

## XIV SIMPÓSIO DE RECURSOS HÍDRICOS DO NORDESTE

### **ANÁLISE DE REDES FLUVIOMÉTRICAS COM EMPREGO DE ENTROPIA**

*Luiz Pádua<sup>1</sup>; Mauro Naghettini<sup>2</sup>; Francisco Silva<sup>3</sup>*

**RESUMO** – A coleta, monitoramento e manutenção de informação contínua e precisa, gerada por estações fluviométricas de uma rede hidrométrica, é de fundamental importância para subsidiar o desenvolvimento de estudos relacionados à hidrologia e à hidráulica, como controle de enchentes e implantação de obras hidráulicas. Neste trabalho, foram empregados métodos que buscam avaliar a quantidade de estações e a qualidade da informação gerada pela rede fluviométrica de uma bacia hidrográfica, utilizando conceitos da teoria da informação. A ideia subjacente é a da chamada rede ótima, cuja função é a de atender, de forma otimizada o objetivo precípuo da hidrometria, qual seja, o de oferecer a informação necessária com um número mínimo de estações corretamente posicionadas na bacia. Os métodos avaliados foram aplicados à sub-bacia do Rio das Velhas pertencente à bacia do Rio São Francisco, no Brasil. Os resultados demonstraram que os métodos analisados, os quais se valem do conceito de entropia, são adequados e eficientes para avaliação de redes fluviométricas existentes, uma vez que permitem a redução de eventuais redundâncias e a maximização da informação gerada pelas estações de monitoramento.

**ABSTRACT**– The collection, monitoring and maintenance of continuous and accurate information, generated by fluviometric stations of a hydrometric network, is fundamental importance to support the development of studies related to hydrology and hydraulics, such as flood control and the implementation of hydraulic structures. In this work, we used models that seek to evaluate the number of stations and the quality of the information generated by the fluviometric network of a hydrographic basin, using information theory concepts. The underlying idea is that of the so-called optimal network, whose function is to optimally meet the primary goal of hydrometry, namely to provide the necessary information with a minimum number of stations correctly positioned in the basin. The evaluated models were applied to the Velhas River subbasin belonging to the São Francisco River basin in Brazil. The results showed that the models analyzed, which use the concept of entropy, are adequate and efficient for the evaluation of existing fluviometric networks, since they allow the reduction of eventual redundancies and seeking the maximization of the information generated.

**Palavras-Chave** – Teoria da Informação, Entropia, Rede Hidrométrica, Rede Fluviométrica, Modelagem Hidrológica, Rio das Velhas.

1) Universidade Federal de Minas Gerais, [lh.padua@gmail.com](mailto:lh.padua@gmail.com)

2) Universidade Federal de Minas Gerais, [mauro.naghettini@gmail.com](mailto:mauro.naghettini@gmail.com)

3) Universidade Federal de Minas Gerais, [fsilva@ehr.ufmg.br](mailto:fsilva@ehr.ufmg.br)

## INTRODUÇÃO

O objetivo principal desse trabalho é avaliar a rede fluviométrica existente da bacia hidrográfica do Rio das Velhas, localizada na região central do Estado de Minas Gerais, Brasil. A bacia hidrográfica do Rio das Velhas, maior afluente em extensão da bacia do Rio São Francisco, tem aproximadamente 801 km de extensão. Sua bacia compreende uma área aproximada de 29.173 km<sup>2</sup>.

A rede hidrométrica, conforme WMO (2008), pode ser definida como um conjunto de instalações desenvolvidas sob com o propósito de permitir a coleta de dados dos diferentes componentes do ciclo hidrológico, projetada e operada para subsidiar a tomada de decisões compatíveis ao amplo contexto da hidrologia e hidráulica. Com particular referência às estações fluviométricas, pretende-se identificar a rede ótima, o que, em síntese, refere-se à identificação dentre as estações já implantadas, daquelas que permitem melhor caracterizar o regime sazonal de vazões (maior geração de informação), e daquelas cuja informação gerada é redundante, e, por conseguinte, podem ser relocadas ou mesmo desativadas.

Em relação à rede fluviométrica existente, foram selecionadas, na área de estudo, sete estações fluviométricas ao longo do curso do Rio das Velhas, operadas pela Agência Nacional de Águas (ANA). Os dados utilizados pelas estações selecionadas referem-se às vazões médias mensais, coletados ao longo de um período de 19 anos, entre os anos de 1994 e 2013.

Para análise da referida de tal rede hidrométrica, serão foram aqui utilizadas as metodologias baseadas na Teoria da Informação, conforme Shannon (1948), as quais que buscam avaliar a quantidade de postos e a qualidade da informação gerada pela rede fluviométrica em bacias hidrográficas.

## METODOLOGIA

### Entropia

Conforme Singh (1997), a entropia representa uma medida da quantidade de caos, ou melhor dizendo, da falta de informações sobre um sistema. O que, conforme esse autor, pode-se afirmar que a entropia representa a quantificação de nossa ignorância sobre um sistema.

De acordo com Shannon e Weaver (1949), a entropia, denotada por  $H(X)$ , de uma variável aleatória discreta  $X$ , com valores discretos  $x_1, x_2 \dots x_n$ , com probabilidades  $p(x_1), p(x_2), \dots p(x_n)$ , onde  $n$  representa o número de elementos, é dada por:

$$H(X) = - \sum_{i=1}^n p(x_i) \log p(x_i) \quad (1)$$

A Equação (1) refere-se a entropia de uma variável discreta  $X$ , com sua respectiva função de probabilidade  $p(x)$ . Dessa forma, a incerteza média de uma variável aleatória discreta  $X$ , com uma distribuição de probabilidade  $p(x)$  é referenciada como a entropia de  $p(x)$ , e representada por  $H(X)$ . O que significa que uma variável com uma entropia  $H(X)$  bits é capaz de prover informação suficiente para se escolher entre  $m = 2^{H(X)}$  alternativas de iguais probabilidades, Stone (2015).

Na Equação (1), a notação  $H$  refere-se à notação padrão para quantidade de informação, conforme Shannon (1948). Nesse trabalho a base logarítmica poderá ser omitida, a menos que seja necessário expressa-la em uma base diferente de 2. Nessa mesma Equação,  $X$  representa uma variável aleatória discreta e sua entropia será denotada por  $H(X)$ . Isso significa que  $X$  não é um argumento de uma função, ele representa somente a indicação dessa variável, para diferenciar de outra possível variável  $Y$ , que teria sua entropia representada por  $H(Y)$ , por exemplo.

Conforme detalhado por Shannon e Weaver (1949), a utilização da quantidade de entropia  $H$  permite vantagens sobre outros possíveis quantificadores, por suas propriedades que a tornam interessante para mensuração de quantidade de informação.

Dentre as propriedades, algumas se destacam como a Entropia Conjunta (EC), ou seja, a quantidade de informação contida entre duas variáveis aleatórias discretas. Para duas variáveis aleatórias,  $X$  e  $Y$ , estocásticamente independentes, a entropia total será a soma das entropias individuais ( $H(X) + H(Y)$ ). Mas se essas variáveis são estocásticamente dependentes, parte da informação contida é compartilhada, e sua entropia conjunta será então, Alfonso *et al.* (2012),

$$H(X, Y) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log p(x_i, y_j) \quad (2)$$

Quando o objetivo é somente medir as dependências entre diversas variáveis aleatórias, ou melhor dizendo, quanta informação é compartilhada por essas variáveis, pode-se computar essa quantidade de informação compartilhada, conforme demonstrado por [Alfonso *et al.*, (2010); McGill, (1954)]. Que conforme esses autores, é possível acessar de forma direta e efetiva a dependência entre múltiplas variáveis aleatórias utilizando o conceito de Correlação Total (CT),  $C$ , pela formulação:

$$C(X_1, X_2, \dots, X_n) = \sum_{i=1}^n H(X_i) - H(X_1, X_2, \dots, X_n) \quad (3)$$

Outra importante relação, capaz de mensurar a dependência entre duas variáveis aleatórias, é a que se refere à Informação Mútua, por vezes, também designada de Transinformação,  $I$ , ou seja, a relação que expressa a redução da incerteza de  $X$  devido ao conhecimento de  $Y$ .

A relação que expressa essa quantificação da Informação Mútua entre duas variáveis aleatórias é computada da seguinte forma,

$$I(X; Y) = \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \quad (4)$$

que pode ser escrita como,

$$I(X; Y) = H(X) - H(X|Y) \quad (5)$$

O termo  $H(X|Y)$  refere-se a Entropia Condicional (ECO), ou seja, a quantidade de informação contida em  $X$  que não está contida em  $Y$ , tendo como  $p(x|y)$ , a probabilidade de se obter  $y$ , conhecendo-se  $x$ . Pode-se definir essa quantidade conforme Cover e Thomas (2012),

$$H(Y|X) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log p(x_i|y_j) \quad (8)$$

### Discretização da série contínua

A metodologia proposta utilizando Entropia, pode ser computada pela discretização das séries de dados contínuos medidas pelas estações fluviométricas.

Neste trabalho optou-se por seguir a metodologia proposta por [Alfonso *et al.* (2010),(2010a),(2012); Li *et al.*, (2012), utilizando a discretização dos dados pela quantização dos mesmos por uma função matemática de nivelamento, onde um valor contínuo da série de dados  $x$ , é convertido em um número inteiro mais próximo, multiplicado por uma constante  $a$ , conforme Equação (9),

$$x_q = a \left( \frac{2x+a}{2a} \right) \quad (9)$$

onde o termo entre parênteses representa a função niveladora. Conforme mostrado por Li *et al.* (2012), a utilização da função niveladora possibilita duas importantes vantagens: (i) ela retira a necessidade de utilização de uma função paramétrica que se ajuste aos dados contínuos; (ii) ela incorpora um

parâmetro físico, representado pela constante  $a$  na sua resolução, diretamente envolvido na natureza da série de dados.

O valor para o parâmetro  $a$  pode ser obtido de diferentes formas de acordo com o contexto do estudo realizado, crucial para o cálculo da Entropia. Neste trabalho, a metodologia utilizada baseou-se na análise das séries de valores obtidos, pela relação entre Entropia Conjunta (EC) e a Correlação Total (CT), buscando o valor ótimo, tal que se obtenha um valor de  $a$  que gerasse um maior valor para EC-CT.

### Análise de redes fluviométricas

O método utilizado para análise da rede fluviométrica, *Maximum Information Minimum Redundancy* (MIMR), proposto por Li *et al.*, (2012), é baseado em análise multiobjetivo, que busca a locação das estações estudadas pela maximização da quantidade de informação que cada uma gera (Entropia), a maximização da Transinformação e a minimização da redundância de informação entre as estações (Total Correlação), conforme esquema a seguir,

$$\left\{ \begin{array}{l} \text{Max. } H(X_1) + \dots + H(X_i) \\ \text{Max. } \sum_{j=1}^{i-1} T(X_i; X_j) \\ \text{Min. } \sum_{j=1}^{i-1} C(X_i; X_j) \end{array} \right. \quad (10)$$

Conforme pode ser observado pela Equação (10), esta indica que o algoritmo resultará como rede ótima, aquela que atingir o melhor ponto de convergência entre a máxima informação efetiva, pela minimização da redundância de informação aferida pela correlação total.

O algoritmo proposto segue as seguintes etapas para seleção da rede ótima Li *et al.*, (2012): (i) Primeiramente faz a leitura dos dados das estações utilizadas; (ii) Realiza a discretização dos dados das estações; (iii) Identifica a estação central, delegada pela estação com maior valor de entropia; (iv) Cria o primeiro conjunto de estações ótimas; (v) Seleciona o segundo conjunto de redes ótimas pela resolução contínua da Equação (10).

### Área de Estudo

A bacia hidrográfica escolhida corresponde a bacia do Rio das Velhas, localizada na região central do Estado de Minas Gerais, Brasil. Sua bacia compreende uma área de 29.173 km<sup>2</sup>, abrangendo 51 municípios mineiros, com uma população próxima de 5 milhões de habitantes, conforme censo do Instituto Brasileiro de Geografia e Estatística (IBGE) de 2010.

Ao longo da bacia do Rio das Velhas foram identificadas 7 estações fluviométricas, com dados superiores ou iguais a 15 anos, com menor número de falhas possíveis, conforme dados da agência Nacional de Águas (ANA), disponibilizados pelo portal eletrônico do Hidroweb ([www.snirh.gov.br/hidroweb](http://www.snirh.gov.br/hidroweb)).

Na Figura 1 é apresentada a bacia do Rio das Velhas e as respectivas estações fluviométricas selecionadas para o estudo. Na Tabela 1 são apresentadas as características das estatísticas básicas de cada uma dessas estações fluviométricas.

Tabela 1 – Estações fluviométricas e características estatísticas das vazões.

ID	Código ANA	Área de Drenagem (km <sup>2</sup> )	Média	Variância	Máximo	Mediana	Mínimo
1	41199998	1.550	31	315	114	25	10
2	41260000	3.730	57	1719	295	43	7
3	41340000	4.860	70	2492	409	54	20
4	41410000	7.080	94	5472	461	64	21
5	41650002	10.700	120	8651	600	83	25
6	41818000	16.600	183	25738	904	109	31
7	41990000	26.500	252	50782	1165	147	37

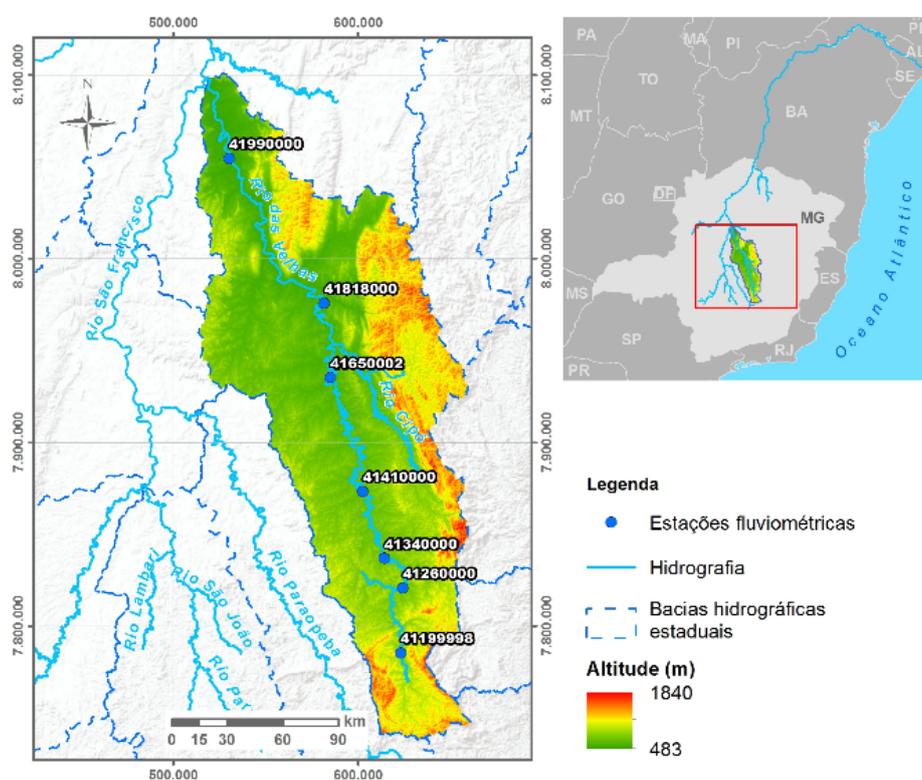


Figura 1 – Mapa de localização do Rio das Velhas e locação das estações fluviométricas selecionadas.

## RESULTADOS

Nessa etapa foi aplicado o método proposto para análise das estações fluviométricas e ranqueamento das mesmas, ou seja, dentro dessas sete estações, quais geram maiores valores de Entropia, e conseqüentemente agregam mais informação à rede, mas ao mesmo tempo mantêm o mínimo possível de informação redundante. Objetivo esse proposto pelo método em análise.

Primeiramente é apresentada a seguir a Tabela 2, referente às entropias individuais das estações, para auxiliar no entendimento dos resultados obtidos para o método de otimização utilizado.

Tabela 2 - Entropia Individual de cada estação fluviométrica.

H (bits)	ID						
	1	2	3	4	5	6	7
	0.34	0.99	1.12	1.45	1.62	2.04	2.36

Os resultados referentes a Entropia Conjunta e Total Correlação, à medida que uma nova estação é adicionada a rede (Passos Interativos), e os valores totais para esses dois parâmetros para todas as sete em conjunto são apresentados na Tabela 3. Esses valores foram obtidos utilizando o valor de  $a$  de  $127 \text{ m}^3/\text{s}$  para discretização das séries contínuas de vazões médias mensais.

Tabela 3 - Entropia Conjunta e Total Correlação das estações fluviométricas ranqueadas por MIMR.

	Passos Interativos						
	1	2	3	4	5	6	7
	MIMR						
ID	7	5	1	3	2	4	6
H	2.36	3.13	3.27	3.69	3.88	4.24	4.72
C	0	0.85	1.06	1.76	2.56	3.66	5.22

Como pode ser observado na Tabela 2, o valor da entropia das estações cresce à medida que as mesmas se aproximam do exutório da bacia (ID1 a ID7), confirmando a similaridade com a variância da series de dados. A entropia conjunta das 7 estações encontrada foi de 4.72 bits, que representa a máxima informação que pode ser extraída delas Li *et al.* (2012). A total correlação obtida, conforme Tabela 3, foi de 5.22 bits, que significa o total de informação duplicada contida no sistema. Conforme Tabela 2, a soma das entropias individuais é de 9.92 bits, valor este aproximadamente 90% superior a Total Correlação, o que mostra que a rede hidrométrica representa um sistema disperso, ou seja, as estações não estão alocadas de forma otimizada.

Importante ponto observado no ranqueamento é que não prevaleceu o sequenciamento dos ranques pelo valor crescente da Entropia individual das estações ou por suas variâncias. Mostrando assim a existência de informação redundante no sistema, tal como foi observado pelo cálculo superior a zero da total correlação. O que mostra que designar as estações de uma rede simplesmente por seus valores de entropia máximos, não levará ao melhor conjunto e poderá gerar mais informação redundante.

Outro ponto refere-se a Entropia Conjunta. Percebe-se pela Tabela 3 que ela aumenta à medida que uma nova estação é incluída na rede. O motivo disso acontecer pode estar relacionado ao baixo número de estações para uma bacia tão grande. O que aplicando parâmetros recomendados na literatura como WMO (2008), para a sub-bacia do Rio das Velhas seriam necessárias cerca de 12 estações fluviométricas.

Sobre o aumento percebido da EC e CT, entre os passos 2-3 e 6-7, percebe-se que fica evidenciado por essa análise, a baixa contribuição obtida pela inserção das estações (ID 1 e ID6) à rede. Após a inclusão das mesmas, o ganho de informação é próximo de 10% para EC, o que a princípio é de grande valia. Porém esse aumento acarreta um acréscimo de quase 50% na CT, indicando assim que a adição dessas novas estações, nesses locais, pouco acrescenta na otimização da rede fluviométrica, ligado a redundância de informação gerada, indicando assim suas possíveis relocações.

Outro ponto refere-se ao conjunto das quatro primeiras estações ranqueadas. Essas quatro estações representam quase 80% da informação do sistema, 3.69 bits. O que mais uma vez evidencia o caráter redundante observado na rede e baixa eficiência da mesma.

Dessa forma, aplicando-se a metodologia, como resultado poder-se-ia recomendar a relocação das Estações ID1 e ID6 de forma a melhorar suas contribuições ao sistema de monitoramento do Rio das Velhas, uma vez que, sob o foco da quantidade de informação, as mesmas estão contribuindo de forma redundante ao sistema. É claro que vale ressaltar novamente, que essa recomendação é feita sob o único critério da Entropia. Não foram levados em consideração possíveis necessidades específicas de monitoramento dessas estações devido a suas localizações geográficas, técnicas, sociais, etc, ou qualquer outra variável que justificasse sua manutenção nesses locais para a bacia hidrográfica.

## CONCLUSÕES

Os resultados encontrados corroboram aqueles encontrados por [Alfonso *et al.* (2010) e Li *et al.* (2012)], mostrando a adequação dessa metodologia, agora analisada em bacias hidrográficas no Brasil de clima tropical.

Cabe destacar que até esse ponto, a expansão da rede fluviométrica está além da capacidade do método avaliado. Futuramente esse ponto pode ser resolvido por simulação. Modelos hidrodinâmicos podem ser gerados para simular uma rede sintética densa de estações. Entretanto, essa tarefa ainda poderá gerar uma grande quantidade de incertezas, visto a baixa quantidade de informação disponível ou da natureza empírica de tais modelos. Etapa a ser avaliada em estudos futuros.

Destaca-se a grande contribuição que tais metodologias agregam à essa linha de pesquisa voltada ao estudo de redes hidrométricas, problemática enfrentada em todos os países que buscam a otimização de suas redes e redução de custos.

## AGRADECIMENTOS

Os autores gostariam de expressar seus sinceros agradecimentos a Leonardo Alfonso, do departamento de Hydroinformatics and Knowledge Management, UNESCO-IHE, Delft, Netherlands e Chao Li, do Department of Biological and Agricultural Engineering, Texas A&M University, College Station, Texas, USA, que proveram parte dos dados que permitiram a realização desse estudo.

Agradecimentos a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte econômico fornecido.

## REFERÊNCIAS BIBLIOGRÁFICAS

ALFONSO, L., A.; LOBBRECHT; R. PRICE. (2010). “*Information theory–based approach for location of monitoring water level gauges in polders*”, *Water Resour. Res.*, 46, W03528, doi:10.1029/2009WR008101.

ALFONSO, L., A. LOBBRECHT, AND R. PRICE. (2010A). “*Optimization of water level monitoring network in polder systems using information theory*”, *Water Resour. Res.*, 46, W12553, doi:10.1029/2009WR008953.

ALFONSO, LEONARDO & HE, LIYAN & LOBBRECHT, ARNOLD & PRICE, ROLAND. (2012). “*Information theory applied to evaluate the discharge monitoring network of the Magdalena River*”. Journal of Hydroinformatics. 15. 211-228. 10.2166/hydro.2012.066.

COVER, T. M., AND J. A. THOMAS (2012). *Information Theory*, John Wiley, New York.

LI, C.; SINGH, V.P.; MISHRA, A.K. (2012). “*Entropy theory-based criterion for hydrometric network evaluation and design: Maximum information minimum redundancy*”. Water Resour. Res. 48, doi:10.1029/2011WR011251.

MCGILL, W. J. (1954). *Multivariate information transmission*, Psychometrika, 19, 97–116.

SHANNON, C.E. A. (1948). “*Mathematical Theory of Communication*”. Bell Syst. Tech. J., 27, 379–423.

SHANNON, C. E., WEAVER W. (1949). “*The Matemactical theory of Communication*”. University of Illinois Press.

SINGH, V. P. (1997). “*The use of entropy in hydrology and water resources*”. Hydrol. Process. 11, 587–626.

STONE, J.V., *Information Theory A Tutorial introduction*. SEBTEL PRESS, 2015. 243 p.

WORLD METEOROLOGICAL ORGANIZATION (WMO) (2008). *Guide to hydrological practices, volume I: Practices hydrology—From measurement to hydrological information*, WMO 168, 16th ed., World Meteorological Organization (WMO), Geneva, Switzerland.