

Sistemas de Informação Geográfica

Manuel Campagnolo

Instituto Superior de Agronomia, Universidade de Lisboa

1	Introdução	6
2	Representar a realidade geográfica em formato digital	11
	• Conceptualização da realidade	12
	• Duas visões básicas para fenómenos espaciais	16
	• Formatos de dados para informação geográfica	19
	• Um exemplo de formato vectorial	20
	• Um exemplo de formato <i>raster</i>	22
3	Bases de dados geográficos e tabelas de atributos	26
	• Esquemas para tabelas de atributos	33
	• Estruturar bases de dados em SIGs	35
	• Cruzamento de tabelas	41
	• Diagramas de operações	47
	• Selecção por atributos	56
4	Análise espacial sobre cdg vectoriais	60
	• Objectos geométricos e coordenadas	63
	• Métodos básicos sobre objectos geométricos	79
	• Métodos para testar relações espaciais	82
	• Selecção por localização	86

• Métodos que suportam análise espacial	94
• Sobreposição vectorial	102
5 Análise de dados matriciais (raster)	123
• Principais características de um cdg matricial	127
• Histogramas e contraste das imagens	129
• Composições coloridas de imagens	132
• Reamostragem de cdg matriciais	133
• Sobreposição matricial	137
• Filtros para processamento de imagem	142
6 Integração vectorial/matricial e interpolação espacial	146
• Restruturação de vectorial para matricial	148
• Restruturação de matricial para vectorial	150
• Interpolação espacial	151
7 Aplicações a modelos digitais de elevações (raster)	164
• Elevação	165
• Algumas técnicas para análise do relevo	168
8 Sistemas de coordenadas de referência	181
• Geoide, elipsoide de referência e <i>datums</i> globais	184

● Coordenadas geográficas	188
● Altitude ortométrica e elipsoidal	197
● Projecções cartográficas	199
● Projecções cartográficas adoptadas na UE	201
● Descrição <code>proj4</code> para sistemas de coordenadas	204
● Transformação de coordenadas	205

9 Imagens multiespectrais

● Imagens de DR: resolução espacial, espectral e temporal	228
● Acesso a dados de deteção remota	231
● Composições coloridas e índices de vegetação	239
● Interpretação de imagens de DR: índices de vegetação	243

Lista de diagramas de operações

Cruzamento de tabelas	44, 55
Tipos de cdg e atributos	47
Atributos e chaves	48, 49
Edição de tabela	55
Selecção por atributos	58
Conversão <i>multipart</i> para <i>singlepart</i>	76, 118
Indicação de CRS e cálculo de áreas	80
Selecção por localização	87, 89
Sobreposição vectorial: intersecção	105, 108
Sobreposição vectorial: união	107, 108
Sobreposição vectorial: geração de buffers	111
Sobreposição vectorial: recorte (<i>clip</i>)	117
Sobreposição vectorial: dissolução	117, 118, 119
Sobreposição vectorial: diferença	121
Sobreposição vectorial: composição de operações	118
Sobreposição matricial	141
Restruturação de vectorial para matricial	149
Restruturação de matricial para vectorial	150

Ciência de Informação geográfica é uma área de engenharia relacionada com a aquisição, processamento, visualização e distribuição de dados georreferenciados provenientes da várias fontes com exactidão posicional conhecida.

Este designação foi criada nos anos 80 e está relacionada com a crescente disponibilidade de dados geográficos e a crescente utilização desses dados. O processamento dessa enorme quantidade de informação requer portanto meios informáticos.

Para estruturar e facilitar o acesso ao manancial de dados geográficos actualmente disponíveis, têm sido desenvolvidas iniciativas para desenvolver infraestruturas e normas de dados espaciais, a nível nacional, continental ou global. Um dos exemplos é a directiva INSPIRE que se aplica à UE e outros países vizinhos (<https://youtu.be/xew6qI-6wNk>).

Disciplinas da Ciência de Informação Geográfica

A Ciência de Informação Geográfica apoia-se em diversas disciplinas tais como:

- ① Informática: processamento automático de informação;
- ② Geodesia: forma da Terra (geoide e elipsoide) e localização;
- ③ Topografia: métodos e instrumentos para medir e representar detalhes da superfície;
- ④ Cartografia: representação simbolizada da realidade geográfica;
- ⑤ Fotogrametria: determinação da posição e forma de objectos à superfície através de medições em imagens;
- ⑥ Detecção remota: observação da Terra do ar ou do espaço;
- ⑦ Sistemas de posicionamento global: determinação da localização 3D de objectos em qualquer local da superfície e em tempo real;
- ⑧ Sistemas de Informação geográfica (SIG): sistemas informáticos para suporte e processamento de dados georreferenciados;
- ⑨ Ontologia: conceptualização e abstracção da realidade por forma a poder ser representada num sistema informático;

Objectivos principais da UC

Pretende-se que os alunos da UC atinjam os seguintes objetivos:

- ① Demonstrar a compreensão dos conceitos que suportam os Sistemas de Informação Geográfica (SIG);
- ② Usar um SIG para aceder a dados locais ou disponíveis na internet, visualizar e pesquisar dados espaciais;
- ③ Criar e editar conjuntos de dados vetoriais, respeitando as normas de SIGs e a topologia da realidade a representar;
- ④ Representar dados geográficos usando as funcionalidades do SIG e aplicando os conhecimentos adquiridos em Cartografia;
- ⑤ Usar a aplicação SIG para analisar dados e saber realizar um projeto SIG do início ao fim: recolha de dados, armazenamento e manipulação de dados, análise, documentação e apresentação.

Referências bibliográficas

- 1 Ciência e Sistemas de Informação geográfica: conceitos e aplicações:
 - ▶ J. Matos, Fundamentos Informação Geográfica, 8^aed. (2008) Lidel. Disponível na BISA
 - ▶ Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). Geographic information science & systems (Fourth edition). Wiley. Disponível na BISA.
 - ▶ J. Campbell and M. Shin, Essentials of Geographic Information Systems, (2011) acessível on-line:
<https://digitalcommons.liberty.edu/textbooks/2>
 - ▶ Chang, K.-T. (2018). Introduction to geographic information systems (Ninth Edition). McGraw-Hill Education Disponível na BISA.
- 2 Topografia e cartografia:
 - ▶ Casaca, J. M., Matos, J. L. de, & Dias, J. M. B. (Eds.). (2006). Topografia Geral: GPS, fotogrametria, detecção remota, modelação numérica do relevo (7^aed.). Lidel.
 - ▶ Gaspar, J. A. (2005). Cartas e projecções cartográficas (3^aed. actualizada e aumentada). Lidel.

Alguns links introdutórios sobre SIGs

- *John Snow and the 1854 Broad Street cholera outbreak*, considerado como uma das primeiras análises de dados geográficos, assim como um trabalho pioneiro em epidemiologia
<https://youtu.be/lNjrAXGRda4>;

Representar a realidade geográfica em formato digital

Simplificação da realidade

A realidade pode ser mais ou menos complexa à escala da representação pretendida.

Chama-se *generalização cartográfica* ao conjunto de processos de simplificação e adequação da informação à escala do mapa.

Paisagem Agrária do Ribatejo

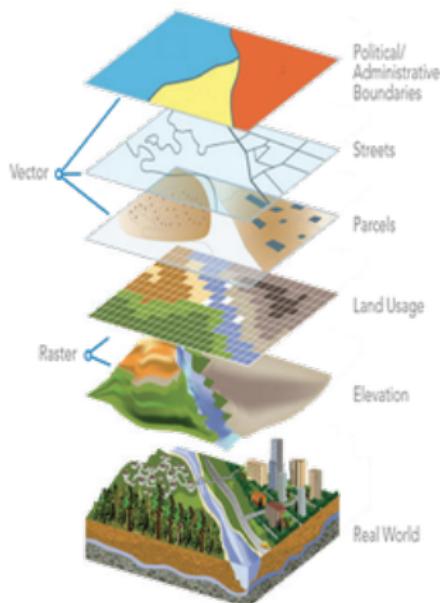


Paisagem agrária da Região de Entre Douro e Minho



Aldeia de Brufe – Terras de Bouro (Barreira de Condensação)

Simplificação da realidade



Conceptualização da realidade

Suponha-se que se pretende criar o cadastro de uma zona rural. Dessa **realidade**, apenas interessa considerar alguns aspectos (fenómenos) como, por exemplo, a localização das parcelas e os seus proprietários. Como é que esses fenómenos do "mundo real" podem ser representados?

Conceito	Definição	Exemplo
entidade	fenómeno do mundo real	<i>parcela de cadastro</i>
atributo	propriedade da entidade	<i>NIF do proprietário</i>
objecto geométrico	representação geométrica da entidade	<i>polígono num SIG</i>
<i>feature</i>	entidade+objecto+atributos	
tipo geométrico	forma de representação	<i>polígono</i>
cdg: conjunto de dados geográficos	coleção de <i>features</i> com um tema e formato comum	<i>representação das parcelas num SIG</i>

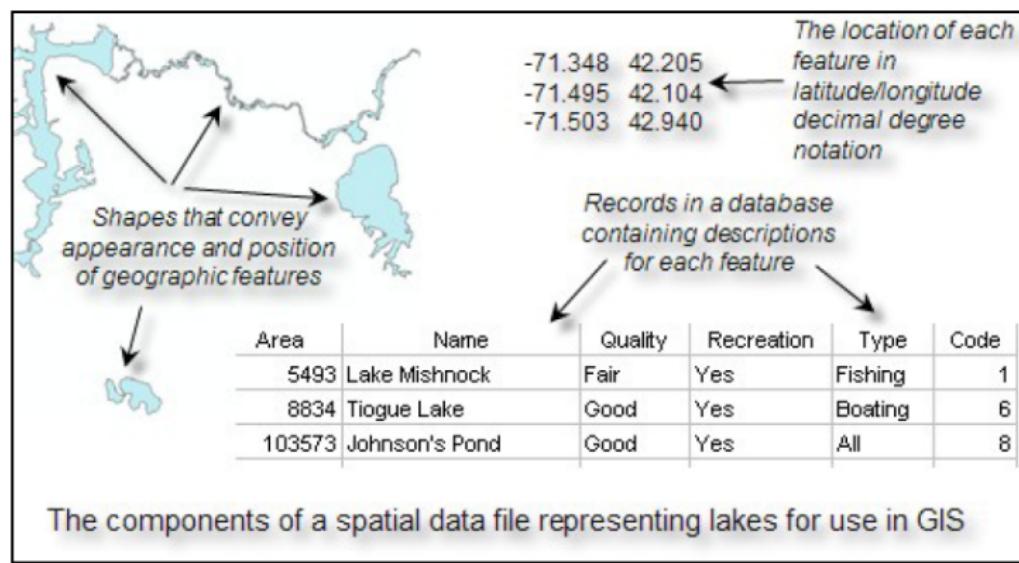
Nota: Matos (2008) sugere “característica” como tradução de *feature*.

Entidades e atributos

Representação de um conjunto de dados geográficos, quando a delimitação espacial das entidades é bem definida:

coleção de objectos geométricos + tabela de atributos

Neste novo exemplo, cada **entidade** (lago) corresponde a uma **linha da tabela**.



Source: <http://gisatbrown.typepad.com/gis/files/spatialdatafiles.pdf>

Duas visões básicas para fenómenos espaciais

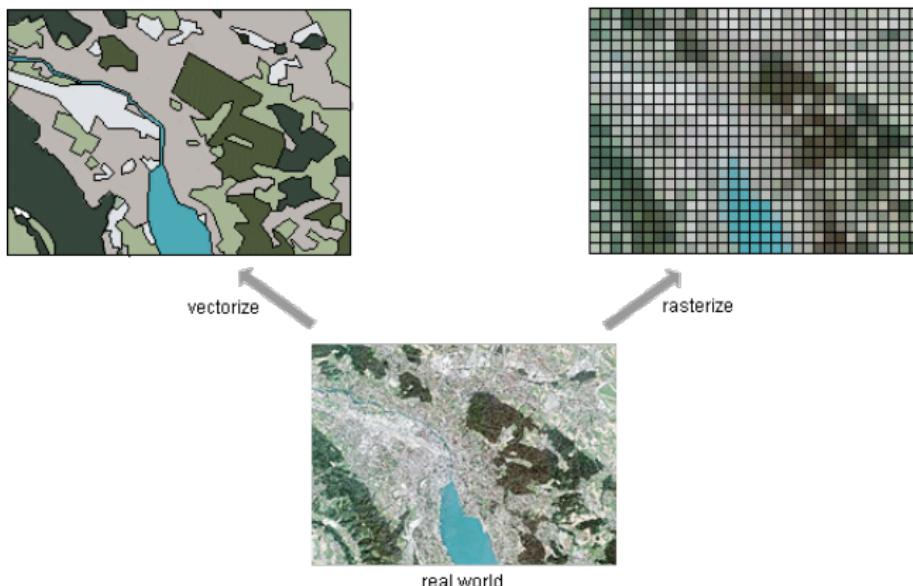
Quando se pretende representar o “mundo real” num sistema de informação geográfica, é necessário escolher o melhor **modelo conceptual de dados** para essa representação.

Para essa escolha é determinante distinguir dois tipos de fenómenos do mundo real:

- ① **Entidades isoladas no espaço, com fronteiras bem definidas**, como por exemplo um conjunto de edifícios, de estradas, de estações meteorológicas ou de parcelas agrícolas. O melhor modelo para representar este tipo de fenómeno é o **modelo vectorial**;
- ② **Fenómenos distribuídos no espaço de forma contínua**, como por exemplo a temperatura da superfície, a quantidade de precipitação ou o índice de área foliar. O modelo mais adaptado para representar este tipo de fenómeno é o **modelo matricial** (também conhecido por *raster*).

Duas visões básicas: vectorial vs matricial

Por vezes a escolha entre modelo vectorial e o modelo matricial para representar um certo fenómeno espacial não é clara. O modelo vectorial permite em geral uma melhor **representação das fronteiras geográficas** das entidades relevantes mas o seu rigor depende da **precisão posicional** dos pontos que definem as fronteiras.



Duas visões básicas para fenómenos espaciais: exemplo

Nesta representação de um conjunto de parcelas agrícolas no Ribatejo sobre uma composição colorida de uma imagem de satélite:

- ① **parcelas agrícolas** (entidades) são representados por polígonos (modelo vectorial);
- ② a **reflectância da superfície** (fenómeno contínuo do espaço) é representada por uma imagem (modelo matricial).



Formatos de dados para informação geográfica

Cada **modelo conceptual de dados** pode ser concretizado numa grande variedade de **formatos de dados**. São esses formatos de dados (tipicamente conhecidos pela sua extensão) que nos permitem armazenar e transferir informação geográfica.

Alguns formatos de ficheiros para conjuntos de dados geográficos:

- ① **Formatos vectoriais**, e.g. OGC GeoPackage (OGC=Open Geospatial Consortium), SpatiaLite, OGC Web Feature Services, GeoJSON (Internet Engineering Task Force); Geography Markup Language (GML), Google Keyhole Markup Language (KML/KMZ), ESRI Shapefile, ESRI Geodatabase,
- ② **Formatos matriciais**, ver GDAL Raster Formats
https://svn.osgeo.org/gdal/tags/gdal_1_2_5/frmts/formats_list.html
- ③ **Tabelas simples** (formato não espacial de dados). e.g. CSV, TXT,
....

Um exemplo de formato do tipo vectorial: *shapefile*

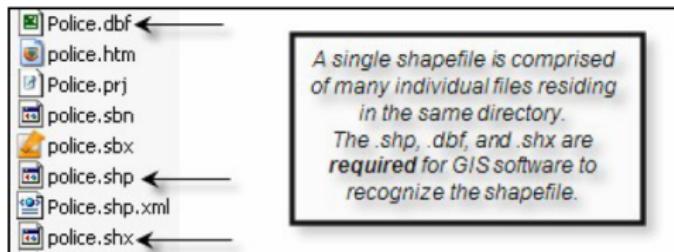
- ① O formato *shapefile* é um formato aberto;
- ② Os objectos geométricos podem ser do tipo **ponto**, **linha** ou **polígono**.
- ③ A escolha do tipo é feita dependendo da **natureza das entidades** a representar, do **detalhe** desejado e das **características das entidades** que se pretende representar.

Exemplos:

- ① **Aglomerados populacionais** podem ser representados como pontos ou como polígonos; A escolha depende do nível de detalhe e da informação que se quer armazenar. Por exemplo, para estudar a dinâmica de zonas urbanas, é necessário ter os contornos dos aglomerados populacionais, i.e. é necessário usar objectos do tipo polígono.
- ② **Linhos de água** ou uma **rede viária** podem ser representados como linhas a uma escala reduzida ou como polígonos a uma escala maior.

Formatos de dados do tipo vectorial: o formato *shapefile* (cont.)

- 1 O formato *shapefile* é na verdade constituído por pelo menos três ficheiros com o mesmo nome e extensões diferentes. Esses ficheiros têm que estar na mesma pasta de dados:
 - ▶ o ficheiro *.shp* armazena os objectos geométricos (e.g. *lagos.shp*)
 - ▶ o ficheiro *.dbf* armazena a tabela de atributos (e.g. *lagos.dbf*)
 - ▶ o ficheiro *.shx* estabelece a relação entre objectos de *.shp* e linhas da tabela *.dbf* (e.g. *lagos.shx*)
- 2 O ficheiro *.prj* é facultativo (e.g. *lagos.prj*) mas muito importante pois contém a informação sobre o sistema de coordenadas em que estão expressos os objectos geométricos que representam as entidades.



Source: <http://gisatbrown.typepad.com/gis/files/spatialdatafiles.pdf>

Dados de tipo matricial: imagens e pixels

Todos os dados de tipo matricial (*raster*) têm uma estrutura semelhante: um **arranjo espacial regular** de valores. Cada elemento do arranjo (ou imagem) é designado por **célula** ou **pixel** e tem um valor numérico ou um não-valor (No data, NA, Null, ...)

9	4	4	4	0	5	9	9	4	4
9	5	4	0	6	0	0	7	4	6
0	7	2	7	8	9	4	7	3	8
6	3	1	1	7	8	7	3	6	1
2	7	6	7	5	7	9	0	7	4
7	6	2	8	7	8	2	8	5	8
7	8	7	3	0	9	0	0	5	2
5	8	5	5	6	5	3	2	2	1
6	2	3	4	5	6	9	0	1	4
6	9	5	1	3	6	6	4	4	1

Figura: Ilustração da estrutura de dados raster. No caso mais simples a georreferenciação é definida pelas coordenadas de um canto da imagem e pelo tamanho do pixel (resolução espacial).

Um exemplo de formato *raster*

Dados do tipo matricial representam o espaço de forma extensiva, e são em geral obtidos por detecção remota ou derivados digitalmente de informação espacialmente contínua.

Os diferentes formatos de dados matriciais – elaborados por diferentes softwares e por diferentes produtores de informação – suportam imagens com uma ou mais várias camadas de informação. Por exemplo, uma imagem multiespectral obtida por detecção remota tem tipicamente várias bandas.

Correntemente o formato mais divulgado (e que é na prática um standard) é o formato **Tag Image File Format (TIFF)**, com extensão `.tiff` ou `.tif`.

O formato *TIFF* admite um conjunto de informação adicional flexível (conhecidos por *tags*), consiste num ficheiro único com tamanho até 4Gb e é um formato do tipo **lossless**, isto é as técnicas de compressão usadas são totalmente reversíveis (não há perda de informação).

Formatos de dados do tipo matricial (*raster*): o formato *GeoTIFF*

O formato *GeoTIFF* é o formato *TIFF* com **geolocation tags**. Essas “etiquetas” permitem geolocalizar os pixels da imagem, sendo essa geolocalização feita num determinado **sistema de coordenadas de referência** que podem ser longitude e latitude ou coordenadas cartográficas.

A associação dos pixels ao sistema de coordenadas pode ser feito de várias formas no formato *GeoTIFF*: associar coordenadas ao canto (0,0) da imagem, transformação linear da grelha, conjunto de pontos de controlo ou coeficientes racionais-polinomiais que permitem fazer a ortorrectificação da imagem.

O formato *GeoTIFF* permite então armazenar:

- ① os **valores** dos pixels da imagem e o respectivo domínios de valores;
- ② a geolocalização dos pixels;
- ③ a **resolução** da imagem, i.e. as distâncias no terreno nas duas direcções coordenadas correspondentes a cada pixel.

Detalhes podem ser encontrados em <http://trac.osgeo.org/geotiff/>

Alguns links para vídeos

- Video sobre pixels e representação digital de imagens em studio.code.org: <https://youtu.be/15aqFQQVBWU>
- Introdução a sistemas de informação geográfica (animação) https://youtu.be/_2r43AWi4YM.

Bases de dados geográficos e tabelas de atributos

SIGs e bases de dados

A maioria dos Sistemas de Informação Geográfica combinam:

- ➊ Um **sistema de gestão de bases de dados**, permitindo realizar pesquisas por atributos (conhecidas por *attribute queries*) nas **tabelas de atributos** em que está armazenada a informação;

Distinguem-se os seguintes tipos de tabelas:

- ▶ **Tabelas de dados geográficos**, associados a *features*, i.e. tabelas em que cada linha corresponde a um objecto geométrico que representa alguma entidade geográfica; Neste caso a pesquisa devolve valores da tabela e objectos geométricos associados.
- ▶ **Tabelas de dados não geográficos**, que são simples tabelas, em que as linhas não estão associadas a objectos geométricos do SIG.

- ➋ Funções de **geometria computational**, permitindo fazer pesquisas (conhecidas por *spatial queries*) e outras operações espaciais sobre os objectos geométricos armazenados no SIG.

Por forma a realizar pesquisas por atributos de forma fiável e rápida, **a base de dados contida na SIG deve estar correctamente estruturada**.

Tabelas, atributos e valores

Cada *feature* de um cdg vectorial está associada a uma **linha** da tabela de atributos.

(Tabela de atributos)

É um arranjo regular de linhas e colunas que deve satisfazer os seguintes requisitos:

- ① Há um único valor em cada célula (i.e. na intersecção de uma linha e uma coluna da tabela);
- ② Os valores de uma coluna (i.e. de um atributo) dizem respeito à mesma propriedade das entidades e têm o mesmo domínio;
- ③ Cada linha é única (i.e. a tabela não pode ter duas ou mais linhas idênticas);
- ④ A ordem das linhas e a ordem das colunas é irrelevante.

Cada tabela de um SIG tem um **nome**, não podendo haver tabelas com nomes repetidos no SIG. Os **atributos** de uma tabela devem ter nomes distintos, mas tabelas distintas podem ter atributos com o mesmo nome.

Tabelas, atributos e valores: exercício

- 1 Indique, entre as tabelas seguintes, quais não verificam as especificidades de uma tabela de atributos. Justifique.

tabela valores:

codigo	valor
0	médio
1	baixo
2	médio
3	0

tabela contentores:

ID	descricao
A1	longo e vazio
D6	baixo e vazio
U2	médio e cheio
A4	elevado e vazio

tabela funcionarios:

numero	iniciais
342	JC
526	MC
567	MJ
612	ER

tabela parcelas:

codigo	atributo
1	cenoura
2	batata
3	João Silva
4	hipotecado

- 2 Um mesmo SIG pode incluir duas tabelas que têm um atributo com nome **codigo**?
- 3 Um mesmo SIG pode incluir duas tabelas com nome **parcelas**?

Valores e domínio dos atributos

Cada atributo tem um **domínio**, ou seja os valores que o atributo pode tomar. O domínio é um conjunto pré-especificado.

Em particular, o domínio pode ser:

- O conjunto de números inteiros com um certo tamanho máximo, e.g. [inteiros; 6]
- O conjunto de números reais com um tamanho máximo e uma determinada precisão (número de casas decimais), e.g. [reais; 8; 2]
- O conjunto das palavras (*strings*) até um certo tamanho, e.g [texto; 20]
- O conjunto de datas num certo formato.

Atributos cujos valores não se repetem

Dado que uma tabela não pode ter linhas repetidas, então existe sempre um conjunto de atributos que permite identificar de forma única as linhas da tabela.

Sem perda de generalidade, vamos admitir que existe sempre pelo menos um atributo da tabela que permite distinguir as linhas da tabela.

Caso uma tabela não contenha um atributo nessa condições, é fácil criar um novo atributo que simplesmente enumera as linhas da tabela por 0, 1, 2,

(Chave candidata e chave primária)

*Um atributo cujos valores não se repetem na tabela é designado por **chave candidata**. Caso haja apenas uma chave candidata, esse atributo também é designado por **chave primária**. Se houver mais de uma chave candidata, qualquer uma delas pode ser escolhida para ser a chave primária.*

Tabela de atributos, chaves candidatas e chave primária: exemplo

A tabela VACC do INE indica o valor acrescentado na actividade primária agrícola e florestal em cada concelho de Portugal Continental, e tem, entre outros, os atributos DTCC e MEuros.

- ① DTCC é um código de 4 dígitos com a indicação do distrito (DT) seguida da indicação do concelho (CC) – por exemplo, 0101 indica o concelho de Águeda no distrito de Aveiro (01), e 1823 indica o concelho de Viseu no distrito de Viseu (18);
- ② MEuros é o valor acrescentado.

O atributo DTCC é uma **chave candidata** pois os seus valores não se repetem (cada linha da tabela diz respeito a um concelho distinto). Como só há uma chave candidata, essa terá que ser a **chave primária**. O atributo MEuros não pode ser uma chave primária pois pode acontecer que dois concelhos distintos tenham o mesmo valor acrescentado.

Representação de uma tabela de atributos por um esquema

Uma tabela de um SIG por ser representada por um **esquema** que descreve os atributos e indica a(s) chave(s) candidatas(s) por um sublinhado.

Exemplo 1: o esquema para a tabela VACc , supondo que apenas tem os atributos DTCC e MEuros , é:

VACc (DTCC, MEuros)

Exemplo 2: a tabela de um cdg conc_2016 que representa os concelhos de Portugal poderia ser descrita pela seguinte esquema:

conc_2016 (Municipio, Distrito, DTCC, area)

Adicionalmente, os **domínios** dos atributos devem ser indicados. Isso pode ser feito da seguinte forma:

- ① Municipio [texto; 254], palavras até 254 caracteres;
- ② Distrito [texto; 254], idem;
- ③ DTCC [texto; 4], palavras até 4 caracteres;
- ④ area [inteiro; 10], números inteiros até 10 dígitos.

Representação de uma tabela de atributos por um esquema: exercícios

Pretende criar-se SIGs que contenham informação relativa aos seguintes conjuntos de dados. Para cada caso, indique o esquema da tabela, os domínios dos atributos e o tipo de objectos geométricos associados às *features*.

- ① Cadastro: conjunto de parcelas para os quais é necessário conhecer o código (NIP), o tipo de parcela (rústico, urbano ou misto) e o NIF do proprietário;
- ② Rios: conjunto de linhas de água para as quais se quer registar a designação e a navegabilidade (navegável ou não);
- ③ Poços: conjunto de poços que têm uma certa localização e diâmetro.

Evitar redundância nas tabelas de atributos: normalização

Um dos principais aspectos a ter em conta no desenho de uma base de dados é assegurar que a sua manutenção e actualização pode ser realizada de forma eficiente e correcta.

Um dos principais problemas para a gestão de uma base de dados é a existência de informação duplicada (redundância). Essa duplicação de informação coloca em causa a integridade da base de dados se esta for sujeita a alterações do seu conteúdo.

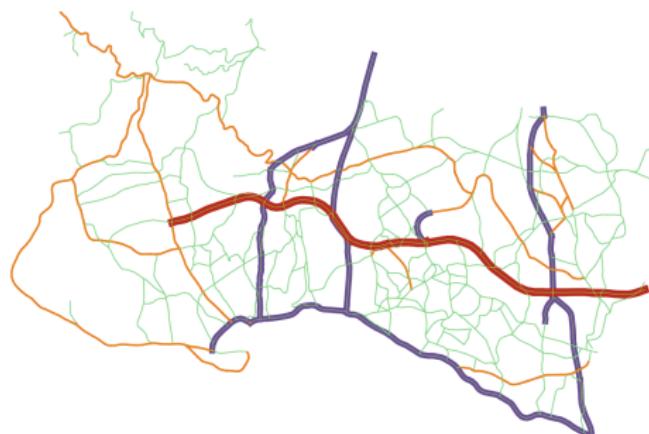
(Normalização)

É o processo de remoção de redundâncias numa base de dados.

Uma das fontes de redundância é a existência de **dependências entre atributos que não são chave candidata da tabela**. Pode eliminar-se esse tipo de dependências cindindo a tabela em duas ou mais tabelas de menores dimensões.

Evitar redundância nas tabelas de atributos: exemplo de normalização

Considere a tabela `RedeViariaProt` (ID, tipo, distancia) associada ao cdg de uma rede viária, em que ID é a identificação da estrada, tipo indica o tipo de estrada (A=auto-estrada, P=est principal, S=est secundária, ou C=caminho) e distancia é uma distância de protecção que depende do tipo de estrada.



Evitar redundância nas tabelas de atributos: exemplo de normalização (continuação)

A tabela original exibe uma dependência entre os atributos `distancia` e `tipo`, não sendo nenhum deles uma chave candidata da tabela, e por isso apresenta redundância de informação.

Para evitar isso pode cindir-se a tabela em duas tabelas criando uma nova base de dados cuja manutenção é mais fácil e mais fiável.

RedeViliariaProt

ID	tipo	distancia
...
1514	S	25
1515	S	25
1516	P	50
1517	A	100
1518	P	50
1519	A	100
1520	C	10
...

RedeViliaria

ID	tipo
...	...
1514	S
1515	S
1516	P
1517	A
...	...

ProtDist

tipo	distancia
C	10
S	25
P	50
A	100

Estruturar bases de dados: exercícios

Proponha tabelas de dados geográficos e não geográficos para estruturar a seguinte informação num SIG:

- ① Localização de linhas de água com indicação do nível de contaminação química; cada linha de água tem uma designação; o seu nível de contaminação pode variar de troço para troço da mesma linha de água; os níveis de contaminação são numerados de 0 a 3, e cada um tem uma descrição desde “não contaminado” até “contaminação elevada”.
- ② Representar um conjunto de árvores notáveis. Registar a localização, o código da árvore, o nome científico, o nome comum, a família, a altura da árvore e o diâmetro do tronco (DAP).
Nota 1: existem diversos indivíduos de cada uma das espécies.
Nota 2: supõe-se que cada espécie tem um único nome comum.

Possíveis soluções para os exercícios do Slide 38

Ex. 1. Cada *feature* corresponde a um segmento de linha de água e pode corresponder a objectos geométricos do tipo linha. Os atributos e seus domínios seriam por exemplo `ID[inteiro; 6]`, `nomeLAGua [texto; 20]` e `CodigoCont [inteiro; 1]`. O esquema da tabela seria:

SegmentosLAGua (ID, nomeLAGua, CodigoCont),

Será necessária uma tabela de dados não geográficos para associar cada código à descrição do nível de contaminação com dois atributos, e.g. `Codigo [integer, 1]` and `Descricao [texto; 20]`:

NiveisContaminacao (Codigo, Descricao)

A tabela NiveisContaminacao tem apenas 4 linhas:

Codigo	Descricao
0	não contaminado
1	contaminação baixa
2	contaminação média
3	contaminação elevada

Possíveis soluções para os exercícios do Slide 38 (continuação)

Ex. 2. É fácil ver que existem dois tipos de atributos: os que dependem de cada árvore, como o **código da árvore**, o **DAP**, a **altura** e o **nome científico** da espécie a que pertence a árvore, e os que podem ser derivados do nome científico da espécie, como o **nome comum** e a **família**. Assim, podemos usar um cdg de pontos em que cada *feature* está associada a uma árvore. O esquema da base de dados é constituído por uma tabela de dados geográficos

Arvores (ID, nomeCientif, Altura, DAP),

e uma tabela de dados não geográficos para registar a família e o nome comum:

Especies (nomeCientif, Familia, nomeComum).

A tabela Especies pode ser algo como:

nomeCientif	Familia	nomeComum
Pinus pinaster	Pinaceae	pinheiro bravo
Quercus suber	Fagaceae	sobreiro
...

Cruzamento de tabelas

Vimos atrás que a melhor estrutura da base de dados pode incluir várias tabelas. Como se pode então aceder ao conjunto dessa informação?

As tabelas de uma base de dados podem ser cruzadas para criar novas formas de ver a informação contida na base de dados.

Como simplificação, iremos considerar que se cruza sempre uma tabela de dados geográficos com uma tabela de dados não geográficos.

Para fazer o cruzamento, é necessário que :

- ① Ambas as tabelas tenham um atributo comum;
- ② Esse atributo seja **chave candidata** da tabela de dados não geográficos.

Nesse caso, o atributo chama-se uma **chave estrangeira** da tabela de dados geográficos, e é indicada com um *sublinhado a tracejado* no esquema da tabela de dados geográficos.

Cruzamento de tabelas: exemplo

Suponha que quer cruzar as seguintes tabelas por forma a adicionar à tabela da esquerda a informação sobre a população do município disponível na tabela da direita:

Municipio	Distrito	DTCC	area
AGUEDA	AVEIRO	0101	335282290
ANADIA	AVEIRO	0103	216639638
AROUCA	AVEIRO	0104	329111969
...

DICO	popM
...	...
0101	47 249
0103	28 501
0104	21 877
...	...

- A tabela da esquerda é a tabela vista no Slide 33:
`conc_2016 (Municipio, Distrito, DTCC, area)`
- A outra tabela é de dados não geográficos que indica a população em cada município; Tem o seguinte esquema:
`Populacao (DICO, popM)`

Os atributos DTCC e DICO tomam os mesmos valores e por isso permitem fazer o cruzamento das tabelas.

Cruzamento de tabelas: exemplo (continuação)

O resultado do cruzamento das tabelas usando DICO como **chave primária** e DTCC como **chave estrangeira** (indicada por um tracejado no esquema da tabela conc_2016) é uma nova tabela:

conc_2016_2(Municipio, Distrito, DTCC, area,DICO, popM)

Na nova tabela conc_2016_2,

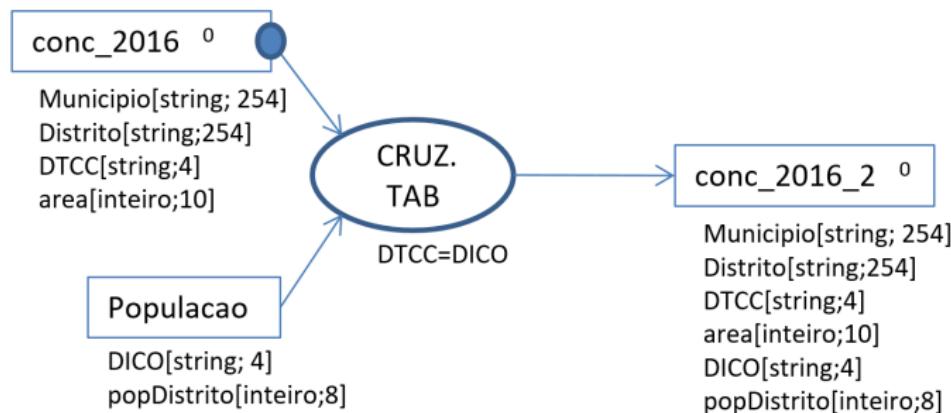
- 1 Os valores dos atributos DTCC e DICO são iguais;
- 2 Cada linha da tabela conc_2016 é aumentada com a linha respectiva da tabela Populacao.

Municipio	Distrito	DTCC	area	DICO	popM
AGUEDA	AVEIRO	0101	335282290	0101	47 249
ANADIA	AVEIRO	0103	216639638	0103	28 501
AROUCA	AVEIRO	0104	329111969	0104	21 877
...

Cruzamento de tabelas: exemplo (continuação)

Quando se trabalha com vários conjuntos de dados e se fazem operações sobre esses dados em SIGs, é conveniente documentar as operações com uma **notação diagramática** simples e compacta.

Nos diagramas de operações deve ser indicado o nome e tipo (ponto, linha ou polígono) dos cdg vectoriais, os seus atributos, e os parâmetros relevantes para as operações. O cruzamento de tabelas no exemplo corresponde ao seguinte diagrama, em que está também assinalado com • o cdg que contém a chave estrangeira.



Tabelas de atributos num SIG: ideias a fixar

- 1 Num SIG existem **tabelas de dados geográficos**, associadas a *features*, e **tabelas de dados não geográficos**;
- 2 As tabelas devem respeitar as condições do Slide 28, e em particular **as linhas da tabela devem ser distintas**; sem perda de generalidade, considera-se que existe pelo menos um atributo que toma valores diferentes em cada linha, ou seja que existe pelo menos uma **chave candidata**;
- 3 Deve evitarse redundância de informação: em particular, não deve haver dependência entre atributos que não são chave candidata da tabela;
- 4 Ao explorar a informação num SIG, as tabelas podem ser cruzadas; em particular, uma tabela de dados geográficos pode ser cruzada com uma tabela de dados não geográficos se a primeira tiver um atributo (**chave estrangeira**) que corresponde a um atributo (**chave primária**) da segunda tabela. Esses dois atributos *não* têm que ter o mesmo nome.

Cruzamento de tabelas: exercício

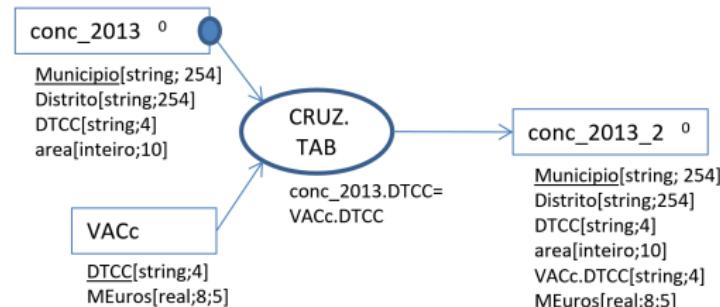
Considere de novo as tabelas

- 1 VACC (DTCC, MEuros), introduzida no Slide 33, e
- 2 conc_2013 (Municipio, Distrito, DTCC, area)

Pretende-se cruzar essas tabelas para aumentar a tabela de atributos do cdg conc_2013 com a informação do valor acrescentado por concelho dada por MEuros.

- 1 Indique a chave estrangeira desse cruzamento de tabelas.
- 2 Represente por uma diagrama essa operação.

Uma solução:



Diagramas de operações (I): tipo de cdg, tabelas e atributos

No diagrama anterior, usam-se caixas para representar tabelas de dados geográficos e não geográficos, e elipses para representar operações. Para cada caixa, indicam-se:

- ① o nome da tabela (coincide com o nome do cdg se for uma tabela de dados geográficos); os nomes no diagrama devem ser **distintos**;
- ② o tipo de cdg se a tabela estiver associada a um cdg:
 - ▶ para um cdg de pontos, usa-se •
 - ▶ para um cdg de linhas, usa-se —
 - ▶ para um cdg de polígonos, usa-se ○

Representação dos atributos de um cdg vectorial:

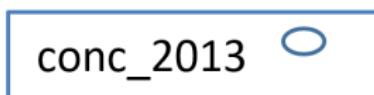
- Os atributos são representados sob o nome do cdg (para maior clareza, as chaves candidatas podem ser sublinhadas);
- O nome do atributo pode ser acompanhado do seu domínio;
- Os nomes dos atributos não se repetem para a mesma tabela; a notação `nomeTabela.nomeAtrib` (e.g. `VACc.DTCC`) permite distinguir atributos originados por cruzamento de tabelas.

Diagramas de operações (II): simplificação do diagrama

Nos diagramas que serão apresentados, convencionase o seguinte.

- 1 Para simplificar, apenas os **atributos que vão ser usados nos parâmetros das operações** (e.g. cruzamento de tabelas) devem obrigatoriamente ser indicados no diagrama;
- 2 Quando um atributo é indicado para um *input* de uma operação, e se o **atributo por preservado pela operação**, esse atributo deve ser também indicado no *output* da operação;
- 3 Se as *features* do cdg são identificadas por uma enumeração 0, 1, 2, ..., essa **enumeração não tem que ser indicada no diagrama**, mesmo que seja uma chave primária da tabela.

Neste caso a
chave primária
é Município



Municipio[string; 254]
Distrito[string;254]
DTCC[string;4]
area[inteiro;10]

codCultura e
área não são
chave primária:
usa enumeração



area [real; 10; 3]
CodCultura [string; 4]

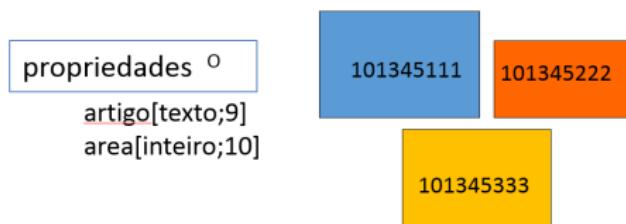
Diagramas de operações e representação de features

Nos exemplos que serão apresentados, por vezes o diagrama será acompanhado por uma representação do cdg, com os seus **objectos geométricos** e **tabela de atributos**.

A **associação** entre cada objecto geométrico e a linha da tabela correspondente pode ser feita

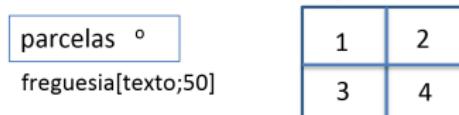
- 1 de forma explícita, usando a chave primária da tabela;

propriedades (artigo, area)



- 2 de forma implícita, por enumeração dos objectos geométricos.

parcelas (ID, freguesia)



Cruzamento de tabelas: Exercício

Suponha que pretende criar um SIG para representar as diversas parcelas de terreno de uma exploração agrícola, em que o número de parcelas é muito superior ao número de culturas. Para cada parcela pretende registar-se:

- ① a localização e área (em ha) das parcelas (Nota: a melhor abordagem seria representar apenas as parcelas no SIG e usar uma funcionalidade do SIG para determinar as áreas: ver próxima secção);
- ② o nome da cultura instalada;
- ③ o tipo e quantidade de fertilizante a aplicar (em $t \times ha^{-1}$) por cultura, que é independente da localização da parcela;
Considera-se que a cada cultura é aplicado um único tipo de fertilizante, mas que os fertilizantes aplicados podem ser distintos entre culturas;
- ④ o custo do fertilizante (em $\text{€} \times t^{-1}$), que depende do tipo de fertilizante.

Cruzamento de tabelas: Exercício (cont.)

Usando um esquema para representar a base de dados e um diagrama para representar as operações a realizar,

- ① Indique como deve ser estruturada a informação relevante para o problema;
- ② Usando cruzamento de tabelas e derivação de novos atributos a partir dos atributos existentes (**edição da tabela**), mostre como se determina o custo do fertilizante a aplicar em cada parcela.

Possível solução para a 1a parte:

Tabela associada a um cdg vectorial de tipo “polígono”:

Parcelas(CodParcela [inteiro;4], area [real;10;3],
CodCultura [texto;4])

Tabelas de dados não geográficos:

Culturas(CodCult [texto;4], Nome [texto;20], Fert_Cod [texto;5],
FertQtt [real;10;3])

Fertilizantes(Codigo [texto;5], CustoTon [real;8;2])

Exemplos de tabelas

Algumas hipotéticas linhas das tabelas do SIG:

Parcelas

CodParcela	area	CodCultura
1	1.2	BA
2	2.3	VI
3	0.5	TO
4	1.3	VI
5	0.7	TO
...

Culturas

CodCult	Nome	FertCod	FertQtt
...
BA	batata	15.10.20	2
TO	tomate	20.10.10	4.5
VI	vinha	15.10.20	1.5
...

Fertilizantes

Codigo	CustoTon
15.10.20	150
20.10.10	138
...	...

Primeiro cruzamento de tabelas

O custo (€) é dado por $\text{custo} = \text{area} (\text{ha}) \times \text{FertQtt} (\text{t/ha}) \times \text{CustoTon} (\text{€/t})$, é necessário dispor de todos esses valores para cada parcela.

Para tal, é necessário cruzar as tabelas para obter a tabela **Parcelas** ampliada com as colunas de **Culturas**. O resultado desse primeiro cruzamento de tabelas (**Parcelas com Culturas**) é:

Parcelas_2

CodParcela	area	CodCultura	CodCult	Nome	FertCod	FertQtt
1	1.2	BA	BA	batata	15.10.20	2
2	2.3	VI	VI	vinha	15.10.20	1.5
3	0.5	TO	TO	tomate	20.10.10	4.5
4	1.3	VI	VI	vinha	15.10.20	1.5
5	0.7	TO	TO	tomate	20.10.10	4.5
...

Segundo cruzamento de tabelas

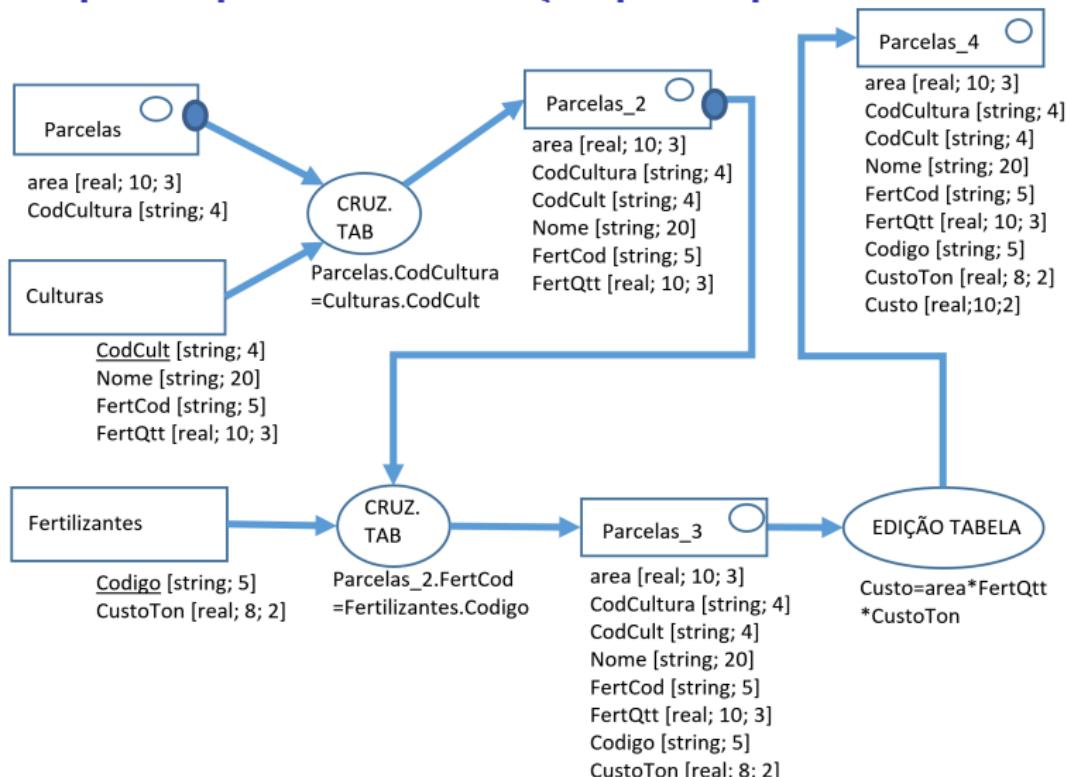
Dado que também é necessário ter o valor de CustoTon para poder calcular $\text{custo} = \text{area} \times \text{FertQtt} \times \text{CustoTon}$, é necessário cruzar a tabela resultante do primeiro cruzamento (Parcelas_2), com a tabela Fertilizantes. O resultado (com algumas colunas removidas para visualização) é:

Parcelas_3

...	area	CodCult	Nome	FertCod	FertQtt	Codigo	CustoTon
1	1.2	BA	batata	15.10.20	2	15.10.20	150
2	2.3	VI	vinha	15.10.20	1.5	15.10.20	150
3	0.5	TO	tomate	20.10.10	4.5	20.10.10	138
4	1.3	VI	vinha	15.10.20	1.5	15.10.20	150
5	0.7	TO	tomate	20.10.10	4.5	20.10.10	138
...

Agora, já está disponível toda a informação necessária para se poder calcular $\text{custo} = \text{area} \times \text{FertQtt} \times \text{CustoTon}$ para cada parcela.

Diagrama para representar a solução para o problema



O atributo **Custo** da tabela **parcelas_4** contém os custos (em €) de fertilizante para cada parcela.

A operação de selecção por atributos ... e um pouco de lógica

(Selecção)

É uma operação sobre um cdg vectorial que devolve um subconjunto de linhas da tabela de atributos e os objectos geométricos correspondentes. A selecção é realizada através de uma expressão lógica que pode tomar valor **V** ou **F** para cada linha da tabela de atributos.

Exemplo: na tabela

conc_2016 (Municipio, Distrito, DTCC, area)

introduzida no Slide 33, pode realizar-se uma selecção dos concelhos com área superior a 1000 km^2 , ou seja $1000\,000\,000 \text{ m}^2$, usando a expressão lógica

$\text{area} > 1000000000$

que toma valor **V** ou **F** para cada linha da tabela.

A operação de selecção por atributos ... e um pouco de lógica

Os operadores lógicos **AND**, **OR** e **NOT** permitem combinar expressões lógicas para criar novas expressões lógicas.

Por exemplo, a expressão

```
area > 1000000000 AND Distrito="EVORA"
```

permite seleccionar os concelhos com mais de 1000 km² do distrito de Évora.

Exercício: como interpreta o resultado da selecção baseada na expressão

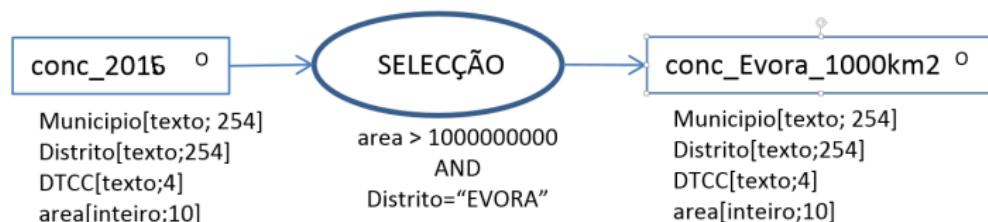
$$(\text{Altura} > 20 \text{ } \mathbf{OR} \text{ } \text{Altura/DAP} > 10) \text{ } \mathbf{AND}$$
$$\mathbf{NOT} (\text{nomeCientif} = \text{"Pinus pinaster"})$$

na tabela Arvores (ID, nomeCientif, Altura, DAP) do Slide 40?

Diagrama para representar a operação de selecção por atributos

A operação de **selecção por atributos** é um operação de selecção em que o valor da expressão lógica depende unicamente dos atributos da tabela. Assim essa operação tem um único *input*, que é o cdg onde é realizada a selecção.

O exemplo do slide anterior pode representar-se pelo seguinte diagrama



em que o *output* foi designado por conc_Evora_1000km2. A tabela de atributos de conc_2016 e de conc_Evora_1000km2 têm as mesmas colunas (mesmos atributos) mas tipicamente a tabela conc_Evora_1000km2 terá um menor número de linhas.

Selecção por atributos: *features seleccionadas*

No exemplo anterior, e porque uma selecção de linhas da tabela de atributos de um cdg vectorial induz uma selecção dos objectos geométricos correspondentes a essas linhas, o resultado da selecção consiste em:

- 1 um subconjunto das linhas da tabela; e
- 2 um subconjunto dos objectos geométricos.

	Municipio	Distrito	DTCC	area
75	ESTREMOZ	ÉVORA	704	513801785
76	ÉVORA	ÉVORA	705	1307081864
77	MONTEMOR-O-N...	ÉVORA	706	1232976775
78	TERRAS DE BOU...	BRAGA	310	277467549
79	ALANDROAL	ÉVORA	701	542683138
80	ARRAIOLOS	ÉVORA	702	683751091
81	BORBA	ÉVORA	703	145187789
82	OVAR	AVEIRO	115	147707197



Análise espacial sobre cdg vectoriais

(Análise espacial)

Conjunto de métodos cujos resultados se alteram quando a localização dos objectos em análise se altera (Longley et al. 2010).



A selecção por atributos não depende da localização das features.

Dados vectoriais: conceitos fundamentais de análise espacial

O OGC OpenGIS Implementation Standard for Geographic information / ISO 19125 define métodos para **objectos geométricos**:

Objectos geométricos do tipo point, line, polygon, multi-point, etc, associados a um sistema de coordenadas;

Métodos sobre objectos geométricos devolvem propriedades tais como dimensão, fronteira, área, centroide, etc (ver Slide 64);

Métodos para testar relações espaciais entre objectos geométricos equals, disjoint, intersects, touches, crosses, within, contains, overlaps and relate, com valor lógico **TRUE** ou **FALSE** (ver Slide 82);

Métodos que suportam análise espacial distância, que devolve uma **distância**, e buffer, convex hull, intersection, union, difference, symmetric difference, que devolvem **novos objectos geométricos** (ver Slide 94).

Fonte: www.opengeospatial.org/standards/sfa

Objectos geométricos e coordenadas

Como discutido anteriormente, cada *feature* de um conjunto de dados geográfico é representado num SIG por um objecto geométrico do tipo "ponto", "linha"ou "polígono".

Um objecto geométrico é portanto definido através:

- ① "pontos": um ou mais pontos de coordenadas (x, y) isolados;
- ② "linhas"ou "polígonos": uma ou mais linhas poligonais, isto é uma sucessão de **vértices** com coordenadas $\{(x_1, x_2), (y_1, y_2), \dots\}$, respectivamente abertas ou fechadas.

em que (x, y) são expressos num determinado **sistema de coordenadas**, que pode ser geográfico ou cartográfico.

Como simplificação, iremos considerar nesta secção que todos os cdg vectoriais têm o mesmo sistema de coordenadas. Assim, o sistema de coordenadas não tem que ser explicitado na descrição das operações espaciais.

Objectos geométricos: dimensão, interior, exterior e fronteira

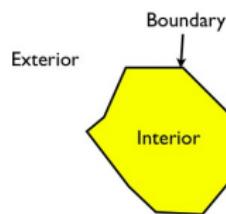
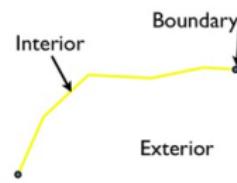
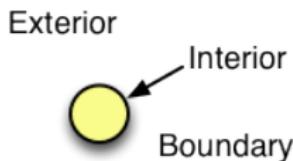
Objectos do tipo “ponto” (p) têm **dimensão 0**,

Objectos do tipo “linha” (L) têm **dimensão 1**,

Objectos do tipo “polígono” (P) têm **dimensão 2**

Os objectos geométricos têm **Interior**, **Fronteira** e **Exterior**:

- ① Se p é do tipo ponto, então $\mathbf{I}(p) = p$, $\mathbf{F}(p)$ é vazio, e $\mathbf{E}(p)$ são todos os pontos que não estão em p ;
- ② Se L é do tipo linha, então $\mathbf{F}(L)$ são os extremos da linha, $\mathbf{I}(L)$ são todos os pontos da linha excepto os extremos, e o exterior $\mathbf{E}(L)$ são os todos os outros pontos;
- ③ Se P é do tipo polígono, $\mathbf{F}(P)$ é a fronteira, $\mathbf{I}(P)$ é o conjunto de pontos de P que não estão na fronteira e $\mathbf{E}(P)$ são os restantes pontos.



Objectos geométricos *singlepart* e *multipart*

Um objecto geométrico (que é a representação geométrica associada a **uma feature**) pode ter **uma ou mais partes**.

- ① Para **objectos do tipo “ponto”** : uma feature está associada a
 - ▶ um único ponto (*singlepart*); ou
 - ▶ mais de um ponto (*multipart*).
- ② Para **objectos do tipo “linha”** : uma feature está associada a
 - ▶ uma única linha poligonal (*singlepart*); ou
 - ▶ uma colecção de linhas poligonais (*multipart*).
- ③ Para **objectos do tipo “polígono”** cada parte pode ou não ter um ou mais *buracos*. Uma feature pode estar associada a
 - ▶ uma única parte positiva (*singlepart*) sem *buracos*;
 - ▶ uma parte positiva com um ou mais *buracos* no seu interior (*singlepart*);
 - ▶ mais do que uma parte positiva com ou sem *buracos* nos seus interiores (*multipart*).

Um “polígono” *singlepart*, com ou sem buracos, é sempre **espacialmente conexo**, i.e. é possível unir dois quaisquer pontos no seu interior. Consequentemente, qualquer *feature* **não conexa** tem que estar associada a um objecto *multipart*.

Objectos geométricos *singlepart* e *multipart*: exemplos

Em seguida ilustram-se alguns casos para objectos do tipo “polígono” que representam **áreas protegidas** em Portugal (ICNF, 2014)

Objecto geométrico *singlepart* com uma única parte e sem *buracos*



Objecto geométrico *singlepart* com uma única parte e 5 *buracos*. Note-se que é **conexo**.

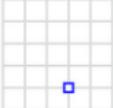
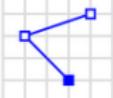


Objecto geométrico *multipart* com várias partes. Note-se que é **não conexo**



Representar objectos geométricos “singlepart”

Geometry primitives (2D)

Type	Examples
Point	 POINT (30 10)
LineString	 LINESTRING (30 10, 10 30, 40 40)
Polygon	 POLYGON ((30 10, 40 40, 20 40, 10 20, 30 10))
	 POLYGON ((35 10, 45 45, 15 40, 10 20, 35 10), (20 30, 35 35, 30 20, 20 30))

Esta representação usa um formato designado por *Well Known Text* (WKT).

Representar conjunto de dados geográficos por uma tabela com geometria

(Representação de um cdg por uma tabela com geometria)

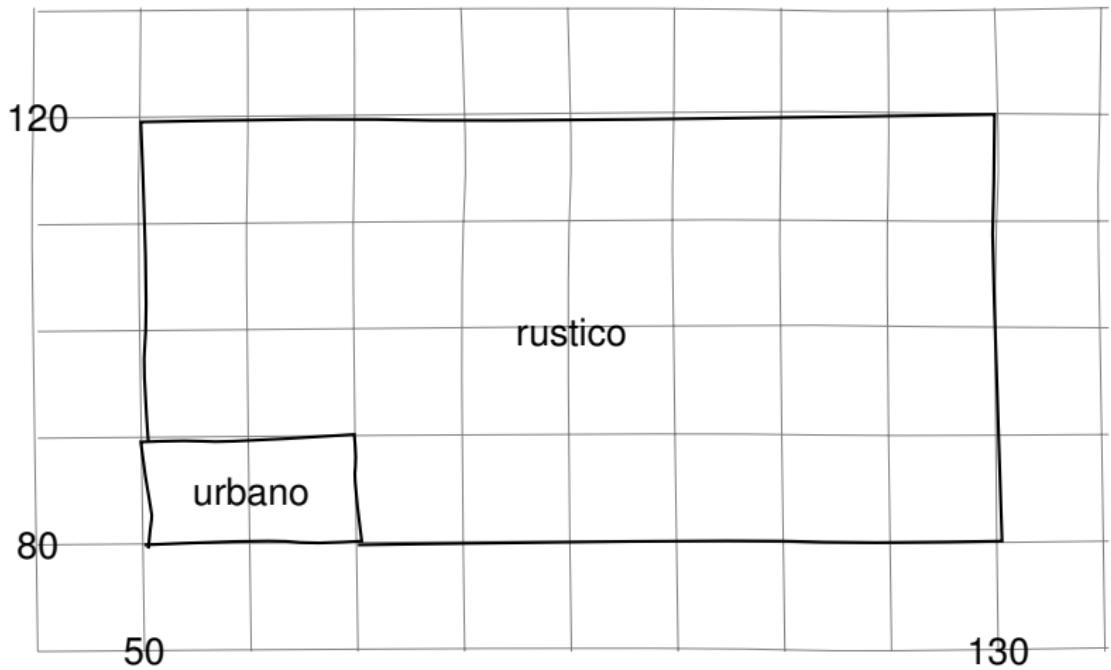
Dado um sistema de coordenadas de referência, um conjunto de dados geográficos (cdg) pode ser descrito por tabela de atributos aumentada com uma coluna geometry que determina a localização geográfica das features.

Exercício. Considere o seguinte cdg, em que a geometria é representada no formato WKT:

Id	Tipo	geometry
1	urbano	POLYGON((50 80,70 80, 70 90,50 90,50 80))
2	rustico	POLYGON((70 80,130 80,130 120,50 120,90 120,70 90,70 80))

Esboce as *features* do cdg, e coloque etiquetas que indiquem o valor do atributo Tipo.

Solução do exercício



Representar conjunto de dados geográficos por uma tabela com geometria. Exemplo de *feature* com um “buraco”

(POLYGON, *rings* exterior e interior)

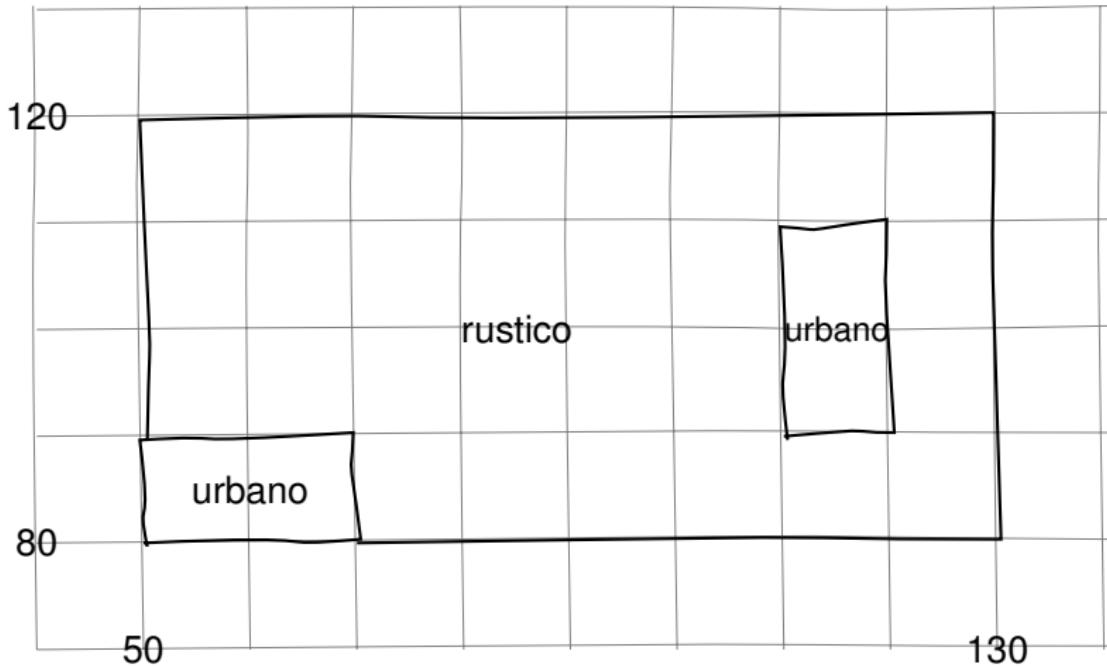
A definição de geometrias do tipo POLYGON, inclui uma ou mais linhas poligonais fechadas (rings) que não se podem intersectar. Existe sempre um único ring exterior. Pode eventualmente existir um ou mais rings interiores que delimitam “buracos” existentes na feature.

Exercício. Considere o seguinte cdg, com geometria em formato WKT. Cada *feature* é do tipo POLYGON, sempre com *ring* exterior e possivelmente com *ring* interior:

Id	Tipo	geometry
1	urbano	POLYGON((50 80,70 80, 70 90,50 90,50 80))
2	rustico	POLYGON((70 80,130 80,130 120,50 120,90 120,70 90,70 80), (110 90,120 90,120 110,110 110,110 90))
3	urbano	POLYGON((110 90,120 90,120 110,110 110,110 90))

Esboce as *features* do cdg, e coloque etiquetas que indiquem o valor do atributo Tipo.

Solução do exercício



Note que a 2a *feature* (*Tipo=rustico*) tem um *ring* exterior e um *ring* interior (um “buraco”). Esse *ring* interior coincide com o *ring* exterior da 3a *feature* (é a mesma linha poligonal fechada).

Validade de objetos geométricos do tipo “polígono”

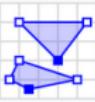
Um objecto geométrico do tipo polígono (`POLYGON`) tem sempre uma linha poligonal envolvente exterior (que irá ser designado por *ring* exterior ou envolvente) e um número arbitrário (0, 1, 2...) de “buracos” (cada um designado por *ring* interior ou envolvido).

Informalmente, as regras para a validade da geometria são as seguintes:

- ① Cada *ring* tem que estar fechado e não pode ter interseções (referido em SIGs como *self-intersection* ou *loop*). Isto visa garantir que quando um *ring* é digitalizado no sentido dos ponteiros do relógio a parte interior fica sempre para a direita e quando é digitalizado no sentido oposto fica sempre para a esquerda;
- ② Os *rings* não se devem cruzar entre si.

Representar objectos geométricos “multipart”

Multipart geometries (2D)

Type	Examples
MultiPoint	 MULTIPPOINT ((10 40), (40 30), (20 20), (30 10)) MULTIPPOINT (10 40, 40 30, 20 20, 30 10)
	 MULTILINESTRING ((10 10, 20 20, 10 40), (40 40, 30 30, 40 20, 30 10))
MultiPolygon	 MULTIPOLYGON (((30 20, 45 40, 10 40, 30 20)), ((15 5, 40 10, 10 20, 5 10, 15 5))) MULTIPOLYGON (((40 40, 20 45, 45 30, 40 40)), ((20 35, 10 30, 10 10, 30 5, 45 20, 20 35), (30 20, 20 15, 20 25, 30 20)))
	 GEOMETRYCOLLECTION (POINT (40 10), LINESTRING (10 10, 20 20, 10 40), POLYGON ((40 40, 20 45, 45 30, 40 40)))

Esta representação usa um formato designado por *Well Known Text* (WKT).

Representar conjunto de dados geográficos *multipart* por uma tabela com geometria do tipo MULTIPOLYGON

(MULTIPOLYGON)

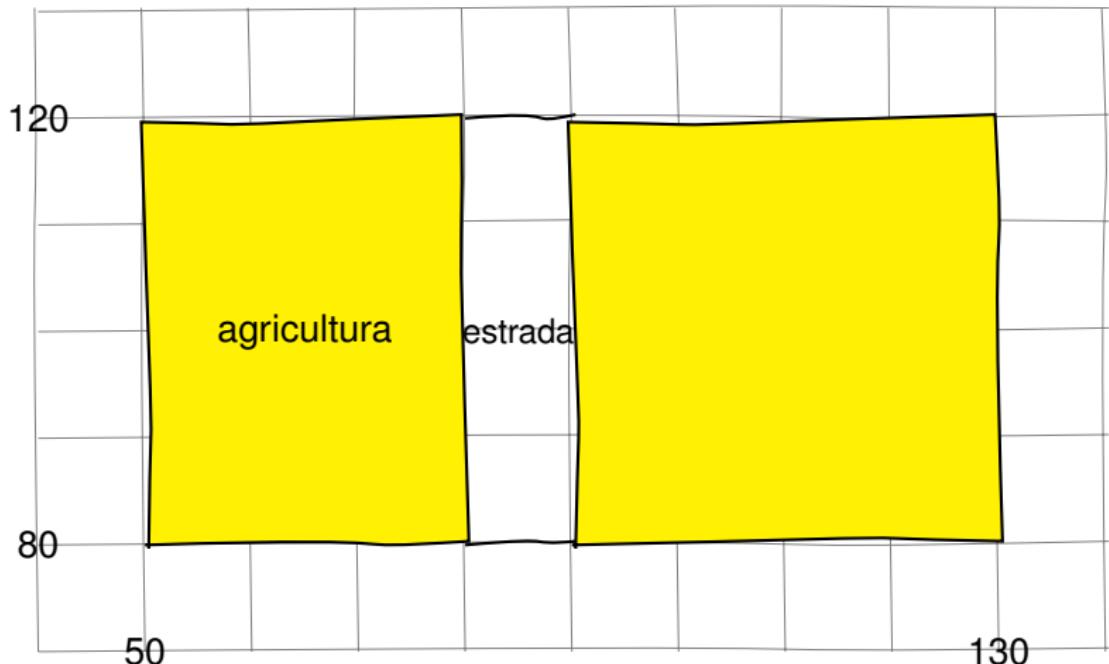
Geometrias do tipo “polígono multipart” são definidas em WKT como $\text{MULTIPOLYGON}(\mathcal{P}_1, \mathcal{P}_2, \dots)$ em que cada \mathcal{P}_i é a representação de um objeto do tipo POLYGON (com ou sem buracos).

Exercício. Considere o seguinte cdg, com geometria em formato WKT. Cada *feature* é do tipo MULTIPOLYGON:

Id	Uso	geometry
1	estrada	$\text{MULTIPOLYGON}((80\ 80, 90\ 80, 90\ 120, 80\ 120, 80\ 80))$
2	agricultura	$\text{MULTIPOLYGON}((50\ 80, 80\ 80, 80, 80\ 120, 50\ 120, 50\ 80), (90\ 80, 130\ 80, 130\ 120, 90\ 120, 90\ 80))$

Esboce as *features* do cdg, e coloque etiquetas que indiquem o valor do atributo Uso.

Solução do exercício



Note que há apenas duas *features*: uma *feature* cuja geometria tem uma única parte (estrada, a branco) e uma *feature* cuja geometria tem duas partes (agricultura, a amarelo).

Conversão de *multipart* em *singlepart*

As aplicações SIG incluem tipicamente uma operação que permite converter *features multipart* em *singlepart*. Tipicamente é usada para garantir que as *features* resultantes são espacialmente conexas.

Exemplo. Num slide anterior foram representadas algumas *features* do cdg de áreas protegidas do ICNF. A operação descrita pelo diagrama seguinte não altera a representação do Parque Natural do Alvão ou da Paisagem Protegida da Serra de Montejunto, dado que são *features singlepart*, mas substitui a *feature* correspondente à Reserva Natural das Berlengas por um conjunto de novas *features singlepart*, sendo cada uma dessas novas *features* espacialmente conexas.

A operação preserva os atributos e os seus valores, replicando-os quando necessário. Neste caso, são criadas várias *features* com designação “Reserva Natural das Berlengas”.



Converter conjunto de dados geográficos com geometria do tipo MULTIPOLYGON para geometria do tipo POLYGON

Exercício. Considere o seguinte cdg, com geometria em formato WKT. Cada *feature* é do tipo MULTIPOLYGON:

Id	Uso	geometry
1	estrada	MULTIPOLYGON(((80 80,90 80, 90 120,80 120,80 80)))
2	agricultura	MULTIPOLYGON(((50 80,80 80,80 120,50 120,50 80)), ((90 80,130 80,130 120,90 120,90 80)))

Qual seria o resultado da conversão *multipart* para *singlepart* do cdg acima? Responda indicando a tabela de atributos ampliada com uma coluna de geometria em formato WKT, em que todas as *features* têm geometria do tipo POLYGON.

Solução do exercício

A geometria descrita por

- MULTIPOLYGON(((50 80, 80 80, 80 120, 50 120, 50 80)),
((90 80, 130 80, 130 120, 90 120, 90 80)))

é convertida em duas geometrias *singlepart*:

- POLYGON((50 80, 80 80, 80 120, 50 120, 50 80)), e
- POLYGON((90 80, 130 80, 130 120, 90 120, 90 80)).

O cdg resultante é:

Id	Uso	geometry
1	estrada	POLYGON((80 80, 90 80, 90 120, 80 120, 80 80))
2	agricultura	POLYGON((50 80, 80 80, 80 120, 50 120, 50 80))
3	agricultura	POLYGON((90 80, 130 80, 130 120, 90 120, 90 80))

Note que POLYGON(...) tem 2 níveis de parênteses enquanto MULTIPOLYGON(((...))) tem 3 níveis de parênteses.

Métodos básicos sobre objectos geométricos: área, perímetro e controide

Todos os objectos geométricos têm algumas propriedades básicas que dependem do seu tipo. Por exemplo:

- ➊ (length) Um objecto do tipo “linha” tem um **comprimento**, que é o comprimento da linha poligonal, ou seja a soma das distâncias entre os seus vértices sucessivos. O comprimento de uma linha *multipart* é a soma dos comprimentos das suas partes:
- ➋ (area) A **área** é uma propriedade de objecto do tipo “polígono”; A área de um objecto *multipart* é a soma das áreas das partes.
- ➌ (centroid) centro de gravidade dos vértices, que pode não estar sobre o objecto;
- ➍ (PointOnSurface) devolve um ponto situado obrigatoriamente sobre o objecto.

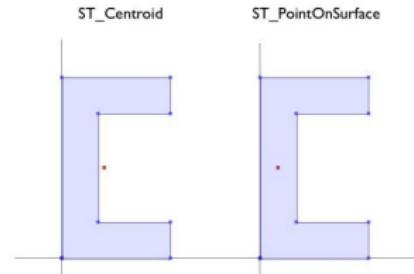


Diagrama de operações: cálculo de área

O cálculo de distâncias e áreas num SIG pode ser realizado sobre um plano, uma esfera ou um elipsoide oblato (achatado).

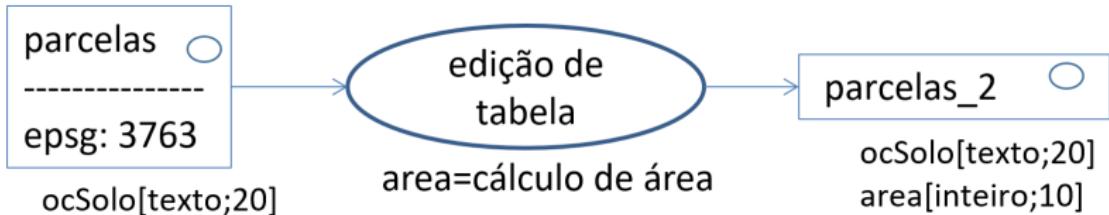
Por omissão, segue-se a recomendação do Open Geospatial Consortium (OGC) é que o cálculo seja realizado sobre o elipsoide global de referência WGS84 (ver Secção sobre sistemas de coordenadas de referência).

Em termos práticos, há portanto duas opções:

- ① Cálcular distâncias e áreas sobre o elipsoide de referência. Os valores estão sempre corretos qualquer que seja o sistema de coordenadas dos dados ou do projeto. Esta é a opção por omissão em QGIS.
- ② Usar o sistema de coordenadas dos dados. O resultado pode ser aproximadamente correto se o sistema de coordenadas for adequado para a região de estudo mas pode haver erros importantes caso não seja.

Diagrama de operações: cálculo de área (cont.)

No diagrama abaixo o CRS do input é ETRS89-TM-PT06 com o código EPSG 3763 (ver Slide 204), para o qual a unidade das coordenadas cartográficas é metro, e a area do output terá também como unidades m², mas o cálculo da área é realizado por omissão sobre o elipsoide. Não é obrigatório indicar o CRS no output pois não é alterado pela operação.



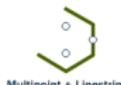
Métodos para testar relações espaciais entre objectos geométricos

A norma OGC/ISO (ver Slide 62) define um conjunto de operadores com valor lógico (**V** ou **F**) para testar **relações espaciais** entre objectos geométricos do tipo “ponto” (p), “linha” (L) ou “polígono” (P).

- ① **Equal**: os objectos coincidem;
- ② **Disjoint**: os objectos não têm pontos em comum;
- ③ **Touches** aplica-se a 2 objectos desde que algum tenha dimensão maior que 0: a Touches b se as fronteiras se intersectam;
- ④ Para a/b de tipo p/L, p/P, L/L or L/P, a **Crosses** b se os objectos se intersectam mas nenhum contém o outro;
- ⑤ a **Within** b se a não intersecta o exterior de b ;
- ⑥ **Overlaps** aplica-se a objectos com mesma dimensão: é **TRUE** se os objectos se intersectam mas nenhum contém o outro;
- ⑦ a **Contains** b se b Within a ;
- ⑧ a **Intersects** b se a e b têm algum ponto em comum;
- ⑨ **Relates** permite definir qualquer relação espacial.

Relações espaciais entre objectos geométricos (ilustração)

Cross



Multipoint & Linestring



Linestring & Linestring



Multipoint & Polygon



Linestring & Multipolygon

Overlap



Multipoint & Multipoint



Linestring & Linestring



Polygon & Polygon

Touch



Point & Linestring



Multipoint & Linestring



Linestring & Linestring



Linestring & Polygon



Point & Polygon



Multipoint & Polygon

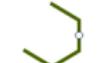
Within/Contains



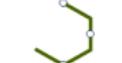
Point & Multipoint



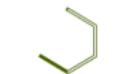
Multipoint & Multipoint



Point & Linestring



Multipoint & Linestring



Linestring & Linestring



Linestring & Polygon



Point & Polygon



Multipoint & Polygon

Disjoint



Point & Multipoint



Multipoint & Multipoint



Point & Linestring



Multipoint & Linestring



Linestring & Linestring



Linestring & Polygon



Multipoint & Polygon



Polygon & Polygon

Intersects



Point & Multipoint



Multipoint & Multipoint



Point & Linestring



Multipoint & Linestring



Linestring & Linestring



Linestring & Polygon



Multipoint & Polygon



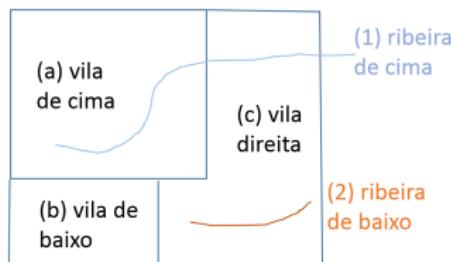
Linestring & Multipolygon

Métodos para testar relações espaciais entre objectos geométricos: resultado como matriz lógica

Considerem-se dois cdg vectoriais

- 1 cdg ribeiras do tipo “linha” com dois objectos *singlepart* (1 e 2) em que cada linha representa uma ribeira;
- 2 cdg freguesias do tipo “políгоно” com 3 objectos *singlepart* (a, b, c) em que cada políгоно representa uma freguesia.

Abaixo apresenta-se o resultado de testar a relação espacial
Intersects: “ribeiras **intersecta** freguesias”

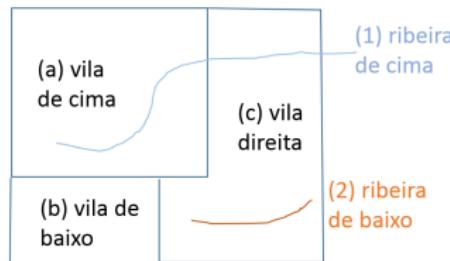


ribeiras

ribeiras	freguesias		
	intersecta	a	b
1	V	F	V
2	F	F	V

Métodos para testar relações espaciais entre objectos geométricos: resultado como matriz lógica (continuação)

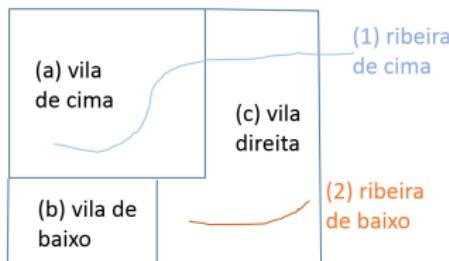
- Relação espacial **Within**: “ribeiras **está contido em** freguesias”



freguesias
ribeiras

está contido em	freguesias		
	a	b	c
1	F	F	F
2	F	F	V

- O exemplo anterior é equivalente a usar a relação espacial **Contains** se se trocar a ordem dos *inputs*: “freguesias **contém** ribeiras”



freguesias
ribeiras

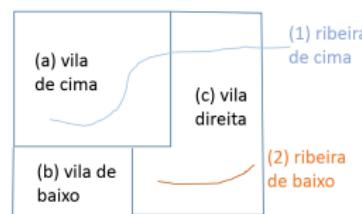
contém	ribeiras	
	1	2
a	F	F
b	F	F
c	F	V

Selecção por localização

No Slide 56 foi definida a operação de selecção sobre cdg vectoriais.

Em complemento à operação de **selecção por atributos** discutida anteriormente, pode usar-se testes a relações espaciais para efectuar uma **selecção por localização**. A diferença com a selecção por atributos é que a expressão lógica usada para fazer a selecção é definida a partir do resultado do teste à relação espacial.

Exemplo: determinar as freguesias que são intersectadas por alguma das ribeiras. A análise do resultado de “freguesias **intersecta** ribeiras”,



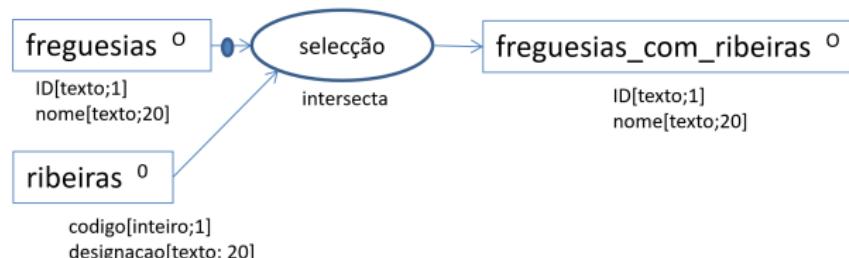
freguesias	ribeiras	1	2
intersecta		1	2
a	V	F	
b	F	F	
c	V	V	

mostra que as freguesias **a** e **c** verificam a relação espacial para **pelo menos um** dos objectos do cdg **ribeiras**. Assim, a selecção das freguesias devolve um novo cdg apenas com as freguesias **a** e **c**.

Diagrama para representar a operação de selecção por localização

A operação de **selecção por localização** é um operação de selecção em que o valor da expressão lógica resulta de um teste a uma relação espacial entre dois cdg vectoriais. Assim essa operação tem dois *inputs*: o cdg em que é realizada a selecção e o cdg auxiliar usado para testar a relação espacial.

O diagrama para exemplo do slide anterior (que usa o teste “freguesias **intersecta** ribeiras”) é:



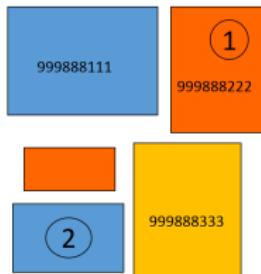
em que o *output* foi designado por *freguesias_com_ribeiras*. Os seus atributos são os mesmos do cdg *freguesias* mas tem tipicamente um menor número de linhas. O símbolo • indica que *freguesias* é o *input* principal, no qual a selecção é efectuada.

Selecção por localização: objectos multipart

Considere o seguinte exemplo com dois cdg.

- ① propriedades, um cdg para do tipo “polígono” para representar o cadastro, em que cada *feature* representa a(s) parcelas(s) de cada proprietário. Existem 3 *features*, duas das quais são *multipart*. Os atributos são o NIF do proprietário e a área da *feature*, determinada como descrito no Slide 79.
- ② pocos, um cdg do tipo “ponto”, em que cada *feature* representa um poço. Os atributos são o código do poço e a sua profundidade em metros.

Pretende determinar-se quais são as propriedades do cadastro que contém algum poço.



propriedades	
NIF	area
999888111	50000
999888222	35000
999888333	40000

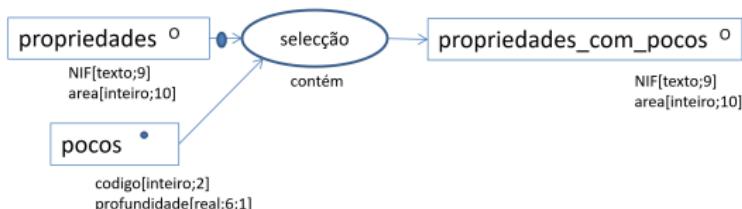
pocos	
codigo	profundidade
1	6.1
2	8.3

Selecção por localização: objectos multipart (continuação)

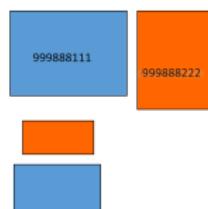
No exemplo anterior a operação de selecção é baseada na relação espacial **contém** (Contains). O resultado do teste “propriedades **contém** pocos” e o diagrama da operação são:

propriedades

pocos		
contém	1	2
999888111	F	V
999888222	V	F
999888333	F	F



Como a 1^a e 2^a features de propriedades verificam a relação **contém** para **alguma feature de pocos**, o cdg propriedades_com_pocos tem essas 2 features e respectivos atributos:



NIF	area
999888111	50000
999888222	35000

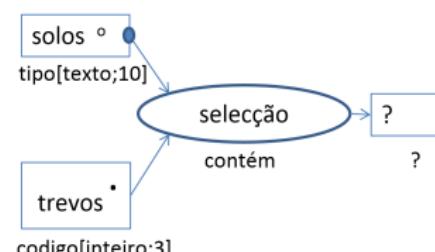
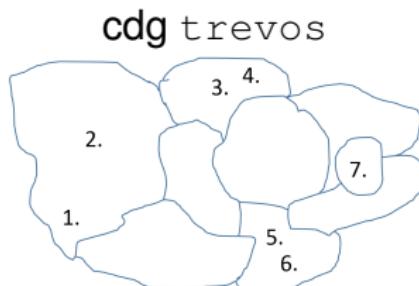
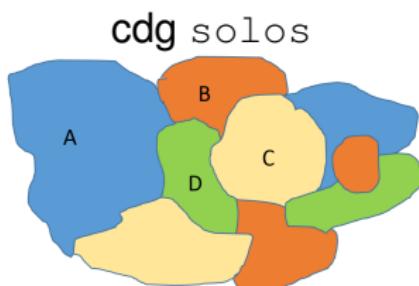
Selecção por localização: exercício

Um SIG contém dois cdg vectoriais representados abaixo:

- 1 um cdg `solos` do tipo “polígono” com 4 *features multipart* que representam 4 tipos de solo; e
- 2 um cdg `trevos` do tipo “ponto” com 7 *features* indicando as localizações de uma espécie de trevos.

Pretende saber-se em que tipos de solo ocorre essa espécie.

Represente o resultado do teste à relação espacial **contém** (Contains) como uma matriz lógica. Represente o *output* da operação de selecção completando o diagrama e indicando quais são as *features* seleccionadas.



Selecção por atributos e por localização: exercícios

Considere dois cdg vectoriais de polígonos concelhos (com os atributos DTCC e designacao, cujos valores são os nomes dos concelhos) e areas_aridas (com o atributo codigo, cujos valores são 1,2, ..., e que enumera as áreas ardidas).

Cada polígono do cdg concelhos representa um concelho de Portugal Continental. Cada polígono de areas_aridas representa uma área ardida num certo ano.

Use operações de selecção para resolver cada um dos problemas abaixo e responda com o correspondente diagrama de operações.

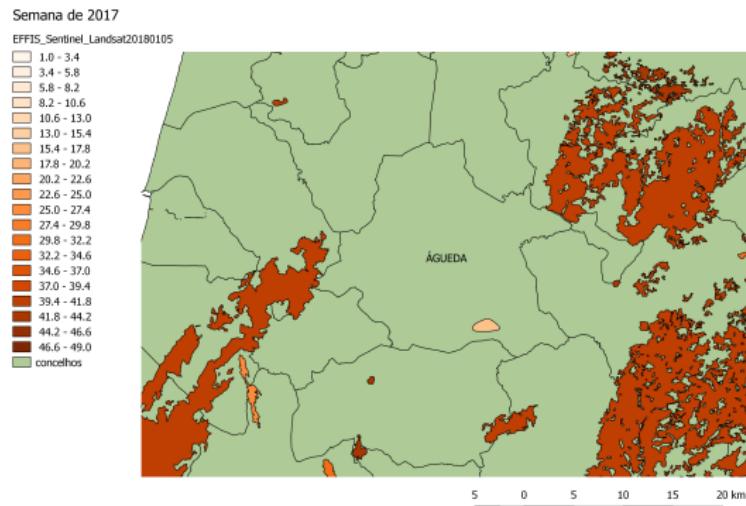
- ① Determinar as áreas ardidas que afectaram o concelho de Águeda;
- ② Determinar os concelhos que foram atingidos pela área ardida 52;
- ③ Determinar as áreas ardidas que não atingiram mais do que um concelho;
- ④ Determinar quais os concelhos que não foram atingidos por incêndios.

Selecção por atributos e por localização: solução exercício

Nota: os dados reais para o exercício podem ser obtidos no Registo Nacional de Dados Geográficos (SNIG):

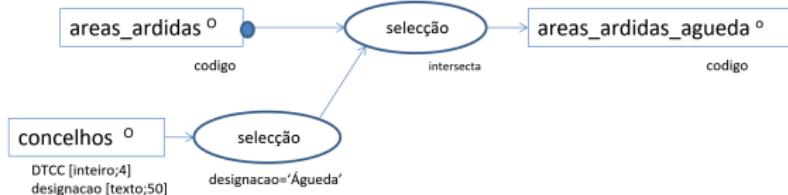
- 1 Os concelhos podem ser identificados na Carta Administrativa Oficial de Portugal (DGT)
- 2 As áreas ardidas podem ser encontradas na Cartografia Nacional de Áreas Ardidas (ICNF)

Na figura abaixo, a legenda corresponde às semanas do ano 2017:

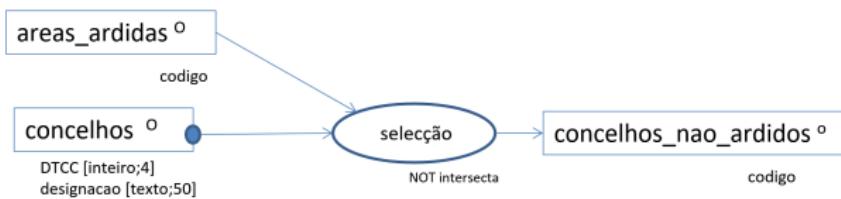


Selecção por atributos e por localização: solução exercício

Alínea 1 (áreas ardidas em Águeda):



Alínea 4 (concelhos não atingidos por incêndios):



Neste último caso, o operador lógico **NOT** indica que o vector lógico resultante do teste à relação espacial `intersecta` é negado. Assim, a selecção dos concelhos é invertida, e os concelhos que não intersectam nenhuma área ardida são devolvidos.

Métodos que suportam análise espacial

A norma OGC/ISO (ver Slide 62) define operadores que devolvem a **distância** entre objectos geométricos ou **novos objectos geométricos** de acordo com alguma relação espacial. Dados objectos *a* e *b*:

- ① **Distance** devolve a distância mais curta entre pontos de *a* e de *b*;
- ② **Buffer** de *a* com distância *d* devolve o objecto geométrico cujos pontos estão uma distância menor ou igual a *d* de *a*;
- ③ **ConvexHull** de *a* devolve um objecto geométrico que é o invólucro convexo de *a*;
- ④ **Intersection** devolve o objecto geométrico cujos pontos estão em *a* e em *b*;
- ⑤ **Union** devolve o objecto geométrico cujos pontos estão em *a* ou em *b*;
- ⑥ **Difference** entre *a* e *b* devolve o objecto geométrico com os pontos que estão em *a* mas não em *b* (operação não comutativa);
- ⑦ **Symmetric Difference** entre *a* e *b* devolve o objecto geométrico cujos pontos satisfazem **apenas uma** das condições: (1) estão em *a* mas não em *b*; (2) estão em *b* mas não em *a*.

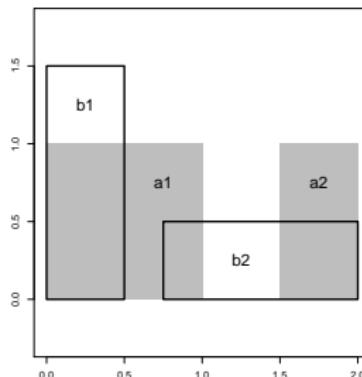
Métodos que suportam análise espacial: distância

O método **Distance** permite calcular distâncias entre objectos geométricos de dois cdg. O resultado é uma **matriz de distâncias**, das distâncias entre objectos *a* do cdg A e os objectos *b* do cdg B.

O exemplo abaixo mostra dois cdg do tipo “polígono” no mesmo sistema de coordenadas:

- 1 cdg A com objectos *singlepart* *a*₁ e *a*₂ (a cinzento),
- 2 cdg B, com objectos *singlepart* *b*₁ e *b*₂ (só contorno)

À direita mostra-se a matriz de distâncias que indica que neste exemplo apenas o par (*b*₁, *a*₂) tem uma distância não nula.

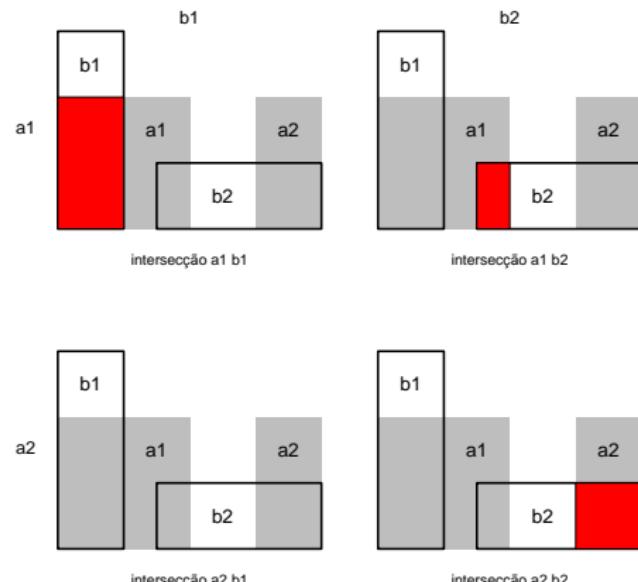
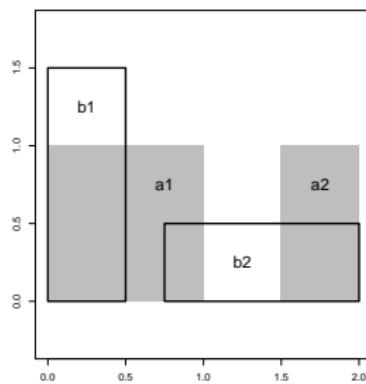


distância	b1	b2
a1	0.00	0.00
a2	1.00	0.00

Métodos que suportam análise espacial: Intersection

O método **Intersection** cria novos objectos geométricos derivados dos objectos geométricos existentes. É aplicado a cada par de objectos (a, b) , em que a pertence ao cdg A e b pertence ao cdg B.

Neste exemplo com 4 pares de objectos *singlepart*, o *output* de **Intersection** é constituído pelos 3 objectos a vermelho abaixo (pois a_2 e b_1 não se intersectam).

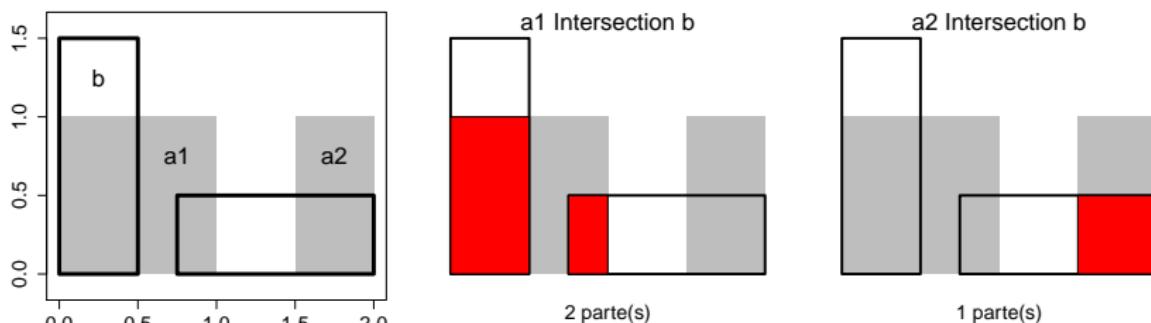


Métodos que suportam análise espacial: Intersection com objectos geométricos *multipart*

No exemplo abaixo,

- o cdg A tem dois objectos geométricos *singlepart* a_1 e a_2 (correspondentes portanto a duas *features*), e
- o cdg B tem um objecto geométrico *multipart* b (corresponde a uma única *feature*).

Como anteriormente a operação é aplicada a todos os pares (a,b) de objectos (a em A e b em B). Neste caso há dois pares (a_1,b) e (a_2,b) e o *output* tem os dois objectos geométricos a vermelho abaixo, sendo o objecto à esquerda *multipart* e o à direita *singlepart*.

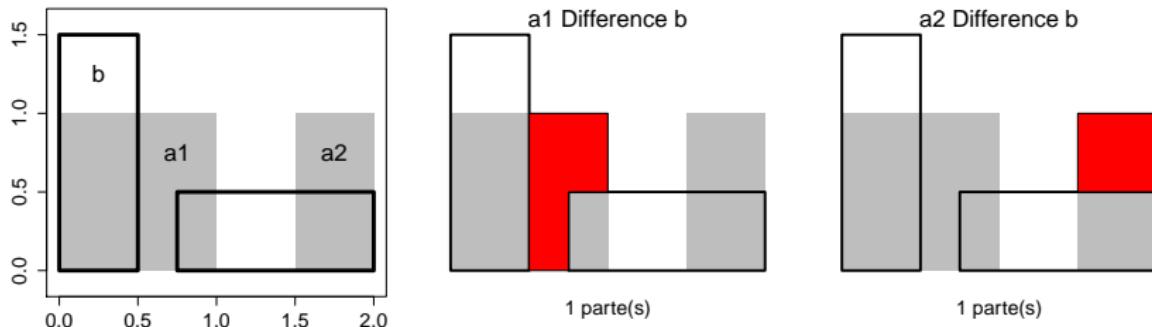


Métodos que suportam análise espacial: Difference

O método **Difference** é também aplicado a todos os pares de objectos de A e B e **depende da ordem dos inputs**. No exemplo abaixo,

- o cdg A tem dois objectos geométricos *singlepart* a_1 e a_2 (correspondentes portanto a duas *features*), e
- o cdg B tem um único objecto geométrico *multipart* b (corresponde portanto a uma única *feature*).

Como neste caso há dois pares (a_1, b) e (a_2, b) de objectos, o *output* de **Difference** tem os dois objectos a vermelho abaixo, sendo neste caso ambos *singlepart*.

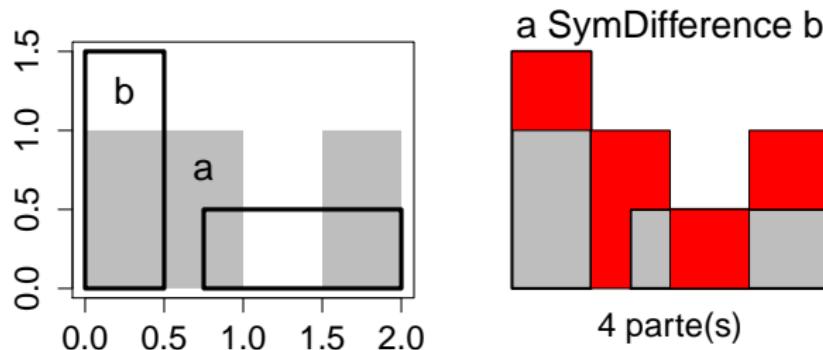


Métodos que suportam análise espacial: Symmetric Difference

Ao contrário de Difference, o resultado do método Symmetric Difference não depende da ordem dos inputs. No exemplo abaixo,

- o cdg A tem um único objecto geométrico *multipart* a (corresponde portanto a uma única *feature*), e
- o cdg B tem um único objecto geométrico *multipart* b (corresponde também a uma única *feature*).

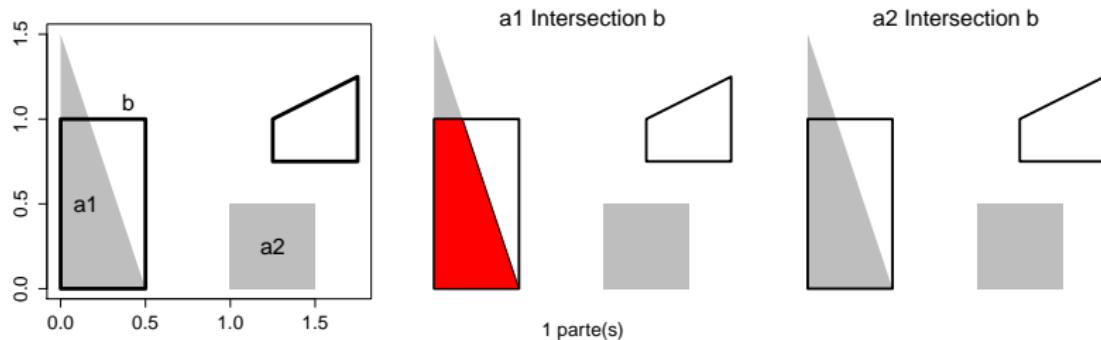
A operação é então aplicada ao único par (a,b) de objectos (a em A e b em B). Neste exemplo, o *output* tem um único objecto *multipart* com 4 partes: as regiões em a e não em b e as regiões em b e não em a.



Métodos que suportam análise espacial: exercício

Considere o cdg A abaixo, com objectos geométricos *singlepart* *a1* e *a2*, a cinzento, e o cdg B, com um único objecto geométrico *multipart* *b* representado pelos 2 polígonos só com contorno. A vermelho, está representado o resultado do método **Intersection**.

- 1 Determine a matriz de distâncias entre A e B;
- 2 Esboce o resultado de “A **Difference** B”;
- 3 Mostre que o resultado de **Symmetric Difference** sobre *a1* e *b* tem 3 partes.



Métodos que suportam análise espacial: *buffer*

A determinação de *buffers* é uma operação muito usada em SIGs pois permite definir a **área de influência** ou **área de protecção** em torno de uma entidade geográfica.

No exemplo mostra-se um *buffer* de 200 m em torno da Paisagem Protegida da Serra de Montejunto. Como discutido no Slide 66, o objecto geométrico que representa esta entidade tem uma parte e 5 *buracos*. Neste exemplo, o *buffer* reduz ou mesmo elimina esses *buracos*.



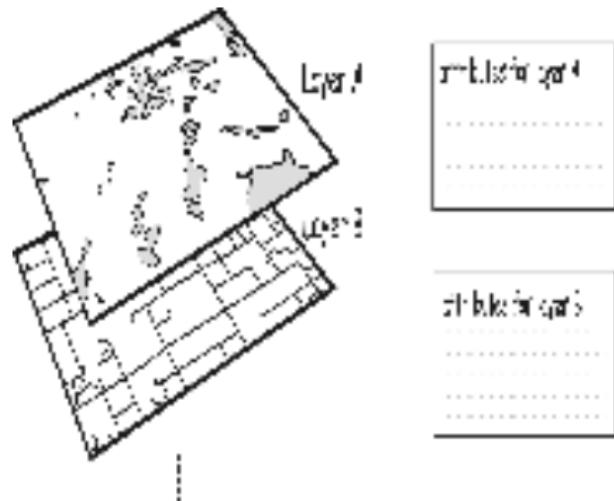
Notas: O argumento “distância” pode ser negativo, dando origem a regiões mais pequenas do que o *input*. Existem parâmetros adicionais que determinam, por exemplo,

- ① a forma do buffer no extremo de uma linha,
- ② a criação de um *buffer* para cada *feature*, ou a criação de um único objecto geométrico para a totalidade do cdg.

Sobreposição vectorial (*map overlay*)

Todas os métodos que suportam análise espacial vistos anteriormente *operam sobre os objectos geométricos mas não operam sobre as tabelas de atributos.*

Designa-se por **sobreposição vectorial** qualquer operação espacial genérica sobre um ou mais cdg com o mesmo sistema de coordenadas, que se aplica às *features* (objectos geométricos e atributos).



Sobreposição vectorial envolve portanto:

- 1 anáise espacial de objectos geométricos dos cdg (ver Slido 04):

Sobreposição vectorial (continuação)

Cada aplicação de SIG (e.g. QGIS, ArcGIS, GeoMedia, etc) contém um grande número de funções de sobreposição vectorial que respondem às necessidades sentidas pelos utilizadores para análise geoespacial.

Algumas dessas operações são muito comuns e são disponibilizadas nas diferentes aplicações de SIG:

- ① **intersecção** (Slide 105);
- ② **união** (Slide 106);
- ③ **geração de buffers** (Slide 111);
- ④ **dissolução** (Slide 115);
- ⑤ **recorte** (clip, ver Slide 116);
- ⑥ **diferença** (Slide 121).

OPERATION	PRIMARY LAYER	OVERLAY POLYGONS	RESULT				
CLIP	<table border="1"><tr><td>1</td><td>2</td></tr><tr><td>3</td><td>4</td></tr></table>	1	2	3	4		
1	2						
3	4						
ERASE	<table border="1"><tr><td>1</td><td>2</td></tr><tr><td>3</td><td>4</td></tr></table>	1	2	3	4		
1	2						
3	4						
SPLIT	<table border="1"><tr><td>1</td><td>2</td></tr><tr><td>3</td><td>4</td></tr></table>	1	2	3	4		
1	2						
3	4						
IDENTITY	<table border="1"><tr><td>1</td><td>2</td></tr><tr><td>3</td><td>4</td></tr></table>	1	2	3	4		
1	2						
3	4						
UNION	<table border="1"><tr><td>1</td><td>2</td></tr><tr><td>3</td><td>4</td></tr></table>	1	2	3	4		
1	2						
3	4						
INTERSECT	<table border="1"><tr><td>1</td><td>2</td></tr><tr><td>3</td><td>4</td></tr></table>	1	2	3	4		
1	2						
3	4						

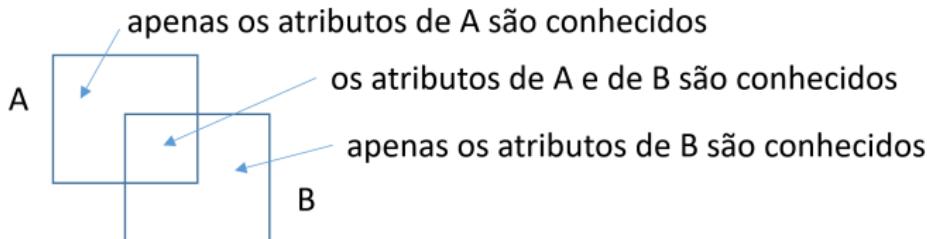
Sobreposição vectorial: intersecção vs união

Ambas as operações de sobreposição vectorial **intersecção** e **união preservam os atributos** dos seus *inputs*. A tabela de atributos do resultado da operação contém por isso todos os atributos dos *inputs*.

A extensão espacial do resultado da operação é:

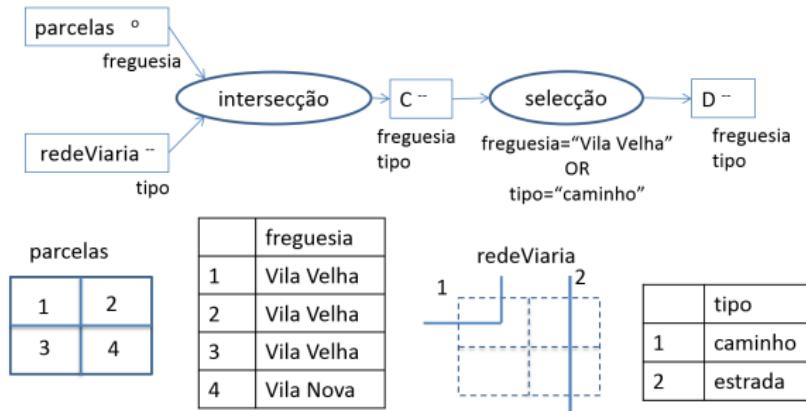
- Para a **intersecção** dos cdg A e B: é conjunto de localizações que estão em A **e** em B.
- Para a **união** dos cdg A e B: é conjunto de localizações que estão em A **ou** em B.

Para a **união**, pode haver localizações onde apenas se conhecem os atributos de A ou de B. Nessas localizações os atributos desconhecidos ficam indefinidos (`null`):



Sobreposição vectorial: intersecção

A operação de intersecção define novas *features* pelo método respetivo (ver Slide 96). A operação **preserva todos os atributos dos dois cdg de input** como ilustrado no diagrama abaixo.



O *output* da operação de intersecção é um cdg com os atributos designacão e tipo:

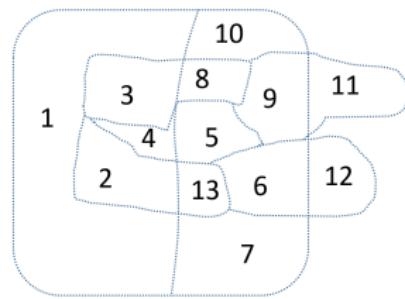
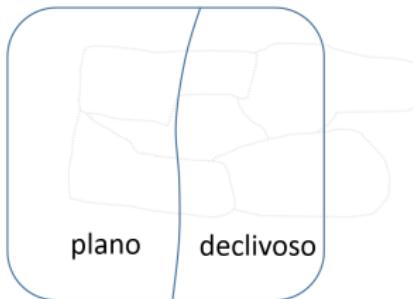
C		freguesia	tipo
	1	Vila Velha	caminho
	2	Vila Velha	estrada
	3	Vila Nova	estrada

D		freguesia	tipo
	1	Vila Velha	caminho
	2	Vila Velha	estrada

Sobreposição vectorial: união

A operação de união define novas *features* que espacialmente correspondem ao resultado de **Intersection** (ver Slide 96) e **Difference** (ver Slide 98). **Todos os atributos dos dois inputs estão no output**.

No exemplo abaixo, o 1º cdg tem atributo `ocSolo` e o 2º cdg tem atributo `clasDeclive`. O terceira figura representa os objectos geométricos resultante da aplicação da operação de **união**.



Na tabela de atributos do cdg de *output*:

- ① 2, 3, 4, 5, 6, 8, 9 e 13 têm valores para ambos os atributos;
- ② 1, 7 e 10 têm apenas valor para atributo `clasDeclive`;
- ③ 11 e 12 têm apenas valor para atributo `ocSolo`.

Sobreposição vectorial: união (continuação)

O exemplo do slide anterior pode ser descrito através do diagrama de operações abaixo, em que o 1º cdg tem nome parcelas e o 2º cdg tem nome declives.

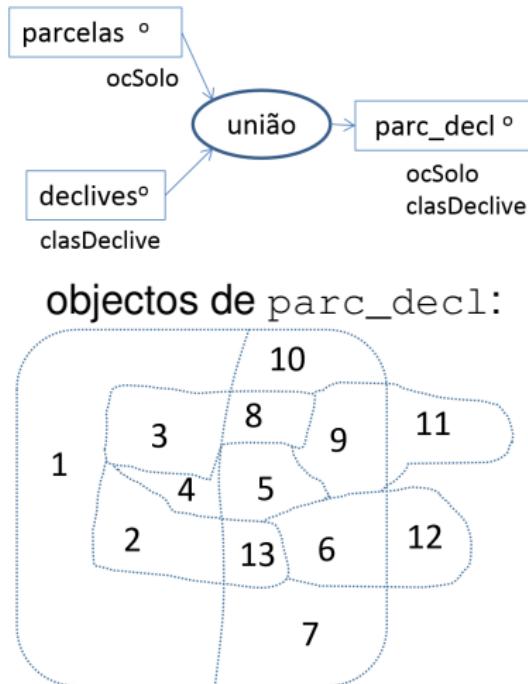
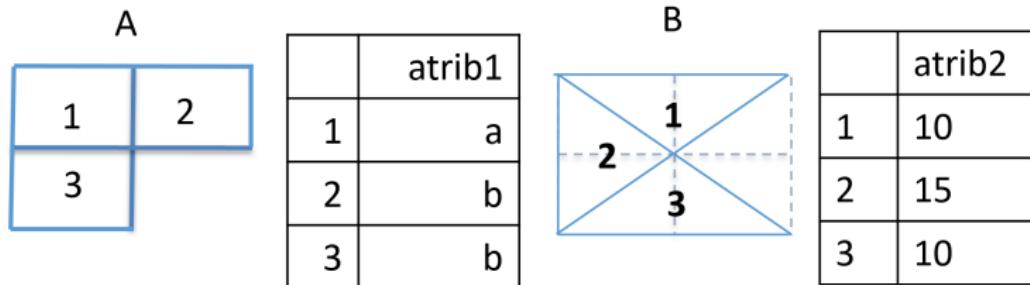


tabela de atributos de parc_decl:

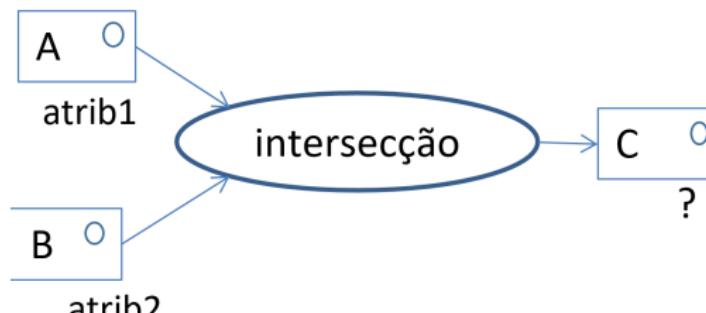
	ocSolo	clasDeclive
1	null	plano
2	mato	plano
3	floresta	plano
4	vinha	plano
5	vinha	declivoso
6	floresta	declivoso
7	null	declivoso
8	floresta	declivoso
9	milho	declivoso
10	null	declivoso
11	milho	null
12	floresta	null
13	mato	declivoso

Sobreposição vectorial: exercício de exame

Considere os conjuntos de dados geográfico (cdg) de polígonos A e B e as suas respectivas tabelas de atributos. As coordenadas geográficas dos 3 vértices (superior esquerdo e direito, e inferior esquerdo) da região representada pelo cdg A coincidem com os de B.



O cdg C é obtidos pela execução do diagrama de operações seguinte. Represente o cdg C a sua tabela de atributos.



Resolução da 1a parte do exercício do slide anterior

A intersecção dos cdg A e B faz-se em dois passos:

- Aplicação do método **Intersection** (ver Slide 96) para determinar os **5** objectos geométricos do cdg C (em baixo, a vermelho);
- Preenchimento da tabela de atributos com os valores da tabela de A e da tabela de B correspondentes aos **5** objectos criados.

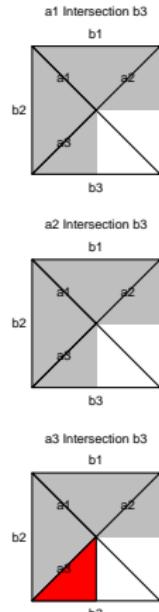
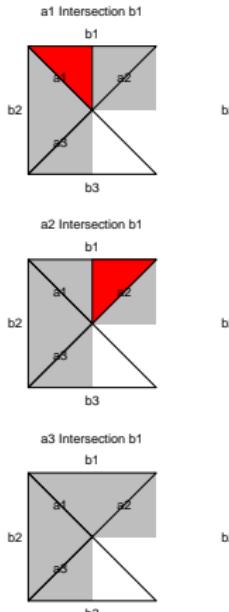


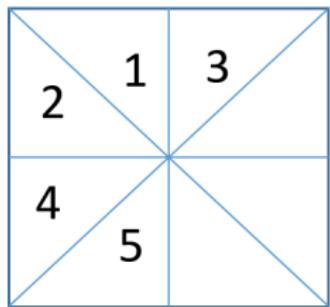
tabela de atributos de C:

	atrib1	atrib2
(a1,b1):	1	a
(a1,b2):	2	a
(a2,b1):	3	b
(a3,b2):	4	b
(a3,b3):	5	b

Nota: as intersecções **pontuais** (e.g. a1 com b3) não originam polígonos. Assim, não originam objectos de C dado que C é um cdg vectorial do tipo “polígono”.

Solução do exercício

- Os objectos geométricos do cdg C e a tabela de atributos são:

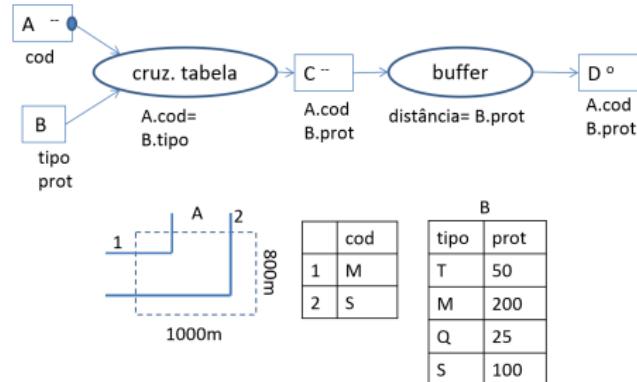


	atrib1	atrib2
1	a	10
2	a	15
3	b	10
4	b	15
5	b	10

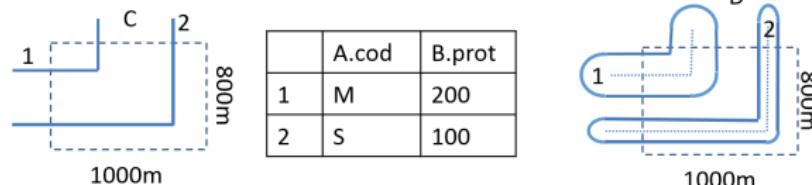
Sobreposição vectorial: geração de buffers

Esta operação tem um único cdg de *input* que pode ser do tipo “ponto”, “linha” ou “polígono”. A distância para buffer pode ser: (1) um parâmetro fixo, ou (2) pode ser definida por um dos atributos do cdg.

Considere, por exemplo, o diagrama e os cdg A e B seguintes:



Os objectos geométricos e atributos de cdg C e D são então:



Sobreposição vectorial: exame de 8 de junho de 2016

Suponha que um SIG tem o seguintes cdg:

- Um cdg do tipo “polígono” com os limites de povoações, código, nome e número de habitantes:
povoacoes (codPov, nomePov, numHab)
- Um cdg do tipo “ponto” com a localização de árvores notáveis, o seu código e a espécie:
notaveis (ID, especie)
- Um cdg do tipo “linha” com a localização, código e nome de rios:
rios (codigo, nomeRio)

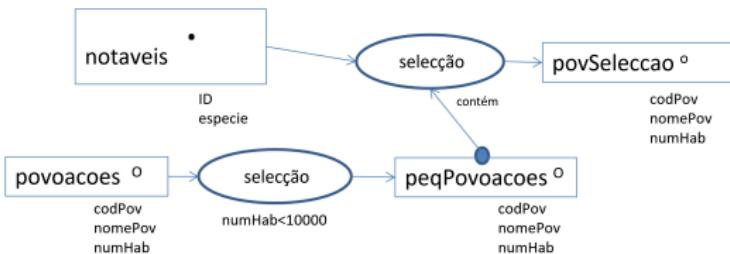
Para cada uma das questões abaixo, apresente um diagrama de operações que mostre como se obtém a respectiva resposta. O diagrama de operações deve indicar os nomes dos conjuntos de dados, o seu tipo e os seus atributos atributos, e também nomes, parâmetros e prioridade (quando se aplique) das operações. As operações que pode usar são *selecção*, *geração de buffers* e *intersecção*.

Sobreposição vectorial: exame de 8 de junho de 2016 (cont.)

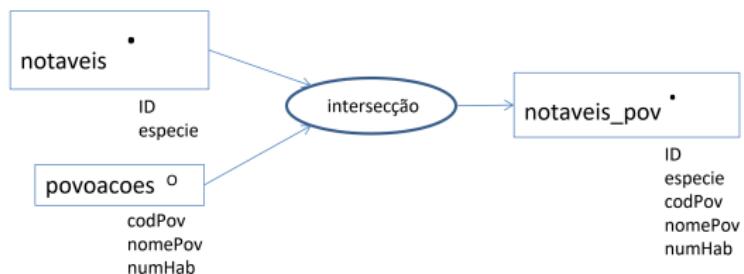
- 1 Identificar as povoações com menos de 10000 habitantes que contém pelo menos uma árvore notável;
- 2 Associar a cada árvore notável o nome da povoação a que pertence, caso pertença a alguma (i.e. a tabela de atributos do resultado deverá ter as colunas ID e nomePov);
- 3 Determinar as árvores notáveis que se situam a menos de 100 m de um rio.

Sobreposição vectorial: uma solução possível para o exercício

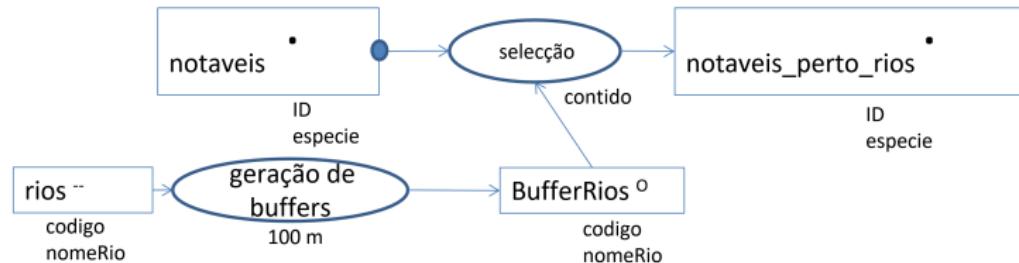
1



2



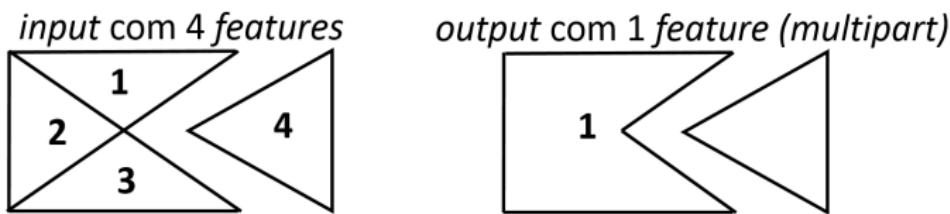
3



Sobreposição vectorial: dissolução

A operação de **dissolução** tem um único cdg de *input* e pode ter como parâmetro adicional um dos seus atributos.

- ➊ **Dissolução sem atributo.** Caso não haja parâmetro adicional, o *output* tem simplesmente uma *feature* (possivelmente *multipart*) que corresponde à totalidade das *features* do *input*, com **as fronteiras interiores eliminadas**:



- ➋ **Dissolução por um atributo.** Neste caso, a operação descrita na alínea anterior é **aplicada a cada subconjunto das features com valor idêntico para esse atributo**. O tabela de atributos do *output* preserva apenas esse atributo e tem tantas linhas como valores distintos que o atributo toma (ver exemplo em Slide 119).

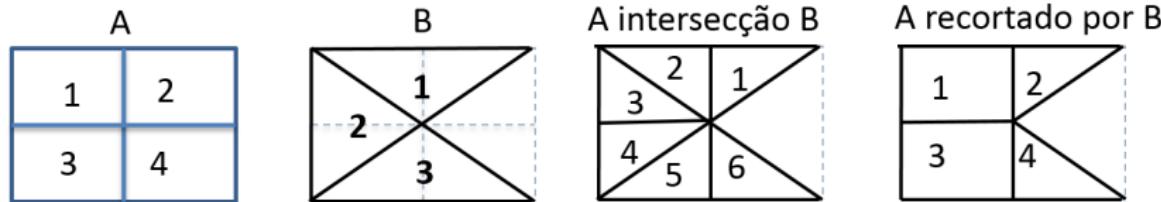
Sobreposição vectorial: recorte

A operação de **recorte**, conhecida pela designação **clip**, tem dois *inputs*: um *input* principal e um *input* auxiliar.

As operações de **recorte** (*clip*) e de **intersecção** originam um *output* com extensão espacial idêntica, mas o *output* da operação de **intersecção** tipicamente **contém um maior número de features** e um maior número de atributos.

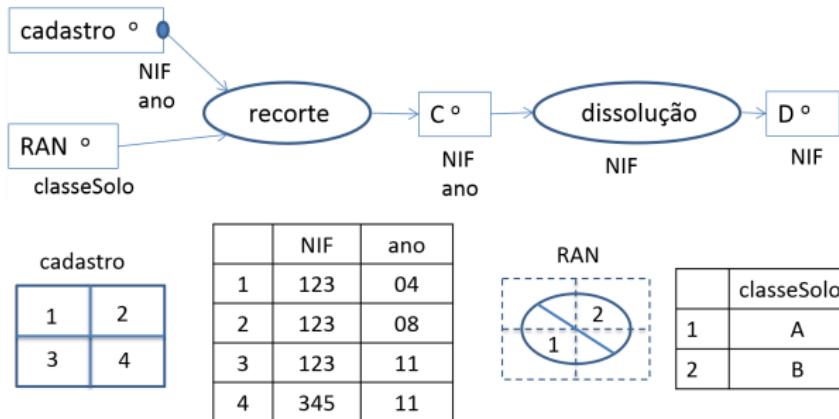
No exemplo abaixo, o cdg A tem 4 *features* e o cdg B tem 3 *features*.

- A intersecção entre A e B contém todas as intersecções entre os objectos de A e os objectos de B (ver Slide 96), mas
- O **recorte** usa B como se fosse uma única *feature multipart* (ver Slide 97), i.e. **usa a dissolução de B (sem atributo) em vez de B**.

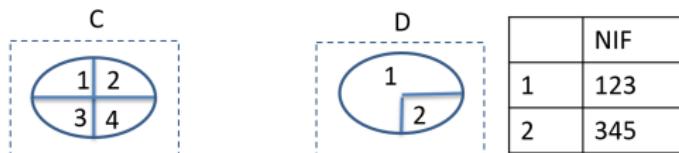


Sobreposição vectorial: exemplo de recorte e dissolução

Neste exemplo, faz-se o recorte do cdg `cadastro` pelo cdg `RAN` (`cdg auxiliar`). Depois, aplica-se a operação de dissolução usando `NIF`.



Note-se que os objectos de `RAN` são irrelevantes para o *output* do **recorte**: só a região definida por `RAN` interessa. Consequentemente, a tabela de atributos de `C` é igual à tabela de atributos de `cadastro`.

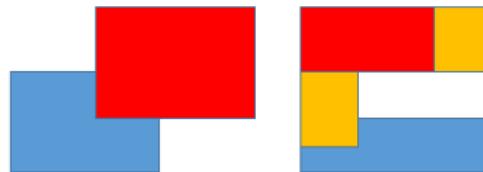


Sobreposição vectorial: dissolução e *multipart* para *singlepart*

Considere o seguinte diagrama de operações sobre o *input* do cdg de tipo “polígono” `regiaoAdeq`:



input (3 objectos *multipart*):



output (2 objectos *singlepart*):



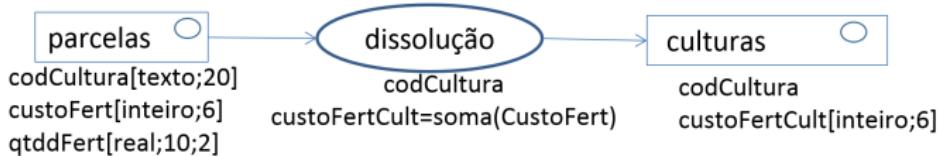
Exercício:

- ① Complete o diagrama;
- ② Explique porque é que os objectos do *output*
 - (a) não tem fronteiras em comum (i.e não se “tocam”);
 - (b) são, cada um deles, espacialmente conexos.

Sobreposição vectorial: dissolução e agrupamento

A operação de **dissolução** origina novas *features* que resultam da agregação de *features* do *input*. Um agregado corresponde a um subconjunto de linhas da tabela de atributos.

A operação de dissolução permite **aplicar uma função aos valores de um atributo** de cada agregado como no exemplo seguinte.



Para cada *feature* do cdg *culturas*,

- 1 o objecto geométrico resulta da dissolução de todas as parcelas com o mesmo valor de *codCultura*;
- 2 o atributo *codCultura* toma essa valor comum;
- 3 o novo atributo *custoFertCult* vai conter a soma de todos os valores de *custoFert* para todas as parcelas com a mesma cultura.

Sobreposição vectorial: dissolução e agrupamento (continuação)

Para o exemplo do slide anterior, as tabelas de atributos dos cdg parcelas e cultura poderiam ser algo como as seguintes tabelas:

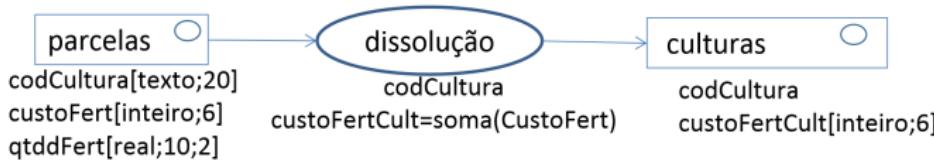
cdg parcelas

	codCultura	custoFert	qtddFert
1	cenoura	2345	250.0
2	batata	1200	220.0
3	nabo	2100	120.0
4	cenoura	4250	320.0
5	cenoura	3450	340.0
6	batata	1430	180.0
7	batata	1520	195.0

cdg culturas

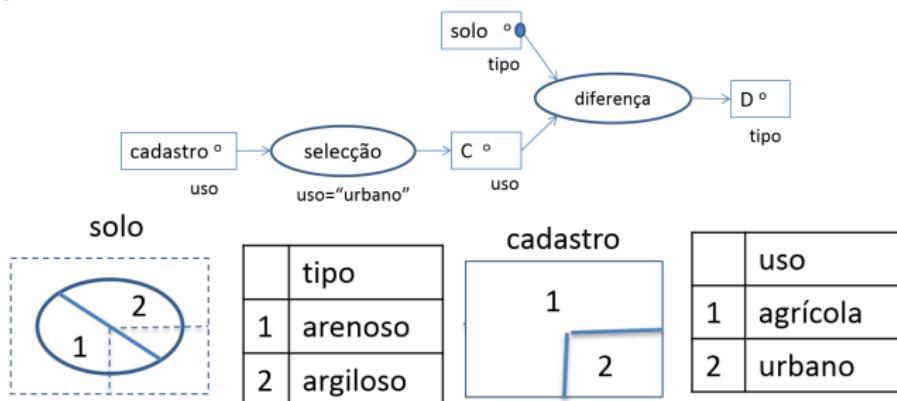
	codCultura	custoFertCult
1	cenoura	10045
2	batata	4150
3	nabo	2100

Confirmar que as tabelas são consistentes com o diagrama da operação:



Sobreposição vectorial: diferença

A operação de **diferença** é a operação complementar à operação **recorte**. Assim, enquanto o *input* auxiliar em **recorte** delimita a região a preservar, o *input* auxiliar da operação **diferença** delimita a região a **eliminar**. Consequentemente, tal como o **recorte**, **diferença** aplica implicitamente **dissolução** ao *input* auxiliar antes de aplicar o método **difference** (ver Slide 98). Os atributos do cdg de *input* principal e do cdg de *output* são idênticos.



solo é o *input* principal da operação de **diferença**. O resultado é:

C

D

Sobreposição vectorial: comparação de operações

Operação	Inputs	Features de output geometria	atributos
Interseção	A,B	intersecções de features de A e de B	atributos de A e de B
União	A,B	idem, mais todas as diferenças A-B e B-A	atributos de A e de B
Recorte ("clip")	A prioritário B auxiliar	recorte das geometrias em A por 'dissolução de B'	atributos de A
Diferença	A prioritário B auxiliar	diferença das geometrias em A por 'dissolução de B'	atributos de A
Buffer	A, atributo ou distância fixa	envolvente de A, sempre do tipo "polígono"	atributos de A
Dissolução	A	1 única feature	sem atributos
Dissolução por atributo	A e atributo	1 feature por valor do atributo	atributo de input
Dissolução c/ agrupamento	A e atributo1 função f e atributo2	1 feature por valor do atributo1	atributo1 $f(\text{atributo2})$

Análise de dados matriciais (*raster*)

Conjuntos de dados geográficos matriciais (*raster*)

O formato matricial é em geral o mais adequado para representar **fenómenos distribuídos no espaço de forma contínua** (ver discussão no Slide 16).

O formato matricial organiza a informação geográfica numa **grelha regular** como descrito no Slide 22.

Os dados geográficos em formato matricial surgem tipicamente nas seguintes situações:

- ① Imagens obtidas por detecção remota;
- ② Dados geográficos compilados numa grelha regular;
- ③ Amostragem regular de dados com distribuição contínua;
- ④ Interpolação espacial de dados pontuais (e.g. precipitação interpolada a partir do observações em estações meteorológicas).

Representação *raster* versus vectorial

A representação vectorial é considerada em geral mais adequada para representar **formas geométricas** pois permite alteração de escala da representação sem perda de precisão e de contraste. A conversão de formato vectorial em *raster* pode ser problemática dada a resolução fixa (tamanho do pixel) do formato *raster*.



A representação precisa dos contornos de entidades geográficas isoladas em formato *raster* pode exigir uma resolução muito fina, correspondente a ficheiros de dados muito pesados. Em contrapartida, as operações de sobreposição de dados são em geral mais eficientes no formato *raster* (**sobreposição matricial**) do que no formato vectorial (**sobreposição vectorial**).

Formatos de dados do tipo matricial (*raster*)

Existem muitos formatos de dados do tipo *raster*, que permitem guardar a informação dos valores de uma imagem de forma eficiente. Os SIGs tem a capacidade de ler muitos formatos distintos.

Um conjunto de dados geográfico matricial (*raster*) é também designado por **imagem** georreferenciada. Como uma imagem tem n linhas e m colunas, o número de pixels (ou células) é $n \times m$, que pode ser um número muito grande. Isto significa que os ficheiros de dados podem ser pesados.

No entanto, dado que existem em geral repetições dos valores dos pixels pode ser mais eficiente armazenar a informação num formato que não é simplesmente uma matriz $n \times m$ de valores.

O formato de dados do tipo *raster* mais comum na actualidade é o formato GeoTIFF discutido no Slide 24.

Principais características de um cdg matricial

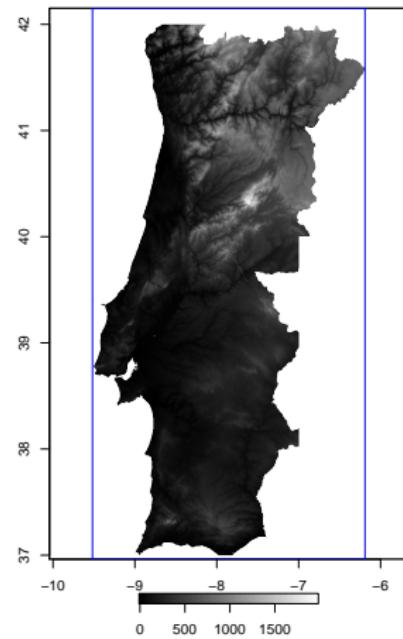
Um cdg matricial é sempre um arranjo regular de pixels que é caracterizado por:

- ① Número de linhas (n) e número de colunas (m) da imagem;
- ② Sistema de coordenadas associado ao conjunto de dados; associa cada posição (i, j) da imagem, em que i é o índice da linha e j é o índice da coluna, uma coordenada geográfica ou cartográfica (x, y) num dado sistema de coordenadas de referência;
- ③ Resolução espacial: distância no terreno que correspondem ao lado do pixel; como as duas dimensões podem não ser iguais, a resolução é geralmente um vector com duas componentes com a resolução na direcção do x e dos y ;
- ④ Extensão da imagem: coordenadas dos cantos da imagem no sistema de coordenadas de referência;
- ⑤ Domínio de valores: conjunto numérico ao qual pertencem os valores da imagem;
- ⑥ null: símbolo arbitrário para os pixels sem valor.

Principais características de um cdg matricial: exemplo

Porção do modelo digital de elevações SRTM (3 arco-segundos) que cobre a quase totalidade de Portugal Continental:

- ① $n = 6231, m = 3993,$
 $n \times m = 24'880'383$ pixels;
- ② Sistema de coordenadas: longitude e latitude em WGS84;
- ③ Resolução espacial: 0.0008333853° nas duas direcções;
- ④ Extensão da imagem:
 $[9.517872^{\circ}W, 6.190164^{\circ}W] \times [36.96097^{\circ}N, 42.15346^{\circ}N];$
- ⑤ Existem $11'519'664$ pixels sem valor associado (i.e. com valor null) que correspondem essencialmente às zonas com água (sem elevação dada pelo SRTM);

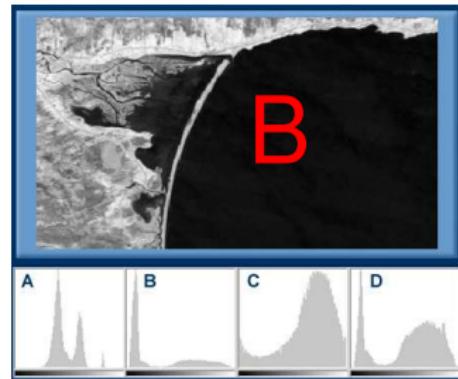
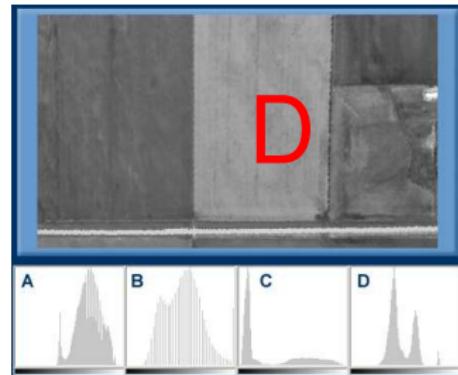
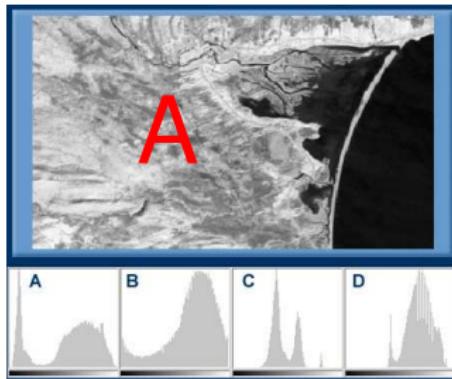


- ⑥ O domínio de valores é $[-20, 1991].$

Histogramas de imagens: exercício

O **histograma** de uma imagem é a distribuição das frequências dos valores dos pixels da imagem.

Exercício: Nas figuras junto, os pixels com menor valor são representados a escuro. Identifique o histograma correcto para cada imagem.



CDG matriciais: uso do histograma para alterar o contraste

(Intensidade de um pixel)

É uma função crescente do valor do pixel que toma valores entre 0 e 1. Tipicamente, a intensidade 0 é representado em preto e a intensidade 1 é representada em branco.

Seja z_{\max} um valor tal que todos os pixels com valor maior do que z_{\max} têm intensidade 1 (pixels brancos). Da mesma forma define-se z_{\min} como o valor tal que todos os pixels com valor abaixo de z_{\min} têm intensidade 0 (pixels pretos).

Assim, a imagem apenas apresenta contraste para valores entre z_{\min} e z_{\max} . A análise do histograma da imagem é útil para escolher z_{\min} e z_{\max} e alterar o contraste. Diz-se então que se faz uma transformação (*stretching*) do histograma.

CDG matriciais: uso do histograma para alterar o contraste – exemplo

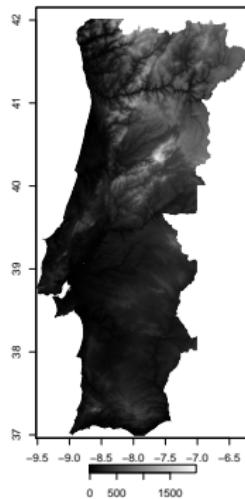


Imagen original sem modificação de contraste.

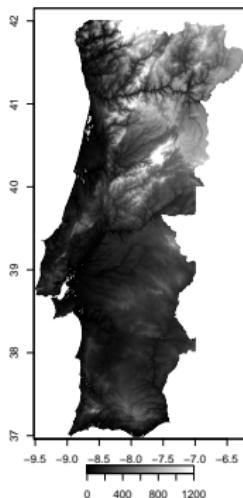
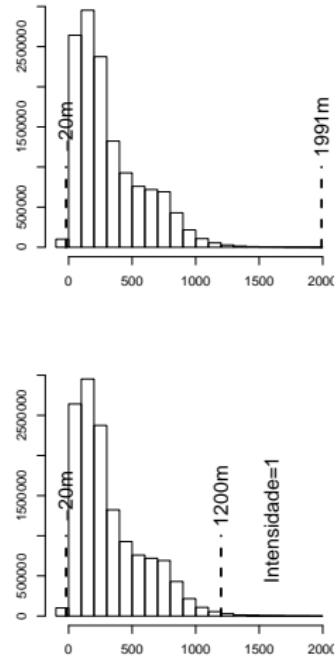
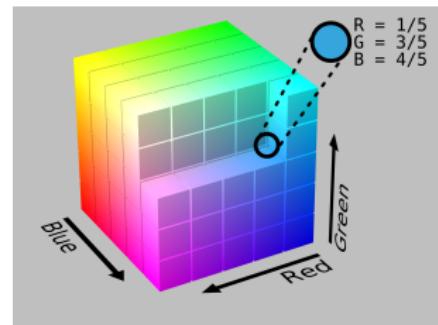


Imagen saturada para valores superiores a 1200m, com aumento de contraste entre -20m e 1200m.

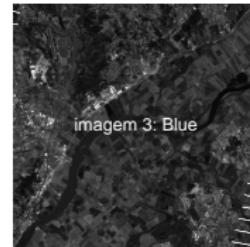
Composições coloridas de imagens

O modelo de cores RGB permite combinar, de forma aditiva, três imagens justapostas para criar uma composição colorida. A composição colorida é definida associando a 1a imagem ao canal *Red*; a 2a ao canal *Green*; e a 3a ao canal *Blue*.

Supondo que para um dado pixel, as três imagens tem intensidades respectivamente R , G e B entre 0 e 1, a cor associada ao pixel na composição colorida corresponde ao esquema junto.



Exemplo
no
Ribatejo:



Reamostragem de cdg matriciais

A reamostragem de um cdg matricial consiste permitir representar esse cdg:

- ① Num novo **sistema de coordenadas de referência**; e/ou
- ② Com uma **extensão**; e/ou
- ③ Com uma nova **resolução**;

de acordo com algum **critério de reamostragem** previamente estabelecido.

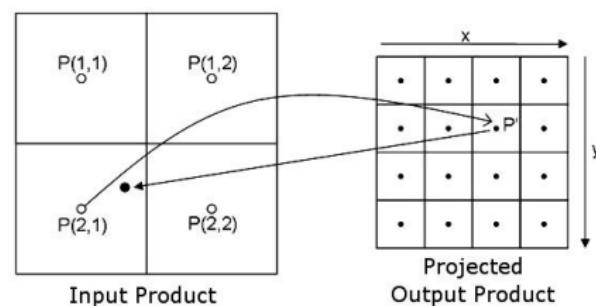
A técnica de reamostragem consiste em:

- ① Definir uma **nova grelha regular, sem valores**, com um determinado sistema de coordenadas, uma determinada extensão e uma determinada resolução;
- ② Usar a imagem a reamostrar para atribuir valores aos pixels da nova grelha.

Reamostragem de cdg matriciais (continuação)

À esquerda está representada a imagem original. À direita está representado *output* da reamostragem: a nova grelha, com o sistema de coordenadas, a extensão e a resolução escolhidos arbitrariamente.

Considere-se que se conhece uma **transformação de coordenadas** (ver Slide 205) que associa cada ponto P' do *output* à localização respectiva (●) na imagem a reamostrar.

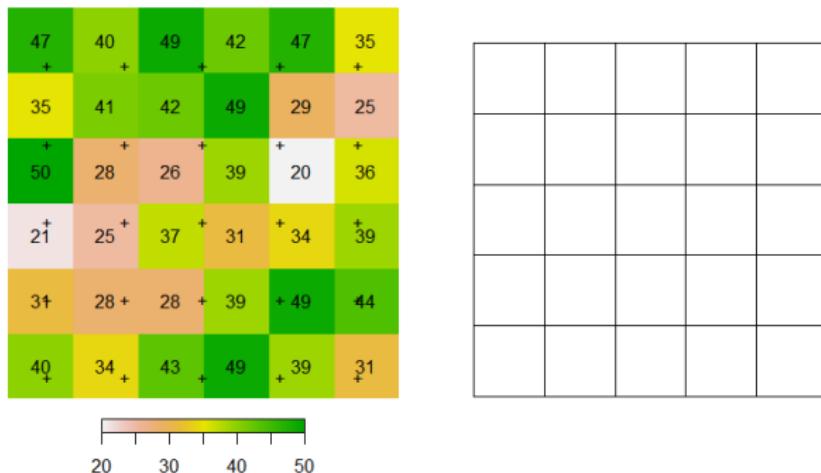


O valor a atribuir ao pixel P' depende do critério de reamostragem.

- Pelo **critério do vizinho mais próximo**, o valor do pixel $P(2, 1)$ é atribuído à célula de P' na imagem transformada dado que o ponto (●) está na célula $P(2, 1)$;
- Pelo **critério de interpolação**, o valor atribuído a P' é uma média ponderada de $P(1, 1)$, $P(1, 2)$, $P(2, 1)$ e $P(2, 2)$. Os coeficientes de ponderação dependem das distâncias entre cada ponto $P(1, 1)$, $P(1, 2)$, $P(2, 1)$ e $P(2, 2)$ e o ponto (●).

Reamostragem de cdg matriciais: exercício

Pretende-se fazer a remostragem do cdg matricial à esquerda A, para um cdg B à direita com resolução mais grosseira. Os centros dos pixels do cdg B estão indicados por “+” no cdg A.

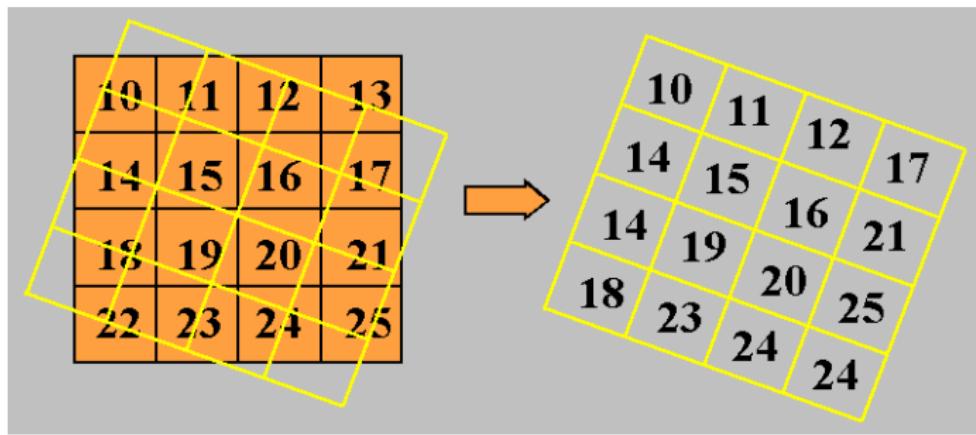


- 1 Determine o cdg B se usar o **critério do vizinho mais próximo**;
- 2 Qual será aproximadamente o valor do pixel no topo da 3a coluna se usar em alternativa um **critério de interpolação** com coeficientes inversamente proporcionais às distâncias?

Reamostragem de cdg matriciais: exercício

Na figura abaixo, a imagem a laranja é a imagem original e a imagem cinzenta é o resultado da sua reamostragem sobre a grelha inclinada.

- 1 Qual é o critério de reamostragem usado neste exemplo?
- 2 Dê uma justificação para os valores 14 e 24 ocorrerem repetidos na imagem que resulta da reamostragem;
- 3 O que ocorre se se tentar “melhorar” a resolução de uma imagem de 100 m para 10 m fazendo uma reamostragem pelo critério do vizinho mais próximo?



Operações sobre cdg matriciais (sobreposição matricial)

O tipo de operações realizadas sobre cdg matriciais tem uma natureza distinta das operações para cdg vectoriais (ver Slide 102).

*No resto desta secção, irá ser considerado que todos os cdg matriciais têm o mesmo sistema de coordenadas, a mesma extensão e a mesma resolução. Caso contrário, pode usar-se uma **reamostragem** para garantir essa condição.*

É habitual classificar as operações sobre cdg matriciais em função dos pixels usados para determinar o resultado da operação no pixel (i, j) :

- ① **operações locais**, só usam os pixels na posição (i, j) ; Exemplo: operações aritméticas ou lógicas sobre uma ou mais imagens;
- ② **operações focais**, usam valores do pixels numa vizinhança do pixel (i, j) ; Exemplo: filtros locais, também conhecidos como janelas móveis sobre a imagem;
- ③ **operações globais** usam em geral todos os pixels da imagem para determinar resultado da operação na posição (i, j) ; Exemplo: mínimo da imagem, segmentação de imagem, determinação do caminho de custo mínimo, mapa de visibilidade sobre um MDE.

Operações locais para cdg matriciais: operações aritméticas e lógicas

Considere cdg matriciais com idênticos CRS, extensão e resolução e operações que devolvem um cdg matricial com igual CRS, extensão e resolução.

- 1 **Comparações**, e.g. `elevacao > 100`, e **expressões Booleanas** como

`(declive>= 5 AND declive<=20) OR elevacao>1000`

devolvem um novo cdg matricial com os mesmos CRS, extensão e resolução, em que pixel tem valor 0 se a expressão acima tem valor lógico **F**, ou valor 1 se a expressão tem valor lógico **V**.

- 2 **Expressões aritméticas**. Por exemplo,

$$1 - \exp(-(pp - ET)^2)$$

devolve um novo cdg matricial com os mesmos CRS, extensão e resolução que os cdg `pp` e `ET`, com valores entre 0 e 1.

Quando o valor do pixel para algum dos cdg de *input* não está definido (`é null`), então o valor do pixel no *output* também é `null`.

Operações locais para cdg matriciais: reclassificação

A reclassificação é uma operação que tem um único cdg de *input*, e que tem como parâmetro uma tabela de reclassificação.

Exemplo: Considere o cdg matricial `declive` com valores entre 0 e 200 (em %), e a função de reclassificação definida pela tabela

declive	novoValor
0	10
]0,100[20
[100, 150[30
≥ 150	40

Note-se que **uma reclassificação pode ser sempre definida por uma expressão aritmética**. Neste exemplo, a expressão é:

```
10 * (declive=0) +
20 * (declive>0 AND declive<100) +
30 * (declive>=100 AND declive<150) +
40 * (declive>=150)
```

em que o valor lógico **V** vale 1 e o valor lógico **F** vale 0.

Operações locais para cdg matriciais: exercício

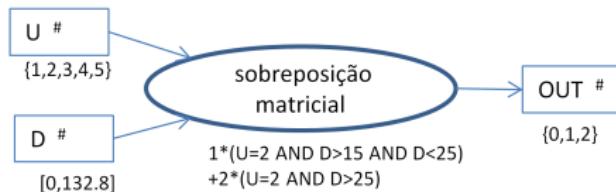
Exercício (adaptado do exercício I.2 do exame de 19 de junho de 2014).

Dispõe-se de dois cdg matriciais U e D de uma dada área de estudo. Ambos os cdg utilizam o mesmo CRS e têm resolução espacial de 90m. O cdg U representa o uso do solo em 2012, segundo a classificação do COS2007; as suas células só podem conter valores inteiros com a seguinte codificação: Territórios artificializados [1]; Áreas agrícolas [2]; Florestas e meios naturais [3]; Zonas húmidas [4]; Corpos de água [5]. O cdg D representa os declives do terreno, em percentagem; as suas células contêm valores numéricos com casas decimais. Os declives do terreno variam entre 0% e 132.8%.

Pretende-se obter um cdg matricial em que cada célula tenha valor 1 nos locais classificados como “Áreas agrícola” onde o declive do terreno é superior a 15% e inferior a 25% e tenha o valor 2 nos locais classificados como “Áreas agrícolas” onde o declive do terreno é superior a 25%. Nas células restantes o valor deve ser 0.

Operações locais para cdg matriciais: exercício (continuação)

- 1 Apresente um diagrama de operações, contendo unicamente operações de sobreposição matricial e de reclassificação, para obter o cdg pretendido. O diagrama deve representar as operações, os seus parâmetros, os domínios de valores, o nome e o tipo de cada cdg.



Uma solução:

- 2 Escreva o resultado final correspondente às células de U e D representadas abaixo e determine nesse extrato a área (indicando as unidades) de “Áreas agrícolas” onde o declive é superior a 15% e inferior a 25%.

U: Uso do solo

2	2	2	2
2	3	2	3
1	2	2	3
1	2	2	3

D: Declive do terreno

11.3	16.6	14.2	12
10.2	19.4	17.6	18.7
9.4	13.7	19.2	22.4
8.3	19.8	27.3	44.6

OUT

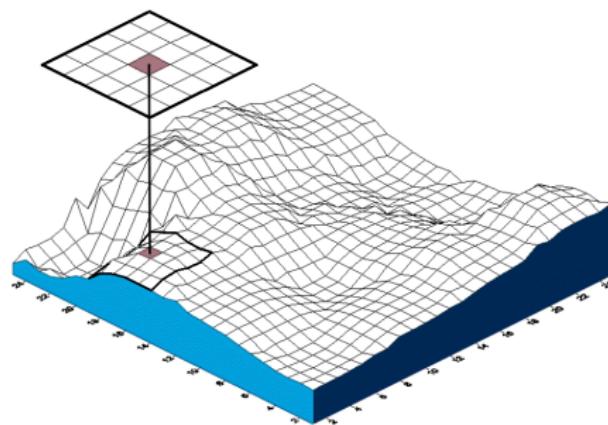
0	1	0	0
0	0	1	0
0	0	1	0
0	1	2	0

A área pedida corresponde a 4 pixels, i.e. $4 \times (90m)^2 = 32400m^2$.

Operações focais sobre cdg matriciais: filtros

Um filtro, também designado por janela móvel, é uma operação que atribui a cada pixel (i, j) do cdg de *output* o valor de uma função dos valores dos pixels do cdg de *input* numa vizinhança de (i, j) .

Essa vizinhança é designada por **janela móvel** e é tipicamente um arranjo de 3×3 , ou 5×5 pixels centrado em (i, j) , mas pode em geral ser uma vizinhança com qualquer forma e dimensão.

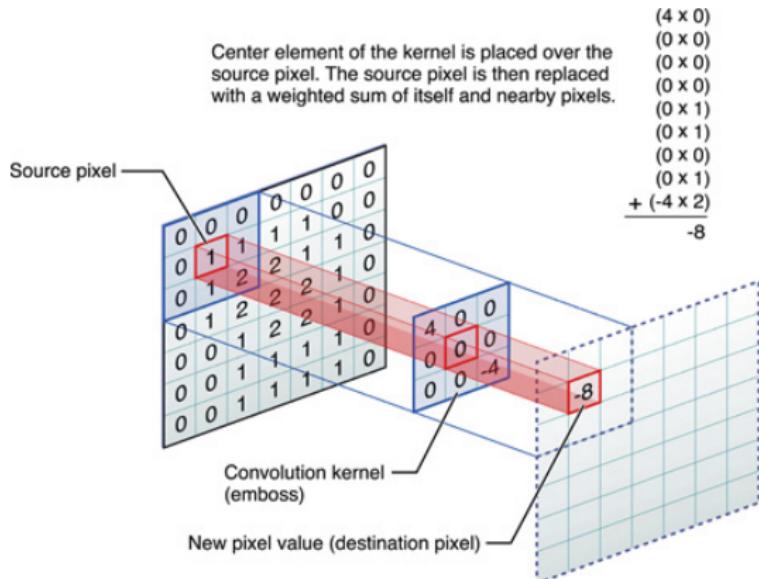


Operações focais sobre cdg matriciais: filtros

É habitual distinguir dois tipos de filtros:

- 1 **filtros não lineares**, quando a função f a aplicar aos valores da janela móvel é uma função não linear; Exemplo: filtro de moda, de mediana ou de máximo;
- 2 **filtros lineares**, quando f é linear (ver Slide 174). Os filtros lineares são muito eficientes computacionalmente.

Os filtros lineares podem ser definidos pelo seu ***kernel***, uma matriz de pesos que identifica a janela móvel e o filtro. Um filtro linear devolve na posição (i, j) a soma ponderada dos valores no ***kernel*** como no exemplo junto.



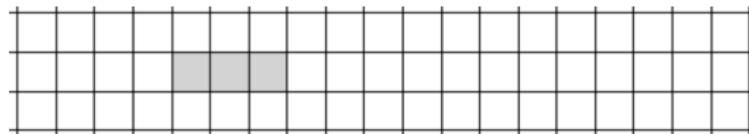
Operações focais sobre cdg matriciais: filtros não lineares

O **filtro de mediana** substitui os valores em cada janela móvel $n \times n$ pelo valor mediano. É tipicamente usado para remover ruído do tipo “sal e pimenta” sobre imagens.

Exercício: a imagem seguinte representa os valores numa secção de uma imagem $0.6\text{-}0.7 \mu\text{m}$ (resolução espacial de 30 m) sobre uma região agrícola no Alentejo:

45	43	40	37	28	31	39	45	38	35	40	36	32	43	45	45	40	34
48	46	47	40	31	25	31	42	38	37	37	37	31	34	39	39	39	36
43	40	39	40	36	30	32	34	35	36	35	34	32	32	38	37	39	38

Aplique filtro de mediana 3×3 e mostre que os valores nas três células assinaladas são, respectivamente, 31, 31 e 32.



Operações focais sobre cdg matriciais: filtros não lineares

No exemplo abaixo, uma imagem de satélite com ruído causado por efeito da atmosfera é processado com um filtro de mediana 3×3 :



Integração vectorial/matricial e interpolação espacial

Restruturação de informação geográfica

Como discutido anteriormente, existem principalmente dois modelos conceptuais de dados: vectorial e matricial. Cada modelo tem vantagens e desvantagens relativamente à representação e ao processamento de informação. Assim, existem vários motivos para se querer restrukturar a informação de um dos modelos para o outro.

- ➊ Combinar várias fontes de informação; por exemplo, pode ser útil converter um cdg vectorial de ocupação do solo – cruzado com a informação da temperatura máxima aconselhada por ocupação – num cdg matricial, para depois combinar esse novo cdg com um cdg matricial de temperatura do solo obtido por detecção remota.
- ➋ Restructurar a informação de um modelo A para um modelo B por forma a poder aplicar uma operação que só pode ser aplicada a B. Por exemplo, converter um cdg matrical num cdg vectorial para depois poder realizar um cruzamento de tabelas.

Restruturar um cdg vectorial num cdg matricial

Como simplificação, considere-se que o sistema de coordenadas do cdg matricial é idêntico ao do cdg vectorial, e que o cdg vectorial é do tipo “polígono”.

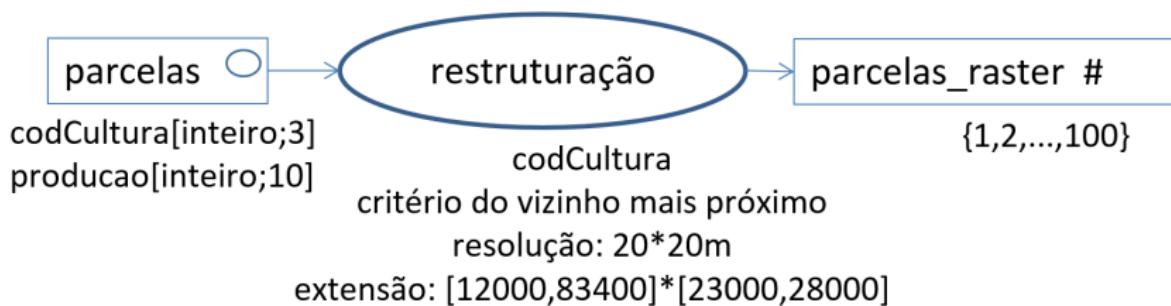
Para restruturar a informação contida no cdg vectorial são necessários os seguintes parâmetros:

- ① **atributo** (numérico) do cdg vectorial a ser restruturado;
- ② **extensão** do cdg matricial de *output*; convenciona-se que é o menor rectângulo que contém o cdg vectorial caso esta extensão não seja indicada;
- ③ **resolução** do cdg matricial;
- ④ **critério de reamostragem**, que é em geral o critério do *vizinho mais próximo*, em que o valor que toma o pixel com centro (x, y) é o valor do atributo do cdg vectorial nessa posição. Caso o cdg vectorial não contenha o ponto (x, y) , o valor do pixel é null.

Restruturar um cdg vectorial num cdg matricial: ilustração



A operação de **restruturação** pode ser descrita por um diagrama como o seguinte



em que os parâmetros da operação indicam a extensão e a resolução do cdg matricial de *output*.

Restruturar um cdg matricial num cdg vectorial

Como o cdg matricial tem um valor por célula, só faz sentido fazer a restruturação se existirem “manchas” de pixels com valor idêntico. Cada uma dessas manchas é convertida num polígono ao qual é associado o valor respectivo. Para esta operação, só é necessário indicar o atributo que resulta da conversão.

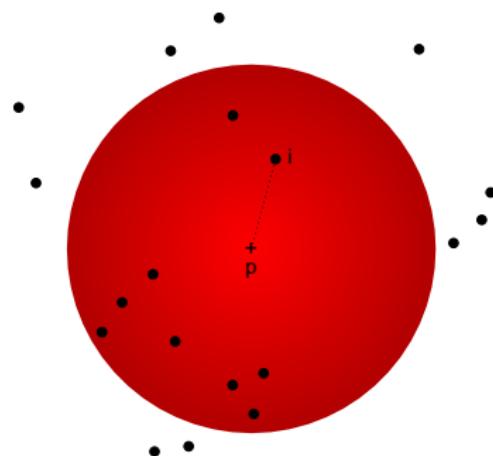


Interpolação espacial

I invoke the first law of geography: everything is related to everything else, but near things are more related than distant things (Tobler, 1970)

A afirmação acima estabelece a **primeira lei da geografia**, que descreve informalmente a ideia de que os fenómenos geográficos têm uma estrutura espacial que depende da distância entre localizações.

Neste exemplo, o ponto mais próximo de p é i , e isso é uma indicação de que p poderá estar mais relacionado com i do que com outros pontos.



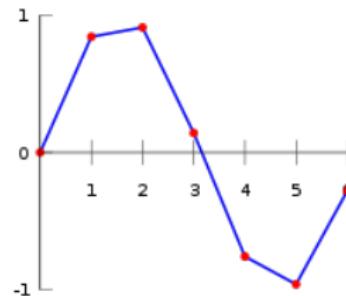
Interpolação espacial de medições pontuais

(Interpolação espacial)

É a interpolação de medições em localizações pontuais para todo o espaço considerado. Mais precisamente, dados valores $\{z_1, \dots, z_n\}$ nas localizações $\{(x_1, y_1), \dots, (x_n, y_n)\}$, a interpolação espacial é uma função $z = f(x, y)$ definida sobre o domínio das coordenadas $\{(x_i, y_i)\}$ que satisfaz algumas condições predefinidas.

A interpolação é **linear** quando a função f é linear por partes e quando $f(x_i, y_i) = z_i$ para todos os pontos, isto é, quando coincide com os valores medidos nos pontos de medição.

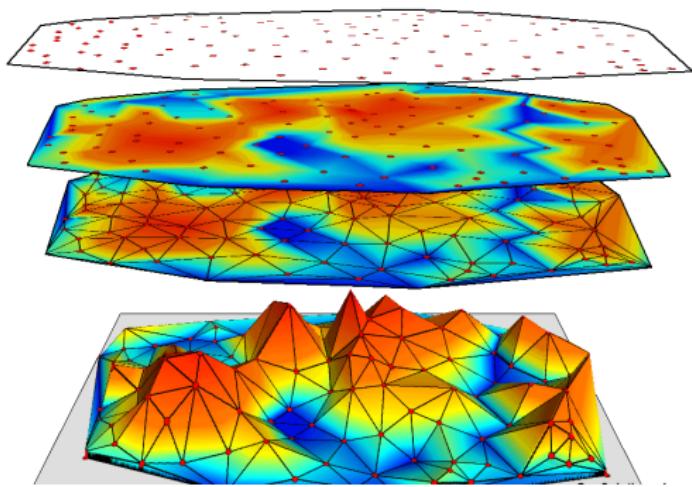
Ilustração de interpolação linear
em uma dimensão:



Interpolação linear sobre uma rede triangular irregular (TIN)

Dado um conjunto de pontos $\{(x_i, y_i)\}$, partitiona-se o seu invólucro convexo em triângulos de Delaunay. A interpolação é **linear** em cada triângulo *abc* com $f(a) = z_a$, $f(b) = z_b$ e $f(c) = z_c$. Esta interpolação adapta-se de forma óptima à densidade de pontos disponíveis.

- ① cdg vectorial de pontos com valores pontuais da variável z ;
- ② representação da magnitude dos valores z_i ;
- ③ triangulação (de Delaunay);
- ④ interpolação linear dos valores de z em cada triângulo.



(adaptado de www.edc.uri.edu)

A interpolação f pode adicionalmente ser convertida num cdg **matricial** em que cada pixel de centro (x, y) toma o valor $f(x, y)$.

Exemplo: interpolação linear e criação de cdg matricial

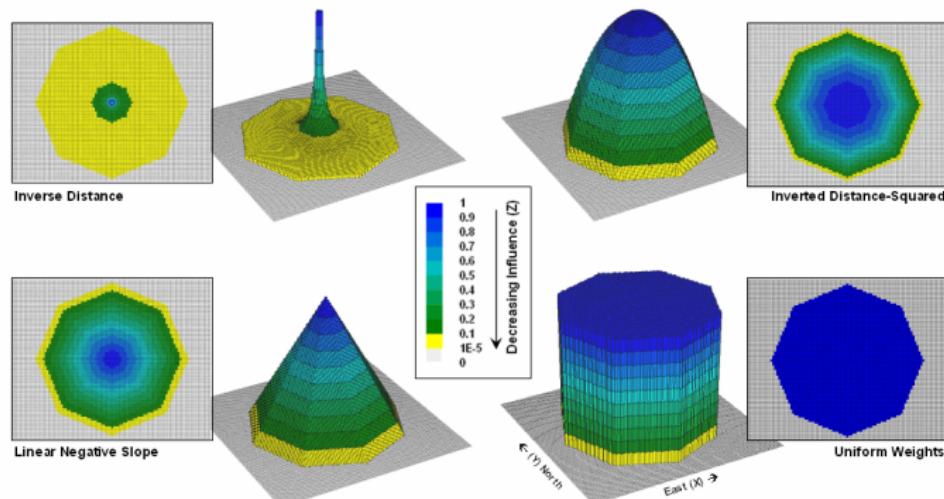
Ilustração da técnica de interpolação linear sobre uma TIN:

- 1 **input**: Um conjunto de pontos com valores da variável z e uma grelha fixa;
- 2 **TIN**: Dos pontos de *input* deriva-se uma triangulação de Delaunay;
- 3 **Para cada pixel da grelha**:
 - ▶ determina-se a qual triângulo pertence o centro do pixel;
 - ▶ realiza-se uma interpolação linear no triângulo para obter a estimativa do valor no centro do pixel (ver Slide 172).

Filtros lineares para interpolação espacial

Como visto no Slide 143, o **kernel** de um filtro linear pode ter **suporte** (células com pesos diferentes de zero) e **pesos** variáveis.

A figura abaixo mostra a forma de alguns exemplos de *kernel* de filtros lineares usados em processamento de imagem para **interpolação espacial**: alguns dão maior peso ao centro do *kernel* e outros têm uma distribuição de pesos mais uniforme.



Filtros lineares para interpolação espacial: distâncias

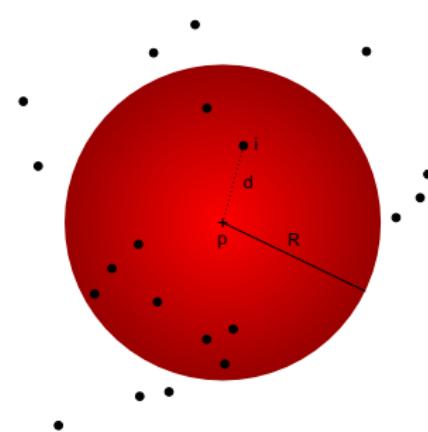
Na figura abaixo, estão indicadas localizações de observações pontuais de uma dada variável z e o **suporte** do kernel com raio R , centrado num ponto p onde se pretende **estimar o valor da variável z** .

Todas as observações que pertencem ao suporte (por exemplo i) são usadas para estimar o valor da variável em p .

Tipicamente, o valor estimado é uma **combinação linear convexa** dos valores pontuais z_i :

$$z_p = \sum_{i \in \text{suporte}} z_i \times c_i$$

em que $\sum_i c_i = 1$ e cada coeficiente c_i depende da distância d entre p e i .



Filtros lineares para interpolação espacial: IDW

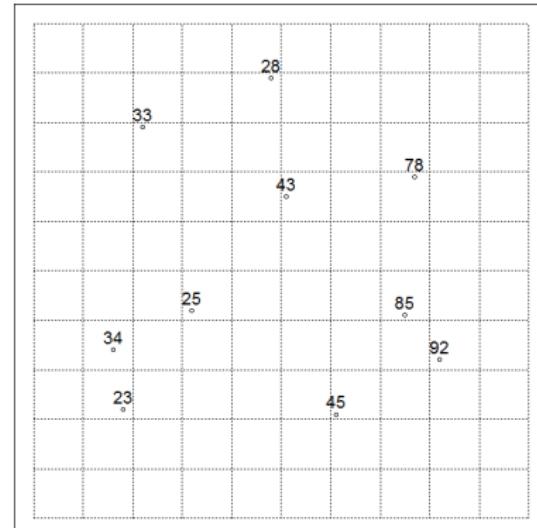
Uma técnica de uso frequente para realizar interpolação espacial é designada por **IDW**: *inverse distance weighting*.

Nesse caso, para cada pixel de centro p , cada coeficiente c_i depende do **inverso** da distância d_i entre p e a i -ésima **observação pontual** que pertence ao suporte.

O resultado depende muito da forma como cada coeficiente c_i depende de d_i : tipicamente escolhe-se um parâmetro ρ (o **expoente**) tal que

$$c_i \text{ é proporcional a } \left(\frac{1}{d_i}\right)^\rho$$

Este processo é realizado em todas as células da grelha que cobre a região de interesse. Como resultado, obtém-se um **cdg matricial**.



Exemplo: IDW com kernel de “raio” ≈ 2.5 pixels

c_i proporcional a $\frac{1}{d_i}$: todos os pontos contribuem razoavelmente.

c_i proporcional a $\left(\frac{1}{d_i}\right)^2$: os pontos mais próximos contribuem muito.

Interpolação espacial com IDW: exercício

No exemplo do slide anterior, suponha que a grelha tem uma resolução espacial de 10 m e considere que as distâncias entre o centro P do píxel mais próximo do ponto com valor 45 (em baixo, na sétima coluna) e os pontos com valores 45, 85 e 92 são respectivamente 5.66 m, 18.87 m e 18.38 m.

Mostre que os valores estimados pela técnica de interpolação IDW são de facto:

- ➊ 61.46 quando o expoente da distância é 1 (figura à esquerda);
- ➋ 51.8 quando o expoente da distância é 2 (figura à direita);

Interpolação espacial: algumas considerações

Existem muitas técnicas de interpolação espacial, que permitem obter estimativas dos valores de uma variável geográfica de interesse em locais em que essa variável não é observada. Exemplos de aplicação: temperatura, precipitação, concentração de poluentes no solo, humidade no solo, ...

É necessário conhecer bem o **comportamento da variável** de interesse para saber se esse tipo de técnicas devem ser aplicadas e para decidir qual a técnica mais adequada.

A **interpolação linear** é uma técnica simples que se adapta bem à amostra de pontos: pode ser adequada se a variável tem um comportamento regular sobre a superfície (por exemplo, elevações numa região em que o relevo não é muito abrupto) mas tem o inconveniente de devolver uma “superfície” não regular (não diferenciável).

A técnica **IDW** depende muito do *kernel* escolhido e, em particular, da forma como os coeficientes dependem das distâncias (expoente igual a 1, 2, ...).

Interpolação espacial: um exemplo de aplicação

Dados climáticos ou ecológicos em formato *raster* e usados em análises espaciais podem em particular corresponder a:

- ① dados genuinamente *raster*, como imagens de satélite;
- ② o resultado de uma **interpolação espacial** de observações pontuais sobre a superfície.

Alguns exemplos de fontes de dados *raster* em estudos globais:

- **Superfície** – e.g. Atlas of the Biosphere, South And Central America Environmental Data, MODIS-Derived Satellite Data;
- **Clima** – e.g. IPCC Data Distribution Centre, Climatic Research Unit of the University Of East Anglia, WorldClim Global Climate Data; variáveis essenciais ESA/CCI;
- **Elevação, batimetria** – e.g. SRTM 30m e 90m Digital Elevation Data e outras fontes em Earth Explorer;
- **Oceanos** – e.g. Permanent Service For Mean Sea Level, World Oceans Atlas.

Interpolação espacial: um exemplo – WorldClim

WorldClim é um conjunto de dados matriciais derivados de medições pontuais de **temperatura** e de **precipitação** ao longo do tempo em estações meteorológicas.

Os dados em formato matricial têm uma resolução de $\approx 1\text{km}$ (mais precisamente, 30 arc segundos), e são muito utilizados para modelação ecológica e monitorização de alterações climáticas. Os dados pontuais originais provém de:

- **Global Historical Climate Network Dataset (GHCN)**: mais de 20000 estações meteorológicas para precipitação e mais de 7000 para temperatura;
- **World Meteorological Organization climatological normals**: mais de 4000 e 3000 estações;
- **FAOCLIM**: mais de 27000 e 20000 estações;
- Outros registos regionais.

No conjunto, e excluindo repetições, Worldclim usa mais do que 47000 localizações para precipitação, 25000 para temperatura média e 14000 para temperaturas mínimas e máximas mensais.

Interpolação espacial: um exemplo – WorldClim

- 1 Apesar de haver um número elevado de medições pontuais, a sua distribuição espacial é muito irregular:

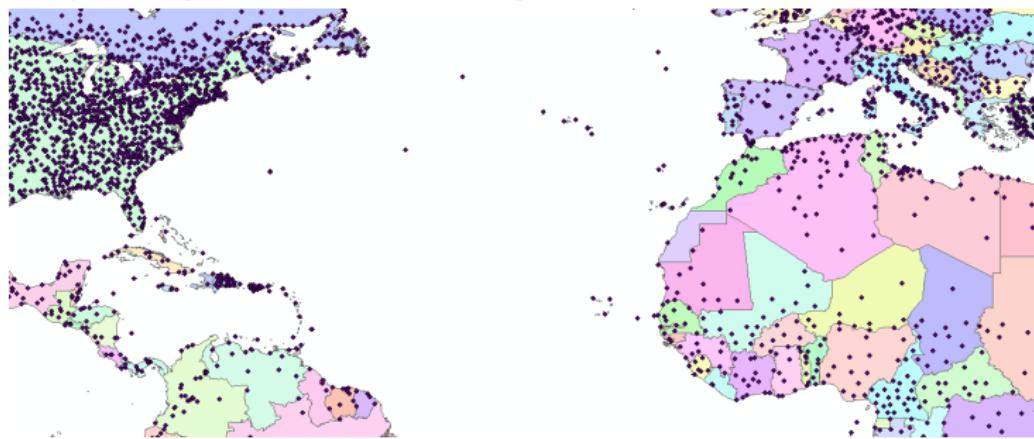


Figura: Estações do Global Historical Climate Network Dataset.

- 2 Worldclim é um conjunto de dados matriciais **derivado de dados pontuais** para as áreas terrestres. Na maior parte das localizações, fornece apenas **estimativas** – obtidas por **interpolação espacial** – das variáveis climáticas.

Aplicações a modelos digitais de elevações (raster)

Dados de elevação

A elevação é a variável essencial para a representação do relevo, a partir da qual podem ser derivadas novas variáveis para caracterização da forma da superfície.

Os dados de elevação podem ser obtidos por várias formas:

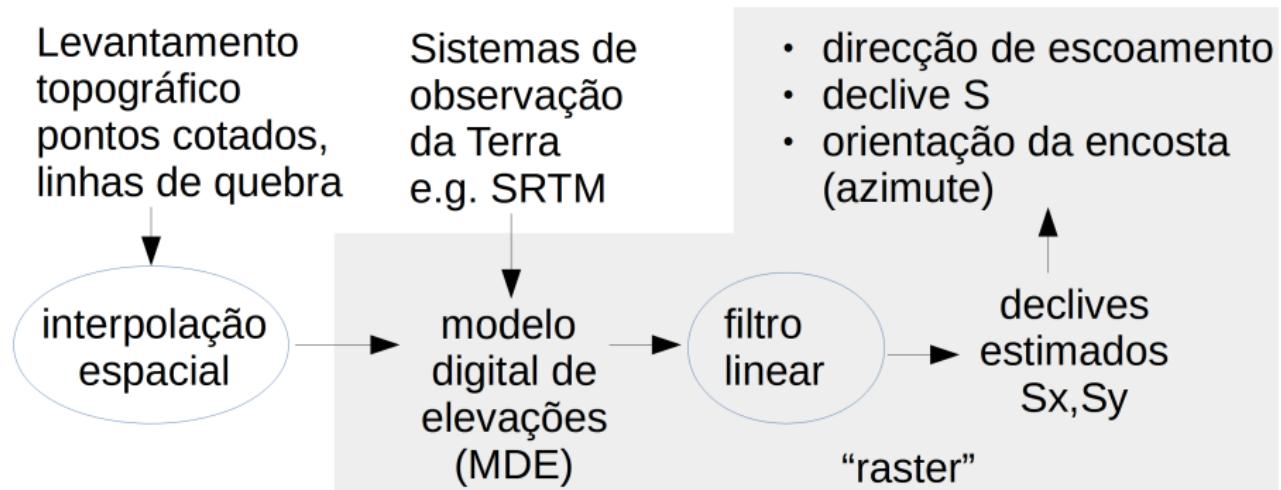
- ① Levantamento topográfico;
- ② Levantamento com técnicas de posicionamento global (e.g. GPS)
- ③ Estereofotogrametria de fotografia aérea ou imagens de satélite;
- ④ Detecção remota “activa” com LIDAR e RADAR.

A **representação do relevo** pode ser feita através de

- ① pontos cotados,
- ② curvas de nível,
- ③ redes triangulares irregulares (TIN),
- ④ **modelos digitais de elevação** (MDE, formato raster, usado nesta neste capítulo).

Algumas técnicas usadas para obter e explorar modelos digitais de elevação em formato raster

Nas secções anteriores foram analisadas as técnicas de **interpolação espacial** e **filtros lineares** que podem ser aplicadas neste contexto.



Filtros lineares permitem derivar de forma expedita nova informação a partir de um modelo digital de elevações (MDE).

Modelos digitais de elevação globais

Existem vários modelos digitais de elevação globais de livre acesso. O modelo mais usado é conhecido por SRTM, obtido em fevereiro de 2000 pela missão “Shuttle Radar Topography Mission”.

O MDE SRTM cobre a maior parte do globo (entre 60°S e 60°N), está disponível com resolução 3 arco-segundos e, desde 2015, 1 arco-segundo. Sobre a Eurásia, 90% dos pixels têm erro posicional inferior a 8.8m e erro de elevação inferior a 6.2m.

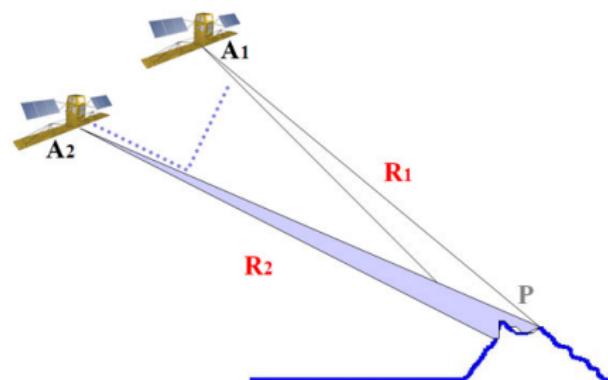


Figura: Ilustração da aplicação da técnica de interferometria por radar de abertura sintética (InSAR) utilizada na missão SRTM.

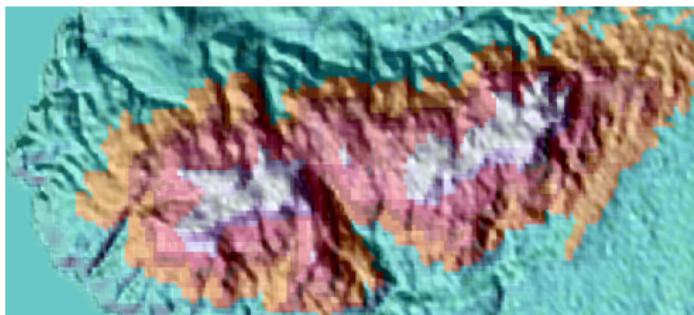
Algumas técnicas para análise do relevo

Algumas das técnicas discutidas em secções anteriores podem ser usadas para analisar o relevo. Nomeadamente:

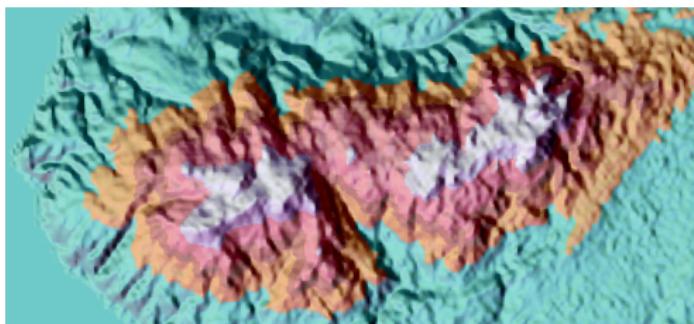
- ① Reamostragem de MDE (ver Slide 133);
- ② Interpolação espacial de valores pontuais de elevação (Slide 153, com eventual incorporação de linhas de quebra (linhas de água, linhas de festo));
- ③ Interpolação espacial do tipo “Inverse distance weighting” (ver Slide 155);
- ④ Derivação de novas variáveis para caracterização do relevo com filtros lineares (ver Slide 143). Por exemplo, filtros lineares permitem derivar o declive em cada direcção (S_x e S_y). A partir de S_x , S_y estima-se a direcção de maior declive e a sua magnitude.

Reamostragem de modelos digitais de elevação

A técnica de reamostragem (Slide 133) converte um *raster* para um novo *raster* com novo CRS, resolução e extensão. Quando o relevo é razoavelmente regular, a reamostragem do relevo para uma resolução mais detalhada pode ser justificada.



Representação do SRTM original sobre a Serra de Sintra (resolução de aproximadamente 90 m).



MDE resultante de uma re-amostragem para 30 m para compatibilidade com imagem Landsat. Neste caso a reamostragem não cria replicações de valores nos pixels pois foi aplicada um critério de interpolação (Lanczos).

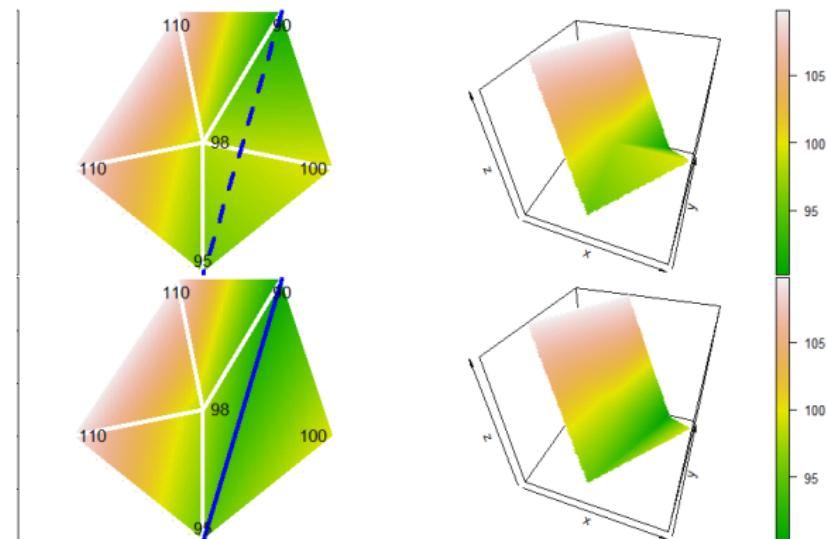
Interpolação linear com linhas de rotura

Linhas de rotura são linhas que não podem ser cruzadas por arestas da triangulação. Um exemplo típico são linhas de água ou linhas de festo numa interpolação linear de valores pontuais de elevação (ver Slide 153).

No exemplo abaixo, a linha a azul representa uma linha de água.

No topo. Triangulação sem linhas de rotura: representação errada do relevo;

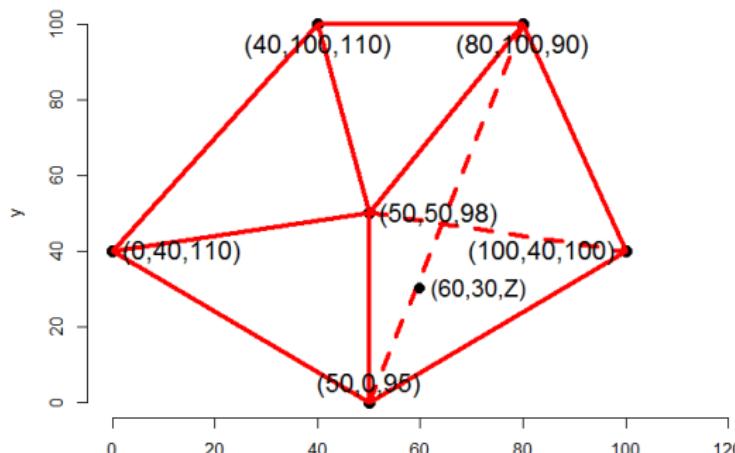
Em baixo. A linha de água é usada como linha de rotura alterando a triangulação: o modelo de elevação é mais rigoroso.



Interpolação linear: exercício

Exercício: Considere a figura abaixo que indica 6 pontos com as suas coordenadas (x_i, y_i, z_i) . Escreva o sistema de equações que permite determinar a coordenada z do ponto com coordenadas $P = (60, 30)$ nas duas situações seguintes, em que se supõe que

- 1 P pertence ao triângulo de vértices $(50,0)$, $(50,50)$ e $(100,40)$;
- 2 A linha a tracejado entre $(50,0)$ e $(80,100)$ é uma linha de rotura e, portanto, P pertence ao triângulo de vértices $(50,0)$, $(80,100)$ e $(100,40)$.



Interpolação linear: solução do exercício

- 1 No primeiro caso, o ponto $P = (60, 30, z)$ pertence ao plano $z = a + bx + cy$ que passa por $(50, 50, 98)$, $(100, 40, 100)$ e $(50, 0, 95)$. Então o valor de z é dado pelo sistema seguinte.

$$\begin{cases} 98 = a + b \times 50 + c \times 50 \\ 100 = a + b \times 100 + c \times 40 \\ 95 = a + b \times 50 + c \times 0 \\ z = a + b \times 60 + c \times 30 \end{cases} \text{ que determina } z = 97.32\text{m}$$

- 2 No segundo caso, o ponto $P = (60, 30, z)$ pertence ao plano $z = a + bx + cy$ que passa por $(80, 100, 90)$, $(100, 40, 100)$ and $(50, 0, 95)$. Então o valor de z é dado pelo sistema seguinte.

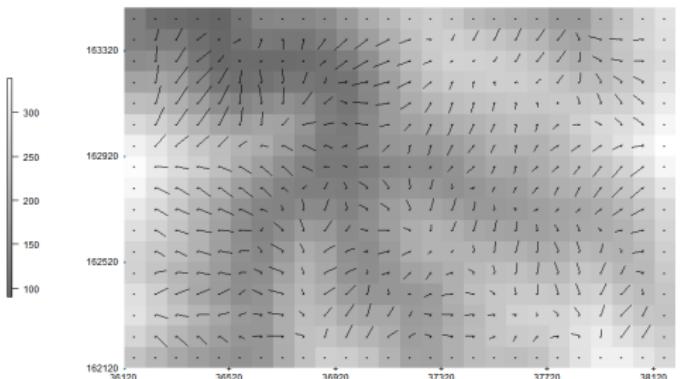
$$\begin{cases} 90 = a + b \times 80 + c \times 100 \\ 100 = a + b \times 100 + c \times 40 \\ 95 = a + b \times 50 + c \times 0 \\ z = a + b \times 60 + c \times 30 \end{cases} \text{ que determina } z = 93.86\text{m}$$

A solução no segundo caso é consistente com a linha de rotura entre $(50, 0)$ e $(80, 100)$.

Derivação de variáveis de caracterização do relevo a partir de MDEs

Num MDE, cada pixel tem um valor que é a elevação média do terreno correspondente ao pixel. A partir de um MDE podem ser derivadas novas variáveis muito relevantes para caracterizar o relevo. Uma dessas variáveis é a **direcção do escoamento** que corresponde à direcção de maior declive em cada ponto.

132	105	101	98	90	113	143	146	146	152	170	204	225	247	255	246	225	213	235	190	178	179	210	241	269	292	
134	126	120	104	92	111	121	118	118	131	158	183	211	237	265	262	254	243	231	220	198	191	216	247	271	289	
153	157	153	129	101	101	101	101	109	109	128	159	195	228	253	265	264	262	254	241	222	287	219	243	265	281	
190	192	160	146	115	106	119	119	117	109	133	145	184	216	242	253	266	265	259	254	244	230	232	244	269	290	
234	228	211	183	154	140	159	159	149	124	155	144	178	202	222	231	236	244	249	248	250	252	256	264	283	308	
278	265	244	218	193	183	191	187	169	131	122	149	175	189	191	208	216	225	234	230	242	261	275	283	300	325	
318	295	274	249	236	215	203	183	184	126	124	149	169	172	173	184	195	212	216	219	228	248	270	296	313	339	
333	307	281	260	235	205	182	159	139	121	119	134	149	154	154	164	183	197	196	206	222	234	240	266	293	323	
323	392	264	239	215	183	157	137	132	130	127	144	163	170	167	161	167	178	183	196	208	215	225	244	267	298	
305	278	250	222	192	161	149	138	159	149	139	166	181	194	193	182	173	173	175	183	196	195	206	224	247	277	
288	265	244	217	184	155	138	133	153	179	174	151	157	184	206	219	208	196	191	181	185	192	201	204	213	237	265
268	299	266	230	213	184	157	144	121	202	191	159	161	176	203	217	217	212	203	208	223	249	268	291	319	349	
207	243	222	201	177	154	155	188	215	198	166	166	169	188	199	209	226	234	238	228	228	245	242	223	225	253	
286	262	229	199	175	156	186	202	224	207	180	168	169	170	179	207	233	246	252	251	249	282	265	243	232	247	
291	266	243	219	194	172	168	206	232	226	204	195	194	176	181	204	226	245	265	269	269	284	288	264	246	244	
271	255	221	201	180	165	167	167	205	236	224	228	219	206	185	207	227	228	245	254	256	279	295	306	289	271	256
249	221	199	181	170	161	161	169	203	236	205	247	241	217	187	185	209	229	235	234	242	268	292	315	313	327	375



À esquerda um MDE na região do Douro. À direita, a direcção e magnitude do maior declive derivado do MDE.

Operações focais sobre cdg matriciais e análise do relevo

Os filtros lineares (ver Slide 143) permitem analisar a vizinhança de cada pixel. Um dos exemplos mais comuns da aplicação desses filtros é a análise do relevo a partir do **modelo digital de elevação** em formato matricial.

As expressões abaixo dão estimativas do declive do terreno nas direcções x e y usando uma vizinhança de 3×3 pixels:

$$S_x = \frac{(z_3 - z_1) + 2(z_6 - z_4) + (z_9 - z_7)}{4 \times (\text{2} \times \text{resolucao em } x)}, \quad S_y = \frac{(z_1 - z_7) + 2(z_2 - z_8) + (z_3 - z_9)}{4 \times (\text{2} \times \text{resolucao em } y)}.$$



Neste exemplo, o pixel tem 20m de lado e as estimativas são:

- 1 $S_x = (1 + 2 \times 3 + 3)/(4 \times 40) = 0.0625, e$
- 2 $S_y = (39 + 2 \times 35 + 37)/(4 \times 40) = 0.9125.$

Operações focais sobre cdg matriciais e análise do relevo (continuação)

Exercício. Verificar que as estimativas de S_x e S_y do Slide 174 são o resultado de aplicação de filtros lineares, em que os *kernel* são os seguintes, e a resolução é de 20 m (i.e. a distância entre a 1a e 3a colunas é 40m, e as distâncias entre a 1a e a 3a linhas também é 40m).

$$S_x : \begin{bmatrix} -\frac{1}{4 \times 40} & 0 & \frac{1}{4 \times 40} \\ -\frac{2}{4 \times 40} & 0 & \frac{2}{4 \times 40} \\ -\frac{1}{4 \times 40} & 0 & \frac{1}{4 \times 40} \end{bmatrix} \quad S_y : \begin{bmatrix} \frac{1}{4 \times 40} & \frac{2}{4 \times 40} & \frac{1}{4 \times 40} \\ 0 & 0 & 0 \\ -\frac{1}{4 \times 40} & -\frac{2}{4 \times 40} & -\frac{1}{4 \times 40} \end{bmatrix}$$

Observação 1: a aplicação destes filtros para estimar os declives pressupõe que x , y e z têm as mesmas unidades (metros neste caso);

Observação 2: dado que S_x e S_y são definidos por filtros lineares, o cálculo das estimativas dos declives é feito de forma muito expedita em SIGs, mesmo para MDE de elevada dimensão.

Operações focais sobre cdg matriciais e análise do relevo (cont.)

A partir das estimativas dos declives na direcção de x e de y pode derivar-se várias variáveis que caracterizam o relevo.

- 1 A **direcção de escoamento**, i.e. a direcção de maior declive é a direcção do vector $(-S_x, -S_y)$;
- 2 O **declive** é o declive segundo a direcção de escoamento

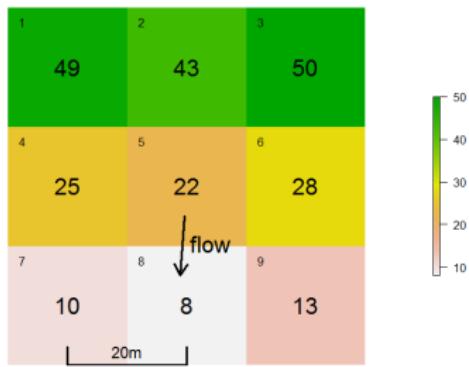
$$S = \sqrt{S_x^2 + S_y^2};$$

- 3 A **orientação da encosta** é a direcção de escoamento expressa como **azimute** (relativamente a Norte, em graus) ou como uma direcção cardinal (N,E,S,W) ou intercardinal (NE, SE, SW, NW). O azimute é $\text{arctan}(S_x/S_y)$ **mais 0° (se $(-S_x, -S_y)$ está no 1º quadrante), 180° (no 2º ou 3º quad.), ou 360° (no 4º quad.)**. Se $S_x = S_y = 0$ a orientação é indefinida (terreno horizontal).

azimute	direcção	azimute	direcção
$337.5^\circ - 22.5^\circ$	N	$22.5^\circ - 67.5^\circ$	NE
$67.5^\circ - 112.5^\circ$	E	$112.5^\circ - 157.5^\circ$	SE
$157.5^\circ - 202.5^\circ$	S	$202.5^\circ - 247.5^\circ$	SW
$247.5^\circ - 292.5^\circ$	W	$292.5^\circ - 337.5^\circ$	NW

Operações focais sobre cdg matriciais e análise do relevo (cont.)

Nas figuras abaixo, o vector *flow* indica a direcção de escoamento.



Neste exemplo (do Slide 174),

- 1 $S_x=0.0625, S_y=0.9125,$
- 2 $S=\sqrt{0.0625^2+0.9125^2}=0.9146379,$
- 3 Como $\arctan(S_x/S_y)=3.918^\circ$, a orientação é 183.918° , i.e. S.

Neste exemplo,

- 1 $S_x=-0.36875, S_y=-0.73125,$
- 2 $S=\sqrt{0.36875^2+0.73125^2}=0.8189647,$
- 3 Como $\arctan(S_x/S_y)=26.76^\circ$, a orientação é 26.76° , i.e. NE.

Operações focais sobre cdg matriciais e análise do relevo

Exercício (adaptado do exercício II.2 do exame de 19 de junho de 2014):

A figura representa um extrato de um modelo digital de elevações em formato raster, com resolução de $18\text{ m} \times 18\text{ m}$. Os valores das células são as altitudes em metros.

- ① Estime a direcção de escoamento para a célula (L,k) ;
- ② Estime o declive para a célula (L,k) ;
- ③ Qual é a orientação (octante) para a célula (L,k) ?

	$k-1$	k	$k+1$
$L-1$	80.7	84	87
L	82.5	87.2	91.8
$L+1$	77.5	82.3	86.6

Resolução do exercício do slide anterior

- 1 Pede-se o vector $(-S_x, -S_y)$ em que S_x e S_y são as estimativas dos declives na direcção do x (Oeste-Este) e do y (Sul-Norte).

A diferença de elevação é estimada pela média ponderada das diferenças em cada direcção sobre a distância horizontal no terreno que é de 36m ($=2 \times 18\text{m}$):

- ▶ Considerando a diferença entre a coluna k+1 e a coluna k-1:
 $S_x = (6.3 + 2 \times 9.3 + 9.1) / (4 \times 36) = 0.236$, e
- ▶ Considerando a diferença entre a linha L-1 e a linha L+1:
 $S_y = (3.2 + 2 \times 1.7 + 0.4) / (4 \times 36) = 0.0486$.

A direcção de escoamento é então a direcção $(-0.236, -0.0486)$;

- 2 O declive é $S = \sqrt{S_x^2 + S_y^2} = \sqrt{0.236^2 + 0.0486^2} = 0.241$, i.e., a altitude decresce em média 24.1m por 100m de distância horizontal no terreno, na direcção de escoamento;
- 3 Como a direcção de escoamento está no 3º quadrante, o azimute deve estar entre 180° e 270° . Como $\arctan(S_x/S_y) = 78.36^\circ$ então o azimute é $78.36 + 180^\circ = 258.36^\circ$. O orientação da encosta é portanto W (oeste) como sugerido pela figura.

Operações focais sobre cdg matriciais: análise do relevo (cont.)

Uma outra característica do relevo que pode ser derivada a partir do MDE é a quantidade de **iluminação** que cada pixel recebe.

A iluminação depende do **declive** e da **orientação de encosta** para cada pixel, e da posição do sol. Em SIGs, a função que devolve a iluminação da superfície como definido acima designa-se tipicamente por *hillshade*.

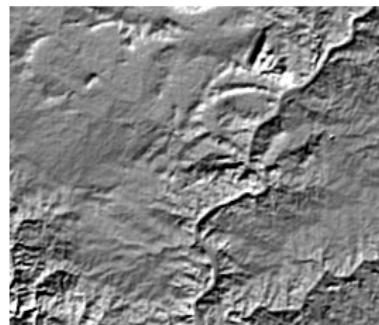
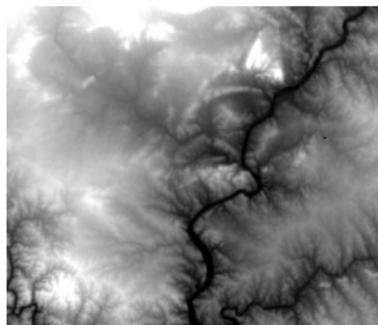


Figura: À esquerda o MDE numa certa região. À direita o valor do coseno do ângulo de incidência quando o sol tem uma elevação de 26.6° e um azimute de 129° : uma cor mais clara indica que o sol incide normalmente sobre a superfície.

Sistemas de coordenadas de referência

Sistemas de coordenadas de referência

Todos os conjuntos de dados geográficos devem estar associados a um **sistema de coordenadas de referência** (CRS) por forma a:

- ① Conhecer a **localização** sobre a superfície terrestre dos fenómenos espaciais representados pelo cdg;
- ② Poder **combinar informação geográfica com diferentes sistemas de coordenadas**. Dados dois cdg com CRS distintos, é necessário poder representá-los no mesmo CRS, com o menor erro posicional possível.

Por exemplo, dados dois cdg matriciais

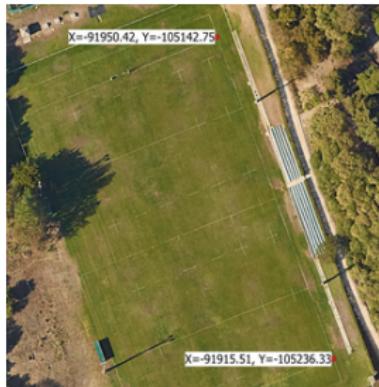
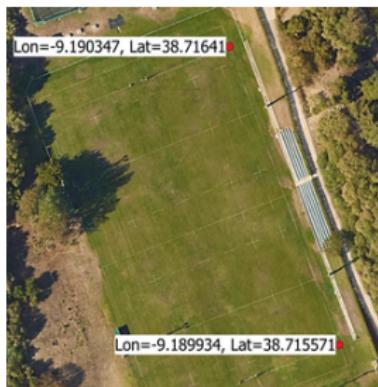
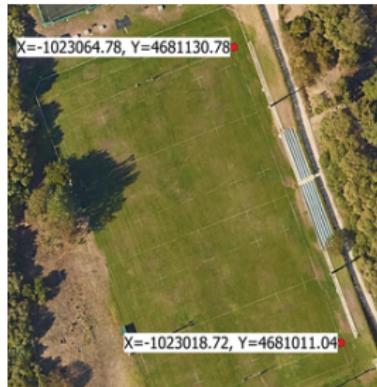
- ▶ cdg A com sistema de coordenadas de referência CRS_A , e
- ▶ cdg B com sistema de coordenadas de referência CRS_B ,

a técnica de **reamostragem** discutida no Slide 133 permite converter um ou ambos cdg para um novo sistema de coordenadas comum (que pode ser CRS_A , CRS_B ou outro CRS arbitrário).

Exemplo: campo de rugby do ISA

A mesma localização sobre o globo pode ser identificada através de coordenadas distintas.

No exemplo abaixo, os mesmos 2 pontos (campo de rugby do ISA) são representados em 3 sistemas de coordenadas de referência distintos.



EPSG:3857 (Google) EPSG:4326 (WGS84) EPSG:3763 (TM-PT06)

Qual é a distância correcta entre os dois pontos?

A forma da Terra: geoide

(Geoide)

É a superfície equipotencial do campo gravítico terrestre que se ajusta melhor ao nível médio das águas do mar.

O geoide é uma superfície irregular devido à heterogeneidade da constituição do globo terrestre.

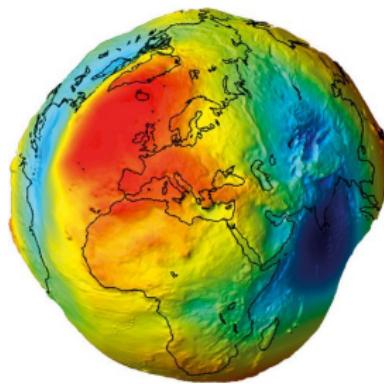


Figura: Representação (muito exagerada) das variações do geoide: o ponto mais elevado situa-se sobre o Oceano Atlântico e o ponto mais baixo sobre o Oceano Índico.

O sistema de referência WGS84

O sistema de referência *World Geodetic System* (WGS84) designa um *datum* e um elipsoide de referência para a Terra usado pelo Departamento de Defesa dos EUA e referência para as coordenadas GPS (*Global Positioning System*).

O sistema WGS84 é um *datum global* pois é ajustado à totalidade da superfície terrestre:

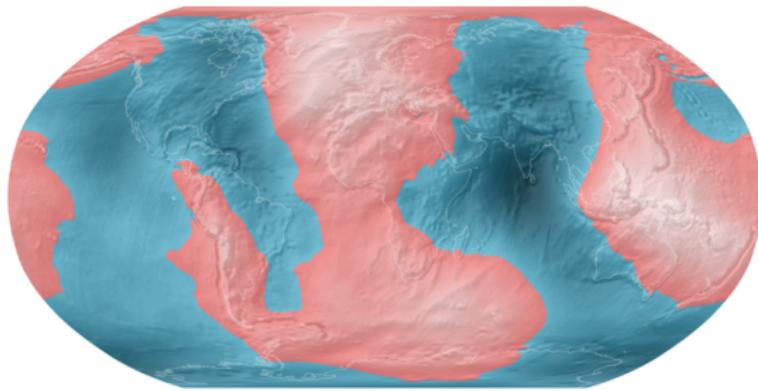
- ① O centro do elipsoide é colocado no centro de gravidade da Terra;
- ② O eixo dos Z passa pelo polo de referência do *International Earth Rotation and Reference Systems Service* (IERS); O eixo dos X passa pela intersecção do meridiano de referência do IERS com o plano que passa pela origem e é normal ao eixo dos Z; o eixo do Y completa o referencial ortogonal;
- ③ A sua superfície é um esferoide com raios $a = 6378137\text{m}$ (equatorial) e $b = 6356752.31\text{m}$ (polar).

Aproximação do geoide pelo datum WGS84

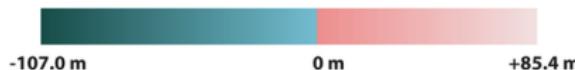
O datum WGS84 dá uma muito boa aproximação para o modelo do geoide pois a diferença máxima entre ambas as superfícies é inferior a 107m.

Deviation of the Geoid from the idealized figure of the Earth

(difference between the EGM96 geoid and the WGS84 reference ellipsoid)



Red areas are above the idealized ellipsoid; blue areas are below.



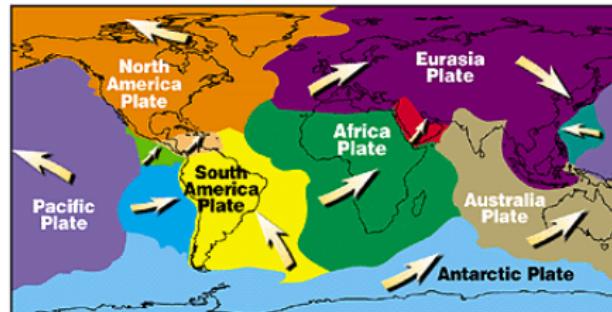
A linha de separação entre vermelho e azul é a linha em que o modelo do geoide intersecta o datum WGS84.

Outros *datums* globais

O *datum* WGS84 não está ligado a nenhuma placa tectónica. Por isso, ao longo do tempo, localizações sobre a superfície sofrem mudanças de coordenadas WGS84. Para minimizar essas alterações, muitos estados adoptam um *datum* global ligado a sua placa tectónica.

- A UE adoptou o *datum* ETRS89 (baseado no elipsoide GRS80), ligado à Placa Eurasiática;
- Os EUA e Canadá adoptaram o *datum* NAD83 (também baseado no elipsoide GRS80), ligado à Placa Norte-Americana.

A Placa Eurasiática tem um movimento de 1 a 2 cm por ano relativamente ao *datum* WGS84.



Sistema de coordenadas de referência

O sistema de coordenadas de referência (CRS) contém dois elementos: o *datum* e o sistema de coordenadas:

- ① O *datum* determina como o sistema de coordenadas está ligado à superfície terrestre (origem, escala e orientação dos eixos coordenados);
- ② O sistema de coordenadas é estabelecido por regras matemáticas que associam a cada localização um vector de coordenadas.

As coordenadas podem ser:

- ① **Cartesianas 3D**, i.e. segundo os três eixos ortogonais do elipsoide: ver Slide 185);
- ② **Geográficas**, i.e. longitude, latitude e altitude normal ao elipsoide;
- ③ **Rectangulares**: coordenadas x , y resultantes de uma projecção cartográfica, e altitude normal ao elipsoide.

Coordenadas geográficas: longitude, latitude

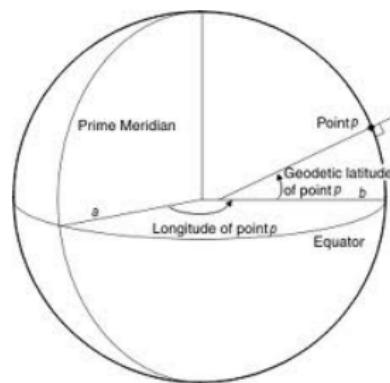
O sistema de coordenadas geográficas tem coordenadas longitude e latitude. Adicionalmente, por ser registada a altitude (ver Slide 197).

Latitude Ângulo ϕ (positivo ou negativo) entre o **plano equatorial** e a linha que passa pelo ponto e que é **normal** à superfície do elipsoide;

Longitude Ângulo λ (positivo ou negativo) entre o **meridiano de referência** e o meridiano que contém o ponto.

Altitude distância positiva ou negativa à superfície de referência.

- A latitude varia entre -90° (S) e 90° (N);
- A longitude varia entre -180° (W) e 180° (E).



Coordenadas geográficas: notação para longitude e latitude

Longitude e latitude são **ângulos**, que podem ser expressos como:

- Graus, minutos e segundos decimais (e.g. lat=38° 50' 7.296"), a notação GMS. O círculo completo tem 360°, cada grau tem 60 minutos, e cada minuto tem 60 segundos;
- Graus decimais (e.g. lat=38.83536°);
- Radianos: o círculo tem 2π radianos. Por isso,
$$1\text{rad} = \frac{360}{2\pi} = \frac{180}{\pi} \approx 57.29577^\circ$$

Conversões:

- GMS para graus decimais: $38^\circ 50' 7.296'' = 38 + 50/60 + 7.296/3600 = 38.83536$ graus decimais; em alternativa, GMS podem ser convertidos em segundos, e segundos em graus:
 $38^\circ 50' 7.296'' = 38 \times 3600 + 50 \times 60 + 7.296 = 139807.296''$ e
 $139807.296''/3600 = 38.83536^\circ$.
- Graus decimais para radianos: multiplicar por $\frac{\pi}{180}$. Por exemplo,
 $38.83536^\circ = \frac{\pi}{180} 38.83536 = 0.6778049\text{rad}$;
- Radianos em graus decimais: multiplicar por $\frac{180}{\pi}$. Por exemplo,
 $0.8\text{rad} = \frac{180}{\pi} 0.8 = 45.83662^\circ$.

Coordenadas geográficas: distâncias em arco-segundos

A resolução de dados matriciais é por vezes indicada em **arco-segundos** (segundos de arco).

Por exemplo, até recentemente, o cdg matricial SRTM (ver Slide 128) era disponibilizado para a generalidade do globo com uma resolução de 3 arco-segundos, i.e. cada pixel SRTM cobre 3 segundos de longitude e 3 segundos de latitude sobre o terreno.

Para converter essa resolução numa distância sobre o terreno **ao longo de um meridiano** pode proceder-se como de seguida:

- 1 Considerar que a superfície terrestre é relativamente bem aproximada por uma esfera com raio de 6371 km;
- 2 Converter 3 arco-segundos em radianos: 3 arco-segundos são $3/3600 = 0.0008333^\circ = 0.000014544\text{rad}$;
- 3 Para um círculo de raio r , o comprimento do arco de círculo de amplitude θ é simplesmente $\theta \times r$. Assim, 3 arco-segundos de latitude corresponde a uma distância no terreno de aproximadamente
 $6371 \times 0.000014544 = 0.09266244\text{km} \approx 92.6\text{m}$.

Coordenadas geográficas e distâncias sobre a esfera: exercício

Exercício. Resolver o problema análogo ao do slide anterior, mas calculando aproximadamente a distância sobre o terreno **ao longo de um paralelo**, e para a latitude de Lisboa ($38^{\circ} 42'49.72''\text{N}$).

Sugestão:

- ① Considerar que a superfície terrestre é relativamente bem aproximada por uma esfera com raio de 6371 km;
- ② Determinar o raio do paralelo de Lisboa;
- ③ Converter a amplitude de 3 arco-segundos numa distância no terreno para esse paralelo.

Solução: aproximadamente 72m.

Observação: A solução do exercício anterior (distância S-N de 92.6m) e deste exercício (distância E-O de 72m) indicam a dimensão aproximada para a área do terreno que corresponde a um pixel SRTM (3 arco-segundos) em Portugal Continental.

Coordenadas geográficas e distâncias sobre o elipsoide

Como se viu no exemplo anterior, um arco de 1° corresponde:

- ① uma distância no terreno **ao longo de um meridiano** (Δ_{lat}^1) que é aproximadamente constante, qualquer que seja o lugar;
- ② uma distância no terreno **ao longo de um paralelo** (Δ_{long}^1) que depende da latitude do lugar.

A tabela abaixo dá os valores precisos (para o elipsoide WGS84) das distâncias no terreno correspondendo a um arco de 1° para um conjunto de latitudes:

latitude (ϕ)	0°	15°	30°	45°	60°	75°	90°
Δ_{lat}^1 (km)	110.574	110.649	110.852	111.132	111.412	111.618	111.694
Δ_{long}^1 (km)	111.320	107.551	96. 486	78.847	55.800	28.902	0

Nota: para a totalidade do globo, o erro máximo cometido por se considerar uma superfície esférica em vez do elipsoide WGS84 é 0.5% da distância real.

Exercício do campo de rugby do ISA (continuação)

Qual é a distância entre os dois extremos do campo de rugby do ISA (ver Slide 183) calculada a partir dos valores de **longitude** e **latitude** sobre o elipsoide WGS84:

Pontos	longitude (WGS84)	latitude (WGS84)
P_1	-9.190347°	38.716410°
P_2	-9.189934°	38.715571°

As distâncias são:

- $\text{dist}(P_1, P_2) = 99.82374 \text{ m}$, sobre o elipsoide WGS84;
- $\text{dist}(P_1, P_2) = 99.93717 \text{ m}$, sobre a esfera de raio 6371 km.

A medição mais precisa é **99.82374 m**.

Coordenadas geográficas: exercício

A grelha *Climate Modeling Grid* (CMG) dos dados MODIS assegura uma cobertura global em coordenadas geográficas (latitude e longitude) com uma resolução de 0.05° .

- 1 Explique por que razão a área na superfície correspondente a uma célula da grelha não é igual para todas as células.
- 2 Mostre que uma cobertura global da superfície do globo contém 25920000 células da grelha.
- 3 Qual é a dimensão aproximada (em km^2) de uma célula da grelha junto ao Equador (considere o modelo esférico para a Terra com raio de 6371 km)?

Coordenadas geográficas: solução do exercício

- ➊ Dado que a resolução da grelha CMG é constante e é dada em graus de latitude e longitude, e dado que um grau de longitude corresponde a uma distância no terreno que varia com a latitude do lugar, uma célula da grelha terá uma área que depende da latitude.
- ➋ A superfície da Terra é coberta por 360×180 células $1^\circ \times 1^\circ$. Cada grau corresponde a 20 células da grelha CMG ($20 \times 0.05^\circ = 1^\circ$). Então, em $1^\circ \times 1^\circ$ existem $20 \times 20 = 400$ células CMG. No total há $360 \times 180 \times 400 = 25920000$ células de 0.05° na grelha CMG.
- ➌ Junto ao Equador, e usando o modelo esférico, as distâncias no terreno correspondentes a 0.05° de latitude e de longitude são iguais. O ângulo é $\theta = 0.05 \times \frac{\pi}{180} = 0.00087266462$ rad e por isso a distância no terreno é $6371 \times 0.00087266462 = 5.559$ km. Portanto, a área de uma célula é aproximadamente $5.559^2 \text{ km}^2 = 30.9 \text{ km}^2$.

Altitude ortométrica e elipsoidal

Existem dois conceitos principais de altitude:

- 1 Altitude ortométrica ou relativa ao nível médio das águas do mar
 H é a distância ao geoide, medida ao longo da linha de fio-de-prumo; Esta é a altitude usada em topografia; A altitude dada pelo SRTM (ver Slide 167) é a altitude ortométrica;
- 2 Altitude elipsoidal h é a distância ao elipsoide de referência; É medida pelos sistemas de posicionamento globais (e.g. GPS).
A diferença $N = h - H$ é designada por *ondulação do geoide*.

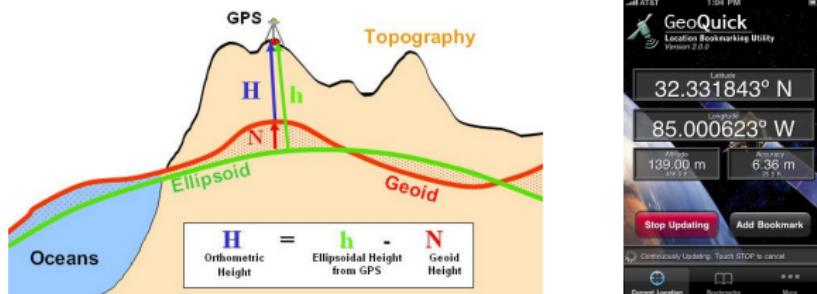


Figura: Ilustração da conversão de altitude elipsoidal medida por GPS numa altitude ortométrica.

Exercício de revisão

WGS84 é um datum global e um sistema de coordenadas de referência (CRS) de uso muito comum.

- 1 As coordenadas WGS84 são coordenadas geográficas ou cartográficas?
- 2 Indique uma tecnologia de utilização generalizada desde 1995 que é suportada por coordenadas WGS84?
- 3 Quais são as duas formas de medir a elevação que podem ser combinadas com as coordenadas WGS84?
- 4 Na UE, o datum oficial é ETRS89. Qual a distinção mais relevante entre esses dois datums (WGS84 e ETRS89)?

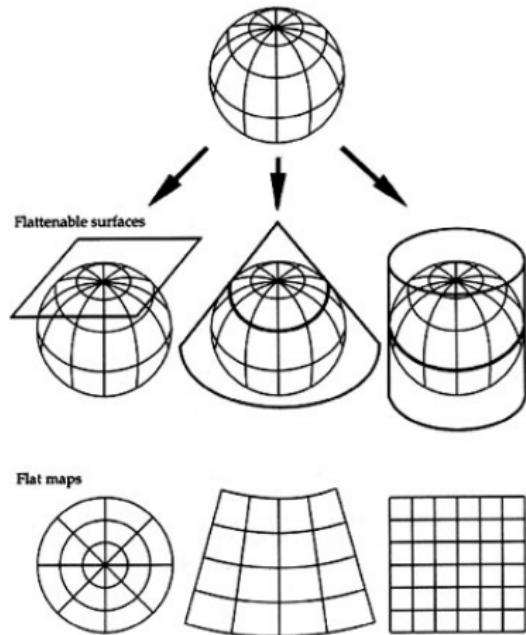
Projeções cartográficas

(Projeção cartográfica)

É uma transformação que faz corresponder, biunivocamente, a cada ponto P numa superfície de referência um ponto P' no plano.

Embora a maioria das projeções usadas não sejam propriamente geométricas, como simplificação pode considerar-se que em geral são adaptações das seguintes projeções geométricas, em que a superfície plana é tangente ou secante à superfície de referência.

- azimutal,
- cónica,
- cilíndrica



Sistemas de coordenadas de referência da UE

A Europa tem uma grande tradição na área da cartografia. Cada país desenvolveu ao longo dos séculos diversas formas de representar a superfície terrestre e diversos sistemas de projecção. No ano 2000, ainda eram usados 5 tipos diferentes de elipsoides de referência e 8 tipos diferentes de projecções cartográficas nos vários países da UE. Em 2003, após um estudo técnico aprofundado, a UE recomendou:

- ① a adopção de ETRS89 como *datum* de referência para a UE;
- ② a utilização de coordenadas geográficas sobre ETRS89 para expressão e arquivar as localizações sobre o território da UE;
- ③ a utilização de 3 projecções cartográficas, suficientes para as necessidades dos programas da UE.

Foi igualmente recomendado que fossem tomadas as disposições necessárias para que os sistemas de coordenadas de referência indicados pudessem coexistir com os sistemas de coordenadas de referência tradicionalmente usados nas regiões e países da UE.

Projecções cartográficas adoptadas na UE

A directiva INSPIRE 2007/2/EC estabelece as seguintes projecções:

- ① ETRS89-TMzn: **ETRS89 – Transversa de Mercator**, uma projecção **conforme** recomendada para escalas maiores do que 1:500 000; depende da zona (zn) em que a projecção é centrada;
- ② ETRS89-LCC: **ETRS89 – Lambert Conic Conformal**, uma projecção **conforme** recomendada para escalas menores do que 1:500 000;
- ③ ETRS89-LAEA: **ETRS89 – Lambert Azimuthal Equal Area**, uma projecção **equivalente** recomendada para cartas estatísticas a qualquer escala.

Propriedades das projecções cartográficas:

- ① Projecção **conforme**: a forma é preservada, i.e. a escala de projecção em torno de um ponto P é a mesma em todas as direcções;
- ② Projecção **equivalente**: a área é preservada.

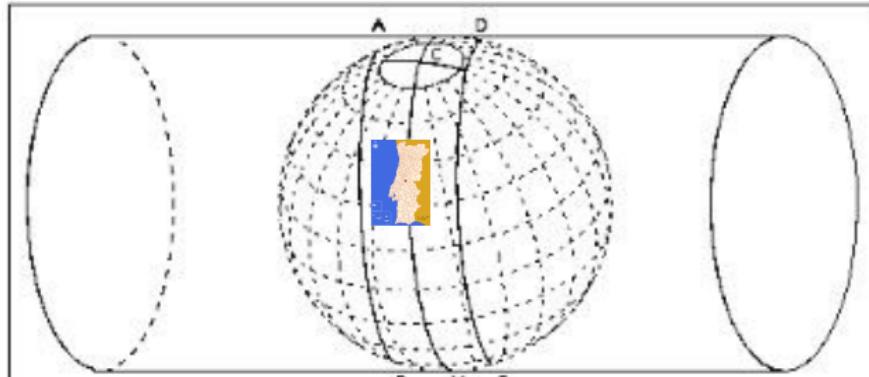
Nota: uma projecção cartográfica não pode ser simultaneamente conforme e equivalente.

A projecção ETRS89-TM-PT06

A projecção transversa de Mercator (i.e., o cilindro sobre o qual é feita a projecção é transverso) é muito adequada para Portugal Continental, dada a forma do seu território.

Por forma a que as distorções de escala sejam mínimas sobre o território de Portugal Continental, o **ponto central e origem das coordenadas cartográficas** da projecção é o seu centro geográfico:

- ① longitude $08^{\circ} 07'59.19''$ W
- ② latitude $39^{\circ} 40'05.73''$ N



As projecções ETRS89-LCC e ETRS89-LAEA

As projecções ETRS89-LCC (conforme) e ETRS89-LAEA (equivalente) são únicas para o território da UE.

- ETRS89-LCC tem dois paralelos padrão (35° N e 65° N). O ponto central tem coordenadas 52° N e 10° E; A origem das coordenadas cartográficas está deslocada de 2 800 000m para sul e de 4 000 000m para oeste.
- ETRS89-LAEA. O ponto central tem coordenadas 52° N e 10° E; A origem das coordenadas cartográficas está deslocada de 3 210 000m para sul e de 4 321 000m para oeste.

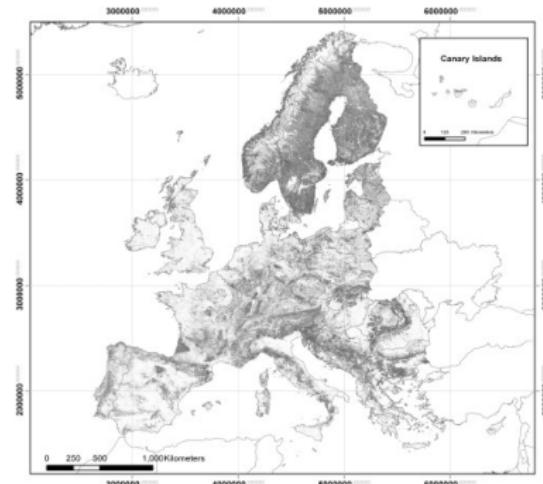


Figura: Um mapa de áreas florestadas usando o CRS ETRS89-LAEA (Pekkarinen *et al.* 2009)

Descrição em formato proj4 dos sistemas de coordenadas

Os CRS podem ser descritos, por exemplo, pelo seu código EPSG ou pela descrição proj4. O datum de referência é WGS84.

- ① **ETRS89-TM-PT06.** EPSG 3763: +proj=tmerc
+lat_0=39.66825833333333 +lon_0=-8.133108333333334
+k=1 +x_0=0 +y_0=0 +ellps=GRS80 +units=m
- ② **ETRS89-LCC.** EPSG 3034: +proj=lcc +lat_1=35 +lat_2=65
+lat_0=52 +lon_0=10 +x_0=4000000 +y_0=2800000
+ellps=GRS80 +units=m
- ③ **ETRS89-LAEA.** EPSG 3035: +proj=laea +lat_0=52 +lon_0=10
+x_0=4321000 +y_0=3210000 +ellps=GRS80 +units=m

em que +proj indica o tipo de projecção, +lat_0 e +lon_0 o centro da projecção, +lat1 e +lat2 as latitudes padrão da projecção cónica, +x_0 e +y_0 a posição da origem das coordenadas cartográficas, +ellps o elipsoide e +units as unidades.

A distorção provocada pela projecção é mínima junto à localização (lon_0,lat_0), variando consoante o tipo de projecção.

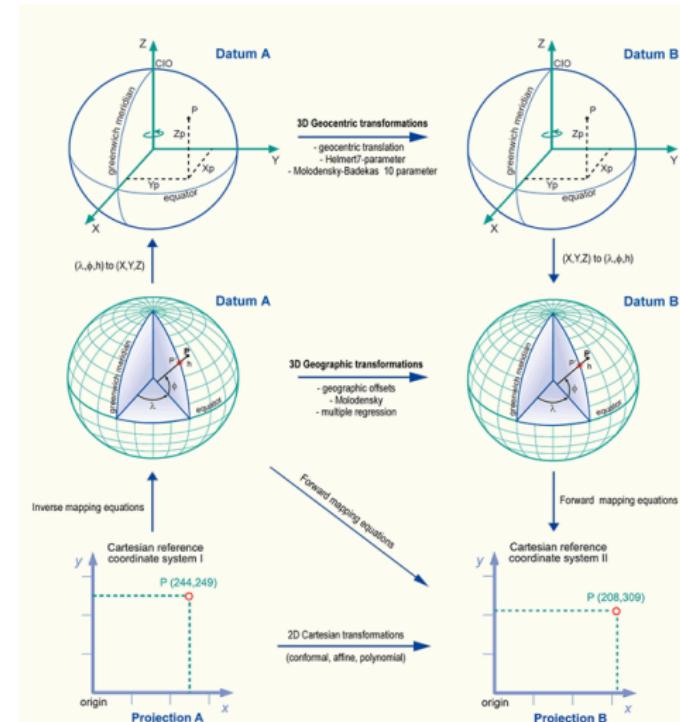
Transformação de coordenadas e transformação de datum

Quando se usa cdg's com sistemas de coordenadas de referência distintos, pode ser necessário converter coordenadas de um CRS_A para um CRS_B.

Quando o *datum* é o mesmo (*datum A*), aplica-se a **projecção inversa** sobre CRS_A seguida da projecção no CRS_B.

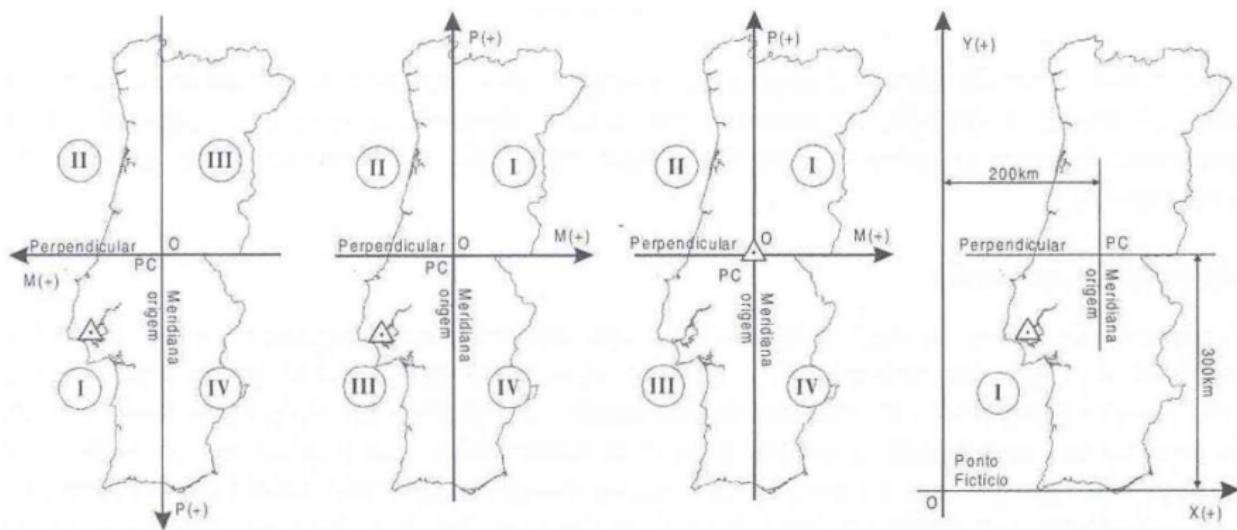
Se os CRS têm **datums diferentes** (*datums A e B*) pode combinar-se as projecções cartográficas com uma **transformação de datum** (A para B) usando:

- 1 coordenadas cartesianas x, y, z ; ou
- 2 coordenadas geográficas.



Sistemas obsoletos de coordenadas para Portugal Continental

Existe muita cartografia do território continental português usando um **datum que não é ETRS89**, mas que pode ser convertida para os sistemas adoptados.



Os CRS representados acima são Bessel-Bonne-Lisboa, Hayford-Gauss-Lisboa (HG-Lx), Hayford-Gauss-Datum73 (HG73) e Hayford-Gauss-Militar (HG-militar) (fonte: Gaspar, 2006).

Transformação de *datum*: exemplo

Considere-se o CRS Hayford–Gauss–Datum73 (ver Slide 206):

```
EPSG 27493: +proj=tmerc +lat_0=39.666666  
+lon_0=-8.131906111111112 +k=1+x_0=180.598  
+y_0=-86.98999999 +ellps=intl +units=m
```

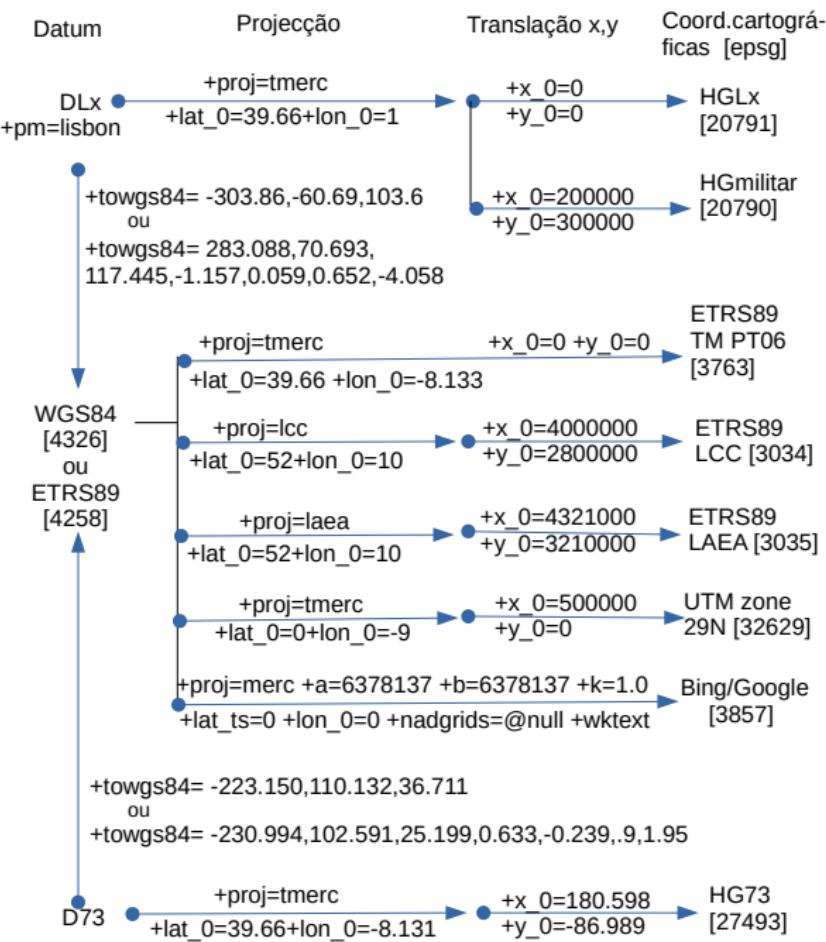
A descrição acima tem que ser completada com os **parâmetros da transformação de *datum*** para o *datum* WGS84, usando coordenadas cartesianas 3D (+towgs84) ou coordenadas geográficas (+nadgrids) para ser compatível com outros CRS.

- ① de 3 parâmetros: +towgs84=-223,110,37,0,0,0,0, com erros posicionais até 2m sobre Portugal Continental;
- ② de 7 parâmetros: +towgs84=-230.994,102.591, 25.199,0.633,-0.239, 0.900,1.950, com erros até 1m;
- ③ baseado na grelha NTv2: +nadgrids=pt73_e89.gsb, com erros até 28cm.

Resumo

Principais projeções e algumas transformações de datum relevantes para a cartografia de Portugal Continental.

Nota: apenas alguns coeficientes são indicados; alguns estão truncados.



Exercício do campo de rugby do ISA (conclusão)

Calcular a distância entre os dois pontos do campo de rugby do ISA (ver Slide 183) a partir das suas coordenadas projectadas:

Pontos	x (EPSG:3857)	y (EPSG:3857)	x (EPSG:3763)	y (EPSG:3763)
P_1	-1023064.78	4681130.78	-91950.42	-105142.75
P_2	-1023018.72	4681011.04	-91915.51	-105236.33

As distâncias euclidianas para as coordenadas acima são:

- $\text{dist}(P_1, P_2) = 128.2934 \text{ m}$, para EPSG:3857 (Google)
- $\text{dist}(P_1, P_2) = 99.87955 \text{ m}$, para EPSG:3763 (ETRS89-TM-PT06)

A medição sobre coordenadas EPSG:3763 é muito precisa. O valor obtido é próximo do valor medido sobre o elipsoide WGS84 que é 99.82374 m.

O valor dado pelas coordenadas “Google” deve-se à distorção gerada pela projecção de Mercator (não transversa), que provoca um afastamento dos pontos que aumenta à medida que se afastam do Equador. Por isso não se deve medir distâncias sobre EPSG:3857.

Exercício

O CRS obsoleto conhecido como “coordenadas militares” para Portugal Continental (Hayford Gauss Militar) tem os seguintes parâmetros:

```
+proj=tmerc +lat_0=39.66666666666666 +lon_0=1 +pm=lisbon  
+x_0=200000 +y_0=300000 +ellps=intl +units=m +k=1  
+towgs84=-283.088,-70.693,117.445,-1.157,0.059,-0.652,-4.05
```

- 1 O datum deste CRS é na prática idêntico ao datum WGS84?
- 2 Em que situação o parâmetro `+towgs84` é opcional? E obrigatório?
- 3 A projecção é conforme. Qual é a propriedade preservada por este projecção?
- 4 Onde se situa o centro da projecção, com distorção nula?
- 5 Qual é a justificação para a escolha dos valores `+x_0=200000 +y_0=300000`?

Imagens multiespectrais

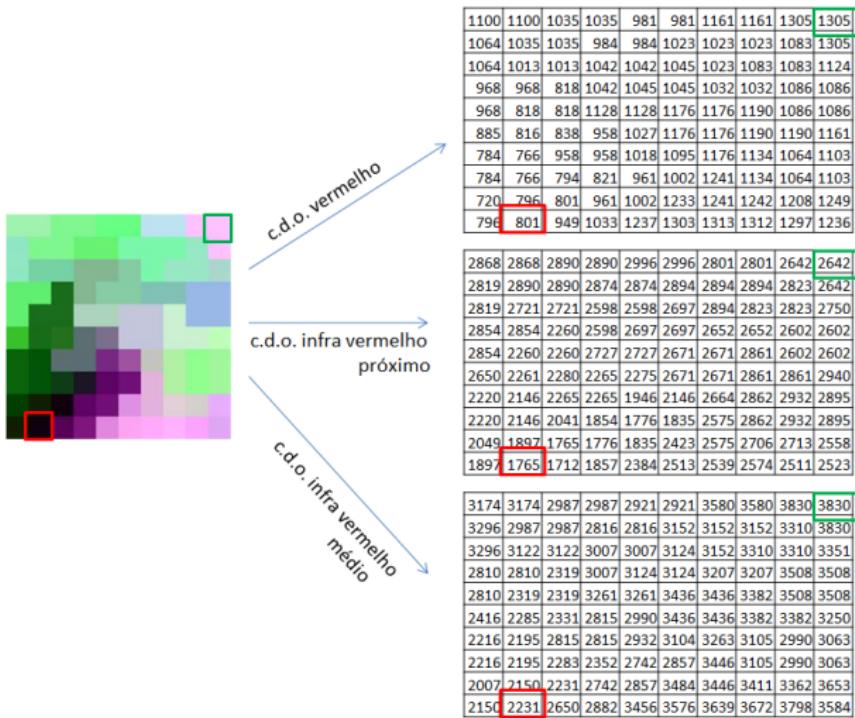
Imagens multiespectrais obtidas por deteção remota

Deteção remota designa o conjunto de técnicas para obtenção de informação sobre um objecto por um sensor que não está em contacto físico com o objecto e para interpretação dessa informação.

Contudo, detecção remota envolve tipicamente informação que pode ser representada por um conjunto de valores de pixels de uma imagem e que pode ser processada digitalmente.

Em geral, as imagens de detecção remota são obtidas através de sensores digitais instalados em satélites, mas podem igualmente ser obtidos por sensores instalados em aviões, em laboratório, ou por sensores portáteis usados em trabalho de campo.

Imagens multiespectrais (ilustração)



As imagens são em geral **multiespectrais**. Contém a resposta ao sinal à superfície para distintos comprimentos de onda.

Regiões do espectro electromagnético

Espectro electromagnético é a gama de comprimentos de onda (ou alternativamente, frequências) da radiação electromagnética.

O espectro electromagnético é habitualmente decomposto em **regiões**, que vão dos raios gama (menor λ , maior frequência e maior energia) até às ondas longas usadas em telecomunicações (maior λ , menor frequência e menor energia).

Para a detecção remota da superfície terrestre as regiões mais importantes são

- a região do **visível** (0.4 a 0.7 μm),
- a região do **infravermelho próximo** (0.7 a 1.2 μm),
- a região do **infravermelho médio** (1.2 a 8 μm),
- a região do **infravermelho térmico** (8 a 14 μm), e
- região das **microondas**, de comprimento de onda superior a 1mm.

Regiões do espectro electromagnético (cont.)

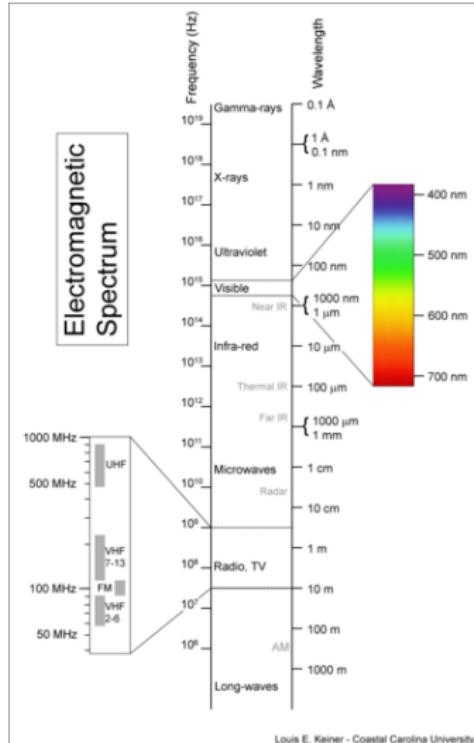


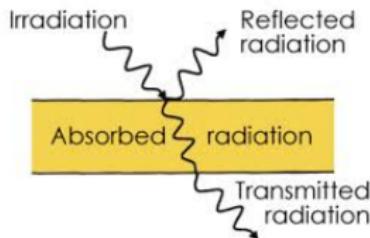
Figura: Regiões do espectro electromagnético

<http://www.scienceprog.com/>

Reflectância, transmitância e absorptância

- **Emissividade** ε : razão entre radiação emitida pelo objecto e a radiação emitida por um corpo negro a igual temperatura. Varia entre 0 e 1 e depende do comprimento de onda.
- **Reflectância** (ρ), **absorptância** (α) e **transmitância** (τ): cada um é a proporção da energia incidente que é respectivamente reflectida em qualquer direcção, absorvida e transmitida. Dependem do comprimento de onda.

$$\rho(\lambda) + \alpha(\lambda) + \tau(\lambda) = 1.$$



Assinaturas espectrais

A **assinatura espectral** é o padrão de reflectância em função do comprimento de onda da radiação incidente sobre o objecto.

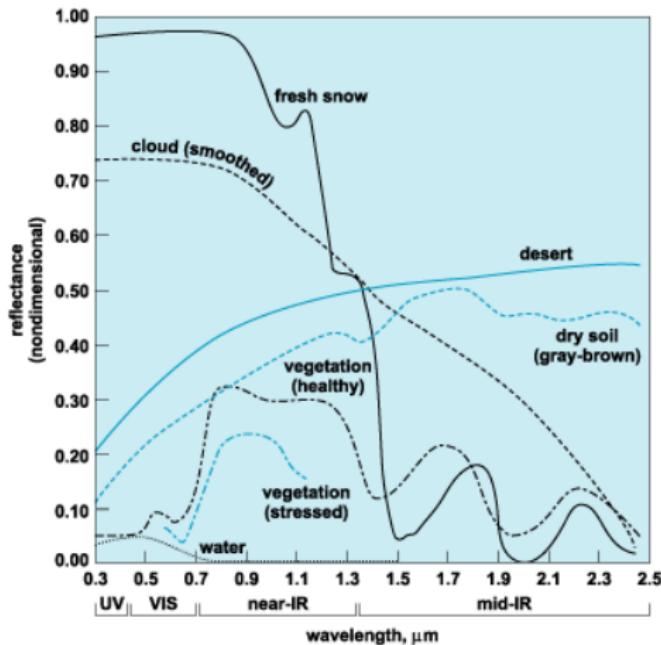


Figura: Assinaturas espectrais típicas de neve, nuvens, deserto, solo nú, vegetação e água, <http://accessscience.com/>

Factores que afectam a assinatura espectral típica em detecção remota

- As imagens são obtidas em determinadas condições de iluminação e observação que afectam em geral o sinal recebido pelo sensor;
- As bandas dos sensores são em número finito (há uma discretização do sinal);
- O sinal no sensor depende não só pelo material à superfície mas também do efeito da atmosfera;
- Um pixel da imagem pode corresponder a uma mistura de diferentes materiais com assinatura espetrais distintas.

Esses factores introduzem “ruído” na assinatura espectral típica e tornam mais difícil a distinção dos materiais à superfície a partir de imagens multiespectrais de detecção remota.

Exemplo de ocupações do solo com assinaturas espectrais semelhantes

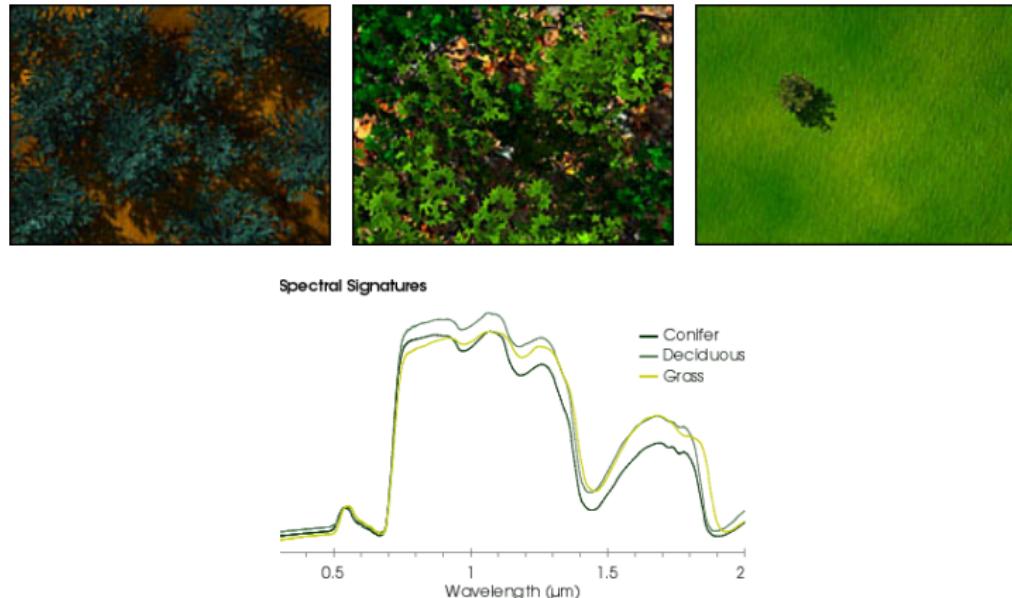


Figura: Fotografias verticais e respectivas assinaturas espectrais de solo coberto com coníferas, folhosas e relva

<http://earthobservatory.nasa.gov/>

Assinaturas espectrais da vegetação

- **região do visível** A clorofila absorve mais na região do azul e do vermelho (com picos de absorção em aproximadamente $0.43\text{-}0.45$ e $0.64\text{-}0.67 \mu\text{m}$) do que no verde (pico de reflexão em $0.54 \mu\text{m}$ aproximadamente). Em situação de senescência ou de stress, o teor de clorofila diminui e a cor das folhas é mais afectada por outros pigmentos como os carotenos, provocando a mudança de cor de verde para vermelho-amarelado como acontece em certas plantas no Outono. Em geral, a reflectância da vegetação na região espectral do visível (0.3 a $0.7 \mu\text{m}$) é baixa.
- **infravermelho próximo** A estrutura interna das folhas causa um pico de reflectância na região do IV próximo (de 0.7 a $1.3 \mu\text{m}$) com valores que podem atingir 50%.
- **infravermelho médio** Existem picos de absorção em 1.4 , 1.9 e $2.7 \mu\text{m}$ aproximadamente, causados pela presença de água nas folhas.

Vegetação: assinaturas espectrais de pigmentos

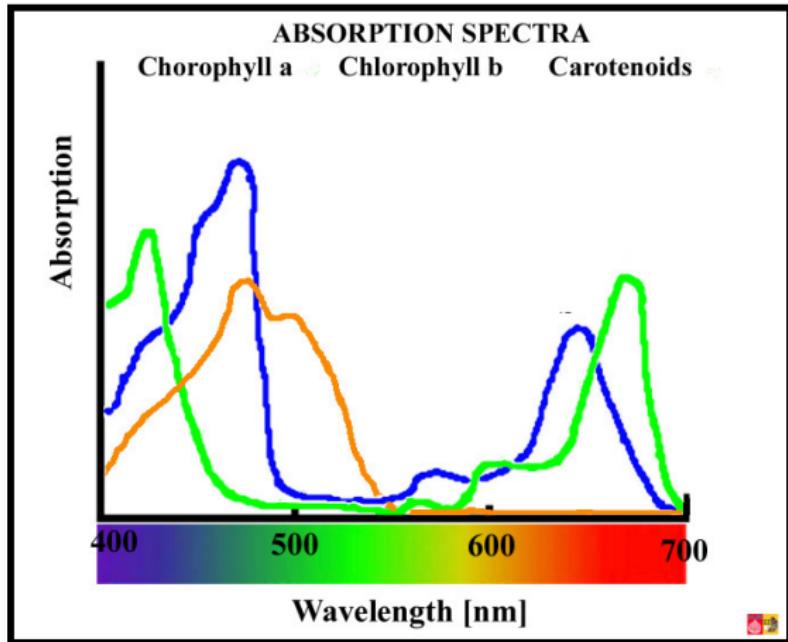


Figura: Assinaturas espectrais de vários pigmentos da vegetação,
<http://www.vcbio.science.ru.nl/>

Assinaturas espectrais da água

A água é caracterizada por absorver ou transmitir a maior parte da radiação que a atinge.

Em geral, quanto maior o comprimento de onda, menor a quantidade reflectida. Para comprimentos de onda superiores a $1 \mu\text{m}$, a reflectância torna-se praticamente nula.

A reflectância de superfícies aquáticas pode aumentar na presença de **sedimentos** – o aumento da reflectância depende da natureza e das propriedades ópticas dos sedimentos (por exemplo, presença clorofila ou de detritos minerais de dimensões variáveis) –, quando **a profundidade é reduzida**, ou quando a superfície de água é muito regular (espelhada) e o ângulo de observação do sensor é elevado.

Assinaturas espectrais do solo

Ao contrário da vegetação, a transmissão da radiação nos solos é baixa, e por isso a assinatura espectral destes materiais depende essencialmente das sua camada superior.

Tipicamente, a reflectância dos solos aumenta com o c.d.o.

Os principais factores que afectam a sua reflectância são a **cor** do solo, a **classe** de solos, o teor de **matéria orgânica**, o teor de **ferro**, a composição **mineralógica**, o teor de **humidade**, e a **textura** do solo.

Em geral, um alto teor de **matéria orgânica** provoca – principalmente na região do visível – uma diminuição de reflectância, e um aumento do **teor de humidade** provoca também uma redução da reflectância.

A reflectância também diminui em geral com o aumento do **tamanho das partículas** (textura grosseira) dado que nessa situação, a radiação pode mais facilmente ficar retida no solo.

Efeito da atmosfera sobre a radiação

A atmosfera é constituída por gases (entre os quais o azoto 78%, oxigénio 21%, argon 0.9%, vapor de água 0.04%, dióxido de carbono 0.033%, e o ozono 0.012%) e por partículas sólidas suspensas nos gases atmosféricos designadas por aerossóis (como sal, poeiras, e partículas diversas resultantes da poluição atmosférica e de fogos).

Alguns gases têm uma elevada absorção para alguns intervalos de comprimento de onda.

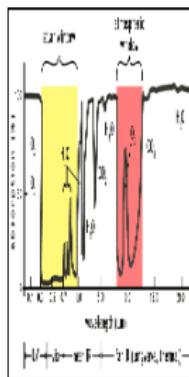


Figura: Efeito de gases na absorção atmosférica e janelas de observação.

Efeitos de Rayleigh e de Mie

A dispersão atmosférica depende da dimensão das partículas da atmosfera que a causam e do c.d.o. (λ) da radiação.

- **Efeito de dispersão de Rayleigh:** é causado pelas moléculas com diâmetro inferior a aproximadamente $\frac{1}{10} \lambda$ e é **inversamente proporcional a λ^4** . Por exemplo, a dispersão causada pela atmosfera na banda do verde é aproximadamente quatro vezes superior à dispersão no infravermelho médio (de maior λ).
- **Efeito de dispersão de Mie:** ocorre para comprimentos de onda semelhantes ao diâmetro de partículas da atmosfera e também depende do comprimento de onda da radiação, embora essa dependência seja atenuada relativamente à dispersão de Rayleigh. A dispersão de Mie é causada sobretudo por **aerossóis**, e em particular por poeiras.
- **Dispersão não selectiva:** é causada por partículas de dimensão maior do que o comprimento de onda (como gotículas de água nas nuvens). Afeta da mesma forma radiação de diferente comprimento de onda.

Efeitos de Rayleigh e de Mie: comparação

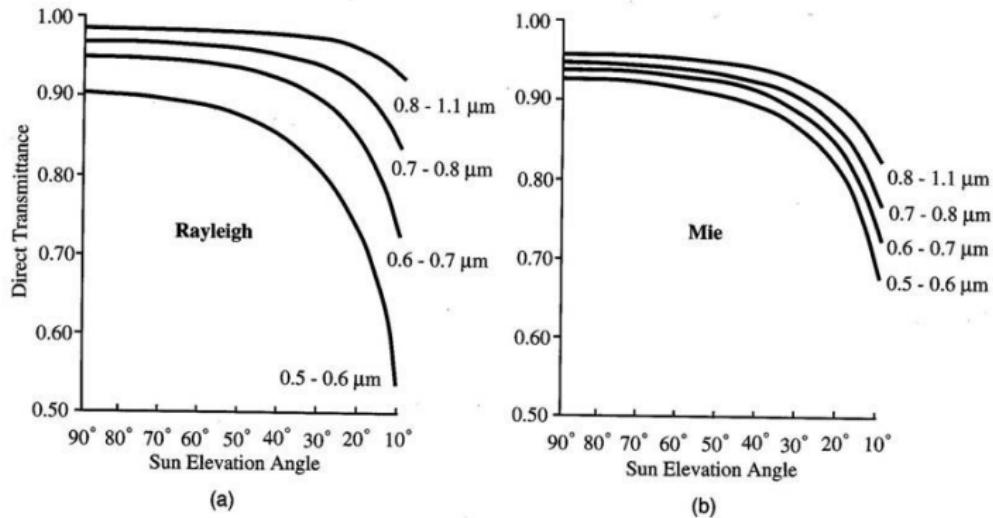


Figure 6-2 (a) Effects of molecular (Rayleigh) scattering and sun angle on incident radiation transmitted to Earth's surface; (b) Effects of aerosol (Mie) scattering and sun angle on radiation transmitted to Earth's surface.

Figura: A dispersão de Rayleigh é mais intensa do que a dispersão de Mie, sobretudo para baixos comprimentos de onda.

Efeitos de Rayleigh e de Mie (ilustração)

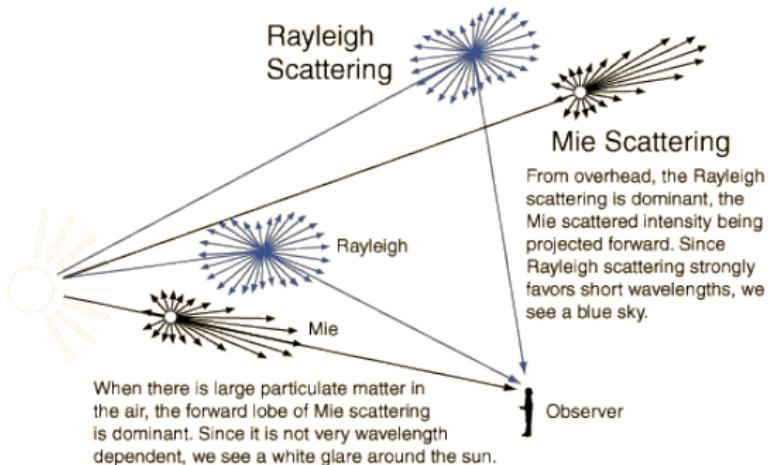


Figura: Ilustração do efeito de Rayleigh e do efeito de Mie sobre a radiação, na atmosfera, <http://hyperphysics.phy-astr.gsu.edu/>

- Porque é que o céu é azul?
- Porque é que o pôr-do-sol é vermelho?
- Porque é que as nuvens são brancas ou cinzentas?

Imagens de detecção remota: resolução espacial, espectral e temporal

Tipos de resolução das imagens

Principais tipos de resolução:

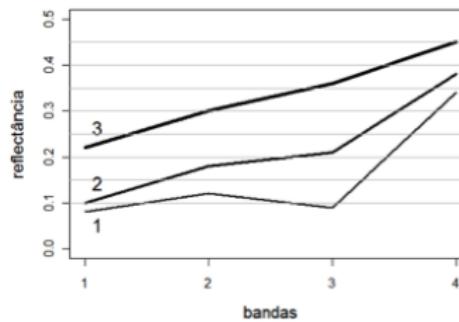
- Resolução espacial: comprimento do lado de uma parcela de terreno que corresponde a um pixel na imagem de satélite quando a observação é vertical. A resolução espacial de imagens pode ser de alguns centímetros (drones) até vários quilómetros (satélites meteorológicos).
- Resolução espectral: número de bandas do sensor e intervalo de comprimentos de onda associado a cada banda.
- Resolução temporal: de um sensor é o intervalo de tempo (medido em dias ou horas) entre observações sucessivas no mesmo local, possivelmente realizadas com ângulos de observação distintos.

Exercício

Considere uma imagem e algumas assinaturas espectrais para o sensor GeoEye1 com as seguintes características.

Band	Wavelength (μm)	Bandwidth (μm)	Resolution (m)	Swath width (km)	Revisit time (days)
Band 1 (VIS)	0.45 to 0.51		1.65	15.2	
Band 2 (VIS)	0.51 to 0.58		1.65	15.2	
Band 3 (VIS)	0.655 to 0.69		1.65	15.2	
Band 4 (NIR)	0.78 to 0.92		1.65	15.2	

Estabeleça uma associação entre cada um dos pontos assinalados (A,B,C) na composição colorida em cor verdadeira e as três assinaturas espectrais.

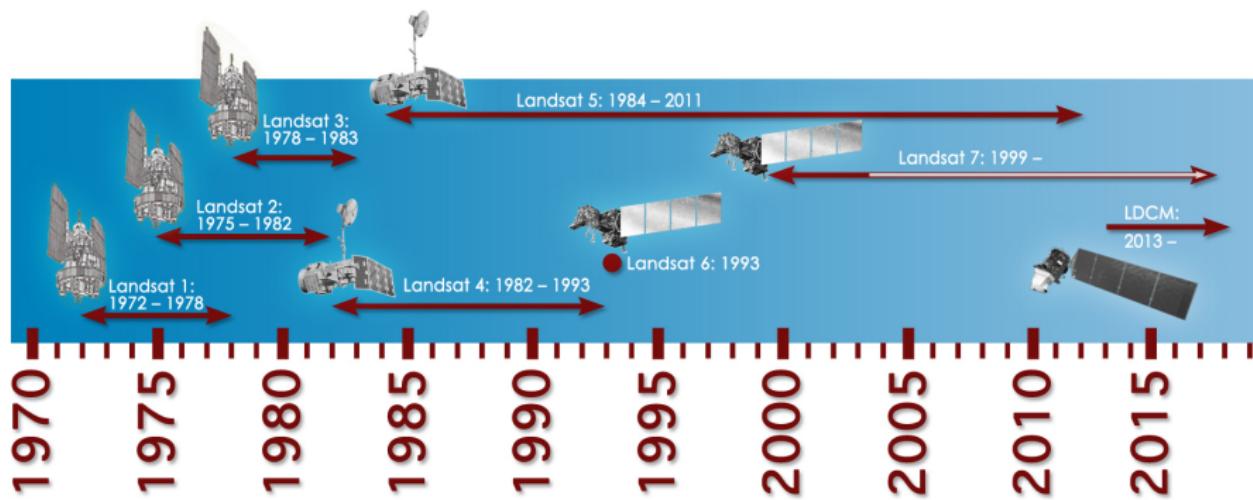


Acesso a dados de deteção remota

Acesso a dados de satélite

Existem diversos sites onde podem ser descarregados dados de satélite disponíveis para a comunidade.

Para além de dados recentes, estão disponíveis séries temporais já bastante longas (desde 1979 para AVHRR, desde 1972 para Landsat-MSS, e desde 1982 para Landsat-TM) que permitem estudos de alterações da superfície nas últimas 4 décadas.



Acesso a dados de satélite: principais sites

Hoje em dia estão disponíveis muitos tipos de produtos de detecção remota, com níveis de processamento diversos. Os dados para os utilizadores finais estão em geral corrigidos

- **geometricamente**, isto é estão ortorrectificados o que significa que estão projectados sobre a superfície por técnicas que minimizam as distorções geométricas;
- **radiometricamente**, isto é estão calibrados e corrigidos quanto aos efeitos da atmosfera.

Ilustração do efeito da ortorrectificação

Esta imagem com resolução espacial de 1m representa um terreno com variação de altitude da ordem de 60 m. Em colorido representa-se a localização correcta da rede viária. Dados não ortorrectificados não podem ser em geral combinados com outras fontes de informação.



Figura: Imagem georreferenciada mas não ortorrectificada.

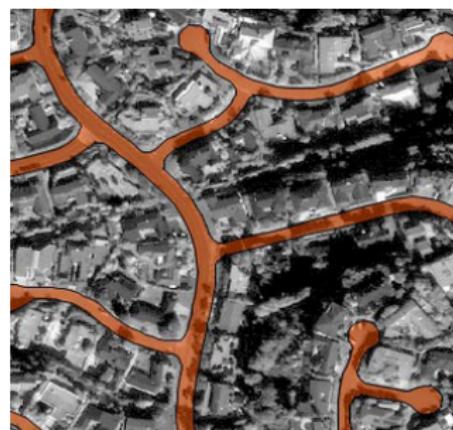
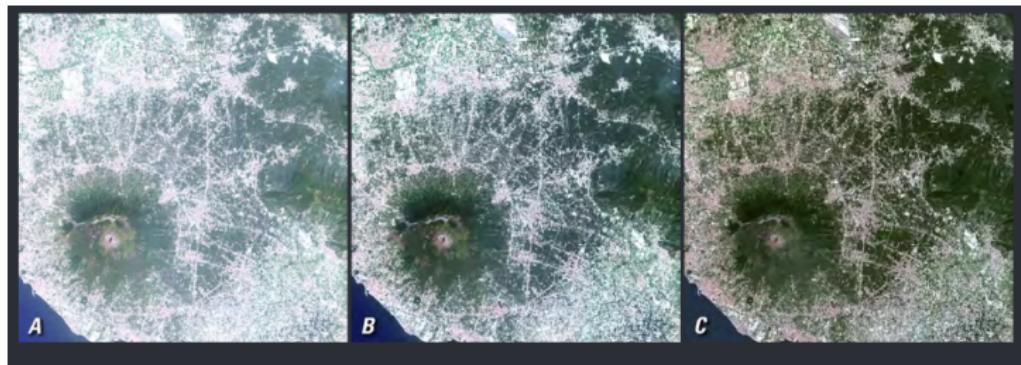


Figura: Imagem georreferenciada e ortorrectificada.

Ilustração de correcções radioméricas

A figura abaixo representa uma composição colorida em cor verdadeira para

- ① números digitais da imagem (esquerda);
- ② valores de reflectância "no topo da atmosfera" (centro) ;
- ③ valores de reflectância à superfície (direita), em que o efeito da atmosfera é corrigido.



These subsets of a Landsat scene of Mount Vesuvius near Naples, Italy, display A, natural color composite; B, top-of-atmosphere reflectance; and C, surface reflectance processing. (Source: Landsat 8 Operational Land Imager [OLI] Path 189 Row 32, acquired May 29, 2013.)

Fonte: <http://pubs.usgs.gov/fs/2015/3034/> (22/5/2015).

Acesso a dados de satélite: dados ESA

Os sites <https://ipsentinel.ipma.pt/> e <https://scihub.copernicus.eu/dhus/> (ESA) permitem aceder a dados Sentinel, incluindo, entre outros, os produtos seguintes.

- ① Sentinel 1 (RADAR),
- ② Sentinel 2 (óptico, 10-20 m, resolução temporal de 5 dias) e
- ③ Sentinel 3 (óptico, 300m e 500 m, dados diários);

A ESA disponibiliza uma aplicação `sen2cor` para correcção radiométrica das imagens Sentinel 2 (conversão para reflectância à superfície) e a aplicação **Sentinel Application Platform** (SNAP) que facilita a obtenção e manipulação de dados Sentinel.

Acesso a dados de satélite: dados NASA e USGS

O site <http://earthexplorer.usgs.gov/> (U.S.Geological Survey) permite o acesso a imagens de

- valores no sensor,
- reflectância no “topo da atmosfera”, e
- reflectância e outras grandezas físicas à superfície

em particular para os sensores

- Landsat 5, 7, 8 e 9;
- MODIS e VIIRS.

O site <https://search.earthdata.nasa.gov/> (EarthData) permite a acesso a dados da NASA, e distribui produtos semelhantes aos anteriores e outros dados e produtos desenvolvidos pela NASA.

Acesso a dados de satélite: interface IP-sentinel

No site basta indicar a região de interesse, as datas e o tipo de produto (satélite, sensor) para poder requisitar as imagens e fazer a transferência.

Por exemplo, a interface para obter dados Sentinel em <https://ipsentinel.ipma.pt/dhus/> é a seguinte:

The screenshot shows the IP-Sentinel web interface. On the left is a search form titled "Advanced Search" with sections for "Sort By" (set to "Descending"), "Sensing period" (From: 2023-01-01 To: 2023-01-02), "Ingestion period" (From: 2023-01-01 To: 2023-01-02), and three mission sections for Sentinel-1, Sentinel-2, and Sentinel-3. The Sentinel-2 section includes a "Cloud Cover % (e.g. [0 TO 5])" input field set to "[0 TO 5.4]". On the right is a map titled "Sentinel Satellites Imagery - Portugal" showing the Iberian Peninsula and surrounding regions. The map features several overlapping orange polygons representing satellite coverage areas, with a dashed red rectangle highlighting a specific region over Portugal.

Composições coloridas e índices de vegetação

Composição colorida de imagens multiespectrais

No Slide 132 foi descrito como se pode combinar três imagens distintas para criar uma composição colorida. Essa é uma forma comum de apresentar imagens multiespectrais pois permite **combinar a informação de várias bandas espectrais numa única imagem.**

- Diz-se que a composição é em **cor verdadeira** quando os canais da composição colorida correspondem às regiões espectrais da radiação visível respectivamente do vermelho, verde e azul. Exemplo: uma composição RGB=321 para o sensor Landsat 5 TM.
- Diz-se que a composição é em **falsa cor** caso contrário. As composições coloridas em falsa cor podem representar informação espectral do infravermelho. Exemplos: composições RGB=432 ou RGB=742 do sensor Landsat 5 TM.

Interpretação de uma composição colorida em falsa cor

Na composição RGB=(IVP,Vermelho,Verde), as cores obtidas tem em geral a seguinte interpretação: **vermelho-magenta** – vegetação vigorosa; **rosa** – vegetação mais esparsa ou com menor vigor vegetativo; **branco** – areia, solo descoberto, pedreiras; **azul escuro ou negro** – água ou estradas recentemente alcatroadas; **cinzento e azul metálico** – zonas urbanas; **castanho** – pastagens pobres e matos esparsos; **bege** – zonas de transição.



Figura: Composição colorida RGB=432 derivada de imagens de reflectância para as bandas do sensor Landsat 5 TM. Reguengos de Monsaraz, 20/05/2010.

Interpretação de uma composição colorida em falsa cor (cont)

A composição colorida é indicada pelo código $RGB=b_1b_2b_3$ em que b_1, b_2, b_3 indicam as bandas do sensor capta as imagens. As bandas “verde”, “vermelho” e IVP são respectivamente as bandas 2, 3 e 4 no sensor Landsat 5 TM e as bandas 3, 4, e 5 no sensor L8 OLI.

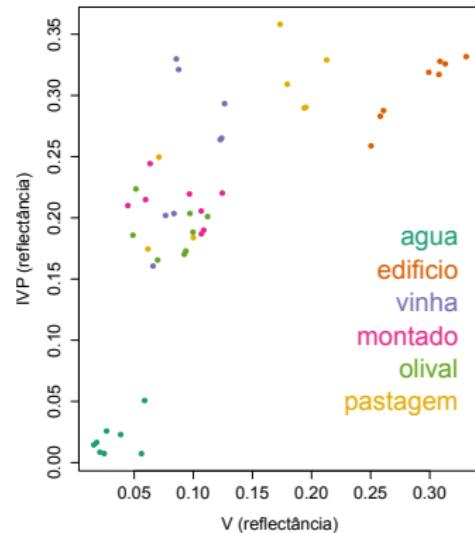
No exemplo anterior, a composição em falsa cor do exemplo anterior tem código $RGB=432$ (Landsat 5 TM), mas a composição colorida à direita, também em falsa cor e que usa as mesmas regiões espectrais, tem código $RGB=543$ (Landsat 8 OLI) pois as bandas do sensor têm “nomes” diferentes.



Representação dos pixels no espaço espectral das bandas do vermelho e do infravermelho próximo

A generalidade dos satélites de observação da terra adquirem informação espectral nas regiões do **vermelho** (aproximadamente $0.6 - 0.7 \mu\text{m}$) e na região do **infravermelho próximo** (aproximadamente $0.7 - 1.3 \mu\text{m}$) pois essas duas bandas fornecem muito informação sobre as materiais que ocorrem à superfície.

Representação das assinaturas espectrais de um conjunto de pixels correspondentes a diferentes ocupações do solo no plano V/IVP. Reguengos de Monsaraz, dados Landsat 8 CDR.

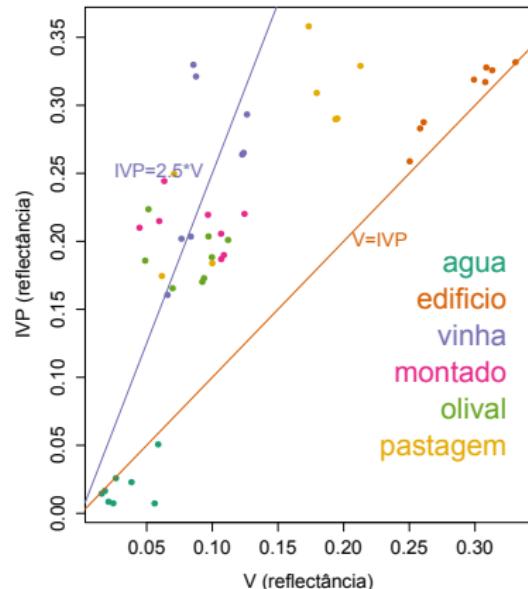


Índices de vegetação

No plano V/IVP podemos considerar as direcções radiais definidas pela relação $IVP = c \times V$ que correspondem por isso a razões constantes entre as reflectâncias nas regiões do vermelho e do infravermelho próximo. A razão V/IVP designa-se por **índice de vegetação**.

As rectas indicam valores constantes da razão V/IVP.
Note-se que raramente se verifica $IVP < V$.

Alguns autores sugerem que $IVP > 3 \times V$ corresponde a um “coberto vegetal denso”.



O índice NDVI (*normalized difference vegetation index*)

Este índice é de utilização generalizada e é definido por

$$\text{NDVI} = \frac{\text{IVP} - \text{V}}{\text{IVP} + \text{V}},$$

em que IVP e V são, respectivamente, as reflectâncias na região do infravermelho próximo e do vermelho.

Propriedades do índice NDVI:

- Se $\text{NDVI} = k$ então $\text{IVP} = \frac{-k-1}{k-1} \text{V}$ (a curva de nível é radial no plano V/IVP);
- NDVI toma sempre um valor entre -1 e 1 , (embora, na generalidade dos cobertos, os valores estejam entre 0 e 1);
- O valor de NDVI é muito instável quando $\text{V} + \text{IVP}$ está próximo de 0 (NDVI não está definido quando $\text{V} = \text{IVP} = 0$);
- Se se multiplicar V e IVP pela mesma constante, NDVI não é alterado.

Exercício

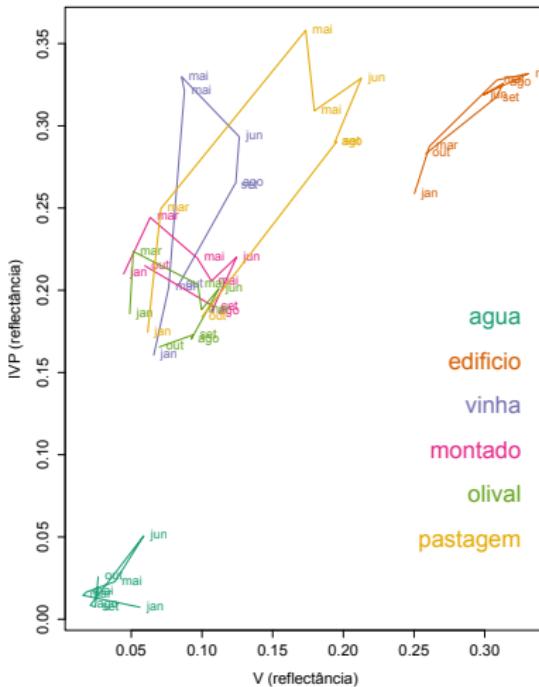
As assinaturas espectrais $a = (.15, .16, .20, .25, .23, .25)$, $b = (.08, .12, .10, .32, .25, .22)$ correspondem às reflectâncias nas bandas 1, 2, 3 (visível), 4 (IVP), e 5 e 7 (infravermelho médio) de dois pixels de uma imagem Landsat TM do final do verão de uma região onde existe pinhal, matos esparsos e culturas anuais de sequeiro.

- 1 Qual das assinaturas espectrais deverá corresponder a uma parcela de terreno coberta pela cultura anual de sequeiro e qual deverá corresponder a uma parcela de pinhal? Justifique.
- 2 Suponha que constrói uma composição colorida em falsa cor (RGB=742) da imagem. Qual das assinaturas espectrais terá uma cor mais verde na composição colorida? Justifique.
- 3 Determine os valores do índice de vegetação NDVI para as assinaturas a e b .

Evolução do padrão V/IVP ao longo do tempo

Os dados de satélite têm em geral uma boa resolução temporal. Por isso, é possível essas imagens para monitorizar a evolução de culturas ao longo do ciclo vegetativo.

Evolução das assinaturas espectrais ao longo de um ano para 6 pixels com diferente ocupação.



Exercício

O seguinte gráfico representa o sinal no Alentejo de três pixels de uma imagem de reflectância de superfície Landsat 8 multiespectral (ver Slide 251) ao longo de 2014. Os números representam o dia juliano (a partir de 1 de janeiro de 2014) em que foram obtidas as observações.

- 1 Indique uma justificação para o facto das observações dos vários pixels terem dispersões distintas;
- 2 Associe cada cor (i.e. cada pixel) a uma das seguintes ocupações do solo: edifício, pastagem e olival.

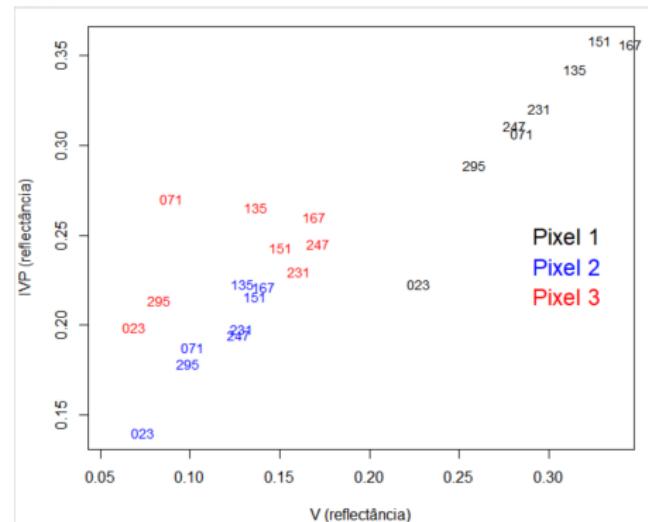
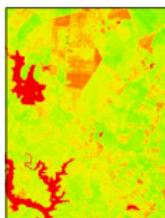
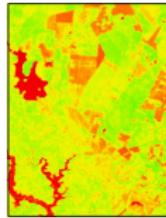


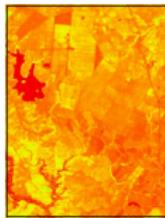
Ilustração da evolução do NDVI ao longo do ano: Reguengos de Monsaraz, de Fevereiro de 2010 a Março de 2011



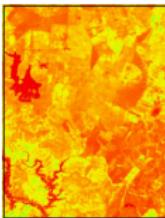
Fev



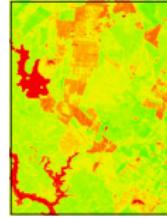
Maio



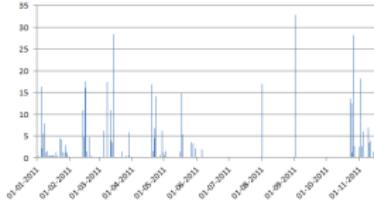
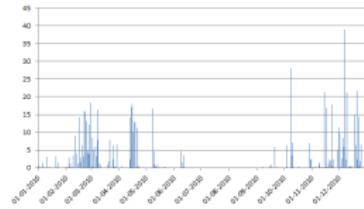
Jun



Nov



Março

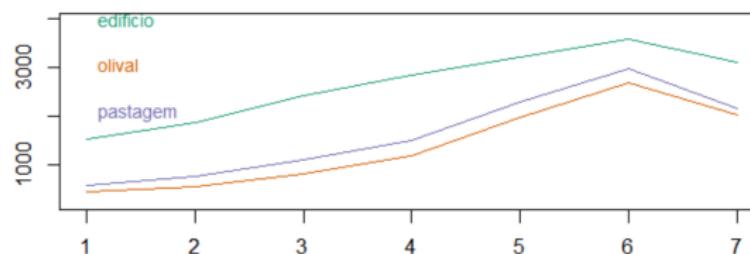
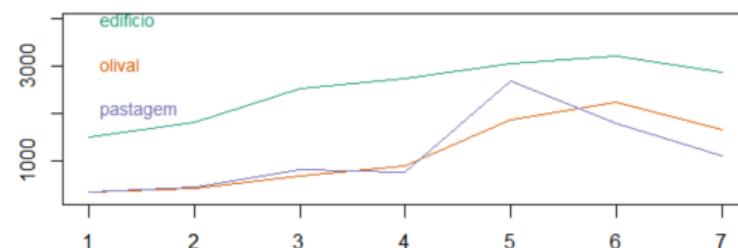


Vermelho representa NDVI=0 e verde indica valores de NDVI próximos de 1. O gráfico de barras indica a precipitação na região.

Exercício

O seguinte gráfico representa assinaturas espectrais em três locais no Alentejo numa imagem de reflectância de superfície Landsat 8 multiespectral com 7 bandas, em que a banda 5 corresponde ao IVP, (ver Slide 251).

Associe cada uma das imagens a uma data, sendo uma das imagens do início de março e outra de meados de agosto.



Exercício

O seguintes quadros descrevem algumas características de dois sensores a bordo do satélite Landsat 8 (Landsat Data Continuity Mission) lançado em 2013. A coluna “swath width” indica a largura da imagem captada ao longo do órbita do satélite.

Band	Wavelength (μm)	Bandwidth (μm)	Resolution (m)	Swath width (km)	Revisit time (days)
Band 1 Coastal/Aerosol (Other)	0.43 to 0.45		30	185	
Band 2 Blue (VIS)	0.45 to 0.51		30	185	
Band 3 Green (VIS)	0.53 to 0.59		30	185	
Band 4 Red (VIS)	0.64 to 0.67		30	185	
Band 5 Near Infrared (NIR)	0.85 to 0.88		30	185	
Band 6 SWIR 1 (SWIR)	1.57 to 1.65		30	185	
Band 7 SWIR 2 (SWIR)	2.11 to 2.29		30	185	
Band 8 Panchromatic (Other)	0.5 to 0.68		15	185	
Band 9 Cirrus (Other)	1.36 to 1.38		30	185	

Band	Wavelength (μm)	Bandwidth (μm)	Resolution (m)	Swath width (km)	Revisit time (days)
Band 10 TIR 1 (TIR)	10.6 to 11.19		100	185	
Band 11 TIR 2 (TIR)	11.5 to 12.51		100	185	

Exercício (cont.)

Considere uma região com ocupação do solo que pode ser eucaliptal ou então zonas de solo descoberto. Suponha que as assinaturas espectrais de dois pixels da imagem para as bandas 3, 4, 5, 6, e 7 (por essa ordem) são A=(0.16,0.20,0.26,0.26,0.27) e B=(0.13,0.12,0.33,0.25,0.22).

- Determine o índice de vegetação NDVI para cada um dos pixels;
- Indique o tipo de ocupação do solo que deverá corresponder a cada um deles.