

# IDENTIFICAÇÃO DE ÁREAS PRIORITÁRIAS PARA RECUPERAÇÃO FLORESTAL COM O USO DE REDE NEURAL DE MAPAS AUTO-ORGANIZÁVEIS

*Identification of Priority Areas for Forest Restoration Using Self-Organizing Maps  
Neural Network*

GUSTAVO FELIPE BALUÉ ARCOVERDE<sup>1</sup>  
CLÁUDIA MARIA DE ALMEIDA<sup>1</sup>  
ARIMATEA DE CARVALHO XIMENES<sup>1</sup>  
EDUARDO EIJI MAEDA<sup>2</sup>  
LUCIANA SPINELLI DE ARAÚJO<sup>1</sup>

<sup>1</sup> Instituto Nacional de Pesquisas Espaciais  
Av. dos Astronautas, Caixa Postal 515- CEP: 12245-970  
São José dos Campos - SP, Brasil  
gustavo@dsr.inpe.br, almeida@dsr.inpe.br, arimatea@dpi.inpe.br  
lucian@dsr.inpe.br

<sup>2</sup> University of Helsinki - Department of Geosciences and Geography  
Gustaf Hällströmin katu 2, 00014, Helsinki, Finland  
eduardo.maeda@helsinki.fi

## RESUMO

O objetivo deste trabalho foi identificar áreas prioritárias para a recuperação florestal e analisar variáveis a elas relacionadas através da rede neural artificial (RNA) de Mapas Auto-Organizáveis (SOM), em duas escalas. Primeiramente, procurou-se identificar uma sub-bacia hidrográfica prioritária para a recuperação florestal na Unidade de Gerenciamento de Recursos Hídricos Paulista (UGRHI) do rio Paraíba do Sul por SOM. Para isto, foram utilizadas variáveis de conectividade ambiental e cobertura florestal. Definiu-se uma sub-bacia hidrográfica situada na represa do Jaguari, município de Igaratá, para estudo em uma escala de maior detalhe. Nas Áreas de Proteção Permanentes (APPs) englobadas nesta sub-bacia hidrográfica, foi realizada uma nova análise por SOM. Neste caso, foram consideradas variáveis de distância a fragmentos florestais, a áreas urbanas, a estradas pavimentadas e a construções rurais, assim como o Índice de Vegetação por Diferença Normalizada e o Potencial Natural de Erodibilidade Laminar. Em ambas as escalas, as áreas prioritárias para a recuperação florestal foram determinadas

através de histogramas do somatório dos valores dos Mapas Auto-Organizáveis de cada variável por agrupamentos delimitados. Por fim, foi gerado um mapa de contribuição de amostras para neurônios vencedores, o que permitiu uma nova abordagem para a análise dos agrupamentos gerados.

**Palavras-Chave:** Redes Neurais Não-Supervisionadas; Recuperação Florestal; Reconhecimento de Padrões Espaciais; Bacia Hidrográfica.

### ABSTRACT

The aim of this work was to identifying priority areas for forest restoration and analyze variables related to such areas at two distinct spatial scales using Self-Organizing Maps neural network (SOM). Initially, a SOM analysis was conducted to detect a watershed suitable for forest restoration within the Management Unit for Hydrological Resources of the Paraíba do Sul river, located in São Paulo State, southeast of Brazil. The variables employed in this analysis were environmental connectivity and forest cover. The Jaguari watershed, located in the municipality of Igaratá, was selected as study area in the second stage of analysis. In the permanent preservation areas along riversides within this watershed, a new SOM analysis was performed to detect suitable areas for forest restoration. At this more refined scale, the regarded variables were distance to forest fragments, urban areas, paved roads, and rural constructions, as well as the NDVI (the Normalized Difference Vegetation Index) and the natural soil erosion potential. At both scales, the priority areas for forest restoration were assessed based on cluster histograms of SOM. Finally, a contributive map of samples for the best matching units was elaborated, and that enabled an insightful approach for the analysis of the generated clusters.

**Keywords:** SOM; Forest Restoration; Spatial Pattern Recognition; Watershed.

### 1. INTRODUÇÃO

A conversão da cobertura natural em espaços alterados pelo homem alcançou magnitude sem precedentes a partir do século XX. As alterações de coberturas naturais provocam desequilíbrios no ciclo de nutrientes, de energia e de permutação gênica em diferentes ecossistemas e biomas. Walker (2004) exemplifica algumas destas conseqüências, tais como: aumento de gases de efeito estufa, perda de biodiversidade, além do assoreamento de rios e lagos. Nesse contexto, um tema de especial interesse é o desmatamento de florestas tropicais. Estas florestas estão presentes na maioria dos países subdesenvolvidos e em desenvolvimento e, portanto, vivenciam um crescimento populacional maior do que o observado nas demais regiões do planeta. Este fato desencadeia uma maior pressão sobre o uso de recursos naturais locais e acelera o desmatamento de ecossistemas de grande biodiversidade (MYERS et al., 2000). Uma das florestas tropicais mais ameaçadas é a Mata Atlântica no Brasil, um dos ecossistemas integrantes da lista dos 25 biomas de alta biodiversidade mais ameaçados do planeta (RODRIGUES et al., 2009). Os remanescentes florestais da Mata Atlântica correspondem a apenas 8,8% da sua

extensão original (SOS MATA ATLÂNTICA, 2009), fato que é igualmente associado a uma alta fragmentação da mesma. Mesmo assim, este bioma proporciona abrigo para várias populações tradicionais e garante abastecimento de água para mais de 100 milhões de pessoas (CONSERVAÇÃO INTERNACIONAL, 2008).

Segundo o Inventário Florestal da Vegetação Natural do Estado de São Paulo, (INSTITUTO FLORESTAL, 2005), a vegetação natural remanescente situada em Unidades de Conservação no estado é de 25%, sendo que a maior parte está localizada em áreas particulares. Portanto, programas de recuperação florestal em propriedades particulares são considerados como de extrema relevância para este bioma. Considerando a rápida dinâmica do desflorestamento na Mata Atlântica, é imprescindível que estes programas tenham forte embasamento técnico-científico na alocação de metas e recursos. Para tanto, estudos devem contemplar o histórico ambiental da área, sua resiliência ambiental, a adequação pedológica para atividades agrícolas, a escolha das melhores sementes de espécimes florestais, as técnicas mais apropriadas de recuperação florestal, os recursos humanos envolvidos, os tipos de convênios para a recuperação, o contexto financeiro para a implementação desses convênios, entre outros (USP, 2009). Todos estes fatores, direta ou indiretamente, estão relacionados com a localização na qual será executado o programa de recuperação florestal.

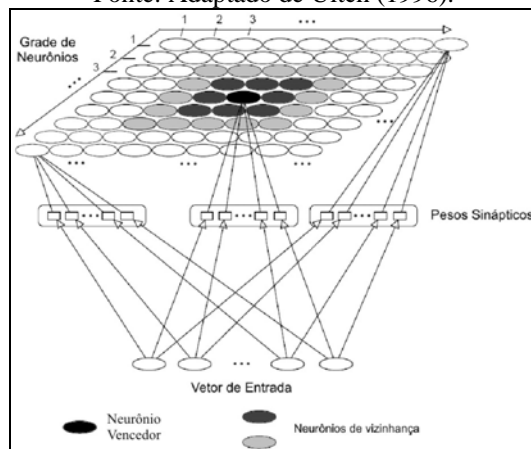
A identificação de áreas prioritárias para a recuperação florestal está estreitamente relacionada com fatores edafo-climáticos, ecológicos e com áreas de influência antrópica atual e potencial. Para tanto, podem ser utilizadas técnicas de reconhecimentos de padrões das variáveis espaciais, a fim de que possam ser ordenadas as áreas mais propícias para projetos de recuperação florestal. Existem diferentes métodos de reconhecimento de padrões que se prestam à análise exploratória de dados e ao ordenamento prioritário dos mesmos. Estes podem ser divididos em métodos lineares e não-lineares. São métodos lineares: análise de principais componentes (APC), análise de correspondência (AC), análise de agrupamentos (AA) e análise discriminante (AD). São métodos não-lineares: ordenamento polar (OP), escalonamento multidimensional não-métrico (EMNM) e rede neural artificial (RNA) de mapas auto-organizáveis, mais conhecida como *SOM*, acrônimo de *Self-Organizing Maps*. A seção seguinte apresenta os conceitos fundamentais relacionados a mapas auto-organizáveis.

## 2. FUNDAMENTAÇÃO TEÓRICA DE MAPAS AUTO-ORGANIZÁVEIS

A rede neural de mapas auto-organizáveis (ou rede do tipo *SOM*), desenvolvida por Kohonen (1982), integra a disciplina de reconhecimento de padrões e se destina principalmente à análise exploratória de dados e de agrupamentos dos mesmos (KOHONEN, 1982). Esta categoria de RNA baseia-se na aprendizagem competitiva e não-supervisionada (HAYKIN, 2001; KOHONEN et al., 2000). Cada variável é mapeada em um espaço finito de neurônios organizados em arranjo, geralmente bidimensional (Mapa de Kohonen) (Figura 1).

O neurônio representa uma abstração formal em RNAs e corresponde concretamente a agrupamentos de valores semelhantes para distintas variáveis, sendo estes valores relativos a um conjunto de uma ou mais unidades territoriais adotadas na análise. Preliminarmente, são distribuídos pesos aleatórios para cada neurônio do Mapa de Kohonen, e o conjunto de valores para cada variável é normalizado (fase de inicialização). Os valores de cada variável nas unidades territoriais representam um vetor de entrada (amostra) na *SOM*. No treinamento seqüencial, a cada iteração, é calculada a distância euclidiana mínima de um vetor de entrada para cada um dos pesos dos neurônios do Mapa de Kohonen; o neurônio que possuir a menor distância entre o vetor e o seu peso será denominado neurônio vencedor (fase de competição). O neurônio vencedor determinará a localização espacial dos neurônios vizinhos (excitados), aproximando-os para si, com base em um raio de influência (fase de cooperação). Finalmente, é realizada a atualização dos pesos sinápticos dos neurônios que foram aproximados para o neurônio vencedor através de uma função gaussiana, baseando-se em taxas de aprendizagem que devem diminuir a cada iteração, de modo que a RNA possa convergir (HAYKIN, 2001).

Figura 1 - Mapa de Kohonen.  
Fonte: Adaptado de Ulch (1996).



O processo de treinamento pode ser dividido em uma primeira fase, de ordenação, e uma segunda, de convergência. Na ordenação, são utilizados raios de influência e taxas de aprendizagem maiores, e na convergência, estes valores são reduzidos. Como critérios de quantificação de erros, são calculados o erro de quantização médio e o erro topológico. O erro de quantização médio é a distância euclidiana da diferença entre todos os vetores de entrada com relação ao vetor do neurônio vencedor na média de todas as iterações. Este erro varia de 0 a 1, caso a normalização dos vetores de entrada tenha ocorrido nesta mesma grandeza. O erro

topológico diz respeito à qualidade da grade do Mapa de Kohonen. Este erro será maior quanto maior for a distância entre o primeiro e o segundo neurônio vencedor para a média de todos os vetores de entrada. O seu valor máximo é 1, e mínimo, 0. Para maiores explicações, podem ser consultados Vesanto et al. (2000) e Haykin (2001).

Vesanto et al. (2000) propuseram algumas heurísticas para a dimensão e cálculo do Mapa de Kohonen, assim como para os parâmetros de treinamento da SOM. Os autores conceberam a criação de uma matriz de distância unificada (Matriz-U), que mostra a distância euclidiana, ao final de todas as iterações executadas, entre os respectivos neurônios e os vetores de entrada para os Mapas de Kohonen de todas as variáveis utilizadas (VESANTO et al., 2000). Na Matriz-U, estão condensadas as informações de todas as variáveis e, portanto, é possível verificar visualmente os possíveis agrupamentos. Além da Matriz-U, outra forma de visualização dos dados que pode ser útil para a descoberta de conhecimento são os chamados Planos de Componentes, que possibilitam visualizar as correlações entre os atributos (variáveis) de entrada do mapa. Para cada atributo, é gerado um Plano de Componente, que é uma representação gráfica (codificação de cores) dos valores de cada variável no vetor de pesos (LEITE et al., 2006). Também é possível delimitar agrupamentos através do método k-médias entre os Mapas de Kohonen existentes. Considerando que a quantidade de agrupamentos é previamente definida pelo usuário da SOM, a escolha do melhor número de agrupamentos pode ser estabelecida conforme o índice de Davies-Bouldin. Este índice visa encontrar agrupamentos esféricos, com compacidade interna e, ao mesmo tempo, com boa separabilidade entre os demais agrupamentos (DAVIES; BOULDIN, 1979). O índice de Davies-Bouldin é dado por:

$$\frac{1}{c} \sum_{k=1}^c \max_{c \neq k} \left\{ \frac{S_c(Q_k) + S_c(Q_l)}{d_{ce}(Q_k, Q_l)} \right\}, \quad (1)$$

em que  $c$  é o número de agrupamentos,  $S_c(Q_k)$  representa a distância intra-agrupamento (dispersão interna do agrupamento  $Q_k$ ), baseada na distância para o centróide,  $d_{ce}(Q_k, Q_l)$  representa a distância entre os agrupamentos  $Q_k$  e  $Q_l$ , também baseada na distância entre os centróides.  $S_c(Q_k)$  é dada por:

$$S_c(Q_k) = \left\{ \frac{1}{N_k} \sum_{j=1}^{N_k} |x_j - c_k|^q \right\}^{1/q}, \quad (2)$$

em que  $x_j \in Q_k$ ,  $N_k$  é o número de amostras no agrupamento  $Q_k$ , e  $c_k = 1/N_k \sum_{x_i \in Q_k} x_i$ .  $d_{ce}$  é dada por:

$$d_{ce}(Q_k, Q_l) = \left\{ \sum_{k=1}^d |c_{ki} - c_{lj}|^p \right\}^{1/p}, \quad (3)$$

em que  $d$  corresponde à dimensão do vetor  $x_k$ . Nos casos em que o expoente  $p = 2$ , tem-se  $d_{ce}$  como a distância euclidiana entre os centróides dos dois agrupamentos.  $S_c$  representa a raiz  $q$ -ésima do  $q$ -ésimo momento dos pontos no agrupamento  $k$  em

torno da média. Se o expoente  $q = 1$ ,  $S_c$  torna-se a média euclidiana dos vetores no agrupamento  $k$  em relação ao centróide deste grupo. Nos casos em que  $q = 2$ ,  $S_c$  torna-se o desvio-padrão da distância das amostras com relação ao centróide do grupo. Por fim, é importante destacar que o menor valor do índice de Davies-Bouldin corresponderá ao número de agrupamentos mais ajustado.

Considerando a temática exposta na Seção de Introdução, os objetivos deste trabalho foram identificar áreas prioritárias para a recuperação florestal através da rede neural *SOM*, bem como analisar a *SOM* como técnica de exploração de variáveis para a recuperação florestal.

### 3. ESTADO DA ARTE: SOM EM APLICAÇÕES FLORESTAIS

Há um número extremamente reduzido de trabalhos na literatura indexada que se dedicaram a explorar rede neural *SOM* em questões florestais. Selamat e Selamat (2006) utilizaram esta categoria de rede neural para detectar agrupamentos de ocorrências de incêndios florestais na Indonésia. Os dados de entrada da rede basearam-se na quantidade de focos de incêndio, que se espalharam em diferentes direções da área de estudo, dentro de um período de 10 meses. Os autores aplicaram inicialmente uma análise por principais componentes para remover a hiperdimensionalidade do banco de dados original, bem como para melhorar a acurácia dos resultados de classificação da rede. A *SOM* produziu enfim um mapa de agrupamentos das ocorrências de incêndios em quatro classes referentes ao risco desses eventos, como a baixa e alta propensão de ocorrência de incêndios, além da baixa e alta magnitude (impacto) desses eventos. Com base nos resultados, os autores foram capazes de identificar um período de tempo em que ocorre a maior quantidade e intensidade de incêndios florestais na região estudada.

Lippitt et al. (2008) empregaram cinco métodos de reconhecimento de padrões para identificar agrupamentos de corte seletivo de madeira no estado de Massachusetts (EUA), três dos quais referentes a redes neurais artificiais dos tipos *SOM*, *Multilayer Perceptron* e *ARTMAP*, e dois deles referentes a árvores de decisão, utilizando os algoritmos de gini e de entropia. Dentre as redes neurais utilizadas, a *SOM* se mostrou como a menos sensível à presença de ruído e de variações nos dados de entrada. Da mesma forma, a rede *SOM* mostrou-se estável e foi capaz de produzir os resultados mais precisos em comparação às demais categorias de rede adotadas na análise.

As conclusões a que chegaram Lippitt et al. (2008) corroboram as obtidas anteriormente por Giraudel e Lek (2001), os quais empregaram *SOM* e alguns dos métodos de reconhecimento de padrões citados na Seção de Introdução (*APC*, *AC*, *OP*, *EMNM*) para o ordenamento de uma comunidade de plantas em Wisconsin (EUA). A finalidade deste estudo era comparar os resultados da *SOM* com os demais métodos usualmente empregados no âmbito da Ecologia. Os autores salientam que nos dois métodos não-lineares, *OP* e *EMNM*, ambos não provêem uma análise simultânea entre as espécies de amostras e as variáveis, como acontece com a *AC* e a *SOM*. Por outro lado, a *AC* é um método linear, ao contrário da *SOM*,

que é não-linear. Em termos de resultados, todas as técnicas apresentaram semelhanças em relação aos dados investigados. Mesmo assim, os autores apontam que a *SOM* seria a opção mais vantajosa para estudos em Ecologia, devido à possibilidade de visualização da proporcionalidade associativa entre amostras e variáveis, além de ser um método mais robusto quando não se conhece o comportamento dos dados, fato comum na Ecologia.

De forma diversa à abordagem metodológica dos trabalhos acima expostos, este estudo apresenta uma contribuição pioneira ao utilizar uma rede neural *SOM* com fins à identificação de áreas prioritárias para a recuperação florestal em um trecho de Mata Atlântica no leste do estado de São Paulo, adotando a bacia hidrográfica como unidade de análise espacial.

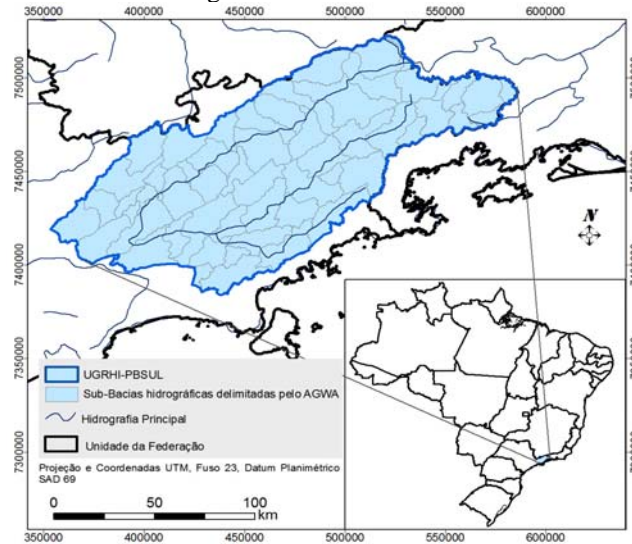
#### **4. MATERIAL E MÉTODOS**

##### **4.1 Área de estudo: escolha e definição**

Este estudo foi executado no âmbito de bacia hidrográfica e em dois diferentes níveis de escala espacial: regional e sub-regional. A bacia hidrográfica em nível regional foi subdividida em sub-bacias hidrográficas, as quais constituíram elementos de análise para a *SOM*, a fim de se identificar aquela com maior prioridade para a recuperação florestal. A sub-bacia hidrográfica identificada em nível regional foi posteriormente analisada em nível sub-regional, considerando apenas as áreas compreendidas como de preservação permanente (APP) de cursos d'água (APP-Água), que incluem trechos de 30 m de largura contados a partir de ambas as margens de cursos d'água, além de um *buffer* de 50 m em torno das nascentes, de acordo com a Resolução CONAMA 303/2002 (BRASIL, 2002).

A escolha da bacia hidrográfica foi baseada em informações existentes de inventário florestal, histórico de degradação ambiental e informações adicionais de sensibilidade ambiental. Considerando que o estado de São Paulo possui inventários florestais periódicos, optou-se por escolher uma bacia hidrográfica nesta unidade da federação. Associado a isto, a bacia hidrográfica do rio Paraíba do Sul detém um longo histórico de ocupação, com alta fragmentação florestal e trechos com alta necessidade de conectividade ambiental. Desta forma, foi escolhida a bacia hidrográfica da Unidade de Gerenciamento de Recursos Hídricos (UGRHI) do rio Paraíba do Sul no seu trecho paulista (UGRHI-PBSUL), com aproximadamente 1.458.210 ha (Figura 2).

Figura 2 – Área de estudo.



#### 4.2 Material

A escolha do conjunto de dados de entrada em uma rede neural artificial é decisiva para o sucesso de seu desempenho, e conseqüentemente, para a qualidade de seus resultados (BOSCHI; GALO, 2007; FERREIRA; GALO, 2010). Na escala regional, foram utilizados os seguintes insumos: (a) mapa de cobertura florestal do Inventário Florestal (INSTITUTO FLORESTAL, 2005); (b) mapa vetorial de hidrografia extraído de cartas topográficas do IBGE em 1:50.000 (IBGE, 1984); (c) “Mapa de Áreas Prioritárias para Incremento da Conectividade” (MAPIC); e (d) dados de altimetria do *Shuttle Radar Topography Mission* (SRTM). O MAPIC foi necessário para avaliar as áreas com maior necessidade de conectividade florestal. Este mapa é um resultado do Projeto “Diretrizes para a Conservação e Restauração da Biodiversidade no Estado de São Paulo”, lançado em 2007, indicando faixas geográficas para compensação ambiental diferenciada, de acordo com a Resolução SMA-SP 85/2008 (PROGRAMA BIOTA/FAPESP, 2007; SÃO PAULO, 2008).

Na escala sub-regional, foram utilizados os seguintes insumos: (a) mapa de cobertura florestal do Inventário Florestal (INSTITUTO FLORESTAL, 2005); (b) mapas vetoriais de hidrografia e altimetria extraídos de cartas topográficas do IBGE em 1:50.000 (IBGE, 1984); (c) mapas vetoriais de rodovias, extraídos a partir de imagens de alta resolução disponíveis no Google Earth (GOOGLE EARTH, 2008); (d) mapa vetorial dos limites de construções rurais, extraídos a partir de imagens de satélite de alta resolução disponíveis no Google Earth (GOOGLE EARTH, 2008);



(e) dados e localização de estações pluviométricas (ANA, 2009); (f) mapa de solos (MME, 1977); e (g) imagem do sensor TM/Landsat de 8 de setembro de 2004 - órbita-ponto 219/076 (INPE, 2004).

### 4.3 Métodos

#### 4.3.1 Escala regional

Primeiramente, foi necessário delimitar as sub-bacias hidrográficas da UGRHI-PBSUL. Para tanto, foram utilizados os dados de altimetria do SRTM, os quais foram processados no programa Automated Geospatial Watershed Assessment - AGWA, uma extensão do programa ArcView 3, que combina dados de um modelo digital de elevação a grades de direção de fluxo e de direção de fluxo acumulado dele derivadas (BURNS et al., 2004). Neste nível de análise regional, foram consideradas como sub-bacias hidrográficas com prioridade para recuperação florestal aquelas que apresentassem o maior percentual de desflorestamento (ou menor percentual de cobertura florestal) aliado à maior necessidade de conectividade ecológica na paisagem. Essa necessidade refere-se ao estabelecimento e proteção de corredores e manchas de vegetação para o trânsito de espécies animais em busca de comida, bem como à regeneração florestal nos mesmos e, conseqüentemente, ao aumento de áreas com maior matriz florestal para o abrigo dessas espécies. Assim, foram utilizadas as seguintes variáveis:

- a) Percentual da cobertura florestal do Inventário Florestal para as APP-Água contidas nas respectivas sub-bacias hidrográficas (%APP-Água\_VEGETAÇÃO). Ou seja, a razão entre área de vegetação natural em APP-Água na respectiva sub-bacia (*a*) e a área total de APP-Água na respectiva sub-bacia (*b*), conforme a Equação 4.

$$\%APP - \text{Água\_VEGETAÇÃO} = \frac{a}{b} * 100 \quad (4)$$

- b) Valor de conectividade ponderado pela área por sub-bacia hidrográfica (Valor\_Conectividade). Ou seja, em cada sub-bacia hidrográfica, este valor é dado pela soma das áreas de cada número de conectividade (*AC<sub>i</sub>*), multiplicadas pelo seu respectivo número de conectividade, (*C<sub>i</sub>*, variando de 1 a 8, segundo o MAPIC), dividida pela área total da sub-bacia hidrográfica (*A\_subbacia*), conforme a Equação 5.

$$\text{Valor\_Conectividade} = \sum_{i=1}^8 \left( \frac{AC_i * C_i}{A\_subbacia} \right) \quad (5)$$

A conectividade corresponde a uma medida de complexidade da teia trófica e representa a proporção entre o número real de ligações existentes na teia nos diferentes níveis tróficos e o número total de ligações possíveis entre as espécies e o substrato que utilizam como alimento (KITCHING, 2000). O número de conectividade (*C<sub>i</sub>*) do MAPIC combina essas informações para diferentes espécies animais e vegetais.

Para a análise por *SOM*, os agrupamentos definidos são identificados de forma mais eficiente se em todas as variáveis a quantificação de grandeza de interesse ocorrer em um mesmo sentido. Sendo assim, foi estipulado que as áreas com prioridade para a recuperação florestal deveriam ter valores altos. Para tanto, foram invertidos os valores físicos da variável “%APP-Água\_VEGETAÇÃO”, fazendo com que os valores mais baixos passassem a ser os mais altos, indicando dessa forma as áreas com menor percentual de cobertura florestal como sendo justamente as de maior prioridade para a recuperação florestal. No caso da variável “Valor\_Conectividade”, existe uma relação de proporcionalidade direta e, assim, os valores de maior prioridade para recuperação florestal são os mais altos, pois estes correspondem às áreas com maior necessidade de conectividade da paisagem.

Além das variáveis já apresentadas, também foram utilizados na *SOM* as coordenadas UTM do centróide de cada sub-bacia hidrográfica. Este procedimento teve o intuito de contemplar a 1a. Lei da Geografia (TOBLER, 1970): “*todas as coisas são parecidas, mas coisas mais próximas se parecem mais que coisas mais distantes*”. Ou seja, os valores de coordenadas (latitude e longitude) passariam a ser treinados na *SOM* com a intenção de agregar vetores de variáveis considerando a proximidade entre os mesmos. Bação et al. (2004) discorrem detalhadamente sobre a importância de se utilizar a informação espacial na *SOM*. O conjunto das variáveis de entrada para essa primeira escala de análise, em nível regional, encontra-se sumariado na Tabela 1. Foram inseridos na *SOM*, portanto, quatro vetores (colunas), correspondendo a cada uma das quatro variáveis de entrada, contendo cada um 56 valores numéricos (linhas), correspondendo aos respectivos valores dessas variáveis para cada uma das 56 sub-bacias hidrográficas existentes na UGRHI-PBSUL, conforme indicado na Figura 2.

Tabela 1 – Conjunto de variáveis de entrada da *SOM*: escala regional.

<b>Código da Variável de Entrada na <i>SOM</i></b>	<b>Descrição da Variável (Unidade de Análise: Sub-Bacia Hidrográfica)</b>
%APP-Água_VEGETAÇÃO	Percentual da cobertura florestal
Valor_Conectividade	Valor de conectividade ponderado pela área
Long	Coordenada X do centróide da sub-bacia
Lat	Coordenada Y do centróide da sub-bacia

A *SOM* foi gerada por treinamento seqüencial no programa SOMToolbox (VESANTO et al., 2000) visando à definição da dimensão do Mapa de Kohonen (número de neurônios), tamanho de raio de vizinhança de inicialização e adaptação e taxa de aprendizagem em ambas as fases de ordenação e convergência. A normalização dos dados de entrada (amostras para cada variável) se deu entre 0 e 1. Os Planos de Componentes para cada uma das variáveis foram analisados visualmente, com a intenção de detectar possíveis redundâncias de informação

mediante uma comparação pareada (por pares) entre planos. Caso houvesse redundância, eliminar-se-ia uma das variáveis do par. Também foram avaliadas visualmente a Matriz-U, assim como a frequência de neurônios vencedores, em que é possível verificar o grau de aderência das variáveis ao resultado da *SOM*.

A definição de agrupamentos da *SOM* foi realizada através do método *k*-médias em todas as variáveis, inclusive as de coordenadas UTM. A escolha da quantidade de agrupamentos foi baseada no índice de Davies-Bouldin (DAVIES; BOULDIN, 1979). Para a identificação de sub-bacias prioritárias para a recuperação florestal, efetuou-se o somatório dos valores normalizados e ordenados resultantes das variáveis em cada agrupamento gerado pela *SOM*. Através de histogramas, foi possível ordenar os agrupamentos mais indicados para a recuperação florestal. Todos os agrupamentos foram relacionados aos polígonos de sub-bacias hidrográficas no mapa geográfico da UGRHI-PBSUL, o que permitiu a escolha da sub-bacia com maior prioridade para recuperação florestal, a ser estudada com maior profundidade na etapa subsequente.

#### 4.3.2 Escala sub-regional

Nesta escala de análise, foram consideradas somente as APP-Água sem ocorrência de vegetação, conforme o Inventário Florestal, e contidas dentro dos limites da sub-bacia hidrográfica com maior prioridade para recuperação florestal. Como unidade de análise, adotaram-se hexágonos regulares com raio de 30 m, dispostos de forma a preencher todo o limite interno dos polígonos irregulares que configuram as APP-Água, visto que o hexágono é um formato de unidade espacial que maximiza as relações de vizinhança. Consideraram-se apenas os hexágonos cujos centróides encontravam-se dentro dos limites das respectivas APP-Água, descartando-se os demais. Seis variáveis ambientais foram selecionadas para integrar a *SOM* nesta etapa, todas elas representadas em formato matricial, com pixels de 30 m de resolução. O valor de cada variável atribuído a um dado hexágono (unidade de análise espacial da *SOM*) corresponde ao valor encontrado no pixel da variável localizado correspondentemente sobre o centróide do hexágono, mediante a sobreposição dos *layers* de variáveis ao *layer* que contém as APP-Água.

As áreas prioritárias para a recuperação florestal nesta escala corresponderiam àquelas com as seguintes características:

- a) áreas que apresentassem menores distâncias a fragmentos florestais, visando tanto à formação de corredores ecológicos como à união destes com matrizes florestais mais consolidadas, denominadas de áreas-núcleo;
- b) áreas mais distantes a zonas urbanizadas consolidadas e a rodovias, ou seja, distantes de pressões antrópicas, sobretudo da influência da especulação imobiliária;
- c) áreas próximas a ocupações rurais de pequeno porte, devido ao fato de que, em projetos de recuperação florestal, existe a necessidade de que as comunidades de pequenos produtores estejam envolvidas no manejo das áreas a serem recuperadas, principalmente no que se refere à remoção de mato para permitir a rebrota de floresta;

d) áreas com menor contingente florestal, pois a prioridade se dá em áreas com formação subarbusativa a arbustiva;

e) áreas mais suscetíveis à perda de solo devido à erosão laminar, uma vez que ocorrendo erosão, a capacidade de recuperação ambiental diminui bruscamente, e além do mais, é precisamente nestas áreas que podem ocorrer os maiores deslizamentos de terra.

Sendo assim, foram gerados *layers* de variáveis que correspondessem às características citadas. Para a característica “a”, foi gerada uma grade de distância a partir dos fragmentos florestais do Mapa do Inventário Florestal (INSTITUTO FLORESTAL, 2005). Esta variável denominou-se “DIST\_FRAG”. Considerando ainda a característica “a”, da união de corredores ecológicos a matrizes florestais mais consolidadas, foi gerada outra grade de distância em relação a fragmentos maiores e com bordas menos perturbadas. Estes fragmentos foram delimitados a partir dos maiores percentis de área absoluta e de uma métrica da paisagem que considera o nível de arredondamento dos fragmentos (*shape*). A área média destes fragmentos foi de 700 ha. Esta variável foi denominada “DIST\_FRAGMAIOR”. Em relação à característica “b”, foi utilizada a imagem TM/Landsat 5, de 8 de setembro de 2004, para mapear manchas urbanas consolidadas com associação às rodovias pavimentadas, sendo gerada uma grade de distância para as mesmas. Esta variável denominou-se “DIST\_URB\_EST”. Para gerar uma variável correspondente à característica “c”, foram utilizados os pontos de edificações das cartas de 1:50.000 do IBGE. Apesar de serem dados da década de 1980, as áreas rurais sofrem pouca alteração em áreas mais distantes das aglomerações urbanas contínuas. Portanto, foram aproveitados apenas os pontos de construções com mais de 2,5 km de distância a estradas pavimentadas e a áreas urbanas contínuas, conforme a imagem TM/Landsat 5. A grade de distância desta informação gerou a variável “DIST\_CONSTRURAIIS”. A característica “d” foi representada pelo Índice de Vegetação por Diferença Normalizada (*Normalized Difference Vegetation Index - NDVI*), gerado a partir da razão da diferença entre as bandas do infravermelho próximo e vermelho (respectivamente, bandas 4e 3 da imagem TM/Landsat 5) pela soma das mesmas. Valores mais baixos deste índice corresponderiam a APPs com formações preponderantemente subarbusativas e pouco florestadas. O vetor de entrada desta variável correspondeu aos valores médios de NDVI extraídos para os hexágonos, considerando-se os pixels cujos centróides encontravam-se contidos no limite interno de cada hexágono

Para a característica “e”, foi necessário gerar uma variável da distribuição espacial do Potencial Natural de Erosão Laminar (PNEL). Esta variável demonstra as áreas mais críticas quanto à perda de solo em t/ha/ano. Este mapa foi gerado a partir de dados do mapa de solos de Oliveira et al. (1999), da rede de estações pluviométricas da Agência Nacional de Águas (ANA, 2009) e dos mapas vetoriais de altimetria e de cursos d’ água em escala 1:50.000 (IBGE, 1984). O mapa de solos, em escala 1:500.000, apenas detalhou duas grandes formações pedológicas: Latossolo Vermelho-Amarelo Distrófico e Argissolo Vermelho-Amarelo Distrófico.

Para a geração do PNEL, foram necessárias as transformações: i) do mapa de solos para o fator  $K$ , segundo valores de Machado et al. (2009); ii) dos dados de pluviometria para o fator  $R$ , segundo equação proposta por Lombardi Neto e Moldenhauer (1992); iii) geração do fator de comprimento de rampa (*Length of Slope – LS*) através de um Modelo Digital de Elevação (MDE). O MDE foi gerado a partir de carta topográfica em escala 1:50.000 (IBGE, 1984) pelo método de interpolação disponível no ArcGIS denominado TOPOGRID. Como é comum que valores de PNEL tenham valores extremos, e para que esta variável pudesse ser significativa na SOM, foi calculado o logaritmo na base 10 desta variável, a fim de que houvesse concentração de valores significativos. A extração do somatório destes valores dentro de cada hexágono constituiu a variável “PNEL\_log”. A Tabela 2 sintetiza o conjunto de variáveis de entrada para a escala de análise sub-regional, sendo as seis primeiras ambientais e as duas últimas de informação espacial.

Tabela 2 – Conjunto de variáveis de entrada da SOM: escala sub-regional.

<b>Código da Variável de Entrada na SOM</b>	<b>Descrição da Variável (Unidade de Análise: Hexágonos das APP-Água)</b>
DIST_FRAG	Distâncias a fragmentos florestais genéricos
DIST_FRAGMAIOR	Distâncias aos maiores fragmentos florestais
DIST_URB_EST	Distâncias a manchas urbanas consolidadas
DIST_CONSTRURAIAS	Distâncias a construções rurais situadas a mais de 2,5 km de estradas pavimentadas
NDVI_MEDIA	Índice de Vegetação por Diferença Normalizada
PNEL_LOG_SOMA	Somatório do Potencial Natural de Erosão Laminar em valores logarítmicos (base 10)
LONG	Coordenada X do centróide dos hexágonos
LAT	Coordenada Y do centróide dos hexágonos

Da mesma forma que na Seção 4.3.1, referente à escala regional, a quantificação de grandeza nas variáveis ocorre em diferentes sentidos, e, portanto, foi necessário aplicar o mesmo recurso de manipulação algébrica da etapa anterior, para que todas as variáveis tivessem os seus maiores valores relacionados a áreas com maior prioridade para a recuperação florestal. As etapas de geração e análise da SOM nesta seção são semelhantes à seção anterior. Nesta etapa, a fim de aprofundar a análise das áreas definidas como prioritárias para a recuperação florestal, foi apresentado sobre o Mapa de Kohonen a frequência de vezes em que os neurônios foram vencedores após 1.000 iterações. Esta frequência foi mapeada geograficamente pelas unidades de análise espacial (hexágonos), ou seja, foi apresentado o valor da frequência do neurônio vencedor pelo qual determinado vetor de entrada (amostra) contribuiu para que o mesmo fosse vencedor. Este dado informa a aderência das respectivas amostras ao Mapa de Kohonen e aos

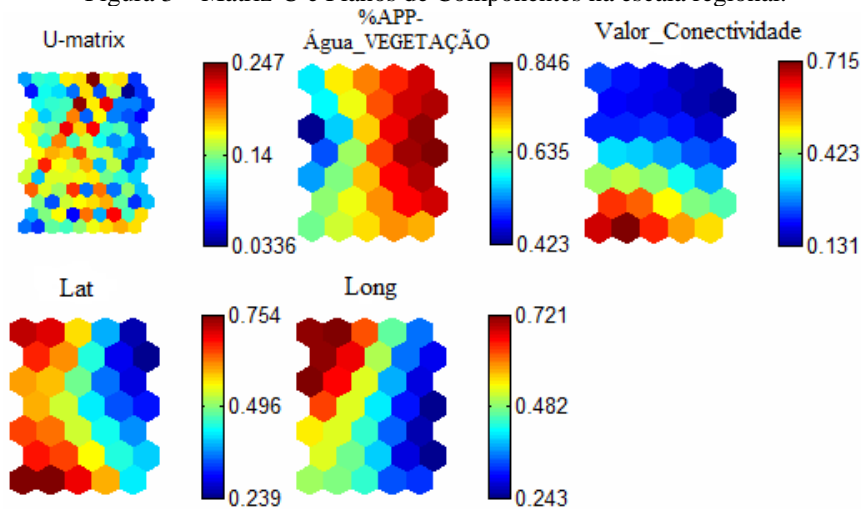
agrupamentos definidos na *SOM*, e, portanto, identifica as amostras mais precisas considerando as variáveis do estudo. Neste trabalho, este mapa foi denominado de Mapa de Contribuição.

## 5. RESULTADOS

### 5.1 Escala Regional

A *SOM* gerada na escala regional obteve erro de quantização médio de 0,219 e erro topológico de 0,036, que são considerados baixos em uma escala de 0 a 1 (VESANTO et al., 2000). O mapa de Kohonen resultou em 35 neurônios, sendo cinco dispostos em linha e sete em coluna. A Figura 3 apresenta a Matriz-U associada e os respectivos Planos de Componentes, nos quais neurônios com cores similares representam comportamento similar. Percebe-se que não houve redundância de informação entre as variáveis. A variável “%APP-Água\_VEGETAÇÃO” apresenta uma correlação inversa com o dado de latitude, ou seja, os valores mais baixos de latitude (norte) concentram as maiores taxas de APP preservada. Os maiores valores de conectividade apresentam-se concentrados para esta componente, visto que, geograficamente, as áreas com números de conectividade elevados estão compreendidas ao longo da Serra da Mantiqueira. A Matriz-U mostra-se confusa, dificultando a visualização de mais de dois aglomerados resultantes.

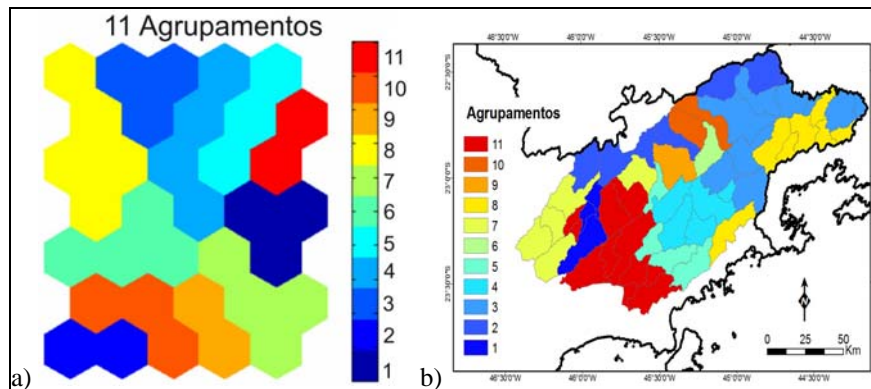
Figura 3 – Matriz-U e Planos de Componentes na escala regional.



A Figura 4 (a) apresenta os agrupamentos delimitados pelo k-médias no Mapa de Kohonen, cujo particionamento alcançou o valor mínimo de 0,61195 para o

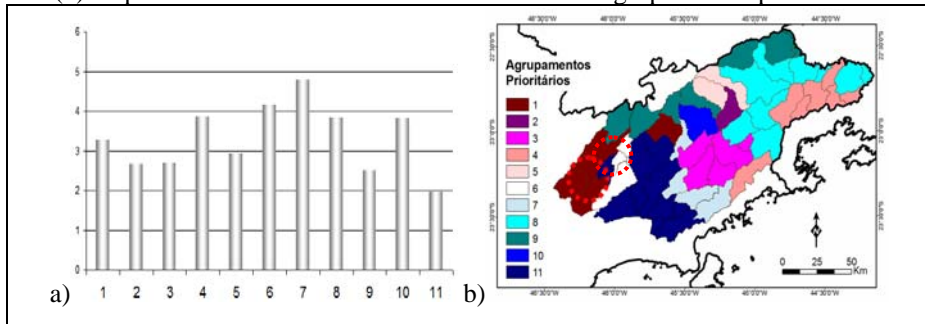
índice de Davies-Bouldin, acusando o número ideal de agrupamentos. Foram delimitados 11 agrupamentos, os quais são representados geograficamente na UGRHI-PBSUL (Figura 4 (b)). Percebe-se que muitos dos agrupamentos na UGRHI-PBSUL encontram-se adjacentes ou próximos, o que é decorrência direta da 1a. Lei da Geografia, conforme exposto na Seção 4.3.1.

Figura 4 – (a) Agrupamentos na escala regional no Mapa de Kohonen; (b) agrupamentos no mapa geográfico da UGRHI-PBSUL.



Conforme a metodologia de *SOM*, construiu-se um histograma da soma dos valores normalizados das componentes, com exceção das componentes de informação espacial (Figura 5 (a)). Percebe-se que o agrupamento 7 foi o que apresentou maiores valores resultantes desta soma, o que indica que é precisamente neste agrupamento onde se encontram as sub-bacias com maior prioridade para a recuperação florestal na UGRHI-PBSUL. É neste agrupamento que se localizam as sub-bacias hidrográficas com as maiores taxas de desflorestamento em APP de cursos d'água (APP-Água) e também aquelas que apresentam a maior necessidade de conectividade ecológica. O mapa da Figura 5 (b) mostra o ordenamento de sub-bacias com maior prioridade de recuperação florestal na UGRHI-PBSUL, tendo em vista o escalonamento de grandezas do histograma da Figura 5 (a). Optou-se por escolher somente uma sub-bacia dentre as mais prioritárias para o andamento do estudo, tendo em vista as considerações do MAPIC. O círculo em vermelho na Figura 5 (b) assinala a sub-bacia escolhida para investigação na escala sub-regional.

Figura 5 – (a) Histograma da soma dos valores normalizados das componentes; (b) mapa da UGRHI-PBSUL com ordenamento de agrupamentos prioritários.

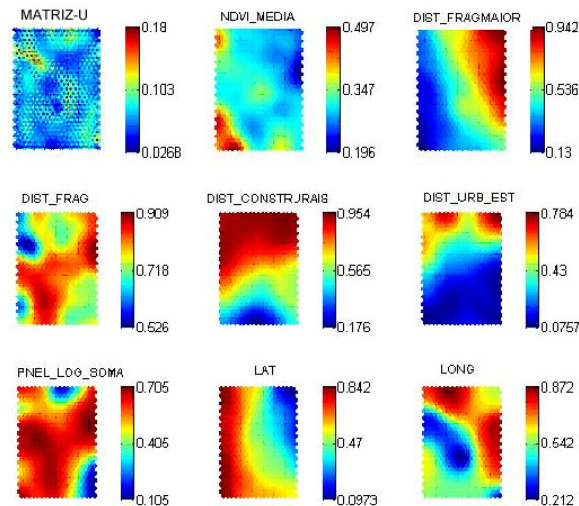


## 5.2 Escala Sub-Regional

A sub-bacia hidrográfica definida está situada na parte alta da represa Jaguari, município de Igaratá (Figura 5(b)), a qual é um manancial estratégico para SP, MG e RJ. Nesta sub-bacia, foram contabilizados 12.923 hexágonos (contidos nas APP-Água), a partir dos quais se extraíram valores para as variáveis da *SOM*.

A partir das rotinas disponíveis no *SOMToolbox*, as componentes da *SOM* foram articuladas em 570 neurônios, obtendo-se erro de quantização médio de 0,226 e erro topológico de 0,023, julgados admissíveis tendo em vista o intervalo de variação de 0 a 1 (VESANTO et al., 2000). A Figura 6 mostra a Matriz-U e os respectivos Planos de Componentes.

Figura 6 - Matriz-U e Planos de Componentes na escala sub-regional.





Percebe-se que não há redundância de informação entre as componentes. Apenas a componente de latitude (LAT) apresenta alguma relação inversa com a componente DIST\_FRAGMAIOR, uma vez que os fragmentos desta encontram-se praticamente todos ao norte da sub-bacia hidrográfica, onde o relevo é mais acidentado. A Matriz-U apresentou resultado satisfatório, pois nela é possível visualizar aglomerados com limites bem definidos.

A Figura 7 (a) mostra a Matriz-U com a sobreposição dos neurônios vencedores (em vermelho) e suas respectivas frequências (tamanho do hexágono). Percebe-se que, de maneira geral, os neurônios vencedores não estão compreendidos em meio a aglomerados, podendo-se observar apenas um aglomerado maior com neurônios vencedores no centro da Matriz-U. Os neurônios vencedores estão compreendidos majoritariamente nos limites da Matriz-U, cujos valores foram baixos. Pode-se inferir que este padrão seja resultante da fase de adaptação e cooperação da SOM, as quais teriam repellido estes neurônios em direção às extremidades. Este comportamento é pertinente, uma vez que o processamento da SOM tende a agrupar neurônios parecidos e afastar neurônios diferentes. Neurônios com menor vizinhança (de borda) tendem a ser um reduto desta especificidade ao longo do treinamento da SOM.

Na Figura 7 (b), estão representados os agrupamentos delimitados pelo k-médias. Foram delimitados 23 agrupamentos, com valor do índice de Davies-Bouldin estimado em 0,7851. Estes agrupamentos estão projetados na sub-bacia hidrográfica na Figura 9 (a). Foi verificado visualmente que os agrupamentos possuem coesão de área e não se apresentam de forma fragmentada. Apesar de haver mais variáveis do que na escala regional, o valor do índice de Davies-Bouldin na escala sub-regional foi apenas sutilmente pior (i.e., ligeiramente maior) em relação ao da escala regional, uma vez que o menor valor deste índice indica o número ideal de agrupamentos. Foi verificado, por meio da imagem TM/Landsat-5, que houve uma separação entre os agrupamentos de acordo com a paisagem, em que áreas de cabeceira de rios, áreas urbanas e entorno de barragens, por exemplo, apresentaram agrupamentos próprios. A Figura 8 mostra o histograma da soma dos valores normalizados de todas as componentes, com exceção daquelas referentes à informação espacial (latitude e longitude). Observa-se que os agrupamentos 3, 10 e 16 foram os que apresentaram maiores valores resultantes desta soma, e é justamente nestes agrupamentos onde se encontram as APP-ÁGUA com maior prioridade para a recuperação florestal nesta sub-bacia hidrográfica, e portanto, foram os escolhidos para formar um único agrupamento para o propósito deste estudo.

Figura 7 – (a) Frequência de neurônios vencedores (em vermelho) sobre Matriz-U na escala sub-regional; (b) agrupamentos no Mapa de Kohonen.

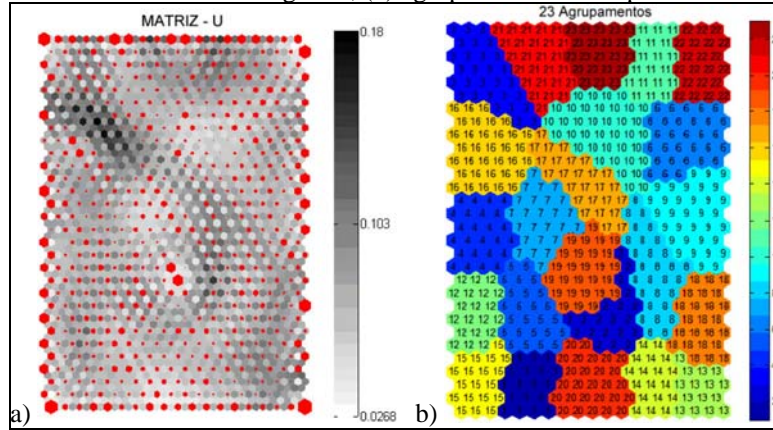
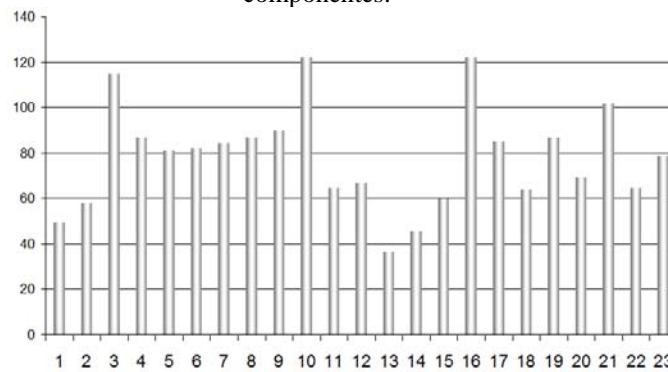


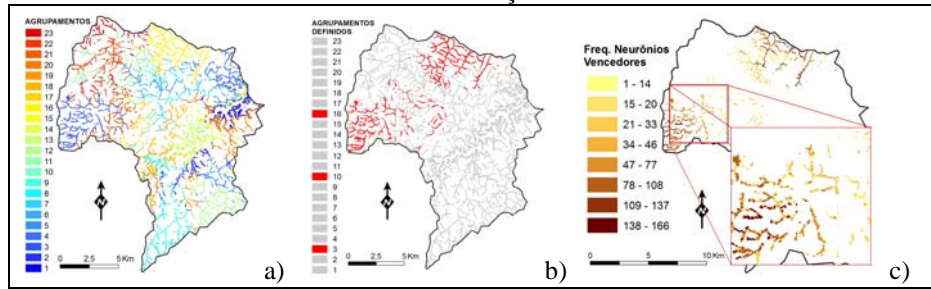
Figura 8 - Histograma da soma dos valores normalizados de todas as componentes.



A Figura 9 (a) apresenta todos os agrupamentos delimitados na sub-bacia hidrográfica, a Figura 9 (b) realça os agrupamentos selecionados prioritariamente para a recuperação florestal em vermelho, e a Figura 9 (c) apresenta o Mapa de Contribuição, em cujo recorte de detalhamento é possível visualizar os hexágonos regulares ao longo das APAs hidrográficas, utilizados como unidades de análise espacial. Neste mapa, é possível notar que a maioria das amostras que contribuiram para neurônios vencedores a oeste constitui amostras de cabeceira, enquanto que na porção leste, estas amostras estão localizadas em áreas de rampas. Esta constatação em ambas as porções indica que os respectivos agrupamentos foram definidos com base em variáveis de alto valor nestas áreas de cabeceira, o que corresponde às

variáveis DIST\_FRAG e PNEL\_LOG\_SOMA, sendo, portanto, mais fidedignas à análise em questão.

Figura 9 - (a) Agrupamentos delimitados na sub-bacia hidrográfica; (b) agrupamentos definidos como prioritários para a recuperação florestal; (c) Mapa de Contribuição.



## 6. CONCLUSÕES E CONSIDERAÇÕES FINAIS

A abordagem metodológica adotada empregando a *SOM* foi eficiente para a identificação do comportamento das variáveis e da redundância de informação entre as mesmas. A *SOM* mostrou-se eficaz na definição de áreas prioritárias para a recuperação florestal. Para tanto, alguns procedimentos de pré-processamento das variáveis de entrada na *SOM* foram essenciais, tais como: normalização, transformação e inversão de valores. Ou seja, a normalização de todas as variáveis entre 0 e 1; a eliminação de valores extremos, como no caso da variável PNEL; e a inversão de valores, de forma que todos os gradientes das variáveis fossem diretamente proporcionais. Estes procedimentos foram decisivos para que a *SOM* obtivesse resultados coerentes.

O uso da Matriz-U, isoladamente, mostrou-se não ser suficiente para a identificação visual de agrupamentos e de variáveis mais adequadas na escala regional. Para contornar este inconveniente, existem ferramentas alternativas de geovisualização no SOMToolbox que são mais eficientes para este objetivo específico, as quais se pretende explorar em trabalhos posteriores.

O Mapa de Contribuição apresentou importância para a identificação de variáveis com maior aderência ao fenômeno estudado, haja vista a constatação da localização dos maiores valores de frequência de neurônios associados aos maiores valores de variáveis que induziriam à recuperação florestal. Além disso, este tipo de dado demonstra que dentro de cada agrupamento existem amostras que são mais fidedignas ao respectivo agrupamento, por apresentarem uma maior frequência de neurônios vencedores.

A escala de análise regional permitiu a escolha da sub-bacia situada na parte alta da represa Jaguari, município de Igaratá, como prioritária para a recuperação

florestal. Esta sub-bacia possui papel estratégico no abastecimento de água da região, pois se localiza dentro da área de contribuição de um manancial que abastece diversas cidades dos estados de São Paulo, Minas Gerais e Rio de Janeiro, inclusive a capital deste último. Neste sentido, é imprescindível que a vegetação florestal em APPs hidrográficas nesta área seja recuperada, o que corresponde exatamente ao objeto de estudo na escala de análise sub-regional, a fim de conter o assoreamento dos corpos d'água, e assim, evitar o comprometimento da qualidade de água desta importante represa.

Estudos futuros deverão ser conduzidos, considerando a temporalidade das variáveis, além de contemplarem análises mais aprofundadas e a quantificação dos dados gerados, sempre com a respectiva projeção dos resultados em mapas geográficos.

## 7. REFERÊNCIAS BIBLIOGRÁFICAS

- AGÊNCIA NACIONAL DE ÁGUAS. *HidroWeb: Sistema de Informações Hidrológicas*. 2009. Disponível em <<http://hidroweb.ana.gov.br/>>. Acesso em: 07 ago. 2009.
- BAÇÃO, F.; LOBO, V.; PAINHO, M. Geo-Self-Organizing Map (Geo-SOM) for building and exploring homogeneous regions. In: EGENHOFER, M.; MILLER, H.; FREKSA, C. (Org.). *GIScience 2004, Lecture Notes in Computer Science*, Berlin: Springer, 2004. p. 22-37.
- BOSCHI, L. S.; GALO, M. L. B. T. O uso de redes neurais artificiais na análise da ambigüidade entre classes de água e plantas aquáticas. *Boletim de Ciências Geodésicas*, v. 10, n. 2, p. 193-212, 2004.
- BRASIL. *Resolução CONAMA n. 303, de 20 de março de 2002*. Dispõe sobre parâmetros, definições e limites de Áreas de Preservação Permanente. Brasília: Diário Oficial da República Federativa do Brasil, 13 de maio de 2002.
- BURNS, I. S.; SCOTT, S.; LEVICK, L.; HERNANDEZ, M.; GOODRICH, D. C.; SEMMENS, D. J.; KEPNER, W. G.; MILLER, S. N. *Automated Geospatial Watershed Assessment (AGWA) - A GIS-Based Hydrologic Modeling Tool: Documentation and User Manual*. 2004. Disponível em <[http://www.epa.gov/esd/land-sci/agwa/pdf/user\\_manual.pdf](http://www.epa.gov/esd/land-sci/agwa/pdf/user_manual.pdf)>. Acesso em: 19 nov. 2009.
- CONSERVAÇÃO INTERNACIONAL. *Sobre o bioma da mata atlântica*. Aliança para a Conservação da Mata Atlântica. São Paulo, 2008. Disponível em <<http://www.aliancamataatlantica.org.br>>. Acesso em: 10 dez. 2009
- DAVIES, D. L. BOULDIN, D. W. A Cluster Separation Measure. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, v. 1, n. 2, p. 224-227, 1979.
- FERREIRA, M. S.; GALO, M. L. B. T. Influência dos dados de entrada na classificação a partir de rede neural artificial. In: III Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação – III SIMGEO. *Anais ...*, 27-30 julho, Recife, PE. Recife: UFPE, 2010.

- GIRAUDEL, J. L.; LEK, S. A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. *Ecological Modelling*, v. 146, n. 1-3, p. 329-339, 2001.
- GOOGLE EARTH. *Imagem do Município de Igaratá*. 2008. Disponível em <<http://www.google.com/earth/index.html>>. Acesso em: 15 ago. 2009.
- HAYKIN, S. *Redes neurais: princípios e prática*. 2 ed. Porto Alegre: Bookman, 2001.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Igaratá: SF-23-Y-D-I-2 MI 2768.2*. Rio de Janeiro, 1984. Carta Topográfica. Escala: 1:50.000.
- INSTITUTO FLORESTAL. *Inventário Florestal da Vegetação Natural do Estado de São Paulo*. São Paulo: Imprensa Oficial, 2005. 200p.
- INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). *Landsat 5 – TM – WRS 219/76 – Data de aquisição: 08/09/04*. São José dos Campos: INPE, 2004. Uma imagem de satélite. 368 MB. GeoTiff. CD-ROM.
- KITCHING, R. L. *Food webs and container habitats: the natural history and ecology of phytotelmata*. Cambridge: Cambridge University Press, 2000. 431 p.
- KOHONEN, T. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, v. 43, p.59-69, 1982.
- KOHONEN T.; KASKI S.; LAGUS K.; SALOJÄRVI J.; HONKELA J.; PAATERO V.; SAARELA A. Self organization of a massive document collection. *IEEE Transactions on Neural Networks*, v. 11, n. 3, p. 574-585, 2000.
- LEITE, A. Y. C. M.; MATOS, L. N.; SILVA, M. A. S. Utilização de rede neural som para visualização e análise de dados censitários. In: 58. Reunião Anual da Sociedade Brasileira para o Progresso da Ciência – 58. SBPC. *Anais ...*, 16-21 julho, Florianópolis, SC. Florianópolis: SBPC, 2006.
- LIPPITT, C. D.; ROGAN, J.; LI, Z.; EASTMAN, J. R.; JONES, T. G. Mapping selective logging in mixed deciduous forest: a comparison of machine learning algorithms. *Photogrammetric Engineering & Remote Sensing*, v. 74, n. 10, p. 1201-1211, 2008.
- LOMBARDI NETO, F.; MOLDENHAUER, W. C. Erosividade da chuva: sua distribuição e relação com perdas de solo em Campinas, SP. *Bragantia*, v. 51, n.2, p.189-196, 1992.
- MACHADO, M. L.; ALVES, J. S.; GOMES, I.; VIEIRA, E. M.; SIMÃO, M. L. R.; NAIME, U. J. Levantamento sistemático dos fatores da Equação Universal de Perdas de Solos (USLE) para o delineamento de áreas com potencial erosivo da bacia PN1 - IGAM, Minas Gerais (resultados parciais). In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14. (SBSR), 2009, Natal. *Anais...* São José dos Campos: INPE, 2009. p. 7733-7739. DVD, On-line. ISBN 978-85-17-00044-7.
- MINISTÉRIO DAS MINAS E ENERGIA (MME). *Projeto RADAMBRASIL: mapa exploratório de solos*. Brasília: MME, 1977.

- MYERS N., MITTERMEIER R. A., MITTERMEIER C. G., DA FONSECA G. A. B., KENT J. Biodiversity hotspots for conservation priorities. *Nature*, v. 403, Feb 2000, pp. 853-858.
- OLIVEIRA, J. B.; CAMARGO, M. N.; ROSSI, M; CALDERANO FILHO, B. *Mapa pedológico do estado de São Paulo: legenda expandida*. Campinas. 1999. 64p. Inclui mapas.
- PROGRAMA BIOTA/FAPESP. *Mapa de Áreas Prioritárias para Incremento da Conectividade do Estado de São Paulo*. São Paulo, SP: Programa BIOTA/FAPESP, Instituto Florestal/SMA, Centro de Referência em Informação Ambiental, 2007. Disponível em <<http://homologa.ambiente.sp.gov.br/21projetos/conectividade.pdf>>. Acesso em: 31 jul. 2009.
- RODRIGUES, R. R.; LIMA, R. A. F.; GANDOLFI, S.; NAVE, A. G. On the restoration of high diversity forests: 30 years of experience in the Brazilian Atlantic Forest. *Biological Conservation*, v. 142, pp. 1242-1251, 2009.
- SÃO PAULO. GOVERNO DO ESTADO. *Resolução SMA-SP 85/2008*. Dispõe sobre os critérios e parâmetros para compensação ambiental de áreas objeto de pedido de autorização para supressão de vegetação nativa no estado de São Paulo. São Paulo: Diário Oficial do Estado, 12 de dezembro de 2008.
- SELAMAT, A.; SELAMAT, M. H. Clustering location of forest fires using self-organizing maps. *Brunei Darussalam Journal of Technology and Commerce*, v. 4, n. 1, p. 113-120, 2006.
- SOS MATA ATLÂNTICA. *Nossos Projetos*. 2009. Disponível em <<http://www.sosmatatlantica.org.br/index.php?section=project&action=listProjects>>. Acesso em: 18 set. 2009.
- TOBLER, W. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, v. 46, p. 234-240, 1970
- ULTSCH, A. Self-organizing neural networks perform different from statistical kmeans clustering. In: Annual Conference of the German Classification Society, n. 19, 1996, [S.l.]. *Proceedings...* [S.l.]: H. H. Bock and W. Polasek, 1996. p. 1-13.
- UNIVERSIDADE DE SÃO PAULO (USP). *Pacto pela restauração da mata atlântica: referencial dos conceitos e ações de restauração florestal*. 1 Ed. São Paulo, Neoband, 2009. 258 p. Disponível em <<http://www.lerf.esalq.usp.br/divulgacao/produzidos/livros/pacto2009.pdf>>. Acesso em: 15 nov. 2009.
- VESANTO, J.; HIMBERG, J.; ALHONIEMI, E.; PARHANKANGAS, J. *SOM toolbox for Matlab 5*. Helsinki University of Technology, 2000. p. 60. Disponível em <<http://www.cis.hut.fi/projects/somtoolbox/package/papers/techrep.pdf>>. Acesso em: 03 dez. 2009.
- WALKER, R. Theorizing land-cover and land-use change: the case of tropical deforestation. *International Regional Science Reviews*. v. 24, n. 3, pp. 247-270, 2004.

Recebido em agosto de 2010. Aceito em setembro de 2011.