

Análise Espacial da Distribuição Sísmica e Respetivos Efeitos na Região do Japão para Determinação do Risco e Perigo Sísmico

Tiago G. Mendes

(fc33684@alunos.fc.ul.pt)

RESUMO

A área da sismologia é uma área bastante discutida atualmente devido à capacidade de destruição de um sismo e número de casualidades por ele provocadas. Para ser possível apostar na prevenção, é necessário analisar o perigo sísmico e o risco sísmico, pelo que é particularmente interessante estudar a distribuição espacial e comportamento dos sismos em regiões sistematicamente afetadas pelos mesmos. O presente projeto analisa os sismos ocorridos recentemente no Japão, por se encontrar junto a uma das zonas geologicamente mais ativas do globo.

A análise de sismologia da região do Japão foi efetuada com recurso a técnicas de análise espacial em *softwares* dedicados. No presente projeto esta análise divide-se em cinco componentes distintas: análise de distribuição de padrões pontuais, que permitem detetar zonas de elevada concentração de pontos, análise de áreas e de autocorrelação espacial, que permitem identificar áreas de intensidade sísmica muito ou pouco correlacionadas, regressão espacial, que permite obter modelos que expliquem interação e comportamento de fenómenos entre si, interpolação, que permite determinar superfícies de tendência através de amostras de pontos e análise de redes, que permite fazer uma análise de proximidade de serviços de urgência para atuação na eventualidade de catástrofes. Para análise das componentes acima referidas foram utilizados dados relacionados com sismos específicos (e.g. intensidade, velocidade de propagação das ondas, entre outros) obtidos por estações sísmicas espalhadas pelo Japão assim como dados de localização de sismos ocorridos na zona nos últimos 20 anos.

No presente projeto foi possível comprovar que a região do Japão é uma região bastante afetada por sismos anualmente, com sismos que podem facilmente chegar a magnitudes consideradas destrutivas (>6.0). Para além disso verificou-se a tendência de ocorrência de sismos ao longo da fronteira entre placas tectónicas. Mais especificamente, a zona com maior tendência para ocorrência de sismos encontra-se numa zona de junção de três placas tectónicas. Foi também possível observar os fenómenos que contribuem para a elevada mortalidade associada a um sismo, comprovando-se que esta não se encontra diretamente relacionada com a intensidade do sismo mas sim com o colapso de edificado e inundações.

1. Introdução

Uma das principais motivações no estudo de sismos e de sismologia prende-se com a destruição causada por um sismo de grande magnitude. Para estudar este efeito, é necessário analisar e distinguir dois conceitos: o perigo sísmico, associado à ocorrência natural de sismos, e o risco sísmico, representando as consequências devastadoras de um sismo, tanto na vida dos seres humanos como nos danos de propriedades. O risco sísmico, contrariamente ao perigo sísmico, é afetado por ações do ser humano (por exemplo, zonas com alto perigo sísmico estarem associadas a um baixo risco pela inexistência de seres humanos).

A ocorrência de sismos encontra-se espacialmente bem definida nas zonas de fronteiras de placas tectónicas (ver **Fig. 1**). Estas podem convergir, divergir ou deslizar entre elas e, apesar de serem geralmente estáveis ao longo do tempo, a fronteira entre elas pode sofrer um deslizamento repentino, e a energia acumulada é libertada sob a forma de um sismo. Dado que as fronteiras de placas se estendem por mais de 150,000 km, a ocorrência de sismos é frequente. A **Tabela 1** descreve a frequência de sismos num dado ano.

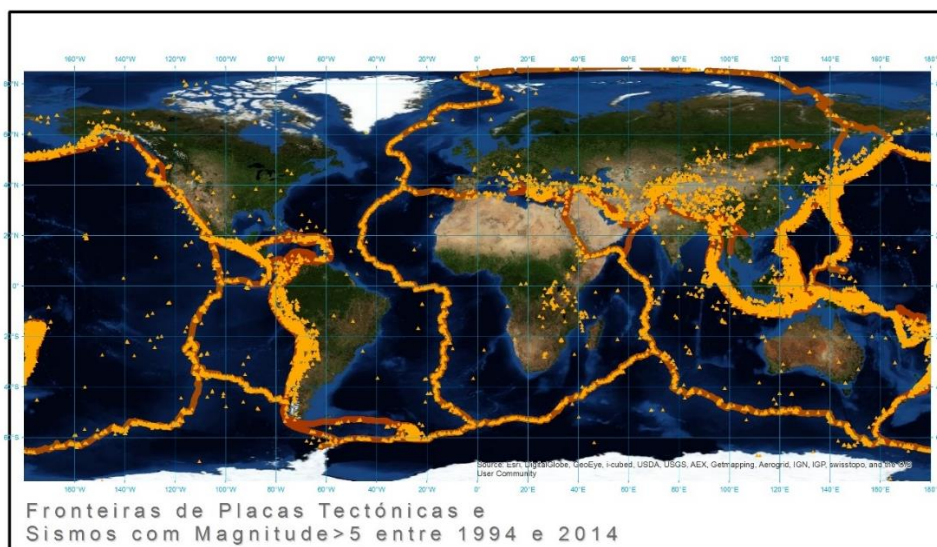


Fig. 1 - Epicentros de sismos com magnitude superior a 5 desde 1994. A laranja: fronteiras de placas tectônicas; A amarelo: epicentros dos sismos

Tabela 1 - Número de sismos por ano [Stein and Wysession, 2003]

Magnitude do Sismo (M _s)	Número de Sismos por Ano	Energia Libertada (10 ¹⁹ J/ano)
>8.0	0-1	0-1,000
7-7.9	12	100
6-6.9	110	30
5-5.9	1,400	5
4-4.9	13,500	1
3-3.9	>100,000	0.2

Como descrito acima, podemos encontrar zonas onde existe um elevado perigo sísmico, cujo risco é reduzido. A 27 de Março de 1964 ocorreu um dos maiores sismos alguma vez registado instrumentalmente, na região do Alasca, com uma magnitude de 9.1. No entanto, apenas 131 pessoas morreram devido ao reduzido número de residentes nesta área. Recentemente, no Haiti, um sismo de magnitude 7.1 causou um número estimado de 316,000 mortes, demonstrando o exemplo de zonas com diferentes níveis de risco sísmico.

O grande motivo para o número de mortes no sismo do Haiti prende-se com o tipo de estruturas existentes. A maioria de mortes relacionadas com sismos resultam do colapso de edifícios, sendo frequente afirmar-se que “sismos não matam pessoas; edifícios matam pessoas” [Stein and Wysession, 2003]. A prevenção do risco sísmico deve ser pensada a nível da engenharia de construção. O sismo de 11 de Março de 2011 no Japão causou um total de 15,884 mortes, apesar de a sua magnitude ser de 9.0. Deste modo se comprova que a engenharia associada ao edificado permite salvar vidas. Estas variáveis serão estudadas na fase da regressão espacial.

A análise espacial constitui uma ferramenta útil na análise de eventos passados de modo a que seja possível prever ocorrências futuras, apoiando na prevenção. No presente projeto serão abordadas diferentes ferramentas que permitem analisar a forma de como os eventos se relacionam, quer espacialmente, quer entre si.

Este projeto foi realizado no âmbito da cadeira de Análise Espacial de Informação Geográfica. Apesar de haver um esforço para garantir a coerência das análises e dos resultados, não se pretende que seja um projeto rigoroso no âmbito da sismologia, sendo mais importante a exploração das ferramentas de análise espacial.

Dada a importância da região do Japão no âmbito do estudo de sismologia e consequentes efeitos, esta foi escolhida para o presente projeto. Além de apresentar uma elevada proximidade com uma zona de fronteira de placas tectónicas, o Japão permite um estudo profundo devido à existência de dados e informação relevantes.

O projeto divide-se em cinco fases distintas:

- Na primeira fase é feita a análise da distribuição espacial de sismos na região do Japão. Nesta fase serão usadas ferramentas de análise de distribuição pontual, densidade pontual e identificação de *clusters*.
- Na segunda fase são analisados os efeitos do sismo de 11 de Março de 2011 em cada região do Japão. Esta análise é feita com base numa rede de estações sísmicas espalhadas pelo país, assim como dados estatísticos das consequências deste sismo. Para este estudo serão utilizadas ferramentas de autocorrelação espacial (índice de Moran e identificação de *Hotspots*).
- Na terceira fase são exploradas ferramentas de regressão espacial, com o intuito de identificar possíveis motivações para a mortalidade associadas à ocorrência de um sismo (edifícios colapsados, intensidade do sismo, etc.).
- A quarta fase consiste na utilização de ferramentas de interpolação para identificação de superfícies de tendência através de uma amostra de dados. Nesta fase é também feita uma interpolação de um DTM SRTM (*Shuttle Radar Topography Mission*) para posterior refinamento da grelha e identificação de zonas de altitude mais baixa, possibilitando a determinação de zonas com mais tendência para sofrerem inundações.
- Por fim, na quinta fase, é feita uma análise de redes para determinar a proximidade de serviços de emergência aos locais afetados por sismos, sendo a análise centrada na região mais afetada pelo sismo de 11 de Março de 2011 – Sendai.

2. Fontes de Informação

Em primeiro lugar foi necessário obter *shapefiles* para apoio visual ao projeto. Foram usadas as seguintes *shapefiles*:

- Dados de batimetria para noção de profundidade e zonas de fronteira de placas.
- Dados de países próximos do Japão para apoio visual.
- Dados de limites administrativos do Japão (regiões).
- Dados de indicação de fronteira de placas.

De seguida procedeu-se à recolha de informação sísmica. Optou-se por recolher informação relativa apenas aos últimos 20 anos, pois é suficiente para uma análise significativa, não representando em si um volume excessivo de dados. Optou-se, também, por apenas obter informação de sismos com magnitude superior a 5.0.

A informação da localização dos epicentros dos sismos encontra-se disponível graças aos centros de sismologia a nível mundial, que mantêm registo destes dados. O *United States Geological Survey* (USGS) [United States Geological Survey, 2014] contém bastante informação de grande utilidade e interesse nesta matéria. Um dos produtos disponíveis é precisamente a localização geográfica dos sismos em conjunto com outros dados. É possível, através desta ferramenta, pesquisar por sismos num dado intervalo de tempo, com um dado intervalo de magnitude dentro de dado retângulo (definido por coordenadas geográficas). Daí é possível obter um ficheiro de texto com informação de latitude, longitude, magnitude, profundidade, etc., necessária para a análise espacial.

Após obtenção destes dados, foi necessário obter registos de intensidades sísmicas sentidas em cada região, num dado sismo. O USGS denomina este tipo de informação de *shake map*. A informação

recolhida vem num ficheiro de texto com informação da latitude, longitude, intensidade sentida, etc. Esta informação é obtida através de estações sísmicas espalhadas pela região.

Os ficheiros de texto foram convertidos para *shapefiles* de pontos no *software* QGIS¹ (pela sua facilidade na conversão). Posteriormente, todas as *shapefiles* foram importadas no ArcGIS² e convertidas do sistema WGS84 (sistema de coordenadas dos ficheiros obtidos) para o sistema JGD2000 (*Japanese Geodetic Grid 2000*) zona 54N. Este era o sistema oficial do Japão até Outubro de 2011, altura em que foi criado o JGD2011 devido à deformação da crosta sofrida pelo sismo de 11 de Março de 2011.

O modelo digital de elevação (MDE) utilizado para interpolação e posterior cálculo de bacias foi obtido através do CSISS (*Center for Spatial Information Science and Systems*)³

Por fim, a rede de estradas do Japão foi obtida no *website* GEOFABRIK⁴, que disponibiliza grandes volumes de dados de mapas OpenStreetMap⁵.

3. Análise da Distribuição Espacial de Padrões Pontuais

A primeira fase do trabalho consiste na análise visual e exploratória dos dados pontuais. Através desta análise é possível detetar e tentar explicar certos padrões ou anomalias nos dados. Em primeiro lugar é efetuada a visualização dos dados com o intuito de identificar e localizar padrões espaciais de interesse, seguindo-se de uma análise exploratória destes dados para quantificar padrões com indicadores estatísticos e, por fim, proceder-se-á à modelação que permite detetar padrões espaciais de interesse.

Esta fase será aplicada à distribuição de sismos na região do Japão.

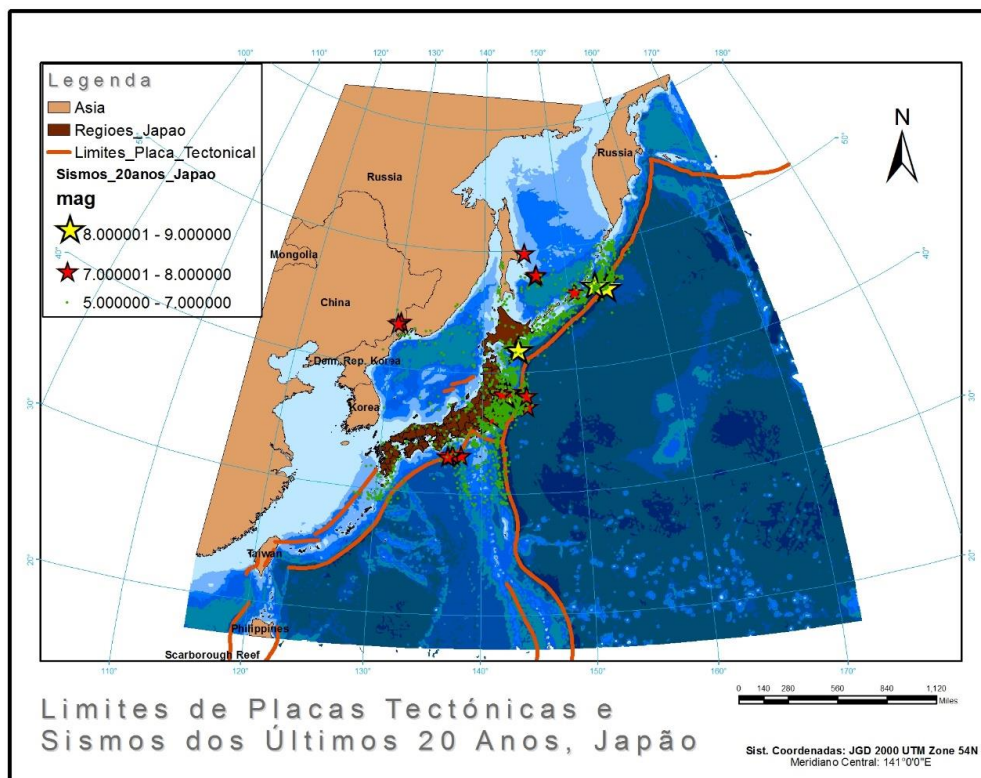


Fig. 2 – Distribuição dos sismos mais recentes na região do Japão; os sismos com magnitude entre 7 e 8 estão representados com estrelas vermelhas e os sismos com magnitude superior a 8 com estrelas amarelas.

¹ www.qgis.org, acessido a 26/06/2014

² <http://www.arcgis.com/>, acessido a 26/06/2014

³ <http://ws.csiss.gmu.edu/DEMExplorer/>, acessido a 15/05/2014

⁴ <http://download.geofabrik.de/asia/japan.html>, acessido a 26/06/2014

⁵ <http://www.openstreetmap.org>, acessido a 26/06/2014

3.1. Visualização

A **Fig. 2** mostra a distribuição de sismos recentes na zona do Japão. Visualmente, a distribuição dos sismos é bastante óbvia, localizando-se na zona de fronteiras de placas tectónicas (descrita pela linha laranja). Outro aspeto que sobressai de imediato, é o facto de a maior densidade de pontos ocorrer junto à zona de fronteira de três placas tectónicas, denominado de junção tripla.

Os próximos passos permitem confirmar com dados concretos as conclusões acima descritas. Foram aplicadas as seguintes operações que permitem descrever padrões de pontos:

- Centro Geométrico (*central feature*) – o ponto mais central dentro do conjunto de dados.
- Média – identifica o centro geográfico para o conjunto de dados
- Mediana – local que representa a menor distância euclidiana a todos os pontos
- Círculo Padrão (*standard distance*) – grau de dispersão dos pontos em torno do centro médio.
- Elipse Padrão (*standard deviational ellipse*) – informação acerca da tendência central, dispersão e direção da distribuição pontual.

Nos casos da elipse padrão e círculo padrão pode ainda ser definido o tamanho da elipse/círculo. Este é dado pelo número de desvios padrão. Neste caso usou-se 1 desvio padrão, contendo cerca de 68% de todos os pontos. A **Fig. 3** mostra os resultados das operações acima descritas.

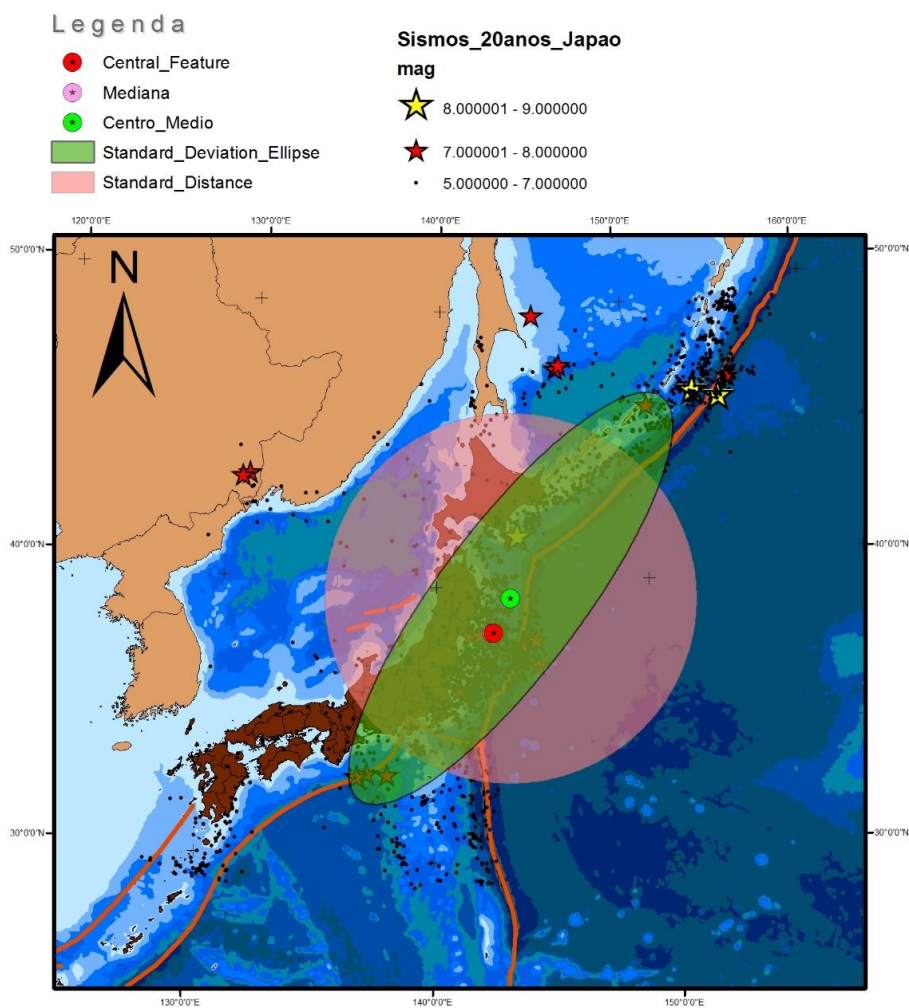


Fig. 3 – Descritores de padrões pontuais (a mediana coincide com o centro geométrico).

Da **Fig. 3** imediatamente se conclui que os pontos estão distribuídos ao longo da linha de junção de placas, sendo que o seu centro se encontra na proximidade do ponto de junção tripla. Esta análise vai de encontro com o que seria expectável em termos teóricos, assim como as conclusões obtidas por análise visual.

Apesar destas operações permitirem uma análise visual imediata e conclusiva, é necessário estudar esta distribuição quanto ao seu padrão, e identificar zonas de maiores aglomerados pontuais.

3.2. Análise Exploratória dos Dados

A análise exploratória dos dados permite-nos concluir acerca do tipo de padrão pontual presente. Neste âmbito é possível analisar efeitos de primeira ordem, ou globais, e eventos de segunda ordem, ou locais.

3.2.1. Contagem por Quadrantes

O método da contagem por quadrantes divide a área que contém os dados numa grelha regular, contando o número de eventos pontuais dentro dela. Esta análise pode ser bastante subjetiva por estar sujeita à definição do tamanho da grelha. Quadrantes grandes não permitem uma boa descrição do padrão e quadrantes demasiado pequenos podem não conter eventos.

Para efetuar o método de contagem por quadrantes foi necessário, em primeiro lugar, criar uma grelha cujos limites correspondiam aos limites dos dados, sendo definido um espaçamento entre quadrantes da grelha. De seguida efetuou-se uma interseção dos dados pontuais dos sismos com a grelha definida. Por forma a obter uma contagem do número de pontos em cada grelha, foi também feito um *summarize* para que houvesse apenas uma instância de cada quadrante contendo a contagem dos pontos.

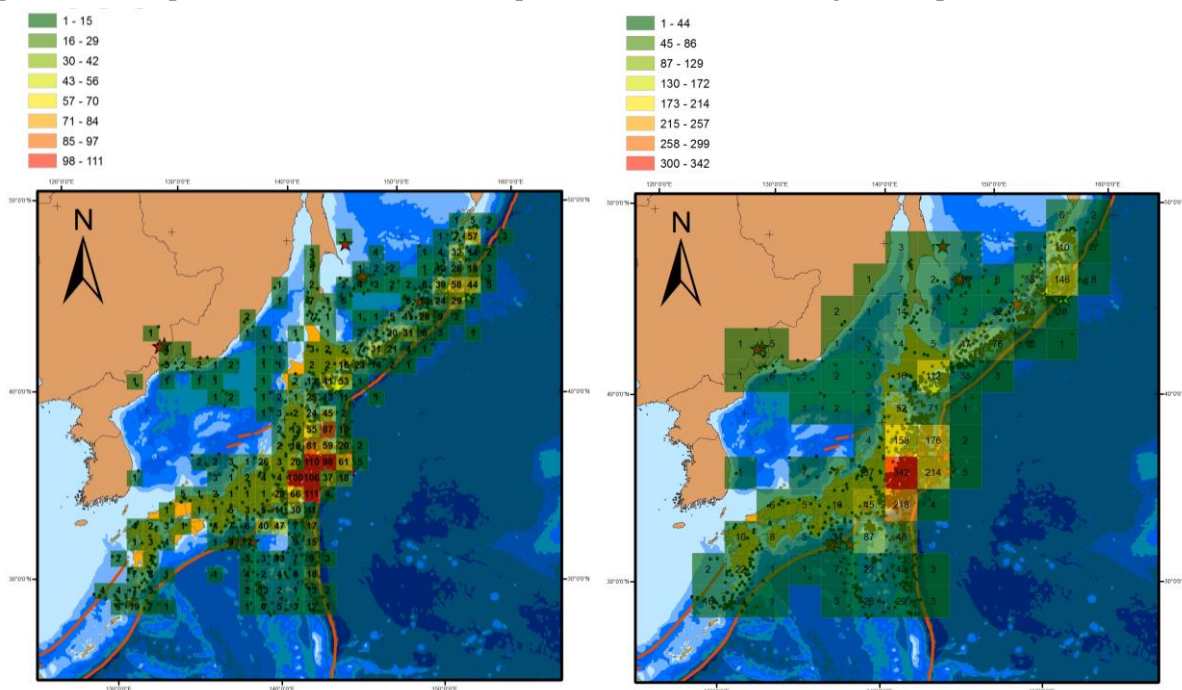


Fig. 4 – Contagem por quadrantes e respetivas estatísticas. Esq.: grelha 100 km × 100 km; Dir.: grelha 200 km × 200 km.

É possível observar na **Fig. 4** que, tal como expectável, a maior concentração de pontos (ou seja, de eventos sísmicos) se encontra junto à fronteira de placas, em particular junto ao ponto triplo. Neste caso, tanto o espaçamento de 200 × 200 km como 100 × 100 km produz resultados semelhantes (embora no primeiro caso estes sejam mais grosseiros).

3.2.2. Estimador de Densidade – *Kernel Density*

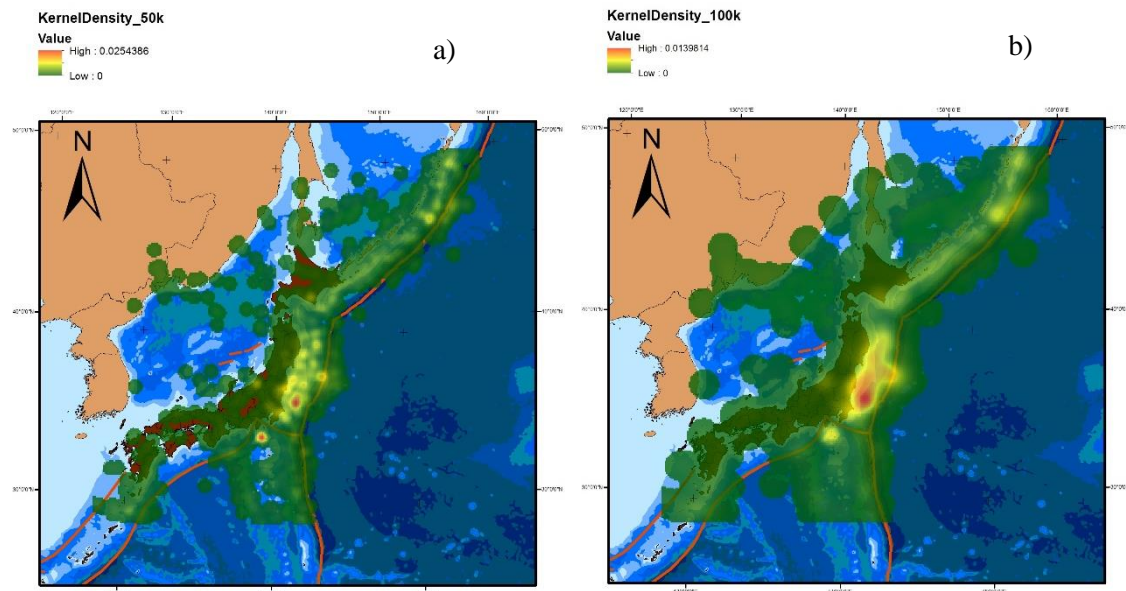
O estimador de densidade explora propriedades de primeira ordem e permite identificar *hot-spots* (zonas de maior densidade pontual). Este consiste em determinar a densidade de pontos numa dada vizinhança, convertendo o conjunto de pontos numa superfície contínua representativa da densidade. Para uma boa

análise, deve ser definida a zona de influência para o estimador, servindo como área para contar todos os pontos em si contidos.

Dado que esta operação identifica zonas com maiores concentrações de pontos, será de esperar que sejam identificados *hot-spots* junto à fronteira de placas tectónicas. Este facto constata-se através da análise da **Fig. 5**.

Foram efetuadas diferentes estimativas de densidade com diferentes áreas de influência. O valor para estas áreas foi determinado empiricamente, através de uma análise preliminar das distâncias entre sismos. Será de esperar que áreas de influência menores resultem numa superfície com pequenas áreas distintas de concentração de pontos, ao passo que áreas de influência maiores tenham tendência para representar uma superfície mais contínua. Este facto permite analisar diferentes características da distribuição de sismos: para valores mais reduzidos, é possível verificar as zonas onde existem maiores concentrações isoladas de pontos. Para áreas maiores é possível observar a tendência de distribuição dos pontos em termos de área.

Na **Fig. 5** verifica-se que ao aumentar a área de influência usada no estimador de densidade, existe uma tendência para agrupar pequenos aglomerados de pontos. De facto, na **Fig. 5d**, o estimador de densidade revela que a distribuição pontual é feita ao longo da linha de fronteira entre placas (existe maior densidade pontual nesta zona). É também notório o facto que os sismos se concentram junto à zona de junção tripla.



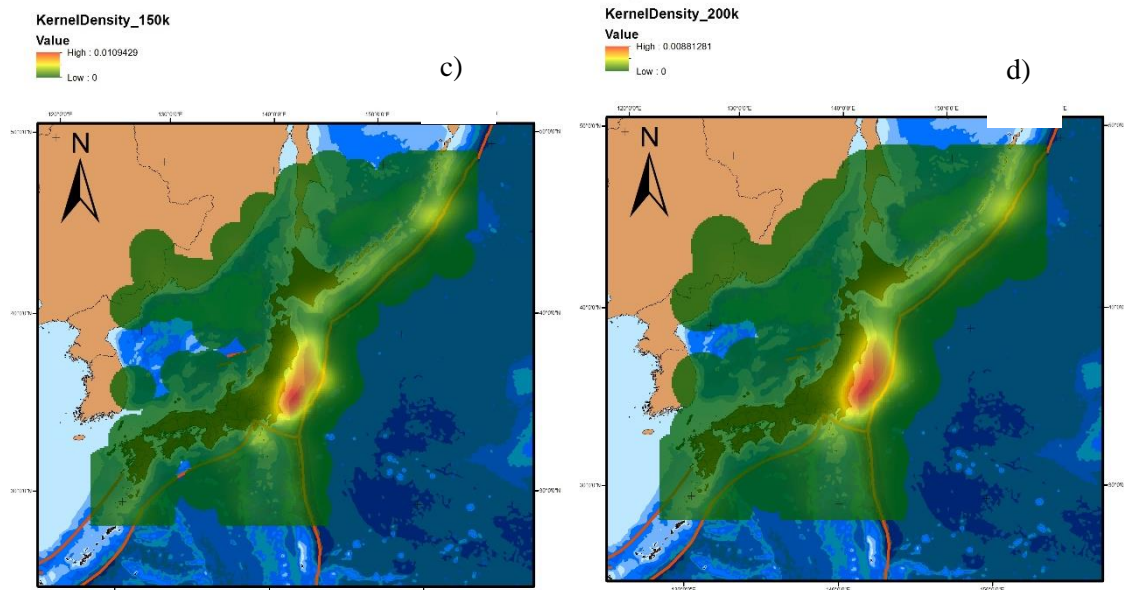
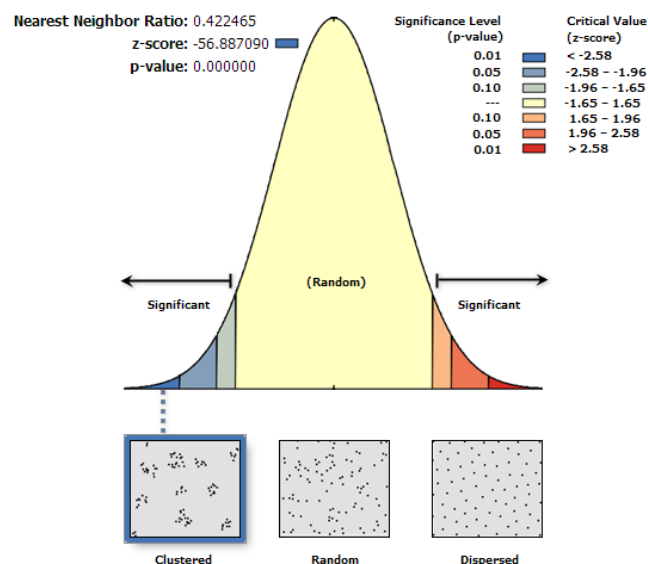


Fig. 5 - Resultados de estimador de densidade com área de influência de a) 50 km, b) 100 km, c) 150 km e d) 200 km

3.2.3. Vizinho Mais Próximo

O método do vizinho mais próximo permite explorar propriedades de 2ª ordem de um processo pontual, sendo este o método mais direto para o fazer. O vizinho mais próximo baseia-se na distância entre eventos do padrão de pontos, medindo a distância entre um dado ponto e o seu ponto mais próximo. Se a média de todas as distâncias de todos os vizinhos mais próximos estiver abaixo de uma distribuição aleatória hipotética, estamos perante um aglomerado, se estiver acima estamos perante dispersão pontual.

Será de esperar, com base na análise já efetuada, que o padrão de distribuição pontual dos dados sísmicos seja de aglomerado. A Fig. 6 comprova este facto.



Given the z-score of -56.8870897562, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Fig. 6 – Resultados da operação do vizinho mais próximo.

A distância média esperada nesta operação era de 22,970 m, sendo que a distância observada foi de 9704 m. Como seria expectável, o resultado do vizinho mais próximo indica um padrão de aglomerado, com menos de 1% de probabilidade deste padrão ser disperso. O *z-score* é uma medida de desvios padrão, estando relacionado com o valor p. Este valor p representa a probabilidade do padrão pontual observado ser resultado de um processo aleatório. A **Tabela 2** explica resumidamente este conceito.

Tabela 2 - Relação entre *z-score*, probabilidade e confiança [ArcGIS, 2014].

Z-score (Desvios Padrão)	p (probabilidade)	Nível de Confiança
<-1.65 ou >1.65	<0.10	90%
<-1.96 ou >1.96	<0.05	95%
<-2.58 ou >2.58	<0.01	99%

Em suma, podemos concluir com 99% de confiança que a distribuição de sismos apresenta um padrão de aglomerado.

3.2.4. Função K

O método do vizinho mais próximo apenas avalia a distância mínima entre dois eventos, podendo ser pouco vantajoso em padrões bastante aglomerados cujas distâncias entre eventos sejam muito pequenas relativamente a outras. A função K permite uma análise mais precisa da distribuição de pontos, especialmente quando se lida com múltiplas distâncias e escalas espaciais.

O passo crítico para obter os melhores resultados da função K consiste em definir corretamente o número de distâncias a avaliar, a distância de partida e os incrementos. O objetivo é encontrar a distância a partir da qual deixa de existir um padrão de aglomerado entre os pontos e passa a existir um padrão disperso.

Neste caso foram feitas várias tentativas, verificando-se que os melhores resultados se obtiveram usando 70 distâncias, começando com uma distância de 100 km em incrementos de 20 km. Para uma primeira abordagem, não foram usados envelopes de simulação. Na **Fig. 7** é possível observar o resultado da aplicação da função K.

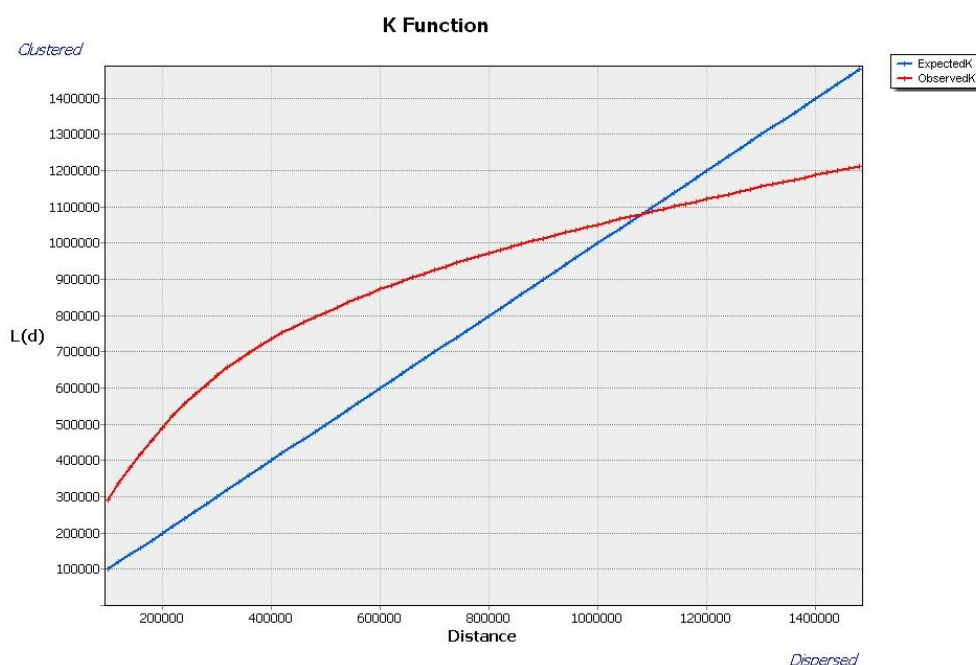


Fig. 7 – Resultado da aplicação da função K sem envelopes de simulação (distâncias em metros).

Quando a linha vermelha (valor observado) se encontra acima da linha azul (valor esperado) significa que existe um padrão de aglomerado. Assim sendo, do gráfico resultante é possível observar que apenas a partir de cerca de 1,100 km é que o padrão pontual tende para a dispersão. Isto significa que existe coerência entre os dois métodos de análise de eventos de segunda ordem, e que os eventos sísmicos apresentam um padrão de aglomerados.

3.3. Modelação do Processo Pontual (*Complete Spatial Randomness*)

O objetivo da modelação do processo pontual consiste em definir um modelo teórico de referência que permita validar a significância estatística dos métodos de exploração pontual.

Em primeiro lugar é necessário definir o modelo de referência, devendo-se de seguida verificar se o padrão pontual se ajusta ao modelo. Esta verificação pode ser feita comparando distribuições teóricas de frequência ou através de gráficos com envelopes de simulação.

O modelo teórico a ser adotado é o da aleatoriedade espacial completa, pressupondo-se que existe igual probabilidade de ocorrência de um evento em qualquer parte da região de estudo (H1) e que existe independência na posição dos eventos (H2).

Assim, para testar a hipótese de aleatoriedade espacial completa é necessário comparar os valores das distribuições de referência com os valores estatísticos obtidos pelos seguintes métodos:

1. Contagem por quadrantes.
2. Vizinho mais próximo.
3. Função K.

3.3.1. Contagem por Quadrantes

Para testar a hipótese de aleatoriedade espacial completa na contagem por quadrantes, é necessário olhar para as estatísticas. O ArcGIS permite obter de forma direta as análises estatísticas de um dado campo na tabela, representadas na **Fig. 8**.

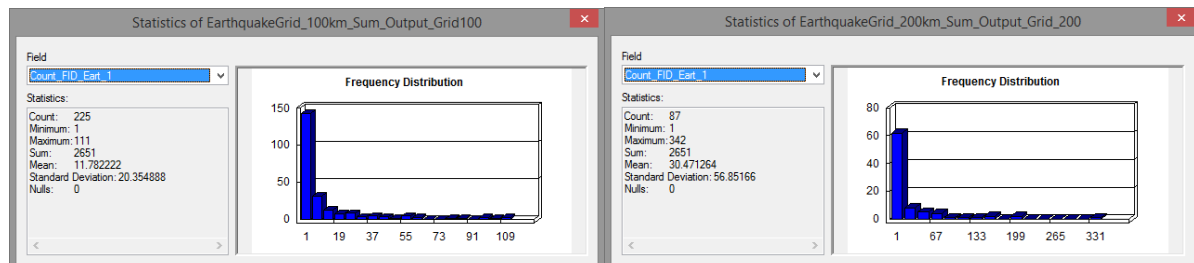


Fig. 8 – Estatísticas do método de contagem por quadrantes. Esq.: Quadrantes 100 × 100 km; Dir.: Quadrantes 200 × 200 km

Para se poder determinar o tipo de padrão pontual existente (aglomerado, disperso ou aleatório), é necessário determinar o rácio variância/média (*variance to mean ratio* – VMTR). Se este valor for superior a 1.0 estamos perante a existência de aglomerados. Se for inferior a 1.0 o padrão é uniforme. Se for igual a 1.0 o padrão é aleatório. Após calcular este valor, é possível obter os valores para o teste estatístico qui-quadrado:

$$\chi^2 = (m - 1) * VMTR \quad (1)$$

Sendo m é o número de quadrantes, sendo $m - 1$ o número de graus de liberdade. A **Tabela 3** contém um sumário das estatísticas fundamentais:

Tabela 3 – Resumo de estatísticas da contagem por quadrantes

Espaçamento de Grelha	m (nº quadrantes)	μ (média)	σ (desvio padrão)	σ^2 (variância)	VTMR	χ^2	$\chi^2_{.99}$	$\chi^2_{.01}$
100 × 100 km	225	11.78	20.35	414.12	35.15	7908.75	177.72	276.16
200 × 200 km	87	30.47	56.85	3231.92	106.07	9228.09	58.46	119.41

O valor VTMR bastante superior a 1.0 indica que existe uma forte presença de padrão de aglomerados nos dados. Este valor confirma o que tinha sido observado visualmente, onde se verificou uma elevada concentração de pontos junto à fronteira de placas tectónicas.

Para testar a hipótese de completa aleatoriedade espacial (H_0) é necessário verificar o valor de qui-quadrado tabelado para 99% de confiança. É possível observar pela tabela que em ambos os casos o valor de χ^2 (7908.75 e 9228.09) é bastante superior ao valor crítico máximo do χ^2_{teste} (276.16 e 119.41, respetivamente). Isto permite rejeitar a hipótese nula (completa aleatoriedade espacial) e afirmar para 99% de confiança que o padrão pontual existente é um padrão de aglomerado.

3.3.2. Vizinho Mais Próximo

Como explicado no ponto 3.2.2, a aleatoriedade espacial é dada pelo *z-score*. A hipótese nula (H_0) equivale a afirmar que o padrão pontual apresenta aleatoriedade espacial completa. Se o valor para o *z-score* se encontrar entre os valores -1.65 e 1.65 (obtidos da tabela de distribuição normal), significa que se aceita a hipótese nula com 90% de confiança.

Dado que o valor *z-score* foi de -56.89, rejeita-se a hipótese nula, podendo-se afirmar com 99% de significância (ver **Tabela 2**) que não se está perante a presença de aleatoriedade espacial completa, mas sim de aglomerados.

3.3.3. Função K

Para analisar a condição de aleatoriedade espacial completa, é necessário obter valores para a função K com envelopes de simulação. Estes envelopes permitem concluir, para um dado nível de confiança, se existe aleatoriedade espacial completa ou não.

O envelope de confiança distribui pontos de forma aleatória na área de estudo, denominado de permutação, calculando a função K para esta distribuição. O número de permutações determina o número de pontos aleatoriamente distribuídos pela área de estudo. Para cada distribuição de pontos, é escolhida para cada distância os valores com maiores desvios (para cima ou para baixo) relativamente aos valores esperados, fixando-se estes valores como intervalo de confiança. As permutações podem, de grosso modo, ser associadas a níveis de confiança:

- 9 permutações equivalem a 90% de confiança.
- 99 permutações equivalem a 99% de confiança.
- 999 permutações equivalem a 99.9% de confiança (não foi possível testar devido à morosidade deste método).

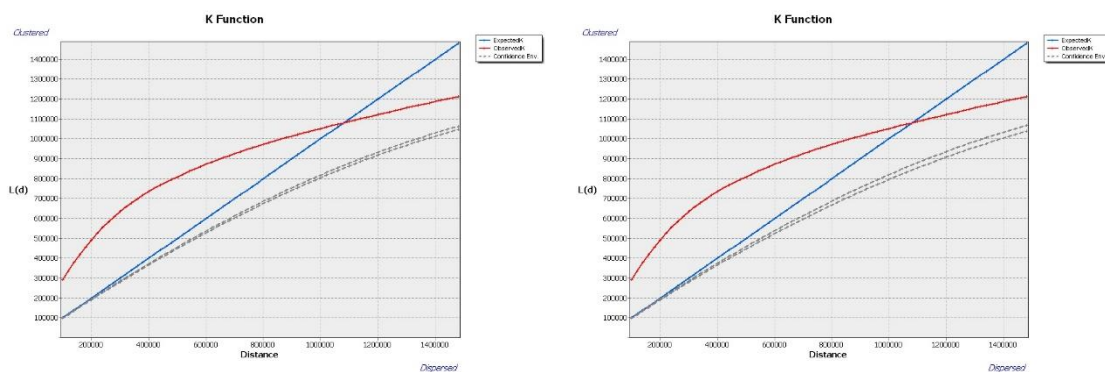


Fig. 9 – Função K com envelopes de simulação para: esq.: 9 permutações, dir.: 99 permutações

Através da **Fig. 9** é possível verificar que o padrão de distribuição pontual é de aglomerado com 90% e 99% de confiança, visto que a reta de valores observados se encontra acima da reta do envelope de confiança.

4. Análise Areal e de Autocorrelação Espacial

Na segunda fase do projeto procede-se à análise dos dados por áreas. Dado que não existem áreas bem definidas relativamente ao epicentro de sismos (inexistência de regiões ou fronteiras), optou-se por analisar as intensidades sentidas em cada região do Japão para determinados sismos. Esta análise é possível devido aos dados existentes e disponíveis das intensidades medidas em várias estações sísmicas no Japão.

Com o estudo por áreas será possível analisar a existência de correlação entre regiões, e detetar possíveis *outliers* nos dados. É também possível determinar se os processos espaciais na região de estudo são ou não estacionários de 1ª ou de 2ª ordem.

Nesta fase é feita, em primeiro lugar, uma visualização dos dados espaciais, seguindo-se de uma análise exploratória dos dados e, por fim, a significância estatística.

4.1. Tratamento dos dados

Por forma a proceder à análise de áreas, foi necessário em primeiro lugar tratar os dados para corresponder à distribuição por regiões. Para este efeito foi necessário intersectar os pontos das estações sísmicas com o mapa das regiões Japonesas. Desta forma, a cada estação sísmica vem associado o nome da região onde se encontra.

O segundo passo consistiu em reduzir o número de estações para que a cada região correspondesse apenas uma estação sísmica. Este passo consistiu em efetuar uma operação de *summarize* no ArcGIS. Para obter um valor único para a intensidade sentida, velocidade máxima de propagação de ondas sísmicas (*peak ground velocity* – PGV) e aceleração máxima de propagação de ondas sísmicas (*peak ground acceleration* – PGA), usou-se o valor médio de todas estas estações. Em situações ideais este estudo torna-se pouco representativo devido aos problemas associados ao uso do valor da média. No entanto, para o estudo envolvido neste projeto, esta abordagem revela-se suficiente.

Por fim, foi apenas necessário efetuar um *join* entre as regiões do Japão e estes valores tabelados, usando o nome da região como chave.

Nesta fase analisou-se a intensidade sísmica, velocidade e aceleração de propagação de ondas sísmicas.

4.2. Visualização dos Dados Espaciais

Através de métodos de corte e gráficos de médias e medianas é possível, de forma semelhante ao que foi efetuado no ponto **3.1**, visualizar de forma imediata como se distribuem os dados na região de estudo.

4.2.1. Visualização por Intervalos Iguais

Este método consiste em definir intervalos mínimo e máximo para verificar a dispersão de dados numa região.

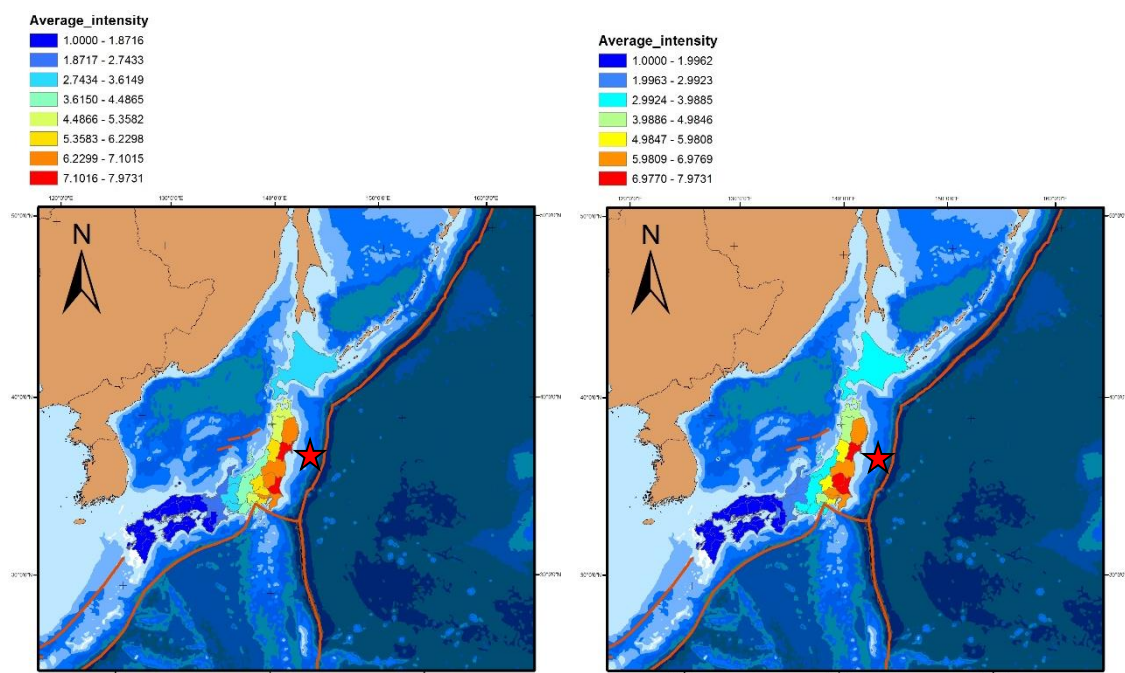


Fig. 10 - Mapa de intensidades sísmicas médias por região. Esq.: sismo de 11 de Março de 2011; Dir.: média de sismos recentes

Ao analisar os mapas de intensidade por intervalos iguais, é possível observar os locais que, no caso do sismo de 11 de Março (**Fig. 10**, esq.), foram mais afetados por esse sismo e, no caso dos sismos mais recentes (**Fig. 10**, dir.) quais os locais que nestes últimos anos têm sofrido mais com sismos.

Relativamente à **Fig. 10**, esq. é óbvio que as zonas onde o sismo foi mais sentido são as zonas que se encontram mais próximo do epicentro (estrela vermelha). No entanto, verifica-se que nos tempos mais recentes, as zonas que têm sido mais afetadas por sismos são as mesmas. Obviamente que o sismo de 11 de Março tem grande influência nas estatísticas dos últimos anos, não só pela grande intensidade do sismo em si, como também pelas réplicas.

É importante também referir que as zonas onde tem havido mais intensidade sísmica encontram-se próximas dos locais onde se detetou maior concentração de pontos (por exemplo, pelo estimador de densidade).

A **Fig. 11** apresenta uma visualização semelhante à **Fig. 10**, neste caso aplicado à intensidade máxima sentida (esq.) e à média da PGV (dir.). As conclusões possíveis de retirar são semelhantes às conclusões já referidas anteriormente. No entanto, é interessante reparar que as zonas onde a velocidade de propagação de ondas sísmicas é maior corresponde à zona onde se sente maior intensidade.

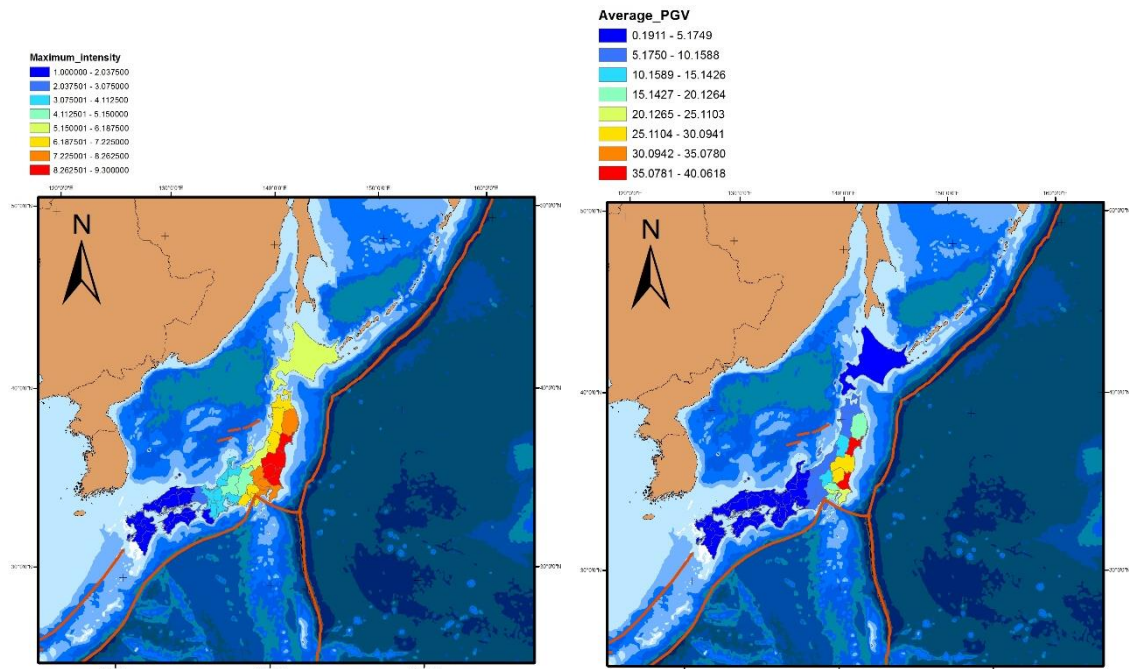


Fig. 11 – Visualização por intervalos iguais. Esq.: Intensidade sísmica máxima por região; Dir.: média da PGV.

4.2.2. Visualização por Percentis

Ao visualizar por percentis é possível ordenar os dados e criar classes para as quais cada grupo contém o mesmo número de elementos. As conclusões que se podem tirar neste caso são semelhantes ao caso anterior, sendo apenas uma forma diferente de visualizar os dados.

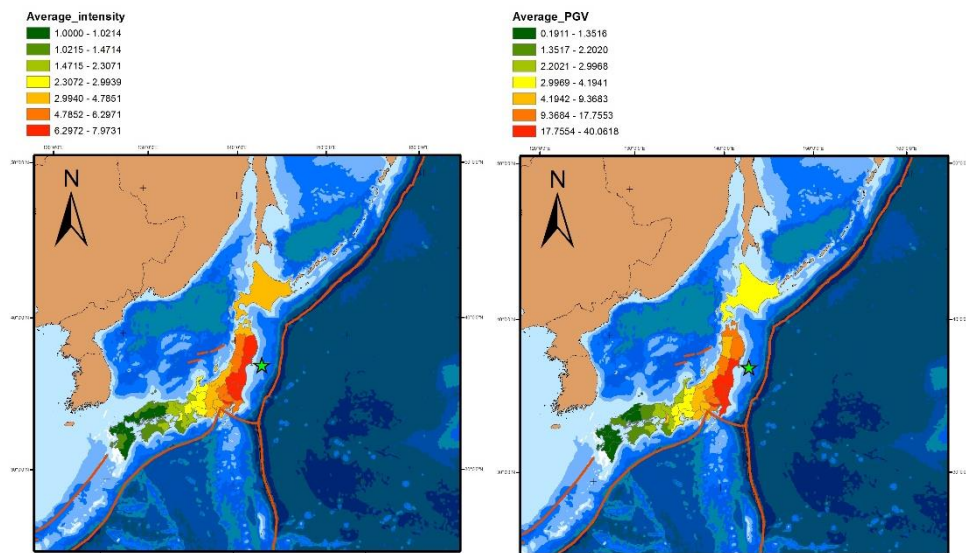


Fig. 12 – Visualização por percentis/quantis. Esq.: Intensidade média dos sismos mais recentes; Dir.: média da PGV dos sismos mais recentes.

4.2.3. Visualização por Desvio Padrão

Este tipo de visualização permite observar diferentes gradações de cores para diferentes valores acima ou abaixo da média, mostrando o comportamento estatístico da variável.

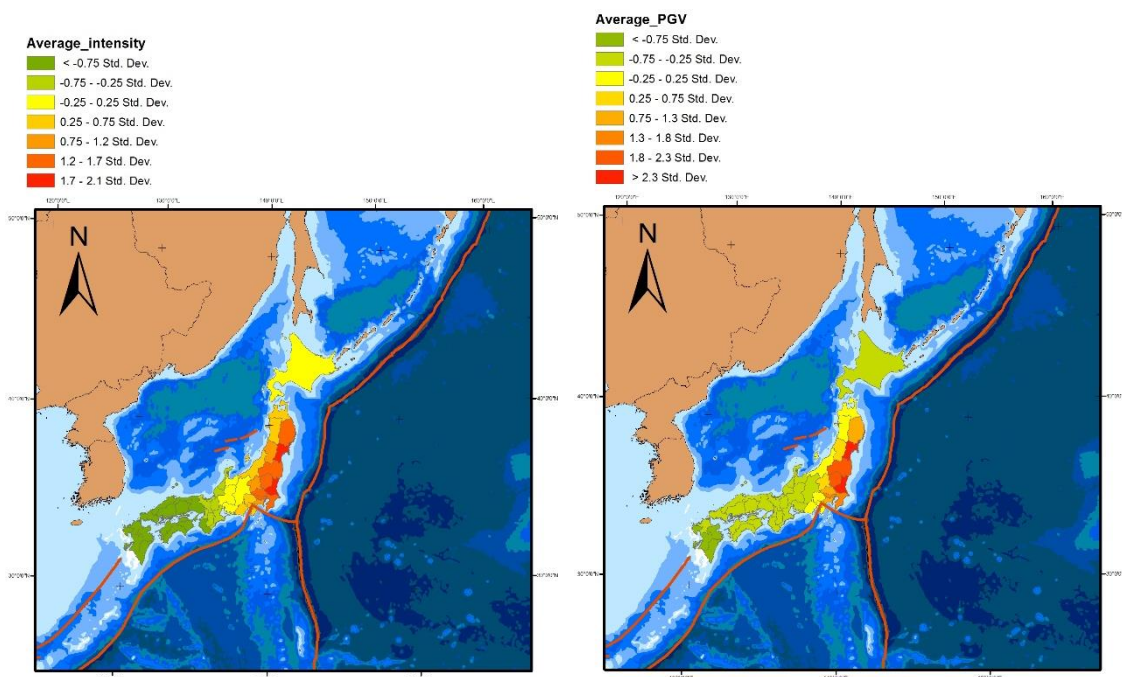


Fig. 13 – Visualização por desvio padrão. Esq.: Intensidade média dos sismos mais recentes; Dir.: Média da PGV nos sismos mais recentes.

Na **Fig. 13** é possível ver os dados já visualizados anteriormente, desta vez por desvio padrão. Neste caso foi definido como linha de corte $\frac{1}{2}$ de desvio padrão.

A visualização dos dados por cortes pode ser bastante útil para obter uma informação visual imediata do comportamento areal das variáveis. No entanto, a sua significância é bastante subjetiva, pois está inteiramente dependente de como são definidos os intervalos. A fase seguinte irá demonstrar como colmatar estes problemas através da análise exploratória.

4.3. Análise Exploratória de Dados Arais

A análise exploratória de áreas permite medir a estrutura de correlação espacial e estimar a magnitude da autocorrelação espacial entre áreas. As estatísticas podem ser divididas em globais ou locais. Nas estatísticas globais pretende-se observar de que forma estão correlacionados os valores no espaço. Nas estatísticas locais o objetivo é identificar *clusters* e *outliers*, através de valores específicos para cada área.

Especificamente, serão analisados os dados por técnicas de correlação e autocorrelação espacial. A correlação espacial consiste numa medida de semelhança entre valores próximos, num conjunto total de amostras. No caso de a semelhança incidir apenas numa variável (medindo o grau de influência sobre si mesma), estamos na presença de autocorrelação.

Se um dado valor influencia outros valores semelhantes na vizinhança, existe autocorrelação positiva. Caso contrário, existe autocorrelação negativa.

4.3.1. Índice de Moran

O índice de Moran é um índice de autocorrelação espacial que permite medir a autocorrelação espacial, comparando o valor de uma dada variável com os valores de outras variáveis em quaisquer localizações. Este considera apenas o primeiro vizinho. Existem várias formas de analisar o índice de Moran. Em primeiro lugar efetuou-se uma análise visual dos resultados, usando o indicador local.

O índice local de Moran é um Indicador Local de Autocorrelação Espacial (*Local Indicators of Spatial Autocorrelation* – LISA). Através do índice local de Moran (Anselin) é possível verificar as zonas onde existe alta correlação entre a variável e si mesma. Na **Fig. 14** é possível verificar que, em qualquer caso existe uma elevada autocorrelação na zona de maior ocorrência de sismos. Esta é representada a

vermelho, indicando que naquelas zonas existem várias regiões onde valores altos estão próximos de valores altos, podendo-se afirmar que as regiões onde se sente maior intensidade sísmica se encontram próximas umas das outras.

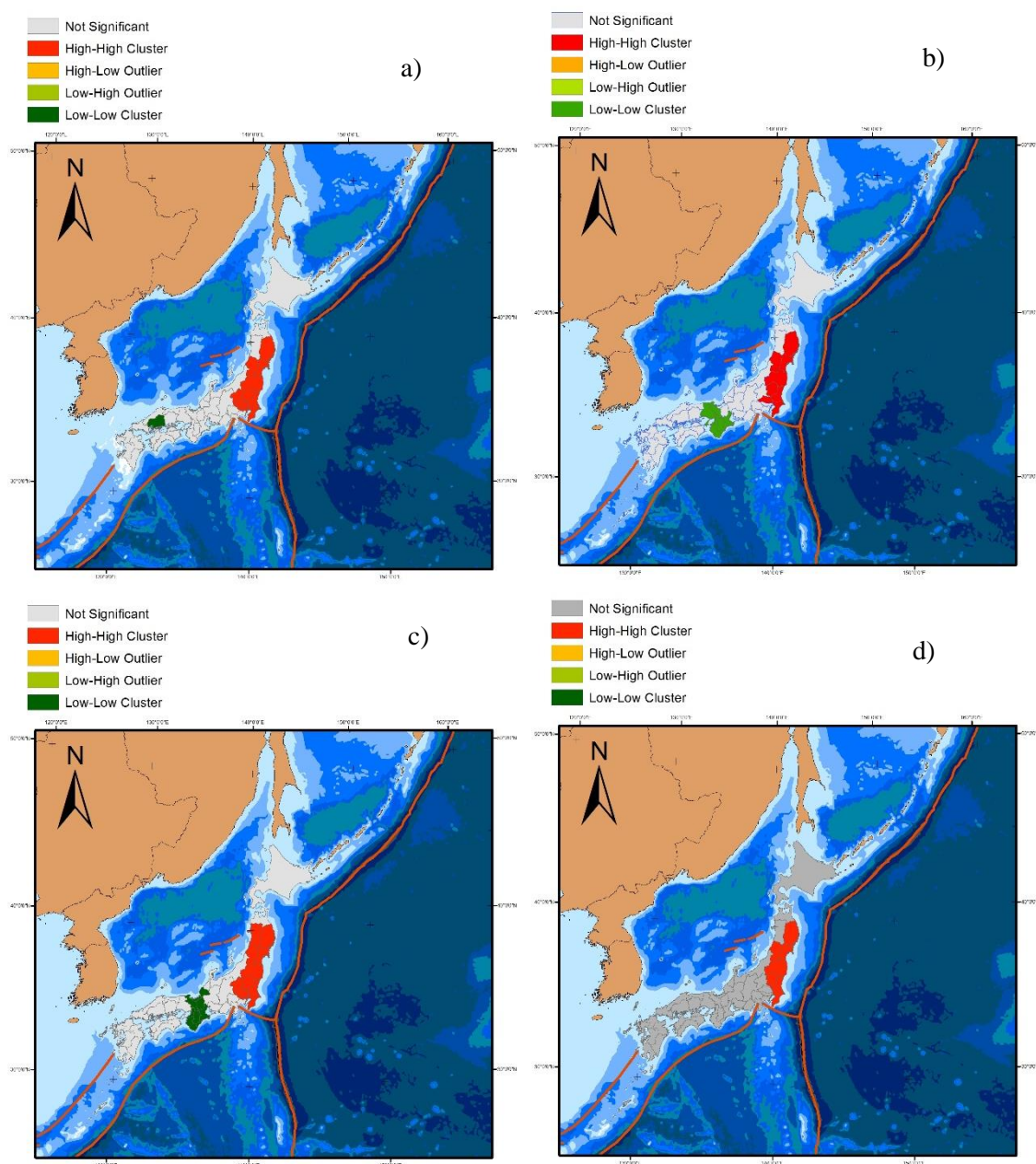


Fig. 14 - Índice Local de Moran. Variáveis: a) intensidades médias do sismo de 11 de Março, b) intensidades médias de sismos recentes, c) intensidades máximas de sismos recentes e d) média da PGV dos sismos recentes.

Da **Fig. 14** também se pode concluir também que existem algumas zonas onde valores baixos se encontram próximos de valores baixos (e, nalguns casos, de valores altos).

O índice de Moran foi também calculado pelo *software* GeoDa⁶, pela sua capacidade de gerar gráficos de significância estatística e de dispersão. Para se calcular o índice de Moran no GeoDa, é necessário em primeiro lugar definir os pesos. Estes são definidos com base na região ou na distância. Neste caso usou-se como parâmetro para a definição da matriz de pesos os limites da região, usando o parâmetro

⁶ <http://geodacenter.asu.edu/> acedido a 26/06/2014

rook contiguity com uma ordem de contiguidade. Estes pesos foram também usados para a regressão espacial (secção 5).

Através do GeoDa é possível comparar os resultados, observando-se uma concordância geral com os resultados apresentados pelo ArcGIS, ainda que com algumas diferenças relativamente aos valores baixos (possivelmente devido à definição da matriz de pesos) (ver Fig. 15).

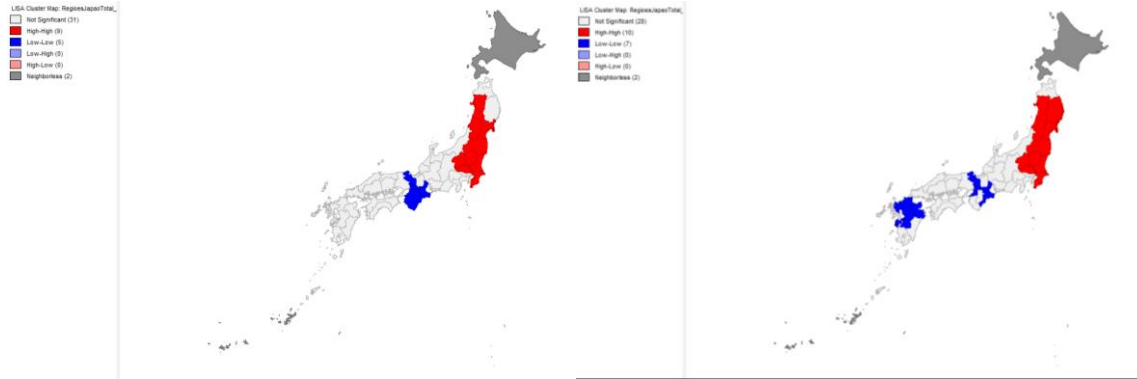


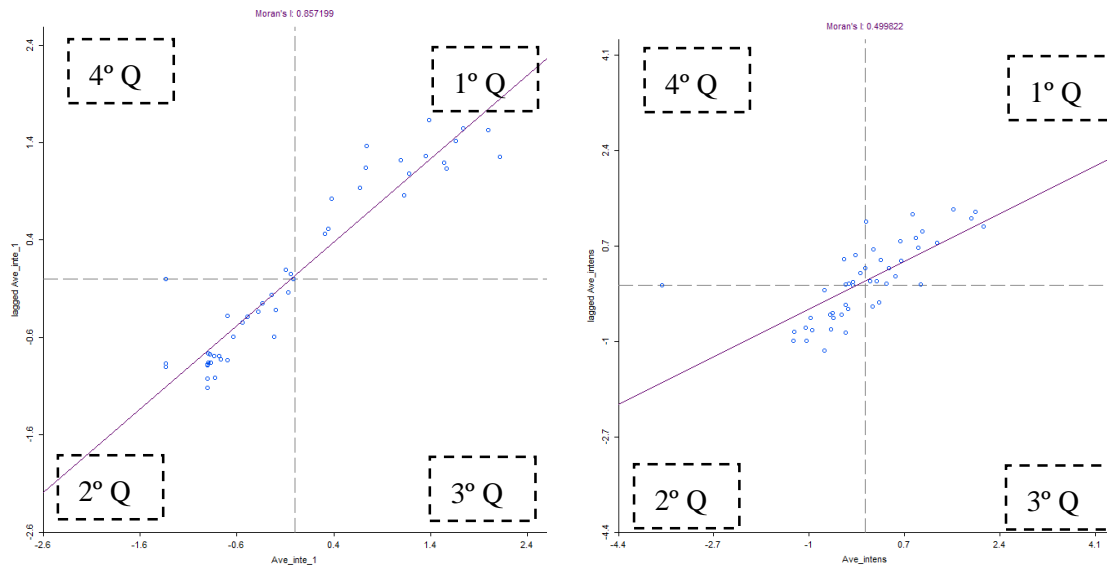
Fig. 15 – Eq. Intensidade média dos sismos recentes; Dir. média da PGV.

Por fim, os gráficos de dispersão de Moran (*Moran scatterplot*) permitem efetuar uma visualização estatística do índice de Moran através da comparação de valores normalizados de um dado atributo numa dada área com a média dos valores normalizados dos vizinhos.

Os gráficos da Fig. 16 permitem observar uma elevada autocorrelação espacial no valor da intensidade média para o sismo de 11 de Março ($I=0.857$). Já para a intensidade média dos sismos recentes e para a média das velocidades máximas de propagação de ondas sísmicas, esta autocorrelação não é tão evidente ($I=0.500$ e $I=0.637$ respetivamente), apesar de ser bastante forte.

Para além do índice, a análise de quadrantes permite identificar locais com autocorrelação positiva (vizinhos com valores semelhantes), localizados no primeiro e segundo quadrante, e locais com autocorrelação negativa, localizados no terceiro e quarto quadrante. Neste caso verifica-se que a grande concentração de valores se encontra no primeiro e segundo quadrante.

Segundo o índice global de Moran, em qualquer uma das variáveis usadas na análise (intensidade média e média das velocidades máximas de ondas sísmicas), existe uma elevada autocorrelação espacial e, consequentemente, padrões de agregados do valor do atributo.



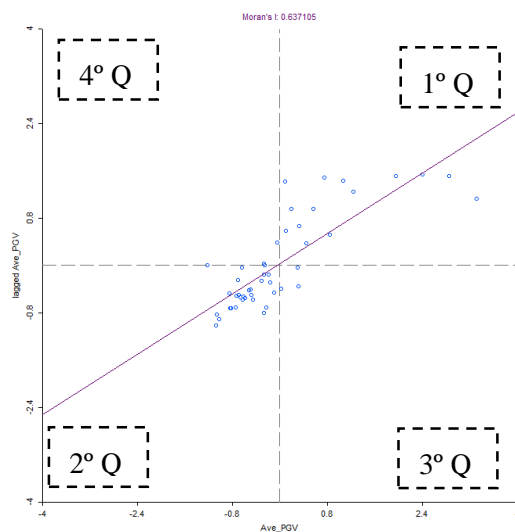


Fig. 16 – Gráficos de dispersão de Moran. Em cima, à esq.: intensidade média do sismo de Honshu. Em cima, à dir.: intensidade média de sismos recentes; em baixo: média da PGV.

4.3.2. Índice G (Getis-Ord) e G_i^* .

De forma semelhante ao índice de Moran, o índice G é um índice global para identificação de autocorrelação espacial. O índice G_i^* é um índice local. Para além do índice G_i^* existe o índice G_i , cuja diferença reside apenas na inclusão ou não da observação i na soma. Ao invés do índice de Moran, que se baseia na covariância entre áreas, o índice G é a soma de valores vizinhos definidos a partir de uma matriz de vizinhança.

O índice G_i^* foi aplicado à intensidade média dos sismos recentes e à média da velocidade máxima de propagação de ondas sísmicas.

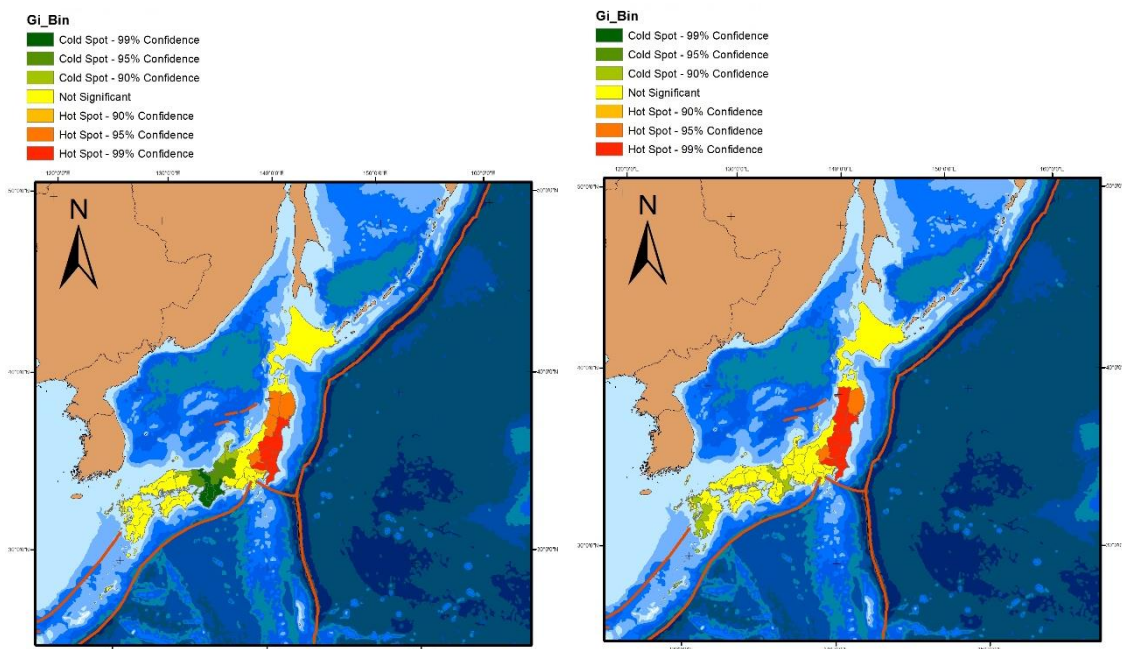


Fig. 17 – Índice G_i^* ; esq.: intensidade média de sismos recentes; dir.: média da PGV.

Dado que o índice G calculado pelo GeoDa traduz resultados semelhantes, o seu resultado não é apresentado no presente projeto.

Da observação da **Fig. 17**, é possível ver que, de forma semelhante ao índice local de Moran, os valores mais elevados se encontram junto à zona de maior concentração de sismos.

4.4. Significância Estatística

Para o índice global de Moran, é possível verificar a significância estatística a partir do teste da pseudo-significância. Este teste usa a técnica da permutação (semelhante ao utilizado para a função K) para criar diferentes arranjos espaciais dos valores, comparando com o índice medido nos dados originais.

Se o valor de i for próximo de -1 ou de 1 estamos perante autocorrelação negativa ou positiva, respetivamente. A figura abaixo apresenta esta técnica usada no *software* GeoDa.

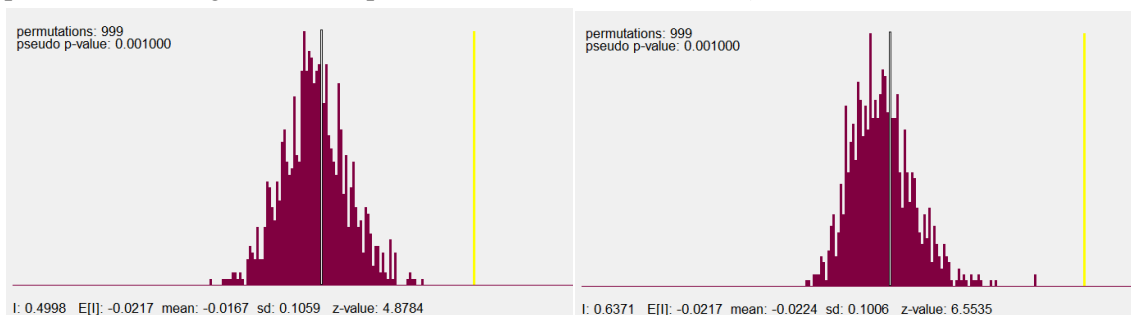


Fig. 18 - Teste da pseudo-significância com 999 permutações; esq.: intensidade média de sismos recentes; dir.: média da PGV.

Analisando a **Fig. 18**, é possível verificar que no primeiro caso (intensidades) o valor do índice encontra-se entre 0 e 1. Com base neste valor, é possível afirmar que existe autocorrelação espacial positiva. No segundo caso, esta verificação é mais evidente, dado o valor do índice de Moran mais próximo de 1.

O índice global de Moran calculado pelo ArcGIS fornece uma indicação da autocorrelação dos dados, assim como a sua significância estatística. Na **Fig. 19** é possível ver o resultado deste índice aplicado às intensidades médias dos sismos recentes e à média das velocidades máximas de propagação das ondas sísmicas.

Da análise dos z -score verifica-se que o z -score em ambos os casos (intensidade e PGV) se encontra acima de 2.58, sendo possível afirmar com 99% de confiança que existe autocorrelação espacial (ver **Tabela 2**).

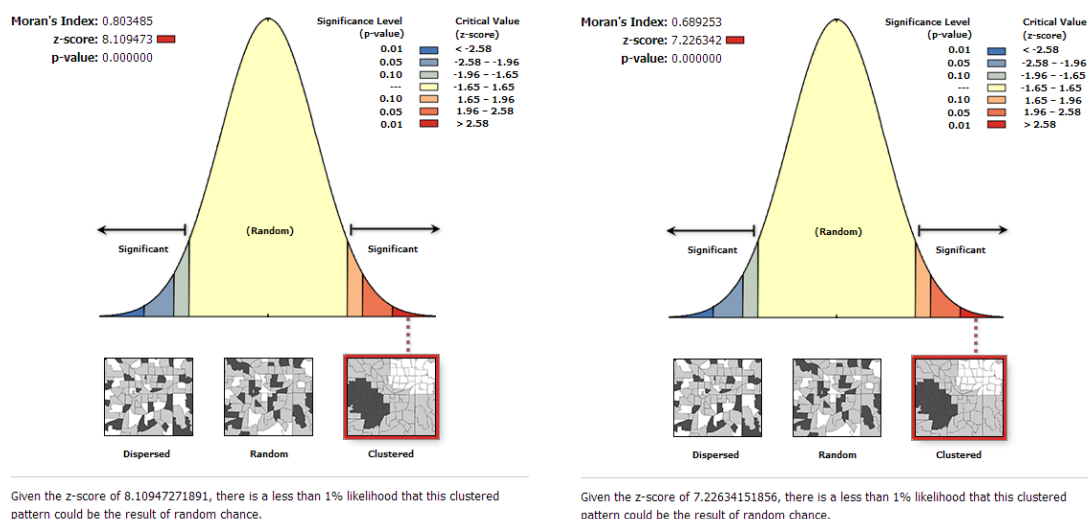


Fig. 19 - Esq. Índice de Moran para intensidade média; Dir. Índice de Moran para a média da PGV.

Para o índice G, o z -score indica que, para 99% de confiança, existe autocorrelação espacial (**Fig. 20**).

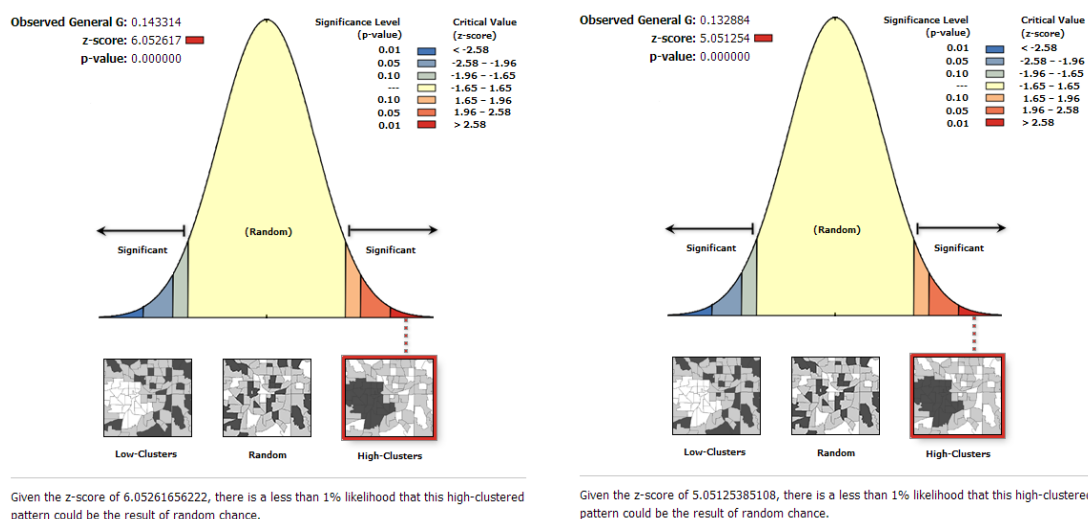


Fig. 20 - Cálculo do índice G para: esq.: intensidades de sismos recentes; dir.: média da PGV.

5. Regressão Espacial

A regressão espacial é uma técnica que permite, estatisticamente, obter um modelo que se ajuste aos valores observados. Desta forma é possível determinar que variáveis contribuem para a relação linear. Neste âmbito é possível analisar regressão espacial simples (apenas uma variável dependente e uma variável independente) e a regressão espacial múltipla (várias variáveis independentes para uma variável independente). Em termos genéricos, a regressão espacial procura encontrar variáveis que contribuam para a ocorrência de outra variável.

Após ser constatada autocorrelação espacial, é possível utilizar outros dois modelos (para além dos modelos de regressão espacial clássica), denominados de *spatial lag models* e *spatial error models*. A diferença entre eles reside no facto de o *spatial lag* atribuir autocorrelação espacial à variável dependente e o *spatial error* atribuir autocorrelação ao erro.

5.1. Diagramas de Dispersão

Por forma a verificar preliminarmente a existência de relação entre variáveis e a sua linearidade, é possível criar diagramas de dispersão entre as variáveis dependentes e independentes. Este constitui o primeiro passo para determinação das variáveis de maior interesse no estudo.

Como explicado na introdução, a queda de edifícios constitui a razão principal para o número de mortos num sismo. Para além deste facto, constatou-se que no sismo de 11 de Março de 2011 95% de mortes ocorreram devido ao *tsunami* que se seguiu ao sismo. Por este motivo procurou-se estudar a relação entre o número de mortos e a queda de edifícios e o número de mortos e áreas inundadas (representados em diagrama de dispersão na **Fig. 21**).

O facto de se verificar um bom ajustamento dos pontos relativamente à reta, permite concluir que estas variáveis são adequadas no estudo da regressão espacial.

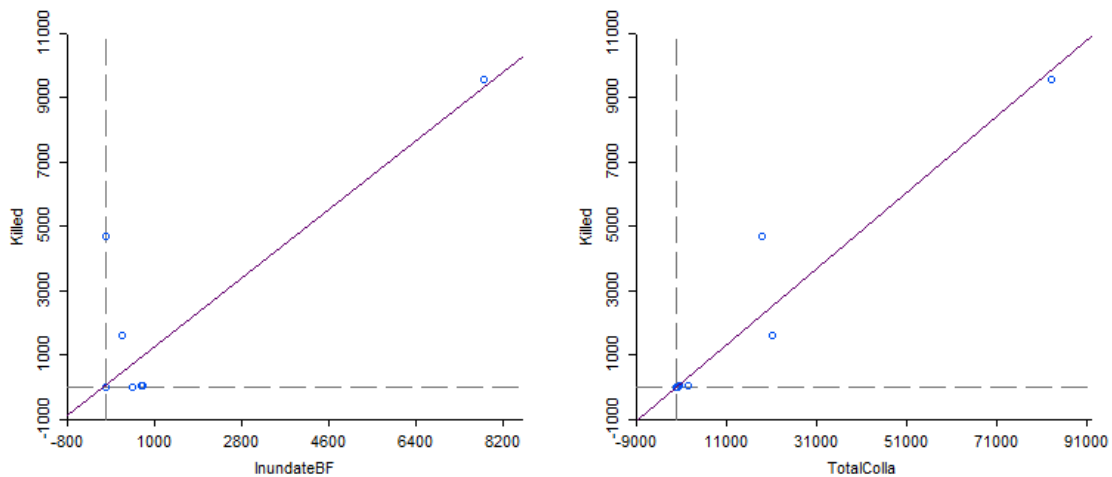


Fig. 21 - Diagramas de dispersão. Esq.: áreas inundadas/número de mortos; dir.: número de edifícios colapsados/número de mortos

5.2. Modelos de Regressão Espacial Clássica

Foram testadas diferentes variáveis dependentes e independentes na procura de modelos de regressão espacial que fossem interessantes de estudar. As questões que originaram a escolha destas variáveis foram as seguintes:

- O número de mortes estará relacionado com o número de edifícios colapsados?
 - Variável dependente: número de mortes.
 - Variável independente: número de edifícios colapsados.
- O número de mortes estará relacionado com as áreas inundadas?
 - Variável dependente: número de mortes.
 - Variáveis independentes: áreas inundadas acima do nível do solo e áreas inundadas abaixo do nível do solo.
- O número de mortes está relacionado com a intensidade sísmica?
 - Variável dependente: número de mortes.
 - Variável independente: intensidade do sismo.
- Os edifícios colapsados estarão relacionados com as velocidades e acelerações de propagação de ondas sísmicas e com a intensidade?
 - Variável dependente: número de edifícios colapsados.
 - Variáveis independentes: velocidade de propagação de ondas sísmicas, aceleração de propagação de ondas sísmicas e intensidade.

Na Fig. 22 são apresentados os resultados obtidos para um modelo de regressão clássica OLS para as variáveis acima descritas. Na análise da qualidade do ajustamento de regressão, é importante analisar o coeficiente de determinação – R^2 . Este valor representa a medida de variação da variável dependente que é explicada pelo modelo. Quando este valor é próximo de 0 estamos perante ausência de correlação e quando se aproxima de 1 aproximamo-nos de correlação perfeita.

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION					SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION				
Data set	: RegioesJapaoTotal				Data set	: RegioesJapaoTotal			
Dependent Variable	: KILLED	Number of Observations	: 47	Dependent Variable	: KILLED	Number of Observations	: 47		
Mean dependent var	: 337.957	Number of Variables	: 2	Mean dependent var	: 337.957	Number of Variables	: 3		
S.D. dependent var	: 1529.89	Degrees of Freedom	: 45	S.D. dependent var	: 1529.89	Degrees of Freedom	: 44		
R-squared	: 0.938267	F-statistic	: 683.941	R-squared	: 0.775170	F-statistic	: 75.8519		
Adjusted R-squared	: 0.936895	Prob(F-statistic)	: 7.46189e-029	Adjusted R-squared	: 0.764951	Prob(F-statistic)	: 5.50525e-015		
Sum squared residual	: 6.79109e+006	Log likelihood	: -345.893	Sum squared residual	: 2.47328e+007	Log likelihood	: -376.267		
Sigma-square	: 150913	Akaike info criterion	: 695.786	Sigma-square	: 562108	Akaike info criterion	: 758.534		
S.E. of regression	: 388.475	Schwarz criterion	: 699.486	S.E. of regression	: 749.739	Schwarz criterion	: 764.085		
Sigma-square ML	: 144491			Sigma-square ML	: 526229				
S.E. of regression ML	: 380.12			S.E. of regression ML	: 725.417				
-----					-----				
Variable	Coefficient	Sgd.Error	t-Statistic	Probability	Variable	Coefficient	Sgd.Error	t-Statistic	Probability
CONSTANT	17.06838	57.97818	0.2943931	0.7698106	CONSTANT	94.41504	113.975	0.8283835	0.4119240
TOTALCOLLA	0.1184725	0.004530103	26.15227	0.0000000	INUNDATEDA	-0.2265701	0.3646762	-0.6212911	0.5376138
					INUNDATEBF	1.194554	0.09703694	12.3103	0.0000000
-----					-----				
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION					SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION				
Data set	: RegioesJapaoTotal				Data set	: RegioesJapaoTotal			
Dependent Variable	: KILLED	Number of Observations	: 47	Dependent Variable	: TOTALCOLLA	Number of Observations	: 47		
Mean dependent var	: 337.957	Number of Variables	: 2	Mean dependent var	: 2708.55	Number of Variables	: 4		
S.D. dependent var	: 1529.89	Degrees of Freedom	: 45	S.D. dependent var	: 12508.5	Degrees of Freedom	: 43		
R-squared	: 0.138330	F-statistic	: 7.22416	R-squared	: 0.725611	F-statistic	: 37.904		
Adjusted R-squared	: 0.119182	Prob(F-statistic)	: 0.0100466	Adjusted R-squared	: 0.706468	Prob(F-statistic)	: 3.84566e-012		
Sum squared residual	: 9.47895e+007	Log likelihood	: -407.84	Sum squared residual	: 2.01779e+009	Log likelihood	: -479.706		
Sigma-square	: 2.10643e+006	Akaike info criterion	: 819.68	Sigma-square	: 4.69255e+007	Akaike info criterion	: 967.411		
S.E. of regression	: 1451.36	Schwarz criterion	: 823.381	S.E. of regression	: 6850.22	Schwarz criterion	: 974.812		
Sigma-square ML	: 2.0168e+006			Sigma-square ML	: 4.29318e+007				
S.E. of regression ML	: 1420.14			S.E. of regression ML	: 6552.24				
-----					-----				
Variable	Coefficient	Sgd.Error	t-Statistic	Probability	Variable	Coefficient	Sgd.Error	t-Statistic	Probability
CONSTANT	-481.3435	371.1276	-1.296976	0.2012503	CONSTANT	1755.477	1996.004	0.8794957	0.3840211
MAX_PGA_1	200.3494	74.54085	2.687779	0.0100466	MAX_PGA_1	311.8214	53.63952	5.813276	0.0000007
					MAX_PGV_1	-256.9697	181.9601	-1.412231	0.1650786
					MAX_INTE_1	-593.0198	772.0788	-0.7680819	0.4466364

Fig. 22 - Resultados de regressão espacial. a) Nº mortes/Nº edifícios colapsados; b) Nº mortes/Nº áreas inundadas; c) Nº de mortes/Intensidade Máxima; d) Nº edifícios colapsados/Intensidade Máxima, PGV e PGA.

Os resultados acima apresentados demonstram que existe uma forte correlação entre as diversas variáveis estudadas (exceto no caso c), onde a correlação é bastante reduzida). Em particular, no caso do estudo da variação do nº de mortes com o nº de edifícios colapsados, o valor de R^2 é quase 1, indicando uma correlação quase perfeita dos dados. Para as restantes variáveis, apesar de valores mais distantes do valor unitário, existe uma forte correlação entre a variável dependente e as variáveis independentes.

Relativamente às duas primeiras questões colocadas no início do capítulo, o número de mortes está de facto ligado ao número de edifícios colapsados e às áreas inundadas. Este facto permite-nos afirmar que a causa para o número de mortes num sismo está ligada ao número de edifícios colapsados e aos efeitos causados por *tsunamis*.

Para além da alta correlação aqui verificada, o valor de R^2 quando se considera o nº de mortes/intensidade máxima demonstra a consistência destas afirmações ($R^2 = 0.138$). Isto significa que o número de mortes está relacionada com os edifícios colapsados e áreas inundadas, e não com a intensidade sísmica.

Os dados de regressão demonstram também que o número de edifícios colapsados depende da intensidade do sismo e da PGV e PGA. Testes efetuados considerando apenas a intensidade do sismo como variável independente ($R^2 = 0.139$) revelaram que o colapso de edifícios não está relacionado com a intensidade. Conclui-se, portanto, que a velocidade e aceleração de propagação de ondas sísmicas influenciam o número de edifícios colapsados.

Dos dados acima é possível escrever as equações que descrevem a ocorrência da variável dependente em termos das variáveis independentes. A equação geral para um modelo com várias variáveis independentes pode ser escrita da seguinte forma:

$$Y_i = a + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i \quad (2)$$

Onde,

Y_i é o valor da variável dependente na i -ésima observação,

a é o valor da variável dependente quando a variável independente é igual a 0,

β_0, \dots, β_p são os parâmetros ou incógnitas do modelo,

X_{i1}, \dots, X_{ip} são os valores das variáveis independentes na i -ésima observação,

ε_i é o termo de erro aleatório com distribuição normal, média zero e variância constante.

Com base neste modelo e nos resultados da regressão é possível criar as equações de regressão:

$$Y = 17.1 + 0.12X_1 + \varepsilon \quad (3)$$

Para o caso da Fig. 20a, onde X_1 é o número de edifícios colapsados.

$$Y = 94.42 - 0.23X_1 + 1.19X_2 + \varepsilon \quad (4)$$

Para o caso da Fig. 20b, onde X_1 é o número de áreas inundadas acima do nível do solo e X_2 é o número de áreas inundadas abaixo do nível do solo.

$$Y = -481.34 + 200.35X_1 + \varepsilon \quad (5)$$

Para o caso da Fig. 20c, onde X_1 é a intensidade máxima.

$$Y = 1755.48 + 311.82X_1 - 256.97X_2 - 593.02X_3 + \varepsilon \quad (6)$$

Para o caso da Fig. 20b, onde X_1 é o valor máximo de PGA, X_2 é o valor máximo de PGV e X_3 é o valor máximo de intensidade.

Apesar do valor do coeficiente de determinação fornecer uma medida de correlação dos dados, este não é suficiente, tornando-se necessário analisar os resultados estatísticos da regressão, apresentados abaixo.

5.3. Diagnósticos de Resíduos do Modelo

Os relatórios de ajustamento do modelo de regressão devem servir como primeira aproximação para análise dos resíduos do modelo. Estes estão apresentados nas figuras abaixo.

Na Fig. 23 devem ser realçados os seguintes valores:

- Condição de Multicolinearidade: este número é indicativo de possíveis problemas na estabilidade do modelo de regressão devido a multicolinearidade (variáveis explicativas demasiado correlacionadas que fornecem informação separada insuficiente). Se este valor se encontrar acima de 30 pode revelar problemas [Anselin, 2005]. Como em qualquer caso este valor é reduzido, este tipo de problemas é inexistente.
- Teste de Jarque-Bera: este é um teste de normalidade dos erros, distribuído como estatísticas qui-quadrado com dois graus de liberdade. Os valores de 0 em qualquer caso sugerem que os erros não assumem uma distribuição normal.
- Detetores de Heterocedasticidade: a heterocedasticidade é um indicador estatístico que permite detetar variâncias não constantes. Como se pode verificar na Fig. 21, apenas o caso b) apresenta ausência de heterocedasticidade. A ausência de heterocedasticidade pode ser preocupante, pois pode invalidar testes de significância estatística que assumem que as variâncias dos erros não variam devido a efeitos de modelação.
- Diagnóstico de Autocorrelação Espacial: os valores reduzidos para o índice de Moran nos casos a) e b) revelam que existe uma elevada autocorrelação espacial entre as variáveis, comprovando a correlação entre o número de mortes e edifícios colapsados/áreas inundadas.

<p>REGRESSION DIAGNOSTICS</p> <p>MULTICOLLINEARITY CONDITION NUMBER 1.239712</p> <p>TEST ON NORMALITY OF ERRORS</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Jarque-Bera</td> <td>2</td> <td>2093.778</td> <td>0.0000000</td> </tr> </tbody> </table> <p>DIAGNOSTICS FOR HETEROSKEDASTICITY</p> <p>RANDOM COEFFICIENTS</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Breusch-Pagan test</td> <td>1</td> <td>2.059607</td> <td>0.1512492</td> </tr> <tr> <td>Koenker-Bassett test</td> <td>1</td> <td>0.1242643</td> <td>0.7244548</td> </tr> </tbody> </table> <p>SPECIFICATION ROBUST TEST</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>White</td> <td>2</td> <td>27.91031</td> <td>0.0000009</td> </tr> </tbody> </table> <p>DIAGNOSTICS FOR SPATIAL DEPENDENCE</p> <p>FOR WEIGHT MATRIX : RegioesJapaoTotal_Rook1.gal (row-standardized weights)</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>MI/DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Moran's I (error)</td> <td>-0.034688</td> <td>-0.0903663</td> <td>0.9279960</td> </tr> <tr> <td>Lagrange Multiplier (lag)</td> <td>1</td> <td>6.1050305</td> <td>0.0134798</td> </tr> <tr> <td>Robust LM (lag)</td> <td>1</td> <td>7.5759619</td> <td>0.0059152</td> </tr> <tr> <td>Lagrange Multiplier (error)</td> <td>1</td> <td>0.1008008</td> <td>0.7508708</td> </tr> <tr> <td>Robust LM (error)</td> <td>1</td> <td>1.5717321</td> <td>0.2099556</td> </tr> <tr> <td>Lagrange Multiplier (SARMA)</td> <td>2</td> <td>7.6767627</td> <td>0.0215284</td> </tr> </tbody> </table>	TEST	DF	VALUE	PROB	Jarque-Bera	2	2093.778	0.0000000	TEST	DF	VALUE	PROB	Breusch-Pagan test	1	2.059607	0.1512492	Koenker-Bassett test	1	0.1242643	0.7244548	TEST	DF	VALUE	PROB	White	2	27.91031	0.0000009	TEST	MI/DF	VALUE	PROB	Moran's I (error)	-0.034688	-0.0903663	0.9279960	Lagrange Multiplier (lag)	1	6.1050305	0.0134798	Robust LM (lag)	1	7.5759619	0.0059152	Lagrange Multiplier (error)	1	0.1008008	0.7508708	Robust LM (error)	1	1.5717321	0.2099556	Lagrange Multiplier (SARMA)	2	7.6767627	0.0215284	<p>REGRESSION DIAGNOSTICS</p> <p>MULTICOLLINEARITY CONDITION NUMBER 1.345252 (Extreme Multicollinearity)</p> <p>TEST ON NORMALITY OF ERRORS</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Jarque-Bera</td> <td>2</td> <td>2095.857</td> <td>0.0000000</td> </tr> </tbody> </table> <p>DIAGNOSTICS FOR HETEROSKEDASTICITY</p> <p>RANDOM COEFFICIENTS</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Breusch-Pagan test</td> <td>2</td> <td>0.5913953</td> <td>0.7440124</td> </tr> <tr> <td>Koenker-Bassett test</td> <td>2</td> <td>0.03596451</td> <td>0.9821785</td> </tr> </tbody> </table> <p>SPECIFICATION ROBUST TEST</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>White</td> <td>5</td> <td>0.2837595</td> <td>0.9979374</td> </tr> </tbody> </table> <p>DIAGNOSTICS FOR SPATIAL DEPENDENCE</p> <p>FOR WEIGHT MATRIX : RegioesJapaoTotal_Rook1.gal (row-standardized weights)</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>MI/DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Moran's I (error)</td> <td>0.002551</td> <td>0.2635140</td> <td>0.7921546</td> </tr> <tr> <td>Lagrange Multiplier (lag)</td> <td>1</td> <td>13.5831418</td> <td>0.0002282</td> </tr> <tr> <td>Robust LM (lag)</td> <td>1</td> <td>22.4791965</td> <td>0.0000021</td> </tr> <tr> <td>Lagrange Multiplier (error)</td> <td>1</td> <td>0.0005454</td> <td>0.9813685</td> </tr> <tr> <td>Robust LM (error)</td> <td>1</td> <td>8.8966001</td> <td>0.0028570</td> </tr> <tr> <td>Lagrange Multiplier (SARMA)</td> <td>2</td> <td>22.4797419</td> <td>0.0000131</td> </tr> </tbody> </table>	TEST	DF	VALUE	PROB	Jarque-Bera	2	2095.857	0.0000000	TEST	DF	VALUE	PROB	Breusch-Pagan test	2	0.5913953	0.7440124	Koenker-Bassett test	2	0.03596451	0.9821785	TEST	DF	VALUE	PROB	White	5	0.2837595	0.9979374	TEST	MI/DF	VALUE	PROB	Moran's I (error)	0.002551	0.2635140	0.7921546	Lagrange Multiplier (lag)	1	13.5831418	0.0002282	Robust LM (lag)	1	22.4791965	0.0000021	Lagrange Multiplier (error)	1	0.0005454	0.9813685	Robust LM (error)	1	8.8966001	0.0028570	Lagrange Multiplier (SARMA)	2	22.4797419	0.0000131
TEST	DF	VALUE	PROB																																																																																																														
Jarque-Bera	2	2093.778	0.0000000																																																																																																														
TEST	DF	VALUE	PROB																																																																																																														
Breusch-Pagan test	1	2.059607	0.1512492																																																																																																														
Koenker-Bassett test	1	0.1242643	0.7244548																																																																																																														
TEST	DF	VALUE	PROB																																																																																																														
White	2	27.91031	0.0000009																																																																																																														
TEST	MI/DF	VALUE	PROB																																																																																																														
Moran's I (error)	-0.034688	-0.0903663	0.9279960																																																																																																														
Lagrange Multiplier (lag)	1	6.1050305	0.0134798																																																																																																														
Robust LM (lag)	1	7.5759619	0.0059152																																																																																																														
Lagrange Multiplier (error)	1	0.1008008	0.7508708																																																																																																														
Robust LM (error)	1	1.5717321	0.2099556																																																																																																														
Lagrange Multiplier (SARMA)	2	7.6767627	0.0215284																																																																																																														
TEST	DF	VALUE	PROB																																																																																																														
Jarque-Bera	2	2095.857	0.0000000																																																																																																														
TEST	DF	VALUE	PROB																																																																																																														
Breusch-Pagan test	2	0.5913953	0.7440124																																																																																																														
Koenker-Bassett test	2	0.03596451	0.9821785																																																																																																														
TEST	DF	VALUE	PROB																																																																																																														
White	5	0.2837595	0.9979374																																																																																																														
TEST	MI/DF	VALUE	PROB																																																																																																														
Moran's I (error)	0.002551	0.2635140	0.7921546																																																																																																														
Lagrange Multiplier (lag)	1	13.5831418	0.0002282																																																																																																														
Robust LM (lag)	1	22.4791965	0.0000021																																																																																																														
Lagrange Multiplier (error)	1	0.0005454	0.9813685																																																																																																														
Robust LM (error)	1	8.8966001	0.0028570																																																																																																														
Lagrange Multiplier (SARMA)	2	22.4797419	0.0000131																																																																																																														
<p>REGRESSION DIAGNOSTICS</p> <p>MULTICOLLINEARITY CONDITION NUMBER 3.192941</p> <p>TEST ON NORMALITY OF ERRORS</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Jarque-Bera</td> <td>2</td> <td>1007.794</td> <td>0.0000000</td> </tr> </tbody> </table> <p>DIAGNOSTICS FOR HETEROSKEDASTICITY</p> <p>RANDOM COEFFICIENTS</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Breusch-Pagan test</td> <td>1</td> <td>94.30232</td> <td>0.0000000</td> </tr> <tr> <td>Koenker-Bassett test</td> <td>1</td> <td>8.190212</td> <td>0.0042117</td> </tr> </tbody> </table> <p>SPECIFICATION ROBUST TEST</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>White</td> <td>2</td> <td>11.71135</td> <td>0.0028636</td> </tr> </tbody> </table> <p>DIAGNOSTICS FOR SPATIAL DEPENDENCE</p> <p>FOR WEIGHT MATRIX : RegioesJapaoTotal_Rook1.gal (row-standardized weights)</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>MI/DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Moran's I (error)</td> <td>0.177341</td> <td>2.1291960</td> <td>0.0332379</td> </tr> <tr> <td>Lagrange Multiplier (lag)</td> <td>1</td> <td>2.0400475</td> <td>0.1532049</td> </tr> <tr> <td>Robust LM (lag)</td> <td>1</td> <td>1.3385720</td> <td>0.2472860</td> </tr> <tr> <td>Lagrange Multiplier (error)</td> <td>1</td> <td>2.6346767</td> <td>0.1045533</td> </tr> <tr> <td>Robust LM (error)</td> <td>1</td> <td>1.9332012</td> <td>0.1644087</td> </tr> <tr> <td>Lagrange Multiplier (SARMA)</td> <td>2</td> <td>3.9732487</td> <td>0.1371577</td> </tr> </tbody> </table>	TEST	DF	VALUE	PROB	Jarque-Bera	2	1007.794	0.0000000	TEST	DF	VALUE	PROB	Breusch-Pagan test	1	94.30232	0.0000000	Koenker-Bassett test	1	8.190212	0.0042117	TEST	DF	VALUE	PROB	White	2	11.71135	0.0028636	TEST	MI/DF	VALUE	PROB	Moran's I (error)	0.177341	2.1291960	0.0332379	Lagrange Multiplier (lag)	1	2.0400475	0.1532049	Robust LM (lag)	1	1.3385720	0.2472860	Lagrange Multiplier (error)	1	2.6346767	0.1045533	Robust LM (error)	1	1.9332012	0.1644087	Lagrange Multiplier (SARMA)	2	3.9732487	0.1371577	<p>REGRESSION DIAGNOSTICS</p> <p>MULTICOLLINEARITY CONDITION NUMBER 11.670593</p> <p>TEST ON NORMALITY OF ERRORS</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Jarque-Bera</td> <td>2</td> <td>198.8897</td> <td>0.0000000</td> </tr> </tbody> </table> <p>DIAGNOSTICS FOR HETEROSKEDASTICITY</p> <p>RANDOM COEFFICIENTS</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Breusch-Pagan test</td> <td>3</td> <td>183.7146</td> <td>0.0000000</td> </tr> <tr> <td>Koenker-Bassett test</td> <td>3</td> <td>30.87945</td> <td>0.0000009</td> </tr> </tbody> </table> <p>SPECIFICATION ROBUST TEST</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>White</td> <td>9</td> <td>33.75724</td> <td>0.0000985</td> </tr> </tbody> </table> <p>DIAGNOSTICS FOR SPATIAL DEPENDENCE</p> <p>FOR WEIGHT MATRIX : RegioesJapaoTotal_Rook1.gal (row-standardized weights)</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>MI/DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Moran's I (error)</td> <td>0.172878</td> <td>2.1875385</td> <td>0.0287031</td> </tr> <tr> <td>Lagrange Multiplier (lag)</td> <td>1</td> <td>0.0027527</td> <td>0.9581573</td> </tr> <tr> <td>Robust LM (lag)</td> <td>1</td> <td>2.6067820</td> <td>0.1064075</td> </tr> <tr> <td>Lagrange Multiplier (error)</td> <td>1</td> <td>2.5037322</td> <td>0.1135769</td> </tr> <tr> <td>Robust LM (error)</td> <td>1</td> <td>5.1077615</td> <td>0.0238190</td> </tr> <tr> <td>Lagrange Multiplier (SARMA)</td> <td>2</td> <td>5.1105142</td> <td>0.0776723</td> </tr> </tbody> </table>	TEST	DF	VALUE	PROB	Jarque-Bera	2	198.8897	0.0000000	TEST	DF	VALUE	PROB	Breusch-Pagan test	3	183.7146	0.0000000	Koenker-Bassett test	3	30.87945	0.0000009	TEST	DF	VALUE	PROB	White	9	33.75724	0.0000985	TEST	MI/DF	VALUE	PROB	Moran's I (error)	0.172878	2.1875385	0.0287031	Lagrange Multiplier (lag)	1	0.0027527	0.9581573	Robust LM (lag)	1	2.6067820	0.1064075	Lagrange Multiplier (error)	1	2.5037322	0.1135769	Robust LM (error)	1	5.1077615	0.0238190	Lagrange Multiplier (SARMA)	2	5.1105142	0.0776723
TEST	DF	VALUE	PROB																																																																																																														
Jarque-Bera	2	1007.794	0.0000000																																																																																																														
TEST	DF	VALUE	PROB																																																																																																														
Breusch-Pagan test	1	94.30232	0.0000000																																																																																																														
Koenker-Bassett test	1	8.190212	0.0042117																																																																																																														
TEST	DF	VALUE	PROB																																																																																																														
White	2	11.71135	0.0028636																																																																																																														
TEST	MI/DF	VALUE	PROB																																																																																																														
Moran's I (error)	0.177341	2.1291960	0.0332379																																																																																																														
Lagrange Multiplier (lag)	1	2.0400475	0.1532049																																																																																																														
Robust LM (lag)	1	1.3385720	0.2472860																																																																																																														
Lagrange Multiplier (error)	1	2.6346767	0.1045533																																																																																																														
Robust LM (error)	1	1.9332012	0.1644087																																																																																																														
Lagrange Multiplier (SARMA)	2	3.9732487	0.1371577																																																																																																														
TEST	DF	VALUE	PROB																																																																																																														
Jarque-Bera	2	198.8897	0.0000000																																																																																																														
TEST	DF	VALUE	PROB																																																																																																														
Breusch-Pagan test	3	183.7146	0.0000000																																																																																																														
Koenker-Bassett test	3	30.87945	0.0000009																																																																																																														
TEST	DF	VALUE	PROB																																																																																																														
White	9	33.75724	0.0000985																																																																																																														
TEST	MI/DF	VALUE	PROB																																																																																																														
Moran's I (error)	0.172878	2.1875385	0.0287031																																																																																																														
Lagrange Multiplier (lag)	1	0.0027527	0.9581573																																																																																																														
Robust LM (lag)	1	2.6067820	0.1064075																																																																																																														
Lagrange Multiplier (error)	1	2.5037322	0.1135769																																																																																																														
Robust LM (error)	1	5.1077615	0.0238190																																																																																																														
Lagrange Multiplier (SARMA)	2	5.1105142	0.0776723																																																																																																														

Fig. 23 - Resultados estatísticos de regressão espacial. a) N° mortes/N° edifícios colapsados; b) N° mortes/N° áreas inundadas; c) N° de mortes/Intensidade Máxima; d) N° edifícios colapsados/Intensidade Máxima, PGV e PGA.

Para além dos diagnósticos apresentados, é possível analisar graficamente a distribuição de resíduos. Estes podem ser analisados através do índice de Moran e estatística Gi, e os resultados para os dois primeiros modelos (n° mortes/edifícios colapsados e n° mortes/áreas inundadas) são apresentados na Fig. 24.

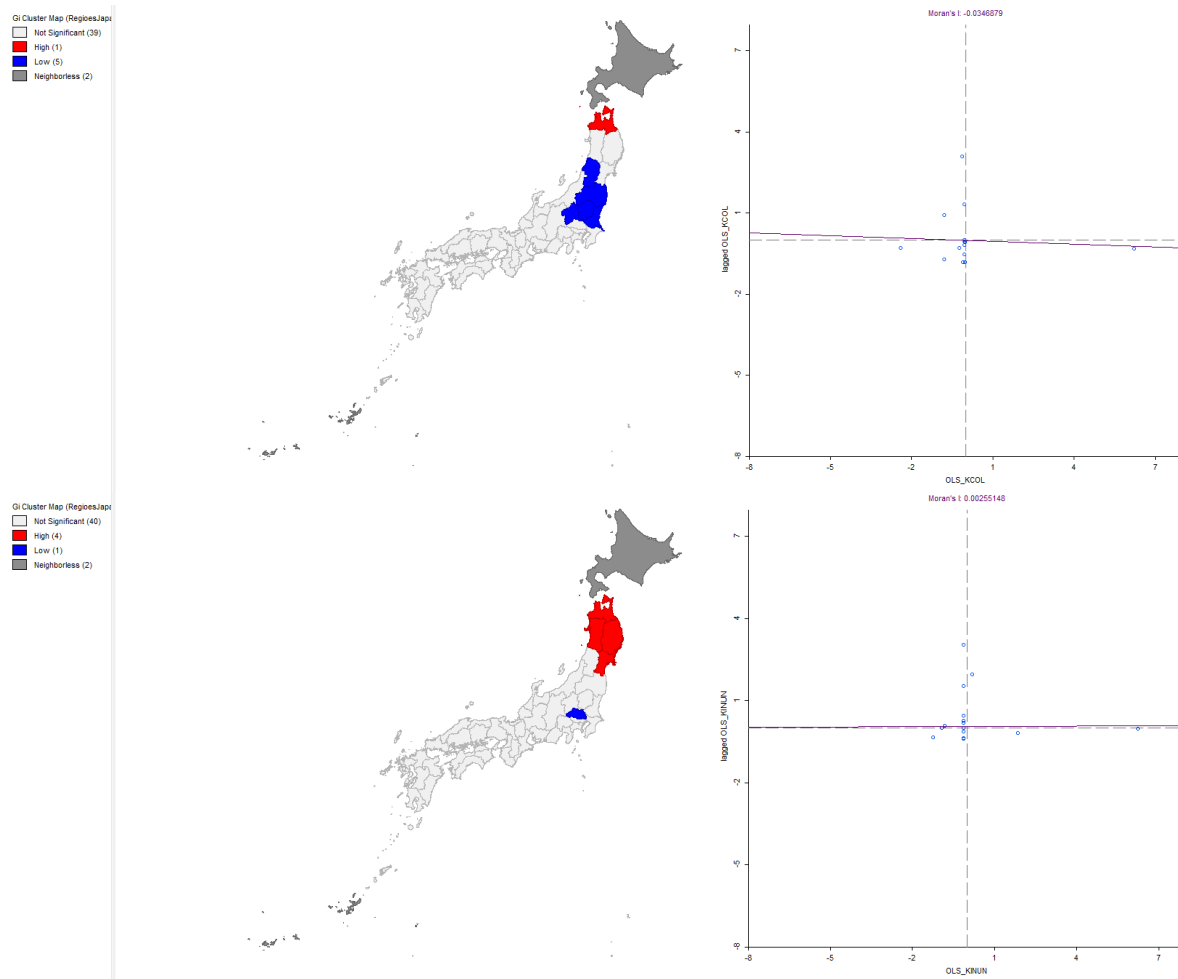


Fig. 24 – Estatística Gi (esq.) e gráfico de dispersão de Moran (dir.). Acima: N° mortos/N° edifícios colapsados; Abaixo: N° mortos/N° áreas inundadas

A **Fig. 24** mostra que existe quase total ausência de correlação espacial entre os resíduos, sendo valor do índice de Moran bastante próximo de 0, indicando que o modelo adotado é adequado em ambos os casos.

5.4. Modelos LAG e Error

Os modelos *Spatial LAG Model* e *Spatial Error Model* permitem verificar se é possível capturar a estrutura de correlação espacial num único parâmetro. Neste caso apenas serão analisados estes modelos aplicados ao n° mortes/n° edifícios colapsados e n° mortes/n° áreas inundadas.

Na **Fig. 25** são apresentados os resultados obtidos em diferentes modelos de regressão. Com base nos dados apresentados é possível escolher o melhor modelo a adotar. Esta escolha deve ser feita com base nos valores de máxima verossimilhança dos diferentes modelos. O logaritmo da máxima verossimilhança constitui a base para outros critérios (como o critério de Akaike ou de Schwarz).

<p>SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION</p> <p>Data set : RegioesJapaoTotal</p> <p>Spatial Weight : RegioesJapaoTotal_Rook1.gal</p> <p>Dependent Variable : KILLED Number of Observations: 47</p> <p>Mean dependent var : 337.957 Number of Variables : 3</p> <p>S.D. dependent var : 1529.89 Degrees of Freedom : 44</p> <p>Lag coeff. (Rho) : 0.160448</p> <p>R-squared : 0.946362 Log likelihood : -342.759</p> <p>Sq. Correlation : - Akaike info criterion : 691.518</p> <p>Sigma-square : 125544 Schwarz criterion : 697.069</p> <p>S.E of regression : 354.322</p>	<p>SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION</p> <p>Data set : RegioesJapaoTotal</p> <p>Spatial Weight : RegioesJapaoTotal_Rook1.gal</p> <p>Dependent Variable : KILLED Number of Observations: 47</p> <p>Mean dependent var : 337.957447 Number of Variables : 2</p> <p>S.D. dependent var : 1529.891833 Degrees of Freedom : 45</p> <p>Lag coeff. (Lambda) : -0.071474</p> <p>R-squared : 0.938502 R-squared (BUSE) : -</p> <p>Sq. Correlation : - Log likelihood : -345.834835</p> <p>Sigma-square : 143939 Akaike info criterion : 695.67</p> <p>S.E of regression : 379.393 Schwarz criterion : 699.37</p>																																																		
<table border="1"> <thead> <tr> <th>Variable</th> <th>Coefficient</th> <th>Std. Error</th> <th>z-value</th> <th>Probability</th> </tr> </thead> <tbody> <tr> <td>W_KILLED</td> <td>0.1604485</td> <td>0.05861263</td> <td>2.737439</td> <td>0.0061921</td> </tr> <tr> <td>CONSTANT</td> <td>-28.54248</td> <td>55.53174</td> <td>-0.513985</td> <td>0.6072625</td> </tr> <tr> <td>TOTALCOLLA</td> <td>0.1145007</td> <td>0.004414072</td> <td>25.93993</td> <td>0.0000000</td> </tr> </tbody> </table>	Variable	Coefficient	Std. Error	z-value	Probability	W_KILLED	0.1604485	0.05861263	2.737439	0.0061921	CONSTANT	-28.54248	55.53174	-0.513985	0.6072625	TOTALCOLLA	0.1145007	0.004414072	25.93993	0.0000000	<table border="1"> <thead> <tr> <th>Variable</th> <th>Coefficient</th> <th>Std. Error</th> <th>z-value</th> <th>Probability</th> </tr> </thead> <tbody> <tr> <td>CONSTANT</td> <td>14.87757</td> <td>53.14912</td> <td>0.2799213</td> <td>0.7795380</td> </tr> <tr> <td>TOTALCOLLA</td> <td>0.1191559</td> <td>0.004365688</td> <td>27.29373</td> <td>0.0000000</td> </tr> <tr> <td>LAMBDA</td> <td>-0.07147402</td> <td>0.1986043</td> <td>-0.3598815</td> <td>0.7189359</td> </tr> </tbody> </table>	Variable	Coefficient	Std. Error	z-value	Probability	CONSTANT	14.87757	53.14912	0.2799213	0.7795380	TOTALCOLLA	0.1191559	0.004365688	27.29373	0.0000000	LAMBDA	-0.07147402	0.1986043	-0.3598815	0.7189359										
Variable	Coefficient	Std. Error	z-value	Probability																																															
W_KILLED	0.1604485	0.05861263	2.737439	0.0061921																																															
CONSTANT	-28.54248	55.53174	-0.513985	0.6072625																																															
TOTALCOLLA	0.1145007	0.004414072	25.93993	0.0000000																																															
Variable	Coefficient	Std. Error	z-value	Probability																																															
CONSTANT	14.87757	53.14912	0.2799213	0.7795380																																															
TOTALCOLLA	0.1191559	0.004365688	27.29373	0.0000000																																															
LAMBDA	-0.07147402	0.1986043	-0.3598815	0.7189359																																															
<p>REGRESSION DIAGNOSTICS</p> <p>DIAGNOSTICS FOR HETEROSKEDASTICITY</p> <p>RANDOM COEFFICIENTS</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Breusch-Pagan test</td> <td>1</td> <td>1.023643</td> <td>0.3116564</td> </tr> </tbody> </table> <p>DIAGNOSTICS FOR SPATIAL DEPENDENCE</p> <p>SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : RegioesJapaoTotal_Rook1.gal</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Likelihood Ratio Test</td> <td>1</td> <td>6.267712</td> <td>0.0122958</td> </tr> </tbody> </table>	TEST	DF	VALUE	PROB	Breusch-Pagan test	1	1.023643	0.3116564	TEST	DF	VALUE	PROB	Likelihood Ratio Test	1	6.267712	0.0122958	<p>REGRESSION DIAGNOSTICS</p> <p>DIAGNOSTICS FOR HETEROSKEDASTICITY</p> <p>RANDOM COEFFICIENTS</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Breusch-Pagan test</td> <td>1</td> <td>2.317949</td> <td>0.1278885</td> </tr> </tbody> </table> <p>DIAGNOSTICS FOR SPATIAL DEPENDENCE</p> <p>SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : RegioesJapaoTotal_Rook1.gal</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Likelihood Ratio Test</td> <td>1</td> <td>0.1163212</td> <td>0.7330592</td> </tr> </tbody> </table>	TEST	DF	VALUE	PROB	Breusch-Pagan test	1	2.317949	0.1278885	TEST	DF	VALUE	PROB	Likelihood Ratio Test	1	0.1163212	0.7330592																		
TEST	DF	VALUE	PROB																																																
Breusch-Pagan test	1	1.023643	0.3116564																																																
TEST	DF	VALUE	PROB																																																
Likelihood Ratio Test	1	6.267712	0.0122958																																																
TEST	DF	VALUE	PROB																																																
Breusch-Pagan test	1	2.317949	0.1278885																																																
TEST	DF	VALUE	PROB																																																
Likelihood Ratio Test	1	0.1163212	0.7330592																																																
<p>SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION</p> <p>Data set : RegioesJapaoTotal</p> <p>Spatial Weight : RegioesJapaoTotal_Rook1.gal</p> <p>Dependent Variable : KILLED Number of Observations: 47</p> <p>Mean dependent var : 337.957 Number of Variables : 4</p> <p>S.D. dependent var : 1529.89 Degrees of Freedom : 43</p> <p>Lag coeff. (Rho) : 0.394372</p> <p>R-squared : 0.840003 Log likelihood : -369.397</p> <p>Sq. Correlation : - Akaike info criterion : 746.794</p> <p>Sigma-square : 374485 Schwarz criterion : 754.195</p> <p>S.E of regression : 611.952</p>	<p>SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION</p> <p>Data set : RegioesJapaoTotal</p> <p>Spatial Weight : RegioesJapaoTotal_Rook1.gal</p> <p>Dependent Variable : KILLED Number of Observations: 47</p> <p>Mean dependent var : 337.957447 Number of Variables : 3</p> <p>S.D. dependent var : 1529.891833 Degrees of Freedom : 44</p> <p>Lag coeff. (Lambda) : 0.010219</p> <p>R-squared : 0.775183 R-squared (BUSE) : -</p> <p>Sq. Correlation : - Log likelihood : -376.266562</p> <p>Sigma-square : 526201 Akaike info criterion : 758.533</p> <p>S.E of regression : 725.397 Schwarz criterion : 764.084</p>																																																		
<table border="1"> <thead> <tr> <th>Variable</th> <th>Coefficient</th> <th>Std. Error</th> <th>z-value</th> <th>Probability</th> </tr> </thead> <tbody> <tr> <td>W_KILLED</td> <td>0.3943724</td> <td>0.08829211</td> <td>4.466677</td> <td>0.0000080</td> </tr> <tr> <td>CONSTANT</td> <td>-24.07856</td> <td>95.91841</td> <td>-0.2510317</td> <td>0.8017897</td> </tr> <tr> <td>INUNDATEDA</td> <td>-0.3235724</td> <td>0.2994673</td> <td>-1.080494</td> <td>0.2799225</td> </tr> <tr> <td>INUNDATEBF</td> <td>1.134119</td> <td>0.07970491</td> <td>14.22898</td> <td>0.0000000</td> </tr> </tbody> </table>	Variable	Coefficient	Std. Error	z-value	Probability	W_KILLED	0.3943724	0.08829211	4.466677	0.0000080	CONSTANT	-24.07856	95.91841	-0.2510317	0.8017897	INUNDATEDA	-0.3235724	0.2994673	-1.080494	0.2799225	INUNDATEBF	1.134119	0.07970491	14.22898	0.0000000	<table border="1"> <thead> <tr> <th>Variable</th> <th>Coefficient</th> <th>Std. Error</th> <th>z-value</th> <th>Probability</th> </tr> </thead> <tbody> <tr> <td>CONSTANT</td> <td>94.54697</td> <td>111.3001</td> <td>0.8494779</td> <td>0.3956153</td> </tr> <tr> <td>INUNDATEDA</td> <td>-0.2236539</td> <td>0.3534021</td> <td>-0.6328595</td> <td>0.5268252</td> </tr> <tr> <td>INUNDATEBF</td> <td>1.190514</td> <td>0.09386125</td> <td>12.68376</td> <td>0.0000000</td> </tr> <tr> <td>LAMBDA</td> <td>0.01021892</td> <td>0.1940869</td> <td>0.05265127</td> <td>0.9580096</td> </tr> </tbody> </table>	Variable	Coefficient	Std. Error	z-value	Probability	CONSTANT	94.54697	111.3001	0.8494779	0.3956153	INUNDATEDA	-0.2236539	0.3534021	-0.6328595	0.5268252	INUNDATEBF	1.190514	0.09386125	12.68376	0.0000000	LAMBDA	0.01021892	0.1940869	0.05265127	0.9580096
Variable	Coefficient	Std. Error	z-value	Probability																																															
W_KILLED	0.3943724	0.08829211	4.466677	0.0000080																																															
CONSTANT	-24.07856	95.91841	-0.2510317	0.8017897																																															
INUNDATEDA	-0.3235724	0.2994673	-1.080494	0.2799225																																															
INUNDATEBF	1.134119	0.07970491	14.22898	0.0000000																																															
Variable	Coefficient	Std. Error	z-value	Probability																																															
CONSTANT	94.54697	111.3001	0.8494779	0.3956153																																															
INUNDATEDA	-0.2236539	0.3534021	-0.6328595	0.5268252																																															
INUNDATEBF	1.190514	0.09386125	12.68376	0.0000000																																															
LAMBDA	0.01021892	0.1940869	0.05265127	0.9580096																																															
<p>REGRESSION DIAGNOSTICS</p> <p>DIAGNOSTICS FOR HETEROSKEDASTICITY</p> <p>RANDOM COEFFICIENTS</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Breusch-Pagan test</td> <td>2</td> <td>0.4863429</td> <td>0.7841371</td> </tr> </tbody> </table> <p>DIAGNOSTICS FOR SPATIAL DEPENDENCE</p> <p>SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : RegioesJapaoTotal_Rook1.gal</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Likelihood Ratio Test</td> <td>1</td> <td>13.74035</td> <td>0.0002099</td> </tr> </tbody> </table>	TEST	DF	VALUE	PROB	Breusch-Pagan test	2	0.4863429	0.7841371	TEST	DF	VALUE	PROB	Likelihood Ratio Test	1	13.74035	0.0002099	<p>REGRESSION DIAGNOSTICS</p> <p>DIAGNOSTICS FOR HETEROSKEDASTICITY</p> <p>RANDOM COEFFICIENTS</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Breusch-Pagan test</td> <td>2</td> <td>0.5855562</td> <td>0.7461877</td> </tr> </tbody> </table> <p>DIAGNOSTICS FOR SPATIAL DEPENDENCE</p> <p>SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : RegioesJapaoTotal_Rook1.gal</p> <table border="1"> <thead> <tr> <th>TEST</th> <th>DF</th> <th>VALUE</th> <th>PROB</th> </tr> </thead> <tbody> <tr> <td>Likelihood Ratio Test</td> <td>1</td> <td>0.001229466</td> <td>0.9720289</td> </tr> </tbody> </table>	TEST	DF	VALUE	PROB	Breusch-Pagan test	2	0.5855562	0.7461877	TEST	DF	VALUE	PROB	Likelihood Ratio Test	1	0.001229466	0.9720289																		
TEST	DF	VALUE	PROB																																																
Breusch-Pagan test	2	0.4863429	0.7841371																																																
TEST	DF	VALUE	PROB																																																
Likelihood Ratio Test	1	13.74035	0.0002099																																																
TEST	DF	VALUE	PROB																																																
Breusch-Pagan test	2	0.5855562	0.7461877																																																
TEST	DF	VALUE	PROB																																																
Likelihood Ratio Test	1	0.001229466	0.9720289																																																

Fig. 25 – *Spatial LAG Model* (esq.) e *Spatial Error Model* (dir.). Acima: nº mortos/edifícios colapsados; abaixo: nº mortos/áreas inundadas.

A Tabela 4 apresenta os diferentes valores para os diferentes métodos. Desta tabela conclui-se que o melhor modelo de regressão a adotar é aquele que possui o menor valor de critério de Akaike. Em ambos os casos corresponde ao modelo *spatial lag*, que é também o que apresenta o maior coeficiente de determinação (R^2).

Por fim, conclui-se que o número de mortes do sismo de 11 de Março de 2011 ao largo da costa do Japão pode ser explicado pelo número de edifícios colapsados e pelo número de áreas inundadas (ou seja, devido ao *tsunami* que se seguiu.

Tabela 4 – Comparação dos vários métodos de regressão.

Variáveis	Modelo	R ²	Log Máxima Verossimilhança	Critério de Akaike	Critério de Shwarz
Nº de Mortes/ Nº de Edifícios Colapsados	OLS	0.94	-345.89	695.79	699.49
	LAG	0.95	-342.76	691.52	697.07
	ERROR	0.94	-345.83	695.67	699.37
Nº de Mortes/ Nº de Áreas Inundadas	OLS	0.78	-376.27	758.53	764.09
	LAG	0.84	-369.40	746.79	754.20
	ERROR	0.78	-376.27	758.53	764.08

6. Interpolação

6.1. Conceitos Teóricos

A interpolação permite estimar valores desconhecidos de pontos da amostra através de valores conhecidos na mesma área ou região [Catita, 2014]. Na prática isto significa que através da amostra de valores de intensidades sísmicas foi feita uma interpolação para estimar valores de intensidades em regiões onde estes são desconhecidos.

A interpolação pode ser classificada quanto à sua extensão espacial, que pode ser global ou local, o seu ajustamento, que pode ser exato ou aproximado, e o seu modelo, que pode ser determinístico ou probabilístico.

Relativamente à extensão espacial, os métodos globais utilizam todos os pontos da amostra, determinando uma função única aplicada por toda a região, e, tendencialmente, resultam em superfícies suavizadas. Como exemplo de interpoladores globais utilizados no projeto podemos referir a interpolação polinomial. Para os métodos locais, a interpolação é aplicada a um pequeno conjunto de pontos vizinhos entre si, pelo que capturam variações locais. Em termos práticos os interpoladores globais são utilizados quando se pretende captar grandes comprimentos de onda (tendências) e os interpoladores locais são utilizados para pequenos comprimentos de onda (variabilidades locais). Os interpoladores locais utilizados no projeto foram o IDW (*Inverse Distance Weighting*), *Splines* e *Kriging*.

Relativamente ao ajustamento, nos métodos exatos o valor estimado pelo interpolador é idêntico ao valor observado na localização. Embora isto represente a situação ideal, a utilização de interpoladores exatos pode ser afetada negativamente por possíveis *outliers* nos dados. Para os métodos aproximados o valor estimado pelo interpolador é diferente do valor observado na localização. Assim, é possível criar uma superfície mais suavizada evitando picos abruptos.

Quanto ao método, os interpoladores podem ser determinísticos ou probabilísticos. Os interpoladores determinísticos ajustam a superfície ao conjunto de valores pontuais, com a premissa que pontos mais próximos se assemelham mais que pontos distantes. Contudo, estes interpoladores não consideram a correlação espacial entre valores e características particulares da sua distribuição espacial. Os métodos probabilísticos permitem modelar a variabilidade espacial de um fenómeno através da análise de autocorrelação espacial.

A Tabela 5 mostra os diferentes métodos de interpolação utilizados e a sua classificação.

Tabela 5 – Métodos de interpolação e a sua classificação. Adaptado de [Catita, 2014].

Extensão Espacial	Global		Local	
	Determinístico	Probabilístico	Determinístico	Probabilístico
Interpolador/Ajustamento	Superfície de Tendência/Polinomial Inexato	Regressão Múltipla Inexato	Polígonos de Thiessen Exato Estimação de Densidade Inexato IDW Exato <i>Splines</i> Exato	Kriging Exato

No presente projeto foram utilizados os seguintes interpoladores:

- Polinomial,
- Polígonos de Thiessen,
- IDW,
- *Spline*

6.2. Interpolação Polinomial

Como referido no ponto 6.1 o interpolador polinomial é um interpolador global, determinístico e inexato. Esta interpolação gera uma superfície suavizada que se ajusta aos pontos da amostra através de um polinómio. A característica global deste interpolador indica que é capaz de capturar os grandes comprimentos de onda da superfície, ou seja, tem uma variação gradual e suave.

O polinómio para uma superfície de primeira ordem é dado por:

$$Z_{xy} = b_0 + b_1x + b_2y \quad (7)$$

No *software* ArcGIS foi utilizado o algoritmo para criar a superfície de tendência aplicada às intensidades registadas no sismo de 11 de Março de 2011, utilizando uma interpolação de segunda ordem e uma interpolação ordem 5, sendo o resultado apresentado na **Fig. 26**.

A superfície de tendência revela a distribuição de intensidades associadas ao sismo acima referido. É possível confirmar, uma vez mais, que as zonas de valores mais elevados encontram-se junto ao local de epicentro do sismo, sendo a intensidade menor para regiões mais afastada do epicentro.

A diferença entre o uso de uma ordem inferior e uma ordem superior é claramente visível ao comparar ambas as imagens. Verifica-se que a ordem superior efetua um ajuste mais refinado, capturando melhor a tendência da intensidade. De fato, quanto maior o grau do polinómio mais fino será o ajuste da superfície.

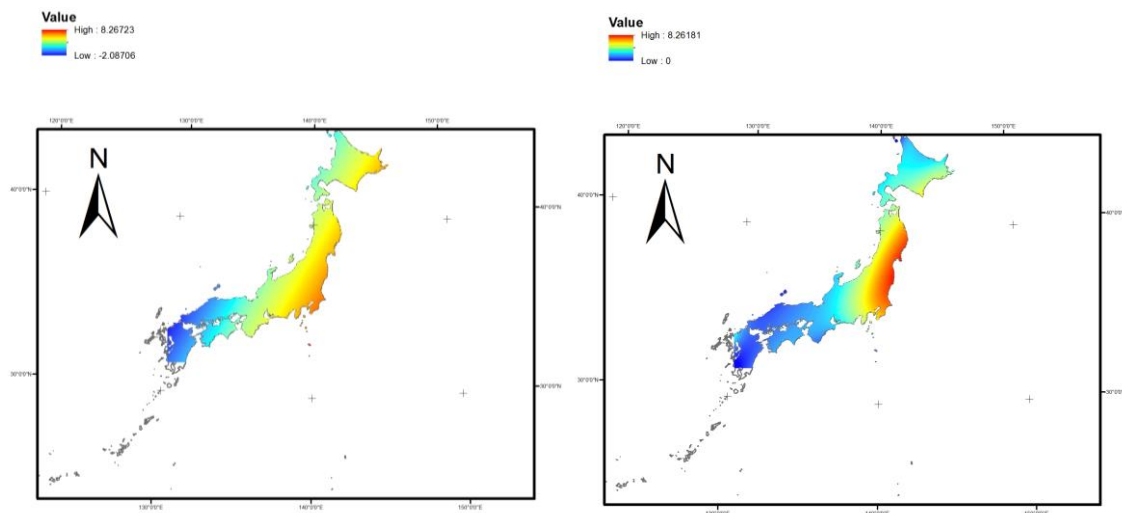


Fig. 26 – Resultado de interpolação polinomial de grau 2 (Esq.) e grau 5 (Dir.).

6.3. Polígonos de Thiessen

Os polígonos de Thiessen são um interpolador local, determinístico e exato, cuja construção segue a regra de que qualquer ponto no interior de um polígono está mais próximo do ponto conhecido desse polígono do que qualquer outro ponto da amostra. Este interpolador serve-se da triangulação de Delaunay, que consiste na rede de ligação entre pontos com uma relação de vizinhança mais próxima.

Os polígonos de Thiessen foram criados no *software* ArcGIS, sendo o seu resultado apresentado na **Fig. 27**

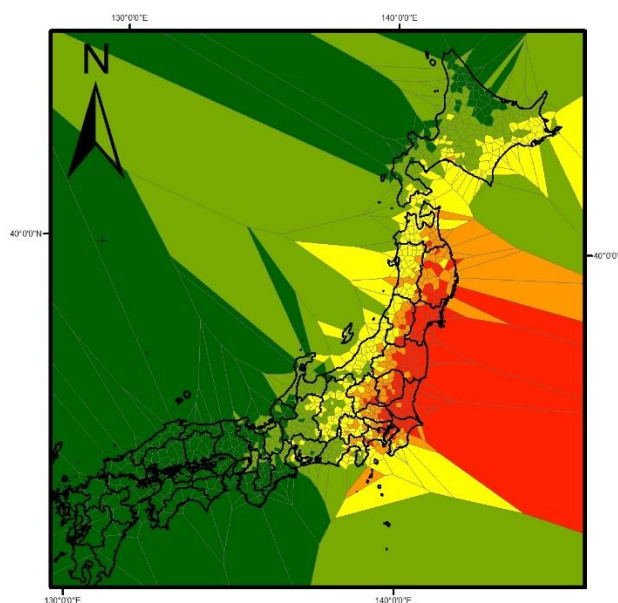
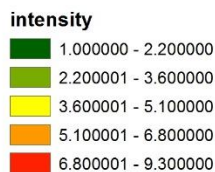


Fig. 27 – Interpolação por polígonos de Thiessen da intensidade.

Os polígonos de Thiessen revelam os locais onde foi sentida a maior intensidade sísmica e as zonas onde essa intensidade não foi tão sentida. Uma vez mais os resultados encontram-se em concordância com o que já havia sido observado.

6.4. Inverse Distance Weighted (IDW)

O método de interpolação ponderada pelo inverso da distância (IDW) é um método local, determinístico e exato. Este método parte do princípio que os pontos conhecidos mais próximos do ponto a estimar influenciam o valor do atributo desconhecido. O IDW determina o valor da célula através de uma combinação linear pesada de um conjunto de amostras pontuais, sendo o peso uma função do inverso da distância.

Para proceder à interpolação por IDW no *software* ArcGIS é necessário definir o número de pontos considerados como vizinhos para efetuar a interpolação e a distância máxima de busca dos pontos próximos. Neste caso foram utilizados os valores por defeito: 12 pontos vizinhos e uma distância máxima igual à diagonal da extensão da amostra. Os resultados são apresentados na **Fig. 28**.

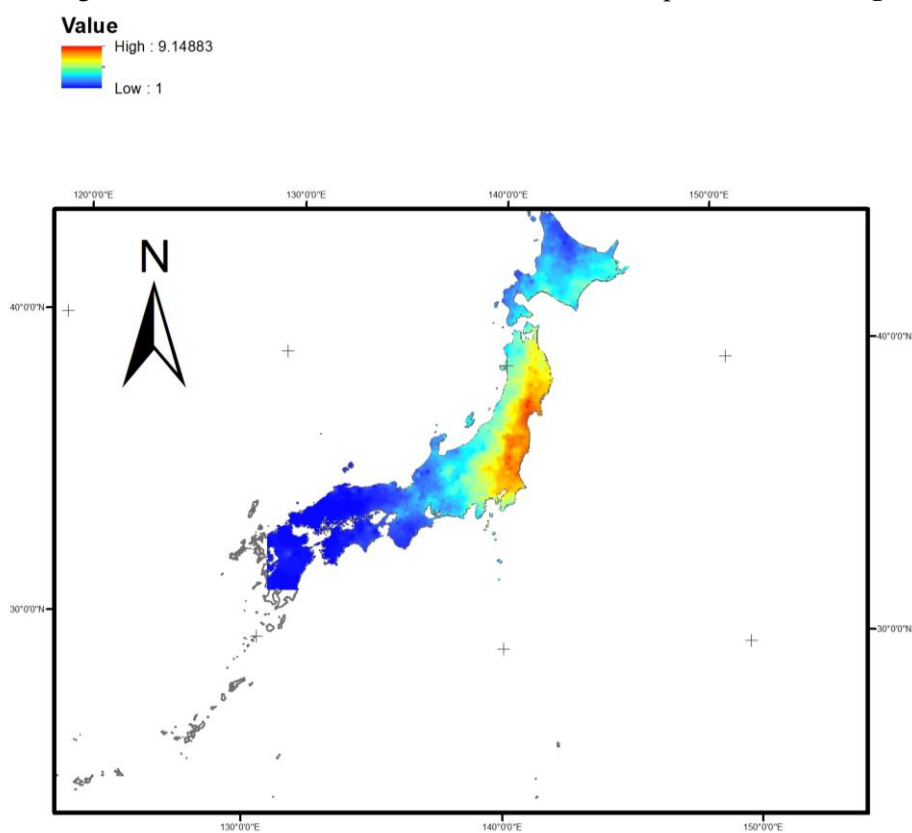


Fig. 28 – Interpolação pelo método IDW.

Comparando a interpolação IDW com as interpolações anteriores, verifica-se que este método permite obter resultados mais refinados, pois baseiam-se na vizinhança do ponto a ser interpolado. É visível a captura de pequenos comprimentos de onda, verificando-se zonas onde são detetadas pequenas variações na superfície interpolada.

6.5. Splines

As *splines* integram-se no grupo de interpoladores locais, determinísticos e exatos. Este interpolador estima valores através de uma função matemática que minimiza a curvatura geral da superfície, resultando numa superfície suavizada que passa exatamente por todos os pontos.

No *software* ArcGIS é necessário definir o peso atribuído das derivadas de terceira ordem da expressão da superfície derivada. O valor aqui utilizado foi o valor por defeito, 0.1. Para além disso foi necessário

definir o número de pontos por região na aproximação local, tendo sido utilizado o valor 12. O resultado é apresentado na **Fig. 29**.

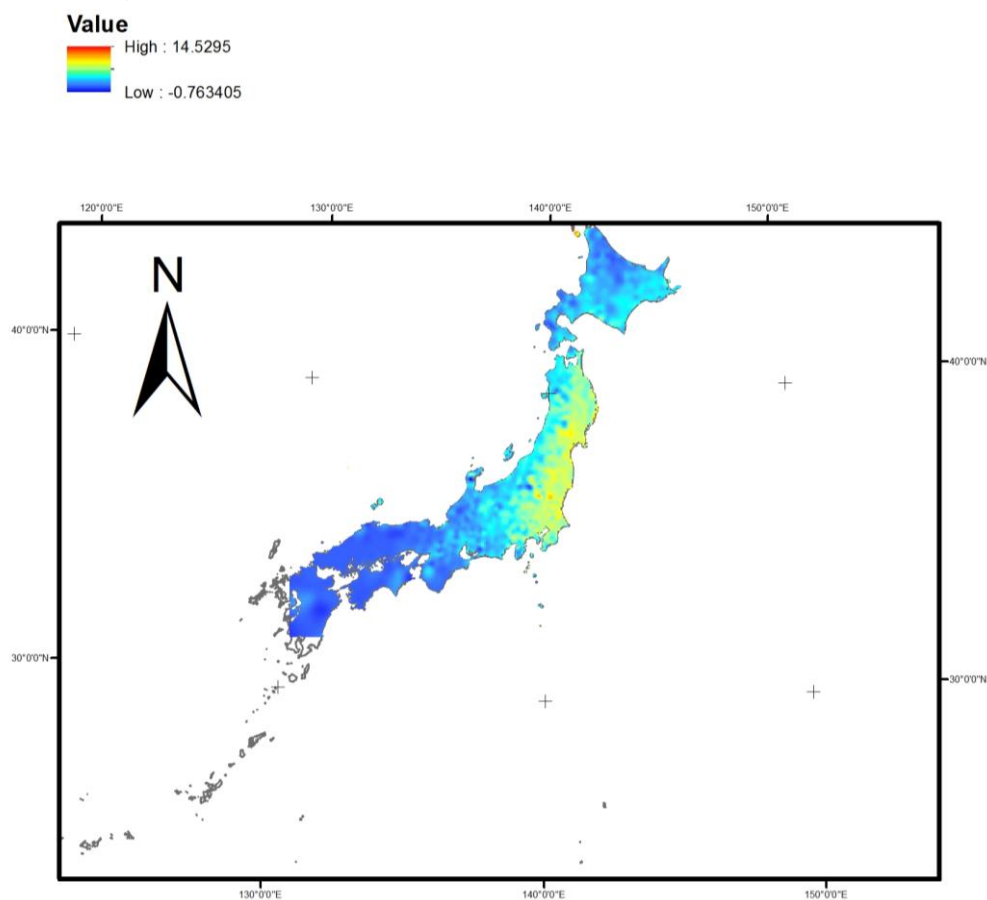


Fig. 29 – Resultados da interpolação por *splines* de mínima curvatura.

É possível observar que são detetadas as mesmas zonas de maior intensidade. No entanto, estas não se encontram tão bem refletidas no caso das *splines* como nos casos anteriores (as zonas de maior intensidade não sobressaem, havendo poucos locais onde aparecem pontos a vermelho).

6.6. Kriging

O interpolador de Kriging enquadra-se num método especial de interpolação, pois é uma interpolação probabilística que requer passos iniciais para obtenção de uma superfície ótima. Estes passos são efetuados por intermédio do semivariograma, que permite caracterizar a autocorrelação das variáveis. Para obter melhores resultados na interpolação, o semivariograma deve ser definido de forma que resulte numa função que melhor se ajuste aos dados da amostra. Esta definição deve entrar em conta com o modelo utilizado (esférico – utilizado neste exemplo, exponencial, Gaussiano ou potência) e com isotropia ou anisotropia. Observando a **Fig. 30**, é possível verificar um comportamento anisotrópico da amostra (preferência direcional), pelo que foi utilizado um modelo anisotrópico, definindo-se a direção preferencial.

Para além do método manual acima descrito foi utilizado um método automático, onde os parâmetros do semivariograma são definidos pelo *software*. Para o método automático é necessário definir o método utilizado (ordinário vs universal), o modelo de semivariograma (esférico, circular, exponencial, gaussiano ou linear) e o raio de busca (número de pontos e distância máxima). Os resultados da interpolação para ambos os métodos (manual e automático) são bastante idênticos, não se verificando diferenças assinaláveis. Por este motivo apenas é apresentado o resultado do método automático na **Fig. 31**.

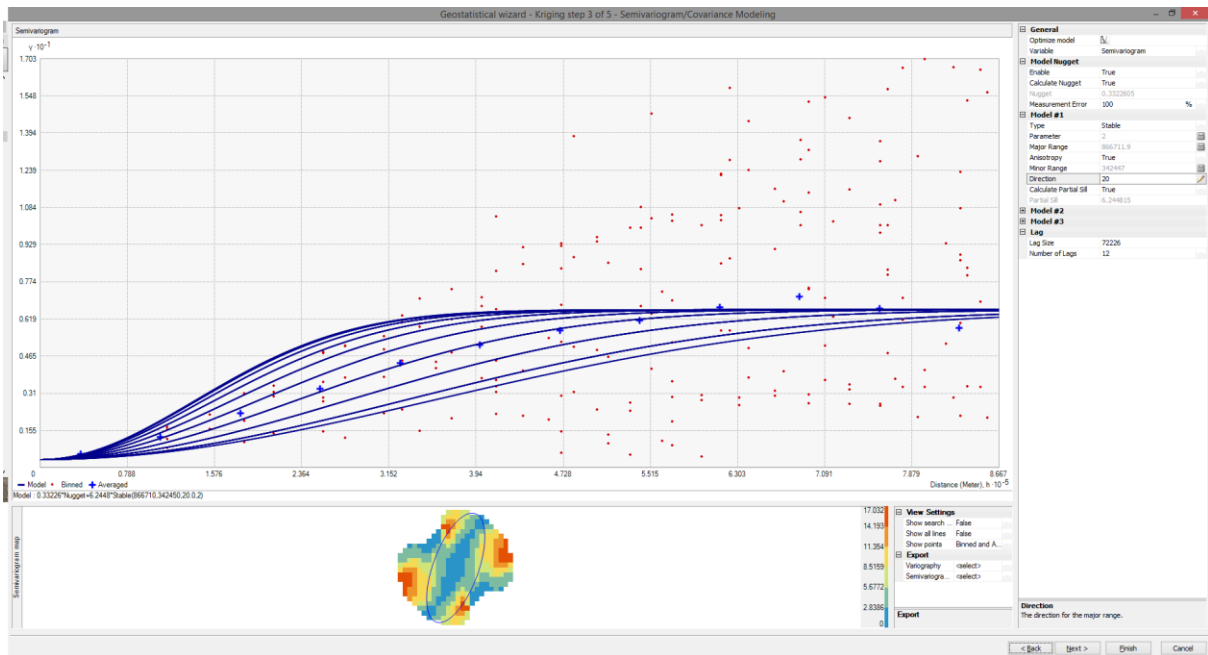


Fig. 30 – Definição de parâmetros para interpolador de Kriging através da análise do semivariograma.

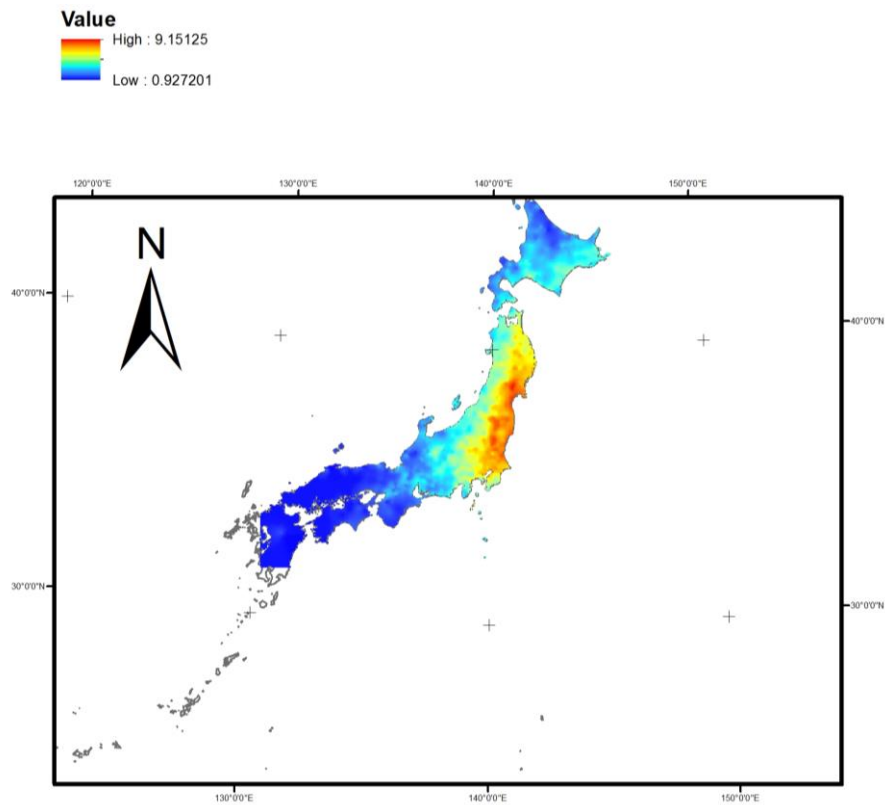


Fig. 31 – Resultado da interpolação pelo método de Kriging manual.

A análise mais interessante que se obtém a partir deste método é a semelhança de resultados quando comparado com o IDW, tendo sido captadas as pequenas variações de intensidade sísmica ao longo de toda a extensão.

6.7. Análise de Resultados

Um método de interpolação exige sempre a análise das precisões associadas ao mesmo. No presente projeto esta análise foi efetuada de duas formas distintas: para o método de Kriging esta análise pode ser extraída de imediato através do ArcGIS, e o seu resultado é apresentado na **Fig. 32**.

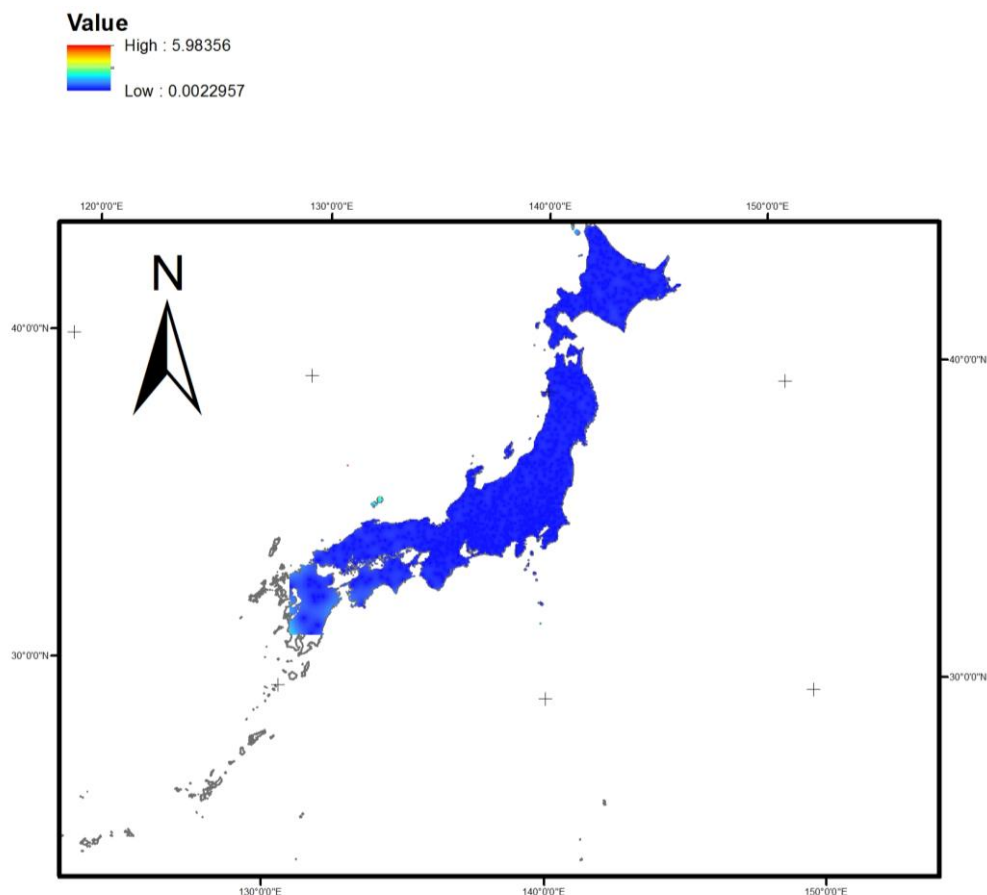


Fig. 32 - Validação da interpolação pelo método de Kriging.

É possível verificar que se atingiu uma boa precisão global de interpolação, pois os valores são extremamente baixos.

Para obter os erros médios quadráticos dos restantes métodos, de forma a avaliar o melhor método de interpolação, foi necessário extrair pontos da superfície interpolada na posição dos pontos da amostra original. Posto isto é feita a comparação dos resultados, que se encontram resumidos na **Tabela 6**.

Tabela 6 – Comparação da precisão dos métodos de interpolação.

Método	Polinomial Grau 2	Polinomial Grau 5	IDW	<i>Spline</i>	Kriging Manual	Kriging Automático
RMS	1.23	0.65	0.14	0.21	0.17	0.17

É possível verificar que o método mais preciso de interpolação aqui verificado foi o IDW. O método de Kriging também apresenta boas precisões, sendo que o método manual apresenta uma precisão igual ao método automático. O método polinomial é o que apresenta a pior precisão, embora o aumento do grau do polinómio tenha resultado numa precisão bastante melhor. O interpolador *spline* também apresenta uma precisão aceitável.

Pela análise da tabela conclui-se que, na presente aplicação, o método IDW seria o melhor método.

7. Identificação de Áreas de Inundação

Como já havia sido concluído, uma das principais razões para o número de mortes na sequência de um sismo, em particular na região do Japão, deve-se às inundações. Por este motivo torna-se importante estudar a variabilidade de altitude na região, identificando as zonas de possíveis cheias e inundações. Para esse efeito é necessário proceder à refinação e interpolação de um MDT (Modelo Digital de Terreno) obtido por SRTM (*Shuttle Radar Topography Mission*).

7.1. Interpolação de MDT SRTM

O MDT obtido através da fonte de dados mencionada em 2 vem num formato *raster* com um espaçamento entre pontos de cerca de 90 m. Para ser possível proceder a uma análise fiável das bacias é necessário refinar e interpolar o *raster* obtido para um espaçamento de 15 m entre pontos. Contudo, tal não foi possível dada a extensão da região a interpolar, tendo-se feito uma interpolação com espaçamento de 500 m entre pontos. Utilizando o ArcGIS procedeu-se à conversão do *raster* para pontos, e posterior projeção (dado que o *raster* vem em coordenadas WGS84) para o sistema de coordenadas JGD2000. Através dos pontos obtidos efetuou-se uma interpolação, optando-se pelo método IDW pelas boas precisões atingidas anteriormente (**Fig. 33**).

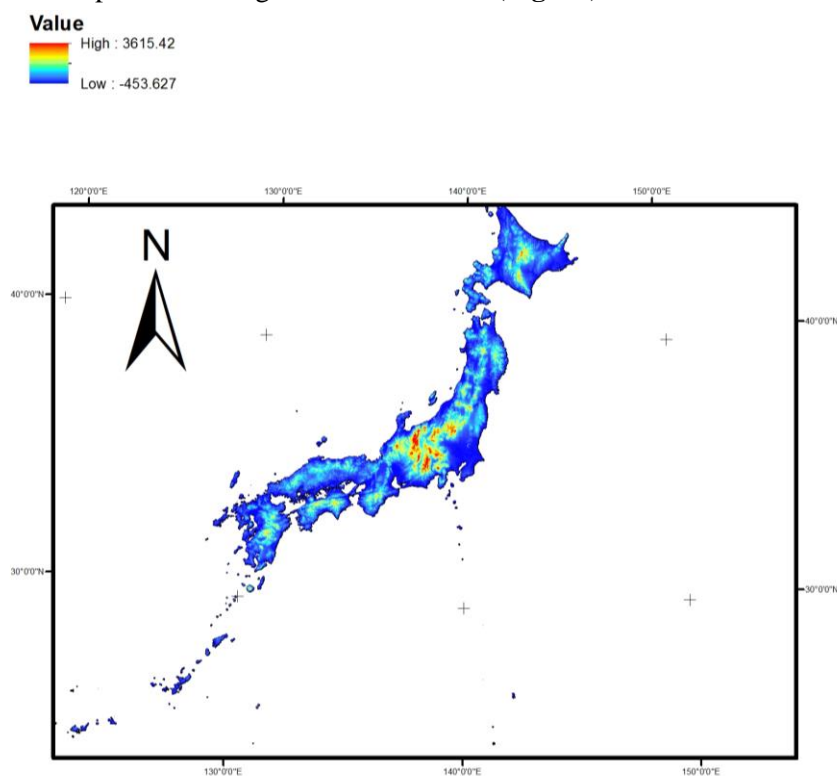


Fig. 33 – MDT do Japão após interpolação pelo método IDW.

Numa situação ideal deveria ser feita a verificação da interpolação através da rede geodésica nacional. Contudo não foi possível aceder a esses dados, pelas dificuldades da linguagem e caracteres específicos. A análise do MDT obtido permite concluir que existe um grande risco de inundação nas zonas do Japão que se encontram junto ao mar, pelo que os planos de emergência no caso de um *tsunami* deverão proceder no sentido de evacuar a população para as zonas mais elevadas.

8. Análise de Redes

Na sequência de um sismo, é de extrema importância determinar a rapidez com que os sistemas de emergência podem chegar aos locais. Esta análise pode ser feita através de redes, onde podem ser identificadas áreas de influência dos serviços de emergência e percursos ótimos até aos locais de emergência.

8.1. Conceitos Teóricos

Uma rede baseia-se no conceito topológico arco/nó, onde os nós são localizações pontuais de início ou fim do fluxo de movimento, e os arcos são os condutores de fluxo entre os nós.

Existem diversos algoritmos de interesse no âmbito do estudo de redes. No presente projeto serão explorados os algoritmos do caminho de custo mínimo, acessibilidade (*closest facility*) e localização (*service area*).

O algoritmo de caminho de custo mínimo permite encontrar o caminho ótimo entre duas ou mais localizações. O algoritmo de acessibilidade permite fornecer uma medida do quão acessível é um lugar a partir de outros lugares. O algoritmo de localização permite encontrar a localização ótima de um conjunto de objetos de modo a que as variáveis encontrem um máximo ou um mínimo.

8.2. Preparação da Rede

Após obtenção de uma rede de estradas é necessário preparar a mesma para proceder à análise dos algoritmos de redes. Dado que seria extremamente moroso proceder a uma análise de redes do Japão, optou-se por analisar apenas a região mais afetada pelos sismos – Sendai.

Em primeiro lugar é necessário corrigir os erros topológicos da rede. Este processo é bastante moroso e, por este motivo, não foi possível aplicar, o que teve implicações importantes como será discutido mais à frente.

Após obter uma rede topologicamente correta é necessário criar uma *network dataset*. Esta operação permite definir uma rede com políticas de conectividade específicas ao mundo real (como por exemplo o sentido das estradas), necessária para correr os algoritmos de análise de redes. O resultado final da rede, incluindo nós e arcos, encontra-se representado na **Fig. 34**, onde estão também representados os centróides dos hospitais para efeitos de análise.

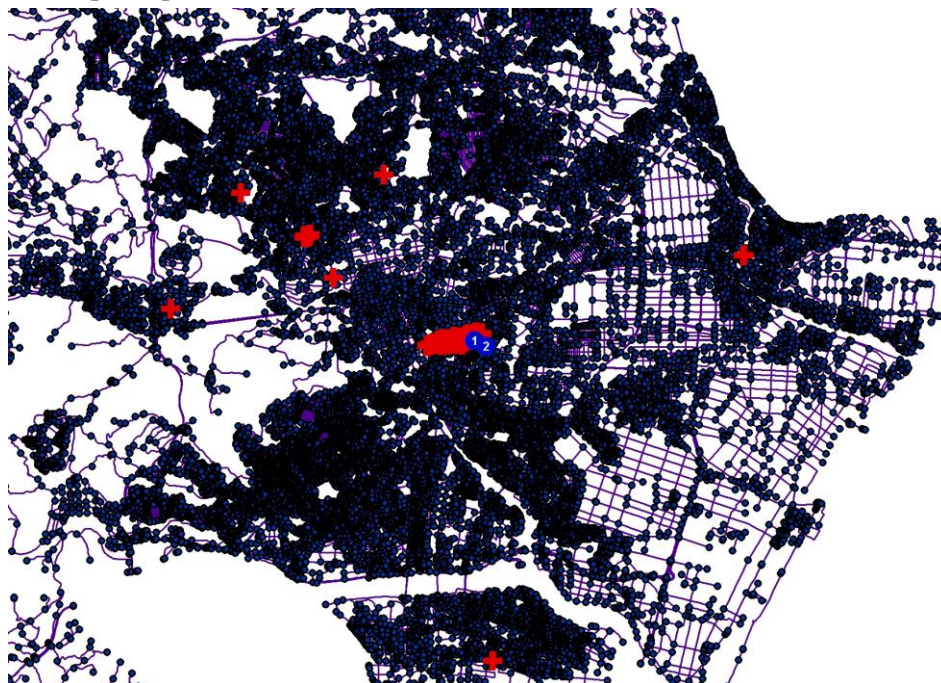


Fig. 34 – Vista geral da rede utilizada com localização de hospitais (cruzes vermelhas).

8.3. Caminho de Custo Mínimo

O caminho de custo mínimo tem como objetivo simplesmente encontrar o caminho ótimo numa rede com base numa origem, destino e impedância (e.g. declive).

Dado que ainda existiam diversos erros topológicos associados à rede obtida, não foi possível correr o algoritmo de custo mínimo. Apenas foi possível observar o caminho de custo mínimo em troços extremamente curtos (**Fig. 35**). Contudo, para referência, refere-se que o objetivo com a análise de custo mínimo seria a determinação do melhor caminho a percorrer pelos serviços de emergência às zonas mais afetadas pelo sismo (junto ao mar).

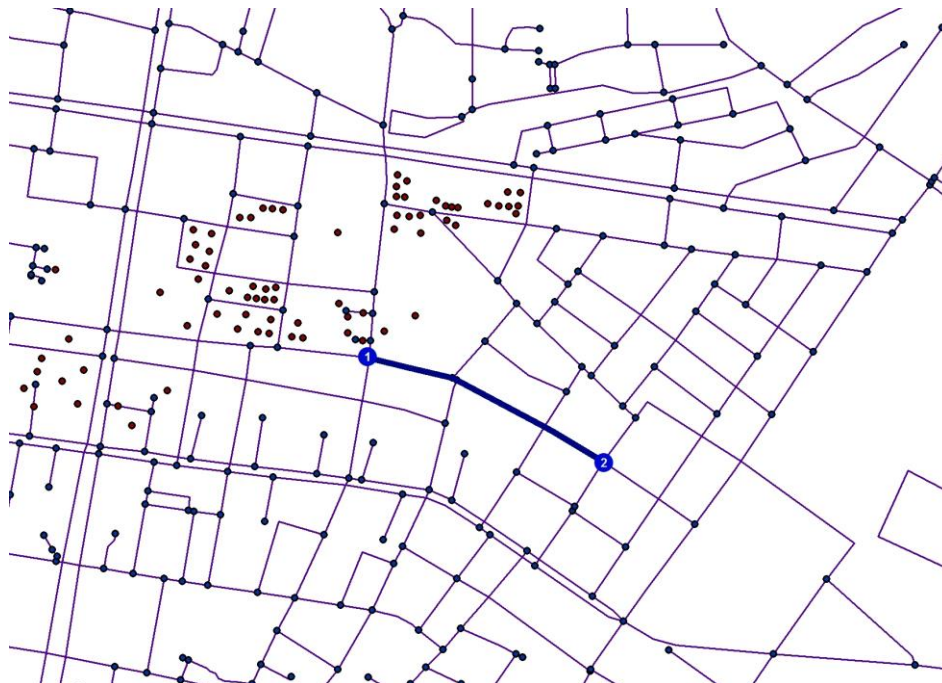


Fig. 35 – Exemplo de um caminho de custo mínimo entre o ponto 1 e o ponto 2.

8.4. *Closest Facility*

A acessibilidade (*Closest Facility*) permite observar o quão acessível é uma localização de todas as outras. Para esse efeito existem vários tipos de medidas de acessibilidade: topológica – conectividade física entre dois pontos no espaço, relativa – medida do grau de conectividade/acessibilidade entre lugares e integral – medida de acessibilidade de um lugar para um número de outros lugares ou atividades.

No âmbito do projeto o objetivo seria analisar os hospitais mais próximos de locais possivelmente mais afetados pelo sismo, nomeadamente zonas ribeirinhas. Na **Fig. 36** encontra-se representada a rede com os locais para análise de acessibilidade.

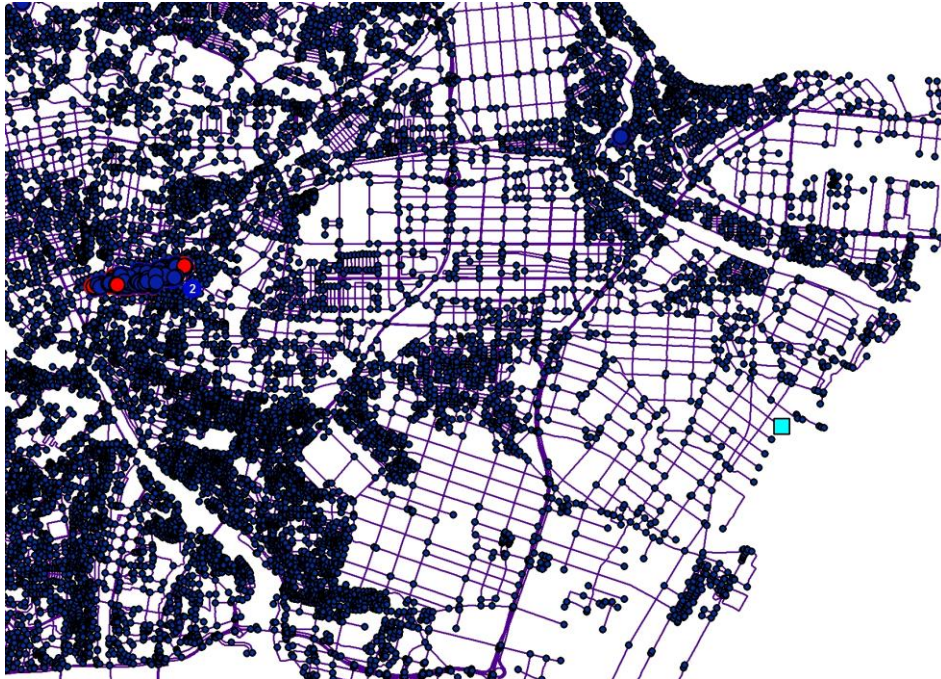


Fig. 36 – Exemplo de *closest facility*.

8.5. Service Area

O *service area* define um polígono que reúne todos os locais que podem ser alcançados a partir de um outro local, entrando em conta com um custo de trajeto (distância ou tempo).



Fig. 37 – Exemplo de *service area*.

Na figura **Fig. 37** é possível observar a área de influência de cada local, com base em duas distâncias: 1500 m e 3000 m. Uma vez mais, devido à topologia, a análise possível de obter é bastante limitada.

9. Conclusões

A análise espacial de informação geográfica fornece uma série de ferramentas úteis na análise dos mais variados efeitos geográficos. No presente projeto foram usadas e discutidas diversas ferramentas no âmbito do estudo de sismologia.

Na primeira fase foi possível observar como as ferramentas de análise espacial apoiam à detecção de padrões pontuais. No caso específico dos sismos, estas ferramentas permitiram comprovar que a distribuição dos sismos se dá junto às zonas de fronteira de placas tectônicas, onde movimentos entre placas podem desencadear uma libertação repentina de energia que provoca os sismos. Especificamente no caso do Japão, foi possível observar o porquê desta zona ser tão ativa em termos sísmicos, devido à existência de várias placas tectônicas ao largo da costa Este. Especificamente, a existência de um ponto triplo faz com que haja uma grande concentração de sismos nessa zona. A detecção de zonas quentes em termos de sismos foi feita graças às ferramentas de detecção de padrões pontuais, como a *kernel density* ou a análise por quadrantes.

Para além de detetar zonas quentes de padrões pontuais, foi possível obter uma análise geral do tipo de padrão existente, especialmente através da função K que demonstra o padrão pontual em função da distância. Comprovou-se que, ao largo da costa do Japão, existe padrão de aglomerado até 1,200 km (que corresponde à extensão da placa ao largo da costa japonesa).

Na segunda fase foi possível estudar de que forma os sismos afetam as regiões do Japão através da análise por áreas. Os dados existentes de intensidades sísmicas sentidas em diferentes estações permitiram analisar de que forma têm sido afetadas essas regiões devido aos sismos mais recentes. Esta análise foi possível graças à detecção de *clusters* e de *hot-spots* com ferramentas de análise espacial (índice de Moran e índice G).

Além de índices globais que permitem, com significância estatística, afirmar a presença ou ausência de aglomerados, índices locais indicaram zonas onde existe autocorrelação espacial. Foram identificadas regiões onde se têm sentido intensidades sísmicas elevadas e reduzidas. Concluiu-se que a costa Nordeste do Japão tem sofrido mais com os sismos recentes, tendo-se sentido maior intensidade sísmica nessa zona.

A regressão espacial permitiu analisar a correlação entre eventos relacionados com os sismos. Dada a importância da prevenção nesta matéria, fez-se uma análise profunda das causas de morte primordiais durante um sismo. Concluiu-se que a intensidade sísmica não é a causa principal para o número de mortes, mas sim a queda de edifícios e os efeitos de *tsunamis*. Foram testados 3 modelos de regressão diferentes para concluir sobre qual o modelo que melhor se ajusta aos dados. Estes modelos podem ser usados para estimar o número possível de mortes em sismos futuros.

Para se conseguir análises mais precisas dos efeitos estudados, seria necessária uma abordagem diferente, especialmente na atribuição de valores de intensidade por região, etc. No entanto o intuito do presente projeto foi a exploração das diferentes técnicas de análise espacial, pelo que esse rigor científico não foi levado ao extremo.

Verificou-se também a importância da interpolação no estudo da sismologia, sendo possível através da mesma observar a distribuição de intensidade observada numa região após ocorrência de um sismo. Neste âmbito os métodos de interpolação IDW e Kriging revelaram ser os métodos mais precisos. Mais importante ainda, a interpolação permite refinar MDTs globais, permitindo um estudo mais preciso de possíveis zonas de inundação, algo importante no âmbito da sismologia.

Por fim, através da análise de redes é possível não só descobrir os melhores caminhos a serem percorridos pelos serviços de emergência no apoio à população, como também as áreas de influência dos serviços de emergência e as medidas de acessibilidade dos mesmos. Esta análise pode apoiar em situações de emergência. Contudo, para que seja possível efetuar uma análise da rede é necessário corrigir a rede de eventuais efeitos de topologia, algo que se relevou extremamente problemático e impeditivo no presente projeto.

O Japão é um país que sofre bastante os efeitos de sismos, e o sismo devastador de 11 de Março de 2011 foi uma prova disso. No entanto, a elevada preparação tecnológica e de engenharia que este país tem

permitiu que o número de vítimas mortais deste sismo fosse menos catastrófica do que teria sido caso não tivesse preparado (especialmente no que diz respeito ao colapso de edifícios). A costa este do Japão representa o local mais crítico, devendo-se concentrar os esforços de planeamento urbanístico nesta zona.

Referências Bibliográficas

Stein, S., Wysession, M. (2003) *An Introduction to Seismology, Earthquakes and Earth Structure*, Blackwell Publishing, Malden, USA.

Catita, C. (2014) *Apontamentos da Cadeira de Análise Espacial da Informação Geográfica*. Texto não publicado, Universidade de Lisboa, Lisboa.

Anselin, L. (2005) *Exploring Spatial Data with GeoDaTM : A Workbook*. Texto não publicado, disponível em <http://geodacenter.asu.edu/system/files/geodaworkbook.pdf>

Wikipedia: The Free Encyclopedia. Wikimedia Foundation, Inc (21/04/2014) *2011 Tohoku earthquake and tsunami*. Acedido a 21/04/2014. http://en.wikipedia.org/wiki/2011_T%C5%8Dhoku_earthquake_and_tsunami

National Police Agency of Japan (10/04/2014) *Damage Situation and Police Countermeasures associated with 2011 Tohoku District – off the Pacific Ocean Earthquake*. Texto não publicado, disponível em http://www.npa.go.jp/archive/keibi/biki/higaijokyo_e.pdf

ArcGis (30/09/2013) *Arcgis Help Library*. Acedido a 19/04/2014. <http://resources.arcgis.com/en/help/main/10.1/index.html#//00qn0000001p000000>

United States Geological Survey (05/03/2014) *Shakemap* Acedido a 15/04/2014. <http://earthquake.usgs.gov/research/shakemap/>

United States Geological Survey (05/03/2014) *Earthquake Archive Search and URL Builder*. Acedido a 15/04/2014. <http://earthquake.usgs.gov/earthquakes/search/>