

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA E GESTÃO DO CONHECIMENTO**

Sylvio Luiz Mantelli Neto

**Metodologia de Modelagem e Representação
Estatística Multivariada do Conhecimento Envolvido
na Identificação Automática dos Padrões Visuais
Atmosféricos, Existentes no Processo de Observação
Sinóptica a Partir da Superfície**

Tese submetida à Universidade Federal de Santa Catarina como parte dos requisitos para a obtenção do grau de Doutor em Engenharia e Gestão do Conhecimento.

Dr. Aldo von Wangenheim- Orientador

Dr. Enio Bueno Pereira - Co-orientador

Florianópolis, Dezembro de 2010

Metodologia de Modelagem e Representação Estatística Multivariada do Conhecimento Envolvido na Identificação Automática dos Padrões Visuais Atmosféricos, Existentes no Processo de Observação Sinóptica a Partir da Superfície

Sylvio Luiz Mantelli Neto

Esta Tese foi julgada adequada para a obtenção do título de Doutor em Engenharia e Gestão do Conhecimento, área de concentração Engenharia do Conhecimento e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento.

Dr. Roberto Carlos dos Santos Pacheco
Coordenador do Programa de Pós-Graduação do Departamento de Engenharia e Gestão do Conhecimento - UFSC

Banca Examinadora

Dr. Aldo von Wangenheim - orientador
Departamentos de Informática e Estatística e Engenharia e Gestão do Conhecimento - UFSC

Dr. Enio Bueno Pereira - co-orientador
Centro de Ciências do Sistema Terrestre do Instituto Nacional de Pesquisas Espaciais - INPE

Dr. João Artur de Souza - membro do EGC
Departamentos de Matemática e Engenharia e Gestão do Conhecimento - UFSC

Dr. Pedro Alberto Barbettá - membro externo ao EGC
Departamento de Informática e Estatística - UFSC

Dra. Marcia Akemi Yamasoe - membro externo à UFSC
Instituto Astronômico e Geofísico, Universidade de São Paulo - USP

Dr. Fernando Ramos Martins - membro externo à UFSC
Centro de Ciências do Sistema Terrestre, Instituto Nacional de Pesquisas Espaciais - INPE

Contents

Contents	iii
List of Figures	vi
List of Tables	x
Lista de Siglas	xiii
1 INTRODUÇÃO	1
1.1 Apresentação do Problema de Pesquisa	1
1.2 Objetivos	2
1.2.1 Objetivo Geral	2
1.2.2 Objetivos Específicos	2
1.3 Justificativa e Relevância do Tema	2
1.4 Ineditismo do Trabalho	3
1.5 Contribuição Teórica	3
1.6 Pressupostos do Trabalho ou Hipóteses	4
1.7 Escopo do Trabalho	4
1.8 Metodologia Científica	5
1.9 Aderência do trabalho proposto à área da Engenharia do Conhecimento e a sua multidisciplinaridade	6
1.10 Estrutura da Tese	6
2 REVISÃO DA LITERATURA	8
2.1 Métodos adotados para a observação a partir da superfície pelos OS	8
2.2 Sensores de medida de Radiação Solar	11
2.2.1 Utilizando o sensor de Brilho Solar para a estimativa das nuvens	12

2.2.2 Utilizando a radiação global (ondas curtas descendente) para a estimativa das nuvens	14
2.2.3 Utilizando a radiação combinada global e difusa (ondas curtas) para a estimativa das nuvens	17
2.2.4 Problemas de representatividade do domínio existente na avaliação dos piranômetros	20
2.2.5 Utilizando o balanço de radiação de ondas curtas e a radiação extraterrestre (ondas longas), para a estimativa das nuvens	22
2.3 SOS utilizando Câmeras	22
2.4 Resumo comparativo das características dos principais métodos utilizando piranômetros e imagens utilizados na estimativa de nuvens encontrados na literatura especializada	26
3 MATERIAIS E MÉTODOS	28
3.1 Aparato experimental	29
3.2 Análise exploratória dos padrões presentes no Hemisfério Celeste do espaço de cor	32
3.3 Ambientes do Agente Inteligente	40
3.4 O Levantamento do Domínio de Tarefa Baseado na Física Óptica-atmosférica dos Fenômenos Observados, e a sua Representatividade	41
4 RESULTADOS	45
4.1 Descrição da Análise Exploratória para a Representação dos Padrões Visuais	45
4.1.1 Padrão MIE	45
4.1.1.1 Definição gráfica do padrão MIE	46
4.1.1.2 Descrição estatística da definição gráfica do padrão MIE . .	46
4.1.1.3 Critério de classificação do padrão MIE	49
4.1.1.4 Limiar de discriminação para o padrão MIE	49
4.1.2 Padrão RAY	50
4.1.2.1 Definição gráfica do padrão RAY	50
4.1.2.2 Descrição estatística da definição gráfica do padrão RAY .	51
4.1.2.3 Critério de classificação do padrão RAY	58
4.1.2.4 Limiar de discriminação para o padrão RAY	58
4.1.3 Padrão Espalhamento Não Seletivo (ENS) ou Nuvens	59
4.1.3.1 Definição gráfica do padrão ENS	59
4.1.3.2 Descrição estatística da definição gráfica do padrão ENS .	60

4.1.3.3	Critério de Classificação do padrão ENS	60
4.1.3.4	Limiar de discriminação para o padrão ENS	60
4.1.4	Difusão do espalhamento Rayleigh ou padrão DRAY	64
4.1.4.1	Definição Gráfica do padrão DRAY	64
4.1.4.2	Observações importantes acerca da ocorrência do padrão DRAY	64
4.1.4.3	Critério de classificação do padrão DRAY	66
4.1.5	Difusão do Espalhamento Não Seletivo (DENS) ou Difusão Nuvens .	66
4.1.5.1	Definição Gráfica do padrão DENS	66
4.1.5.2	Critério de Classificação do padrão DENS	68
4.2	Ajustes Empíricos no Modelo de Representação Matemática dos Padrões Visuais, resumo dos critérios de classificação	68
4.2.1	Limitações no uso do intervalo de confiança de bonferroni no Critério de Classificação dos padrões ENS e RAY	70
4.2.2	Resumo dos Critérios de Classificação Estabelecidos	70
4.3	Síntese dos Resultados Gerais	71
4.4	Análise dos resultados	74
4.5	Comparação dos resultados com outros métodos encontrados na literatura especializada	75
4.6	Validação dos dados utilizando imagens de Satélite	79
5	Conclusões e sugestões para trabalhos futuros	89
5.1	Conclusões	89
5.2	Sugestões de Trabalhos Futuros	90
Bibliography		92
Glossário		97
Apêndice A		100
Anexo A		158
Anexo C		172

List of Figures

2.1	Critérios utilizados para classificação das nuvens.	10
2.2	Exemplos típicos de sensor de brilho solar.	13
2.3	Exemplos de tipos de piranômetros que medem diferentes componentes da radiação solar.	14
2.4	Detalhes da resposta espectral produzida pelo sol e a sensibilidade dos sensores típicos de radiação solar.	15
2.5	Metodologia básica para estimativa do tipo de nuvem proposta por (DUCHON; O'MALLEY, 1999) utilizando piranômetro.	16
2.6	Exemplo ilustrativo de modelamento do Cloud Forcing para os dados obtidos a partir da estação do LABSOLAR da UFSC de Florianópolis para o mês de Abril de 2004.	20
2.7	Problemas relacionados à distribuição espacial e representatividade do céu. .	21
2.8	Exemplos típicos de SOS em operação.	23
2.9	Efeito do filtro de difusão discretizando e acentuando as bordas. Figura extraída de (MANTELLI, 2001) capítulo 4.	25
2.10	Esquema de cores do espaço RGB. A utilização somente dos canais vermelho e azul restringe o domínio das variáveis independentes a cores que variam somente de preto a magenta. Figura extraída de (GONZALEZ; WOODS, 2002) capítulo 6.	26
3.1	Esquema geral ilustrando todas as fases, aplicativos e procedimentos utilizados durante as pesquisas desenvolvidas na presente tese.	30
3.2	Remoção dos fatores que interferem na análise da imagem através de mascaraamento (redução dos valores dos píxeis a zero), utilizando a ferramenta GT tool do LAPIX. Construção da máscara das obstruções horizontais, banda de sombreamento móvel e da máscara resultante.	31
3.3	Interface gráfica desenvolvida para análise, implementação das técnicas desenvolvidas e visualização dos resultados.	32
3.4	Ilustração do fragmento de código gerado a partir da manipulação simbólica do MAPLE para a racionalização das equações lineares do cálculo da distância de MAHALANOBIS (equação 3.9), que reduziu as equações lineares a um conjunto simples de somas e subtrações.	32

3.5 Ilustração da ocorrência do <i>locus</i> típico do padrão de nuvens observado em uma imagem do TSI, com a sua visualização nos espaços de cores. As indicações "Ra", "Mi" e "Bo", marcam o <i>locus</i> típico para os padrões Rayleigh, Mie e a borda + banda de sombreamento respectivamente. As unidades são em intensidades relativa dos píxeis para o espaço de cores RGB, e normalizados para o HSL. Figura obtida de (MANTELLI et al., 2010)	33
3.6 Ilustração da ocorrência do <i>locus</i> típico do padrão de céu observado na imagem do TSI, com a sua visualização nos espaços de cores. As indicações "Ra", "Mi" e "Bo", marcam o <i>locus</i> típico para os padrões Rayleigh, Mie e a borda + banda de sombreamento respectivamente. As unidades são em intensidades relativa dos píxeis para o espaço de cores RGB, e foram normalizados para o HSL. Figura obtida de (MANTELLI et al., 2010)	34
3.7 Análise utilizada no espaço de cor RGB, para a classificação dos píxeis das imagens do TSI utilizando a EGD e a EGP em relação à diagonal principal. As unidades são em intensidade relativa dos píxeis.	35
3.8 Hiperelipse como superfície de classificação de pontos no espaço tridimensional. Adaptada a partir da figura 4.28 (DUDA; HART; STORK, 2001). . .	38
3.9 Exemplos dos fenômenos físicos observáveis no CH e o mapeamento no <i>locus</i> típico no espaço de cor RGB. Figura obtida de (MANTELLI; WANGENHEIN; PEREIRA, 2009)	43
3.10 Exemplos de padrões de cores ópticos-atmosférico causados pela luz solar. . .	44
 4.1 Definição do padrão MIE através das imagens e a representação do <i>locus</i> típico dos píxeis nos espaços de cores RGB e HSL. Na primeira linha de figuras a imagem original. Na segunda, a imagem subtraída dos fatores que não representam informação útil. Na terceira fileira a imagem indicando somente a presença do espalhamento MIE.	47
4.2 AED do padrão Mie, após a eliminação dos <i>outliers</i>	48
4.3 Visualização da composição do padrão Rayleigh em dois níveis de luminosidade diferentes no espaço de cores RGB.	51
4.4 Definição do padrão RAY através das imagens e a representação do <i>locus</i> típico dos píxeis nos espaços de cores RGB e HSL. Na primeira linha de figuras a imagem original, na segunda a imagem subtraída dos fatores que não representam informação útil, na terceira fileira a imagem indicando somente a presença do espalhamento Rayleigh com eliminação de outliers através da EGD e dos pixeis saturados.	52
4.5 Definição do padrão Rayleigh através das imagens e a representação do <i>locus</i> típico dos píxeis nos espaços de cores RGB e HSL. Na primeira linha de figuras a imagem original, na segunda a imagem subtraída dos fatores que não representam informação útil, na terceira fileira a imagem indicando somente a presença do espalhamento Rayleigh.	53
4.6 AED do padrão RAY, após a eliminação dos <i>outliers</i> da imagem 200211031130. .	54
4.7 AED do padrão RAY, após a eliminação dos <i>outliers</i> da imagem 200211031500. .	55

4.8	AED do padrão RAY, após a eliminação dos <i>outliers</i> e padrões espúrios e resultante da concatenação das imagens 200211031130 e 200211031500.	57
4.9	Definição do padrão ENS através das imagens e a representação do <i>locus</i> típico dos píxeis nos espaços de cores RGB e HSL. Na primeira linha de figuras a imagem original. Na segunda, a imagem subtraída dos fatores que não representam informação útil. Na terceira fileira a imagem indicando somente a presença do espalhamento ENS.	61
4.10	AED do padrão ENS, após a eliminação dos <i>outliers</i>	62
4.11	Definição do padrão difusão Rayleigh através das imagens e a representação do <i>locus</i> típico dos píxeis nos espaços de cores RGB e HSL. Na primeira linha de figuras a imagem original. Na segunda, a imagem subtraída dos fatores que não representam informação útil. Na terceira fileira a imagem indicando somente a presença da difusão do espalhamento Rayleigh.	65
4.12	Definição do padrão ENS através das imagens e a representação do <i>locus</i> típico dos píxeis nos espaços de cores RGB e HSL. Na primeira linha de figuras a imagem original. Na segunda, a imagem subtraída dos fatores que não representam informação útil. Na terceira fileira a imagem indicando somente a presença do espalhamento DENS.	67
4.13	Gradiente de profundidade óptica entre os padrões RAY e ENS.	69
4.14	Ajuste dos limiares de discriminação para os padrões RAY e ENS.	69
4.15	Figura indicando a redução da variabilidade do intervalo de confiança de bonferroni, para a construção da dimensão vermelha do padrão ENS, com os seguintes parâmetros: $F_{p,n-p,(\alpha=0,01)} = 6,24$ e $\sigma_i^2 = 28,68$	70
4.16	Gráfico indicando a composição dos padrões em porcentagem presentes no céu ao longo do dia no eixo vertical para o dia 03 nov. 2002. O eixo horizontal indica o horário GMT	71
4.17	Porcentagem dos padrões distribuídos ao longo do dia.	73
4.18	Representação esquemática indicando regiões e pontos notáveis dos padrões abordados no presente trabalho no espaço de cores RGB.	76
4.19	Correlação entre LONG, EGD e Mahalanobis para céu claro.	76
4.20	Correlação entre LONG, EGD e Mahalanobis para céu encoberto.	77
4.21	Diferenças absolutas LONG (10,42 a - 24,13) , EGD(+ 0,03 a -31,63) e Mahalanobis para céu claro.	78
4.22	Diferenças absolutas LONG (10,42 a - 24,13) , EGD(+ 0,03 a -31,63) e Mahalanobis para céu claro.	79
4.23	Distribuição dos píxeis da imagem de satélite nos espaços de cores RGB e HSL.	80
4.24	Comparação da estimativa da cobertura de nuvens LONG , EGD , MAHA e SAT para céu claro.	86
4.25	Diferenças absolutas entre LONG , EGD , MAHA e SAT para céu claro.	87

4.26 Correlação entre todos os métodos e o método de estimativa de cobertura de nuvens de satélite.	88
5.1 Ilustrações das imagens originais.	160
5.2 cont. Ilustrações das imagens originais.	161
5.3 Ilustrações das imagens originais mascaradas.	162
5.4 cont. Ilustrações das imagens originais mascaradas.	163
5.5 Ilustrações das imagens analisadas.	164
5.6 cont. Ilustrações das imagens analizadas.	165
5.7 Legenda de cores dos padrões classificados.	166
5.8 Imagens do satélite GOES 8 disponíveis entre 09:39 e 11:09 (b),(e),(h),(k), obtidas em um horário intermediário da imagens do imageador de nuvens antes (a),(d),(g),(j) e depois (c),(f),(i),(l).	167
5.9 Imagens do satélite GOES 8 disponíveis entre 11:39 e 14:09 (b),(e),(h),(k), obtidas em um horário intermediário da imagens do imageador de nuvens antes (a),(d),(g),(j) e depois (c),(f),(i),(l).	168
5.10 Imagens do satélite GOES 8 disponíveis entre 14:39 e 16:39 (b),(e),(h),(k), obtidas em um horário intermediário da imagens do imageador de nuvens antes (a),(d),(g),(j) e depois (c),(f),(i),(l).	169
5.11 Imagens do satélite GOES 8 disponíveis entre 17:09 e 19:39 (b),(e),(h),(k), obtidas em um horário intermediário da imagens do imageador de nuvens antes (a),(d),(g),(j) e depois (c),(f),(i),(l).	170
5.12 Imagens do satélite GOES 8 disponível entre 20:09 (b) e 2039 (e), obtida em um horário intermediário da imagens do imageador de nuvens antes (a), (d) e depois (c), (f).	171

List of Tables

2.1 Resumo comparativo dos principais métodos utilizando piranômetros encontrados na literatura especializada. RT é a resolução temporal em minutos, RE é a resolução espacial em porcentagem, OS é o método utilizado pelos operadores sinópticos, PB é o método utilizando o sensor de Brilho Solar utilizado por (PALLÉ; BUTLER, 2002), DO é o método utilizando piranômetro desenvolvido por (DUCHON; O'MALLEY, 1999), KC é o método utilizando piranômetro desenvolvido por (KARSTEN; CZEPLAK, 1980), HCH é o método utilizando piranômetro desenvolvido por (HARRISON; CHALMERS; HOGAN, 2008)	27
2.2 Resumo comparativo dos principais métodos utilizando imagens encontrados na literatura especializada. SE é o método de estimativa de nuvens utilizado por (SOUZA-ECHER et al., 2006), LO é o método utilizando piranômetro desenvolvido por (LONG; SLATER; TOOMAN, 2001), MA1 é o método utilizando piranômetro desenvolvido por (MANTELLI, 2001), CA é o método utilizando piranômetro desenvolvido por (CAZORLA; OLMO; ALADOS-ARBOLEDAS, 2008), MA2 é o método utilizando piranômetro desenvolvido por (MANTELLI et al., 2010)	27
4.1 Comparação da variabilidade dos intervalos de confiança de bonferroni com a variabilidade do desvio padrão para o padrão ENS.	70
4.2 Resumo dos métodos de classificação estabelecidos. H é a ordem hierárquica, P é o padrão.	71
4.3 Resumo dos resultados gerais.	72
4.4 Resumo dos resultados gerais (cont.).	74
4.5 Níveis de luminância natural encontrados em alguns ambientes externos. Extraído de (REINHARD et al., 2006) capítulo 1 tabela 1.1.	74
4.6 Matrix média de Intensidade para 6 imagens, dos píxeis ao redor da estação de Florianópolis, caracterizando o céu claro.	80
4.7 Matrix de repetitividade da estimativa da média de Intensidade de céu claro para 6 imagens, dos píxeis ao redor da estação de Florianópolis no ponto 151 x 45.	81

4.8	Matrix de valores limiares de céu claro ja arredondados para inteiros, para teste de hipótese unilateral dos píxeis ao redor da estação de Florianópolis no ponto 151 x 45.	81
4.9	Matrix de valores limiares de céu claro LIM já arredondados para inteiros, para teste de hipótese unilateral dos píxeis ao redor da estação de fpolis. X e Y são as coordenadas dos píxeis na imagem, MED é a média dos valores dos píxeis para céu claro, STD o desvio padrão, POS é a posição do pixel em relação ao píxel central C no ponto 151 x 45. Os pontos foram mapeados nas posições	82
4.10	Valores das imagens de satélite relativo à estação de Florianópolis	83
4.11	Resultados da comparação dos resultados da validação.	84
4.12	Valores da cobertura das nuvens a partir das imagens de satélite	85
4.13	Coeficientes de correlação os métodos de estimativa da cobertura de nuvens e o método de satélite.	86
5.1	Dados das comparações entre os métodos encontrados na literatura dos métodos (LONG; SLATER; TOOMAN, 2001)(L), (MANTELLI et al., 2010b)(E) e o proposto no presente trabalho(M).	159

"I'd like to share a revelation that I've had during my time here. It came to me when I tryed to classify your species and I realized that you're not actually mammals. Every mammal on this planet instinctively develops a natural equilibrium with the surrounding environment. But you humans do not. You moved to an area multiply, and multiply, until every natural resource is consumed. The only way you can survive is to spread to another area. There is another organism on this planet that follows the same pattern. Do you know what it is? A virus. Human beings are a disease. A cancer of this planet. You are a plague. And we are the cure."

Fragment of a speech made by agent Smith on the movie matrix, when he was torturing Morpheus to find out the location of Zion. Humans are remarkably

sensitive to the statistical regularities of their environment. (STEYVERS; GRIFFITHS; DENNIS, 2006) p. 333.

Listas de Siglas

AI Agente Inteligente

CCD Charged Coupled Device

CIRAM Centro de Informações de Recursos Ambientais e de Hidrometeorologia de Santa Catarina da EPAGRI
<http://ciram.epagri.sc.gov.br/portal/website/>

CF cloud forcing

CPTEC Centro de Previsão do Tempo e Estudos Climáticos do INPE
<http://www.cptec.inpe.br/>

EGC Departamento de Engenharia e Gestão do Conhecimento da UFSC
<http://www.egc.ufsc.br/>

EGD Distância Euclideana Geométrica

EGP Projeção Euclideana Geométrica

EMC Departamento de Engenharia Mecânica da UFSC
<http://www.emc.ufsc.br/>

EPAGRI Empresa de Pesquisa Agropecuária de Santa Catarina
<http://www.epagri.rct-sc.br/>

F_c fator de nuvens

FD Fração Difusa

F_s fator de brilho solar

GMT Greenwich Mean Time

GT Padrão ouro ou ground truth

HC Hemisfério Celestial

HSL HUE, SATURATION and LIGHTNESS ou matix, saturação e brilho. Espaço de cores tridimensional no formato de um cilindro

IA Inteligência artificial

INMET Instituto Nacional de Meteorologia <http://www.inmet.gov.br/>

INE Departamento de Informática e Estatística da UFSC <http://www.inf.ufsc.br/>

INPE Instituto Nacional de Pesquisas Espaciais <http://www.inpe.br>

IQD Distância Interquartil

Kt Índice de claridade

LABSOLAR Laboratório de Energia Solar do Departamento de Engenharia Mecânica da UFSC <http://www.lepten.ufsc.br/>

LAPIX Laboratório de Processamento de Imagens e Computação Gráfica do Departamento de Informática e Estatística da UFSC <http://www.lapix.ufsc.br/>

LASER Light Amplification by Stimulated Emittion of Radiation

LIDAR LASER Detection and Ranging

MODTRAN Modelo físico atmosférico desenvolvido pela força aérea Americana <http://www.modtran.org/>

OES Observatório Espacial do Sul do INPE em São Martinho da Serra, RS <http://www.inpe.br/crs/>

Op opacidade

OS Observador Sinóptico

pirradiômetro Sensor combinado para medir simultaneamente a radiação global e o albedo.

RADAR Radio Detection and Ranging

RE Resolução Espacial

RT Resolução Temporal

RGB RED, GREEN e BLUE ou vermelho, verde e azul. Espaço de cores tridimensional no formato de um cubo (cartesiano)

SODAR Sound Detection and Ranging, ou radar acústico

SOS Sistema de Observação Sinóptica

TOA Valor da radiação no topo da atmosfera

TSI Total Sky Imager

UFSC Universidade Federal de Santa Catarina <http://www.ufsc.br/>

WMO World Meteorological Organization - Organização Meteorológica Mundial <http://www.wmo.int/>

Chapter 1

INTRODUÇÃO

1.1 Apresentação do Problema de Pesquisa

As nuvens são um conjunto de fenômenos atmosféricos de grande impacto nas atividades diárias do cidadão comum. Além do claro indicativo da chegada de chuvas, descargas atmosféricas etc., elas influenciam em muito o clima local (HU; STAMNES, 2000) são o maior fator modulador da energia solar que atinge a superfície da Terra; contribuindo para o aquecimento (HARRISSON et al., 1993) ou resfriamento (YAMANOUCHI, 1993) da superfície do planeta. As nuvens também se interpoem como obstáculo no campo de observação de determinados campos da ciência dificultando a interpretação do objeto de pesquisas; como por exemplo Geoprocessamento, Astronomia, Sensoriamento Remoto, etc. Nos modelos atmosféricos de previsão e levantamento de recursos renováveis que utilizam imagens de satélite, a presença das nuvens causam um aumento na variabilidade das incertezas dos modelos, nas condições de cobertura parcial ou total do céu. Por isto nestes modelos, torna-se necessária a determinação mais refinada não somente das nuvens mas também de outros fenômenos que precisam ser considerados de forma mais apropriada. Mas uma estimativa mais refinada da quantidade das nuvens, para ser considerada nestes modelos ainda é um desafio, devido às variações nas condições climáticas e de observação. As nuvens podem ser avaliadas a partir de satélites ou da superfície. O presente trabalho se concentrará nos aspectos de observação a partir da superfície, cuja metodologia de avaliação precisa ser refinada para apresentar um desempenho melhor. A observação das nuvens e sua estimativa de altitude e distribuição vertical também é importante para muitas atividades, especialmente para a aviação e as aplicações meteorológicas.

As nuvens não são o único fenômeno físico observável a partir das imagens no domínio da radiação visível. Outros fenômenos também se manifestam oriundos da interação da energia solar com os constituintes básicos existentes na atmosfera como o vapor de água, moléculas, aerossóis, etc. São justamente estes fenômenos que precisam ser melhor avaliados, por que causam erros de interpretação na análise das imagens pelos sistemas automáticos. Buscar uma metodologia apropriada para determinar não apenas a quantidade de nuvens mas também representar todos estes fenômenos de forma automática para consideração nos modelos mencionados é o principal desafio a ser abordado no presente trabalho.

1.2 Objetivos

1.2.1 Objetivo Geral

O objetivo do presente trabalho é estabelecer uma metodologia mais precisa de interpretação, representação e classificação automática a partir dos valores dos píxeis coloridos das imagens para mapeá-los nos respectivos padrões de conhecimento estabelecidos pela teoria física atmosférica; visando substituir os observadores sinópticos por um agente inteligente.

1.2.2 Objetivos Específicos

- Revisão bibliográfica visando um levantamento dos diferentes métodos automáticos de observação das nuvens existentes na literatura especializada a partir da superfície.
- Propor um domínio de trabalho apropriado, para representar os padrões coloridos correspondentes aos fenômenos atmosféricos observáveis;
- Efetuar uma análise geométrica no espaço multidimensional de cores, verificando a ocorrência do *locus* típico dos padrões.
- Efetuar uma análise exploratória de dados nos padrões identificáveis através das imagens;
- Determinar um modelo matemático para representar os dados relativos a estes padrões identificáveis;
- Implementar um aplicativo para validar a metodologia proposta;
- Comparar a metodologia apresentada com outros métodos existentes, apresentando conclusões e avaliando de forma crítica o método proposto.
- Validar a metodologia desenvolvida utilizando imagens de satélite.
- Apresentar trabalhos futuros visando a continuidade ou refinamento da técnica desenvolvida.

1.3 Justificativa e Relevância do Tema

A quantificação das nuvens pode ser efetuada a partir de imagens de satélites ou da superfície. Mas cada um destes pontos de observação possuem limitações inerentes. Existe um consenso geral de que para reduzir as incertezas dos dois métodos de observação deve-se utilizá-los simultaneamente de forma a combinar as suas melhores características. Entretanto existe um problema de escala de observação a ser considerado. Os satélites possuem escala de observação continental, enquanto que a observação de superfície possui a escala de observação local. A correlação entre uma estação de superfície de medida de radiação solar e a radiação média nas estações vizinhas é de 0.9 para uma área de aproximadamente 30 km de distância (BARNNET et al., 1998). Nestas condições a observação a partir da superfície seria utilizada como um ponto de validação, representativo de alguns píxeis na imagem. A cobertura total

de grandes extensões de terra para validação na superfície é inviável. Entretanto com vários pontos distribuídos ao longo do território seria possível obter uma representação suficiente para reduzir as incertezas da observação utilizando extração. Este procedimento seria o mesmo que o adotado atualmente para validação das imagens de satélite através de medidas no solo (i.e. temperatura, umidade, vento, radiação solar, etc.). Uma análise das limitações inerentes aos métodos de observação a partir de satélite e da superfície, já foi efetuada pelo autor em sua dissertação de mestrado (MANTELLI, 2001). O presente trabalho se concentrará principalmente em novos paradigmas de observação e classificação dos padrões a partir da superfície, conforme sugerido como trabalhos futuros no texto da dissertação.

1.4 Ineditismo do Trabalho

O ineditismo do presente trabalho, encontra-se nos seguintes aspectos.

- Utilizar técnicas de análise e modelamento da área de Engenharia do conhecimento para a solução dos problemas tecnológicos que envolvem aspectos transdisciplinares nas seguintes áreas: meteorologia, física da atmosfera, processamento de imagens, inteligência artificial e representação matemática.
- Propor um novo paradigma de interpretação e análise automática dos padrões atmosféricos a partir de câmeras de superfície, como alternativa à análise restrita utilizada atualmente.
- Estabelecer um modelo para representar matematicamente o domínio de observação cognitivo dos Observador Sinóptico (OS).
- Utilizar o conceito de agentes inteligentes Agente Inteligente (AI) (RUSSELL; NORVIG, 2003) na solução do problema.
- Permitir a expansão de características observáveis atualmente utilizadas, através da redução das simplificações matemáticas que limitam a utilização de maiores recursos existentes, expandindo as dimensões disponíveis e não utilizadas no espaço vetorial de cores.

1.5 Contribuição Teórica

A modelagem cognitiva de padrões de observação humana é um assunto que vem sendo pesquisado de maneira a substituir atividades desgastantes desenvolvidas pelos OS. Substituir operadores humanos por sistemas automáticos é sempre um desafio. O desafio envolvido na presente pesquisa está na capacidade de reproduzir a habilidade do observador sinóptico de superfície (WMO, 1995) em classificar as condições do céu. A observação sinóptica de superfície é executada a partir do solo por um operador humano ou sistema automático. No presente trabalho o operador humano será convencionado como observador sinóptico OS e o sistema automático como sistema de observação sinóptico Sistema de Observação Sinóptica (SOS). No presente trabalho além da modelagem cognitiva de uma observação sinóptica, também se buscará uma modelagem estatística, uma vez que os padrões atmosféricos se repetem de forma sistemática. Se tais padrões são repetitivos eles podem ser

modelados. Não foi encontrada na literatura especializada durante a pesquisa bibliográfica, nenhuma contribuição utilizando técnicas estatísticas multivariadas para o modelamento cognitivo dos padrões visuais atmosféricos.

Redes neurais tem sido utilizadas nesta tarefa, entretanto em muitos casos as redes neurais apenas mapeiam um conjunto de treinamento cuidadosamente preparados em uma solução. A rede é treinada para reproduzir e ajustar ciclicamente os resultados até atingir um erro de classificação aceitável. Muitas vezes neste processo não existe a preocupação de se utilizar uma representação matemática rigorosa dos padrões (ou categorias, conjunto de treinamento) a serem mapeados, somente na reprodução dos mesmos. Nesta fase do desenvolvimento tecnológico a solução deste problema esta além do uso apenas de uma rede neural. Em muitos casos, as soluções utilizando as redes neurais tem sido implementadas como uma "caixa-preta", apresentando uma incapacidade de identificar a relevância das variáveis independentes e gerar um conjunto de regras para expressar um modelo operacional (ZHANG, 2007). (ZHANG, 2007) também menciona que, em muitos casos, as redes neurais não oferecem o mesmo tipo de entendimento e interpretação dos problemas, tão boas quanto os modelos estatísticos. A utilização de redes neurais será considerada como trabalho futuro depois que a metodologia proposta for estabelecida e validada. A teoria estatística além de ser um modelo matemático mais rigoroso, permite a construção de um conhecimento acerca do domínio de dados utilizado (ZHANG, 2007).

1.6 Pressupostos do Trabalho ou Hipóteses

Demonstrar que os padrões cognitivos observáveis na atmosfera representados através dos valores dos píxeis presentes nas imagens (variáveis independentes) podem ser modelados através da estatística multivariada nos padrões correspondentes conforme a teoria estabelecida pela física atmosférica (variáveis dependentes), automaticamente.

1.7 Escopo do Trabalho

O presente trabalho envolve a análise de imagens obtidas a partir da superfície para substituir um observador humano. A análise Bayesiana será utilizada por que estes métodos tem sido utilizados recentemente na representação do conhecimento humano ao invés das técnicas tradicionais, permitindo um modelamento supervisionado dos padrões observáveis. Serão utilizadas somente imagens diurnas, por que as atividades desenvolvidas considerarão somente a parte visível do espectro de ondas. As imagens possuem um ângulo de visão de 160 graus por limitações do imageador de nuvens. O modelamento matemático utilizará a análise multivariada restrita a 3 dimensões por que os píxeis coloridos possuem representação tridimensional. O principal espaço de cor a ser considerado no presente trabalho é o RED GREEN BLUE (RGB), por ser o padrão utilizado na maioria dos dispositivos de obtenção e reprodução de imagens. Eventualmente será utilizado o espaço de cor HUE SATURATION LIGHTNESS (HSL) para se obter informações complementares na análise das cores presentes nas imagens. O HSL fornece uma representação de cores muito próxima da percepção humana. No HSL os padrões de interesse para a classificação automática não apresentam distribuição geométrica linear e distribuição estatística normal ao longo deste espaço de cor, dificultando o modelamento matemático conforme pode ser observado nas figuras 3.5 (f)

e 3.6 (f). Será considerado somente a distribuição normal para todos os dados correspondentes aos padrões atmosféricos, conforme será mostrado na Análise exploratória de dados. O problema da ocorrência de pequenos desvios na normalidade dos padrões analisados é contornado pela quantidade massiva de dados envolvidos na análise (MONTGOMERY; RUNGER, 2003).

Embora deva ser avaliada a variabilidade da iluminação natural ao longo do dia em relação ao ajuste automático efetuado pela câmera, no presente trabalho será mencionado como sugestão de um trabalho de pesquisas no futuro devido à sua complexidade.

Atualmente existem diversos métodos de estimativa de nuvens a partir da superfície. Dentre os métodos em questão não serão considerados os métodos de identificação que utilizam sistemas complexos como : são equipamentos que utilizam LASER através de tecnologia LIDAR, utilizado para determinar a altitude da base das nuvens. Em geral estão presentes nos aeroportos e indicam o "teto" de vôo dos aviões. (ceilômetros), : LASER Detection and Ranging. O LIDAR é um equipamento que utiliza o mesmo princípio de funcionamento do RADAR, mas utiliza LASER ao invés de ondas de rádio. O LIDAR emite periodicamente um feixe de LASER na posição zenital. O LIDAR analisa o meio de propagação do feixe de LASER através de um fenômeno chamado retroespelhamento (backscattering). Através da análise das reflexões retroespelhadas da luz consegue-se determinar a distância e a atenuação do meio de propagação do LASER (LIDAR), Radio Detection and Ranging (RADAR), etc., por possuírem custo de aquisição elevado comparado aos sensores avaliados na presente tese, exigem uma equipe técnica e científica especializada para operação e interpretação dos dados e também apresentam restrições no ângulo de observação (zenital). Por estes motivos a presente tese se limitará aos piranômetros e câmeras de superfície.

O presente método não utilizará dados obtidos a partir de modelos físicos que utilizam programas do tipo : MODerate resolution atmospheric TRANsmision. Modelo computacional utilizado para modelar a propagação de ondas eletromagnéticas na atmosfera na faixa espectral 100-50,000cm⁻¹ (0.2 to 100um). (MODTRAN) desenvolvido pela força aérea Americana, devido à restrições de custos para o licenciamento do produto.

1.8 Metodologia Científica

A metodologia Bayesiana (aprendizado supervisionado por indução e análise) será utilizada no presente trabalho da seguinte maneira. Os padrões (ou categorias) a serem utilizados serão definidos a partir da teoria Física da Atmosfera. A partir destes padrões serão determinadas, as manifestações visuais identificáveis nas imagens obtidas a partir da superfície. Após a determinação será feita uma análise exploratória no espaço vetorial de cores para verificar o *locus* típico de ocorrência destes padrões. Este *locus* será caracterizado estatisticamente através de uma análise exploratória utilizando técnicas multivariadas. A caracterização na análise será utilizada para determinar os padrões em outras imagens, através da distância estatística ou distância de (MAHALANOBIS, 1936).

1.9 Aderência do trabalho proposto à área da Engenharia do Conhecimento e a sua multidisciplinaridade

Muitas soluções tecnológicas não apresentam resultados satisfatórios por utilizarem uma abordagem inapropriada para a solução do problema em questão. Substituir operadores humanos que interpretam fenômenos físicos baseados em análise subjetiva (WMO, 2008), por sistemas automáticos é um desafio. No caso da presente tese, modelar o domínio dos dados representados pelos valores dos píxeis coloridos e mapeá-los em conhecimento representados pelos padrões atmosféricos exige um conjunto de técnicas e equipamentos que transcende o campo de conhecimento e atuação da meteorologia ou do processamento de imagens. Os SOS baseados em imagens existentes no mercado utilizam técnicas triviais e simplificadas (LONG; SLATER; TOOMAN, 2001), que indicam que a solução foi encontrada em apenas uma única área de conhecimento. Tal solução se reflete no desempenho e na incerteza do sistema.

Para refinar o método de análise e reduzir as incertezas envolvidas no processo automático de estimativa das nuvens, o presente trabalho utilizará o conceito de agente inteligente, no qual se busca um artefato que visa substituir um operador humano. Entretanto substituir uma atividade de monitoração humana por um sistema automático com um certo nível de inteligência embutida é uma pesquisa que envolve diversos campos do conhecimento. Entre elas pode-se citar o conhecimento específico do domínio de aplicação (Física da Atmosfera, Sensoriamento Remoto, Meteorologia), cognição humana (no estabelecimento do domínio e na modelagem dos padrões visuais), Inteligência Artificial (IA), processamento de imagens, cálculos matemáticos e estatístico além de aspectos de análise e implementação computacional. A área da Ciência na qual todos estes procedimentos transdisciplinares podem ser modelados apropriadamente é a Engenharia do Conhecimento.

1.10 Estrutura da Tese

Para cumprir os objetivos propostos, o presente trabalho apresenta-se organizado da seguinte maneira.

O primeiro capítulo apresenta uma introdução do assunto em um contexto geral, seguido dos objetivos gerais, dos específicos, a contextualização do trabalho proposto e sua aderência à área de Engenharia do Conhecimento.

No segundo capítulo será apresentada a revisão da literatura, visando buscar uma fundamentação teórica ao trabalho que esta sendo desenvolvido.

No terceiro capítulo será apresentada a metodologia, os materiais e métodos utilizados na presente tese.

No quarto capítulo serão apresentados os resultados obtidos durante a presente pesquisa, o modelamento dos padrões e a sua representação. A análise dos resultados e a comparação com os outros métodos encontrados na literatura especializada. Por último a validação dos métodos através das imagens de satélite.

No quinto capítulo serão apresentadas as conclusões e as sugestões de trabalhos futuros.

As referência bibliográficas encontram-se depois do capítulo 5.

O apêndice A apresenta os resultados intermediários obtidos durante a presente pesquisa e divulgados através de publicações científicas. Os apêndices A1 e A2, são de trabalhos aceitos e já publicados. Os apêndices A3 e A4 são de trabalhos aceitos mas ainda não publicados. Os apêndices A5 e A6 são de trabalhos ainda em fase de submissão.

Apêndice A1, XXIX Congresso da Sociedade Brasileira de Computação - Colóquio em Informática Brasil INRIA. Esta publicação indica a importância da análise correta do domínio e estabelecimento dos critérios de classificação.

Apêndice A2, IEEE Computer Based Medical Systems. Esta publicação descreve aspectos de análise e modelamento de dados importantes na presente tese.

Apêndice A3, Journal of Applied Meteorology and Climatology. Esta publicação descreve resultados preliminares e importantes alcançados durante a pesquisa.

Apêndice A4, 1st Congress of Computational Interdisciplinary Sciences. Resumo do apêndice 6 apresentado em painel no congresso.

Apêndice A5, Computer Vision and Applications (submitted). Esta publicação ilustra o mesmo modelamento proposto, inclusive co análise sensorial, aplicado em outro campo de conhecimento.

Apêndice A6, Journal of Computational Interdisciplinary Sciences (submitted). Esta publicação ilustra a comparação do método utilizando imagens com sensores de superfície para avaliação quantitativa das nuvens.

O apêndice B apresenta um conjunto de imagens obtidos a partir do imageador de nuvens, utilizados para testar a metodologia utilizada.

O anexo A apresenta o conjunto de imagens originais, o conjunto de imagens mascaradas, o conjunto de imagens analisadas a partir da metodologia proposta. Apresenta também uma síntese tabular dos resultados das comparações numéricas e as imagens do satélite GOES-8.

O anexo B apresenta as equações utilizadas no cálculo da fração difusa, opacidade e índice de claridade.

O anexo C apresenta as equações utilizadas no cálculo das efemérides solar.

A tese também apresenta uma lista de siglas e um Glossário para facilitar a leitura e interpretação do texto.

Chapter 2

REVISÃO DA LITERATURA

No presente capítulo será feito um levantamento dos métodos e técnicas automáticas empregados para a observação das nuvens a partir da superfície, descritas na literatura especializada. Serão descritas e analisadas de forma resumida, as técnicas de observação utilizadas pelos OS; pelos sensores de medida de radiação solar individualmente, pelos sensores de radiação combinados; e pelos SOS que utilizam métodos de análise em imagens obtidas a partir de câmeras digitais.

De acordo com as normas estabelecidas pela Organização Mundial de Meteorologia (WMO - World Meteorological Organization) (WMO, 1995), são utilizados atualmente dois métodos de observação a partir da superfície. Os executados através de observadores sinópticos humanos ou OS, e os executados por sistemas observadores sinópticos SOS (utilizando câmeras ou não). Os OS registram de forma regular e sistemática as condições atmosféricas vigentes utilizando para isto um treinamento cognitivo especializado. O registro das condições do céu, das nuvens e a sua classificação é feito de acordo com um padrão estabelecido no guia (WMO, 2008) e nos manuais (WMO, 1975) e (WMO, 1987). Em geral os OS estão presentes nas estações meteorológicas padronizadas mantidos pela aeronáutica (aeroportos), marinha (em embarcações e bases navais), e em outras instituições que possuem estações meteorológicas mantidas por órgãos relacionados aos centros de previsão do tempo (INPE, INMET, EMBRAPA, EPAGRI-CIRAM, etc.). Nas próximas subseções serão analisados alguns dos métodos que são empregados para quantificar as nuvens de forma continua sem intervenção humana, existentes na literatura especializada.

2.1 Métodos adotados para a observação a partir da superfície pelos OS

(WMO, 2008) no capítulo 15, descreve um conjunto de critérios para classificar a quantidade, a altura e o tipo das nuvens de acordo com as condições atmosféricas pelos OS. O guia apresenta um conjunto de regulamentações técnicas internacionalmente adotadas, com o objetivo de padronizar a geração e a apresentação de dados e as informações acerca do tempo e do clima. Esta regulamentação técnica é suplementada com uma série de manuais e guias que descrevem detalhadamente, as práticas e procedimentos a serem adotados por seus signatários. Os guias são recomendações, enquanto que os manuais são procedimentos obrigatórios. Por isto as informações acerca da classificação das nuvens são suplementadas por

duas referências (WMO, 1975) e (WMO, 1987). Estes dois suplementos fornecem uma descrição científica e ilustrativa, para auxiliar na identificação dos tipos de nuvens. Os métodos recomendados nestas publicações tem por objetivo: estimar a altura da base, a quantidade (ou cobertura), e classificar os tipos das nuvens observadas. Entretanto este tipo de observação é altamente complexa e subjetiva, por envolver um conhecimento tácito adquirido somente através de treinamento especializado.

A quantidade de nuvens é estimada subjetivamente (WMO, 2008) (15.2.1) em oitavas (valores de 0 a 8), sendo que o valor zero é atribuído para céu claro e o valor oito para céu encoberto. A representação em oitavas indica uma resolução de aprox. 12,5 % ou 1 octa. A indicação em octas é devido à subdivisão dos quadrantes de observação utilizando os pontos cardinais. O valor 9 é reservado para indicar que a observação não pode ser feita por ocorrência de neblina, falta de luminosidade, etc. As observações em geral são executadas em turnos de observação, nos horários sinópticos a cada 3 horas, . Nestas condições pode-se perceber que existe uma grande incerteza para a estimativa da quantidade de nuvens, por que o estabelecimento das oitavas é virtual e cada observador possui um critério pessoal de avaliação. Esta incerteza é acompanhada de uma inconsistência (diferença significativa de avaliação) que existe quando o mesmo céu é observado por observadores distintos (HOYT, 1978). O efeito pacote também é outro tipo de incerteza causada pela super avaliação da quantidade de nuvens quando elas se encontram próximas ao horizonte (HOLLE; MACKAY, 1975).

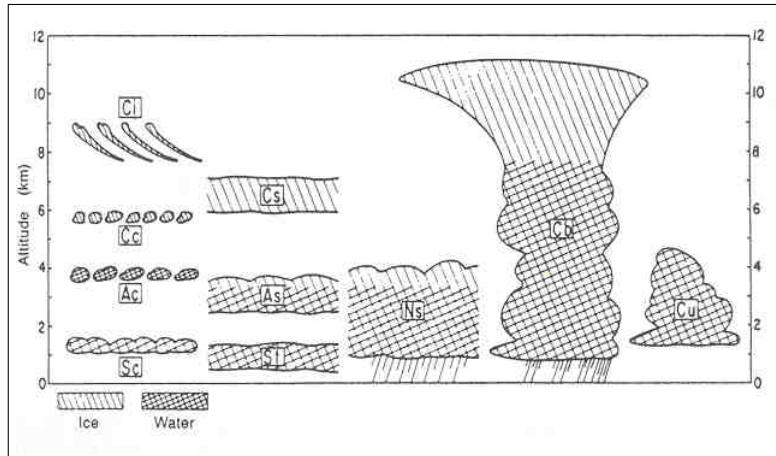
Da mesma forma que a cobertura do céu, a altitude da nuvem também é estimada de forma subjetiva. Em geral recomenda-se o uso do "bom senso" nas estimativas comparando a altitude das nuvens com obstáculos topográficos relativos como montanhas, marcos geográficos de altitude conhecida e também a forma genérica e o tipo da nuvem em questão (WMO, 2008) (15.2.3). Nestas condições pode-se perceber também que existe uma grande incerteza para a estimativa da altitude das nuvens, pois elas podem ocorrer em diferentes camadas a quilômetros de distância do observador. Apesar da subjetividade da estimativa da altura das nuvens, as incertezas podem ser reduzidas através da determinação típica da ocorrência de certos tipos da nuvem em determinadas altitudes.

A identificação das nuvens (ou sua classificação) é uma tarefa nada trivial e envolve uma análise qualitativa que demanda treinamento especializado e uma comparação através de um conjunto de figuras (WMO, 2008) (15.1.4.3). Para o desempenho da tarefa é utilizado basicamente o conhecimento tácito. Um exemplo das possibilidades de classificação das nuvens pode ser visualizado na figura 2.1. Nenhum método de observação automático presente nos SOS conseguiu ainda substituir plenamente o operador sinóptico nesta tarefa com baixo custo.

A avaliação quantitativa dos padrões envolvidos no procedimento dos OS chega a ser grosseira para os padrões tecnológicos atuais. Mas apesar de apresentarem problemas de inconsistências, as observações dos OS ainda são muito utilizadas especialmente pelos aspectos qualitativos da validação, e por serem os únicos dados históricos disponíveis para estudar tendências climáticas a longo prazo (KOTARBA, 2009). Estas e outras inconsistências aliados à fadiga, desgaste nos turnos de observação estimulam em muito as pesquisas para a utilização de métodos alternativos. Especialmente considerando-se o fato de que estas práticas de observação tem sido utilizadas desde o século XIX (WMO, 2008), (HARRISON; CHALMERS; HOGAN, 2008) *apud* (HAMBLYN, 2001). Para contornar os problemas dos OS, algumas alternativas vem sendo procuradas. Dentre elas estão os trabalhos científicos que utilizam sensores de radiação solar, SOS baseados em câmeras, pulsos de luz, RADAR

Altitude	Forma	Gênero	Espécie	Variedade
BAIXAS 0 - 1,5 km	CUMULIFORM	CUMULUS (Cu)	HUMILIS MEDIocris CONGESTUS FRACTUS	
		CUMULONIMBUS (Cb)	CAPILLATUS CALVUS	
	STRATIFORM	STRATUCUMULUS (Sc)	FLOCCUS CASTELLANUS STRATIFORMIS LENTICULARIS	OPACUS TRANSLUCIDUS MODULATUS RADIATUS PERLUCIDUS
		STRATUS (St)	NEBULOSUS FRACTUS	OPACUS TRANSLUCIDUS UNDULATUS
		NIMBOSTRATUS (Ns)		
		ALTOSTRATUS (As)		OPACUS TRANSLUCIDUS UNDULATUS RADIATUS DUPLICATUS
MEDIAS 0 - 6 km	CUMULIFORM	ALTOCUMULUS (Ac)	CASTELLANUS FLOCCUS STRATIFORMIS LENTICULARIS	OPACUS TRANSLUCIDUS UNDULATUS RADIATUS- PERLUCIDUS DUPLICATUS LACUNOSUS
ALTAS 6-12 km	CIRRIFORM	CIRRUS (Ci)	UNCINUS SPISSATUS FLOCCUS CASTELLANUS	RADIATUS DUPLICATUS INTORTUS VERTEBRATUS
	STRATIFORM	CIRROSTRATUS (Cs)	NEBULOSUS FIBRATUS	DUPLICATUS
	CUMULIFORM	CIRROCUMULUS (Cc)	STRATIFORMIS FLOCCUS CASTELLANUS LENTICULARIS	UNDULATUS DUPLICATUS LACUNOSUS

(a) Tabela sumária de acordo com a altitude da base da nuvem, forma, gênero, espécie e variedade, indicada por (WMO, 2008) cap. 15.



(b) Ilustração acerca da distribuição da altitude das nuvens adaptada de (LENOBLE, 1993) fig 16.4.

Figure 2.1: Critérios utilizados para classificação das nuvens.

(baseados na tecnologia de ondas de rádio), ceilômetros ou LIDAR (baseados na tecnologia a LASER), SODAR (baseados na tecnologia de ondas de áudio), balões (sondas meteorológicas), etc. Todos estes SOS possuem custo de aquisição elevado e equipe operacional especializada.

No presente capítulo serão analisados somente os métodos que utilizam os sensores de radiação solar. Dentre os principais métodos serão citados o sensor de brilho solar, o piranômetro para as componentes da radiação global, direta, difusa ou combinação destes. Em seguida serão descritos os poucos trabalhos existentes na literatura dos métodos que utilizam câmeras.

2.2 Sensores de medida de Radiação Solar

As incertezas e inconsistências inerentes ao registro e análise dos OS, motivaram a busca de outros métodos de avaliação das nuvens, não somente na área de meteorologia, mas também na climatologia. Os climatologistas se preocupam muito com a variabilidade e as tendências climáticas a longo prazo, especialmente por causa do impacto das atividades antropogênicas. As nuvens desempenham um papel importante neste estudo. Entretanto o uso das informações obtidas pelos OS nem sempre são apropriadas, por que as estimativas são discretas e não estão em um formato apropriado para identificar tendências ou comparações entre outras variáveis atmosféricas (PALLÉ; BUTLER, 2002).

Existem muitos dados históricos disponíveis que foram, cuidadosa e sistematicamente obtidos a partir dos sensores de radiação solar. Os dados destes sensores são extremamente afetados pela presença das nuvens, pois estas se interpõem entre a radiação solar no topo da atmosfera e a superfície modulando a radiação solar. Estes sensores também podem fornecer uma boa indicação histórica acerca da variabilidades das nuvens ao longo do tempo. A obtenção da cobertura de nuvens a partir de sensores de superfície assume como premissa as seguintes condições. Que o período de tempo que o sensor é sombreado pelas nuvens é proporcional à quantidade de nuvens existente no céu. Por isto a metodologia utiliza os valores de radiação obtidos em janelas de amostragem em um determinado período de tempo (janelas móveis) (DUCHON; O'MALLEY, 1999), (PALLÉ; BUTLER, 2002), (HARRISSON et al., 1993), (ORSINI et al., 2002), etc. Durante este período, considera-se que a distribuição espacial das nuvens é uniforme durante a observação; mas isto nem sempre ocorre.

Os resultados das estimativas dos sensores de radiação solar, são sempre comparados qualitativamente e quantitativamente aos OS (DUCHON; O'MALLEY, 1999), (PALLÉ; BUTLER, 2002), (HARRISSON et al., 1993), (ORSINI et al., 2002), etc. Os dados quantitativos fornecidos por esta metodologia mostraram-se mais apropriados para os estudos de tendências climáticas do que os dados dos OS. Vale a pena salientar que as medidas são representativas apenas de uma proporção da cobertura de nuvens presente no céu, correspondente a 5 graus de ângulo cônico de abertura do hemisfério celeste relativo ao Sol. Nestas condições os dados podem ser considerados uma amostragem de todo o hemisfério celeste ao longo de um determinado período de tempo. Entretanto, a variabilidade da amostragem da população (representada pelo hemisfério celeste), fica restrita somente à faixa de percurso da trajetória solar. Em geral os dados são utilizados neste tipo de análise são avaliados com uma resolução temporal e espacial convenientemente escolhidas pelo autor. Como exemplo pode-se citar as observações dos OS mencionadas no trabalho de (DUCHON; O'MALLEY, 1999). Elas foram executadas a cada hora, limitadas somente nos quadrantes SE e SW. Neste caso foram consideradas somente as nuvens que ocorriam próximas do sol! Entretanto, o piranômetro utilizado era sensível em um ângulo de 180 graus de observação. Em outro experimento executado por (PALLÉ; BUTLER, 2002) os dados dos piranômetros foram comparados às observações efetuadas em uma ou duas vezes por dia às 09:00 am e/ou 09:00

pm no horário GMT.

2.2.1 Utilizando o sensor de Brilho Solar para a estimativa das nuvens

O sensor de brilho solar (ou CAMPBELL-STOKES) foi inventado em 1853. O sensor mede o brilho solar através de uma espécie de lente esférica que queima um papel graduado a partir do qual se determina somente a quantidade de horas de brilho solar através de um fator de brilho solar ou F_s (PALLÉ; BUTLER, 2002). O F_s é definido como a quantidade de horas que o brilho solar está presente e pode ser obtido através da determinação do número de horas de queima do papel (N_{ss}) dividido pela quantidade de horas possíveis de sol a partir de uma elevação solar acima de 3 graus para o local da estação (N_s) conforme pode ser observado na equação 2.1.

$$F_s = \frac{N_{ss}}{N_s} \quad (2.1)$$

Pela forma com que F_s é calculada, pode-se concluir que F_s representa uma resolução temporal diária dos dados. O CAMPBELL-STOKES é muito suscetível à umidade do papel (IQBAL, 1983). Quando úmido o papel começa a ser queimado com aproximadamente 280 W/m^2 e quando seco a aproximadamente 70 W/m^2 ((IQBAL, 1983) pag. 369. Isto causa um problema de limiar de medida que depende das condições atmosféricas vigentes aumentando em muito as incertezas com diferenças de até 20% entre equipamentos. Por muitos anos o limiar inferior de 210 W/m^2 foi adotado como padrão. O limiar exato de queima do sensor também não é muito bem definido (HARRISON; CHALMERS; HOGAN, 2008) *apud* (BIDER, 1958). O CAMPBELL-STOKES não é sensível à radiação difusa, nem à presença de nuvens do tipo cirrus. Estas nuvens quando presentes, não chegam a interromper a queima do papel (PALLÉ; BUTLER, 2002). Neste caso existe uma divergência nos dados, quando o operador indica uma cobertura do céu por ocorrência de cirrus e o CAMPBELL-STOKES não registra.

Os novos sensores de brilho solar a semicondutores não são suscetíveis à presença da umidade como o Campbell-Stokes. De acordo com a (WMO, 2008), o sensor a fotodiodo começa a funcionar somente a partir de 120 W/m^2 de valores de radiação direta. Quando se utilizam os pirâmetros para a determinação do brilho solar o limiar de radiação equivalente estabelecido é de 120 W/m^2 . O sensor de brilho solar é ilustrado na figura 2.2.

O Sensor de Brilho Solar indica binariamente a presença da luz solar ou não, este sensor possui restrições como indicador de cobertura de nuvens. Ele indica as horas de brilho solar disponíveis durante o dia, e para indicar a cobertura de nuvens é utilizado um outro fator chamado fator de nuvens (F_c), obtido a partir de F_s conforme pode ser observado na equação 2.2.

$$F_c = 1 - F_s \quad (2.2)$$

(PALLÉ; BUTLER, 2002) em um trabalho que buscava encontrar as tendências das variações inter-anuais nos dados de cobertura de nuvens, em um longo período de tempo (1881 - 2000) fez um estudo comparativo entre os dados do sensor de brilho solar e as

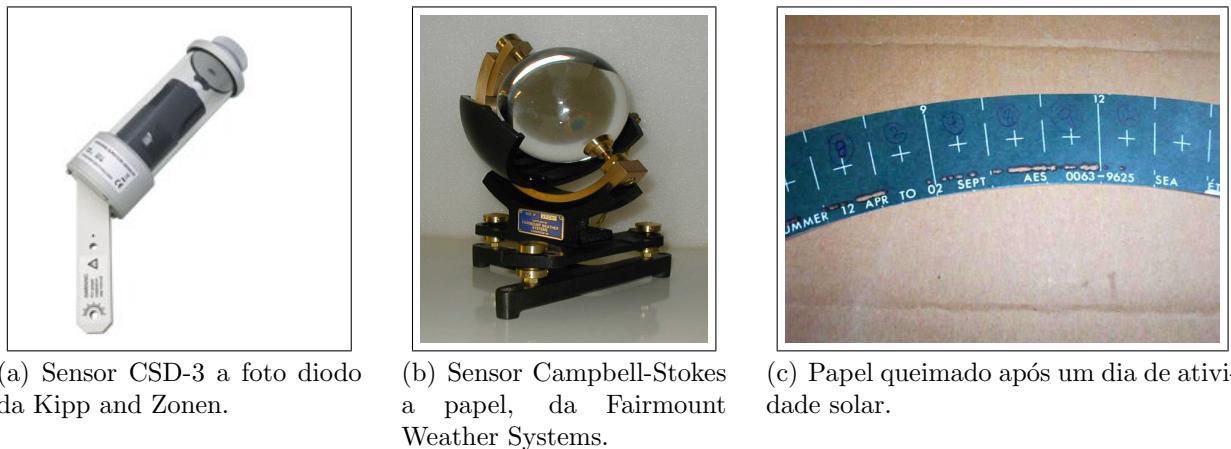


Figure 2.2: Exemplos típicos de sensor de brilho solar.

observações dos OS em três estações de observação distintas na Irlanda. O trabalho de pesquisa indicou uma determinada consistência nas estimativas das nuvens dos OS, mas com certo erro sistemático (bias) que inviabilizou qualquer conclusão mais significativa acerca do estudo nas tendências da cobertura das nuvens a longo prazo. Entretanto estas tendências eram menos acentuadas quando utilizadas no estudo da freqüência de ocorrência de céu claro. O estudo concluiu que os sensores de brilho solar forneciam indicadores quantitativos mais confiáveis para o estudo de variabilidade da cobertura das nuvens a longo prazo do que os fornecidos pelos OS. O estudo ressaltou entretanto que as nuvens do tipo Cirrus não são expressivas o suficiente para causarem alterações no registro da radiação solar do sensor de brilho solar.

Existem dois pontos importantes a serem considerados na análise e validação dos dados; a resolução temporal (RT), e a resolução espacial (RE) dos dados. A resolução temporal dos dados obtidos por (PALLÉ; BUTLER, 2002) é diária nas seguintes circunstâncias: os dados de radiação solar são considerados somente com o sol acima de um determinado ângulo de elevação (três graus); não são considerados os fatores sazonais e geográficos envolvidos nesta avaliação. Estes dados são validados e comparados aos dados de resolução temporal instantânea gerados pelos OS às 09:00 GMT. Esta comparação apresenta uma discrepância na resolução temporal que acredita-se inserir um grande "bias" na incerteza e análise dos resultados.

Da mesma forma uma discrepância espacial também está presente por que a avaliação na cobertura de todo o hemisfério celeste fica limitado somente ao ângulo da auréola solar. O ângulo sólido correspondente a um hemisfério é de 2π sr. O sol ocupa um ângulo sólido de aproximadamente $6,8 \cdot 10^{-5}$ sr (LENOBLE, 1993), ou 5 graus (REINHARD et al., 2006) pag. 398. A representatividade do sol em relação ao hemisfério celeste é de aproximadamente $(6,8 \cdot 10^{-5})/(2\pi)$ ou 0,00034 % do hemisfério celeste.

Por estar tanto tempo em operação, existe uma grande quantidade de dados históricos do Campbell-Stokes acumulados em muitas estações, algumas desde o século XIX. Muitas estações ainda mantém o equipamento operacional. Convém destacar que os dados destes sensores foram obtidos de forma cuidadosa, sistemática e recentemente vêm sendo resgatados e digitalizados para uso científico, inclusive nas pesquisas relacionadas à atividade solar e incidência de raios cósmicos (PALLÉ; BUTLER, 2001).



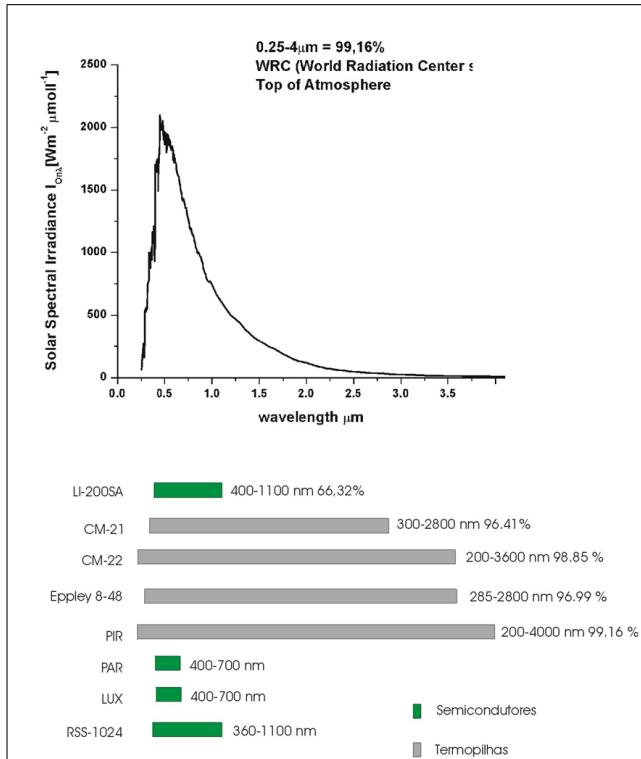
Figure 2.3: Exemplos de tipos de piranômetros que medem diferentes componentes da radiação solar.

2.2.2 Utilizando a radiação global (ondas curtas descendente) para a estimativa das nuvens

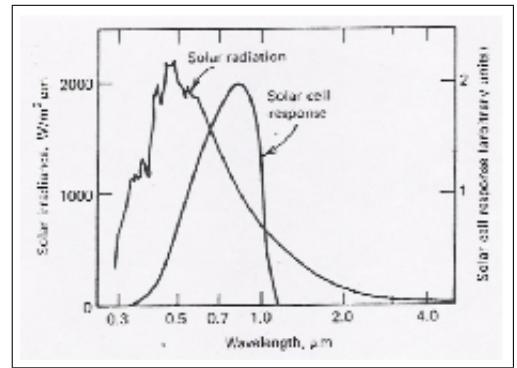
Piranômetros são sensores utilizados nas estações meteorológicas, e convertem a radiação solar em sinais elétricos. O sinal elétrico possui resposta linear em relação à radiação solar incidente, e é modulado pela presença ou ausência das nuvens. Um exemplo típico de uma estação de monitoração utilizando piranômetros, além de diversos tipos de sensores pode ser observado na figura 2.3.

Estes sensores tem dois tipos básicos de construção: a termopilha e a semicondutor. As diferenças básicas entre os dois tipos de sensores é a resposta espectral. Os sensores a semicondutor embora mais baratos não possuem uma resposta espectral linear e completa do espectro da radiação solar conforme pode ser observado na figura 2.4. Por isto para atividades que envolvem precisão com menores incertezas de medida, são utilizados os sensores à base de termopilhas. Dados de piranômetros a semicondutor não serão considerados na presente tese.

(DUCHON; O'MALLEY, 1999) estabeleceram um método para avaliação não super-



(a) Exemplos de resposta espectral de alguns piranômetros a semicondutores e a termopilha em relação ao espectro solar padrão.

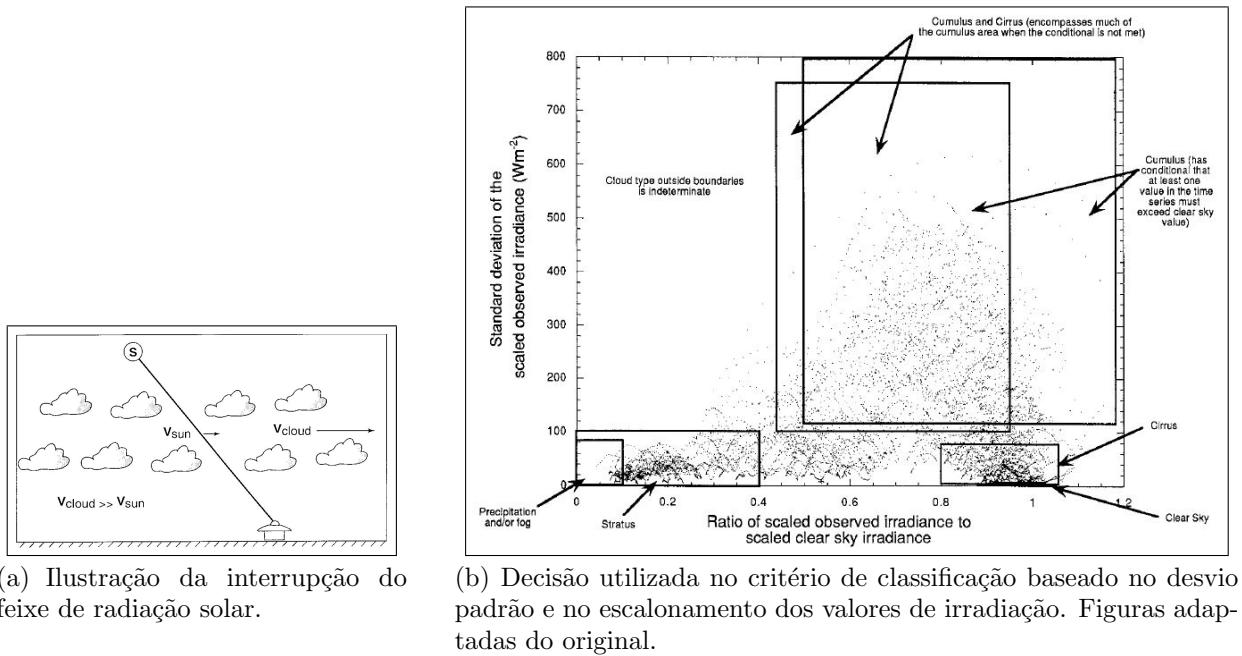


(b) Detalhe da não linearidade na resposta espectral do sensor a semicondutor. Figuras adaptadas de (DUFFIE; A., 2006), figura 2.3.3 pag. 54.

Figure 2.4: Detalhes da resposta espectral produzida pelo sol e a sensibilidade dos sensores típicos de radiação solar.

visionada (automática) de custo reduzido do tipo de nuvens a partir da superfície, através de séries temporais de dados. Os autores utilizaram o cálculo de um modelo atmosférico de céu claro e os dados de um piranômetro, para através das propriedades estatísticas (média móveis e desvio padrão escalados) e manipulações matemáticas, determinar de maneira subjetiva sete categorias entre tipos de nuvem e condições presentes no céu. O trabalho aproveita-se do fato, de que as nuvens naturalmente se interpõem entre o sensor e a radiação solar direta. Como as séries temporais dos dados não são estacionárias elas são proporcionalmente escaladas para gerar séries estacionadas. A partir desta características são monitorados os desvios padrão do sinal em relação ao céu claro em janelas móveis de 21 minutos, e relacionados com os seguintes tipos de nuvens cirrus, cumulus, cirrus e cumulus, stratus, precipitação e neblina. Os dados dos piranômetros são comparadas às observações dos OS. As observações dos tipos de nuvens foram efetuadas de hora em hora, mas somente nas direções SW e SE dos quadrantes de observação. Os dados dos piranômetros foram obtidos através de médias móveis de 21 minutos em relação a um ângulo sólido correspondente ao círculo solar. O método considera a distribuição de nuvens uniforme das nuvens no céu ao longo do tempo, (DUCHON; O'MALLEY, 1999) (5.a). Uma figura ilustrando as condições de obtenção dos dados e o método de classificação sugerido por (DUCHON; O'MALLEY, 1999), pode ser visualizado na figura 2.5.

Os resultados desta técnica foram comparados com os dados levantados pelos OS e encontraram um índice de concordância de 45%. As diferenças foram justificadas pelas



(a) Ilustração da interrupção do feixe de radiação solar.

(b) Decisão utilizada no critério de classificação baseado no desvio padrão e no escalonamento dos valores de irradiação. Figuras adaptadas do original.

Figure 2.5: Metodologia básica para estimativa do tipo de nuvem proposta por (DUCHON; O'MALLEY, 1999) utilizando piranômetro.

seguintes razões.

- O método utilizando o piranômetro é mais influenciado pelas nuvens interpostas entre o sol e o sensor, enquanto que o OS avalia todo o céu.
- A presença de aerossóis induz os piranômetros a superestimar a presença de cirrus e cirrus mais cumulus.

Por estas razões, o método do piranômetro subestima a presença de cumulus, especialmente nas condições de céu com nuvens esparsas, quando comparados com os OS, (DUCHON; O'MALLEY, 1999) (5.a). A publicação conclui que o método proposto se mostra muito útil quando os OS não estão disponíveis. Os desvios entre os valores modelados e os medidos ficaram ao redor de $\pm 30\%$. Entretanto estas diferenças dobram $\pm 75\%$ na presença de aerossóis, que são classificados como cirrus. Aerossóis e cirrus tem desvios similares, entretanto uma pequena correção é necessária no modelo de céu claro para os dias claros com turbidêz acentuada. É interessante notar o seguinte.

- Que dentre as categorias propostas pelos autores esta o tipo indeterminável; ou seja o critério estabelecido claramente não consegue classificar parte dos casos existentes.
- Das seis categorias propostas, quatro (duas a duas) não são mutuamente exclusivas e apresentam ambiguidade de classificação, conforme pode ser observado na figura 2.5 (b).

Estas observações indicam que o critério proposto não é suficiente para classificar satisfatoriamente todos os casos propostos. Mas o trabalho chama a atenção para duas características

importantes que podem ser obtidas a partir dos dados dos piranômetros para a determinação das nuvens. A variabilidade temporal e as diferenças em relação ao modelo de céu claro. O método proposto apesar de limitado, mostra-se útil quando nenhuma outra opção encontra-se disponível.

A RT dos dados obtidos por (DUCHON; O'MALLEY, 1999) a partir dos piranômetros é de 21 minutos. Nesta publicação, os dados dos piranômetros são novamente validados e comparados aos dados de resolução temporal instantânea gerados pelos OS a cada hora. Esta comparação apresenta também uma discrepância na resolução temporal nas duas observações. Esta discrepância pode inserir um grande "bias" na incerteza na análise dos resultados. (DUCHON; O'MALLEY, 1999) apresenta uma concordância de 45% entre as duas observações. O autor justifica estas diferenças pelas discrepâncias nas RT e RE. A RT de 21 minutos da janela de observação dos dados é mais refinada no método de (DUCHON; O'MALLEY, 1999) do que a utilizada pelos sensores de brilho solar. Mas a RE é a mesma que a utilizada pelo sensor de brilho solar, e a mesma discrepância na RE está presente. Vale a pena salientar que o método considera a cobertura de nuvens uniforme durante o intervalo de observação, e que a velocidade de deslocamento das nuvens é muito maior que a velocidade de deslocamento solar. Mas isto nem sempre ocorre, nestes casos o erro sistemático da metodologia se acentua.

2.2.3 Utilizando a radiação combinada global e difusa (ondas curtas) para a estimativa das nuvens

(KARSTEN; CZEPLAK, 1980) utilizaram 10 anos de dados das radiações global e difusa comparadas aos dados gerados pelos OS para obter um índice de cobertura de nuvens para o observatório Meteorológico de Hamburgo na Alemanha. (KARSTEN; CZEPLAK, 1980) compararam as médias horárias das radiações global e difusa e com observações horárias dos OS. A partir destes dados, foi obtida uma equação paramétrica (KARSTEN; CZEPLAK, 1980) (equação 4 da publicação), conforme transcrita na equação 2.3. Mas, segundo os autores, os resultados obtidos são altamente dependentes da posição geográfica do local de observação. Para outros locais, um novo levantamento precisaria ser executado. Os dados indicaram que o levantamento é altamente dependente (ou apresenta grande variabilidade) do tipo e da ocorrência das nuvens. O autor indica também que a turbidéz atmosférica não foi considerada no levantamento. A pesquisa desenvolvida por (KARSTEN; CZEPLAK, 1980) é altamente especializada para o local e não pode ser generalizada para outras regiões com clima distinto. O autor também trabalha com RT horária comparada as observações instantâneas dos OS.

$$\frac{D_h}{G_h} = 0,3 + 0,7 \left(\frac{N}{8} \right)^2 \quad (2.3)$$

onde

- D_h : é a média horária da radiação difusa
- G_h : é a média horária da radiação global
- N: é o índice de cobertura de nuvens em oitavas

(DUCHON; O'MALLEY, 1999) na seção 2.2.2 descrevem e utilizam um modelo físico de céu claro para ser utilizado como referência. Outros autores utilizam outros índices para estabelecer uma referência de céu claro, para ser utilizado nos modelos. (HARRISON; CHALMERS; HOGAN, 2008) utilizam médias de 5 minutos de resolução temporal dos dados de radiação global, difusa, a radiação estimada no topo da atmosfera para determinar os seguintes índices: fração difusa (Df) conforme indicado na equação 2.4, e opacidade (Op) e índice de claridade (Kt) da atmosfera conforme indicado nas equações 2.5, 2.6 e 2.7. Além destes índices (HARRISON; CHALMERS; HOGAN, 2008) utilizam também, os valores das médias e os desvios padrão obtidos dos piranômetros global e difusa além das observações dos OS. 13 amostras dos dados de radiação centralizadas no horário de observação do OS foram obtidas, para efetuar a análise em um período de tempo total de 65 minutos. Nestas condições, os autores podem determinar a quantidade e o tipo das nuvens em duas categorias: cumuliformes e estratiformes; de acordo com altitude média e baixa respectivamente. Na categoria cumuliformes foram incluídas nuvens do tipo cumulus e altocumulus. Na categoria estratiformes foram incluídas as nuvens do tipo stratus e altostratus. A metodologia desenvolvida, não permitiu que pudessem ser identificadas nuvens do tipo stratus e stratocumulus de maneira significativa, mas permitiu identificação de céu claro, encoberto e as categorias cumuliforme e estratiforme.

$$Df = \frac{D}{G} \quad (2.4)$$

$$Op = 1 - Kt \quad (2.5)$$

$$Kt = \frac{G}{E} \quad (2.6)$$

$$E = \frac{S_0 \cos(\theta)}{d^2} \quad (2.7)$$

onde:

- Df : é a fração difusa instantânea
- D : é a radiação difusa instantânea
- G : é a radiação global instantânea
- Kt : é o índice de claridade instantâneo
- E : é a radiação solar no Topo da Atmosfera
- S_0 : é a radiação solar média total
- θ : é o ângulo zenital
- d : é a distância entre o Sol e a Terra em unidades astronômicas [AU]

Maiores detalhes acerca das equações e dos cálculos utilizados nestes parâmetros pode ser consultado nos Anexos B e C.

Apesar de utilizar médias dos sensores obtidas a cada 5 minutos o método utilizado por (HARRISON; CHALMERS; HOGAN, 2008) utiliza uma RT análise de 65 minutos ao redor do intervalo de observação dos OS. A RE continua limitada nos mesmos parâmetros que os métodos anteriores. (HARRISON; CHALMERS; HOGAN, 2008) consegue uma melhor representação dos tipos de nuvens presentes no céu mas segundo o próprio autor não é um método completamente objetivo e também apresenta ambiguidades e incertezas. O autor recomenda uma investigação mais detalhada utilizando outros sensores que podem apresentar uma indicação mais precisa da altitude das nuvens como por exemplo os pirgeômetros de onda longa.

(LONG; ACKERMAN, 1996) utiliza os dados dos sensores de radiação global e difusa para determinar as condições de céu claro a partir de um outro conceito importante utilizado na literatura especializada chamado de *cloud forcing* (CF). Este termo é utilizado para retratar a diferença existente na radiação global entre o céu claro e a radiação medida na superfície no instante de observação (LONG; ACKERMAN, 1996). É um modelo empírico baseado nas condições de observação local de céu claro. Os valores em geral são negativos, por que indicam a atenuação da energia que chega à superfície devido à presença das nuvens. Neste paradigma de análise não são utilizadas as classificações das nuvens recomendados pela (WMO, 2008), devido à dificuldade envolvida no processo de identificação automático. O CF é utilizado em muitos trabalhos de pesquisa para determinar a influência das nuvens através do balanço de energia que chega ao solo. Para a obtenção do valor da radiação de céu claro são utilizados modelos empíricos ou físicos de transferência radiativa (WALISER; COLLINS; ANDERSON, 1996). Os modelos empíricos levam em consideração as condições meteorológicas locais, enquanto que os modelos físicos requerem em geral dados acerca da quantidade de aerossóis, temperatura e umidade, bem como o albedo da superfície (LONG; ACKERMAN, 1996). Mas nem sempre todos estes dados estão disponíveis para o cálculo dos modelos, por isto são utilizados os modelos empíricos. Para a determinação do céu claro são selecionados dados de irradiação solar global de dias especialmente claros. Estes dias são confirmados através de imagens de satélite, observações na superfície, etc. Em seguida os dados de irradiação global são normalizados utilizando-se a distância média entre o Sol e a Terra. Depois, são construídos gráficos para as condições de céu claro em função do ângulo zenital solar. A partir do gráfico é calculada uma função de ajuste polinomial a partir do método dos mínimos quadrados. Um exemplo do cálculo do cloud forcing para a estação do LABSOLAR em Florianópolis, utilizando os dados de Abril de 2004 pode ser observado na figura 2.6, para um critério de céu claro descrito na equação 2.8. Maiores detalhes de implementação poderão ser obtidos junto a (WALISER; COLLINS; ANDERSON, 1996). O autor em suas conclusões sugere que o método apresentado é bem representativo como valor de análise para o balanço de energia que chega à superfície. CF é um bom indicativo comparativo da metodologia apresentada na presente tese, mas para isto é necessário um experimento de campo simultâneo para estabelecer uma correlação ou método de validação. Fica como sugestão para trabalhos futuros comparativos da metodologia proposta.

$$G - Dir. * \cos\theta \leq 50W/m^2 \quad (2.8)$$

onde:

- G: é a radiação Global

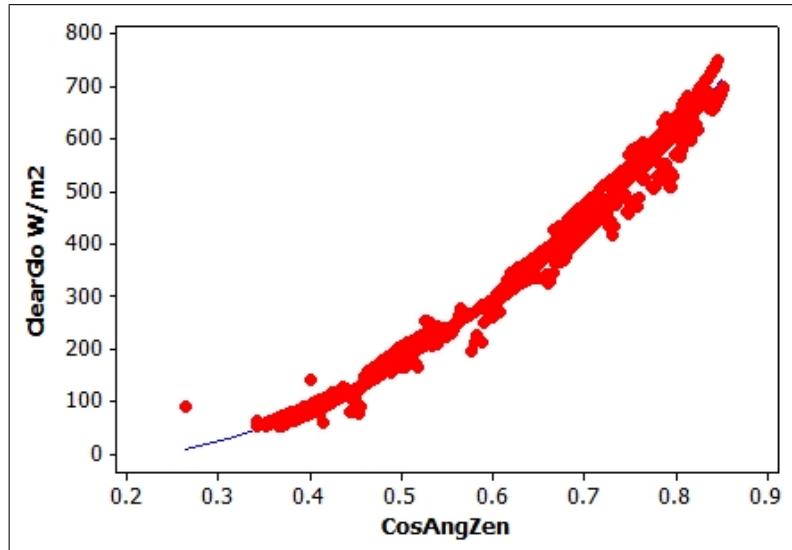


Figure 2.6: Exemplo ilustrativo de modelamento do Cloud Forcing para os dados obtidos a partir da estação do LABSOLAR da UFSC de Florianópolis para o mês de Abril de 2004.

- Dir: é a radiação direta
- θ : é o ângulo zenital solar

A RT utilizada para estabelecer CF é de 10 minutos. Os dados são validados com o WSI (Whole Sky Imager) que possui RT de aquisição de dados instantânea. A discrepância na RT é a menor dos métodos vistos até o presente momento. RE do CF possui uma resolução espacial mais refinada, por que utiliza dois sensores. O sensor de radiação difusa mede a componente espacial da radiação incidente na superfície complementar ao ângulo sólido representado pelo Sol. Neste caso a RE do método é de 2π sr. A RE do WSI é de um ângulo θ de 160 graus. o ângulo sólido Ω de um cone é independente de seu raio é dado pela seguinte fórmula 2.9:

$$\Omega = 2\pi(1 - \cos(\theta/2)) = 3.3054\text{sr} \quad (2.9)$$

O cálculo da representatividade do ângulo de abertura do WSI em relação ao hemisfério esta ilustrado na equação 2.10:

$$RE = \left(\frac{3.3054}{2\pi}\right)100 = 52,61\%. \quad (2.10)$$

2.2.4 Problemas de representatividade do domínio existente na avaliação dos piranômetros

Os sensores de radiação ou piranômetros apresentam um problema de representatividade do domínio que esta sendo analisado. Este problema é comum a todos os métodos (DUCHON; O'MALLEY, 1999), (HARRISON; CHALMERS; HOGAN, 2008), (KARSTEN; CZEPLAK, 1980). Os modelos que utilizam os piranômetros estão baseados na avaliação dos fluxos de radiação baseados no tipo e na quantidade de nuvens (KARSTEN; CZEPLAK,

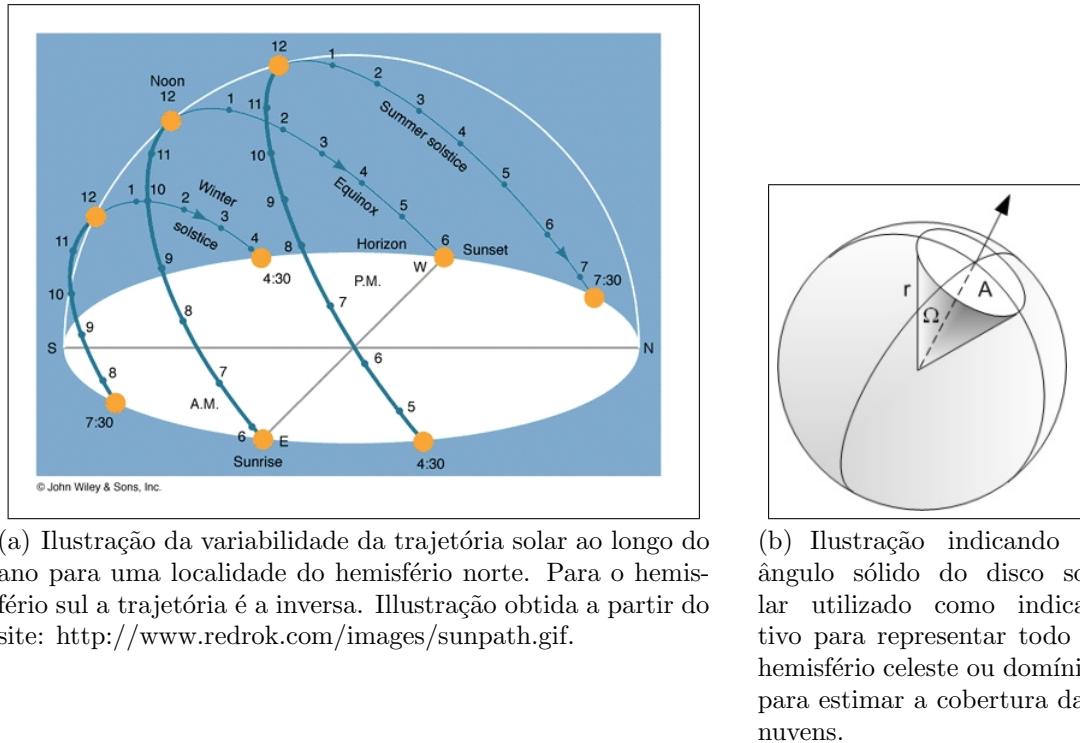


Figure 2.7: Problemas relacionados à distribuição espacial e representatividade do céu.

1980). Em geral estes modelos consideram as diferenças existentes entre a parametrização do céu claro e das condições atuais das nuvens, como os modelos desenvolvidos por (DUCHON; O'MALLEY, 1999), (HARRISON; CHALMERS; HOGAN, 2008),

Mas os métodos baseados nos piranômetros possuem uma deficiência na amostragem do domínio de observação (MANTELLI et al., 2010), causados pela trajetória sistemática do sol conforme ilustrada na figura 2.7 (a) e pela distribuição espacial das nuvens Figure 2.7 (b). Na figura 2.7 (a), a ocorrência das nuvens no hemisfério norte será completamente descartada na avaliação da cobertura do céu. Este problema é o inverso para localidades presentes no hemisfério sul. Tal fato também foi notado por (DUCHON; O'MALLEY, 1999), e durante o seu experimento de validação os OS ignoraram as observações das nuvens na metade sul do céu. Com isto este método não se torna muito representativo da avaliação das nuvens presentes no céu. A figura Figura 2.7 (b) também indica um outro aspecto do domínio de amostragem. De acordo com (LENOBLE, 1993) o disco solar possui um ângulo de abertura de 5,5 graus ou um ângulo sólido de $6.8 \cdot 10^{-5}$ sr, representando apenas 0.000354 % do hemisfério celeste. Este fato indica que os métodos baseados nos piranômetros são mais sensíveis às nuvens que ocorrem na linha de visada do sol, indicando que a amostragem do hemisfério celeste não ser muito representativo do domínio de observação (MONTGOMERY, 2005). A utilização de médias de intervalos de dados reduzem os efeitos das diferenças causadas por estes problemas mas aumentam os erros aleatórios e as incertezas das medidas em até 65% (MANTELLI et al., 2010a)

2.2.5 Utilizando o balanço de radiação de ondas curtas e a radiação extraterrestre de ondas longas, para a estimativa das nuvens

(ORSINI et al., 2002) ressalta que os fluxos das componentes da radiação solar e infra-vermelha obtidas na superfície são fortemente influenciadas pelos gases atmosféricos, aerossóis e principalmente pelas nuvens. Em um experimento realizado pelos autores em uma base Antártica, estas componentes foram obtidas a partir de um albedômetro e de um pirradiômetro, tentando determinar a relação destas componentes da radiação atmosférica com quantidade e o tipo de nuvens.

(ORSINI et al., 2002) utilizam o mesmo método proposto por (DUCHON; O'MALLEY, 1999) para através de um par de valores (uma média escalada do valor de ondas curtas e do desvio padrão) obtidos em um intervalo medida de 21 minutos determinar pelo menos três tipos de nuvens.

(MARTY; PHILIPONA, 2000) descreve um método que determina a cobertura de nuvens, a partir de um índice de céu claro e de parâmetros atmosféricos locais (temperatura e umidade) e de dados de um pirgeômetro (onda longa descendente) com RT de 50 minutos. Mas estes parâmetros são calculados utilizando um modelo físico computacional determinado pelo programa MODTRAN desenvolvido pela Força Aérea Americana. O uso do MODTRAN esta fora de escopo do presente trabalho, devido à dificuldades de licenciamento do produto.

2.3 SOS utilizando Câmeras

Além da estimativa das nuvens a partir de piranômetros, pirgeômetros, ou combinação destes existem as técnicas de observação de superfície que utilizam câmeras. A partir das câmeras obtém-se imagens digitais e a partir destas imagens tenta-se obter uma estimativa da cobertura de nuvens, procurando utilizá-las para substituir as observações feitas pelos OS. A estimativa das nuvens a partir dos piranômetros é obtida a partir de uma média de dados obtidas em um período de tempo. Mas esta estimativa é limitada na resolução temporal do período de amostragem e pela distribuição espacial das nuvens (LONG, 1997). A resolução temporal esta relacionada ao deslocamento das nuvens durante o período de amostragem. A resolução espacial está limitada ao ângulo sólido representado pelo feixe da radiação direta do raio solar. As câmeras efetuam uma amostragem instantânea das condições do céu simultaneamente às imagens de satélite, eliminando discrepâncias por diferença de resolução temporal durante a aquisição das imagens.

Nesta seção serão descritos os métodos de identificação das nuvens existentes na literatura especializada. Em todos estes métodos, os resultados são comparados de forma quantitativa às observações efetuadas pelos OS. Qualitativamente os métodos apresentados pelos piranômetros utilizando estatística, apresentam algum resultado qualitativo com relação ao tipo de nuvem existente no céu.

Os SOS são equipamentos recentes, existentes em poucos locais e possuem custos elevados. Os SOS utilizam câmeras, sistemas refletores, dispositivos de sombreamento para evitar a exposição direta da luz solar além de programas de controle e classificação. Fotos ilustrativas de SOS podem ser observado na figura 2.8.



(a) WSI (Whole sky Imager) instalado no Observatório Meteorológico de Lindenberg, em Lindenberg Alemanha.

(b) TSI (Total Sky Imager) instalado no Instituto Nacional de Pesquisas Espaciais - Observatório Espacial do Sul em São Martinho da Serra RS.

Figure 2.8: Exemplos típicos de SOS em operação.

A utilização do SOS reduz a incerteza causada por diversos aspectos subjetivos de análise, do custo de manutenção e o desgaste dos OS em turnos de observação. A observação das nuvens a partir da superfície envolve aspectos quantitativos (quantidade das nuvens) e qualitativos (tipo e altura das nuvens).

A incerteza na avaliação da quantidade de nuvens no céu feita pelos OS possui uma resolução de 12,5% ou um octa segundo (HARRISON; CHALMERS; HOGAN, 2008) na página 20. Os SOS que utilizam câmeras baseiam-se na quantidade de pixels e possuem uma resolução muito menor. Por exemplo, uma imagem com resolução de 352 x 288 contém 101376 pixels. Considerando que somente 50% destes são úteis para a análise, a resolução em porcentagem é aproximadamente $101376 \times 0,5 = 0.002\%$ da cobertura. Comparando as resoluções dos dois sistemas pode-se concluir que a estimativa da quantidade de nuvens pelos SOS é muito melhor.

Uma comparação entre os dois métodos indica que os operadores sinópticos possuem uma análise qualitativa muito superior na classificação dos padrões de observação presentes no céu. Enquanto que os sistemas automáticos apresentam uma análise quantitativa com incerteza muito menor e uma análise qualitativa limitada. Isto ocorre por que os sistemas automáticos ou imageadores utilizam métodos de análise de imagens triviais, simplificados. As inconsistências são perceptíveis até mesmo a partir de uma simples inspeção visual. Outro grande problema é que estes métodos de análise buscam somente resultados binários (céu e nuvem) e desconsideram outros fenômenos atmosféricos presentes na imagem que induzem ao erro o processo de classificação dos pixels.

(SOUZA-ECHER et al., 2006) em um experimento realizado na estação Antártica Comandante Ferraz, utilizou uma câmera colorida para fotografar o céu e comparou os dados com os de observação de um OS. Neste experimento utilizou-se uma imagem plana com um ângulo de visualização de 62 graus, na posição zenital, evitando sempre a exposição

direta da luz solar. As imagens foram analisadas no espaço de cor HSL (Hue, Saturation and Lightness em coordenadas cilíndricas). Somente o canal de saturação foi utilizado na análise que caracterizou três padrões distintos: céu, nuvens e um terceiro para casos indeterminados obtido por exclusão dos outros dois padrões. A função de discriminação utilizada para classificar estes padrões, foi baseada em um limiar de três desvio-padrão a partir de médias obtidas em imagens de referência. Uma imagem de referência para céu claro e outra para céu encoberto. Esta abordagem permite uma análise limitada a somente três padrões estabelecidos. A eliminação do sol do domínio de observação, restringe a análise somente a padrões de baixa luminosidade.

(LONG et al., 2006) utilizou um critério distinto e dois equipamentos diferentes o TSI (Total Sky Imager) e o WSI (Whole Sky Imager), ambos podem ser visualizados na figura 2.8. TSI utiliza uma imagem refletida a partir de um espelho e o WSI obtém a imagem diretamente a partir do céu; ambos com um ângulo de observação de 160 graus. Um descrição detalhada dos equipamentos pode ser obtida a partir de (LONG et al., 2006) e (LONG; SLATER; TOOMAN, 2001). Embora as câmeras utilizadas nos equipamentos obtenham imagens com 24 bits/pixel no espaço RGB, a classificação dos padrões está restrita a um limiar da razão de 0,6 R/G dos canais de cores (LONG et al., 2006). Este critério restringe o domínio de análise somente dos padrões de cor que variam de preto a magenta, ignorando qualquer característica adicional que possa ser obtida a partir do canal verde. Estas informações adicionais poderiam ser utilizadas para classificar ou analisar outros padrões atmosféricos que pudessem ser obtidos a partir da mesma imagem. (LONG et al., 2006) também indica a dificuldade na identificação de pequenas diferenças devido a outros componentes atmosféricos. Píxeis azuis de alto brilho correspondente ao céu azul na região de transição entre o espalhamento molecular e a turbidéz atmosférica são dificilmente identificáveis com a metodologia utilizada. Esta dificuldade é devida à uma abordagem inapropriada conforme indicado nos trabalhos anteriormente desenvolvidos pelo autor em (MANTELLI, 2001) e (MANTELLI; WANGENHEIN; PEREIRA, 2005), ou pela simplificação do domínio de amostragem reduzindo sua dimensão original com a eliminação do canal verde (MANTELLI et al., 2010). Muitos outros autores utilizam a mesma técnica de classificação de céu/nuvens mas com aparelhos experimentais distintos para a obtenção das imagens, tais como (KREUTER et al., 2009), (FEISTER et al., 2000), (SCHADE et al., 2009) e muitos outros mais. Este critério de classificação é o mais utilizado nos grandes centros de pesquisa localizados na Europa e nos Estados Unidos.

(MANTELLI, 2001) em sua tese de mestrado utilizou redes neurais para identificação de céu e nuvem, pré-processado por filtros de difusão anisotrópica, normalizados a partir do modelo de céu claro sugerido por (IQBAL, 1983). O filtro de difusão acentuou as bordas discretizou as nuvens, conforme pode ser observado na figura 2.9. Apesar de conseguir bons resultados as redes neurais não conseguiram classificar altos valores de luminosidade nas condições de céu claro, para píxeis próximos do sol. A presença de reflexões ao redor do disco solar induz à erros de classificação (SABBURG; WONG, 1999). Estes píxeis são resultantes do espalhamento da luz solar nos aerossóis e vapor de água (turbidéz) variável de acordo com as condições atmosféricas. Estes píxeis são classificados errôneamente como nuvens. Durante a pesquisa, (MANTELLI, 2001) observou que em geral o estabelecimento de um resultado binário (céu e nuvem) para ser utilizado em qualquer técnica de classificação dos padrões existentes no céu é limitado em desempenho, por que existem mais do que dois padrões visíveis no domínio de análise. Por exemplo a cor laranja/vermelha devido ao espalhamento Mie, a absorção (nuvens cinza) e o espalhamento (nuvens brancas) da luz solar devido à densidade óptica das nuvens, difusão solar através das nuvens etc. Como consequência o sistema classi-

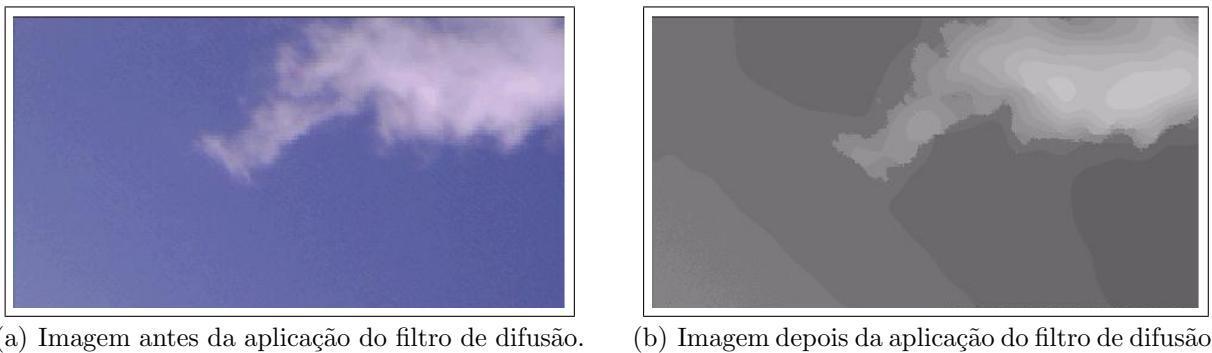


Figure 2.9: Efeito do filtro de difusão discretizando e acentuando as bordas. Figura extraída de (MANTELLI, 2001) capítulo 4.

ficará aleatoriamente estes padrões em uma das categorias estabelecidas resultando em erros de classificação. (MANTELLI, 2001), (MANTELLI; WANGENHEIN; PEREIRA, 2005) e (MANTELLI; WANGENHEIN; PEREIRA, 2009) sugere que um domínio mais apropriado seja estabelecido para a classificação de todos os padrões presentes no céu para reduzir os erros de classificação de abordagem binária.

(CAZORLA; OLMO; ALADOS-ARBOLEDAS, 2008) utilizou uma rede neural estabelecendo preliminarmente 18 parâmetros de entrada obtidos a partir de uma matriz 9x9 para cada um dos píxeis da imagem. A quantidade destes parâmetros (ou a configuração da entrada da rede neural) foi otimizada utilizando-se algoritmos genéticos para identificar quais parâmetros eram mais significativos para obter os resultados de céu, nuvem e nuvens finas. Nenhuma indicação na publicação foi apresentada, justificando a escolha destas categorias. Os parâmetros obtidos a partir desta otimização foram o valor médio do canal azul, o valor médio do canal vermelho e a variância do canal vermelho da matriz da vizinhança do pixel. Mas apesar de apresentarem bons resultados, a pesquisa apresenta a mesma limitação do trabalho de (LONG; SLATER; TOOMAN, 2001). A análise ignora o canal verde, restringindo o domínio de cores a variações de preto ao magenta, conforme ilustrado na figura 2.10 . (LILLESAND; KIEFER, 1994) capítulo 1 descreve que as nuvens espalham quantidades iguais de vermelho, verde e azul e as define como "espalhamento não seletivo". Mas este método possui também os mesmos erros de abordagem na definição das categorias de classificação dos resultados apontados por (MANTELLI, 2001). Algoritmos genéticos são muito utilizados na otimização dos sistemas (RUSSELL; NORVIG, 2003). Mas uma análise crítica qualitativa apresentada pelo algoritmo genético, indica que a otimização induziu a um domínio de análise limitado. As redes neurais são muito utilizadas na classificação, mas o domínio e o resultado de análise devem ser corretamente estabelecidos (JAIN; DUIN; MAO, 2000). O conhecimento envolvido no processo de análise precisa ser melhor avaliado, com uma definição apropriada das variáveis independentes, intermediárias e dependentes envolvidas na pesquisa (CRESSWELL, 1994). Um detalhe importante acerca da rede neural é que nenhum modelo matemático formal ou físico entre o domínio dos dados de entrada e a saída categorizada de dados é formalmente estabelecido. O mapeamento entre o domínio e os resultados são mimetizados e corrigidos por repetição, até que os erros de classificação atinjam níveis satisfatórios. Se a mimetização for errôneamente estabelecida, o problema se mantém.

(MANTELLI et al., 2010) utilizou um método Bayesiano (NIST, 2009) visando a utilização de agentes inteligentes (RUSSELL; NORVIG, 2003) para classificar padrões atmos-

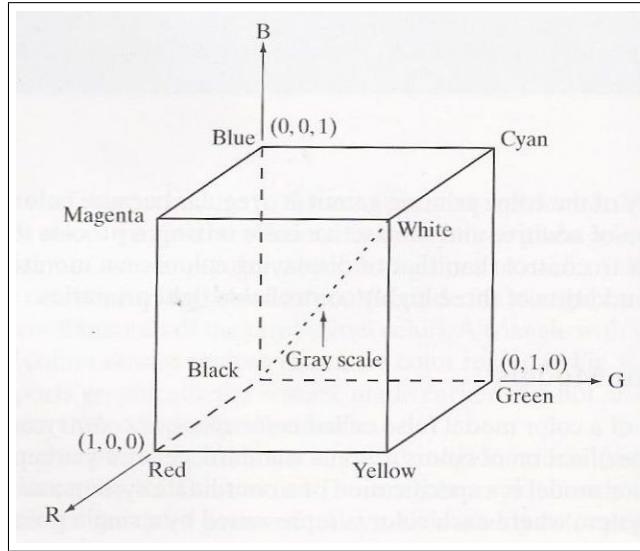


Figure 2.10: Esquema de cores do espaço RGB. A utilização somente dos canais vermelho e azul restringe o domínio das variáveis independentes a cores que variam somente de preto a magenta. Figura extraída de (GONZALEZ; WOODS, 2002) capítulo 6.

féricos. (MANTELLI et al., 2010) observou que os padrões de céu e nuvem ocorrem em um *locus* característico no espaço de cores, paralelos à diagonal principal do cubo de cores RGB. A localização típica destes padrões foi determinada empiricamente através da otimização dos fatores envolvidos (MONTGOMERY, 2005) e da análise exploratória de dados a partir da Distância e Projeção Euclidiana na diagonal principal do cubo de cores RGB. O método mostrou-se invariante às oscilações de luminosidade dos padrões presentes no céu, e também em relação à elevação solar ao longo do dia, por que se baseia em atributos de cor. O trabalho de (MANTELLI et al., 2010) é um dos resultados preliminares apresentados nesta pesquisa, no capítulo 4.

2.4 Resumo comparativo das características dos principais métodos utilizando piranômetros e imagens utilizados na estimativa de nuvens encontrados na literatura especializada

Um resumo dos principais métodos baseados em piranômetros presentes na literatura especializada encontra-se na tabela 2.1.

Um resumo dos principais métodos baseados em imagens presentes na literatura especializada encontra-se na tabela 2.2.

Table 2.1: Resumo comparativo dos principais métodos utilizando piranômetros encontrados na literatura especializada. RT é a resolução temporal em minutos, RE é a resolução espacial em porcentagem, OS é o método utilizado pelos operadores sinópticos, PB é o método utilizando o sensor de Brilho Solar utilizado por (PALLÉ; BUTLER, 2002), DO é o método utilizando piranômetro desenvolvido por (DUCHON; O'MALLEY, 1999), KC é o método utilizando piranômetro desenvolvido por (KARSTEN; CZEPLAK, 1980), HCH é o método utilizando piranômetro desenvolvido por (HARRISON; CHALMERS; HOGAN, 2008)

Método	RT	RE	Limitações
OS	180	12,5%	subjetividade
PB	diária	100%	papel
DO	21	< 1%	domínio não representativo, distribuição espacial
KC	60	< 1%	domínio não representativo, distribuição espacial
HCH	35	< 1%	domínio não representativo, distribuição espacial

Table 2.2: Resumo comparativo dos principais métodos utilizando imagens encontrados na literatura especializada. SE é o método de estimativa de nuvens utilizado por (SOUZA-ECHER et al., 2006), LO é o método utilizando piranômetro desenvolvido por (LONG; SLATER; TOOMAN, 2001), MA1 é o método utilizando piranômetro desenvolvido por (MANTELLI, 2001), CA é o método utilizando piranômetro desenvolvido por (CAZORLA; OLMO; ALADOS-ARBOLEDAS, 2008), MA2 é o método utilizando piranômetro desenvolvido por (MANTELLI et al., 2010)

Método	Discriminação	Classificação	Limitações
SE	3σ	binária	1-D, baixa luminosidade, domínio
LO	linear	binária	2-D, domínio
MA1	NN	binária	1-D, domínio
CA	NN	binária	2-D, domínio
MA2	estatística	binária	binária, domínio

Chapter 3

MATERIAIS E MÉTODOS

Quando o céu é observado a partir da superfície, é difícil ignorar a visão telúrica do azul claro ao profundo, a diversidade e a forma das nuvens brancas e escuras, a composição surpreendente de coloração amarelo-alaranjada durante o nascer e o por do sol, os raios-de-sol através das nuvens, etc. Os observadores sinópticos regularmente analisam e classificam estes padrões baseados em um treinamento cognitivo especializado (WMO, 2008), (WMO, 1975), (WMO, 1987). O presente trabalho visa buscar técnicas computacionais para substituir estas observações por sistemas automáticos, onde é altamente desejável evitar a análise e a avaliação subjetiva destes padrões. Apesar do avanço das técnicas computacionais em nossos dias, os sistemas automáticos atuais ainda utilizam técnicas triviais e obtém resultados binários (céu e nuvem), ignorando os demais padrões presentes nas cenas naturais e as dificuldades envolvidas na sua identificação. Esta abordagem precisa ser refinada e uma nova metodologia investigada para representar de maneira mais apropriada os fenômenos presentes no domínio de observação.

Os métodos de identificação automática levantados na literatura especializada subutilizam os dados fornecidos pelas câmeras digitais por simplificação matemática, conforme destacado na revisão da literatura e resumido na tabela 2.2. A maior parte dos métodos não utilizam todas as dimensões disponíveis no espaço de cores, ou analisam o domínio de forma limitada. Esta limitação leva à perda de informação e à redução do desempenho dos métodos propostos e acentuando os erros de classificação (JAIN; DUIN; MAO, 2000).

No presente capítulo serão descritos os métodos utilizados para a identificação automática baseados na metodologia Bayesiana, na estatística multivariada e no paradigma dos agentes inteligentes. O modelamento Bayesiano tem sido utilizado cada vez mais na representação do conhecimento humano, por causa do rigor matemático envolvendo técnicas estatísticas na representação da análise subjetiva (CHARTER; TENENBAUN; A., 2006). Segundo (CHARTER; TENENBAUN; A., 2006), *"a matemática da probabilidade é uma ferramenta vital na construção das teorias cognitivas, fornecendo um campo muito vasto para a cognição aprendizado e raciocínio"*. O reconhecimento dos padrões propostos na presente tese representam regularidades existentes da natureza (CHARTER; TENENBAUN; A., 2006) em um espaço semântico de alta dimensão (ou hiperespaço) (STEYVERS; GRIFFITHS; DENNIS, 2006) e (DUDA; HART; STORK, 2001). (STEYVERS; GRIFFITHS; DENNIS, 2006) também menciona que *"os seres humanos são muito suscetíveis às regularidades estatísticas do seu ambiente"*. A metodologia Bayesiana foi portanto a escolhida para o presente trabalho, por representar melhor as regularidades existentes nas cenas naturais observadas

pelos seres humanos.

A metodologia Bayesiana compreende dois passos análise supervisionada e classificação (JAIN; DUIN; MAO, 2000) e (NIST, 2009). A análise supervisionada será representada pela identificação manual dos padrões existentes nas imagens, pelas técnicas de análise exploratória de dados (AED) e otimização de experimentos (MONTGOMERY, 2005). Os padrões serão definidos a partir de modelos matemáticos sugeridos por (JOHNSON; WICHERN, 2007). A classificação utilizará os critérios baseados na manifestação dos padrões de acordo com a física atmosférica (IQBAL, 1983), (LENOBLE, 1993) e (LILLESAND; KIEFER, 1994). A discriminação será feita utilizando critérios estatísticos de controle de qualidade, a partir dos graus de liberdade (MONTGOMERY, 2001) levantados no processo e da distância de Mahalanobis (MAHALANOBIS, 1936). Redes neurais também utilizam estes passos, entretanto uma das grandes críticas das redes neurais é a abordagem "caixa preta" e a interpretação sintática na solução do problemas na qual o desenvolvedor em geral, não toma o cuidado de modelar rigorosamente o domínio dos eventos, estando sujeitos a erros de semântica no campo de aplicação (ZHANG, 2007) (JAIN; DUIN; MAO, 2000). A metodologia proposta na presente tese vai além da reprodução ou mapeamento sintático dos padrões existentes nas imagens, buscando uma representação matemática rigorosa do domínio analisado através de métodos estatísticos estabelecidos e consolidados. Uma ilustração geral descrevendo as atividades da metodologia utilizada, pode ser observada na figura 3.1.

Nas próximas subseções serão detalhados os seguintes tópicos. Descrição do aparato experimental de aquisição e manipulação de imagens. Análise exploratória de dados ou EDA (NIST, 2009). Otimização da análise dos dados experimentais (MONTGOMERY, 2005). A definição do domínio de observação ou a "*representação do conhecimento sobre alguma parte do mundo sobre o qual desejamos expressar algum conhecimento*", (RUSSELL; NORVIG, 2003) pag. 245. A definição do ambiente de tarefa onde estão essencialmente os problemas para os quais os agentes são as soluções", (RUSSELL; NORVIG, 2003) pag. 39 (NEWELL; SIMON, 1972) para ser utilizado pelos sistemas automáticos baseados em agentes inteligentes (RUSSELL; NORVIG, 2003). A correspondência entre os padrões observados e os fenômenos físicos atmosféricos descritos na literatura especializada (IQBAL, 1983), (LENOBLE, 1993), (LILLESAND; KIEFER, 1994), (RICHARDS, 1995), etc.

3.1 Aparato experimental

As imagens utilizadas na parte experimental do presente trabalho, foram obtidas a partir de um imageador de nuvens disponível comercialmente. O Total Sky Imager ou TSI-440 da Yankee Environmental. Mas a análise desenvolvida no presente trabalho, pode ser aplicada a qualquer câmera devidamente adaptada a um sistema de sombreamento automático. O TSI gera imagens automaticamente no formato padrão JPEG (Joint Photograph Group) com uma resolução de 352 x 288 pixels e 24-bits no espaço RGB. Uma foto deste imageador já foi ilustrada na figura 2.8-(b). O TSI possui um programa residente que foi configurado para gerar imagens a cada 15 minutos do nascente ao poente. As imagens foram registradas a partir do ano, mês, dia e hora GMT da obtenção, indicados no nome do arquivo. 50 imagens no total foram utilizadas. O programa residente não permite que o TSI gere imagens com o sol a menos de 5 graus de elevação. A imagem refletida representa um ângulo de visualização de 160 graus. As imagens não são obtidas diretamente do HC, mas a partir de

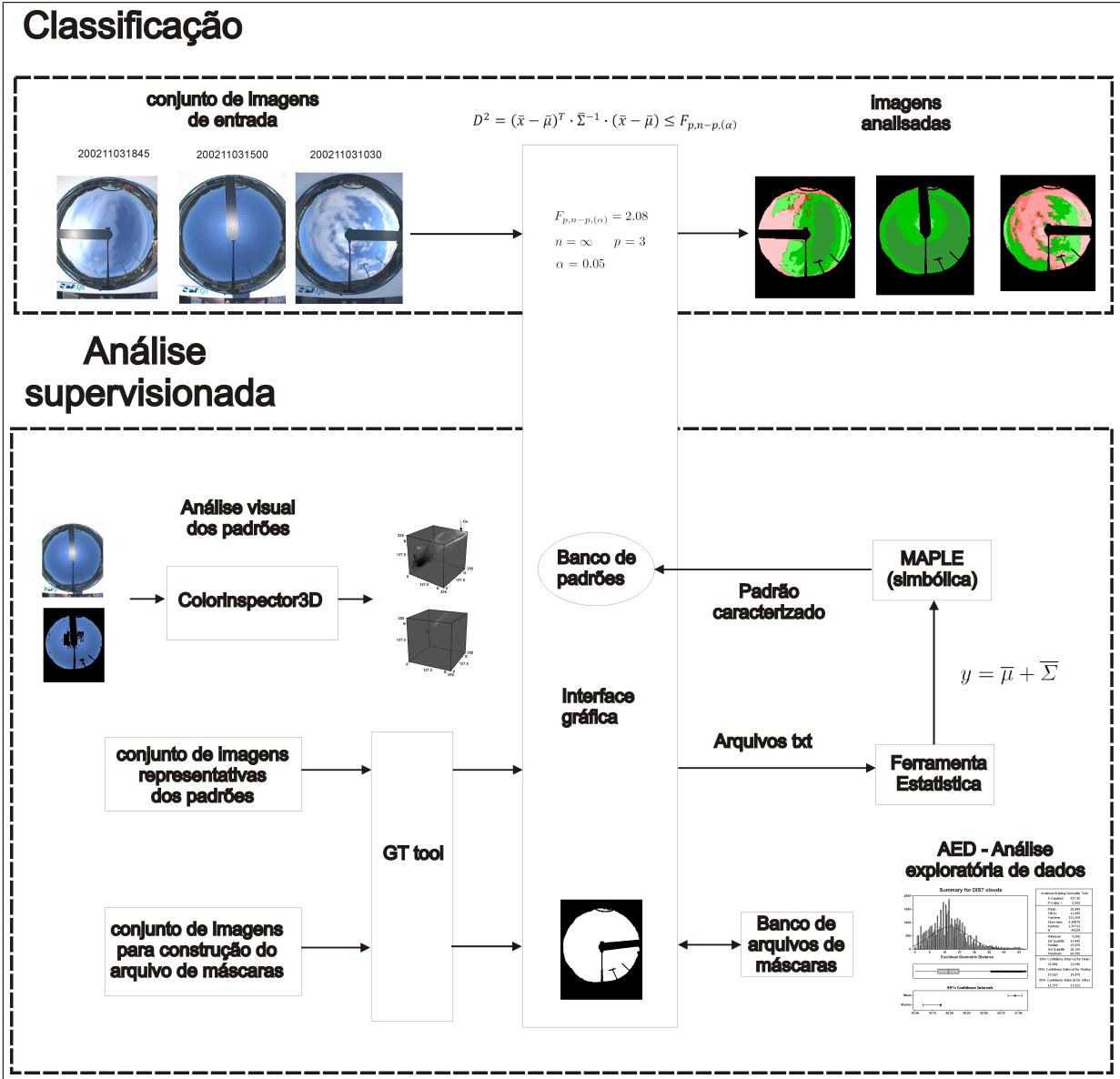


Figure 3.1: Esquema geral ilustrando todas as fases, aplicativos e procedimentos utilizados durante as pesquisas desenvolvidas na presente tese.

um refletor no formato de domo. O refletor possui uma banda de sombreamento aplicada com um adesivo especial e um sistema de rastreamento solar, para evitar danos à câmera de vídeo por exposição direta à luz solar. As imagens foram obtidas nas dependências do LABSOLAR-EMC-UFSC (lat. 27° 32' S, lon. 48 ° 31' W). No local também está instalada uma estação BSRN (Baseline Surface Radiation Network), que monitora os dados de radiação solar continuamente, e forneceu os dados adicionais de radiação solar para avaliar a metodologia proposta.

Para a analisar preliminarmente as imagens nos diversos espaços de cor, utilizou-se uma ferramenta de visualização chamada ColorInspector3D disponibilizada gratuitamente na internet para fins educacionais e científicos. Esta ferramenta permite, a partir do posicionamento do mouse sobre as imagens, mapear o *locus* geométrico dos padrões de interesse das imagens no espaço de cor RGB ou HSL. Um exemplo da visualização e mapeamento

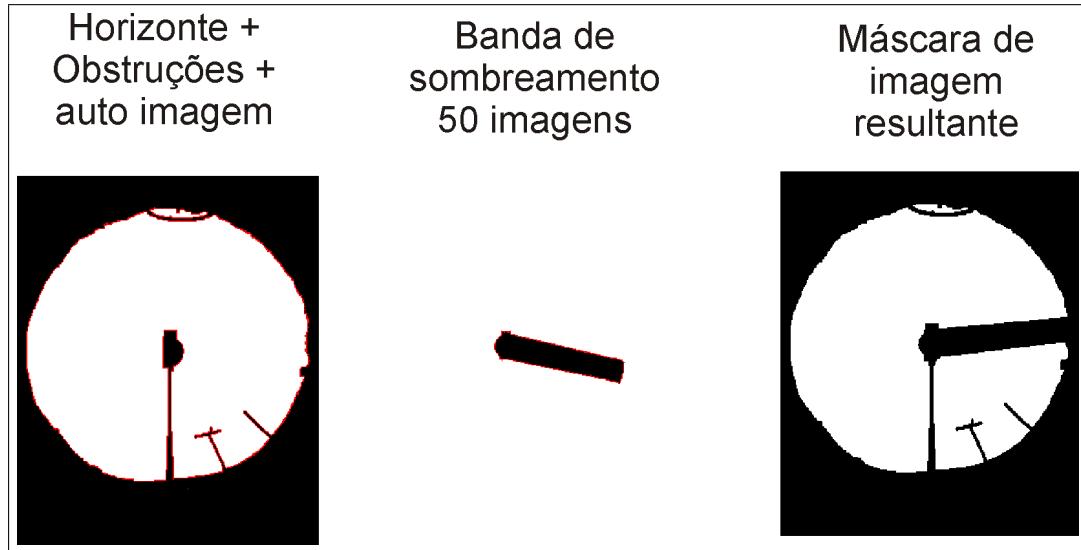


Figure 3.2: Remoção dos fatores que interferem na análise da imagem através de mascaramento (redução dos valores dos píxeis a zero), utilizando a ferramenta GT tool do LAPIX. Construção da máscara das obstruções horizontais, banda de sombreamento móvel e da máscara resultante.

destes padrões pode ser observado na figuras 3.5, 3.6 e 3.9.

Uma outra ferramenta de análise desenvolvida no LAPIX chamada GT-tool (Ground Truth Generator Tool) também foi utilizada para extrair manualmente (píxel a píxel) regiões específicas das imagens que não significavam informações úteis na análise tais como: obstruções do horizonte, postes, edificações, etc. GT-tool também foi utilizado para selecionar nas imagens, os padrões de interesse que seriam caracterizados na AED. Na figura 3.2 esta ilustrada a remoção dos píxeis relativo às obstruções.

Para facilitar a manipulação, extração dos valores dos píxeis para análise estatística, visualização dos resultados, testes e implementação das técnicas utilizadas durante a presente pesquisa, foi desenvolvida uma interface gráfica em linguagem C++ para a plataforma WINDOWS utilizando o VISUAL STUDIO e a biblioteca wxWIDGETS. A figura 3.3 ilustra a interface gráfica implementada.

Para efetuar os cálculos estatísticos (médias, matrizes de co-variancia, etc) utilizadas na AED utilizou-se um pacote estatístico comercialmente disponível.

Para facilitar a manipulação das equações lineares da distância de MAHALANOBIS a partir dos parâmetros dos padrões definidos durante a AED, e também simplificar a implementação dos algoritmos e reduzir o tempo de processamento computacional, foi utilizado o recurso de racionalização simbólica disponível no MAPLE. Esta racionalização permitiu que a resolução das equações lineares proposta na equação 3.9 durante a classificação para cada píxel da imagem, ficasse reduzida a um conjunto simples de somas e subtrações conforme exemplificado na figura 3.4.

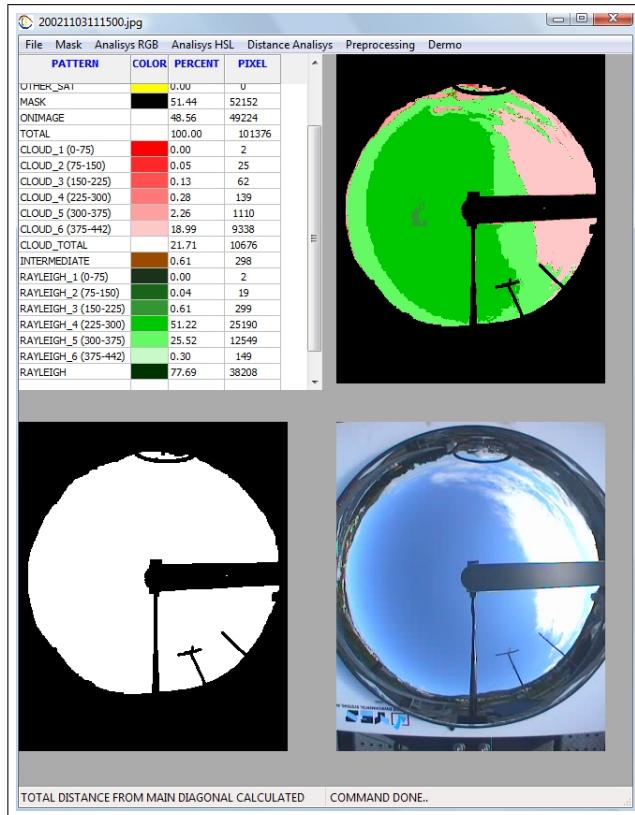


Figure 3.3: Interface gráfica desenvolvida para análise, implementação das técnicas desenvolvidas e visualização dos resultados.

```
D1 = (150.829 - r) * (-0.076688833 - 0.0533026416250343554 * r + 0.0766774818699234030 * g - 0.0243992141479335671 * b);
D2 = (162.128 - g) * (0.124059320 + 0.0766774818699233614 * r - 0.121880714523458569 * g + 0.0456346148407959368 * b);
D3 = (176.862 - b) * (0.063594348 - 0.0243992141479335151 * r + 0.0456346148407958744 * g - 0.0213846621162948822 * b);
D = D1 + D2 + D3;
```

Figure 3.4: Ilustração do fragmento de código gerado a partir da manipulação simbólica do MAPLE para a racionalização das equações lineares do cálculo da distância de MAHALANOBIS (equação 3.9), que reduziu as equações lineares a um conjunto simples de somas e subtrações.

3.2 Análise exploratória dos padrões presentes no Hemisfério Celeste do espaço de cor

Durante a observação e análise sistemática de diversas imagens do HC, pode-se perceber claramente a ocorrência dos padrões de interesse em um *locus* típico no espaço de cor (MANTELLI et al., 2010b). Uma ilustração deste fato pode ser observado na figura 3.5 para as nuvens e na figura 3.6 para o céu. A ferramenta de visualização Color inspector 3D indicou que os píxeis das nuvens são tipicamente cinza e brancos, estão distribuídos linearmente próximos da diagonal principal do cubo do espaço RGB e possuem uma pequena componente da cor azul presente no padrão. Os píxeis do céu azul também indicaram uma ocorrência semelhante embora um pouco mais afastados da diagonal principal. Os mesmos padrões apresentaram uma distribuição não linear nos píxeis do espaço HSL. Baseado nessas observações, optou-se por executar as operações matemáticas no espaço RGB evitando

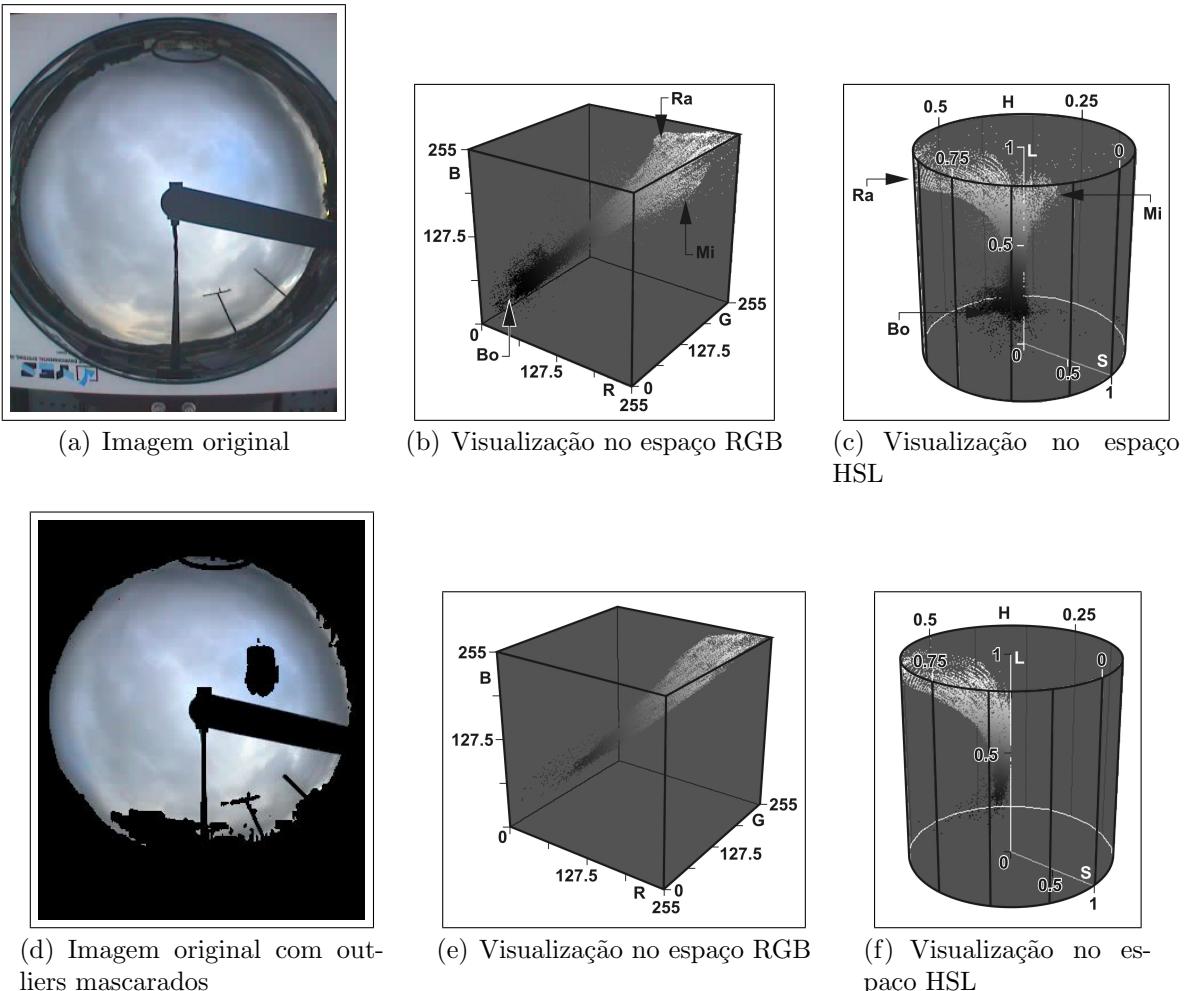


Figure 3.5: Ilustração da ocorrência do *locus* típico do padrão de nuvens observado em uma imagem do TSI, com a sua visualização nos espaços de cores. As indicações "Ra", "Mi" e "Bo", marcam o *locus* típico para os padrões Rayleigh, Mie e a borda + banda de sombreamento respectivamente. As unidades são em intensidades relativa dos píxeis para o espaço de cores RGB, e normalizados para o HSL. Figura obtida de (MANTELLI et al., 2010)

as não linearidades do HSL. Mas o HSL continuou a ser consultado ao longo da presente pesquisa como uma análise complementar, por que a sua representação esta bem próxima à da percepção de cor humana. Outros espaços de cores não foram considerados.

As observações descritas no parágrafo anterior durante a presente pesquisa, levaram ao desenvolvimento de um método de classificação utilizando distância euclidiana geométrica ou EGD e a projeção euclidiana geométrica ou EGP dos padrões de interesse em relação à diagonal principal do cubo RGB (MANTELLI et al., 2010b). Este método apresenta a vantagem de classificar os padrões de céu e nuvem independente das variações de luminosidade ao longo do dia. Para ilustrar o método proposto, será considerado um píxel genérico, conforme ilustrado na figura 3.7.

Como os padrões de céu e nuvem possuíam uma ocorrência paralela em duas distâncias distintas à diagonal do cubo RGB, a EGD foi utilizada para a descrição do *locus* geométrico destes padrões em relação à diagonal. As distâncias entre os píxeis podem ser determinadas

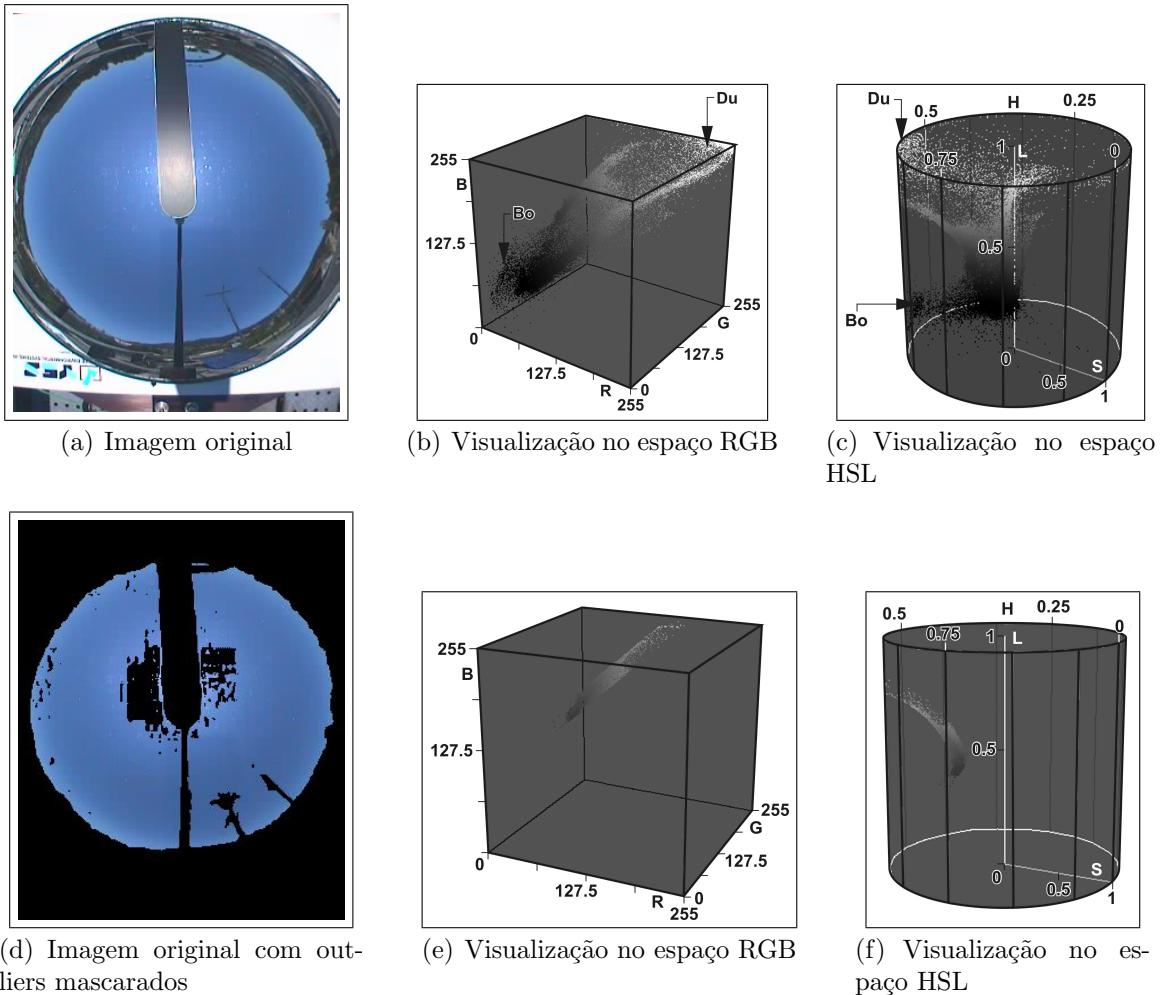


Figure 3.6: Ilustração da ocorrência do *locus* típico do padrão de céu observado na imagem do TSI, com a sua visualização nos espaços de cores. As indicações "Ra", "Mi" e "Bo", marcam o *locus* típico para os padrões Rayleigh, Mie e a borda + banda de sombreamento respectivamente. As unidades são em intensidades relativa dos píxeis para o espaço de cores RGB, e foram normalizados para o HSL. Figura obtida de (MANTELLI et al., 2010)

considerando-os vetores. A distância foi utilizada como parâmetro de discriminação e a projeção como brilho do pixel representando estes padrões, atribuindo aos valores dos píxeis um significado físico que permite correlacioná-los aos sensores do tipo piranômetros. Estas grandezas podem ser obtidas a partir das equações 3.1 e 3.2.

$$PROJ = |\mathbf{D}| \cos(a) \quad (3.1)$$

$$DIST = |\mathbf{D}| \sin(a) \quad (3.2)$$

Através do teorema de Al-Kashi (ou a Lei dos Cossenos) e da fórmula da soma dos vetores, podemos obter as equações 3.3, 3.4 e 3.5.

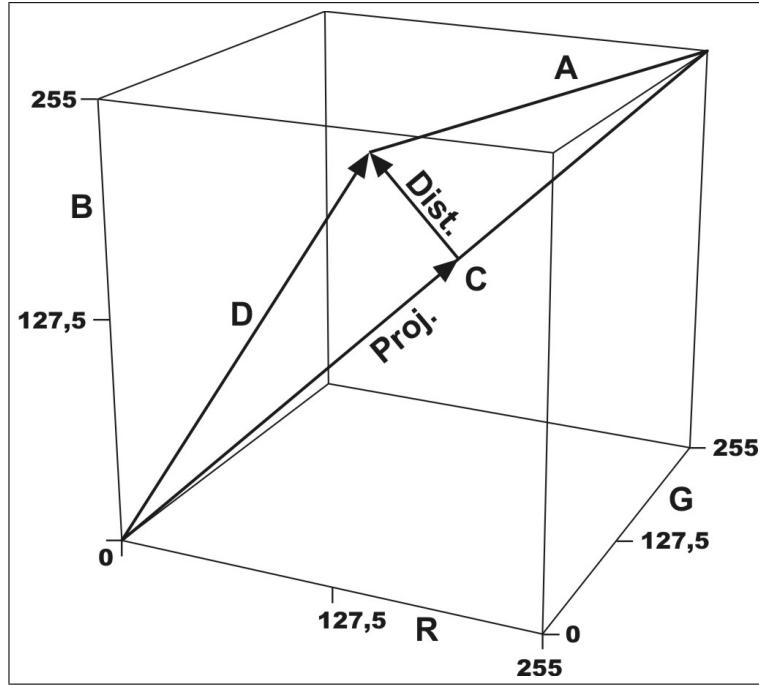


Figure 3.7: Análise utilizada no espaço de cor RGB, para a classificação dos píxeis das imagens do TSI utilizando a EGD e a EGP em relação à diagonal principal. As unidades são em intensidade relativa dos píxeis.

$$|\mathbf{A}|^2 = |\mathbf{D}|^2 + |\mathbf{C}|^2 - 2 |\mathbf{D}| |\mathbf{C}| \cos(a) \quad (3.3)$$

$$|\mathbf{C}| = |\mathbf{D}| + |\mathbf{A}| \quad (3.4)$$

$$|\mathbf{C}| = \sqrt[3]{255^2 + 255^2 + 255^2} = 441,673 \quad (3.5)$$

onde

- a: é o ângulo entre o vetor do píxel e a diagonal principal
- **A**: é o vetor complementar que vai das coordenadas genéricas do pixel de cor (r, g, b) ao vértice da diagonal principal (255, 255, 255).
- **C**: é o vetor correspondente à diagonal principal com as coordenadas (255, 255, 255)
- **D**: é o valor do pixel tendo as coordenadas genéricas (r, g, b).

Substituindo os termos na equação do teorema de Al-Kashi e rearranjando a equação chegamos ao resultado apresentado pela equação 3.6

$$\cos(a) = \frac{|\mathbf{C} - \mathbf{D}|^2 - |\mathbf{D}|^2 - |\mathbf{C}|^2}{-2 |\mathbf{D}| |\mathbf{C}|} \quad (3.6)$$

A equação 3.6 ilustra que somente o valor (r , g , b) do píxel é necessário para calcular a sua distância e a sua projeção sobre a diagonal do cubo RGB. O $\text{sen}(a)$ pode ser obtido facilmente a partir do teorema de Pitágoras ($\text{sen}^2(a) + \cos^2(a) = 1$). Maiores detalhes acerca desta técnica desenvolvida no decorrer do presente trabalho com a análise dos resultados, poderão ser obtidas na publicação (MANTELLI et al., 2010b).

Mas a EGD na maneira em que foi concebida, ainda está limitada a resultados de classificação binários dos padrões analisados. A presença de vários padrões em diferentes *loci* típicos no espaço de cor induziu a pesquisa a um modelamento matemático estatístico multivariado. Desta maneira os padrões puderam ser caracterizados em 3 dimensões através de modelos paramétricos. A vantagem de se utilizar técnicas estatísticas, é que estes métodos permitem utilizar as mesmas ferramentas disponíveis e históricamente estabelecidas. Como por exemplo definição de populações, teste de hipóteses, inferências, índices e intervalos de confiança, etc. A literatura utilizada para o desenvolvimento do trabalho foi baseado em (PEDHAZUR; SCHMELKIN, 1991), (JOHNSON; WICHERN, 2007), (MONTGOMERY, 2005), (MONTGOMERY; RUNGER, 2003), (MONTGOMERY, 2001), (WILKS, 2005) e (NIST, 2009).

Os padrões a serem identificados serão caracterizados a partir de uma análise exploratória, utilizando uma ferramenta estatística devido ao grande volume de dados a ser manipulado. Inicialmente serão selecionadas as imagens que tipicamente representem cada padrão de interesse. Dentro destas imagens serão extraídos somente os píxeis que tipicamente representem estes padrões, conhecidos como "ground truth" ou GT (FERNANDEZ-GARCIA et al., 2008). A extração será executada a partir da ferramenta GT desenvolvida no LAPIX. Nesta ocasião também serão removidos todos os padrões espúrios (i.e. postes, elementos geográficos, borda do equipamento, banda de sombreamento, etc.) por que não representam nenhuma informação e interferem na caracterização do padrão (MONTGOMERY, 2005). Após a caracterização dos padrões será efetuada a remoção de pontos conhecidos como "outliers" através de uma técnica chamada distância inter-quartil, para aproximar a mediana da média e reduzir a variância do padrão. Desta forma estarão caracterizados os padrões de interesse na imagem.

A EDA dos padrões de interesse, indicou que eles possuem uma distribuição que pode ser assumida como normal no espaço de cor, mas com covariâncias distintas em suas dimensões. Isto indica que a distância Euclidiana não deve ser utilizada para a classificação dos padrões. A distância Euclidiana utilizada nestas circunstâncias, incluiria uma grande quantidade de "outliers" na classificação dos padrões (JOHNSON; WICHERN, 2007) pag. 30. Estes "outliers" também causariam uma degradação no desempenho da classificação e um aumento na variância nos dados analisados. Por esta razão, será utilizado no presente trabalho a distância estatística ou de Mahalanobis (MAHALANOBIS, 1936).

A distância de Mahalanobis considera de maneira apropriada os efeitos das diferentes covariâncias caracterizadas no padrão. Os padrões selecionados como referência ou os GT, formam no espaço de cores um agrupamento de pontos que pode ser caracterizado por um modelo linear. Este modelo estatístico que descreve o dado a partir da EDA, pode ser representado a partir de um vetor médio, uma matriz de covariância, e um erro, conforme sugerido por (MONTGOMERY, 2005) ilustrado na equação 3.7.

$$y_p = \mu + \Sigma + \epsilon \quad (3.7)$$

$$y_p = \begin{vmatrix} \bar{r}_p \\ \bar{g}_p \\ \bar{b}_p \end{vmatrix} + \begin{vmatrix} \sigma_{rr} & \sigma_{rg} & \sigma_{rb} \\ \sigma_{gr} & \sigma_{gg} & \sigma_{gb} \\ \sigma_{br} & \sigma_{bg} & \sigma_{bb} \end{vmatrix} + \begin{vmatrix} e_r \\ e_g \\ e_b \end{vmatrix}$$

onde:

- y_p : é o padrão correspondente com p=3 dimensões (RGB)
- μ : é o vetor médio do padrão
- Σ : é a matriz de covariância do padrão
- \mathbf{e} : é a matriz do erro aleatório do padrão
- \bar{r}_p é a média da componente vermelha dos píxeis do padrão.
- \bar{g}_p é a média da componente verde dos píxeis do padrão.
- \bar{b}_p é a média da componente azul dos píxeis do padrão.
- e_r é o erro padrão da amostra devido à componente vermelho
- e_g é o erro padrão da amostra devido à componente verde
- e_b é o erro padrão da amostra devido à componente azul

O erro padrão da amostra representada pelos padrões será calculado pela expressão descrita na equação 3.8 para cada componente de cor.

$$e_r = \frac{s_r}{\sqrt{n}}$$

$$e_g = \frac{s_g}{\sqrt{n}} \quad (3.8)$$

$$e_b = \frac{s_b}{\sqrt{n}}$$

onde:

- s_r é o desvio padrão da amostra devido à componente vermelho
- s_g é o desvio padrão da amostra devido à componente verde
- s_b é o desvio padrão da amostra devido à componente azul
- n : é o número de observações que define o padrão

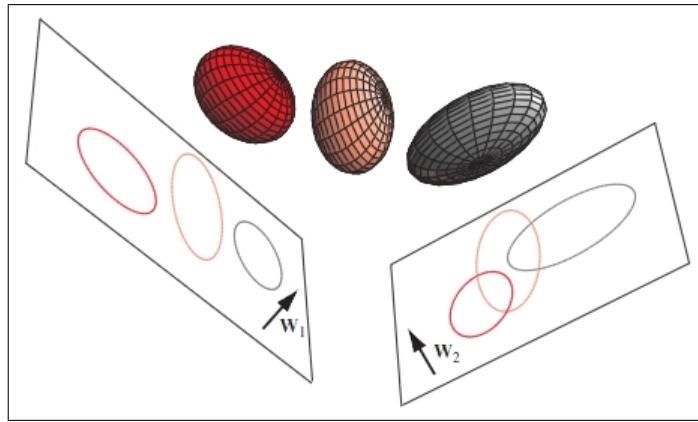


Figure 3.8: Hiperelipse como superfície de classificação de pontos no espaço tridimensional. Adaptada a partir da figura 4.28 (DUDA; HART; STORK, 2001).

A discriminação de cada padrão é feita através de um volume de contorno ao redor do padrão, no formato de uma hiperelipsóide (DUDA; HART; STORK, 2001). Os eixos desta hiperelipsóide são caracterizados a partir dos autovetores unitários da matriz de covariância do padrão. Os autovalores determinam o comprimento dos eixos e o vetor médio determina o centro da hiperelipsóide, conforme pode ser observado na figura 3.8.

A classificação dos píxeis é feita utilizando a distância de Mahalanobis a partir de um limiar obtido em uma tabela estatística encontrada facilmente na literatura, nos mesmos moldes dos testes de hipóteses clássicos. A distância de Mahalanobis possui distribuição F e pode ser calculada de acordo com os graus de liberdade, determinados a partir da formula sugerida por (JOHNSON; WICHERN, 2007) cap. 5 e calculado de acordo com a equação 3.9.

$$D^2 = n(\mathbf{x} - \mu)^T \cdot \Sigma^{-1} \cdot (\mathbf{x} - \mu) \leq \frac{(n - 1)p}{(n - p)} F_{p, n-p, (\alpha)} \quad (3.9)$$

onde:

- D^2 : é a distância de Mahalanobis genérica entre o pixel que esta sendo classificado e o do padrão definido como GT.
- \mathbf{x} (r,g,b) : é o pixel representado pelo vetor no qual se deseja calcular a distância em relação à população.
- μ : é o vetor médio do padrão (ou da população)
- T : é a operação de transposição de matriz
- $^{-1}$: é a operação de inversão de matriz
- n : é o número de píxeis utilizados para determinar o padrão.
- p : são os graus de liberdade correspondente ao número de dimensões do espaço de cor($p = 3$).

- $n-p$: são os graus de liberdade correspondente à população amostrada. Se $(n-p) \geq 120$ a tabela considera os graus de liberdade como infinitos (∞).
- $\alpha = 0,05$: é o nível de confiança estabelecido para a avaliação do teste
- $F_{p,n-p,(\alpha)}$: é o valor do limite de discriminação do padrão. Este valor é obtido a partir do ponto percentual de uma tabela da distribuição F. $F_{3,\infty,(0,1)} = 3,78$; $F_{3,\infty,(0,05)} = 2,61$; $F_{3,\infty,(0,01)} = 2,08$

Outros critérios de classificação como intervalos de confiança simultâneos de bonferroni conforme sugerido por (MONTGOMERY; RUNGER, 2003), (MONTGOMERY, 2001) e (JOHNSON; WICHERN, 2007) podem ser utilizados. No caso multidimensional o intervalo pode ser obtido a partir da equação 3.10 para cada dimensão do espaço de cor, de acordo (JOHNSON; WICHERN, 2007) p. 226.

$$\left(\mu_i \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p,(\alpha=0,01)}} \sqrt{\frac{\sigma_i^2}{n}} \right) \quad (3.10)$$

As imagens das figuras 3.5 e 3.6 foram utilizadas como padrão de modelamento. O efeito da remoção dos píxeis espúrios pode ser visualizada nas imagens 3.5 (d) e 3.6 (d) para os padrões de nuvens e céu respectivamente. Nos painéis 3.5(e) e 3.5(f) e 3.6(e) e 3.6(f), podem ser observados os efeitos da retirada dos padrões espúrios e dos "outliers" das imagens nos espaços RGB e HSL.

Os "outliers" foram retirados a partir da Distância Interquartil IQD. A distância interquartil é a distância entre os quartis superior e inferior dos dados (NIST, 2009). A IQD é obtida a partir da análise estatística e da diferença entre o percentil inferior de 25% dos dados ou Q1 e do percentil superior de 75% dos dados ou Q3. Os pontos localizados fora dos limites inferior (IQI) e superior (IQS) além do ponto médio são eliminados. Um sumário das equações utilizadas para eliminar os "outliers" estão descritos nas equações 3.11, 3.12 e 3.13 a seguir.

$$IQD = Q3 - Q1 \quad (3.11)$$

$$IQI = Q1 - 1.5 \times IQD \quad (3.12)$$

$$IQS = Q3 + 1.5 \times IQD \quad (3.13)$$

onde:

- IQD: Distância interquartil
- IQI: Limite interquartil inferior
- IQS: Limite interquartil superior

Além dos píxeis eliminados como "outliers" através da distância interquartil, outros 3 padrões presentes na imagem foram considerados como espúrios (MONTGOMERY, 2005): a borda, as obstruções horizontais e a banda de sombreamento móvel. A borda não é relevante para a análise da imagem, por que representa uma auto imagem do equipamento. As obstruções horizontais são objetos fixos na imagem como postes, torres, árvores, edificações e montanhas, etc. que estão presentes na imagem e não representam nenhuma informação útil. A banda móvel de sombreamento também foi mascarada. Este mascaramento foi executado à mão, pixel a pixel, antes da análise exploratória de dados na GT "tool". A borda e as obstruções horizontais eram fixos e foram removidos facilmente. Mas a banda de sombreamento é móvel e foi necessária a remoção distinta para cada uma das imagens produzidas ao longo do dia. A saída da GT "tool" é uma máscara negra em um arquivo no formato BMP. As máscaras foram concatenadas e carregadas dinamicamente pelo aplicativo em C++ durante a fase de análise das imagens.

3.3 Ambientes do Agente Inteligente

Um grande avanço tem se observado no modelamento da percepção humana para ser utilizado por agentes inteligentes (RUSSELL; NORVIG, 2003) na capacidade de percepção, análise e atuação em um determinado ambiente. Nestes sistemas o processamento e interpretação dos dados é baseado em um modelo cognitivo de maneira a fornecer a informação e o conhecimento acerca de uma observação. Em outras palavras o agente inteligente efetua um mapeamento dos dados (através de interpretação) em conhecimento. Para interpretar os dados é necessário uma definição apropriada do que deve ser percebido do ambiente de acordo com as suas propriedades, da capacidade dos sensores e da limitação do processamento a ser realizado.

A definição apropriada do ambiente de atuação do agente inteligente é um ponto chave na solução do problema. O ambiente será definido de acordo com a versão padrão proposta por (NEWELL; SIMON, 1972). Segundo estes autores a definição começa com a "*a descrição do espaço do problema no qual a atividades se desenvolvem*" (NEWELL; SIMON, 1972) p. 50. Este **ambiente de pesquisa** tem 4 componentes principais: o estado inicial, o estado final (ou objetivo), os operadores (que permitem as mudanças entre estados) e as regras (que disciplinam ou regulam as mudanças de estado). Mas a natureza subjetiva do projeto e a grande variedade de soluções existentes podem levar a uma boa ou má definição do problema. Visando restringir a grande quantidade de interpretações e soluções, foi também definido por (NEWELL; SIMON, 1972) o **ambiente de tarefa**. O **ambiente de tarefa** antevê todas as possibilidades que podem ser acessadas pelo agente para a sua atuação, mas se restringe somente no conjunto de opções necessárias para o agente resolver um tipo particular de problema. Se o **ambiente de pesquisa** for muito restrito, o agente atinge estados inacessíveis na solução geral. Foi justamente este o problema que aconteceu nas soluções anteriormente utilizadas e atualmente vigentes para a interpretação dos padrões existentes no HC de acordo com (MANTELLI, 2001).

A especificação do agente de acordo com (RUSSELL; NORVIG, 2003) será definida da seguinte maneira. No presente trabalho o **tipo de agente** é um sistema de análise de imagens. A entrada de dados do agente ou **sensor**, ocorre através de um dispositivo CCD presente nas câmeras automáticas que geram os dados brutos. O **domínio dos valores** dos dados esta limitado aos 8 bits de resolução da CCD com valores inteiros no intervalo

[0,255], em um espaço vetorial multidimensional (espaço de cores). Os dados apresentam-se na forma de diferentes níveis dos píxeis compondo uma imagem que forma padrões de cores (informação). Estes padrões formam o **domínio semântico** ou de **domínio de tarefa** a ser interpretado pelo agente. As imagens ocorrem no HC e representam um campo de observação no espectro de luz visível em um ângulo sólido de 180 graus. A **atuação** do agente ocorrerá na interpretação e classificação dos fenômenos físicos atmosféricos observados, e o mapeamento destes padrões. O **objetivo** do agente é categorizar corretamente os possíveis padrões de cores visando a substituição de um OS.

As propriedades do **ambiente de tarefa** foram determinadas da seguinte maneira.

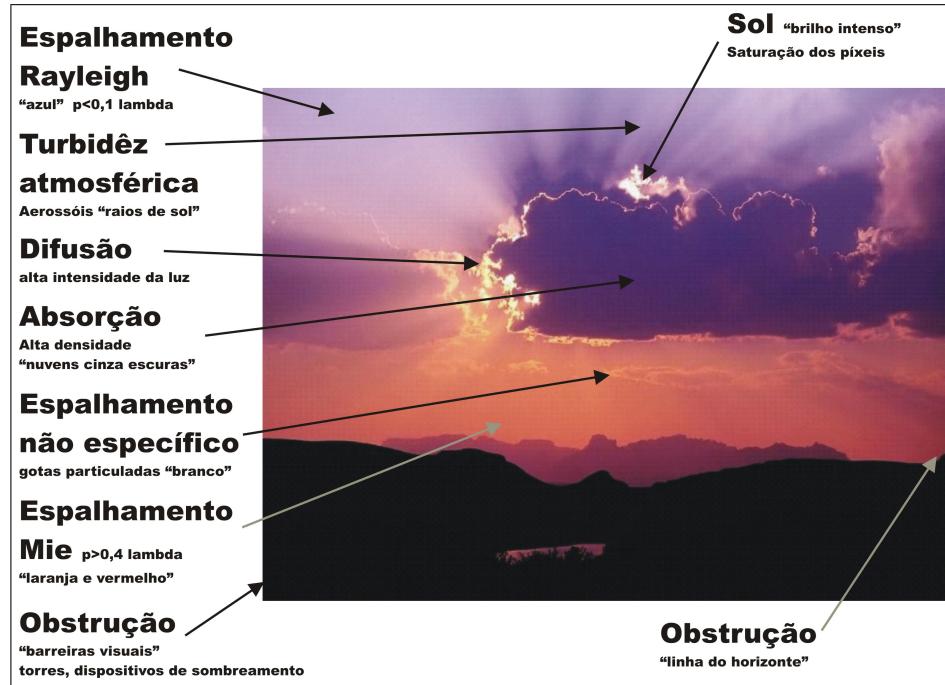
- O HC é um ambiente de tarefa **completamente observável**, pois não há nenhuma necessidade de retenção de informação de outros dispositivos ou sensores relevantes para a sua atuação.
- O HC é um ambiente de tarefa **determinístico**, pois o seu comportamento pode ser determinado por leis físicas. Os fenômenos que são observados são determinísticos, mas a sua classificação é realizada de forma padronizada e previamente definida.
- O HC é um ambiente de tarefa **episódico**, a atuação do agente não depende de um estado anterior. Os dados dos eventos são independentes.
- O HC é um ambiente de tarefa **estático**, por que os dados não se alteram durante o processo de análise. Não há necessidade de análise em tempo real.
- O HC é um ambiente de tarefa **discreto**, por que as imagens (ou fotos digitais) são obtidas periodicamente, com uma resolução temporal entre dois dados é de 15 minutos.

O agente a ser especificado é um agente do tipo **reativo**, que atuará a partir da imagens alimentadas por um sistema automático, baseado em um modelo Bayesiano de **aprendizagem supervisionada**. A modelagem Bayesiana cumpre os requisitos "operadores" e as "regras" sugeridos por (NEWELL; SIMON, 1972). Os operadores compreendem a distância de Mahalanobis dos múltiplos padrões estabelecidos do pixel sendo avaliado. As regras são estabelecidas a partir dos critérios estatísticos (teste de hipóteses ou nível de confiança multivariados a serem utilizados).

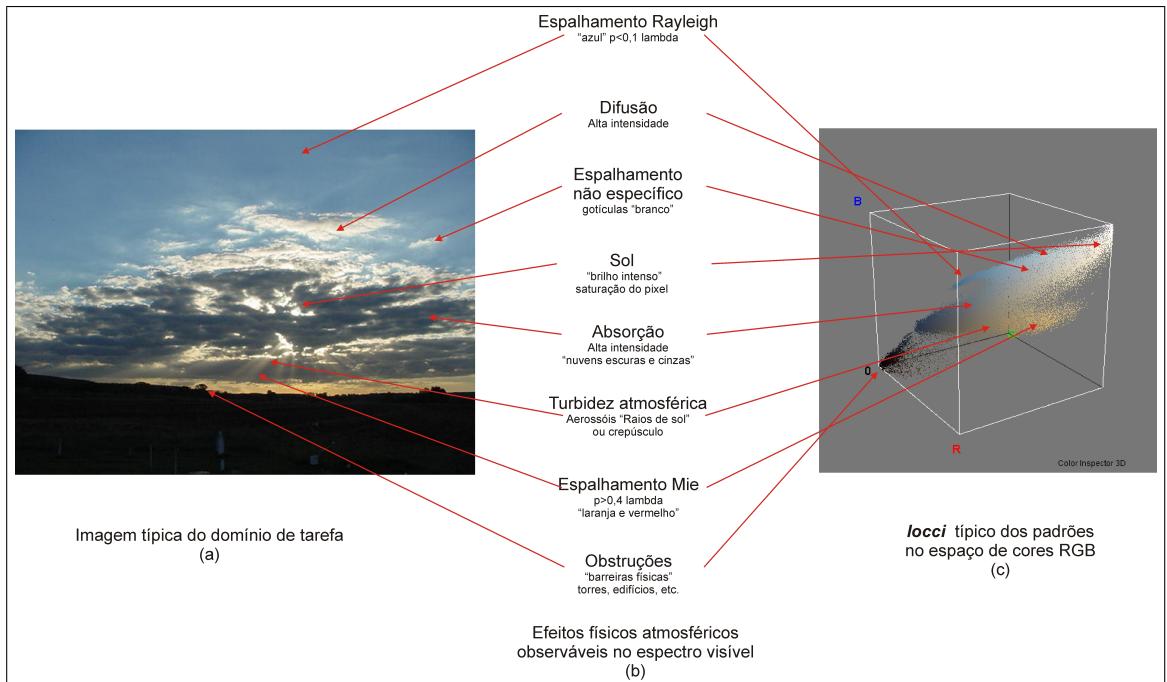
3.4 O Levantamento do Domínio de Tarefa Baseado na Física Óptica-atmosférica dos Fenômenos Observados, e a sua Representatividade

Dois fenômenos físicos principais são predominantes na interação da radiação solar com os componentes atmosféricos: o espalhamento e a absorção. Ambos influenciam diretamente a radiação solar extraterrestre modificando consideravelmente a energia espectral que atravessa a atmosfera (IQBAL, 1983). O espalhamento é a difusão dos fótons da luz solar pelas moléculas e partículas presentes na atmosfera (LENOBLE, 1993). A absorção é a perda de energia solar para os elementos presentes na atmosfera, em geral na forma de calor. A absorção da radiação ultra-violeta pelo ozônio e a diminuição da radiação solar pela presença das nuvens são exemplos deste fenômeno. Para um observador na superfície, os principais efeitos destes fenômenos observáveis a olho nu no HC são os seguintes.

- **Sol.** É a fonte principal de luz de alta intensidade e é caracterizado pela saturação dos valores dos píxeis. A luz do Sol aparece com uma coloração predominantemente branca, mas também apresentando componentes resultantes de amarelo e vermelho. Amarelo por que perde parte da componente azul devido ao espalhamento Rayleigh. Vermelho por que também parte da componente verde por espalhamento durante o pôr-do-sol. No presente trabalho o sol é obstruído por sistemas de sombreamento, para que não haja danos à câmera CCD. Quando o sol aparece na imagem, é por que o equipamento apresenta problemas de alinhamento do sistema de sombreamento automático.
- **Espalhamento Rayleigh, azul.** Ocorre quando a luz é espalhada por átomos e moléculas de gases presentes no ar (i.e. N₂, O₂, etc.) com diâmetros de até aproximadamente um décimo do comprimento de onda da luz (IQBAL, 1983). É o tipo de fenômeno que predomina nas condições de céu claro e limpo onde os comprimentos de onda de luz violeta e azul são espalhados em maior quantidade; sendo a principal razão pela qual a cor do céu é azul. Quanto mais saturado, mais predominante é o efeito do espalhamento Rayleigh. Saturado no presente contexto, significa menor quantidade dos componentes verde e vermelho. Durante este espalhamento não ocorre nenhum ganho ou perda de energia. Vale a pena salientar que na ausência do espalhamento Rayleigh um observador na superfície da terra observaria no céu as estrelas com um fundo negro, mesmo durante o dia. Um exemplo do espalhamento Rayleigh pode ser observado na figura 3.9.
- **Espalhamento Mie vermelho-alaranjado.** A coloração vermelho-alaranjada durante o pôr-do-sol que aparece nas condições de céu claro, também é devida ao espalhamento Rayleigh. Durante o pôr-do-sol a camada atmosférica chega a ficar até doze vezes mais densa quando comparada ao meio dia. Próximo ao horizonte nas condições de céu claro, a luz do sol atravessa uma camada atmosférica mais espessa, com moléculas cujo diâmetro é um pouco maior como por exemplo os pequenos aerossóis. Nestas circunstâncias ocorre o espalhamento parcial da componente verde da luz, resultando uma cor de tom vermelho-alaranjada. Estas moléculas apesar de serem um pouco maiores, ainda se encontram no critério de ser muito menor que o comprimento de onda da luz incidente (RICHARDS, 1995). Um exemplo do espalhamento Rayleigh vermelho-alaranjado pode ser observado na figura 3.9.
- **Espalhamento próximo à superfície.** Ocorre quando a luz interage com partículas e moléculas com tamanho maior do que quatro vezes o comprimento de onda. Partículas de poeira, pólen, fumaça, névoa, vapor de água, etc. estão entre as principais causas. O espalhamento Mie tende a acontecer nas partes baixas da atmosfera, próximo à superfície, onde as partículas maiores são abundantes. Em dias de céu claro com o Sol próximo ao meio dia, pode-se notar claramente no HC um gradiente na cor azul do céu entre a posição zenital e a horizontal. Esta perda de saturação na cor azul do céu se dá por causa do acréscimo do brilho causando espalhamento Mie. Em dias de céu claro com o Sol pouco acima do horizonte, o brilho por espalhamento Mie também está presente, acentuando a baixa saturação de cor da luz solar. Um exemplo do espalhamento Rayleigh vermelho-alaranjado pode ser observado nas figuras 3.9 e 3.10(b).
- **Espalhamento por turbidêz atmosférica na vizinhança solar.** Ocorre em três situações diferentes, causando uma variabilidade na área da imagem correspondente ao círculo solar. A primeira ocorre quando nas condições de céu claro ao redor do sol,



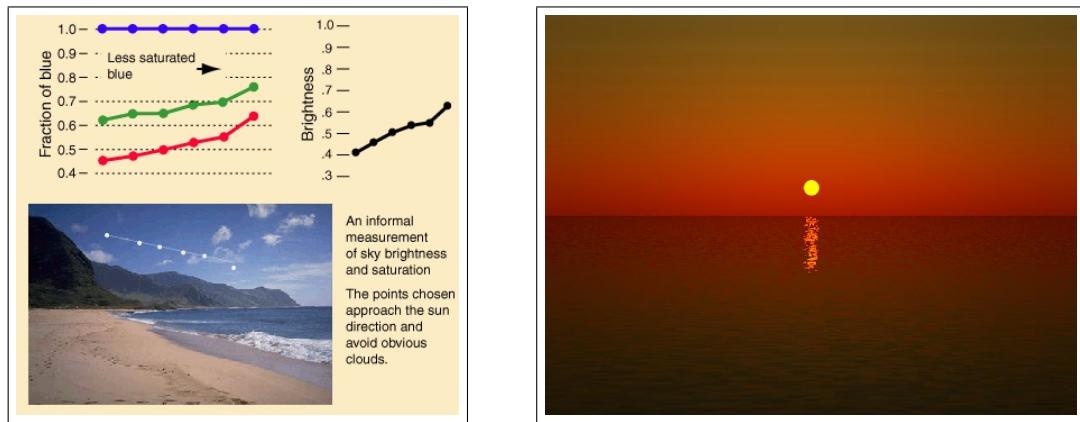
(a) Exemplos dos fenômenos físicos observáveis no CH



(b) Mapeamento dos fenômenos físicos no espaço de cor RGB

Figure 3.9: Exemplos dos fenômenos físicos observáveis no CH e o mapeamento no locci típico no espaço de cor RGB. Figura obtida de (MANTELLI; WANGENHEIN; PEREIRA, 2009)

devido à grande quantidade de partículas (poeira, névoa, maresia, etc.) na atmosfera. A segunda quando o sol está um pouco acima do horizonte, causando um borrão branco ao redor de sua imagem. A terceira quando aparece na imagem os raios do crepúsculo solar entre as nuvens. Um exemplo do espalhamento Mie por turbidêz atmosférica na



(a) Efeito Rayleigh e perda de saturação por efeito Mie, com o sol em altos ângulos de elevação.
 (b) Efeito Mie de cor vermelho-alaranjado, com o sol em pequenos ângulos de elevação.

Figure 3.10: Exemplos de padrões de cores ópticos-atmosférico causados pela luz solar.

vizinhança solar pode ser observado nas figuras 3.9 e 3.10.

- **Espalhamento não seletivo ou nuvem branca.** Ocorre quando a luz do sol atravessa a neblina e as gotículas de água das nuvens, espalhando quantidades equivalentes de azul, verde e vermelho resultando na cor branca característica das nuvens (LILLE-SAND; KIEFER, 1994). Um exemplo do espalhamento Mie não seletivo ou nuvem branca pode ser observado nas figuras 3.9 e 3.10.
- **Espalhamento por difusão da luz através das nuvens.** Ocorre quando a luz do sol brilha intensamente atrás e através das nuvens brancas. Um exemplo do espalhamento Mie por difusão da luz através das nuvens pode ser observado nas figuras 3.9 e 3.10.
- **Absorção das nuvens** Ocorre quando a luz do sol é absorvida e transformada em outra forma de energia (geralmente calor). A absorção é seletiva em relação ao comprimento de onda incidente. No caso das nuvens, algumas são tão expressas que a luz não passa. No espectro visível se manifesta na forma de nuvens de cor cinza e escura. Um exemplo da absorção das nuvens pode ser observado nas figuras 3.9 e 3.10.

Chapter 4

RESULTADOS

Neste capítulo estarão descritos os resultados obtidos durante o desenvolvimento da pesquisa, utilizando a metodologia descrita no capítulo 3. Nem todos os padrões propostos puderam ser representados, pela ausência de características distinguíveis no espaço de cores suficientes para discriminar e classificar um padrão. Outros padrões não puderam ser identificados devido a limitações existentes na câmera, no programa desenvolvido pelo fabricante ou no conjunto de dados disponível. Os resultados da presente pesquisa, estão distribuídos nas subseções do presente capítulo na seguinte seqüencia. A primeira seção apresenta a definição, análise exploratória supervisionada e da representação dos padrões visuais propostos na metodologia. As subseções apresentam os passos intermediários desenvolvidos para cada padrão, a análise supervisionada utilizada nas imagens representativas do padrão. A segunda seção descreve os problemas encontrados na representação matemática dos padrões visuais, os ajustes empíricos e um resumo dos critérios de classificação utilizados. A terceira seção descreve uma síntese dos resultados gerais obtidos a partir dos métodos propostos. Na quarta seção esta descrita a análise dos resultados obtidos. Na quinta uma comparação dos dados com outros métodos encontrados na literatura especializada. Na sexta e última,a validação dos dados utilizando imagens de satélite.

4.1 Descrição da Análise Exploratória para a Representação dos Padrões Visuais

A análise supervisionada consiste na definição gráfica, na caracterização estatística, no estabelecimento dos critérios de classificação e na discriminação dos padrões, baseados na metodologia apresentada no capítulo anterior. Nas subseções a seguir estarão descritos os processos utilizados.

4.1.1 Padrão MIE

A definição do padrão MIE foi desenvolvida observando as imagens adquiridas próximas à ocorrência do crepúsculo matutino ou vespertino. Nesta ocasião são observados somente os comprimentos de onda maiores de laranja e vermelho (LILLESAND; KIEFER, 1994) página 10. Este fenômeno ocorre por que a luz solar percorre um caminho maior onde o

espalhamento e a absorção dos comprimentos de onda menores (Rayleigh) é praticamente completa.

4.1.1.1 Definição gráfica do padrão MIE

Na presente subseção estão descritos os passos necessários para a representação gráfica do padrão definido como MIE. Na imagem selecionada, o espalhamento MIE estava presente em menor quantidade e foi necessária a eliminação de outros padrões da imagem antes de representá-lo de maneira apropriada. A principal característica, é o pixel apresentar um atributo de cor em um ângulo distinto na dimensão H do espaço de cores HSL. Após a observação da ocorrência deste padrão no espaço de cores HSL, utilizou-se um índice empírico de limiarização de 0,5 na dimensão H para separá-lo e em seguida caracterizá-lo no espaço RGB. A imagem utilizada para representar o padrão Mie esta ilustrada na figura 4.1.

4.1.1.2 Descrição estatística da definição gráfica do padrão MIE

Na presente subseção estão descritos os passos necessários para a definição estatística do padrão definido como efeito MIE. A AED do padrão MIE utilizando-se um pacote estatístico é descrita nos gráficos da figura 4.2.

A eliminação dos *outliers* do padrão MIE utilizando a distância interquartil, está descrito nas equações 4.1.

vermelho

$$\begin{aligned} IQD_R &= Q3 - Q1 = 198 - 159 \Rightarrow IQD_R = 39 \\ IQI_R &= Q1 - 1,5 \times IQI_R = 159 - 58,5 \Rightarrow IQI_R = 100,5 \\ IQS_R &= Q3 + 1,5 \times IQS_R = 198 + 58,5 \Rightarrow IQS_R = 256,5 \end{aligned}$$

verde

$$\begin{aligned} IQD_G &= Q3 - Q1 = 195 - 161 \Rightarrow IQD_G = 34 \\ IQI_G &= Q1 - 1,5 \times IQI_G = 161 - 51 \Rightarrow IQI_G = 110 \\ IQS_G &= Q3 + 1,5 \times IQS_G = 195 + 51 \Rightarrow IQS_G = 242 \end{aligned} \quad (4.1)$$

azul

$$\begin{aligned} IQD_B &= Q3 - Q1 = 187 - 154 \Rightarrow IQD_B = 33 \\ IQI_B &= Q1 - 1,5 \times IQI_B = 154 - 49,5 \Rightarrow IQI_B = 104,5 \\ IQS_B &= Q3 + 1,5 \times IQS_B = 187 + 49,5 \Rightarrow IQS_B = 236,5 \end{aligned}$$

O cálculo do erro padrão para o padrão MIE de acordo com a equação 3.8 esta descrito na equação 4.2.

$$\begin{aligned} e_r &= \frac{s_r}{\sqrt{n_{MIE}}} = \frac{28}{\sqrt{3247}} = 0,491 \\ e_g &= \frac{s_g}{\sqrt{n_{MIE}}} = \frac{25,52}{\sqrt{3247}} = 0,448 \\ e_b &= \frac{s_b}{\sqrt{n_{MIE}}} = \frac{23,85}{\sqrt{3247}} = 0,418 \end{aligned} \quad (4.2)$$

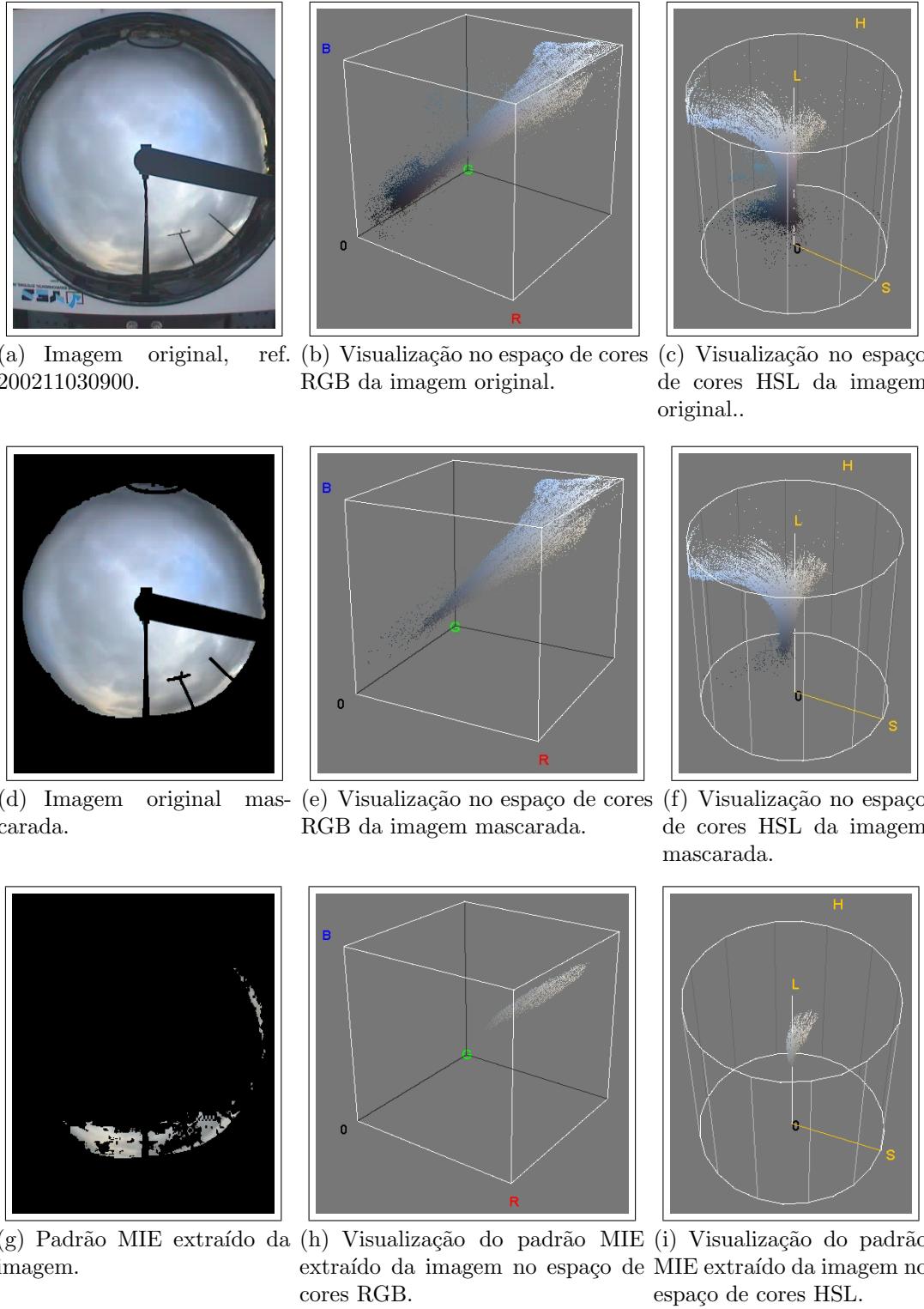
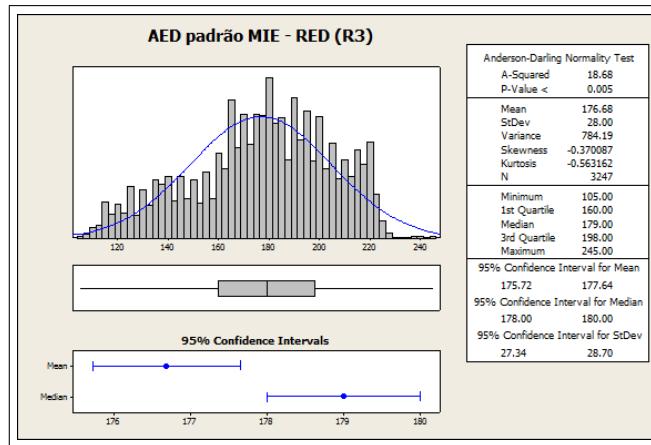
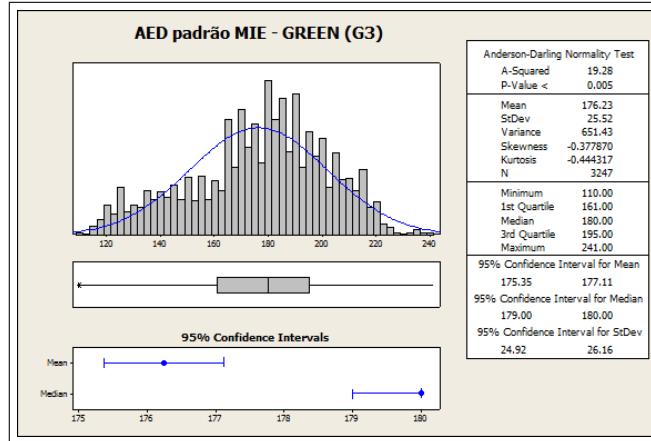


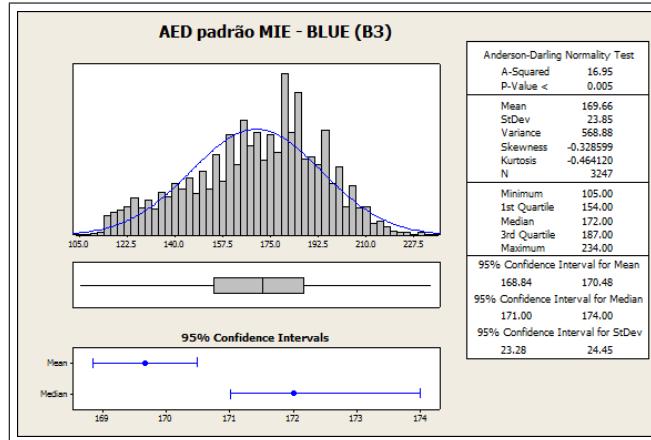
Figure 4.1: Definição do padrão MIE através das imagens e a representação do *locus* típico dos píxeis nos espaços de cores RGB e HSL. Na primeira linha de figuras a imagem original. Na segunda, a imagem subtraída dos fatores que não representam informação útil. Na terceira fileira a imagem indicando somente a presença do espalhamento MIE.



(a) AED padrão vermelho.



(b) AED padrão verde.



(c) AED padrão azul.

Figure 4.2: AED do padrão Mie, após a eliminação dos *outliers*.

$$y_{MIE} = \begin{vmatrix} 176,681 \\ 176,233 \\ 169,657 \end{vmatrix} + \begin{vmatrix} 784,194 & 701,700 & 628,364 \\ 701,700 & 651,431 & 594,634 \\ 628,364 & 594,634 & 568,876 \end{vmatrix} + \begin{vmatrix} 0,491 \\ 0,448 \\ 0,418 \end{vmatrix} \quad (4.3)$$

A representação estatística do padrão MIE esta descrita na equação 4.3.

4.1.1.3 Critério de classificação do padrão MIE

A implementação da distância de MAHALANOBIS foi calculada a partir da equação 3.9. A racionalização da solução da equações lineares pelo aplicativo MAPLE para o padrão MIE utilizada na interface gráfica encontra-se descrita na equação 4.4

$$\begin{aligned} D_{MIE}^2 = & \\ & (176,681 - r) * (-0,399255783 - 0,0440891671506497066 * r + \\ & 0,0662450227090553290 * g - 0,0205448945044711916 * b) + \\ & (176,233 - g) * (0,56287227 + 0,0662450227090553706 * r - \\ & 0,133010092533842578 * g + 0,0658602857463068958 * b) + \\ & (169,657 - b) * (0,.150879911 - 0,0205448945044712228 * r + \\ & 0,0658602857463069374 * g - 0,0479069130426699740 * b) \end{aligned} \quad (4.4)$$

4.1.1.4 Limiar de discriminação para o padrão MIE

A discriminação do padrão Espalhamento MIE será baseado no limiar da distribuição F, que é obtida a partir da consulta em uma tabela estatística (MONTGOMERY; RUNGER, 2003) anexo A, conforme descrito na equação 3.9. O Valor do limiar obtido a partir de tabelas estatísticas está descrito na equação 4.5

$$DD^2 \leq \frac{(n-1)p}{(n-p)} F_{p,n-p,(\alpha)} = \frac{(3247-1)3}{(3247-3)} F_{p,n-p,(\alpha)} = 3,002 F_{p,n-p,(\alpha)} \quad (4.5)$$

onde:

- $F_{p,n-p,(\alpha)}$ = 2,08($\alpha = 0,1$) ; 2,61($\alpha = 0,05$) ; 3,78($\alpha = 0,01$)
- DD^2 : é o limiar de discriminação para o padrão com os seguintes valores 6,24($\alpha = 0,1$) ; 7,83($\alpha = 0,05$) ; 11,34($\alpha = 0,01$).
- p: é o número de dimensões do espaço de cores RGB = 3
- n: é o número de amostras representativas do padrão obtidas na AED após a eliminação dos outliers = 3247
- α : é o coeficiente de confiança = 0,1 (0,05 ; 0,01)

Vale a pena salientar que o limiar de discriminação é muito próximo ao utilizado por outros padrões. A razão desta proximidade, é que o índice n-p é muito maior do que 120. Acima deste valor são considerados infinitos graus de liberdade nas tabelas estatísticas da distribuição F (MONTGOMERY; RUNGER, 2003). Os píxeis que estiverem abaixo do limiar de discriminação calculados a partir da equação 4.4, para um coeficiente de confiança de 0.01 serão considerados como padrão MIE.

O cálculo do intervalo de confiança de bonferroni definido na equação 3.10 para o padrão MIE esta descrito na equação 4.6.

vermelho

$$\left(\mu_r \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p,(\alpha=0,01)}} \sqrt{\frac{\sigma_r^2}{n}} \right) = \left(176,68 \pm \sqrt{\frac{3(3247-1)}{(3247-3)} 3,78} \sqrt{\frac{28^2}{3247}} \right)$$

$$= (176,68 \pm 5,5722,4)$$

verde

$$\left(\mu_g \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p,(\alpha=0,01)}} \sqrt{\frac{\sigma_g^2}{n}} \right) = \left(176,23 \pm \sqrt{\frac{3(3247-1)}{(3247-3)} 3,78} \sqrt{\frac{25,52^2}{3247}} \right) \quad (4.6)$$

$$= (176,23 \pm 5,0788)$$

azul

$$\left(\mu_b \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p,(\alpha=0,01)}} \sqrt{\frac{\sigma_b^2}{n}} \right) = \left(169,66 \pm \sqrt{\frac{3(3247-1)}{(3247-3)} 3,78} \sqrt{\frac{23,85^2}{3247}} \right)$$

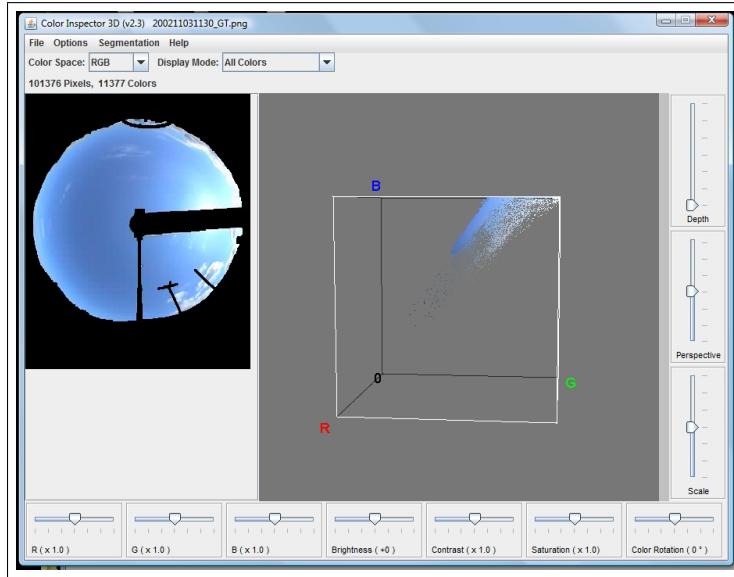
$$= (169,66 \pm 4,7464)$$

4.1.2 Padrão RAY

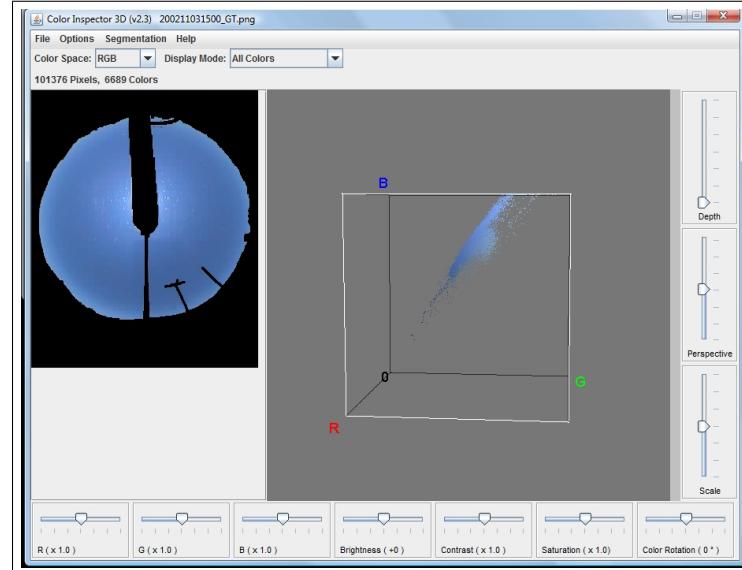
Segundo (LILLESAND; KIEFER, 1994) ”o céu azul é uma manifestação do espalhamento Rayleigh. À medida que a luz do sol interage com a atmosfera terrestre, é espalhada predominantemente comprimentos de onda menores (azul). Conseqüentemente observamos o céu azul”. A definição do padrão Rayleigh foi efetuada observando as imagens adquiridas próximas ao meio dia solar. Nesta ocasião a luz solar percorre um caminho menor até a superfície.

4.1.2.1 Definição gráfica do padrão RAY

Na presente subseção estão descritos os passos necessários para a definição gráfica do padrão definido como espalhamento Rayleigh. Para este padrão duas imagens foram necessárias para representar toda a variabilidade dos píxeis em seu domínio, para definir completamente este padrão. As imagens do padrão RAY apresentaram uma certa variância ao longo do dia entre horários diferentes, de forma que somente um horário não é suficientemente representativo de todos os casos de ocorrência de luminosidade deste padrão. As



(a) Imagem ref. 200211031130.



(b) Imagem ref. 200211031500.

Figure 4.3: Visualização da composição do padrão Rayleigh em dois níveis de luminosidade diferentes no espaço de cores RGB.

imagens utilizadas como referência para este padrão estão ilustradas na figura 4.3.

Em ambas as imagens de referência, o padrão RAY estava representado em grande quantidade e também foi necessária a eliminação de outros padrões espúrios da imagem antes de utilizá-la como padrão. A primeira imagem utilizada para representar o padrão RAY esta ilustrada na figura 4.4 e a segunda na figura 4.5.

4.1.2.2 Descrição estatística da definição gráfica do padrão RAY

Na presente subseção estão descritos os passos necessários para a definição estatística do padrão definido como espalhamento Rayleigh. A AED do padrão Rayleigh esta descrita

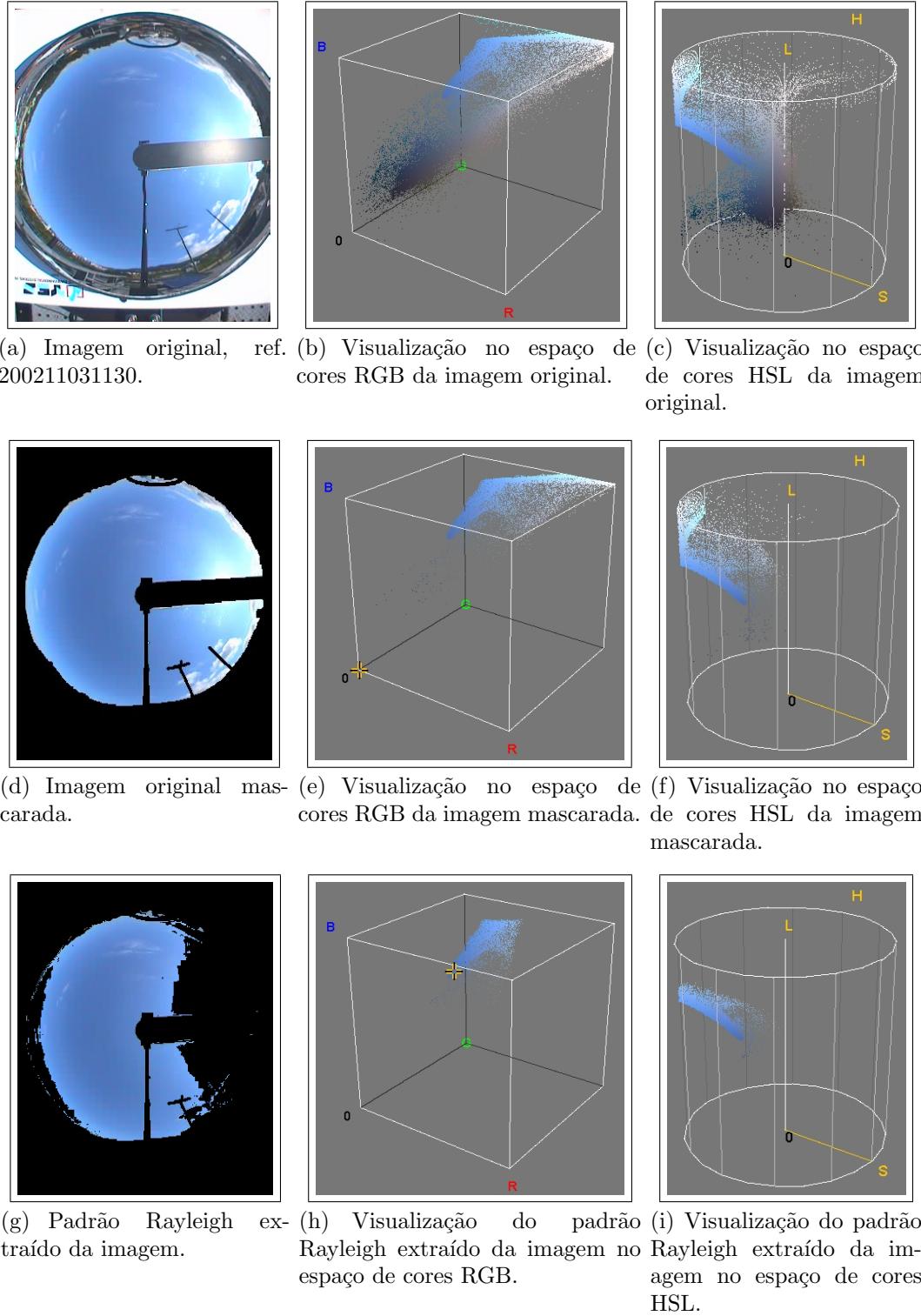


Figure 4.4: Definição do padrão RAY através das imagens e a representação do *locus* típico dos píxeis nos espaços de cores RGB e HSL. Na primeira linha de figuras a imagem original, na segunda a imagem subtraída dos fatores que não representam informação útil, na terceira fileira a imagem indicando somente a presença do espalhamento Rayleigh com eliminação de outliers através da EGD e dos pixels saturados.

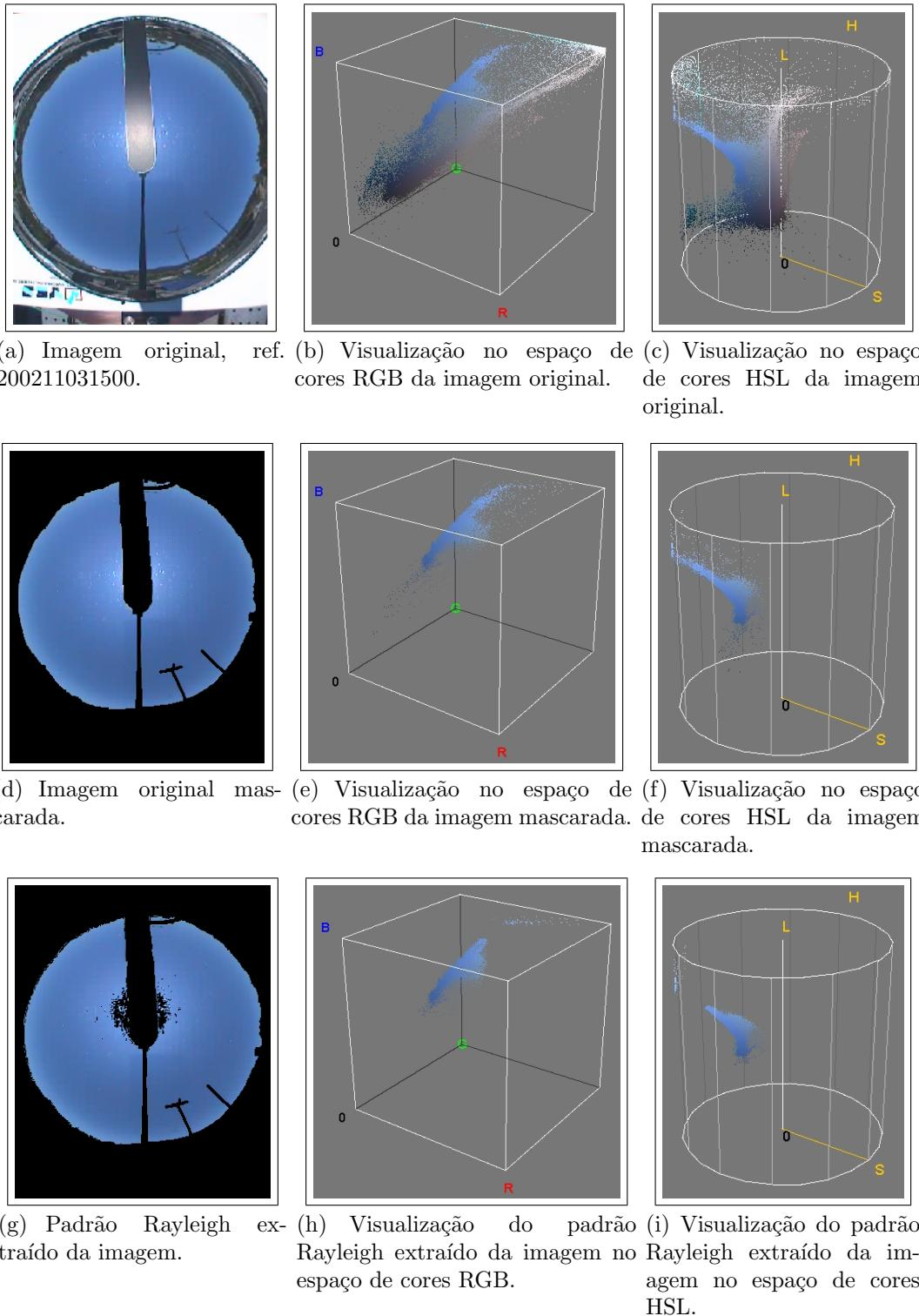
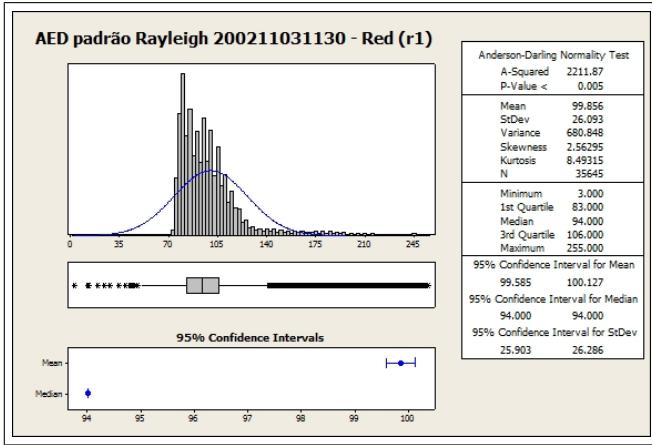
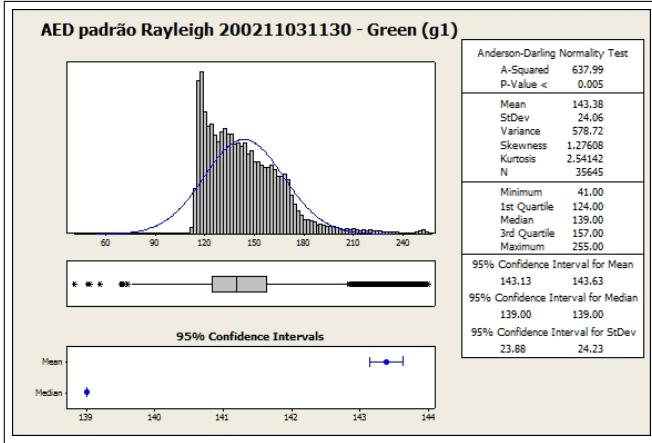


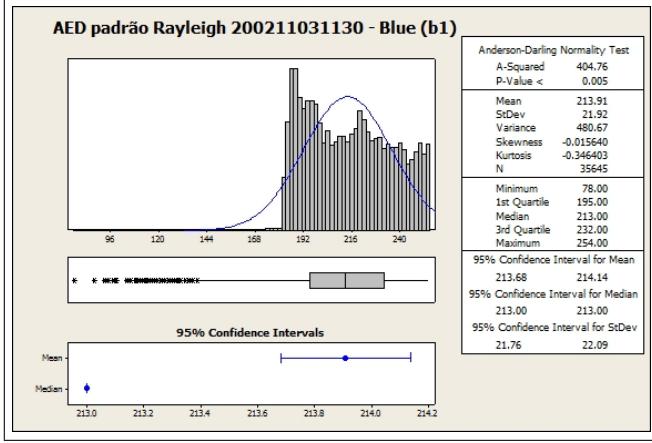
Figure 4.5: Definição do padrão Rayleigh através das imagens e a representação do *locus* típico dos píxeis nos espaços de cores RGB e HSL. Na primeira linha de figuras a imagem original, na segunda a imagem subtraída dos fatores que não representam informação útil, na terceira fileira a imagem indicando somente a presença do espalhamento Rayleigh.



(a) AED cor vermelha.



(b) AED cor verde.

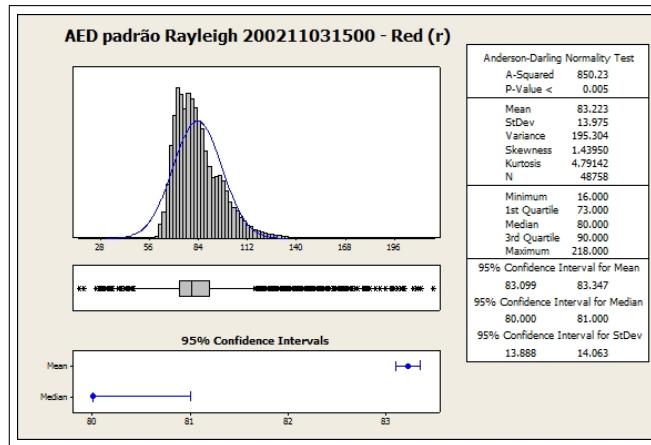


(c) AED cor azul.

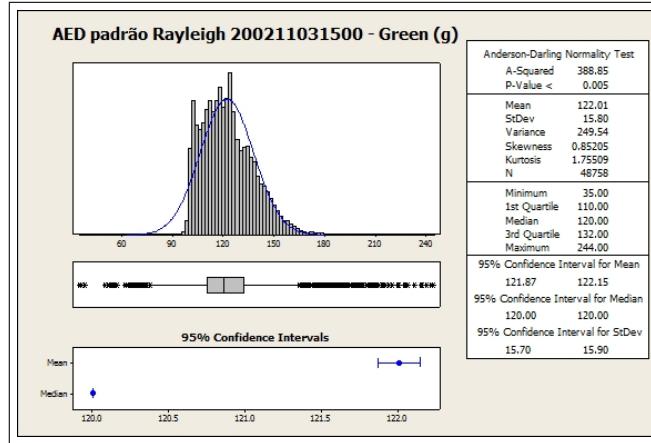
Figure 4.6: AED do padrão RAY, após a eliminação dos *outliers* da imagem 200211031130.

nas figuras 4.6 e 4.7.

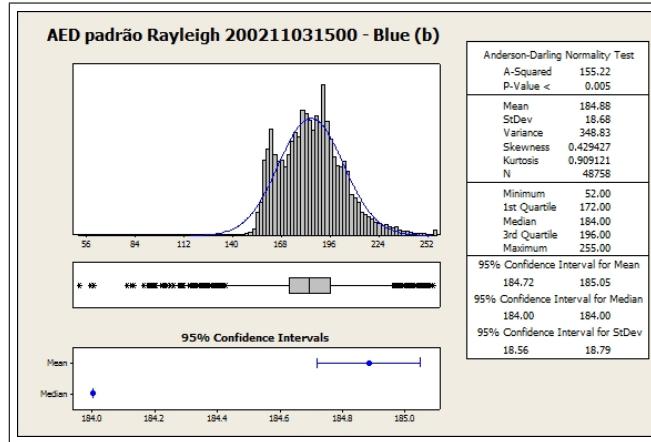
A eliminação dos *outliers* através da distância interquartil, do padrão RAY utilizado está descrito nas equações 4.7 e 4.8. Além dos *outliers*, também foram removidos das imagens os píxeis saturados que representavam o padrão difusão. Após a remoção dos píxeis espúrios os dados foram concatenados e uma nova AED efetuada para a definição do padrão



(a) AED cor vermelha.



(b) AED cor verde.



(c) AED cor azul.

Figure 4.7: AED do padrão RAY, após a eliminação dos *outliers* da imagem 200211031500.

vermelho

$$\begin{aligned} IQD_R &= Q3 - Q1 = 106 - 83 \Rightarrow IQD_R = 23 \\ IQI_R &= Q1 - 1,5 \times IQI_R = 83 - 34,5 \Rightarrow IQI_R = 48,5 \\ IQS_R &= Q3 + 1,5 \times IQS_R = 106 + 34,5 \Rightarrow IQS_R = 140,5 \end{aligned}$$

verde

$$\begin{aligned} IQD_G &= Q3 - Q1 = 157 - 124 \Rightarrow IQD_G = 33 \\ IQI_G &= Q1 - 1,5 \times IQI_G = 124 - 49,5 \Rightarrow IQI_G = 74,5 \\ IQS_G &= Q3 + 1,5 \times IQS_G = 157 + 49,5 \Rightarrow IQS_G = 206,5 \end{aligned} \quad (4.7)$$

azul

$$\begin{aligned} IQD_B &= Q3 - Q1 = 232 - 195 \Rightarrow IQD_B = 37 \\ IQI_B &= Q1 - 1,5 \times IQI_B = 195 - 55,5 \Rightarrow IQI_B = 139,5 \\ IQS_B &= Q3 + 1,5 \times IQS_B = 232 + 55,5 \Rightarrow IQS_B = 287,5 \end{aligned}$$

vermelho

$$\begin{aligned} IQD_R &= Q3 - Q1 = 90 - 73 \Rightarrow IQD_R = 17 \\ IQI_R &= Q1 - 1,5 \times IQI_R = 73 - 25,5 \Rightarrow IQI_R = 47,5 \\ IQS_R &= Q3 + 1,5 \times IQS_R = 90 + 25,5 \Rightarrow IQS_R = 115,5 \end{aligned}$$

verde

$$\begin{aligned} IQD_G &= Q3 - Q1 = 132 - 110 \Rightarrow IQD_G = 22 \\ IQI_G &= Q1 - 1,5 \times IQI_G = 110 - 33 \Rightarrow IQI_G = 77 \\ IQS_G &= Q3 + 1,5 \times IQS_G = 132 + 33 \Rightarrow IQS_G = 165 \end{aligned} \quad (4.8)$$

azul

$$\begin{aligned} IQD_B &= Q3 - Q1 = 196 - 172 \Rightarrow IQD_B = 24 \\ IQI_B &= Q1 - 1,5 \times IQI_B = 172 - 36 \Rightarrow IQI_B = 136 \\ IQS_B &= Q3 + 1,5 \times IQS_B = 196 + 36 \Rightarrow IQS_B = 232 \end{aligned}$$

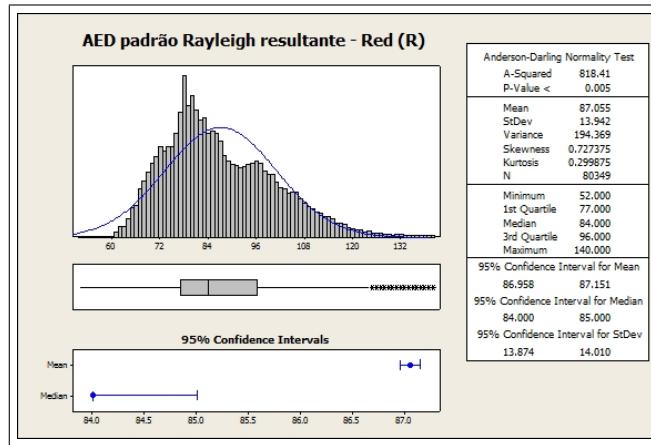
Ray resultante. A AED do padrão RAY concatenado esta ilustrado na figura 4.8.

O cálculo do erro padrão para o padrão RAY de acordo com a equação 3.8, foi efetuado após a concatenação dos dois padrões conforme esta descrito na equação 4.9.

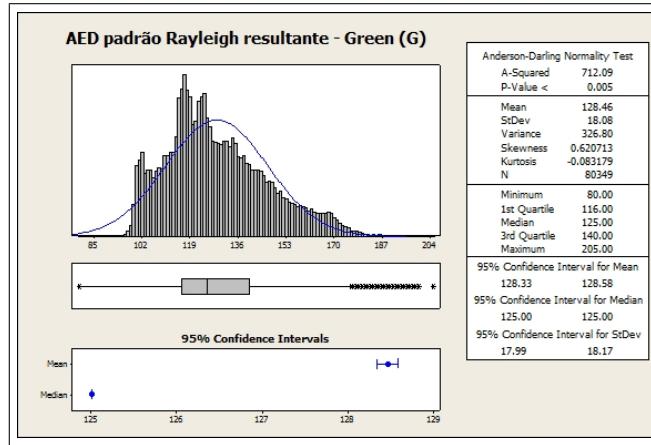
$$\begin{aligned} e_r &= \frac{s_r}{\sqrt{n_{RAY}}} = \frac{13,942}{\sqrt{80349}} = 0,0492 \\ e_g &= \frac{s_g}{\sqrt{n_{RAY}}} = \frac{18,08}{\sqrt{80349}} = 0,0638 \\ e_b &= \frac{s_b}{\sqrt{n_{RAY}}} = \frac{23,53}{\sqrt{80349}} = 0,0830 \end{aligned} \quad (4.9)$$

A representação estatística do padrão RAY esta descrita na equação 4.10.

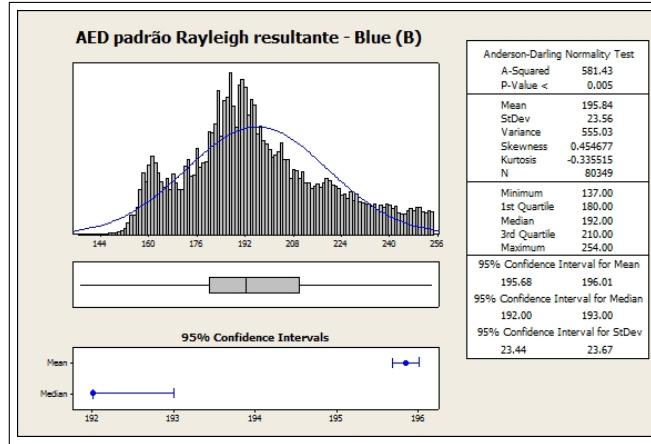
$$y_{RAY} = \left| \begin{array}{c} 87,055 \\ 128,46 \\ 195,84 \end{array} \right| + \left| \begin{array}{ccc} 194,369 & 237,908 & 286,716 \\ 237,908 & 326,799 & 404,879 \\ 286,716 & 404,879 & 555,027 \end{array} \right| + \left| \begin{array}{c} 0,0492 \\ 0,0638 \\ 0,0830 \end{array} \right| \quad (4.10)$$



(a) AED cor vermelha.



(b) AED cor verde.



(c) AED cor azul.

Figure 4.8: AED do padrão RAY, após a eliminação dos *outliers* e padrões espúrios e resultante da concatenação das imagens 200211031130 e 200211031500.

$$\begin{aligned}
D_{RAY}^2 = & \\
(87.055 - r) * (-0.047292481 - 0.0500454887327130352 * r + & \\
0.0457586785481874302 * g - 0.00752735572219554762 * b) + & \\
(128.46 - g) * (-0.414656769 + 0.0457586785481874510 * r - & \\
0.0736361821426030212 * g + 0.0300778133515898424 * b) + & \\
(195.84 - b) * (0.679760317 - 0.00752735572219556063 * r + & \\
0.0300778133515898528 * g - 0.0198542803660602536 * b)
\end{aligned} \tag{4.11}$$

4.1.2.3 Critério de classificação do padrão RAY

A implementação da distância de MAHALANOBIS após a racionalização da solução das equações lineares pelo aplicativo MAPLE para o padrão RAY encontra-se descrita na equação 4.11

4.1.2.4 Limiar de discriminação para o padrão RAY

A discriminação do padrão RAY será baseado no limiar da distribuição F, que é obtida a partir da consulta em uma tabela estatística (MONTGOMERY; RUNGER, 2003) anexo A, conforme descrito na equação 3.9. O Valor do limiar obtido a partir de tabelas estatísticas está descrito na equação 4.12

$$DD^2 \leq \frac{(n-1)p}{(n-p)} F_{p,n-p,(\alpha)} = \frac{(80349-1)3}{(80349-3)} F_{p,n-p,(\alpha)} = 3,00007 F_{p,n-p,(\alpha)} \tag{4.12}$$

onde:

- $F_{p,n-p,(\alpha)}$ = 2,08($\alpha = 0,1$) ; 2,61($\alpha = 0,05$) ; 3,78($\alpha = 0,01$)
- DD^2 : é o limiar de discriminação para o padrão com os seguintes valores 6,24($\alpha = 0,1$) ; 7,83($\alpha = 0,05$) ; 11,34($\alpha = 0,01$).
- p: é o número de dimensões do espaço de cores RGB = 3
- n: é o número de amostras representativas do padrão obtidas na AED após a eliminação dos outliers = 80349
- α : é o coeficiente de confiança = 0,1 (0,05 ; 0,01)

Vale a pena salientar que o mesmo limiar de discriminação é utilizado por outros padrões. A razão deste critério é que o índice n-p é muito maior do que 120. Acima deste valor são considerados infinitos graus de liberdade nas tabelas estatísticas da distribuição F. Os píxeis que estiverem abaixo do limiar de discriminação calculados a partir da equação 4.4, para um coeficiente de segurança de 0,01 serão considerados como espalhamento Mie.

Cálculo do intervalo de confiança de bonferroni na equação 4.13.

vermelho

$$\left(\mu_r \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p,(\alpha=0,01)}} \sqrt{\frac{\sigma_r^2}{n}} \right) = \left(87,06 \pm \sqrt{\frac{3(80349-1)}{(80349-3)} 3,78} \sqrt{\frac{13,94^2}{80349}} \right)$$

$$= (87,06 \pm 0,053370)$$

verde

$$\left(\mu_g \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p,(\alpha=0,01)}} \sqrt{\frac{\sigma_g^2}{n}} \right) = \left(128,46 \pm \sqrt{\frac{3(80349-1)}{(80349-3)} 3,78} \sqrt{\frac{18,08^2}{80349}} \right) \quad (4.13)$$

$$= (128,46 \pm 0,72330)$$

azul

$$\left(\mu_b \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p,(\alpha=0,01)}} \sqrt{\frac{\sigma_b^2}{n}} \right) = \left(195,44 \pm \sqrt{\frac{3(80349-1)}{(80349-3)} 3,78} \sqrt{\frac{23,55^2}{80349}} \right)$$

$$= (195,44 \pm 0,942136)$$

4.1.3 Padrão Espalhamento Não Seletivo (ENS) ou Nuvens

O espalhamento Espalhamento Não Seletivo ENS foi definido utilizando as imagens com nuvens. De acordo com (LILLESAND; KIEFER, 1994) página 10 "as nuvens espalham quantidades iguais da radiação dos comprimentos de onda vermelho verde e azul, por isto a sua resultante é branca". Mas de acordo com o observado por (MANTELLI et al., 2010b), as nuvens possuem um pequeno componente do azul devido ao componente azul. Isto faz com que o padrão das nuvens ocorra ligeiramente deslocado para a região da cor azul, ao invés de ao redor do eixo principal do cubo no espaço de cores RGB, ou eixo central do HSL. Os píxeis que representam as nuvens mais densas que absorvem uma maior quantidade de luz ocorrem mais próximos ao eixo principal. Mas à medida que a densidade vai reduzindo eles vão se afastando do eixo principal em direção à cor que representa o espalhamento Rayleigh. Isto permite a caracterização do *locus* geométrico utilizando a distância euclideana (EGD) como indicativo de densidade óptica , e a sua projeção sobre o eixo principal (EGP) como a sua intensidade luminosa (MANTELLI et al., 2010b). O ponto exato de separação entre os dois padrões ficará determinado por critérios de diferenças estatísticas entre as duas amostras, utilizando-se a EGD. Estudos mais refinados para determinação do ponto exato de discriminação entre os dois padrões para determinar quando um píxel deixa de representar uma nuvem e passa a representar o céu claro fica como sugestão para trabalhos futuros.

4.1.3.1 Definição gráfica do padrão ENS

Na presente subseção estão descritos os passos necessários para a representação gráfica do padrão definido como ENS. Na imagem selecionada, ENS estava presente em grande quantidade, mas foi necessária a eliminação de outros padrões (MIE e RAY) presentes na imagem para representar o ENS de maneira apropriada. A imagem utilizada para representar

o padrão ENS, foi a mesma utilizada para representar o padrão MIE e esta ilustrada na figura 4.9.

4.1.3.2 Descrição estatística da definição gráfica do padrão ENS

Na presente subseção estão descritos os passos necessários para a definição estatística do padrão definido como ENS. A AED do padrão ENS esta descrita na figura 4.10.

A eliminação dos *outliers* através da distância interquartil, do padrão ENS utilizado está descrito nas equações 4.14.

O cálculo do erro padrão para o padrão ENS de acordo com a equação 3.8 esta descrito na equação 4.15.

$$\begin{aligned} e_r &= \frac{s_r}{\sqrt{n_{ENS}}} = \frac{28,68}{\sqrt{40542}} = 0,1424 \\ e_g &= \frac{s_g}{\sqrt{n_{ENS}}} = \frac{29,45}{\sqrt{40542}} = 0,1463 \\ e_b &= \frac{s_b}{\sqrt{n_{ENS}}} = \frac{32,14}{\sqrt{40542}} = 0,1596 \end{aligned} \quad (4.15)$$

A representação estatística do padrão ENS esta descrita na equação 4.16.

$$y_{ENS} = \left| \begin{array}{c} 157,54 \\ 169,36 \\ 184,99 \end{array} \right| + \left| \begin{array}{ccc} 822,531 & 826,204 & 817,627 \\ 826,204 & 867,389 & 902,137 \\ 817,627 & 902,137 & 1033,119 \end{array} \right| + \left| \begin{array}{c} 0,1424 \\ 0,1463 \\ 0,1596 \end{array} \right| \quad (4.16)$$

4.1.3.3 Critério de Classificação do padrão ENS

A implementação da distância de MAHALANOBIS após a racionalização da solução das equações lineares pelo aplicativo MAPLE para o padrão ENS encontra-se descrita na equação 4.17

$$\begin{aligned} D_{ENS}^2 &= \\ &(157,542 - r) * (0,006313390 - 0,0579920289793638410 * r + \\ &0,0817419639824926714 * g - 0,0254826418669714498 * b) + \\ &(169,361 - g) * (0,089365846 + 0,0817419639824927132 * r - \\ &0,127776841899000432 * g + 0,0468849958573261950 * b) + \\ &(184,990 - b) * (0,096027434 - 0,0254826418669714812 * r + \\ &0,0468849958573262228 * g - 0,0217413419809087024 * b) \end{aligned} \quad (4.17)$$

4.1.3.4 Limiar de discriminação para o padrão ENS

A discriminação do padrão Espalhamento ENS será baseado no limiar da distribuição F, que é obtida a partir da consulta em uma tabela estatística (MONTGOMERY; RUNGER,

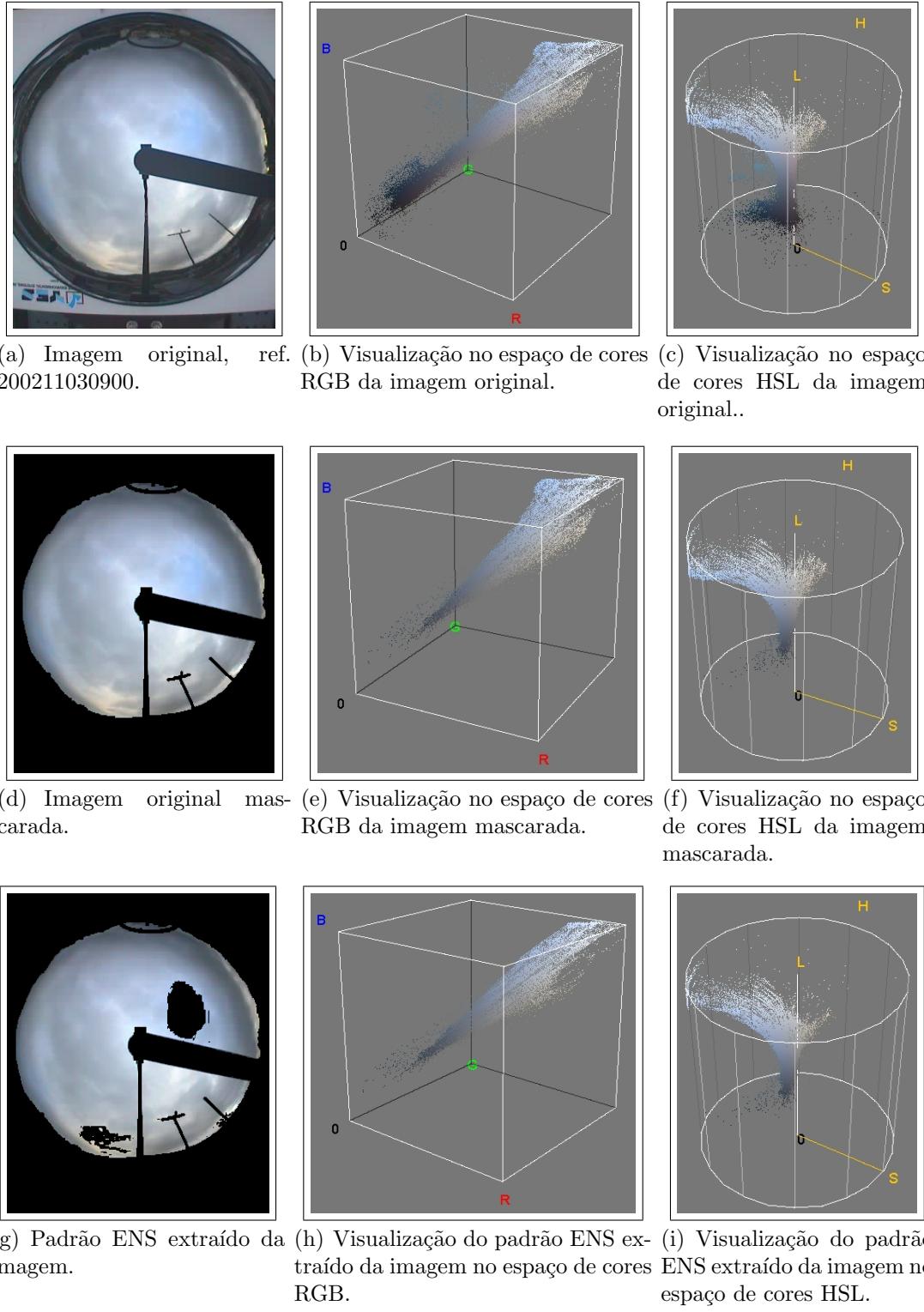
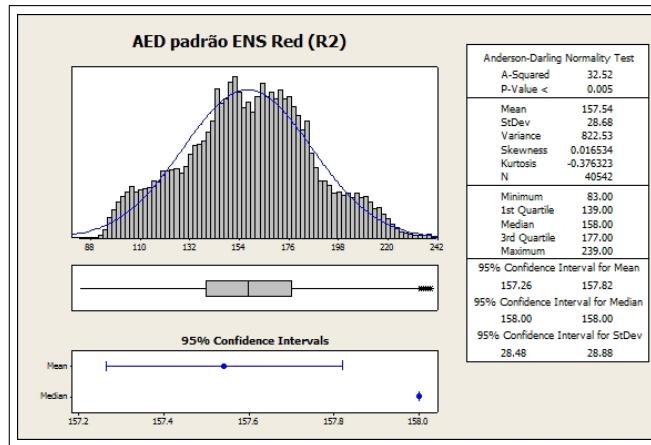
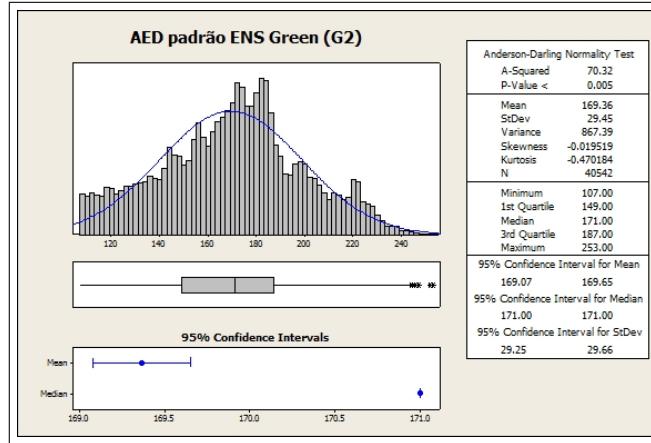


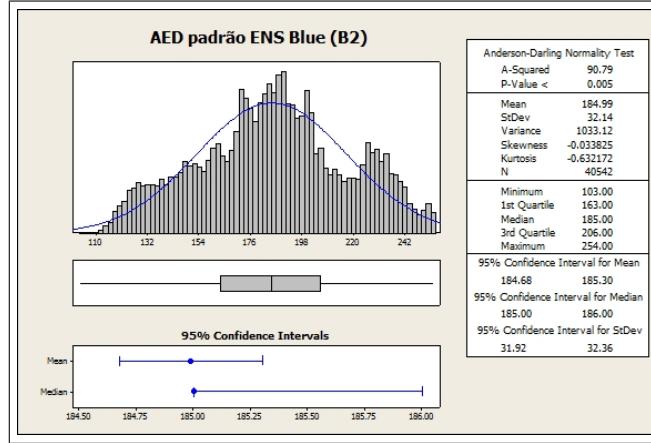
Figure 4.9: Definição do padrão ENS através das imagens e a representação do *locus* típico dos píxeis nos espaços de cores RGB e HSL. Na primeira linha de figuras a imagem original. Na segunda, a imagem subtraída dos fatores que não representam informação útil. Na terceira fileira a imagem indicando somente a presença do espalhamento ENS.



(a) AED cor vermelha.



(b) AED cor verde.



(c) AED cor azul.

Figure 4.10: AED do padrão ENS, após a eliminação dos *outliers*.

vermelho

$$\begin{aligned} IQD_R &= Q3 - Q1 = 174 - 126 \Rightarrow IQD_R = 48 \\ IQI_R &= Q1 - 1,5 \times IQI_R = 126 - 72 \Rightarrow IQI_R = 54 \\ IQS_R &= Q3 + 1,5 \times IQS_R = 174 + 72 \Rightarrow IQS_R = 246 \end{aligned}$$

verde

$$\begin{aligned} IQD_G &= Q3 - Q1 = 185 - 137 \Rightarrow IQD_G = 48 \\ IQI_G &= Q1 - 1,5 \times IQI_G = 137 - 72 \Rightarrow IQI_G = 65 \\ IQS_G &= Q3 + 1,5 \times IQS_G = 185 + 72 \Rightarrow IQS_G = 257 \end{aligned} \quad (4.14)$$

azul

$$\begin{aligned} IQD_B &= Q3 - Q1 = 202 - 149 \Rightarrow IQD_B = 53 \\ IQI_B &= Q1 - 1,5 \times IQI_B = 149 - 79,5 \Rightarrow IQI_B = 105,5 \\ IQS_B &= Q3 + 1,5 \times IQS_B = 202 + 79,5 \Rightarrow IQS_B = 281,50 \end{aligned}$$

2003) anexo A, conforme descrito na equação 3.9. O Valor do limiar obtido a partir de tabelas estatísticas está descrito na equação 4.18

$$DD^2 \leq \frac{(n-1)p}{(n-p)} F_{p,n-p,(\alpha)} = \frac{(40542-1)3}{(40542-3)} F_{p,n-p,(\alpha)} = 3,0001 F_{p,n-p,(\alpha)} \quad (4.18)$$

onde:

- $F_{p,n-p,(\alpha)}$ = 2,08($\alpha = 0,1$) ; 2,61($\alpha = 0,05$) ; 3,78($\alpha = 0,01$)
- DD^2 : é o limiar de discriminação para o padrão com os seguintes valores 6,24($\alpha = 0,1$) ; 7,83($\alpha = 0,05$) ; 11,34($\alpha = 0,01$).
- p: é o número de dimensões do espaço de cores RGB = 3
- n: é o número de amostras representativas do padrão obtidas na AED após a eliminação dos outliers = 40542
- α : é o coeficiente de confiança = 0,1 (0,05 ; 0,01)

Vale a pena salientar que o mesmo limiar de discriminação é utilizado por outros padrões. A razão deste critério é que o índice n-p é muito maior do que 120. Acima deste valor são considerados infinitos graus de liberdade nas tabelas estatísticas da distribuição F. Os píxeis que estiverem abaixo do limiar de discriminação calculados a partir da equação 4.4, para um coeficiente de segurança de 0.99 serão considerados como espalhamento Mie.

Cálculo do intervalo de confiança de bonferroni na equação 4.19.

vermelho

$$\left(\mu_r \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p,(\alpha=0,01)}} \sqrt{\frac{\sigma_r^2}{n}} \right) = \left(157,54 \pm \sqrt{\frac{3(40542-1)}{(40542-3)} 3,78} \sqrt{\frac{28,68^2}{40542}} \right)$$

$$= (157,54 \pm 0,089566)$$

verde

$$\left(\mu_g \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p,(\alpha=0,01)}} \sqrt{\frac{\sigma_g^2}{n}} \right) = \left(169,36 \pm \sqrt{\frac{3(40542-1)}{(40542-3)} 3,78} \sqrt{\frac{29,45^2}{40542}} \right) \quad (4.19)$$

$$= (169,36 \pm 0,090794)$$

azul

$$\left(\mu_b \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p,(\alpha=0,01)}} \sqrt{\frac{\sigma_b^2}{n}} \right) = \left(184,99 \pm \sqrt{\frac{3(40542-1)}{(40542-3)} 3,78} \sqrt{\frac{32,14^2}{40542}} \right)$$

$$= (184,99 \pm 0,094815)$$

4.1.4 Difusão do espalhamento Rayleigh ou padrão DRAY

A difusão do espalhamento Rayleigh (DRAY) ocorre nas seguintes condições de céu claro. Píxeis na proximidade do círculo solar, píxeis próximos à superfície com o sol em pequenos ângulos de elevação ou píxeis na vizinhança das nuvens do tipo cirrus quando presentes. Nos sistemas de reconhecimento automáticos de imagens com técnicas simplificadas, a má caracterização destes píxeis pode levá-los a serem classificados errôneamente como nuvens (MANTELLI, 2001). A sua quantidade também pode ser associada à ocorrência dos aerossóis atmosféricos em suspensão, quando os píxeis representativos deste padrão ocorrem próximo à superfície. É um padrão diferenciado das nuvens por que ocupa um *locus* distinto no espectro de cores, como será demonstrado nas subseções a seguir.

4.1.4.1 Definição Gráfica do padrão DRAY

Na presente subseção estão descritos os passos necessários para a definição gráfica do padrão definido como difusão do espalhamento Rayleigh. Na imagem selecionada, além do padrão DRAY, também estavam presentes o padrão RAY e o padrão Espalhamento não Específico (nuvens), e foi necessário retirá-los de acordo com os critérios de classificação mencionados mais adiante. A imagem representativa de pontos do padrão pode ser obtida conforme ilustrado a figura 4.11.

4.1.4.2 Observações importantes acerca da ocorrência do padrão DRAY

A partir da análise da imagem da figura 4.11 que contém a difusão Rayleigh pode-se observar o seguinte no espaço de cores RGB. O padrão DRAY ocorre tipicamente em uma

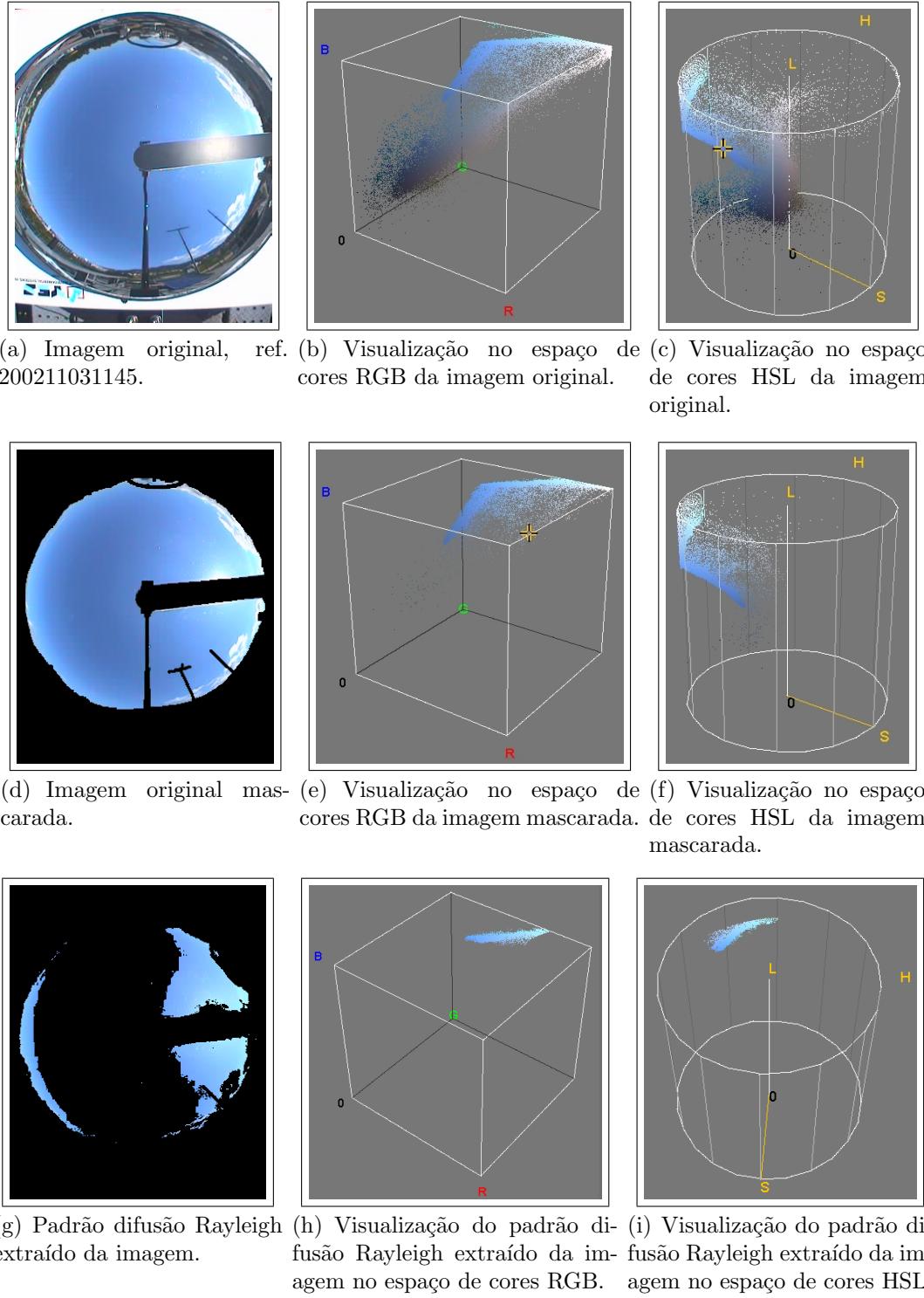


Figure 4.11: Definição do padrão difusão Rayleigh através das imagens e a representação do *locus* típico dos píxeis nos espaços de cores RGB e HSL. Na primeira linha de figuras a imagem original. Na segunda, a imagem subtraída dos fatores que não representam informação útil. Na terceira fileira a imagem indicando somente a presença da difusão do espalhamento Rayleigh.

região que se distribui desde o final do espalhamento Rayleigh, até o final da escala do espaço de cores. Os valores dos píxeis vão saturando progressivamente até o fundo da escala. Em primeiro lugar satura a cor azul depois a verde e por último o vermelho, conforme pode ser observado nas figuras 4.11. Estes píxeis ocupam a região de transição entre os padrões de espalhamento Rayleigh e a difusão das nuvens.

4.1.4.3 Critério de classificação do padrão DRAY

O critério de classificação do padrão DRAY é bem simples. O padrão ocorre próximo ao fundo de escala do espaço de cores e a partir da saturação total da componente azul do pixel, conforme pode ser observado nas figuras 4.11. Qualquer pixel com valor de saturação de azul no fundo de escala pode ser classificado neste padrão, desde que as outras dimensões não tenham chegado à saturação simultaneamente. Para que os píxeis não sejam confundidos com os píxeis de DENS foi estabelecida uma hierarquia no processo de classificação. Em primeiro lugar é feita a verificação se o píxel pertence ao padrão DENS, para depois verificar se o píxel pertence ao padrão DRAY. Para a classificação foi utilizada uma expressão lógica baseada no limiar dos valores, conforme indicado na equação 4.20.

$$D_{Ray} = (((R \geq 250) \text{AND} (G \geq 250) \text{AND} (B \geq 250)) \text{ AND } (EGD \geq 52, 5)) \quad (4.20)$$

onde:

- R: é o valor do componente vermelho do pixel
- G: é o valor do componente verde do pixel
- B: é o valor do componente azul do pixel
- EGD distância Euclideana Geométrica

4.1.5 Difusão do Espalhamento Não Seletivo (DENS) ou Difusão Nuvens

A Difusão do Espalhamento Não Seletivo DENS foi definida utilizando as imagens com o sol brilhando atrás das nuvens. É um padrão que ocupa o fundo de escala do espaço de cores RGB. O DENS se confunde com o DRAY após a saturação das cores azul e verde, por causa da alta intensidade do sol.

4.1.5.1 Definição Gráfica do padrão DENS

Na presente subseção estão descritos os passos necessários para a representação gráfica do padrão definido como DENS. Na imagem selecionada, o DENS estava presente em menor quantidade e foi necessário a eliminação de outros padrões da imagem antes de representá-lo de maneira apropriada. A imagem utilizada para representar o padrão DENS está ilustrada na figura 4.12.

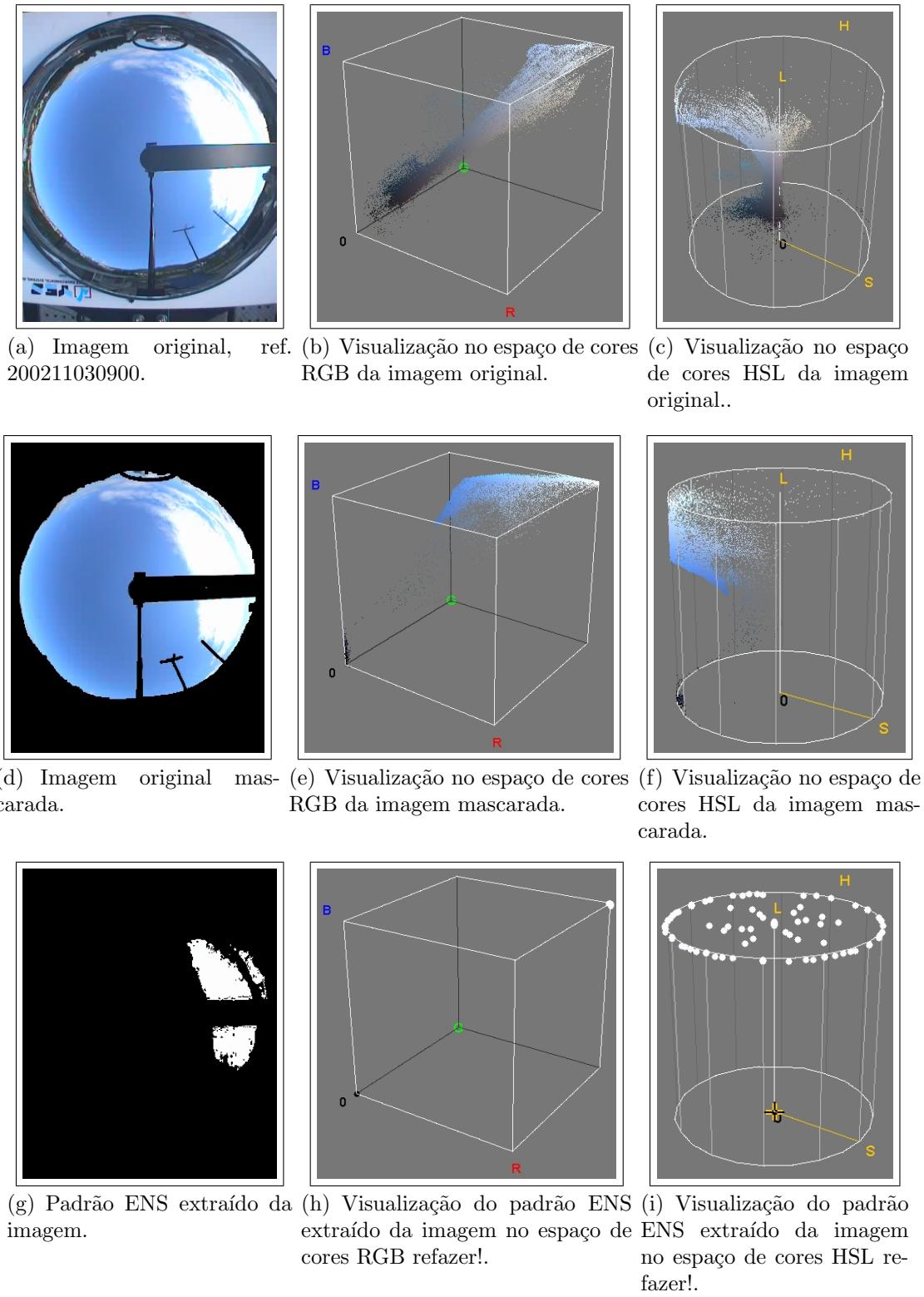


Figure 4.12: Definição do padrão ENS através das imagens e a representação do *locus* típico dos píxeis nos espaços de cores RGB e HSL. Na primeira linha de figuras a imagem original. Na segunda, a imagem subtraída dos fatores que não representam informação útil. Na terceira fileira a imagem indicando somente a presença do espalhamento DENS.

4.1.5.2 Critério de Classificação do padrão DENS

Não foi efetuada a descrição estatística por que os píxeis ocupam um *locus* no final da diagonal escala de cores, bastando para isto uma limiarização simples. Neste ponto, o critério de classificação do padrão DENS é similar ao DRAY, pelas mesmas razões. O padrão ocorre próximo ao fundo de escala do espaço de cores e a partir da saturação total das cores, conforme pode ser observado na figura 4.12. Qualquer pixel com valor de saturação acima de um limiar de escala pode ser classificado neste padrão, a saturação é atingida simultaneamente em todas as dimensões. Neste caso foi utilizada um único limiar dos valores, conforme indicado na equação 4.21.

$$D_{Ray} = (((R \geq 250) \text{AND}(G \geq 250) \text{AND}(B \geq 250)) \text{ AND } (EGD \leq 52,5)) \quad (4.21)$$

4.2 Ajustes Empíricos no Modelo de Representação Matemática dos Padrões Visuais, resumo dos critérios de classificação

Dois ajustes empíricos no modelo matemático foram necessários. O primeiro ajuste foi devido à presença de um conjunto de píxeis intermediários entre os padrões ENS e RAY, que não estava sendo classificado. O segundo foi o estabelecimento uma hierarquia mutuamente exclusiva no critério de classificação.

O primeiro ajuste foi necessário pelas seguintes razões. As imagens selecionadas, caracterizaram os padrões ENS e RAY, de acordo com a teoria proposta no capítulo 3. Entretanto notou-se também a ocorrência de píxeis relativos a um padrão intermediário distinto e entre estes dois padrões, conforme ilustrado nas figura 4.13. Este padrão intermediário compreende píxeis que ocorrem desde a cor azul com intensidade luminosa maior (que poderiam ser classificados como RAY) até a cor acinzentada com intensidade luminosa maior (que poderiam ser classificados como ENS). O exato limiar de discriminação entre os dois padrões será determinado empiricamente, por que o conceito do exato valor de que um píxel deixa de ser nuvem e passa a ser céu claro é altamente subjetivo. Para a determinação exata do limiar de classificação entre os dois padrões de uma maneira não subjetiva poderia ser realizada por exemplo, através da função de discriminação multivariada de fisher (JOHNSON; WICHERN, 2007), Devido à sua complexidade ficará recomendada como sugestão em trabalhos futuros. O ajuste foi executado, adicionando incrementalmente 10 % do valor original do limiar de discriminação. Inicialmente com o padrão ENS e em seguida com o padrão RAY.

O efeito do ajuste progressivo do limiar de discriminação pode ser observado nas imagens da figura 4.14 (a) e (b).

O segundo ajuste nos dados, foi o estabelecimento de uma hierarquia na seqüência de classificação dos padrões (JOHNSON; WICHERN, 2007). Em primeiro lugar é feita a separação dos píxeis de alta intensidade, para em seguida ser classificada a difusão dos padrões DENS DRAY. e o o padrão DENS. Os píxeis correspondentes a estes padrões encontram-se próximos ao fundo de escala simultaneamente nas três dimensões. A separação entre estes dois padrões é feita através da EGD. Em segundo lugar deve ser classificado o padrão MIE. Em terceiro o padrão ENS, em quarto a classificação do padrão RAY. Por último ficará o

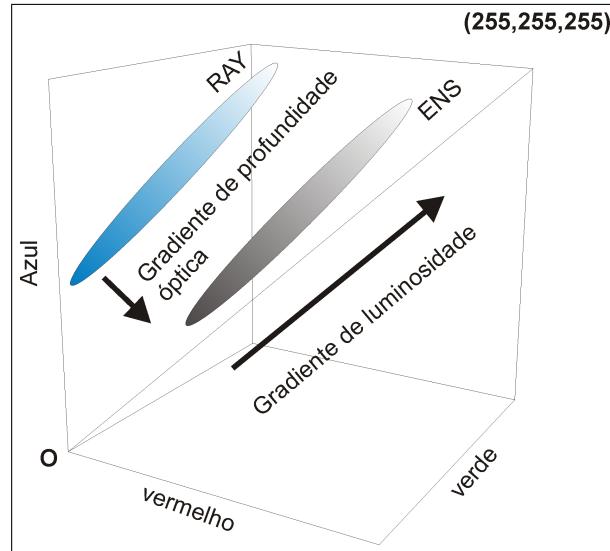
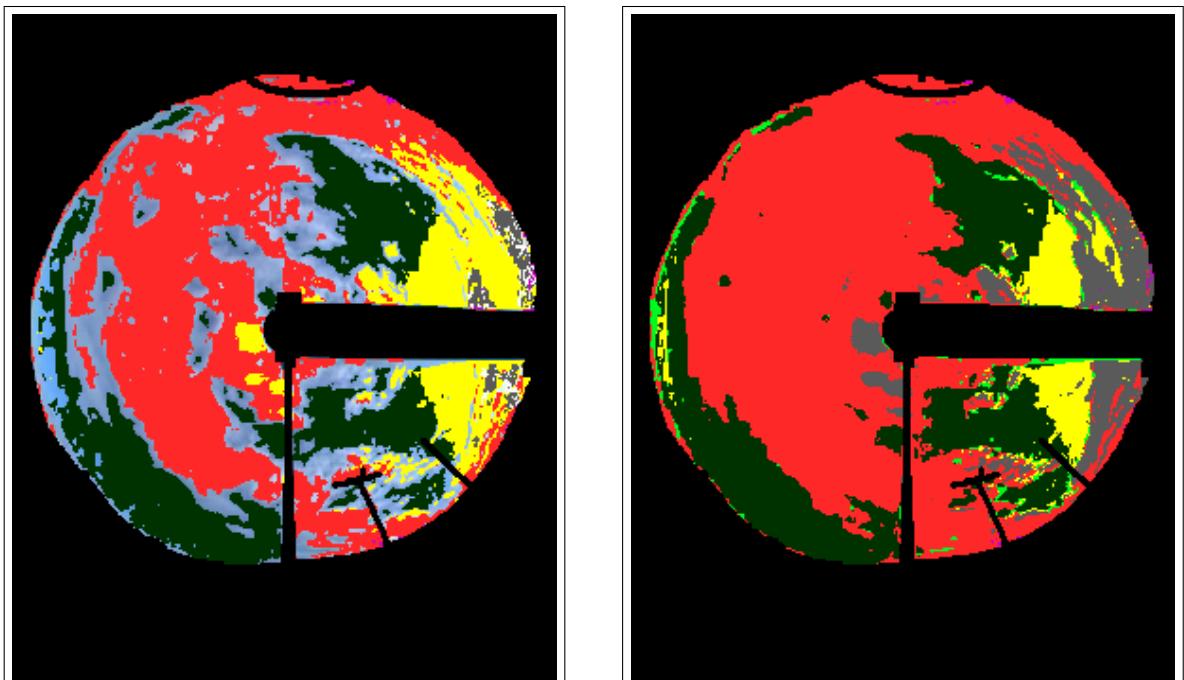


Figure 4.13: Gradiente de profundidade óptica entre os padrões RAY e ENS.



(a) Visualização da imagem para $D_{RAY}^2 \leq 11,34$ e $D_{ENS}^2 \leq 11,34$.

(b) Visualização da imagem para $D_{RAY}^2 \leq 22,68$ e $D_{ENS}^2 \leq 29,01$.

Figure 4.14: Ajuste dos limiares de discriminação para os padrões RAY e ENS.

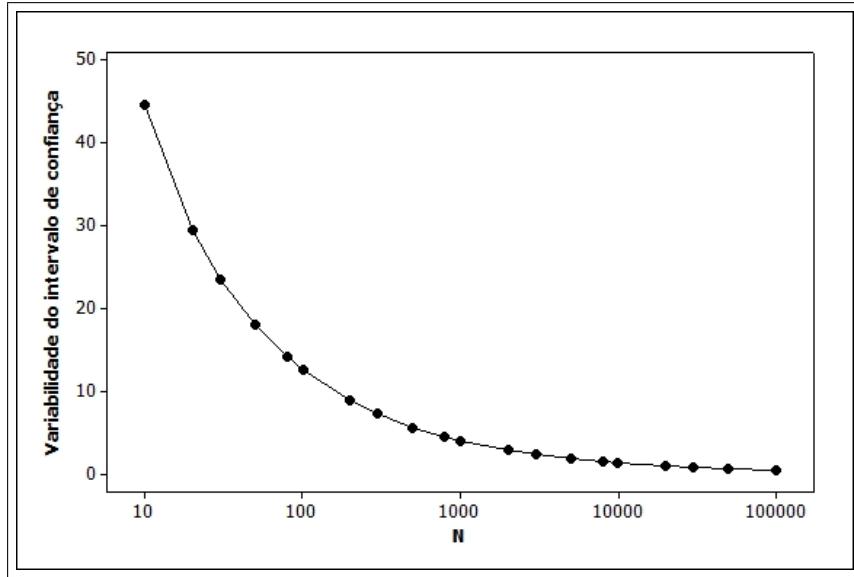


Figure 4.15: Figura indicando a redução da variabilidade do intervalo de confiança de bonferroni, para a construção da dimensão vermelha do padrão ENS, com os seguintes parâmetros: $F_{p,n-p,(\alpha=0,01)} = 6,24$ e $\sigma_i^2 = 28,68$

Table 4.1: Comparação da variabilidade dos intervalos de confiança de bonferroni com a variabilidade do desvio padrão para o padrão ENS.

Critério	R	G	B	Prob.
bonf.	$157,54 \pm 0,089566$	$169,36 \pm 0,090794$	$184,99 \pm 0,094815$	0,99
σ_{ENS}	$157,54 \pm 28,68$	$169,36 \pm 29,45$	$184,99 \pm 32,14$	0,68
$2\sigma_{ENS}$	$157,54 \pm 57,36$	$169,36 \pm 58,9$	$184,99 \pm 64,28$	0,95
$3\sigma_{ENS}$	$157,54 \pm 86,04$	$169,36 \pm 88,35$	$184,99 \pm 96,42$	0,99

padrão NC (Não Classificado) para abrigar os píxeis que não se enquadram em nenhuma das categorias anteriores presentes nas imagens, pela saturação e degeneração de cores.

4.2.1 Limitações no uso do intervalo de confiança de bonferroni no Critério de Classificação dos padrões ENS e RAY

Como alternativa para contornar o problema de ajuste empírico da distância de mahalanobis como limiar de discriminação poderia ser utilizado o conceito de intervalo de confiança de bonferroni indicado na equação 3.10. Entretanto para uma grande amostra da população a variabilidade do intervalo de confiança praticamente desaparece, inviabilizando o seu uso conforme pode ser observado no cálculo dos intervalos de confiança ilustrado na figura 4.15 e na tabela 4.1.

4.2.2 Resumo dos Critérios de Classificação Estabelecidos

Um resumo dos critérios utilizados na classificação dos padrões após os ajustes empíricos, esta apresentado na tabela 4.2.

Table 4.2: Resumo dos métodos de classificação estabelecidos. H é a ordem hierárquica, P é o padrão.

H	P	Critério de classificação
1	DENS	$((R \geq 250) \text{ AND } (G \geq 250) \text{ AND } (B \geq 250)) \text{ AND } (EGD \geq 52,5)$
2	DRAY	$((R \geq 250) \text{ AND } (G \geq 250) \text{ AND } (B \geq 250)) \text{ AND } (EGD \leq 52,5)$
3	MIE	$D_{MIE}^2 \leq F_{p,n-p,(\alpha)} = 11,34$
4	ENS	$D_{ENS}^2 \leq F_{p,n-p,(\alpha)} = 29,01$
5	RAY	$D_{RAY}^2 \leq F_{p,n-p,(\alpha)} = 22,68$
6	NC	Não classificado em nenhum dos casos anteriores

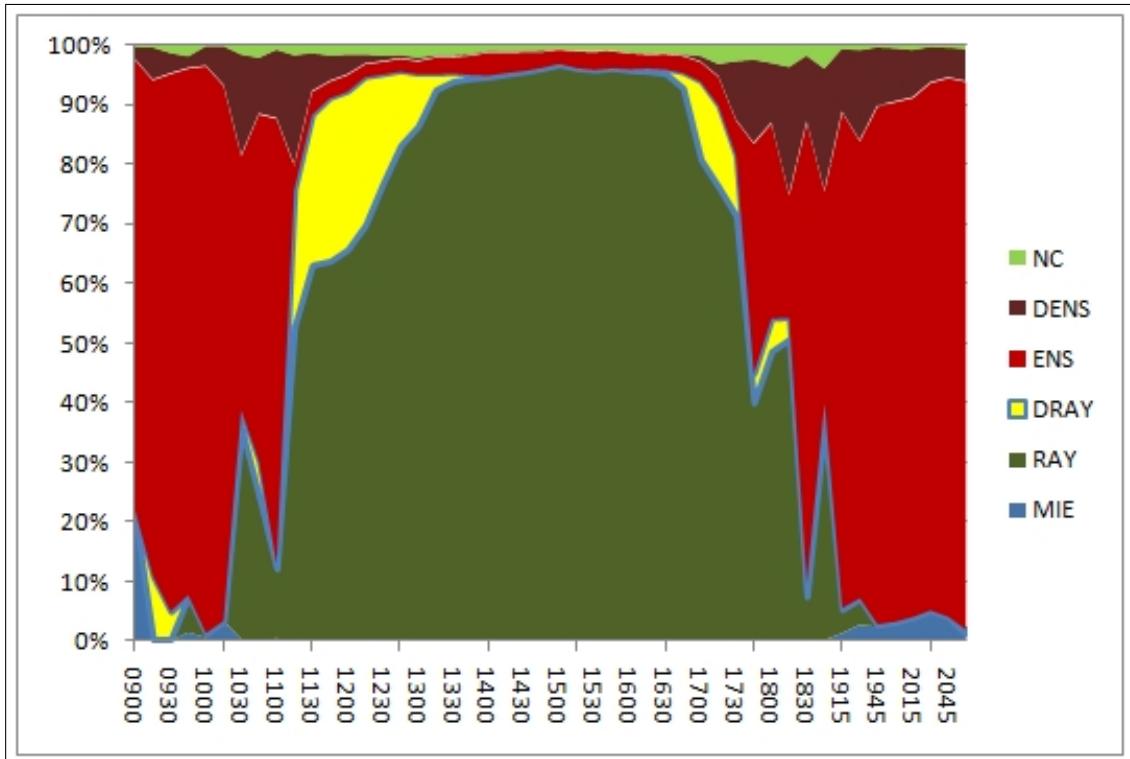


Figure 4.16: Gráfico indicando a composição dos padrões em porcentagem presentes no céu ao longo do dia no eixo vertical para o dia 03 nov. 2002. O eixo horizontal indica o horário GMT

4.3 Síntese dos Resultados Gerais

Os resultados completos da aplicação do modelo proposto nas imagens, encontram-se nas tabelas 4.3, 4.4. A figura 4.16 indica os padrões resumidos em um único gráfico ao longo do dia. A figura 4.17 indica a distribuição dos padrões individualmente separados ao longo do dia. As imagens individualmente geradas com a classificação dos padrões propostos encontram-se no anexo A nas tabelas 5.5 e 5.6.

Table 4.3: Resumo dos resultados gerais.

imagem	MIE	RAY	DRAY	CÉU	ENS	DENS	NUVENS	NC
200211030900	20.76	0	1.25	1.25	75.63	2.14	77.77	0.23
200211030915	0.43	0.01	10.61	10.62	83.2	5.43	88.63	0.32
200211030930	0.08	0.43	4.61	5.04	90.17	3.42	93.59	1.29
200211030945	1.51	6.26	0.09	6.35	88.34	1.95	90.29	1.85
200211031000	0.62	0.82	0.01	0.83	95.12	3.29	98.41	0.13
200211031015	3.47	0	0.04	0.04	89.8	6.57	96.37	0.12
200211031030	0.39	35.36	3.12	38.47	42.65	16.95	59.6	1.54
200211031045	0.18	24.08	6.07	30.14	58.22	9.4	67.61	2.09
200211031100	0.48	11.79	0.85	12.64	74.65	11.51	86.16	0.73
200211031115	0.08	52.79	22.99	75.78	4.02	18.45	22.48	1.68
200211031130	0.06	63.1	25.26	88.36	3.88	6.41	10.28	1.3
200211031145	0.03	63.74	27.44	91.18	2.77	4.31	7.08	1.7
200211031200	0.02	65.7	26.56	92.27	2.81	3.34	6.15	1.56
200211031215	0.02	69.77	24.94	94.71	2.17	1.56	3.73	1.53
200211031230	0	76.82	18.38	95.2	2.14	0.91	3.05	1.75
200211031245	0.03	83.13	12.55	95.68	2.13	0.49	2.62	1.67
200211031300	0.01	86.42	8.84	95.25	2.17	0.54	2.71	2.03
200211031315	0.01	92.51	2.76	95.28	2.76	0.26	3.02	1.69
200211031330	0.01	93.91	1.5	95.41	2.71	0.26	2.96	1.62
200211031345	0.03	94.34	0.78	95.12	3.22	0.23	3.45	1.41
200211031400	0.03	94.62	0.4	95.01	3.87	0.21	4.08	0.88
200211031415	0.04	95.06	0.36	95.42	3.41	0.21	3.62	0.92
200211031430	0	95.52	0.29	95.8	3.16	0.15	3.3	0.89
200211031445	0	96.03	0.3	96.33	2.64	0.24	2.88	0.79
200211031500	0	96.7	0.31	97.01	2.36	0.04	2.39	0.6
200211031515	0	96.07	0.22	96.29	2.76	0.12	2.88	0.83
200211031530	0	95.82	0.19	96.01	3	0.17	3.16	0.83
200211031545	0	96.08	0.19	96.27	2.84	0.1	2.94	0.79
200211031600	0	95.74	0.37	96.11	2.63	0.18	2.81	1.08
200211031615	0	95.58	0.65	96.23	2.3	0.16	2.47	1.3
200211031630	0	95.2	0.85	96.06	2.55	0.1	2.66	1.29
200211031645	0	92.76	2.71	95.46	2.83	0.19	3.03	1.51
200211031700	0.01	80.8	13.17	93.97	3.29	0.95	4.24	1.79
200211031715	0.03	76.18	13.73	89.91	4.83	2.08	6.92	3.14
200211031730	0.05	71.31	10.12	81.43	6.01	9.82	15.83	2.69
200211031745	0.11	39.94	4.97	44.91	38.56	14.02	52.58	2.41

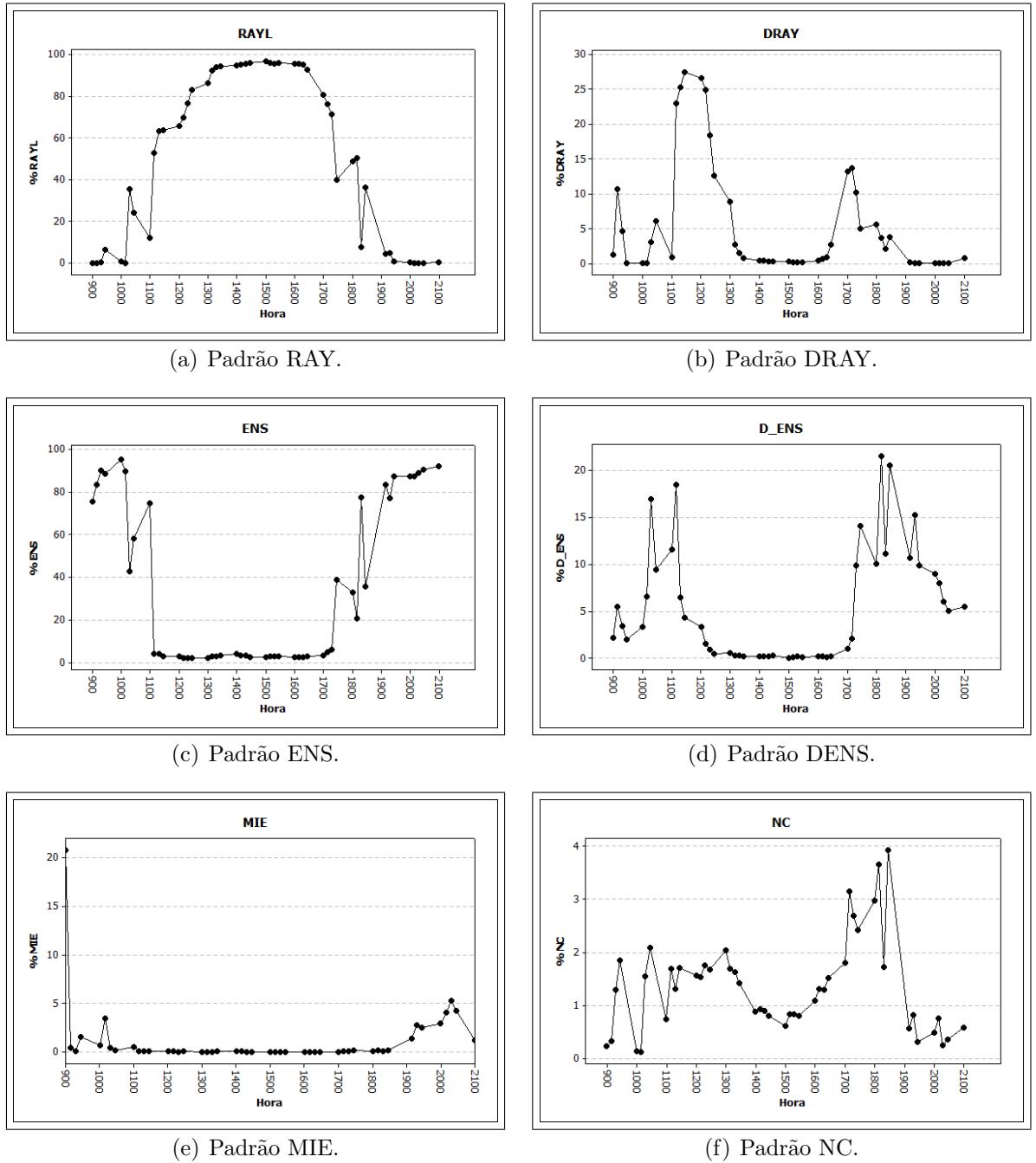


Figure 4.17: Porcentagem dos padrões distribuídos ao longo do dia.

Table 4.4: Resumo dos resultados gerais (cont.).

imagem	RAY	DRAY	CÉU	ENS	DENS	NUVENS	MIE	NC
200211031800	0.06	48.65	5.56	54.21	32.72	10.05	42.76	2.97
200211031815	0.18	50.55	3.64	54.19	20.53	21.44	41.97	3.66
200211031830	0.08	7.55	2.11	9.66	77.43	11.11	88.54	1.72
200211031845	0.18	36.24	3.73	39.97	35.42	20.51	55.92	3.92
200211031915	1.34	4.13	0.21	4.34	83.25	10.61	93.86	0.56
200211031930	2.78	4.55	0.02	4.58	76.97	15.24	92.21	0.82
200211031945	2.45	0.47	0.02	0.5	87.07	9.89	96.96	0.31
200211032000	2.9	0.43	0.07	0.49	87.4	8.99	96.39	0.48
200211032015	4.08	0.01	0.09	0.1	87.27	8.01	95.28	0.75
200211032030	5.27	0	0.02	0.02	88.8	6.03	94.83	0.24
200211032045	4.19	0.02	0.01	0.03	90.58	5.05	95.63	0.36
200211032100	1.14	0.18	0.75	0.93	91.91	5.46	97.37	0.57

Table 4.5: Níveis de luminância natural encontrados em alguns ambientes externos. Extraído de (REINHARD et al., 2006) capítulo 1 tabela 1.1.

Condição	iluminância cd/m^2
Luz das estrelas	10^{-3}
Luz da Lua	10^{-1}
Iluminação interna	10^2
Luz do Sol	10^5
Máxima intensidade dos monitores CRT	10^2

4.4 Análise dos resultados

O modelo apresentado efetuou a classificação e discriminação dos padrões propostos, entretanto alguns pontos importantes devem ser ressaltados. Existe uma dificuldade na análise de cenas com iluminação natural por causa da presença de uma grande ordem de magnitude de valores conforme pode ser observado nos dados apresentados na tabela 4.5. De acordo com estes dados, não é possível para os monitores CRT (Cathod Ray Tube) reproduzir todas as intensidades luminosas presentes nas cenas naturais (KOSLOF, 2006). Embora os monitores LCD, aparentem ser mais brilhantes o seu brilho não consegue alcançar até o presente momento, ordens de magnitude muito acima da tecnologia existente (REINHARD et al., 2006). De uma certa forma toda a tecnologia de obtenção armazenamento e reprodução de imagens desenvolvida recentemente, foi construída sobre esta limitação. Esta restrição denominada de LDR (Low Dynamic Range), norteou todo o processo de aquisição, armazenamento de informações e reprodução das imagens, utilizadas na presente tese. A tecnologia que permite obtenção de imagens em ordens de magnitude de iluminância superior é denominada HDR (High Dynamic Range). A utilização da tecnologia HDR aplicada ao imageamento das nuvens, fica como sugestão de pesquisas para trabalhos futuros.

As câmeras fotográficas conseguem ajustar-se às condições de iluminação das cenas naturais através do controle do tempo de exposição e da abertura do obturador, de maneira

a acomodar as escalas de luminosidade presentes na cena. Mas isto causa uma segmentação na escala de luminosidade. Quando na mesma imagem ocorrem cenas de luminosidade de ordem superior, ocorre também a saturação e a distorção dos valores. A partir das câmeras entretanto, é possível através de uma composição de imagens cobrir toda a faixa de luminosidade contida nas diversas cenas em uma técnica denominada *High Dynamic Range Imaging* ou HDRI (INANCI, 2005) (MOECK; ANAOKAR, 2006), (DEBEVEC; MALIK, 1997).

Entretanto a tecnologia de armazenamento (formato dos arquivos) e reprodução de imagens HDRI ainda não está suficientemente disponível para ser utilizada. Os seres humanos adaptam-se muito mais facilmente a muitas ordens de magnitude de iluminância natural. Sendo assim neste quesito, ainda não é possível com a tecnologia disponível, substituir completamente o observador humano sem restrições.

A limitação das imagens LDR afetou o desenvolvimento do presente trabalho através de distorções nas imagens, quando a ordem de magnitude dos valores monitorados era maior do que 10^2 . Todos os padrões propostos, foram afetados por esta limitação, especialmente os padrões DENS e DRAY por ocorrerem próximos à saturação do domínio de tarefa. Esta saturação causou uma distorção na distribuição espacial dos padrões no espaço de cores que dificultou de maneira geral o modelamento dos fenômenos, tornando-se necessário ajustes empíricos, conforme o apresentado na subseção 4.2. Uma figura esquemática ilustrando o problema pode ser observada na figura 4.18. Nesta figura SB, SG e ST são os pontos de saturação das dimensões azul, verde e triplo. Estes pontos foram utilizados como limiares de discriminação. A linha EGD representa a fronteira de discriminação entre os padrões ENS e RAY do método de Distância Euclideana Geométrica (MANTELLI et al., 2010b). (LONG; SLATER; TOOMAN, 2001) utiliza somente as dimensões de cores vermelha e azul para a classificação dos padrões. A diagonal principal do cubo indica a escala logarítmica que melhor representa as ordens de magnitude de iluminância das cenas naturais.

4.5 Comparação dos resultados com outros métodos encontrados na literatura especializada

Para possibilitar a comparação dos dados, foi necessário efetuar um agrupamento dos padrões para reduzi-los a duas categorias e compará-los com os métodos (LONG; SLATER; TOOMAN, 2001) (LONG) e (MANTELLI et al., 2010b) (EGD), por que se restringem a uma classificação binária. Os padrões classificados no presente trabalho (MAHA) também foram agrupados de maneira que RAY e DRAY representem as condições de céu claro e ENS e DENS representem as condições de céu encoberto. O padrão MIE não foi classificado na categoria céu claro ou encoberto pelo seguinte motivo. O programa residente desenvolvido pelo fabricante do imageador de nuvens, não coleta as imagens com o sol abaixo de 5 graus de ângulo de elevação. Com este critério o programa residente descarta a maioria das imagens que registram este padrão. Desta maneira faltaram elementos nos dados utilizados para definir apropriadamente uma classificação para o padrão MIE. Quando este padrão for detectado nas imagens ele causará um *bias* na avaliação. Fica como sugestão de trabalho futuro abordar este aspecto da pesquisa.

Os dados dos métodos LONG e EGD utilizados na comparação foram obtidos a partir de trabalhos já publicados pelo autor no âmbito da presente pesquisa (MANTELLI et al.,

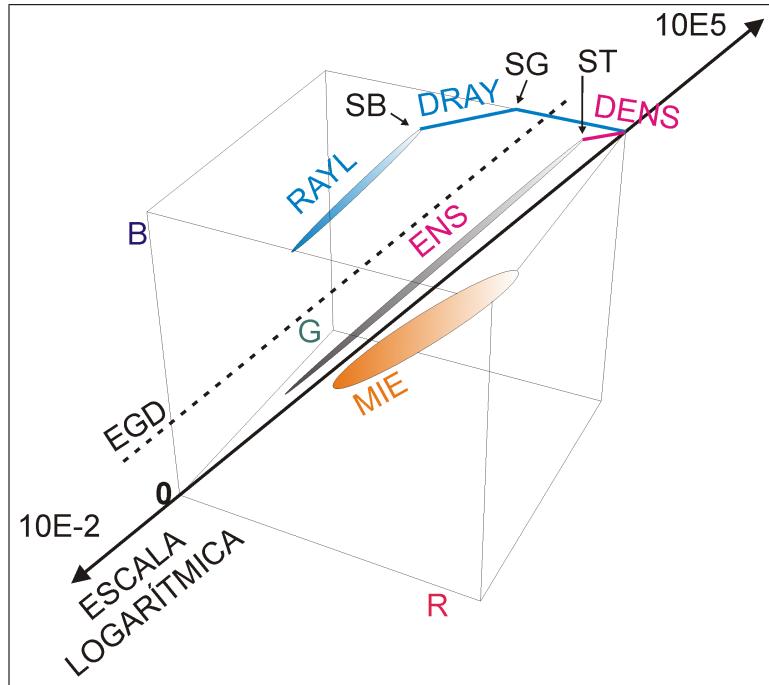


Figure 4.18: Representação esquemática indicando regiões e pontos notáveis dos padrões abordados no presente trabalho no espaço de cores RGB.

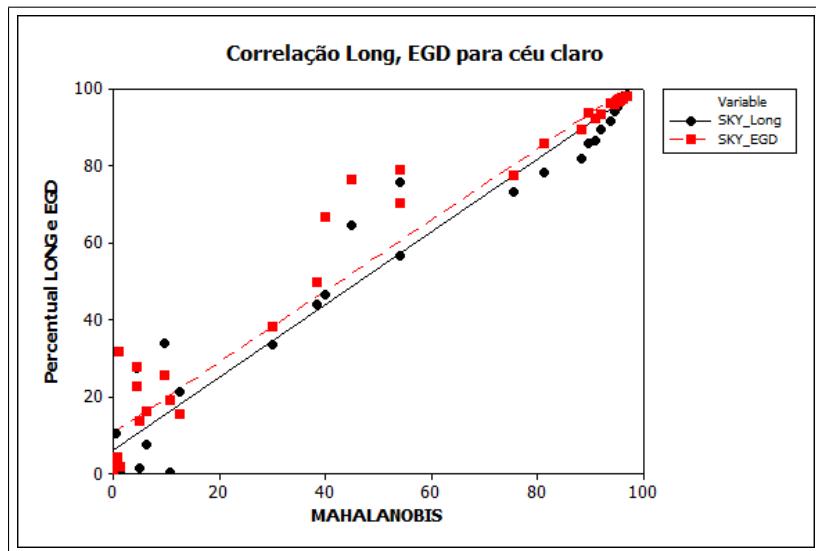


Figure 4.19: Correlação entre LONG, EGD e Mahalanobis para céu claro.

2010b) e reproduzidas no Anexo A, na tabela 5.1. Juntamente com estes dados, encontram-se os padrões agrupados para realizar a avaliação de céu claro e céu encoberto que utiliza a distância de Mahalanobis e as diferenças absolutas entre os métodos para determinação do céu claro e encoberto.

Uma comparação pareada dos resultados de classificação para céu claro pode ser observada através da correlação apresentada na figura 4.19. Os dados numéricos utilizados na comparação dos resultados entre os três métodos estão descritos no anexo A, tabela 5.1. A partir da análise da figura 4.19, pôde-se perceber que os dados indicaram uma forte corre-

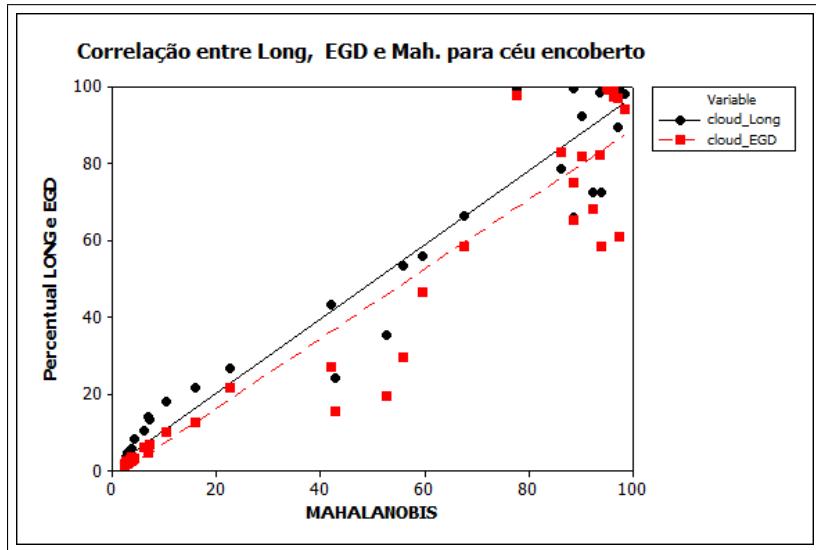


Figure 4.20: Correlação entre LONG, EGD e Mahalanobis para céu encoberto.

lação positiva entre os três métodos com um coeficiente de correlação na ordem de 0,97 para a estimativa de céu claro. Em situações aonde ocorrem condições mistas de céu e nuvem e na presença do padrão MIE na mesma cena, a dispersão é maior. O gráfico da figura 4.19, também apresentou um *offset* ou *bias* em favor do método LONG e um pouco mais além para o método EGD na avaliação de baixas porcentagens céu claro. Indicando que EGD e LONG avaliam o céu claro em uma quantidade ligeiramente acima do MAHA. Esta diferença se justifica pelo fato do MAHA classificar outros padrões presentes no domínio além do céu e da nuvem, que não são classificados como céu claro. Outro fator que influencia nesta análise é a degeneração progressiva do padrão por saturação das cores originais conforme destacado na figura 4.11. Esta distorção é mais comum, quando ocorrem píxeis de alta intensidade especialmente nas condições de céu claro e na proximidade do círculo solar. O critério de classificação MAHA remove da análise estes píxeis degenerados, por que eles não mantém o mesmo atributo de cor durante a distorção. Por último pode-se concluir também que quando o céu se aproxima na condição 100 % de céu claro, a dispersão diminui e as diferenças entre os 3 métodos se reduzem ligeiramente.

Uma outra comparação pareada dos resultados de classificação para céu encoberto precisou ser analisada, por que o método MAHA classifica mais de dois padrões e a análise não pode ser assumida como binomial complementar. A correlação para céu encoberto pode ser observada através do gráfico apresentado na figura 4.20. Os dados numéricos utilizados na comparação dos resultados entre os três métodos, também estão descritos no anexo A, tabela 5.1. A partir da análise da figura, pôde-se perceber que os dados indicaram uma forte correlação positiva entre os três métodos com um coeficiente da ordem de 0,98 entre LONG e MAHA e 0,96 entre EGD e MAHA para a estimativa de céu encoberto. O gráfico da figura 4.20, também apresentou um *offset* ou *bias* negativo em relação ao método EGD. Não foi notada a existência de um *bias* em relação ao método LONG na avaliação de céu encoberto. Em situações aonde ocorrem condições mistas de céu e nuvem e na presença do padrão MIE na mesma cena, a dispersão também foi maior, pelas mesmas razões indicadas na avaliação do céu claro. Por último pode-se concluir também que quando o céu se aproxima na condição 100 % de céu encoberto, a dispersão aumenta e as diferenças de avaliação entre os 3 métodos se acentuam mais do que a avaliação complementar para céu claro. Esta diferença, como no

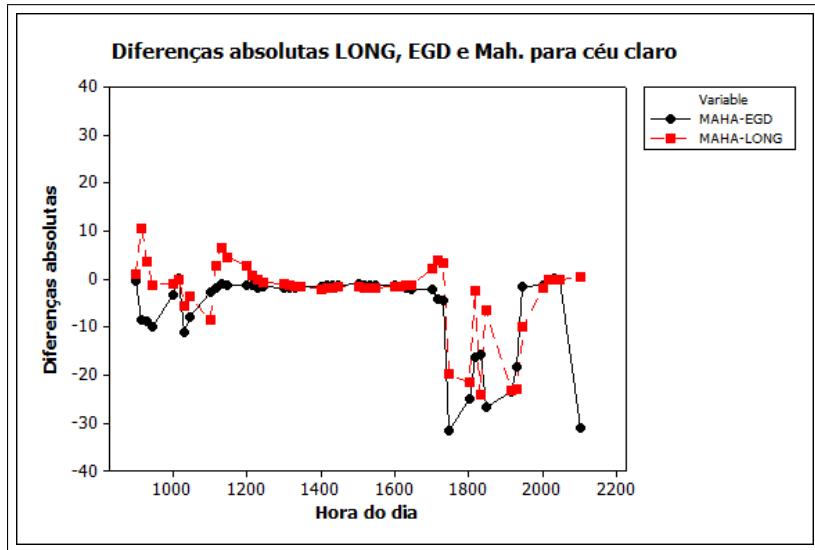


Figure 4.21: Diferenças absolutas LONG (10,42 a -24,13) , EGD(+ 0,03 a -31,63) e Mahalanobis para céu claro.

caso de céu claro, também se justifica pelo fato do MAHA classificar outros padrões presentes no domínio além do céu e da nuvem, que não são classificados como céu claro.

Além da correlação entre os métodos também foram analisadas as diferenças percentuais absolutas entre LONG, EGD e MAHA para a avaliação da cobertura do céu claro conforme ilustrado no gráfico da figura 4.21 e encoberto no gráfico da figura 4.22.

Uma análise das diferenças no resultado de avaliação da figura 4.21 entre os dois métodos para a quantidade de céu claro indicou o seguinte. Valores de diferenças percentuais em relação ao MAHA de no mínimo de -31,63 % e um máximo da ordem 0,03 % em relação ao método EGD. Os valores das diferenças percentuais em relação ao MAHA de no mínimo -24,13 % e um máximo de 10,42 % em relação ao método LONG. A partir do gráfico pode-se notar que parte destas diferenças são acentuadas nas condições mistas de céu e nuvem e maiores na presença de nuvens finas. Para o céu totalmente claro (das 12:15 às 16:45) as diferenças se mostram abaixo de 5 % entre os 3 métodos.

Uma análise das diferenças no resultado de avaliação da figura 4.22 entre os dois métodos para a quantidade de céu encoberto indicou o seguinte. Valores de diferenças percentuais em relação ao MAHA de no mínimo de -20,01 % e um máximo da ordem 36,39 % para o céu encoberto em relação ao método EGD. Os valores das diferenças percentuais em relação ao MAHA de no mínimo -22,04 % e um máximo de 22,33 % em relação ao método LONG. As diferenças de avaliação foram em geral ligeiramente mais discrepantes com a estimativa do EGD do que o método LONG para céu encoberto. A maior diferença negativa foi devido à presença do espalhamento MIE. Na presença do espalhamento MIE, os métodos LONG e EGD classificaram este padrão como céu encoberto. Acredita-se que com esta característica, ambos os métodos também classifiquem como nuvem, o padrão MIE correspondente ao céu claro devido à limitações do critério de classificação binária.

Como comentário geral pode-se afirmar que parte das diferenças existentes entre os três métodos, é devido à classificação arbitrária que os métodos de classificação binária (ou binomial) desempenham quando encontram um terceiro padrão. Outra diferença a ser no-

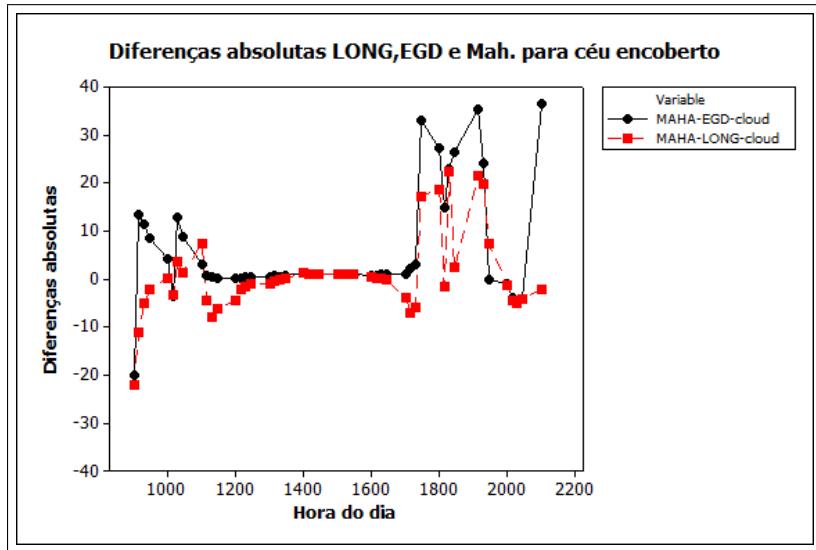


Figure 4.22: Diferenças absolutas LONG (10,42 a - 24,13) , EGD(+ 0,03 a -31,63) e Mahalanobis para céu claro.

tada é que os métodos EGD e MAHA trabalham com a imagem no espaço tridimensional, permitindo uma maior especificidade dos elementos contidos na imagem. O método desenvolvido por LONG descarta uma das dimensões do espaço de cores, enquanto que o EGD analisa as três dimensões. Outro ponto importante é que as maiores diferenças ocorrem na presença de nuvens finas, devido à variações no limiar de discriminação entre os três métodos. O ponto exato de discriminação quando um píxel deixa de representar o padrão RAY e passa a representar o padrão ENS precisa ser melhor estudado e fica como sugestão para trabalhos futuros.

4.6 Validação dos dados utilizando imagens de Satélite

Além da comparação do método proposto com técnicas similares, também está prevista a validação do método proposto utilizando imagens de satélite. Como as imagens estão dentro do espectro visível, optou-se por utilizar o canal visível do GOES-8. As imagens foram obtidas a partir do site da agência americana NOAA (National Oceanographics and Atmospheric Administration). Inicialmente as imagens foram separadas para as coordenadas 47 a 49 graus de longitude Oeste e 26 a 28 graus de latitude Sul. As coordenadas geográficas do local aonde foram adquiridas as imagens de superfície (LABSOLAR) são as seguintes (27 Graus 36 minutos 6,1 segundos SUL e 48 Graus 31 minutos 4,2 segundos OESTE). As imagens obtidas para a validação do trabalho estão ilustradas nas figuras 5.8, 5.9, 5.10, 5.11, 5.12 do anexo A. As imagens de satélite disponíveis, não foram obtidas no exato instante de obtenção do imageador de nuvens na superfície. De maneira que parte das discrepâncias pode ser justificada pela diferença temporal na aquisição da imagem. Nas figuras do anexo, as imagens de satélite aparecem entre as imagens obtidas no intervalo anterior e posterior da imagem do imageador de nuvens. Infelizmente as imagens de satélite disponíveis já vieram previamente marcadas com o contorno do continente e da ilha de Santa Catarina, interferindo na qualidade da análise.

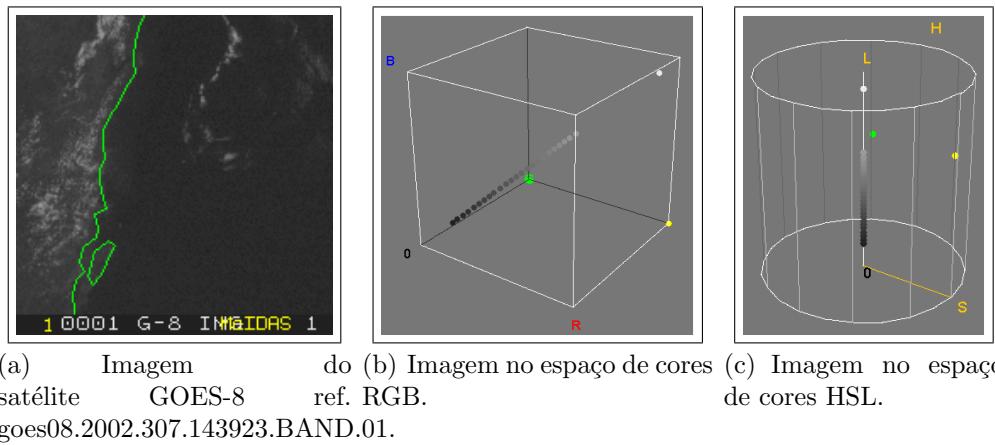


Figure 4.23: Distribuição dos píxeis da imagem de satélite nos espaços de cores RGB e HSL.

Table 4.6: Matrix média de Intensidade para 6 imagens, dos píxeis ao redor da estação de Florianópolis, caracterizando o céu claro.

-	150	151	152
44	61,5	64	64,833
45	60,667	63,167	63,167
46	57,333	57,333	60,667

As imagens de resolução 187 x 189 píxeis, eram relativas a um quadrado de 200 x 200 km. A título de ilustração, uma das imagens utilizadas como referência para céu claro, foi analisada na ferramenta ColorInspector3D nos espaços de cores RGB e HSL, encontrase na figura 4.23. Nesta imagem pode-se perceber em verde o padrão correspondente ao entorno da ilha e do continente, os padrões correspondentes à legenda da imagem e os dados.

A partir das imagens de satélite, foram separados grupos de 9 píxeis em uma matriz 3 x 3, cujo píxel central era correspondente à coordenadas geográficas da estação de Florianópolis, aonde foi instalado o imageador. O pixel central correspondente à estação nas coordenadas da imagem (45, 151) foi convencionado como C e os circundantes adjacentes foram associados aos pontos cardinais N, S, E, O, NO, NE, SO, SE conforme ilustrado na tabela 4.9.

As seguintes seis imagens de satélite foram definidas como padrão de céu claro: 13:09, 13:39, 14:09, 15:39, 16:09 e 16:39, a partir dos seguintes critérios. A coincidência do céu claro do imageador, e a indicação simultânea dos métodos MAHA, EGD e LONG de pelo menos 95% de céu claro. A partir dos valores dos píxeis extraídos destas seis imagens foi definida a matrix média de céu claro para a estação, conforme pode ser verificado na tabela 4.6.

A estimativa da repetitividade (ALBERTAZZI; SOUZA, 2008) da estimativa de céu claro para os píxeis correspondentes estão descritos segundo a equação 4.22. Os resultados do cálculo da repetitividade estão descritos na tabela 4.7.

$$R_e = t \ u = t \sqrt{\frac{\sum_{i=1}^n (I_i - \bar{I})^2}{n-1}} = t \sqrt{\frac{\sum_{i=1}^6 (I_i - \bar{I})^2}{6-1}} \quad (4.22)$$

Table 4.7: Matrix de repetitividade da estimativa da média de Intensidade de céu claro para 6 imagens, dos píxeis ao redor da estação de Florianópolis no ponto 151 x 45.

-	150	151	152
44	2,739	8,367	7,360
45	6,831	4,916	7,360
46	4,082	4,472	5,164

Table 4.8: Matrix de valores limiares de céu claro ja arredondados para inteiros, para teste de hipótese unilateral dos píxeis ao redor da estação de Florianópolis no ponto 151 x 45.

-	150	151	152
44	64	72	72
45	68	68	70
46	61	62	66

onde:

- R_e : é a repetitividade que define a faixa dentro da qual espera-se que o erro aleatório da estimativa esteja contido, para o pixel de céu claro.
- t : é o de "student" 1,943 para 5 graus de liberdade e nível de confiança $\alpha = 0,05$, teste unilateral.
- n : número medições repetitivas efetuadas dos píxeis utilizados para determinação do céu claro.
- I_i : é a i-ésima indicação
- \bar{I} : é a média das "n" indicações
- u : é a incerteza padrão, corresponde à estimativa do desvio padrão da distribuição dos erros de medição.
- $n-1$: é o número de graus de liberdade

Limiar de discriminação do valor do pixel a partir das imagens de satélite foi estabelecido da seguinte maneira. Foi considerado como céu encoberto as diferenças dos valores maiores do que o apresentado na tabela 4.8.

Depois de definidos os critérios de classificação e os limiares de discriminação foram extraídos das imagens de satélite disponíveis os píxeis a serem analisados, conforme ilustrado na tabela 4.10.

Como as imagens de satélite estavam entre duas estimativas de céu claro e encoberto. Foi utilizado o valor médio entre estas duas estimativas para os três métodos MAHA, EGD e LONG, para ser comparada com a estimativa da imagem de satélite ou SAT. Os resultados deste passo intermediário pode ser observado na tabela 4.11.

Os resultados numéricos da estimativa de cobertura de nuvens para todas as imagens de satélite pode ser observado na tabela 4.12.

Table 4.9: Matrix de valores limiares de céu claro LIM já arredondados para inteiros, para teste de hipótese unilateral dos píxeis ao redor da estação de fpolis. X e Y são as coordenadas dos píxeis na imagem, MED é a média dos valores dos píxeis para céu claro, STD o desvio padrão, POS é a posição do pixel em relação ao píxel central C no ponto 151 x 45. Os pontos foram mapeados nas posições

X	Y	MED.	STD	POS	LIM
44	150	61.5	2.739	NO	64
45	150	60.667	6.831	O	68
46	150	57.333	4.082	SO	61
44	151	64	8.367	N	72
45	151	63.167	4.916	C	68
46	151	57.333	4.082	S	62
44	152	64.833	7.36	NE	72
45	152	63.167	7.36	E	70
46	152	60.667	5.164	SE	66

A partir dos dados numéricos puderam ser obtidos gráficos para efetuar uma comparação dos métodos na figura 4.24, das diferenças absolutas entre os métodos conforme ilustrado na figura 4.25 e na correlação entre eles conforme ilustrado na figura 4.26 e resumido na tabela 4.13.

Analizando os resultados da validação utilizada pode-se concluir o seguinte. A correlação entre os 3 métodos e a validação de satélite não foi muito acentuada, conforme pode ser observada nos resultados da tabela 4.13. Parte das diferenças pode ser justificada pela diferença de tempo existente entre a aquisição da imagens de satélite e a aquisição da imagem do imageador de nuvens. O imageador somente obteve imagens nove minutos antes e seis minutos depois da imagem de satélite. A partir da análise das comparações das figuras 4.24, 4.26 e 4.26 acredita-se que esta diferença deva se acentuar à medida que a velocidade do deslocamento das nuvens aumente entre os intervalos de avaliação.

Table 4.10: Valores das imagens de satélite relativo à estação de Florianópolis

pos	1009	1039	1109	1139	1309	1339	1409	1439	1539	1609	1639	1709	1739	1909	1939	2009	2039
NO	80	54	48	54	59	59	64	59	64	64	59	64	74	59	74	69	54
O	86	69	59	54	54	59	59	59	59	69	69	64	74	64	69	64	48
SO	86	86	80	59	54	54	54	54	59	59	64	59	74	64	69	64	54
N	80	64	54	54	54	69	64	64	74	69	69	74	64	74	64	54	54
C	86	69	43	54	59	59	59	59	64	69	69	69	74	64	69	59	54
S	80	64	80	54	54	54	54	54	59	59	64	69	80	64	64	69	54
NE	74	64	48	54	59	64	54	69	69	69	74	74	74	64	74	64	54
E	80	54	43	54	59	59	54	69	74	64	69	74	74	64	69	64	59
SE	74	54	64	54	54	64	54	64	64	64	69	74	64	69	59	54	54

Table 4.11: Resultados da comparação dos resultados da validação.

hora	sat	MAHA-A	MAHA-D	MAHA-M	EGD-A	EGD-D	EGD-M	LONG-A	LONG-D	LONG-M	S-M	S-L	S-E
1009	100	98.4	96.4	97.4	94.2	99.9	97.05	98.2	99.7	98.95	2.6	1.05	2.95
1039	44.4	59.6	67.6	63.6	46.6	58.7	52.65	55.9	66.3	61.1	-19.2	-16.7	-8.25
1109	22.2	86.2	22.5	54.35	83.1	21.7	52.4	78.8	26.8	52.8	-32.15	-30.6	-30.2
1139	0	10.3	7	8.65	9.9	6.9	8.4	18.2	13.4	15.8	-8.65	-15.8	-8.4
1309	0	2.7	3	2.85	2.4	2.3	2.35	3.7	3.4	3.55	-2.85	-3.55	-2.35
1339	0	3	3.4	3.2	2.4	2.7	2.55	3.1	3.2	3.15	-3.2	-3.15	-2.55
1409	0	4.1	3.6	3.85	2.7	2.7	2.7	2.8	2.6	2.7	-3.85	-2.7	-2.7
1439	0	3.3	2.9	3.1	2.8	1.9	2.35	2.2	1.9	2.05	-3.1	-2.05	-2.35
1539	11.1	3.16	2.9	3.03	2.1	1.84	1.97	2.1	1.8	1.95	8.07	9.15	9.13
1609	33.3	2.8	2.5	2.65	2	1.6	1.8	2.3	2.3	2.3	30.65	31	31.5
1639	55.5	2.6	3	2.8	1.7	2	1.85	2.5	3.2	2.85	52.7	52.65	53.65
1709	55.5	4.2	6.9	5.55	3.15	4.9	4.025	8.13	14.2	11.165	49.95	44.335	51.475
1739	100	15.8	52.6	34.2	12.7	19.6	16.15	21.8	35.3	28.55	65.8	71.45	83.85
1909	22.2	92.2	93.9	93.05	68.2	58.6	63.4	72.4	72.4	72.4	-70.85	-50.2	-41.2
1939	88.9	92.2	96.7	94.45	68.2	97	82.6	72.4	89.6	81	-5.55	7.9	6.3
2009	33.3	96.4	95.3	95.85	97.4	99.3	98.35	97.7	99.7	98.7	-62.55	-65.4	-65.05
2039	100	94.8	95.6	95.2	99.9	99.8	99.85	99.9	99.7	99.8	4.8	0.2	0.15

Table 4.12: Valores da cobertura das nuvens a partir das imagens de satélite

pos	1009	1039	1109	1139	1309	1339	1409	1439	1539	1609	1639	1709	1739	1909	1939	2009	2039
NO	16	-10	-16	-10	-5	-5	0	-5	0	0	-5	0	10	-5	10	5	-10
O	18	1	-9	-14	-14	-14	-9	-9	-9	1	1	-4	6	-4	1	-4	-20
SO	25	25	19	-2	-7	-7	-7	-7	-2	-2	3	-2	13	3	8	3	-7
N	8	-8	-18	-18	-18	-18	-3	-8	-8	2v-3	-3	2	-8	2	-8	-18	
C	18	1	-25	-14	-9	-9	-9	-9	-9	-4	1	1	6	-4	1	-9	-14
S	18	2	18	-8	-8	-8	-8	-8	-3	-3	2	7	18	2	2	7	-8
NE	2	-8	-24	-18	-13	-8	-18	-18	-3	-3	-3	2	2	2	-8	2	-8
E	10	-16	-27	-16	-11	-11	-16	-1	4	-6	-1	4	4	4	-6	-1	-11
SE	8	-12	-2	-12	-12	-12	-2	-12	-2	-2	-2	-2	3 8	-2	3v-7	-12	
COB	100	44.4	22.2	0	0	0	0	11.1	33.3	55.5	55.5	100	22.2	88.9	33.3	100	

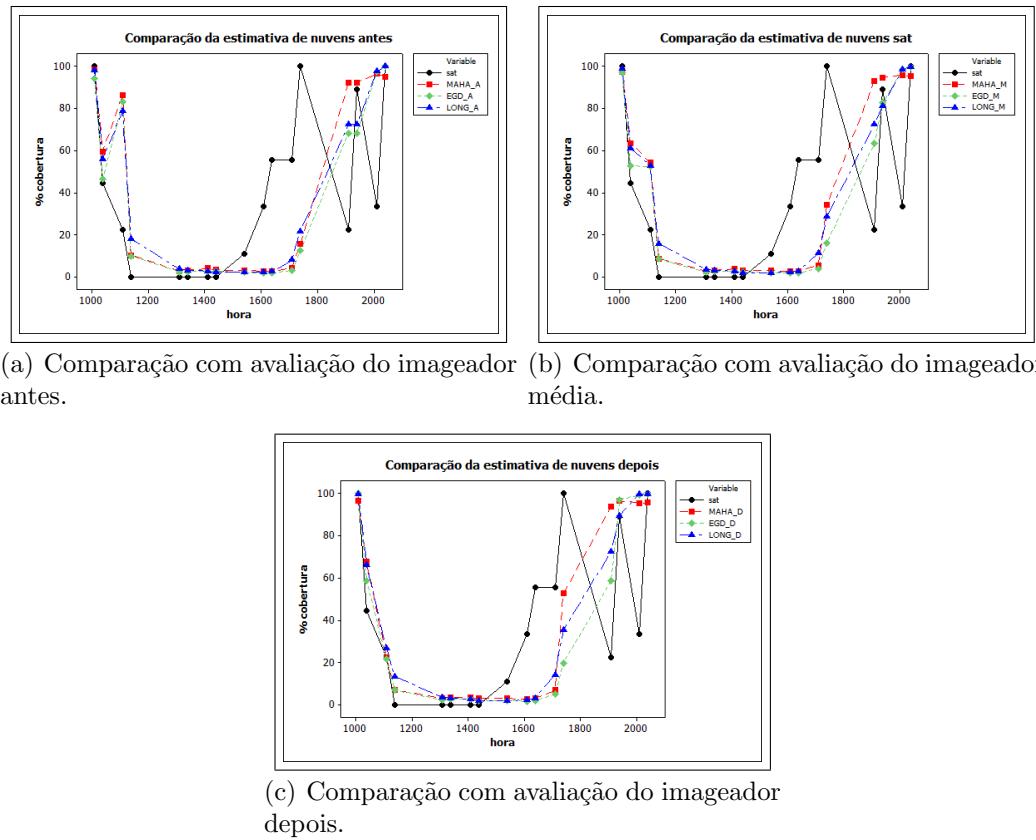
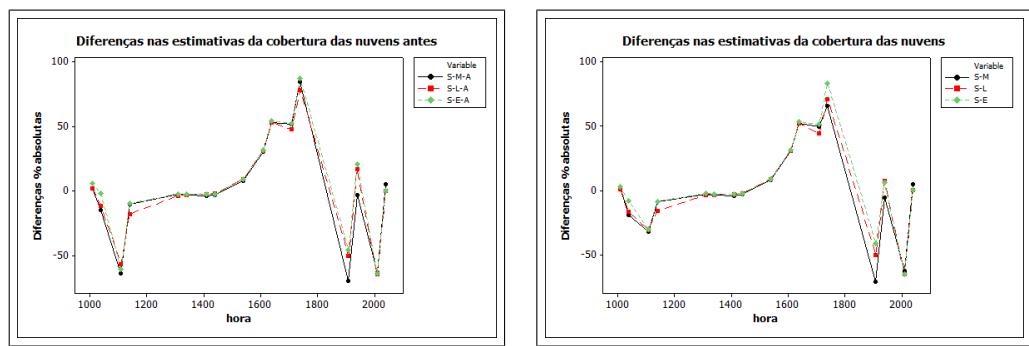


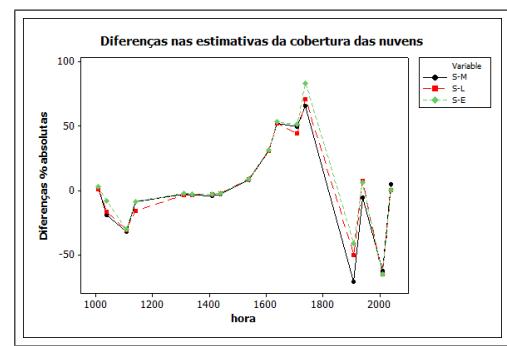
Figure 4.24: Comparação da estimativa da cobertura de nuvens LONG , EGD , MAHA e SAT para céu claro.

Table 4.13: Coeficientes de correlação os métodos de estimativa da cobertura de nuvens e o método de satélite.

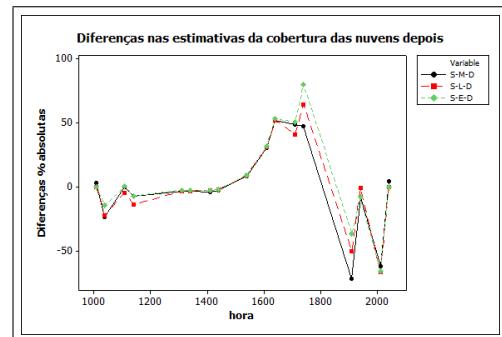
método	coef	p
malanobis média	0.588	0.013
egd média	0.581	0.014
long média	0.603	0.010



(a) Comparação com avaliação do imageador antes.



(b) Comparação com avaliação do imageador média.



(c) Comparação com avaliação do imageador depois.

Figure 4.25: Diferenças absolutas entre LONG , EGD , MAHA e SAT para céu claro.

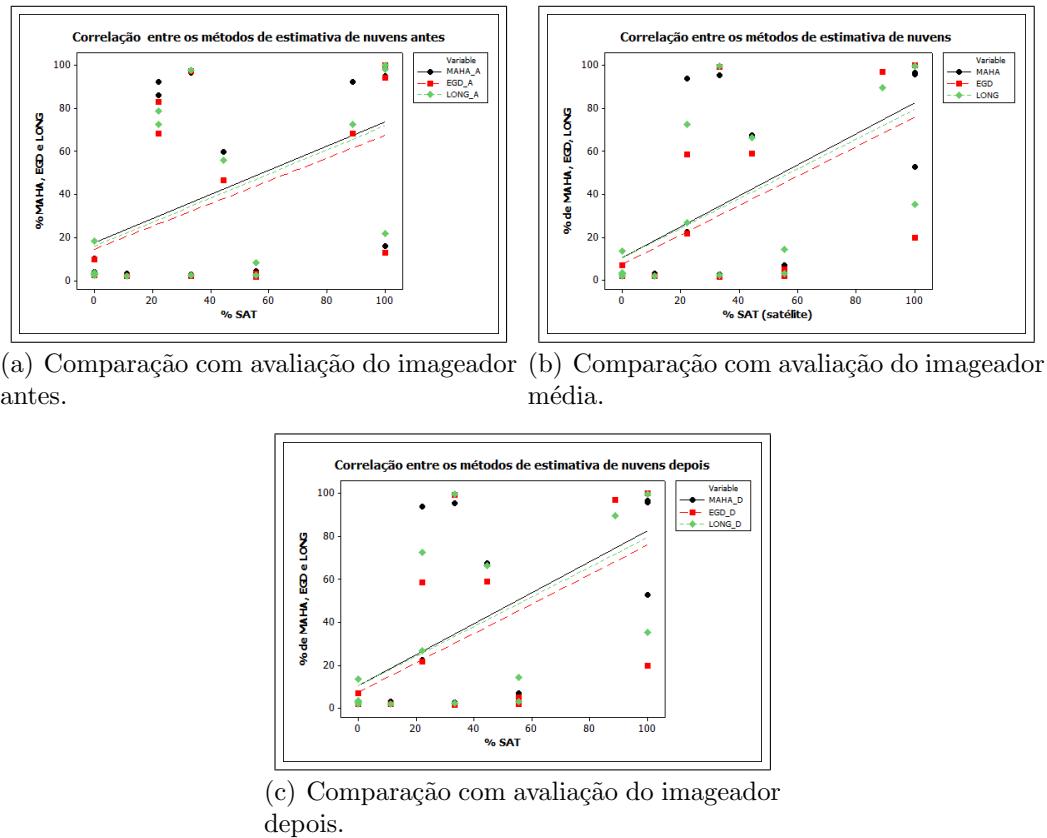


Figure 4.26: Correlação entre todos os métodos e o método de estimativa de cobertura de nuvens de satélite.

Chapter 5

Conclusões e sugestões para trabalhos futuros

5.1 Conclusões

O presente trabalho apresentou ao longo de seu curso, o desenvolvimento de uma metodologia original e alternativa para representar matematicamente os padrões atmosféricos visuais observados obtidos através câmeras, de maneira a substituir os observadores sinópticos no processo de validação das imagens de satélite a partir da superfície. O capítulo 2 apresentou o levantamento bibliográfico acompanhado de uma análise crítica, ressaltando as deficiências e limitações acerca das principais técnicas disponíveis existentes na literatura especializada utilizando instrumentos para a automatização deste processo visando reduzir a subjetividade envolvida no processo de observação. O capítulo 3 apresentou a metodologia desenvolvida de maneira a modelar o processo de mapeamento dos dados representados pelos valores dos pixels existentes no domínio de observação das imagens nos atributos de cores em conhecimento. A metodologia bayesiana (modelamento supervisionado e análise) através de métodos estatísticos multivariados foi utilizada para cumprir os objetivos em função da alta susceptibilidade humana às irregularidades estatística existentes no seu ambiente. O capítulo 4 apresenta os resultados obtidos a partir da aplicação das técnicas desenvolvidas em um conjunto de dados disponível.

Como resultado da presente pesquisa, também foram produzidos trabalhos científicos refinando as técnicas propostas. Dentre elas seria importante mencionar a comparação da estimativa de nuvens entre os métodos de piranômetros e as imagens, o desenvolvimento da técnica EGD, entre outras. As publicações produzidas no decorrer da presente pesquisa que são requisitos indispensáveis do programa de Pós Graduação do EGC estão descritos no Apêndice A.

A partir dos análise dos resultados pode-se observar que os métodos utilizados, apesar de muito próximos em resultados quantitativos, apresentam uma análise qualitativa mais refinada, que a binária predominante na comunidade científica. A abordagem dicotômica utilizada pelos principais métodos disponíveis, leva a perda de informação e aumento das incertezas na interpretação dos fenômenos observáveis. O método desenvolvido permitiu a obter mais informações e um conhecimento mais refinado a partir do mesmo conjunto de dados, do mesmo fenômeno físico observado. Isto pode ser demonstrado pela análise compar-

ativa, redução de dimensionamento e mapeamento dos novos padrões propostos nos padrões antigos correspondentes. A análise não foi mais refinada, devido às limitações na obtenção, armazenamento e reprodução das imagens. A tecnologia de baixo alcance dinâmico ou LDR vigente limita a análise. Estas limitações causavam saturação progressiva das dimensões e distorção de cores nos valores dos píxeis próximos ao fundo de escala. Este problema somente será solucionado com as novas tecnologias de alto alcance das imagens HD em introdução no mercado.

As diferenças entre o método proposto e os métodos existentes, foram devido à classificação arbitrária que os métodos de classificação binária (ou binomial) desempenham quando encontram um terceiro padrão. Outra diferença a ser notada é que os métodos EGD e MAHA trabalham com a imagem no espaço tridimensional, permitindo uma maior especificidade dos elementos contidos na imagem. O método desenvolvido por LONG descarta uma das dimensões do espaço de cores, enquanto que o EGD analisa as três dimensões.

As diferenças notadas entre a validação dos métodos de superfície com as imagens de satélite, foram causadas pelas diferenças temporais de aquisição das imagens (9 e 6 minutos). Estas diferenças se acentuaram significativamente entre as imagens de satélite e as superfície especialmente em cenas de rápidas mudanças na cobertura de nuvens ao longo do tempo. Este problema pode ser minimizado se as imagens de satélite e as de superfície forem obtidas simultaneamente.

Conforme demonstrado ao longo do trabalho, a dicotomia dos resultados utilizados pela maioria dos métodos existentes, dificulta a análise dos resultados induzindo o sistema automático a erros de interpretação. Por este motivo um novo paradigma de interpretação do conhecimento envolvido no processo foi utilizado baseado na teoria da física atmosférica ao invés da interpretação dicotômica dos padrões. No decorrer do presente trabalho, observou-se que a análise qualitativa do domínio de observação dos operadores sinópticos é muito melhor que o do agente, devido ao conhecimento tácito na análise de um ambiente tridimensional. Por outro lado a análise quantitativa do agente é muito mais precisa, desde que as metodologias de reconhecimento dos padrões visuais sejam apropriadamente interpretadas e possam ser mapeadas automaticamente pelo sistema. O presente trabalho ainda está aquém da substituição total do observador humano por sistemas automáticos com os recursos e técnicas disponíveis. Mas apresentou um avanço significativo nas técnicas de análise, interpretação e modelamento do domínio, reduzindo as incertezas e limitações inerentes ao processo em relação às metodologia trivial e unidisciplinar utilizadas até o presente momento pela comunidade científica envolvida.

5.2 Sugestões de Trabalhos Futuros

Os métodos e ferramentas desenvolvidas no presente trabalho possibilitaram o desenvolvimento de uma nova técnica de análise em área aplicada. Entretanto outros pontos precisam ser abordados de maneira a refinar o trabalho desenvolvido, e contornar algumas dificuldades técnicas encontradas. Apesar de representar um avanço, o uso da metodologia Bayesiana com análise multivariada pode ser expandida ou combinada a mais de um espaço vetorial. A análise multivariada também pode ser investigada a partir de outras variáveis que podem contribuir com o refinamento do fenômeno como os dados oriundos de modelos físicos ou de outros sensores como por exemplo os dados da varredura de um fotômetro solar. Outra linha de investigação poderia comparar o levantamento desenvolvido para no espectro

visível com outros sensores que atual na faixas do espectro infra vermelho.

Uma das fontes de incerteza dos métodos estudados está determinação do ponto exato de discriminação entre os dois padrões (céu e nuvem) para determinar quando um píxel deixa de representar uma nuvem e passa a representar o céu claro. (LONG; SLATER; TOOMAN, 2001) utiliza um critério empírico de discriminação, enquanto que (SOUZA-ECHER et al., 2006) e (MANTELLI et al., 2010b) utilizam critérios estatísticos. Entretanto o afastamento do alinhamento da distribuição em relação à diagonal principal dos padrões demanda a utilização de um método mais refinado. A sugestão é a utilização do método de fisher (fisher discriminant function) (JOHNSON; WICHERN, 2007).

Não foi possível analisar de maneira satisfatória a separação do padrão espalhamento MIE, por falta de dados apropriados. O programa de aquisição de imagens limita a obtenção somente a partir de um ângulo de elevação solar de 5 graus, justamente no horário de maior ocorrência deste padrão. Fica como sugestão de trabalho futuro a obtenção e análise de dados que permitam um levantamento mais refinado e apropriada deste padrão.

A escala de luminosidade de um ambiente natural é logarítmica. Em geral estão presentes várias décadas de intensidade luminosas na mesma cena causando distorções e degeneração de cores. Para contornar este problema em trabalhos futuros poderia se utilizar a técnica de imageamento de alto alcance dinâmico utilizando uma câmera tempo de exposição e abertura de obturador programáveis. A partir de através da composição de imagens poderia contornar o problema da escala móvel de luminosidade das cenas naturais.

Outra sugestão de trabalhos futuros seria a utilização da textura das nuvens como indicativo da altitude. A textura tem o potencial de agregar um componente qualitativo adicional na determinação da ocorrência da altitude de determinados tipos de nuvens.

O conhecimento do domínio de tarefa mais refinado desenvolvido no presente trabalho, abre a possibilidade de utilizar-se apropriadamente soluções baseadas também em redes neurais, por que o domínio de trabalho com todas as suas características foi detalhadamente estudado.

Finalmente poderia ser sugerida a aquisição simultânea de imagens de superfície e de satélite, de maneira a reduzir as discrepâncias de avaliação das imagens nas condições de grande dinâmica das condições do tempo.

Bibliography

ALBERTAZZI, A. G.; SOUZA, A. R. *Metrologia Científica e Industrial*. [S.l.]: Editora Manolo, 2008. 407 p.

BARNNET, T. P. et al. On the space-time scales of the surface solar radiation field. *J. Climate*, v. 11, p. 88–96, 1998.

BIDER, M. Über die genauigkeit der registrierungen de sonnenscheinautographen campbell-stokes. a precisão do registro gráfico do brilho do sol. *Arch. Meteorol. Geophys. Bioklimatol.*, v. 9, p. 199–230, 1958.

CAZORLA, A.; OLMO, F. J.; ALADOS-ARBOLEDAS, L. Development of a sky imager for cloud cover assessment. *Journal of the Optical Society of America*, OSA, v. 25, n. 1, p. 29–39, 2008.

CHARTER, N.; TENENBAUN, J. B.; A., Y. Probabilistic models of cognition: Conceptual foundations. *Trends in cognitive sciences*, v. 10, p. 287–291, 2006.

CRESSWELL, J. W. *Research Design Qualitative and Quantitative Approaches*. [S.l.]: Sage Publications Inc, 1994. 228 p.

DEBEVEC, P. E.; MALIK, J. *Recovering High Dynamic Range Radiance Maps from Photographs*. [S.l.], 1997. Disponível em: <<http://www.cs.berkeley.edu/~malik/papers-debevec-malik97.pdf>>.

DUCHON, C. E.; O'MALLEY, M. S. Estimating cloud type from pyranometer observations. *Journal of Applied Meteorology*, v. 39, n. 1, Jan 1999.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Patterns Classification*. 2nd. ed. [S.l.: s.n.], 2001. 654 p.

DUFFIE, J. A.; A., B. W. *Solar Engineering of Thermal Process*. 3rd. ed. [S.l.]: John Wiley & Sons Inc., 2006.

FEISTER, U. et al. Ground-based cloud images and sky radiances in the visible and near infrared region fromwhole sky imager measurements. In: *Proceedings of Climate Monitoring - Satellite Application Facility Trainning Workshop*. [S.l.: s.n.], 2000.

FERNANDEZ-GARCIA, N. et al. Automatic generation of consensus ground truth for the comparison of edge detection techniques. *Image and Vision Computing*, v. 26, 2008.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. 2nd. ed. [S.l.]: Prentice Hall, 2002. 793 p.

- HAMBLYN, R. *The invention of clouds*. [S.l.]: Picador, 2001.
- HARRISON, R. G.; CHALMERS, N.; HOGAN, R. J. Retrospective cloud determinations from surface solar radiation measurements. *Atmospheric Research*, v. 90, p. 54–62, 2008.
- HARRISSON, E. F. et al. *Radiation Budget at the Top of the Atmosphere. Atlas of Satellite Observations Related to Global Change*. [S.l.]: Cambridge Univ. Press, 1993. pp.19-38 p.
- HOLLE, R.; MACKAY, S. Tropical cloudiness from all-sky cameras on barbados and adjacent atlantic ocean. *Journal of Applied Meteorology*, v. 14, 1975.
- HOYT, D. V. V. interannual cloud-cover variations in the contiguous united states. *Journal of Applied Meteorology*, v. 17, 1978.
- HU, Y.; STAMNES, K. Climate sensitivity to cloud optical properties. *Tellus*, v. 52, p. 81 – 93, 2000.
- INANCI, M. N. Evaluation of high dynamic range photography as a luminance data acquisition system. *Ligthting Res. Technol.*, v. 38,2, p. 123–136, 2005.
- IQBAL, M. *An Introduction to Solar Radiation*. [S.l.]: New York Academic Press, 1983.
- JAIN, A.; DUIN, R.; MAO, J. Statistical pattern recognition: a review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 22, n. 1, p. 4–37, Jan 2000. ISSN 0162-8828.
- JOHNSON, A. R.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. 6th. ed. [S.l.]: Pearson Education International, 2007. 773 p.
- KARSTEN, F.; CZEPLAK, G. Solar and terrestrial radiation dependent on the amount and type of cloud. *Solar Energy*, v. 24, p. 177–189, 1980.
- KOSLOF, T. *Visualizing High Dynamic Range Images*. [S.l.], 2006. Disponível em: <<http://vis.berkeley.edu/courses/cs294-10-sp06/wiki/images/5/52/ToddWriteup.pdf>>.
- KOTARBA, A. Z. A comparison of modis-derived cloud amount with visual surface observations. *Atmospheric Research*, v. 92, 2009.
- KREUTER, A. et al. All-sky imaging: a simple, versatile system for atmospheric research. *Applied Optics*, v. 48, p. 1091 – 1097, 2009. Disponível em: <http://www-opticsinfobase.org/DirectPDFAccess/503B44E9-BDB9-137E-C455BA62A9D969C4_176587-.pdf?da=1&id=176587&seq=0>.
- LENOBLE, J. *Atmospheric Radiative Transfer*. [S.l.]: A. Deepak Publishing, 1993. 532 p.
- LILLESAND, T. M.; KIEFER, R. W. *Remote Sensing and Image Interpretation*. [S.l.]: John Wiley & Sons, 1994. 750 p.
- LONG, C. N. Studies of the effect of clouds on solar irradiance using automated clear sky detection and empirical fitting algorithms. In: *Proceedings of the sixth atmospheric radiation measurement (ARM) team meeting*. [s.n.], 1997. p. 457–460. Disponível em: <http://www.arm.gov/publications/proceedings/conf07/extended_abs/long_cn.pdf>.

- LONG, C. N.; ACKERMAN, T. P. Detection of clear skies using total and difuse shortwave irradiance calculatiobs of shortwave cloud forcing and clear skies diffuse ratio. In: *Proceedings of the sixth atmospheic radiation measurement (ARM) team meeting*. [s.n.], 1996. p. 179–183. Disponível em: <http://www.arm.gov/publications/proceedings/conf06-extended_abs/long_cn.pdf>.
- LONG, C. N. et al. Retrieving cloud characteristics from ground-based daytime color all-sky images. *Journal of Atmospheric and Oceanic Technology*, v. 23, p. 633–652, may 2006.
- LONG, C. N.; SLATER, D. W.; TOOMAN, T. *Total Sky Imager Model 880 Status and Testing Results*. [S.l.], November 2001. [Http://www.arm.gov/publications/tech_reports/arm-tr-006.pdf](http://www.arm.gov/publications/tech_reports/arm-tr-006.pdf).
- MAHALANOBIS, P. C. On the generalized distance in statistics. *Proceedings National Institute of Science. India*, 1936. Disponível em: <<http://ir.isical.ac.in/dspace/handle/1-1268>>.
- MANTELLI, S. *Desenvolvimento de uma Nova metodologia para a estimativa da cobertura de nuvens usando uma câmera de superfície e comparando com imagens de satélite*. Dissertação (Mestrado) — Universidade Federal de Santa Catarina, Departamento de Informática e Estatística, 2001.
- MANTELLI, S. et al. Preliminary comparison of short-wave radiometer data and euclidean geometric distance-, image-based method for cloud cover evaluation from surface. In: (PACIS), P.-A. A. of C. I. S. (Ed.). [S.l.]: Pan-American Association of Computational Interdisciplinary Sciences, 2010.
- MANTELLI, S. L. et al. The use of euclidean geometric distance on rgb color space for classification of sky and cloud patterns. *Journal of Atmospheric and Oceanic Technology: In Press*, n. Early Online Release, 2010. Disponível em: <<http://ams.allenpress.com/perlserve/?request=get-abstract&doi=10.1175%2F2010JTECHA1353.1>>.
- MANTELLI, S. L.; WANGENHEIN, A. v.; PEREIRA, E. B. Modelo preliminar de estimativa de cobertura de nuvens, no espaço de cores rgb obtidas a partir de imageador automático. In: XII SYMP. BRASILEIRO DE SENSORIAMENTO REMOTO, GOIANIA BRAZIL. [S.l.], 2005. p. 4123–4131.
- MANTELLI, S. L.; WANGENHEIN, A. v.; PEREIRA, E. B. Methodology for automatic observation of sky patterns. *Colloque d'Informatique: Bréazil / INRIA, Coopérations, Avancées et Défis.*, p. 71 – 75, 2009.
- MANTELLI, S. L. et al. The use of euclidean geometric distance on rgb color space for classification of sky and cloud patterns. *Journal of Applied Meteorology and Climatology*, 2010.
- MARTY, C.; PHILIPONA, R. The clear-sky index to separate clear-sky from cloudy-sky situations in climate research. *Geophysical Research Letters*, v. 27, p. 2649 – 2652, 2000. Disponível em: <<ftp://ftp.pmodwrc.ch/pub/publications/grl%20csi.pdf>>.
- MOECK, M.; ANAOKAR, S. Illuminance analysis from high dynamic range images. *Leukos*, v. 2, n. 3, p. 211–228, 2006.

- MONTGOMERY, D. C. *Introdução ao Controle Estatístico de Qualidade*. 4 th. ed. [S.l.: s.n.], 2001. 513 p.
- MONTGOMERY, D. C. *Design and Analisys of Experiments*. [S.l.]: John Wiley and Sons Inc., 2005. 643 p.
- MONTGOMERY, D. C.; RUNGER, R. C. *Applied Statistics and Probability for Engineers*. 3 rd. ed. [S.l.: s.n.], 2003. 976 p.
- NEWELL, A.; SIMON, H. A. *Human Problem Solving*. [S.l.]: Englewood Cliffs, 1972. 920 p.
- NIST. *Engineering Statistics Handbook*. NIST - National Institute of Standards, 2009. Disponível em: <<http://www.itl.nist.gov/div898/handbook/eda/eda.htm>>.
- ORSINI, A. et al. Cloud cover classification through simultaneous ground-base measurements of solar and infra-red measurements. *Atmospheric Research*, v. 61, p. 251 – 275, 2002.
- PALLÉ, E.; BUTLER, C. J. Sunshine records from ireland: Cloud factors and possible links to solar activity and cosmic rays. *International Journal of Climatology*, v. 21, p. 709–729, 2001.
- PALLÉ, E.; BUTLER, C. J. Comparison of sunshine records and synoptic cloud observations: a case study for ireland. *Physics and Chemistry of the Earth*, n. 27, january 2002.
- PEDHAZUR, E. J.; SCHMELKIN, L. P. *Measurement, Design and Analysis*. [S.l.]: Lawrence earlbaum associates, 1991. 819 p.
- REINHARD, E. et al. *High Dynamic Range Imaging*. [S.l.]: Elsevier, 2006. 502 p.
- RICHARDS, J. A. *Remote Sensing Digital Image Analysis*. 2nd. ed. [S.l.: s.n.], 1995. 340 p.
- RUSSELL, S.; NORVIG, P. *Inteligência Artificial*. [S.l.]: Prentice Hall, 2003. 1021 p.
- SABBURG, J.; WONG, J. Evaluation of ground-based sky camera system for use in surface irradiance measurements. *Journal of Atmospheric and Oceanic Technology*, v. 16, p. 752–759, 1999.
- SCHADE, N. H. et al. Total and partial cloud amount detection during summer 2005 at westerland (sylt, germany). *Atmospheric Chemistry and Physics*, v. 9, p. 1143 – 1150, 2009. Disponível em: <<http://www.atmos-chem-phys.org/9/1143/2009/acp-9-1143-2009.pdf>>.
- SOUZA-ECHER, M. P. et al. A simple method for the assessment of the cloud cover state in high-latitude regions by a ground-based digital camera. *Journal of Atmospheric and Oceanic Technology*, v. 23, n. 3, p. 437–447, 2006.
- STEYVERS, M.; GRIFFITHS, T. L. T. L.; DENNIS, S. Probabilistic inference in human semantic memory. *Trends in Cognitive Sciences*, v. 10, p. 327–334, 2006.
- WALISER, D. E.; COLLINS, W. D.; ANDERSON, S. P. An estimate of the surface shortwave cloud forcing over the western pacific during toga coare. *Geophysical Research Letters*, v. 23, n. 5, p. 519–522, march 1996.
- WILKS, D. S. *statistical methods in the atmospheric sciences*. [S.l.]: Academic Press, 2005. 627 p. (International Geophysics Series).

WMO. *International Cloud Atlas: Manual on the observation of clouds and other meteors.* World Meteorological Organization, 7bis, avenue de la Paix, Case postale 2300, CH-1211 Geneva 2, Switzerland, 1975. WMO-No. 407. I ISBN 9263104077.

WMO. *International Cloud Atlas.* World Meteorological Organization, 7bis, avenue de la Paix, Case postale 2300, CH-1211 Geneva 2, Switzerland, 1987. ISBN 92-63-15306-X, Volume II WMO-No. 407.

WMO. *Manual on Codes, International Codes.* World Meteorological Organization, 7bis, avenue de la Paix, Case postale 2300, CH-1211 Geneva 2, Switzerland, 1995. Disponível em: <<http://www.wmo.int/pages/prog/www/WMOCodes/Manual/Volume-I-selection-ManualCodesSel.html>>.

WMO. *Guide to Meteorological Instruments and Methods of Observations.* 7. ed. World Meteorological Organization, 7bis, avenue de la Paix, Case postale 2300, CH-1211 Geneva 2, Switzerland, 2008. WMO-No. 8 I.15-1 - I.15-11. Disponível em: <http://www.wmo.int/pages/prog/gcos/documents/gruanmanuals/CIMO/CIMO_Guide-7th_Edition-2008.pdf>.

YAMANOUCHI, T. Variations of incident solar flux and snow albedo on the solar zenith angle and cloud cover at mizuho station antartica. *Journal of the Meteorological Society of Japan*, v.61, n. 6, p. 879–893, 1993.

ZHANG, P. G. Avoiding pitfalls in neural network research. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS PART C APPLICATIONS AND REVIEWS.*, v. 37, p. 3–16, January 2007.

Glossário

absorção: é a perda de energia solar para os elementos presentes na atmosfera.

albedômetro: medidor de albedo da superfície.

ambiente de pesquisa: O ambiente são todas as opções possíveis e imagináveis sem restrição que podem ser definidas para a atuação do agente.

ambiente do problema: O ambiente do problema que se restringe somente no conjunto de opções necessárias para o agente resolver um tipo particular de problema.

ambiente de tarefa: O ambiente de tarefa contém todas as possibilidades que podem ser acessadas pelo agente para a sua atuação.

antropogênicas: Efeitos, processos, objetos ou materiais antropogênicos, são aqueles derivados das atividades humanas, em oposição a aqueles que ocorrem em ambientes naturais sem influência humana. Este termo é muitas vezes utilizado no contexto de externalidades ambientais na forma de resíduos químicos ou biológicos que são produzidos como subprodutos de atividades humanas. Por exemplo, é largamente aceito que o aumento de dióxido de carbono na atmosfera com origem antropogênica é o factor principal por detrás das alterações climáticas.

AU- Unidade astronômica : é a distância média entre o sol e a terra ou seja $1,496 \times 10^8$ km.

BSRN: é um programa das Nações Unidas para a instalação de um conjunto de estações de alta precisão para monitoração do clima <http://www.bsrn.awi.de/> e <http://www.gewex.org/bsrn.html>

ceilômetros: são equipamentos que utilizam LASER através de tecnologia LIDAR, utilizado para determinar a altitude da base das nuvens. Em geral estão presentes nos aeroportos e indicam o "teto" de vôo dos aviões.

cirrus: tipo de nuvem fina que ocorre em altas camadas atmosféricas (> 7000 m). O Anexo A apresenta ilustrações acerca deste tipo de nuvem.

cloud forcing: é um índice que indica a diferença entre a radiação de céu claro e a radiação medida.

ColorInspector3D: desenvolvido pela <http://www.f4.fhtw-berlin.de/~barthel/ImageJ/ColorInspector//help.htm>

conhecimento tácito: Tácito é o tipo de conhecimento difícil de ser transferido para outra pessoa através de meios escritos ou formais.

difusa: é a radiação que é recebida na superfície horizontal a partir de um ângulo sólido de 2π com exceção do ângulo sólido compreendido pelo Sol. Iqbal (1983) pag. 215. Sua origem esta no espalhamento causado pelas moléculas de ar e pelos aerossóis.

direta: é radiação solar que chega à superfície diretamente do círculo solar Iqbal (1983) pag. 108.

EDA: A análise exploratória de dados é um conjunto de suposições, princípios e técnicas necessárias para se obter maiores informações acerca dos dados em questão

EPPLEY: Eppley Laboratories <http://www.eppleylab.com>

EKO: EKO Instruments Co. Ltd. <http://www.eko-usa.com>

espalhamento: é a difusão dos fótons da luz solar pelas moléculas e partículas presentes na atmosfera

Fairmount Weather Systems: Unit 4, Whitecroft Road, Meldreth, Hertfordshire, SG8 6NE, UK. <http://www.fairmountweather.com/>

fator de brilho solar: Fs - é definido como a quantidade de horas que o brilho solar esta presente (queimando o papel) dividido pela quantidade de horas possíveis de sol a partir de uma elevação acima de 3 graus para o local da estação Pallé e Butler (2002)

fator de nuvens: Fc - é definido como o número de horas complementar ao de brilho solar onde $Fc = 1 - Fs$

FD: Fração difusa - é definida como a razão entre a radiação difusa D e a global G

global: radiação global (ou total) é a soma das radiações direta e difusa

Greenwich Mean Time: é a hora oficial do tempo baseada no fuso horário de Greenwich na Inglaterra. É utilizado como horário padrão nas estações de observação sinóptica, imagens de satélite, etc. Não sofre alterações com o horário de verão vigente no local de observação.

Kipp and Zonen: Delftechpark 36, 2628 XH Delft , The Netherlands. www.kippzonen.com

LABSOLAR-EMC-UFSC: (<http://www.lepten.ufsc.br/home/solar.html>)

LIDAR: LASER Detection and Ranging. O LIDAR é um equipamento que utiliza o mesmo princípio de funcionamento do RADAR, mas utiliza LASER ao invés de ondas de rádio. O LIDAR emite periodicamente um feixe de LASER na posição zenital. O LIDAR analisa o meio de propagação do feixe de LASER através de um fenômeno chamado retroespalhamento (backscatering). Através da análise das reflexões retroespalhadas da luz consegue-se determinar a distância e a atenuação do meio de propagação do LASER

MAPLE: (<http://www.maplesoft.com/>)

onda longa: é a radiação medida na superfície na faixa de infra vermelho distante (3,5 a 50 μm).

LUX: (ou luxímetro) é a radiação na faixa de sensibilidade da luz visível para os seres humanos.

MODTRAN: MODerate resolution atmospheric TRANsmission. Modelo computacional utilizado para modelar a propagação de ondas eletromagnéticas na atmosfera na faixa espectral 100-50,000cm⁻¹ (0.2 to 100um).

nuvens: são um agregado de pequenas partículas de água, cristais de gelo ou uma mistura de ambos, cuja base pode ser percebida a partir do ponto de observação. O diâmetro máximo é de 200 μm . Gotas ou partículas acima deste valor são considerados como chuva ou granizo. WMO (2008) (15.1.1).

pirradiômetro: Sensor de radiação combinado descendente e ascendente.

pirgeômetro: Sensor utilizado na medida da radiação de onda longa.

pirheliômetro: Sensor utilizado na medida da radiação. direta

PAR: (Photosynthetically Active Radiation) é a radiação fotosinteticamente ativa na faixa de sensibilidade das plantas

repetitividade: "A repetitividade define a faixa dentro da qual, para uma dada probabilidade, o erro aleatório é esperado." (ALBERTAZZI; SOUZA, 2008)

resolução espacial: refere-se à precisão da medida de distância mínima perceptível entre dois objetos distintos em uma imagem.

resolução temporal: refere-se à precisão de uma medida em relação a um período de tempo.

sensor de brilho solar: o termo brilho solar esta associado ao brilho do disco solar que excede a radiação difusa de fundo. Em geral está relacionada à radiação visível e geralmente associado à presença de sombra atrás de objetos iluminados. O limiar de sensibilidade do sensor é de 120 W/m². WMO (2008) (8.1). O sensor foi inventado por John Francis Campbell em 1853 e aperfeiçoado por George Gabriel Stokes em 1879. É composto de uma esfera de quartzo que concentra os raios de sol em um papel graduado.

SONDA: Sistema de Organização Nacional de Dados Ambientais para o setor de Energia. Rede de monitoração de radiação solar para validação imagens de satélite e modelos físicos atmosféricos <http://sonda.cptec.inpe.br/>

VISUAL STUDIO: www.microsoft.com

WINDOWS: www.microsoft.com

wxWIDGETS: <http://www.wxwidgets.org/>

Yankee Environmental:(<http://www.yesinc.com>)

zenital: Ângulo solar relativo à normal do ponto de observação.

Publicações desenvolvidas como resultados intermediários no decorrer do presente trabalho

Apêndice A1

- XXIX Congresso da Sociedade Brasileira de Computação - Colóquio em Informática Brasil INRIA. Methodology for Automatic Observation of sky patterns.

Poster e resumo extendido.

Contribuição da publicação em relação à pesquisa desenvolvida. A Publicação basicamente destaca a importância da modelagem apropriada do domínio cognitivo a ser investigado por um agente inteligente.

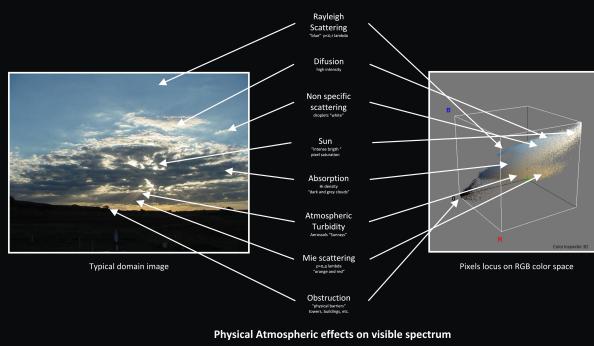
METHODOLOGY FOR AUTOMATIC OBSERVATION OF SKY PATTERNS



Sylvio Luiz Mantelli Neto^{1*}, Aldo von Wangenheim², Enio Bueno Pereira¹

¹ - Image Processing and Computer Graphics Group and Knowledge Engineering Department
Knowledge Engineering Department and Department of Statistics and Computer Sciences
Federal University of Santa Catarina – Florianópolis / Brazil

² - INPE/CPTEC - Center for Weather Forecasts and Environment Studies
Brazilian Institute for Space Research – São José dos Campos / Brazil



ETAGE	FORM	GENUS	SPECIES	VARIETY
LOW-STAGE	CUMULIFORM	CUMULUS	HABITUS MEDICRIS CASTELLANUS FRACTUS	
		CUMULONIMBUS	CAPILLATUS CAVUS	
	STRATIFORM	STRATOCUMULUS	FLABELLARIS PASTILLARUM LENTICULARIS	GRACILIS TRANSILICIOS DEDICIOS RADICATUS PERILOCUTUS
		STRATUS	NEBULOSUS FRACTUS	GRACILIS TRANSILICIOS UNDULATUS
MID-STAGE		NIMBOSTRATUS		
		ALTOSTRATUS		GRACILIS TRANSILICIOS UNDULATUS RADICATUS DUPLOCUTUS
	CUMULIFORM	ALTICUMULUS	CASTELLANUS TRANSILICIOS STRATIFORMIS UNIFORMIS	GRACILIS TRANSILICIOS UNDULATUS RADICATUS DUPLOCUTUS LACUNOSUS
HIGH-STAGE	CIRIFORM	CIRRUS	UNICUS SPISSATUS FERENS CASTELLARUM	RADICATUS DUPLOCUTUS VULNERIS VERTIBRATUS
	STRATIFORM	CIRROSTRATUS	NEBULOSUS FRACCUS	DUPLOCUTUS
	CUMULIFORM	CIRROCUMULUS	FLACCUS PASTILLARUM LENTICULARIS	UNDULATUS DUPLOCUTUS LACUNOSUS

Abstract:

Surface observers on monitoring meteorological stations regularly classify and register sky patterns based on human cognitive training. Substitute those observers by automatic systems is highly desirable to reduce subjective analysis. Most automatic systems use thresholding methods and ignores other physical phenomena existent in atmosphere obtaining only binary results (clouds and sky). The current work proposes a method that will establish the correspondence between observations percept from those patterns and theory using Bayesian approach embedded in Intelligent Agents in the following steps. Model the cognitive patterns on color space obtained from surface cameras and analysis. Validation and refinement of system results could be done using sun photometers and LIDAR systems.

Keywords:

Automatic pattern recognition, Cognitive Bayesian model, Multivariate statistics, atmospheric patterns.

Introduction:

- Atmospheric and physical models have high variability and reduced performance on the presence of cloudy conditions.
- Satellite and surface based observations have limitations on providing data for model validation. Simultaneous observations are desirable to reduce uncertainties inherent to both observation systems.
- Human observers provide good qualitative, but bad quantitative analysis. In automatic systems problems are inverse.
- Replace human observers is desirable to reduce shift loads, costs and subjective aspects of observations.
- Existant automatic surface observations systems does not present satisfactory results (binary) due to limited approach.
- Advances on automated system performance has to deal with proper characterization on domain of physical phenomena, considering cognitive aspects using appropriate inference engines in order to match World Meteorological Organization observation standards.

Methodology

- Bayesian methodology based on supervised learning and analysis.
- Cognitive modeling based on Exploratory Data Analysis using multivariate statistics mathematical models to characterize specific patterns on color space.
- Field physical validation.

Expected Scientific Results

- Development of appropriate methodology for surface based satellite image validation on visible bands for cloudy conditions.
- Improvement of automatic atmospheric observation systems.

Collaboration proposals to INRIA teams

- PERCEPTION: Representation and interpretation of sky patterns.
- SELECT: Multivariate statistical modeling on Bayesian approach
- CLIME: Validation of atmospheric physics issues



INPE/CPTEC - Brazilian Institute for Space Research
Center for Weather Forecasts and Environment Studies
www.cptec.inpe.br

LAPIX – Image Processing and Computer Graphics Laboratory
(48) 3721-9516
www.lapix.ufsc.br



UFSC - Federal University of Santa Catarina
University Campus – Trindade/Florianópolis
www.ufsc.br

Methodology for Automatic Observation of sky patterns

Sylvio Luiz Mantelli Neto^{1,2,3}, Aldo Von Wangenheim^{2,3}, Eros Comunello²,
Enio Bueno Pereira¹

¹INPE-CPTEC Instituto Nacional de Pesquisas Espaciais - Centro de Previsão do
Tempo e Estudos Climáticos. Av. dos Astronautas 1758 São José dos Campos SP
Brasil 12227-010

²UFSC-INE-LAPIX Universidade Federal de Santa Catarina - Departamento de
Informática e Estatística - Laboratório de Processamento de Imagens. Florianópolis SC
Brasil.

³UFSC-EGC Universidade Federal de Santa Catarina - Departamento de Engenharia e
Gestão do Conhecimento Florianópolis SC Brasil.

sylvio@lepten.ufsc.br, awangen@inf.ufsc.br,
enio.pereira@cptec.inpe.br

Abstract. *Surface observers on monitoring meteorological stations regularly classify and register sky patterns based on human cognitive training. Substitute those observers by automatic systems is highly desirable to reduce subjective analysis. Most automatic systems use thresholding methods and ignores other physical phenomena existent in atmosphere obtaining only binary results (clouds and sky). The current work proposes a method that will establish the correspondence between observations percept from those patterns and theory using Bayesian approach embedded in Intelligent Agents in the following steps. Model the cognitive patterns on color space obtained from surface cameras and analysis. Validation and refinement of system results could be done using sun photometers and LIDAR systems.*

Keywords: Automatic pattern recognition, Cognitive Bayesian model, Multivariate statistics, atmospheric patterns.

1. Introduction

The main purpose of the current work is the characterization of sky patterns on Celestial Hemisphere (CH) observable from the surface. This characterization will improve the classification and qualification of observed optical atmospheric phenomena allowing the use of intelligent agents (IA) (NORVIG and RUSSEL, 2003). The most widely used characterization of sky patterns present on CH is established by World Meteorological Organization WMO. Synoptic observers use WMO (1975, 1987 and 1996) criteria to classify cloud type, height and amount according current atmospheric conditions. The human observer is intended to be replaced from that task by a surface camera equipped with “fish eye” lens and image interpretation algorithms, reducing efforts imposed by regular human working shifts, subjective aspects and misinterpretation. Sky monitoring by cameras and related methods are also a part of the efforts to environment monitoring activities (WCRP Report 2007). The automatic sky imagers available on the market usually presents binary evaluation of CH patterns (sky and clouds), reducing the

dimension of color space restricting the capabilities present on observed color domain (LONG et al., 2006). WMO classification was established to be used by an observer that has a very well developed and trained cognitive system to watch, classify and register the CH patterns. The synoptic observer has a good performance on qualitative analysis and poor performance on quantitative of the present CH conditions. With automatic systems is the inverse. The defined rules established by WMO are not suitable to be used by IA to replace human observation. With current paradigms IA are not capable yet to fully classify different patterns from those ones defined on the WMO domain environment¹. If the domain is restricted, so is the program analysis or IA action. The classification criteria established by WMO could be used as a target objective for IA classification, but not as a domain. The sensed domain to be considered in that case is broader than cloud observation parameters established by WMO, leading IA's to misinterpretation. This misinterpretation increases the uncertainty of the observation and imposes a disappointment on automatic systems outcomes. The current work will show that WMO standards are not suitable to be used as a domain environment, but as a target outcome. It proposes an alternative domain based on optical atmospheric phenomena percept by humans to be used by IA. To achieve that an exploratory data analysis (EDA) is proposed to be performed in color space to find atmospheric patterns typical locus on a multivariate context. As the pattern identification is highly oriented to human interpretation, cognitive modeling using a Bayesian approach (supervised learning and analysis) related tools is intended to be performed. To validate the aforementioned model a field experiment is also proposed to validate and refine the proposed patterns analysis using LIDARS and SUNPHOTOMETERS. The next sections presents a little further details about the proposed method.

2. Cognitive Environment Definition and Physical Context

Proper environment definition is a key issue in the current work. The environment will be defined according to the work proposed by Newell & Simon (1972). In that work it is suggested the definition of three different levels to be taken into account by the IA: research, task and problem environments. Restrictions on those environments could lead to inaccessible states on the problem solution (MANTELLI, 2001). The current studies consider only the visible part of radiation, because sensor and human cognitive systems only respond to visible part of the electromagnetic spectrum. Two principal physical phenomena dominate the interaction of sun light with the atmosphere, scattering and absorption (LILLESAND and KIEFER, 1994; LENOBLE, 1993; IQBAL, 1984). The principal observable scatter causes are: Rayleigh, Mie, aerosol atmospheric turbidity, non selective and sun light diffusion. The principal observable effect of absorption is the dark and grey levels of clouds. More detailed information about proposed environment could be observed on figure 1. In this context an EDA of those patterns on multivariate color space need to be performed to characterize their typical locus.

¹ Domain is a section of the world about which we wish to express some knowledge. P. 197 (RUSSEL AND NORVIG 2003).

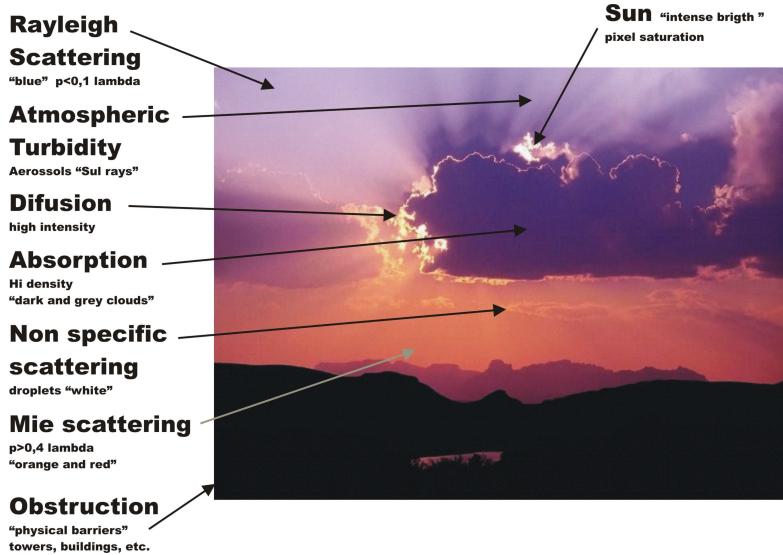


Figure 1. A typical scenario of proposed observable sky patterns from surface.

Categorization of clouds according to WMO is also an issue to be deal with. It is a point to be investigated that clouds could be discriminated by texture. It is considered that cloud texture is the same but different altitudes of same texture causes different properties that could be distinguished. Fourier analysis, wavelets and similar methods is intended to be used to investigate that. From physical context above several patterns may be proposed to take part on the environment and could be easily identified by its results. If not defined properly automatic systems are far behind to percept patterns properly. Task environment defined that way is further beyond WMO standards. The first suggestion is to reduce the identified patterns and map into a WMO goal as suggested on physical context. The difference now is that mapping is done specifically by the system and not unpredictably and improperly by the software.

3. State-of-the art and future work

Based on the aforementioned reasons it was proposed an environment defined by induction using Bayesian cognitive modeling (TENENBAUM, 2006). In that context current research leaded to multivariate statistical exploratory data analysis (EDA) of proposed image patterns. Although the current research is working in EDA definition the appropriate IA inference still have to be defined. The model validation is proposed in two steps. The first step is comparing the system performance against a synoptic observer. The second one, by using advanced field equipments and atmospheric modeling. Among then it is proposed a sun photometer and a LIDAR² techniques to measure simultaneous atmospheric and visual parameters such as clouds distances by texture.

4. International cooperation and French-Brazilian teams

The state-of-the-art of sky imaging is being addressed by several research institutes and is being targeted by international scientific community (WCRP Report 2007). The

² LIDAR – Laser Detection and Ranging.

original approach and modeling could be done in a joint cooperation between Brazil and France. Several INRIA Research groups could be involved in these tasks.

The first one in formal modeling, representation and interpretation of sky patterns into its symbolic atmospheric meaning. Checking if the proposed method is the more appropriate to be applied on the specific cognitive area. The effects of image illumination variation along the day should also be investigated. (INRIAS's PERCEPTION research group)

The second one is the model selection and statistical learning for pattern recognition in the multivariate image domain of the atmospheric patterns. Patterns should be investigated for their typical distribution (normal or non-normal) on color space. The study could be extended by adding and verifying the effects (or factors) of typical solar variables (global, direct and diffuse) on image. (INRIAS's SELECT research group)

The third potential research group could be on model validation. Validation could be done in several ways. It will be proposed two methods. The first one demands high cost equipments not available to the Brazilian counterpart (i.e. sun photometers, trackers and LIDARS) or satellite images. Some of these manufacturers are based in France (CIMEL), and are present in several French research centers. For satellite images INRIA has a research group working on environment data processing and computer vision techniques (INRIAS's CLIME research group).

5. Expected Scientific results

Several results are expected from current research as listed. Development of a suitable IA in applied area. An improvement of automatic sky imaging observation methodology to be employed on surface monitoring stations. Cost reduction and rationalization of current synoptic operations. Test of a suitable monitoring point for cross validation of satellite images on visible spectrum for clear, cloudy and mixed conditions skies. A better evaluation of atmospheric models on mixed cloudy conditions. A comparison of proposed methodologies based on tropical and tempered climates.

6. Collaboration proposals.

Actually UFSC-INE-LAPIX (Laboratory of Image Processing) do not have cooperation program with any official INRIA partners. Current potential Brazilian Partners includes INPE-CPTEC (Brazilian Institute for Space Research – Center of Weather Forecasts and Environment Studies). The proposed activities for Brazilian partners involve modeling of domain and target sky patterns. Joint affords could be done with PERCEPTION and SELECT groups on cognitive aspects of synoptic observation and statistics modeling. CLIME group could support on validation and atmospheric physics issues of atmospheric modeling. The proposed activities with INRIA partners involve modeling and operational support for field validation of proposed modeling.

7. Conclusions

Although Celestial Hemisphere pattern classification has a satisfactory result when done by a trained operator, automatic classification using surface cameras are subject to uncertainties due to inappropriate approach. Current work proposes a new paradigm on classification matching physical and optical phenomena based on current atmospheric

conditions. This approach is believed to open a new perspective based on human cognitive observations to automatic systems. Once the factors that cause uncertainties are tagged, it is easier to be addressed and overcome them. Independently of the **how** the solution will be implemented **what** to do to solve the problem is now established. Define which intelligent agent to do the job and the best technique to solve it is the next step. The new approach presented here is believed to be a suitable approach for classification of Celestial Hemisphere patterns.

References

- Iqbal, M., Introduction to Solar Radiation. Academic Press 1984.
- Lenoble, J., Atmospheric Radiative Transfer. 1993 A Deepak Publishing.
- Lillesand, T. M., Kiefer, R. W. Remote Sensing and Image Interpretation 1994 John Wiley & Sons Inc.
- Long C. N., Sabburg J. M., Calbó J., Pagès D., 2006: Retrieving Cloud Characteristics from Ground-Based Daytime Color All-Sky Images. *J. of Atmos. and Oceanic Technology* 23, issue 5, 633–652
- Mantelli S. L., Desenvolvimento de Metodologia para a estimativa da cobertura de nuvens usando uma câmera de superfície e comparando com as imagens de satélite. 125f. Dissertação (Mestrado em Ciências da Computação). Universidade Federal de Santa Catarina, Departamento de Informática e Estatística, Florianópolis, 2001.
- Newell, A., Simon H. A., Human Problem Solving. Englewood Cliffs, NJ: Prentice Hall 1972.
- Russel S. J. and Norvig P., Artificial Intelligence a Modern Approach. 2003. Prentice Hall.
- Tenenbaum J. B., Griffiths T. L. and Kemp C., 2006: Theory-based Bayesian models of inductive learning and reasoning. Special Issue: Probabilistic models of cognition. *TRENDS in Cognitive Sciences*, 10, 309-318
- WCRP World Climate Research Program, 2007: *Section of the Baseline Surface Radiation Network*, Lindenberg Germany, Informal Rep. No. 9 54 pp. [Available online at <http://www.gewex.org/BSRN/BSRN-9-REPORT.pdf>.]
- World Meteorological Organization, 1975: International Cloud Atlas: Manual on the observation of Clouds and Other Meteorol. Volume I, WMO-No. 407, Geneva.
- World Meteorological Organization, 1987: International Cloud Atlas. Volume II, WMO-No. 407, Geneva.
- World Meteorological Organization, 1996: Guide to Meteorological Instruments and Methods of Observation, WMO-No. 8, Geneva.

Apêndice A2

IEEE Computer Based Medical Systems

Contribuição da publicação em relação à pesquisa desenvolvida. A publicação basicamente indica a distribuição típica dos padrões de interesse no espaço de cores, a utilização da metodologia Bayesiana, aplicação de métodos estatísticos multivariados, a importância das funções de discriminação (Distância Euclidiana, Distância de Mahalanobis e Polinomial Mahalanobis) apropriadas para a classificação dos padrões de interesse.

Learning a Nonlinear Color Distance Metric for the Identification of Skin Immunohistochemical Staining

Antonio Carlos Sobieranski¹, Sylvio Luiz Mantelli Neto^{1,2,3}, Leandro Coser^{1,3}, Eros Comunello^{1,4}
Aldo von Wangenheim^{1,3}, Eduardo Cargnin-Ferreira⁵, Gabriella Di Giunta⁶

¹LAPIX - Laboratory of Image Processing and Computer Graphics
UFSC - Federal University of Santa Catarina

²CPTEC - Center for Weather Forecasts and Environment Studies
INPE - Brazilian Institute for Space Research

³ECC - Knowledge Engineering Department - UFSC - Federal University of Santa Catarina

⁴MCA - Graduate Program in Applied Computer Science - UNIVALI - Itajai Valley University

⁵BEG - Laboratory of Histological Markers, Department of Biology, Embriology
and Genetics - UFSC - Federal University of Santa Catarina

⁶Pathology Service - UFSC - Federal University of Santa Catarina

{asobieranski, leandro, eros}@cyclops.ufsc.br, sylvio@lepten.ufsc.br, awangenh@inf.ufsc.br

Abstract

This paper presents a semiautomatic method for the identification of immunohistochemical (IHC) staining in digitized samples. The user trains the system by selecting on a sample image some typical positive stained regions that will be used as a reference for the construction of a distance metric. In this learning process, the global optimum is obtained by induction employing higher polynomial terms of the Mahalanobis distance, extracting nonlinear features of the IHC pattern distributions. The results of the proposed method showed a high correlation to a pathologist's manual analysis, which was used as a golden standard, presenting a more robust discrimination between stained and non-stained areas with little bias.

1 Introduction

Immunohistochemistry (IHC) is a technique regularly used in histopathology diagnostic to study the distribution and intensity of biomarkers in tissues [6][5]. This technique uses specific antibodies to stain proteins in situ. The most common use of IHC in medical practice is on the diagnosis and/or prognosis evaluation of tumor markers.

The assessment of protein expression using immunohistochemical is in most cases, performed in a subjective man-

ner. Although IHC as performed nowadays has various advantages, such as low cost and wide availability, the main difficulty is the need to employ analysis methods that retain a subjective character, like slide scores [8]. In cases where the objective evaluation of protein expression is required this procedure becomes labor-intensive, error-prone and observer-dependent.

In the last years, computer image analysis tools have reduced the effect of observer biases, and increased the usefulness of IHC as a tissue-based procedure for evaluation of diseases [11].

In [15] it is presented a review of recently image analysis methods for IHC identification and quantification. At least seven computational methods are presented. Most of the IHC image analysis methods described employ some kind of preprocessing, color information from the HSI or RGB color spaces and threshold cutoff.

Other related approaches are based in spatial domain information: In [13], a k-means clustering technique with linear thresholding is used to identify classes of stained tissues. In [11], the CMYK color space is used to quantify IHC stains in hematoxylin counterstained histological sections. A quantitative analysis of precancerous lesions is presented in [1]. The later work combines basic image processing methods such as median and low pass filtering, region growing segmentation, thresholding and morphological operations.

Although these methods can achieve good results for a number of specific application areas, generally they are based upon linear color space models and it is unlikely to provide a general solution [4]. Their current development status reached a point where large clinical studies are still required to validate their effectiveness [15]. Preprocessing and manual intervention is still necessary in order to resolve particularly challenging or ambiguous cases, as well as to provide high-level supervision of the produced diagnostic results [15].

In this paper we present a semiautomatic computer assisted approach that uses a new method for the identification of IHC patterns on RGB image samples. A preliminary identification of typical staining areas of a given IHC examination is performed by the pathologist, whose marking is used to train a nonlinear distance metric. This distance metric is later used to match similar remaining IHC color vectors quantitative analysis.

2 Mathematical Background

In the review presented by [15] it is possible to observe that the majority of computer tools use preprocessing methods with some kind of Euclidean distance-based linear thresholding segmentation in a specific color space to perform the image analysis. The main objective of any of the methods described is to discriminate a typical target pattern on multivariate color space due to IHC color marking. Euclidean distance is properly used to indicate similarity only if variances among dimensions are the same and normally distributed [7]. This approach includes too much outliers in the analysis, as could be notice latter in figure 2-b.

If a target pattern has a multivariate distribution and the variances of dimensions are not the same (but still normally distributed) a Mahalanobis distance (MD) can be used to indicate vector similarity [12]. The MD weights the distance calculation according to the statistical variation of each dimension, given by:

$$d_M(x, y) = \sqrt{(x - y)^T A^{-1} (x - y)} \quad (1)$$

where: $d(x, y)$ is the Mahalanobis distance between two color vectors $x(r_i, g_i, b_i)$ and $y(r_j, g_j, b_j)$ on RGB color space. A^{-1} is the inverted covariance matrix of a specific pattern and T is the matrix transpose operation. However, the MD has its performance reduced if the pattern is non-linearly distributed on color space.

A distance metric with good quality should be able to identify and discriminate relevant and irrelevant features. In the past years various research studies are being performed on this issue, more specifically in distance metric learning (DML). The most relevant DML methods are summarized in a survey presented by Yang [9], and can be clas-

sified in supervised and non-supervised ones. For the supervised ones, a large number of DML are knn-based methods learned by means of pairwise constraints¹ [2]. Additional work on DML using pairwise constraints is presented by Xiang [14], where a MD is learned for data clustering and pattern classification. Xiang compares his classification method with other ones described in the Yang's survey, indicating that pairwise classification approaches could be used in IHC context.

But differently from the aforementioned DML methods, it would be desirable to select only the similar data points instead of the pairs of similar and dissimilar ones. It is more natural and comfortable for the pathologist to specify only positive IHC data points that belongs to the same class. In the section 2.1 it will be described how the implementation of this feature was possible.

2.1 Polynomial Mahalanobis Distance

In our approach, we avoided the limitations of the pairwise constraints by using the Polynomial Mahalanobis Distance (PMD) [3]. This distance metric is able to capture the nonlinear features of the IHC pattern by using only similar data points in the learning step. The nonlinear discrimination is guided by a polynomial degree (q -order), which determines how rigorous the similarity must be in face of IHC pattern distribution. Its q -order also acts as better outlier elimination for high polynomial orders, increasing the robustness of the classification method.

The PMD can be obtained from the Mahalanobis distance described in equation 1, which is also the first of q -orders in a PMD sequence. The main concept of the PMD is mapping the input data into higher order polynomial terms, creating very high polynomial maps. Let $S : p_1, p_2, p_3, \dots, p_N$ be a set of m -dimensional color vectors, where $p_i \in \Re^m$ corresponds to the similar data points (input data), and N being the size of the set S . The first step is compute the MD between any two color vectors x and y (by equation 1) using the input data S in the covariance calculation, obtaining the first q -order (1-order, or the MD between x and y). The second step is calculate the second polynomial projection (2-order), by mapping all m -dimensional basis $p_k = p_{k1}, p_{k2}, \dots, p_{km}$ of S , for $k = 1, \dots, N$, into all polynomial terms of order q or less. For example, consider a two dimensional vector $p_k = p_{k1}, p_{k2}$. The mapping of this two dimensional vector into 2-order polynomial terms is $(p_{k1}, p_{k2}, p_{k1}^2, p_{k2}^2, p_{k1}p_{k2})$.

From a computational point of view, the q -order PMD could be obtained by directly mapping all data points into their respective polynomial terms, and so using it in the equation 1. However, for large m and q , the dimension of

¹The user defines which pairs of data points are similar (S) and dissimilar (D), in the form of $S : (p_i, p_j)$ and $D : (p_i, p_j)$.

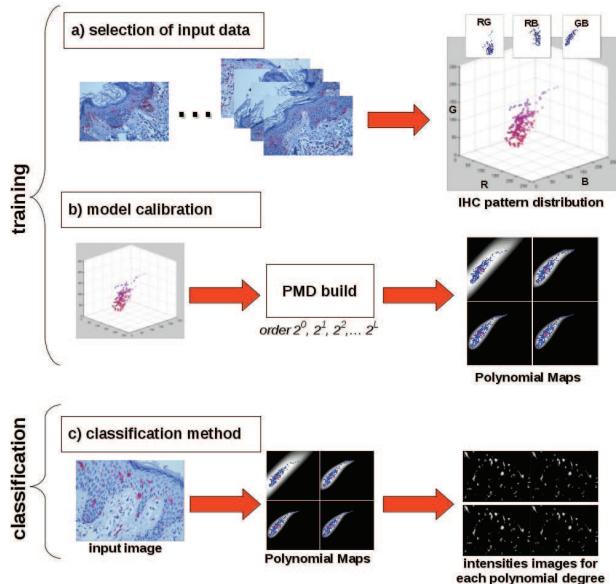


Figure 1. General overview of the proposed approach: (a) input data definition during training and corresponding points on RGB color space; (b) learning the PMD distance metric; (c) classification and discrimination of IHC patterns

the terms combination makes this distance metric infeasible. A way to compute very large polynomial terms is by using the following framework proposed by [3]:

$$d_{PM}(x, y) = d_{M_{\sigma^2}}(x, y) + \sum_{l=1}^L d_{M_{\sigma^2}}(g_l^i, g_l^j) \quad (2)$$

where $d_{M_{\sigma^2}}(x, y)$ is the Mahalanobis distance with a small positive value σ^2 used to avoid inversion limitation, $L > 0$ is the q -order parameter ($q = 2^L$), and the arguments g_l^i and g_l^j are the next projections ($q + 1$) of x and y into their polynomial terms. The PMD is better explained in details on the original paper [3].

3 IHC Identification Method

The proposed approach is based on Bayesian models [10] and is performed in two steps: training and classification. The diagram of figure 1 shows a general overview of the proposed approach.

The training step is a supervised learning process subdivided into two phases: selection of input data points and model calibration, as shown in figure 1-a) and 1-b), respectively.

As first phase, the input data points are defined by the pathologist, which selects a few typical IHC areas from a specific sample image, or from a set of IHC images of the same session. This data points are later used as input data (S) in the distance metric learning. As we can note in figure 1-a) at right side, the selected input data points compose a three-dimensional histogram on the RGB color space. This histogram corresponds to the IHC pattern distribution we are interested in modeling into a useful discrimination function.

As second phase of training step, the model calibration is where a distance metric is established. The figure 1-b) illustrates this phase as follows.

- The IHC pattern distributions previously defined in phase a) are used as input data points (S) (section 2.1) in the PMD metric learning (PMD build).
- The parameter L of equation 2 needs be defined. As early explained, this parameter controls the number of polynomial maps obtained, one for each q -order until L -order. PMD metric is a projective method; all lower-orders built ($q \leq L$) are available to be used without new training requirements. In figure 1-b) at right side is demonstrated the polynomial maps obtained by the PMD metric, where the number of projections is $L + 1$ because the first order (2^0) is always calculated. These polynomial maps will be better detailed later, in figure 2.

Once the distance metric is learned and the polynomial maps are obtained on the calibration model they can be used to identify other related IHC expressions. This is done in the last step of our approach, designated classification (figure 1-c). This step can explained in the follow manner.

- The pathologist defines the image or set of images that will be classified by the polynomial maps. One of the advantages of this approach is the possibility to select a different set of images to be classified, distinct of the images used in the training step, and using the same distance metric. The only restriction is to select images of a same context, having the same tissue and biomarker properties.
- Once defined, the images are classified by computing the similarity of each color vector in a specific q -order map by the learned distance metric (figure 1-c), at center). As previously described, the lower q -order polynomial maps are also available for the user to be chosen, including the original MD distance (linear). However, the MD distance still requires some kind of threshold value if the operator wants to use it.

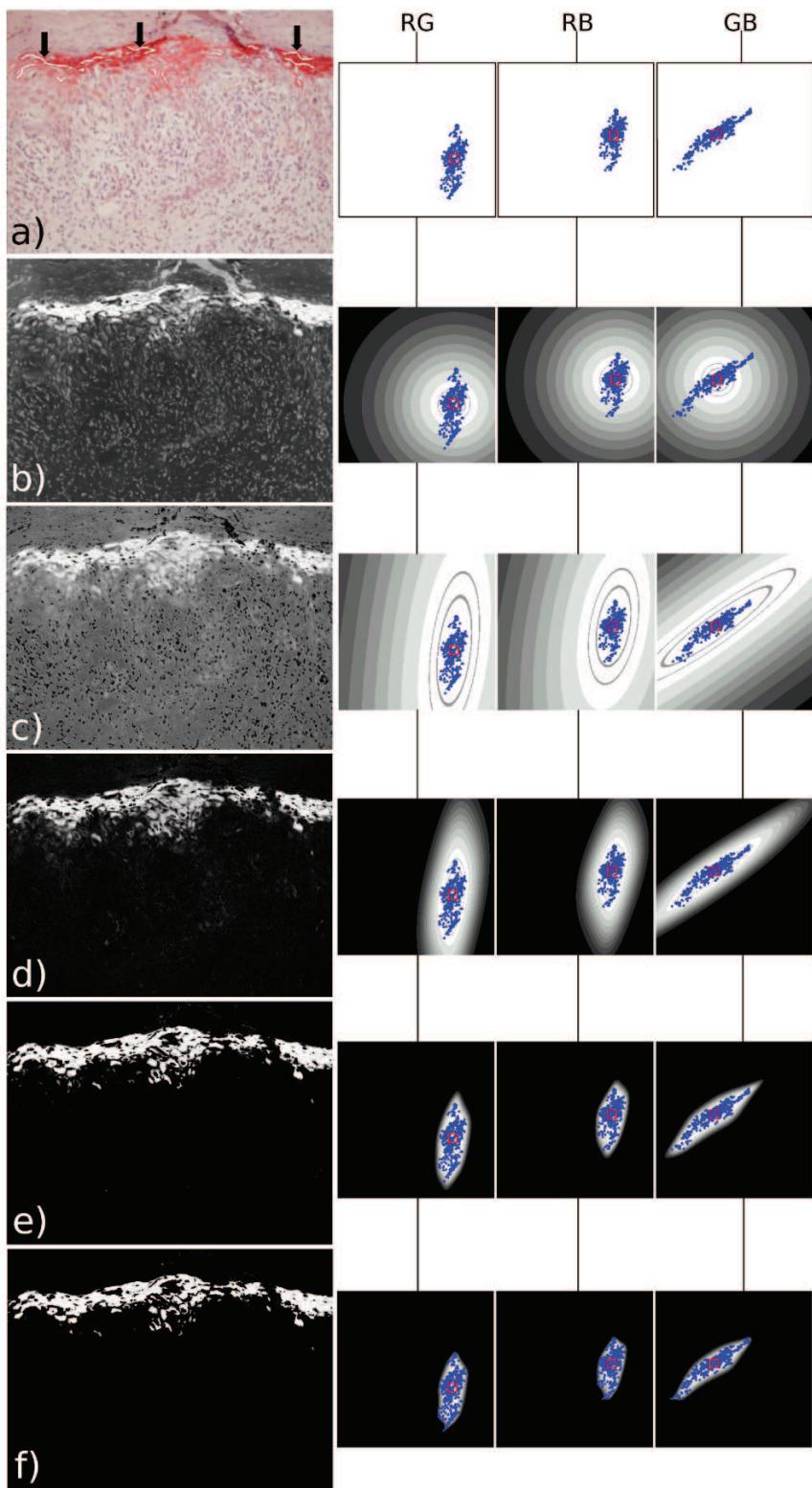


Figure 2. Training step and evolution of classification model. Left side: sample image and classification results. Right side: color distribution and the linear / polynomial maps.

- Finally, for each sample image a similarity map corresponding to calibrated IHC intensities is obtained. This similarity maps also defines the segmented areas related to the IHC positive staining.

The higher polynomial map selected by pathologist provides higher responses to non-similar color vectors. Its values increases depending on q -order selected. This natural property was explored in our approach in order to avoid the use of thresholding and outliers usually present in related approaches [15]. A normalization is made available to visually enhance the discrimination function obtained by the method, given by:

$$I_{(x,y)} = e^{(-\lambda d_{PM}(x,y))} \quad (3)$$

where d_{PM} is the PMD described in the equation 2, and $\lambda > 0$ is a contrast parameter value which is used to refine the polynomial map degree. This normalization result on values varying on a range [0-1]. Similar color vector are as near as possible to 1, and the non-similar ones to 0.

The figure 2 presents the results of the training and classification steps and a brief comparison with Euclidean, MD and PMD metric. This analysis was obtained from one specific image sample, where the training and the classification steps were performed using the parameters $L = 5$ and contrast value $\lambda = 1$.

The left column shows the image sample (in a) and the obtained similarity maps (b, c, d, e and f). In a) it is also possible to observe the input data points defined by the pathologist as white paths (black arrows). The right side shows the bi-dimensional perspectives of the IHC pattern distribution (in a-right). All the others bi-dimensional perspectives are combined from the RGB color space (RG, RB, GB) for better visualization.

In row a) the input data points over the sample image corresponding to the IHC pattern distribution are shown. This distribution is decomposed into their RGB combination as the blue dots at the right side.

In row b) it is presented the Euclidean distance metric to exemplify two typical problems: the needing of a rough thresholding value, and the presence of many outliers. The image at left side shows the similarity of stained areas (white regions) over the IHC pattern defined in a). The increasing or relaxing of the thresholding parameter does not provide a robust classification of stained and non-stained areas.

In row c) it is presented the MD metric (first q -order). It is important to notice that this map also would needs a threshold value in order cut-of IHC areas. At c)-right side it is evident the elliptical surface generated by the covariance of the multivariate input data points, eliminating some outliers points.

In row d) it is shown the 2-order result and their respective polynomial map decomposition. In this q -order is pos-

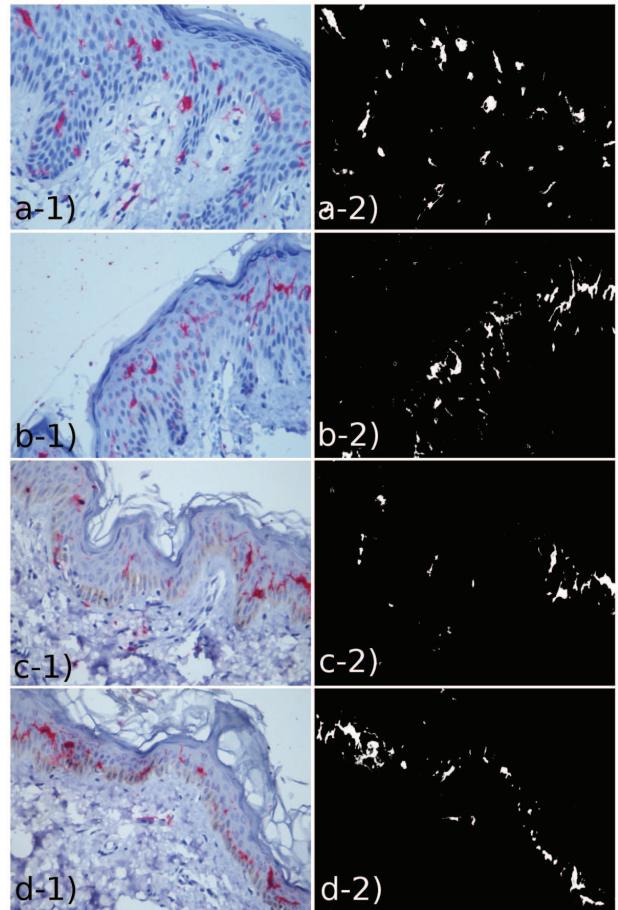


Figure 3. IHC identification results.

sible to obtain a very good discrimination between positive areas (white) and negative ones (black). This polynomial map avoids the needing of a thresholding value, and at this point the next polynomial maps describe nonlinear discrimination surfaces.

In the row sequences (e) and (f), we present the 4-order and 16-order, respectively (2^2 and 2^4). It is clear to note that, the higher the polynomial q -order map, the tighter the distance metric is. In other words, more rigorous the classification process is over the IHC pattern distribution.

Figure 3 depicts the classification results of four IHC samples using the 4-order polynomial map. In the next section we present an evaluation of these and others results.

4 Experimental Setup and Evaluation

Experiment images used to validate the proposed method were done using human skin biopsies submitted to immunohistochemistry using CD1a and iNOS as primary antibodies. The reaction was visualized with Permanent Red Chro-

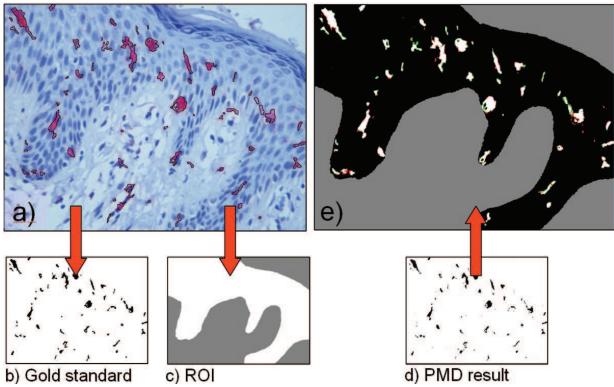


Figure 4. Example of a marking process performed by pathologist for GT generation.

mogen (Dako Carpinteria, California, USA).

Digital images were acquired using the Sight DS-2Mv digital camera connected to a light microscope. From each case, photomicrographs of 1600x1200 pixels of resolution were obtained from no coincident consecutive fields, throughout all the epidermis at x200 magnification (for iNOS antibody) or x400 magnification (for CD1a antibody).

The gold standard evaluation was carried out by one pathologist that manually performed the quantitative and qualitative evaluation of the IHC samples. The positive stained areas were delimited employing a Ground Truth Generator tool (GT-tool) developed specially for general gold standard extraction (GT), and using a digital tablet.

The diagram of figure 4 shows the evaluation process as follows: in a) the pathologist marks the positive stained areas and they are extracted as golden standard pattern in (b). It is also defined the region of interest (ROI) in (c), in which the comparison will be restricted. In this approach, the histological examination was restricted to epidermal region.

The results of the proposed method were performed by an operator independently marking the IHC patterns of the same images using a graphical interface where the PMD method was implemented. The operator's activities were performed into an environment designed to be friendly user. An efficient implementation based on the framework proposed in equation 3 was embedded in the computational tool to obtain very higher polynomial maps. A paired comparison was done on 15 different images and the results are shown in the next section.

The comparison between pathologist and operator was performed using the results of the proposed approach, represented by figure 4-d). The cross-correlation was employed to identify the coincident areas between (b) and (d) in the specified ROI (c). In e) is illustrated the final result of this comparison. The white regions are the true-positive

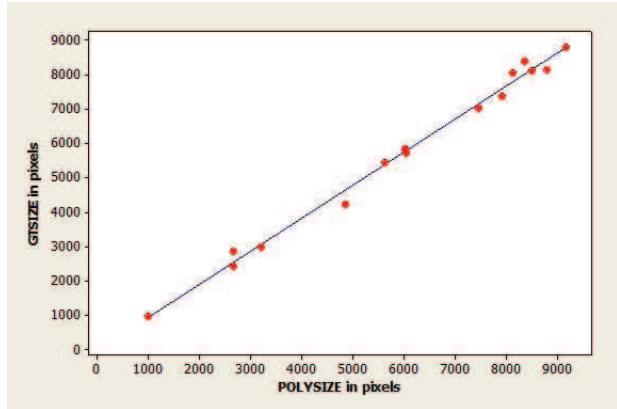


Figure 5. Scatter plot differences between pathologist analysis and PMD analysis.

stained areas, identified by both, the method and by the pathologist. The black region is the false positive regions rejected by the method and by the pathologist. Green areas are false negative estimated by the pathologist and the red ones are the false positive indicated by the method.

5 Preliminary Results

A reduced set of images showing results from the use of proposed method is illustrated on figure 3. As general analysis we can state that the Pearson correlation coefficient between GT marked by pathologist and PMD showed an index of 0.966. The scatter plot between the two methods is illustrated on figure 5. Probability plot of the differences between both analysis (GT-PMD) showed a normal distribution with a mean bias of -265 pixels (STD=239.15, SE=61.74). This bias is probably related to the pathologist's subjective tendency to ignore very small stained individual regions, which then appear as false positives in the validation of the automated method against the manually generated GT. Other indexes obtained by comparing the GT and the proposed approach are the specificity=0.85, sensitivity=0.99 and fitness=0.92.

6 Conclusion and Discussion

In this paper we presented a semiautomatic method for the identification of IHC expression based upon the Polynomial Mahalanobis Distance (PMD). The comparisons against GTs produced by a pathologist's analysis show an excellent correlation with a little bias due to pathologist subjective definition of pattern borders. The proposed approach has an accurate discrimination between stained and

non stained areas of the IHC patterns, avoiding the need of thresholding value commonly used on related approaches. Other advantage is the ability to describe non-linear patterns even in the RGB color space. The PMD method show a good performance for red patterns even in the presence of closely related maroon color, as shown in figure 3.

Qualitatively human operator has the cognitive training knowledge to define IHC patterns. But quantitatively the proposed method has demonstrated an excellent performance on supervised quantification and marking. It is a valuable tool that could be used to reduce the intensive work of time consuming evaluation activities of IHC analysis for a large number of samples. In this paper we perform a preliminary evaluation using CD1a and iNOS biomarkers. As a future work this approach should be evaluated with other IHC biomarkers.

7 Acknowledgements

We would like to thank the grants of FINEP - Financiadora de Estudos e Projetos, SEBRAE -Servio Brasileiro de Apoio s Micro e Pequenas Empresas and CNPQ - Conselho Nacional de Desenvolvimento Cientfico e Tecnolgico.

References

- [1] A.S. Jadhav, S.Banerjee, P.K.Dutta. Quantitative analysis of histopathological features of precancerous lesion and condition using image processing technique. *CMBS 2006*, pages 231–236.
- [2] E.P. Xing, A.Y. Ng, M.I. Jordan, S. Russell. Distance metric learning with application to clustering with side-information. *Advances in NIPS, Cambridge, Mam USA*, pages 505–512, 2003.
- [3] G. Grudic, J. Mulligan. Outdoor path labeling using polynomial mahalanobis distance. *Robotics: Science and Systems II Conference*, 2006.
- [4] J. Gil, H-S. Wu. Applications of Image Analysis to Anatomic Pathology: Realities and Promises. *Cancer Investigation*, 21(6):950–959, January 2003.
- [5] J.A. Kiernan. *Histological and Histochemical Methods. Theory and Practice*. London, 3rd Ed. Arnold, 2004.
- [6] J.C. Jenette. Immunohistology in diagnostic pathology. *CRC Press, Boca Raton*, 1989.
- [7] R. Johson and D. Wichern. *Applied Multivariate Statistical Analysis (4th ed.)*. Upper Saddle River, NJ: Prentice-Hall, 1998.
- [8] J.S. Ross, J.A. Fletcher, K.J. Bloom, G.P. Linette, J. Stec, W.F. Symmans, L. Pusztai, G.N. Hortobagyi. Targeted therapy in breast cancer. the her-2/neu gene and protein. *Mol Cell Proteomics*, 3:379–398, 2004.
- [9] L.Yang, R. Jin. Distance metric learning: a comprehensive survey. technical report. michigan state university. (<http://www.cse.msu.edu/~yangliu1>). 2006.
- [10] N. Chater, J.B. Tenenbaum and A. Yuille. Probabilistic models of cognition: Conceptual foundations. *Special Issue: Probabilistic models of cognition. TRENDS in Cognitive Sciences*, 10(7), 2006.
- [11] Pham et. al. Quantitative image analysis of immunohistochemical stains using a CMYK color model. *Diagn Pathol*, 8(2), January 2007.
- [12] D. S. R.O Duda. *Pattern Classification (2nd ed.)*. John Wiley and Sons, 2001.
- [13] Wu et. al. Semi-automated image processing system for micro-to macro-scale analysis of immunohistopathology: Application to ischemic brain tissue. *Computer Methods and Programs in Biomedicine*, 78:75–86, 2005.
- [14] Xiang S., Nie F., Zhang C. Learning a mahalanobis distance metric for data clustering and classification. *Pattern Recognition*, 2008.
- [15] Z. Theodosiou, I.N. Kasampalidis, G. Livanos, M. Zervakis, I. Pitas, K. Lyroudia. Automated Analysis of FISH and Immunohistochemistry Images: A Review. *Cytometry Part A* 71A, pages 439–450, January 2007.

Apêndice A3

Journal of Applied Meteorology and Climatology. (In Press)

Contribuição da publicação em relação à pesquisa desenvolvida. A publicação ilustra uma metodologia estatística de eliminação de outliers, a utilização da metodologia Bayesiana, a ocorrência dos padrões em um *locus* típico no espaço de cores, a utilização de métodos geométricos na definição e classificação dos padrões, a invariância da luminosidade na determinação dos padrões durante o dia, a utilização de critérios mais rigorosos de classificação dos padrões, modelamento correto do domínio de atuação de um agente computacional no modelamento. Abstract aceito e publicação submetida ao Journal

The Use of Euclidean Geometric Distance on RGB Color Space for the Classification of Sky and Cloud Patterns

SYLVIO LUIZ MANTELLI NETO

Earth System Sciences Center (CCST-INE), National Institute for Space Research, São José dos Campos, São Paulo, and Knowledge and Engineering Department (EGC), and Image Processing and Graphics Computing Lab (LAPIX), and Solar Energy Lab (LABSOLAR), Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil

ALDO VON WANGENHEIM

Image Processing and Graphics Computing Lab (LAPIX), Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil

ENIO BUENO PEREIRA

Earth System Sciences Center (CCST-INE), National Institute for Space Research, São José dos Campos, São Paulo, Brazil

EROS COMUNELLO

University of Itajai Valley (UNIVALI), São José, Santa Catarina, Brazil

(Manuscript received 1 July 2009, in final form 17 November 2009)

ABSTRACT

The current work describes the use of multidimensional Euclidean geometric distance (EGD) and Bayesian methods to characterize and classify the sky and cloud patterns present in image pixels. From specific images and using visualization tools, it was noticed that sky and cloud patterns occupy a typical *locus* on the red-green-blue (RGB) color space. These two patterns were linearly distributed parallel to the RGB cube's main diagonal at distinct distances. A characterization of the cloud and sky patterns EGD was done by supervision to eliminate errors due to outlier patterns in the analysis. The exploratory data analysis of EGD for sky and cloud patterns showed a Gaussian distribution, allowing generalizations based on the central limit theorem. An intensity scale of brightness is proposed from the Euclidean geometric projection (EGP) on the RGB cube's main diagonal. An adapted EGD-based method was compared with existing classification methodologies, which limit the color-space domain examined. Elimination of this limitation was considered a sufficient criterion for a classification system that has resource restrictions. The EGD-adapted results showed a correlation of 97.9% for clouds and 98.4% for sky when compared to established classification methods. It was also observed that EGD was able to classify cloud and sky patterns invariant to their brightness attributes and with reduced variability because of the sun zenith angle changes. In addition, it was observed that Mie scattering could be noticed and eliminated (together with the reflector's dust) as an outlier during the analysis. Although Mie scattering could be classified with additional analysis, this is left as a suggestion for future work.

AU1

1. Introduction

Automatic cloud evaluation from the surface is an important issue in meteorology to reduce subjective aspects and operational costs of synoptic observers (SO). Several research groups are demanding new techniques

for automatic cloud and sky detection to replace SO using automatic cameras (World Climate Research Program 2007). Substituting the SO evaluation with the automatic system defined by the World Meteorological Organization (2008) as a synoptic observation system (SOS) is not a trivial task. It involves aspects of human perception, atmospheric sciences, mathematics, computer artificial intelligence, etc., in the design of an "artifact" called an intelligent agent (IA; Russell and Norvig 2003, chapter 2). The existing image analysis artifacts used as SOS do not match the qualitative performance of SO, and to find

Corresponding author address: Sylvio Mantelli, UFSC-EMC-LABSOLAR, Campus Trindade, Florianopolis SC Brazil, 88040-900.

E-mail: sylvio@lepten.ufsc.br

better solutions, improvements to the classification techniques must be made.

SO observations are normalized by the World Meteorological Organization (2008, section I-15). These observations usually describe cloud type and amount, but the evaluation is highly subjective. With regards to cloud amount, for instance, inconsistencies exist when the same sky is evaluated by distinct operators (Hoyt 1978). The packing effect is another inconsistency, which is caused by an overestimation of the amount of clouds near the horizons (Holle and Mackay 1975). The natural human lack of consistency, due to the operator's fatigue and the effects of shifts of observation teams, stimulate the use of automatic systems. However, most camera-based SOS methods are still compared to SOs for their validation on the qualitative analysis of sky conditions (Souza-Echer et al. 2006; Long et al. 2006).

Clouds can be evaluated from satellite images, avoiding some of the previously mentioned problems but introducing others (Rossow 1982)—such as the mentioned pixel geometric distortion, clouds that cannot be detected between layers because of the vertical distribution, seasonal surface variation, etc. Any satellite-based assessment (e.g., energy balance, temperature, radiation, wind, clouds, etc.) must take into account the surface observations' "ground truths" (GTs) to reduce modeling uncertainties. Furthermore, there is a consensus that for cloud evaluation a complementary observation will be the best way to reduce the limitations of both techniques.

SOs are trained to develop cognitive skills for sky pattern identification, but perform poorly when determining the precise amount of clouds. Camera-based SOSs have a better performance when determining cloud amounts than a surface observer, but perform worse for pattern identification. A Camera-based SOS also relies on methodologies like simplified dimension thresholding (Souza-Echer et al. 2006) and the reduction of multivariate color spaces (Long et al. 2001). Those were the only methodologies found in the literature that used image analysis. Short wave, long wave, and other sensors used for that purpose are out of the scope of the current work, but they could be investigated as a cross-comparison analysis in the future. The present work will consider only the comparison of the two equivalent methods existing in the literature for surface image analysis.

Souza-Echer et al. (2006) used a flat image with a 62° field of view (FOV) camera on a zenith mount, always avoiding the direct sun light. Only the saturation dimension of hue, saturation, and lightness (HSL, cylindrical coordinates) was used on the characterization of the three patterns: sky, clouds, and a third class obtained by exclusion. The discrimination function to classify sky and clouds is based on three standard deviation level thresholding

from the pattern average. Only these three patterns and their amount are produced by this approach. The elimination of the sun from the observation domain restricts the analysis to small brightness patterns only.

Long et al. (2006) employed a different criteria and an experimental setup using two pieces of equipment: the Total Sky Imager (TSI) and the Whole Sky Camera (WSC). TSI uses an image from a reflected mirror and WSC a direct image of the sky, both with a 160° FOV. A detailed description about the experimental setup and analysis is provided by Long et al.'s (2006) paper and in Long et al. (2001). Although the cameras obtained images in a 24 bits pixel⁻¹ red-green-blue (RGB) file format, the classification is restricted to 0.6 threshold R-B dimensions ratio (Long et al. 2006). This criterion reduces the domain color analysis from black to magenta only (Gonzales and Woods 2002), discarding any reference or additional data that could be gathered from the green channel, which might help to classify or analyze further information of atmospheric patterns seen from images. The paper also points out the difficulty of identifying small differences in patterns due to atmospheric contents. Brighter-blue pixels representing blue skies in the transition between molecular scatterings to turbidity are difficult to classify with the proposed method, and they probably would be difficult to detect from a reduced dimension thresholding classification (Mantelli Neto 2001, 2005).

The aforementioned methodologies represent some important pioneering techniques aiming at the replacement of sky-state observers in meteorological stations. However, both methodologies do not use all the possibilities available for image analysis. A 24-bit image allows $2^{24} = 16\,777\,216$ different color combinations that can be grouped, analyzed, and combined with a great potential to be explored as a domain. Souza-Echer et al. (2006) use only 8 bits or $2^8 = 256$ lightness possibilities to identify clouds. Long et al. (2001) use 16 bits or $2^{16} = 65\,536$ colors. Clouds, however, are white with equally likely components of red, blue, and green. Lillesand and Kiefer (1994) define clouds as nonselective, equally scattering all color components. Our work also presents a new point of view on sky pattern analysis based on a Bayesian methodology (Tenenbaum et al. 2006; Chater et al. 2006; Russell and Norvig 2003, chapter 20; National Institute Of Standards 2006, section 1) to improve surface automatic observation of the atmosphere. AU10

The methodology of the present work uses multivariate color space features (Johnson and Wichern 2007) to classify clouds and sky pixels by means of a pattern statistical characterization using the Euclidean geometric distance (EGD). It also intends to propose a scale based on the brightness projection from the Euclidean geometric

projection (EGP) value of pixels on the RGB diagonal cube.

The next sections will show in a Bayesian approach the exploratory data analysis (EDA) of the pixel pattern domain and the mathematical approach to find the geometric position of the target patterns. From that approach, a solution was implemented on a graphical user interface (GUI) and input images were analyzed using the proposed method. A preliminary analysis and a color-space dimension reduction have been made to allow a comparison with other methods.

2. Material and methods

a. Experimental setup and preliminary analysis

Images were collected using a commercially available sky imager (TSI-440, available online at <http://www.yesinc.com/products/data/tsi440/index.html>) in standard Joint Photographic Experts Group (JPEG) file format at a 352×288 resolution with 24-bit colors. Images were obtained not directly from the sky, but by a dome-shaped reflector every 15 min from dawn to dusk, according to TSI program parameters. The reflected image represents a 160° hemispheric angle of view. An adhesive, moving shading band was applied on the reflector surface to avoid damage on the camera by direct exposition to sunlight. A preliminary image processing was performed on images to eliminate spurious and systematic patterns that might interfere in the image analysis (Montgomery 2005). Three patterns are defined as spurious: border effects, horizontal obstructions, and the moving shading band. Borders are not relevant for image analysis because they represent the equipment's self image. Horizontal obstructions are fixed objects like poles, towers, trees, buildings, and geographic features that are present in the image and do not represent any useful information. The equipment's moving shading band was also masked. Obstructions and the horizon have a fixed position and are easily masked on the image. But the shading band is a dynamic feature that moves according to the solar movement and for every image. A different mask file was obtained according to each specific time position. Masking was performed by hand-marking patterns pixel by pixel using the software GT generator tool developed by the research group. The output of the GT tool is a black mask in a bitmap (BMP) file. Mask files are loaded dynamically by the GUI software tool during the analysis phase for every input image.

The images were obtained in the facilities of the Solar Energy Laboratory of the Federal University of Santa Catarina, Florianopolis, Brazil, (LABSOLAR, available online at <http://www.lepten.ufsc.br/>) located at $27^\circ 32' S$, $48^\circ 31' W$. The site is also a Baseline Surface Radiation

Network (BSRN, available online at <http://www.bsrn.awi.de/>) station site. A set of preliminarily images were analyzed on the RGB and HSL color spaces using the Color Inspector 3D visualization tool (available online at <http://xxxxxxxx>). AU2

During visual inspection of several collected images, a typical *locus* of cloud and sky patterns can be noticed in the color space. The typical locus related to pattern presence is illustrated in Fig. 1 for a cloudy sky and Fig. 2 F1|F2 for a blue sky. Color inspector 3D showed that cloud pixels are typically gray and white, distributed linearly, and closely parallel to the RGB diagonal cube. It is also possible to notice a luminous gradient-scale distribution along the main diagonal. Sky patterns also showed a linear behavior in the RGB cube, but pixels were located a bit farther from the RGB diagonal. Sky and cloud patterns of the same images showed a nonlinear behavior in the HSL color space. Based on those observations, it was decided to perform mathematical operations in the RGB color space, avoiding the HSL nonlinearity. But some analyses were still performed on HSL because its representation is more closely related to the human perception of colors. The other color spaces were not considered.

b. Cloud and sky pattern characterization with exploratory data analysis

A preliminary statistical analysis was used on EGD to characterize the target patterns identified visually at two different distances from the RGB main diagonal. The selected images represent typical samples of cloud and sky patterns. The same original cloud and sky images seen on Figs. 1 and 2 (and Figs. 4 and 5) were used as reference ground truths (Jiang et al. 2006) for those two patterns (Fernandez-Garcia et al. 2008). Then an EDA was performed for each pattern to determine their typical EGD from RGB cube diagonal. F3

To explain the EGD method, a generic pixel in the RGB color space is shown in Fig. 3. the pixel distances can be determined by considering them as vector modules. Projection and distance of pixels from the main diagonal can be easily calculated by

$$\text{PROJ} = |\bar{\mathbf{D}}| \cos(a)$$

$$\text{DIST} = |\bar{\mathbf{D}}| \sin(a).$$

By means of the Al-Kashi theorem (also known as the law of cosines) and the sum of vectors formula we obtain

$$|\bar{\mathbf{C}}| = \sqrt{255^2 + 255^2 + 255^2} = 441.673$$

$$|\bar{\mathbf{A}}|^2 = |\bar{\mathbf{D}}|^2 + |\bar{\mathbf{X}}|^2 - 2|\bar{\mathbf{D}}||\bar{\mathbf{C}}| \cos(a)$$

$$\bar{\mathbf{C}} = \bar{\mathbf{D}} + \bar{\mathbf{A}},$$

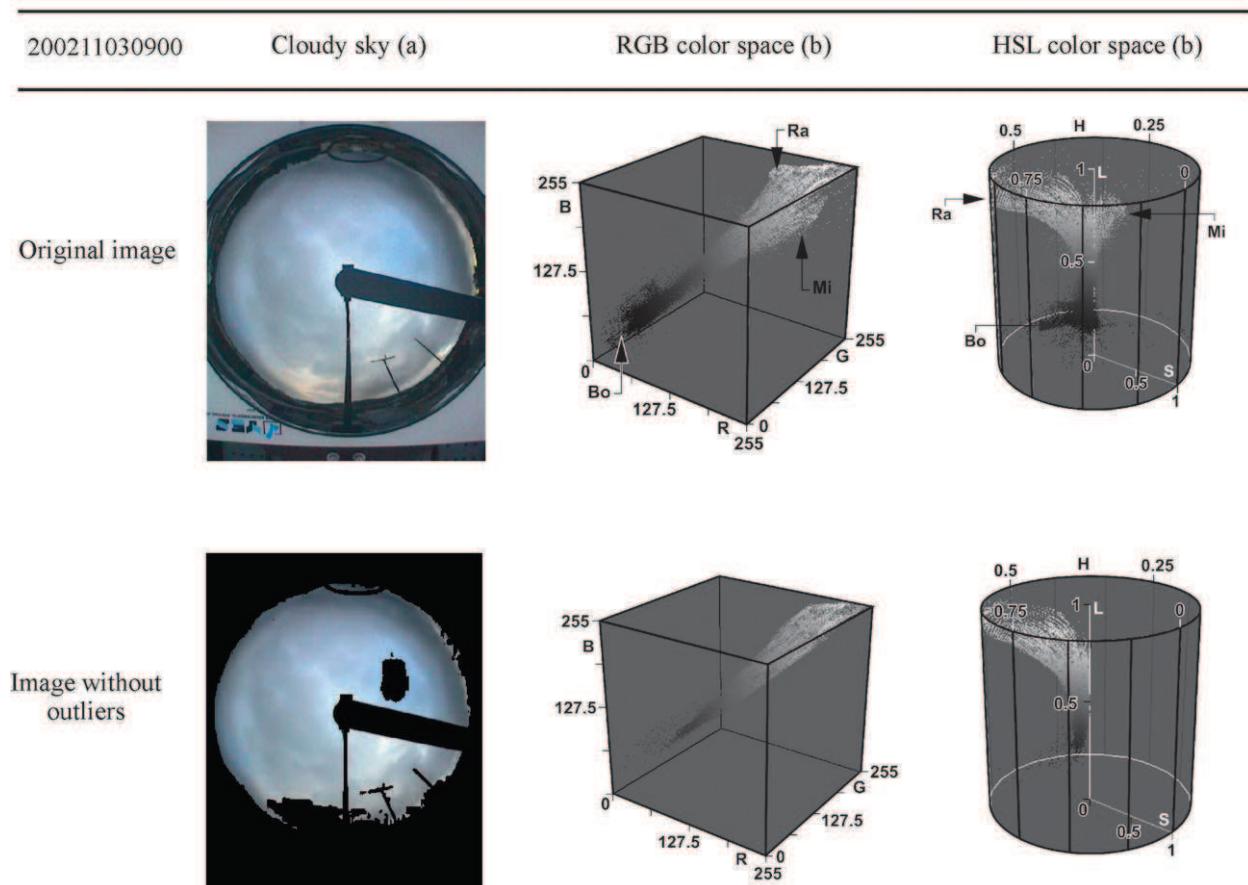


FIG. 1. (a) The typical locus of the cloud patterns observed in images are shown in (b) RGB color-space and (c) HSL color space. The first row is related to the original image and the second row to masked images without outliers. Outliers were masked to black. “Ra,” “Mi,” and “Bo” labels on the color space indicate, respectively, the typical locus of Rayleigh- and Mie-scattering patterns, the equipment’s border, and [AU5] the shading band. Units are in pixel relative intensity for column (b) and pixel-normalized relative intensity for column (c).

where

- a is the angle between the analyzed vector pixel value and the main diagonal;
- $\bar{\mathbf{A}}$ is the complementary vector from the color pixel to the RGB cube vertex;
- $\bar{\mathbf{C}}$ is the color cube main diagonal vector, with a value of (255 255 255), and
- $\bar{\mathbf{D}}$ is the pixel having (R, G, and B) values.

Replacing $\bar{\mathbf{A}}$ in the Al-Kashi theorem and rearranging the formula leads to

$$\cos(a) = \frac{|\bar{\mathbf{C}} - \bar{\mathbf{D}}|^2 - |\bar{\mathbf{D}}|^2 - |\bar{\mathbf{C}}|^2}{-2|\bar{\mathbf{D}}||\bar{\mathbf{C}}|}.$$

The formulation above demonstrates that only the pixel value is necessary to calculate its distance and projection in the main RGB diagonal. The distance values’

text files were generated in a GUI interface from images and loaded into a commercially available statistical analysis tool. After being analyzed, the distances of the two patterns were considered as normally distributed continuous variables, with a typical mean and variance distance from the cube’s diagonal. Image characterization for the cloud pattern obtained from RGB coordinates are illustrated in Fig. 4.

To eliminate outlier points, the interquartile distance (IQD) range was used (which is the distance between the upper and lower quartile of data, available online at <http://www.itl.nist.gov/div898/handbook/eda/section3/boxplot.htm>). The range is obtained from the calculated points of the lower 25% (Q1) and the upper 75% (Q3) of data statistics. Points outside the inferior and superior interquartile limits (IQI and IQS, respectively) of the average point were eliminated. A summary of the equations used in outlier elimination is described below:

[F4]

[AU3]

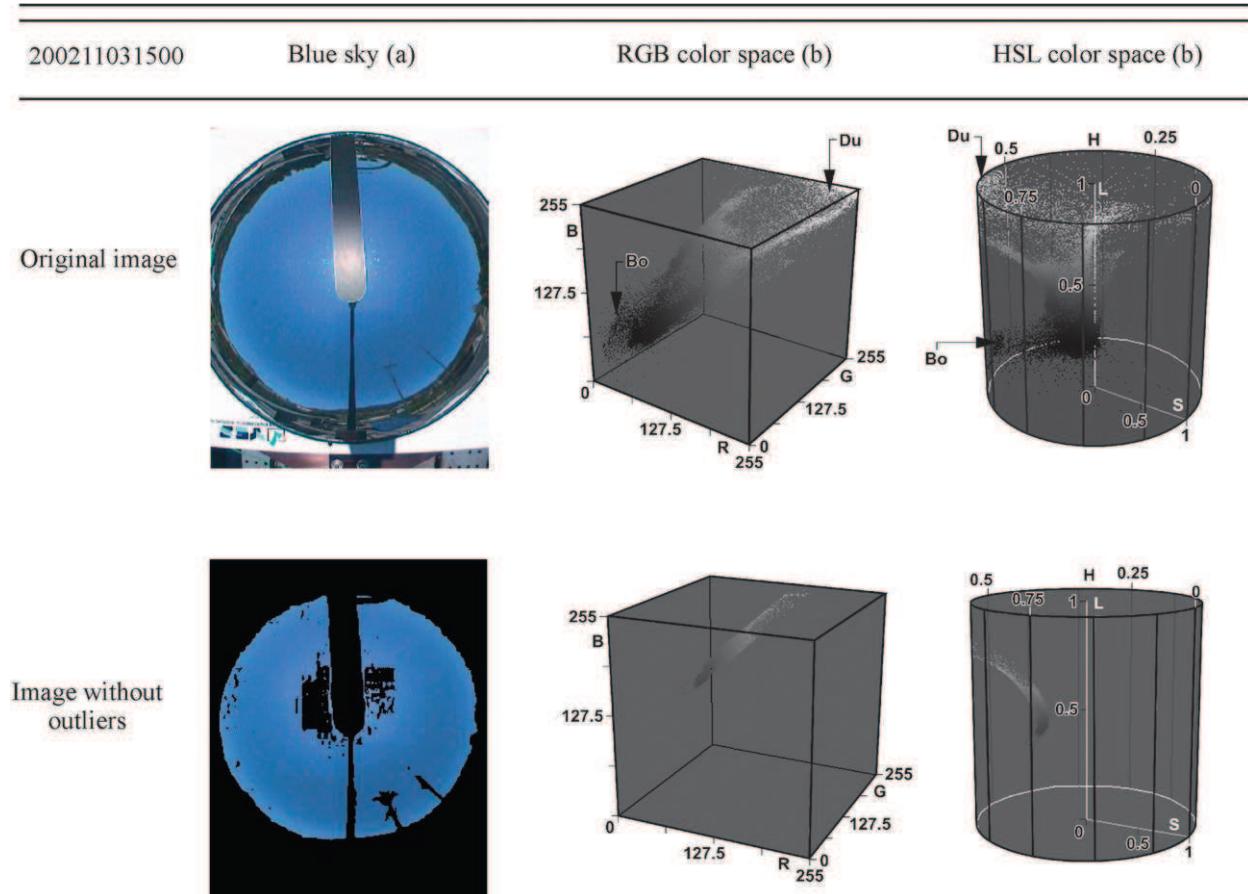


FIG. 2. (a) The typical locus of sky patterns observed in images are shown in (b) RGB color-space column and (c) HSL color space. The first row is related to the original image and the second row to masked images without outliers. Outliers were masked to black. “Du” and “Bo” on the color space indicate, respectively, the typical locus of Dust pattern and the equipment’s border and shading band. Units are [AU6] in pixel relative intensity for column (b) and pixel normalized relative intensity for column (c).

$$IQD = Q_3 - Q_1 \quad IQD: \text{Interquartile distance},$$

$$IQI = Q_1 - 1.5 \times IQD \quad IQI: \text{Inferior inter quartile limit, and}$$

$$IQS = Q_3 + 1.5 \times IQD \quad IQS: \text{Superior inter quartile limit.}$$

Figure 4 illustrates the statistical analysis that was performed with the pixel values extracted from masked cloud images and the pixel values converted into text files. Outlier extraction is made in the following way by using a statistical analysis

$$IQD = Q_3 - Q_1 = 12.75 \quad \text{and}$$

$$IQS = Q_3 + 1.5 \times IQD = 45.315.$$

Cloud outlier pixels were discarded if distance > 45.315 . Only the superior interquartile limit value was considered

because the distribution of clouds pixels starts in the main diagonal. After excluding the outlier values from cloud images, the remaining pixels from the original image can be visualized on the second row of Fig. 1. A summary for cloud pattern limits is illustrated in Table 1.

For the sky pattern, pixel values from the masked image were extracted and converted into text files to perform the statistical analysis shown in Fig. 5. From analysis of the figure, a statistical outlier extraction was developed using the following conditions:

$$IQD = IQ_3 - IQ_1 = 8.26,$$

$$IQS = IQ_3 + 1.5 \times IQD = 89.240, \quad \text{and}$$

$$IQI = IQ_1 - 1.5 \times IQD = 56.200.$$

Clear-sky outlier pixels were discarded if they were not in the interval $56.200 \leq \text{distance} \leq 89.240$. After excluding outliers from the image, the remaining pixels can be visualized in the image of Fig. 2 in the second row.

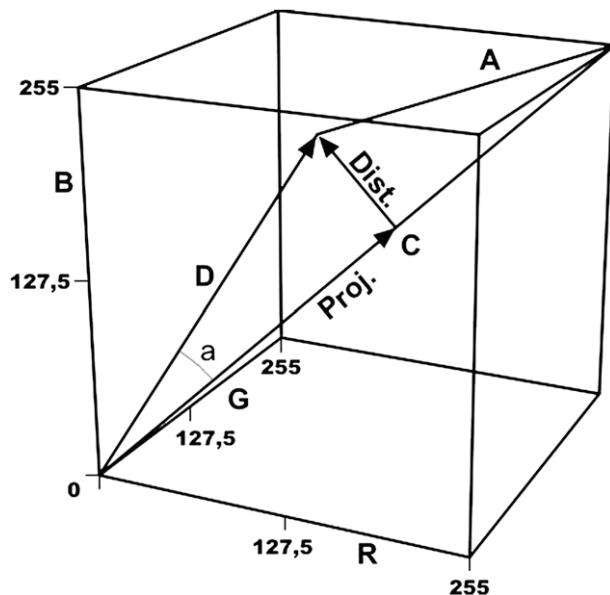


FIG. 3. Typical analysis used on generic pixel determination of EGD on RGB color-space diagonal. Units are in pixel relative intensity.

A discrimination boundary within a 3 standard deviation from the average was established for both patterns to include 95% of pixels in the classification. A summary of boundary values used for sky patterns is illustrated in Table 1. As can be noticed in Table 1, there is an overlapping region of distance limits between 55.30 and

54.01, resulting in error type I (classification of clouds as Rayleigh) or type II (classification of Rayleigh as clouds) according to the pattern analyzed. Type I and II errors are due to misclassification and are caused by the superposition of the end tails of two neighboring Gaussian distributions. A “real visual boundary” between cloud and sky patterns in the image is defined here by statistical exclusion. This overlapping region was considered an intermediate pattern, tagged as the “I” category classification and colored brown Fig. 6c.

F6

c. Brightness scale

Pixel EGPs for cloud or sky patterns were ranked in submultiples of the RGB cube’s main diagonal maximum value. These EGP values were divided into six slots, in categories C1–C6 for clouds and R1–R6 for Rayleigh scattering. Categories were selected based on previous work done by the author (Mantelli Neto 2005). This work conceptually considered blue sky as shades of the Rayleigh-scattering effect (Iqbal 1983; Lenoble 1993); clouds as nonselective scattering of solar light in white (Lillesand and Kiefer 1994); and the total simply by sky. EGP values increase from “darker” to “brighter” according to their brightness value, as shown in Table 2. Shades of green were employed for the sky and shades of red for clouds to avoid color confusion with the original patterns for the results presented in Fig. 6. The discrimination between clouds and Rayleigh scattering was done

T2

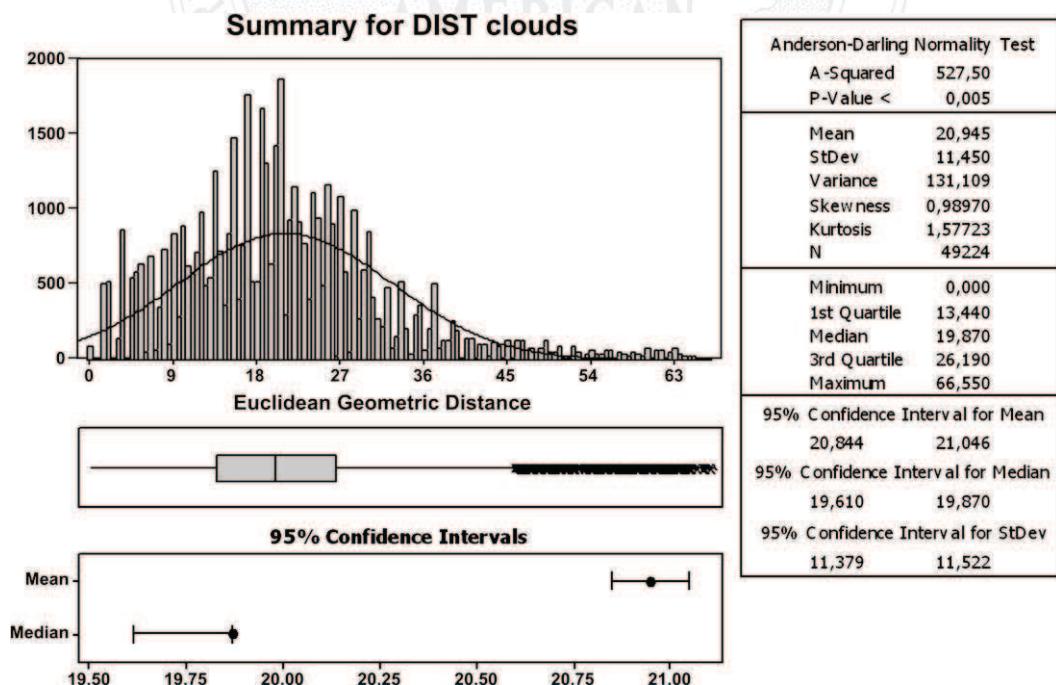


FIG. 4. EDA used for the characterization of the cloud pattern. It includes EGD statistical summary, histogram, box plot, and confidence interval. Horizontal-axis units are in pixel intensity normal to RGB cube’s main diagonal.

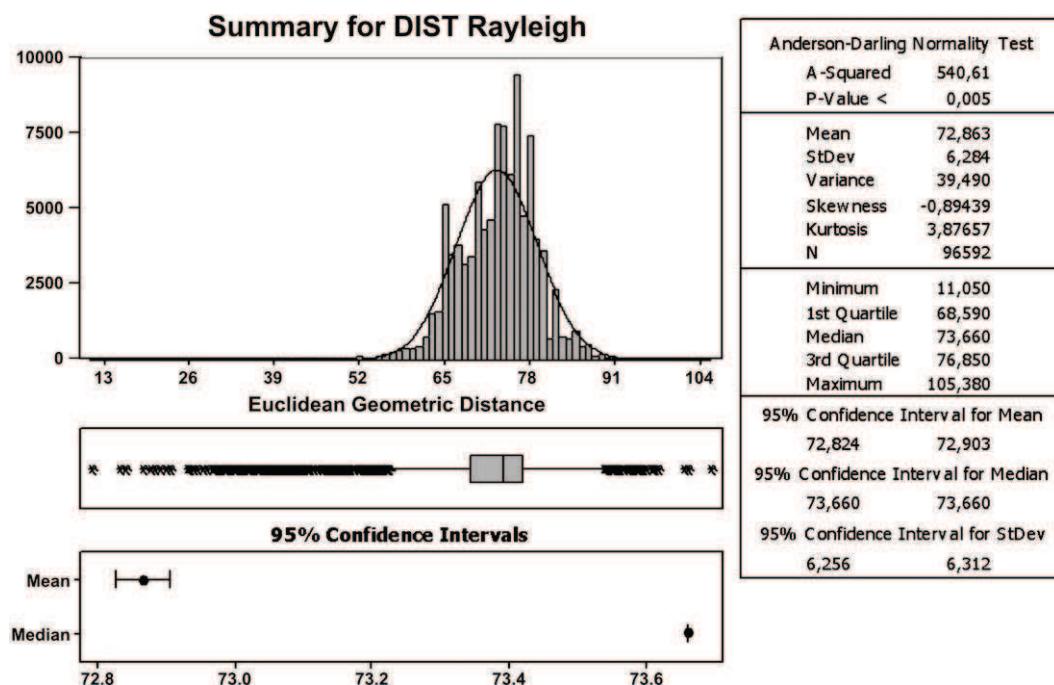


FIG. 5. EDA used for the characterization of the Rayleigh pattern. It includes EGD statistical summary, histogram, box plot, and confidence interval. Horizontal-axis units are in pixel intensity normal to RGB cube's main diagonal.

by statistical methods based on the Gaussian distribution. All the assumptions made on statistical hypothesis tests for additional pixel values were used in the discrimination analysis and have the same meaning. After the characterization of cloud and sky patterns, their typical geometric information was implemented as confidence intervals and loaded into the GUI for pixel classification of other images. Some results of image classification are presented in Fig. 6 and analyzed in the next section.

3. Results

After supervised learning, a set of 49 images relative to one day of observations made on 11 March 2002 every 15 min were analyzed using EGD (the 1900 LT image was missing). Results are presented as the percentage of coverage for each pattern. Images were obtained from output files generated by the GUI tool. A general summary showing pattern data can be observed in Fig. 7, describing the diurnal variability of sky patterns. Data shows dominant categories of pixels mostly between R3

and R6 for sky and C3 and C6 for clouds for the specific image set observed. An increased amount of "I" patterns occur mostly in mixed conditions because a threshold (or frontier) between cloud and sky patterns is not formally defined but determined by statistical parameters.

It is important to keep in mind that clouds are not purely white or gray; otherwise, they would have a distribution of points concentrated around the RGB cube's main diagonal or around the HSL cylinder's main axis. The following could explain why they contain some amount of color. Possible colors could be blue due to Rayleigh scattering or red-orange due to Mie scattering. This fact could also be confirmed with additional verification using the HSL color space, as shown in Fig. 8. Rayleigh and Mie effects are present in different sectors of the HSL color-space angle as can be seen in Fig. 8b. For a clear sky, the blue color is limited to a well-defined sector as can be seen in the figure's first row. For a cloudy sky pixel values span one sector with a component of blue and in another sector with red-orange (circled in dashed on second row). For the TSI used in the experiment, there

F8

AU8

TABLE 1. Summary of cloud and sky population patterns limits.

Pattern	Mean	Std dev	N	Q1	Q3	Distance lower limit $\mu - 3\sigma$	Distance upper limit $\mu + 3\sigma$
Cloud	20.945	11.45	49224	13.44	26.19	0	55.30
Sky	72.863	6.284	96592	68.59	76.85	54.01	—

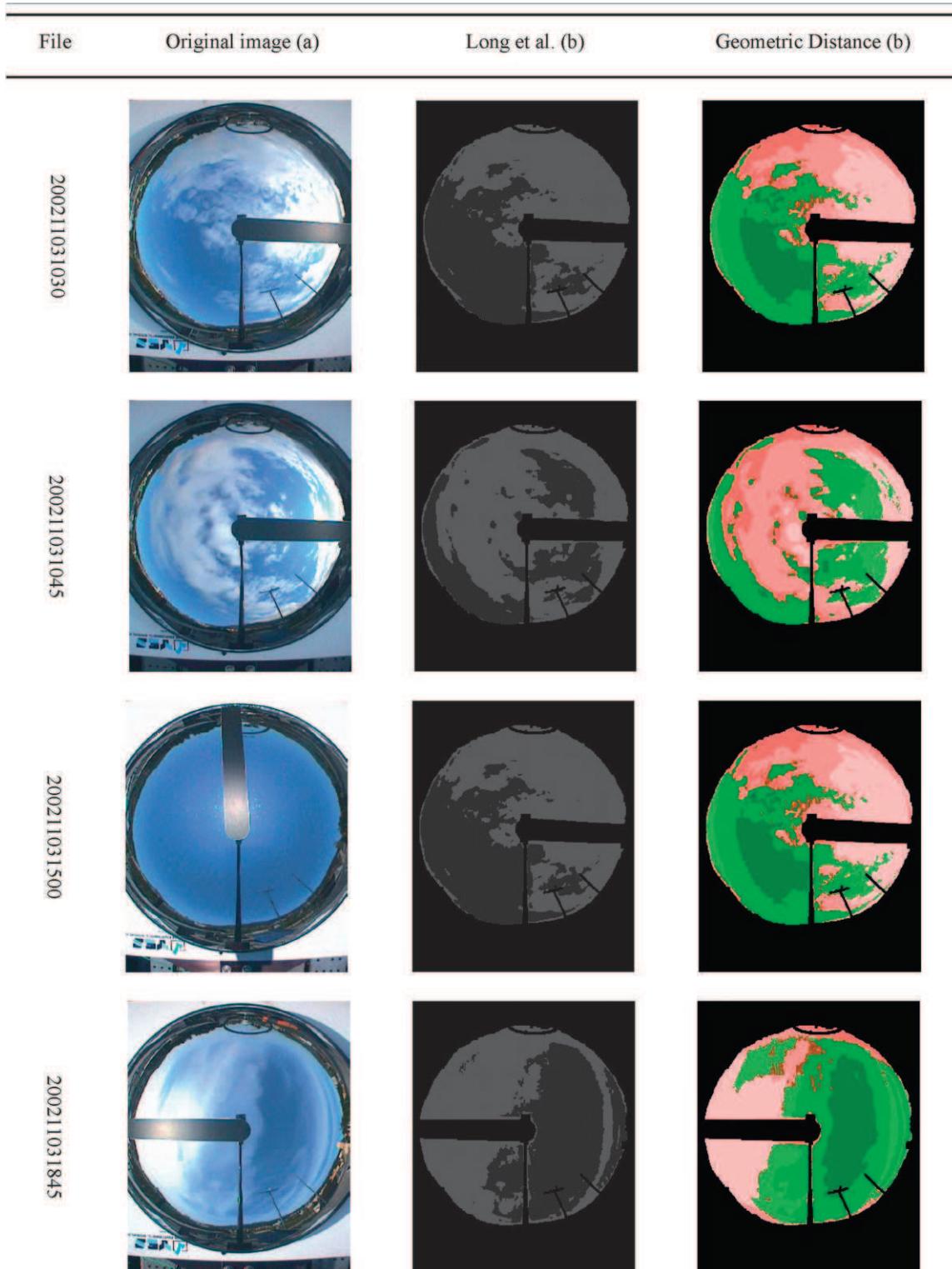


FIG. 6. (a) Original images were compared to (b) images analyzed using Long's method and (c) the geometric distance method. In column (b), clouds were marked in clear gray, and skies were marked in dark gray. In column (c), clouds were colored in shades of red and skies in shades of green. Intermediate patterns were colored in brown in (c).

TABLE 2. Assignments for the pixels' projection values into category slots.

Assigned category	R1 or C1	R2 or C2	R3 or C3	R4 or C4	R5 or C5	R6 or C6
Projection value	0–75	75–150	150–225	225–300	300–375	375–442

are some limitations on Mie-scattering observations because the software takes images only above 5° of solar elevation, thus avoiding the major occurrences of Mie scattering.

Noteworthy here is the robustness of the characterization method when eliminating outliers. Although the elimination of pixels in the middle of the cloud pattern (the second row of Fig. 1) might seem strange, it has an explanation. By using Color Inspector 3D, it was observed that Rayleigh and Mie scattering were intermixed in the original images of clouds, as explained in the previous paragraph. That feature helps to identify and eliminate these patterns from cloud images.

The Rayleigh scattering observed in Fig. 7 shows a diurnal variation with a clearly noticeable gradient in the brightness. It is speculated that this gradient could be associated to other atmospheric contents (aerosols or water vapor), causing diffusion of sun light. For an observer on the surface, Rayleigh scattering is brighter at higher zenith angles than lower ones, indicating that brightness is more intense near the surface (R5) than the lower zenith angles (R3). This is probably due to the diffusion of sun

light due to a higher concentration of atmospheric constituents near the surface.

EGD is a distance whose values span from overcast (near the main diagonal) to clear sky, defined here as Rayleigh scattering independent of sun-light variation. This indicates that EGD could also be used to provide information on sky clearness or a similar index. EGD and pixel EGP on the main diagonal (brightness) could be used to support the spatial evaluation of radiation processes from the atmosphere toward the surface. The advantage of the EGD-derived values employed in this analysis is the instantaneous temporal resolution in contrast to indexes obtained from a long temporal series of data. However, further evaluation and careful investigation should be performed to validate and convert pixel values to physical units.

4. Comparisons to related work

There is a limitation in comparing the current methodology with the results of other related works because

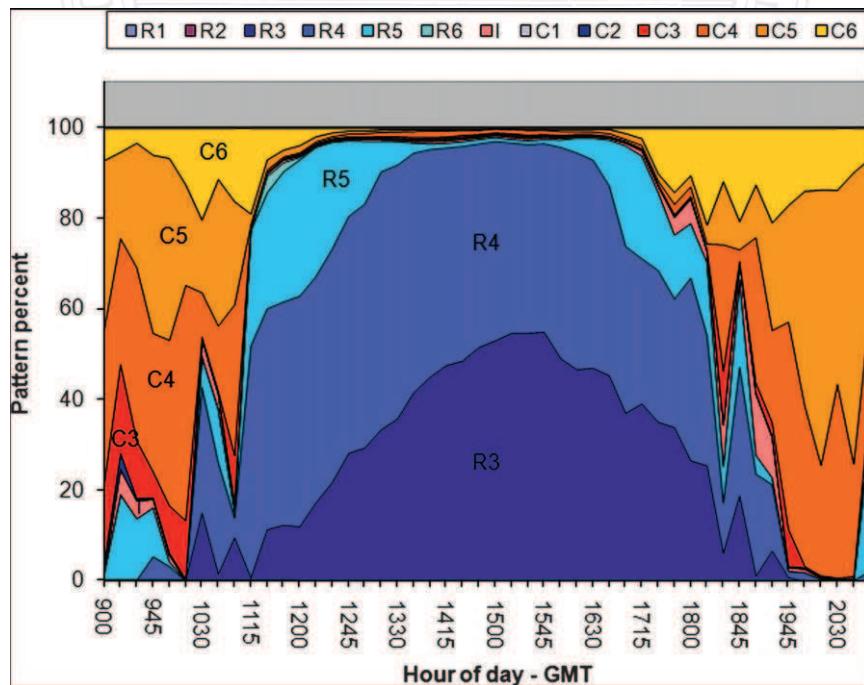


FIG. 7. Sky pattern graph showing the proportion or fraction of sky patterns obtained by the geometric distance method vs time of day. Images were taken from 0900 to 2100 LT every 15 min. The legend categories are in Table 2.

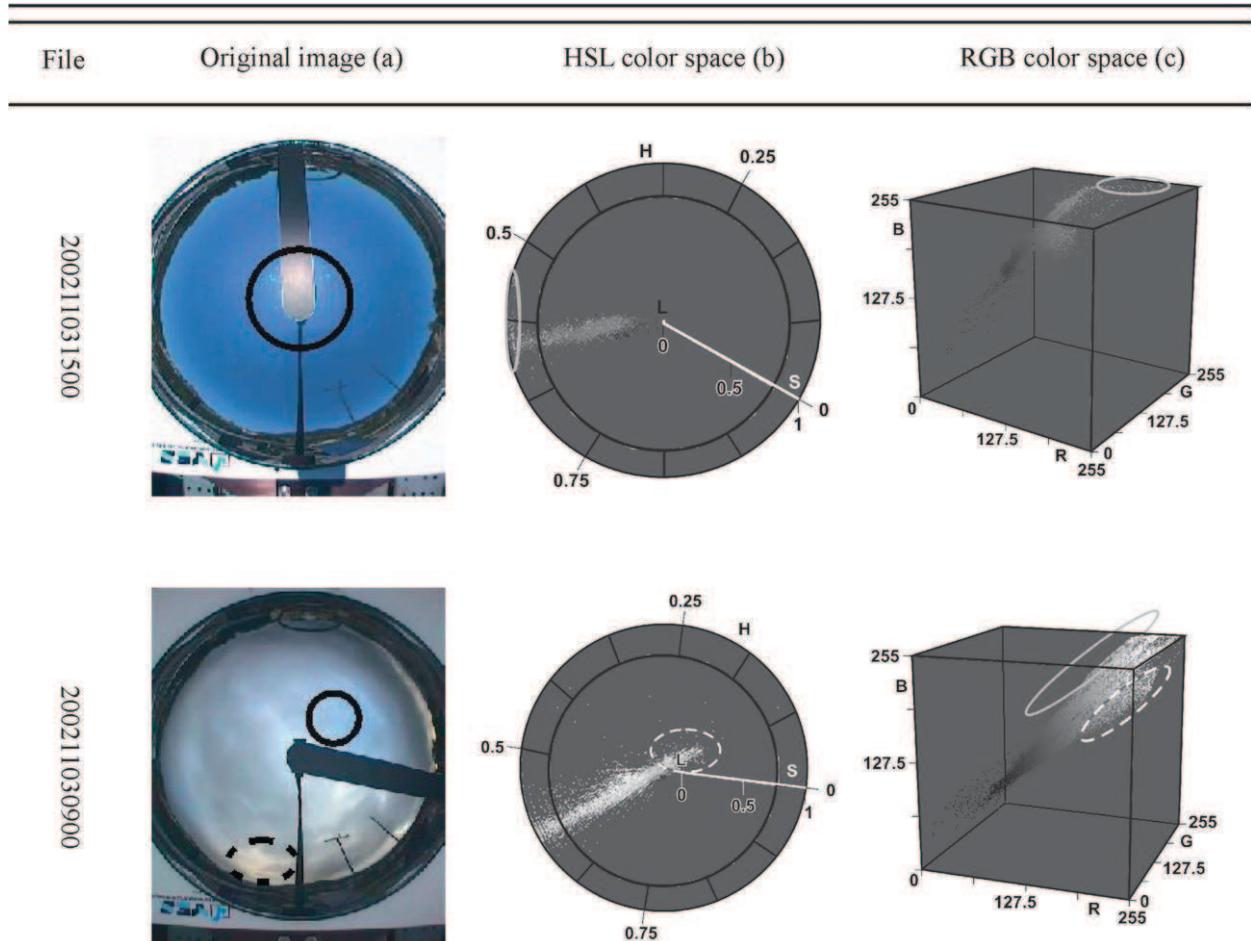


FIG. 8. The first line is used as a comparison showing Rayleigh scattering and the effect of the reflector's surface dust on color space and image, outlined in solid and dashed lines. The second line shows (a) obstruction-masked original images with identification of Mie scattering in dashed lines and Rayleigh scattering in solid lines intermixed in the cloud pattern. (b) The respective HSL color space reduced to H, S dimensions only showing the presence of Mie scattering in dashed lines is illustrated. Note that the Rayleigh scattering was superimposed by cloud pixels in that perspective view and was not outlined. (c) The respective RGB color space showing typical locus of Rayleigh scattering in solid lines and Mie scattering in dashed lines is illustrated.

of distinct experimental setups and domains used for the desired outcomes.

In Souza-Echer et al. (2006), images were taken from a common camera with observations close to small zenith angles and the sun was always kept out of the FOV. In that condition, pixels representing blue and cloudy skies with high bright values were not analyzed. In high-brightness conditions the saturation values for cloud/sky patterns are mixed with each other, making the discrimination based only on saturation values incomplete. In those conditions, the discriminating function adopted by Souza-Echer et al. (2006) works only on low-brightness pixel values. This is illustrated in Figs. 9a and 9b, where 9a represents an overcast sky and 9b a clear sky. Only the hue and saturation dimensions of the HSL color space are shown, and saturation discriminating values as proposed

by Souza-Echer et al. (2006) could be seen as concentric circles. In Fig. 9, the angle H represents different colors on the image domain, and the radius is its saturation S . High-brightness pixels for sky and clouds are mixed with each other in the same saturation radius, leading to misclassification errors. Therefore, because of the limitation of Souza-Echer et al. (2006) with high-brightness pixels, the classification was not implemented in the GUI tool to be compared with the proposed EGD method.

Long et al.'s (2006) methodology used the same experimental setup as the one presented in this paper but considered only two of the three dimensions available for data analysis. The domain used by Long et al. (2006) is the two-dimensional (2D) coordinates using the red and blue channels and not considering the green channel features. This approach is distinct from the approach used for the

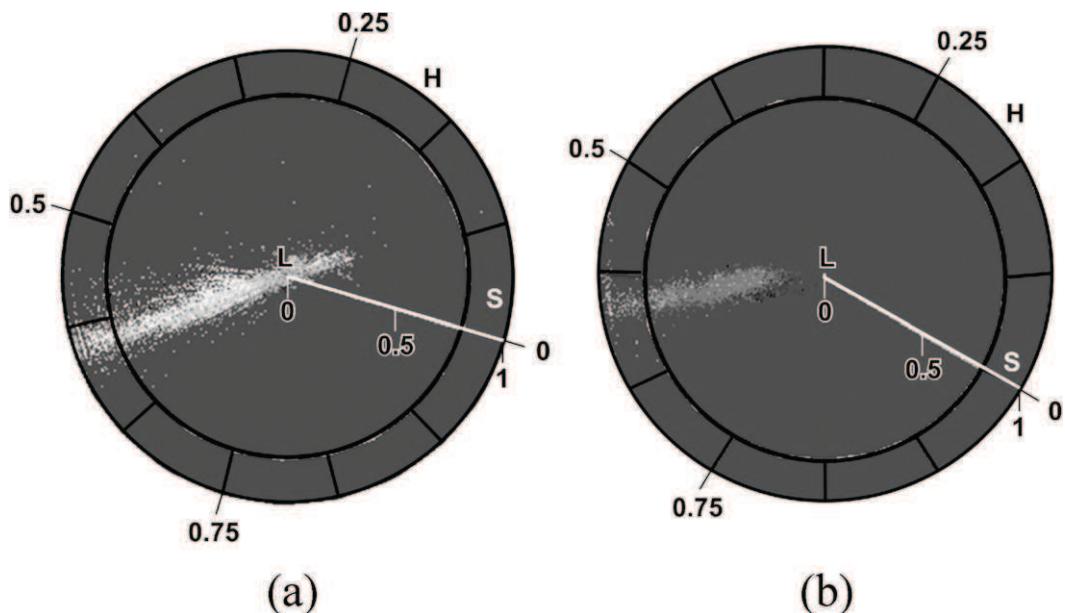


FIG. 9. Differences in the pixel distribution for a cloudy sky from Fig. 1a and a clear sky from Fig. 2b reduced to S and H dimensions of HSL color space. Circles are placed by the visualization tool and are an example illustrating the type of discrimination function used by Souza-Echer et al. (2006) for each case. Units are in pixel-normalized relative intensity.

F10 present paper, which uses a three-dimensional (3D) domain. We applied the same methodology as Long et al. (2006) to try and reproduce the same technique for the detection of clouds and clear sky. The histograms illustrated in Fig. 10 with the distribution of values are related to the same cloud and sky patterns images of Figs. 1, 2, 4, and 5. As in Long et al.'s (2006) original paper, the same threshold of 0.6 was used for the separation of the cloud and clear-sky patterns.

From that implementation, a selected group of images was analyzed and the comparative results are shown in Table 2. Only a few cases were shown because of space constraints. A complete set of data, including all results and image analyses, can be seen on the Image Processing and Graphics Computing Lab (LAPIX) home page (available online at <http://www.lapix.ufsc.br/Clouds/CloudsGeometricDistance/CloudsGeometricDistance.html>).

F11 The numeric data obtained from the EGD analysis of 49 images were grouped and reduced to be compared with the Long et al. (2006) method. A summarized proportion or fraction of the sky and cloud pattern along the day comparing Long et al. (2006) and EGD method are shown in Fig. 11.

F12 Figure 12 illustrates a comparative correlation between the two methods for sky and cloud proportions along the day. Pearson's correlation coefficient showed a 0.979 ($r^2 = 0.958441$) correlation for cloud proportion and a 0.984 ($r^2 = 0.968256$) correlation for sky proportion.

Small differences between the two methods can be explained. The main difference is that Long et al. (2006) established a classification method based on a "reduced dimension empirical discrimination proportion value." The

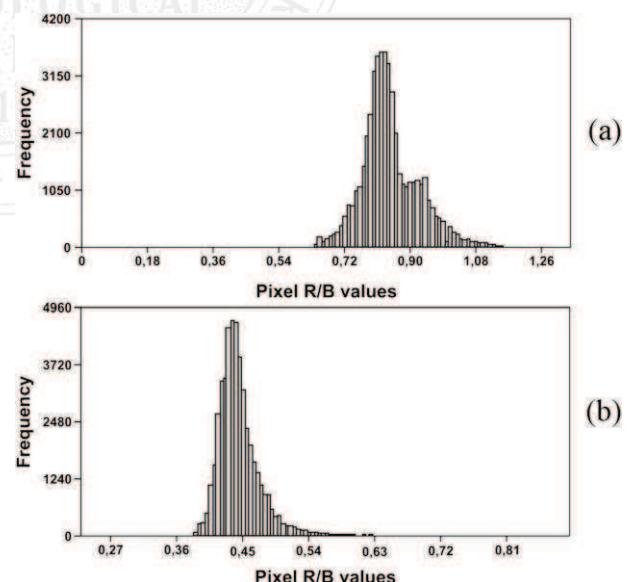


FIG. 10. Histograms of patterns according to the Long et al. (2006) method implemented on the GUI for (a) clouds and (b) clear skies. Horizontal scale is nondimensional.

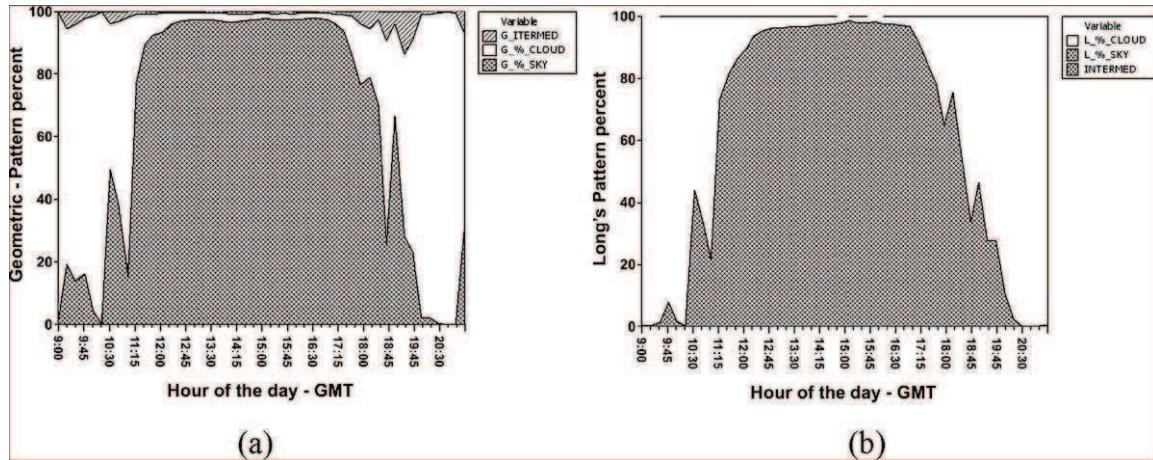


FIG. 11. Comparison of proportions or fraction of sky pattern along the day obtained from (a) geometric distance and (b) Long et al. (2006). Images were taken from 0900 to 2100 at 15-min intervals.

present method is based on “statistics-based confidence intervals.” The geometric distance locus considers the tail superposition of statistical characterization due to the already known type I and II errors, not mentioned or considered in the Long et al. (2006) paper. Long et al. (2006) used a hard discrimination criterion, making use of intermediate values classified as either clouds or sky. But for clear-sky and cloud conditions both methods agree. The squared correlation coefficient indicates that the 95.84% variability in clouds and the 96.83% in sky detection in the Long et al. (2006) method are associated with cloud variability in the EGD method. Differences between the two methods are greater, especially when transitory cases are

present, because of the occurrence of mixed blue-sky and cloudy conditions in the image. Figure 12 also shows that differences between the two methods tend to increase as the cloud proportion increases.

Looking at the graphics of Fig. 12, a small offset can be observed. It can be inferred that the Long et al. (2006) method assigns more pixels to the cloud proportion than the EGD method. This is confirmed by inspecting the accumulated analysis of both methods throughout the day. Long et al.’s (2006) method indicates a higher amount of cloud proportion and a smaller sky proportion than the EGD method. This explains the observed bias between the two methods. In fact, the determination of

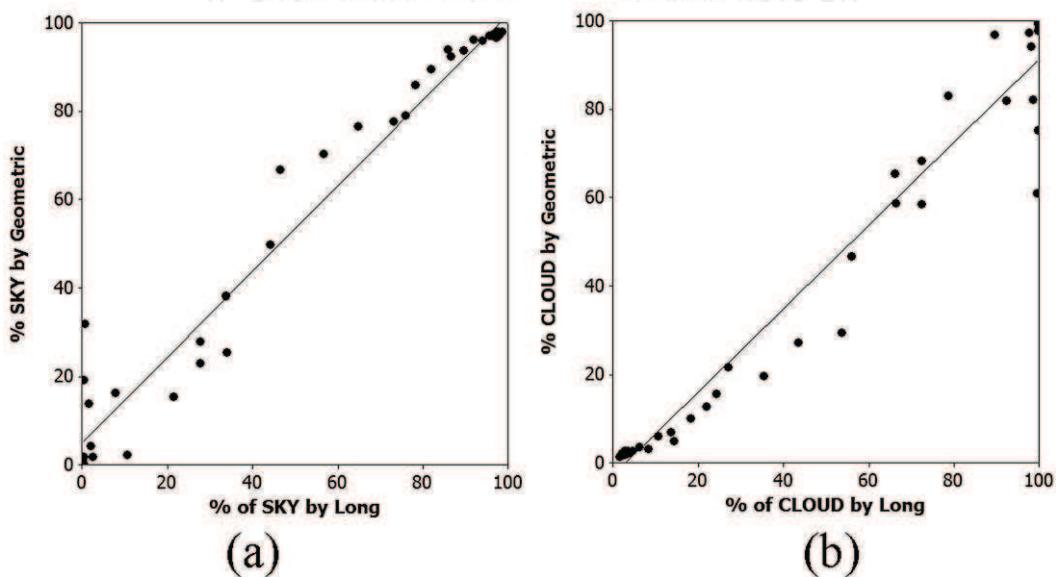


FIG. 12. The comparative correlations between Long et al. (2006) and the EGD methods for (a) sky pattern on and (b) clouds pattern on.

cloud proportions is sensitive to the established discrimination values; as a result, distinct criteria can lead to different results. The differences could be reduced if a characteristic pixel value could be established as the boundary in transition between clouds to sky. Image-preprocessing filtering on smooth transitions like the ones found in the current domain could help to determine that boundary. This is a subject that is recommended to be investigated in future.

5. Future work

Future work based on the Bayesian methodology could explore the use of geometric loci to model other features in the images, especially the ones identifiable by observers but not by automatic systems, which would allow the improvement of qualitative analyses performed by automatic systems. Because of the Rayleigh gradient observed in the sky and its quantification with the brightness EGP scale, we intend to investigate a correlation between the pattern variations to the sun according to photometer measurements. EGD- and EGP-derived values could be correlated to surface radiometers to support the evaluation of a radiative surface flux. In that case, a comparison between EGD's clearness, cloud forcing, and other established time series-based indexes existent in the literature should be investigated. The current methodology is being tested long term at the BSRN São Martinho da Serra station in Southern Brazil.

6. Conclusions

The purpose of the work described in this paper was to develop a methodology to improve the automatic classification of clouds and sky patterns from surface images. EGD allowed not only a classification of those patterns comparable to existing methods but also provided new features in the images based on their color attributes, like Mie scattering and dust removal from the reflector. Those new features could be observed with the 2D (used by Long et al. 2006) and the 1D (used by Souza-Echer et al. 2006) color space-based approaches, because they were masked out by dimensional reduction and mathematical simplifications. The 3D approaches expand the domain analysis of color space to its limit, allowing new potential features to be investigated.

The use of EGD allowed for the classification of clouds and Rayleigh-scattering patterns invariant to their brightness, reducing problems due to solar disk presence, solar variations of the zenith angle, and in the amount of images necessary to model typical occurrences of patterns. Statistical methodologies applied to the image analysis supported those assumptions by

means of the Gaussian distribution of patterns and the central limit theorem.

The Bayesian model using supervised learning and analysis applied to EGD patterns proved to be highly correlated to the Long et al. (2006) approach, even when simplified by dimension reduction. This high correlation proved that the Bayesian model used in the present research is a useful tool and could be employed to identify other patterns based on color attributes in future research works. These patterns representing physical phenomena in different color attributes are the same ones perceived by human vision. The mathematical modeling used in the present work matches the theory of the probabilistic model of cognition, as described by Tenenbaum et al. (2006) and Chater et al. (2006), and uses statistical learning and statistical inferences to classify patterns.

Acknowledgments. This work was made possible thanks to the SONDA project sponsored by FINEP (22.01.0569.00) and by PETROBRAS (0050.0029348.07.4). We acknowledge graduate students Leandro Coser, Antonio Carlos Sobieransky, and Adiel Mittmann from LAPIX-UFSC for their support and fruitful discussions on the current subject, and also Professor Sergio Colle from LABSOLAR-EMC-UFSC for providing facilities for the experimental setup.

REFERENCES

- Chater, N., J. B. Tenenbaum, and A. Yuille, 2006: Probabilistic models of cognition: Conceptual foundations. *Trends Cognit. Sci.*, **10**, 287–291.
- Fernandez-Garcia, N. L., A. Carmona-Poyato, R. Medina-Carnicer, and F. J. Madrid-Cuevas, 2008: Automatic generation of consensus ground truth for the comparison of edge detection techniques. *Image Vis. Comput.*, **26**, 496–511.
- Gonzales, R. C., and R. E. Woods, 2002: *Digital Image Processing*. Prentice Hall, 793 pp.
- Holle, L. R., and S. A. Mackay, 1975: Tropical cloudiness from all-sky cameras on Barbados and adjacent Atlantic Ocean area. *J. Appl. Meteor.*, **14**, 1437–1450.
- Hoyt, D. V., 1978: Interannual cloud-cover variations in the contiguous United States. *J. Appl. Meteor.*, **17**, 354–357.
- International Organization for Standardization, 1995: Guide to the expression of uncertainty in measurement, ISO, 112 pp. AU4
- Iqbal, M., 1983: *Introduction to Solar Radiation*. Academic Press, 390 pp.
- Jiang, X., C. Marti, C. Irniger, and H. Bunke, 2006: Distance measures for image segmentation evaluation. *EURASIP J. Appl. Signal Process.*, **2006**, 209.
- Johnson, R. A., and D. W. Wichern, 2007: *Applied Multivariate Statistical Analysis*. 4 ed. Prentice Hall, 773 pp.
- Lenoble, J., 1993: *Atmospheric Radiative Transfer*. A. Deepak Publishing, 532 pp.
- Lillesand, T. M., and R. W. Kiefer, 1994: *Remote Sensing and Image Interpretation*. John Wiley & Sons, 750 pp.

- Long, C. N., D. W. Slater, and T. Tooman, 2001: Total sky imager model 880 status and testing results, ARM Rep. DOE/SC-ARM/TR-006, 36 pp. [Available online at http://www.arm.gov/publications/tech_reports/arm-tr-006.pdf.]
- , J. M. Sabburg, J. Calbó, and D. Pagès, 2006: Retrieving cloud characteristics from ground-based daytime color all-sky images. *J. Atmos. Oceanic Technol.*, **23**, 633–652.
- Mantelli Neto, S. L., 2001: Development of a methodology for cloud coverage estimation using surface cameras compared to satellite images. M.Sc. dissertation, Computer Science Department, Federal University of Santa Catarina, 125 pp.
- , A. V. Wangenhein, and E. B. Pereira, 2005: Preliminary model for clouds estimation on RGB color space using automatic imager. *Proc. XII Brazilian Remote Sensing Symp.*, Goiania, Brazil, National Institute of Space Research, 4123–4131. [Available online at <http://marte.dpi.inpe.br/col/ltid.inpe.br/sbsr/2004/11.16.17.23/doc/4123.pdf>.]
- Montgomery, D. C., 2005: *Design and Analysis of Experiments*. John Wiley and Sons, 643 pp.
- National Institute of Standards, cited 2010: E-handbook of statistical methods. [Available online at <http://www.itl.nist.gov/div898/handbook/>.]
- Rossow, W. B., 1982: Clouds. *Atlas of Satellite Observations Related to Global change*, R. J. Gurney et al., Eds., International Satellite Cloud Climatology Project, 141–162.
- Russell, S., and P. Norvig, 2003: *Artificial Intelligence: A Modern Approach*. 2nd ed. Elsevier, 1021 pp.
- Souza-Echer, M. P., E. B. Pereira, L. S. Bins, and M. A. R. Andrade, 2006: A simple method for the assessment of the cloud cover state in high-latitude regions by a ground-based digital camera. *J. Atmos. Oceanic Technol.*, **23**, 437–447.
- Tenenbaum, J. B., T. L. Griffiths, and C. Kemp, 2006: Theory-based Bayesian models of inductive learning and reasoning. *Trends Cognit. Sci.*, **10**, 309–318.
- World Climate Research Program, 2007: Section of the baseline surface radiation network. WCRP Informal Rep. 9, 54 pp.
- World Meteorological Organization, 2008: Guide to meteorological instruments and methods of observations. 7 ed. WMO No. 8 I.15-1 - I.15-11, 681 pp.

Apêndice A4

Computer Vision and Applications (submitted).

Contribuição da publicação em relação à pesquisa desenvolvida. Comparação da metodologia desenvolvida com metodologias previamente estabelecidas e a comparação preliminar dos desempenhos.

Multivariate Bayesian cognitive modeling for unsupervised quality control of baked pizzas

Sylvio Luiz Mantelli Neto · Daniel Besen de Aguiar · Bianca Sens dos Santos · Aldo von Wangenheim

Received: date / Accepted: date

Abstract The present article describes a Bayesian multivariate methodology developed for unsupervised quality control of pizzas based on RGB color attributes. A sensory experiment was done to define the readiness point ground truth. During the validation phase, different pizza samples were baked at a different temperature. The cheese and crust color patterns were statistically compared against the ground truth to check the readiness point. Results show that the proposed methodology presents a good performance demonstrating that color attributes can be used as an unsupervised quality control using traditional statistical methods.

Keywords Bayesian Modeling · Multivariate Statistics · Geometric Color Space Locus

1 Introduction

Large scale production in industry is tied to a massive control that aims to guarantee product quality. Many food production lines rely on supervised (i.e., human) quality control [9]. In that case, if the production rate

increases, so does the demand for supervised analysis. In order to reduce the operational overload, quality control in large-scale production is usually carried out by sampling. Supervised quality control that depends on human activities has several limitations. Among them could be mentioned: operator training, subjective analysis, lack of consistency among operators and repetitive observation stress. Training demands time to be spent on several operators and also working shifts to perform their activities. Subjective analyses are always based on the operator's personal criteria of quality. Consistency is also relevant, since different criteria are used by different operators. Repetitive stress is caused by long shifts of annoying visual tasks. To avoid those shortcomings, new solutions are being seek for automatic food quality control that soundly support mass production on industry.

Quality control applied to food production has specific features to be analyzed [4]. Most quality control tasks are performed by visual supervision, leading to the same limitations mentioned above. In baked food production, for example, visual or human supervision could also cause food contamination due to manual inspection or sample destruction due to homogenization for color analysis [1]. For these reasons, several applications are making use of digital resources like cameras, computers and advanced automatic systems [6]. But replacing human supervision by an automated system with a certain level of intelligence is an activity that involves aspects of human cognition, artificial intelligence (AI), image processing and computer analysis. All of these fields are related to Knowledge Engineering, which deals with the transference of certain domains of human expertise to machines (or complex systems). The following research works in the related literature were considered in the pursuit of this objective.

Sylvio Luiz Mantelli Neto
INPE-CCST and UFSC-EGC/INCOD/LEPTEN
Campus Trindade Florianopolis SC Brazil 88040-900
Tel.: + 55 (48) 3721-9516 ext. 17
E-mail: sylvio@lepten.ufsc.br

Daniel Besen de Aguiar
UFSC-INCOD/LEPTEN
Campus Trindade Florianopolis Sc Brazil 88040-900
Bianca Sens dos Santos
UFSC-EMC-LEPTEN
Campus Trindade Florianopolis Sc Brazil 88040-900
Aldo von Wangenheim
UFSC-EGC/INCOD
Campus Trindade Florianopolis Sc Brazil 88040-900

[25] used an image-based food quality control system, but their main objective was qualitative and still relied on supervised analysis of the CMYK, RGB and L* a* b* color spaces. They limited their work to a PHOTOSHOP color appearance analysis; there was no reference to its practical use or implementation in automated quality control.

[21] and [22] used fuzzy logic-based automation to check the distribution of tomato sauce and pizza shape. They then extended their analysis to the distribution cheese, ham, olives and oregano. The work made use of AI resources in the assembly analysis of raw pizzas.

[5] compared three classification methods for the analysis of pizza toppings using the RGB, HSV, I1I2I3, CIE L*a*b* and YCbCr color spaces. Even with the use of neural networks and computer algorithms, their data analysis was simplified by reducing the dimension of color spaces. But dimension reduction leads also to a reduction of classification performance by an increase of classification errors and information loss [10]. [24] produced an assessment of baking conditions for cooking cheese. In the research of [24], color images were used during experiments but the dimension of data was again reduced to a grey scale level in order to establish the browning factor. Variable dimension reduction simplifies mathematical manipulation but also limits the feature domain that can be explored in the experiment.

The work of [20] uses CCD camera images and measurements of other variables to monitor macroscopic changes and structural modifications that occur during baking to refine the bread baking process. The camera system was used only to monitor bread expansion during baking, by thresholding the images obtained. No analysis of the readiness point in the baking process was made.

[17] studied the kinetics of bread browning during baking. In their research, images were analyzed in the L*, a*, b* color space. They used the Euclidean distance (referred to as total color change) between the averages of two color bread cooking point patterns, and developed a mathematical model correlating the Euclidean distance to weight loss.

[7] in their review outline the importance of computer vision as a tool for food quality control. But proper digital inspection of the food quality process requires more resources than those provided by computer vision only. Computer vision is one element in a chain that starts with modeling the domain and variability of the operator's senses (visual cognitive system) and the matching of the kind of sensor to the domain of data to be used. In order to extract the desired information from obtained data, appropriate computer vision algorithms, such as segmentation, must be

used. These algorithms are typically complex and time consuming, posing a special challenge for mass production environments. Finally, appropriate tools (i.e., artificial intelligence, decision tools, Bayesian methods, neural networks, genetic algorithms, fuzzy logic, etc.) must be chosen to interpret and classify the information into the desired knowledge field to perform the desired task. All these tasks are performed by a system, usually called intelligent agent [18], that interprets (or maps) the contents of images to the desired knowledge application field. All these activities that deals with knowledge modeling and insertion into machines constitute a process called digital convergence.

The goal of this research is to use the Bayesian approach [23], [3], [8] and [16] in order to propose a multivariate methodology that can be used as an automatic quality control system for pizzas, using CCD cameras to obtain its data. The following sections will describe how the data obtained by the camera and all related instrumentation during the cooking process were analyzed by a group of users in order to establish a preferred readiness point and to define a ground truth pattern Jiang et al.[11]. Then image processing techniques were used to carry on a mathematical and color space analysis in order to calibrate the model of readiness representation. Validation and analysis of the proposed model was performed on separate samples, baked at different temperatures. Results show that the proposed statistical model makes up a good methodology to be used as an automatic quality control tool of the readiness point, based on color attributes.

2 Material and Methods

The present method was used to evaluate the performance of a heat pipe oven and the readiness point of pizza in a M.Sc. dissertation at LABTUCAL¹ by [19]. It was also associated with an undergraduate project² carried out by [2] and supervised by this article's corresponding author. The methodology proposed by the authors was Bayesian, which is one of the main trends among the research groups involved. Neural Networks (NN) were also considered as an indicator for the readiness point. But NNs merely mimic the patterns they were presented in the training stage and do not provide the necessary mathematical foundation to calibrate the model. Like NNs, the Bayesian approach is also a supervised calibration method, but the latter provides a

¹ LABTUCAL: Heat Pipe Laboratory, Mechanical Engineering Department of the Federal University of Santa Catarina, www.labtucal.ufsc.br.

² Department of Computer and Statistics of the Federal University of Santa Catarina, www.ine.ufsc.br

more rigorous statistical and mathematical modeling, very close to the practice of quality control used in industry. The present work proposes a definition of the readiness point based on a sensory analysis. It starts by acquiring images with CCD cameras and by applying a preliminary image analysis in the RGB color space. Other color spaces, like HLS, etc. were not considered due to the non-linear distribution in the target patterns in those spaces. The desired patterns present in images were characterized by multivariate statistics. Modeling was guided and calibrated by sensory analysis made by users, who determined the readiness point. Then new images were acquired and analyzed statistically to check if the readiness point was achieved during the cooking process validation. A general diagram of the procedure used can be observed in Figure 1.

2.1 Experimental setup

The pizza images analyzed in this article were obtained with a Cannon EOS Digital Rebel camera. The lighting conditions were controlled. Two lamps of 6500 K were used as the lighting source to minimize variation along the experiment. These lamps, measuring 45 cm each, were installed 40 cm away from the pizza base and 50 cm from each other. The distance between the camera lens and the pizza center was 30 cm. The pizza images were obtained under the following conditions: RGB color space, focal distance of 28 mm, diaphragm aperture of f/5.6, and exposition of 1/200 seconds. The image file format was JPEG with a resolution of 2048 × 1360 for the sensory analysis and 640 × 800 during the validation phase. Images were obtained taking the pizza out of the oven and returning it to the cooking chamber immediately afterwards.

Three different pizzas were baked in an electric oven of size 242 × 375 × 298 mm at 200 ± 25 °C for the sensory analysis phase and three others were baked at 250 ± 25 °C for the validation phase. Temperatures were measured with a K-type thermocouple in an HP 34970 data logger controlled by the software LABVIEW. A typical temperature distribution observed during the experiment, showing the influence of image acquisition, can be seen in Figure 2.

2.2 Sensory analysis

Sensory analysis was performed to determine the satisfactory baking point. The goal of this analysis was to establish the color pattern that should be used as the ground truth to define the pizza readiness point. The main attribute used to represent the readiness point of

pizzas was color, defined by comparing images of different pizzas. For the sensory analysis, the cooking temperature of pizzas was reduced to 200 °C to allow the preparation of at least ten different samples from the same baking lot for the analysis.

Mozzarella pizzas were baked in an electric oven, registered by the CCD camera and code marked at fixed times: 0 (raw), 4, 8, 12, 16, 20, 24, 28, 32 and 36 minutes. Acceptance categories for the color of pizzas were determined according to 5 categories (1 through 5) of a Hedonic scale based on the following classification: disliked a lot, disliked, regular, liked and liked a lot.

Sixty-eight untrained users from the Mechanical Engineering Department (students, professors and staff) of the Federal University of Santa Catarina, Florianopolis, Brazil, with ages varying from 20 to 60 years, volunteered for the analysis. Pizza images were sent by e-mail (randomly mixed) and a template form was used to score every pizza. Users classified every sample according to the scale provided and thus the appearance of pizzas was analyzed. Pictures showing the aspect of the pizzas are illustrated in Figure 3.

Statistical tests of acceptance were done by analysis of variance (ANOVA) and the averages were compared with the Tukey test with 5% of significance. Data were evaluated using histograms. The Hedonic classification used in the acceptance index is given by

$$AI = \frac{100N_{AI}}{5}, \quad (1)$$

where AI is the acceptance index, N_{AI} is the average score taken from the acceptance test for a specific sample. For a sample to be accepted, it should score a minimum of 70%. The results of tests are provided in Table 1.

Some of the samples in the sensory analysis presented statistically similar acceptance levels. In such cases, the samples were discarded from the set of patterns of the cognitive ground truth. As a conclusion of the sensory test, the range from 8 to 16 minutes (200 ± 25 °C) of baking time was determined to be the most appropriate according to the acceptance index. The corresponding pictures were used as the ground truth for modeling the readiness point.

2.3 Analysis of color evolution during baking for calibration of the statistical model

The chosen pizza flavor has several image elements (or patterns) to be monitored during the baking process: background, crust, cheese, olive and sauce. A preliminary analysis of these elements was done during the baking process in the RGB color space using the tool

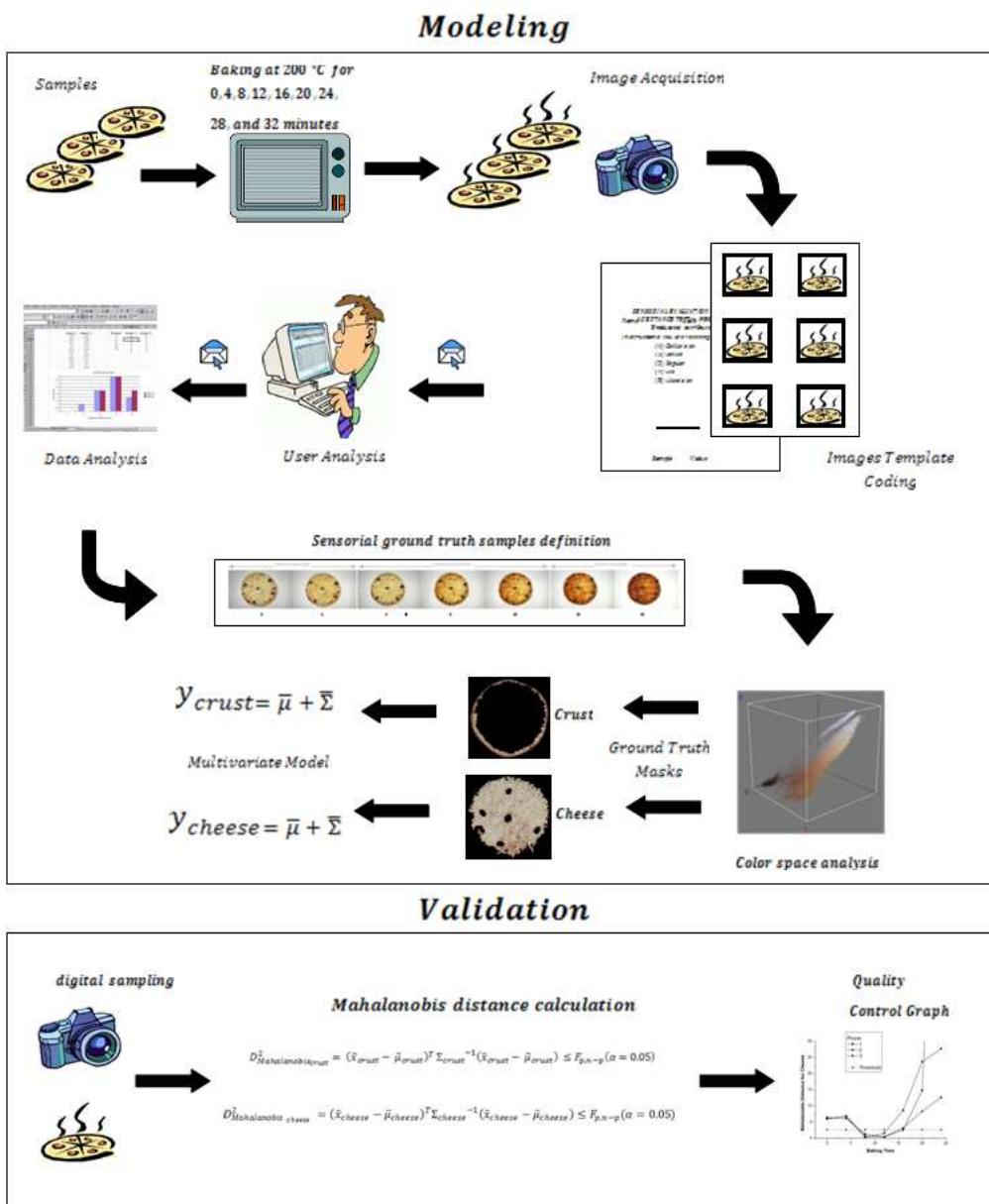


Fig. 1 General overview of the proposed Bayesian methodology for quality control. Both modeling and validation phases are shown. See Equations 2 through 6 for description of formula terms.

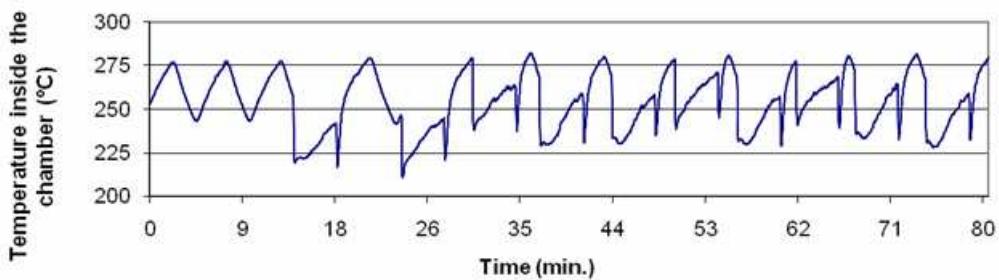


Fig. 2 Typical distribution of oven temperature during cooking experiments.

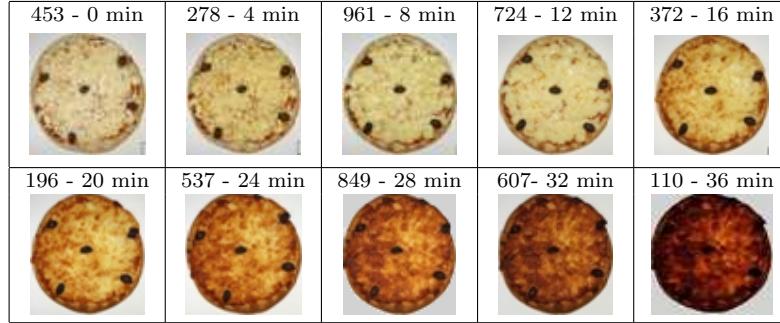


Fig. 3 Pictures of a pizza set submitted to sensory analysis. Cooking time was omitted during sensory tests.

Table 1 Results obtained from the evaluation of pizzas by 68 users, based on the acceptance for Tukey tests with a 5% level of significance

Sample Order	Pizza Code	Cooking Time	Average	Std. Dev.	Acceptance Index	Tukey Classification
6	P453	0	1.7059	0.8987	34.12	Not significant
1	P278	4	2.2941	0.8297	48.88	
3	P961	8	3.7059	0.9934	74.12	
4	P724	12	3.5882	0.8506	71.76	
2	P372	16	4.2059	0.8735	84.12	
5	P196	20	3.5294	1.1778	70.59	
4	P537	24	2.4706	0.9219	49.41	
8	P849	28	1.5294	0.6572	30.59	
9	P607	32	1.1471	0.4324	22.94	
10	P110	36	1,0000	0.0000	20.00	

ColorInspector3D³. This analysis indicated the following observations, as can be seen in Figure 4. Background was discarded, because it was not the object of observation. The black color of the olive element does not show significant change during the baking process. Sauce does not change either, since it was mostly under the cheese and covered by its melting. Crust and cheese patterns changed considerably during the baking process. A shift of those patterns was noticed during baking from raw to overcooked due to browning. Crust borders browned and burned earlier than cheese. According to what was observed in the analysis, there is a typical cooking point that can be related to pixel values in the RGB color space and a typical *locus* that corresponds to a preferred baking point for both cheese and crust.

In order to improve the modeling, all the discarded patterns were removed from the data set by manipulation to minimize their influence on the analysis [15]. The removal was done manually, pixel by pixel, using a Ground Truth Generation tool developed at LAPIX⁴. The variation of pixels in the crust and cheese during the baking process are illustrated in Figure 5. The characterization of that typical *locus* was obtained

³ ColorInspector3D is available at <http://rsb.info.nih.gov/ij/plugins/color-inspector.html>

⁴ LAPIX: Laboratory for Image Processing and Computer Graphics. Department of Informatics and Statistics, Federal University of Santa Catarina. www.lapix.ufsc.br

by a multivariate exploratory data analysis (EDA) [16] performed for every color dimension by a statistical software. This characterization can be expressed as

$$y = \bar{\mu} + \bar{\Sigma}, \quad (2)$$

where y is the specific image element for the preferred baking point, $\bar{\mu}$ is the pixel multivariate average mean vector and $\bar{\Sigma}$ is the covariance matrix representing the random variable in the dimensions of the RGB color space associated with the observation [15].

As indicated by the sensory analysis, the preferred interval of baking time lies between 8 and 16 minutes at 200 ± 25 °C. Three pizza images belonged to that interval. Pixel values, separated in crust and cheese patterns, were extracted from these three images and then grouped and expressed in one single pattern representing the global readiness point, according to Table 2.3 and Equations 3 and 4.

$$y_{\text{cheese}} = \begin{vmatrix} \bar{\mu}_1 \\ \bar{\mu}_2 \\ \bar{\mu}_3 \end{vmatrix} + \begin{vmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{vmatrix} \quad (3)$$

$$y_{\text{crust}} = \begin{vmatrix} \bar{\mu}_1 \\ \bar{\mu}_2 \\ \bar{\mu}_3 \end{vmatrix} + \begin{vmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{vmatrix} \quad (4)$$

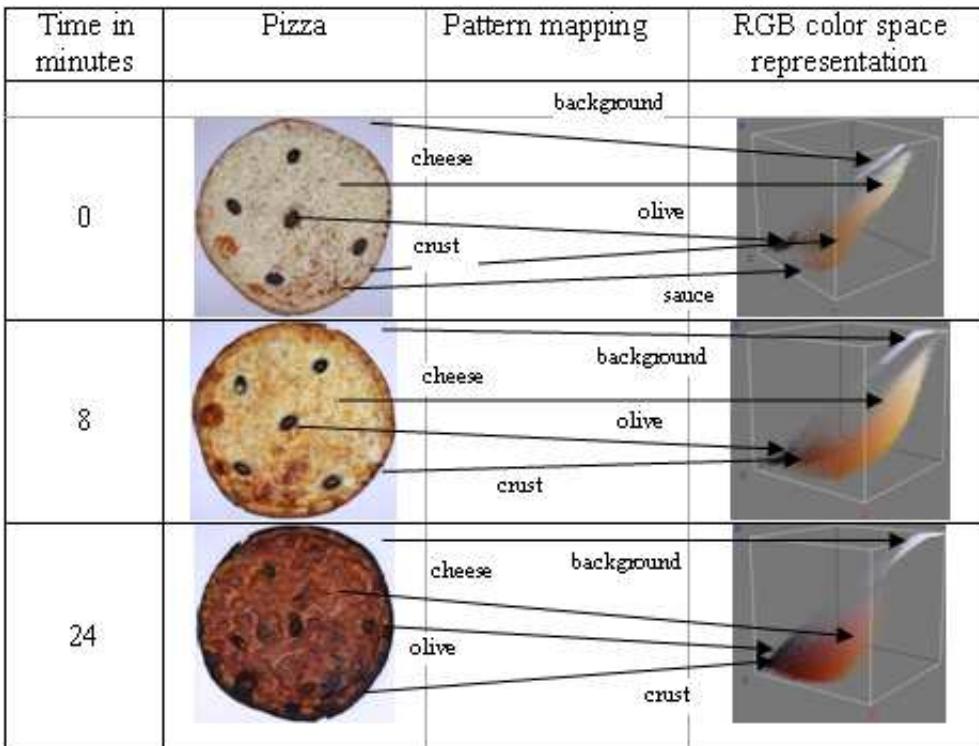


Fig. 4 The mapping of background, cheese, olive, crust and sauce elements into patterns in the RGB color space, and their evolution during the baking process at selected times.

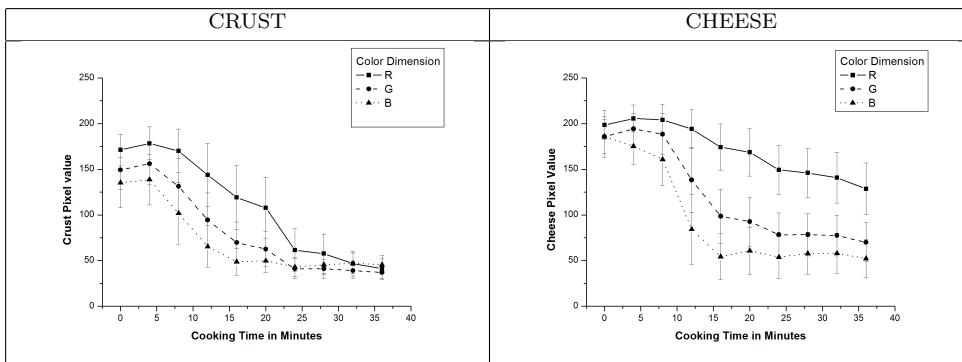


Fig. 5 Evolution of the R, G and B average values for the three pizzas during the baking process along time, in minutes, for crust and cheese patterns.

These two patterns extracted from the images showed a multivariate normal distribution or near normal in their dimensions. The present analysis assumed data to be normal, because the sample sizes are very large [12].

3 Multivariate validation of the readiness point

A second, independent experiment with 3 pizzas, baked at $T = 250$ °C was executed to independently analyze the baking process. The image resolution was changed from 2048×1360 , used in the sensory analysis, to 640×480 in order to reduce the massive amount of data.

Validation tests for the baking point of pizzas were carried out by checking the Mahalanobis distance⁵ [13] from the target pattern to the pattern defined in the sensory analysis. Hotelling quality control graphs [14] and

[12] were used as graphic illustration of the process. The Mahalanobis distance used in the current analysis for quality control charts can be described by

$$D^2 = (\bar{x} - \bar{\mu})^T * \bar{\Sigma}^{-1} * (\bar{x} - \bar{\mu}) F_{p,n-p,(\alpha)}, \quad (5)$$

⁵ Mahalanobis or statistical distance is used instead of Euclidean distance due to distinct covariances existent among color dimensions for target patterns [12]

Table 2 Exploratory data analysis of pizza element values (crust and cheese) obtained from the sensory analysis performed by consumers. Here, n is the number of pixels evaluated in the analysis.

Pattern	Dimension	Mean μ	Std. Dev.	Covariance Matrix S		
Cheese n=191302	Red	201.40	17.81	317.077	507.519	415.699
	Green	162.31	35.30	507.519	1246.003	1242.256
	Blue	83.682	39.128	415.699	1242.256	1531.020
Crust n=31260	Red	152.85	24.84	617.210	585.236	335.757
	Green	99.73	27.912	585.236	779.060	595.032
	Blue	45.539	25.253	335.757	595.032	637.706

where D^2 is the generic Mahalanobis distance, n is the sample size, $\bar{x} = (x_1, x_2, x_3)$ is the sample average matrix, $\bar{\mu}$ is the pattern defined in the sensory analysis, $\bar{\Sigma}$ is the pattern covariance matrix, p is the variable dimension and α is the level of confidence. M^T is used to denote the transpose of the matrix M and M^{-1} the inverse matrix. $F_{p,n-p}$ is the upper (100α) percentile critical value of the F distribution at p and $n - p$ degrees of freedom. A univariate quality control chart was obtained along with critical values based on the multivariate F distribution. The critical value or threshold used in quality control is obtained from statistical tables according to Equation 6 for the following conditions: degrees of freedom $p = 3$ (Red, Green and Blue dimensions), $n = \infty$ (number of pixels analyzed is very large), and level of confidence $\alpha = 0.05$.

$$F_{p,n-p,(\alpha)} = 2.08 \quad (6)$$

Figure 6 contains quality control charts based on the validation set.

4 Analysis of results

Results obtained in the experiments are related to the comparison of the validation set to the previously established readiness point modeled by the sensory analysis, using multivariate statistics and univariate quality control charts. The temperature used for the validation set was higher than that of the sensory analysis set ($250 \pm 25^\circ\text{C}$). This difference caused the preferred baking point to be reached earlier during the validation phase and later in the training phase. This expected result was confirmed by the chart in Figure 6, where the preferred baking point, according to the Mahalanobis distance, was reached between 7 and 16 minutes in the case of cheese. The beginning of the interval was very similar for all three pizza samples. Variances at the end of the interval might have been caused by uncontrolled and unpredicted factors related to cheese melting. It is speculated that the non-uniform cheese distribution on the

surface of the pizzas may cause the cheese portions located nearer the metal parts to burn faster. These large variances could probably be significantly reduced in a more controlled environment. These improvements are left as a future work in a new, more refined experiment.

A little overshoot was noticed during the baking of cheese, but not crust, as can be seen in the graphs of Figures 5 and 6. This overshoot was due to the fat present on the cheese surface, mainly in the beginning of the baking process. In the case of the crust, the overshoot was negligible, since it has much less fat than the cheese. Pictures taken during this phase tend to reflect more light, which is registered by the camera. This fact could be better monitored by increasing the temporal resolution of pictures during the experiment. The observation of the general appearance of the charts leads to the speculation that the typical equation describing the baking process could be empirically obtained from the experiment. That confirmation, however, is also left as future work for new experiments.

Multivariate analysis of crust and cheese patterns showed a normal or near normal distribution in their dimensions. The large amount of data samples used in the experiment guarantees that the current approach can be used as a sufficient summary and as a criteria for analysis of normal data samples. Attention should be paid, however, to the partial offset caused by loss of sample information introduced by summary statistics of averages and by covariance matrices [12].

One important point to be outlined is that the analysis of the readiness point is based on pixel values occurring in a typical geometric *locus* in the RGB color space. Therefore, one pizza provides thousands of pixel samples representing typical occurrences of the pattern for the readiness point of crust and cheese.

Some adjustments seem to be required in the case of crust patterns, since there was an offset in the quality control graph of Figure 6. It is possible to adjust the threshold according to the quality control system and the illumination conditions. Validation has shown that statistical quality control based on the Mahalanobis distance (or any other technique like hypothesis testing or

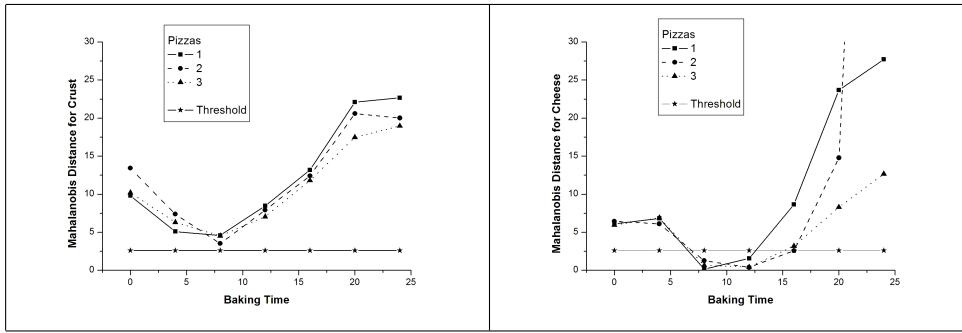


Fig. 6 Quality control charts based on the Mahalanobis distance along the baking process. The distance is based on multivariate quality readiness point of pizza baking for crust and cheese after the validation of the sensory analysis

multiple confidence intervals and so on) using a specific population has a great potential to be applied in large scale color-based quality control systems. Data showed a clear evolution of baking point patterns towards the optimal, as established by the ground truth, and then away from it as overbaking takes place. Differences between the readiness point of crust and cheese during baking is also clearly noticeable in the charts. It is assumed that this makes no difference, since crust browning is accepted more commonly as indicating the readiness point than cheese browning. In the charts of Figure 6, the crust did not reach the critical (or threshold) value for readiness, but the overall crust area is very small when compared to the cheese surface area, and the readiness point is more clearly indicated by the cheese. The crust area corresponds to only $13.50 \pm 2.45\%$ of the pizza in our experiment. If properly configured, the methodology presented by this article may work on a sample area of the pizza, thus increasing the monitoring speed of the quality control system.

It can also be observed that the patterns of both crust and cheese tend to black (or zero) in the graph of Figure 5.

5 Conclusions

A new methodology has been developed for the unsupervised determination of patterns for the readiness point of pizzas based on color attributes and using Bayesian cognitive modeling. The same methodology may be applied not only to baked pizza, but also in other cases where quality control that relies on color attributes. The method presented here uses a temporal univariate graph to monitor multivariate color variables. The threshold used for the readiness point may be optimally adjusted to accommodate implementation or system details. Other statistical tools, such as multiple simultaneous confidence intervals, Bonferroni multiple comparisons, etc. may also be applied to this method in

quality control systems. A reduced set of pixels taken from pizza samples may also be used to reduce the amount of data, but this was considered a refinement of the method presented in this article and was left as future work.

Acknowledgements The authors would like to thank the technicians involved at LABTUCAL activities, who helped in the experimental setup.

References

1. Abdullah MZ (2008) Quality evaluation of bakery products. In: Sun DW (ed) Computer Vision Technology for Food Quality Evaluation, Academic Press, Amsterdam, pp 481 – 522, DOI DOI: 10.1016/B978-012373642-0.50023-5
2. de Aguiar DB (2008) Metodologia bayesiana para o controle de qualidade de pizzas (bayesian methodology for quality control of pizzas - undergraduate project - ufsc - ine - informatic and statistics department). Tech. rep., Federal University of Santa Catarina, Brazil
3. Chater N, Tenenbaum JB, Yuille A (2006) Probabilistic models of cognition: Conceptual foundations. Trends in Cognitive Sciences 10(7):287 – 291, DOI DOI: 10.1016/j.tics.2006.05.007, special issue: Probabilistic models of cognition
4. Du CJ, Sun DW (2004) Recent developments in the applications of image processing techniques for food quality evaluation. Trends in Food Science & Technology 15(5):230 – 249, DOI DOI: 10.1016/j.tifs.2003.10.006
5. Du CJ, Sun DW (2005) Comparison of three methods for classification of pizza topping using different colour space transformations. Journal of Food Engineering 68(3):277 – 287, DOI DOI: 10.1016/j.jfoodeng.2004.05.044

-
6. Du CJ, Sun DW (2005) Pizza sauce spread classification using colour vision and support vector machines. *Journal of Food Engineering* 66(2):137 – 145, DOI DOI: 10.1016/j.jfoodeng.2004.03.011
7. Du CJ, Sun DW (2006) Learning techniques used in computer vision for food quality evaluation: a review. *Journal of Food Engineering* 72(1):39 – 55, DOI DOI: 10.1016/j.jfoodeng.2004.11.017
8. Duda RO, Hart PE, Stork DG (2001) Patterns Classification., 2nd edn. John Wiley and Sons
9. Gunasekaran S (1996) Computer vision technology for food quality assurance. *Trends in Food Science & Technology* 7(8):245 – 256, DOI DOI: 10.1016/0924-2244(96)10028-5
10. Jain A, Duin R, Mao J (2000) Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22:4
11. Jiang X, Irniger C, Bunke H (2006) Distance measures for image segmentation evaluation. *EURASIP Journal on Applied Signal Processing* 2006:1 – 10
12. Johnson RA, Wichern DW (1998) Multivariate Statistical Analysis. Prentice Hall
13. Mahalanobis PC (1936) On the generalized distance in statistics. National Institute of Science pp 49 – 55
14. Montgomery DC (2001) Introduction to Statistical Quality Control. John Wiley & Sons Inc.
15. Montgomery DC (2005) Design and Analysis of Experiments. John Wiley & Sons Inc.
16. NIST (2006) Engineering statistics handbook. Tech. rep., NIST - National Institute Of Standards
17. Purlis E, Salvadori VO (2007) Bread browning kinetics during baking. *Journal of Food Engineering* 80(4):1107 – 1115, DOI DOI: 10.1016/j.jfoodeng.2006.09.007
18. Russel S, Norvig P (2004) Artificial Inteligence. Elsevier
19. Santos BS (2009) Determinação das condições térmicas de cocção e das propriedades termo-físicas da pizza (determination of thermal cooking and thermophysics properties of pizza). Master's thesis, UFSC - Universidade Federal de Santa Catarina
20. Sommier A, Chiron H, Colonna P, Valle GD, Rouill J (2005) An instrumented pilot scale oven for the study of french bread baking. *Journal of Food Engineering* 69(1):97 – 106, DOI DOI: 10.1016/j.jfoodeng.2004.07.015
21. Sun DW, Brosnan T (2003) Pizza quality evaluation using computer vision–part 1: Pizza base and sauce spread. *Journal of Food Engineering* 57(1):81 – 89, DOI DOI: 10.1016/S0260-8774(02)00275-3
22. Sun DW, Brosnan T (2003) Pizza quality evaluation using computer vision–part 2: Pizza topping analysis. *Journal of Food Engineering* 57(1):91 – 95, DOI DOI: 10.1016/S0260-8774(02)00276-5
23. Tenenbaum JB, Griffiths TL, Kemp C (2006) Theory-based bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences* 10(7):309 – 318, DOI DOI: 10.1016/j.tics.2006.05.009, special issue: Probabilistic models of cognition
24. Wang HH, Sun DW (2003) Assessment of cheese browning affected by baking conditions using computer vision. *Journal of Food Engineering* 56(4):339 – 345, DOI DOI: 10.1016/S0260-8774(02)00159-0
25. Yam KL, Papadakis SE (2004) A simple digital imaging method for measuring and analyzing color of food surfaces. *Journal of Food Engineering* 61(1):137 – 142, DOI DOI: 10.1016/S0260-8774(03)00195-X, applications of computer vision in the food industry

Apêndice A5

Eleventh BSRN Scientific Review and Workshop Queenstown, New Zealand 13-16 April 2010

Contribuição da publicação em relação à pesquisa desenvolvida. Poster de divulgação dos resultados parciais das pesquisas em um "workshop" organizado pela WMO Organização Mundial de Meteorologia.

Automatic classification of atmospheric patterns

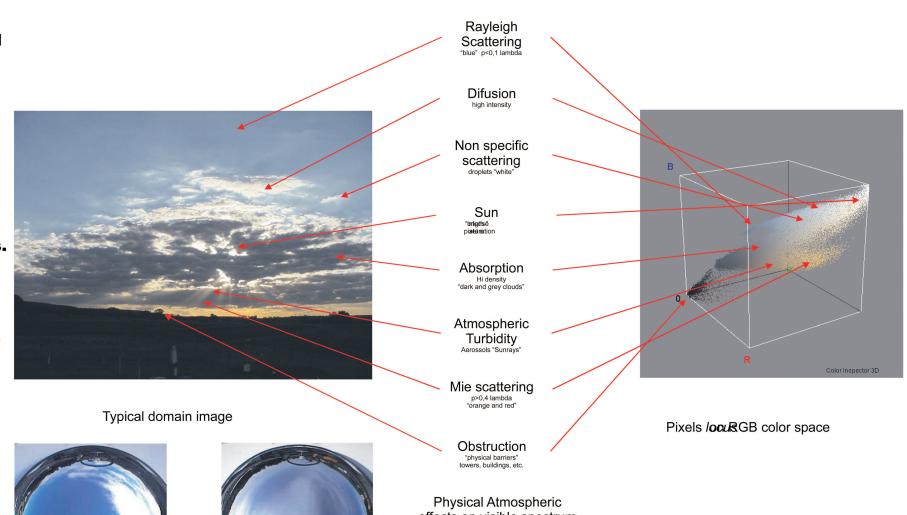
Sylvio Luiz Mantelli Neto, Enio Bueno Pereira, Aldo von Wagenheim
INPE-CCST Brazilian Institute for Space Research-Center for Science and Earth Systems,
UFSC-LAPIX-EGC Federal University of Santa Catarina
Computer Graphics and Image Processing Lab and Knowledge Engineering Department
sylvio@lepten.ufsc.br

Abstract

Surface observers on monitoring stations regularly classify and register sky patterns based on human cognitive skills. Substitute those observers by automatic systems is highly desirable to reduce subjective analysis. Most automatic systems uses trivial classification methods, simplifies mathematical evaluation by domain restrictions, expects only binary results (cloud and sky) and finally ignore other physical phenomena existent in atmosphere. These approaches restricts available resources, reduces the overall performance and causes information loss [1]. This research proposes a method that will establish the correspondence between observations percept [2] from those atmospheric patterns [3] and theory using Bayesian [6] multivariate [4] approach embedded in intelligent agents[5]. Preliminary results of that approach already obtained a classification method invariant to solar elevation. Full domain interpretation allowed also the analysis of other phenomena like Mie scattering, Cirrus and other factors not evaluated with previous paradigm analysis.

Motivation

- Atmospheric and physical models have high variability and reduced performance on mixed cloudy conditions.
- Simultaneous surface and satellite observations are desirable for a better atmospheric evaluation and reduction of uncertainties inherent to both observation methods.
- Surface observers (SO) provide good qualitative, but bad quantitative analysis. In automatic systems these limitations are inverse.
- Methodologies used on image based Surface Observation Systems (SOS) restricts available resources to analyze atmospheric patterns.
- Binary results on analysis also reduces system performance, due to existence of further patterns on the observed domain.
- Advances in automated system methodologies could improve the performance of SOS by proper characterization of physical domain.



Methodology

- Analysis of color attributes on color space of a specific atmospheric phenomena, indicated a typical occurrence (or locus) on Red, Green Blue (RGB) representing that pattern fig. 1.
- Locci relative to multiple patterns could be characterized Geometrically or statistically using image pixel values.
- Pixels Euclidean Geometric Distance (EGD) and Projection (EGP), together with statistical methods were used on 3D color space to analyse cloud and sky patterns.
- Bayesian[2][6], exploratory data analysis (EDA)[6], central limit theorem[6], multivariate statistical methods[4], feature extraction[7] and cognitive analysis[2] were used on current research method.
- Other color spaces where investigated, but patterns did not showed geometric linear distribution. They where used on auxiliary analysis.



Figure 4 sky image showing piranometer overestimation of cloud coverage, point A in the graphic, due to clouds spatial distribution.



Figure 5 sky image showing piranometer underestimation of cloud coverage, point B in the graphic, due to low sensitivity to thin clouds.

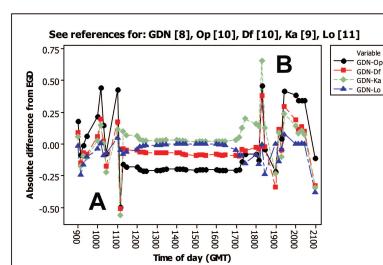


Figure 3 graphic indicating differences between Euclidean Geometric Distance and piranometer based methods for estimation of cloud coverage.

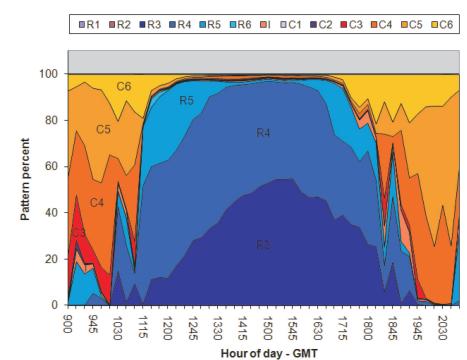


Figure 2 typical evolution of Davy atmospheric patterns obtained from surface cameras, categorized as cloud (C) and Rayleigh scattering(R). Figure adapted from [8]

Results and conclusions

- Patterns EGD showed Gaussian distribution, allowing the use of statistical distance on discrimination.
- Feature analysis based on full color attributes of pixels, allowed a discrimination independent of solar elevation (or brightness).
- Saturation variation on blue color (Rayleigh scattering) due to atmospheric contents (water vapor, aerosols) near surface could be evaluated fig. 2.
- EGD allowed evaluation of more features present on images than reduced domain approaches, exploring full color domain.
- Preliminary comparison to clouds evaluation with piranometer based methods indicated large differences on fig. 3. One due to spatial distribution and temporal resolution of clouds, shown on figs. 3 and 4. Other due to low sensitivity of piranometers to thin clouds, shown on figs. 3 and 5.

Future work

- Multivariate characterization of patterns.
- Extend intercomparison to historical local clear sky models (i.e. cloud forcing, etc.)
- Extend intercomparison to physical models (pirgeometer, atmospheric parameters, etc.).
- Extend intercomparison to sun photometer data.

References

- Jain, A., Duin, R., Mao J. Statistical Pattern Recognition: a review. *IEEE Transactions on Pattern Recognition Analysis and Machine Intelligence* jan. 2000, vol. 22, n. 12.
- Tenenbaum, J. B., Griffiths, T. L., Kemp, C. Theory-based Bayesian models of inductive learning and reasoning. *Trends in cognitive Sciences* 2006, 10(7) 309-318, special issue: Probabilistic models of cognition.
- Lenoble, J. Atmospheric Radiative Transfer. *A Deepak* 1993 532pp.
- Johnson, R. A. and Wichern D. W. *Applied Multivariate Statistical Analysis*, Pearson -Prentice Hall 2007, 773 pp.
- Russell, S. J. and Norvig, P. *Artificial Intelligence, a Modern Approach*. Prentice-Hall 1995, 932 pp.
- NIST/SEMATECH e-Handbook of Statistical Methods, <http://www.itl.nist.gov/div898/handbook/>, mar 2010.
- Montgomery, D. C. *Design and Analysis of Experiments*. John Wiley and Sons 2005, 643 pp.
- Neto SLM, von Wagenheim A, Pereira EB, Comunello E. The use of Euclidean geometric distance on RGB color space for classification of sky and cloud patterns. *Journal of Atmospheric and Oceanic Technology*: In Press EOR jan. 2010.
- Karsten F. and Czeplak G. Solar and Terrestrial Radiation Dependent on the Amount and Type of Cloud. *Solar Energy* 1980 Vol. 24 pp. 177-189.
- Harrison R. G., Chalmers N., Hogan R. J. Retrospective Cloud Determinations from Surface Radiation Measurements 2008 *Atmospheric Research* vol. 90 pp. 54-62.
- Long C., Slater D. W., Tooman T. Total Sky Imager Model 880 Status and Testing Results November 2001, ARM TR-006



INPE - Brazilian Institute for Space Research
www.inpe.br



CST - Center of Terrestrial System Science



LAPIX - Financer of Studies and Projects
www.lipex.ufsc.br

LAPIX Laboratory of Image Processing and Computer Graphics
www.lipex.ufsc.br



UFSC - Federal University of Santa Catarina
www.ufsc.br



INCoD - National Research Institute on Digital Convergence
www.incod.ufsc.br



Knowledge Engineering Department
www.egc.ufsc.br

Apêndice A6

1st Congress of Computational Interdisciplinary Sciences (accepted).

Contribuição da publicação em relação à pesquisa desenvolvida. A aplicação da metodologia Bayesiana, a utilização de métodos estatísticos multivariados para modelagem de padrões cognitivos para serem utilizados em visão computacional.

Poster

Comparison of Bayesian Euclidean Geometric Distance image-based method and short wave radiometer data for cloud coverage evaluation from surface camera images

Sylvio Luiz Mantelli Neto^{1,2}, Leandro Coser^{1,2}, Enio Bueno Pereira¹, Aldo von Wagenheim^{1,2}

¹INPE-CCST Brazilian Institute for Space Research - Center for Science and Earth Systems,

²UFSC-INCoD/EGC Federal University of Santa Catarina

National Research Institute on Digital Convergence and Knowledge Engineering Department

sylvio@lepten.ufsc.br

Abstract.

The purpose of this research is the evaluation of method used for representation of the knowledge existent on surface images taken from an all sky camera to monitor atmospheric conditions, in order to be used as ground truth for the validation of satellite images. Most used ground truth classification of sky conditions are done by synoptic observers (SO) based on highly human-oriented subjective analysis [1]. Replace SO evaluation by automatic synoptic observation systems (SOS) is desirable to reduce the uncertainty of subjective analysis and the working shifts loads. Current image-based SOS's are expensive, have a reduced performance due to simplified classification and dimension reduction methods [2] leading to information loss of available knowledge that could be obtained from images [3]. Present work makes a comparison between a few shortwave radiometer (SW) based methods [4] [5] with image Bayesian Euclidean Geometric Distance Method (EGD) [6], that uses full dimension domain to perform cloud coverage evaluation. Data analysis confirmed that SW radiometers are much more sensitive to clouds in the sunlight pathway than out of pathway [7], making them more sensitive to spatial distribution of clouds, increasing the evaluation differences in sky cover from -56% to 65% in certain cases. As a conclusion, current research indicated that EGD, is much less susceptible to cloud spatial distribution than SW radiometers. Its is inferred from current research that image based methods, have a better performance on cloud amount assessment than SW and is more suitable to be used for cloud cover ground truth validation.

Motivation

- ◆ Satellite-based models have higher variability and reduced performance on mixed cloudy conditions.
- ◆ Satellite cloud assessments in visible channel is validated by surface observations of SO, SW or SOS in order to reduce the uncertainties of cloud evaluation methods.
- ◆ Clouds have to be properly evaluated, in order to be considered on these models.
- ◆ SO provide good qualitative, but bad quantitative analysis of clouds. In automatic systems these limitations are inverse.
- ◆ Advances in automated system methodologies could improve the performance of SOS by proper characterization of physical domain.
- ◆ Current work makes a quantitative comparison between EGD image based methods and SW based methods to compare their evaluation differences.

Methodology

- ◆ [4] parametrized clouds from surface based on 10 years SW observation data, proposing an expression described on following equation.

Where:

C_{kc}: Cloud cover in percent

D: Difuse Radiation

$$C_{kc} = \sqrt{\left(\frac{D}{G} - 0.3\right) \frac{1}{0.7}}$$

- ◆ [5] also parametrized clouds from surface based on 10 years SW observations using two indexes, Opacity(Op) and Difuse Fraction (Df), described on the following equations set.

$$\begin{aligned} D_f &= \frac{D}{G} \\ O_{p=1-K_t} &= \\ K_t &= \frac{G}{E} \\ E &= \frac{S_o \cos \theta}{d^2} \end{aligned}$$

Where

Df: Difuse Fraction

D: Difuse Radiation

G: Global Radiation

Op: Opacity Index

Kt: Clearness Index

E: Top of Atmosphere Radiation

So: Average Solar Constant

d: Sun-Earth distance

θ: Zenith Angle

Results and conclusions (cont.)

- ◆ Preliminary comparison between EGD and with SW based methods on cloud amount indicated large differences in certain cases as illustrated on graphics of fig. 2.
- ◆ One difference is due to cloud spatial distribution, over estimating cloud amount as shown on fig. 3. Larger averaging intervals reduce these problems related to temporal resolution effects.
- ◆ Another difference is due to low sensitivity of SW piranometers to thin clouds under estimating their amount, as shown on fig. 4.
- ◆ Current work was restricted to evaluation of clouds amounts only, but more features could be explored and are under investigation on richer information that could be obtained from images. Fig. 5 illustrates the potential data that could be obtained like blue sky gradient related to atmospheric contents and optical density of clouds based on full color space feature analysis.

Future work

- ◆ Further investigation is being done to expand the current analysis to other SW and image based methods and longer data series.



Figure 3 sky image showing piranometer overestimation of cloud coverage, point A in the graphic, due to clouds spatial distribution.



Figure 4 sky image showing piranometer underestimation of cloud coverage, point B in the graphic, due to low sensitivity to thin clouds.

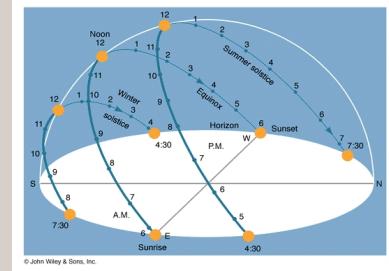


Figure 1 Typical example of annual variation of Solar path for a northern hemisphere site. Illustration obtained from: <http://www.redrok.com/images/sunpath.gif>.

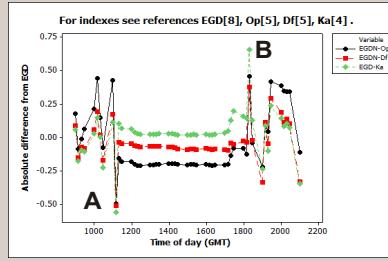


Figure 2 graphic indicating differences between Euclidean Geometric Distance and piranometer based methods for estimation of cloud coverage.

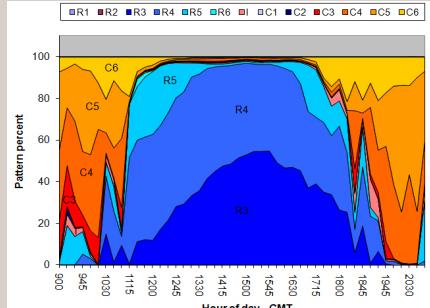


Figure 5 typical evolution of Daily atmospheric patterns obtained from surface cameras, categorized as cloud (C) and Rayleigh scattering(R). Figure adapted from [6].

References

- [1] WMO. Guide to Meteorological Instruments and Methods of Observations. 7. ed. World Meteorological Organization, 2008. WMO-No. 8 I.15-1 I.15-11.
- [2] Long, C. N. et al. Retrieving cloud characteristics from ground-based daytime color all-sky images. Journal of Atmospheric and Oceanic Technology, 23: 633-652, 2006.
- [3] Jain, A., Duin, R., Mao, J. Statistical Pattern Recognition: a review. IEEE Transactions on Pattern Recognition Analysis and Machine Intelligence Jan. 2000, vol. 22, n. 12.
- [4] Karsten F. and Czeplak G. Solar and Terrestrial Radiation Dependent on the Amount and Type of Cloud. Solar Energy 1980 Vol. 24 pp. 177-189.
- [5] Harrison R. G., Chalmers N., Hogan R. J. Retrospective Cloud Determinations from Surface Radiation Measurements 2008 Atmospheric Research vol. 90 pp. 54-62.
- [6] Neto SLM, von Wagenheim A, Pereira EB, Comunello E. The use of Euclidean geometric distance on RGB color space for classification of sky and cloud patterns. Journal of Atmospheric and Oceanic Technology: In Press EOR jan. 2010.
- [7] Duchon, C. E. and O'Malley, M. S. Estimating cloud type from piranometer observations. Journal of Applied Meteorology v. 38, p. 132-141, 1999.
- [8] Tenenbaum, J.B., Griffiths, T. L., Kemp, C. Theory based Bayesian Models of Inductive Learning and reasoning. Trends in Cognitive Sciences 2006, 10(7) 309-318, special issue: Probabilistic models of cognition.
- [9] NIST/SEMATECH e-Handbook of Statistical Methods. <http://www.itl.nist.gov/div898/handbook/>, mar 2010.
- [10] Johnson, R. A. and Wichern D. W. Applied Multivariate Statistical Analysis. Pearson -Prentice Hall 2007, 773 pp.
- [11] Montgomery, D. C. Design and Analysis of Experiments, John Wiley and Sons 2005, 643 pp.

Results and conclusions

- ◆ Current work confirmed that SW based methods are more sensitive to clouds on sunlight pathway than image based methods, due to systematic trajectory of the sun. For northern hemisphere observation sites, southern occurrence of clouds are under evaluated by SW methods. On southern hemisphere observation sites the problem is inverse as illustrated on fig. 1.



INPE - Brazilian Institute for Space Research
www.inpe.br



CST - Centro de Terrestrial System Science
Centro de Ciência do Sistema Terrestre



FINEP - Financer of Studies and Projects
www.finep.gov.br

First Conference on Computational Interdisciplinary Sciences
São José dos Campos - SP - Brazil
August 23-27, 2010
<http://www.lac.inpe.br/CCIS/>



UFSC - Federal University of Santa Catarina
www.ufsc.br



INCoD - National Research Institute on Digital Convergence and Knowledge Engineering Department
www.incod.ufsc.br



Knowledge Engineering Department
www.egc.ufsc.br

Apêndice A6

Journal of Computational Interdisciplinary Sciences (submitted).

Contribuição da publicação em relação à pesquisa desenvolvida. Uma descrição mais detalhada acerca. Comparação da metodologia desenvolvida com metodologias previamente estabelecidas e a comparação preliminar dos desempenhos.

jcis@epacis.org

Preliminary Comparisson of Short Wave Radiometer Data and Euclidean Geometric Distance Image Based Method for Cloud Cover Evaluation from Surface

Sylvio Luiz Mantelli Neto^{a,b,c,1}, Aldo von Wangenheim^{b,c}, Enio Bueno Pereira^a, and Leandro Coser^{b,c}

^aBrazilian Institute for Space Research, São José dos Campos, SP, Brazil

^bNational Intitute of Science and Technology for Digital Convergence,

Universidade Federal de Santa Catarina, Florianópolis, SC, Brazil

^c Departamento de Engenharia e Gestão do Conhecimento and LABSOLAR at Universidade Federal de Santa Catarina, Florianópolis, SC, Brazil

Received on July 30, 2010 / accepted on *****, 2010

Abstract

The purpose of this research is the comparison of two methods commonly used as a ground truth validation for cloud coverage of satellite images. Most used ground truth classifications of sky conditions are done by synoptic observers (SO) based on highly human-oriented subjective analysis [1]. Replace SO evaluation by automatic synoptic observation systems (SOS) is not a trivial task due to the uncertainty involved in the subjective SO evaluation methods. Present work makes a quantitative comparison between SOS methods using shortwave radiometers (SW) data [2] [3], with image based Euclidean Geometric Distance (EGD) [4]. All of them used to evaluate cloud coverage from surface. Comparison analysis indicated that SW methods are much more sensitive to the clouds in the sunlight pathway than out of pathway, causing a bias related to spatial distribution of clouds. Preliminary analysis indicated an overestimation of the sky cover up to 65% in certain cases. SW also underestimate the presence of thin clouds, with noticed differences up to -56% on the evaluation. As a conclusion current research indicated that EGD, and it is believed that any other image based method, is much less susceptible to cloud spatial distribution than SW radiometers, and could be used as a better evaluation method for cloud cover ground truth validation.

Keywords: Cognitive knowledge modeling and evaluation, Bayesian multivariate statistical discriminator.

¹E-mail Corresponding Author: sylvio.mantelli@inpe.br

1. INTRODUCTION

Ground truth validation for satellite-based physical computational models are always used to reduce the uncertainties involved at the remote sensing [5], weather forecasting, large scale assessments, climatologic trends etc. One of the methods used for clouds evaluation is based on qualitative and quantitative analysis by surface observations according to criteria recommended by World Meteorological Organization (WMO) [1]. Most variables used by surface validation is based on quantitative data like temperature, atmospheric pressure, humidity, wind, etc. obtained from synoptic stations. Surface sensors in general, are more precise than satellite evaluations, but satellites have advantage of larger spatial resolution coverage. Nevertheless both evaluation methods (satellite and surface) should be considered complementary and used simultaneously in the applied research areas as recommended by WMO [1]. Combined analysis is also recommended to avoid disadvantages of each observation system and to reduce uncertainties involved in the process [6]. Surface data in current work is intended to be used to support qualitative analysis of satellite numerical models [1], in clouds coverage. One important aspect of satellite image based methods is the evaluation of clouds. Clouds plays a major role on those models and its effects are very important modulating the amount of energy that reaches the surface, transporting large amount of water vapor and heat, obstructing remote sensing targets, etc. Cloud effects are easier to evaluate by satellite models under stable conditions of clear or fully overcast skies. But for mixed and intermediate conditions uncertainties are accentuated due to the difficulty on the evaluation of the effects and identification of the occurrence of scattered clouds [9]. Surface validations of clouds could be done by Synoptic Observers (SO) or Synoptic Observation Systems (SOS) [1]. SO evaluations follow WMO [1] procedures and are highly subjective on the qualitative and quantitative analysis of now casting conditions. To avoid the problems involved in SO evaluations [7], SOS are being employed in the systematic assessment of cloud conditions. There are several SOS automatic methods being used for clouds evaluation from surface like ceilometers, Infra red (IR) or long wave radiometers, short wave (SW) radiometers and sky imagers (SKI). In the current work only SW and EGD SKI methods will be compared. SW cloud evaluations are based on parametric models considering the dependence of radiation fluxes on cloud amount and type [2] or cloud forcing [7]. Usually these models take into account the difference between clear-sky parametrization and current (or all-

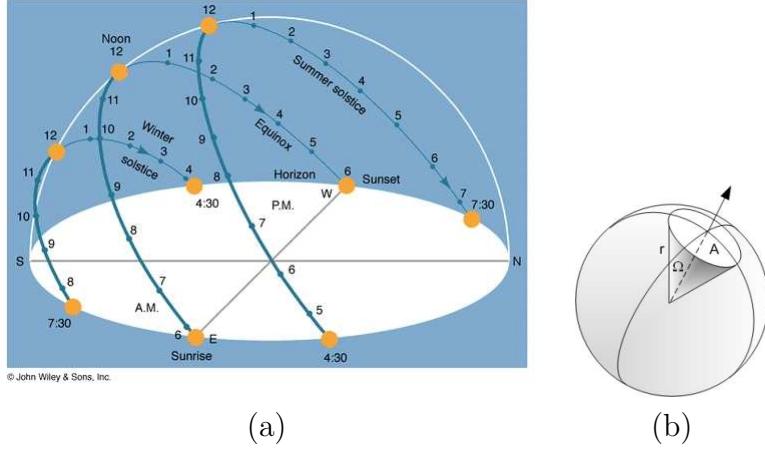


Figure 1: Problems related to spatial distribution of clouds. (a) Typical example of annual variation of Solar path for a northern hemisphere site. Illustration obtained from: <http://www.redrok.com/images/sunpath.gif>. (b) Illustration showing solar disk solid angle used to represent the whole sky domain to estimate cloud coverage.

sky) conditions, like the models developed by [8], [3], [10], [11], [12], [14].

Parametrization could be done by evaluation Top of Atmosphere (TOA) irradiance combined to Global and Difuse surface measurements [3], or using surface time series characterization of clear-sky [12]. But SW methods lacks of proper domain sampling and data seeding only to clouds located in solar path to surface sensors as illustrated on Fig. 1 (a). As could be noticed on example illustrated on Fig. 1 (a), northern occurrence of clouds will be discarded on sky cover evaluations (on northern hemispheric sites). That fact was also outlined by [8], and its validation by SO ignored the observations at southern half of the sky making that method not very representative of cloud/sky evaluation. Fig. 1 (b) also indicates another aspect of sampling domain. Solar disk has an opening angle of 5.5 or 6.8 10^{-5} sr according to [15] of sky hemisphere solid angle representing only 0.000354 % of the hemispheric domain. That method makes the SW-based methods more sensitive to clouds in the sun pathway, making the sampling measurements not completely representative of the whole celestial hemisphere [27].

Data averaging and longer integrating intervals reduce the effects of er-

rors, but increase the evaluation uncertainties. [8] also considered uniform clouds distribution and that clouds velocity are faster than sun at observation site. These conditions is not always present, making SW based models even more dependent of site specific atmospheric conditions.

SKI are SOS that evaluates the sky using surface cameras like [16], [17], [18] or camera systems [31]. Although the device used for acquisition and accessories are different, the main issue is the method used to detect the patterns present on images. [16] uses supervised statistical training on saturation channel only of Hue Saturation Lightness (HSL) color space. Discrimination function is based on three standard deviation (3σ) saturation values of established patterns classified as cloud, sky and undetermined. Hue and Lightness channels are not analyzed and sun was kept always out of camera field of view (FOV), leaving high intensity patterns out of domain analysis. [19] used a linear discrimination based on empirical ratio of 0.6 Red and Green components (0.6 R/G) of RGB color space for classification of cloud, sky patterns. Blue channel is not analyzed and image is obtained from a camera pointed to a reflector. A solar tracker is also used to avoid direct sun light by a shading band. [19] method restricts the observation domain to R and G color channels only, leading to information loss on pattern analysis [22]. [17] used neural network (NN) normalized by parametric clear sky model suggested by [20] on RGB color space images. Discrimination function was based on supervisioned NN-training of typical pixel values of cloud and clear sky images. But binary outcome approach expected by this modeling leads to classification errors due to occurrence of more than two different visual patterns present in atmospheric domain [22], [28]. [18] used genetic algorithms (GA) optimization of 3 from 18 input parameters obtained from a 1 and 9 pixel window. Parameters obtained by GA analysis were average values inside of Red and Blue windows and the variance of the Red component of pixel values. Discrimination function classifies patterns using a supervised-trained Multilayered Perceptron Network (MLP) in 3 categories of clear sky, opaque and thin clouds. Green channel was not considered leading again to domain information loss due to dimensioning reduction [22], [28]. This fact could be confirmed by [21] observation that clouds is the result of equally likely scattering of red, green and blue color components. [4] used Bayesian geometric distance methods considering the typical occurrence of cloud and sky patterns on red, green and blue (RGB) color space. Patterns where characterized by supervised multivariate statistics and classified by typical Euclidean Geometric Distance (EGD) of these patterns from main

diagonal. This method allowed statistical-based discrimination of patterns invariant from lightness of blue skies and clouds patterns. EGD also used the three dimensions of RGB color channels and the full information available on color domain. Thats the reason why EGD was selected to be compared to SW based methods. Current work will made comparison among SW-based methods and EGD image-based methods for cloud assessments and outline their main differences on surface cloud amount evaluations. It will start with a short description of SW methods used and EGD method. Then it will be made a graphical comparissson of their data followed by the final conclusions.

2.MATERIAL AND METHODS

SW data of a Global and Difuse radiometers, used for the experiment were taken from a one selected observation day in LABSOLAR BSRN² station facilities at Federal University of Santa Catarina Florianopolis SC Brazil. The 49 images used on the experiment were taken from a Total Sky Imager (TSI) model 440³ co-allocated in the same site. Sky images were taken every 15 minutes interval and the SW Global and Difuse irradiance data sampled every second and collected in one minute interval averages.

[2] parameterized clouds from surface based on 10 years observation data taken from global and difuse SW piranometers at Hamburg-Germany meteorological observatory. They compared the SW data with SO observation, proposing an expression relating Global, Difuse and cloud cover. Rearranging their expression as a function of cloud cover, and using percent instead of eights, we arrived to the equation 1.

$$C_{KC} = \sqrt{\left(\frac{D}{G} - 0.3\right) \frac{1}{0.7}} \quad (1)$$

Where:

- CKC: is the adapted [2] cloud cover in percent
- D: Difuse Radiation
- G: Global Radiation

²www.bsrn.ethz.ch

³www.yankeeenvironmental.com

[5] also parametrized clouds from surface based on 10 years SW Global and Difuse data using two indexes, Opacity(Op) and Difuse Fraction (Df), described on the equations 2 to 5.

$$D_f = \frac{D}{G}, \quad (2)$$

$$O_p = 1 - K_t, \quad (3)$$

$$K_t = \frac{D}{E}, \quad (4)$$

$$E = \frac{S_0 \cos \theta}{d^2} \quad (5)$$

Where:

- Df: is the [3] Diffuse Fraction index
- D: Difuse Radiation
- G: Global Radiation
- Op: is the [3] Opacity Index
- Kt: Clearness Index
- E: Top of Atmosphere Radiation
- S_0 : Average Solar Constant
- d: Sun-Earth distance
- θ : Sun Zenith Angle

EGD [4] used images to evaluate cloud amount based on a typical occurrence (or locus) on RGB color space. Specific pixel values represented typical sky and cloud patterns. A Bayesian cognitive modelling [24] was applied on image patterns of cloud and sky. An exploratory data analysys (EDA) with outlier elimination [25] and feature extraction [27], were used to characterize these patterns by multivariate statistics [26] during supervised learning phase [28]. Central limit theorem [25], multivariate statistical were used on

current research method. Pixels patterns typical locus were characterized by Euclidean Geometric Distance (EGD) and Projection (EGP), on 3D color space to define cloud and sky patterns. Discrimination function was based on pattern typical distance from main diagonal of color space. Fig. 2 describes the characterization of typical geometric distances on color space.

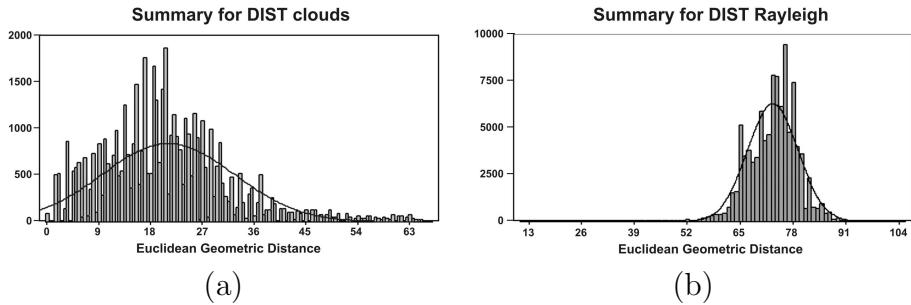


Figure 2: EGD characterization of cloud on (a) and sky on (b) patterns. Figure adapted from [4], where Rayleigh scattering is referenced as sky.

3. RESULTS AND CONCLUSIONS

The evaluation of cloud amount of the three methods was implemented. Data comparison confirmed that SW based methods are more sensitive to clouds on sunlight pathway than image based methods [8], due to systematic trajectory of the sun. For northern hemisphere observation sites, southern occurrence of clouds are under evaluated by SW methods. On southern hemisphere observation sites the problem is inverse. Preliminary comparison between EGD and with SW based methods on cloud amount also indicated large differences in certain cases as illustrated on graphics of fig. 3(a). One difference is due to cloud spatial distribution, over estimating cloud amount as shown on fig. 3(b). Larger averaging intervals of SW data reduce these problems related to temporal resolution effects. Another difference is due to low sensitivity of SW pyranometers to thin clouds under estimating cloud amount, as shown on fig. 3(c). For clear skies, correspondent to the central region of the graph on fig. 3(a), data indicated a positive bias of 2% for Ka, and negative ones -8 % for Df and -20% for Op. For intermediate cloud conditions, correspondent to extreme parts of the graph, preliminary analysis indicated an overestimation of the sky cover up to 65% for Ka, 37 % for Df

and 45 % for Op. The effects of over estimation were smaller for Df index. SW also underestimate the presence of thin clouds, with noticed differences up to -56% for Ka, -51% for Df and -49% for Op. on the evaluation. Df indicated the smallest over estimation but the differences of the three methods have almost the same result. As a conclusion current research indicated that EGD, and it is believed that any other image based method, is much less susceptible to cloud spatial distribution than SW radiometers, and could be used as a better evaluation method for cloud cover ground truth validation.

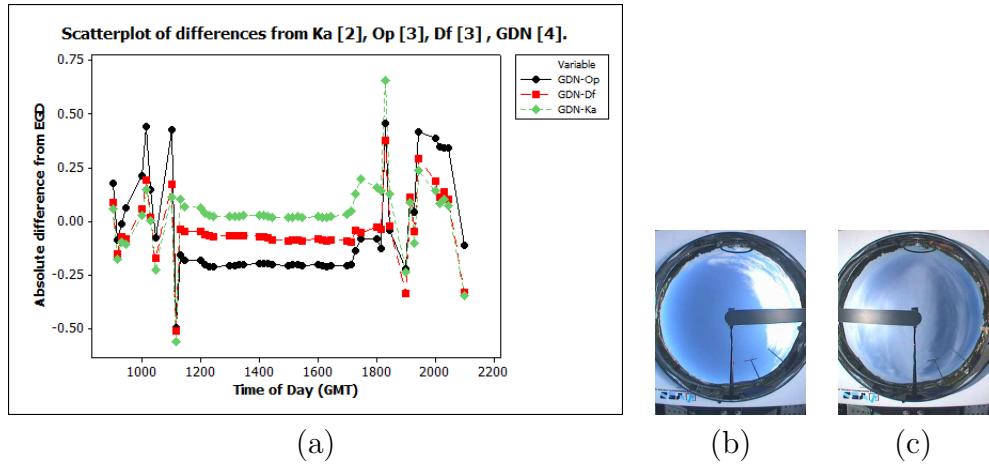


Figure 3: Graphic indicating differences between Euclidean Geometric Distance and piranometer based methods for estimation of cloud coverage (a). On (b) an sky image showing piranometer overestimation of cloud coverage, responsible for positive differences on the graphic, due to clouds spatial distribution. On (c) an sky image showing piranometer underestimation of cloud coverage, responsible for negative differences on the graphic, due to low sensitivity of SW to thin clouds.

ACKNOWLEDGMENTS: This work was made possible thanks to the SONDA project sponsored by FINEP (22.01.0569.00) and by PETROBRAS (0050.0029348.07.4) Authors also thank to Prof. Sergio Colle from LABSOLAR-EMC-UFSC for providing facilities for the experimental set-up.

References

- [1] WMO. Guide to Meteorological Instruments and Methods of Observations. 7. ed. World Meteorological Organization, WMO-No. 8, pp. 1.15.1-1.15.11, 2008.
- [2] KARSTEN F. and CZEPLAK G. Solar and Terrestrial Radiation Dependent on the Amount and Type of Cloud. *Solar Energy*, vol. 24, pp. 177-189, 1980.
- [3] HARRISON R. G., CHALMERS N., HOGAN R. J. Retrospective Cloud Determinations from Surface Radiation Measurements. *Atmospheric Research*, vol. 90, pp. 54-62, 2008.
- [4] MANTELLI NETO S. L. , von WANGENHEIM A., PEREIRA E. B., COMUNELLO E. The use of Euclidean Geometric Distance on RGB Color Space for Classification of Sky and Cloud Patterns. *Journal of Atmospheric and Oceanic Technology*: In Press EOR jan. 2010.
- [5] HARRISON E. F., MINNIS P., BARKSTROM B. R., GIBSON G. G.. Radiation Budget at the Top of the Atmosphere. *Atlas of Satellite Observations Related to Global Change*, 35 p. 1995.
- [6] ROSSOW, W and SCHIFFER R. A. Advances in Understanding Clouds From ISCCP. *Bulletin of the American Meteorological Society* v. 80, pp. 2261-2287, 1999.
- [7] NORRIS J. R. Multidecadal changes in near-global cloud cover and estimated cloud cover radiative forcing. *Journal of Geophysical Research*, v. 110, 17 p., 2005
- [8] DUCHON, C. E. and O'MALEY, M. S. Estimating Cloud Type from Piranometer Observations. *Journal of Applied Meteorology* v. 38, pp. 132-141, 1999.
- [9] MARTINS F. R., SILVA S. A. B., PEREIRA E. B. and ABREU S. L. The influence of cloud cover index on the accuracy of solar irradiance estimates. *Meteorology and Atmospheric Physics*, v. 99, pp. 169-180, 2008.

- [10] CESS R. D., ZHANG M. H., MINNIS P., CORSETTI L., DUTTON E.G., FORGAN B.W., GARBER B.P., GATES W.L., HACK J.J., HARRISON E.F., JING X., KIEHI J.T., LONG C.N., MORCRETTE J.J., POTTER G.L., RAMANATHAN V., SUBASILAR B., WHITLOCK C.H., YOUNG D.F., ZHOU Y. Absorption of Solar Radiation by Clouds: Observations Versus Models. *Science* v. 267(5197), pp. 496-499, 1995.
- [11] LONG, C. N. Studies of the Effect of Clouds on Solar Irradiance Using Automated Clear Sky Detection and Empirical Fitting Algorithms. In: Proceedings of the sixth atmospheric radiation measurement (ARM) team meeting. [s.n.], 1997. p. 457-460. Available in: http://www.arm.gov/publications/proceedings/conf07/extended_abs/long_cn.pdf.
- [12] LONG, C. N.; ACKERMAN, T. P. Detection of Clear Skies Using Total and Difuse Shortwave Irradiance Calculations of Shortwave Cloud Forcing and Clear Skies Difuse Ratio. In: Proceedings of the Sixth Atmospheric Radiation Measurement (ARM) Team Meeting. [s.n.], p. 179-183 1996. Available in: http://www.arm.gov/publications/proceedings/conf06/extended_abs/long_cn.pdf.
- [13] LONG. C.N., ACKERMAN T.P., GAUSTAD K.L., COLE J.N.S. Estimation of Fractional Sky Cover from Broadband Shortwave Radiometer Measurements. *Journal of Geophysical Research*. v. 111, n. D111, D11204.1-D11204.11, 2006.
- [14] WALISSER, D.E., COLLINS W.D., ANDERSON S.P. An Estimate of the Surface Shortwave Cloud Forcing Over the Western Pacific During TOGA COARE. *Geophysical Research Letters* v. 23 n. 5 pp. 519-522. 1996.
- [15] LENOBLE, J. *Atmospheric Radiative Transfer*. A. Deepak Publishing, 532 p. 1993.
- [16] SOUZA-ECHER M.P., PEREIRA E.B., BINS L.S., ANDRADE M.A.R. A Simple Method for the Assessment of the

Cloud Cover State in High-Latitude Regions by a Ground-Based Digital Camera. *Journal of Atmospheric and Oceanic Technology*, v. 23, n. 3, p. 437-447, 2006.

- [17] MANTELLI, S. L. Desenvolvimento de uma Nova metodologia para a estimativa da cobertura de nuvens usando uma câmera de superfície e comparando com imagens de satélite. Dissertao da Universidade Federal de Santa Catarina, Departamento de Informática e Estatística, 2001.
- [18] CAZORLA, A.; OLMO, F. J.; ALADOS-ARBOLEDAS, L. Development of a Sky Imager for Cloud Cover Assessment. *Journal of the Optical Society of America, OSA*, v. 25, n. 1, pp. 29-39, 2008.
- [19] LONG C. N., SABBURG J.M., CALBO J., PAGÈS D. Retrieving cloud characteristics from ground-based daytime color all-sky images. *Journal of Atmospheric and Oceanic Technology*, v. 23 pp. 633-652, 2006.
- [20] IQBAL, M. *An Introduction to Solar Radiation*. New York Academic Press, 1983.
- [21] LILLESAND, T. M.; KIEFER, R. W. *Remote Sensing and Image Interpretation*. John Wiley and Sons, 750 p., 1994.
- [22] JAIN A. , DUIN R., MAO J. Statistical Pattern Recognition: a review. *IEEE Transactions on Pattern Recognition Analysis and Machine Intelligence*, vol. 22, n. 12, jan. 2000.
- [23] ZHANG P. G. Avoiding Pitfalls in Neural Network Research. *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews*, vol. 37, NO. 1, Jan. 2007.
- [24] TENENBAUN J.B., GRIFFTHS T. L., KEMP C. Theory based Bayesian Models of Inductive Learnning and reasoning. *Trends in Cognitive Sciences*, special issue: Probabilistic models of cognition. 10(7), pp. 309-318, 2006.
- [25] NIST/SEMATECH e-Handbook of Statistical Methods, <http://www.itl.nist.gov/div898/handbook/>, mar 2010.

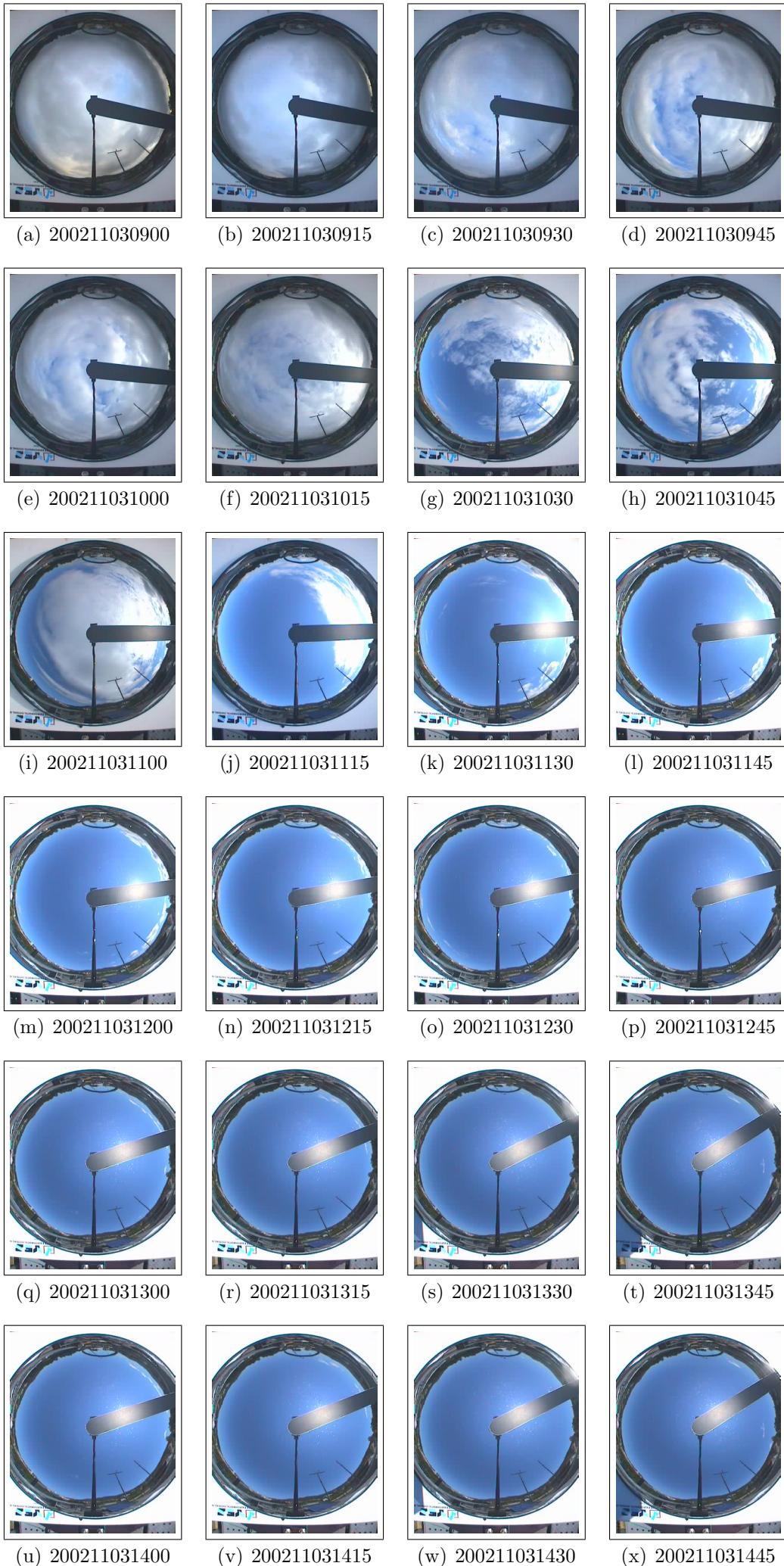
- [26] JOHNSON, R. A. and WICHERN D. W. *Applied Multivariate Statistical Analysis*, Pearson-Prentice Hall, 773 pp., 2007.
- [27] MONTGOMERY, D. C. *Design and Analysis of Experiments*, John Wiley and Sons, 643 pp., 2005.
- [28] Zhang H., Fritts J. E., GOLDMAN S. A. *Image Segmentation Evaluation: A Survey of Unsupervised Methods*. v. 110,2 pp. 260-280, 2008.

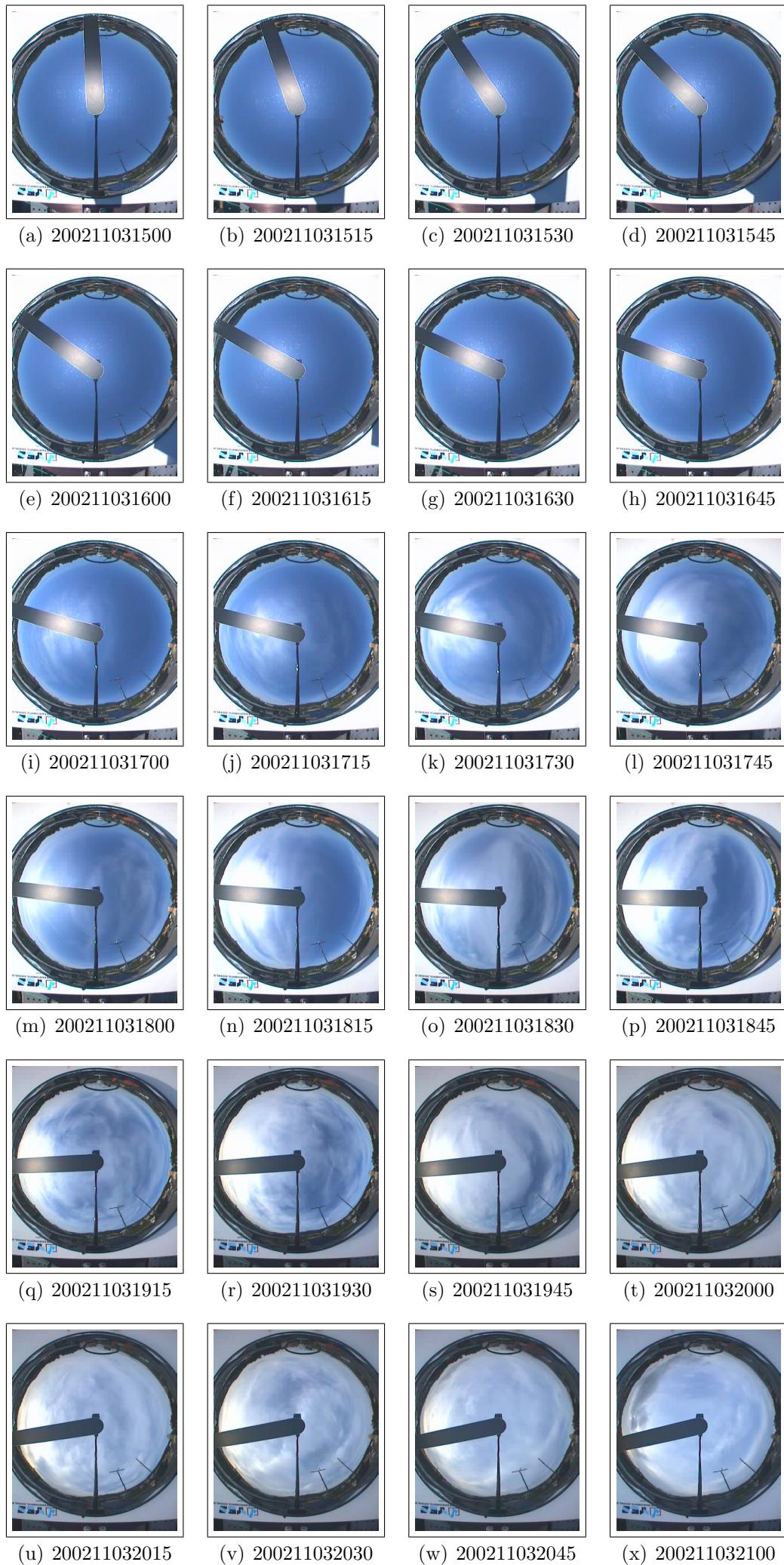
Anexo A Conjuntos de Imagens Originais, Mascaradas, Analisadas e do Satélite GOES-8

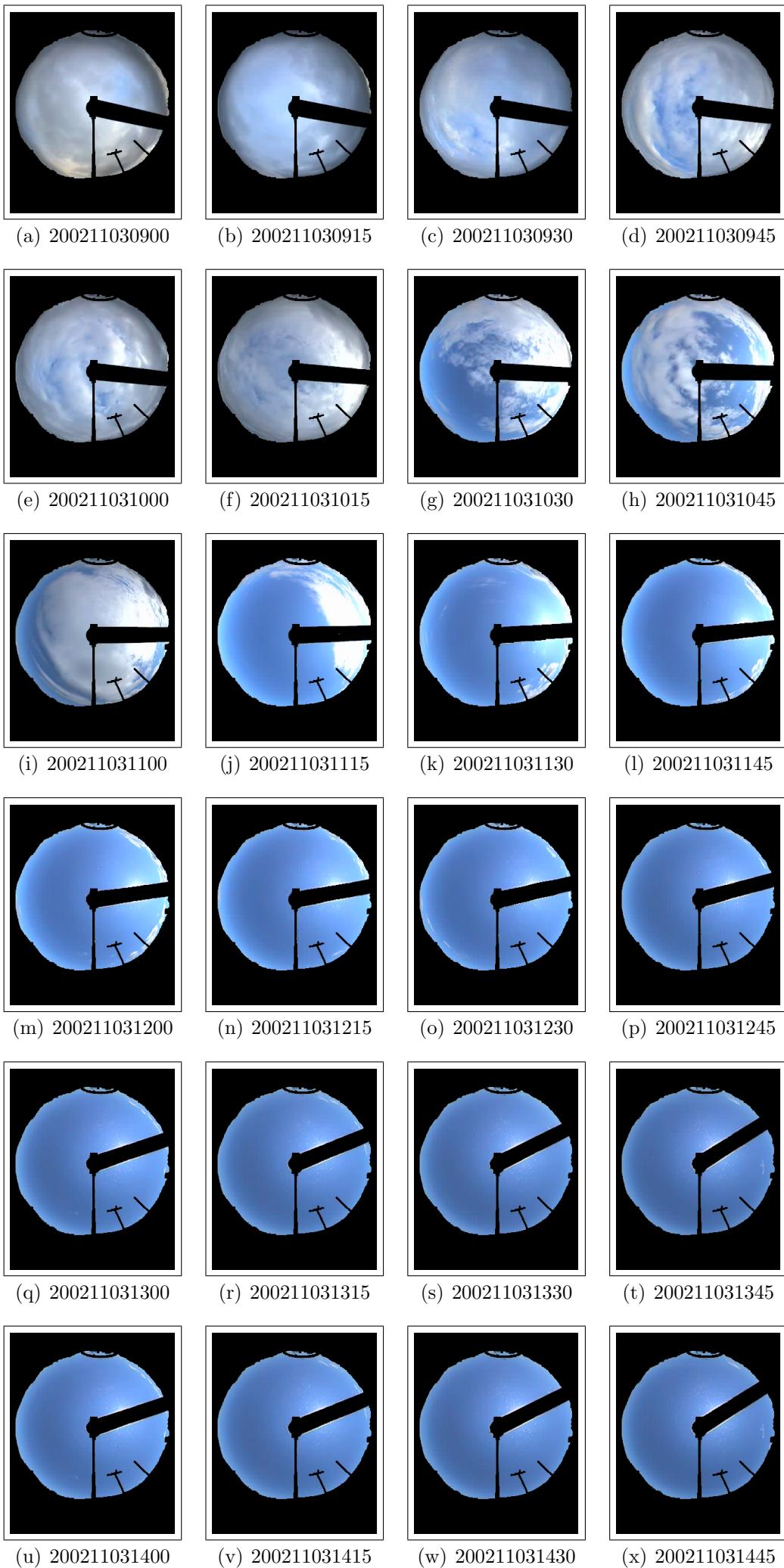
No presente anexo estão ilustradas imagens em quatro conjuntos de figuras. O primeiro conjunto, 5.1 e 5.2 contem as imagens originais obtidas do imageador de nuvens. O segundo conjunto, 5.3 e 5.4, contem as imagens mascaradas dos fatores que interferem na análise da imagem. O terceiro conjunto, 5.5 e 5.6 contem as imagem analisadas segundo código descrito na legenda da figura 5.7, ao final do anexo. Uma síntese tabular dos resultados encontra-se no capítulo 4, nas tabelas 4.3 e 4.4. O quarto conjunto de figuras 5.8, 5.9, 5.10, 5.11 e 5.12, ilustra as imagens do satélite GOES-8 que foram obtidas próximas às do imageador de nuvens.

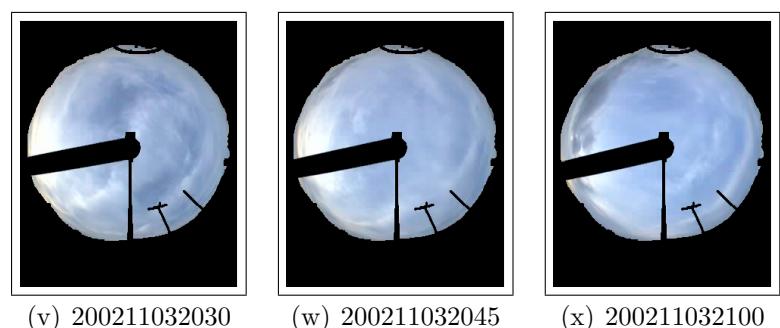
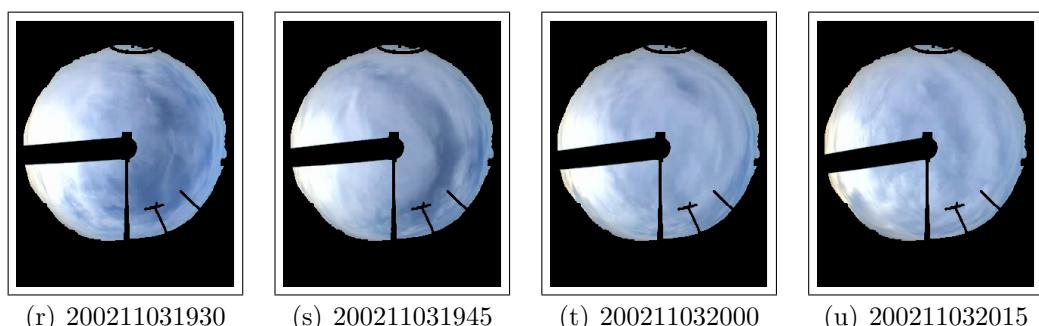
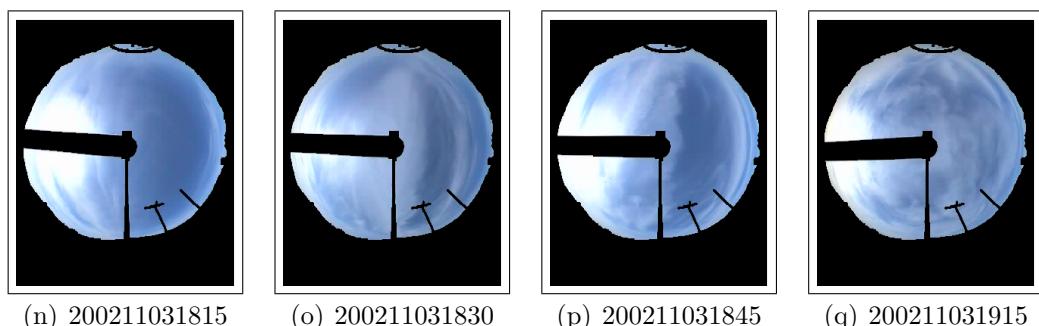
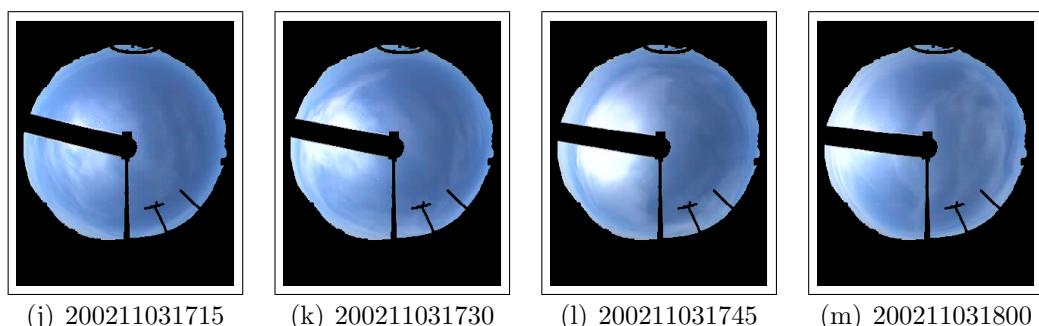
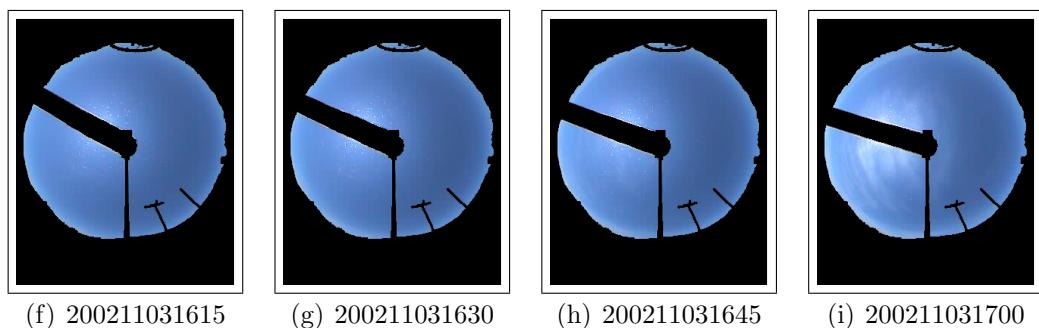
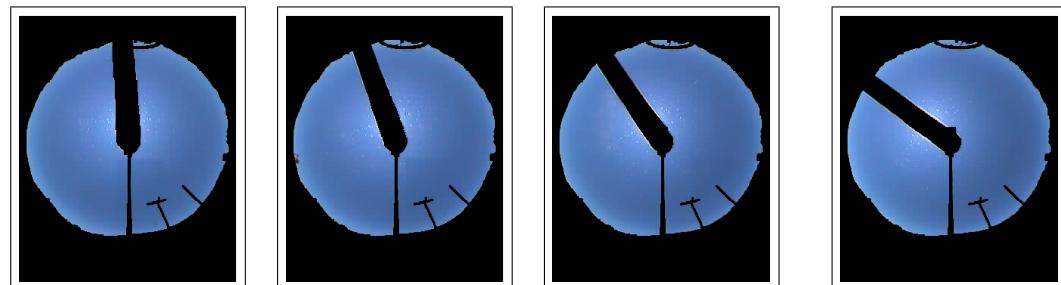
Table 5.1: Dados das comparações entre os métodos encontrados na literatura dos métodos (LONG; SLATER; TOOMAN, 2001)(L), (MANTELLI et al., 2010b)(E) e o proposto no presente trabalho(M).

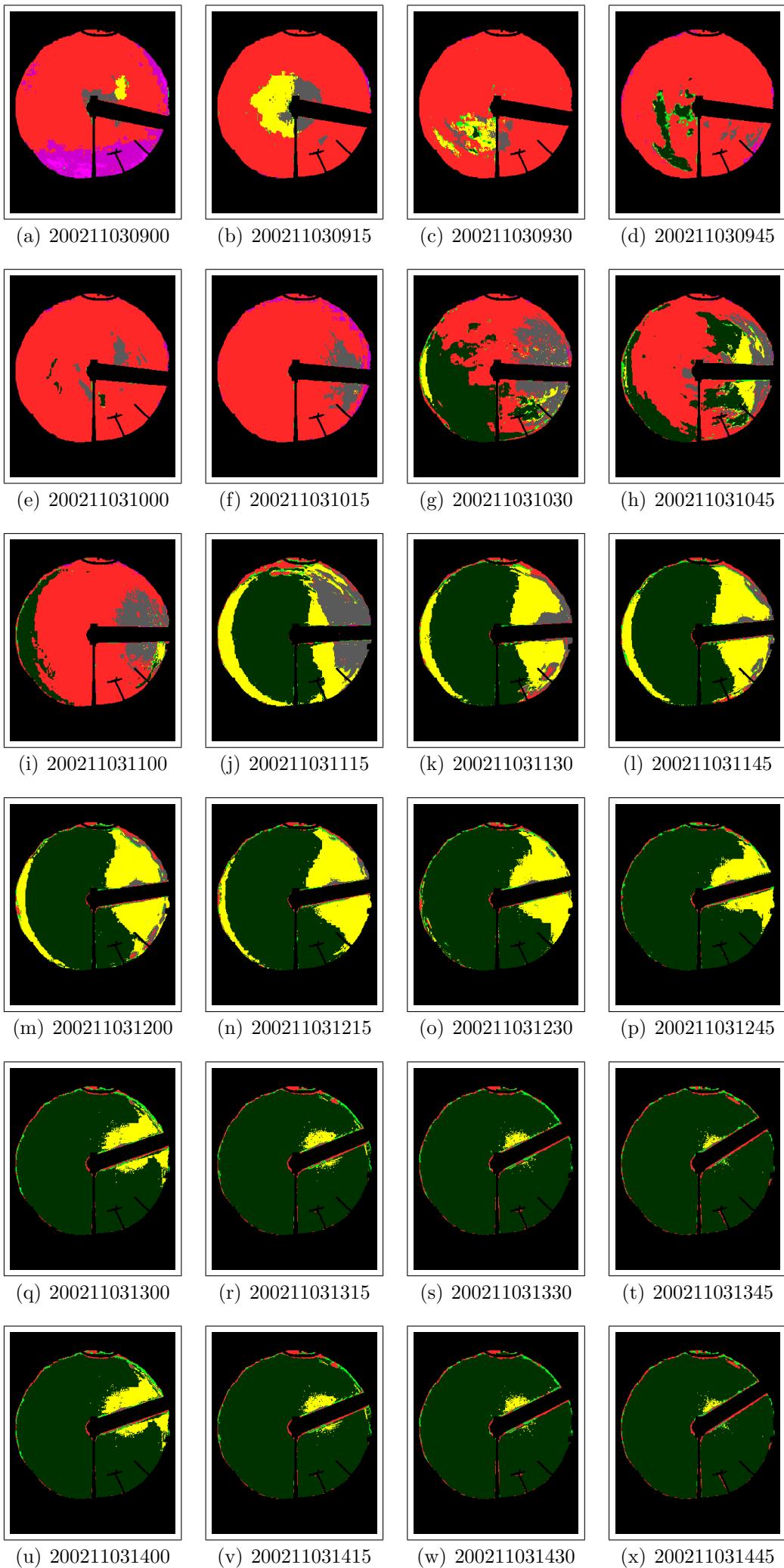
hora	sky L	sky E	sky M	sky M-E	sky M-L	cloud L	cloud E	cloud M	cloud M-E	cloud M-L
900	0.19	1.81	1.25	-0.56	1.06	99.81	97.78	77.77	-20.01	-22.04
915	0.20	19.18	10.62	-8.56	10.42	99.80	75.27	88.63	13.36	-11.17
930	1.32	13.83	5.04	-8.79	3.72	98.68	82.24	93.59	11.35	-5.09
945	7.63	16.25	6.35	-9.90	-1.28	92.37	81.89	90.29	8.40	-2.08
1000	1.80	4.29	0.83	-3.46	-0.97	98.20	94.25	98.41	4.16	0.21
1015	0.28	0.01	0.04	0.03	-0.24	99.72	99.93	96.37	-3.56	-3.35
1030	44.07	49.72	38.47	-11.25	-5.60	55.93	46.63	59.60	12.97	3.67
1045	33.75	38.19	30.14	-8.05	-3.61	66.25	58.66	67.61	8.95	1.36
1100	21.24	15.37	12.64	-2.73	-8.60	78.76	83.06	86.16	3.10	7.40
1115	73.15	77.69	75.78	-1.91	2.63	26.84	21.71	22.48	0.77	-4.36
1130	81.79	89.46	88.36	-1.10	6.57	18.21	9.94	10.28	0.34	-7.93
1145	86.58	92.51	91.18	-1.33	4.60	13.42	6.91	7.08	0.17	6.34
1200	89.47	93.64	92.27	-1.37	2.80	10.53	5.99	6.15	0.16	-4.38
1215	94.08	96.05	94.71	-1.34	0.63	5.92	3.65	3.73	0.08	-2.19
1230	95.48	97.05	95.20	-1.85	-0.28	4.52	2.63	3.05	0.42	-1.47
1245	96.42	97.36	95.68	-1.68	-0.74	3.57	2.28	2.62	0.34	-0.95
1300	96.32	97.24	95.25	-1.99	-1.07	3.68	2.36	2.71	0.35	-0.97
1315	96.58	97.26	95.28	-1.98	-1.30	3.41	2.34	3.02	0.68	-0.39
1330	96.87	97.20	95.41	-1.79	-1.46	3.13	2.42	2.96	0.54	-0.17
1345	96.78	96.85	95.12	-1.73	-1.66	3.22	2.73	3.45	0.72	0.23
1400	97.16	96.69	95.01	-1.68	-2.15	2.84	2.74	4.08	1.34	1.24
1415	97.39	96.82	95.42	-1.40	-1.97	2.61	2.66	3.62	0.96	1.01
1430	97.75	97.19	95.80	-1.39	-1.95	2.24	2.27	3.30	1.03	1.06
1445	98.06	97.67	96.33	-1.34	-1.73	1.94	1.90	2.88	0.98	0.94
1500	98.51	98.10	97.01	-1.09	-1.50	1.49	1.45	2.39	0.94	0.90
1515	98.12	97.65	96.29	-1.36	-1.83	1.87	1.83	2.88	1.05	1.01
1530	97.88	97.38	96.01	-1.37	-1.87	2.11	2.14	3.16	1.02	1.05
1545	98.15	97.65	96.27	-1.38	-1.88	1.85	1.84	2.94	1.10	1.09
1600	97.71	97.55	96.11	-1.44	-1.60	2.29	2.03	2.81	0.78	0.52
1615	97.72	97.95	96.23	-1.72	-1.49	2.28	1.68	2.47	0.79	0.19
1630	97.46	97.93	96.06	-1.87	-1.40	2.54	1.68	2.66	0.98	0.12
1645	96.83	97.54	95.46	-2.08	-1.37	3.17	2.05	3.03	0.98	-0.14
1700	91.86	96.26	93.97	-2.29	2.11	8.13	3.15	4.24	1.09	-3.89
1715	85.88	94.02	89.91	-4.11	4.03	14.12	4.87	6.92	2.05	-7.20
1730	78.23	86.03	81.43	-4.60	3.20	21.77	12.68	15.83	3.15	-5.94
1745	64.74	76.54	44.91	-31.63	-19.83	35.26	19.61	52.58	32.97	17.32
1800	75.79	79.06	54.21	-24.85	-21.58	24.21	15.58	42.76	27.18	18.55
1815	56.58	70.42	54.19	-16.23	-2.39	43.42	27.18	41.97	14.79	-1.45
1830	33.79	25.46	9.66	-15.80	-24.13	66.21	65.44	88.54	23.10	22.33
1845	46.46	66.71	39.97	-26.74	-6.49	53.54	29.52	55.92	26.40	2.38
1915	27.60	27.90	4.34	-23.56	-23.26	72.40	58.58	93.86	35.28	21.46
1930	27.55	22.89	4.58	-18.31	-22.97	72.45	68.21	92.21	24.00	19.76
1945	10.43	2.15	0.50	-1.65	-9.93	89.57	96.98	96.96	-0.02	7.39
2000	2.28	1.86	0.49	-1.37	-1.79	97.72	97.35	96.39	-0.96	-1.33
2015	0.33	0.32	0.10	-0.22	-0.23	99.67	99.32	95.28	-4.04	-4.39
2030	0.14	0.00	0.02	0.02	-0.12	99.86	99.92	94.83	-5.09	-5.03
2045	0.27	0.06	0.03	-0.03	-0.24	99.73	99.77	95.63	-4.14	-4.10
2100	0.54	31.91	0.93	-30.98	0.39	99.46	60.98	97.37	36.39	-2.09

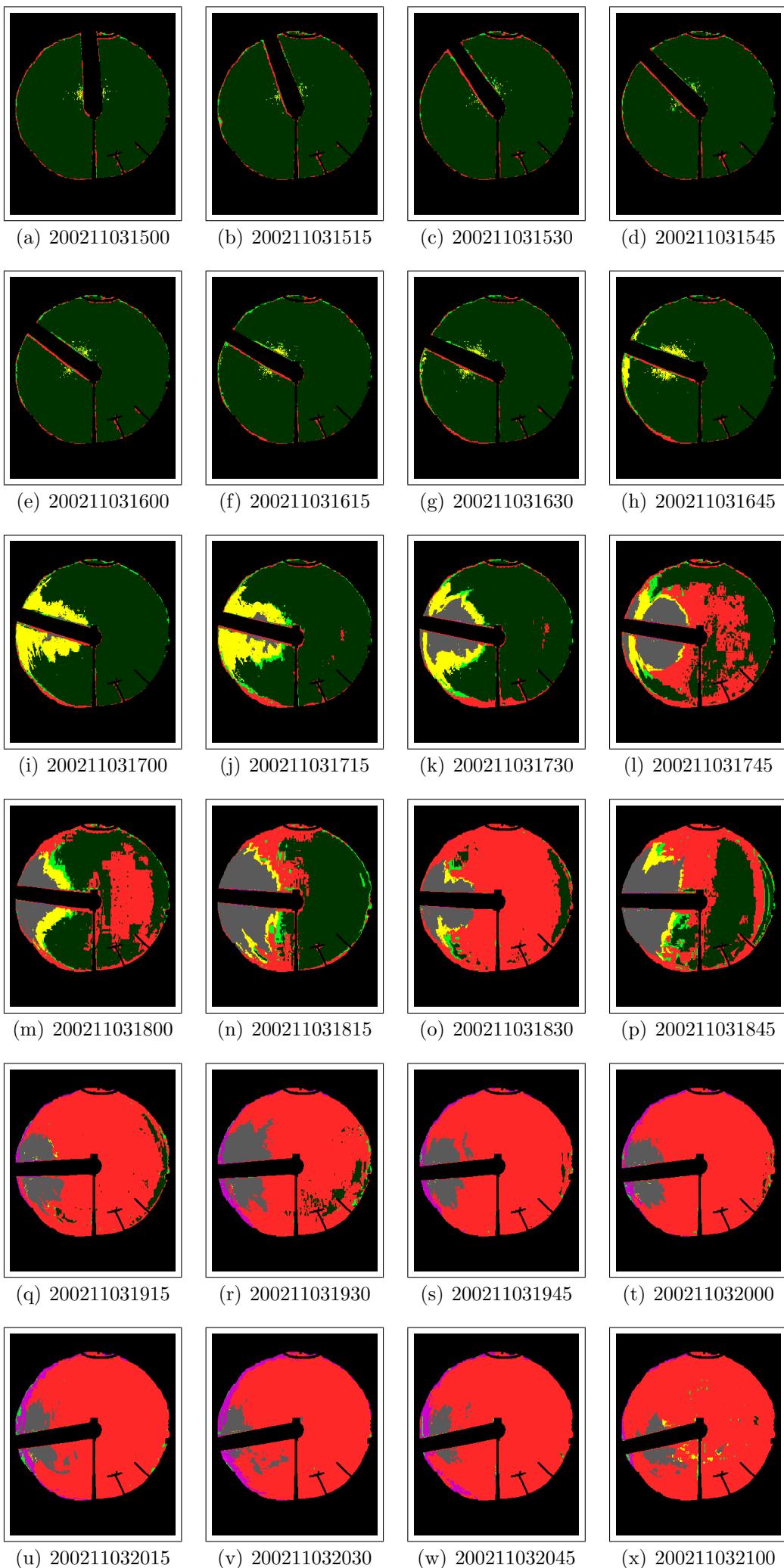












PATTERN	COLOR
MASK	Black
IMAGEDATA	
TOTAL	
NONE	
MIE	Magenta
DENS	Grey
ENS	Red
ENS_TOT	White
INT	Brown
DRAY	Yellow
RAYL	Dark Green
RAY_TOT	White
NC	Green

Figure 5.7: Legenda de cores dos padrões classificados.

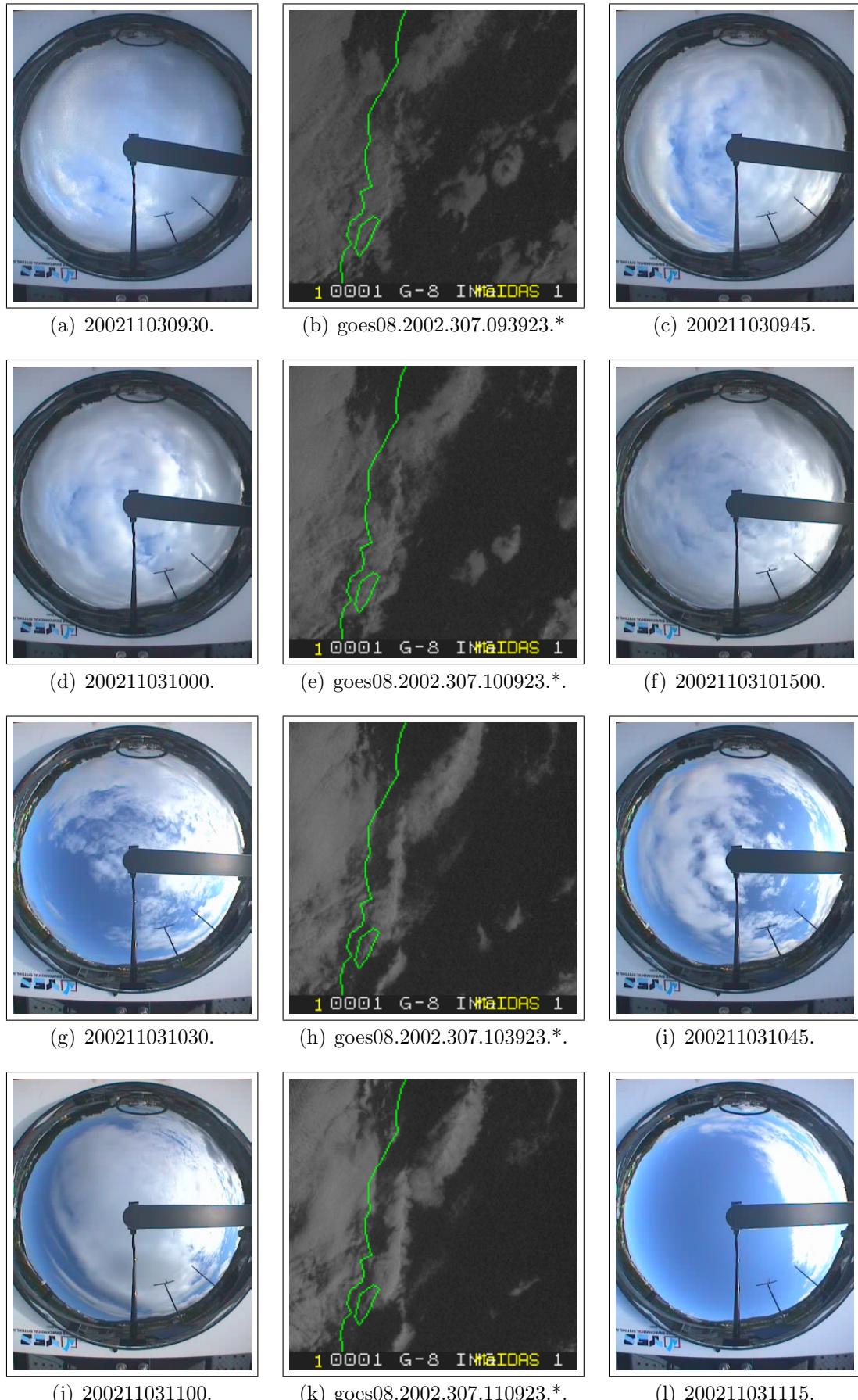


Figure 5.8: Imagens do satélite GOES 8 disponíveis entre 09:39 e 11:09 (b),(e),(h),(k), obtidas em um horário intermediário da imagens do imageador de nuvens antes (a),(d),(g),(j) e depois (c),(f),(i),(l).

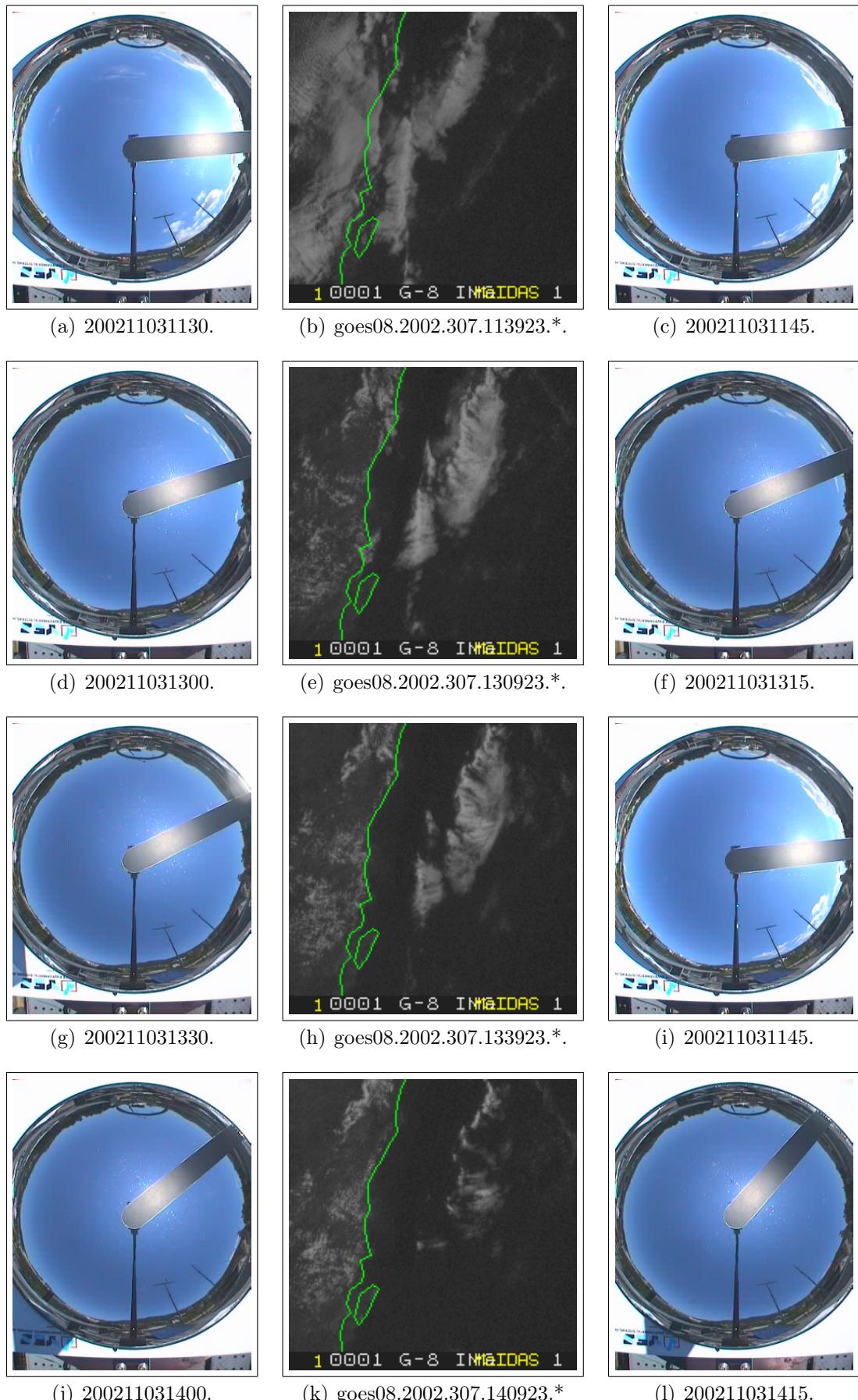


Figure 5.9: Imagens do satélite GOES 8 disponíveis entre 11:39 e 14:09 (b),(e),(h),(k), obtidas em um horário intermediário da imagens do imageador de nuvens antes (a),(d),(g),(j) e depois (c),(f),(i),(l).

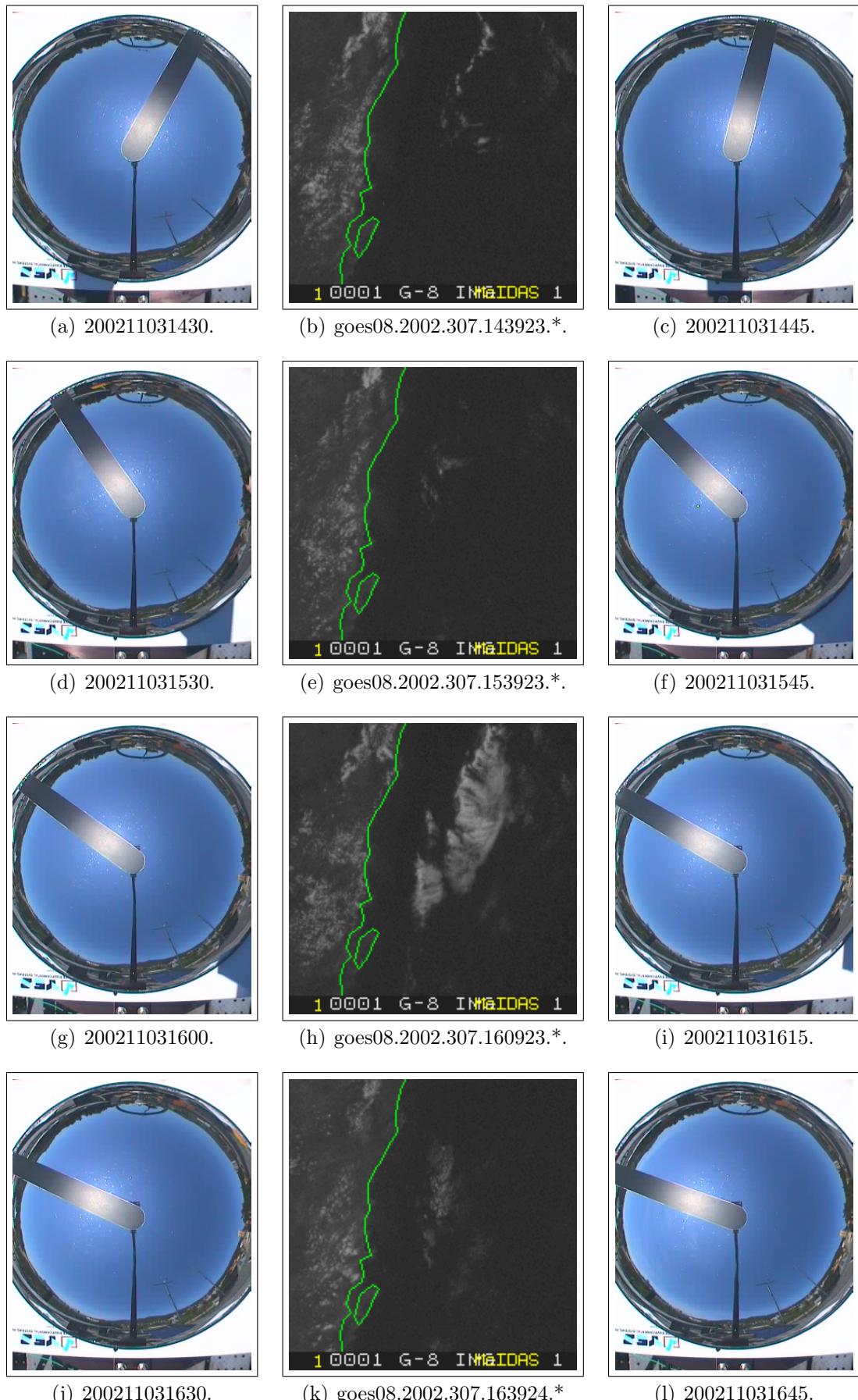


Figure 5.10: Imagens do satélite GOES 8 disponíveis entre 14:39 e 16:39 (b),(e),(h),(k), obtidas em um horário intermediário da imagens do imageador de nuvens antes (a),(d),(g),(j) e depois (c),(f),(i),(l).

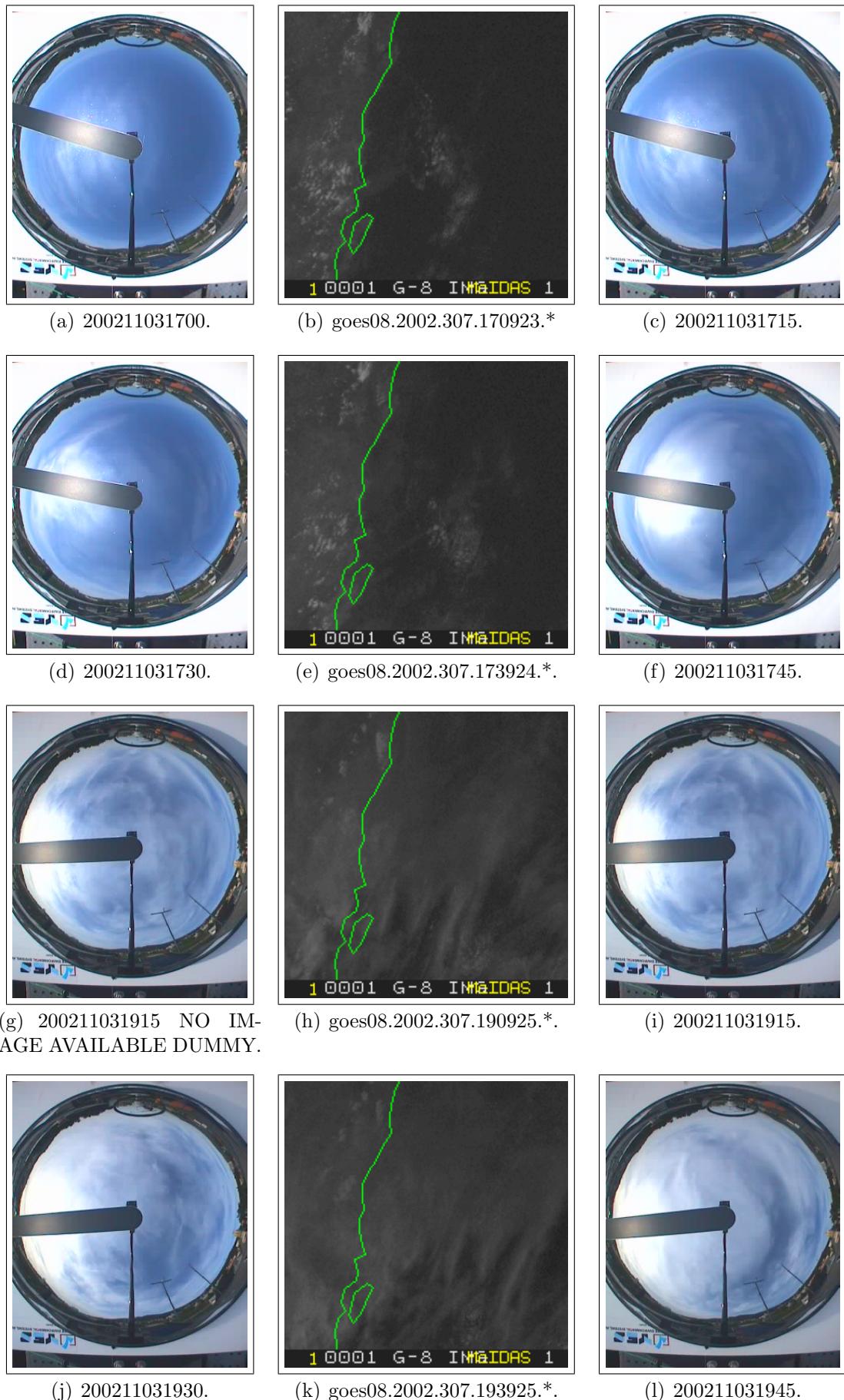


Figure 5.11: Imagens do satélite GOES 8 disponíveis entre 17:09 e 19:39 (b),(e),(h),(k), obtidas em um horário intermediário da imagens do imageador de nuvens antes (a),(d),(g),(j) e depois (c),(f),(i),(l).

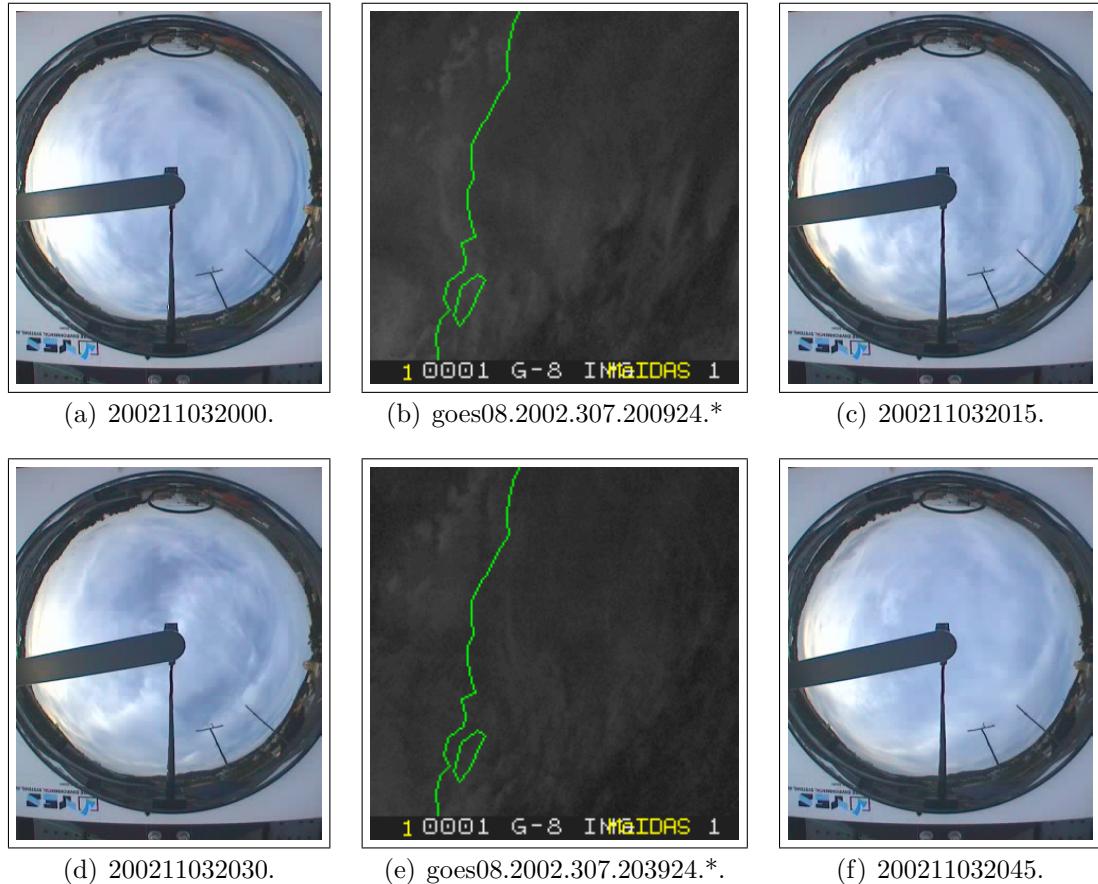


Figure 5.12: Imagens do satélite GOES 8 disponível entre 20:09 (b) e 2039 (e), obtida em um horário intermediário da imagens do imageador de nuvens antes (a), (d) e depois (c), (f).

Anexo C Equações utilizadas para o cálculo das efemérides solar

No presente anexo estão descritas as equações utilizadas para o cálculo das efemérides solar utilizada no presente trabalho.

- Dia angular obtido a partir de (IQBAL, 1983) equação (1.22).

$$\Gamma = \frac{2\pi(dn-1))}{365}$$

onde:

Γ : é o dia angular em radianos,
 dn : é o dia do ano em uma escala de 0 a 365.

- Fator de excentricidade obtido a partir de (IQBAL, 1983) equação (1.2.1) .

$$E_o = \frac{r}{r_0} = 1,000110 + 0.034221\cos(\Gamma) + 0.001280\sin(\Gamma) + 0,000719\cos(2\Gamma) + 0,000077\sin(2\Gamma)$$

onde:

E_o : é o fator de excentricidade da órbita terrestre para o dia da observação.
 r : é a distância entre a terra e o sol no instante de observação,
 r_0 : é a distância média entre a terra e o sol $1,496 \times 10^8$ km,
 Γ : é o dia angular em radianos.

- Ângulo de declinação solar obtido a partir de (IQBAL, 1983) equação (1.3.1)

$$\delta = 0,006918 - 0.399912\cos(\Gamma) + 0.070257\sin(\Gamma) - 0,006758\cos(2\Gamma) + 0,000907\sin(2\Gamma) - 0,002697\cos(3\Gamma) + 0.00148\sin(3\Gamma)$$

onde:

δ : é o ângulo de declinação solar para o dia da observação.
 Γ : é o dia angular em radianos.

- Equação do tempo obtido a partir de (IQBAL, 1983) eq. (1.4.1).

$$E_t = (0,000075 - 0,001868\cos(\Gamma) - 0,032077\sin(\Gamma) - 0,014615\cos(2\Gamma) - 0,04089\sin(2\Gamma))(228,18)$$

onde:

E_t : é a equação do tempo para o dia da observação.

Γ : é o dia angular em radianos.

- Correção da latitude local obtido a partir de (IQBAL, 1983) equação (1.4.2).

$$LAT = L_s + 4(L_s - L_e)$$

onde:

LAT : é o horário local aparente em minutos no instante de observação.

L_s : é o horário local em hora e décimos de hora no instante de observação.

L_e : é a correção da longitude do local de observação.

- Cosseno do ângulo zenital obtido a partir de (IQBAL, 1983) equação (1.5.1).

$$\cos\theta_z = \sin\delta\sin\phi + \cos\delta\cos\phi\cos\omega$$

onde:

θ_z : é o ângulo zenital em graus no instante de observação.

δ : é o ângulo de inclinação da terra em graus no dia da observação.

ϕ : é o ângulo da latitude em graus do local da observação, que para Florianópolis é de -27,5 graus.

ω : é o ângulo horário em graus no instante da observação.

- O ângulo de declinação da terra em graus foi obtido a partir de (IQBAL, 1983) equação (1.3.1).

$$\delta = (0,006918 - 0,399912\cos(\Gamma) + 0,070257\sin(\Gamma) - 0,006758\cos(2\Gamma) + 0,00097\sin(2\Gamma) - 0,002697\cos(3\Gamma) + 0,00148\sin(3\Gamma))(180/\pi)$$

- O ângulo horário ω pode ser obtido pela seguinte equação

$$\omega = (LAT - 12)15$$

- O horário solar ou horário local aparente (ou Local Apparent Time) foi obtido a partir de (IQBAL, 1983) equação (1.4.2).

$$LAT = LST + 4(L_s - L_e) + E_t$$

onde:

LAT : é o horário local aparente em graus e décimos de graus.

LST : é o horário local.

L_s : é a longitude padrão de Florianópolis que é de 45 graus.

L_e : é a longitude local de Florianópolis que é de 48,5 graus.

E_t : é a equação do tempo.