



***CARACTERIZAÇÃO DA OCUPAÇÃO DO SOLO
COM RECURSO À APLICAÇÃO DE MODELOS
DE MISTURAS ESPECTRAIS EM SÉRIES
MULTI-TEMPORAIS DE IMAGENS MODIS***

Pedro André de Oliveira

Dissertação apresentada como requisito parcial para obtenção do
grau de Mestre em Ciência e Sistemas de Informação Geográfica

Instituto Superior de Estatística e Gestão de Informação
da Universidade Nova de Lisboa

CARACTERIZAÇÃO DA OCUPAÇÃO DO SOLO COM RECURSO À
APLICAÇÃO DE MODELOS DE MISTURAS ESPECTRAIS EM SÉRIES
MULTI-TEMPORAIS DE IMAGENS MODIS

Dissertação orientada por

Professor Doutor Mário Caetano

Junho de 2005

ao meu filho
à minha mulher
ao meu pai
à minha mãe
e ao meu irmão

AGRADECIMENTOS

Gostaria de expressar o meu mais profundo agradecimento a todos os que tornaram este trabalho possível.

Ao Professor Doutor Mário Caetano, não só pelos seus ensinamentos que permitiram a realização deste trabalho, mas também pela disponibilidade, empenhamento e grande interesse e dedicação com que acompanhou, analisou, discutiu orientou e debateu o mesmo.

Ao Engenheiro António Perdigão, na qualidade de chefe de Divisão de Cartografia e Informação Geográfica do IDRHA, local onde exerço a minha actividade profissional, por todo o interesse e disponibilidade demonstrada para a concretização desta dissertação.

Ao Dr. Paulo Gonçalves do INRIA Rhône-Alpes -França pelo apoio prestado na análise, revisão e discussão dos resultados durante a realização do trabalho. Ao colega Ricardo Armas do ISEGI, pelo apoio prestado na análise das imagens MODIS e VEGETATION, na aprendizagem do software ERDAS e PCI e nas etapas de processamento de dados raster.

À Cláudia, minha mulher e melhor amiga, pela paciência, incentivo, carinho, compreensão, apoio, amizade e muito amor que me deu em todos os momentos. Ao Francisco, meu filho de 8 meses, pela grande alegria que erradia sempre que me vê e que me deu forças para poder concretizar este trabalho. Ao meu irmão, ao meu Pai e à minha Mãe, pelo seu grande amor, carinho, amizade, apoio, compreensão e dedicação. Sem a minha Família, com todo o seu apoio e com TUDO o que me deram e fizeram, esta dissertação nunca teria sido possível, a eles dedico com muito amor.

Muito Obrigado

Pedro André de Oliveira

RESUMO

Realizou-se uma investigação sobre a aplicação do Modelo de Mistura Linear (MML) numa série multi-temporal. Depois de aplicado a uma sequência anual completa de imagens MODIS com 500 metros de resolução especial, o MML gerou um conjunto de imagens fracção com uma variação intra-anual da abundância de componentes puros vegetação, solo e sombra na mistura do pixel.

No pré-processamento das imagens de MOD09A1 utilizou-se uma abordagem do Compósito do Valor Máximo, para obter os dados de base para *input* do modelo de MML. Baseado em 36 imagens fracção, caracterizaram-se 304 amostras, correspondendo cada amostra a uma de onze classes de ocupação do solo definidas neste estudo. Estabeleceu-se uma relação entre estas classes e o CLC2000.

Para a classificação das unidades de ocupação do solo, baseadas na variação do perfil temporal dos componentes puros no pixel das amostras, recorreu-se à utilização de uma rede neuronal de mapas auto-organizados (SOM). A análise da qualidade dos resultados resultou da construção da matriz de convergência, sendo calculados os índices de exatidão global, do produtor e do utilizador para avaliação de resultados.

O trabalho reporta uma análise descritiva dos resultados obtidos segundo a metodologia proposta e apresenta-se como um estudo preliminar para análise de mistura espectrais, numa vertente multi-temporal, para caracterização da ocupação do solo.

ABSTRACT

A multi-temporal extension of Linear Mixture Models (LMM) is investigated. Applied to a full year sequence of MODIS 500 meters resolution images, time-varying Linear Mixture Models provide us with the inter-annual evolution of “vegetation”, “soil” and “shadow” sub-pixel components.

A maximum value composite approach was used for pre-processing MOD09A1 data, to obtain basic data for input in the LMM. Based on the 36 fractions images, 304 sample data was characterized, each corresponding to one of the eleven Land Cover classes defined in this study. A relation of this classes and the CLC2000 was established. For classification land cover units, based on the temporal variation profile of the endmember on the sample's pixel, a Self Organized Maps (SOM) neural network was used. Quality analysis was performed with confusion matrix, and global, user and producer quality precision index was calculated for evaluation of the results.

Reported work is a descriptive analysis of the results and stands as a preliminary study towards automatic multi-temporal LMM based characterization and classification of land cover.

PALAVRAS CHAVE

Detecção remota

Modelo de mistura linear

Cartografia de ocupação e de uso do solo

MODIS

Análise multi-temporal

KEY WORDS

Remote sensing

Linear Mixture Model

Land cover cartography

MODIS

Multi-temporal analysis

ACRÓNIMOS

AVHRR – Advanced Very High Resolution Radiometer
AVIRIS – Airborne Visible/Infrared Imaging Spectrometer
Cdo – Comprimento de onda
CLC2000 –CORINE Land Cover 2000
CNES – Centre National d’Études Spatiales
CNIG – Centro Nacional de Informação Geográfica
COS’90 – Cartografia de Ocupação do Solo 90
CORINE – Co-ordination of Information on the Environment
DN – *Digital Number* (número digital)
EEA – European Environment Agency
EM – Energia electromagnética
ENVISAT – Earth Observation ENVironmental SATellite
ERE – Effective Resolution Element
EROS – Earth Resource Observation Satellite
ERS – European Remote Sensing Satellites
ERTS – Earth Resource Technology Satellite
ESA – Agencia Espacial Europeia (*European Spatial Agency*)
ETM⁺ - Enhanced Thematic Mapper Plus
EUA – Estados Unidos da América
FAO - Food And Agriculture Organization Of The United Nations
GMES – Global Monitoring of Environment and Security
HRV – Haute Résolution dans le Visible
HRV-IR – High Resolution Visible Infrared
IFOV – Instantaneous Field of View
IGBP - International Geosphere–Biosphere Programme
INE – Instituto Nacional de Estatística
INGA – Instituto Nacional de Garantia Agrária
IRS – Indian Remote Satellite

LAC – Local Area Coverage
LMM - Linear Mixture Model
MARS – Monitoring Agriculture by Remote Sensing
MERIS - Medium Resolution Imaging Spectrometer
MML – Modelo de Mistura Linear
MODIS – Moderate Resolution imaging Spectroradiometer
MSS - Multispectral Scanner
NASA – National Aeronautics and Space Administration
NDVI – Normalised Difference Vegetation Index
NNbC - *Neural Network based Cluster* (cluster baseado em redes neurais)
NOAA – National Oceanographic and Atmospheric Administration
PAC – Política Agrícola Comum
PCA – Principle Component Analysis
PNA – Plano Nacional da Água
POES - Polar Operational Environmental Satellite
RBV - Return Beam Vidicon
RGA – Recenseamento Geral Agrícola
SOM – *Self Organizing Maps* (Neural Network based Cluster)
SPOT – Système Probatoire d’Observation de la Terre
TM – Thematic Mapper
UE – União Europeia
UEA – Unidade Espacial de Análise

ÍNDICE DO TEXTO

AGRADECIMENTOS	iv
RESUMO	v
ABSTRACT	vi
PALAVRAS CHAVE	vii
KEY WORDS	vii
ACRÓNIMOS	viii
ÍNDICE DE FIGURAS	xiii
ÍNDICE DE QUADROS	xv
1 – INTRODUÇÃO	1
1.1 – Enquadramento: detecção remota e caracterização da ocupação do solo	1
1.2 – Objectivos	6
1.3 – Organização do trabalho.....	8
2 – SENSORES DE BAIXA E MÉDIA RESOLUÇÃO ESPACIAL: POTENCIALIDADES E LIMITAÇÕES PARA A CARACTERIZAÇÃO DA OCUPAÇÃO DO SOLO	10
2.1 – Breve resenha histórica	10
2.2 – Princípios e fundamentos da detecção remota.....	12
2.3 – Sensores de resolução espacial moderada e baixa	16
2.3.1 - Sensor LANDSAT ETM+	17
2.3.2 - Sensor NOAA-AVHRR.....	19
2.3.3 - Sensor SPOT VEGETATION.....	20
2.3.4 - Sensor MERIS	21
2.3.5 - Sensor MODIS.....	22
2.4 – Métodos de extracção de informação	24
2.4.1 – Sistemas de nomenclatura	25
2.4.1.1 - CORINE Land Cover.....	25
2.4.1.2 - United States Geological Survey	26

2.4.1.3 - International Geosphere–Biosphere Programme	27
2.4.1.4 - FAO Land Cover Classification System	28
2.4.1.5 - AVHRR Global Land Cover Classification.....	29
2.4.2 – Algoritmos de extracção de informação.....	30
2.4.3 – Restrições das abordagens tradicionais na extracção de informação	36
2.4.4 – Análise de Misturas Espectrais.....	37
2.4.4.1 - Componentes Puros Imagem <i>versus</i> Referência.....	39
2.4.4.2 - Linear <i>versus</i> Não linear	41
2.5 – Estudos para a Caracterização da Ocupação do Solo	42
3 - APLICAÇÃO DO MODELO DE MISTURAS LINEARES A SÉRIES MULTI-TEMPORAIS DE IMAGENS MODIS 500	51
3.1 – Base de Dados	52
3.1.1 - Imagens MODIS Reflectância à Superfície.....	53
3.1.2 - CLC2000.....	55
3.2 – Método.....	56
3.2.1 – Pré-processamento	56
3.2.1.1 - Constituição de compósitos semanais	57
3.2.1.2 - Aplicação de uma projecção cartográfica	58
3.2.1.3 - Corte de imagem à fronteira nacional	59
3.2.1.4 - Compósitos Mensais.....	60
3.2.2 – Nomenclatura e amostras para caracterização do solo	62
3.2.2.1 - Nomenclatura.....	62
3.2.2.2 - Selecção de amostras.....	64
3.2.3 – Modelo de Mistura Linear	65
3.2.3.1 - Selecção de Componentes Puros.....	66
3.2.3.2 - Algoritmo do Modelo de Mistura Linear.....	69
3.2.3.3 - Imagens Fracção	71
3.2.3.4 - Erro Médio Quadrático (RMS).....	71
3.2.4 - Caracterização da ocupação do solo	73

3.2.4.1 – Perfis temporais das abundâncias dos componentes puros....	73
3.2.4.1.1 - Extração de informação a partir de imagens fracção	74
3.2.4.1.2 - Caracterização das classes de ocupação do solo	75
3.2.4.2 – Análise de Clusters com algoritmo de redes neuronais (SOM)	76
3.2.5 – Avaliação da qualidade da classificação.....	79
4- RESULTADOS E DISCUSSÃO	81
4.1 – Série mensal de imagens MODIS.....	81
4.2 –Seleção de amostras	86
4.3 – Seleção dos componentes puros.....	91
4.4 – Resultado da aplicação do MML.....	93
4.4.1 – Imagens fracção	93
4.4.2 – Imagens do erro médio quadrático	97
4.5 – Análise de comportamentos não regulares nas amostras	101
4.6 – Perfil Temporal de Classes de Ocupação do solo	102
A - Classe “Massas de água”	103
I - Classe “improdutivo”	105
U – Classe “Zonas Urbanas (contínuas) e Indústria”	106
B – Classe “Pastagens naturais”	108
M – Classe “matos e espécies subarbustivas”	110
P- Classe “florestas resinosas”	111
F – Classe “floresta de folhosas”	113
R – Classe “Agricultura de regadio”	114
S – Classe “Agricultura de sequeiro”	116
Z – Classe “áreas florestais ardidas”	118
H – Classe “cortes e novas plantações”	120
4.7 - Análise clusters (SOM).....	123
5 – CONCLUSÕES	132
REFERÊNCIAS BIBLIOGRÁFICAS	137
ANEXO 1 - Quality Factor e série multi-temporal de MOD09A1	150
ANEXO 2.....	157

ÍNDICE DE FIGURAS

Figura 2.1 - Esquema de rede neuronal de ligações FFNN (adaptado de Malheiro et al., 2004)	34
Figura 2.2 - Contribuição dos componentes puros no <i>pixels</i> mistos (adaptado de Asner, 2000).....	39
Figura 3.1 – Portal de Internet de acesso aos dados MODIS	53
Figura 3.2 – Esquema de ligação das imagens MOD09A1 necessárias para cobrir o território de Portugal (verde – imagem 1; amarelo – imagem 2) (Fonte: NASA,2002).....	57
Figura 3.3 – Limitação da área de estudo através da aplicação de uma máscara	60
Figura 3.4 – Esquema de criação das imagens compósito mensal.....	61
Figura 3.5 – Relação e dependências das Classes de Ocupação do Solo definidas.....	64
Figura 3.6 - Fluxograma do MML (adaptado de Caetano, 2000)	66
Figura 3.6 - Reflectância dos principais elementos da superfície terrestre (Fonseca, 2004).....	67
Figura 4.1 – Pixels não contaminados com nuvens na área de estudo	83
Figura 4.2 - Histogramas da banda do <i>Quality Factor</i> nos compósitos mensais.....	84
Figura 4.3 – Composição colorida RGB214 dos compósitos mensais.....	87
Figura 4.4 - Selecção de amostras em, na malha de 500 por 500 m, com composição colorida RGB 432 de LANDSAT 7 ETM+	88
Figura 4.5 - Selecção de amostras, sobre unidades do CLC2000	89
Figura 4.6 - Distribuição, no território nacional, das amostras	90
Figura 4.7 – Reflectância (%) dos componentes puros seleccionados, nas bandas do MOD09A1.....	92
Figura 4.8 - Composição RGB de imagens de fracção solo, vegetação e sombra na série intra-anual de Fevereiro a Julho de 2000	95

Figura 4.9 - Composição RGB de imagens de fracção solo, vegetação e sombra na série intra-anual de Agosto de 2000 a Janeiro de 2001	96
Figura 4.10 – Imagens de erro médio quadrático de Fevereiro a Julho de 2000	99
Figura 4.11 – Imagens de erro médio quadrático de Agosto de 2000 a Janeiro de 2001	100
Figura 4.12 - Histograma de análise do termo do RMS do erro para a série multi-temporal	101
Figura 4.13 – Perfil temporal médio da fracção de vegetação, solo e sombra para as amostras de treino na classe “massas de água”	104
Figura 4.14 – Perfil temporal médio das fracções vegetação, solo e sombra nas amostras de treino da classe “improdutivo”	106
Figura 4.15– Perfil temporal médio das amostras da classe “urbano”.	107
Figura 4.16 – Perfil temporal médio das amostras da classe “pastagens pobres”	109
Figura 4.17 – Perfil das amostras da classe matos e espécies subarbustivas	111
Figura 4.18 – Perfil temporal médio da classe “florestas resinosas”.	112
Figura 4.19 - Perfil temporal médio da classe “floresta folhosa”	114
Figura 4.20 – Perfil temporal médio da classe “regadio”	115
Figura 4.21 – Perfil temporal médio da classe “sequeiro”.	117
Figura 4.22 – Perfis temporais das amostras 9, 27 e 29 na classe “áreas ardidas”	119
Figura 4.23 – Perfis temporais da classe cortes e novas plantações	121

ÍNDICE DE QUADROS

Quadro 2.1 - Características do ETM ⁺ (fonte: NASA, 1999)	18
Quadro 2.2 - Características do sensor AVHRR, instalado a bordo do NOAA-15 (fonte: Cracknel, 1997)	19
Quadro 2.3 – Características do VEGETATION (fonte: CNES, 2000)	21
Quadro 3.1 – Resolução espectral e espacial de MOD09A1	54
Quadro 3.1 – Parâmetros para projecção das imagens MOD09A1	59
Quadro 3.2 - Relação entre as classes da nomenclatura eleita e as classes do CLC2000	63
Quadro 4.1 – Pixels contaminados com nuvens nos compósitos mensais	82
Quadro 4.2 – <i>Quality Factor</i> : Conversão de DN para binário	84
Quadro 4.3 -Reflectância espectral (DN e %) dos componentes puros seleccionados	91
Quadro 4.4 – Resumo de análise de clustering	102
Quadro 4.5. – Valores médios (m) e desvio padrão (std) das imagens fracção na classe “massas de água”	104
Quadro 4.6 – Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “improdutivo”.	105
Quadro 4.7 – Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “urbano”.	107
Quadro 4.8 – Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “pastagens pobres”.	109
Quadro 4.9 – Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “matos e espécies subarbustivas”	110
Quadro 4.10– Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “florestas resinosas”	112
Quadro 4.11 – Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “floresta de folhosas”	113

Quadro 4.12 – Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “regadio”	115
Quadro 4.13 – Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “sequeiro”	117
Quadro 4.14 – Análise de clusters, das amostras, com produção de 28 aglomerados	124
Quadro 4.15 – Matriz de confusão e índices de qualidade da análise SOM com 28 clusters	125
Quadro 4.16 – Análise de clusters, das amostras, com produção de 33 aglomerados	126
Quadro 4.17 – Matriz de confusão e índices de qualidade da análise SOM com 33 clusters	126
Quadro 4.18 – Análise de clusters, das amostras, com produção de 36 aglomerados	127
Quadro 4.19 – Matriz confusão e índices de qualidade da análise SOM com 36 clusters	128
Quadro 4.20 – Análise de clusters, das amostras, com produção de 36 aglomerados	129
Quadro 4.21 – Matriz de confusão e índices de qualidade da análise SOM com 36 clusters	129
Quadro 4.22 – Matriz de confusão e índices de qualidade da análise SOM para 3 datas	130
Quadro 4.23 – Matriz de confusão e índices de qualidade da análise SOM para 1 data	131

1 – INTRODUÇÃO

1.1 – Enquadramento: detecção remota e caracterização da ocupação do solo

Os organismos responsáveis pela definição e aplicação de políticas e/ou directivas de ordenamento e planeamento do território carecem de dados sobre a estrutura da superfície terrestre, com uma periodicidade elevada e com carácter actualizado, para a fundamentação e suporte das suas tomadas de posições.

Numa época onde a utilização racional e sustentável dos recursos naturais está na ordem do dia, é importante dispor de informação que traduza a estrutura e a forma como estes recursos estão disponíveis. Conservar o território e disciplinar as actividades humanas, é uma tarefa que resulta do conhecimento da situação actual e de uma definição de linhas estratégicas para a regulamentação dos diferentes sectores de actividade que interagem, directa ou indirectamente, com as diferentes unidades da paisagem.

A caracterização do território, de acordo com a sua dimensão funcional ou utilização socio-económica, presente e futura, permite definir diferentes unidades de ocupação do solo. Estas unidades ou classes (e.g., zonas agrícolas, florestais, (semi)naturais, húmidas e massas de água) resultam de diferentes tipos de utilização do solo (e.g., residencial, industrial, recreativo, florestal, agrícola) (CCE, 2004), a qual é muitas vezes condicionada por alguns fenómenos naturais (e.g., fogos e inundações) ou humanos (e.g., desmatção, reflorestação, urbanização, inundações) e consequentemente conduz a alterações das respectivas classes de ocupação entretanto identificadas, para uma dada posição geográfica. Estas alterações permitem-nos afirmar que as classes de ocupação do solo apenas são válidas durante um dado período de tempo.

A identificação das classes de ocupação do solo, num dado período e para uma localização espacial, resulta essencialmente da aplicação de um ou vários métodos científicos e/ou estatísticos para a classificação dos dados/informação de base utilizados (JARS, 1996).

Diferentes abordagens podem ser utilizadas na recolha dos dados de base para a caracterização da ocupação do solo. O recurso a inventários e trabalhos de campo, recenseamentos estatísticos gerais e agrícolas, fotografias aéreas ou imagens de satélites são as principais formas para a aquisição de dados sobre a paisagem. Mediante a estratégia escolhida, as unidades de ocupação de solo podem variar em número, forma e contexto.

A nomenclatura, o número de classe de ocupação do solo e a unidade mínima cartografada condicionam a escala de publicação dos dados. Estes últimos podem ser produzidos para uma escala global, regional ou local, consoante o fim em vista para a sua utilização futura. Independentemente da abrangência dos dados, estes podem ser encarados na sua vertente de informação história ou actualizada.

A publicação de dados para caracterização da ocupação do solo ao nível global ou nacional está associada à cartografia de escalas pequenas, com cobertura de grandes áreas (continentes). A nomenclatura apresenta classes de índole generalista e esta informação apresenta um grau de actualização variável. É um tipo de dados bastante utilizado como *input* para modelações globais, na detecção de alterações nas unidades da paisagem, para a extracção de informação estatística para caracterização do território ou ainda como base de estudos com vista à definição de políticas de âmbito global.

A caracterização de ocupação do solo de carácter local resulta de dados cuja escala base apresenta um grande detalhe, ou seja, grandes escalas. Este tipo de dados tem normalmente associada uma nomenclatura pormenorizada com um elevado número de classes. O seu grau de

actualização varia consoante se trate de caracterização do solo em áreas urbanas ou rurais, estando esta dependente da disponibilidade de informação de base e do custo de execução. Estes dados podem ser utilizados com fins de planeamento urbano, gestão de recursos naturais, avaliação de impactos ambientais ou operações de emergência.

A detecção remota, é uma fonte actualizada de dados, e portanto adequada para a caracterização da ocupação do solo. Para áreas relativamente grandes, as imagens de satélite apresentam, face a outros dados de base, vantagens relativamente ao preço de aquisição, à frequência de disponibilidade de dados e ao detalhe. Ainda quando comparado com outros dados, as imagens de satélite apresentam um custo reduzido relativamente ao conjunto total dos processos para a aquisição, processamento, análise e publicação de resultados. Esta vantagem competitiva resulta, entre vários factores, do facto das imagens de satélite poderem ser submetidas a métodos de classificação automáticos, reduzindo a morosidade do processo, os encargos com pessoal e a complexidade da cadeia de produção.

A disponibilidade actual de sensores e as suas características ao nível da resolução espacial e temporal, permitem caracterizar a ocupação do solo com uma periodicidade e um detalhe que se adequa e responde às necessidades dos utilizadores.

O maior ou menor detalhe da caracterização da ocupação do solo (características e número das unidades taxonómicas) depende das resoluções espaciais, espectrais e temporais utilizadas. Em função das duas primeiras assim a publicação de documentos cartográficos de escala generalizada ou pormenorizada.

Os sensores LANDSAT e *Système Probatoire d'Observation de la Terre - Haute Résolution dans le Visible* (SPOT-HRV) apresentam uma resolução espacial grande, adequada para a caracterização da ocupação do solo ao nível regional e têm sido amplamente utilizados em diversos

projectos e estudos por todo o globo, como é exemplo a iniciativa do CORINE Land Cover da Agencia Europeia do Ambiente da União Europeia (Instituto do Ambiente, 2005). O aparecimento dos sensores Quickbird e IKONUS, de muito grande resolução espacial, revolucionou a capacidade de discriminação dos diferentes objectos da superfície terrestre. Quer para estes últimos sensores, quer ainda para os primeiros, a resolução temporal e os custos de aquisição dos seus dados são, em muitos dos casos, uma das restrições principais à sua utilização.

O sensor *Advanced Very High Resolution Radiometer* (AVHRR), da série *National Oceanographic and Atmospheric Administration* (NOAA), foi largamente utilizado para a caracterização da vegetação devido à sua elevada resolução temporal e à resolução espectral disponível (equipado com sensores espectrais nas bandas do vermelho e infravermelho próximo que permitem analisar o vigor da vegetação). Apesar do AVHRR ter uma baixa resolução espacial, quando explorada em simultâneo na sua vertente multi-temporal, mostrou ser um sensor com vantagens para a identificação e acompanhamento de diferentes tipos de vegetação. Ainda devido à versatilidade deste sensor, foram produzidos, a partir dos seus dados, alguns produtos para a caracterização da ocupação do solo a nível global (e.g., *AVHRR Global Land Cover Classification*) (Hansen, 2000).

Mais recentemente, também a Agência Espacial Europeia (ESA) desenvolveu um sensor de baixa resolução espacial - o SPOT VEGETATION, cujo objectivo principal visa a caracterização, acompanhamento e monitorização, ao nível global, da cobertura do solo/recursos naturais e, em particular, da vegetação.

No virar do século XX surgiram os sensores multi-espectrais *Moderate Resolution imaging Spectroradiometer* (MODIS) e *Medium Resolution Imaging Spectrometer* (MERIS), com uma resolução espacial moderada, e com resolução temporal muito elevada (1 a 3 dias). Estas duas

características permitem acompanhar, com elevada periodicidade e com um detalhe acentuado, os fenómenos terrestres globais e regionais.

O MERIS disponibiliza informação em 15 bandas espectrais, todas com resolução espacial de 300 m, mas a política adoptada para a distribuição das suas imagens dificulta a sua utilização. No MODIS estão disponíveis 32 bandas espectrais, cuja resolução espacial varia entre os 250 m e 1 km. O carácter público e gratuito como forma de distribuição das imagens do MODIS, diversificou a sua utilização e áreas de aplicação.

O MODIS e o MERIS são assim um grupo de sensores, com custos moderados ou mesmo nulos, com um potencial de utilização muito elevado e que permitem monitorizar os ciclos naturais dos diferentes componentes atmosféricos, terrestres e dos oceanos.

A disponibilização gratuita das imagens e as características em termos de resolução espectral, espacial e temporal, faz dos produtos MODIS um elemento de trabalho adequado para analisar, no âmbito de estudos académicos e/ou de baixos recursos financeiros, o comportamento dos objectos da superfície terrestre. Neste ultimo contexto os produtos MODIS foram seleccionados como dados de base para a realização deste trabalho de Dissertação de Mestrado.

O sensor MODIS apenas disponibiliza dados desde Fevereiro de 2000, e embora a sua aquisição seja desprovida de custos, o número de estudos efectuados com estes dados, para a caracterização da ocupação do solo, têm sido em número reduzido. Diferentes metodologias foram aplicadas aos dados MODIS para estudar e caracterizar a superfície terrestre, nomeadamente a aplicação de algoritmos de classificação assistida (Törmäa, 2004; Chuvieco et al, 2004), ou a exploração temporal dos índices de vegetação com classificadores não assistidos (Ferreira et al, 2003).

Uma das potencialidades para a exploração de dados MODIS para a caracterização da ocupação do solo é a de análise de séries temporais tal como o tem sido para as imagens do sensor de baixa resolução espacial

AVHRR (e.g., Pan, 2003; Giri, 2000). Tentando aproveitar a resolução moderada do MODIS e conciliando com a sua elevada resolução temporal, alguns estudos utilizam estes dados para caracterizar a vegetação, com base nas variações inter-anuais do índice de vegetação, ao longo do período de desenvolvimento das cultura (Ferreira, 2005; Matshuoka, 2004). A caracterização da ocupação do solo a partir de imagens de satélite apresenta, assim, bastantes vantagens, embora possa ser comprometida pelo facto de diferentes objectos da superfície de solos apresentarem assinaturas espectrais semelhantes, num determinado comprimento de onda (cdo) e num dado momento, ou pelo facto de algumas classes de uso do solo/ocupações de solo não serem identificadas ao nível do *pixel*. Verifica-se, então, uma mistura de informação espectral, dos vários objectos da superfície, ao nível do pixel da imagem de satélite, originando um pixel misto. Este constrangimento poderá ser ultrapassado ao associar algoritmos de análise da composição/contribuição espectral ao nível do sub-pixel com análise de séries multi-temporais de dados.

A associação destas duas técnicas (análise multi-temporal e análise de misturas espectrais) permitirá que o problema do *pixel* misto e da resposta espectral semelhante possam ser ultrapassados e consequentemente contribuirá para uma melhor discriminação dos objectos da superfície. Com base neste pressupostos define-se uma metodologia que permite a exploração de imagens de resolução espacial moderada e possibilita a discriminação e caracterização de classes de ocupação do solo com uma dinâmica temporal, ou seja, com características espectrais que apresentam variações temporais.

1.2 – Objectivos

O estudo realizado no âmbito da Dissertação de Tese de Mestrado de Ciência & Sistemas de Informação Geográfica teve como principal objectivo o desenvolvimento de uma metodologia para a exploração de

uma série intra-anual de imagens do sensor *Moderate Resolution Imaging Spectroradiometer* (MODIS), através de um processo bi-etápico, como forma de caracterização da ocupação do solo. Para atingir este objectivo recorreu-se, numa primeira fase, à aplicação do Modelo de Mistura Linear (MML) para determinação da contribuição espectral dos componentes puros - solo, vegetação e sombra - nas imagens multi-espectrais; numa segunda fase, procedeu-se à análise das abundâncias dos componentes puros na série anual de imagens fracção, com vista à caracterização de classes de ocupação de solo. Este último passo permitiu identificar padrões (comportamentos típicos), resultantes da variação temporal das características das classes, em função dos componentes puros estudados.

Após a extracção da informação das imagens MODIS pela aplicação do MML recorreu-se à utilização de um classificador não assistido, baseado numa rede neuronal de *clustering* (*Self Organized Maps*) com o objectivo de identificar as classes de ocupação do solo, utilizando a informação derivada das bandas espectrais, i.e, as abundâncias dos componentes puros - vegetação, solo e sombra- ao longo dos 12 meses. A avaliação da qualidade dos resultados obtidos baseou-se na matriz de contingência (matriz de erro) com o cálculo da precisão global, do utilizador e do produtor.

Para a realização deste trabalho utilizou-se o produto de nível 3 do MODIS, imagens de compósitos de 8 dias com reflectância à superfície (MOD09A1). Este produto apresenta 7 bandas espectrais, todas com uma resolução espacial de 500 m. Embora o sensor MODIS apresente um produto com melhor resolução espacial (250 m) este só disponibiliza 2 bandas espectrais (vermelho e infravermelho próximo). Utilizando-se um maior número de bandas espectrais, a resolução do modelo de mistura linear, possibilita a definição de um maior número de componentes puros para modelar a mistura espectral do pixel, razão pela qual se optou por um produto com menor resolução espacial mas com maior resolução espectral.

1.3 – Organização do trabalho

A presente Tese encontra-se estruturada em cinco capítulos, sendo o conteúdo de cada um deles descrito em seguida de forma sumária.

Neste primeiro capítulo de “Introdução” faz-se uma breve abordagem à caracterização da ocupação do solo, à importância da detecção remota como fonte de dados base, aos diferentes sensores, suas características e aplicações, e à definição dos critérios para a selecção do sensor a utilizar neste estudo. Definiram-se os objectivos do trabalho, bem como, a sua relação e enquadramento com outros trabalhos de detecção remota.

O enquadramento do trabalho proposto é expresso no segundo capítulo. Fruto de uma revisão bibliográfica, apresenta-se uma breve resenha histórica da detecção remota, os seus fundamentos teóricos, os principais projectos espaciais e alguns dos sensores de baixa e média resolução espacial. Abordam-se pontos relacionados com a caracterização da ocupação do solo, sensores utilizados para este fim, escalas de produção, nomenclaturas desenvolvidas e métodos de classificação das imagens. Revêem-se, também, os fundamentos teóricos do MML para a caracterização da ocupação do solo e a problemática da multi-temporalidade neste estudos. Por fim, analisaram-se problemas/questões colocados por diferentes autores na variabilidade espacial, espectral e temporal das imagens de satélite.

O terceiro capítulo diz respeito aos “Métodos”. Descreve-se a metodologia de pré-processamento dos dados, os pressupostos para a aplicação do MML e os critérios de selecção dos componentes puros. Propõe-se uma metodologia para a exploração das imagens fracção obtidas no MML e para caracterização de classes de ocupação do solo com dinâmica temporal. Por fim, propõe-se a exploração dos resultados do MML através da aplicação de um classificador não assistido baseado numa rede neuronal com vista a uma avaliação global dos resultados.

No quarto capítulo expõem-se os resultados da aplicação da metodologia. Discute-se o tratamento e análise temporal do resultado do pré-processamento, dos componentes puros seleccionados e das imagens fracção obtidas como fruto da aplicação do MML aos compósitos mensais de MOD09A1. Caracterizam-se os padrões temporais determinados para as classes de ocupação do solo propostas, e apresentam-se os resultados de validação para a metodologia proposta. São ainda expressas algumas considerações finais.

O quinto e último capítulo é constituído pelas conclusões e reflexões finais da Dissertação .

A estruturação da Tese foi, assim, elaborada de modo a dar encadeamento aos assuntos expostos, de forma a obter-se um melhor entendimento na abordagem dos mesmos e das diferentes fases constituintes do trabalho.

2 – SENSORES DE BAIXA E MÉDIA RESOLUÇÃO ESPACIAL: POTENCIALIDADES E LIMITAÇÕES PARA A CARACTERIZAÇÃO DA OCUPAÇÃO DO SOLO

Neste capítulo explicam-se sumariamente os princípios e os conceitos associados à caracterização de ocupação do solo. Faz-se uma breve resenha histórica sobre a detecção remota, descrevem-se os seus fundamentos teóricos, e apresentam-se as características dos principais satélites e sensores de média e baixa resolução espacial. Expõem-se os métodos utilizados para a extracção de informação das imagens de satélite, de média e baixa resolução, e os sistemas de nomenclatura, para caracterizar a ocupação do solo, mais difundidos. Apresenta-se os fundamentos do Modelo de Mistura Linear e por fim aborda-se a exploração multi-temporal de imagens de média e baixa resolução espacial, bem como, a respectiva aplicação do Modelo de Mistura Linear a estes dados.

2.1 – Breve resenha histórica

Desde muito cedo o Homem sentiu o desejo de conhecer o ambiente que o envolve nas suas actividades de forma a modelá-lo às suas necessidades. Com a evolução da espécie humana, os cinco sentidos foram-se adaptando e aperfeiçoando com o intuito de identificar, diferenciar e apreender informação sobre as entidades e dos fenómenos que ocorrem no mundo em seu redor. O Homem, através do engenho e da arte, desenvolveu ferramentas que permitiram complementar suas faculdades sensoriais.

A observação e registo dos diferentes elementos da superfície terrestre continua a ser uma disciplina onde o Homem despende tempo e recursos para melhor analisar e interpretar os vários fenómenos. A utilização dos olhos como sensor, para adquirir informação sobre a paisagem circundante, é a forma mais simples que está ao dispor do Homem. O registo do que os seus olhos deslumbram sobre diferentes

suportes (e.g., pedra, papiro, tela, mosaico) foi, até há pouco tempo, uma das formas de perpetuar, retratar e expressar o que se passa no globo.

A realização de descobertas distintas, nos campos da óptica e da química, propiciaram o aparecimento da fotografia no séc. XIX, um facto histórico, que contribuiu para o aparecimento da ciência que mais tarde veio a ser conhecida como Detecção Remota. Um passo importante ocorreu em França, quando, durante um voo de demonstração pelo piloto Wilbur Wright, o passageiro L.P. Bonvillain fotografou a superfície terrestre. Posteriormente, esse mesmo piloto, efectuou outro voo em Itália, onde realizou a primeira sequência de fotografias aéreas (Barata, 2001).

Os vários episódios das grandes guerras mundiais e o aparecimento dos diferentes programas espaciais levaram ao desenvolvimento de tecnologias para aquisição de imagens/fotográficas inseridas em plataformas espaciais ou aerotransportadas, quer para objectivos de estratégia militar, quer na monitorização dos diferentes ecossistemas terrestres ou interplanetários.

Esta é a época áurea da detecção remota, onde a indústria espacial foi, ao longo das últimas quatro décadas, um motor constante na criação e lançamento de diversos satélites e plataformas espaciais equipadas com sensores para a monitorização da esfera celeste, da superfície e do subsolo terrestre.

Muito embora tenham sido os russos a dar início à aventura espacial, com o lançamento, em 1957, do primeiro satélite – *Sputnik*; foram os americanos, através da *National Aeronautics and Space Administration* (NASA) e do Departamento de Interior dos Estados Unidos da América (EUA), que mais contribuíram para a monitorização da superfície terrestre. A NASA desenvolveu o programa *Earth Resources Observation Satellite* (EROS), cujo principal objectivo era reunir e recolher dados dos recursos da superfície terrestre, com base em instrumentos colocados a bordo de satélites para observação da Terra. O início deste programa espacial que,

mais tarde, ficou designado de LANDSAT, deu-se com o lançamento do primeiro de sete satélites, o *Earth Resources Technology Satellite* ERTS 1, em Julho de 1972. Este é o programa mais antigo que ainda hoje está operacional.

Quer a NASA, quer a Agência Espacial Europeia (ESA), quer ainda outras agências espaciais espalhadas pelo globo, têm definido vários programas espaciais para o desenvolvimento de sensores com fins concretos, dedicados, nomeadamente, à monitorização dos recursos naturais e em particular à inventariação, acompanhamento e detecção das alterações da ocupação do solo.

2.2 – Princípios e fundamentos da detecção remota

Ao longo das 24 horas, dos 365 dias do ano, os satélites ao descreverem as suas órbitas, observam a superfície terrestre, comportando-se como uma fonte permanente e remota de informação. Segundo Horler e Barber (1981) citado por Caetano (2002), a detecção remota pode ser definida como o processo para obter informação sobre um objecto sem estar em contacto directo com ele, e inclui medições de energia eléctrica, magnética, electromagnética e vibracional (acústica).

Um sistema de detecção remota é composto, na sua essência, por um fenómeno de estudo (objecto), por um fluxo de energia electromagnética (EM) e por um sensor concebido especificamente para captar e registar o fluxo de energia proveniente do objecto. Este sistema, embora simples, pode apresentar alguma complexidade, pois a interacção da radiação com os objectos depende do seu comprimento de onda (cdo), das propriedades do objecto e das interacções entre os dois.

De acordo com o Princípio de Conservação de Energia a radiação incidente num dado elemento/objecto pode dar origem a três reacções: penetrar no objecto e ser absorvida por ele (radiação absorvida), passar através do objecto (radiação transmitida) ou ainda retornar do objecto

depois de interagir com ele (radiação reflectida). As diferentes percentagens de cada uma delas permitem caracterizar o objecto, mas apenas a radiação emitida ou reflectida pelo objecto pode ser captada e registada pelos sensores dos diferentes satélites. À quantidade de radiação incidente que é reflectida designa-se de reflectância espectral.

O sol é a grande fonte emissora de energia luminosa, a qual se faz na forma de radiação electromagnética. A teoria corpuscular e a teoria ondulatória dão um fundamento científico à radiação electromagnética, a qual, por definição, é constituída por fotões que se deslocam à velocidade da luz num movimento ondulatório (Bird,1991).

O espectro electromagnético cobre uma gama de valores, de frequência das ondas electromagnéticas, que vão desde as ondas rádio (alto cdo e baixa energia) até aos raios gama (baixo cdo e alta energia), sendo que os nossos olhos apenas captam uma parte muito pequena do espectro que corresponde à zona da banda do visível (0,38 – 0,72 μm).

A quantidade de energia que é emitida por um objecto depende da sua temperatura, sendo esta relação definida pela Lei de Stefem-Bolzman. O quociente entre a composição espectral emitida por um objecto e a sua temperatura é estabelecida pela Lei de Wien. De acordo com estas leis o sol comporta-se como um corpo negro, emitindo em quase todos os comprimentos de onda, sendo que 99% daquela radiação é emitida nas bandas do ultravioleta, visível e infravermelho. Estas são também as bandas com maior interesse para a detecção remota.

A intensidade, a energia e a composição espectral da radiação reflectida pelos elementos da superfície da terra que chega aos sensores é condicionada pela atmosfera. Ao atravessá-la, a energia EM sofre dois tipos de interacções: dispersão ou absorção. O primeiro caso ocorre quando gases e partículas existentes na atmosfera provocam a dispersão da radiação. A absorção atmosférica resulta numa perda de energia para os constituintes da atmosfera (absorvem energia). As zonas do espectro

onde não ocorre absorção, por parte dos constituintes da atmosfera, designam-se de janelas atmosféricas (Bird, 1991 a).

A energia registada pelos sensores dos satélites traduz as características espectrais da superfície e da atmosfera. Constituintes com comportamentos iguais num cdo poderão ser diferentes noutro e, portanto, observáveis distintamente num sensor com várias bandas espectrais (multi-espectrais).

Em função da altitude, orientação e rotação relativamente à Terra, ou seja, em relação à órbita, os satélites classificam-se de geo-estacionários ou não geo-estacionários. Os primeiros observam sempre a mesma área da Terra; os segundos não se mantêm fixos sobre o mesmo ponto da superfície terrestre e deslocam-se sobre um plano que forma um determinado ângulo com o equador. Nestes últimos, quanto mais baixa for a órbita maior será a velocidade e menor será o tempo necessário para o satélite dar a volta completa à Terra e voltar a passar no mesmo ponto. Para o acompanhamento de processos globais e dinâmicos a opção correcta é a escolha de satélites geo-estacionários, contudo, no caso de se pretender um detalhe elevado da superfície deve optar-se por satélites não geo-estacionários e de órbita polar ou quase polar, pois deslocam-se num plano perpendicular ao equador e podem obter imagens que cobrem praticamente toda a Terra durante o seu ciclo (Caetano, 2002).

Através de modelos de calibração interna do sensor, a quantidade de energia reflectida a partir de uma dada área da superfície e coberta por cada pixel, é convertida em números digitais (DN). Para as diferentes bandas espectrais do sensor é gerada uma matriz com a informação radiométrica de cada pixel, dando origem a uma imagem multi-espectral. Alguns sensores, designados de pancromáticos, recolhem dados de radiância apenas numa zona do espectro electromagnético, não subdividindo a radiação nas suas componentes espectrais (JARS, 1996).

Uma das especificações dum sensor é a resolução. Segundo Jensen (1986) e citado em Cartalis (2000), existem quatro tipos de resoluções: resolução espectral, resolução radiométrica, resolução temporal e resolução espacial.

Por resolução espectral entende-se a dimensão e o número de intervalos do espectro electromagnético possíveis de serem registados por um dado sensor. Quanto mais bandas existirem e menor forem os seus intervalos melhor será a discriminação dos objectos e da sua envolvente.

A resolução radiométrica define a sensibilidade que o sensor tem para diferenciar a intensidade de energia EM captada num dado comprimento de onda, seja ela reflectida ou emitida. Esta grandeza define o número de níveis digitais que o sensor tem disponível para exprimir os dados recolhidos. As imagens são então representadas por um DN positivo, que varia de 0 a uma determinada potência de 2 menos 1 (e.g., numa imagem de 8 bits existem 256 níveis de resolução e o DN varia de 0 a 255).

O espaço de tempo que medeia duas aquisições sucessivas, de imagens sobre a mesma área e com o mesmo ângulo de observação do local, designa-se de resolução temporal. Se apenas se exigir que as imagens sejam do mesmo local, sem considerar o ângulo e a frequência que decorre entre as aquisições, qualifica-se de resolução temporal relativa.

A resolução espacial define-se como o menor ângulo, ou a distância mínima entre dois objectos, que pode ser diferenciada pelo sensor. Um dos modos de a definir é através do campo de visão instantâneo (IFOV – *Instantaneous Field of View*) do sensor, o qual representa a dimensão da parcela de terreno que é captada pelo sensor, à medida que este se vai deslocando ao longo da sua órbita e adquirindo informação da superfície terrestre. Contudo, a resolução espacial está, também, relacionada com a dimensão dos próprios objectos. Para que o objecto se possa destacar tem que ter, pelo menos, o dobro da dimensão do IFOV, na sua menor

dimensão, e terá também que possuir uma resposta espectral diferente dos elementos que o rodeiam, caso contrário não será possível diferenciá-lo. Resultou então a ideia de elemento efectivo de resolução (ERE – *Effective Resolution Element*), que se define como a dimensão da área de terreno na qual pode ser atribuído um valor único de radiância, com intervalo de confiança de 5% em relação à radiância relativa efectiva (Baio, 1996).

2.3 – Sensores de resolução espacial moderada e baixa

A caracterização da ocupação do solo é um dos objectivos de alguns dos programas espaciais actualmente em funcionamento. Em variados projectos, levados a cabo pela NASA e a ESA, conceberam-se sensores para esse fim, sendo de destacar, entre eles, os de moderada (média) e baixa resolução espacial e que têm tido uma ampla utilização.

O LANDSAT não se enquadra, actualmente, nesta categoria de sensores (pertence aos sensores de alta resolução espacial), mas é, talvez, o mais utilizado nos estudos de caracterização da ocupação do solo. Foram, também, utilizadas imagens deste sensor, como informação auxiliar na realização deste trabalho de Dissertação, pelo que a sua caracterização se torna pertinente.

Situados entre os sensores de alta resolução (e.g., LANDSAT e SPOT) e os sensores de baixa resolução (e.g., NOAA-AVHRR e VEGETATION), o MODIS e o MERIS adequam-se ao estudo dos recursos naturais e dos ciclos biológicos. O MERIS e MODIS, apresentam, assim, uma resolução espacial moderada ou média, sendo no primeiro de 300 m em todas as bandas, e no segundo varia, consoante a banda, entre os 250 m e os 1000 m.

Ao nível da resolução espectral, estes dois sensores, de resolução espacial moderada, têm a capacidade de registar entre 15 e 32 cdo diferentes, colocando-os na transição entre os sensores multi e hiper-espectrais.

As diferentes estratégias adoptadas, pelas Agências Espaciais, na distribuição dos produtos obtidos dos diferentes sensores propicia ou restringe a sua utilização em estudos e projectos científicos.

As imagens dos sensores de resolução espacial moderada têm uma resolução temporal e espectral elevada, apresentando custos de aquisição relativamente baixos ou mesmo nulos. Estes sensores têm um potencial tecnológico elevado e apresentam bastantes subprodutos. A caracterização intra-anual da ocupação do solo pode ser obtida a partir do processamento de uma série temporal de imagens de satélite, as quais sendo de média resolução espacial tem um custo (imagens + processamento + tempo) relativamente baixo comparativamente às de alta resolução espacial. Com base nestes pontos elegeu-se os sensores de resolução espacial moderada (MODIS e MERIS) como os mais adequados a este estudo.

Seguidamente apresenta-se, resumidamente, as características dos satélites de baixa e média resolução espacial considerados relevantes para a Dissertação.

2.3.1 - Sensor LANDSAT ETM+

Os sensores do LANDSAT são actualmente classificados como de alta resolução espacial. Resulta de um projecto espacial de observação da terra e dos recursos naturais que teve início em 1962. A *National Aeronautics and Space Administration* (NASA), a *National Oceanic and Atmospheric Administration* (NOAA) e o *United States Geological Survey* (USGS) gerem actualmente o programa LANDSAT e, devido às suas características, é o sensor mais utilizado para caracterizar a ocupação do solo.

O primeiro satélite foi lançado em 1972, e desde então, existe pelo menos um LANDSAT em órbita que recolhe imagens ininterruptamente, tornando-o o programa mais antigo e com maior número de imagens

recolhidas ainda em operação, no grupo dos satélites de observação da Terra.

Inicialmente denominado ERTS (*Earth Resources Technology Satellite*), o LANDSAT possuía dois sensores de média resolução espacial: a câmara RBV (*Return Beam Vidicon*) e o MSS (*Multispectral Scanner*). Destaca-se o 5º LANDSAT por ter a bordo os sensores MSS e *Thematic Mapper* (TM), sendo este último de alta resolução espacial. O sensor LANDSAT 6, devido a problemas técnicos nunca chegou a adquirir imagens, sendo lançado o LANDSAT 7. Este último está equipado com o *Enhanced Thematic Mapper Plus* (ETM⁺) e o *Advanced Landsat Sensor* (ALS), ambos sensores de alta resolução espacial. Este último sensor apresenta desde o ano de 2004 problemas radiométricos.

No Quadro 2.1 encontram-se as características do sensor ETM⁺, o qual apresenta uma resolução temporal de 16 dias e em que a aquisição das imagens têm um preço relativamente acessível. O *swath* e o número de bandas do ETM⁺ permitem efectuar inúmeros estudos ligados ao acompanhamento da vegetação, oceanos, entre outros (NASA, 1999).

Quadro 2.1 - Características do ETM⁺ (fonte: NASA, 1999)

Banda (Nº)	Banda (cdo)	R. espectral (µm)	R. espacial (m)	Swath (km)
1	Azul	0,450 - 0,515	30	185
2	Verde	0,525 - 0,605	30	
3	Vermelho	0,630 - 0,690	30	
4	IV Prox.	0,75 - 0,90	30	
5	IV Médio 1	1,15 - 1,75	30	
6	IV Termico	10,4 - 12,5	60	
7	IV Médio 2	2,09 - 2,35	30	
Pan	Pancrom.	0,54 - 0,90	15	

2.3.2 - Sensor NOAA-AVHRR

Dentro dos satélites meteorológicos lançados pelos EUA, alguns da série NOAA (*National Oceanographic and Atmospheric Administration*), foram equipados com o sensor *Advanced Very High Resolution Radiometer* (AVHRR).

A utilização dos satélites meteorológicos é anterior à dos satélites de monitorização dos recursos terrestres. Devido à elevada resolução temporal e às características espectrais e de geo-posicionamento muitos cientistas usam-nos para a monitorização dos recursos terrestres.

Numa altitude de 830 km, estes satélites têm uma órbita quase polar e são heliocêntricos. A órbita tem uma inclinação de 99,09° e uma duração nominal de 102 minutos, traduzindo-se em 14,1 órbitas por dia, ou seja uma resolução temporal de 9 dias. O sensor AVHRR apresenta uma resolução espacial de 1,1 km (no *nadir*), classificando-o de baixa resolução espacial (Cracknell, 1997) (Quadro 2.2)

Quadro 2.2 - Características do sensor AVHRR, instalado a bordo do NOAA-15 (fonte: Cracknel, 1997)

Banda (nº)	Banda (cdo)	R. espectral (µm)	R. espacial (m)	Swath (km)
1	Visível (verm)	0,58 - 0,68	500	2940
2	IV próximo	0,72 - 1,0	1100	
3	IV médio	1,58 - 1,64 (dia)		
4		3,55 - 3,93 (noite)		
5	IV térmico 1	10,3 - 11,3		
6	IV térmico 2	11,5 - 12,5		

Uma das características que distingue este sensor doutros são as possíveis utilizações das suas bandas. A banda do vermelho do visível é utilizada para análise de nuvens e cartografia de superfície; a banda do infravermelho próximo tem aplicações na delimitação de corpos de água e

em cartografia de cobertura de vegetação; a banda do infravermelho médio é utilizada para determinação da temperatura da superfície do mar e da cobertura noturna de nuvens; as bandas térmicas 1 e 2 têm ambas aplicação na determinação da temperatura da superfície, sendo que a primeira também serve para distinguir nuvens. Estas especificidades evidência as capacidades do AVHRR em aplicações de longo prazo (e.g., análise multi-temporal) ligadas à monitorização da vegetação, do solo e à oceanografia (Hill, 1991).

2.3.3 - Sensor SPOT VEGETATION

O programa SPOT (*Système Probatoire d'Observation de la Terre*) foi desenvolvido pelo CNES (*Centre Nacional d'Études Spatiales*), em França, em colaboração com a Suécia e a Bélgica. O intuito deste programa foi desenvolver um satélite adaptado à monitorização das parcelas na Europa, em que os seus sensores apresentassem uma melhor resolução espacial do que os concebidos pelos americanos.

O VEGETATION, sensor criado no âmbito do referido programa pelo CNES e pela ESA, é um instrumento de baixa resolução espacial, que ficou operacional em Junho de 1998, a bordo do SPOT 4. Uma segunda versão, com melhoramentos ao nível da tecnologia e do sistema óptico, foi instalada no último satélite da série (SPOT 5). É um sensor passivo, composto por quatro câmaras independentes, cuja resolução radiométrica se encontra distribuída pela banda do azul, vermelho, infravermelho próximo e médio (Quadro 2.3). As câmaras apresentam um ângulo de cobertura de 101° e um *swath* de 2250 km, o que faz com que o sensor consiga cobrir a superfície terrestre praticamente num dia.

De acordo com CNES (2000) a finalidade geral do sensor VEGETATION são as medições de características básicas do coberto vegetal, o que o torna num instrumento vital para o estudo da vegetação global do planeta.

Quadro 2.3 – Características do VEGETATION (fonte: CNES, 2000)

Banda (Nº)	Banda (cdo)	R. spectral (μm)	R. espacial (m)	Swath (km)
(B0)	Vermelho	0,45 – 0,52	1165	2250
(B2)	IV Prox.1	0,61 – 0,68		
(B3)	IV Prox.2	0,78 – 0,89		
(B4)	IV médio	1,58 – 1,75		

2.3.4 - Sensor MERIS

A 1 de Março de 2002 foi lançado, pela Agência Espacial Europeia (ESA), o satélite de órbita helio-síncrona denominado ENVISAT-1 (Earth Observation ENVironmental SATellite). De entre os vários instrumentos que transporta a bordo, encontra-se o *Medium Resolution Imaging Spectrometer* (MERIS), um sensor de alta resolução radiométrica e espectral e de média resolução espacial.

O desenvolvimento deste sensor visou: o estudo dos oceanos e, em particular, a detecção de fitoplanton, clorofila e materiais em suspensão; o estudo da atmosfera ao nível do balanço da radiação e dos constituintes, em especial dos aerossóis; e o estudo dos continentes, ao nível da dinâmica dos processos bioquímicos, das alterações ambientais e dos ciclos biológicos, em particular na vegetação.

O sensor MERIS tem um IFOV de 68.5° no *nadir* e cobre um *swath* de 1150 km, com cinco câmaras ópticas idênticas, cobrindo o espectro entre os 390 e os 1040 nm, ou seja, no visível e infravermelho próximo. O tempo de duração da órbita deste satélite é de 100 minutos, cobrindo a totalidade da superfície terrestre em três dias, sendo, contudo, 35 dias o tempo que medeia a passagem do satélite sobre o mesmo local.

A imagem da terra é disponibilizada numa resolução espacial de 300 m no *nadir*, a qual também é degradada/reduzida, ao nível do sensor e de forma contínua, para 1200 m por uma combinação de quatro colunas de varrimento adjacentes por quatro linhas.

Existem três níveis de produtos derivados da informação adquirida pelo sensor MERIS, sendo que o primeiro (nível 1B) é a informação corrigida radiometricamente e geometricamente, e os produtos dos níveis 2 e 3 provêm dos produtos do nível 1. Uma das desvantagens da utilização dos dados deste sensor relaciona-se com as possíveis restrições face às políticas da ESA para a disponibilização dos produtos.

2.3.5 - Sensor MODIS

Desenvolvido pela NASA, o sensor *Moderate Resolution Imaging Spectrometers* (MODIS) visa o estudo e monitorização da superfície dos continentes, dos oceanos e dos constituintes atmosféricos.

O primeiro sensor MODIS (*ProtoFlight Model* ou PFM) foi integrado a bordo da plataforma espacial *TERRA* (EOS AM-1), com lançamento a 18 de Dezembro de 1999. O segundo (*Flight Model 1* ou FM1) foi lançado em 4 de Maio de 2002, a bordo da plataforma espacial *AQUA* (EOS PM-1).

O sensor multi-espectral MODIS tem uma sensibilidade radiométrica elevada (radiância), medida ao longo de 36 bandas espectrais, que cobrem, na totalidade, a região entre os 0.4 μm e os 14.4 μm . Este sensor, de resolução espacial moderada, apresenta uma resolução radiométrica elevada (12 bits).

A radiação electromagnética é registada pelo MODIS em quatro sensores (câmaras) independentes, cuja sensibilidade espectral abrange o visível (VIS), o infravermelho próximo (IVP), médio (IVM) e longo (IVL). As duas primeiras bandas (620 - 679 nm e 841 - 876 nm) apresentam uma resolução espacial nominal de 250 m no *nadir* sendo utilizadas, principalmente, para a definição das fronteiras terra/nuvens/aerossóis e

acompanhamento da vegetação. As cinco bandas seguintes (459 - 479 nm, 545 - 565 nm, 1230 -1250 nm, 1628 -1652 nm e 2105 - 2155 nm) distinguem-se das anteriores pela resolução do *pixel* de 500 m. As restantes 29 bandas registadas pelo MODIS, com uma resolução espacial de 1 km, no *nadir* são usualmente aproveitadas no estudo dos oceanos e dos componentes da atmosfera.

Com a órbita a 705 km de altura e um ângulo de varrimento de $\pm 55^\circ$, o MODIS consegue um *swath* de 2330 km, o que permite uma disponibilização de uma cobertura global da terra em cada 1 a 2 dias. Este sensor dispõe a bordo, de quatro calibradores internos (*v-groove blackbody*, *Solar Diffuser*, *Solar Diffuser Stability Monitor* e *Spectroradiometric Calibration Assembly*) e de uma “vista para o espaço” (vista para a lua e vista para o espaço longínquo) que lhe permite a produção de imagens georeferenciadas com radiância calibrada.

Actualmente a NASA, através do *MODIS Science Team*, tem um vasto grupo de investigação, que processa hierarquicamente os dados e explora as potencialidades deste sensor a vários níveis. Neste momento existem disponíveis cerca de 40 produtos diferentes, que compreendem produtos de baixo nível (1A – radiância nas 36 bandas espectrais), de alto nível (1B – radiância calibrada e imagem georeferenciadas) ou combinações dos dois (produtos nível 2 e 3).

O MODIS 09 – reflectância à superfície corrigida atmosféricamente, e o MODIS 43 – função de difusão da reflectância bi-direcional e albedo, são dois produtos do nível 2, que apresentam um interesse elevado e cuja combinação origina o produto de nível 3, utilizado neste estudo (MOD09A1).

A política de distribuição pública e gratuita adoptada pela NASA, e em particular pelo *MODIS Science Team*, para diferentes produtos gerados, potencia o seu uso em diferentes estudos científicos.

Um subproduto deste sensor, disponibilizado anualmente, que caracteriza a ocupação do solo no globo é o MOD12. A obtenção deste é feita aplicando um classificador de rede neuronal a outros produtos MODIS. O MOD12 caracteriza a ocupação do solo, de acordo com quatro sistemas de nomenclatura diferentes, no entanto, uma das suas principais limitações é a sua resolução espacial de 1 km. O *MODIS science team* também disponibiliza, com uma periodicidade de 16 dias, um sub-produto (MOD13), com o índice de vegetação de diferença normalizada (NDVI), que permite acompanhar o desenvolvimento da vegetação (estes e outros produtos estão disponíveis na página de *internet* em *MODIS Data Product* - <http://modis.gsfc.nasa.gov>).

2.4 – Métodos de extracção de informação

Na detecção remota existem diversas técnicas que permitem extrair informação das imagens de satélite para caracterizar a ocupação do solo.

A técnica de classificação permite transformar uma imagem de satélite numa carta temática, e no decorrer desta, os objectos da superfície são agrupados de acordo com características espectrais, espaciais ou temporais homogéneas, atribuindo-se a cada *pixel* uma determinada classe ou categoria previamente definida (Baio, 1996).

Segundo Caetano (2002), o primeiro passo para a extracção de informação temática de imagens de satélite consiste na definição de uma nomenclatura, também designado de esquema de classificação da cartografia a produzir.

Podem distinguir-se os sistemas de classificação de nomenclaturas em dois tipos base: o hierárquico e o não-hierárquico. A maioria apresenta sistemas hierárquicos, pois são sistemas mais consistentes. Têm a capacidade de incluir diferentes níveis de informação, começando com classes gerais, subdividindo-se, sistematicamente, em subclasses mais detalhadas. Em cada nível, as classes são mutuamente exclusivas e

independentes, sendo que, as classes mais altas apresentam critérios de carácter geral, e sempre que se desce no nível, o número e a complexidade dos critérios aumenta (Gregório, 2000).

A opção por um determinado sistema de nomenclatura deverá ter em consideração o estudo em causa, a unidade mínima identificável, a escala de produção, a disponibilidade e natureza da informação de base e ainda as respectivas características espectrais, espaciais e temporais.

As classes da nomenclatura devem ter uma terminologia clara e bem definida, por forma a evitar termos vagos, devendo estar adaptadas aos futuros utilizadores (Perdigão, 1997).

2.4.1 – Sistemas de nomenclatura

De entre os principais sistemas de nomenclatura definidos internacionalmente e utilizados, ao nível global, para a caracterização da ocupação do solo, apresentam-se cinco sistemas que são amplamente referidos nos diferentes estudos desta área da detecção remota:

- CORINE Land Cover (ao nível Europeu);
- Nomenclatura do United States Geological Survey;
- Nomenclatura do International Geosphere–Biosphere Programme;
- Nomenclatura FAO Land Cover Classification System;
- AHVRR Global Land Cover Classification .

2.4.1.1 - CORINE Land Cover

Na Europa é definido, como nomenclatura de referência, o sistema CORINE Land Cover (CLC). Este sistema resultou do programa “*Coordination of Information on the Environment - CORINE*” desenvolvido pela Agência Europeia do Ambiente (EEA) da União Europeia.

Este é um sistema hierárquico com três níveis de nomenclatura, em função do qual o número de classes é definido:

- Nível I – composto por 5 classes, de carácter generalista que tentam descrever os grandes tipos de ocupação do solo no planeta.
- Nível II – composto por 15 classes, em que o objectivo é a utilização desta cartografia entre a escala 1: 500 000 e a 1: 100000.
- Nível III – composto por 44 classes, tendo por objectivo a sua aplicação a projectos à escala 1:100000. Informação adicional está disponível no terceiro nível, com um carácter administrativo.

O CORINE é assim, obtido com base em imagens de alta resolução espacial, contudo, carece de informação auxiliar para a determinação das unidades do segundo e terceiro nível. O primeiro nível permite diferenciar superfícies artificiais, áreas agrícolas, florestas e áreas de vegetação semi-natural, zonas húmidas e corpos de água.

Em 2000 iniciou-se a produção do IMAGE e CORINE *Land Cover* (I&CLC2000) baseada na interpretação visual de imagens de alta resolução espacial apoiadas em ortofotomapas. Esta produção regeu-se por uma metodologia definida pela Agência Europeia do Ambiente (AEA) e pelo Joint Research Center (JRC). Em Portugal, o Instituto do Ambiente (IA) é a entidade responsável pela publicação do CLC2000. Esta cartografia foi pensada para apoiar o ordenamento e planeamento do território (Informação mais detalhada sobre o programa CORINE e os seus produtos encontra-se disponível na pagina de internet <http://image2000.jrc.it>).

2.4.1.2 - United States Geological Survey

Como resultado da necessidade de existência de uma nomenclatura normalizada para os EUA, Anderson et al. (1976) desenvolveram um sistema de classificação de ocupação do solo, o qual foi adoptado pelo *United States Geological Survey* (USGS). Este é um sistema hierárquico com quatro níveis. O primeiro nível apresenta nove classes de índole generalista (*water, developed, barren, Florest upland, shubland, non-nature*

woody, herbaceous upland natural /semi-natural vegetation, Herbaceous planted/cultivated, wetlands). O segundo nível apresenta 24 classes, com um detalhe elevado, que permitem a produção de cartografia a uma escala inferior a 1:80000. As classes do terceiro e quarto nível apenas podem ser derivadas a partir de alta resolução espacial ou de ortofotomapas devido ao seu elevado detalhe. Nestes níveis é produzida cartografia a escalas superiores a 1:20000 (mais informação sobre este sistema esta disponível em <http://landcover.usgs.gov>).

Com base no sistema de nomenclatura desenvolvido por Anderson et al. (1976) foi construído um sistema mais simples, designado de *National Land Cover Data* (NLCD), e que apenas adopta os dois primeiros níveis do USGS Land Cover (com pequenas diferenças), pelo que passou a ser adoptado por todo os EUA (University of Maryland, 1997).

2.4.1.3 - International Geosphere–Biosphere Programme

Segundo Hansan (2000), em 1992 é concebido por Townshend um sistema de classificação baseado no conjunto de imagens de NOAA-AVHRR, com 1km de resolução espacial, seguindo as recomendações do *International Geosphere–Biosphere Programme* (IGBP) para a investigação e desenvolvimento de programas para detecção de alterações, a nível global, na superfície terrestre. Este é actualmente designado de *IGBP Data and Information System Cover* (IGBP-DISCover), tratando-se também de um sistema hierárquico com 17 classes de ocupação do solo.

Ao contrário do CORINE e do USGS, aquele sistema permite a distinção das classes de ocupação do solo, com base em parâmetros biofísicos (tipo e longevidade da folha, altura média da vegetação e grau de cobertura), facultando a sua inclusão em processos de modelação bio-ecológica.

Das 17 classes da nomenclatura do IGBP, 11 unidades correspondem a vegetação natural, 3 correspondem a classes

desenvolvidas e de mosaico, e as restantes a classes de não vegetação. Nas classes de vegetação natural distinguem-se, com base em critérios biofísicos, espécies de folha permanente (*evergreen*) e de folha caduca (*deciduous*); florestas de espécies de folha larga (*broadleaf*) e de folha fina (*needleleaf*); florestas mistas (*mixed forests*); matos densos (*closed shrublands*) de matos pouco densos/abertos (*open shrublands*); savanas e montado (*woody savanas*); pastagens (*grassland*); sapais ou pântanos (*wetland*). Nas três classes mosaico ou desenvolvidas distinguem-se culturas anuais (*croplands*); urbano e áreas em construção (*urban and built-uplands*); mosaicos de vegetação natural e culturas anuais (*cropland/natural vegetation mosaic*). Nas três classes de não vegetação a nomenclatura IGBP identifica unidades de neve e gelo (*snow and ice*); corpos de água (*water bodies*) e solo nu (*barren land*).

As características dos três sistemas de nomenclatura da ocupação do solo anteriormente referidos foram adoptados por vários estudos, com dados de diferentes sensores. Tal procedimento foi também adoptado pelo *MODIS Science Team*, para o produtos MOD12Q1 – Global Land Cover data, disponibilizando dados de caracterização da ocupação do solo, classificados segundo cinco sistemas de nomenclatura diferentes.

2.4.1.4 - FAO Land Cover Classification System

A utilização de diferentes sistemas de nomenclatura e de diferentes sensores, dificulta a comparação de dados de ocupação do solo entre períodos e locais diferentes. Esta não uniformização conduz a uma série de limitações, como seja a diferença de definições das unidades taxonómicas identificadas e a exactidão relativamente ao conteúdo, contexto e forma dessas mesmas unidades (Waser, 2004).

Para ultrapassar alguns destes constrangimentos a *Food and Agriculture Organization of the United Nations* (FAO) desenvolveu uma base de dados de classificação de ocupação do solo para todo o mundo. O

FAO *Land Cover Classification System* (LCCS) é um sistema de classificação *à priori* normalizado, ou seja, onde as classes estão previamente definidas. O que torna este sistema único é o facto da sua nomenclatura poder ser comparável com todos os sistemas e em todo o mundo. É um sistema composto por duas etapas/fases de classificação: a fase dicotómica e a fase hierárquica (Gregorio et al, 2000).

Numa primeira etapa (fase dicotómica) oito classes genéricas são definidas em: áreas territoriais cultivadas ou gestionadas; territórios de vegetação natural e semi-natural; áreas aquáticas ou regulamente inundáveis cultivadas, áreas aquáticas ou regulamente inundáveis com vegetação natural ou semi-natural; superfícies artificiais, áreas solo nu ou improdutivo, corpos de água artificiais e corpos de água natural incluindo gelo e neve.

Numa segunda etapa (fase hierárquica – modelar) são geradas as classes de ocupação, de acordo com classificadores pré-definidos. Estes últimos dependem das características das oito classes genéricas (informação mais detalhada sobre este sistema está disponível em <http://www.fao.org/sd/Eidirect/Eire0019.htm>).

2.4.1.5 - AVHRR Global Land Cover Classification

Como resultado do elevado potencial de utilização de alguns dos sensores, vários sistemas de nomenclatura para a caracterização da cobertura do solo foram desenvolvidos.

Os dados do sensor AVHRR, instalado a bordo dos NOAA conduziram à definição do *AVHRR Global Land Cover Classification* (GLCC), o qual foi desenvolvido na Universidade de *Maryland*, EUA, pelo que é também designado de UMD 1km Global Land Cover. É um sistema hierárquico com 13 classes de ocupação do solo, das quais 10 são dedicadas à vegetação natural. As 3 classes de inexistência de vegetação são: a água (*water*), o solo nu (*bare soil*) e o urbano e edificado (*urban and*

built.up). Na vegetação natural podem distinguir-se: quatro classes de floresta (*evregreen needleleaf forest*, *evergreen broadleaf forest*, *deciduous needleleaf forest* e *deciduous broadleaf forest*), uma classe mista de floresta (*mixed forest*), uma classe de terras florestais degradadas (*woodland*), uma classe de montado (*wooded grassland*), uma classe de matos densos (*closed shrubland*) e uma classe de matos dispersos (*open shrubland*). Apresenta ainda uma classe de pastagem (*grassland*) e uma classe de culturas anuais (*cropland*).

Esta nomenclatura está associada a um produto de classificação da cobertura do solo, com o mesmo nome, o qual é derivado dos dados do AVHRR, tem 1 km de resolução espacial, e está acessível a toda a comunidade científica (Hansen, 1998) (informação mais detalhada sobre este sistema está disponível na página de *internet* do *Glogal Land Cover Facility*, em <http://glcf.umiacs.umd.edu/data/landcover/>).

2.4.2 – Algoritmos de extracção de informação

Segundo JARS (1996), a extracção de informação pode ser realizada directamente por intervenção humana ou por intermédio de operações e análises computacionais (classificação automática).

A interpretação visual ou foto-interpretação é uma técnica que se baseia na identificação visual, na delimitação e na classificação (atribuição de código) de zonas com o mesmo tipo de ocupação de solo. Nesta técnica os objectos da superfície são agrupados em classes, mediante as suas características espectrais e espaciais, nomeadamente, através da sua forma, textura, contexto e tamanho, entre outras (Baio, 1996). É ainda, comum utilizar-se informação auxiliar (e.g., trabalhos de campo, fotografias aéreas) para apoio à interpretação e extracção de informação das imagens de satélite.

A classificação automática de imagens consiste na utilização de algoritmos de reconhecimento de padrões espectrais e/ou espaciais, que convertem os dados da imagem em classes. Esta conversão é feita através da comparação da imagem com elementos padrão, utilizando regras de classificação ou funções discriminantes (JARS, 1996).

Independentemente do tipo de classificação, surge o conceito de unidade espacial de análise (UEA), a qual pode ser o *pixel* ou o objecto. No primeiro caso, a classificação baseia-se única e exclusivamente nas características espectrais individuais do *pixel*; no segundo caso, considera-se o objecto, ou seja, o conjunto de *pixels* que partilham uma determinada característica (espectrais e espaciais) e cuja heterogeneidade não ultrapassa um dado valor (Caetano, 2002).

Os algoritmos de classificação podem distinguir-se em rígidos e relaxados (*soft*). Os rígidos têm como pressuposto que cada UEA apenas é atribuída a uma só classe. Contudo, na realidade a paisagem é heterogénea, pelo que existe a necessidade de encarar o valor da UEA como pertencendo a vários objectos da superfície e, conseqüentemente, a várias classes. Definiram-se algoritmos de classificação relaxada, onde existe uma probabilidade de pertença de uma UEA para cada classe (JARS, 1996).

Podemos ainda distinguir a classificação em dois tipos: a assistida ou não assistida. A classificação não assistida, (*unsupervised classification*) tem como pressuposto que não se define *à priori* o número de classes que correspondem à real ocupação da superfície. Trata-se de um método exploratório de dados, que agrupa os *pixels* homogéneos, sob o ponto de vista espectral, num mapa de agregados (*clusters*), de acordo com um determinado critério. Numa segunda fase atribui-se uma nomenclatura aos agregados definidos. O método da distância euclidiana simples é o mais utilizado para a determinação da similaridade entre as assinaturas espectrais. Para a formação dos agregados dos *pixels* existem diversos

algoritmos (e.g.: K-médias, método do histograma multi-dimensional); o algoritmo ISODATA (*Iterative Self-Organizing Data Analysis Technique*) é o mais utilizado, cujos detalhes podem ser consultados em Caetano (2002).

Segundo Barata (2001), dois tipos de dificuldades podem estar associados a este tipo de classificação. A primeira resulta da dificuldade de conseguir estabelecer a correcta correspondência entre as classes espectrais e as reais. A segunda advém do facto de ser frequente existirem, na classificação final, diferentes classes espectrais que pertencem ao mesmo tipo de classe de ocupação do solo.

Na classificação assistida (*supervised classification*), o utilizador escolhe previamente uma nomenclatura e identifica um conjunto de classes de ocupação do solo. A fase de treino consiste na recolha de amostras para identificação de assinaturas espectrais que caracterizem as classes (áreas de treino). A fase de classificação agrupa todos os *pixels* da imagem de acordo com o comportamento espectral das classes previamente definidas na fase de treino. Para que as áreas de treino sejam representativas da classe, devem ser unimodais e representativas da dispersão da classe na imagem, o que se consegue seleccionando várias amostras, bem distribuídas pela área de estudo e subdividindo as classe multi-modais (Fonseca, 2004).

Existem diversos algoritmos para a fase da classificação, sendo os mais difundidos, o método das distâncias mínimas às médias, o método do paralelepípedo e o método da máxima verosimilhança, cujos detalhes podem ser consultados em Caetano (2002).

Outra forma de extracção de informação das imagens de satélite é através do recurso a redes neuronais e das árvores de decisão.

As redes neuronais artificiais (*artificial neural network* - ANN) são modelos computacionais, preditivos não lineares, criados com o intuito de emular o funcionamento do cérebro humano. Pretende-se que as ANN tenham capacidade de aprendizagem, de adaptação e de generalização.

A estrutura fundamental numa rede neuronal é o neurónio. Cada neurónio é estimulado ou seja, recebe sinais dos neurónios vizinhos, enviando sinais após processamento, para outros neurónios. As redes neuronais com ligações para a frente (FFNN) constituem uma classe especial de ANN, nas quais todos os neurónios de uma determinada camada l estão ligados a todos os neurónios da camada $l-1$, (Figura 2.1) onde o algoritmo de aprendizagem, “aprende” com base nos dados de *input*.

Uma rede com ligações para a frente é constituída tipicamente por uma camada de entrada, que corresponde aos dados que entram na rede, uma camada escondida, cujos neurónios recebem os dados produzidos pelos neurónios da camada de entrada e uma camada de saída, cujos neurónios recebem dados da camada escondida e que correspondem à saída da rede. As conexões entre os neurónios são designados de pesos ou ponderações.

Os resultados de saída da rede neuronal dependem dos dados de entrada, dos valores iniciais dos parâmetros da rede e da relação entre os próprios neurónios. Essa relação é representada pelo produto da matriz de pesos que incide nesse neurónio, W (e.g., o elemento $w_{S,R}$ da matriz, corresponde ao sinal propagado pelo neurónio R da camada de entrada para o neurónio S da camada escondida), pelos valores de entrada na rede, I , ao qual se adiciona um *bias* (termo de polarização, i.e., valor limiar a partir do qual o neurónio é inibido - b_S) associado a cada neurónio. A esse resultado finalmente aplica-se uma função de activação (f), de acordo com o problema em questão (Malheiro et al., 2004). Da aplicação da função de activação resultará um valor que será propagado para os neurónios da camada seguinte.

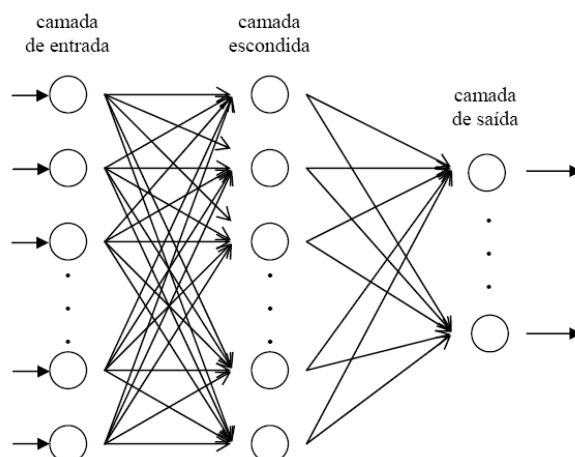


Figura 2.1 - Esquema de rede neuronal de ligações FFNN (adaptado de Malheiro et al., 2004)

O objectivo da rede neuronal é validar correctamente os dados de entrada, i.e. produzir as saídas mais adequadas para os exemplos que são introduzidos na rede. Para tal é necessário configurar os parâmetros da rede, ou seja, os pesos sinápticos, as taxas de aprendizagem e os parâmetros opcionais de *bias*. Para ajustar estes parâmetros, é necessário que a rede seja devidamente treinada. Tal qual na classificação, as redes podem ser treinadas de forma assistida e não assistida.

Existem tipos de redes neuronais que não possuem camada escondida. No *Self Organized Maps* (SOM) os neurónios encontram-se dispostos em forma de matriz, mas, como é comum a todo o tipo de redes, está presente uma matriz de pesos, conectando duas camadas.

Na análise de *clusters* com redes neuronais (e.g., SOM) são aplicadas um conjunto de metodologias de estatística multivariada que, a partir de um conjunto de informação sobre um grupo de entidades, procura organizá-las em grupos relativamente homogéneos, determinando uma “estrutura” de semelhanças entre as unidades (Bação, 2002). Este tipo de análise difere da classificação tradicional porque não depende de classes pré-definidas, sendo os registos agrupados apenas com base nas suas semelhanças.

As árvores de decisão são ferramentas de classificação e previsão. Ao contrário das redes neuronais, o modelo estrutura-se em forma de

ramos de árvore, que representa um conjunto de decisões, as quais geram regras para a classificação dos dados (Bação, 2002).

Nas árvores de decisão procura-se, a partir de um conjunto de dados independentes (variáveis de *input*), definir regras que permitem isolar, em grupos, as observações que possuem valores idênticos para a variável dependente ou o *target* (*output*). Cada registo (*input*) é apresentado na raiz da árvore, onde é aplicado um teste, que determina a que nó filho esse registo pertence. Ao longo da árvore são aplicados, em cada nó, testes que melhor discriminam cada uma das classes *target*. O processo é feito, através de um caminho único, até que o registo chegue a uma folha (nó terminal), onde todos os registos desse nó estão classificados da mesma forma. Nós diferentes podem dar origem a uma mesma classe, mas cada nó terminal produz essa classificação com um conjunto de regras diferentes. Os métodos de árvores de regressão mais conhecidos são *Classification and Regression Trees* (CHART) e o *Chi Square Automatic Interaction Detection* (CHAID).

Existem diversos algoritmos de redes neuronais e de árvores de decisão, cujos detalhes podem ser consultados em Bação (2002). Os métodos de classificação descritos não são as únicas técnicas de extracção de informação das imagens. O recurso a índices de vegetação (Ferreira et al, 2003) ou de neve (Scharfen, 2005, Klein, 1998) permite avaliar, monitorizar e caracterizar a evolução da ocupação do solo. Outros algoritmos, como o modelo de mistura linear, permitem expressar a composição espectral de uma imagem com base em imagens fracção que expressam a abundância de um determinado número de elementos puros (Adams et al., 1986, Smith et al, 1990).

2.4.3 – Restrições das abordagens tradicionais na extracção de informação

Alguns estudos classificam imagens de resolução espacial moderada baseando-se na reflectância espectral dos componentes da superfície e em índices de vegetação. Estes últimos exprimem as características intrínsecas da vegetação, nomeadamente, a alta reflectividade na banda do Infravermelho durante o período de desenvolvimento vegetativo. Contudo, estas duas abordagens e os processos de classificação utilizados para extrair informação das imagens apresentam alguns problemas, que enumeram-se em seguida.

O problema do *pixel* misto é reportado em vários estudos (Quarmby et al., 1992, Ashton, 1998) onde existe mais que um tipo de ocupação na área coberta pelo *pixel*, implicando falhas na identificação de objectos que se encontram ao nível do *subpixel*. A radiância registada pelo satélite é uma integração da soma das radiâncias de todos os materiais que estão no ângulo de observação (*IFOV*) do sensor (Shimaburuko e Smith, 1988), pelo que as várias ocupações registadas num único pixel traduzem-se numa ocupação com características espectrais complexas. Nesta sequência pode-se verificar que, para além dos materiais da superfície, a mistura também contará com a contribuição dos constituintes atmosféricos. Assim, a proporção de materiais (objectos da superfície e partículas atmosféricas) num *pixel* poderá variar de *pixel* para *pixel* (Settle, 1991).

A resposta espectral mista dos *pixels*, que se encontram nas zonas de transição entre grandes áreas de unidades cartográficas pode originar um segundo tipo de problema/erro na caracterização das classes de ocupação do solo (Cross, 1991). Nestas zonas de transição é possível ter várias combinações de objectos, os quais podem ser incorrectamente agrupados numa classe de objectos puros.

Um outro constrangimento relaciona-se com o efeito topográfico da superfície, que resulta em formas de iluminação/brilho da imagem obtida, o que implica que um mesmo objecto da superfície possa apresentar reflectâncias diferentes. Os índices de vegetação, por serem normalizados, reduzem o efeito topográfico ou seja a geometria da iluminação. (Rouse et al., 1973, citado em Asner, 2000). Mas para classes de ocupação do solo dependentes da topografia, o uso de índices de vegetação conduz a erros de classificação, sendo necessário introduzir dados de simulação relativos ao relevo, para ajudar à sua identificação (Caetano, 2000).

A atribuição de classes aos processos de classificação não assistida, ou a definição de classes de treino na classificação assistida é difícil, pois resulta de um processo manual e carece de dados de campo para apoio à identificação das mesmas (Caetano, 2000).

Outro dos obstáculos relaciona-se com o facto da classificação obedecer a regras e critérios cujo processo de selecção, depende da sensibilidade de cada um, tornando-se num método subjectivo (Verbaly, 2004). No caso da classificação não assistida é necessário focar a relação entre as classes preliminares e as classes do sistema de nomenclatura seleccionado. No caso da classificação assistida, as áreas de treino poderão não representar integralmente as unidades de um país ou continente.

2.4.4 – Análise de Misturas Espectrais

Uma das metodologias de extracção de informação baseia-se na Análise de Misturas Espectrais (AME) e, em particular, no Modelo de Mistura Linear.

A base para a AME assenta no pressuposto físico de que a maioria da variedade espectral existente numa imagem de satélite resulta da

mistura de um pequeno número de componentes com diferentes reflectâncias (Adams et al., 1986).

A mistura ao nível do *sub-pixel* estipula que a reflectância registada por um *pixel* não pode ser apenas interpretada pelas propriedades de um tipo de cobertura do solo. Pelo contrário, a reflectância da paisagem deve ser determinada com base nas variações das proporções de cada tipo de cobertura de solo (Asner, 1998). Determinar as variações ao nível do *sub-pixel* é uma etapa essencial da análise da unidade espacial em zonas heterogéneas de ocupação do solo

A determinação da componente espectral dos objectos da superfície registados num dado *pixel* pode ser modelada através de um conjunto de reflectâncias de elementos designados, na literatura, de componentes puros (*Endmember*). Estes podem ser comparados às classes de treino do modelo, as quais são escolhidas com base em unidades “puras” da superfície (um só tipo de ocupação), por exemplo, vegetação e solo.

Os componentes puros (*endmember*) definem as características espectrais das unidades não contaminadas com misturas de outros objectos/elementos da superfície, num conjunto de imagens. Segundo Adams et al. (1986), os componentes puros podem ser definidos como um conjunto de elementos, cujas respectivas reflectâncias espectrais permitem, através de regras apropriadas, reproduzir a reflectância de todos os *pixels* mistura. A Figura 2.2 ilustra esquematicamente a contribuição de três componentes puros – sombra, solo, vegetação - nos *pixels* mistos representativos dos diferentes objectos da superfície.

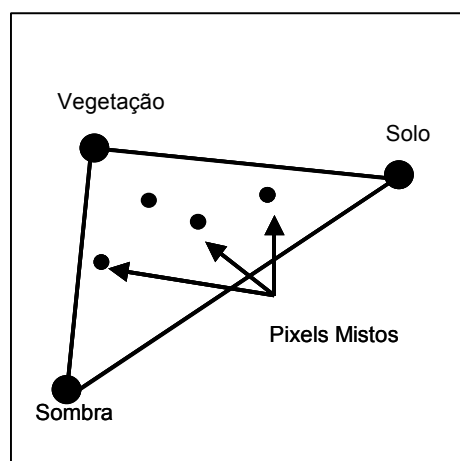


Figura 2.2 - Contribuição dos componentes puros no *pixels* mistos (adaptado de Asner, 2000).

2.4.4.1 - Componentes Puros Imagem *versus* Referência

Os componentes puros podem ser derivados das imagens, de trabalhos de campo ou através estudos espectrais em laboratório.

Os componentes puros imagem são aqueles que são obtidos directamente a partir de pontos das imagens de satélite. A maioria das aplicações de AME utiliza este tipo de selecção de componentes puros, uma vez que são fáceis de obter, representam medidas espectrais à mesma escala temporal e espacial que os dados e providenciam uma medida simples da abundância dos mesmos (Roberts et al., 1999).

Os componentes puros de referência são obtidos em trabalhos de campo ou em análise de laboratório e retransmitem as características espectrais de elementos verdadeiramente puros. A adopção deste tipo de componentes puros tem vantagens, as quais se traduzem no facto de serem independentes da data, do tempo, do espaço/área, do satélite/sensor utilizado e das condições atmosféricas (Adams, 1986). No entanto, a sua aplicação é mais difícil e apresenta alguns problemas. O primeiro resulta do facto da reflectância dos *endmember*, obtida a partir de livrarias espectrais, corresponder a materiais puros, os quais poderão não estar presentes nas imagens. Um segundo problema advém do facto da

informação ser recolhida a uma escala diferente da imagem (Roberts et al., 1999). A utilização de componentes puros de referência requer a calibração da radiância da imagem para reflectância à superfície, uma vez que ambos têm de estar nas mesmas unidades (Caetano, 2000).

Na maioria das aplicação de AME é utilizado um componente puro virtual, a sombra, que permite a simulação dos *pixels* que têm a mesma abundância que os materiais da superfície, mas têm diferentes reflectâncias devido à diferente geometria de iluminação da cena (Adams et al., 1986, Smith et al, 1990).

A identificação dos vectores de componentes puros, juntamente com as imagens de satélite escolhidas são *inputs* da AME. Como resultado do modelo são geradas imagens fracção para cada componente puro escolhido. Estas imagens fracção expressam a abundância/contribuição que cada componente puro tem para a informação espectral do *pixel*. Uma vez que, a composição espectral de um material da superfície depende da contribuição de cada componente puro, o resultado do modelo, expresso nas imagens fracção, relaciona a contribuição dos componentes puros dentro de cada *pixel*. Neste modelo obtém-se tantas imagens fracção quanto o número de componentes puros.

A AME, ao contrário de outros modelos, incorpora um sistema de auto-avaliação que se traduz pela determinação do erro expresso através de uma imagem de resíduos, a qual apresenta a estimativa do erro (RMS) relativo à modelação da contribuição dos componentes puros na mistura do *pixel* (Adams, 1990). Este modelo apresenta a vantagem de não necessitar de informação auxiliar para a análise e determinação dos resultados. A imagem de RMS permite, assim, avaliar a capacidade dos componentes puros para modelar a mistura dos *pixels* mistos na imagem, mas não permite inferir acerca da adequabilidade dos componentes puros para caracterizar as diferentes ocupações do solo (Roberts et al., 1990)

2.4.4.2 - Linear *versus* Não linear

A AME pode classificar-se em Mistura Linear e Mistura Não Linear de componentes. A diferença entre os dois tipos de análise de misturas baseia-se no tipo de interações que os fotões têm com o material da superfície, abrangido por um *pixel*.

A mistura é linear se um fotão interage apenas com um componente da superfície, no seu caminho entre o sol e o sensor, ou seja, antes de ser reflectido da superfície (Malinmgreau e Tucker, 1988, citado em Cross, 1991). Num modelo de misturas não lineares, o fotão interage com mais de um componente/objecto, antes de ser reflectido da superfície e registado pelo sensor (Caetano, 2000).

A simplicidade do Modelo de Mistura Linear tem contribuído para a generalização da sua utilização em detrimento do Modelo Não Linear. As principais fontes de não linearidade da radiação EM podem ter quatro origens (Caetano, 2000):

- efeitos de adjacência de *pixels* brilhantes junto a *pixels* escuros, o que se reflecte numa fonte adicional de radiância;
- estratificação vertical dos componentes puros, ou seja, fotões que são reflectidos pelo solo e que podem interagir com a vegetação existente num nível superior antes de serem reflectidos da superfície;
- difusão múltipla e transmissões no copado das plantas;
- irregularidade da zona onde o modelo é aplicado devido a:
 - irradiância difusa do céu ;
 - irradiância directa ou difusa reflectida de zonas adjacentes.

Embora os processos reais de transmissão de energia EM entre os objectos da superfície, e em especial na vegetação, apresentem um comportamento não linear, os modelos lineares conseguem muitas vezes modelar a radiância captada pelo satélite.

2.5 – Estudos para a Caracterização da Ocupação do Solo

Considerando os diferentes estudos para a caracterização da ocupação do solo abordados nesta Dissertação, as diferentes metodologias serão apresentados da seguinte forma: 1) estudaremos a diferença entre a opção por uma abordagem de análise efectuada em séries de dados uni-temporais (uma única imagem) e multi-temporais (várias imagens no período de estudo); 2) como resultado das vantagens que a multi-temporalidade propicia à caracterização da ocupação do solo, estudaremos os diferentes métodos de pré-processamento e os objectivos da sua aplicação; 3) diferenciaremos os estudos em função da variável de estudo e do algoritmo de extracção de informação utilizado nas imagens de satélite.

Os estudos de caracterização de ocupação do solo realizados com uma única imagem (uni-temporal) são sobretudo efectuados com dados de sensores de alta resolução espacial, do tipo LANDSAT, ou resultante da utilização de sensores aerotransportados. Estudos como o de Robert et al. (1990), realizados na zona da reserva natural de *Japer Ridgel*, nos EUA, onde é adquirida uma única imagem do sensor *Airborn visible/infrared imaging spectroemter* (AVIRIS) ou mais recentemente Aboelghar (2002), que utiliza uma imagem Landsat ETM⁺ para estudar a ocupação do solo no Egipto, evidenciam a abordagem uni-temporal. Os estudos multi-temporais têm como vantagem permitir a distinção dos diferentes elementos à superfície, com base na variação espectral dos objectos ao longo do tempo.

Devido à facilidade de aquisição e ao preço das imagens de média e baixa resolução, este tipo de dados praticamente só são utilizados em séries multi-temporais. A utilização de uma série de imagens ao longo de um dado período de tempo (abordagem multi-temporal) também permite analisar a evolução do comportamento espectral dos diferentes elementos da superfície do globo.

O número de imagens utilizado nos diferentes estudos com séries multi-temporais, varia conforme os tipos de ocupação do solo que se pretendem caracterizar. No âmbito das imagens de alta resolução, Tateishi (2004) constrói uma série com cinco imagens de Landsat ETM⁺, respectivamente uma na primavera, duas no verão e outras duas no Outono. Já na baixa resolução espacial, o conjunto de dados utilizado é muito superior. No trabalho realizado por Agrawal (2003), são utilizadas as imagens diárias do sensor VEGETATION (VGT) como dados de base da série temporal.

Giri et al. (2000) utilizam dois sensores (NOAA e VGT) para a constituição da série temporal, com o intuito de caracterizar a ocupação do solo e detectar as suas alterações. Uma abordagem semelhante é utilizada por Chuvieco et al. (2005), onde a série temporal é constituída por compósitos de 8 dias de imagens NOAA-AVHRR e MODIS durante os anos de 2002 a 2004, com o objectivo de caracterizar as áreas ardidas.

Nos estudos multi-temporais, diferentes autores recorrem ao pré-processamento das imagens com o objectivo de reduzir os problemas atmosféricos e reduzir o volume de dados. É uma fase importante da exploração de dados para caracterização do solo, e onde várias abordagens metodológicas podem ser utilizadas.

Moigne et al. (2001) explora dois métodos de pré-processamento de uma série intra-anual de imagens de LANDSAT ETM⁺. Na primeira abordagem utilizam a análise de componentes principais (PCA), a qual permite identificar a inter-dependência entre canais espectrais e consequentemente reduzir o número de bandas espectrais a explorar. Posteriormente utilizam o algoritmo *Regression Variable Subtitution* (RVS) para determinar uma combinação linear (vector de substituição) entre os canais, permitindo substituir uma banda espectral por outra.

Algumas estratégias para o pré-processamento dos dados são executadas com o propósito de reduzir o efeito das nuvens e da atmosfera (também denominado de ruído) no conjunto de dados da série temporal.

Para análise da vegetação na Índia, Agrawal (2003), utiliza o *Maximum Value Composite* (MVC) como forma de redução do ruído das imagens diárias de VGT. Para o efeito, recorre a uma abordagem pixel a pixel e, para um período de cinco dias, elege o pixel com NDVI máximo nas imagens que formam o compósito mensal, para representar o período. Contudo, o ruído residual, de origem atmosférica bem como de anisotropias dos constituintes da superfície, mantém-se nas imagens. Pan (2003) utiliza a mesma técnica (máximo NDVI) para constituição de compósitos de uma série multi-temporal (12 imagens), gerados de NOAA-AHVR.

Giri (2000) gera compósitos de NOAA-AVHRR, servindo-se do parâmetro de qualidade (QF - *Quality Factor*) com a finalidade de seleccionar as melhores imagens de Verão (em termos de cobertura de nuvens e qualidade dos dados). As falhas de dados na melhor imagem foram preenchidas com dados das outras imagens do compósito.

Quarmby (1992) dispo de uma série temporal de 20 imagens, realiza o pré-processamento dos dados de NOAA-AVHRR com o objectivo de eliminar a presença de nuvens. Para a constituição de nove imagens compósito também recorre ao processo do valor máximo compósito (valor máximo NDVI no mês) mas complementa-o com a substituição dos pixels com nuvens através da interpolação dos pixels vizinhos.

Nos estudos multi-temporais para caracterização da ocupação do solo foram utilizadas diferentes variáveis de estudo e diferentes algoritmos de extracção de informação. A abordagem mais simples reside na utilização directa das bandas espectrais, submetendo-as ao algoritmo de classificação assistido de máxima verosimilhança, para identificar 27 classes de ocupação do solo, agrupada em 9 grandes classes, segundo uma nomenclatura hierárquica de dois níveis (Ahand, 1989). Outros

estudos apresentam como variáveis de estudo, informação derivada das bandas espectrais, como seja os índices de vegetação ou as imagens fracção resultante da aplicação de MML.

Dentro destes dois últimos tipos de variáveis, a aplicação de algoritmos de classificação não assistida é uma das abordagens mais utilizada. Para a definição das classes de ocupação, Agrawal (2003) recorre à aplicação do algoritmo de classificação não assistida (ISOCCLASS) a nove imagens mensais de MCV de imagens NDVI derivadas de AVHRR, com o objectivo de diferenciar vários tipos de ocupação do solo, e em especial a vegetação. Numa segunda fase aplica a nomenclatura GLC2000 da FAO para classificação do resultado da análise *cluster*, sendo obtidas 45 classes de ocupação do solo. Segundo este autor, o perfil temporal de NDVI é um reflexo da cadência de crescimento da vegetação, sendo mais fiável a classificação da vegetação com base em séries multi-temporais de NDVI do que em classificação espectral.

Também Pan (2003), recorre ao algoritmo de *clustering* ISODATA para explorar 12 imagens de máximo NDVI, e identificar classes de vegetação de acordo com nomenclatura do IGBP, utilizando, contudo, informação auxiliar (e.g.: modelos de elevação/terreno, clima, solo) na caracterização de ocupação do solo. O recurso a este tipo de dados é, normalmente, empregue pós-classificação, para separação das classes preliminares de vegetação heterogéneas em classe de ocupação homogéneas (Loveland *et al.* 1997, 2000, citado em Pan, 2003).

Outro tipo de metodologia é a aplicação de redes neuronais e árvores de decisão. Recorrendo a estas ultimas, Matsuoka (2004) define uma métrica multi-temporal para caracterização das classes de ocupação do solo na série de compósitos mensais derivados do MOD09A1. Os critérios de identificação das classes são definidos manualmente, em função do comportamento temporal do NDVI e da reflectância, o que pode originar um défice ou excesso de área das classes identificadas. A

metodologia aplicada evidencia dificuldades na distinção entre vegetação natural e culturas agrícolas, devido ao comportamento semelhante (Matsuoka, 2004),

Harahsheh e Tateishi (1997), recorrem à classificação assistida, através do algoritmo de árvore de decisão, para classificar uma série temporal de NDVI obtida de 36 compósitos de AVHRR. Para definir as regras de decisão da árvore recorrem aos valores máximos e mínimos de NDVI para 57 regiões (áreas de teste), as quais representam 13 classes de ocupação do solo.

A análise de misturas espectrais é utilizada, em diversos estudos, para a exploração e extracção de informação dos dados de sensores de várias resoluções espaciais.

A utilização de compósitos, na análise de misturas espectrais, permite aumentar a precisão do modelo, uma vez que os componentes puros não serão afectados pelas nuvens (Quarmby, 1992).

Tateishi (2004) explora uma série temporal de 5 imagens Landsat 7 ETM+ através da análise de misturas espectrais, decompondo-as em quatro fracções – vegetação “verde”, vegetação não fotossintética, solo (áreas cultivadas e areia (deserto) - por imagens compósito. O algoritmo ISOCLUSTER é aplicado às imagens fracção para avaliação do comportamento das fracções ao longo do tempo. O estudo realiza-se numa zona de acentuados contrastes nas unidades da paisagem (lago Baikal – Rússia). Este estudo avalia a precisão, para caracterização da vegetação natural, do modelo de mistura linear quando comparado com a análise baseada na evolução temporal do NDVI e na análise espectral, revelando-se como o mais eficaz do ponto de vista da exactidão do utilizador e do produtor.

Ustin (1993) também aborda a caracterização da ocupação do solo através da decomposição espectral de imagens de LANDSAT e AVIRIS recorrendo à análise de misturas espectrais, de três datas diferentes.

Contudo o objectivo deste estudo é determinar os componentes puros que melhor caracterizam a área estudada, o qual é alcançado através da análise das imagens de resíduos do modelo de misturas.

Como já for referido anteriormente, a identificação de componentes puros pode ser obtida a partir de livrarias espectrais, recolha dos dados em campo, ou de identificação na imagem.

Roberts et al. (1990) utilizam livrarias espectrais para definir os componentes puros do MML. Esta opção poderá implicar a não existência dos componentes na série temporal estudada.

Quarmby et al. (1992) determinam os componentes puros com base em duas imagens de alta resolução espacial (SPOT HRV), que, depois de classificadas com o algoritmo da máxima verosimilhança, são degradadas com o filtro de majoração do pixel de 5 por 5. Desta forma conseguem reclassificar pixels isolados pertencentes a um tipo de ocupação para a mesma classe a que pertencem a maioria dos pixels vizinhos. Com base na sobreposição das classes obtidas com as imagens de AVHRR, seleccionam-se os componentes puros e respectivas respostas espectrais.

Cross et al. (1991) aplicam uma abordagem semelhante: depois de utilizarem uma classificação assistida em imagens LANDSAT TM, degradam o resultado para sobreposição com AVHRR. Os componentes puros e as respectivas assinaturas espectrais são seleccionados com base nos dados de ocupação do solo derivados da alta resolução espacial.

A análise dos resultados da aplicação das diferentes metodologias permite avaliar o seu desempenho na caracterização da ocupação do solo. Giri (2000), conseguiu realizar a discriminação das classes de ocupação do solo, com uma resposta espectral semelhante, com base na análise de evolução do índice de vegetação (NDVI de VGT). Este foi útil na discriminação de zonas florestais e não florestais, enquanto que, a discriminação baseada exclusivamente na informação espectral do sensor (AVHRR) providenciou informação complementar sobre floresta de folha

permanente e folha caduca. A metodologia proposta, com sensores de baixa resolução espacial e a utilização de apenas algumas das bandas espectrais, traduz-se na pouca capacidade para discriminação das classes, originando sub-estimação da área da classe. Além disso, os erros de geometria das imagens e incorrecção na definição das classes afectam a exactidão do processo de classificação (Giri, 2000).

Uma metodologia para avaliação da sombra (*shadow*) nas áreas florestais (vegetação “cerrado”) a partir de imagens de LANDSAT MSS é feita com recurso ao MML, onde a sombra é considerada um componente principal do *pixel* (Shimabukuro e Smith, 1988). Segundo estes autores, a contribuição da fracção sombra varia em função da altura do povoamento, uma vez que, dentro do eucalipto foram identificadas várias concentrações/proporções de sombra em função da idade do povoamento florestal. A estrutura do povoamento também influencia a sombra, bem como a componente topográfica. Igualmente Robert et al. (1990) analisa as alterações de iluminação nas imagens de AVIRIS, e conclui que a introdução de um componente puro virtual (sombra) permite modelar, quer diferentes tipo de ocupação do solo, com comportamentos espectrais semelhantes, quer diferenças de iluminação em objectos iguais da superfície, reduzindo o efeito topográfico. Concluem que, apesar de os materiais da superfície não terem sido modelados como componentes puros, podem ser identificados na imagem de resíduos, após a sua análise.

No caso de exploração de séries temporais com MML, a análise das imagens resíduos permitiu identificar a pureza dos componentes puros escolhidos, face à informação espectral do conjunto de dados utilizados (Ustin, 1993). Segundo este autor a imagem de resíduos permite identificar componentes “mais” puros, caso os escolhidos apresentem informação espectral mista.

Na análise de resultados, Quarmby et al. (1992) concluíram que componentes puros com características espectrais e ciclos culturais muito

semelhantes (e.g., milho e algodão), após aplicação do modelo de misturas espectrais, conduzem a estimativas deficientes da sua abundância. Também classes de ocupação do solo, com comportamento semelhante (e.g., pastagens e pântanos) apresentam discernimento complexo e de difícil (Zhu e Tateishi, 2001).

Na análise da evolução de índice de vegetação, derivados de MODIS, para a caracterização de ocupação do solo, com análise de árvores de decisão, Matsuoka (2005) identifica problemas na distinção de classes de floresta natural e culturas agrícolas, pois apresentam características espectrais semelhantes.

Ainda, no que concerne à análise dos resultados apresentada pelos diferentes autores, a maioria dos casos apenas é qualitativa sendo omitidos os índices de qualidade (exactidão global, do produtor e do utilizador) e consequentemente uma análise qualitativa dos erros de comissão e de omissão das classes de caracterização do solo obtidas pelos diferentes processos de extração de informação. Contrariando esta tendência, o relatório do CLC2000, (IA, 2005) faz-se uma descrição detalhada dos índices de qualidade e consequentemente da fiabilidade desta cartografia de ocupação do solo. A construção, interpretação e leitura da matriz de confusão e o cálculo dos índices de qualidade é descrito por Caetano (2002) e WC (2004). Verbyla (2004) apresenta como a interpretação dos índices de qualidade e da matriz pode, em termos quantitativos, ser erroneamente interpretado.

Tendo por base os diferentes estudos apresentados, considera-se que a metodologia proposta para a exploração de uma série temporal de imagens MODIS, com 7 bandas espectrais de 500 m de resolução espacial para caracterizar a ocupação do solo, irá permitir superar algumas das limitações identificadas.

O problema do *pixel* misto, será ultrapassado com recurso ao Modelo de Mistura Linear, cujo pressuposto é determinar as componentes

dessa mistura, considerando as assinaturas espectrais dos vários componentes puros utilizados (Shimakazi, 2001).

Ao contrário da abordagem com índices de vegetação, na aplicação do MML a identificação das zonas de transição entre classes de ocupação do solo, será tida em conta, uma vez que, a abundância dos diferentes materiais da superfície será caracterizada por este modelo. Desta forma serão contempladas as zonas de gradiente dos diferentes materiais, independentemente da resolução espacial utilizada.

O efeito topográfico será modelado com base na existência de um componente puro virtual (a sombra), permitindo modelar a geometria de iluminação diferente da cena/imagem. A informação espectral dos diferentes elementos poderá ser desagregada da variação espectral devido ao efeito topográfico (Roberts et al., 1990).

A questão relativa à interpretação dos resultados dos classificadores convencionais será também ultrapassada, uma vez que, o resultado da análise de misturas espectrais traduz-se em proporções (0 a 100%) de cada um dos componentes puros no *pixel*, não sendo por isso necessário pensar em termos de reflectância ou números digitais.

A aplicação do MML a séries multi-temporais permitirá definir, ao longo do tempo, as variações da abundância dos componentes puros no *pixel* e consequentemente permitir, após análise, a identificação de padrões característicos dos diferentes elementos da ocupação do solo.

Após determinadas as características para cada uma das classes de ocupação do solo, com base nas diferentes contribuições de cada um dos três componentes puros, será constituído, através de uma classificação baseadas numa análise de *cluster* de redes neuronais (*Self Organized Maps* – SOM), um mapa dos diferentes elementos que caracterizam a ocupação do solo.

3 - APLICAÇÃO DO MODELO DE MISTURAS LINEARES A SÉRIES MULTI-TEMPORAIS DE IMAGENS MODIS 500

Neste capítulo apresentam-se a base de dados e a metodologia desenvolvida para a exploração de uma série temporal de imagens MODIS com 500 m de resolução espacial. Por ultimo apresenta-se a estratégia desenvolvida para a caracterização da ocupação do solo em Portugal com base numa classificação de uma série multi-temporal de imagens fracção.

Tal como já foi referido anteriormente, o objectivo que se propõe alcançar neste estudo é o explorar uma série intra-anual de imagens do sensor *Moderate Resolution Imaging Spectroradiometer* (MODIS), para a caracterização da ocupação do solo.

Constituiu-se, então, uma base de dados composta por uma série temporal de imagens de compósitos semanais do produto MOD09A1, a qual foi submetida a um conjunto de procedimentos de pré-processamento para a constituição de compósitos mensais. Após a identificação dos componentes puros nas imagens de compósitos mensais procedeu-se à aplicação do MML para determinar a contribuição espectral dos componentes puros - solo, vegetação e sombra - nas imagens multi-espectrais. Numa segunda fase procedeu-se à análise das abundâncias dos componentes puros na série anual de imagens fracção com vista à caracterização de classes de ocupação de solo. Este passo permitiu identificar o comportamento típico (padrão) da variação temporal das classes, em função da abundância dos componentes puros estudados.

Através da análise das imagens fracção e das imagens resíduos, caracterizou-se, com base em amostras criteriosamente definidas, as onze classes de ocupação definidas. Por fim procedeu-se à análise de *clusters*, recorrendo a um algoritmo de redes neuronais não assistida, com o objectivo de classificar as amostras de informação derivada (abundância relativa), em função das classes de ocupação do solo definidas no estudo.

Para a avaliação dos resultados determinaram-se e analisaram-se os índices de qualidade precisão global e exactidão do utilizador e do produtor.

3.1 – Base de Dados

A área de estudo definida para a realização deste trabalho corresponde a todo o território de Portugal continental.

A exploração da série intra-anual de dados para caracterizar a ocupação do solo, utiliza imagens de um sensor de média resolução espacial. Os dados dos sensores MERIS e MODIS, disponibilizados respectivamente pela ESA e pela NASA, tem especificações técnicas apropriadas para este estudo. Apesar dos produtos de nível 1 do primeiro sensor terem uma resolução espacial de 300 m em todas as 16 bandas espectrais, e do segundo variar entre os 250 m e 1 km, em função dos 32 canais espectrais, optou-se por utilizar os dados do sensor *Moderate Resolution Imaging Sepctrometers* (MODIS) essencialmente devido à estratégia de distribuição gratuita adoptada pela NASA.

Elegeu-se assim, o produto MOD09A1 para constituir a série temporal do estudo a desenvolver no âmbito da Dissertação de Mestrado.

O acesso aos diferentes produtos MODIS faz-se através do portal de internet <http://modis.gsfc.nasa.gov>, havendo disponível um repositório de dados em <http://edcimswww.cr.usgs.gov/pub/imswelcome/> (Figura3.1).

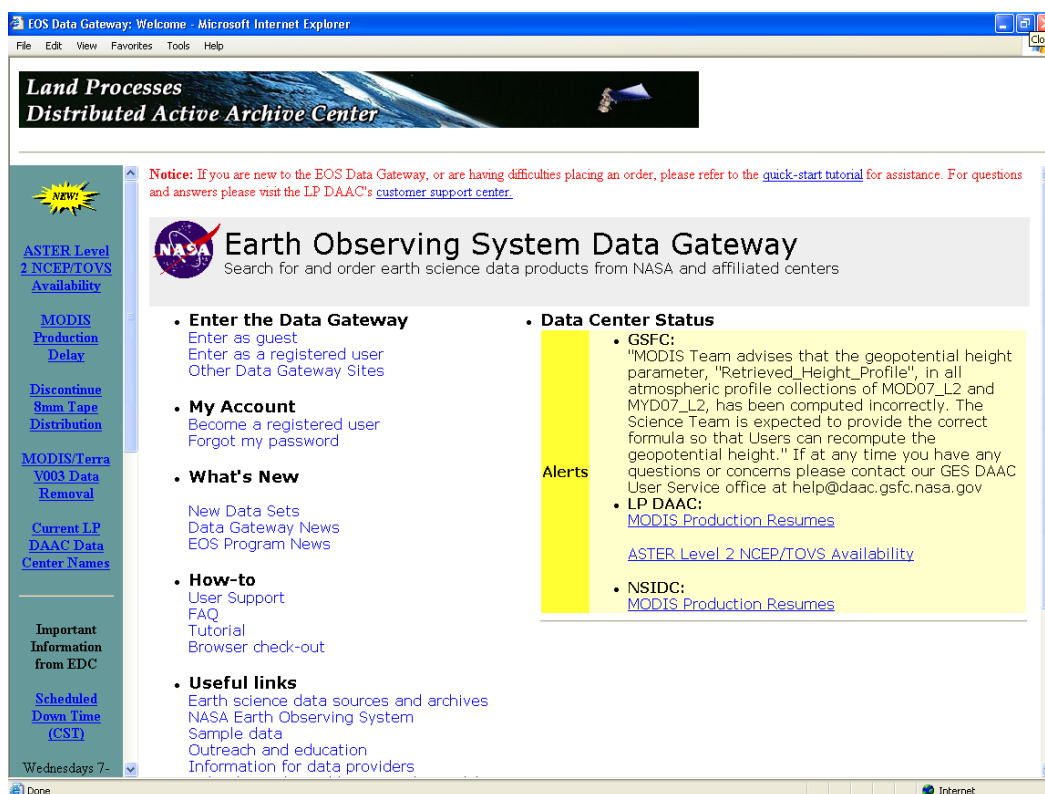


Figura 3.1 – Portal de Internet de acesso aos dados MODIS

3.1.1 - Imagens MODIS Reflectância à Superfície

Para os diferentes produtos foram desenvolvidos, *pelo MODIS Science Team*, uma série de algoritmos para correcção e/ou agregação dos dados obtidos pelo sensor MODIS.

O MOD09A1 é um produto compósito de 8 dias, de nível 3 do programa MODIS, cuja reflectância é determinada à superfície. Este produto apresenta 7 bandas espectrais (Vis+ IVpróx+ IVmédio) e em todas elas uma resolução espacial de 500 m.

A escolha deste produto baseou-se no facto de ser geometricamente e atmosféricamente corrigido, apresentando valores de reflectância calculados para a superfície terrestre. Apresenta, ainda, uma resolução temporal adequada ao estudo em vista, uma vez que, para este, não é

necessário dispor de dados diários. Esta última opção prende-se exclusivamente com a gestão de um grande volume de dados.

No MOD09A1 as bandas 1 e 2, respectivamente, o cdo do vermelho e do infravermelho próximo, foram generalizadas para 500 m, de acordo com a metodologia descrita no documento “*Algorithm Technical Background Document*”(ATBD) do MOD09 (Vermote e Vermueulen, 1999). No Quadro 3.1 encontram-se as especificações espectrais e espacial das 7 bandas que constituem o MOD09A1.

Quadro 3.1 – Resolução espectral e espacial de MOD09A1

Banda (nº)	Banda (cdo)	R. espectral (µm)	R. espacial (m)	Swath (km)
1	Vermelho	0,62 – 0,67	500	2330km
2	IV próx.	0,841 – 0,876		
3	Azul	0,459 – 0,479		
4	Verde	0,545 – 0,565		
5	IV médio 1	1,23 – 1,25		
6	IV médio 2	1,628 – 1,652		
7	IV médio 3	2,105 – 2,155		

Na criação do compósito de 8 dias todas as observações (*pixel*) recebem um índice que reflecte a qualidade da observação, e que varia entre 0 e 5 (inclusive). Estes valores traduzem-se no seguinte:

- 0 – valor de inicialização dos dados e o pior registo de todos;
- 1 – os dados da banda 3 estão indisponíveis;
- 2 – indica dados com nuvens (informação retirada do produto MODIS 1 km *State Quality Assurance* ou *Band Quality Assurance* de 500 ou 250 m);
- 3 – aquisição de dados num ângulo solar zenital elevado (> 75°);
- 4 – dados na área de sombra das nuvens;

- 5 – melhor valor e dados que não correspondem a nenhum dos critérios anteriores.

Nesta sequência os compósitos são gerados dentro da órbita (*within orbit*) e ao longo das órbitas (*across orbit*). No primeiro caso, para todas as observações da mesma órbita, e que pertencem a um intervalo de 8 dias, é seleccionado, para cada *pixel*, o compósito que contém o melhor índice. Uma vez seleccionado o “melhor” compósito em cada órbita, agrupam-se as melhores órbitas.

Juntamente com as 7 bandas do compósito de 8 dias é disponibilizada, *pixel a pixel*, informação de controlo de qualidade (*quality factor - QF*), codificada a 16 bits, onde se apresenta o estado (*flag*) dos parâmetros de qualidade relativos aos algoritmos de correcção dos valores do *pixel*. O Anexo 1 apresenta os possíveis estados deste parâmetro (*flag*) e o seu significado.

Para caracterizar o comportamento anual da ocupação do solo adquiriram-se 86 imagens compósito (8 dias), das quais 8 são do mês de Janeiro de 2001. A razão destas últimas resulta do facto das imagens MODIS estarem disponíveis apenas desde o mês de Fevereiro de 2000.

No Anexo 1 apresenta-se ainda a listagem das imagens de MOD09A1 que foram adquiridas para constituir a base de dados deste trabalho.

3.1.2 - CLC2000

Recorreu-se à utilização dos produtos nacionais do programa IMAGE & CORINE Land Cover (I&CLC2000) como informação auxiliar, uma vez que o programa disponibiliza em cada país da União Europeia dois tipos de dados: (1) conjunto de imagens LANDSAT ETM+, obtidas para o ano 2000 em diversas datas; (2) cartografia nacional de ocupação do solo (CLC2000) à escala 1:100000 e com uma unidade mínima

cartografada de 25 hectares, obtida através de foto-interpretação das imagens de satélite LANDSAT ETM+, ortofotomapas e recurso a dados auxiliares (IA, 2005).

3.2 – Método

A metodologia proposta para atingir o objectivo desta dissertação, está estruturada em seis etapas:

- pré-processamento dos dados e criação de imagens compósitos mensais de MODIS 500;
- definição de uma nomenclatura e selecção de amostras (áreas características)
- definição dos componentes puros e aplicação do modelo de mistura linear;
- caracterização da ocupação do solo com base em perfiz temporais tipo, definidos a partir da análise de informação extraída das imagens fracção, nas áreas amostradas;
- classificação da informação de abundância espectral das varias amostras através da aplicação de um algoritmo de *clustering*, baseado numa rede neuronal;
- avaliação de resultados da classificação através da análise de índices de qualidade.

3.2.1 – Pré-processamento

O pré-processamento das imagens MODIS é constituído pelos seguintes passos: (1) constituição de mosaicos de compósitos semanais para Portugal continental; (2) aplicação de uma projecção cartográfica ao mosaico de imagens compósitos semanais, (3) corte das imagens pela fronteira nacional, através da aplicação de uma máscara com a área de estudo (4) constituição dos compósitos mensais.

3.2.1.1 - Constituição de compósitos semanais

Para cobrir a totalidade da área de estudo (Portugal) estão disponíveis, por cada data do compósito semanal, duas imagens/cenas (*data granule*); uma cobre a parte norte da península ibérica e outra a sul, como ilustra a Figura 3.2.

As imagens MOD09A1 são disponibilizadas no formato *Hierarchical Data Format* (HDF). Este formato é independente da plataforma de software utilizado e é compatível com outros formatos normalizados tendo sido desenvolvido para armazenar grandes volumes de dados, incluindo os metadados associados. Para constituir a série intra-anual para o ano 2000, são necessários 86 *data granule*, cada um com um volume aproximado de 200 Mbytes.

Após a importação das 86 imagens, em formato HDF, para o formato proprietário (IMG - Image) do software utilizado (ERDAS®), procedeu-se à constituição de uma única imagem para cada data (criação de mosaico). Para o efeito recorreu-se a uma função que permite juntar/concatenar várias imagens numa única, de acordo com as suas coordenadas relativas (função *Mosaic* do ERDAS®), obtendo-se um conjunto de dados base composto por 43 imagens de compósitos de 8 dias.

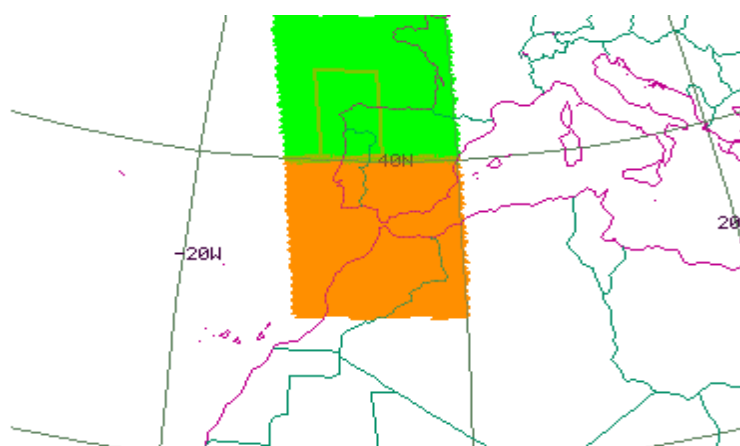


Figura 3.2 – Esquema de ligação das imagens MOD09A1 necessárias para cobrir o território de Portugal (verde – imagem 1; amarelo – imagem 2) (Fonte: NASA,2002)

3.2.1.2 - Aplicação de uma projecção cartográfica

As imagens MOD09A1 estão referenciadas segundo um sistema de coordenadas astronómicas (latitude e longitude). A aplicação de uma projecção cartográfica permite passar de um sistema cartesiano polar, definido com base no elipsóide, para um sistema de coordenadas cartesianas rectangular. Esta passagem faz-se através de um processo de redução da superfície de referencia a dimensões apropriadas (efeito de escala), seguido da planificação desse modelo reduzido, através de formulas de transformação características da projecção utilizada (efeito de deformação para planificação do elipsóide).

Para se trabalhar simultaneamente as imagens MODIS com dados de outras fontes (e.g., CAOP e CORINE), foi necessário uniformizar todos os dados segundo uma única projecção cartográfica. Embora as imagens apresentem um volume de dados superior, optou-se por realizar a transformação do sistema de coordenada na série temporal, pois poder-se-á, no futuro, integrar estas imagens de satélite com outros dados produzidos em Portugal. A projecção das imagens MOD09A1 nas coordenadas do sistema cartográfico utilizado (Sistema Hayford-Gauss, *datum* Lisboa Militar) foi efectuada recorrendo-se à função de transformação de coordenadas.

O quadro 3.1 apresenta os parâmetros do elipsóide, *datum*, factor de escala, origem da latitude e da longitude do ponto central, e distância da falsa origem para o sistema de projecção utilizado para projectar as imagens da série multi-temporal de MOD09A1.

Quadro 3.1 – Parâmetros para projecção das imagens MOD09A1

Parâmetro	Valor
Elipsoide	Hayford
Datum	Lisboa
Factor de escala do meridiano central	1,0
Latitude do ponto central da projecção	8:07:54,000 W
Longitude do ponto central da projecção	39:40:00 N
Falsa origem Este	200000 metros
Falsa origem Norte	300000 metros

3.2.1.3 - Corte de imagem à fronteira nacional

A terceira fase do pré-processamento consistiu na aplicação de uma máscara ao conjunto de dados base, com o limite da fronteira de Portugal continental acrescido de uma zona de influência (*buffer*) de 5 km ao longo da linha de costa (Figura3.3). Para o efeito recorreu-se à álgebra de mapas, realizando a multiplicação da máscara binária (com o limite da área de estudo) por cada uma das imagens dos compósitos semanais. O que se pretende com a aplicação da máscara é a restrição/ corte da informação espectral do conjunto de dados à área de estudo. A zona de influência de 5 km foi incluída na área de estudo para minimizar os problemas de fronteira continente/oceano ao longo da série temporal e para garantir a existência de pixels de água, fundamentais para a aplicação MML, nomeadamente para a simulação do componente puro sombra/água.

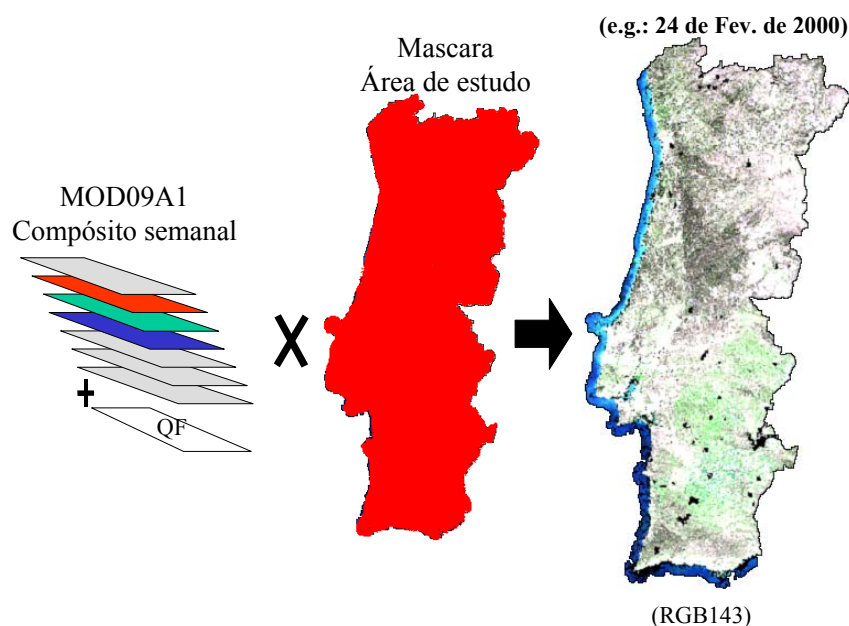


Figura 3.3 – Limitação da área de estudo através da aplicação de uma máscara

3.2.1.4 - Compósitos Mensais

Para identificar padrões específicos dos diferentes objectos da superfície terrestre, com vista à caracterização da ocupação do solo, considerou-se que uma imagem por mês era suficiente, estando, assim, a pressupor-se que se conseguem identificar classes de ocupação do solo exclusivamente com variações espectrais baseadas numa cadência mensal.

Embora a criação dos compósitos mensais permita reduzir o volume e conjunto de dados a processar, o seu principal objectivo é minimizar o efeito das nuvens e dos constituintes da atmosfera nas imagens.

Para a constituição dos compósitos recorreu-se, simultaneamente, ao algoritmo do Compósito do Valor Máximo (Pan, 2003) o qual se baseia no índice de vegetação de diferença normalizada (NDVI), e à análise dos parâmetros do factor de qualidade (*QF*) do *pixel* (Giri, 2000). A figura 3.4 ilustra o processo de criação dos compósitos mensais a partir das imagens dos compósitos semanais de um dado mês.

Para o efeito, definiu-se um algoritmo no software ERDAS® (componente *Modeler* do software) que gera o compósito mensal a partir de todas as imagens do mês. Este algoritmo efectua uma análise *pixel a pixel*, respeitando as seguintes regras, pela ordem em que se encontram:

- valor máximo de NDVI;
- ausência de nuvens e neblina (*cirrus*) no parâmetro de QF.

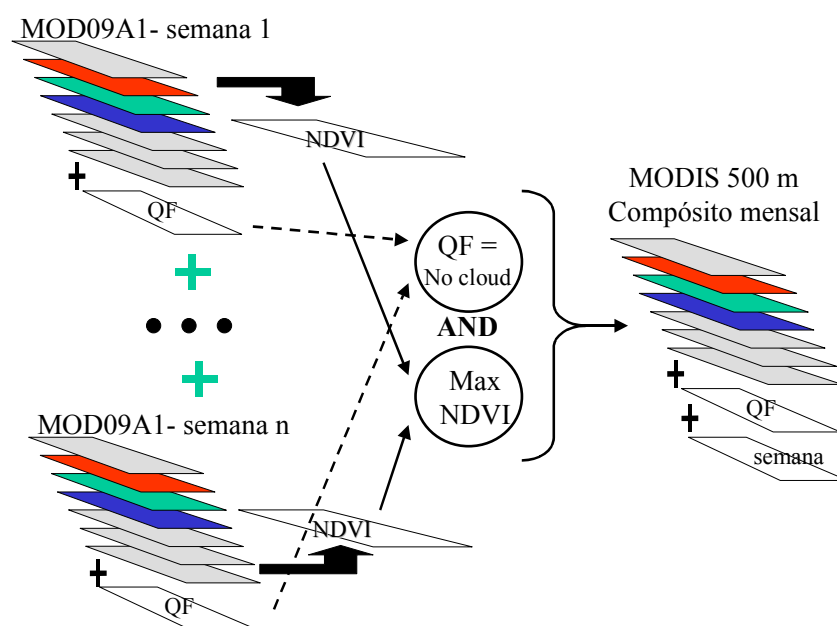


Figura 3.4 – Esquema de criação das imagens compósito mensal

Relativamente à primeira regra, o algoritmo definido calcula o NDVI para cada compósito semanal, e selecciona, *pixel a pixel*, para o conjunto de imagens de um dado mês, o *pixel* com o valor máximo de NDVI. Numa segunda fase é avaliada a presença de nuvens nesse *pixel*.

Para determinar todos os números digitais (DN) que correspondem às zonas com nuvens, recolheram-se os histogramas das imagens para a banda do QF e traduziram-se os números digitais em decimais para binário. Desta forma, determinaram-se todos os valores (DN) do canal QF onde se verifica a ausência de nuvens e/ou neblina. Estes valores foram definidos no modelo por forma a que os pixels contaminados fossem excluídos.

Caso o *pixel* com o NDVI máximo apresentasse nuvens e/ou neblina, ele não seria considerado elegível, e o compósito final terá informação de *NoData*. Optou-se por não recorrer a nenhum método de substituição do *pixel*, uma vez que se pretende realizar uma análise ao subpixel. A utilização de metodologias como análise de vizinhança ou interpolação, poderá alterar o resultado da análise de misturas espectrais.

Como resultado do pré-processamento obtiveram-se 12 imagens (uma para cada mês Fev. 2000 a Jan. 2001) com 9 bandas (7 bandas espectrais + QF + nº do mês). Como pressuposto prouhou-se estudar a caracterização da ocupação do solo durante um ciclo completo (12 meses) e uma vez que o sensor MODIS apenas disponibiliza dados desde Fevereiro, o mês de Janeiro é composto pelas imagens do ano de 2001.

3.2.2 – Nomenclatura e amostras para caracterização do solo

A caracterização da ocupação do solo foi realizada em áreas que foram criteriosamente escolhidas, e que são representativas da classe de ocupação do solo onde se inserem. O processo de identificação das referidas áreas (selecção de amostras) foi procedido pela definição do sistema de classificação (nomenclatura).

3.2.2.1 - Nomenclatura

A definição da nomenclatura adoptada, resultou de uma análise cuidada das especificações de cada uma das unidades dos diferentes sistemas abordados no Capítulo II, e em especial com o CORINE Land Cover. O sistema de nomenclatura adoptado prevê a identificação/diferenciação de classes com um comportamento dinâmico ao longo do tempo, pelo que é possível a caracterização de áreas de vegetação natural e resultante da actividade humana (zonas agrícolas), áreas edificadas, áreas inertes e massas de água. A nomenclatura adoptada assenta nas

unidades do CLC2000, pelo que a relação entre as duas está definida no Quadro 3.2.

O sistema de classificação (nomenclatura) construído para caracterizar a ocupação do solo identifica onze classes, dos quais: 1 corresponde a massas de água, 1 a zonas urbanas e indústria, 1 a zonas improdutivas (rocha, inertes e solo nu), 4 a áreas de vegetação natural (pastagens pobres, matos e espécies subarbustivas, floresta de folhosas e floresta de resinosa), 2 a áreas agrícolas (sequeiro e regadio) e 2 a áreas de alteração do coberto (fogos em florestas e matos e áreas florestais de cortes e novas plantações). Estas duas últimas caracterizam-se por uma descontinuidade/ quebra na componente vegetação durante o período estudado.

Quadro 3.2 - Relação entre as classes da nomenclatura eleita e as classes do CLC2000

Classes do CLC2000	Novas Classe	Designação da nomenclatura adoptada
511 e 512	A	Massas de água
11 e 121	U	Zonas urbanas (contínuas) e industria
131, 331 e 332	I	Improdutivos (rochas, inertes)
231	B	Pastagens naturais
322	M	Matos e espécies subarbustivas
311	F	Floresta de folhosas
312	P	Floresta de resinosas
211,222, 223, 241,242,243 e 244	S	Agricultura de sequeiro
212 e 213	R	Agricultura de regadio
324	H	Cortes e Novas plantações florestais
334	Z	Áreas ardidas de matos e floresta

A Figura 3.5 apresenta, em síntese, as relações entre as unidades de solo do sistema de classificação adoptado neste estudo.

Na nomenclatura adoptada não foi integrada nenhuma classe de áreas húmidas (wetlands) do CLC2000, uma vez que em Portugal estas áreas não tem uma grande expressão.

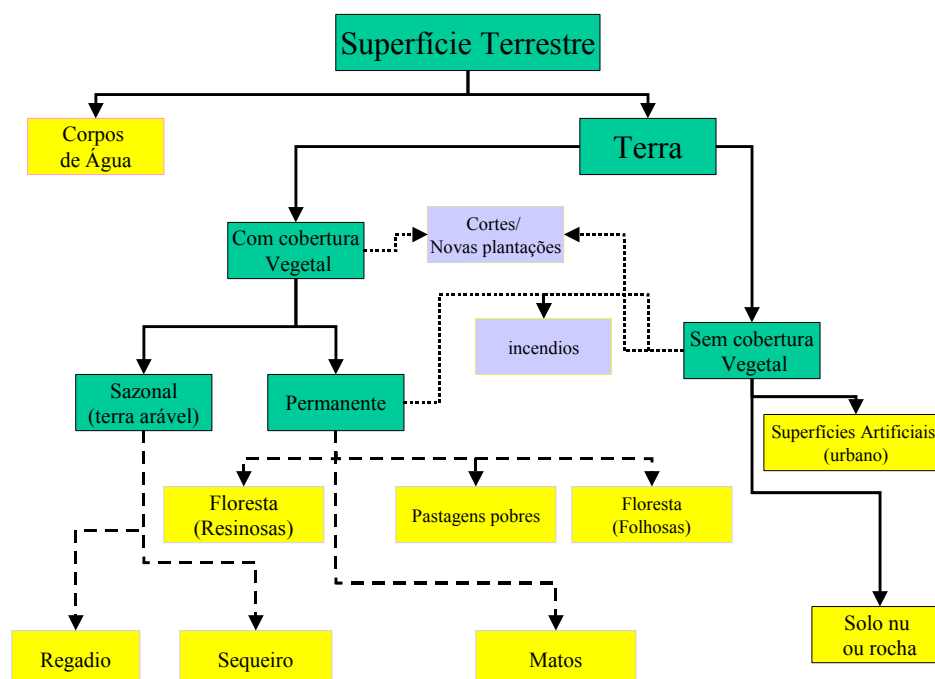


Figura 3.5 – Relação e dependências das Classes de Ocupação do Solo definidas.

3.2.2.2 - Selecção de amostras

A partir das 12 imagens compósitos gerou-se uma grelha/máscara (matriz) que permitiu identificar apenas os *pixels* que, durante toda a série temporal, ficaram sem cobertura de nuvens ou neblina. A constituição desta máscara foi feita recorrendo a uma inquirição dos atributos (*Query*), identificando, para cada pixel, e ao longo dos 12 meses a existência de nuvens. A máscara resultou de uma validação booleana (V/F ou 0/1) da *Query*. Bastava que num mês o pixel apresentasse nuvens, e obter-se-ia um valor falso na *Query*, e consequentemente a máscara teria o valor zero.

Apenas os *pixels* sem nuvens na série temporal (valor V ou 1) podem ser caracterizados ao longo dos 12 meses, e por sua vez a selecção das amostras apenas pôde ser feita dentro desta grelha/máscara.

Com base na relação estabelecida no Quadro 3.1, utilizou-se o I&CLC2000 (imagens de LANDSAT ETM⁺ e unidades taxonómicas do CORINE) para estabelecer um conjunto de amostras para cada uma das classes de ocupação do solo definidas.

Para cada uma das classes de ocupação do solo foram definidas 30 unidades de amostra, exceptuando-se a classe água e improdutivo (15 amostras apenas) devido à dificuldade de identificar áreas puras

A selecção das amostras foi realizada através de interpretação visual e de forma distribuída na área de estudo. Procurou-se garantir que as amostras fossem representativas da classe, independentes entre si e sem correlação espacial. Foi ainda utilizado o critério de selecção do *pixel* central numa matriz 3*3 para minimizar os efeitos e defeitos da georeferenciação das imagens MODIS ao longo da série temporal.

3.2.3 – Modelo de Mistura Linear

Para o desenvolvimento/resolução do Modelo de Mistura Linear consideraram-se três pressupostos:

- os componentes/elementos da superfície são suficientemente largos e/ou opacos de modo a que os fotões (energia EM) apenas interagem com um componente da superfície;
- o sinal recebido pelo sensor é uma mistura linear das reflectâncias dos vários componentes de *pixels* puros (*endmember*), cujos pesos são a percentagem da área do *pixel* que cada elemento apresenta (Adams et al, 1990).
- a soma das abundância dos componentes puros em cada *pixel*, expresso pelas das imagens fracção, é igual à unidade.

A optimização da aplicação do modelo de mistura linear resulta de um processo iterativo, uma vez que a existência de um erro médio quadrático permite avaliar, para cada *pixel*, o desempenho do modelo com base nos componentes puros seleccionados. A Figura 3.6 apresenta o fluxograma do Modelo de Mistura Linear.

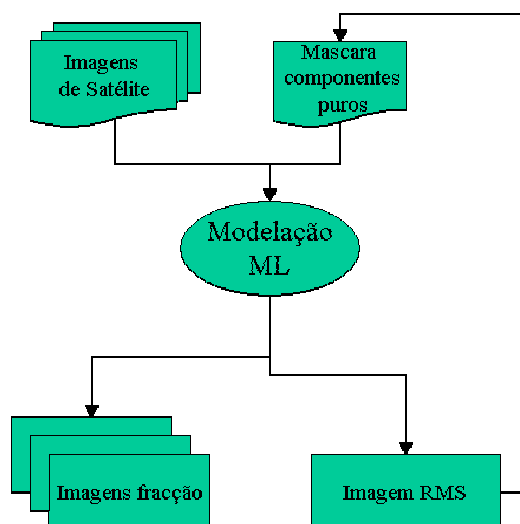


Figura 3.6 - Fluxograma do MML (adaptado de Caetano, 2000)

3.2.3.1 - Selecção de Componentes Puros

Neste estudo recorreu-se à selecção de componentes puros (*endmember*) na imagem pois:

- o uso de componentes puros na imagem é suficiente para atingir os objectivos propostos;
- todas as imagens são do mesmo sensor, e o recurso a componente puros de referência é fundamental, quando se pretende comparar sensores e satélites;
- o uso de componentes de referência conduziria à modelação do efeito da atmosfera, da irregularidade da superfície e do comportamento dos objectos da superfície;

- na área de Portugal continental e à escala das imagens MODIS existem várias zonas onde é possível encontrar materiais da superfície puros.

A utilização de componentes puros de imagem tem ainda a grande vantagem de dispensar dados auxiliares ou de modelação complexa de fenómenos da atmosfera e superfície.

Para a caracterização da ocupação do solo elegeram-se três componentes puros – vegetação, solo e sombra. Este último componente foi modelado com base no comportamento espectral dos corpos de água, pois a sua reflectividade poderia ser alterada, unicamente, devido a variações de iluminação da imagem. A escolha da sombra foi assim utilizada para modelar os efeitos topográficos e diferenças de iluminação.

Cada um dos componentes puros reflecte um comportamento espectral teórico que está caracterizado na literatura. A Figura 3.6 ilustra o perfil espectral teórico e expectável para os componentes puros definidos.

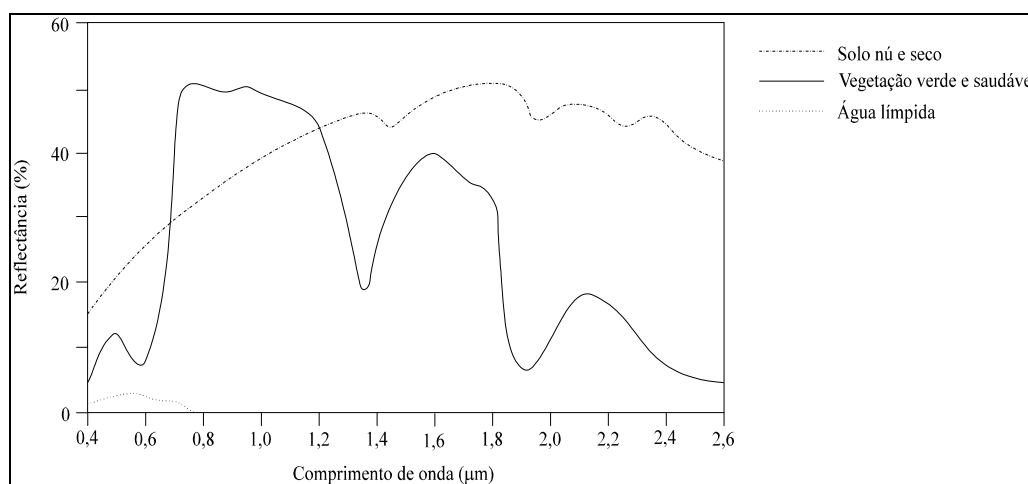


Figura 3.6 - Reflectância dos principais elementos da superfície terrestre (Fonseca, 2004).

A selecção dos componentes puros e respectivas características espectrais, na série multi-temporal de compósitos mensais, foi feita com recurso a um algoritmo desenvolvido em ambiente ERDAS® (componente

Modeler do software). A estratégia adotada visou definir um *pixel*, por cada componente puro, que respeitasse o comportamento apresentado na sua curva espectral teórica (Figura 3.6).

Atendendo às bandas espectrais do produto MOD09A1, e ao comportamento espectral teórico destes três elementos (água, solo e vegetação), estabeleceram-se as seguintes regras:

- Componente puro de vegetação – de entre todos os *pixels* da série temporal, aquele que apresenta o valor máximo de NDVI e simultaneamente o valor mínimo de IV médio (diferente de zero).
- Componente puro de solo – de entre todos os *pixels* da série temporal, aquele que apresenta o máximo valor na banda do IV médio (banda 7) e simultaneamente o mínimo valor absoluto NDVI ($<0,1$ ou 0).
- Componente puro de água /sombra – de entre todos os *pixels* da série temporal, aquele que apresenta, simultaneamente, o mínimo valor absoluto (diferente de zero) de NDVI e IV médio e a reflectância nos canais do cdo do visível (verde, azul e vermelho) inferior a 5 %.

Identificados os três *pixels*, correspondentes aos três componentes puros, estes foram mantidos para as 12 imagens compósito mensais da série temporal estudada, pois só assim as imagens fracção dos diferentes meses são comparáveis. Para garantir esta premissa, as 12 imagens de compósitos mensais foram editadas, modificando três *pixels* com a informação espectral dos componentes puros. Desta forma introduziram-se, de uma forma artificial, os componentes puros nos compósitos mensais que serão utilizados como *input* do modelo de mistura linear.

3.2.3.2 - Algoritmo do Modelo de Mistura Linear

Os *inputs* para o MML são a própria imagem e a assinatura espectral de cada um dos componentes puros (ficheiro de assinaturas). O MML baseia-se no pressuposto de que a radiância de cada *pixel* é uma combinação linear das características de um pequeno e finito número de componentes puros (Smith et al., 1990). É, portanto, um método que não pressupõe a hipótese do *pixel* puro. Matematicamente, o modelo pode definir-se através de:

$$DN_c = \sum_1^N F_n DN_{n,c} + E_c \quad (3.1)$$

onde as seguintes variáveis representam:

DN_c – reflectância da imagem na banda c

N – nº de componentes puros

F_n – fracção do componente n

$DN_{n,c}$ – reflectância do componente n na banda c

E_c – termo erro da banda c em resultado de n componentes espectrais

A expressão (3.1) pode ser escrita sob a forma de matriz para a totalidade do conjunto de M bandas e N componentes puros, e para um dado *pixel*:

$$L = P * F + E \quad (3.2)$$

onde L representa a coluna de radiância da matriz com M linhas, uma para cada banda do satélite; P é a matriz do componente puro (é uma matriz de $M \times N$, cujas colunas são a reflectância espectral do componente puro); F é a fracção do componente puro expresso num vector de N elementos, um para cada componente puro; e E é o vector erro com M elementos, um por cada banda.

Uma vez que os componentes puros são conhecidos, o modelo de misturas estima a abundância de cada componente no *pixel*, minimizando o erro. Matematicamente, o objectivo consiste em determinar o vector F de N

dimensões e o valor mínimo do vector E , de forma a que a relação $L = X_j * P + E$ se mantenha exacta. Assim, X_j é a incógnita a determinar, a qual se pode resolver invertendo a equação, o que se traduz em:

$$X_j = L * P^{-1} - E \quad (3.3)$$

O método do *Constrained Least Squares* (CLS) resolve a equação (3.3) minimizando a soma dos erros quadráticos (equação 3.4), mas sujeitando o modelo de mistura a duas condições de “constrangimento” (Shimabukuro e Smith, 1988):

- para qualquer *pixel* a proporção de valores de X_j não pode ser negativa, traduzindo-se na relação 3.5 ;
- para qualquer *pixel* a soma das proporções deverá ser igual à unidade, condição expressa pela equação 3.6.

$$\sum_{i=1}^M E_i^2 = \sum_{i=1}^M (L_i - (f_1 p_{i1} + f_2 p_{i2} + \dots + f_n p_{in}))^2 \quad (3.4)$$

$$0 \leq F_n \leq 1 \quad (3.5)$$

$$\sum F_n = 1 \quad (3.6)$$

Um outro tipo de algoritmo é apresentado por Caetano, (2000), o qual não obriga a que as proporções das fracções sejam sempre positivas e iguais à unidade. Segundo este autor, relaxando o modelo, chega-se a uma solução mais próxima da realidade, permitindo que existam fracções negativas e outras maiores do que a unidade. Nestes casos poder-se-á detectar, com o próprio resultado do modelo, elementos mais puros que os seleccionados.

Para aplicação da metodologia proposta nesta Dissertação optou-se pela utilização do *software* IDRISI Kilimanjaro®, onde está disponível o algoritmo CLS para a resolução do MML.

Ao contrário do ERDAS® (software que processa imagens de 8, 16 e 32 bits), o IDRISI® apenas consegue aplicar os modelos internos a imagens em formato 8 bits. Foi, assim, necessário realizar uma função de

transformação linear simples de 16 bits para 8 bits, através da função/comando *Stretch* deste software, obtendo-se 12 imagens de compósitos, com 256 níveis de radiância. Foi ainda efectuada a conversão de 16 para 8 bits através de uma transformação Gaussiana. A avaliação de diferentes métodos de conversão de 16bits para 8 bits consiste, em identificar qual deles permite minimizar a perda do poder discriminativo (resolução) da informação espectral após o processo de transformação.

Aos compósitos mensais em 8 bits, contendo os três *pixels* artificiais, com as características espectrais dos componentes puros seleccionados, foi aplicado o algoritmo MML. Da aplicação do modelo resultou a determinação das imagens fracção e imagens (matriz) dos erros médios quadráticos (RMS).

3.2.3.3 - Imagens Fracção

Para cada componente puro seleccionado como *input* do modelo foi gerada uma imagem fracção. A imagem fracção de cada componente puro expressa, *pixel a pixel*, a contribuição que este componente tem na assinatura espectral de um dado *pixel* (Adams et al., 1989). O intervalo de valores deste tipo de imagem varia entre 0 e 1, ou seja entre 0 e 100%. Alguns autores, especialmente quando relaxam o modelo, apresentam uma transformação do intervalo de valores, para uma escala que varia de 100 a 200. Desta forma, quando a imagem fracção apresenta valores menores que 100, indica que há uma sub-modelação, ou seja, há valores originais menores que 0; no caso de ser maior que 200, está-se na presença de uma sobre-modelação, ou seja valores originais maiores que 1.

Neste estudo, a abundância de cada componente puro, ao longo de cada uma das 36 imagens fracção obtidas (12 meses * 3 componentes puros) é expressa através de um intervalo que varia entre 0 a 1.

3.2.3.4 - Erro Médio Quadrático (RMS)

A raiz média quadrática do erro (*root mean squared error* - RMS _E) para uma imagem multi-espectral mede a quantidade da variedade espectral que não é explicada com o MML e com os componentes puros seleccionados.

O RMS do erro pode ser calculado ao nível do *pixel*, através da seguinte relação matemática:

$$RMS = \sum_{k=1}^m \frac{\left(\sqrt{\frac{\sum_{j=1}^n (DN_{jk} - DN'_{jk})^2}{n}} \right)}{m} \quad (3.7)$$

onde:

DN_{jk} – reflectância modelada

DN'_{jk} – reflectância medida no *pixel*

n – nº de bandas espectrais

m – nº total de *pixels* da imagem

j – linha da imagem matriz

k – coluna da imagem matriz

De acordo com a equação anterior podemos concluir que o erro médio quadrático ao longo de todas as bandas, e para um dado *pixel* é dado pela diferença entre o valor de radiância actual do *pixel* e aquele que é esperado se a mistura realmente acontecer no *pixel*. Quanto maior o valor do RMS menor é a “veracidade” da mistura calculada pelo modelo de mistura linear e expressa no *pixel* face à que existe na realidade

O erro médio quadrático também pode ser calculado ao nível da imagem, através do valor médio do erro em todos os *pixels*. Contudo, a média não é um bom aferidor estatístico, não dando informação acerca de quão verdadeiro e refinado é o modelo. Para o RMS ao nível do *pixel*, pode ser gerado uma imagem contendo o respectivo valor de erro em cada um dos *pixels*, estimado ao longo de todas as bandas da imagem original.

As unidades do valor do erro são expressas em números digitais. Valores baixos do erro reflectem a boa construção do modelo e dos seus *inputs*, enquanto que, valores altos exprimem uma incorrecta modelação e consequentemente incorrecta construção dos dados do modelo.

Uma terceira forma para análise do erro é, com base nas imagens resíduo, as quais são geradas uma para cada banda das imagens originais. Estas imagens exprimem a diferença entre a radiância do sensor do satélite e a radiância estimada pela aplicação do modelo de MML (Roberts et al, 1990).

O modelo do software IDRISI ® gera uma imagem de RMS por cada data de imagem compósito, a qual apresenta a estimativa (soma) do erro ao modelar o sinal do pixel ao longo das 7 bandas da imagem original.

3.2.4 - Caracterização da ocupação do solo

A caracterização da ocupação do solo de cada uma das classes da nomenclatura adoptada foi feita segundo duas formas diferentes: a primeira consistiu na determinação de padrões temporais, resultantes das variações intra-anuais das fracções de cada um dos componentes puros estudados; a segunda forma consiste na análise de *clusters* com algoritmo de redes neuronais, com o objectivo de identificar grupos homogéneos de dados de variação da abundância espectral dos componentes puros ao longo dos 12 meses.

3.2.4.1 – Perfis temporais das abundâncias dos componentes puros

Com base na informação contida nas imagens fracção pretendeu-se caracterizar a ocupação do solo em cada uma das 11 classes da nomenclatura definida. Esta caracterização operacionalizou-se três etapas:

- Extracção de informação das amostras nas 12 imagens fracção;
- Identificação de comportamentos não naturais para a classe;

- Análise temporal das variações das três fracções, e caracterização de cada classe de ocupação do solo.

3.2.4.1.1 - Extracção de informação a partir de imagens fracção

Esta etapa consistiu em extrair a informação relativa à variação da abundância dos componentes puros nos *pixels* amostra. Começou-se por recorrer ao software ArcInfo da ESRI®, para extrair a abundância de cada componente puro, por imagem compósito e por classe. Para tal utilizou-se a função *Sample*. Como resultado deste comando obtém-se um ficheiro *ascii*, onde em cada linha é apresentado o identificador do *pixel* (ID), as coordenadas absolutas, e o valor de abundância das 3 fracções em cada um dos meses, ou seja, o valor do *pixel* nas 36 variáveis.

Identificação de comportamentos não naturais para a classe

Apesar do cuidado tido na selecção das amostras, onde se tentou garantir que estas fossem representativas da classe, é de esperar que algumas tenham que ser excluídas pois:

- As classes do CLC são identificadas, para além de imagens de satélite, com dados auxiliares, donde a característica de uma área de uma dada classe poderá resultar de dados auxiliares e não de informação espectral.
- Apenas se utilizou um conjunto de imagens LANDSAT ETM+, de uma única data, pelo que não é garantida que se verifique a dinâmica temporal na área de uma dada amostra.
- Existem amostras que podem ter sido seleccionados em classes do CLC2000 com baixo índice de qualidade global, onde as suas unidades se encontram mal classificadas, e consequentemente não

correspondem, na realidade, à classe que o CORINE define para a unidade.

A selecção das áreas de amostra, baseada na informação do CLC2000 e nas imagens LANDSAT não garante que, por exemplo, uma área de regadio ou de sequeiro, seja efectivamente explorada nesse ano, ou que o corte e plantação de um povoamento florestal tenha sido realizado no período estudado, ou ainda que o fogo tenha ocorrido nesse ano. Estes e outros exemplos traduzem a necessidade de analisar o comportamento de cada uma das amostras seleccionadas face ao comportamento padrão de cada uma das classe de ocupação do solo.

Para garantir um comportamento semelhante, em termos de abundância das amostras, optou-se pela aplicação de um algoritmo de *clustering* (*Self Organized Maps*), para identificar aquelas que apresentam padrões diferentes do “normal” da classe (*outliers*). Desta forma tentou-se garantir que as amostras fossem o mais homogéneo possível.

As classes de fogos em áreas florestais e matos e a de cortes e novas plantações não foram sujeitas a este tipo de análise, pois não é de esperar um comportamento médio (ou normal), uma vez que o acontecimento pode ocorrer em diferentes alturas do ano. A segunda classe referida, ainda pode ter um padrão bastante irregular, uma vez que o corte pode ser faseado no tempo, e a plantação poderá ocorrer apenas no ano seguinte.

3.2.4.1.2 - Caracterização das classes de ocupação do solo

Posteriormente os dados *ascii*, com a contribuição de cada uma das fracções ao longo dos 12 meses, foram integrados em folha de cálculo MS Excel. Os elementos identificados como não tendo um comportamento “normal” face à classe foram referenciados com um parâmetro “NO”, de forma a não ser considerado para a caracterização da classe. Cada

conjunto de amostras, de cada classe, foi estruturado e dividido aleatoriamente em dois conjuntos (treino e validação), através de um programa Macro de MS Excel.

Para cada conjunto de dados da classe, e por mês, foram definidos o máximo, o mínimo, a média, o desvio padrão, a variância e a mediana. Com base na evolução temporal destes parâmetros estatísticos definiram-se curvas típicas e caracterizaram-se as classes de ocupação do solo. Identificou-se o padrão de cada classe e os critérios para distinguir cada uma delas. Para a classe fogos em áreas florestais e matos e para a classe cortes e novas plantações foram identificados padrões característicos, com base em algumas amostras mais representativas. Nestas duas classes não foi possível determinar um perfil médio devido à heterogeneidade de ocorrência do acontecimento no tempo.

3.2.4.2 – Análise de Clusters com algoritmo de redes neuronais (SOM)

Com o objectivo de classificar os dados resultantes do MML (abundância dos componentes puros extraída das imagens fracção), em função das classes de ocupação do solo definidas no estudo, recorreu-se a uma análise de clusters, baseada numa rede neuronal não assistida (*Self Organized Maps*), para agrupar as amostras com características semelhantes.

Neste processo excluíram-se as classes de ocupação do solo de cortes e novas plantações e fogos em áreas florestais e matos, uma vez que não apresentam um comportamento *standard* em termos temporais. Para as restantes 9 classes utilizaram-se, na classificação, todas as amostras anteriormente seleccionadas (incluindo os *outliers*), uma vez que, o objectivo é avaliar a capacidade deste classificador para lidar com a heterogeneidade de cada classe.

O *Self Organized Maps* (SOM) ou mapa auto-organizacional de Kohonen é uma rede neuronal, não supervisionada (não assistida) com

capacidade para agrupar os dados de entrada em aglomerados (grupos ou subconjuntos), sem conhecimento *à priori* de critério de aglutinação dos dados.

A aprendizagem da rede de modo não assistida, ajusta os pesos/ponderações, sem qualquer medida externa de erro. A estrutura de rede resultante reflecte características importantes dos dados de entrada, o que pode ser usado como base para um esquema de classificação. Desta forma, através de um processo de auto-organização, a rede aprende a formar representações internas dos dados de entrada de modo a codificar tais características.

O SOM pode ser visto, esquematicamente, como uma camada de neurónios de *input* e uma camada de neurónios de *output*, formando esta última, uma grelha designada de “mapa topológico”. Cada ponto do mapa representa um neurónio, o qual corresponde a um vector de valores, e cuja dimensão depende do número de variáveis de *input*. Cada neurónio tem um vector de pesos w associado, que inicialmente são distribuídos de forma aleatória pelos dados de entrada (indivíduos). Durante o treino, cada neurónio do arranjo recebe sinais de entrada. Considerando cada um destes sinais, como componentes de um vector x , na iteração t , este vector é designado por $x(t)$. O sinal de saída de um neurónio i é calculado por:

$$D_i(t) = f(\|x(t) - w_i(t)\|) \quad (3.8)$$

onde f é uma função real do tipo gaussiano em torno de zero (função de vizinhança). $\|...\|$ representa o módulo da diferença entre o vector de pesos w e o vector de entrada x .

Os neurónios competem de acordo com uma estratégia do tipo o “vencedor leva tudo”, na qual o neurónio de melhor resposta é seleccionado, isto é, o neurónio de menor desvio/distância. Por outras

palavras, a unidade escolhida é aquela cujo vector $w(t)$ de pesos está mais próximo a $x(t)$.

Cada neurónio possui uma vizinhança, a qual resulta da distância que o separa do indivíduo, definindo a taxa de aprendizagem - $\alpha(d(t))$ (varia entre 0 e 1). O tamanho desta vizinhança - $d(t)$ - e a taxa de aprendizagem são gradualmente reduzidos.

O algoritmo de ajuste pode ser resumido por:

- passo 1: iniciar os vectores de pesos w com valores aleatórios pequenos e a distância máxima (d_{max}) do neurónio ao indivíduo com valor suficientemente grande para incluir todas as unidades; Definir a taxa de aprendizagem.
- passo 2: apresentar à rede cada vector $x(t)$;
- passo 3: em cada iteração t , seleccionar a unidade j de menor desvio $D_j(t)$;
- passo 4: a unidade j vencedora tem seus pesos ajustados por $w_j(t+1) = w_j(t) + \alpha(0) [x(t) - w_j(t)]$
- e cada unidade i , a uma distância $d(t)$ da unidade j menor do que d_{max} , tem seus pesos ajustados por
- $w_i(t+1) = w_i(t) + \alpha(d(t)) [x(t) - w_i(t)]$,
- onde $0 < \alpha(d(t)) < 1$ é uma função decrescente com a distância $d(t)$;
- passo 5: diminuir d_{max} (raio da vizinhança) e actualizar taxa de aprendizagem;
- passo 6: repetir a partir do passo 2 para todos os vectores de entrada.

Assim, é produzido um mapa topográfico bidimensional do espaço de dados de entrada para a rede, no qual cada indivíduo é identificado com o vector sinal de entrada, que obteve melhor resposta no treino. Isto

permite que os indivíduos sejam agrupadas em regiões ou grupos, em que a resposta para um dado vector de entrada é máxima.

Este modelo encontra-se programado num Macro de *Visual Basic for Applications* (VBA) desenvolvido em MS Excel e está disponível em <http://www.geocities.com/adotsaha/>.

Numa primeira fase, para 242 amostras (excluíram-se as amostras da classe Z e Classe H), definiu-se a taxa de aprendizagem inicial de 0,99 e a final de 0,1; a distância da vizinhança da função Gaussiana tem uma taxa de convergência inicial de 50% e final de 1%; o número de iterações da rede é 100 e os dados foram apresentados à rede de forma aleatória.

Para cada *cluster* obtido, foi associada uma classe de ocupação do solo potencial, em função do maior número de amostra das classes iniciais.

No caso de *clusters* complexos, e sem uma tendência de ocupação do solo bem definida, foram identificadas as amostras que o constituíam, e posteriormente, apenas estas foram novamente apresentadas à rede. Neste caso mantiveram-se as condições dos parâmetros, alterando-se apenas o agrupamento dos dados em 10 *clusters*. Foi ainda estudada e realizada a possibilidade de associar as amostras das classes de ocupação do solo com o objectivo de melhorar a classificação.

3.2.5 – Avaliação da qualidade da classificação

De acordo com o resultado da agregação das amostras para cada classe de ocupação do solo, obtido no processo de clustering, procedeu-se a avaliação da qualidade do mapa obtido da classificação. Para o efeito construiu-se a matriz de contingência (matriz de erro ou confusão) com base nas amostras definidas. Nesta, os elementos da diagonal principal exprimem a concordância entre os dados classificados (mapa) e os dados de referencia (realidade). Foi ainda possível identificar quer o número de amostras que não foram atribuídas a uma dada classe (erros de omissão)

quer ainda aquelas que foram incorrectamente atribuídas a uma classe (erros de comissão).

De seguida calcularam-se os seguintes índices de qualidade da classificação e a precisão de cada uma das classes:

- Exactidão global – proporção de unidades de amostras classificadas correctamente face ao total de amostras;
- Exactidão do produtor – proporção de unidades de amostras correctamente atribuídas à sua classe;
- Exactidão do utilizador – proporção de unidades de amostras que pertencem à classe a que forma atribuídas.

O facto da classificação realizada não utilizar uma fase de treino (como na classificação assistida) não impossibilita a utilização das amostras classificadas no processo de avaliação de exactidão temática.

Com base nos resultados das diferentes etapas da metodologia proposta, concluiu-se acerca da efectividade do método para a caracterização da ocupação do solo.

4- RESULTADOS E DISCUSSÃO

A apresentação e discussão dos resultados está repartida em cinco secções. A primeira resulta da análise do resultado do pré-processamento; a segunda analisa a selecção dos componentes puros utilizados para *input* do modelo de mistura linear; a terceira secção corresponde ao resultado da aplicação do MML, ou seja, à análise das imagens fracção e das imagens de erro médio quadrático; a quarta secção analisa os padrões temporais para as diferentes classes de ocupação do solo e apresenta a sua caracterização; a ultima secção apresenta o resultado da análise de clusters e a avaliação da qualidade desta classificação.

4.1 – Série mensal de imagens MODIS

O algoritmo de Compósito Valor Máximo, a análise do estado da *flag* da banda do Quality Factor, relativo a contaminação com nuvens, e a limitação da área de estudo com uma máscara, produziram os seguintes resultados:

Excluíram-se pixels da área de estudo, pelo facto de estes estarem parcialmente ou totalmente contaminados com nuvens ao longo da série de imagens de compósitos semanais, utilizadas na criação do compósito mensal. O Quadro 4.1 apresenta o número de pixels (número absoluto e percentagem) que não são utilizados, em cada mês da série intra-anual e dentro da área de estudo, devido à existência de nuvens. Como se pode inferir do quadro anterior o mês de Abril de 2000 apresenta cerca de um quinto da área de estudo contaminada com nuvens, o que limitará algumas etapas deste estudo.

A localização geográfica, em cada mês, dos *pixels* contaminados com nuvens é alterada perante a variação da disposição das mesmas no território nacional, razão pela qual se definiu uma máscara binária indicando todos os pixels, que durante toda a série intra-anual, não

estiveram contaminados por este elemento atmosférico, ou que não tem anomalias referentes ao deficiente funcionamento do sensor. Esta máscara, definida com base na banda do *Quality Factor*, permitiu, identificar todas as possíveis áreas de estudo para a caracterização da ocupação do solo, ou seja, elegíveis como potenciais áreas de escolha na selecção das amostras. A figura 4.1 ilustra o critério utilizado e o resultado desta análise sob a forma de máscara binária, onde aos pixels contaminados lhes foi atribuído o valor de “nodata” (representado a preto).

Quadro 4.1 – Pixels contaminados com nuvens nos compósitos mensais

Mês	Total Pixels / composito	Pixels usados (nº)	Nº de Pixels contaminados	% pixels contaminados
Fev – 00	454745	434717	20028	4,40
Mar – 00	454745	440750	13995	3,08
Abr – 00	454745	372543	82202	18,80
Mai – 00	454745	439876	14869	3,27
Jun - 00	454745	444122	10623	2,34
Jul – 00	454745	444142	10603	2,33
Ago – 00	454745	447228	7517	1,65
Set – 00	454745	438312	16433	3,61
Out – 00	454745	436761	17984	3,95
Nov – 00	454745	435531	19214	4,23
Dez – 00	454745	425886	28859	6,35
Jan - 01	454745	443617	11128	2,45

Uma análise pormenorizada dos *pixels* contaminados e não contaminados permite-nos afirmar que, para uma análise multi-temporal, a presença de nuvens restringiu a área efectiva de estudo para 76,3% da inicialmente definida, ou seja, embora o compósito do mês de Abril apresente 18,3 % de área contaminada, na globalidade dos compósitos mensais, o território teve uma cobertura de nuvens em 23,7%.

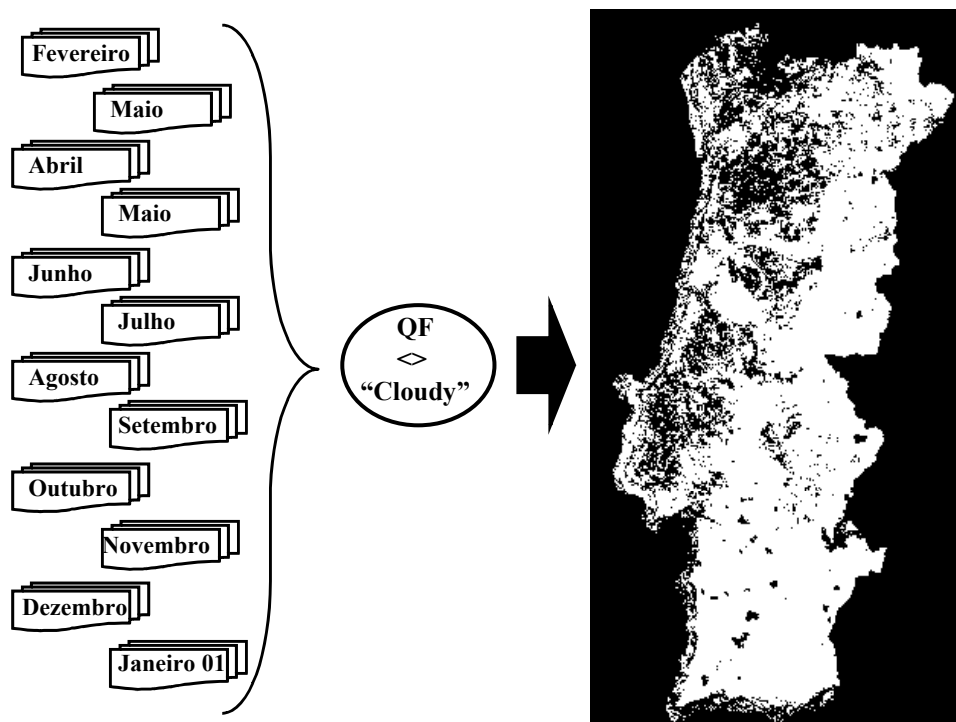


Figura 4.1 – Pixels não contaminados com nuvens na área de estudo

Na Figura 4.2 representam-se os histogramas da banda do *Quality Factor* nos 12 compósitos mensais, resultantes da metodologia de pré-processamento das imagens de MOD09A1. De acordo com os valores das classes do histograma, e com as definições dos parâmetros do *Quality Factor* (Anexo 1), os compósitos apenas apresentam *pixels*, onde não existem nuvens, tal qual se exemplifica com os valores de ND iguais a 72, 74 e 136.

O Quadro 4.2 apresenta a conversão destes valores de decimal para binário. De acordo com o Anexo 1 e com o documento MOD09A1 ATDB, disponível em <http://modis.gsfc.nasa.gov>, aferiu-se que o parâmetro (*flag*) relativo à presença de nuvens, obtido pelos histogramas, apresenta o valor zero, o que indica a sua ausência.

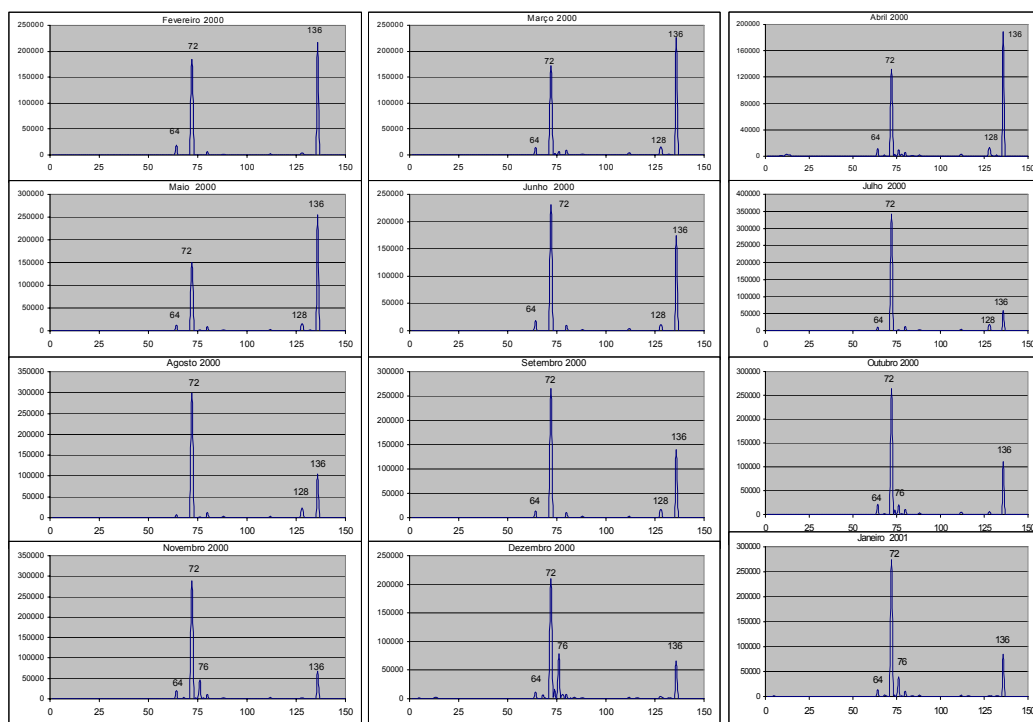


Figura 4.2 - Histogramas da banda do *Quality Factor* nos compósitos mensais

Quadro 4.2 – *Quality Factor*: Conversão de DN para binário

DN (Décimal)	Binário (16Bits)
72	0000000001001000
74	0000000001001010
136	0000000010001000

Um segundo ponto que resulta da metodologia de pré-processamento é a própria criação dos compósitos mensais. Com base nos critérios definidos e nas imagens de compósitos semanais disponíveis para cada mês, foram geradas 12 imagens de compósitos mensal (Fevereiro de 2000 a Janeiro de 2001), com 9 bandas de informação, das quais 7 contém a informação espectral, 1 apresenta o valor do *Quality Factor*, e 1 é indicativa do número da imagem semanal donde provem cada um dos pixels que compõem o compósito mensal.

Estas 12 imagens são o ponto de partida para se atingir o objectivo proposto nesta Dissertação e o primeiro resultado efectivo obtido com os

métodos propostos. Seguidamente apresentamos uma análise temporal, baseada exclusivamente numa interpretação visual dos resultados obtidos.

Para tal ilustram-se uma composição colorida (RGB 214) com a evolução temporal dos compósitos mensais (Figura 4.3). Para avaliar o desenvolvimento e o vigor da vegetação atribui-se à cor vermelho a banda do infravermelho próximo, o verde à banda do vermelho e o azul à banda do verde, pois nesta composição colorida é possível diferenciar as espécies vegetais naturais e conduzidas pelo homem do resto da paisagem envolvente.

No que concerne à ausência de pixels, devido à contaminação destes por nuvens, verifica-se que o mês com maior número de “falhas” foi o de Abril de 2000.

A análise das imagens compósito permite-nos assim, verificar a evolução do desenvolvimento da vegetação, entre os meses de Inverno/Primavera, na zona norte do país e nas áreas de montanha do centro e sul. Esta evolução traduz-se pelo aumento e/ou predominio da reflectividade na banda do infravermelho próximo. Este comportamento pode-se observar nos meses de primavera, em particular nas planícies alentejanas onde florescem as pastagens e vegetação natural e as culturas cerealíferas de sequeiro

Nos meses de Verão verifica-se ainda a presença das áreas regadas (grandes regadios colectivos), onde as pequenas áreas com reflectividade elevada na banda do IVprox. são circundadas por áreas de sequeiro, o que traduz o desenvolvimento vegetativo das culturas regadas

Ainda na série temporal das composições coloridas é possível identificar as áreas florestais do centro do País, devido ao comportamento relativamente homogéneo (reflectividade da vegetação), ao longo do tempo, na banda do infravermelho próximo.

O comportamento anteriormente descrito, para áreas com vegetação, era expectável, uma vez que, os compósitos são gerados com

base no valor máximo do índice de vegetação do *pixel*. Este critério elege os *pixels* que maior abundância de vegetação têm, na série dos dados mensais.

4.2 –Seleccção de amostras

Com base na metodologia proposta para a selecção das amostras para caracterização das 11 classes de ocupação do solo, identificaram-se 304 áreas. A Figura 4.4 apresenta a selecção das amostras, utilizando uma composição colorida RGB432 de LANDSAT ETM+ e a Figura 4.6 apresenta a distribuição, no território nacional, das amostras identificadas para cada classe de ocupação do solo adoptada.

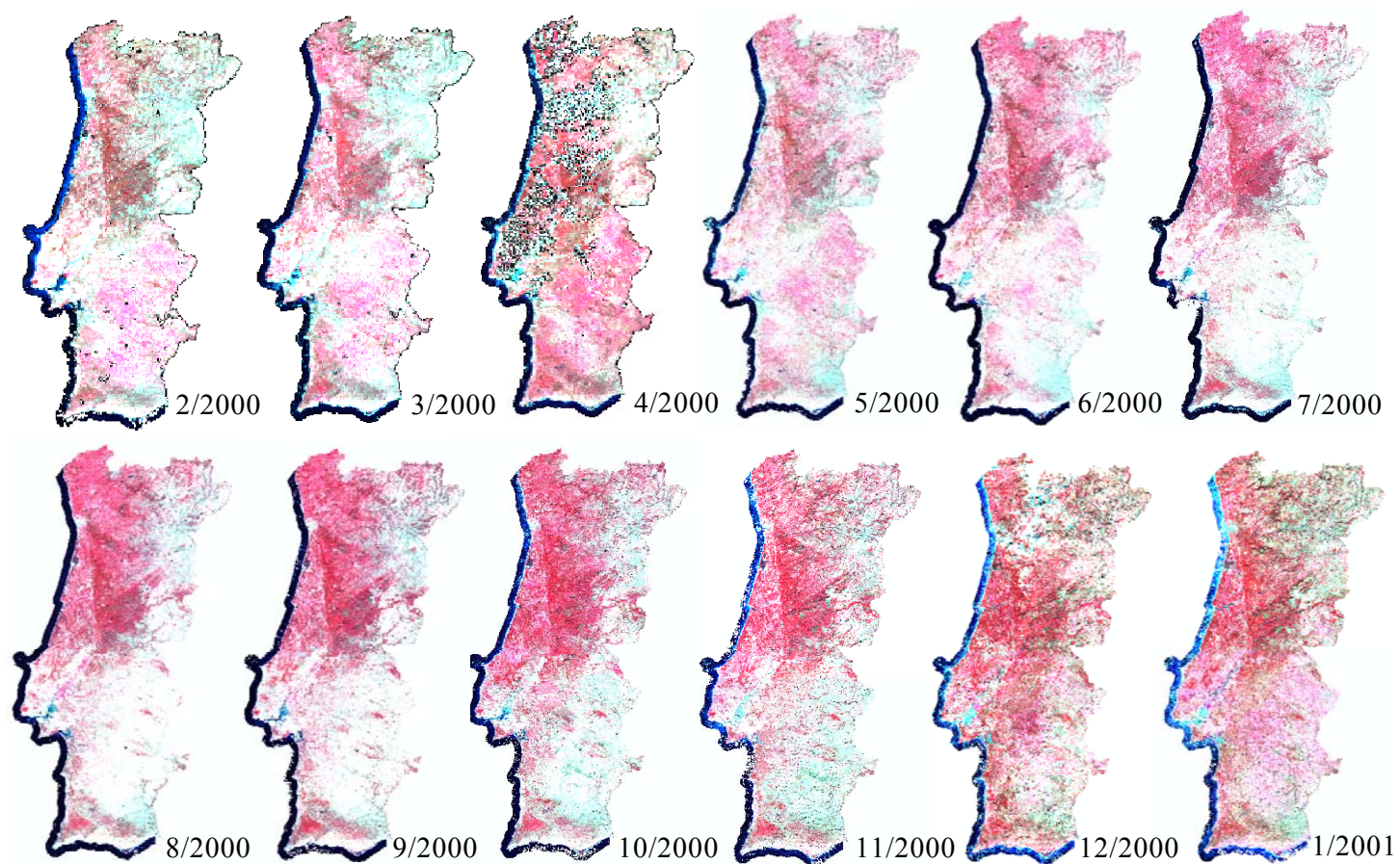


Figura 4.3 – Composição colorida RGB214 dos compósitos mensais.

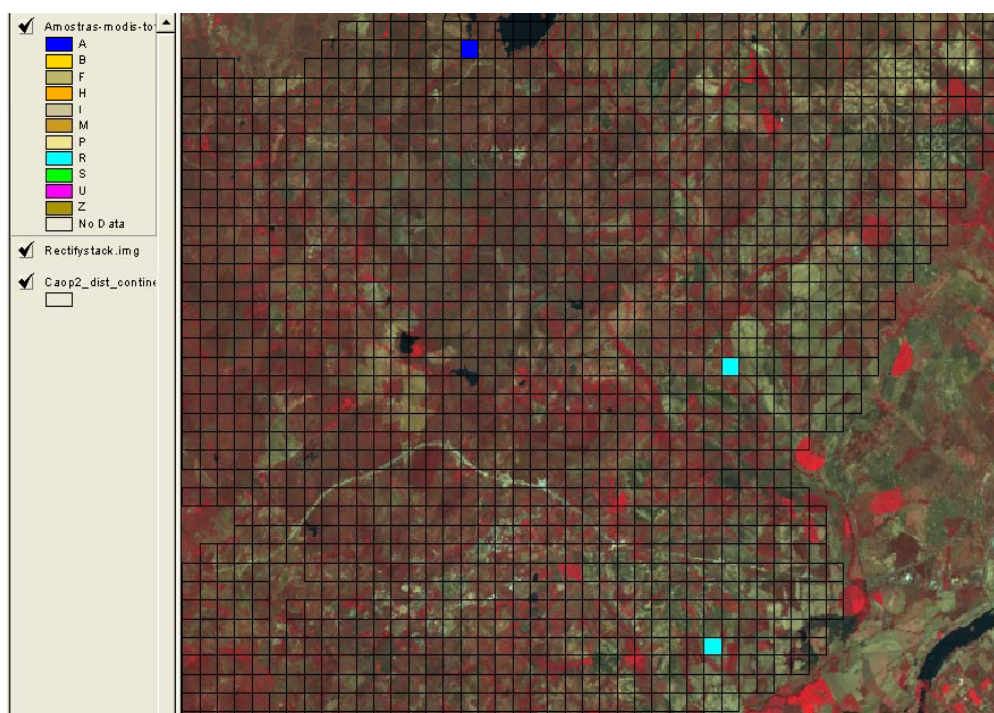
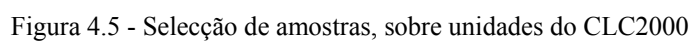


Figura 4.4 - Selecção de amostras em, na malha de 500 por 500 m, com composição colorida RGB 432 de LANDSAT 7 ETM+

A selecção das amostras foi feita atendendo à máscara que identifica os pixels que podem ser acompanhados/estudados, durante toda a série intra-anual (Figura 4.2) e às unidades do CLC2000. A identificação de amostras, sem a presença de nuvens, e que se encontrem como *pixel* central de uma malha de 3x3 de uma zona com uma ocupação do solo homogénea, mostrou ser uma tarefa difícil de executar (Figura 4.4).

Apesar deste constrangimento, procurou-se, durante a selecção das amostras, garantir que não existisse correlação espacial entre as amostras. Assim, procurou-se distribuir, de uma forma uniforme, as amostras no território nacional (Figura 4.6), sem se seleccionar amostras dentro da mesma unidade do CLC (Figura 4.5).

A escolha das amostras não resultou de um processo aleatório, mas pelo contrário, de um processo exaustivo, onde se tentou encontrar *pixels*, que através de interpretação visual, se situavam em zonas homogéneas



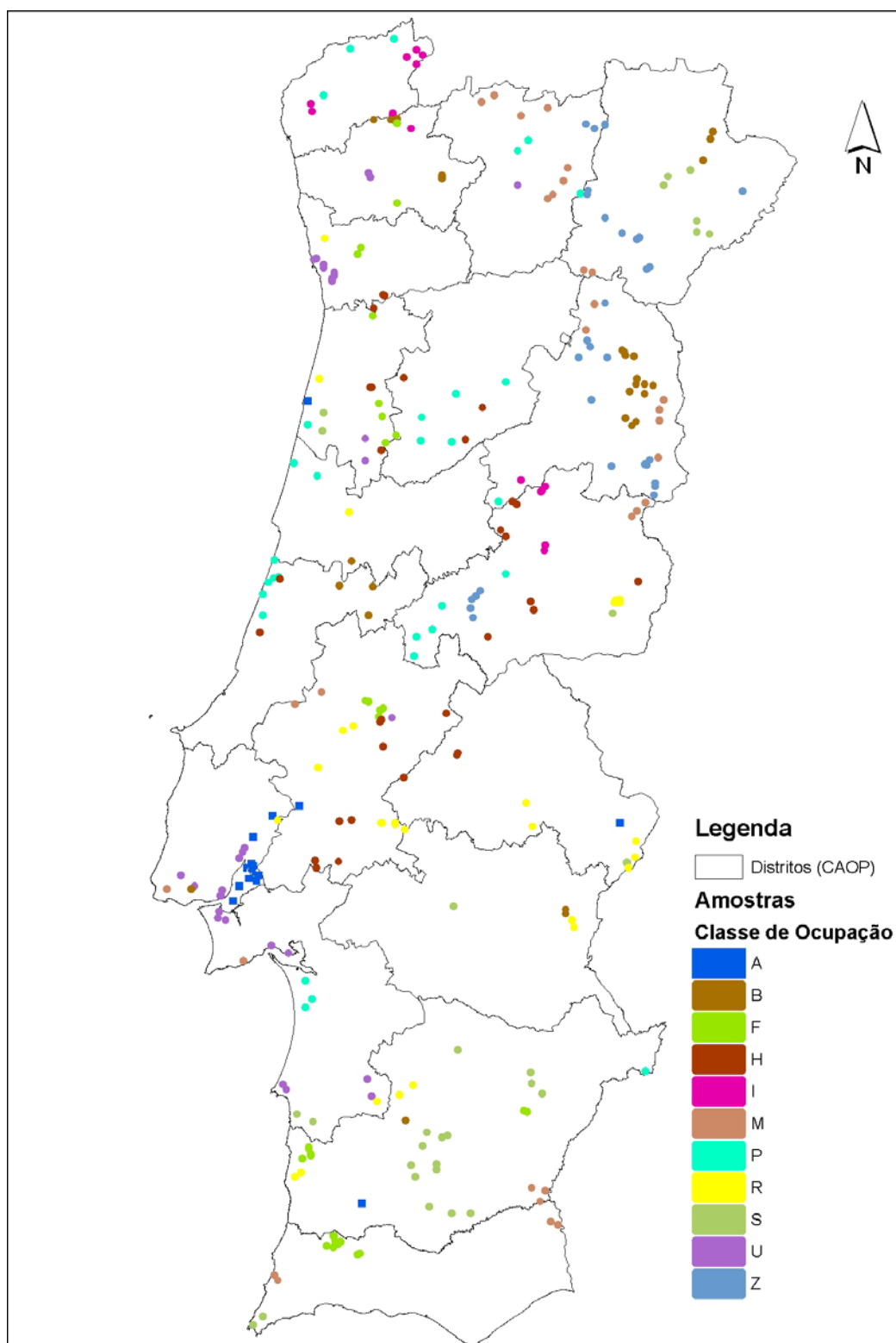


Figura 4.6 - Distribuição, no território nacional, das amostras

4.3 – Selecção dos componentes puros

De acordo com a metodologia descrita, seleccionaram-se os componentes puros através do método iterativo.

O componente puro vegetação foi identificado, na imagem de Junho de 2000, na zona dos pivots de rega do concelho do Montijo. O componente puro solo foi identificado, na imagem de Março de 2000, na zona das planícies alentejanas do concelho de Aljustrel. Para o componente puro água foram identificados vários *pixels* na zona da costa portuguesa (oceano), todos com as mesmas características espectrais, pelo que foi seleccionado aleatoriamente um dos *pixels*.

O Quadro 4.3 apresenta os valores de reflectância, expressos em números digitais e em percentagem, dos componentes puros seleccionados e a Figura 4.7 apresenta a reflectância (%) nas bandas de MOD09A1.

Quadro 4.3 -Reflectância espectral (DN e %) dos componentes puros seleccionados

	Vegetação		Solo		Sombra/água	
	%	DN	%	DN	%	DN
azul (b 3)	1,7	167	8,9	873	3,3	320
verde (b 4)	4,0	389	13,9	1359	2,4	237
Vermelho (b 1)	0,0	2	18,6	1816	0,6	60
IV Prox. (b 2)	41,5	4051	29,4	2865	0,6	62
IV M1 (b 5)	32,5	3171	39,5	3851	1,2	115
IV M2 (b 6)	15,8	1540	46,5	4535	1,1	111
IV M3 (b 7)	5,7	558	41,0	3999	0,5	53

A análise da Figura 4.7 permite inferir que os componentes puros seleccionados representam correctamente os elementos puros água, solo e vegetação, pois o perfil traçado é semelhante às curvas teóricas determinadas para estes elementos (Figura 3.6).

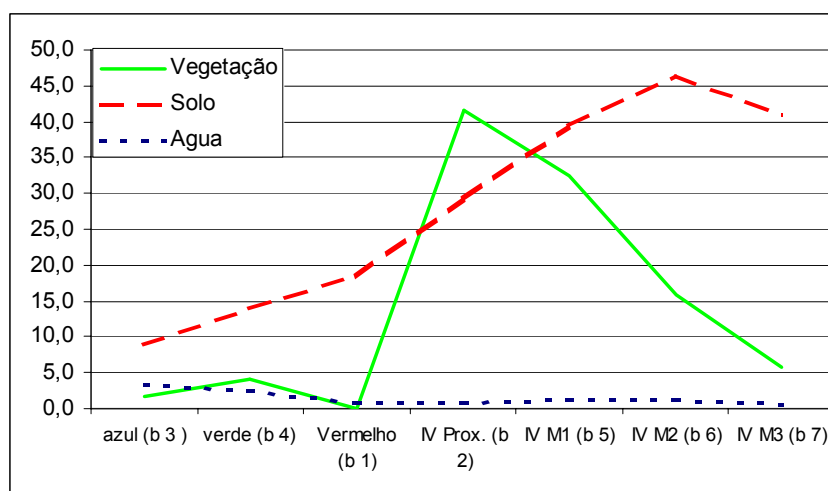


Figura 4.7 – Reflectância (%) dos componentes puros seleccionados, nas bandas do MOD09A1

Resumidamente podemos caracterizar os componentes puros da seguinte forma:

- Vegetação – apresenta baixa reflectância (< 5%) nas bandas do visível (cdo do azul, do verde e do vermelho) e valores elevados na banda do Infravermelho próximo. Dentro dos canais do visível, a banda do vermelho apresenta a reflectividade mais baixa, enquanto que a banda do verde tem o valor mais elevado (explicando a cor verde das plantas). Dentro da gama do Infravermelho médio, o canal de menor cdo tem valores de reflectância elevados. Nesta gama de cdo, a reflectividade diminui à medida que diminui a energia.
- Solo – À medida que aumenta o cdo, do visível para o infravermelho médio, aumenta reflectividade. Exceptua-se a reflectividade no último canal do produto MOD09A1.
- Água – Apresenta valores muito baixos de reflectividade, nas várias bandas do infravermelho. Apesar de também apresentarem valores muito baixos de reflectância nos canais do visível, é a banda do azul que tem o valor mais alto (explica a cor azul do Oceano).

Através da edição de cada uma das imagens de compósitos mensais, foram inseridos três pixels artificiais, cada um com a assinatura espectral do respectivo componente puro, pois a modelação de todos os compósitos é feita com os mesmos componentes puros.

4.4 – Resultado da aplicação do MML

Para se proceder à aplicação do modelo de mistura linear, procedeu-se à conversão de cada um dos compósitos mensais, do formato *Unsigned 16Bits*, para *Signed 8Bits*, de acordo com o definido na metodologia. A análise dos resultados, mostrou que não havia ganho de resolução, quando se recorria à função que melhor descrevia o histograma de distribuição dos DN da imagem (transformação de função gaussiana) face à utilização de uma transformação linear. Verificou-se, ainda, que qualquer que fosse a transformação aplicada, esta conduzia sempre à perda de resolução de discriminação nas assinaturas espectrais dos componentes puros. Face ao exposto, o método utilizado para transformar as 12 imagens compósitos foi o da transformação linear simples.

Como resultado da aplicação do MML às 12 imagens de compósito mensais, com as assinaturas espectrais dos três componentes puros vegetação, solo e sombra, obtiveram-se 36 imagens fracção e 12 imagens de erro médio quadrático.

4.4.1 – Imagens fracção

Para se proceder à interpretação das imagens fracção, constitui-se, mensalmente, uma composição colorida RGB da fracção solo, vegetação e sombra/água, ou seja atribuiu-se a cor vermelha à componente solo, a cor verde à componente vegetação e a cor azul à componente sombra/água. As Figuras 4.8 e 4.9 ilustram a referida composição colorida para os 12 meses estudados, contudo ressalva-se que para os *pixels* onde o MML não

foi aplicado (onde o QF indicava problemas) estes encontram-se a branco (transparente).

A interpretação da composição colorida RGB solo/vegetação/sombra permite concluir que:

- Nos meses de Fevereiro, as grandes zonas agrícolas do Ribatejo e dos grandes perímetros de rega colectivos do Alentejo, apresentam uma elevada percentagem de fracção solo, pois é a altura de preparação do terreno e da sementeira das culturas de Primavera/Verão. Nestas mesmas zonas, identifica-se o aumento da fracção de vegetação nos meses de Maio a Agosto. Neste último o contraste da fracção vegetação com as unidades de paisagem envolvente é bastante evidente.
- Nos meses de Fevereiro a Maio as zonas de vegetação natural, grandes pastagens e zonas de cereais de sequeiro, que caracterizam a paisagem do Alentejo são bem evidentes, e estão representadas pelo verde da composição colorida. A ceifa dos cereais, a senescência da vegetação natural e das pastagens, reflecte-se no aumento da componente solo, o que se traduz na passagem de verde a vermelho da composição colorida;
- A norte do rio Tejo, a variação temporal da vegetação é evidente, verificando-se que, nos meses de Primavera, onde o desenvolvimento vegetativo é superior aos outros períodos, há um aumento da intensidade de cor verde na composição colorida. A diminuição de intensidade é evidente nos meses de Verão, acompanhada com um aumento da tonalidade azul (ensombramento). Esta situação resulta do facto da floresta marcar as características da paisagem a norte do rio Tejo;

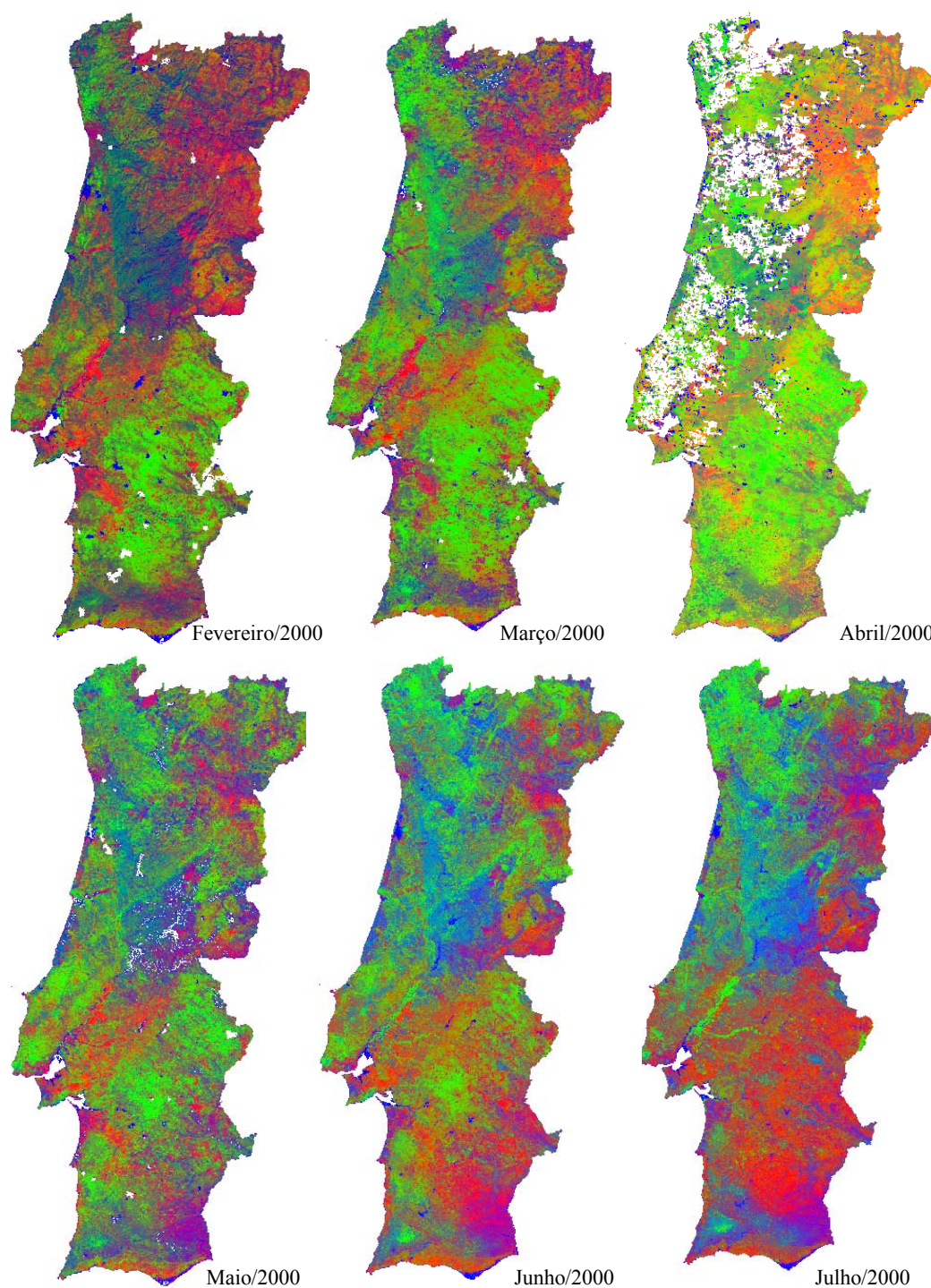


Figura 4.8 - Composição RGB de imagens de fracção solo, vegetação e sombra na série intra-anual de Fevereiro a Julho de 2000

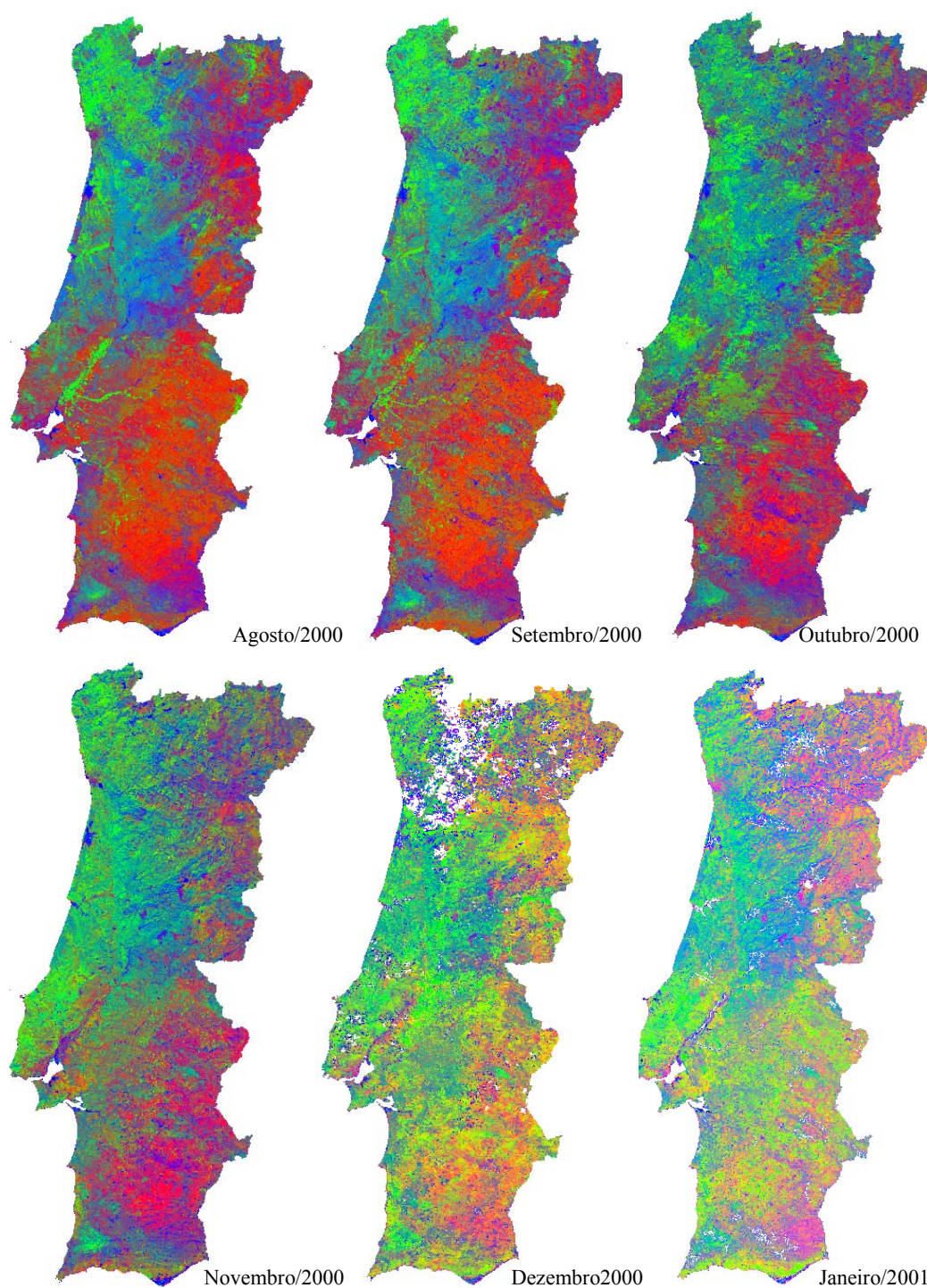


Figura 4.9 - Composição RGB de imagens de fracção solo, vegetação e sombra na série intra-anual de Agosto de 2000 a Janeiro de 2001

- A variação da vegetação natural nas unidades da paisagem das “terras quentes” de Trás-os-Montes é bem evidente, após uma

análise temporal das composições coloridas. A cor laranja/vermelho, durante os meses de Outono/Inverno, deste território interior de Portugal, poderá explicar-se com base na floresta de folha caduca, característica desta região, onde a senescência da vegetação é acompanhada por um aumento da percentagem do solo exposto.

- As zonas urbanas de Portugal apresentam sempre a mesma tonalidade durante toda a série temporal da composição colorida. Estas áreas (e.g.: Lisboa e Porto) são facilmente identificadas no período de maior desenvolvimento vegetal das zonas envolventes;
- Por fim, da análise visual efectuada aos compósitos, verifica-se que as regiões montanhosas de Portugal, nomeadamente as da cordilheira central (Meseta Ibérica) e a cadeia montanhosa do Algarve (serras de Monchique e do Caldeirão) apresentam uma percentagem elevada de sombra (tom azulado), o que traduz a correcta modelação da componente sombra (*shadow*) na simulação do efeito da diferente geometria de iluminação das imagens de compósito e o efeito topográfico.

Assim, conclui-se que as variações do comportamento temporal das fracções, permitem inferir a existência de diferentes classes de ocupação do solo, com dinâmica temporal, e permitiu ainda modelar o efeito topográfico e de iluminação das zonas, o qual é visível pela interpretação das zonas de montanha.

4.4.2 – Imagens do erro médio quadrático

Para avaliar a capacidade do MML explicar a variabilidade espectral das imagens MODIS, foram analisadas as imagens do erro médio quadrático de cada mês (Figuras 4.10 e 4.11). Estas estão representadas numa escala 8 *bits* de tons de cinzentos, onde a cor escura representa erro

de modelação elevado e cor clara erro reduzido. As imagens estão expressas em reflectância (%)

A interpretação das referidas figuras permite afirmar que não existem padrões associados à modelação das misturas espectrais, exceptuando-se os meses de Fevereiro e Novembro, onde existe um padrão espacial associado ao varrimento (scanner) da superfície terrestre pelo sensor.

As imagens, em geral, apresentam um erro médio quadrático baixo (na ordem dos 2% em reflectância) evidenciando uma correcta modelação dos componentes puros escolhidos para explicar a variabilidade espectral das 12 imagens compósito. Nas imagens de Março, Agosto e Setembro o erro de modelação na zona do Alentejo apresenta valores um pouco mais altos (reflectância com valores num intervalo entre 4% e os 5%). Sendo a zona alentejana caracterizada pela existencia de grandes áreas de vegetação, este aumento do valor do RMS médio das 7 bandas, evidencia que o componente puro vegetação seleccionado, por ser genérico para toda a vegetação natural e conduzida pelo Homem (floresta, pastagens, regadio, sequeiro) não consegue explicar/modelar, de forma igual, a mistura espectral todos estes tipos de unidades da paisagem.

Muito embora também no Minho se verificar, nos meses de inverno, um erro de modelação superior, a análise global de todas as imagens apresentam valores baixos, indiciando uma escolha de componentes puros apropriados para simular a mistura espectral do pixel.

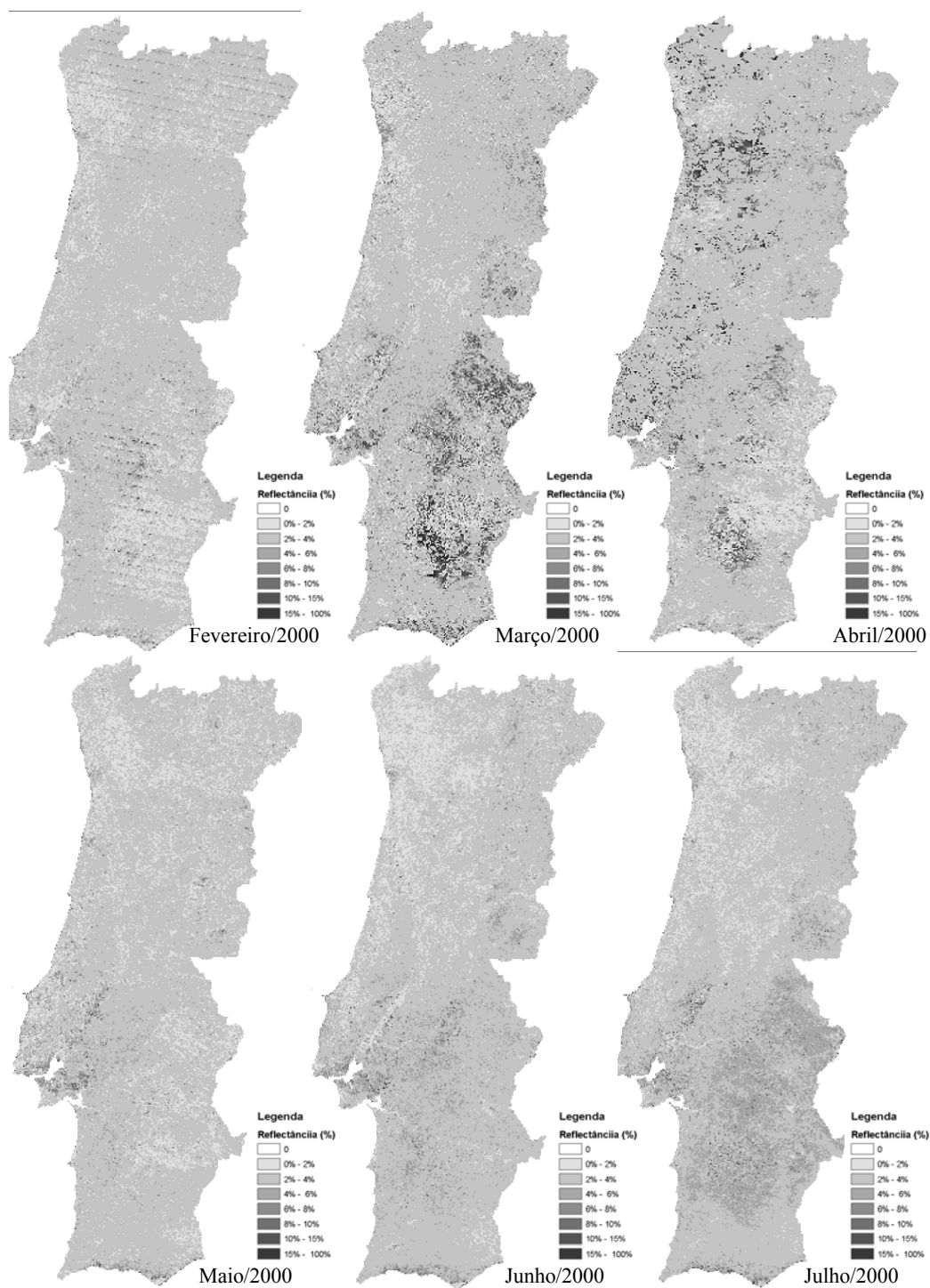


Figura 4.10 – Imagens de erro médio quadrático de Fevereiro a Julho de 2000

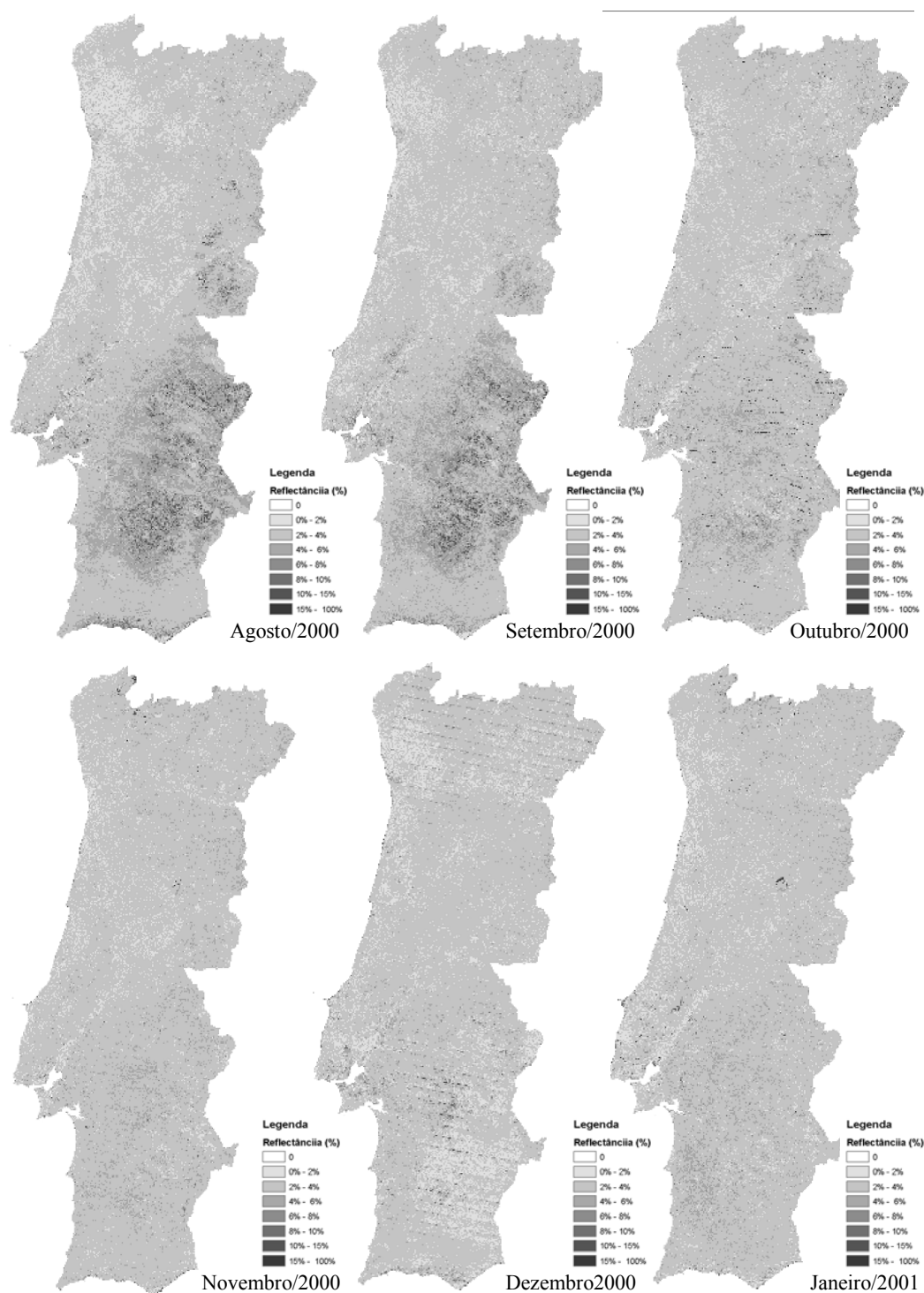


Figura 4.11 – Imagens de erro médio quadrático de Agosto de 2000 a Janeiro de 2001

Outra forma de avaliar o erro de modelação da mistura do *pixel*, foi através da análise do histograma de distribuição de erros médios quadráticos para cada um dos meses (Figura 4.12)

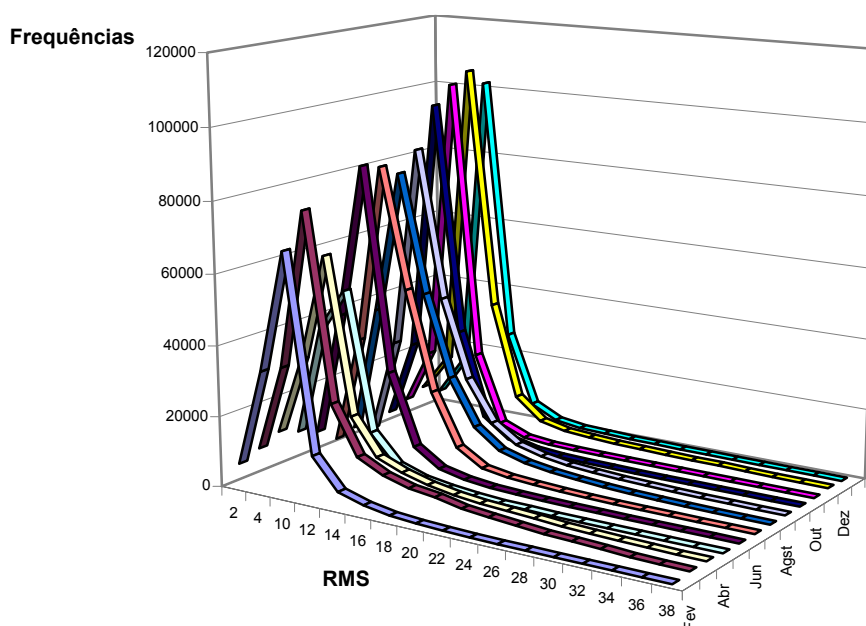


Figura 4.12 - Histograma de análise do termo do RMS do erro para a série multi-temporal

Na maioria dos meses, o valor médio do erro médio quadrático é de reflectância de 3% ou seja, a frequência dos valores de reflectância dos pixels é bastante elevada no intervalo de 1% e 5%.

4.5 – Análise de comportamentos não regulares nas amostras

Após a extracção das diferentes abundâncias dos componentes puros nas 304 áreas amostradas, e ao longo dos 12 meses, procurou-se, dentro de cada classe de ocupação do solo, comportamentos não uniformes (excepto na classe dos fogos em áreas florestais e matos e na classe dos cortes e novas plantações). Recorreu-se para isso, à análise de clustering (SOM) para identificar as amostras que não apresentam o mesmo comportamento que a média dos pontos amostrados para a classe, sendo que essas amostras, não foram utilizados para a definição dos perfis tipos de caracterização das classes de ocupação do solo. A definição dos

perfiz foi feita panas com metade das amostras finais, sendo que estas foram divididas de forma aleatória em treino e validação. O Quadro 4.4 apresenta um resumo desta análise.

Quadro 4.4 – Resumo de análise de clustering

Classe	Amostras Iniciais (nº)	Amostras excluídas(nº)	Amostras finais (nº)	Amostras treino (nº)	Amostras de validação(nº)
Água	15	5	10	5	5
Improdut.	15	6	9	4	5
Urbano	31	5	26	13	13
Matos	30	4	26	13	13
Folhosa	30	2	28	14	14
Resinosas	30	5	25	12	13
Pastagens	30	7	23	11	12
Sequeiro	31	3	28	14	14
Regadio	30	6	24	12	12

4.6 – Perfil Temporal de Classes de Ocupação do solo

Nesta secção define-se os perfis temporais das 11 classes de ocupação do solo estudadas, com base nas oscilações dos valores dos componentes puros. O perfil de cada uma das classes deriva unicamente dos dados do conjunto de treino, gerado a partir da escolha aleatória (Macro MS Excel) das amostras, de acordo com o método definido no Capítulo III.

A constituição dos compósitos mensais das imagens MODIS baseou-se, entre vários critérios, na selecção dos *pixels* com o máximo NDVI para o mês. É, portanto, de esperar que a fracção de vegetação apresente valores mais elevados, pois esta componente estará maximizada no *pixel* misto.

A nomenclatura adoptada agrupa as classes em grandes grupos: massas de água, zonas improdutivos (rocha, inertes e solo nu), zona urbanas e indústria; áreas de vegetação natural (pastagens pobres, matos

e espécies subarborescentes, floresta de resinosa e floresta de folhosas), áreas agrícolas (regadio e sequeiro) e classes com alteração do coberto (fogos florestais, cortes e novas plantações).

A análise que se segue baseia-se no esquema da Figura 3.7 (Cap. III), onde se expressa a comparação entre as grandes classes e uma descrição detalhada de cada uma das 11 classes de ocupação do solo definidas.

A abordagem efectuada apresenta uma descrição qualitativa, contudo deve ser entendida de forma mais quantitativa, pois como veremos os desvios padrões associados a cada um dos componentes puros, ao longo da série temporal das imagens de fracção, são relativamente elevados e irregulares, o que se explica pela variabilidade das amostras seleccionadas. A variabilidade das amostras resulta também do esforço de que as áreas seleccionadas fossem o mais representativo da classe e dessa mesma variabilidade inter-classe.

A - Classe “Massas de água”

O primeiro grande grupo a caracterizar, é designado de “massas de água”, e nele se englobam amostras obtidas em diferentes tipos de planos de água: albufeiras, estuários, oceano e linhas de água. A elevada percentagem da fracção água/sombra na mistura do *pixel* permite distinguir esta classe dos outros grandes grupos.

De acordo com a Figura 4.13 e com os dados do Quadro 4.5, na classe “corpos de água” a fracção água/sombra, durante todo o período de estudo, apresenta valores acima dos 60%. As fracções solo e vegetação apresentam, respectivamente, intervalos de variação entre 9 e 20%, e 5 e 22%. A contribuição espectral destas duas fracções para a mistura do *pixel* desta classe, ao longo dos 12 meses, é relativamente constante. O desvio

padrão, para a água, apresenta valores médios na ordem de 12,8%. Estes valores elevados resultam do número reduzido de amostras identificadas.

Quadro 4.5. – Valores médios (m) e desvio padrão (std) das imagens fracção na classe “massas de água”.

	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Jan
m (vegt)	0,05	0,22	0,20	0,14	0,22	0,07	0,14	0,16	0,14	0,11	0,17	0,13
m (solo)	0,09	0,18	0,20	0,15	0,16	0,09	0,14	0,12	0,14	0,11	0,15	0,13
m (água)	0,87	0,60	0,60	0,72	0,62	0,84	0,72	0,72	0,72	0,78	0,68	0,74

STD(vegt)	0,06	0,17	0,16	0,11	0,15	0,08	0,04	0,11	0,07	0,08	0,1	0,11
STD(solo)	0,07	0,12	0,06	0,08	0,11	0,09	0,1	0,06	0,06	0,07	0,02	0,09
STD(água)	0,11	0,23	0,12	0,15	0,2	0,15	0,1	0,11	0,1	0,08	0,1	0,09

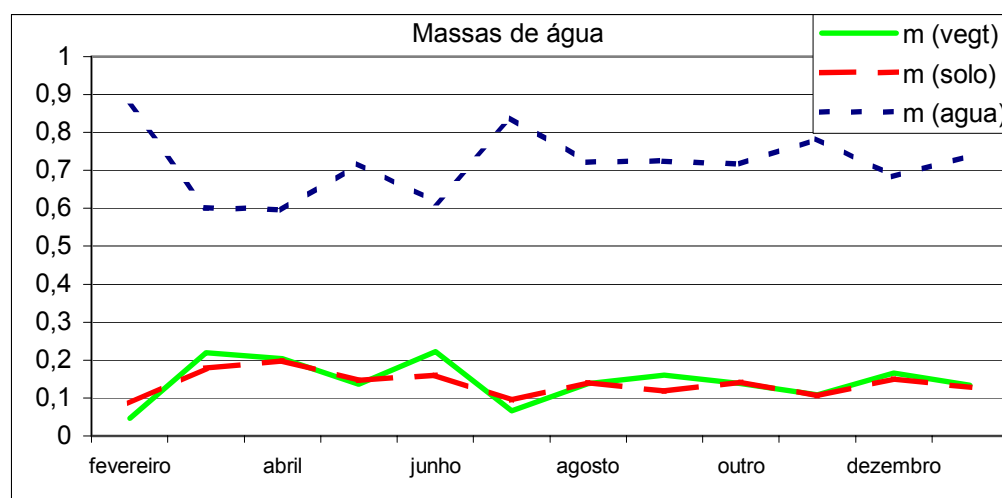


Figura 4.13 – Perfil temporal médio da fracção de vegetação, solo e sombra para as amostras de treino na classe “massas de água”.

A existência de alguma mistura de solo e de vegetação nos *píxels* desta classe, poderá resultar de uma deficiente modelação, pois a selecção das amostras foi feita nos estuários dos grandes rios e em albufeiras.

I - Classe “improdutivo”

Em segundo lugar analisaremos a classe “improdutivo”. Nela se inclui as zonas de inertes, rochas e solo nu. A análise desta classe baseia-se nos dados do Quadro 4.6 e na Figura 4.14.

Quadro 4.6 – Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “improdutivo”.

	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Jan
m (vegt)	0,33	0,34	0,46	0,38	0,39	0,36	0,38	0,37	0,36	0,31	0,36	0,36
m (solo)	0,33	0,38	0,32	0,35	0,33	0,34	0,30	0,30	0,30	0,27	0,21	0,29
m (água)	0,35	0,28	0,22	0,28	0,28	0,31	0,33	0,34	0,35	0,42	0,44	0,35
STD(vegt)	0,04	0,17	0,05	0,02	0,09	0,04	0,06	0,04	0,06	0,05	0,04	0,11
STD (solo)	0,05	0,03	0,11	0,06	0,09	0,07	0,05	0,11	0,05	0,08	0,08	0,13
STD(água)	0,08	0,14	0,09	0,05	0,07	0,04	0,06	0,14	0,02	0,09	0,1	0,24

As três fracções têm uma contribuição constante para esta classe, mas verifica-se o domínio da fracção de vegetação na composição do *pixel*, durante quase todo o período de estudo. A fracção vegetação contribui, em média, com cerca de 37% para a informação do *pixel*. Exceptua-se o período de desenvolvimento vegetativo (meses de Primavera) onde existe um pico com valor de 46% e desvio padrão (STD) associado de 0,05. Nos meses de Inverno atinge-se o valor mínimo (31% e STD de 0,05). Esta diminuição é acompanhada pelo aumento da sombra (passa de valores médios de 30% para 44%), que poderá reflectir diferenças do ângulo de incidência solar ou simplesmente a compensação imposta pelo MML para garantir 100% na soma das fracções.

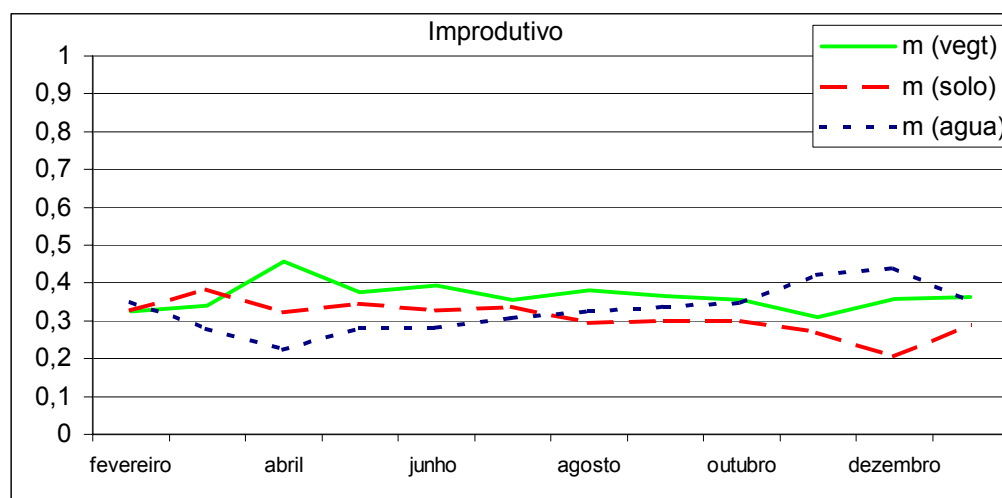


Figura 4.14 – Perfil temporal médio das frações vegetação, solo e sombra nas amostras de treino da classe “improdutivo”.

Tal como as outras fracções, a componente solo apresenta pequenas variações na sua contribuição para a mistura dos *pixels* desta classe de ocupação. O valor mínimo atinge-se nos meses de Inverno, passando de um valor médio de 30% para um valor mínimo de 21% com desvio padrão associado de 0,08.

U – Classe “Zonas Urbanas (contínuas) e Indústria”

O terceiro grande grupo a analisar é a classe “urbano” - das grandes zonas urbanas, onde se incluem as zonas urbanas contínuas e as zonas industriais. Esta classe apresenta um comportamento similar à anterior (improdutivo), sendo difícil a sua distinção unicamente com base na evolução temporal das fracções. Estes dois grandes grupos (classe improdutivo e urbano) apresentam um comportamento bem distinto relativamente às classes com vegetação natural. Os dados para análise desta classe expressam-se no Quadro 4.7 e na Figura 4.15.

Quadro 4.7 – Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “urbano”.

	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Jan
m (vegt)	0,23	0,38	0,37	0,34	0,31	0,26	0,28	0,29	0,31	0,27	0,32	0,30
m (solo)	0,37	0,38	0,31	0,37	0,42	0,47	0,41	0,43	0,32	0,31	0,21	0,28
m (água)	0,40	0,24	0,32	0,29	0,27	0,27	0,30	0,29	0,37	0,42	0,47	0,42
STD (vegt)	0,16	0,19	0,16	0,13	0,09	0,12	0,14	0,15	0,18	0,09	0,18	0,12
STD (solo)	0,09	0,1	0,1	0,1	0,12	0,11	0,12	0,15	0,12	0,09	0,08	0,09
STD (água)	0,12	0,1	0,15	0,11	0,1	0,1	0,11	0,14	0,1	0,08	0,19	0,11

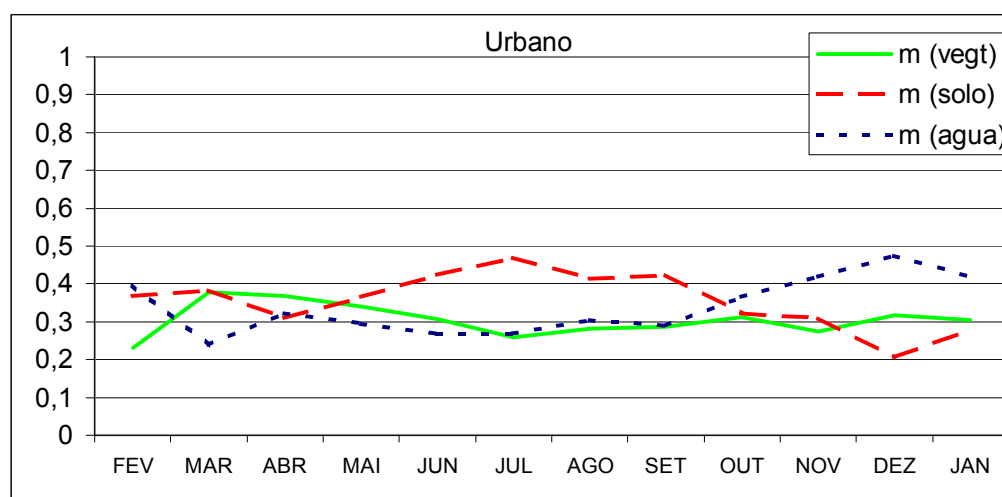


Figura 4.15– Perfil temporal médio das amostras da classe “urbano”.

A classe “urbano” caracteriza-se, assim, pelo domínio da fracção de solo na composição do *pixel*. Durante os meses de Primavera/Verão existe um aumento deste componente, que varia de um mínimo de 31%, no mês de Abril, a um máximo de 47% no mês de Julho (o desvio padrão varia entre os 10% e 11%, respectivamente).

Do mês de Julho a Dezembro, esta classe, caracteriza-se por uma diminuição da componente solo, acompanhada pelo aumento de ensombramento (fracção sombra de 27% para 47%, com desvio padrão de período homónimo de 0,1 para 0,19), o que poderá ser explicado por

alteração do ângulo de incidência solar e pelo escurecimento da imagem devido à sombra das estruturas das cidades e zonas industriais.

A vegetação nas zonas urbanas apresenta um comportamento relativamente constante, com um aumento significativo na fase de desenvolvimento vegetativo das espécies vegetais, o que se pode dever à existência de espaços verdes nos *pixels* seleccionados.

B – Classe “Pastagens naturais”

Passamos a analisar as classes de vegetação natural e de carácter agrícola/florestal. Estas classes distinguem-se das anteriores, pois durante uma parte do período estudado a vegetação é a fracção dominante.

Iniciaremos a caracterização pelo grande grupo da vegetação natural, e dentro deste, pelas pastagens naturais/pobres e pelos matos e espécies subarbustivas. Estas duas classes distinguem-se pelas oscilações temporais dos valores dos três componentes puros, sendo que na classe de pastagens pobres, a amplitude das variações é superior.

No grupo das classes de ocupação do solo com vegetação natural, as pastagens e os matos, caracterizam-se pelo facto das contribuições das fracção solo e vegetação oscilarem simetricamente. As duas outras classes (floresta de folhosas e floresta de resinosa) apresentam um comportamento relativamente constante, com pequenas variações, mas onde as fracções solo e vegetação não tem alternância entre si.

Analisando os dados da Figura 4.16 e do Quadro 4.7, conclui-se que esta classe apresenta um domínio da fracção solo durante os meses de Verão/Outono. A contribuição do componente solo no pixel varia entre os 40% (desvio padrão de 0,1) em Junho e os 49% em Agosto (desvio padrão de 0,17), voltando a decrescer até Novembro, onde atinge o valor 33% (desvio padrão de 0,08).

Quadro 4.8 – Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “pastagens pobres”.

	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Jan
m (vegt)	0,32	0,39	0,41	0,47	0,35	0,27	0,23	0,21	0,21	0,30	0,41	0,37
m (solo)	0,34	0,43	0,40	0,31	0,40	0,44	0,49	0,47	0,42	0,33	0,35	0,26
m (água)	0,34	0,18	0,19	0,21	0,25	0,29	0,28	0,31	0,37	0,37	0,24	0,37
STD(vegt)	0,13	0,11	0,09	0,11	0,12	0,16	0,17	0,18	0,17	0,14	0,08	0,13
STD (solo)	0,07	0,08	0,09	0,09	0,1	0,12	0,17	0,13	0,11	0,08	0,07	0,06
STD(água)	0,12	0,1	0,08	0,11	0,05	0,08	0,16	0,12	0,09	0,13	0,07	0,17

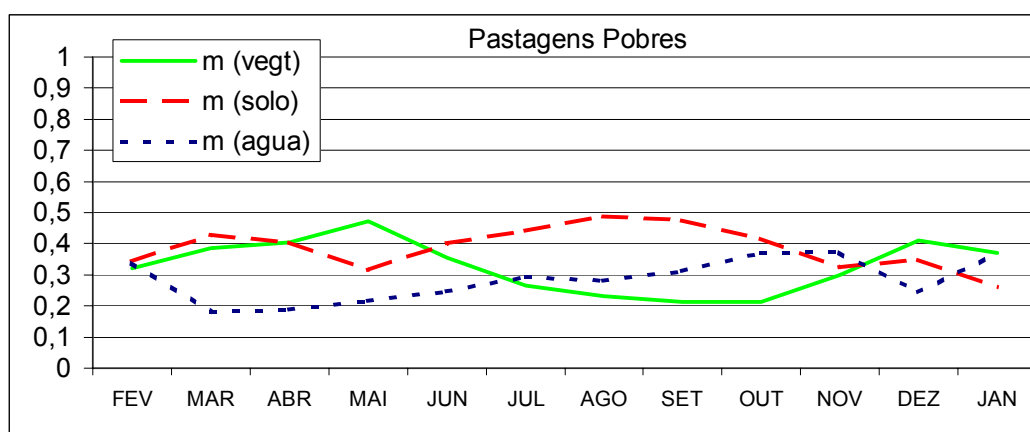


Figura 4.16 – Perfil temporal médio das amostras da classe “pastagens pobres”.

Nesta classe a fracção vegetação tem um comportamento simétrico ao solo, exceptuando-se os meses de Primavera, onde o desenvolvimento natural da vegetação faz aumentar a percentagem deste componente puro com o máximo em Maio (47% e STD de 0,11). É neste mês que se atinge o valor mínimo da fracção solo (31% STD de 0,09). A sombra (fracção água) tem um comportamento tendencialmente crescente ao longo do ano (Fevereiro a Dezembro), apresentando um acréscimo de cerca de 19% neste período, exceptuando-se algumas quebras.

M – Classe “matos e espécies subarbustivas”

Como se pode verificar da Figura 4.17 e do Quadro 4.9, a amplitude da oscilação da abundância dos componentes puros, nesta classe, é inferior à classe anterior.

Quadro 4.9 – Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “matos e espécies subarbustivas”

	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Jan
m (vegt)	0,33	0,37	0,39	0,40	0,39	0,33	0,29	0,27	0,30	0,29	0,42	0,33
m (solo)	0,31	0,35	0,32	0,30	0,34	0,37	0,42	0,38	0,33	0,26	0,28	0,24
m (água)	0,36	0,28	0,29	0,30	0,27	0,30	0,29	0,35	0,36	0,45	0,30	0,43
STD(vegt)	0,08	0,09	0,05	0,11	0,15	0,18	0,14	0,1	0,14	0,06	0,1	0,07
STD (solo)	0,1	0,12	0,13	0,07	0,11	0,13	0,16	0,14	0,13	0,1	0,11	0,08
STD(água)	0,1	0,13	0,12	0,14	0,11	0,15	0,11	0,11	0,11	0,08	0,17	0,1

Observa-se uma variação cíclica e simétrica entre o solo e a vegetação, mas com uma amplitude muito reduzida. Com valores que variam, durante os meses de Primavera/Verão, entre os 33% e 40%, (desvio padrão de 0,08 e 0,11 respectivamente). A classe de vegetação é a que mais contribui para a informação espectral desta classe. O solo mantém um comportamento aproximadamente constante ao longo do ano (34%), exceptuando-se os meses de Julho a Setembro, onde apresenta o valor máximo de 40% (STD de 0,16)

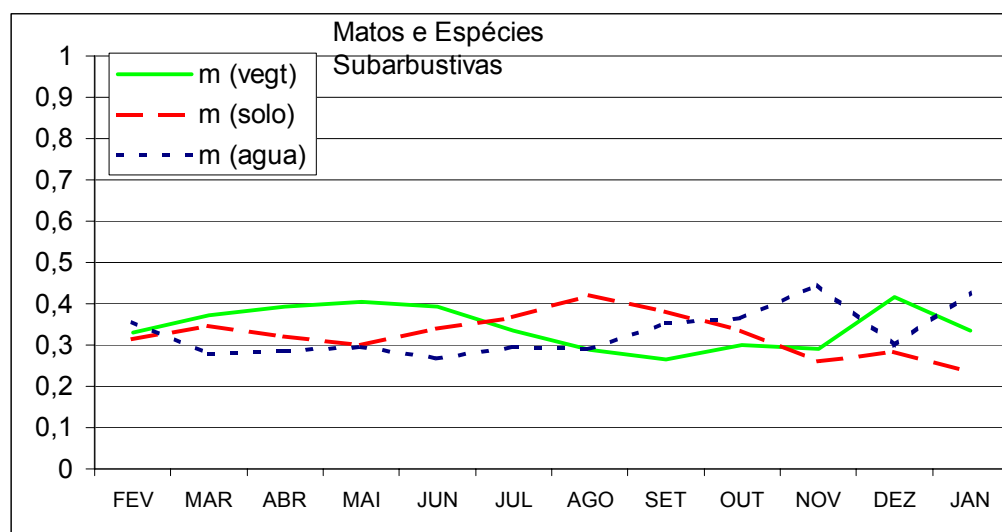


Figura 4.17 – Perfil das amostras da classe matos e espécies subarbustivas

À semelhança da classe de “improdutivos” a fracção sombra/água aumenta nos meses de Inverno. Como nesta classe não existem espécies com porte elevado, o aumento dessa fracção poderá resultar dos diferentes ângulos de iluminação das imagens.

P- Classe “florestas resinosas”

No grupo de classes da vegetação natural, as classes que correspondem à floresta (resinosa e folhosa) distinguem-se, das duas anteriores (pastagens e matos), pelo elevado valor da fracção de vegetação, cujo padrão é constante ao longo do tempo. Esta fracção domina a mistura do *pixel*, mas a distinção entre as duas classes de floresta está na sombra.

Os dados do Quadro 4.10 e a Figura 4.18 estão na base da caracterização da classe de ocupação do solo “floresta de resinosa”.

Quadro 4.10– Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “florestas resinosas”

	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Jan
m (vegt)	0,42	0,51	0,53	0,44	0,45	0,48	0,45	0,46	0,51	0,46	0,49	0,42
m (solo)	0,12	0,14	0,13	0,13	0,15	0,15	0,12	0,10	0,06	0,05	0,06	0,04
m (água)	0,46	0,36	0,34	0,43	0,40	0,38	0,43	0,44	0,44	0,49	0,46	0,55
STD(vegt)	0,07	0,05	0,12	0,07	0,08	0,1	0,07	0,08	0,12	0,1	0,14	0,11
STD (solo)	0,05	0,07	0,06	0,05	0,06	0,04	0,04	0,04	0,03	0,05	0,07	0,03
STD(água)	0,06	0,08	0,13	0,08	0,08	0,1	0,07	0,08	0,12	0,1	0,17	0,1

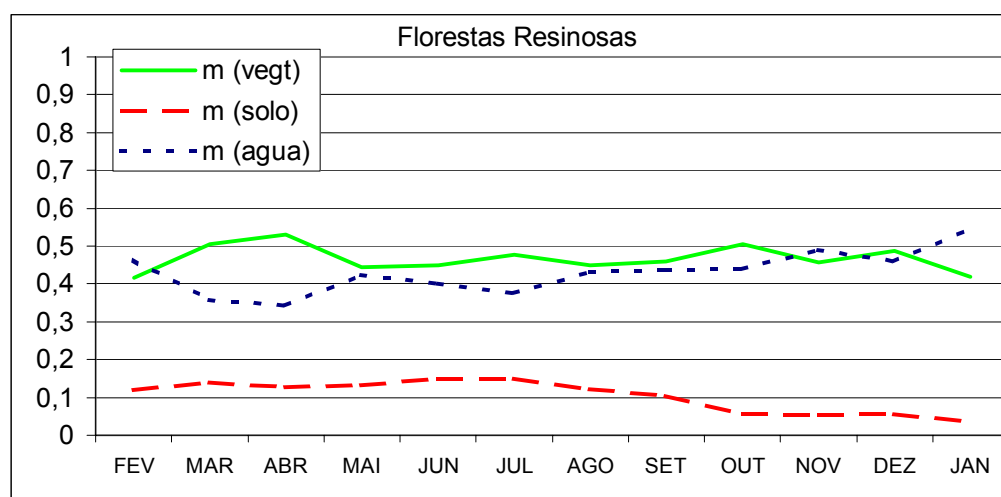


Figura 4.18 – Perfil temporal médio da classe “florestas resinosas”.

Os *pixels* característicos da floresta de resinosas apresentam baixos valores da fracção solo e elevados valores das fracções de sombra e de vegetação (valor médio de 42 e 46% respectivamente). Durante todo o período de estudo as três fracções apresentam um comportamento constante, exceptuando-se na Primavera e no Inverno. No primeiro caso, devido ao desenvolvimento vegetativo, a fracção de vegetação torna-se mais dominante, com um valor de 53% (STD de 0,12). No segundo período a fracção sombra aumenta (valor médio de 40% passa para o máximo de 48%) e é acompanhada pela diminuição da fracção solo, o que se pode

explicar quer pela alteração do ângulo de iluminação (altura solar) quer pela perda de vigor da vegetação e consequente compensação da fracção sombra para respeitar as condições do CLS do MML.

A fracção solo tem uma baixa contribuição para a caracterização da classe floresta de resinosa (oscilam entre 4 e os 15%)

F – Classe “floresta de folhosas”

A análise do perfil temporal desta classe é suportada pelos dados do Quadro 4.11 e pela Figura 4.19. A classe “floresta de folhosas” caracteriza-se por um comportamento constante das três fracções sendo a fracção de vegetação a que mais contribui para caracterizar esta classe. O predomínio desta fracção resulta da densidade elevada dos povoamentos florestais destas espécies, o que se traduz numa baixa reflectância do solo por baixo das copas.

Quadro 4.11 – Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “floresta de folhosas”

	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Jan
m (vegt)	0,50	0,63	0,69	0,56	0,56	0,54	0,53	0,56	0,62	0,53	0,70	0,60
m (solo)	0,13	0,12	0,10	0,15	0,16	0,15	0,15	0,11	0,05	0,04	0,05	0,03
m (água)	0,37	0,26	0,21	0,29	0,29	0,30	0,32	0,34	0,33	0,43	0,25	0,37
STD(vegt)	0,06	0,15	0,1	0,12	0,11	0,08	0,09	0,1	0,15	0,06	0,22	0,12
STD(solo)	0,07	0,06	0,04	0,07	0,08	0,08	0,07	0,07	0,04	0,04	0,04	0,04
STD(água)	0,1	0,16	0,11	0,1	0,11	0,09	0,09	0,1	0,16	0,09	0,21	0,12

O ensombramento provocado pelo porte elevado das árvores, resulta em valores da fracção sombra que variam entres os 20 e os 40%. Esta fracção apresenta um comportamento simétrico à fracção vegetação

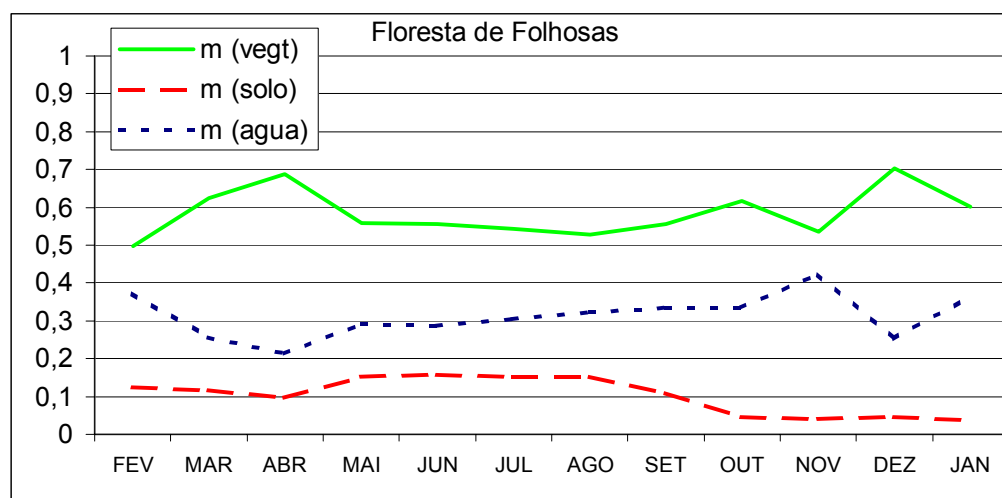


Figura 4.19 - Perfil temporal médio da classe “floresta folhosa”.

Nos meses de Janeiro a Abril ocorre o maior desenvolvimento das espécies florestais, pelo que a fracção de vegetação aumenta. Durante este período, esta fracção apresenta uma distribuição normal, com um valor máximo de 69% e um desvio padrão de 10%. Por oposição, a fracção solo pouco contribui para a caracterização desta classe (abundância entre os 5 e 15 %)

R – Classe “Agricultura de regadio”

O grupo das áreas agrícolas distingue-se de todos os anteriores por apresentar grandes amplitudes de variação das fracções (chegando a valores superiores a 50%) durante um período do ano. Dentro deste grupo encontramos as classes de regadio e de sequeiro. O comportamento padrão destas classes é similar, mas o critério de distinção, entre as duas, reside no desfasamento temporal do aumento da componente vegetação no *pixel* da imagem, relativamente ao seu período natural de desenvolvimento vegetativo.

A caracterização desta classe de ocupação do solo (agricultura de regadio) baseia-se nos dados do Quadro 4.12 e da Figura 4.20.

O regadio tem um comportamento cíclico e caracteriza-se por apresentar um domínio da fracção vegetal, variando entre os 41% (STD de 0,21) em Outubro e os 83% em Agosto (STD de 0,1), exceptuando os meses de Fevereiro e Novembro.

Quadro 4.12 – Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “regadio”

	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Jan
m (vegt)	0,37	0,47	0,44	0,42	0,56	0,77	0,83	0,67	0,41	0,28	0,43	0,49
m (solo)	0,43	0,47	0,37	0,46	0,40	0,21	0,16	0,29	0,38	0,35	0,32	0,21
m (água)	0,20	0,06	0,19	0,12	0,04	0,03	0,02	0,05	0,21	0,38	0,25	0,30
STD(vegt)	0,21	0,22	0,22	0,12	0,23	0,1	0,1	0,18	0,21	0,18	0,13	0,11
STD(solo)	0,16	0,21	0,16	0,15	0,22	0,07	0,09	0,16	0,17	0,13	0,11	0,07
STD(água)	0,1	0,06	0,13	0,1	0,07	0,05	0,02	0,04	0,15	0,15	0,1	0,08

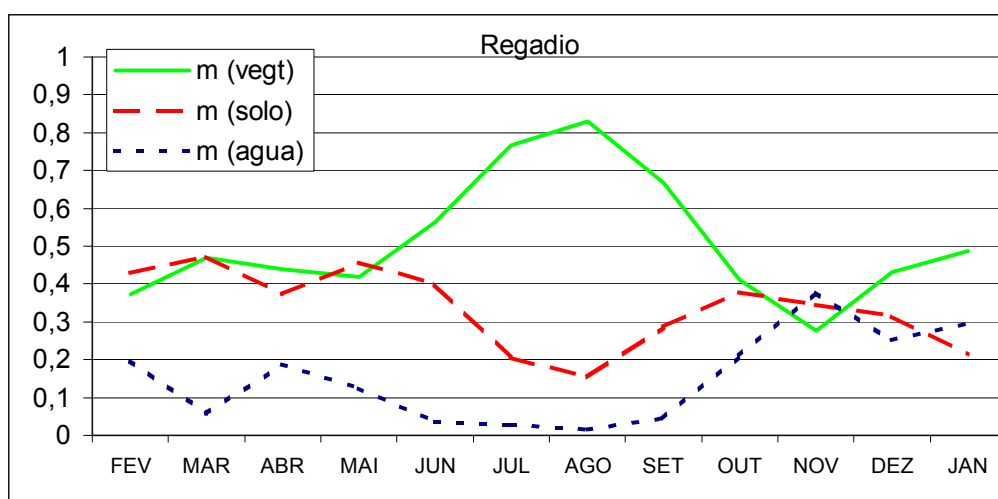


Figura 4.20 – Perfil temporal médio da classe “regadio”.

Tratam-se de culturas anuais, onde a intervenção humana, através da condução da rega, permite desencontrar a fase de maior desenvolvimento vegetativo com o período das chuvas, o que explica

elevado valor do componente vegetação na mistura do *pixel*, durante os meses de Verão. O componente vegetação apresenta uma função de distribuição gaussiana (ou normal) e tem um comportamento simétrico em relação à contribuição das fracções solo e sombra. Este facto resulta de um coberto vegetal denso das culturas extensivas e de intensidade cultural, sendo na maioria dos casos de cobertura total do solo. A contribuição espectral do solo, expressa na respectiva imagem fracção, apresenta um comportamento constante nas fases onde não existe desenvolvimento de culturas de regadio.

O reduzido valor da fracção sombra/água, resulta do facto das culturas de regadio apresentarem um porte reduzido e, também, porque não são influenciadas pelo efeito topográfico. Apenas se verifica um aumento desta componente nos meses correspondentes ao pousio do solo, provavelmente, devido à alteração do ângulo solar

S – Classe “Agricultura de sequeiro”

A classe de ocupação do solo “agricultura de sequeiro” tem um comportamento semelhante à do regadio, mas o desenvolvimento da vegetação coincide com a altura de maior precipitação nas latitudes portuguesas. No Quadro 4.13 e na Figura 4.21 apresentam-se os dados que caracterizam esta classe.

Esta classe exhibe duas fases distintas, onde a vegetação e o solo apresenta um comportamento simétrico. Na primeira fase (Dezembro a Abril), a contribuição da vegetação para a mistura do *pixel* é dominante (79%). Na segunda fase existe uma inversão deste comportamento, sendo a contribuição da fracção solo dominante (68%), e da vegetação reduzida .

Quadro 4.13 – Valores médios (m) e desvio padrão (std) das imagens fracção nas amostras de treino da classe “sequeiro”

	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Jan
m (vegt)	0,79	0,75	0,76	0,60	0,41	0,31	0,30	0,30	0,17	0,19	0,50	0,59
m (solo)	0,17	0,20	0,19	0,26	0,47	0,63	0,67	0,68	0,60	0,42	0,33	0,22
m (água)	0,04	0,05	0,05	0,14	0,12	0,06	0,04	0,02	0,24	0,39	0,17	0,19
STD (vegt)	0,16	0,13	0,11	0,13	0,12	0,10	0,15	0,12	0,12	0,12	0,17	0,19
STD (solo)	0,10	0,07	0,10	0,10	0,13	0,11	0,13	0,13	0,13	0,10	0,09	0,09
STD (água)	0,08	0,08	0,06	0,08	0,08	0,05	0,05	0,03	0,10	0,08	0,13	0,12

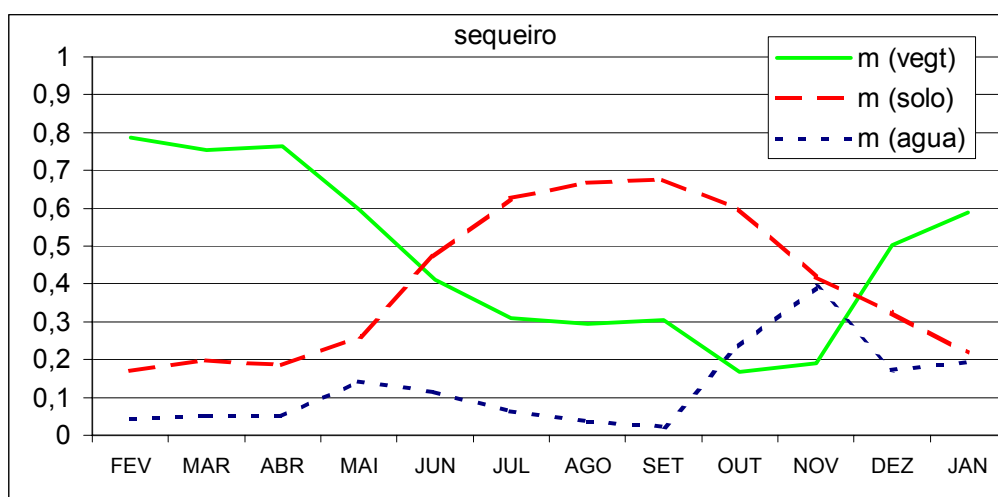


Figura 4.21 – Perfil temporal médio da classe “sequeiro”.

Durante os meses de Junho a Novembro, a abundância da componente solo, na mistura do *pixel*, varia entre os 47% (STD de 0,11) em Junho e os 68% em Setembro (STD de 0,13).

Tal como o “regadio”, esta classe é muito pouco afectada pela fracção sombra. Este componente é utilizado principalmente para simular os efeitos de diferentes ângulos de iluminação. Os acréscimos de valores percentuais de sombra nos meses de Outubro e Novembro, poderão resultar de diferentes ângulos de observação e de iluminação da imagem.

Z – Classe “áreas florestais ardidas”

O último grupo de classes de ocupação em estudo corresponde a alterações do coberto durante o ano. Neste grupo enquadram-se a classe de fogos em áreas de floresta e matos e a classe de cortes e novas plantações.

Estas classes caracterizam-se pela existência de uma quebra acentuada na fracção de vegetação, em dois meses sucessivos e durante um período do ano. Esta evidência permite distinguir este grupo de classes de todas as outras classes de ocupação estudadas. Nas áreas de sequeiro e de regadio também existe uma quebra da fracção vegetação, mas esta tem um comportamento gradual, devido à senescência da vegetação e ao amadurecimento dos frutos (no caso dos cereais).

A distinção entre estas duas classes (Cortes e novas plantações e fogos florestais) baseia-se em três critérios. O primeiro reside no facto de após um fogo, a componente sombra aumentar na razão inversa à quebra verificada na fracção vegetação. O segundo resulta do facto de, nos cortes e novas plantações, a fracção solo aumentar logo no mês seguinte ao evento, e a componente vegetação aumentar progressivamente nos meses seguintes. O terceiro critério apenas é válido para países com uma condição geográfica e climática semelhante à de Portugal, e resulta do facto dos cortes ocorrerem, na maioria, na Primavera, e os fogos florestais terem lugar no Verão.

Tratando-se de características que não ocorrem simultaneamente no espaço e no tempo, não é possível determinar um comportamento médio. Nesta sequência, para caracterizar estas classes analisou-se individualmente cada uma das amostras e elegeram-se aquelas que melhor descreviam o fenómeno.

Das amostras de treino, foram seleccionadas a 9ª, a 27ª e 29ª para traduzir o comportamento padrão desta classe e a alteração da cobertura do solo (floresta a solo nu), como se ilustra na Figura 4.22.

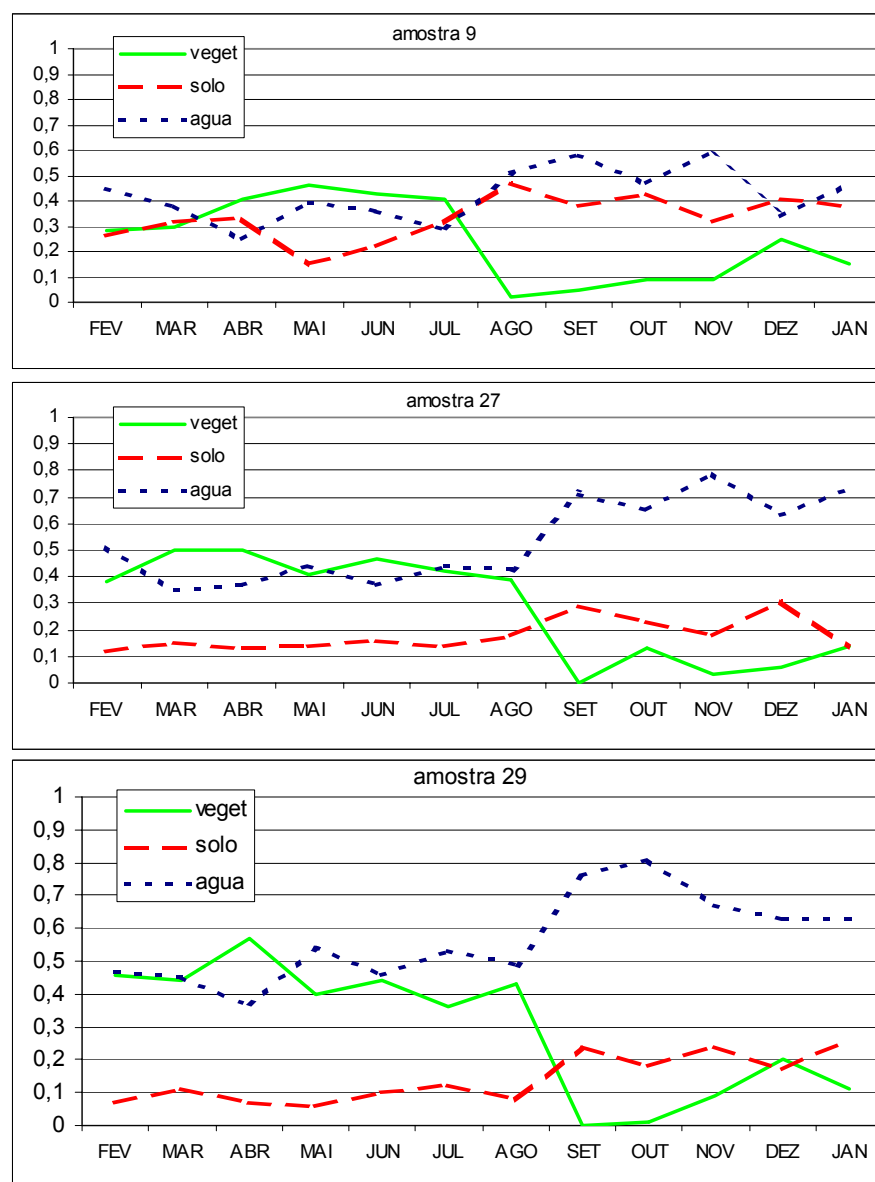


Figura 4.22 – Perfis temporais das amostras 9, 27 e 29 na classe “áreas ardidas”.

As amostras 27 e 29 desta classe ilustram um comportamento típico de floresta de resinosa até ao período onde se verifica o fogo, ou seja, a quebra acentuada na contribuição da fracção de vegetação do *pixel*. Durante este período as fracções apresentam um comportamento constante ao longo do tempo, sendo que, a fracção solo apresenta um valor três vezes inferior às fracções de vegetação e sombra.

Com a ocorrência do fogo florestal, durante os meses de Verão, dá-se uma quebra em 40% do valor da fracção de vegetação e verifica-se um acréscimo em cerca de 30% da fracção sombra. Esta última resulta do acumular de cinza na superfície do solo

A amostra 9 apresenta um comportamento, até ao fogo, de floresta de folhosas. No mês seguinte à ocorrência do fogo verifica-se uma quebra acentuada da fracção de vegetação, mas o aumento da componente sombra é acompanhado pelo aumento da fracção solo. .

Nos dois tipos de floresta, a diminuição da componente sombra nos meses de Dezembro e Janeiro poderá estar relacionado pelo arrastamento das cinzas pela água das chuvas.

H – Classe “cortes e novas plantações”

Os cortes florestais caracterizam-se por uma quebra muito acentuada da contribuição da fracção vegetação durante os meses de Primavera. Esta quebra resulta de uma diminuição da percentagem da fracção vegetação num valor igual ou superior a 20%, entre dois meses consecutivos, e para as nossas latitudes, durante o período que medeia os meses de Fevereiro e de Maio.

No mês onde se verifica a quebra da vegetação, existe, por oposição, um aumento da fracção solo, que se traduz num acréscimo médio de 15% deste face ao que existia antes do corte. Nos meses seguintes, o valor da fracção solo aumenta progressivamente, chegando a atingir cerca de 30% a mais do que o valor na altura do corte, tal qual como é apresentado nas amostras 11 e 22 da Figura 4.23, o que pode ser consequência de os resíduos dos cortes serem retirados do local.

A fracção solo apresenta um comportamento constante, exceptuando-se o caso da amostra 11, onde se deve ter mobilizado o terreno para instalar um novo povoamento florestal. Este facto pode ser justificado com o

crescimento da fracção vegetação, desde o mês de Maio (0%) até ao final do período de estudo (25%).

Neste tipo de classe de ocupação, a fracção sombra pode ser utilizada pelo modelo para garantir 100% de mistura de componentes puros no *pixel*.

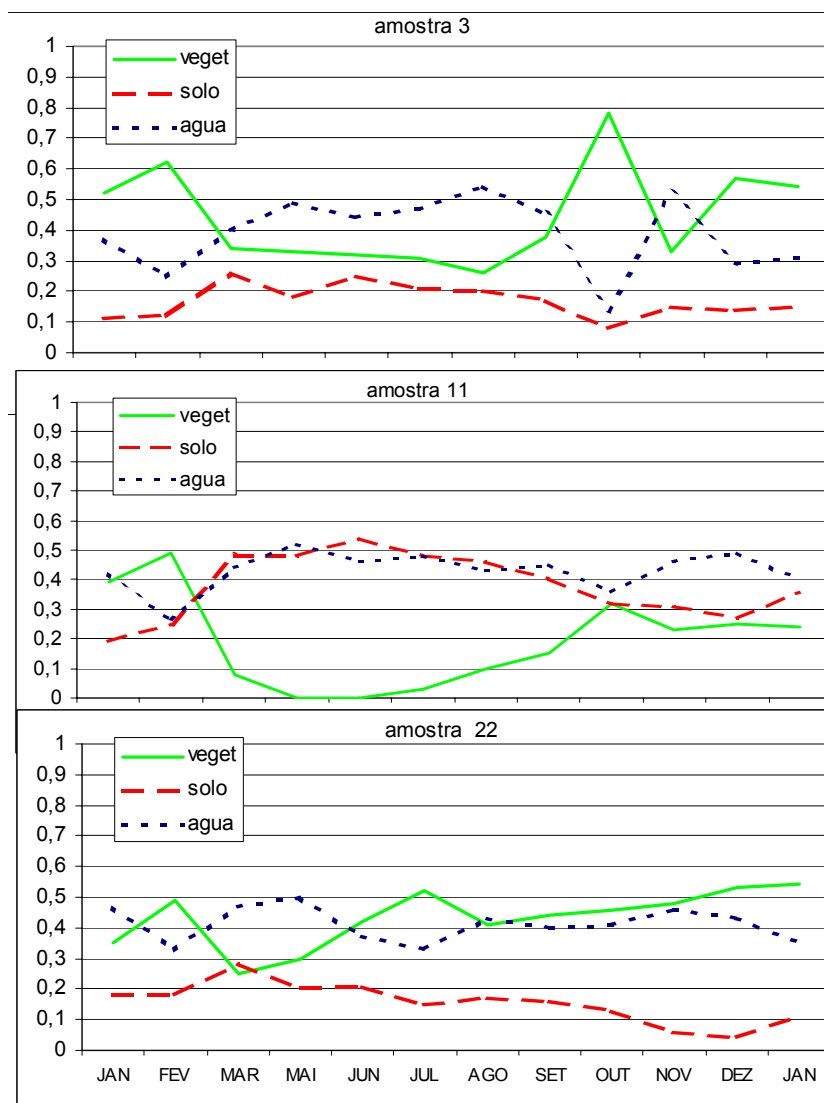


Figura 4.23 – Perfis temporais da classe cortes e novas plantações

Em resumo podemos definir os seguintes critérios para caracterizar as diferentes classes de ocupação do solo:

- A classe de Massas de água apresenta uma fracção de sombra superior a 60 %,
- A classe urbano caracteriza-se pelo domínio da fracção solo durante 9 meses do ano (exclui-se a primavera);
- A classe improdutivo apresenta uma contribuição constante no tempo das três fracções, cada uma com cerca de um terço da composição do pixel;
- O domínio da fracção vegetação caracteriza as classes de ocupação do solo com vegetação natural e agrícola. A distinção entre elas faz-se de acordo com os seguintes critérios:
 - distribuição gaussiana (devido ao facto de se utilizar a média de várias amostras) dos componentes vegetação e solo (simétricos entre si) cuja contribuição é de cerca de um terço e com pequenas amplitudes na variação, caracteriza a classe de matos e espécies subarbustivas;
 - as pastagens pobres também apresentam uma distribuição gaussiana para a variação da componente vegetação e solo, mas as amplitudes são cerca de 20% relativamente aos valores extremos;
 - O regadio apresenta uma distribuição gaussiana da fracção vegetação, com uma amplitude de mais 50% relativamente ao valor mínimo, e onde o valor máximo é atingido nos meses de Verão. A fracção solo e a sombra tem um comportamento relativamente constante, contudo, a última tem um valor muito baixo (aprox. 5%).
 - A classe sequeiro tem um comportamento semelhante ao regadio, mas o valor máximo da fracção vegetação ocorre no final dos meses de Inverno, e no verão a fracção solo tem um acréscimo em 50% face ao valor mínimo.

- A floresta de resinosas caracteriza-se por um comportamento constante da fracção vegetação, cuja contribuição ronda os 45%. A fracção sombra contribui na mesma proporção, enquanto o solo ronda apenas os 10% .
- Na floresta de folhosas também se verifica uma contribuição constante da fracção vegetação (cerca de 60%), contudo a sombra tem um peso inferior (cerca de 30%). A fracção solo também é constante, mas com baixa percentagem na mistura espectral do pixel (aprox. 10%);
- Uma quebra acentuada da fracção vegetação, em dois meses sucessivos, caracteriza as classes de fogos em áreas florestais e matos e em cortes e novas plantações, as quais se distinguem por:
 - No mês seguinte ao fogo, a fracção de vegetação decresce em mais de 40% acompanhado com o aumento da fracção sombra no mesmo valor. A fracção solo tem um comportamento constante e os seus valores são baixos.
 - No corte e novas plantações a vegetação decresce, mas é acompanhada pelo aumento da fracção solo. A sombra é pouco significativa na mistura espectral do pixel e apresenta um comportamento constante ao longo do tempo

4.7 - Análise clusters (SOM)

No Quadro 4.14 apresenta-se a distribuição das todas as amostras de cada uma das 9 classes de ocupação do solo estudadas nesta abordagem (todas excluindo cortes e novas plantações e fogos florestais) pelos *clusters* obtidos pelo SOM. A cada aglomerado foi atribuída uma classe provável, sendo posteriormente determinado os índices de qualidade da classificação (Quadro 4.15). Nesta tabela, e nas restantes que contém o mesmo tipo de informação, a informação das colunas corresponde à

realidade e a informação das linhas corresponde ao resultado da classificação. Neste estudo considerou-se ainda que, para efeitos da classificação dos agregados/clusters do SOM, o CLC2000 representa a realidade.

Quadro 4.14 – Análise de clusters, das amostras, com produção de 28 aglomerados

cluster	Água	Urbano	Improd.	Matos	Pastagens	Folhosas	Resinosas	Sequeiro	Regadio	TOTAL	classe provável
1	2									2	água
2	1									1	água
3	1									1	água
4	1									1	água
5	2									2	água
6	2									2	água
7	3									3	água
8	1						3			4	resinosas
9				1						1	matos
10		1			1					2	pastagens
11		9	2	1						12	urbano
12		9		4	3		1	1		18	urbano
13		1	2	2	7					12	pastagens
14		3	3	4	5			2		17	pastagens
15		1		2	3			6		12	sequeiro
16			2		1			12		15	sequeiro
17		1						3		4	sequeiro
18		1						2		3	sequeiro
19								1		1	sequeiro
20								1	27	28	regadio
21								1		1	sequeiro
22				1				1	1	3	matos
23	1							1		2	água
24		2		1	1				1	5	urbano
25			2	1	5					8	pastagens
26		1		1	1				1	4	matos
27		1		1						2	matos
28	1	1	4	11	3	30	26			76	folhosas
TOTAL	15	31	15	30	30	30	30	31	30	242	

O índice de exactidão global para esta classificação é relativamente baixo (57%), verificando-se que as classes áreas Improdutivas, matos e resinosas apresentam um erro de omissão muito elevado. As amostras da classe “urbano” são classificadas, principalmente como pastagens e sequeiro, concluindo que este erro advém do facto de as imagens compósito serem formadas com base no valor máximo de NDVI, ou seja o vigor da vegetação nas áreas urbanas é maximizado e também pela capacidade que o SOM tem para identificar os diferentes padrões temporais. São ainda classificados nesta classe um número elevado de amostras de pastagens e de matos. Na classe improdutivos as amostras foram essencialmente classificadas como matos e pastagens. A classe dos matos apresenta uma elevada percentagem de amostras classificadas como florestas e como pastagens. As resinosas estão classificadas como

folhosas, concluindo-se que, na primeira iteração, o algoritmo SOM não consegue identificar a variabilidade das classes

Com base nestas “confusões” de classificação entre as classes, facilmente se explica a baixa exactidão do utilizador e do produtor para cada classe (Quadro 4.15).

Quadro 4.15 – Matriz de confusão e índices de qualidade da análise SOM com 28 clusters

	Água	Urbano	Improdutivo	Matos	Pastagens	Folhosas	Resinosas	Sequeiro	Regadio	Total	Exactidão utilizador
Água	14							1		15	93,3
Urbano		20	2	6	4		1	1	1	35	57,1
Improd.			0							0	0,0
Matos		1		3	2			1	2	9	33,3
Pastagens		5	7	8	17			2		39	43,6
Folhosas	1	1	4	11	3	30	26			76	39,5
Resinosas							3			3	100,0
Sequeiro		4	2	2	4			25		37	67,6
Regadio								1	27	28	96,4
Total	15	31	15	30	30	30	30	31	30	242	
exactidão do produtor	93,3	64,5	0,0	10,0	56,7	100,0	10,0	80,6	90,0		57%

O facto das amostras da floresta de resinosas estarem agrupadas no cluster da classe floresta de folhosas, as amostras deste *cluster* (28) foram novamente submetidas à rede neuronal, e o resultado foi anexado à anterior classificação. O resultado desta abordagem é expresso no Quadro 4.16 (seis novos clusters identificados com a cor salmão) e os índices de qualidade no Quadro 4.17. Os resultados mostram que esta abordagem permitiu melhorar a exactidão do produtor para a classe “resinosas” não comprometendo a classe “folhosas”, ou seja, o número de amostras correctamente classificadas, na sua verdadeira classe, contribuiu para melhorar a exactidão global do modelo (65%).

A exactidão do utilizador, na generalidade das classes diminuiu, no entanto, o número de amostras que efectivamente pertencem à classe “resinosas” aumentou, conduzindo a um acréscimo do erro de comissão.

Quadro 4.16 – Análise de clusters, das amostras, com produção de 33 aglomerados

cluster	Água	Urbano	Improdutivo	Matos	Pastagens	Folhosas	Resinosas	Sequeiro	Regadio	TOTAL	classe provável
1	2									2	água
2	1									1	água
3	1									1	água
4	1									1	água
5	2									2	água
6	2									2	água
7	3									3	água
8	1						3			4	resinosas
9				1						1	matos
10		1			1					2	pastagens
11		9	2	1						12	urbano
12		9		4	3		1	1		18	urbano
13		1	2	2	7					12	pastagens
14		3	3	4	5			2		17	pastagens
15		1		2	3			6		12	sequeiro
16			2		1			12		15	sequeiro
17		1						3		4	sequeiro
18		1						2		3	sequeiro
19								1		1	sequeiro
20								1	27	28	regadio
21								1		1	sequeiro
22				1				1	1	3	matos
23	1							1		2	água
24		2		1	1				1	5	urbano
25			2	1	5					8	pastagens
26		1		1	1				1	4	matos
27		1		1						2	matos
28	1									1	água
29			3	3	1	2				9	improdutivo
30		1		4	1		3			9	matos
31				1		1	12			14	resinosas
32						15	10			25	folhosa
33			1	3	1	12	1			18	folhosas

Podemos ainda observar no Quadro 4.17, que as mostras das classes matos e pastagens encontram-se distribuídas por outras classes, pelo que evidencia a dificuldade de serem correctamente classificadas, traduzindo-se nos baixos índices de qualidade (exactidão do produtor e do utilizador)

Quadro 4.17 – Matriz de confusão e índices de qualidade da análise SOM com 33 clusters

	Água	Urbano	Improdutivo	Matos	Pastagens	Folhosas	Resinosas	Sequeiro	Regadio	Total	Exactidão utilizador
Água	14							1		15	93,3
Urbano		20	2	6	4		1	1	1	35	57,1
Improd.		1	3	3	1	2				10	30,0
Matos		2		8	2		3	1	2	18	44,4
Pastagens		5	7	7	18			2		39	46,2
Folhosas			1	3	1	27	11			43	62,8
Resinosas	1			1		1	15			18	83,3
Sequeiro			2	2	4			25		33	75,8
Regadio		3						1	27	31	87,1
Total	15	31	15	30	30	30	30	31	30	242	
exactidão do produtor	93,3	64,5	20,0	26,7	60,0	90,0	50,0	80,6	90,0		65%

O mesmo tipo de raciocínio (nova apresentação das amostras do *cluster* à rede neuronal) foi aplicado aos clusters 12, 13 e 14, uma vez que

as amostras das classes “improdutivos” e “matos” estão distribuídas por estes clusters (classe urbano e pastagens).

Os resultados (Quadro 4.18 e 4.19) obtidos da terceira apresentação dos dados à rede neuronal, mantendo constante todos os parâmetros definidos nos métodos, originou seis novos *clusters* identificados com a cor roxa. Relativamente aos matos não seria de esperar grande melhoria uma vez que as suas amostras estão distribuídas por todas as classes.

Quadro 4.18 – Análise de clusters, das amostras, com produção de 36 aglomerados

cluster	Agua	Urbano	Improdutivo	Matos	Pastagens	Folhosas	Resinosas	Sequeiro	Regadio	TOTAL	classe provavel
1	2									2	agua
2	1									1	agua
3	1									1	agua
4	1									1	agua
5	2									2	agua
6	2									2	agua
7	3									3	agua
8	1						3			4	resinosas
9				1						1	matos
10		1			1					2	pastagens
11		9	2	1						12	urbano
12		1								1	urbano
13			1							1	improdutivo
14		1	1	1			1			4	matos
15		1		2	3					6	pastagens
16		2		1	1			1		5	urbano
17		8	3	6	11			2		30	pastagens
18		1		2	3			6		12	sequeiro
19			2		1			12		15	sequeiro
20		1						3		4	sequeiro
21		1						2		3	sequeiro
22								1		1	sequeiro
23								1	27	28	regadio
24								1		1	sequeiro
25				1				1	1	3	matos
26	1							1		2	água
27		2		1	1				1	5	urbano
28			2	1	5					8	pastagens
29		1		1	1				1	4	matos
30		1		1						2	matos
31	1									1	agua
32			3	3	1	2				9	improdutivo
33		1		4	1		3			9	matos
34				1		1	12			14	resinosas
35						15	10			25	folhosa
36			1	3	1	12	1			18	folhosas
TOTAL	15	31	15	30	30	30	30	31	30	242	

Em termos de exactidão global do modelo esta iteração não provocou alteração, uma vez que, o aumento de elementos, correctamente classificados na classe improdutivos e matos, foi acompanhado pela diminuição o número de elementos das classes urbano e pastagens (diagonal principal). Os “matos” e os “improdutivos” continuam a ter baixa exactidão do produtor (amostras correctamente classificadas), contudo

conseguiu-se melhorar, ao nível da exactidão do utilizador, a qualidade de algumas das outras classes.

Nas classes improdutivo, matos, urbano e pastagens verifica-se, de um modo genérico, uma dificuldade extrema de este classificador as diferenciar com base na variação temporal das fracções, o que se traduz no facto de a exactidão do utilizador ter sido de novo degradada, devido ao aumento do erro de comissão.

Quadro 4.19 – Matriz confusão e índices de qualidade da análise SOM com 36 clusters

	Água	Urbano	Improdutivo	Matos	Pastagens	Folhosas	Resinosas	Sequeiro	Regadio	Total	Exactidão utilizado
Água	14							1		15	93,3
Urbano		14	2	3	2			1	1	23	60,9
Improd.			4	3		2				9	44,4
Matos		4	1	9	2		4	1	2	23	39,1
Pastagens			5	9	20			2		46	43,5
Folhosas			1	3		27	11			42	64,3
Resinosas	1			1	2	1	15			20	75,0
Sequeiro		3	2	2	4			26		37	70,3
Regadio									27	27	100,0
Total	15	31	15	30	30	30	30	31	30	242	
exactidão do produtor	93,3	45,2	26,7	30,0	66,7	90,0	50,0	83,9	90,0		64%

Por fim procedeu-se à agregação das amostras da classe “matos” e “pastagens” numa única classe (P+M), uma vez que ambas são classes de vegetação natural não florestal, com o intuito de melhorar a qualidade da classificação (Quadro 4.20 e 4.21).

Como foi verificado na leitura e análise dos perfis temporais estas são classes com variação temporal homogénea e pouco acentuada e com proporções semelhante para a contribuição das 3 fracções, pelo que a sua distinção/diferenciação é difícil.

A abordagem de submeter novamente os dados dos *clusters* “confusos” à rede neuronal permitiu melhorar os três índices de qualidade da classificação, embora apenas 71% das amostras definidas, para as 9 classes de ocupação do solo, estejam correctamente classificadas. As classes das folhosas, regadio, e água são aquelas onde a maioria das amostras estão correctamente classificadas (cerca de 90%).

Quadro 4.20 – Análise de clusters, das amostras, com produção de 36 aglomerados

cluster	Água	Urbano	Improdutivo	Past+Matos	Folhosas	Resinosas	Sequeiro	Regadio	TOTAL	classe provável
1	2			0					2	água
2	1			0					1	água
3	1			0					1	água
4	1			0					1	água
5	2			0					2	água
6	2			0					2	água
7	3			0					3	água
8	1			0					4	resinosas
9				1		3			1	Past+Matos
10		1		1					2	Past+Matos
11		9	2	1					12	urbano
12		1		0					1	urbano
13			1	0					1	improdutivo
14		1	1	1		1			4	Past+Matos
15		1		5					6	Past+Matos
16		2		2			1		5	urbano
17		8	3	17			2		30	Past+Matos
18		1		5			6		12	sequeiro
19			2	1			12		15	sequeiro
20		1		0			3		4	sequeiro
21		1		0			2		3	sequeiro
22				0			1		1	sequeiro
23				0			1	27	28	regadio
24				0			1		1	sequeiro
25				1			1	1	3	Past+Matos
26	1			0			1		2	água
27		2		2				1	5	urbano
28			2	6					8	Past+Matos
29		1		2				1	4	Past+Matos
30		1		1					2	Past+Matos
31	1			0					1	água
32			3	4	2				9	Past+Matos
33		1		5		3			9	Past+Matos
34				1	1	12			14	resinosas
35				0	15	10			25	folhosa
36			1	4	12	1			18	folhosas
TOTAL	15	31	15	60	30	30	31	30	242	

Quadro 4.21 – Matriz de confusão e índices de qualidade da análise SOM com 36 clusters

	Água	Urbano	Improdutivo	Past+Matos	Folhosas	Resinosas	Sequeiro	Regadio	Total	Exactidão utilizador
Água	14						1		15	93,3
Urbano		14	2	5			1	1	23	60,9
Improd.			1		2	4			7	14,3
Past+Matos		14	9	47			3	2	75	62,7
Folhosas			1		27	11			39	69,2
Resinosas	1			2	1	15			19	78,9
Sequeiro		3	2	6			26		37	70,3
Regadio								27	27	100,0
Total	15	31	15	60	30	30	31	30	242	
exactidão do produtor	93,3	45,2	6,7	78,3	90,0	50,0	83,9	90,0		71%

Foi possível também aperfeiçoar a exactidão do produtor em todas as classes, simplesmente devido à redução do número amostras de vegetação natural não florestal que estavam incorrectamente classificada nas análises anteriores. Verifica-se que a classe de “improdutivos” é de

difícil definição, quer em termos de classificação, quer ao nível da identificação das amostras representativas da classe.

Realizou-se ainda uma análise cluster utilizando apenas dados de um mês de primavera, de um mês de verão e de um mês de outono (respectivamente Maio, Julho, e Outubro) cujo resultado dos índices de qualidade se apresenta no Quadro 4.22. Como se verifica quer a exactidão global (42%), quer a exactidão do produtor e do utilizador apresentam valores inferiores, pelo que se conclui que não é suficiente a escolha de três data para a caracterização da ocupação do solo.

Quadro 4.22 – Matriz de confusão e índices de qualidade da análise SOM para 3 datas

	Água	Urbano	Improdutivo	Matos	Pastagens	Folhosas	Resinosas	Saqueiro	Regadio	Total	Exactidão utilizador
Água	12	1		5						18	66,7
Urbano		16	7	4	6	1	1			35	45,7
Improd.		4	4		4			1	3	16	0,0
Matos				0						0	0,0
Pastagens		2	1	3	5			1	4	16	31,3
Folhosas	1	2		9	5	12	5			34	35,3
Resinosas	1		1	4		17	24			47	51,1
Saqueiro	1	6	2	5	10			29	23	76	38,2
Regadio									0	0	0,0
Total	15	31	15	30	30	30	30	31	30	242	
Exactidão do produtor	80,0	51,6	26,7	0,0	16,7	40,0	80,0	93,5	0,0		42%

Realizou-se ainda uma análise uni-temporal (uma única data), usando para o efeito os dados das amostras do composito do mês de Agosto. O quadro 4.23 apresenta os resultados e pode-se inferir que a análise de multi-temporal associada às imagens de fracção do modelo de mistura linear permite caracterizar a ocupação do solo com maior exactidão face à análise uni-temporal, o que se traduz por esta ultima apresentar valores dos parâmetros de qualidade bastante inferiores (exactidão global é de 37% e a exactidão do produtor e do utilizador tem valores muito baixos).

Quadro 4.23 – Matriz de confusão e índices de qualidade da análise SOM para 1 data

	Água	Urbano	Improdutivo	Matos	Pastagens	Folhosas	Resinosas	Sequeiro	Regadio	Total	Exatidão utilizada
Água	12									12	100,0
Urbano		3		1	1					5	60,0
Improd.			0							0	0,0
Matos		1		2	1		1	1	2	8	12,5
Pastagens		1	1	1	1			1		5	20,0
Folhosas		4	4	10	9	12	3		3	45	26,7
Resinosas			1	2	1	15	25		1	47	53,2
Sequeiro	2							1		3	27,1
Regadio		20	7	11	16		1	26	15	96	27,1
Regadio	1	2	2	3	1	3		3	9	24	12,5
Total	15	31	15	30	30	30	30	31	30	242	
exatidão do produtor	80,0	9,7	0,0	6,7	3,3	40,0	83,3	83,9	30,0		37%

Globalmente podemos concluir que a análise de *clusters* das imagens fracção de uma série temporal de imagens MODIS 500, baseada no algoritmo não assistido do *Self Organized Maps*, identifica, de forma satisfatória, as classes de ocupação do solo definidas neste trabalho.

5 – CONCLUSÕES

Para a área de estudo deste trabalho a metodologia definida atingiu o objectivo proposto para esta Dissertação, ou seja, a caracterização da ocupação do solo, com base numa série multi-temporal de imagens de média resolução espacial, através de um processo bi-etápico. Numa primeira fase aplicou-se o MML para a determinação da abundância dos componentes puros vegetação, solo e sombra na mistura do pixel das imagens MODIS 500; numa segunda fase procedeu-se à exploração da variação temporal das fracções para a identificação de 11 classes de ocupação do solo, com base no algoritmo SOM.

A aplicação do MML para determinar a composição da mistura dos pixels de média resolução espacial do MODIS mostrou ser uma técnica eficaz. A simbiose entre esta técnica e a abordagem de análise multi-temporal permitiu diferenciar classes de ocupação com características espectrais semelhantes, num dado período do ano (e.g., pastagens pobres e matos). Também as classes resultantes de alteração brusca da ocupação do solo (fogos florestais e cortes e novas plantações) são claramente identificadas com base no período de ocorrência do fenómeno e na variação temporal das três fracções solo, vegetação e sombra. A utilização deste último componente puro mostrou-se essencial para distinguir florestas de folhosas e resinosas e minimizar o efeito topográfico e de iluminação.

A principal vantagem desta metodologia é a caracterização da ocupação do solo com base em imagens de média resolução espacial que têm uma resolução temporal diária e cujo custo de aquisição é nulo, flexibilizando o período de estudo.

Uma vantagem, que também ressalta do uso do MML, é o facto de se transformar as imagens originais de MODIS expressas em reflectância em cada uma das bandas numa unidade expressa em percentagem de

ocupação de vegetação, solo e sombra/água, que é facilmente entendível por qualquer utilizador.

Na etapa de pré-processamento, a criação dos compósitos mensais, baseada no algoritmo do Compósito do Valor Máximo maximizou a componente vegetação na mistura do pixel da imagem, sobrestimando a sua abundância em termos de fracção e consequentemente desvirtuando o resultado final em termos de perfil médio de cada classe de ocupação do solo. Embora se tenha procurado identificar todos os pixels contaminados por nuvens, o método utilizado é relativamente simples e apresenta deficiências, pois percorrer todos os DN da banda de QF para as 12 imagens é um processo exaustivo, moroso e sujeito a erros. Possivelmente foram excluídos pixels que não se apresentavam contaminados e que poderiam ser potenciais componentes puros. Também o facto de a criação do compósito mensal utilizar compósitos semanais, cujo critério para criação se baseia num parâmetro que consideramos pouco claro, pode introduzir ruído nas imagens base do estudo.

A selecção das amostras representativas de cada classe foi feita com base nas unidades taxonómicas do CORINE Land Cover 2000. Segundo o relatório técnico do Instituto do Ambiente, as classes 211 (culturas anuais de sequeiro), 212 (culturas anuais de regadio), 231 (pastagens), 321 (pastagens naturais), 322 (matos) e 323 (vegetação esclerofítica) tiveram problemas de classificação, pelo que apresentam índices de qualidade baixos. Como resultado, as classes do CLC200 utilizadas como de base (referência da verdade) para a identificação de 4 das 11 unidades taxonómicas de classificação adoptadas neste estudo apresentam problemas de classificação na cartografia CLC. Acrescido a este facto esteve, também, a dificuldade de identificar pixels que fossem utilizáveis ao longo de toda a série temporal, pelo que, as amostras poderão não ser totalmente representativas, em termos espectrais, das 11 classes definidas, condicionando o resultado final.

O resultado da análise do MML, é função da complexidade de mistura do *pixel*. Consoante a “pureza” dos componentes puros seleccionados melhor será a modelação da mistura linear do *pixel* expressa pelas imagens fracções. Ao aplicar o MML, a abundância das fracções na mistura do *pixel* pode ser sobrestimada ou subestimada devido à reflectância aditiva dos componentes que constituem o *pixel* misto, ou seja, a linearidade do modelo conduz à soma dos erros relativos à incorrecta escolhas dos componentes puros.

Outra limitação do MML provem do uso de um comportamento espectral genérico de cada componente puro. As propriedades da vegetação variam com a espécie vegetal e com as condições do meio ambiente onde se insere. As propriedades ópticas do solo também variam de acordo com os componentes que o constituem (matéria orgânica, minerais, etc.) e com o teor de humidade no solo. No MML apenas se pode utilizar um comportamento espectral por cada componente puro. Esta situação, imposta pelo modelo, força um comportamento linear da radiação com os elementos que constituem a mistura do *pixel*. A impossibilidade deste algoritmo para modelar os fenómenos não lineares entre a radiação e os objectos, conduz a erros de modelação da mistura espectral. Existem também erros relativos ao comportamento óptico do sensor (resolução espectral) e a propriedades espectrais dos objectos da superfície que não foram consideradas neste estudo.

Uma fonte de perda de exatidão dos resultados resulta da resolução espacial da imagem de base. Numa imagem de *pixel* de 500 por 500 m, é natural que existam diferentes áreas com várias classes de ocupação do solo. Nesse caso é possível ter diferentes unidades de ocupação mas que apresentam um contínuo em termos das imagens fracção, ou seja, um *pixel* pode cobrir duas unidades distintas e opostas, sendo por exemplo metade da área ocupada por uma classe de vegetação densa e outra de solo nu; contudo, quando sujeito ao MML o *pixel* pode originar uma única classe de

ocupação, com um comportamento misto. Por outro lado, também se podem subestimar classes de ocupação do solo, quando classes de ocupação não apresentam uma área e/ou um comportamento espectral significativo que seja representativo na informação do *pixel*.

Ainda como resultado da aplicação do MML verificou-se a presença de *outliers* nos perfis tipo traçados para cada uma das classes. Este tipo de dados contribuiu para a alteração do valor médio e desvio das fracções num dado momento e consequentemente comportamentos dificilmente explicáveis nas classes, e demonstram a heterogeneidade espacial e espectral das mesmas.

Outra restrição do MML traduz-se no facto da soma das fracções ser sempre igual à unidade (CLS), o que obriga a que uma alteração do valor de uma fracção devido a ruído na imagem, obrigatoriamente se repercute nas outras duas fracções.

O algoritmo de *clustering* baseado na rede neuronal, conseguiu modelar a variabilidade da informação espectral derivada (fracções) e agregar as amostras correspondentes a cada uma das 8 classes de ocupação do solo estudadas de uma forma satisfatória (71%). Este valor pode, na verdade ser mais alto, se se pensar que algumas das áreas utilizadas como amostras podem estar mal classificadas de origem. Refira-se que a cartografia utilizada na selecção das amostra, i.e. CLC2000, tem um erro aproximadamente de 15%.

Como desenvolvimento futuro propõe-se a utilização de dados diários de imagens MODIS com radiância à superfície terrestre e a realização de um estudo da evolução da ocupação do solo com uma periodicidade semanal. Dever-se-á, também, estudar outros métodos de constituição dos compósitos, em especial, recorrer a métodos que eliminem valores de *outliers* nas bandas espectrais e que permitam ultrapassar o

problema de contaminação dos dados da superfície com resposta espectral das nuvens.

A identificação das características de ocupação do solo de cada uma das amostras deve ser confirmada com trabalhos de campo para garantir uma verdade absoluta sobre a realidade.

O aumento do número de componentes puros (variáveis do MML), com o objectivo de tirar partido de todas as bandas do produto MOD09A1, e a opção pela utilização de livrarias espectrais para a definição das assinaturas espectrais poderá contribuir para melhorar a modelação da mistura espectral dos pixels da imagem.

Ainda relativamente à análise espectral de misturas, relaxando o modelo, permitindo que existam fracções negativas e outras maiores do que a unidade, poderá chegar-se a uma solução mais próxima da realidade. Nestas condições, poder-se-ão detectar, com o próprio resultado do modelo, elementos mais puros que os seleccionados.

Numa abordagem futura dever-se-á, ainda, ensaiar outros classificadores, como árvores de regressão, para a caracterização da ocupação do solo utilizando para o efeito uma série multi-temporal de imagens de média resolução espacial.

REFERÊNCIAS BIBLIOGRÁFICAS

ABOELGHAR, M.; TATEISHI, R.; TSOLMON, R., 2002, Land-cover sub-pixel classification using Linear Mixture Model on LANDSAT ETM⁺ data in Egypt [versão electrónica]. In *Proceedings of 23rd Asian Conference on Remote Sensing* (Japão).

(URL: <http://www.gisdevelopment.net/aars/acrs/2002>, consulta em Março de 2004).

ADAMS, J.B.; SMITH, M.O., 1986, Spectral mixture modelling: a new analysis of rock and soil types at the Viking Lander 1 site. *Journal of Geophysical Research*, **91**, 8098-8112.

ADAMS, J.B.; SMITH, M.O.; GILLESPIE, A.R., 1990, Imaging spectroscopy: data analysis and interpretation based on spectral mixture analysis. In *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, editado por Pieters e Englert, pp. 3-39.

AGRAWAL, S.; JOSKI, P.K.; SHUKLA, Y.; ROY, P.S., 2003, SPOT VEGETATION multi-temporal data for classifying vegetation in south central Asia. *Current Science*, **84**, 1440-1448.

ANDERSON J. R.; HARDY, E. E.; ROACH, J.T.; WITMER R. E., 1976, A land use and land cover classification system for use with remote sensor data [versão electrónica]. *U.S. Geological Survey Circular* **671**, USGPO, USA.
(URL: <http://landcover.usgs.gov/pdf/anderson.pdf>).

ASHTON, E.A.; SCHAUM, A., 1998, Algorithms for the detection of sub-pixel targets in multispectral imagery. *Photogrammetric Engineering & Remote Sensing*, **64**, pp 723-733.

ASNER, G.P.; LOBELL, D.B., 2000, A bio geophysical approach for automated SWIR unmixing of soils and vegetation. *Remote Sensing of Environment*, **74**, 99-112.

ASNER, G.P.; HICKE, J. A.; LOBELL, D.B., 2000, Per-pixel analysis of forest structure: vegetation indices, spectral mixture analysis and canopy reflectance modelling.

(URL:www.nrel.colostate.edu/~jhicke/pubs/Asner_etal_Forest-S_chapter.pdf, consulta em Dezembro de 2004).

BAÇÃO, F., 2002, Apontamentos de apoio às aulas da disciplina de Data Mining do Mestrado de C & SIG. UNIGIS, ISEGI-UNL, Lisboa.

BAKKER, W.H.; PARODI, G.N.; TIMMERMANS, W.J., 1997, NOAA AVHRR: An application of a cloud-detection technique. *ITC Journal*, **1**, 64-73.

BAIO, M., 1996, *A detecção remota como fonte de informação geográfica*. Dissertação de Mestrado, IST-UTL, Lisboa.

BARATA, M.T.A., 2001, Classificação do coberto vegetal em imagem de detecção remota através de uma metodologia baseada na morfologia matemática.

Dissertação de Doutoramento, IST-UTL, Lisboa.

BIRD, A.C., 1991, Principles of Remote Sensing: electromagnetic radiation, reflectance and emissivity. In *Remote Sensing and Geographical Information System for Resource Management in Development Countries*, Vol 1, editado por A.S. Belward e C.R. Valenzuela (The Netherlands: Kluwer Academic Pubs), pp 1-15.

BIRD, A.C., 1991a, Principles of Remote Sensing: Interaction of electromagnetic radiation with the atmosphere and the earth. In *Remote Sensing and Geographical Information System for Resource Management in Development Countries*, Vol 1, editado por A.S. Belward e C.R. Valenzuela (The Netherlands: Kluwer Academic Pubs), pp 17-30.

BISHOP, J; CONGALTON, R. G.; BECKER, M. L., 2004, Monitoring biodiversity of select restoration sites in New Zealand.

(URL:http://www.globe.gov/boulder2004docs/congalton_paper_GLOBE_04_NZpaper.pdf , consulta em Janeiro de 2005).

BOLESA, S. H.; XIAOA, X.; LIUB, J.; ZHANGA, Q.; MUNKHTUYAC, S.; CHENB, S.; OJIMAD, D., 2004, Land cover characterization of Temperate East Asia using multi-temporal VEGETATION sensor data. *Remote Sensing of Environment*, **90**, 477-489.

BURKE, S., 1997, Statistics refresher: understanding the structure of scientific data. *Scientific Data Management*, **1**, 32-38.

CAETANO, M., 1995, *Burned vegetation mapping in mountains areas with satellite remote sensing*. Master thesis, University of California, Sta. Barbara, USA.

CAETANO, M., 2000, Characterization of coniferous forest undestory with remotely sense data. Dissertação de Doutoramento, IST-UTL, Lisboa.

CAETANO, M., 2002, Apontamentos de apoio às aulas da disciplina de detecção remota do Mestrado de C & SIG. UNIGIS, ISEGI-UNL, Lisboa.

CAETANO, M.; SANTOS, T.; GONÇALVES, L., 2002, Cartografia de ocupação do solo com imagens de satélite: o estado da arte. In *Proceedings of the ESIG'2002-VII Encontro de Utilizadores de Informação Geográfica* (Oeiras), [CD-ROM].

CARTALIS, C.; FEIDAS, H.; GLEZAKOU, M.; PROEDROU, M.; CHRYSOULAKIOS, N., 2000, Use of earth observation in support of environmental impact assessments: prospects and trends. *Environmental Science & Policy*, **3**, 287-294.

CNES, 2000, *Vegetation Programme*. Technical Information.
(URL: <http://vegetation.cnes.fr>, consulta em Novembro de 2002).

Chuvieco, E.; Ventura, G.; Pilar, M. M.; Gómez, I., 2005, Assessement of multitemporal compositing techniques of MODIS e AVRRH imagens for burned land mapping. *Remote Sensing of Environment*, vol **94**, pp450 -462.

CCE - COMISSÃO DAS COMUNIDADES EUROPEIAS, 2004, Proposta de Directiva do Parlamento Europeu e do Conselho que estabelece uma infra-estrutura de informação espacial na Europa (INSPIRE), CCE, Bruxelas.

CRACKNELL, P. A. (Ed.), 1997, *The Advanced Very High Resolution Radiometer*, (London: Taylor & Francis) .

CROSS, A. M.; SETTLE, J.J.; DRAKE; N.A.; PAIVINEN, R.T.M., 1991, Sub pixel measurement of tropical forest cover using AVHRR data. *International Journal of Remote Sensing*, **12**, 1119-1129.

EGBERT, S.; MARTÍNEZ-MEYER, E.; ORTEGA-HUERTA, M.; PETERSON, T., 2002, Use of data sets derived from time-series AVHRR imagery as surrogates for

land cover maps in predicting species distribution [versão electrónica]. In *Proceedings IGRASS 2002*.

(URL: www.specifysoftware.org/informatics/bios/bioatownperson/Eetal_igrass_2002.pdf , consulta em Janeiro de 2005).

EMMU – Environmental Mapping and Modelling Unit, 1996, Mapping European Land Cover with NOAA-AVHRR Data. Space Application Institute, Joint Research Centre, European Commission.

ESA, 2004. Envisat Caring for the Earth.

(URL: <http://envisat.esa.int> , consulta em Janeiro de 2004)

FARAKLIOTI, M., 1998, Mixed pixels in remote sensing: linear mixing model – vegetation index.

(URL: <http://www.ee.surrey.ac.uk/personal/M.Faraklioti/project.html>, consulta em Setembro de 2004).

FERREIRA, L.; YOSHIOKA, H; HUETE, A; SANO, E., 2003, Seasonal landscape and spectral vegetation index dynamics in the Brazilian Cerrado: an analysis of large scale biosphere-atmosphere experiment in Amazônia (LBA). *Remote Sensing of Environment*, **87**, 534-550.

FONSECA, A.D.; FERNANDES J. C. (Eds.), 2004, *Detecção Remota*, (Lisboa: LIDEL Edições Técnicas).

GIRI, C.; MOE, T. A.; SHRESTHA, S., 2000, Multi-temporal analysis of the VEGETATION data for land cover assessment monitoring in Indochina [versão electrónica]. In *Proceedings of 1st VEGETATION Annual Conference*, Japão.

(URL: www.spot-vegetation.com , consulta em Maio de 2003).

GREGORIO, A.D.; JANSEN, L.J.M., 2000, Land Cover Classification System: Classification Concepts and User Manual. FAO: SD/SDR (Itália: FAO).

GUERSCHAMAN, J.P.; PARUELO, J.M.; DI BELLA, C.; GIALLORENZI, M.C; PACIN, F., 2003, Land cover classification in Argentine Pampas using multitemporal Landsat TM data. *International Journal of Remote Sensing*, **24**, 3381-3402.

HAMMOND, T.O.; VERBYLA, D.L., 1996, Optimistic bias in classification accuracy assessment. *International Journal of Remote Sensing*, **17**, 1261-1266.

HANSEN, M., DEFRIES, R., TOWNSHEND, J. R. G.; SOHLBERG, R., 2000, Global land cover classification at 1km resolution using a decision tree classifier. *International Journal of Remote Sensing*, **21**, 1331-1365.

Harahsheh, H; Tateishi, R., 1997, Land cover map of west Asia using 1 km AVHRR data [versão electrónica]. In *Proceedings of 18 th Asian Conference on Remote Sensing*, Malaysia.

(URL: www.gisdevelopment.net , consulta em Dezembro de 2004).

HILL, J., 1991, Remote sensing systems: sensors and platforms. In *Remote Sensing and Geographical Information System for Resource Management in Development Countries*, Vol 1, editado por A.S. Belward e C.R. Valenzuela (The Netherlands: Kluwer Academic Pubs), pp 55-70.

HILL, J., 1994, Land degradation and soil erosion hazard mapping in Mediterranean environments with operational earth observation satellites. In *Proceedings of International Symposium "Operationalization of Remote Sensing"*, (The Netherlands: ITC), pp 41- 52.

IA - Instituto do Ambiente, 2005, CLC2000 Portugal, Technical Report.

JARS – JAPAN ASSOCIATION OF REMOTE SENSING, 1996, *Remote Sensing*.
(URL:<http://www.profc.udec.cl/~gabriel/tutoriales/rsnote/contents.htm>,
consulta em Fevereiro de 2005).

KLEI, A.; HALL, D. ; RIGGS, G., 1998, Global snow cover monitoring using MODIS [versão electrónica]. In *Proceedings of 22nd International Symposium On Remote Sensing Environment* (Abidjan, Côte D'Ivoire).
(URL: <http://geog.tamu.edu/klein/publications/proceedings/> , consulta em Fevereiro de 2005).

LOBO, A.; PINEDA, N., 2000, SPATEM: the analysis of annual sequences of Vegetation data at the landscape scale [versão electrónica]. In *Proceedings of 1st VEGETATION Annual Conference* (Japão).
(URL: www.spot-vegetation.com , consulta em Maio de 2003).

LUNETTA, R.S. E ELVIDGE, C.D. (Eds), 1999, Remote Sensing Change Detection: Environmental Monitoring Methods and Applications, (London: Taylor & Francis).

MALHEIRO, R.; PAIVA, R.; MENDES, A.; MENDES, T.; CARDOSO, A., 2004. *Sistemas de classificação Musical com redes neuronais* In *Gestão e Desenvolvimento*”, Univ. Católica Portuguesa, Vol 12, pp 167- 195.
(URL:http://www4.crb.ucp.pt/Biblioteca/GestaoDesenv/GD12/gestaodesenvolvimento12_167.pdf , consulta em Fevereiro 2005).

MARÉE, R.; GEURTS, P.; PIATER, J.; WEHENKEL, L., 2004, , A generic approach for image classification based on decision tree ensembles and local sub-windows.

(URL: <http://www.montefiore.ulg.ac.be/~maree/papers/maree-accv2004.pdf>, consulta em Janeiro 2005).

Matsuoka, M.; Hayasaka, T.; Fukushima, Y; Honda Y., 2004, Analysis of the land cover and its change over yellow river basin using satellite data [versão electrónica]. In *Proceedings of XX th International Society for Photogrammetry and Remote Sensing Congress*, **Vol VII** (Istambul).

(URL: <http://www.isprs.org/istanbul2004/comm7/papers/164.pdf>, consulta em Dezembro de 2004).

MOIGNE, J.; LAPORTE, N; NETANYAHU, N., 2001, Enhancement of tropical land cover mapping with wavelet-based fusion and unsupervised clustering of SAR and Landsat image data. In *Proceedings of International Society for Optical Engineering* (Toulouse).

NASA, 1999, Welcome to LANDSAT Programme.

(URL: <http://geo.arc.nasa.gov/sge/landsat/landsat.html>, consulta em Janeiro de 2005).

NASA, 2002, Welcome to MODIS.

(URL: <http://www.gsfc.nasa.gov> , consulta em Janeiro de 2004).

NUNES, A.; CAETANO, M.; SANTOS, T.; GONÇALVES, L., 2002, Generalização de cartografia de ocupação do solo derivada de classificação de imagens de satélite ao nível do pixel. In *Proceedings of the ESIG'2002-VII Encontro de Utilizadores de Informação Geográfica* (Oeiras), [CD-ROM].

OKI, K.; OGUMA, H.; SUGITA, M., 2002, Subpixel classification of alder trees using multitemporal Landsat thematic mapper imagery. *Photogrammetric Engineering & Remote Sensing*, **68**, 77-82.

OKIN, G.S.; ROBERTS, D.A., Remote sending in arid regions: challenges and opportunities. In *The Manual of Remote Sensing*, Vol 4, editado por Ustin, S. (American Society of Photogrammetry and Remote Sensing), pp 1-30.

PAN, Y.; GONG, P.; HE, C.; SHI, P.; PU, R., 2003, An integrative classification of vegetation in China based on NOAA-AVHRR and vegetation-climate indices of the Holdridge life zone. *International Journal of Remote Sensing*, **24**, 1009-1027.

PERDIGÃO, V.; ANONNI, A., 1997, Technical and Methodological Guide for Updating CORINE Land Cover Data Base (CEC Editions).

QUARMBY, N.A.; TOWNSHEAD, J.R.G.; SETTLE, J.J.; WHITE, K.N.; MILNES, M.; HINDLE, T.L.; SILLEOS, N., 1992, Linear mixture modelling applied to AVHRR data for crop area estimation. *International Journal of Remote Sensing*, **13**, 415-425.

RATANA, P.; HUETE, A.R.; FERREIRA, L., 2005, Analysis of Cerrado physiognomies and conversion in the MODIS saesonal-temporal domain [Versão electrónica]. *Earth Interactions*, **9**, 1-22.

(URL: <http://ams.allenpress.com/amsonline/> , consulta em Fevereiro de 2005).

ROBERTS, D. A., SMITH, M.O.; ADAMS, J.B.; SABOL, D.E.; GILLESPIE, A.R., WILLIS, S.C., 1990, Isolating woody plant material and senescent vegetation from vegetation in AVIRIS data. In *Proceedings of the Airborne Science Workshop: AVIRIS*, JPL Pub. 90-54, pp 243-270.

ROBERTS, D. A.; BATISTA, G.T.; PEREIRA, J.L.G.; WALER, E.K.; NELSON, B. W., 1999, Change identification using multitemporal spectral mixture analysis: applications in eastern Amazonia. In *Remote Sensing Change Detection - Environmental Monitoring Methods and Applications*, editado por LUNETTA, R.S. e ELVIDGE, C.D. (London: Taylor & Francis), pp 137-159.

SHARFEN, G.; KHALSA, S., 2005, Assessing the utility of MODIS for monitoring snow and sea ice extent [versão electrónica]. In *Proceedings of the European Association of Remote Sensing Laboratories*, vol 2.

URL: <http://las.physik.uni-oldenburg.de/proceedings/> consulta em Fevereiro de 2005).

SETTLE, J.J, DRAKE, N.A., 1993, Linear mixing and the estimation of ground cover proportions. *International Journal of Remote Sensing*, **14**, 1159-1177.

SHIMABUKURO, Y.E.; SMITH, J.A., 1988, Shade images of forested areas obtained from Landsat MSS data. In *Proceedings of 22nd International Symposium On Remote Sensing Environment* (Abidjan, Côte D'Ivoire), pp 807-811.

SHIMABUKURO, Y.E.; NOVO, E.M.; PONZONI, F.J., 1988, Índice de vegetação e modelo linear de misturas no monitoramento da região do Pantanal. *Pesquisa Agropec. Brasileira*, **33**, 1729-1737.

STRAHLER, A.; MUCHONEY, D.; BORAK, J.; FRIEDL, M.; GOPAL, S.; LAMBIN, E.; MOODY, A., 1999, MODIS Land Cover Product. Algorithm Theoretical Basis Document (ATBD), Version 5.0 [software] .

(URL: <http://daac.gsfc.nasa.gov/data/dataset/MODIS> , consulta em Janeiro de 2005).

TATEISHI, R.; SHIMAZAKI, Y.; GUNIN, P. D., 2004, Spectral and temporal linear mixing model for vegetation classification. *International Journal of Remote Sensing*, **25**, 4203-4218.

TÖRMÄÄ, M.; LUMMEA, J.; PYYSSALOA, U.; PATRIKAINENB, N.; LUOJUSB, K., 2004, Tree species classification using ERS SAR and MODIS NDVI images [versão electrónica]. In *Proceedings of XXth International Society for Photogrammetry and Remote Sensing Congress*, **Vol VII** (Istambul).

(URL: <http://www.isprs.org/istanbul2004/comm7/papers/181.pdf>, consulta em Dezembro de 2004)

WC - Waters Corporation, 2004, *The Validation Process*, Waters Corporation ed. (URL: [Http://www.waters.com/WatersDivision/](http://www.waters.com/WatersDivision/), consulta em Novembro de 2004).

UNIVERSITY OF MARYLAND, 1997, GLCF – Global Land Cover Facility.

(URL: <http://glcf.umiacs.umd.edu/index.shtml> , consulta em Dezembro de 2005).

USTIN, S.L.; SMITH, M. O.; ADAMS, J.B., 1993, Remote sensing of ecological processes: A strategy for developing and testing ecological models using spectral mixture analysis. In *Scaling Physiological Process: Leaf to Globe*, editado por FHLERINGER, J.R. e FIELD C.B. (Academic Press), pp 339-357.

VERBYLA, D. L.; HAMMOND, T. O., 2001, *How to lies with an error matrix*.
(URL: <http://nrm.salrm.uaf.edu/~dverbyla/online/errormatrix.html>, consulta em Outubro de 2004).

VERMOTE, E. F.; VERMEULEN, A., 1999, MODIS Atmospheric Correction Algorithm: Spectral Reflectances (MOD09), Algorithm Theoretical Background Document (ATBD), Version 4.0 [software].

ZHU, L.; TATEISHI, R., 2001, Application of linear mixture models to time series of NOAA AVHRR data [versão electrónica]. In *Proceedings of 22nd Asian Conference on Remote Sensing* (Japão).
(URL: <http://www.gisdevelopment.net> , consulta em Março de 2004).

WASER, L.; ZIMMERMANN, N.; KELLER, M.; SCHWARZ, M., 2004, Rapid assessment of land cover in the european alps - evaluation of MODIS mod12 data.
(URL: http://www.wsl.ch/staff/markus.schwarz/MODIS_waser_01.pdf , consulta em Fevereiro de 2005)

ZHU, L.; TATEISHI, R., 2002, Linear mixture modelling for quantifying vegetation cover using time series NDVI data [versão electrónica]. In *Proceedings of 23rd Asian Conference on Remote Sensing* (Japão).
(URL: <http://www.gisdevelopment.net> , consulta em Março de 2004).

ANEXO 1 - Quality Factor e série multi-temporal de MOD09A1

Possíveis estados do parâmetro do Factor de Qualidade (banda 12) do produto MOD09A1

Bit	Descrição
0-1	cloud state (estado de nuvens); 00 -- clear 01 -- cloudy 10 -- mixed 11 -- not set, assumed clear
2	cloud shadow (sombras das nuvens); 0 -- no 1 -- yes
3-5	land/water flag; class definitions; 000 -- shallow ocean 001 -- land 010 -- ocean coastlines and land shorelines 011 -- shallow inland water 100 -- ephemeral water 101 -- deep inland water 110 -- continental/moderate ocean 111 -- deep ocean
6-7	aerosol quality (qualidade dos aerossóis); 00 -- climatology 01 -- low 10 -- average 11 -- high
8-9	cirrus detected (detecção neblinas); 00 -- none 01 -- small 10 -- average 11 -- high

10	internal cloud algorithm flag; 0 -- clear 1 -- cloudy
11	internal fire algorithm flag; 0 -- no fire 1 -- fire
12	MOD35 snow/ice flag; 0 -- no 1 -- yes
13-14	BRDF correction performed; 00 -- no 01 -- Montana methodology 10 -- Boston methodology
15	internal snow algorithm flag; 0 -- no snow 1 -- snow

Na banda 12 do MOD12, os numeros digitais (DN) estão expressos em decimal. Para interpretar os DN é necessário converte-los para binário. Como exemplo, um valor decimal de 72 (ou 0000.0000.0100.1000 em binário) para o estado do QA de um pixel codifica os seguintes dados:

Bit	Descrição	Valor
0-1	cloud state;	00 = limpo (clear)
2	cloud shadow;	0 = não (no)
3-5	land/water flag;	001 = Terra (land)
6-7	aerosol quantity;	01 = baixo (low)
8-9	cirrus detected;	00 = ausência (none)
10	internal cloud algorithm flag;	0 = sem nuvens (no cloud)

11	internal fire algorithm flag;	0 = sem fogo (no fire)
12	MOD35 snow/ice flag;	0 = não (no)
13-14	BRDF correction performed;	00 = sem correcção (none)
15	internal snow algorithm flag;	0 = sem neve (no snow)

Informação específica sobre o MOD09 está disponível no servidor de ftp em <ftp://modular.nascom.nasa.gov/pub/LatestFilespecs/>.

Série multi-temporal de MOD09A1

Para constituir a série multitemporal de imagens MOD09A1 – compositos de 8 dias com reflectância à superfície, foram utilizados todos os *data granule* existentes para o período de 2000/02/23 a 2001/01/31. Os dados foram obtidos em <http://edcimswww.cr.usgs.gov/pub/imswelcome/>.

Apresenta-se em seguida a listagem de todos os *data granules* utilizados:

- MOD09A1.A2000057.h17v04.004.2002357114838.hdf
- MOD09A1.A2000057.h17v05.004.2002357130937.hdf
- MOD09A1.A2000065.h17v04.004.2002359193133.hdf
- MOD09A1.A2000065.h17v05.004.2002359195900.hdf
- MOD09A1.A2000073.h17v05.004.2002362102144.hdf
- MOD09A1.A2000073.h17v04.004.2002362063314.hdf
- MOD09A1.A2000081.h17v04.004.2002363213532.hdf
- MOD09A1.A2000081.h17v05.004.2002363220646.hdf
- MOD09A1.A2000089.h17v04.004.2003001112412.hdf
- MOD09A1.A2000089.h17v05.004.2003001112924.hdf
- MOD09A1.A2000097.h17v04.004.2003003192934.hdf
- MOD09A1.A2000097.h17v05.004.2003003195911.hdf

- MOD09A1.A2000105.h17v04.004.2003005091545.hdf
- MOD09A1.A2000105.h17v05.004.2003005102905.hdf
- MOD09A1.A2000113.h17v04.004.2003008010038.hdf
- MOD09A1.A2000113.h17v05.004.2003008004220.hdf
- MOD09A1.A2000121.h17v04.004.2003009063447.hdf
- MOD09A1.A2000121.h17v05.004.2003009064310.hdf
- MOD09A1.A2000129.h17v04.004.2003012091149.hdf
- MOD09A1.A2000129.h17v05.004.2003012103058.hdf
- MOD09A1.A2000137.h17v04.004.2003014012454.hdf
- MOD09A1.A2000137.h17v05.004.2003014025410.hdf
- MOD09A1.A2000145.h17v04.004.2003015215626.hdf
- MOD09A1.A2000145.h17v05.004.2003015213552.hdf
- MOD09A1.A2000153.h17v04.004.2003018084434.hdf
- MOD09A1.A2000153.h17v05.004.2003018101116.hdf
- MOD09A1.A2000161.h17v04.004.2003021003953.hdf
- MOD09A1.A2000161.h17v05.004.2003021040426.hdf
- MOD09A1.A2000169.h17v04.004.2003310003056.hdf
- MOD09A1.A2000169.h17v05.004.2003309234638.hdf
- MOD09A1.A2000177.h17v04.004.2003028162854.hdf
- MOD09A1.A2000177.h17v05.004.2003028165123.hdf
- MOD09A1.A2000185.h17v04.004.2003297134037.hdf
- MOD09A1.A2000185.h17v05.004.2003297135057.hdf
- MOD09A1.A2000193.h17v04.004.2003032130534.hdf
- MOD09A1.A2000193.h17v05.004.2003032131610.hdf
- MOD09A1.A2000201.h17v04.004.2003307192227.hdf
- MOD09A1.A2000201.h17v05.004.2003307192431.hdf
- MOD09A1.A2000209.h17v04.004.2003308051422.hdf
- MOD09A1.A2000209.h17v05.004.2003308051644.hdf

- MOD09A1.A2000217.h17v05.004.2003308140852.hdf
- MOD09A1.A2000217.h17v04.004.2003308140900.hdf
- MOD09A1.A2000225.h17v04.004.2003309180133.hdf
- MOD09A1.A2000225.h17v05.004.2003309180617.hdf
- MOD09A1.A2000233.h17v04.004.2003321043945.hdf
- MOD09A1.A2000233.h17v05.004.2003321044936.hdf
- MOD09A1.A2000241.h17v04.004.2003043120333.hdf
- MOD09A1.A2000241.h17v05.004.2003043141129.hdf
- MOD09A1.A2000249.h17v04.004.2003049195700.hdf
- MOD09A1.A2000249.h17v05.004.2003049200556.hdf
- MOD09A1.A2000257.h17v04.004.2003051063950.hdf
- MOD09A1.A2000257.h17v05.004.2003051073921.hdf
- MOD09A1.A2000265.h17v05.004.2003055183517.hdf
- MOD09A1.A2000265.h17v04.004.2003055183610.hdf
- MOD09A1.A2000273.h17v04.004.2003059103315.hdf
- MOD09A1.A2000273.h17v05.004.2003059110523.hdf
- MOD09A1.A2000281.h17v04.004.2003059194301.hdf
- MOD09A1.A2000281.h17v05.004.2003059194959.hdf
- MOD09A1.A2000289.h17v04.004.2003063133002.hdf
- MOD09A1.A2000289.h17v05.004.2003063133913.hdf
- MOD09A1.A2000297.h17v05.004.2003064025645.hdf
- MOD09A1.A2000297.h17v04.004.2003064024738.hdf
- MOD09A1.A2000305.h17v05.004.2003067161344.hdf
- MOD09A1.A2000305.h17v04.004.2003067175802.hdf
- MOD09A1.A2000313.h17v04.004.2003068112022.hdf
- MOD09A1.A2000313.h17v05.004.2003068113236.hdf
- MOD09A1.A2000321.h17v04.004.2003071142031.hdf
- MOD09A1.A2000321.h17v05.004.2003071141955.hdf

- MOD09A1.A2000329.h17v04.004.2003071225513.hdf
- MOD09A1.A2000329.h17v05.004.2003071222338.hdf
- MOD09A1.A2000337.h17v04.004.2003074025915.hdf
- MOD09A1.A2000337.h17v05.004.2003074030710.hdf
- MOD09A1.A2000345.h17v04.004.2003074191410.hdf
- MOD09A1.A2000345.h17v05.004.2003074195656.hdf
- MOD09A1.A2000353.h17v05.004.2003076005725.hdf
- MOD09A1.A2000353.h17v04.004.2003076004649.hdf
- MOD09A1.A2000361.h17v04.004.2003076211141.hdf
- MOD09A1.A2000361.h17v05.004.2003076223438.hdf
- MOD09A1.A2001001.h17v04.004.2003078032022.hdf
- MOD09A1.A2001001.h17v05.004.2003078032640.hdf
- MOD09A1.A2001009.h17v05.004.2003079050415.hdf
- MOD09A1.A2001009.h17v04.004.2003079050002.hdf
- MOD09A1.A2001017.h17v04.004.2003079213256.hdf
- MOD09A1.A2001017.h17v05.004.2003079190208.hdf
- MOD09A1.A2001025.h17v05.004.2003083081144.hdf
- MOD09A1.A2001025.h17v04.004.2003083081359.hdf

ANEXO 2

MACRO_1

'Macro de MS Excel para separar aleatoriamente as amostras
'retiradas com o comando sample do ArcInfo (ESRI), sobre as
'imagens fracções de vegetação , solo e agua, resultado da
'aplicação de modelo de misturas lineares a uma série temporal
'de imagens Modis 9A (500m) para o ano de 2000.
'Os pixels utilizados para a constituição máscara (shp e GRID) de
'amostras foram estraidos do CLC2000 e com auxilio de imagens
'Landsat TM 2000

OBJECTIVO

'separa aleatoriamente em dois grupos (treino e validação) as
'diferentes amostras obtidas para cada tipo de classe de ocupação
'do solo.
'são eleiminadas as amostras que não são homogeneas da classe
'para tal é indicado na primeira coluna "no" no caso de serem a escluir.
'a exclusão faz-se com base no resultado do Cluster de redes neuronais,
'macro excel SOM de Angshuman Saha.
'depois de separar as amostras faz a separação em cada uma
'das fracções (vegetação, solo, agua). Calcula a média, o desvio
'padrão e a variancia (desv_padr^2). Calcula ainda a media+desv_prd;
'a média; e a média-desv_prd. Estes três ultimos é para fazer os
'gráficos de análise multi-temporal das fracções e a sua dispersão ao
'longo da série temporal. os gráficos deverão ser feitos manualmente.

Realizado por Pedro Oliveira @COSoft em Jan_2005

'Option Base 1 'indice das tabelas começam em 1
Option Explicit

'Const nr_elementos = 30

Sub Processa()

Dim i, contador, indice
Dim nr_elementos, tot_nr_elementos
Dim linha
Dim tabela(9999)

've quantos elementos (nº de records) tem a tabela

nr_elementos = 0

linha = 2

tot_nr_elementos = 0

While Cells(linha, 2).Value <> ""

tot_nr_elementos = tot_nr_elementos + 1

If Cells(linha, 1) <> "no" Then

nr_elementos = nr_elementos + 1

tabela(tot_nr_elementos) = 0

Else

tabela(tot_nr_elementos) = 2 ' os elementos deste valor não sao processados

End If

```

    linha = linha + 1
Wend

'inicia a os elementos da tabela para a série 0

contador = 0
'divide os (nº amostras - por defeito é 30) nr_elementos em 2 séries
While contador < Int(nr_elementos / 2)
    indice = Int(Rnd() * (nr_elementos - 1)) + 1 ' indice --> [0..29]+1
    If tabela(indice) = 0 Then
        tabela(indice) = 1
        contador = contador + 1
    End If
Wend

'Vamos então popular as 2 novas tabelas

Dim orig_table_start_row, orig_table_start_column
Dim dest_table_start_row, dest_table_start_column ' Variaveis que definem a origem das
tabelas
Dim Serie ' Define as 2 tabelas de destino para o qual foram divididos os
valores aleatoriamente
Dim NewSheet, originalSheet ' vars para selecção das worksheets
Dim orig_table_column_offset, dest_table_column_offset, dest_table_row_offset
Dim dest_table_cells_space ' define o espaco entre as tabelas de destino
Dim linhas_tabela_destino

    orig_table_start_row = 2 'Selection.Cells.Row
    orig_table_start_column = 2 'Selection.Cells.Column
    dest_table_start_column = orig_table_start_column - 1

linhas_tabela_destino = Int(nr_elementos / 2)
dest_table_cells_space = Int(nr_elementos / 2) + 1 + 15 ' 15 linhas para colocar o min, o
max, as médias e a variancia

' define a folha original
Set originalSheet = ActiveSheet

    ' processar cada das series
    For Serie = 0 To 1

        'cria a nova worksheet
        Set NewSheet = Sheets.Add(Type:=xlWorksheet)

' dá o nome à nova folha (sheet)
Select Case Serie
Case 0
NewSheet.Name = "validação"
Case 1
NewSheet.Name = "treino"
End Select

```

```

' posiciona na 2 linha da nova worksheet
dest_table_start_row = 2
' contador para a linha da nova serie
dest_table_row_offset = 0

' copia o titulo para cada uma das tabelas finais
' aqui copia o'am_f
NewSheet.Cells(dest_table_start_row - 1, dest_table_start_column).Value =
originalSheet.Cells(orig_table_start_row - 1, orig_table_start_column).Value
NewSheet.Cells(dest_table_start_row - 1 + (dest_table_cells_space),
dest_table_start_column).Value = originalSheet.Cells(orig_table_start_row - 1,
orig_table_start_column).Value
NewSheet.Cells(dest_table_start_row - 1 + (dest_table_cells_space * 2),
dest_table_start_column).Value = originalSheet.Cells(orig_table_start_row - 1,
orig_table_start_column).Value
' aqui copia os meses_x, onde x é v,s ou a
orig_table_column_offset = 3
dest_table_column_offset = 1
While originalSheet.Cells(orig_table_start_row - 1, orig_table_start_column +
orig_table_column_offset).Value <> ""
NewSheet.Cells(dest_table_start_row - 1, dest_table_start_column +
dest_table_column_offset).Value = originalSheet.Cells(orig_table_start_row - 1,
orig_table_start_column + orig_table_column_offset).Value '_v
NewSheet.Cells(dest_table_start_row - 1 + (dest_table_cells_space),
dest_table_start_column + dest_table_column_offset).Value =
originalSheet.Cells(orig_table_start_row - 1, orig_table_start_column +
orig_table_column_offset + 1).Value '_s
NewSheet.Cells(dest_table_start_row - 1 + (dest_table_cells_space * 2),
dest_table_start_column + dest_table_column_offset).Value =
originalSheet.Cells(orig_table_start_row - 1, orig_table_start_column +
orig_table_column_offset + 2).Value '_a

orig_table_column_offset = orig_table_column_offset + 3 ' passa para a proxima
serie de meses na tabela original
dest_table_column_offset = dest_table_column_offset + 1 ' passa para a proxima
coluna na tabela destino
Wend

' copia os elementos da tabela para a serie indicada
For i = 1 To tot_nr_elementos
If tabela(i) = Serie Then

NewSheet.Cells(dest_table_start_row + dest_table_row_offset,
dest_table_start_column).Value = originalSheet.Cells(orig_table_start_row + i - 1,
orig_table_start_column).Value '_v
NewSheet.Cells(dest_table_start_row + dest_table_row_offset +
(dest_table_cells_space), dest_table_start_column).Value =
originalSheet.Cells(orig_table_start_row + i - 1, orig_table_start_column).Value '_s
NewSheet.Cells(dest_table_start_row + dest_table_row_offset +
(dest_table_cells_space * 2), dest_table_start_column).Value =
originalSheet.Cells(orig_table_start_row + i - 1, orig_table_start_column).Value '_a

```

```

orig_table_column_offset = 3
dest_table_column_offset = 1
While originalSheet.Cells(orig_table_start_row + i - 1, orig_table_start_column +
orig_table_column_offset).Value <> ""
    NewSheet.Cells(dest_table_start_row + dest_table_row_offset,
dest_table_start_column + dest_table_column_offset).Value =
originalSheet.Cells(orig_table_start_row + i - 1, orig_table_start_column +
orig_table_column_offset).Value ' _v
    NewSheet.Cells(dest_table_start_row + dest_table_row_offset +
(dest_table_cells_space), dest_table_start_column + dest_table_column_offset).Value =
originalSheet.Cells(orig_table_start_row + i - 1, orig_table_start_column +
orig_table_column_offset + 1).Value ' _s
    NewSheet.Cells(dest_table_start_row + dest_table_row_offset +
(dest_table_cells_space * 2), dest_table_start_column + dest_table_column_offset).Value =
originalSheet.Cells(orig_table_start_row + i - 1, orig_table_start_column +
orig_table_column_offset + 2).Value ' _a
    orig_table_column_offset = orig_table_column_offset + 3 ' passa para a proxima
serie de meses na tabela original
    dest_table_column_offset = dest_table_column_offset + 1 ' passa para a proxima
coluna na tabela destino
Wend

dest_table_row_offset = dest_table_row_offset + 1 ' avança para a proxima linha
da nova tabela
End If
Next i

' calculo do mínimo, máximo, média, desvio padrao, variancia, m-dv, m, m+dv

Dim tabela_nr, a
Dim var As String ' serve para indicar qual a origem dos dados para eixo de gráfico

For tabela_nr = 0 To 2 ' percorre as 3 tabelas

    ' acrescenta no final da linha a que endmember da imagens (MML) as estatisticas
calculadas
    If tabela_nr = 1 Then
        var = " (solo)"
    Else
        If tabela_nr = 2 Then
            var = " (agua)"
        Else
            var = " (vegt)"
        End If
    End If

    a = tabela_nr * dest_table_cells_space ' calcula a linha em que começa cada tabela
' coloca o rótulo dos valores estatísticos a calcular na célula (x,1)
    NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 2 + a, 1).Value =
"Minimo"
    NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 3 + a, 1).Value =
"Máximo"

```

```

    NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 4 + a, 1).Value =
"Média"
    NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 5 + a, 1).Value =
"Desv_Padr"
    NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 6 + a, 1).Value =
"Variância"
    'deixa-se uma linha em branco
    NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 8 + a, 1).Value = "m-
dv" & var
    NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 9 + a, 1).Value = "m"
& var
    NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 10 + a, 1).Value =
"m+dv" & var

    For i = 1 To 12 ' caculo por cada mes
        NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 2 + a, i + 1).Value =
"=MIN(" & Chr(i + 65) & (dest_table_start_row + a) & ":" & Chr(i + 65) &
(dest_table_start_row + dest_table_row_offset - 1 + a) & ")"
        NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 3 + a, i + 1).Value =
"=MAX(" & Chr(i + 65) & (dest_table_start_row + a) & ":" & Chr(i + 65) &
(dest_table_start_row + dest_table_row_offset - 1 + a) & ")"
        NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 4 + a, i + 1).Value =
"=AVERAGE(" & Chr(i + 65) & (dest_table_start_row + a) & ":" & Chr(i + 65) &
(dest_table_start_row + dest_table_row_offset - 1 + a) & ")"
        NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 5 + a, i + 1).Value =
"=STDEV(" & Chr(i + 65) & (dest_table_start_row + a) & ":" & Chr(i + 65) &
(dest_table_start_row + dest_table_row_offset - 1 + a) & ")"
        NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 6 + a, i + 1).Value =
"=VAR(" & Chr(i + 65) & (dest_table_start_row + a) & ":" & Chr(i + 65) &
(dest_table_start_row + dest_table_row_offset - 1 + a) & ")"
        NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 8 + a, i + 1).Value =
"=" & Chr(i + 65) & (dest_table_start_row + linhas_tabela_destino + 4 + a) & "-" & Chr(i +
65) & (dest_table_start_row + linhas_tabela_destino + 5 + a)
        NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 9 + a, i + 1).Value =
"=" & Chr(i + 65) & (dest_table_start_row + linhas_tabela_destino + 4 + a)
        NewSheet.Cells(dest_table_start_row + linhas_tabela_destino + 10 + a, i + 1).Value =
"=" & Chr(i + 65) & (dest_table_start_row + linhas_tabela_destino + 4 + a) & "+" & Chr(i +
65) & (dest_table_start_row + linhas_tabela_destino + 5 + a)
    Next i
    Next tabela_nr

    Next Serie

End Sub

```

```

' _____MACRO_2_____
'
' macro para separar os valores medios dos clusters obtidos pela
' aplicação do programa SOM ( macro de excel desenvolvido por
' Anshuman Saha). Objectivo é fazer gráficoa individuais por cluster.
'
' _____
' requisito:
' deveram ser copiados os dados da folha de output do SOM cluster
' para uma nova folha designada de "folha1", e colocadas na celula(1,1)
' depois copiar o macro para o editor de VB ,<num módulo e correr
'
' organização do xls:
' só existem 3 fracções para cada mês ou seja, todos os meses estão da forma
' genero: mes_v (veget); mes_s (Solo) e mes_a (água ou sombra). estão
' em coluna e para cada mes estão os três seguidos. A primeira
' linha corresponde à indicação do cuslter, a celula 1,1 não tem nada
' _____ cluster_x cluster_y .... cluster_z
'   mes_v
'   mes_s
'   mes_a
'   mes+1_v
'
' _____
' é necessário fazer manualmente os gráficos de acordo com a série temporal
'
' ***** desenvolvido por Pedro Oliveira @ fev 2005*****
'
' _____

```

Option Explicit

Dim sep_linhas, sep_linhas2, sep_linhas3 As Integer

Dim i, j, x As Integer
 Dim valor1, valor2, valor3
 Dim cab_1

Dim coluna, cluster, amostras, coluna_nova As Integer

Dim row_in, coln_in As Integer

Sub teste()

' coloca activa a folha designada de folha1 (para onde se copiou os dados)
 Worksheets("folha1").Activate

' define a linhe e a coluna inicial
 row_in = 1
 coln_in = 1

' substitui o valor da celula 1 de vazio ao valor -
 ' para o programa não parar logo
 Cells(1, 1).Value = "-"

```
' vê o número de cluster
While Cells(row_in, coln_in) <> ""
    coln_in = coln_in + 1
Wend

cluster = coln_in - 1

' coloca novamente a posição inicial do cursor em celula(1,1)
row_in = 1
coln_in = 1
' vê o numero de amostras
While Cells(row_in, coln_in) <> ""
    row_in = row_in + 1
Wend
amostras = row_in - 1

' coloca novamente a posição inicial do cursor em celula(1,1)
row_in = 1
coln_in = 1

' sep_linhas é onde a posição em linha irá começar a ser escrito a nova tablea
sep_linhas = amostras + cluster
sep_linhas2 = sep_linhas + cluster + 2
sep_linhas3 = sep_linhas2 + cluster + 2

' coloca a especificação de cada uma das fracções
Cells(sep_linhas, 1).Value = "fraccao_ vegetacao"
Cells(sep_linhas2, 1).Value = "fraccao_ solo"
Cells(sep_linhas3, 1).Value = "fraccao_ sombra"

' copia cabecalhos e separa no numero de clusters

For x = 1 To cluster
cab_1 = Cells(1, x).Value
Cells(sep_linhas + x, 1).Value = cab_1
Cells(sep_linhas2 + x, 1).Value = cab_1
Cells(sep_linhas3 + x, 1).Value = cab_1
Next x

' copia os valores por mês e por fracção, e distribui
' pelas celulas correspondetes às fracções.
' começa em dois pois a primeira linha é do cabecalho

coluna_nova = 2

For i = 2 To amostras Step 3
For j = 1 To cluster

    valor1 = Cells(i, j).Value
    valor2 = Cells(i + 1, j).Value
```

```
valor3 = Cells(i + 2, j).Value

' no caso de se estar na segunda coluna, não se pode ir para trás
' isto é o mes de fevereiro, tem de ser ercito imediatamente ao
' indice dos cluster, enquanto para os seguintes passa-se de 3, em 3

If j = 1 Then
' caso seja a coluna um (mês) distribui o titulo do mes pelas tres fraccoes
Cells(sep_linhas + j, coluna_nova).Value = valor1
Cells(sep_linhas2 + j, coluna_nova).Value = valor1
Cells(sep_linhas3 + j, coluna_nova).Value = valor1
Else
'distribui os valores de cada um dos cluster pelas fraccoes e mes
Cells(sep_linhas + j, coluna_nova).Value = valor1
Cells(sep_linhas2 + j, coluna_nova).Value = valor2
Cells(sep_linhas3 + j, coluna_nova).Value = valor3

End If

Next j
' muda de coluna onde está a escrever os novos valores
coluna_nova = coluna_nova + 1
Next i

End Sub
```