

# MODELOS ESTOCÁSTICOS UNIVARIADOS E MULTIVARIADOS NA DETECÇÃO DE MUDANÇAS CLIMÁTICAS (\*)

Teresinha de Maria Bezerra S. Xavier (1)

**(\*) Palestra em mesa-redonda no II-SIC/SBMET  
(Simpósio Internacional de Climatologia), Nov. 2007**

**(1) ACECI Academia Cearense de Ciências (Diretora Científica); Profa. UFC / Universidade Federal do Ceará / Centro de Tecnologia; Professora/Pesq. e Conferencista Visitante no Depto. de Ciências Atmosféricas, IAG-USP; Membro do Fórum Brasileiro de Mudanças Climáticas.**

**E-Mail: [txavier@secrel.com.br](mailto:txavier@secrel.com.br) Rua Oswaldo Cruz 176/400, Fortaleza-CE, CEP 60.125-150, Cel. (0xx85-9983.7150),**

**A Climatologia, levando em conta definições e bases históricas, como também a maior parte da pesquisa nela desenvolvida, possui fundamentação e objetivos fortemente calcados em conceitos probabilísticos e estatísticos.**

**Por outro lado, a partir da segunda metade do século XX, seu estudo enfatiza a avaliação da ocorrência de flutuações, variações e mudanças climáticas, em especial nas últimas décadas, em que a preocupação da ciência se volta para aspectos ligados a mudanças climáticas de origem antrópica, envolvendo o “efeito estufa”.**

**Assim, nesta palestra, tratamos técnicas estocásticas, univariadas e multivariadas, direcionadas á detecção de variações, flutuações e, especialmente, mudanças climáticas. Não insistimos nas possíveis distinções entre tais conceitos. Por outro lado, as mesmas técnicas podem aplicar-se nestes vários contextos.**

**Variações e flutuações climáticas dizem respeito, é certo, a alterações do clima relacionadas a efeitos de natureza rítmica, donde não permanentes, ou em princípio, reversíveis. Em geral, relacionados a fatores que atuem de forma não persistente.**

**Enquanto se reservaria a designação de Mudanças climáticas às que se esperam sejam permanentes ou quase irreversíveis, em função de fatores naturais ou decorrentes de uma ação antrópica, continuada.**

**A técnicas a serem consideradas, são as seguintes (trata-se de lista não exaustiva).**

(i) Técnicas Univariadas clássicas, paramétricas e não-paramétricas. **Entre elas: a Análise de Variância (ANOVA) e a Regressão Linear e Curvilínea Simples, estas consideradas como procedimentos complementares, entre si;**

(ii) Técnicas Multivariadas. **Tais como: Análise de Variância Múltipla (MANOVA), Regressão Múltipla e Análise de Covariância;**

(iii) Modelagem **das variáveis envolvidas através de Leis Probabilísticas Discretas, Contínuas e Mistas;**

**[ *continua* ]**

(iv) Processos Estocásticos, **entre os quais** Análise de Séries Temporais, Modelos Markovianos, Modelos de Extremos, **etc**;

(v) **finalmente**, Métodos Fatoriais (**ACP, Análise Discriminante, etc.**), Métodos de Classificação Automática, rígida e “fuzzy”, e “Data Mining” (**mineração de dados**).

**Como seria impossível tratar sistematicamente, num curto espaço de tempo, todas estes assuntos, decidiu-se pela estratégia de mostrar variedade significativa de exemplos ou de sugestões para sua utilização, no contexto da análise de flutuações e/ou mudanças climáticas. Tais exemplos e/ou sugestões são extraídas dos nossos próprios trabalhos de pesquisa e/ou de nossa experiência pessoal, ou ainda da bibliografia nacional e internacional.**

## EXEMPLIFICAÇÃO

1. Análise de Variância (ANOVA) e Regressão tomadas como técnicas complementares no estudo de Mudanças Climáticas Urbanas

**Como primeiro exemplo, consideramos mudanças climáticas urbanas verificadas a partir dos dados da Estação Meteorológica do IAG/USP, com relação ao comportamento da temperatura de superfície e também da umidade relativa do ar.**

**De fato, a EM do IAG / USP considera-se como um laboratório natural no estudo de alterações ligadas à urbanização e à poluição.**

**Há há uma explicação para isto. Nos idos de 1936 o Parque do Estado podia considerar-se como “extra-muros” à cidade de São Paulo. Porém, com o avanço da urbanização, ocorreu por assim dizer um “sequestro” das condições climáticas originais do Parque do Estado, que aos poucos se torna um enclave verde rodeado por áreas densamente edificadas.**

**Ora, conforme as FIGURAS 1 & 2, verifica-se uma tendência nítida para o aumento da temperatura e, paralelamente, diminuição da umidade relativa do ar, ao longo do período de 70 anos de observações, a saber, 1936 até 2005.**



**Demais, por um outro motivo, os resultados podem ser considerados indubitáveis, pois as séries climáticas submetidas às análises são confiáveis, no sentido de envolverem contínuo e rigoroso controle de qualidade durante as várias fases de aquisição, transcrição e salvaguarda dos dados respectivos. (\*)**

**Note-s que todos estes resultados e figuras foram tomadas de Xavier, Pereira & Xavier (2007), artigo apresentado neste II-SIC/SBMET.**

**(\*) Com efeito, os resultados serão garantidos, apenas se as técnicas forem aplicadas corretamente e, ao mesmo tempo, se os dados forem confiáveis !**

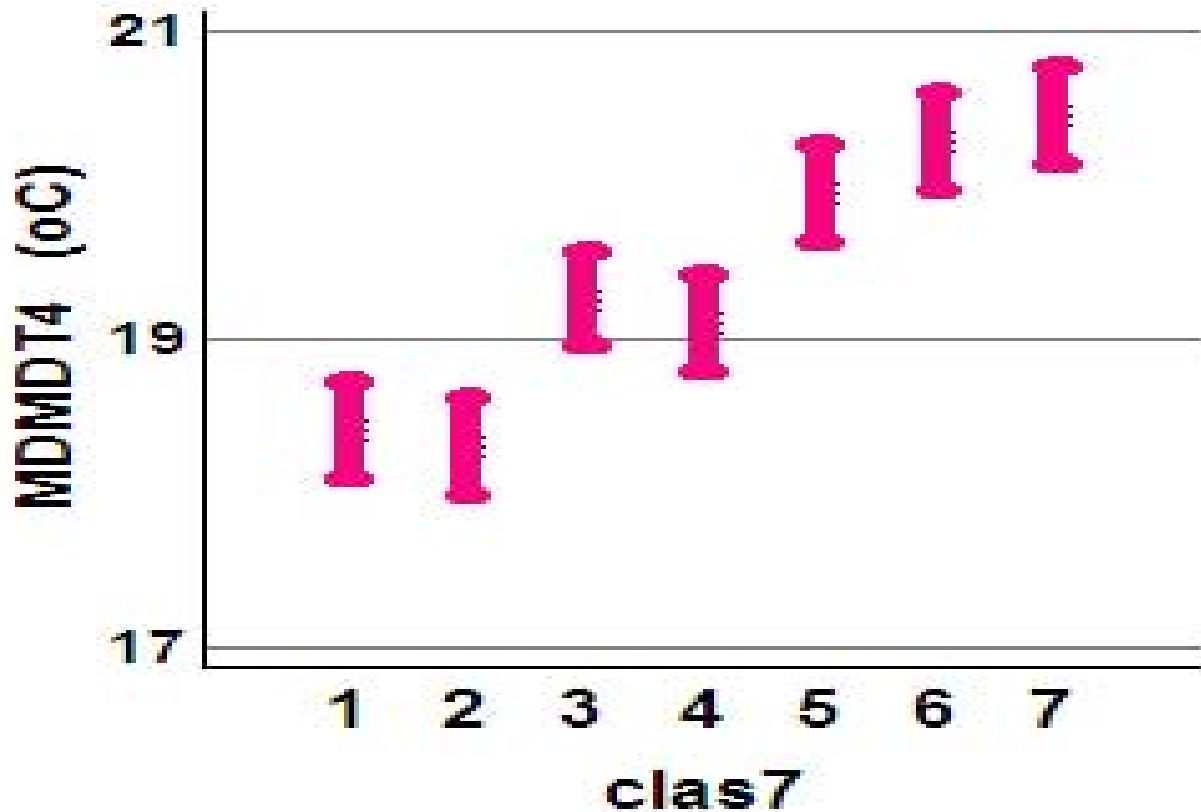


Figura 1 ANOVA EM/IAG/USP 1936 a 2005  
temperaturas médias no trimestre 4 K=7 classes

**Na Figura 1 precedente denota-se, claramente, o aumento progressivo das temperaturas (trimestre 4 = outubro-dezembro) levando em conta as posições relativas dos intervalos de confiança para as classes consecutivas: 1936-45, 46-55, 56-65, 66-75, 76-85, 86-95, 96-2005.**

**A probabilidade de erro na ANOVA foi praticamente nula:  $p=0,0000$ , donde o resultado obtido possui uma elevadíssima significância do ponto de vista estatístico.**

**Em seguida, passa-se a examinar a mesma questão, sob outra perspectiva. Isto é, pela regressão linear.**

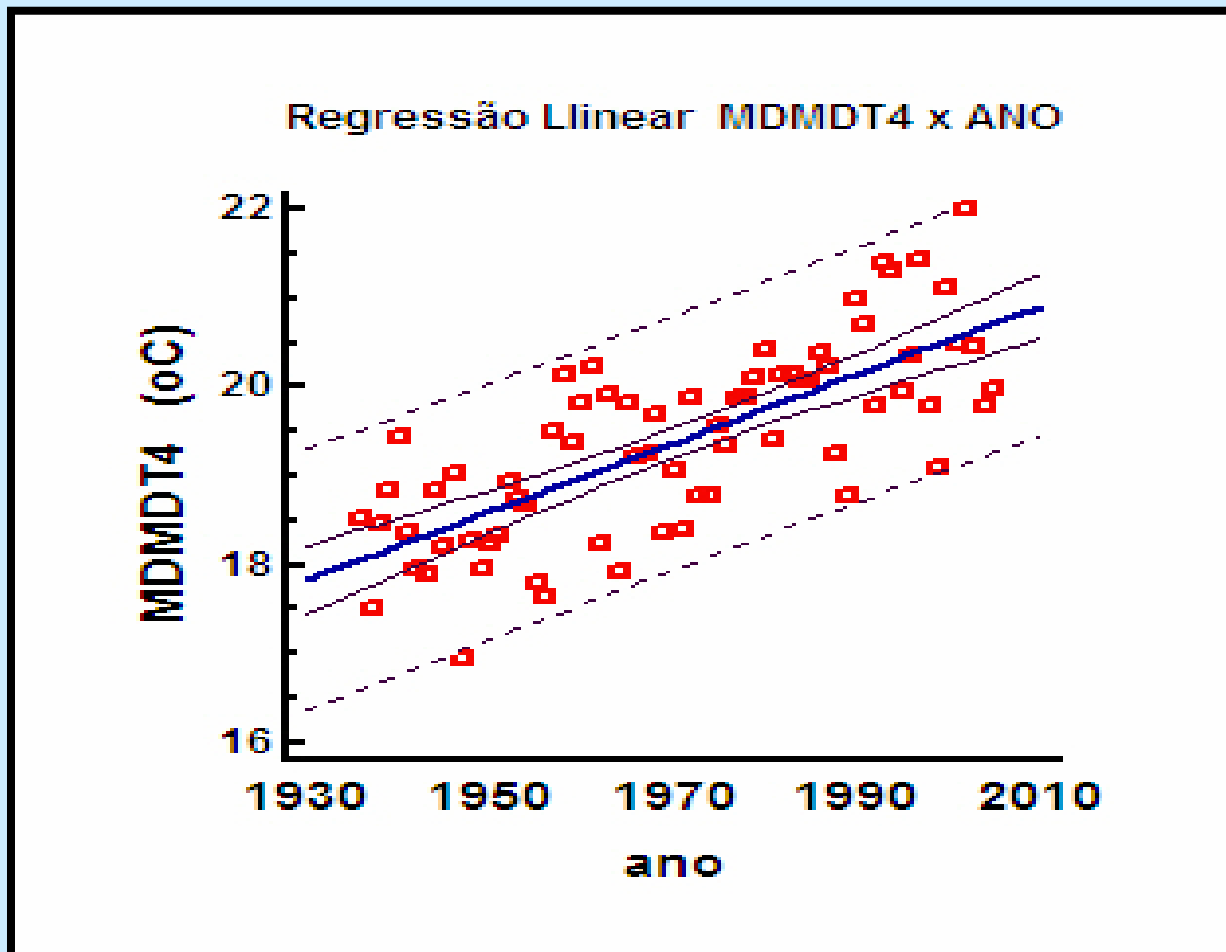


Figura 2 Regressão linear EM/IAG/USP

Evidência de crescimento da temperatura de superfície no 4º trimestre, período 1964-2005 (regressão com  $p = 0,0000$ )

**De fato, nos quatro trimestres do ano (jan.-mar.,..., out.-dez.) em 1936-2005, os incrementos globais foram de 1,96; 2,26; 1,76 e 2,68 graus C, resp., com probabilidade de erro  $p=0,0000$ , com base nas regressões.**

**O cálculo (estimativa) é simples: a medida da “inclinação da reta” (ou “slope”) estima o aumento ano a ano. Logo, multiplicando por 70 (número de anos) obtemos a estimativa do incremento de temperaturas para todo o período 1936-2005.**

**Análogos resultados valem para outras faixas sazonais (meses, bimestres, etc.) e, ainda, para temperaturas mínimas e médias.**

**Deve-se ter em mente, contudo, que mediante uma análise de variância obtém-se, apenas, informação sobre o comportamento global das médias respeito às várias classes.**

**Assim, quando da rejeição da “hipótese nula”, o teste somente informa que as classes ou grupos não podem ser considerados “iguais” ou provindos de uma mesma “população”. Porém, nada é adiantado sobre a exata localização das eventuais diferenças entre as médias e, em especial, sobre suas significâncias estatísticas, individualizadas.**

**Para esta última finalidade estão disponíveis testes complementares para comparações múltiplas de médias, em geral incorporados aos módulos de ANOVA existentes em vários “pacotes computacionais”.**

**Ademais, neste procedimento estatístico estão disponíveis várias saídas tabulares e gráficas, como:**

- 1. “tabela da análise de variância”, que consolida um resumo da execução do procedimento e, inclusive, com informação sobre a significância estatística alcançada globalmente (explicitada por uma “probabilidade de erro”  $p$ );**

2. **tabela para o comportamento estatístico das classes (médias, variabilidade, etc.);**
3. **gráfico com os intervalos de confiança das médias, como na Figura 1 ;**
4. **enfim, saídas para os testes de comparações múltiplas, também incorporando suas significâncias estatísticas, individualizadas; etc.**



**Note-se que a análise de variância clássica, ou ANOVA, designa-se como paramétrica. Pois depende de pressupostos, como normalidade e homocedasticidade (homogeneidade das variâncias), embora seja relativamente imune a pequenos ou médios afastamentos dos pressupostos por se tratar de procedimento estatístico considerado “robusto”.**

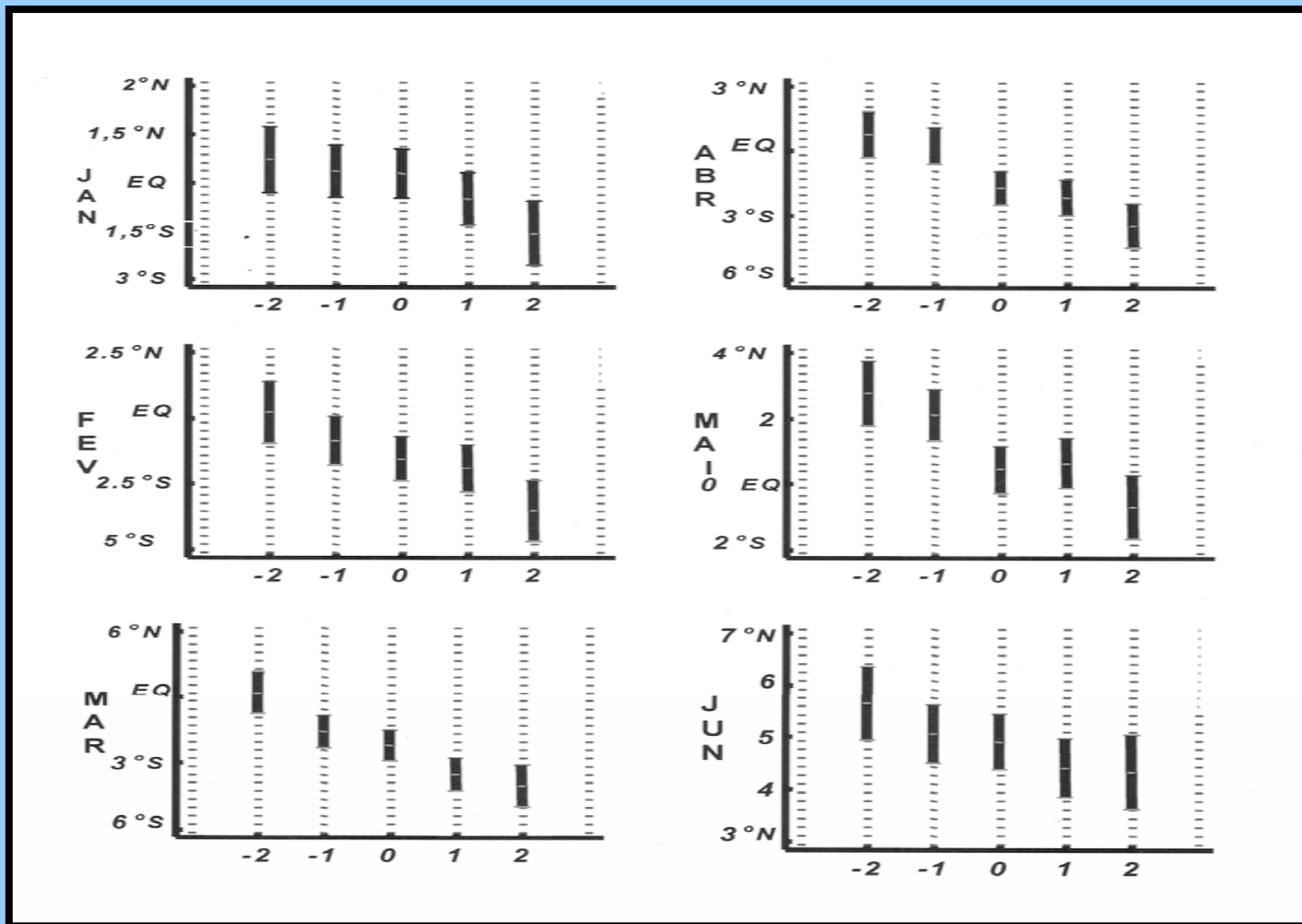
**Contudo, para dirimir eventuais dúvidas, pode-se empregar uma análise de variância não-paramétrica, não submetida a tais pressupostos. Trata-se do teste de Kruskal-Wallis (de fato, nele, testa-se a igualdade das medianas e não das médias). Como uma referência geral sobre testes não-paramétricos, veja-se *Conover (1971 / 2001)*; também, *Siegel (1956)*.**

## 2. Análise de Variância (ANOVA) para “flagrar” o papel da posição da ZCIT sobre as chuvas durante a “quadra chuvosa” no Ceará e Nordeste Setentrional

**Neste contexto, consideramos** Intervalos de Confiança para as Latitudes Médias da Posição da ZCIT ao longo do Meridiano de 37°W, de Janeiro a Junho, **relativamente às 5 classes para a chuva :**

**-2 = MS = Muito Seco, -1 = S = Seco, 0 = N = Normal, 1 = C = Chuvoso e 2 = MC = Muito Chuvoso, durante a “quadra chuvosa” (fevereiro-maio) na Bacia Metropolitana (Ceará), conforme os dados pluviométricos sobre a referida bacia, no período 1964-1998.**

**Veja-se a FIGURA 3.**



**FIGURA 3** BACIA METROPOLITANA - ESTADO do CEARÁ  
 Intervalos de Confiança para as Latitudes Médias da ZCIT  
 em função da chuva (5 classes) na referida bacia

**Os resultados, do ponto de vista gráfico, na Figura 3, são fáceis de interpretar. Ou seja, nos vários meses ocorre uma tendência da ZCIT permanecer numa posição média mais ao norte, quando o ano for MUITO SECO ou SECO; ou descer ao sul da linha equatorial, em média, se o ano for CHUVOSO ou MUITO CHUVOSO.**

**De fato, tratam-se de “gráficos de intervalos de confiança” na ANOVA (Análise de Variância). No caso, confiança de 95%, donde com erro máximo da ordem de 5%.**

**Resultados análogos foram obtidos para outras bacias, no Ceará. fonte – Xavier et al. (2003), Rev. Bras. Recursos Hídricos-ABRH, Vol. 8 (2), pp. 19-37.**

### 3. Análise de Variância e Regressão para detectar "mudanças climáticas na superfície" dos oceanos

**Um artigo está sendo apresentado neste II-SIC-SBMET sobre mudanças climáticas no Atlântico Intertropical em termos da TSM - temperatura da superfície do mar: Xavier, Xavier, Silva-Dias e Silva-Dias (2007). De fato, dando continuidade a artigos anteriores, ou seja :**

I-SIC-SBMET-2005, e CBMET-Congresso Brasileiro de Meteorologia-2006, **etc.**

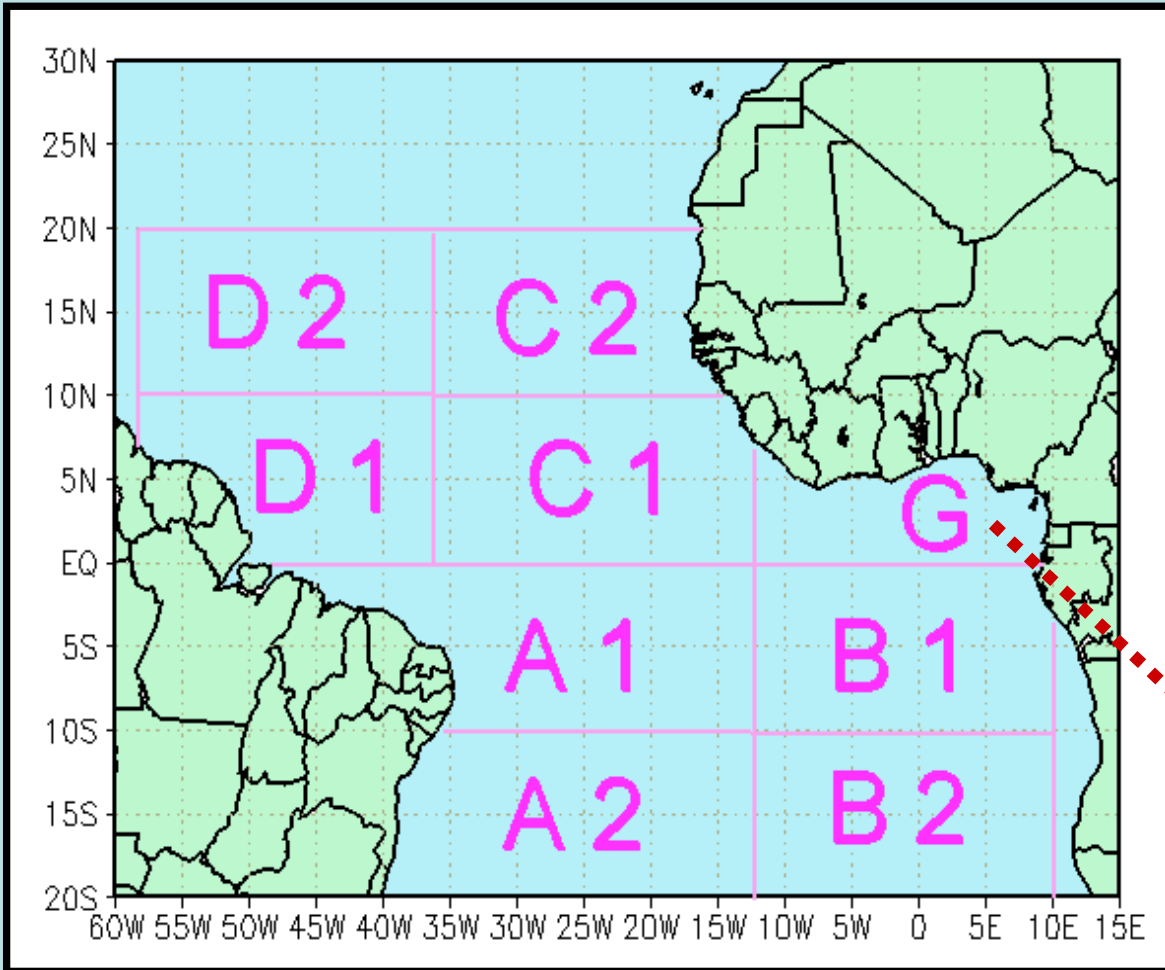
**O mesmo se aplica à análise de mudanças no regime de ventos, como também da TSM, conforme artigo de Xavier, A.F.S., Xavier, T. de Ma. B.S. & J.Ma.B. Alves, no CBMET-Congresso Bras. de Meteorologia-2000.**

**No “slide” seguinte, são mostradas alterações da TSM no “Golfo da Guiné”, em termos de aumentos nos meses de março e dezembro, no período 1964-2005.**

**Na FIGURA 4, temos a indicação da localização do Golfo da Guiné, ns costa da África. Na FIGURA 5, os Intervalos de confiança para os valores da TSM obtidos pela ANOVA, com  $K = 3$  classes. Ou seja:**

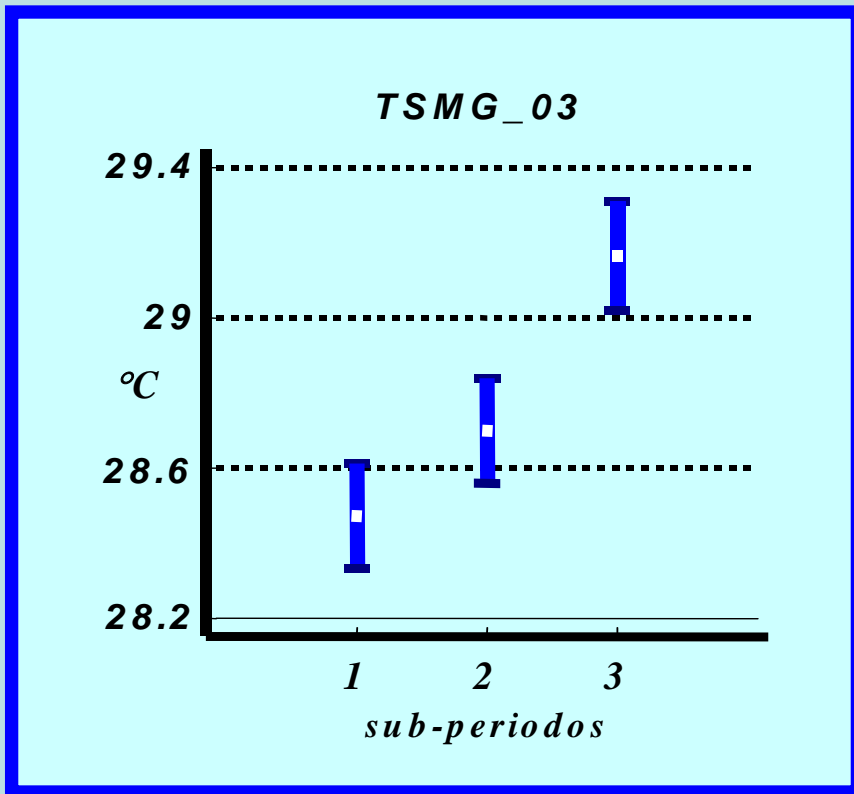
**[ no 1o semestre, 1 = 1964-77, 2 = 1978-91 e 3 = 1992/2005; no 2o semestre, 1 = 1963-76, 2 = 1977-90, 3 = 1991-2004 ]**

Figura 4

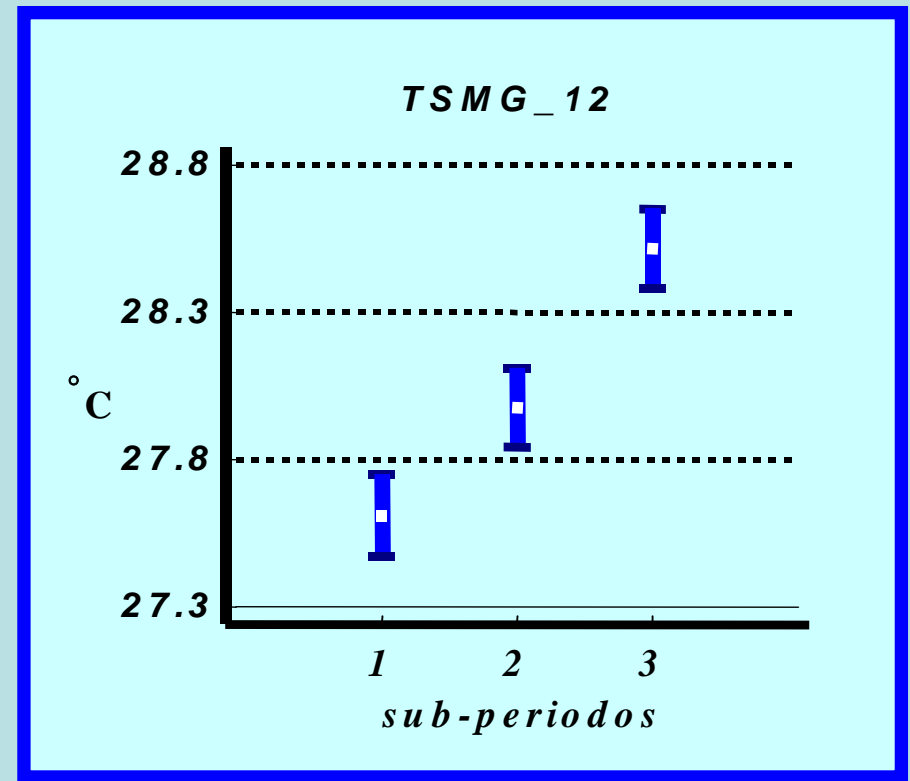


XAVIER & XAVIER (1995), XAVIER (2001)

G = Golfo da Guiné



(a) março



(b) dezembro

## Figura 5

**TSM na Área G (Golfo de Guiné) : Intervalos de Confiança para TSMG03 (março) e TSMG12 (dezembro) com respeito aos sub-períodos 1, 2, 3 em 1964-2005**



Por outro lado, na Figura 6, tem-se o gráfico de dispersão das observações no plano (com  $x$ =ano,  $y$ =tsmg12), para o mês de dezembro

Através dessa última figura denota-se claramente que na primeira metade do período 1964-2004, as temperaturas em dezembro na Área G foram todas inferiores a  $28,4^{\circ}\text{C}$ , ao passo que na segunda metade, a partir de 1985, de vinte observações onze (52,4%) ultrapassaram aquele valor.

E mais, de 1993 em diante, dez dentre doze médias de temperaturas (83,3%) foram superiores a  $28,4^{\circ}\text{C}$ .

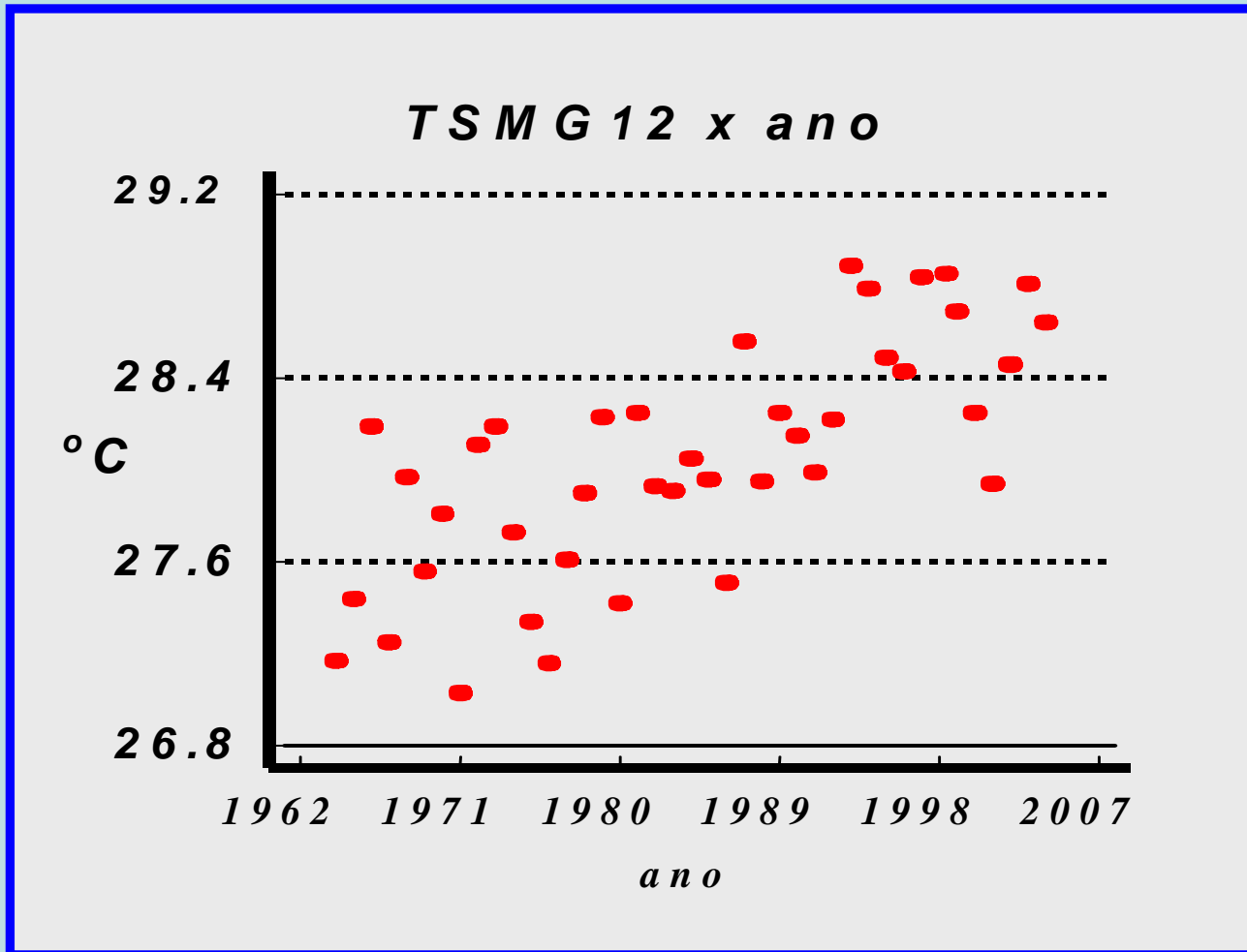


Figura 6

Varição da TSM/SST na Área G (Golfo da Guiné) em dezembro, entre 1964 e 2004.

#### **4. Regressão Múltipla “stepwise” (ou “passo a passo”) e Análise Discriminante**

**São úteis para modelos de previsão (ver: “TEMPO DE CHUVA”). Note-se que na Regressão Múltipla o modelo de regressão fornece valores previstos (calculados) que permitem “a posteriori” avaliar a qualidade da previsão pela comparação com os valores observados.**

**Já a Análise Discriminante permite um previsão por “classes” da chuva, ou de outra variável que venha a ser prevista. Note-se que classes “tercísticas” são mais pobres para a previsão, embora seja um procedimento usual. Classes “quintísticas” (ou ainda as classes cf. Pinkayan) permitem melhor discriminação.**

## 5. Análise de Classificação ("classificação automática", "rígida" ou "fuzzy")

**Uma aplicação possível refere-se à caracterização de estados do oceano Pacífico (eventos El Niño, La Niña e estados neutros) ou do Atlântico (dipolos).**

**Os resultados obtidos, comparando com a chuva, permitiriam estabelecer um modelo de previsão para a chuva, digamos no Ceará (e por extensão para o Nordeste setentrional). Em última instância, servindo para acoplar a um modelo discriminante de previsão.**

**Quanto à classificação "fuzzy" tem sido pouco usada. Mas oferece interesse no sentido de tratar situações em que "níveis de pertinência" sejam mais apropriados do que "probabilidades" (em especial se estas não possam ser estimadas).**

Versões “fuzzy” **dos** algoritmos de Forgy **e de** “Nuvens Dinâmicas” de Diday **foram ensaiados por Xavier & Xavier** [notas de aula para **Cursos no IAG/USP/ Mestrado e Doutorado em Meteorologia, e também na UFC/Mestrado em Recursos Hídricos; além de artigos em Congressos; também, objeto de dissertações de alunos do Prof. Airton F. S. Xavier, nos Mestrado de Matemática/UFC e de Computação Profissionalizante Aplicada-MPCOM/UECE/CEFET**].

**Além disso, há possibilidade de aplicar técnicas “fuzzy” para prognósticos climáticos.**

## 7. Análise de Séries Temporais.

**Técnicas de Análise de Séries Temporais têm sido amplamente empregadas para o estudo de variações e mudanças climáticas.**

**Contudo, faz-se necessário cautela ao interpretar determinados resultados, pois “periodicidades fantasmas” ou “espúrias” podem ser incorporados aos resultados. Em especial quando são aplicadas técnicas de Análise de Fourier, Espectrais e similares, sem o devido cuidado.**

**Isto decorre, principalmente, do fato de que para a chuva e outras variáveis geofísicas terrestres não existem propriamente ciclos, mas quase ciclos.**

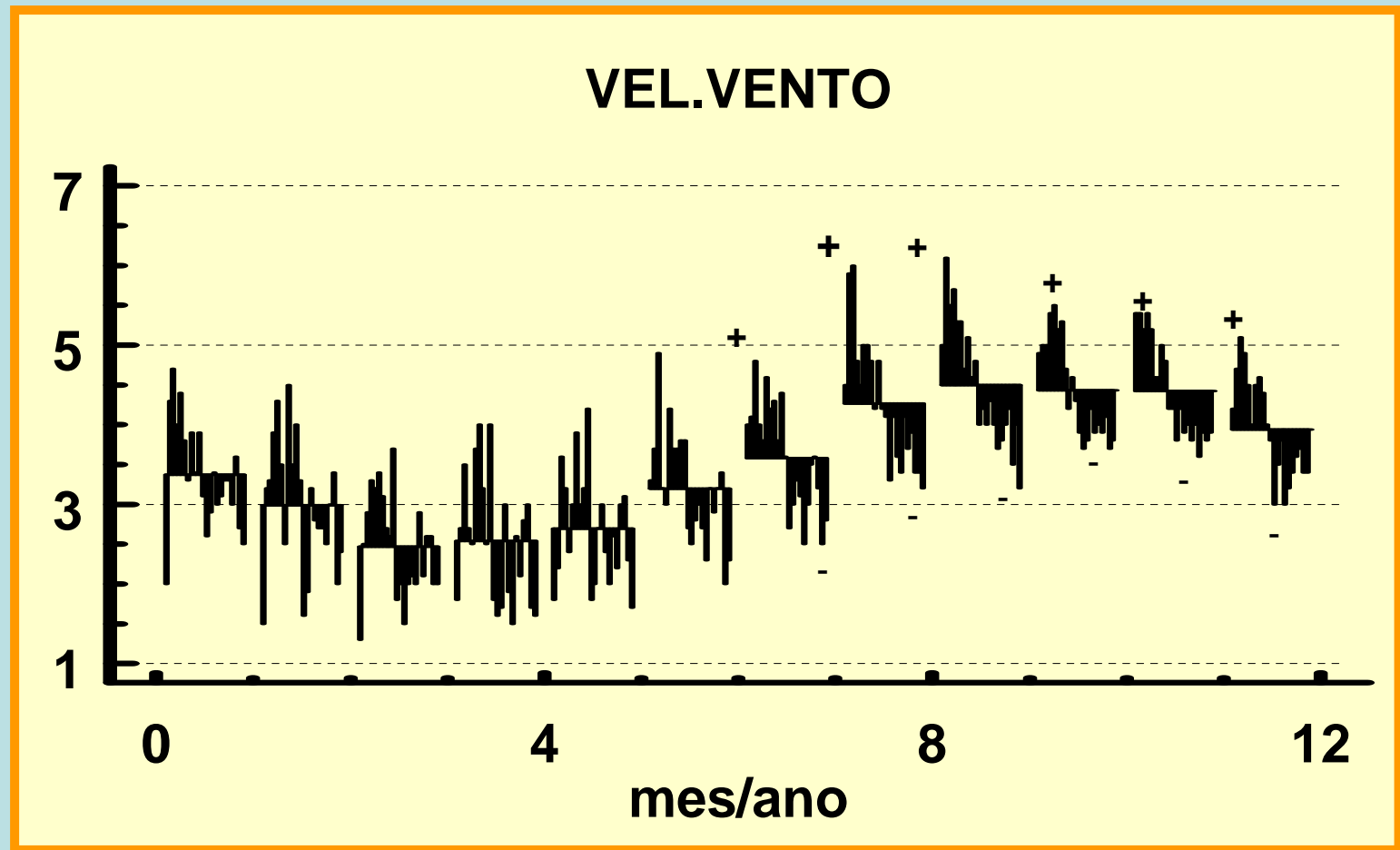
**Com a exceção de ciclos sazonais propriamente ditos ou aqueles que são comandados por fatores astronômicos, de natureza nitidamente cíclica.**

**Alguns métodos gráficos para representação de séries temporais são bastante úteis, em particular aqueles relacionados ao estudo de sub séries sazonais, que apesar de clássicos e constando de muitos “pacotes” estatísticos, são geralmente sub utilizados.**

**Veja-se, assim, a FIGURA 7 com a variabilidade sazonal da força do vento na estação meteorológica da FUNCEME, onde fica evidenciada a sua sazonalidade, a par de uma diminuição progressiva ao longo dos anos, em especial no 2º semestre, ou seja, semestre em que os ventos são mais fortes.**

**Esta diminuição não se deve decerto a qualquer alteração de ordem global ou regional, no vento, mas, simplesmente, pela criação de “falésias” artificiais constituídas pelo paredão de edificações muito elevadas, na orla marinha, originada da especulação imobiliária e de política errônea das autoridades municipais.**





**FIGURA 7** *Estação Meteorológica da FUNCEME – Fortaleza-CE 1974-95*  
*Evolução sazonal /anual : Médias Mensais do Vento*  
*fonte : Xavier & Xavier, 1996; Xavier, 2001*

## 8. Modelos Markovianos

Tais modelos são extremamente úteis. Em especial, para tratar questões ligadas à evolução de séries diárias de variáveis meteorológicas, climáticas, etc. Conforme indicam vários estudos existentes (a bibliografia é muito ampla, a começar pelos estudos pioneiros de Gabriel & Neuman), são apropriados na análise de ocorrências para “sequências de dias chuvosos” e, por complementação, “sequências de dias secos” (ou “veranicos”). No caso da análise de temperaturas diárias, analogamente, para avaliar as ocorrências de “ondas de calor”, “ondas de frio”, etc.

**Estas técnicas foram objeto de artigos (em reuniões científicas) e de material didático para disciplinas de pós graduação (Curso de Especialização ministrado na Paraíba, Mestrado em Recursos Hídricos / UFC, etc.).**

**Por outro lado, no pós-doutorado na Universidade de Lille-1, França, estudou-se uma família de Leis de Poisson Generalizadas, bem como, Processos Pontuais Quase-Poissonianos. As Leis Generalizadas de Poisson são úteis no estudo de eventos raros em Meteorologia, Climatologia, Hidrologia, Oceanografia, etc. Bem como, em Epidemiologia.**

## 9. “Data Mining” (ou “Mineração de Dados”)

**“Data Mining” não constitui propriamente uma técnica. Porém uma designação genérica para o emprego de recursos teóricos muito variados, como : Classificação Automática, Modelos Fatoriais, Redes Neurais, Modelos Genéticos, Inteligência Artificial, etc., a par de toda uma parafernália de “software” altamente especializado.**

**Na verdade, trata-se de prover recursos para retirar informações de enormes massas de dados. Contudo, há necessidade de ter em mente que tais recursos não operam milagres, donde poderem conduzir a resultados decepcionantes, se mal empregados.**

**Em particular, chamo atenção para procedimentos que sejam empregados como “caixas-pretas”, como é o caso das “redes neurais”, em determinadas situações, em especial para a pesquisa climática. Nada substitui o “namoro” com os dados que se deseja trabalhar.**

**O emprego das minhas “caixinhas” (“boxes”) (veja-se a FIGURA 4) assume, aliás, um significado específico. Pois desejo conhecer, exatamente como meus dados funcionam, de sorte que confio pouco em “variáveis artificiais”... É claro que estas podem se mostrar úteis, eventualmente. É o caso das ACP’s (ou EOF’s) que me ajudam a construir “óculos multi-dimensionais”...**

## 9-A. "Data Mining" antes do "Data Mining"...

**A este respeito, considero a reavaliação da cronologia de secas e anos excepcionalmente chuvosos, apresentada de forma integrada em [Xavier, 2001], Cap. 7, "Secas Ocorridas em 1964-2001 e seus Graus de Severidade", pp. 229-245, e Cap. 8, "Complementos à Cronologia das Secas do Ceará e Nordeste Setentrional", pp. 247-290.**

**Para o Ceará e Nordeste setentrional, considero em particular 1910-1979, comparando nossos resultados, objetivos, com cronologias oficiais e registros históricos. Omitimos, aqui, referências a outros autores, porém comparecem na bibliografia do referido livro e dos demais trabalhos nossos.**

**A rigor, com nossa investigação sobre a cronologia das secas dispõe-se, pela primeira vez, de levantamento em bases inteiramente objetivas ou científicas, além de completas, nos períodos considerados. Assim, quem queira documentar-se a respeito e dar continuidade a tal modalidade de investigação, obrigatoriamente teria de passar por esses trabalhos. Analisamos no Cap. 8, ainda, ocorrências de secas antes de 1910, com dados de *Fortaleza-CE* a partir de 1849, complementados com os de *Quixeramobim-CE* desde 1865 e *Mossoró-RN* desde 1889.**

**Para exemplificar nossas instâncias sobre cronologias de secas, apresenta-se um levantamento para o Ceará na década 1930-1939, conforme a Tabela 1 à página seguinte, extraída de [Xavier & Xavier, 1987] e de [Xavier, 2001].**

ESTATÍSTICAS para os PLUVIÔMETROS  
[ CEARÁ-BRASIL ] 1930-1939

	mu seco	seco	normal	chuvoso	mu chuvoso	total
1930 :	47 34.06%	53 38.41%	28 20.29%	9 6.52%	1 0.72%	138
1931 :	25 19.38%	74 57.36%	29 22.48%	1 0.78%	0 0.00%	129
1932 :	131 87.33%	13 8.67%	6 4.00%	0 0.00%	0 0.00%	150
1933 :	7 4.76%	32 21.77%	82 55.78%	25 17.01%	1 0.68%	147
1934 :	1 0.55%	1 0.55%	17 9.39%	67 37.02%	95 52.49%	181
1935 :	1 0.51%	2 1.01%	48 24.24%	81 40.91%	66 33.33%	198
1936 :	84 40.98%	66 32.20%	44 21.46%	11 5.37%	0 0.00%	205
1937 :	8 3.90%	57 27.80%	108 52.68%	31 15.12%	1 0.49%	205
1938 :	23 11.50%	60 30.00%	93 46.50%	18 9.00%	6 3.00%	200
1939 :	8 3.92%	45 22.06%	95 46.57%	43 21.08%	13 6.37%	204

fonte : Xavier, T. de Ma.B. S. & Xavier, A.F.S. (1987), Rev. Bras. de Engenharia – Caderno de Recursos Hídricos, Vol. 5, N.2, pp. 7-31

☐ = “m.seco”/”seco”  
 cercadura sem sombra = “chuvoso”/”m.chuvoso”  
 ondulado = “normal”; duplo = “chuvoso” ou “m.chuvoso”

Tabela 1



Nessa década, 1930 e 1931 haviam sido SECOS (em 1930, 34,2% dos postos na categoria “muito seco”, 38,4% na categoria “seco”, ao passo que em 1931 esses percentuais foram 19,4% e 57,4% , resp.). Segue-se, então, a “Grande Seca” de 1932 (com 87,3% dos postos na categoria “muito seco”). Em seguida, verifica-se que 1933 é predominantemente “normal”, enquanto 1934 e 1935 caem nas categorias “chuvoso” ou “muito chuvoso”. Já em 1936 retorna a situação de chuvas escassas, com o Ceará predominantemente “muito seco” ou “seco” (41,0% e 32,2% dos postos nessas categorias, resp.). Não obstante, esse ano não é referido como seca na maioria das cronologias de outras fontes consultadas; contudo, [Pompeu So, 1953] realmente anotara 1936 como “escassamente pluvioso”, mencionando duas outras fontes.

**Em nosso livro, são exibidas todas as tabelas para as décadas de 1910-1919 a 1970-1979. Dispomos de análogas tabelas para outros Estados nordestinos, como *Piauí, Rio Grande do Norte, Paraíba e Pernambuco*. Bem como, uma tabela resumindo a situação, para estes Estados.**

**Note-se que nessa modalidade de pesquisa não caberia utilizar atabalhoada ou irrefletidamente modernos procedimentos automatizados de “data mining”, como “classificação automática” (“rígida” ou “fuzzy”), “métodos fatoriais” ou a “caixa preta” das “redes neurais”, etc., vez que nos estudos climáticos exige-se o concurso em paralelo de procedimentos peculiares, além da preparação preliminar dos dados, conforme cada propósito ou objetivo específico. Evidentemente, todos aqueles métodos automatizados serão sempre bem-vindos, mas com a ressalva precedente.**

## 10. Modelos de Extremos

**Remetemos, neste contexto, a nosso último livro, Xavier, Xavier & Alves (2007), “Quantis e Eventos Extremos: Aplicações a Ciências da Terra e Ambientais.**

**Evidentemente, mudanças climáticas costumam estar associadas a ocorrências de extremos.**