

Projecto LANDAU - Metodologia geoestatística para a caracterização da incerteza espacial e actualização de cartas de ocupação do solo

Classificação uni-temporal da ocupação do solo através de dados de muito elevada resolução espacial - QUICKBIRD

Action 3.2 - Task 3

Relatório de execução

Joel Dinis, Pedro Rodrigues, Rita Nicolau



Setembro 2012

ÍNDICE

1	Introdução	5
2	Métodos	5
2.1	Dados e área de aplicação.....	5
2.2	Nomenclatura	6
2.3	Algoritmos de classificação ensaiados.....	7
2.3.1	Amostragem destinada ao treino dos algoritmos de classificação	7
2.4	Validação dos mapas produzidos.....	8
2.4.1	Amostragem destinada à validação	9
3	Resultados	9
3.1	Área de estudo E.....	10
3.2	Área de estudo J	13
3.3	Área de estudo A.....	16
3.4	Análise conjunta das áreas estudadas.....	19
4	Conclusões	21
5	Referências Bibliográficas	22

ANEXO I – Matrizes de erro / confusão associadas à classificação da ocupação do solo na área de estudo E, com base em dados QUICKBIRD.....23

ANEXO II – Matrizes de erro / confusão associadas à classificação da ocupação do solo na área de estudo J, com base em dados QUICKBIRD33

ANEXO III – Matrizes de erro / confusão associadas à classificação da ocupação do solo na área de estudo A, com base em dados QUICKBIRD43

1 Introdução

O presente relatório dá continuidade à descrição das actividades desenvolvidas no âmbito da *Action 3.2 - Task 3* do projecto LANDAU. As actividades em causa visaram a classificação automática da ocupação do solo em Portugal Continental a partir de dados de satélite de muito elevada resolução espacial (imagens QUICKBIRD), de acordo com a nomenclatura de ocupação do solo LANDAU.

Na produção de mapas de ocupação de solo a partir de dados QUICKBIRD testaram-se sete metodologias de classificação automática distintas, sendo objecto do presente estudo a avaliação do desempenho de tais metodologias.

2 Métodos

No presente ponto descrevem-se os dados, a área de aplicação, a nomenclatura e as metodologias empregues na classificação automática da ocupação do solo a partir de dados de satélite.

2.1 Dados e área de aplicação

A classificação automática da ocupação do solo foi realizada a partir de dados QUICKBIRD, que representam o nível detalhe máximo (pixeis de 2.4 m) facultado pelas três escalas de análise previstas pelo projecto LANDAU (média, elevada e muito elevada resolução espacial).

O trabalho foi aplicado a três áreas de estudo: E, J e A (vide Figura 1), que se consideraram representativas da diversidade paisagística do Continente (Nicolau *et al.*, 2011):

- Área de estudo E, situada na região Norte (concelhos de Monção, Valença, Paredes de Coura e Arcos de Valdevez);
- Área de estudo J, situada junto ao mar na região Centro (concelhos de Peniche e Lourinhã);
- Área de estudo A, localizada no Alentejo (concelhos de Cartaxo, Azambuja, Salvaterra de Magos e Benavente).

As imagens QUICKBIRD, empregues na classificação da ocupação do solo das três áreas citadas, são relativas ao período 2004-2005, designadamente aos meses de Setembro de 2004 para E, Agosto de 2004 para J e Setembro de 2005 para A.

A informação de referência utilizada para apoiar a recolha de amostras de treino dos

classificadores e de amostras de validação da cartografia produzida para as áreas de estudo mencionadas, foi a seguinte:

- Imagens aéreas orto-rectificadas, com uma resolução espacial de 50 cm e uma resolução espectral de 4 bandas (1995, 2005, 2007);
- Inventário Florestal – IF (2005);
- Cartografia CORINE Land Cover (2000, 2006).

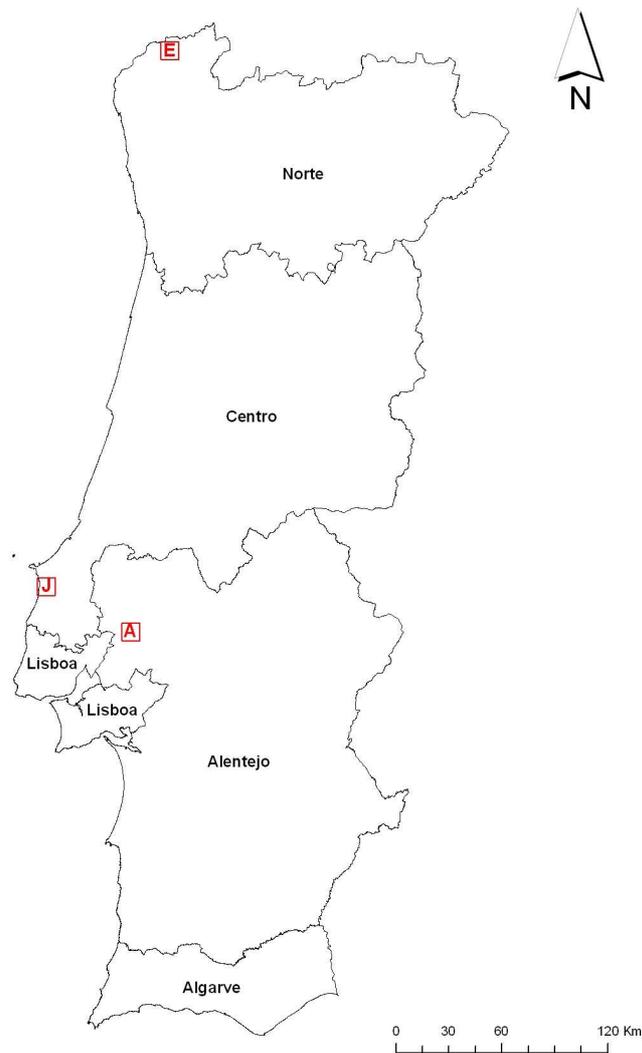


Figura 1 - Áreas de estudo LANDAU: E, J e A

2.2 Nomenclatura

A nomenclatura de ocupação do solo adoptada no presente trabalho foi anteriormente proposta no contexto do projecto (Dinis *et al.*, 2012a), sendo por isso designada de LANDAU. A sua proposta emergiu da necessidade de obtenção de uma nomenclatura compatibilizada com o sistema de classificação Land Cover Classification System (LCCS), que fosse funcional às múltiplas escalas de análise utilizadas no projecto.

Trata-se de uma nomenclatura que na sua versão mais desagregada inclui 15 classes de ocupação/uso do solo (vide Quadro 1).

Quadro 1 – Nomenclatura LANDAU

LANDAU – Nível 1	LANDAU – Nível 2
1 Artificial Areas (Territórios Artificializados)	1.1 Continuous Artificial Areas (Áreas Artificiais Contínuas)
	1.2 Discontinuous Artificial Areas (Áreas Artificiais Descontínuas)
2 Croplands (Áreas Agrícolas)	2.1 Irrigated Agriculture (Agricultura de Regadio)
	2.2 Non-irrigated Agriculture (Agricultura de Sequeiro)
	2.3 Rice Crops (Arrozais)
3 Natural and Semi-natural Vegetated Areas (Florestas e Meios Naturais e Semi-naturais)	3.1 Broadleaved Forest (Floresta de Folhosas)
	3.2 Coniferous Forest (Floresta de Resinosas)
	3.3 Mixed Forest (Floresta Mista)
	3.4 Grassland (Vegetação Herbácea)
	3.5 Shrubland (Matos)
	3.6 Baren to Sparsely Vegetated Areas (Vegetação Esparsa)
4 Bare Land (Solo Nu)	4 Bare Land (Solo Nu)
5 Burnt Areas (Áreas Ardidas)	5 Burnt Areas (Áreas Ardidas)
6 Wetlands (Zonas Húmidas)	6 Wetlands (Zonas Húmidas)
7 Water bodies (Corpos de Água)	7 Water bodies (Corpos de Água)

2.3 Algoritmos de classificação ensaiados

Os algoritmos ensaiados na classificação da ocupação do solo foram os seguintes: Maximum Likelihood (ML), Linear Discriminant Classifier (LDC), Diagonal Quadratic Discriminant Classifier (DQDC), K-Nearest Neighbours (KNN), Parzen Classifier (PARZEN), Classification and Regression Tree (CART) e Backpropagation Multilayer Perceptron (BMP).

Estes algoritmos foram apresentados num trabalho anterior (Dinis *et al.*, 2012b), que visou a classificação automática da ocupação do solo através da nomenclatura LANDAU, com recurso a dados MERIS.

2.3.1 Amostragem destinada ao treino dos algoritmos de classificação

Para o treino dos algoritmos de classificação recolheram-se amostras distintas por área de estudo. Na área de estudo E recolheu-se uma amostra formada por 8546 observações (pontos). A amostra obtida para a área de estudo J era constituída por 9129 pontos e a amostra da área de estudo A incluía 10980 pontos. As amostras referidas foram deterministicamente recolhidas sobre a Cartografia CORINE Land Cover (2006), na sequência da sua reclassificação em classes de nível 2 da nomenclatura LANDAU.

O Quadro 2 disponibiliza informação sobre o número de pontos amostrados por área

de estudo, segundo classes de nível 2 da nomenclatura LANDAU.

Quadro 2 – Nº de pontos de treino amostrados por área de estudo, segundo classes da nomenclatura LANDAU

Nomenclatura LANDAU nível 2	Área E: Nº de pontos amostrados	Área J: Nº de pontos amostrados	Área A: Nº de pontos amostrados
1.1 Áreas Artificiais Contínuas	-	-	-
1.2 Áreas Artificiais Descontínuas	1372	632	450
2.1 Agricultura de Regadio	380	380	1269
2.2 Agricultura de Sequeiro	1360	3978	4475
2.3 Arrozais	-	-	750
3.1 Floresta de Folhosas	709	702	740
3.2 Floresta de Resinosas	980	515	971
3.3 Floresta Mista	-	-	-
3.4 Vegetação Herbácea	1142	1104	461
3.5 Matos	1058	350	1014
3.6 Vegetação Esparsa	350	192	-
4 Solo Nu	350	260	260
5 Áreas Ardidas	550	-	-
6 Zonas Húmidas	-	-	190
7 Corpos de Água	295	1016	400
TOTAL	8546	9129	10980

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

2.4 Validação dos mapas produzidos

Para comparação dos mapas produzidos e análise da sua utilidade para aplicações específicas importa conhecer a exactidão na classificação da ocupação do solo produzida por cada algoritmo, ou seja a exactidão temática de cada mapa. Neste sentido, na presente secção descreve-se o método de amostragem utilizado para validar os mapas produzidos, bem como as medidas de exactidão temática, derivadas da matriz de erro/confusão, que retratam o desempenho na classificação da ocupação do solo de cada algoritmo testado.

Sendo inviável avaliar o erro cometido por cada algoritmo de classificação em todas as posições do domínio espacial, a validação da classificação da ocupação do solo produzida por cada algoritmo é usualmente realizada a partir de uma amostra representativa das classes de ocupação/uso do solo presentes no domínio espacial. Esta amostra deverá ser preferencialmente recolhida de modo aleatório por forma a possibilitar generalizações sobre a área de interesse.

2.4.1 Amostragem destinada à validação

Para validação dos mapas de ocupação do solo foram recolhidas amostras formadas por 561 observações (pontos) na área E, 501 observações na área J e 559 observações na área A. Tratou-se de uma amostragem aleatória, estratificada pelas classes de nível 2 da nomenclatura LANDAU que à escala de análise adoptada tinham representação em cada área de estudo. A dimensão da amostra foi determinada considerando a significância estatística e o esforço operacional necessário para a sua construção.

Através da consulta da informação de referência listada em 2.1 foi identificada a classe de espaço que com maior probabilidade descreve cada ponto amostrado, bem como uma segunda classe de espaço que poderá igualmente descrever a ocupação/uso do solo no mesmo ponto, destinando-se esta última a situações em que há ambiguidade na atribuição da primeira classe de espaço.

A matriz de erro/confusão (ou tabela de contingência) foi obtida por confronto da classificação de cada observação da amostra de validação com a classificação atribuída pelo algoritmo para a mesma posição do domínio espacial. Admitiu-se a existência de concordância na classificação quando uma das duas classes de espaço atribuídas à observação amostrada, a partir da consulta de informação de referência, coincidia com a classe de espaço indicada pelo algoritmo de classificação.

As medidas de avaliação do desempenho na classificação estimadas a partir da matriz de erro/confusão, foram a exactidão global (\hat{P}), a exactidão do produtor (\hat{P}_p) e a exactidão do utilizador (\hat{P}_u). O modo de obtenção destas medidas já foi explicitado num trabalho de classificação análogo ao presente, baseado em dados MERIS (Dinis *et al.*, 2012b).

3 Resultados

No presente capítulo descrevem-se os principais resultados decorrentes da classificação da ocupação do solo de cada área de estudo, através dos sete algoritmos ensaiados. Nos Anexos I, II e III disponibilizam-se as matrizes de erro/confusão associadas aos classificadores empregues por área de estudo (E, J e A, respectivamente). Nos Quadros 3, 4 e 5 compararam-se as medidas de exactidão auferidas na classificação das classes da nomenclatura LANDAU (nível 2), representadas nas áreas de estudo E, J e A (respectivamente).

3.1 Área de estudo E

Dos resultados obtidos para a área de estudo E (vide Quadro 3) constata-se que o algoritmo com melhor desempenho global na classificação do uso do solo foi o ML ($\hat{P} = 82\%$). O desempenho global dos algoritmos LDC, KNN, PARZEN e CART foi similar ($\hat{P} = 71\%$), ainda que inferior ao do ML. O BMP foi o algoritmo que denotou pior desempenho global ($\hat{P} = 63\%$).

Uma análise por classes de nível 2 da nomenclatura LANDAU possibilita destacar que:

- A escala de análise facultada pela imagem QUICKBIRD impossibilitou o reconhecimento das seguintes classes de espaço: Áreas Artificiais Contínuas (1.1), Arrozais (2.3), Floresta Mista (3.3) e Zonas Húmidas (6).
- As Áreas Artificiais Descontínuas (1.2) foram melhor classificadas do ponto de vista do produtor pelo algoritmo BMP ($\hat{P}_p = 95\%$), embora através dos algoritmos KNN e CART também se tenham obtido exactidões aceitáveis na sua classificação ($\hat{P}_p = 82\%$ para ambos). Do ponto de vista do utilizador, a exactidão na classificação destas áreas foi maior quando se utilizaram os algoritmos ML e LDC ($\hat{P}_u = 77\%$ para ambos).
- Uma exactidão do produtor máxima ($\hat{P}_p = 100\%$) foi alcançada na classificação de áreas de Agricultura de Regadio (2.1) através dos algoritmos ML, LDC e BMP. Estas áreas foram igualmente classificadas com boa exactidão do produtor pelos algoritmos DQDC e CART ($\hat{P}_p = 98\%$ para ambos). A maior exactidão do utilizador obtida na classificação destas áreas ($\hat{P}_u = 92\%$) foi proporcionada pelo algoritmo BMP.
- Na classificação da Agricultura de Sequeiro (2.2) o ML foi o algoritmo que permitiu classificar este tipo de agricultura com maior exactidão do produtor ($\hat{P}_p = 79\%$). A maior exactidão do utilizador obtida na classificação destas áreas resultou da aplicação do algoritmo CART ($\hat{P}_u = 66\%$), cujo desempenho foi apenas ligeiramente superior ao dos algoritmos LDC e ML ($\hat{P}_u = 65\%$ para ambos).
- Na classificação da Floresta de Folhosas (3.1) o algoritmo ML foi o que revelou

melhor desempenho, apresentando não só a maior exactidão do produtor ($\hat{P}_p=84\%$) mas também a maior exactidão do utilizador ($\hat{P}_u=94\%$).

- Na classificação de Florestas de Resinosas (3.2) as mais elevadas exactidões do produtor e do utilizador foram igualmente obtidas pelo algoritmo ML ($\hat{P}_p=78\%$; $\hat{P}_u=75\%$).
- O algoritmo ML foi, de entre os testados, o que permitiu classificar a Vegetação Herbácea (3.4) com maior exactidão do produtor ($\hat{P}_p=77\%$). A maior exactidão do utilizador foi alcançada pela aplicação dos algoritmos KNN e CART ($\hat{P}_u=78\%$ para ambos).
- O algoritmo ML destacou-se mais uma vez na classificação dos Matos (3.5), denotando simultaneamente a maior exactidão do produtor ($\hat{P}_p=77\%$) e a maior exactidão do utilizador ($\hat{P}_u=87\%$). Na classificação desta classe de espaço evidenciaram-se ainda os algoritmos BMP e LDC, facultando uma exactidão do produtor (no caso do BMP) e uma exactidão do utilizador (no caso do LDC) da ordem de grandeza das obtidas através do ML.
- No que concerne à classificação da Vegetação Esparsa (3.6), a maior exactidão do produtor ($\hat{P}_p=83\%$) foi obtida através do algoritmo ML e a maior exactidão do utilizador ($\hat{P}_u=85\%$) deveu-se à aplicação do algoritmo CART.
- Na classificação do Solo Nu (4) o ML foi o algoritmo que permitiu obter melhores resultados: uma exactidão do produtor (\hat{P}_p) de 85% e uma exactidão do utilizador máxima ($\hat{P}_u=100\%$), igualmente proporcionada pelo algoritmo LDC.
- A máxima exactidão do produtor ($\hat{P}_p=100\%$) obtida na classificação das Áreas Ardidias (5) resultou da aplicação do algoritmo ML. Este algoritmo, a par do LDC, também facultou a maior exactidão do utilizador ($\hat{P}_u=86\%$).
- Embora a maioria dos algoritmos testados tenha revelado boa performance na classificação de Corpos de Água (7), o melhor desempenho coube ao algoritmo LDC ($\hat{P}_p=\hat{P}_u=100\%$). Na classificação desta classe destacaram-se igualmente com máxima exactidão do utilizador ($\hat{P}_u=100\%$) os algoritmos ML e CART.

Quadro 3 – Medidas de exactidão temática dos algoritmos de classificação testados – Área de estudo E

	ML		LDC		DQDC		KNN		PARZEN		CART		BMP	
	\hat{P}_p	\hat{P}_u												
\hat{P}_p = Exactidão do Produtor														
\hat{P}_u = Exactidão do Utilizador														
1.1 Áreas Artificiais Contínuas	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.2 Áreas Artificiais Descontínuas	69%	77%	79%	77%	18%	37%	82%	36%	79%	37%	82%	34%	95%	30%
2.1 Agricultura de Regadio	100%	88%	100%	88%	98%	83%	96%	84%	93%	84%	98%	90%	100%	92%
2.2 Agricultura de Sequeiro	79%	65%	52%	65%	72%	52%	66%	64%	67%	62%	65%	66%	37%	39%
2.3 Arrozais	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3.1 Floresta de Folhosas	84%	94%	79%	94%	68%	78%	74%	86%	74%	86%	81%	82%	83%	93%
3.2 Floresta de Resinosas	78%	75%	49%	75%	39%	53%	66%	63%	66%	64%	56%	65%	39%	66%
3.3 Floresta Mista	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3.4 Vegetação Herbácea	77%	74%	61%	74%	66%	59%	72%	78%	70%	77%	66%	78%	36%	39%
3.5 Matos	77%	87%	57%	87%	74%	67%	62%	82%	63%	83%	63%	74%	77%	70%
3.6 Vegetação Esparsa	83%	81%	81%	81%	58%	51%	41%	82%	43%	83%	66%	85%	0%	0%
4 Solo Nu	85%	100%	83%	100%	81%	93%	76%	95%	77%	91%	64%	94%	67%	97%
5 Áreas Áridas	100%	86%	89%	86%	79%	79%	74%	70%	79%	71%	84%	84%	89%	74%
6 Zonas Húmidas	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7 Corpos de Água	93%	100%	100%	100%	93%	90%	90%	96%	90%	96%	93%	100%	90%	93%
\hat{P} = Exactidão Global	82%		71%		66%		71%		71%		71%		63%	

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

3.2 Área de estudo J

Da análise dos resultados obtidos para a área de estudo J (vide Quadro 4) verifica-se que os algoritmos que tiveram melhor desempenho global na classificação do uso do solo foram o ML e o CART ($\hat{P} = 85\%$). O algoritmo BMP foi o que revelou pior desempenho global ($\hat{P} = 67\%$) neste contexto.

Uma análise da classificação por classes de nível 2 da nomenclatura LANDAU permite realçar que:

- O detalhe proporcionado pela escala da QUICKBIRD não possibilitou o reconhecimento das seguintes classes de espaço: Áreas Artificiais Contínuas (1.1), Arrozais (2.3), Floresta Mista (3.3) e Zonas Húmidas (6). Também não se identificaram Áreas Áridas (5), pelo que se crê que eram inexistentes à data da imagem utilizada (Agosto 2004).
- As Áreas Artificiais Descontínuas (1.2) foram melhor classificadas pelos algoritmos ML e CART, quer do ponto de vista do produtor ($\hat{P}_p = 93\%$), quer do ponto de vista do utilizador ($\hat{P}_u = 79\%$).
- Uma exactidão do produtor máxima ($\hat{P}_p = 100\%$) foi alcançada na classificação da Agricultura de Regadio (2.1) através dos algoritmos ML, CART e DQDC. A maior exactidão do utilizador obtida na classificação desta agricultura ($\hat{P}_u = 96\%$) foi proporcionada pelo algoritmo DQDC, com um valor que apenas foi ligeiramente superior ao produzido pelos algoritmos ML e CART ($\hat{P}_u = 95\%$).
- O BMP foi o algoritmo que permitiu classificar a Agricultura de Sequeiro (2.2) com maior exactidão do produtor ($\hat{P}_p = 92\%$). A maior exactidão do utilizador obtida na classificação destas áreas resultou da aplicação do algoritmo PARZEN ($\hat{P}_u = 88\%$).
- Na classificação da Floresta de Folhosas (3.1) o algoritmo BMP foi responsável pela obtenção de uma exactidão do produtor máxima ($\hat{P}_p = 100\%$), não obstante o bom desempenho dos algoritmos ML e CART ($\hat{P}_p = 98\%$). A maior exactidão do utilizador ($\hat{P}_u = 98\%$) na classificação deste tipo de floresta foi obtida pelos algoritmos ML, CART e PARZEN.

- Na classificação de Florestas de Resinosas (3.2) os algoritmos ML, CART e PARZEN denotaram a mais elevada exactidão do produtor ($\hat{P}_p=100\%$). A maior exactidão do utilizador obtida na classificação desta classe ($\hat{P}_u=94\%$) foi gerada pelos algoritmos ML e CART.
- Os algoritmos LDC e BMP foram, de entre os testados, os que permitiram classificar a Vegetação Herbácea (3.4) com maior exactidão do produtor ($\hat{P}_p=91\%$). A maior exactidão do utilizador foi alcançada pela aplicação dos algoritmos ML e CART ($\hat{P}_u=67\%$).
- Embora reduzida, a maior exactidão do produtor ($\hat{P}_p=47\%$) obtida na classificação dos Matos (3.5) deveu-se à aplicação dos algoritmos ML e CART. O algoritmo DQDC destacou-se na obtenção da maior exactidão do utilizador ($\hat{P}_u=78\%$).
- Na classificação da Vegetação Esparsa (3.6), os algoritmos ML e CART revelaram a maior exactidão do produtor ($\hat{P}_p=92\%$). A maior exactidão do utilizador ($\hat{P}_u=89\%$) foi disponibilizada pelo algoritmo DQDC.
- Na classificação do Solo Nu (4), o algoritmo LDC evidenciou-se pela apresentação da mais elevada exactidão do produtor ($\hat{P}_p=89\%$). A máxima exactidão do utilizador ($\hat{P}_u=100\%$) foi alcançada pelos algoritmos DQDC, KNN e PARZEN.
- Todos os algoritmos testados revelaram boa performance na classificação de Corpos de Água (7), mas somente cinco (ML, DQDC, KNN, PARZEN e CART) produziram em simultâneo máximas exactidões do produtor e do utilizador ($\hat{P}_p=\hat{P}_u=100\%$).

Quadro 4 – Medidas de exactidão temática dos algoritmos de classificação testados – Área de estudo J

	ML		LDC		DQDC		KNN		PARZEN		CART		BMP	
	\hat{P}_p	\hat{P}_u												
\hat{P}_p = Exactidão do Produtor \hat{P}_u = Exactidão do Utilizador														
1.1 Áreas Artificiais Contínuas	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.2 Áreas Artificiais Descontínuas	93%	79%	43%	70%	41%	45%	86%	56%	86%	64%	93%	79%	16%	23%
2.1 Agricultura de Regadio	100%	95%	86%	90%	100%	96%	91%	95%	95%	95%	100%	95%	91%	95%
2.2 Agricultura de Sequeiro	80%	80%	76%	79%	76%	70%	83%	77%	82%	88%	80%	80%	92%	54%
2.3 Arrozais	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3.1 Floresta de Folhosas	98%	89%	94%	87%	94%	94%	92%	96%	92%	98%	98%	98%	100%	85%
3.2 Floresta de Resinosas	100%	94%	72%	82%	94%	88%	94%	87%	100%	89%	100%	94%	85%	83%
3.3 Floresta Mista	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3.4 Vegetação Herbácea	79%	67%	91%	50%	85%	54%	74%	58%	88%	62%	79%	67%	91%	56%
3.5 Matos	47%	76%	12%	25%	43%	78%	27%	56%	38%	68%	47%	76%	0%	0%
3.6 Vegetação Esparsa	92%	85%	84%	84%	64%	89%	76%	88%	82%	85%	92%	85%	0%	0%
4 Solo Nu	78%	91%	89%	96%	78%	100%	48%	100%	59%	100%	78%	91%	78%	95%
5 Áreas Áridas	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6 Zonas Húmidas	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7 Corpos de Água	100%	100%	100%	94%	100%	100%	100%	100%	100%	100%	100%	100%	100%	98%
\hat{P} = Exactidão Global	85%		74%		77%		78%		83%		85%		67%	

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

3.3 Área de estudo A

Os resultados relativos à área de estudo A (vide Quadro 5) mostram que o algoritmo com melhor desempenho global na classificação do uso do solo foi o ML ($\hat{P} = 86\%$), tendo o algoritmo BMP sido o que revelou pior desempenho global ($\hat{P} = 68\%$).

Uma análise da classificação obtida por classes de nível 2 da nomenclatura LANDAU indica que:

- O detalhe proporcionado pela escala da QUICKBIRD não possibilitou o reconhecimento das seguintes classes de espaço: Áreas Artificiais Contínuas (1.1), Floresta Mista (3.3), Vegetação Esparsa (3.6) e Áreas Áridas (5).
- O melhor desempenho na classificação das Áreas Artificiais Descontínuas (1.2) foi assegurado pelos seguintes algoritmos: CART, do ponto de vista do produtor ($\hat{P}_p = 65\%$), e BMP, do ponto de vista do utilizador ($\hat{P}_u = 88\%$).
- Na classificação da Agricultura de Regadio (2.1) qualquer um dos algoritmos testados produziu uma exactidão do produtor máxima ($\hat{P}_p = 100\%$), mas a mais elevada exactidão do utilizador obtida na classificação destas áreas ($\hat{P}_u = 96\%$) apenas foi assegurada pelos algoritmos ML, LDC e BMP.
- Os algoritmos BMP e KNN foram os que permitiram classificar a Agricultura de Sequeiro (2.2) com maior exactidão do produtor ($\hat{P}_p = 98\%$). A maior exactidão do utilizador obtida na classificação destas áreas resultou da aplicação do algoritmo LDC ($\hat{P}_u = 64\%$).
- Os algoritmos ML, LDC e DQDC foram, de entre os testados, os que permitiram classificar os Arrozaes (2.3) com maior exactidão do produtor ($\hat{P}_p = 92\%$). A maior exactidão do utilizador foi alcançada pela aplicação do algoritmo ML ($\hat{P}_u = 98\%$).
- Na classificação da Floresta de Folhosas (3.1) os algoritmos ML e BMP possibilitaram a obtenção de uma exactidão do produtor máxima ($\hat{P}_p = 100\%$), não obstante o bom desempenho dos algoritmos LDC e DQDC ($\hat{P}_p = 98\%$). A maior exactidão do utilizador ($\hat{P}_u = 94\%$) na classificação deste tipo de floresta foi obtida pelo algoritmo ML.

- Na classificação de Florestas de Resinosas (3.2), destacou-se o algoritmo ML pela apresentação da mais elevada exactidão do produtor ($\hat{P}_p=100\%$) e da maior exactidão do utilizador ($\hat{P}_u=89\%$).
- O algoritmo ML foi também, de entre os testados, o que obteve maior exactidão do produtor ($\hat{P}_p=82\%$) na classificação da Vegetação Herbácea (3.4). A máxima exactidão do utilizador ($\hat{P}_u=100\%$) na classificação deste tipo de vegetação resultou da aplicação do algoritmo BMP.
- O algoritmo ML evidenciou-se igualmente na classificação dos Matos (3.5), onde foi responsável pelas maiores exactidões do produtor ($\hat{P}_p=98\%$) e do utilizador ($\hat{P}_u=91\%$).
- Na classificação do Solo Nu (4), a mais elevada exactidão do produtor ($\hat{P}_p=89\%$) foi gerada pelo algoritmo ML e a maior exactidão do utilizador ($\hat{P}_u=90\%$) foi alcançada pelo algoritmo BMP.
- O algoritmo LDC foi, de entre os testados, o que permitiu classificar as Zonas Húmidas (6) com maior exactidão do produtor ($\hat{P}_p=73\%$). A mais elevada exactidão do utilizador ($\hat{P}_u=100\%$) obtida na classificação destas zonas, foi no entanto disponibilizada por cinco algoritmos distintos: ML, KNN, PARZEN e BMP.
- A mais elevada exactidão do produtor ($\hat{P}_p=91\%$) obtida na classificação de Corpos de Água (7) foi proporcionada pelo algoritmo BMP. A máxima exactidão do utilizador ($\hat{P}_u=100\%$) resultou da aplicação do algoritmo ML.

Quadro 5 – Medidas de exactidão temática dos algoritmos de classificação testados – Área de estudo A

	ML		LDC		DQDC		KNN		PARZEN		CART		BMP	
	\hat{P}_p	\hat{P}_u												
\hat{P}_p = Exactidão do Produtor \hat{P}_u = Exactidão do Utilizador														
1.1 Áreas Artificiais Contínuas	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.2 Áreas Artificiais Descontínuas	64%	82%	32%	67%	33%	46%	62%	70%	44%	65%	65%	69%	61%	88%
2.1 Agricultura de Regadio	100%	96%	100%	96%	100%	94%	100%	93%	100%	93%	100%	94%	100%	96%
2.2 Agricultura de Sequeiro	93%	62%	96%	64%	90%	63%	98%	49%	96%	49%	93%	51%	98%	35%
2.3 Arrozais	92%	98%	92%	88%	92%	94%	68%	89%	68%	89%	74%	95%	52%	81%
3.1 Floresta de Folhosas	100%	94%	98%	89%	98%	74%	96%	75%	96%	75%	88%	76%	100%	91%
3.2 Floresta de Resinosas	100%	89%	96%	82%	68%	69%	69%	74%	69%	74%	78%	70%	92%	74%
3.3 Floresta Mista	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3.4 Vegetação Herbácea	82%	76%	76%	81%	77%	66%	70%	88%	70%	85%	65%	77%	10%	100%
3.5 Matos	98%	91%	96%	82%	59%	75%	86%	78%	84%	78%	84%	79%	96%	71%
3.6 Vegetação Esparsa	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4 Solo Nu	61%	88%	50%	77%	48%	82%	47%	82%	51%	64%	53%	87%	39%	90%
5 Áreas Áridas	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6 Zonas Húmidas	71%	100%	73%	83%	59%	67%	57%	100%	57%	100%	48%	95%	2%	100%
7 Corpos de Água	85%	100%	89%	93%	87%	98%	87%	95%	87%	95%	90%	96%	91%	86%
\hat{P} = Exactidão Global	86%		82%		74%		77%		75%		77%		68%	

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

3.4 Análise conjunta das áreas estudadas

Uma avaliação da execução dos algoritmos testados nas três áreas de estudo (E, J e A), possibilita evidenciar o algoritmo ML pelo melhor desempenho global na classificação do uso do solo em qualquer das áreas estudadas. De facto, a exactidão global proporcionada pelo ML no conjunto das três zonas estudadas, oscilou entre 82% e 86%, apresentando um valor médio de 84%, valor que se destaca da exactidão global média de 78%, obtida pelo algoritmo, CART, posicionado em segundo lugar neste contexto.

Procede-se seguidamente à análise da classificação, segundo classes da nomenclatura LANDAU, realizada pelo algoritmo ML para o conjunto das três áreas estudadas. A Figura 2 descreve os valores médios de exactidão produzidos pelo ML para cada classe da nomenclatura citada.

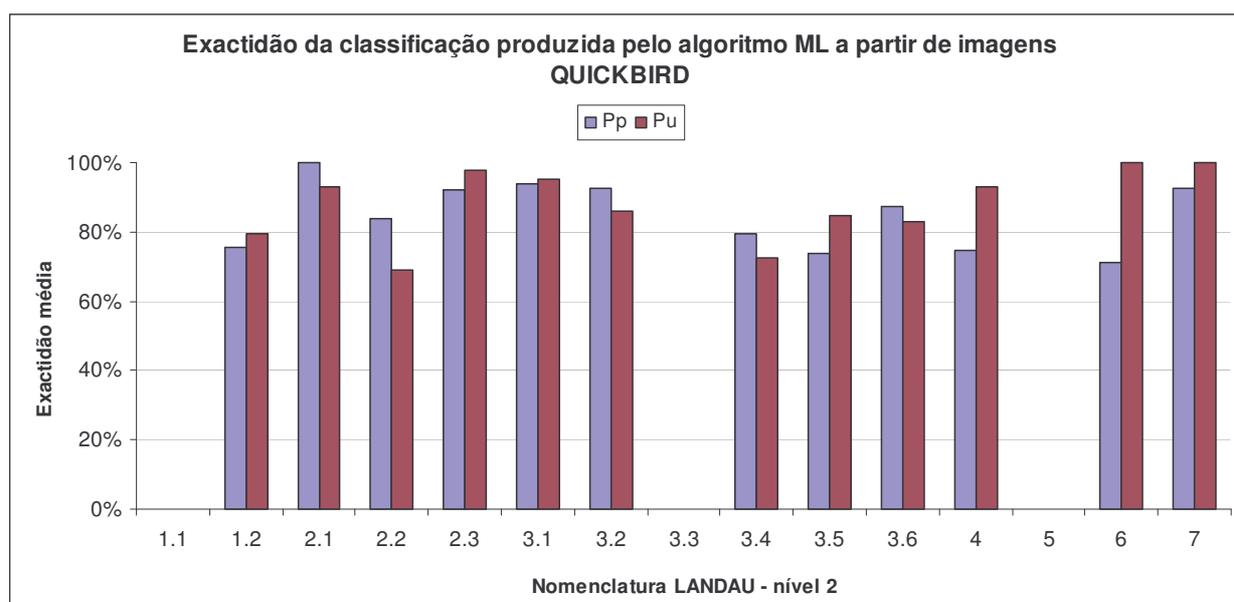


Figura 2 – Valores médios de exactidão do produtor (Pp) e do utilizador (Pu), segundo classes da nomenclatura LANDAU, obtidos pelo algoritmo ML para o conjunto das três áreas estudadas

Como já mencionado, o elevado detalhe proporcionado pelas imagens QUICKBIRD não permitiu o reconhecimento de Áreas Urbanas Contínuas (1.1) em qualquer das zonas estudadas. Tal deve-se ao facto do tecido urbano não ser suficientemente denso nas área geográficas consideradas.

Na classificação das Áreas Urbanas Descontínuas (1.2), O algoritmo ML obteve exactidões variáveis entre 64% a 93% do ponto de vista do produtor, e entre 77% a 82% do ponto de vista do utilizador, atingindo exactidões médias de 75% no contexto

do produtor e de 79% no contexto do utilizador.

No que concerne às Áreas Agrícolas (2), verificou-se que a Agricultura de Regadio (2.1) foi classificada pelo ML com maior exactidão do que a Agricultura de Sequeiro (2.2). Na classificação do Sequeiro obteve-se uma exactidão média do produtor de 84%, enquanto que na classificação do Regadio aquela exactidão foi de 100%. Já a exactidão do média do utilizador foi de 69% na classificação do Sequeiro e de 93% na classificação do Regadio.

Na classificação dos Arrozais (2.3), apenas identificados na área A, a exactidão do utilizador obtida pelo ML foi de 98% e a exactidão do produtor foi de 92%.

As exactidões alcançadas pelo algoritmo ML na classificação das Florestas e Meios Naturais e Semi-naturais (3), denotaram grande variabilidade, oscilando entre 47% e 100% no caso da exactidão do produtor, e entre 67% e 98% no caso da exactidão do utilizador. Neste grande grupo, o ML denotou pior desempenho na classificação da Vegetação Herbácea (3.4) e dos Matos (3.5), particularmente na área J. As exactidões médias do produtor e do utilizador apresentadas pelo ML foram, respectivamente, 74% e 85% para a Vegetação Herbácea (3.4) e 79% e 72% para os Matos (3.5).

O elevado detalhe proporcionado pela escala de análise adoptada impossibilitou o reconhecimento da Floresta Mista (3.3) nas áreas estudadas, facto que provavelmente originou uma avaliação por excesso das Florestas de Folhosas (3.1) e de Resinosas (3.2). Na classificação das primeiras (3.1), o ML obteve exactidões médias de 94%, no contexto do produtor, e de 95% no contexto do utilizador. Na classificação das segundas (3.2), o ML obteve uma exactidão média do produtor de 93% e uma exactidão média do utilizador de 86%.

No que diz respeito à classificação da Vegetação Esparsa (3.6) que apenas foi identificada em E e J, o ML alcançou uma exactidão média do produtor igual a 88% e uma exactidão média do utilizador igual a 83%.

A aplicação do ML na classificação de Solo Nu (4) originou uma exactidão média do produtor igual a 75% e uma exactidão média do utilizador igual a 93%.

Na classificação de Áreas Áridas (5), apenas identificadas em E, a exactidão do utilizador obtida pelo ML foi de 86% e a exactidão do produtor foi de 100%.

Na classificação das Zonas Húmidas (6), que à resolução espacial adoptada somente tinham representação na área A, a exactidão do utilizador obtida pelo ML foi de 100% e

a exactidão do produtor foi de 71%.

Na classificação de Corpos de Água (7), a exactidão do utilizador obtida através do ML foi de 100% em todas as áreas estudadas e a exactidão média do produtor foi de 93%.

4 Conclusões

Os resultados relativos à classificação do uso do solo em três zonas do Continente (E, J e A), revelam que o algoritmo ML (Maximum Likelihood) se destacou pelo melhor desempenho global. Deste modo, o presente estudo comprova que, à escala das imagens QUICKBIRD, o ML denota globalmente melhores resultados do que os restantes algoritmos de classificação testados. Estes resultados são concordantes com os resultados obtidos pelo projecto STALOG, que envolveu a comparação de múltiplos classificadores em diferentes aplicações (Mitchie *et al.*, 1994; citado por: Hastie *et al.*, 2009), em que o ML figurou consistentemente entre os três algoritmos com melhor desempenho na classificação de diversos conjuntos de dados.

A superioridade do ML na classificação de espécies florestais a partir de imagens de muito elevada resolução (IKONOS), foi também evidenciada por Plantier (2006) num estudo aplicado a três áreas distintas de Portugal Continental. Segundo esta autora, as espécies florestais mais comuns em Portugal Continental apresentam respostas espectrais semelhantes em imagens de muito elevada resolução, o que ocasiona alguma dificuldade na sua discriminação com procedimentos automáticos.

A comparação dos resultados do presente estudo com os obtidos à escala de uma imagem MERIS (Dinis *et al.*, 2012a), sugere que o aumento da resolução espacial proporcionou um melhor desempenho na classificação da ocupação do solo realizada pela maioria dos algoritmos testados (ML, DQDC, KNN, PARZEN e CART). Ao invés da maioria dos algoritmos, o LDC e o BMP evidenciaram um decréscimo do desempenho na classificação com o aumento da resolução espacial.

No que diz respeito às áreas estudadas, a área E, sendo a mais povoada por floresta, meios naturais e semi-naturais e a menos urbanizada das três analisadas, foi aquela em que todos os algoritmos testados cometeram maior erro na classificação do uso do solo, provavelmente devido à maior diversidade de classes, à elevada fragmentação da paisagem e à proximidade espectral que as classes do grupo florestal denotam.

Por outro lado, a área de estudo A é, de entre as três consideradas, a mais intervencionada pelo homem, não só devido à apresentação de maior proporção de

ocupação agrícola e agro-florestal, mas também devido à presença de territórios artificializados. Trata-se igualmente da área que apresenta padrões de ocupação do solo de maiores dimensões e homogeneidade, facto que poderá fundamentar a obtenção de menor erro na classificação do uso do solo, por parte da maioria dos algoritmos ensaiados. Adicionalmente, a composição da paisagem apresenta classes passíveis de menor confusão espectral, tais como arrozais, sequeiro, urbano descontínuo, do que as restantes áreas de estudo, facto que contribui efectivamente para uma melhoria dos resultados.

5 Referências Bibliográficas

Dinis J, Gonçalves M, Nicolau R, Reis R. Proposta de uma Nomenclatura de Ocupação do Solo. Relatório de execução do projecto LANDAU, Action 3.1 - Task 3. Lisboa, Instituto Geográfico Português, Janeiro 2012a.

Dinis J, Rodrigues P, Gonçalves M, Nicolau R, Reis R. Classificação uni-temporal da ocupação do solo através de dados de média resolução espacial – MERIS. Relatório de execução do projecto LANDAU, Action 3.2 - Task 3. Lisboa, Instituto Geográfico Português, Abril 2012b.

Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: Data mining, inference and prediction. Springer, 2nd edition, 2009.

Mitchie M, Spiegelhater DJ, Taylor CC. Machine learning, neural and statistical classification. Ellis Horwood series in artificial intelligence, 1994.

Nicolau R, Dinis J, Reis R. Definição das áreas de estudo e selecção dos dados de base. Relatório de execução do projecto LANDAU, Tarefa 1. Lisboa, Instituto Geográfico Português, Junho 2011.

Plantier T A. (2006). Análise de padrões espectrais e espaciais de imagens de muito grande resolução espacial para caracterização da ocupação do solo. Dissertação de Mestrado. Instituto Superior Técnico, Universidade Técnica de Lisboa. 2006.

ANEXO I – Matrizes de erro / confusão associadas à classificação da ocupação do solo na área de estudo E, com base em dados QUICKBIRD

Quadro I.1 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo Maximum Likelihood – ML

		Informação de Referência														Total	\hat{P}_u (%)		
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6	
Classificação	4	44															44	100%	
	5		19		2		1											22	86%
	6			-														-	-
	7				28													28	100%
	1.1					-												-	-
	1.2	2					27		2		1	1		1		1		35	77%
	2.1							44			4			2				50	88%
	2.2	4					8		54		1	4		6	2	4		83	65%
	2.3									-								-	-
	3.1										58	3				1		62	94%
	3.2										5	62				16		83	75%
	3.3												-					-	-
	3.4						1		10						34		1	46	74%
	3.5											9				62		71	87%
3.6	2					2		2					1			30	37	81%	
Total		52	19	-	30	-	39	44	68	-	69	79	-	44	81	36	561		
\hat{P}_p (%)		85%	100%	-	93%	-	69%	100%	79%	-	84%	78%	-	77%	77%	83%		$\hat{P} = 82\%$	

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro I.2 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo Linear Discriminant Classifier – LDC

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	43															44	100%
	5		17				1					2			2		22	86%
	6			-													-	-
	7		2		30							1					28	100%
	1.1					-											-	-
	1.2	4					31		18							4	35	77%
	2.1							45			6						50	88%
	2.2	2					1		38		5	8		13		3	83	65%
	2.3									-							-	-
	3.1								2		54	17			8		62	94%
	3.2										2	39			22		83	75%
	3.3												-				-	-
	3.4								13			1		22	4		46	74%
	3.5										1	11		1	48		71	87%
3.6	3					6		2							29	37	81%	
Total		52	19	-	30	-	39	45	73	-	68	79	-	36	84	36	561	
\hat{P}_p (%)		83%	89%	-	100%	-	79%	100%	52%	-	79%	49%	-	61%	57%	81%		$\hat{P} = 71\%$

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro I.3 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo

Diagonal Quadratic Discriminant Classifier - DQDC

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	42					3										45	93%
	5		15		2		1							1			19	79%
	6			-													-	-
	7		3		28												31	90%
	1.1					-											-	-
	1.2	7					7	1	1		3						19	37%
	2.1							43			7				2		52	83%
	2.2	3					14		51			9		10	1	11	99	52%
	2.3									-							-	-
	3.1										47	10				3	60	78%
	3.2										11	31				17	59	53%
	3.3												-				-	-
	3.4		1				1		13						27		46	59%
	3.5										1	29				60	90	67%
3.6						13		6					1			41	51%	
Total		52	19	-	30	-	39	44	71	-	69	79	-	41	81	36	561	
\hat{P}_p (%)		81%	79%	-	93%	-	18%	98%	72%	-	68%	39%	-	66%	74%	58%		$\hat{P} = 66\%$

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro I.4 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo K Nearest Neighbours – KNN

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	41							1						1	43	95%	
	5		14		3		1								1	1	20	70%
	6			-													-	-
	7		1		27												28	96%
	1.1					-											-	-
	1.2	11	2				32	2	17			6		3	5	12	90	36%
	2.1							43			8						51	84%
	2.2	2					4		44		1	6		8		4	69	64%
	2.3									-							-	-
	3.1										51	6			2		59	86%
	3.2										8	52			22		82	63%
	3.3												-				-	-
	3.4						1		5					31	1	2	40	78%
	3.5		1								1	9			51		62	82%
3.6		1				1							1		14	17	82%	
Total		54	19	-	30	-	39	45	67	-	69	79	-	43	82	34	561	
\hat{P}_p (%)		76%	74%	-	90%	-	82%	96%	66%	-	74%	66%	-	72%	62%	41%		$\hat{P} = 71\%$

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro I.5 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo **PARZEN**

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	41					1	1	1						1	45	91%	
	5		15		3		1								1	1	21	71%
	6			-													-	-
	7		1		27												28	96%
	1.1					-											-	-
	1.2	10	1				31	2	16			5		3	5	11	84	37%
	2.1							42			8						50	84%
	2.2	2					4		45		1	7		9		5	73	62%
	2.3									-							-	-
	3.1										51	6			2		59	86%
	3.2										8	52			21		81	64%
	3.3												-				-	-
	3.4						1		5					30	1	2	39	77%
	3.5		1								1	9			52		63	83%
3.6		1				1							1		15	18	83%	
Total		53	19	-	30	-	39	45	67	-	69	79	-	43	82	35	561	
\hat{P}_p (%)		77%	79%	-	90%	-	79%	93%	67%	-	74%	66%	-	70%	63%	43%		$\hat{P} = 71\%$

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro I.6 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo

Classification and Regression Tree - CART

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	34							1							1	36	94%
	5		16		2							1					19	84%
	6			-													-	-
	7				28												28	100%
	1.1					-											-	-
	1.2	17	1				32	1	19		2	2		6	7	6	93	34%
	2.1							44			5						49	90%
	2.2	2					4		47		1	8		5		4	71	66%
	2.3									-							-	-
	3.1										56	9			3		68	82%
	3.2										4	44			20		68	65%
	3.3												-				-	-
	3.4		1				1		4					25		1	32	78%
	3.5		1								1	15		1	52		70	74%
3.6						2		1					1		23	27	85%	
Total	53	19	-	30	-	39	45	72	-	69	79	-	38	82	35	561		
\hat{P}_p (%)	64%	84%	-	93%	-	82%	98%	65%	-	81%	56%	-	66%	63%	66%		$\hat{P} = 71\%$	

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro I.7 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo

Backpropagation Multilayer Perceptron - BMP

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	35					1										36	97%
	5		17		3		1					2					23	74%
	6			-													-	-
	7		2		27												29	93%
	1.1					-											-	-
	1.2	16					37		32					1		36	122	30%
	2.1							45			4						49	92%
	2.2	1							28		7	14		20	2		72	39%
	2.3									-							-	-
	3.1										57	3			1		61	93%
	3.2										1	31			15		47	66%
	3.3												-				-	-
	3.4								16			2		12	1		31	39%
	3.5											27			64		91	70%
3.6															0	0	0%	
Total	52	19	-	30	-	39	45	76	-	69	79	-	33	83	36	561		
\hat{P}_p (%)	67%	89%	-	90%	-	95%	100%	37%	-	83%	39%	-	36%	77%	0%		$\hat{P} = 63\%$	

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

ANEXO II – Matrizes de erro / confusão associadas à classificação da ocupação do solo na área de estudo J, com base em dados QUICKBIRD

Quadro II.1 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo **Maximum Likelihood – ML**

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	21							2								23	91%
	5		-														-	-
	6			-													-	-
	7				50												50	100%
	1.1					-											-	-
	1.2	5					41		2					1		3	52	79%
	2.1							21							1		22	95%
	2.2	1					1		82					3	14	1	102	80%
	2.3									-							-	-
	3.1										49			1			50	98%
	3.2											50			3		53	94%
	3.3												-				-	-
	3.4								13					41	7		61	67%
	3.5								3		1			4	26		34	76%
3.6						2							2	4	46	54	85%	
Total		27	-	-	50	-	44	21	102	-	50	50	-	52	55	50	501	
\hat{P}_p (%)		78%	-	-	100%	-	93%	100%	80%	-	98%	100%	-	79%	47%	92%		$\hat{P} = 85\%$

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro II.2 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo Linear Discriminant Classifier – LDC

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	24							1								25	96%
	5		-														-	-
	6			-													-	-
	7				50		3										53	94%
	1.1					-											-	-
	1.2	1					19		6					1			27	70%
	2.1							19							2		21	90%
	2.2	2					6	1	77					1	4	7	98	79%
	2.3									-							-	-
	3.1							2	1		48				4		55	87%
	3.2								1			36		1	6		44	82%
	3.3												-				-	-
	3.4						15		12					52	24	1	104	50%
	3.5										3	14		1	6		24	25%
3.6						1		3					1	3	42	50	84%	
Total		27	-	-	50	-	44	22	101	-	51	50	-	57	49	50	501	
\hat{P}_p (%)		89%	-	-	100%	-	43%	86%	76%	-	94%	72%	-	91%	12%	84%		$\hat{P} = 74\%$

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro II.3 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo

Diagonal Quadratic Discriminant Classifier – DQDC

		Informação de Referência														Total	\hat{P}_u (%)		
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6	
Classificação	4	21															21	100%	
	5		-															-	-
	6			-														-	-
	7				50													50	100%
	1.1					-												-	-
	1.2	5					18		7						2		8	40	45%
	2.1							22									1	23	96%
	2.2	1					15		75						3	4	9	107	70%
	2.3									-								-	-
	3.1											51	3					54	94%
	3.2											2	44		1	3		50	88%
	3.3													-				-	-
	3.4						8		14						50	20	1	93	54%
	3.5								2		1				3	21		27	78%
3.6						3		1								32	36	89%	
Total		27	-	-	50	-	44	22	99	-	54	47	-	59	49	50	501		
\hat{P}_p (%)		78%	-	-	100%	-	41%	100%	76%	-	94%	94%	-	85%	43%	64%		$\hat{P} = 77\%$	

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro II.4 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo K Nearest Neighbours – KNN

		Informação de Referência														Total	\hat{P}_u (%)		
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6	
Classificação	4	13															13	100%	
	5		-															-	-
	6			-														-	-
	7				50													50	100%
	1.1					-												-	-
	1.2	11					38		4					3	2	10	68	56%	
	2.1							20							1		21	95%	
	2.2	2					4	2	85					6	9	2	110	77%	
	2.3									-								-	-
	3.1										48	1		1			50	96%	
	3.2											47			7		54	87%	
	3.3												-					-	-
	3.4						2		10						39	16	67	58%	
	3.5								3		4	2		2	14		25	56%	
3.6	1												2	2	38	43	88%		
Total		27	-	-	50	-	44	22	102	-	52	50	-	53	51	50	501		
\hat{P}_p (%)		48%	-	-	100%	-	86%	91%	83%	-	92%	94%	-	74%	27%	76%		$\hat{P} = 78\%$	

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro II.5 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo **PARZEN**

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	16															16	100%
	5		-														-	-
	6			-													-	-
	7				50												50	100%
	1.1					-											-	-
	1.2	9					38		2					1	1	8	59	64%
	2.1							21							1		22	95%
	2.2	1					3	1	81					1	4	1	92	88%
	2.3									-							-	-
	3.1										47			1			48	98%
	3.2											50			6		56	89%
	3.3												-				-	-
	3.4						2		13					51	16		82	62%
	3.5								3		4			2	19		28	68%
3.6	1					1							2	3	41	48	85%	
Total	27	-	-	50	-	44	22	99	-	51	50	-	58	50	50	501		
\hat{P}_p (%)	59%	-	-	100%	-	86%	95%	82%	-	92%	100%	-	88%	38%	82%		$\hat{P} = 83\%$	

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro II.6 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo

Classification and Regression Tree – CART

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	21							2								23	91%
	5		-														-	-
	6			-													-	-
	7				50												50	100%
	1.1					-											-	-
	1.2	5					41		2					1		3	52	79%
	2.1							21							1		22	95%
	2.2	1					1		82					3	14	1	102	80%
	2.3									-							-	-
	3.1										49			1			50	98%
	3.2											50			3		53	94%
	3.3												-				-	-
	3.4								13					41	7		61	67%
	3.5								3		1			4	26		34	76%
3.6						2							2	4	46	54	85%	
Total	27	-	-	50	-	44	21	102	-	50	50	-	52	55	50	501		
\hat{P}_p (%)	78%	-	-	100%	-	93%	100%	80%	-	98%	100%	-	79%	47%	92%		$\hat{P} = 85\%$	

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro II.7 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo

Backpropagation Multilayer Perceptron - BMP

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	21					1										22	95%
	5		-														-	-
	6			-													-	-
	7				50										1		51	98%
	1.1					-											-	-
	1.2	4					7		2					1		17	31	23%
	2.1							20							1		21	95%
	2.2	2					27	2	93					3	15	29	171	54%
	2.3									-							-	-
	3.1										55	7		1	2		65	85%
	3.2											39			8		47	83%
	3.3												-				-	-
	3.4						9		6					52	22	4	93	56%
	3.5														0		0	0%
3.6															0	0	0%	
Total	27	-	-	50	-	44	22	101	-	55	46	-	57	49	50	501		
\hat{P}_p (%)	78%	-	-	100%	-	16%	91%	92%	-	100%	85%	-	91%	0%	0%		$\hat{P} = 67\%$	

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

ANEXO III – Matrizes de erro / confusão associadas à classificação da ocupação do solo na área de estudo A, com base em dados QUICKBIRD

Quadro III.1 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo Maximum Likelihood – ML

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	30			1		2							1			34	88%
	5		-														-	-
	6			29													29	100%
	7				40												40	100%
	1.1					-											-	-
	1.2	4		1	1		32							1			39	82%
	2.1							50		2							52	96%
	2.2	10		3			9		50	1				8			81	62%
	2.3								1	46							47	98%
	3.1				2						50			1			53	94%
	3.2	1		3	1							50			1		56	89%
	3.3												-				-	-
	3.4	4		2	1		7		3	1					56		74	76%
	3.5			3	1									1	49		54	91%
3.6																-	-	
Total		49	-	41	47	-	50	50	54	50	50	50	-	68	50	-	559	
\hat{P}_p (%)		61%	-	71%	85%	-	64%	100%	93%	92%	100%	100%	-	82%	98%	-		$\hat{P} = 86\%$

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro III.2 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo Linear Discriminant Classifier – LDC

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	24			1		6										31	77%
	5		-														-	-
	6			30			4					1		1			36	83%
	7	1		2	42												45	93%
	1.1					-											-	-
	1.2	7					16							1			24	67%
	2.1							50		2							52	96%
	2.2	12					10		53	2				6			83	64%
	2.3						5			46				1			52	88%
	3.1	2			1						50	1		2			56	89%
	3.2	1		5							1	47		1	2		57	82%
	3.3												-				-	-
	3.4	1			1		8		2					51			63	81%
	3.5			4	2		1							4	49		60	82%
3.6																-	-	
Total		48	-	41	47	-	50	50	55	50	51	49	-	67	51	-	559	
\hat{P}_p (%)		50%	-	73%	89%	-	32%	100%	96%	92%	98%	96%	-	76%	96%	-		$\hat{P} = 82\%$

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro III.3 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo
Diagonal Quadratic Discriminant Classifier – DQDC

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	23			1		4										28	82%
	5		-														-	-
	6			24	1		1							1	9		36	67%
	7			1	40												41	98%
	1.1					-											-	-
	1.2	15					16		2					2			35	46%
	2.1							50		2	1						53	94%
	2.2	6					14		53	1				10			84	63%
	2.3	2							1	46							49	94%
	3.1			1	1						50	15		1			68	74%
	3.2	1			2							34			12		49	69%
	3.3												-				-	-
	3.4	1		7	1		13		3	1				50			76	66%
	3.5			8								1		1	30		40	75%
3.6															-	-	-	
Total	48	-	41	46	-	48	50	59	50	51	50	-	65	51	-	559		
\hat{P}_p (%)	48%	-	59%	87%	-	33%	100%	90%	92%	98%	68%	-	77%	59%	-		$\hat{P} = 74\%$	

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro III.4 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo K Nearest Neighbours – KNN

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	23			1		4										28	82%
	5		-														-	-
	6			24													24	100%
	7			2	41												43	95%
	1.1					-											-	-
	1.2	6		3			31							4			44	70%
	2.1				1			50		2				1			54	93%
	2.2	18		2			12		54	14				10			110	49%
	2.3						1		1	34				2			38	89%
	3.1										49	14		1	1		65	75%
	3.2	1		3	1						2	34		1	4		46	74%
	3.3												-				-	-
	3.4	1			2		1							46	2		52	88%
	3.5			8	1		1					1		1	43		55	78%
3.6															-	-	-	
Total		49	-	42	47	-	50	50	55	50	51	49	-	66	50	-	559	
\hat{P}_p (%)		47%	-	57%	87%	-	62%	100%	98%	68%	96%	69%	-	70%	86%	-		$\hat{P} = 77\%$

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro III.5 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo PARZEN

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	25			1		13										39	64%
	5		-														-	-
	6			24													24	100%
	7			2	41												43	95%
	1.1					-											-	-
	1.2	5		3			22							4			34	65%
	2.1				1			50		2				1			54	93%
	2.2	17		2			12		53	14				10			108	49%
	2.3						1		1	34				2			38	89%
	3.1										49	14		1	1		65	75%
	3.2	1		3	1						2	34		1	4		46	74%
	3.3												-				-	-
	3.4	1			2		1		1					46	3		54	85%
	3.5			8	1		1					1		1	42		54	78%
3.6															-	-	-	
Total		49	-	42	47	-	50	50	55	50	51	49	-	66	50	-	559	
\hat{P}_p (%)		51%	-	57%	87%	-	44%	100%	96%	68%	96%	69%	-	70%	84%	-		$\hat{P} = 75\%$

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro III.6 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo

Classification and Regression Tree – CART

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	26			1		3										30	87%
	5		-														-	-
	6			19												1	20	95%
	7				2	43											45	96%
	1.1					-											-	-
	1.2	7					31		3		2				2		45	69%
	2.1							50		2					1		53	94%
	2.2	12		1	1		10		54	10					17		105	51%
	2.3						1		1	37							39	95%
	3.1	1			1						45	11			1		59	76%
	3.2	1		3	1						4	38			1	6	54	70%
	3.3												-				-	-
	3.4	2		6			3			1					43	1	56	77%
	3.5			9	1										1	42	53	79%
3.6																-	-	
Total	49	-	40	48	-	48	50	58	50	51	49	-	66	50	-	559		
\hat{P}_p (%)	53%	-	48%	90%	-	65%	100%	93%	74%	88%	78%	-	65%	84%	-		$\hat{P} = 77\%$	

- classe de espaço não representada na área de estudo, à resolução espacial adoptada

Quadro III.7 – Matriz de erro / confusão associada à classificação da ocupação do solo produzida pelo algoritmo
Backpropagation Multilayer Perceptron – BMP

		Informação de Referência														Total	\hat{P}_u (%)	
		4	5	6	7	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5			3.6
Classificação	4	19					2										21	90%
	5		-														-	-
	6			1													1	100%
	7	1		6	43												50	86%
	1.1					-											-	-
	1.2	3					28							1			32	88%
	2.1							50		2							52	96%
	2.2	24		11	2		14		63	22				42			178	35%
	2.3	1					2		1	26				2			32	81%
	3.1										51	4		1			56	91%
	3.2	1		12	1							46			2		62	74%
	3.3												-				-	-
	3.4													6			6	100%
	3.5			11	1									8	49		69	71%
3.6															-	-	-	
Total	49	-	41	47	-	46	50	64	50	51	50	-	60	51	-	559		
\hat{P}_p (%)	39%	-	2%	91%	-	61%	100%	98%	52%	100%	92%	-	10%	96%	-		$\hat{P} = 68\%$	

- classe de espaço não representada na área de estudo, à resolução espacial adoptada