

Investigação e Aplicação de Ferramentas para Análise de Séries Temporais em Ambiente de *Software* Livre

¹Barbosa, E.B.M., ²Rosa, R.R., ²Vijaykumar, N.L.

¹Centro de Previsão de Tempo e Estudos Climáticos – CPTEC
Instituto Nacional de Pesquisas Espaciais – INPE
Rod. Presidente Dutra, km. 39 – 12630-000, Cachoeira Paulista, SP
E-mail: eduardo@cptec.inpe.br

²Laboratório Associado de Computação e Matemática Aplicada – LAC
Instituto Nacional de Pesquisas Espaciais – INPE
Av. dos Astronautas, 1758 – 12227-010, São José dos Campos, SP
E-mail: reinaldo@lac.inpe.br, vijay@lac.inpe.br

Resumo

A análise de séries temporais não-estacionárias em geral, comuns na natureza, exige a utilização de abordagens robustas da matemática. Duas novas técnicas estatísticas computacionais que tem introduzido novos paradigmas à teoria da análise de séries temporais e merecem destaque são a transformada em ondeletas e a análise de padrões-gradiente. O presente trabalho tem por objetivo investigar as características dessas técnicas, em um ambiente comum de programação seguindo o conceito de software científico livre como, por exemplo, o Scilab. As técnicas foram aplicadas em séries temporais de temperatura do ar medidas em reservatórios hidrelétricos do Brasil. Análises preliminares em periodogramas e no espectro global de ondeleta permitem a identificação de padrões de variabilidade, possivelmente, associados ao ciclo diurno da temperatura. A análise de padrões-gradiente apresentam os momentos estatísticos e o valor médio do parâmetro F_A muito semelhantes. Os resultados preliminares mostraram-se coerentes quando comparados com as implementações originais, desenvolvidas em ambientes distintos de programação como, por exemplo, Matlab e IDL.

Palavras-chave: *software científico livre, scilab, análise de padrões-gradiente, transformada em ondeletas*

1. Introdução

A partir do avanço tecnológico ocorrido nas últimas décadas como, por exemplo, o desenvolvimento de novos materiais e instrumentos de medida e dispositivos computacionais avançados, as séries temporais têm sido analisadas dentro de um novo paradigma. Devido as altas resoluções temporal e espacial, as atenções voltaram-se para a análise e modelagem da variabilidade complexa das amplitudes que constituem a amostra ou sub-amostragens das séries.

Um exemplo clássico de dados que apresentam variabilidade complexa são aqueles relacionados à observação de variáveis atmosféricas e do meio ambiente.

A análise de séries temporais (ST) não-estacionárias exige abordagens matemáticas robustas, com a finalidade de compreender os processos físicos que atuam no sistema subjacente. A partir da literatura, pode-se observar que a utilização de novas técnicas estatísticas computacionais, continuamente investigadas e desenvolvidas nos últimos anos, tem introduzido novos paradigmas à teoria da análise de ST com variabilidade complexa. Duas destas novas técnicas que merecem destaque são a Transformada em Ondeletas (*Wavelets Transform*) e a Análise de Padrões-Gradiente (*Gradient Pattern Analysis*).

O presente trabalho tem por objetivo investigar as características da transformada em ondeletas e da análise de padrões-gradiente quando aplicadas,

em ambiente de *software* científico livre, na análise e interpretação de ST obtidas a partir de dados de ciências atmosféricas.

2. Dados e Metodologia

2.1 Dados do Sistema Integrado de Monitoramento Ambiental

Os reservatórios destinados a geração hidrelétrica têm sido apontados, pela comunidade científica, como potenciais causadores de mudanças climáticas devido a emissão de gases de efeito estufa na atmosfera [12].

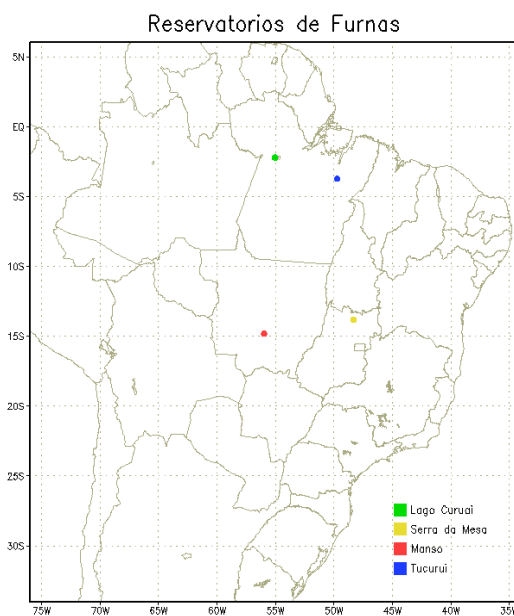


Figura 1 – Localização dos Reservatórios selecionados para Estudo

O Sistema Integrado de Monitoramento Ambiental (SIMA), projeto desenvolvido em conjunto entre a Universidade do Vale do Paraíba e o INPE, possui um conjunto de plataformas de coleta de dados (PCD) instaladas em reservatórios hidrelétricos do Brasil (Figura 1). A finalidade do projeto é monitorar diferentes variáveis hidrometeorológicas, avaliar suas interações com o ambiente e determinar se os reservatórios hidrelétricos podem ser considerados fonte limpa de geração de energia. Os instrumentos do SIMA realizam medições, em frequência horária, desde janeiro de 2004 para as seguintes variáveis: clorofila, PH, concentração de O₂, nitrato, amônia, temperatura da água, pressão atmosférica, umidade relativa, temperatura do ar, velocidade e direção do vento, radiação solar e fluxo e direção da água.

Os reservatórios escolhidos para realização das investigações foram: Serra da Mesa e Manso, localizados na região Centro-Oeste do Brasil, e Lago Curuai e Tucuruí, localizados na região Norte do país.

Os conjuntos de dados utilizados durante a realização deste trabalho correspondem à duas ST de temperatura do ar, constituídas por 604 pontos, medidas nos reservatórios de Serra da Mesa e Manso, com resolução de 4 horas (Figuras 2 e 3), durante a estação de outono (março, abril e maio) de 2004.

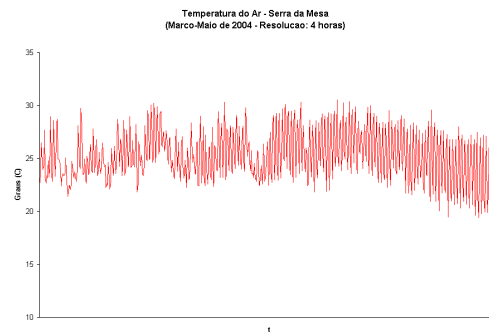


Figura 2 – Temperatura do ar (4 horas) entre março e maio de 2004 (Serra da Mesa)

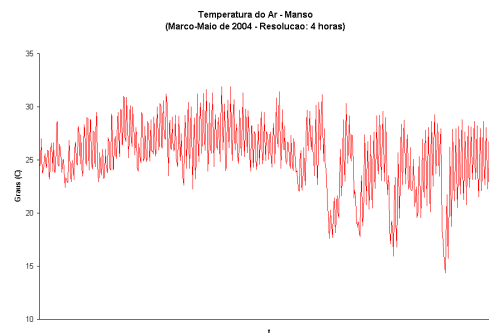


Figura 3 – Temperatura do ar (4 horas) entre março e maio de 2004 (Manso)

2.2 Transformada em Ondeletas

A transformada em ondeletas pode ser aplicada na análise de ST obtidas a partir de qualquer sistema físico e tornou-se uma ferramenta comum, uma vez que possibilita o entendimento das variações locais de uma série. A idéia central da análise de ondeletas, no contexto de análises de sinais, consiste em decompor uma ST em diferentes níveis de resolução tempo-frequência e, então, determinar as componentes de variabilidade dominantes [4].

Uma família generalizada de ondeletas pode ser descrita da seguinte maneira,

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right). \quad (1)$$

O parâmetro a se refere à escala, b é o parâmetro de translação ou localização da função ondeleta ($\psi_{a,b}$), geralmente, chamada de função ondeleta mãe. A variação do parâmetro a tem o efeito de dilatação ($a > 1$) e de contração ($a < 1$) na função ondeleta mãe. Com isto, podem-se analisar, respectivamente, aspectos de longos e curtos períodos na série. À medida que b varia, o sinal é analisado localmente em torno desse ponto.

Funções ondeletas são classificadas como discreta ou contínua. A função ondeleta discreta mais simples é a de Haar [6] (2), utilizada na decomposição e filtragem de qualquer série temporal.

$$\psi(t) = \begin{cases} 1, & 0 \leq t < 1/2 \\ -1, & 1/2 \leq t < 1 \\ 0, & \text{caso contrário} \end{cases} \quad (2)$$

A onda produzida pela função de Haar é descontínua, por isso considera-se pobre o resultado correspondente à localização de frequência produzida, sendo esta função, em geral, empregada no entendimento básico da análise de ondeletas [5].

Em funções ondeletas contínuas, os parâmetros de escala e localização assumem valores contínuos. Nesse sentido, muitas das informações apresentadas referentes à escala ou tempos próximos são redundantes e isso faz com que este tipo de função possua um alto custo computacional. Uma das funções contínuas mais comum é a de Morlet [7] (3), a qual consiste de uma onda plana modulada por uma Gaussiana.

$$\psi(t) = \pi^{-1/4} (e^{i\omega t} - e^{-\omega^2/2}) e^{-t^2/2} \quad (3)$$

A ondeleta de Morlet é largamente utilizada na análise das frequências pertencentes ao sinal e na identificação de relações entre tempo e frequência [4]. Estas funções são utilizadas na análise de ST, onde suavidade e variações contínuas na amplitude da ondeleta são esperadas [11].

A ondeleta mãe conhecida como Chapéu Mexicano (4) é a derivada de segunda ordem da função de densidade de probabilidade Gaussiana.

$$\psi(t) = \frac{2}{\sqrt{3\pi^{1/4}}} e^{-t^2/2} (1-t^2). \quad (4)$$

Um dos fatores críticos na análise de ondeletas é a escolha arbitrária da função mãe, $\psi_0(t)$. Ainda

não existe na literatura um consenso de qual a melhor função ondeleta a ser utilizada. Comumente se aceita que a função ondeleta a ser utilizada possua um formato característico próximo das características encontradas na série temporal [5,11].

2.3 Análise de Padrões-Gradiente

Uma maneira eficiente para caracterizar a variabilidade de padrões em séries curtas e não-estacionárias, é a utilização do operador extraído da técnica computacional análise de padrões-gradiente (GPA) [9]. Em linhas gerais, o GPA realiza investigações numéricas de padrões espaço-temporal complexos a partir do operador de fragmentação assimétrica da amplitude (AAF).

Ao contrário dos métodos estatísticos convencionais, o GPA depende unicamente das propriedades simétricas locais presentes no padrão gradiente do sinal. Por isso, apresenta a relevante propriedade de ser menos dependente do comprimento da ST, quando comparado às técnicas convencionais [3].

Para aplicar o GPA em uma série unidimensional $X(t)$ é necessário convertê-la em seqüências de matrizes 3×3 . Dessa maneira, altas resoluções temporais podem ser obtidas com o agrupamento de dados em matrizes elementares, chamadas de padrão de variabilidade.

O campo gradiente, composto por nove vetores, é utilizado com a finalidade de caracterizar diferentes máximos e mínimos. Os vetores são distribuídos espacialmente no campo de modo que os pares com a mesma magnitude sejam removidos. Após a remoção, o número de vetores é denotado por L . Toda quebra de simetria pode ser quantificada pela conexão de pontos médios dos vetores L a partir de linhas I , gerando o campo de triangulação de Delaunay [1]. A medida de assimetria é a diferença $I-L$ normalizada por L , então chamada de AAF e abreviada pelo parâmetro $F_A \equiv (I-L)/L$ com $I \geq L > 0$ [10].

Deve-se considerar que para uma dada ST $X(t)$ composta por N pontos é possível calcular o inteiro $N/9$. Este número representa a quantidade de F_A , e pertencerá ao seguinte domínio {0.778; 0.889; 1.000; 1.111; 1.222; 1.333}. Isto significa que o padrão de variabilidade mais complexo pode ser detectado a partir de 54 pontos de uma ST. Entretanto, para caracterizar a robustez de um regime, pela média de $\langle F_A \rangle$, é necessário pelo menos dois intervalos de 54 pontos, então a técnica começa a ser consistente a partir de $N \geq 108$ [2].

3. Resultados Preliminares

O conceito de *software* livre, introduzido pela *Free Software Foundation* (FSF) na década de oitenta, tem apresentado uma crescente popularidade nos últimos anos. Sua idéia central consiste em garantir a qualquer usuário a liberdade para cópia, utilização e modificação dos programas conforme suas necessidades. Atualmente, projetos derivados desse conceito são definidos sob o título *Open Source*, cuja finalidade é reunir diferentes movimentos de *software* livre.

O Scilab, criado no início dos anos noventa por pesquisadores do *Institut de Recherche en Informatique et en Automatique* (INRIA) e *École Nationale des Ponts et Chaussées* (ENPC), é definido como um pacote de *software* científico destinado a resolução de problemas numéricos [8]. Com um poderoso ambiente de programação pode ser utilizado em aplicações científicas de diferentes áreas de conhecimento. Atualmente desenvolvido e mantido pelo consórcio Scilab, o *software* pode ser obtido gratuitamente a partir da Internet, seguindo o conceito de *software* livre.

Para realização deste trabalho, optou-se pela adaptação das técnicas de transformada em ondeletas e GPA, em Matlab e IDL, respectivamente, ao ambiente de programação do Scilab. O conjunto de programas fornecido pela biblioteca de *softwares Wavelab* auxiliou a adaptação da transformada em ondeletas ao ambiente Scilab. O GPA foi adaptado ao ambiente de *software* científico livre com o auxílio de um extenso conjunto de programas desenvolvidos pelo LAC.

Após as adaptações algorítmicas em um ambiente comum de programação, as duas técnicas foram aplicadas nos conjuntos de dados e os resultados comparados com aqueles gerados pelos programas originais.

3.1 Resultados da Transformada em Ondeletas

Análises foram realizadas em periodogramas (Figuras 4 e 6), gerados pelo pacote *Wavelab* em *software* Matlab, visando conhecer as relações existentes entre as componentes de diferentes frequências em função da escala temporal do sinal. Com adaptações nos programas foi possível obter, em ambientes Matlab e Scilab, o espectro global de ondeletas (*global wavelet spectrum*), a partir do cálculo das variâncias em cada período da série. Em geral, esses gráficos auxiliam na identificação de quais os períodos associados à

maior energia [4, 5] e seus resultados são apresentados nas Figuras 5 e 7.

O gráfico abaixo (Figura 4) representa o periodograma obtido a partir da transformada em ondeletas onde, o eixo *x* é a escala temporal, em dias, e o eixo *y* são os períodos presentes na série. A linha branca contínua no periodograma é chamada cone de influência e os períodos fora deste cone devem ser negligenciados por não possuírem confiança estatística. A barra de cores abaixo do gráfico representa a energia associada a cada período da série.

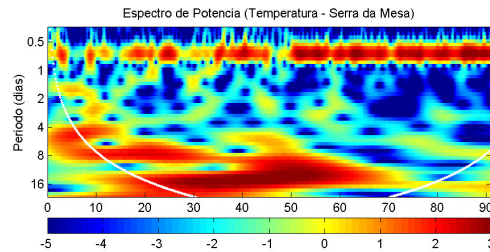


Figura 4 – Periodograma (reservatório de Serra da Mesa)

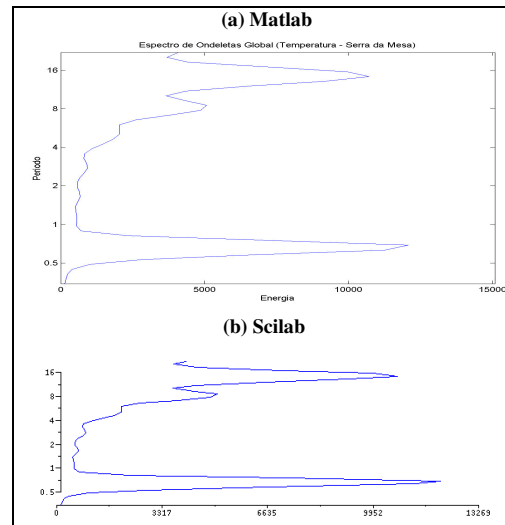


Figura 5 – Gráficos de GWS em (a) Matlab e (b) Scilab para o reservatório de Serra da Mesa

A partir das análises no periodograma, referente conjunto de dados de Serra da Mesa (Figura 4), e com auxílio do espectro global de ondeletas (GWS) (Figura 5) pode-se identificar durante toda a ST um sinal persistente em escala de aproximadamente 12 horas, possivelmente, associado ao período diurno de dados, quando a temperatura, em geral, é mais alta se comparada ao período noturno. Além disso, o GWS identifica um ciclo em escalas de 4 a 16 dias durante os meses de outono, principalmente, durante o mês de abril em que os valores de

temperatura mostram-se mais intensos sobre o reservatório.

As Figuras 6 e 7 representam os resultados obtidos para o conjunto de dados do reservatório de Manso. A partir das análises, podem-se identificar algumas características semelhantes àquelas do conjunto de dados do reservatório de Serra da Mesa. Ou seja, observou-se durante todo o período de dados um sinal persistente em escala de aproximadamente 12 horas, entretanto, menos intenso se comparado com Serra da Mesa. O GWS (Figura 7) também indica um ciclo em escalas de 4 a 16 dias com intensidade semelhante a Serra da Mesa, no entanto, deslocado entre o final de abril e durante o mês de maio.

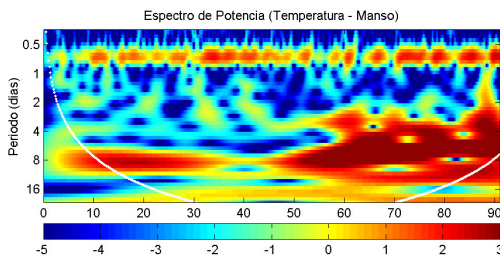


Figura 6 – Periodograma (reservatório de Manso)

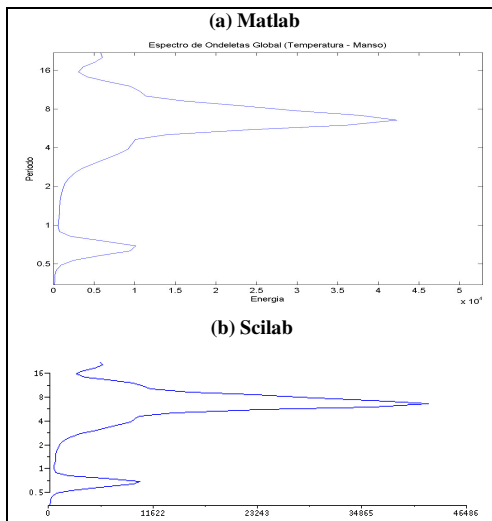


Figura 7 – Gráficos de GWS em (a)Matlab e (b)Scilab para o reservatório de Manso

4.2 Resultados da Análise de Padrões-Gradiente

A técnica do GPA foi aplicada nas ST de temperatura, medidas nos reservatórios de Serra da Mesa e Manso, com a finalidade de comparar os resultados das implementações em diferentes ambientes de programação, isto é, IDL (original)

e Scilab. Os resultados são apresentados nas Tabelas 1 e 2.

Tabela 1 – Estatísticas do GPA para o reservatório de Serra da Mesa

	Média	Variância	Curtose	Assimetria	F_A
IDL	1.089	0.039	-0.262	-1.303	1.111
Scilab	1.087	0.040	-0.304	-1.291	1.111

Tabela 2 – Estatísticas do GPA para o reservatório de Manso

	Média	Variância	Curtose	Assimetria	F_A
IDL	1.112	0.022	-0.507	-0.613	1.111
Scilab	1.112	0.022	-0.555	-0.639	1.111

A partir dos resultados obtidos, pode-se observar que os momentos estatísticos, bem como, os valores de F_A são muito semelhantes para as duas implementações em análise, bem como, para os dois conjuntos de dados em estudo.

4. Conclusões

Os resultados obtidos a partir da aplicação das técnicas de transformada de ondeletas e análise de padrões-gradiente, em ambiente comum de programação seguindo o conceito de *software* científico livre, nos dados de temperatura do ar do SIMA, mostraram-se bem coerentes se comparados àqueles obtidos a partir dos programas originais.

Os gráficos de GWS em Scilab são muito semelhantes aos do Matlab e conseguiram identificar as energias associadas ao ciclo diurno da temperatura dos reservatórios, que variam em escalas de 12 horas até 16 dias. Os resultados obtidos com a aplicação do GPA nos conjuntos de dados, apresentaram os momentos estatísticos e valores médios de F_A muito semelhantes para os dois ambientes comparados, isto é, Scilab e IDL.

Algumas pesquisas recentes na literatura atual, indicam um extenso conjunto de trabalhos de investigações, bem como, para aplicações das técnicas de transformadas de ondeletas e análise de padrões-gradiente em diversos contextos. Entretanto, nenhuma iniciativa foi observada com a finalidade de integrá-las numa única ferramenta em ambiente científico público.

Diante desse contexto, o desenvolvimento desse trabalho pretende contribuir para a compilação e integração de diferentes técnicas estatísticas computacionais em ambientes de programação integrados, seguindo o conceito de *software* científico livre como, por exemplo, Scilab e Octave.

5. Referências Bibliográficas

- [1] Adelsbrunner, H. GEOMETRIC ALGORITHMS. Hand Book of Convex Geometry A, Elsevier, 1993.
- [2] Assireu, A.T., Rosa, R.R., Lorenzetti, J.A., Vijaykumar, N.L., Rempel, E.L., Ramos, F. M., Sá, L.D. Abreu, Bolzan, M.J.A., Zanandrea, A. GRADIENT PATTERN ANALYSIS OF SHORT NONSTATIONARY TIME SERIES: AN APPLICATION TO LAGRANGIAN DATA FROM SATELLITE TRACKED DRIFTERS. *Physica D*, v. 168-169, 397-403 pp., 2002.
- [3] Assireu, A.T., Lorenzetti, J.A., Novo, E.M.L.M., Stech, J.L., Braga, C.Z.F., Lima, I.B.T. APLICAÇÃO DO OPERADOR DE FRAGMENTAÇÃO ASSIMÉTRICA (FA) NA CARACTERIZAÇÃO DE CONTROLES GEOMORFOLÓGICOS EM RESERVATÓRIOS HIDROELÉTRICOS. *Revista Brasileira de Geociências*, v.32, n.4., 2004.
- [4] Barbosa, E.B.M., Rosa, R.R., Vijaykumar, N.L., Bolzan, M.J.A., Tomasella, J. CARACTERIZAÇÃO POR ONDELETAS DE PROCESSOS FÍSICOS NÃO-LINEARES NA MICRO-BACIA AMAZÔNICA. *Anais do XXVIII Congresso Nacional de Matemática Aplicada e Computacional*, CD-ROM, 2005.
- [5] Bolzan, M.J.A. ANÁLISE DA TRANSFORMADA EM ONDELETAS APLICADA O SINAL GEOFÍSICO. *Revista Brasileira de Ensino em Física*, v.26, n.1, 37-41 pp., 2004.
- [6] Haar, A. ZUR THEORIE DER ORTHOGONALEN FUNKTIONEN-SYSTEME. *Math. Ann.*, v.69, 331-371 pp., 1910.
- [7] Morlet, J. SAMPLING THEORY AND WAVE PROPAGATION. NATO ASI Series, Issues in Acoustic Signal/Image Processing and Recognition, n.1, 233-261 pp., 1983.
- [8] Pires, P.S.M., Rogers, D.A. FREE/OPEN SOURCE SOFTWARE: AN ALTERNATIVE FOR ENGINEERING STUDENTS. 32ND ASEE/IEEE Frontiers in Education Conference, 3TG, 2002.
- [9] Rosa, R.R., Sharma, A.S., Valdivia, J.A. CHARACTERIZATION OF LOCALIZED TURBULENCE IN PLASMA EXTENDED SYSTEM, *Physica A* 257 509-514 pp., 1998.
- [10] Rosa, R.R., Sharma, A.S., Valdivia, J.A. CHARACTERIZATION OF ASYMMETRIC FRAGMENTATION PATTERNS IN SPATIALLY EXTENDED SYSTEMS. *Int. J. Mod. Phys. C* 10 (1), 147-163 pp., 1999.
- [11] Torrence, C., Compo, G.P. A PRACTICAL GUIDE TO WAVELET ANALYSIS. *Bulletin of The American Meteorological Society*, v.79, n.1, 61-78 pp., 1998.
- [12] Novo, E.M.L.M., Tundisi, J.G. CONTRIBUTION OF REMOTE SENSING TECHNIQUE TO ASSESSMENT OF METHANE EMISSION FROM LARGE TROPICAL RESERVOIRS. *Remote Sensing Review*, 10, 143-153 pp, 1994.