

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

Action 3.2 - Task 3

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

Joel Dinis, Pedro Rodrigues, Rita Nicolau

Dezembro 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	Apresentação e Discussão de 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	21

ANEXO I – Matrizes de erro / confusão associadas à classificação da ocupação do solo na área de estudo E, com base em dados LANDSAT23

ANEXO II – Matrizes de erro / confusão associadas à classificação da ocupação do solo na área de estudo J, com base em dados LANDSAT33

ANEXO III – Matrizes de erro / confusão associadas à classificação da ocupação do solo na área de estudo A, com base em dados LANDSAT43

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 consistiram na classificação automática da ocupação do solo de três áreas estudo de Portugal Continental, de acordo com a nomenclatura de ocupação do solo LANDAU, e com recurso a dados de satélite de elevada resolução espacial (imagens LANDSAT).

O principal objectivo do estudo é a avaliação do desempenho das sete metodologias de classificação automática ensaiadas na produção de mapas de ocupação de solo, desenvolvidos a partir de dados LANDSAT.

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 LANDSAT, que apresentam um nível detalhe intermédio (pixels de 30 metros de resolução geométrica) de entre as 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 LANDSAT, empregues na classificação da ocupação do solo das três áreas citadas, são relativas ao ano de 2007, designadamente ao mês de Junho para E, e Julho para J e 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 – CLC (2000, 2006)
- Carta de Ocupação/uso do Solo –COS (2007).

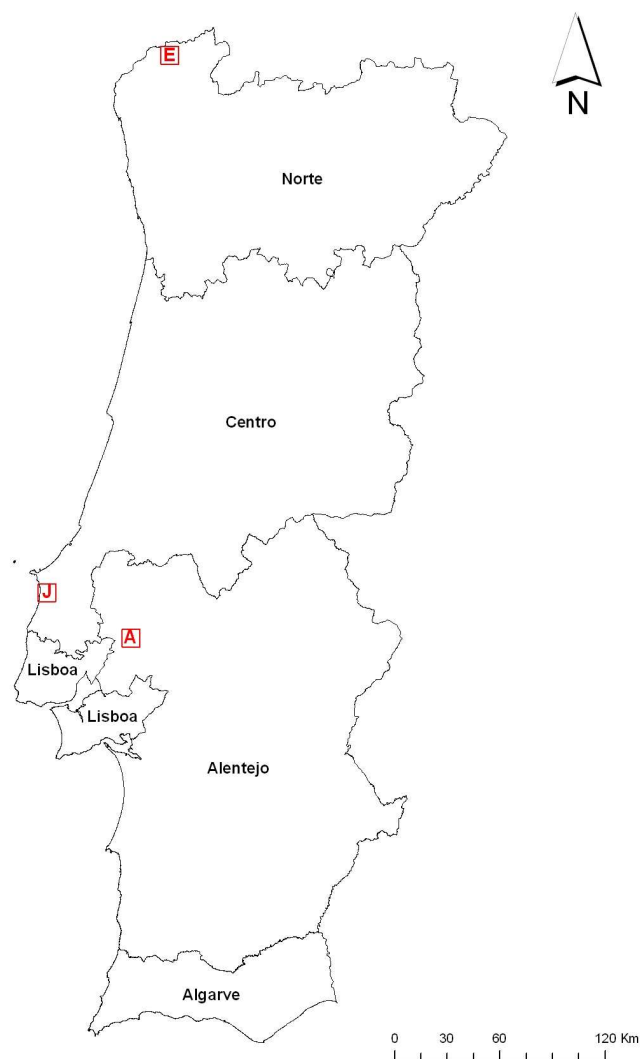


Figura 1 - Áreas de estudo LANDAU: E, J e A

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 três 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 8764 observações (pontos). A amostra obtida para a área de estudo J era constituída por 9182 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	218	73	-
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 Áridas	550	-	-
6 Zonas Húmidas	-	-	190
7 Corpos de Água	295	1016	400
TOTAL	8764	9182	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 constituídas por 575 observações (pontos) na área E, 665 observações na área J e 550 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 o risco do produtor ($< 15\%$) e o risco do utilizador ($< 5\%$), mas também o esforço operacional de concretização.

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 Apresentação e Discussão de 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 obtidas 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) verifica-se que o algoritmo com melhor desempenho global na classificação do uso do solo foi o ML ($\hat{P} = 75\%$). O DQDC foi o algoritmo que denotou pior desempenho global ($\hat{P} = 64\%$). Com um desempenho global ligeiramente inferior ao do ML, o algoritmo BMP também se destacou neste âmbito ($\hat{P} = 73\%$), apesar da sua deficiente performance na classificação de manchas de Floresta Mista (3.3).

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

- Na área de estudo J não se encontravam presentes as classes Áreas Artificiais Contínuas (1.1), Arroçais (2.3), Áreas Áridas (5) e Zonas Húmidas (6).
- A maior exactidão do produtor obtida na classificação da Áreas Artificiais Descontínuas (1.2) foi disponibilizada pelo algoritmo ML ($\hat{P}_p = 64\%$), que disponibilizou igualmente a mais elevada exactidão do utilizador ($\hat{P}_u = 79\%$).
- Uma exactidão do produtor máxima ($\hat{P}_p = 100\%$) foi alcançada na classificação de áreas de Agricultura de Regadio (2.1) através dos algoritmos LDC e BMP. Estas áreas foram igualmente classificadas com boa exactidão do produtor pelos algoritmos ML e DQDC ($\hat{P}_p = 97\%$ para ambos). A maior exactidão do utilizador obtida na classificação destas áreas ($\hat{P}_u = 67\%$) foi proporcionada pelo algoritmo ML.
- Na classificação da Agricultura de Sequeiro (2.2) o algoritmo ML evidenciou-se pela obtenção das maiores exactidões, tanto do ponto de vista do produtor ($\hat{P}_p = 74\%$) como do utilizador ($\hat{P}_u = 76\%$).
- Na classificação da Floresta de Folhosas (3.1) o ML foi igualmente o algoritmo que revelou melhor desempenho, apresentando não só a maior exactidão do produtor ($\hat{P}_p = 78\%$) mas também a maior exactidão do utilizador ($\hat{P}_u = 69\%$), igualmente apresentada pelo algoritmo BMP.

- Na classificação de Florestas de Resinosas (3.2) a mais elevada exactidão do produtor ($\hat{P}_p=78\%$) resultou da aplicação do algoritmo LDC e a mais elevada exactidão do utilizador resultou da aplicação dos algoritmos ML e KNN ($\hat{P}_u=93\%$ para ambos).
- Na classificação da Floresta Mista (3.3) o algoritmo DQDC obteve simultaneamente a maior exactidão do produtor e a maior exactidão do utilizador ($\hat{P}_p=\hat{P}_u=71\%$).
- Os algoritmos PARZEN e KNN foram, de entre os testados, os que possibilitaram classificar a Vegetação Herbácea (3.4) com maior exactidão do produtor ($\hat{P}_p=86\%$), mas a maior exactidão do utilizador foi alcançada pela aplicação do algoritmo ML ($\hat{P}_u=77\%$).
- O algoritmo BMP destacou-se na classificação dos Matos (3.5) pela apresentação da maior exactidão do produtor ($\hat{P}_p=70\%$), mas foi através do ML que se atingiu a maior exactidão do utilizador ($\hat{P}_u=67\%$).
- No que concerne à classificação da Vegetação Esparsa (3.6), a maior exactidão do produtor ($\hat{P}_p=71\%$) e a maior exactidão do utilizador ($\hat{P}_u=67\%$) foram obtidas através do algoritmo BMP.
- Na classificação do Solo Nu (4), o algoritmo ML originou a mais elevada exactidão do produtor ($\hat{P}_p=95\%$), tendo a maior exactidão do utilizador ($\hat{P}_u=94\%$) sido assegurada pelo algoritmo DQDC.
- A maior exactidão do produtor ($\hat{P}_p=88\%$) obtida na classificação das Áreas Áridas (5) resultou da aplicação dos algoritmos ML e BMP. O algoritmo LDC facultou a maior exactidão do utilizador ($\hat{P}_u=76\%$).
- Na classificação de Corpos de Água (7) todos os algoritmos testados obtiveram uma exactidão do utilizador máxima ($\hat{P}_u=100\%$), mas a maior exactidão do produtor ($\hat{P}_p=86\%$) foi somente obtida pelos algoritmos PARZEN e KNN.

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	\hat{P}_u	\hat{P}_p	\hat{P}_u	\hat{P}_p	\hat{P}_u	\hat{P}_p	\hat{P}_u	\hat{P}_p	\hat{P}_u	\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%	79%	57%	67%	52%	63%	57%	58%	57%	60%	60%	65%	60%	65%
2.1 Agricultura de Regadio	97%	67%	100%	59%	97%	49%	85%	58%	85%	58%	96%	58%	100%	66%
2.2 Agricultura de Sequeiro	74%	76%	56%	69%	49%	55%	61%	72%	61%	74%	63%	75%	66%	75%
2.3 Arrozais	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3.1 Floresta de Folhosas	78%	69%	59%	68%	75%	65%	63%	64%	62%	67%	77%	63%	76%	69%
3.2 Floresta de Resinosas	71%	93%	78%	80%	66%	80%	70%	93%	67%	91%	62%	91%	74%	85%
3.3 Floresta Mista	64%	64%	64%	50%	71%	71%	64%	45%	67%	48%	40%	40%	0%	0%
3.4 Vegetação Herbácea	71%	77%	72%	62%	57%	43%	86%	7%	86%	67%	76%	70%	61%	76%
3.5 Matos	65%	67%	61%	56%	66%	61%	62%	66%	62%	65%	66%	61%	70%	62%
3.6 Vegetação Esparsa	67%	65%	56%	56%	40%	50%	55%	63%	57%	63%	54%	63%	71%	67%
4 Solo Nu	95%	85%	84%	90%	82%	94%	86%	83%	88%	84%	88%	84%	88%	86%
5 Áreas Áridas	88%	70%	73%	76%	68%	75%	80%	67%	80%	65%	80%	70%	88%	72%
6 Zonas Húmidas	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7 Corpos de Água	77%	100%	73%	100%	64%	100%	86%	100%	86%	100%	82%	100%	64%	100%
\hat{P} = Exactidão Global	75%		69%		64%		70%		70%		70%		73%	

- 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 o algoritmo que apresentou melhor desempenho global na classificação do uso do solo foi o ML ($\hat{P} = 79\%$). O algoritmo BMP foi o que revelou pior desempenho global ($\hat{P} = 62\%$) neste contexto.

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

- Na área de estudo J não se encontravam presentes as classes Áreas Artificiais Contínuas (1.1), Arrozaís (2.3), Áreas Áridas (5) e Zonas Húmidas (6).
- As Áreas Artificiais Descontínuas (1.2) foram melhor classificadas do ponto de vista do produtor ($\hat{P}_p = 79\%$) pelos algoritmos KNN e PARZEN. A mais elevada exactidão do utilizador obtida na classificação destas áreas ($\hat{P}_u = 87\%$) foi proporcionada pelo algoritmo DQDC.
- Atingiu-se uma exactidão do produtor máxima ($\hat{P}_p = 100\%$) na classificação da Agricultura de Regadio (2.1) por recurso à grande maioria dos algoritmos ensaiados (com excepção para o algoritmo DQDC). A maior exactidão do utilizador obtida na classificação desta agricultura ($\hat{P}_u = 92\%$) foi produzida pelo algoritmo PARZEN.
- O BMP foi o algoritmo que permitiu classificar a Agricultura de Sequeiro (2.2) com maior exactidão do produtor ($\hat{P}_p = 93\%$). A maior exactidão do utilizador obtida na classificação destas áreas ($\hat{P}_u = 74\%$) resultou da aplicação do algoritmo ML.
- Na classificação da Floresta de Folhosas (3.1) os algoritmos KNN e CART proporcionaram a mais elevada exactidão do produtor ($\hat{P}_p = 87\%$). A maior exactidão do utilizador ($\hat{P}_u = 85\%$) na classificação deste tipo de floresta foi obtida pelo algoritmo LDC.
- Na classificação de Florestas de Resinosas (3.2), a maior exactidão do produtor ($\hat{P}_p = 82\%$) foi alcançada pelo algoritmo LDC e a maior exactidão do utilizador

($\hat{P}_u=80\%$) foi produzida pelo algoritmo ML.

- O melhor desempenho na classificação da Floresta Mista (3.3) foi possibilitado pelos algoritmos: DQDC no que diz respeito à exactidão do produtor ($\hat{P}_p=72\%$) e PARZEN no que diz respeito à exactidão do utilizador ($\hat{P}_u=50\%$).
- O algoritmo LDC foi o que permitiu classificar a Vegetação Herbácea (3.4) com maior exactidão do produtor ($\hat{P}_p=88\%$), tendo a maior exactidão do utilizador ($\hat{P}_u=94\%$) resultado da aplicação do algoritmo ML.
- A maior exactidão do produtor ($\hat{P}_p=79\%$) obtida na classificação dos Matos (3.5) deveu-se à aplicação do algoritmo BMP. O algoritmo LDC destacou-se na obtenção da maior exactidão do utilizador ($\hat{P}_u=81\%$).
- Na classificação da Vegetação Esparsa (3.6), o algoritmo ML foi responsável pela obtenção das maiores exactidões do produtor ($\hat{P}_p=63\%$) e do utilizador ($\hat{P}_u=74\%$).
- Na classificação do Solo Nu (4), os algoritmos KNN e PARZEN evidenciaram-se pela apresentação da máxima exactidão do produtor ($\hat{P}_p=100\%$) e da mais elevada exactidão do utilizador ($\hat{P}_u=81\%$). O algoritmo ML também obteve máxima exactidão do produtor ($\hat{P}_p=100\%$), mas uma exactidão do utilizador inferior à supracitada ($\hat{P}_u=71\%$).
- A maior exactidão do produtor ($\hat{P}_p=88\%$) obtida na classificação de Corpos de Água (7) resultou da aplicação dos algoritmos KNN, PARZEN, CART e BMP. Com excepção do BMP, todos os algoritmos testados disponibilizaram uma exactidão do utilizador máxima ($\hat{P}_u=100\%$).

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	\hat{P}_u	\hat{P}_p	\hat{P}_u	\hat{P}_p	\hat{P}_u	\hat{P}_p	\hat{P}_u	\hat{P}_p	\hat{P}_u	\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	78%	84%	73%	65%	67%	87%	79%	71%	79%	71%	68%	68%	65%	57%
2.1 Agricultura de Regadio	100%	79%	100%	79%	96%	73%	100%	85%	100%	92%	100%	77%	100%	48%
2.2 Agricultura de Sequeiro	92%	74%	77%	73%	84%	63%	83%	71%	83%	71%	81%	60%	93%	67%
2.3 Arrozais	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3.1 Floresta de Folhosas	83%	80%	64%	85%	62%	70%	87%	64%	84%	67%	87%	52%	0%	0%
3.2 Floresta de Resinosas	72%	80%	82%	69%	40%	74%	68%	74%	71%	71%	57%	74%	2%	25%
3.3 Floresta Mista	52%	42%	64%	36%	72%	38%	26%	45%	32%	50%	25%	24%	0%	0%
3.4 Vegetação Herbácea	76%	94%	88%	83%	67%	86%	62%	87%	63%	87%	50%	80%	85%	87%
3.5 Matos	76%	79%	72%	81%	63%	74%	68%	66%	68%	65%	64%	70%	79%	46%
3.6 Vegetação Esparsa	63%	74%	37%	61%	41%	38%	59%	69%	61%	70%	46%	67%	8%	63%
4 Solo Nu	100%	71%	95%	78%	95%	78%	100%	81%	100%	81%	87%	80%	89%	62%
5 Áreas Áridas	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6 Zonas Húmidas	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7 Corpos de Água	82%	100%	84%	100%	84%	100%	88%	100%	88%	100%	88%	100%	88%	88%
\hat{P} = Exactidão Global	79%		74%		68%		74%		74%		67%		62%	

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

3.3 Área de estudo A

Os resultados disponibilizados no Quadro 5 mostram que o LDC foi o algoritmo com melhor desempenho global na classificação do uso do solo da área de estudo A ($\hat{P} = 82\%$), correspondendo a segunda posição ao algoritmo ML ($\hat{P} = 81\%$). O DQDC foi dos algoritmos ensaiados o que revelou pior desempenho global ($\hat{P} = 76\%$) neste contexto.

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

- Na área de estudo A não se encontravam presentes as classes Áreas Artificiais Contínuas (1.1), Vegetação Esparsa (3.6) e Áreas Ardidias (5). O detalhe disponibilizado pelo sensor LANDSAT não permitiu discriminar a classe Floresta Mista (3.3).
- O melhor desempenho na classificação das Áreas Artificiais Descontínuas (1.2) foi assegurado pelos seguintes algoritmos: BMP, do ponto de vista do produtor ($\hat{P}_p = 85\%$), e DQDC, do ponto de vista do utilizador ($\hat{P}_u = 94\%$).
- Na classificação da Agricultura de Regadio (2.1) o algoritmo LDC produziu a maior exactidão do produtor ($\hat{P}_p = 98\%$), mas a mais elevada exactidão do utilizador obtida na classificação destas áreas ($\hat{P}_u = 84\%$) foi assegurada pelos algoritmos KNN e PARZEN.
- O DQDC foi o algoritmo que permitiu classificar a Agricultura de Sequeiro (2.2) com maior exactidão do produtor ($\hat{P}_p = 79\%$). A maior exactidão do utilizador alcançada na classificação desta agricultura resultou da aplicação dos algoritmos KNN e PARZEN ($\hat{P}_u = 86\%$).
- O DQDC foi igualmente o algoritmo que permitiu classificar os Arrozais (2.3) com maior exactidão do produtor ($\hat{P}_p = 83\%$), mas a maior exactidão do utilizador foi obtida através do algoritmo ML ($\hat{P}_u = 92\%$).
- Na classificação da Floresta de Folhosas (3.1) o algoritmo ML foi responsável pela obtenção da maior exactidão do produtor ($\hat{P}_p = 83\%$). A maior exactidão do

utilizador ($\hat{P}_u = 88\%$) foi disponibilizada pelo algoritmo LDC.

- Na classificação de Florestas de Resinosas (3.2), destacaram-se os algoritmos: LDC pela apresentação da maior exactidão do produtor ($\hat{P}_p = 95\%$) e ML pela apresentação da maior exactidão do utilizador ($\hat{P}_u = 89\%$).
- O LDC foi também, de entre os algoritmos testados, o que revelou maior exactidão do produtor ($\hat{P}_p = 93\%$) na classificação da Vegetação Herbácea (3.4). A máxima exactidão do utilizador ($\hat{P}_u = 80\%$) na classificação deste tipo de vegetação resultou da aplicação do algoritmo BMP.
- O algoritmo LDC destacou-se mais uma vez pela apresentação da maior exactidão do produtor ($\hat{P}_p = 84\%$) na classificação dos Matos (3.5), embora a maior exactidão do utilizador tivesse sido obtida pelos algoritmos ML e BMP ($\hat{P}_u = 91\%$).
- Na classificação do Solo Nu (4), quer o LDC, quer o DQDC, produziram uma exactidão do produtor máxima ($\hat{P}_p = 100\%$). Contudo a maior exactidão do utilizador ($\hat{P}_u = 89\%$) só foi assegurada pela aplicação do algoritmo LDC.
- O algoritmo ML foi o que permitiu classificar as Zonas Húmidas (6) com maior exactidão do produtor ($\hat{P}_p = 86\%$). A mais elevada exactidão do utilizador ($\hat{P}_u = 87\%$) obtida na classificação destas zonas foi disponibilizada pelo algoritmo CART.
- Na classificação de Corpos de Água (7), a mais elevada exactidão do produtor ($\hat{P}_p = 97\%$) foi proporcionada pelos algoritmos LDC e BMP. A máxima exactidão do utilizador ($\hat{P}_u = 95\%$) resultou da aplicação do algoritmo CART.

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	\hat{P}_u	\hat{P}_p	\hat{P}_u	\hat{P}_p	\hat{P}_u	\hat{P}_p	\hat{P}_u	\hat{P}_p	\hat{P}_u	\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	71%	86%	74%	73%	33%	94%	84%	63%	84%	63%	82%	56%	85%	70%
2.1 Agricultura de Regadio	79%	80%	98%	72%	91%	81%	95%	84%	95%	84%	93%	81%	89%	75%
2.2 Agricultura de Sequeiro	71%	77%	64%	82%	79%	66%	51%	86%	50%	86%	49%	82%	67%	78%
2.3 Arrozais	81%	92%	62%	90%	83%	70%	81%	79%	81%	79%	79%	83%	76%	69%
3.1 Floresta de Folhosas	83%	81%	69%	88%	66%	73%	71%	71%	73%	71%	70%	70%	79%	79%
3.2 Floresta de Resinosas	93%	89%	95%	80%	79%	74%	86%	71%	86%	73%	88%	72%	93%	78%
3.3 Floresta Mista	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3.4 Vegetação Herbácea	84%	76%	93%	77%	87%	73%	86%	74%	86%	75%	81%	76%	82%	80%
3.5 Matos	77%	91%	84%	88%	75%	83%	72%	82%	72%	82%	66%	78%	83%	91%
3.6 Vegetação Esparsa	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4 Solo Nu	97%	63%	100%	89%	100%	71%	97%	79%	97%	79%	94%	78%	94%	79%
5 Áreas Áridas	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6 Zonas Húmidas	86%	72%	71%	79%	53%	78%	67%	82%	67%	82%	79%	87%	0%	0%
7 Corpos de Água	86%	94%	97%	93%	92%	93%	87%	93%	87%	93%	87%	95%	97%	83%
\hat{P} = Exactidão Global	81%		82%		76%		78%		78%		77%		79%	

- 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 das áreas estudadas. A exactidão global proporcionada pelo ML para o conjunto das três zonas estudadas, oscilou entre 75% e 81%, apresentando um valor médio de 78%. O algoritmo LDC ocupou a segunda posição no ranking dos classificadores ensaiados, denotando uma exactidão global média de 75% no conjunto das três zonas estudadas.

Apresenta-se de seguida uma análise da classificação segundo classes da nomenclatura LANDAU, realizada pelo algoritmo ML para o conjunto das três áreas estudadas. A Figura 2 complementa as ilações apresentadas.

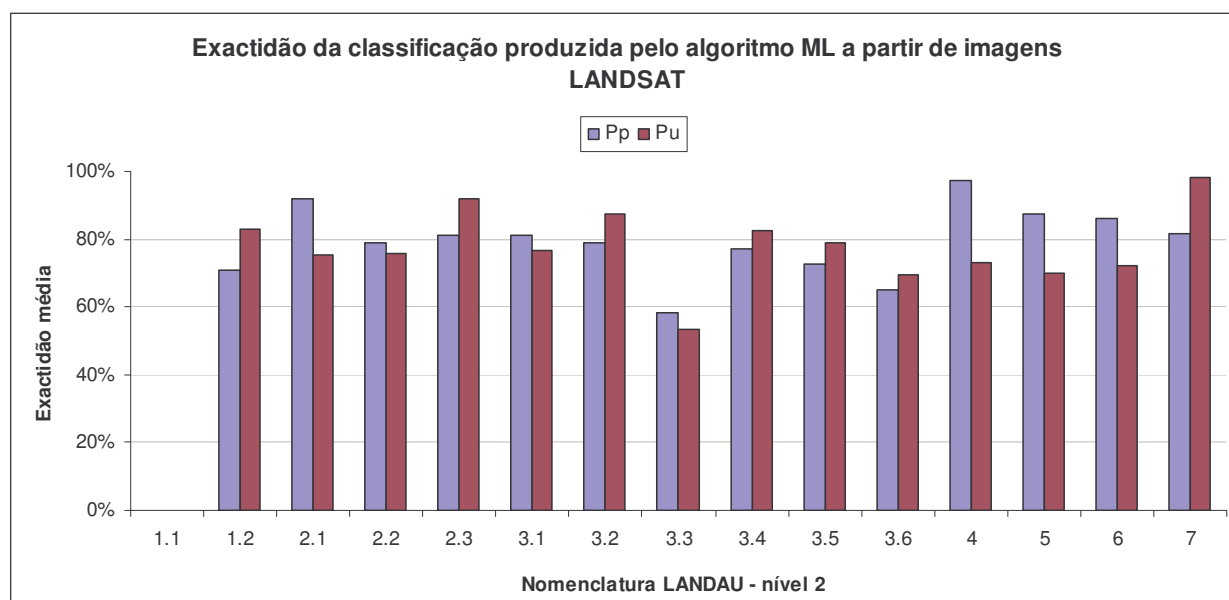


Figura 2 – Valores médios de exactidão do produtor e do utilizador, segundo classes da nomenclatura LANDAU, obtidos pelo algoritmo ML para o conjunto das três áreas estudadas

As imagens LANDSAT não viabilizaram o reconhecimento de Áreas Urbanas Contínuas (1.1) em qualquer das zonas estudadas. Esta situação, já ocorrida num trabalho similar (Dinis *et al.*, 2012c) em que se utilizaram imagens de muito elevada resolução (QUICKBIRD), é devida à baixa densidade deste tipo de ocupação naquelas áreas.

Na classificação das Áreas Urbanas Descontínuas (1.2), O algoritmo ML proporcionou exactidões variáveis entre 64% a 71% do ponto de vista do produtor, e entre 79% a 86% do ponto de vista 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 média do produtor (92%) do que a Agricultura de Sequeiro (2.2) (79%). Já a exactidão média do utilizador obtida pelo ML na classificação do Sequeiro (76%) foi quase da mesma ordem de grandeza da auferida na classificação do Regadio (75%).

Os Arrozais (2.3), somente identificados na área A, foram classificados pelo ML com uma exactidão do utilizador de 92% e uma exactidão do produtor foi de 81%.

No grande grupo das Florestas e Meios Naturais e Semi-naturais (3), o algoritmo ML teve pior desempenho na classificação da Floresta Mista (3.3), que apenas foi identificada em E e J. Neste tipo de floresta, o ML obteve exactidões do utilizador e do produtor que não ultrapassaram os 64%.

O ML evidenciou maior fidelidade na classificação das Florestas de Folhosas (3.1) e de Resinosas (3.2). Nas primeiras (3.1) obteve exactidões do produtor iguais ou superiores a 78% e exactidões do utilizador iguais ou superiores a 69%. Nas segundas (3.2) produziu exactidões do produtor iguais ou superiores a 71% e exactidões do utilizador iguais ou superiores a 80%.

No que diz respeito à classificação da Vegetação Herbácea (3.4), o ML obteve uma exactidão média do produtor de 77% e uma exactidão média do utilizador de 82%.

Na classificação dos Matos (3.5), as exactidões proporcionadas pelo ML oscilaram entre: 67% a 91% do ponto de vista do utilizador, e 65% a 77% do ponto de vista do produtor.

A Vegetação Esparsa (3.6), apenas identificada em E e J, foi a segunda classe de espaço pior classificada pelo ML, que obteve exactidões médias do produtor e do utilizador de 65% e 70%, respectivamente.

A aplicação do ML na classificação de Solo Nu (4) possibilitou alcançar uma exactidão média do produtor de 97% e uma exactidão média do utilizador de 73%.

Na classificação de Áreas Áridas (5), apenas identificadas em E, a exactidão do utilizador obtida pelo ML foi de 70% e a exactidão do produtor foi de 88%.

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 72% e a exactidão do produtor foi de 86%.

Na classificação de Corpos de Água (7), a exactidão média do utilizador obtida através do ML foi de 98% e a exactidão média do produtor foi de 82%.

4 Conclusões

Todos os algoritmos testados evidenciaram maior exactidão global na classificação do uso do solo na área A, comparativamente às restantes áreas de estudo. Tal como já referido num trabalho precedente (Dinis *et al.*, 2012c), a área de estudo A contem uma superfície agrícola e agro-florestal superior à das restantes áreas e é das estudadas a mais intervencionada pelo homem. Trata-se igualmente da área que apresenta manchas de ocupação do solo de maiores dimensões e com maior homogeneidade espectral, facto que poderá fundamentar a obtenção de melhores resultados na sua classificação.

Ao invés, a área de estudo E é das três estudadas a sujeita a menor pressão humana. A dominância de Matos e de Floresta, a par da elevada fragmentação da paisagem, originam uma menor homogeneidade espectral que poderá constituir justificação para que a totalidade dos algoritmos testados tenham alcançado piores resultados na classificação da ocupação do solo desta área.

O presente estudo sugere que, à escala das imagens LANDSAT, o algoritmo ML (Maximum Likelihood), denota globalmente melhores resultados para o território nacional do que os restantes algoritmos de classificação testados.

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.

Dinis J, Rodrigues P, Nicolau R. Classificação uni-temporal da ocupação do solo através de dados de muito elevada resolução espacial – QUICKBIRD. Relatório de execução do projecto LANDAU, Action 3.2 - Task 3. Lisboa, Instituto Geográfico Português, Setembro 2012c.

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.

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

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	55			1		8									1	65	85%	
	5		49		2						1	9			5	4	70	70%	
	6			-													-	-	
	7				17													17	100%
	1.1					-												-	-
	1.2	1					34		3					1		4	43	79%	
	2.1							29	6		2			3	3		43	67%	
	2.2	1					8		50					2	2	3	66	76%	
	2.3									-								-	-
	3.1										31	2	5		7		45	69%	
	3.2				1							55			3		59	93%	
	3.3				1						1	1	9		2		14	64%	
	3.4								4					27	2	2	35	77%	
	3.5						1	3			4	9		4	46	2	69	67%	
3.6	1	7				3		2		1	1		1	1	32	49	65%		
Total		58	56	-	22	-	53	30	68	-	40	77	14	38	71	48	575		
\hat{P}_p (%)		95%	88%	-	77%	-	64%	97%	74%	-	78%	71%	64%	71%	65%	67%		$\hat{P} = 75\%$	

- 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	47					5									52	90%	
	5		41									6			5	2	54	76%
	6			-													-	-
	7				16												16	100%
	1.1					-											-	-
	1.2	8					30									7	45	67%
	2.1						1	30	6		4			5	5		51	59%
	2.2	1					10		35					2	1	2	51	69%
	2.3									-							-	-
	3.1										23		5		6		34	68%
	3.2		3		5						1	60			5	1	75	80%
	3.3				1						1		9		6	1	18	50%
	3.4								16					31	1	2	50	62%
	3.5		1				1		3		9	11		5	45	6	81	56%
3.6		11				6		3		1					27	48	56%	
Total		56	56	-	22	-	53	30	63	-	39	77	14	43	74	48	575	
\hat{P}_p (%)		84%	73%	-	73%	-	57%	100%	56%	-	59%	78%	64%	72%	61%	56%		$\hat{P} = 69\%$

- 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	47					3										50	94%
	5		39		2							5			1	5	52	75%
	6			-													-	-
	7				14												14	100%
	1.1					-											-	-
	1.2	8					27		3							5	43	63%
	2.1							29	5		2			15	7	1	59	49%
	2.2						14		30		1			1	1	8	55	55%
	2.3									-							-	-
	3.1										30	7	3		6		46	65%
	3.2		2		5						1	51	1		4		64	80%
	3.3				1							1	10		2		14	71%
	3.4		2				2		20					27	3	9	63	43%
	3.5		4					1	2		6	13		4	47		77	61%
3.6	2	10				6		1							19	38	50%	
Total	57	57	-	22	-	52	30	61	-	40	77	14	47	71	47	575		
\hat{P}_p (%)	82%	68%	-	64%	-	52%	97%	49%	-	75%	66%	71%	57%	66%	40%		$\hat{P} = 64\%$	

- 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	50					7		1						2	60	83%	
	5		45		1					1	8				6	6	67	67%
	6			-													-	-
	7				19												19	100%
	1.1					-											-	-
	1.2	8	2				30		6		1					5	52	58%
	2.1						2	23	5		3			4	3		40	58%
	2.2						9		39					2	2	2	54	72%
	2.3									-							-	-
	3.1				1			1			25	3	5		4		39	64%
	3.2				1							53			3		57	93%
	3.3										2	3	9		6		20	45%
	3.4							3	8					38	3	4	56	7%
	3.5		2						3		8	9			46	2	70	66%
3.6		7				5		2						1	26	41	63%	
Total		58	56	-	22	-	53	27	64	-	40	76	14	44	74	47	575	
\hat{P}_p (%)		86%	80%	-	86%	-	57%	85%	61%	-	63%	70%	64%	86%	62%	55%		$\hat{P} = 70\%$

- 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	52					7		1						2	62	84%	
	5		45		1						1	10			6	6	69	65%
	6			-													-	-
	7				19												19	100%
	1.1					-											-	-
	1.2	7	2				30		6		1					4	50	60%
	2.1						2	23	5		3			4	3		40	58%
	2.2						8		39					2	2	2	53	74%
	2.3									-							-	-
	3.1				1			1			24	2	4		4		36	67%
	3.2				1							51	1		3		56	91%
	3.3										2	3	10		6		21	48%
	3.4						1	3	8					38	3	4	57	67%
	3.5		2						3		8	10			46	2	71	65%
3.6		7				5		2						1	26	41	63%	
Total		59	56	-	22	-	53	27	64	-	39	76	15	44	74	46	575	
\hat{P}_p (%)		88%	80%	-	86%	-	57%	85%	61%	-	62%	67%	67%	86%	62%	57%		$\hat{P} = 70\%$

- 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	51					10										61	84%
	5		45		1							6			4	8	64	70%
	6			-													-	-
	7				18												18	100%
	1.1					-											-	-
	1.2	6					32		2		1			2		6	49	65%
	2.1							26	7		2			4	5	1	45	58%
	2.2						8		41					1	2	3	55	75%
	2.3									-							-	-
	3.1				1						34	6	5		8		54	63%
	3.2				2							48			3		53	91%
	3.3										2	3	4		1		10	40%
	3.4						1	1	9					32	1	2	46	70%
	3.5		2						4		5	14	1	3	48	2	79	61%
3.6	1	9				2		2						1	26	41	63%	
Total	58	56	-	22	-	53	27	65	-	44	77	10	42	73	48	575		
\hat{P}_p (%)	88%	80%	-	82%	-	60%	96%	63%	-	77%	62%	40%	76%	66%	54%		$\hat{P} = 70\%$	

- 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	51					7		1								59	86%	
	5		50		5							8			2	4	69	72%	
	6			-													-	-	
	7				14												14	100%	
	1.1					-											-	-	
	1.2	5					32		3		1				2	6	49	65%	
	2.1							29	5		2				5	3	44	66%	
	2.2	1						9		42					2	1	1	56	75%
	2.3										-							-	-
	3.1				1							34	1	6		7		49	69%
	3.2		1		2							2	57	1		4		67	85%
	3.3													0				0	0%
	3.4								5						25	2	1	33	76%
	3.5								5		6	11		8	52	2		84	62%
3.6	1	6				5		3						1	1	34	51	67%	
Total	58	57	-	22	-	53	29	64	-	45	77	7	41	74	48	575			
\hat{P}_p (%)	88%	88%	-	64%	-	60%	100%	66%	-	76%	74%	0%	61%	70%	71%		$\hat{P} = 73\%$		

- 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 LANDSAT

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	22			1		5							1		2	31	71%
	5		-														-	-
	6			-													-	-
	7				41												41	100%
	1.1					-											-	-
	1.2						49		4						2	3	58	84%
	2.1							23							6		29	79%
	2.2						8		100					6	6	16	136	74%
	2.3									-							-	-
	3.1										44	4	6		1		55	80%
	3.2				5				1		1	33			1		41	80%
	3.3										8	7	13		3		31	42%
	3.4								3					72	1	1	77	94%
	3.5				2		1					2	4	11	81	1	102	79%
	3.6				1				1				2	5	5	40	54	74%
Total		22	-	-	50	-	63	23	109	-	53	46	25	95	106	63	655	
\hat{P}_p (%)		100%	-	-	82%	-	78%	100%	92%	-	83%	72%	52%	76%	76%	63%		$\hat{P} = 79\%$

- 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	18					5										23	78%			
	5		-														-	-			
	6			-													-	-			
	7				42												42	100%			
	1.1					-											-	-			
	1.2						45		8					1	2	13	69	65%			
	2.1							23					1		5		29	79%			
	2.2	1							11			82		1	1	6	10	112	73%		
	2.3											-					-	-			
	3.1												34	1	2		3	40	85%		
	3.2													40	2	1	3	9	58	69%	
	3.3												18	6	16		4	45	36%		
	3.4															91	1	3	109	83%	
	3.5																	73	4	90	81%
3.6																			23	38	61%
Total		19	-	-	50	-	62	23	107	-	53	49	25	104	101	62	655				
\hat{P}_p (%)		95%	-	-	84%	-	73%	100%	77%	-	64%	82%	64%	88%	72%	37%		$\hat{P} = 74\%$			

- 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	18					4		1								23	78%
	5		-														-	-
	6			-													-	-
	7				42												42	100%
	1.1					-											-	-
	1.2						39		1					1	2	2	45	87%
	2.1							22						1	7		30	73%
	2.2	1					14	1	97					11	2	29	155	63%
	2.3									-							-	-
	3.1										33	4	2		8		47	70%
	3.2				3						1	17	1		1		23	74%
	3.3				1						17	12	21		5		56	38%
	3.4								8					66	1	2	77	86%
	3.5				3		1				2	5	3	5	66	4	89	74%
3.6				1				8			5	1	15	12	26	68	38%	
Total	19	-	-	50	-	58	23	115	-	53	43	29	98	104	63	655		
\hat{P}_p (%)	95%	-	-	84%	-	67%	96%	84%	-	62%	40%	72%	67%	63%	41%		$\hat{P} = 68\%$	

- 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	17					4										21	81%	
	5		-														-	-	
	6			-													-	-	
	7				44												44	100%	
	1.1					-											-	-	
	1.2						49		7					1	2	10	69	71%	
	2.1							23						1	3		27	85%	
	2.2				1		8		90			1		8	9	10	127	71%	
	2.3									-							-	-	
	3.1										48	10	5	2	9	1	75	64%	
	3.2				2					1		2	34	1		3	3	46	74%
	3.3											3	3	5			11	45%	
	3.4						1		5					1	61	1	1	70	87%
	3.5				3				3		2	2	6	17	71	3	107	66%	
3.6								2					1	9	6	40	58	69%	
Total		17	-	-	50	-	62	23	108	-	55	50	19	99	104	68	655		
\hat{P}_p (%)		100%	-	-	88%	-	79%	100%	83%	-	87%	68%	26%	62%	68%	59%		$\hat{P} = 74\%$	

- 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	17					4										21	81%
	5		-														-	-
	6			-													-	-
	7				44												44	100%
	1.1					-											-	-
	1.2						49	7						1	2	10	69	71%
	2.1							23							2		25	92%
	2.2				1		8		90			1		8	9	10	127	71%
	2.3									-							-	-
	3.1										46	9	4	1	8	1	69	67%
	3.2				2				1		4	36	1		4	3	51	71%
	3.3										3	3	6				12	50%
	3.4						1		5				1	61	1	1	70	87%
	3.5				3				3		2	2	6	17	68	3	104	65%
3.6								3				1	9	6	44	63	70%	
Total		17	-	-	50	-	62	23	109	-	55	51	19	97	100	72	655	
\hat{P}_p (%)		100%	-	-	88%	-	79%	100%	83%	-	84%	71%	32%	63%	68%	61%		$\hat{P} = 74\%$

- 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	20					4								1		25	80%
	5		-														-	-
	6			-													-	-
	7				44												44	100%
	1.1					-											-	-
	1.2						41		8					1		10	60	68%
	2.1							23				1		2	4		30	77%
	2.2	2					12		88		1	1		18	7	18	147	60%
	2.3									-							-	-
	3.1				2						48	15	5	6	12	5	93	52%
	3.2				1						3	29	2		4		39	74%
	3.3				1						3	4	4	3		2	17	24%
	3.4						2		7				1	51	2	1	64	80%
	3.5				1				3			1	3	16	62	2	88	70%
3.6	1			1		1		2				1	5	5	32	48	67%	
Total	23	-	-	50	-	60	23	108	-	55	51	16	102	97	70	655		
\hat{P}_p (%)	87%	-	-	88%	-	68%	100%	81%	-	87%	57%	25%	50%	64%	46%		$\hat{P} = 67\%$	

- 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	16					9								1		26	62%	
	5		-														-	-	
	6			-													-	-	
	7				44						3	1					2	50	88%
	1.1					-												-	-
	1.2	2					41		2					3	1	23	72	57%	
	2.1						1	23			6	2	6		10		48	48%	
	2.2						12		101			3		7	9	19	151	67%	
	2.3									-							-	-	
	3.1										0						0	0%	
	3.2										2	1	1				4	25%	
	3.3												0				0	0%	
	3.4				1				5					85	2	5	98	87%	
	3.5				4				1		37	39	15	5	92	5	198	46%	
3.6				1							1			1	5	8	63%		
Total		18	-	-	50	-	63	23	109	-	48	47	22	100	116	59	655		
\hat{P}_p (%)		89%	-	-	88%	-	65%	100%	93%	-	0%	2%	0%	85%	79%	8%		$\hat{P} = 62\%$	

- 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 LANDSAT

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	32					7		12							51	63%	
	5		-													-	-	
	6			38	7			7			1					53	72%	
	7			3	50											53	94%	
	1.1					-										-	-	
	1.2	1			1		36		4							42	86%	
	2.1			1				33		5	2					41	80%	
	2.2						8			51				5	2	66	77%	
	2.3			2				1			35					38	92%	
	3.1							1				35	2		1	4	43	81%
	3.2										2	2	40			1	45	89%
	3.3													-			-	-
	3.4								5	1	1				48	8	63	76%
	3.5										1	1		3		50	55	91%
3.6																-	-	
Total		33	-	44	58	-	51	42	72	43	42	43	-	57	65	-	550	
\hat{P}_p (%)		97%	-	86%	86%	-	71%	79%	71%	81%	83%	93%	-	84%	77%	-		$\hat{P} = 81\%$

- 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	34					1		3								38	89%
	5		-														-	-
	6			27						7							34	79%
	7			5	63												68	93%
	1.1					-											-	-
	1.2						37		14								51	73%
	2.1			4			1	42		7	4						58	72%
	2.2						9		45					1			55	82%
	2.3			2						26	1						29	90%
	3.1						1				29	1		1	1		33	88%
	3.2				2					1	6	41			1		51	80%
	3.3												-				-	-
	3.4								7	1	1			56	8		73	77%
	3.5						1	1	1		1	1		2	53		60	88%
3.6															-	-	-	
Total		34	-	38	65	-	50	43	70	42	42	43	-	60	63	-	550	
\hat{P}_p (%)		100%	-	71%	97%	-	74%	98%	64%	62%	69%	95%	-	93%	84%	-		$\hat{P} = 82\%$

- 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		
Classificação	4	35					5		9							49	71%
	5		-													-	-
	6			21	3					1	2					27	78%
	7			4	57											61	93%
	1.1					-										-	-
	1.2						16		1							17	94%
	2.1			3				39		5	1					48	81%
	2.2				1		27		58					2		88	66%
	2.3			9				4		35	1	1				50	70%
	3.1			2							29	6		1	2	40	73%
	3.2			1	1						9	34			1	46	74%
	3.3												-			-	-
	3.4								5	1	1			52	12	71	73%
	3.5						1				1	2		5	44	53	83%
3.6															-	-	
Total		35	-	40	62	-	49	43	73	42	44	43	-	60	59	-	550
\hat{P}_p (%)		100%	-	53%	92%	-	33%	91%	79%	83%	66%	79%	-	87%	75%	-	$\hat{P} = 76\%$

- 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	33					2		6						1		42	79%
	5		-														-	-
	6			28	5						1						34	82%
	7			4	53												57	93%
	1.1					-											-	-
	1.2	1					42		22		1			1			67	63%
	2.1			1				41		6	1						49	84%
	2.2						5	1	37								43	86%
	2.3			7	1					34	1						43	79%
	3.1							1			29	5		1	5		41	71%
	3.2			2	2					1	7	37			3		52	71%
	3.3												-				-	-
	3.4								7	1				49	9		66	74%
	3.5						1		1		1	1		6	46		56	82%
3.6															-	-	-	
Total		34	-	42	61	-	50	43	73	42	41	43	-	57	64	-	550	
\hat{P}_p (%)		97%	-	67%	87%	-	84%	95%	51%	81%	71%	86%	-	86%	72%	-		$\hat{P} = 78\%$

- 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	33					2		6						1		42	79%
	5		-														-	-
	6			28	5						1						34	82%
	7			4	53												57	93%
	1.1					-											-	-
	1.2	1					42		22		1			1			67	63%
	2.1			1				41		6	1						49	84%
	2.2						5	1	36								42	86%
	2.3			7	1					34	1						43	79%
	3.1							1			30	5		1	5		42	71%
	3.2			2	2					1	6	37			3		51	73%
	3.3												-				-	-
	3.4								7	1					50	9	67	75%
	3.5						1		1		1	1		6	46		56	82%
3.6																-	-	
Total		34	-	42	61	-	50	43	72	42	41	43	-	58	64	-	550	
\hat{P}_p (%)		97%	-	67%	87%	-	84%	95%	50%	81%	73%	86%	-	86%	72%	-		$\hat{P} = 78\%$

- 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	32					2		6					1			41	78%
	5		-														-	-
	6			33	4						1						38	87%
	7			3	52												55	95%
	1.1					-											-	-
	1.2	2					41		28						2		73	56%
	2.1			1				39		6	2						48	81%
	2.2						6	2	36								44	82%
	2.3			5	2					34							41	83%
	3.1						1	1			30	2			9		43	70%
	3.2				2					2	8	38		1	2		53	72%
	3.3												-				-	-
	3.4								3	1	1			48	10		63	76%
	3.5										1	3		7	40		51	78%
3.6															-	-	-	
Total		34	-	42	60	-	50	42	73	43	43	43	-	59	61	-	550	
\hat{P}_p (%)		94%	-	79%	87%	-	82%	93%	49%	79%	70%	88%	-	81%	66%	-		$\hat{P} = 77\%$

- 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	30					2		6								38	79%
	5		-														-	-
	6			0													0	0%
	7			14	67												81	83%
	1.1					-											-	-
	1.2	2			1		44		15							1	63	70%
	2.1			2			1	39		7	2			1			52	75%
	2.2						5		50		1			6	2		64	78%
	2.3			12						31	1	1					45	69%
	3.1			2				5			34	1			1		43	79%
	3.2			3	1					2	4	40				1	51	78%
	3.3												-				-	-
	3.4								4	1				45	6		56	80%
	3.5										1	1		3	52		57	91%
3.6																-	-	
Total		32	-	33	69	-	52	44	75	41	43	43	-	55	63	-	550	
\hat{P}_p (%)		94%	-	0%	97%	-	85%	89%	67%	76%	79%	93%	-	82%	83%	-		$\hat{P} = 79\%$

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