

# Análise da Correlação Cruzada Através da Distribuição Lognormal Três Parâmetros

Wellington Luis de Oliveira<sup>1</sup>; Débora Dias Jardim Penna<sup>2</sup>; Jorge Machado Damázio<sup>3</sup> & Maria Elvira Piñeiro Maceira<sup>4</sup>;

**RESUMO** – Na geração de séries sintéticas multivariadas por modelos autoregressivos, a estratégia comumente adotada é, inicialmente, sortear aleatoriamente ruídos independentes com uma dada distribuição de probabilidades, e em seguida aplicar uma transformação linear para torná-los correlacionados. Esta transformação linear é chamada de *Matriz de Carga* e é, em geral, obtida a partir da decomposição de Cholesky da matriz de correlações cruzadas histórica. Para a geração de séries sintéticas de vazões (ou energias) afluentes aos reservatórios utiliza-se uma transformação exponencial em ruídos normalmente distribuídos e correlacionados através da matriz de carga. Esta transformação exponencial tem como objetivo transformar os ruídos normais em ruídos com distribuição lognormal 3 parâmetros, de tal forma que se evite a geração de vazões negativas. Essa transformação pode degradar a correlação existente nos ruídos normais, interferindo assim na reprodução das correlações cruzadas históricas. Este trabalho tem o propósito de apresentar a metodologia de geração de cenários via a distribuição lognormal 3 parâmetros, e propor uma nova transformação linear (matriz de carga) para correlacionar os ruídos normais. Esta nova matriz de carga faz com que os ruídos com distribuição lognormal reproduzam com maior precisão a correlação histórica.

**ABSTRACT** – In the multivariate generation of synthetic series by autoregressive models, the strategy adopted is to sample independent noises with a given distribution of probabilities, and then apply a linear transformation to make them correlated. This linear transformation is called load matrix and is generally obtained from the Cholesky decomposition of the historical cross-correlations matrix. For the generation of streamflows it is needed to apply an exponential transformation on correlated normal noises to avoid negative values. This transformation may degrade the correlation existing in the normal noise, thus interfering with the historical cross-correlations reproduction. This paper aims to present the methodology used for generating synthetics series by the lognormal 3 parameters distribution, and propose a new load matrix to correlate the normal noises. This new load matrix makes the lognormal distributed noises reproduce more accurately the historical correlation.

**Palavras-chave:** Geração de séries sintéticas, correlação cruzada, distribuição lognormal 3 parâmetros.

---

<sup>1</sup> Pesquisador do Centro de Pesquisas de Energia Elétrica - CEPEL, Av. Horácio Macedo, 354 – Cidade Universitária. E-mail: wlo@cepel.br

<sup>2</sup> Pesquisadora do Centro de Pesquisas de Energia Elétrica. E-mail: debora@cepel.br

<sup>3</sup> Pesquisador do Centro de Pesquisas de Energia Elétrica e professor da Universidade do Estado do Rio de Janeiro. E-mail: damazio@cepel.br

<sup>4</sup> Pesquisadora do Centro de Pesquisas de Energia Elétrica e professora da Universidade do Estado do Rio de Janeiro. E-mail: elvira@cepel.br

## 1 – INTRODUÇÃO

O planejamento da operação hidrotérmica do Sistema Interligado Nacional - SIN utiliza séries sintéticas do processo estocástico de afluições aos aproveitamentos hidroelétricos obtidas pelo Modelo de Geração de Séries Sintéticas de Vazões e Energias – GEVAZP, desenvolvido pelo CEPEL, Maceira *et al.* (1997, 2000 e 2005). O modelo GEVAZP gera séries sintéticas de vazões mensais afluentes com um modelo autorregressivo periódico de ordem  $p \in \mathbb{N}^{12}$  denominado  $PAR(p)$ . Para maiores informações, ver Salas *et al.* (1980); Kelman *et al.* (1992) e Hipel & McLeod (1994).

As séries sintéticas geradas pelo modelo GEVAZP são utilizadas no Modelo Estratégico de Geração Hidrotérmica a Subsistemas Equivalentes Interligados – NEWAVE, Maceira *et al.* (1998), no Modelo de Simulação às Usinas Individualizadas para Subsistemas Hidrotérmicos Interligados – SUSHI-O, Maceira *et al.* (1995), e no Modelo de Determinação da Coordenação da Operação à Médio Prazo – DECOMP, Maceira *et al.* (2002), todos desenvolvidos pelo CEPEL e utilizados pelo Operador Nacional do Sistema Elétrico (ONS) para operar o SIN no curto e médio prazo, com exceção do modelo SUSHI-O.

Para cada mês  $m$  do ano, o GEVAZP estima a ordem  $p(m)$  do modelo e os coeficientes autorregressivos  $\Phi_j^m$ . Representando por  $Y_m^i$  a  $i$ -ésima afluição gerada para o mês  $m$ , e por  $Z_m^i$  o valor padronizado de  $Y_m^i$  (que corresponde à subtração de  $Y_m^i$  pela média e divisão pelo desvio-padrão históricos do mês  $m$ ), a geração dos cenários obedece a equação (1)

$$Z_m^i = \sum_{j=1}^{p(m)} \Phi_j^m Z_{m-j}^i + a_m^i, \quad (1)$$

onde  $a_m^i$  é um ruído aleatório extraído de uma distribuição lognormal 3 parâmetros com média  $\tilde{\mu}_m$ , variância  $\tilde{\sigma}_m^2$  e deslocamento  $\Delta_m$  (representa-se neste texto como  $a_m \sim LN(\tilde{\mu}_m, \tilde{\sigma}_m^2, \Delta_m)$ ). A incorporação do deslocamento na distribuição  $a_m \sim LN(\tilde{\mu}_m, \tilde{\sigma}_m^2, \Delta_m)$  tem como motivação evitar gerar séries sintéticas com ocorrências de afluições negativas, sendo modelado como função da tendência hidrológica representada pelo somatório do lado direito de (1).

Quando a geração de séries se dá com uma condição inicial desfavorável, a tendência hidrológica fornecida para a geração impacta fortemente a assimetria da distribuição lognormal do ruído. Nestas situações, a transformação lognormal 3 parâmetros impacta a reprodução da correlação cruzada histórica. Por este motivo, Damázio *et al.* (2009) propuseram uma heurística capaz de melhorar a reprodução das correlações cruzadas entre as energias afluentes aos subsistemas no planejamento da operação energética de médio prazo.

Pretende-se com este texto apresentar alguns resultados utilizados na geração de séries sintéticas de afluências pelo modelo GEVAZP, e propor uma nova matriz de carga para a geração de cenários espacialmente correlacionados. Diferentemente de *Damázio et al. (2009)*, a matriz de carga proposta é obtida a partir da função inversa que relaciona as correlações cruzadas das distribuições lognormais, com as correlações cruzadas das distribuições normais subjacentes.

Este trabalho está organizado como se segue: é tratada na Seção 2 a distribuição lognormal 3 parâmetros. A definição e estimação dos parâmetros desta distribuição são realizadas na Seção 3. A Seção 4 propõe uma nova transformação linear para correlacionar os ruídos e, finalmente, são apresentados nas Seções 5 e 6 alguns resultados numéricos e as principais conclusões do trabalho, respectivamente.

## 2 – DISTRIBUIÇÃO LOGNORMAL 3 PARÂMETROS

Nesta seção são apresentados os momentos de primeira e segunda ordem da distribuição univariada lognormal 3 parâmetros, bem como sua função de densidade. É considerada a correlação entre duas variáveis lognormais 3 parâmetros e a sua relação com a correlação das distribuições normais subjacentes.

*Definição 1.* Seja  $\Delta \in \mathbb{R}$  um parâmetro. Se  $X \equiv \ln(\tilde{X} - \Delta)$  tem distribuição normal com média  $\mu$  e variância  $\sigma^2$ , i.e.,  $X \sim N(\mu, \sigma^2)$ , então é dito que  $\tilde{X}$  tem distribuição lognormal 3 parâmetros. Escreve-se  $\tilde{X} \sim LN(\tilde{\mu}, \tilde{\sigma}^2, \Delta)$ .

Quando  $\Delta = 0$ , diz-se que  $\tilde{X}$  tem distribuição lognormal 2 parâmetros, dados por  $\tilde{\mu}$  e  $\tilde{\sigma}^2$ .

Outra maneira equivalente de definir uma distribuição lognormal 3 parâmetros  $\tilde{X}$  a partir de uma distribuição normal  $X$ , pode ser obtida fazendo  $\tilde{X} = e^X + \Delta$ .

De acordo com *Yevjevich (1972)*, a função de densidade de uma distribuição lognormal 3 parâmetros é dada pela Definição 2.

*Definição 2.* Se  $X \sim N(\mu, \sigma^2)$ , então a função de densidade da distribuição lognormal 3 parâmetros  $\tilde{X} = e^X + \Delta$  é dada por

$$f_{\tilde{X}}(\tilde{x}) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma(\tilde{x} - \Delta)} e^{-\frac{(\ln(\tilde{x} - \Delta) - \mu)^2}{2\sigma^2}} & \text{se } \tilde{x} > \Delta \\ 0, & \text{se } \tilde{x} \leq \Delta. \end{cases} \quad (2)$$

Os dois primeiros momentos da distribuição log-normal 3 parâmetros são dados pela Proposição 1.

*Proposição 1.* Se  $X = \ln(\tilde{X} - \Delta) \sim N(\mu, \sigma^2)$ , então  $\tilde{X} \sim LN(\tilde{\mu}, \tilde{\sigma}^2, \Delta)$ , com

$$\tilde{\mu} = e^{\mu + \frac{\sigma^2}{2}} + \Delta, \quad (3)$$

e

$$\tilde{\sigma}^2 = e^{2(\mu + \sigma^2)} - e^{2\mu + \sigma^2}. \quad (4)$$

A prova desta Proposição pode ser obtida em *Degrott e Scheruish (2002)*. Vale ressaltar que, por (3) e (4), o parâmetro de deslocamento  $\Delta$  não influencia a variância da distribuição lognormal  $\tilde{X}$  (somente influencia a média). A seguir passa-se ao estudo da correlação cruzada.

## 2.1 – Correlação entre duas variáveis lognormais e sua dependência com a correlação das distribuições normais

Uma das principais preocupações na geração de séries sintéticas multivariadas é a reprodução da correlação cruzada histórica. Mais especificamente para o planejamento hidrotérmico, a correlação cruzada determina a dependência hidrológica entre os aproveitamentos hidroelétricos. Por exemplo, aflúências às hidroelétricas na mesma bacia são positivamente correlacionadas.

No modelo GEVAZP a metodologia utilizada para a geração de séries de aflúências sintéticas correlacionadas entre si corresponde à execução do seguinte pseudo-algoritmo:

1. identificação e estimação dos parâmetros do modelo autorregressivo periódico -  $Par(p)$ ;
2. estimação dos momentos de primeira e segunda ordem das distribuições lognormal 3 parâmetros  $(\tilde{\mu}_i, \tilde{\sigma}_i^2)$ , onde  $i$  varia em 1 até o número de séries sintéticas correlacionadas que se deseja gerar; e cálculo do parâmetro de deslocamento  $\Delta_i$  em função das condições hidrológicas antecedentes;
3. cálculo dos momentos de primeira e segunda ordem da distribuição normal  $(\mu_i, \sigma_i^2)$ ;
4. geração aleatória de ruídos  $\xi^i$  independentes e normalmente distribuídos (*iid*) com média zero e variância um -  $\xi^i \sim N(0,1)$ ;

5. correlação dos ruídos  $\xi^i$  através de uma transformação linear chamada *Matriz de Carga* -  $M$ , obtida pela decomposição de Cholesky da matriz de correlações cruzadas histórica. Matematicamente,  $\bar{\xi} = \xi M$ ;
6. obtenção dos ruídos lognormais  $a^i = e^{\sigma \bar{\xi}^i + \mu} + \Delta$ ;
7. aplica-se o modelo  $Par(p)$  e os ruídos  $a^i$  como em (1), para gerar as séries sintéticas;
8. retorno ao Passo 3 para o próximo intervalo de tempo.

Nesta subseção são considerados os passos 5 e 6. Como a transformação não-linear  $a^i = e^{\sigma \bar{\xi}^i + \mu} + \Delta$  pode impactar a correlação dos ruídos  $\xi^i$ , o principal interesse deste trabalho é estudar as propriedades da matriz de carga a ser utilizada no passo 5 do pseudo-algoritmo, de modo a proporcionar ruídos lognormais  $a^i$  capazes de gerar a correlação histórica com maior precisão.

A seguir é estabelecida uma relação entre a correlação cruzada entre duas variáveis lognormais com a correlação cruzada entre duas variáveis normais subjacentes.

Sejam  $\tilde{X}_1 \sim LN(\tilde{\mu}_1, \tilde{\sigma}_1, \Delta_1)$  e  $\tilde{X}_2 \sim LN(\tilde{\mu}_2, \tilde{\sigma}_2, \Delta_2)$ . Segue que  $X_1 = \ln(\tilde{X}_1 - \Delta_1) \sim N(\mu_1, \sigma_1)$

e  $X_2 = \ln(\tilde{X}_2 - \Delta_2) \sim N(\mu_2, \sigma_2)$ . Seja também a correlação cruzada das variáveis normais  $X_1$  e  $X_2$  dada por  $\rho \equiv Corr(X_1, X_2)$ . Então, de acordo com *Matalas (1967)*, a correlação cruzada entre as variáveis lognormais 3 parâmetros é determinada por

$$\tilde{\rho} = \frac{e^{\rho \sigma_1 \sigma_2} - 1}{\sqrt{e^{\sigma_1^2} - 1} \sqrt{e^{\sigma_2^2} - 1}}. \quad (5)$$

Assim sendo, pode-se afirmar que a correlação cruzada de variáveis lognormal 3 parâmetros é função dos desvios-padrão e da correlação cruzada das variáveis normais, i.e.,

$\tilde{\rho} = g(\sigma_1, \sigma_2, \rho)$ , para  $g : \mathbb{R}^3 \rightarrow \mathbb{R}$  dada por

$$g(\sigma_1, \sigma_2, \rho) = \frac{e^{\rho \sigma_1 \sigma_2} - 1}{\sqrt{e^{\sigma_1^2} - 1} \sqrt{e^{\sigma_2^2} - 1}}. \quad (6)$$

No processo de geração de séries sintéticas pelo modelo GEVAZP, assume-se a hipótese de que o histórico de afluições possui distribuição lognormal. Por este motivo, torna-se interessante escrever a correlação cruzada  $\tilde{\rho}$  em função dos 3 parâmetros desta distribuição. A seção seguinte trata deste assunto.

### 3 – ESTIMAÇÃO DOS 3 PARÂMETROS DA DISTRIBUIÇÃO LOGNORMAL PELO MODELO GEVAZP

Na equação (1), o modelo GEVAZP considera

$$Z_m = \frac{Y_m - \hat{\mu}_m}{\hat{\sigma}_m},$$

onde  $\hat{\mu}_m$  e  $\hat{\sigma}_m^2$  são os estimadores da média e da variância do histórico de vazões (ou energias) afluentes  $\{Y_m\}_{ano=1}^N$ . Utilizando a formulação do modelo  $Par(p)$ , a geração de séries sintéticas de afluências obedece à equação

$$Y_m = \underbrace{\hat{\mu}_m + \hat{\sigma}_m \sum_{j=1}^{p(m)} \Phi_j^m \frac{(Y_{m-j} - \hat{\mu}_{m-j})}{\sigma_{m-j}}}_{PD_m} + \underbrace{\hat{\sigma}_m a_m}_{PA_m}, \quad (7)$$

onde  $PD_m$  é a parte determinística do modelo e  $PA_m$  é a parte aleatória.

As afluências sintéticas geradas devem ser não-negativas, i.e.,  $Y_m \geq 0$ . Para garantir a generalidade do desenvolvimento seguinte, assume-se um limite inferior  $L_m \geq 0$  para o qual se procura satisfazer a desigualdade  $Y_m \geq L_m$ . Isto fornece a condição que deve ser satisfeita pela variável  $a_m$ :

$$a_m \geq \frac{L_m - PD_m}{\hat{\sigma}_m}. \quad (8)$$

Sendo  $E(\cdot)$  o operador correspondente ao valor esperado, uma condição clássica que deve ser satisfeita pelos modelos autorregressivos é a seguinte:

$$E(a_m) = 0. \quad (9)$$

Pelas condições (8) e (9), a variável  $a_m$  necessita ter uma distribuição lognormal 3 parâmetros, i.e.,

$$a_m \sim LN(\tilde{\mu}_m, \tilde{\sigma}_m, \Delta_m), \text{ para } \Delta_m \equiv \frac{L_m - PD_m}{\hat{\sigma}_m}. \quad (10)$$

Os parâmetros  $\tilde{\mu}_m$  e  $\tilde{\sigma}_m$  da distribuição lognormal 3 parâmetros de  $a_m$  podem ser estimados pelas estatísticas amostrais,

$$\tilde{\mu}_m \equiv \sum_{ano=1}^N \frac{a_m^{ano}}{N} \text{ e } \tilde{\sigma}_m^2 \equiv \sum_{ano=1}^N \frac{(a_m^{ano} - \tilde{\mu})^2}{N}, \quad (11)$$

onde o resíduo  $a_m^{ano}$  é obtido através do histórico fazendo

$$a_m^{ano} \equiv \frac{Y_m^{ano} - PD_m^{ano}}{\hat{\sigma}_m^2}. \quad (12)$$

Portanto, os 3 parâmetros da distribuição lognormal  $a_m$  são definidos por (10) e (11), resta determinar os momentos de primeira e segunda ordem da variável normal subjacente. Os resultados a seguir são baseados em *Charbeneau (1978)*.

Desconsiderando o índice  $m$  para não sobrecarregar a notação, pela Proposição 1 segue que

$$\begin{aligned} \tilde{\sigma}^2 &= (e^{\mu+0.5\sigma^2})^2(e^{\sigma^2} - 1) \\ &= e^{2\mu}e^{\sigma^2}(e^{\sigma^2} - 1) \\ &= e^{2\mu}(\phi^2 - \phi), \end{aligned} \quad (13)$$

para

$$\phi \equiv e^{\sigma^2}. \quad (14)$$

Deste modo, segue de (13) que  $e^{2\mu} = \frac{\tilde{\sigma}^2}{\phi^2 - \phi}$ . Então,

$$\mu = \frac{1}{2} \ln \left( \frac{\tilde{\sigma}^2}{\phi^2 - \phi} \right). \quad (15)$$

Novamente pela Proposição 1, a média da distribuição de  $a_m$  é dada por

$$\begin{aligned} \tilde{\mu} &= e^{\mu+0.5\sigma^2} + \Delta \\ &= e^{\mu} \sqrt{e^{\sigma^2}} + \Delta \\ &= e^{\mu} \sqrt{\phi} + \Delta \\ &= \sqrt{\frac{\tilde{\sigma}^2}{\phi^2 - \phi}} \phi + \Delta \\ &= \sqrt{\frac{\tilde{\sigma}^2}{\phi - 1}} + \Delta. \end{aligned} \quad (16)$$

Por (13) e (16) verifica-se que os momentos de primeira e segunda ordem da variável  $a_m$  estão em função do parâmetro  $\phi$ . Resta então determinar  $\phi$ . De (16) segue que

$$\frac{(\tilde{\mu} - \Delta)^2}{\tilde{\sigma}^2} = \frac{1}{\phi - 1},$$

ou seja,

$$\phi = 1 + \frac{\tilde{\sigma}^2}{(\tilde{\mu} - \Delta)^2} \quad (17)$$

Pela definição (14) e por (17), a variância da variável normal fica determinada por:

$$\sigma^2 = \ln \left( 1 + \frac{\tilde{\sigma}^2}{(\tilde{\mu} - \Delta)^2} \right). \quad (18)$$

Desta maneira, por (11), (15) e (18) pode-se estimar os momentos de primeira e segunda ordem da distribuição normal associada à distribuição  $a_m \sim LN(\tilde{\mu}_m, \tilde{\sigma}_m^2, \Delta_m)$ .

Na Seção 2 foi apresentada uma função para a correlação cruzada  $\tilde{\rho}$  entre duas variáveis lognormais 3 parâmetros, que depende da correlação cruzada  $\rho$  e das variâncias das variáveis normais subjacentes. A seguir a correlação cruzada  $\tilde{\rho}$  é, alternativamente, estabelecida em função dos 3 parâmetros da distribuição lognormal  $a_m$ .

### 3.1 – Correlação cruzada em função dos parâmetros da distribuição lognormal

Da equação (5) segue que a correlação entre duas variáveis lognormais é dada por

$$\tilde{\rho} = \frac{e^{\rho\sigma_1\sigma_2} - 1}{\sqrt{e^{\sigma_1^2} - 1}\sqrt{e^{\sigma_2^2} - 1}} = \frac{e^{\rho\sqrt{\ln(\phi_1)}\sqrt{\ln(\phi_2)}} - 1}{\sqrt{\phi_1 - 1}\sqrt{\phi_2 - 1}} = \frac{\left(e^{\sqrt{\ln(\phi_1)}\sqrt{\ln(\phi_2)}}\right)^\rho - 1}{\sqrt{\phi_1 - 1}\sqrt{\phi_2 - 1}},$$

onde  $\phi_i = e^{\sigma_i^2}$ , para  $i = 1, 2$ . Alternativamente, utilizando a representação (17) pode se escrever

$$\tilde{\rho} = \frac{\left( e^{\sqrt{\ln\left(1 + \frac{\tilde{\sigma}_1^2}{(\tilde{\mu}_1 - \Delta_1)^2}\right)} \sqrt{\ln\left(1 + \frac{\tilde{\sigma}_2^2}{(\tilde{\mu}_2 - \Delta_2)^2}\right)}} - 1 \right)^\rho}{\sqrt{\frac{\tilde{\sigma}_1^2}{(\tilde{\mu}_1 - \Delta_1)^2}} \sqrt{\frac{\tilde{\sigma}_2^2}{(\tilde{\mu}_2 - \Delta_2)^2}}}. \quad (19)$$

A representação (19) é função dos 3 parâmetros das distribuições lognormais de  $a_m^1$  e de  $a_m^2$ , e da correlação cruzada  $\rho$  das variáveis normais subjacentes. Esta representação pode ser utilizada para gerar aflúências capazes de reproduzir a correlação cruzada histórica. Por exemplo, se é desejado gerar uma amostra bivariada  $\{a_m^{1,i}\}_{i=1}^N, \{a_m^{2,i}\}_{i=1}^N$  de variáveis lognormais  $a_m^1$  e  $a_m^2$ , com correlação cruzada  $\tilde{\rho}$ , basta definir a função inversa de (19) com relação a  $\rho$ :

$$\rho \equiv \frac{\ln\left(\tilde{\rho} \sqrt{\frac{\tilde{\sigma}_1^2}{(\tilde{\mu}_1 - \Delta_1)^2}} \sqrt{\frac{\tilde{\sigma}_2^2}{(\tilde{\mu}_2 - \Delta_2)^2}} + 1\right)}{\sqrt{\ln\left(1 + \frac{\tilde{\sigma}_1^2}{(\tilde{\mu}_1 - \Delta_1)^2}\right)} \sqrt{\ln\left(1 + \frac{\tilde{\sigma}_2^2}{(\tilde{\mu}_2 - \Delta_2)^2}\right)}} \text{ se } \tilde{\rho} > \frac{-1}{\sqrt{\frac{\tilde{\sigma}_1^2}{(\tilde{\mu}_1 - \Delta_1)^2}} \sqrt{\frac{\tilde{\sigma}_2^2}{(\tilde{\mu}_2 - \Delta_2)^2}}} \quad (20)$$

e substituir  $\rho$  (se  $-1 \leq \rho \leq 1$ ) na matriz de correlações cruzadas, antes de aplicar a decomposição de Cholesky para o cálculo da matriz de carga.

A seguir são apresentados alguns resultados da correlação cruzada utilizando duas usinas hidroelétricas do sistema interligado nacional.

### 3.2 – Análise de sensibilidade – Resultados numéricos

A Figura 1 contém as correlações entre variáveis lognormais, obtidas pela fórmula (19), para as usinas de Itutinga e Salto Caxias, quando a correlação  $\rho$  (das variáveis normais) varia no intervalo  $[-1,1]$ . Os dois primeiros parâmetros das distribuições lognormais são estimados pelo modelo GEVAZP e dados por:  $\tilde{\sigma}_{Itutinga} = 0,7480$ ,  $\tilde{\mu}_{Itutinga} \approx 0$ ,  $\tilde{\sigma}_{S.Caxias} = 0,8710$ , e  $\tilde{\mu}_{S.Caxias} \approx 0$ .

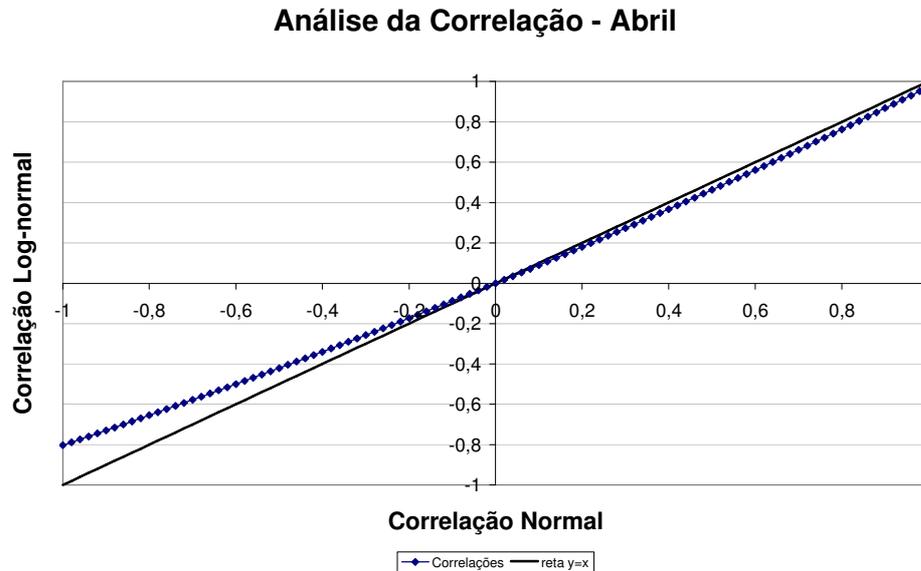
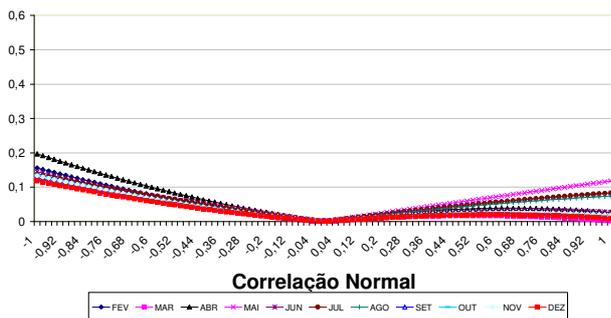


Figura 1: Correlações lognormais do mês de Abril. Tendência hidrológica igual a 100% da média histórica. Itutinga  $\times$  Salto Caxias - PMO 08/2007.

Dada esta configuração, é importante notar que correlações lognormais  $\tilde{\rho}$  inferiores a -0,81 são impossíveis de se obter com  $\rho \in [-1,1]$ .

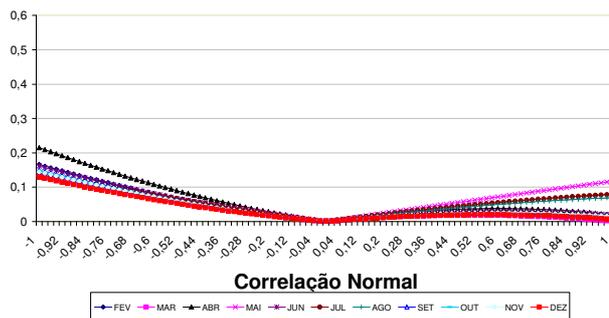
Como pode ser verificado em (19) a correlação  $\tilde{\rho}$  depende dos 3 parâmetros da distribuição lognormal. Dadas as equações (7) e (10), o deslocamento  $\Delta$  depende diretamente do passado recente utilizado pelo modelo autorregressivo periódico. A seguir é analisada a importância que a tendência hidrológica (passado recente) desempenha na reprodução das correlações cruzadas lognormais. São apresentados na Figura 2 os erros absolutos entre as correlações normais e lognormais (3 parâmetros), utilizando a fórmula (19), quando a tendência hidrológica (TH) considerada é igual a: 100%, 80%, 50% e 30% da média histórica (Média de Longo Termo - MLT). São analisados os erros  $|\tilde{\rho} - \rho|$ , quando a correlação normal  $\rho$  é alterada no intervalo  $[-1,1]$ .

**Erros Absolutos da Correlação**



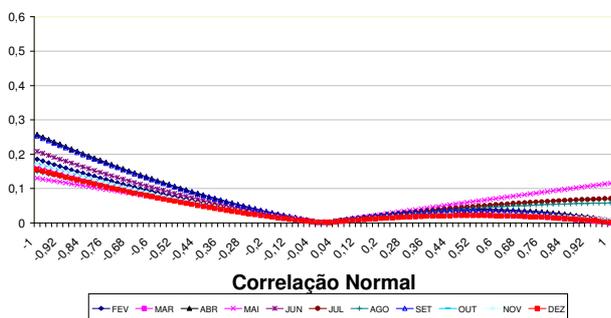
(a) TH igual a 100% da MLT.

**Erros Absolutos da Correlação**



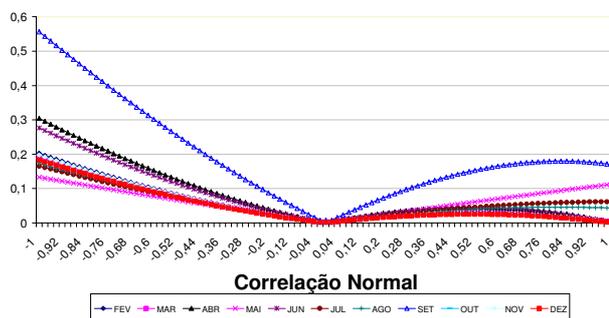
(b) TH igual a 80% da MLT.

**Erros Absolutos da Correlação**



(c) TH igual a 50% da MLT.

**Erros Absolutos da Correlação**



(d) TH igual a 30% da MLT.

Figura 2: Itutinga × Salto Caxias: FEV – DEZ (variando a correlação normal).

Resultados numéricos mostram que o efeito da variação da tendência hidrológica na representação das correlações cruzadas das distribuições lognormais é bastante significativo. Os erros maiores são verificados para as correlações negativas próximas de -1. Já os erros na reprodução das correlações positivas pela transformação lognormal são menores.

Chama-se a atenção na Figura 2(d) aos altos índices de erros na reprodução da correlação cruzada para o mês de setembro. São apresentadas nas Figuras 3 e 4 a correlação lognormal dada em função da correlação normal (equação (19)) para o mês de setembro quando a tendência hidrológica é igual a 100% e 30% da média histórica, respectivamente.

### Análise da Correlação - Setembro

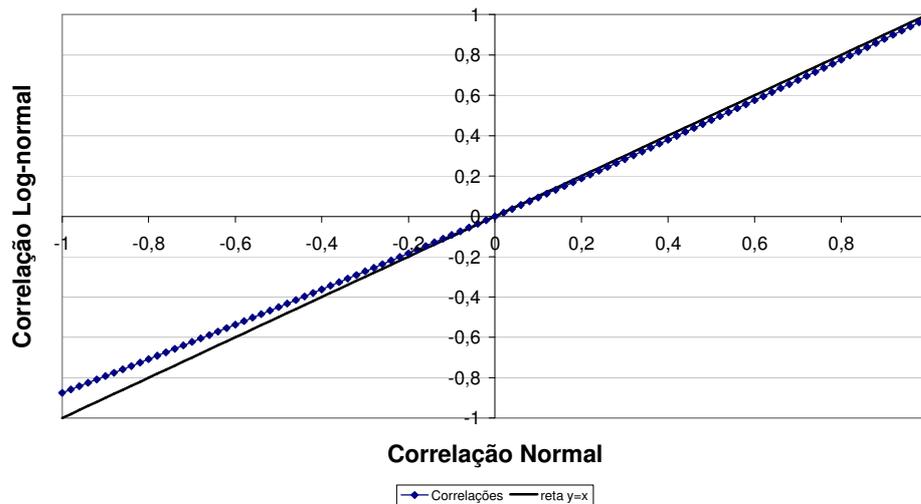


Figura 3: Setembro: tendência hidrológica igual a 100% da MLT.

### Análise da Correlação - Setembro

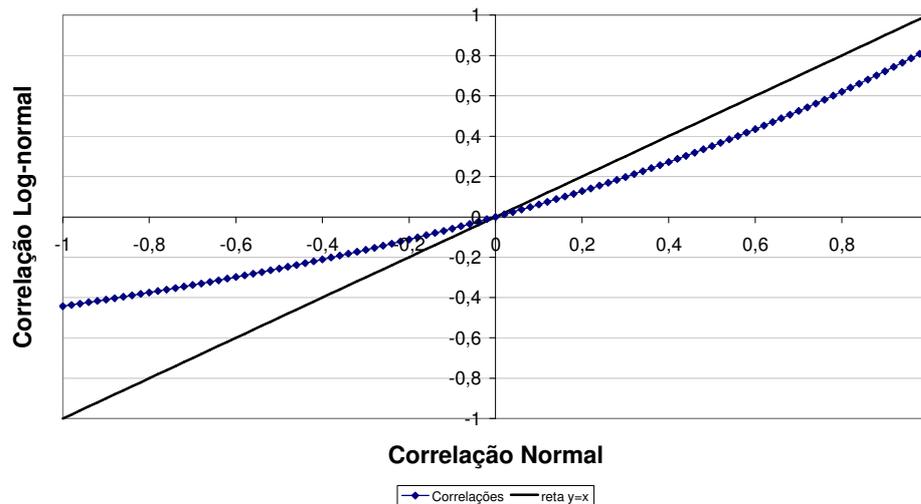


Figura 4: Setembro: tendência hidrológica igual a 30% da MLT.

Dada a configuração utilizada, pode-se verificar na Figura 4 que é impossível obter (com  $\rho$  variando em  $[-1,1]$ ) uma correlação menor do que  $-0,43$  para as variáveis lognormais 3 parâmetros.

A seguir é apresentada uma nova proposta para a matriz de carga capaz de melhorar, quando possível, a reprodução da correlação cruzada pelos ruídos lognormais.

#### 4 – ALTERAÇÃO DA NA MATRIZ DE CORRELAÇÕES CRUZADAS HISTÓRICAS

Dados os 3 parâmetros  $\tilde{\mu}, \tilde{\sigma}^2$  e  $\Delta$  da variável lognormal  $a_m$ , seja a constante  $\kappa$  definida por

$$\kappa \equiv \phi - 1 = \frac{\tilde{\sigma}^2}{(\tilde{\mu} - \Delta)^2}.$$

Pela formulação (19) pode-se escrever a correlação cruzada  $\tilde{\rho}$  como  $\tilde{\rho} = h(\rho)$ , onde a função  $h: \mathbb{R} \rightarrow \mathbb{R}$  é definida por

$$h(\rho) = \frac{\left( e^{\sqrt{\ln(1+\kappa_1)} \sqrt{\ln(1+\kappa_2)}} \right)^\rho - 1}{\sqrt{\kappa_1 \kappa_2}}. \quad (21)$$

Deste modo, definindo a inversa de  $h(\cdot)$  como

$$h^{-1}(\tilde{\rho}) \equiv \frac{\ln(\tilde{\rho} \sqrt{\kappa_1 \kappa_2} + 1)}{\sqrt{\ln(1+\kappa_1)} \sqrt{\ln(1+\kappa_2)}}, \text{ se } \tilde{\rho} > \frac{-1}{\sqrt{\kappa_1 \kappa_2}} \quad (22)$$

torna-se possível a alteração da correlação das variáveis normais para que, quando aplicada a transformação exponencial, as variáveis lognormais possuam a correlação desejada  $\tilde{\rho}$  (se  $\tilde{\rho} > -1/\sqrt{\kappa_1 \kappa_2}$ ). A alteração das correlações segue a relação seguinte

$$\rho = \varphi(\tilde{\rho}) \equiv \begin{cases} 1 & \text{se } h^{-1}(\tilde{\rho}) > 1, \\ h^{-1}(\tilde{\rho}) & \text{se } -1 \leq h^{-1}(\tilde{\rho}) \leq 1, \\ -1 & \text{se } h^{-1}(\tilde{\rho}) < -1, \text{ ou } \tilde{\rho} \leq \frac{-1}{\sqrt{\kappa_1 \kappa_2}}. \end{cases} \quad (23)$$

A seguir é explicado resumidamente o processo de correlação dos ruídos normais utilizados na geração de séries sintéticas.

#### 4.1 – Processo de correlação dos ruídos normais

Seja  $H_m \in \mathbb{R}(T, n)$  a matriz com valores históricos do mês  $m$ , onde  $T$  é o número de observações do histórico, e  $n$  é o número de objetos, por exemplo, usinas hidroelétricas. Sem perda de generalidade, pode-se supor que cada coluna de  $H_m$  tem média zero e variância igual a um. Assim sendo,

$$\Xi_H \equiv \frac{1}{T} H_m' H_m \in \mathbb{R}(n, n)$$

é a matriz de correlações cruzadas entre os  $n$  objetos.

Seja  $\Lambda \in \mathbb{R}(N, n)$  uma matriz contendo  $n$  amostras com  $N$  ruídos independentes e normalmente distribuídos. Quando  $N$  é suficientemente grande,

$$\frac{1}{N} \Lambda' \Lambda = I \in \mathbb{R}(n, n)$$

é a matriz identidade. A proposição seguinte é importante para o entendimento do processo de geração multivariada de séries sintéticas.

Proposição 3. Seja  $M_H \in \mathbb{R}(n, n)$  uma decomposição da matriz  $\Xi_H$ , tal que  $M_H' M_H = \Xi_H$ .

Então, a matriz de ruídos  $\bar{\Lambda} \in \mathbb{R}(N, n)$  obtida a partir da transformação linear  $\bar{\Lambda} \equiv \Lambda M_H$

possui matriz de correlação  $\Xi_H$ .

Prova: Como cada coluna da matriz  $\Lambda$  tem média zero e variância igual a um, e  $\bar{\Lambda}$  é obtida a partir da transformação linear  $\Lambda M_H$ , então cada coluna de  $\bar{\Lambda}$  tem média zero e variância um. Logo,

$$\text{Corr}(\bar{\Lambda}) = \frac{1}{N} \bar{\Lambda}' \bar{\Lambda} = \frac{1}{N} (\Lambda M_H)' (\Lambda M_H) = M_H' \frac{\Lambda' \Lambda}{N} M_H = M_H' I M_H = \Xi_H,$$

como se queria demonstrar.

Como apresentado na Seção 3, o resultado da Proposição 3 não necessariamente vale para a matriz de ruídos lognormais obtidos a partir de  $\Lambda$ . No entanto, para

$$\Xi(i, j) \equiv \varphi(\Xi_H(i, j)), \quad i = 1, \dots, T \text{ e } j = 1, \dots, n,$$

se existir a decomposição  $M'M = \Xi$ , pela definição da função  $\varphi(\cdot)$ , os ruídos lognormais obtidos a partir dos ruídos normais  $\bar{\Lambda} \equiv \Lambda M$  proporcionam uma matriz de correlação cruzada igual (ou bastante próxima) a  $\Xi_H$ .

A única adversidade desta modelagem é a possibilidade de não existir a decomposição  $M'M = \Xi$ . Isso acontece se a matriz  $\Xi$  não é semidefinida positiva, isto é, se  $\Xi$  possui ao menos um autovalor negativo. Para contornar esta situação, é proposto a seguir um algoritmo que constrói uma matriz  $M_\varepsilon$  que permite a reprodução (quando possível) das correlações lognormais.

#### 4.2 – Uma nova proposta para matriz de carga

Objetiva-se nesta seção obter uma matriz  $M_\varepsilon$  satisfazendo a seguinte relação

$$M_\varepsilon' M_\varepsilon = \begin{cases} \Xi & \text{se } \Xi \geq 0, \\ \Xi_\varepsilon & \text{caso contrário,} \end{cases}$$

onde  $\varepsilon \geq 0$  é uma tolerância que mede (através de uma norma apropriada) a distância entre as matrizes  $\Xi$  e  $\Xi_\varepsilon$ . O algoritmo a seguir constrói a matriz de carga proposta neste trabalho.

Algoritmo.

**Passo 1.** Calcule a matriz de correlações históricas  $\Xi_H$ .

**Passo 2.** Calcule a matriz de correlações alteradas  $\Xi$  fazendo

$$\Xi(i, j) \equiv \varphi(\Xi_H(i, j)), \quad i, j = 1, \dots, n$$

para  $i \neq j$ , e defina  $t = 1$  e  $\Xi_\varepsilon \equiv \Xi$ .

**Passo 3.** Compute o menor autovalor  $\lambda_{\min}$  de  $\Xi_\varepsilon$ .

**Passo 3'.** Se  $-\varepsilon \leq \lambda_{\min}$  realize a decomposição espectral  $U'DU = \Xi_\varepsilon$ , onde a matriz diagonal  $D$  contém todos os autovalores, e  $U$  contém todos os autovetores de  $\Xi_\varepsilon$ .

Faça, para  $i = 1, \dots, n$ ,  $D(i, i) = 0$ , se  $-\varepsilon \leq D(i, i) < 0$ , e  $M_\varepsilon \equiv U'\sqrt{D}$ .

Vá ao Passo 4.

**Passo 3''.** Caso contrário, faça  $t = \frac{t}{2}$ , e atualize a matriz  $\Xi_\varepsilon$  da seguinte maneira:

$$\Xi_\varepsilon = (1-t)\Xi_H + t\Xi.$$

Volte ao Passo 3.

**Passo 4.** Finalize o algoritmo.

O algoritmo termina em um número finito de passos se  $\varepsilon > 0$ , pois

$$t \rightarrow 0 \Rightarrow \Xi_\varepsilon \rightarrow \Xi_H,$$

e  $\Xi_H$  possui todos os autovalores não-negativos (esta afirmação pode ser confirmada levando em conta que a matriz de correlação é uma matriz de *Gram*, logo é não-negativa, Lima (1999)).

Uma diminuição moderada do fator de combinação convexa  $t$  no Passo 3'' do algoritmo pode ser, por exemplo,  $t = \frac{t}{4}$ .

A decomposição espectral pode ser obtida utilizando o Método de Jacobi, Press et al. (1992) (capítulo 11), para calcular os autovetores  $U$  e autovetores  $D$  da matriz simétrica  $\Xi_H$ . Se  $\Xi_H$  possui todos os autovalores positivos, o algoritmo pode obter uma matriz  $\Xi_\varepsilon$  com todos os autovalores positivos. Deste modo, a decomposição de Cholesky pode ser empregada em vez da decomposição espectral.

Por (21) segue que  $h(\cdot)$  é uma função crescente. Sejam  $\rho = \Xi_H(i, j)$  para quaisquer dois índices distintos  $i$  e  $j$  fixos,  $\rho_a \equiv h^{-1}(\rho)$ , e  $\rho_t = (1-t)\rho + t\rho_a$ . Ou bem

$$\rho_a \leq \rho_t \leq \rho \Rightarrow \rho = h(\rho_a) \leq h(\rho_t) \leq h(\rho),$$

ou

$$\rho_a \geq \rho_t \geq \rho \Rightarrow \rho = h(\rho_a) \geq h(\rho_t) \geq h(\rho).$$

De qualquer forma,

$$|\rho - h(\rho_t)| \leq |\rho - h(\rho)|. \quad (24)$$

Isto significa que a matriz de carga  $M_\varepsilon$  fornecida pelo algoritmo proposto reproduz de forma mais precisa a correlação histórica. Mais especificamente,

1. se  $t = 0$  a matriz de carga  $M_\varepsilon$  é equivalente à matriz de carga  $M$ , com a diferença que  $M_\varepsilon$  é obtida pela decomposição espectral;
2. se  $0 < t \leq 1$ , os ruídos lognormais obtidos utilizando a matriz de carga  $M_\varepsilon$  reproduzem de maneira mais precisa as correlações históricas.

## 5 – RESULTADOS NUMÉRICOS

Nesta seção a eficiência do algoritmo proposto é comparada com a eficiência do *método original* utilizando a decomposição de Cholesky da matriz de correlações cruzadas históricas. É considerada uma configuração do planejamento energético de médio prazo, com 4 subsistemas: Sudeste (SE), Sul (S), Nordeste (NE) e Norte (N). Foram geradas 2000 séries sintéticas de energias afluentes a cada subsistema, para um horizonte de 24 meses.

São apresentados na Figura 5 os erros absolutos da correlação cruzada obtidas a partir de 2000 séries sintéticas geradas utilizando o método original. A tendência hidrológica para a geração foi considerada igual a 100% da MLT. A Figura 6 contém a mesma análise para o *método proposto*.

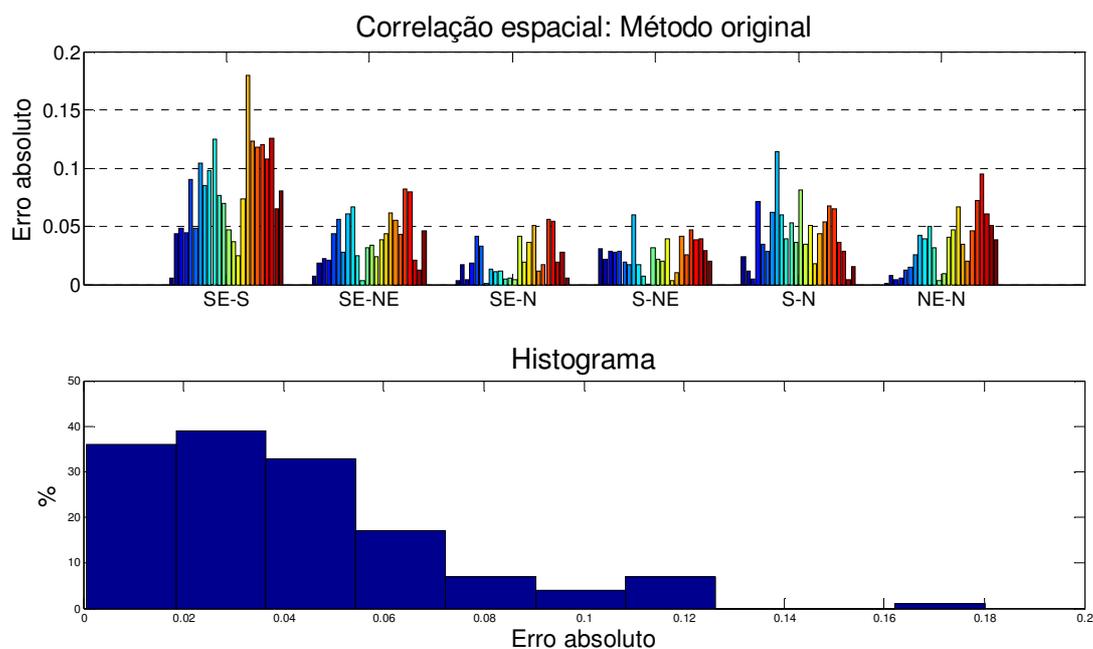


Figura 5: Tendência hidrológica utilizada igual a 100% da MLT.

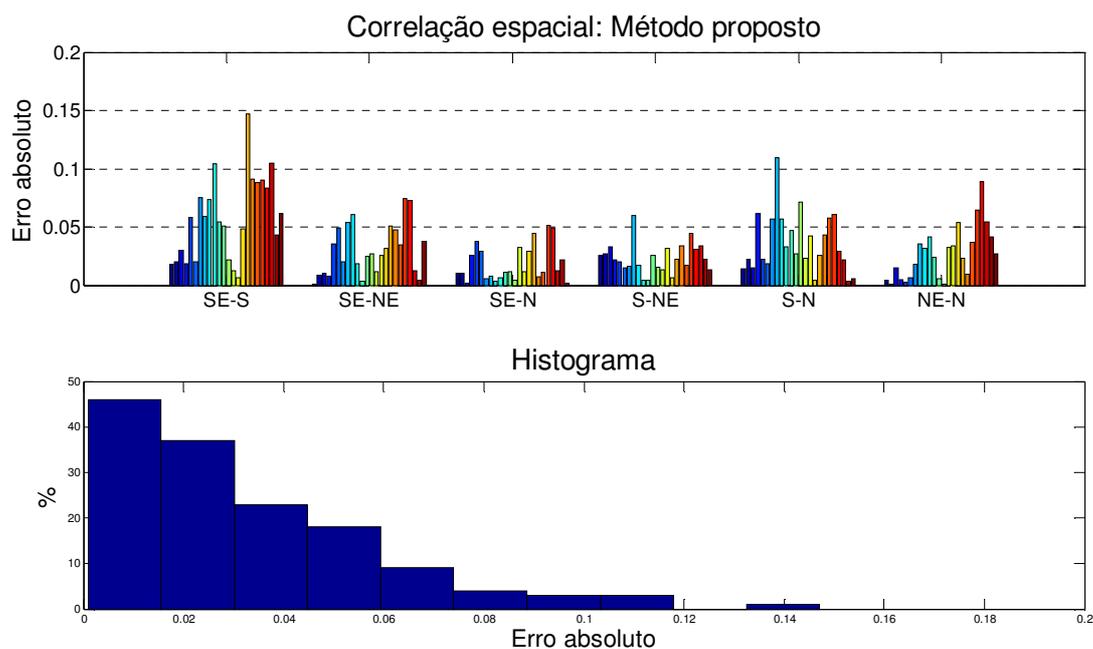


Figura 6: Tendência hidrológica utilizada igual a 100% da MLT.

É importante notar que os erros fornecidos pelo método proposto são inferiores aqueles apresentados pelo método original. Por exemplo, para pouco mais de 35% da população os erros cometidos pelo método original são inferiores a 0,02; enquanto que para o método proposto, o percentual da população é superior a 45%. Esta análise fica mais clara quando são analisadas as distribuições acumuladas dos erros, apresentadas na Figura 7.

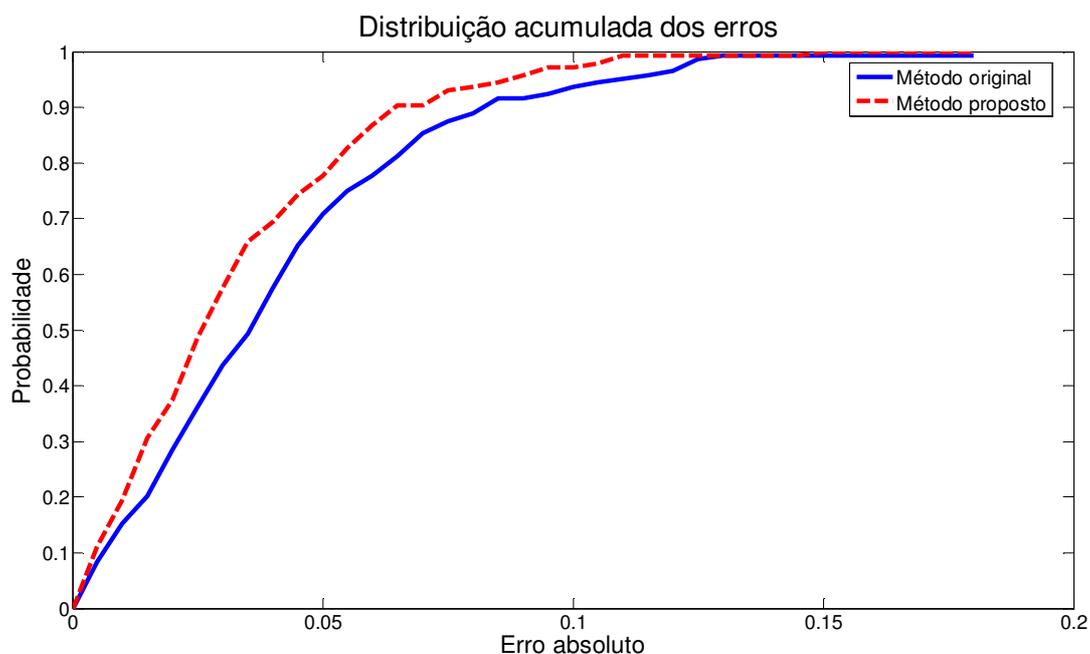


Figura 7: Tendência hidrológica igual a 100% MLT.

Como a linha tracejada na Figura 7 esta sempre acima da linha contínua, concluí-se que o método proposto reproduz melhor a correlação cruzada histórica do que o método original, uma vez que os erros encontrados são menores.

São apresentadas na Figura 8 as distribuições dos erros absolutos de correlação cruzada quando a tendência hidrológica utilizada para a geração das 2000 séries é igual a 10% da média histórica.

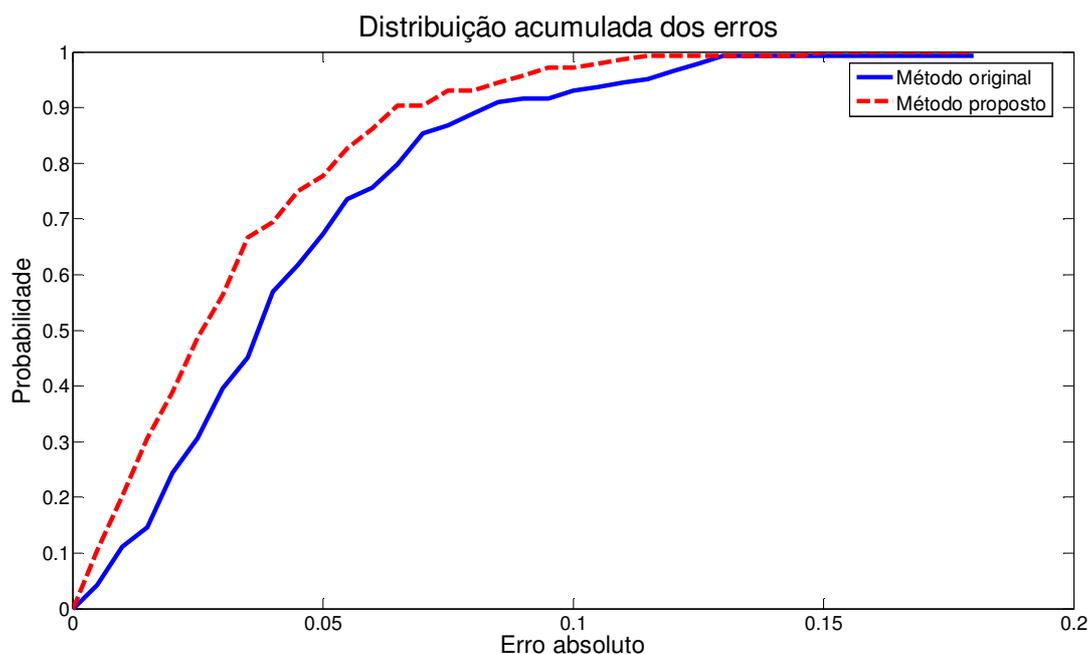


Figura 8: Tendência hidrológica igual a 10% da MLT.

Para os quatro subsistemas analisados, a menor correlação cruzada histórica é igual a -0,122, entre os subsistemas Sul e Norte. Como mencionado na Seção 3.2, correlações inferiores a esta são mais difíceis de serem reproduzidas utilizando a matriz de carga baseada na decomposição de Cholesky da matriz de correlações históricas. Neste sentido, espera-se que a matriz de carga proposta forneça resultados comparativos melhores do que aqueles apresentados nas Figuras 7 e 8.

## 6 – CONCLUSÕES

Foi apresentada neste trabalho uma análise estatística do tratamento da distribuição lognormal para a geração de séries sintéticas utilizando modelos autorregressivos periódicos. A metodologia utilizada no tratamento dos ruídos normais, lognormais e no tratamento da correlação cruzada foi detalhada. Contudo, este trabalho não se restringiu somente ao aspecto teórico. Foram apresentados na Subseção 3.2 alguns resultados de sensibilidade no processo de correlação cruzada, demonstrando que a tendência hidrológica desempenha um papel significativo na reprodução das correlações cruzadas pelas séries sintéticas geradas, utilizando a formulação lognormal três parâmetros.

Foi apresentado na Seção 4 um algoritmo que constrói uma transformação linear (matriz de carga) capaz de melhorar a reprodução das correlações cruzadas. Neste aspecto, os resultados teóricos e numéricos comprovam a melhoria desta nova transformação linear em relação à matriz de carga obtida pela decomposição de Cholesky da matriz de correlações cruzadas históricas.

## BIBLIOGRAFIA

DAMÁZIO, J.M.; MACEIRA, M.E.P; COSTA, F.S.; PENNA, D.D.J.; DUARTE, V.S. (2009) “*Representação da Correlação Cruzada na Geração de Cenários de Afluências para o Planejamento da Operação Energética de Médio Prazo*”. XI Simpósio de Especialistas em Planejamento da Operação e Expansão Elétrica - SEPOPE, Belém, Brasil.

DEGROTT, M. H.; SCHERUISH, M. J. (2002) “*Probability and Statistics*” - 3rd ed.. Assison Wesley, 816 pp.

DINIZ, A. L., SANTOS, T. N., MACEIRA, M. E. P. (2006). “*Short term security constrained hydrothermal scheduling for large scale systems considering transmission losses*”, IEEE/PES Transm. Distr. Conf. Expos. Latin America, Caracas, Venezuela, pp 2-6.

HIPEL, K.W.; McLEOD, A.I. (1998). “*Time Series Modelling of Water Resource and Environmental Systems*”. Amsterdam, The Netherlands, Elsevier Science.

- KELMAN, J.; OLIVEIRA, G.C.; MACEIRA, M.E.P. (1992). “*Modelo de Séries Hidrológicas (MSH) – Manual de Metodologia*”, Relatório Técnico CEPEL nº 242/92.
- LIMA, E.L. (1999). “*Algebra Linear*”. Rio de Janeiro: Instituto Nacional de Matemática Pura e Aplicada. Terceira edição, 357 pp.
- MACEIRA, M.E.P.; BEZERRA, C.V. (1997). “*Stochastic Streamflow model for Hydroelectric Systems*” In: Proceedings of 5th International Conference on Probabilistic Methods Applied to Power Systems, Vancouver, Canada, Sep., pp. 305-310.
- MACEIRA, M.E.P.; JARDIM, D.L.D.D. (2000). “*Modelo de Geração de Séries Sintéticas de Energia e Vazões – GEVAZP – Manual de Referência*” In: Relatório Técnico DPP/PEL 83/2000, CEPEL, Rio de Janeiro.
- MACEIRA, M.E.P.; MERCIO, C.B.; GORENSTIN, B.G.; CUNHA, S.H.F.; SUSANO, C.; SACRAMENTO, M.C.; KLIGERMAN, A. (1998). “*Application of the NEWAVE model in the energy evaluation of the Brazilian North/Northeast and South/Southeast interconnected systems*”, VI Symposium of Specialists in Electric Operational and Expansion Planning, Salvador, Brazil.
- MACEIRA, M. E. P.; PENNA, D.D.J.; DAMÁZIO, J.M. (2005). “*Geração de Cenários Sintéticos de Energia e Vazão para o Planejamento da Operação Energética*” In: XVI Simpósio Brasileiro de Recursos Hídricos, João Pessoa, Paraíba, Brasil, pp. 3-12.
- MACEIRA, M.E.P.; TERRY, L.A.; DAMAZIO, J.M.; COSTA, F.S., MELO, A.C.G. (2002). “*Chain of Models for Setting the Energy Dispatch and Spot Price in the Brazilian System, Power*” System Computation Conference – PSCC’02, Sevilla, Spain, June 24-28, pp. 1-4.
- MATALAS, N.C. (1967). “*Mathematical Assessment of Synthetic Hydrology*”. Water Resources Research, Vol.3, No 4.
- PAPOULIS, A. (1965) “*Probability, Random Variables, and Stochastic Processes*” - 3rd ed. McGraw-Hill series in electrical engineering.
- PRESS, W.H.; FLANNERY, B.P.; TEUKOLSKY, S.A.; VETTERLING, W.T. (1992) “*Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*”. Second Edition, Hardcover.
- SALAS, J.D; DELLEUR, J. W.; YEVIEVICH, V.; LANE W. L. (1980). “*Applied Modeling of Hydrologic Time Series*”, Water Resources Publications, p. 484.
- YEVJEVICH, V. (1972) “*Probability and Statistics in Hydrology*”. Water Resources Publications.