

Identificação de cenários de concentração de poluentes no município do Rio de Janeiro utilizando redes neurais

Use of neural networks in the identification of pollutant concentration scenarios in the city of Rio de Janeiro

Flávia Ribeiro Villela*
Marcos Antonio Cruz Moreira**

Estudos realizados em diversos países têm relatado associação entre a poluição atmosférica e vários indicadores de morbidade e mortalidade, mesmo quando as concentrações de poluentes estão abaixo dos limites padronizados. Este trabalho tem como local de estudo o município do Rio de Janeiro e tem como objetivo identificar cenários locais e globais caracterizados por dias de alta ou baixa poluição e dias meteorológicos típicos ou atípicos. Busca também verificar as associações com a distribuição estatística das contagens de eventos de saúde. O método usado na identificação desses cenários foram os mapas topológicos de Kohonen, baseados em redes neurais.

Palavras-chave: Poluição do ar. Mapas de Kohonen. Redes Neurais.

Studies carried out in several countries have reported an association between air pollution and several indicators of morbidity and mortality, even when pollutant concentrations are below standard limits. This work has as geographic place of study the city of Rio de Janeiro, and aims to identify local and global scenarios characterized by high or low pollution days and typical or atypical weather days. It also verifies the associations with the statistical distribution of health event counts. The method used to identify these scenarios was Kohonen topological maps based on neural networks.

Keywords: Air Pollution. Kohonen Maps. Neural Network.

1 Introdução

A preocupação com os efeitos da poluição atmosférica não é recente, tendo inspirado políticas oficiais de controle já na década de 1950. Naquela época, na Europa, era grande a concentração de indústrias nas grandes cidades e a combustão de carvão era largamente utilizada para aquecimento dos domicílios. No episódio que se tornou conhecido como o grande *fog* londrino de 1952, durante uma inversão térmica, o dramático aumento da poluição foi acompanhado de grandes incrementos na mortalidade e na morbidade, principalmente por doenças respiratórias e cardíacas, sobretudo em idosos (ANDERSON et al., 1995; POLONIECKI et al., 1997).

* Doutoranda em Engenharia de Reservatório e de Exploração (UENF). Professora da Universidade Federal do Rio de Janeiro (UFRJ) - Macaé (RJ) - Brasil. E-mail: flaviavillela@macae.ufrj.br.

** Doutor em Engenharia Elétrica (UFRJ). Professor do Instituto Federal de Educação, Ciência e Tecnologia Fluminense (IFFluminense, campus Macaé) - Macaé (RJ) - Brasil. E-mail: macruz@iff.edu.br.

Apesar dos esforços para reduzir o impacto da poluição urbana na saúde, diversos estudos epidemiológicos continuaram reportando associação significativa entre os níveis de poluição e o número de internações hospitalares por doenças respiratórias e cardiovasculares.

O município do Rio de Janeiro possui áreas poluídas devido à atividade industrial e ao tráfego intenso de caminhões e outros veículos, mas também possui áreas litorâneas e rurais quase livres de fontes de poluição. Além disso, diferentes cenários de poluição atmosférica podem emergir devido aos maciços existentes na região (DAUMAS et al., 2004).

Partindo dessa premissa, o objetivo deste trabalho consiste em desenvolver um estudo para o município do Rio de Janeiro com a finalidade de identificar cenários locais e globais caracterizados por dias de alta ou baixa poluição e dias meteorológicos típicos ou atípicos, bem como verificar as associações entre a distribuição desses dias e a dos dias cujas contagens de eventos de saúde excedem as esperadas para aquela época do ano e dia da semana. Na identificação desses cenários foram utilizados os mapas topológicos de Kohonen.

2 Metodologia

Foram utilizadas três variáveis de entrada: material particulado com diâmetro aerodinâmico de até $10 \mu\text{g}/\text{m}^3$ (PM_{10}), temperatura média e umidade relativa do ar.

Neste estudo foram utilizados dados sobre poluição por Partículas Inaláveis (PM_{10}) mensurados a cada seis dias e dados meteorológicos de temperatura e umidade relativa do ar. Os dados referentes às internações hospitalares foram coletados diretamente de bancos de dados informatizados, disponibilizados pelo Ministério da Saúde para os hospitais conveniados ao Sistema Único de Saúde (SUS). Esses bancos de dados contêm informações de todas as internações realizadas no âmbito do SUS através das Autorizações de Internação Hospitalar (AIH). Para análise foram selecionadas as internações hospitalares de residentes no município do Rio de Janeiro por doenças do aparelho circulatório (DAC) e internações hospitalares de idosos maiores de 65 anos por doenças do aparelho respiratório (DAR65).

Para melhor entendimento das relações entre as variáveis de poluição e meteorologia, foram calculados os coeficientes de correlação linear de Pearson para esses dados no período total de estudo (Tabela 1).

Tabela 1: Padrão de correlação linear entre os dados de poluição do ar e meteorologia. Correlação de Pearson e p-valor

	PM_{10}	TEMP	UMID
PM_{10}	1	0.198	-0.270
		0.000	0.000

Fonte: Autoria própria (2016)

Note que as correlações entre umidade e os poluentes são negativas e estatisticamente significativas, mas de forma geral são baixas. Já a temperatura é correlacionada positivamente com os poluentes.

3 Mapas Auto-Organizáveis de Kohonen

O Mapa Auto-Organizável de Kohonen (SOM) pertence à classe das redes neurais sem treinamento supervisionado e foi desenvolvido tendo como modelo o processo biológico de reconhecimento de padrões visuais (KOHONEN, 2001).

O SOM converte relações estatísticas complexas e não lineares entre dados multidimensionais em relações geométricas simples, numa grade regular de representação de baixas dimensões. Como ocorre uma compressão da informação nesse processo, preservando-se, entretanto, as mais importantes relações topológicas e métricas entre os dados originais na grade de representação, é possível se produzir representações abstratas dos itens de dados (KOHONEN, 2001). Estes dois aspectos, visualização e abstração, podem ser utilizados de várias maneiras em aplicações complexas tais como diagramas de estado de processos e máquinas, gráficos de similaridade de tabelas estatísticas de aplicações de *Data Mining* etc.

O SOM consiste, usualmente, numa grade regular bidimensional de nós. Durante o processamento do SOM, o modelo de cada observação analisada é associado a cada nó. O algoritmo SOM computa os modelos de forma que eles representem de maneira ótima o domínio (discreto ou contínuo) das observações (vetores de entrada). Os modelos são organizados numa ordem bidimensional na qual modelos semelhantes estão mais próximos uns dos outros do que de modelos dissimilares. Nesse aspecto, o SOM é um gráfico de similaridades e também um diagrama de agrupamentos.

O algoritmo de auto-organização é composto por quatro etapas:

- Inicialização do mapa;
- Processo competitivo;
- Processo cooperativo;
- Adaptação sináptica.

Na inicialização do mapa, atribui-se um vetor de pesos aleatórios iniciais às conexões entre os neurônios das camadas de entrada e saída. Todavia, nem sempre tal procedimento produz o melhor resultado, tendo em vista que alguns vetores de pesos podem assumir valores muito diferentes dos padrões de treinamento. Dessa forma, alguns nodos da rede tornam-se inutilizáveis, pois poderão nunca representar um padrão do conjunto de exemplos. Uma alternativa é usar amostras extraídas aleatoriamente do conjunto de exemplos como atribuição inicial aos pesos.

No processo competitivo, os neurônios competem entre si para definir o mais semelhante ao padrão apresentado, sendo que apenas um neurônio será o vencedor. Entretanto não existem conexões físicas entre eles. Os neurônios são localmente interconectados por uma relação de

vizinhança, determinando a topologia do mapa (linear, em anel, hexagonal e retangular).

A determinação do neurônio vencedor em uma rede SOM pode ser feita verificando qual neurônio possui a menor distância euclidiana em relação ao padrão apresentado. Essa distância pode ser calculada da seguinte forma:

$$\text{Distância: } \sum_{i=1}^n \|x_i - w_{ji}\|$$

onde X_i é a i -ésima entrada do padrão apresentado à rede e W_{ji} é o peso do neurônio (referente à entrada).

No processo cooperativo, o neurônio vencedor do processo competitivo tende a influenciar o estado dos neurônios vizinhos. Assim, a vizinhança é inicializada com um valor de largura d_0 e vai diminuindo a cada iteração, segundo a relação:

$$d_t = d_0 \left(1 - \frac{t}{T}\right)$$

onde d_t é a largura da vizinhança para a iteração t atual e T é o número total de iterações.

O neurônio na posição (x, y) da matriz de saída será considerado como pertencente à vizinhança de x_v, y_v , se:

$$(v - d_t) < x < (v + d_t) \text{ e } (v - d_t) < y < (v + d_t)$$

Cada neurônio tem pesos sinápticos correspondentes a cada uma das entradas. Se forem consideradas três entradas de cores R, G e B (*red, green e blue*), por exemplo, haverá um vetor de pesos com três componentes, como pode ser visto na Figura 1.

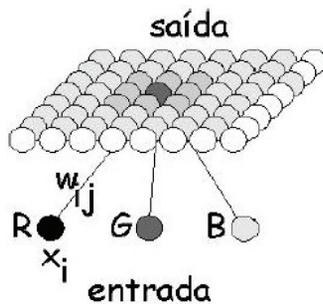


Figura 1: Pesos associados em uma rede SOM

Na etapa de adaptação sináptica, os pesos do neurônio vencedor e seus vizinhos são atualizados de modo a se aproximarem espacialmente do padrão de entrada.

Os elementos a serem agrupados são apresentados, um por vez, aos neurônios de entrada. A cada apresentação, os estímulos gerados pelo elemento são capturados pela camada de entrada e transmitidos igualmente a todos os neurônios da camada do mapa. O neurônio que reagir mais

fortemente aos estímulos do elemento apresentado ganha-o para si.

A cada nova apresentação de um elemento ao mapa, o perfil de sensibilidade dos neurônios vai se alterando e isso é chamado de treinamento.

4 Resultados

Os Mapas Auto-Organizáveis de Kohonen foram utilizados para identificação dos vários cenários anuais bem como do cenário geral de poluição atmosférica. Todos os procedimentos foram realizados utilizando o *software* R. O pacote kohonen R visa fornecer funções para mapas auto-organizáveis simples de usar, com especial ênfase na visualização. Neste trabalho a função *som* foi utilizada para definir o mapeamento dos objetos no treinamento definido para as unidades do mapa (RON; LUTGARDE, 2007; YAN, 2004).

Para a formação dos grupos, foram definidos mapas com dimensões 3x3 neurônios (para cada ano) e 5x3 neurônios (todos os anos), organizados em forma hexagonal, função de distância euclidiana, taxa de aprendizagem (α) para iniciar a partir de 0,05 e para parar a 0,01 durante as atualizações do número de iterações e decrescimento linear. O treinamento decorreu durante 2000 épocas e o tamanho da vizinhança foi escolhido de forma a que dois terços de todas as distâncias do mapa caíssem dentro desse número; depois de um terço das iterações apenas as unidades vencedoras são adaptadas.

Com diversas tentativas, procurou-se encontrar os parâmetros da função *som* necessários que permitissem definir os *clusters* adequadamente de modo que a rede convergisse. Além disso, a formação dos grupos com a topologia utilizada revelou padrões interessantes que serão apresentados posteriormente.

Elaborou-se um diagrama (Figura 2) com o mapeamento. O gráfico do canto superior esquerdo apresenta o mapa dividido nos 15 grupos. Em cada grupo foi elaborado um gráfico de setor, indicando a contribuição de cada variável em estudo. No gráfico “cenários de poluição do ar”, a similaridade média dos objetos mapeados para a unidade vencedora pode ser visualizada. O gráfico de contagens mostra o número de vezes que o neurônio vencedor foi ativado em cada grupo. E finalmente o gráfico inferior direito mostra onde cada objeto é mapeado; os rótulos indicam quais os dias que pertencem a cada grupo.

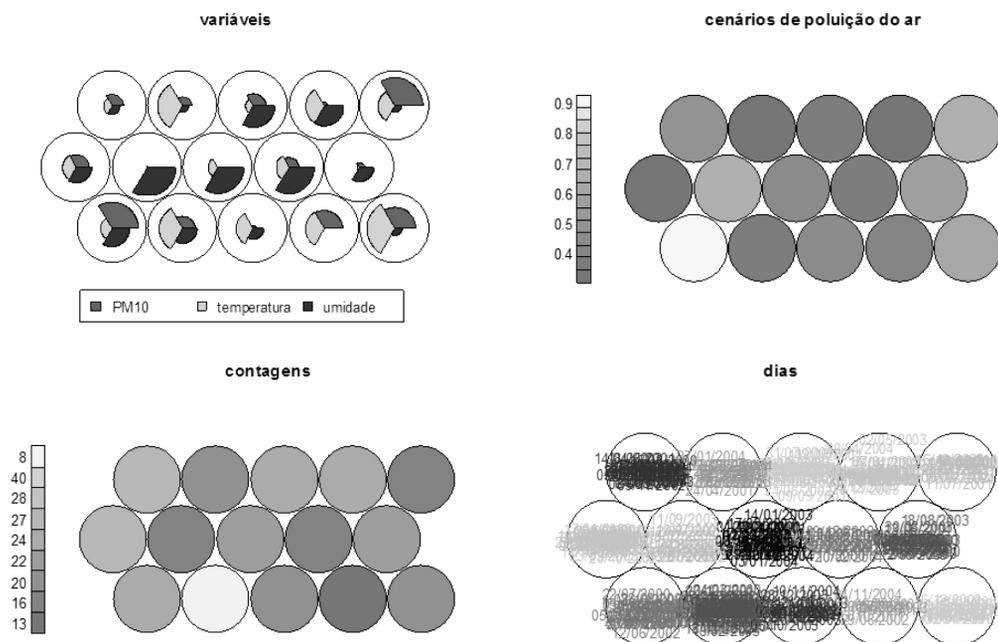


Figura 2: Cenários de poluição do ar

Fonte: Autoria própria (2016)

| 72 |

Os perfis dos grupos foram identificados de acordo com as médias das variáveis meteorológicas e de poluição do ar em estudo (Tabela 2).

Tabela 2: Perfil dos cenários de poluição (por grupo), segundo médias

Grupo	N	PM ₁₀	TEMP	UMID
1	16	57,59	28,41	69,89
2	27	65,55	24,43	79,12
3	24	64,70	21,96	84,57
4	24	44,89	26,06	82,90
5	13	101,47	25,17	70,03
6	28	63,02	22,38	72,60
7	13	36,45	19,08	90,88
8	22	35,36	22,87	88,50
9	13	58,83	24,11	88,87
10	20	48,08	20,16	77,86
11	24	97,77	23,31	80,77
12	40	65,07	27,84	77,85
13	16	41,81	25,58	74,17
14	16	84,10	29,74	69,86
15	8	77,44	26,21	64,32

Fonte: Autoria própria (2016)

Note que os grupos 14 e 15 apresentaram valores mais altos de poluição do ar com temperaturas altas e umidades mais baixas. Já os grupos 5 e 11 apresentaram valores mais altos de poluição do ar com umidades altas e temperaturas mais baixas. E o terceiro padrão apresenta um comportamento mais confuso: poluição do ar alto e/ ou baixo e temperatura e umidade altas.

Para descrever os padrões revelados por esses agrupamentos, foram realizadas análises dos padrões (Tabela 3) baseadas na contribuição das variáveis de estudo nos grupos que apresentaram padrão de poluição atmosférica alto.

Tabela 3: Padrões revelados

	Poluição	Umidade	Temperatura
Padrão 1	↑	↓	↑
Padrão 2	↑	↑	↓
Padrão 3	↓ ↑	↑	↑

Fonte: Autoria própria (2016)

Deve-se destacar que a análise de agrupamentos foi realizada tanto utilizando todos os anos, quanto para cada ano separadamente (5 análises de agrupamentos, portanto, foram realizadas). Dessa forma, foi possível verificar o efeito de cenários atípicos de poluição do ar na saúde tanto anuais como em geral (todos os anos).

Para estudar o comportamento dos grupos de dias encontrados na análise de agrupamentos em relação à poluição do ar e associação desta com a saúde, foram feitas comparações entre as médias de contagens de eventos de saúde e concentrações médias dos níveis de PM_{10} dos dias que formaram esses grupos. A fim de verificar o efeito após um dia de um aumento significativo de poluição do ar na saúde, foram consideradas também séries com contagem de internações hospitalares (AIH) dos dias consecutivos àqueles em que foram mensurados os níveis de poluição ($t+1$), conforme Tabela 4.

Tabela 4: Média de contagens de AIH, segundo grupo – todos os anos

Grupo	DAR65		DAC	
	t	t+1	t	t+1
1	10	10	44	40
15	9	7	45	44
3	9	8	46	41
4	9	13	48	52
5	11	9	54	49
6	9	9	42	41
7	8	8	36	43
8	8	8	36	45
9	10	9	45	38
10	9	10	47	48
11	10	10	54	48
12	10	9	43	49
13	7	9	31	42
14	11	10	47	47
15	8	9	38	38

Fonte: Autoria própria (2016).

| 74 |

Em geral, note que os grupos de menor tamanho e com as maiores médias dos níveis de poluição atmosférica (Tabela 2) praticamente apresentaram as maiores médias de internações hospitalares, tanto nos dias correntes (t) pertencentes aos respectivos grupos de dias considerados poluídos, como no dia seguinte passados esses dias (t+1).

Em seguida, para verificar essa associação foram calculados coeficientes de correlação linear de Pearson (Tabela 5), ponderados pelos tamanhos dos grupos, das médias de PM_{10} com as médias dos dados diários de contagem de internações hospitalares por causas específicas na cidade do Rio de Janeiro para o mesmo dia e com um dia de defasagem (t+1).

**Tabela 5: Padrão de correlação (ponderado pelos tamanhos dos grupos).
Coeficiente de Correlação de Pearson**

	Ano	DAR 65		DAC	
		t	t+1	t	t+1
PM ₁₀	1	0,656	0,161	0,605	0,090
	2	0,812	-0,144	0,945	0,183
	3	0,053	0,197	0,683	0,477
	4	0,655	0,133	0,972	0,799
	5	0,505	0,015	0,537	-0,202
	todos os anos	0,861	0,317	0,885	0,484

Fonte: Autoria própria (2016).

Note que as correlações entre as médias de internações hospitalares de doenças do aparelho respiratório em idosos com as médias de PM_{10} são positivamente altas quando não há defasagem de tempo (t). Observe também que as correlações das médias de PM_{10} com as médias de internações hospitalares por doenças do aparelho circulatório para o ano 4 tanto sem defasagem de tempo (t) como com defasagem de tempo (t+1) foram altas.

Foram construídos gráficos de dispersão das médias de material particulado contra as médias de internações hospitalares dos grupos encontrados na análise de agrupamentos (Figura 3), a fim de verificar se os grupos com dias mais poluídos tinham médias de internação hospitalares maiores.

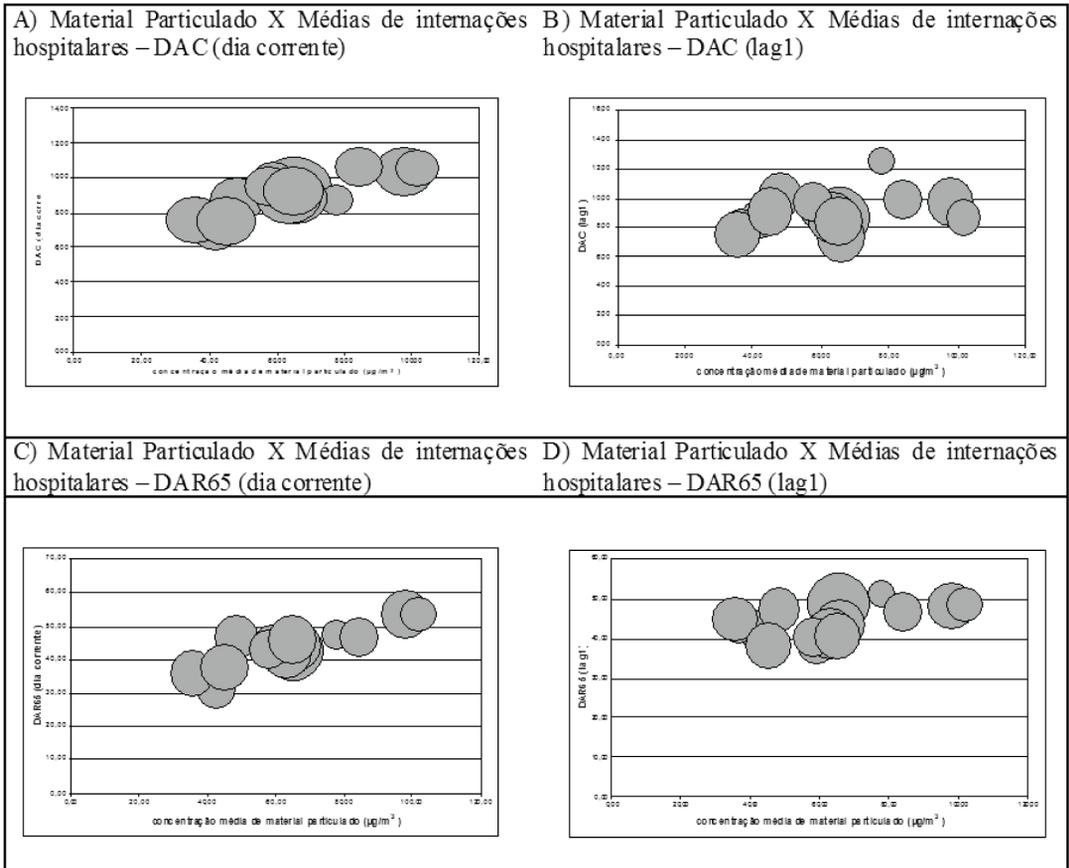


Figura 3: Médias dos escores dos grupos no componente x Médias de internações hospitalares dos grupos.

Fonte: Autoria própria (2016)

Note que os menores grupos com as maiores médias de níveis de poluição atmosférica (dias atípicos) praticamente apresentaram as maiores medianas de internações hospitalares por causas e faixas etárias específicas tanto nos dias correntes (t) pertencentes aos respectivos grupos de dias considerados poluídos como no dia seguinte passados esses dias (t+1).

5 Conclusões

Os Mapas Auto-Organizáveis de Kohonen são um método de classificação bastante eficiente na identificação de grupos, pois possuem a vantagem de uma visualização simples e rápida dos agrupamentos formados. Neste caso, pôde-se, com a ajuda do Mapa, identificar cenários de concentração de poluentes no município do Rio de Janeiro e pôde-se constatar se realmente os dias de alta poluição atmosférica correspondiam a aumentos de internações hospitalares por causas específicas.

Na formação dos grupos foram revelados padrões interessantes e satisfatórios ao analisarmos a contribuição das variáveis em estudo nos grupos que apresentaram padrão de poluição atmosférica alto, tornando-se, portanto, mais “fácil” e menos subjetiva a interpretação dos dados meteorológicos e de poluição do ar.

Além disso, observou-se que os dias tendem a se agrupar em massa (dias típicos) ou em pequenos grupos (dias atípicos).

Observou-se para o município do Rio de Janeiro, no período de estudo como um todo, uma associação positiva entre as médias das admissões hospitalares de doenças do aparelho circulatório e respiratório em idosos e as médias dos níveis de poluição atmosférica (PM₁₀).

Referências

ANDERSON, H. R.; LIMB, E. S.; BLAND, J. M.; PONCE DE LEON, A.; STRACHAN, D.P.; BOWER, J.S. *The Health Effects of an Air Pollution Episode in London, December 1991*. London: St George's Hospital Medical School, 1995. 184p.

CID 10. 10a Classificação Internacional de Doenças. Disponível em: <<http://www.datasus.gov.br/cid10/webhelp/cid10.htm>>. Acesso em: 9 jan. 2017.

DATASUS. Ministério da Saúde. Informações de Saúde. Assistência a Saúde. *Internações por Especialidades e Local de Internação*. Disponível em: <<http://www.tabnet.datasus.gov.br>>. Acesso em: 9 jan. 2017.

DAUMAS, R. P.; MENDONÇA, G. A. S.; PONCE DE LEON, A. Poluição do ar e mortalidade em idosos no Município do Rio de Janeiro: análise de série temporal. *Cad. Saúde Pública*, v.20, n.1, p.311-319, jan./fev. 2004.

KOHONEN, T. *Self-Organizing Maps*. 3 ed. Berlin: Springer-Verlag, 2001. (Springer Series in Information Kohonen, T., Self-Organizing Maps; n.30)

POLONIECKI, J. D.; ATKINSON, R. W.; PONCE DE LEON, A. et al. *Daily time series for cardiovascular hospital admissions and previous day's air pollution in London, UK*. [S1]: Occupational and Environmental Medicine, 1997. p. 535-540.

RONW; LUTGARDE, M.C. *Self- and Super-organizing Maps in R*. [SI]: The kohonen Package, 2007.

YAN, J. *som: Self-Organizing Map*. R package version 0.3-4. 2004. .Disponível in: <URL <http://CRAN>>. Acesso em: 2016.