

# C&S SIG

***RELAÇÃO ENTRE AS VARIÁVEIS SOCIAIS,  
ECONÔMICAS E AMBIENTAIS COM O PADRÃO DA  
DISTRIBUIÇÃO ESPAÇO-TEMPORAL DOS CASOS  
DE DENGUE POR MUNICÍPIO NO BRASIL:***

***De 2008 até 2012, utilizando o SOM***

João Felipe Campos Villar

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Ciência e Sistemas de Informação Geográfica

**RELAÇÃO ENTRE AS VARIÁVEIS SOCIAIS,  
ECONÔMICAS E AMBIENTAIS COM O PADRÃO DA  
DISTRIBUIÇÃO ESPAÇO-TEMPORAL DOS CASOS  
DE DENGUE POR MUNICÍPIO NO BRASIL:**

De 2008 até 2012, utilizando o SOM

Dissertação Orientada por:

Professor Doutor  
Roberto André Pereira Henriques

Fevereiro 2015

*"É preciso sonhar, mas com a condição de crer em nosso sonho, de observar com atenção a vida real, de confrontar a observação com nosso sonho, de realizar escrupulosamente nossas fantasias. Sonhos, acredite neles."*

(Lenin, 1902)

## **Agradecimentos**

Primeiramente gostaria dedicar meus agradecimentos ao Professor Dr. Pedro Cabral que me auxiliou, incentivou e me acolheu de maneira muito generosa desde o primeiro dia que iniciei meus estudos nessa instituição.

À Portugal, sobretudo seus cidadãos que me receberam de braços abertos, fazendo por mim muitas vezes o que minha terra natal o Brasil não tornou possível nem viável.

A todos os meus companheiros de classe, especialmente ao Hugo Alexandre Dias da Silva que me acolheu como irmão. Não esquecendo também do Artur, Filipe e Kátia que estiveram sempre comigo em todo o trajeto desse curso.

Aos meus orientadores Dr. Fernando Martin Biscione e Professor Dr. Roberto André Pereira Henriques, por aceitarem esse desafio.

Aos meus pais (Carlos e Andrea) por se esforçarem ao máximo na minha educação e orientação sendo o alicerce de toda minha vida pessoal, profissional e acadêmica.

À minha companheira Anastasiia Razumova que vem sendo meu porto seguro, e que relativamente nesse trabalho fez com que eu o observasse através de outro ponto de vista.

Ao Agrocampus Ouest e a cidade de Angers na França que me possibilitaram um novo conhecimento e suas aplicações na Ciência e Sistemas de Informação Geográfica (SIG).

À todos os meus amigos do projeto BH 1, os quais foram fundamentais no desenrolar da minha juventude (Leandro, André, Pedro, Thiago, Kirk e Leonardo).

A todos os meus ex-colegas de trabalho da Unimed-BH (Nelson, Renata, Cláudia, Márcia e Flávia), especialmente a gestora Dra. Ana Paula Franco Viegas Pereira que tornou possível a realização desse curso em conjunto com minhas atribuições profissionais e ao Dr. Rômulo Paes que despertou toda minha sede de conhecimento na área de SIG.

Aos meus colegas de trabalho da Arya inventário territorial que também foram muito importantes, sobretudo me dando apoio e incentivo na etapa de conclusão desse trabalho.

Finalmente para todas as figuras históricas que me serviram de exemplo e inspiração, e que muitas vezes me fazem seguir adiante com seus exemplos de vida. Entre eles: Ayrton Senna da Silva, Ernesto Guevara de La Sierna, Sergei Pavlovich Korolev, Albert Einstein, Alberto Santos Dumont, Irineu Evangelista de Sousa vulgo Barão de Mauá, Howard Robard Hughes Jr, dentre outros.

# RELAÇÃO ENTRE AS VARIÁVEIS SOCIAIS, ECONÔMICAS E AMBIENTAIS COM O PADRÃO DA DISTRIBUIÇÃO ESPAÇO-TEMPORAL DOS CASOS DE DENGUE POR MUNICÍPIO NO BRASIL:

De 2008 até 2012, utilizando o SOM

## Resumo

Atualmente, um dos principais desafios que afeta a saúde pública no Brasil é a crescente evolução no número de casos e epidemias provocados pelo vírus da dengue. Não existem estudos suficientes que consigam elucidar quais fatores contribuem para a evolução das epidemias de Dengue. Fatores como condições sanitárias, localização geográfica, investimentos financeiros em infraestrutura e qualidade de vida podem estar relacionados com a incidência de Dengue. Além disso, outra questão que merece um maior destaque é o estudo para se identificar o grau de impacto das variáveis determinantes da dengue e se existe um padrão que está correlacionado com a taxa de incidência. Desta forma, este trabalho tem como objetivo principal a correlação da taxa de incidência da dengue na população de cada município brasileiro, utilizando dados relativos aos aspectos sociais, econômicos, demográficos e ambientais. Outra contribuição relevante do trabalho, foi a análise dos padrões de distribuição espacial da taxa de incidência de Dengue e sua relação com os padrões encontrados utilizando as variáveis socioeconômicas e ambientais, sobretudo analisando a evolução temporal no período de 2008 até 2012. Para essa análises, utilizou-se o Sistema de Informação Geográfica (SIG) aliado com a mineração de dados, através da metodologia de rede neural mais especificamente o mapa auto organizável de Kohonen ou *self-organizing maps* (SOM). Tal metodologia foi empregada para a identificação de padrão de agrupamentos dessas variáveis e sua relação com as classes de incidência de dengue no Brasil (Alta, Média e Baixa). Assim, este projeto contribui de forma significativa para uma melhor compreensão dos fatores que estão associados à ocorrência de Dengue, e como essa doença está correlacionada com fatores como: meio ambiente, infraestrutura e localização no espaço geográfico.

# **Relation between social, economic and environmental variables with space-temporal pattern of dengue cases by municipality in Brazil:**

From 2008 until 2012, using SOM

## **Abstract**

One of the main challenges that affect the public health in Brazil nowadays is the growing up evolution on the number of cases and epidemics made by the dengue virus. Do not exist sufficient studies to elucidate which kind of factors contribute for dengue epidemics evolution. Therefore, some factors such as sanitary conditions, geographic location, and financial investments in infrastructure and health quality, can be related with the dengue incidence. Another great question that deserves a more prominent focus is the study to identify the impact level of some dengue determinant variables and if exists one pattern that can be related with the incidence rate. In this way, this study had as main objective the correlation of dengue incidence rate in the population of each Brazilian municipality, using social aspects, economics, demographics and environmental data related. Another relevant contribution of this study was the spatial distribution pattern analysis of the dengue incidence tax and its relation with the found patterns utilizing socioeconomic and environmental variables, especially analyzing the temporal evolution on the period from 2008 until 2012. For such, the geographic information system (GIS) allied with the data mining utilizing the neural networks methodology more specifically the Kohonen self-organizing map (SOM), was used for grouping pattern identification for such variables and their relation with the dengue incidence rate classes (High, Medium, Low) in Brazil. In this way, this project contributes significantly for one better factors compression that are associated with the dengue occurrence, and how this disease is correlated with some factors, such as: environment, infrastructure and geographic space location.

## **Palavras-Chave**

Dengue

Epidemiologia

Sistema de Informação Geográfica

Self-Organizing Map

## **Keywords**

Dengue

Epidemiology

Geographic Information System

Self-Organizing Map

## Acrónimos

**BMU** – Best Matching Unit

**CE** – Comissão Europeia

**EQM** – Erro de Quantização Médio

**ET** – Erro Topológico

**FUNASA** – Fundação Nacional de Saúde

**IBGE** – Instituto Brasileiro de Geografia e Estatística

**LIRAA** – Levantamento rápido do Índice de Infestação por *Aedes aegypti*

**MS** – Ministério da Saúde do Brasil

**DATASUS** – Departamento de informática do SUS

**RNA** – Rede Neuronal Artificial

**SOM** – Self-organizing Map

**SINAN** - Sistema de informação de Agravos de Notificação

**SUS** – Sistema Único de Saúde

**U-Mat** – Unified Distance Matrix



# Índice

<b>Agradecimentos</b> .....	iv
<b>Resumo</b> .....	v
<b>Abstract</b> .....	vi
<b>Palavras-Chave</b> .....	vii
<b>Keywords</b> .....	vii
<b>Acrónimos</b> .....	viii
<b>Índice de Tabelas</b> .....	xi
<b>Índice de Figuras</b> .....	xii
<b>Índice de Gráficos</b> .....	xiv
<b>1. Introdução</b> .....	
1.1 Enquadramento .....	1
1.2 Hipótese .....	2
1.3 Objetivo principal .....	2
1.4 Objetivos específicos .....	2
1.5 Estrutura .....	3
<b>2. Dengue</b> .....	
2.1 Introdução .....	4
2.2 Histórico do dengue no Brasil .....	7
2.3 Aplicação dos Sistemas de Informação Geográfica .....	10
<b>3. Comportamento geográfico do Dengue e sua relação com variáveis determinantes</b> .....	11
3.1 Introdução .....	11
3.2 Fatores determinantes do dengue no Brasil .....	13
3.3 Relação das variáveis determinantes com o espaço .....	14
3.4 Relação entre as variáveis determinantes com a taxa de incidência .....	16
<b>4. Rede neural SOM</b> .....	17
4.1 Introdução .....	17
4.2 Algoritmo SOM e sua parametrização .....	23
4.3 Interpretação dos resultados do SOM .....	27
4.3.1 Visualização da estrutura do SOM .....	28
4.3.2 Visualização dos dados do SOM .....	30
4.3.3 Integração geográfica da visualização dos resultados .....	31
4.4 Medidas de qualidade do SOM .....	32
4.4.1 Erro médio de quantização .....	32
4.4.2 Erro topológico .....	33
4.4.3 Observações sobre medidas de qualidade .....	33
<b>5. Metodologia</b> .....	34
5.1 Recolhimento e tratamento dos dados .....	34
5.1.1 Revisão Bibliográfica .....	34
5.1.2 Tratamento dos dados .....	38
5.2 Estruturação e elaboração da base de dados .....	41
5.3 Análise exploratória dos dados .....	41
5.4 Análise espacial dos dados .....	43
5.5 Análise multidimensional e mineração dos dados .....	46
5.5.1 Variáveis selecionadas para análise no SOM .....	46
5.5.2 Ajustamento dos dados .....	48
5.5.3 Definição do tamanho e da topologia .....	50
5.5.4 Critérios de treinamento .....	50
5.6 Avaliação do modelo .....	51
<b>6. Resultados</b> .....	53
6.1 Resultados da análise exploratória .....	54
6.2 Resultados da análise espacial .....	58

6.3 Resultados dos clusters .....	71
6.3.1 Caracterização do perfil dos clusters .....	71
6.3.2 Caracterização geográfica .....	75
<b>7. Limitações.....</b>	<b>80</b>
<b>8. Discussão e conclusão .....</b>	<b>81</b>
<b>ANEXO A .....</b>	<b>87</b>
<b>ANEXO B .....</b>	<b>107</b>
<b>ANEXO C .....</b>	<b>108</b>

## Índice de Tabelas

TABELA 1 – Macro e Micro determinantes de transmissão da dengue.....	12
TABELA 2 – Classificação da taxa de incidência da dengue.....	43
TABELA 3 – Parâmetros utilizados na definição dos dois treinos do SOM realizados para cada ano do período analisado.....	50
TABELA 4 – Erros encontrados para a rede 1 e rede 2, geradas por processos de aprendizagem diferentes em todos os anos do período analisado.....	51
TABELA 5 – Correlação da taxa de incidência da dengue com o cluster identificado pelo SOM em 2008.....	72
TABELA 6 – Correlação da taxa de incidência da dengue com o cluster identificado pelo SOM em 2009.....	72
TABELA 7 – Correlação da taxa de incidência da dengue com o cluster identificado pelo SOM em 2010.....	73
TABELA 8 – Correlação da taxa de incidência da dengue com o cluster identificado pelo SOM em 2011.....	74
TABELA 9 – Correlação da taxa de incidência da dengue com o cluster identificado pelo em 2012.....	74

## Índice de Figuras

FIGURA 1 – Imagem do mosquito <i>Aedes aegypti</i> .....	4
FIGURA 2 – Distribuição global da dengue: áreas infestadas com o mosquito vetor da dengue estão em laranja e as áreas com dengue epidêmica estão em vermelho.....	5
FIGURA 3 – Mapa da infestação por município pelo vetor <i>Aedes aegypti</i> de 1995 a 2010 no Brasil.....	6
FIGURA 4 – Mapa da infestação pelo <i>Aedes aegypti</i> no Brasil, período entre 1967 até 1984.....	8
FIGURA 5 – Mapa da infestação pelo <i>Aedes aegypti</i> no Brasil, período entre 1965 até 1985, origem Paraguai.....	9
FIGURA 6 – Mapa da infestação pelo <i>Aedes aegypti</i> no Brasil, período entre 1965 até 1985, origem Guianas e Venezuela.....	9
FIGURA 7 – Divisão municipal do Brasil conforme resolução do IBGE em 2007.....	15
FIGURA 8 – Representação gráfica de um neurônio biológico ou natural e seus principais componentes.....	17
FIGURA 9 – Modelo de rede neural simples, alimentada adiante com camada única.....	18
FIGURA 10 – Desenho esquemático do fluxo de funcionamento de uma rede SOM e seus constituintes básicos: dados de entrada (3D ou +) e Dados de Saída (2D).....	21
FIGURA 11 – Esquema comparativo de diferentes métodos de segmentação.....	22
FIGURA 12 – Tamanho e topologia de um SOM 2D 3x3 retangular e Hexagonal.....	23
FIGURA 13 – Figura esquemática apresentando o ajustamento do BMU e de sua vizinhança através dos dados de entrada (X).....	25
FIGURA 14 – Diagrama de Voronoi. Divisão do espaço onde todos os pontos internos estão próximos ao ponto gerador correspondente.....	26
FIGURA 15 – Funções bubble e gaussiana.....	27
FIGURA 16 – Esquema dos diversos métodos de visualização do SOM.....	28
FIGURA 17 – Método da Matriz-U, representação da projeção dos neurônios.....	29
FIGURA 18 – Formas de apresentação da matriz-u, gráfico de pizza à esquerda e gráfico de barras à direita.....	29
FIGURA 19 – Exemplos de um plano de componente relativos a 3 variáveis observadas (X, Y e salário).....	30
FIGURA 20 – Exemplo das diferentes formas de se visualizar as observações do SOM...	31
FIGURA 21 – Mapa com a distribuição espacial das estações meteorológicas automáticas (EMA) no Brasil.....	36
FIGURA 22 – Modelo de espacialização das chuvas gerado pela krigagem.....	39

FIGURA 23 – Esquema explicativo da equação utilizada no modelo de regressão.....	42
FIGURA 24 – Equações formuladas para explanação do índice local de Moran.....	45
FIGURA 25 – Matriz-U de cada ano do período analisado (2008 até 2012).....	51
FIGURA 26 – Segmentação da Matriz-U de cada ano do período analisado (2008 até 2012).....	53
FIGURA 27 – Divisão do território brasileiro em regiões e unidades federativa (estados).....	54
FIGURA 28 – Análise do índice local de Moran para o ano de 2008.....	61
FIGURA 29 – Análise do índice local de Moran para o ano de 2009.....	62
FIGURA 30 – Análise do índice local de Moran para o ano de 2010.....	63
FIGURA 31 – Análise do índice local de Moran para o ano de 2011.....	64
FIGURA 32 – Análise do índice local de Moran para o ano de 2012.....	65
FIGURA 33 – Taxa de incidência da dengue no Brasil em 2008.....	66
FIGURA 34 – Taxa de incidência da dengue no Brasil em 2009.....	67
FIGURA 35 – Taxa de incidência da dengue no Brasil em 2010.....	68
FIGURA 36 – Taxa de incidência da dengue no Brasil em 2011.....	69
FIGURA 37 – Taxa de incidência da dengue no Brasil em 2012.....	70
FIGURA 38 – Distribuição espacial dos clusters identificados pelo GeoSOM Suite em 2008.....	75
FIGURA 39 – Distribuição espacial dos clusters identificados pelo GeoSOM Suite em 2009.....	76
FIGURA 40 – Distribuição espacial dos clusters identificados pelo GeoSOM Suite em 2010.....	77
FIGURA 41 – Distribuição espacial dos clusters identificados pelo GeoSOM Suite em 2011.....	78
FIGURA 42 – Distribuição espacial dos clusters identificados pelo GeoSOM Suite em 2012.....	79

## Índice de Gráficos

GRÁFICO 1 – Correlação de variáveis ambientais com os casos da dengue no Sudeste do Brasil em 2008.....	55
GRÁFICO 2 – Correlação de variáveis ambientais com os casos da dengue no Sudeste do Brasil em 2009.....	55
GRÁFICO 3 – Correlação de variáveis ambientais com os casos da dengue no Sudeste do Brasil em 2010.....	56
GRÁFICO 4 – Correlação de variáveis ambientais com os casos da dengue no Sudeste do Brasil em 2011.....	56
GRÁFICO 5 – Correlação de variáveis ambientais com os casos da dengue no Sudeste do Brasil em 2012.....	57
GRÁFICO 6 – Índice global de Moran para todos os municípios do Brasil utilizando a taxa de incidência da dengue em 2008.....	58
GRÁFICO 7 – Índice global de Moran para todos os municípios do Brasil utilizando a taxa de incidência da dengue em 2009.....	59
GRÁFICO 8 – Índice global de Moran para todos os municípios do Brasil utilizando a taxa de incidência de dengue em 2010.....	59
GRÁFICO 9 – Índice global de Moran para todos os municípios do Brasil utilizando a taxa de incidência da dengue em 2011.....	60
GRÁFICO 10 – Índice global de Moran para todos os municípios do Brasil utilizando a taxa de incidência da dengue em 2012.....	60

# 1. Introdução

## 1.1 Enquadramento

A epidemiologia paisagística, uma das áreas da epidemiologia, parte da premissa que, se as exigências ambientais dos transmissores e reservatórios de doenças são conhecidas, é possível prever riscos epidemiológicos através do conhecimento das variáveis ambientais, auxiliando a tomada de medidas profiláticas e de controle da enfermidade. A possibilidade de se combinar as variáveis ambientais com os locais de incidência de doenças trazidas pelo avanço tecnológico do geoprocessamento e sensoriamento remoto, vem transformando a epidemiologia paisagística nas últimas décadas (APARÍCIO, 2001). Geralmente são utilizados dados cartográficos inerentes a dois eixos: dados relativos a população infectada (ex: localização dos casos notificados da dengue) e os dados relativos aos vetores transmissores (ex: locais com nível de infestação pelo vetor transmissor elevado).

No eixo relativo à população, podem ser utilizadas informações acerca da população atingida pela epidemia. No Brasil, o Ministério da Saúde, através do DATASUS, disponibiliza essas informações no repositório de dados Sistema Nacional de Agravos de Notificação (SINAN), sendo possível assim obter o endereço de residência juntamente com o prontuário médico relativo a cada infectado pela dengue. Outra fonte de informação bastante útil são os levantamentos censitários, pois ajudam no estabelecimento de correlação dos casos de infecção com fatores socioeconômicos.

Já o eixo relativo aos vetores transmissores podem ser subsidiados de informações relativas ao meio ambiente, proveniente de informações cartográficas como a carta topográfica disponibilizadas nas escalas de 1:25.000 até 1:1.000.000 pelo IBGE. Outra fonte de informações sobre o mosquito vetor, é o mapeamento rápido dos índices de infestação por *Aedes aegypti* (LIRAA), disponibilizados pelo Ministério da Saúde. Informações ligadas ao tempo e clima também são outra fonte de informação importante quando se deseja estudar o comportamento da epidemia e do mosquito, especificamente da relação da sazonalidade com o aumento de vetores, através de variáveis como o índice de pluviosidade, humidade e temperatura. Recentemente a Detecção Remota vem sendo empregada frequentemente em estudos que desejam obter maiores informações acerca do ambiente e do *habitat* em que os vetores da dengue vivem e se reproduzem e também a população que está predominantemente exposta ao contato com o ambiente de proliferação do vetor. Dessa maneira os dados cartográficos e de cunho espacial são extremamente importantes na elucidação do relacionamento destes dois eixos de investigação, realizando assim um trabalho ecológico dentro da epidemiologia.

Grande parte desses estudos utilizam áreas político-administrativas, que representam recortes do espaço geográfico, valendo-se dessas unidades para detecção de padrões de transmissão. Contudo, elas nem sempre representam a realidade envolvida na dinâmica epidemiológica da doença. Diante disso, a estratificação territorial permite dimensionar espacialmente os eventos através de procedimento de agregação que ocorre por homogeneidade de características e de desagregação dos territórios devido à heterogeneidade destas. Nesse sentido, a maioria dos estudos têm utilizado a análise de *clusters* na busca de padrões espaciais de eventos e caracterização de áreas homogêneas.

Analisar o papel das populações humanas e da infestação pelo mosquito transmissor da dengue em cada território, considerando-se as condições socioeconômicas e do ambiente onde se processa a interação entre eles, contribui de maneira significativa na identificação do papel de cada um na manutenção da circulação viral, podendo acrescentar elementos ao debate das estratégias de prevenção, controle e combate à dengue. Dessa forma, com essa dissertação pretende-se identificar e caracterizar os *clusters* espaciais identificados através da correlação de variáveis sociais, ambientais e econômicas para posterior correlação com os casos da dengue, levando em consideração o espaço geográfico como ponto fundamental na organização e distribuição dos fenômenos estudados.

## **1.2 Hipótese**

Existe um padrão e uma relação de dependência entre as variáveis sociais, econômicas e ambientais com o perfil da distribuição espaço temporal dos casos da dengue ocorridos nos municípios brasileiros ao longo do período de 2008 a 2012.

## **1.3 Objetivo principal**

Identificar o perfil e comportamento da dengue no Brasil baseados nas variáveis sociais, econômicas e ambientais, relacionando-as com a taxa de incidência da dengue na população, no período de 2008 até 2012.

## **1.4 Objetivos específicos**

- Calcular a taxa de incidência da dengue nos municípios da série temporal analisada;
- Investigar a existência de relação da taxa de incidência da dengue com os padrões encontrados nas variáveis analisadas;
- Dimensionar e detectar padrões nas variáveis analisadas utilizando mineração de dados através das redes neuronais do tipo *Self-Organizing Maps* (SOM);
- Analisar e identificar a existência de padrão na distribuição espacial e temporal da taxa de incidência da dengue correlacionada com as variáveis analisadas.



## 1.5 Estrutura

Nesse trabalho de mestrado são aplicadas diferentes técnicas e métodos provenientes de diversas áreas do conhecimento, sobretudo da Ciência da Informação Geográfica, epidemiologia e da mineração de dados.

Para a aplicação desses diversos conhecimentos de áreas tão abrangentes foi necessária a elaboração de uma estrutura subdividida em dois eixos: o embasamento teórico que subsidia os elementos para a construção da pesquisa científica e do referencial bibliográfico, e a aplicação prática que reproduz e descreve as metodologias utilizadas em etapas para se alcançar os objetivos propostos. Dessa forma essa dissertação perpassa por seis capítulos.

No primeiro e presente capítulo é apresentado o tema a ser investigado, o enquadramento metodológico e a estrutura desse trabalho.

O segundo capítulo descreve e conceitua um dos principais problemas de saúde pública da atualidade no Brasil: a dengue, uma doença infecciosa viral aguda.

No terceiro capítulo é descrito o comportamento espacial da dengue bem como a relação das variáveis determinantes da doença com o espaço geográfico.

No quarto capítulo o conceito de redes neuronais, especificamente as do tipo SOM, são explanados e as principais características sintetizadas.

No quinto capítulo é apresentada a metodologia empregada no trabalho, descrevendo as fontes, os dados, tratamento e a parametrização da rede neuronal SOM.

O sexto capítulo aborda a apresentação dos resultados e a análise desses resultados, que visam a identificação de *clusters* e padrões das variáveis analisadas e sua relação com a ocorrência de casos da dengue.

No sétimo capítulo são descritas as limitações relativas ao presente estudo.

O último capítulo apresenta as conclusões, bem como as discussões acerca das limitações e aplicações encontradas, propondo recomendações para futuros estudos e trabalhos.

## 2. Dengue

### 2.1 Introdução

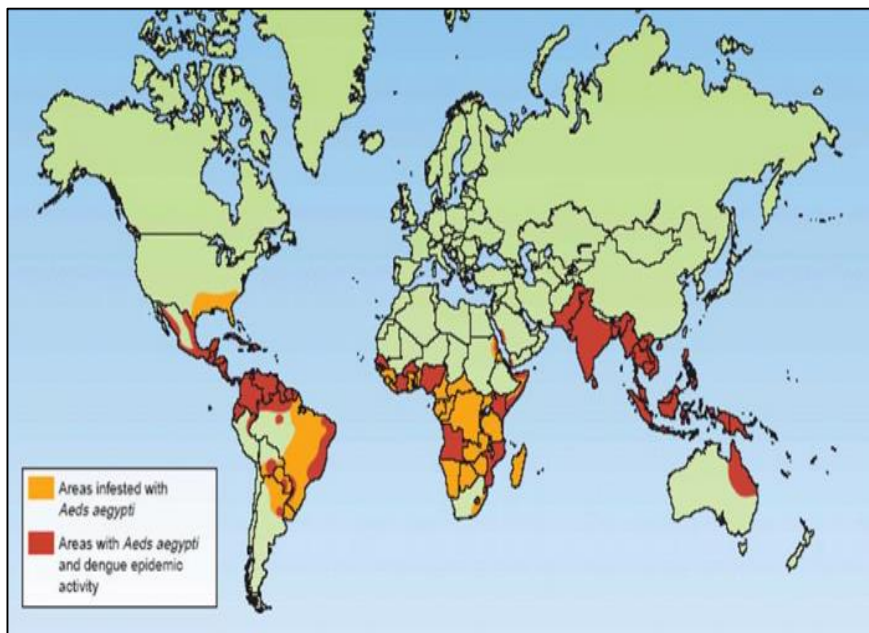
As doenças transmitidas por vetores constituem, ainda hoje, importante causa de morbidade no Brasil e no mundo. A dengue é considerada a principal doença reemergente nos países tropicais e subtropicais (TAUIL , 2001), onde aspectos do meio físico-biótico (ambientais) como o regime de chuvas, a temperatura, vegetação, coleções hídricas e a umidade favorecem a proliferação das populações de vetores e agentes infecciosos (CORREIA, MONTEIRO, *et al.*, 2007). Os padrões espaço-temporais da dengue, e de outras doenças transmitidas por vetores, estão relacionados a fatores de ordem biológica, geográfica, ambiental, social e econômica que atuam sinergicamente. Os fatores climáticos influenciam o ciclo de vida dos mosquitos transmissores de doenças como a dengue, entretanto, estes fatores não são determinantes para explicar a ocorrência da doença (REITER, 2001). A relação entre características ambientais e ocorrência de doenças é mediada pela estrutura demográfica da população e pelo perfil socioeconômico local, atuando sobre a suscetibilidade e as possibilidades de contato com vetores e reservatórios (WERNECK, RODRIGUES, *et al.*, 2002). Devem ser considerados ainda os hábitos culturais e comportamentais que contribuem para a ocorrência de padrões epidemiológicos diferenciados das enfermidades. Os primeiros relatos da dengue coincidem com o período de expansão geográfica do mosquito hematófago, *Aedes aegypti*, da África para os trópicos. O *Aedes aegypti* (FIG. 1) é o principal vetor envolvido na transmissão urbana do vírus da dengue ao hospedeiro humano (GUBLER, 1998). As características antropofílicas, o habitat urbano doméstico e a alta eficiência na transmissão do vírus da dengue, tornam o *Aedes aegypti* um vetor altamente competente para a disseminação da doença em ambientes urbanos e peri-urbanos. As espécies *Aedes albopictus* e o *Aedes polynesiensis* também podem sustentar a transmissão do vírus da dengue ao hospedeiro humano (GUBLER, 1998).



**FIGURA 1 – Imagem do mosquito *Aedes aegypti***

Fonte: Portal da Saúde: <http://portal.saude.gov.br>

O mosquito *Aedes aegypti* é altamente susceptível às temperaturas extremas e climas secos e encontra-se amplamente disseminado entre os limites de latitude 40°N e 40°S, oferecendo um grande risco de novas epidemias para as populações residentes nessas áreas (FIG. 2) (MACKENZIE, GUBLER e PETERSEN, 2004). Durante o repasto sanguíneo, fêmeas adultas de mosquitos do gênero *Aedes* transmitem o vírus da dengue ao hospedeiro humano. O ciclo de transmissão envolve a ingestão de sangue de um hospedeiro virêmico (i.e., com partículas virais circulantes no sangue periférico) pelo mosquito vetor e, após um período de incubação extrínseco, ocorre a transmissão do vírus a um novo hospedeiro vertebrado (GUBLER, 1998). O período de incubação extrínseco no vetor dura de 8 a 12 dias e é necessário para que ocorra a replicação e disseminação viral na hemocele e em todos os tecidos do inseto, até finalmente infectar as glândulas salivares e poder transmitir o vírus para um novo hospedeiro humano susceptível. O mosquito permanece infectado e assintomático durante toda sua vida e é capaz de transmitir o vírus da dengue para diversas pessoas, devido a sua capacidade de se alimentar em múltiplos indivíduos em sucessão (GUBLER, 1998). Outra forma importante de manutenção do vírus da dengue entre os mosquitos do gênero *Aedes* é a transovariana, em que o vírus é transmitido diretamente da fêmea infectada para a prole, dispensando o homem no ciclo mantenedor (HENCHAL e PUTNAK, 1990).

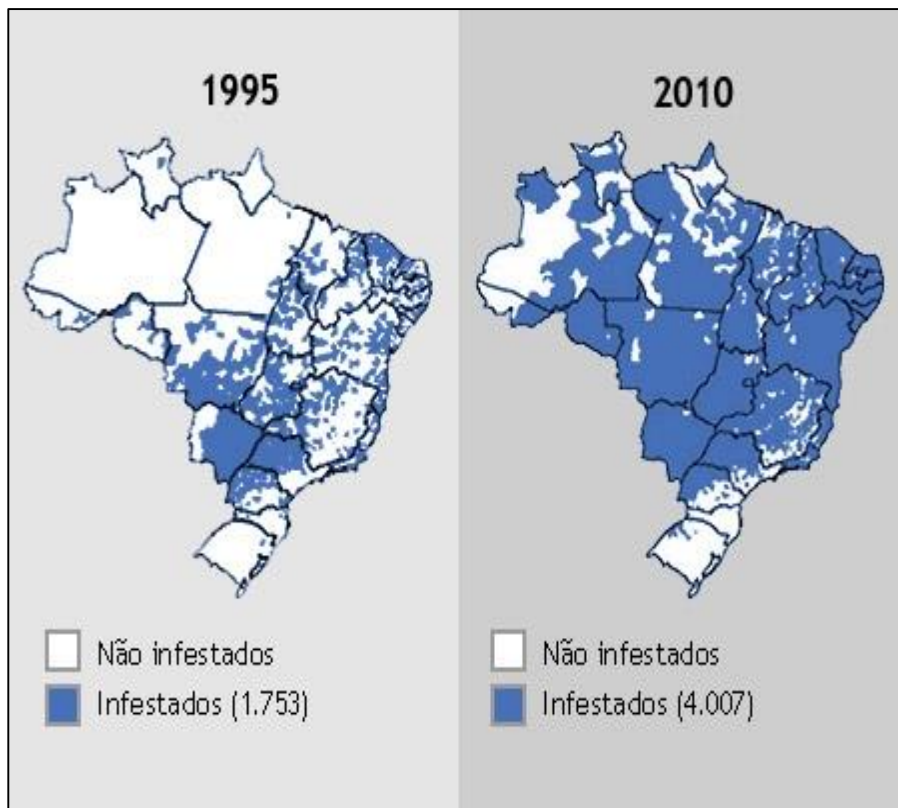


**FIGURA 2 – Distribuição global da dengue:** áreas infestadas com o mosquito vetor da dengue estão em laranja e as áreas com dengue epidêmica estão em vermelho. Fonte: Mackenzie et al., 2004

O hospedeiro humano torna-se virêmico após um período de incubação de 3 a 14 dias, após a picada por um mosquito infectado, concomitante ao início do período febril (GUBLER, 1998).

A manutenção da epidemia da dengue está relacionada à alterações demográficas e sociais ocorridas nos últimos 50 anos, que incluem o aumento da população global associada à

urbanização não-controlada, à falência dos programas de combate ao vetor e às mudanças nas políticas de saúde pública. Todos esses fatores têm contribuído para o aumento de epidemias da dengue, o desenvolvimento da hiperendemicidade e o aumento de casos de Febre Hemorrágica da dengue (GUBLER, 1998). A Organização Mundial de Saúde (OMS) tem demonstrado muita preocupação com o elevado número de epidemias, crescente hiperendemicidade e com o aumento da gravidade dos casos da dengue. No Brasil, a dengue é um dos principais problemas de saúde pública. Do ano de 1995 até 2010 observa-se o aumento expressivo da infestação pelo vetor transmissor da dengue, sobretudo nos estados de Rondônia, Acre, Minas Gerais, São Paulo, Paraná, Mato Grosso do Sul, Goiás (FIG. 3). Assim, estudos que contribuam para o monitoramento e controle dos fatores ambientais que possam estar associados com a disseminação da doença, devem ser utilizados pelos órgãos de saúde pública locais.



**FIGURA 3 – Mapa da infestação por município pelo vetor *Aedes aegypti* de 1995 a 2010 no Brasil.** Fonte:

Ministério da Saúde 2010.

## 2.2 Histórico da dengue no Brasil

As primeiras referências na literatura brasileira sobre a ocorrência da dengue podem ser encontradas em meados do século XIX. Entretanto, pelo comportamento clínico da doença ser muito similar com outras doenças causadas por arbovírus, especificamente do gênero Flavivírus, e da família Flaviviridae, sobretudo com a febre amarela, que no passado esteve amplamente presente no Brasil. Portanto, ao avaliarmos a questão histórica nesse prisma, e levando em consideração que a dengue tem um forte relacionamento com ambientes urbanos, a primeira epidemia de febre amarela urbana ocorreu em 1685, nas cidades de Recife, Pernambuco e Salvador, ainda no Brasil colonial sobre domínio do império português (FRANCO, 1969). Ainda durante o século XIX, podem ser encontrados indícios de doenças com características e sintomas semelhantes à dengue no Brasil. Especificamente no período entre 1846-1848, são verificados relatos de epidemias nos estados do Rio de Janeiro, Bahia, Pernambuco e algumas partes no norte do país (MARZOCHI, 1994).

A fase de expansão do vetor *Aedes aegypti* é iniciada entre os anos de 1850 a 1899 no Brasil império, fato comprovado pela expansão da febre amarela urbana no mesmo período. De acordo com Brasil (2001), “O *Aedes aegypti* se propaga pelo [império], seguindo caminhos de navegação marítima e fluvial, o que leva a ocorrência de epidemias de [febre amarela urbana] em quase todas as províncias [...], desde o Amazonas até o Rio Grande do Sul” (BRASIL, 2001, p.23). Nessa mesma época são encontrados mais relatos com sintomas semelhantes à dengue, pois a cidade de São Paulo foi atingida por uma epidemia da dengue no período entre 1851 até 1853 (CORDEIRO, 2008).

Os registros surgem novamente no século início XX, em 1916 novamente na cidade de São Paulo (LIMA, 1985). Em 1917 foi publicado um artigo que tratava da incursão de casos dentro do estado do Rio Grande do Sul, na cidade de Santa Maria. Já em 1923 pode ser encontrado um relato com a confirmação clínica realizada pelo médico Antônio Pedro (VASCONCELOS, 1999), na cidade de Niterói no estado do Rio de Janeiro. Esses teriam sido os últimos casos registrados da dengue no país, permanecendo assim sem registros da doença por um período de quase sessenta anos, até a reemergência em 1981 (TAUIL, 2001). Contudo, mesmo nesse período de ausência de registros existe uma suspeita de circulação viral. Um inquérito sorológico realizado na Amazônia nos anos 50 detectou marcas de infecções por vírus da dengue, corroborando com a suspeita (DONALISIO, 1999).

Alguns anos antes da reemergência da doença no país pode ser observada em 1976 uma reinfestação pelo principal vetor da doença no Brasil (*Aedes aegypti*), dessa vez definitiva, tendo como porta de entrada a cidade de Salvador na Bahia. Em 1977 o vetor também foi identificado no Rio de Janeiro, e posteriormente nos anos seguintes em algumas capitais e grandes centros nordestinos (FIG. 4). No início da década de 80, período coincidente com a reemergência da doença, algumas cidades fronteiriças com o Paraguai foram infestadas (FIG. 5). No Norte do Brasil alguns municípios limítrofes a Venezuela e as Guianas (FIG. 6) também se tornaram foco do vetor (CATÃO, 2012).

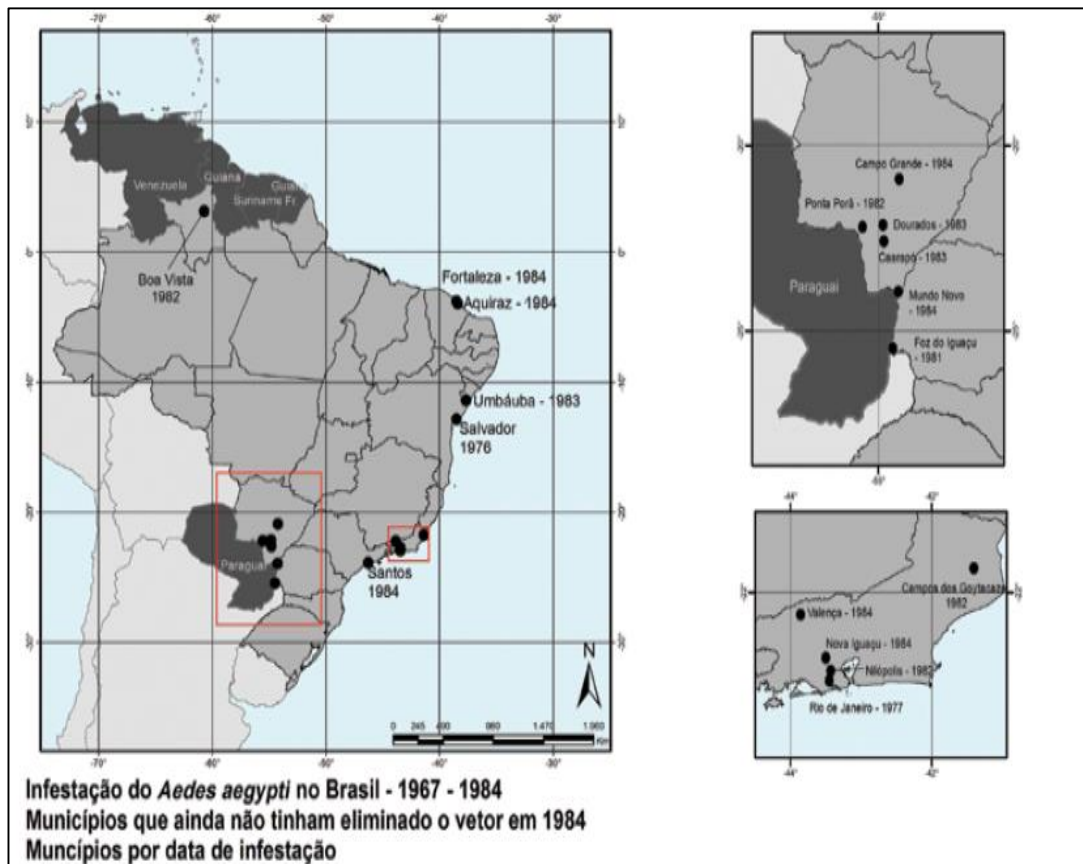


FIGURA 4 – Mapa da infestação pelo *Aedes aegypti* no Brasil, período entre 1967 até 1984. Fonte: Lima (1985), adaptado de Catão (2012)

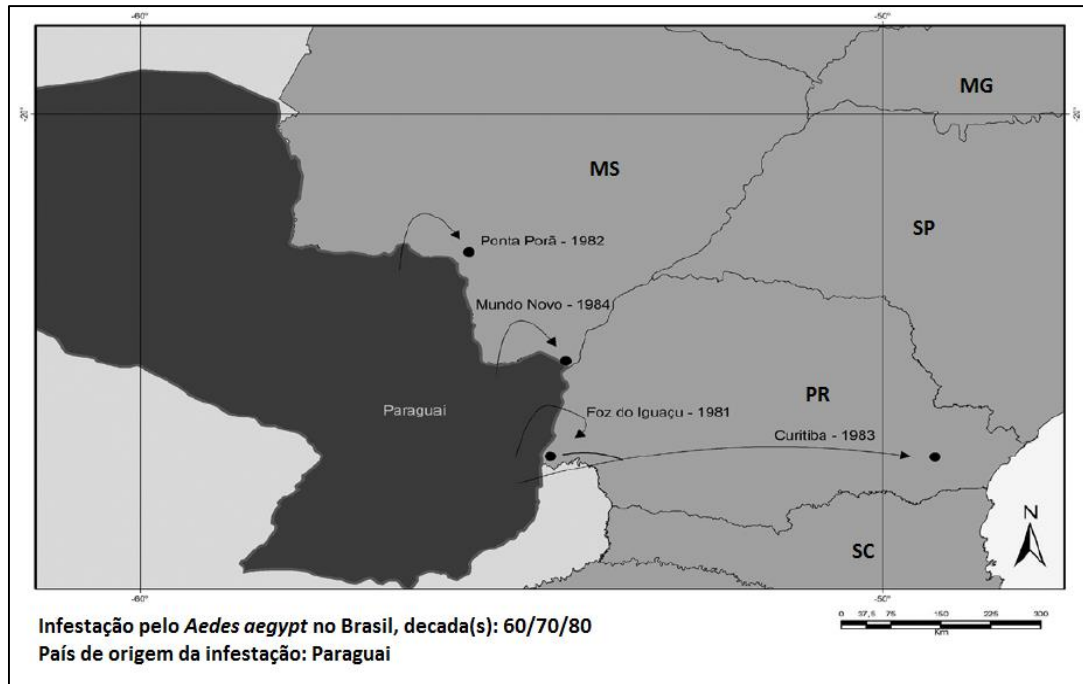


FIGURA 5 – Mapa da infestação pelo *Aedes aegypti* no Brasil, período entre 1965 até 1985, origem Paraguai.

Fonte: Lima (1985), adaptado de Catão (2012)

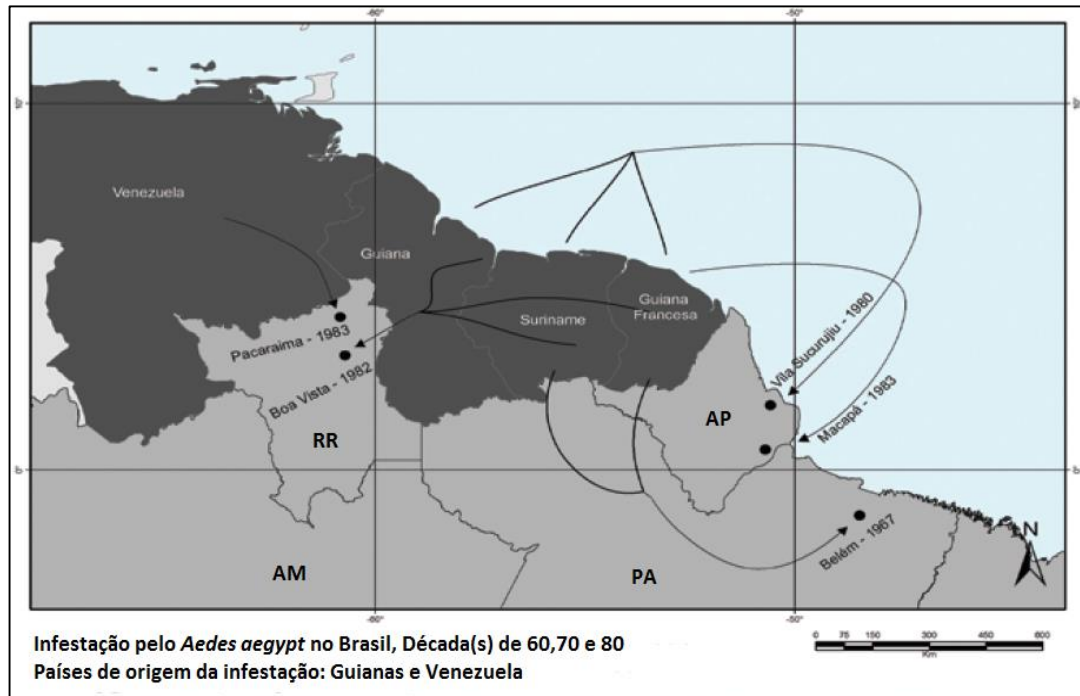


FIGURA 6 – Mapa da infestação pelo *Aedes aegypti* no Brasil, período entre 1965 até 1985, origem Guianas

e Venezuela. Fonte: Lima (1985), adaptado de Catão (2012)

### **2.3 Aplicação dos Sistemas de Informação Geográfica**

O uso de técnicas de geoprocessamento em estudos epidemiológicos da dengue tem se tornado cada vez mais frequente. Os Sistemas de Informação Geográfica demonstram ser uma importante ferramenta nos estudos dessa doença, uma vez que ela apresenta relações com as condições ambientais e socioeconômicas. Os estudos que utilizam geoprocessamento podem fornecer informações e evidências para a identificação e mapeamento dos habitats potenciais para os mosquitos vetores da doença, prever alterações nas populações dos mosquitos vetores, monitorar quantitativamente e qualitativamente alterações nos respectivos *habitats* e identificar áreas de maior e menor risco da doença (HUGH-JONES, 1989). Além disso, permitem também monitorar variáveis ambientais no tempo e no espaço em diversas escalas e subsidiar as ações de vigilância, planejamento e controle da dengue. Ferramentas de mineração de dados também podem auxiliar na identificação de comportamento e padrão espacial na distribuição dos casos de infecção, bem como identificar as possíveis correlações entre variáveis econômicas, sociais e ambientais que possam também influenciar a distribuição dos casos de infecção pelo vírus da dengue.

Neste contexto, torna-se imprescindível o emprego das técnicas oferecidas pela ciência da informação geográfica em conjunto com os SIG auxiliado por ferramentas e técnicas de data mining, para elucidar o perfil da distribuição espaço-temporal da população infectada pela dengue em uma dada região ou localidade.



### **3. Comportamento geográfico da dengue e sua relação com variáveis determinantes**

#### **3.1 Introdução**

O vírus da dengue é transmitido a partir de um inter-relacionamento estabilizado do vírus, vetor, pessoas (infectadas e susceptíveis) e o espaço geográfico. Porém, existem muitos outros fatores que contribuem para o aumento de ocorrência e do risco da doença, principalmente os de caráter espacial (sociais e ambientais), e esses atuam de forma ativa na distribuição desigual da dengue no espaço geográfico. De acordo com a Organização Pan-Americana de Saúde (OPAS) (1997):

*O risco é um conceito utilizado para medir a probabilidade de um evento futuro com consequências negativas como uma infecção pelo vírus da dengue ou um surto. A probabilidade depende da presença de uma ou mais características ou determinantes do evento. A dinâmica da transmissão dos vírus da dengue é determinada pela interação do ambiente, do agente, da população de hospedeiros, e de vetores, que existem juntos num habitat específico. A magnitude e a intensidade da interação irão definir a transmissão da dengue em uma comunidade, região ou país (OPAS, 1997, p.19).*

Dessa forma, para uma melhor compressão da dengue no território brasileiro se faz necessário identificar e localizar as suas determinantes no espaço, bem como entender a maneira com que essas se combinam em cada localidade, desta maneira é possível ampliar o entendimento da doença. Para uma melhor compreensão da doença os fatores determinantes podem ser subdivididos em macro e micro determinantes (ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE, 1997). Nos macrodeterminantes se encontram as áreas geográficas onde o vetor se desenvolve e entra em contato com os hospedeiros. Podem ser relacionados diversos componentes ambientais que possibilitam a ocorrência do vetor como latitude, altitude, temperatura, umidade relativa do ar (TAB. 1). Outro ponto chave a ser observado é de que se esses componentes forem ausentes em determinada zona, os mesmos podem servir como barreira de transmissão, limitando e até mesmo restringindo o ciclo de vida do vetor, como no caso de altas altitudes ou de baixas temperaturas. Porém em países e regiões tropicais esses fatores de risco estão presentes em quase toda extensão territorial, e sozinhos não conseguem explicar a distribuição espacial da doença (CATÃO, 2012). Também estão ligados nesse fator os componentes sociais que são relativos à produção do espaço sobretudo ao espaço urbano.

TABELA 1

Macro e Micro determinantes de transmissão da dengue.

Fator	Macro determinante		Micro determinante		
	Ambiental	Social	Hospedeiro	Agente	Vetor
Componente	*Latitude; *Longitude; *Altitude; *Amplitude térmica; *Umidade relativa do ar;	*Densidade populacional; *Condições sanitárias e socioeconômicas; *Educação;	*Sexo; *Idade; *Condições imunológicas e de saúde; *Ocupação;	*Nível da viremia;	*Densidade de fêmeas adultas; *Abundância e locais de criação do mosquito; *Frequência de alimentação; *Hospedeiro preferencial; *Disponibilidade de hospedeiros; *Suscetibilidade inata à infecção;

Fonte: Adaptado de Catão (2012, p123 -124).

Já os microdeterminantes estão relacionados com o agente etiológico (tipos virais da dengue), aos vetores (*Aedes aegypti* e *Aedes albopictus* no Brasil) e aos hospedeiros (humanos imunes ou não) (ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE, 1997).

É ainda importante ressaltar que alguns autores destacam a importância da associação entre a frequência e distribuição dos casos da dengue com a plasticidade e o poder de adaptação do vetor, especificamente o mosquito *Aedes aegypti*, a um ambiente mais urbano e com maior concentração de seres humanos (BARCELLOS e BASTOS, 1996). É necessário também destacar a importância de se focar a análise sobre a questão da produção dos vetores e de disponibilidade dos criadouros, tendenciada pela baixa educação e conhecimento da população, baixa infraestrutura urbana e falhas no controle vetorial (BARCELLOS, 2003). E que também os fatores climáticos e orográficos (precipitação, umidade relativa, amplitude térmica, altitude, etc.) são chaves para delimitar a intensidade das epidemias e densidade populacional dos vetores.

### **3.2 Fatores determinantes da dengue no Brasil**

Os estudos ambientais correlacionados com os de saúde vêm demonstrando sua efetividade no estabelecimento de causas, impactos e distribuição espacial dos mais diversos tipos de enfermidades dos seres humanos. Na maioria das vezes, os métodos tradicionais de pesquisa não conseguem estabelecer relações inteiramente confiáveis, o que acaba dificultando a espacialização e a distribuição das doenças em uma localidade. Este fato prejudica a aplicação de técnicas adequadas e inovadoras para o combate ou erradicação das doenças, para futuras medidas preventivas e corretivas. Por esse motivo grande parte das pesquisas epidemiológicas de caráter ecológico no Brasil utilizaram análises baseadas no emprego dos sistemas de informação geográfica na identificação de variáveis chave, no que concerne aos fatores determinantes da dengue (FILHO e ROUQUAYROL, 2006). Para tais análises, destaca-se a aplicação dos SIG, agregando dados ambientais e sociais em conjunto com dados de saúde para promover a qualidade da caracterização de cada um destes fatores (FLAUZINO, SOUZA-SANTOS e OLIVEIRA, 2009). Na maioria desses estudos foram empregadas como variáveis explanatórias sexo, idade, densidade populacional, situação/estado civil, condição de moradia, nível de instrução, renda, instalação sanitária de água e de esgoto, lixo (cobertura do serviço). Assim a dengue pode ser enquadrada como uma doença de transmissão essencialmente urbana.

A alta taxa de crescimento urbano no Brasil aliada a dinâmica de uso e ocupação do solo em grandes centros urbanos, que por sua vez reflete um uso não homogêneo e com formas de ocupação econômico-social bem distintas, determina fortemente as condições ambientais (moradia), adensamento populacional e saneamento ambiental. Essas condições são fatores de risco para ocorrência da dengue e estão fortemente correlacionadas com os dados relativos à incidência da doença, que podem auxiliar na detecção dos surtos epidêmicos no Brasil (DA COSTA e NATAL, 1998). Os dados de incidência mais utilizados podem ser encontrados através do serviço de notificação de casos por meio de um serviço de saúde (dados secundários), e os dados obtidos através de inquéritos soropidemiológicos (dados primários). Entretanto, existe um consenso de que o número de casos notificados da dengue no país não representa a realidade, essa subnotificação reside no fato de que a doença pode se manifestar de forma branda e assintomática em alguns indivíduos. Já os inquéritos são mais precisos e identificam com maior exatidão os indivíduos acometidos pela doença, contudo esses são considerados muito dispendiosos e são realizados em amostras populacionais (VASCONCELOS, 1999).

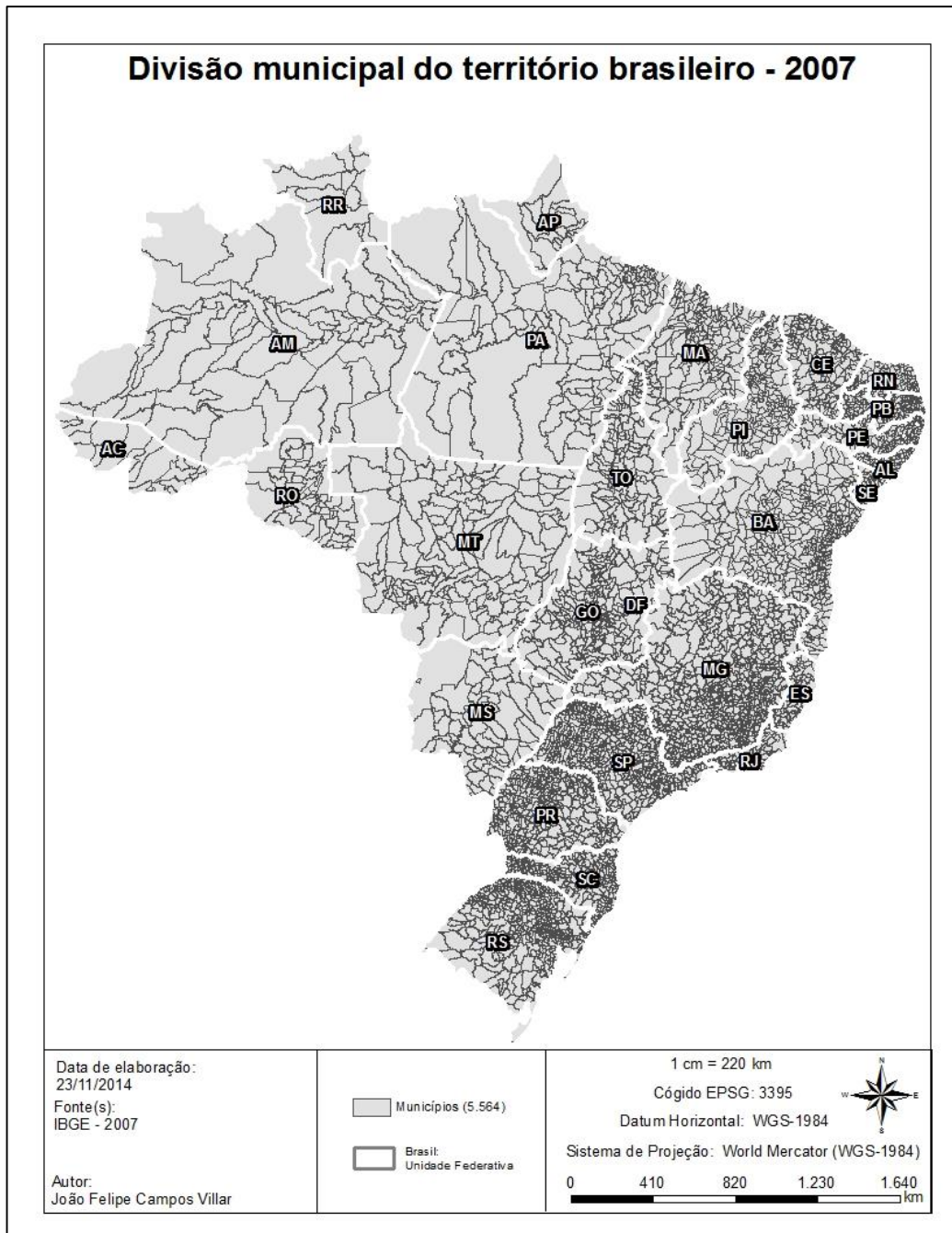
Ao se escolher uma determinada fonte de dados em detrimento de outra se obtém uma análise de distintas realidades, essas podem ser observadas nos resultados de uma dada

pesquisa, tanto pela quantidade de pessoas infectadas quanto pelos principais determinantes envolvidos. Por essa razão é extremamente necessária a verificação das notificações e registros de saúde em consonância com a sua respectiva análise de distribuição espacial, para que seja possível uma efetiva correlação com fatores determinantes que também possuem uma forte relação com o componente geográfico.

### **3.3 Relação das variáveis determinantes com o espaço**

As ferramentas de geoprocessamento e de cartografia digital são amplamente empregadas para a seleção e posterior análise dos fatores determinantes da dengue, conforme vasta pesquisa de artigos, dissertações, teses e livros na literatura científica brasileira (CATÃO, 2012). Uma grande parte dessas publicações estabelece correlações entre a taxa de incidência da doença em determinado local (extensões que podem variar de escala local, regional e nacional) com a presença ou ausência dos fatores determinantes da dengue. Outro aspecto relevante é que a análise em diferentes escalas geográficas evidenciou diferentes correlações nos fatores determinantes da dengue. Em alguns trabalhos foi identificado risco em populações com melhores indicadores socioeconômicos na região Nordeste do Brasil. Em outros, com nível de agregação mais local, foi observada uma correlação positiva dependendo do nível de escolaridade. Já em alguns estudos que analisaram regiões densamente povoadas ou ao longo dos eixos de transporte público, foi possível estabelecer uma correlação positiva com os surtos da doença. Outros ainda encontraram maior risco em áreas com melhores condições socioeconômicas. Contudo a maioria dessas pesquisas utiliza a escala local com unidades de agregação dos dados para pequenas extensões territoriais (distrito, bairros, setores, etc.). Com exceção de dois artigos que possuem como enfoque o município para unidade de análise. Em escalas locais são muitas as particularidades de cada região, resultado direto do elevado número de determinantes que atuam de diferentes formas de acordo com a escala e se combinam de maneiras diferenciadas em cada localidade. Sendo assim é extremamente necessária uma visão mais abrangente para que seja possível a observação do comportamento em conjunto dos fatores determinantes da dengue, mesmo que isso implique em um primeiro momento em perda de detalhes.

Em escala nacional é necessária uma abordagem de visão do conjunto do problema, analisando como se distribuem e como se combinam os principais fatores determinantes da dengue. Para tal é preciso utilizar diversos indicadores que são disponibilizados através da unidade de análise município no Brasil. O município é a menor unidade política-administrativa existente no Brasil, sendo todo o território dividido em municípios (FIG. 7), com exceção de algumas áreas como o distrito federal (BRASIL, 1988). Por essa razão a maioria dos indicadores e ações de vigilância sanitária e epidemiológica utilizam o município como unidade mínima de análise.



**FIGURA 7 – Divisão municipal do Brasil conforme resolução do IBGE em 2007. Fonte: IBGE (2007)**

Dessa forma os fatores determinantes mais significativos por município (CATÃO, 2012) são densidade demográfica, população, temperatura, amplitude térmica, mobilidade urbana, renda, nível de escolaridade e condições sanitárias.

### **3.4 Relação entre as variáveis determinantes com a taxa de incidência**

A maioria dos estudos e pesquisas na literatura brasileira que buscam a correlação dos determinantes com a taxa de incidência da dengue utilizam os dados secundários provenientes dos órgãos oficiais de saúde (FLAUZINO, SOUZA-SANTOS e OLIVEIRA, 2009). As variáveis socioeconômicas comumente utilizadas para estabelecer esse relacionamento são aquelas provenientes dos indicadores socioeconômicos preconizados pela Rede Interagencial de informações para Saúde (RIPSA-REDE INTERAGENCIAL DE INFORMAÇÕES PARA SAÚDE, 2008). Já as variáveis ambientais são aquelas preconizadas pela Fundação Nacional de Saúde (FUNASA) e estão relacionadas com a drenagem urbana, limpeza pública, e pontos estratégicos para a vigilância do vetor (BRASIL, 2001). De acordo com FLAUZINO, ET AL (2009 (FLAUZINO, SOUZA-SANTOS e OLIVEIRA, 2009)), os indicadores utilizados nesses estudos foram sexo, idade, densidade demográfica, situação civil, razão de dependência, razão de sexos, condições de moradia, instrução, renda, índice de pobreza, instalação sanitária de água e esgoto, coleta de lixo, índice de infestação e criadouros do vetor, índice pluviométrico, umidade relativa, temperatura e estação climáticas no momento do estudo. Alguns desses estudos identificaram como fatores de risco para taxa de incidência elevada a faixa etária e o sexo. Um fator de risco para incidência mais elevada, também pode ser identificado em áreas com maior concentração de indivíduos, grande diversidade econômica, forte adensamento populacional, alta densidade domiciliar, alta densidade de lotes vagos, grandes vias de circulação e de pessoas, falta de oferta de emprego e áreas deficientes em coleta de esgoto e lixo. Os estudos que tiveram como foco o município como unidade de análise espacial mostram resultados similares evidenciando um maior risco em faixas etárias mais elevadas, não registrando diferença de risco para os sexos masculino e feminino. Conforme Torres (2008), as populações que são afetadas pela dengue em regiões não indenes anteriormente, são populações de faixas etárias mais elevadas e mulheres por conta da característica domiciliar do vetor e essa população permanecer mais tempo diário nesse tipo de ambiente. Em relação a variável ambiental é possível encontrar associação de alta taxa de incidência com a presença de criadouros para o vetor. Portanto, a maioria dos estudos evidenciam que a dengue é uma doença relacionada diretamente com as características sócio ecológicas do ambiente, para além das características do indivíduo, que também é uma componente chave para o estabelecimento de relações com a taxa de incidência.

## 4. Rede neural SOM

### 4.1 Introdução

Qualquer rede neural artificial (RNA) é baseada em um modelo de rede biológico, mas isso não quer dizer necessariamente que os modelos artificiais possuem comportamento idêntico ao biológico. Entretanto, diversas características provenientes das redes biológicas podem ser aplicadas à um modelo RNA. No neurônio natural ou biológico, os sinais elétricos e químicos são recebidos e processados, através das sinapses (FIG. 8). As sinapses são pontos de ligação entre os neurônios, alguns podem possuir até 200.000 sinapses. Os sinais elétricos são recebidos por um neurônio através das sinapses e esse pode disparar ou não seu sinal através de um axônio que frequentemente está conectado a milhares de outros neurônios. A sinapse está localizada entre o axônio do neurônio que envia um sinal e do dendrito do neurônio que recebe o sinal. Dessa forma todo neurônio pode ser considerado um mecanismo de processamento e de armazenamento da informação (SCHULLER, 2000).

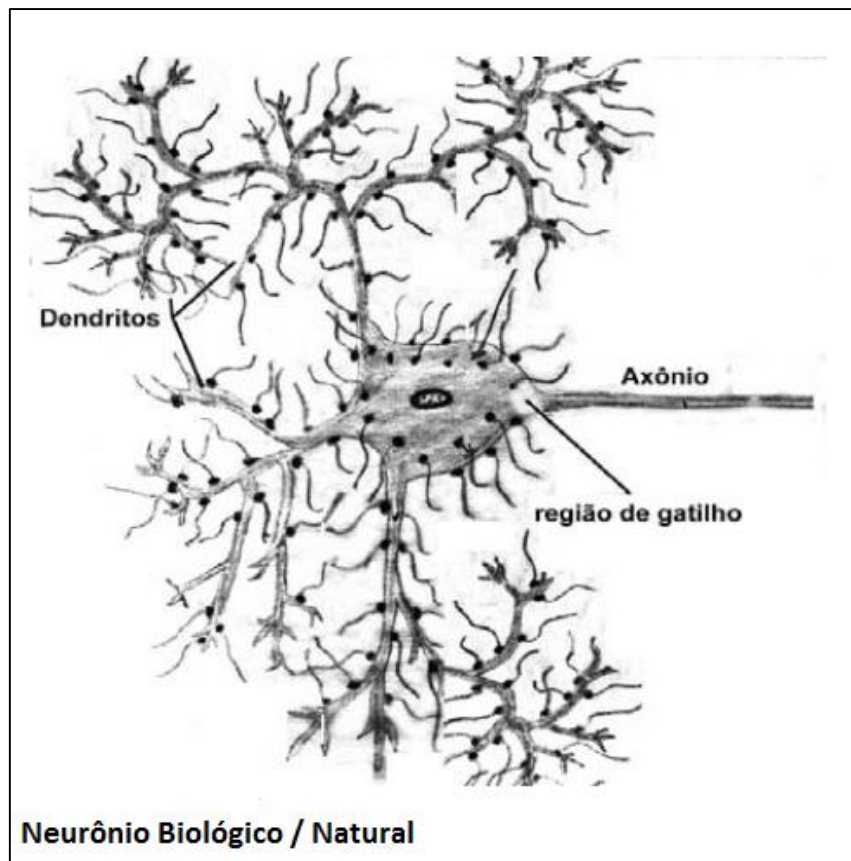
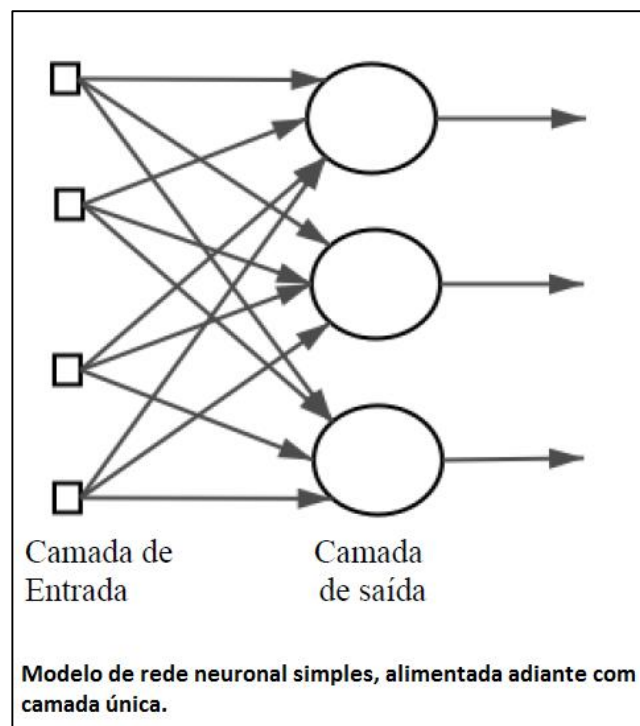


FIGURA 8 – Representação gráfica de um neurônio biológico ou natural e seus principais componentes.

Fonte: Schuler (2000), adaptado de Schuler (2000)

O gigantesco desenvolvimento na área da ciência da informação nas últimas décadas, que possibilitou o compartilhamento e o armazenamento de dados e informações nunca antes imaginados, aliado com as características de manipulação do sinal elétrico pelos neurônios, foram desenvolvidas as RNAs. Tais redes foram desenvolvidas no intuito de responder ao grande desafio de hoje na área dos sistemas de informação, que é a transformação dos dados em informação e da informação em conhecimento útil aplicado à um tema específico.

Fundamentalmente as RNAs nada mais são que modelos matemáticos inspirados na metodologia de funcionamento de neurônios em um cérebro humano, sobretudo na forma com que esse complexo sistema se organiza e se adapta para o aprendizado de um problema específico. A base de operação de um modelo mais simples se baseia em um conjunto de conexões entre unidades computacionais, denominadas “neurônios”, formando uma rede neural (CARDON e MÜLLER, 1994). A rede está estruturada em camadas, sendo duas as principais camadas operativas, podendo existir mais camadas, dependendo da complexidade do problema a ser respondido (FIG. 9). A primeira camada denominada entrada ou *input*, que é formada por neurônios tem como função principal a apresentação dos dados de um determinado problema à rede. A segunda é a camada de saída ou *output*, constituída de neurônios que geram o resultado da rede.



**FIGURA 9 –Modelo de rede neural simples, alimentada adiante com camada única.** Fonte: Schuler (2000), adaptado de Schuler (2000)



A grande semelhança de uma rede artificial com uma rede natural está no fato que o complexo ajustamento da rede depende do curso de um processo de aprendizagem. Portanto as RNA's herdaram das redes biológicas a capacidade de aprender através de exemplos, construindo interpolações através desse aprendizado (BARRETO, 2002).

Dessa maneira a rede neural realiza uma adaptação de suas conexões de acordo com a informação adquirida pelo processo de aprendizagem. Portanto, para se aplicar na prática um estudo levando a cabo uma rede neural como ferramenta de mineração de dados ou *data mining*<sup>1</sup>, se faz necessária a execução do processo de aprendizagem.

As RNAs possuem tipos diferentes de aprendizagem e esses são classificados em dois grandes grupos. Um grupo engloba as redes de aprendizagem supervisionada, onde os exemplos de problemas apresentados a rede são inseridos no conjunto de dados de entrada, os quais determinam um resultado nos dados de saída que já são previamente conhecidos, de acordo com Barreto (2002) pode ser definido como:

*Neste caso o 'professor' indica explicitamente um comportamento bom ou ruim. Por exemplo, seja o caso de reconhecimento de caracteres e para simplificar seja reconhecer entre um A ou X. Escolhe-se uma rede direta, com dois neurônios na camada de saída, uma ou várias camadas internas e um conjunto de neurônios na camada de entrada capaz de representar com a precisão desejada a letra em questão. Apresentam-se estas letras sucessivamente a uma retina artificial constituída por uma matriz de elementos fotossensíveis, cada um ligado a um neurônio de uma RNA direta (feedforward). Observa-se qual dos dois neurônios de saída está mais excitado. Se for o que se convencionou representar a letra que for apresentada nada deve ser corrigido, caso contrário modifica-se os valores das conexões sinápticas no sentido de fazer a saída se aproximar da desejada. Foi exatamente isto que fez Rosenblatt com o seu Perceptron. Como a cada exemplo apresentado uma correção é introduzida depois de observar a saída da rede este é um caso de ensino supervisionado (BARRETO, 2002, p.23).*

Já no grupo da aprendizagem não supervisionada não são utilizados dados de exemplo para efeito de comparação nos dados de saída com os dados de entrada, ou seja, a rede não possui nenhum resultado previamente conhecido. A rede utiliza um tipo de treino que força a rede a adaptar-se às idiosincrasias dos dados num processo não orientado, fazendo com

---

<sup>1</sup> "Data Mining is a decision support process where we look in a large data bases for unknown and unexpected patterns of information" (FERRUZZA, 1998).

que sejam encontrados nos próprios dados relações estruturais.

Estas relações podem ser encontradas através de medidas de similaridades observadas em alguns dados de exemplo da própria rede, conforme Barreto (2002):

*Aprendizado não Supervisionado é quando para fazer modificações nos valores das conexões sinápticas não se usa informações sobre se a resposta da rede foi correta ou não. Usa-se por outro lado um esquema, tal que, para exemplos de coisas semelhantes, a rede responda de modo semelhante. Aprendizado não supervisionado se chama também descobridor de regularidades ou redes auto organizadas devido à propriedade básica de seu funcionamento (BARRETO, 2002, p.23).*

Essa capacidade de aprendizagem faz com que a rede seja customizável para diferentes tipos de aplicação, sobretudo para o reconhecimento de padrões ou classificação de dados (GURNEY, 1997). As RNAs não se constituem de um método apenas, mas toda uma família de modelos aplicados. Um ótimo exemplo de rede que realiza uma projeção não linear de dados de entrada multidimensionais em um arranjo de neurônios (unidades) é o *Self-Organizing Map (SOM)* ou mapa auto organizável de Kohonen (HENRIQUES, 2010).

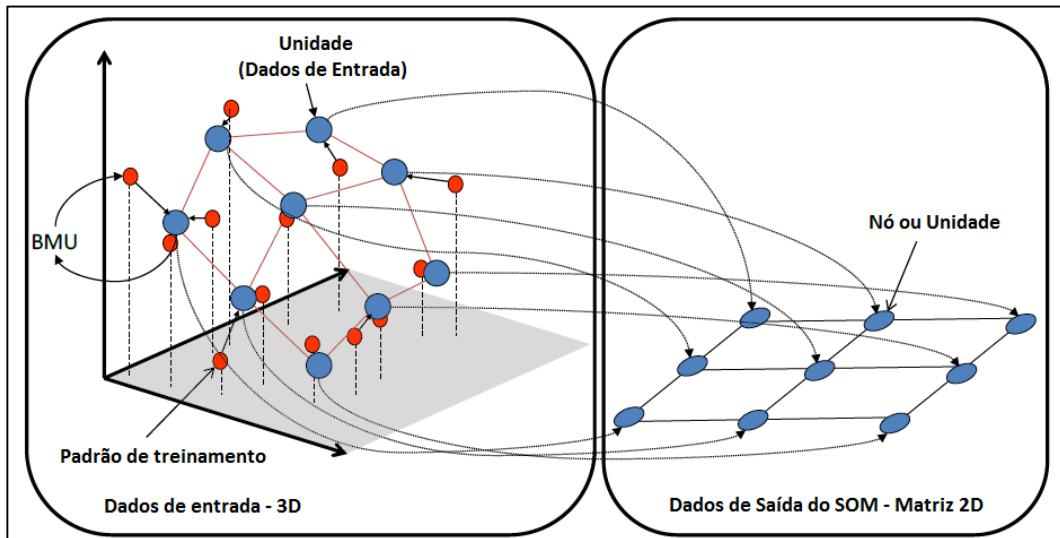
Um SOM caracteriza-se por ser uma rede de aprendizagem não supervisionada, que tem como ideia básica o mapeamento de dados multidimensionais em uma ou duas dimensões, mantendo as relações topológicas de acordo com os padrões nos encontrados dados. Seu objetivo elementar é o de extrair e ilustrar as estruturas essenciais em um conjunto de dados, através de um mapeamento resultante de um processo de classificação não supervisionada (KANSKI, NIKKILÄ e KOHONEN, 1998).

Os *Self-organizing maps (SOMs)*, foram propostos por Teuvo Kohonen no início da década de 80 (KOHONEN, 1982), sendo aplicado como uma ferramenta para a visualização de dados com elevada multidimensionalidade, ele implementa um mapeamento ordenado de uma distribuição de alta dimensionalidade em uma matriz regular de baixa dimensionalidade.

O SOM também realiza a conversão de relações estatísticas complexas e não lineares entre itens de dados multidimensionais em relações geométricas simples utilizando uma visualização de baixa dimensionalidade. Dessa maneira, é feita a compressão da informação preservando-se as mais importantes relações topológicas e métricas dos itens inerentes aos dados primários nessa visualização, sendo útil para a produção de algumas abstrações dos dados analisados (KOHONEN, 1999).

Essas duas interfaces visualização e abstração, podem ser aplicadas de diversas maneiras em tarefas computacionais complexas tal como análise de processo, percepção da máquina, controle e comunicação.

Basicamente um SOM é constituído por uma matriz regular bidimensional (2D) composta por nós (neurônios ou unidades) (FIG. 10), proveniente dos dados multidimensionais de entrada (3D ou +).



**FIGURA10 – Desenho esquemático do fluxo de funcionamento de uma rede SOM e seus constituintes básicos: dados de entrada (3D ou +) e Dados de Saída (2D).** Fonte: Henriques (2010), adaptado de Henriques (2010)

Uma das principais particularidades do SOM é a possibilidade de se criar uma ordenada segmentação topológica utilizando as observações organizadas em segmentos sobre uma matriz regular mantendo preservada a relação de vizinhança entre as mesmas (FIG. 11).

Pode ser observado que a posição 1 está próxima da 2 e que as duas estão distantes da posição 9, posição mais extrema e inversamente proporcional a 1 (SKUPIN e AGARWAL, 2008).

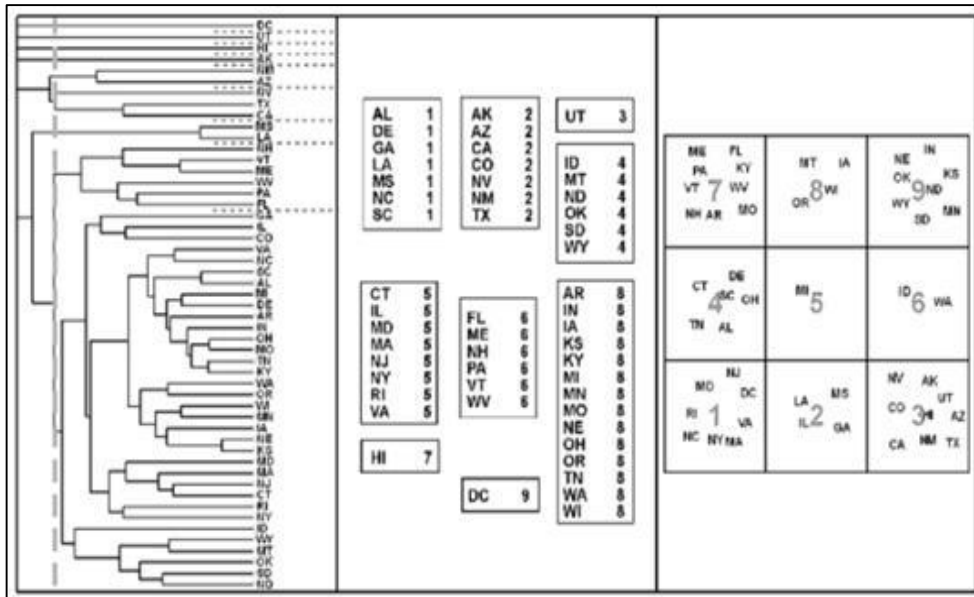


FIGURA 11 – Esquema comparativo de diferentes métodos de segmentação (Da esquerda para a direita: Hierárquica, Média k e SOM). Fonte: Skupin e Agaward (2010)

A elaboração de um espaço ordenado topologicamente através das observações iniciais pode ser vista como a representação das observações em um espaço de dimensionalidade menor. Ou seja, o SOM pode ser considerado um método de redução do espaço dos dados de entrada, sendo assim as variáveis iniciais são reduzidas conservando a fiabilidade do resultado gerado (dados de saída). O SOM se diferencia dos demais métodos de *clustering*<sup>2</sup> pela propriedade que o mesmo tem em não considerar as observações iniciais como objetos discretos mas como amostras de um espaço n-dimensional, concebido através de treinamento sobre os próprios dados de entrada, gerando no final um mapa discreto de dimensão menor que representa esse espaço n-dimensional inicial (SKUPIN e AGARWAL, 2008).

<sup>2</sup> *Clustering* ou Clusterização é o ato de se agrupar objetos de uma dada base de dados, de modo que objetos mais similares fiquem no mesmo cluster e objetos menos similares sejam alocados para clusters distintos (OCHI, DIAS e ALVES, 2004).

## 4.2 Algoritmo SOM e sua parametrização

O Algoritmo da rede SOM leva em consideração duas definições chaves para seu correto emprego: o tamanho da rede e a sua topologia (KOHONEN, 1999). Assim sendo, o primeiro parâmetro a ser definido em um SOM é o tamanho da rede com sua respectiva topologia, o que corresponde ao número de unidades (neurônios) que serão utilizadas e da dimensão da matriz que será utilizada. Por exemplo, para se obter uma rede com um total de 18 neurônios de duas dimensões (2D) deverá ser utilizado um SOM de 3 x 3. A definição da dimensão da matriz topológica pode ser algumas vezes difícil de ser definida devido a subjetividade inerente a esse processo, entretanto podem ser utilizadas duas metodologias para auxiliar essa definição: o *k-means* SOM, na qual o número de neurônios da rede deverá ser igual ao número de segmentos esperados e o SOM emergente onde é utilizado um elevado número de neurônios para a obtenção de uma matriz com segmentos mais bem definidos (DOS SANTOS, 2013).

Na topologia são estabelecidas as relações de vizinhança entre os neurônios da rede, sendo a topologia retangular a mais amplamente empregada. Nesse tipo de topologia cada neurônio possui 4 vizinhos. Outro exemplo topológico bastante difundido é o da topologia hexagonal, onde cada neurônio possui outros 6 neurônios vizinhos (DOS SANTOS, 2013).

A topologia traduz as relações de vizinhança entre os neurônios sendo frequentemente usadas a topologia retangular ou topologia hexagonal (FIG. 12). Numa topologia retangular cada neurônio tem 4 neurônios vizinhos e na topologia hexagonal a vizinhança é formada por 6 neurônios (DOS SANTOS, 2013).

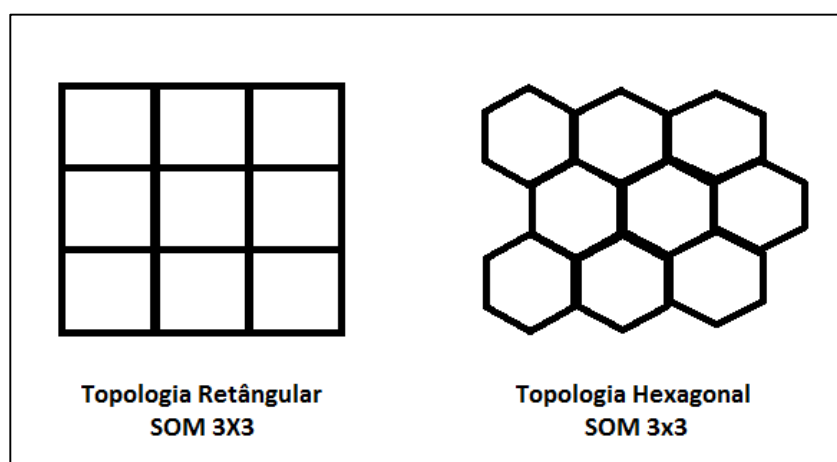


FIGURA 12 – Tamanho e topologia de um SOM 2D 3x3 retangular e Hexagonal, da esquerda para direita.

Fonte: dos Santos (2013), adaptado de dos Santos (2013)

Depois de definida a topologia da rede cada neurônio recebe um vetor com valores que terá a mesma dimensão da quantidade de variáveis utilizadas nos dados de *input*. Se pretendemos utilizar 5 variáveis a serem utilizadas na rede, cada neurônio trabalhará com um vetor de 5 dimensões. Esse vetor é construído de maneira randômica, uma vez que as propriedades dos dados de entrada serão utilizadas pelos neurônios no processo de aprendizagem. Entretanto, podem ser atribuídos valores específicos para os neurônios com o objetivo de melhorar a aprendizagem do algoritmo (KOHONEN, 2001). O algoritmo também pode ser facilmente descrito através da seguinte função:

*For all training patterns*

*Compute the distance to all units*

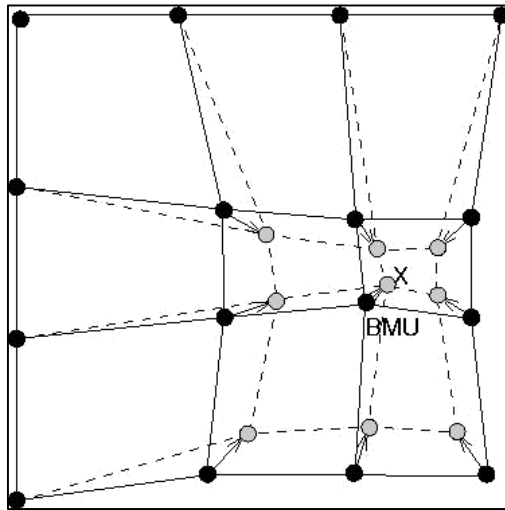
*Find the closest unit*

*Update that unit and its neighbors*

*Repeat this process until a given stopping criteria is met*

Como pode ser observado na função, é necessário realizar um processo de treino antes de se iniciar o cálculo da rede, e é nele que o segundo parâmetro principal (taxa inicial de aprendizado) vai ser definido no SOM (HENRIQUES, 2010).

O treino nada mais é que um processo interativo onde cada observação é selecionada de maneira aleatória e apresentada para a rede neural, para isso o neurônio mais similar a essa observação ou *best matching unit* (BMU) deve ser encontrado para serem ajustados os vetores desse BMU e dos respectivos neurônios vizinhos mais próximos (FIG. 13). Esse ajustamento consiste em se deslocar os neurônios em direção da observação utilizada, fazendo com que essa seja melhor representada na rede (VESANTO, 1997).



**FIGURA 13 – Figura esquemática apresentando o ajustamento do *BMU* e de sua vizinhança através dos dados de entrada (X). Fonte: Vensato (1997)**

O BMU pode variar de 0 até 1, e pode ser ajustado de acordo com a taxa de aprendizado, sendo que uma taxa maior implica diretamente em um BMU com direção mais próxima da observação. A distância dos neurônios vizinhos até o BMU está diretamente relacionada com a extensão do ajustamento necessária para eles mesmos (BAÇÃO, LOBO e PAINHO, 2005).

O algoritmo SOM básico consiste em três fases principais: competição, cooperação e atualização (KOHONEN, 2001). Na primeira fase de competição, todos os neurônios competem para encontrar um BMU específico para um padrão de treinamento. Na fase de competição, a vizinhança do BMU é definida baseada no raio de vizinhança. Finalmente na fase de atualização, os pesos dos BMUs são atualizados ao longo seus neurônios vizinhos para se aproximar ao máximo do padrão nos dados (HENRIQUES, 2010). Esse processo é repetido diversas vezes, fazendo com que observações semelhantes se associem com os neurônios relativamente mais próximos. A taxa de aprendizagem e o número de vizinhos decrescem ao longo do treinamento até que restem poucos neurônios, facilitando a convergência dos resultados. Portanto na fase de atualização os valores dos neurônios não são mais aleatórios e refletem as propriedades estatísticas dos dados de entrada originais (VESANTO, 1997).

No SOM é possível utilizar duas formas diferentes de treinamento. Na primeira, chamado treinamento sequencial, é possível encontrar um BMU para cada padrão de treinamento selecionado aleatoriamente apresentado para rede. Assim o BMU é atualizado de acordo com os pesos do padrão de treinamento e respectiva taxa de aprendizagem. Inicialmente a taxa de aprendizagem pode ser alta para permitir ajustes pesados nos neurônios. A mobilidade do neurônio diminui na medida em que a taxa de aprendizagem diminui.

Baseado na taxa de vizinhança, um agrupamento de unidades circundantes é movido para

um padrão de treinamento próximo.

Esse procedimento pode ser descrito com a seguinte função:

Let

$X$  be the set of  $n$  training patterns  $x_1, x_2, \dots, x_n$

$W$  be a  $p \times q$  grid of units  $w_{ij}$  where  $i$  and  $j$  are their coordinates on that grid

$\alpha$  be the learning rate, assuming values in  $]0, 1[$ , initialized to a given initial learning rate

$r$  be the radius of the neighborhood function  $h(w_{ij}, w_{mn}, r)$ , initialized to a given initial radius

1 Repeat

2 For  $k=1$  to  $n$

3 For all  $w_{ij} \in W$ , calculate  $d_{ij} = \|x_k - x_{ij}\|$

4 Select the unit that minimize  $d_{ij}$  as the winner  $W_{winner}$

5 Update each unit  $w_{ij} \in W$ :  $w_{ij} = w_{ij} + \alpha h(W_{winner}, w_{ij}, r) \|x_k - x_{ij}\|$

6 Decrease the value of  $\alpha$  and  $r$

7 Until  $\alpha$  reaches 0

A segunda forma de treinamento chamada de *batch training* (treinamento em lote), comparado com o sequencial, esse treinamento tem como principal diferença o processo de atualização e da não obrigatoriedade de se apresentar aleatoriamente os padrões de treinamento para a rede (HENRIQUES, 2010). Nesse tipo de treinamento algumas vezes a taxa de aprendizagem pode ser omitida (VESANTO, HIMBERG, *et al.*, 2000). O algoritmo desse treinamento atualiza os neurônios apenas depois de uma época, ou seja, depois de todas os padrões de treinamento serem apresentados assim. Em cada época o espaço de entrada é construído usando o diagrama de Voronoi (FIG. 14), também chamado de polígonos de Thiessen (BRASSEL e REIF, 1978).

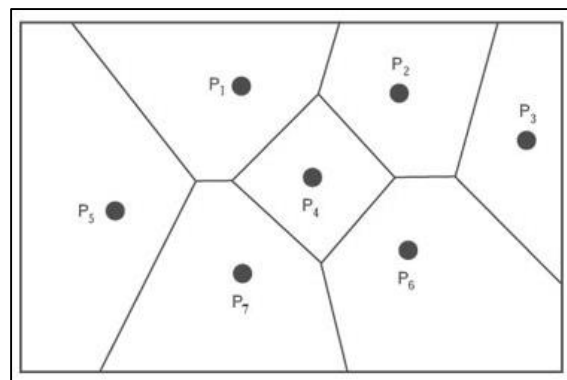


FIGURA 14 – Diagrama de Voronoi. Divisão do espaço onde todos os pontos internos estão próximos ao ponto gerador correspondente. Fonte: da Silva e Bacha (2011)



Depois de definido a taxa de aprendizagem, o terceiro e ultimo principal parâmetro a ser definido é o raio e função de vizinhança.

A função de vizinhança ( $h$ ), pode receber valores entre 0 e 1 e é uma função da posição de duas unidades (vencedora, outra unidade) e um raio ( $r$ ) definido, que habitualmente decresce com o tempo. Na maioria das vezes, a função de vizinhança é uma função radial com o máximo no seu centro, que regularmente diminui até o raio de vizinhança. O raio de vizinhança pode ter valores entre 0 (quando apenas a BMU for atualizada) e o tamanho máximo da rede.

Dois funções de vizinhança podem ser consideradas como as mais utilizadas (FIG. 15). Uma é a função *bubble* que define uma constante em torno de toda a vizinhança da BMU sendo a região restante igual a zero. A outra é a função gaussiana, em que o raio da vizinhança em torno da BMU é maior no início e diminui linearmente durante o treino, ela oferece resultados melhores porém com uma carga computacional bem mais alta (VESANTO, 1997).

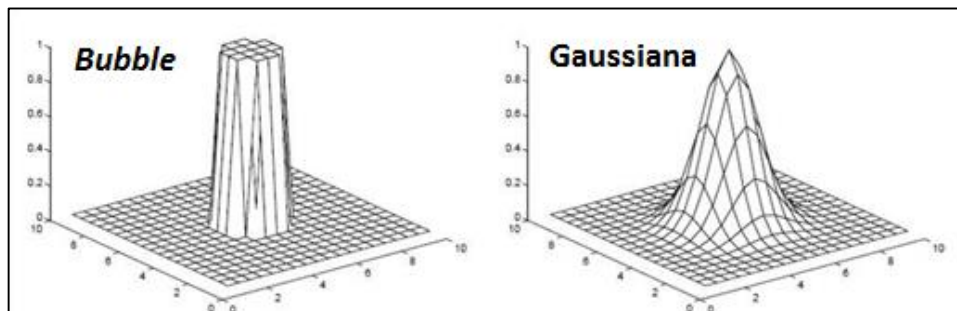


FIGURA 15 – Funções *bubble* e gaussiana. Fonte: Vensato (1997), adaptado de Vensato (1997)

### 4.3 Interpretação dos resultados do SOM

Terminada a parametrização e a aprendizagem do SOM, os próximos passos são a exploração e interpretação dos resultados. Para essa etapa diferentes métodos de visualização podem ser utilizados.

Normalmente para se explorar a estrutura do SOM, utiliza-se a visualização para analisar a existência de padrões ocultos nos dados, sempre tentado se executar a descrição da topologia elaborada, identificação de diferentes segmentos no mapa, correlacionar as variáveis usadas no processo de aprendizagem. Os dados do SOM também podem ser analisados através de outras representações visuais como por exemplo os mapas geográficos utilizados na análise exploratória sobre outra dimensão (SKUPIN e AGARWAL, 2008).

### 4.3.1 Visualização da estrutura do SOM

Existem várias formas de se visualizar o SOM para que uma melhor compressão dos padrões nos dados. Os métodos de visualização do SOM podem ser classificados em três grandes agrupamentos de acordo com o espaço a ser visualizado. Esses métodos podem exibir o espaço de entrada, o espaço de saída ou os dois espaços simultaneamente. Na FIG. 16 pode ser observado um esquema de classificação dos diversos métodos utilizados para visualizar o SOM.

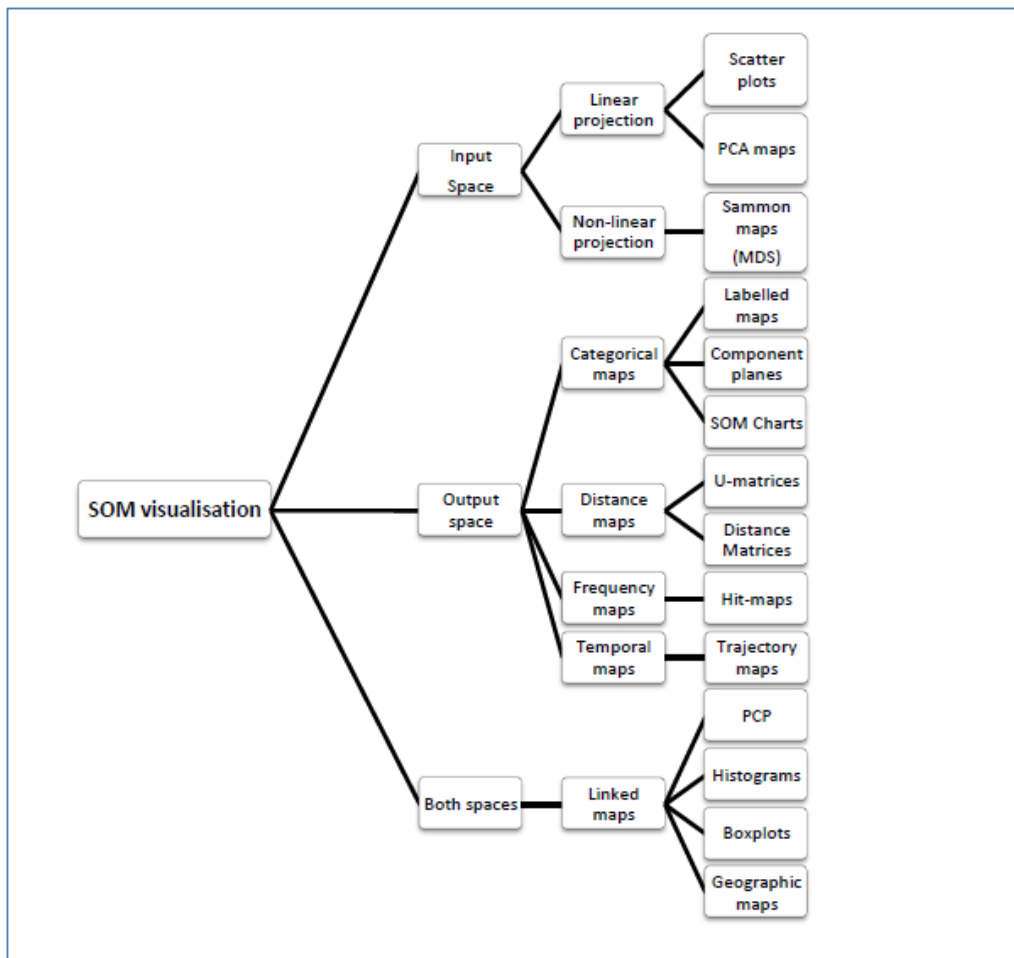


FIGURA 16 – Esquema dos diversos métodos de visualização do SOM. Fonte: Henriques (2010)

Uma dentre várias características do SOM é a preservação das principais relações de topologia dos dados quando eles são representados em um espaço de dimensionalidade menor. Todavia, pode ocorrer uma distorção na proximidade geométrica entres os objetos do espaço de entrada, devido a um efeito de contração do espaço em áreas em que o número de observações é pequeno ou ainda de expansão em áreas com alta densidade de observações (SKUPIN e AGARWAL, 2008).

Essas duas propriedades do SOM somadas, fornecem uma visualização do grau de distorção em um espaço ordenado topologicamente. Esse tipo de visualização pode ser efetuado por um método denominado de *U-Matrix* ou Matriz U, podendo ser classificado como um método de visualização do espaço de saída através de mapas de distâncias (ULTSCH, GUIMARÃES, *et al.*, 1993). Esse método consiste basicamente em se projetar os neurônios em uma matriz regular (FIG. 17), a qual possibilita a representação visual da distância entre os neurônios, fornecendo assim uma interpretação mais simples dos resultados. A identificação de segmentos depende diretamente da interpretação de quem os analisa, podendo ser subjetiva perante as diferentes magnitudes de distorção do espaço entrada (HENRIQUES, BACAO e LOBO, 2012).

Os elementos da rede neural expandidos possibilitam a construção da Matriz-U, que possui o objetivo de quantificar e codificar as distâncias entre os neurônios da rede. Na codificação das distâncias é empregado um recurso de diferenciação através de um esquema de cores, também sendo utilizadas diversas outras técnicas de apresentação, como o gráfico de pizza e o de barras na FIG. 18.

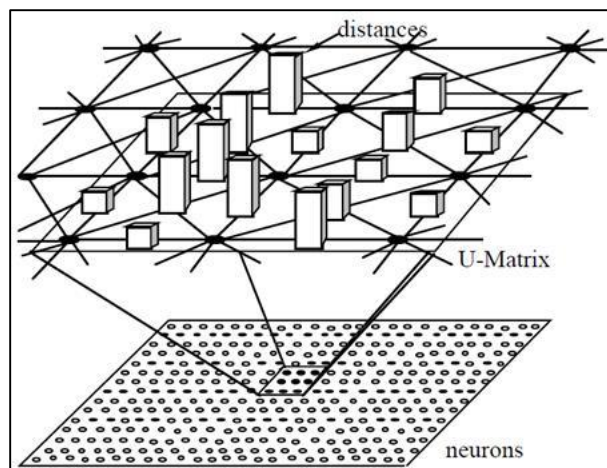


FIGURA 17 – Método da Matriz-U, representação da projeção dos neurônios. Fonte: Ultsch *et al.* (1993)

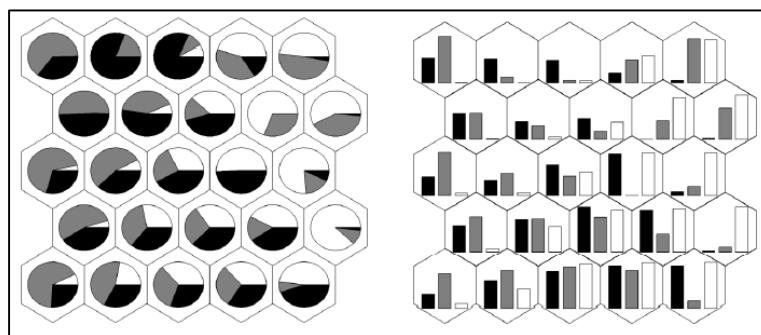
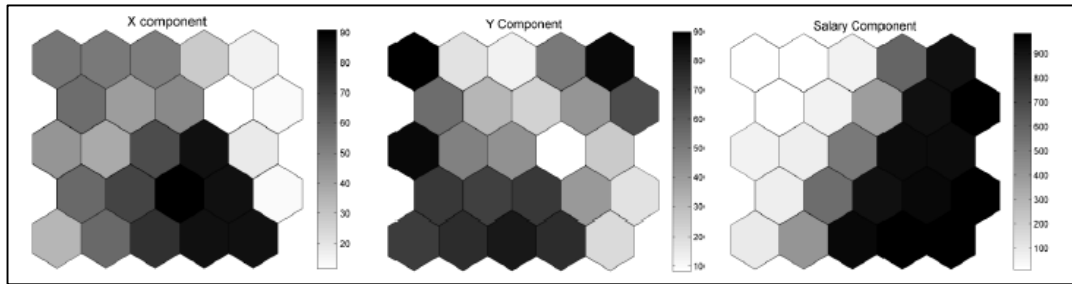


FIGURA 18 – Formas de apresentação da matriz-u, gráfico de pizza à esquerda e gráfico de barras à direita. Fonte: Henriques (2010)

Depois de se obter uma análise geral dos dados, devem ser checadadas as informações detalhadas de cada neurônio da rede. Esse procedimento pode ser realizado utilizando a visualização do plano de componentes. Esse consiste na visualização da distribuição de uma variável em toda a matriz do SOM. Também permite caracterizar diferentes regiões no mapa, tornando possível a definição do comportamento de cada região e uma possível correlação entre variáveis. Correlações que são detectadas através de padrões semelhantes encontrados nas mesmas posições no plano (KOHONEN, 2001).



**FIGURA 19 – Exemplos de um plano de componente relativos a 3 variáveis observadas (X, Y e salário).**

Fonte: Henriques (2010)

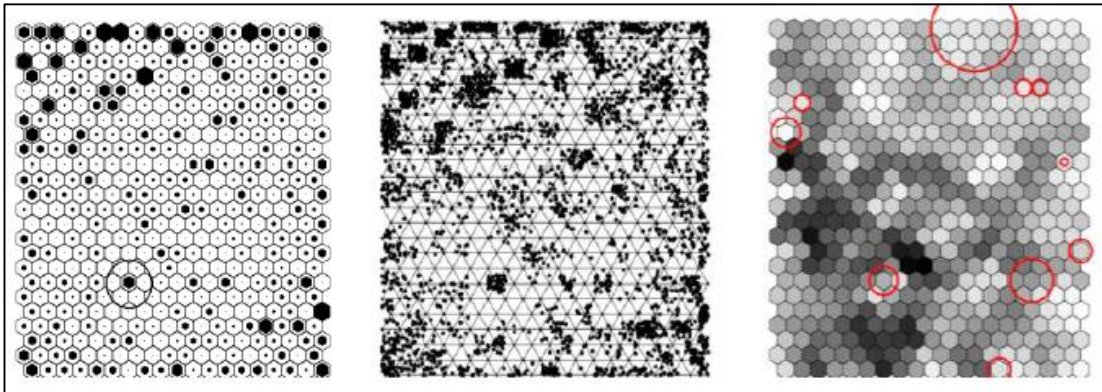
O plano de componentes disponibiliza seus resultados da mesma forma da matriz-u, através de um esquema de cores, sendo mais comumente através de uma escala de cinza (FIG. 19).

A sobreposição do plano de componentes e a matriz-u possibilita conhecer com mais coerência as observações do espaço de entrada. Assim, através do plano de componentes é possível identificar as características que tornam um conjunto de observações distintas, utilizando para isso um segmento de observações que se encontra distante dos demais (VESANTO, 1997).

#### **4.3.2 Visualização dos dados do SOM**

A maneira mais usual de visualização dos dados no SOM é a partir de uma amostra de observações que destaque no mapa o respectivo BMU, ajustando o realce de acordo com quanto maior o número de observações associadas à esse BMU. Essa abordagem mais simplista da informação possui apenas a desvantagem não conseguir representar o nível de similaridade entre as observações e o BMU (erro de quantização) de maneira adequada.

Uma forma de contornar o problema é utilizar a localização de um marcador que indique o erro de quantização através de sua dimensão e o BMU através de sua localização (VESANTO, 1997).



**FIGURA 20 – Exemplo das diferentes formas de se visualizar as observações do SOM.** Fonte: Vensato (1999)

A FIG. 20 mostra três exemplos de se visualizar as observações de um SOM. A primeira imagem da esquerda pode se observar a localização das observações e está representada no preenchimento do neurônio no mapa associado ao BMU, onde o realce é maior na medida em que se aumenta o número de observações representados. A imagem do centro posiciona cada observação é no respectivo BMU com um pequeno deslocamento aleatório para poder destacar a observação das demais. A da direita mostra círculos vermelhos com o centro coincidente ao centro do BMU de um conjunto de observações e o diâmetro desses representa a distância média do BMU às observações representadas por esse. A textura de fundo representa a matriz de distâncias entre neurônios (VESANTO, 1997).

### **4.3.3 Integração geográfica da visualização dos resultados**

Recentemente o SOM vem sendo amplamente empregado em conjunto com outras ferramentas. Essa integração fornece uma exploração maior dos resultados com maneiras diferenciadas na representação visual, facilitando o processo de construção de conhecimento. Quando desejamos utilizar dados de caráter geográfico com o SOM fica explícita a necessidade de uma ferramenta que disponibilize uma melhor interpretação dos padrões e estrutura desses dados, fica assim evidente a necessidade de representação em um mapa geográfico. Através do mapa geográfico é possível uma representação visual dos resultados de forma inovadora e simplificada, também esse é disponibilizado em uma estrutura de dados compatível com os Sistemas de Informação Geográfica (SKUPIN e AGARWAL, 2008).

Devido a esse fato foi desenvolvida a aplicação GeoSOM suíte (amplamente empregada nesse trabalho) pela *Nova Information Management School* (NovalIMS), no intuito de se analisar a natureza espacial dos dados, tal como a visualização dos resultados do SOM em um mapa geográfico (HENRIQUES, BACAO e LOBO, 2012).

#### **4.4 Medidas de qualidade do SOM**

Um aspecto muito importante do SOM é o de avaliar a sua qualidade, podendo assim constituir uma análise mais confiável dos dados. Como a estrutura do SOM é definida ainda no início do processo de aprendizagem com a predefinição de pressupostos sobre os parâmetros do mapa, como a sua topologia, tamanho e a inicialização dos neurônios. Pressupostos esses que se tratados de forma divergente conduzirão a um mapa divergente para com os dados.

Para se avaliar a qualidade são consideradas a exatidão e a preservação topológica do mapa (URIARTE e MARTIN, 2005). A exatidão é medida pelo erro de quantização médio e a preservação do mapa pelo erro topológico.

##### **4.4.1 Erro médio de quantização**

Tendo em consideração que ao final do processo de aprendizagem do SOM cada observação é exatamente apresentada pelo seu respectivo BMU, nesse caso ideal o SOM estaria perfeitamente ajustado aos dados de entrada. Entretanto, a realidade não segue esse preceito de forma tão exata, pois ocorre na realidade um aumento significativo das da distância entre as observações e os BMUs, fazendo com que a exatidão do mapa diminua. Dessa maneira, o erro médio de quantização serve para mensurar o grau de ajustamento do SOM aos dados e corresponde a distância média medida entre cada observação e o BMU correspondente (URIARTE e MARTIN, 2005). Deve-se ainda, considerar que este erro é inerente a todo processo de aprendizagem, uma vez que o BMU é ajustado no sentido de cada observação. Uma boa resolução para o SOM seria aquela em que as observações posicionadas em áreas distantes do espaço de entrada não são mapeadas pelos neurônios vizinhos, que pode ser definido conforme mapa topológico.

#### **4.4.2 Erro topológico**

Através do erro topológico é possível mensurar a integridade topológica da rede utilizando a observação de descontinuidades na projeção das observações utilizadas nos neurônios do SOM. Pode ser descrita como a quantidade de observações que um vetor consegue representar, onde o primeiro e segundo BMUs são adjacentes (ASTUDILLO e OOMMEN, 2014). Pode ocorrer uma descontinuidade topológica do mapa gerando um erro topológico quando o primeiro e segundo BMU de uma dada observação não são adjacentes. Esse erro topológico consegue expressar a proporção de vizinhos locais que estão corretamente mapeados (KIVIULOTO, 1996).

#### **4.4.3 Observações sobre medidas de qualidade**

Analisados em conjunto, o erro topológico e o erro médio de quantização são excelentes medidas de qualidade da aprendizagem de um SOM. Entretanto, podem ser encontradas algumas limitações em sua utilização e novas alternativas estão sendo pesquisadas (ASTUDILLO e OOMMEN, 2014).

Uma forte limitação do erro médio de quantização é sua inutilidade para comparação de mapas com dimensões diferentes. Mapas de maior dimensão possuem tendência a ter um erro menor pois esses utilizam um número maior de neurônios para a representação das observações analisadas, fazendo com que cada um desses esteja mais próximo das observações. Já o erro topológico possui uma limitação quando se faz necessária a comparação entre redes de diferentes topologias. O erro topológico tem forte tendência a aumentar com a dimensão da rede, devido ao fato de ser complexo ordenar um maior número de neurônios (ASTUDILLO e OOMMEN, 2014).

## **5. Metodologia**

Para se alcançar os objetivos propostos pelo estudo foi necessário seguir um conjunto de etapas em ordem cronológica de execução, visando a construção metodológica para o correto desenvolvimento do trabalho. Essas etapas consideram a aquisição de dados, homogeneização, elaboração e estruturação de uma base de dados consolidada para a correta aplicação da mineração de dados e suas respectivas análises, podendo elas serem observadas nos tópicos seguintes.

### **5.1 Recolhimento e tratamento dos dados**

#### **5.1.1 Revisão Bibliográfica**

Foi realizada a revisão bibliográfica conforme objetivos do trabalho, utilizando a Biblioteca do conhecimento online (b-on) e o Google Acadêmico. Procurou-se a dar ênfase a pesquisa dos seguintes termos: dengue, sistema de informação geográfica, variáveis: sociais, econômicas e ambientais, mineração de dados, redes neurais auto organizadas. Pesquisas adicionais ao escopo original do projeto foram necessárias, sobretudo aos temas dos dados ambientais e epidemiológicos. Uma vez que foi necessária à pesquisa de métodos de interpolação dos dados ambientais bem como métodos epidemiológicos utilizados para a definição de epidemias e endemias em um conjunto de dados analisado. Assim sendo a etapa de revisão Bibliográfica foi uma das mais importantes etapas, sendo importante destacar que talvez esta etapa foi revisitada algumas vezes devido a entraves que foram encontrados ao longo do trabalho.

#### **5.1.2 Levantamento dos dados**

O levantamento dos dados compreendeu a busca e recolha de informações referentes às variáveis sociais, econômicas e ambientais nas instituições abaixo, conforme escopo do trabalho.

##### **A) Ministério da Saúde (MS)**

Todas as informações de caráter epidemiológico da dengue foram levantadas através do portal online do departamento de informática do Sistema Único de Saúde do Brasil (DATASUS), mais especificamente no SINAN (Sistema de Informação de Agravos de Notificação). Neste foi possível coletar informações importantes, acerca dos casos da



dengue, extrapolando em muito a expectativa inicial do projeto que era a recolha de alguns dados básicos como o número de casos da dengue. Além deste dado foi possível obter informações mais específicas como: evolução dos casos (óbito ou cura), casos por tipo de carga viral (DEN-1, DEN-2, DEN-3 ou DEN-4), principais complicações (Alterações neurológicas, Disfunção cardiorrespiratória, Insuficiência hepática, Plaquetas <50.000 mm<sup>3</sup>, Hemorragia digestiva, Derrames cavitários, Leucometria < 1000), manifestações hemorrágicas, faixa etária, raça/cor e sexo.

Através do Ministério da Saúde (MS) também foi possível obter informações relativas ao nível de infestação urbana pelo vetor da Dengue (*Aedes Aegypti*), através do indicador Levantamento rápido do Índice de Infestação por *Aedes aegypti* (LIRAA). Entretanto, esta informação necessitou ser requisitada através de um tramite legal e de um sistema de requisição de informações do governo brasileiro baseado na Lei nº 12.527<sup>3</sup>. Por esse motivo foi necessária uma maior atenção na execução dessa atividade, pois foi necessário elaborar, cadastrar e monitorar a solicitação formal junto ao ministério da saúde, através da plataforma online de acesso a informação do MS – Governo do Brasil.

#### **B) Instituto Brasileiro de Geografia e Estatística (IBGE)**

Foram levantadas as informações demográficas contidas nos censos e pesquisas de base populacional para obter informações de cunho social e econômico. Contudo, as informações censitárias no Brasil estão referenciadas em um período intercensitário de 10 anos (ano 2000 ou ano 2010). Somente as informações demográficas básicas como: população total, população masculina, população feminina e população por faixa etária pode ser obtida para o período de análise (2008 até 2010). Foi necessária a aquisição das demais informações em outros órgãos governamentais de acesso mais restrito e moroso, e as quais serão especificadas como fontes fora do escopo do trabalho.

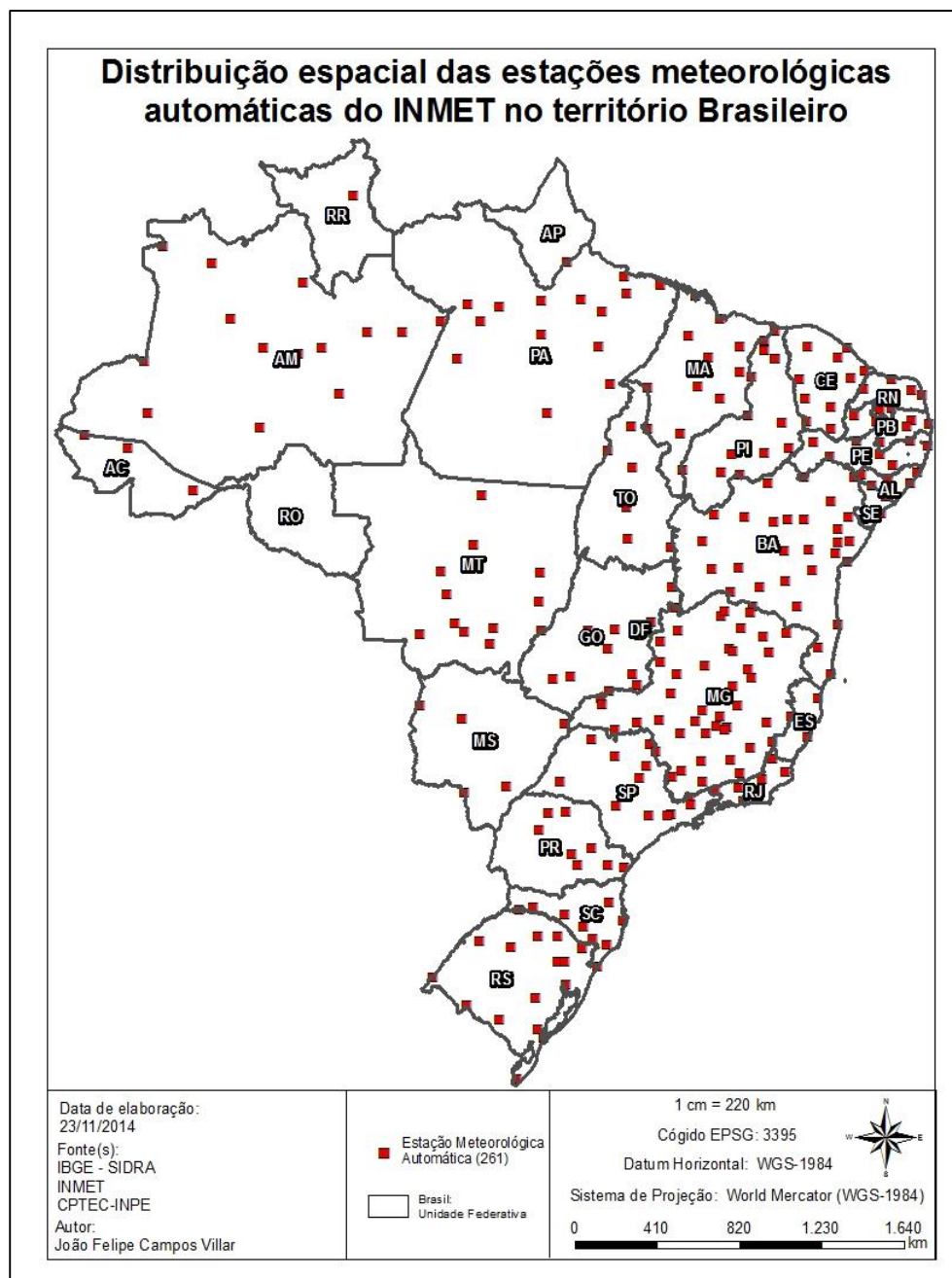
#### **C) Instituto de Meteorologia Brasileiro (INMET)**

Dados de caráter ambiental (temperatura, umidade e precipitação) foram coletados neste órgão.

---

<sup>3</sup> A Lei nº 12.527 regulamenta o direito constitucional de acesso dos cidadãos às informações públicas. Ao regulamentar esse direito, a Lei torna essencial o princípio de que o acesso é a regra, e o sigilo é a exceção, consolida e define o marco regulatório sobre o acesso à informação pública sob a guarda do Estado e estabelece procedimentos para que a Administração Pública responda a pedidos de informação do cidadão.

Neste era possível apenas descarregar os dados relativos a cada estação, ou seja, foi necessário realizar o mesmo procedimento de *download* para cada estação meteorológica automatizada (EMA) pertencente a rede de 265 estações meteorológicas de aquisição de dados automáticas (EMA) existentes no Brasil, sendo que dessas, apenas 261 (FIG. 21) foram utilizadas, devido ao fato de que quatro estações se encontravam fora de funcionamento regular.



**FIGURA 21 – Mapa com a distribuição espacial das estações meteorológicas automatizadas (EMA) no Brasil.**

Algumas fontes de dados fora do escopo inicial do trabalho foram utilizadas, devida à carência de algumas informações nos órgãos previamente analisados. A seguir são descritos os órgãos com as respectivas informações adquiridas.

### **I) Ministério do Trabalho e Emprego (MTE)**

O Ministério do Trabalho e Emprego (MTE) possui informações de todas as empresas e empregados formais no Brasil. Essas informações são de envio obrigatório e são registradas na Relação Anual de Informações Sociais (RAIS). Assim foi possível obter informações bastante confiáveis a respeito da condição socioeconômica dos empregados de carteira assinada no Brasil. Através do sistema WEB foi possível coletar as informações sociais e econômicas (Escolaridade e Remuneração) da população PEA (Pessoas Economicamente Ativas) do Brasil no período de tempo analisado (2008 até 2012).

### **II) IPC Marketing Editora**

Através do sistema online *IPC Maps*, que é atualizado anualmente pela empresa privada IPC Marketing Editora, foi possível obter os dados relativos ao Produto Interno Bruto (PIB) de cada município brasileiro para o período de análise estipulado pelo projeto (2008 até 2012). Convém destacar que os dados são disponibilizados de forma gratuita no sítio web da empresa e estes são baseados em dados divulgados pelas pesquisas econômicas realizadas anualmente pelo IBGE, caracterizando assim o caráter público da informação utilizada. A empresa somente tem o trabalho de organizar e redistribuir as informações através de um meio eletrônico mais eficiente do que os relatórios em formato digital e em papel fornecidos pelo IBGE.

### **III) Programa das Nações Unidas para o Desenvolvimento (PNUD)**

O PNUD disponibiliza em seu *site da web* o Atlas do Desenvolvimento Humano dos Municípios, onde foi possível realizar a consulta dos dados do Índice de Desenvolvimento Humano Municipal (IDH-M), que é uma medida composta de indicadores de três dimensões do desenvolvimento humano: longevidade, educação e renda. Esse índice consegue demonstrar com muita qualidade a condição sócio econômica de uma dada região. Também foram consultados outros indicadores relativos a educação, habitação, saúde, trabalho renda e vulnerabilidade. Todos os dados do atlas são baseados nos censos do IBGE de 1991, 2000, 2010, mas para o presente trabalho foram utilizados somente os dados de 2010.

### 5.1.2 Tratamento dos dados

Uma das etapas mais importantes do trabalho é o tratamento dos dados. Nessa etapa foram padronizados os dados provenientes de algumas instituições. Os dados que obtiveram um tratamento mais detalhado podem ser observados abaixo:

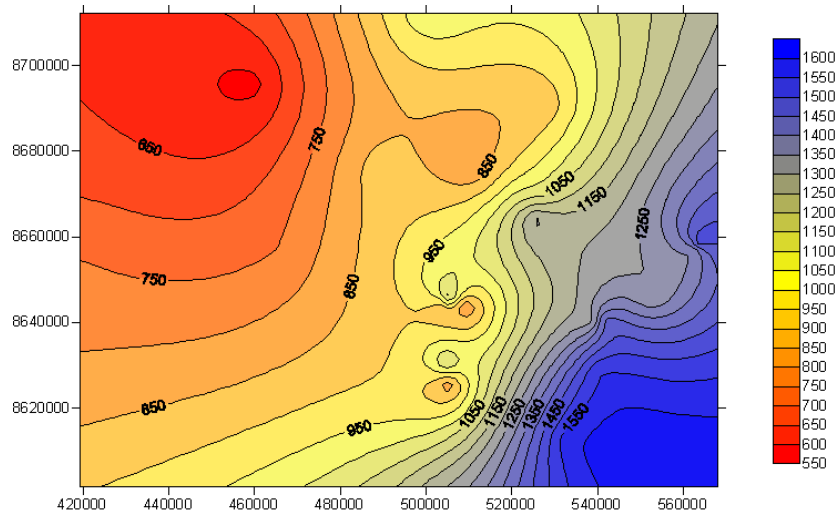
#### a) Dados meteorológicos

Os dados relativos a temperatura média, precipitação acumulada e umidade relativa foram calculados a partir dos dados das estações (EMA). Houve um grande volume de trabalho e esforço nestes dados uma vez que os mesmos se encontravam em formato ASCII texto, todos segregados pelo respectivo código da estação de referência. Estes dados, relativos as 265 estações meteorológicas, foram unificados, padronizados e tratados utilizando o *software* de planilhas eletrônicas Microsoft Excel®.

Mesmo assim os dados coletados e tratados não eram suficientes para representar os mais de 5.000 municípios existentes no Brasil. Para sanar esse problema metodológico foi realizada a interpolação espacial dos dados utilizando a rede de estações existente. Para se interpolar os dados foi necessária uma ampla pesquisa de referências bibliográficas no Brasil que indicassem qual a melhor metodologia a ser empregada para a solução deste tipo de problema.

Entre os métodos empregados rotineiramente para a interpolação de dados meteorológicos, podem ser destacados o inverso ponderado pela distância, krigagem, curvatura mínima (*spline*) e a triangulação com interpolação linear. Cada interpolador tem características próprias que os distinguem uns dos outros. Esses podem ser do tipo global, onde todos os pontos de dados conhecidos são ajustados simultaneamente em toda a superfície a ser interpolada. Podem variar quanto à transição (abrupta ou gradual) ou ainda podem ter caráter determinístico ou probabilístico. Quanto à exatidão podem ser exatos ou inexatos, destacando que os inexatos assumem incertezas (erros) nos dados existentes (DE SOUZA, GOMES, *et al.*, 2011).

Conforme Souza, Gomes, *et al.* (2011) os melhores resultados para a interpolação de dados de precipitação podem ser obtidos pela krigagem, devido ao fato desse ser um interpolador que utiliza a geoestatística como base, o que possibilita a esse modelo ser o mais apto na representação e geração de cartas de isoietas, fazendo com que essas possuam boa fiabilidade em relação aos dados amostrados (FIG. 22).



**FIGURA 22 – Modelo de espacialização das chuvas gerado pela krigagem.** Fonte: de Souza, Gomes, *et al.* (2011)

A interpolação por krigagem teve sua base criada pelo geólogo sul africano Danie Krieg e posteriormente foi formulada pelo matemático francês Georges Matheron. Trata-se de um método de regressão baseado em valores observados Z que circundam os pontos dos dados, possuindo um peso de acordo com os valores de covariância espacial. Algumas das vantagens desse método são: auxilia na compensação de agrupamento de dados, fornece um indicador para a estimativa de erro na interpolação, viabiliza uma base para a execução de simulações estocásticas dos dados (BOHLING, 2006).

A krigagem ainda hoje se constitui em um dos melhores métodos de interpolação aplicado aos dados climáticos (temperatura máxima, temperatura mínima, precipitação e umidade), sobretudo com a krigagem ordinária que responde muito satisfatoriamente para algumas análises meteorológicas (BARBOSA, 2006).

Tendo como base essas referências bibliográficas, foi utilizado o método de Krigagem ordinária para interpolar todos os dados meteorológicos provenientes do INMET. Para isso foi necessária a utilização da extensão *Geo Spatial Analyst* do software ArcGIS© 10.2 da ESRI®, onde foram realizadas as interpolações de todas as variáveis e todos os anos analisados (2008 até 2012), constituindo mais de 10 superfícies de krigagem conforme o ANEXO A.

Após a interpolação foi realizado o processo de análise espacial correlacionando a camada de dados contendo os pontos com a sede de cada município do Brasil com a superfície de krigagem, gerando assim uma tabela no formato dbf, contendo os valores respectivos para cada variável meteorológica utilizada (temperatura máxima, temperatura mínima, amplitude térmica média, umidade relativa acumulada, precipitação acumulada) agregados por município e respectivo ano de análise.

#### **b) Dados epidemiológicos**

Todos os dados epidemiológicos se encontravam em formato CSV e agregados pelo respectivo mês e ano epidemiológico. Sendo assim os dados tiveram apenas que ser agrupados, tratados e calculados utilizando o *software* Microsoft Excel®.

#### **c) Dados ambientais**

Os dados relativos a taxa de infestação urbana pelo mosquito da dengue se encontravam muito desorganizados e com muitas inconsistências mesmo sendo disponibilizados em formato de planilha eletrônica xls (Microsoft Excel®). Foi necessário um trabalho meticuloso para tratamento dos dados. Os nomes dos municípios muitas vezes estavam duplicados e divergentes do formato de nome e código estipulados pelo IBGE, portanto foi necessário realizar a padronização dos valores de cada campo da base.

#### **d) Demais dados**

Todos os demais dados tiveram praticamente o mesmo tipo de tratamento, importação e estruturação no *software* Microsoft Excel, compatibilização dos nomes dos municípios para o formato do IBGE (Maiúsculo e sem acentos), cruzamento com a tabela de códigos dos municípios e padronização dos campos numéricos para o padrão utilizado no Brasil.

## 5.2 Estruturação e elaboração da base de dados

Nesta etapa os dados que foram coletados, tratados e armazenados em formato de planilha eletrônica foram organizados e importados em uma base de dados. Esta foi referenciada espacialmente através dos limites políticos administrativos municipais do Brasil. Como padrão de dados espaciais foi utilizado o sistema de coordenadas geográficas sem nenhuma projeção e o *datum* horizontal WGS-84, para que fosse possível compartilhar o projeto com instituições brasileiras e portuguesas.

O formato da base de dados escolhido foi o *Personal Geospatial Data Base* (PGDB) da ESRI®, devido ao fato de que a base de dados foi analisada posteriormente em softwares que exigem o formato de banco de dados compatível com o *mdb* (Microsoft Access) e com o arquivo cartográfico digital *shapefile* criado pela ESRI®, sendo agora um formato compatível com o padrão *Open Geospatial Consortium* (OGC).

Os títulos das tabelas dentro da base de dados, foram estruturados e organizados baseados na tipologia das variáveis, e essas foram agrupadas pelos seguintes eixos temáticos: economia, demografia, cartografia, meio-ambiente e epidemiologia. Portanto a elaboração da base de dados foi fundamental para a etapa de análise exploratória dos dados, onde foi possível identificar as variáveis mais relevantes para serem utilizadas no SOM.

A base de dados ainda possibilitou a elaboração de uma consulta *Structured Query Language* (SQL), visando a exportação dos dados em formato adequado (*shapefile*) e com as variáveis selecionadas para serem utilizados posteriormente no *software* GeoSOM, para a construção da rede neural SOM.

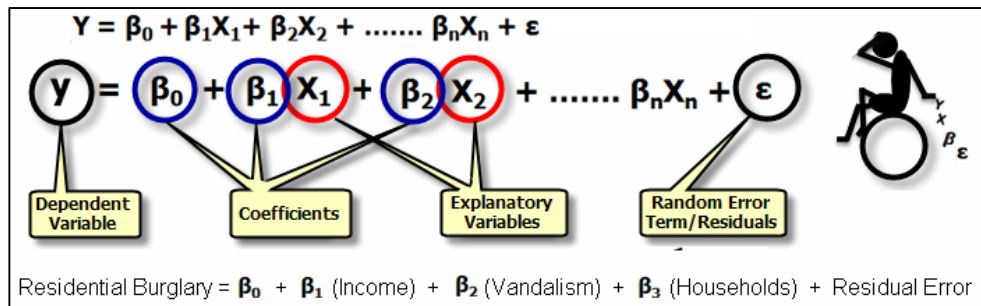
## 5.3 Análise exploratória dos dados

Foi realizada nesta etapa uma análise estatística descritiva no intuito de se verificar o comportamento de alguns dados, sobretudo os dados meteorológicos que possuem estreita relação com os casos da dengue, conforme discutido no capítulo 2. Essa análise exploratória pode ser dividida em duas fases.

Na primeira utilizando o Microsoft Excel foi realizada uma análise com a estatística descritiva para cada uma das variáveis selecionadas tendo como referência os fatores determinantes da dengue, também apresentados no capítulo 2.

A segunda fase consistiu na utilização da ferramenta *Exploratory Regression* do *Software ArcGIS 10.2*, a metodologia de regressão exploratória se mostra uma ferramenta robusta de grande utilidade quando se pretende avaliar o grau de relacionamento que variáveis explicativas (independente) possuem com uma dada variável dependente (alvo) (ROSENSHEIN, SCOTT e PRATT, 2011). Portanto a equação de regressão possui como elementos principais (FIG. 23):

- **Variável dependente (Y):** aquilo que se deseja modelar ou prever;
- **Variáveis explicativas (X):** variáveis que podem influenciar ou auxiliar na explicação da variável dependente;
- **Coefficiente (β):** valores, computados pela ferramenta de regressão, refletem o relacionamento e a força de cada variável explanatória para com a variável dependente;
- **Resíduos (ε):** a parte da variável dependente que não pode ser explicada pelo modelo de regressão, são as sobre e sob predições do modelo.



**FIGURA 23 – Esquema explicativo da equação utilizada no modelo de regressão.** Fonte: Rosenhein (2011)

Através de seis checagens disponibilizadas nessa ferramenta é possível avaliar melhor o modelo de regressão, o que possibilita identificar as variáveis mais importantes para um determinado modelo.

No caso desse trabalho a técnica de mineração de dados através da regressão exploratória foi apenas utilizada como suporte no processo exploratório dos dados, auxiliando a definição das variáveis a serem utilizadas na rede neural SOM. Uma vez que a escolha das variáveis a serem utilizadas foram identificadas na vasta literatura brasileira referente a correlação de variáveis socioeconômicas e ambientais com a epidemia da dengue.

Também foi realizada uma pesquisa na literatura brasileira para determinar o método que melhor estabelecesse o cálculo e a classificação da taxa de incidência da dengue na população. Foi de extrema importância o cálculo dessa variável, no trabalho uma vez que os valores absolutos dos casos da dengue refletem de forma significativa os valores da população absoluta de cada município (TORRES, 2008), gerando assim uma análise extremamente enviesada dos dados. Portanto para definir a taxa de incidência foi utilizado o parâmetro disponibilizado pelo Departamento de Vigilância Epidemiológica do Ministério da



Saúde do Brasil, que estabelece a seguinte equação matemática para o cálculo da taxa (MINISTÉRIO DA SAÚDE - BRASIL, 2010):

$$\alpha = \frac{dy}{dx} \times 100.000$$

**EQUAÇÃO 1 – Taxa de incidência da dengue.**

Onde:

**$\alpha$** : Taxa de incidência da dengue

**$dy$** : Número de casos da dengue

**$dx$** : Valor absoluto da população

Com a utilização dessa fórmula foi possível identificar em cada município do Brasil o número de casos da dengue em cada 100.000 habitantes. Ainda é preconizada nos parâmetros a classificação da taxa de incidência da dengue em três intensidades: alta, média e baixa (TAB. 2), sendo ainda criada uma nova variável contendo a respectiva classificação (MINISTÉRIO DA SAÚDE - BRASIL, 2010).

**TABELA 2**

Classificação da taxa de incidência da dengue.

Classificação	Parâmetro
Baixa	Até 100 casos/100 mil hab.
Média	De 101 a 300 casos/100 mil hab
Alta	Acima de 300 casos/100 mil hab

Fonte: Ministério da Saúde – Brasil (2011, p3).

## 5.4 Análise espacial dos dados

Essa etapa consistiu na detecção do comportamento e dos padrões espaço temporais dos dados analisados. Segundo a primeira lei da geografia proposta por Waldo Tobler (1970), todos os dados estão inter-relacionados, mas dados que estão mais próximos espacialmente um dos outros estão mais relacionados entre si do que dados mais distantes (LONGLEY, GOODCHIELD, *et al.*, 2003).

Portando, levando em consideração que os dados geográficos vinculam lugar, tempo e atributo (LONGLEY, GOODCHIELD, *et al.*, 2003), também sabendo do caráter altamente

geográfico da dengue, conforme o capítulo 2 deste trabalho, se fez necessária a utilização de algumas técnicas e métodos da Ciência da Informação Geográfica, especialmente na área da estatística espacial para que fosse possível identificar o padrão de distribuição espacial inerente da taxa de incidência da dengue, e correlacionar esse com o padrão de distribuição espacial encontrado nos dados resultantes da mineração de dados através da rede neural SOM.

A análise espacial simples dos dados não revela muitos padrões espaciais relevantes, sendo na maioria das vezes indicada uma análise da dependência espacial<sup>4</sup> dos dados. Dessa forma, optou-se nesse trabalho por realizar análise do índice global de Moran e do índice local de Moran (Moran-LISA)<sup>5</sup>, para a taxa de infestação da dengue e dos resultados encontrados pelo SOM, *clusters*, em cada ano do período analisado (2008 a 2012)

Utilizando técnicas como o índice local de Moran (FIG.24), pode ser analisada com bastante eficiência a dependência espacial de cada variável em estudo, fornecendo assim maiores informações acerca do comportamento espacial dos dados e sua possível inter-relação geográfica.

No caso do índice global de Moran os valores variam de -1 a +1. Os valores iguais a zero indicam a inexistência de autocorrelação espacial entre os dados, já os próximos a zero estabelecem um autocorrelação espacial não significativa. Os valores positivos indicam uma autocorrelação positiva, ou seja, o valor da variável do objeto tende a ser semelhante ao valor da mesma variável de objetos vizinhos. Já os valores negativos indicam autocorrelação espacial negativa, o valor de uma variável em uma dada região não é dependente de valores da mesma variável em áreas diferentes (MARQUES, HOLZSHUH, *et al.*, 2010).

Apesar do índice global de Moran ser um bom indicador para o comportamento espacial dos dados, ele apresenta apenas um valor único para medir a associação espacial em toda a área de estudo. Assim sendo também é necessário utilizar um método que seja capaz de expressar com mais detalhes o regime de associação espacial em um número mais elevado de áreas. Dessa forma o índice local de Moran se mostra um método muito eficiente na identificação e cálculo da associação local de objetos tendo como base a mesma variável de análise (MARQUES, HOLZSHUH, *et al.*, 2010).

---

<sup>4</sup> A dependência espacial pode ser entendida como a tendência a que o valor de uma variável associada a uma determinada localização assemelhe-se mais ao valor de suas amostras vizinhas do que ao restante das localizações do conjunto amostral. Fonte: Longley, Goodchild, *et al.* (2003)

<sup>5</sup> No índice local de Moran a autocorrelação espacial é calculada a partir do produto dos desvios em relação a média como uma unidade de covariância, permitindo a decomposição do indicador global de Moran em um indicador local de associação espacial (LISA). Fonte: Longley, Goodchild, *et al.* (2003)

No caso específico desse trabalho foram utilizadas as ferramentas *Global Moran's I* (para o cálculo do índice global de Moran) e *Anselin Local Moran's I* (para o computar o índice local de Moran) através da extensão *Spatial Statistics*® no Software ArcGIS 10.2©.

A ferramenta *Anselin Local Moran's I* possibilita a detecção de *clusters* em feições com magnitude similar (ANSELIN, 1995). Ela também identifica *outliers*<sup>6</sup> espaciais, utilizando para isso o cálculo do valor para o índice local de Moran, o score Z e o valor-p, bem como um código representando o tipo de *cluster* para cada feição. O score Z e o valor-p representam a significância estatística para o valor índice computado pela ferramenta (MITCHELL, 2005).

The Local Moran's  $I$  statistic of spatial association is given as:

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^n w_{i,j} (x_j - \bar{X}) \quad (1)$$

where  $x_i$  is an attribute for feature  $i$ ,  $\bar{X}$  is the mean of the corresponding attribute,  $w_{i,j}$  is the spatial weight between feature  $i$  and  $j$ , and:

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^n w_{ij}}{n - 1} - \bar{X}^2 \quad (2)$$

with  $n$  equating to the total number of features.

The  $z_{I_i}$ -score for the statistics are computed as:

$$z_{I_i} = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}} \quad (3)$$

where:

$$E[I_i] = \frac{\sum_{j=1, j \neq i}^n w_{ij}}{n - 1} \quad (4)$$

$$V[I_i] = E[I_i^2] - E[I_i]^2 \quad (5)$$

FIGURA 24 – Equações formuladas para explanação do índice local de Moran. Fonte: Mitchel (2005)

<sup>6</sup> Na estatística um *outlier* corresponde a um valor atípico, ou seja é uma observação em um dado conjunto de observações que apresenta um grande afastamento das demais, tanto em termos superiores quanto em inferiores. (GLADWELL, 2008)

## 5.5 Análise multidimensional e mineração dos dados

Essa etapa consistiu na análise e correlação das variáveis estudadas com os aspectos epidemiológicos da dengue. Utilizando a metodologia de mineração de dados baseada em redes neurais, especificamente o mapa auto organizável de Kohonen ou SOM (*Self Organizing Map*), foi possível realizar o agrupamento ou *clustering* dos dados. Para tal foi utilizado o *software* GeoSOM suíte, visando estabelecer o agrupamento não só pela análise das variáveis, mas também em relação a localização dos dados no espaço geográfico. Nessa etapa é importante destacar que foi necessário realizar a normalização dos dados a serem introduzidos na ferramenta, sendo esta uma tarefa de grande relevância para o trabalho.

### 5.5.1 Variáveis selecionadas para análise no SOM

Foram selecionadas 17 variáveis de acordo com as referências bibliográficas acerca dos estudos epidemiológicos ecológicos aplicados a dengue e com as análises exploratórias explicadas no tópico 5.3. Essas variáveis foram agrupadas de acordo com os três temas a seguir:

#### A) Ambientais

##### 1) LIRAA:

Indicador do nível de infestação pelo mosquito vetor da dengue/ *Aedes aegypti*) valores elevados indicam alto índice de infestação.

##### 2) Precipitação acumulada no ano:

Quantidade de chuva em mm<sup>3</sup> (milímetros cúbicos) acumulada durante todo o ano.

##### 3) Temperatura máxima do ano:

Maior temperatura máxima registrada em todo o ano em graus Celsius (°C).

##### 4) Temperatura mínima do ano:

Menor temperatura mínima registrada em todo o ano em graus Celsius (°C).

##### 5) Umidade relativa média no ano:

Razão entre o conteúdo real de umidade de uma amostra de ar e a quantidade de umidade que o mesmo volume de ar pode conservar na mesma temperatura e pressão quando saturado (AYOANDE, 2010). Esse é expresso na forma de

percentual. Foi utilizado nesse caso a média dessa medida dos todos os dados do ano.

**6) Amplitude térmica no ano:**

Diferença entre a temperatura máxima do ano e a temperatura mínima do ano expressa em graus Celsius (°C).

**B) Socioeconômicas**

**1) Índice de GINI:**

Mede o grau de desigualdade em uma dada população, ele aponta a diferença entre os rendimentos dos mais pobres e dos mais ricos. Numericamente, varia de 0 a 1, o zero representa a situação de igualdade de renda.

**2) IDHM:**

Índice de desenvolvimento humano do município, varia de 0 até 1, quanto mais próximo do zero piores são as condições de desenvolvimento urbano, mais próximo do um 1 melhores são as condições.

**3) PIB per capita:**

Produto interno bruto em milhões de reais (R\$) dividido pela população absoluta.

**4) Remuneração média:**

Remuneração média em salários mínimos (no Brasil atualmente próximo de U\$ 350,00) de toda a população PEA (População Economicamente Ativa).

**5) THEIL:**

Medida estatística da distribuição de renda, possui valor entre 0 e 1 e quanto maior este valor, pior a distribuição de renda.

**6) Taxa de analfabetismo:**

Percentual de pessoas analfabetas em relação ao total de pessoas de um mesmo, no caso desse trabalho foi utilizado o percentual sobre o valor absoluto da população municipal

## C) Geográficas

### 1) Área:

Área em quilômetros quadrados (km<sup>2</sup>) do município, calculada utilizando a projeção transversa de mercator.

### 2) Densidade demográfica:

Valor absoluto da população dividido pela área.

### 3) Indicador de mobilidade urbana:

Medida estatística para expressar o grau de acesso dos cidadãos aos meios de transporte em dada localidade. Varia de 0 a 1, quanto mais próximo do 0 piores as condições de infraestrutura urbana que favoreçam o acesso e deslocamento dos cidadãos dentro da área de agregação estuda, no caso desse estudo o município.

### 4) Latitude:

Latitude em graus decimais, baseada na sede municipal, utilizando como referência o *datum* WGS-84.

### 5) Longitude:

Longitude em graus decimais, baseada na sede municipal, utilizando como referência o *datum* WGS-84.

## 5.5.2 Ajustamento dos dados

Após a escolha das variáveis é necessário fazer o ajustamento dos dados, etapa primordial para poder se utilizar os dados nos processos de segmentação e modelação. No ajustamento dos dados são realizados alguns procedimentos tais como: identificação e tratamento de *outliers* (valores extremos) ou omissos, redução da dimensionalidade, normalização dos dados, eliminação de redundâncias, etc. Também foram identificados municípios que não possuíam dados em mais de 60% das dimensões (variáveis) analisadas, e esses excluídos da análise.

Como grande parte das variáveis utilizadas nesse trabalho possuem escalas completamente diferentes, foi extremamente necessária a normalização dos dados. Foram utilizados dois métodos de normalização para posterior efeito comparativo de dois tipos de treinamento na montagem de modelos distintos de SOM para cada ano do período analisado.

O primeiro método utilizado foi o *range* ou Min-Max, disponível no aplicativo GeoSOM Suite,

que basicamente consiste em transformar a amplitude dos valores de todas as dimensões em uma amplitude entre os valores 0 e 1.

O método possui como vantagem principal a preservação das relações entre valores distintos, pois ele mantém a distribuição original dos dados, não introduzindo assim qualquer enviesamento (DOS SANTOS, 2013). Esse método pode ser descrito pela seguinte equação:

$$y' = \left( \frac{y - \min1}{\max1 - \min1} \right) \times (\max2 - \min2) + \min2$$

#### **EQUAÇÃO 2 – Normalização pelo Range.**

Onde:

**y**: valor original

**y'**: novo valor normalizado

**min1**: valor mínimo da variável original

**max2**: valor máximo da variável original

Como segundo método para efeito comparativo entre os treinos, foi utilizada a normalização *var* ou *z-score*. Essa normalização transforma os valores originais da variável de entrada, de maneira que sua média seja 0 e desvio padrão (variância) igual a 1. Esse pode ser descrito através da equação a seguir:

$$y' = \frac{y - \text{média}}{DP}$$

#### **EQUAÇÃO 3 – Normalização pelo Z-score.**

Onde:

**y'**: novo valor normalizado

**y**: valor original

**DP**: desvio padrão da variável original

**média**: média da variável original

Esse método de normalização funciona bem quando os valores mínimos e máximos da variável de entrada não são conhecidos, ou quando existem *outliers* que podem inviabilizar a utilização do método *range* (HENRIQUES, 2010).

Terminada a etapa de ajustamento dos dados, foi iniciada a elaboração da rede SOM, para tal foi necessária a definição da estrutura e dos parâmetros para o treinamento como o tamanho, topologia, forma de inicialização, taxa de aprendizagem, grau de vizinhança etc. A seguir é descrita a construção dos parâmetros mais relevantes para o processo de montagem da rede, que são o tamanho da rede, topologia, taxa de aprendizagem e grau de vizinhança.

### **5.5.3 Definição do tamanho e da topologia**

É possível definir o tamanho do SOM através de duas formas, o *K-means* e o SOM emergente.

É importante definir o número correto de segmentos a serem utilizados pela rede, uma vez que um número baixo pode truncar a informação presente nos dados. Por essa razão foi escolhida a elaboração de uma rede SOM de elevada dimensão, para que fosse construído um mapa topológico com resolução adequada para a identificação de regiões de particularidades específicas, facilitando assim o processo de delimitação de clusters, tornando-os mais homogêneos.

Por esse motivo foi definida uma dimensão de 20 x 20 para ambos os tipos de rede SOM elaboradas para cada ano do período analisado. Dessa forma todas as redes neurais elaboradas possuem 400 neurônios.

Como topologia foi definido o formato hexagonal, devido a esse formato ser o mais largamente utilizado na literatura de redes neurais SOM (DOS SANTOS, 2013), e por oferecer também uma maior dimensão de vizinhança para cada neurônio.

### **5.5.4 Critérios de treinamento**

É fundamental realizar o processo de variação de parâmetros de aprendizagem com diferentes valores em uma rede neural SOM. Isso se torna necessário para poder se alcançar melhores resultados através de um modelo de rede mais bem ajustado aos dados analisados. Assim é possível determinar qual processo de aprendizagem gera resultados validados.

Portanto foi realizado um processo de aprendizagem executado em duas etapas de treinamento onde os padrões da rede foram apresentados aleatoriamente, fazendo com que a taxa de aprendizagem tivesse várias tentativas na construção da rede SOM.

Como pode ser observado na TAB. 3, o primeiro treino utilizou um grau de vizinhança de valor 30 e um número de interações igual a 200, taxa de aprendizagem de 0,6 e o tipo de normalização range. No segundo treino foi definido um grau de vizinhança de valor 15, número de interações igual a 300, taxa de aprendizagem igual a 0,3, tipo de normalização var e função de vizinhança gaussiana.



**TABELA 3**

Parâmetros utilizados na definição dos dois treinos do SOM realizados para cada ano do período analisado.

Parâmetros	2008		2009		2010		2011		2012	
	Treino 1	Treino 2	Treino 1	Treino 2	Treino 1	Treino 2	Treino 1	Treino 2	Treino 1	Treino 2
Taxa de aprendizagem	0,6	0,3	0,6	0,3	0,6	0,3	0,6	0,3	0,6	0,3
Número de interações	200	300	200	300	200	300	200	300	200	300
Grau de vizinhança	30	15	30	15	30	15	30	15	30	15
Tipo de normalização	Range	Var	Range	Var	Range	Var	Range	Var	Range	Var

## 5.6 Avaliação do modelo

Nessa fase, foi realizada a escolha das redes SOM que melhor representaram os dados analisados. Através da comparação entre o erro topológico e o erro médio de quantização, da interpretação da rede e na matriz-u que melhor apresentaram resultados para os dados de entrada.

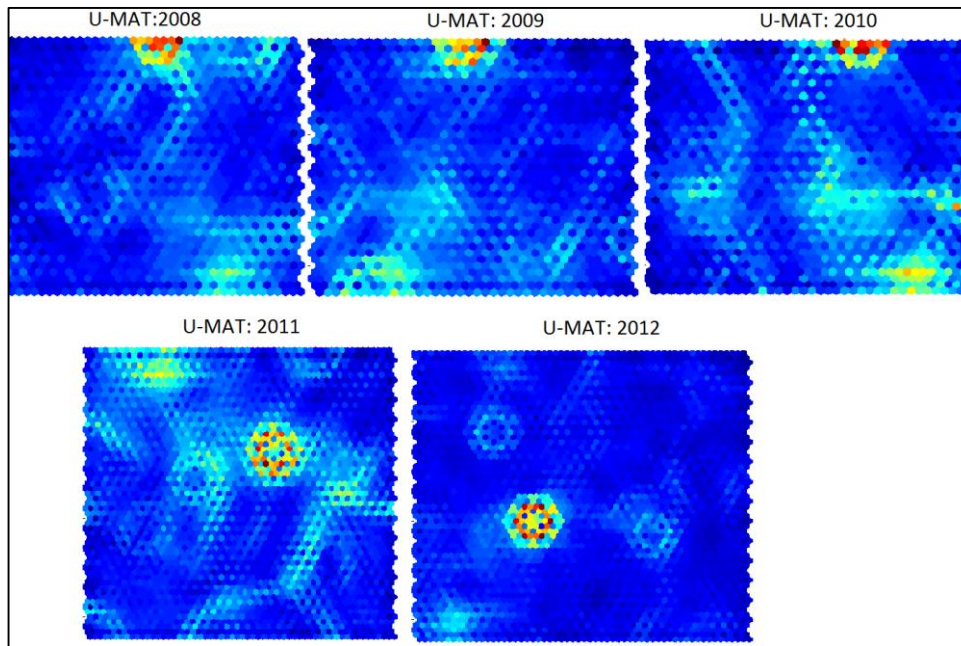
Conforme observado na TAB. 4, a rede que apresentou melhores resultados para todos os anos foi a do tipo1. Vale destacar que o erro topológico (et) se manteve quase que constante nas duas redes e em todos os anos, ficando em torno de 0,03. O erro médio de quantização sempre foi melhor na rede 1 em todos os anos, variando de 0,14 até 0,19.

**TABELA 4**

Erros encontrados para a rede 1 e rede 2, geradas por processos de aprendizagem diferentes em todos os anos do período analisado.

Tipo de Erro	2008		2009		2010		2011		2012	
	Rede 1	Rede 2	Rede 1	Rede 2	Rede 1	Rede 2	Rede 1	Rede 2	Rede 1	Rede 2
EMQ	0,19	1,66	0,15	1,59	0,16	1,57	0,16	1,39	0,14	1,44
ET	0,03	0,04	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,02

Devido a essa análise de tipo de erros, foi escolhido nesse estudo o padrão dos parâmetros de aprendizagem da rede 1 para realizar análise e definição de clusters na matriz-u em cada ano do período de 2008 até 2012.



**FIGURA 25 – Matriz-U de cada ano do período analisado (2008 até 2012).**

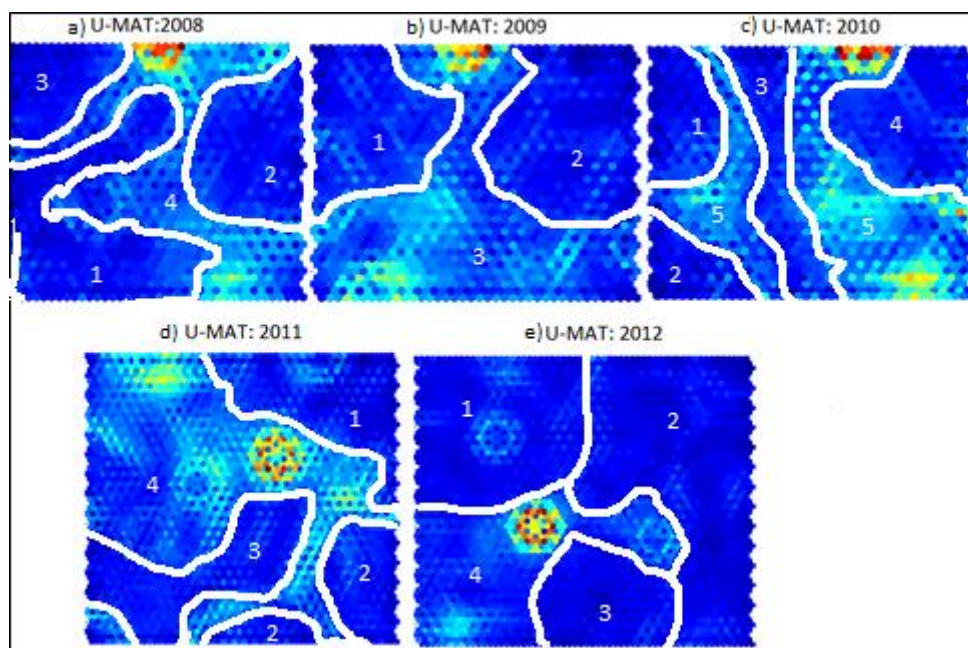
Através da FIG. 25 é possível observar a matriz-u para cada ano, pelas quais foi possível definir o cluster relativo a cada ano. A cor azul representa baixa distância entre os neurônios vizinhos, e a cor vermelha representa uma maior distância. Uma descontinuidade abrupta de cores é sempre utilizada como delimitador do cluster

## 6. Resultados

Nesse capítulo são apresentados os resultados obtidos com esse estudo em ordem cronológica de execução, conforme a metodologia estabelecida no capítulo 5.

O primeiro resultado obtido pode ser considerado a delimitação de clusters em cada matriz-U, onde foi possível identificar diferentes agrupamentos para cada ano do período analisado.

No ano de 2008 foi possível identificar a presença de 4 clusters, devido a segmentação de 4 áreas na matriz-u, do respectivo ano (FIG. 26, a).



**FIGURA 26 – Segmentação da Matriz-U de cada ano do período analisado (2008 até 2012).**

Já no ano de 2009 apenas 3 clusters foram definidos, conforme as 3 áreas que puderam ser segmentadas na matriz-u (FIG. 26, b).

No ano de 2010 foi possível delimitar mais clusters, 5 no total, pois a matriz-u desse ano possui limites mais bem definidos para a segmentação (FIG. 26, c).

Em 2011 pode ser facilmente delimitadas 4 áreas na segmentação, devido aos limites bem mais definidos nesse ano, gerando um resultado de 4 clusters (FIG.26, d).

Finalmente em 2012 foram novamente delimitadas 4 áreas de segmentação que forneceram 4 clusters, mas nesse ano devido os limites não estarem bem definidos, um grau de dificuldade maior teve de ser superado para a delimitação dos clusters (FIG. 26, e).

## 6.1 Resultados da análise exploratória

A análise exploratória dos dados revelou alguns aspectos e comportamentos pertinentes ao estudo de incidência da dengue. Primeiramente, foi possível observar que as variáveis ligadas ao meio ambiente estão fortemente correlacionadas com os casos da dengue em todo o período de análise e em todas as regiões do Brasil (FIG. 27), abaixo serão citados alguns dos exemplos mais pertinentes que demonstram essa forte correlação.

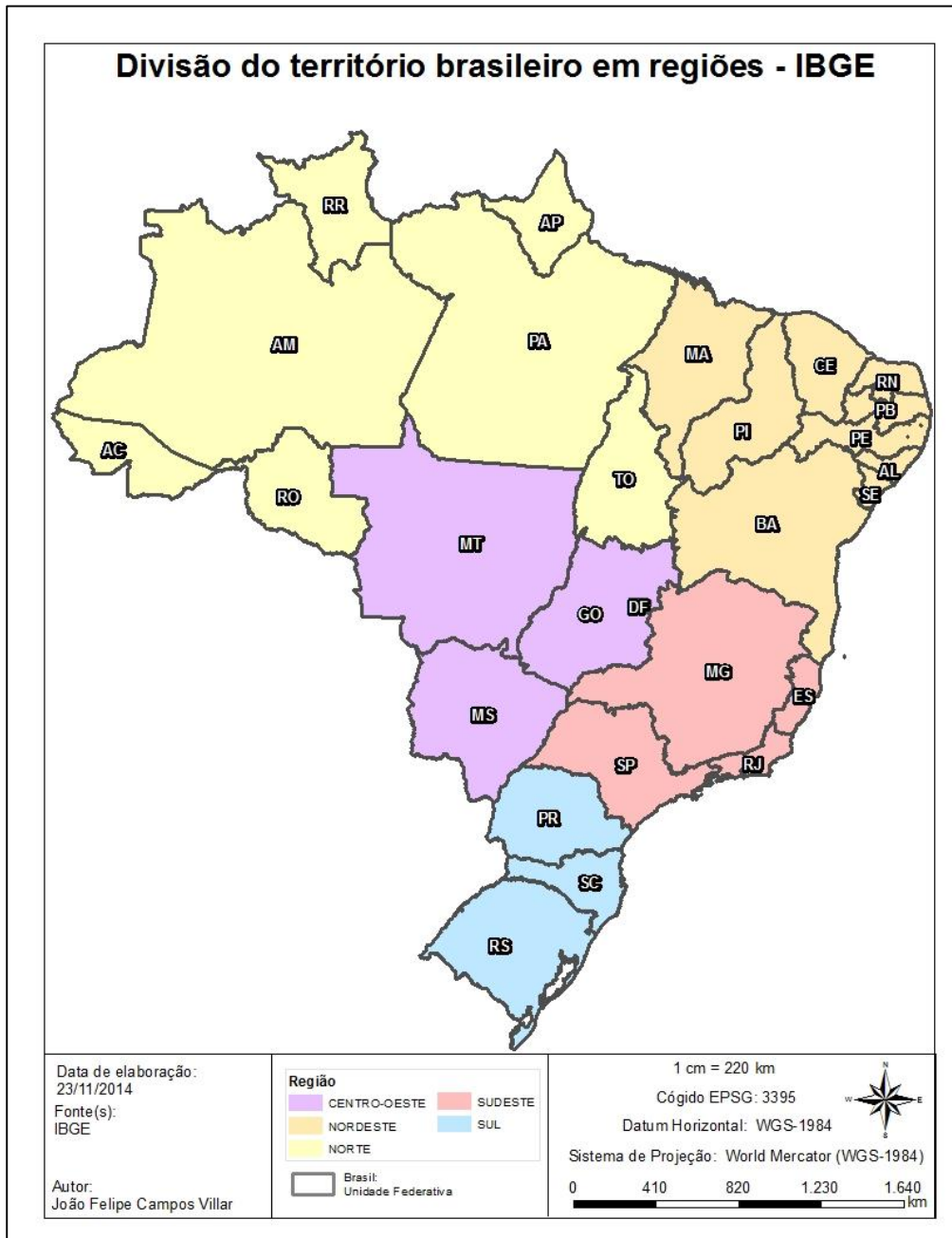


FIGURA 27 – Divisão do território brasileiro em regiões e unidades federativa (estados). Fonte: IBGE (2007)

Pode ser observado no GRAF. 1 que na região Sudeste entre os meses de junho, julho e agosto de 2008 ocorreu uma redução acentuada nos casos da dengue e simultaneamente ocorreu uma forte queda nos valores da precipitação acumulada, seguida da umidade relativa e da amplitude térmica (diferença entre a máxima e a mínima temperatura registradas).

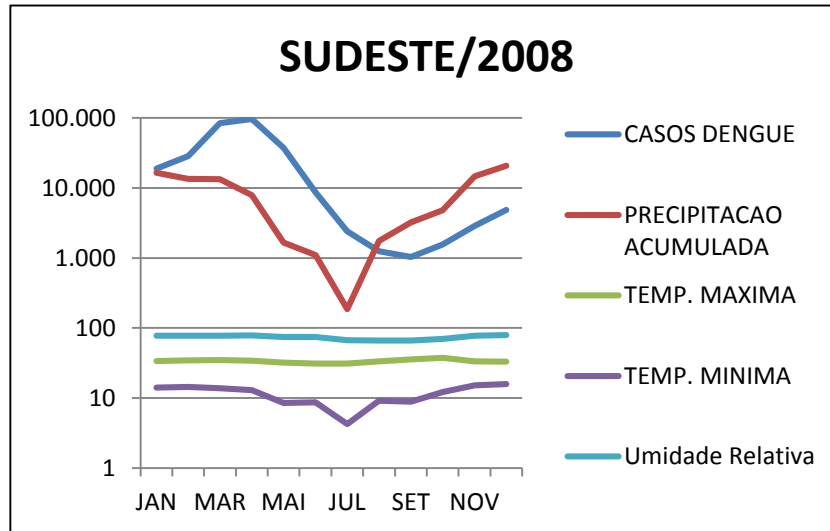


GRÁFICO 1 – Correlação de variáveis ambientais com os casos da dengue no Sudeste do Brasil em 2008.

Entretanto pode ser observado um comportamento oposto quanto a precipitação na região Sul no ano de 2008, em que ocorre uma diminuição nos casos (entre os meses de junho, julho e agosto) e um aumento expressivo na precipitação, mas o comportamento da amplitude térmica é o mesmo, quando a mesma aumenta o número de casos sofrem uma queda (GRAF. 2).

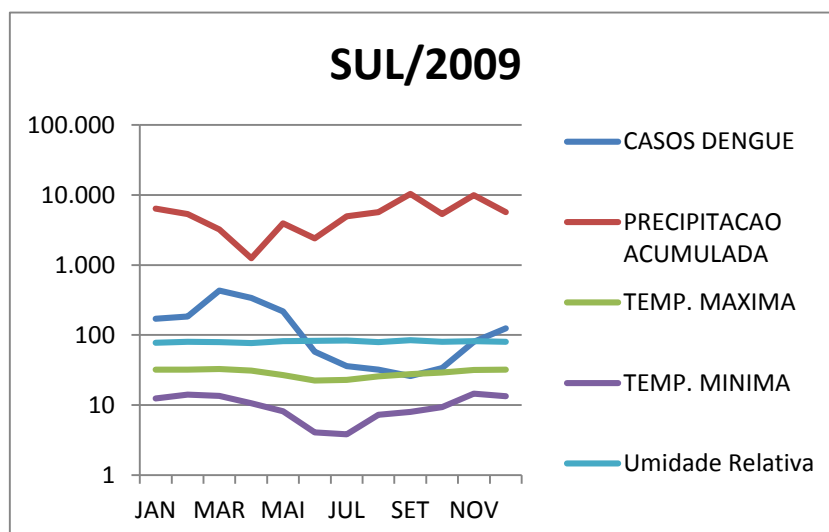
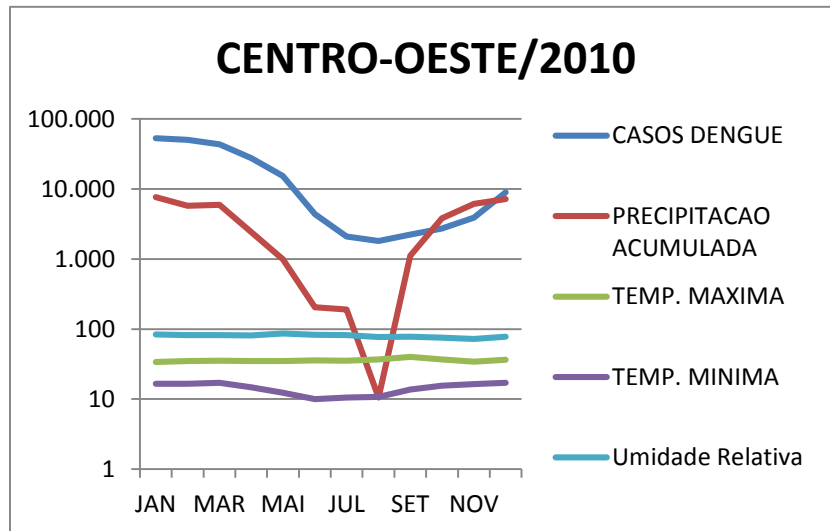


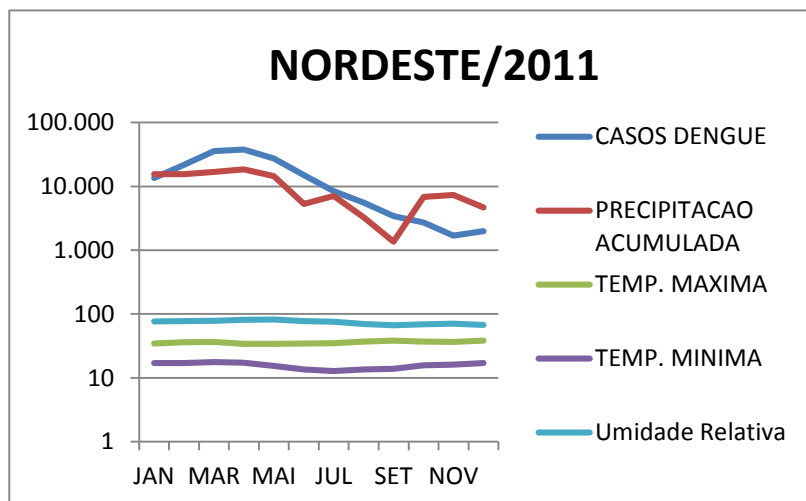
GRÁFICO 2 – Correlação de variáveis ambientais com os casos da dengue no Sul do Brasil em 2009.

Já na região Centro-Oeste em 2010 foi possível observar que a precipitação acumulada cai drasticamente entre os meses de julho agosto e setembro, contudo os casos da dengue não sofrem uma alteração na mesma proporção, sendo esses influenciados por um período anterior, o que pode evidenciar que os casos da dengue possuem um limite mínimo de precipitação para se tornarem estáveis (GRAF. 3)

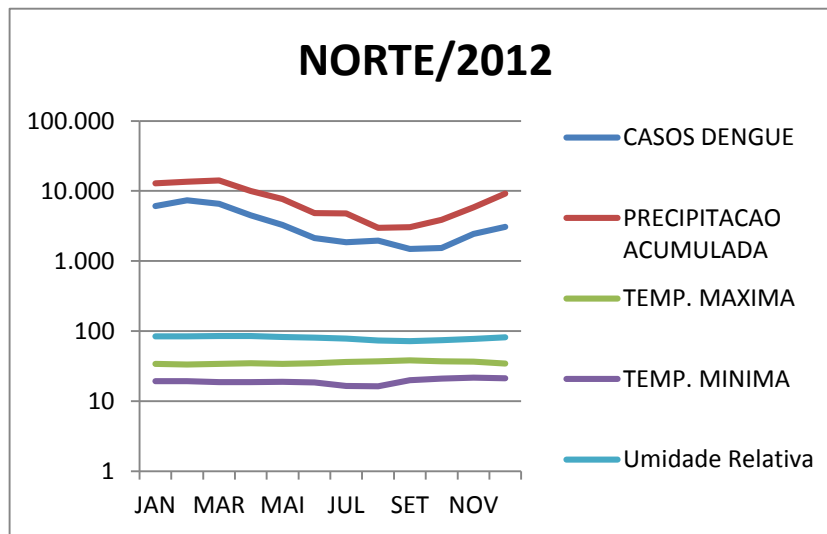


**GRÁFICO 3 – Correlação de variáveis ambientais com os casos da dengue no Centro-oeste do Brasil em 2010.**

Em 2011 a região Nordeste apresenta comportamento que corrobora os dados apresentados pela região sul e centro-oeste anteriormente (GRAF. 4). A precipitação acumulada sofre uma queda e os casos da dengue sofrem uma queda até um certo limite e se mantém estáveis, posteriormente a precipitação acumulada se eleva até um determinado valor mas os casos da dengue continuam se mantendo quase que constantes, o que evidencia mais uma vez um limite máximo e mínimo de flutuação dos casos da dengue conforme a precipitação, ou seja, aparentemente a relação não é tão linear como algumas referências bibliográficas citadas no capítulo 2 desse trabalho sugerem.



**GRÁFICO 4 – Correlação de variáveis ambientais com os casos da dengue no Centro-oeste do Brasil em 2011.**



**GRÁFICO 5 – Correlação de variáveis ambientais com os casos da dengue no Norte do Brasil em 2012.**

É importante destacar ainda que a região norte em 2012 apresentou um comportamento extremamente proporcional entre os casos da dengue a precipitação, evidenciando um certo comportamento espacial para a relação dessas variáveis, sendo que as demais variáveis são praticamente constantes devido ao clima tropical da floresta Amazônica em que está inserida a maior parte dessa região.

As demais variáveis, pertencentes aos grupos de variáveis socioeconômicas e geográficas, possuem um relacionamento com os casos da dengue bastante subjetivo. Por essa razão o comportamento dessas variáveis pode ser melhor compreendido através do método de regressão exploratória dos dados.

Na regressão exploratória foi possível identificar que oito variáveis possuem significância estatística de mais de 50 por cento em todos os anos para a variável dependente utilizada, que no caso foi a taxa de incidência da dengue. Dessas, cinco variáveis possuem significância positiva maior que 50 por cento, são elas: indicador de mobilidade urbana, LIRAA, temperatura mínima, amplitude térmica, temperatura máxima. Ou seja, valores elevados nessas variáveis contribuem para um aumento significativo dos valores da variável dependente. Outras 3 variáveis (precipitação acumulada, IDHM, umidade relativa) possuem significância negativa maior que 50 por cento. Assim essas variáveis contribuem para uma diminuição significativa nos valores da variáveis dependentes quando seus valores aumentam muito. No ANEXO B é possível verificar o sumário dos resultados da regressão exploratória feita para todo o período de análise. O aspecto espacial do dado também foi explorado e será melhor descrito no tópico a seguir.

## 6.2 Resultados da análise espacial

A análise espacial dos dados evidenciou uma forte autocorrelação espacial nos dados relativos a taxa de incidência da dengue, corroborando dessa forma com a literatura brasileira apresentada no capítulo 2, que define uma forte relação entre a doença e as características geográficas.

Através do relatório gerado utilizando a análise do índice global de moran, foi possível notar que todos os anos apresentaram padrão clusterizado, com o índice sempre positivo para todos os anos analisados.

O ano de 2008 apresentou o valor de 0,119 para índice global de Moran (IGM) e z-score de 103, o que revela uma significância do índice para a existência de cluster (GRAF. 6).

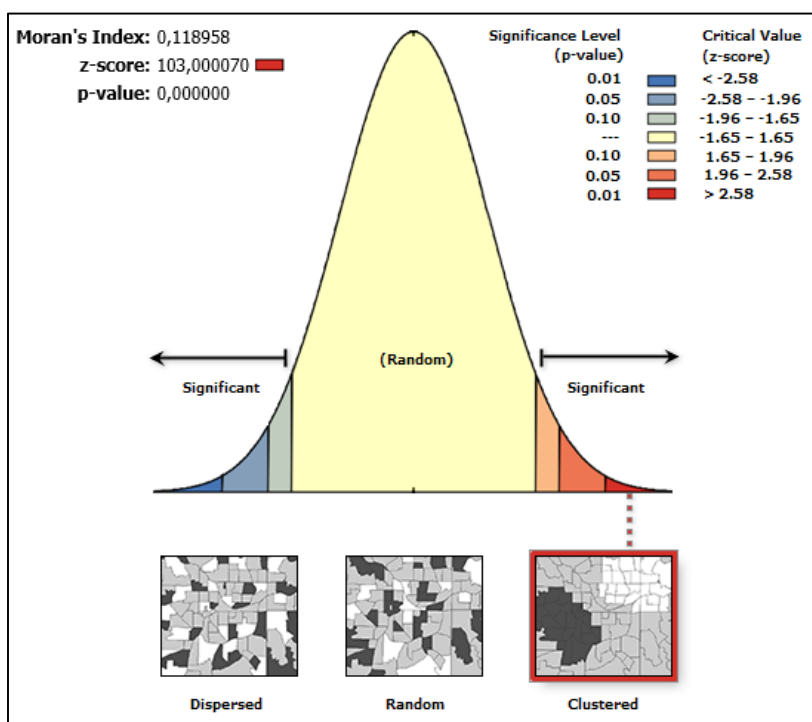


GRÁFICO 6 – Índice global de Moran para todos os municípios do Brasil utilizando a taxa de incidência da dengue em 2008.

No ano de 2009 é possível verificar um valor de 0,105 no IGM e z-score de 91, evidenciando uma autocorrelação espacial positiva para a área do Brasil (GRAF. 7).



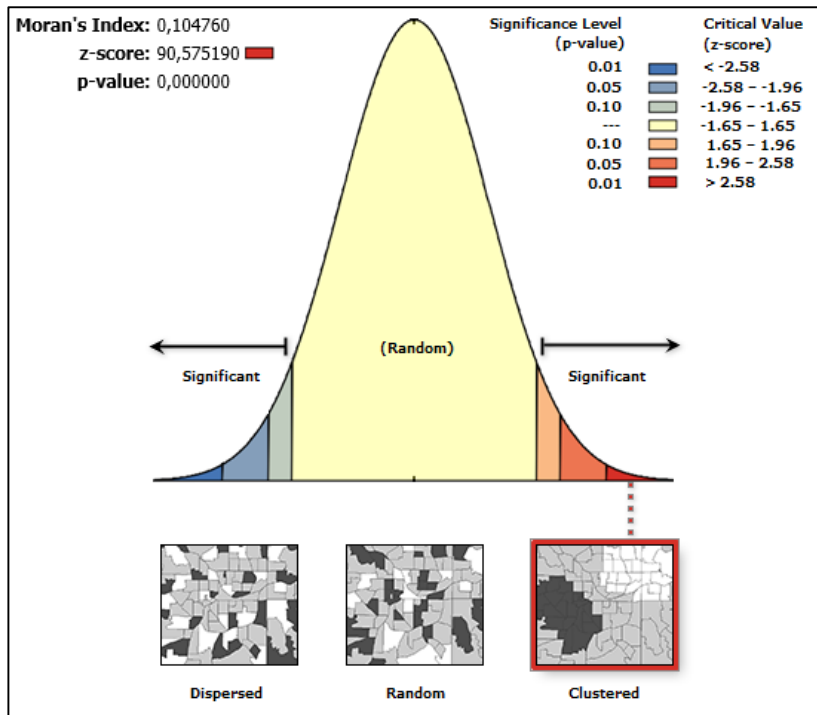


GRÁFICO 7 – Índice global de Moran para todos os municípios do Brasil utilizando a taxa de incidência da dengue em 2009.

Para o ano de 2010 observa-se um valor de 0,114 no IGM e z-score de 98, demonstrando mais uma vez o padrão clusterizado dos dados (GRAF. 8).

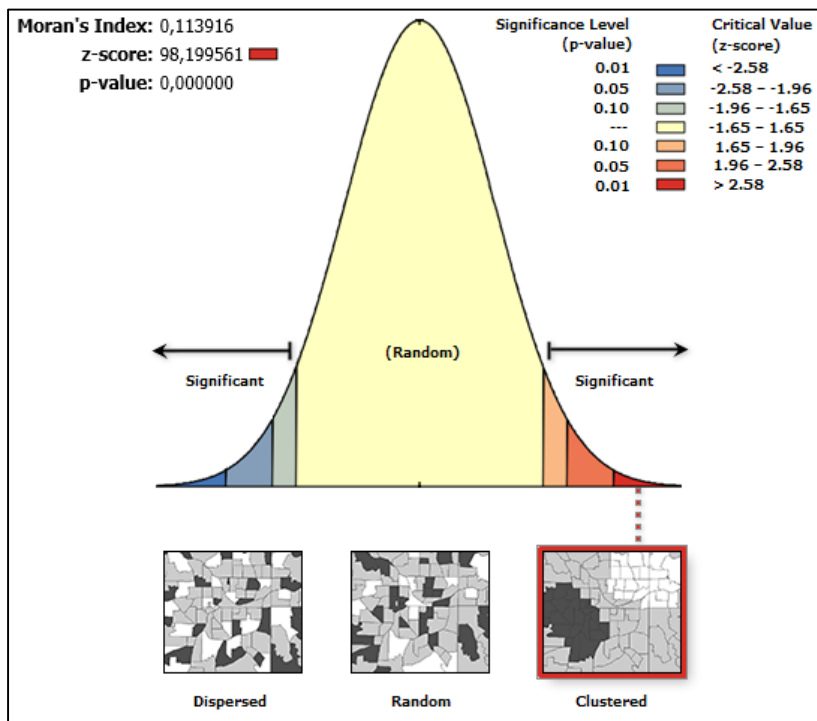


GRÁFICO 8 – Índice global de Moran para todos os municípios do Brasil utilizando a taxa de incidência da dengue em 2010.

Em 2011 foi encontrado um valor de 0,075 no IGM e no z-score 64, demonstrando que nesse ano os dados tem uma autocorrelação positiva e significância menor, portanto menos clusterizados do que os dois anos anteriores (GRAF. 9).

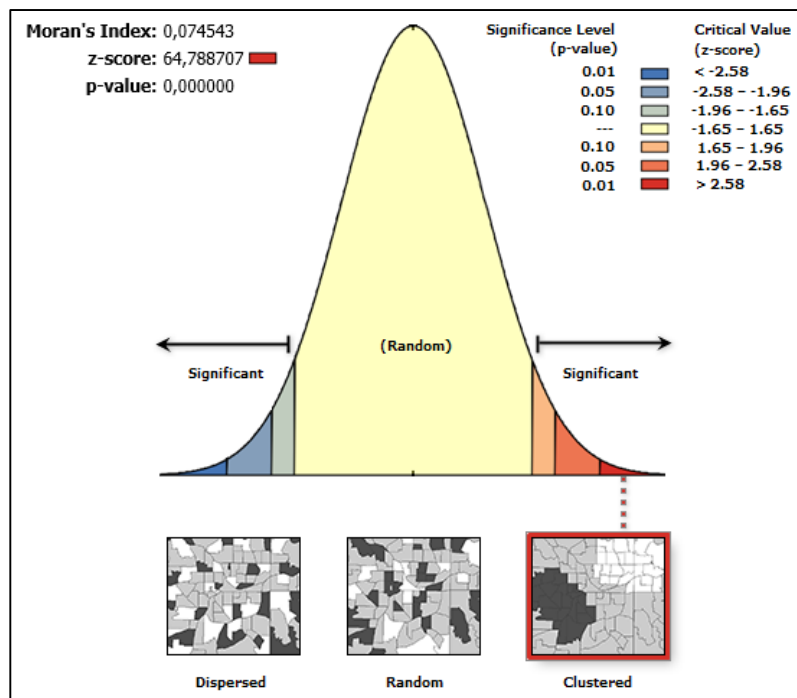


GRÁFICO 9 – Índice global de Moran para todos os municípios do Brasil utilizando a taxa de incidência da dengue em 2011.

Já 2012 apresentou o valor de 0,085 para o IGM e z-score 73, valores muito próximos aos encontrados no ano inicial do período de análise, evidenciando um possível ciclo no comportamento espacial geral dos dados (GRAF. 10).

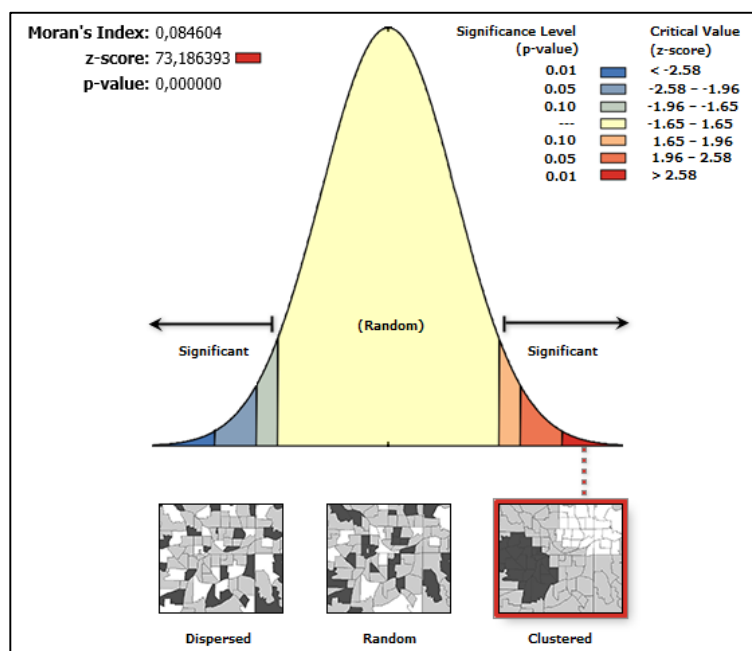
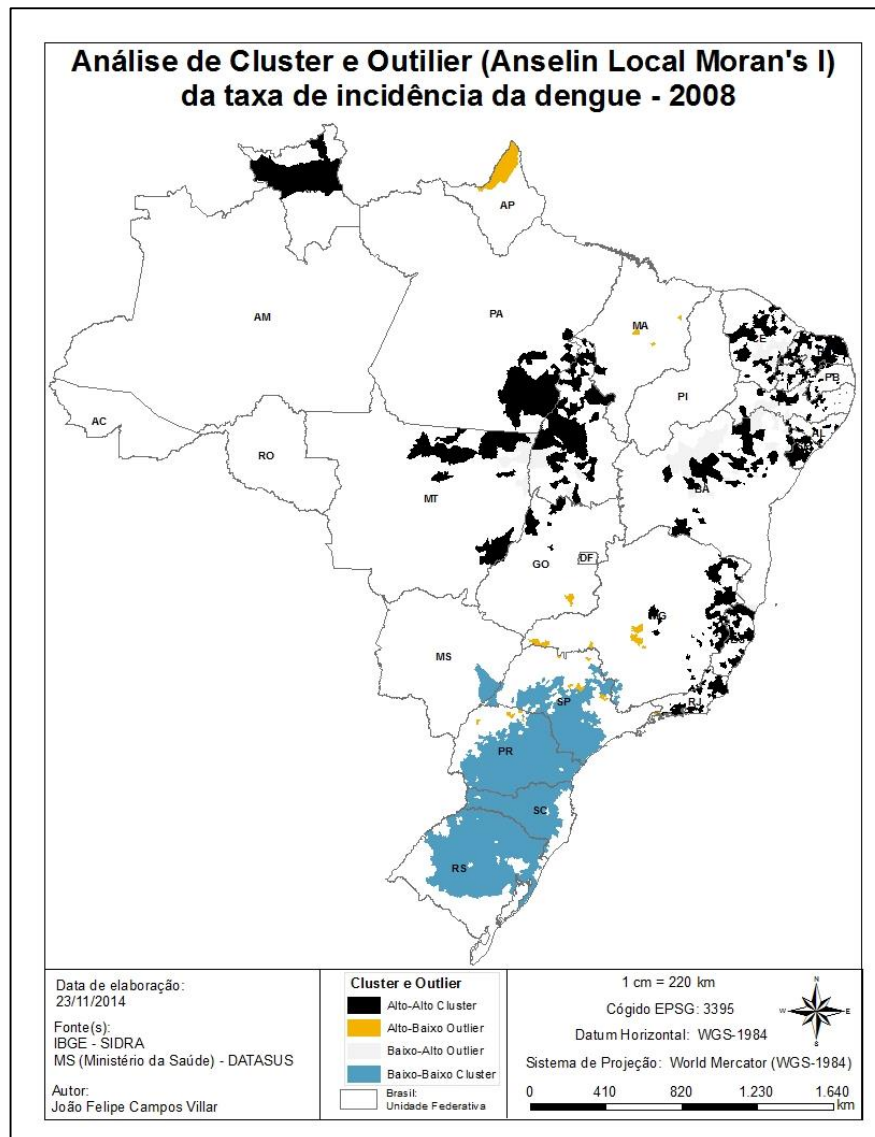


GRÁFICO 10 – Índice global de Moran para todos os municípios do Brasil utilizando a taxa de incidência da dengue em 2011.

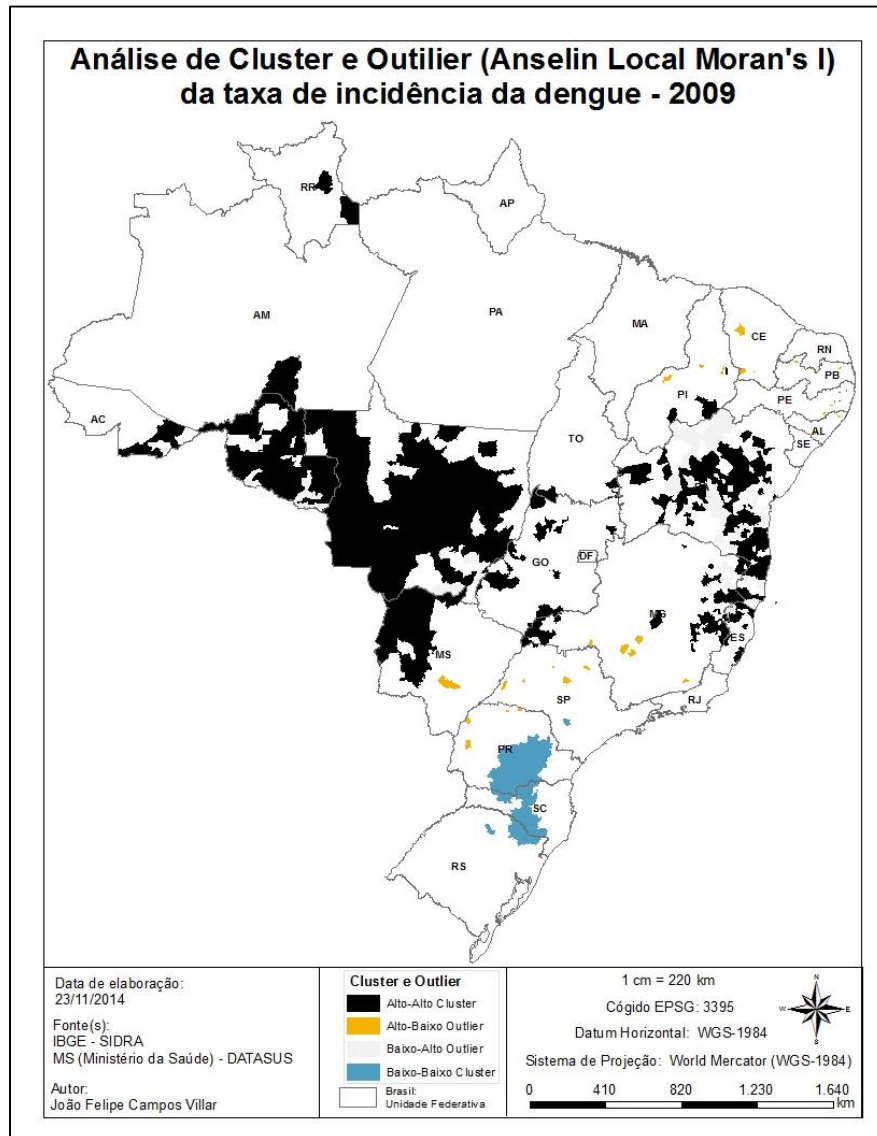
Atráves do indicador local de Moran (LISA) foram identificados padrões distintos para algumas regiões em anos específicos. Dessa forma esse indicador forneceu um detalhe ainda maior na autocorrelação espacial dos dados relativos a taxa de incidência da dengue no Brasil.

No ano de 2008 pode ser observado que na região Sul do Brasil, maior parte do estado de São Paulo (SP) e uma pequena região do Mato Grosso Sul (MS) existe a ocorrência de um cluster com baixos valores de taxa de incidência. Já os clusters de valores elevados da taxa podem ser observados praticamente em todo o estado do Rio de Janeiro (RJ), porção extrema oriental de Minas Gerais (MG), na região Nordeste do Brasil, no estado de Roraima (RR) e na porção central da região Norte e Centro-Oeste do Brasil. Ocorrem alguns *outliers* de extremo superior e inferior de forma dispersa no território do país (FIG. 28).

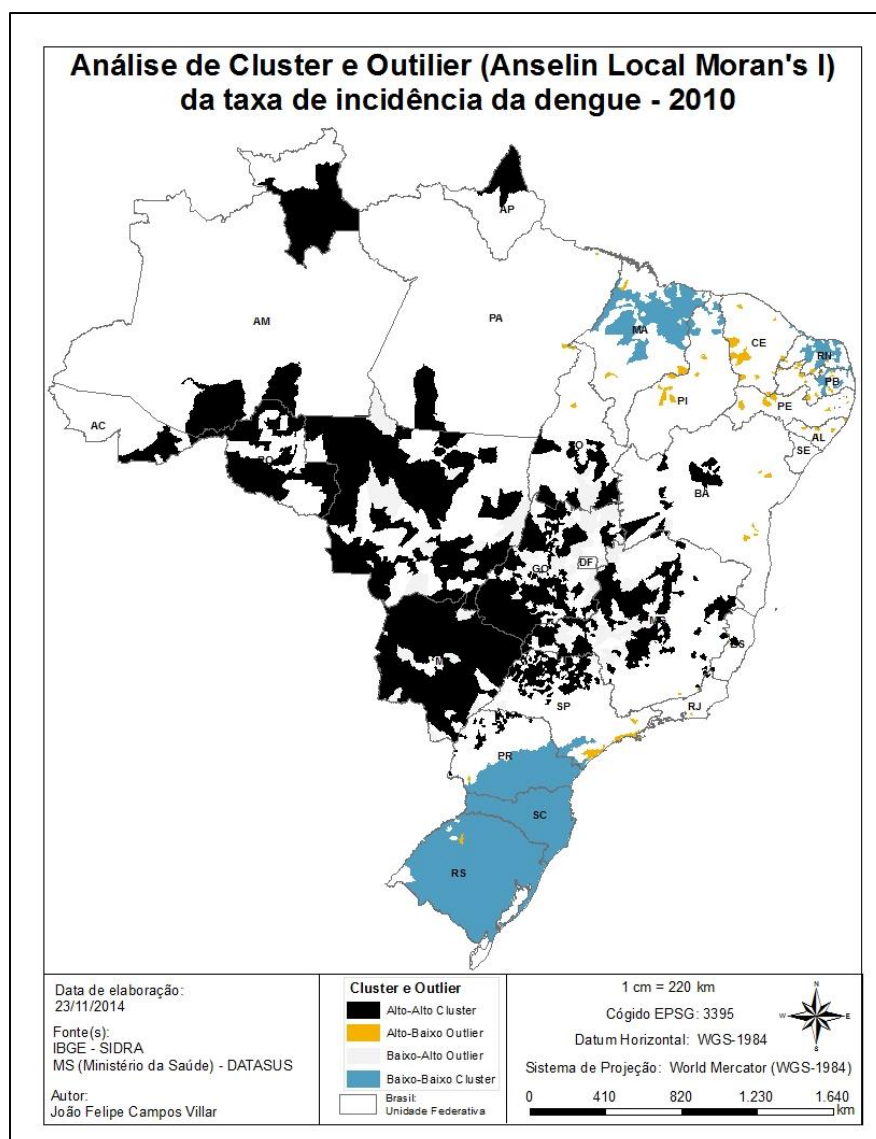


**FIGURA 28 – Análise do índice local de Moran para o ano de 2008.**

Em 2009 ocorreu um comportamento bem distinto em relação ao ano anterior. Na região Sul e em SP, houve uma redução na área de clusters com valores baixos. Houve um aumento significativo de clusters dos valores elevados da taxa na Região Centro-Oeste do Brasil, sobretudo na porção mais ocidental. Alguns poucos *outliers* do extremo superior podem ser identificados nos estados de MG, SP, MS, Ceará (CE) e Pernambuco (PE) porque alguns estados apenas com siglas e outros não? Padronizar. Já os do extremo inferior estão concentrados na região Nordeste (FIG. 29).



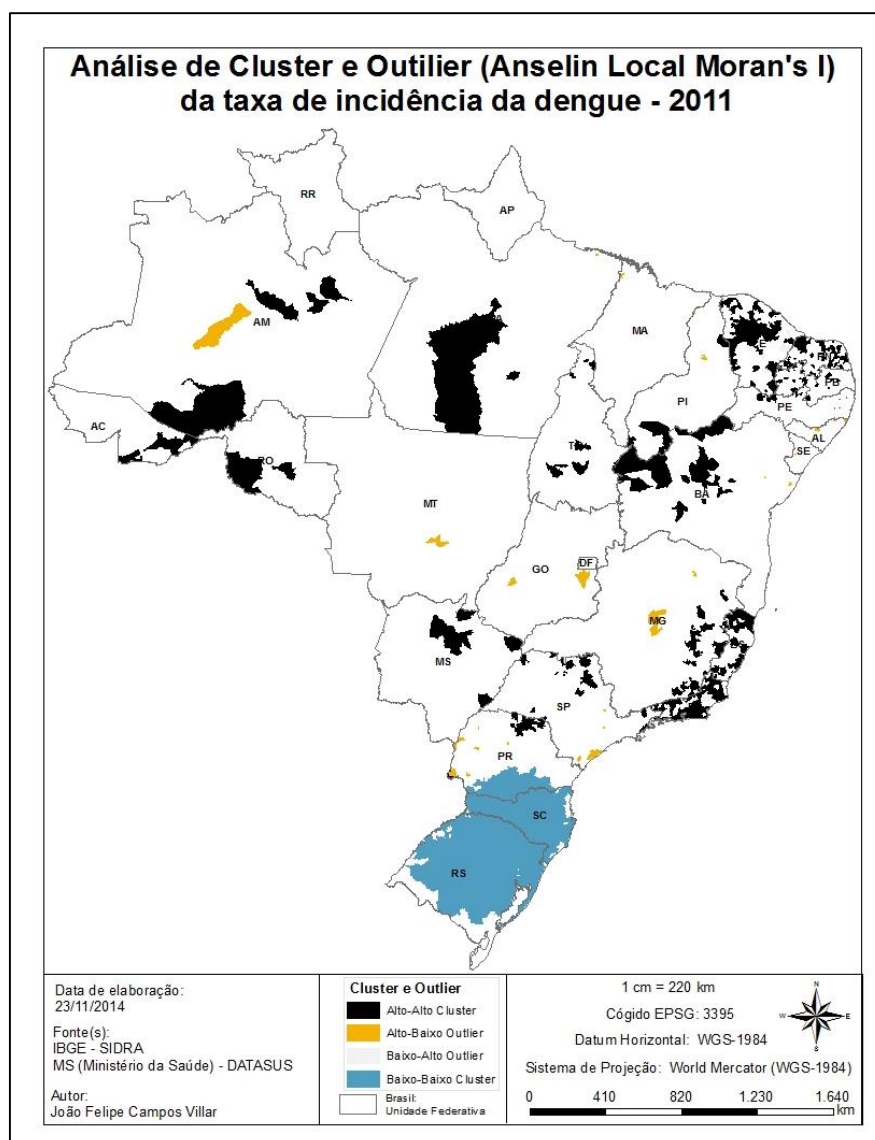
**FIGURA 29 – Análise do índice local de Moran para o ano de 2009.**



**FIGURA 30 – Análise do índice local de Moran para o ano de 2010.**

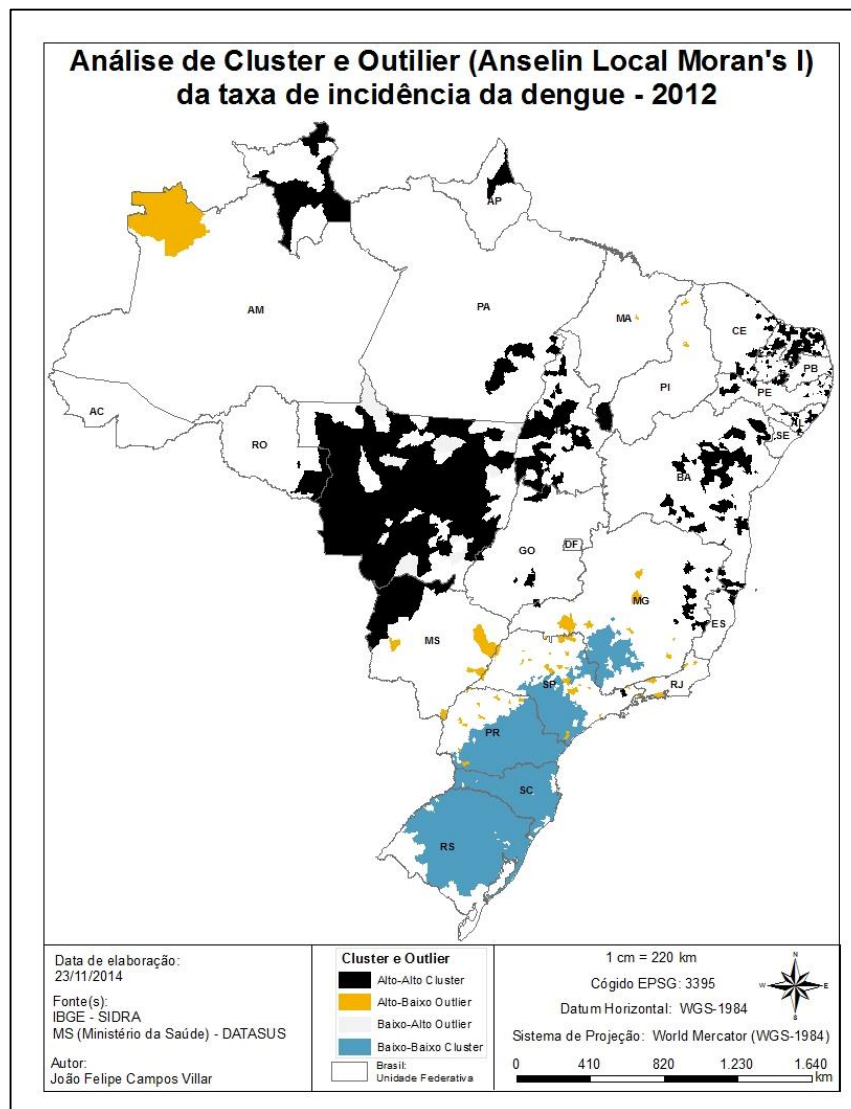
Já em 2010 aconteceu um forte avanço do cluster de valores elevados da taxa de incidência para a porção oriental e setentrional do Brasil, sobretudo para os estados de MG, SP, Amazonas (AM) e Acre (AC). O cluster de baixo valor da taxa ganhou força novamente na região Sul e ocorreu em alguns estados da região Nordeste em regiões mais próximas do litoral do Brasil. Alguns *outliers* do extremo superior podem ser detectados na região Nordeste e em SP (FIG. 30).

Comparado com o ano anterior, em 2011 ocorreu um drástico desaparecimento do cluster de valores elevados, tendo esse apresentado um comportamento mais disperso no território do país. Somente em alguns estados pode ser observada a concentração espacial desse clusters como observado no RJ, Espírito Santo (ES) e alguns outros estados da Região Nordeste. Houve uma discreta redução do cluster de valores baixos da taxa na região Sul, sobretudo no estado do Paraná (PR). Alguns *outliers* do extremo superior novamente ocorreram de forma mais dispersa no país, sendo que alguns poucos *outliers* do extremo inferior puderam ser identificados (FIG. 31).



**FIGURA 31 – Análise do índice local de Moran para o ano de 2011.**

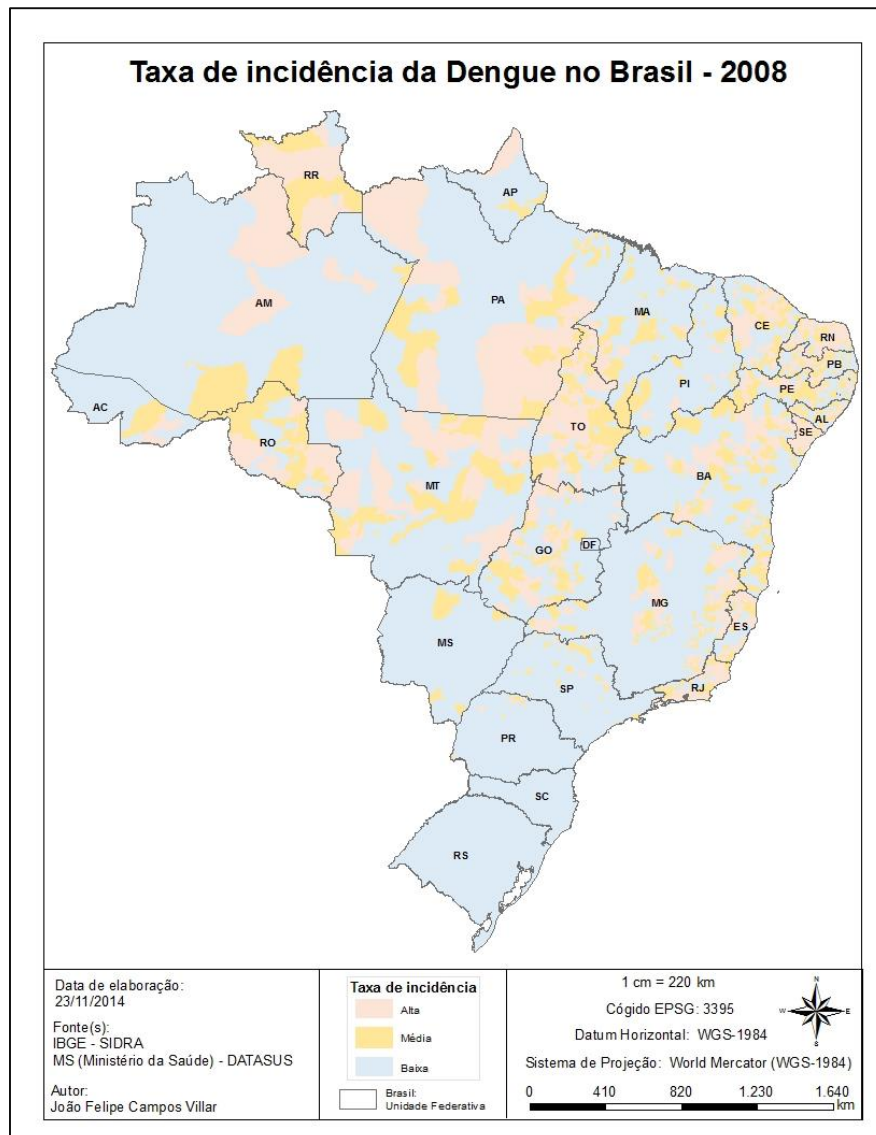
Finalmente em 2012, pode ser observado um comportamento bem distinto em relação aos anos anteriores. O cluster de valores baixos da taxa de incidência avançou de forma significativa da região sul em direção aos estados de SP e MG. Já o cluster de valores elevados sofreu forte redução territorial ao se comparar com o ano de 2010, uma vez que a epidemia da dengue se manifesta em períodos aproximados de 2 anos (MINISTÉRIO DA SAÚDE - BRASIL, 2010), o ano anterior pode não ser uma boa base comparativa, sugerindo até um aumento na área do cluster. Entretanto ocorreu um aumento desse mesmo cluster na região Nordeste, especialmente nos estados com maior porção de terra em área litorânea. Os *outliers* do extremo superior ocorreram de forma um pouco mais clusterizada (agrupada), especialmente na região Sudeste do país (FIG. 32).



**FIGURA 32 – Análise do índice local de Moran para o ano de 2012.**

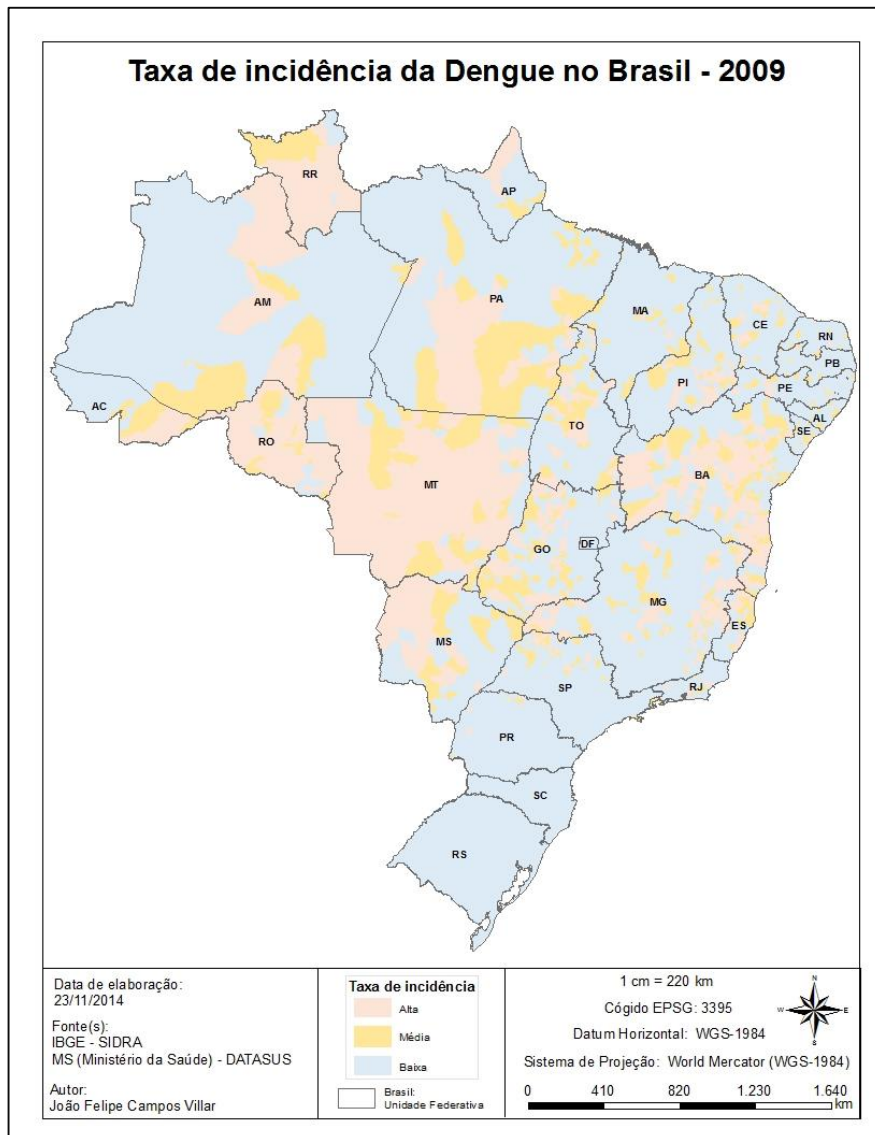
Realizada a análise do índice local de Moran, a análise espacial final foi dedicada aos valores absolutos da taxa de incidência da dengue em todos os anos. Foi observado um comportamento mais severo em anos específicos seguido no ano posterior de um comportamento mais brando, corroborando em muito com a literatura brasileira que indica ciclos epidêmicos a cada dois anos.

O ano de 2008 pode ser considerado um ano de baixa incidência da dengue na maior parte do território brasileiro. Pode ser observada nesse ano uma ocorrência de média e alta taxa de incidência de forma mais dispersa na porção setentrional do país tomando como referência o estado de São Paulo (FIG. 33).



**FIGURA 33 – Taxa de incidência da dengue no Brasil em 2008.**

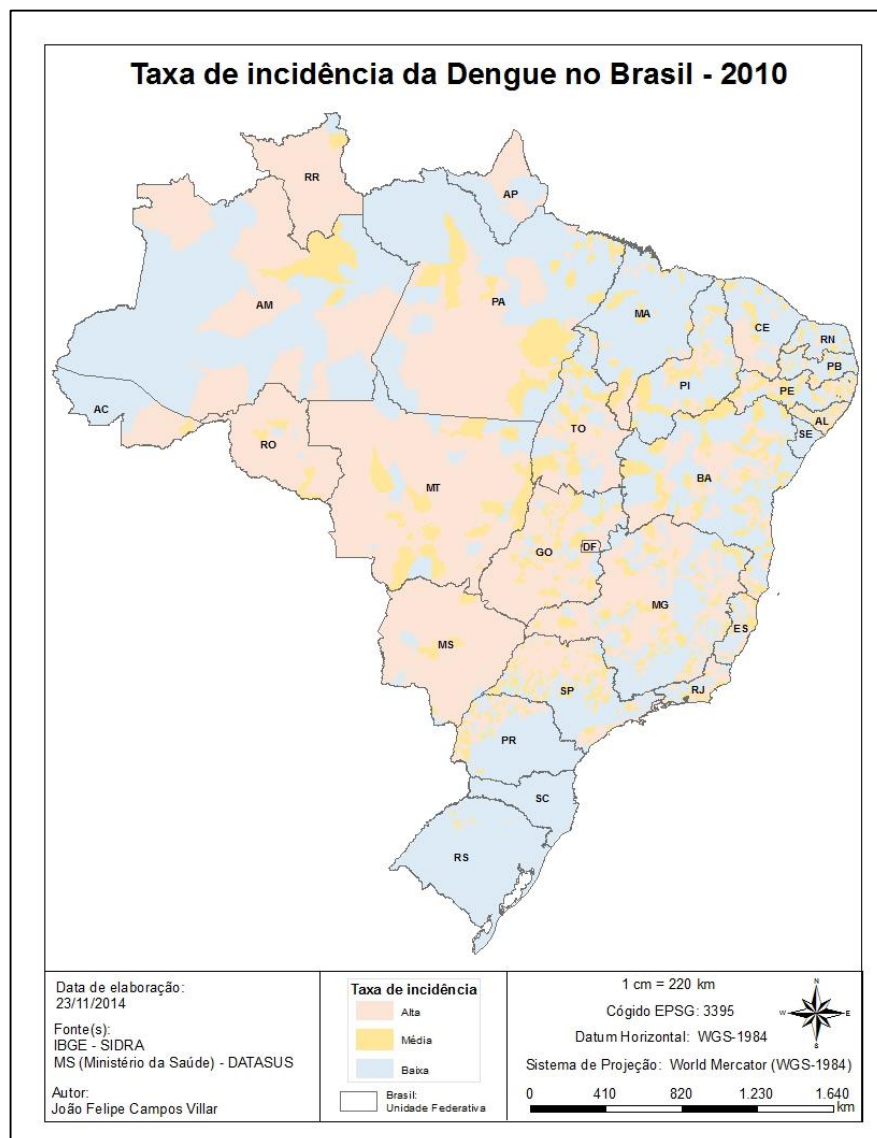




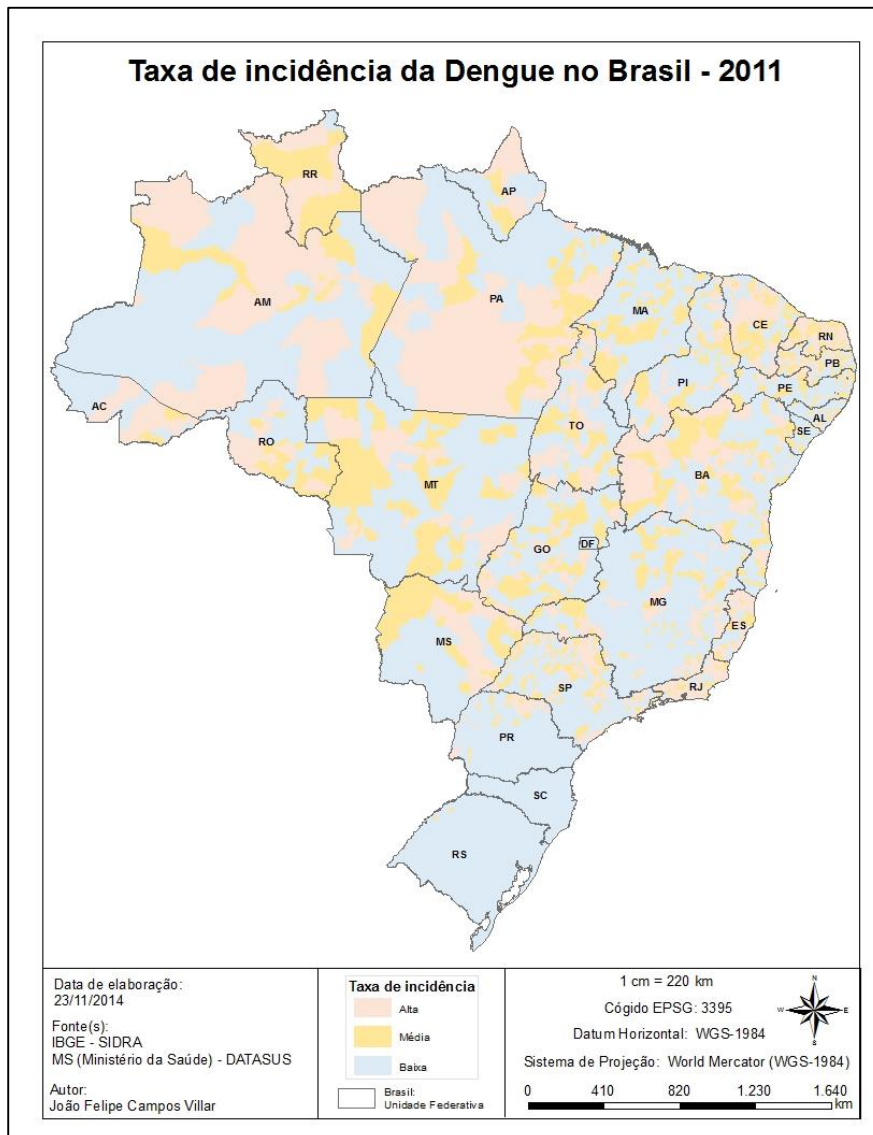
**FIGURA 34 – Taxa de incidência da dengue no Brasil em 2009.**

No ano de 2009 pode ser observado uma expansão da alta taxa de incidência na região Centro-Oeste do Brasil, entretando o padrão disperso da alta e média taxa continua atuando na porção setentrional do país. A maioria do território nacional continua sobre o domínio da baixa taxa de incidência (FIG. 34).

O ano de 2010 pode ser considerado um ano epidêmico para a dengue no Brasil (MINISTÉRIO DA SAÚDE - BRASIL, 2010), uma vez que a maior parte do território brasileiro está sobre o domínio da alta taxa de incidência, compactuando com informações inerentes aos relatórios do MS. Entretanto, essa alta taxa continua atuando na porção setentrional do país a partir do estado de São Paulo. Na região sul a baixa taxa de incidência ainda predomina, somente o estado do Paraná (PR) sofreu na porção noroeste uma alteração para a alta taxa de incidência (FIG. 35).

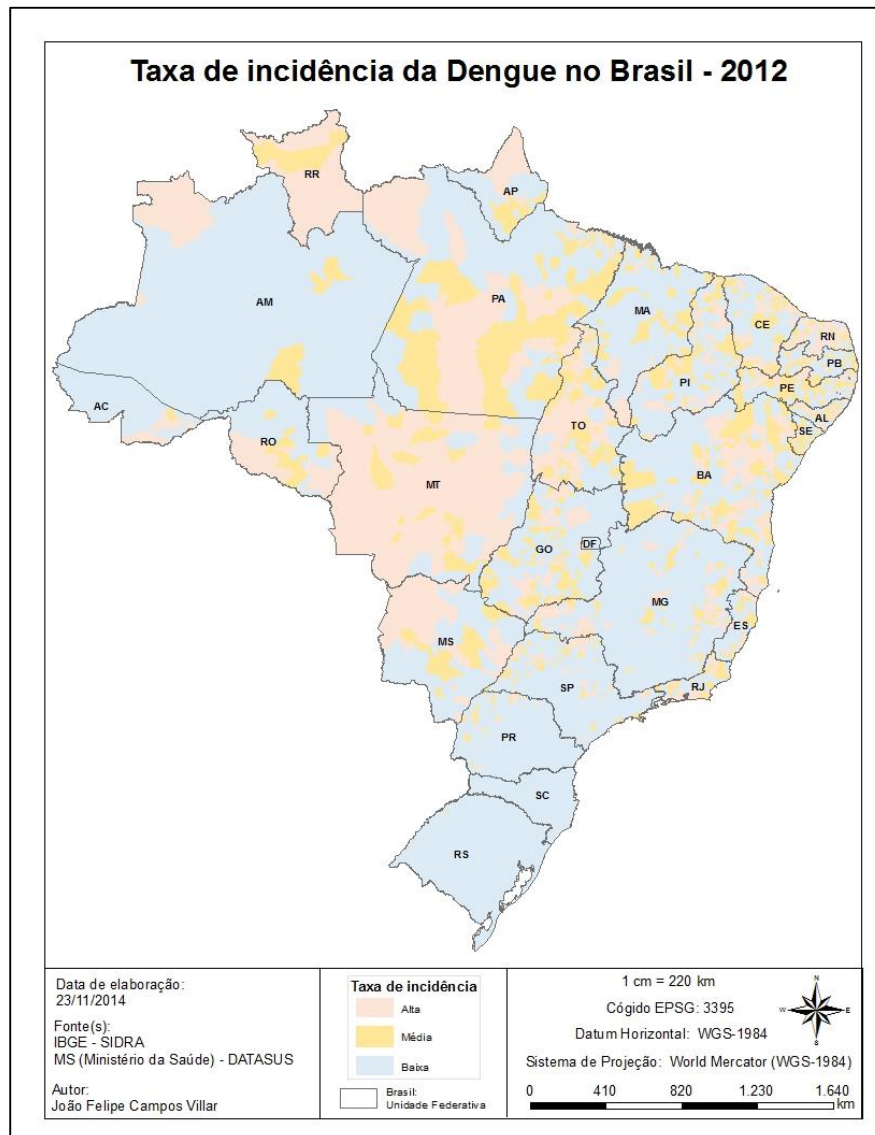


**FIGURA 35 – Taxa de incidência da dengue no Brasil em 2010.**



**FIGURA 36 – Taxa de incidência da dengue no Brasil em 2011.**

No ano de 2011 pode ser observada uma contração na área afetada pela alta taxa de incidência, especialmente na região Centro-Oeste e no estado de Minas Gerais. Porém os estados do Rio de Janeiro e Espírito Santo obtiveram um aumento na alta taxa de incidência. As regiões Norte, Nordeste e Sul continuaram com um comportamento semelhante ao do ano anterior, somente com uma discreta diminuição da taxa alta no estado do Paraná (FIG. 37).



**FIGURA 37 – Taxa de incidência da dengue no Brasil em 2011.**

Já o ano de 2012 apresentou um comportamento semelhante ao ano anterior, somente os estados de Mato Grosso e Tocantins apresentaram uma maior área afetada pela alta taxa de incidência da dengue. Na região Norte, mais especificamente no estado do Amazonas, pode ser observado uma transição de alta para baixa taxa de incidência. Sendo que nas demais regiões o comportamento se manteve mais ou menos estável (FIG. 37).

De maneira geral os resultados da análise espacial mostraram uma forte relação da taxa de incidência da dengue com a localização geográfica em conjunto com o ano analisado. Demonstrando dessa forma que os casos da dengue sofrem forte influência de um componente espaço temporal.

## **6.3 Resultados dos clusters**

A descrição dos clusters encontrados pelo SOM é extremamente necessária para o objetivo proposto por esse estudo. A avaliação desses resultados foi subdividida em dois tópicos.

No primeiro tópico foi caracterizado o perfil de cada cluster encontrado de acordo com as médias de cada variável analisada. O segundo tópico caracteriza geograficamente a distribuição espacial desses clusters para cada ano do período analisado no intuito de se identificar padrões espaciais.

### **6.3.1 Caracterização do perfil dos clusters**

Através da rede neural SOM foi possível a identificação de clusters para cada ano do período de análise. Correlacionando um dado cluster com a média de cada variável selecionada dentro do respectivo cluster, foi possível determinar padrões e alguns comportamentos dessas variáveis. Foi também possível relacionar a classificação da taxa de incidência com cada cluster, definindo assim um perfil para cada classe da taxa, baseado nos padrões encontrados em cada cluster.

No ANEXO C, pode ser observado o padrão de cada cluster para cada ano, através da análise da média de cada variável do grupo socioeconômico, ambiental e geográfico.

O ano de 2008 apresentou 4 clusters, sendo que foi possível identificar padrões nos clusters de número 0, 2 e 3. O de número 0 apresenta um padrão extremamente compatível com a taxa alta e média de incidência da dengue (TAB. 5), sendo observado nesse um LIRAA elevado se comparado com a média do ano, precipitação 100mm<sup>3</sup> acima da média geral, baixa amplitude térmica, taxa de analfabetismo elevada, grande área geográfica e elevada densidade demográfica.

No ano de 2009 foi possível identificar 3 clusters, similar ao número de classes da taxa de incidência da dengue. Entretanto a correlação mostrou-se favorável apenas no cluster 1 em relação a uma taxa de incidência da dengue média. O padrão observado no cluster 1 é de baixo PIB per capita e baixa remuneração média. Já o cluster de número 0 apresentou-se bem distribuído entre todas as classes (Alta, Média e Baixa) (TAB. 6).

TABELA 5

Correlação da taxa de incidência da dengue com o cluster identificado pelo SOM em 2008.

INCIDÊNCIA	CLUSTER	N. MUNICÍPIOS	% TOTAL GERAL	% SOBRE INCIDÊNCIA
Alta	0	291	5,23%	34,28%
	1	370	6,65%	43,58%
	2	182	3,27%	21,44%
	3	6	0,11%	0,71%
<b>Alta Total</b>	-	<b>849</b>	<b>15,26%</b>	-
Média	0	263	4,73%	35,11%
	1	301	5,41%	40,19%
	2	175	3,15%	23,36%
	3	10	0,18%	1,34%
<b>Média Total</b>	-	<b>749</b>	<b>13,46%</b>	-
Baixa	0	986	17,72%	24,86%
	1	717	12,89%	18,08%
	2	660	11,86%	16,64%
	3	1.603	28,81%	40,42%
<b>Baixa Total</b>	-	<b>3.966</b>	<b>71,28%</b>	-
<b>Total Geral</b>	-	<b>5.564</b>	<b>100,00%</b>	-

TABELA 6

Correlação da taxa de incidência da dengue com o cluster identificado pelo SOM em 2009.

INCIDÊNCIA	CLUSTER	N. MUNICÍPIOS	% TOTAL GERAL	% SOBRE INCIDÊNCIA
Alta	0	204	3,67%	31,10%
	1	226	4,06%	34,45%
	2	226	4,06%	34,45%
<b>ALTA Total</b>	-	<b>656</b>	<b>11,79%</b>	-
Média	0	156	2,80%	30,23%
	1	218	3,92%	42,25%
	2	142	2,55%	27,52%
<b>MEDIA Total</b>	-	<b>516</b>	<b>9,27%</b>	-
Baixa	0	804	14,45%	18,31%
	1	1261	22,66%	28,71%
	2	2327	41,82%	52,98%
<b>BAIXA Total</b>	-	<b>4392</b>	<b>78,94%</b>	-
<b>Total Geral</b>	-	<b>5.564</b>	<b>100,00%</b>	<b>100,00%</b>

O ano de 2010 foi o que mais apresentou clusters (5). Foi também o ano onde foi possível correlacionar com mais força um cluster (0) com a alta taxa de incidência de dengue. O padrão encontrado para esse cluster foi o de um LIRAA elevado relativamente a média do ano, THEIL acima da média, alta taxa de analfabetismo, grande área geográfica e densidade demográfica elevada. A baixa taxa de incidência de dengue se apresentou bem distribuída entre todos os clusters e a média se apresentou mais concentrada no cluster 0 e 4 (TAB. 7).

TABELA 7

Correlação da taxa de incidência da dengue com o cluster identificado pelo SOM em 2010.

INCIDÊNCIA	CLUSTER	N. MUNICÍPIOS	% TOTAL GERAL	% SOBRE INCIDÊNCIA
ALTA	0	651	11,70%	43,31%
	1	1	0,02%	0,07%
	2	272	4,89%	18,10%
	3	296	5,32%	19,69%
	4	283	5,09%	18,83%
<b>ALTA Total</b>	-	<b>1.503</b>	<b>27,01%</b>	-
MÉDIA	0	287	5,16%	36,94%
	1	3	0,05%	0,39%
	2	118	2,12%	15,19%
	3	128	2,30%	16,47%
	4	241	4,33%	31,02%
<b>MÉDIA Total</b>	-	<b>777</b>	<b>13,96%</b>	-
BAIXA	0	973	17,49%	29,63%
	1	750	13,48%	22,84%
	2	296	5,32%	9,01%
	3	449	8,07%	13,67%
	4	816	14,67%	24,85%
<b>BAIXA Total</b>	-	<b>3.284</b>	<b>59,02%</b>	-
<b>Total Geral</b>	-	<b>5.564</b>	<b>100,00%</b>	-

Em 2011 foram identificados novamente 4 clusters, foi possível encontrar uma boa correlação entre os clusters 0 e 1 com as taxas alta e média de incidência de dengue. Diferente de outros anos, nesse foi possível estabelecer também uma boa correlação entre o cluster 3 e a baixa taxa de incidência de dengue (TAB. 8).

O padrão das variáveis no cluster 0 é o de um LIRAA elevado se comparado com a média, grande área geográfica e elevada densidade demográfica. No cluster 1 uma alta taxa de analfabetismo. Já no cluster 3 foi observado um LIRAA bem abaixo da média, temperatura mínima baixa, amplitude térmica elevada, um PIB per capita um pouco acima da média, baixa taxa de analfabetismo, pequena área geográfica, pouca mobilidade urbana e baixa densidade demográfica.

TABELA 8

Correlação da taxa de incidência da dengue com o cluster identificado pelo SOM em 2011.

INCIDÊNCIA	CLUSTER	N. MUNICÍPIOS	% TOTAL GERAL	% SOBRE INCIDÊNCIA
Alta	0	326	5,86%	33,68%
	1	321	5,77%	33,16%
	2	228	4,10%	23,55%
	3	93	1,67%	9,61%
<b>ALTA Total</b>	-	<b>968</b>	<b>17,40%</b>	-
Média	0	291	5,23%	32,70%
	1	307	5,52%	34,49%
	2	223	4,01%	25,06%
	3	69	1,24%	7,75%
<b>MEDIA Total</b>	-	<b>890</b>	<b>16,00%</b>	-
Baixa	0	923	16,59%	24,91%
	1	760	13,66%	20,51%
	2	566	10,17%	15,27%
	3	1457	26,19%	39,31%
<b>BAIXA Total</b>	-	<b>3.706</b>	<b>66,61%</b>	-
<b>Total Geral</b>	-	<b>5.564</b>	<b>100,00%</b>	-

TABELA 9

Correlação da taxa de incidência da dengue com o cluster identificado pelo SOM em 2012.

INCIDÊNCIA	CLUSTER	N. MUNICÍPIOS	% TOTAL GERAL	% SOBRE INCIDÊNCIA
Alta	0	261	4,69%	31,03%
	1	322	5,79%	38,29%
	2	215	3,86%	25,56%
	3	43	0,77%	5,11%
<b>ALTA Total</b>	-	<b>841</b>	<b>15,12%</b>	-
Média	0	288	5,18%	36,59%
	1	282	5,07%	35,83%
	2	174	3,13%	22,11%
	3	43	0,77%	5,46%
<b>MEDIA Total</b>	-	<b>787</b>	<b>14,14%</b>	-
Baixa	0	991	17,81%	25,18%
	1	784	14,09%	19,92%
	2	628	11,29%	15,96%
	3	1533	27,55%	38,95%
<b>BAIXA Total</b>	-	<b>3.936</b>	<b>70,74%</b>	-
<b>Total Geral</b>	-	<b>5.564</b>	<b>100,00%</b>	-

Finalmente em 2012 foi definido pelo SOM novamente 4 clusters. Foi observado um comportamento da correlação entre os clusters a taxa de incidência de dengue muito similar ao ano anterior, em que dois clusters estão fortemente relacionados com a taxa média e alta de incidência e o cluster 3 muito relacionado a baixa taxa (TAB. 9). Os padrões encontrados foram praticamente os mesmos do ano anterior reforçando a relação de LIRAA elevado relativo a média do ano, baixa amplitude térmica, alta taxa de analfabetismo, grande área



geográfica e densidade demográfica elevada para a alta e média taxa de incidência. Amplitude térmica alta, baixa temperatura mínima, baixa taxa de analfabetismo, pequena área geográfica, pouca mobilidade urbana e baixa densidade demográfica novamente relaciona com uma baixa taxa de incidência de dengue.

### 6.3.2 Caracterização geográfica

Depois de correlacionada a taxa de incidência com os clusters definidos através da análise da matriz-u da rede neural SOM, foi ainda possível elaborar uma representação geográfica desses clusters para cada ano do período de análise. É possível observar que no ano de 2008 o cluster 0 está localizado em sua grande parte na região Norte do Brasil, disperso no estado de Minas Gérias e concentrado nos estados

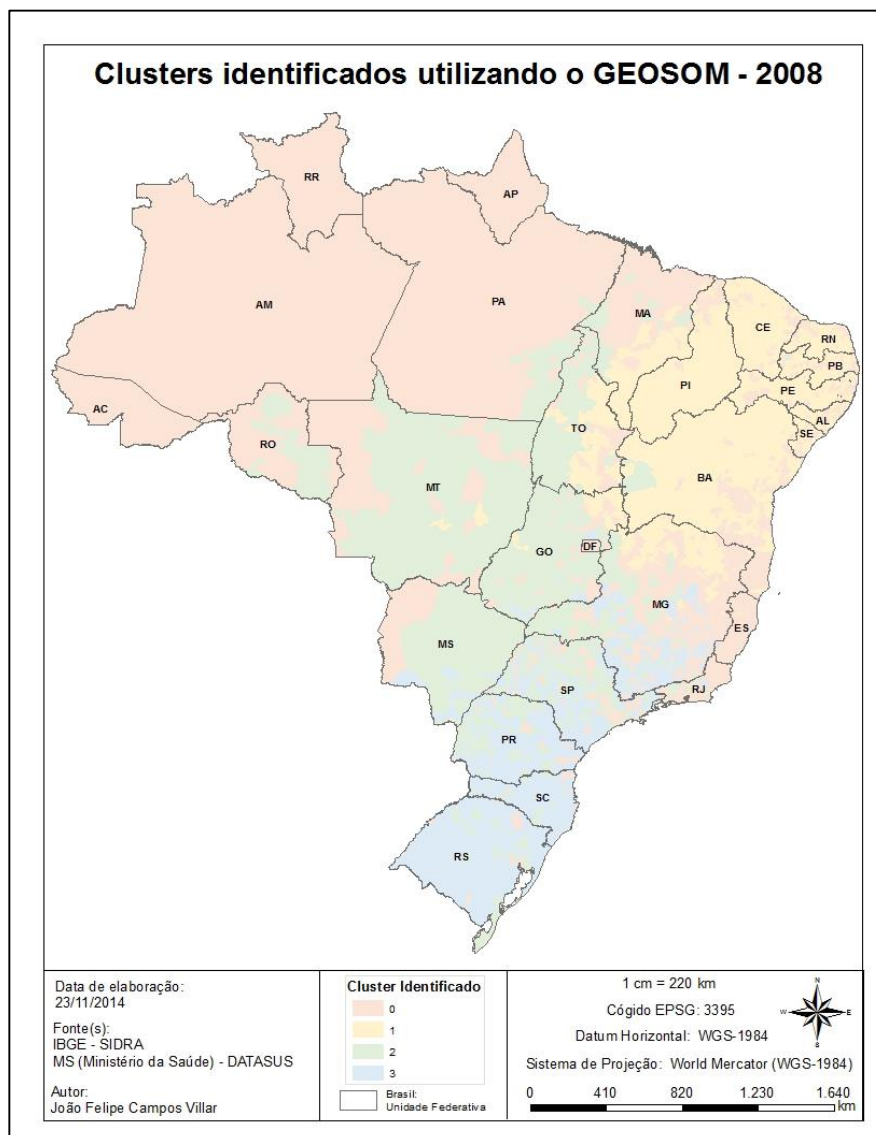
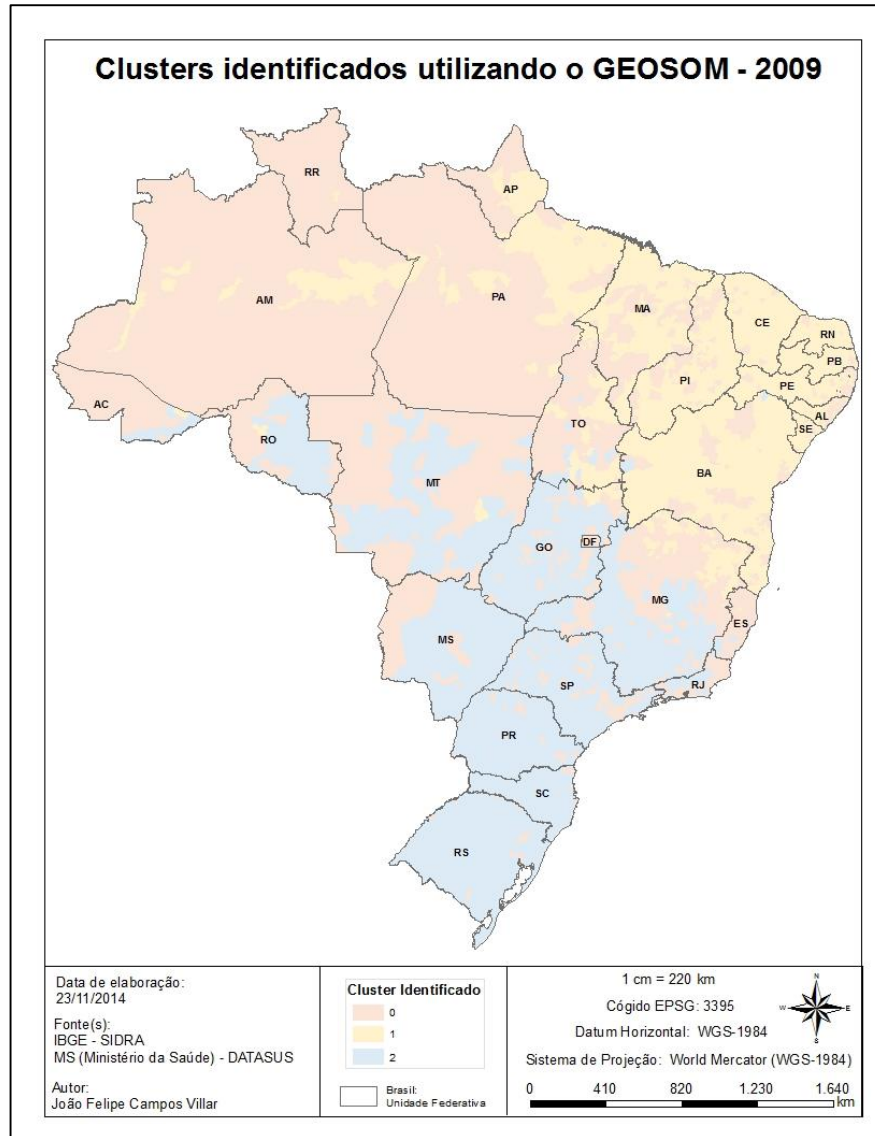


FIGURA 38 – Distribuição espacial dos clusters identificados pelo GeoSOM Suite em 2008.

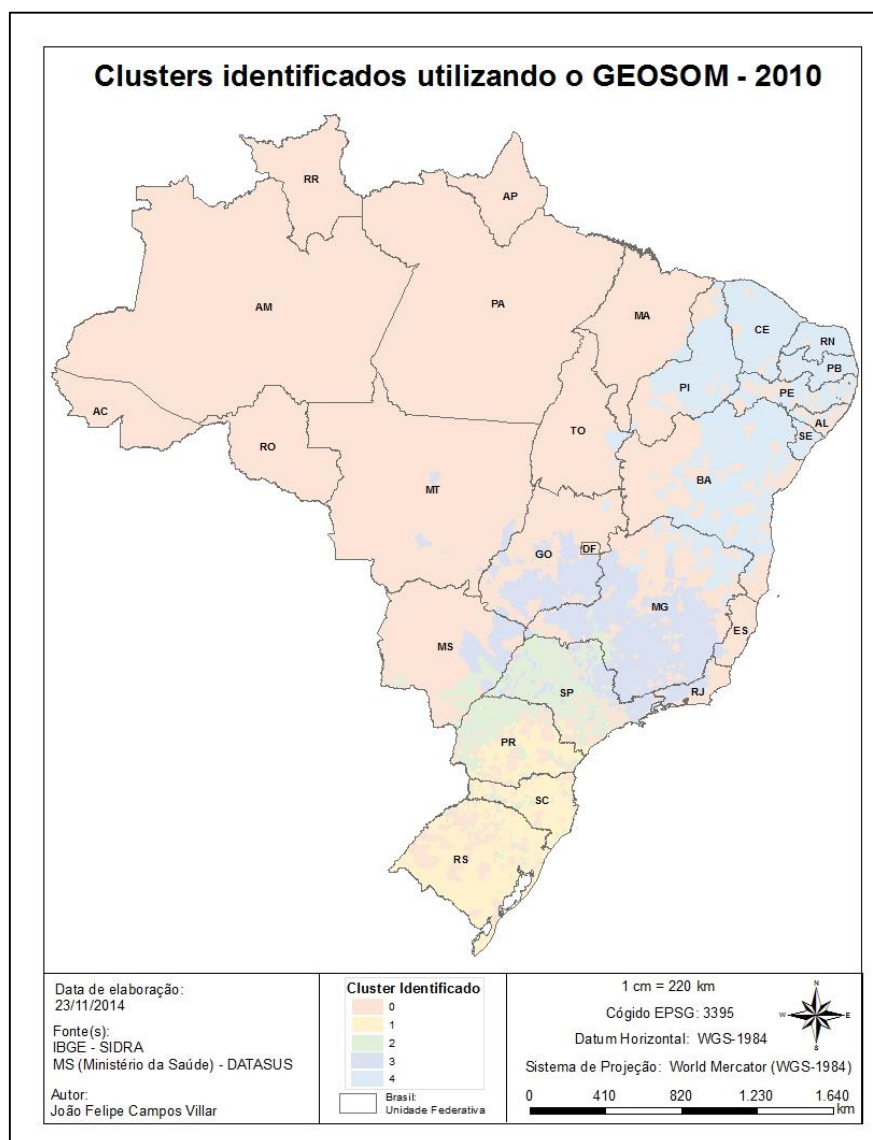
do Rio de Janeiro e Espírito Santo. O 1 está concentrado na região Nordeste, seguido do 2 na região Centro-Oeste. O cluster de número 3 está localizado em sua maior porção na região Sul. Os Estados de Minas Gerais e São Paulo são mais heterogêneos em relação a todos os clusters (FIG. 38).



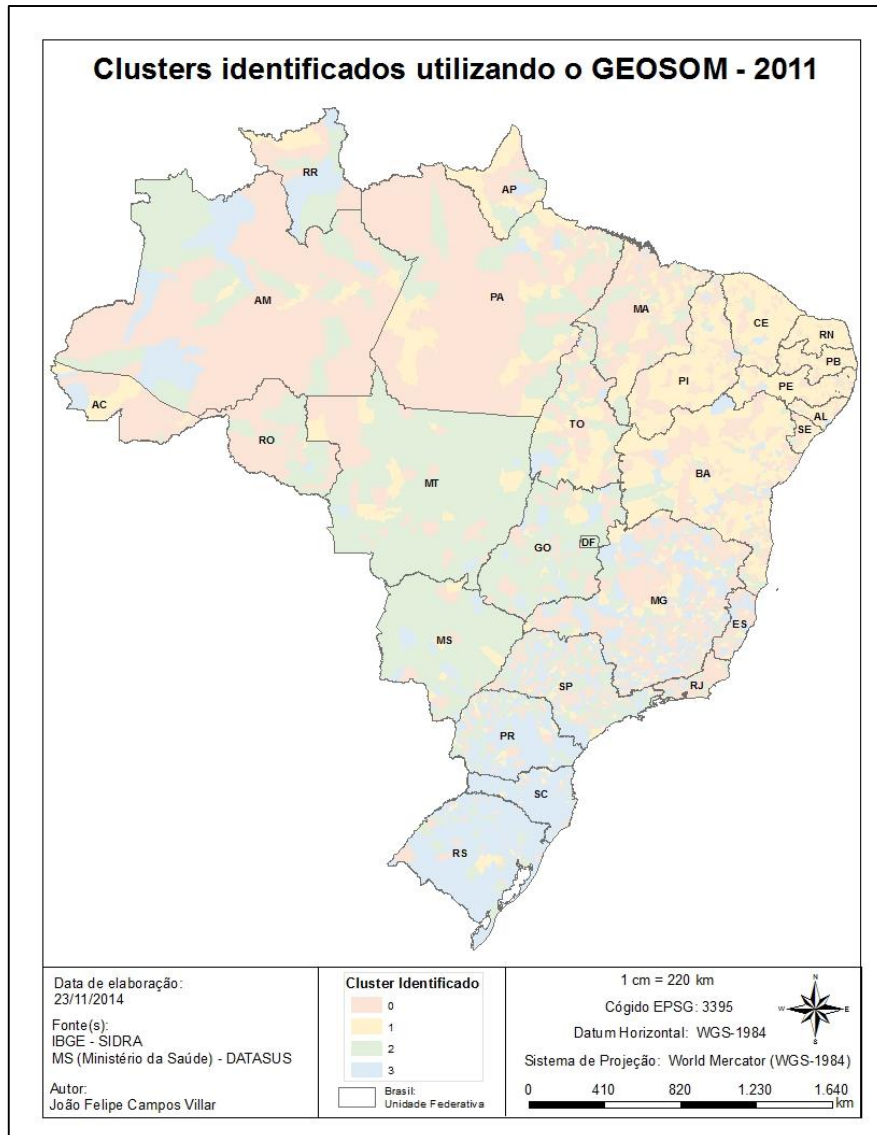
**FIGURA 39 – Distribuição espacial dos clusters identificados pelo GeoSOM Suite em 2009.**

Em 2009 os clusters encontrados possuem um comportamento ainda mais agrupado e concentrado em relação ao ano anterior. O Cluster 0 novamente predomina na região Norte e nos estados do Rio de Janeiro e Espírito Santo. O cluster 1 no nordeste. A principal diferença está no fato do cluster 2 está localizado na região Centro-Oeste, Sudeste e Sul, possivelmente devido ao menor número de clusters encontrados comparativamente ao ano de 2008.

Já 2010 foi um ano que apresentou uma maior quantidade e heterogeneidade espacial em relação aos clusters, porém pode ser observado que o cluster de número 0 está fortemente localizado nas regiões Norte e Centro-Oeste e nos estados do Rio de Janeiro e Espírito Santo. O estado de Minas Gerais apresenta predomínio maior do cluster 4, mas com fortes resquícios do cluster 0. São Paulo possui alta variabilidade de clusters. Na região Nordeste predomina o cluster 4. A região sul se caracteriza fortemente pelo cluster 1, mas o estado do Paraná apresenta comportamento mais heterogêneo em relação aos clusters (FIG. 40).



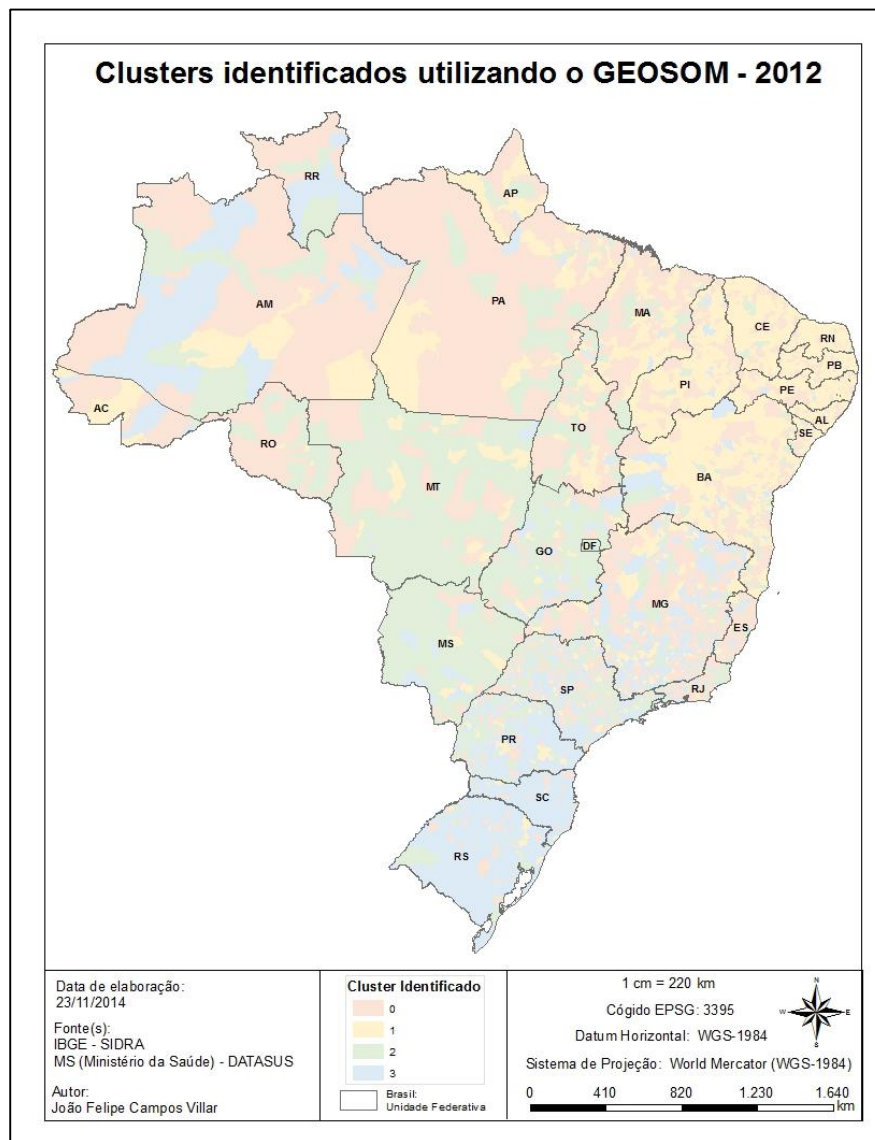
**FIGURA 40 – Distribuição espacial dos clusters identificados pelo GeoSOM Suite em 2010.**



**FIGURA 41 – Distribuição espacial dos clusters identificados pelo GeoSOM Suite em 2011.**

O ano de 2011 foi claramente, o que apresentou uma maior heterogeneidade espacial quanto a distribuição de clusters. Somente na região Centro-Oeste pode ser observado um predomínio maior do cluster 2, seguido da região Sul pelo cluster de número 3. É interessante destacar que novamente os estados de Minas Gerais, São Paulo e Paraná apresentaram um comportamento bem diversificado em relação aos clusters. (FIG. 41).

Em 2012 novamente pode ser observado um comportamento extremamente heterogêneo na distribuição de clusters pelo território brasileiro. Todavia a região Sul apresentou um comportamento mais homogêneo em relação ao cluster de número 3. O estado de Minas Gerais mais uma vez apresenta diversidade no número de clusters, porém nesse ano os estados do Amazonas e Pará apresentaram esse mesmo comportamento. São Paulo obteve uma menor diversidade de clusters, sendo que o cluster de número 2 está mais presente em seu território. Na região Centro-Oeste o cluster de número 2 ainda é predominante e na região Nordeste o cluster 1 se apresenta mais marcante (FIG. 42)



**FIGURA 42 – Distribuição espacial dos clusters identificados pelo GeoSOM Suite em 2012.**

## 7. Limitações

Como em todo estudo epidemiológico ecológico baseado em dados de registros disponibilizados pelos órgãos de saúde, uma subestimação de casos pode ocorrer.

Os dados meteorológicos não puderam ser encontrados para cada município analisado, sendo necessária a interpolação dos dados para todo o território brasileiro. Essa interpolação pode gerar alguns resultados um pouco desviados da realidade sobretudo em regiões com menor densidade de pontos.

Relativamente ainda aos dados meteorológicos, devida a grande quantidade de informação e recursos computacionais utilizados, bem como a escala de trabalho ser anual para todos os dados analisados, não foi possível a análise desses dados de forma mensal, sendo os mesmos agregados anualmente. Essa agregação de dados pode ocultar alguns resultados no tocante a sazonalidade, não sendo possível detectar alguns comportamentos micro climáticos.

## 8. Discussão e conclusão

Atualmente no Brasil a dengue se apresenta como um dos principais problemas de saúde pública, por se tratar de uma doença multifatorial e que está evoluindo rapidamente nos últimos anos. Estudos que auxiliem na elucidação dos principais fatores determinantes da dengue, sobretudo aqueles relacionados com a epidemiologia ecológica se mostram extremamente úteis para uma melhor compreensão e futuro auxílio na criação de mecanismos de controle para a doença.

A utilização dos Sistemas de Informação Geográfica (SIG) em conjunto com a mineração de dados, especificamente as redes neurais SOM se mostraram extremamente úteis para análise espaço temporal dos casos da dengue no Brasil (BARCELLOS, 2003).

Foi possível estabelecer padrões baseados no SOM correlacionados com a taxa de incidência da dengue. Através do agrupamento de variáveis determinantes da dengue, que foram selecionadas através da literatura e obtiveram comprovação através de análises exploratória, foi possível a aplicação eficiente da rede neural como ferramenta para agrupamento de observações. Também foi possível através dessa, a identificação de comportamentos específicos para algumas variáveis que persistentemente apresentaram o mesmo comportamento correlacionado com a taxa de incidência em anos diferentes. Por exemplo, a variável LIRAa sempre que relacionada com uma alta taxa de incidência apresentou valores elevados, caracterizando assim a importância do vetor no ciclo epidêmico da doença.

Já as variáveis ambientais se mostraram extremamente importantes na incidência da doença. Temperaturas mínimas e amplitudes térmicas altas sempre estão relacionadas com uma baixa taxa de incidência da dengue, isso evidencia mais uma vez fatores que possam gerar interferência no vetor impactam no ciclo da doença (CATÃO, 2012).

Variáveis socioeconômicas também podem interferir na taxa de incidência da doença, uma vez que à alta e a média taxa de incidência apresentaram relação com taxa de analfabetismo elevada, PIB per capita baixo e inversamente proporcional a taxa de incidência baixa apresentou taxa de analfabetismo baixa e PIB per capita alto (DONALISIO, 1999). Resultados que corroboram para a interferência de questões mais inerentes a população afetada como grau de conhecimento sobre a doença e condições de vida.

Contudo, variáveis de cunho mais geográfico mostraram que fatores como o tamanho da área afetada e densidade demográfica são fortemente relacionados a taxa de incidência da

doença, evidenciando assim o caráter geográfico e demográfico que influencia praticamente todas as doenças do tipo infecto contagiosas. Dessa maneira o objetivo principal desse estudo foi alcançado, pois foi possível de forma bem produtiva identificar o perfil e o comportamento da dengue no Brasil no período de 2008 até 2012.

Para mais além, o SOM aliado ao SIG pode de maneira significativa vir a auxiliar os órgãos de vigilância epidemiológica regionais e locais para o efetivo controle e monitoramento da dengue, através da identificação de perfil e comportamento ao longo do tempo e do espaço geográfico (BARCELLOS e BASTOS, 1996).

Entretanto, algumas considerações devem ser feitas, como a dependência de fontes confiáveis de informação nesse tipo de estudo, uma vez que a subestimação de dados pode alterar de forma mais significativa os resultados gerados por um conjunto de observação pequeno.

Portanto, o método SOM pode colaborar mais significativamente em escalas que favoreçam um número maior de registros por variável (HENRIQUES, 2010), onde terão menos efeito fatores como o sub registro de dados. Também é de extrema importância a correta normalização dos dados, uma vez que o SOM cruza informações entre variáveis para o aprendizado efetivo da rede, assim variáveis de escalas muito diferentes podem afetar diretamente os resultados.



## Referências Bibliográficas

- ANSELIN, L. Local Indicators of Spatial Association – LISA, v. 27, n. 2, p. 93-115, 1995.
- APARÍCIO, C. **Utilização de Geoprocessamento e Senoriamento Remoto Orbital para análise espacial de paisagem com incidência de Leishmaniose Tegumentar Americana.** USP. São Paulo, p. 104. 2001.
- ASTUDILLO, C.; OOMMEN, J. Topology-oriented self-organizing maps: a survey. **Pattern Analysis and Applications**, London, v. 17, n. 2, 2014.
- AYOANDE, J. O. **Introdução à climatologia para os trópicos.** 13. ed. Rio de Janeiro: Bertrand Brasil, 2010.
- BAÇÃO, F.; LOBO, V.; PAINHO, M. The self-organizing map, the Geo-SOM, and relevant variants for geosciences. **Computers & Geosciences**, n. 31, 2005. ISSN 155-163.
- BARBOSA, J. P. M. UTILIZAÇÃO DE MÉTODO DE INTERPOLAÇÃO PARA ANÁLISE E ESPACIALIZAÇÃO DE DADOS CLIMÁTICOS: O SIG COMO FERRAMENTA1. **CAMINHOS DE GEOGRAFIA**, São Paulo, v. 9, n. 17, p. 85 - 96, 2006.
- BARCELLOS, C. A saúde nos Sistemas de Informação Geográfica: apenas uma camada a mais? **Caderno prudentino de Geografia**, Presidente Prudente, v. 25, p. 29-43, 2003.
- BARCELLOS, C.; BASTOS, F. I. Geoprocessamento, ambiente e saúde: uma união possível? **Cadernos de Saúde Pública**, Rio de Janeiro, v. 3, n. 12, p. 387-7, 1996.
- BARRETO, J. **Introdução às Redes Neurais Artificiais.** Universidade Federal de Santa Catarina. Florianópolis, p. 50. 2002.
- BOHLING, G. **KRIGING.** Kansas Geological Survey. Kansas, p. 1-20. 2006.
- BRASIL. **Constituição Federal da República Federativa do Brasil.** Distrito Federal: Senado Federal, 1988. 292 p.
- BRASIL. **Dengue:** instruções para pessoal de combate ao vetor: manual de normas técnicas. 3. ed. Brasília: Ministério da Saúde, 2001.
- BRASSEL, K.; REIF, D. A Procedure to Generate Thiessen Polygons. **Geographical Analysis**, v. 11, n. 3, p. 289-303, 1978.
- CARDON, A.; MÜLLER, D. N. **Introdução Às Redes Neurais Artificiais.** Universidade Federal do Rio Grande do Sul. Porto Alegre, p. 30. 1994.
- CATÃO, R. D. C. **Dengue no Brasil:** Abordagem Geográfica na Escala Nacional. 1. ed. São Paulo: Cultura Acadêmica, v. I, 2012. 178 p.
- CORDEIRO, M. T. **Evolução da dengue no estado de Pernambuco, 1987-2006: epidemiologia e caracterização molecular dos sorotipos circulantes.** Fundação Oswaldo Cruz. Recife, p. 225. 2008.

CORREIA, V. R. D. M. et al. Uma aplicação do sensoriamento remoto para a investigação de endemias urbanas. **Caderno de saúde pública**, Rio de Janeiro, p. 14, 2007.

DA COSTA, I. P.; NATAL, D. Distribuição espacial da dengue e determinantes socioeconômicos em localidade urbana no Sudeste do Brasil. **Revista de Saúde Pública**, São Paulo, v. 3, n. 32, p. 232-6, 1998.

DA SILVA, R. R.; BACHA, C. J. C. Polígonos de Voronoi como alternativa aos problemas das áreas mínimas comparáveis: uma análise das mudanças populacionais na região norte do Brasil. **Revista Brasileira de Estudos da População**, São Paulo, v. 28, n. 1, p. 133 - 151, Jan/Jun 2011.

DE SOUZA, J. L. L. et al. **Avaliação de métodos de interpolação aplicados à espacialização das chuvas no território identidade Portal do Sertão / Bahia**. Anais XV Simpósio Brasileiro de Sensoriamento Remoto - SBSR. Curitiba: [s.n.]. 2011. p. 4295-4302.

DONALISIO, M. R. **O dengue no espaço habitado**. Hucitec/Funcraf. São Paulo. 1999.

DOS SANTOS, J. M. E. C. **Avaliação da evolução do perfil regional em inovação: O caso das regiões europeias 2003/2009**. ISEGI - Universidade Nova de Lisboa. Lisboa, p. 69. 2013.

FERRUZZA, G. DECISION TECHNOLOGIES IN DATABASE MARKETING: PART IX. **The On-Line Executive Journal for Data-Intensive Decision Support**, San Diego, v. 2, n. 10, p. 1-4, 1998.

FILHO, D. A.; ROUQUAYROL, M. Z. **Introdução à EPIDEMIOLOGIA**. 4. ed. Rio de Janeiro: KOOGAN, 2006.

FLAUZINO, R. F.; SOUZA-SANTOS, R.; OLIVEIRA, R. M. Dengue, geoprocessamento e indicadores socioeconômicos e ambientais: um estudo de revisão. **Revista Panamericana de Saúde Pública**, Washington, v. 25, n. 5, p. 456-461, 2009.

FRANCO, O. **História da febre-amarela no Brasil**. Departamento Nacional de Endemias Rurais. Rio de Janeiro, p. 208. 1969.

GLADWELL, M. **Fora de série: outliers**. 1. ed. Rio de Janeiro: Sextante, 2008.

GUBLER, D. J. Dengue and Dengue Hemorrhagic Fever. **Clinical Microbiology Reviews**, v. 11, n. 3, p. 16, 1998.

GURNEY, K. **An Introduction to Neural Networks**. 1. ed. Bristol, PA, USA: Taylor & Francis, Inc., v. 1, 1997.

HENCHAL, E. A.; PUTNAK, J. R. The Dengue Viruses. **Clinical Microbiology Reviews**, v. 3, n. 4, p. 21, 1990.

HENRIQUES, R. A. P. **Artificial Intelligence in Geospatial Analysis: applications of Self-Organizing Maps in the context of Geographic Information Science**. Universidade Nova de Lisboa. Lisboa, p. 1-451. 2010.

HENRIQUES, R.; BACAO, F.; LOBO, V. Exploratory geospatial data analysis using the GeoSOM suite. **Computers, Environment and Urban Systems**, n. 36, p. 218-232, 2012.

- HUGH-JONES, M. Applications of remote sensing to the identification of the habitats of parasites and disease vectors. **Parasitol Today**, v. 5, p. 8, 1989.
- KANSKI, S.; NIKKILÄ, J.; KOHONEN, T. **Methods for Interpreting a Self-Organized Map in Data Analysis**. 6th European Symposium on Artificial Neural Networks. Helsinki: Helsinki University of Technology. 1998. p. 1-6.
- KIVIULOTO, K. **Topology preservation in self-organizing maps**. IEEE Neural Networks Council (ed) Proceedings of international conference on neural networks (ICNN'96). New Jersey: [s.n.]. 1996. p. 294 - 299.
- KOHONEN, T. Self organizing formation of topologically correct feature maps. **RecMap: rectangular map approximations**, v. 41, n. 1, p. 59-69, 1982.
- KOHONEN, T. The self-Organizing Map (SOM). **Quinquenal Report 1994 - 1998**, Helsinki, 1999. 12 - 17.
- KOHONEN, T. **Self-Organizing Maps**. 3. ed. Berlim: Springer, 2001. 502 p.
- LIMA, J. F. Risco de Urbanização da febre amarela no Brasil. **Cadernos de Saúde Pública**, Rio de Janeiro, 1, n. 3, 1985. 377-384.
- LONGLEY, P. A. et al. **Sistema e Ciência da Informação Geográfica**. 3. ed. Porto alegre: Bookman, 2003. 540 p.
- MACKENZIE, J. S.; GUBLER, D. J.; PETERSEN, L. R. Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. **Nature Medicine Supplement**, v. 10, n. 12, p. 12, 2004.
- MARQUES, A. P. D. S. et al. **Análise exploratória de dados de área para índices de furto na mesoregião de presidente prudente-SP**. III Simpósio de Ciências Geodésicas e Tecnologias da Geoinformação. Recife: [s.n.]. 2010. p. 001-008.
- MARZOCHI, K. B. F. Dengue in Brazil - situation, transmission and control - a proposal for ecological control, Rio de Janeiro, 89, 1994. 133-141.
- MINISTÉRIO DA SAÚDE - BRASIL. **Experiências Nacionais em Monitoramento e Avaliação - Dengue**. Departamento de Vigilância Epidemiológica. Rio de Janeiro, p. 30. 2010.
- MITCHELL, A. **The ESRI Guide to GIS Analysis**. Redlands: ESRI press, v. 2, 2005.
- OCHI, L. S.; DIAS, C. R.; ALVES, S. S. F. **Clusterização em Mineração de Dados**. Universidade Federal Fluminense. Niterói, p. 46. 2004.
- ORGANIZAÇÃO PAN-AMERICANA DE SAÚDE. **Dengue and Dengue Haemorrhagic Fever in The Americas: guidelines for prevention and control**. Organização Pan-Americana de Saúde. Washinton. 1997.
- REITER, P. Climate Change and Mosquito-Borne Disease. **Environmental Health Perspectives**, v. 109, p. 21, 2001.

RIPSA-REDE INTERAGENCIAL DE INFORMAÇÕES PARA SAÚDE. **Indicadores básicos de saúde no Brasil: conceitos e Aplicações**. 2. ed. Brasília: Organização Pan-Americana de Saúde, 2008.

ROSENSHEIN, L.; SCOTT, L.; PRATT, M. **Exploratory Regression: A tool for modeling complex phenomena**. ESRI. Redlands, p. 1. 2011.

SCHULLER, J. P. S. **Computação Quântica e Biológica**. Universidade Federal do Rio Grande do Sul. Porto alegre, p. 50. 2000.

SKUPIN, A.; AGARWAL, P. **Self-organising Maps: Applications in Geographic Information Science**. Chichester: Wiley, 2008. 205 p.

TAUIL, P. L. Urbanização e ecologia do dengue. **Caderno de saúde pública**, Rio de Janeiro, v. 17, p. 4, 2001.

TORRES, E. M. Dengue. **Estudos Avançados**, São Paulo, v. 22, n. 64, p. 33-52, 2008.

ULTSCH, . et al. **Knowledge Extraction from Artificial Neural Networks and Applications**. Proc. Transputer Anwender Treffen/ World Transputer Congress TAT/WTC 93. Aachen: Springer. 1993. p. 1-16.

URIARTE, E. A.; MARTIN, F. D. Topology Preservation in SOM. **International Journal of Applied Mathematics and Computer Sciences**, v. 1, n. 1, p. 19-22, 2005.

VASCONCELOS, P. F. D. C. **Estudo de epidemias de dengue: uso e significado dos inquéritos soropidemiológicos transversais**. Universidade federal da Bahia. Salvador, p. 224. 1999.

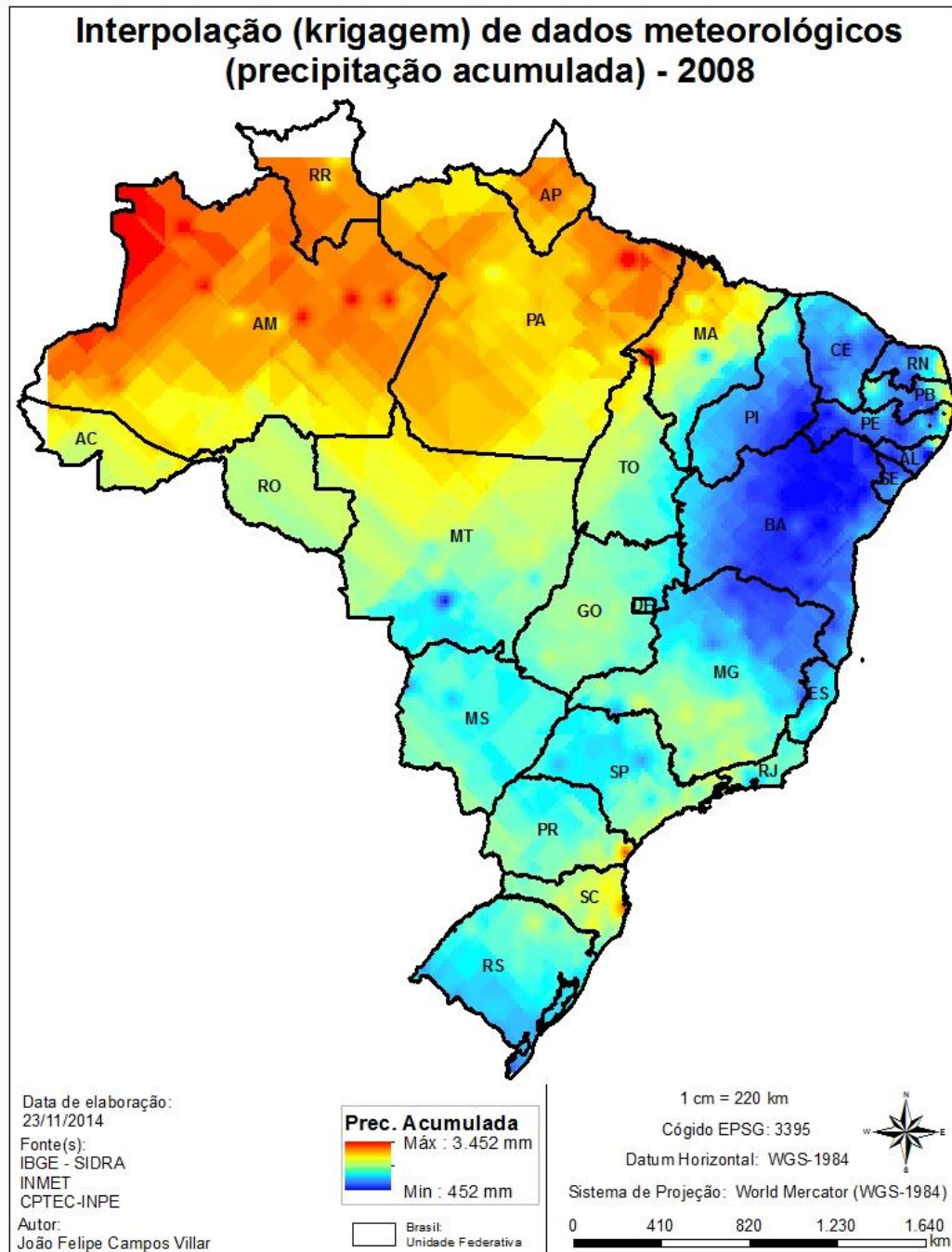
VESANTO, J. **Data mining techniques based on the self-organizing map**. Helsinki University of Technology. Helsinki, p. 63. 1997.

VESANTO, J. et al. **Som Toolbox for Matlab 5**. Helsinki University of Technology. Helsinki, p. 59. 2000.

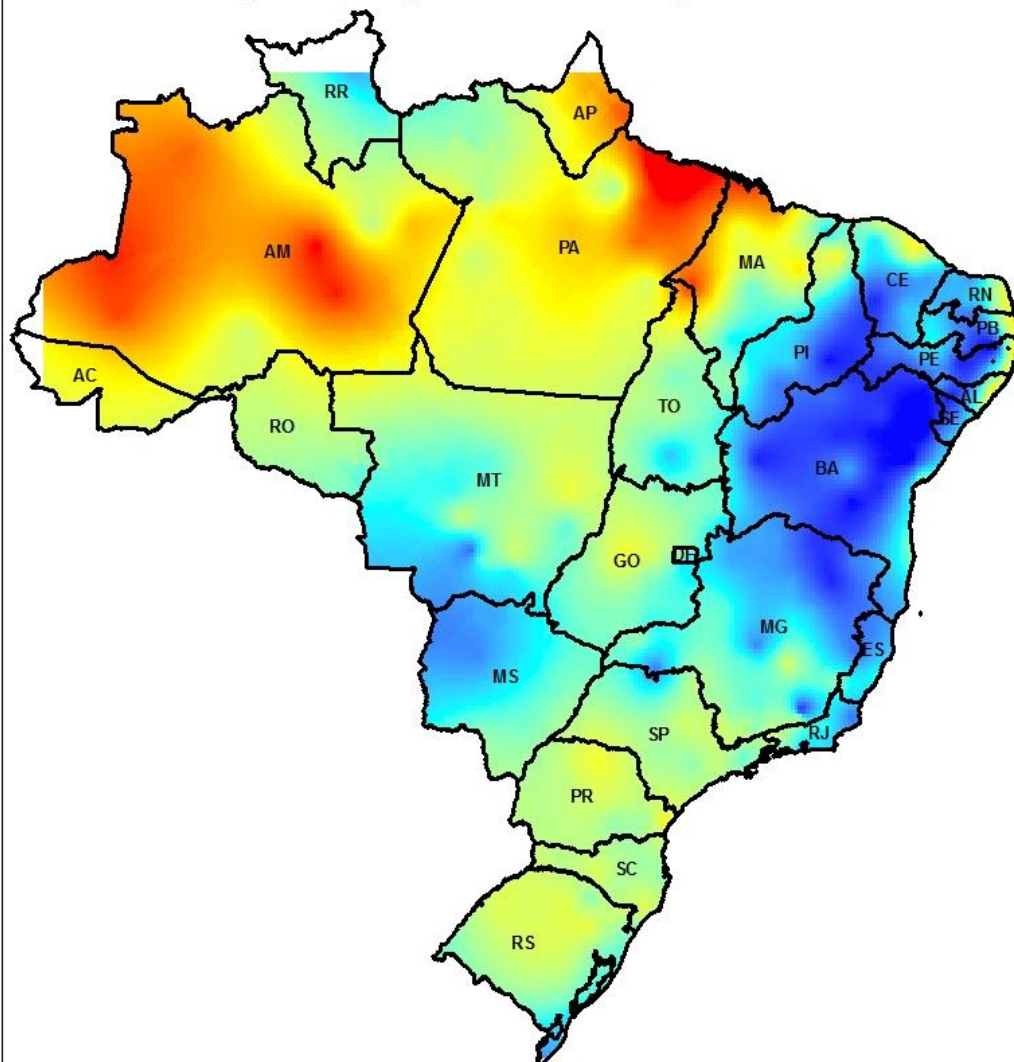
WERNECK, G. L. et al. The burden of Leishmania chagasi infection during an urban outbreak of visceral leishmaniasis in Brazil. **Acta Tropica**, v. 83, n. 1, p. 6, 2002.

## ANEXO A

Interpolação por krigagem de todos os dados meteorológicos de 2008 até 2012.



## Interpolação (krigagem) de dados meteorológicos (precipitação acumulada) - 2009



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar

**Prec. Acumulada**  
Máx : 3.845 mm  
Min : 497 mm

Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

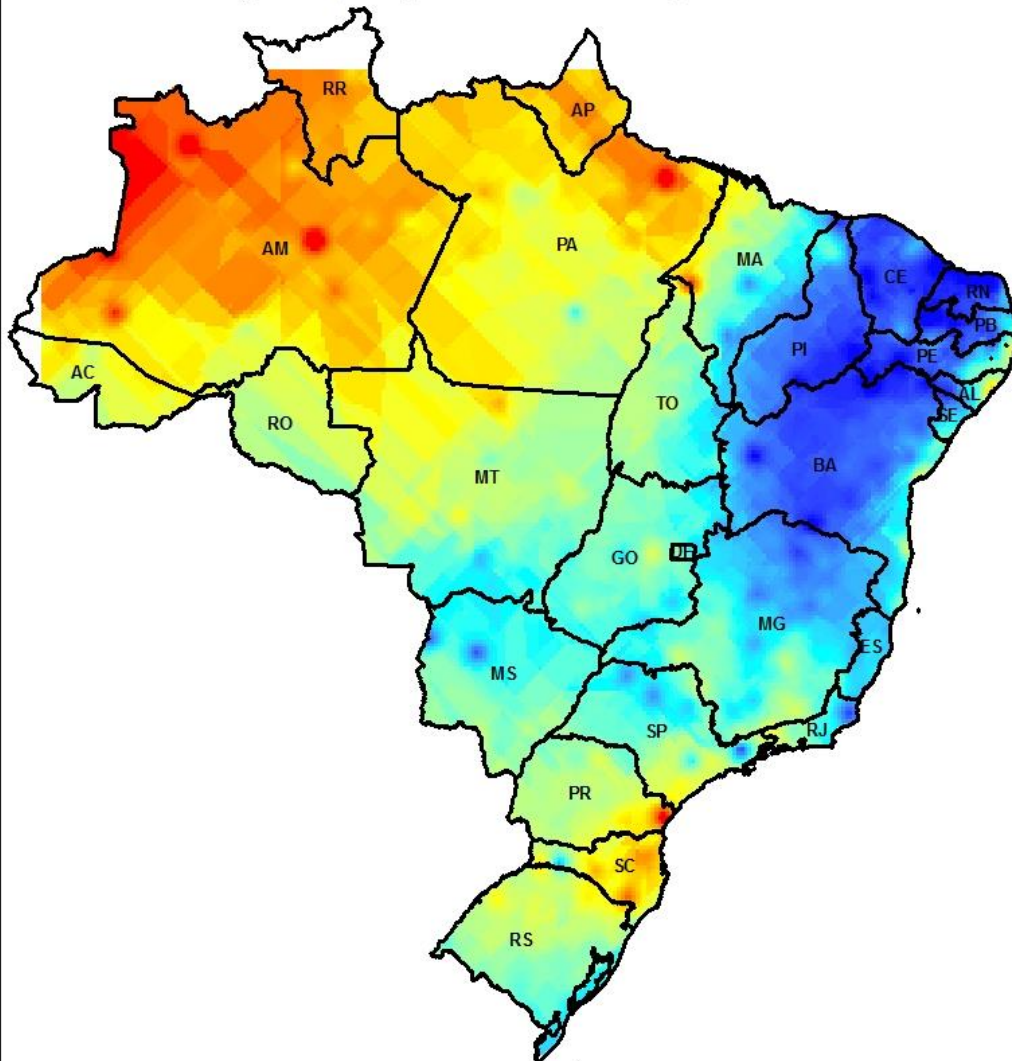
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)

0 410 820 1.230 1.640 km



## Interpolação (krigagem) de dados meteorológicos (precipitação acumulada) - 2010



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar

**Prec. Acumulada**  
Máx : 3.151 mm  
Min : 289 mm

Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

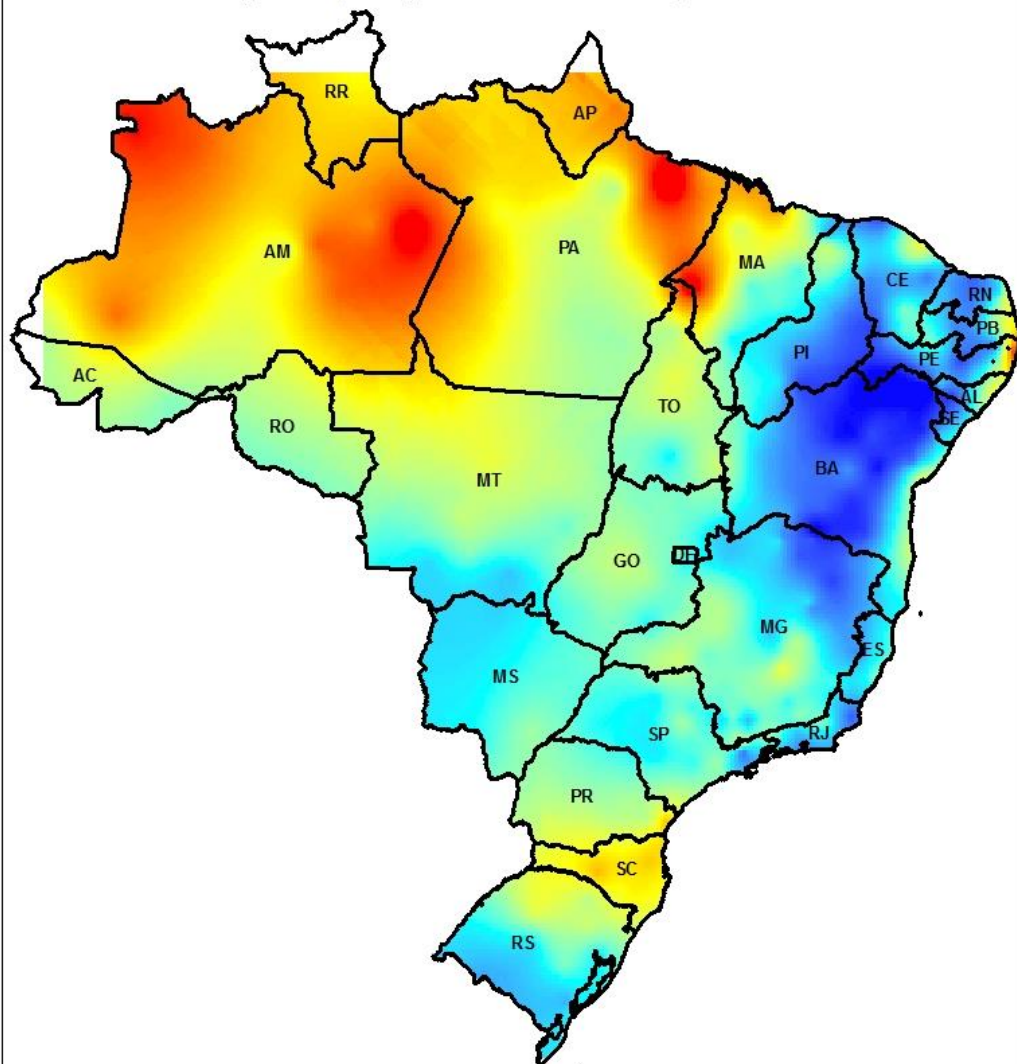
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)

0 410 820 1.230 1.640 km



## Interpolação (krigagem) de dados meteorológicos (precipitação acumulada) - 2011



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar

**Prec. Acumulada**  
Máx : 3.564 mm  
Min : 295 mm

Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

Datum Horizontal: WGS-1984

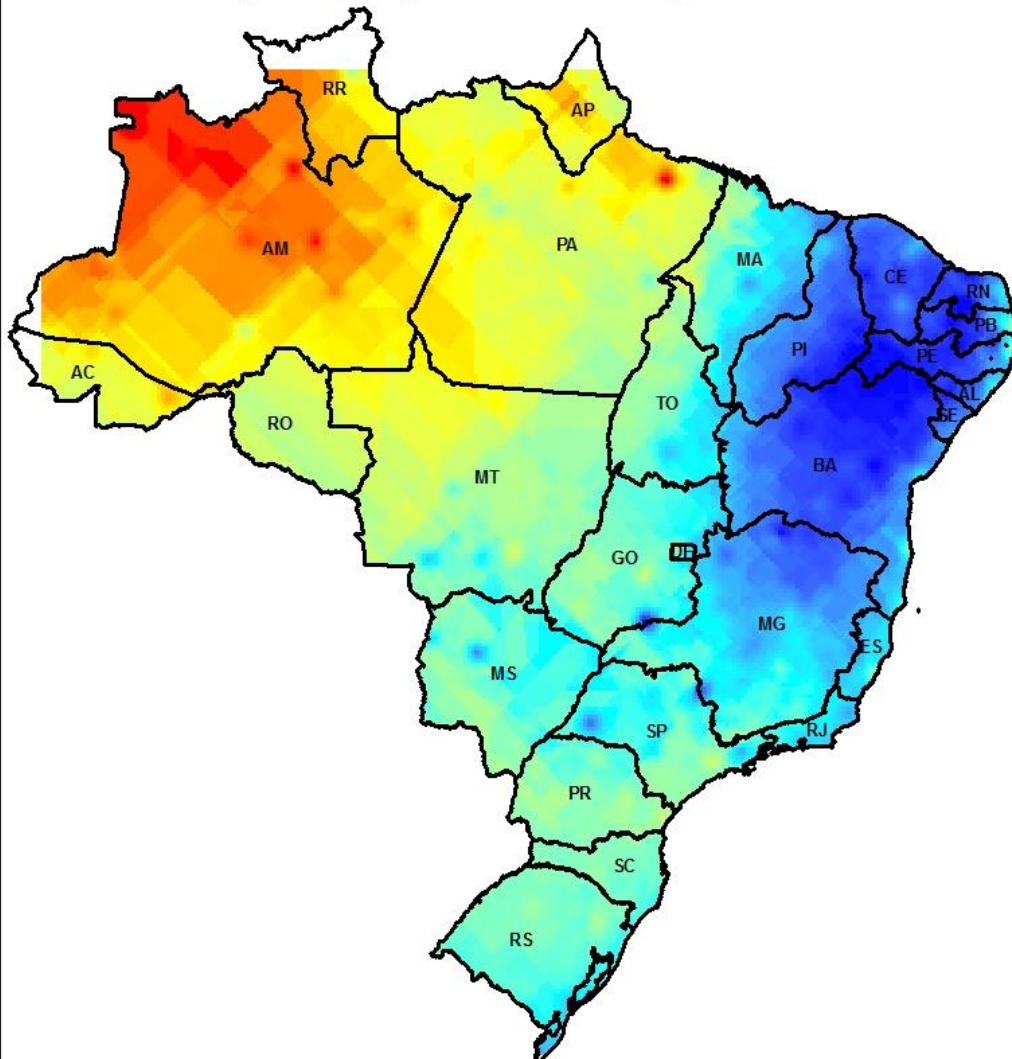
Sistema de Projeção: World Mercator (WGS-1984)

0 410 820 1.230 1.640 km





## Interpolação (krigagem) de dados meteorológicos (precipitação acumulada) - 2012



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar

**Prec. Acumulada**  
Máx : 3.819 mm  
Min : 154 mm

Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

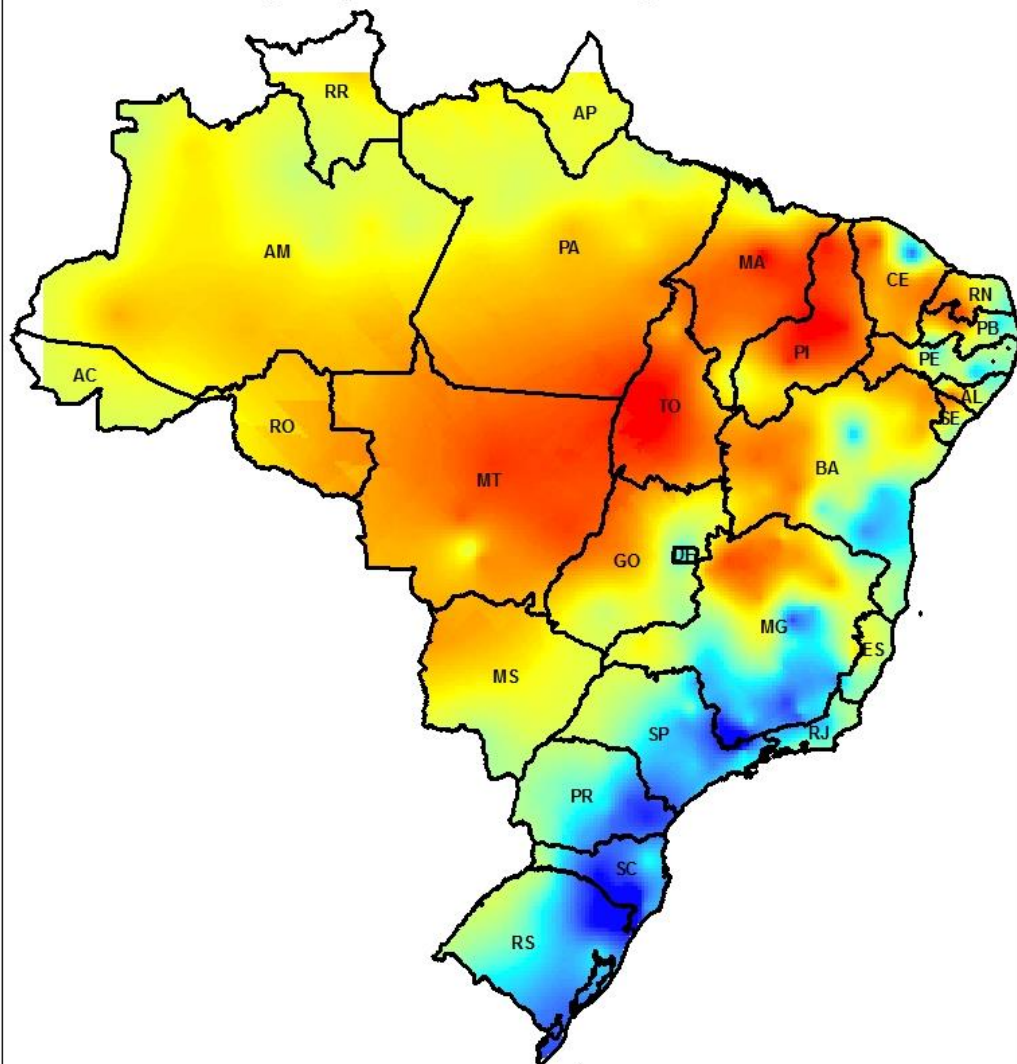
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)

0 410 820 1.230 1.640 km



## Interpolação (krigagem) de dados meteorológicos (temperatura máxima) - 2008



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar



Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

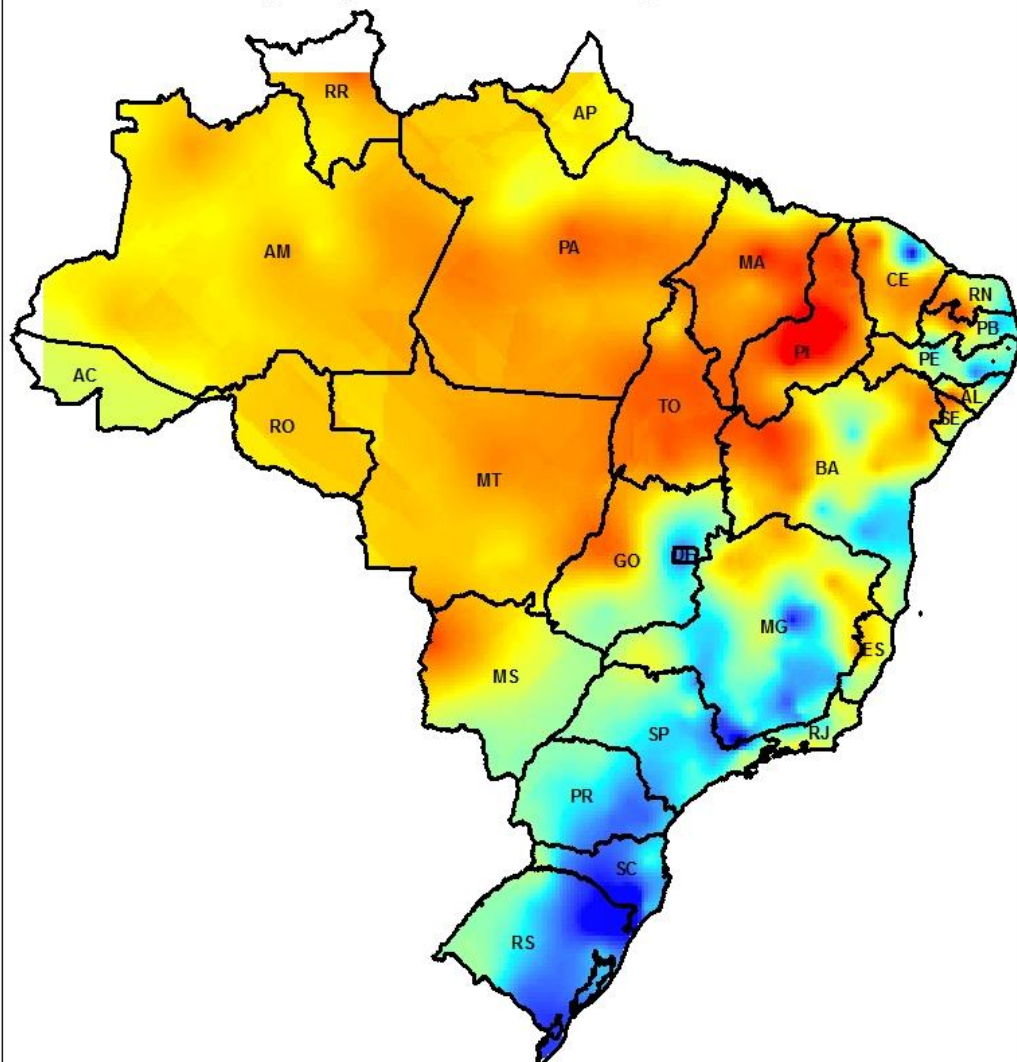
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)



0 410 820 1.230 1.640 km

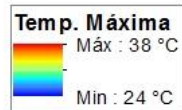
## Interpolação (krigagem) de dados meteorológicos (temperatura máxima) - 2009



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar



Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

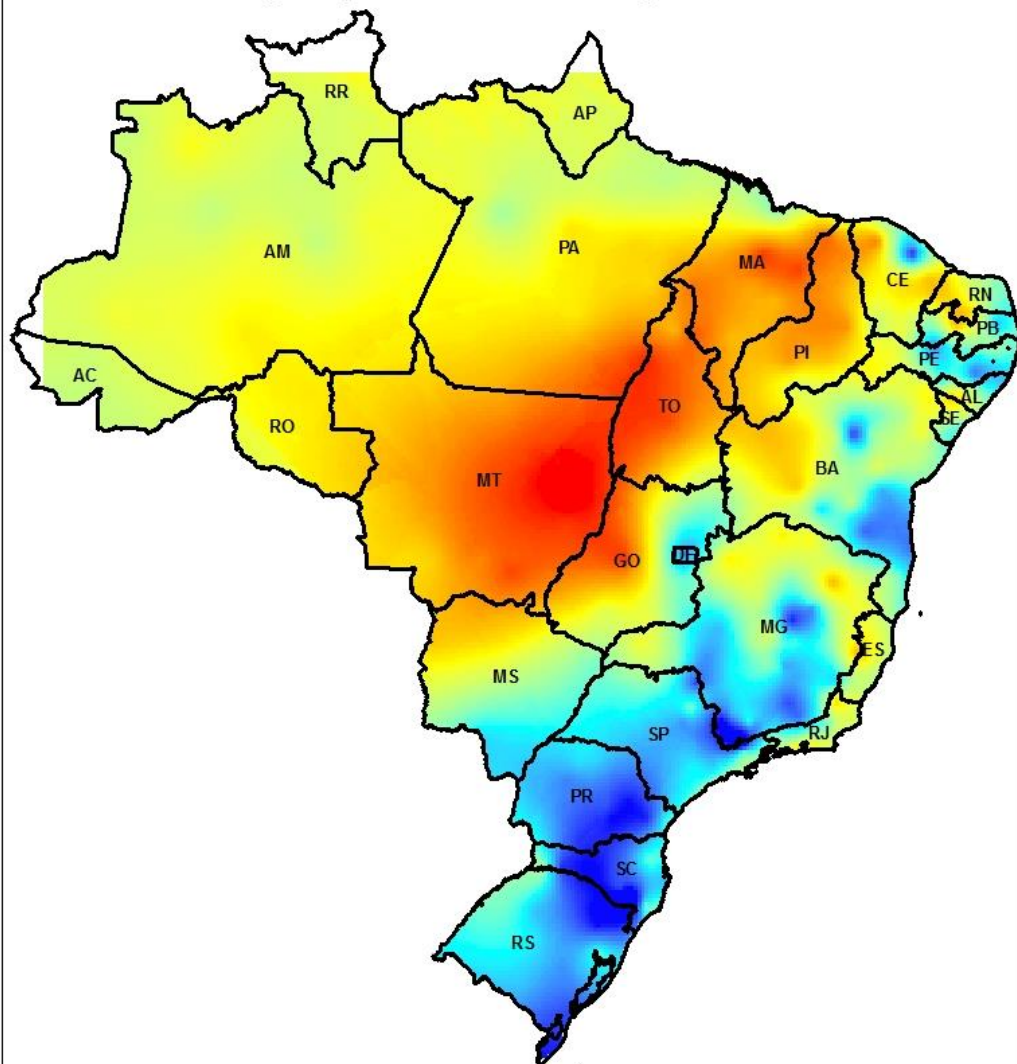
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)



0 410 820 1.230 1.640 km

## Interpolação (krigagem) de dados meteorológicos (temperatura máxima) - 2010



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar

**Temp. Máxima**



Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

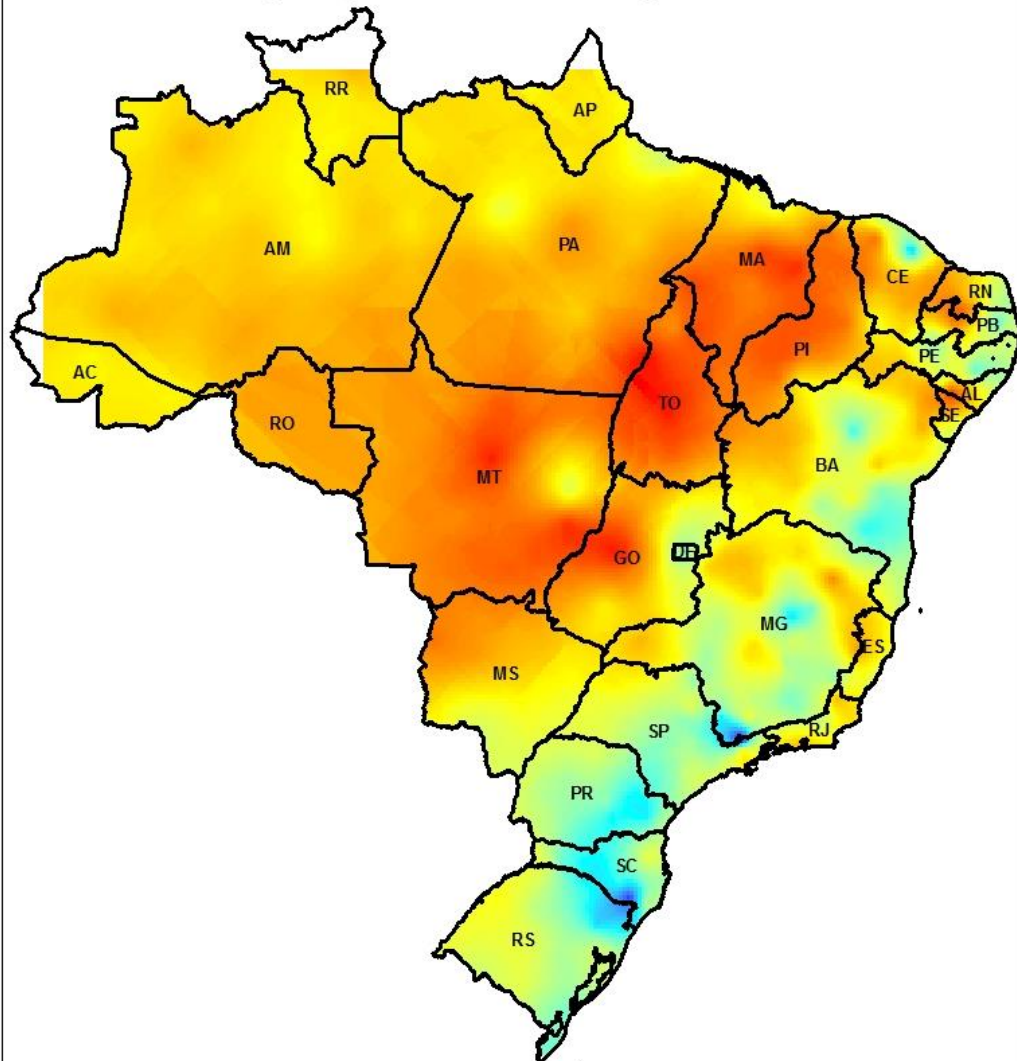
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)



0 410 820 1.230 1.640 km

## Interpolação (krigagem) de dados meteorológicos (temperatura máxima) - 2011



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar

**Temp. Máxima**  
Máx : 39 °C  
Min : 23 °C

Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

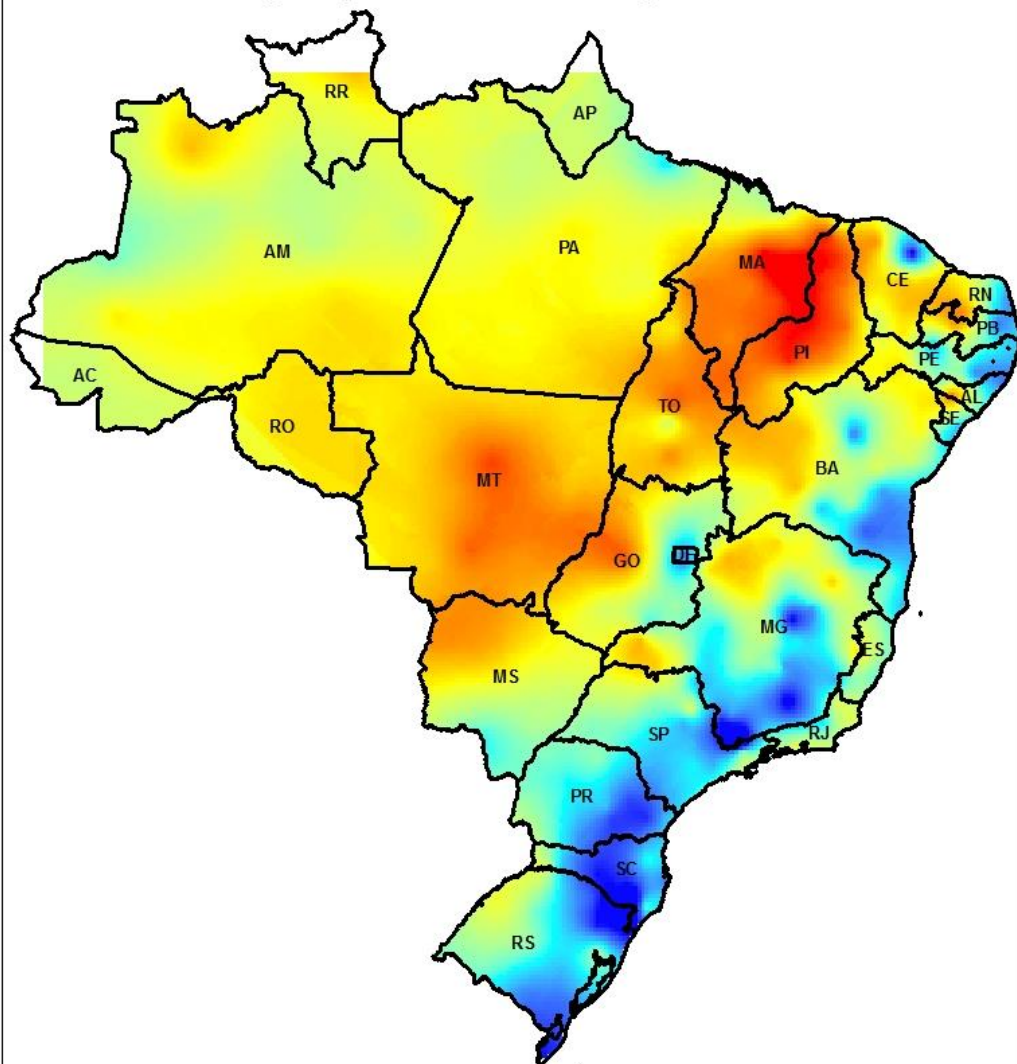
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)

0 410 820 1.230 1.640 km



## Interpolação (krigagem) de dados meteorológicos (temperatura máxima) - 2012



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar

**Temp. Máxima**  
Máx : 40 °C  
Min : 25 °C

Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

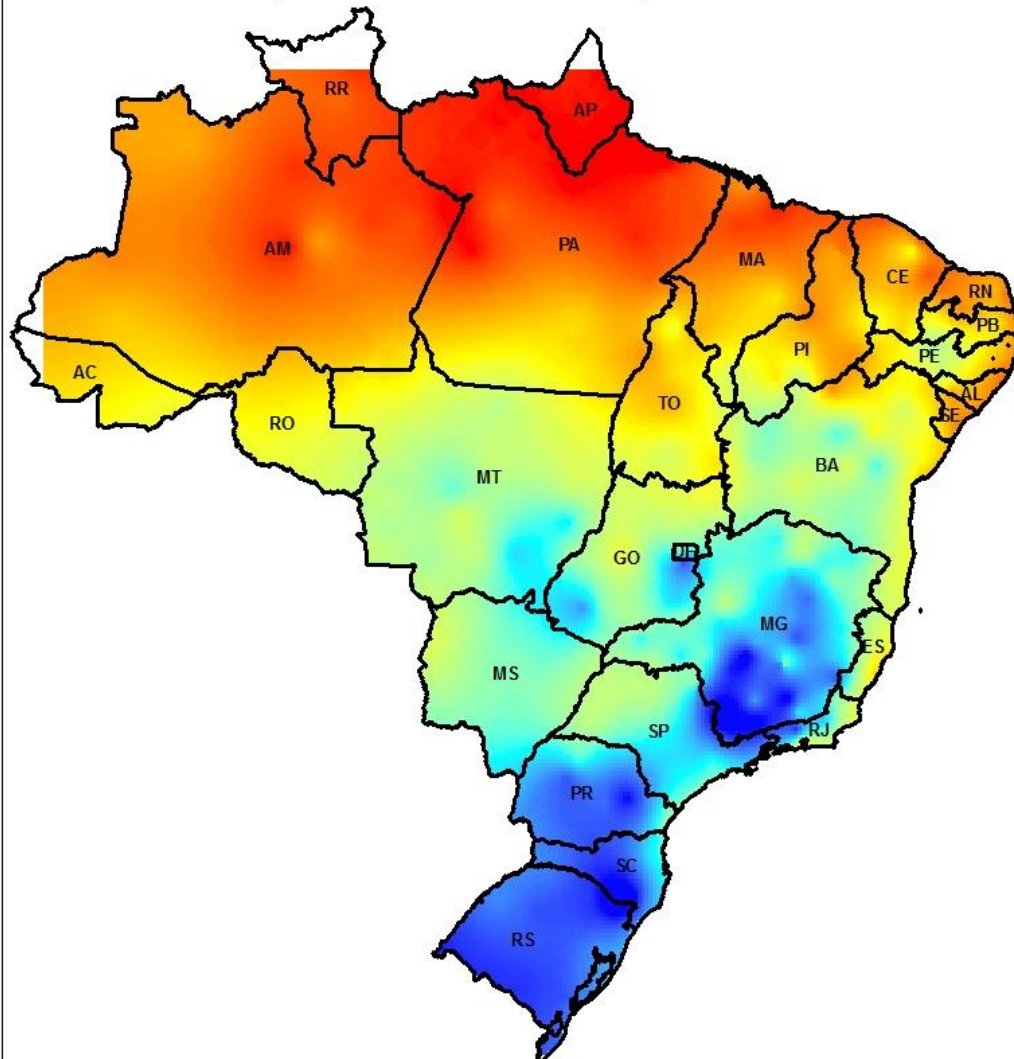
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)

0 410 820 1.230 1.640 km



## Interpolação (krigagem) de dados meteorológicos (temperatura mínima) - 2008



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar

**Temp. Mínima**  
Máx : 23 °C  
Min : 4 °C

Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

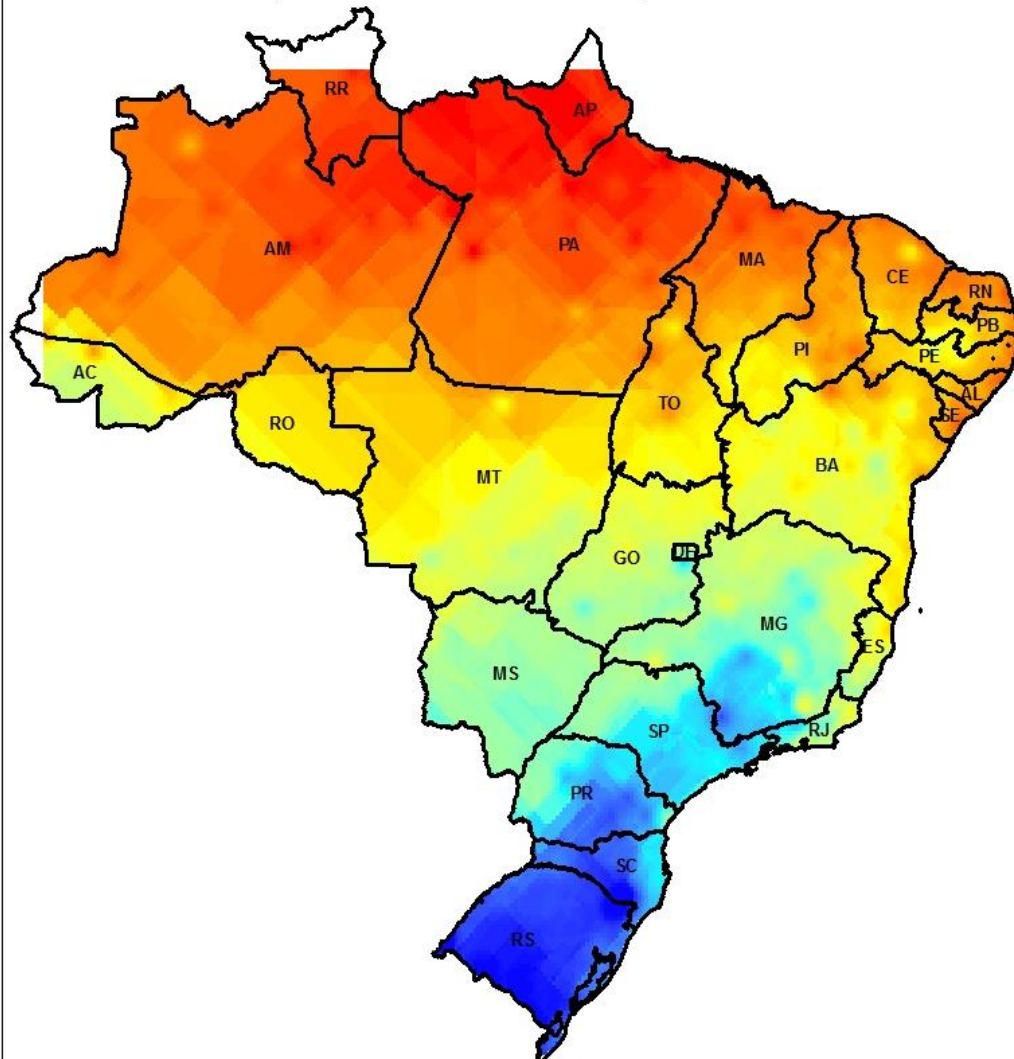
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)

0 410 820 1.230 1.640 km



## Interpolação (krigagem) de dados meteorológicos (temperatura mínima) - 2009



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar



Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

Datum Horizontal: WGS-1984

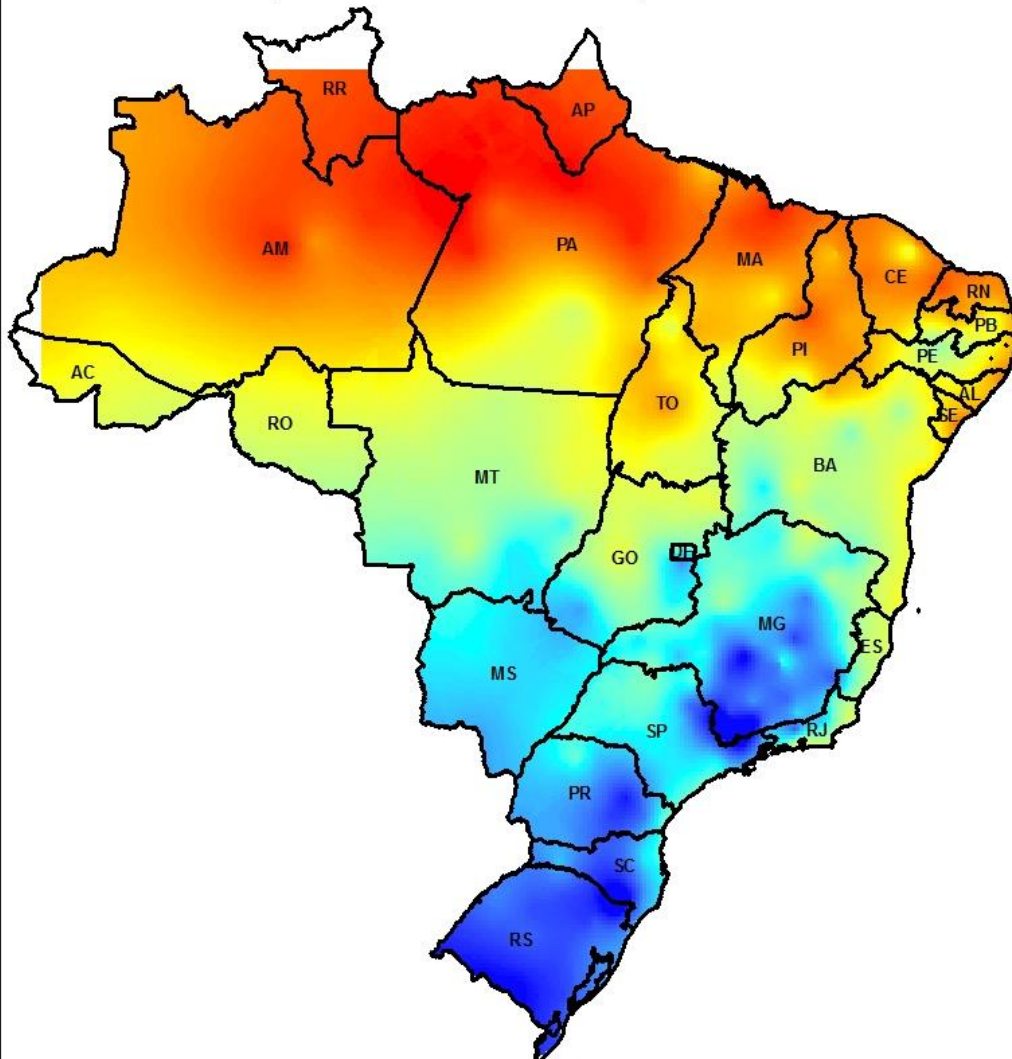
Sistema de Projeção: World Mercator (WGS-1984)



0 410 820 1.230 1.640 km



## Interpolação (krigagem) de dados meteorológicos (temperatura mínima) - 2010



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar

**Temp. Mínima**  
Máx : 24 °C  
Min : 5 °C

Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

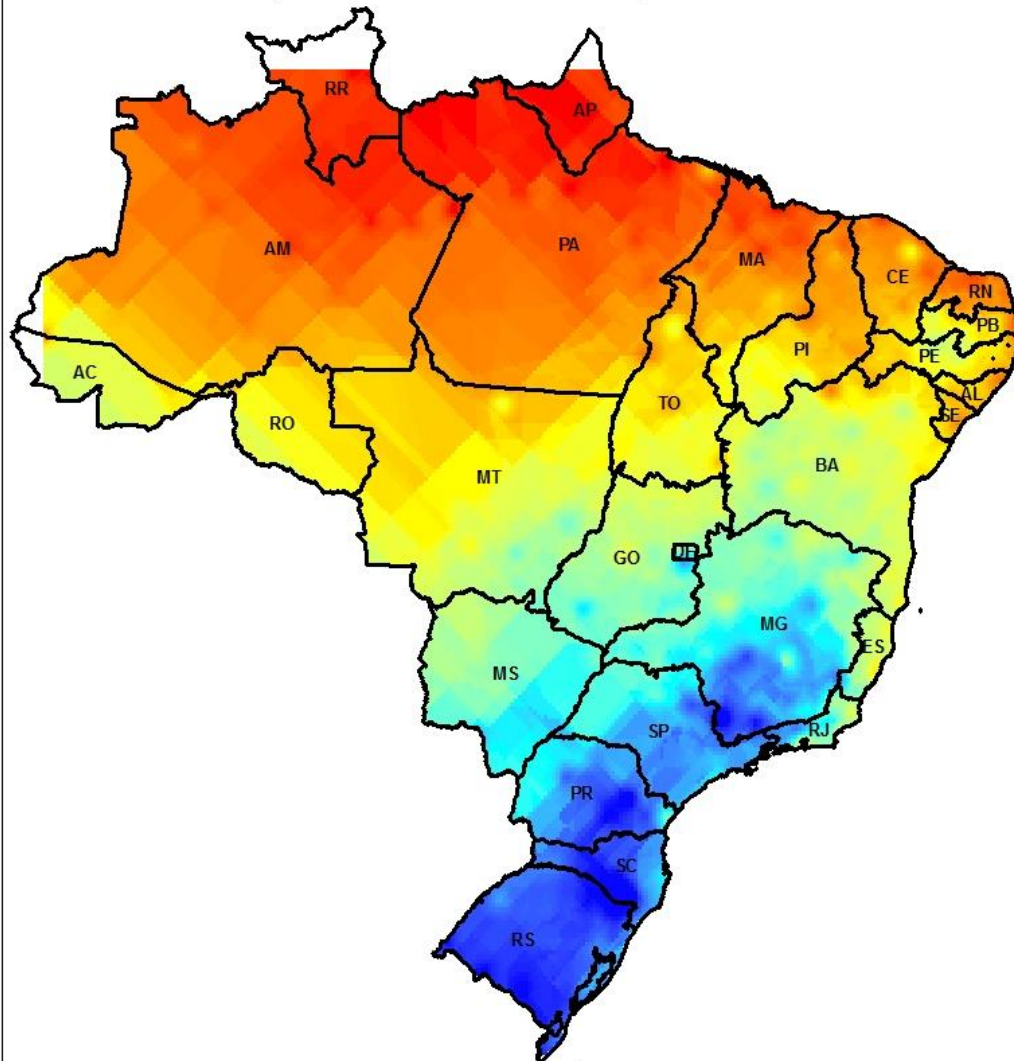
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)

0 410 820 1.230 1.640 km



## Interpolação (krigagem) de dados meteorológicos (temperatura mínima) - 2011



Data de elaboração:  
23/11/2014  
Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE  
Autor:  
João Felipe Campos Villar

**Temp. Mínima**  
Máx : 24 °C  
Min : 5 °C

Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

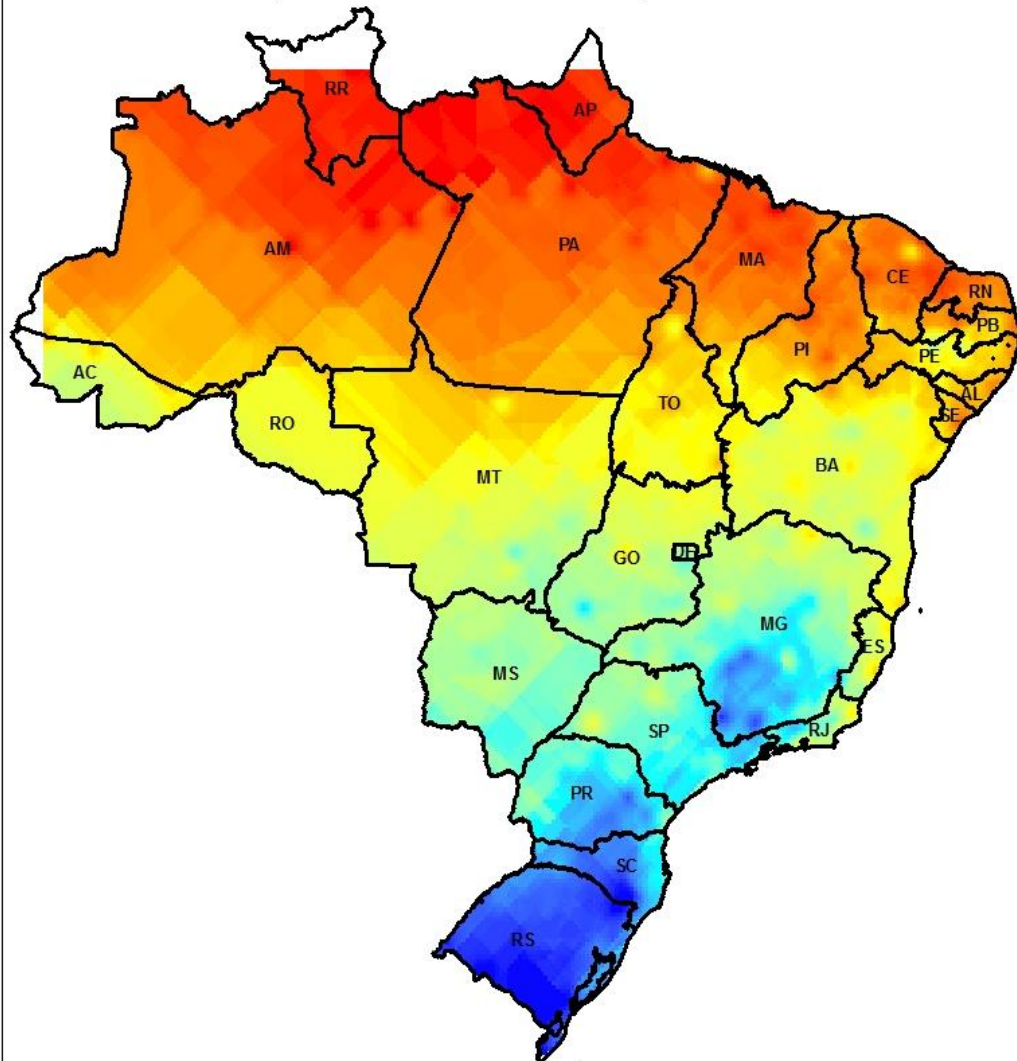
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)



0 410 820 1.230 1.640 km

## Interpolação (krigagem) de dados meteorológicos (temperatura mínima) - 2012



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar

**Temp. Mínima**  
Máx : 24 °C  
Min : 5 °C

Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

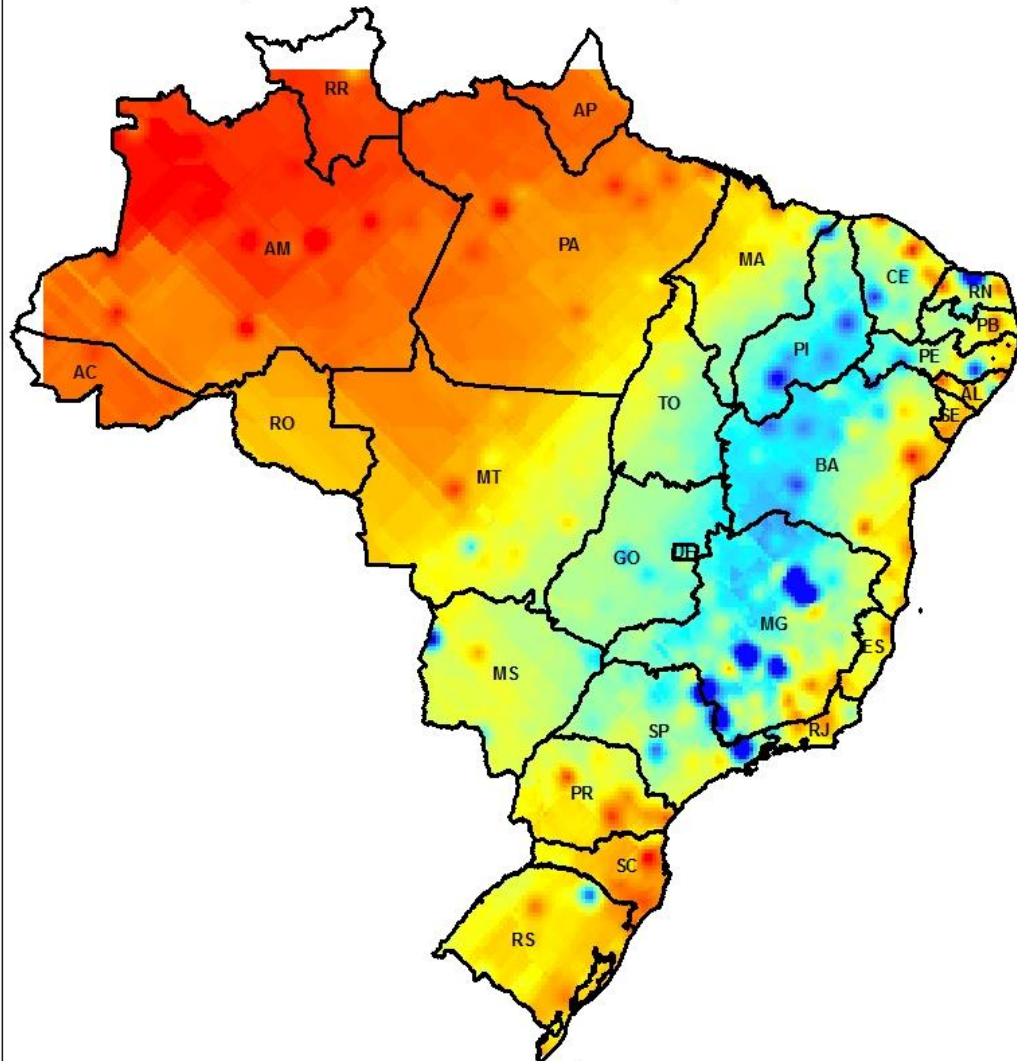
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)

0 410 820 1.230 1.640 km



## Interpolação (krigagem) de dados meteorológicos (umidade relativa média) - 2008



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar

**Umida. Rel. média**

Máx : 89 %  
Min : 22 %

Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

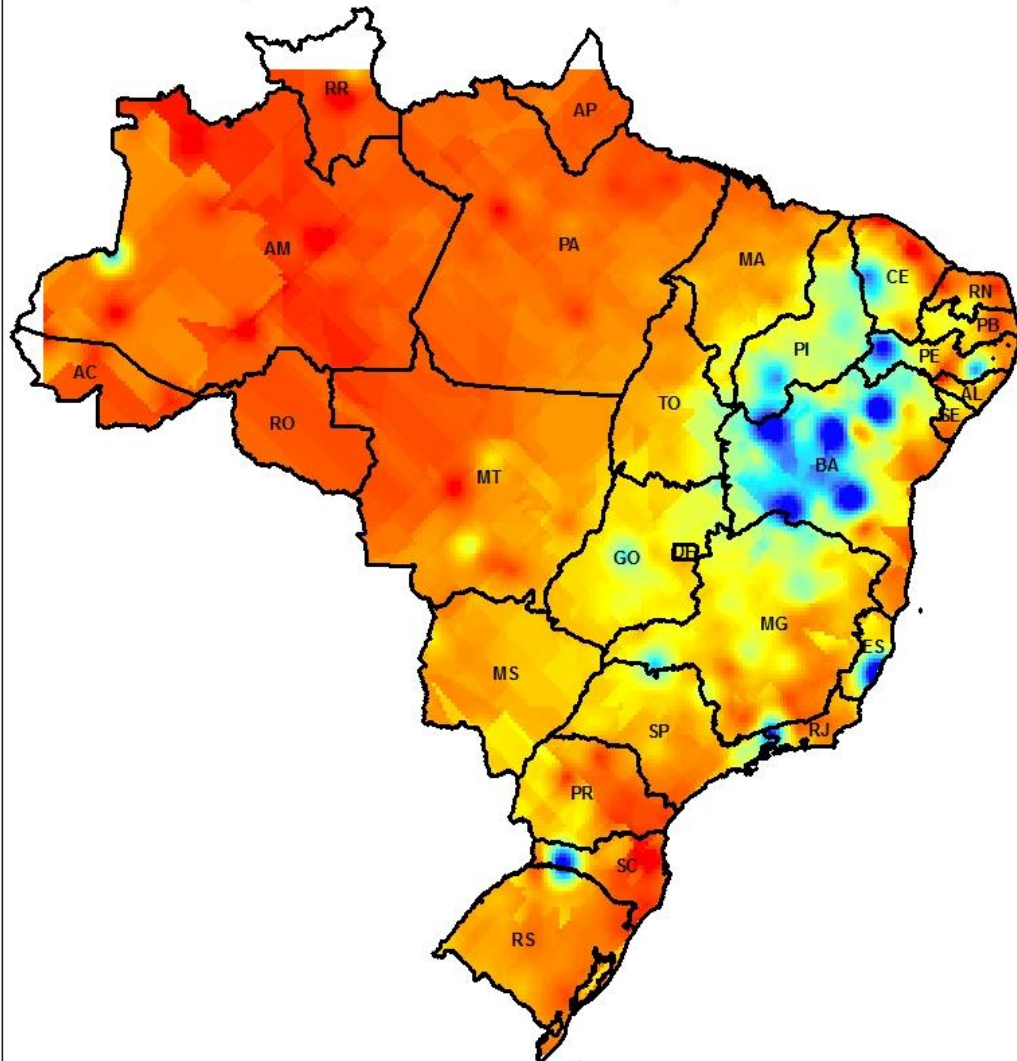
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)

0 410 820 1.230 1.640 km



## Interpolação (krigagem) de dados meteorológicos (umidade relativa média) - 2009



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar

### Umida. Rel. média



Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

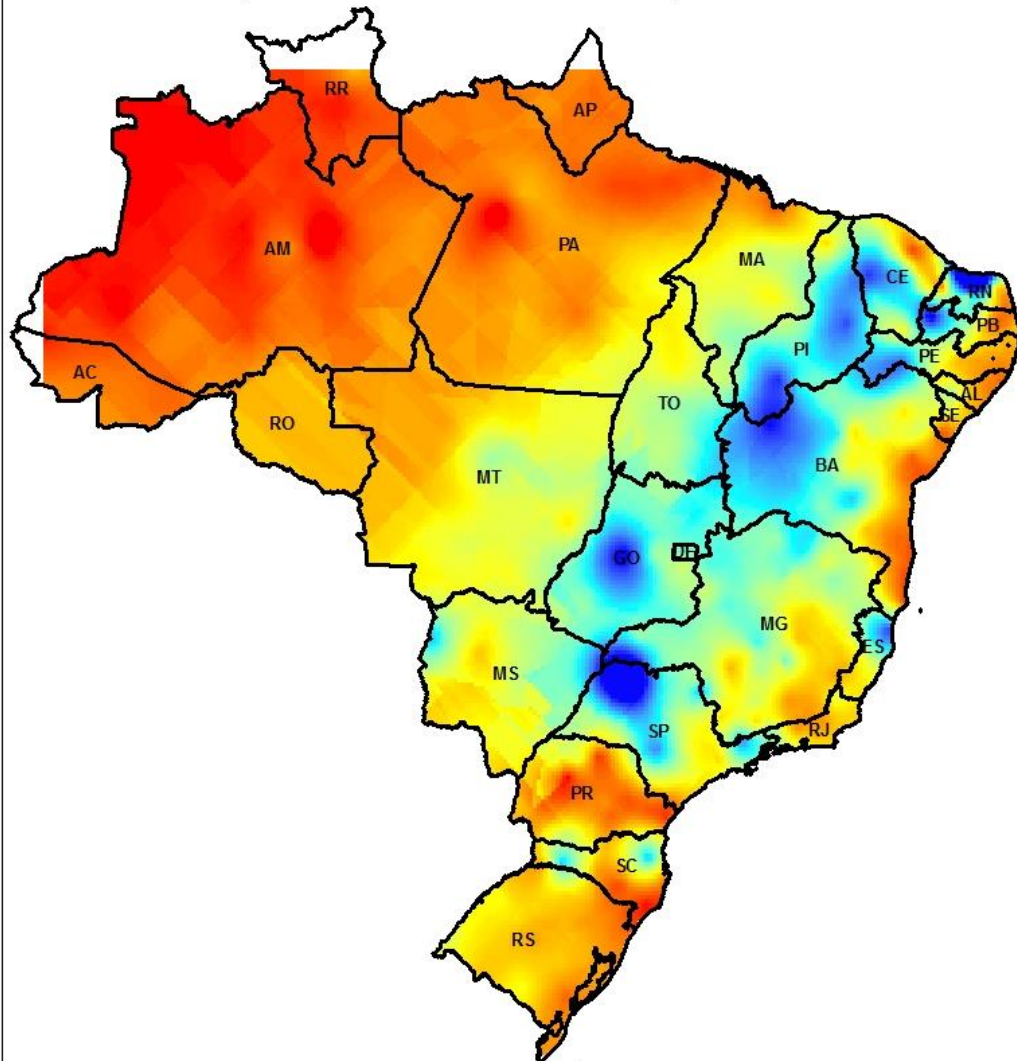
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)

0 410 820 1.230 1.640 km



## Interpolação (krigagem) de dados meteorológicos (umidade relativa média) - 2010



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar

### Umida. Rel. média

Máx : 88 %  
Min : 10 %

Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

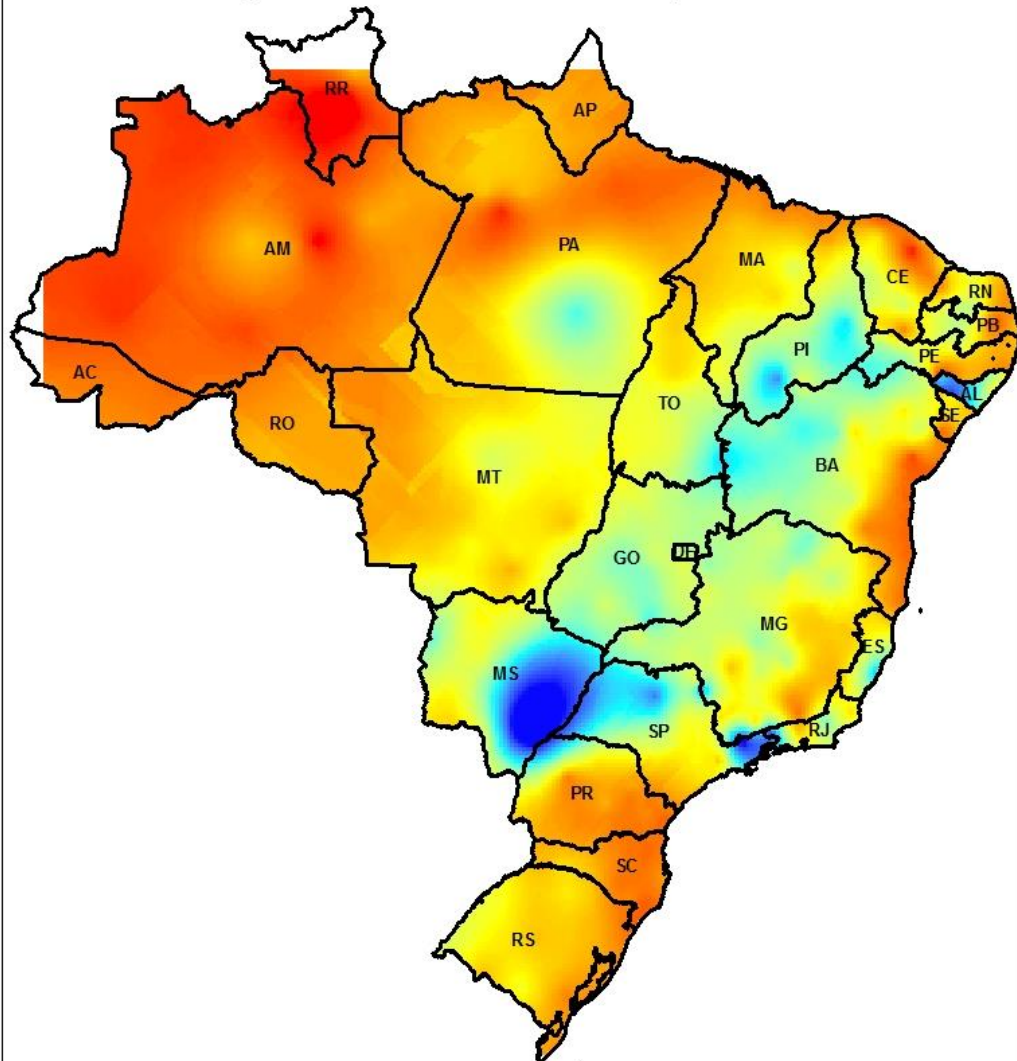
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)

0 410 820 1.230 1.640 km



## Interpolação (krigagem) de dados meteorológicos (umidade relativa média) - 2011



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar

### Umida. Rel. média

Máx : 92 %  
Min : 23 %

Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

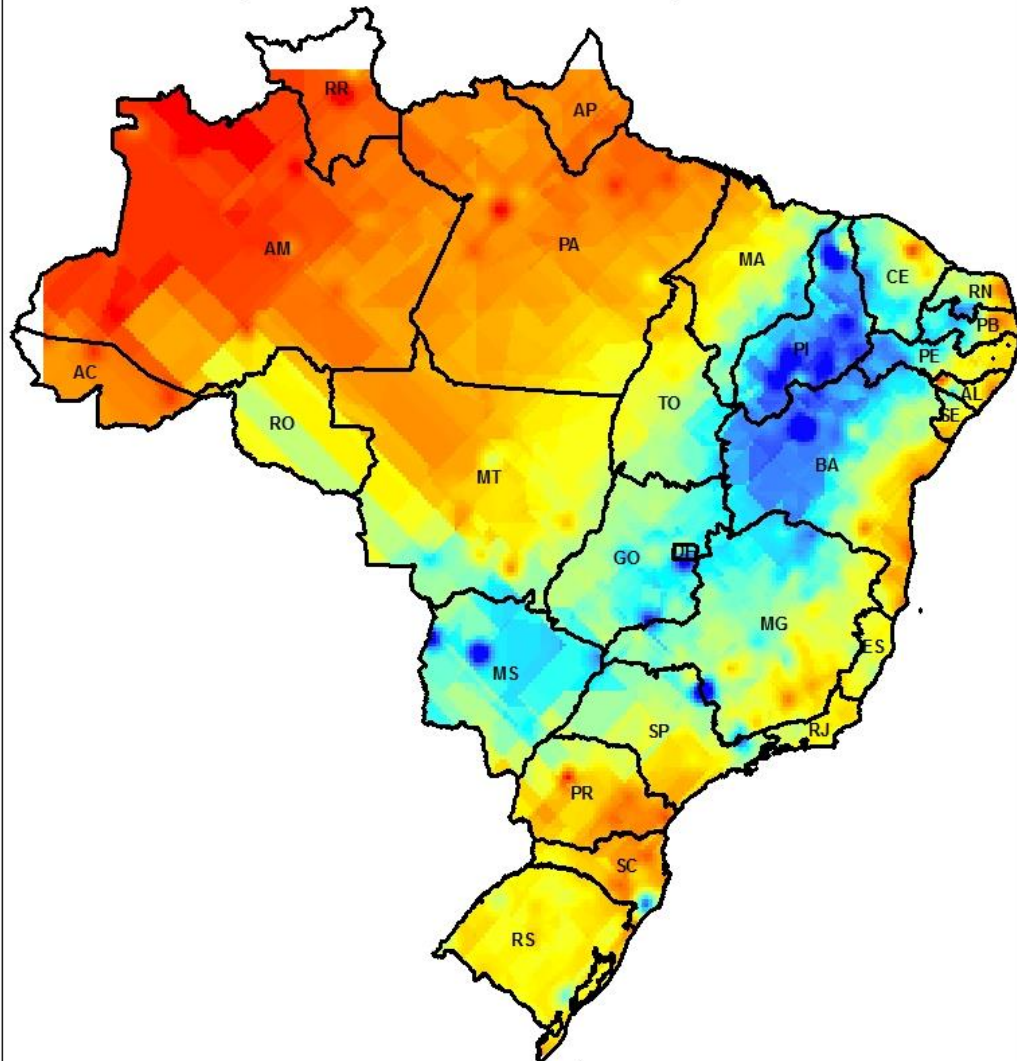
Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)

0 410 820 1.230 1.640 km



## Interpolação (krigagem) de dados meteorológicos (umidade relativa média) - 2012



Data de elaboração:  
23/11/2014

Fonte(s):  
IBGE - SIDRA  
INMET  
CPTEC-INPE

Autor:  
João Felipe Campos Villar

### Umida. Rel. média



Brasil:  
Unidade Federativa

1 cm = 220 km

Código EPSG: 3395

Datum Horizontal: WGS-1984

Sistema de Projeção: World Mercator (WGS-1984)

0 410 820 1.230 1.640 km





## ANEXO B

Sumário global e detalhado da regressão exploratória de todo o período de análise (2008 até 2012).

Sumário Global da Regressão Exploratória (TAXA_INCIDENCIA_DENGUE)				
Percentual de critérios de pesquisa que passaram na regressão				
Critério de pesquisa	Corte	Ensaio	# Passou	% Passou
Min R-Quadrado Ajustado	> 0,50	32646,00	0,00	0,00
Max Coeficiente do Valor-P	< 0,05	32646,00	945,00	2,89
Max Valor de VIF	< 7,50	32646,00	21497,00	65,85
Min Valor-P de Jarque-Bera	> 0,10	32646,00	0,00	0,00
Min Valor-p de auto correlação espacial	> 0,10	38,00	0,00	0,00

Sumário de significância das variáveis analisadas			
Variável	% Significância	% Negativo	% Positivo
IND_MOBILIDADE	100,00	0,00	100,00
LIRAA	100,00	0,00	100,00
PRECIPITACAO	100,00	100,00	0,00
TEMPERATURA_MINIMA	100,00	0,00	100,00
AMPLITUDE_TERMICA	81,56	44,42	55,58
TEMPERATURA_MAXIMA	74,92	25,25	74,75
IDHM	67,53	51,22	48,78
UMIDADE_RELATIVA	53,31	99,53	0,47
DENSIDADE_DEMOGRAFICA	35,30	2,38	97,62
GINI	27,82	16,44	83,56
AREA	22,26	70,43	29,57
THEIL	20,51	22,25	77,75
REMU_MEDIA	17,93	67,77	32,23
PIB_PERCAP	11,19	83,11	16,89
TAXA_ANALFABETISMO	8,96	9,49	90,51

## ANEXO C

Correlação de clusters identificados no SOM com as variáveis analisadas (ambientais, socioeconômicas e geográficas).

**TABELA 1**

Correlação em 2008

CLUSTER	VARIÁVEIS AMBIENTAIS						VARIÁVEIS SOCIOECONOMICAS						VARIÁVEIS GEOGRÁFICAS		
	LIRA	PRECIPITAÇÃO	TEMPERATURA MÁXIMA	TEMPERATURA MÍNIMA	UMIDADE RELATIVA	AMPLITUDE TÉRMICA	ÍNDICE DE GINI	THEIL	IDHM	PIB PER CAPTA	TAXA DE ANALFABETISMO	REMUNERAÇÃO MÉDIA	ÁREA	DENSIDADE DEMOGRÁFICA	INDICADOR DE MOBILIDADE
0	0,1092	1.655	32	17	71	16	0,52	0,49	0,64	12	1.781	1,98	2.994	261	72
1	0,0042	1.099	34	19	69	16	0,53	0,53	0,59	5	941	1,56	915	57	69
2	0,0129	1.672	33	14	70	20	0,48	0,41	0,71	13	943	2,11	1.689	69	92
3	0,0006	1.623	30	10	71	20	0,46	0,38	0,70	13	622	2,04	391	55	4
Média Geral	0,0338	1.510	32	15	70	18	0,49	0,46	0,66	11	1.081	1,92	1.479	115	55

**TABELA 2**

Correlação em 2009

CLUSTER	VARIÁVEIS AMBIENTAIS						VARIÁVEIS SOCIOECONOMICAS						VARIÁVEIS GEOGRÁFICAS		
	LIRA	PRECIPITAÇÃO	TEMPERATURA MÁXIMA	TEMPERATURA MÍNIMA	UMIDADE RELATIVA	AMPLITUDE TÉRMICA	ÍNDICE DE GINI	THEIL	IDHM	PIB PER CAPTA	TAXA DE ANALFABETISMO	REMUNERAÇÃO MÉDIA	ÁREA	DENSIDADE DEMOGRÁFICA	INDICADOR DE MOBILIDADE
0	0,2271	1.643	33	17	70	17	0,51	0,48	0,65	15	2.198	2,08	3.872	344	63
1	0,0049	1.486	34	20	71	15	0,53	0,54	0,59	5	1.173	1,60	1.053	57	61
2	0,0013	1.868	31	11	74	20	0,46	0,39	0,71	14	688	2,02	716	56	39
Média Geral	0,0496	1.704	32	15	72	17	0,49	0,46	0,66	11	1.153	1,91	1.479	116	51

**TABELA 3**

Correlação em 2010

CLUSTER	VARIÁVEIS AMBIENTAIS						VARIÁVEIS SOCIOECONOMICAS						VARIÁVEIS GEOGRÁFICAS		
	LIRA	PRECIPITAÇÃO	TEMPERATURA MÁXIMA	TEMPERATURA MÍNIMA	UMIDADE RELATIVA	AMPLITUDE TÉRMICA	ÍNDICE DE GINI	THEIL	IDHM	PIB PER CAPTA	TAXA DE ANALFABETISMO	REMUNERAÇÃO MÉDIA	ÁREA	DENSIDADE DEMOGRÁFICA	INDICADOR DE MOBILIDADE
0	0,0956	1.501	34	16	71	18	0,54	0,53	0,65	14	1.413	1,95	3.278	197	76
1	0,0000	1.778	30	10	76	21	0,45	0,37	0,71	17	470	2,08	416	59	4
2	0,0296	1.484	31	12	65	19	0,43	0,33	0,73	18	521	2,15	366	96	95
3	0,1363	1.386	32	11	70	21	0,47	0,40	0,69	14	864	1,89	599	94	73
4	0,0266	890	34	19	68	15	0,51	0,50	0,59	6	1.314	1,61	656	62	75
<b>Média Geral</b>	<b>0,0643</b>	<b>1.371</b>	<b>33</b>	<b>15</b>	<b>70</b>	<b>18</b>	<b>0,49</b>	<b>0,46</b>	<b>0,66</b>	<b>13</b>	<b>1.065</b>	<b>1,90</b>	<b>1.479</b>	<b>117</b>	<b>68</b>

**TABELA 4**

Correlação em 2011

CLUSTER	VARIÁVEIS AMBIENTAIS						VARIÁVEIS SOCIOECONOMICAS						VARIÁVEIS GEOGRÁFICAS		
	LIRA	PRECIPITAÇÃO	TEMPERATURA MÁXIMA	TEMPERATURA MÍNIMA	UMIDADE RELATIVA	AMPLITUDE TÉRMICA	ÍNDICE DE GINI	THEIL	IDHM	PIB PER CAPTA	TAXA DE ANALFABETISMO	REMUNERAÇÃO MÉDIA	ÁREA	DENSIDADE DEMOGRÁFICA	INDICADOR DE MOBILIDADE
0	0,2091	1585	33	16	72	17	0,51	0,48	0,65	9	816	1,91	2.184	130	78
1	0,1836	1338	33	18	73	15	0,52	0,51	0,61	6	1.027	1,76	1.040	101	80
2	0,1968	1657	33	14	68	19	0,49	0,45	0,69	13	685	2,13	2.321	158	93
3	0,0334	1726	31	11	73	20	0,47	0,40	0,70	13	540	2,07	657	97	27
<b>Média Geral</b>	<b>0,1494</b>	<b>1577</b>	<b>33</b>	<b>15</b>	<b>72</b>	<b>18</b>	<b>0,49</b>	<b>0,46</b>	<b>0,66</b>	<b>10</b>	<b>764</b>	<b>1,96</b>	<b>1.479</b>	<b>118</b>	<b>66</b>

**TABELA 5**

Correlação em 2012

CLUSTER	VARIÁVEIS AMBIENTAIS						VARIÁVEIS SOCIOECONOMICAS						VARIÁVEIS GEOGRÁFICAS		
	LIRA	PRECIPITAÇÃO	TEMPERATURA MÁXIMA	TEMPERATURA MÍNIMA	UMIDADE RELATIVA	AMPLITUDE TÉRMICA	ÍNDICE DE GINI	THEIL	IDHM	PIB PER CAPTA	TAXA DE ANALFABETISMO	REMUNERAÇÃO MÉDIA	ÁREA	DENSIDADE DEMOGRÁFICA	INDICADOR DE MOBILIDADE
0	0,2998	1.183	33	16	69	17	0,51	0,48	0,65	14	798,87	1,92	2.206	162	72
1	0,3902	788	34	18	67	16	0,52	0,51	0,61	8	940,72	1,76	1.095	95	72
2	0,3427	1.407	34	15	69	19	0,49	0,44	0,69	20	643,80	2,11	2.007	155	84
3	0,0437	1.366	32	11	71	21	0,47	0,40	0,70	18	489,26	2,00	787	78	21
<b>Total Geral</b>	<b>0,2557</b>	<b>1.179</b>	<b>33</b>	<b>15</b>	<b>69</b>	<b>18</b>	<b>0,49</b>	<b>0,46</b>	<b>0,66</b>	<b>15</b>	<b>715,82</b>	<b>1,94</b>	<b>1.479</b>	<b>120</b>	<b>60</b>

# C& SIG