

# **Comparação da performance de vários algoritmos na classificação automática da ocupação/uso do solo em imagens de satélite de distintas resoluções espaciais**

**Rita Nicolau, Joel Dinis,  
Pedro Rodrigues, Márcia Gonçalves, Vasco Nunes**

**Projeto LANDAU**

**Metodologia geoestatística para a caracterização da incerteza espacial e  
atualização de cartas de ocupação do solo  
(PTDC/CTE-SPA/103872/2008 )**



# Tarefas da responsabilidade do IGP/DGT

1. Seleção de três áreas de estudo representativas da diversidade paisagística e da ocupação/uso do solo (LULC) do Continente.
2. Proposta de uma nomenclatura para classificação de LULC, que seja funcional a diferentes resoluções espaciais e que seja compatível com a cartografia nacional já disponível.
3. Produção de cartografia LULC para as áreas de estudo selecionadas, através da classificação automática de imagens de distintas resoluções espaciais com diversos algoritmos
  - a. Validação da cartografia produzida.
  - b. Comparação de resultados.

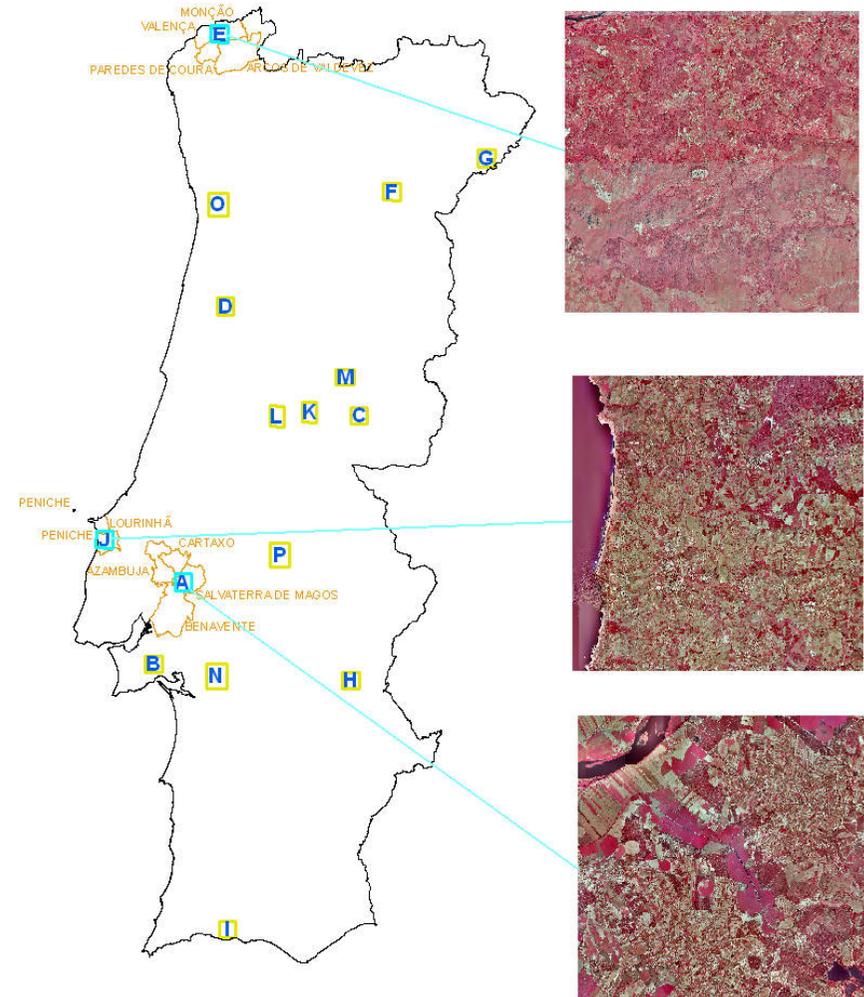
# 1. Seleção de três áreas de estudo

As áreas de estudo deveriam retratar os diferentes padrões de LULC do Continente, respeitando a variabilidade da paisagem e a distinta fragmentação existente entre o norte e o sul.

- Com base na disponibilidade de imagens de muito alta resolução (QuickBird) existentes no IGP, selecionaram-se 16 áreas de estudo candidatas.
- Através da sobreposição destas áreas com a COS 2007, calcularam-se os seguintes parâmetros por área candidata:
  - Nº de classes de nível 3 da COS 2007 por hectare;
  - Mediana da área dos polígonos de nível 3 da COS 2007 (em ha);
  - Nº de classes de nível 3 da COS 2007 partilhadas entre áreas candidatas.

# 1. Seleção de três áreas de estudo

- Desenvolvimento de um indicador multicritério que expressa a soma ponderada das pontuações de cada área candidata nos parâmetros anteriormente listados.
- Comparação do valor do indicador multicritério entre áreas candidatas e seleção das áreas de estudo E, J e A.



0 26.000 52.000 104.000 156.000 208.000 Meters

## 2. Proposta de nomenclatura para classificação da ocupação/uso do solo

A nomenclatura deveria ser:

- funcional a diferentes resoluções espaciais;
  - compatível com os diversos produtos cartográficos nacionais existentes (COS e CLC);
  - compatível com o “Land Cover Classification System - LCCS”.
- Com recurso a dados de satélite de média resolução espacial (MERIS) procedeu-se à análise da separabilidade espectral de classes de LULC, associadas aos níveis 2 e 3 da cartografia CLC 2006:
    - esta análise foi desenvolvida a partir do cálculo da distância de Jeffries-Matusita (Richards & Jia, 2006, pag. 273), que reflete a probabilidade de duas classes serem confundidas na classificação.

## 2. Proposta de nomenclatura para classificação da ocupação/uso do solo

- Com recurso a dados de satélite de muito elevada resolução espacial (QuickBird) e a dois classificadores (LDC - *Linear Discriminant Classifier* e o QDC - *Quadratic Discriminant Classifier*), produziram-se mapas de LULC para cada uma das três áreas de estudo do projeto por forma a:
  - detetar possíveis problemas de classificação que não têm representatividade à escala da MERIS, mas que são representativos à escala das QuickBird.
- Com base nas evidências produzidas pela análise da separabilidade espectral (média resolução) e pelos mapas preliminares (muito elevada resolução), procedeu-se por fim à proposta da nomenclatura de ocupação/uso do solo.

## 2. Proposta de nomenclatura para classificação da ocupação/uso do solo

### Nomenclatura LANDAU

- A nomenclatura LANDAU é hierárquica, sendo composta por dois níveis.
- Na sua versão mais desagregada inclui 15 classes de ocupação/uso do solo.
- Trata-se de uma nomenclatura muito similar à nomenclatura GlobCorine

(Defourny & Bontemps & Bogaert, 2010).

1. Territórios Artificializados
  - 1.1 Áreas Artificiais Contínuas
  - 1.2 Áreas Artificiais Descontínuas
2. Áreas Agrícolas
  - 2.1 Agricultura de Regadio
  - 2.2 Agricultura de Sequeiro
  - 2.3 Arrozais
3. Florestas e Meios Naturais e Seminaturais
  - 3.1 Floresta de Folhosas
  - 3.2 Floresta de Resinosas
  - 3.3 Floresta Mista
  - 3.4 Vegetação Herbácea
  - 3.5 Matos
  - 3.6 Vegetação Esparsa
4. Solo Nu
5. Áreas Ardidas
6. Zonas Húmidas
7. Corpos de Água

### 3. Produção de cartografia LULC

- A classificação automática da ocupação/uso do solo com base na nomenclatura LANDAU, foi independentemente desenvolvida para cada área de estudo a partir de imagens de:
  - Média resolução espacial (300x300 m)-> dados MERIS de 2005 (superfície classificada = Portugal Continental);
  - Resolução espacial intermédia (30x30 m)-> dados LANDSAT de 2007 (superfície classificada = 30 272 ha);
  - Muito elevada resolução espacial (2.4x2.4 m) -> dados QUICKBIRD de 2004-2005 (superfície classificada = 30 266 ha).
- Para cada resolução espacial e área de estudo ensaiaram-se sete algoritmos de classificação:
  - 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).

### 3. Produção de cartografia LULC

A recolha de amostras destinadas ao treino dos algoritmos e à validação da classificação foi desenvolvida independentemente para cada escala de análise e área de estudo

- As **amostras de treino** foram deterministicamente recolhidas sobre a cartografia CORINE Land Cover (2006), na sequência da sua reclassificação em classes de nível 2 da nomenclatura LANDAU.
- Para as **amostras de validação** utilizou-se uma amostragem aleatória, estratificada pelas classes de nível 2 da nomenclatura LANDAU, que à escala de análise adotada tinham representação em cada área de estudo (Exemplo: à escala da QUICKBIRD não se amostraram as classes Áreas Artificiais Contínuas e Floresta Mista).

Dimensão das amostras (nº pontos)	Escala de análise		
	MERIS	LANDSAT	QUICKBIRD
Treino	11957	10882	28340
Validação	750	1780	1571

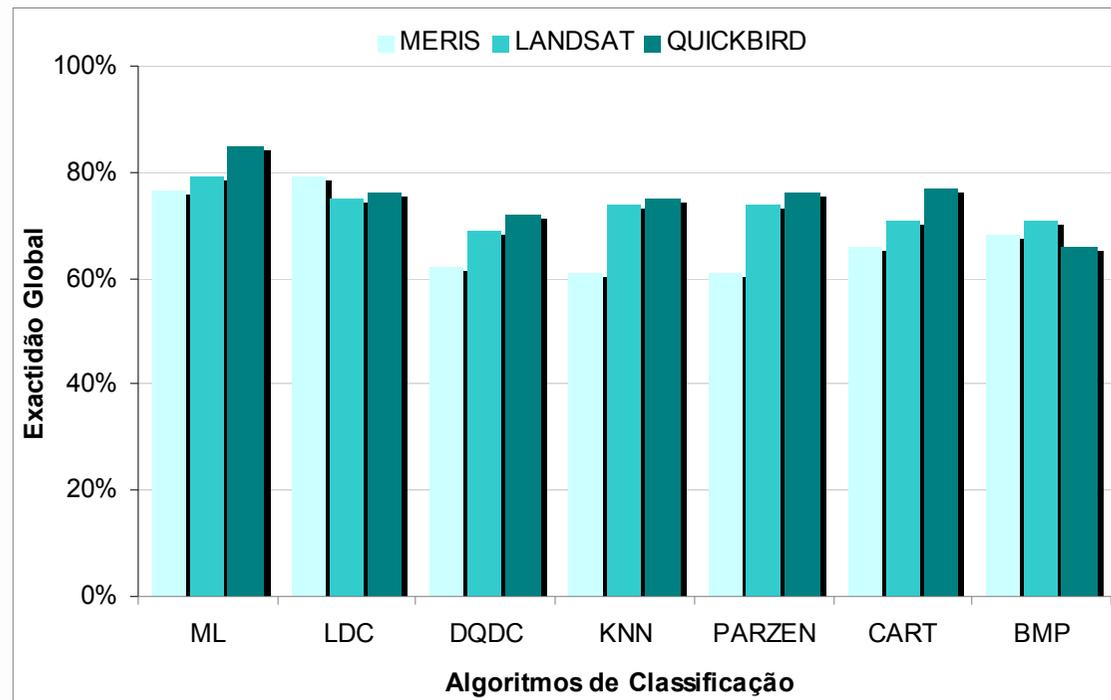
### 3. Produção de cartografia LULC

- As medidas utilizadas para avaliação do desempenho dos algoritmos testados na classificação do uso do solo foram a exatidão global ( $\hat{P}$ ), a exatidão do produtor ( $\hat{P}_p$ ) e a exatidão do utilizador ( $\hat{P}_u$ ).
- Na comparação do desempenho dos classificadores a cada escala de análise e entre escalas de análise utilizou-se apenas a exatidão global ( $\hat{P}$ ), tendo sido desenvolvidos testes para avaliar a significância estatística das diferenças:
  - entre exatidões obtidas por diferentes algoritmos à mesma escala de análise (tendo em conta a dependência das amostras empregues, utilizou-se o teste McNemar com correção de continuidade\*);
  - entre exatidões obtidas pelos algoritmos a diferentes escalas de análise (utilizou-se o teste para diferenças entre duas proporções resultantes de amostras independentes\*).

### 3. Produção de cartografia LULC

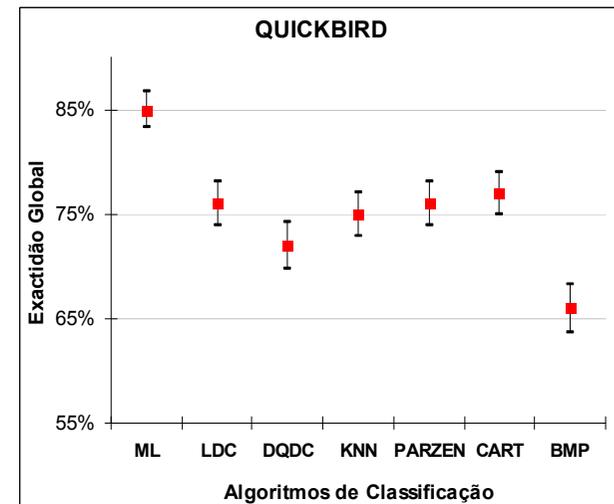
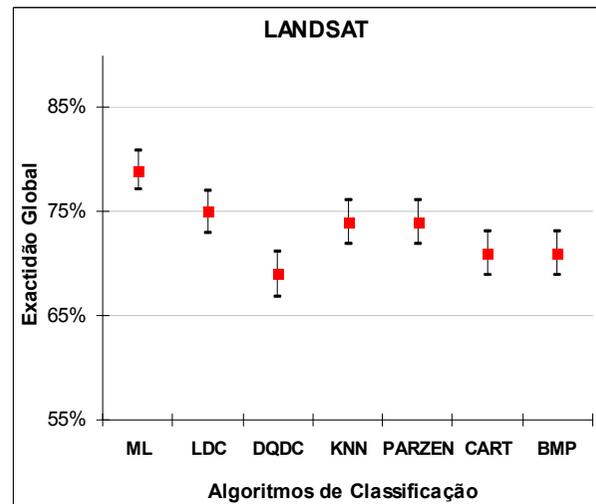
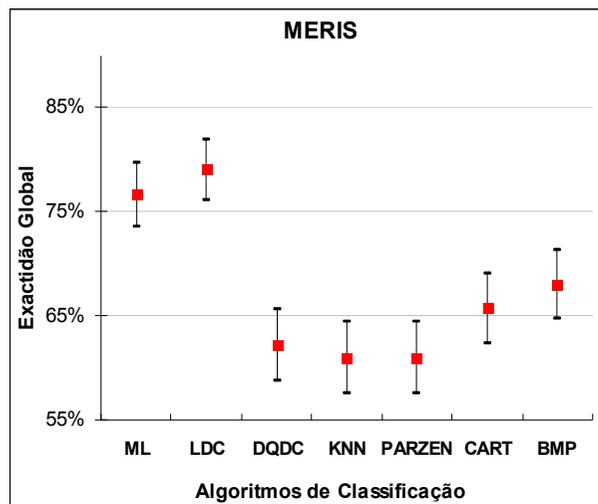
Desempenho global dos algoritmos de classificação a diferentes escalas de análise:

- Para a maioria dos algoritmos a exatidão global cresceu com o nível de detalhe dos dados de satélite a classificar.
- O algoritmo ML destacou-se pelo melhor desempenho global na classificação do uso do solo, embora a exatidão disponibilizada pelo ML tenha sido superada pela obtida pelo algoritmo LDC à escala da MERIS.



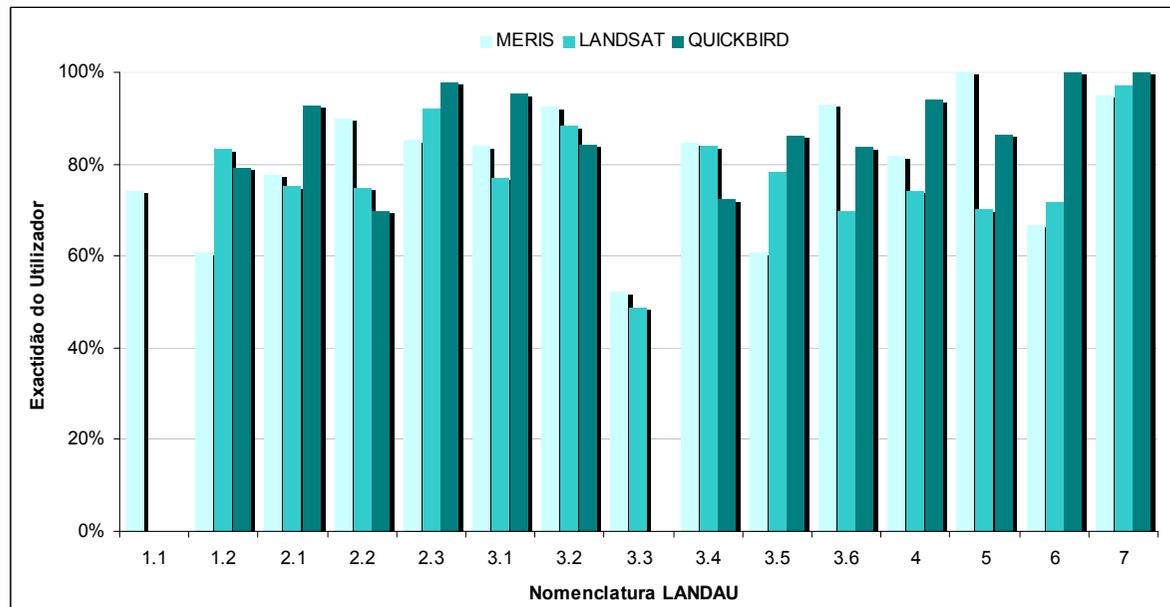
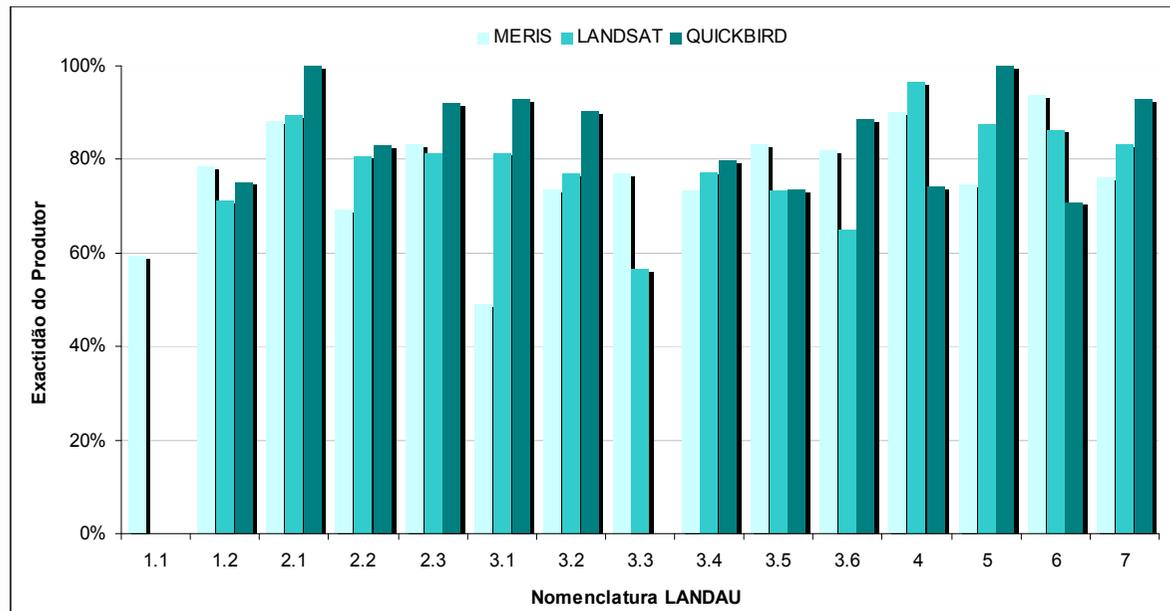
### 3. Produção de cartografia LULC

- À escala da MERIS, as exatidões globais disponibilizadas pelo ML e pelo LDC foram superiores às disponibilizadas pelo restantes classificadores. Contudo não se comprovou que as exatidões globais proporcionadas pelo ML e pelo LDC fossem significativamente diferentes entre si.
- Tanto à escala da LANDSAT, como à escala da QUICKBIRD, comprovou-se que a exatidão global obtida pelo ML é significativamente diferente das exatidões globais obtidas pelos restantes classificadores.
- o algoritmo ML teve melhor desempenho global na classificação do uso do solo por recurso a dados QUICKBIRD.



### 3. Produção de cartografia LULC

Desempenho do algoritmo ML na identificação de classes da nomenclatura LANDAU a diferentes escalas de análise



### 3. Produção de cartografia LULC

Exatidões do produtor obtidas pelo ML na identificação de classes LULC a diferentes escalas de análise:

- As exatidões do produtor foram na sua maioria mais elevadas quando se utilizaram imagens QUICKBIRD.
- A imagem MERIS permitiu obter maiores exatidões do produtor na classificação de Territórios Artificializados (1.1 e 1.2), Floresta Mista (3.3), Matos (3.5) e Zonas Húmidas (6).
- As imagens LANDSAT apenas proporcionaram uma maior exatidão do produtor na classificação do Solo Nu (4).

### 3. Produção de cartografia LULC

Exatidões do utilizador obtidas pelo ML na identificação de classes LULC a diferentes escalas de análise:

- As imagens QUICKBIRD permitiram a obtenção de maiores exatidões do utilizador na classificação de Agricultura de Regadio (2.1), Arrozais (2.3), Florestas de Folhosas (3.1), Matos (3.5), Solo Nu (4), Zonas Húmidas (6) e Corpos de Água (7).
- A imagem MERIS permitiu obter maiores exatidões do utilizador na classificação de Agricultura de Sequeiro (2.2), Floresta de Resinosas (3.2), Floresta Mista (3.3), Vegetação Herbácea (3.4), Vegetação Esparsa (3.6) e Áreas Ardidadas (5).
- As imagens LANDSAT apenas proporcionaram uma maior exatidão do utilizador na classificação de Áreas Artificiais Descontínuas (1.2).

# Conclusões

- Os resultados destacam o algoritmo ML pelo melhor desempenho global na classificação do uso do solo e demonstram que à medida que a resolução espacial aumenta, o ML revela ganhos na qualidade da classificação.
- Embora o desempenho global do algoritmo LDC não tenha sido significativamente diferente do do ML à escala de menor de detalhe (MERIS), o LDC evidenciou um decréscimo de desempenho na classificação com o aumento da resolução espacial.
  - na comparação dos resultados obtidos à escala da QUICKBIRD com os obtidos à escala da MERIS, observou-se um ganho de qualidade dos classificadores que definem fronteiras mais complexas (ML, CART, KNN, PARZEN) e uma perda de qualidade dos que definem fronteiras menos complexas (LDC, BMP).

# Conclusões

- O presente estudo demonstrou uma vantagem dos classificadores mais simples (ML e LDC) sobre classificadores mais sofisticados (CART, KNN, BMP)
  - este resultado é consistente com o descrito por alguns investigadores (Wilkinson, 2005; Hand, 2006) que referem que o aparente sucesso no desenvolvimento de algoritmos de classificação cada vez mais sofisticados não se traduz em sucesso na qualidade dos resultados obtidos, particularmente quando tais classificadores são utilizados no processamento de dados reais.

# Conclusões

No que diz respeito à escala de análise mais ajustada à classificação do uso do solo com base na nomenclatura LANDAU, as exatidões do utilizador e do produtor obtidas no presente estudo apontam para que se empreguem:

- imagens de menor detalhe (MERIS) na classificação de classes com assinaturas espectrais complexas tais como as Áreas Artificiais Contínuas (1.1) e a Floresta Mista (3.3);
- imagens LANDSAT (elevada resolução espacial) na classificação de Áreas Artificiais Descontínuas (1.2);
- imagens de muito elevada resolução espacial (QUICKBIRD) na classificação da Agricultura de Regadio (2.1), da Floresta de Folhosas (3.1), da Vegetação Esparsa (3.6) e dos Corpos de Água (7).

Obrigada pela sua atenção

**Medidas de exatidão temática dos algoritmos de classificação testados a partir de dados MERIS**  
(área de estudo: Portugal Continental)

$\hat{P}_p$  = Exatidão do Produtor

$\hat{P}_u$  = Exatidão do Utilizador

	ML		LDC		DQDC		KNN		PARZEN		CART		BMP	
	$\hat{P}_p$	$\hat{P}_u$												
<b>1.1 Áreas Artificiais Contínuas</b>	59%	74%	56%	61%	40%	47%	61%	35%	35%	55%	46%	64%	42%	60%
<b>1.2 Áreas Artificiais Descontínuas</b>	78%	61%	58%	55%	42%	34%	52%	80%	73%	44%	70%	44%	62%	50%
<b>2.1 Agricultura de Regadio</b>	88%	78%	88%	98%	78%	89%	89%	76%	71%	88%	92%	87%	84%	81%
<b>2.2 Agricultura de Sequeiro</b>	69%	90%	78%	90%	65%	87%	74%	60%	52%	81%	63%	79%	69%	81%
<b>2.3 Arrozais</b>	83%	85%	71%	85%	60%	69%	78%	51%	57%	67%	55%	82%	57%	88%
<b>3.1 Floresta de Folhosas</b>	49%	84%	82%	88%	54%	71%	72%	47%	47%	67%	59%	68%	74%	81%
<b>3.2 Floresta de Resinosas</b>	73%	92%	92%	71%	89%	64%	62%	75%	88%	63%	70%	69%	83%	80%
<b>3.3 Floresta Mista</b>	77%	52%	80%	69%	45%	38%	46%	63%	38%	33%	60%	48%	66%	63%
<b>3.4 Vegetação Herbácea</b>	73%	85%	92%	63%	67%	43%	52%	60%	75%	59%	60%	55%	77%	69%
<b>3.5 Matos</b>	83%	61%	81%	71%	43%	68%	38%	89%	77%	58%	83%	42%	94%	37%
<b>3.6 Vegetação Esparsa</b>	82%	93%	90%	96%	66%	56%	55%	46%	66%	49%	44%	65%	54%	82%
<b>4 Solo Nu</b>	90%	82%	64%	91%	64%	74%	77%	46%	48%	71%	54%	90%	26%	87%
<b>5 Áreas Áridas</b>	75%	100%	88%	85%	78%	74%	83%	65%	65%	79%	82%	95%	75%	100%
<b>6 Zonas Húmidas</b>	94%	67%	92%	90%	67%	60%	59%	65%	65%	55%	73%	75%	86%	68%
<b>7 Corpos de Água</b>	76%	95%	72%	97%	72%	95%	90%	52%	50%	89%	72%	90%	64%	82%
$\hat{p}$ = Exatidão Global	77%		79%		62%		61%		61%		66%		68%	

**Medidas de exatidão temática dos algoritmos de classificação testados a partir de dados LANDSAT**  
(área de estudo: E, J, A)

$\hat{p}_p$  = Exatidão do Produtor

$\hat{p}_u$  = Exatidão do Utilizador

	ML		LDC		DQDC		KNN		PARZEN		CART		BMP	
	$\hat{P}_p$	$\hat{P}_u$												
<b>1.1 Áreas Artificiais Contínuas</b>	-	-	-	-	-	-	-	-	-	-	-	-	-	-
<b>1.2 Áreas Artificiais Descontínuas</b>	71%	83%	68%	68%	52%	78%	73%	64%	73%	65%	70%	63%	70%	64%
<b>2.1 Agricultura de Regadio</b>	89%	75%	99%	69%	94%	66%	94%	75%	94%	76%	96%	72%	95%	63%
<b>2.2 Agricultura de Sequeiro</b>	81%	75%	68%	74%	74%	62%	68%	74%	67%	74%	67%	67%	78%	71%
<b>2.3 Arrozais</b>	81%	92%	62%	90%	83%	70%	81%	79%	81%	79%	79%	83%	76%	69%
<b>3.1 Floresta de Folhosas</b>	81%	77%	64%	80%	67%	69%	75%	66%	74%	68%	79%	59%	50%	74%
<b>3.2 Floresta de Resinosas</b>	77%	88%	83%	77%	63%	77%	73%	80%	73%	78%	67%	79%	59%	80%
<b>3.3 Floresta Mista</b>	56%	49%	64%	40%	72%	44%	42%	45%	47%	48%	31%	30%	0%	0%
<b>3.4 Vegetação Herbácea</b>	77%	84%	86%	77%	71%	69%	74%	77%	75%	77%	65%	76%	79%	83%
<b>3.5 Matos</b>	73%	78%	72%	74%	67%	72%	67%	70%	67%	69%	65%	69%	77%	58%
<b>3.6 Vegetação Esparsa</b>	65%	70%	45%	58%	41%	42%	57%	67%	59%	67%	49%	65%	36%	66%
<b>4 Solo Nu</b>	96%	74%	91%	88%	90%	82%	92%	81%	93%	82%	90%	81%	90%	79%
<b>5 Áreas Áridas</b>	88%	70%	73%	76%	68%	75%	80%	67%	80%	65%	80%	70%	88%	72%
<b>6 Zonas Húmidas</b>	86%	72%	71%	79%	53%	78%	67%	82%	67%	82%	79%	87%	0%	0%
<b>7 Corpos de Água</b>	83%	97%	88%	96%	84%	97%	87%	97%	87%	97%	86%	97%	89%	86%
$\hat{p}$ = Exatidão Global	79%		75%		69%		74%		74%		71%		71%	

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

**Medidas de exatidão temática dos algoritmos de classificação testados a partir de dados QUICKBIRD**  
(área de estudo: E, J, A)

$\hat{p}_p$  = Exatidão do Produtor

$\hat{p}_u$  = Exatidão do Utilizador

	ML		LDC		DQDC		KNN		PARZEN		CART		BMP	
	$\hat{P}_p$	$\hat{P}_u$												
<b>1.1 Áreas Artificiais Contínuas</b>	-	-	-	-	-	-	-	-	-	-	-	-	-	-
<b>1.2 Áreas Artificiais Descontínuas</b>	75%	79%	50%	61%	31%	44%	76%	50%	68%	51%	79%	55%	56%	39%
<b>2.1 Agricultura de Regadio</b>	100%	93%	97%	92%	99%	90%	97%	90%	97%	90%	99%	93%	98%	94%
<b>2.2 Agricultura de Sequeiro</b>	83%	70%	73%	67%	78%	62%	82%	63%	81%	66%	79%	66%	76%	44%
<b>2.3 Arrozais</b>	92%	98%	92%	88%	92%	94%	68%	89%	68%	89%	74%	95%	52%	81%
<b>3.1 Floresta de Folhosas</b>	93%	95%	89%	79%	85%	81%	86%	85%	86%	85%	88%	85%	93%	90%
<b>3.2 Floresta de Resinosas</b>	91%	84%	69%	74%	62%	69%	75%	73%	76%	74%	74%	75%	66%	74%
<b>3.3 Floresta Mista</b>	-	-	-	-	-	-	-	-	-	-	-	-	-	-
<b>3.4 Vegetação Herbácea</b>	80%	72%	78%	60%	77%	59%	72%	73%	76%	73%	70%	73%	47%	54%
<b>3.5 Matos</b>	74%	86%	56%	71%	61%	71%	59%	76%	62%	78%	64%	76%	62%	71%
<b>3.6 Vegetação Esparsa</b>	88%	84%	83%	79%	62%	69%	62%	87%	66%	85%	81%	85%	0%	0%
<b>4 Solo Nu</b>	74%	94%	72%	92%	68%	91%	59%	92%	64%	82%	63%	91%	59%	95%
<b>5 Áreas Áridas</b>	100%	86%	89%	77%	79%	79%	74%	70%	79%	71%	84%	84%	89%	74%
<b>6 Zonas Húmidas</b>	71%	100%	73%	83%	59%	67%	57%	100%	57%	100%	48%	95%	2%	100%
<b>7 Corpos de Água</b>	93%	100%	96%	93%	94%	97%	93%	98%	93%	98%	95%	98%	94%	92%
$\hat{P}$ = Exatidão Global	85%		76%		72%		75%		76%		77%		66%	

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