



V SBQEE

Seminário Brasileiro sobre Qualidade da Energia Elétrica

17 a 20 de Agosto de 2003

Aracaju – Sergipe – Brasil



Código: AJU 04 104

Tópico: Análise, diagnósticos e soluções

METODOLOGIAS PARA IDENTIFICAR ASSOCIAÇÃO ENTRE PADRÕES CLIMÁTICOS E QUALIDADE DE FORNECIMENTO DE ENERGIA ELÉTRICA

Vilson R. Wronski
Sílvia M. Nassar

Jorge Coelho
Edward Gauche

Hermes Queirós

Marilene de Lima
Cristina Lourenço

INE/UFSC

EEL/UFSC

CELESC

EPAGRI/CLIMERH

RESUMO

Este artigo se propõe, a partir de um banco de dados com informações meteorológicas e de um banco de dados com índices de desempenho de empresas do setor elétrico, apresentar a metodologia utilizada para identificar padrões climáticos que influenciam a qualidade do fornecimento da região em estudo (índices de continuidade de frequência e duração da falha, número de atendimentos e tempo médio de atendimento).

PALAVRAS-CHAVE

Índices de Continuidade, Redes Neurais, Qualidade da Energia, Estatística Multivariada, Data Base.

1.0 - INTRODUÇÃO

No atual modelo adotado para o setor elétrico brasileiro é evidente a preocupação das empresas em minimizar as interrupções do fornecimento de energia, seja devido a aspectos técnicos, econômicos, seja devido a problemas de sua imagem perante o consumidor ou devido às exigências do órgão regulador. Assim, as empresas distribuidoras de energia elétrica devem ter uma noção clara dos motivos das interrupções ocorridas em seus alimentadores, de modo a precaver-se contra as interrupções intempestivas e, portanto, não programadas [1]. Este artigo apresenta uma metodologia que equaciona a influência da condição climática nos índices de continuidade de fornecimento de

energia elétrica. Esta metodologia envolve a Estatística e RNA (Redes Neurais Artificiais).

A Estatística é uma ferramenta de análise de dados que permite extrair informações de observações de uma realidade, e assim realizar um diagnóstico situacional [2]. E para caracterizar os agrupamentos de dias em relação as variáveis climáticas foram utilizados, principalmente, procedimentos estatísticos de análise multivariada.

Uma outra técnica utilizada é Redes Neurais Artificiais que tenta reproduzir em computadores digitais baseados em instrução (CBI) a rede de neurônios cerebrais, com todas suas características e potencialidades [3].

2.0 - ANÁLISE ESTATÍSTICA

Os métodos estatísticos multivariados são úteis para identificar grupos de variáveis e de indivíduos, por exemplo: análise de correspondência múltipla, análise por componentes principais, e análise de *clusters*. O primeiro procedimento permite identificar a associação entre variáveis e é adequado para um conjunto de variáveis qualitativas ou categóricas, e os outros dois para variáveis quantitativas. Tais métodos utilizam toda a informação dos dados estruturados em uma base de dados, onde as linhas representam os indivíduos e as colunas as variáveis. A análise estatística multivariada busca identificar padrões de associação considerando a variabilidade dos dados, dessa forma aumentando a capacidade de diagnóstico estatístico de uma realidade de interesse.

2.1 Análise por Componentes Principais

A Análise por Componentes Principais consiste em obter conhecimento na forma digital e analógica. Isto é, representar os dados de um conjunto de variáveis quantitativas, que contém todas as informações coletadas de um fenômeno, em tabelas de frequência e representações gráficas interpretáveis.

Na Análise de Componentes Principais existe uma reta com origem no centro de gravidade e na direção de maior espalhamento, isto é, a direção da reta onde ocorre a maior variância (ou maior inércia). Nesta reta está o primeiro Eixo Fatorial ou a primeira Componente Principal. E, ortogonalmente a esta, na direção do segundo maior espalhamento está o segundo Eixo Fatorial ou a segunda Componente Principal. A rigor existirão tantas componentes quantas forem as variáveis envolvidas na análise. Mas, como o grau de representatividade da Componente Principal é a sua inércia, então em geral, as primeiras componentes têm inércias altas, restando para as demais componentes representatividades pouco significantes.

Os autovalores de uma matriz estão associados à inércia explicada pelas Componentes Principais e são obtidos pela aplicação da álgebra matricial sobre a matriz de dados. Para maiores formalidades do método, ver as obras referenciadas [4]-[8].

2.2 Análise de Cluster

A utilização do método de análise de *cluster* ou de agrupamentos permite que sejam identificados agrupamentos de variáveis e de indivíduos. Algebricamente, considera-se uma matriz com p variáveis e n indivíduos e definem-se duas questões: a medida de similaridade e o método de agrupamento.

Como medidas de similaridade podemos ter a distância Euclidiana, distância Minkowski e a distância Chebychev e como método de agrupamento o hierárquico e o não hierárquico. Como resultados estatísticos temos sempre a matriz de distância e o dendograma.

O método de agrupamento hierárquico por agregação pode ser: simples (vizinho mais próximo), completa (vizinho mais distante), centróide ou média e de Ward.

O método de agrupamento não-hierárquico, K -médias, é utilizado quando o número de objetos é muito grande. É necessário estabelecer a priori o número K de agrupamentos desejados, escolher o centro inicial de cada grupo e a agregação. Existem diferentes algoritmos, por exemplo: maximização de distâncias entre os agrupamentos, classificação das distâncias e consideração de observações com intervalos

constantes e escolha das primeiras K (número de) observações para agrupamentos.

3.0 - REDES NEURAIAS ARTIFICIAIS

A Inteligência Artificial Conexionista (RNA) é mais utilizada na solução de problemas mal definidos, isto é, onde falta o conhecimento explícito (modelagem matemática/função de transferência) para solucioná-los [3]. Pode-se então afirmar que a técnica de RNA deve aportar qualidade à solução de problemas onde se disponha de um banco de dados das variáveis relacionadas ao problema, e não seu modelo matemático (na maioria das vezes difícil de ser obtido).

Em Redes Neurais Artificiais a aprendizagem (aquisição de conhecimento) se dá a partir de um conjunto de dados das variáveis relacionadas ao problema. Uma vez concluído o treinamento da RNA, esta se encontra apta a solucionar o problema para o qual foi treinada (com a precisão desejada) [9].

O problema de determinação de condições climáticas semelhantes é um problema típico para solução com RNA de treinamento não supervisionado. As RNA que se baseiam neste tipo de aprendizado foram popularizadas por Kohonen, sendo assim denominadas Mapas Auto-Organizáveis de Kohonen ou *Kohonen's Self-Organizing Maps (Kohonen's SOM)* [10].

4.0 - OBTENÇÃO DOS DADOS

Em geral os dados são estruturados numa base de dados, onde nas linhas estão contidos os indivíduos e nas colunas as variáveis [2], que são os atributos das diferentes características avaliadas.

Para analisar a associação entre condições climáticas e qualidade do fornecimento de energia elétrica enfrenta-se o seguinte problema: extrair e concatenar dados que estão em bases de dados diferentes, e que foram criadas para atender necessidades informacionais também diferentes. A dificuldade inicial é a concepção da estrutura da base necessária à análise estatística onde o que necessita ser analisado deve constituir uma variável para análise estatística e, portanto deverá estar disposto numa coluna da base de dados. Isto leva a complicações na extração dos dados nas bases originais e sua formatação em uma nova base de dados.

4.1 Base de Dados – Variáveis Climáticas

No presente estudo foram utilizados dados meteorológicos da estação meteorológica de São José, no Estado de Santa Catarina pertencentes ou sob a manutenção da Empresa de Pesquisa

Agropecuária e Extensão Rural de SC (EPAGRI). Esta estação é padronizada conforme determinação da Organização Meteorológica Mundial (OMM), com leituras diárias e simultâneas na rede de estações meteorológicas em todo o Globo, em três horários padrões, chamados sinóticos (12:00TMG, 18:00TMG e 00:00TMG – TMG: Tempo Médio de Greenwich) correspondendo as 09:00h, 15:00h e 21:00h, horário de Brasília, respectivamente.

Da base original de dados climáticos definiu-se a extração das seguintes variáveis: precipitação acumulada em 24 horas, temperatura mínima e temperatura máxima, umidade relativa e velocidade do vento.

4.2 Base de Dados – Variáveis de Fornecimento

A Companhia de Energia Elétrica de Santa Catarina (CELESC) coleta os dados de atendimento e de interrupção no fornecimento e armazena no sistema de Manutenção e Operação da Distribuição numa base de dados *Oracle*.

Da base original de dados de desempenho foram extraídas as seguintes variáveis de fornecimento: número de atendimentos, DEC, FEC e TMA (valores diários). Foi solicitado ainda, sistema afetado e causa da interrupção.

Nessa coleta de dados não foram consideradas as interrupções programadas, interrupções no Sistema de Transmissão e Geração (Sistema externo) e interrupções menores que três minutos.

5.0 - CARACTERIZAÇÃO CLIMÁTICA

Nesse artigo focaremos a análise para o mês de dezembro de 2001. Neste mês, as chuvas ficaram abaixo da média climatológica em todo o Estado, inclusive na cidade de São José, onde choveu apenas 65% do que normalmente chove. Em relação às temperaturas, estas ficaram ligeiramente acima do normal. O mês de dezembro de 2001 foi considerado um mês atípico em relação a suas normais climatológicas

6.0 - VARIÁVEIS TÉCNICAS - DESCRITIVAS

Para o período anual analisado observa-se, na tabela 1, que a média do número de atendimentos (NA) foi de 17,12 atendimentos/dia e que no mês de Dezembro a média de NA foi superior à média anual. A variabilidade do NA diário, mostrada pelo desvio padrão (dp) foi maior, também, no mês de Dezembro.

Na tabela 1 pode-se observar que o tempo médio diário de atendimento TMA foi de 83,50 minutos, apresentando considerável variabilidade (desvio padrão igual a 42,35 minutos).

Tabela 1. Distribuição diária de variáveis técnicas para os meses de Dezembro de 2001 e média Anual. São José, Santa Catarina

Indicadores por Dia	DEZEMBRO		ANUAL	
	média	dp	média	dp
NA	20,70	11,33	17,12	7,54
TMA (min)	105,01	66,20	83,50	42,35
DEC (min)	6,3603	26,0223	2,2176	8,6183
FEC	0,0514	0,1492	0,0289	0,0741

Destaca-se que dentre todos os meses, em dezembro ocorreu o maior tempo médio de atendimento. Observa-se ainda que o impacto de falhas de continuidade de fornecimento no mês de dezembro foi aproximadamente três vezes maior, tanto em DEC quanto em FEC, que a média anual.

7.0 - INVESTIGAÇÃO DE AGRUPAMENTOS

Num primeiro momento da investigação de agrupamentos de variáveis climáticas utilizou-se o procedimento Análise Fatorial por Componentes Principais, que permite explorar a associação entre variáveis quantitativas, como é o caso das variáveis climáticas: velocidade do vento, umidade, precipitação e temperatura; todas analisadas para os valores diário máximo e mínimo, exceto a precipitação que foi a acumulada para as 24 horas.

Na figura 1 observa-se no plano fatorial delimitado pelos dois primeiros fatores onde há dois grupos de variáveis bem caracterizados: velocidade do vento (vmin e vmax), e as relativas a umidade do ar e precipitação (lp24, umin e umax).

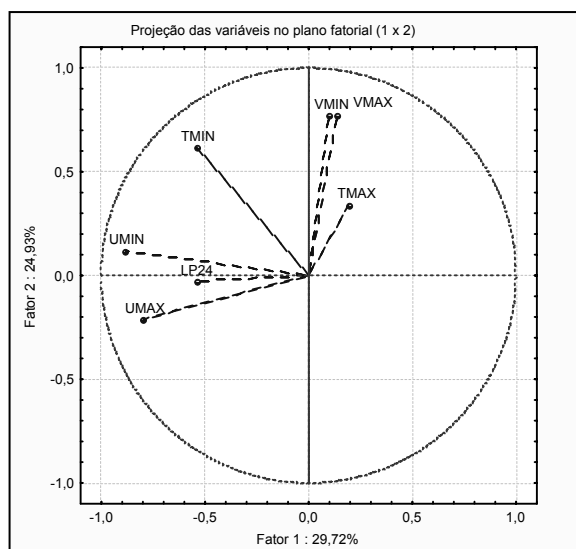


Figura 1. Análise Fatorial por Componentes Principais para as variáveis climáticas. Mês de Dezembro de 2001. São José, Santa Catarina.

Um terceiro agrupamento formado pelas temperaturas (tmin e tmax) está mais bem representado no terceiro fator. Estes três agrupamentos aqui serão chamados de condição: vento, chuva e temperatura, respectivamente, havendo um maior contraste entre as condições vento e chuva. Como os dias em análise são para o mês de dezembro, fica intuitivo tal contraste, pois é um mês de verão onde as temperaturas são menos heterogêneas.

Este agrupamento de variáveis é corroborado pelos resultados estatísticos na tabela 2, onde para três fatores obtém-se um percentual de inércia de 74,55%. Lembrando que o percentual de inércia informa quanto da variabilidade dos dados é explicada pela representação em cada fator.

Tabela 2. Descrição dos autovalores e inércia no espaço fatorial para as variáveis climáticas. Mês de Dezembro de 2001, São José, Santa Catarina.

Fator	Autovalor	% de Inércia
1	2,08	29,72
2	1,74	24,93
3	1,39	19,90
4	0,71	10,10

Por outro lado, decidiu-se também investigar o agrupamento de variáveis utilizando outro procedimento estatístico chamado de Análise de Cluster, considerando a distância Euclidiana para a medida de distância. Na figura 2 observam-se três grupos formados pela velocidade do vento (vmin e vmax), pela umidade (umin e umax) e pelas temperaturas (tmin e tmax). A precipitação diária (lp24) apresenta as maiores distâncias, isto é, maiores contrastes dentro do mês analisado.

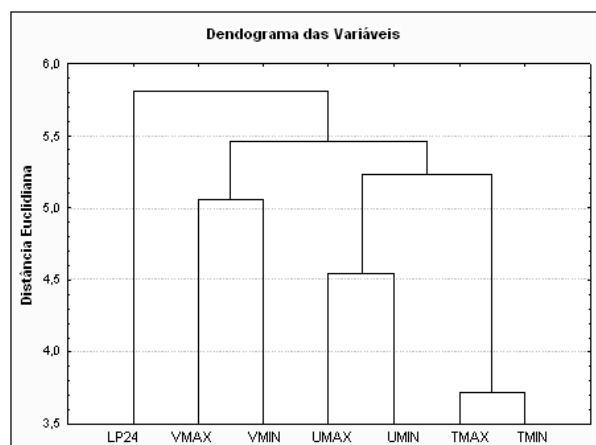


Figura 2. Análise de Cluster para as variáveis climáticas. Mês de Dezembro de 2001, São José, Santa Catarina.

Dando seqüência a análise dos dados será agora apresentado os resultados estatísticos para a identificação de grupos de dias com condições climáticas similares. Foram utilizados dois

métodos na Análise de Cluster: o método hierárquico e o método das K-médias, ambos considerando a distância Euclidiana como a medida de similaridade.

No método hierárquico observa-se a similaridade (distância) entre todos os dias do mês e verifica-se que as condições climáticas dos dias 22 e 23 contrastam com as dos demais dias. E que os dias 1, 26 e 27 apresentaram condições climáticas mais semelhantes no mês de dezembro.

Para identificar os dias de cada agrupamento de dias utilizou-se o método das K-médias estabelecendo que o número de clusters seria igual a 3. Na figura 3 observa-se as médias das variáveis climáticas para os três agrupamentos de dias, onde aparece com maior intensidade a condição vento, isto é as variáveis velocidade do vento máxima e mínima (vmax e vmin).

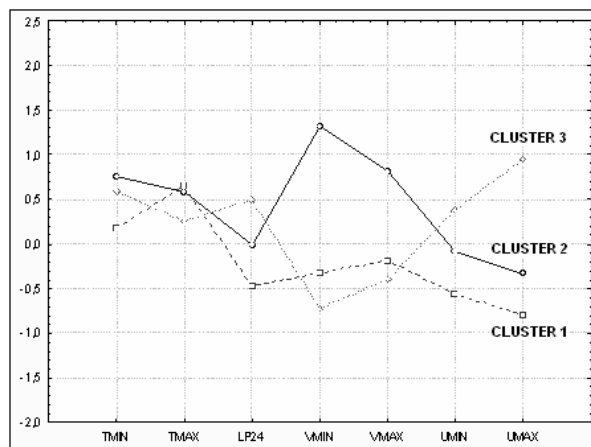


Figura 3. Análise de Cluster pelo método K-médias para agrupamento de dias segundo as variáveis climáticas. Mês de Dezembro de 2001, São José, Santa Catarina.

A análise estatística para avaliar a distância entre os agrupamentos de dias em relação as variáveis climáticas foi feita utilizando o procedimento ANOVA. Na tabela 3 observa-se que a variável temperatura máxima (tmax) não foi estatisticamente significativa para os agrupamentos de dias, e as demais foram as que contribuíram para formação dos agrupamentos de dias.

Tabela 3. Resultados estatísticos do teste ANOVA para as variáveis climáticas na formação dos agrupamentos de dias. Mês de Dezembro de 2001, São José, Santa Catarina.

Variáveis	Rótulo	Estatística F	p-valor
Log Precipitação 24 hs	LP24	2,617	0,090831
Temperatura min	TMIN	3,385	0,048218
Temperatura max	TMAX	1,244	0,303643
Velocidade Vento-min	VMIN	41,464	0,000000
Velocidade Vento-max	VMAX	7,967	0,001824
Umidade - min	UMIN	4,581	0,019009
Umidade - max	UMAX	22,999	0,000001

Como resultado da análise de clusters pelo método K-médias foi identificados três grupos de dias (grupo 1, 2 e 3), os quais podem ser observados na tabela 4. Cada grupo contém 14, 9 e 8 dias, respectivamente.

Tabela 4. Identificação de agrupamentos de dias segundo o método das K-médias (E) e SOM (N).
Mês de Dezembro de 2001. São José, Santa Catarina.

DIA	E	N	DIA	E	N	DIA	E	N
1	1	3	11	3	3	21	1	1
2	1	1	12	3	1	22	2	2
3	1	1	13	3	1	23	2	2
4	3	1	14	2	3	24	1	1
5	3	3	15	2	2	25	3	3
6	3	3	16	1	1	26	1	1
7	1	1	17	1	1	27	1	3
8	2	2	18	1	1	28	2	1
9	2	2	19	2	1	29	1	1
10	3	1	20	2	2	30	1	1
			31	1	1			

Os agrupamentos de dias segundo os padrões climáticos foram chamados de:

- Cluster 1: predominância da variável Temperatura;
- Cluster 2: predominância da variável Vento;
- Cluster 3: predominância da variável Precipitação e Umidade (ou simplesmente Chuva).

8.0 - VALIDAÇÃO POR RNA

As técnicas de análise estatística multivariada e Kohonen's SOM trabalharão no sentido de uma validação concorrente dos padrões descobertos. A tabela 4 apresenta a identificação dos dias segundo cada cluster de condições climáticas (em número de 3) obtidos pela técnica SOM. Entre as duas técnicas K-médias (E) e SOM (N), observa-se que para o mês de Dezembro houve uma concordância de 71% dos dias.

No entanto, a comparação para 10 (dez) meses analisados obteve-se uma taxa média de concordância entre as duas técnicas de 93,73%, destacando-se taxas de concordância de 100% para os meses de Junho e Julho [11].

9.0 - VARIÁVEIS TÉCNICAS - ASSOCIAÇÃO

Inicialmente, investiga-se de maneira exploratória, o comportamento das variáveis técnicas de fornecimento: número de Atendimentos(NA), DEC, FEC e TMA, nos agrupamentos de dias identificados pelas variáveis climáticas.

Na tabela 5 são descritos os indicadores de fornecimento segundo os três grupos climáticos de dias, caracterizados predominantemente pela

temperatura (cluster 1), vento (cluster 2) e pela chuva (cluster 3).

Tabela 5. Distribuição diária de variáveis técnicas segundo os três agrupamentos climáticos de dias.
São José, Santa Catarina.

Variáveis Técnicas		Clusters			
		1-Temp.	2-Vento	3-Chuva	Anual
	N	158	81	126	365
NA	Média	16,41	17,95	17,49	17,12
	DP	7,28	9,42	6,40	7,54
TMA	Média	78,35	92,72	84,04	83,50
	DP	32,93	60,02	38,30	42,35
DEC	Média	1,7961	4,0035	1,8345	2,3061
	DP	5,1101	16,6730	2,9617	8,7774
FEC	Média	0,0262	0,0415	0,0274	0,0300
	DP	0,0689	0,1015	0,0617	0,0753

Observa-se que há evidências de maior número de atendimentos e maior tempo médio de atendimento nas chamadas emergenciais para a condição vento e chuva, e um menor número para a condição temperatura.

Verifica-se também, que há evidências de maior duração e frequência equivalente por consumidor nas chamadas emergenciais para a condição vento, e um menor número para a condição temperatura e chuva.

Para poder explorar a influência das condições climáticas, os indicadores de fornecimento nos três grupos de dias do mês devem ser comparados com o indicador mensal da tabela 1. Posteriormente, podem-se destacar as variáveis que mais influenciaram na classificação e analisá-las graficamente como na figura 4, a fim de confirmar os achados estatísticos.

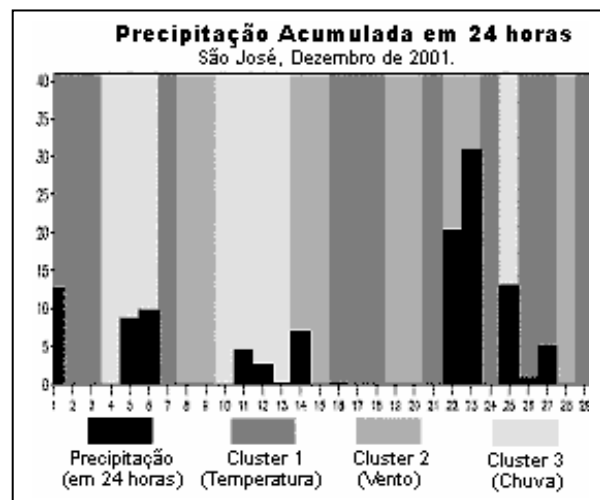


Figura 4. Precipitação acumulada em 24 horas e os agrupamentos de dias do mês de Dezembro de 2001.
São José, Santa Catarina.

Deste modo, a partir do conhecimento dos clusters climáticos e de posse da previsão de

variáveis climáticas, a empresa responsável pela distribuição de energia da localidade pode dimensionar as equipes de emergência para atender aos consumidores com os índices de qualidade e satisfação desejados.

Na figura 5 apresenta a estrutura inicial da rede bayesiana que expressa o espaço de associação entre as variáveis climáticas e o número diário de atendimentos de emergência.

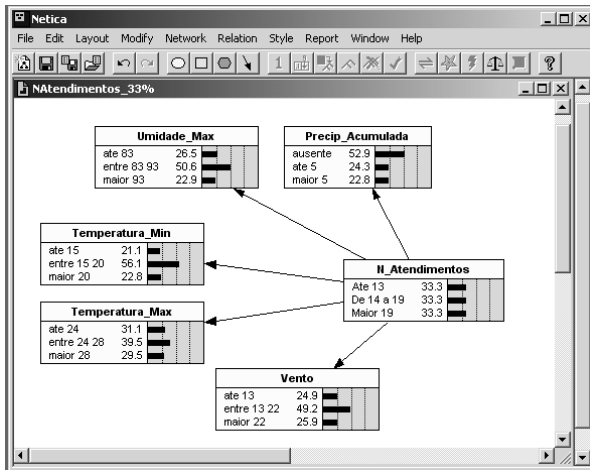


Figura 5. Rede bayesiana de associação entre as variáveis climáticas e o número diário de atendimentos emergenciais. São José, Santa Catarina.

Esta rede bayesiana será utilizada para prever a chance de cada faixa de número de atendimentos a partir das informações climáticas fornecidas pelo CLIMERH.

10.0 - AGRADECIMENTOS

Esta pesquisa está sendo desenvolvida com o trabalho cooperativo de profissionais de diversas áreas: Epagri / Climerh (Empresa de Pesquisa Agropecuária e Extensão Rural de SC / Centro Integrado de Meteorologia e Recursos Hídricos), CELESC (Centrais Elétricas de SC) e UFSC (Universidade Federal de SC) com profissionais dos departamentos de Engenharia Elétrica e de Informática e Estatística.

11.0 - CONCLUSÕES

Esta metodologia foi desenvolvida com forte caráter exploratório e que alcançou significativamente seus objetivos. Nesse aspecto podemos destacar o controle da empresa sobre os índices de satisfação de seus clientes e alcance de metas estipuladas pelo órgão regulador à empresa de distribuição. A empresa pode ainda, definir com mais segurança a quantidade de suas equipes de apoio para ficar de sobreaviso a possíveis anormalidades no

fornecimento provocadas por fenômenos climáticos.

Em relação às redes neurais artificiais o treinamento da rede SOM de Kohonen não é o foco principal desse artigo, porque foi amplamente discutida em [11].

A metodologia estatística apresenta certa vantagem sobre redes neurais, porque não precisa de um especialista com vasto conhecimento no assunto para o treinamento da rede. Por esse motivo RNA está sendo utilizada para validar os achados estatísticos.

Essa metodologia está sendo aplicada em duas cidades litorâneas e duas cidades do interior para descobrir o efeito de possíveis variações climáticas nas variáveis de desempenho no fornecimento de energia elétrica.

12.0 - REFERÊNCIAS BIBLIOGRÁFICAS

- [1] J. Coelho, E. Gauche, S.M. Nassar, V.R. Wronscki, H. Queiroz, M. de Lima, "Influence of weather variables in continuity levels of electrical power supply through artificial neural networks", in VIII SEPOPE, 2002, Brasília, Brazil, in CD.
- [2] M. Ohira, S.M. Nassar, A. H. Catapan, M.M. Reis, R.C. Bastos, V. R. Wronscki, "SEstat – Sistema de Ensino-Aprendizagem em Estatística", Version 2.0, Dezembro de 2002. [Online]. Disponível em <http://www.inf.ufsc.br/lea/>.
- [3] BARRETO, J. M., Inteligência Artificial no Limiar do Século XXI, 291 páginas, Duplic Edições, Florianópolis 1997.
- [4] JOHNSON, R. A. & WICHERN, D. W. Applied Multivariate Statistical Analysis, USA: New York, Prentice Hill, 4ª ed., 1998.
- [5] LEBART, L. & MORINEAU, A. & PIRON, M. Statistique Exploratoire Multidimensionnelle, France: Dunod, 1995.
- [6] BANET, T. A. & MORINEAU A. Aprender de los Datos: El Análisis de Componentes Principales, Espanha: Barcelona, EUB, 1ª ed. 1999.
- [7] ESCOFIER, B. & PAGÈS J. Análises Factorielles Simples et Multiples, France: Paris, Dunod, 2ª ed. 1990.
- [8] MANLY, B. F. J. Multivariate Statistical Methods, New Zealand: Otago, Chapman & Hall, 2ª ed, 1997.
- [9] Edward Gauche, Reconfiguração de Sistemas de Distribuição para Minimização de Perdas Resistivas com o Auxílio de Redes Neurais Artificiais, Tese de Doutorado, UFSC 1999.
- [10] Teuvo Kohonen, "Self-Organizing Maps", Springer-Verlag, 3rd Edition, 2001.
- [11] J. Coelho, E. Gauche, S.M. Nassar, V.R. Wronscki, H. Queiroz, M. de Lima, "Reliability Diagnosis of Distribution System under Adverse Weather Conditions", in IEEE BOLOGNA POWER TECH, June 2003, Italy, in CD.