

COMBINADO SISTEMAS DE INFORMAÇÃO GEOGRÁFICA E REDES NEURAIS GMDH PARA PREVER DENSIDADES DEMOGRÁFICAS

Nair Cristina Margarido Brondino¹
Odney Carlos Brondino²

Resumo: O objetivo deste trabalho é propor uma nova abordagem para estimar densidades demográficas, a partir da combinação de um Sistema de Informações Geográficas e de Redes Neurais GMDH. O modelo proposto consiste na divisão do espaço em uma malha retangular composta de células com 0.01km^2 de área. As previsões futuras são feitas com base nas densidades demográficas de cada célula e de suas vizinhas adjacentes em determinado momento. Apesar da pouca disponibilidade de dados na etapa de modelagem, a utilização do método proposto, aplicado ao caso de uma cidade média brasileira, apresentou resultados promissores.

Palavras-Chave: Sistema de Informações Geográficas; Redes Neurais Artificiais; Densidade Demográfica.

COMBINING GEOGRAPHIC INFORMATION SYSTEMS AND GMDH NEURAL NETWORKS TO PREDICT DEMOGRAPHIC DENSITY

Abstract: The aim of this work is to advance a new approach for estimating demographic density, through combining a Geographic Information System with GMDH Neural Networks. The model that is suggested parts the analyzed space into a rectangular grid formed by multiple cells measuring 0.01 km^2 each. The forecasts are elaborated based on the demographic density in each cell and in its neighboring cells at a given time. Despite the limited availability of data during the modeling phase, the utilization of this method for studying a Brazilian medium-sized city presented promising results.

Keywords: Geographic Information System; Artificial Neural Networks; demographic density.

¹ Professora Assistente Doutor do Departamento de Matemática da UNESP-BAURU. brondino@fc.unesp.br

² Professor dos Cursos de Engenharia de Produção do Instituto Paulista de Ensino Superior Unificado - IPESU e da Universidade Paulista - UNIP. ocbondino@ig.com.br.

Estudos Geográficos, Rio Claro, 8(2): 4-14, jul./dez., 2010 (ISSN 1678—698X)
<http://www.periodicos.rc.biblioteca.unesp.br/index.php/estgeo>

INTRODUÇÃO

O presente trabalho trata de uma nova abordagem para prever a tendência de crescimento espacial de cidades.

Um modelo que seja capaz de efetuar previsões de crescimento demográfico mais realistas tem sido o foco de pesquisa de muitos planejadores urbanos. A disponibilidade de uma ferramenta que proporcione previsões de boa qualidade conduz a um planejamento mais eficiente e sustentável, uma vez que pode permitir a escolha adequada dos lugares em que serão instalados equipamentos públicos, como hospitais e escolas.

Desta forma, modelos que apresentem características espaço-temporais, como os Autômatos Celulares (AC) e alguns modelos estatísticos, parecem ser uma escolha adequada, devido às características inerentes a este tipo de fenômeno. Dentre os pesquisadores que fizeram uso de autômatos celulares para estudar processos de dinâmica urbana, pode-se citar Brondino & Brondino (2005), Barredo, J. I. *et. al.* (2003), Bessusi *et al.* (1998) e Li *et al.* (2003). Neste tipo de modelo, onde o espaço é dividido em uma grade uniforme composta de várias células, as dinâmicas são geradas por um processo iterativo no qual, em toda interação, a célula muda de estado, a partir da utilização de alguma(s) regra(s) de transformação. Tais regras dependem do que está acontecendo na célula de estudo e em sua vizinhança imediata em determinado momento. De certa forma, todas as dinâmicas podem ser baseadas em tal representação, o que, segundo BATTY *et. al.* (1999), motiva o seu uso no estudo dos processos de dinâmica urbana. Entretanto, na maioria das vezes, Autômatos Celulares são utilizados para prever direções de crescimento, mas não fornecem uma previsão em termos de valores.

Neste contexto, o presente trabalho propõe um método para prever densidades demográficas. O método proposto aqui representa o espaço urbano através de uma malha retangular. A partir da divisão do espaço em células, utiliza-se uma Rede Neural Artificial GMDH (Group Method of Data Handling) para prever a densidade demográfica em cada célula da rede. A escolha do modelo GMDH, cujos aspectos serão melhor detalhados na seção 2 deste trabalho, consiste no fato de que, dentre outras possíveis aplicações, ela é indicada para fazer previsões de curto e médio prazo e tem se mostrado como substituto potencial aos clássicos modelos estatísticos. Apesar da pouca disponibilidade de dados na etapa de modelagem, a utilização do método proposto, aplicado ao caso de uma cidade média, apresentou resultados promissores, os quais são apresentados na seção 3. A seção 4 traz as considerações finais.

O MODELO GMDH

A rede neural GMDH é um tipo de rede auto-organizável, que, por sua estrutura, pode ser utilizada para fazer previsões de curto e médio prazo. Neste tipo de rede, uma relação geral entre entrada e saída pode ser descrita na forma de um modelo matemático, que também é chamado de função referência. Usualmente, esta descrição é considerada como uma forma discreta da Série Funcional de Volterra, apresentada na expressão (1).

$$y = a_0 + \sum_{i=1}^m a_i x_i + \sum_{i=1}^m \sum_{j=1}^m a_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m a_{ijk} x_i x_j x_k + \dots \quad (1)$$

onde y é a saída, $\mathbf{x} = (x_1, x_2, \dots)$ é o vetor de entrada e \mathbf{a} é o vetor de pesos. Como pode ser observado, a função dada em (1) é linear nos pesos, mas não linear nas variáveis. A forma parcial desta função, como um estado ou função soma, é desenvolvida em cada unidade simulada e então é ativada em paralelo para a construção mais complexa (Madala & Ivakhnenko, 1994).

Desta forma, cada unidade simulada k recebe um par de variáveis de entrada $(x_i, x_j) \subset \mathbf{x}$, $i \neq j$ e gera uma função $f(\cdot)$, que, por sua vez, é uma forma parcial da função de referência. A equação (2) mostra a função parcial $f(x_i, x_j)$.

$$f(x_i, x_j) = v_0 + v_1 x_i + v_2 x_j + v_3 x_i x_j + v_4 x_i^2 + v_5 x_j^2 \quad (2)$$

onde os $v^{(k)}$ representam os pesos. Desta forma, se m variáveis de entrada forem consideradas, então o número total de funções parciais como a mostrada em (2), que serão geradas na primeira camada intermediária, será dado por $\binom{m}{2}$. Se denotarmos por o a saída desejada, então o erro estimado em cada unidade é dado pela equação (3).

$$e_p = y_p - o_p; p \in N_A, \quad (3)$$

onde N_A denota o conjunto de treinamento.

O erro quadrático de treinamento para cada vetor de entrada é dado pela expressão (4). Os pesos $v^{(k)}$ devem ser tais que conduzam ao erro quadrático mínimo e são obtidos pelo Método dos Mínimos Quadrados. Cada camada contém um grupo de unidades que estão interconectadas com as unidades da próxima camada. Os pesos em cada camada devem minimizar o erro E . O valor de uma função objetivo é usado como limiar para ativar ou desativar cada unidade. Esta função é comparada com a saída gerada quando outro subconjunto de dados (conjunto de teste), denotado por N_B é apresentado à rede. As unidades ativadas em uma camada são, então, utilizadas como entrada na próxima camada e assim por diante.

$$E = \sum_{p \in N_A} e_p^2 \quad (4)$$

RESULTADOS E MÉTODOS

Como já havia a disponibilidade de dados dos Censos de 1980, 1991 e 2000 em formato digital, a cidade de São Carlos, localizada no Estado de São Paulo (Brasil), foi escolhida para o estudo de caso realizado neste trabalho.

Inicialmente houve a divisão do espaço em uma grade de 162 linhas e 140 colunas, totalizando 22680 células, cada qual representando uma área de 0.01 km².

A divisão foi realizada por intermédio da utilização de um Sistema de Informações Geográficas - SIG (no caso, o TransCAD). Na sequência, os dados das densidades demográficas de 1980, 1991 e 2000, que já se encontravam disponíveis em formato digital, foram adicionados a cada célula, através de um recurso do próprio SIG. Os dados foram, então, exportados para serem utilizados na simulação da rede GMDH, a qual foi implementada através de uma rotina programada no software Matlab e denominada GMDHNET. A vantagem em escolher o software Matlab nesta simulação reside no fato de que, devido à sua capacidade de memória, ele permite tratar, simultaneamente, várias matrizes de grande dimensão ao mesmo tempo.

A rotina GMDHNET considera como entrada da rede, um vetor de cinco componentes, cada qual representando as densidades demográficas de 1980, em cada célula de estudo e nas quatro células imediatamente adjacentes a ela, de acordo com o esquema apresentado na Figura 1. A escolha deste tipo de vizinhança deu-se devido ao aumento de complexidade do modelo GMDH quando existem muitos nós de entrada. Como saída da rede, foram usadas as densidades de 1991.

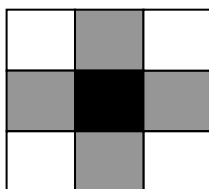


Figura 1 – Célula de estudo (em preto) e vizinhança escolhida para a construção da rede GMDH (em cinza).

Após ler os arquivos das densidades de 1980 e 1991, a rotina separou o conjunto total de dados em duas partes, reservando 90% para a fase de treinamento e os 10% remanescentes para a fase de teste do modelo.

Definidas entradas e saídas, iniciou-se o processo de construção da rede. O algoritmo de treinamento escolhido foi o indutivo. Desta forma, a saída de cada neurônio da primeira camada intermediária consiste de uma função não linear de duas variáveis de entrada, como a dada pela equação (2), e ajustada pelo Método dos Mínimos Quadrados Ordinários.

Para a construção do modelo utilizado aqui, foi construída uma rede com 10 neurônios na primeira camada intermediária, cada um combinando as variáveis de entrada duas a duas, de acordo com a equação apresentada em (2). A seguir, o Erro Quadrático Médio (EQM) foi calculado para cada neurônio da primeira camada intermediária, tanto para o conjunto de treinamento como para o de teste. A escolha dos neurônios que serviriam de entrada para a próxima camada intermediária foi feita através de uma comparação dos valores da função, contidos no conjunto de teste, e de um valor de EQM previamente escolhido. O diagrama esquemático da rede GMDH construída é apresentado na Figura 2, onde x_{ij} denota a célula localizada na linha i e coluna j da grade.

Como indicado na Figura 2, as simulações realizadas conduziram à construção de uma rede com três camadas intermediárias. Os neurônios que apresentaram os menores EQM na primeira camada intermediária foram aqueles que relacionaram a célula i em estudo e suas vizinhas, os quais estão

representados pelos neurônios de números 2, 5, 8 e 9 na Figura 2. As saídas destes neurônios serviram como entradas para a segunda camada intermediária e assim por diante, até a última camada, cuja saída culminou na melhor função ajustada. Vale ressaltar que a função ajustada constitui-se de um modelo cuja finalidade é estimar a densidade em cada célula num período de 10 em 10 anos.

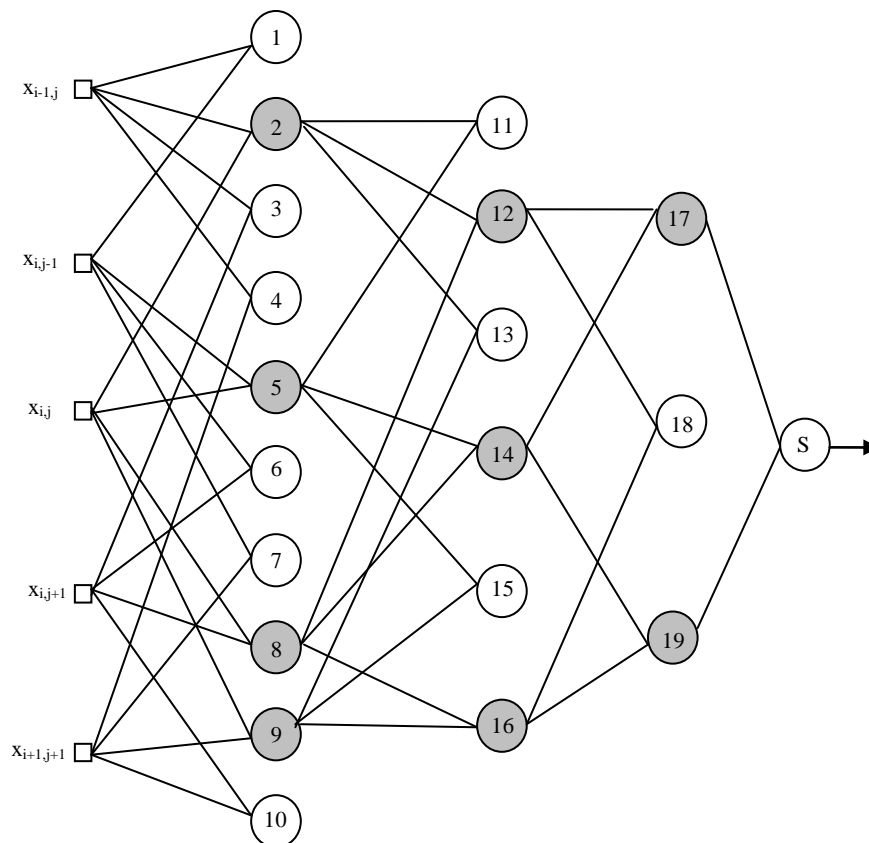


Figura 2 – Estrutura da rede construída para esta aplicação.

Ao final da simulação, os valores da função ajustada para cada célula para o ano de 1991 foram exportados para o SIG, com o objetivo de construir uma representação georeferenciada do comportamento do modelo. Os mapas apresentados na Figura 3 apresentam o comportamento das densidades reais e estimadas para o ano de 1991. As diferenças na geometria dos mapas apresentados indicam que algumas regiões habitadas em 1991 não foram identificadas pelo modelo, como, por exemplo, nas regiões Noroeste, Sudoeste e Sul. Vale ressaltar que o desenvolvimento observado nas regiões Sudoeste e Sul ocorreu em virtude da construção de um bairro nesta região, que está localizado a uma grande distância do entorno da cidade e que no de 1991 a densidade demográfica do setor censitário correspondente a esta região era inferior a 0.01 hab./km². Devido à particularidade deste fenômeno, que não segue uma tendência natural e, conseqüentemente não pôde ser captado pelo modelo, uma vez que este foi gerado pela configuração observada em 1980, tais células aparecem vazias nos resultados apresentados na Figura 3b. Tal região, hachurada na cor rosa na Figura 3a, corresponde a um total de 1183 células, 22,5% de um total de 5238 células não vazias em 1991. Apesar desse fato, 2660 células foram classificadas adequadamente dentro das faixas de densidades. Um comparativo entre densidades reais e previstas para o ano de 1991

pode ser observado no mapa mostrado na Figura 4. A Figura 5 apresenta as diferenças entre os valores de densidade previsto e real para o ano de 1991. Os resultados indicam que em 67% das células, o erro de previsão ficou na faixa que vai de - 10 a 10 habitantes por km², erro este que foi considerado aceitável para este estudo.

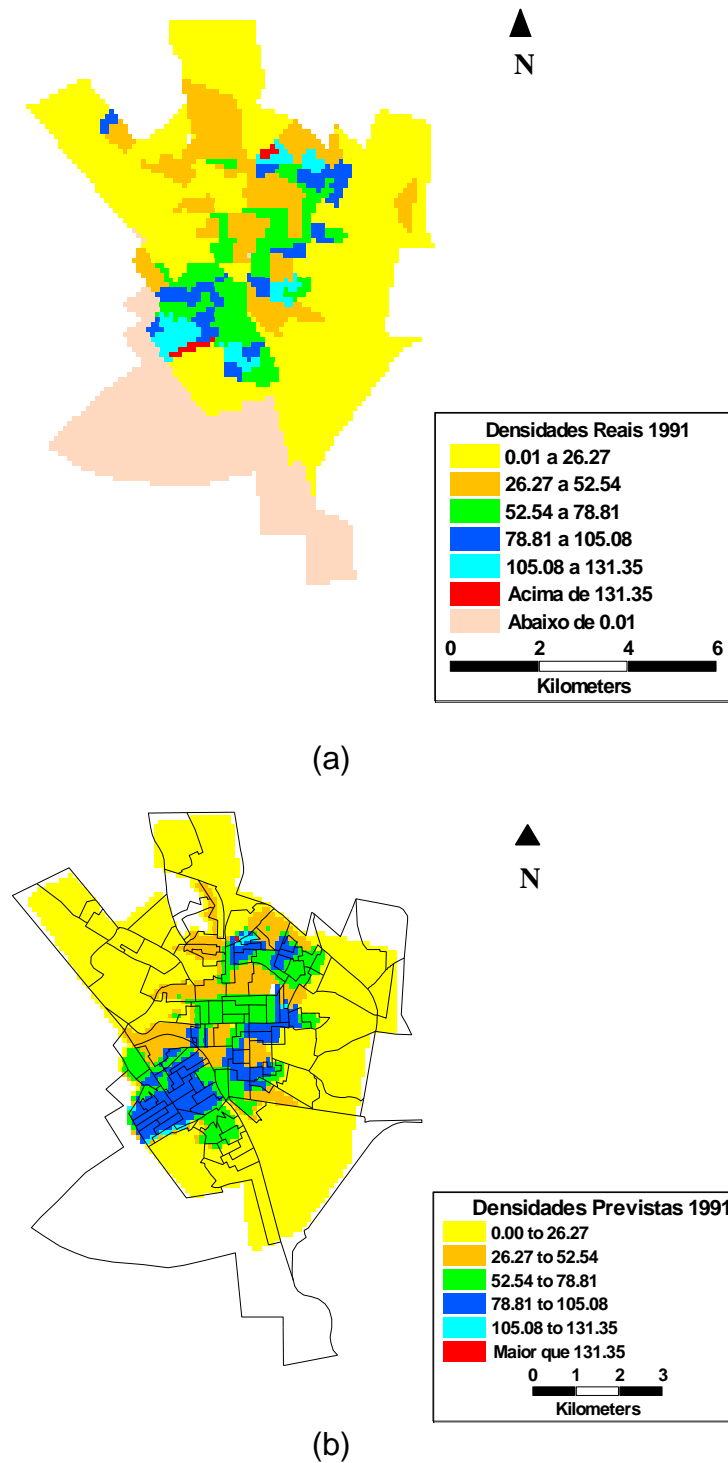


Figura 3 – Densidades Reais (a) e Estimadas (b) para o ano de 1991.

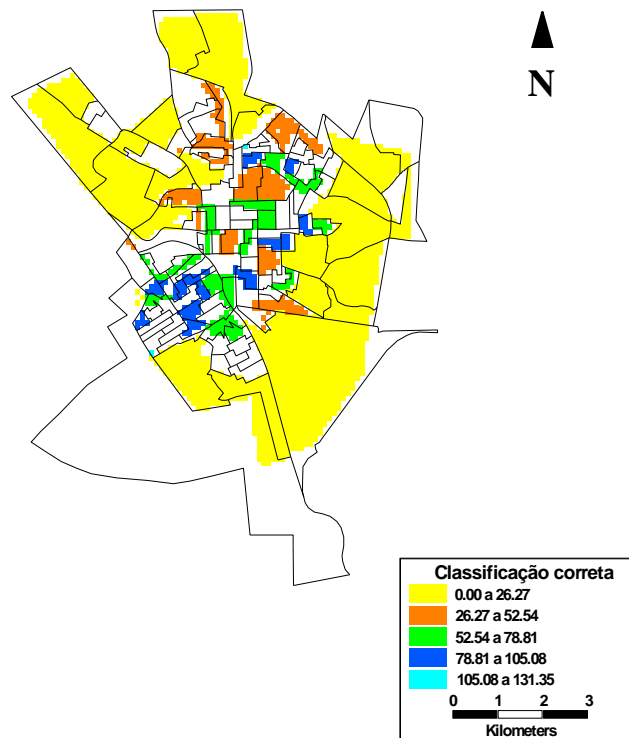


Figura 4 – Coerência entre valores reais e previstos para o ano 1991.

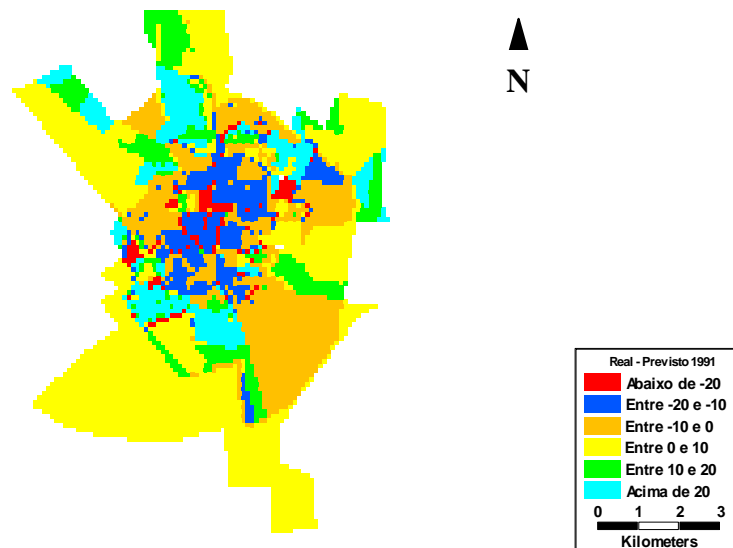


Figura 5 – Diferença entre valores previsto e real para o ano de 1991.

Tendo em vista o que se discutiu anteriormente, o modelo ajustado foi considerado satisfatório. Desta forma, as densidades estimadas para o ano de 1991 em cada célula foram utilizadas para o cálculo das projeções para o ano 2000. Os comportamentos das densidades reais e estimadas para o ano 2000 podem ser observados nas Figuras 6a e 6b. Os resultados observados indicaram que 67% das células foram classificadas adequadamente pelo modelo dentro das faixas de

Estudos Geográficos, Rio Claro, 8(2): 4-14, jul./dez., 2010 (ISSN 1678—698X)
<http://www.periodicos.rc.biblioteca.unesp.br/index.php/estgeo>

densidade especificadas. Este comportamento pode ser observado no mapa apresentado na Figura 7.

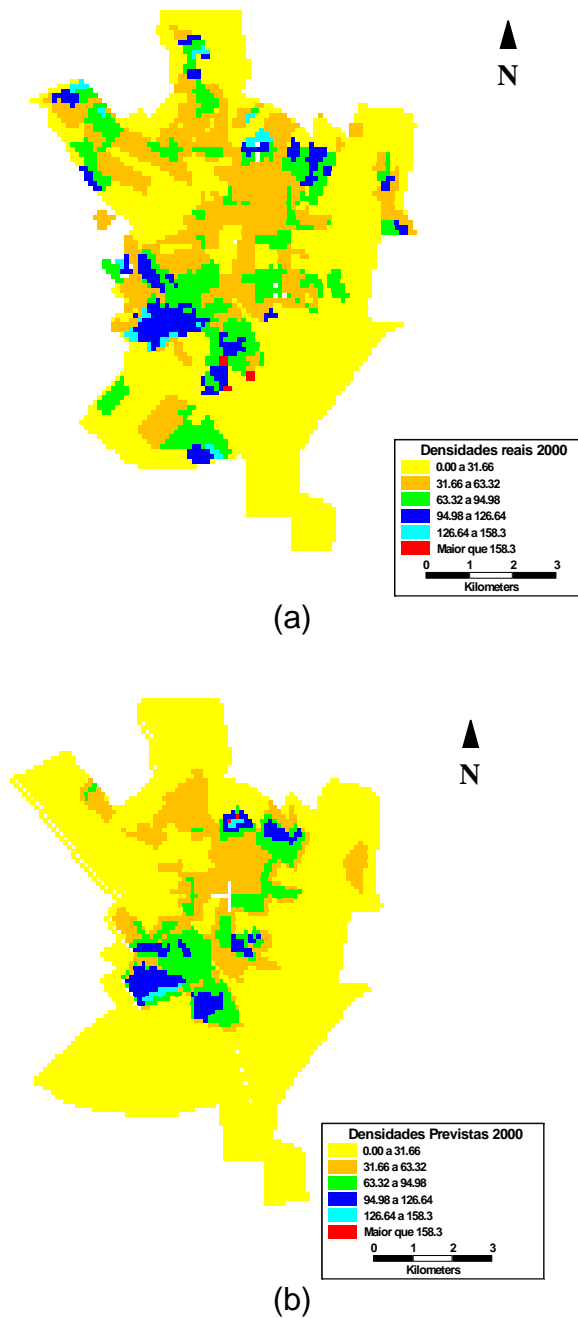


Figura 6 – Comportamento das Densidades Reais e Previstas no ano 2000.

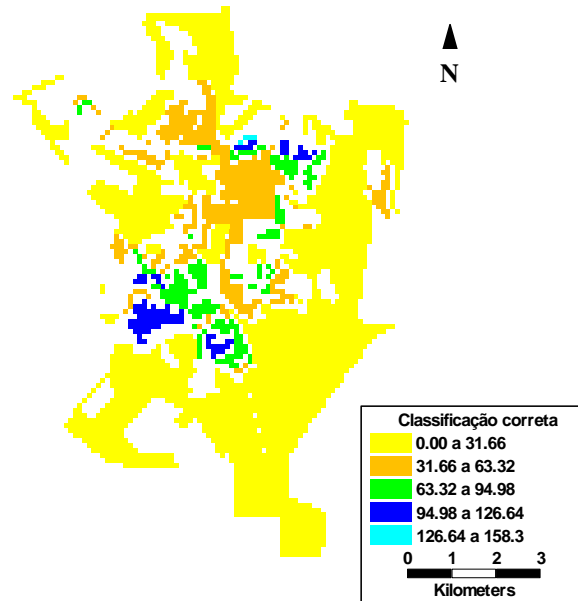


Figura 7 – Coerência entre valores reais e previstos para o ano 2000.

As diferenças entre as densidades previstas e reais para o ano 2000 são apresentadas na Figura 8. Os resultados observados indicam que em 57% das regiões, os erros de estimação localizaram-se na faixa entre -10 e 10 habitantes por quilômetro quadrado e que em 24% das regiões, o erro foi maior que 20 habitantes por km².

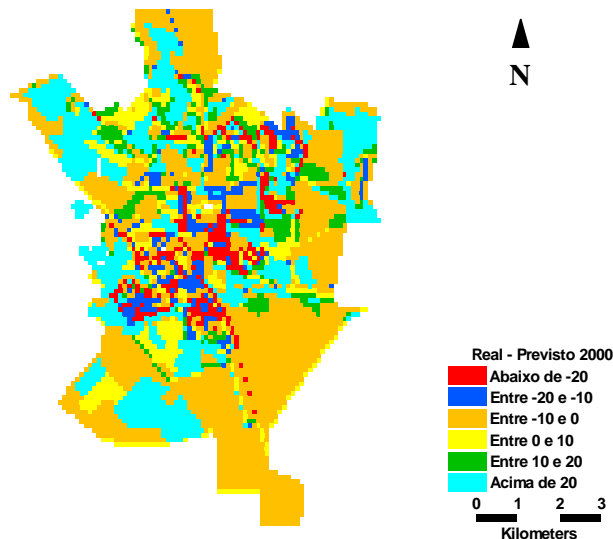


Figura 8 – Diferença entre valores previsto e real para o ano 2000.

CONSIDERAÇÕES FINAIS

Este trabalho trouxe a proposta da aplicação de um modelo que combina ferramentas de um Sistema de Informações Geográficas e uma rede neural do tipo GMDH com vistas à previsão de densidades demográficas em uma cidade média brasileira. O modelo foi elaborado a partir dos dados dos Censos de 1980 e 1991. Na sequência, os valores estimados pelo modelo para o ano de 1991 foram utilizados para realizar as projeções para o ano 2000. Tais projeções foram comparadas aos valores reais obtidos em 2000 e os resultados observados indicaram que o modelo classificou adequadamente 67% das regiões com relação às faixas de densidade especificadas. Com relação aos erros de previsão, pôde-se observar que em 57% das células não vazias em 2000, este ficou na faixa de -10 a 10 habitantes por km². Desta forma, apesar da escassez de dados disponíveis, o modelo mostrou-se promissor, uma vez que foi gerado a partir de dados coletados em duas etapas com diferença de 10 anos. Uma vez que se trata de um modelo que envolve uma série temporal, na medida em que mais dados forem coletados e disponibilizados, sua sensibilidade pode ser melhorada significativamente.

REFERÊNCIAS

- BARREDO, J. I.; Kasanko, M.; McCormick, N.; Lavallo, C. (2003) **Modelling dynamic spatial processes: simulation of urban future scenarios through cellular automata**. *Landscape and Urban Planning*, 64, pp. 145-160.
- BATTY, M.; Xie, Y.; Sun, Z. (1999) **Modeling urban dynamics through GIS-based cellular automata**. *Computers, Environment and Urban Systems*, 23, p. 205-233.
- BESUSSI, E.; Cecchini, A.; Rinaldi, H. (1998) **The Diffused City of the Italian North-East: Identification of Urban Dynamics Using Cellular Automata Urban Models**. *Computers, Environment and Urban Systems*, vol. 22, n.5, pp. 497-523.
- BRONDINO, N. C. M.; Brondino, O. C. (2005). **Uso de Autômatos Celulares para Analisar a Tendência de Crescimento de uma Cidade Média Brasileira**. Anais do 1º Congresso Luso-Brasileiro para o Planejamento Urbano, Regional, Integrado, sustentável - Pluris 2005. São Carlos. CD-ROM.
- MADALA, H. R.; Ivakhnenko, A. G. (1994). **Inductive Learning Algorithms for Complex System Modeling**. CRC Press. Florida.
- LI, L.; Sato, Y.; Haihong, Z. (2003). **Simulating spatial urban expansion based on a physical process**. *Landscape and Urban Planning*, 64, pp. 67–76.

Artigo submetido em: 16/02/2011

Aceito para publicação em: 03/09/2012

Publicado em: 03/09/2012