

# Modelagem de dados de sensoriamento remoto por Grades Numéricas Generalizadas: aplicação em sistemas limnológicos

Thalita Biazuz Veronese  
Ramon Morais de Freitas  
Murilo da Silva Dantas  
Nandamudi L. Vijaykumar  
Reinaldo Roberto Rosa

Instituto Nacional de Pesquisas Espaciais - INPE  
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil  
thalitabv@gmail.com, aeroramon@gmail.com, {murilo, vijay, reinaldo}@lac.inpe.br

**Abstract.** Databases in remote sensing applications are often composed of different kinds of information collected from distinct high resolution measurement devices. Like in all areas of natural science, analysis of these multivariate data has become a fundamental task for remote sensing researchers, increasing the demand for development of expert methods able to treat such systems characterized by large variability and extension. In this paper, we apply a new formalism to represent multi-source remotely sensed data systems: a mathematical multi-source data modeling based on the concept of generalized numerical lattices (GNL). We define a generalized numerical lattice  $\mathcal{L}(\alpha, \ell_k, M_i)$  as a function of coefficients representing the structural  $(\alpha, \ell_k)$  and physical  $(M_i)$  properties, reducing any multivariate database in a closed set of classified time series in  $N$  dimensions. In a preliminary investigation, we have aggregated, as a GNL data modeling, a collection of CBERS remote sensing images and SIMA time series, both obtained from FURNAS hydroelectric reservoirs. Our first goal is the implementation of a computational interface that will allow users to enter a data set composed of files independently stored and return a well structured formal GNL set aggregating all information needed to characterize the system in a unified way.

**Palavras-chave:** data mining, information systems, time series, signal analysis; mineração de dados, sistemas de informação, séries temporais, análise de sinais.

## 1. Introdução

Atualmente, devido à crescente demanda e disponibilidade de dados e à necessidade de convertê-los em informações de efetiva aplicabilidade, diversos trabalhos têm sido desenvolvidos com o objetivo de integrar dados de sensoriamento remoto advindos de múltiplas fontes (Solberg et al., 1994; Petit e Lambin, 2001; Bouman, 1992). Duas principais razões podem ser apontadas como motivadoras desta situação. A primeira, de aspecto primordialmente computacional, consiste na dificuldade em armazenar e processar de forma organizada e fisicamente reduzida os conjuntos de dados que se acumulam ao longo das coletas, medidas e cálculos envolvidos na tarefa de análise de um sistema real. Neste aspecto, muitos progressos vêm sendo alcançados através de pesquisas em mineração de dados e processamento distribuído. Uma segunda motivação consiste na agregação de dados coletados de múltiplas fontes na tentativa de atingir resultados melhores e mais confiáveis, uma vez que informações obtidas de forma diferente podem ser complementares. Neste caso, nota-se uma grande limitação: na maioria das vezes, os dados a serem agregados devem ser do mesmo tipo. Assim, podem-se combinar atributos de imagens obtidas por satélite e por radar, ou calcular parâmetros a partir de séries de temperatura, umidade e pH, mas não se associam explicitamente informações obtidas por dados de diferentes características estruturais como dimensionalidade, por exemplo.

Neste trabalho, propomos a integração de imagens CBERS e séries temporais SIMA, todas obtidas dos reservatórios da hidroelétrica de FURNAS, reduzindo toda a informação caracterizadora do sistema através da modelagem de uma representação formal denominada grade numérica generalizada (GNL) (Veronese et al., 2008). Em sensoriamento remoto, assim como em muitas áreas da Ciência, é possível coletar dados de um sistema natural de diversas formas, a fim de extrair informações e realizar diferentes tipos de análise sobre ele. Frequentemente isso gera uma grande quantidade de conjuntos de dados, armazenados independentemente e compostos por informação espaço-temporal em uma, duas e três dimensões, representando muitas medidas possíveis tomadas sobre o mesmo sistema. Através do Projeto Balanço de Carbono, implementado pela hidroelétrica de Furnas, são gerados diversos arquivos de dados provenientes de observações obtidas por diferentes instrumentos de alta resolução. Muitos desses dados representam processos físicos distintos possivelmente relacionados. Utilizando a modelagem GNL, é possível representá-los como sistemas de dados organizados de forma coerente.

A próxima seção descreve os dados a serem utilizados e a metodologia de grades numéricas generalizadas para modelagem dos mesmos. Os resultados de uma investigação preliminar e as considerações finais são apresentados, respectivamente, nas Seções 3 e 4.

## 2. Dados e Metodologia

Desenvolvido a partir de uma parceria entre a Universidade do Vale do Paraíba (UNIVAP) e o Instituto Nacional de Pesquisas Espaciais (INPE), o Sistema para Monitoramento Ambiental (SIMA) consiste em um conjunto hardware/software projetado para aquisição de dados e monitoramento em tempo real de sistemas hidrológicos (Stevenson et al., 1993, Stech et al., 2006). Formado por uma bóia ancorada à qual são afixados sensores, sistemas de armazenamento de dados, bateria e antena de transmissão, o sistema coleta dados em intervalos de tempo pré-programado e os transmite via satélite em tempo quase-real para qualquer usuário em um raio de 2500 km do ponto de aquisição.

Neste trabalho, consideramos as seguintes séries, representando 21 parâmetros coletados pelo SIMA no período de fevereiro de 2006 a fevereiro de 2007: direção, intensidade e velocidades zonal e meridional do vento; temperatura da água; temperatura e umidade relativa do ar; pressão atmosférica; radiação incidente e refletida; velocidade zonal e meridional da corrente; temperatura da sonda; condutividade; concentração de oxigênio dissolvido,  $NH_4^+$  e  $NO_3^-$ ; pH; turbidez; e clorofila. As Figuras 1 (a) e (b) ilustram, respectivamente, as séries correspondentes às 2390 medidas de temperatura do ar e direção do vento neste período.

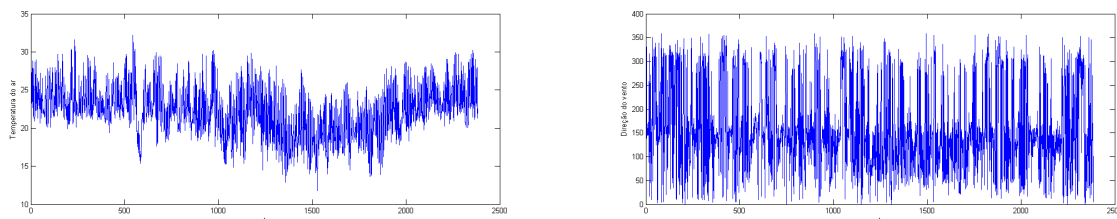


Figura 1. Série temporal representando as medidas de temperatura do ar ( $^{\circ}C$ ) e direção do vento ( $^{\circ}NV$ ) no período de fevereiro de 2006 a fevereiro de 2007.

Os dados bidimensionais, tomados aqui como exemplo, são compostos por seis imagens CBERS obtidas de 2005 a 2008, redimensionadas em  $64 \times 64$  pixels, ilustradas na Figura 2.

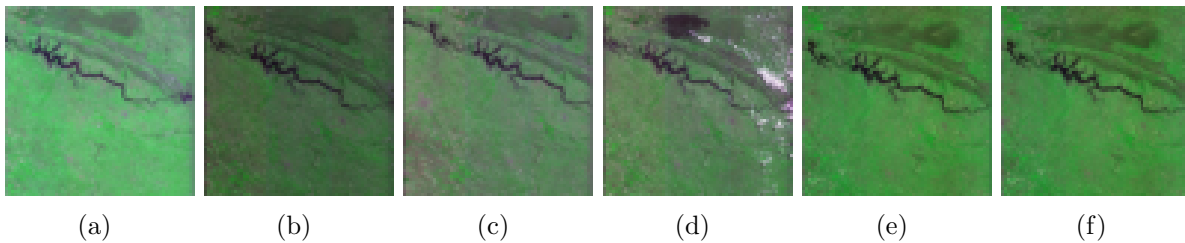


Figura 2. Imagens CBERS obtidas em 2005 (a), 2006 (b, c), 2007 (d) e 2008 (e, f).

A capacidade de tratar dados heterogêneos, como no caso daqueles obtidos nos reservatórios de Furnas, é fundamental para a análise e a representação apropriadas da informação e seus níveis de abstração (Datcu e Seidel, 1999). A metodologia proposta tem como principal objetivo uniformizar a informação através do formalismo proporcionado pelas grades numéricas generalizadas. Para a modelagem GNL do conjunto de dados SIMA/CBERS, primeiramente obtêm-se as medidas desejadas na forma de variáveis sobre o sistema observável; a informação contida no sistema de dados é então parametrizada e analisada para gerar a representação em grades numéricas. O modelo final é obtido através da união de todas as grades numéricas – uma para cada tipo de dado – em uma única representação, como ilustrado no diagrama abaixo.

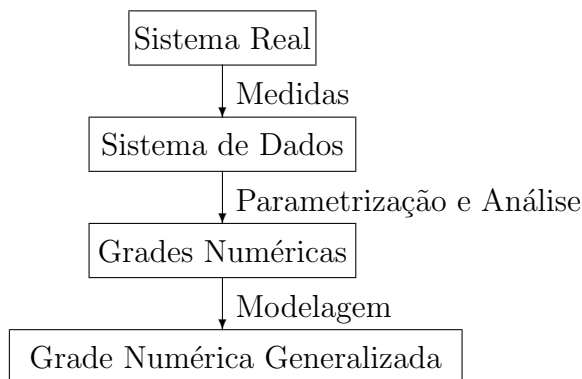


Figura 3. Diagrama do processo de modelagem GNL.

Uma GNL é definida como uma função de três coeficientes, representando as propriedades estruturais e físicas do sistema, como mostra a Equação 1:

$$\mathcal{L} = f(\alpha, \ell_\kappa, M_i) \quad , \quad (1)$$

onde  $\alpha$  se refere ao grau variacional degree, que é a quantidade de variáveis de estado nos domínios fundamentais tempo e espaço tridimensional;  $\ell_\kappa$  indica os coeficientes de extensão, dados pela quantidade de medidas discretas em cada domínio usual; e  $M_i$  é o conjunto de propriedades físicas caracterizando o sistema. O grau variacional indica quantos tipos de informação estão disponíveis no sistema de dados. Uma medida  $f(t)$  obtida sobre o

tempo  $t$  tem, por exemplo, duas observáveis, o que corresponde ao valor de  $\alpha = 2$ . Quando a informação é medida em uma, duas ou três dimensões, o grau variacional passa a ser, respectivamente, 3, 4 or 5. Se uma nova variável for incorporada, gerando um observável que é função do tempo, do espaço 3-D e desta nova variável,  $\alpha$  passa a ter valor 6, e assim sucessivamente, como na Tabela 2.

Tabela 2. Grau variacional

$\alpha$	Variáveis
2	$U_1 = t; U_2 = f(t)$
3	$U_1 = t; U_2 = x; U_3 = f(x, t)$
4	$U_1 = t; U_2 = x; U_3 = y; U_4 = f(x, y, t)$
5	$U_1 = t; U_2 = x; U_3 = y; U_4 = z; U_5 = f(x, y, z, t)$
6	$U_1 = t; U_2 = x; U_3 = y; U_4 = z; U_5 = g(U_j, j \leq \alpha - x); U_6 = f(x, y, z, U_5, t)$
$\vdots$	$\vdots$

O coeficiente de extensão  $\ell_\kappa$  refere-se ao tamanho do conjunto de dados em cada variável medida. Assim,  $\ell_0$  denota o número de pontos  $N$  que compõem o vetor  $f(t)$ ;  $\ell_1$  corresponde ao tamanho da dimensão  $x$  no espaço;  $\ell_2$  é o tamanho da variável na dimensão seguinte, e assim sucessivamente. Para o elemento  $M_i$ , podemos calcular pelo menos três propriedades físicas relevantes a uma estrutura espaço-temporal. A primeira delas é o fator de acoplamento, que quantifica a intensidade do acoplamento entre medidas da grade em ambos os domínios: espaço e tempo (Mecke e Stoyan, 2000). O coeficiente de correlação pode ser apontado como um exemplo desta propriedade. Como um segundo tipo de propriedade a ser analisada, podemos citar as escalas de flutuação estatística, como os coeficientes de autocorrelação e leis de potência, para identificar escalas nas quais a grade apresenta correlação mais forte. Uma terceira propriedade é dada por medidas morfométricas, como funcionais de Minkowski e a Análise de Padrões Gradientes (GPA) (Rosa et al., 2007, 2008), caracterizando os aspectos estruturais do padrão observado no tempo e/ou espaço.

### 3. Resultados Preliminares

A representação gerada pela modelagem GNL contém todas as grades numéricas relativas a um sistema particular, distribuídas em linhas e colunas dispostas em ordem crescente dos valores de  $\alpha$  e  $\ell_0$ , respectivamente, como ilustrado no exemplo da Figura 3. A notação  $\mathcal{L}^{\alpha, N, L_x, L_y, L_z}$  indica os valores de  $\alpha$  e  $\ell_i, i = 0, 1, \dots, \alpha - 2$  em cada grade numérica. As quantidades dentro dos parênteses mostram as propriedades físicas calculadas sobre os dados.

$\mathcal{L}^{2,1650}$ (0.1, -0.01, $M_3$ )	$\mathcal{L}^{2,1690}$ (0.4, -0.25, $M_3$ )	$\mathcal{L}^{2,2270}$ (0.1, -0.30, $M_3$ )	$\mathcal{L}^{2,2310}$ (0.1, -0.2, $M_3$ )	$\mathcal{L}^{2,2390}$ (0.1, -0.2, $M_3$ )	...	22
$\mathcal{L}^{4,1,64,64}$ (0.8, $M_2$ , 1.989)	$\mathcal{L}^{4,2,64,64}$ (0.1, $M_2$ , 1.988)	$\mathcal{L}^{4,1,64,64}$ (0.1, $M_2$ , 1.995)	$\mathcal{L}^{4,2,64,64}$ (0.1, $M_2$ , 1.989)			4
2	2	2	2	1	...	26

Figura 4. Modelagem GNL sobre um sistema de dados de sensoriamento remoto.

Esta modelagem pode também ser utilizada em diversas outras aplicações envolvendo dados de sensoriamento remoto, incorporando à aplicação uma nova abordagem de mineração

de dados baseada em propriedades físicas.

#### 4. Conclusão

O sensoriamento remoto envolve, na maioria das vezes, a obtenção e análise de grandes quantidades de dados associados a medidas de alta resolução coletadas sobre sistemas naturais. Neste trabalho, foi proposta a utilização de um novo conceito para organização e análise de dados, que nós chamamos de representação em grades numéricas generalizadas. Através desta técnica inovadora, esperamos modelar e entender sistemas de dados complexos de forma mais eficiente, evitando análise e armazenamento redundantes de informação. A implementação do ambiente computacional para automatizar o processo de modelagem GNL já está em desenvolvimento e seus resultados devem ser publicados em breve, confirmando a aplicabilidade e grande utilidade do método em sensoriamento remoto e outras possíveis áreas.

#### Agradecimentos

Os autores agradecem ao CNPq pelo apoio financeiro.

#### Referências Bibliográficas

Bouman, B. A. M. Linking physical remote sensing models with crop growth simulation models, applied for sugar beet. **International Journal of Remote Sensing**, v. 13, n. 14, p. 2565-2581, 1992.

Datcu, M.; Seidel, K. Bayes. Bayesian methods: applications in information aggregation and image data mining. In: International Archives of Photogrammetry and Remote Sensing, 2005, Vallavoid, Espanha.

Cladis, P. E.; Palffy-Muhoray, P. **Spatio-Temporal Patterns in Nonequilibrium Systems**, Addison-Wesley, 1995.

Mecke, K. R.; Stoyan, D. **Statistical Physics and Spatial Statistics**, Springer-Verlag, 2000.

Petit, C. C.; Lambin, E. F. Integration of multi-source remote sensing data for land cover change detection **International Journal of Geographical Information Science**, v. 15, n. 8, p. 785-803, 2001.

Rosa, R. R.; Baroni, M. P. M. A.; Zaniboni, G.T.; Ferreira da Silva, A.; Roman, L. S.; Pontes, J.; Bolzan, M. J. A. Structural Complexity of disordered surfaces: Analyzing the porous silicon SFM patterns, **Physica A**, v.2, n. 386, p. 666-673, 2007.

Rosa, R. R.; Karlick, M.; Veronese, T. B.; Vijaykumar, N. L.; Dantas, M. S.; Borgazzi, A. I.; Sawant, H. S.; Sych, R. A.; Mendes, O. **Gradient pattern analysis of short solar radio bursts**, **Advances in Space Research**,v. 42, p. 844-851, 2008.

Solberg, A. H. S.; Jain, A. K.; Taxt, T. Multisource classification of remotely sensed data:

fusion of Landsat TM and SAR images, **IEEE Transactions on Geoscience and Remote Sensing**, v. 32, n. 4, p. 768-778, 1994.

Veronese, T. B.; Vijaykumar, N. L.; Rosa, R. R. Generalized Numerical Lattice: a formalism for representing time series complexity measures, **Journal of Computational Interdisciplinary Sciences**, submitted, 2008.