

## IMPACTO DA APLICAÇÃO DA TÉCNICA DE AGREGAÇÃO NA CONSTRUÇÃO DA ÁRVORE DE CENÁRIOS PARA O PLANEJAMENTO DA OPERAÇÃO DE CURTO PRAZO

*Felipe Treistman<sup>1\*</sup> & Débora Dias Jardim Penna<sup>2</sup> & Hugo Santarém de Araújo<sup>3</sup> & Maria Elvira Piñeiro Maceira<sup>4</sup> & Jorge Machado Damázio<sup>5</sup>*

**Resumo** – O planejamento da operação de sistemas hidrotérmicos tem como principal incerteza as afluições. Para avaliar os diversos critérios probabilísticos adotados, são usados cenários sintéticos de vazões que reproduzem corretamente as estatísticas do registro histórico. Na construção da árvore de cenários para o planejamento da operação de curto prazo brasileiro aplica-se a técnica de agregação K-Means, de forma a reduzir a dimensão do problema, preservando a qualidade de representação do processo estocástico das afluições. Este artigo tem como objetivo fazer um estudo comparativo entre a aplicação do método de agrupamento diretamente nos cenários de vazões ou nos ruídos que irão ser utilizados para a geração dos cenários de vazões. Na segunda opção é possível realizar uma correção do desvio padrão da amostra agregada, o que resulta em uma melhor reprodução dessa característica. Conclui-se que os cenários de afluição gerados com a aplicação do processo de agregação nos ruídos representam melhor a variabilidade condicionada, e consequentemente o processo estocástico das afluições.

**Palavras-Chave** – geração de cenários sintéticos; técnicas de agregação; planejamento da operação de curto prazo

## CLUSTERING TECHNIQUES IMPACT ASSESSMENT IN THE CONSTRUCTION OF THE SCENARIO TREE APPLIED TO THE SHORT- TERM OPERATION PLANNING

**Abstract** – The operation planning of hydrothermal systems has the streamflow as the main uncertainty. To evaluate the various probabilistic criteria adopted, streamflow scenarios must be generated in order to reproduce the statistics of the historical streamflow record. After the construction of a scenario tree for the Brazilian short-term planning, a clustering technique is applied to reduce the dimensionality problem, while still preserving the quality of the streamflow stochastic process. This paper intends to compare the application between the clustering techniques directly to the streamflow scenarios or to the noise that will be incorporated into the stochastic process representation. The second option has the advantage of correcting the standard deviation of the aggregated sample through the application of a correction factor, resulting in a better representation of this statistic. It is concluded that streamflow scenarios generated by the clustering technique applied to the noise produces a superior representation of the conditional variability, and consequently, of the streamflow stochastic process.

**Keywords** – synthetic scenario generation; clustering techniques; short-term operation planning

<sup>1</sup>Centro de Pesquisas de Energia Elétrica – CEPEL – felipe\_contrato.puc@cepel.br

<sup>2</sup>Centro de Pesquisas de Energia Elétrica – CEPEL – debora@cepel.br

<sup>3</sup>Centro de Pesquisas de Energia Elétrica – CEPEL – hugosa@cepel.br

<sup>4</sup>Centro de Pesquisas de Energia Elétrica – CEPEL, Universidade Estadual do Rio de Janeiro - UERJ – elvira@cepel.br

<sup>5</sup>Centro de Pesquisas de Energia Elétrica – CEPEL, Universidade Estadual do Rio de Janeiro - UERJ – damazio@cepel.br

## INTRODUÇÃO

O Sistema Interligado Nacional – SIN – caracteriza-se por ser um sistema hidrotérmico de grande porte, cujo planejamento se defronta com forte acoplamento espacial e temporal entre as usinas hidrelétricas, tendo como principal fonte de incerteza as afluições hidrológicas futuras. Devido a sua alta complexidade, o planejamento da operação do SIN é dividido em diferentes etapas que consideram distintos horizontes de tempo e graus de detalhamento da representação do SIN, sendo resolvido através da cadeia de modelos desenvolvida pelo Centro de Pesquisas de Energia Elétrica – CEPEL, Maceira *et al* (2002).

A modelagem estocástica das afluições é empregada para representar a principal incerteza desse problema, tanto na etapa de planejamento de curto prazo com a construção de uma árvore de cenários, quanto na etapa de médio/longo prazo através da geração de cenários em paralelo, capturando as principais características das séries históricas de vazões. Atualmente, o modelo oficial do setor elétrico em ambas etapas do planejamento para a geração de cenários sintéticos de energia e vazões – GEVAZP, Maceira *et al* (2005)- utiliza um modelo autorregressivo periódico mensal.

Sendo um problema de grande porte, torna-se interessante trabalhar com o menor número de cenários possíveis, porém sem perda de qualidade na representação do comportamento hidrológico. Para isso são utilizadas técnicas de agregação na geração de cenários de afluição. Atualmente, o modelo GEVAZP faz o uso do método de agrupamento K-Means. Esse artigo tem como objetivo fazer uma análise comparativa entre a aplicação da técnica de agregação após a etapa de sorteio dos ruídos do modelo e após o cálculo das vazões, para a construção da árvore de cenários para o planejamento da operação de curto prazo.

## METODOLOGIA

### Modelo Autorregressivo Periódico

Séries hidrológicas mensais possuem um comportamento periódico de suas propriedades probabilísticas, como por exemplo, média, variância, assimetria e estrutura de autocorrelação. Modelos autorregressivos periódicos conseguem capturar efetivamente as características da série histórica com o intuito de gerar cenários sintéticos de afluição.

Estes modelos, referenciados como PAR(p), Box & Jenkins (1970), onde p é a ordem do modelo, podem ser matematicamente descritos por:

$$\left( \frac{Z_t - \mu_m}{\sigma_m} \right) = \phi_1^m \left( \frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \phi_{p_m}^m \left( \frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + a_t \quad (1)$$

onde:

$Z_t$  é uma série sazonal de período s;

N é o número de anos;

t é o índice do tempo,  $t = 1, 2, \dots, sN$ , função do ano T ( $T = 1, 2, \dots, N$ ) e do período m ( $m = 1, 2, \dots, s$ );

$\mu_m$  é a média sazonal de período s;

$\sigma_m$  é desvio-padrão sazonal de período s;

s é o número de períodos ( $s = 12$  para séries mensais);

$\phi^m(B)$  é o operador autorregressivo de ordem  $p_m$ ;

$p_m$  é a ordem do operador autorregressivo do período  $m$ ;  
 $a_t$  série de ruídos independentes com média zero e variância  $\sigma_a^{2(m)}$ ;

### Técnica de Agregação K-Means

No planejamento da operação de curto prazo conduzido pelo Operador Nacional do Sistema Elétrico no PMO e suas revisões, realizado pelo modelo DECOMP, Xavier *et al.*(2005), a representação da incerteza hidrológica é feita através de uma árvore de cenários de aflúências, como ilustrado na Figura 1.

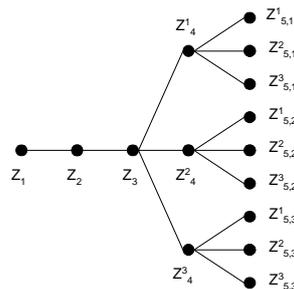


Figura 1 – Estrutura de Árvore

Para correta representação do processo estocástico de aflúências, é necessário que essa árvore possua muitas aberturas descrevendo adequadamente as características extraídas da série histórica. Porém, a utilização de um número muito grande de cenários poderia tornar a solução do problema de planejamento de curto prazo inviável computacionalmente. Uma forma de contornar esse problema é através da aplicação de alguma técnica de agregação. Essas técnicas têm como objetivo reduzir o número de cenários simulados através da escolha de um conjunto representativo segundo uma métrica.

Atualmente o GEVAZP faz uso do método de agregação não hierárquico K-Means. Para esta aplicação, uma alternativa é agrupar os vetores de vazões naturais afluentes sintéticas compostos pelas vazões geradas para cada usina hidroelétrica considerada na configuração do sistema hidroelétrico brasileiro. Estes vetores de vazões são agregados de maneira multivariada conforme descrito em Penna *et al.* (2001) e tendem a preservar as características do registro histórico, tais como a média e desvio padrão. Como métrica de similaridade entre os objetos (cenários hidrológicos multivariados) e os centróides do grupo é utilizada a distância Mahalanobis, de forma a preservar a correlação espacial entre as usinas hidroelétricas (UHEs). A correta reprodução das correlações espaciais entre as UHEs é de extrema importância para a geração de cenários sintéticos de aflúências utilizados no planejamento do SIN, por conta de sua complementaridade hídrica. Para a agregação dos vetores de vazões as variáveis devem ser previamente padronizadas para que vazões com diferentes ordens de grandeza não prejudiquem processo de agrupamento dos vetores de vazões.

Atualmente, a cada estágio do horizonte de estudo e a cada ramo da árvore de aflúências é gerada uma amostra com um número grande de cenários hidrológicos e equiprováveis para o processo de agregação, chamada de amostra original. Após o processo de agregação, os cenários selecionados possuem probabilidades distintas de ocorrência, proporcional ao número de cenários que representa. O processo de agregação está ilustrado na Figura 2.

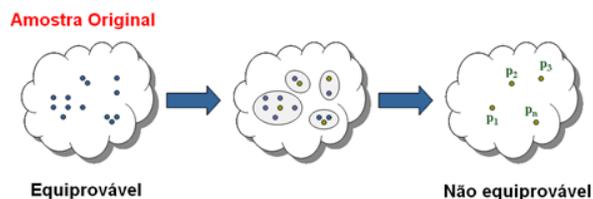


Figura 2 – Processo de agregação

O número de aberturas em cada estágio é o número de agrupamentos obtidos a partir da amostra original. Para cada mês do ano, existem valores pré-determinados para o número de aberturas a serem realizadas em estudos oficiais do Programa Mensal de Operação (PMO), definidos conforme a Tabela 1:

Tabela 1 – Número de aberturas para a árvore de cenários sintéticos para cada mês do PMO

	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
N° de aberturas	136	116	143	143	193	267	513	353	393	259	228	153

Outra forma de se aplicar o método de agregação é a utilização do mesmo nos vetores de ruídos normais não correlacionados. Neste caso a métrica de similaridade se resume a própria distância Euclidiana. Além disso, não é necessário fazer o processo de padronização, dado que os ruídos normais sorteados para todas as usinas possuem a mesma média (zero) e o mesmo desvio padrão (um).

### Fator de correção do desvio padrão agregado

No processo de agregação dos ruídos independentes, o desejado é obter amostras de ruídos normais com desvio padrão igual a um. O desvio padrão dos ruídos normais que compõe a amostra original é, devido ao grande tamanho da amostra, praticamente igual a um. Na agregação ocorre uma degradação do desvio padrão e para corrigir esta degradação propõe-se o cálculo do seguinte fator de correção:

$$FC = \frac{\sum_{i=1}^{N_{UHE}} \frac{\text{desvio padrão desejado}}{\sigma_i}}{N_{UHE}} = \frac{\sum_{i=1}^{N_{UHE}} \frac{1}{\sigma_i}}{N_{UHE}} \quad (2)$$

onde:

- $\sigma_i$  é o desvio padrão dos ruídos agregados para a usina  $i$ ;
- $N_{UHE}$  é o número total de usinas participantes do processo de agregação;
- $FC$  é o fator de correção do desvio padrão;

Este fator de correção é então aplicado a todos os ruídos agregados, para todas as usinas, tornando o desvio padrão da amostra agregada aproximadamente igual a um.

### Métricas de avaliação

Para analisar o impacto da aplicação da técnica de agregação K-Means nas vazões e nos ruídos, serão realizadas gerações de cenários sintéticos de vazões para cada usina do Programa PMO do ano de 2016. Cada caso possui em torno de cento e cinquenta UHEs. Em seguida serão apresentadas algumas métricas utilizadas para avaliar suas diferenças. Como a geração é realizada condicionada ao passado recente, definem-se os valores teóricos dos modelos como os valores condicionados:

$$E(Z_t | Z_{t-}) = \left[ \phi_1^m \left( \frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \phi_{p_m}^m \left( \frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) \right] \sigma_m + \mu_m \quad (3)$$

$$\sigma_{Z_t|Z_{t-}} = \sigma_m \sigma_a^{(m)} = \sigma_m \sqrt{1 - \phi_1^m \rho^m(1) - \phi_2^m \rho^m(2) - \dots - \phi_{p_m}^m \rho^m(p_m)} \quad (4)$$

onde:

$E(Z_t | Z_{t-})$  é a média condicionada do modelo autorregressivo ajustado;

$\sigma_{Z_t|Z_{t-}}$  é o desvio padrão condicionado do modelo autorregressivo ajustado;

De forma a avaliar os resultados encontrados, será analisado um índice definido como a razão das estatísticas agregadas pelos respectivos valores teóricos ( $\omega = \frac{\hat{\theta}_{ag}}{\theta_{vt}}$ ), para todas as usinas. Quanto mais próximo do valor unitário, melhor o processo de geração/agregação. Também será avaliada a reprodução das correlações espaciais entre as usinas. Para isso, será calculada a diferença entre a correlação espacial histórica e a correlação espacial sintética obtida após o processo de agregação ( $\Delta = \rho_{xy,h} - \rho_{xy,ag}$ ), para todas as usinas.

Serão apresentados os resultados para um mês pertencente ao período úmido do SIN (fevereiro) e um mês representante do período seco (agosto). Os gráficos do tipo Box-plot permitirão fazer uma análise da distribuição do índice  $\omega$  de todas as usinas avaliadas para um determinado mês, tanto para o caso de agregação das vazões, como para a agregação nos ruídos.

## RESULTADOS

Na Figura 3 é apresentado um gráfico Box-plot do índice  $\omega$  para a média de cada uma das usinas. Observa-se que o processo de agregação aplicado nas vazões e nos ruídos, apresentam distribuições semelhantes, tendo em geral valores muito próximos a um, demonstrando que ambos reproduzem corretamente os valores teóricos. Observa-se ainda que nos meses secos, que possuem uma amostra agregada maior, há uma menor dispersão dessa métrica. Esse resultado era esperado, dado que com um maior número de aberturas, é possível representar melhor o processo estocástico. A Tabela 2 apresenta a média das distribuições para todos os meses do ano. É possível verificar que apesar da agregação nas vazões já apresentar resultados médios próximos ao valor teórico, a agregação nos ruídos consegue ser ligeiramente superior nesse quesito ao longo de todo o ano.

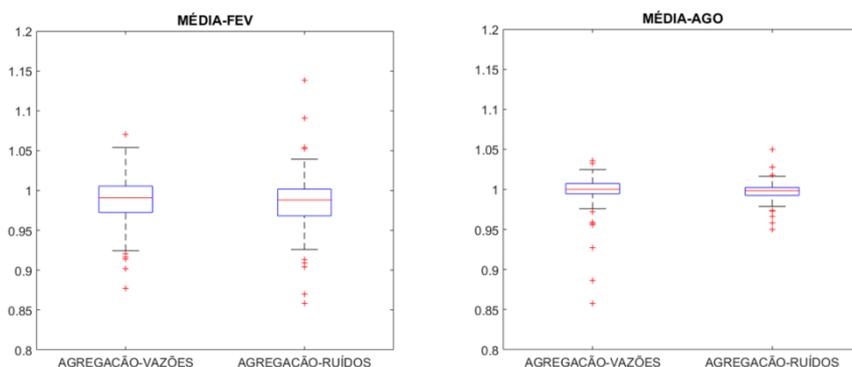


Figura 3 – Box-plot do índice  $\omega$  para avaliação da média, para os meses de fevereiro e agosto

Tabela 2 – Média do índice da avaliação das médias sintéticas

	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
Ag. Vazões	0.963497	0.968964	0.986534	0.982588	0.961528	0.982139	0.995696	0.995865	0.996592	0.980641	0.958697	0.972881
Ag. Ruídos	0.975184	0.96919	0.990704	0.988137	0.990653	0.995508	0.996402	0.997121	0.998343	0.999728	1.015613	0.985519

A seguir, é apresentado na Figura 4 os resultados para o desvio padrão para os meses selecionados anteriormente. É possível observar que as amostras agregadas nos ruídos possuem valores mais próximos ao valor teórico do desvio padrão. Encontra-se o mesmo comportamento para os demais meses do ano. Mais uma vez verifica-se o efeito do número de aberturas na dispersão apresentada no gráfico. Na Tabela 3 são apresentados os resultados para todos os meses da média do índice utilizado para avaliar o desvio padrão. Observa-se um ganho significativo na reprodução do desvio padrão teórico em todos os meses analisados, quando utilizada a técnica de agregação nos ruídos. Também é possível perceber que os meses do período úmido possuem um ganho maior. Esse fato está relacionado ao menor número de aberturas consideradas.

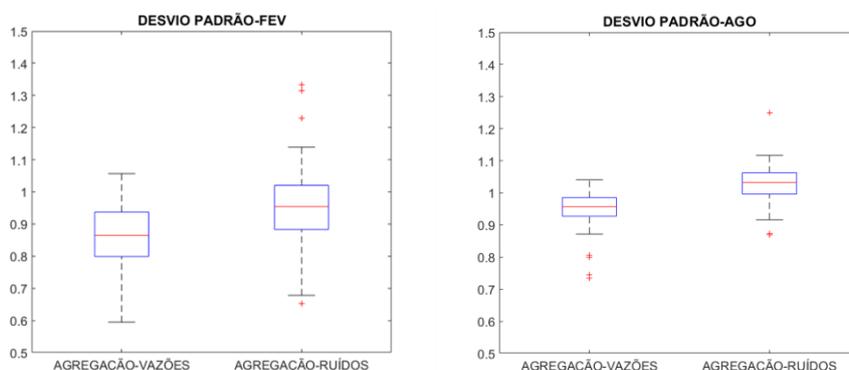


Figura 4 – Box-plot do índice  $\omega$  para avaliação do desvio padrão, para os meses de fevereiro e agosto

Tabela 3 – Média do índice da avaliação dos desvios padrão sintéticos

	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
Ag. Vazões	0.867347	0.853197	0.876125	0.869928	0.878739	0.928711	0.970766	0.950291	0.949227	0.90922	0.886543	0.878489
Ag. Ruídos	0.905644	0.927301	1.032449	1.027236	0.95453	0.985183	1.006455	1.026794	1.025992	0.994384	1.00926	0.967715

Para identificar o efeito dessas alterações na geração dos cenários sintéticos de vazões, as Figuras 5a e 5b apresentam as distribuições acumuladas empíricas das usinas de Tucuruí e Água Vermelha para o mês de fevereiro e as Figuras 5c e 5d para o mês de agosto. Pode-se observar que os cenários resultantes do processo de agregação nos ruídos possuem um desvio padrão maior (e mais próximo ao valor teórico), sua distribuição possui caudas mais longas, o que representa a geração de cenários mais severos de secas e cheias, principalmente para fevereiro.

A geração de cenários severos, principalmente cenários secos, é de grande relevância para o planejamento da operação de curto prazo. Analisando-se as distribuições, também é perceptível que o mês de agosto possui distribuições muito mais próximas entre as duas metodologias. Essa constatação deve-se ao fato de no mês de agosto a agregação ser feita para 353 aberturas, mais que o triplo do número de aberturas do mês de fevereiro. Além disso, os meses secos possuem um desvio padrão menor, o que pode ser observado também no formato das distribuições.

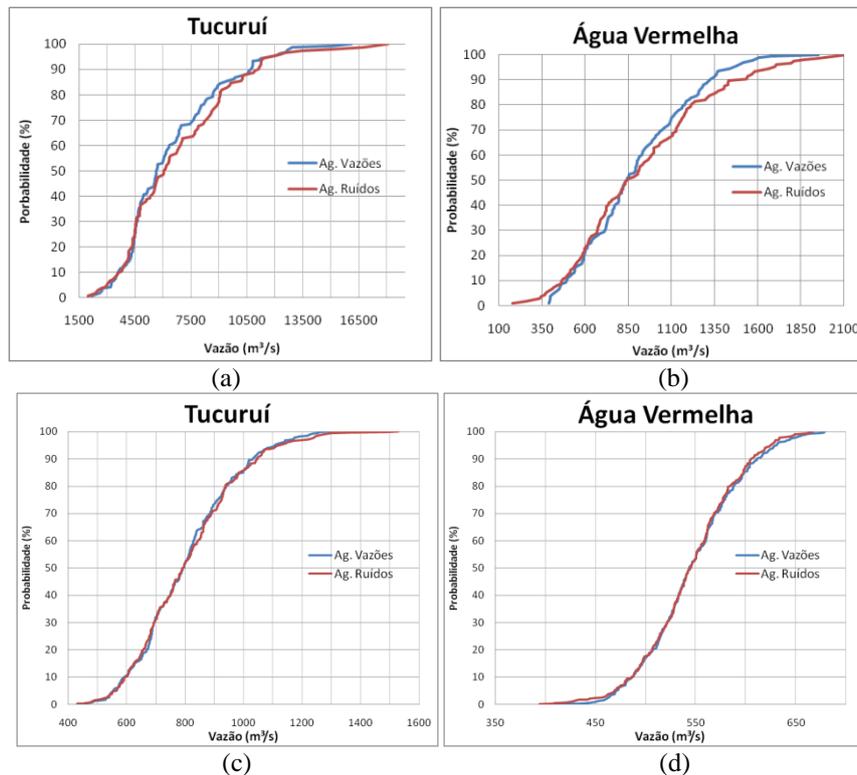


Figura 5 – Dist.acum.dos cenários gerados para as UHEs Tucuruí e Água Vermelha: (a) e (b) fevereiro, (c) e (d) agosto

A próxima análise, referente à Figura 6, apresenta as diferenças entre a correlação espacial dos cenários sintéticos e a correlação histórica entre as usinas para os meses de fevereiro e agosto. Observa-se que não há impacto nas correlações espaciais quando se utiliza a técnica de agregação nas vazões ou nos ruídos. A Tabela 4 apresenta a média dessas distribuições. É possível verificar que as correlações espaciais são reproduzidas satisfatoriamente em ambas as metodologias.

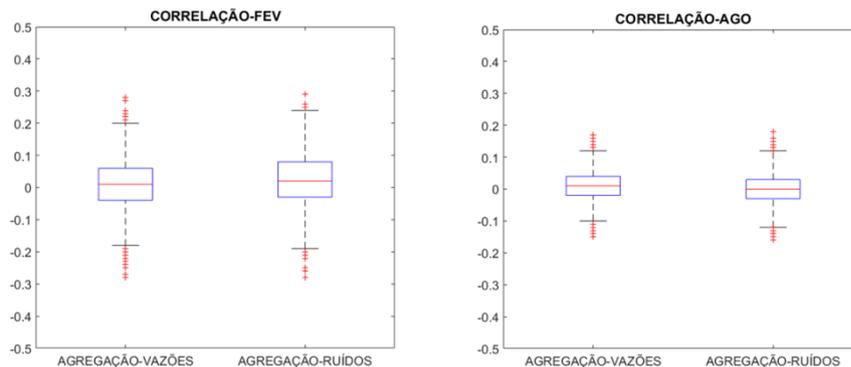


Figura 6 – Box-plot das diferenças entre a correlação cruzada histórica e a correlação cruzada agregada ( $\Delta$ )

Tabela 4 – Média das diferenças entre a correlação espacial histórica e agregada

	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
Ag. Vazões	0.003619	0.012714	0.008257	0.007372	0.014593	0.007433	0.007495	0.010373	0.017915	0.015174	0.011196	0.018855
Ag. Ruídos	0.030564	0.021288	-0.01654	-0.01675	0.012153	0.011873	0.008435	0.00257	0.001582	0.025824	0.021736	0.019651

Vale a pena destacar que a geração de cenários com agregação nos ruídos teve um tempo computacional que chegou a ser 400% menor em alguns casos quando comparado com a geração de cenários sintéticos com agregação nas vazões. Esse menor esforço computacional pode estar

relacionado ao fato de durante o processo de agregação de ruídos não ser necessário realizar a padronização das vazões, nem o cálculo da distância Mahalanobis. Adicionalmente, na agregação de vazões, é necessário realizar uma etapa adicional referente à geração da amostra original de vazões para a posterior agregação.

## CONCLUSÕES

Esse artigo teve o objetivo de analisar o impacto da aplicação da técnica de agregação K-Means na construção da árvore de cenários sintéticos de vazões utilizadas no modelo de planejamento energético de curto prazo. Foram avaliadas duas formas de aplicação dessa técnica: após a geração dos cenários de vazões, ou após o sorteio dos ruídos normais não correlacionados a serem utilizados na geração das vazões. Para a segunda aplicação propõe-se o cálculo, e posterior uso, de um fator de correção para o desvio padrão da amostra agregada, que em geral sofre alguma degradação durante o processo de agregação.

Foram gerados cenários sintéticos de afluência para os doze PMOs do ano de 2016, com a configuração completa do parque hidráulico brasileiro, com o intuito de avaliar as diferenças de resultados ao longo do ano entre as duas metodologias. Segundo as métricas utilizadas, demonstrou-se que houve um ganho significativo na reprodução do desvio padrão teórico quando realizada a agregações nos ruídos e aplicado o fator de correção proposto. As médias apresentaram pequeno ganho em todos os meses, enquanto a correlação espacial entre as usinas não apresentou alterações significativas. Nos meses em que o número de aberturas é maior, no período seco, a diferença entre as duas metodologias é menor, porém ainda é possível afirmar que a agregação nos ruídos gera resultados melhores. Adicionalmente, observou-se que a aplicação da técnica de agregação nos ruídos chegou a ter um tempo computacional 400% menor quando comparado com o processo de agregação nas vazões.

O principal aprimoramento dessa metodologia foi produzir cenários sintéticos de vazões com um desvio padrão mais próximo ao valor teórico condicionado. Com isso, podem-se obter cenários de vazões mais severos como pode ser observado pelas distribuições acumuladas empíricas. A geração de cenários severos, principalmente de secas, é de grande importância para o planejamento da operação de curto prazo. Sendo assim, espera-se que com a utilização dessa técnica obtenham-se ganhos na tomada de decisão do planejamento de curto prazo.

## REFERÊNCIAS

- BOX, G.E.P., JENKINS, G.M. (1970). Time Series Analysis - Forecasting and Control, Holden-Day.
- MACEIRA, M. P., TERRY, L. A., COSTA, F. S., DAMÁZIO, J. M., & MELO, A. C. G. (2002). Chain of optimization models for setting the energy dispatch and spot price in the Brazilian system. In Proceedings of the Power System Computation Conference-PSCC , Vol. 2, pp. 24-28.
- MACEIRA, M.E.M; PENNA, D.D.J.; DAMÁZIO, J.M. (2006). Geração de cenários sintéticos de energia e vazão para o planejamento da operação energética. Cadernos do IME – Série Estatística, 21(2), 11 p.
- PENNA, D.D.J.; MACEIRA, M.E.P.; FALCÃO, D.M. (2001). Stochastic Streamflow Model for Hydroelectric Systems Using Clustering Techniques, In Anais do IEEE Porto Power Tech Conference, vol.3, Porto, Portugal, Sep.
- XAVIER, L. N., DINIZ, A. L., COSTA, F. S., et al, (2005). Aprimoramento da modelagem da função de produção energética das usinas hidroelétricas no modelo DECOMP: metodologia e resultados, XVIII SNPTEEE - Seminário Nacional de Produção e Transmissão de Energia Elétrica, Curitiba, Out. 2005.