

Mineração de dados espaço-temporais para reconhecimento de padrões relacionados à vegetação aquática submersa

Ligia Flávia Antunes Batista ^{1,2}
Nilton Nobuhiro Imai ¹
Luiz Henrique da Silva Rotta ¹
Fernanda Sayuri Yoshino Watanabe ¹
Edivaldo Domingues Velini ³

¹ Universidade Estadual Paulista – UNESP
Faculdade de Ciências e Tecnologia – 19060-900 – Presidente Prudente - SP, Brasil
nnimai@fct.unesp.br, {luizhrotta, fernandasyw}@yahoo.com.br

² Universidade Tecnológica Federal do Paraná – UTFPR
86036-370 Londrina - PR, Brasil
ligia@utfpr.edu.br

³ Universidade Estadual Paulista – UNESP
Faculdade de Ciências Agrônômicas – 18603-970 – Botucatu-SP, Brasil
velini@fca.unesp.br

Abstract. This work presents the application of spatio-temporal visual data mining applied to submerged aquatic vegetation which may infest water bodies, a big problem in many reservoirs of Brazil. Thousands of points were acquired by hydroacoustic method in a reservoir of hydroelectric power plant within a period of five months. The aim of this work is to identify unknown patterns that are difficult of discovering only with traditional tools of Geographical Information Systems (GIS). The study area is a portion of Taquaruçu reservoir, located in Paranapanema river, near to Santo Inácio city, state of Paraná. The technique of mining used is clustering through self-organizing maps, coupled to a parallel coordinate plot (PCP), a reorderable matrix and a geographical multiple display, available in VIS-STAMP tool. The advantages of the applied method was to support user interactions at different levels, that allow to focus on detailed views of data. It was possible to identify homogeneous and heterogeneous areas in relation to dimensions space and time. The reorderable matrix highlights temporal dimension and the map matrix focuses on the spatial one. Besides, PCP represents an easy way to characterize clusters when it is necessary to deal with multivariate data. The results shows that VIS-STAMP is a great approach to overview complex patterns present in the data.

Palavras-chave: patterns, self-organizing maps, parallel coordinate plot, macrophytes, padrões, mapas auto-organizáveis, gráficos de coordenadas paralelas, macrófitas.

1. Introdução

A crescente aplicação de instrumentos de aquisição de dados, tais como sensores a bordo de satélites, aeronaves ou mesmo em nível terrestre, facilitam sobremaneira atividades que envolvem monitoramento espaço-temporal. Com isso, uma grande quantidade de informações vai se acumulando e o desafio passa a ser a análise desses dados, os quais são valiosos na medida em que podem revelar padrões úteis para processos de tomada de decisão. Neste contexto, a possibilidade de extrair conhecimento pela aplicação de técnicas de mineração de dados mostra-se bastante promissora, conforme apontam Loucks e Beek (2005).

Neste trabalho será abordado um problema de vários reservatórios brasileiros em que ocorre infestação de vegetação aquática, as macrófitas submersas, como descrito por Bini e Thomaz (2005) e Bini (1999). O mapeamento de macrófitas aquáticas submersas é importante para que se possa acompanhar seu ritmo de crescimento e com isso implantar estratégias de controle, uma vez que o crescimento excessivo gera problemas de cunho ecológico e econômico. Em termos ecológicos, há um desequilíbrio ambiental que, em situações extremas, pode causar mortandade de peixes devido à anoxia noturna. A questão econômica

está relacionada à necessidade de interromper a geração de energia elétrica devido ao entupimento de turbinas, gerando prejuízos.

Para realizar este mapeamento, uma das técnicas mais eficazes utiliza sensoriamento hidroacústico, em que pulsos sonoros são enviados e o retorno deste pulso gera informação para estimar altura de plantas submersas e batimetria. Por meio desta técnica, foi feito o acompanhamento da evolução das plantas de março a agosto de 2010, gerando milhares de pontos georreferenciados com dados de profundidade e altura de plantas.

Deste modo, este trabalho tem como objetivo caracterizar essa massa de dados por meio da metodologia de mineração de dados por visualização de padrões espaço-temporais multivariados.

A seção 2 contextualiza brevemente as técnicas de mineração de dados. Na seção 3 apresenta-se o domínio da aplicação, macrófitas submersas. Material e método são relatados na seção 4. A seção 5 descreve e discute os resultados obtidos. As conclusões estão na seção 6.

2. Mineração de Dados

Adriaans e Zantinge (1996) descrevem a mineração de dados como uma fase de um processo mais amplo denominado de descoberta de conhecimento. Os passos desse processo envolvem: seleção, limpeza, transformação, mineração e análise. A seleção consiste em filtrar os dados a serem processados. A etapa de limpeza envolve compatibilização dos dados, eliminação de duplicidades e valores nulos ou inválidos. A transformação converte os dados em formatos que facilitem a extração de conhecimento. A mineração envolve a aplicação de técnicas para busca de padrões, as quais dividem-se em várias categorias: classificação, agrupamento, associação, visualização, regressão, entre outras. Cada uma dessas categorias pode ser implementada com vários algoritmos, detalhados em Amo (2004), Loucks e Beek (2005) e Adriaans e Zantinge (1996).

Dentre as abordagens aplicadas, as técnicas de Inteligência Computacional têm se mostrado bastante úteis. Em particular os métodos baseados em aprendizado de máquina, como redes neurais, apresentam-se bastante promissores. Uma das classes de redes neurais são os mapas auto-organizáveis, que transformam de modo adaptativo um padrão de dimensão arbitrária em um mapa discreto uni- ou bidimensional, como descrito por Haykin (2001), o qual apresenta os modelos, propriedades e algoritmos dessa estrutura.

Diversos trabalhos cuja abordagem envolve mineração de dados espaciais têm sido realizados. Martins e Fonseca (2009) aplicam classificação orientada a objetos e mineração de dados para regiões de atividade agrícola. Korting et al. (2009) descrevem uma ferramenta para mineração de dados de sensoriamento remoto chamada GeoDMA, a qual implementa, além de outras funcionalidades, a extração de atributos espaciais e espectrais para serem usados na etapa de classificação. Freitas et al. (2009) utilizam a abordagem de mapas auto-organizáveis para seleção de feições e análise exploratória de dados multitemporais de sensoriamento remoto. Shimabukuro (2004) apresenta novas representações para conjuntos de dados com atributos espaciais e temporais para apoiarem a análise exploratória. Guo et al. (2006) apresentam um sistema para visualização para padrões espaço-temporais multivariados, cuja metodologia é aplicada no trabalho ora descrito.

3. Vegetação aquática submersa

As macrófitas aquáticas são plantas que podem ser consideradas daninhas em potencial, pois, se as condições ambientais são favoráveis, ela tenderá a proliferar-se ocupando extensas áreas. A colonização excessiva normalmente ocorre em ambientes com grande interferência antrópica, tais como eutrofização artificial e modificações hidrológicas (MONTEIRO; PEZZATO; HENRY-SILVA, 2003). No Brasil há vários registros da macrófita submersa *Egeria sp.*, como nos reservatórios de Itaipu (BINI; THOMAZ, 2005) e Jupuí (VELINI et al., 2005), a qual é a espécie predominante na área deste estudo.

Registra-se a presença de plantas submersas em profundidades da zona eufótica, em que chega luz, as quais variam conforme as condições de transparência e de sedimentos em suspensão na água. Macrófitas são de difícil mapeamento, pois, em virtude de serem submersas, complicam sua identificação por sensoriamento remoto óptico, uma vez que a radiância registrada pelos sensores não traz muita informação da coluna d'água, principalmente em grandes profundidades. Desta forma, o monitoramento deve ser feito diretamente em campo.

4. Material e método

A área de estudo compreende aproximadamente 1.5km² de área do reservatório de Taquaruçu, rio Paranapanema, no município de Santo Inácio, estado do Paraná, entre as coordenadas 51°44'49"W, 22°38'26"S e 51°43'2"W, 22°37'9"S, como ilustrado na Figura 1.

Os dados utilizados neste trabalho foram obtidos em levantamentos que mapearam profundidade e altura das plantas com a utilização de um ecobatímetro *Biosonics DT-X*,

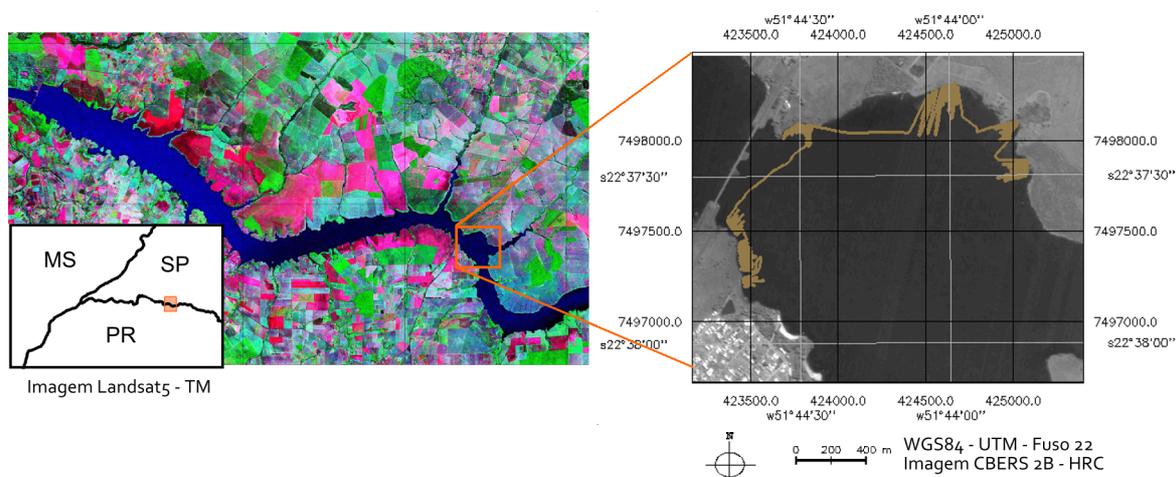


Figura 1. Área de estudos

acoplado a um GPS Garmin, instalado a 50cm de profundidade na lateral de um barco. Os levantamentos consistiram em percorrer a área de interesse delineando transectos. Na Figura 1 as linhas em marrom mostram o trajeto do barco durante o mapeamento. Observa-se que se percorre áreas das margens em direção a regiões mais profundas até um certo limite, em que não há mais presença de macrófitas submersas devido à profundidade.

Foram realizadas cinco campanhas de campo nas datas de 10/03/2010, 07/04/2010, 23/04/2010, 18/06/2010, 05/08/2010, identificadas respectivamente como campos 1, 2, 3, 4 e 5 neste texto. Cada uma gerou de 2000 a 6500 pontos amostrais com dados de profundidade e altura de plantas. As atividades de geoprocessamento, após cada trabalho de campo envolveram:

a) Transformação das coordenadas geodésicas para o sistema UTM com o aplicativo ProGrid e geração de arquivos no formato ASCII-Spring;

- b) Importação dos dados para o aplicativo Spring representados como Modelo Numérico de Terreno, gerando um plano de informação para cada data;
- c) Interpolação dos dados com o método de média ponderada;
- d) Recorte da área dos levantamentos, os quais se restringiam a regiões de possível ocorrência de plantas até profundidades de aproximadamente 7m, delimitadas por um polígono, conforme Figura 2(a);
- e) Exportação dos planos de informação de altura de plantas e profundidade em cada levantamento para formato TIFF;
- f) Com a ferramenta TerraView, o polígono da área de interesse foi importado e a partir dele a área foi discretizada pela divisão em células de 50 por 50m, totalizando 234 células, conforme Figura 2(b);
- g) Preenchimento de células com o *plugin* FillCell do TerraView, usando como dados de entrada os arquivos TIFF anteriormente exportados e a estratégia de preenchimento de valores médios;
- h) Exportação das células preenchidas para o formato *shapefile* e importação para o aplicativo VIS-STAMP. Essa ferramenta, proposta por Guo et al. (2006), baseia-se na abordagem de mapas auto-organizáveis de Kohonen para agrupar e visualizar dados espaço-temporais multivariados, a qual integra métodos visuais e cartográficos para explorar e entender os padrões espaço-temporais. Tal abordagem integra os seguintes componentes

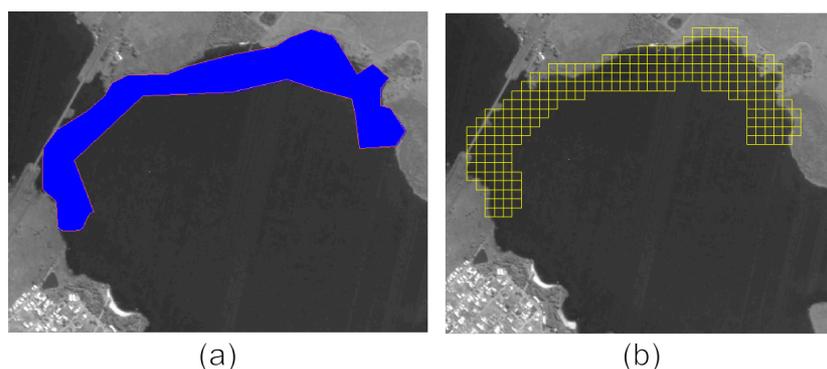


Figura 2. Etapas de processamento

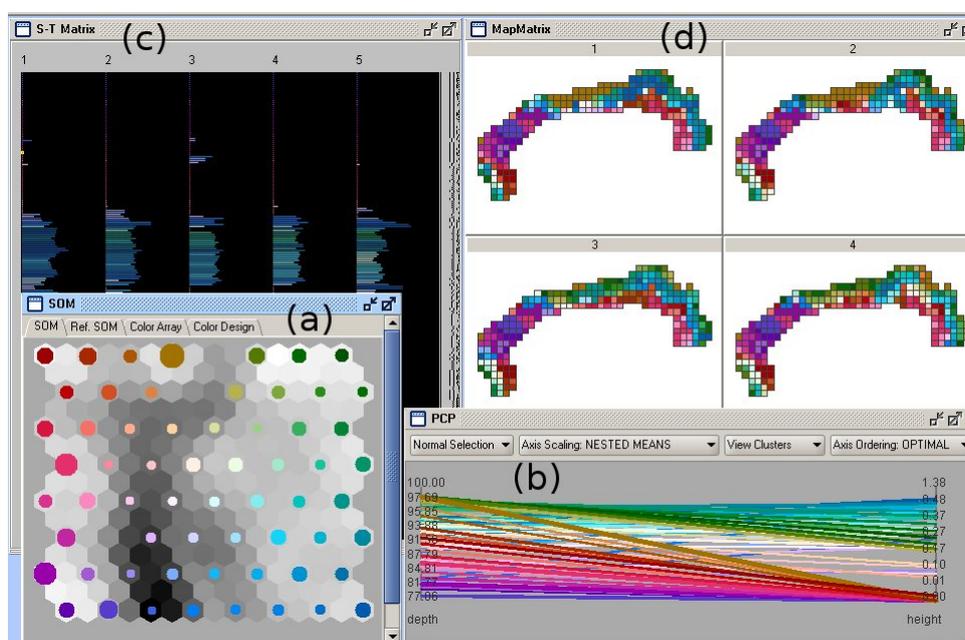


Figura 3. Módulos da metodologia proposta por Guo (2006)

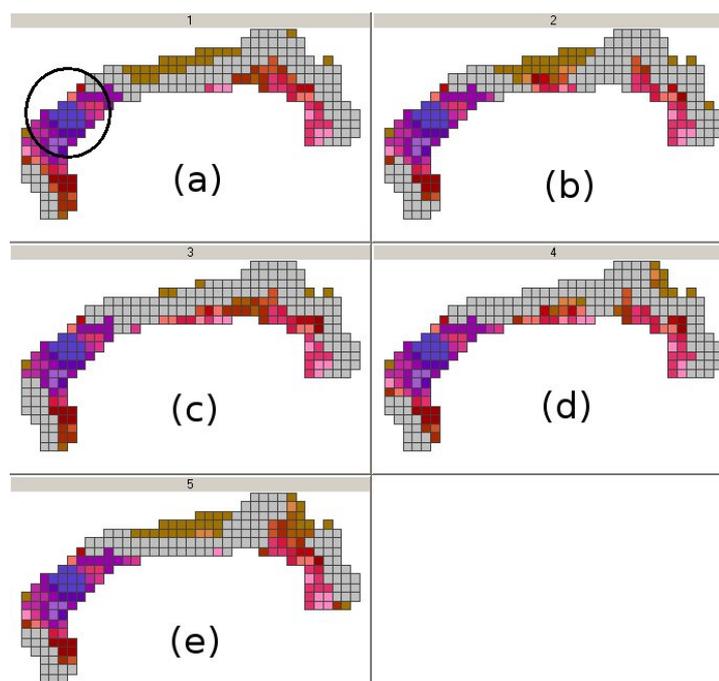


Figura 4. Seleção de grupos sem plantas

visuais: (a) mapa auto-organizável para agrupamento e visualização multivariada; (b) gráfico de coordenadas paralelas (GCP) que permite a identificação dos grupos gerados, como uma legenda; (c) matriz reordenável para organizar os padrões espaço-temporais; (d) matriz reordenável de mapas que permite avaliar os padrões espaciais. Os quatro módulos estão ilustrados na Figura 3, correspondentes respectivamente às janelas *SOM* (a), *PCP* (b), *S-T Matrix* (c), e *MapMatrix* (d).

A ferramenta VIS-STAMP requer valores positivos para os dados multivariados. Por isso adicionou-se 100 unidades aos dados de profundidade, o que pode ser observado pela Figura 5. Neste trabalho, ao lado do eixo de profundidade das figuras que ilustram GCP, foi colocado um eixo com a altura real (negativa) para facilitar as análises. A variável altura não foi modificada.

O GCP pode ser considerado o principal componente, o qual retrata um espaço de dimensão k em um espaço visual bidimensional, usando k eixos equidistantes paralelos. Cada eixo está associado a um atributo, sobre o qual são mapeados os respectivos intervalos de valores de dados, conforme explica Shimabukuro (2004).

5. Resultados e discussão

Para explorar os resultados, inicialmente selecionou-se as áreas sem plantas, as quais estão com células diferentes da cor cinza na Figura 4. A região marcada na Figura 4 (a) com o círculo mostra onde está o talvegue. Pelo GCP ilustrado na Figura 5, vê-se que áreas sem plantas estão, de modo geral, relacionadas a todas as profundidades. O mesmo pode ser visto na Figura 4, em que as células sem plantas aparecem tanto nas margens quanto nas regiões do talvegue. Entretanto, verifica-se que as regiões sem plantas concentram-se no canal (talvegue) e nas áreas limítrofes do levantamento, mais rasas ou mais profundas.

A Figura 6 ilustra áreas com presença de plantas, descritas pelo GCP da Figura 7. Vê-se que os tons em azul escuro representam plantas maiores, com aproximadamente 0.5m de altura. Já os tons em ciano e rosa claro possuem altura de até 0.2m.

A metodologia proposta por Guo (2006) possibilita ainda identificar e isolar áreas que não apresentaram grandes alterações ao longo do tempo, distinguindo-as de áreas heterogêneas.

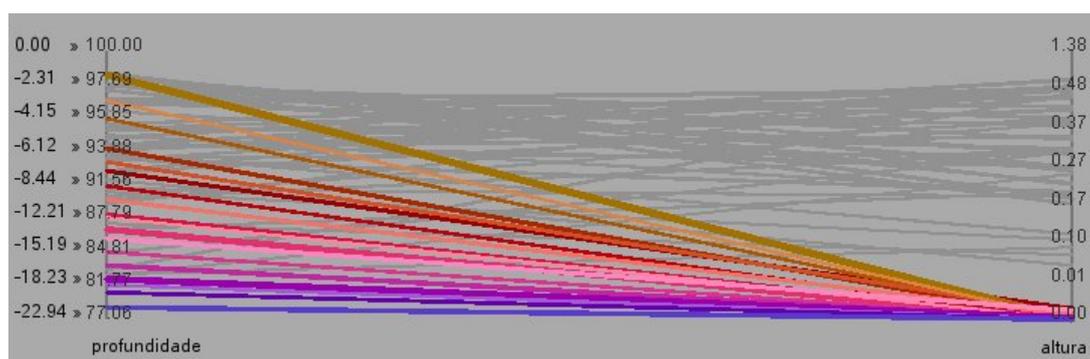


Figura 5. GCP de grupos sem plantas

Essa seleção é feita por meio da ferramenta matriz reordenável, ilustrada na Figura 8, a qual permite, ainda, avaliar a evolução temporal do objeto de estudo. Nesta figura, cada coluna representa um levantamento, rotulado como campo 1 a 5. O comprimento das barras horizontais coloridas, em cada coluna, indica a altura das plantas. A cor identifica o agrupamento ao qual pertence cada célula.

Juntos, a matriz de mapas, o GCP e a matriz reordenável permitem uma análise integrada das dimensões espaço e tempo. Percebe-se, por exemplo, que na matriz reordenável da Figura 8, as células azul escuras concentradas no retângulo 1 possuem altura de plantas praticamente

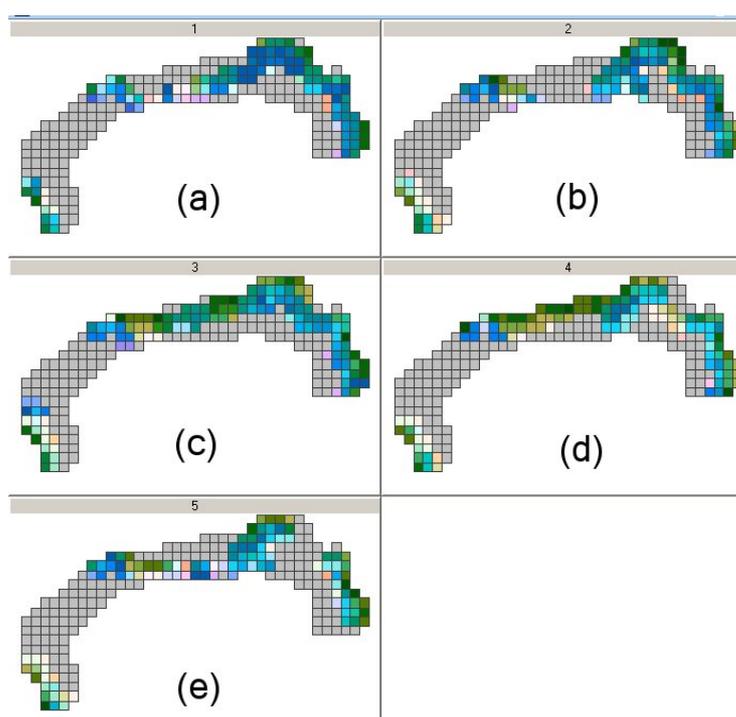


Figura 6. Células com plantas (altura maior que zero)

estável ao longo do tempo e estão localizadas em regiões de profundidades entre 4 e 6m, o que é constatado pela análise do GCP da Figura 7. A localização das células com tais características pode ser encontrada pela matriz de mapas da Figura 6, em profundidades intermediárias da região leste.

Ainda na Figura 8, o retângulo 2 destaca células com comportamento temporal bastante heterogêneo. Nas colunas referentes aos campos 1, 2 e 5 deste retângulo, a maioria das linhas horizontais não está visível, indicando altura de planta igual a zero. No campo 3 (23/abril), as

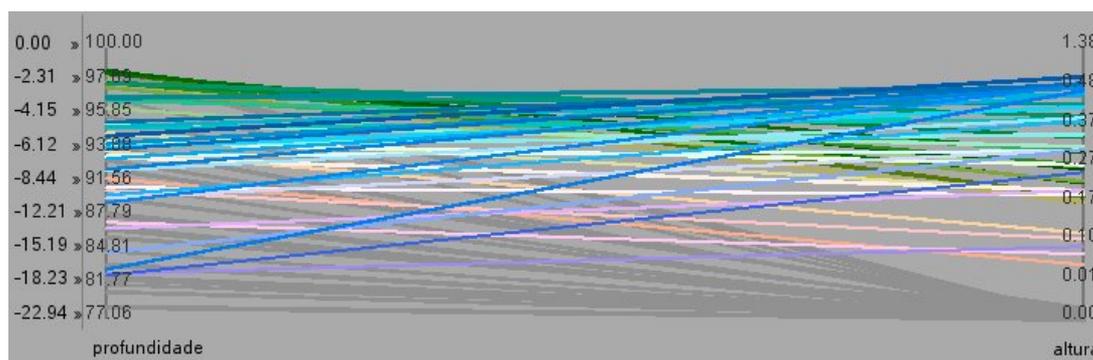


Figura 7. GCP de áreas com plantas

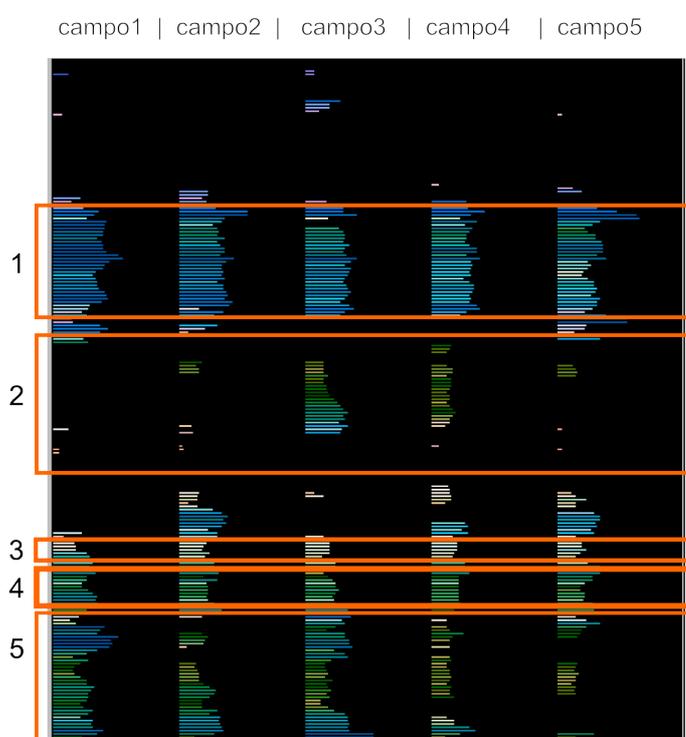


Figura 8. Matriz reordenável

plantas desta área tem a maior altura observada, a qual decresce no campo 4 (18/junho) e volta a ficar nula no campo 5 (05/agosto).

Com relação ao retângulo 3, a maior parte das células tem altura de plantas entre 0,15 e 0,4m, localizadas em profundidades de 5 a 6m, mantendo-se sem grandes alterações ao longo dos cinco levantamentos. O retângulo 4 também apresenta células homogêneas em relação ao tempo e esparsamente distribuídas no espaço. São células de maior altura (entre 0,3 e 0,5m) situadas em menores profundidades (2 e 4m).

O retângulo 5 novamente mostra comportamento espaço-temporal bastante heterogêneo, em que as células sem plantas alternam-se com as células com plantas de diferentes alturas, em profundidades variadas.

Enfatiza-se que essas análises foram realizadas selecionando cada retângulo no componente da ferramenta e a caracterização de altura e profundidade foi feita pelos respectivos gráficos de GCP de cada seleção.

6. Conclusões

Este trabalho apresentou a aplicação da metodologia de visualização de padrões espaço-temporais multivariados proposta por Guo (2006) a dados de macrófitas aquáticas submersas. Esta técnica de mineração permitiu a análise do grande volume de dados adquiridos durante mapeamentos realizados para avaliar o desenvolvimento da vegetação.

O método e a ferramenta computacional que o implementam mostraram-se bastante promissores para aplicações espaço-temporais, uma vez que permitem, de modo interativo, a aplicação de filtros, seleções e detalhamentos durante a busca de padrões, auxiliando sobremaneira as análises exploratórias.

Trabalhos futuros com este método incluem a consideração de outras variáveis físicas, climáticas e morfométricas que possam explicar o desenvolvimento da vegetação submersa em estudo.

Agradecimentos

Ao Programa de Pós-graduação em Ciências Cartográficas e ao Departamento de Cartografia da Faculdade de Ciências e Tecnologia da Universidade Estadual Paulista, Campus de Presidente Prudente. À UTFPR, por oferecer condições para realização deste trabalho.

Referências

- Adriaans, P.; Zantinge, D. **Data mining**. Harlow: Addison-Wesley Longman, 1996.
- Amo, S. Técnicas de Mineração de Dados. In: XXIV Congresso da Sociedade Brasileira de Computação. Jornada de Atualização em Informática (JAI), 2004, Salvador. **Anais...** Salvador: Universidade Federal da Bahia, 2004. p. 195-236.
- Bini, L. M.; Thomaz, S. M. Prediction of *Egeria najas* and *Egeria densa* occurrence in a large subtropical reservoir (Itaipu Reservoir, Brazil-Paraguay). **Aquatic Botany**, v. 83, p. 227-238, 2005.
- Bini, L. M. et al. Aquatic macrophyte distribution in relation to water and sediment conditions in the Itaipu reservoir, Brazil. **Hydrobiologia**, v. 415, p. 147-154, 1999.
- Freitas, R. M.; Adami, M.; Rudorff, B. F. T.; Shimabukuro, Y. E.; Rosa, R. R.; Silva, J. D. S. da; Arcoverde, G. F. B. Análise exploratória de dados multitemporais de sensoriamento remoto usando mapas auto-organizáveis: estudo em áreas de expansão da cana-de-açúcar. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 14., 2009, Natal. **Anais...** São José dos Campos: INPE, 2009. Artigos, p. 7797-7804.
- Guo, D.; Chen, J.; MacEachren, A. M.; Liao, Ke. A visualization system for space-time and multivariate patterns (VIS-STAMP). **IEEE Transactions on Visualization and Computer Graphics**. v. 12, n. 6, p.1461-1474, 2006.
- Haykin, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001.
- Korting, T. S.; Fonseca, L. M. G.; Escada, M. I. S.; Câmara, G. GeoDMA – Um sistema para mineração de dados de sensoriamento remoto. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 14., 2009, Natal. **Anais...** São José dos Campos: INPE, 2009. Artigos, p. 7813-7820.
- Loucks, D. P.; Beek, E. **Water Resources Systems Planning and Management**. Paris: United Nations Educational, Scientific and Cultural Organization, 2005. (Studies and Reports in Hydrology).
- Martins, V. A.; Fonseca, L. M. G. Classificação de uso de solo baseada na análise orientada a objeto e mineração de dados utilizando imagens SPOT/HRG-5. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 14., 2009, Natal. **Anais...** São José dos Campos: INPE, 2009. Artigos, p. 7837-7844.
- Shimabukuro, M. H. **Visualizações Temporais em uma Plataforma de Software Extensível e Adaptável**. 2004. 123p. Tese (Doutorado em Ciências da Computação e Matemática Computacional) - Universidade de São Paulo, São Paulo, 2004.