



Universidade de Coimbra
Departamento de Engenharia Civil

Integração da incerteza na classificação e avaliação da exactidão temática de imagens multiespectrais

Aplicação à avaliação do estado de conservação do património edificado da Baixa de Coimbra

Luísa Maria da Silva Gonçalves

Tese apresentada para a obtenção do grau de Doutor em Engenharia Civil
na Especialidade de Urbanismo, Ordenamento do Território e Transportes
pela Universidade de Coimbra

Coimbra 2009



Universidade de Coimbra
Departamento de Engenharia Civil

Integração da incerteza na classificação e avaliação da exactidão temática de imagens multiespectrais

Aplicação à avaliação do estado de conservação do património edificado da Baixa de Coimbra

Luísa Maria da Silva Gonçalves

Doutoramento em Engenharia Civil
Especialização em Urbanismo, Ordenamento do Território e Transportes

Orientador: Eduardo Nuno Brito Santos Júlio
(Professor Auxiliar c/ Agregação da FCTUC)

Orientador: Mário Sílvio Rochinha de Andrade Caetano
(Professor Auxiliar Convidado do ISEGI UNL)

Tese apresentada para a obtenção do grau de Doutor em Engenharia Civil
na Especialidade de Urbanismo, Ordenamento do Território e Transportes

pela Universidade de Coimbra

Coimbra 2009

Resumo

Esta tese pretende avaliar a vantagem da integração da incerteza no processo de produção de informação temática, a partir de imagens multiespectrais de muito grande resolução espacial. Com este objectivo, a informação da incerteza é aplicada nas seguintes fases do processo de produção: (1) extracção e classificação de informação temática; (2) avaliação do comportamento do classificador; (3) avaliação da exactidão temática.

A abordagem de classificação, que combina a análise de padrões espectrais ao nível do *pixel* com a análise de padrões espaciais ao nível dos objectos, tem mostrado ser adequada para a identificação das classes de unidade de paisagem que contêm uma variedade de coberturas de ocupação do solo. Neste estudo, é desenvolvida e testada uma abordagem metodológica que integra, no processo de classificação combinada *pixel/objecto*, a informação da incerteza. O novo método desenvolvido envolve uma primeira classificação, ao nível do *pixel*, para extrair os elementos de superfície (e.g. copas de sobreiro, edifícios). Nesta primeira fase utiliza-se um classificador não rígido, o que permite obter os graus de atribuição das classes aos *pixels* e o cálculo de incerteza dessa atribuição. Num segundo passo, procede-se a uma outra classificação, mas ao nível do objecto, para identificar as unidades de paisagem (e.g. Floresta de Sobreiro, Área Urbana). A informação sobre a incerteza é aplicada na classificação das unidades de paisagem através de um conjunto de regras de decisão que consideram: os elementos de superfície classificados existentes dentro de cada objecto e o grau de incerteza associado a essa atribuição. Com a aplicação desta abordagem metodológica pretende-se avaliar se a integração da incerteza associada à classificação dos elementos de superfície pode melhorar a classificação das unidades de paisagem.

Os mapas temáticos resultantes do processo de classificação de imagens multiespectrais são usualmente sujeitos a um processo de validação para que seja atribuído um índice de fidelidade aos resultados obtidos. Neste estudo, é investigada a utilidade da aplicação das medidas de incerteza na avaliação da classificação de imagens de detecção remota. Em particular, analisa-se se as medidas de incerteza de não-especificidade e entropia podem ser usadas para avaliar as dificuldades do classificador e estimar a exactidão da classificação. Para dar resposta a este objectivo, desenvolveram-se dois índices de incerteza e avaliou-se a correlação existente entre a

informação obtida com estes dois índices e os índices de exactidão estatísticos do produtor e do utilizador.

A detecção remota tem demonstrado ser uma fonte de informação adequada em várias áreas, nomeadamente na monitorização dos recursos naturais e na produção de cartografia de ocupação do solo para gestão e planeamento urbano e florestal. Com o aumento contínuo da resolução espacial das imagens de satélite, abrem-se horizontes a novas aplicações. Dada a importância crescente que tem vindo a ser atribuída à conservação do património edificado, que engloba não só os monumentos mas também as construções dos centros históricos, considerou-se relevante abordar esta temática no âmbito deste trabalho. Assim, investiga-se o potencial das imagens de muito grande resolução espacial para extrair informação de suporte à avaliação do estado de conservação das coberturas do espaço edificado e testa-se a nova metodologia de classificação desenvolvida. A área escolhida para teste e implementação do estudo situa-se na Baixa de Coimbra. Para a sua realização foram utilizadas imagens aéreas com uma resolução espacial de 0.5 m e imagens do satélite Quickbird com uma resolução de 2.4 m. Os resultados obtidos evidenciaram que a grande resolução espacial das imagens multiespectrais utilizadas, embora não permita identificar os diferentes tipos de anomalias, permite uma boa identificação da presença de anomalias não estruturais das coberturas e dos seus diferentes materiais de revestimento. A aplicação da metodologia desenvolvida permitiu obter um mapa de anomalias do edificado com uma Exactidão Global de 78%, tendo como informação de referência os dados de campo resultantes do levantamento efectuado pela Universidade de Coimbra.

A integração da informação sobre a incerteza demonstrou ser importante no processo de classificação combinada, contribuindo para melhorar significativamente os resultados finais da classificação. A aplicação das medidas de incerteza na avaliação da classificação revelou uma correlação positiva entre a informação dada pelas medidas de incerteza e os índices estatísticos de exactidão, principalmente com a exactidão do utilizador. Os resultados obtidos demonstraram que a aplicação das medidas de incerteza, embora não substitua os índices de exactidão, pode ser utilizada como indicador do comportamento do classificador e, consequentemente, dos resultados finais da classificação.

Abstract

The aim of this thesis is to evaluate the usefulness of the uncertainty integration in the production of thematic information using Very High Spatial Resolution (VHSR) images. To achieve this goal, the uncertainty information is integrated in the following steps of the production process: (1) extraction and classification of thematic information; (2) evaluation of the thematic accuracy.

The hybrid classification approach that combines the spectral patterns analysis, at pixel bases, with spatial patterns analysis, at object bases, has shown to be suitable for the identification of land unit classes that contain a variety of land coverage elements. In this study, a new classification approach was developed and tested, which incorporates uncertainty information in the pixel/object hybrid classification. This new classification method involves a first classification, where the spatial unit is the pixel, to obtain the surface elements (e.g. crown trees, parts of buildings). At this stage, the pixel-based classification is performed with a soft classifier, which allows the calculation and the assignment to each pixel of different degrees of attribution to the several classes and allows the assessment of the uncertainty of this assignment. A second classification is performed to identify the land units (e.g, Crock Forest, Urban Area), where the spatial unit is the object. The uncertainty information was used in the classification of the land units, considering a set of decision rules that incorporate the arrangement of the surface elements classification within each object and their degree of uncertainty obtained by the uncertainty measures. The main goal of this new approach is to evaluate if the uncertainty information associated with the classification of surface elements into the classification of landscape units can improve the final accuracy of classification results.

Producing thematic maps, using multispectral classification images process, requires the assessment of their accuracy since the decision of whether a particular map is fit for use within a certain application is based on the results of the accuracy assessment. This study investigates the usefulness of the uncertainty measures to analyse the thematic classification results of remote sensing images. In particular, it is analysed if the entropy and the non-specificity uncertainty measures can be used to identify the difficulties found by the classifier and to estimate the classification accuracy. Two uncertainty indexes are proposed, and the correlation

between the information given by them and the user's and producer's accuracy is determined, to assess the relation and compatibility of both sources of information.

Remote sensing is an important source of information in several areas, such as the monitoring of natural resources and the production of land cover maps for urban and forest management and planning. With the continuous improvement of the spatial resolution of remote sensing sensors, a broad number of new applications are possible. Due to the growing importance of heritage conservation, which includes not only the monuments, but also the construction of historic centres, it was considered important to address this issue in this thesis. This study intends to evaluate the applicability of VHRS multispectral images, to extract information concerning the state of conservation of building roofs and to test the new hybrid classification approach. The case-study area selected was the historical city centre of Coimbra. The test was carried out with a multi-spectral aerial image and a QuickBird satellite image with spatial resolutions of respectively 0.5 m and 2.4 m. The obtained results showed that even though the very high spatial resolution of the images did not allowed the identification of different types of pathologies they allow a good identification of the existence of roofing pathologies and the identification of their different materials. The hybrid pixel-object classification methodology developed allowed the production of a building pathology map with a global classification accuracy of 78%. The results were validated using the data obtained from a survey conducted by the University of Coimbra.

The integration of uncertainty information proved to be valuable in the hybrid classification process, contributing significantly to improve the final classification results. The evaluation of the thematic accuracy with uncertainty information showed that there is a positive correlation between the information given by the uncertainty measures and the user's and producer's accuracy but mainly with the user's accuracy. The results showed that the uncertainty measures may be used as indicators of the classifier performance and consequently as indicators of the final classification results, however, they cannot replace the accuracy indexes.

Agradecimentos

Desejo manifestar o meu sincero reconhecimento aos meus orientadores científicos Doutor Eduardo Nuno Brito Santos Júlio e ao Doutor Mário Sílvio Rochinha de Andrade Caetano, pelas suas sábias sugestões e pelo estímulo e apoio que sempre demonstraram.

Uma palavra de agradecimento muito especial à Doutora Cidália Costa pela amizade oferecida, durante a realização desta tese, que foi fundamental para minimizar os momentos de desânimo e pela sua disponibilidade, de valor inestimável, em acompanhar colaborar e incentivar as várias etapas deste trabalho.

Ao Instituto Geográfico Português pela cedência de imagens aéreas multiespectrais e imagens IKONOS.

À Câmara Municipal de Coimbra pela cedência das imagens QuickBird e informação cartográfica necessária para a realização do estudo.

À Universidade de Coimbra e ao Doutor Romeu da Silva Vicente pela disponibilização dos dados alfanuméricos relativos ao Processo de Reabilitação e Renovação da Baixa de Coimbra.

Gostaria ainda de agradecer à Engenheira Tatiana Plantier, Dr. Hugo Carrão e Engenheiro José Nogueira pela disponibilidade e colaboração prestadas em algumas tarefas, que constituíram uma ajuda valiosa.

Estendo também os meus agradecimentos a todas as pessoas, colegas do Departamento de Engenharia Civil, amigos e familiares que, de alguma forma, me apoiaram durante o período de execução deste trabalho.

Índice

| | |
|---|-----------|
| Capítulo 1 | 1 |
| Introdução..... | 1 |
| 1.1 Enquadramento..... | 1 |
| 1.2 Objectivos..... | 6 |
| 1.3 Organização do documento | 9 |
| Capítulo 2 | 13 |
| Incerteza | 13 |
| 2.1 Introdução..... | 13 |
| 2.2 Teorias de Incerteza: breve referência histórica..... | 17 |
| 2.3 Medidas de Incerteza | 23 |
| 2.3.1 Entropia | 25 |
| 2.3.2 Medidas de não-especificidade..... | 26 |
| 2.3.3 . Medida do desvio máximo relativo | 29 |
| 2.4 Avaliação comparativa do comportamento das medidas de incerteza..... | 32 |
| 2.4.1 Medidas aplicadas a distribuições de possibilidade..... | 32 |
| 2.4.2 Medidas aplicadas a distribuições de probabilidade..... | 38 |
| 2.5 Conclusões..... | 43 |
| Capítulo 3 | 46 |
| Aplicação de medidas de incerteza no processo de produção de mapas de unidade de paisagem..... | 46 |
| 3.1 Introdução..... | 46 |
| 3.2 Revisão bibliográfica | 49 |
| 3.3 Área de estudo e dados..... | 56 |
| 3.4 Abordagem metodológica..... | 57 |
| 3.4.1 Protocolos de amostragem | 61 |
| 3.4.2 Nomenclatura | 63 |
| 3.4.3 Mapa de elementos de superfície | 65 |

| | |
|---|------------|
| 3.4.3.1 Classificação..... | 65 |
| 3.4.3.2 Quantificação da incerteza..... | 68 |
| 3.4.3.3 Avaliação dos classificadores..... | 69 |
| 3.4.3.4 Avaliação da exactidão temática do mapa | 72 |
| 3.4.3.5 Distribuição espacial da incerteza | 75 |
| 3.4.4 Mapa de unidades de paisagem | 76 |
| 3.4.4.1 Abordagem de classificação combinada com integração da incerteza..... | 76 |
| 3.4.4.2 Abordagem de classificação combinada sem integração de incerteza..... | 79 |
| 3.4.4.3 Avaliação da exactidão temática do mapa de unidades de paisagem | 80 |
| 3.5 Resultados e discussão..... | 81 |
| 3.5.1 Avaliação dos classificadores | 81 |
| 3.5.1.1 Avaliação dos classificadores com base em índices de exactidão estatísticos. | 81 |
| 3.5.1.2 Avaliação dos classificadores com base em índices de incerteza | 84 |
| 3.5.2 Avaliação da exactidão temática do mapa de elementos de superfície | 97 |
| 3.5.2.1 Avaliação da exactidão temática do mapa de elementos de superfície com base em índices de exactidão estatísticos | 97 |
| 3.5.2.2 Avaliação da informação temática do mapa de elementos de superfície com base em índices de incerteza..... | 99 |
| 3.5.3 Distribuição espacial da incerteza..... | 102 |
| 3.5.4 Avaliação da exactidão temática dos mapas de unidades de paisagem com e sem integração de incerteza: análise comparativa dos resultados..... | 105 |
| 3.6 Conclusões..... | 109 |
| Capítulo 4 | 111 |
| Utilização de imagens multiespectrais na avaliação do estado de conservação das coberturas da Baixa de Coimbra | 111 |
| 4.1 Introdução..... | 111 |
| 4.2 Área de estudo e dados..... | 114 |
| 4.3 Metodologia | 117 |
| 4.3.1 Protocolo de amostragem | 119 |
| 4.3.2 Análise espectral | 120 |
| 4.3.3 Classificação | 121 |
| 4.4 Resultados e discussão..... | 125 |
| 4.4.1 Análise espectral | 125 |
| 4.4.2 Classificação | 126 |

| | |
|--|------------|
| 4.4.3 Análise comparativa dos resultados: abordagem proposta versus métodos tradicionais | 130 |
| 4.5 Conclusões..... | 132 |
| Capítulo 5 | 134 |
| Conclusões..... | 134 |
| 5.1 Síntese do trabalho realizado | 134 |
| 5.2 Principais conclusões..... | 136 |
| 5.3 Perspectivas de trabalho futuro..... | 142 |
| Bibliografia | 144 |
| ANEXOS | 158 |
| Anexo A – Imagem IKONOS | 159 |
| Anexo B – Regras de Classificação do MUP..... | 161 |
| Anexo C – Regras de Classificação do MAC..... | 170 |

Índice de Figuras

| | |
|---|----|
| Figura 1.1 – Síntese da organização da tese..... | 9 |
| Figura 2.1 – Esquema sumário das medidas de incerteza (adaptado de Klir e Folger 1988)..... | 24 |
| Figura 2.2 – Valores de NSp, Un e RI, ordenados por ordem crescente dos valores de NSp, para um conjunto de distribuições de possibilidade associadas a objectos numerados de 1 a 37, considerando três possíveis classes. | 33 |
| Figura 2.3 – Valores de NSp, Un e RI, ordenados por ordem crescente dos valores de NSp, obtidos para um conjunto de distribuições de possibilidade associadas a objectos numerados de 1 a 37, considerando nove classes como teoricamente possíveis e possibilidades nulas para as classes da quarta à nona. | 34 |
| Figura 2.4 – Comportamento das medidas de incerteza NSp, Un e RI para distribuições de possibilidade com diferentes graus de dispersão para todas as classes. | 35 |
| Figura 2.5 – Comportamento das medidas de incerteza NSp, Un e RI para distribuições de possibilidade em que os diferentes graus consecutivos apresentam uma variação de valores de possibilidade idêntica..... | 36 |
| Figura 2.6 – Comportamento das medidas de incerteza NSp, Un e RI para distribuições de possibilidade em que todos os graus de possibilidade são iguais em cada distribuição de possibilidade. | 37 |
| Figura 2.7 – Valores de Er e RI, ordenados por ordem crescente dos valores de RI, obtidos para distribuições de probabilidade numeradas de 1 a 12, considerando quatro possíveis classes. | 39 |
| Figura 2.8 – Valores de Er e RI obtidos para distribuições de probabilidade numerados de 1 a 12 considerando dez possíveis classes e probabilidades nulas para | |

| | | |
|---------------|---|----|
| | as classes da quarta à décima. | 40 |
| Figura 2.9 – | Comportamento das medidas de incerteza E_r e R_I para distribuições de probabilidade associadas a pixels numerados de 1 a 27, correspondendo a diferentes graus de dispersão para todas as classes, ordenados por ordem crescente dos valores de R_I | 41 |
| Figura 2.10 – | Comportamento das medidas de incerteza E_r e R_I para distribuições de probabilidade associadas a pixels numerados de 1 a 19, correspondendo a diferentes graus de dispersão para todas as classes, ordenados por ordem crescente dos valores de R_I | 42 |
| Figura 3.1 – | Extracto de uma imagem IKONOS (RGB 432): a) estrutura matricial da imagem em que é visível a grelha regular de células designadas por pixels; b) representação de objectos na estrutura matricial sobrepostos pelos objectos vectoriais representados por polígonos (contornos a azul)..... | 51 |
| Figura 3.2 – | Área de estudo: Imagem IKONOS (RGB 432)..... | 56 |
| Figura 3.3 – | Fluxograma da metodologia utilizada para o estudo da eleição dos classificadores e produção do MES..... | 60 |
| Figura 3.4 – | Fluxograma da metodologia utilizada para a produção do MUP..... | 61 |
| Figura 3.5 – | Extractos da imagem IKONOS (RGB 432) que ilustram corpos de água com diferentes reflectâncias. | 64 |
| Figura 3.6 – | a) Exemplo ilustrativo da classificação assistida com o classificador fuzzy da distância mínima à média; b) função de pertença sigmoide usada no classificador fuzzy baseado no classificador da distância mínima à média. | 67 |
| Figura 3.7 – | Esquema de classificação das unidades de paisagem..... | 79 |
| Figura 3.8 – | Índices de Exactidão do Utilizador (EU) para os classificadores fuzzy (F_DM) e probabilístico (MV_B). | 82 |
| Figura 3.9 – | Índices de Exactidão do Produtor (EP) para os classificadores fuzzy (F_DM) e probabilístico (MV_B)..... | 83 |

- Figura 3.10 – Incerteza média por classe obtida com a aplicação das medidas de incerteza E_r e RI aos resultados da classificação efectuada com o classificador probabilístico, para o conjunto amostral constituído pelas amostras de teste..... 84
- Figura 3.11 – Distribuição da incerteza da classificação probabilística, obtida com as medidas E_r e RI , por cinco níveis de incerteza correspondendo a 0% e a quatro quantis..... 85
- Figura 3.12 – Correlação por classe obtida entre as medidas E_r e RI , para as amostras de teste. 85
- Figura 3.13 – Visualização da distribuição espacial da incerteza da classificação probabilística, obtida com as medidas E_r e RI , considerando cinco níveis de incerteza correspondendo a 0% e a quatro quantis: a) incerteza obtida com a medida RI ; b) imagem IKONOS da área de estudo (RGB 432); c) incerteza obtida com a medida E_r 86
- Figura 3.14 – Incerteza média por classe obtida com a aplicação das medidas de incerteza E_r e RI aos resultados da classificação efectuada com o classificador fuzzy (F_{DM}), para o conjunto amostral constituído pelas amostras de teste..... 87
- Figura 3.15 – Distribuição da incerteza da classificação fuzzy, obtida com as medidas E_r e RI , considerando cinco níveis de incerteza, correspondendo a 0% e a quatro quantis..... 87
- Figura 3.16 – Correlação por classe obtida entre as medidas RI , N_{Sp} e Un , para as amostras de teste..... 88
- Figura 3.17 – Visualização da distribuição espacial da incerteza da classificação fuzzy, obtida com as medidas Un , N_{Sp} , RI , considerando cinco níveis de incerteza correspondendo a 0% e a quatro quantis: a) incerteza obtida com a medida N_{Sp} ; b) imagem IKONOS da área de estudo (RGB 432); c) incerteza obtida com a medida Un ; d) incerteza obtida com a medida RI 89
- Figura 3.18 – Índices de incerteza $I_{1-\overline{RI}}$ e $I_{1-\overline{Er}}$ e Exactidão do Utilizador (EU) e do

| | |
|--|----|
| Produtor (EP) para o classificador probabilístico. | 90 |
| Figura 3.19 – Índices de incerteza $I_{RI<0.5}$ e $I_{Er<0.5}$ e Exactidão do Utilizador (EU) e do Produtor (EP) para o classificador probabilístico. | 91 |
| Figura 3.20 – Visualização da correlação entre o índice de Exactidão do Utilizador (EU) e os índices de incerteza I_{1-RI} , I_{1-Er} | 92 |
| Figura 3.21 – Visualização da correlação entre o índice de Exactidão do Utilizador (EU) e os índices de incerteza $I_{RI<0.5}$ e $I_{Er<0.5}$ | 92 |
| Figura 3.22 – Índices de incerteza I_{1-NSp} , I_{1-Un} , I_{1-RI} e Exactidão do Utilizador (EU) e do Produtor (EP) para o classificador F_DM. | 93 |
| Figura 3.23 – Índices de incerteza $I_{RI<0.5}$, $I_{NSp<0.5}$, $I_{Un<0.5}$ e Exactidão do Utilizador (EU) e do Produtor (EP) para o classificador F_DM. | 93 |
| Figura 3.24 – Correlação entre o índice de Exactidão do Utilizador (EU) e os índices de incerteza $I_{RI<0.5}$, $I_{NSp<0.5}$, $I_{Un<0.5}$ e, I_{1-NSp} , I_{1-Un} , I_{1-RI} | 95 |
| Figura 3.25 – Soma dos valores dos três melhores graus de possibilidade associados às unidades amostra classificadas no mapa como classes Herbáceas (H). | 96 |
| Figura 3.26 – Probabilidade Condicionada pelo Mapa (PCM) e Probabilidade Condicionada pela Referência (PCR) do Mapa de Elementos de Superfície. | 98 |
| Figura 3.27 – Incerteza média por classe obtida com a aplicação das medidas de incerteza Er e RI aos resultados da classificação efectuada com o classificador probabilístico, para o conjunto amostral constituído pelas amostras de referência. | 99 |
| Figura 3.28 – Distribuição da incerteza da classificação probabilística, obtida com as medidas Er e RI, considerando cinco níveis de incerteza correspondendo a 0% e a quatro quantis. | 99 |
| Figura 3.29 – Correlação por classe obtida entre as medidas Er e RI para as amostras de | |

| | |
|--|-----|
| referência..... | 100 |
| Figura 3.30 – Índices de incerteza, I_{1-RI} e I_{1-Er} , Probabilidade Condicionada pelo Mapa (PCM) e a Probabilidade Condicionada pela Referência (PCR)..... | 101 |
| Figura 3.31 – a) imagem IKONOS (RGB 432); b) c) d) e e) mostram, respectivamente, o MES-MV produzido com o classificador probabilístico (MV_B) e a distribuição espacial da incerteza obtida com as medidas de RI, Er e E _x . As imagens f), g), h) e i) mostram, respectivamente, o MES-F produzido com o classificador fuzzy (F_DM) e a distribuição espacial da incerteza obtida com as medidas de RI, Un e E _x | 104 |
| Figura 3.32 – Probabilidade Condicionada pelo Mapa (PCM). | 105 |
| Figura 3.33 – Probabilidade Condicionada pela Referência (PCR). | 106 |
| Figura 3.34 – Regras de classificação aplicadas: a) imagem IKONOS (RGB 432) segmentada; b) objectos sobrepostos ao MES obtido com o classificador probabilístico (MV_B); c) objectos sobrepostos à informação sobre a incerteza; d) MUP obtido com a aplicação das regras..... | 107 |
| Figura 3.35 – MUP obtido com ambas as metodologias: a) classificação sem integração da incerteza; b) classificação com integração da incerteza; c) extracto da imagem IKONOS (RGB 432). Os círculos a branco marcam unidades de paisagem ocupadas por agricultura que foram correctamente classificadas quando a incerteza foi considerada nas regras de classificação..... | 108 |
| Figura 3.36 – Mapa de unidades de paisagem final..... | 109 |
| Figura 4.1 – Imagem de coberturas no centro histórico da Baixa de Coimbra..... | 111 |
| Figura 4.2 – Limite da área de estudo sobre: a) imagem aérea (RGB 321); b) imagem QuickBird (RGB 321)..... | 116 |
| Figura 4.3 – Extracto da imagem aérea (RGB 321) com a sobreposição da informação cartográfica dos edifícios..... | 116 |
| Figura 4.4 – Esquema da metodologia utilizada para a produção do MAC..... | 119 |

- Figura 4.5 – Exemplo das amostras de treino em extractos da imagem aérea (RGB 321) (a) e da imagem QuickBird (RGB 321) (b); as diferentes cores das unidades amostra correspondem a diferentes classes. 120
- Figura 4.6 – Fluxograma de Classificação..... 124
- Figura 4.7 – Diagrama de dispersão da imagem aérea e QuickBird onde as elipses representam as classes dos elementos de superfície dos materiais: Telha Cerâmica Escura (TC_E); Telha Cerâmica Clara (TC_C); Fibrocimento (F); Chapa Metálica (CM) e Sombra (S). 125
- Figura 4.8 – Diagrama de dispersão das imagens aérea e QuickBird onde as elipses representam as classes dos elementos de superfície de telha cerâmica com anomalia (TC-A); Não Degradada (TC-ND) e Telha Cerâmica Sombra (TC-S). 125
- Figura 4.9 – Índices de Exactidão do Produtor (EP) e do Utilizador (EU) dos Mapas de Elementos de Superfície dos Materiais de revestimento (MES-M), produzidos com a imagem aérea e a QuickBird. 127
- Figura 4.10 – Índices de Exactidão do Produtor (EP) e do Utilizador (EU) dos Mapas de Elementos de Superfície das Anomalias da telha cerâmica (MES-A) produzidos com a imagem aérea e a QuickBird. 127
- Figura 4.11 – Incerteza média por classe dos Mapas de Elementos de Superfície dos Materiais de revestimento (MES-M), produzidos com a imagem aérea e a QuickBird..... 128
- Figura 4.12 – Incerteza média por classe da classificação dos Mapas de Elementos de Superfície das Anomalias da telha cerâmica (MES-A) produzidos com a imagem aérea e a QuickBird..... 129
- Figura 4.13 – Índices de Exactidão do Utilizador (EU) e do Produtor (EP) do mapa final (MAC) obtido com a abordagem combinada aplicada às imagens aéreas. 129
- Figura 4.14 – Mapa de Anomalias das Coberturas dos edifícios com telha cerâmica (MAC): a) mapa obtido pela Universidade de Coimbra por métodos tradicionais; b) mapa obtido de forma automática com a abordagem

- desenvolvida; c) diferença entre os dois mapas..... 130
- Figura 4.15 – a) Extracto da imagem aérea (RGB 321) com a sobreposição dos edifícios com coberturas não degradadas resultante do levantamento efectuado pela Universidade de Coimbra; b) Extracto da imagem aérea (RGB 321) com sobreposição dos edifícios com coberturas não degradadas resultante do processo de classificação automático. Os círculos assinalam alguns dos edifícios com coberturas reabilitadas e que foram bem identificados pelo método automático de classificação. 131
- Figura 4.16 – Classificação das unidades de edifícios a partir dos elementos de superfície classificados. a) extracto da imagem aérea (RGB 321); b) classificação ao nível do pixel dos elementos de superfície Telha Cerâmica Escura (TC_E), Telha Cerâmica Clara (TC_C) e Sombra (S); c) classificação ao nível do edifício em Edifícios com Cobertura de Telha Cerâmica (E-C). Os círculos assinalam os edifícios que apresentam cobertura de telha cerâmica com sombra e com diferente brilho e que foram correctamente identificados com a abordagem de classificação desenvolvida..... 132

Índice de Tabelas

| | |
|--|-----|
| Tabela 2.1 – Medidas de incerteza. | 31 |
| Tabela 3.1 – Características da imagem IKONOS. | 56 |
| Tabela 3.2 – Nomenclatura dos elementos de superfície. | 63 |
| Tabela 3.3 – Nomenclatura das unidades de paisagem. | 64 |
| Tabela 3.4 – Matriz de erro. | 70 |
| Tabela 3.5 – Parâmetros da segmentação. | 77 |
| Tabela 3.6 – Regras de classificação. | 78 |
| Tabela 3.7 – Exactidão Global (EG) e coeficiente de KHAT (KHAT) para os classificadores fuzzy (F_DM) e probabilístico (MV_B). | 81 |
| Tabela 3.8 – Matriz de confusão para avaliação do classificador probabilístico (MV_B). | 82 |
| Tabela 3.9 – Matriz de confusão para avaliação do classificador fuzzy (F_DM). | 82 |
| Tabela 3.10 – Valores de correlação entre as diferentes medidas de incerteza. | 89 |
| Tabela 3.11 – Média dos valores do melhor grau e segundo grau de possibilidade associados aos pixels amostra. | 94 |
| Tabela 3.12 – Correlação entre os índices de incerteza e os IEE para os classificadores fuzzy (F_DM) e probabilístico (MV_B). | 97 |
| Tabela 3.13 – Índices de Exactidão Gobal (EG e KHAT) e Probabilidade Global (PG). | 97 |
| Tabela 3.14 – Soma das distribuições de probabilidade por classe. | 102 |
| Tabela 3.15 – Correlação entre os índices de incerteza e os IPEE | 102 |

| | |
|---|-----|
| Tabela 4.1 – Regras de classificação aplicadas aos mapas MES-M e MES-A para obter o mapa MAC..... | 123 |
| Tabela 4.2 – Índices de Exactidão Global da classificação dos MES-M e MES-A..... | 127 |

Lista de Acrónimos, Abreviaturas e Símbolos

| | |
|--------|--|
| AML | <i>Arc Macro Language</i> |
| B | <i>Blue</i> (banda espectral azul) |
| bit | Digito binário que toma valores 0 e 1 |
| CAD | <i>Computer Aided Design</i> |
| CELPA | Associação da Indústria Papeleira Portuguesa |
| $C(x)$ | Função característica |
| DGRF | Direcção Geral dos Recursos Florestais |
| EG | Exactidão Global |
| EP | Exactidão do Produtor |
| ERTS-1 | <i>Earth Resources Technology Satellite-1</i> |
| Er | Entropia Relativa |
| EU | Exactidão do Utilizador |
| F | Conjunto <i>fuzzy</i> |
| F_DM | Classificador <i>Fuzzy</i> da Distância Mínima à Média |
| FNEA | <i>Fractal Net Evolution Approach</i> |
| G | <i>Green</i> (banda espectral verde) |
| $H(p)$ | Medida de Entropia de Shannon |
| HRV | <i>High Resolution Visible</i> |

| | |
|------------------------|--|
| IDRISI | <i>Software</i> |
| IEE | Índices de Exactidão Específicos |
| IEG | Índices de Exactidão Global |
| IGP | Instituto Geográfico Português |
| IMGRE | Imagens de Muito Grande Resolução Espacial |
| IKONOS | (satélite) |
| IPEE | Índices Probabilísticos de Exactidão Específicos |
| IPEG | Índices Probabilísticos de Exactidão Global |
| $I_{1-\overline{Er}}$ | Incerteza média por classe obtida com a medida de entropia relativa |
| $I_{Er<0.5}$ | Índice de incerteza baseado nos baixos valores de incerteza obtida com a medida de entropia relativa, corresponde à percentagem de objectos por classe com incerteza inferior a 0.5 |
| $I_{1-\overline{NSp}}$ | Incerteza média por classe obtida com a medida de não-especificidade |
| $I_{NSp<0.5}$ | Índice de incerteza baseado nos baixos valores de incerteza obtida com a medida de não-especificidade, corresponde à percentagem de objectos por classe com incerteza inferior a 0.5 |
| $I_{1-\overline{RI}}$ | Incerteza média por classe obtida com o <i>ratio</i> de incerteza |
| $I_{RI<0.5}$ | Índice de incerteza baseado nos baixos valores de incerteza obtida com o <i>ratio</i> de incerteza, corresponde à percentagem de objectos por classe com incerteza inferior a 0.5 |
| $I_{1-\overline{Un}}$ | Incerteza média por classe obtida com a medida de incerteza-U normalizada |
| $I_{Un<0.5}$ | Índice de incerteza baseado nos baixos valores de incerteza, corresponde à |

percentagem de objectos por classe com incerteza inferior a 0.5

| | |
|---------|--|
| Landsat | (satélite) |
| MAC | Mapa de Anomalias das Coberturas |
| MES | Mapa de Elementos de Superfície |
| MES-M | Mapa de Elementos de Superfície dos Materiais |
| MES-A | Mapa de Elementos de Superfície das Anomalias |
| MSS | <i>Multispectral Scanner</i> (sensor) |
| MUP | Mapa de Unidades de Paisagem |
| MV_B | Classificador da Máxima Verosimilhança não Rígido |
| NDVI | <i>Normalized Difference Vegetation Index</i> |
| $N(A)$ | Medida de necessidade de ocorrência do acontecimento A |
| NSp | <i>Non-Specificity (Não-especificidade)</i> |
| NSp(II) | Medida de incerteza de Não-especificidade para uma distribuição ordenada de possibilidades Π |
| OrbView | (satélite) |
| p | Função p da distribuição de probabilidade |
| $P(A)$ | Medida de probabilidade de ocorrência do acontecimento A |
| Pan | <i>Panchromatic</i> (sensor) |
| PCM | Probabilidade Condicionada pelo Mapa |
| PCR | Probabilidade Condicionada pela Referência |
| PG | Probabilidade Global |

| | |
|--------------|---|
| p_i | Probabilidade da ocorrência |
| $P(i)$ | Probabilidade <i>a priori</i> da classe i |
| $p(x)$ | Distribuição de probabilidade da variável x |
| $p(x_k i)$ | Função densidade de probabilidade de um <i>pixel</i> x_k pertencer à classe i |
| QuickBird | (satélite) |
| R | <i>Red</i> (banda espectral vermelha) |
| RGB | <i>Red-Green-Blue</i> |
| RI | <i>Ratio</i> de Incerteza |
| RPH | <i>Relative Probability Entropy</i> |
| SIG | Sistemas de Informação Geográfica |
| Sp | <i>Specificity</i> (Especificidade) |
| Sp (II) | Medida de incerteza de especificidade para uma distribuição ordenada de possibilidades II |
| SPOT | (satélite) |
| TM | <i>Thematic Mapper</i> (sensor) |
| UEA | Unidade Espacial de Análise (<i>pixel</i> ou objecto) |
| UEAs | Unidades Espaciais de Análise |
| u_i | Vector média para a classe i |
| UMC | Unidade Mínima Cartográfica |
| $U(II)$ | Medida de incerteza-U para uma distribuição ordenada de possibilidades II |
| Un | Medida de incerteza-U normalizada |

| | |
|-----------------|---|
| V_i | Matriz variância – covariância para a classe i |
| V_i^{-1} | Inversa da matriz variância – covariância para a classe i |
| VHR | <i>Very High Resolution</i> (sensor) |
| X_k | Vector obtido a partir da resposta espectral do <i>pixel</i> x_k em todas as bandas |
| XS | Modo de funcionamento multiespectral do sensor HRV |
| α_{\max} | O maior valor de $\Pi(x)$ |
| $\mu_F(x)$ | Grau de pertença de x ao conjunto <i>fuzzy</i> F |
| Π | Medida de possibilidade |
| $\Pi(A)$ | Medida de possibilidade de ocorrência do acontecimento A |
| $ \Pi_\alpha $ | Cardinalidade de Π_α |
| Π_α | Nível de possibilidade α de Π |
| π | Função de distribuição de possibilidade |
| $\pi(x)$ | Distribuição de possibilidade da variável x |
| X | Conjunto Universal |

Capítulo 1

Introdução

1.1 Enquadramento

Desde o lançamento do primeiro satélite de observação da Terra, o *Earth Resources Technology Satellite-1* (ERTS-1), em 1972, que as imagens obtidas por detecção remota têm servido como fonte de informação em várias áreas, nomeadamente na monitorização dos recursos naturais e na produção de cartografia de ocupação do solo. Como exemplo, refira-se as abordagens que utilizam processos de inferência em que as categorias de uso do solo e os valores da radiação espectral dos *pixels* imagem são relacionadas com uma variável urbana (e.g., densidade populacional) (e.g. Lo, 1995; Wu *et al.*, 2005; Lu *et al.*, 2006), a delimitação ou estudo das zonas morfológicas urbanas a partir da análise espacial dos padrões das várias categorias de uso do solo (e.g. Donnay e Thomsin, 1994; Tenedorio *et al.*, 2005), a modelação da variabilidade espacial e temporal da biomassa utilizando índices de vegetação e imagens hiperspectrais (e.g. Foody *et al.*, 2003; Lu, 2006; Cho e Skidmore, 2009).

No início do século XXI, encontravam-se em órbita cerca de 31 satélites que disponibilizavam dados de ocupação do solo com uma resolução que variava entre 1 e 30 m (Stoney, 2002; Gonçalves e Caetano, 2008). O lançamento de satélites com sensores ópticos, que recolhem Imagens de Muito Grande Resolução Espacial (IMGRE) da superfície terrestre, como o IKONOS e o QuickBird, vieram disponibilizar uma nova fonte de dados importante colocando novos desafios à comunidade científica, o que levou ao desenvolvimento de novos estudos e ao aparecimento de novas tecnologias. O aumento da resolução espacial potenciou a sua utilização no contexto das áreas urbanas a grande escala abrindo, igualmente, horizontes a novas aplicações (Blaschke *et al.*, 2004); como exemplo, refira-se a sua utilização para extracção de informação de suporte à avaliação do estado de conservação do património construído. Para definir estratégias de intervenção no espaço construído, seja à escala da cidade, do quarteirão ou do edifício, é necessário proceder ao levantamento prévio da realidade existente. As técnicas habitualmente

utilizadas neste processo revelam-se trabalhosas e morosas e a dificuldade de aceder à informação de algumas zonas dos edifícios, como sejam as coberturas, dificulta o processo de avaliação. Neste contexto, e atendendo às actuais tendências de aumento significativo das acções de reabilitação urbana, é importante explorar ferramentas alternativas de observação e procurar desenvolver métodos mais rápidos e mais económicos que permitam apoiar os agentes envolvidos no processo de reabilitação urbana (Hemmler *et al.*, 2005; Vicente, 2008). Neste âmbito, as IMGRE podem constituir uma ferramenta de dados a explorar para extrair informação de suporte às acções de planeamento e/ou diagnóstico do espaço edificado.

Teoricamente, as IMGRE permitem a identificação de objectos e unidades de paisagem mais pequenos permitindo a produção de mapas a uma escala maior do que outras imagens de menor resolução espacial (e.g. Landsat-TM), uma vez que o seu nível de resolução espacial corresponde às escalas de análise, entre 1:5 000 e 1:25 000 (Donnay *et al.*, 2001; Gonçalves e Caetano, 2004). Apesar das enormes potencialidades que o aumento da resolução espacial apresenta, verificam-se também algumas limitações na sua aplicação devidas, por exemplo, à maior variabilidade espectral e à acentuada presença de sombras, o que provoca erros na classificação (Goetz *et al.*, 2003). Como a resolução espectral é mais baixa do que a dos sensores com mais baixa resolução espacial (e.g. Landsat-TM), existem algumas limitações, como por exemplo na caracterização da ocupação florestal, o que dificulta a separabilidade das diferentes espécies florestais (Goetz *et al.*, 2003). Além disso, a maioria das teorias e métodos de análise quantitativa e de processamento foram desenvolvidas para efectuar classificações das imagens de satélite ao nível do *pixel*, com uma resolução espacial de 10 a 100m (Landsat TM, Landsat MSS, SPOT HRV, SPOT PAN). Este tipo de abordagens não permite explorar e extrair o máximo de informação contida nas IMGRE, obtidas por sensores remotos. A variabilidade espacial inerente à superfície urbana e não urbana continua a ser um problema na extracção da informação a partir dos dados de detecção remota de muito grande resolução espacial, pelo que o desenvolvimento de metodologias e técnicas que melhorem a extracção de informação é ainda necessário.

As abordagens de classificação amplamente aplicadas para extrair informação temática das imagens baseiam-se, principalmente, no processo de classificação assistida, possuindo em comum três importantes fases: (1) a identificação na imagem de áreas de treino para cada classe, as quais são utilizadas como descritores das suas características espectrais; (2) a partir dos resultados obtidos na primeira fase, é efectuada a atribuição de uma classe a cada unidade espacial da imagem (um *pixel* ou um objecto), no caso de se utilizarem classificadores rígidos,

ou de várias classes, no caso de se utilizarem classificadores não rígidos (e.g. *fuzzy*) (e.g. Wang, 1990; Foody, 2000b; Ibrahim *et al.*, 2005; Doan e Foody, 2007) (as expressões, *classificadores não rígidos* e *classificadores rígidos* são uma tradução das expressões *soft classifiers* e *hard classifiers* utilizadas na literatura anglófona); (3) a validação dos dados classificados para que seja atribuído um índice de exactidão aos resultados obtidos.

No processo de classificação existem várias fontes de incerteza que influenciam o resultado final. O comportamento do classificador, por exemplo, varia em função de uma série de variáveis, tais como: a resposta espectral das classes, as características das áreas de treino e a própria natureza do classificador. A avaliação da classificação está também sujeita a várias fontes de incerteza. O processo genericamente utilizado para estimar a exactidão temática da cartografia produzida consiste em comparar, para um conjunto de pontos amostra, os resultados da classificação com dados de referência considerados como verdade no terreno (e.g. Foody, 2002). Esta comparação é realizada com recurso a uma matriz de confusão em que, geralmente, os dados de referência são representados nas colunas da matriz e comparados com os dados do mapa produzido representados nas linhas da matriz. A informação de referência é adquirida através de observações de campo ou utilizando outras fontes como, por exemplo, imagens com maior resolução, utilizando técnicas de interpretação visual. Os resultados obtidos para a exactidão temática, com esta abordagem tradicional, são altamente dependentes do tamanho da amostra, do protocolo de amostragem e dos dados de referência utilizados, os quais, quando dependentes de um intérprete humano, possuem a subjectividade inerente à interpretação visual. Este aspecto é realçado por Goodchild (2003) que refere que, devido à heterogeneidade da ocupação de solo da superfície terrestre, não é espectável que diferentes intérpretes cheguem a uma classificação idêntica.

Outra limitação inerente às técnicas de classificação assistida para extrair informação de imagens de detecção remota, tem a ver com o facto destas terem sido desenvolvidas para serem aplicadas na classificação de classes com uma representação discreta no espaço, assumindo que cada *pixel* da imagem é puro e representa uma única classe (Foody e Arora, 1996). Contudo, o *pixel* é uma unidade espacial que pode representar uma área do terreno composta por mais do que uma classe de ocupação do solo. Por outro lado, a ocupação do solo raramente apresenta transições abruptas entre classes havendo um *continuum* natural da paisagem, existindo por isso classes que apresentam fronteiras difusas (Gopal e Woodcock, 1994).

Como existe incerteza no processo de produção dos mapas temáticos, essa incerteza pode-se

propagar aos produtos que destes resultam e afectar as decisões neles baseadas. Importa, portanto, que a sua influência no resultado final seja avaliada para que possam ser gerados produtos mais fiáveis. Consequentemente, a representação da incerteza é importante e útil no processo de decisão dos utilizadores em vários domínios (e.g. Leitner e Buttenfield, 2000; Foody e Atkinson, 2002). Contudo, o estudo da incerteza é difícil, dado que está relacionado com o que não se conhece, ou não se conhece com certeza, ou com o que é ambíguo ou vago (Atkinson e Foody, 2002). Na segunda metade do século XX foram desenvolvidas as teorias matemáticas dos conjuntos *fuzzy*, lógica *fuzzy* e a teoria das possibilidades que permitem modelar e processar este tipo de informação. Estas teorias têm sido extensivamente utilizadas com bons resultados em várias áreas, nomeadamente Engenharia, Medicina, Ciências da Computação, entre outras, o que leva a considerar que podem ser utilizadas para modelar e processar a incerteza presente na classificação de imagens de satélite ou aéreas.

Um dos focos de investigação, durante as últimas décadas, tem sido o desenvolvimento de novas abordagens de classificação, para extrair a informação das imagens de detecção remota, que permitam associar mais do que uma classe de informação a um *pixel* (e.g. Maselli *et al.*, 1995; Brown *et al.*, 2000; Foody, 2000b; Zhang e Foody, 2001; Ibrahim *et al.*, 2005; Doan e Foody, 2007). Estes novos métodos utilizam classificadores não rígidos que atribuem a cada Unidade Espacial de Análise (UEA), diferentes graus de probabilidade, possibilidade ou pertença a uma ou várias classes, dependendo do tipo de classificador utilizado. No resultado da classificação pode ficar associada a cada UEA mais do que uma classe, por exemplo, a classe *Solo nu* com um grau de probabilidade de 0.7 e a classe *Vegetação Herbácea Esparsa* com um grau de probabilidade de 0.3. No resultado da classificação com os classificadores rígidos apenas é atribuída uma única classe a cada UEA que corresponde à classe para a qual, por exemplo, o valor de probabilidade foi mais elevado, não sendo disponibilizado o valor dessa atribuição. O resultado da classificação com métodos não rígidos, para além de permitir obter informação adicional sobre a ocupação do solo, permite também a determinação de informação sobre a incerteza da classificação. Apesar do crescente interesse da comunidade académica na aplicação da informação sobre a incerteza, no âmbito da detecção remota, nos estudos efectuados até ao momento esta tem sido essencialmente utilizada como um indicador da confiança da classificação (Maselli *et al.*, 1994; Foody, 1996; Zhu, 1997; Ricotta, 2005), existindo ainda vários aspectos que necessitam de ser aperfeiçoados. Refira-se, como exemplo, a análise da influência da incerteza no processo de classificação e avaliação dos resultados e o desenvolvimento de novas metodologias para que esse conhecimento possa ser usado na produção de mapas mais exactos e, adicionalmente, disponibilizar ao utilizador informação mais fiável acerca da qualidade

dos mapas gerados. Por exemplo, a utilidade da sua integração na fase do processo de classificação, em que se atribui uma classe a cada unidade espacial da imagem, quando se utiliza uma combinação de métodos, nunca foi investigada.

Nas duas últimas décadas, a investigação tem sido considerável ao nível do desenvolvimento de classificadores não rígidos, para extrair informação das imagens de detecção remota, o que constitui apenas uma das fases do processo. Paralelamente, têm sido desenvolvidas igualmente novas abordagens para avaliar a exactidão das classificações (e.g. Foody, 1996; Arora e Foody, 1997; Binaghi *et al.*, 1999; Woodcock e Gopal, 2000; Oki *et al.*, 2004; Pontius e Cheuk, 2006). Contudo, a maioria destas abordagens baseia-se em matrizes, a partir das quais são inferidos índices semelhantes aos amplamente usados na avaliação das classificações obtidas com classificadores rígidos. A construção destas matrizes, para além de ser dispendiosa e morosa, atendendo a que requer a identificação da classe de ocupação e/ou uso do solo de cada amostra de referência, apenas possibilita uma estimativa da Exactidão Global, não fornecendo informação sobre a distribuição espacial do erro. Além disso, a subjectividade inerente ao processo de atribuição das classes de referência às amostras pode alterar consideravelmente a exactidão do mapa a que se referem sem que, por este motivo, este esteja mais (ou menos) correcto. O desenvolvimento de novos métodos ou a combinação de métodos existentes para avaliar a exactidão da classificação continua a ser uma área de pesquisa.

1.2 Objectivos

O trabalho a desenvolver nesta tese de doutoramento compreende duas linhas de investigação. A primeira visa avaliar a aplicabilidade e a utilidade da informação sobre a incerteza em duas fases do processo de produção automática de cartografia temática, a partir de Imagens Multiespectrais de Muito Grande Resolução Espacial (IMGRE). Na fase de extracção e classificação da informação, pretende-se avaliar se a integração da informação sobre a incerteza no processo de classificação permite melhorar os resultados da exactidão do mapa final. Na fase de avaliação da exactidão temática, pretende-se investigar se a informação sobre a incerteza, obtida a partir das medidas de incerteza, pode ser utilizada para avaliar o comportamento dos classificadores e como indicador da exactidão da classificação.

A segunda linha de investigação visa explorar as potencialidades das IMGRE, para extracção de informação de suporte à avaliação do estado de conservação do espaço construído, através da identificação dos materiais de revestimento das coberturas e caracterização das suas anomalias.

Para avaliar a utilidade da informação sobre incerteza no processo de classificação, foi desenvolvida uma abordagem que adopta o princípio da classificação combinada *pixel/objecto*, em que se testou a integração da informação sobre a incerteza associada à classificação ao nível do *pixel*, no processo de classificação ao nível do objecto.

O uso da abordagem de classificação que combina *pixels* e objectos tem mostrado ser adequado para a identificação das classes de unidade de paisagem, que contêm uma variedade de coberturas de ocupação do solo, a partir das IMGRE (e.g. Wang *et al.*, 2004; Plantier e Caetano, 2007), mas a integração da informação sobre a incerteza neste método de classificação nunca foi efectuado.

Para extrair a informação temática ao nível do *pixel* foram testados dois classificadores não rígidos o que permitiu obter a informação necessária para quantificar a incerteza dessa atribuição, com a aplicação de diferentes medidas de incerteza. Neste estudo pretende-se apenas quantificar a incerteza da informação temática que pode surgir devido à confusão ou dificuldade do classificador em atribuir, no processo de classificação, uma classe a um *pixel*.

A nova metodologia de classificação desenvolvida, teve como objectivo subjacente contribuir para reduzir, na abordagem combinada, a influência da incerteza associada à classificação ao nível do *pixel* e melhorar significativamente os resultados finais da classificação ao nível do

objecto.

A metodologia proposta foi desenvolvida e testada em dois casos de estudo: (1) numa área florestal, para obtenção de um mapa de unidades de paisagem, em que foram utilizadas imagens IKONOS; e (2) no centro histórico da Baixa de Coimbra, para obtenção de um mapa de anomalias e tipo de materiais das coberturas, tendo sido utilizadas imagens aéreas e imagens QuickBird. No primeiro caso de estudo, investigou-se se a introdução da incerteza no processo de classificação combinada melhora a exactidão do mapa produzido. No segundo caso de estudo avaliou-se a aplicabilidade da abordagem metodológica na monitorização do património edificado, onde a detecção remota não tem sido utilizada.

Para avaliar a utilidade da informação sobre a incerteza no processo de avaliação do comportamento e adequação dos classificadores e como indicador da exactidão temática da classificação, foram utilizadas diferentes medidas de incerteza, nomeadamente entropia e não-especificidade. A partir destas medidas foram desenvolvidos dois índices de incerteza, um baseado na média e outro baseado nos baixos valores de incerteza. Os índices propostos foram testados em duas classificações efectuadas, com dois classificadores assistidos não rígidos: (1) o classificador paramétrico probabilístico da máxima verosimilhança não rígido, tendo sido obtidos os graus de probabilidade *a posteriori* utilizando o teorema de Bayes, em que a cada *pixel* ficou associada uma distribuição de probabilidade às várias classes e (2) o classificador *fuzzy*, baseado na distância mínima à média, em que a cada *pixel* ficou associada uma distribuição de possibilidade. Para avaliar a utilidade da informação sobre a incerteza no processo de avaliação foi efectuada ainda uma comparação entre a informação fornecida pelos índices de incerteza desenvolvidos e a dos índices de exactidão estatísticos. No âmbito da detecção remota, as medidas de incerteza mais utilizadas para indicar o grau de confiança associado à atribuição das classes aos *pixels* e indicar a fiabilidade da classificação ao nível do *pixel* têm sido as medidas de entropia, nomeadamente, a entropia de Shannon (e.g. Maselli *et al.*, 1994; Foody, 1995; Foody, 1996; Zhu, 1997). Contudo, quando o resultado da classificação é obtido a partir de um classificador *fuzzy* em que a cada *pixel* ficam associados graus de possibilidade, estas medidas não são adequadas (Ricotta, 2005). Neste estudo, é testada a utilidade das medidas de não-especificidade no processo de avaliação de uma classificação *fuzzy* e investiga-se se estas medidas permitem avaliar a ambiguidade em especificar uma única solução. Como as medidas de incerteza de não-especificidade e de entropia podem ser aplicadas a todas as unidades espaciais, foi ainda utilizada a distribuição espacial da incerteza na análise dos resultados da classificação obtidos.

Para dar resposta à segunda linha de investigação, o estudo foi aplicado a uma área urbana, com cerca de 800 edifícios, situada no centro histórico da Baixa de Coimbra, tendo sido utilizadas imagens aéreas com 0.5 m de resolução e imagens do satélite QuickBird com uma resolução de 2.4 m. A utilização de diferentes imagens com diferentes resoluções teve como objectivo avaliar a importância da resolução espacial neste tipo de aplicações. A extracção da informação foi efectuada a partir de ambos os tipos de imagem, utilizando a abordagem de classificação combinada com integração de incerteza. A cada edifício, obtido a partir da cartografia 1:1 000, ficou associada informação sobre: o tipo de revestimento principal da cobertura, a existência de anomalias nas coberturas com revestimento de telha cerâmica e o seu estado de conservação. A informação acerca do revestimento de telha cerâmica das coberturas e da presença de anomalias foi comparada com a obtida a partir de um levantamento efectuado pela Universidade de Coimbra.

Com o estudo realizado, pretende-se contribuir para: (1) a integração de medidas de incerteza no processo de produção de cartografia temática a partir de imagens multiespectrais; (2) o desenvolvimento de uma metodologia que integre a incerteza no processo de classificação e que melhore a extracção da informação a partir das imagens multiespectrais de muito grande resolução; (3) potenciar a aplicação dos dados de detecção remota de muito grande resolução à área da reabilitação do património construído.

1.3 Organização do documento

A tese encontra-se dividida em 5 capítulos cuja estrutura está relacionada com as linhas de investigação descritas na secção 1.2. Na figura 1.1 apresenta-se a estrutura e organização desta tese.

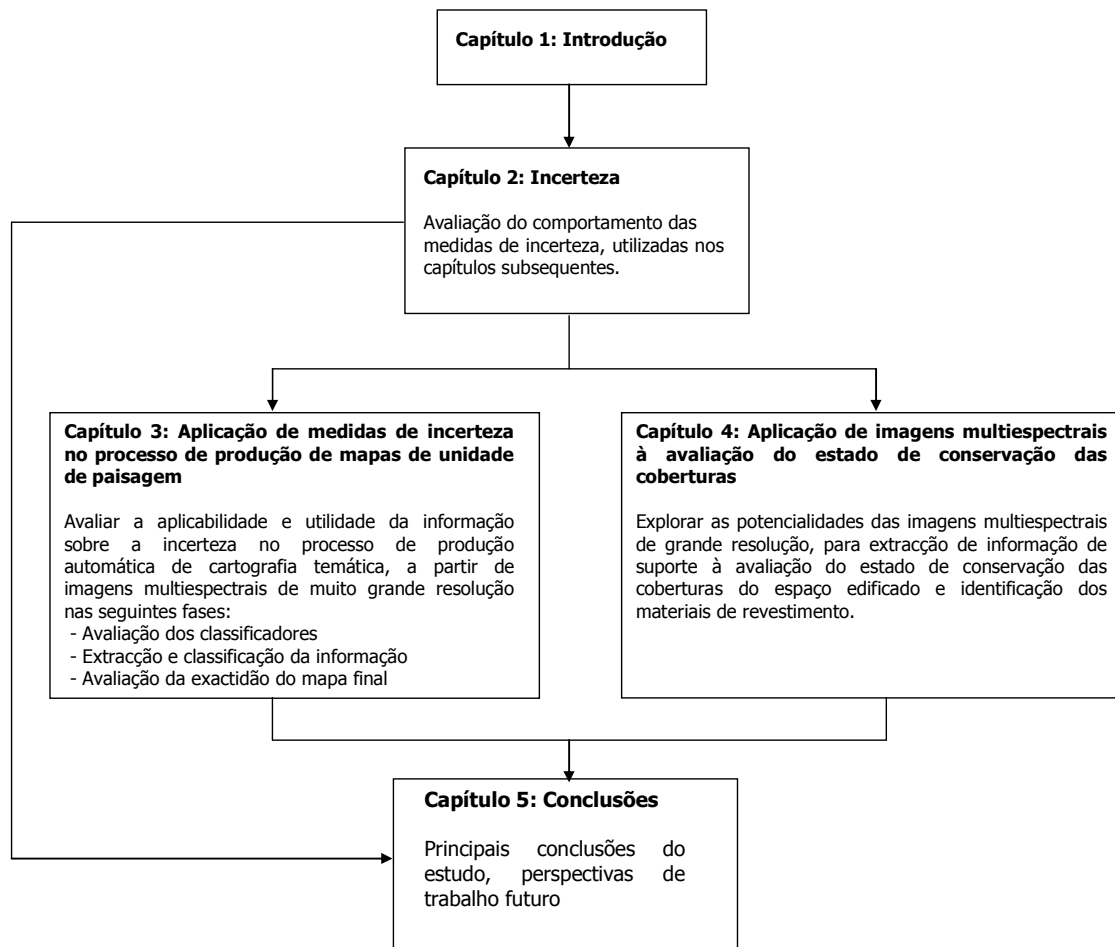


Figura 1.1 – Síntese da organização da tese.

No presente Capítulo, apresentam-se o enquadramento do tema abordado, as motivações do trabalho realizado, os objectivos enunciados e a estrutura da tese.

No Capítulo 2, são descritos alguns fundamentos teóricos relativos às medidas de incerteza, considerados pertinentes para o estudo desenvolvido e apresentado nos capítulos subsequentes. Inicialmente são abordadas as fontes de incerteza em detecção remota, seguindo-se uma revisão de literatura sobre as medidas clássicas de quantificação da incerteza e as medidas de não-especificidade, aplicáveis a conjuntos *fuzzy*, e distribuições de possibilidade, propostas por

Yagger (1982) e Higashi e Klir (1982). Por fim, apresenta-se um estudo do comportamento das medidas de não-especificidade, entropia de Shannon e uma medida disponível no *software* comercial IDRISI. Para a realização do estudo, são utilizados dados reais e dados sintéticos que simulam as distribuições de probabilidade e possibilidade ou os graus de pertença que são atribuídos aos *pixels* no processo de classificação temática utilizando algoritmos não rígidos. São ainda expostas as características mais pertinentes das várias medidas no contexto da sua aplicação à detecção remota.

O Capítulo 3 tem como perspectiva central investigar a aplicabilidade e utilidade da informação sobre a incerteza, na fase da classificação de imagens multiespectrais, para aumentar a exactidão dos mapas finais, e na fase de validação, a qual constitui uma das linhas de investigação desta tese. É apresentado um caso de estudo utilizando imagens de grande resolução do satélite IKONOS, em que é implementada e testada uma metodologia que combina classificação ao nível do *pixel* e do objecto com integração de incerteza, com o fim de obter mapas de unidades de paisagem. No processo de classificação ao nível do *pixel* são aplicados dois classificadores não rígidos e expõe-se a avaliação do seu comportamento com base em dois índices de incerteza propostos e nos índices de exactidão estatísticos. É também avaliada a utilidade da informação disponibilizada pela espacialização da incerteza no processo de avaliação da classificação. No final do capítulo, é apresentada uma análise comparativa dos resultados alcançados com a metodologia de classificação desenvolvida e uma metodologia semelhante mas na qual a informação da incerteza não é considerada.

No Capítulo 4, procede-se à avaliação das potencialidades das imagens multiespectrais de grande resolução para extracção de informação de suporte à avaliação do estado de conservação das coberturas do espaço construído, que constitui a segunda linha de investigação. É apresentado o caso de estudo da Baixa de Coimbra, em que são utilizadas imagens do satélite QuickBird, com 2.4 m de resolução espacial, e imagens aéreas, com 0.5 m de resolução. Inicialmente, são descritos os dados imagem, a nomenclatura, os protocolos de amostragem, os dados auxiliares e são apresentadas as várias etapas da metodologia. Seguidamente, expõem-se os resultados da extracção da informação dos materiais das coberturas e a identificação de existência de anomalias, efectuada a partir de ambos os tipos de imagem, ao nível do *pixel*. Apresentam-se ainda os resultados da classificação automática dos objectos ou edifícios, obtidos com base num conjunto de regras desenvolvidas que permitiram obter mapas com a informação do revestimento das coberturas, da presença de anomalias e do estado de conservação do edificado. Os resultados alcançados com a metodologia de

classificação desenvolvida são ainda comparados com os obtidos a partir de um levantamento das anomalias efectuado pela Universidade de Coimbra. Na parte final do capítulo, é feita uma discussão sobre os resultados obtidos com os dois tipos de imagem sendo referidas as suas potencialidades e limitações neste tipo de aplicações.

No Capítulo 5, resumem-se as principais conclusões deste trabalho, capítulo a capítulo, fornecendo ainda algumas perspectivas relevantes de trabalho futuro.



A investigação realizada no âmbito desta tese de doutoramento encontra-se parcialmente publicada em revistas internacionais e em conferências com publicação em actas com revisão, nomeadamente:

- [1] Gonçalves, L. M. G., Fonte, C. C., Júlio, E. N. B. S., Caetano, M., *Evaluation of soft possibilistic classifications with non specificity uncertainty measures. International Journal of Remote Sensing. (Aceite 15 Dezembro 2008), In press.*
- [2] Gonçalves, L. M. G., Fonte, C. C., Júlio, E. N. B. S., Caetano, M., *Assessment of the state of conservation of buildings through roof mapping using very high spatial resolution images. Construction and Building Materials, 23 (8): 2795-2802. doi:10.1016/j.conbuildmat.2009.03.002*
- [3] Gonçalves, L. M. G., Fonte, C. C., Júlio, E. N. B. S., Caetano, M., *A method to incorporate uncertainty in the classification of remote sensing images. International Journal of Remote Sensing - special issue "uncertainty in remotely sensed information, spatial analysis, and process modeling". 30, (20): 5489-5503. doi: 10.1080/01431160903130929*
- [4] Gonçalves, L. M. G., Fonte, C. C., Júlio, E. N. B. S., Caetano, M., 2009, *Evaluation of remote sensing images classifiers with uncertainty measures. In Spatial Data Quality From Process to Decisions, Proceedings of 6th International Symposium on Spatial Data Quality (Rodolphe Devillers and Helen Goodchild editors) CRC Press, Canadá, Julho 2009, pp. 163-177.*
- [5] Gonçalves, L. M. G., Fonte, C. C., Júlio, E. N. B. S., Caetano, M., 2009, *On the information provided by uncertainty measures in the classification of Remote Sensing images. Sessão temática "Fuzzy Geographical Information" integrada no "International Fuzzy Systems Association World Congress 2009 " (IFSA 2009) e na "European Society for Fuzzy Logic and Technology Conference" (EUSFLAT 2009), Lisboa, Julho 2009. In press. (posteriores desenvolvimentos em execução para submeter à revista International Journal of Remote Sensing).*

Parte do conteúdo do capítulo 2 faz parte do artigo [1]. No artigo é abordada a avaliação do comportamento das medidas de não-especificidade desenvolvidas por Higashi e Klir (1982) e por Yager (1982), utilizando dados reais e dados sintéticos, e investigada a utilidade das medidas

de não-especificidade no processo de avaliação da exactidão de uma classificação fuzzy. Neste capítulo, o estudo comparativo do comportamento das medidas é complementado com a avaliação do comportamento de um indicador sobre a incerteza da classificação, disponível no software comercial IDRISI. Faz-se também uma análise comparativa do comportamento da medida de entropia relativa proposta por (Maselli et al., 1994) e a medida de incerteza disponibilizada no software comercial IDRISI.

O conteúdo do capítulo 3 faz parte dos artigos [3], [4] e [5]. No artigo [4] é investigada a compatibilidade da informação dos índices de incerteza com a informação dos índices estatísticos na avaliação do comportamento de dois classificadores não rígidos. Neste capítulo o estudo é complementado com aplicação dos índices de incerteza no processo de avaliação da classificação.

O conteúdo do capítulo 4 faz parte do artigo [2]. No artigo é apresentado o caso de estudo da Baixa de Coimbra em que são utilizadas apenas imagens aéreas. No capítulo 4 apresenta-se o caso de estudo aplicado também a imagens do satélite QuikBird e faz-se uma análise comparativa dos resultados obtidos com estes dois tipos de imagens.

Capítulo 2

Incerteza

2.1 Introdução

O objectivo da detecção remota é inferir informação sobre objectos à superfície terrestre, a partir de medições efectuadas de uma localização remota, frequentemente situada no espaço. O processo de inferência não é exacto e, conseqüentemente, existe uma certa incerteza associada aos resultados produzidos (Woodcock, 2002). As fontes de erro e incerteza que afectam a informação temática obtida por classificação automática de imagens de satélite ou aéreas são várias e distribuem-se pelas diversas fases, desde a captura das imagens ao seu processamento, extracção da informação e sua representação (Woodcock, 2002). Schott (1997) apresenta as várias etapas do processo de detecção remota como elos de uma cadeia em que é subjacente a ideia de que a robustez de todo o processo é influenciada pela fragilidade entre as ligações dos elos da cadeia. Este modelo ilustra de uma forma simples como as várias etapas, desde a captura da imagem, ao seu processamento, classificação e representação da informação temática, estão interligados e como a limitação numa etapa influencia todo o processo. Segundo Lu e Weng (2007), para se conseguir atingir uma boa classificação a partir de imagens de detecção remota seria necessário entender as relações entre os diferentes níveis da classificação, identificar os elos mais frágeis da cadeia e conseguir melhorá-los. A limitação da resolução espacial e espectral das imagens, por exemplo, e as condições atmosféricas no momento de aquisição das imagens, podem provocar incerteza nos resultados obtidos a partir dos dados de detecção remota *per se*. A rectificação geométrica ou o registo de imagens podem provocar incerteza posicional e os algoritmos usados para calibrar os efeitos atmosféricos e topográficos podem provocar erros radiométricos (Lu e Weng, 2007). No processo de classificação existem também muitas fontes de incerteza que influenciam os resultados finais da classificação. A incerteza temática que pode surgir na atribuição de uma classe a uma Unidade Espacial de Análise (UEA) no processo de classificação assistida, pode ser provocada pela existência de unidades espaciais mistas, ou pelo facto de diferentes ocupações de solo

apresentarem assinaturas espectrais semelhantes ou, ainda, por diferentes métodos de classificação associarem classes diferentes aos mesmos níveis de radiância (Ibrahim *et al.*, 2005). A classificação da mesma imagem com classificadores assistidos diferentes, utilizando o mesmo protocolo de amostragem para treino dos classificadores, produz frequentemente resultados diferentes. Isto deve-se ao facto de cada método de classificação assistido definir, de forma diferente, no espaço de atributos de base, uma região correspondente a cada classe em função dos aglomerados obtidos na fase de treino. Assim, é crucial avaliar a exactidão da classificação para controlar a qualidade dos mapas obtidos. No entanto, as metodologias tradicionalmente utilizadas para avaliar a exactidão também estão sujeitas a incerteza, podendo esta variar em função de vários factores relacionados com a escolha dos pontos de referência e a recolha de informação sobre a realidade existente no terreno.

Com o aumento da utilização de mapas obtidos por processos automáticos, há um reconhecimento crescente, no domínio das Ciências de Informação Geográfica, de que a qualidade da informação espacial é extremamente importante (Gahegan e Ehlers, 2000; Brown *et al.*, 2009). Assim, a cartografia temática elaborada a partir da detecção remota deve ser fiável para poder ser integrada num Sistema de Informação Geográfica (SIG) e utilizada nos processos de tomada de decisão. Todos os erros e incerteza existentes nos dados são herdados pelos SIG, onde se misturam e combinam, podendo a sua propagação ter consequências negativas nas decisões baseadas na informação gerada (Fonte, 2003; Yong Ge *et al.*, 2008).

Nesta perspectiva, a incerteza constitui um problema central no âmbito da detecção remota e um dos principais desafios consiste em, por um lado, entender melhor a natureza e as fontes de incerteza gerados nos diferentes níveis do processo e, por outro lado, conseguir caracterizar e quantificar adequadamente a incerteza existente nos mapas produzidos, para que o utilizador possa ter informação acerca da fiabilidade das zonas classificadas. A disponibilização desta informação permitiria ao utilizador controlar a propagação da incerteza nos processos de análise e cruzamento dos dados (Fonte, 2003).

Existem vários tipos de incerteza e o significado atribuído ao termo *incerteza* aparece com diferentes interpretações na literatura. Por exemplo Heuvelink (1998), refere-se à incerteza como sinónimo de erro; Zhang (1996) refere-se à incerteza como estando relacionada com ambiguidades na avaliação de uma quantidade; Klir e Yuan (1995) referem que a incerteza envolvida numa qualquer situação é o resultado de alguma informação deficiente (e.g. incompleta, vaga, imprecisa ou contraditória); Fonte (2003) refere que a incerteza expressa uma

falta de conhecimento completo acerca do valor a ser determinado, referindo ainda que a incerteza pode estar associada à representação de fenómenos vagos ou ambíguos, ou à determinação do erro quando não é possível calcular o seu valor deterministicamente.

Neste trabalho considera-se que o termo incerteza expressa uma falta de conhecimento completo, resultado de alguma informação deficiente, podendo estar associado à representação de fenómenos vagos ou ambíguos. Em geral, o termo *vago* é associado à dificuldade em estabelecer distinções claras ou precisas entre fenómenos do mundo físico que se pretende representar, tal como a dificuldade em identificar fronteiras precisas entre diferentes classes de ocupação de solo (Klir, 1987; Foody, 1996). O termo *ambiguidade*, por outro lado, está associado às relações de um para muitos, *i.e.* situações que podem ter uma ou mais alternativas pouco específicas (Klir, 1987; Foody, 1996).

Klir (2000) considera que existem pelo menos três tipos de incerteza reconhecidos. O primeiro é designado como não-especificidade (ou imprecisão). Este tipo de incerteza reflecte a ambiguidade em especificar a solução exacta. Por exemplo, se se utilizar um sistema de análise de imagem que identifica, de forma automática, o número da face que, num lançamento do dado, fica voltado para cima e o sistema apenas sugere que o número pode ser 5 ou 6, mas não consegue ser mais específico, este tipo de incerteza resulta da limitação do sistema e é designada por não-especificidade. Em detecção remota este tipo de incerteza pode ocorrer devido, por exemplo, à resolução do sensor (Pal e Bezdek, 2000). O segundo tipo de incerteza, com origem na teoria das probabilidades, tem sido designado na literatura por diferentes termos, tais como Incerteza Probabilística (Pal e Bezdek, 2000) e Incerteza do tipo Entropia (*entropy-like uncertainty*) (Klir, 2000); o termo deriva da teoria das probabilidades em que a incerteza é quantificada pela medida de entropia introduzida por Shannon (1948). Este tipo de incerteza pode ocorrer devido a fenómenos aleatórios, à redundância ou ao acaso. Por exemplo, se no lançamento de um dado se pretender saber o número associado à face que fica voltada para cima, o problema tem que ser abordado em termos de uma distribuição de probabilidade e ao resultado dessa determinação está associada uma quantidade de incerteza que é devida ao fenómeno ser aleatório. Supondo que se pretende interpretar o valor do topo da face do dado utilizando o conceito de 'elevado' ou 'baixo', a incerteza aqui presente ocorre devido a imprecisões linguísticas. Este terceiro tipo de incerteza é designado por *fuzziness*, por estar relacionado com conjuntos *fuzzy*, e resulta de linguística imprecisa ou vaga, contrariamente aos outros dois tipos de incerteza que resultam apenas de informação deficiente (Klir, 2000).

Cada um dos tipos de incerteza requer uma medida adequada através da qual o grau de incerteza pode ser determinado. As medidas que permitem avaliar a incerteza que resulta de fenómenos vagos são designadas por medidas de *fuziness* (Klir e Folger, 1988). Medidas relacionadas com a incerteza que resulta de fenómenos ambíguos dividem-se em três tipos: medidas de não-especificidade, medidas de dissonância na evidência e medidas de confusão na evidência (Klir, 1987).

Como existem várias fontes responsáveis pelos erros e incertezas nos dados de detecção remota e, como para proceder à sua quantificação é importante aplicar os métodos adequados a cada caso, este estudo apenas aborda a incerteza da informação temática que pode surgir devido à dificuldade do classificador em atribuir, no processo de classificação, uma classe a um *pixel*.

Não se pretende, neste capítulo, apresentar um estudo exaustivo sobre as várias teorias matemáticas de incerteza e as várias medidas no âmbito de cada teoria. Pretende-se, sim, analisar o comportamento das medidas que são aplicadas nos capítulos subsequentes, de forma a compreender melhor os resultados da sua aplicação no processo de classificação e avaliação da qualidade de informação temática. Assim, começa-se por apresentar uma breve referência histórica sobre a evolução das teorias da incerteza. Relativamente às medidas para quantificação da incerteza, são apenas abordadas as medidas aplicáveis a distribuições de probabilidade, nomeadamente, a entropia de Shannon e a medida de entropia relativa proposta por Maselli *et al.* (1994), e as medidas aplicáveis a distribuições de possibilidade e a conjuntos *fuzzy*, tais como as medidas de não-especificidade desenvolvidas por Higashi e Klir (1982) e por Yager (1982). Apresenta-se, de seguida, um estudo em que foi avaliado o comportamento e a informação que se pode inferir com a aplicação das medidas de entropia relativa e não-especificidade, sendo expostas as características mais pertinentes para o estudo apresentado nos capítulos subsequentes. Para a sua realização, foram utilizados dados reais e dados sintéticos que simulam as distribuições de probabilidade ou possibilidade atribuídas às UEA, no processo de classificação com algoritmos não rígidos. No estudo efectuado é ainda considerado uma medida de incerteza da classificação disponível no *software* comercial IDRISI.

2.2 Teorias de Incerteza: breve referência histórica

Tradicionalmente, a incerteza era considerada, por parte da comunidade técnico-científica, indesejável, não havendo interesse no seu estudo (Klir, 2001). Esta realidade sofreu alguma alteração no início do século XX, resultante de vários desenvolvimentos relacionados, por exemplo, com a estatística mecânica e o princípio da incerteza proposto por Heisenberg na mecânica quântica. Contudo, e apesar dos desenvolvimentos ocorridos, pouco mudou durante a primeira metade do século XX. Embora se reconhecesse que a incerteza era relevante em algumas áreas, foi considerado que a única forma de a avaliar era através da teoria das probabilidades.

Segundo Klir (2004) existem, actualmente, duas teorias reconhecidas como teorias clássicas de incerteza, as quais emergiram na primeira metade do século XX, e que foram formalizadas em termos dos conjuntos clássicos. A teoria mais antiga é baseada na noção de *possibilidade* e, a mais recente, e que se tornou consideravelmente mais visível, é baseada na noção de *probabilidade*.

Na segunda metade do século XX, surgiram duas importantes generalizações na Matemática que provocaram uma grande alteração no reconhecimento e estudo da incerteza abrindo novas perspectivas (Klir, 2001). Uma delas foi a generalização da teoria dos conjuntos clássicos para a teoria dos conjuntos *fuzzy*, introduzidos por Zadeh (1965). A outra foi, segundo Klir (2001), a generalização da teoria das medidas clássicas para a teoria das medidas monótonas, introduzidas por Choquet em 1953.

A teoria dos conjuntos *fuzzy*, proposta por Zadeh (1965), teve como intuito caracterizar a capacidade do cérebro humano em lidar com situações vagas. Oferece uma alternativa à teoria das probabilidades, no que diz respeito à modelação da incerteza, tendo sido encarada inicialmente como uma concorrente da teoria das probabilidades (Pal e Bezdek, 1994). A motivação de Zadeh (1965) resultou do facto de, muito frequentemente, as classes de objectos existentes no mundo real como, por exemplo, "alto", "declive elevado" ou conceitos como "Urbano" e "Não urbano", não apresentarem critérios de pertença definidos com precisão. Na teoria clássica de conjuntos, a pertença de cada objecto ou elemento a um conjunto é apresentada como uma questão de veracidade ou falsidade, não permitindo uma transição gradual entre a pertença e a não pertença. Pelo contrário, na teoria dos conjuntos *fuzzy*, a pertença de um elemento a um conjunto *fuzzy* é apresentada como uma questão de grau.

Na teoria clássica de conjuntos, a cada conjunto F está associada uma função característica $c(x)$, definida no conjunto universal X , e que pode apenas tomar os valores zero ou um, sendo:

$$c(x) = \begin{cases} 0 & \text{se } x \notin F \\ 1 & \text{se } x \in F \end{cases} \quad (2.1)$$

Na teoria dos conjuntos *fuzzy* é feita uma generalização da função característica, dando origem à chamada *função de pertença* de um elemento x do conjunto universal a um conjunto F , $\mu_F(x)$, que permite associar a cada elemento do conjunto universal X graus de pertença a F .

Nestas condições, o conjunto F é denominado conjunto *fuzzy*, sendo constituído pelos pares ordenados $(x, \mu_F(x))$ e caracterizado pela função de pertença que lhe está associada (Klir e Yuan, 1995). O contradomínio da função de pertença pode ser qualquer um mas, normalmente, por simplicidade e convenção, é considerado o intervalo $[0,1]$ tendo-se:

$$\mu_F(x) : X \longrightarrow [0,1] \quad (2.2)$$

Esta teoria tem sido aplicada com sucesso em áreas como controlo industrial (e.g. Sousa e Bose, 1994; Bonissone *et al.*, 1995; Reznik *et al.*, 2000), classificações de padrões e processamento de imagem (e.g. Ishibuchi *et al.*, 1999; Zhang e Foody, 2001).

Na literatura, são apresentados vários significados para os graus de pertença a um conjunto *fuzzy*. Dubois e Prade (1997), por exemplo, apresentam três interpretações para os graus de pertença: graus de semelhança, graus de preferência e graus de incerteza.

A interpretação de graus de pertença como graus de semelhança significa que $\mu_F(x)$ pode ser considerado como o grau de proximidade de x aos elementos ideais de F . Esta interpretação baseia-se no conceito de agrupar elementos em conjuntos caracterizados pelas propriedades de um ou vários elementos, considerados ideais, sendo assumido que existe pelo menos um elemento ideal, cujo grau de pertença ao conjunto é 1, ou por propriedades representativas do conjunto. Os graus de pertença dos outros elementos ao conjunto são determinados avaliando a semelhança das suas propriedades com as propriedades do elemento ideal ou as propriedades

do conjunto (Fonte, 2003). Esta interpretação é usada na classificação de padrões, análises de regressão, análise de *cluster*, entre outras (Dubois e Prade, 1997).

A interpretação de graus de pertença como graus de preferência significa que F representa o conjunto dos objectos preferidos, ou valores de decisão de uma variável u , e $\mu_F(x)$ representa a intensidade da preferência a favor do objecto x , ou a viabilidade de seleccionar x como um valor de u .

A interpretação como grau de incerteza foi proposta por Zadeh (1978) quando introduziu a teoria das possibilidades em termos de conjuntos *fuzzy*. Segundo esta interpretação, $\mu_F(x)$ é o grau de possibilidade de uma variável u tomar o valor x , dado o incompleto estado de conhecimento " u é F " (considera-se que o conjunto de valores que a variável u pode tomar é representado por um conjunto *fuzzy* F). Esta interpretação tem sido usada em sistemas inteligentes e na inteligência artificial. De notar que a palavra *possibilidade*, ligada aos conjuntos *fuzzy* nesta última interpretação, pode transmitir um significado lógico em que *possível* significa "compatível com a informação disponível" ou um significado físico em que *possível* significa "fácil de alcançar". Dubois e Prade (1997) consideram que a interpretação de $\mu_F(x)$ como um grau de incerteza apenas se refere à interpretação lógica.

A partir dos anos 70 do século XX, foram propostas várias teorias de incerteza, começando a existir um reconhecimento crescente de que o conceito de incerteza é demasiado vasto para ser apenas avaliado pela teoria das probabilidades (Klir, 2001). Entre as várias abordagens existentes para a modelação e tratamento de incerteza encontram-se, por exemplo, a teoria das possibilidades (Zadeh, 1978) e a teoria da evidência Dempster-Shafer (Dempster, 1967; Shafer, 1976). A ideia de medir a incerteza sem recorrer às probabilidades foi proposta por Deluca e Termini (1972). Zadeh (1978) foi quem primeiro reconheceu que a teoria das possibilidades é a ferramenta natural para representar e manipular informação expressa em termos de proposições *fuzzy*, tendo iniciado o desenvolvimento da teoria das possibilidades em termos dos conjuntos *fuzzy* (Klir e Yuan, 1995).

Na teoria das possibilidades, uma medida de possibilidade Π num conjunto X (e.g. o conjunto dos números reais) é caracterizada pela distribuição de possibilidade $\pi: X \rightarrow [0,1]$, sendo definida por (Dubois *et al.*, 2004):

$$\Pi(A) = \sup\{\pi(x), x \in A\} \quad \forall A \subseteq X \quad (2.3)$$

Para conjuntos finitos, é definida por:

$$\Pi(A) = \max\{\pi(x), x \in A\} \quad \forall A \subseteq X \quad (2.4)$$

e satisfaz os seguintes axiomas,

$$\Pi(\emptyset) = 0 \quad (2.5)$$

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)) \quad (2.6)$$

Na teoria das possibilidades, uma distribuição de possibilidade π , associada a uma variável u que toma valores x no universo X , representa o que se conhece sobre o valor da variável, sendo associado aos vários elementos de X um valor no intervalo $[0,1]$.

Quando $\pi(x) = 0$, significa que $u = x$ é impossível, ou seja, que x é um valor impossível da variável u à qual π está associado. Quando $\pi(x) = 1$, apenas significa que x é um dos valores mais plausível dessa variável (Dubois *et al.*, 2004). De notar que pode associar-se a diferentes elementos X um grau de possibilidade igual a 1.

A teoria das possibilidades permite modelar a incerteza de uma forma flexível, captura perfeitamente a ignorância, permitindo que a possibilidade de qualquer evento seja igual a 1 (Dubois *et al.*, 2004).

Pode estabelecer-se uma ligação entre a teoria das possibilidades e a teoria dos conjuntos *fuzzy* quando se considera que o conjunto de valores que a variável u pode tomar é representado por um conjunto *fuzzy* F , podendo interpretar-se $\mu_F(x)$ como o grau de possibilidade da variável u tomar o valor x , *i.e.*, considerar para todo o $x \in X$, $\pi_u(x) = \mu_F(x)$. Segundo Dubois e Prade (1993), o facto da distribuição de possibilidade poder ser determinada pela função de pertença de um conjunto *fuzzy* não significa que os dois conceitos (graus de pertença e distribuição de possibilidade) sejam coincidentes. De notar que $\pi_u(x)$ estima a possibilidade da variável u ser igual a x , sabendo o estado incompleto de conhecimento " u é F " e $\mu_F(x)$ estima o grau de compatibilidade da informação $u = x$ com a afirmação que se pretende avaliar " u é F ".

Nos últimos anos têm sido demonstradas várias ligações entre as probabilidades, a teoria das possibilidades e a teoria dos conjuntos *fuzzy* que sugerem que estas teorias são complementares e não antagónicas (Dubois e Prade, 1993). Uma análise comparativa entre a teoria das probabilidades, a teoria das possibilidades e a teoria dos conjuntos *fuzzy* pode ser encontrada, por exemplo, em Klir e Yuan (1995), Dubois *et al.* (2000) e Fonte (2003). Neste trabalho não se desenvolve este assunto, sendo apenas referidas algumas diferenças básicas entre as diferentes teorias.

Uma confusão frequente que ocorre entre funções de pertença e medidas de probabilidade deve-se à forma como as definições são entendidas (Dubois e Prade, 1993). Segundo Zadeh (1965) um conjunto *fuzzy* F num universo X é definido por uma função de pertença, conforme acima exposto, em que $\mu_F(x)$ é o grau de pertença do elemento x a F . Em contraste, uma medida de probabilidade P atribui um número $P(A)$ a cada subconjunto de U (conjunto finito) e satisfaz os axiomas:

$$P(U) = 1; P(\emptyset) = 0; \tag{2.7}$$

$$\text{Se } A \cap B = \emptyset \text{ } P(A \cup B) = P(A) + P(B) \tag{2.8}$$

em que $P(A)$ é a probabilidade de um determinado valor da variável u , que toma valores no universo X , tomar os valores do conjunto A bem conhecido. Uma confusão habitual entre conjuntos *fuzzy* e probabilidades consiste em confundir $P(A)$ com um grau de pertença. Segundo Dubois e Prade (1993), quando se considera um grau de pertença $\mu_F(x)$, o elemento $x \in X$ é fixo e conhecido, não sendo o conjunto F bem definido. Pelo contrário, quando se considera $P(A)$, o conjunto A é bem definido, sendo desconhecido o valor da variável u , a que a probabilidade está ligada. A função de pertença $\mu_F(x)$ tem mais semelhança com uma distribuição de probabilidade $P(\{x\}) = p(x)$, apresentando, no entanto, a diferença de ser $\sum_{x \in X} p(x) = 1$, enquanto que $\sum_{x \in X} \mu_F(x)$ pode assumir qualquer valor não negativo (Dubois e Prade, 1993).

Outro ponto que distingue as probabilidades dos graus de pertença dos conjuntos *fuzzy* é que a teoria das probabilidades é a teoria dos acontecimentos aleatórios, enquanto que a teoria dos conjuntos *fuzzy* não está relacionada com acontecimentos, mas sim com conceitos não quantificados, como por exemplo, "perigoso", "alto" ou "quente", quantificando-se apenas até

que ponto determinado elemento está de acordo com o conceito em questão (Fonte, 2003).

Uma diferença fundamental entre a teoria das probabilidades e a teoria das possibilidades é que, na primeira, a incerteza de um acontecimento é representada por uma única medida, a medida de probabilidade $P(A)$, a qual apresenta a seguinte propriedade:

$$P(A) + P(\bar{A}) = 1 \tag{2.9}$$

o que significa que a probabilidade de um acontecimento determina a probabilidade do acontecimento contrário em que \bar{A} é o complementar de A .

Na teoria das possibilidades a incerteza de um acontecimento A é representada utilizando duas medidas: uma medida de possibilidade, $\Pi(A)$, que representa o grau de possibilidade de ocorrência de um acontecimento; e uma medida de necessidade $N(A)$, que representa o grau de impossibilidade de ocorrência do acontecimento contrário, $N(A) = 1 - \Pi(\bar{A})$, *i.e.*, a certeza de A reflecte a impossibilidade do seu complementar. Estas medidas apresentam as seguintes propriedades:

$$\Pi(A) + \Pi(\bar{A}) \geq 1 \tag{2.10}$$

$$N(A) + N(\bar{A}) \leq 1 \tag{2.11}$$

Quando existe uma total ignorância acerca de um fenómeno A , tem-se $\Pi(A) = \Pi(\bar{A}) = 1$ (e $N(A) = N(\bar{A}) = 0$) enquanto $\Pi(A \cap \bar{A}) = 0$ e $N(A \cup \bar{A}) = 1$. Esta capacidade para modelar a ignorância (falta de conhecimento) é uma característica típica da teoria das possibilidades (Dubois e Prade, 1993; Dubois *et al.*, 2004).

Outra diferença fundamental entre as duas teorias é o facto do axioma da aditividade da teoria das probabilidades:

$$P(A \cup B) = P(A) + P(B) \tag{2.12}$$

para todos os conjuntos A e B tais que $A \cap B = \emptyset$, ser substituído, na teoria das possibilidades, por:

$$\Pi(A \cup B) = \sup\{\Pi(A), \Pi(B)\} \quad (2.13)$$

o qual é um axioma menos exigente (Klir e Yuan, 1995).

As teorias das probabilidades e das possibilidades têm papéis distintos ao descrever a incerteza. Há fenómenos que podem ser representados pela teoria das probabilidades e não o podem ser pela teoria das possibilidades, sendo igualmente verdade o inverso (e.g. uma distribuição de probabilidade não expressa ignorância e uma distribuição de possibilidade não expressa fenómenos aleatórios) (Dubois e Prade, 1993). Segundo Dubois *et al.* (2004), a teoria das probabilidades disponibiliza um modelo quantitativo para modelar fenómenos aleatórios (*randomness*) e indecisão (*indecisiveness*); ao passo que a teoria das possibilidades disponibiliza um modelo qualitativo para modelar conhecimento incompleto (Dubois *et al.*, 2004). O facto de a teoria das possibilidades e da teoria das probabilidades modelarem diferentes facetas da incerteza leva a que sejam consideradas, por vários autores (e.g. Klir, 2001; Dubois *et al.*, 2004), como teorias complementares da incerteza.

2.3 Medidas de Incerteza

Várias medidas de incerteza foram desenvolvidas no âmbito da teoria dos conjuntos clássicos, teoria das probabilidades, teoria das possibilidades, teoria dos conjuntos *fuzzy* (Figura 2.1). Em cada teoria da incerteza, esta é representada por uma função, designada por função de incerteza, a qual atribui um valor numérico a cada conjunto de alternativas, no intervalo [0,1], traduzindo o grau de evidência (e.g. verosimilhança, semelhança) de que a verdadeira alternativa esteja no conjunto (Klir e Smith, 2001). Esse valor é suposto medir a quantidade de incerteza de determinado tipo, numa dada teoria de incerteza.

Das várias medidas de incerteza existentes, as duas consideradas actualmente como clássicas são a entropia de Shannon (1948), que se aplica à incerteza formalizada em termos da teoria clássica das probabilidades, e a medida de Hartley (1928), que se aplica à incerteza formalizada em termos da teoria clássica das possibilidades (Klir, 1999).

A partir das medidas clássicas de incerteza foram desenvolvidas, por diversos investigadores, métricas no âmbito de outras teorias (Klir, 1999) (ver Figura 2.1). Por exemplo, Higashi e Klir (1982) propuseram uma generalização da função proposta por Hartley (1928) para a teoria dos

conjuntos *fuzzy*, a qual designaram por *U-uncertainty*. Outra medida alternativa de não-especificidade, também aplicável a conjuntos *fuzzy* finitos não vazios, foi proposta por Yager (1982). De acordo com Klir (2000), Hohle propôs em 1982 uma generalização da função proposta por Shannon (1948), a qual designou por medida de confusão. Yagger (1983) também propôs uma generalização da função proposta por Shannon (1948) para a teoria da evidência, a qual designou por medida de dissonância. Dubois e Prade (1985), por exemplo, propuseram uma generalização da função proposta por Hartley (1928) para a teoria da evidência, para medir a não-especificidade. Segundo Klir (2000), das várias medidas estabelecidas nas várias teorias, as mais estudadas e melhor compreendidas são a entropia de Shannon e as medidas de não-especificidade.

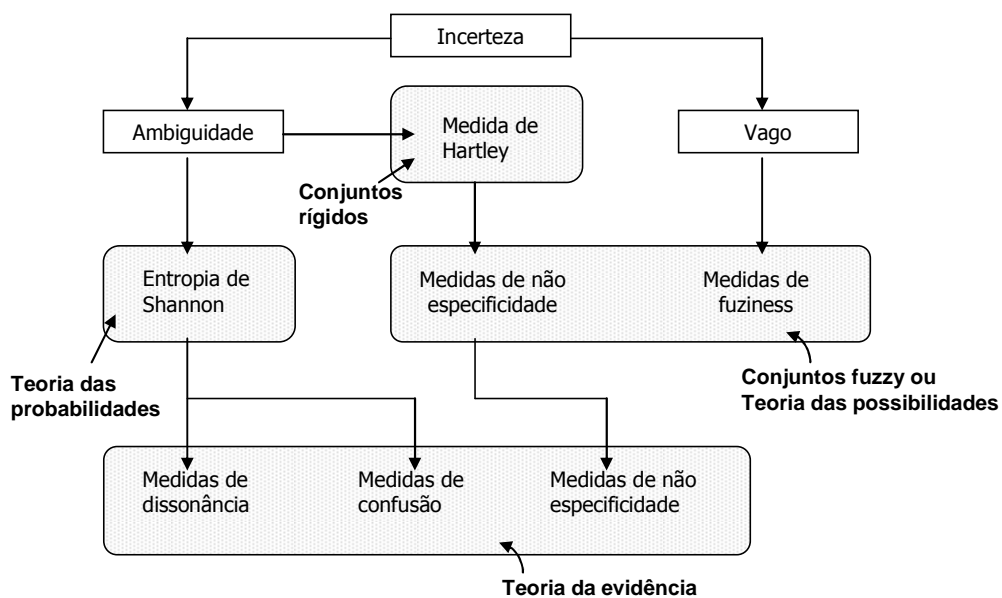


Figura 2.1 – Esquema sumário das medidas de incerteza (adaptado de Klir e Folger 1988).

No âmbito da detecção remota, as medidas de incerteza mais utilizadas têm sido as medidas de entropia, nomeadamente a entropia de Shannon. Esta medida tem sido utilizada por vários autores para indicar o grau de confiança associado à atribuição das classes aos *pixels* e indicar a fiabilidade da classificação ao nível do *pixel* (e.g. Maselli *et al.*, 1994; Foody, 1995; Foody, 1996; Zhu, 1997). Um elevado valor de entropia significa que a UEA apresenta uma atribuição partilhada por várias classes, tomando o valor máximo quando a UEA é atribuída a todas as classes com um grau de probabilidade idêntico e igual a $1/n$ (n é o número de classes). Um valor baixo de entropia significa que a UEA apresenta um elevado grau de probabilidade de pertença a uma classe, tomando o valor mínimo (entropia = 0) quando a UEA é atribuída apenas a uma classe com o grau de probabilidade de 1 (Shalan *et al.*, 2003). O valor da

entropia, como indicador do grau de confiança da classificação, é baseado no pressuposto de que, numa classificação exacta, a cada UEA fica atribuída uma única classe com um elevado valor de probabilidade (Foody, 1995). Como esta medida implica uma interpretação probabilística dos graus associados às UEAs, a sua aplicação não é apropriada a métodos de classificação que utilizam classificadores *fuzzy* e em que os graus associados às UEAs representam distribuições de possibilidade. A escolha das medidas de incerteza depende do classificador não rígido aplicado no processo de classificação. No âmbito da detecção remota, apenas recentemente tem sido dada atenção à aplicação das medidas de não-especificidade. Ricotta (2004) propôs, de forma teórica, a sua aplicação para quantificar a incerteza associada a distribuições de possibilidade. Nesta tese analisa-se a utilidade das medidas de não-especificidade no processo de avaliação de uma classificação *fuzzy*, e verifica-se se estas medidas permitem avaliar a ambiguidade em especificar uma solução única e, com base nessa informação, avalia-se o comportamento do classificador e estima-se a exactidão da classificação. Informação sobre estas medidas e um estudo comparativo sobre o seu comportamento é apresentado nas secções 2.3.2 e 2.4.1.

2.3.1 Entropia

Vários autores referem que a entropia de Shannon (1948) é a única medida que capta, de forma adequada, a incerteza no âmbito da teoria das probabilidades (e.g. Klir, 2000).

A medida clássica proposta por Shannon aplica-se à incerteza formalizada em termos de uma distribuição de probabilidade ($p(x) | x \in X$) definida num conjunto finito X . Shannon (1948) demonstrou que a forma adequada de medir a quantidade de incerteza em qualquer situação caracterizada pela distribuição de probabilidade p , é através da função H definida por:

$$H(p) = -\sum_{i=1}^n p_i \log_2 p_i \quad (2.14)$$

Do ponto de vista matemático, o significado da entropia de Shannon é realçado pelo facto de ser uma função que satisfaz um conjunto de axiomas considerados necessários para que uma medida de incerteza seja reconhecida no enquadramento da teoria das probabilidades (Ricotta, 2005).

Para uma distribuição de probabilidade p , esta medida varia no intervalo $[0, \log_2 n]$, em que o

mínimo é obtido para $p = (1, 0, \dots, 0)$ e o máximo para $p = (1/n, 1/n, \dots, 1/n)$.

Maselli *et al.* (1995) propuseram uma nova métrica baseada na entropia de Shannon, que designaram por *Relative Probability Entropy* (RPH) e que, neste estudo, é designada por Entropia Relativa (Er), dada por:

$$Er = -\frac{1}{\log_2 n} \sum_{i=1}^n p_i \log_2 p_i, \quad (2.15)$$

Para uma distribuição de probabilidade p , esta medida varia no intervalo $[0,1]$, em que o mínimo é obtido para $p = (1, 0, \dots, 0)$ e o máximo para $p = (1/n, 1/n, \dots, 1/n)$.

Em detecção remota, a utilização desta medida é apenas apropriada no caso de ser usada como um indicador da qualidade da classificação não rígida, obtida a partir de modelos probabilísticos, em situações em que uma UEA apresenta, no resultado da classificação, uma atribuição partilhada a várias classes, embora no terreno se admita que existe apenas uma classe (Foody, 1995). Contudo, quando os graus de probabilidade associados à UEA representam pertenças parciais às classes, nos casos das unidades espaciais mistas, *i.e.*, quando a pertença parcial às classes é uma característica da classificação e do terreno, a aplicação directa da entropia não é apropriada (Foody, 1996).

2.3.2 Medidas de não-especificidade

As medidas de não-especificidade foram primeiramente introduzidas por Hartley (1928), em termos das medidas clássicas possibilísticas. A medida proposta por Hartley quantifica a incerteza que resulta da ambiguidade envolvida na selecção de um elemento, a partir de um conjunto finito de alternativas possíveis, e em que apenas uma pode ser considerada verdadeira. Baseia-se no pressuposto de que as alternativas possíveis num dado universo podem ser distinguidas das não possíveis (Klir, 1987). Desta medida, resulta que a incerteza associada ao conjunto de alternativas possíveis deve-se à falta de especificidade em caracterizar a alternativa verdadeira (Klir, 2004). Quanto maior é o conjunto de alternativas possíveis, menos específica é a sua caracterização e uma especificidade total é obtida quando apenas uma alternativa é possível. Este tipo de incerteza é designado pelo termo *não-especificidade* (na literatura anglófona, *nonspecificity*) (Klir, 1999; Klir, 2004).

Existem duas medidas de não-especificidade desenvolvidas no contexto da teoria dos conjuntos *fuzzy* que foram usadas neste estudo para determinar a incerteza resultante das classificações *fuzzy*, por quantificarem a ambiguidade em especificar uma solução única: a medida de especificidade (Sp), proposta por Yager (1982), e a medida de incerteza-U, proposta por Higashi e Klir (1982) (Pal e Bezdek, 2000). Ambas podem ser aplicadas a conjuntos *fuzzy* e a distribuições de possibilidade ordenadas.

O conceito de especificidade (Sp), de uma distribuição de possibilidade foi proposto por Yager para resumir o grau até ao qual uma distribuição de possibilidade permite um, e apenas um, elemento como possível. Esta medida pode também ser usada para indicar até que ponto uma UEA contém uma, e uma só, classe temática (Ricotta, 2005).

Para uma distribuição ordenada de possibilidades π , definida num conjunto universal X , em que a distribuição de possibilidade verifica a condição $\pi(x_1) \geq \pi(x_2) \geq \dots \geq \pi(x_n)$, a medida de especificidade proposta por Yager (1982) é dada por:

$$Sp(\pi) = \int_0^{\alpha_{\max}} \frac{1}{|\pi_\alpha|} d\alpha \tag{2.16}$$

onde π_α é o nível de possibilidade α de π , que corresponde ao subconjunto de elementos que têm pelo menos um valor de possibilidades igual a α , o que em termos matemáticos pode ser expresso por: $\pi_\alpha = \{x | \pi(x) \geq \alpha\}$, onde $|\pi_\alpha|$ é a cardinalidade de π_α , *i.e.*, o número de elementos de π_α ; e α_{\max} é o maior valor de $\pi(x)$ para todo os elementos do conjunto universal X , o qual pode ser expresso como $\alpha_{\max} = \max_x \pi(x), \forall x \in X$.

A medida de especificidade Sp pode ser usada para definir a medida de não-especificidade NSp dada pela equação:

$$NSp(\pi) = 1 - Sp(\pi) \tag{2.17}$$

Substituindo (2.16) em (2.17) vem:

$$NSp(\pi) = 1 - \int_0^{\alpha_{\max}} \frac{1}{|\pi_\alpha|} d\alpha \tag{2.18}$$

e, resolvendo o integral, obtém-se (Ricotta, 2005):

$$NSp(\pi) = 1 - \sum_{i=1}^n [\pi(x_i) - \pi(x_{i+1})] \frac{1}{i} \quad (2.19)$$

onde n é o número de elementos no conjunto universal e $\pi(x_{n+1})$ assume o valor zero.

Higashi and Klir (1982) propuseram a medida de imprecisão designada por incerteza-U, a qual pode ser interpretada como uma medida de não-especificidade (Klir, 2000; Pal e Bezdek, 2000).

Para uma distribuição ordenada de possibilidades π , tem-se:

$$U(\pi) = \int_0^{\alpha_{\max}} \log_2 |\pi_\alpha| d\alpha + (1 - \alpha_{\max}) \log_2 n \quad (2.20)$$

onde π_α , $|\pi_\alpha|$, α_{\max} e n têm o mesmo significado anteriormente apresentado.

Resolvendo o integral (2.20), obtém-se (Mackay *et al.*, 2003):

$$U(\pi) = [1 - \pi(x_1)] \log_2 n + \sum_{i=2}^n [\pi(x_i) - \pi(x_{i+1})] \log_2 i \quad (2.21)$$

onde $\pi(x_{n+1})$ assume o valor zero.

Em Pal e Bezdek (2000) ou em Klir (2000) podem ser encontradas as propriedades de ambas as medidas. Destas salientam-se as a seguir indicadas:

- Para uma distribuição de possibilidade π , $0 \leq NSp(\pi) \leq 1$, em que o mínimo é obtido para $\pi(x_1, x_2, \dots, x_n) = (1, 0, \dots, 0)$ e o máximo para $\pi(x_1, x_2, \dots, x_n) = (0, 0, \dots, 0)$;
- Para um conjunto universal X e uma distribuição de possibilidade π , $0 \leq U(\pi) \leq \log_2 n$, em que o mínimo é obtido para $\pi(x_1, x_2, \dots, x_n) = (1, 0, \dots, 0)$ e o valor máximo para $\pi(x_1, x_2, \dots, x_n) = (\alpha, \alpha, \dots, \alpha)$, $\forall \alpha \in [0, 1]$.

No contexto da detecção remota, a classificação não rígida possibilística gera uma distribuição de possibilidade associada a cada unidade espacial. Considere-se o exemplo da distribuição de possibilidade $\pi(c_1, c_2, c_3, c_4) = (1, 0.8, 0.6, 0.4)$, associada a uma UEA. Os graus de possibilidade 1, 0.8, 0.6 e 0.4 expressam a possibilidade de a unidade espacial poder pertencer à classe c_1 , c_2 , c_3 e c_4 . O conjunto universal é o conjunto de todas as classes, que são as n classes consideradas e que, neste exemplo, é constituído por quatro classes. Para esta distribuição de

possibilidade, o valor de $NSp(\pi)$ e $U(\pi)$ é obtido considerando as equações (2.19) e (2.21) e é calculado por (2.22) e (2.23).

$$NSp(\pi) = 1 - [(1 - 0.8) + (0.8 - 0.6)\frac{1}{2} + (0.6 - 0.4)\frac{1}{3} + (0.4 - 0)\frac{1}{4}] = 0.53 \quad (2.22)$$

$$U(\pi) = [1 - 1]\log_2 4 + [(0.8 - 0.6)\log_2 2 + (0.6 - 0.4)\log_2 3 + 0.4\log_2 4] = 1.32 \quad (2.23)$$

As análises desenvolvidas neste estudo requerem a comparação dos resultados obtidos por estas duas medidas, *i. e.*, $NSp(\pi)$ e $U(\pi)$. Para possibilitar esta comparação, foi adoptada a medida incerteza-U, normalizada dada por:

$$U_n(\pi) = \frac{U(\pi)}{\log_2 n} \quad (2.24)$$

Esta nova métrica foi adoptada porque, por um lado, NSp e U têm diferentes intervalos de variação, o que não permite uma comparação entre os resultados obtidos por cada uma das medidas e, por outro lado, a informação dada pela medida normalizada permite uma mais fácil interpretação por parte do utilizador. Assim, o valor zero significa que não existe ambiguidade na atribuição da unidade espacial a uma classe e, portanto, não existe incerteza e o valor 1 significa que a incerteza é máxima. Por exemplo, a distribuição de possibilidade ordenada $\pi(c_1, c_2, c_3, c_4) = (1, 0, 0, 0)$, atribuída a uma unidade espacial, significa que a incerteza é mínima e que $NSp(\pi) = U_n(\pi) = 0$. Para $\pi(c_1, c_2, c_3, c_4) = (0, 0, 0, 0)$, a incerteza é máxima, pois nenhuma classe foi adequadamente atribuída à unidade espacial e $NSp(\pi) = U_n(\pi) = 1$.

2.3.3. Medida do desvio máximo relativo

Apesar do reconhecimento crescente da importância da informação sobre a incerteza, continuam a ser raros os programas comerciais de SIG e de processamento e classificação de imagem que disponibilizam o acesso à informação sobre a incerteza e sua visualização associada aos resultados da classificação temática. O *software* IDRISI disponibiliza uma medida de incerteza desenvolvida por Clark Lab. (2001), designada por *Relative Maximum Probability Deviation* (Bo e Wang, 2008), que neste estudo é designada por Medida de Desvio Máximo Relativo sendo também referida por *Ratio* de Incerteza (RI). Esta medida varia no intervalo [0, 1] e pode ser aplicada a distribuições de probabilidade e de possibilidade (Eastman, 2006), ou seja, a medida

não considera como hipótese condições que excluam a sua aplicação quer a distribuições de probabilidade, quer a distribuições de possibilidade (o que não acontece com outras medidas, as quais só podem ser aplicadas a distribuições de probabilidade ou a distribuições de possibilidade). Quando aplicada a distribuições ordenadas de probabilidade a medida é dada por:

$$RI = 1 - \frac{\max_{i=1,..,n}(p_i) - \frac{\sum_{i=1}^n p_i}{n}}{1 - \frac{1}{n}} \quad (2.25)$$

onde p_i ($i=1, \dots, n$) representa os graus de probabilidade das várias classe associadas às UEAs, n é o número de classes consideradas, o mínimo é obtido para $p=(1,0,\dots,0)$ e o máximo é obtido para $p=(1/n, 1/n, \dots, 1/n)$. Quando aplicada a distribuições de probabilidade, $\sum_{i=1}^n p_i = 1$ e a medida depende do valor máximo de probabilidade associada às UEAs e do número total de classes.

Quando aplicada a distribuições ordenadas de possibilidade a medida é dada por:

$$RI = 1 - \frac{\max_{i=1,..,n}(\pi_i) - \frac{\sum_{i=1}^n \pi_i}{n}}{1 - \frac{1}{n}} \quad (2.26)$$

onde π_i ($i=1, \dots, n$) representa os graus de possibilidade das várias classe associadas às UEAs e n tem o mesmo significado anteriormente apresentado. De salientar ainda, que para uma distribuição ordenada de possibilidades: (1) o mínimo é obtido para $\pi=(1,0,\dots,0)$; (2) o valor máximo é obtido para $\pi=(\alpha, \alpha, \dots, \alpha)$, $\forall \alpha \in [0,1]$; (3) $\sum_{i=1}^n \pi_i \geq 0$; e (4) a medida depende do valor máximo de possibilidade, da soma de todas as possibilidades associadas às UEAs e do número total de classes. O numerador do segundo termo das equações (2.25) e (2.26) representa a diferença entre o máximo grau de probabilidade/possibilidade com que uma classe é atribuída a uma UEA e o valor mínimo que o melhor grau de probabilidade/possibilidade pode

assumir se ocorrer uma dispersão total, ou seja, quando todas as classes apresentam igual probabilidade/possibilidade de ficarem atribuídas a uma UEA. Quando aplicada a distribuições de probabilidade, esta situação só ocorre quando todas as classe são atribuídas à UEA com o grau de probabilidade de $1/n$. O denominador, do segundo termo das equações (2.25) e (2.26), representa a situação extrema da diferença entre o máximo grau de probabilidade/possibilidade com que uma classe pode ser atribuída a UEA, 1 (total atribuição de uma só classe), e o valor mínimo se ocorrer uma dispersão total por todas as classes. O *ratio* destas duas quantidades expressa o grau de compatibilidade a uma classe específica (Eastman, 2006).



A Tabela 2.1 apresenta uma síntese das medidas abordadas, as quais são utilizadas nos capítulos 3 e 4. As diferentes interpretações e significados dos valores obtidos com as diferentes medidas são subsequentemente analisadas na secção 2.4. deste capítulo.

Tabela 2.1 – Medidas de incerteza.

| Nome | Fórmula | Significado dos símbolos |
|------------------------|--|---|
| Entropia de Shannon | $H(p) = -\sum_{i=1}^n p_i \log_2 p_i$ | p: distribuição de probabilidade |
| Entropia Relativa | $Er = -\frac{1}{\log_2 n} \sum_{i=1}^n p_i \log_2 p_i$ | |
| Incerteza-U | $U(\pi) = [1 - \pi(x_1)] \log_2 n + \sum_{i=2}^n [\pi(x_i) - \pi(x_{i+1})] \log_2 i$ | π : distribuição de possibilidade |
| Un | $U_n(\pi) = \frac{U(\pi)}{\log_2 n}$ | |
| Não-especificidade NSp | $NSp(\pi) = 1 - \sum_{i=1}^n [\pi(x_i) - \pi(x_{i+1})] \frac{1}{i}$ | |
| Ratio de Incerteza RI | $RI = 1 - \frac{\max_{i=1..n}(p_i) - \frac{1}{n}}{1 - \frac{1}{n}}$ | $\sum_{i=1}^n \pi_i$ p: distribuição de probabilidade π : distribuição de possibilidade |
| | $RI = 1 - \frac{\max_{i=1..n}(\pi_i) - \frac{\sum_{i=1}^n \pi_i}{n}}{1 - \frac{1}{n}}$ | |

2.4 Avaliação comparativa do comportamento das medidas de incerteza

Nesta secção é analisado o comportamento das medidas de não-especificidade, NSp e Un, entropia relativa, Er, e a medida de incerteza RI, considerando vários aspectos que incluem: (1) a sensibilidade das medidas à variação da compatibilidade entre as características das UEAs e as características que definem as classes; (2) a sensibilidade à dispersão dos graus de possibilidade ou probabilidade pelas várias classes atribuídas às UEAs. A noção de compatibilidade é expressa através da magnitude dos graus de possibilidade ou probabilidade associados a cada UEA no processo de atribuição das classes (dependendo do classificador utilizado). Graus elevados de possibilidade ou probabilidade significam uma grande compatibilidade, enquanto valores baixos significam uma baixa compatibilidade. O termo dispersão refere-se ao número de possibilidades ou probabilidades positivas atribuídas a cada UEA; quanto maior o número de possibilidades ou probabilidades positivas maior o número de classes possíveis ou prováveis. Como a medida RI, segundo Eastman (2006), pode ser aplicada a distribuições de probabilidade e possibilidade, neste estudo é efectuada uma análise comparativa do seu comportamento com as medidas de não-especificidade e com a medida de entropia relativa.

2.4.1 Medidas aplicadas a distribuições de possibilidade

Para ilustrar e avaliar o comportamento das medidas de incerteza, foram utilizados vários conjuntos de valores, reais e sintéticos, que simulam distribuições de possibilidade atribuídas a um conjunto de UEA, aos quais foram aplicadas as medidas de não-especificidade e a medida de incerteza RI. Os dados reais resultaram da classificação de uma imagem IKONOS, cujas características se apresentam no Anexo A, em que foi utilizada uma abordagem de classificação *fuzzy* orientada por objectos e em que a unidade espacial de análise é o objecto, por vezes também referido no texto como *entidade geográfica*.

A Figura 2.2 mostra os resultados, ordenados por ordem crescente dos valores de NSp, obtidos para um conjunto de distribuições de possibilidade, constituídas por valores reais, associadas a objectos classificados, escolhidos de forma aleatória.

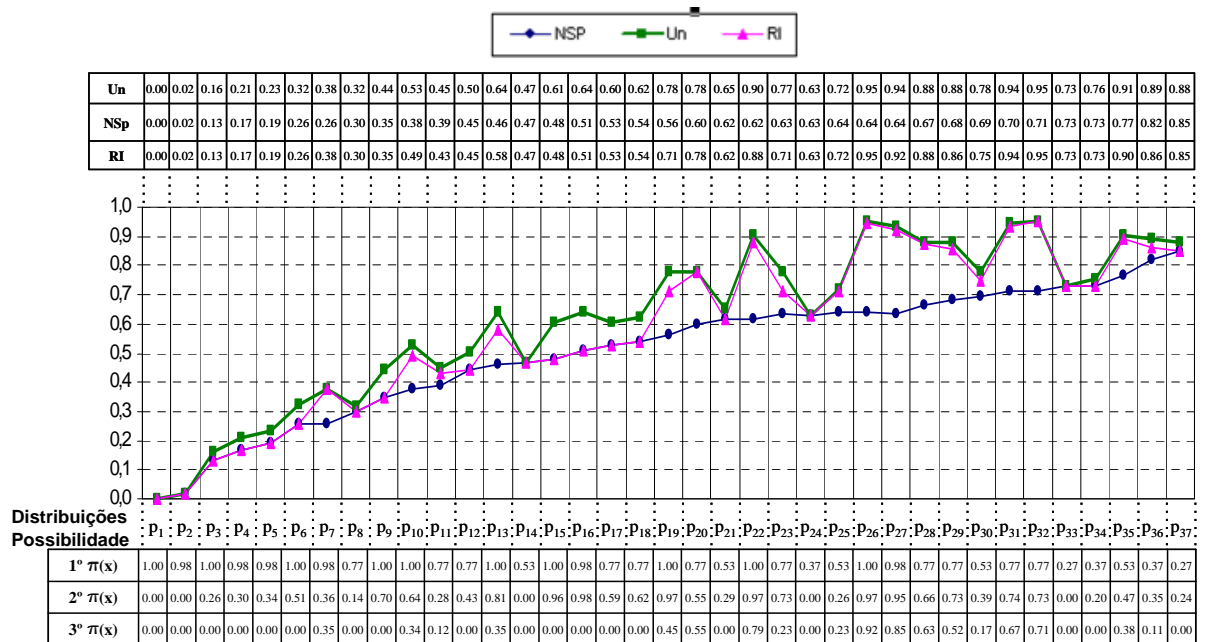


Figura 2.2 – Valores de NSp, Un e RI, ordenados por ordem crescente dos valores de NSp, para um conjunto de distribuições de possibilidade associadas a objectos numerados de 1 a 37, considerando três possíveis classes.

Pode-se observar que todas as medidas apresentam uma tendência semelhante verificando-se um aumento da incerteza para distribuições $\pi(x)$ com baixos graus de possibilidade. Contudo, existem algumas diferenças consideráveis entre elas. Para as distribuições de possibilidade apresentada na Figura 2.2, Un e RI dão valores mais elevados do que a medida NSp. Só são obtidos valores iguais quando apenas um grau de possibilidade diferente de zero está associado à entidade geográfica, o que significa que apenas uma classe é atribuída com um qualquer grau de possibilidade. A medida de incerteza RI dá resultados semelhantes à medida Un, quando o objecto é atribuído às três classes possíveis com um grau de possibilidade diferente de zero. Nas restantes situações aproxima-se da medida NSp. Para algumas entidades geográficas, os valores dados pelas medidas NSp e Un diferem consideravelmente. Por exemplo, para o objecto p_{26} , em que os três graus de possibilidade são 1, 0.97 e 0.92, o valor da incerteza com a medida NSp é de 0.64, enquanto com as medidas Un e RI é de 0.95.

Para analisar o comportamento das medidas quando o número teoricamente possível de classes aumenta, foram consideradas nove classes como teoricamente possíveis em que os valores não nulos da distribuição de possibilidade são os mesmos que no exemplo apresentado na Figura 2.2 e todos os graus de possibilidade adicionais são iguais a zero. Esta distribuição de possibilidade significa que a dispersão da classificação diminui, pois o mesmo número de possibilidades

positivas é obtido considerando um maior número de classes teoricamente possíveis. Os resultados obtidos apresentam-se na Figura 2.3. Os valores de NSp mantêm-se inalterados, relativamente aos apresentados na Figura 2.2 em que foram consideradas apenas três classes, enquanto os valores de Un e RI diminuem consideravelmente, tornando-se, em geral, inferiores aos valores obtidos com a medida NSp. Este facto mostra que as medidas Un e RI são sensíveis à dispersão da classificação sobre todas as classes consideradas como teoricamente possíveis, contrariamente à medida NSp. Pode concluir-se que Un e RI dependem da dispersão de possibilidades considerando todas as classes e em que menos dispersão significa menos incerteza, enquanto NSp apenas avalia a incerteza tendo em consideração as classes com graus positivos de possibilidade. A medida RI embora acompanhe a tendência da medida Un, apresenta uma maior sensibilidade à variação da dispersão que a medida Un, apresentando os valores mais baixos. Por exemplo, para o objecto p_{15} , ao qual foram atribuídas apenas duas classes com graus de possibilidade diferentes de zero e muito próximos, 1 e 0.96, o valor da incerteza com a medida NSp é de 0.48, com a medida Un é de 0.30, enquanto com a medida RI o valor da incerteza é apenas de 0.12.

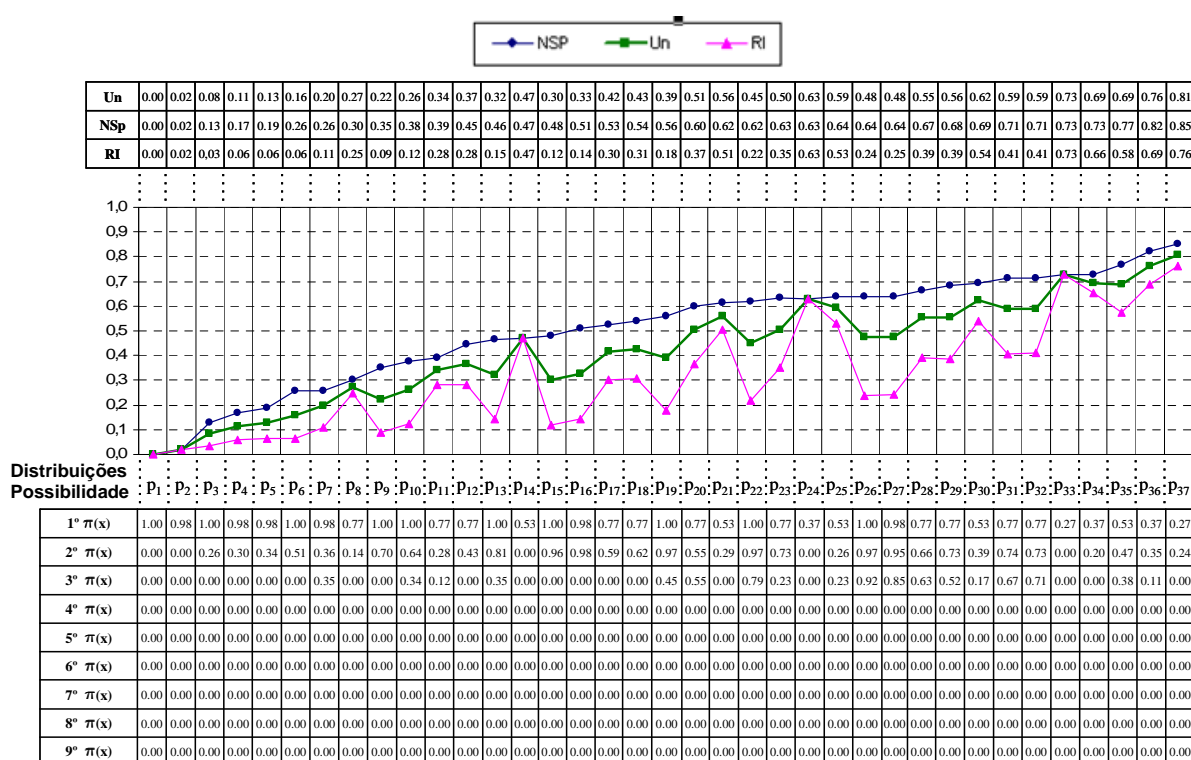


Figura 2.3 – Valores de NSp, Un e RI, ordenados por ordem crescente dos valores de NSp, obtidos para um conjunto de distribuições de possibilidade associadas a objectos numerados de 1 a 37, considerando nove classes como teoricamente possíveis e possibilidades nulas para as classes da quarta à nona.

Para permitir avaliar melhor o comportamento e a sensibilidade das medidas à dispersão da classificação, foram efectuadas experiências adicionais com valores simulados. Foram consideradas, igualmente, 9 classes como teoricamente possíveis e seleccionadas 10 distribuições de possibilidade com diferentes graus de dispersão sobre todas as classes, em que os valores para os três melhores graus de possibilidade são sempre iguais e muito elevados, e os restantes valores vão aumentando gradualmente para cada distribuição de possibilidade. O objectivo consistiu em avaliar o comportamento das medidas NSp, Un e RI com a variação da dispersão. Os resultados obtidos apresentados na Figura 2.4 mostram que, com a variação da dispersão, a variação dos valores de incerteza com as medidas Un e RI é maior do que a obtida com a medida NSp. Para esta medida, a diferença dos valores da incerteza obtidos, por exemplo, para as distribuições de possibilidade p_2 e p_{10} é apenas de 0.17, enquanto para Un é de 0.40 e para RI é de 0.60.

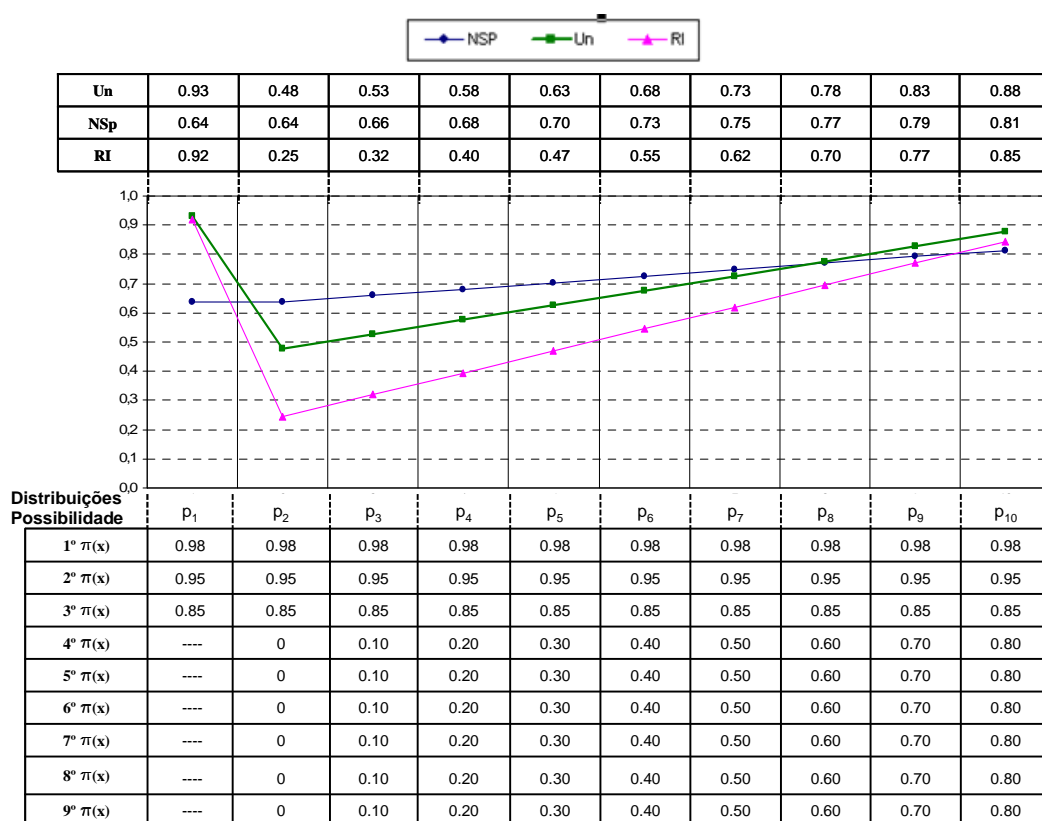


Figura 2.4 – Comportamento das medidas de incerteza NSp, Un e RI para distribuições de possibilidade com diferentes graus de dispersão para todas as classes.

Outro aspecto a salientar é o comportamento das medidas relativamente à compatibilidade entre os resultados da classificação e as classes. A Figura 2.5 apresenta o comportamento das medidas NSp, Un e RI, para distribuições de possibilidade com diferentes graus de

compatibilidade às classes, mas com a mesma variação entre os graus consecutivos de possibilidade. De notar que as medidas de incerteza Un e RI mantêm-se inalteradas, o que significa que não são sensíveis à compatibilidade entre as características dos objectos às classes mas à variação entre os graus de possibilidade da distribuição de possibilidade, enquanto o valor da incerteza com a medida NSp aumenta com a diminuição dos valores dos graus de possibilidade.

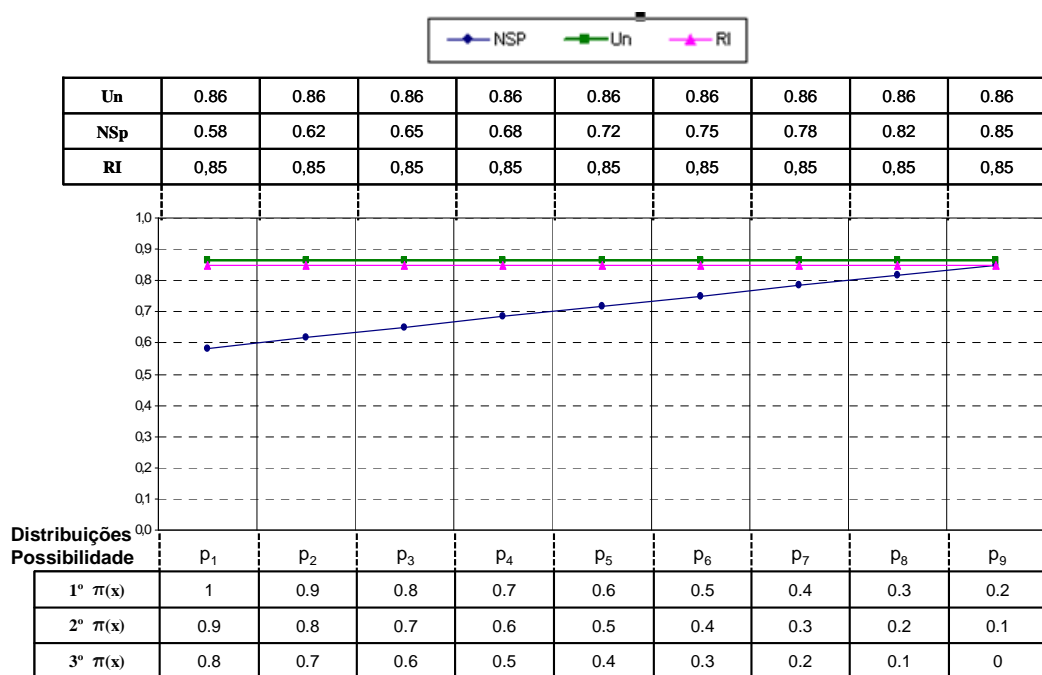


Figura 2.5 – Comportamento das medidas de incerteza NSp , Un e RI para distribuições de possibilidade em que os diferentes graus consecutivos apresentam uma variação de valores de possibilidade idêntica.

A Figura 2.6 ilustra o comportamento das medidas considerando diferentes distribuições de possibilidade em que, para cada uma, os valores são todos iguais, ou seja, são atribuídas possibilidades iguais a todas as classes, o que significa que a confusão entre elas é máxima.

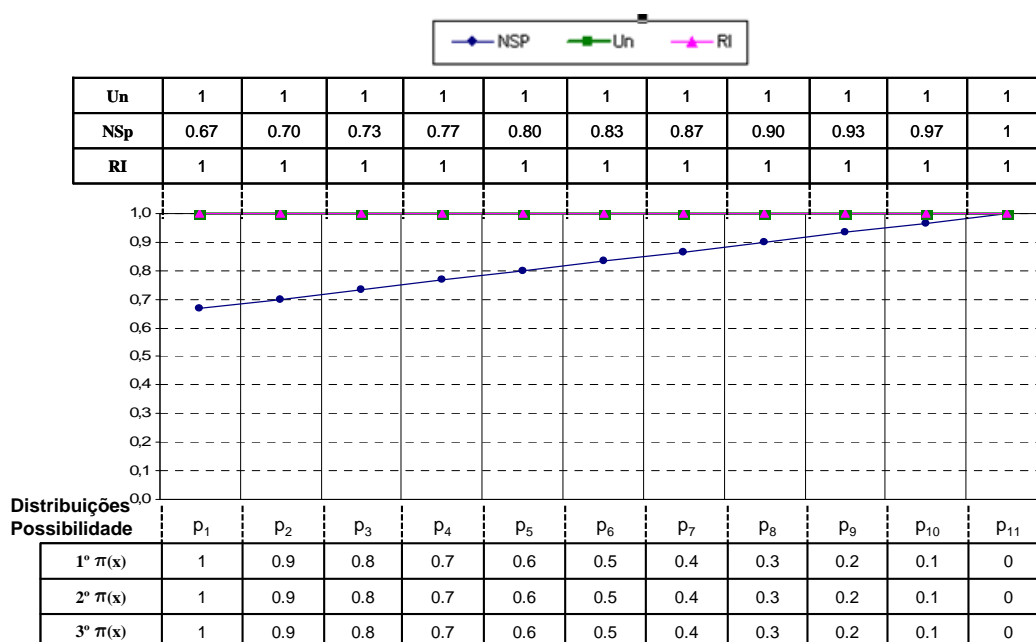


Figura 2.6 – Comportamento das medidas de incerteza NSp, Un e RI para distribuições de possibilidade em que todos os graus de possibilidade são iguais em cada distribuição de possibilidade.

Nesta situação, as medidas de incerteza Un e RI assumem o valor unitário para todas as distribuições de possibilidade consideradas, o que significa que, para estas medidas, a incerteza é máxima, contrariamente à medida NSp que apenas assume o valor máximo quando todos os valores da distribuição de possibilidade são zero. A partir da Figura 2.6, é também visível que as medidas Un e RI, quando existe a mesma variação entre todos os graus consecutivos de possibilidade, não são sensíveis aos valores dos graus de possibilidade, mas apenas à variação entre os diferentes graus de possibilidade que constituem a distribuição de possibilidade.

Por outro lado, NSp varia entre 0.67 e 1, o que demonstra a sua sensibilidade à variação da compatibilidade entre as características dos objectos e as classes, mesmo que a variação entre todos os graus consecutivos de possibilidade seja igual.

Estes resultados evidenciam que a informação disponibilizada por estas medidas ao utilizador é diferente. NSp é mais sensível à compatibilidade entre as classes e as características dos objectos, enquanto as medidas Un e RI são mais sensíveis à diferença entre os graus consecutivos de possibilidade e à dispersão da classificação sobre o total de classes consideradas teoricamente possíveis.

Da análise efectuada, pode-se concluir que a sensibilidade das medidas de incerteza Un e RI à dispersão de possibilidades pelas classes teoricamente possíveis implica que, quanto mais

pequeno o número de classes com possibilidades positivas atribuídas a uma entidade geográfica, menor é o valor de incerteza. A medida RI é mais sensível à variação da dispersão da classificação do que a medida Un. A medida de incerteza NSp é mais sensível aos valores dos graus de possibilidade, avalia principalmente a compatibilidade da entidade geográfica com as classes e não a dispersão pelas várias classes. Mais ainda, as medidas Un e RI assumem valores elevados quando: (1) todos os valores na distribuição de possibilidade são iguais; (2) as diferenças entre todos os valores dos graus de possibilidade consecutivos são pequenas; (3) são atribuídos graus de possibilidade com valores elevados a todas as classes consideradas teoricamente possíveis; e (4) são atribuídos graus de possibilidade com valores baixos a todas as classes consideradas teoricamente possíveis. A medida NSp assume valores elevados quando: (1) são atribuídos graus de possibilidade com valores baixos a todas as classes consideradas teoricamente possíveis; e (2) são atribuídos graus de possibilidade com valores elevados a todas ou à maioria das classes consideradas teoricamente possíveis. Estas medidas atribuem baixa incerteza às entidades geográficas quando a possibilidade mais elevada é 1, ou próximo de 1, e os restantes valores da distribuição de possibilidade são zero ou perto de zero.

2.4.2 Medidas aplicadas a distribuições de probabilidade

Nesta secção é feita uma análise ao comportamento da medida de entropia relativa, E_r , e do coeficiente de incerteza RI quando aplicados a distribuições de probabilidade. Para avaliar e ilustrar o comportamento destas duas medidas de incerteza, foram utilizados vários conjuntos de valores reais e sintéticos, os quais simulam distribuições de probabilidade atribuídas a um conjunto de UEA. O conjunto de valores reais resultou da classificação de uma imagem IKONOS (ver Figura 3.2) cujas características se apresentam na Tabela 3.1, em que foi utilizado o classificador da máxima verosimilhança não rígido, tendo sido obtidos os graus de probabilidade *a posteriori*, e em que a unidade espacial de análise é o *pixel*.

Na Figura 2.7 apresenta-se o comportamento das medidas RI e E_r para os valores sintéticos considerados. Foram consideradas 10 classes como teoricamente possíveis e seleccionadas 12 distribuições de probabilidade, que simulam situações em que ocorre dispersão sobre todas as classes i.e., todas as classes apresentam uma probabilidade de pertencerem a uma dada UEA. Para cada distribuição de probabilidade os valores sintéticos vão variando gradualmente até à situação extrema em que todas as classes apresentem igual probabilidade de ficarem atribuídas a uma dada UEA. O objectivo consistiu em avaliar o comportamento das medidas E_r e RI com a

variação dos valores dos graus de probabilidade quando a dispersão é total. Os resultados obtidos encontram-se ordenados por ordem crescente dos valores de RI. Pode-se observar que ambas as medidas apresentam uma tendência semelhante, verificando-se um aumento da incerteza com a diminuição dos graus de probabilidade, atingindo-se o máximo para a distribuição de probabilidade em que os valores são todos iguais. Consta-se também que Er dá valores muito mais elevados do que a medida RI, apresentando valores de incerteza superiores a 0.50 para todas as distribuições de probabilidade, à excepção da primeira, em que o primeiro grau de probabilidade associado à entidade geográfica é 0.90. Contudo, à medida que o valor do primeiro grau de probabilidade, associado à entidade geográfica, vai diminuindo e a diferença entre os graus consecutivos é menor, o comportamento da medida RI aproxima-se de Er.

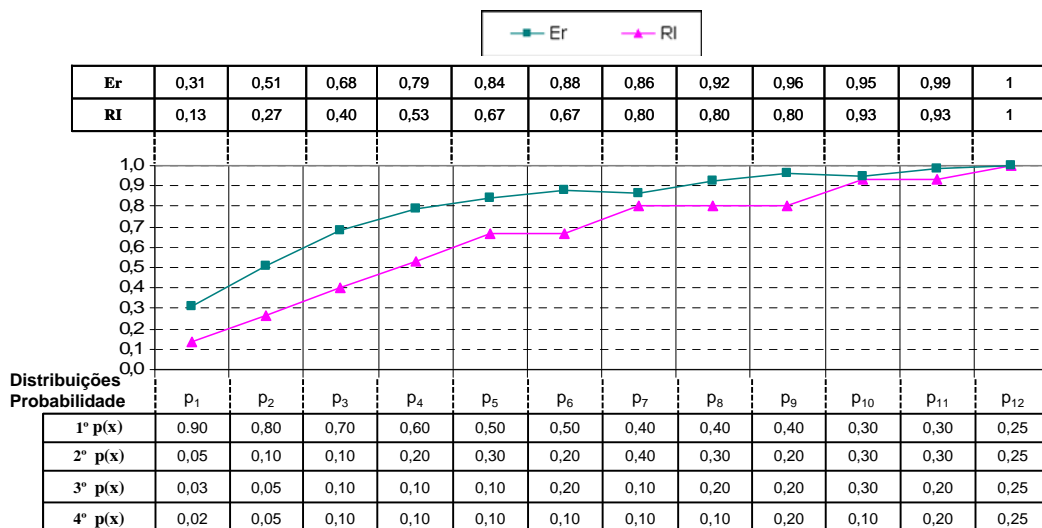


Figura 2.7 – Valores de Er e RI, ordenados por ordem crescente dos valores de RI, obtidos para distribuições de probabilidade numeradas de 1 a 12, considerando quatro possíveis classes.

Para analisar o comportamento das medidas quando o número teoricamente possível de classes aumenta, foram consideradas, igualmente, 10 classes como teoricamente possíveis em que os valores não nulos da distribuição de probabilidade são os mesmo que os do exemplo anterior e todos os graus de probabilidade adicionais são iguais a zero. Os resultados obtidos apresentam-se na Figura 2.8.

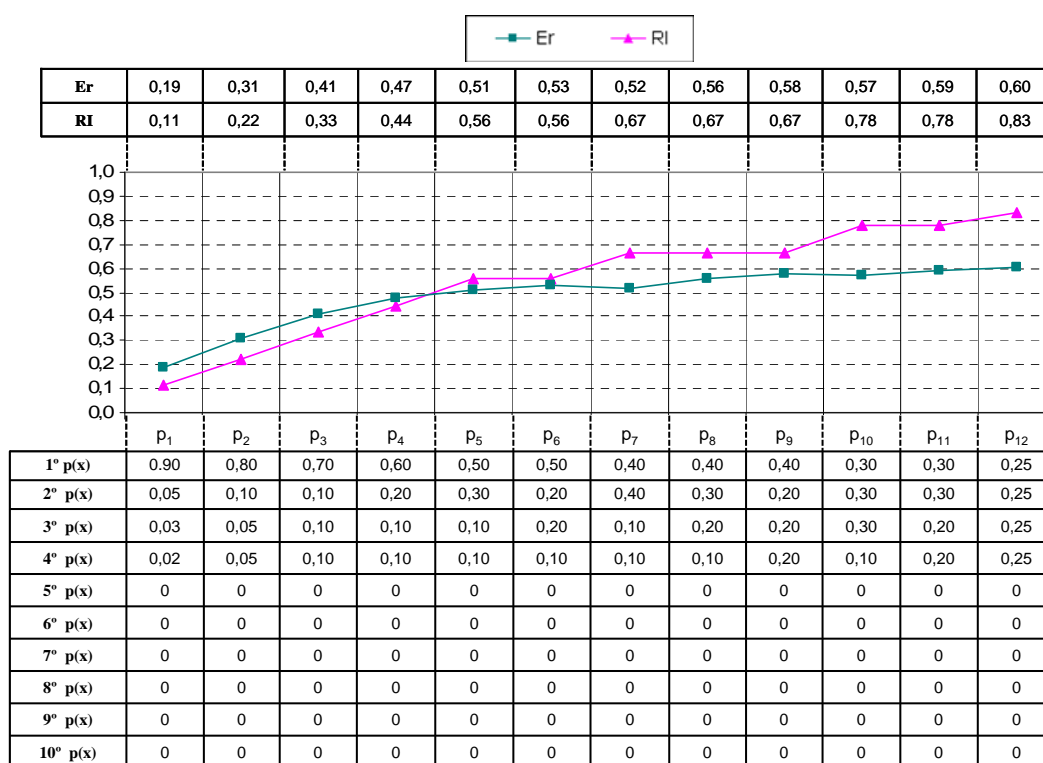


Figura 2.8 – Valores de Er e RI obtidos para distribuições de probabilidade numerados de 1 a 12 considerando dez possíveis classes e probabilidades nulas para as classes da quarta à décima.

Para estas distribuições de probabilidade, em que a dispersão da classificação é menor, pois o mesmo número de probabilidades positivas é obtido considerando um maior número de classes teoricamente possíveis, os valores de ambas as medidas diminuem consideravelmente. Este facto mostra que ambas as medidas são sensíveis à dispersão da classificação sobre todas as classes consideradas como teoricamente possíveis e dependem da dispersão de probabilidades considerando todas as classes, em que menos dispersão significa menos incerteza. A medida RI apresenta valores mais elevados do que Er, à medida que o valor do primeiro grau de incerteza diminui e a diferença entre os graus consecutivos é menor. Este facto significa que a medida RI, quando aplicada a distribuições de probabilidade, embora apresente sensibilidade à dispersão, é também muito sensível à compatibilidade entre as características da entidade geográfica e as classes, nomeadamente, ao valor do primeiro grau de probabilidade associado à entidade geográfica. Pode ainda concluir-se que o comportamento destas medidas é muito semelhante para distribuições de probabilidade em que a dispersão é baixa e o máximo grau de probabilidade é elevado (e.g. distribuições de probabilidade p_1, p_2, p_3, p_4). Contudo, à medida que o valor do maior grau de probabilidade diminui, o comportamento das medidas difere significativamente. Por exemplo, para a incerteza obtida para a distribuição de probabilidade p_4 , em que o maior grau de probabilidade é de 0.60, a medida Er apresenta um valor de 0.47 e a

medida RI de 0.44, ao passo que para a distribuição de probabilidade p_{12} , em que o máximo grau de probabilidade é de 0.25, o valor da incerteza com a medida Er é de 0.60 e com a medida RI é de 0.83.

A Figura 2.9 e a Figura 2.10 mostram os resultados, ordenados por ordem crescente dos valores de RI, obtidos para um conjunto de distribuições de probabilidade, constituídas por valores reais, associadas a *pixels* classificados, escolhidos de forma aleatória.

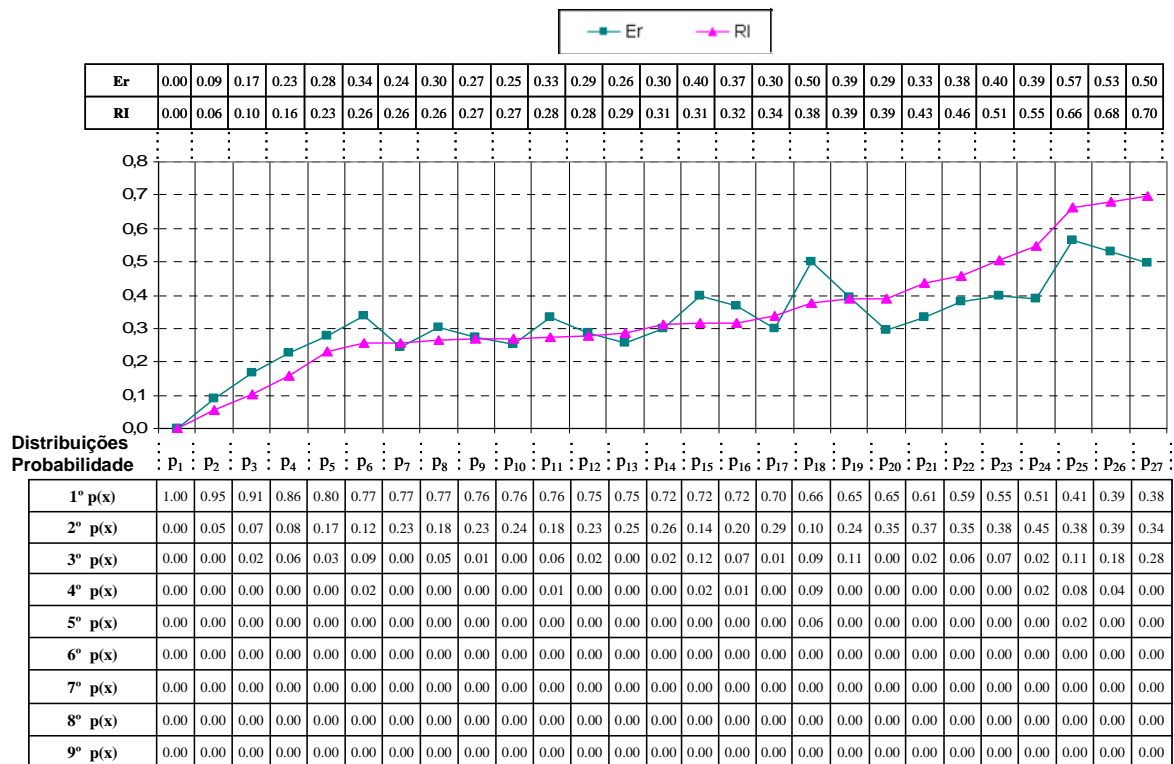


Figura 2.9 – Comportamento das medidas de incerteza Er e RI para distribuições de probabilidade associadas a *pixels* numerados de 1 a 27, correspondendo a diferentes graus de dispersão para todas as classes, ordenados por ordem crescente dos valores de RI.

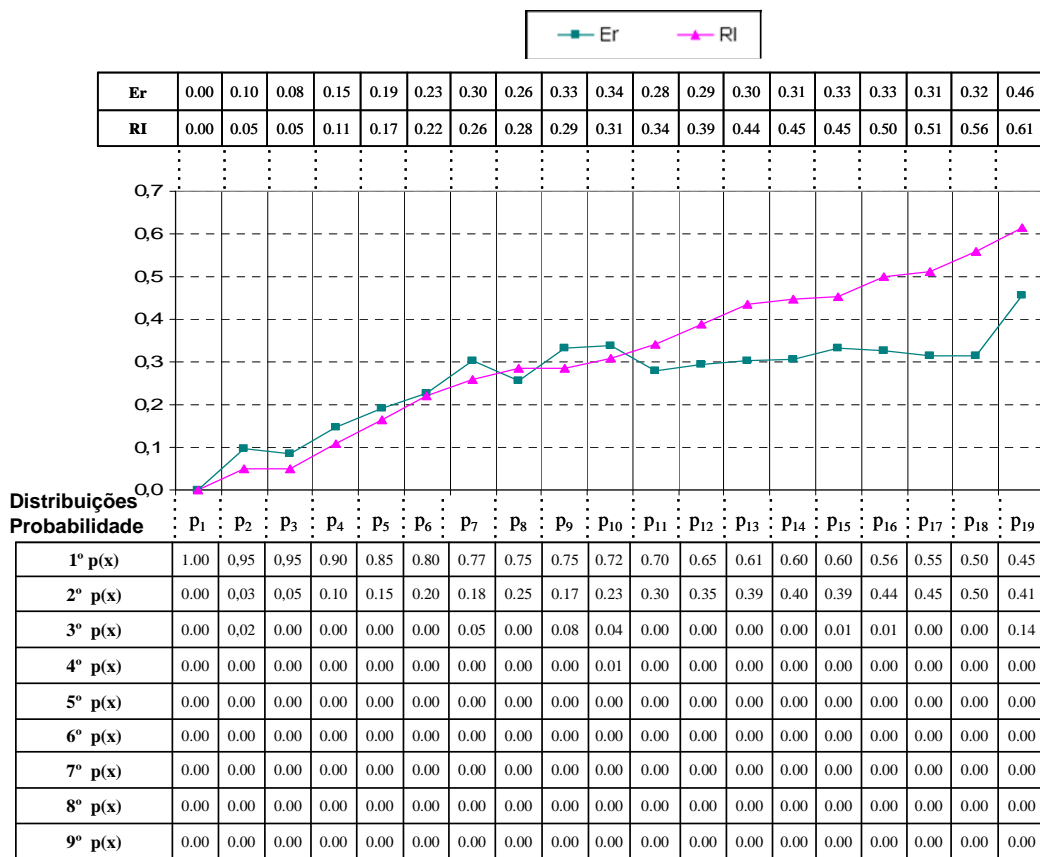


Figura 2.10 – Comportamento das medidas de incerteza Er e RI para distribuições de probabilidade associadas a pixels numerados de 1 a 19, correspondendo a diferentes graus de dispersão para todas as classes, ordenados por ordem crescente dos valores de RI.

Estas figuras evidenciam que a medida RI apresenta valores mais elevados do que Er, à medida que o valor do primeiro grau de incerteza diminui e a diferença entre os graus consecutivos é menor. Por exemplo, a incerteza obtida para a distribuição de probabilidade p_{18} da Figura 2.10, em que o maior grau de probabilidade é de 0.50, o valor da incerteza com a medida Er é de apenas 0.32 e com a medida RI é de 0.56 mas, para a distribuição de probabilidade p_6 , em que o máximo grau de probabilidade é de 0.80, o valor da incerteza com a medida Er é de 0.23 e com a medida RI é de 0.22. Das Figuras 2.9 e 2.10 constata-se que, quando o valor do maior grau de probabilidade é elevado, estas duas medidas apresentam um comportamento semelhante mas, quando o primeiro grau de probabilidade começa a diminuir e apresenta valores abaixo dos 0.70, os comportamentos das duas medidas distanciam-se significativamente. Na Figura 2.9 pode também constatar-se que a medida Er é mais sensível à dispersão do que a medida RI. Por exemplo, para a distribuição de probabilidade p_{18} que apresenta uma atribuição partilhada por 5 classes e em que o primeiro grau de probabilidade é 0.66, o valor de Er é de 0.50 e o valor de RI é de 0.38.

Da análise efectuada pode-se concluir que a sensibilidade das medidas de incerteza E_r e R_I à dispersão de probabilidades pelas classes teoricamente possíveis implica que, quanto menor o número de classes com probabilidades positivas atribuídas a uma UEA, menor é o valor de incerteza. Pode ainda concluir-se que as medidas E_r e R_I assumem valores elevados quando: (1) todos os valores na distribuição de probabilidade são iguais a $1/n$, em que n representa o número de classes; (2) são atribuídos graus de probabilidade com valores baixos a todas as classes consideradas teoricamente possíveis. A medida R_I é mais sensível ao valor do primeiro grau de pertença e a medida E_r é mais sensível à dispersão. Estas medidas atribuem baixa incerteza às UEAs quando a probabilidade mais elevada é um, ou próximo de um, e os restantes valores da distribuição de probabilidade são zero ou perto de zero.

2.5 Conclusões

No processo de produção automática de cartografia temática a partir das IMGRE, a obtenção de diferentes graus de probabilidade, possibilidade ou pertença a uma ou várias classes, atribuídos às Unidades Espaciais de Análise (UEAs) (*pixels* ou *objectos*), depende da natureza dos classificadores aplicados. Os graus de atribuição às classes, associados às UEAs podem ser interpretados como a dificuldade do classificador em atribuir uma classe à UEA. Esta interpretação é válida assumindo que cada UEA conterà apenas uma classe e que a classe que lhe é atribuída será aquela a que corresponde o maior grau de probabilidade ou possibilidade, dependendo do classificador aplicado. Nesta situação, as distribuições de probabilidade (ou possibilidade, quando for caso disso) associadas a cada UEA, permitem obter informação sobre a incerteza da classificação que foi feita. A aplicação de medidas de incerteza a cada distribuição de probabilidade ou possibilidade vai permitir captar a compatibilidade e/ou dispersão da atribuição das classes a cada UEA e obter um valor que fica associado a cada UEA, o qual reflecte a incerteza dessa atribuição. A escolha das medidas de incerteza a aplicar depende do método de classificação e do tipo de informação sobre incerteza que se pretenda avaliar. Se a cada UEA ficar associada uma distribuição de possibilidade, para obter informação sobre a incerteza dessa atribuição, podem ser aplicadas medidas de não-especificidade. Se a cada UEA ficar associada uma distribuição de probabilidade, as medidas baseadas na entropia de Shannon são as apropriadas. Recentemente, foi proposta por Clark Lab. (2001) uma medida que pode ser aplicada a distribuições de possibilidade ou probabilidades, disponível no *software* IDRISI. De referir que esta medida não considera como hipóteses condições que excluam a sua aplicação

quer a distribuições de probabilidade, quer a distribuições de possibilidade, ao contrário de outras medidas, as quais só podem ser aplicadas a distribuições de probabilidade ou a distribuições de possibilidade.

As medidas de não-especificidade que podem ser aplicadas a conjuntos *fuzzy* e a distribuições de possibilidade foram propostas por Yagger (1982) (especificidade Sp e não-especificidade NSp) e por Higashi e Klir (1982) (incerteza- U). Ricotta (2005) propôs a aplicação das medidas NSp e incerteza- U na análise da incerteza da classificação de imagens de detecção remota, fazendo uma abordagem apenas teórica. No estudo apresentado neste capítulo, procedeu-se a uma análise comparativa das três medidas (Un , NSp e RI) tendo sido efectuada a sua aplicação a distribuições de possibilidade obtidas a partir de uma classificação *fuzzy*. O estudo permitiu concluir que as várias medidas têm comportamentos diferentes e disponibilizam informação diferente ao utilizador. Foi proposta uma nova métrica, Un , a qual corresponde à normalização da medida de incerteza- U , tendo sido usada para permitir a comparação com as medidas NSp e RI e para estabelecer um intervalo de variação mais compreensível para o utilizador.

A análise do comportamento das medidas NSp , Un e RI mostrou que:

- A medida de incerteza NSp é sensível aos valores absolutos da distribuição de possibilidade, enquanto as medidas Un e RI são sensíveis à diferença entre os graus consecutivos de possibilidades e à dispersão;
- A medida NSp avalia a incerteza considerando apenas as classes às quais foram atribuídas possibilidades positivas, enquanto as medidas Un e RI são sensíveis à dispersão entre todas as classes consideradas teoricamente possíveis, mesmo que tenham sido atribuídos graus de possibilidade com valores iguais a zero a alguma destas.

Pelas razões atrás enunciadas, pode concluir-se que: para avaliar a dispersão ou a variação entre os graus consecutivos de possibilidade, a medida Un é a mais adequada; e para avaliar a compatibilidade entre as características das UEAs e as classes, deverá ser adoptada a medida NSp . A medida RI apresenta um comportamento próximo da medida Un , embora seja mais sensível à variação da dispersão.

A entropia de Shannon (1948) é aplicável a distribuições de probabilidade, sendo considerada a única medida que capta, de forma adequada, a incerteza no âmbito da teoria das probabilidades (Klir, 2000). Maselli *et al.* (1995) propuseram uma métrica que corresponde à normalização da entropia de Shannon para estabelecer um intervalo de variação mais compreensível para o

utilizador, a qual foi adoptada neste estudo e designada por *entropia relativa*, E_r . A análise comparativa do comportamento das medidas E_r e R_I aplicadas a distribuições de probabilidade, permitiu concluir que nem sempre estas medidas apresentam comportamentos semelhantes.

A análise comparativa do comportamento das medidas E_r e R_I , mostrou que:

- as duas medidas são sensíveis à dispersão. Contudo, a medida E_r é mais sensível do que a R_I . Por outro lado, a medida R_I é mais sensível ao valor atribuído ao melhor grau de probabilidade do que a medida E_r .

A medida R_I , quando aplicada a distribuições de probabilidade, é mais sensível ao valor atribuído ao melhor grau de probabilidade. Este comportamento deve-se ao facto de, na expressão (2.25), o somatório dos vários graus de probabilidade assumir o valor 1, $\sum_{i=1}^n p_i = 1$ e de a medida passar a depender apenas do valor atribuído ao melhor grau de probabilidade e do número de classes consideradas.

O estudo comparativo do comportamento das várias medidas, efectuada neste capítulo, permitiu evidenciar as características de cada uma, tendo assumido uma importância significativa na tomada de decisão relativamente à escolha das medidas a adoptar nas fases seguintes da investigação, apresentadas nos capítulos 3 e 4 desta tese. Outro aspecto relevante desta análise foi o ter clarificado o tipo de informação que se pode inferir sobre a incerteza quando se aplica este tipo de medidas aos resultados da classificação, no âmbito da detecção remota, obtidos com classificadores não rígidos.

Capítulo 3

Aplicação de medidas de incerteza no processo de produção de mapas de unidade de paisagem

3.1 Introdução

O trabalho apresentado neste capítulo teve como objectivos: (1) investigar a utilidade da integração da informação sobre a incerteza na abordagem de classificação combinada para aumentar a exactidão da classificação de unidades de paisagem; e (2) analisar o potencial da aplicação das medidas de incerteza na avaliação do comportamento de classificadores e da classificação de imagens multiespectrais, obtidas a partir de classificadores não rígidos.

No estudo realizado, desenvolveu-se e testou-se uma abordagem que adopta o princípio da classificação combinada *pixel/objecto* mas que integra a informação da incerteza associada à classificação ao nível do *pixel* no processo de classificação ao nível do objecto. A abordagem de classificação que combina *pixels* e objectos tem mostrado ser adequada para a identificação das classes de unidade de paisagem que contêm uma variedade de coberturas de ocupação do solo, a partir das IMGRE (e.g. Wang *et al.*, 2004), tendo sido testada com sucesso por Plantier e Caetano (2007) na identificação das principais espécies florestais. No entanto, a integração da informação sobre a incerteza neste método de classificação nunca foi efectuada. Na nova abordagem desenvolvida, estudou-se se a integração da informação da incerteza no processo de classificação combinada contribui para reduzir a influência da incerteza associada à classificação ao nível do *pixel* e para melhorar significativamente os resultados finais da classificação ao nível do objecto. Como é difícil discriminar as diferentes espécies florestais com as IMGRE (e.g. Goetz *et al.*, 2003; Carleer e Wolff, 2004), neste estudo é analisado se a introdução da incerteza no processo de classificação combinada melhora os resultados da identificação das principais espécies florestais em Portugal, obtidos a partir deste tipo de imagens.

Existem vários algoritmos para classificar imagens de detecção remota. A avaliação da sua

adequação, a um objectivo específico, pode ser efectuada através de índices estatísticos obtidos a partir de matrizes de confusão tradicionais considerando um conjunto de unidades amostra (e.g. *pixels*, áreas). Tradicionalmente, para avaliar o comportamento do classificador considera-se um conjunto de amostras nas regiões representativas das classes, do tipo das utilizadas para treino do classificador. Para avaliação da classificação, qualquer Unidade Espacial de Análise (UEA) da imagem pode ser seleccionada para ser incluída no protocolo de amostragem. Este tipo de avaliação requer a existência de dados de referência, os quais correspondem aos dados considerados como verdade no terreno para o conjunto de unidades amostra, e, além disso, que a sua aplicação seja limitada a um conjunto de amostras, não permitindo a espacialização do erro. Este estudo pretende investigar se, para classificadores para os quais se possa obter informação sobre a incerteza, a avaliação do comportamento dos classificadores e da classificação também pode ser feita com índices de incerteza.

Para dar resposta ao segundo objectivo, na classificação ao nível do *pixel* foram aplicados dois classificadores não rígidos: um classificador probabilístico, tendo sido obtidos os graus de probabilidade *a posteriori* utilizando o teorema de Bayes, e um classificador *fuzzy*, baseado na distância mínima à média. A cada *pixel* ficaram associados diferentes graus de probabilidade e possibilidade às várias classes, de acordo com o tipo de classificador aplicado. Esta informação adicional pode ser interpretada como graus parciais de pertença às classes, se se considerar que o *pixel* contém mais do que um tipo de ocupação do solo e estes graus podem ser usados como indicadores da proporção relativa de cada uma das classes (Shi *et al.*, 1999; Ibrahim *et al.*, 2005). Contudo, se se considerar que um *pixel* apenas contém uma classe, esta informação adicional associada aos *pixels* pode ser utilizada como indicador da dificuldade em atribuir correctamente uma dada classe a um *pixel*. Neste estudo, os resultados dos classificadores foram interpretados como graus de probabilidade *a posteriori* e graus de possibilidade e, portanto, a cada *pixel* ficou associada uma distribuição de probabilidade ou uma distribuição de possibilidade, permitindo obter informação sobre a dificuldade em atribuir uma única classe a um *pixel*. Pretendeu-se assim obter uma classificação rígida à qual fica associada a informação sobre a incerteza da classificação. Às distribuições de probabilidade e possibilidade associadas aos *pixels* foram ainda aplicadas diferentes medidas de incerteza e analisada a relação e compatibilidade entre a informação fornecida pelas diferentes medidas e os índices de exactidão estatísticos. No âmbito da detecção remota, as medidas de incerteza mais utilizadas para indicar o grau de confiança associado à atribuição das classes aos *pixels*, e indicar a fiabilidade da classificação ao nível do *pixel*, têm sido as medidas de entropia, nomeadamente, a entropia de Shannon (e.g. Maselli *et al.*, 1994; Foody, 1995; Foody, 1996; Zhu, 1997). Contudo, quando o

resultado da classificação é obtido a partir de um classificador *fuzzy*, em que a cada *pixel* ficam associados graus de possibilidade, estas medidas não são adequadas (Ricotta, 2005). No âmbito da detecção remota apenas recentemente tem sido dada atenção à aplicação das medidas de não-especificidade apresentadas no Capítulo 2. Ricotta (2005) propôs, apenas de forma teórica, a aplicação das medidas de não-especificidade desenvolvidas por Yager (1982) e por Higashi e Klir (1982) para quantificar a incerteza associada a distribuições de possibilidade. Neste estudo, é testada e analisada a contribuição das medidas de não-especificidade no processo de avaliação de uma classificação *fuzzy*. Para avaliar o potencial da informação disponibilizada pelos índices de incerteza propostos, é feita uma análise comparativa com os resultados obtidos com os índices de exactidão estatísticos.

Neste capítulo começa-se por apresentar uma breve revisão bibliográfica sobre abordagens de classificação temática. Em seguida, apresenta-se uma descrição das imagens de satélite utilizadas, da nomenclatura dos mapas, dos protocolos de amostragem e das várias etapas da abordagem metodológica desenvolvida. Por fim, procede-se à apresentação e discussão dos resultados em que se expõe a avaliação do comportamento dos classificadores e da classificação com base na informação da incerteza utilizando medidas de não-especificidade, entropia relativa e desvio máximo relativo, abordadas no capítulo 2. Apresenta-se ainda uma análise comparativa dos resultados da avaliação efectuada com os índices de incerteza propostos e com os índices de exactidão estatísticos. No final do capítulo é apresentada uma análise comparativa dos resultados alcançados com a metodologia desenvolvida, que integra a incerteza no processo de classificação, e com uma metodologia semelhante mas em que a informação da incerteza não é considerada.

Para a realização do estudo foram utilizadas várias ferramentas, nomeadamente o *software*: (1) IDRISI, no qual estão disponíveis os classificadores aplicados e a medida de incerteza designada nesta tese por RI; (2) eCognition V3.2, para extrair os padrões espaciais da imagem (objectos) através de um processo de segmentação da imagem; (3) Matlab V 7.1, para o desenvolvimento de sub rotinas que permitiram aplicar as medidas de não-especificidade, N_{Sp} e U_n , e entropia relativa, ER, a todos os *pixels* da imagem classificada; (4) ArcInfo Workstation, para a programação das regras de classificação em linguagem AML (*Arc Macro Language*); (6) ArcGIS V 9.2 e Geomedia Professional V6.0, para as operações de cálculo e de análise espacial; e (7) PCI Geomatics V.10, para definição das áreas de treino, teste e de referência, e para o estudo da separabilidade espectral das classes.

3.2 Revisão bibliográfica

As potencialidades da detecção remota, como meio de disponibilização de dados para produção de cartografia de ocupação do solo, são amplamente reconhecidas dadas as suas vantagens que se prendem com: a facilidade em obter dados actualizados, a disponibilização em formato digital e a aquisição de informação de extensas zonas da superfície terrestre, incluindo locais de difícil acesso ou mesmo inacessíveis. Estes exemplos ilustram as capacidades oferecidas pela detecção remota, às quais os métodos convencionais de levantamento de campo para aquisição da informação temática, não conseguem dar resposta de forma adequada. Além disso, constituem uma alternativa muito dispendiosa, sobretudo pelos recursos humanos que requerem e pela morosidade do processo de foto-interpretação, sendo também muito condicionados pela falta de coerência entre os foto-intérpretes (Goodchild, 2003). No entanto, as vantagens da detecção remota por satélite não são absolutas, dependendo do tipo de imagens, do campo de aplicação, da nomenclatura a utilizar e do tipo de abordagem.

Com o objectivo de encontrar uma metodologia que possibilite a extracção da informação das imagens de satélite de forma automática, capaz de reduzir as limitações inerentes aos métodos de interpretação visual, tem-se assistido ao desenvolvimento de várias teorias e métodos de análise quantitativa e de processamento, destacando-se, por exemplo, abordagens de análise contextual (Binaghi *et al.*, 1997; Stuckens *et al.*, 2000; Laha *et al.*, 2006), redes neuronais (Foody *et al.*, 1995; Foody e Arora, 1997; Foody, 2001; Erbek *et al.*, 2004), árvores de decisão (Friedl e Brodley, 1997; Friedl *et al.*, 1999), análise de imagem orientada por objectos (Benz *et al.*, 2004; Gonçalves e Caetano, 2004; Walter, 2004; Taubenböck *et al.*, 2006), redes neuro-fuzzy (Gamba e Dell'acqua, 2003; Qiu e Jensen, 2004), métodos SVM (Support Vector Machine)(Brown *et al.*, 2000; Foody e Mathur, 2006), algoritmos baseados em conhecimento (Hung e Ridd, 2002; Schmidt *et al.*, 2004) e abordagens que combinam múltiplos classificadores (Huang e Lees, 2004; Doan e Foody, 2007; Foody *et al.*, 2007).

Apesar dos consideráveis desenvolvimentos conseguidos, a exactidão dos mapas temáticos obtidos a partir de dados de detecção remota é muitas vezes considerada baixa para o seu uso operacional (Foody, 2008a). Existe uma série de razões que influenciam a qualidade dos mapas finais, como sejam: a natureza das classes, a resolução espectral e radiométrica dos dados de detecção remota e os métodos usados na extracção da informação (Foody, 1999). A classificação das imagens não é um processo simples pois nem sempre os objectos no terreno (e.g. cobertura de edifício, copa de árvore) têm a resposta espectral esperada, tanto devido a

condições de iluminação, como as condições atmosféricas (Baio, 1996; Erbek *et al.*, 2004), entre outras razões. O facto do uso do solo, nomeadamente nas áreas urbanas, ser definido mais em termos da sua função do que da sua forma, torna as relações entre o uso do solo e as respostas espectrais gravadas nas imagens muito complexa.

Com os progressos conseguidos nos últimos anos ao nível da detecção remota, uma variedade de sensores disponibilizam agora imagens de grande resolução espacial nas quais se enquadram as imagens adquiridas pelos satélites IKONOS, QuickBird e OrbView. Estudos realizados por vários autores têm feito notar que o aumento da resolução espacial levanta alguns problemas, tornando difícil a extracção da informação do uso do solo com base unicamente em informação espectral e que os métodos de classificação ao nível do *pixel* não são adequados para a identificação de alguns tipos de classes de ocupação do solo a partir das IMGRE (e.g. Schiewe *et al.*, 2001; Carleer e Wolff, 2004). A complexidade das relações entre *pixels* e unidades de paisagem, assim como o aumento de unidades de paisagem que são mosaicos de pequenas entidades ou arranjos espaciais de classes de ocupação do solo, como, por exemplo, a cobertura Agro-florestal, colocam novos desafios e impõem a necessidade de desenvolvimento de métodos que incorporem forma e contexto, à semelhança do processo de foto-interpretação (e.g. Wang *et al.*, 2004; Plantier e Caetano, 2007), e que permitam de igual modo uma maior integração com os Sistemas de Informação Geográfica (SIG). Neste sentido, vários estudos têm vindo a ser desenvolvidos com o objectivo de avaliar o potencial destas imagens para diferentes áreas de aplicação e investigar as abordagens mais adequadas para extracção da informação (e.g. Goetz *et al.*, 2003; Carleer e Wolff, 2004; Chubey *et al.*, 2006; Lu e Weng, 2007).

Pode considerar-se que as várias abordagens de análise dos dados multiespectrais desenvolvidas podem agrupar-se em quatro tipos: (1) classificação ao nível do *pixel* apenas com base na informação espectral; (2) classificação ao nível do *pixel* combinada com análise contextual e/ou da textura; (3) classificação ao nível do objecto; (4) classificação combinada *pixel*/objecto.

O termo objecto é aqui considerado como a representação física de uma entidade espacial do mundo real (e.g. edifício, copa de árvore). A sua forma de representação espacial depende da estrutura ou modelo conceptual utilizado. Se for utilizada uma estrutura vectorial, os objectos são representados graficamente por pontos, linhas ou polígonos. A estrutura matricial também pode ser utilizada para representar objectos através de um conjunto de células contíguas a que corresponde o mesmo atributo ou através de regiões às quais fica associado um valor de determinado fenómeno (ver Figura 3.1).

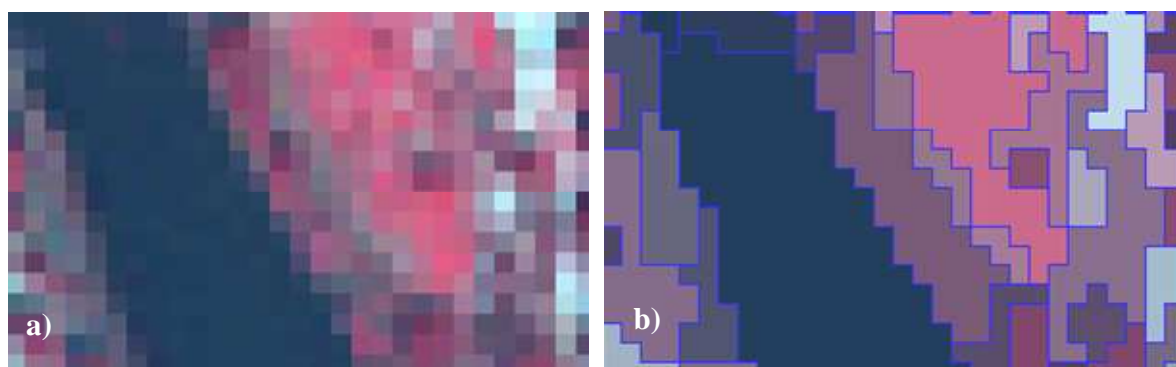


Figura 3.1 – Extracto de uma imagem IKONOS (RGB 432): a) estrutura matricial da imagem em que é visível a grelha regular de células designadas por pixels; b) representação de objectos na estrutura matricial sobrepostos pelos objectos vectoriais representados por polígonos (contornos a azul).

Algumas das motivações para o desenvolvimento de técnicas para a extracção de objectos da imagem, de forma automática, e de métodos de classificação ao nível do objecto, devem-se ao facto de (1) as imagens possuírem características de textura que são negligenciadas nos métodos mais comuns de classificação, (2) permitirem a identificação de unidades mínimas de ocupação do solo superiores à unidade do *pixel*, o que se adequa mais à conversão para formato vectorial onde os mapas têm uma determinada unidade mínima de representação de acordo com a escala; (3) facilitarem a integração de informação de contexto; (4) em muitas situações, a informação necessária para interpretar uma imagem não é representada ao nível do simples *pixel*, mas no significado dos objectos da imagem e nas suas relações mútuas; (5) aproximarem-se mais dos processos cognitivos humanos do que as análises baseadas no *pixel*. A visão humana geralmente divide primeiro as imagens em áreas ou regiões homogéneas, caracterizando-as posteriormente. Com efeito, a nossa percepção do conteúdo da imagem é principalmente baseada em objectos (Blaschke *et al.*, 2000). Os métodos de classificação, em que a unidade espacial é o objecto, utilizam uma abordagem semelhante, extraem primeiro os objectos, através de algoritmos de segmentação da imagem, procedendo posteriormente à sua classificação. A segmentação da imagem consiste na sua subdivisão em zonas, ou regiões, homogéneas e uniformes em função de um determinado atributo da imagem (Benz *et al.*, 2004; Blaschke *et al.*, 2004).

Hay *et al.* (2005), propuseram uma abordagem ao nível do objecto e desenvolveram um método de segmentação *Size Constrained Region Merging* (SCRM) para obter os objectos a diferentes escalas. O estudo pretendeu investigar as potencialidades do novo método para extrair informação das copas de árvores e unidades de paisagem florestal a partir de imagens IKONOS. Segundo os autores, os resultados obtidos com a análise exploratória desenvolvida permitiram

concluir que este tipo de abordagem apresenta boas potencialidades para a realização de inventários florestais.

Cleve *et al.* (2008) compararam uma abordagem de classificação ao nível do *pixel* com uma abordagem ao nível do objecto sobre imagens aéreas de grande resolução espacial para obter uma carta de ocupação de solo. O melhor resultado foi obtido com a classificação ao nível do objecto em que o valor da Exactidão Global foi cerca de 40% superior. Contudo, os autores concluem que, para se atingirem bons resultados de exactidão nos mapas, é importante seleccionar o método de classificação adequado às características da imagem e aos objectivos a atingir.

Guerrero *et al.* (2006) utilizaram uma abordagem ao nível do objecto para produzir um mapa de ocupação de solo utilizando imagens QuickBird. Os resultados obtidos evidenciaram dificuldades em discriminar as espécies florestais e o tipo de unidades de paisagem Mediterrânicas com este tipo de abordagem usando apenas informação multiespectral.

Vários autores propuseram métodos que combinam informação espectral e espacial como forma de melhorar a classificação a partir de imagens de muito grande resolução. Wang *et al.* (2004) investigaram três abordagens de classificação para extrair informação de diferentes tipos de espécies de árvores tropicais a partir de imagens IKONOS: (1) ao nível do *pixel* com o algoritmo da máxima verosimilhança; (2) ao nível do objecto com o algoritmo do vizinho mais próximo; e (3) um método de classificação combinada *pixel/objecto* em que a classificação ao nível do *pixel* foi utilizada para extrair a informação das classes espectralmente separáveis e a classificação ao nível do objecto foi usada para extrair as restantes classes. O melhor resultado da Exactidão Global foi obtido com o método de classificação combinada. Os autores referem também que existem vantagens e desvantagens associadas a ambas as abordagens de classificação *pixel* e objecto. Concluem que, na abordagem de classificação ao nível do objecto, a segmentação é um aspecto crítico pois não é possível que os objectos gerados contenham apenas *pixels* de uma única classe, gerando objectos mistos, principalmente nas zonas de fronteira entre dois tipos de ocupação de solo.

Plantier e Caetano (2007) utilizaram uma abordagem combinada *pixel/objecto* para classificar imagens IKONOS, visando a produção automática de mapas florestais à escala 1:10 000. Os resultados do estudo mostraram que este tipo de abordagem é adequada para transformar a informação espacialmente muito detalhada das IMGRE num mapa temático em formato vectorial, com uma Unidade Mínima Cartográfica (UMC) superior à do *pixel*. Contudo, na

transição do Mapa de Elementos de Superfície (MES) para o Mapa de Unidades de Paisagem (MUP), a exactidão decresceu em cerca de 20%, pelo que os autores sugerem que uma melhoria dos processos de segmentação e regras de classificação pode contribuir para aumentar a exactidão do mapa produzido.

Durante as últimas décadas, um dos focos de investigação tem sido o desenvolvimento e aplicação de métodos de classificação não rígidos, ou *fuzzy*, que permitam atribuir mais do que uma classe de informação a um *pixel* ou objecto (e.g. Maselli *et al.*, 1995; Brown *et al.*, 2000; Foody, 2000b; Zhang e Foody, 2001; Ibrahim *et al.*, 2005; Doan e Foody, 2007). O resultado da classificação com estes métodos não rígidos, para além de permitir obter informação adicional sobre a ocupação do solo, permite também a determinação de informação sobre a incerteza da classificação. Têm sido propostas abordagens de classificação totalmente *fuzzy* (*fully fuzzy classifications*), *i.e.*, que estimam a incerteza em todas as fases do processo de produção, considerando áreas de treino *fuzzy*, classificadores *fuzzy* e medidas *fuzzy* de avaliação da exactidão (e.g. Zhang e Foody, 2001; Ibrahim *et al.*, 2005). Apesar dos progressos alcançados, a aplicação dos classificadores não rígidos ou *fuzzy*, para extracção de informação a partir de imagens multiespectrais, continua a ser limitada pela falta de métodos adequados e bem sucedidos para avaliação da exactidão dos resultados da classificação, fase que, pela sua importância, deve ser considerada uma parte integrante do processo de classificação (Binaghi *et al.*, 1999; Carrão *et al.*, 2007).

A abordagem amplamente aceite e utilizada para estimar a qualidade da informação temática, obtida no processo de classificação automática de imagens de satélite e aéreas, baseia-se na determinação da matriz de erro ou confusão, a partir da qual se podem deduzir os Índices de Exactidão Temática Global (Exactidão Global, Coeficiente de Khat) e os Índices de Exactidão Temática Específica (Exactidão do Produtor e Exactidão do Utilizador) (Congalton e Green, 1999). As estatísticas globais descrevem o erro temático através de um único valor que é obtido com base no nível de concordância daquilo que se pretende avaliar com uma referência. Como é impraticável a avaliação da exactidão de todo o mapa, por comparação com uma referência, a avaliação é feita para uma amostra (*e.g. pixels*) que deve cumprir um conjunto de requisitos para que seja representativa e possa ser generalizada para todo o mapa (Stehman e Czaplewski, 1998; Foody, 2008b). Embora este tipo de informação relativamente à qualidade dos dados seja importante, apresenta algumas limitações conforme já referido na secção 3.1: (1) é influenciada pelo método de amostragem e tamanho da amostra; (2) não permite a visualização da distribuição espacial do erro; (3) quando a referência está dependente de um foto-intérprete

tem associada toda a subjectividade inerente à interpretação visual; e (4) a cada área do mapa apenas pode ser atribuída uma única classe de ocupação do solo com toda a certeza. Além disso, existem estudos que mostram que o valor do erro não é uniforme em todo o mapa, ou seja, existe uma variação espacial do erro (Liu *et al.*, 2004; Foody, 2005; Brown *et al.*, 2009) que não pode ser modelada de forma adequada pelas estatísticas de exactidão. Outra limitação deste tipo de abordagem é que apenas se adequa às classificações em que se assume que em cada localização do mapa existe uma única classe (Woodcock e Gopal, 2000; Lu e Weng, 2007). Este pressuposto nem sempre se verifica especialmente nas classificações temáticas de imagens de baixa resolução, em que numa área geográfica pode existir mais do que uma classe de ocupação de solo, ou quando são efectuadas classificações que fazem uso da lógica *fuzzy*. Embora a aplicação de estatísticas locais permitisse modelar de forma mais adequada a distribuição espacial da variação da qualidade (Foody, 2005), é difícil de implementar e, além disso, não permite um entendimento fácil por parte do utilizador (Brown *et al.*, 2009).

Gopal e Woodcock (1994) e Woodcock e Gopal (2000) desenvolveram uma abordagem para avaliação da exactidão temática baseada numa escala linguística com o objectivo de incorporar, na estimação da exactidão do mapa, a heterogeneidade da própria paisagem e a incerteza humana que existe na identificação das classes de ocupação de solo de referência. Esta abordagem disponibiliza uma informação mais alargada acerca da fonte de erros pois permite compreender a origem dos erros do mapa e verificar se esses erros são devidos à natureza heterogénea da paisagem, ao processo de classificação automática ou à incerteza inerente ao processo de elaboração de bases de dados de referência. Contudo, apresenta algumas desvantagens porque a avaliação é baseada na escolha de uma amostra, sendo ainda influenciada pela subjectividade do foto-intérprete, não permite uma representação espacial do erro e considera apenas a incerteza dos dados de referência.

Foody (1996) investigou a aplicação de duas medidas de proximidade (*measures of closeness*) que medem a '*distância*' entre a classificação e os dados de referência para inferirem o grau de semelhança entre a representação temática obtida a partir de uma classificação *fuzzy* e a existente no terreno. Uma das medidas é baseada na distância euclidiana, *i.e.*, a distância entre a proporção da classe atribuída a um *pixel* no terreno e a proporção obtida com a classificação *fuzzy*.

Binaghi *et al.* (1999) propõem um método que, baseando-se nas matrizes de erro tradicionais, alarga a sua aplicação à avaliação das classificações efectuadas com abordagens não rígidas ou

fuzzy. Um ponto crítico referido pelos autores relativamente à análise e avaliação das classificações *fuzzy* é a recolha de dados de referência. Mais estudos são necessários para encontrar uma abordagem que avalie os resultados das classificações *fuzzy*.

Recentemente começou-se a explorar a possibilidade de se disponibilizar de forma espacial informação sobre a incerteza ou confiança da atribuição de uma classe a um *pixel*. Neste tipo de abordagens é disponibilizado, para cada *pixel*, informação sobre a classe atribuída e uma medida de confiança dessa atribuição (e.g. Zhu, 1997). Maselli *et al.* (1994) investigaram a aplicabilidade da entropia relativa para estimar e representar graficamente a incerteza da classificação obtida, a partir de uma imagem Landsat TM, com o algoritmo da máxima verosimilhança. Segundo os autores, os graus de probabilidade *a posteriori*, obtidos para cada *pixel* com o algoritmo da máxima verosimilhança, podem expressar informação acerca do grau de confiança da atribuição do *pixel* à classe ou às várias classes. Esta abordagem é apenas apropriada para situações em que, no processo de classificação obtida com um classificador, um *pixel* pode ser atribuído com determinado grau de probabilidade a uma ou várias classes mas, no terreno, considera-se que existe apenas uma classe de ocupação do solo.

Segundo Schiewe e Ehlers (2007), a quantificação e visualização da incerteza contribuem para uma melhor percepção por parte do utilizador da qualidade e fiabilidade dos resultados produzidos. Contudo, e apesar dos desenvolvimentos ocorridos nas duas últimas décadas (e.g. Gahegan e Ehlers, 2000), o conhecimento sobre a incerteza existente nos dados geográficos continua a ser um problema (e.g. Veregin, 1995; Schiewe e Ehlers, 2007). No estudo desenvolvido no âmbito desta tese, investigou-se a utilidade das medidas de incerteza nomeadamente, as medidas de não-especificidade, NSp e Un, entropia relativa, ER, e a medida de incerteza RI, apresentadas no capítulo 2, para obter informação sobre a incerteza existente na classificação temática de imagens multiespectrais através de um processo de produção automático. Analisou-se se estas medidas permitem avaliar a ambiguidade em especificar uma solução única no processo de atribuição das classes a cada *pixel* da imagem e com base nessa informação avaliar o comportamento do classificador e igualmente estimar a exactidão da classificação. O processo de análise foi complementado com a espacialização da informação sobre a incerteza da classificação produzida.

3.3 Área de estudo e dados

A área de estudo situa-se no sul do País, numa zona com um relevo suave e ocupada essencialmente por sistemas agro-florestais, áreas agrícolas, pastorícia e floresta, e em que as principais espécies são o eucalipto, pinheiro manso e sobreiro.

A imagem de satélite utilizada foi obtida pelo sensor IKONOS no modo pancromático (Pan), com uma resolução espacial de 1m, e no modo multiespectral (XS), com resolução espacial de 4m. O estudo foi efectuado utilizando as imagens multiespectrais. Apresentam-se na Figura 3.2 a localização da área abrangida pela imagem de satélite e na Tabela 3.1 as suas características.

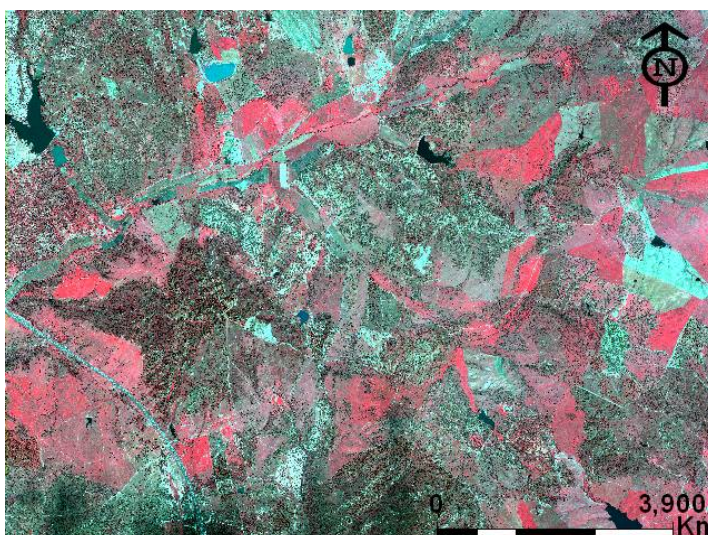


Figura 3.2 – Área de estudo: Imagem IKONOS (RGB 432).

Tabela 3.1 – Características da imagem IKONOS.

| | |
|------------------------------------|---------------|
| Data de aquisição | 06/04/2004 |
| Ângulo de elevação do satélite (°) | 74.8 |
| Ângulo de elevação do sol (°) | 55.5 |
| Dimensão (m x m) | 11884 x 14432 |
| Bits/pixel | 11 |

A correcção geométrica das imagens multiespectrais consistiu na sua ortorrectificação. O erro médio quadrático para a correcção geométrica foi 1.39 m, inferior a metade do tamanho do

pixel, o que garante uma correcta georreferenciação. Os *pixels* da imagem foram guardados com um formato de 16 *bits* para não perder a informação contida nos 11 *bits* da imagem original. Como em estudos unitemporais, sem efeito topográfico significativo e com atmosfera uniforme na área coberta pela imagem, não é necessária correcção radiométrica, por não melhorar os resultados (Caetano, 1995), no estudo descrito esta não foi efectuada.

3.4 Abordagem metodológica

O principal objectivo da classificação neste estudo é obter um Mapa de Unidades de Paisagem (MUP), a uma escala 1:10 000, usando uma abordagem de classificação combinada *pixel/objecto* com integração de incerteza. Com esta abordagem pretende-se investigar se a introdução da incerteza, associada à classificação dos elementos de superfície, pode melhorar os resultados da classificação das unidades de paisagem do MUP.

O método desenvolvido envolve uma classificação a dois níveis: a primeira ao nível do *pixel*, utilizando um classificador não rígido, para obter um Mapa de Elementos de Superfície (MES), em que a unidade básica é o *pixel*; e a segunda para obter o MUP, com uma escala 1:10 000, em que a unidade básica é o objecto. O primeiro mapa é executado para obter as entidades elementares, como por exemplo copas das árvores, partes das coberturas de edifícios, designadas por elementos de superfície, que são as unidades básicas das unidades de paisagem. O objectivo subjacente à produção do MES é obter uma classificação em que se pressupõe que a cada *pixel* corresponde um elemento de superfície no terreno.

Para produzir o MES testaram-se dois classificadores não rígidos: o classificador probabilístico da máxima verosimilhança, tendo-se obtido os graus de probabilidade *a posteriori* utilizando o teorema de Bayes (MV_B), e um classificador *fuzzy* baseado no princípio lógico do classificador da distância mínima à média (F_DM).

Para obter a incerteza da informação temática devido à ambiguidade ou dificuldade do classificador em atribuir no processo de classificação uma classe a um *pixel*, foram aplicadas às duas classificações diferentes medidas de incerteza. Ao resultado da classificação probabilística foram aplicadas duas medidas, a de Entropia Relativa (E_r), proposta por Maselli *et al.* (1994), e a medida de incerteza, desenvolvida por Clark Lab. (2001), designada neste estudo por *Ratio* de Incerteza (RI). Ao resultado da classificação obtida com o classificador *fuzzy*, foram aplicadas

três medidas: a medida de não-especificidade, NSp, desenvolvida por Yager (1982), a medida Un e a medida RI. A nova métrica, Un, baseia-se na medida desenvolvida por Higashi e Klir (1982) e foi proposta no segundo capítulo desta tese, para possibilitar uma comparação dos resultados com os obtidos com a aplicação da medida NSp. O indicador RI pode ser aplicado a distribuições de probabilidade e de possibilidades (Eastman 2006), ou seja, a medida não considera como hipótese condições que excluam a sua aplicação quer a distribuições de probabilidade, quer a distribuições de possibilidade. No estudo comparativo apresentado no Capítulo 2, pôde concluir-se que a medida RI, quando aplicada a distribuições de possibilidade, apresenta um comportamento próximo da medida Un, embora seja mais sensível à variação da dispersão. Quando aplicada a distribuições de probabilidade, embora apresente um comportamento próximo da medida Er, é mais sensível ao valor atribuído ao melhor grau de probabilidade.

Para elaborar o MES e avaliar os classificadores, utilizaram-se amostras representativas dos elementos de superfície, as quais foram divididas de forma aleatória em dois conjuntos: um para treino e outro para teste dos classificadores. A avaliação do comportamento dos dois classificadores foi efectuada com base num conjunto estratificado de amostras, com as quais se calcularam os Índices de Exactidão Global (IEG) e os Índices de Exactidão Específicos (IEE) e dois índices de incerteza obtidos a partir das medidas de incerteza, um baseado na média da incerteza por classe e outro nos baixos valores de incerteza.

A avaliação da exactidão temática do mapa obtido com o melhor classificador fez-se com base num conjunto aleatório estratificado de amostras, com as quais se calcularam os Índices Probabilísticos de Exactidão Global (IPEG) e os Índices Probabilísticos de Exactidão Específicos (IPEE), e nos dois índices de incerteza.

Desta fase resultou a eleição do classificador que obteve melhores resultados e, conseqüentemente, do melhor mapa de elementos de superfície. Paralelamente, foi efectuada uma análise comparativa entre os resultados obtidos com os índices de incerteza e os indicadores estatísticos para avaliar a compatibilidade de ambas as informações.

O segundo mapa, MUP, é executado para obter os padrões espaciais das unidades de paisagem. A informação da incerteza é usada no processo de transformação do MES num MUP, com base num conjunto de regras de decisão que consideram: os elementos de superfície classificados existentes dentro de cada objecto e o grau de incerteza associado à atribuição dos *pixels* às várias classes de elementos de superfície. Com a integração da informação sobre a incerteza no

processo de decisão, pretende-se evitar que os elementos de superfície classificados com maior grau de incerteza sejam utilizados na fase da classificação das unidades de paisagem.

Para avaliar se a introdução da incerteza associada à classificação dos elementos de superfície pode melhorar os resultados da classificação das unidades de paisagem, foi aplicado um método semelhante mas sem considerar a incerteza da classificação. O primeiro método (Figuras 3.3 e 3.4) integra a incerteza no processo de classificação e é composto pelas seguintes etapas: (1) classificação dos elementos de superfície ao nível do *pixel* usando classificadores não rígidos; (2) determinação da incerteza temática da classificação; (3) avaliação do comportamento dos classificadores e da classificação dos elementos de superfície com base em índices de incerteza e nos índices de exactidão estatísticos; (4) segmentação da imagem para obtenção dos objectos imagem; (5) classificação dos objectos com base em regras de decisão; e (6) avaliação da exactidão do MUP final com base em índices estatísticos. O segundo método de classificação, o qual não integra a incerteza no processo de classificação, é composto pelas seguintes etapas: (1) classificação não rígida ao nível do *pixel*; (2) avaliação do classificador e da classificação com base em índices estatísticos; (3) segmentação da imagem; (4) classificação dos objectos com base em regras de decisão; e (5) avaliação da exactidão do MUP com base em índices estatísticos.

Como o segundo método de classificação combinada é em tudo análogo ao primeiro método, excepto na introdução da incerteza, no trabalho aqui apresentado não se procede à exposição da segunda metodologia.

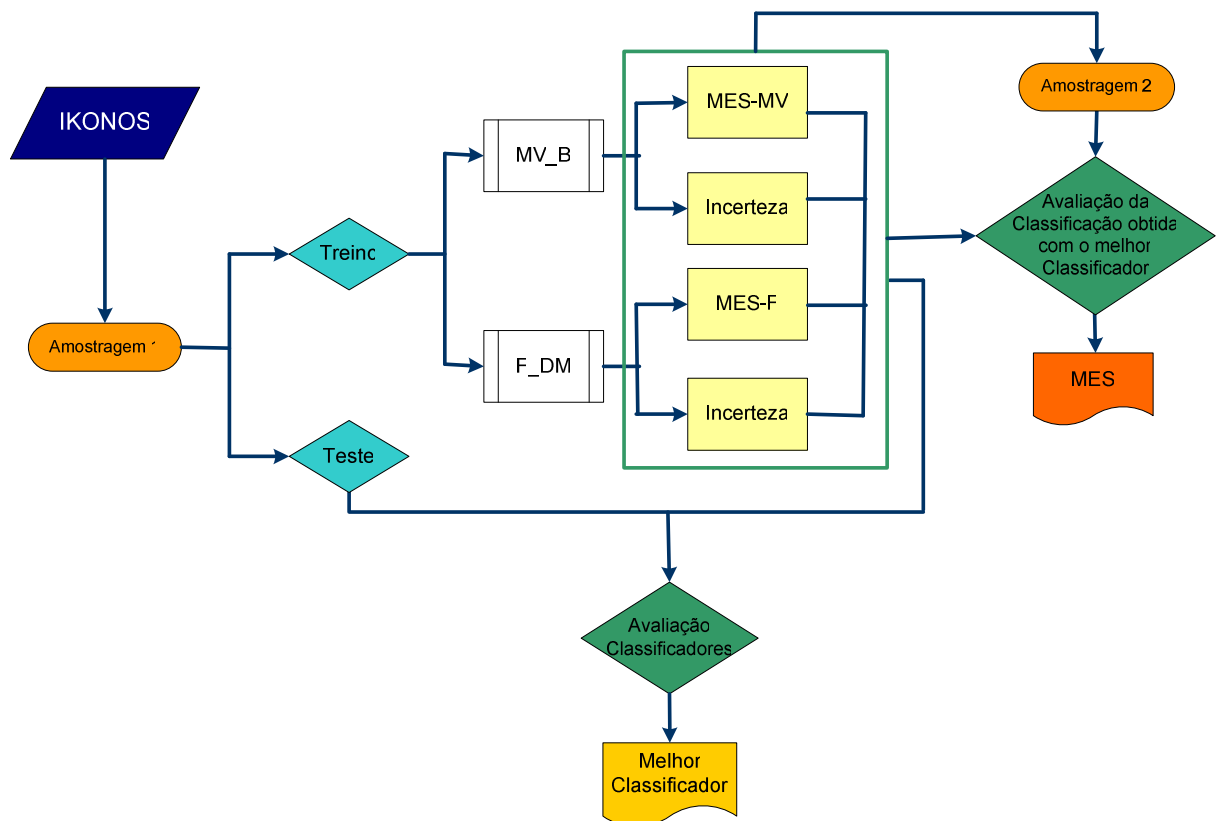


Figura 3.3 – Fluxograma da metodologia utilizada para o estudo da eleição dos classificadores e produção do MES.

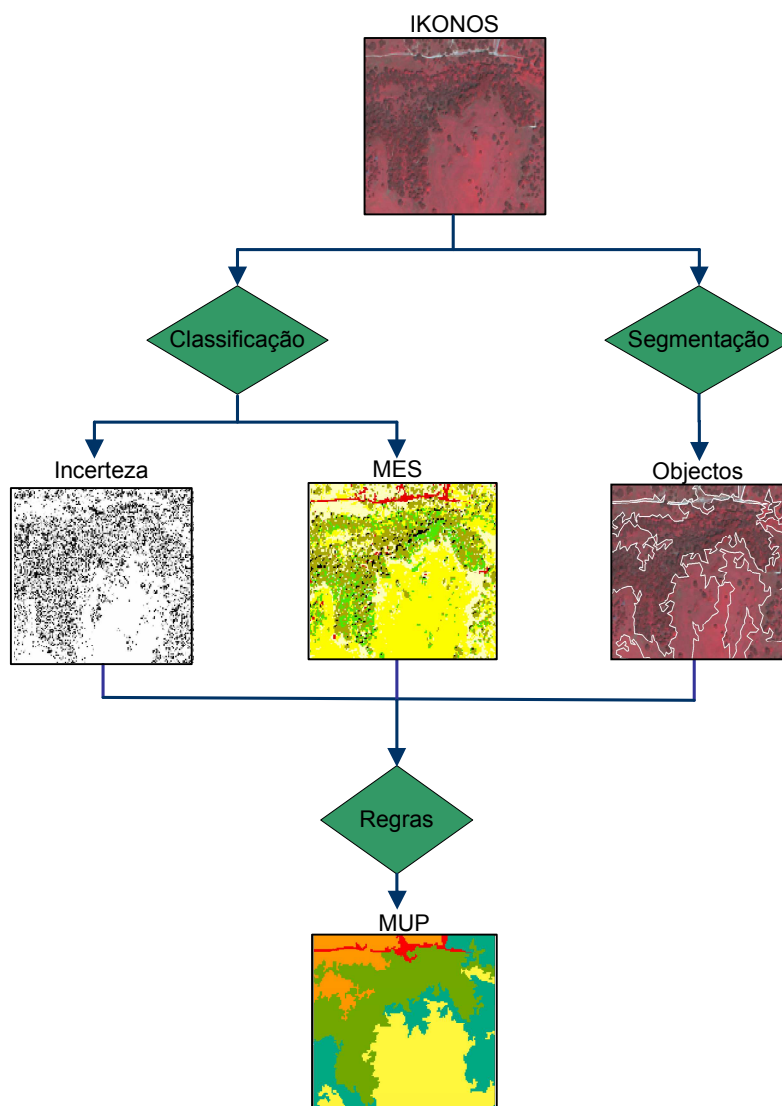


Figura 3.4 – Fluxograma da metodologia utilizada para a produção do MUP.

3.4.1 Protocolos de amostragem

No estudo foram definidos três protocolos de amostragem, um para treino e teste dos classificadores, aplicado na classificação ao nível do *pixel*, outro para avaliação dos MES e um terceiro para avaliação do MUP. A definição do primeiro protocolo de amostragem baseou-se na recolha das amostras representativas dos vários elementos de superfície, tendo sido utilizado o *pixel* como unidade amostra. Um foto-intérprete desenhou, em média, 25 polígonos representativos de cada classe de interesse, na imagem IKONOS, e sobre estes foi lançada uma amostragem aleatória, estratificada. Cada classe ficou representada por aproximadamente 250

amostras (*pixels*) (Plantier e Caetano, 2007). O total de amostras de cada classe foi posteriormente dividido aleatoriamente em dois conjuntos de dimensão idêntica: um foi utilizado para treino e outro para teste do classificador assistido. Desta amostra resultaram, em média, 120 áreas de treino para cada classe de elementos de superfície.

Para a avaliação da exactidão dos MES, foi lançada uma amostra aleatória estratificada por classe para que qualquer *pixel* da imagem pudesse ser seleccionado para referência. Desta forma, não se utilizaram amostras idênticas às de treino e teste do classificador, o que permite obter uma avaliação da exactidão do mapa mais realista. Segundo Stehman e Czaplewski (1998), para que se obtenha uma avaliação estatística rigorosa é fundamental que seja utilizado um protocolo de amostragem probabilístico, o que requer que todas as unidades escolhidas na amostra tenham associadas uma probabilidade de inclusão conhecida e maior do que zero. Com este segundo protocolo de amostragem, cada classe ficou representada por 100 amostras, tendo o *pixel* como unidade amostra. A dimensão da amostra de 100 *pixels* por classe é a recomendada para garantir um erro padrão na ordem dos 5%, assumindo um valor de 50% para a Exactidão Global (Stehman, 2001; Wickham *et al.*, 2004).

Para a avaliação da exactidão da classificação, a unidade amostra pode ser um pixel, um polígono ou uma área fixa, não existindo consenso sobre qual o melhor tipo de unidade. Segundo Stehman e Czaplewski (1998), a escolha da unidade amostra deve ser efectuada em função das características da paisagem, do processo de classificação, ou dos objectivos do projecto, entre outros factores.

Os protocolos para treino e teste do classificador e para avaliação da classificação do MES tiveram como objectivo avaliar a classificação associada a mapas classificados ao nível do *pixel*. A escolha do *pixel* como unidade amostra foi efectuada em função das características do mapa produzido. Com a abordagem desenvolvida para a obtenção do MUP, produziu-se um mapa em formato vectorial com uma UMC de 0.5 ha. Neste caso, a escolha da unidade amostra para a avaliação da exactidão do MUP, em formato vectorial, foi também efectuada em função das características do mapa produzido, razão pela qual se optou por escolher para unidades amostra regiões com uma área fixa de 0.5 ha, com um formato quadrangular. O método de amostragem adoptado foi igualmente o método aleatório estratificado, tendo sido lançadas 40 unidades amostrais por classe do MUP. Com a dimensão de 40 amostras por classe, e assumindo um valor de 50% para a Exactidão Global como o pior cenário, foi possível garantir um erro padrão da ordem dos 8% para os valores da exactidão. Quanto à construção da referência, utilizou-se a

técnica muito utilizada (e.g. Wilkinson, 2005; Plantier e Caetano, 2007), de atribuir duas classificações às amostras sempre que ocorresse dúvida sobre a verdadeira ocupação no terreno. Nestes casos, consideraram-se as amostras como correctamente classificadas, caso o mapa estivesse de acordo com uma das duas classificações de referência.

3.4.2 Nomenclatura

Como no estudo foram utilizadas imagens em que o *pixel* possui uma grande resolução espacial e não traduz unidades de paisagem (e.g. Área Agro-florestal, Floresta), mas antes componentes das unidades de paisagem (e.g. copa de árvore, sombra), adoptou-se uma nomenclatura baseada em elementos de superfície por ser a que melhor se adequa à informação contida ao nível do *pixel* numa IMGRE. A nomenclatura de elementos de superfície adoptada neste estudo, assim como a nomenclatura de unidades de paisagem, são idênticas às elaboradas por Plantier e Caetano (2007). Para a produção do mapa de elementos de superfície foram consideradas as seguintes classes: Copa de Eucalipto (CE); Copa de Sobreiro (CS); Copa de Pinheiro Manso (CPM); Sombra (S); Água Profunda (AP); Água Pouco Profunda (APP); Herbáceas (H), Herbáceas Esparsas (HE) e Área Não Vegetada (ANV) (Tabela 3.2). A opção de considerar mais do que uma classe de água deve-se ao facto de existir uma grande dispersão no espectro provocada por existirem na área de estudo corpos de água com diferentes profundidades (Figura 3.5).

Tabela 3.2 – Nomenclatura dos elementos de superfície.

| Classe | Símbolo |
|---------------------------|---------|
| Água Profunda/escuro | AP |
| Água Pouco Profunda/clara | APP |
| Área não Vegetada | ANV |
| Sombra | S |
| Herbáceas | H |
| Herbáceas Esparsas | HE |
| Copa Eucalipto | CE |
| Copa Pinheiro Manso | CPM |
| Copa Sobreiro/Azinheira | CS |

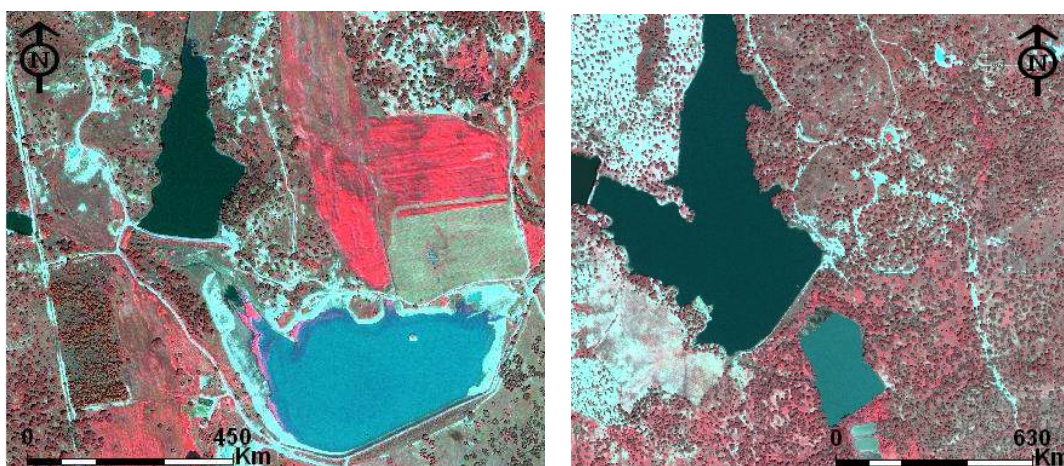


Figura 3.5 – Extractos da imagem IKONOS (RGB 432) que ilustram corpos de água com diferentes reflectâncias.

Para a produção do MUP foram consideradas oito unidades de paisagem cuja descrição se apresenta na Tabela 3.3. As unidades de paisagem consideradas são compostas por diferentes tipos de elementos de superfície, assumindo a proporção destes igualmente um aspecto relevante na sua caracterização.

Tabela 3.3 – Nomenclatura das unidades de paisagem.

| Classe | Símbolo | Descrição |
|----------------------------|---------|--|
| Urbano/Solo Nu | USN | Áreas artificiais, solo nu |
| Agro-Pastoricia | AP | Áreas com ocupação agrícola, pastagens, zonas com vegetação herbácea |
| Corpos de Água | CA | Superfícies de água, naturais ou artificiais |
| Floresta de Eucalipto | FE | Vegetação composta principalmente por árvores em que pelo menos 75% das copas pertencem à espécie eucalipto. |
| Floresta de Pinheiro Manso | FPM | Vegetação composta principalmente por árvores em que pelo menos 75% das copas pertencem à espécie Pinheiro Manso |
| Floresta de Sobreiro | FS | Vegetação composta principalmente por árvores em que pelo menos 75% das copas pertencem à espécie Sobreiro. |
| Agro-Florestal | AF | Culturas anuais ou pastagens e terras incultas abrangendo pelo menos 50% da superfície, com ocupação de copas de árvores das quais mais de 80% sejam de sobreiro/azinheira |
| Floresta Mista | FM | Vegetação composta por árvores de diferentes espécies florestais e em que nenhuma ocupa uma área superior a 75%. |

3.4.3 Mapa de elementos de superfície

3.4.3.1 Classificação

Para identificar os elementos de superfície procedeu-se a uma classificação assistida e utilizaram-se dois classificadores não rígidos: classificador *fuzzy* baseado no princípio do classificador da distância mínima à média (F_DM) e o classificador probabilístico da máxima verosimilhança, tendo-se obtido os graus de probabilidade *a posteriori* utilizando o teorema de Bayes (MV_B). Numa classificação assistida, o treino do classificador é feito de modo assistido, *i.e.*, admite-se conhecida a classe correspondente a cada padrão do conjunto de treino gerado. O classificador é treinado a replicar a decisão correcta para todos os padrões de treino. A razão pela qual foram escolhidos estes classificadores deve-se ao facto de estarem disponíveis na maioria dos *softwares* comerciais e permitirem que, no resultado final da classificação, fique associado, a cada *pixel*, graus de probabilidade e possibilidade a uma ou várias classes, o que possibilita a determinação da incerteza da classificação e a aplicação de diferentes medidas de incerteza.

O classificador da máxima verosimilhança é um dos classificadores mais utilizados e que tem demonstrado bons resultados em estudos aplicados ao coberto florestal com imagens IKONOS (Goetz *et al.*, 2003; Wang *et al.*, 2004). Este classificador tem sido usado essencialmente na sua versão tradicional em que, no resultado final da classificação, a cada *pixel* fica atribuído apenas a classe para a qual o valor de probabilidade obtido pela função densidade (Equação 3.1) foi mais elevado. Contudo, podem ser obtidos todos os valores de probabilidade *a posteriori*, ficando associado a cada *pixel* uma distribuição de probabilidade às várias classes (e.g. Wang, 1990; Foody, 1996; Foody, 2004). Por exemplo, a um *pixel* pode ficar associado um grau de probabilidade de 0.68 à classe de Copas de Eucalipto, 0.29 a Herbáceas e 0.03 a Copas de Pinheiro Manso, sendo o somatório dos graus de cada classe potencial igual a 1 (Lillesand *et al.*, 2004). No resultado final obtém-se uma imagem por classe e o valor em cada *pixel* expressa a probabilidade *a posteriori* do *pixel* pertencer a essa classe.

A função densidade é dada por (Foody *et al.*, 1992):

$$p(x_k | i) = \frac{1}{(2\pi)^{n/2} |V_i|^{1/2}} \exp\left(-1/2[(X_k - u_i)^T V_i^{-1} (X_k - u_i)]\right) \quad (3.1)$$

Onde $p(x_k | i)$ é uma função densidade de probabilidade de um *pixel* x_k pertencer à classe i ,

n é o número de bandas, X_k é o vector obtido a partir da resposta espectral do *pixel* x_k em todas as bandas, V_i é a matriz variância-covariância para a classe i e u_i é o vector média para a classe i considerando todos os *pixels*. O termo $(X_k - u_i)^T V_i^{-1} (X_k - u_i)$ é a distância Mahalanobis entre o *pixel* e o centróide da classe i .

As probabilidades *a posteriori*, de acordo com o teorema de Bayes, são obtidas por:

$$p(i | x_k) = \frac{P(i) p(x_k | i)}{\sum_{i=1}^t P(i) p(x_k | i)} \quad (3.2)$$

em que $p(x_k | i)$ é a função densidade de probabilidade e $p(i | x_k)$ representa a probabilidade *a posteriori* de um *pixel* pertencer à classe i (probabilidade de a hipótese ser verdadeira dada a evidência, *i.e.*, uma vez conhecidos os dados experimentais), $P(i)$ é a probabilidade *a priori* da classe i (probabilidade da hipótese ser verdadeira independentemente da evidência, ou seja, antes da análise de um conjunto de dados) e t representa o número total de classes. Para cada *pixel*, a soma das probabilidades *a posteriori* é 1 (Foody *et al.*, 1992).

Os graus de probabilidade *a posteriori* podem ter uma interpretação diferente em função do tipo de classificação efectuada, conforme referido na secção 3.1. Estes graus podem ser interpretados como graus parciais de pertença às classes se se considerar que são o resultado de uma classificação não rígida. Esta interpretação pressupõe que o analista considera que o *pixel* contém mais do que um tipo de ocupação do solo e as probabilidades são usadas como indicador da proporção relativa de cada uma. Esta interpretação é muito utilizada na classificação de *pixels* mistos, sendo designada por *sub-pixel classification* (e.g. Foody, 1996; Foody, 2000a). Contudo, se a classificação for considerada rígida, os graus de probabilidade *a posteriori* podem ser interpretados como a probabilidade de atribuir correctamente uma classe a um *pixel*, representando uma incerteza probabilística. No estudo, a obtenção dos graus de probabilidade *a posteriori* foi efectuada com o objectivo de se obter informação sobre a dificuldade em atribuir correctamente uma dada classe a um *pixel*. Pretendeu-se, assim, obter uma classificação rígida, à qual fica associada a informação sobre a incerteza da classificação (Eastman e Laney, 2002).

O segundo método de classificação usa um classificador *fuzzy* assistido, baseado no classificador da distância mínima à média, disponível no *software* IDRISI (Eastman e Laney, 2002). Com este

classificador, a imagem é classificada com base na informação contida nas assinaturas espectrais das áreas de treino e num parâmetro designado por *Z-score distance* introduzido pelo utilizador. Depois de definido um conjunto representativo de amostras para cada classe e os espaços característicos, o conjunto de pertença *fuzzy* é calculado com base na distância Euclidiana entre a assinatura espectral de cada *pixel*, em cada banda, e a média das assinaturas espectrais das áreas de treino para cada classe (Figura 3.6 a)), usando uma função de pertença sigmoide (Figura 3.6 b)). Este método baseia-se no pressuposto de que a média das assinaturas espectrais das áreas de treino para cada classe, representa o valor ideal da classe. A atribuição dos *pixels* a todas as classes é efectuada assumindo valores no intervalo $[0,1]$, em que o valor zero significa nenhuma atribuição e o valor 1 total atribuição.

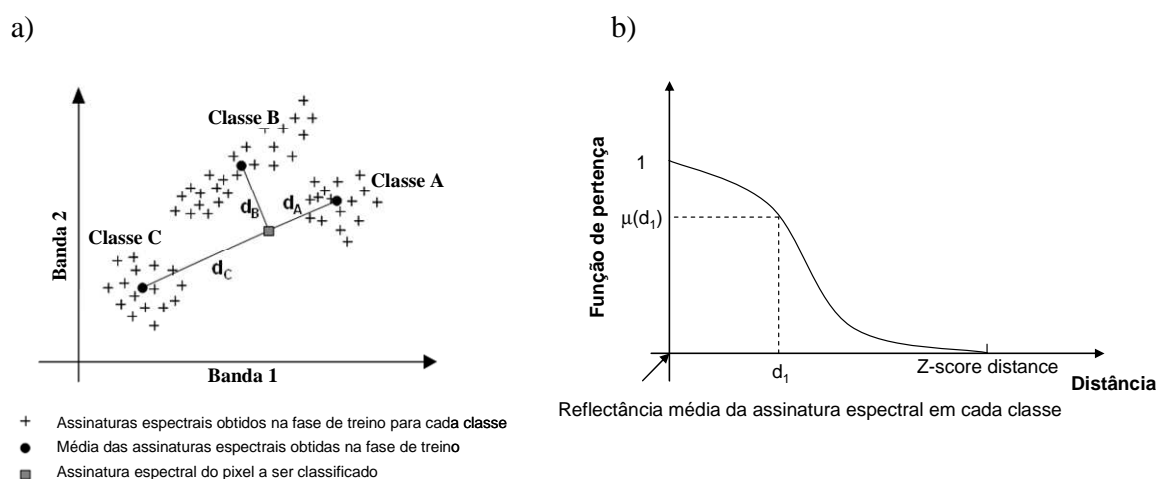


Figura 3.6 – a) Exemplo ilustrativo da classificação assistida com o classificador fuzzy da distância mínima à média; b) função de pertença sigmoide usada no classificador fuzzy baseado no classificador da distância mínima à média.

Quanto mais próxima a assinatura espectral de um *pixel* estiver da média das assinaturas espectrais das amostras representativas de uma classe, maior será o grau de pertença a essa classe. À medida que a distância aumenta, o grau de pertença diminui. Quando a distância atinge o valor atribuído pelo utilizador ao parâmetro de distância *Z-score*, o grau de pertença decresce para zero (Figura 3.6 b)).

Neste estudo, o valor adoptado para o parâmetro *Z-score* foi o valor médio de separabilidade das classes. Para avaliar a separabilidade espectral entre as classes definidas em cada uma das imagens, foi usada a informação das áreas de treino e recorreu-se à medida de separabilidade Distância de Battacharrya (B-Distance).

No resultado final da classificação obtém-se uma imagem por classe e cada imagem expressa a possibilidade de pertença dos *pixels* que a constituem à classe (o total dos graus de todas as classes potenciais pode ser superior a 1).

3.4.3.2 Quantificação da incerteza

Para a quantificar a incerteza da classificação foram utilizadas as medidas de não-especificidade (NSp, Un), a entropia relativa (Er) e o *ratio* de incerteza (RI), disponível no *software* comercial IDRISI. Mais informação sobre estas medidas de incerteza encontra-se no Capítulo 2. Apesar dos graus de probabilidade e de possibilidade, associados aos *pixels* no resultado do processo de classificação temática, disponibilizarem informação acerca da incerteza da classificação dos elementos de superfície, a aplicação de medidas de incerteza a cada distribuição de probabilidade ou possibilidade permite obter informação adicional. Este tipo de indicadores permite captar a compatibilidade e/ou dispersão da atribuição das classes a cada UEA e obter um valor que fica associado a cada UEA e que reflecte a incerteza dessa atribuição.

A medida de entropia relativa introduzida por Maselli *et al.* (1994) e o *ratio* de incerteza disponível no *software* IDRISI foram utilizadas para quantificar a incerteza da classificação ao nível do *pixel* obtida com o classificador probabilístico. A medida disponível no *software* comercial é dada pela expressão (3.3), onde p_i ($i=1, \dots, n$) representa os graus de atribuição das classes aos *pixels*, e n é o número de classes (Eastman, 2006),

$$RI = 1 - \frac{\max_{i=1, \dots, n}(p_i) - \frac{\sum_{i=1}^n p_i}{n}}{1 - \frac{1}{n}} \quad (3.3)$$

Esta medida pode ser aplicada a graus de possibilidade e graus de probabilidade. No caso de ser

aplicada a graus de probabilidade, $\sum_{i=1}^n p_i = 1$.

A medida de entropia relativa, utilizada no estudo, é semelhante à introduzida por Maselli *et al.* (1994), e é dada por:

$$Er = - \frac{\sum_{i=1}^n p_i \log_2 p_i}{\log_2 n} \quad (3.4)$$

Ambas as medidas, RI e Er, assumem valores no intervalo [0,1], significando o valor zero que não existe ambiguidade na atribuição da unidade especial a uma classe e que, portanto, não existe incerteza. O valor 1 significa que a incerteza é máxima.

Para quantificar a incerteza da classificação obtida com o classificador *fuzzy*, utilizaram-se as medidas de não-especificidade (NSp), desenvolvida por Yager (1982), dada por:

$$NSp(\pi) = 1 - \sum_{i=1}^n [\pi(x_i) - \pi(x_{i+1})] \frac{1}{i} \quad (3.5)$$

e a versão normalizada, proposta no capítulo 2, da medida incerteza-U, desenvolvida por Higashi e Klir (1982), dada por:

$$Un(\pi) = \frac{[1 - \pi(x_1)] \log_2 n + \sum_{i=2}^n [\pi(x_i) - \pi(x_{i+1})] \log_2 i}{\log_2 n} \quad (3.6)$$

Em qualquer uma das equações, (3.5) e (3.6), π representa uma distribuição de possibilidade definida num conjunto Universal X , a qual verifica a condição $\pi(x_1) \geq \pi(x_2) \geq \dots \geq \pi(x_n)$, onde n é o número de elementos no conjunto universal e $\pi(x_{n+1})$ assume o valor zero.

A versão normalizada da medida incerteza-U foi utilizada para as duas medidas terem o mesmo intervalo de variação [0,1] e para facilitar a comparação dos resultados obtidos. Conforme referido no Capítulo 2, as medidas de não-especificidade são adequadas para determinar a incerteza resultante das classificações *fuzzy* possibilísticas, permitindo quantificar a ambiguidade em especificar uma solução única (Pal e Bezdek, 2000). Neste estudo, o objectivo é quantificar a incerteza que está associada à atribuição de um elemento de superfície a um *pixel*.

3.4.3.3 Avaliação dos classificadores

3.4.3.3.1 Avaliação dos classificadores com base em índices de exactidão estatísticos

O procedimento para a avaliação dos classificadores consistiu em cruzar os dois mapas

produzidos com as áreas de teste. As áreas de teste resultaram da divisão das amostras do primeiro protocolo de amostragem, em que apenas foram considerados *pixels* da imagem abrangidos pelos polígonos das classes de interesse. Consequentemente, no protocolo de amostragem utilizado para teste do classificador, ficaram excluídos da amostragem todos os *pixels* da imagem não abrangidos pelos polígonos de interesse. Por este motivo, este é apenas adequado para avaliar os classificadores e não os mapas (Wilkinson, 2005).

O processo de validação dos mapas produzidos foi composto pelos seguintes passos: (1) recolha de uma amostra da população (áreas de teste); (2) recolha da informação de referência relativa às áreas amostradas; (3) cálculo da matriz de erro e respectivos índices de exactidão.

Os procedimentos de análise dos resultados foram baseados em matrizes de erro, conforme se apresenta na Tabela 3.4, a partir das quais foram deduzidos os Índices de Exactidão Específicos (IEE), o Índice de Exactidão do Utilizador (EU) e o Índice de Exactidão do Produtor (EP) e os Índices de Exactidão Global (IEG) (Exactidão Global e coeficiente de KHAT).

Tabela 3.4 – Matriz de erro.

| | | Classes Referência | | | | Total |
|----------------------------|----------------|--------------------|----------------|-----|----------------|----------|
| | | H ₁ | H ₂ | ... | H _k | |
| Classes Mapa (estratos) | H ₁ | n_{11} | n_{12} | ... | n_{1k} | n_{1+} |
| | H ₂ | n_{21} | n_{22} | ... | n_{2k} | n_{2+} |
| | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| | H _k | n_{k1} | n_{k2} | ... | n_{kk} | n_{k+} |
| Total | | n_{+1} | n_{+2} | ... | n_{+k} | n |

A matriz de erro é uma tabela de contingência, em que n_{ij} é o número de unidades amostra cuja categoria ou classe no mapa é i ($i=1,2,\dots,k$) e cuja classe de referência é j ($j=1,2,\dots,k$). Os elementos da diagonal representam o número de indivíduos correctamente atribuídos à sua classe e os elementos fora da diagonal representam classificações incorrectas. As classificações incorrectas podem ser divididas em dois tipos: erros de omissão e erros de comissão. Os erros de omissão correspondem a indivíduos que, pertencendo a uma determinada classe, não foram incluídos nela. Os erros de comissão correspondem a indivíduos que foram classificados numa determinada classe mas que na realidade pertencem a outra.

As somas marginais da amostra no mapa e na referência são dadas, respectivamente, por:

$$n_{i+} = \sum_{j=1}^k n_{ij} \quad (3.7)$$

$$n_{+j} = \sum_{i=1}^k n_{ij} \quad (3.8)$$

A Exactidão Global dos mapas foi obtida por:

$$EG = \frac{\sum_{i=1}^k n_{ii}}{n} \quad (3.9)$$

em que n_{ii} é o número de *pixels* da amostra da classe i no mapa que correspondem à mesma classe na referência e n é o número total de *pixels* da amostra.

A Exactidão do Utilizador (EU) para cada classe é obtida através da expressão:

$$EU = \frac{n_{ii}}{n_{i+}} \quad (3.10)$$

em que n_{ii} é o número de *pixels* da amostra da classe i no mapa que correspondem à mesma classe na referência e n_{i+} é o número de *pixels* da amostra da classe i no mapa.

A Exactidão do Produtor (EP) para cada classe é obtida por:

$$EP = \frac{n_{jj}}{n_{+j}} \quad (3.11)$$

em que n_{jj} é o número de *pixels* da amostra da classe j na referência que correspondem à mesma classe no mapa e n_{+j} representa o número de *pixels* da amostra pertencentes à classe de referência j .

3.4.3.3.2 Avaliação dos classificadores com base em índices de incerteza

Para avaliar se a informação obtida pelas medidas de incerteza pode ser usada como indicador do comportamento do classificador, foram considerados dois tipos de índices de incerteza. Um índice é baseado na incerteza média por classe. O outro índice de incerteza é baseado nos baixos valores de incerteza, *i.e.*, corresponde à percentagem de *pixels* por classe com incerteza inferior a 0.5 ($I_{NSp < 0.5}$, $I_{Un < 0.5}$, $I_{RI < 0.5}$, $I_{Er < 0.5}$).

Como a incerteza toma valores no intervalo $[0,1]$, em que o valor 1 significa máxima incerteza e o valor 0 nenhuma incerteza e, como os índices estatísticos variam entre 0 e 100%, em que 100% significa muita exactidão e 0% baixa exactidão, optou-se por considerar para o índice de incerteza baseado na média, o complementar da incerteza média por classe (I_{1-NSp} , I_{1-Un} , I_{1-Rl} , I_{1-Er}), para permitir uma comparação com os índices de exactidão. Desta forma, o valor 1 passa a significar total certeza e o valor 0 total incerteza. O índice de incerteza baseado nos baixos valores de incerteza possui uma abordagem semelhante à dos índices de exactidão estatísticos, em que os valores são obtidos através da contagem do número de amostras correspondentes às classes no mapa e no terreno. Neste estudo o coeficiente de correlação entre os índices de incerteza e os IEE foi também calculado.

Apesar das medidas de incerteza poderem ser aplicadas a toda a imagem, para avaliar a relação e a compatibilidade entre a informação fornecida pelas medidas de incerteza e os IEE, utilizaram-se apenas os valores obtidos para os *pixels* amostra para que os resultados não fossem influenciados pela representatividade da amostra. Esta metodologia foi usada porque a matriz de erro é gerada com base apenas na informação obtida a partir de um conjunto de amostras.

Outro aspecto igualmente considerado foi a análise dos graus de probabilidade e possibilidade associados aos *pixels*. Esta informação permite avaliar o nível de incerteza da atribuição das classes aos *pixels*. Por exemplo, a comparação entre os graus de probabilidade ou possibilidade consecutivos permite avaliar a capacidade do classificador em separar os objectos, o que reflecte a ambiguidade em especificar uma solução única. Se um objecto for atribuído, por exemplo, à classe Áreas Residenciais Contínuas com um grau possibilístico de 0.9, para o melhor valor da classificação, e atribuído à classe Áreas Residenciais Descontínuas com um grau de 0.8, para o segundo melhor valor da classificação, significa que existe ambiguidade ou incerteza, atendendo a que os valores são muito próximos.

3.4.3.4 Avaliação da exactidão temática do mapa

3.4.3.4.1 Avaliação da classificação com base em índices de exactidão estatísticos

Para a avaliação do mapa, utilizou-se um segundo protocolo de amostragem e um método de amostragem em que todas as classes têm associada uma probabilidade de inclusão conhecida e

diferente de zero (Stehman e Czaplewski, 1998). Para o efeito, foi definida uma amostragem aleatória estratificada, tendo o *pixel* como unidade amostra e em que qualquer *pixel* da imagem pudesse ser seleccionado para ser incluído no protocolo de amostragem. Esta avaliação foi apenas efectuada para o mapa produzido com o classificador que apresentou melhores resultados no estudo do comportamento dos classificadores.

A avaliação da exactidão do MES foi efectuada a partir da matriz de erro (ver Tabela 3.4), em que n_{ij} é o número de unidades amostra cuja categoria ou classe no mapa é i e a classe de referência é j . Segundo Card (1982), quando se utiliza uma amostra aleatória estratificada, a Exactidão Global (EG) de um mapa não deve ser apenas estimada através do somatório das entradas na diagonal da matriz de erro (número de *pixels* que foram bem classificados) a dividir pelo número total de amostras, devido ao enviesamento introduzido no caso de existir uma diferente contribuição da amostra nas diferentes classes. Por este motivo, os IEG e os IEE estimados, por classe, usados para avaliar o mapa, devem incluir as áreas conhecidas de cada classe do mapa de forma a melhorar a estimativa da proporção dos indivíduos correctamente classificados no mapa (Carrão *et al.*, 2007). O facto de se considerarem proporções, permite obter resultados menos enviesados dado que se estão a extrapolar, para toda a população (área de estudo), os índices de exactidão em função da proporção que cada classe nela ocupa (Stehman e Czaplewski, 1998; Zhu *et al.*, 2000; Carrão *et al.*, 2007). Quando se aplicam proporções, a Matriz de Confusão (MC) é designada por Matriz de Confusão Probabilística (MCP) e a Exactidão Global (EG) é designada por Probabilidade Global (PG). O índice de Exactidão do Utilizador (EU) corresponde à probabilidade condicionada de um dado *pixel*, classificado no mapa como i , pertencer realmente à classe i na referência. Este índice é referido neste estudo como Probabilidade Condicionada pelo Mapa (PCM). A Exactidão do Produtor (EP) corresponde à probabilidade condicionada de um dado *pixel*, classificado como j na referência, ser classificado como classe j no mapa, sendo referida neste estudo como Probabilidade Condicionada pela Referência (PCR).

Na definição da amostragem aleatória estratificada, cada estrato corresponde às categorias ou classes do mapa e o número de amostras em cada estrato n_{i+} ($i=1,2,\dots,k$) é definido previamente antes da recolha das amostras e de forma independente de estrato para estrato. Deste modo, na matriz de erro, o total por linhas é fixo e previamente conhecido n_{i+} , mas o total por colunas n_{+j} ($j=1,2,\dots,k$) é variável. Considerando que os mapas temáticos produzidos

são constituídos por N *pixels*, o número total de *pixels* de cada classe no mapa N_{i+} é conhecido, mas o total de *pixels* para cada classe da referência N_{+j} é desconhecido.

A PG estimada para o mapa, *i.e.*, a proporção de *pixels* correctamente classificados no mapa, foi calculada por:

$$PG = \frac{1}{N} \sum_{i=1}^k \frac{N_{i+}}{n_{i+}} n_{ii} \quad (3.12)$$

em que N é o número de *pixels* do mapa; N_{i+} é o número de *pixels* da classe i no mapa; n_{i+} é o número de *pixels* da amostra da classe i no mapa; e n_{ii} é o número de *pixels* da amostra da classe i no mapa que correspondem à mesma classe na referência.

A probabilidade condicionada de um dado *pixel*, classificado no mapa como i , pertencer realmente à classe i na referência (PCM) é dado por:

$$PCM = \frac{n_{ii}}{n_{i+}} \quad (3.13)$$

A estimativa da exactidão do produtor de uma amostra aleatória estratificada deve ser calculada usando estimadores não enviesados para obter as quantidades desconhecidas N_{+j} (número de *pixels* da classe j na referência) e N_{ii} (número de *pixels* da classe i no mapa que correspondem à mesma classe na referência). Estas quantidades foram obtidas por:

$$\hat{N}_{ii} = \frac{N_{i+}}{n_{i+}} n_{ii} \quad (3.14)$$

$$\hat{N}_{+j} = \sum_{i=1}^k \frac{N_{i+}}{n_{i+}} n_{ij} \quad (3.15)$$

onde n_{ij} representa o número de unidades amostra cuja categoria ou classe no mapa é i ($i = 1, 2, \dots, k$) e cuja classe de referência é j ($j = 1, 2, \dots, k$).

A probabilidade condicionada de um dado *pixel*, classificado como j na referência, ser classificado como classe j no mapa (PCR) é, assim, dada pelo *ratio* entre estes dois estimadores:

$$PCR = \frac{\hat{N}_{ii}}{\hat{N}_{+j}} \quad (3.16)$$

De notar que, se a avaliação é baseada considerando o *pixel* como unidade amostra, é indiferente considerar no cálculo das proporções a área dos *pixels*, uma vez que esta é igual para todos, ou o número de *pixels* amostra (Stehman e Czaplewski, 1998).

3.4.3.4.2 Avaliação da classificação com base em índices de incerteza

Para avaliar se a informação obtida pelas medidas de incerteza pode ser usada como indicador da exactidão da classificação e se existe correlação com os resultados da exactidão da classificação obtidos com os índices PCM e PCR, foram utilizados os índices de incerteza expostos na secção 3.4.3.3.2., nomeadamente: o índice baseado na média, que corresponde ao complementar da incerteza média por classe, e o índice de incerteza baseado nos baixos valores de incerteza, que corresponde à percentagem de objectos por classe com incerteza inferior a 0.5.

3.4.3.5 Distribuição espacial da incerteza

Com a aplicação de classificadores não rígidos, pode ficar associada uma distribuição de probabilidade ou possibilidade, a cada *pixel* da imagem, conforme o classificador utilizado, cujos valores podem ser interpretados como o grau de compatibilidade das características do *pixel* com as características da classe ou das classes à qual foi atribuído, e um indicador sobre a incerteza da classificação. Esta informação pode ser representada geograficamente permitindo a visualização da distribuição espacial da incerteza e a sua relação espacial com as classes atribuídas a cada *pixel*, podendo ser útil do ponto de vista do utilizador. Quando no resultado final da classificação fica apenas atribuído ao *pixel* a classe correspondente ao melhor grau de probabilidade, de possibilidade ou de pertença, considerada a melhor classe, são cometidos erros que podem ser designados, também neste contexto, de omissão e de comissão. Os erros de comissão ocorrem porque, quando é atribuído ao *pixel* a melhor classe e o grau de compatibilidade é inferior a 1, está a ser considerada uma compatibilidade superior à que realmente existe (Zhu, 1997). Os erros de omissão ocorrem quando, no processo de classificação, existe compatibilidade entre as características do *pixel* com várias classes e apenas se considera a classe que apresenta melhor grau de compatibilidade, sendo omitidos todos os outros possíveis graus de compatibilidade desse *pixel* com as restantes classes. Para estimar os

erros de comissão, utilizou-se uma medida proposta por Zhu (1997), designada por *exaggeration uncertainty measure*, e que se traduziu por Medida de Exagero (E_x). Esta medida, quando aplicada aos resultados de uma classificação obtida com um classificador probabilístico, é dada por:

$$E_x = 1 - p(x_1) \quad (3.17)$$

em que $p(x_1)$ representa o maior grau de probabilidade.

Quando aplicada a distribuições de possibilidade, $p(x_1)$ é substituído na expressão (3.17) pelo maior grau de possibilidade $\pi(x_1)$.

Para estimar os erros de omissão da classificação, foram utilizadas as medidas de incerteza de entropia relativa (E_r) e de não-especificidade (NSp , Un) e a medida RI . A análise apresentada na secção 2.4, permitiu concluir que estes indicadores são sensíveis à variação da compatibilidade entre as características das UEAs e as várias classes, assim como à dispersão da atribuição das várias classes a uma UEA. A dispersão ocorre quando existe compatibilidade entre as características do *pixel* com várias classes. A informação obtida com estas medidas pode ser utilizada como um indicador da dificuldade do classificador em atribuir uma solução única e, conseqüentemente, como indicador da informação omitida no resultado final da classificação.

3.4.4 Mapa de unidades de paisagem

3.4.4.1 Abordagem de classificação combinada com integração da incerteza

O MUP foi construído através da combinação do MES que apresentou melhores resultados de exactidão, da incerteza da sua classificação e dos objectos obtidos utilizando um método de segmentação. Na fase de segmentação, toda a imagem é subdividida em zonas (ou regiões) homogéneas e uniformes que correspondem aos padrões espaciais da imagem. A extracção dos objectos foi efectuada utilizando o algoritmo de segmentação designado por "*Fractal Net Evolution Approach*" (FNEA), implementado no *software* eCognition, também denominado por "segmentação multi-resolução" (Batz e Schape, 2000). O método baseia-se na técnica de fusão de regiões começando cada *pixel* por formar um objecto ou região. A fusão dos objectos contíguos é decidida com base em critérios de homogeneidade local. Os objectos contíguos são

fundidos num único objecto se a heterogeneidade do objecto resultante da fusão, não ultrapassar um valor máximo de heterogeneidade espectral. O parâmetro que determina a heterogeneidade máxima é, por esta razão, designado por “parâmetro escala” (Batz e Schape, 2000) e é atribuído pelo utilizador. Consequentemente, o tamanho dos objectos imagem, resultantes da fusão, depende do valor atribuído a esse parâmetro. O processo pára quando não houver mais objectos para fundir. Este método permite a extracção de objectos da imagem, em qualquer resolução ou escala que se pretenda, podendo ser registados em níveis diferentes, permitindo construir uma rede hierárquica de objectos imagem que representam simultaneamente a informação da imagem com diferentes resoluções espaciais. Uma desvantagem deste método é que a decisão final sobre o valor adequado a atribuir ao parâmetro escala é efectuada com base numa análise visual do resultado da segmentação e não num critério quantitativo. Neste estudo foi utilizado apenas um nível de segmentação da imagem e os parâmetros utilizados encontram-se na Tabela 3.5. O critério utilizado na escolha do parâmetro escala foi obter uma imagem cujo tamanho médio dos objectos fosse 0.5 ha e que permitisse dividir a imagem em regiões com significado, correspondendo aos padrões espaciais das unidades de paisagem.

Tabela 3.5 – Parâmetros da segmentação.

| Peso das bandas | Escala | Cor | Forma | Suavização | Compacidade |
|-----------------|--------|-----|-------|------------|-------------|
| 1 | 500 | 0.9 | 0.1 | 0.1 | 0.9 |

O passo seguinte consistiu no desenvolvimento das regras para permitir transformar os elementos de superfície em unidades de paisagem. As regras foram desenvolvidas de forma a incorporarem: (1) o arranjo dos elementos de superfície, obtidos com a classificação ao nível do *pixel*, dentro de cada objecto; e (2) o grau de incerteza dessa atribuição. A construção das regras requer uma análise preliminar dos graus de probabilidade atribuídos aos elementos de superfície e a sua respectiva incerteza, de forma a escolher o limiar adequado.

A transformação do MES num MUP é semelhante a uma árvore de decisão com uma estrutura hierárquica constituída por vários níveis. A cada nível é aplicado um conjunto de regras, as quais constituem um conjunto de questões relativas aos atributos das observações. Com a aplicação das regras, os dados pertencentes a cada nível são divididos em subconjuntos com atributos idênticos, de forma a que, em cada subconjunto, todos os indivíduos pertençam à mesma classe (Unidade de Paisagem). A interrupção da divisão dos diferentes subconjuntos do conjunto de dados inicial é condicionada por diferentes factores como, por exemplo: (1) a informação

disponível não permitir a divisão em subconjuntos mais puros; (2) o facto dos indivíduos de determinado subconjunto pertencerem todos à mesma classe.

Na Figura 3.7 apresenta-se o esquema de classificação hierárquica desenvolvido e, na Tabela 3.6, as regras aplicadas em cada nível hierárquico. O objectivo da regra 1 é permitir a distinção entre 'Floresta' e 'Não Floresta'. O critério utilizado para diferenciar estas duas classes baseou-se na definição da CELPA (Associação da Indústria Papeleira Portuguesa), a qual considera como 'Floresta' regiões em que a percentagem de coberto florestal é superior a 10%. A regra 2 atribui os objectos considerados 'Não Floresta' a uma de três unidades de paisagem: Agro-Pastorícia, Corpos de Água, Urbano/Solo Nu. A regra 3 distingue as regiões florestais em 'Floresta Densa' e 'Floresta Não Densa'. A regra 4 atribui cada objecto classificado como 'Floresta Densa' a uma das quatro unidades de paisagem: Floresta de Eucalipto, Floresta de Pinheiro, Floresta de Sobreiro, Floresta Mista. A regra 5 atribui os objectos considerados 'Floresta Não Densa' a uma de duas classes: Áreas Agro-Florestais e Floresta Mista. Na definição das regras de classificação houve um esforço no sentido de produzir regras gerais, embora a escolha do limiar para distinguir os elementos de superfície a serem atribuídos às classes, em função da baixa incerteza, esteja muito dependente da qualidade da classificação ao nível do *pixel*. As linhas de código em AML são apresentadas no Anexo B.

Tabela 3.6 – Regras de classificação.

| Nº | Regras | Classe |
|----|---|-------------------------|
| 1 | Os objectos têm mais de 10% de ES classificados como copas de árvores com uma incerteza inferior a 0.25 | Floresta |
| | Os objectos não satisfazem o teste anterior | Não-Floresta |
| 2 | Se a moda dos ES classificados dentro de cada objecto com uma incerteza inferior a 0.25, for Água Profunda ou Água Pouco Profunda | Corpos de Água |
| | Se a moda dos ES classificados dentro de cada objecto com uma incerteza inferior a 0.25 for Herbáceas ou Herbáceas Esparsas | Agro-Pastoricia |
| | Se a moda dos ES classificados dentro de cada objecto com uma incerteza inferior a 0.25, for Área não Vegetada ou Sombra | Urbano/Solo Nu |
| 3 | Os objectos têm mais de 75% de ES classificados como copas de árvores com uma incerteza inferior a 0.25, independentemente da espécie | Floresta Densa |
| | Os objectos não satisfazem o teste anterior | Floresta Não Densa |
| 4 | Se os objectos possuem mais de 75% de ES classificados como Copa de Eucalipto | Floresta de Eucalipto |
| | Se os objectos possuem mais de 75% de ES classificados como Copa de Pinheiro Manso | Floresta Pinheiro Manso |
| | Se os objectos possuem mais de 75% de ES classificados como Copa de Sobreiro | Floresta de Sobreiro |

| | | |
|---|---|----------------|
| | Os objectos não satisfazem o teste anterior e nem Copa de Eucalipto nem Copa de Pinheiro Manso são espécies dominantes | Floresta Mista |
| 5 | A percentagem de ES classificados como copas de árvores é inferior a 50%; a percentagem de herbáceas e herbáceas esparsas é superior à de copas de sobreiro e 80% das copas de árvores são copas de sobreiros com uma incerteza inferior a 0.25 | Agro-Florestal |
| | Os objectos não satisfazem o teste anterior | Floresta Mista |

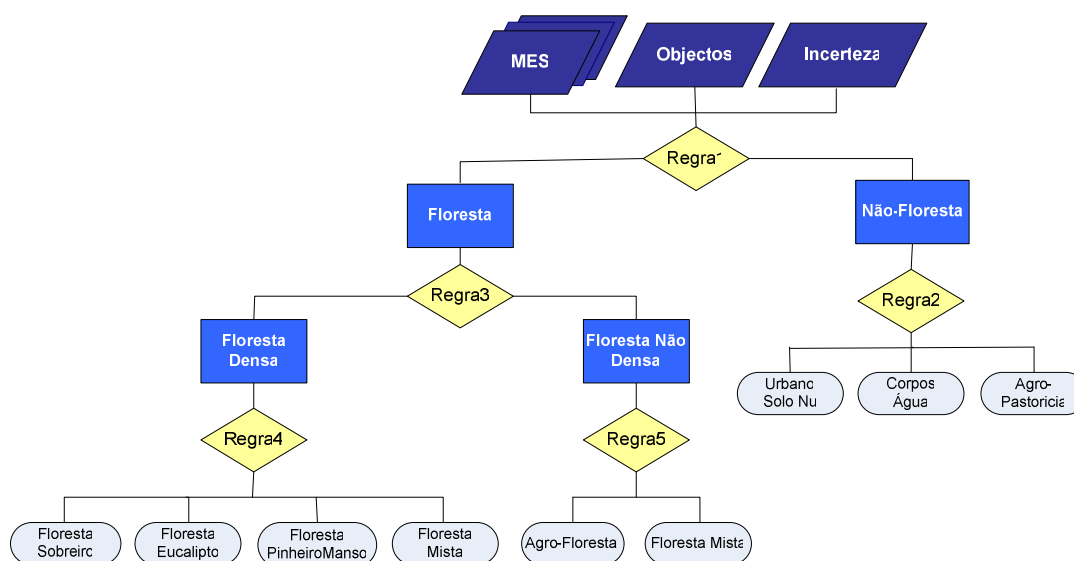


Figura 3.7 – Esquema de classificação das unidades de paisagem.

3.4.4.2 Abordagem de classificação combinada sem integração de incerteza

Para avaliar se a utilização da incerteza no processo de classificação combinada melhora os resultados finais da classificação, foi considerada a aplicação de um método semelhante ao previamente descrito, sem utilizar a informação da incerteza. A identificação dos elementos de superfície foi efectuada com o classificador da máxima verosimilhança, tendo sido usada a mesma nomenclatura de classificação e os mesmos protocolos de amostragem para treino do classificador, mas não foram obtidos os graus de probabilidade *a posteriori*. O classificador da máxima verosimilhança atribui a cada *pixel* a classe mais provável, em função da dimensão das elipses formadas pelas áreas de treino. A cada *pixel* fica apenas associada a informação da classe a que foi atribuído com maior probabilidade. As regras de transformação do MES no MUP foram muito semelhantes mas não incluem critérios baseados na incerteza. A avaliação da exactidão do MES e do MUP resultante foi efectuada com os mesmos protocolos de amostragem expostos na secção 3.4.1.

3.4.4.3 Avaliação da exactidão temática do mapa de unidades de paisagem

Segundo Stehman e Czaplewski (1998), a escolha da unidade da amostra deve ser efectuada, por exemplo, em função das características da paisagem, das características do processo de obtenção do mapa ou dos objectivos do projecto. Como neste estudo o objectivo consistia em avaliar a exactidão de um MUP vectorial com uma UMC média de 0.5 ha, a escolha da unidade amostra para a avaliação da exactidão do mapa foram regiões com uma área fixa de 0.5 ha conforme referido na secção 3.4.1. Para a escolha das amostras foi definida uma malha regular quadrada, com 0.5 ha por quadrícula, e escolhidas 40 unidades amostrais por classe do MUP, utilizando o método de amostragem aleatório estratificado, o que garantiu um erro padrão da ordem dos 8% para os valores dos IPEE (PCM, PCR) estimados para cada classe, conforme referido na secção 3.4.1. Este valor é obtido assumindo que a exactidão da classificação é superior a 50% (Wickham *et al.*, 2004), o que é aceitável pois a construção do MUP envolveu uma classificação ao nível do *pixel* e uma análise do terreno.

A referência foi determinada através de interpretação visual de imagens aéreas com 0.5 m de resolução. Na atribuição das classes às amostras foi tida em consideração não só a área de 0.5 ha contida na amostra, mas também a sua envolvência. Quando as unidades de amostra continham mais do que uma unidade de paisagem, eram repartidas em polígonos, em função das unidades de paisagem nela contidas. A repartição das unidades amostra fez-se através de interpretação visual, tendo em conta a UMC e a ocupação em redor da amostra.

A avaliação da exactidão foi efectuada com uma matriz de confusão probabilística em que os valores de entrada p_{ij} na matriz de erro são as proporções da área que é classe i no mapa e classe j na referência em vez do número de amostras. A partir da matriz de confusão foram obtidos os parâmetros de exactidão PCM e PCR (Stehman e Czaplewski, 1998).

3.5 Resultados e discussão

3.5.1 Avaliação dos classificadores

3.5.1.1 Avaliação dos classificadores com base em índices de exactidão estatísticos

A Tabela 3.7 apresenta os valores dos Índices de Exactidão Global (EG e KHAT) derivados das matrizes de confusão dos MES para cada um dos classificadores.

Tabela 3.7 – Exactidão Global (EG) e coeficiente de KHAT (KHAT) para os classificadores fuzzy (F_DM) e probabilístico (MV_B).

| (%) | EG | KHAT |
|------|-------|-------|
| MV_B | 96.75 | 96.32 |
| F_DM | 84.54 | 82.55 |

Os valores obtidos para a EG e KHAT foram bastante elevados para os dois classificadores verificando-se que, para o classificador probabilístico MV_B, obtêm-se valores superiores. É de notar que os elevados valores da exactidão obtidos com as áreas de teste podem não traduzir a verdadeira exactidão dos mapas. Por um lado, as áreas que foram utilizadas para avaliar o comportamento do classificador (áreas de teste) possuem uma assinatura espectral, para cada classe, muito semelhante à do conjunto amostral utilizado para treino do classificador. Por outro lado, neste conjunto amostral foram apenas utilizados *pixels* puros representativos dos elementos de superfície que se pretendiam classificar (Pontius e Cheuk, 2006). Estes factos normalmente têm como consequência uma sobrestimação dos valores da exactidão obtidos.

Na Tabela 3.8 e na Tabela 3.9 apresentam-se as matrizes de confusão para a avaliação dos classificadores MV_B e F_DM. Na Figura 3.8 e na Figura 3.9 apresentam-se os resultados dos Índices de Exactidão do Utilizador (EU) e Exactidão do Produtor (EP) de ambos os classificadores, ordenados por ordem crescente dos valores obtidos para o classificador *fuzzy*.

Tabela 3.8 – Matriz de confusão para avaliação do classificador probabilístico (MV_B).

| | | Ocupação na Referência | | | | | | | | | EU (%) |
|------------------|-----|------------------------|-------|------|------|------|-------|------|------|------|---------------|
| | | AP | APP | ANV | CE | S | H | CS | CPM | HE | |
| Ocupação no Mapa | AP | 166 | | 1 | | | | | | | 99.4 |
| | APP | 1 | 74 | | | | | | | | 98.7 |
| | ANV | 1 | | 177 | | 1 | | | | | 98.9 |
| | CE | | | | 130 | | | | 3 | | 97.7 |
| | S | 1 | | 1 | | 138 | | | | | 98.6 |
| | H | | | | | | 201 | | | | 100.0 |
| | CS | | | 1 | 1 | | | 160 | | 14 | 90.9 |
| | CPM | | | | 8 | | | | 162 | | 95.3 |
| | HE | | | | | | | 12 | | 131 | 91.6 |
| EP (%) | | 98.2 | 100.0 | 98.3 | 93.5 | 99.3 | 100.0 | 93.0 | 98.2 | 90.3 | 96.75% |

Tabela 3.9 – Matriz de confusão para avaliação do classificador fuzzy (F_DM).

| | | Ocupação na Referência | | | | | | | | | EU (%) |
|------------------|-----|------------------------|------|------|------|------|------|------|------|------|---------------|
| | | AP | APP | ANV | CE | S | H | CS | CPM | HE | |
| Ocupação no Mapa | AP | 153 | | | | 2 | | | | | 98.7 |
| | APP | | 45 | 4 | | | | | | | 91.8 |
| | ANV | 16 | 28 | 150 | 5 | 4 | 15 | 7 | 15 | 22 | 57.3 |
| | CE | | | | 118 | | | 2 | 19 | | 84.9 |
| | S | | | | | 131 | | | | | 100.0 |
| | H | | | | | | 173 | | 1 | | 99.4 |
| | CS | | | | 2 | | | 145 | | 20 | 86.8 |
| | CPM | | | | 14 | | | | 130 | | 90.3 |
| | HE | | | | | | | 18 | | 103 | 85.1 |
| EP (%) | | 90.5 | 61.6 | 97.4 | 84.9 | 95.6 | 92.0 | 84.3 | 78.8 | 71.0 | 84.54% |

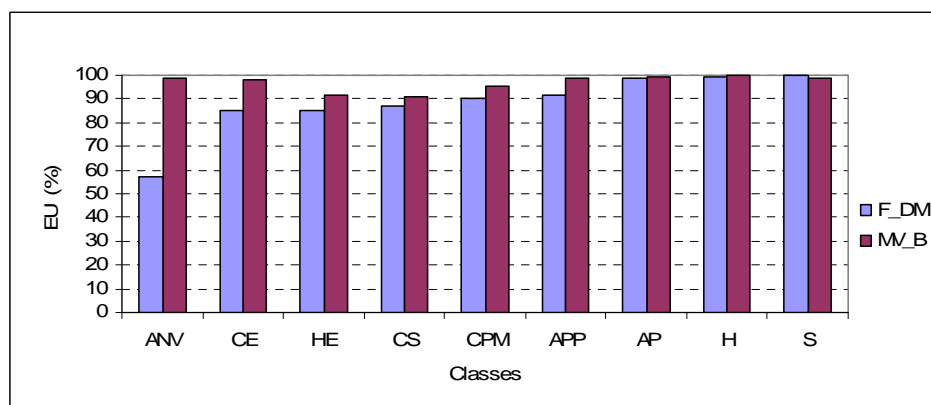


Figura 3.8 – Índices de Exactidão do Utilizador (EU) para os classificadores fuzzy (F_DM) e probabilístico (MV_B).

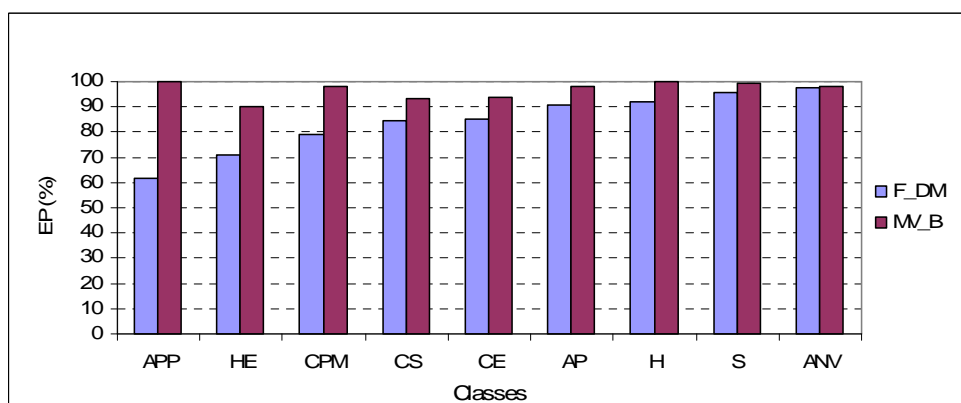


Figura 3.9 – Índices de Exactidão do Produtor (EP) para os classificadores fuzzy (F_DM) e probabilístico (MV_B).

Pela análise da Tabela 3.8 e Tabela 3.9 e da Figura 3.8 e Figura 3.9, pode constatar-se que, com o classificador MV_B, os valores de EU e EP são, para todas as classes, superiores a 90%. As classes que apresentam valores mais baixos de EU e EP são as classes Copa de Sobreiro (CS) e Herbáceas Esparsas (HE). A classe Copa de Sobreiro é a que apresenta mais erros de comissão e a classe Herbáceas Esparsas a que apresenta mais erros de omissão. Os valores dos índices de Exactidão Específicos (IEE), obtidos por classe, foram bastante elevados, o que indica uma boa capacidade do classificador em identificar as classes consideradas.

Relativamente ao classificador *fuzzy*, apenas as classes Herbáceas (H), Sombra (S) e Água Profunda apresentam valores de EU e EP superiores a 90%. Todas as classes, à excepção da classe Sombra (S), apresentam valores mais baixos de EU e EP, comparativamente aos obtidos com o classificador MV_B. A classe que apresenta o valor mais baixo de EP (62%) é a classe Água Pouco Profunda (APP), o que significa que é a classe que apresenta mais erros de omissão. A Água Pouco Profunda omitida foi classificada como Áreas Não Vegetadas. A classe que apresenta o valor mais baixo de EU (57%) é a classe Áreas Não Vegetadas (ANV) cujas comissões se dão com todas as classes conforme se pode constatar na matriz de confusão (Tabela 3.9). Este facto evidencia uma grande dificuldade do classificador *fuzzy* em classificar correctamente a classe Áreas Não Vegetadas (ANV) e a classe Água Pouco Profunda (APP). Os baixos resultados obtidos estão relacionados com: (1) a variabilidade espectral destas classes; (2) o facto da amostra apresentar uma grande variância da resposta espectral; e (3) a abordagem utilizada pelo classificador F_DM que considera, no processo de atribuição de uma dada classe a um *pixel*, a distância entre a sua resposta espectral e um valor ideal que é obtido a partir da média da resposta espectral das áreas de treino. O que contribui para que a classe Área Não Vegetada (ANV) apresente uma grande dispersão espectral é o facto de englobar

todos os tipos de material urbano como, por exemplo, estradas, telhados de zinco que apresentam uma elevada reflectância em todas as bandas e ainda todos os tipos de solo nu. Por outro lado, e ao contrário do que seria de esperar, a classe Água também possui uma grande dispersão espectral, o que pode ser explicado pelo facto de as zonas de água existentes nesta imagem possuírem diferentes profundidades.

Pelos resultados obtidos, pode concluir-se que o classificador probabilístico MV_B produz melhores resultados, tanto a nível global como na identificação das espécies florestais, superando o classificador *fuzzy*, baseado na distância mínima à média.

3.5.1.2 Avaliação dos classificadores com base em índices de incerteza

Para possibilitar a comparação dos índices de incerteza com os resultados obtidos através dos IEE, as medidas de incerteza foram aplicadas apenas ao conjunto amostral das áreas de teste, constituído por 1384 *pixels* amostra. A Figura 3.10 apresenta os valores da incerteza média por classe, obtidos com a aplicação das medidas RI e Er aos resultados da classificação probabilística e ordenados por ordem crescente dos valores de RI. A Figura 3.11 apresenta a incerteza distribuída por cinco níveis, correspondendo a 0% e aos quatro quantis: 1-25%, 26-50%, 51-75% e 76-100%. Na Figura 3.12 apresenta-se a correlação entre ambas as medidas para cada classe.

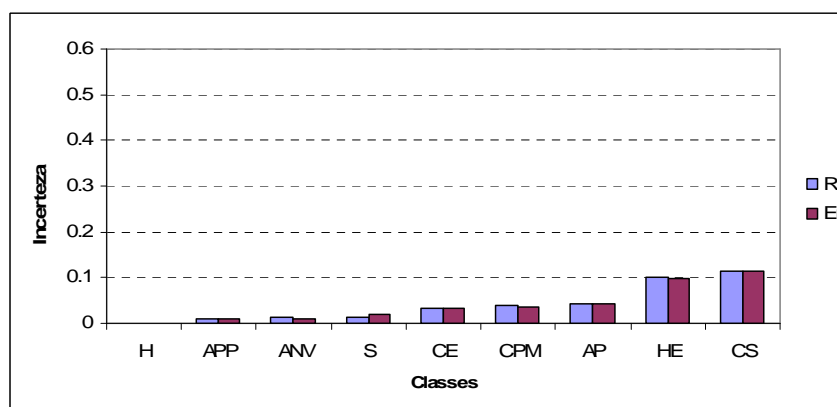


Figura 3.10 – Incerteza média por classe obtida com a aplicação das medidas de incerteza Er e RI aos resultados da classificação efectuada com o classificador probabilístico, para o conjunto amostral constituído pelas amostras de teste.

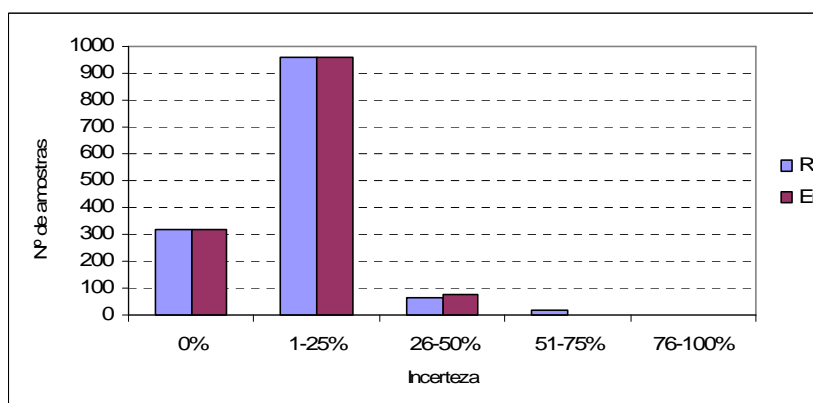


Figura 3.11 – Distribuição da incerteza da classificação probabilística, obtida com as medidas Er e RI, por cinco níveis de incerteza correspondendo a 0% e a quatro quantis.

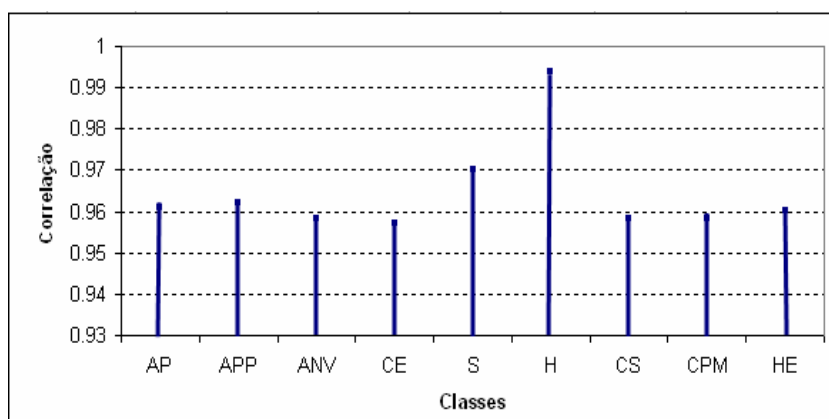


Figura 3.12 – Correlação por classe obtida entre as medidas Er e RI, para as amostras de teste.

Pela análise da Figura 3.10 e Figura 3.11 pode constatar-se que os resultados obtidos com a medida Er são quase idênticos aos obtidos com a medida RI, para o conjunto constituído pelas amostras de teste, registando a medida RI valores ligeiramente superiores para todas as classes, à excepção da classe Sombra (S). Os valores da incerteza média por classe obtidos com ambas as medidas é bastante baixo (inferior a 0.12) e o maior número de amostras apresenta uma incerteza compreendida entre 1-25%. Apenas com a medida RI foram obtidos resultados no intervalo 51-75%. As classes que apresentam valores mais elevados de incerteza são as classes Copas de Sobreiro (CS) e Herbáceas Esparsas (HE). Na análise de correlação entre as duas medidas por classe (Figura 3.12), observam-se valores elevados para todas as classes (superiores a 0.95), registando-se para a classe Herbáceas o valor mais elevado.

A proximidade da informação sobre a incerteza com as duas medidas deve-se ao facto de o máximo grau de probabilidade associada aos *pixels* amostra ser em geral bastante elevado; 88% dos *pixels* amostra apresentam valores para o máximo grau de probabilidade superior a 0.9. Isto

vem confirmar que, para distribuições de probabilidade em que a dispersão é baixa e o máximo grau de probabilidade é elevado, o comportamento destas duas medidas é muito semelhante, conforme referido no Capítulo 2.

Na Figura 3.13 é visível a distribuição espacial da incerteza da classificação obtida com as duas medidas. A partir da análise visual, é possível observar que a distribuição da incerteza com ambas é superior nas zonas florestais, em que há uma maior heterogeneidade de diferentes elementos de superfície. Verifica-se contudo que, com a medida Er, a distribuição da incerteza, nestas zonas, apresenta valores mais homogêneos do que com a medida RI, em que um considerável número de elementos de superfície apresentam incerteza superior a 0.50.

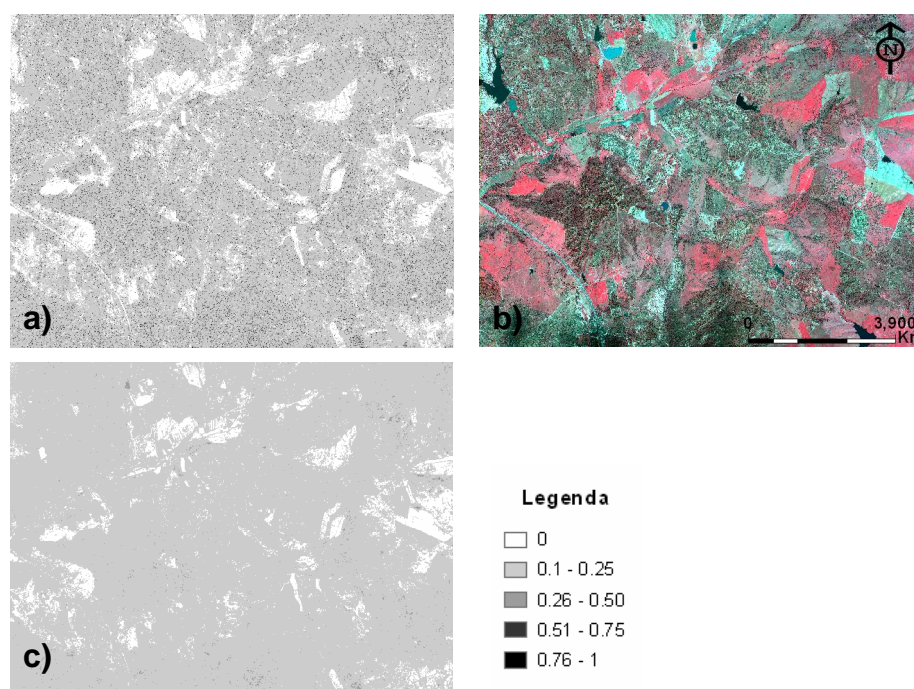


Figura 3.13 – Visualização da distribuição espacial da incerteza da classificação probabilística, obtida com as medidas Er e RI, considerando cinco níveis de incerteza correspondendo a 0% e a quatro quantis: a) incerteza obtida com a medida RI; b) imagem IKONOS da área de estudo (RGB 432); c) incerteza obtida com a medida Er.

A Figura 3.14 apresenta a incerteza média por classe, obtida com a aplicação das medidas NSp, Un e RI aos resultados da classificação obtidos com o classificador F_DM, ordenados por ordem crescente dos valores de NSp. A Figura 3.15 apresenta a incerteza distribuída por cinco níveis, correspondendo a 0% e a quatro quantis (1-25%, 26-50%, 51-75% e 76-100%). Na Figura 3.16 apresenta-se a correlação entre as medidas para cada classe.



Figura 3.14 – Incerteza média por classe obtida com a aplicação das medidas de incerteza *Er* e *RI* aos resultados da classificação efectuada com o classificador fuzzy (*F_DM*), para o conjunto amostral constituído pelas amostras de teste.

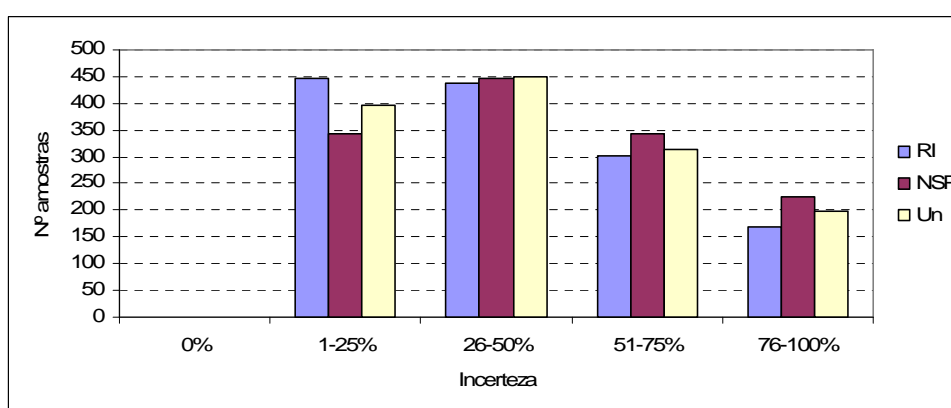


Figura 3.15 – Distribuição da incerteza da classificação fuzzy, obtida com as medidas *Er* e *RI*, considerando cinco níveis de incerteza, correspondendo a 0% e a quatro quantis.

Os valores da incerteza média por classe, obtidos com as medidas *NSp*, *Un* e *RI* aplicados à classificação efectuada com o classificador *F_DM*, apresentados na Figura 3.14, variam entre baixo (0.3) e médio (0.55). Embora o comportamento das várias medidas seja muito próximo, verifica-se que, com a medida *NSp* são obtidos, em geral, valores de incerteza superiores aos obtidos com as medidas *Un* e *RI*. A maior diferença dos valores obtidos entre *NSp* e *Un* foi de 0.09 e entre *NSp* e *RI* foi de 0.19, apresentando a medida *RI* os valores mais baixos. Verificou-se ainda algumas discrepâncias na tendência apresentada pelas várias medidas: (1) com a medida *NSp* e *Un*, a classe que apresentou o valor médio de incerteza mais baixo foi a Água Profunda (*AP*); e (2) com a medida *RI*, foi a classe Herbáceas Esparsas (*HE*). As classes Área Não Vegetada (*ANV*) e Água Pouco Profunda (*APP*) foram as que apresentaram níveis de incerteza superiores, com todas as medidas. Uma análise mais detalhada permitiu constatar que, com a medida *NSp*, 17% dos *pixels* apresentaram incerteza superior a 0.75. Com as medidas *Un*

e RI, a percentagem de *pixels* que apresentaram incerteza superior a 0.75, foi de 15% e 13% respectivamente, verificando-se que os *pixels* que apresentaram este valor de incerteza com a medida RI também apresentaram com as medidas Un e NSp. Foi realizada, ainda, uma análise aos valores dos graus de possibilidade associados aos *pixels* que apresentaram um valor de incerteza com a medida de NSp superior a 0.75, tendo-se constatado que o máximo grau de possibilidade era inferior a 0.47 e que 24% deste *pixels* eram atribuídos apenas a uma classe. Para a medida Un e RI observou-se que o máximo grau de possibilidade associado aos *pixels*, com este nível de incerteza, foi de 0.37 e 0.25, respectivamente. A classe Herbáceas Esparsas (HE) foi aquela em que se observou maiores discrepâncias entre os valores obtidos com as várias medidas. A análise de correlação entre as medidas por classe (Figura 3.16) reforça esta constatação. Efectivamente, embora os valores de correlação para todas as classes sejam elevados (superiores a 0.95), existe uma maior correlação entre as medidas RI e UN, do que entre as medidas RI e NSp e a classe para a qual o valor é menor é a classe Herbáceas Esparsas (HE).

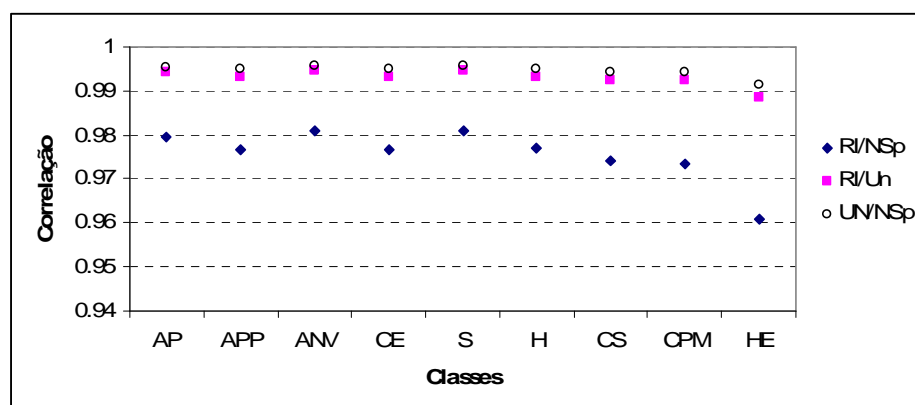


Figura 3.16 – Correlação por classe obtida entre as medidas RI, NSp e Un, para as amostras de teste.

Uma análise dos graus de possibilidade permitiu constatar que 27% dos *pixels* atribuídos à classe HE apresentavam um valor para o melhor grau de pertença inferior a 0.6, apresentando uma dispersão apenas entre duas classes, e que 63% destes pontos apresentavam uma diferença entre os valores atribuídos aos dois primeiros graus consecutivos superior a 0.3. Como a medida NSp é muito sensível à compatibilidade e pouco sensível à dispersão e à diferença entre os valores consecutivos de possibilidade, este facto explica que os valores da incerteza desta medida sejam ligeiramente mais elevados do que os obtidos com as medidas Un e RI. Além disso, e conforme referido na secção 2.4, embora a medida RI apresente um comportamento próximo da medida Un, é mais sensível à variação da dispersão. O facto de neste caso de estudo existir baixa dispersão e a diferença entre os valores consecutivos da

distribuição de probabilidade associada a cada *pixel* não revelar uma forte proximidade, explica os baixos valores da medida RI relativamente à medida Un e, especialmente, à medida NSp, que se traduzem numa menor correlação (ver Tabela 3.10).

Tabela 3.10 – Valores de correlação entre as diferentes medidas de incerteza.

| | Un | NSp | RI |
|-----|------|------|------|
| Un | 1.00 | 0.91 | 0.92 |
| NSp | 0.91 | 1.00 | 0.67 |
| RI | 0.92 | 0.67 | 1.00 |

Os valores apresentados na Tabela 3.10 mostram que a correlação entre as medidas RI e Un, é muito próxima da correlação entre NSp e Un, registando-se uma menor correlação entre as medidas RI e NSp.

Na Figura 3.17 é visível a distribuição espacial da incerteza da classificação obtida com as medidas Un, NSp e RI. A partir da análise visual é possível verificar que existe uma distribuição da incerteza por toda a imagem em que um elevado número de elementos de superfície apresenta valores de incerteza superiores a 0.5, sendo esta constatação comum às três medidas.

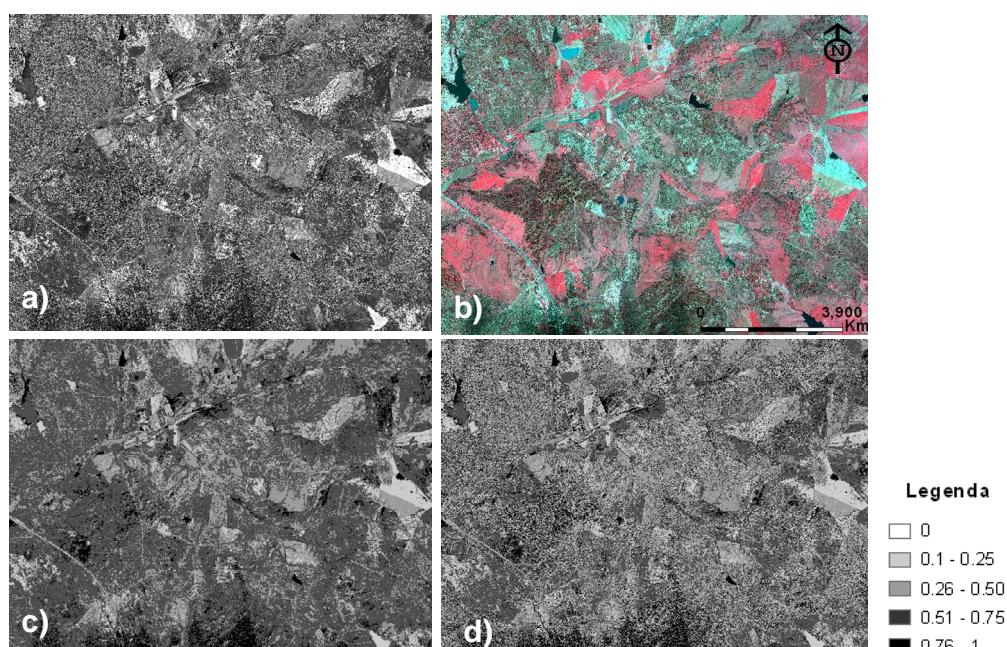


Figura 3.17 – Visualização da distribuição espacial da incerteza da classificação fuzzy, obtida com as medidas Un, NSp, RI, considerando cinco níveis de incerteza correspondendo a 0% e a quatro quantis: a) incerteza obtida com a medida NSp; b) imagem IKONOS da área de estudo (RGB 432); c) incerteza obtida com a medida Un; d) incerteza obtida com a medida RI.

Os resultados acima apresentados corroboram as conclusões do estudo apresentado no Capítulo 2 de que a medida NSp é essencialmente sensível aos graus com valores positivos de possibilidade e que a medida Un é essencialmente sensível à variação entre os valores consecutivos de possibilidade e à dispersão da classificação sobre todas as classes consideradas teoricamente possíveis. A medida RI apresenta um comportamento mais próximo da medida Un do que da medida NSp, embora seja sensível aos graus com valores positivos de possibilidade. Quando aplicada a distribuições de possibilidade, a medida RI é fortemente sensível à dispersão da classificação sobre todas as classes consideradas como teoricamente possíveis.

Da análise global dos resultados apresentados na Figura 3.10, Figura 3.11, Figura 3.14 e Figura 3.15 verifica-se que, com o classificador probabilístico MV_B, a incerteza média por classe é significativamente inferior à obtida com o classificador *fuzzy* F_DM, podendo concluir-se que o classificador MV_B produz melhores resultados, o que é consistente com as conclusões obtidas com a avaliação baseada nos IEE.

Na Figura 3.18, apresenta-se uma análise comparativa entre os valores de EU e EP e os índices I_{1-RI} e I_{1-Er} , ordenados por ordem crescente de I_{1-RI} , com a aplicação do classificador MV_B.

Na Figura 3.19, apresentam-se os valores de EU e EP e os índices $I_{RI<0.5}$ e $I_{Er<0.5}$, ordenados por ordem crescente dos valores de $I_{RI<0.5}$.

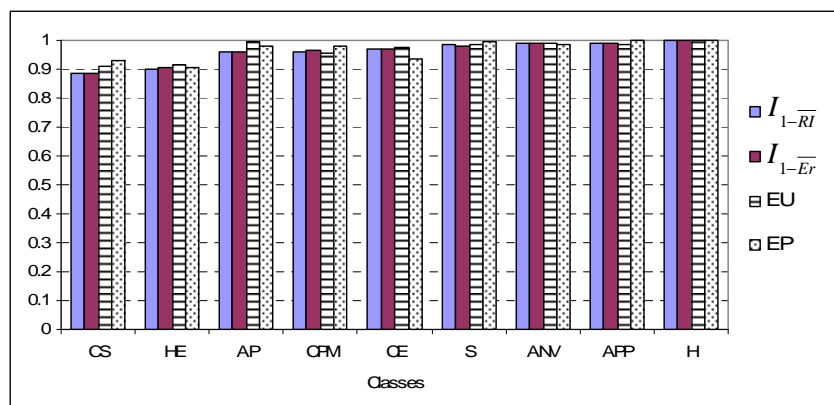


Figura 3.18 – Índices de incerteza I_{1-RI} e I_{1-Er} e Exactidão do Utilizador (EU) e do Produtor (EP) para o classificador probabilístico.

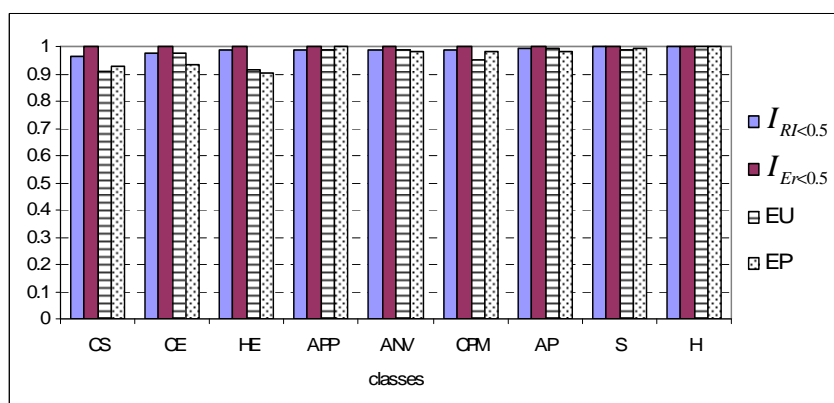


Figura 3.19 – Índices de incerteza $I_{RI<0.5}$ e $I_{Er<0.5}$ e Exactidão do Utilizador (EU) e do Produtor (EP) para o classificador probabilístico.

Dos resultados apresentados na Figura 3.18 e Figura 3.19 pode constatar-se que os valores dos índices de incerteza e dos IEE são muito elevados para todas as classes e muito semelhantes. Contudo, verifica-se que o valor da correlação entre os índices de incerteza, I_{1-RI} e I_{1-Er} , e os índices EP e EU é ligeiramente superior à obtida com os índices $I_{RI<0.5}$ e $I_{Er<0.5}$ (ver Tabela 3.12). Pode observar-se também que os índices de incerteza I_{1-Er} e I_{1-RI} dão resultados próximos mas são evidentes ligeiras discrepâncias entre os valores de $I_{RI<0.5}$ e $I_{Er<0.5}$. Com o índice $I_{Er<0.5}$ obteve-se total certeza para todas as classes, ou seja, todas as unidades amostra apresentaram uma incerteza inferior a 0.5, o que não se verificou com o outro índice. Na Figura 3.20, apresenta-se a correlação por classe entre EU e os índices I_{1-RI} e I_{1-Er} , onde se pode visualizar que existe uma boa correlação por classe e que a correlação entre cada um dos índices de incerteza e EU é muito semelhante. Todas as classes se encontram muito próximas da diagonal, à excepção da classe Água Profunda (AP). Na Figura 3.21, pode visualizar-se que a correlação por classe entre EU e o índice $I_{RI<0.5}$ não apresenta resultados tão bons, as classes Herbáceas Esparsa (HE), Copa de Sobreiro (CS) e Copa de Pinheiro Manso (CPM) são as classes que apresentam menor correlação. Entre EU e $I_{Er<0.5}$ a correlação é bastante baixa.

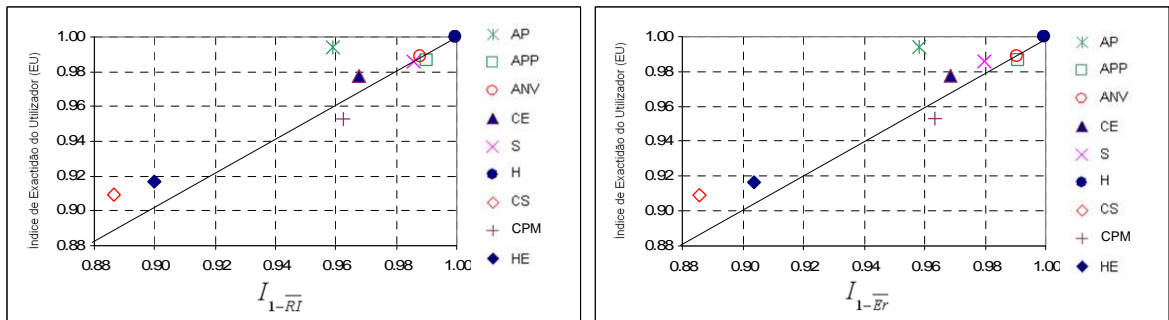


Figura 3.20 – Visualização da correlação entre o índice de Exactidão do Utilizador (EU) e os índices de incerteza I_{1-RI} , I_{1-Er} .

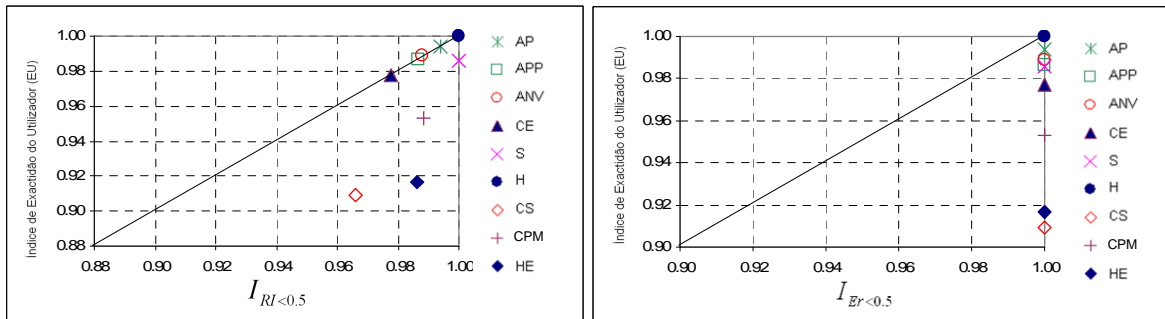


Figura 3.21 – Visualização da correlação entre o índice de Exactidão do Utilizador (EU) e os índices de incerteza $I_{RI < 0.5}$ e $I_{Er < 0.5}$.

Este comportamento do índice $I_{Er < 0.5}$ deve-se ao facto de o melhor grau de probabilidade associada aos *pixels* amostra ser, em geral, bastante elevado, 88% apresenta valores superiores a 0.9. Além disso, existe também uma baixa dispersão da classificação; 58% dos *pixels* apresentam uma dispersão em apenas três classes. Consequentemente, o facto de a medida Er ser muito sensível à dispersão explica que os valores da incerteza resultantes sejam todos inferiores a 0.5. A medida RI, quando aplicada a distribuições de probabilidade, embora seja sensível à dispersão, é mais sensível à compatibilidade, nomeadamente ao valor do maior grau de probabilidade, do que a medida Er. Por exemplo, para a seguinte distribuição de probabilidade (0.45,0.40,0.15,0,0,0,0,0), o valor de incerteza Er é 0.46 e, com a medida de incerteza RI, é 0.61. Os índices de incerteza I_{1-RI} e I_{1-Er} fornecem, para este caso de estudo, uma informação mais próxima da obtida com os IEE, nomeadamente com EU, tendo-se obtido uma avaliação do classificador MV_B muito semelhante. As classes que apresentam maior e menor exactidão são também as que apresentam maior e menor certeza.

Na Figura 3.22, apresenta-se uma comparação entre os valores de EU e EP e os índices I_{1-NSp} ,

$I_{1-\overline{Un}}$ e $I_{1-\overline{RI}}$, ordenados por ordem crescente de $I_{1-\overline{NSp}}$, com a aplicação do classificador *fuzzy*. Na Figura 3.23, apresentam-se os valores de EU e EP e os índices $I_{NSp<0.5}$, $I_{Un<0.5}$ e $I_{RI<0.5}$, ordenados por ordem crescente dos valores de $I_{NSp<0.5}$.

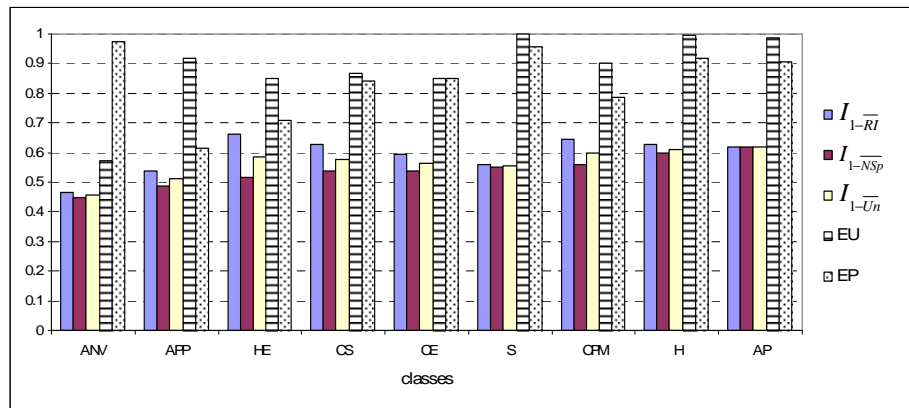


Figura 3.22 – Índices de incerteza $I_{1-\overline{NSp}}$, $I_{1-\overline{Un}}$, $I_{1-\overline{RI}}$ e Exactidão do Utilizador (EU) e do Produtor (EP) para o classificador F_{DM} .

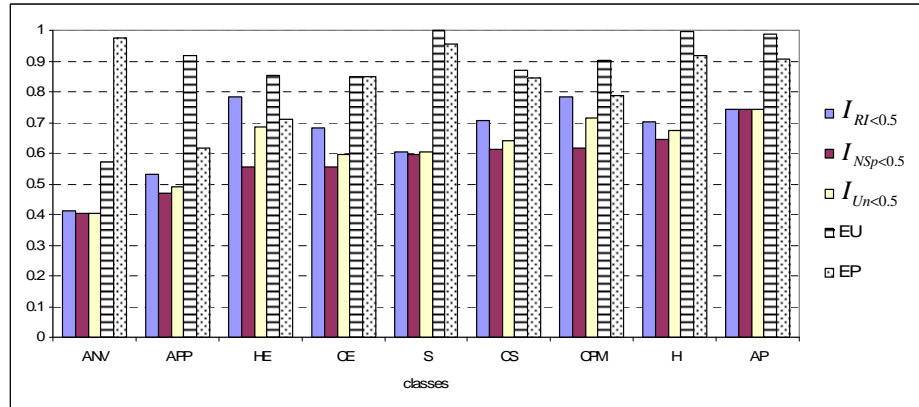


Figura 3.23 – Índices de incerteza $I_{RI<0.5}$, $I_{NSp<0.5}$, $I_{Un<0.5}$ e Exactidão do Utilizador (EU) e do Produtor (EP) para o classificador F_{DM} .

Os aspectos principais evidenciados pelos gráficos da Figura 3.22 e Figura 3.23 são: (1) para todas as classes, os valores da certeza variam entre baixo e médio; (2) a correlação entre os índices de incerteza e os IEE é menos evidente do que a correlação observada quando aplicados aos resultados da classificação obtida com o classificador MV_B (ver Figura 3.18, Figura 3.19 e Tabela 3.12). Para procurar compreender a razão dos valores dos índices de incerteza serem tão baixos, foi efectuada uma análise aos valores da distribuição de possibilidade associados aos *pixels* amostra das áreas de teste. A análise mostrou que, para todas as classes, existe uma

elevada percentagem de amostras que apresentam um baixo valor para o maior grau de possibilidade. Por exemplo, 25% dos *pixels* amostra atribuídos a Água Profunda (AP) foram atribuídos com um grau de possibilidade inferior a 0.5, o que explica os baixos valores de certeza. Verificou-se também que nenhuma das amostras foi classificada com o valor mais elevado de possibilidade (1). Para todos os *pixels*, o valor do segundo maior grau foi atribuído à classe Áreas Não Vegetadas (ANV) com valores muito próximos dos atribuídos ao maior grau de possibilidade, excepto para os *pixels* cujo melhor grau de pertença foi atribuído à classe ANV (ver Tabela 3.11). Isto significa que existe uma grande confusão entre a classe ANV e as restantes, o que constitui um indicador de que um dos grandes problemas da aplicação deste classificador, a este caso de estudo, é a dificuldade em discriminar a classe ANV das restantes classes.

Tabela 3.11 – Média dos valores do melhor grau e segundo grau de possibilidade associados aos *pixels* amostra.

| Melhor Classe | Média $\pi(x_1)$ | Média $\pi(x_2)$ | Segunda Classe | % de <i>pixels</i> com segunda classe |
|---------------|------------------|------------------|----------------|---------------------------------------|
| APP | 0.62 | 0.19 | ANV | 100 |
| HE | 0.63 | 0.38 | ANV | 100 |
| S | 0.38 | 0.03 | ANV | 100 |
| H | 0.57 | 0.16 | ANV | 95 |
| CS | 0.54 | 0.22 | ANV | 90 |
| CE | 0.48 | 0.14 | ANV | 88 |
| CPM | 0.52 | 0.22 | ANV | 82 |
| AP | 0.53 | 0.003 | ANV | 26 |
| ANV | 0.41 | 0.07 | H | 16 |

Esta conclusão foi também obtida através da matriz de confusão (Tabela 3.9). A classe ANV apresenta uma diferença muito elevada entre os valores de EU (57,3%) e EP (97,4%), o que significa que esta classe apresenta um elevado número de erros de comissão e, portanto, amostras que foram atribuídas à classe ANV deveriam ter sido atribuídas a outras classes como, por exemplo, às classes Herbáceas Esparsas (HE) e Água Pouco Profunda (APP), provocando um aumento dos erros de omissão destas últimas.

Na Figura 3.24, apresenta-se a correlação por classe entre o índice EU e os índices de incerteza, $I_{NSp < 0.5}$, $I_{Un < 0.5}$, $I_{RI < 0.5}$ e, I_{1-NSp} , I_{1-Un} , I_{1-RI} , onde se pode visualizar que existe uma menor

correlação entre este índice de exactidão e os índices de incerteza, quando aplicados aos resultados da classificação obtida com o classificador F_DM.

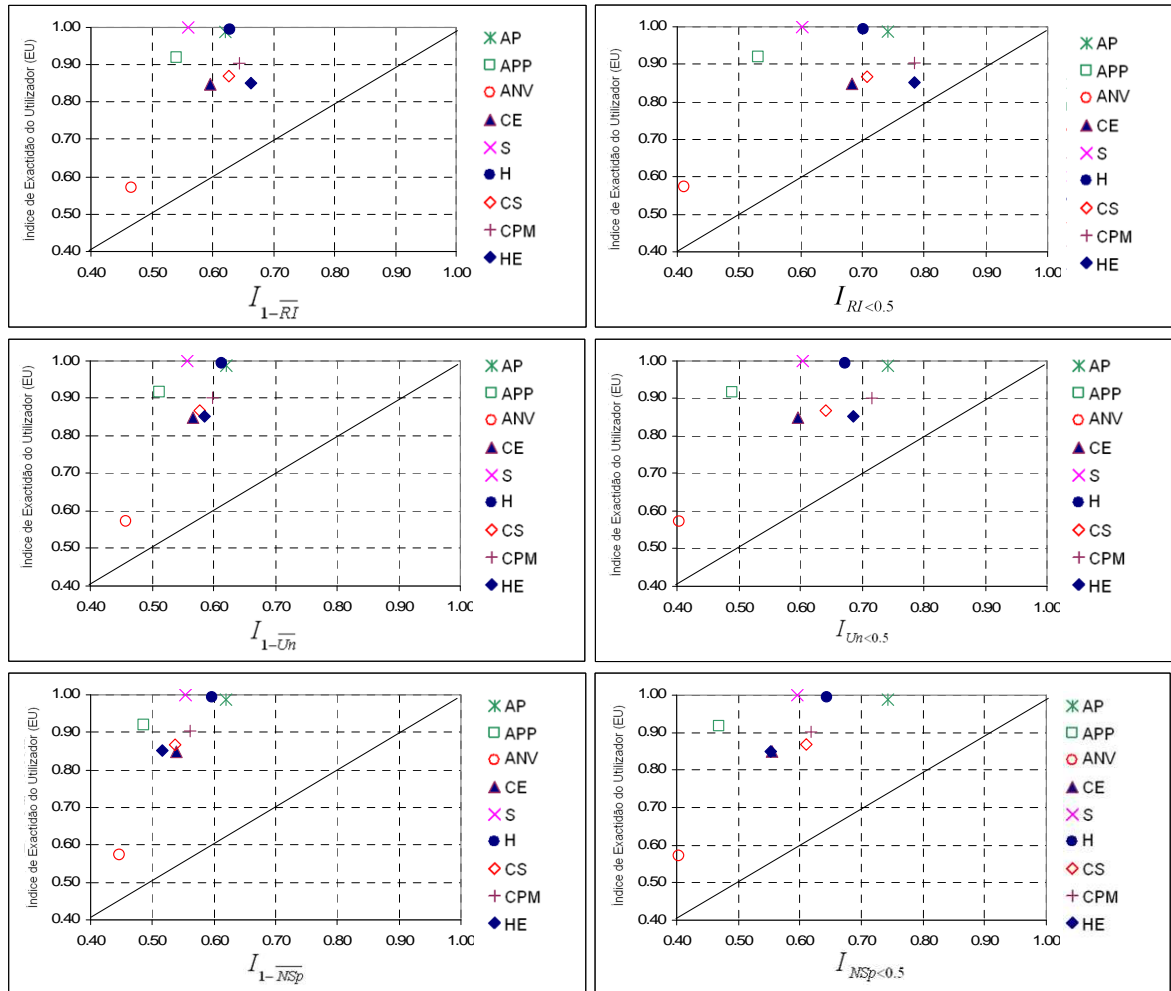


Figura 3.24 – Correlação entre o índice de Exactidão do Utilizador (EU) e os índices de incerteza

$$I_{RI<0.5}, I_{N\bar{S}p<0.5}, I_{U\bar{n}<0.5} \text{ e } I_{1-N\bar{S}p}, I_{1-U\bar{n}}, I_{1-RI}.$$

Relativamente à menor correlação entre a informação da incerteza e EU, obtida com a aplicação do classificador F_DM (ver Tabela 3.12), as maiores discrepâncias verificaram-se nas classes Sombra (S), Herbáceas (H) e Água Pouco Profunda (APP). A classe Herbáceas, por exemplo, é uma das classes que, com este classificador, apresenta os valores de exactidão mais elevados, o que significa que, de acordo com os índices estatísticos, foi bem classificada embora com níveis de certeza inferiores. Na Figura 3.25 apresenta-se a soma dos valores dos três melhores graus de possibilidade associados às 174 unidades amostra classificadas no mapa como Herbáceas (H). A soma do melhor grau de possibilidade destas unidades amostra à classe Herbáceas é apenas de 111.17. O segundo melhor grau de possibilidade foi atribuído à classe ANV, para

todas as unidades amostrais, e a sua soma foi de 14.41. O terceiro melhor grau de possibilidade foi atribuído à classe Copas de Pinheiro Manso (CPM) e a sua soma foi de 0.04. Estes resultados mostram que uma quantidade considerável de *pixels* foi atribuída à classe H com valor baixo de possibilidade e que existe uma confusão significativa com a classe ANV.

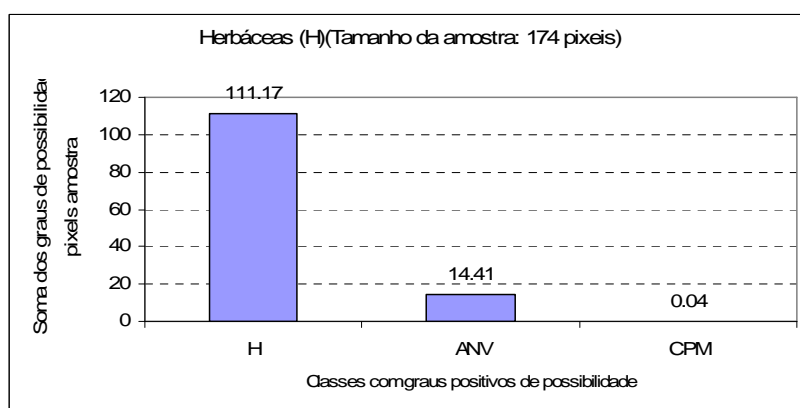


Figura 3.25 – Soma dos valores dos três melhores graus de possibilidade associados às unidades amostra classificadas no mapa como classes Herbáceas (H).

Uma análise à forma como o classificador funciona, permite explicar a razão dos graus de possibilidade obtidos com este classificador terem sido tão baixos. Com a aplicação do classificador F_DM os graus de possibilidade reflectem a proximidade da resposta espectral do conjunto de amostras a um valor ideal. Esse valor ideal é a média das respostas espectrais obtidas com as áreas de treino, que corresponde a um ponto. Os graus de possibilidade são então obtidos calculando a distância entre a resposta espectral de cada amostra de teste ao valor médio ideal. Consequentemente, se o valor de uma amostra de teste for relativamente distante da média, mesmo que a sua resposta espectral seja próxima de alguns valores obtidos para as áreas de treino, é atribuído um valor baixo de possibilidade. Os graus de possibilidade reflectem igualmente as características da nuvem de pontos que compõe a amostra de treino, pois a grandeza da variância da amostra de treino vai-se reflectir no cálculo do valor médio ideal e, consequentemente, nos valores dos graus de possibilidade obtidos.

Os resultados obtidos com os dois classificadores mostram que a informação dos índices de incerteza, conjuntamente com a análise dos graus de possibilidade ou probabilidade, permite detectar os principais problemas do comportamento do classificador que, consequentemente, são responsáveis pelo resultado final da classificação e que podem ser causados por vários factores como sejam: as imagens, o algoritmo ou a abordagem de classificação, a nomenclatura, o protocolo de treino e/ou o teste da classificação. Com a aplicação dos dois classificadores,

obteve-se uma correlação positiva entre os índices de incerteza e os IEE, especialmente com EU, conforme se pode constatar da Tabela 3.12, o que permite inferir que os índices de incerteza podem ser muito úteis no processo de avaliação do comportamento do classificador e no processo de decisão do classificador mais adequado.

Tabela 3.12 – Correlação entre os índices de incerteza e os IEE para os classificadores fuzzy (F_DM) e probabilístico (MV_B).

| | Classificador F_DM | | | | | | Classificador MV_B | | | |
|----|------------------------|-----------------------|-----------------------|---------------|--------------|--------------|-----------------------|-----------------------|--------------|--------------|
| | $I_{1-\overline{NSp}}$ | $I_{1-\overline{Un}}$ | $I_{1-\overline{RI}}$ | $I_{NSp<0.5}$ | $I_{Un<0.5}$ | $I_{RI<0.5}$ | $I_{1-\overline{RI}}$ | $I_{1-\overline{Er}}$ | $I_{En<0.5}$ | $I_{RI<0.5}$ |
| EU | 0.80 | 0.76 | 0.59 | 0.76 | 0.69 | 0.58 | 0.94 | 0.93 | ----- | 0.69 |
| EP | 0.26 | -0.01 | -0.27 | 0.21 | -0.03 | -0.21 | 0.86 | 0.84 | ----- | 0.67 |

3.5.2 Avaliação da exactidão temática do mapa de elementos de superfície

3.5.2.1 Avaliação da exactidão temática do mapa de elementos de superfície com base em índices de exactidão estatísticos

A avaliação da exactidão da classificação do MES foi efectuada apenas para o mapa obtido com o melhor classificador. Na tabela 3.13 apresentam-se os valores da Exactidão Global (EG), coeficiente de KHAT (KHAT) e Probabilidade Global (PG). Na Figura 3.26 apresentam-se os valores da Probabilidade Condicionada pelo Mapa (PCM) e da Probabilidade Condicionada pela Referência (PCR), ordenados por ordem crescente dos valores da PCM.

Tabela 3.13 – Índices de Exactidão Gobal (EG e KHAT) e Probabilidade Global (PG).

| (%) | EG | KHAT | PG |
|-----|------|------|------|
| | 71.4 | 67.8 | 66.5 |

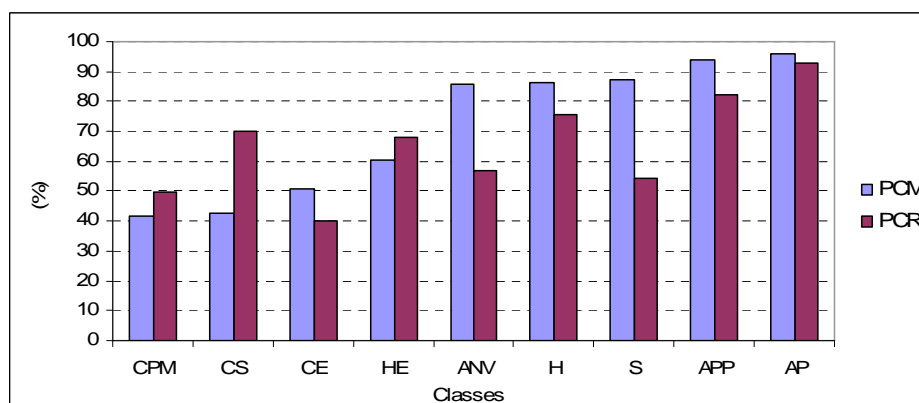


Figura 3.26 – Probabilidade Condicionada pelo Mapa (PCM) e Probabilidade Condicionada pela Referência (PCR) do Mapa de Elementos de Superfície.

Os valores para a exactidão da PCM e da PCR por classe, obtidos com as amostras de referência, foram bastante inferiores aos obtidos com as áreas de teste. Esta diminuição está relacionada com a definição do protocolo de amostragem de referência. Nesta avaliação, conforme acima exposto, não se utilizaram amostras idênticas às do treino do classificador. Foi considerado um conjunto amostral aleatório estratificado por classe, lançado por toda a imagem para que qualquer *pixel* pudesse ser seleccionado, incluindo os *pixels* mistos.

As classes Água (AP e APP) e Herbáceas (H) foram bem identificadas, podendo-se constatar que as classes não vegetadas apresentam melhores resultados do que as classes vegetadas. Uma confusão significativa ocorreu entre as classes Copas de Sobreiro (CS) e Herbáceas Esparsas (HE) e entre Copas de Pinheiro Manso (CPM) e Herbáceas (H). Com efeito, 22% dos *pixels* que, na referência, foram considerados Herbáceas Esparsas foram classificados como Copas de Sobreiro e 24% dos *pixels* que, na referência, foram considerados Herbáceas (H) foram classificados como Copas de Pinheiro Manso. Esta confusão deve-se à proximidade das suas assinaturas espectrais. As classes que apresentaram valores mais baixos para a exactidão da PCM e PCR foram as classes Copas de Pinheiro Manso (CPM) e Copas de Eucalipto (CE). As classes que apresentaram maiores erros de comissão foram as classes Copas de Sobreiro (CS) e Copas de Pinheiro Manso (CPM). A classe Copas de Sobreiro (CS) apresentou comissões com todas as classes à excepção da classe Águas Profundas (AP), as principais omissões dão-se com a classes Copas de Eucalipto.

3.5.2.2 Avaliação da informação temática do mapa de elementos de superfície com base em índices de incerteza

Com base na informação da incerteza das amostras de referência ao nível do *pixel* foi calculada a incerteza média por classe, obtida com a aplicação das medidas RI e Er (Figura 3.27). A Figura 3.28 apresenta a incerteza distribuída por cinco níveis, correspondendo a 0% e a quatro quantis (1-25%, 26-50%, 51-75% e 76-100%). Na Figura 3.29 apresenta-se a correlação entre ambas as medidas para cada classe.

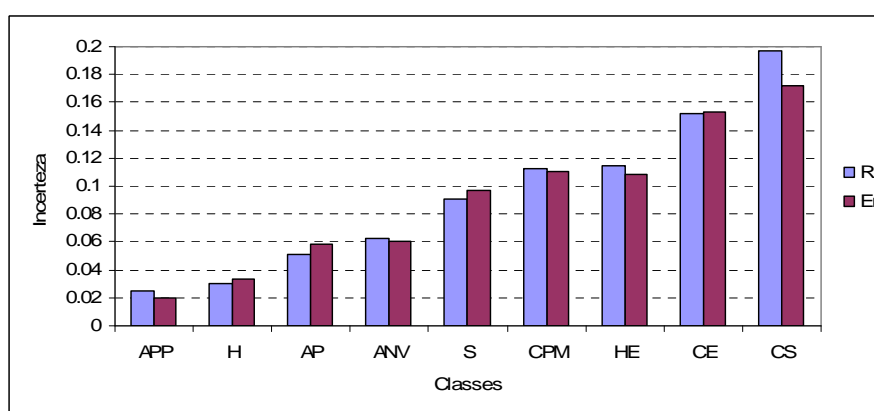


Figura 3.27 – Incerteza média por classe obtida com a aplicação das medidas de incerteza Er e RI aos resultados da classificação efectuada com o classificador probabilístico, para o conjunto amostral constituído pelas amostras de referência.

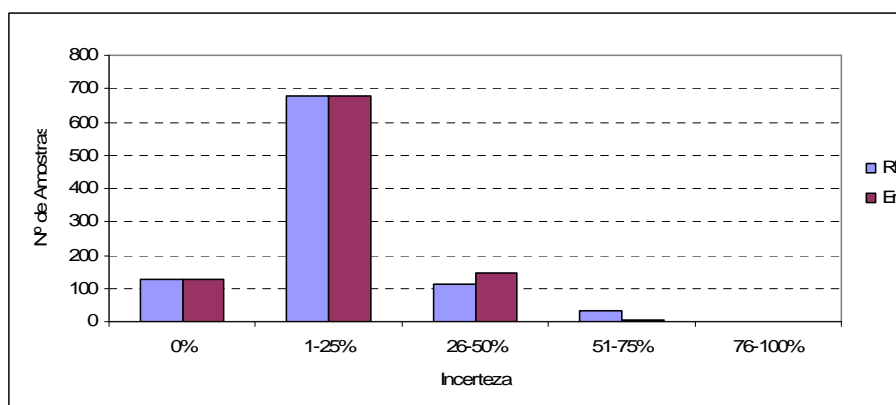


Figura 3.28 – Distribuição da incerteza da classificação probabilística, obtida com as medidas Er e RI, considerando cinco níveis de incerteza correspondendo a 0% e a quatro quantis.

Da análise da Figura 3.27 e Figura 3.28 pode constatar-se que os resultados obtidos com as medidas Er e RI apresentam, para a maioria das amostras, uma incerteza cujo valor pertence ao intervalo 1-25%. Verifica-se que, com a medida RI, um maior número de amostras apresenta

resultados no intervalo 51-75%. Os valores da incerteza média por classe, obtidos com as duas medidas, é bastante baixo, inferior a 0.2. Os valores obtidos com ambas são bastante próximos, à excepção da classe Copa de Sobreiro (CS), em que a medida RI regista valores superiores. Conforme seria expectável, verifica-se que, com as amostras de referência, todas as classes apresentam valores médios da incerteza superiores aos obtidos com as amostras de teste. Verifica-se ainda que as classes Copa de Eucalipto (CE) e Copa de Sobreiro (CS) são as que apresentam um maior aumento dos valores da incerteza.

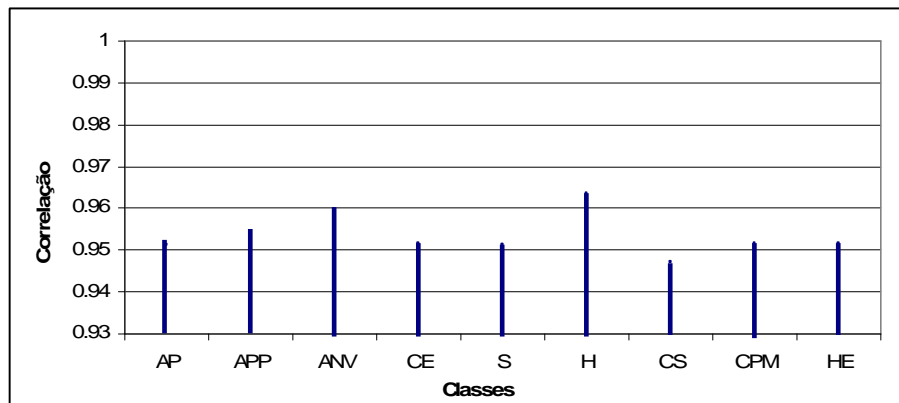


Figura 3.29 – Correlação por classe obtida entre as medidas E_r e RI para as amostras de referência.

Na análise de correlação entre as duas medidas por classe, observam-se valores elevados para todas as classes mas consta-se que são ligeiramente inferiores aos obtidos quando se consideram as amostras de teste (ver Figura 3.12). Continua a registar-se, para a classe Herbáceas (H), o valor mais elevado.

A comparação da informação da incerteza média por classe (Figura 3.27) com os Índices Probabilístico de Exactidão Específica (IPEE) apresentados na Figura 3.26 permite constatar que as espécies florestais são as que apresentam valores mais elevados de incerteza, sendo igualmente as que apresentam os valores mais baixos de exactidão. Na Figura 3.30 apresentam-se os valores da PCM e PCR e os índices I_{1-RI} , ordenados por ordem crescente de I_{1-RI} .

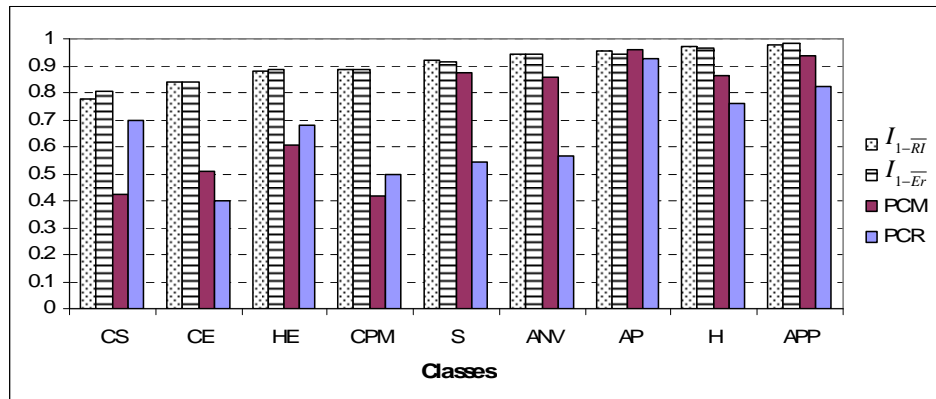


Figura 3.30 – Índices de incerteza, $I_{1-\bar{R}I}$ e $I_{1-\bar{E}r}$, Probabilidade Condicionada pelo Mapa (PCM) e a Probabilidade Condicionada pela Referência (PCR).

Os resultados apresentados na Figura 3.30 evidenciam que os valores obtidos com o índice de incerteza baseado na média são, de uma forma geral, superiores aos dos IPEE. As classes não vegetadas são as que apresentam uma maior concordância entre os índices de incerteza e os IPEE e são também aquelas que apresentam valores superiores com ambos. Contudo, nas classes florestais existe uma grande discrepância entre os dois tipos de índices. Estas classes apresentam valores de exactidão que variam entre médio e baixo (40% e 70%) mas valores elevados de certeza na classificação (80% a 90%).

Efectuando uma análise aos valores da distribuição de probabilidade associada a cada *pixel* amostra, constata-se que as classes florestais (CS, CE, CPM) e a classe Herbáceas Esparsas (HE) são as que apresentam maior dispersão (Tabela 3.14). Contudo, embora as classes florestais sejam as que apresentam maior dispersão verifica-se, simultaneamente, que os valores dos melhores graus de probabilidade são elevados. Por exemplo, a classe Copa de Eucalipto (CE), embora apresente uma dispersão por 6 classes, os valores são residuais (ver Tabela 3.14) e 58% dos *pixels* amostra apresentam valores para o melhor grau de probabilidade superiores a 0.90, o que explica os valores elevados obtidos com as medidas de incerteza.

A avaliação efectuada com o índice de incerteza baseado na média, conjuntamente com a análise dos graus de probabilidade, permitiu concluir que as classes APP, H, AP, ANV e S são aquelas que apresentaram valores mais elevados, podendo inferir-se que foram as classes em que há maior compatibilidade e menos incerteza na atribuição de uma solução única aos *pixels*. Para estas classes, verificou-se uma boa concordância entre os IPEE e o índice de incerteza baseado na média.

Tabela 3.14 – Soma das distribuições de probabilidade por classe.

| | | Somatório dos Graus de Probabilidade | | | | | | | | | Nº Amostras <i>n</i> | $\frac{\sum_{i=1}^n p_{ii} - n}{n}$ |
|------------------|-----|--------------------------------------|------|------|------|------|-------|------|------|------|-------------------------|-------------------------------------|
| | | AP | APP | ANV | CE | S | H | CS | CPM | HE | | |
| M A P A | AP | 98,36 | 3,7 | 0,2 | 0 | 0,7 | 0 | 0 | 0 | 0 | 103 | 0.95 |
| | APP | 1,9 | 96,8 | 0,3 | 0 | 0 | 0 | 0 | 0 | 0 | 99 | 0.98 |
| | ANV | 0 | 0 | 94,5 | 0 | 0,1 | 0 | 1,5 | 0 | 3,8 | 99 | 0.95 |
| | CE | 0 | 0 | 0,01 | 90,5 | 0,05 | 0,14 | 7,9 | 4,4 | 1 | 104 | 0.87 |
| | S | 0 | 0 | 2,8 | 0,07 | 95,2 | 0 | 4,9 | 0 | 0 | 103 | 0.92 |
| | H | 0 | 0 | 0,03 | 0 | 0 | 111,9 | 0,2 | 1,1 | 1,8 | 115 | 0.97 |
| | CS | 0 | 0 | 0,6 | 3,3 | 2,6 | 0,7 | 93,7 | 2,3 | 9,8 | 113 | 0.83 |
| | CPM | 0 | 0 | 0 | 1,73 | 0 | 3,8 | 4,7 | 95,4 | 0,3 | 106 | 0.90 |
| | HE | 0 | 0 | 1,73 | 0,14 | 0 | 3,3 | 5,9 | 0,03 | 98,1 | 109 | 0.90 |

Apesar das discrepâncias verificadas entre os IPEE e os índices de incerteza, sobretudo nas classes florestais, existe uma correlação positiva entre eles, nomeadamente entre a PCM (ver Tabela 3.15).

Tabela 3.15 – Correlação entre os índices de incerteza e os IPEE

| | Correlação | | | |
|-----|-----------------------|-----------------------|--------------|--------------|
| | $I_{1-\overline{RI}}$ | $I_{1-\overline{Er}}$ | $I_{En<0.5}$ | $I_{RI<0.5}$ |
| PCM | 0.87 | 0.87 | 0.91 | 0.79 |
| PCR | 0.45 | 0.51 | 0.30 | 0.25 |

Existe um aspecto relevante no processo de avaliação da exactidão por processos estatísticos, que é o facto de a classificação de referência estar dependente de um intérprete. Neste estudo, a informação da referência foi adquirida utilizando uma imagem de maior resolução utilizando técnicas de interpretação visual. O processo de construção da referência tem sempre associada a subjectividade inerente à interpretação visual afectando as exactidões do mapa, o que não acontece com os índices de incerteza.

3.5.3 Distribuição espacial da incerteza

Na Figura 3.31 b), apresenta-se a classificação obtida com o classificador MV_B (MES_MV) e,

nas imagens c), d) e e), a distribuição espacial dos valores da incerteza obtidos, respectivamente, com as medidas de RI, Er e E_x . Na Figura 3.31 f), encontra-se a classificação obtida com o classificador F_DM (MES_F) e as imagens g), h) e i) ilustram, respectivamente, a distribuição espacial sobre a incerteza obtida com as medidas RI, Un e E_x .

A partir da Figura 3.31 b) e f), pode observar-se que os resultados da classificação com o classificador MV_B e F_DM são muito diferentes. A avaliação do comportamento dos classificadores com os índices de exactidão estatísticos e com os índices de incerteza indicou que o classificador F_DM apresenta piores resultados do que o classificador MV_B. A visualização do resultado da classificação reforça claramente esta avaliação. Com o classificador F_DM, a maior parte da imagem foi classificada como Área Não Vegetada (ANV), enquanto que com o classificador MV_B obteve-se uma maior diversidade de classes, o que apresenta uma maior conformidade com a Figura 3.31 a).

As imagens apresentadas na Figura 3.31 c) e g) correspondem à distribuição espacial da informação sobre a incerteza obtida com a medida RI para as duas classificações. As zonas com valores de incerteza mais elevados, que correspondem às zonas mais escuras na imagem, são aquelas em que no processo de classificação ocorreu uma compatibilidade entre as características do *pixel* com várias classes, ou seja, regiões em que ocorreu mais dispersão. Estas imagens ilustram também os erros de omissão cometidos quando o *pixel* apresenta compatibilidade com várias classes mas, no resultado final, é apenas disponibilizada a informação sobre a classe para a qual ocorreu maior grau de compatibilidade. A partir da comparação visual destas duas imagens, verifica-se que os valores de incerteza mais elevados para o MES_MV ocorrem, essencialmente, nas zonas compostas por uma maior diversidade de diferentes elementos de superfície, como sejam as zonas florestais (zona indicada com seta preta) e Herbáceas Esparsas. No MES-F, verifica-se que os elevados valores de incerteza ocorrem, de uma forma generalizada, por toda a imagem. A mesma conclusão pode ser inferida a partir das imagens d) e h) da Figura 3.31. De notar a grande conformidade que existe entre a distribuição espacial de incerteza do MES-F, obtida com as medidas RI e Un. Este comportamento foi também relatado na secção 2.4, com a análise comparativa das diferentes medidas.

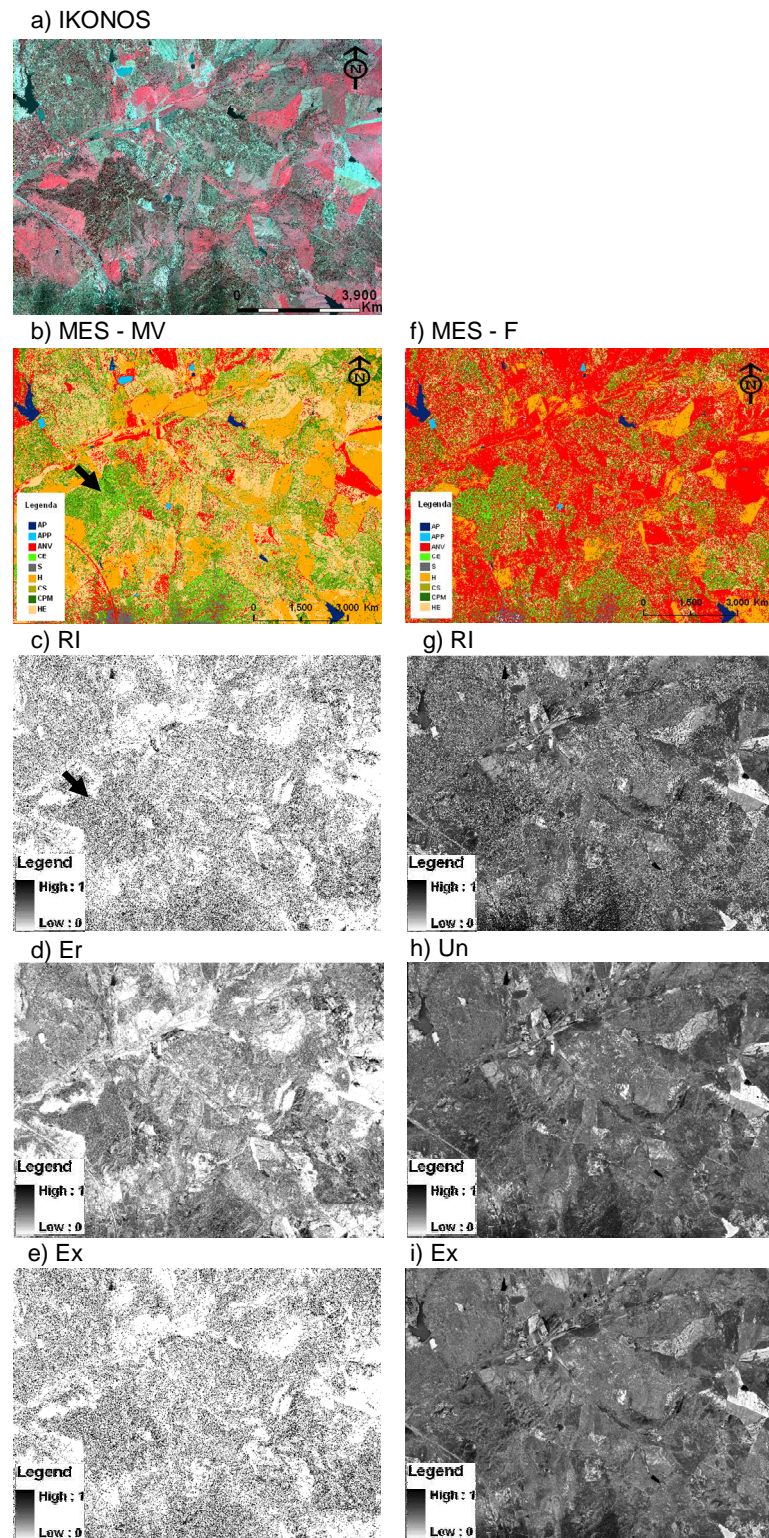


Figura 3.31 – a) imagem IKONOS (RGB 432); b) c) d) e) mostram, respectivamente, o MES-MV produzido com o classificador probabilístico (MV_B) e a distribuição espacial da incerteza obtida com as medidas de RI, Er e Ex. As imagens f), g), h) e i) mostram, respectivamente, o MES-F produzido com o classificador fuzzy (F_DM) e a distribuição espacial da incerteza obtida com as medidas de RI, Un e Ex.

As Figuras 3.31 e) e i) ilustram os erros de comissão que ocorreram com a escolha da melhor classe, ou seja, quando é considerada uma compatibilidade entre as características da classe e as do *pixel* superior à que realmente existe, o que se verifica quando o grau de compatibilidade da melhor classe escolhida é inferior a 1. Estas imagens mostram igualmente que, com o classificador *fuzzy*, foram obtidos erros de comissão superiores aos obtidos com o classificador MV_B. Comparando as imagens c) e e), constata-se que as zonas em que ocorreram maiores comissões foram também aquelas em que ocorreram maiores valores de omissão.

3.5.4 Avaliação da exactidão temática dos mapas de unidades de paisagem com e sem integração de incerteza: análise comparativa dos resultados

O valor da Probabilidade Global da classificação para o MUP produzido com o método que integra a informação sobre a incerteza foi de 66% e o valor obtido com o método que não integra essa informação foi de 54%. Apesar dos valores da exactidão não serem elevados, devido fundamentalmente a dificuldades em discriminar os diferentes elementos de superfície que representam as copas das espécies florestais e a problemas na identificação dos limites de algumas unidades de paisagem na fase de segmentação, obteve-se um melhoramento de 12% para o MUP obtido com a integração da informação sobre a incerteza.

A Figura 3.32 e a Figura 3.33 permitem a comparação entre os resultados dos IPEE obtidos na avaliação da exactidão da classificação com a aplicação das duas metodologias. Os resultados da classificação obtidos com o novo método, desenvolvido para integrar a informação sobre a incerteza, são consideravelmente melhores para todas as classes e essa constatação é mais evidente nas unidades de paisagem relativas às classes florestais.

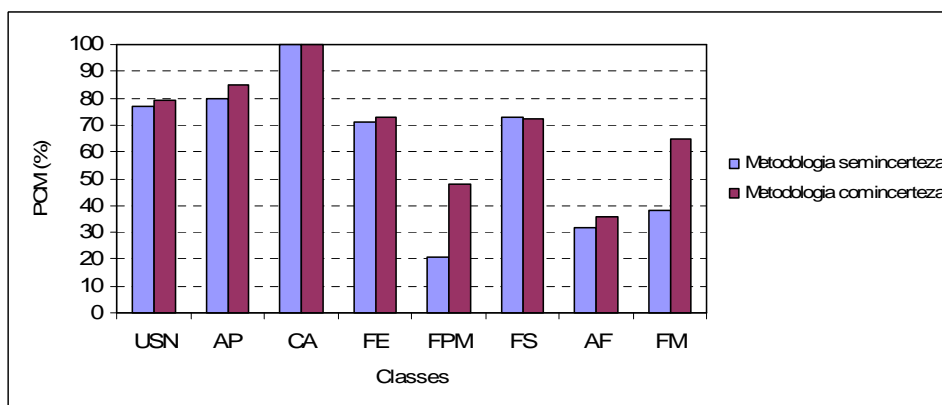


Figura 3.32 – Probabilidade Condicionada pelo Mapa (PCM).

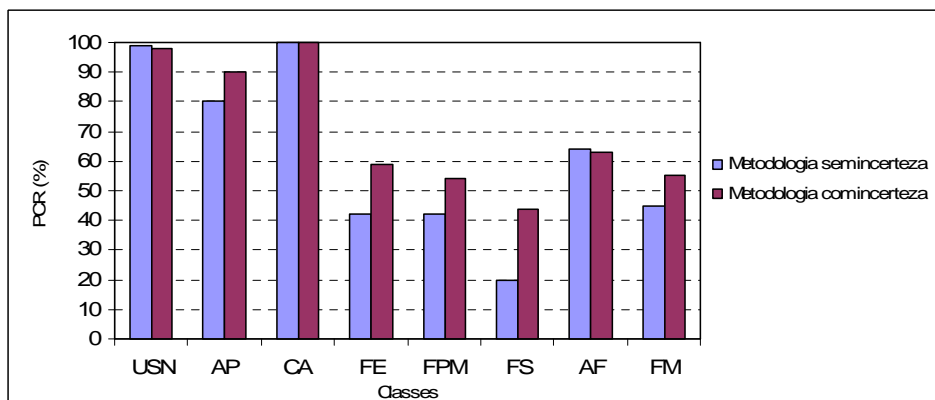


Figura 3.33 – Probabilidade Condicionada pela Referência (PCR).

Uma comparação entre a Figura 3.27, a Figura 3.32 e a Figura 3.33 mostra que as classes de unidade de paisagem que evidenciam um aumento mais significativo da exactidão com a integração da informação sobre a incerteza, foram as classes que resultaram do arranjo dos elementos de superfície que apresentavam elevados valores de incerteza como, por exemplo, Floresta de Eucalipto (FE) e Floresta de Pinheiro Manso (FPM). Este facto revela que quando são excluídos, do processo de transformação do MES no MUP, os *pixels* aos quais foram atribuídas classes de elementos de superfície com graus de probabilidade mais baixos e, consequentemente, com maior incerteza, as unidades de paisagem são melhor identificadas. Por exemplo, ao excluir os elementos de superfície Copa de Eucalipto (CE) e Copa de Pinheiro Manso (CPM) classificados com maior incerteza, as unidades de paisagem Floresta de Eucalipto (FE) e Floresta de Pinheiro Manso (FPM) são melhor identificadas.

Na Figura 3.34, apresentam-se alguns resultados exemplificativos das regras aplicadas aos elementos de superfícies contidos dentro das regiões definidas pela segmentação. A Figura 3.35 mostra um exemplo de uma unidade de paisagem que só foi correctamente classificada como agricultura quando a incerteza da classificação dos elementos de superfície foi considerada.

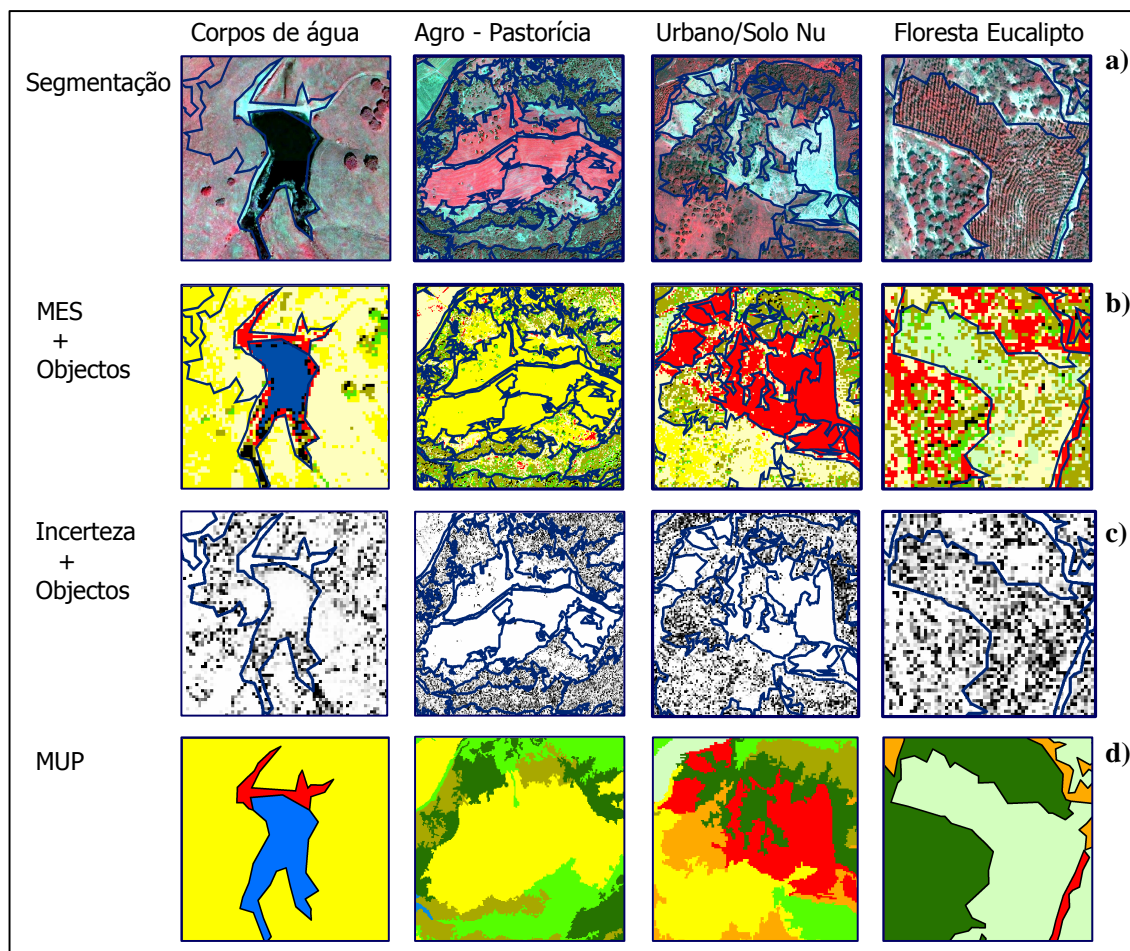


Figura 3.34 – Regras de classificação aplicadas: a) imagem IKONOS (RGB 432) segmentada; b) objectos sobrepostos ao MES obtido com o classificador probabilístico (MV_B); c) objectos sobrepostos à informação sobre a incerteza; d) MUP obtido com a aplicação das regras.

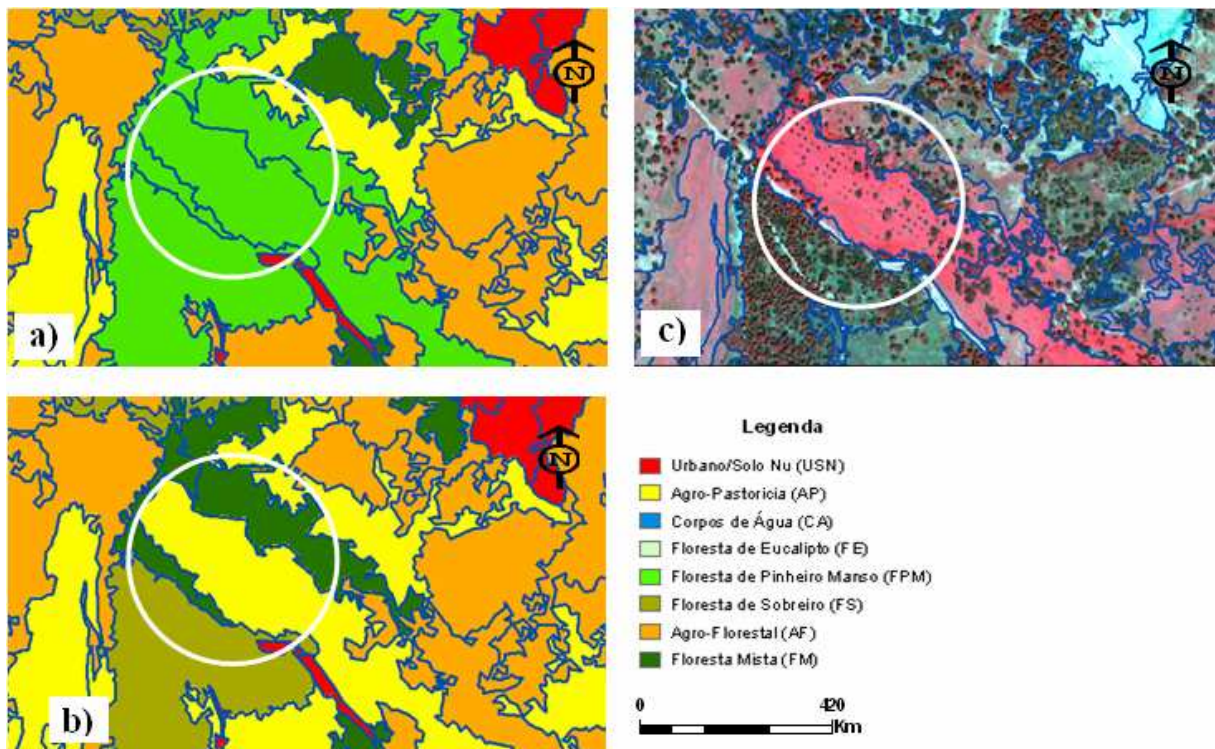


Figura 3.35 – MUP obtido com ambas as metodologias: a) classificação sem integração da incerteza; b) classificação com integração da incerteza; c) extracto da imagem IKONOS (RGB 432). Os círculos a branco marcam unidades de paisagem ocupadas por agricultura que foram correctamente classificadas quando a incerteza foi considerada nas regras de classificação.

Este novo método de classificação demonstra que a informação sobre a incerteza da classificação permite a identificação dos elementos de superfície em que ocorreu maior dificuldade em atribuir uma solução única, possibilitando assim que se exclua o seu uso no processo de transformação do MES no MUP. Com a aplicação deste método, obtém-se um mapa (Figura 3.36) em que cada unidade geométrica (polígono) se encontra classificada com uma só unidade de paisagem.

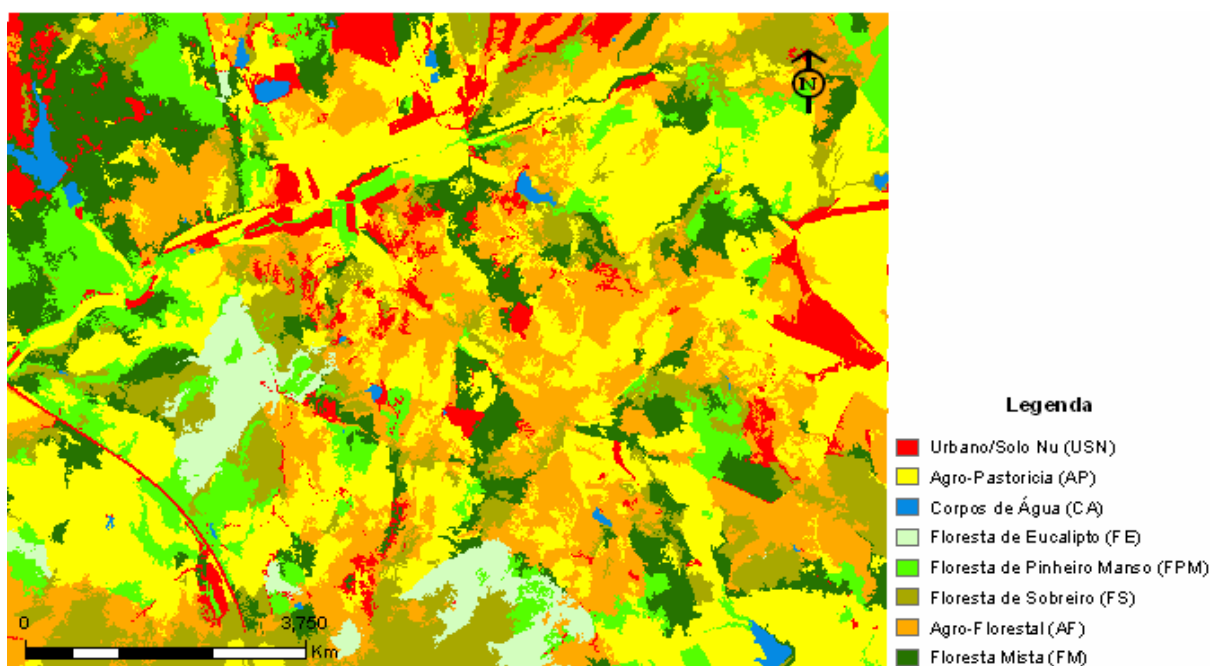


Figura 3.36 – Mapa de unidades de paisagem final.

3.6 Conclusões

Os objectivos do estudo descrito neste capítulo eram os seguintes: (1) investigar a influência e a utilidade que a informação da incerteza associada à classificação dos elementos de superfície tem, se integrada no processo de classificação das Unidades de Paisagem (UP); e (2) estudar o potencial da aplicação das medidas de incerteza na avaliação do comportamento de dois classificadores não rígidos e na avaliação da classificação temática produzida.

Os resultados obtidos demonstraram que a utilização da informação de incerteza é importante no processo de classificação combinada *pixel/objecto*. O valor da probabilidade global do MUP produzido com o método de classificação combinada *pixel/objecto* com integração de incerteza, melhorou 12% comparativamente ao obtido com a classificação utilizando um método semelhante mas sem a integração da incerteza. A determinação da incerteza permitiu identificar os elementos de superfície pior classificados e evitar o seu uso na identificação das UP, contribuindo para uma melhoria significativa dos resultados. Do ponto de vista metodológico, a abordagem combinada *pixel/objecto* com integração da incerteza demonstrou ser adequada à transformação do Mapa de Elementos de Superfície num Mapa de Unidades de Paisagem em formato vectorial para integração num Sistema de Informação Geográfica.

Para determinar se as medidas de incerteza podiam ser utilizadas como indicadores para avaliar o comportamento dos classificadores e a exactidão da classificação, foi efectuada uma comparação entre os resultados dados pelas medidas de incerteza e a informação obtida a partir das matrizes de confusão. Os resultados obtidos mostram que a informação sobre a incerteza da classificação de regiões consideradas representativas das várias classes pode ser usada, conjuntamente com os graus de probabilidade ou possibilidade, para detectar os principais problemas dos classificadores. Com os índices de exactidão e a informação sobre a incerteza foram obtidas as mesmas conclusões sobre o comportamento dos dois classificadores. As análises mostraram que existe uma correlação positiva entre a informação obtida pelas medidas de incerteza e os índices de exactidão estatísticos, principalmente com o índice de exactidão do utilizador. Por estas razões, pode concluir-se que a informação sobre a incerteza pode ser utilizada para avaliar o comportamento dos classificadores e, conseqüentemente, como um indicador da exactidão da classificação mas não substitui os índices de exactidão estatísticos na avaliação da exactidão da classificação.

O estudo mostrou que a aplicação das medidas de incerteza a distribuições de probabilidade ou possibilidade apresenta um conjunto de vantagens importantes nomeadamente: (1) é uma abordagem simples; (2) permite uma mais rápida e económica avaliação dos classificadores do que os índices de exactidão; (3) não são influenciadas pela subjectividade que pode existir nos dados de referência; (4) podem ser aplicadas a todo o mapa, o que permite uma representação espacial da incerteza. Outra importante vantagem destas medidas é que constituem uma ferramenta extremamente útil a utilizar numa fase preliminar, antes de se avançar para a recolha da informação de referência e para construção das matrizes de erro (que é um processo moroso e caro). Por exemplo, ao permitir identificar regiões e/ou classes com elevado nível de incerteza, pode auxiliar na identificação de problemas que podem conduzir à implementação de medidas que reduzam a incerteza da classificação, com, conseqüentemente, melhoramento da exactidão.

Capítulo 4

Utilização de imagens multiespectrais na avaliação do estado de conservação das coberturas da Baixa de Coimbra

4.1 Introdução

Para se definirem estratégias de intervenção no espaço construído, é necessário proceder a um levantamento prévio das características arquitectónicas, construtivas e estruturais dos edifícios que fazem parte da zona em estudo, incluindo o seu estado de conservação (Hemmler *et al.*, 2005; Vicente, 2008). Esta fase de levantamento e recolha de informação desempenha um papel muito importante na definição das operações de conservação e reabilitação a efectuar (Vicente, 2008). As técnicas tradicionais usadas para obter a informação e para identificar as anomalias (estruturais e não estruturais) são *low-tech*, requerem uma análise individual e *in situ* de cada edifício, o que as torna trabalhosas e morosas (Hemmler *et al.*, 2005). De referir que a geometria dos edifícios é habitualmente obtida por recurso a medições à fita, combinadas com levantamentos topográficos. Acresce ainda o facto de algumas partes dos edifícios serem, por vezes, de difícil acesso, nomeadamente as coberturas, o que dificulta o seu processo de inspecção e diagnóstico (Figura 4.1).



Figura 4.1 – Imagem de coberturas no centro histórico da Baixa de Coimbra.

Neste contexto, atendendo ao aumento crescente de acções de reabilitação urbana, é importante explorar outras fontes de informação e desenvolver novos métodos que permitam identificar os materiais construtivos e as anomalias existentes (ausência de material, fissuração, degradação química, colonização biológica, entre outras) e que possam ser vias alternativas, ou complementares, aos métodos manuais tradicionais sem apresentarem as desvantagens acima referidas. Estudos experimentais efectuados em paredes de alvenaria de monumentos com revestimentos de pedra (Inkpen *et al.*, 2001; Lerma, 2001; Ruiz *et al.*, 2002; Lerma, 2005) demonstraram que é possível obter informação fiável sobre os tipos de materiais e detectar anomalias nas fachadas, através da classificação automática de imagens fotográficas multiespectrais, pelo que a utilização desta fonte de dados na área da reabilitação é bastante promissora.

Lerma (2001) investigou a viabilidade de extrair informação, de forma automática, relativa à degradação do revestimento de duas fachadas de um monumento em que o material de revestimento predominante era pedra calcária. Numa das fachadas, utilizou apenas as bandas do visível e, noutra, utilizou também a do infravermelho próximo. A informação foi classificada com o classificador da máxima verosimilhança. Segundo este autor, a utilização da banda do infravermelho é essencial para estudos deste tipo, tendo permitido um aumento de 12% para o valor da Exactidão Global. Embora os resultados finais tenham sido bastante satisfatórios, um aspecto que é salientado pelo autor prende-se com a dificuldade em distinguir as áreas de sombra das áreas escuras provocadas pela poluição. Num estudo posterior (Lerma, 2005) a investigação foi alargada a fachadas de monumentos cujo revestimento era constituído por diversos materiais (e.g. pedra, alvenaria, tijolo, argamassas), tendo sido utilizadas quatro bandas multiespectrais (três bandas do visível e a banda do infravermelho próximo). Neste estudo foi também comparado o comportamento do classificador da máxima verosimilhança com o classificador da distância mínima à média. A avaliação da exactidão dos resultados obtidos com o classificador da máxima verosimilhança (88%) foi superior à obtida com o classificador da distância mínima à média (67%).

Ruiz *et al.* (2002) desenvolveram uma abordagem para cartografar os materiais e anomalias das fachadas no interior de edifícios históricos, como meio de suporte à planificação de trabalhos de conservação. Para a sua realização foram utilizadas bandas do visível e infravermelho próximo, informação de textura (variação dos valores dos níveis radiométricos dos *pixels*, medida numa dada vizinhança local) e uma combinação da informação espectral com a informação de textura. Os melhores resultados da classificação foram obtidos usando a informação espectral combinada

com a informação de textura. Um dos aspectos evidenciados neste estudo prende-se com a influência que a iluminação das fachadas tem neste tipo de aplicações. Para atenuar o seu efeito, as fachadas foram iluminadas usando luz difusa.

Os resultados promissores que foram obtidos com a utilização de imagens multiespectrais, obtidas automaticamente através de fotogrametria terrestre, sobre o tipo de materiais e anomalias do revestimento das fachadas dos edifícios, leva a considerar que podem ser utilizadas imagens multiespectrais aéreas ou de satélite para obter informação semelhante relativa às coberturas dos edifícios. Existem estudos sobre a utilização de IMGRE de zonas urbanas para identificar o tipo de materiais das coberturas, como por exemplo coberturas de telha vermelha (e.g. Hofmann, 2001; Fonseca, 2004), mas não para identificar as anomalias dos revestimentos deste tipo de coberturas.

No estudo que se apresenta neste capítulo, avaliou-se o potencial das imagens aéreas multiespectrais, comparando-o com o das imagens de satélite de muito grande resolução espacial, para extrair informação sobre o tipo de materiais de revestimento das coberturas e sobre as anomalias que estas apresentam. Para avaliar o potencial dos dois tipos de imagens, o estudo foi aplicado a cerca de 800 coberturas de edifícios situados no centro histórico da Baixa de Coimbra. Foram utilizadas imagens do satélite QuickBird, com uma resolução de 2.4 m, e imagens aéreas digitais, com uma resolução espacial de 0.5 m, em quatro bandas (azul, verde, vermelho e infravermelho próximo) com características espectrais comparáveis às da imagem QuickBird.

Apesar das imagens multiespectrais de grande resolução apresentarem um enorme potencial, conforme já referido no capítulo 3, permitindo identificar elementos de superfície mais pequenos do que as imagens de média e baixa resolução, também apresentam algumas limitações como, por exemplo, a grande variabilidade espectral, assim como a grande quantidade de sombras e o facto de o arranjo espacial dos *pixels* assumir uma elevada importância no processo de classificação. Além disso, as imagens multiespectrais aéreas ou de satélite levantam problemas adicionais devido, por exemplo, às distorções atmosféricas quando comparadas com as imagens fotográficas terrestres. A utilização desta fonte de dados em novas aplicações coloca assim desafios, exigindo o desenvolvimento de novas metodologias e a integração de diferentes métodos de extracção da informação de forma automática e que permitam também uma fácil integração da informação em Sistemas de Informação Geográfica (SIG). Existem várias técnicas de extracção de informação que vão desde a classificação ao nível do *pixel*, ao nível do objecto

(e.g. Blaschke *et al.*, 2004; Gonçalves e Caetano, 2004) ou uma combinação das duas (e.g. Wang *et al.*, 2004; Plantier e Caetano, 2007). Esta última é também referida por alguns autores como classificação ao nível da parcela, quando os objectos utilizados não são obtidos a partir de um processo de segmentação automática mas de informação vectorial digital existente em formato CAD (*Computer Aided Design*) ou num SIG (e.g. Aplin *et al.*, 1999).

No presente estudo, é aplicada uma abordagem de classificação semelhante à exposta no Capítulo 3, que combina a classificação ao nível do *pixel* com a classificação ao nível do objecto, integrando a incerteza associada à classificação ao nível do *pixel*, com o objectivo de obter um mapa das anomalias das coberturas da Baixa de Coimbra (centro histórico, extra-muros, da cidade). Esta abordagem apresenta também algumas semelhanças com a abordagem ao nível da parcela, pois os objectos são os contornos dos edifícios obtidos a partir da cartografia digital à escala 1:1 000. A obtenção do mapa, acima referido, é feita através do desenvolvimento de um conjunto de regras que consideram: (1) as diferentes proporções dos *pixels* classificados como tendo anomalias, (2) o tipo de revestimento, (3) o valor dos graus de probabilidade dos *pixels* classificados e (4) a quantificação da informação acerca da incerteza, associada à classificação de cada *pixel*, obtida com a medida de incerteza disponível no *software* IDRISI designada, nesta tese, por *Ratio* de Incerteza (RI). Os resultados obtidos com a classificação automática das imagens multiespectrais são posteriormente comparados com os dados resultantes do levantamento efectuado pela Universidade de Coimbra.

4.2 Área de estudo e dados

A área de estudo situa-se na Baixa de Coimbra, tem uma área aproximada de 15 ha e estende-se ao longo de uma faixa compreendida entre a colina e o rio (Figura 4.2). Inclui a freguesia de S. Bartolomeu e parte da freguesia de Santa Cruz, sendo delimitada a nascente pela muralha da cidade, a sul pelo Largo da Portagem, a poente pelas avenidas Emídio Navarro e Fernão de Magalhães e a norte pela Rua do Carmo e Largo do Arnado. Envolve um universo total estimado de cerca de 842 famílias residentes (1979 indivíduos), 1280 alojamentos (481 vagos) e cerca de 800 edifícios, maioritariamente antigos, alguns dos quais remontam ao séc. XVI.

Para a realização do estudo foram utilizadas: (1) imagens aéreas multiespectrais, com uma resolução espacial de 0.5 m e uma resolução radiométrica de 8 bits; e (2) imagens multiespectrais, obtidas pelo satélite QuickBird, com uma resolução espectral de 2.4 m no modo

(XS). As imagens aéreas pertencem ao conjunto de imagens multiespectrais digitais da Direcção Geral dos Recursos Florestais (DGRF), adquiridas com as 4 bandas multiespectrais, que cobrem Portugal Continental, obtidas por avião entre 2005 e 2006, com o sensor UltraCamTM da Vexcel e orto-rectificadas pelo Instituto Geográfico Português (IGP).

As imagens QuickBird foram adquiridas em Dezembro de 2004 e a sua correcção geométrica foi efectuada com recurso à ortorrectificação, tendo-se obtido um erro médio quadrático inferior a metade do *pixel*. Como o estudo não inclui uma componente multitemporal e, dada a pequena dimensão da área de estudo e homogeneidade da atmosfera, não foi aplicado qualquer tipo de correcção radiométrica. No estudo foram utilizadas as quatro bandas multiespectrais (banda do azul, verde, vermelho e infra-vermelho próximo).

Para realçar a informação que se pretende extrair das imagens, foi utilizado um índice de vegetação, *Normalized Difference Vegetation Index* (NDVI), como banda adicional. Este índice resulta do *ratio* entre a diferença e a soma das bandas do vermelho e infra-vermelho próximo:

$$NDVI = (\rho_{nir} - \rho_{red}) / (\rho_{nir} + \rho_{red}) \quad (4.1)$$

onde ρ_{nir} e ρ_{red} representam a reflectância das bandas do infravermelho próximo e do vermelho.

Para além das imagens foi também utilizada informação cartográfica à escala 1:1 000, em formato vectorial, a partir da qual foram extraídos os polígonos dos edifícios (Figura 4.3). A informação correspondente ao edificado, depois de tratada, foi importada para o ambiente SIG. As linhas de acção empreendidas na concepção do projecto SIG, que incidiu sobre a área de estudo, foram as seguintes: (1) criação do projecto e dos respectivos ficheiros; (2) estruturação de conteúdos; e (3) criação da base de dados geográfica. A base de dados geográfica desenvolvida contém a informação, disponibilizada pela Universidade de Coimbra, sobre as anomalias e tipo de materiais das coberturas do edificado resultante das acções de campo efectuadas. Na tese, esta informação é designada por informação de referência.

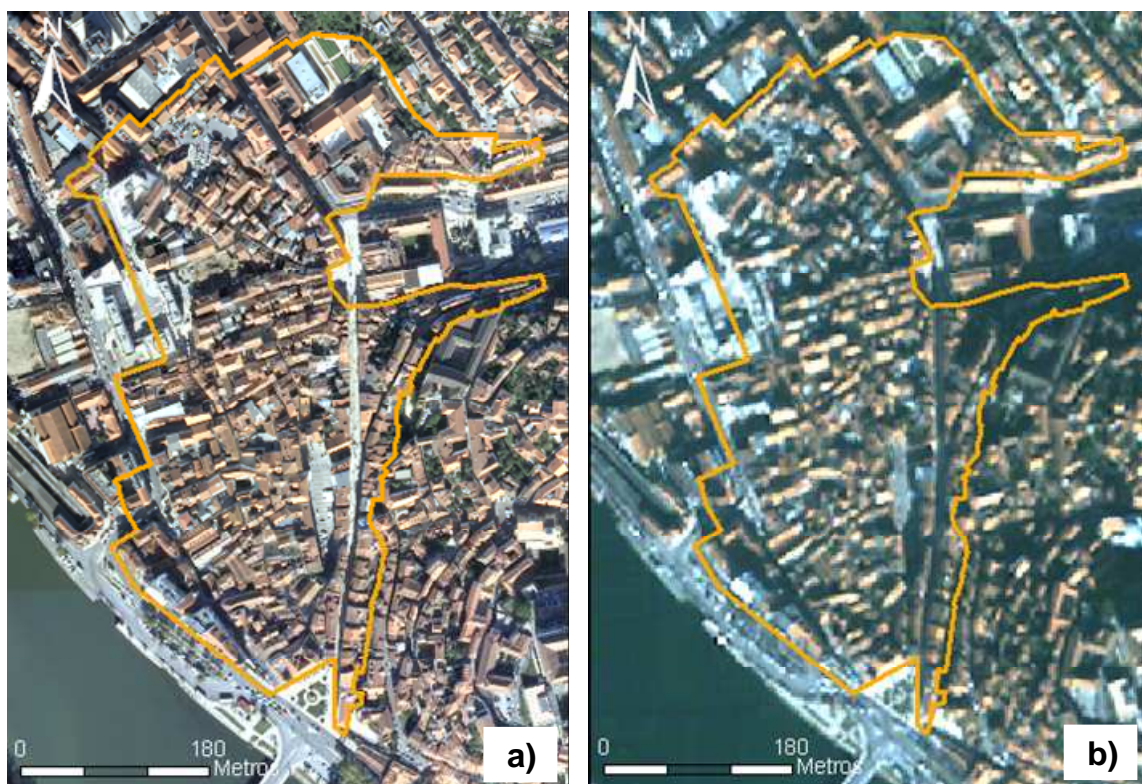


Figura 4.2 – Limite da área de estudo sobre: a) imagem aérea (RGB 321); b) imagem QuickBird (RGB 321).



Figura 4.3 – Extracto da imagem aérea (RGB 321) com a sobreposição da informação cartográfica dos edifícios.

4.3 Metodologia

Antes da definição e implementação da metodologia, procedeu-se a uma análise das imagens aéreas e de satélite para avaliar se é possível, por foto-interpretção, identificar os diferentes tipos de revestimentos das coberturas, com e sem anomalias, e os diferentes tipos de anomalias existentes. A partir da análise visual concluiu-se que, apesar da grande resolução espacial, estas imagens apenas permitiam a identificação das coberturas, com e sem anomalia, não possibilitando a diferenciação dos seus diferentes tipos. Além disso, a maioria dos telhados apresenta vários tipos de anomalias, o que dificulta a definição das áreas de treino (ou de referência), usadas para treino do classificador e validação da classificação, mesmo com o *pixel* com uma resolução espacial de 0.5 m. Após esta análise prévia, procedeu-se à definição do método, à definição das classes e à identificação das áreas de treino e referência.

Na área de estudo, os tipos de revestimentos mais frequentes são: telha cerâmica, fibrocimento e chapa metálica. Os resultados obtidos no levantamento efectuado pela Universidade de Coimbra revelam que, dos 681 edifícios estudados (representando cerca de 82% dos edifícios da área de estudo), 90% possui um revestimento da cobertura em telha cerâmica. Por esta razão, optou-se por direccionar o estudo para a identificação de anomalias das coberturas com este tipo de revestimento. A metodologia de classificação foi desenvolvida de forma a obter um Mapa de Anomalias das Coberturas com revestimento de telha cerâmica (MAC) e o estado de conservação em função da proporção de anomalias. O método de classificação consiste numa abordagem combinada, incluindo duas classificações da imagem ao nível do *pixel* utilizando um classificador não rígido, a incerteza das duas classificações ao nível do *pixel* e uma classificação ao nível do edifício.

Numa primeira fase, foram efectuadas as duas classificações assistidas ao nível do *pixel*. Uma, para identificar os elementos de superfície correspondentes aos diferentes tipos de materiais utilizados nos revestimentos das coberturas como, por exemplo, telha cerâmica e fibrocimento, obtendo-se um Mapa de Elementos de Superfície dos Materiais (MES-M) e, outra, para identificar elementos de superfície correspondentes a telha cerâmica com e sem anomalias, obtendo-se um Mapa de Elementos de Superfície de telha cerâmica com anomalias (MES-A). A opção de utilizar um classificador não rígido deveu-se ao facto, referido no Capítulo 3, deste tipo de classificadores atribuir a cada *pixel* diferentes graus de probabilidade ou possibilidade às várias classes consideradas, permitindo o cálculo da incerteza da classificação.

Como a unidade básica de representação da informação de referência adoptada no estudo é o edifício, era fundamental que, com esta nova abordagem de classificação, a unidade básica de representação fosse igualmente ao nível do edifício, para ser possível uma comparação dos resultados obtidos com as duas abordagens. A fase seguinte da classificação consistiu em obter o MAC, em que a unidade de representação espacial passa a ser o edifício em vez do *pixel*. A informação dos edifícios foi obtida a partir da cartografia em formato vectorial à escala 1:1 000, convertida para o formato raster, obtendo-se um mapa de edifícios em formato raster. Este mapa serviu, posteriormente, como máscara para que apenas fosse utilizada a informação do MES-M e do MES-A e da respectiva incerteza, dentro dos edifícios, para obtenção do mapa final. A identificação de anomalias na cobertura de cada edifício foi efectuada considerando um conjunto de regras que incluem: (1) o arranjo dos elementos de superfície, obtidos com a classificação ao nível do *pixel*, dentro de cada edifício; e (2) o grau de incerteza associada a essa atribuição.

A introdução da informação da incerteza associada à classificação dos elementos de superfície dos diferentes tipos de materiais e dos elementos de superfície com e sem anomalias, foi efectuada para evitar o uso de *pixels* mal classificados no processo de transformação dos MES-M e MES-A no MAC porque, conforme se demonstrou no capítulo 3, este tipo de abordagem permite melhorar a exactidão da classificação.

Em síntese, a metodologia aplicada à imagem aérea e do satélite QuickBird (Figura 4.4), consiste nos seguintes passos: (1) definição das classes de informação; (2) análise da separabilidade espectral das classes; (3) classificação não rígida ao nível do *pixel* para obtenção do MES-M e do MES-A; (4) avaliação da exactidão das classificações; (5) avaliação da incerteza de ambas; (6) conversão da informação dos edifícios para o formato matricial; (7) classificação das coberturas com telha cerâmica, com e sem anomalias, efectuada com base em regras de decisão; (8) avaliação da exactidão da classificação; e (9) comparação dos resultados obtidos com a informação recolhida pela Universidade de Coimbra.

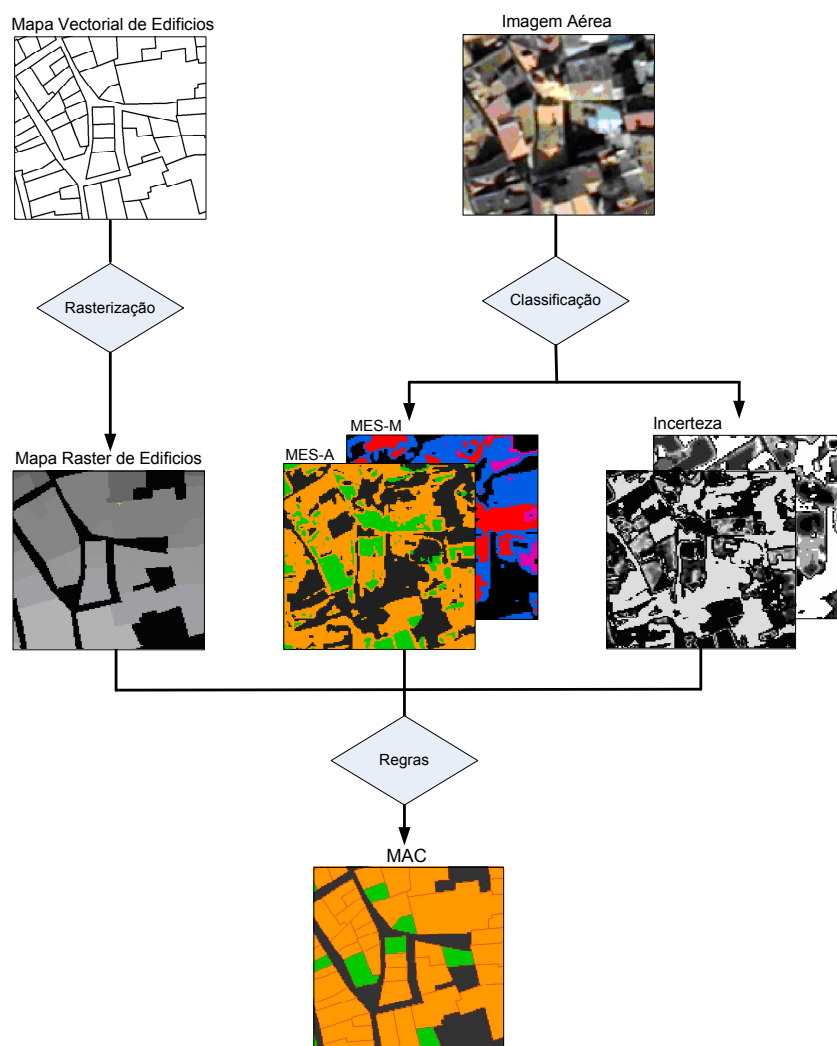


Figura 4.4 – Esquema da metodologia utilizada para a produção do MAC.

4.3.1 Protocolo de amostragem

Uma classificação assistida exige, à partida, a identificação de um conjunto de classes bem definidas. Este método envolve duas fases distintas: treino e classificação. Na fase de treino, consideram-se amostras representativas das várias classes que se pretendem identificar e efectua-se uma análise estatística dos atributos espectrais que melhor descrevem cada tipo de elementos de superfície a representar em cada classe. Esta fase exige a delimitação prévia, efectuada pelo operador, de áreas suficientemente representativas da variabilidade espectral de cada classe.

Para a fase de treino e validação da classificação das duas imagens, foram definidos três

protocolos de amostragem; dois para treino e avaliação do MES-M e do MES-A e outro para avaliação do MAC. A definição dos protocolos de amostragem utilizados para treino e avaliação do MES-M e do MES-A baseou-se na recolha das amostras representativas dos vários elementos de superfície, tendo sido utilizado o *pixel* como unidade amostra. Para cada classe, foram considerados cerca de 15 edifícios, dentro dos quais foi lançada uma amostra aleatória estratificada de 100 amostras para treino do classificador. Para avaliar a qualidade dos mapas obtidos (MES-M, MES-A, MAC), foi considerada uma amostra aleatória estratificada composta por 100 amostras por classe considerando todos os edifícios. Para a avaliação da exactidão, foi gerada uma matriz de erro, construída com base na verificação da classificação efectuada nos pontos seleccionados sobre o mapa temático pelo método de amostragem aleatória estratificada. Nesta fase, o levantamento efectuado pela Universidade de Coimbra foi utilizado como informação de referência.

Como a imagem QuickBird e a imagem aérea têm resoluções diferentes, houve por vezes necessidade de deslocar o *pixel* amostrado para o elemento de superfície mais próximo, devido à posição da amostra na Imagem Aérea não corresponder ao mesmo elemento de superfície (Figura 4.5).

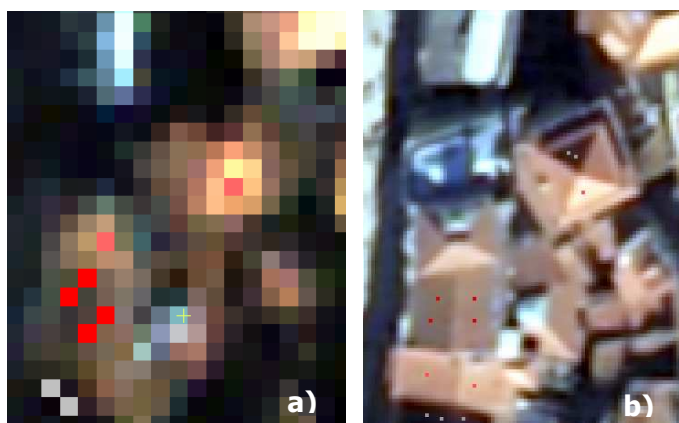


Figura 4.5 – Exemplo das amostras de treino em extractos da imagem aérea (RGB 321) (a) e da imagem QuickBird (RGB 321) (b); as diferentes cores das unidades amostra correspondem a diferentes classes.

4.3.2 Análise espectral

A informação das áreas de treino foi também usada para estudar a separabilidade das classes ao nível do *pixel*. Para analisar e comparar a separabilidade espectral entre as classes definidas em cada uma das imagens, recorreu-se a um diagrama de dispersão bidimensional e à medida de

separabilidade Distância de Battacharrya (B-Distance). Para eixos de visualização, escolheram-se as bandas do vermelho e infra-vermelho próximo, atendendo a que são as que, devido à presença de vegetação pioneira e musgos nalguns telhados, permitem uma melhor diferenciação das classes.

A B-Distance foi adoptada como medida de separabilidade por ter uma relação muito próxima com o resultado da classificação, de acordo com vários estudos em que foi utilizada (Herold *et al.*, 2003; Wang *et al.*, 2004). Este indicador apresenta valores no intervalo $[0, 2]$, onde 0 indica completa sobreposição, valores superiores a 1.9 significam boa separabilidade e 2 representa total separabilidade.

4.3.3 Classificação

O método de classificação desenvolvido no estudo tem várias etapas. Numa primeira fase, foram efectuadas duas classificações assistidas não rígidas, ao nível do *pixel*, para obtenção do MES-M e do MES-A, tendo-se aplicado o classificador probabilístico da máxima verosimilhança e obtido, *a posteriori*, os graus de probabilidade da pertença de cada *pixel* às várias classes, utilizando o teorema de Bayes (Wang, 1990; Foody *et al.*, 1992; Foody, 1996; Foody, 2004). No resultado final, obtém-se uma imagem por classe onde o valor associado a cada *pixel* expressa a probabilidade, *a posteriori*, do *pixel* pertencer a essa classe. Na classificação dos MES-M, foram consideradas as quatro bandas multiespectrais e, na classificação do MES-A, foi ainda considerado o índice de vegetação como banda adicional. A escolha deste classificador deve-se ao facto de permitir obter, de acordo com outros estudos, bons resultados nas classificações e de estar disponível na maior parte do *software* de processamento e classificação de imagens (Eastman e Laney, 2002).

As classes usadas no estudo para o MES-M foram: Telha Cerâmica Escura (TC_E); Telha Cerâmica Clara (TC_C); Fibrocimento (F); Chapa Metálica (CM) e Sombra (S). Para o MES-A foram: Telha Cerâmica com Anomalia (TC-A); Telha Cerâmica Não Degradada (TC-ND) e Telha Cerâmica Sombra (TC-S). As classes Sombra (S) e Telha Cerâmica Sombra (TC-S) foram consideradas devido a uma significativa quantidade de sombras que as imagens multiespectrais apresentavam. Numa primeira fase foi efectuada uma classificação preliminar sem considerar estas duas classes e, com base numa análise visual da incerteza da classificação, verificou-se que as zonas de maior incerteza coincidiam com zonas de sombra. O facto de se considerar

estas classes no processo de desenvolvimento das regras permitiu reduzir a incerteza e melhorar a classificação dos materiais e anomalias ao nível do edifício.

Para determinar a incerteza da classificação foi utilizada a medida RI, disponível no *software* IDRISI, dada por:

$$RI = 1 - \frac{\max_{i=1,\dots,n}(p_i) - \frac{\sum_{i=1}^n p_i}{n}}{1 - \frac{1}{n}} \quad (4.2)$$

onde p_i ($i=1,\dots,n$) representa a probabilidade *a posteriori* da pertença de cada *pixel* às várias classes e n o número de classes consideradas. Como a classificação foi efectuada com um classificador probabilístico, $\sum_{i=1}^n p_i = 1$, assumindo este indicador valores no intervalo $[0,1]$. A

razão da escolha desta medida deve-se ao facto de ter mostrado, no estudo apresentado no capítulo 3, que permite disponibilizar informação sobre a dificuldade do classificador em atribuir apenas uma classe a cada *pixel*. Quando aplicada a distribuições de probabilidade é uma medida que, embora seja sensível à dispersão, é mais sensível à compatibilidade, nomeadamente ao valor atribuído ao melhor grau de probabilidade. Mais informação sobre esta medida pode ser encontrada nas secções 2.3.3 e 2.4.2.

A fase seguinte consistiu na construção do MAC, através da combinação do MES-M, do MES-A, da incerteza da classificação e dos polígonos que representam os edifícios. Para tal, foi necessário desenvolver um conjunto de regras que incluem: (1) a informação dos *pixels* classificados dentro dos edifícios; e (2) o grau de incerteza associado a essa atribuição, obtido com a medida de incerteza, de forma a evitar que os *pixels* classificados com muita incerteza sejam utilizados e influenciem negativamente a classificação ao nível do edifício. A construção das regras requer uma análise preliminar das probabilidades atribuídas a cada *pixel* e do grau de incerteza da classificação para definir um limiar de decisão.

As classes usadas no MAC foram: "Edifícios com Cobertura de Telha Cerâmica com Anomalias" (E-C-A), podendo assumir vários estados de conservação, desde o "Estado de Conservação 1" (EC1), correspondendo ao pior estado de conservação, até ao "Estado de Conservação 4" (EC4), correspondendo ao melhor estado de conservação (com anomalias); e " Edifícios com Cobertura

de Telha Cerâmica Não Degradados” (E-C-ND), correspondendo ao “Estado de Conservação 5” (EC5), considerado o perfeito estado de conservação. Na Tabela 4.1 apresenta-se as regras de classificação e na Figura 4.6, o esquema de classificação hierárquico desenvolvido.

O processo de classificação é semelhante a uma árvore de decisão, *i.e.*, a uma estrutura hierárquica em que é aplicado um conjunto de regras em cada nível, permitindo agrupar os objectos em subconjuntos de indivíduos sucessivamente menos heterogéneos.

O objectivo das regras 1 a 3 é distinguir entre “Edifícios com Cobertura de Telha Cerâmica” (E-C) e “Edifícios sem Cobertura de Telha Cerâmica” (E-NC). As regras 4 a 6 distinguem os “Edifícios com Cobertura de Telha Cerâmica” (E-C) numa de duas classes: “Edifícios com Telha Cerâmica com Anomalia” (E-C-A) e “Edifícios com Telha Cerâmica Não Degradada” (E-C-ND), ao qual corresponde o “Estado de Conservação 5” (EC5). A regra 7 distingue “Edifícios com Telha Cerâmica com Anomalia” (E-C-A) numa de quatro classes: “Estado de Conservação 1” (EC1), “Estado de Conservação 2” (EC2), “Estado de Conservação 3” (EC3) e “Estado de Conservação 4” (EC4). As linhas de código em AML são apresentadas no Anexo C.

Tabela 4.1 – Regras de classificação aplicadas aos mapas MES-M e MES-A para obter o mapa MAC.

| Nº | Regras | Se Verdadeira | Se Falsa |
|----|---|-----------------------|----------------------|
| 1 | Se mais de 50% dos <i>pixels</i> do MES-M, dentro dos edifícios, foram classificados como “Telha Cerâmica” com incerteza inferior a 0.25. | Atribuir Classe E-C | Aplicar Regra 2 |
| 2 | Se a percentagem dos <i>pixels</i> do MES-M, dentro dos edifícios, classificados como “Telha Cerâmica” com uma incerteza inferior a 0.25, for superior à percentagem de <i>pixels</i> classificados quer como: “Fibrocimento”; “Chapa Metálica”; “Sombra”. | Atribuir classe E-C | Aplicar Regra 3 |
| 3 | Se a percentagem dos <i>pixels</i> do MES-M, dentro dos edifícios, classificados como “Telha Cerâmica” com uma incerteza inferior a 0.25, for superior à percentagem de <i>pixels</i> classificados como: “Fibrocimento” e “Chapa Metálica” e se a percentagem de <i>pixels</i> classificados como “Sombra” é superior a 50%. | Atribuir classe E-C | Atribuir classe E-NC |
| 4 | Se mais de 50% dos <i>pixels</i> do MES-A, dentro dos edifícios, classificados como “Edifícios com Cobertura de Telha Cerâmica”, forem “Telha Cerâmica com Anomalia” com uma incerteza inferior a 0.25. | Atribuir classe E-C-A | Aplicar Regra 5 |
| 5 | Se a percentagem dos <i>pixels</i> do MES-A, classificados como “Telha Cerâmica com Anomalia”, dentro dos edifícios classificados como “Edifícios com Cobertura de Telha Cerâmica”, for superior à percentagem de “Telha Cerâmica Não Degradada” e superior à percentagem “Telha Cerâmica Sombra”. | Atribuir classe E-C-A | Aplicar Regra 6 |

| | | | |
|---|--|-----------------------|---------------------------|
| 6 | Se a percentagem dos <i>pixels</i> do MES-A, classificados como “Telha Cerâmica com Anomalia, dentro dos edifícios classificados como “Edifícios com Cobertura de Telha Cerâmica”, for superior à percentagem de “Telha Cerâmica Não Degradada” e se a percentagem de Telha Cerâmica Sombra” for superior a 50%. | Atribuir classe E-C-A | Atribuir classe EC_ND=EC5 |
| 7 | Edifícios classificados como ‘Edifícios com cobertura de Telha Cerâmica com Anomalia’ têm mais de 75% dos <i>pixels</i> classificados como “Telha Cerâmica com Anomalias” com uma incerteza inferior a 0.25. | EC1 | |
| | Edifícios classificados como ‘Edifícios com cobertura de Telha Cerâmica com Anomalia’ têm entre 50% e 75% dos <i>pixels</i> classificados como “Telha Cerâmica com Anomalias” com uma incerteza inferior a 0.25. | EC2 | |
| | Edifícios classificados como ‘Edifícios com cobertura de Telha Cerâmica com Anomalia’ têm entre 25% e 50% dos <i>pixels</i> classificados como “Telha Cerâmica com Anomalias” com uma incerteza inferior a 0.25. | EC3 | |
| | Edifícios classificados como ‘Edifícios com cobertura de Telha Cerâmica com Anomalia’ têm entre 0 e 25% dos <i>pixels</i> classificados como “Telha Cerâmica com Anomalias” com uma incerteza inferior a 0.25. | EC4 | |

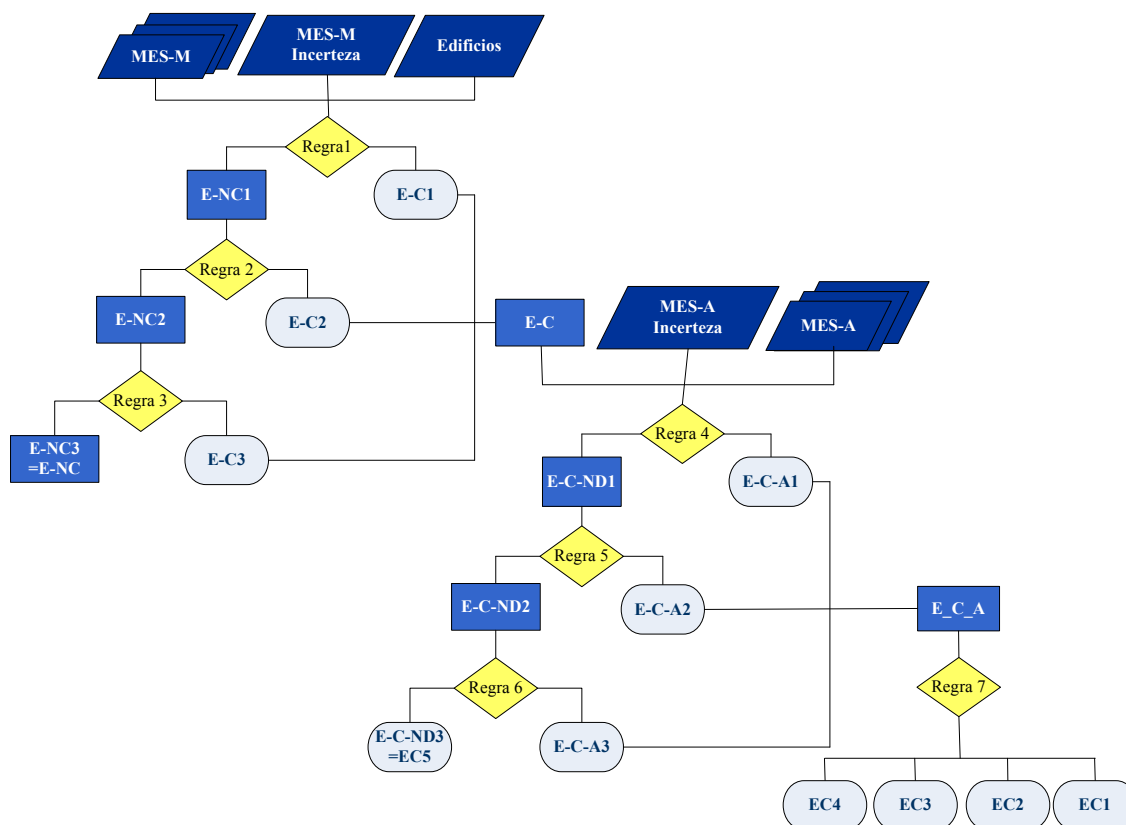


Figura 4.6 – Fluxograma de Classificação.

4.4 Resultados e discussão

4.4.1 Análise espectral

A Figura 4.7 e a Figura 4.8 mostram, para cada imagem, as elipses de distribuição das classes dos elementos de superfície das coberturas, correspondente aos diferentes materiais de revestimento, e das classes correspondentes aos elementos de superfícies das coberturas com e sem anomalias, respectivamente.

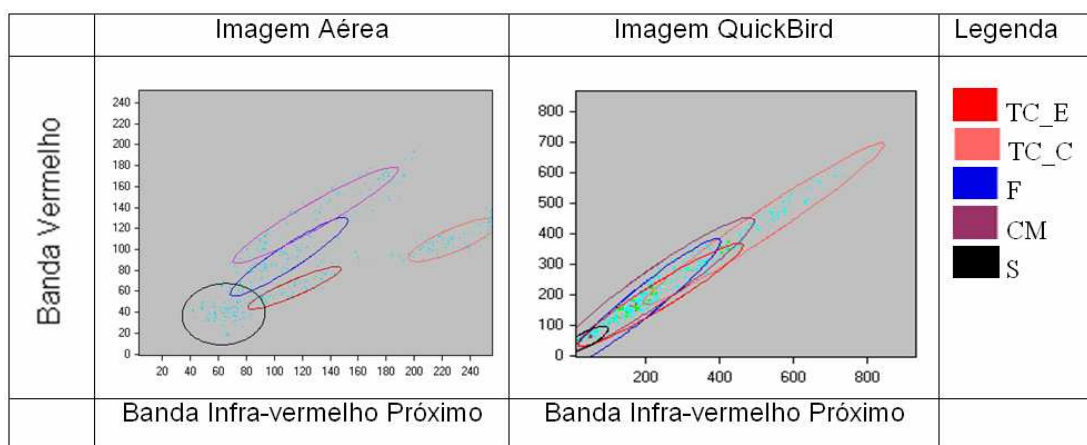


Figura 4.7 – Diagrama de dispersão da imagem aérea e QuickBird onde as elipses representam as classes dos elementos de superfície dos materiais: Telha Cerâmica Escura (TC_E); Telha Cerâmica Clara (TC_C); Fibrocimento (F); Chapa Metálica (CM) e Sombra (S).

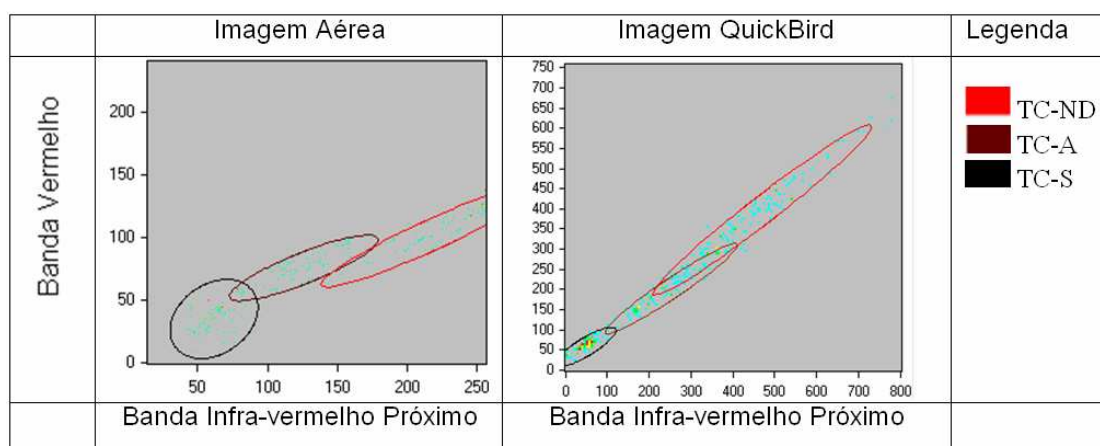


Figura 4.8 – Diagrama de dispersão das imagens aérea e QuickBird onde as elipses representam as classes dos elementos de superfície de telha cerâmica com anomalia (TC-A); Não Degradada (TC-ND) e Telha Cerâmica Sombra (TC-S).

Da análise comparativa dos diagramas de dispersão, pode ser observado que, para as bandas representadas, a sobreposição da assinatura espectral das classes diminui com o aumento da resolução espacial das imagens. Na imagem QuickBird, verifica-se uma quase total sobreposição das elipses correspondentes aos elementos de superfície dos materiais, o que significa que a proximidade das assinaturas espectrais destas classes é muito grande. Na imagem aérea, pelo contrário, as elipses correspondentes aos elementos de superfície dos materiais apresentam uma separação quase total, podendo ser observadas pequenas sobreposições entre a classe Sombra (S) e as classes Fibrocimento (F) e Telha Cerâmica Escura (TC_E).

Relativamente às elipses de distribuição das classes dos elementos de superfície das coberturas com e sem anomalias, verifica-se principalmente uma maior sobreposição, na imagem QuickBird, entre as assinaturas espectrais das classe Telha Cerâmica Não Deteriorada (TC-ND) e Telha Cerâmica com Anomalia (TC-A).

Os resultados quantitativos de separabilidade dados pela B-Distance reforçam os resultados obtidos com os diagramas de dispersão. A medida de separabilidade média dos elementos de superfície das coberturas correspondentes aos materiais é, na imagem aérea, de 1.87 e, na imagem QuickBird, de 1.72. Relativamente ao valor médio de separabilidade dos elementos de superfície das coberturas correspondentes às anomalias, na imagem aérea obteve-se um valor de 1.93 e na imagem QuickBird de 1.90. Estes resultados são um indicador de que existe uma maior dificuldade em separar as classes do MES-M do que em separar as classes do MES-A e que essa dificuldade é maior com a imagem do satélite QuickBird.

4.4.2 Classificação

Os resultados obtidos para a Exactidão Global (EG) e Coeficiente de Khat do MES-M e do MES-A, produzidos com as duas imagens, utilizando o mesmo método de classificação, apresentam-se na Tabela 4.2. Os valores dos índices de exactidão evidenciam que o classificador teve um bom comportamento quando aplicado às imagens multiespectrais de muito grande resolução e que a exactidão dos mapas obtidos a partir da imagem aérea é ligeiramente superior à obtida a partir da imagem satélite. Como a imagem aérea possui características espectrais comparáveis à imagem do satélite QuickBird, pode concluir-se que o aumento da exactidão se deve ao facto de a resolução espacial da imagem aérea ser superior. O que demonstra que, para este tipo de aplicações, um aumento da resolução espacial do *pixel* melhora a identificação e extracção deste

tipo de informação. Comparando os resultados obtidos nos diferentes mapas, constata-se ainda que o mapa de elementos de superfície das anomalias da telha cerâmica, para as duas imagens, apresenta valores de exactidão ligeiramente superiores aos do mapa dos revestimentos, o que é consistente com os resultados dados pela B-Distance. Na Figura 4.9 e na Figura 4.10, apresentam-se os resultados obtidos para a exactidão do Utilizador e do Produtor, para cada uma das classes.

Tabela 4.2 – Índices de Exactidão Global da classificação dos MES-M e MES-A.

| | Imagem Aérea | | Imagem QuickBird | |
|-------|--------------|------|------------------|------|
| | EG | Khat | EG | Khat |
| MES-M | 89% | 86% | 78% | 72% |
| MES-A | 94% | 92% | 89% | 84% |

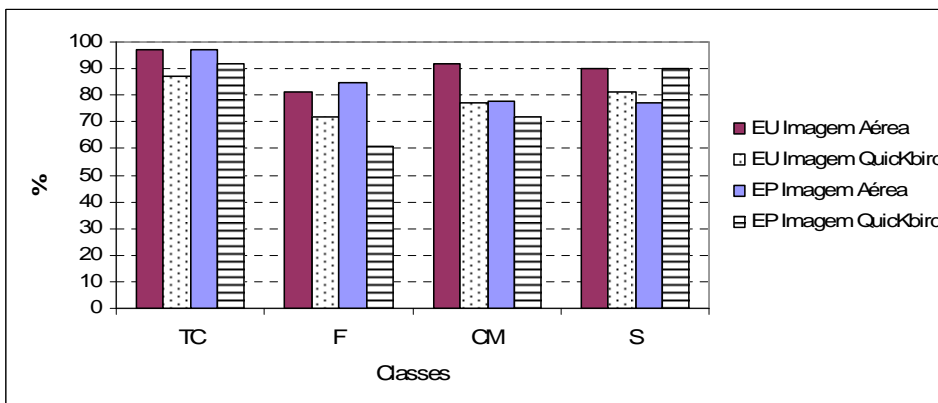


Figura 4.9 – Índices de Exactidão do Produtor (EP) e do Utilizador (EU) dos Mapas de Elementos de Superfície dos Materiais de revestimento (MES-M), produzidos com a imagem aérea e a QuickBird.

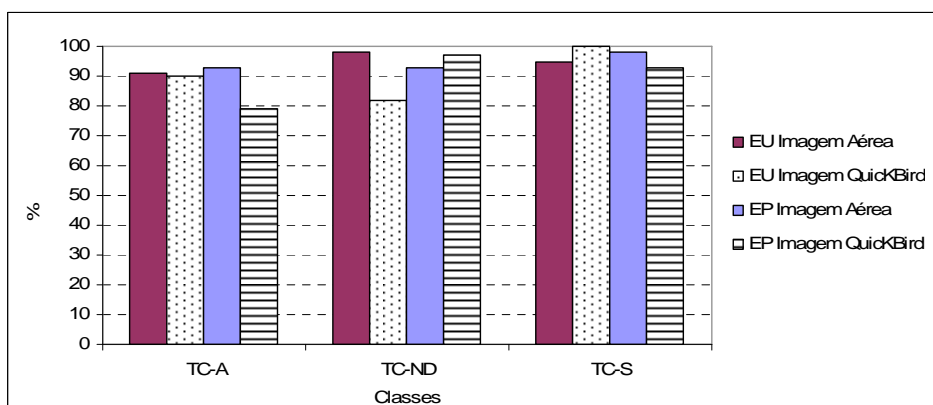


Figura 4.10 – Índices de Exactidão do Produtor (EP) e do Utilizador (EU) dos Mapas de Elementos de Superfície das Anomalias da telha cerâmica (MES-A) produzidos com a imagem aérea e a QuickBird.

Os resultados da exactidão do Produtor e Utilizador das classes do MES-M mostram que a classe

Telha Cerâmica (TC-C e TC-E) foi identificada com grande exactidão. Observa-se alguma confusão entre as classes Fibrocimento (F), Chapa Metálica (CM) e Sombra (S), devido às confusões espectrais entre estas classes, o que resulta em valores de exactidão inferiores.

Relativamente aos resultados das classes do MES-A, constata-se que a classe Telha Cerâmica com Anomalias (TC-A) apresenta, com a imagem aérea, valores superiores para a exactidão do Produtor face à do Utilizador, acontecendo o contrário com a classe Telha Cerâmica Não Deteriorada (TC-ND). Este resultado revela que os *pixels* que correspondem a Telha Cerâmica com Anomalia (TC-A) no terreno foram bem identificados (menos erros de omissão). Contudo, outros *pixels* que no terreno não apresentavam anomalia, foram igualmente classificados como pertencendo à classe Telha Cerâmica com Anomalias (TC-A) (erros de comissão). A classificação da imagem satélite apresenta um comportamento oposto, ou seja, a classe TC-A apresenta mais erros de omissão do que a classe TC-ND. Esta diferença de comportamento pode dever-se ao facto de a imagem QuickBird apresentar mais sombras do que a imagem aérea (Figura 4.2).

A Figura 4.11 e a Figura 4.12 mostram a incerteza média por classe do MES-M e do MES-A, respectivamente. A comparação dos resultados da incerteza com os da exactidão, apresentada na Figura 4.9 e na Figura 4.10, mostra que existe elevada consistência entre eles. Com efeito, as classes Telha Cerâmica e Sombra (S) apresentam os valores mais baixos de incerteza e os mais altos valores de exactidão, enquanto as outras duas classes registam os valores mais elevados de incerteza e os mais baixos de exactidão.

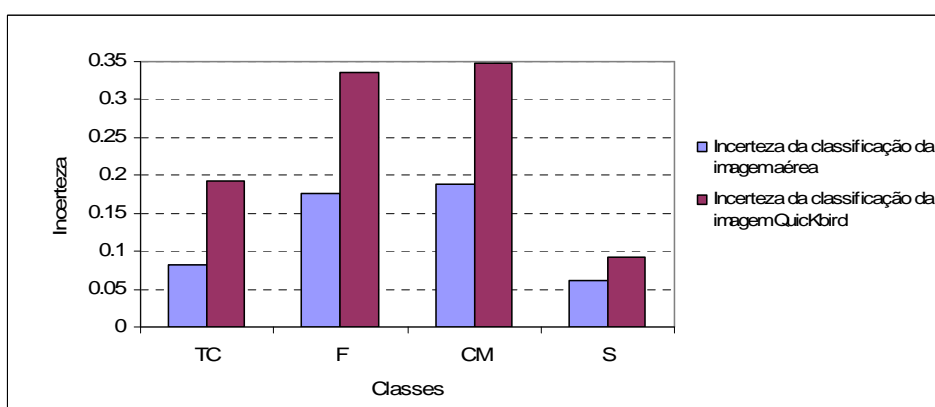


Figura 4.11 – Incerteza média por classe dos Mapas de Elementos de Superfície dos Materiais de revestimento (MES-M), produzidos com a imagem aérea e a QuickBird.

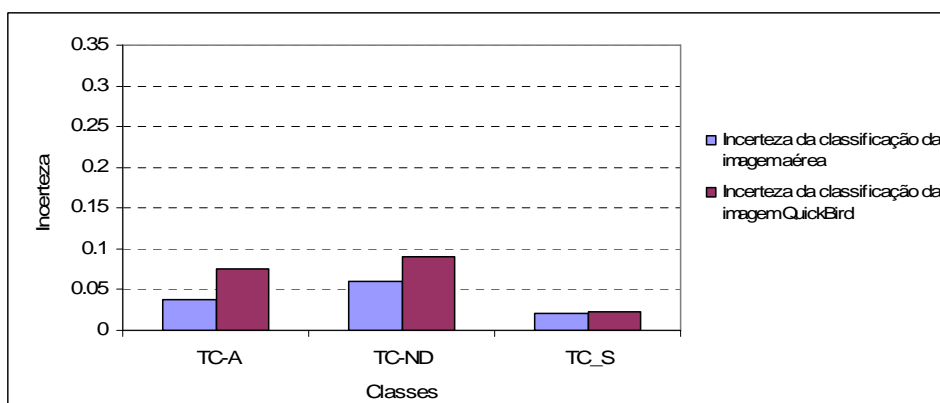


Figura 4.12 – Incerteza média por classe da classificação dos Mapas de Elementos de Superfície das Anomalias da telha cerâmica (MES-A) produzidos com a imagem aérea e a QuickBird.

O mapa final (MAC) apresentou para a EG um valor de 78%. Na Figura 4.13, apresentam-se os valores de Exactidão do Utilizador e do Produtor obtidos para o MAC.

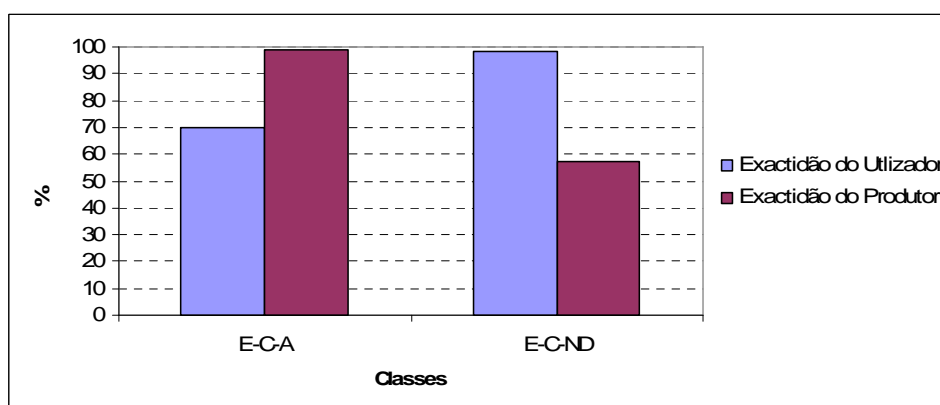


Figura 4.13 – Índices de Exactidão do Utilizador (EU) e do Produtor (EP) do mapa final (MAC) obtido com a abordagem combinada aplicada às imagens aéreas.

Os resultados mostram que a classe Edifícios com Cobertura de Telha Cerâmica com Anomalias (E-C-A), apresenta valores de exactidão do produtor bastante elevados e superiores aos valores obtidos para a exactidão do utilizador, o que significa que, embora as coberturas de telha cerâmica com anomalias existentes na área de estudo tenham sido bem identificadas (valor reduzido de erros de omissão), existem coberturas não degradadas segundo os dados de referência que também foram classificadas como tendo anomalias (erros de comissão).

4.4.3 Análise comparativa dos resultados: abordagem proposta versus métodos tradicionais

O estudo de referência analisou 826 edifícios dentro da área de estudo. Com o método tradicional de levantamento e recolha da informação, 619 edifícios (75%) foram identificados como tendo cobertura em telha cerâmica e, com a nova abordagem, utilizando imagens multiespectrais com uma resolução de 0.5 m, foram identificados 624 edifícios (76%), o que corresponde a um ligeiro aumento na identificação deste tipo de material. Uma análise comparativa dos resultados obtidos pelos dois métodos, efectuada através de uma operação de intersecção espacial dos dois mapas, utilizando ferramentas de SIG, mostrou que, dos edifícios identificados como possuindo cobertura em telha cerâmica na referência, 532 (85%) foram também identificados com esta nova abordagem.

Da análise comparativa entre a informação do MAC obtido de forma automática, com a abordagem desenvolvida, e o mapa de referência, utilizando ferramentas SIG de análise espacial, constatou-se que, das coberturas que possuíam revestimento de Telha Cerâmica com anomalias na referência, 77% também foram identificadas com a nova abordagem. Na Figura 4.14 apresenta-se o MAC obtido com a metodologia desenvolvida e o mapa obtido pela Universidade de Coimbra.

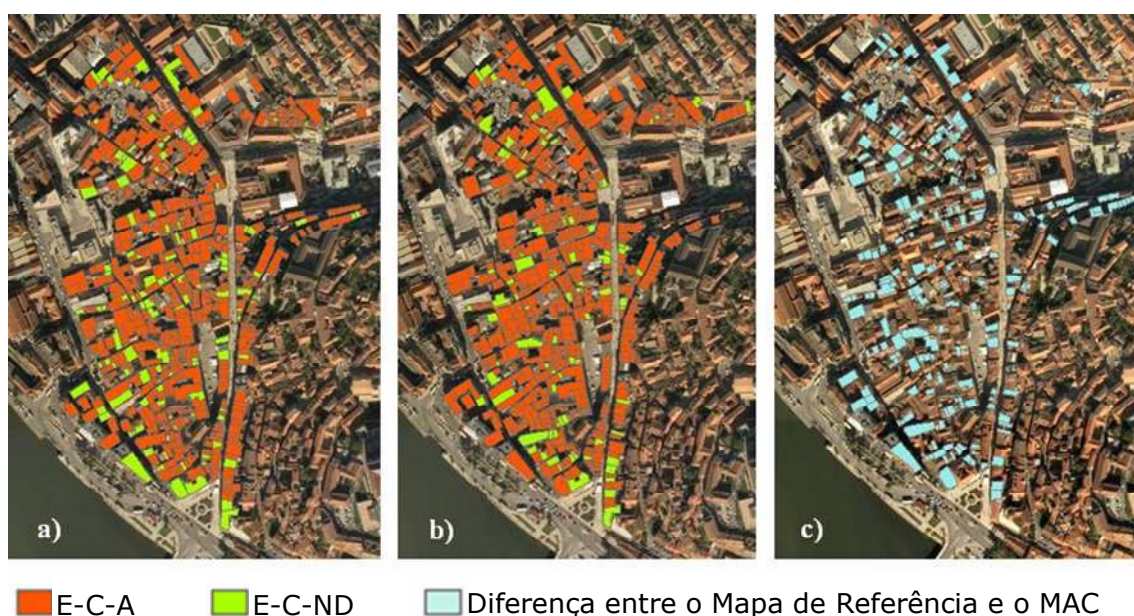


Figura 4.14 – Mapa de Anomalias das Coberturas dos edifícios com telha cerâmica (MAC): a) mapa obtido pela Universidade de Coimbra por métodos tradicionais; b) mapa obtido de forma automática com a abordagem desenvolvida; c) diferença entre os dois mapas.

A análise das diferenças existentes entre os dois mapas de anomalias, permitiu concluir que estas se devem, em parte, à presença significativa de zonas de sombra na imagem e ao facto de não existir um bom ajuste entre os polígonos dos edifícios extraídos da cartografia e as imagens (Figura 4.3). De referir ainda que, o facto de o levantamento por métodos tradicionais ter sido efectuado entre 2003 e 2004 e as imagens usadas na abordagem desenvolvida serem de 2005 e 2006, é outro dos factores que influênciam os resultados.

Como a assinatura espectral da classe Telha Cerâmica Sombra (TC-S) apresenta alguma sobreposição com a classe Telha Cerâmica com Anomalias (TC-A), os edifícios que apresentam elevada quantidade de sombras podem ser incorrectamente classificados como possuindo anomalias, facto que pode explicar os erros de comissão e omissão apresentados pelas classes TC-A e TC-ND (Figura 4.13). Os desfasamentos entre a imagem e a informação cartográfica levam a que alguns *pixels* situados no interior dos polígonos, supostamente correspondentes aos edifícios, efectivamente não lhes pertençam e, como consequência, geram conclusões erradas na aplicação das regras. Como existe uma discrepância temporal entre os dados usados em ambos os estudos e como ocorreram algumas alterações durante esse período nas coberturas dos edifícios, existem diferentes resultados obtidos que podem estar correctos nos dois estudos. Por exemplo, durante este período, algumas coberturas foram rehabilitadas, não apresentando nenhuma anomalia na imagem utilizada, o que foi correctamente identificado pelo processo automático, conforme se pode ver na (Figura 4.15). Por outro lado, alguns telhados que não apresentavam anomalia em 2003-2004 podem ter desenvolvido ou agravado algum tipo de anomalia desde esse período tornando-se visível em 2005-2006.

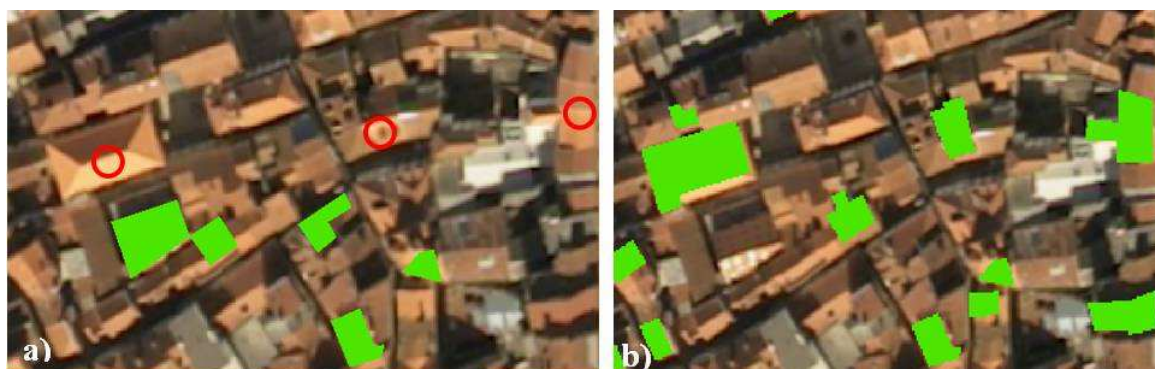


Figura 4.15 – a) Extracto da imagem aérea (RGB 321) com a sobreposição dos edifícios com coberturas não degradadas resultante do levantamento efectuado pela Universidade de Coimbra; b) Extracto da imagem aérea (RGB 321) com sobreposição dos edifícios com coberturas não degradadas resultante do processo de classificação automático. Os círculos assinalam alguns dos edifícios com coberturas rehabilitadas e que foram bem identificados pelo método automático de classificação.

A grande resolução espacial das imagens multiespectrais utilizadas no estudo, apesar de permitir a identificação dos materiais e a detecção de anomalias, também permite a identificação de outras características, como exemplo, regiões com telha cerâmica que pertencem à mesma cobertura mas que apresentam diferente brilho, ou sombra, devido à inclinação e diferente orientação relativamente ao sol das várias águas. Estes aspectos representam um problema significativo para este tipo de aplicações. Contudo, o método de classificação combinada proposto demonstrou ser adequado para ultrapassar este tipo de dificuldades dado que permitiu captar, a partir dos elementos de superfície classificados e do seu arranjo, a informação para classificar com resultados bastante bons, o tipo de revestimento e a existência de anomalias das coberturas, ao nível do edifício (Figura 4.16).

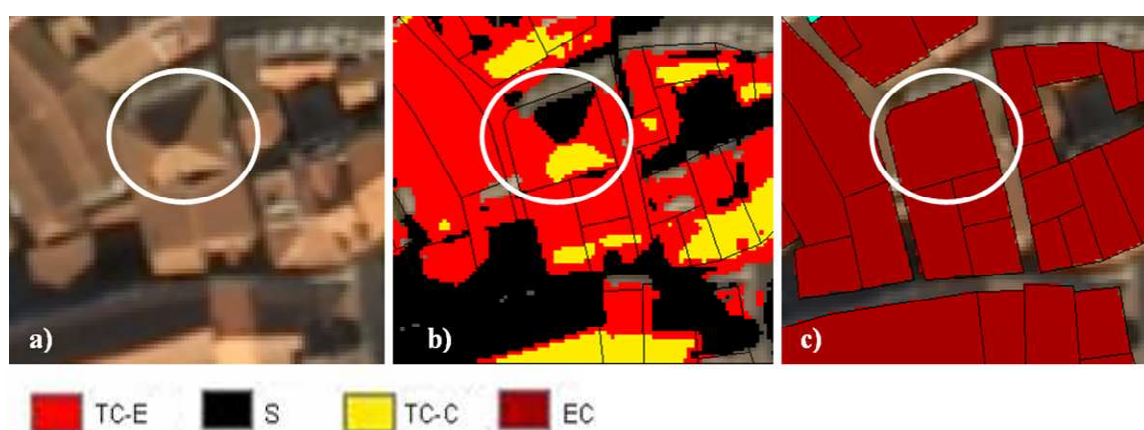


Figura 4.16 – Classificação das unidades de edifícios a partir dos elementos de superfície classificados. a) extracto da imagem aérea (RGB 321); b) classificação ao nível do pixel dos elementos de superfície Telha Cerâmica Escura (TC_E), Telha Cerâmica Clara (TC_C) e Sombra (S); c) classificação ao nível do edifício em Edifícios com Cobertura de Telha Cerâmica (E-C). Os círculos assinalam os edifícios que apresentam cobertura de telha cerâmica com sombra e com diferente brilho e que foram correctamente identificados com a abordagem de classificação desenvolvida.

4.5 Conclusões

Os resultados obtidos revelam que a utilização das imagens de muito grande resolução espacial (IMGRE) para extracção de informação de suporte à avaliação do estado de conservação das coberturas é bastante promissora. Com a imagem aérea, com 0.5 m de resolução espacial, foi possível detectar a presença de anomalias e identificar os diferentes materiais de revestimento das coberturas com uma exactidão cerca de 10 % superior à obtida com a imagem QuickBird com 2.4 m de resolução especial. Como as duas imagens possuem resolução espectral idêntica,

os resultados obtidos traduzem a importância que a resolução espacial tem neste tipo de aplicações. Ao nível da metodologia, a abordagem desenvolvida mostrou-se adequada para transformar a informação de elementos de superfície em informação em que a unidade de representação é o edifício e resolver algumas dificuldades como, por exemplo, a influência das sombras e da diferente iluminação dos vários elementos de superfície no processo de classificação. A representação ao nível do edifício, neste tipo de abordagem, possibilita a comparação dos resultados com métodos tradicionais e facilita a integração da informação num SIG. Um aspecto determinante para a obtenção de bons resultados com este tipo de metodologia combinada *pixel-object*, é existir um bom ajustamento entre a informação vectorial e a imagem utilizada. Dada a influência que as zonas de sombra na imagem têm neste tipo de aplicações, a adopção de estratégias como, por exemplo, a altura do ano e do dia em que as imagens são captadas, é igualmente importante, uma vez que permite reduzir a presença de sombras na imagem, com consequente melhoria dos resultados. A Exactidão Global do MAC foi de 78%, o que evidencia as potencialidades de aplicação da metodologia desenvolvida à reabilitação do espaço edificado.

Capítulo 5

Conclusões

5.1 Síntese do trabalho realizado

No âmbito desta tese de doutoramento, estudou-se a vantagem da integração da incerteza no processo de extracção de informação temática, a partir de Imagens Multiespectrais de Muito Grande Resolução Espacial (IMGRE), assim como na posterior avaliação da exactidão temática.

Numa primeira fase, investigou-se o comportamento de um conjunto de medidas de incerteza que permitem quantificar a ambiguidade em especificar uma solução única, *i.e.*, em atribuir uma única classe a uma Unidade Espacial de Análise (UEA) da imagem, nomeadamente, as medidas de não-especificidade, entropia e um *ratio* de incerteza disponibilizado no programa comercial IDRISI. O estudo foi efectuado com o objectivo de avaliar: (1) a sensibilidade das medidas à variação da compatibilidade entre as características das UEAs da imagem e as classes; e (2) a sensibilidade das medidas à dispersão, *i.e.*, quando a UEA apresenta, no resultado da classificação, uma atribuição partilhada a várias classes.

Numa segunda fase, desenvolveu-se uma abordagem metodológica que integra, no processo de classificação combinada *pixel/objecto*, a informação da incerteza. A informação sobre a incerteza associada à classificação ao nível do *pixel* foi integrada no processo de classificação ao nível do objecto. Esta abordagem metodológica é semelhante a uma árvore de decisão, com uma estrutura hierárquica constituída por vários níveis, aos quais são aplicados um conjunto de regras. Ao nível da integração da incerteza no processo de avaliação da exactidão temática, desenvolveram-se dois índices de incerteza: um índice baseado na média, correspondendo ao valor complementar da incerteza média por classe, e um índice de incerteza baseado nos baixos valores de incerteza, *i.e.*, correspondendo à percentagem de objectos por classe com incerteza inferior a 0.5. Para avaliar o potencial da informação disponibilizada pelos índices de incerteza foi feita uma análise comparativa com os resultados obtidos com os índices de exactidão estatísticos, Índice de Exactidão do Utilizador e Índice de Exactidão do Produtor,

tradicionalmente utilizados.

Dada a tendência de aumento significativo das acções de reabilitação urbana e a importância que a fase de diagnóstico representa na fase de planeamento, o desenvolvimento de metodologias e o teste de novas ferramentas de observação que permitam apoiar os vários agentes envolvidos no processo de reabilitação urbana, assume uma importância significativa. Neste contexto, considerou-se relevante abordar esta temática no âmbito deste trabalho, tendo-se estudado, igualmente, o potencial das IMGRE para extrair informação de suporte à avaliação do estado de conservação das coberturas do espaço edificado.

A metodologia de classificação proposta foi desenvolvida e testada primeiro no caso de uma área florestal, para obtenção de um Mapa de Unidades de Paisagem (MUP), tendo sido utilizadas imagens IKONOS multiespectrais com 4 m de resolução espacial. Posteriormente, foi aplicada ao caso do centro histórico da Baixa de Coimbra para obtenção de um mapa de anomalias e tipo de materiais das coberturas, tendo sido utilizadas imagens aéreas multiespectrais com 0.5 m de resolução espacial e imagens multiespectrais do satélite QuickBird com 2.4 m de resolução espacial.

A escolha do primeiro caso de estudo, prendeu-se com o facto de a produção de cartografia de ocupação de solo ser uma área em que as imagens obtidas por detecção remota têm servido com sucesso como fonte de informação, mas em que o desenvolvimento de métodos de classificação que permitam atingir melhores resultados continuam a ser uma área de investigação, assim como o desenvolvimento de métodos de avaliação da exactidão. Neste caso, o objectivo consistiu em investigar se a introdução da incerteza no processo de classificação combinada melhora a exactidão do mapa produzido e se os índices de incerteza podem disponibilizar alguma informação sobre a qualidade dos classificadores e/ou da classificação.

No segundo caso de estudo, avaliou-se a aplicabilidade da abordagem metodológica para a produção automática de um Mapa de Anomalias das Coberturas (MAC) do património edificado, onde a detecção remota não tem sido utilizada. O inventário e levantamento dos edifícios da Baixa de Coimbra são uma tarefa complexa, agravada pelos problemas de acessibilidade aos edifícios antigos que dificultam as acções de inspecção. Por este motivo, é importante a procura de novas ferramentas de observação e novas soluções que possam servir de apoio ao diagnóstico e inspecção, indispensáveis na fase de definição de uma estratégia e planeamento de uma acção de reabilitação. Esta preocupação torna-se mais relevante quando se pretende inspecionar um número elevado de edifícios (cerca de 800 edifícios no caso da Baixa de

Coimbra), num curto espaço de tempo.

Considera-se que este trabalho contribuiu para mostrar a utilidade da informação sobre a incerteza no âmbito do processo de produção da cartografia temática, bem como para melhorar o conhecimento sobre o comportamento das medidas de incerteza quando aplicáveis aos resultados de classificação temática, obtidos a partir de classificadores assistidos *fuzzy* e probabilísticos. Adicionalmente, considera-se que este trabalho contribuiu para a promoção da utilização dos dados de detecção remota de muito grande resolução espacial na área da reabilitação do património construído. Com efeito, a informação que é possível extrair destas imagens pode ser extremamente útil como apoio à fase de levantamento, diagnóstico e monitorização do edificado antigo ao nível das coberturas. Esta convicção é fundamentada com base nas conclusões, indicadas na secção 5.2, para cada uma das áreas de trabalho desenvolvidas nesta tese de doutoramento.

5.2 Principais conclusões

As principais conclusões dos trabalhos de investigação realizados são apresentadas em cada um dos capítulos respectivos, encontrando-se resumidas nos parágrafos seguintes.

O **segundo capítulo** é dedicado ao estudo comparativo das medidas de incerteza, nomeadamente: a medida de não-especificidade, NSp, propostas por Yagger (1982); uma nova métrica proposta na tese, Un, que corresponde à normalização da medida de incerteza-U proposta por Higashi e Klir (1982); a medida de entropia relativa, Er, proposta por Maselli *et al.* (1995); e o *ratio* de incerteza, RI, desenvolvido por Clark Lab. (2001). As medidas NSp, Un e RI foram aplicadas a distribuições de possibilidade associadas a UEA, obtidas com um classificador *fuzzy*, e as medidas Er e RI foram aplicadas a distribuições de probabilidade associadas a UEA, obtidas com um classificador não rígido probabilístico. Para ilustrar melhor o comportamento das medidas foram também utilizados conjuntos de valores sintéticos que simulam distribuições de probabilidade ou possibilidade.

Da análise comparativa do comportamento das três medidas Un, NSp e RI, sublinham-se os seguintes aspectos:

- A medida de incerteza NSp é sensível aos valores absolutos da distribuição de

possibilidade, enquanto as medidas Un e RI são sensíveis à diferença entre os graus consecutivos de possibilidades e à dispersão;

- A medida NSp avalia a incerteza considerando apenas as classes às quais foram atribuídas possibilidades positivas, enquanto as medidas Un e RI são sensíveis à dispersão entre todas as classes consideradas teoricamente possíveis, mesmo que tenham sido atribuídos graus de possibilidade com valores iguais a zero a alguma destas.

Do estudo realizado com as medidas Er e RI, a análise comparativa do seu comportamento demonstrou que as duas medidas são sensíveis à dispersão. Contudo, a medida Er é mais sensível do que a medida RI. Por outro lado, a medida RI é mais sensível ao valor atribuído ao melhor grau de probabilidade do que a medida Er.

As conclusões enunciadas revelam que, para avaliar a dispersão ou a variação entre os graus consecutivos de possibilidade, a medida Un é a mais adequada; ao passo que, para avaliar a compatibilidade entre as características das UEAs e as classes, deverá ser adoptada a medida NSp. A medida RI apresenta um comportamento próximo da medida Un, sendo ainda mais sensível à variação da dispersão. Relativamente ao estudo de comportamento entre as medidas RI e Er, as diferenças de comportamento verificadas não são significativas, não permitindo concluir que uma é mais apropriada do que outra. Contudo, constatou-se que a medida RI, quando aplicada a distribuições de probabilidade, é mais sensível ao valor atribuído ao melhor grau de probabilidade, do que a medida Er.

A escolha das medidas de incerteza a aplicar depende do classificador utilizado e do tipo de informação sobre incerteza que se pretende avaliar. Em detecção remota, a utilização das medidas baseadas na entropia e das medidas de não-especificidade é apenas apropriada em situações em que uma UEA apresenta, no resultado da classificação, uma atribuição partilhada por várias classes embora, no terreno, se admita que existe apenas uma classe. As medidas de não-especificidade NSp e Un, aplicáveis a conjuntos *fuzzy* e distribuições de possibilidade, apenas permitem quantificar a ambiguidade em especificar uma solução única.

O estudo comparativo do comportamento das várias medidas evidenciou as características de cada uma e permitiu concluir que têm comportamentos diferentes e disponibilizam informação diferente ao utilizador. Permitiu, igualmente, clarificar o tipo de informação que se pode inferir sobre a incerteza, quando se aplica este tipo de medidas aos resultados da classificação temática, obtida com classificadores não rígidos probabilísticos ou *fuzzy*, o que é útil do ponto de

vista do utilizador na tomada de decisão sobre que medida de incerteza aplicar.

No **capítulo 3** investigou-se a influência e utilidade que a informação da incerteza associada à classificação dos elementos de superfície teria se integrada no processo de classificação das Unidades de Paisagem. Foi também analisado o potencial da aplicação das medidas de incerteza na avaliação do comportamento de dois classificadores e da exactidão do mapa de ocupação do solo produzido.

As conclusões mais relevantes sobre a utilidade da integração da informação sobre a incerteza no processo de classificação são as seguintes:

- A determinação da informação sobre a incerteza da classificação temática do Mapa de Elementos de Superfície (MES), através da aplicação da medida de incerteza RI aos graus de probabilidade *a posteriori* obtidos com o classificador de MV_B, permitiu identificar os elementos de superfície em que ocorreu maior ambiguidade no processo de classificação e evitar o seu uso na identificação das Unidades de Paisagem;
- A utilização da informação sobre incerteza revelou-se importante no processo de classificação combinada *pixel/objecto*. Os resultados obtidos na avaliação do MUP mostraram que o valor da Exactidão Global, com o método de classificação combinada *pixel/objecto* com integração da informação de incerteza, melhorou 12% comparativamente ao obtido com a classificação utilizando um método semelhante mas sem a integração da incerteza;
- Do ponto de vista metodológico, a abordagem combinada *pixel/objecto* com integração da incerteza demonstrou ser apropriada à transformação do Mapa de Elementos de Superfície num Mapa de Unidades de Paisagem em formato vectorial, adequado a uma integração num Sistema de Informação Geográfica.

Para determinar se as medidas de incerteza podem ser utilizadas como indicadores para avaliar o comportamento dos classificadores e a exactidão da classificação, foi efectuada uma comparação entre os resultados dados pelas medidas de incerteza e a informação obtida a partir das matrizes de confusão. As conclusões mais relevantes relativamente a esta parte do trabalho são as seguintes:

- Com os índices de exactidão e a informação sobre a incerteza foi possível inferir conclusões idênticas sobre o comportamento dos dois classificadores. Ambos

permitiram concluir que: (1) o classificador *fuzzy*, baseado na distância mínima à média, não é adequado para a classificação da IMGRE, com a nomenclatura definida, devido à grande variabilidade espectral das classes na imagem; (2) o classificador probabilístico MV_B classifica bem os *pixels* para o qual foi treinado. Tal permite afirmar que a informação sobre a incerteza da classificação de regiões consideradas representativas das várias classes pode ser usada, conjuntamente com os graus de probabilidade ou possibilidade, para detectar os principais problemas dos classificadores;

- As análises mostraram que existe uma correlação positiva entre a informação obtida pelas medidas de incerteza e os índices de exactidão estatísticos, principalmente com o índice de exactidão do utilizador. Verificou-se também que essa correlação é maior entre o índice de exactidão do utilizador e o índice de incerteza baseado na média, do que entre o índice de incerteza baseado nos baixos valores de incerteza;
- A aplicação das medidas de incerteza a distribuições de probabilidade ou possibilidade é uma abordagem simples, permitindo uma mais rápida e económica avaliação dos classificadores do que os índices de exactidão;
- Os índices de incerteza não são influenciados pela subjectividade que pode existir nos dados de referência usados nos índices de exactidão;
- Os índices de incerteza podem ser aplicados a todo o mapa;
- A informação sobre a incerteza pode ser utilizada para avaliar o comportamento dos classificadores e, conseqüentemente, como um indicador da exactidão da classificação;
- Os índices de incerteza constituem uma ferramenta extremamente útil a utilizar numa fase preliminar, antes de se avançar para a recolha da informação de referência (processo moroso e dispendioso) e construção das matrizes de erro;
- A identificação de regiões e/ou classes com elevado nível de incerteza pode auxiliar, numa fase preliminar, na identificação de problemas que podem conduzir à implementação de medidas que reduzam a incerteza da classificação, com conseqüente melhoramento da exactidão.

O **capítulo 4** aborda a utilização das IMGRE como ferramenta de suporte para extracção de informação necessária à avaliação do estado de conservação das coberturas do espaço edificado. O primeiro procedimento consistiu na exploração espectral das imagens aéreas e das imagens Quickbird com, respectivamente, 0.5 m e 2.4 m de resolução espacial, para estudar a separabilidade espectral entre as classes de elementos de superfície das coberturas, nas quais se incluem os principais tipos de materiais de revestimento, existentes na área de estudo, que são chapa metálica, fibrocimento e telha cerâmica, assim como o revestimento de telha cerâmica não deteriorada e com anomalias. O estudo de separabilidade revelou que os elementos de superfície dos materiais de revestimento na imagem QuickBird se confundem entre si, o que significa uma grande proximidade das assinaturas espectrais destas classes nesta imagem. Pelo contrário, na imagem aérea, os elementos de superfície dos materiais apresentam uma separação quase total, verificando-se apenas pequenas sobreposições entre as classes Fibrocimento, Telha Cerâmica Escura e Sombra.

Relativamente às classes dos elementos de superfície das coberturas de telha cerâmica não deteriorada e com anomalias, o estudo revelou uma boa separabilidade entre as classes, nas duas imagens, verificando-se apenas uma maior sobreposição, na imagem QuickBird, entre as assinaturas espectrais das classes Telha Cerâmica Não Deteriorada e Telha Cerâmica com Anomalia.

Para a produção do MAC, aplicou-se a metodologia combinada *pixel/objecto* com integração de incerteza, a qual apresenta algumas semelhanças com a abordagem ao nível da parcela, pois os objectos são os contornos dos edifícios obtidos a partir da cartografia digital à escala 1:1 000. Dos resultados expostos no capítulo 4 salientam-se as seguintes conclusões:

- Os Mapas de Elementos de Superfície das Anomalias (MES-A), e os Mapas de Elementos de Superfície dos Materiais de Revestimento (MES-M) das coberturas dos edifícios, obtidos a partir da classificação da imagem aérea, permitiram identificar a presença de anomalias e os diferentes materiais de revestimento das coberturas com uma exactidão cerca de 10 % superior à obtida com a imagem QuickBird. Os resultados obtidos revelam a importância que a resolução espacial tem neste tipo de aplicações, uma vez que as duas imagens possuem resolução espectral idêntica;
- A Exactidão Global do Mapa de Anomalias das Coberturas (MAC) da Baixa de Coimbra produzido foi de 78%, o que evidencia as potencialidades de aplicação da metodologia desenvolvida à reabilitação do espaço edificado;

- Um aspecto determinante para a obtenção de bons resultados com este tipo de metodologia é existir um bom ajustamento entre a informação cartográfica vectorial e a imagem utilizada. Os desfasamentos entre a imagem e a informação cartográfica levam a que alguns *pixels* situados no interior dos polígonos supostamente correspondentes aos edifícios efectivamente não lhes pertençam e, como consequência, geram conclusões erradas na aplicação das regras;
- O estudo mostrou que a utilização das IMGRE para extracção de informação de suporte à avaliação do estado de conservação das coberturas é bastante promissora. Contudo, apesar da grande resolução espacial, estas imagens apresentam ainda algumas limitações atendendo a que apenas permitem a identificação das coberturas com e sem anomalia, não possibilitando a diferenciação dos seus diferentes tipos;
- Ao nível da metodologia, a abordagem desenvolvida mostrou-se adequada para transformar a informação de elementos de superfície, em que a unidade mínima de representação considerada é o *pixel*, na informação em que a unidade de representação é o edifício. A aplicação de regras permitiu resolver algumas dificuldades, com significativo sucesso, como por exemplo a influência das sombras e da diferente iluminação dos vários elementos de superfície no processo de classificação.

A análise comparativa entre a informação do MAC, obtido de forma automática com a abordagem desenvolvida, e o mapa de referência, obtido através de métodos tradicionais pela Universidade de Coimbra, permitiu constatar que:

- dos edifícios identificados como possuindo cobertura de telha cerâmica na referência, 85% foram também identificados com esta nova abordagem;
- das coberturas que possuíam revestimento de Telha Cerâmica com anomalias na referência, 77% também foram identificadas com a nova abordagem;
- atendendo à discrepância temporal entre as imagens usadas e os dados de referência, alguns dos diferentes resultados obtidos podem estar correctos nos dois estudos, por terem ocorrido algumas alterações durante esse período nas coberturas dos edifícios.

5.3 Perspectivas de trabalho futuro

Os trabalhos de investigação realizados permitiram identificar algumas linhas de desenvolvimentos futuros, dos quais se salientam:

- A aplicação da metodologia de classificação proposta a outros casos de estudo. Investigar igualmente se a utilização de outros classificadores, como por exemplo, redes *neuro-fuzzy* e aplicação de medidas de não-especificidade, pode melhorar os resultados finais do mapa de unidades de paisagem ou o mapa de anomalias das coberturas;
- Tendo-se constatado que a distribuição espacial da incerteza disponibiliza informação extremamente útil que pode ser usada no processo de classificação, a sua integração no processo de combinação de múltiplos classificadores é outra possível área de investigação;
- No processo de classificação poderá avaliar-se se a combinação dos resultados de vários classificadores, com base na informação sobre a incerteza obtida para cada um, permite obter classificações mais exactas;
- Investigar o comportamento de outras medidas de incerteza, como por exemplo, as desenvolvidas no âmbito da teoria da evidência, e testar a sua aplicação para quantificar a incerteza existente nos mapas produzidos;
- Incluir a informação acerca da semelhança de cada ponto do conjunto de treino às características ideais das classes e usar essa informação no processo de classificação. Para tal, seria necessário efectuar algumas adaptações aos métodos existentes ou desenvolver novos métodos de classificação de forma a modelar a informação acerca da incerteza das áreas de treino, por exemplo, com conjuntos *fuzzy* e desenvolver classificadores que sejam capazes de usar essa informação;
- Na utilização das IMGRE na identificação das anomalias das coberturas do edificado, um dos aspectos que influencia significativamente os resultados produzidos está relacionado com a existência de zonas de sombra na imagem. De forma a atenuar a sua influência, podem ser adoptadas estratégias como, por exemplo, utilizar imagens captadas em determinadas alturas do ano e do dia de forma a reduzir a presença de

sombras na imagem e a melhorar os resultados finais. A utilização de imagens hiperespectrais, obtidas com sensores a bordo de aviões que captam imagens cujos comprimentos podem variar entre os 400 nm e os 2500 nm, constitui igualmente uma área a investigar;

- Estando em curso uma tese de doutoramento sobre a utilização de classificação automática de imagens fotográficas multiespectrais para identificar anomalias em estruturas de betão aparente (e.g. fissuração; colonização biológica; manchas de humidade; áreas reparadas; carbonatação; ataque de cloretos; entre outras), avaliar a utilidade da integração da incerteza na fase de classificação e na avaliação dos resultados obtidos, através da aplicação dos índices de incerteza propostos na presente tese de doutoramento.

B i b l i o g r a f i a

- APLIN, P., ATKINSON, P. and CURRAN, P., 1999, Per-field classification of land use using the forthcoming very fine spatial resolution satellite sensors: problems and potential solutions *Advances in remote sensing and GIS analysis*. P. M. Atkinson and N. J. Tate (eds), (Chichester: John Wiley & Sons), pp. 219-239.
- ARORA, M. K. and FOODY, G. M., 1997, Log-linear modelling for the evaluation of the variables affecting the accuracy of probabilistic, fuzzy and neural network classifications. *International Journal of Remote Sensing*, **18** (4): 785-798.
- ATKINSON, P. and FOODY, G. M., 2002, Uncertainty in remote sensing and GIS: fundamentals. *Uncertainty in remote sensing and GIS*. Giles M. Foody and Peter M. Atkinson (eds), (Chichester: John Wiley & Sons), pp. 1-18.
- BAATZ, M. and SCHAPE, A., 2000, Multiresolution segmentation - an optimization approach for high quality multi-scale image segmentation. In *Angewandte Geographische Informationsverarbeitung XII, Beiträge zum AGIT-Symposium Salzburg 2000*, J. Strobl, T. Blaschke and G. Griesebner (Ed.), (Herbert Wichmann Verlag), pp. 12-23.
- BAIO, M., 1996, Avaliação de alterações de ocupação do solo por imagens numéricas multiespectrais, Tese de Mestrado, Universidade Nova de Lisboa, Lisboa.
- BENZ, U. C., HOFMANN, P., WILLHAUCK, G., LINGENFELDER, I. and HEYNEN, M., 2004, Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing*, **58** (3-4): 239-258.
- BINAGHI, E., BRIVIO, P. A., GHEZZI, P. and RAMPINI, A., 1999, A fuzzy set-based accuracy assessment of soft classification. *Pattern Recognition Letters*, **20** (9): 935-948.
- BINAGHI, E., MADELLA, P., MONTESANO, M. G. and RAMPINI, A., 1997, Fuzzy contextual classification of multisource remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, **35** (2): 326-340.

- BLASCHKE, T., BURNETT, C. and PEKKARINEN, A., 2004, Image segmentation methods for object-based analysis and classification. *Remote Sensing Image Analysis: Including the Spatial Domain*. M. de Jong Steven and D. van der Meer Freek (eds), (Dordrecht: Springer), pp. 211-236.
- BLASCHKE, T., LANG, S., LORUP, E., STROBL, J. and ZEIL, P., 2000, Object-Oriented Image Processing in an Integrated GIS/Remote Sensing Environment and Perspectives for Environmental Applications. *Environmental Information for Planning, Politics and the Public*. A. Cremers and k. Greve (eds), (Marburg: Metropolis Verlag), pp. 555-570.
- Bo, Y. C. and WANG, J. F., 2008, A general method for assessing the uncertainty in classified remotely sensed data at pixel scale. In *8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Shanghai, China, D. Li, Y. Ge and G. M. Foody (Ed.), (World Acad Union-World Acad Press), pp. 186-194.
- BONISSONE, P. P., BADAMI, V., CHIANG, K. H., KHEDKAR, P. S., MARCELLE, K. W. and SCHUTTEN, M. J., 1995, Industrial applications of fuzzy logic at General Electric. *Proceedings of the IEEE*, **83** (3): 450-465.
- BROWN, K. M., FOODY, G. M. and ATKINSON, P. M., 2009, Estimating per-pixel thematic uncertainty in remote sensing classifications. *International Journal of Remote Sensing*, **30** (1): 209-229.
- BROWN, M., LEWIS, H. and GUNN, S., 2000, Linear spectral mixing models and support vector machines for remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, **28** (5): 2346-2360.
- CAETANO, M., 1995, Burned vegetation mapping in mountainous areas with satellite remote sensing, Ma Thesis, University of California, Santa Barbara.
- CARD, D. H., 1982, Using known map category marginal frequencies to improve estimates of thematic map accuracy. *Photogrammetric Engineering and Remote Sensing*, **48** (3): 431-439.
- CARLEER, A. and WOLFF, E., 2004, Exploitation of very high resolution satellite data for tree species identification. *Photogrammetric Engineering and Remote Sensing*, **70** (1): 135-140.

- CARRÃO, H., CAETANO, M. and COELHO, P., 2007, Sample Design and Analysis for Thematic Map Accuracy Assessment: An Approach Based on Domain Estimation for the Validation of Land Cover Products. *32nd International Symposium on Remote Sensing of Environment*, San Jose, Costa Rica.
- CHO, M. A. and SKIDMORE, A. K., 2009, Hyperspectral predictors for monitoring biomass production in Mediterranean mountain grasslands: Majella National Park, Italy. *International Journal of Remote Sensing*, **30** (2): 499-515.
- CHUBEY, M. S., FRANKLIN, S. E. and WULDER, M. A., 2006, Object-based analysis of Ikonos-2 imagery for extraction of forest inventory parameters. *Photogrammetric Engineering and Remote Sensing*, **72** (4): 383-394.
- CLARK LAB., 2001, IDRISI guide to GIS and image processing.
- CLEVE, C., KELLY, M., KEARNS, F. R. and MORLTZ, M., 2008, Classification of the wildland-urban interface: A comparison of pixel-and object-based classifications using high-resolution aerial photography. *Computers Environment and Urban Systems*, **32** (4): 317-326.
- CONGALTON, R. G. and GREEN, K., 1999, Assessing the accuracy of remotely sensed data: principles and practices. (Boca Raton, FL: CRC/ Lewis Press).
- DELUCA, A. and TERMINI, S., 1972, A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. *Information and Control*, **20** (4): 301-312.
- DEMPSTER, A. P., 1967, Upper and Lower Probabilities Induced by a Multivalued Mapping. *Annals Mathematics Statistics*, **38**: 325-339.
- DOAN, H. T. X. and FOODY, G. M., 2007, Increasing soft classification accuracy through the use of an ensemble of classifiers. *International Journal of Remote Sensing*, **28** (20): 4609-4623.
- DONNAY, J.-P., BARNSLEY, M. and LONGLEY, P., 2001, Remote sensing and urban analysis. *Remote sensing and urban analysis*. J.-P. Donnay, M. Barnsley and P. Longley (eds), (New York: Taylor & Francis), pp. 3-18.

- DONNAY, J.-P. and THOMSIN, L., 1994, Urban remote sensing and statistics: prospective research and application. In *Proceedings of the Symposium: New Tools for Spatial Analysis*, M. Painlo (Ed.), (Eurostat (Luxembourg:Office for Official Publication of the European Communities)), pp. 137-145.
- DUBOIS, D., MAURIS, L. and PRADE, H., 2004, Probability-Possibility Transformations, Triangular Fuzzy Sets, and Probabilistic Inequalities. *Reliable Computing*, **10** (4): 273-297.
- DUBOIS, D., OSTASIEWICK, W. and PRADE, H., 2000, Possibility theory, probability and fuzzy sets. *Fundamentals of Fuzzy Sets*. D. Dubois and H. Prade (eds), (The Handbook of Fuzzy Sets Series, Kluwer Acad. Publ.), pp. 343-438.
- DUBOIS, D. and PRADE, H., 1985, A note on measures of specificity for fuzzy sets. *International Journal of General Systems*, **10** (4): 279-283.
- DUBOIS, D. and PRADE, H., 1993, Fuzzy sets and probability: Misunderstanding, bridges and gaps. *Second IEEE International Conference on Fuzzy Systems*, San Francisco, CA, USA.
- DUBOIS, D. and PRADE, H., 1997, The three semantics of fuzzy sets. *Fuzzy Sets and Systems*, **90** (2): 141-150.
- EASTMAN, J. R., 2006, IDRISI Andes guide to GIS and image processing. Clark Lab., Clark University.
- EASTMAN, J. R. and LANEY, R. M., 2002, Bayesian soft classification for sub-pixel analysis: A critical evaluation. *Photogrammetric Engineering and Remote Sensing*, **68** (11): 1149-1154.
- ERBEK, F. S., OZKAN, C. and TABERNER, M., 2004, Comparison of maximum likelihood classification method with supervised artificial neural network algorithms for land use activities. *International Journal of Remote Sensing*, **25** (9): 1733-1748.
- FONSECA, A. M., 2004, Classificação da ocupação do solo urbano utilizando funções de pertença sobre uma imagem IKONOS da cidade de Lisboa. In *III Conferência Nacional de Cartografia e Geodesia*, L. Bastos and J. Bastos (Ed.), (Lidel - Edições Técnicas Lda.), pp. 135-145.
- FONTE, C. C., 2003, Entidades Geográficas Difusas, Tese Doutorado, Universidade de Coimbra, Coimbra.

- FOODY, G. M., 1995, Cross-entropy for the evaluation of the accuracy of a fuzzy land cover classification with fuzzy ground data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **50** (5): 2-12.
- FOODY, G. M., 1996, Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data. *International Journal of Remote Sensing*, **17** (7): 1317-1340.
- FOODY, G. M., 1999, The continuum of classification fuzziness in thematic mapping. *Photogrammetric Engineering & Remote Sensing*, **65** (4): 443-451.
- FOODY, G. M., 2000a, Estimation of sub-pixel land cover composition in the presence of untrained classes. *Computers & Geosciences*, **26** (4): 469-478.
- FOODY, G. M., 2000b, Mapping land cover from remotely sensed data with a softened feedforward neural network classification. *Journal of Intelligent & Robotic Systems*, **29** (4): 433-449.
- FOODY, G. M., 2001, Thematic mapping from remotely sensed data with neural networks: MLP, RBF and PNN based approaches. *Journal of Geographical Systems*, **3** (3): 217-232.
- FOODY, G. M., 2002, Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, **80** (1): 185-201.
- FOODY, G. M., 2004, Sub-Pixel Methods in Remote Sensing. *Remote Sensing Image Analysis: Including the Spatial Domain* Steven M. de Jong and Freek D. Van Der Meer (eds), (Dordrecht: Springer), pp. 37-70.
- FOODY, G. M., 2005, Local characterization of thematic classification accuracy through spatially constrained confusion matrices. *International Journal Remote Sensing*, **26** (6): 1217-1228.
- FOODY, G. M., 2008a, Harshness in image classification accuracy assessment. *International Journal of Remote Sensing*, **29** (11): 3137-3158.
- FOODY, G. M., 2008b, Sample size determination for image classification accuracy assessment and comparison. In *8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Shanghai, China, D. Li, Y. Ge and G. M. Foody (Ed.), (World Academic Union Liverpool), pp. 154-162.

- FOODY, G. M. and ARORA, M. K., 1996, Incorporating mixed pixels in the training, allocation and testing stages of supervised classifications. *Pattern Recognition Letters*, **17** (13): 1389-1398.
- FOODY, G. M. and ARORA, M. K., 1997, An evaluation of some factors affecting the accuracy of classification by an artificial neural network. *International Journal of Remote Sensing*, **18** (4): 799-810.
- FOODY, G. M. and ATKINSON, P., 2002, *Uncertainty in remote sensing and GIS*. (Hoboken, NJ: John Wiley & Sons).
- FOODY, G. M., BOYD, D. S. and CUTLER, M. E. J., 2003, Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sensing of Environment*, **85** (4): 463-474.
- FOODY, G. M., BOYD, D. S. and SANCHEZ-HERNANDEZ, C., 2007, Mapping a specific class with an ensemble of classifiers. *International Journal of Remote Sensing*, **28** (7-8): 1733-1746.
- FOODY, G. M., CAMPBELL, N. A., TRODD, N. M. and WOOD, T. F., 1992, Derivation and Applications of Probabilistic Measures of Class Membership from the Maximum-Likelihood Classification. *Photogrammetric Engineering and Remote Sensing*, **58** (9): 1335-1341.
- FOODY, G. M. and MATHUR, A., 2006, The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. *Remote Sensing of Environment*, **103** (2): 179-189.
- FOODY, G. M., MCCULLOCH, M. B. and YATES, W. B., 1995, Classification of remotely-sensed data by an artificial neural-network - issues related to training data characteristics. *Photogrammetric Engineering and Remote Sensing*, **61** (4): 391-401.
- FRIEDL, M. A. and BRODLEY, C. E., 1997, Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, **61** (3): 399-409.
- FRIEDL, M. A., BRODLEY, C. E. and STRAHLER, A. H., 1999, Maximizing land cover classification accuracies produced by decision trees at continental to global scales. *IEEE Transactions on Geoscience and Remote Sensing*, **37** (2): 969-977.

- GAHEGAN, M. and EHLERS, M., 2000, A framework for the modelling of uncertainty between remote sensing and geographic information systems. *ISPRS Journal of Photogrammetry and Remote Sensing*, **55** (3): 176-188.
- GAMBA, P. and DELL'ACQUA, F., 2003, Increased accuracy multiband urban classification using a neuro-fuzzy classifier. *international Journal of Remote Sensing*, **24** (4): 827-834.
- GOETZ, S. J., WRIGHT, R. K., SMITH, A. J., ZINECKER, E. and SCHAUB, E., 2003, IKONOS imagery for resource management: Tree cover, impervious surfaces, and riparian buffer analyses in the mid-Atlantic region. *Remote Sensing of Environment*, **88** (1-2): 195-208.
- GONÇALVES, L. M. and CAETANO, M., 2004, Classificação das imagens do satélite IKONOS utilizando uma abordagem orientada por objectos. In *III Conferência Nacional de Cartografia e Geodesia*, Lisboa, L. Bastos and J. Matos (Ed.), (Lidel), pp. 287-298.
- GONÇALVES, L. M. S. and CAETANO, M., 2008, Estudo sobre um futuro aumento da resolução espacial dos sensores de muito grande resolução espacial. *X Encontro de Sistemas de Informação Geográfica (ESIG' 2008)*, Oeiras, Portugal.
- GOODCHILD, M. E., 2003, Geographic information science and systems for environmental management. *Annual Review of Environment and Resources*, **28**: 493-519.
- GOPAL, S. and WOODCOCK, C., 1994, Theory and methods for accuracy assessment of thematic maps using fuzzy-sets. *Photogrammetric Engineering and Remote Sensing*, **60** (2): 181-188.
- GUERRERO, I., M.TANASE, MANAKOS, I. and I.GITAS, 2006, A semi-operational approach for land cover mapping in the Mediterranean. *Proceedings of the 26th EARSeL Symposium: New developments and challenges in remote sensing*, Warsaw, Poland.
- HARTLEY, R. V. L., 1928, Transmission of information. *Bell System Technical Journal*, **7** (3): 535-563.
- HAY, G. J., CASTILLA, G., WULDER, M. A. and RUIZ, J. R., 2005, An automated object-based approach for the multiscale image segmentation of forest scenes. *International Journal of Applied Earth Observation and Geoinformation* **7**(4): 339-359.

- HEMMLER, M., WERTZ, F. and MAIERHOFER, C., 2005, Damage detection on buildings surfaces with multi-spectral techniques. *XX CIPA International Symposium*, Torino, Italy.
- HEROLD, M., GARDNER, M. E. and ROBERTS, D. A., 2003, Spectral resolution requirements for mapping urban areas. *IEEE Transactions on Geoscience and Remote Sensing*, **41** (9): 1907-1919.
- HEUVELINK, G., 1998, Error propagation in environmental modelling with GIS. (London: Taylor & Francis).
- HIGASHI, M. and KLIR, G., 1982, Measures of uncertainty and information based on possibility distributions. *International Journal of General Systems*, **9** (1): 43-58.
- HOFMANN, P., 2001, Detecting informal settlements from Ikonos image data using methods of object oriented image analysis—an example from Cape Town (South Africa). In *Remote Sensing of Urban Areas / Fernerkundung in urbanen Räumen. (Regensburger Geographische Schriften, Heft 35)*. , Regensburg/Germany, C. JÜRGENS (Ed.), (Institut für Geographie an der Universität Regensburg), pp. 41-42.
- HUANG, Z. and LEES, B. G., 2004, Combining non-parametric models for multisource predictive forest mapping. *Photogrammetric Engineering and Remote Sensing*, **70** (4): 415-425.
- HUNG, M. C. and RIDD, M. K., 2002, A subpixel classifier for urban land-cover mapping based on a maximum-likelihood approach and expert system rules. *Photogrammetric Engineering and Remote Sensing*, **68** (11): 1173-1180.
- IBRAHIM, M. A., ARORA, M. K. and GHOSH, S. K., 2005, Estimating and accommodating uncertainty through the soft classification of remote sensing data. *International Journal of Remote Sensing*, **26** (14): 2995-3007.
- INKPEN, R. J., FONTANA, D. and COLLIER, P., 2001, Mapping decay: Integrating scales of weathering within a GIS. *Earth Surface Processes and Landforms, Spatial Issue: Weathering 2000*, **26** (8): 885-900.
- ISHIBUCHI, H., NAKASHIMA, T. and MURATA, T., 1999, Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, **29** (5): 601-618.

- KLIR, G. J., 1987, Where do we stand on measures of uncertainty, ambiguity, fuzziness and the like? *Fuzzy Sets and Systems*, **24** (2): 141-160.
- KLIR, G. J., 1999, Uncertainty and Information Measures for Imprecise Probabilities: An Overview. *1^a International Symposium on Imprecise Probabilities and Their Applications*, Ghent, Belgium.
- KLIR, G. J., 2000, Measures of uncertainty and information. *Fundamentals of fuzzy sets*. D. Dubois and H. Prade (eds), (Springer), pp. 439-457.
- KLIR, G. J., 2001, Foundations of fuzzy set theory and fuzzy logic: A historical overview. *International Journal of General Systems*, **30** (2): 91-132.
- KLIR, G. J., 2004, Generalized information theory: aims, results, and open problems. *Reliability Engineering & System Safety* **85**: 21-38.
- KLIR, G. J. and FOLGER, T., 1988, Fuzzy Sets, Uncertainty, and Information. (Englewood Cliffs, New Jersey: Prentice Hall).
- KLIR, G. J. and SMITH, R. M., 2001, On measuring uncertainty and uncertainty-based information: Recent developments. *Annals of Mathematics and Artificial Intelligence*, **32** (1-4): 5-33.
- KLIR, G. J. and YUAN, B., 1995, Fuzzy sets and fuzzy logic-theory and applications. (New Jersey: Prentice Hall PTR).
- LAHA, A., PAL, N. R. and DAS, J., 2006, Land cover classification using fuzzy rules and aggregation of contextual information through evidence theory. *IEEE Transactions on Geoscience and Remote Sensing*, **44** (6): 1633-1641.
- LEITNER, M. and BUTTENFIELD, B., 2000, Guidelines for the display of attribute certainty. *Cartography and Geographic Information Science*, **27** (1): 3-14.
- LERMA, J. L., 2001, Automatic Feature Recognition Technique on Stone Monuments Using Visible and IR Photography. *International Cultural Heritage Informatics Meeting (ICHIM)*, Milano, Italy.
- LERMA, J. L., 2005, Automatic Plotting of Architectural Facades with Multispectral Images. *Journal of Surveying Engineering*, **131** (3): 73-77.

- LILLESAND, T. M., KIEFER, R. W. and JONATHAN, W. C., 2004, Remote Sensing and Image Interpretation. (U.S.A: John Wiley & Sons).
- LIU, W. G., GOPAL, S. and WOODCOCK, C. E., 2004, Uncertainty and confidence in land cover classification using a hybrid classifier approach. *Photogrammetric Engineering and Remote Sensing*, **70** (8): 963-971.
- LO, C. P., 1995, Automated population and dwelling unit estimation from high-resolution satellite images: a GIS approach. *International Journal of Remote Sensing*, **16** (1): 17-34.
- LU, D., 2006, The potential and challenge of remote sensing-based biomass estimation. *International Journal of Remote Sensing*, **27** (7): 1297-1328.
- LU, D. and WENG, Q., 2007, A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, **28** (5): 823-870.
- LU, D. S., WENG, Q. H. and LI, G. Y., 2006, Residential population estimation using a remote sensing derived impervious surface approach. *International Journal of Remote Sensing*, **27** (16): 3553-3570.
- MACKAY, D., SAMANTA, S., AHL, D., EWERS, B., GOWER, S. and BURROWS, S., 2003, Automated parameterization of land surface process models using fuzzy logic. *Transaction in GIS*, **7** (1): 139-153.
- MASELLI, F., CONESE, C., FILIPPIS, T. D. and NORCINI, S., 1995, Estimation of forest parameters through fuzzy classification of TM data. *IEEE Transactions on Geoscience and Remote Sensing*, **33** (1): 77-84.
- MASELLI, F., CONESE, C. and PETKOV, L., 1994, Use of probability entropy for the estimation and graphical representation of the accuracy of maximum likelihood classifications. *ISPRS Journal of Photogrammetry and Remote Sensing*, **49** (2): 13-20.
- OKI, K., UENISHI, T. M., OMASA, K. and TAMURA, M., 2004, Accuracy of land cover area estimated from coarse spatial resolution images using an unmixing method. *International Journal of Remote Sensing*, **25** (9): 1673-1683.

- PAL, N. R. and BEZDEK, J. C., 1994, Measuring fuzzy uncertainty. *IEE Transactions on Fuzzy Systems*, **2** (2): 107-118.
- PAL, N. R. and BEZDEK, J. C., 2000, Quantifying different facets of fuzzy uncertainty. *Fundamentals of Fuzzy Sets* D. Dubois and H. Prade (eds), (Springer), pp. 459-480.
- PLANTIER, T. and CAETANO, M., 2007, Mapas do Coberto Florestal: Abordagem Combinada Pixel/Objecto. In *V Conferência Nacional de Cartografia e Geodesia*, L. Bastos and J. Matos (Ed.), (Lidel - Edições Técnicas Lda.), pp. 157-166.
- PONTIUS, R. G. and CHEUK, M. L., 2006, A generalized cross-tabulation matrix to compare soft-classified maps at multiple resolutions. *International Journal of Geographical Information Science*, **20** (1): 1-30.
- QIU, F. and JENSEN, J. R., 2004, Opening the black box of neural networks for remote sensing image classification. *International Journal of Remote Sensing*, **25** (9): 1749-1768.
- REZNIK, L., GHANAYEM, O. and BOURMISTROV, A., 2000, PID plus fuzzy controller structures as a design base for industrial applications. *Engineering Applications of Artificial Intelligence*, **13** (4): 419-430.
- RICOTTA, C., 2004, Evaluating the classification accuracy of fuzzy thematic maps with a simple parametric measure. *International Journal of Remote Sensing*, **25** (11): 2169-2176.
- RICOTTA, C., 2005, On possible measures for evaluating the degree of uncertainty of fuzzy thematic maps. *International Journal of Remote Sensing*, **26** (24): 5573-5583.
- RUIZ, L. A., LERMA, J. L. and GIMENO, J., 2002, Application of Computer Vision Techniques to Support in the Restoration of Historical Buildings *Photogrammetric Computer Vision, ISPRS Commission III, Symposium 2002*, Graz, Austria
- SCHIEWE, J. and EHLERS, M., 2007, Fuzzy models for handling uncertainty in the integration of high resolution remotely sensed data and GIS. *Geographic Uncertainty in Environmental Security*. A. Morris and S. Kokhan (eds), (Springer Netherlands), pp. 89-106.

- SCHIEWE, J., TUFTE, L. and EHLERS, M., 2001, Potential and problems of multi-scale segmentation methods in remote sensing. *GeoBIT/GIS*, **6**: 34-39.
- SCHMIDT, K. S., SKIDMORE, A. K., KLOOSTERMAN, E. H., VAN OOSTEN, H., KUMAR, L. and JANSSEN, J. A. M., 2004, Mapping coastal vegetation using an expert system and hyperspectral imagery. *Photogrammetric Engineering and Remote Sensing*, **70** (6): 703-715.
- SCHOTT, J. R., 1997, *Remote Sensing: The Image Chain Approach* (New York: Oxford University Press).
- SHAFER, G., 1976, *A mathematical theory of evidence*. (Princeton: Princeton University Press).
- SHALAN, M. A., ARORA, M. K. and GHOSH, S. K., 2003, An evaluation of fuzzy classifications from IRS 1C LISS III imagery: a case study. *International Journal of Remote Sensing*, **24** (15): 3179-3186.
- SHANNON, C. E., 1948, A mathematical theory of communication. *Bell System Technical Journal*, **27** (3): 379-423.
- SHI, W. Z., EHLERS, M. and TEMPFLI, K., 1999, Analytical modelling of positional and thematic uncertainties in the integration of remote sensing and geographical information systems. *Transactions in GIS*, **3** (2): 119-136.
- SOUSA, G. C. D. and BOSE, B. K., 1994, A fuzzy set theory based control of a phase-controlled converter DCmachine drive. *IEEE Transactions on industry Applications*, **30** (1): 34-44.
- STEHMAN, S. V., 2001, Statistical rigor and practical utility in thematic map accuracy assessment. *Photogrammetric Engineering & Remote Sensing*, **67** (6): 727-734.
- STEHMAN, S. V. and CZAPLEWSKI, R. L., 1998, Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of Environment*, **64** (3): 331-344.
- STONE, W. E., 2002, Remote sensing in the 21st century: outlook for the future. Available on line at: http://rst.gsfc.nasa.gov/Sect21/Sect21_1.html (accessed February 2008).
- STUCKENS, J., COPPIN, P. R. and BAUER, M. E., 2000, Integrating contextual information with per-pixel classification for improved land cover classification. *Remote Sensing of Environment*, **71** (3): 282-296.

- TAUBENBÖCK, H., THOMAS, E. and ACHIM, R., 2006, An urban classification approach based on an object-oriented analysis of high resolution satellite imagery for a spatial structuring within urban areas. *1st EARSeL Workshop of the SIG Urban Remote Sensing*, Humboldt-Universität zu Berlin, 2-3 March 2006.
- TENEDORIO, J. A., ENCARNACAO, S., ESTANQUEIRO, R. and ROCHA, J., 2005, Great Lisbon Metropolitan Area land use/cover characterization through multi-temporal and multi-resolution VIS components analysis. In *25th Symposium of the European-Association-of-Remote-Sensing-Laboratories (EARSeL)*, Porto, PORTUGAL, A. Marcal (Ed.), (Millpress Science Publishers), pp. 495-502.
- VEREGIN, H., 1995, Developing and testing of an error propagation model for GIS overlay operations *International Journal of Geographical Information Systems*, **9** (6): 595-619.
- VICENTE, R. D. S., 2008, Estartégias e metodologias para intervenção urbana - Avaliação da vulnerabilidade e risco sísmico do edificado da Baixa de Coimbra, Tese Doutorado, Universidade de Aveiro, Aveiro.
- WALTER, V., 2004, Object-based classification of remote sensing data for change detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, **58** (3-4): 225-238.
- WANG, F., 1990, Improving remote sensing image analysis through fuzzy information representation. *Photogrammetric Engineering and Remote Sensing*, **56** (8): 1163-1169.
- WANG, L., SOUSA, W. P. and GONG, P., 2004, Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery. *International Journal of Remote Sensing*, **25** (24): 5655-5668.
- WICKHAM, J. D., STEHMAN, S. V., SMITH, J. H., WADE, T. G. and YANG, L., 2004, A priori evaluation of two-stage cluster sampling for accuracy assessment of large-area land cover maps. *International Journal of Remote Sensing*, **25** (6): 1235-1252.
- WILKINSON, G. G., 2005, Results and implications of a study of fifteen years of satellite image classification experiments. *IEEE Transactions on Geoscience and Remote Sensing*, **43** (3): 433-440.

- WOODCOCK, C. E., 2002, Uncertainty in remote sensing. *Uncertainty in remote sensing and GIS*. Giles M. Foody and Peter M. Atkinson (eds), (Chichester: John Wiley & Sons), pp. 19-24.
- WOODCOCK, C. E. and GOPAL, S., 2000, Fuzzy set theory and thematic maps: accuracy assessment and area estimation. *International Journal of Geographical Information Science*, **14** (2): 153-172.
- WU, S.-S., QIU, X. and WANG, L., 2005, Population Estimation Methods in GIS and Remote Sensing: A Review. *GIScience & Remote Sensing*, **42** (1): 80-96.
- YAGER, R. R., 1982, Measuring tranquility and anxiety in decision making: an application of fuzzy sets. *International Journal of General Systems*, **8** (3): 139-146.
- YAGER, R. R., 1983, Entropy and specificity in a mathematical theory of evidence. *International Journal of General Systems*, **9** (4): 249-260.
- YONG GE, SNAPING, L., DUAN, R., BAI, H. and CAO, F., 2008, Multi-level measurements for uncertainty in classified remotely sensed imagery. In *Accuracy in Geomatics* D. Li, Y. Ge and G. Foody (Ed.), (World Academic Union Liverpool), pp. 171-178.
- ZADEH, L. A., 1965, Fuzzy sets. *IEEE Transactions Information and Control*, **8** (3): 338-353.
- ZADEH, L. A., 1978, Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, **1**: 3-28.
- ZHANG, J., 1996, A surface approach to the handling of uncertainties in an integrated spatial database environment. *Cartographica*, **33** (1): 23-31.
- ZHANG, J. and FOODY, G. M., 2001, Fully-fuzzy supervised classification of sub-urban land cover from remotely sensed imagery: statistical and artificial neural network approaches. *International Journal of Remote Sensing*, **22** (4): 615-628.
- ZHU, A. X., 1997, Measuring uncertainty in class assignment for natural resource maps under fuzzy logic. *Photogrammetric Engineering and Remote Sensing*, **63** (10): 1195-1202.
- ZHU, Z., YANG, U., STEHMAN, S. V. and CZAPLEWSKI, R. L., 2000, Accuracy Assessment for the U.S. Geological Survey Regional Land-Cover Mapping Program: New York and New Jersey Region. *Photogrammetric Engineering & Remote Sensing*, **66** (12): 1425-1435.

A N E X O S

Anexo A – Imagem IKONOS

As Imagem IKONOS utilizadas foram as CARTERRA™ GEO obtidas a partir do sensor IKONOS-2 no modo multiespectral (XS), com resolução espacial de 4m. Apresentam-se no quadro A.1 as características das imagens.



Figura A.1- Imagem IKONOS, (RGB 321), do Concelho da Marinha Grande à escala 1:50 000

Quadro A.1 Características da imagem IKONOS.

| | |
|---|-----------------------------|
| Data de aquisição | 2000/09/02 |
| Hora de aquisição | 11h 14 m |
| Ângulo de elevação solar (graus) | 53.37588 |
| Valor nominal da elevação solar (graus) | 81.06776 |
| Ângulo azimutal solar (graus) | 145.0290 |
| Valor nominal do azimute solar (graus) | 50.9100 |
| Banda 1 (azul) | 0.45 μ m - 0.52 μ m |
| Banda 2 (verde) | 0.52 μ m - 0.60 μ m |
| Banda 3 (vermelho) | 0.63 μ m - 0.69 μ m |
| Banda 4 (infravermelho próximo) | 0.76 μ m - 0.90 μ m |
| Resolução espacial | 4m (XS) |
| Resolução radiométrica | 8 bits |
| Dimensão (m x m) | 10928x7460 |

Anexo B – Regras de Classificação do MUP

```

/*#####
/*#
/*#          CLASSIFICA OBJECTOS (REGIÕES) COM BASE          #
/*#          NOS ELEMENTOS DE SUPERFICIE NELES PRESENTES      #
/*#                                                         #
/*#####
/*
/*
/*          DEFINE O CAMINHO DAS IMAGENS
/*
&Workspace e:\Doutoramento\\imagemIkonos\Projecto
grid
/*
/*:~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~:
/*::                                     ::
/*::  CALCULA OS ELEMENTOS DE SUPERFÍCIE DE CADA CLASSE ATRÍBUIDOS  ::
/*::    AOS PIXELS COM UMA INCERTEZA INFERIOR A 0.25               ::
/*::                                     ::
/*:~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~:
/*
/* if ((AgProf >= 0.75) and (incerteza < 0.25)) GPAP = 1
endif
/*
if ((AgPProf >= 0.75) and (incerteza < 0.25)) GPAPP = 1
endif
/*
if ((Anv >= 0.75) and (incerteza < 0.25)) GPANV = 1
endif
/*
if ((Euc >= 0.75) and (incerteza < 0.25)) GPEUC = 1
endif
/*
if ((Somb >= 0.75) and (incerteza < 0.25)) GPSOMB = 1
endif
/*
if ((Herb >= 0.75) and (incerteza < 0.25)) GPHERB = 1
endif
/*
if ((Sobr >= 0.75) and (incerteza < 0.25)) GPSOBR = 1
endif
/*
if ((PM >= 0.75) and (incerteza < 0.25)) GPPM = 1
endif
/*
if ((HEsp >= 0.75) and (incerteza < 0.25)) GPHE = 1
endif
/*
/*:~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~:
/*::                                     ::
/*::          TRANSFORMA OS NULL DAS GRIDS                       ::
/*::          ANTERIORES EM ZERO                                  ::
/*::                                     ::
/*:~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~~::~:
/*
GPAPn = con (isnull (GPAP), 0, GPAP)
/*
GPAPPn = con (isnull (GPAPP), 0, GPAPP)

```

```

/*
GPANVn = con (isnull (GPANV), 0, GPANV)
/*
GPEUCn = con (isnull (GPEUC), 0, GPEUC)
/*
GPSOMBn = con (isnull (GPSOMB), 0, GPSOMB)
/*
GPHERBn = con (isnull (GPHERB), 0, GPHERB)
/*
GPSOBRn = con (isnull (GPSOBR), 0, GPSOBR)
/*
GPPMn = con (isnull (GPPM), 0, GPPM)
/*
GPHEn = con (isnull (GPHE), 0, GPHE)
/*
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::
/*::   CONTA O N° DE ELEMENTOS DE SUPERFÍCIE DE CADA CLASSE   ::
/*::   COM INCERTEZA INFERIOR A 0.25 DENTRO DAS REGIÕES     ::
/*::
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*
/*
ZSAPn = zonalsum (obj1, GPAPn, data)
/*
ZSAPPn = zonalsum (obj1, GPAPPn, data)
/*
ZSANVn = zonalsum (obj1, GPANVn, data)
/*
ZSEUCn = zonalsum (obj1, GPEUCn, data)
/*
ZSSOMBn = zonalsum (obj1, GPSOMBn, data)
/*
ZSHERBn = zonalsum (obj1, GPHERBn, data)
/*
ZSSOBRn = zonalsum (obj1, GPSOBRn, data)
/*
ZSPMn = zonalsum (obj1, GPPMn, data)
/*
ZSHEn = zonalsum (obj1, GPHEn, data)
/*
/*
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::
/*::   CALCULA A % DE ELEMENTOS DE SUPERFÍCIE DE CADA CLASSE   ::
/*::   COM INCERTEZA INFERIOR A 0.25 DENTRO DE CADA REGIÃO   ::
/*::
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*
/*
/*CRIA UMA GRID EM QUE CADA PIXEL ASSUME O VALOR 1
/* E CONTA O N° DE PIXELS EM CADA OBJECTO
/*.....
/*
grid1 = con (isnull (incerteza), 0, 1)
/*
ZSobj = zonalsum (obj1, grid1, data)
/*
/*
/* CALCULA A % DE ELEMENTOS DE SUPERFÍCIE DE CADA CLASSE EM CADA OBJECTO
/*.....
/*

```



```

if (GPHT == 1) HT = 11
ENDIF
/*
/*
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::
/*::      ATRIBUI OS CODIGOS ÀS CLASSES
/*::
/*::
/*::
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*
if (GPANV == 1) ANV = 3
endif
/*
if (GPEUC == 1) EUC = 4
endif
/*
if (GPSOMb == 1) Somb = 5
endif
/*
if (GPSObr == 1) Sobr = 7
endif
/*
if (GPPM == 1) PM = 8
endif
/*
/*
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::
/*::      JUNTA TODAS AS CLASSES ÁGUA E SOMBRA, COM INCERTEZA
/*::      INFERIOR A 0.25, NUM ÚNICO MAPA COM AS SEGUINTE PRIORITYADES
/*::
/*::      10: ÁGUA TOTAL
/*::      8: PINHEIRO MANSO
/*::      4: EUCALIPTO
/*::      7: SOBREIRO
/*::      11: HERBACEA TOTAL
/*::      3: ÁREA NÃO VEGETADA
/*::      5: SOMBRA
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*
/*
I_MV = merge (AGT,EUC,PM,Sobr, HT, ANV, Somb)
/*
/*
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::
/*::      CLASSIFICA OBJECTOS QUE NÃO SÃO FLORESTA
/*::
/*::
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::
/*::      ATRIBUI AOS OBJECTOS "NÃO FLORESTA"
/*::      A UNIDADE DE PAISAGEM CORRESPONDENTE À MODA
/*::      DOS ELEMENTOS DE SUPERFÍCIE
/*::      COM GRAU DE INCERTEZA <0.25
/*::
/*::
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*
setmask NForest
/*
cl = zonalmajority (OBJ1, I_Mv, data)

```

```

/*
/* .....
/* .          CLASSES DE          .          CLASSES          .
/* .  ELEMENTOS DE SUPERFÍCIE .  .  UNIDADE DE PAISAGEM .
/* .....
/*
/*      3- ÁREA NÃO VEGETADA      1- uRBANO/SOLO nU
/*      5- SOMBRA                  2- AGRICULTURA / PASTAGEM
/*     10- ÁGUA TOTAL              3- CORPOS DE ÁGUA
/*     11- HERBÁCEAS TOTAL
/*
/*
/*          .....
/*          .  CORRESPONDÊNCIAS  .
/*          .....
/*
/*
/*URBANO/SOLO NU (1) = ÁREA NÃO VEGETADA (3)
/*
/*AGRICULTURA/PASTAGEM (2) = HERBÁCEAS TOTAL (11)+ SOMBRA (5)
/*
/*CORPOS DE ÁGUA (3) = ÁGUA TOTAL (10)
/*
if (Cl == 3) UrbsN = 1
endif
/*
if (Cl == 11 | Cl == 5) AgrPast = 2
endif
/*
if (Cl == 10) CA = 3
endif
/*
setmask off
/*
/*
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::                                                                    ::
/*::          CLASSIFICA OBJECTOS QUE SÃO FLORESTA                      ::
/*::                                                                    ::
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::                                                                    ::
/*::          ATRIBUI AOS OBJECTOS "FLORESTA"                            ::
/*::A UNIDADE DE PAISAGEM CORRESPONDENTE, CONSIDERANDO UM CONJUNTO     ::
/*::  DE REGRAS APLICADAS SOBRE OS ELEMENTOS DE SUPERFÍCIE            ::
/*::                                                                    ::
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*
/*
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::                                                                    ::
/*::          CALCULA A PERCENTAGEM TOTAL DE HERBÁCEAS EM CADA OBJECTO  ::
/*::                                                                    ::
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*
HT = con (I_mv == 11, 1)
/*
HTN = con (isnull(HT), 0, HT)
/*
ZSHT = zonalsum (OBJ1, HTN, data)
/*
PHT = ZSHT div ZSobj
/*

```



```
/*
SETMASK OFF
/*
/*
/*:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*                                                                           ::
/*    IDENTIFICA ESPAÇOS AGRO FLORESTAIS e FLORESTAS MISTAS          ::
/*                                                                           ::
/*:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*
/*
SETMASK NFPMASK
/*
if ((PCopas <= 0.5) AND (PCSobr >= 0.80) & (pCOPAS < PHT) AgroFl = 7
    else FMISTA2 = 8
endif
/*
SETMASK OFF
/*
/*
/*
/*:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::                                                                           ::
/*::             JUNTA TODAS AS CLASSES NUM ÚNICO MAPA                 ::
/*::             COM AS SEGUINTEs PRIORIDADES                           ::
/*::                                                                           ::
/*::             1:             URBSN      - CLASSE (1)                  ::
/*::             2:             AGRPAST    - CLASSE (2)                  ::
/*::             3:             CA          - CLASSE (3)                  ::
/*::             4:             FEUC       - CLASSE (4)                  ::
/*::             5:             FPM        - CLASSE (5)                  ::
/*::             6:             FSOBR     - CLASSE (6)                  ::
/*::             7:             AgroFl     - CLASSE (7)                  ::
/*::             8:             FMISTA1    - CLASSE (8)                  ::
/*::             9:             FMISTA2    - CLASSE (8)                  ::
/*::                                                                           ::
/*:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*
/*
I_Map = merge (UrbSN, AgrPast, CA, FEUC, FPM, FSOBR, AgroFl, FMISTA1, FMISTA2)
/*
/*
/*:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::                                                                           ::
/*::             SEPARA O MAPA POR CLASSES                               ::
/*::                                                                           ::
/*:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*
/*             CLASSE 1 ->
/*.....
Map_UrbSN = select (I_Map, 'value = 1')
/*
/*             CLASSE 2 ->
/*.....
Map_AgrPt = select (I_Map, 'value = 2')
/*
/*             CLASSE 3 ->
/*.....
MAP_CA = select (I_Map, 'value = 3')
/*
/*             CLASSE 4 ->
```

```
/*.....  
Map_FEUC = select (I_Map, 'value = 4')  
/*  
/*          CLASSE 5 ->  
/*.....  
MAP_FPM = select (I_Map, 'value = 5')  
/*  
/*          CLASSE 6 ->  
/*.....  
MAP_FSobr = select (I_Map, 'value = 6')  
/*  
/*          CLASSE 7 ->  
/*.....  
MAP_AgrFl = select (I_Map, 'value = 7')  
/*  
/*          CLASSE 8 ->  
/*.....  
MAP_FM = select (I_Map, 'value = 8')  
/*
```

Anexo C – Regras de Classificação do MAC

```

/*#####
/*#
/*#  CLASSIFICA O REVESTIMENTO DAS COBERTURAS DOS EDIFICIOS      #
/*#                                COM BASE NOS ELEMENTOS DE SUPERFICIE #
/*#                                NELES PRESENTES                     #
/*#                                                                #
/*#####
/*
/*
/*                                DEFINE O CAMINHO DAS IMAGENS
/*
&Workspace
C:\imagens_IGP\Ortos_Coimbra\Projecto_Baixa\class_2Objrevest_mv\S_lidar
grid
/*
/*
/*:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::
/*::  CALCULA OS ELEMENTOS DE SUPERFÍCIE DOS REVESTIMENTOS DAS   ::
/*::  COBERTURAS DOS EDÍFIÇOS, ATRÍBUÍDOS AOS PIXELS              ::
/*::                                COM UMA INCERTEZA INFERIOR A 0.25  ::
/*::                                                                ::
/*::                                                                ::
/*:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*
/*
/*                                CRIA MÁSCARAS DAS CLASSES
/* .....
if ((TCERCl >= 0.75) and (incerteza < 0.25)) GTCERCL = 1
endif
/*
if ((TCERes >= 0.75) and (incerteza < 0.25)) GTCERes = 1
endif
/*
if ((TCERsomb >= 0.75) and (incerteza < 0.25)) GTCERSomb = 1
endif
/*
if ((Fibro >= 0.75) and (incerteza < 0.25)) GFibro = 1
endif
/*
if ((Chapmt >= 0.75) and (incerteza < 0.25)) GChapmt = 1
endif
/*
/*
/*:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::
/*::                                TRANSFORMA OS NULL DAS GRIDS   ::
/*::                                ANTERIORES EM ZERO                ::
/*::                                                                ::
/*:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*
TCERCLn = con (isnull (GTCERCL), 0, GTCERCL)
/*
TCEResn = con (isnull (GTCERes), 0, GTCERes)
/*
Sombn = con (isnull (GTSomb), 0, GTSomb)
/*
Fibron = con (isnull (GFibro), 0, GFibro)
/*
Chapmtn = con (isnull (GChapmt), 0, GChapmt)
/*

```

```

/*:.....:
/*:::                                     ::
/*:::          CRIA A CLASSE Tcer          ::
/*:::          Tcer = TCERcl + TCERes      ::
/*:::                                     ::
/*:.....:
/*
Tcer = sum (TCERCLn, TCEResn)
/*
/*
/*:.....:
/*:::                                     ::
/*:::          CRIA A CLASSE NTcer        ::
/*:::          NTcer = Fibro + Chapmt     ::
/*:::                                     ::
/*:.....:
/*
NTcer = sum (Fibron, Chapmtn)
/*
/*
/*:.....:
/*:::                                     ::
/*:::  CONTA O N° DE ELEMENTOS DE SUPERFÍCIE DE CADA CLASSE  ::
/*:::          COM UMA INCERTEZA INFERIOR A 0.25              ::
/*:::          DENTRO DOS EDÍFICIOS                          ::
/*:::                                     ::
/*:.....:
/*
/*
ZTCER = zonalsum (g_edificio, Tcer, data)
/*
ZNTCER = zonalsum (g_edificio, NTcer, data)
/*
ZSOMB = zonalsum (g_edificio, Sombn, data)
/*
/*
/*:.....:
/*:::                                     ::
/*:::  CALCULA A % DE ELEMENTOS DE SUPERFÍCIE DE CADA CLASSE  ::
/*:::          DENTRO DE CADA EDIFICIO                          ::
/*:::                                     ::
/*:.....:
/*
/*
/*  CRIA UMA GRID EM QUE CADA PIXEL ASSUME O VALOR 1
/*  E CONTA O N° DE PIXELS DENTRO DE CADA EDIFICIO
/*.....:
/*
grid1 = con (isnull (incerteza), 0, 1)
/*
ZSobj = zonalsum (g_edificio, grid1, data)
/*
/*
/*  CALCULA A % DE ELMENTOS DE SUPERFÍCIE DE CADA CLASSE
/*  DENTRO DE CADA EDÍFICIO
/*.....:
/*
PTCER = ZTCER div ZSobj
/*
PNTCER = ZNTCER div ZSobj
/*

```

```
PSOMB = ZSOMB div ZSobj
/*
/*
/*              REGRA DE CLASSIFICAÇÃO
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::
/*::           CLASSIFICA AS COBERTURAS DOS EDIFÍCIOS COM E SEM      ::
/*::           REVESTIMENTO CERÂMICO                                ::
/*::                                                                 ::
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*
/*
if (PTCER >= 0.50) EDTcer1 = 1
  else if (PNTCER >= 0.50) EDNTcer1 = 2
    else edmask1 = 1
endif
/*
/*
setmask edmask1
/*
if ((PTCER > PNTCER) and (PTCER > PSOMB)) EDTcer2 = 1
  else if ((PNTCER >= PTCER) and (PNTCER > PSOMB)) EDNTcer2 = 2
    else edMask2 = 1
endif
setmask off
/*
/*
setmask edMask2
/*
if ((PTCER > PNTCER) and (PSOMB > 0.5)) EDTcer3 = 1
  else ((PNTCER > PTCER) and (PSOMB > 0.5)) EDNTcer3 = 2

endif
setmask off
/*
/*
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::
/*::           JUNTA TODAS AS CLASSES                                ::
/*::           NUM ÚNICO MAPA                                        ::
/*::                                                                 ::
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*
/*
MAP_ED_Revest = merge (EDTcer1,EDTcer2,EDTcer3,EDNTcer1,EDNTcer2,EDNTcer3)
/*
/*
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*::
/*::           SEPARA O MAPA   POR CLASSES                          ::
/*::                                                                 ::
/*::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
/*
/*
/*           CLASSE 1 ->
/* .....
MapED_TCER = select (MAP_ED_Revest, 'value = 1')
/*
/*
/*           CLASSE 2 ->
/* .....
MapED_NTCEr = select (MAP_ED_Revest, 'value = 2')
/*
```

```
/*#####
/*#
/*#          CLASSIFICA OS EDÍCIOS COM REVESTIMENTO DE          #
/*#          TELHA CERÂMICA COM E SEM ANOMALIAS, COM BASE      #
/*#          NOS ELEMENTOS DE SUPERFÍCIE NELES PRESENTES      #
/*#          #
/*#####
/*
/*          DEFINE O CAMINHO DAS IMAGENS
/*
&Workspace
C:\imagens_IGP\Ortos_Coimbra\Projecto_Baixa\class_2Objpat_mv\S_lidar
grid
/*
/*
/*
/*:::
/*::          CALCULA OS ELEMENTOS DE SUPERFÍCIE DAS COBERTURAS,    ::
/*::          COM E SEM ANOMALIAS ATRIBUÍDOS AOS PIXELS COM      ::
/*::          UMA INCERTEZA INFERIOR A 25%                       ::
/*::          #
/*::
/*:
/*
/*
if ((pattcer >= 0.75) and (incerteza < 0.25)) Gpatcer = 1
endif
/*
if ((Non_pattcer >= 0.75) and (incerteza < 0.25)) GNpatcer = 1
endif
/*
if ((somb >= 0.75) and (incerteza < 0.25)) Gsomb = 1
endif
/*
/*
/*
/*:::
/*::          TRANSFORMA OS NULL DAS GRIDS                      ::
/*::          ANTERIORES EM ZERO                                ::
/*::          #
/*::
/*:
patcern = con (isnull (Gpatcer), 0, Gpatcer)
/*
Npatcern = con (isnull (GNpatcer), 0, GNpatcer)
/*
sombn = con (isnull (Gsomb), 0, Gsomb)
/*
/*:::
/*::          CONTA O N° DE ELEMENTOS DE SUPERFÍCIE DE CADA CLASSE  ::
/*::          COM UMA INCERTEZA INFERIOR A 0.25                ::
/*::          DENTRO DOS EDÍFIOS                               ::
/*::          #
/*::
/*:
/*
Zpatcern = zonalsum (g_edificio, patcern, data)
```

```
/*
ZNpatcern = zonalsum (g_edificio, Npatcern, data)
/*
Zsombn = zonalsum (g_edificio, sombn, data)
/*
/*
/*:~::~:
/*:~:   CALCULA A % DE ELEMENTOS DE SUPERFÍCIE DE CADA CLASSE :~:
/*:~:   DENTRO DE CADA EDIFÍCIO :~:
/*:~:   :~:
/*:~:   :~:
/*:~:~::~:
/*
/*
/*   CRIA UMA GRID EM QUE CADA PIXEL ASSUME O VALOR 1
/*   E CONTA O Nº DE PIXELS DENTRO DE CADA EDIFÍCIO
/*.....
/*
/*
grid1 = con (isnull (incerteza), 0, 1)
/*
ZSobj = zonalsum (g_edificio, grid1, data)
/*
/*
/*   CALCULA A % DE CADA CLASSE EM CADA EDIFÍCIO
/*.....
/*
Pcpat = Zpatcern div ZSobj
/*
Pspat = ZNpatcern div ZSobj
/*
Psomb = Zsombn div ZSobj
/*
/*
/*   REGRAS DE CLASSIFICAÇÃO
/*:~:~::~:
/*:~:   CLASSIFICA AS COBERTURAS DOS EDIFÍCIOS :~:
/*:~:   COM REVESTIMENTO CERÂMICO COM E SEM ANOMALIA :~:
/*:~:   :~:
/*:~:   :~:
/*:~:~::~:
/*
/*
if (Pspat >= 0.50) EDSPAT1 = 2
    else if (Pcpat > 0.50) EDCPAT1 = 1
        else edmask1 = 1
endif
/*
setmask edMask1
/*
if ((Pspat > Pcpat) and (Pspat > Psomb)) EDSPat2 = 2
    else if ((Pcpat > Pspat) and (Pcpat > Psomb)) EDCPat2 = 1
        else edMask2 = 1
endif
setmask off
/*
setmask edMask2
/*
```

```

if ((Pspat > Pcpat) and (Psomb > 0.5)) EDSPAT3 = 2
    else ((Pcpat > Pspat) and (Psomb > 0.5)) EDCPat3 = 1
endif
setmask off
/*
/*
/*:.....
/*:
/*:          JUNTA TODAS AS CLASSES NUM ÚNICO MAPA          ::
/*:
/*:
/*:.....
/*
/*
setmask MapED_TCER
/*
MapEDPat_Tcer = merge (EDSPAT1,EDSPAT2,EDSPAT3,EDCPAT1,EDCPat2,EDCPat3)
/*
setmask off
/*
/*:.....
/*:
/*:          SEPARA O MAPA POR CLASSES          ::
/*:
/*:
/*:.....
/*
/*          CLASSE 1 ->
/*.....
MapED_Cpat = select (MapEDPat_Tcer, 'value = 1')
/*
/*          CLASSE 2 ->
/*.....
MapED_Spat = select (MapEDPat_Tcer, 'value = 2')
/*
/*          REGRAS ESTADO DE CONSERVAÇÃO
/*:.....
/*:
/*: + DE 75% DE ANOMALIA          - EC1          ::
/*: DE 50% A 75% DE ANOMALIA - EC2          ::
/*: DE 25% A 50% DE ANOMALIA - EC3          ::
/*: DE 0% A 25% DE ANOMALIA - EC4          ::
/*: MapED_Spat          - EC5          ::
/*:
/*:.....
/*
/*
setmask MapED_Cpat
if (Pcpat >= 0.75) EC1 = 1
    else if ( (Pcpat >= 0.50) and (Pcpat < 0.75) ) EC2 = 2
        else if ( (Pcpat >= 0.25) and (Pcpat < 0.50) ) EC3 = 3
            else EC4 = 4
        endif
/*
setmask off
/*
EC5 = MapED_Spat
/*
/*

```

```
/*:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::  
/*:::                                                                    :::  
/*:::          JUNTA TODAS AS CLASSES NUM ÚNICO MAPA                    :::  
/*:::                                                                    :::  
/*:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::  
/*  
/*EC_MV = merge (EC1,EC2,EC3,EC4,EC5)
```