman, A.; Carlin, J.B.; Stern, H.S. & Rubin, D.B. 2003. *Bayesian data analysis – 2ª ed.* London, Chapman and Hall, 689p.
- Gelman, A.; Carlin, J.B.; Stern, H.S.; Rubin, D.B.; Dunson, D. & Vehtari, A. 2013. *Bayesian data analysis – 3ª ed.* London, Chapman and Hall, 639p.
- Geman, S. & Geman, D. 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on pattern analysis and machine intelligence, 6*: 721-741.
- Hastings, W.K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*: 97-10.
- Le Coz, J.; Renard, B.; Bonnifait, L.; Branger, F.; Le Boursicaud, R. 2014. Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: a Bayesian approach. *Journal of Hydrology, 509*: 573-587.
- McMillan, H.; Freer, J.; Pappenberger F.; Krueger, T. & Clark, M. 2010. Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. *Hydrological Processes, 24*: 1270-1284.
- Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H. & Teller, E. 1953. Equations of state calculations by fast computing machines. *The Journal of Chemical Physics, 21*: 1087-1091.
- Montanari, A. & Brath, A. 2004. A stochastic approach for assessing the uncertainty of rainfall-runoff simulations. *Water Resour, 40*: W01106.
- Montanari, A. & Grossi, G. 2008. Estimating the uncertainty of hydrological forecasts: A statistical approach. *Water Resour, 44*: W00B08.
- Moyeed, R.A. & Clarke, R.T. 2005. The use of Bayesian methods for fitting rating curves, with case studies. *Advances in Water Resources, 28*: 807-818.
- Overleir, A.P. 2004. Accounting for heteroscedasticity in rating curve estimates. *Journal of Hydrology, 292*: 173-181.
- R Core Team. 2012. *R: a language and environment for statistical computing.* R Foundation for Statistical Computing, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Rantz, S.E. 1982. *Measurement and computation of streamflow: Volume 2. Computation of discharge.* Geological Survey Water-Supply Paper 2175. Washington, United States Government Printing Office, 118p.
- Schmidt, A.R. 2002. *Analysis of stage-discharge relations for open-channel flows and their associated uncertainties.* University of Illinois, Urbana-Champaign, Tese de Doutorado, 329p.
- Silva, F.E.; Naghettini, M. & Fernandes, W. 2014. Avaliação bayesiana das incertezas nas estimativas dos parâmetros de um modelo chuva-vazão conceitual. *Revista Brasileira de Recursos Hídricos, 19*: 148-159.
- Westerberg, I.; Guerrero, J.L.; Seibert, J.; Beven, K.J. & Halldin, S. 2011. Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River. *Hydrological Process, 25*: 603-613.