

UNIVERSIDADE FEDERAL DO AMAZONAS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE APOIO À PESQUISA
PROGRAMA INSTITUCIONAL DE INICIAÇÃO CIENTÍFICA

**MONITORAMENTO DE CLUSTERS NO ESPAÇO - TEMPO:
MÉTODOS, ESTUDOS COMPARATIVOS E APLICAÇÕES.**

Bolsita: Renatha dos Anjos Frazão, FAPEAM.

MANAUS
2014

UNIVERSIDADE FEDERAL DO AMAZONAS
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
DEPARTAMENTO DE APOIO À PESQUISA
PROGRAMA INSTITUCIONAL DE INICIAÇÃO CIENTÍFICA

RELATÓRIO FINAL

PIB-E/0022/2013

**MONITORAMENTO DE CLUSTERS NO ESPAÇO - TEMPO:
MÉTODOS, ESTUDOS COMPARATIVOS E APLICAÇÕES.**

Bolsita: Renatha dos Anjos Frazão, FAPEAM.
Orientador: Prof. Dr. Max Sousa de Lima.

MANAUS
2014

Todos os direitos deste relatório são reservados à Universidade Federal do Amazonas, ao Núcleo de Estudo e Pesquisa em Ciência da Informação e aos seus autores. Parte deste Relatório só poderá ser reproduzido para fins acadêmicos ou científicos.

Esta pesquisa, financiada pela Fundação de Amparo à Pesquisa do Estado do Amazonas - FAPEAM, através do Programa Institucional de Bolsas de Iniciação Científica da Universidade Federal do Amazonas, foi desenvolvida pelo Núcleo de Estudo e Pesquisa em Ciência da Informação e se caracteriza como sub projeto do projeto de pesquisa Bibliotecas Digitais.

Resumo

Neste trabalho foi realizado um estudo aplicado e comparativo entre métodos para detecção de clusters emergentes os quais são importantes, na área de saúde pública, para detectar aglomerados de casos de uma determinada doença, indicando o surgimento de um período epidêmico ou surto. Nossa metodologia consistiu na modelagem, monitoramento e detecção de aglomerado de novos casos de Hanseníase confirmados no Estado do Amazonas no período 2008-2012. Os dados foram obtidos no Sistema de Informações de Agravos de Notificação do DATASUS e e-siga-AM. Para análise espaço-temporal dos dados, utilizamos os métodos Cusum Adaptativo, Scan Prospectivo e Scan Score. Estes métodos avaliaram sequencialmente todos os possíveis candidatos a aglomerados no espaço-tempo controlando adequadamente a taxa de falsos alarmes de 5%. Os resultados obtidos foram similares para os Métodos Cusum Adaptativo e Scan Prospectivo o que reforçou a boa habilidade dos métodos em detectar cluster emergentes. Do ponto de vista da saúde pública, análises mais detalhada nos mostraram que a região sul do Estado do Amazonas passou recentemente por vários períodos epidêmicos, que talvez possam ser reflexos da vulnerabilidade social da região. Nossos experimento computacionais sugeriram que em estudos epidemiológicos usando dados históricos o método Scan Prospectivo deve ser recomendado. No entanto, em problemas de vigilância epidemiológica, o Cusum Adaptativo é preferível.

Palavras chaves: Cluster Emergente, Cusum Adaptativo, Scan Prospectivo, Scan Score, Vigilância epidemiológica.

Sumário

| | | |
|----------|---|-------|
| 1 | INTRODUÇÃO | p. 1 |
| 2 | REVISÃO BIBLIOGRÁFICA | p. 2 |
| 3 | MÉTODOS UTILIZADOS | p. 4 |
| 3.1 | Métodos Para detecção de Clusters no Espaço-tempo | p. 4 |
| 4 | RESULTADOS & DISCUSSÕES | p. 5 |
| 4.1 | Base de dados e Modelo | p. 5 |
| 4.2 | Aplicação dos Métodos Cusum Adaptativo, Scan Prospectivo e Scan Score no problema de Vigilância Epidemiológica | p. 7 |
| 4.2.0.1 | Results | p. 7 |
| 4.3 | Estudo comparativo entre os Métodos Cusum Adaptativo, Scan Prospe- ctivo e Scan Score | p. 8 |
| 5 | CONCLUSÕES | p. 10 |
| 6 | REFERÊNCIA BIBLIOGRÁFICA | p. 11 |
| 7 | CRONOGRAMA DE ATIVIDADES | p. 13 |

1 INTRODUÇÃO

Métodos estatísticos aplicados á análise de dados, obtidos periodicamente pelos sistemas de monitoramento e/ou vigilância em saúde pública são importantes para detectar *cluster*(em português: aglomerados) de eventos, os quais podem indicar uma rápida mudança no padrão dos dados observados. Geralmente, um *cluster* é uma inesperada aglomeração de eventos no espaço, tempo ou no espaço-tempo. Quando o período de monitoramento é pré-especificado e os eventos são avaliados de acordo com sua distribuição espacial, um *cluster* é representado por um subconjunto de localizações espaciais da região em estudo onde a taxa de ocorrência de tais eventos é elevada de forma significativa. Se o espaço é ignorado, representamos o *cluster* pelo grupo (ou sequência) de observações que causam uma mudança no padrão temporal dos dados. No entanto, se o espaço e o tempo são monitorados simultaneamente, um *cluster* espaço-temporal é identificado por um par (localizações espaciais, intervalo de tempo) para o qual ocorreu a mudança. Por exemplo, se os eventos são relacionados à saúde, um *cluster* ou aglomerado espaço-temporal de casos de doença é representado por um conjunto de áreas (bairros, municípios, etc.) e um intervalo de tempo (semana, mês, etc.) que representa o período de duração do aglomerado. Neste caso, este aglomerado pode ser atribuído a fatores demográficos, ambientais ou a padrões sazonais.

Seja na escala espacial ou temporal, o problema de detecção e monitoramento de um *cluster* pode contribuir para o esclarecimento dos mecanismos responsáveis pela geração dos eventos e uma rápida detecção seria benéfica tanto para os indivíduos quanto para a sociedade, por exemplo, no sentido de reduzir despesas com medicamentos ou evitar que a doença se espalhe. Para estudar estes tipos de problemas, neste projeto de pesquisa, foi realizado um estudo comparativo entre os métodos: Scan prospectivo (*Kulldorff*, 2001), *Scan-Score*(*Tango*, 2011) e o GMAX-ARL (*Lima e Duczmal*, 2014) denominado neste trabalho "Cusum Adaptativo". Os métodos avaliados também foram aplicados no problema de vigilância epidemiológica em um conjunto de dados referente aos novos casos de Hanseníase observados no estado do Amazonas no período 2008-2012.

2 REVISÃO BIBLIOGRÁFICA

O problema de monitoramento e detecção de *clusters* tem sido o foco de uma variedade de sistemas de vigilância: Sistema de vigilância em crimes (*Gorr and Harries, 2003*) para detectar *clusters* emergentes de eventos relacionados a algum tipo de crime como violentos, à mão armada, etc.; Sistema de vigilância em doenças para monitorar dados de saúde pública como, por exemplo, visitas em hospitais (*Sabhnani et al., 2005*) ou *cluster* de casos de doenças como dengue, malária, hanseníase (*Kulldorff, 2001*); Sistemas de monitoramento ambiental para detectar altos níveis de poluição no ar e água (*Ailamaki et al., 2003*) ou detectar *cluster* emergentes de queimadas em florestas (*Tuia et al., 2008*).

Este problema pode ser formulado da seguinte forma: considere que a área de estudo é representada por uma mapa \mathcal{S} particionado em L -áreas A_l contíguas identificadas por um ponto s_l . Vamos assumir a partir de agora que s_l é o centróide do polígono que delinea e identifica a área A_l . Considere sobre \mathcal{S} o monitoramento (ou vigilância) estatístico(a) de um processo estocástico $\mathbf{X} = \{X_t(s_l), t = 1, 2, \dots \text{ e } l = 1, 2, \dots L\}$. A cada tempo discreto $t \geq 1$ observamos um vetor L -variado $\mathbf{X} = (X_t(s_1), X_t(s_2), \dots, X_t(s_L))'$, em que $X_t(s_l)$ representa algum valor observado na l -ésima área. Sejam \mathcal{Z} uma coleção de conjuntos conexos de localizações espaciais Z em \mathcal{S} , $\mathcal{D}_n = \{X_1, \dots, X_n\}$ o conjunto de dados acumulados até o instante n . O objetivo do monitoramento é verificar se existe algum *cluster* que começou em um tempo desconhecido $k \leq n$ e que está presente até o estágio atual n . O *cluster* no espaço-tempo é expresso em função de uma mudança no padrão da distribuição do sub-processo $X_t = \{X_t(s_l) : s_l \in Z, t \geq k\}$ de modo que $C_n(Z) = \{s_l \in \mathcal{S} : s_l \in Z\} \times [k, n]$ representa o *cluster* no espaço-tempo também chamado de *cluster* emergente. Nesse trabalho assumimos que se Z é um círculo de raio r centrado em um ponto qualquer s_0 do mapa. Então $C_n(Z)$ é um cilindro de base circular centrada em s_0 com raio r e altura $n - k$. As áreas A_l pertencentes a este cilindro serão representadas pelo conjunto $Z = \{s_l \in \mathcal{S} : d(s_l, s_0) \leq r\}$. Em que d é a distância euclidiana entre s_0 e o centróide s_l de A_l .

A detecção da mudança ou do *cluster* emergente é obtida através da avaliação dos

dados em \mathcal{D}_n sobre a classe $\mathcal{C} = \{C_n(Z) : Z \in \mathcal{Z}\}$ de todos os candidatos a clusters. O momento da mudança é o primeiro ponto no tempo para o qual existem evidências suficientes de que uma mudança tenha ocorrido. Matematicamente, o monitoramento é feito usando uma função de alarme $\Psi_n = \Psi(\mathcal{D}_n)$ e um limite de controle B , tais que o tempo de alarme, τ , de uma possível mudança é dada por,

$$\tau(B) = \min\{n : \Psi_n > B\}.$$

Se Ψ_n excede B , então há uma indicação de uma possível mudança no processo e um alarme é soado. Se não há mudanças no processo, todos os alarmes são falsos. O valor de B é geralmente especificado de forma a minimizar a ocorrência de alarmes falsos. Geralmente \mathcal{Z} é representado por alguma forma geométrica regular, como círculos, elipses ou por formas irregulares tal que a cardinalidade de \mathcal{Z} é excessivamente grande. Dessa forma, uma função de monitoramento Ψ_n deve ser ajustada para o problema de compações múltiplas, ter um pequena atraso na detecção, e ter uma boa capacidade para detectar corretamente pelo menos uma parte substância do *cluster* emergente.

Alguns métodos para modelagem e detecção de *cluster* no espaço-tempo têm sido propostos: *Kulldorff* (2001) usa um método baseado na estatística Scan espaço-tempo. A significância estatística deste método é realizada através do *valor_p* que não é uma medida adequada para este problema. *Rogerson* (2001) propôs o uso de uma estatística de Knox local (*Knox*, 1963) combinada com métodos de somas cumulativas (CUSUM) usando aproximações pela distribuição normal para monitoramento de *cluster* em processos pontuais. *Marshall* (2007) discute vários problemas ao se utilizar estas aproximações. Variantes da Estatística Scan espaço-temporal são apresentadas em *Neill* (2006,2009). *Sonesson* (2007) usa a estatística CUSUM para detecção de cluster no espaço-tempo em dados de área. Baseado em uma partição da região em estudo, *Porter* (2007) utiliza versões adaptativas dos métodos CUSUM e *Shiryayev-Roberts* para detecção de mudanças no espaço-tempo em processos de Poisson. *Assunção e Corrêa* (2009) monitoram a função de intensidade de um processo de Poisson heterogêneo no espaço-tempo através da estatística de *Shiryayev-Roberts*. *Tango* (2011) usa uma estatística scan combinada com um teste score para detecção de *cluster* no espaço-tempo em uma janela temporal de comprimento variável. *Lima e Duczmal*(2009,2011) propõem o uso de um Fator de Bayes Cumulativo para monitoramento de *cluster* em dados de área com taxa padrão conhecida. Combinando algoritmos para detecção de *cluster* no espaço com técnicas de monitoramento na escala temporal *Lima* (2014) apresenta várias abordagens adaptativas para detecção de *cluster* no espaço-tempo.

3 MÉTODOS UTILIZADOS

3.1 Métodos Para detecção de Clusters no Espaço-tempo

Os métodos para detecção de clusters emergentes utilizados neste projeto são baseados no seguinte teste de hipótese: $H_0 : \nu > n$ and $H_{k,z} : \nu = k, Z = z$ onde a hipótese H_0 é tal que o padrão de eventos é modelado por $\theta_0^t = \{\theta_{0,l}^t : s_l \in \mathcal{S}\}$ e sob $H_{k,z}$ o padrão de eventos muda depois de um tempo ν in Z para $\theta_1^t = \{\theta_{1,l}^t : s_l \in Z\} \cup \{\theta_{0,l}^t : s_l \notin Z\}$, com $\theta_{0,l}^t$ e $\theta_{1,l}^t$ considerado fixo no tempo t . Assumimos que o processo $\mathbf{X} = \{X_t(s_l), t = 1, 2, \dots \text{ e } l = 1, 2, \dots, L\}$ é Poisson com média $\theta_{0,l}^t = \mu_t(s_l)$ sob H_0 e sob $H_{k,z}, \theta_{1,l}^t = \delta_{t,z} \mu_t(s_l)$ para algum $\delta_{t,z} > 1$ estimado adaptativamente por Máxima Verossimilhança (Lima and Duczmal, 2014). Desta forma, pode-se mostrar (Lima, 2011) que as funções de monitoramento para os métodos CUSUM Adaptativo/ADC (Lima and Duczmal, 2014), Scan Prospectivo/SP (Kulldorff, 2001) e Scan-Score/SS (Tango, 2011) são dada, respectivamente, por:

$$\Psi_n^{ADC} = \max_{Z \in \mathcal{Z}} \left\{ \max_{1 \leq k \leq n} \prod_{t=k}^n (\delta_{t,z})^{\sum_{s_l \in Z} X_t(s_l)} \exp \left(- \sum_{s_l \in Z} [\mu_t(s_l) \delta_{t,z} - \mu_t(s_l)] \right) \right\}$$

$$\begin{aligned} \Psi_n^{SP} &= \sup_{C_n(Z) \in \mathcal{C}} \left(\frac{X(C_n(Z))}{\mu(C_n(Z))} \right)^{X(C_n(Z))} \left(\frac{X - X(C_n(Z))}{X - \mu(C_n(Z))} \right)^{X - X(C_n(Z))} \\ &\times I \left\{ \left(\frac{X(C_n(Z))}{\mu(C_n(Z))} \right) > \left(\frac{X - X(C_n(Z))}{X - \mu(C_n(Z))} \right) \right\}. \end{aligned}$$

Onde $I\{\}$ é a função indicadora, X é o número total de eventos observados no espaço-tempo, $X(C_n(Z)) = \sum_{s_l \in Z} \sum_{t=k}^n X_t(s_l)$ e $\mu(C_n(Z)) = \sum_{s_l \in Z} \sum_{t=k}^n \mu_t(s_l)$ denotam, respectivamente, o número de eventos e o valor esperado para o cilindro $C_n(Z)$ definido no capítulo 2 e,

$$\Psi_n^{SS} = \frac{\sum_{t=k}^n \sum_{s_l \in Z} (X_t(s_l) - \mu_t(s_l))}{\sqrt{\sum_{t \in I_w} \sum_{s_l \in \xi} \mu_t(s_l)}}$$

4 RESULTADOS & DISCUSSÕES

4.1 Base de dados e Modelo

Nesta seção será ilustrada a utilização dos métodos estatísticos proposto para detecção de *cluster* emergentes. Os dados utilizados são referentes aos novos casos de Hanseníase no Estado do Amazonas, no período de 2008-2012. Foram notificados 4.135 casos de hanseníase neste período. Os valores observados foram obtidos no Sistema de Informações de Agravos de Notificação (Sinan), disponível pelo DATASUS, e no e-siga/AM. A Hanseníase é uma doença infecto contagiosa causada pelo bacilo *Mycobacterium leprae* que afeta os nervos e a pele, provocando danos severos. Um dos estados do país em que nos últimos anos o coeficiente de detecção foi observado uma queda gradativa, é o Estado do Amazonas, em que a detecção média do período ficou em torno de 2,35 por 10.000 habitantes. Apesar desse indicador, o estado ainda é considerado de alta endemicidade, segundo o Ministério da Saúde, e por isso, o desenvolvimento de métodos estatísticos para vigilância epidemiológica, detectar e identificar se há um aumento acima do esperado no número de ocorrências de um evento no espaço e no tempo é importante, pois pode contribuir para o esclarecimento dos mecanismos responsáveis pela geração dos eventos e, uma rápida detecção seria benéfica tanto para os indivíduos quanto para a sociedade, por exemplo, no sentido de reduzir despesas com medicamentos ou evitar que a doença se espalhe. Nesta aplicação, assumimos que $X_t(s_l)$ representa o número de novos casos de Haseníase observados nos 62 municípios do Estado do Amazonas/2008-2012. O período de vigilância é mensal durante 5 anos e portanto o processo estocástico sob vigilância é representado por $X = \{X_t(s_l) : l = 1, 2, \dots, 62 \text{ e } t = 1, 2, \dots, 60\}$, e $X_t(s_l)$ assumido poisson com média $\mu_t(s_l) = n_t(s_l)\lambda$. Onde s_l é o identificador do l -ésimo município do Estado, $n_t(s_l)$ representa a população em risco no tempo t em s_l e λ representa a taxa de referência global para doença fornecida pelo Ministério da Saúde cujo valor é 0.0001 (um caso para cada 10000 habitantes). A evolução temporal da taxa de casos de Haseníase por 10000hab no Estado do Amazonas nos últimos quatro é apresentada na Figura 1(A), onde observa-se uma tendência decrescente na taxa global no Estado. No entanto, para

os municípios de Manaus e Itacoatiara, Figura 1(B), essa mesma taxa apresenta um padrão diferente do global. Esses padrões localizados, em mapas, diferentes do global são comuns em problemas de mapeamento de doenças, e um dos objetivos da utilização dos métodos de monitoramento no espaço-tempo é exatamente determinar se esses padrões localizados diferentes do global representam evidências do surgimento de pequeno surto da doença. Na Figura 1(C) é apresentada as coordenadas espaciais (latitude, longitude) que identificam a posição no mapa de cada município e na Figura 1(D) mostramos a distribuição espacial das taxas de casos por 10000hab. Nota-se nessa figura que a maioria dos municípios possui taxa de casos acima de 3 para cada 10000hab. Isso, de acordo com os parâmetros do Ministério da Saúde, reflete um período de alta epidemia na região. Mas, devemos chamar atenção ao fato que no Estado do Amazonas, existe uma grande heterogeneidade espacial da população o que pode levar a uma alta variação nas taxas e isso acarreta em interpretação errônea dos resultados. Este fato justifica o monitoramento dos casos desta doença através de técnicas mais robustas do que apenas visualização de taxas.

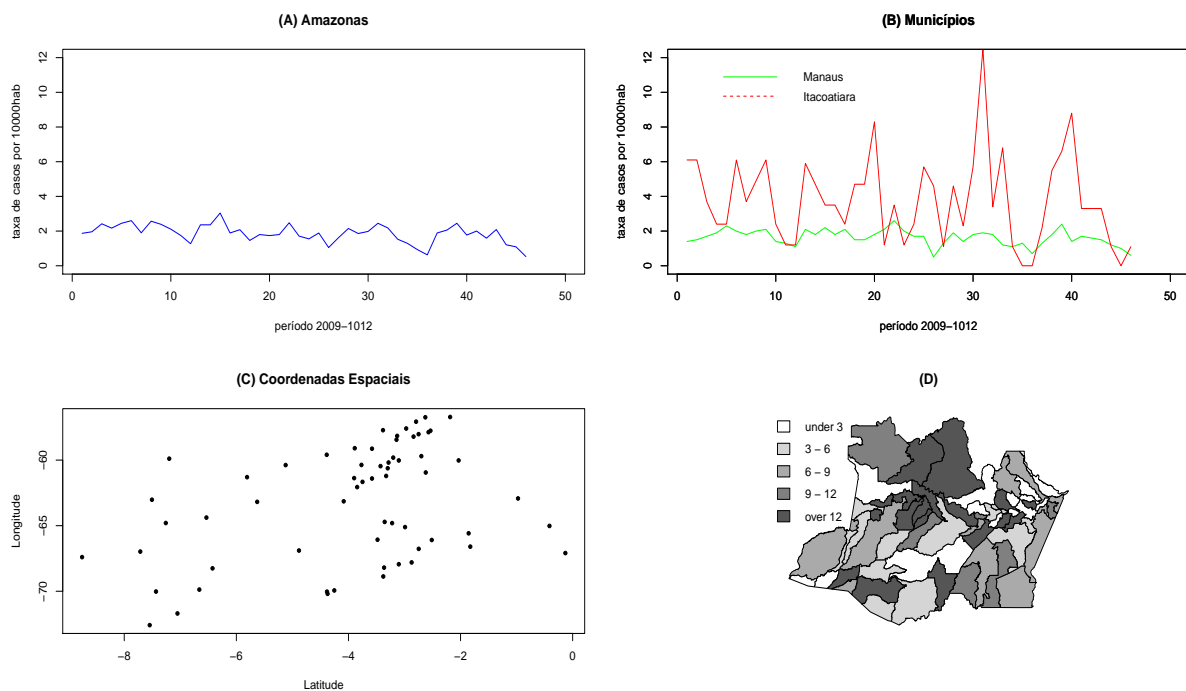


Figura 1: (A):Evolução Temporal dos casos de Haseníane no Amazonas; (B):Evolução Temporal dos casos de Haseníane nos municípios de Manaus e Itacoatiara; (C):Coordenadas Espaciais s_l que idedentificam cada Município do Estado; Distribuição espacial dos casos de Haseníase no Estado do Amazonas para cada 10000hab

Tabela 1: Cluster Emergente de casos de Hanseníase no Estado do Amazonas usando os métodos Cusum Adaptativo, Scan Prospectivo e Scan Score.

| Método | Localização Estimada(\hat{Z}) | intervalo de tempo | função de alarme |
|------------------|-----------------------------------|--------------------|------------------|
| Cusum Adaptativo | Guajará | [58,60] | 92.779 |
| Scan Prospectivo | Guajará | [58,60] | 5.007 |
| Scan Score | Humaitá | [58,60] | 4.309 |

4.2 Aplicação dos Métodos Cusum Adaptativo, Scan Prospectivo e Scan Score no problema de Vigilância Epidemiológica

4.2.0.1 Results

Os gráficos de monitoramento, usando as funções de alarme Ψ_n^{ADC} , Ψ_n^{SP} , Ψ_n^{SS} (veja, seção 3.1) para os casos de Hanseníase no Estado do Amazonas são apresentados na Figura 2 com limite $\log(B)=3$ o que implica em uma taxa de falso alarme igual 0.05. Ou seja, espera-se um falso alarme a cada 20 períodos de monitoramento. Essa figura também mostra a parte espacial do cluster emergente estimado em alguns pontos no tempo como $t = 33$ e $t = [58, 60]$. O tempo $t = 1$ representa janeiro de 2008 enquanto que $t = n = 60$ representa dezembro de 2012. Nota-se na figura que os métodos Cusum Adaptativo, Scan Prospectivo e Scan Score dispararam o alarme no período final de estudo $n = 60$. O método Cusum Adaptativo é mais sensível a mudança no padrão de casos indicando que existe no mapa um possível *cluster* emergente. Neste caso, os valores obtidos pelas funções de alarme foram 92.779 para o Cusum Adaptativo e 5.007 e 4.309, respectivamente para o Scan Prospectivo e Scan Score. Nesses casos, as regiões que dispararam o alarme foram o município de Guajará para os métodos Cusum Adaptativo e Scan Prospectivo e Humaitá para o Scan Score. O tempo de alarme estimado para os clusters foram $t = 58$ de modo que o *cluster* no espaço-tempo estimado pelo Cusum Adaptativo e Scan Prospectivo é representado pelo cilindro $C_{60}(\hat{Z}) = \{Guajará\} \times [58, 60]$, isto é $\hat{Z} = \{Guajará\}$ e $\hat{k} = 58$. O cilindro estimado pelo Scan Score é $C_{60}(\hat{Z}) = \{Humaitá\} \times [58, 60]$ com $\hat{Z} = \{Humaitá\}$. Uma síntese dos resultados para os últimos períodos de vigilância são apresentados na Tabela 1.

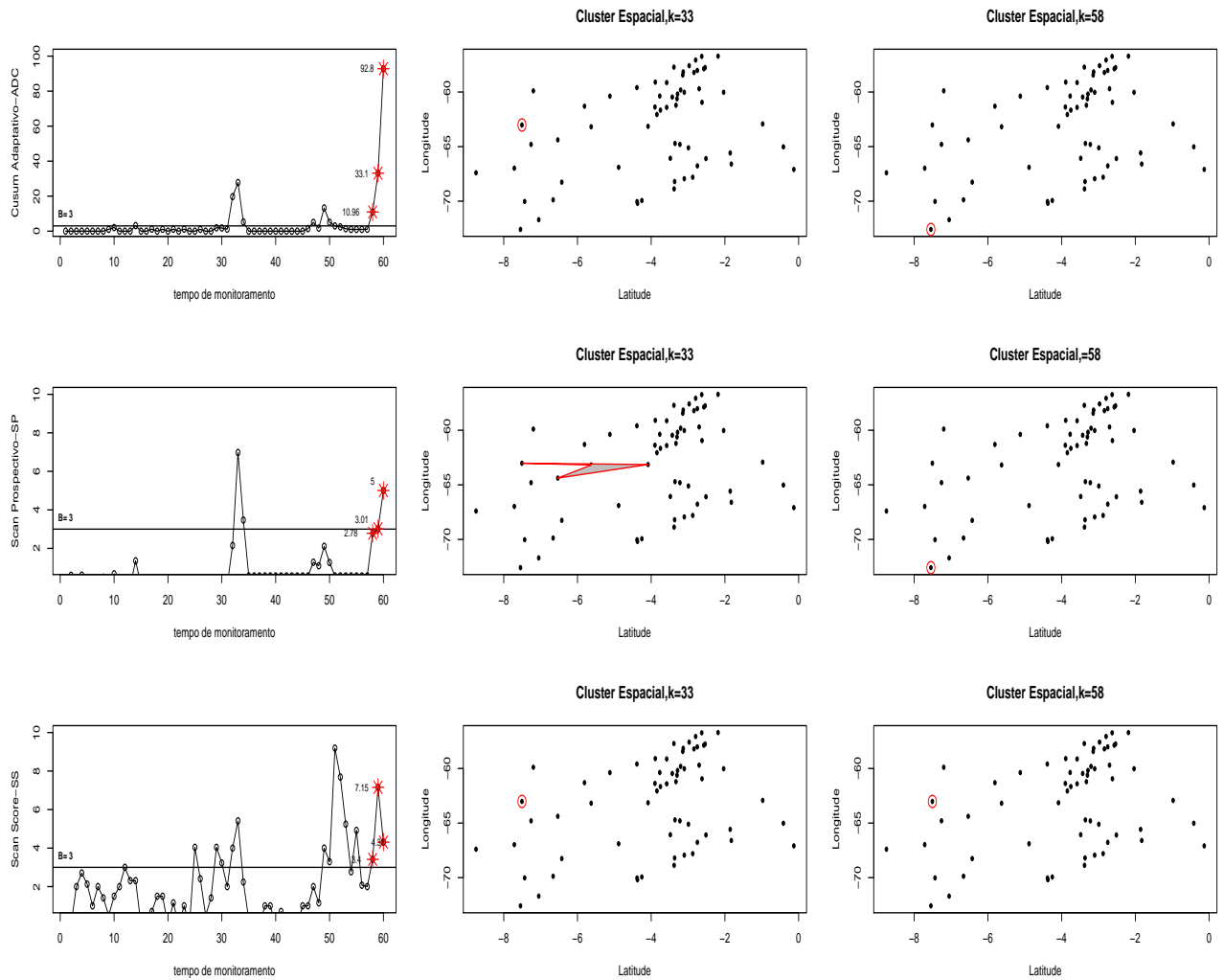


Figura 2: Vigilância On-line dos casos de Haseniasa no Amazonas usando os métodos Cusum Adaptativo, Scan Prospectivo e Scan Score, com $\log(B)=3$

4.3 Estudo comparativo entre os Métodos Cusum Adaptativo, Scan Prospectivo e Scan Score

Nesta seção nós comparamos a performance dos Métodos Cusum Adaptativo, Scan Prospectivo e Scan Score na detecção de *clusters* emergentes. O experimento computacional descrito na seção 5.3 de Lima and Duczmal (2014) é repetido para o método Scan Score. São geradas 1000 réplicas de Monte Carlo sobre as hipóteses H_0 e $H_{k,z}$ com $\delta_{k,z} = 1.5, 1.75, 2.0$ (veja, seção 3.1). Usando as réplicas sob a hipótese H_0 , estimamos o limite de controle B para o método Scan Score usando uma taxa de falso alarme α igual a 0.05 ou 5%. Usando as réplicas geradas sob hipótese $H_{k,z}$, comparamos os métodos usando as medidas de precisão para localização do cluster: Sensitividade (SS) e Valor Predito Positivo (PPV). Essas medidas avaliam a performance dos métodos para localizar a

Tabela 2: Valores Estimados para Sensitividade (**SS**), Valor Predito Positivo(**PPV**) e Atraso na detecção (**ADD**) para diferentes valores de $\delta_{t,z}$ e taxa de falso alarme ($\alpha = 0.05$), usando os Métodos Cusum Adaptativo, Scan Prospectivo e Scan Score.

| | | Medidas de Performance | | | |
|------------------|-----------------|------------------------|-----------|------------|------------|
| | | $\delta_{t,z}$ | SS | PPV | ADD |
| Cusum Adaptativo | $\alpha = 0.05$ | 1.50 | 0.799 | 0.514 | 0.955 |
| | | 1.75 | 0.857 | 0.540 | 0.591 |
| | | 2.00 | 0.882 | 0.497 | 0.118 |
| Scan Prospectivo | $\alpha = 0.05$ | 1.50 | 0.763 | 0.726 | 3.451 |
| | | 1.75 | 0.765 | 0.799 | 1.743 |
| | | 2.00 | 0.785 | 0.851 | 1.011 |
| Scan Score | $\alpha = 0.05$ | 1.50 | 0.571 | 0.634 | 3.785 |
| | | 1.75 | 0.510 | 0.742 | 3.462 |
| | | 2.00 | 0.341 | 0.679 | 2.917 |

parte espacial do cluster emergente. Também é avaliado o atraso na detecção, denotado por \widehat{ADD} e que representa uma medida para avaliar a velocidade do método na detecção da parte temporal do *cluster*. O ideal é que a **SS**, **PPV** sejam próximos de 1 e que o atraso seja pequeno. Para mais detalhes dessas medidas veja Lima and Duczmal (2011,2014). Chamamos atenção para o fato que as réplicas foram geradas apenas para o Método Scan Score. Estes resultados são comparados com os resultados para o Cusum Adaptativo e Scan Prospectivo, obtidos em Lima e Duczmal (2014). O método Cusum Adaptativo neste texto é o mesmo GMAX-ALR usado em Lima e Duczmal (2014).

As medidas de performance para os três métodos são descritos na Tabela 2. Como esperado, no Scan Prospectivo o atraso (\widehat{ADD}) na detecção decresce e a precisão na localização cresce quando o valor $\delta_{t,z}$ cresce. Quando $\delta_{t,z} = 2$ o atraso estimado é pequeno ($\widehat{ADD} = 1.011$), e os valores obtidos para precisão são **SS**=0.785 e **PPV**=0.851. Em comparação, o método Cusum Adaptativo tem o menor atraso e maior **SS** em todos os casos. No entanto, possui menor (**PPV**). Isso ocorre porque nesse método a parte espacial do cluster é estimada com os dados acumulados até o instante anterior e isso aumenta a velocidade de detecção do método. Ainda no Cusum Adaptativo, observa-se que a **SS** é sempre maior que o **PPV** indicando que na teoria o método detecta um *cluster* maior que o verdadeiro, o inverso ocorre no Scan Score e no Scan Prospectivo, quando $\delta_{k,z}$ cresce. O método Scan Score, possui o maior atraso e tem melhor (**PPV**) que o Cusum Adaptativo.

5 CONCLUSÕES

Neste trabalho nós avaliamos computacionalmente e através de aplicações os métodos Cusum Adaptativo, Scan Prospectivo e Scan Score para vigilância estatística em mapas de doenças em que o foco principal é a detecção de cluster emergentes no espaço-tempo. Em essência, O Cusum Adaptativo é baseado na Razão de Verossimilhança adaptativa, o Scan Prospectivo é baseado na Razão de Verossimilhança Generalizada e o Scan Score é baseado na Estatística Score. Esses métodos combinam técnicas de vigilância na escala temporal e algoritmos para detecção de *cluster* espacial de doenças. Os resultados obtidos em nossa aplicação foram similares para os Métodos Cusum Adaptativo e Scan Prospectivo o que reforça a boa habilidade dos métodos em detectar cluster emergentes. Ou seja, podemos concluir que esses métodos são efetivos para solucionar o problema de vigilância epidemiológica no espaço-tempo. Do ponto de vista da saúde pública uma análise mais detalhada nos gráficos de monitoramento mostraram que a região do sul do Estado do Amazonas passou recentemente por vários períodos epidêmicos, que talvez possam ser reflexos da vulnerabilidade social da região.

Nos experimentos computacionais observamos que existe um conflito entre o atraso e a precisão na detecção do cluster. Os resultados mostraram que o Método Cusum Adaptativo possui o menor atraso entre todos os métodos mas, tem um menor **PPV** quando comparado com o Scan Prospectivo. O Scan Score possui maior atraso em todos os casos. O Scan Prospectivo possui o maior **PPV**. Na literatura, em estudos epidemiológicos usando dados históricos, a precisão na localização espacial do cluster é considerada mais relevante e, neste caso o Scan Prospectivo deve ser recomendado. No entanto, em problemas de vigilância epidemiológica, a precisão na localização espacial pode ser sacrificada em função de um menor atraso na detecção. Neste caso, o Cusum Adaptativo é preferível. Embora o Scan Score não tenha obtido bons resultados nos experimentos computacionais, ressaltamos que para efeito de padronização nas comparações, usamos apenas o Scan Score baseado no modelo Poisson. No entanto, poderíamos ter obtidos resultados diferentes se tivéssemos usado o Scan Score baseado no modelo Binomial Negativo.

6 REFERÊNCIA BIBLIOGRÁFICA

1. Assunção, R., Corrêa, T (2009). Surveillance to detect emerging space-time clusters. *Computational Statistics and Data Analysis*, 53, 2817-2830.
2. Ailamaki, A., Faloutsos, C., Fischbeck, P (2003). An environmental sensor network to determine drinking water quality and security. *SIGMOD Record*, 32, 47-52.
3. Kulldorff, M., (2001) Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of Royal Statistical Society, A.*, 164, 61-72.
4. Lima, M.S, Duczmal, L.H, (2009) Endemic Disease Surveillance Using Bayes Factor. *International Society for Disease Surveillance Eighth Annual Conference*, Miami, EUA.
5. Lima, M.S, (2011). Métodos Adaptativos para detecção de cluster no espaço-tempo. *Tese de Doutorado*, Universidade Federal de Minas Gerais.
6. Lima, M.S, Duczmal, L.H., (2011). Surveillance and Detection of Space-Time Clusters using Adaptive Bayes Factor. In: *Cancer Clusters*. Editor: Bradley D. Germano, Series: *Cancer Etiology, Diagnosis and Treatments*, Nova Science Publishers, ISBN: 978-1-61209-516-5.
7. Lima, M.S, Duczmal, L.H., (2014). Adaptive likelihood ratio approaches for the detection of space-time disease clusters. *Computational Statistics and Data Analysis*, v. 77, p. 352-370.
8. Marshall, J.B., Spitzner, D.J., Woodall, W.H., (2007). Use of the local Knox statistic for the prospective monitoring of disease occurrences in space and time. *Statistics in Medicine*, 26, 1576-1593.
9. Mood, A. M.; Graybill, F. A.; Boes, D. C. *Introduction to the Theory of Statistics*, 3. ed.: Mcbraw-Hill, 1974.

10. Neill, D.B, (2006). Detection of Spatial and Spatio-Temporal Clusters. Ph.D. Thesis, University Pittsburgh.
11. Neill, D.B, (2009). Expectation-based scan statistics for monitoring spatial time. International Journal of Forecasting , 25, 498-517.
12. Porter,D.M, (2007) Some Adaptive approaches for space-time anomaly detection. First International Workshop in Sequential Methodologies, IWSM.
13. Sabhnani, M., Neill, D.B., Moore, A. W., Tsui, F.C., (2005). Detecting anomalous clusters in pharmacy retail data. In Proceedings of the KDD, workshop on data mining methods for anomaly detection 2005, 58-61.
14. Sonesson, C., (2007). A CUSUM framework for detection of space time disease clusters using scan statistic. Statistics in Medicine, 26, 4770- 4789.
15. Tango, T., Takahashi, K., Kohriyama, Kazuaki. (2011). A Space-Time Scan Statistic for Detecting Emerging Outbreaks. Biometrics, 67, 106-115.

