

**COMPARAÇÃO DE MÉTODOS E TÉCNICAS DE ANÁLISE MULTIVARIADA  
LINEAR (ACP) E NÃO LINEAR (SOFM) EM SIG**  
**Ensaio de modelação geográfica de indicadores de apoio a políticas públicas territoriais**

**MORGADO, Paulo; Nuno, ROCHA, Jorge; COSTA; COSTA, Eduarda**

**Resumo**

A competitividade territorial é um assunto premente. Há medida que o tempo passa, as cidades e as regiões tornaram-se mais competitivas entre si. Para esta situação, diversas questões contribuíram maioritariamente, como a globalização e o incremento dos relacionamentos entre territórios, mas também entre as instituições (públicas, privadas e Organizações Não Governamentais – ONG's), as empresas e a população. Como uma consequência (ou como uma causa) o território actua como um camaleão, sempre em mudança e a tentar adaptar-se ao ambiente circundante. Donde se conclui que, um conhecimento mais preciso e mais detalhado sobre a geografia que nos cerca, como também as linhas estratégicas para o território e as opções de localização para as empresas, ou mesmo as estratégias de políticas públicas definidas para o território, permanecem fundamentais para um modelo multivariado de avaliação, capaz de refinar processos de decisão. Neste artigo destacam-se os métodos e as técnicas para além do processo e sublinha-se o que sobressai dos resultados, que são geralmente o objectivo mais importante para a maioria dos trabalhos em SIG. Como Alan Wilson afirmou "nós vivemos no mundo da informação, onde as enormes quantidades de dados de origem e natureza diversas se arvoram em várias esferas da vida. Para retirar informação úteis destes dados, devem-se aplicar técnicas especiais de generalização e de visualização da informação contida numa determinada série de dados". Deve-se notar que existem outras questões relevantes para os resultados, como: que tipo de informação se aplica ao objectivo do trabalho; que tipo de tecnologia está disponível; que tipo de métodos estão a ser aplicados; o que é que eles fazem ; que vantagens e desvantagens têm e como podem influenciar o trabalho. Neste artigo são focados todos estes pontos de discussão emergentes e explica-se cada um deles, numa busca por novos padrões territoriais que poderiam advir enquanto contributo de uma política para o planeamento.

**Palavras-chave**

Geografia, sistemas de informação geográfica, sistemas complexos, modelo de avaliação multivariada, *self organizing feature map* (SOFM), políticas públicas territoriais e sistemas de suporte à decisão.

**Introdução**

Para enfrentar este novo desafio colectivo do desenvolvimento territorial, têm que se determinar políticas nacionais e regionais, ou, as existentes serem mais precisas sobre o que tem que ser feito: o planeamento e os modelos têm que definir, discutir, ser construídos e reconstruídos à medida que as mudanças ocorrem, para que as estratégias de desenvolvimento e as decisões sejam mais eficientes. As metodologias clássicas de planeamento, baseadas num conjunto de Planos Directores Municipais (PDM), determinísticas e de aplicação *top-down*, carecem de uma coordenação intergovernamental que seja fundamental para resolver os problemas que vão para além do nível local. Em resumo, dois pontos remanescem cruciais: os dados (mesmo os não espaciais), que representam informação e consequentemente conhecimento, e as tecnologias ou, mais precisamente, os modelos computacionais aplicados, que são ferramentas para lidar com situações complexas e para a implementação de novas políticas públicas. De facto, as tecnologias podem desempenhar um papel importante, porque potenciam as capacidades dos programas e influenciam positivamente os avanços científicos, nomeadamente com reflexos directos no desenvolvimento de algoritmia mais avançada, robusta e eficaz, e de programas de exploração e análise de dados que integram, tratam e cruzam uma grande quantidade de informação com facilidade e rapidez, uma tarefa impossível quando se recorre a metodologias tradicionais.

Para o planeamento integrado do território, a análise espacial dos dados e os SIG são as ferramentas com maior capacidade de suportar, gerir, e produzir conhecimento (ou, como a maioria dos casos, para descobrir o conhecimento) porque dão origem a tomadas de decisão mais exactas, fornecendo, não uma única solução, mas um espectro de soluções que vai dos casos mais plausíveis às situações mais improváveis que podem ocorrer em situações extremas. Neste artigo focaremos os conceitos gerais sobre análise de dados (visualização, operação, exploração e modelação), análise espacial e SIG (a contribuição de uma maior interactividade entre SIG e a análise espacial de dados), planeamento e complexidade (focando a importância da análise multivariada, que confronta técnicas estatísticas lineares e não-lineares) e finalmente, um ponto de vista crítico e comparativo sobre os principais resultados e as diferentes metodologias e parâmetros processuais.

### **Dados espaciais e políticas públicas para um desenvolvimento regional**

Para todas as organizações – públicas ou privadas – focarem-se na política, significa articular objectivos e indicadores (i.e. indicadores que possam reflectir a dinâmica e a dimensão). Muito embora, no domínio das políticas públicas, nós compreendamos os estudos, as plantas, as leis, as acções e as decisões que de uma certa maneira influenciam directamente a vida das pessoas, quando se faz uma pesquisa na Internet sobre o tema "políticas públicas", pode-se encontrar uma surpreendente quantidade de domínios que se destacam. Desde o socio-económico, independentemente dos diversos sectores, aos domínios ambientais e tecnológicos, todos referem as políticas públicas como a causa, o efeito ou a motivação do seu bem ou mal sucedido crescimento e desenvolvimento, do seu colapso ou mesmo da sua fase estacionária, emergindo como consequência da comparação entre regiões e/ou territórios. Diversos autores, provenientes de domínios científicos diferentes (e.g. Lopes, Orsini, Peyrefitte, Gilbert, Hall, entre muitos outros), têm apontado o papel e a importância das políticas públicas para o desenvolvimento regional.

Não obstante, durante muitos anos o planeamento foi realizado sem dados apropriados (sem recorrer à escala apropriada da análise ou mesmo com dados obsoletos), às vezes com uma falta enorme de dados e, mais frequentemente, sem uma articulação entre os objectivos e os indicadores de desempenho. No limite, e somente afluando o assunto, poder-se-ia dizer que para fazer uma política pública apropriada se deve ter uma base de dados orientada para objecto, bem definida e architectada, contendo o *histórico* de informação referente à área em questão. Felizmente, a temática deste trabalho circunscreve muito mais os métodos e as técnicas, para municiar uma grande quantidade de dados, aplicados ao suporte de decisões ou para resolução de problemas, no que concerne ao território e ao desenvolvimento regional. Para uma leitura adicional sobre a teoria e a ideia conceptual do território e da sua base de dados histórica deve-se rever o desenvolvimento regional (Lopes, 1987) e a integração espacial (Neto, 1999).

Muito do propósito deste artigo relaciona-se com os problemas com que nos deparámos ao longo do ensaio proposto, e que vão desde os problemas tidos em identificar e recolher a informação até aos problemas de a organizar e poder descobrir “novos conhecimentos” e apontar diferentes padrões espaciais, e encontrar explicações para diferenças territoriais crónicas independentemente de mudanças de políticas públicas. Consequentemente, muito do núcleo central do que segue será baseada nos dados espaciais adquiridos. Pelo estudo da bibliografia de muitos autores referenciados e pela nossa própria experiência, concluímos que cinco categorias principais são sempre referidas quando se trata de desenvolvimento regional: social, económico, uso e ocupação do solo, demografia, ambiente e mais recentemente, com as políticas emergentes da sociedade do conhecimento e da informação, a ciência e investigação. Neste artigo, e de maneira a usar os resultados para comparar a eficiência e as características de metodologias diferentes, experimentaram-se todos em contextos diferentes. Para o primeiro grupo das experiências utilizou-se um conjunto de dados adquirido de todos estes domínios, mas para o segundo círculo de experiências decidiu-se refinar os dados de entrada e ser mais selectivo, bem como variar parâmetros processuais nas metodologias adoptadas de forma a calibrar o modelo de avaliação desenhado, tendo em conta somente três domínios: dinâmica social; dinâmica demográfica e dinâmica económica.

**Dinâmica social** – inclui: a percentagem de pessoas com um curso superior em comparação com as que têm mais de 25 anos; percentagem de pessoas com um curso superior em comparação com as que têm mais de 15 anos. Esta escolha decorreu da forte regressão linear entre regiões de elevado desenvolvimento e o nível da instrução dos seus habitantes. Estes dados foram adquiridos para o ano de 2001, no Instituto Nacional de Estatística (INE).

**Económico** – inclui: Valor Acrescentado Bruto (VAB) *per capita* porque é esse que melhor reflecte a dinâmica económica interna, bem como as mudanças que ocorreram no sector. Para este, escolheu-se o PIB para 2001 e a variação do VAB *per capita* entre os anos de 1991 e 2001, a preços de 2003; nesta categoria também se incluiu a taxa de desemprego para 2001 e também sua variação na mesma década (1991-2001).

**Demográfico** – inclui: população residente em 2001; variação da população residente entre 1991 e 2001; índice de dependência dos velhos (ou índice de envelhecimento - rácio entre a população com mais de 65 e a população com as idades entre os 15 e os 64 anos); Taxa de actividade (percentagem população activa – entre 15 e 64 anos) e taxa de analfabetismo.

No que concerne à escala de estudo, escolheu-se a sub-região, que é equiparável com a unidade territorial estatística III (NUT III) (figura 1). Esta é a escala da análise que melhor corresponde ao compromisso entre dados disponíveis, escala territorial comparativa e tempo de processamento.



**Figura 1 – NUT's II e III para Portugal Continental, 2001**

## SIG e análise espacial dos dados

Diversos autores têm escrito sobre o conceito e as possíveis aplicações da análise espacial e dos SIG. De Goodchild, a Batty, Longley, Dangermond, Tomlinson e muitos outros, apesar de não existir consenso sobre o conceito de SIG, toda a comunidade concorda que foi devido à emergência, ou revolução, dos SIG (Yano, 2001), que os dados se tornaram espaciais ou *geodados* – à medida que os conjuntos de dados se tornaram georreferenciados. É este conceito que estabelece a diferença entre uma análise estatística e uma análise espacial. Não obstante, sempre se estabeleceu um forte relacionamento entre a manipulação de dados e os SIG (Batty, 2003), desde que a geografia evoluiu, na década de 70, para a geografia quantitativa, a qual apesar do forte criticismo que foi alvo, cresceu cada vez mais forte, até dar lugar ao seu sucessor: O SIG. A nova geografia (quantitativa) deu origem a uma geografia nova. Mais a mais, nos anos 90, à medida que as técnicas de manipulação de dados se tornaram mais avançadas e o progresso “corre” sobre tecnologia informática, permitindo o alargamento dos métodos de modelação espacial e viabilizando a transição dos SIG para as Ciências de Informação Geográfica (CIG). De facto, as CIG permitiram a inclusão de novos métodos e conceitos na geografia, no planeamento territorial, e mesmo, nos sistemas de decisão espacial – conceitos de modelos de dados complexos, emergência, modelos estocásticos, conceitos (aleatoriedade e sobreposição de *feedback*) inerentes a "outras formas de análise que são puramente determinísticas" (Batty, 2003). Mas, em primeiro lugar é necessário definir o que se entende por análise espacial.

A análise espacial envolve a exacta descrição dos dados que se relacionam com um processo que opera no espaço, bem como a exploração de padrões e relações entre eles. Assim, no que respeita ao contexto dos SIG, a análise espacial é o “passo seguinte”. Depois de superada a fase inicial de digitalização, edição e validação da informação geográfica, e a criação dos elementos que se encontram agora preparados para integrar um projecto SIG, a informação começa a tornar-se útil. Gerir, cruzar, adicionar, suprimir, modificar, intersectar, unir, fundir, sobrepor, medir, são somente algumas das funções disponibilizadas pelos SIG. Como Andy Mitchel afirma, é agora que todo o trabalho duro de digitalização, de construção de uma base de dados, de verificação de erros e de operar com os detalhes das projecções e de sistemas coordenados finalmente compensa em termos de resultados e melhores decisões.

Neste momento, poderíamos pedir ao sistema para fornecer alguns resultados baseados nas perguntas básicas: **Onde é? ou está.** Esta é uma pergunta de curiosidade básica, mas fundamental para saber que objectos/fenómenos espaciais estão em numa posição geográfica exacta. Este desejo conduz à pergunta seguinte.

**O quê?** Poderíamos encontrar num lugar em particular. Tão importante como a primeira, esta pergunta permite-nos fazer um inventário dos objectos/fenómenos que existem nesse lugar em particular.

**Como?** Podemos estabelecer a conexão entre os objectos e seus atributos, ou são mais físicos ou são mais abstractos. Se as duas primeiras perguntas somente requererem funções simples de inquirição, esta já requer princípios básicos da análise espacial e de conhecimento relacionado com o *software* e com a informação.

**Quando?** Ocorreu o fenómeno, ou começaram as transformações, ou... Esta pergunta transporta-nos à “máquina do tempo”. Tão importante como a dimensão espacial, a dimensão temporal permite a monitorização do território e o que acontece através do tempo (análise de séries temporais).

**E se?** Este é o presente/futuro da análise espacial em SIG e da geocomputação. A simulação espacial baseada em modelos complexos como os autómatos celulares, redes neuronais ou a inteligência artificial. Estes modelos realistas, estão a começar a ter um papel fundamental no planeamento de riscos ambientais, expansão urbana, doenças epidémicas, influência dos transportes e das comunicações na configuração do território, etc..

## MODELO CONCEPTUAL PARA AVALIAÇÃO DAS POLÍTICAS PÚBLICAS TERRITORIAIS

Um modelo é sempre uma generalização da realidade, fortemente marcada pela forma como a vemos ou segundo o objectivo do nosso trabalho. Assim, o modelo conceptual para avaliação de políticas públicas territoriais (figura 2) é naturalmente também uma representação imperfeita da realidade, em consequência do grau de artificialidade que lhe é conferido pelas considerações adoptadas para este projecto, independentemente da pluridisciplinidade da equipa de trabalho. Não obstante, é preciso partir de axiomas e esta é a metodologia adoptada para iniciar o ensaio de modelação geográfica eleito para este projecto e de forma a desenvolver mecanismos de análise e representação espacial de indicadores. Todavia, o objectivo do projecto está circunscrito à análise comparativa de métodos e técnicas eg. análise multivariada linear (ACP e AC) e não linear (SOFM), com vista a poder contribuir-se com uma metodologia orientada para apoio à decisão no que respeita às políticas públicas territoriais e ao exercício do planeamento, ou seja, estruturar um modelo computacional capaz de **Avaliar Políticas Públicas Territoriais (APPT)**.

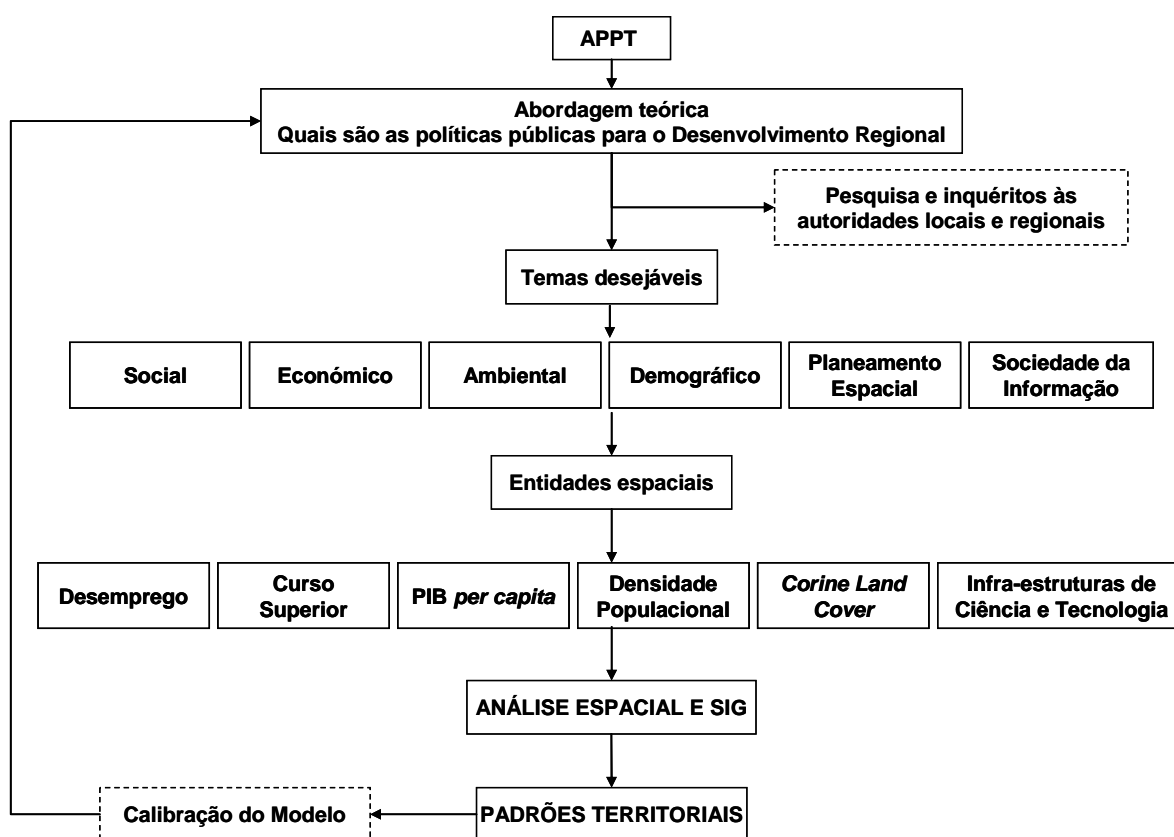


Figura 2: Modelo conceptual para avaliação de políticas públicas territoriais (APPT)

Em síntese, podemos reter daqui dois conceitos chave:

- O exercício da modelação, aperfeiçoado pelos avanços científicos e tecnológicos, confirma-se cada vez mais como o mais ajustado para uma avaliação multivariada. Efectivamente, é através da modelação que melhor se pode confrontar espaço (análise multiescala) e tempo (passado, presente e similar cenários futuros).
- Também, pelo efeito dos avanços científicos e tecnológicos, novos métodos e técnicas de análise emergiram no quadro da investigação e desenvolvimento, nomeadamente aqueles que aqui procuramos aplicar: SIG e análise espacial, e outros decorrentes da teoria da complexidade e redes neuronais. Em conjunto, estes dois conceitos chave balizam todo o trabalho de ensaio decorrido ao

longo deste projecto, conforme a seguir se expõem.

Assim assumido, as questões que o projecto agora nos exige são: que técnicas e métodos de SIG e análise especial melhor se adequam aos objectivos definidos? E como é que interagem com o modelo conceptual desenhado?

## MÉTODOS DE ANÁLISE LINEAR E NÃO LINEAR DE DADOS

### *Primeira fase experimental*

#### **Análise Multivariada e SIG**

Actualmente vivemos num Mundo em constante mudança e conseqüentemente de desafios mais complexos, que exigem respostas imediatas e de grande eficácia. A emergência de novas tecnologias, como os SIG, veio permitir desenvolver mecanismos de apoio à decisão mais musculados que os vigentes até então e mais adaptados à realidade dos problemas que se colocam aos decisores e agentes com intervenção no território, aprimorando a eficiência dos mesmos e qualificando os resultados, que no geral devem contribuir para a melhoria da qualidade de vida dos cidadãos. De facto, e advogando o contrário a que muitas vezes da oposição (que no decorrer da última década tem quase desaparecido) vinham criticando, os SIG são muitos mais do que instrumentos para a visualização de informação geográfica e a criação rápida de mapas temáticos; as suas funcionalidades e o crescente desenvolvimento neles operado por inclusão de algoritmia avançada de análise e computação modular, confere-lhes actualmente um papel fundamental em todo o espectro de actividades que usem informação geográfica ou seja, a grande maioria, quer no domínio público, quer no domínio privado.

Não obstante, muito caminho existe ainda por percorrer e desempenhos a desenvolver e aperfeiçoar. Os SIG não devem ser vistos como uma panaceia para os problemas que afligem o território ou o negócio daqueles que agem no território. Na realidade, e à medida que desejamos aplicar metodologias mais avançadas e complexas, os SIG mais comerciais ainda não nos conseguem auxiliar em toda a plenitude do problema (também porque são programas de banda larga e não orientados a situações pontuais e ou pessoais, embora seja assim que eles têm evoluído), sendo necessário recorrer a software livre ou *open source*, ou na maioria dos casos a desenvolver o nosso próprio software ou algoritmia em linguagem que seja reconhecida pelo programa SIG em que operamos, caso queiramos continuar a trabalhar nesse interface, e depois aí ‘correr’ os processos correspondentes.

Mais directamente relacionado com o nosso projecto, e quanto a métodos de análise de dados, optamos pela análise exploratória multivariada eg. análise em componentes principais (ACP) e análise de *clusters*, enquanto técnicas de combinação linear; e ainda, por contraponto a análise de redes neuronais artificiais (SOFM), mediante técnicas de não linearidade e natureza competitiva. Relembrando o objectivo definido, o pretendido era estabelecer uma comparação de métodos por confrontação e interpretação espacial dos resultados obtidos, em ambiente SIG. Por isso, vamos analisar cada uma das situações *per se*, de forma a uma melhor compreensão daquilo em que consistem, e da suas funções.

Relativamente ao SIG, optou-se por trabalhar a informação segundo uma estrutura de dados de natureza matricial. Na realidade, esta não foi tanto uma opção, mas uma imposição do sistema, na medida em que só neste tipo de estrutura de dados o programa permite ‘correr’ os procedimentos de análise multivariada. Tal é compreensível, pela capacidade de flexibilidade e de cálculo que caracterizam as estruturas matriciais face às estruturas em vectorial, e que se traduz em rapidez de processamento.

Formalmente, podemos ainda afirmar que uma estrutura de dados matricial implica um conjunto de observações  $Y_i$ ,  $i = 1 \dots, n$ , numa superfície espacial de atributos contínuos,  $S_i$ , para a região de estudo,  $R$ . É igualmente fundamental que, para todos as situações, i.e. para todos os atributos considerados, seja preciso conferir aspectos relacionados com a variabilidade desses atributos, como forma de

dissipar a incerteza, mediante procedimentos de interpolação ou predição dos valores que os atributos podem assumir noutros locais, que não o inicial. Quando mencionamos incerteza e predição, falamos da necessidade de estimar valores próximos do valor real, o qual desconhecemos à partida; não obstante, outra razão pela qual recorreremos a técnicas de incerteza neste nosso ensaio prende-se com a necessidade de normalizar os atributos multifonte que constituem o *input* da análise multivariada. Neste caso específico recorreremos à técnica *fuzzy* (figura 3) segundo aplicação de métodos linear ou sigmoidal, consoante a natureza dos atributos. Desta forma foi-nos possível dispor todos os atributos numa escala binária continua de dados de intervalo entre 0 e 1 (também podia ser entre 0 e 255), e normalizar todos os atributos, conforme exigência do programa para ‘correr’ o processo de análise multivariada.

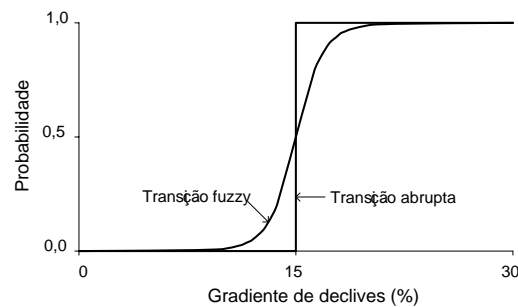


Figura 3: Exemplo da técnica *Fuzzy*

Quanto à resolução espacial da estrutura de dados matricial, esta é resultado de um acordo entre a escala de origem da informação, o objectivo do trabalho e o rigor espacial que implica, assim como as capacidades do nosso computador, quer no que respeita ao processador, quer no que respeita à RAM. Em síntese, optou-se por uma resolução de 1km, para trabalhar atributos espaciais para todo o território de Portugal Continental (92km<sup>2</sup>).

Esclarecida a situação das questões relacionadas com o SIG, estamos agora perante a problemática da análise dos dados. Embora já tenhamos afirmado anteriormente que a opção recaiu na análise multivariada, e mais concretamente na ACP e AC, para uma técnica linear e, SOFM, quando aplicado uma técnica não linear, importa referir também a importância de que se revestem estes métodos em geral, e muito particularmente para o ensaio experimental aplicado neste projecto. Em ambos os métodos, o objectivo consiste na identificação de um restrito número de variáveis (conjunto menor e resumido de variáveis independentes), por combinação linear e não linear do universo de variáveis de origem, sem perda de informação significativa.

Relativamente à técnica de análise exploratória multivariada de componentes principais, esta caracteriza-se pela capacidade em transformar um conjunto de variáveis correlacionadas num subconjunto de variáveis independentes, por combinação linear das variáveis originais, eliminando assim redundância de informação. Às variáveis independentes designamos de componentes principais. Formalmente, a expressão põe-se da seguinte forma:

$$\begin{aligned}
 C_1 &= b_{11}V_1 + b_{12}V_2 + b_{13}V_3 + \dots + b_{1n}V_n \\
 C_2 &= b_{21}V_1 + b_{22}V_2 + b_{23}V_3 + \dots + b_{2n}V_n \\
 &\dots \\
 C_n &= b_{n1}V_1 + b_{n2}V_2 + b_{n3}V_3 + \dots + b_{nn}V_n
 \end{aligned}
 \tag{1}$$

Em que  $C$  é a componente,  $b$  o parâmetro de ajustamento e  $V$  a variável.

Assim, para este primeiro exercício com base na ACP (tabela 1), utilizámos 8 variáveis de *input* que se agrupam nas seguintes classes: infra-estruturas de redes de estradas e ferrovias e tecido urbano, ambas retiradas das classes de uso do solo da carta *Corine* 2000; espaços de ciência, ensino e investigação;

densidade populacional e taxa de desemprego 2001; valor acrescentado bruto *per capita* 2001, a sua variação entre 2001 e 1991 e a população com educação ao nível do ensino superior.

Tabela 1: Matriz da ACP

Variáveis	C1	C2	C3	C4	C5	C6	C7	C8
idri_cor2_trapfuzz	0,049	-0,164	-0,969	0,054	-0,169	0,020	0,000	0,000
idri_ctec_fuzz	0,252	-0,673	-0,059	-0,462	0,489	0,167	0,003	-0,001
idri_cor2_urbfa_fuz	0,239	-0,672	0,236	-0,107	-0,639	0,127	0,002	-0,001
idri_desmp01_fuzz	0,681	0,358	-0,045	-0,512	-0,149	-0,300	-0,177	-0,022
idri_dpop01_fuzz	0,541	-0,517	0,051	0,463	0,165	-0,428	-0,112	-0,003
idri_txesup_fuzz	0,962	0,108	0,000	0,041	0,010	-0,031	0,207	-0,134
idri_vab_01_fuzz	0,968	0,140	-0,002	0,000	-0,002	-0,013	0,121	0,166
idri_var_vab_fuzz	0,769	0,174	0,014	0,257	0,055	0,527	-0,178	-0,019
Vector Próprio	3,33	1,39	1,00	0,77	0,73	0,60	0,13	0,05
% Total de variância	41,63	17,38	12,50	9,63	9,13	7,50	1,63	0,63
% Variância acumulada	41,63	59,01	71,51	81,14	90,27	97,77	99,40	100,00

A ACP permite-nos extrair três componentes principais que explicam mais do que 71% do total de variância. A primeira componente é composta pelos atributos socio-económicos que revela correlação forte com os atributos nível de ensino superior, VAB *per capita* e variação do VAB. Os *scores* desta componente mostram-nos porções do território nacional onde os contingentes populacionais denotam maiores habilitações, economia com uma forte performance e simultaneamente uma variação económica entre 1991 e 2001 positiva. A segunda componente revela-nos uma correlação negativa entre áreas com predomínio de tecido urbano e a presença de espaços de ciência e investigação. Os *scores* negativos elevados mostra-nos os territórios predominantemente urbanos e de maior fulgor tecnológico. Finalmente, a Terceira componente está fortemente correlacionada com os atributos infra-estruturas de transporte. Os *scores* permitem-nos identificar territórios com elevado nível de conexão.

Esta ACP, em ambiente SIG, já nos permitiu de forma célere e relativamente fácil estabelecer algumas diferenciações territoriais, com base em três vectores cruciais potenciais ao desenvolvimento: economia e recursos Humanos; urbanização e ciência e investigação; e acessibilidade. (figura 4).

O próximo passo experimental consiste na análise de *clusters* (AC) com base nas componentes principais obtidas pela matriz atrás apresentada. Do resultado, esperamos obter padrões territoriais que ampliem a nossa leitura apenas com base na ACP, conforme descrito abaixo:

- Com o identificador 3 - o território mais urbano, e com mais recursos Humanos e potencial económico;
- Com o identificador 2 – territórios urbanos com potencial de ciência e investigação;
- Com os identificadores 1 e 4 – territórios com pequenas diferenças em termos de potencial económico e de recursos Humanos.

Apesar do alento retirado da obtenção desta primeira experiência, que revelou padrões territoriais diferenciados, outras experiências podem e devem ser feitas, não só porque existem e como tal se enquadram no objectivo deste ensaio, como também porque nesta experiência algumas questões ficaram em aberto, nomeadamente a forte diferenciação entre as várias sub-regiões Alentejanas, ou os pequenos pontos vermelhos em toda a envolvente à cidade de Beja e que são praticamente omissos em Évora. Assim, optou-se por iniciar uma segunda experiência com as mesmas variáveis originais com base noutra metodologia (Kohonen SOFM - Cherkassky e Mulier, 1994) e técnicas (não linearidade), por modo a poderemos confrontar resultados e obter respostas e a confirmar ou refutar hipóteses.

Não obstante as razões acima apontadas para se iniciar nova experiência, outras relevam de natureza menos empírica. Várias têm sido as vozes (Wilson, Batty, Haggett, Openshaw entre outros) que no



decorso das últimas duas décadas têm aludido ao facto dos sistemas urbanos e regionais (Willshaw e von der Malsburg, 1976) serem classificados como sistemas complexos e como tal com características perfeitamente ajustadas aos conceitos decorrentes da teoria da complexidade e suas propriedades de emergência e não linearidade. Desta forma, e por mais esta razão, pretendemos com esta nova experiência olhar e analisar o território, pelas variáveis aqui adoptadas, como um todo, na procura de descobrir novos padrões que a técnica linear encobre e por conseguinte não deixa visualizar quando se espacializam; procuramos descobrir idiosincrasias territoriais e responder ao que os métodos mais clássicos até aqui têm deixado por responder. Ainda com base nas palavras de Sir Alan Wilson (Wilson, 2000) “When we solve a set of linear equations, the solution, if there is one, is unique. If the equations are nonlinear, it turns out there are typically very large numbers of solutions.” e que “...the actual solution will be determined by the initial conditions. We would usually begin with a search for equilibrium solutions”, que demos início a uma nova vaga experimental na análise espacial deste ensaio.

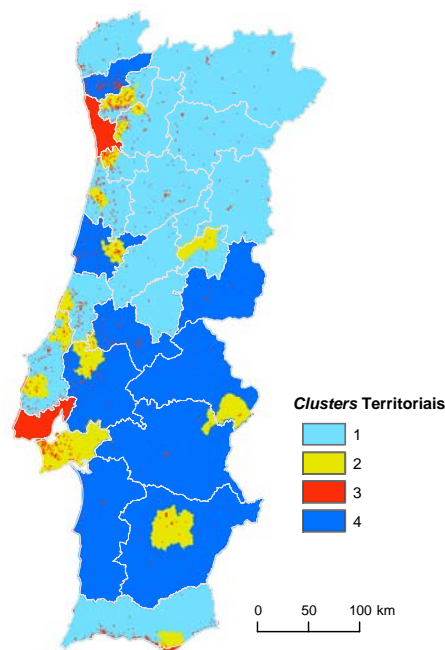


Figura 4: *Clusters* regionais resultantes da análise multivariada, para Portugal Continental

### Self organizing feature map (SOFM)

SOFM transforma as variáveis de *input* de qualquer dimensão num *output* de uma ou duas dimensões, com base em parâmetros de restrição como o raio de vizinhança topológica, o número de neurónios (daqui em diante passamos a designar por nós) e a taxa de aprendizagem, e sem perda do padrão inicial da informação. Os atributos de *input* são computados segundo a técnica não supervisionada (alguns autores também apelidem de não dirigida) de aprendizagem de Kohonen, o que significa que a rede neuronal reconhece alguns padrões, regularidades e correlações dos dados originais e ‘aprende’ a reflectir essas propriedades na saída (Estanqueiro, 2005). O *output* do SOFM pode depois, eventualmente servir de *input* para uma análise neuronal artificial supervisionada, tal como a *multi layer perceptron* (MLP).

Resumindo, a vantagem do SOFM para este ensaio é a sua capacidade de agrupamento de dados, com base em processos de aprendizagem das suas unidades. Desta forma é capaz de reduzir significativos volumes de informação em entidades representativas do material original (à semelhança do que ACP faz), mas baseado na técnica não linear, o que permite de certa forma fugir à maior generalização causada pelas técnicas lineares (figura 5).

O método SOFM (Haggett, 2001; Kohonen, 1996; Kohonen, 1993; Kohonen, 1995c) consiste num

algoritmo de rede neuronal usado num largo espectro de aplicações, sobretudo no domínio das engenharias, mas também na resolução de problemas relacionados com análise espacial de dados e na tentativa de fabricar agrupamentos de dados (Back *et al*, 1996; Orsini, 1998). Para uma melhor compreensão do que é, e para que serve, podemos encontrar informações esclarecedoras e com exemplos no domínio da análise exploratória de dados e na comparação de resultados entre SOFM (Kohonen, 1993) e outros algoritmos. Nalguns estudos sobre *Data mining* podemos encontrar referência a dois tipos de instrumentos de análise de dados: métodos de *clustering* e métodos de projecção. O SOFM é o único caso especial em que é apresentado como valioso instrumento em ambos os métodos, dada a capacidade que detém em reduzir significativos volumes de informação e simultaneamente ser capaz de projectar esses grupos num ou dois mapas de saída.

Torna-se assim evidente as razões pela qual optámos por usar esta metodologia SOFM no nosso ensaio de análise espacial de dados, e ainda que, como comprovámos no nosso exercício, a saída gráfica (mapa) não seja o principal objectivo do SOFM, só as tabelas com os agrupamentos (*clustering*) dos dados é já de inestimável riqueza informativa (Cherkassky e Mulier, 1994). Não obstante, e porque é relevante para explicar os resultados obtidos, vejamos como funciona o SOFM e a importância de alguns dos seus parâmetros de processamento.

O número de nós num SOFM precisa de ser suficientemente grande para se perceber a capacidade de redução do seu algoritmo. Estes pequenos conjuntos dos nós fornecem uma representação subsimbólica dos grupos de dados. No processo do treino estes formam automaticamente pequenos conjuntos de nós transformando-se em protótipos, ordenados de acordo com sua similaridade. A figura 5 descreve como o SOFM reduz a dimensionalidade dos dados, de dois para um, comparativamente ao uso de uma componente principal.

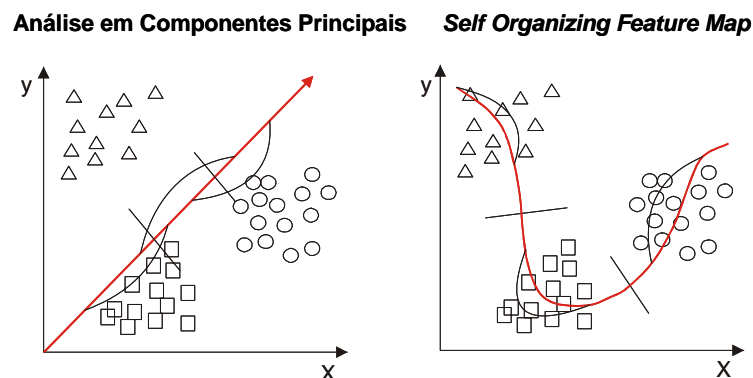


Figura 5: Redução da dimensionalidade num caso bidimensional. Se apenas for utilizado o primeiro componente principal, nenhuma clara fronteira de decisão será encontrada, e a redução da dimensionalidade não melhora a análise. O SOFM pode fornecer uma "curva" na qual as ocorrências (pontos de dados) podem ser apropriadamente projectados tornando fáceis os agrupamentos. Neste sentido o SOFM comporta-se de maneira similar às curvas principais (Batty, 2003; Gilbert, 1996).

Depois de treinados os nós, cada vector de dados pode ser atribuído (classificado) a um dos nós. Uma vez que os dados originais estejam agrupados nesta maneira, as propriedades estatísticas das variáveis de treino podem ser calculadas e visualizadas.

A aprendizagem competitiva, relacionada com este processo, é um procedimento adaptativo em que os nós numa rede neuronal se tornam gradualmente sensíveis às diferentes categorias da entrada, constituídas por conjuntos de amostras num domínio específico do espaço da entrada (Amari, 1980; Kohonen, 1995b; Kohonen, 1996; Peyrefitte, 1998). Um tipo de divisão de trabalho emerge na rede quando os diferentes nós se especializam de forma a representar tipos diferentes de entradas.

A especialização é reforçada pela competição entre os nós: quando uma entrada é activada, o nó que a melhor pode representar ganha a competição e "é autorizado" a aprendê-la ainda melhor, como será

descrito abaixo.

Se existir uma ordem entre os nós, i.e. os nós estão localizados numa malha discreta, o algoritmo de aprendizagem (o SOFM) competitivo pode ser generalizado; caso contrário, não somente o nó vencedor mas também os seus vizinhos da malha poderão aprender, os nós vizinhos especializar-se-ão gradualmente de forma a representar entradas similares, e as representações tornar-se-ão ordenadas na malha do mapa. Esta é a essência do algoritmo SOFM:

1. Para nosso projecto, os exercícios com a rede neuronal foram estruturados em 3 fases:
2. Aplicação do SOFM às 8 variáveis iniciais, para que fosse possível confrontar os resultados com a primeira análise multivariada (tabela 2, figura 4);
3. Aplicação do SOFM às 11 variáveis (baseadas nas categorias, demográfica, social e económica) para confrontar com os resultados da segunda análise multivariada (análise factorial e de componentes principais);
4. Aplicação do SOFM aos 3 resultados principais da segunda análise multivariada para que se possa comparar a sua robustez e o comportamento dos dados.

No entanto, nesta fase e no que concerne ao SOFM, este permanece somente adstrito ao ponto número 1 destas etapas: uma experiência com 8 variáveis. A rede é criada a partir de uma malha de nós 2D, cada um dos quais inteiramente conectado à camada da entrada. A figura 6 mostra uma pequena rede de Kohonen representando um vector bidimensional.

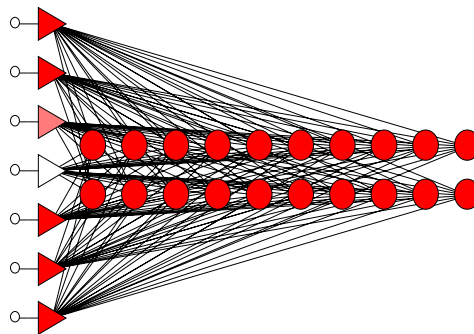


Figura 6: SOFM – Rede neuronal de Kohonen com base em 8 variáveis

Durante o período de treino, cada nó com uma actividade positiva dentro da vizinhança do nó vencedor, participa no processo de aprendizagem. Um elemento processador vencedor é determinado para cada vector de entrada, baseado na similaridade entre o vector de entrada e o vector ponderado. Considere-se o vector de entrada,  $X$ , tal que:

$$X = [x_1, x_2, \dots, x_p]^t \quad (2)$$

O vector ponderado,  $W_j$ , correspondente ao neurónio  $j$  da camada da saída pode ser definido:

$$W_j = [w_{j1}, w_{j2}, \dots, w_{jp}]^t \quad j = 1, 2, \dots, n \quad (3)$$

Os nós representam as entradas com os vectores da referência  $X_i$ , cujos componentes correspondem aos pesos sinápticos. Um vector da referência é associado a cada nó, chamado unidade, num ajuste mais abstracto. A unidade, indexada com  $c$ , cujo vector de referência é o mais próximo à entrada, é o vencedor da competição:

$$c = c(x) = \arg \min_i \{ \|x - w_i\|^2 \} \quad (4)$$

Normalmente recorre-se ao uso da medida euclidiana, embora outros métodos possam constituir alternativas plausíveis.

A unidade vencedora e os nós vizinhos adaptaram-se para melhor poderem representar os dados de *input*, mediante modificação sobre os seus vectores de referência. O conhecimento aprendido pela unidade vencedora é gerido mediante uma vizinhança do tipo filtro de Kernel  $h$ , o qual decresce à medida que a distância dos nós à unidade vencedora aumenta. Caso as posições dos nós  $i$  e  $j$  na superfície matricial e organizadas bidimensionalmente pelos vectores  $r_i$  e  $r_j$ , respectivamente, então tem-se:

$$h_{ij}(t) = h(\|r_i - r_j\|; t), \text{ em que } t \text{ representa o tempo.} \quad (5)$$

Durante o processo de aprendizagem num determinado tempo  $t$ , os vectores de referência modificam-se iterativamente de acordo com a seguinte regra de adaptação:

Em que  $x(t)$  representa o *input* num determinado tempo  $t$  e  $c = c(x(t))$  é o índice da unidade vencedora:

$$w_i(t+1) = w_i(t) + h_{ci}(t)[x(t) - w_i(t)] \quad (6)$$

Na prática, o filtro de vizinhança foi escolhido no início do processo de aprendizagem de forma a garantir uma ordenação global das variáveis de *input*, bem como pelo facto de este método melhor se ajustar ao tipo de redes neuronais que são as Kohonen ou SOFM, isto é redes neuronais não supervisionadas, em que o conjunto de variáveis de treino correspondem a todas as variáveis de *input*.

O processo de aprendizagem consiste na selecção da unidade vencedora, conforme a equação 5, e adaptação dos pesos sinápticos, pela equação 6, e pode ser modelado segundo uma estrutura de rede neuronal, na qual os nós formam agrupamentos mediante conexões de restrição (Goodchild, 1992; Kohonen, 1981). Como resultado, o mapa de Portugal Continental (figura 7).

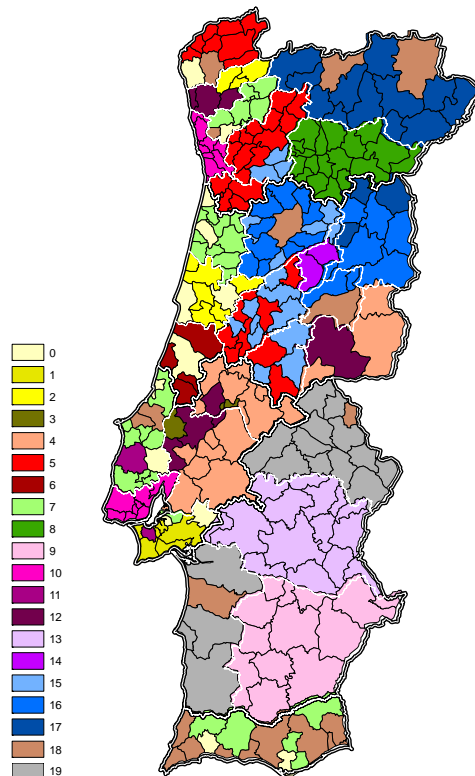


Figura 7: SOFM Clusters regionais para Portugal Continental

## Primeiras conclusões

Os resultados obtidos com a presente metodologia de análise permitem salientar dois aspectos fundamentais:

1. O método de agregação permitiu chegar a 19 grupos distintos tradutores da realidade concelhia do território continental. Tendo por base o grau de urbanização, o sistema de povoamento e de ordenamento do território, assim como as dinâmicas económica, social e de inovação;
2. Os 19 grupos obtidos representam situações regionais diferenciadas, que denotam um nível de “ajustamento” às características dos territórios muito superior ao obtido na metodologia anterior (ACP) onde se obtiveram apenas 5 grupos.

Este ajustamento às diferentes realidades regionais é particularmente visível a dois níveis: a individualização do fenómeno urbano com diferentes níveis de serviços e individualização de bacias económicas ou de aglomeração, que traduzem a base económica e a bacia de emprego de cada um dos territórios.

Assim, na Região Norte é possível individualizar claramente:

- O Vale do Douro (Grupo 8);
- O território transmontano (Grupo 17), apenas interrompido pela matriz urbana de Chaves e Bragança (Grupo 18);
- A Área Metropolitana do Porto (Grupo 10);
- Um território de características urbano-difusas, onde coexiste uma matriz rural com uma forte densidade populacional e uma crescente expansão urbano-industrial. Aqui está o grupo 5, que engloba parte do Minho-Lima e do Ave e Tamega;
- O maior grau de urbanização do eixo Esposende-Barcelos-Braga aparece evidenciado no Grupo 12 e parte do Vale do Ave, onde se inclui Guimarães, surge igualmente individualizado (Grupo 7).

No caso da região Centro, os resultados da análise são igualmente muito favoráveis, pois permitem individualizar claramente os sistemas territoriais do Litoral, com um padrão de ocupação mais denso e urbano-difuso, dos do interior, com um padrão de ocupação menos denso e mais polarizado. Dentro de cada um destes é ainda possível encontrar vários sub-sistemas regionais:

- O sistema de Aveiro (Grupos 0 e 7);
- O sistema de Viseu (Grupo 18 e 16);
- O sistema de Leiria (Grupo 0 e 6);
- O sistema de Coimbra (Grupo 0 e 2) onde são visíveis já os prolongamentos da sua bacia de emprego a Ansião, Alvaiázare, Miranda do Corvo e Poiães que se incluem no Grupo 5;
- O Pinhal Interior (Grupo 15, “interrompido” pelos concelhos de Góis, Sertã e Proença-a-Nova);
- O sub-sistema formado por Seia-Gouveia-Oliveira do Hospital;
- A Beira Interior Norte, incluindo aqui a Covilhã e Manteigas (Grupo 16);
- A Beira Interior Sul, onde se inclui o Fundão, face ao alargamento da bacia de emprego e das relações funcionais de Castelo Branco.

Na Região de Lisboa encontramos:

- A Área Metropolitana de Lisboa norte (incluindo Almada) (Grupo 10);
- A Área Metropolitana Sul, cuja demarcação se deve ao diferente modelo de ocupação do solo (Grupo 1);
- O prolongamento urbano para norte, evidenciando o *urban sprawl* correspondente aos Grupo

7 e 12, que caracterizam igualmente as envolventes de Aveiro e da AMP no alinhamento de Braga e Guimarães.

No caso do Alentejo, para além deste surgir individualizado do restante Interior do país (evidenciado as diferenças no sistema de povoamento e na importância da actividade agrícola) as diferenças permitem individualizar 4 grandes grupos:

- A Lezíria e o Médio Tejo (predominância do Grupo 4);
- O Alentejo Central (Grupo 13);
- O sul do Alentejo (Grupo 9);
- O Norte Alentejano em associação ao Alentejo litoral, associação esta que decorre do fraco grau de urbanização destes territórios, visível numa baixa densidade populacional e numa reduzida taxa de urbanização.

No caso do Algarve, os grupos obtidos dão uma hierarquia que responde afirmativamente á organização do território:

- A demarcação da “capitalidade” de Faro e Portimão (Grupo 0);
- O grau de urbanização dos concelhos da faixa litoral (Grupo 18);
- Os restantes.

Não obstante já termos testado a importância da metodologia adoptada para a definição de agrupamentos regionais e para descobrir particularidades territoriais com significância nos padrões comportamentais regionais, ainda cremos que a metodologia pode e deve ser submetida a mais robustos testes, e o seu modelo calibrado por um refinamento das variáveis de *input*, de forma a voltar verificar semelhanças e diferenças nos métodos de análise linear multivariada (ACP e Análise de *Clusters*) e análise não linear (SOFM).

## ANÁLISE LINEAR E NÃO LINEAR DE DADOS

### *Segunda experiência*

#### Análise multivariada e SIG

Para esta segunda fase experimental (tabela 2), com base em apenas 11 variáveis representativas do quadro demográfico, económico e social, vamos iniciar o teste com ACP, com rotação *varimax*.

Tabela 2: Matriz dos *loadings* do ACP

<i>Loadings</i>	Componente		
	1	2	3
Pop2001	<b>0.744618</b>	-0.465908	0.048275
Var_POP	0.131401	<b>-0.883302</b>	0.021180
VAB_Cap00_100	<b>0.847134</b>	-0.246550	-0.005954
IDV_01	-0.148536	<b>0.962598</b>	-0.039568
Var_VAB	0.026892	0.100529	<b>0.797971</b>
TxAnalf01	-0.221066	<b>0.846784</b>	-0.336564
TxEnsup01	<b>0.938331</b>	-0.144491	0.117100
TxEnsSec_01	<b>0.940228</b>	-0.162182	0.056162
%_15_64	0.314800	<b>-0.928271</b>	0.021970
Desemp_01	0.103667	0.341703	<b>-0.832773</b>
Var_91_00	-0.303128	0.061326	<b>-0.828161</b>
Variância explicada	<b>3.327239</b>	<b>3.741213</b>	<b>2.151123</b>
Percentagem do Total	<b>0.302476</b>	<b>0.340110</b>	<b>0.195557</b>

O resultado da análise deu-nos 3 componentes capazes de explicar mais de 84% do total de variância. A primeira componente está relacionada com a dimensão demográfica e económica (População e VAB per capita) em associação com altos níveis educacionais (secundário e universitário); a segunda componente sublinha a dinâmica económica e o baixo nível educacional (as regiões menos dinâmicas são aquelas que denotam maior iliteracia e menor capital activo humano); e a terceira componente mostra-nos a dinâmica económica, revelada pela variação do VAB em oposição ao desemprego e à variação do desemprego.

A análise de *clusters*, dos *loadings* permitem-nos identificar 5 grupos (figura 8):

- **Grande Lisboa** – caracterizada pela dimensão populacional e económica, assim como pelos altos níveis educacionais e dinâmica económica;
- **Alentejo** – caracterizado pelos altos índices de população idosa e pelos indicadores negativos referentes à dinâmica económica e ao elevado desemprego;
- **Região Litoral** – caracterizada pelas dinâmicas demográficas e económicas;
- **Região Interior** – caracterizada pelos indicadores de envelhecimento e fracos índices de literacia;
- **Norte Litoral** – caracterizado por um pequeno subgrupo de indicadores de população jovem (forte no capital activo humano) e baixos índices de educação.

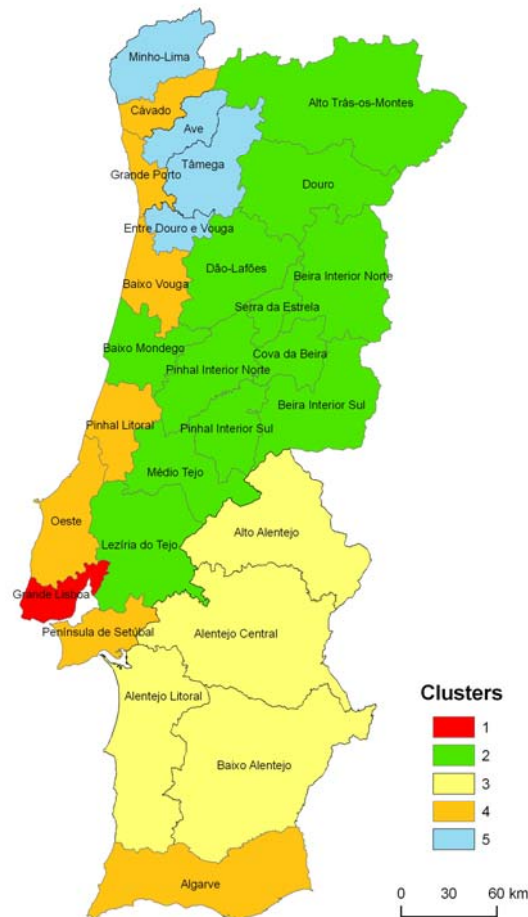


Figura 8: *Clusters* Regionais de Portugal Continental a partir da ACP *Varimax*





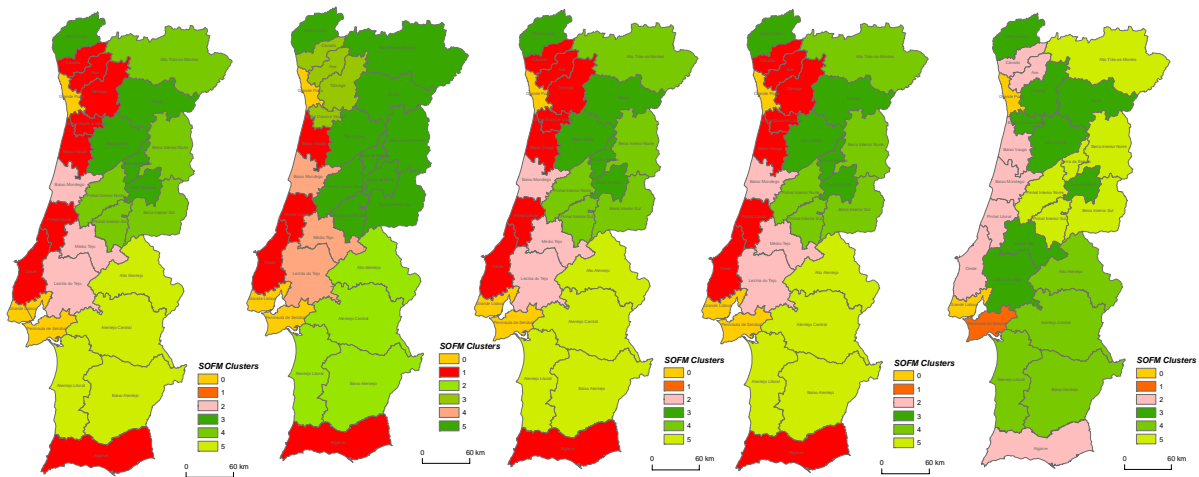


Figura 10: *SOFM Clusters Regionais* com base nas 11 variáveis originais, para Portugal (resultados de várias experiências obtidos por mudanças nos parâmetros específicos)

Várias considerações podem, a partir dos resultados destas experiências, serem feitas (e muitas mais especulações), com base na comparação de metodologias e técnicas de análise adoptadas. Todavia, o que nos interessa mais é apreciação dos resultados em análise devidamente orientados pela premissa retirada das afirmações de Sir Alan Wilson (Wilson, 2000) acerca de procedimentos matemáticos não lineares: “If the equations are nonlinear, it turns out that there are typically very large numbers of solutions” and also that we “...would usually begin with a search for equilibrium solutions”. Foi com esta premissa por base que fomos sempre avançando nas análises experimentais desenvolvidas ao longo deste projecto e cujos resultados obtidos nos permitem agora, com fundamentos científicos afirmar que três conclusões sobressaem claramente:

1. Sub-regiões do Alentejo e a Grande Lisboa representam o *core* territorial de Portugal Continental. Tal facto é nos revelados pela inalteração (equilíbrio) destas regiões à medida que experiências e parâmetros foram sendo feitas e modificados, respectivamente.
2. A região do Algarve sempre, à medida que alterações foram sendo experimentadas, permaneceu no agrupamento com as Sub-regiões do Litoral. Sobretudo, as Regiões do Oeste e do Pinhal Litoral, que são contíguas à Região da Grande Lisboa. Também bastante interessante, e merecedor de nota de relevo, é o facto de que este padrão especial não emergiu aquando da experiência de ‘correr’ o SOFM com base nos três *cores* territoriais da ACP. Nesta experiência, a Região do Algarve aparece no agrupamento composto pela Península de Setúbal e Grande Porto;
3. As Sub-regiões que estão ligadas pelo eixo indirecto Nordeste – Sudoeste são as mais instáveis. Podemos deste facto analítico concluir que estas são muito susceptíveis a perturbações que possam ocorrer, sobretudo nos domínios das dinâmicas económicas e sociais.

Como notas finais, e após este exercício experimental de comparação de métodos de análise linear e não linear, por interpretação dos resultados, podemos afirmar que estes métodos e técnicas, assim como a interacção com os SIG, são fundamentais instrumentos de suporte à decisão e cruciais ao processo de planeamento num mundo caracterizado por transformações constantes. A modelação geográfica e a geografia computacional são recursos técnicos e científicos actuais imprescindíveis aos decisores e planeadores que se defrontam com problemas de complexidade cada vez mais acrescida e aos quais precisam responder encontrando soluções em tempos cada vez mais imediatos.

Os sistemas *top-down* baseados em teorias e metodologias deterministas e reducionistas já não são capazes de responder cabalmente aos problemas contemporâneos. É preciso renovar e adoptar nossas metodologias e técnicas; os sistemas *bottom-up*, heurísticos e baseados na teoria da complexidade, recuperada dos anos 70 e desenvolvida com ajuda dos avanços na tecnologia informática, constituem a

corrente actual de investigação científica, nos vários domínios sociais, humanidades, ou ainda nas chamadas ciências duras.

## **BIBLIOGRAFIA**

- Amari, S., Topographic organization of nerve fields. *Bulletin of Mathematical Biology*, 42, 1980, 339-364.
- Back, B., Sere, K., e Vanharanta, H., Data mining accounting numbers using self-organizing maps. In Alander, J., Honkela, T., and Jakobsson, M., editors, *Proceedings of STEP'96, Finnish Artificial Intelligence Conference*, pages 35-47. Finnish Artificial Intelligence Society, Vaasa, Finland, 1996.
- Bailey, T., Gatrell, A, *Interactive spatial data analysis*, Prentice Hall, England, 1995, pp.409.
- Batty, M., *Network geography in Working papers series*. Centre for advanced spatial analysis. Paper 63. London. 2003. pages 1-3.
- Cherkassky, V. e F. Mulier. 'Self-Organizing Networks for Nonparametric Regression', In Cherkassky et al. pp. 188-212, 1994. (7)
- Estanqueiro, R., *A emergência das redes neuronais artificiais em geografia*. 2005. (por publicar)
- Gilbert, G., *Le Fédéralism Financier*, *Revue Economique*, nº 2, 1996.
- Goodchild, M., *Geographical information science*. *International Journal of Geographical Information Systems* 6(1), 1992, 31–45.
- Haggett, P., *Geography. A global synthesis*. Prentice Hall, England, 2001, pp.400-406.
- Hastie, T. e W. Stuetzle. 'Principal Curves', *Journal of the American Statistical Association*. 84, pp. 502-516, 1989. (6)
- Kaski, S. and Kohonen, T., Winner-take-all networks for physiological models of competitive learning. *Neural Networks*, 7, 1994, 973-984.
- Kohonen, T. et al, *Engineering applications of the self-organizing map*. *Proceedings of the IEEE*, 84:1358-1384, 1996b.
- Kohonen, T. et al., *SOM\_PAK: the self-organizing map program package*. Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland, 1996a.
- Kohonen, T., *Construction of similarity diagrams for phonemes by a self-organizing algorithm*. Report TKK-F-A463, Helsinki University of Technology, Espoo, Finland, 1981.
- Kohonen, T., *Emergence of invariant-feature detectors in self-organization*. In Palaniswami, M., Attikiouzel, Y., Marks II, R. J., Fogel, D., and Fukuda, T., editors, *Computational intelligence. A dynamic system perspective*, pages 17-31. IEEE Press, New York, NY, 1995b.
- Kohonen, T., *Emergence of invariant-feature detectors in the adaptive-subspace self-organizing map*. *Biological Cybernetics*, 75, 1996, 281-291.
- Kohonen, T., *Physiological interpretation of the self-organizing map algorithm*. *Neural Networks*, 6, 1993, 895-905.
- Kohonen, T., *Self-Organization and Associative Memory*. (3rd edition 1989). Springer, Berlin, 1984.
- Kohonen, T., *Self-organized formation of topologically correct feature maps*. *Biological Cybernetics*, 43:-69, 1982..
- Kohonen, T., *Self-Organizing Maps*. Springer, Berlin, 1995c.
- Kohonen, T., *Self-organizing maps: optimization approaches*. In Kohonen, T., Mäkisara, K., Simula, O., and Kangas, J., editors, *Artificial Neural Networks. Proceedings of ICANN'91, International Conference on Artificial Neural Networks*, volume II, pages 981-990, North-Holland, Amsterdam, 1991.
- Kohonen, T., *The adaptive-subspace SOM (ASSOM) and its use for the implementation of invariant feature detection*. In Fogelman-Soulié, F. and Gallinari, P., editors, *Proceedings of ICANN'95, International Conference on Artificial Neural Networks*, volume 1, pages 3-10. EC2 & Cie, Paris, 1995a.
- Kohonen, T., *The Self-Organizing Map*. *Proceedings of the IEEE*, 78, 1990,1464-1480.
- Longley, P.; Goodchild, M. et al., *Geographic information systems ad science*. John Willey & Sons., London, 2001.
- Martin, D., *Geographic information systems. Socioeconomic applications*. 2ª Ed. Routledge, London,

- 1996.
- Mitchell, W. J., e-topia. "URBAN LIFE, JIM – BUT NOT AS WE KNOW IT". MIT Press. London, 2000.
- Nioche, J.P., De l'Évaluation à l'Analyse des Politiques Publiques, Revue Française de Science Politique, vol. 32, n°1, Fevereiro, 1982.
- Orsini, A., "La Problématique Générale de L'Évaluation", Revue d'Économie Régionale et Urbaine, n° 3, 1998.
- Peyrefitte, M., La Prise en Compte des Logiques d'Acteurs dans L'Évaluation d'une Politique Publique Territoriale", Revue d'Économie Régionale et Urbaine, n° 3, 1998.
- Serrano, A. et al., Cidades e Territórios do Conhecimento. Um novo referencial para a competitividade. Edições Sílabo. APOC. Lisboa, 2005, 177pp.
- Serrano-Cinca, C., Self-organizing neural networks for financial diagnosis. To appear in Decision Support Systems, 1996.
- Swindale, N. W., A model for the formation of ocular dominance stripes. Proceedings of the Royal Society of London, B, 208, pages 243-264, 1980
- Willshaw, D. J. e C. von der Malsburg., How patterned neural connections can be set up by self-organization. In Proceedings of the Royal Society London, volume B194, pages 431-445, 1976.
- Wilson, A, Complex spatial systems. The modeling foundations of urban and regional analysis. Pages 1, 41-46. 2000
- Yano, K., GIS an quantitative geography in Geojournal n°52. Kluwer Academic Publishers. Netherlands. 2001. pp.173-180