SNPTEE
SEMINÁRIO NACIONAL
DE PRODUÇÃO E
TRANSMISSÃO DE
ENERGIA ELÉTRICA

GPL - 24
16 a 21 Outubro de 2005
Curitiba - Paraná

GRUPO VII
GRUPO DE ESTUDO DE PLANEJAMENTO DE SISTEMAS ELÉTRICOS - GPL

APLICAÇÃO DE TÉCNICAS DE AGRUPAMENTO SOBRE SÉRIES DE POTÊNCIA PARA AVALIAÇÃO DE CONFIABILIDADE EM SISTEMAS HIDROTÉRMICOS

Marcia Endler Kranen* — **IM/NCE-UFRJ**

Adriano Joaquim de Oliveira Cruz — **IM/NCE-UFRJ**

Albert Cordeiro Geber de Melo — **Cepel**

RESUMO

O objetivo deste trabalho é propor uma metodologia para reduzir o esforço computacional na avaliação de confiabilidade, considerando a variação das potências disponíveis nas usinas hidrelétricas em função da aleatoriedade das vazões afluentes aos reservatórios. A metodologia proposta consiste na utilização de técnicas de agrupamento, baseadas em lógica nebulosa, aplicadas às Séries de Potência Disponível (SPD) obtidas a partir de estudos hidrotérmicos de operação do sistema elétrico brasileiro (SEB). A meta é obter o menor número de séries que reflita o comportamento randômico da hidrologia e que preserve a dependência espacial e temporal dos dados originais. A análise de confiabilidade baseada neste conjunto reduzido de cenários hidrológicos deve manter a precisão na estimativa dos índices de confiabilidade obtidos quando os estudos são realizados a partir dos dados originais.

PALAVRAS-CHAVE

Análise de Agrupamentos, Análise de Confiabilidade, Lógica Nebulosa, Medidas de Validação.

1.0 - INTRODUÇÃO

O objetivo de um sistema elétrico de potência é produzir, transmitir e distribuir energia elétrica com segurança, com custo mínimo e com um grau adequado de continuidade e qualidade do serviço. A fim de garantir que estas condições serão atendidas, torna-se necessário que o planejamento e a operação do sistema mantenham reservas adequadas de geração que minimizem o risco de falhas no suprimento de energia. (1) Uma das técnicas utilizadas para mensurar o risco de ocorrência de falhas em um sistema elétrico é a analise de confiabilidade. Esta técnica possui como principal resultado um conjunto de índices de confiabilidade, calculados por meio de métodos estatísticos, que serão utilizados por especialistas como base para a verificação do estado do sistema. (2)
Ao avaliar a confiabilidade do sistema elétrico é importante considerar o comportamento estocástico das variáveis envolvidas no problema, tais como saídas forçadas dos equipamentos de geração e de transmissão, assim como a variação da carga ao longo do tempo. Os índices de confiabilidade são obtidos analisando-se o desempenho de diversos estados do sistema obtidos pela combinação dos níveis da carga e da disponibilidade dos equipamentos, condicionadas a um cenário hidrológico específico. (1)
Nos sistemas de potência em que a capacidade de produção é predominantemente hidrelétrica, a produção de energia depende do volume de água armazenado nos reservatórios das usinas. Uma vez que este volume armazenado depende das condições hidrológicas observadas durante um período de tempo, pode-se afirmar que a potência disponível das usinas hidrelétricas é também função da hidrologia, que por sua vez também é uma variável com comportamento aleatório e cuja previsão é complexa. Logo, neste tipo de sistema a hidrologia também desempenha um papel importante nos níveis da confiabilidade por causa da variação que esta impõem a potência disponível. (2, 3)

*Rua Guilherme Baptista, 490 / 201 - CEP 22790-160 – Rio de Janeiro - RJ - BRASIL
Tel.: (21) 2487-4709 - Fax: (21) 3393-8504 - e-mail: marcia_kranen@yahoo.com.br

Estudos de avaliação de confiabilidade composta são realizados pela análise da performance de vários estados do sistema, resultantes da combinação dos níveis de carga e disponibilidade dos equipamentos. Normalmente, o impacto das condições hidrológicas sobre a potência disponível dos geradores é considerado quando a análise de cada estado selecionado é feita utilizando um fluxo de potência ótimo (fpo) linear.
Por outro lado, este impacto não costuma ser avaliado utilizando fpo não linear. O esforço computacional associado a estes estudos pode ser muito alto, e algumas vezes proibitivo, devido ao grande número de condições hidrológicas que devem ser consideradas, associada à complexidade dos algoritmos utilizados para verificar a adequação de cada uma dos estados selecionados.
O objetivo deste trabalho é propor um processo que reduza o esforço computacional proveniente da avaliação dos diversos cenários hidrológicos nos estudos de confiabilidade. Este processo consiste na aplicação das técnicas agrupamento, baseadas em lógica nebulosa e em técnicas estatísticas, sobre as SPD das usinas hidrelétricas, derivadas dos cenários hidrológicos determinados por estudos de planejamento energético.
O processo proposto reflete o comportamento aleatório das séries hidrológicas em um número menor de SPD mantendo a dependência temporal e espacial existente nos dados originais. Assim, a análise de confiabilidade baseada neste número reduzido de cenários hidrológicos resulta em um esforço computacional reduzido e mantém a exatidão na estimativa dos índices de confiabilidade.
A metodologia proposta foi aplicada a uma configuração do subsistema Sudeste/Centro-Oeste do SEB, composto por 94 usinas hidrelétricas, 2293 barras e 3554 circuitos. A avaliação de confiabilidade foi realizada de duas formas: Na primeira, a avaliação de confiabilidade foi repetida para cada uma das SPD, baseadas nos dados originais; Na segunda, foram utilizados os resultados obtidos a partir dos processos de agrupamento. Estes resultados foram discutidos em termos da acuidade dos resultados obtidos e do esforço computacional necessário para realização dos diferentes estudos.

2.0 - O PROBLEMA DO PLANEJAMENTO DA OPERAÇÃO

O SEB é um sistema de grande porte com predominância de geração hidrelétrica (mais de 90% de sua produção) e é caracterizado por possuir grandes reservatórios com capacidade de regulação plurianual. Sendo que estes reservatórios encontram-se posicionados em cascata ao longo de diversas bacias hidrográficas (4). O fato destes reservatórios possuírem volumes limitados adicionado às incertezas associadas à variação das afluências futuras gera uma inter-relação entre as decisões tomadas num determinado instante e suas conseqüências futuras. O problema do planejamento energético da operação do SEB tem as seguintes características: possui dependência temporal uma vez que uma decisão tomada no presente afeta os custos futuros de operação; é essencialmente estocástico, devido as incertezas associadas às afluências futuras; as séries históricas de vazão possuem longos períodos de seca; os reservatórios possuem capacidade de regulação plurianual; a operação de uma usina a montante afeta a operação de outras usinas a jusante; e existe um compromisso entre o custo de operação e a confiabilidade do sistema. Além de todas estas características, a programação diária do despacho deve levar em conta o compromisso entre a utilização de unidades termo ou hidrelétricas. Como conseqüência, o planejamento energético varia deste a otimização plurianual até o despacho horário.
Por causa destas características o planejamento energético da operação tem de ser realizado utilizando uma cadeia de modelos para os diversos horizontes de planejamento e graus de detalhe na representação do sistema (5). Esta cadeia de modelos foi desenvolvida pelo CEPEL (Centro de Pesquisa em Energia Elétrica) e é utilizada pelo ONS (Operador Nacional do Sistema) e CCEE (Câmara de Comercialização de Energia Elétrica). Neste trabalho foram utilizados dois destes modelos, o NEWAVE e o SUISHI-O (4-6).

2.1 Modelo Newave

Este modelo (4-6) calcula, para cada mês do período de planejamento (5 a 10 anos), a alocação ótima dos recursos térmicos e hidráulicos, que minimize o custo esperado de operação (custo de geração térmica, adicionado as penalidades por falha no suprimento). Os reservatórios que compõem o sistema são agregados em reservatórios equivalentes representando os subsistemas Sul, Sudeste, Norte e Nordeste do SEB. A metodologia de solução é obtida pela aplicação de programação dinâmica dual estocástica. Neste algoritmo a energia armazenada e a tendência hidrológica representam um estado, e um esquema de simulação de Monte Carlo é utilizado para construir interativamente a função de custo futuro do sistema.
O NEWAVE possui 4 módulos básicos: Módulo de Cálculo dos Sistemas Equivalentes (para cada subsistema, os reservatórios são agregados de forma a obter um reservatório equivalente e as vazões são agregadas em vazões equivalentes), Módulo de Cálculo do Modelo Estocástico de Energias Afluentes (estima os parâmetros do modelo estocástico auto-regressivo periódico (PAR(p)) e gera séries sintéticas de energias afluentes), Módulo de Cálculo da Política de Operação Hidrotérmica (determina a estratégia de operação mais econômica para os subsistemas equivalentes, levando em conta as incertezas nas afluências e níveis de carga), e Módulo de Simulação da Operação (simula a operação do sistema equivalente ao longo do período de planejamento para cenários distintos de afluências futuras, falhas dos componentes e variações da demanda e calcula indicadores estatísticos de desempenho do sistema, tais como, a probabilidade de déficit de energia, o valor esperado do custo total de operação, o valor esperado da energia não suprida e os custos marginais de operação).

Entre os resultados do modelo estaremos considerando a função de custo futuro das SPD. Como estes resultados encontram-se agregados por subsistema, será necessário utilizar o modelo SUISHI-O para desagregá-los pelas usinas hidrelétricas.

2.2 Modelo SUISHI-O

O SUISHI-O (4-6) é um modelo detalhado de simulação/otimização da operação de usinas para sistemas hidrotérmicos interconectados. Ele possui a capacidade de simular sistema hidrotérmicos eletricamente interconectados por uma rede elétrica, desde que sejam hidraulicamente desconectados. Ele pode ser acoplado ao modelo Newave através da função de custo futuro para cada estágio, e ele pode considerar restrições operativas relativas aos usos múltiplos da água, tais como: vazão máxima para controle de cheias, vazão mínima para saneamento ou navegação e desvio de vazão do rio para irrigação. Ele simula séries de vazão históricas ou sintéticas, de forma a produzir os índices probabilísticos de performance do sistema. Este modelo pode calcular a energia firme ou a energia assegurada em a partir de uma probabilidade de déficit pré-estabelecida. Além de permitir a representação da operação de bacias especiais.
Dentre todos os resultados obtidos pelo SUISHI-O, neste trabalho estaremos interessados nas SPD calculadas para cada usina hidrelétrica. A Figura 1 apresenta um exemplo destas séries de potência.

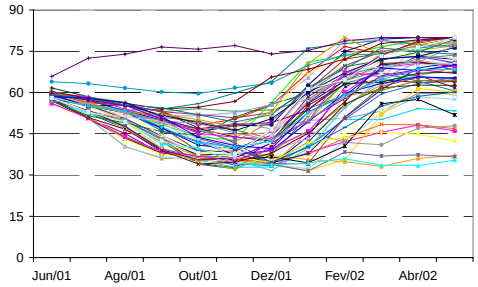


FIGURA 1: Exemplo de série de potência disponível (Usina de Caconde)

3.0 - CARACTERÍSTICAS DA SOLUÇÃO PROPOSTA

Uma forma de considerar o impacto das condições hidrológicas sobre os estudos de confiabilidade é efetuar os estudos de confiabilidade condicionados a cada um dos cenários hidrológicos possíveis e tomar os valores médios dos índices de confiabilidade calculados (3). Para tanto, é necessário executar os modelos relacionados acima pertencentes à cadeia de planejamento da operação para obter as SPD associadas a cada usina hidrelétrica.
No caso do SEB, os modelos a serem utilizados são o Newave e o Suishi-O. Por sua vez, a avaliação de confiabilidade composta será realizada pelo modelo NH2 (2, 7), também desenvolvido pelo CEPEL. A Figura 2 apresenta um esquema do fluxo de informação entre os sistemas sem a utilização de processo de agrupamento.

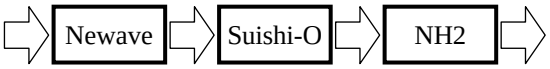


FIGURA 2 – Fluxo de informação sem a utilização do processo de agrupamento

Como mencionado anteriormente, o esforço computacional necessário para realizar a análise desta maneira é grande mesmo que esta análise seja realizada somente para um mês. Por exemplo, para o SEB, os estudos de confiabilidade, que envolvem a avaliação de condições hidrológicas são normalmente realizados levando em conta dados histórico (65 séries hidrológicas) ou dados sintéticos (2000 séries calculadas a partir do histórico).
Neste trabalho está sendo proposta uma metodologia para reduzir este esforço computacional. O proposto é utilizar um processo de agrupamento, baseado em lógica nebulosa e em métodos estatísticos, a ser aplicado sobre as SPD, obtidas a partir do Newave e do Suishi-O. É esperado que o processo de agrupamento produza um número menor de cenários hidrológicos e, conseqüentemente, um número menor de estudos de confiabilidade. De forma a manter a acuidade dos índices de confiabilidade, este processo de agrupamento deve manter as dependências temporal e espacial existentes nos dados originais. A Figura 3 apresenta o fluxo de informação entre os 3 sistemas incluindo o processo de agrupamento.

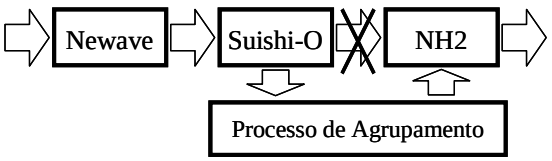


FIGURA 3 – Fluxo de informação com a utilização do processo de agrupamento

4.0 - TÉCNICAS DE AGRUPAMENTO

As técnicas de agrupamento são uma forma não supervisionada de particionamento de dados. Este tipo de técnica já foi utilizado para resolver problemas em vários contextos e por pesquisadores de diversas áreas. Estas características refletem a sua vasta aplicação e usabilidade como uma etapa da análise exploratória de dados (16).
O processo de agrupamento que está sendo proposto poderia utilizar uma grande variedade de técnicas, tais como: métodos estatísticos (k-Means, k-Modes, PAM, CLARA, etc) (8-11), métodos nebulosos (FCM, GG, GK, Fanny, etc) (12-17), e redes neurais artificiais (LQV, SON, ART, etc) (18-19). Neste trabalho, será analisada a performance do FCM e do PAM, que serão apresentados a seguir.
Em muitos algoritmos de agrupamento é necessário informar o número de agrupamentos que se deseja obter a priori. De maneira a auxiliar na determinação do melhor número de agrupamentos para um determinado conjunto de dados, são utilizadas as medidas de validação. Existe um grande número de trabalhos sobre medidas de validação entre eles os trabalhos apresentados em (20-23). O FCM e o PAM são algoritmos deste tipo. Por este motivo foram selecionadas a ICC para realizar a validação do FCM, e a Silhueta média global para validar o PAM.

4.1 FCM (12, 16, 19)

O FCM (Fuzzy c-Means) é um método de agrupamento que permite que um dado pertença a mais de um agrupamento. Este algoritmo foi criado por Dunn em 1973, mas foi Bezdek, em 1981, quem o definiu na forma como este é conhecido (12). Este algoritmo é baseado na minimização da função objetivo apresentada na equação 1.

$$J_m = \sum_{i=1}^{n}\sum_{j=1}^{c}(\mu_{ij})^m . \left\| x_i - c_j \right\|^2,\ 1 \le \mathrm{m} < \infty \qquad (1)$$

Onde *m* é um número real maior que 1, u_{ij} é o grau de pertinência de uma amostra x_i no agrupamento *j*, x_i é a i-ésima amostra pertencente ao espaço de dados que está sendo agrupado, c_j é o j-ésimo centro de agrupamento, e ||*|| é uma medida de similaridade entre as amostras e o centro do agrupamento.
O particionamento é efetuado pela otimização iterativa da equação 1, onde os graus de pertinências são atualizados como mostrado na equação 2 e os centros dos agrupamento como mostra a equação 3. O processo iterativo irá parar quando $\max_{ij}\{|u_{ij}^{(k+1)} - u_{ij}^{k}|\} < \varepsilon$, onde ε é um critério de parada definido entre 0 e 1, e *k* é o número de iterações.

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|}\right)^{\frac{2}{m-1}}} \qquad (2)$$

$$c_j = \frac{\sum_{i=1}^{n} u_{ij}^m x_i}{\sum_{i=1}^{n} u_{ij}^m} \qquad (3)$$

4.1.1 ICC (23)

O ICC (Inter Class Contrast) foi definido por Franco em 2002 para ser uma medida de validação tanto para algoritmos de agrupamento rígidos ou nebulosos. Esta medida é determinada pela equação 4.

$$ICC = \frac{S_{Be}}{n} . d_{\min} . \sqrt{c} \qquad (4)$$

Onde S_{Be} representa a compacidade dos agrupamentos e d_{min} representa a separação entre os agrupamentos. O termo $\sqrt{c}$ foi colocado na equação de forma a garantir que a medida cresça de acordo com o número de clusters que está sendo definido.
A estratégia para determinar o melhor número de agrupamento utilizando a ICC é calculando o seu resultado para um intervalo $2 \le c \le c_{\max}$ e determinando o seu valor máximo. Sendo que c_{max} é determinado de acordo com a quantidade de dados disponíveis.

4.2 PAM (9-11)

Este algoritmo de particionamento de dados rígido foi definido por Kauffman e Rousseeuw em 1990 (9). Comparado ao k-Means, o PAM tem as seguintes características: (i) utiliza uma matriz de dissimilaridades e (ii) é mais robusto, pois ele minimiza a soma das dissimilaridades entre todos os dados ao invés da soma das distâncias Euclidianas aos centros dos agrupamentos.
Em muitos problemas envolvendo agrupamentos, pode-se estar interessado em selecionar como representante do agrupamento um objeto pertencente ao conjunto de dados. O PAM trabalha desta forma, ou seja, ele seleciona *medoids*, que podem ser definidos como um objeto do agrupamento que possui a menor média dissimilaridade em relação a todos os outros objetos do agrupamento. Na literatura sobre o assunto, este tipo de objeto também é chamado de *centrótipo*.
O algoritmo do PAM é realizado em duas etapas.Na primeira etapa, são selecionados os *k* objetos mais centrais, ou seja, são selecionados os objetos cujas dissimilaridades médias para o restante do conjunto é mínima. Estes

serão os centrótipos iniciais. Os demais objetos do conjunto de dados serão alocados aos agrupamentos representados pelo centrótipo mais próximo ao objeto, segundo a equação 5. Na segunda etapa, a função objetivo, dada pela equação 6, é minimizada realizando a troca de cada um dos centrótipos por um objeto não selecionado.

$d(i, mv_i) \leq d(i, m_w)$ para todo w = 1,..., k (5)

$$J = \sum d(i, mv_i) \quad (6)$$

4.2.1 Silhueta Média Global ($\overline{S}$) (9-11)

A $\overline{S}$ é uma medida de validação de agrupamentos definida especialmente para ser utilizada em conjunto com o PAM. Ela foi criada para auxiliar a decisão sobre o número de agrupamentos apropriado e para facilitar a distinção entre um agrupamento ruim ou não. Cada agrupamento é representado por uma silhueta, mostrando quais objetos encentram-se dentro do agrupamento e quais estão localizados em posições entre agrupamentos. A forma como as silhuetas são construídas é apresenta da a seguir.

Considere um objeto x_i pertencente a um conjunto de dados, e seja A o agrupamento ao qual o objeto pertence. A silhueta do objeto i é calculada como mostra a equação 7.

$$a(\mathbf{x}_i) = \frac{1}{|A|-1}\left(\sum_{j=1, j\neq i}^{|A|-1} D(\mathbf{x}_i, \mathbf{x}_j)\right) \quad (7)$$

$$D(\mathbf{x}_i, C) = \frac{1}{|C|}\sum_{k=1}^{|C|} D(\mathbf{x}_i, \mathbf{x}_k) \quad (8)$$

onde D($\mathbf{x}_i$,$\mathbf{x}_j$) representa a dissimilaridade entre os objetos $\mathbf{x}_i$ e $\mathbf{x}_j$ e $|A|$ é o número de amostras existentes em A.

Dando continuidade, considere C como um agrupamento diferente de A e calcule a D($\mathbf{x}_{ii}$,C), que representa a dissimilaridade média entre o objeto x_i e os k objetos pertencentes a C, como mostrado na equação 8. Calcule D(x_i,C) para todos os agrupamentos diferentes de A, e selecione o menor valor obtido, como mostrado na equação 9. A silhueta de x_i pode ser definida pela equação 10.

$$b(\mathbf{x}_i) = \min_{C\neq A} D(\mathbf{x}_i, C) \quad (9)$$

$$s(\mathbf{x}_i) = \frac{b(\mathbf{x}_i) - a(\mathbf{x}_i)}{\max\{a(\mathbf{x}_i), b(\mathbf{x}_i)\}} \quad (10)$$

A silhueta pode ser interpretada como:

- s(x_i) ≈ 1 - x_i foi bem alocado ao agrupamento.
- s(x_i) ≈ 0 - x_i está em uma região intermediária entre agrupamentos.
- s(x_i) ≈ -1 - x_i não foi bem alocado ao agrupamento.

Para determinar o melhor número de agrupamentos é necessário determinar a $\overline{S}$ como mostra a equação 11.

$$\overline{S} = \frac{1}{n}\sum_{i=1}^{n} s(\mathbf{x}_i) \quad (11)$$

5.0 - APLICAÇÃO DA SOLUÇÃO PROPOSTA

A metodologia proposta será aplicada a uma configuração do SEB. O estudo do planejamento da coordenação hidrotérmica utilizou os dados de uma configuração completa do sistema, incluindo a representação dos intercâmbios entre os subsistemas. Nele foram utilizados os modelos Newave e Suishi-O, utilizado um horizonte de planejamento de 5 anos discretizado mensalmente. O resultado obtido foram 65 SPD, associadas às séries históricas de vazão, para cada uma da 94 usinas hidrelétricas representadas neste estudo de caso. Neste trabalho, será considerado somente o primeiro ano do horizonte de estudo o que representa séries de 12 meses.

A análise de confiabilidade foi realizada utilizando a avaliação de confiabilidade composta disponibilizada pelo NH2. Neste caso, foi utilizada uma configuração do subsistema Sudeste/Centro-Oeste, composta por 2293 barras, 3554 circuitos e 56 usinas hidrelétricas. Desta 56 usinas, somente 30 apresentaram alguma variação nos valores de suas SPD, sendo por isso selecionadas para participar do processo de agrupamento. Um mês do primeiro ano do período de planejamento foi selecionado para a realização da análise de confiabilidade, considerando uma condição de carga pesada. Os diferentes estados do sistema foram selecionados utilizando simulação de Monte Carlo, e a adequação de cada estado selecionado foi verificada utilizando um fluxo de potência ótimo baseado em um algoritmo de pontos interiores.

5.1 Análise sem Processo de Agrupamento (*APA*)

Neste enfoque da solução, foram executados 65 estudos de confiabilidade, tendo sido adotado como critério de convergência da simulação de Monte Carlo o coeficiente de variação 5% para LOLP, o que resultou em uma amostragem média de 3000 estados, formados pela combinação das falhas de geradores, de transformadores e de linhas de transmissão. A Figura 4 e a Tabela 1 apresentam um resumo destes resultados.

5.2 Agrupamento de Matriz Única (*AMU*)

No *AMU*, todas as séries de potência de todas as usinas hidrelétricas que participam do processo são colocadas em uma única matriz de dados, onde o critério para formação desta matriz foi: as primeiras 12 colunas representam os dados das séries de potência referentes à primeira usina, as colunas 13 a 24 representam os dados da segunda usina e assim sucessivamente até que todas as usinas foram representadas. O fato de todos os dados, de todas as usinas que participam do processo de agrupamento, serem processados em uma única

execução do algoritmo garante que as dependências temporal e espacial existentes nos dados foram mantidas. A Figura 5 apresenta uma possível representação da matriz de dados deste processo, onde a célula com a representação u30 $_{2,12}$, indica o valor do décimo segundo mês da série número 2 da trigésima usina representada no conjunto de dados.

TABELA 1: *APA* – Estatísticas da LOLP

	LOLP
Máximo	0,2626
Mínimo	0,1829
Média	0,2123

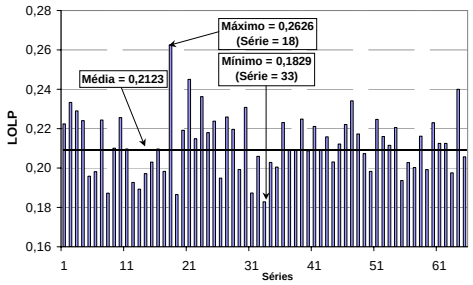


FIGURA 4: *APA* – Resultados para LOLP

Séries	Usina 1				Usina 30			
	M01	M02	...	M12	M01	M02	...	M12
1	$u1_{1,1}$	$u1_{1,2}$	...	$u1_{1,12}$	$u30_{1,1}$	$u30_{1,2}$	...	$u30_{1,12}$
2	$u1_{2,1}$	$u1_{2,2}$	...	$u1_{2,12}$	$u30_{2,1}$	$u30_{2,2}$	...	$u30_{2,12}$
65	$u1_{65,1}$	$u1_{65,2}$	...	$u1_{65,12}$	$u30_{65,1}$	$u30_{65,2}$	...	$u30_{65,12}$

FIGURA 5 – Representação da matriz de dados do *AMU*

Para determinar o melhor número de agrupamentos (*c*), é necessário adotar um valor para o coeficiente de nebulização *m*, que será utilizado no FCM. Neste trabalho, foram adotados os seguintes valores para *m*: 1,25; 1,50; 1,75 e 2,00. A Tabela 2 apresenta para cada um dos valores de m, o melhor número de agrupamentos determinado e o valor do ICC associado. Para *m* = 1,25, o número de agrupamentos obtido foi *c* = 10. Entretanto, este número diminuiu para 3 e 2 para os valores maiores de *m*.

A Tabela 2 também apresenta o tempo de processamento de cada uma das execuções do processo de agrupamento, em relação ao tempo de processamento associado com o cálculo dos índices de confiabilidade para uma condição hidrológica, ou seja, em comparação a uma execução do programa NH2. Como pode ser verificado, o tempo total para executar o AMU para todos os 4 valores de m fica em torno de 6% do tempo necessário para executar uma avaliação de confiabilidade. Ou seja, o acréscimo de tempo por causa da inclusão do processo é pequeno.

TABELA 2 – Resultados do *AMU*

m	***c***	***ICC***	**Tempo de processamento relativo (%)**
1.25	10	5.0135	1.4
1.50	3	2.8394	1.5
1.75	2	2.1076	1.5
2.00	2	1.5866	1.4

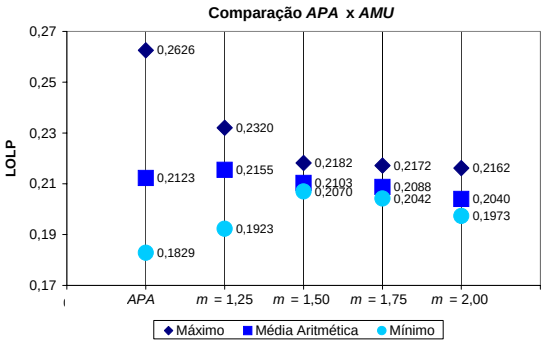


FIGURA 6 – Comparação dos resultados APA e AMU

A Figura 6 apresenta os valores do máximo, da média e do mínimo para LOLP obtidos para o *AMU* para cada um dos valores de *m*. Também são apresentados, para fins de comparação, os valores obtidos pelo *APA*. Pode ser observado que as médias das LOLPs estão sempre muito próximas, o que indica que este método consegue preservar o valor médio da LOLP. Entretanto, este processo tende a concentrar os valores da LOLP ao redor da média.

Em termos do tempo de execução dos casos de confiabilidade, uma vez que não há como escolher o melhor resultado antes de executar todos os casos determinados pelo processo, seria necessário executar 17 casos de confiabilidade em lugar de 65, o que resultaria em uma redução de 74% no número de casos, o que bastante bom.

5.3 Agrupamento por Mapeamento de Resultados com Seleção do Coeficiente de Nebulização (*AMRM*)

Neste segundo método, os espaços de dados das usinas hidrelétricas são agrupados em duas etapas interligadas por um *Mapa de Resultados*. Na primeira etapa, o FCM é utilizado de forma a manter a dependência temporal existente nas SPD das usinas. Os resultados obtidos nesta primeira etapa, são mapeados de forma a manter a dependência espacial existente entre os dados das usinas. Este mapa será tratado na segunda etapa do processo. Na segunda etapa do *AMRM*, o PAM foi utilizado em conjunto com a Silhueta Média Global.

A Figura 7 apresenta um exemplo do espaço de dados utilizado no *AMRM*, cuja interpretação é semelhante a da Figura 5. Já o mapa de resultados é apresentado na Figura 8, onde, a célula apresentada por $c5_{u30,2}$ indica que a segunda série de potência disponível da trigésima usina (u30) foi alocada ao agrupamento 5 (c5).

	Séries	**M01**	**M02**		**M12**
Usina 1	1	$u1_{1,1}$	$u1_{1,2}$		$u1_{1,12}$
	2	$u1_{2,1}$	$u1_{2,2}$		$u1_{2,12}$
	65	$u1_{65,1}$	$u1_{65,2}$		$u1_{65,12}$
Usina 30	1	$u30_{1,1}$	$u30_{1,2}$		$u30_{1,12}$
	2	$u30_{2,1}$	$u30_{2,2}$		$u30_{2,12}$
	65	$u30_{65,1}$	$u30_{65,2}$		$u30_{65,12}$

FIGURA 7 – Espaço de dados da primeira etapa

Séries	**Usina 1**	**Usina 2**		**Usina 30**
1	$c1_{u1,1}$	$c1_{u2,1}$		$c5_{u30,1}$
2	$c1_{u1,2}$	$c1_{u2,2}$		$c5_{u30,2}$
65	$c2_{u1,65}$	$c1_{u2,65}$		$c3_{u30,65}$

FIGURA 8 – Mapa de Resultados

A principal diferença entre este processo e o *AMU* está no fato que além de determinar o melhor número de agrupamentos a *ICC* também foi utilizada para encontrar o melhor valor do coeficiente de nebulização utilizado para cada um dos espaços de dados. Ou seja, o resultado final foi somente um conjunto de centrótipos.

A Tabela 3 apresenta o número de agrupamentos determinado pelo *AMRM* e o tempo de processamento relativo do processo calculado da mesma forma que na Tabela 2. Como pode ser verificado, o tempo de processamento relativo do *AMRM* (5%) ficou próximo ao tempo obtido para o *AMU*, portanto também pode ser considerado muito pequeno.

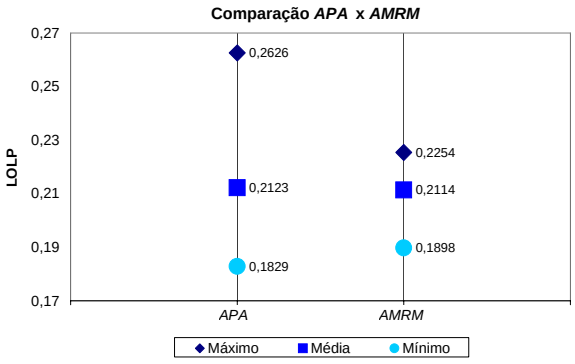
A Figura 9, assim como a Figura 6, apresenta os resultados das estatísticas da LOLP do *APA* e *AMRM* de forma a permitir a sua comparação. Novamente pode ser verificado que este processo foi eficiente na determinação do valor médio da LOLP, mas ainda concentrando os resultados do máximo e do mínimo muito próximos à média.

Em termos do tempo de execução dos casos de confiabilidade, neste caso seria necessário executar somente 4 casos de confiabilidade em lugar de 65, o que resultaria em uma redução de 94% no número de casos, representando uma grande redução inclusive em relação ao *AMU*.

TABELA 3 – Resultados do *AMRM*

c	$\overline{S}$	**Relative Time (%)**
4	0.2166	5.3

FIGURA 9 – Comparação dos resultados *APA* e *AMRM*



6.0 - CONCLUSÕES

Este trabalho apresentou uma metodologia para reduzir o esforço computacional da avaliação de confiabilidade levando em conta a variação hidrológica. A metodologia proposta consiste em aplicar técnicas de agrupamento sobre as SPD de usinas hidrelétricas. Onde as SPD foram obtidas a partir de estudos de planejamento da coordenação hidrotérmica. O objetivo era obter um modelo equivalente que refletisse o comportamento hidrológico em um número menor de SPD preservando as dependências temporal e espacial dos dados originais.

Dois processos de agrupamento foram descritos: Um que trata os dados em uma única etapa (*AMU*) e outro que trata os dados em duas fases utilizando um Mapa de Resultados para interligá-las (*AMRM*).
A metodologia proposta foi aplicada a uma caso representando uma configuração real do SEB, e os resultados da analise de confiabilidade composta foram comparados em relação as resultados obtidos sem utilizar os processo de agrupamento (*APA*) em termos de acuidade e esforço computacional.
Foi observado que ambos os processos foram eficientes em calcular o valor médio da LOLP ficando este próximo ao valor médio determinado pelo *APA*. Entretanto, ambos os processos tendem a concentrar os resultados da LOLP ao redor da média, reduzindo assim a dispersão apresentada pelos resultados obtidos pelo *APA*.
Tanto o *AMU* quanto o *AMRM* foram eficientes em termos da redução do número de casos de confiabilidade, permitindo uma redução de 65 casos (avaliados pelo *APA*) para 17 e 4 casos, respectivamente. Esta redução permitiu a redução no tempo de processamento de 73% e 94%, respectivamente.

7.0 - REFERÊNCIAS BIBLIOGRÁFICAS

(1) R.Billinton, R.N. Allan, *Reliability Evaluation of Power Systems*, Pitman Advanced Publishing Program, 1984.
(2) A. C. G. Melo, et al. “Probabilistic Analysis in Power System Performance Evaluation – New Developments in the NH2 Model”, VI Symposium of Specialists in Opera-tional and Expansion Planning (VI SEPOPE), Salvador, BA, May 1998.
(3) A.M. Leite da Silva, A.C.G. Melo, S.H.F. Cunha, "Generation Capacity Reliability Evaluation of Large Scale Hydrothermal Systems”, IEE Proceedings, Part C, Vol. 138, No. 1, January 1991.
(4) S. M. Amado, et al., “Making Hydro More Competitive Through Integrated Technologies”. Hydro 2000, Bern, Switzerland, October, 2000.p. 375-383.
(5) M. E. Maceira, et al. “Chain of Optimization Models for Setting the Energy Dispatch and Spot Price in the Brazil-ian System”. 14th. PSCC, Sevilha, Spain, 2002.
(6) M. A. M. Rodrigues, et al, “Sistema de Encadeamento de Modelos Energéticos”. In: XVI SNPTEE, Grupo VII, Planejamento de Sistemas Elétricos, Outubro, 2001.
(7) A. C. G. Melo, et al. “Probabilistic Adequacy Evaluation of Large Scale Power Systems – A Brazilian Case”, IEEE Stockholm Power Tech, Stockholm, Sweden, July 18- 25, 1995.
(8) K. C. Gowda, E. Diday, “Symbolic Clustering Using a New Similarity Measure”. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 22, No. 2, March/April 1992, p. 368-378.
(9) L. Kauffman, P. J. Rousseeuw, “Finding Groups in Data: An Introduction to Cluster Analysis”. New York: Willey Series, LTD, 1990. 368 p.
(10) P. S. Nagpaul, “Guide to Advanced Data Analysis”, Using IDANS Software. New Delhi, India. Available in: http://www.unesco.org/webworld/idams/ advguide/Chapt7-1-1.htm. Accessed: 22/10/2004.
(11) M. J. Van Der Laan, K. S. Pollard, J. Brian, “A New Partitioning Around Medoids Algorithm”. U. C. Berkeley Division of Biostatistics Working Paper Series, 2002. Available in: http://www.bepress.com/ ucbbio-stat/paper105. Accessed: 22/10/2004.
(12) J. C. Bezdek, R. Ehrlich, W. Full, “FCM: Fuzzy C-Means Algorithm”, Computer and Geoscience, Vol. 10, 1984, p. 191-203.
(13) F. Höppner, F. Klawonn, R. Kruse, T. Runkler, “Fuzzy Cluster Analysis – Methods for Classification, Data Analysis and Image Recognition”. Chihester: John Willey & Sons, LTD, 1999, 289 p.
(14) R. N. Davé, R. Krishnapuram, “Robust Clustering Meth-ods: A Unified View”. IEEE Transactions on Fuzzy Sys-tems, Vol. 5, No. 2, May 1997, p. 270-293.
(15) W. Härdle, Y. Mori, P. Vieu, “Statistical Case Studies”. MD Tech, 2004. SOFYAN, H., cap 19: “Fuzzy Cluster-ing”, March 20, 2003. Available in: http://www.quantlet.com/mdstat/scripts/xcs/pdf/xcspdf.pdf. Accessed: 22/10/2004.
(16) A. K. Jain, M. N. Murty, P. J. Flyn, “Data Clustering: A Review”. ACM Computing Surveys, Vol. 31, p. 264 - 323, 1999.
(17) F. Klawonn, F. Höppner, “What is Fuzzy About Fuzzy Clustering? Understanding and Improving the Concept of the Fuzzifier”. Advances in Intelligent Data Analysis, p. 254-264, 2003.
(18) J. S. R. Jang, C. T. Sun, E. Mizutani, ‘Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence”. Prentice Hall, Upper Saddle River, New Jersey, 1997.
(19) J.C. Dunn, “A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clus-ters”, Journal of Cybernetics 3: 32-57. 1973.
(20) M. Halkidi, Y. Batistakis, M. Vargiannis, “On Clustering Validation Techniques”. Journal of Intelligent Information Systems, Kluwer Academic Publishers, vol. 17, 2/3, Hol-land, 2001a, p. 107-145.
(21) M. Halkidi, M. Vargiannis, “Clustering Validity Assess-ment: Finding the optimal partitioning of a data set”. In: IEEE International Conference on Data Mining (ICDM 2001) Proceedings, San Jose, California, USA, 2001b, p. 187-194.
(22) M. K. Pakhira, S. Bandyopadhyay, U. Maulik, “Validity Index for Crisp and Fuzzy Clusters”. Pattern Recognition, Vol. 37, p. 487-501, 2004.
(23) C. R. Franco, L. S. Vidal, A. J. O. Cruz, “A Validity Measure for Hard and Fuzzy Clustering Derived From Fisher’s Linear Discriminant”. In: 2002 IEEE Interna-tional Conference on Fuzzy Systems, Honolulu, May, 2002.