

Aplicação de Mapas Auto-Organizáveis para análise de dados de geosensores: estudo de caso na Mata Atlântica

Alessandra de França Santana¹
Rafael Duarte Coelho dos Santos¹
Gerson da Penha Neto¹
Nandamudi Lankalapalli Vijaykumar¹

¹ Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
{alessandra.santana, rafael.santos, gerson.neto, vijay}@lac.inpe.br

Abstract. A geosensor network can collect environmental data, process it and relay it for further analysis. One important step on this data analysis task is to verify the consistency and coherence of the collected data. The purpose of this paper is to demonstrate the use of a neural network as a tool for analyzing and visualizing data collected from a geosensor network for preliminary quality assessment. We adopted the model of Self-Organizing Maps, which aims to form clusters of data whose characteristics are similar to each other. Through these clusters is possible to highlight differences and similarities in the collected measurements, and show the behavior of the region in which the geosensor was allocated.

Palavras-chave: Self-Organizing Maps, geosensors, Atlantic Forest, Mapas Auto-Organizáveis, geosensores, Mata Atlântica.

1. Introdução

Estudos sobre o meio ambiente envolvem diversas áreas, como matemática, biologia, geografia e muitas outras. O meio ambiente também é foco importante na política, na mídia e nas pesquisas acadêmicas. Com o derretimento de geleiras, extinção de espécies, enchentes e secas intensas, percebemos que o planeta apresenta um desequilíbrio ambiental jamais registrado antes. Um exemplo bem típico é o desmatamento observado na Mata Atlântica. Correspondendo a cerca de 15% do território nacional, a Mata Atlântica se estende por uma área de 1.300.000 km² e apresenta uma porcentagem de devastação assustadora de 93% de sua formação original, segundo dados da Fundação SOS Mata Atlântica (2010). Pela sua relevância é importante estudar e compreender os fenômenos que ocorrem na floresta, em especial os relacionados à troca de carbono e influenciados pela ação antrópica e por fenômenos meteorológicos.

Para poder observar de perto alguns dos fenômenos presentes na Mata Atlântica, de forma constante e minuciosa, podem ser utilizados dispositivos chamados geosensores, que tem a capacidade de coletar dados, efetuar algum processamento básico e enviá-los para análise posterior. Também conhecidos como motes (*remote sensors*), Reis (2005), estes dispositivos podem ser compostos de vários tipos de sensores e uma unidade de coleta, armazenamento e repasse dos dados, que também armazena a data e a hora das medidas dos sensores.

Os motes coletam séries temporais de dados que devem ser analisados de acordo com o seu objetivo científico, mas freqüentemente é necessário verificar a qualidade dos dados coletados antes de usá-los. Como o volume de dados coletados pode ser muito grande (vários motes com vários sensores, alta freqüência de coleta de dados) é importante automatizar total ou parcialmente a análise preliminar dos dados, não só para garantir a qualidade dos mesmos mas para obter informações sobre o funcionamento dos sensores para possível manutenção.

O objetivo deste artigo é demonstrar o uso de uma rede neural (Mapas Auto-Organizáveis de Kohonen) como ferramenta de análise e visualização de dados desta natureza. Uma ferramenta baseada neste algoritmo permite evidenciar discrepâncias e semelhanças nas medidas coletadas, bem como mostrar o comportamento da região em que o mote foi alocado.

2. Dados

Em um experimento em geosensores, desenvolvido pela Universidade de São Paulo com apoio da FAPESP, Microsoft Research e Universidade Johns Hopkins, foram distribuídos 53 motes, cada um com três sensores de temperatura e um de umidade relativa do ar, em uma reserva próxima de São Luiz do Paraitinga, estado de São Paulo (Figura 1). Cada sensor fez leituras em intervalos de 30 segundos, coletando por dia em torno de 2.880 registros. Multiplicando este valor pelo número de sensores de temperatura, temos para um dia de coleta, um total de 8.640 registros. Ainda multiplicando este valor pelo número de motes, 53, temos para apenas um dia, um somatório aproximado de 457.920 leituras de temperaturas. Somando a isso as coletas de umidade relativa do ar, totaliza-se 610.560 dados de quatro sensores, para 53 motes, no período de um dia.

Foi criado um banco de dados com quase 25 milhões de dados, dentre esse total, 15 milhões de leituras de temperatura, 5 milhões de leituras de umidade relativa do ar, e mais 5 milhões de registros de data e hora da coleta. É importante ressaltar que por falhas nos sensores, nem todos os motes contém medidas para todos os instantes de tempo considerados no banco de dados.

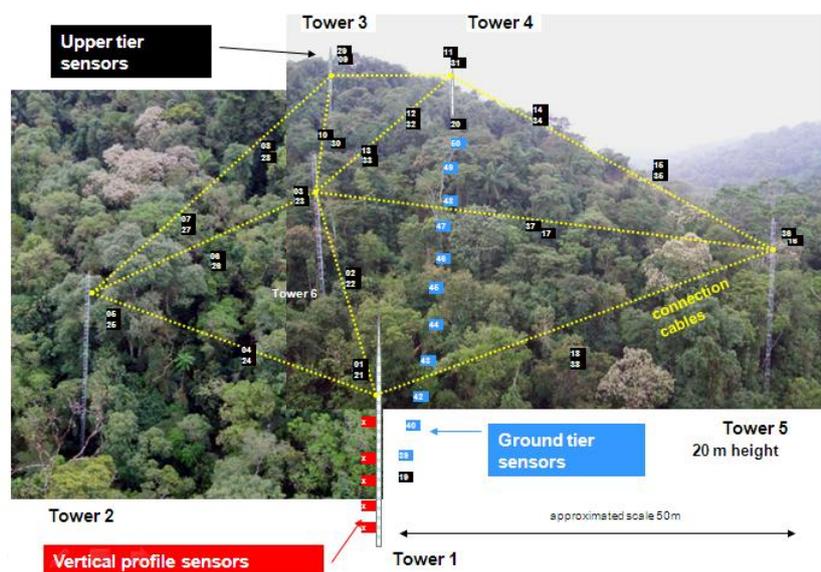


Figura 1. Disposição dos geosensores no sítio de coleta (fonte: Humberto Rocha, USP)

O banco foi modelado da seguinte forma: uma tabela que contém dados de leituras das temperaturas, umidade relativa do ar, data da coleta e identificação do mote, e uma tabela que contém apenas informações do mote, como sua identificação e coordenadas x e y.

As duas tabelas foram relacionadas através do campo de identificação do mote, em um relacionamento do tipo $n:1$, n leituras para cada mote. O banco de dados foi estruturado de forma a permitir o acesso dos dados pela ferramenta que fará a visualização dos mesmos.

3. Metodologia

3.1 Mapas Autos-Organizáveis

Dado o grande volume de dados coletados no experimento e a necessidade de visualizá-los para identificar agrupamentos, optou-se pelo uso dos Mapas Auto-Organizáveis, cujo intuito é agrupar os dados de entrada (que podem ter múltiplas dimensões, como no exemplo, sendo séries temporais) em um mapa bi-dimensional, preservando a topologia dos dados – em suma, dados que são próximos no espaço de atributos original permanecerão próximos no

espaço bi-dimensional criado pelo mapa, o que facilitará a sua visualização e identificação. Através da inspeção visual desses agrupamentos, será possível observar o comportamento dos sensores, se estão funcionando corretamente, e observar também algumas características das regiões de coleta em função destes dados.

Mapas Autos-Organizáveis, também conhecidos como Mapas de Kohonen, ou ainda Rede SOM (*Self-Organizing Maps*), é um modelo de Rede Neural Artificial, considerado como não supervisionada e competitiva. Resumidamente ela é caracterizada pela formação de um mapa topológico dos padrões de entrada em que a localização espacial dos neurônios no mapa gerado são indicativos de características estatísticas intrínsecas contidos nos padrões de entrada, Haykin (1999).

O funcionamento da rede pode ser descrito da seguinte forma: a rede recebe um conjunto de dados, que será o conjunto de padrões de entrada. Primeiramente é feita uma normalização dos padrões e a rede começa a ser treinada. O conjunto de padrões de entrada é adicionado a uma lista de padrões, e para cada iteração da época, encontra-se o vencedor, atualiza-se seu peso e de seus vizinhos e, por fim, remove-se um padrão da lista de padrões, até que a época termine, ou seja, até que não reste padrão algum na lista de padrões. Terminada a época, verifica-se se o erro mínimo foi obtido e, em caso contrário, inicia-se uma nova época até que o erro mínimo seja atingido.

Para o cálculo do vencedor, foi adotada a distância Euclidiana, Kohonen (2001), dada pela seguinte equação (Equação 1):

$$c = \operatorname{arg} \min \{d(x, m_i)\}, \quad (1)$$

onde x representa um padrão de entrada, e m_i um vetor de pesos.

Para a atualização dos pesos, Kohonen (2001), utilizou-se a equação (Equação 2):

$$m_i(t+1) = m_i(t) + \alpha(t) h_{ci}(t) [x(t) - m_i(t)], \quad (2)$$

onde m_i é um vetor de pesos; t é um inteiro representando uma coordenada de tempo; $h_{ci}(t)$ é a função de vizinhança, com a importância de definir o tamanho da região a ser atualizada; $\alpha(t)$ é a taxa de aprendizado, decrementada ao longo do treinamento da rede; $x(t)$ é o padrão de entrada no instante t .

Para o cálculo da vizinhança foi utilizada a função Gaussiana, Kohonen (2001), (Equação 3):

$$h_{ci}(t) = \alpha(t) \cdot \exp\left(-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}\right) \quad (3)$$

onde $\alpha(t)$ é a taxa de aprendizado no instante t ; $\sigma(t)$ define a largura do núcleo, também decrementada ao longo do treinamento da rede; r_c e r_i correspondem às coordenadas topológicas do neurônio vencedor e do neurônio da iteração, respectivamente.

3.2 Ferramentas

Para o desenvolvimento das ferramentas de visualização, utilizamos como linguagem de programação o C# e o Asp.Net.

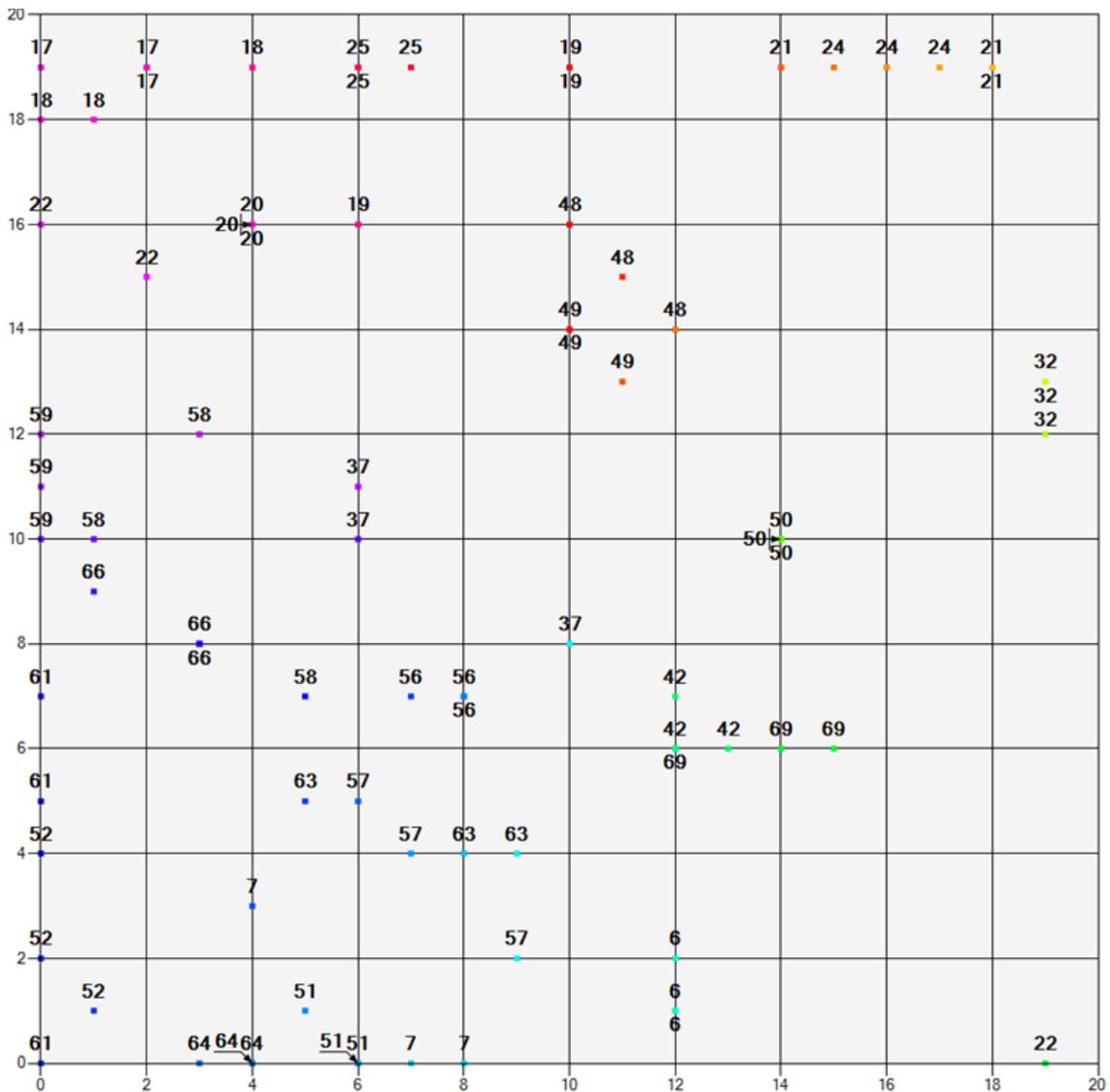
Usamos a IDE *Visual Studio 2010*; o SGDB *SQL Server 2008*. Também foram utilizadas tecnologias como LINQ (*Language Integrated Query*) para conexão com o banco de dados e o WCF (*Windows Communication Foundation*) como *Web Service*.

A aplicação foi desenvolvida obedecendo ao padrão MVC (*Model-View-Controller*), o que nos permite trabalhar com módulos independentes, diminuindo o acoplamento e provendo uma flexibilidade na edição de código. Conseguimos, com isso, uma aplicação eficiente, reutilizável e com uma manutenção facilitada, dada a independência de seus componentes.

4. Resultados

A ferramenta, usada em conjunto com o banco de dados coletados dos motes, pode ser configurada para extrair séries temporais a partir dos dados armazenados, em diversas configurações. Para demonstrar o algoritmo e sua implementação para os fins propostos, foi selecionado um conjunto de dados em um período de 3 dias, utilizando dados em intervalos de 4 horas. Somente os motes que apresentavam todos os dados, sem falhas, para todos os sensores no período foram considerados.

O objetivo deste primeiro experimento é verificar se as séries temporais (cada uma com 18 pontos) podem ser mapeadas em duas dimensões no mapa gerado pela ferramenta. Executou-se o algoritmo da rede neural e obteve-se o seguinte mapa, mostrado na figura 2:



A Figura 2 representa os estados finais dos neurônios da rede, dispostos como um mapa bi-dimensional. Cada ponto no mapa representa a série temporal correspondente a um sensor de temperatura. Como cada mote possui 3 sensores de temperatura, existem três rótulos numéricos para cada mote, em um total de 81 pontos.

A premissa deste trabalho é que usando o Mapa de Kohonen para visualizar os *clusters* formados com as séries temporais, será simples verificar se elas tem um comportamento semelhante ou diferente de outras séries correspondentes a outros motes ou mesmo ao mesmo mote. Para ilustrar as comparações visuais, algumas anotações foram feitas manualmente na Figura 2, resultando na Figura 3.

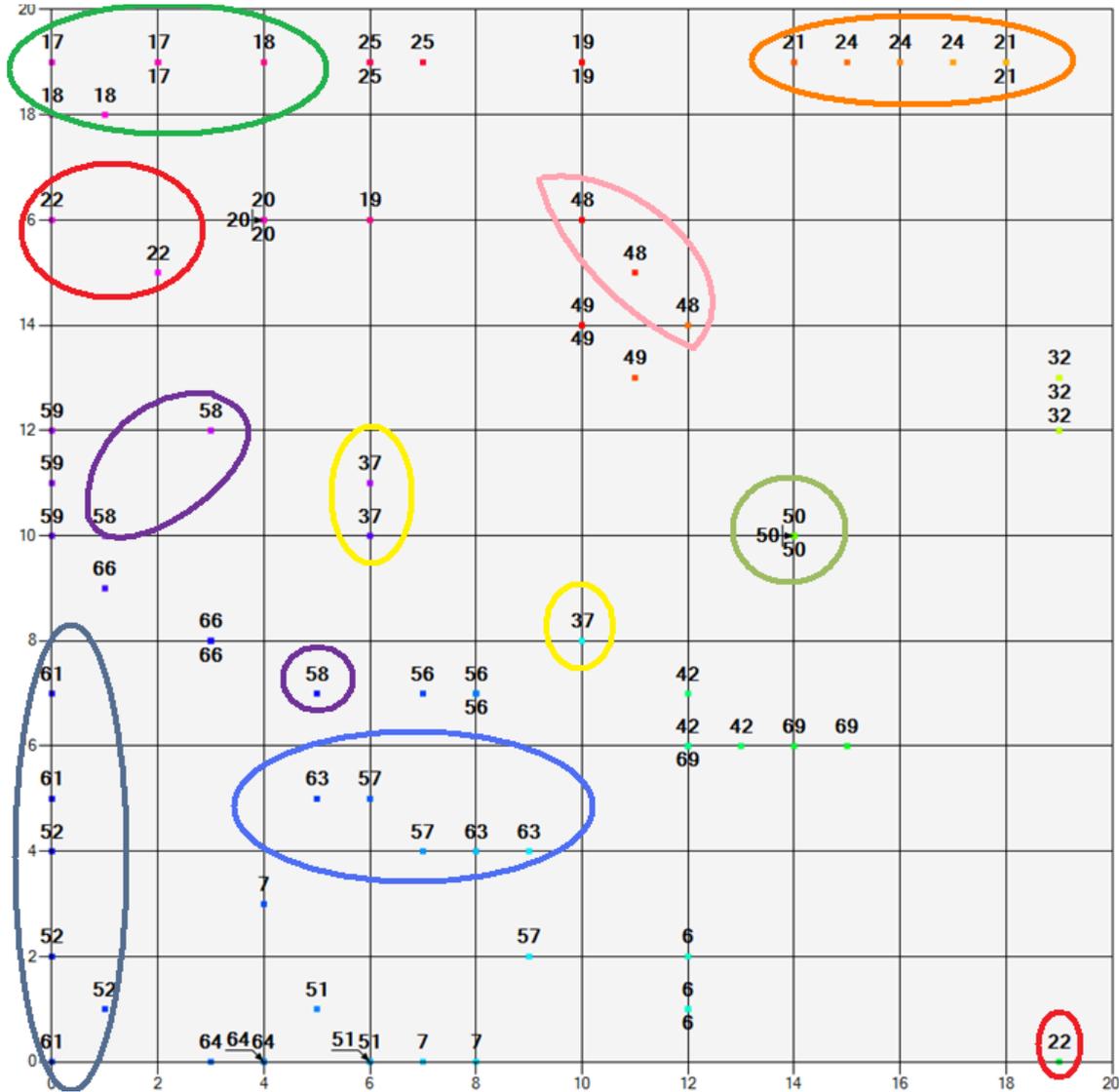


Figura 3. Formação de *clusters*

As anotações na Figura 3 mostram motes com comportamento notável para os fins deste estudo. Os dados usados para o treinamento da rede também foram usados para gerar gráficos correspondentes às séries temporais.



Figura 4. Mote 22

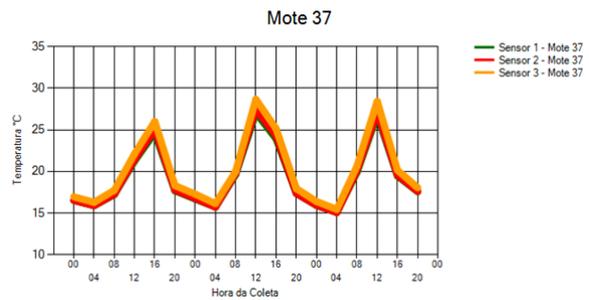


Figura 5. Mote 37

A Figura 4 mostra as três séries temporais correspondentes aos três sensores de temperatura do mote 22. Pode-se observar que um destes sensores, que é fisicamente próximo dos outros, apresenta leituras incompatíveis com o esperado e com o medido pelos outros dois sensores, indicando uma falha no sensor em si. Esta discrepância dos valores é refletida no mapa mostrado na Figura 3: os três pontos correspondentes ao mote 22 não estão próximos no mapa. A Figura 5, mostra as três séries temporais correspondentes aos três sensores de temperatura do mote 37 – nota-se alguma diferença entre as séries temporais, e correspondentemente, os três pontos do mote 37 apresentam-se relativamente próximos no mapa de Kohonen.

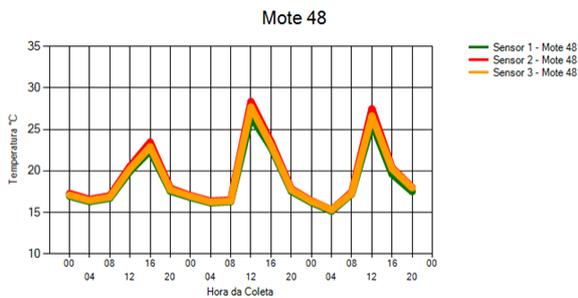


Figura 6. Mote 48

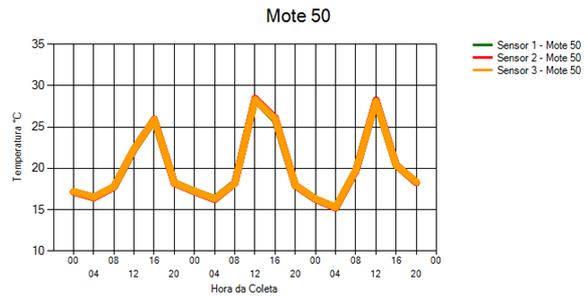


Figura 7. Mote 50

Os dados dos motes 48 e 50 são mostrados respectivamente nas Figuras 6 e 7. Podemos observar que os três sensores destes motes apresentam-se bem próximos no mapa de Kohonen e que suas séries temporais são praticamente idênticas.

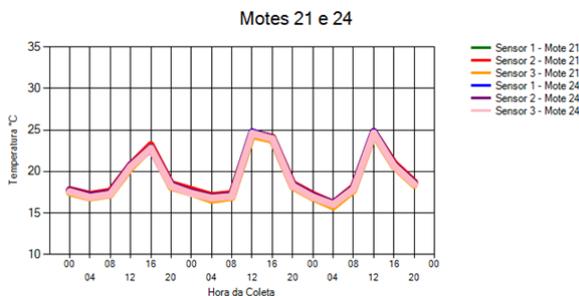


Figura 8. Motes 21 e 24

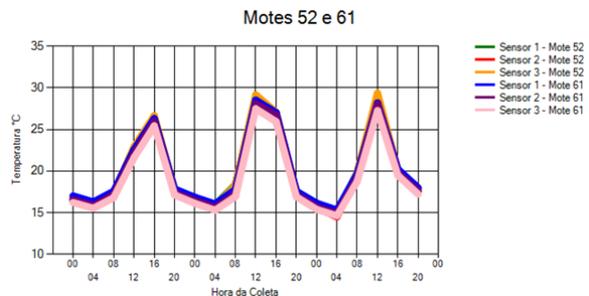


Figura 9. Motes 52 e 61

Outro tipo de visualização pode ser feito tomando-se todas as séries temporais de dois motes. A Figura 8 mostra as séries temporais dos motes 21 e 24, que apresentam comportamento praticamente idêntico e que estão próximas no mapa de Kohonen – estes motes estão localizados geograficamente próximos um do outro, e é de se esperar que suas leituras sejam semelhantes. A Figura 9 mostra as séries temporais dos motes 52 e 61, que

também são semelhantes e que estão próximos no Mapa de Kohonen, mas não no sítio onde foram instalados, indicando uma correlação que pode ser coincidental.

5. Conclusões

A ferramenta que implementa a visualização das séries com um mapa de Kohonen possibilita a análise de quão similares ou diferentes são os comportamentos dos sensores dos motes, avaliando se suas medições estavam dentro do padrão esperado, verificando quais motes tem comportamento semelhante, quais sensores do mesmo mote se comportam de forma diferente, etc.

Outras formas de visualização estão sendo consideradas e implementadas, em particular a visualização dos motes sobre a grade das posições geográficas dos mesmos. Usando uma técnica de colorização dos motes de acordo com sua posição no mapa de Kohonen será possível identificar visualmente de forma imediata motes cujos valores dos sensores não se comportam como os sensores geograficamente vizinhos.

Próximas versões da ferramenta possibilitarão a exploração das séries temporais em intervalos e amostragens diversos, assim como o uso de uma variante da rede neural que permite o processamento de séries com valores incompletos.

6. Agradecimentos

Os autores agradecem o apoio da FAPESP (Instituto Microsoft Research – FAPESP de Pesquisas em TI) no contexto do projeto 09/53154-0 (*Development and Application of a Geosensor Network for Environmental Monitoring*). Os autores também agradecem ao Professor Humberto Ribeiro Rocha do Departamento de Ciências Atmosféricas da Universidade de São Paulo por ter disponibilizado os dados de teste.

7. Referências

Reis, I. A. Roteamento Hierárquico de Dados em Redes de Geo-Sensores utilizando Conglomerados Espacialmente Homogêneos. 2005. 78 p. Proposta de Tese de Doutorado em Sensoriamento Remoto - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2005.

Fundação SOS Mata Atlântica - Portal SOS Mata Atlântica. Disponível em:

<<http://www.sosmatatlantica.org.br/index.php?section=info&action=mata>> Acesso em: 3.nov.2010.

Kohonen, T. **The Self-Organizing Maps**. Springer Series in Information Sciences. Springer, Berlin, Heidelberg, v. 30, ed. 3, 2001.

Haykin, S. **Neural Networks: A Comprehensive Foundation**. Prentice Hall. 2ed. New Jersey. 1999.