

## XXIII SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS

### **AValiação DO EFEITO DE MÉTODOS DE PRÉ-TRATAMENTO DE DADOS EM AGRUPAMENTOS DE CHEIAS HISTÓRICAS**

*Patrícia Cristina Steffen*<sup>1</sup>; *Júlio Gomes*<sup>2</sup>

**RESUMO** – Cheias constituem o desastre natural com maior número de mortes no mundo, o que levanta discussões a respeito de como prevenir situações de risco, bem como, minimizar seus impactos. Na tentativa de reduzir os impactos oriundos destes fenômenos, a análise de agrupamento de cheias pode favorecer o desenvolvimento de sistemas de previsão e alerta de cheias. Porém, para que a análise de agrupamento possa ser melhor executada, é necessário aplicar um tratamento inicial dos indicadores de cheia selecionados, a fim de reduzir as diferenças físicas e de unidade de medida existentes entre eles. Desta forma, tentando aprimorar o estudo relacionado ao tema, o presente estudo propõe a aplicação de diferentes métodos de pré-tratamento dos dados, analisando o consequente comportamento do agrupamento das cheias observadas no município de União da Vitória, Paraná. Os métodos aplicados foram: *autoscaling*, *range scaling*, *Pareto scaling*, *vast scaling* e *level scaling*. Os resultados mostraram que os diferentes métodos produziram agrupamentos distintos das cheias. Verificou-se ainda que o método *vast scaling* apresentou um melhor desempenho, considerando-se os eventos observados no local do estudo.

**ABSTRACT** – Floods are the global natural disaster with the highest number of deaths, which raises discussions about how to prevent risk situations, as well as minimize their impacts. In the attempt to reduce the impacts from those phenomena, flood-clustering analysis may favor the development of flood forecasting and warning systems. However, in order to the clustering analysis to be better executed, it is necessary to apply an initial treatment of the selected flood indicators to reduce the physical and measurement unit differences among them. Thus, in order to improve the study related to the subject, the present study proposes the application of different methods of data initial treatment, analyzing the consequent behavior of the observed floods clustering in the municipality of União da Vitória, Paraná. The applied methods were *autoscaling*, *scaling range*, *Pareto scaling*, *vast scaling* and *level scaling*. The results showed that the different methods produced distinct clustering of floods. It was also noted that the *vast scaling* method presented a better performance, considering the events observed at the study site.

**Palavras-Chave** – Pré-tratamento de dados. Agrupamento de cheias. *Fuzzy c-means*.

#### **INTRODUÇÃO**

Apesar de representarem a menor parte do período hidrológico de um corpo hídrico, os eventos extremos máximos, quando ocorrem, podem gerar consequências de grandes proporções para a região atingida (TUCCI; BERTONI, 2003). Historicamente, há inúmeros registros de eventos extremos máximos, não apenas no Brasil, e uma vez que representam o desastre natural com maior número de mortes ocasionadas, requerem estudos direcionados, a fim de tentar minimizar seus impactos.

---

1) Mestra, Doutoranda no Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental, UFPR, patriciaacsteffen@yahoo.com.br  
2) Doutor, Professor no Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental, UFPR, jgomes.dhs@ufpr.br

Segundo Steffen e Gomes (2018), a análise de agrupamento favorece o avanço no conhecimento dos processos de transformação chuva-vazão em uma determinada localidade e, conseqüentemente, pode contribuir para o desenvolvimento de sistemas de previsão e alerta de cheias, a partir da calibração de modelos para os grupos obtidos, com o objetivo de auxiliar na minimização das conseqüências originadas pelos eventos extremos máximos.

Com base nos trabalhos de Wang *et al.* (2014) e Steffen (2017), utilizou-se o método de agrupamento *fuzzy c-means (fcm)*, que permite a definição de grupos de eventos de cheia que possuam características semelhantes entre si, desde que haja uma definição prévia de um conjunto de indicadores de cheia. Em resumo, fundamentado nos conceitos da lógica difusa, o *fcm* determina níveis de proximidade dos dados da amostra (eventos de cheia) com cada um dos grupos gerados.

Assim, dados os indicadores de cheia a serem analisados no presente estudo (vazão e nível de água máximos, volume total extravasado, tempo acima do limiar de extravasamento, tempo entre início do evento e ultrapassagem do limiar de extravasamento e taxas médias de ascensão e recessão do hidrograma de cheia), sabe-se que, ao compará-los, existem diferenças físicas e de unidade de medida inerentes, que precisam ser eliminadas antes de se iniciar o processo de agrupamento.

Para eliminar as diferenças físicas e de unidades entre os indicadores de cheia, Wang *et al.* (2014) e Steffen (2017) utilizaram o método da normalização, porém a literatura ainda apresenta outros métodos capazes de diminuir tais divergências. Van den Berg *et al.* (2006) apresentam três classes de métodos de tratamento inicial de informações: *centering*, *scaling* e *transformations*. Os métodos da classe *scaling* são, de forma genérica, baseados em parâmetros estatísticos para sua definição, como média e desvio-padrão dos dados da amostra. Já a classe *centering* converte os valores da amostra para intervalos próximos de zero, sendo, em geral, aplicado em combinação com os métodos do tipo *scaling*. Os métodos tipo *transformations*, por sua vez, constituem-se de transformações não lineares da amostra. Neste estudo, o enfoque é dado aos métodos do tipo *scaling*, apresentados por Van den Berg *et al.* (2006), e mais bem descritos na seção de métodos.

Desta forma, o objetivo do presente trabalho é avaliar o efeito que os métodos de tratamento inicial das informações (*autoscaling*, *range scaling*, *Pareto scaling*, *vast scaling* e *level scaling*) exercem sobre o agrupamento de cheias observadas no município de União da Vitória, Paraná, Brasil, no período de 1931 a 2015. Além disso, a comparação usou como base o trabalho de Steffen (2017), que aplica o método da normalização à mesma região e que agrupou as cheias em 4 grupos distintos.

## ÁREA DE ESTUDO

A área de estudo é o município de União da Vitória, na região sudeste do estado do Paraná. A bacia do rio Iguaçu se subdivide em três trechos: alto, médio e baixo Iguaçu, localizando-se o município de União da Vitória no seu trecho médio. A Figura 1 apresenta a localização de União da

Vitória no Estado do Paraná e o trecho médio da bacia do rio Iguaçu, representado pelo número 11 na referida figura.

A estação fluviométrica União da Vitória (65310000), com série de vazões observadas de 1931 a 2015, foi usada para o estudo. As vazões observadas, a partir de 1980, foram corrigidas a partir dos dados de nível de água observados na estação fluviométrica R5 Porto Vitória (65365800), devido ao início de operação da Usina Hidrelétrica Foz do Areia, localizada no rio Iguaçu, a jusante da cidade de União da Vitória. Steffen (2017) apresenta a série de dados utilizada no presente estudo, bem como, traz informações complementares sobre a correção efetuada nas vazões.



Figura 1 - Localização da cidade de União da Vitória no estado do Paraná e na bacia hidrográfica do rio Iguaçu (STEFFEN; GOMES, 2018; adaptado de ÁGUASPARANÁ, 2016)

A partir da série de vazões da estação fluviométrica União da Vitória (65310000), foram então selecionados os eventos de cheia anuais, conforme apresenta Steffen (2017). Foram selecionados 85 eventos, caracterizados por cheias com ou sem extravasamento, representando as máximas cheias observadas em cada ano do período em análise.

## MÉTODOS

Na Estação Fluviométrica União da Vitória (65310000), o zero da régua corresponde à cota de 739,61 m, enquanto a cota de desapropriação das áreas ribeirinhas ao rio Iguaçu em União da Vitória é equivalente a 744,50 m, o que corresponde a uma leitura de régua de 4,89 m para o extravasamento. A partir da família de curvas de descarga da estação fluviométrica, apresentada em Castanharo e Buba (2008), a vazão correspondente ao extravasamento é de 1.387 m<sup>3</sup>/s.

## Indicadores de cheia

A análise de agrupamento requer a definição de indicadores de cheia. Aqui, os indicadores de cheia utilizados foram os mesmos adotados por Steffen (2017): vazão e nível de água máximos,

volume total extravasado, tempo acima do limiar de extravasamento, tempo entre início do evento e ultrapassagem do limiar de extravasamento e taxas médias de ascensão e recessão do hidrograma de cheia. Steffen (2017) calculou os referidos indicadores para todos os eventos selecionados e apresenta informações detalhadas sobre o processo.

### Tratamento inicial das informações

O tratamento inicial foi aplicado aos sete indicadores de cheia, efetuado a partir de cinco processos distintos: *autoscaling*, *range scaling*, *Pareto scaling*, *vast scaling* e *level scaling*. Os equacionamentos que seguem são encontrados em Van den Berg *et al.* (2006).

*Autoscaling*, também chamado de *unit* ou *unit variance scaling* (VAN DEN BERG *et al.*, 2006) é um dos métodos comumente aplicado, que se baseia no desvio-padrão (MINGOTI, 2005; VAN DEN BERG *et al.*, 2006). Após a aplicação do *autoscaling*, os elementos constituintes de cada indicador geram um desvio padrão igual a 1 (*unit*), e passam a ser valores adimensionais. Van den Berg *et al.* (2006) destacam que o método *autoscaling* inflaciona a medida de erros. A Equação (1) apresenta a aplicação prática do *autoscaling* aos indicadores da amostra:

$$x_{ij} = \frac{x_{ij}^0 - \bar{x}_j}{s_j} \quad (1)$$

*Range scaling*, assim como o *autoscaling*, utiliza de medidas de dispersão da amostra, e, além de tornar os valores da amostra adimensionais e inflacionar as medidas de erro, é sensível aos *outliers*. A Equação (2) apresenta a aplicação do *range scaling* aos indicadores da amostra:

$$x_{ij} = \frac{x_{ij}^0 - \bar{x}_j}{x_{j,max}^0 - x_{j,min}^0} \quad (2)$$

*Pareto scaling*, por sua vez, é muito semelhante ao *autoscaling*, porém, ao invés de utilizar o desvio-padrão da amostra não tratada, utiliza-se de sua raiz quadrada como fator de escala. Desta forma, os dados não se tornam adimensionais após o tratamento, mas sim, mantêm a sua unidade de medida original, mantendo os valores da amostra mais próximos da medida original que o *autoscaling*. A Equação (3) representa a obtenção da amostra tratada pelo método de *Pareto scaling*:

$$x_{ij} = \frac{x_{ij}^0 - \bar{x}_j}{\sqrt{s_j}} \quad (3)$$

*Vast scaling* tem seu nome originado em “*variable stability scaling*”, e consiste em uma variação do *autoscaling*. O método foca em variáveis estáveis e utiliza o desvio-padrão e o coeficiente de variação como fatores de escala, como mostra a Equação (4):

$$x_{ij} = \left( \frac{x_{ij}^0 - \bar{x}_j}{s_j} \right) \frac{\bar{x}_j}{s_j} \quad (4)$$

*Level scaling* se enquadra em uma subclasse dentre os métodos do tipo *scaling*, uma vez que, diferentemente dos demais métodos, utiliza uma medida de tamanho, ao invés de uma medida de dispersão, porém, também inflaciona as medidas de erros e fornece valores tratados adimensionais. A Equação (5) apresenta o processo de tratamento pelo método *level scaling*:

$$x_{ij} = \frac{x_{ij}^0 - \bar{x}_j}{s_j} \quad (5)$$

A Equação (6) apresenta o método de Normalização, utilizado por Steffen (2017) e Steffen e Gomes (2018), cujos resultados são comparados aos métodos empregados no presente estudo:

$$x_{ij} = \frac{x_{ij}^0 - x_{j,min}^0}{x_{j,max}^0 - x_{j,min}^0} \quad (6)$$

Nas Equações de (1) a (6),  $x_{ij}$  é o valor tratado da  $i$ -ésima observação do  $j$ -ésimo indicador;  $x_{ij}^0$  é o valor amostral (não tratado) da  $i$ -ésima observação do  $j$ -ésimo indicador;  $\bar{x}_j$  é a média amostral do  $j$ -ésimo indicador;  $s_j$  é o desvio-padrão amostral do  $j$ -ésimo indicador;  $x_{j,max}^0$  e  $x_{j,min}^0$  são, respectivamente, os valores máximos e mínimos amostrais do  $j$ -ésimo indicador;  $\sqrt{s_j}$  é a raiz quadrada do desvio-padrão amostral do  $j$ -ésimo indicador;  $\frac{\bar{x}_j}{s_j}$  corresponde ao coeficiente de variação amostral do  $j$ -ésimo indicador.

### Agrupamento dos dados

O agrupamento de dados por meio do método *fuzzy c-means (fcm)* é apresentado em detalhes por Steffen (2017), sendo representado resumidamente pelas seguintes etapas:

1. Arbitra-se o número de grupos ( $K$ ) e o parâmetro de incerteza do método ( $r$ );
2. Estima-se aleatoriamente a matriz de partição difusa inicial;
3. Calcula-se a matriz de centroides dos grupos;
4. Calcula-se a matriz de distâncias Euclidianas dos dados amostrais aos centroides dos grupos;
5. Calcula-se o valor da função objetivo do método;
6. Atualiza-se a matriz de partição difusa;
7. Compara-se a matriz de partição difusa do passo em execução (matriz final) com a matriz de partição difusa do passo anterior (matriz inicial);
8. Calcula-se o erro existente entre os elementos das matrizes de partição difusa inicial e final;
9. Compara-se o erro calculado com o erro tolerável, adotado como igual a  $10^{-5}$ ;
10. Volta-se à etapa 3, caso o erro calculado seja maior que o tolerável, ou conclui-se o algoritmo, caso contrário.

É importante ressaltar que, apesar da definição do número de grupos ( $K$ ) e do grau de incerteza ( $r$ ) ser uma etapa essencial para obtenção do agrupamento, ela não faz parte do processo iterativo do

*fcm*, ou seja, é necessário definir estes parâmetros previamente, pois o *fcm* visa somente definir os elementos constituintes (eventos de cheia) de cada grupo. No entanto, como não se conhece inicialmente os valores de  $K$  e  $r$ , o que se faz, na prática, é adotar-se valores de  $K$  e  $r$  e realizar o agrupamento diversas vezes, buscando estabelecer  $K$  e  $r$  que produzem o agrupamento “ótimo”.

No presente estudo, adotou-se os parâmetros  $F_c$ ,  $(1 - F_c)$  e  $H_c$  para a definição do agrupamento “ótimo”, mesmos parâmetros usados em Steffen (2017). O agrupamento ótimo corresponde aos valores de  $K$  e  $r$  que produzem o valor máximo de  $F_c$  e os valores mínimos de  $(1 - F_c)$  e  $H_c$ . A forma de cálculo dos referidos parâmetros pode ser encontrada em Steffen (2017) e Steffen e Gomes (2018).

Por fim, o que determina a qual grupo cada evento pertence é a matriz de partição difusa que estabelece um nível de pertencimento (variando de zero a 1) de cada evento a cada um dos grupos, em maior ou menor grau, ou seja, cada evento tem similaridade com todos os grupos, porém é considerado pertencente ao grupo com o qual apresenta o maior (máximo) grau de pertencimento.

### Análise comparativa dos agrupamentos de cheias

Após a definição dos grupos, usando o método *fcm* e os diferentes métodos de pré-tratamento, efetuou-se uma análise comparativa dos resultados obtidos, inclusive, com os resultados apresentados em Steffen (2017), que utilizou o método de normalização como forma de tratamento inicial.

## RESULTADOS

Para cada método de tratamento inicial das informações, foram definidos o grau de incerteza ( $r$ ) e o número de grupos ( $K$ ), a partir dos parâmetros  $F_c$ ,  $(1 - F_c)$  e  $H_c$ , cujos resultados são apresentados na Tabela 1.

Tabela 1 – Definição do grau de incerteza ( $r$ ) e do número de grupos ( $K$ )

Método	Máximo $F_c$	Mínimo $(1 - F_c)$	Mínimo $H_c$	Grau de incerteza ( $r$ )	Número de grupos ( $K$ )
<i>Autoscaling</i>	0,9478	0,0522	0,0911	1,25	3
<i>Range scaling</i>	0,9358	0,0642	0,1092	1,25	3
<i>Pareto scaling</i>	0,9739	0,0261	0,0504	1,25	3
<i>Vast scaling</i>	0,9770	0,0230	0,0447	1,25	4
<i>Level scaling</i>	0,9515	0,0485	0,0954	1,25	4
<b>Normalização*</b>	<b>0,9317</b>	<b>0,0683</b>	<b>0,1274</b>	<b>1,25</b>	<b>4</b>

Obs: \* Resultados obtidos por Steffen (2017).

A partir da Tabela 1, observa-se que os valores máximos de  $F_c$  são sempre superiores a 0,90 e, conseqüentemente, os mínimos de  $(1 - F_c)$  e  $H_c$  aproximam-se de zero. Por isso, admite-se que os graus de incerteza e os números de grupos obtidos são aceitáveis. Nota-se que, para todos os métodos, o grau de incerteza ( $r$ ) ideal foi de 1,25, porém houve variação do número de grupos ( $K$ ) ideal, com metade dos métodos sugerindo 3 grupos e a outra metade sugerindo 4 grupos.

Procurou-se avaliar também a matriz de partição difusa obtida ao final do processo de agrupamento, a fim de identificar que método inicial de tratamento produziu o melhor agrupamento.

A matriz de partição difusa estabelece o grau de pertencimento dos eventos a cada um dos grupos. O valor do grau de pertencimento de um evento a um determinado grupo varia entre 0 (zero) e 1, e quanto mais próximo de 1, mais fortemente um evento pertence a um determinado grupo. Desta forma, foram analisados os máximos graus de pertencimento de cada evento aos grupos definidos pelo *fcm*, considerando-se os agrupamentos definidos pelos parâmetros apresentados na Tabela 1.

A Tabela 2 apresenta uma análise estatística dos máximos graus de pertencimento obtidos a partir das matrizes de partição difusa dos agrupamentos definidos pelos parâmetros apresentados na Tabela 1, a fim de avaliar os níveis máximos de similaridade dos eventos aos grupos.

Tabela 2 – Análise estatística dos máximos graus de pertencimento das matrizes de partição difusa do *fcm*

Método	Mínimo	Máximo	Amplitude	Média	Desvio-padrão
<i>Autoscaling</i>	0,5234	1,0000	0,4766	0,9633	0,0982
<i>Range scaling</i>	0,5179	1,0000	0,4821	0,9524	0,1172
<i>Pareto scaling</i>	0,6601	1,0000	0,3399	0,9845	0,0470
<i>Vast scaling</i>	0,5913	1,0000	0,4087	0,9851	0,0571
<i>Level scaling</i>	0,6960	1,0000	0,3040	0,9714	0,0625
<b>Normalização*</b>	<b>0,4086</b>	<b>1,0000</b>	<b>0,5914</b>	<b>0,9520</b>	<b>0,1197</b>

Obs: \*Análise realizada a partir dos resultados obtidos por Steffen (2017).

Observa-se, a partir da Tabela 2, que os mínimos valores dos máximos graus de pertencimento variaram entre aproximadamente 0,40 e 0,70. O método *level scaling* apresentou o maior valor (0,6960), enquanto o método de Normalização, usado em Steffen (2017) e Steffen e Gomes (2018), apresentou o pior valor (0,4086). Como consequência, os métodos *level scaling* e normalização apresentaram respectivamente a maior e a menor amplitude dos máximos graus de pertencimento.

Analisando-se os valores médios dos máximos graus de pertencimento, verifica-se que estes valores variaram entre aproximadamente 0,95 e 0,98, o que indica um alto grau de pertencimento dos eventos aos grupos, em termos médios. O melhor desempenho é obtido pelo método *vast scaling* (0,9851), com o método *Pareto scaling* apresentando um valor muito próximo (0,9845). O pior desempenho é novamente apresentado pelo método da normalização (0,9520), contudo muito próximo do desempenho apresentado pelo método *range scaling* (0,9524).

Ainda em relação à Tabela 2, o desvio-padrão dos máximos graus de pertencimento variou entre 0,04 e 0,12 aproximadamente, sendo que os métodos *Pareto scaling*, *vast scaling* e *level scaling*, nesta ordem, tiveram os menores desvios-padrão, também muito próximos entre si. O maiores valores de desvio-padrão foram obtidos respectivamente pelos métodos de normalização e *range scaling*. Pequenos valores de desvio-padrão são desejáveis, principalmente se forem verificados altos valores médios de graus de pertencimento, como o observado no presente trabalho.

A Tabela 3 apresenta as frequências relativas dos máximos graus de pertencimento, segundo classificação em 4 intervalos de classe (0,40-0,60; 0,60-0,80; 0,80-0,90 e 0,90-1) e considerando-se os 85 eventos utilizados neste estudo.

Tabela 3 – Frequência relativa (%) dos máximos graus de pertencimento

Método	Classes de máximo grau de pertencimento			
	0,40 – 0,60	0,60 – 0,80	0,80 – 0,90	0,90 – 1
<i>Autoscaling</i>	3,5	3,5	4,7	88,3
<i>Range scaling</i>	4,7	4,7	2,3	88,3
<i>Pareto scaling</i>	0,0	1,2	3,5	95,3
<i>Vast scaling</i>	1,2	1,2	1,2	96,4
<i>Level scaling</i>	0,0	4,7	3,5	91,8
<b>Normalização*</b>	<b>4,7</b>	<b>1,2</b>	<b>8,2</b>	<b>85,9</b>

Obs: \*Análise realizada a partir dos resultados obtidos por Steffen (2017).

A partir dos resultados apresentados na Tabela 3, verifica-se que o método *vast scaling* apresentou a maior frequência relativa (96,4%) para o intervalo de valores de grau de pertencimento superior a 0,90, seguido pelo método *Pareto scaling* (95,3%). O método de normalização apresentou a menor frequência relativa (85,9%) para o referido intervalo.

Os resultados obtidos mostraram que, de modo geral, os métodos *vast scaling* e *Pareto scaling* resultaram nos melhores desempenhos, considerando-se a análise dos máximos graus de pertencimento, extraídos das matrizes de partição difusa. Por outro lado, o método da normalização, usado por Steffen (2017) e Steffen e Gomes (2018), apresentou o pior desempenho, comparativamente aos demais métodos.

A Tabela 4 mostra os agrupamentos obtidos pelos métodos *vast scaling* e *Pareto scaling*, bem como, o agrupamento produzido por Steffen (2017), usando o método da normalização. Na referida tabela, os grupos estão numerados em ordem crescente de severidade. O Grupo 1 contém os eventos menos severos, enquanto o Grupo 4 contém os eventos mais severos. Em destaque, estão os eventos de cheia que, nos métodos *vast scaling* e *Pareto scaling*, apresentam classificação distinta da obtida pelo método da normalização. No caso específico do método *Pareto scaling*, por apresentar número diferente de grupos, considerou-se o Grupo 2 equivalente aos Grupos 2 e 3 da normalização e o Grupo 3 equivalente ao Grupo 4 da Normalização.

A comparação dos agrupamentos obtidos pelos métodos de normalização e *vast scaling*, que resultaram em 4 grupos, mostra que a maior diferença se encontra no modo como os métodos distribuíram os eventos nos Grupos 2 e 3. O método *vast scaling* concentrou os eventos no Grupo 2, enquanto o método da normalização distribuiu os eventos de modo mais equitativo entre os Grupos 2 e 3. Vale ainda destacar que o método *vast scaling* resultou em um Grupo 1 com eventos exclusivamente sem extravasamento, enquanto o método de normalização apresentou um evento com extravasamento em seu Grupo 1, justamente o evento do ano de 1931, classificado no Grupo 2 pelo método *vast scaling*. Ainda, o método *vast scaling* produziu um Grupo 4 com somente 4 eventos em comparação ao método da normalização que resultou em 6 eventos no mesmo grupo. No método *vast scaling*, os eventos dos anos 1957 e 1998 foram classificados no Grupo 3.

Tabela 4 – Agrupamentos de cheias gerados pelos métodos *vast scaling*, *Pareto Scaling* e normalização

	Grupo	n	Ano da cheia
<i>Vast scaling</i>	1	32	1933, 1934, 1936, 1940, 1941, 1942, 1943, 1944, 1945, 1949, 1951, 1952, 1956, 1958, 1959, 1960, 1962, 1963, 1964, 1967, 1968, 1969, 1974, 1977, 1978, 1985, 1986, 1991, 1994, 2002, 2003, 2006
	2	33	<b>1931</b> , 1932, 1937, 1939, 1947, 1948, 1950, 1953, 1955, 1961, 1965, 1966, 1970, 1972, 1973, 1975, 1976, 1979, 1980, 1981, 1984, 1988, 1989, 1996, 1999, 2001, 2004, 2005, 2007, 2008, 2009, 2012, 2015
	3	16	1938, 1946, 1954, <b>1957</b> , 1971, 1982, 1987, 1990, 1993, 1995, 1997, <b>1998</b> , 2000, 2010, 2011, 2013
	4	4	1935, 1983, 1992, 2014
<i>Pareto Scaling</i>	1	44	1931, 1933, 1934, 1936, <b>1937</b> , 1940, 1941, 1942, 1943, 1944, 1945, <b>1948</b> , 1949, 1951, 1952, <b>1953</b> , 1956, 1958, 1959, 1960, 1962, 1963, 1964, <b>1965</b> , <b>1966</b> , 1967, 1968, 1969, <b>1970</b> , 1974, <b>1976</b> , 1977, 1978, <b>1979</b> , 1985, 1986, 1991, 1994, <b>1996</b> , 2002, 2003, <b>2004</b> , 2006, <b>2008</b>
	2	37	1932, 1938, 1939, 1946, 1947, 1950, 1954, 1955, <b>1957</b> , 1961, 1971, 1972, 1973, 1975, 1980, 1981, 1982, 1984, 1987, 1988, 1989, 1990, 1993, 1995, 1997, <b>1998</b> , 1999, 2000, 2001, 2005, 2007, 2009, 2010, 2011, 2012, 2013, 2015
	3	4	1935, 1983, 1992, 2014
Normalização*	1	33	1931, 1933, 1934, 1936, 1940, 1941, 1942, 1943, 1944, 1945, 1949, 1951, 1952, 1956, 1958, 1959, 1960, 1962, 1963, 1964, 1967, 1968, 1969, 1974, 1977, 1978, 1985, 1986, 1991, 1994, 2002, 2003, 2006
	2	22	1937, 1939, 1948, 1953, 1961, 1965, 1966, 1970, 1973, 1975, 1976, 1979, 1980, 1981, 1988, 1996, 1999, 2004, 2007, 2008, 2011, 2012
	3	24	1932, 1938, 1946, 1947, 1950, 1954, 1955, 1971, 1972, 1982, 1984, 1987, 1989, 1990, 1993, 1995, 1997, 2000, 2001, 2005, 2009, 2010, 2013, 2015
	4	6	1935, 1957, 1983, 1992, 1998, 2014

Obs: \* Resultado obtido por Steffen (2017).

A comparação dos agrupamentos obtidos pelos métodos de normalização e *Pareto scaling*, que resultaram em 4 grupos e 3 grupos, respectivamente, mostra que metade dos eventos que pertenciam ao Grupo 2 no método da normalização passaram a ser classificados como pertencentes ao Grupo 1 no método *Pareto scaling*, que contém, portanto, eventos sem e com extravasamento. O Grupo 2 do método *Pareto scaling* é composto basicamente pelos eventos classificados nos Grupos 2 e 3 do método da normalização, excluindo-se os 11 eventos que passaram a ser classificados no Grupo 1. Ainda, o método *Pareto scaling* produziu um Grupo 3 igual ao Grupo 4 do método *vast scaling*.

## CONCLUSÃO

Os resultados mostraram que os diferentes métodos de tratamento inicial das informações produziram agrupamentos distintos no método *fuzzy c-means*. Contudo, observando-se unicamente os máximos graus de pertencimento, obtidos da matriz de partição difusa, entende-se que todos os métodos são aplicáveis devido aos altos valores dos máximos graus de pertencimento observados.

Os métodos *autoscaling*, *range scaling* e *Pareto scaling* resultaram em uma classificação em 3 grupos distintos, enquanto os métodos *vast scaling*, *level scaling* e normalização, usado por Steffen (2017) e Steffen e Gomes (2018), resultaram em uma classificação em 4 grupos. Analisando-se mais profundamente o resultado dos agrupamentos gerados, verificou-se que o Grupo 3 nos métodos

*autoscaling*, *range scaling* e *Pareto scaling* praticamente equivale ao Grupo 4 obtidos nos demais métodos. Por outro lado, o Grupo 1 obtido pelos métodos *vast scaling*, *level scaling* e normalização foi composto majoritariamente por eventos sem extravasamento, enquanto o Grupo 1 produzido pelos métodos *autoscaling*, *range scaling* e *Pareto scaling* resultou em uma combinação de eventos com e sem extravasamento. De modo geral, o Grupo 2 dos métodos *autoscaling*, *range scaling* e *Pareto scaling* resultou da combinação dos Grupos 2 e 3 dos demais métodos, excluindo-se os eventos que passaram a ser classificados no Grupo 1 dos referidos métodos.

Embora os resultados de todos os métodos avaliados tenham sido bons, a análise estatística dos máximos graus de pertencimento mostrou um desempenho superior dos métodos *vast scaling* e *Pareto scaling* em relação ao demais. No entanto, para o presente caso, a recomendação seria o uso do método *vast scaling*, principalmente pelo fato de conter no grupo de eventos menos críticos, Grupo 1, somente eventos sem extravasamento, o que parece mais razoável em se tratando dos eventos de cheia e seus impactos observados no local de estudo.

## AGRADECIMENTOS

À CAPES pela concessão de bolsa de Doutorado ao primeiro autor.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ÁGUASPARANÁ – INSTITUTO DAS ÁGUAS DO PARANÁ. Disponível em: <<http://www.aguasparana.pr.gov.br>>. Acesso em: 19 jul. 2016.
- CASTANHARO, G.; BUBA, H. (2008). Revisão do comportamento hidráulico do rio Iguaçu na região de União da Vitória e seu impacto na operação hidráulica dos reservatórios de Foz do Areia e Segredo. *Espaço Energia*, n. 8, pp. 28-38.
- MINGOTI, S. A. (2005). Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. Editora UFMG: Belo Horizonte.
- STEFFEN, P. C. (2017). Agrupamento de cheias históricas. Dissertação (Mestrado em Engenharia de Recursos Hídricos e Ambiental). 178 f. Universidade Federal do Paraná, Curitiba.
- STEFFEN, P. C.; GOMES, J. (2018). Clustering of historical floods observed on Iguaçu River, in União da Vitória, Paraná. *Revista Brasileira de Recursos Hídricos*, v. 23, pp. 1-12.
- TUCCI, C. E. M; BERTONI, J. C. (2003). Inundações urbanas na América do Sul. ABRH – Associação Brasileira de Recursos Hídricos: Porto Alegre.
- VAN DEN BERG, R. A.; HOESFLOOT, H. C. J.; WESTERHUIS, J. A.; SMILDE, A. K.; VAN DER WERF, M. J. (2006). Centering, scaling, and transformations: improving the biological information content of metabolomics. *BMC Genomics*, v. 7.
- WANG, L. N.; CHEN, X. H.; SHAO, Q. X.; LI, Y. (2014). Flood indicators and their clustering features in Wujiang River, South China. *Ecological Engineering*, v. 76, pp. 66-74.