

MODELAÇÃO EPIDÉMICA USANDO FORMATO VECTORIAL

Joana Margarida

CASA, University College of London, 1-19 Torrington Place, Gower Street
London WC1E 6BT
Tel: +44 (0) 20 7679 4287, Fax: +44 (0) 20 7813 2843
j.simoes@ucl.ac.uk

Modelos Complexos e Sistemas de Informação Geográfica

Os Sistemas de Informação Geográfica (SIG) podem ser definidos como uma ferramenta para revelar padrões na informação geográfica, que de outra forma poderiam ser imperceptíveis (Longley, P., Goodchild, M., Maguire, D., Rhid D, 2001). A sua utilidade na modelação espacial é indiscutível e a verdade é que têm vindo a assumir um papel cada vez mais importante nesta área.

Tradicionalmente, os dados espaciais têm sido divididos em duas grandes classes: *raster* e *vectorial* (Worboys, M. F., 1995).

Os modelos *raster* são computacionalmente mais simples¹ e por essa razão estão mais difundidos. Essa simplicidade traduz-se nos conceitos, algoritmos e suporte de hardware. São eficientes para dados com uma distribuição homogénea uma vez que não lidam bem com a variação espacial. Por outro lado, tendem a ocupar muito espaço de armazenamento e têm problemas de apresentação uma vez que não suportam bem as ampliações de escala (*zoom in*).

Os modelos *vectoriais* têm um processamento mais lento, o que obriga a uma optimização dos algoritmos que efectuam operações sobre eles, mas a representação de polígonos irregulares (como lagos ou fronteiras de países) é mais realista e mais precisa²

¹http://www.soton.ac.uk/~mdw/work_modellingwithinGIS.shtml

²http://graphics.lcs.mit.edu/~seth/pubs/taskforce/paragraph3_6_2_0_1.html

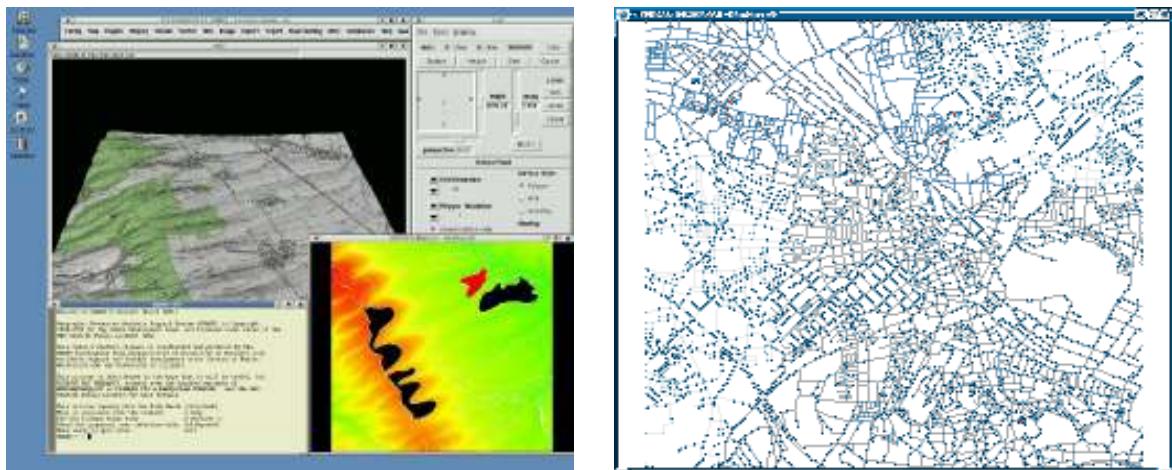


Figura 1 - Exemplos de representação *raster* (esquerda) e vectorial (direita) no sistema GRASS³.

Os problemas geográficos são frequentemente problemas de uma complexidade organizada, que tem sido inapropriadamente modelados como problemas simples, usando as leis da física ou como problemas de uma complexidade desorganizada, usando a generalização estatística⁴.

Contudo, os progressos nas capacidades dos computadores e na própria modelação computacional têm trazido técnicas que proporcionam a modelação

e compreensão destes problemas de complexidade organizada. Exemplos destas técnicas são os autómatos celulares (AC) e os modelos baseados em agentes (MBA), que pelas suas características se adaptam muito bem aos SIG *raster* e vectorial.

Estes modelos têm em comum uma ênfase na modelação do comportamento ao nível prático mais baixo, e um interesse na emergência de padrões espaciais e interacções entre agentes, que podem resultar na evolução de estratégias de interacção; esta interacção pode ser do agente com o meio ou deste com outros agentes⁵.

Ao contrário dos modelos tradicionais que estão focados em encontrar parâmetros que permitam a homeostase do sistema, os modelos baseados em agentes tendem a concentrar-se no problema da morofogénese: mudança de estrutura a uma macro-escala, como resultado de interações a uma micro-escala.

O uso de SIG com estas tecnologias de modelação permite simular dinamicamente fenómenos

3<http://grass.itc.it>

4<http://www.glue.umd.edu/~cdibble/cath/geocomp96.html>

5http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/kohler_tim/kohler.html

evolucionários e não lineares, como aqueles que encontramos em sistemas físicos, biológicos e sobretudo em sistemas sociais.

Os AC são uma classe particular de modelos dinâmicos espaciais, com regras muito específicas. Nestes modelos, a unidade básica é a célula, que por sua vez está situada num plano a 2D - esta estrutura torna-os muito semelhantes aos SIG *raster*⁶.

Podem considerar-se três graus de integração entre o SIG e o modelo⁷; do menor para o maior grau de integração: isolado, agregado e embebido.



Figura 2 - Graus de integração entre o modelo e o SIG.

No modelo isolado, a integração é nula e não há qualquer ligação com o SIG. No modelo agregado, pode haver uma integração fraca, que se faz através de ficheiros de dados, ou uma integração forte, que já inclui um *interface* comum. Finalmente no modelo embebido, o modelo e o SIG são um único produto, o que quer dizer que ou se adiciona um SIG simples a um modelo ou o modelo é implementado no SIG usando a sua *API* (*Advanced Program Interface*) ou uma linguagem de *script*.

Relativamente aos modelos agregados de integração fraca existe a vantagem de se poderem usar modelos existentes mas os dados têm de ser formatados para a troca, o que envolve um esforço adicional da parte do utilizador. Nos modelos agregados de integração forte todo este processo de troca de dados é automático, mas o modelo tem de ser escrito tendo em conta a compatibilidade com o SIG o que envolve um esforço de programação¹.

O modelo embebido dentro do SIG tem a vantagem de ser mais fácil e mais rápido de

6<http://www.geog.ucsb.edu/~kclarke/ucime/banff2000/78-mu-paper.htm>

7<http://www.css.cornell.edu/courses/620/lecture23.ppt>

implementar, porque aproveita funções já feitas, mas existe a grande desvantagem de ser mais lento e de não ser tão flexível como um programa feito de raiz.

A solução de embeber um SIG simples dentro do modelo, apresenta-se como a mais rápida e a mais flexível em termos de *output* e de alterações futuras, e foi por esta razão adoptada no modelo que se descreve em seguida.

A programação orientada para o objecto (OO), surge como uma maneira natural de implementar os agentes⁸.

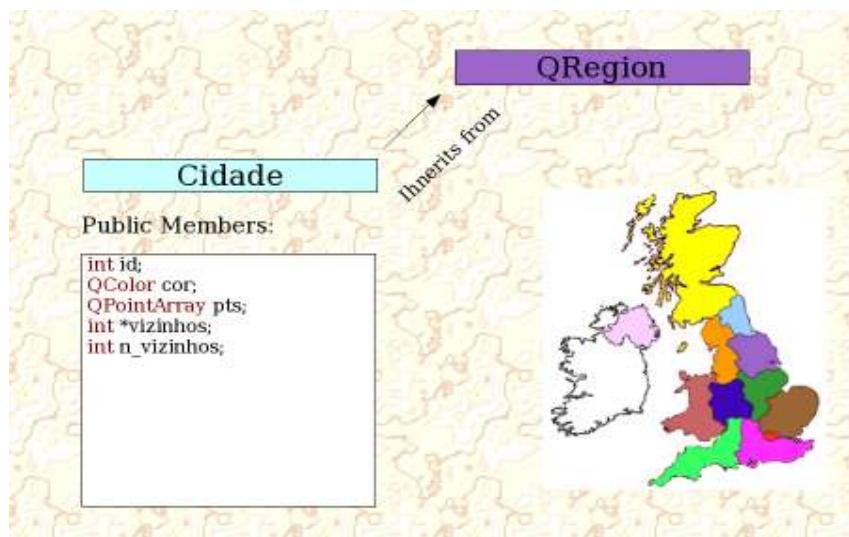


Figura 3 – Exemplo de uma classe: “cidade”, que deriva da classe “Qregion” da API do Qt.⁹

Os objectos são estruturas do programa que guardam tanto os dados (*slots*) como os procedimentos para operar sobre esses dados (métodos). Na maior parte das linguagens OO, os objectos são criados a partir de *templates* chamadas classes, que especificam a composição dos objectos, os tipos de dados que podem ter e os métodos que podem usar.

As classes estão organizadas hierarquicamente, herdando métodos e *slots* de classes superiores. Todos os objectos derivados da mesma classe têm os mesmos métodos e *slots*, embora os dados que guardam possam variar.

Uma vez que as classes necessárias para o modelo estão criadas, os indivíduos são gerados criando-se instâncias da classe.

⁸<http://www.uni-koblenz.de/~kgt/Learn/Textbook/node97.html>

⁹<http://doc.trolltech.com/3.0/classes.html>

A vantagem da abordagem OO é que as *slots* podem representar os estados internos do agente (num modelo epidémico, se está infectado ou não) e os métodos podem implementar a interpretação da regra (por exemplo, o processo de infecção). Ao especificar regras para a classe significa que todos os indivíduos instanciados dessa classe vão partilhar as mesmas regras, embora com valores diferentes.

Uma vez que o paradigma da orientação ao objecto se adequa tão bem à modelação baseada em agentes, para implementar o modelo que se descreve em seguida também foi escolhida uma linguagem OO: c++.

Modelo Epidémico

Neste estudo pretendeu-se modelar a propagação espacial de uma doença infecciosa, do ponto de vista de sistemas complexos.

Uma vez que a doença neste caso se propaga em populações humanas, está subjacente o problema do movimento, o que obrigou à conceptualização de dois modelos: um de movimento e outro de infecção.

A primeira abordagem foi considerar uma classe especial de AC adequada à descrição de movimento: Autómatos Celulares com Troca de Posição (ACTP).

Como é característico dos AC, o domínio do modelo (1) é uma malha regular, sendo que cada indivíduo ocupa uma posição na grelha e um estado (2, 3).

$$\Omega = \{0, 1, \dots, p\}^{\Lambda} \quad (1)$$

$$(\mathbf{a}_{(i, j)})_{(i, j) \in \Lambda} \quad (2)$$

$$\mathbf{a} \in \{0, 1, \dots, p\} \quad (3)$$

No modelo SIR (Susceptível-Infectado-Removido) existem três estados: os susceptíveis que são os indivíduos capazes de adquirir a doença, os infectados que são os indivíduos infectados com a doença e os removidos: aqueles que estão mortos, imunes ou recuperados da doença.

A matriz Λ contém um sub conjunto de unidades σ , sendo que $\Lambda \subseteq \sigma$.

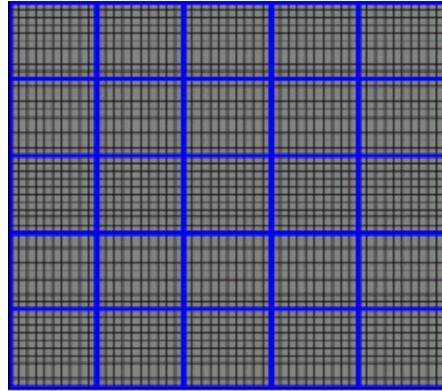


Figura 4 – Matriz Λ com as sub unidades σ (a azul).

Estas unidades foram criadas com a intenção de representarem uma unidade espacial que agrupa um conjunto de indivíduos (por exemplo, um concelho, ou uma cidade).

Os ACTP consideram em cada passo temporal, duas fases: mistura e contágio (figura 5).

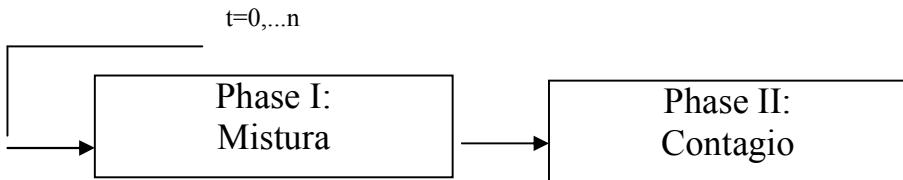


Figura 5 – Fases do modelo.

Isto significa que a regra do modelo está dividida em duas sub-regras, uma para descrever o movimento dos indivíduos e outra para descrever o processo de infecção.

Se $\tau((a_{(i,j)}))_{(i,j) \in \Lambda}$ for a imagem do elemento $(a_{(i,j)})_{(i,j) \in \Lambda}$ sobre a aplicacao de τ , a regra de mistura $\tau: \Omega \rightarrow \Omega$ é uma funcao que satisfaz as condições (4) e (5).

Isto significa que a regra do modelo está dividida em duas sub-regras, uma para descrever o movimento dos indivíduos e outra para descrever o processo de infecção.

Se $\tau((\mathbf{a}_{(i,j)}))_{(i,j) \in \Lambda}$ for a imagem do elemento $(\mathbf{a}_{(i,j)})_{(i,j) \in \Lambda}$ sobre a aplicação de τ , a regra de mistura $\tau: \Omega \rightarrow \Omega$ é uma função que satisfaz as condições (4) e (5).

$$(\mathbf{a}_{(i,j)})_{(i,j) \in \Lambda} \neq 0 \Rightarrow \tau((\mathbf{a}_{(i,j)}))_{(i,j) \in \Lambda} \neq 0 \quad (4)$$

Para cada $x_1, x_2 \in \Lambda$, $x_1 \neq x_2$:

$$a_{x1}, a_{x2} \neq 0 \Rightarrow \tau(a_{x1}) \neq \tau(a_{x2}) \quad (5)$$

As expressões acima estabelecem que cada indivíduo pode deslocar-se apenas para uma célula vazia e que dois indivíduos não podem deslocar-se simultaneamente para a mesma célula.

A regra de movimento neste caso, usa um modelo de movimento que não é totalmente aleatório, mas que considera taxas de migração.

São assim considerados quatro tipos de movimento, com diferentes amplitudes: vizinhança, intra região, inter região e aleatório, sendo que a região tem o mesmo significado que σ .

O movimento na vizinhança está limitado às células que constituem a vizinhança do indivíduo; neste caso considera-se a vizinhança de *Moore* que são as oito células adjacentes. Este tipo de movimento pretende representar o movimento de um indivíduo na sua rua ou bairro, e é lhe associada uma probabilidade elevada.

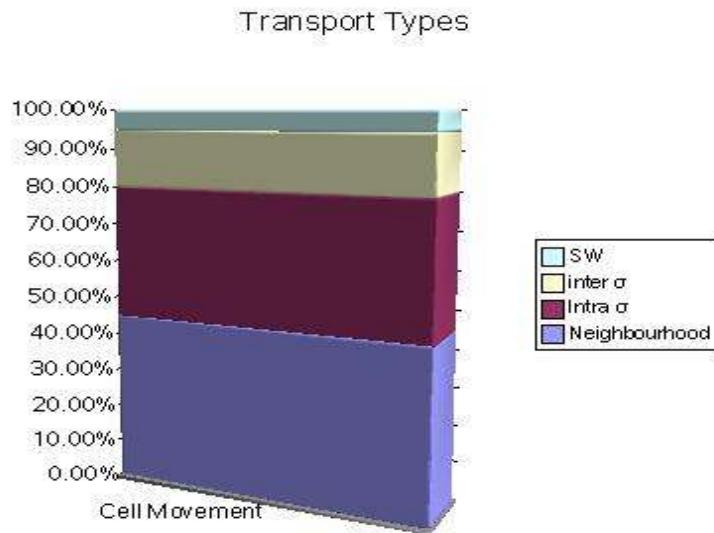


Figura 6 – Probabilidade de cada tipo de movimento, considerado na regra de transporte.

O movimento intra região está limitado à região onde o indivíduo se encontra e pretende representar as deslocações para trabalhar ou lazer dentro da cidade; a probabilidade associada ao movimento intra região é inferior à do movimento na vizinhança.

O movimento inter região é o movimento para regiões vizinhas, pretendendo representar a deslocação para cidades próximas, por exemplo para visitar familiares ou fazer compras. A probabilidade associada a este tipo de movimento é inferior às probabilidades dos movimentos referidos anteriormente.

Finalmente, e com uma probabilidade muito baixa, está o movimento aleatório. Este tipo de movimento não tem qualquer tipo de restrições e pode efectuar-se para qualquer parte do domínio.

Com o movimento aleatório pretende ilustrar-se o facto de que um pequeno numero de indivíduos é responsável por deslocações de grande curso, que têm um impacto muito grande na estrutura da rede de contactos.

De facto, esta componente é responsável pela criação de atalhos na rede, que lhe conferem uma configuração muito característica das redes sociais complexas: "Small Worlds" (Newman, M. E. J., 1999; Watts, Duncan, J., 2003).

Para a regra de contágio, foi considerado um modelo SIR, com a vizinhança de Moore.

$$(\mathbf{s}_{i,j})^{t+1} = (\mathbf{s}_{i,j})^t, (\mathbf{s}_{i-1,j})^t, (\mathbf{s}_{i+1,j})^t, (\mathbf{s}_{i,j-1})^t, (\mathbf{s}_{i,j+1})^t, (\mathbf{s}_{i-1,j-1})^t, (\mathbf{s}_{i+1,j+1})^t \quad (6)$$

A regra de contágio é determinística e totalística, baseada na regra 1022 de Wolfram (Wolfram, S., 2002). Em regras totalísticas, o estado de uma célula é função dos estados dos vizinhos. Considerando os estados como $s=I, S, R$ e 1 como uma célula infectada e 0 como uma célula não infectada, a transição de S para I é dada por (7).

$$(\mathbf{s}_{i,j})^{t+1} = \begin{cases} 1 & \text{if } \sum_{l=i-1}^{l=i+1} \sum_{c=i-1}^{c=i+1} s_{l,c} \geq 1 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

A transição de I para R é dada por (8)

$$(\mathbf{s}_{i,j})^t = I \Rightarrow (\mathbf{s}_{i,j})^{t+1} = R \quad (8)$$

A regra (7) estabelece que se pelo menos um dos vizinhos de um susceptível estiver infectado, o indivíduo fica infectado. A regra (8) estabelece que o estado removido depende apenas do estado individual no passo de tempo anterior: isto é, um indivíduo infectado passa sempre a removido.

Implementação do Modelo

A implementação do modelo CATP foi feita em Visual Basic, utilizando o *activeX Map Objects*¹⁰ para providenciar as funções SIG de visualização e análise espacial.

¹⁰<http://www.esri.com/software/mapobjects/>

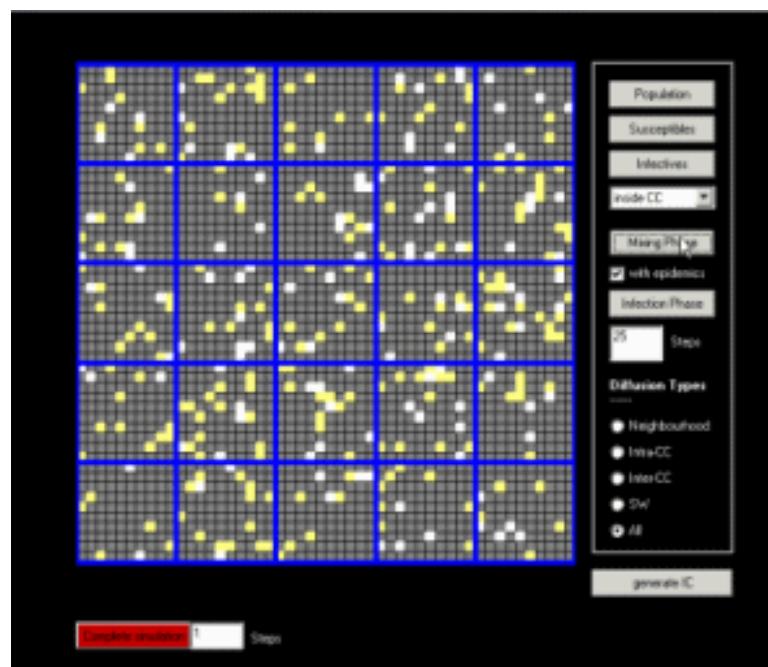
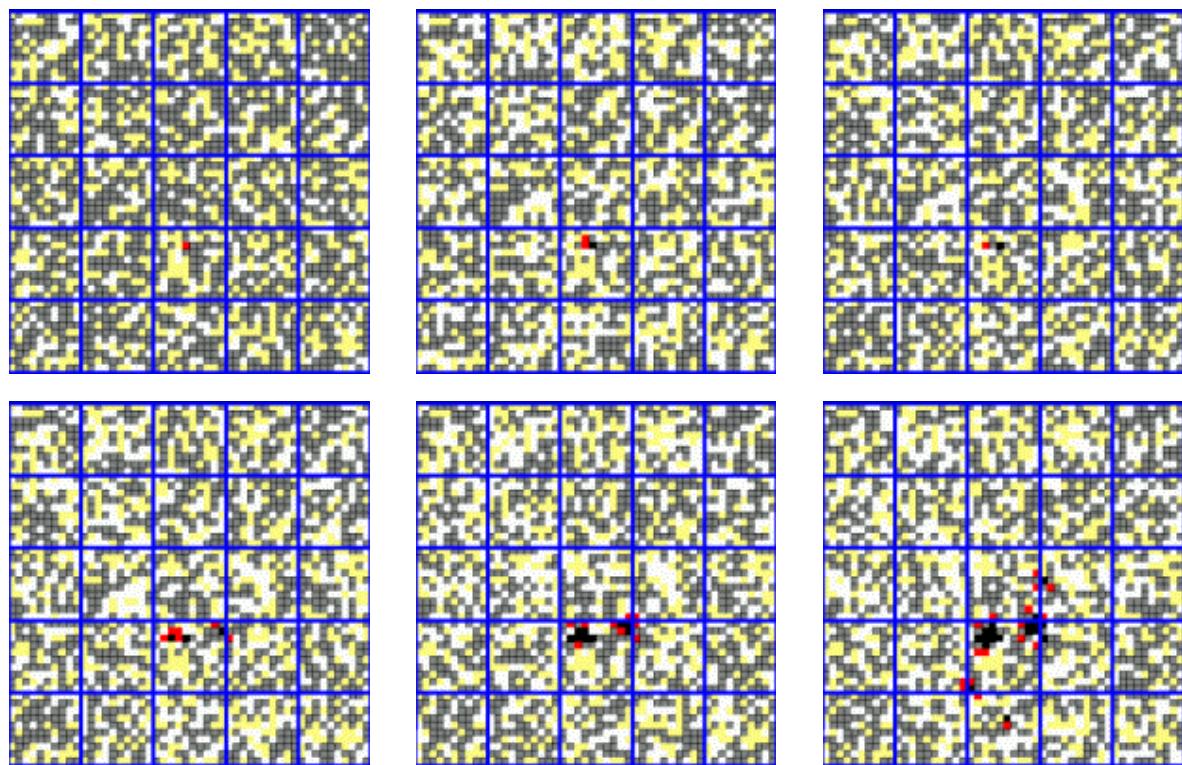


Figura 7 - Modelo de AC com as regiões a azul e os indivíduos susceptíveis a amarelo.



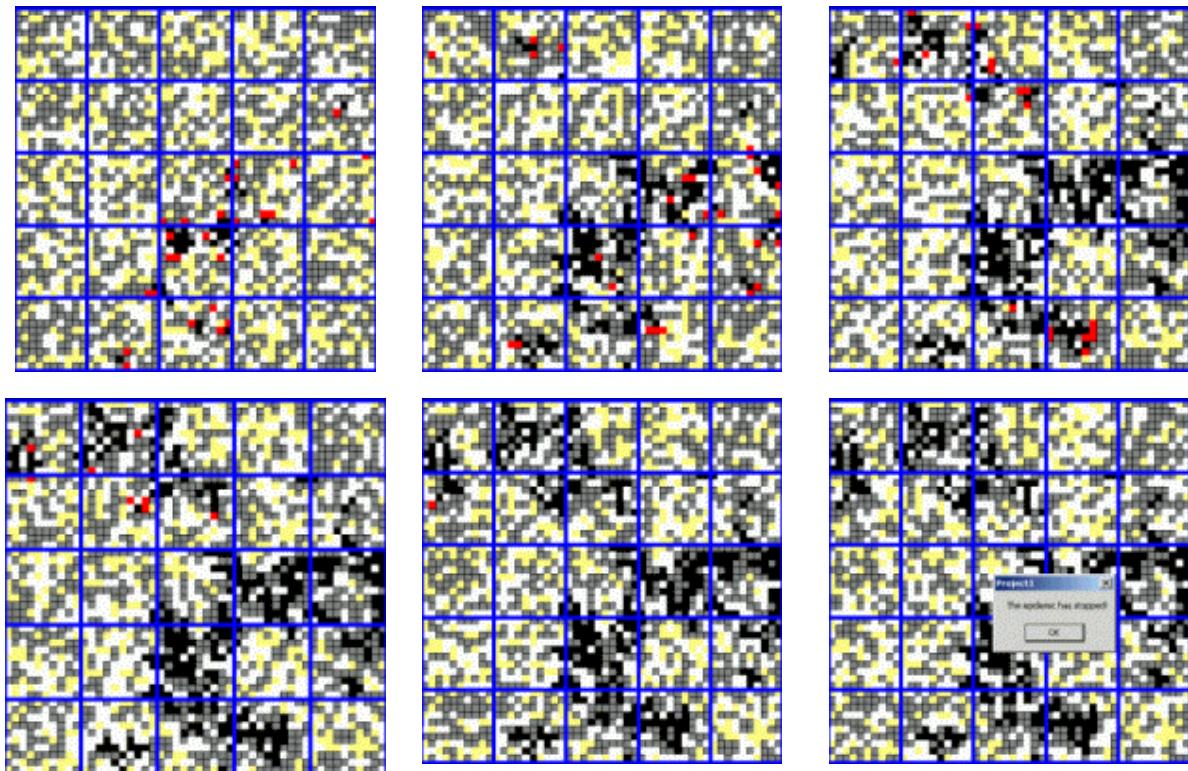


Figura 8 – simulação do modelo ACTP.

Como se tratava de um modelo de AC, foi considerada uma representação *raster* onde os indivíduos e as regiões foram representados conforme aparece na figura 7.

Na figura 8, é apresentada uma simulação que usa como condições iniciais um indivíduo infectado localizado aleatoriamente na grelha. Os indivíduos susceptíveis estão indicados a amarelo, os infectados a vermelho e os removidos a preto

A abordagem *raster/AC*, apresenta algumas limitações num modelo epidémico com representação de movimento. Essas limitações, têm essencialmente a ver com a representação discreta e homogénea imposta pelos AC.

O espaço é considerado homogêneo, o que quer dizer que as regiões são polígonos regulares, cuja artificialidade das fronteiras é bastante evidente. A vizinhança de cada indivíduo e de cada região é também homogênea em todo o domínio.

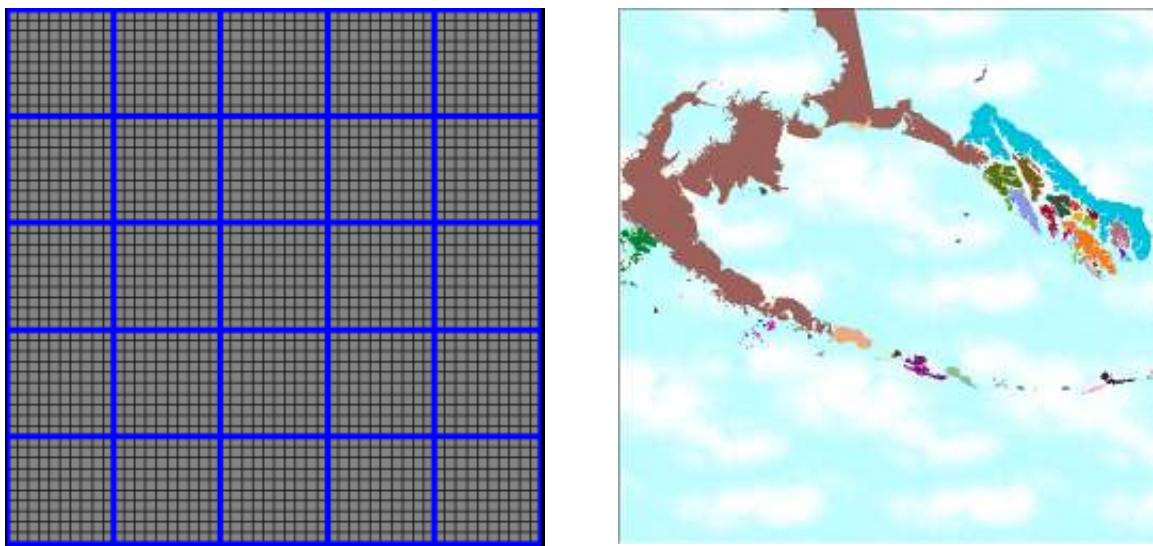


Figura 9 - Evolução da representação das regiões do modelo de AC (esquerda) para o MBA (direita).

Por outro lado, uma vez que os AC são vocacionados para modelar fenómenos locais, a representação do movimento é pobre, obrigando a variações discretas num espaço, que sabemos ser contínuo.

Por estas razões, o modelo foi re-implementado num espaço vectorial, usando agentes em vez de autómatos celulares.

Neste caso foi considerada uma camada (*layer*) de indivíduos (pontos) movendo-se livremente sobre uma camada de regiões (polígonos irregulares).

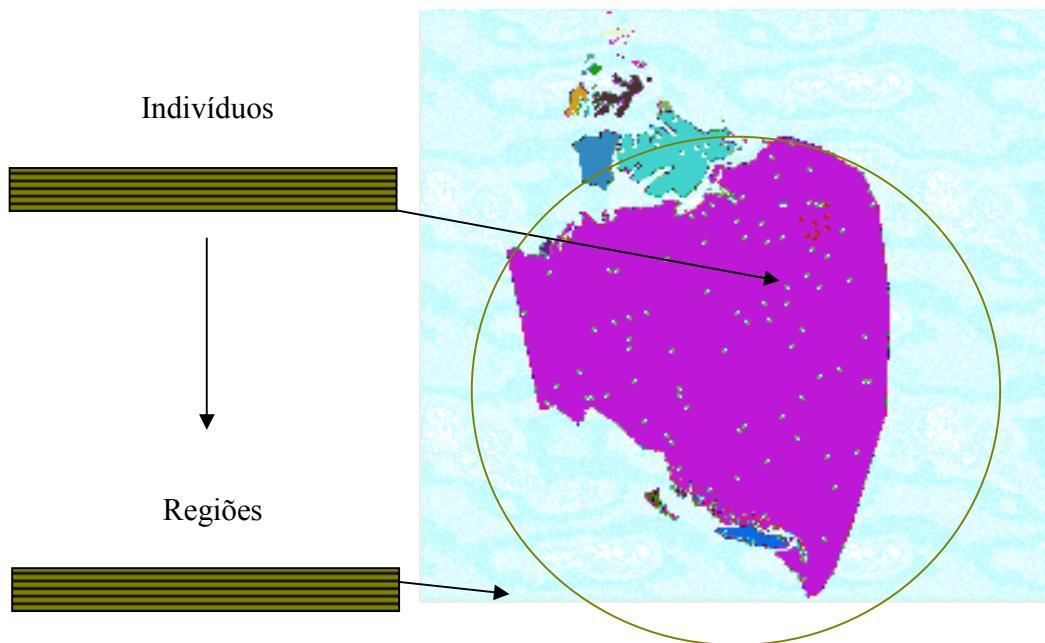


Figura 10 – Camadas que interagem neste modelo.

O modelo de movimento é em tudo semelhante ao usado nos ACTP. O modelo de contágio é SLIR (susceptível-latente-infectado-removido). Este modelo difere do anterior na medida que considera um estado latente entre o susceptível e o infectado.

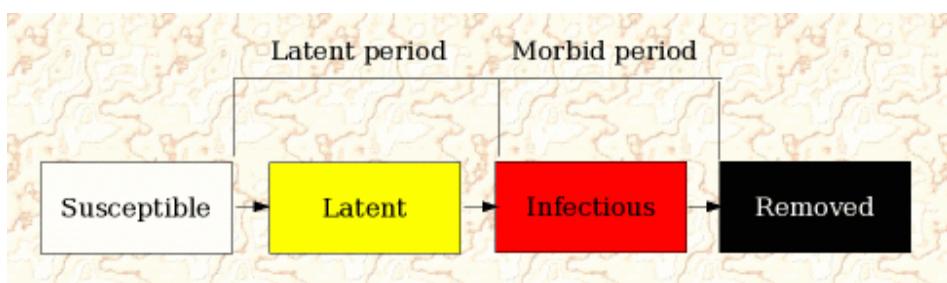


Figura 11 – modelo de contágio utilizado no MBA.

Este modelo foi integralmente programado em C++ usando a API do Qt para implementar a parte

gráfica e a Shapelib¹¹ para providenciar o acesso de baixo nível as *shapefiles*.

A Shapelib é um conjunto de bibliotecas escritas em linguagem C, que permitem fazer programas que leiam, escrevam e alterem shapefiles e respectivo ficheiro de atributos.

O programa final é constituído por uma parte principal e diversos módulos que implementam o *user interface* (UI), o modelo epidémico e as funções SIG (figura 12).

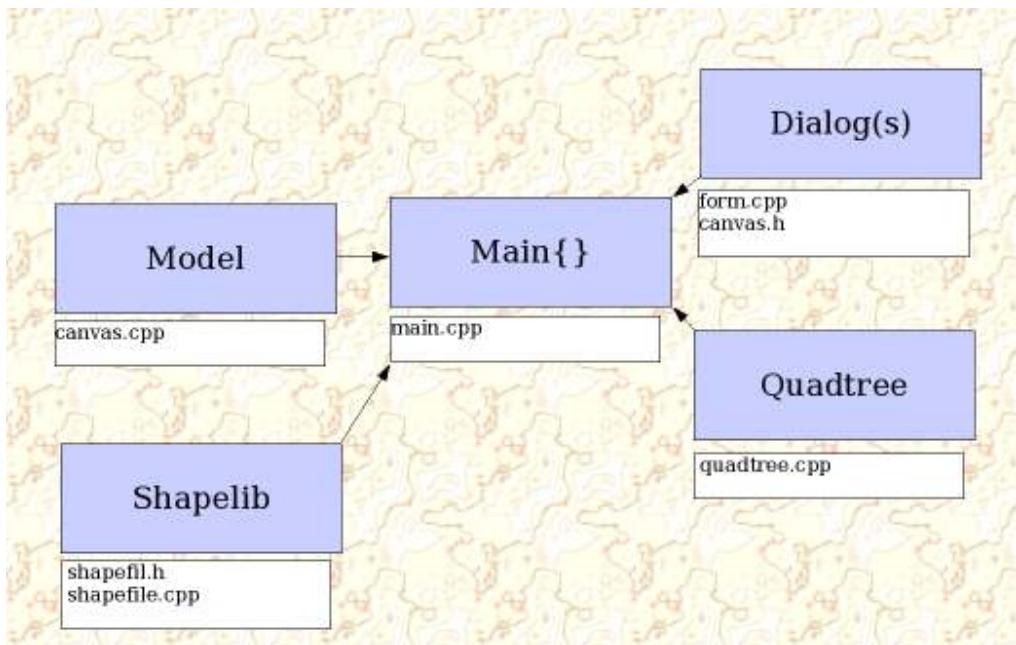


Figura 12 – Estrutura final do programa.

Como foi referido anteriormente, o processamento de dados vectoriais é mais lento que o de dados *raster* e uma vez que a simulação computacional é muito intensiva é preciso um cuidado adicional com os algoritmos que se implementam.

Neste modelo existem operações que são efectuadas muitas vezes, tais como: a busca do polígono em que se encontra um ponto, a busca dos polígonos na vizinhança de um polígono ou a busca dos pontos na vizinhança de um ponto.

De forma a optimizar estas tarefas para que o modelo não fosse demasiado lento, foram implementados alguns algoritmos espaciais que se descrevem em seguida.

A *quadtree* é um método utilizado tradicionalmente para armazenar e aceder a dados *raster*. Trata-se de um *index* espacial que decompõe recursivamente o dataset em células quadrangulares

11 <http://shapelib.maptools.org/>

de diferentes tamanhos, até cada célula ter um valor homogéneo¹².

Uma vez que este modelo é vectorial, foi implementada uma "adaptação" da *quadtree* (figura 13).

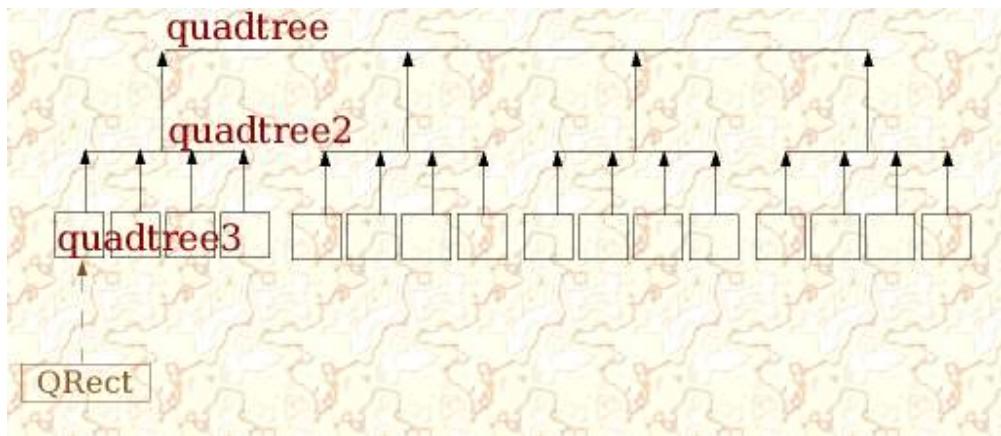


Figura 13 – Estrutura da *quadtree* vectorial, implementada neste programa.

Esta *quadtree* divide o espaço em três níveis; para cada nível, ela divide o espaço em quatro quadrados, e no ultimo nível guarda o conjunto de regiões que intersectam cada quadrado. Como por vezes uma região pode intersectar mais de um quadrado, algumas regiões são guardadas mais do que uma vez, o que não constitui um problema.

Uma vez criada a *quadtree*, para encontrar uma região o algoritmo apenas tem de percorrer a *quadtree* e no ultimo nível procurar numa coleção de regiões.

Numa camada com 100 polígonos, com nove polígonos no ultimo nível da *quadtree*, no pior dos casos a busca irá demorar 17 passos (4+4+9) contra 100 numa busca não optimizada, o que é aproximadamente cinco vezes mais rápido.

O modelo de movimento usa o conceito de "região vizinha" (cidade vizinha, concelho vizinho,...). Esta é uma região que é contígua aquela onde o indivíduo se encontra, numa analogia ao conceito de vizinhança dos AC.

12 <http://en.mimi.hu/gis/quadtree.html>

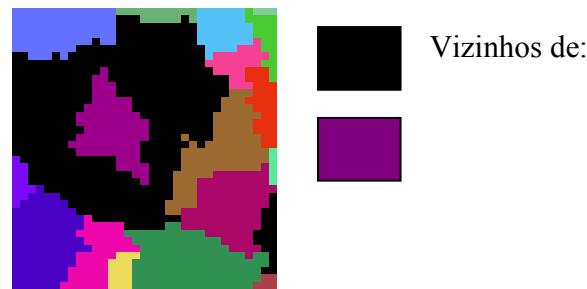


Figura 14 – Região (a violeta) com os seus vizinhos (a preto).

Uma vez que a busca de vizinhos de cada polígono (polígonos com arestas comuns) ia ser utilizada intensivamente no modelo, foi criada uma função para gerar os polígonos vizinhos, no início da simulação.

Para tornar a busca mais eficiente, ele gera apenas os vizinhos de polígonos que contém indivíduos, que são afinal aqueles que interessam ao modelo.

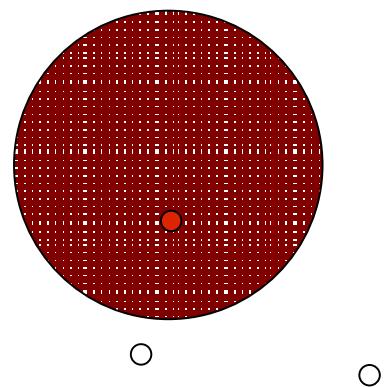


Figura 15 – *Buffer* de infecção de um indivíduo.

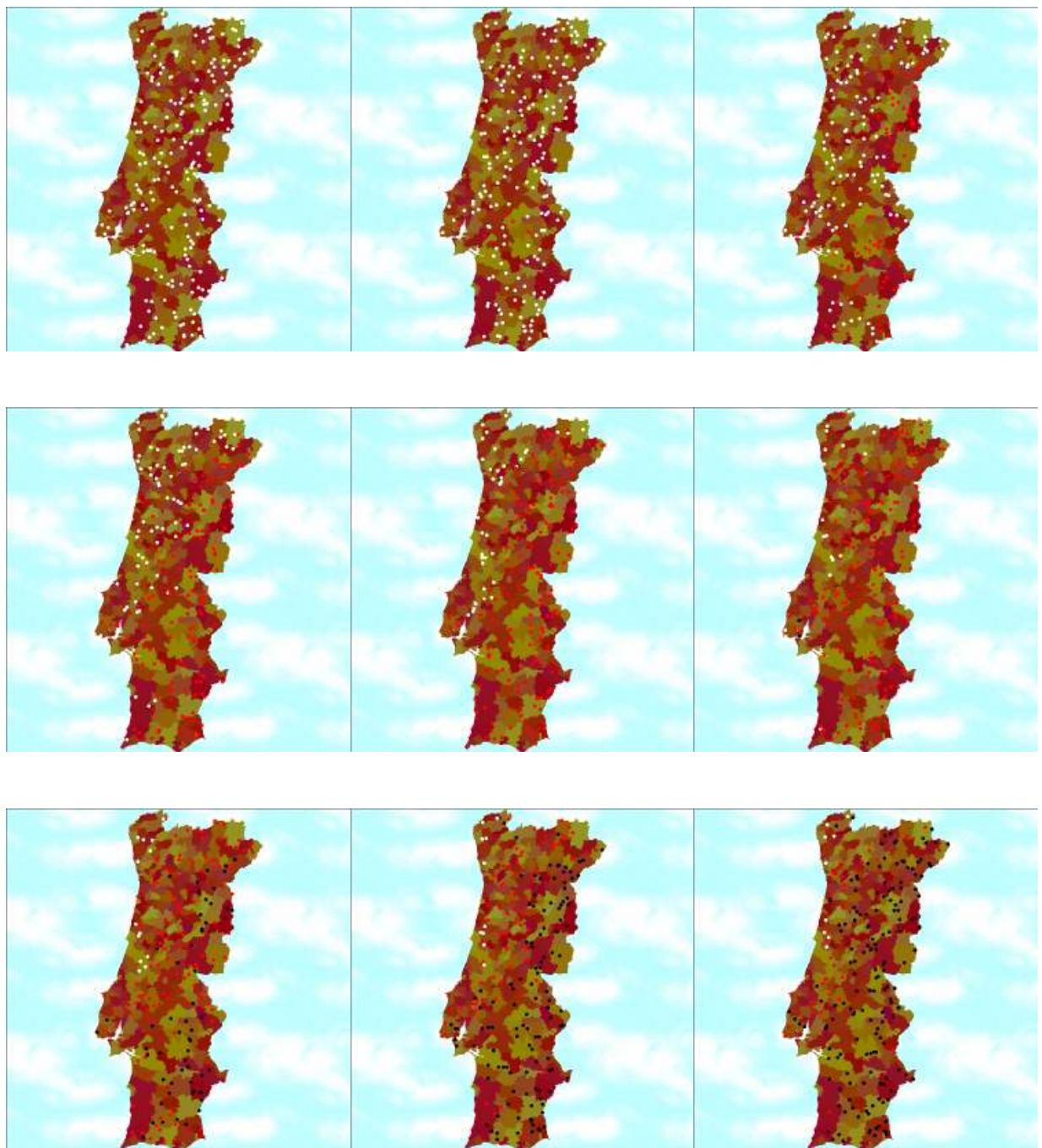


Figura 16 – Simulação $t=0, \dots, 16$ (na figura são apresentados apenas de dois em dois passos).

O modelo de infecção procura pontos dentro da vizinhança de um indivíduo. Ao contrário do que acontece com os AC, onde a vizinhança é definida por um certo número de células, aqui ela é definida através de um raio (raio de infecção).

No programa esta busca foi implementada, criando um círculo centrado no ponto e com raio igual ao raio de infecção. Este círculo é o buffer do indivíduo: qualquer indivíduo suscetível que se encontre dentro do *buffer* vai ser infectado por ele.

Na figura 16 é apresentada uma simulação utilizando o mapa de Portugal como domínio, os concelhos como regiões Ω e um indivíduo infectado como condição inicial.

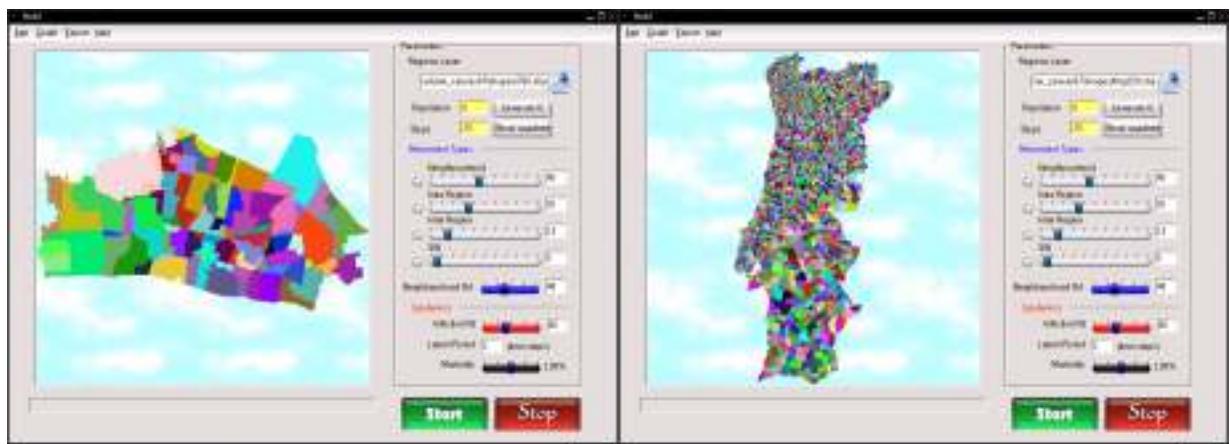
Considerações finais

Embora este estudo ainda não esteja concluído e o modelo necessite ainda de ser refinado e testado, é possível desde já tirar algumas conclusões.

Neste caso de modelação epidémica em populações humanas, parece evidente que a abordagem MBA é bastante mais realista que a de AC. A representação das fronteiras geográficas é mais realista e desta forma é possível avaliar a influência da configuração das regiões no comportamento do modelo.

A representação do movimento também é mais realista neste modelo e processa-se num espaço contínuo de coordenadas geográficas, em vez de através saltos discretos numa matriz.

É importante tornar a salientar que embora o modelo de AC e MBA sejam conceptualmente semelhantes, o facto de assentarem em estruturas de dados diferentes (*raster* e vectorial) levou a implementações completamente diferentes. Como já foi referido, a implementação de operações de análise espacial em formato vectorial obrigou à criação de algoritmos de optimização (como a *quadtree*) para que o modelo fosse eficiente, e portanto obrigou a uma programação mais cuidada.



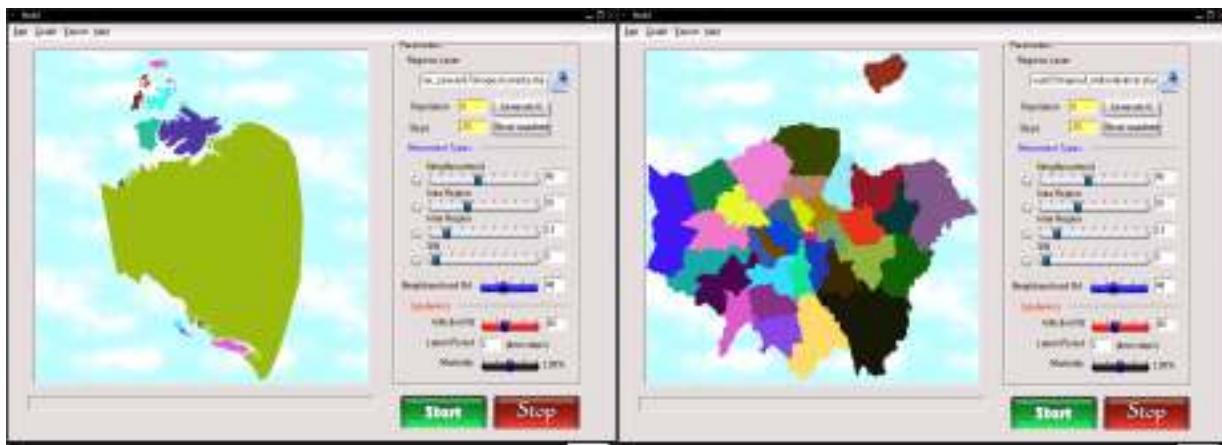


Figura 17 – Exemplos da diversidade geográfica que pode ser considerada no modelo.

A rapidez do modelo tambem está relacionada com o facto de estar programado de raiz, seguindo o modelo embebido de SIG. É de referir que em simulação computacional, a *performance* do modelo é um ponto importante, uma vez que se utiliza um grande volume de dados e um grande numero de operações.

Para terminar, e embora esteja um pouco fora do âmbito deste artigo, é de referir que a abordagem de sistemas complexos já permitiu capturar algumas propriedades interessantes da dinâmica do sistema, pelo que se revela promisora no estudo deste e de outros fenómenos geográficos.

Agradecimentos

Gostaria de agradecer a Aidan Slingsby e Michael Batty pelas sugestões na versão inglesa deste artigo e a Carlos Pinto pela revisão da versão final.

Bibliografia

- Longley, P., Goodchild, M., Maguire, D., Rhid D., *Geographic Information Systems and science*. John Wiley & Sons, England, 2001.
 Mansilla, R., Gutierrez, J., Deterministic site exchange cellular automata models for the spread

of diseases in human settlements¹³

Margarida, J., A Artificial Life Approach to Spatial Epidemics, ALERGIC talk in University of Sussex, Brighton, 28th April, 2004.

Worboys, M. F., *GIS A computing perspective*. Taylor & Francis, 1995.

Watts, Duncan. J., *Six degrees – the science of a connected age*. W.W. Norton Company, Inc., N.y. 368 p, 2003.

Wolfram, S. *A New Kind of Science*, Wolfram Media INC, 2002.

Newman, M. E. J. Small Worlds: the Structure of Social Networks. Working Pappers of Santa Fe Institute, 1999.

¹³Fonte: <http://arxiv.org/pdf/nlin.CG/0004012>