



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENERGIA NUCLEAR
COMISSÃO NACIONAL DE ENERGIA NUCLEAR
CENTRO REGIONAL DE CIÊNCIAS NUCLEARES DO NORDESTE
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS ENERGÉTICAS E
NUCLEARES

MANOEL HENRIQUES DE SÁ CAMPOS FILHO

**MODELAGEM E PREDIÇÃO DE FLUTUAÇÕES DA IRRADIAÇÃO SOLAR DE
CURTA DURAÇÃO**

Recife

2019

MANOEL HENRIQUES DE SÁ CAMPOS FILHO

**MODELAGEM E PREDIÇÃO DE FLUTUAÇÕES DA IRRADIAÇÃO SOLAR DE
CURTA DURAÇÃO**

Tese submetida ao Programa de Pós-Graduação
em Tecnologias Energéticas e Nucleares para
obtenção do título de Doutor em Ciências.

Área de Concentração: Fontes Renováveis de
Energia.

Orientador: Prof. Dr. Chigueru Tiba

Recife

2019

Catálogo na fonte
Bibliotecário Carlos Moura, CRB-4 / 1502

C198m Campos Filho, Manoel Henriques de Sá.
Modelagem e predição de flutuações da irradiação solar de curta duração. / Manoel Henriques de Sá Campos Filho. - Recife, 2019.
136 f. : il.

Orientador: Prof. Dr. Chigueru Tiba.

Tese (doutorado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Tecnologias Energéticas e Nucleares, 2019.
Inclui referências.

1. Engenharia de energia solar. 2. Modelos de predição da irradiação solar. 3. Modelagem da irradiação solar. 4. Fração de cobertura de nuvens. 5. Geração fotovoltaica. 6. Processamento de imagens do céu. I. Tiba, Chigueru, orientador. II. Título.

MANOEL HENRIQUES DE SÁ CAMPOS FILHO

MODELAGEM E PREDIÇÃO DE FLUTUAÇÕES DA IRRADIAÇÃO SOLAR DE CURTA DURAÇÃO

Tese apresentada ao Programa de Pós-Graduação em Tecnologias Energéticas e Nucleares da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciências.

Aprovada em: 18/07/2019.

BANCA EXAMINADORA

Prof^o. Dr. Chigueru Tiba (Orientador)
Universidade Federal de Pernambuco (DEN)

Prof^a. Dr^a. Olga de Castro Vilela (Examinadora Interna)
Universidade Federal de Pernambuco (DEN)

Prof^o. Dr. Carlos Alberto Brayner de Oliveira Lira (Examinador Interno)
Universidade Federal de Pernambuco (DEN)

Prof^o. Dr. José Bione de Melo Filho (Examinador Externo)
Universidade de Pernambuco - UPE / CHESF

Prof^o. Dr. Tsang Ing Ren (Examinador Externo)
Universidade Federal de Pernambuco (CIN)

AGRADECIMENTOS

Aos meus pais Manoel Henriques e Ana Maria que desde a tenra idade me incentivaram e apoiaram no estudo das Ciências em geral. Ao meu pai, particularmente, por me apresentar a beleza e a importância das áreas de Matemática, Física e Química.

À minha querida esposa Fabiana por todo o apoio, estímulo e afeto durante essa jornada.

Às minhas adoradas filhas Nathália e Amanda que, a seu modo, contribuíram e me encorajaram a enfrentar os desafios que se apresentaram.

Ao meu orientador, Prof. Chigueru Tiba, um agradecimento especial pela orientação, apoio, incentivo e solidariedade durante todas as etapas, especialmente diante de imprevistos ao longo deste trabalho.

Aos amigos do Laboratório de Energia Solar na área de testes, Rinaldo, Guilherme, Marcelo e ao colega de pós-graduação Fábio pelas inúmeras vezes que me auxiliaram ao curso dessa trajetória, notadamente no processo de instalação de hardware e durante a coleta de dados.

A todas as bancas de professores que avaliaram cada etapa deste trabalho, contribuindo para aperfeiçoar o estudo.

Ao professor Raimundo Jaildo dos Anjos, Coordenador INMET/3º DISME, pela inestimável colaboração com seu conhecimento profundo na Ciência da Meteorologia.

A todos os amigos e colegas que, direta ou indiretamente, contribuíram para realização deste trabalho.

À UFPE - Universidade Federal de Pernambuco e ao DEN – Departamento de Energia Nuclear pela oportunidade de realizar o doutorado.

À FACEPE - Fundação de Amparo à Ciência e Tecnologia de PE pela viabilização e incentivo através da bolsa concedida.

RESUMO

Atualmente, um dos maiores desafios na geração de energia solar é garantir que a energia gerada seja fornecida de forma segura e confiável. Em consequência, tornou-se essencial elaborar modelos para prever a incidência de radiação solar dentro de vários intervalos de tempo de interesse. O objetivo principal deste trabalho foi o desenvolvimento de uma metodologia composta por algoritmos e um conjunto de instrumentos que permitissem fazer previsões sobre a incidência de radiação solar dentro de curtos horizontes temporais. Para esta pesquisa, foram utilizadas imagens de uma câmera de lente *fisheye* de baixo custo e dados da estação solarimétrica. As imagens foram armazenadas juntamente com os dados de irradiância em um banco de dados. Durante a etapa de processamento de dados, cada fotografia do céu foi submetida a um tratamento digital para aprimoramento de contraste. As imagens resultantes foram, então, segmentadas com o auxílio de novos atributos matemáticos desenvolvidos a partir de espaços de cores conhecidos (RGB, HSV, YCrCb e Lab). Com esses atributos, construiu-se um modelo através de Máquina de Vetores de Suporte (SVM) para classificação, ou SVC, o que resultou em uma taxa de acerto na classificação de nuvens com precisão de 99,2% pelo método de 10 folders de validação cruzada. Em seguida calculou-se a fração de cobertura de nuvens. Os resultados obtidos por imagens foram usados como entrada para modelar a irradiância total e direta por regressão via SVR e validados pelo método de 10 folders. O coeficiente de correlação de Pearson atingiu 98% considerando toda a faixa de incursão dos sinais de irradiância dentro das 6236 amostras do período. Um modelo final com a inclusão das irradiâncias entre os dados de uma série temporal foi então desenvolvido por meio de SVR para predição das irradiâncias total e direta em oito horizontes temporais curtos (entre 1 e 30 minutos). Em síntese, os resultados demonstraram que com um equipamento simples e barato como a câmera *fisheye* foi possível: calcular a fração de cobertura de nuvens no céu; modelar a irradiância solar total e direta; estimar a velocidade (em pixels) das nuvens; e com adição de dados solarimétricos foram realizadas predições em curtos espaços de tempo com acurácia comparável a alguns trabalhos realizados dentro de uma metodologia similar no que se refere a coleta de dados e aos horizontes de predição espaço-temporal.

Palavras-chave: Modelos de predição da irradiação solar. Modelagem da irradiação solar. Fração de cobertura de nuvens. Geração fotovoltaica. Processamento de imagens do céu.

ABSTRACT

Nowadays, one of the biggest challenges in solar power generation is to ensure that the power generated is delivered safely and reliably. As a consequence, it has become essential to develop models to predict the incidence of solar radiation within time intervals of interest. The main objective of this work was the development of a methodology composed by algorithms and a set of instruments that allowed making predictions about the incidence of solar radiation within short time frames. In order to carry this research, images from a low-cost fisheye lens camera and data from the solarimetric station were used. The images were stored together with the irradiance data in a database. During the data processing stage, each image of the sky was submitted to a digital treatment for contrast enhancement. The resulting images were then segmented using new mathematical attributes developed from known color spaces (RGB, HSV, YCrCb and Lab). Based on these attributes, a model was built using the Support Vector Machine (SVM) for classification, or SVC, which resulted in a cloud classification accuracy rate of 99.2% by the method of 10 folders for cross-validation. The fraction of cloud cover was then calculated. The results obtained from the images were used as input to model the total and direct irradiance by regression via SVR and validated by the method of 10 folders. The Pearson correlation coefficient reached 98% regarding the entire incursion range of the irradiance signals within the 6236 samples of the period. A final model including irradiance data as part of a time series was then developed by using an SVR to predict total and direct irradiance in eight short time horizons (from 1 to 30 minutes). In summary, the results demonstrated that with a simple and inexpensive equipment such as the fisheye camera, it was possible to: calculate the fraction of cloud cover in the sky; model the total and direct solar irradiances; estimate the speed (in pixels) of the clouds; and with the addition of solarimetric data, predictions were made in a short period of time with comparable accuracy to other published works using similar methodologies with respect to dataset building and spatio-temporal prediction horizons.

Keywords: Prediction models for solar irradiation. Solar irradiation modeling. Fraction of cloud cover. Photovoltaic power generation. Sky image processing.

LISTA DE FIGURAS

Figura 1 – Evolução das instalações FV no mundo.....	19
Figura 2 – Irradiância solar medida na localidade de Coruripe, Alagoas em 01/01/2009	22
Figura 3 – Dias com diferentes características meteorológicas e de irradiâncias	23
Figura 4 – Boletim do ONS: a geração solar em abril de 2018.....	24
Figura 5 – Curva do pato ou <i>duck curve</i> exibindo desequilíbrio entre geração e carga	25
Figura 6 – Diagrama: despacho x horizontes de previsão x modelos	26
Figura 7 – Arranjo de 8 piranômetros em um semicírculo com $r=6m$	30
Figura 8 – Sinal de um par de sensores ref. à série temporal da irradiância medida.....	31
Figura 9 – Esquema de uma borda linear com triplos sensores.....	32
Figura 10 – Evolução da autocorrelação do índice de claridade k^*	35
Figura 11 – Exemplo LES: magnitude da vorticidade em um contorno longitudinal	36
Figura 12 – Exemplo de dispositivo <i>sky imager</i>	44
Figura 13 – Imagem de nuvens à esquerda e vetores de movimento à direita	45
Figura 14 – Horizontes temporais, espaciais, imagens e modelos de previsão.....	46
Figura 15 – Imagem com aplicação da razão vermelho para azul (RBR).....	47
Figura 16 – Etapas para segmentação da imagem segundo a razão vermelho/azul	50
Figura 17 – Imagem para critérios de decisão	50
Figura 18 – Exemplo da aplicação da saturação (do espaço HSV) a uma imagem	51
Figura 19 – Câmera <i>fisheye</i> Vivotek modelo FE8174V	54
Figura 20 – Imagem obtida pela câmera no teto do laboratório	55
Figura 21 – Método simplificado da aplicação do método RBR para detecção de nuvens	55
Figura 22 – Ofuscamento, saturação, "flare" e outros efeitos em imagens	55
Figura 23 – Estação solarimétrica do grupo FAE.....	57
Figura 24 – Pireliômetro sNIP	58
Figura 25 – Piranômetro CMP-21	58

Figura 26 – Espectro de frequências eletromagnéticas e cores associadas	58
Figura 27 – Datalogger CR3000.....	59
Figura 28 – Fluxograma de entrada de dados no computador do laboratório	60
Figura 29 – Programa desenvolvido para captura de imagens	61
Figura 30 – Mapeamento de velocidades por cores (à esquerda) e vetores (à direita).....	62
Figura 31 – Fluxo ótico esparsa aplicado ao monitoramento de veículos.....	63
Figura 32 – Deslocamento radial de um ponto na nuvem e sua imagem em forma de pixel ...	64
Figura 33 – Velocidade radial de um ponto na nuvem e sua imagem em forma de pixel.....	66
Figura 34 – Representação da imagem do deslocamento tangencial de um ponto na nuvem..	68
Figura 35 – Máscara para aplicação do algoritmo FAST	73
Figura 36 – Campo de visão da câmera olho de peixe ou <i>fisheye</i>	74
Figura 37 – Câmera <i>Pinhole</i>	75
Figura 38 – Representação geométrica em 3D do modelo adotado	76
Figura 39 – Imagem sem distorção (a), com distorção tipo almofada (b) e barril (c).....	78
Figura 40 – SVM linear com dados linearmente separáveis	81
Figura 41 – Mapeamento do espaço de entrada no espaço de características	81
Figura 42 – Fluxograma para criação do espaço vetorial de segmentação	85
Figura 43 – Máscara utilizada para segmentação.....	86
Figura 44 – A radiação direta e sua projeção sobre a vertical local. “A” é o azimute	87
Figura 45 – Ilustração gráfica do método de Otsu.....	90
Figura 46 – Imagem usada para calibrar e extrair parâmetros intrínsecos da câmera.....	95
Figura 47 – Calibragem da imagem <i>fisheye</i> à direita, original à esquerda	96
Figura 48 – Determinação da posição e da trajetória solar ao longo de um dia	97
Figura 49 – Aplicação do algoritmo CLAHE à imagem à esquerda e resultado à direita.....	98
Figura 50 – Método usado para classificar pixels como nuvem ou céu	98
Figura 51 – Histograma do atributo <i>f0</i>	100
Figura 52 – Histograma do atributo <i>f1</i>	100

Figura 53 – Histograma do atributo <i>f2</i>	100
Figura 54 – Histograma do atributo <i>f3</i>	101
Figura 55 – Histograma do atributo <i>f4</i>	101
Figura 56 – Histograma do atributo <i>f5</i>	101
Figura 57 – Histograma do atributo <i>f6</i>	102
Figura 58 – Imagem segmentada (em tons de laranja) à direita	102
Figura 59 – Mapeamento do deslocamento de pixels após algoritmo Gunnar Farneback	103
Figura 60 – Algoritmo FAST para encontrar pontos coincidentes em imagens seguidas	105
Figura 61 – Vetor velocidade média à esquerda e imagem segmentada à direita	106
Figura 62 – Nebulosidade e trajetória do Sol no dia 18/04/2019 às 14h00min	107
Figura 63 – Sequência de fotografias do hemisfério no dia 18/04 em ordem cronológica	108
Figura 64 – Imagens e gráficos durante a modelagem das irradiâncias total e direta	111
Figura 65 – Gráficos das irradiâncias total (vermelho) e direta (verde) para o dia 18/04	113
Figura 66 – Gráficos das irradiâncias total (vermelho) e direta (verde) para o dia 19/04	113
Figura 67 – Gráficos das irradiâncias total (vermelho) e direta (verde) para o dia 22/04	114
Figura 68 – Gráficos das irradiâncias total (vermelho) e direta (verde) para o dia 23/04	114
Figura 69 – Gráficos das irradiâncias total (vermelho) e direta (verde) para o dia 21/05	115
Figura 70 – Gráfico RSME(GHI) prev. x Horizonte de previsão	116
Figura 71 – Gráfico RSME(DNI) prev. x Horizonte de previsão	117
Figura 72 – Modelos (DNI): persistência e novo método (linhas pontilhadas/cheias)	118
Figura 73 – Gráfico de previsão para o dia 05/06 (horizonte de 5 minutos)	119
Figura 74 – Gráfico do modelo SVR de previsão para o horizonte de 1 minuto	119
Figura 75 – Gráfico do modelo SVR de previsão (DNI) para o horizonte de 5 minutos	120
Figura 76 – Método setor em escada ajustável para previsão da DNI	121
Figura 77 – Gráfico da variação da DNI relativo ao da figura anterior	121
Figura 78 – Dias com pequena variabilidade: (a) céu claro e (b) nublado	122
Figura 79 – Dias com pequena variabilidade: céu claro	123

Figura 80 – Dias com pequena variabilidade: nublado	124
Figura 81 – Gráfico do modelo SVR de predição (GHI) para o horizonte de 1 minuto	124
Figura 82 – Gráfico do modelo SVR de predição (GHI) para o horizonte de 5 minutos.....	125

LISTA DE TABELAS

Tabela 1 – Classificação dos 10 maiores países em produção de energia solar elétrica	18
Tabela 2 – Exemplo de correlação entre pares de sensores.....	31
Tabela 3 – Atenuação das velocidades radiais e tangenciais de pixels	70
Tabela 4 – Resultados das métricas aplicadas a cada atributo	99
Tabela 5 – Resultados dos modelos para as irradiâncias total e direta.....	109
Tabela 6 – Resultados dos modelos de predição para a irradiância total (GHI)	115
Tabela 7 – Resultados dos modelos de predição para a irradiância direta (DNI).....	116

LISTA DE ABREVIATURAS E SIGLAS

ANEEL	Agência Nacional de Energia Elétrica
ANSI	Acrônimo para <i>American National Standards Institute</i>
AR	Modelo de auto regressão
ARIMA	Modelo de auto regressão integrado e de médias móveis
ARX	Modelo de auto regressão com entradas exógenas
BEM	Desvio médio, também denominado viés
C++/C#	Linguagens de programação C “mais mais” e C Sharp
CNN	Rede neural convolucional
DBN	Acrônimo para <i>Deep Belief Network</i> , classe de redes neurais profundas (DL)
DEN	Departamento de Energia Nuclear da UFPE
DHI	Irradiância solar difusa incidente em uma superfície horizontal
DL	Rede neural genérica do tipo <i>deep learning</i>
DNI	Irradiância solar incidente diretamente em uma superfície normal aos raios solares
DRNN	Rede neural do tipo <i>deep recurrent neural network</i>
FAE	Grupo de estudos do DEN voltado a pesquisas sobre fontes alternativas de energia
FFT	<i>Fast Fourier Transform</i> ou transformada rápida de Fourier, algoritmo eficiente para cálculo da transformada discreta de Fourier
FOV	Acrônimo para <i>Field of Vision</i> ou campo de visão
FV	Fotovoltaico(a)(s)
GHI	Global Horizontal Irradiance, ou irradiância total. Soma da irradiância solar difusa (DHI) com a componente normal da irradiância direta (DNI) do hemisfério incidentes sobre uma superfície horizontal
HD	Disco rígido de computador, representa uma mídia magnética
HSV	Espaço de cores com representação em coordenadas cilíndricas.
IEA PVPS	Agência Internacional de Energia – Programa de Sistemas Energéticos Fotovoltaicos
KNN	Rede neural dos K vizinhos mais próximos (ou K-Nearest Neighbors)
Lab	Espaço de cores em que L é a intensidade, “a” é uma componente de cor na faixa entre verde e vermelho e “b” é outra componente na faixa entre azul e amarelo
LDA	Linear Discriminant Analysis, algoritmo de classificação relacionado ao aprendizado de máquina
LES	Large Eddy Simulation, modelo matemático para turbulência usado em dinâmica computacional de fluidos
LIBSVM	Biblioteca para uso do algoritmo SVM de aprendizado de máquina
LIDAR	Tecnologia óptica de detecção remota e acrônimo para <i>Light Detection And Ranging</i>

MAE	Métrica: desvio médio absoluto
MBE	Métrica: desvio médio ou viés
ML	Abreviatura para <i>Machine Learning</i> ou aprendizado de máquina
MLP	Rede neural do tipo <i>Perceptron</i> multi camadas
MLP-BP	Rede neural <i>Perceptron</i> multi camadas com retropropagação
MSE	Métrica: desvio médio quadrático
ONS	Operador Nacional do Sistema Elétrico
OpenCV	Biblioteca de código aberto de Visão Computacional
P&B	Preto e branco, usado para descrever imagens em tons de cinza
PC	Computador pessoal
PCA	<i>Principal Component Analysis</i> ou análise de componentes principais
PSPI	Acrônimo em alusão ao modelo inteligente de persistência baseado na Física
RGB	Modelo de cores aditivas em que as cores vermelho, verde e azul são combinadas para formarem uma grande quantidade de cores compostas.
RMSE	Métrica: raiz quadrada do desvio médio quadrático (MSE)
RNA	Rede neural artificial, máquina de aprendizado também representada pela sigla ANN
SMO	Sequential Minimal Optimization, aperfeiçoamento voltado às SVM's
SPS	<i>Superpixel Segmentation</i> , técnica largamente utilizada para segmentação de imagens
SVM	<i>Support Vector Machine</i> ou Máquina de vetores de suporte, algoritmo de aprendizado de máquina (ML) supervisionado
SVC	Máquina de vetores de suporte voltada para problemas de classificação
SVR	Máquina de vetores de suporte voltada para problemas de regressão
WRNN	Rede neural recorrente associada a <i>wavelet</i>
WSI	refere-se ao dispositivo do tipo <i>Sky imager</i> completo
TCP/IP	Conjunto de protocolos de comunicação entre computadores em rede
TSI	Refere-se ao dispositivo do tipo <i>Sky imager</i> total
YCrCb	Espaço de cores em que Y é a luminância e Cr assim como Cb são componentes de cromaticidade.

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Contextualização	16
1.2	Caracterização do problema	21
1.3	Objetivos e metas.....	27
2	REVISÃO DE LITERATURA	28
2.1	Deslocamento de nuvens e redes de sensores terrestres discretos	28
2.2	Modelos baseados em redes neurais, estatísticas e outras abordagens	34
2.2.1	Satélites e <i>sky imagers</i>	42
2.3	Pixels, espaços de cores e segmentação	46
3	MATERIAL E MÉTODOS.....	53
3.1	Câmera tipo <i>fisheye</i>	53
3.2	Estação solarimétrica e <i>datalogger</i>	56
3.3	Captura de imagens e coleção de dados de irradiâncias	59
3.4	Fluxo ótico e velocidade das nuvens	61
3.4.1	Fluxo ótico denso	61
3.4.2	Fluxo ótico esparsos	62
3.5	Calibragem da câmera <i>fisheye</i>	73
3.6	Determinação da posição do Sol	78
3.7	Máquina de aprendizado supervisionado SVM	79
3.8	Segmentação de nuvens	82
3.9	Modelo para estimar as irradiâncias a partir das imagens.....	86
3.10	Modelo para predições de irradiâncias	91
4	RESULTADOS E DISCUSSÃO	94
4.1	Calibragem da câmera <i>fisheye</i>	94
4.2	Determinação da posição do Sol	96
4.3	Segmentação de nuvens	97

4.4	Fluxo ótico e velocidade das nuvens	103
4.4.1	Fluxo ótico denso	103
4.4.2	Fluxo ótico esparso	104
4.5	Modelo para estimar as irradiâncias a partir das imagens.....	106
4.6	Modelo para predições em curtos intervalos de tempo	112
5	CONCLUSÕES E CONSIDERAÇÕES FINAIS	127
	REFERÊNCIAS	130

1 INTRODUÇÃO

1.1 Contextualização

A energia solar é uma realidade técnica e econômica no Brasil e no mundo. Em todo o planeta há mais de 500GW instalados que continuam crescendo em taxas elevadas, especialmente em alguns países. Além disso, existe um forte apelo ecológico à sua expansão devido ao fato de ser menos agressiva ao meio ambiente, em sintonia com a época de estudos crescentes sobre impactos ambientais, aquecimento global e aumento da frequência de fenômenos extremos. Isso não significa que tanto a fabricação quanto o descarte de painéis solares, inversores e demais elementos usados na geração não produzam resíduos e materiais poluentes, mas possuem menor dano à natureza quando se compara à geração de energia a partir de outros recursos, como o de combustíveis fósseis ainda largamente utilizados. Dentro dessa conjuntura, a sociedade cobra cada vez mais vigorosamente aos seus respectivos governos por mais empenho e esforços por acordos internacionais para redução das emissões de CO₂ na atmosfera, o que coloca novamente a energia solar em um patamar de destaque como excelente alternativa de solução.

Qualquer mudança na matriz energética de um país está intimamente relacionada e dependente dos investimentos tecnológicos e econômicos necessários. Matérias publicadas na mídia impressa ou digital, oriundas de várias fontes (empresas, instituições governamentais, não governamentais e especialistas), preveem em um futuro relativamente próximo um quadro de rentabilidade igual ou superior de produção de energia solar sobre outras fontes. A revista econômica Forbes, por exemplo, através de estudo fundamentado em levantamentos feitos na Alemanha, noticiou que a energia solar poderá se tornar a fonte de energia mais barata por volta de 2025 em muitas regiões (GUERRINI, 2015). Esse tipo de sinalização ao mercado de investidores contribui para impulsionar ainda mais toda a cadeia produtiva que movimenta o mercado de energia solar, o que realimenta todo o processo e aumenta o otimismo financeiro nesse setor.

É relevante complementar a discussão com uma breve visão da expansão da energia solar nos âmbitos nacional e internacional com o respaldo de informações e números baseados em estudos e relatórios. No âmbito nacional é importante observar as necessidades existentes, características peculiares de sua introdução em larga escala e, dessa forma, como a energia solar tem se apresentado como uma das respostas aos problemas enfrentados. Assim, ao mesmo

tempo em que há um número crescente de centrais de energia solar no exterior influenciado pelo barateamento dos custos de implantação, no Brasil há reincidentes momentos de apreensão na geração elétrica. No cerne dessa questão estão o aumento da demanda por energia elétrica e a intermitência dos recursos hídricos sujeitos ao regime de chuvas. Isso em um país cuja geração elétrica tem natureza predominantemente hidráulica. Por conseguinte, o preço da energia tem experimentado sucessivas elevações que representam um obstáculo ao crescimento nacional com sérios reflexos no setor produtivo através do aumento de custos na indústria e comércio.

No cerne entre produção e demanda de energia elétrica é importante avaliar a evolução do uso de energia solar no Brasil na medida em que a ANEEL (ANEEL, 2018) tem divulgado sua expansão. Por exemplo, no leilão de geração “A4” cuja fonte de informações relata ter sido em 04/04/2018, houve um deságio de 59,07% em relação aos preços-teto estabelecidos, representando uma economia de R\$ 9,73 bilhões para os consumidores de energia. O certame movimentou ao todo R\$ 6,74 bilhões em contratos, equivalentes a um montante de 54.094.749,6/MWh de energia. Foram registrados os menores preços já negociados para usinas eólicas e solares. Uma análise desses resultados por fonte apresentou para energia solar 29 empreendimentos com preço médio final de R\$ 118,07/MWh, representando a contratação de 40.060.620MWh de energia. Com a construção de usinas nos estados do Ceará (14), Piauí (6), Minas Gerais (6) e Pernambuco (3), o total de investimento será de R\$ 4,2 bilhões e 1.032,54 megawatts-pico (MWp) contratados. A potência que será injetada na rede após a conversão de corrente contínua (energia solar) em corrente alternada (energia elétrica para a rede de distribuição) será de 806,64 MW.

Se forem comparados dados dos primeiros leilões em relação a esse realizado em 2018 verifica-se que o preço médio do MWh solar aproximou-se de 1/3 dos valores iniciais a grosso modo, sem calcular dados e índices econômicos como a inflação do período. Porém, deve-se observar que a produção em megawatts-pico estimada do referido leilão de 2018 apresentou valor similar a dos primeiros certames. Esse contexto reflete claramente o aumento de competitividade da energia solar na medida em que aproximadamente a mesma quantidade de MWp experimentou uma redução significativa.

Há ainda o tempo gasto para que novas fontes geradoras de energia entrem em funcionamento, sendo um dado importante que merece ser observado. Dois exemplos servem bem ao propósito de salientar a questão temporal relacionada à operação efetiva desses empreendimentos. A usina de Itaipu, 2^a maior do mundo em geração elétrica, possui a capacidade de 14 GW e a maior do mundo, a Hidrelétrica de Três Gargantas na China, tem a capacidade de 22,5 GW. Então, em apenas 2014 (IEA, 2015), foram instalados em energia solar

o equivalente a mais do que a soma dessas duas maiores usinas reunidas. Outra informação a ser destacada é que Itaipu levou quase 10 anos para começar a operar desde o início dos contratos, enquanto a Hidrelétrica de Três Gargantas consumiu 13 anos quando entraram em operação apenas as primeiras turbinas.

O panorama de crescimento de energia elétrica fotovoltaica nacional, apesar dos saltos após todos esses projetos, ainda está longe de ser comparado com a capacidade instalada e acumulada em alguns países, como a Índia, China, Japão e Estados Unidos. Esses países ostentam a marca de dezenas de giga watts instalados acumulados de energia solar. Contudo, a retrospectiva mostra que já houve ocasião em que o Brasil passou a ingressar no *ranking* mundial de potência solar instalada. O fato ocorreu em 2017, conforme análise datada de 06/2018 e apresentada de forma segmentada por Sauaia (SAUAIA, 2019). A Tabela 1 a seguir contém a classificação mais atual relativa aos 10 países mais bem colocados em termos de potência instalada e acumulada até 2018.

Tabela 1 – Classificação dos 10 maiores países em produção de energia solar elétrica

Lugar (posição)	Capacidade instalada em 2018		Capacidade acumulada até 2018	
	País	Potência (GW)	País	Potência (GW)
1º	China	45,0	China	176,1
2º	Índia	10,8	Estados Unidos	62,2
3º	Estados Unidos	10,6	Japão	56,0
4º	Japão	6,5	Alemanha	45,4
5º	Austrália	3,8	Índia	32,9
6º	Alemanha	3,0	Itália	20,1
7º	México	2,7	Reino Unido	13,0
8º	Coréia do Sul	2,0	Austrália	11,3
9º	Turquia	1,6	França	9,0
10º	Holanda	1,3	Coréia do Sul	7,9

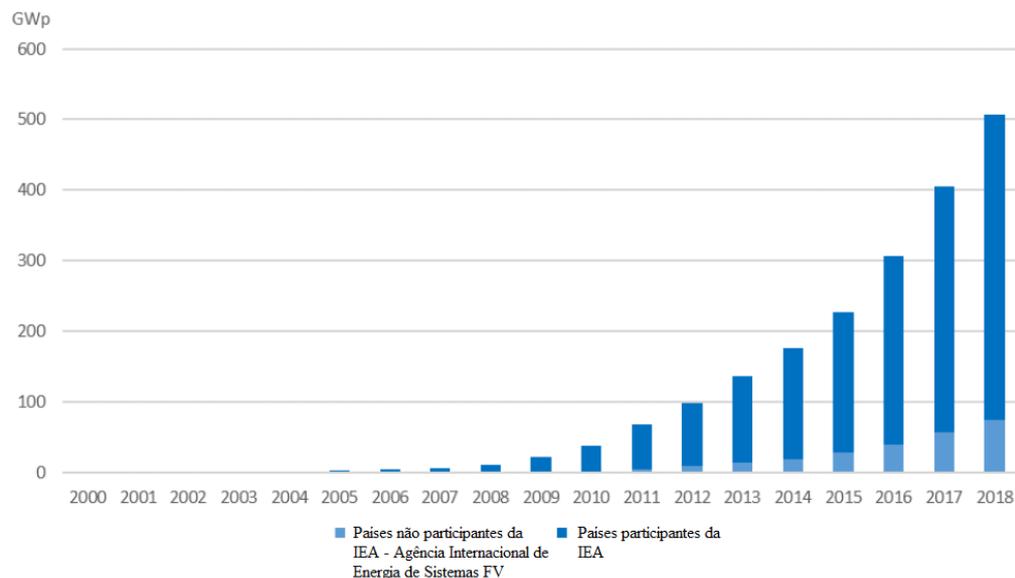
Fonte: (IEA, 2019)

De acordo com os dados mais recentes levantados pela IEA – *International Energy Agency PVPS*, da qual o Brasil faz parte como membro, alguns aspectos da evolução do

emprego de sistemas FV também merecem destaque tendo como base o período mais recente, o ano de 2018 (IEA, 2019):

- A capacidade total de energia FV instalada no mundo cruzou a barreira de 500GW, ou seja, $\frac{1}{2}$ TW (IEA, 2019);
- O mercado global FV para o ano de 2018 foi de, pelo menos, 97,9GW, mas estima-se que possa ter sido algo em torno de 99,9GW;
- A China, líder mundial em capacidade instalada de cerca de 176,1GW, sofreu uma retração em 2018 de 53GW para 45GW instalados;
- A Índia progrediu significativamente com um crescimento de 10,8GW, tornando-se o segundo maior mercado;
- O mercado nos Estados Unidos decresceu ligeiramente para 10,6GW;
- A União Europeia instalou 8,3GW e o restante da Europa algo em torno de 1,1GW. O maior mercado europeu foi a Alemanha (3GW);
- O Japão estabeleceu-se na quarta posição com 6,5GW instalados em 2018;
- Outros mercados cresceram espetacularmente, como a Austrália com algo próximo a 3,8GW, México com 2,7GW e Coreia do Sul com 2GW;
- Ainda de acordo com o IEA, a evolução das instalações FV para o ano de 2018 pode ser vista através do gráfico da Figura 1.

Figura 1 – Evolução das instalações FV no mundo



Fonte: (IEA, 2019)

A entrada inicialmente gradual do Brasil no contexto fotovoltaico revela-se atualmente como intensa com perspectivas de grandes somas de investimentos que vem ocorrendo a cada leilão. Enfim, um quadro com forte estímulo, o que remete às análises mais atuais da instituição IEA-PVPS em que alguns países experimentaram recentemente um crescimento “espetacular” de mercado. Até mesmo aquelas aplicações em menor escala são incentivadas, pois o contexto indica um forte despertar no cenário nacional. É preciso obviamente ressaltar a importância das políticas públicas para alavancar esse setor a exemplo do que vem acontecendo em outros países.

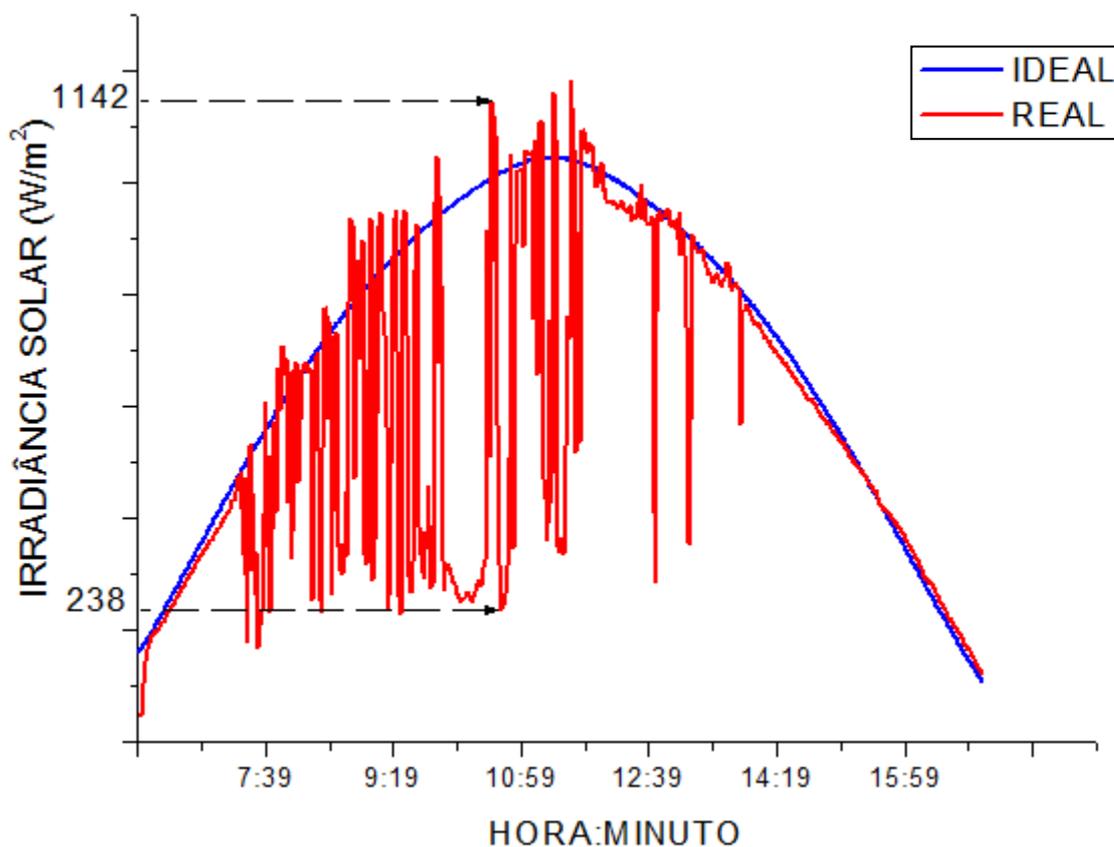
Do ponto de vista tecnológico ainda há muito por fazer. Por exemplo, a eficiência na conversão de energia solar em elétrica vem melhorando progressivamente ao longo dos últimos tempos. Em outro aspecto, a aplicação em grande escala do uso desse tipo de geração elétrica vem acompanhada por novos desafios tecnológicos, de modo que sempre está surgindo a necessidade de novas pesquisas para solucioná-los. E do ponto de vista da aplicação local, tão importante quanto mostrar fatos e alguns números que corroborem a percepção desse momento de mudança nas matrizes energéticas é o entendimento de que existem alguns obstáculos e suas soluções estão abertas às novas ideias e propostas inovadoras. Também merece ser enfatizada a importância da nacionalização, o quanto possível, de soluções tecnológicas por se tratarem de conquistas estratégicas do setor energético e pela possibilidade de reduzir o custo de aquisição. Nesses termos, existem alguns desafios que ocupam posições chave na produção de energia solar. Um deles está na interrupção noturna na produção de energia. Mas a produção diurna também encontra dificuldades semelhantes quando há diminuição significativa das irradiâncias total e direta.

1.2 Caracterização do problema

A quantidade de energia produzida em uma planta solar sofre variações ao longo do dia, ano e localidade. Essa flutuação tem duas componentes básicas: uma determinística e outra de natureza aleatória ou estocástica. A determinística está associada ao movimento astronômico Terra-Sol de acordo com a geometria bem estabelecida que descreve o movimento dos corpos celestes. É uma visão que pode ser simplificada e refletir o ponto de vista do movimento relativo do Sol diante do referencial de um observador na Terra (RABL, 1985). A contribuição estocástica se refere às alterações na atmosfera que podem ter como causa uma série de fatores como aqueles relacionados ao clima da região. Considerando as medições ao nível da superfície terrestre, essa variabilidade aleatória deu origem a uma definição encontrada na literatura pelo termo “taxa de rampa” (HARPER, 2010). Numericamente é calculada como a razão entre as grandezas físicas: (1) diferença de potência (por unidade de área) e (2) intervalo de tempo de medição da rampa. Aplicada à irradiância solar assume normalmente as dimensões $W/(m^2 \cdot min)$, significando a variação da irradiância solar (potência por unidade de área) em um determinado período de tempo.

Mais precisamente, as nuvens são a maior causa de variabilidade na produção FV (PELLAND, GALANIS e KALL, 2013). O intervalo de tempo durante a passagem de uma nuvem sobre um sistema de captação de energia solar é influenciado por diversos elementos, dentre os quais sua altura, tamanho e forma bem como pela velocidade da nuvem. Essa passagem pode produzir quedas na geração de energia da ordem de 60% em poucos segundos (MILLS e WISER, 2010). O sombreamento de um sistema de captação de energia solar, em um intervalo menor que 10 minutos, produz flutuações que são diretamente absorvidas pelo sistema elétrico, mas acarreta variações de frequência (RAMALHO, 2015). Porém, se essas variações ultrapassarem certos limiares, pode ocorrer o risco de falha no sistema de energia (MARCOS *et al.*, 2014). Para representar visualmente esse comportamento aleatório, o gráfico da Figura 2 apresenta uma comparação entre o valor ideal da irradiância na localidade de Coruripe (Alagoas) para um dia perfeitamente claro e o valor efetivamente medido ao longo desse dia. A observação da irradiância nesse exemplo demonstra o quanto seus valores podem se afastar das medidas ideais assumindo características erráticas, distorcidas e comumente chamadas de transientes da irradiância solar.

Figura 2 – Irradiância solar medida na localidade de Coruripe, Alagoas em 01/01/2009

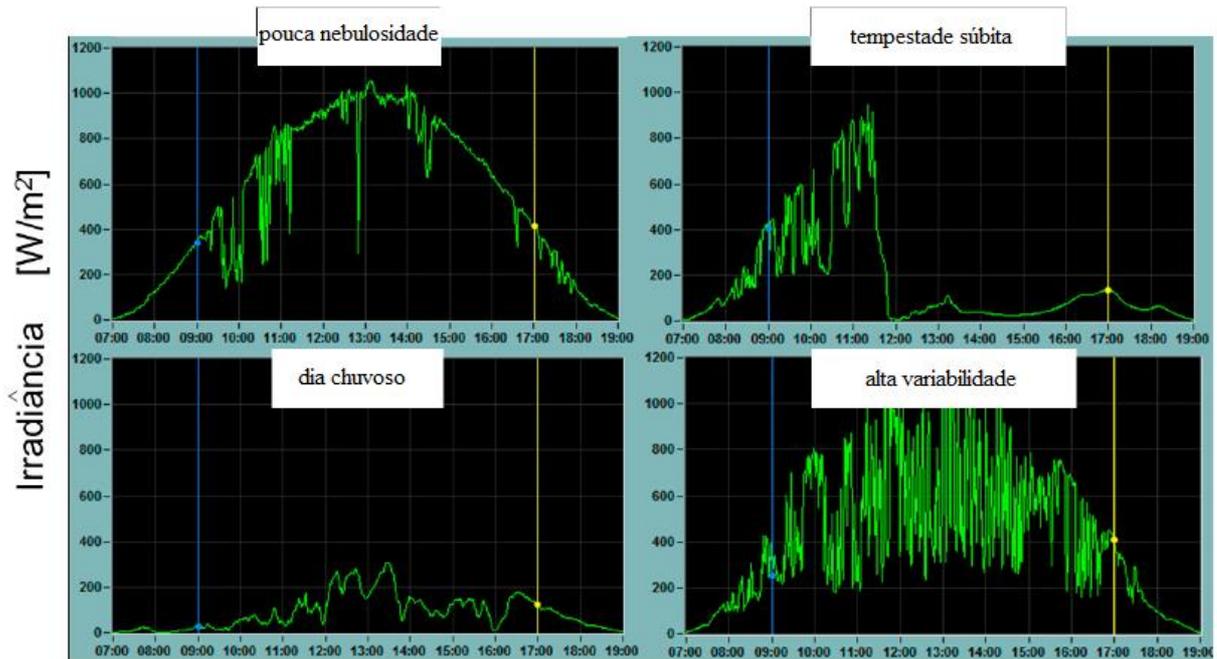


Fonte: (RAMALHO, 2015)

Também é possível perceber que o avanço da nebulosidade sobre uma instalação solar pode resultar em uma queda abrupta de produção de energia de um ponto máximo a valores ínfimos. Por exemplo, no gráfico anterior (Figura 2) observa-se em um curto intervalo de tempo a redução da irradiação solar de 1142 para apenas 238 W/m^2 , o que representa uma perda de potência aproximada de 80%.

A variabilidade encontra-se associada diretamente às características meteorológicas do dia em estudo, de modo que os gráficos das irradiâncias assumem contornos e aspectos peculiares da aleatoriedade que marcou o período. O trabalho de Reindl (REINDL *et al.*, 2017), para exemplificar as mudanças atmosféricas que ocorrem na zona de convergência intertropical de Cingapura onde houve o experimento, incluiu o gráfico representado pela Figura 3 para demonstrar a intensidade e variabilidade da irradiância total (GHI) em quatro dias contendo diferentes situações: com pouca nebulosidade, com uma tempestade súbita, em um dia chuvoso e em um dia com alta variabilidade.

Figura 3 – Dias com diferentes características meteorológicas e de irradiâncias

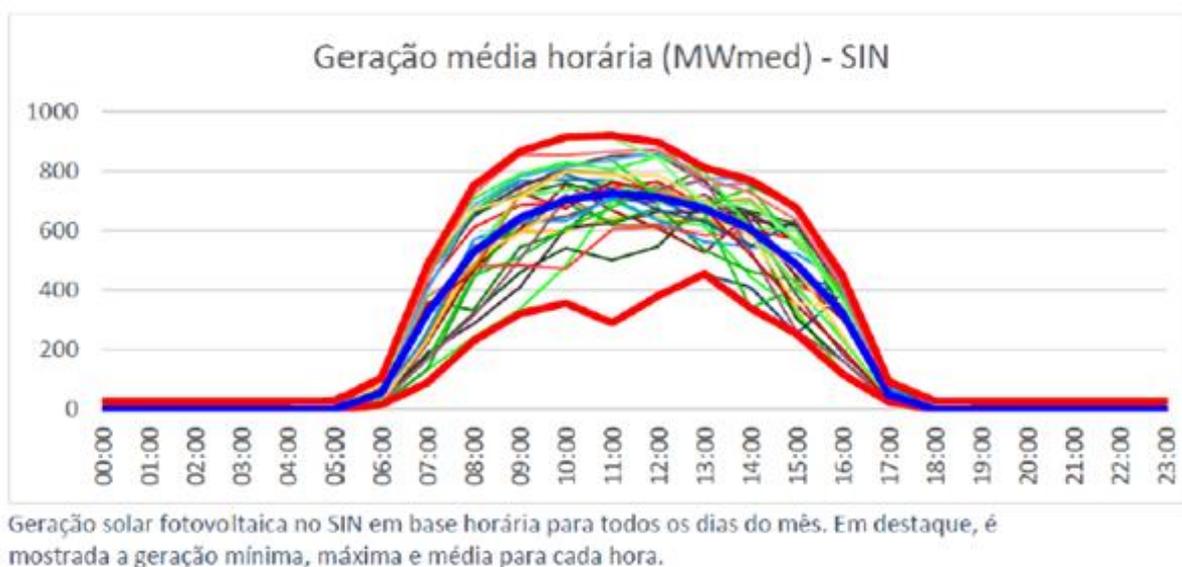


Fonte: Reindl (REINDL *et al.*, 2017)

Através de boletins obtidos do Operador Nacional do Sistema Elétrico – ONS torna-se evidente o impacto da variância da irradiação solar sobre plantas destinadas à geração elétrica FV. Em particular, o boletim relativo à geração solar do mês de abril de 2018 contém um gráfico, Figura 4, que não deixa dúvidas sobre quanto essa intermitência já afeta atualmente a geração dentro de uma escala de centenas de MW (em média). Em vermelho estão representadas (acima e abaixo) as curvas que exibem respectivamente os valores máximo e mínimo daquele período e destacada em azul em uma faixa intermediária a curva que representa o valor médio. As demais cores representam outros dias daquele período.

Figura 4 – Boletim do ONS: a geração solar em abril de 2018

A Geração Solar em Abril de 2018



Fonte: boletim mensal do ONS de geração FV para 04/2018

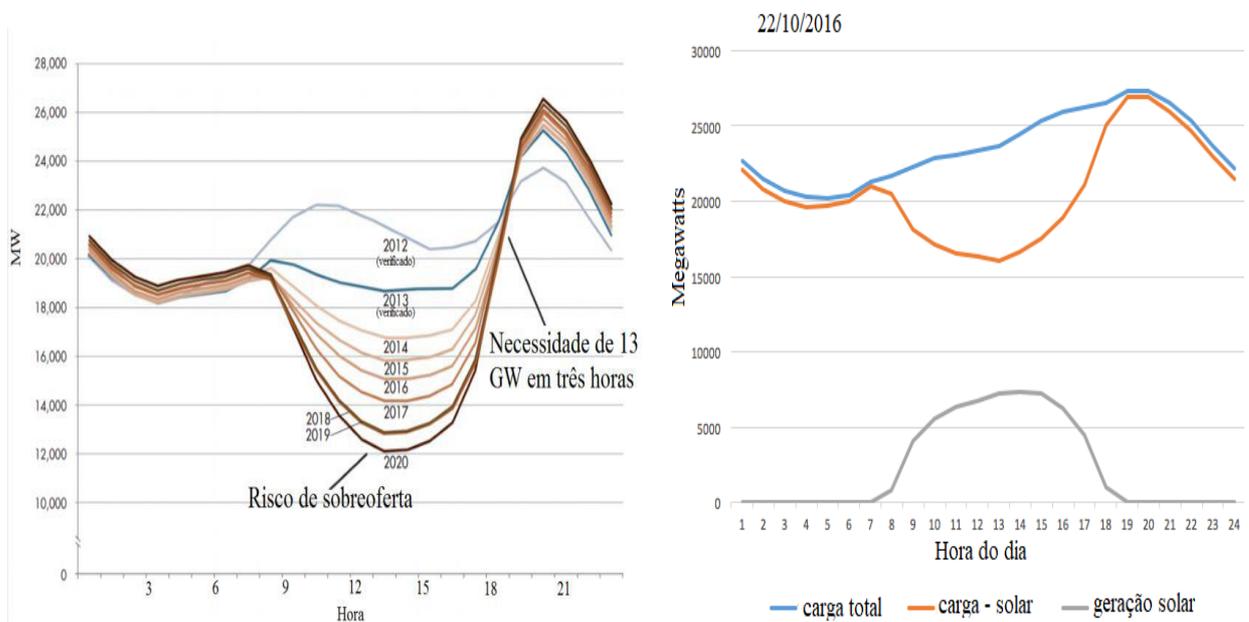
Portanto, uma vez identificado e conhecido o impacto causado pelas nuvens sobre a potência de saída de um sistema fotovoltaico ou qualquer sistema de captação de energia solar, torna-se necessário o desenvolvimento de alternativas que possam mitigar essas perturbações. Considerando a produção em larga escala em plantas de produção de energia solar, talvez esse problema envolva um dos maiores desafios da indústria fotovoltaica na atualidade: entregar energia segura e confiável. Historicamente, sabe-se que a demanda por eletricidade é largamente não controlada (WEST *et al.*, 2014). Por outro lado, com a rápida inserção da energia FV nas malhas da rede elétrica, faz-se necessário o ajuste da geração de acordo com essa demanda. Para mitigar ou gerenciar essas flutuações ou intermitências da irradiância, normalmente os operadores de sistemas de energia elétrica recorrem aos sistemas de reserva para minimizar os efeitos da variabilidade de potência, o que vem estar associado ao conceito de “despachabilidade” das reservas de potência. Essas reservas são definidas como a capacidade de potência que pode ser usada dentro de um determinado prazo com o propósito de auxiliar na geração, de efetuar balanceamentos de carga e no controle de frequência. É chamada de “reserva de regulação” para um intervalo de variabilidade entre segundos a vários minutos, sendo usada para manter a frequência e tensão do sistema dentro da tolerância permitida durante os períodos

de flutuações (SHEDD *et al.*, 2012). Dessa forma o conceito de despachabilidade encontra-se diretamente vinculado aos seguintes aspectos:

- Descasamento entre geração e demanda (*duck curve* – Figura 5);
- Existência de picos/redução de potência seja na geração ou na carga;
- Regulação de frequência e tensão do grid;
- Chaveamento de fontes energéticas não intermitentes;
- Horizontes de despachabilidade: horas, minutos ou segundos;
- Armazenamento de energia para compensação.

A Figura 5 apresenta a curva do pato (ou *duck curve*) com dados obtidos da empresa California ISO (ou *California Independent System Operator.*) em que é fácil constatar um desequilíbrio constante e crescente, ano após ano, entre a carga total e a geração solar, com risco de sobreoferta cada vez maior em determinado intervalo horário. Outros pontos que merecem atenção são aqueles que se encontram nas laterais da região de sobreoferta quando a geração solar encontra-se nula ou em seus patamares mínimos e a requerer fontes auxiliares para compensação visando a manutenção das exigências da carga. Em ambas as situações a necessidade de despachabilidade de potência torna-se clara.

Figura 5 – Curva do pato ou *duck curve* exibindo desequilíbrio entre geração e carga

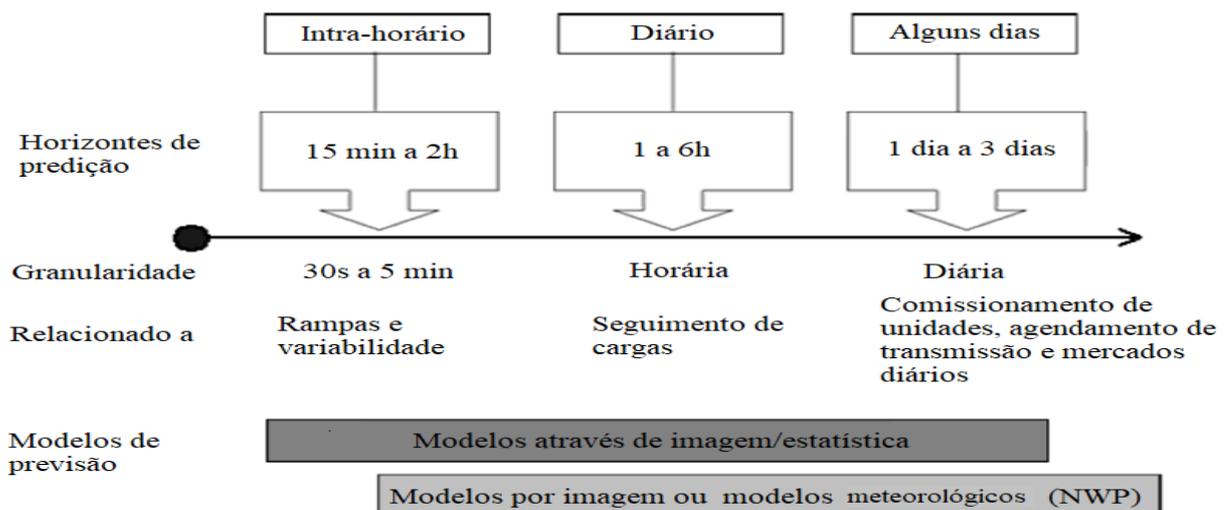


Fonte: *California Independent System Operator*

A capacidade de previsão do recurso solar surge então nesse contexto como uma ferramenta extremamente importante para prover otimização no uso de sistemas de reservas para uma melhor integração à rede elétrica, possibilitando a tomada de ações e emprego de medidas corretivas em tempo hábil. Em particular, um sistema de predição em um horizonte curto de tempo, objeto deste trabalho de tese, encontra aplicação não apenas no uso de sistemas de reserva, mas também no planejamento operacional e em chaveamentos de fontes de energia, além do gerenciamento do sistema durante picos de consumo (WU, CHEN e RAHMAN, 2014). Além do aspecto técnico, vale salientar, em termos de viabilidade de uma planta solar, que a redução dos custos de implantação, dos custos operacionais e de integração depende de poder antecipar-se às variações de potência da geração elétrica solar (PEDRO e COIMBRA, 2012; YANG *et al.*, 2014; DAGNE *et al.*, 2013).

A Figura 6, adaptada a partir do trabalho de Diagne (DIAGNE *et al.*, 2013), apresenta um diagrama que procura sintetizar todo o problema da variabilidade na geração de energia solar FV. Esse mecanismo contém as inter-conexões desde as rampas até a ponta final representada pela minimização do problema através dos modelos de predição em coordenação com os demais elementos para mitigar as consequências indesejáveis. Encontra-se segmentado pelos horizontes de predição/granularidade e inclui o elemento indispensável que diz respeito à despachabilidade: seguimento de cargas, comissionamento de unidades, agendamento de transmissão e mercados diários. Por fim, observa como parte integrante da solução os modelos baseados em imagem/estatística para curtos espaços de tempo, foco deste trabalho.

Figura 6 – Diagrama: despacho x horizontes de predição x modelos



Fonte: (DIAGNE *et al.*, 2013)

1.3 Objetivos e metas

O propósito deste trabalho foi desenvolver um sistema de modelagem e predição de flutuações de curta duração na irradiância solar.

Para atingir esse objetivo, foram previstas as seguintes metas:

- Seleção de modelo para calibragem e levantamento de parâmetros intrínsecos da câmera escolhida (cuja lente é do tipo olho de peixe ou *fisheye*);
- Seleção de modelo para determinação das posições solares ao longo dos dias;
- Segmentação de imagens por ML: identificação de pixels como nuvem ou céu;
- Modelagem da fração de cobertura do céu e do Sol;
- Modelagem das irradiâncias (total e direta) por imagens;
- Desenvolvimento de sistema de predição para as irradiâncias total e direta.

2 REVISÃO DE LITERATURA

Existem muitos modelos de previsão da irradiação solar. Em sua maioria, esses modelos compreendem tanto abordagens estatísticas, quanto abordagens baseadas em predição numérica meteorológica (ou NWP). Modelos estatísticos se caracterizam pela análise de dados históricos. Entre eles, estão modelos através de séries temporais, os com base em dados de satélites, outros através de imagens do céu registradas da superfície da Terra, modelos com base em redes neurais artificiais, modelos através de transformadas *wavelet*, etc. Modelos fundamentados em dados meteorológicos, por sua vez, não apenas baseiam-se em dados históricos, mas também na reprodução de informações físicas (DIAGNE *et al.*, 2013). Combinações dos modelos anteriores também podem ser encontradas em diversos artigos, formando modelos híbridos (MARQUEZ, PEDRO e COIMBRA, 2013), que surgem pela necessidade de superar deficiências encontradas em modelos individuais. Essa combinação entre diferentes métodos pode levar a uma maior precisão em predições, ao mesmo tempo em que agrega características de linearidade, não linearidade ou ambas (DIAGNE *et al.*, 2013). Pela importância que os métodos estatísticos apresentam, será um assunto explorado mais à frente em um tópico específico. As imagens de satélites e câmeras também têm se mostrado extremamente valiosas em vários estudos que envolvem modelos estatísticos, de modo que esse tipo de abordagem particular merece atenção por meio de outro tópico.

Por fim, ainda há métodos preditivos de características determinísticas que foram encontrados na literatura e que não poderiam deixar de ser mencionados. Destinam-se a estimar a ocorrência (ou não) de flutuações a partir da velocidade (na forma de vetor sobre o solo) de passagem das sombras de nuvens sobre sensores. Dessa forma, serão alvo de comentários logo a seguir.

2.1 Deslocamento de nuvens e redes de sensores terrestres discretos

A identificação da direção e da velocidade das nuvens merece atenção especial neste trabalho. As pesquisas bibliográficas revelaram uma variedade de artigos com métodos desenvolvidos para essa finalidade. Fazem uso de diferentes recursos matemáticos, estatísticos e computacionais, além de diferentes tecnologias de sensoriamento.

A velocidade das nuvens tem sido a principal entrada em modelos para estudos de predição e variabilidade em instalações FV, considerando horizontes curtos de tempo (CHOW *et al.*, 2011). Existe, por exemplo, uma estimativa simples para a taxa de rampa máxima possível em uma planta solar através da expressão (Eq. 1) (BOSCH, ZHENG e KLEISSL, 2013):

$$\frac{dP}{dt} = P_{csk} K_t v / l \quad \text{Eq. 1}$$

Em que P_{csk} é a saída de potência em céu claro, v é a velocidade escalar da (sombra da) nuvem (ou módulo de seu vetor velocidade) e l o comprimento da planta na direção de deslocamento da nuvem. K_t , conhecido como índice de claridade ou coeficiente de transmitância, é definido como a razão entre a irradiação global que atinge a superfície terrestre e a irradiação que incide no topo da atmosfera. Dessa forma, a taxa de rampa é função de quanto tempo leva uma nuvem para cobrir a planta, assumindo que a nuvem seja maior que a planta.

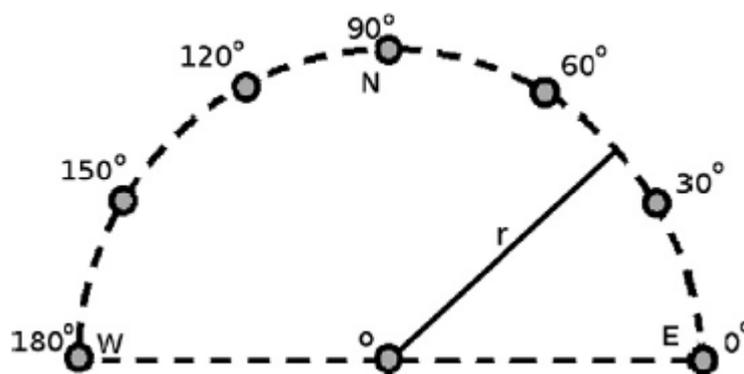
Para a obtenção do vetor de deslocamento/velocidade de uma nuvem (ou então de sua sombra sobre o solo) várias técnicas podem ser empregadas. Frequentemente são usadas técnicas por imagens de satélites (HAMMER *et al.*, 1999). Essas imagens podem detectar padrões de nuvens tanto no espectro visível quanto no infravermelho. Assumindo que as características visuais das nuvens não mudem significativamente sobre um curto intervalo, esse vetor pode ser calculado a partir da sequência de sucessivas imagens. Não linearidades no deslocamento de nuvens bem como os fenômenos de formação e dissipação fazem com que modelos de predição numérica superem aqueles baseados em vetores de movimento em horizontes de tempo mais extensos. No entanto, não apenas dados provenientes de imagens de satélite são complexos para adquirir e processar, mas a baixa atualização das imagens (a cada 15-30 min.) e o atraso de transferência desses dados podem não permitir a detecção de variações importantes, como aquelas originadas pelas correntes de ar convectivas.

Métodos variados têm sido propostos ou usados para obter a velocidade das nuvens a partir de sensores baseados no solo. Em um desses experimentos, Hammer e Stolzenburg empregaram uma rede de sensores ao longo de algumas edificações e analisaram a relação cruzada de sinais para obter a velocidade e direção das nuvens (HAMMER e STOLZENBURG, 1993). Porém, um grande número de sensores foi utilizado e os resultados da relação cruzada precisaram ser analisados manualmente, pois o método de automatização desenvolvido não foi bem-sucedido. De modo similar, Hinkelman e colaboradores determinaram a velocidade das

nuvens analisando o atraso na máxima relação cruzada entre sensores alinhados na direção da nuvem, mas esse método também não pôde ser automatizado desde que não podia ser conhecida a direção da nuvem a priori (HINKELMAN *et al.*, 2011). Baldwin e Collins propuseram a construção de uma rede com dois círculos concêntricos de sensores em volta de uma planta FV para detecção prévia e medição da velocidade de nuvens, mas nenhum algoritmo para essa medição foi incluído (BALDWIN e COLLINS, 2011).

Em outro estudo (BOSCH, ZHENG e KLEISSL, 2013), dois métodos foram apresentados por meio de sensores para medição da irradiância global no plano horizontal. O primeiro método, dito como aquele do par mais correlacionado, é o que seleciona os dois sensores com a maior relação cruzada na série temporal da irradiância. Conforme se pode visualizar pela rede de sensoriamento na Figura 7, a estrutura geométrica adotada tem a forma de um semicírculo. Então oito sensores foram agrupados em sete pares, todos compartilhando o sensor central. A direção do vento será então identificada pelo par mais correlacionado. O número de sensores está relacionado à resolução, no caso 30° para os sete pares. Se houvesse mais sensores, maior seria a resolução. Esse aumento de resolução na direção de movimento da nuvem também se traduz em maior precisão em sua velocidade. Considerando essa estrutura, os autores quantificaram que um erro de $\pm 30^\circ/2$ (equivalente a $\pm 15^\circ$) na direção resultará em um erro máximo de $(1-\cos 15^\circ)$, aproximadamente 3,4%, aceitável dentro da proposta do artigo.

Figura 7 – Arranjo de 8 piranômetros em um semicírculo com $r=6m$



Fonte: (BOSCH, ZHENG e KLEISSL, 2013)

A Figura 8 é usada para visualizar os sinais representados pelos sensores S_a e S_b que representam um dos pares acima, estando separados por uma distância D . É assumida a premissa de que estão fisicamente alinhados com o movimento da nuvem. As irradiâncias de

cada sensor, I_a e I_b , estão fortemente correlacionadas, mas com um atraso de tempo t_{ab} . Uma vez que o atraso tenha sido determinado pela análise da relação cruzada, a velocidade da nuvem pode ser calculada pela expressão:

$$v = D/t_{ab} \quad \text{Eq. 2}$$

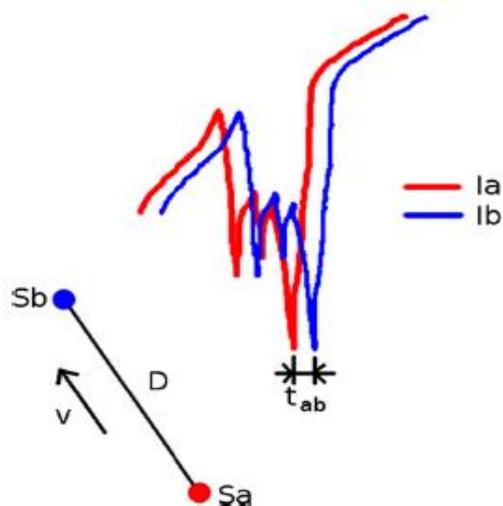
A Tabela 2, a seguir, exemplifica a seleção do melhor par correlacionado diante dos possíveis ângulos. Observa-se que houve uma melhor correlação para o par de número 5.

Tabela 2 – Exemplo de correlação entre pares de sensores

Par	Ângulo θ	Correlação	(t_{ab}) (s)	V(m/s)
1	0°	0,714	1,65	3,60
2	30°	0,7062	1,70	3,50
3	60°	0,7433	1,35	4,40
4	90°	0,8123	0,80	7,50
5	120°	0,8483	1,05	5,70
6	150°	0,8268	0,85	7,10
7	180°	0,7629	0,30	20,00

Fonte: (BOSCH, ZHENG e KLEISSL, 2013)

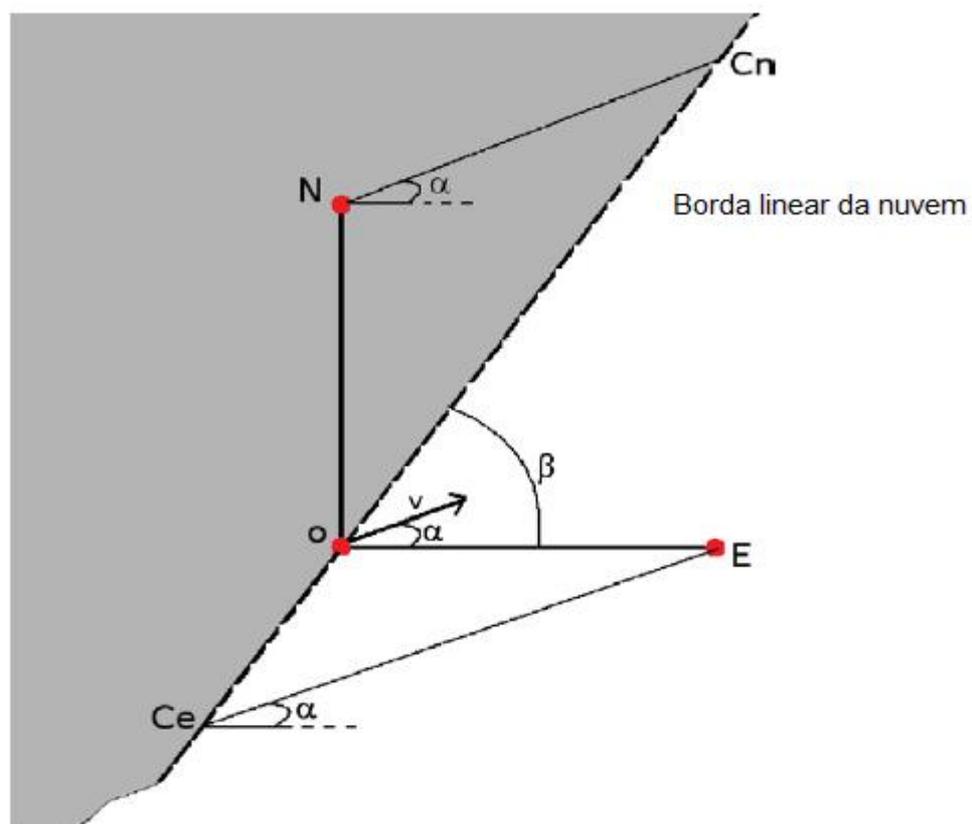
Figura 8 – Sinal de um par de sensores ref. à série temporal da irradiância medida



Fonte: (BOSCH, ZHENG e KLEISSL, 2013)

Um segundo método alternativo é proposto através de um sensoriamento triplo. Vetores de movimento são calculados a partir da borda linear de sombreamento passando pelo arranjo. O conceito de passagem de uma borda de sombreamento linear, que é mais intuitivo imaginar, é, contudo, uma implementação baseada em máximos e mínimos da série temporal da irradiância. Esse outro método usa os sensores rotulados como **O**, **N** e **E**, onde as distâncias de **N** e **E** para **O** são iguais ao raio **r** (6m). É assumido (i) um vetor de velocidade constante da nuvem, (ii) uma borda linear de sombreamento e (iii) que a nuvem é grande o suficiente para cobrir os três sensores. A condição (i) é aceita pela proximidade dos sensores, enquanto (ii) e (iii) são geralmente satisfeitas, levando-se em conta as dimensões da nuvem de $\sim 10^2 - 10^3$ m comparada ao arranjo. De acordo com o esquema apresentado na Figura 9, “ t_{on} ” é o tempo necessário para a nuvem se mover da posição **N** para **Cn** e “ t_{oe} ” o tempo para se mover de **Ce** para **E**. A direção do movimento é dada pelo ângulo α com relação à horizontal (eixo x). O ângulo da borda linear de sombreamento com a horizontal (eixo x) é representado por β . O vetor de movimento que segue a direção dada pelo segmento \overline{NCn} ou \overline{CeE} , é representado pelo vetor **v**.

Figura 9 – Esquema de uma borda linear com triplos sensores.



Fonte: (BOSCH, ZHENG e KLEISSL, 2013)

Através da cinemática, se obtém a expressão para o cálculo do vetor velocidade, onde D é o valor do raio r e v o módulo do vetor velocidade:

$$t_{on} \sin \alpha + t_{oe} \cos \alpha = D/v \quad \text{Eq. 3}$$

Desde que “ t_{on} ” e “ t_{oe} ” podem ser obtidos das medições, restam duas incógnitas, α e v . A informação necessária para resolver a Equação 3 pode ser obtida por uma segunda borda de nuvem passando pelos sensores, com os mesmos α e v , mas com um β (que representa o ângulo da borda com o eixo horizontal x) diferente. Atribuindo índices diferentes para as nuvens 1 e 2, resulta que:

$$t_{on1} \sin \alpha + t_{oe1} \cos \alpha = D/v \quad \text{Eq. 4}$$

$$t_{on2} \sin \alpha + t_{oe2} \cos \alpha = D/v \quad \text{Eq. 5}$$

O cálculo de α pelas Equações (4) e (5) traz o resultado:

$$\alpha = \arctan\left(-\frac{(t_{oe2}-t_{oe1})}{(t_{on2}-t_{on1})}\right) \quad \text{Eq. 6}$$

A partir dessa última expressão obtém-se a velocidade escalar das nuvens pela substituição de α nas equações anteriores.

A primeira abordagem, por obtenção da máxima correlação, foi considerada mais robusta, de modo que representaria “a verdade sobre o solo”. Contudo, o segundo método é viável diante de eventuais restrições de orçamento, quando o custo de instalação e manutenção de um grande arranjo de piranômetros for proibitivo. Os resultados relativos às direções das nuvens apresentados em ambas os métodos foram consistentes entre si e com a análise visual das imagens. As velocidades calculadas também foram similares e consistentes com dados provenientes de sondas atmosféricas. Assim como acontece com os experimentos de medição

via dados de satélite, a precisão desses métodos é difícil de assegurar devido à falta de dispositivos de alta qualidade como referência.

2.2 Modelos baseados em redes neurais, estatísticas e outras abordagens

Há um leque de possibilidades no que se refere à escolha de modelos que usam uma abordagem estatística direta ou indireta. Então foram selecionadas para a apresentação as principais abordagens que têm conquistado evidência em trabalhos de pesquisa. Serão mencionadas sem seguir um critério de ordenação cronológica de desenvolvimento ou categorização quanto à relevância científica.

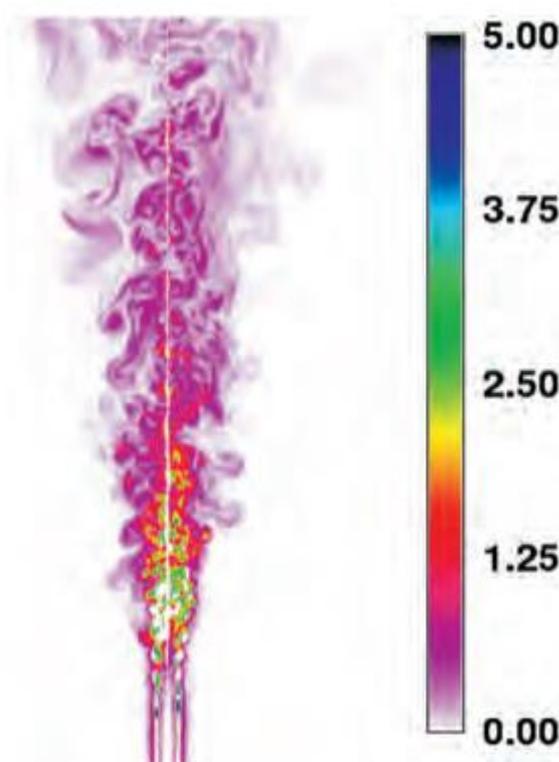
Os modelos de predição de irradiância solar através de séries temporais utilizam medições recentes de irradiância como uma entrada básica complementada por variáveis correlatas (TROCCOLI, DUBUS e HAUPT, 2014). A dependência funcional entre valores prévios e valores previstos é aprendida por meio de algoritmos estatísticos em uma fase de treinamento sobre dados históricos, assumindo que padrões nos conjuntos de dados históricos se repitam no futuro e assim possam ser explorados para o objetivo de predição. Esses modelos tiram vantagem da alta correlação em espaços de tempo curtos em séries temporais de irradiância e nebulosidade. A Figura 10 mostra os coeficientes de autocorrelação do índice de céu claro (k^*), que é definido como a razão entre a irradiância global medida e a irradiância em condições de céu claro. Esse índice pode ser considerado como a medida inversa da nebulosidade. A autocorrelação é alta para um atraso de poucas horas, mas decresce rapidamente com o aumento do atraso, limitando a efetividade da abordagem da série temporal para horizontes além de algumas horas. Para horizontes mais largos é essencial levar em conta fenômenos dinâmicos como movimento e formação (ou dissipação) de nuvens, o que torna o modelamento físico necessário, de forma que não pode ser determinado com base apenas nas medições locais.

Contudo, qualquer modelo tem uma incerteza associada independentemente do horizonte de predição. Isso é causado, por exemplo, pelos limites da resolução espacial e temporal, por incertezas sobre os parâmetros de entrada, além das simplificações assumidas dentro do modelo. Portanto, para uma escala de tempo de uma ou duas horas à frente, previsões baseadas em medições precisas *in loco* são vantajosas. Para previsões de curto tempo, seja intra-horária ou ainda com a antecedência de 2 ou 3 horas, as técnicas de aprendizado estocástico podem ser muito competitivas e relativamente simples de configurar (PELLAND *et al.*, 2013).

As abordagens por séries temporais também são feitas por métodos de regressão clássicos como o modelo simples autorregressivo (AR) e o modelo autorregressivo integrado de média móvel (ARIMA).

Com forte fundamento em análises estatísticas, o trabalho de Jamaly (JAMALY e KLEISSL, 2017) aplicou o método denominado espaço-temporal Krigging aplicado à predição de irradiâncias. Os dados tiveram como origem o assim chamado LES – Large Eddy Simulation, modelo matemático para turbulência usado em dinâmica computacional de fluidos (Figura 11). Mostrou-se, por meio de simulações de irradiâncias, com acurácia (RMSE normalizado) superior a 11,33% e 13,93% para duas variações desse método denominadas, respectivamente, IST (isotrópica espaço-temporal) e SP (espacial). Uma terceira variação, porém, AST (anisotrópica espaço-temporal), apresentou-se como não confiável diante de longos períodos ou de medições de baixa acurácia da velocidade de nuvens.

Figura 11 – Exemplo LES: magnitude da vorticidade em um contorno longitudinal



Fonte: (MATHEW, 2010)

Outros métodos numéricos que extraem características preditivas também podem ser integrados ao estudo, tais como redes neurais artificiais (RNA) e algoritmos genéticos. Para

qualquer modelo de série temporal, a seleção e a disponibilidade de variáveis apropriadas de entrada tanto quanto o pré-processamento para otimizar (e normalizar) esses dados são de importância relevante para um resultado preditivo acurado.

No trabalho de Durrani e colegas (DURRANI, *et al.*, 2018), a proposta de desenvolvimento teve como objetivo a construção de algoritmos preditivos para gerenciamento eficiente de energia e monitoramento do grid residencial integrado à rede de distribuição de energia. Esse sistema é baseado em um modelo de previsão de irradiância associado a um modelo fotovoltaico. Para tanto usaram múltiplas redes neurais totalizando sete redes do tipo *feed-forward* com três camadas: camada de entrada, camada escondida e camada de saída dos dados. As variações captadas pelos *grids* residenciais desempenharam importante papel na entrada de dados para o modelo, considerando a disponibilidade do local onde foi realizado o trabalho, Stuttgart - Alemanha, privilegiado pela quantidade de residências a utilizarem seus geradores de energia FV ligados ao grid de distribuição. Ao final do trabalho os autores conseguiram alcançar uma taxa para o erro médio absoluto de 3,4% para a irradiância total global, GHI, em dias ensolarados e 23% em dias nublados. Afirmaram ainda que esse modelo baseado em previsão de irradiância através de redes neurais obteve desempenho muito melhor que o modelo de previsão baseado em persistência.

Há um trabalho que ressalta a importância acerca do uso de redes neurais artificiais sobre dados de natureza estatística de energia solar obtidos no nordeste do Brasil (LIMA, *et al.*, 2016). Nessa pesquisa os autores demonstraram com dados obtidos no período entre 2009 a 2011, a capacidade de ajuste e refinamento dos modelos pelo uso dessa ferramenta de previsão. Encontram-se nesse trabalho tabelas comparativas pelo uso de algumas métricas – coeficiente de correlação, MSE e RMSE – que apresentam sistemática melhora de *performance* pelo emprego da RNA. Nesse estudo foi utilizada a RNA *multilayer perceptron*, ou MLP, com três camadas - entrada, escondida ou intermediária e saída - e a função do tipo sigmoide selecionada foi a tangente hiperbólica para a camada intermediária. Ainda segundo os autores, a RNA do tipo MLP é a arquitetura mais utilizada em pesquisas meteorológicas.

Reikard (REIKARD, 2009) desenvolveu um estudo comparativo entre métodos de séries temporais para previsão da radiação solar em alta resolução. Os experimentos computacionais usam seis conjuntos de dados de entrada de irradiância global horizontal, com resolução temporal de 5, 15, 30 e 60 minutos. Os dados exibem variações não lineares devido às variações atmosféricas como a nebulosidade. Essas séries incluem três bases de dados mantidas pelo Laboratório Nacional de Energias Renováveis em Golden, Colorado, compreendendo 35.040 observações, 8.760 por ano em 4 anos. Foram avaliadas várias

abordagens: uma através de regressão por logaritmos, outra pelo uso do método autorregressivo integrado de média móvel (ARIMA), outra pelo método denominado por componentes não observados tomando o exemplo do que fez Harvey (HARVEY, 1989), outra avaliação através de redes neurais, além de métodos por funções de transferência e modelos híbridos como uma associação entre o método de regressão e aquele por RNA. Todos os modelos foram estimados considerando-se sua história anterior ao início do horizonte de predição. Depois os valores de predição foram calculados e comparados aos valores medidos. Em quase todos os casos os melhores resultados foram obtidos pelo método ARIMA com coeficientes variantes no tempo. As exceções aconteceram diante de algumas características particulares. Em resoluções consideradas altas, a função de transferência usada foi superior ao ARIMA em um cenário de nebulosidade. Em resoluções mais altas, da ordem de 5 min., alguns casos usando redes neurais ou modelos híbridos apresentaram melhores desempenhos. O autor do artigo reporta uma dificuldade inicial de conciliar os resultados favoráveis das redes neurais que então se apresentaram superiores ao método ARIMA apenas no caso de altas resoluções, em aparente oposição a outros estudos que se mostraram mais favoráveis às RNA's de forma mais abrangente. Paralelamente, também foram relatadas dificuldades colocadas como substanciais no treinamento das redes, exigindo milhões de interações. Além disso, em experimentos de predição sobre longos intervalos, o número de interações precisou ser limitado a um dado número de épocas e o resultado é que algumas instâncias convergiram para um mínimo local, o que pode ter contribuído para um desempenho menor em relação a outros métodos dentro das mesmas circunstâncias de análise. Uma vez que existem várias alternativas de uso no campo de redes neurais, esse trabalho não discutiu eventuais testes ou análises teóricas sobre quais seriam as possíveis ou mais apropriadas redes para confrontar o problema em questão, não revelando se houve um projeto criterioso nesse sentido. Na verdade, não foi mencionado nem mesmo qual foi o tipo de rede utilizada, nem as suas características de implementação, treinamento e validação.

Outro estudo realizou uma comparação entre abordagens preditivas (PEDRO e COIMBRA, 2012). Foram usados dados de entrada não exógenos para predição de uma planta solar FV de 1 MWp, com rastreamento em um eixo, operando em Merced, Califórnia. Essa usina solar fornecia algo em torno de 17 a 20% do consumo anual da Universidade da Califórnia. Os dados usados nesse trabalho correspondem à média horária da potência no intervalo entre novembro de 2009 e agosto de 2011. Dados anteriores a janeiro de 2011 são usados para o treinamento dos modelos para predição da potência média horária entre uma e duas horas de antecipação. Os modelos usados foram: modelo persistente, ARIMA, modelo dos

k vizinhos mais próximos, modelo de rede neural artificial e modelo de rede neural modificado por algoritmo genético. A precisão dos modelos é avaliada através da determinação de parâmetros estatísticos, tais como o erro absoluto médio, o desvio das médias e o coeficiente de correlação (r^2) entre os valores previstos e medidos dentro do período de janeiro a agosto de 2011. Os resultados mostraram que os modelos de predição baseados em redes neurais apresentaram melhor desempenho que as outras técnicas, que aperfeiçoamentos substanciais podem ser alcançados por meio de otimização de parâmetros da rede com algoritmos genéticos e que a precisão de todos os modelos depende fortemente da sazonalidade característica da variabilidade solar.

O exemplo citado a seguir explora o fato de que modelos estatísticos podem incluir não apenas dados medidos no local em estudo, mas também dados adicionais exógenos, ou seja, obtidos fora da planta solar, tais como aqueles provenientes de modelos numéricos de predição meteorológica ou oriundos de informações (preditivas) com base em dados de satélites. A combinação adotada no estudo de Bacher e colaboradores (BACHER, MADSEN e NIELSEN, 2009) combinando modelos de predição físicos e estatísticos reafirma essa premissa bastante vantajosa, possibilitando não só estender o horizonte de predição como ainda aumentar significativamente a precisão dos dados previstos. Nesses termos, o método desenvolvido foi voltado para predições em uma janela de tempo de 36 horas. Os dados foram extraídos de 21 painéis FV montados sobre um teto em uma pequena cidade da Dinamarca. O método sugerido tem dois estágios. Primeiramente é feita, a partir da potência de saída, uma normalização estatística por meio de um modelo de céu claro. O modelo para céu claro é encontrado usando técnicas estatísticas de suavização de curvas. Então são realizadas predições da potência de saída normalizada usando modelos adaptativos de séries temporais. Em seguida, modelos autorregressivos (AR) e AR com entrada exógena (ARX) são avaliados, onde o último usa um modelo de predição meteorológica como entrada. Os resultados indicam que dentro de uma janela preditiva de 2 horas, a entrada mais importante é aquela relativa às observações da potência de saída do conjunto FV, enquanto para horizontes mais longos a entrada significativa é a referente aos dados do modelo meteorológico. Uma melhora de 35% referente ao parâmetro RMSE foi obtida pelo modelo ARX em comparação com o modelo de referência.

As transformadas *wavelet* também têm sido utilizadas para predição dos recursos solares, normalmente associadas a outros métodos numéricos de análise. Um exemplo dessa abordagem está no trabalho de Lave e colaboradores no desenvolvimento de um modelo para simular plantas solares FV, dada a série temporal da irradiância de um sensor e usando correlações espaço-temporais (LAVE, KLEISSL e STEIN, 2013). É simulada a redução da

variabilidade que ocorre em expandir em escala a saída de um sensor em um único ponto até uma planta solar FV, combinada com a transformada *wavelet* da série temporal desse sensor, produzindo então uma saída simulada de potência de toda a planta. O método foi validado através dos dados obtidos em uma planta FV de 2 MW em Ota City, Japão. O resultado das simulações concordou bem com a potência de saída da planta diante das variabilidades. Comparado a outros modelos de simulação, esse método também foi satisfatório.

Em outro trabalho, as transformadas *wavelet* foram utilizadas em conjunto com redes neurais para o mesmo propósito de predição da radiação solar dentro de uma metodologia chamada de redes neurais recorrentes *wavelet* (ou WRNN) (CAPIZZI, NAPOLI e BONANNO, 2012). Esse estudo lida com o problema de predição da radiação solar com base em dados meteorológicos observados. Um intervalo de predição de 2 dias foi obtido usando essa metodologia (WRNN). Segundo os autores, a característica inovadora dessa abordagem é que a técnica WRNN obtém a predição no domínio da *wavelet* e adicionalmente, através da transformada inversa, fornece o sinal previsto como saída. Os resultados obtidos pelas simulações apresentaram um desvio RMSE muito baixo quando comparado a outros modelos obtidos por redes neurais híbridas.

Pelo estudo anterior, é natural a concordância sobre a união vantajosa dessas técnicas: enquanto *wavelets* têm a propriedade de localização em termos de tempo e frequência, além de características focais, as redes neurais aliam as qualidades de adaptabilidade, tolerância a falhas, robustez e a forte habilidade em inferir (BRAGA, CARVALHO e LUDERMIR, 2014). Essa combinação de técnicas tem grande potencial para ganhar espaço cada vez maior considerando esse tipo de problema de previsão.

Os estudos de Mellit (MELLIT, BENGHANEM e KALOGIROU, 2006) e Diagne (DIAGNE, DAVID, *et al.*, 2013) que assim como outras pesquisas relacionam-se às irradiâncias e à predição, corroboram com a ideia de que há um número razoável de trabalhos que encontraram soluções eficientes pelo uso de redes neurais do tipo *feed-forward* perceptron multicamadas (MLP). Mas como podem existir RNA's mais eficientes para um problema em especial, há trabalhos em que as RNA's do tipo MLP são testadas e confrontadas com outros algoritmos de aprendizado, como no caso do estudo de Alfadda (ALFADDA, RAHMAN e PIPATTANASOMPORN, 2018) em que ML's do tipo MLP, SVM e KNN foram utilizadas para predição de irradiância por meio de aerossóis.

Acredita-se que o uso frequente das RNA's do tipo MLP seja decorrente de vários fatores a destacar:

- Facilidade de implementação dessas redes através de expressiva quantidade de linguagens de programação, *softwares* de análise e bibliotecas de programação;
- Resultados bem-sucedidos reportados em várias publicações;
- Pelo fato de ser uma das RNA's mais conhecidas, divulgadas e com mais tempo de utilização no meio científico.

No entanto, outras opções têm sido sugeridas como mais eficientes ou superiores. Há, por exemplo, um trabalho realizado por Alzahrani (ALZHRANI *et al.*, 2017) que realiza uma comparação entre alguns tipos de algoritmos de aprendizado: MLP, SVM e DRNN para predição da irradiância total. A última saiu-se melhor quando calculados os valores resultantes das métricas RMSE e MBE, considerando que os dados foram agrupados em quatro categorias de condições meteorológicas: céu claro, poucas nuvens, nuvens esparsas e céu sobrecarregado de nuvens. Mas chamou atenção o intervalo muito reduzido de medição entre as amostras de irradiância: 10ms. Na realidade 10ms já representa uma média de amostras coletadas a cada 1ms. É comum trabalhos utilizarem intervalos que compreendem segundos, minutos e frações de hora assim como horas, dias, etc., mas a necessidade de intervalos tão reduzidos não ficou satisfatoriamente justificada, especialmente perante a velocidade das nuvens. Para esclarecer melhor, segundo a escala Saffir–Simpson baseada na velocidade dos ventos e que é usada para classificar furacões, um furacão de maior escala, de classificação (5), significaria ventos com velocidades superiores a 70 m/s. Com essa velocidade uma nuvem percorreria apenas 70 cm em 10 ms. Em condições meteorológicas diárias/comuns esse intervalo de tempo provavelmente não conseguiria captar mudanças significativas nas irradiâncias. Também não ficou claro o intervalo à frente de predição. Intervalos curtos ou longos podem alterar substancialmente a capacidade de predição das RNA's e conseqüentemente poderiam provocar menores ou maiores discrepâncias em função de cada modelo bem como dos parâmetros individuais de ajustes.

O estudo de Lingfors (LINGFORS *et al.*, 2018) explora informações obtidas através de dados obtidos diretamente de sistemas FV e pelo dispositivo denominado Lidar – *Light Detection and Ranging* - posicionado sobre um satélite para avaliação do decréscimo de performance na produção de energia FV. O fator sombreamento foi mais uma vez identificado como a principal causa das flutuações com impacto direto na produção/controle de energia solar. Dois modelos foram criados dentro da metodologia do sistema, sendo aquele abastecido com dados do Lidar o que apresentou resultados mais acurados. Por outro lado, o segundo modelo denominado *QCPV-Tuning* baseado na parametrização dos módulos em relação à

geração FV apresentou imprecisões significativas segundo os pesquisadores devido a problemas originados dessa parametrização. Por fim, após aperfeiçoamentos relatados no trabalho, os autores consideraram os métodos como de confiança para avaliação dos recursos solares. Relataram ter obtido valores para o coeficiente de correlação de Pearson em uma faixa de 0,28 a 0,57 no melhor caso.

A corroborar com os trabalhos anteriores, o estudo levado a termo na Finlândia por Lappalainen (LAPPALAINEN e VALKEALAHTI, 2017) não apenas reafirma o sombreamento como um problema em parques solares FV, como ainda acrescenta que o sombreamento causado pelas bordas de nuvens tem um comprimento médio da ordem de 150m, o que significa que mesmo as maiores plantas solares serão afetadas pelo evento. Somando-se às demais perdas associadas ao sombreamento, afirmam os pesquisadores que as transições de irradiâncias podem levar a falhas relativas ao ponto de rastreamento de máxima potência e causar flutuações significativas na saída de potência de sistemas FV.

Entre as soluções propostas também houve menção às que usaram o algoritmo SVM de núcleo radial escolhido para pesquisas também relacionadas à predição de radiação solar. Foi o caso do estudo de Jiang (JIANG e DONG, 2017) cujo objetivo do estudo foi a previsão horária da irradiância global, um estudo de caso da área do Tibet. Através de uma variante do algoritmo SVM, denominado pelos autores como KSVM-SVS, o estudo alega ter atingido excelente performance nos resultados.

2.2.1 Satélites e *sky imagers*

Muitas predições de irradiância solar têm sido realizadas mediante modelos numéricos e observações de satélites por causa da larga cobertura espacial e temporal oferecida pelas imagens tomadas acima da atmosfera. Essa abordagem fornece informações prévias com antecedência de várias horas ou dias, entretanto podem ser significativamente tendenciosas e levarem a erros aleatórios do ponto de vista da estimação da irradiância (MATHIESEN e KLEISSL, 2011; LORENZ *et al.*, 2009). Há outro aspecto a se considerar favorável ao uso de satélites: por exemplo, se for assumida para as nuvens uma velocidade máxima de 160 km/h, isso seria aproximadamente equivalente a uma área de 2000 x 2000 km² em um horizonte de 6 h. Dados de satélite para essa situação seriam mais apropriados devido à sua ampla área de cobertura.

Sabe-se que dados de satélites e modelos numéricos de previsão meteorológica não têm resoluções espacial e temporal suficientes para registrar rápidas flutuações da irradiância na superfície terrestre (INMAN, PEDRO e COIMBRA, 2013). Uma alternativa é a utilização de imagens baseadas no solo para captar as condições meteorológicas locais. Assim, predições de curta duração baseadas em um tipo equipamento conhecido genericamente por “*sky imager*” (Figura 12), tornaram-se um novo campo para pesquisas (TROCCOLI, DUBUS e HAUPT, 2014). Entre eles existem modelos como o “*sky imager total*” (TSI) ou o “*sky imager inteiro*” (WSI) encontrados em diversos artigos publicados. Com alta resolução espacial e temporal, os *sky imagers* têm o potencial de capturar essas variações mais acentuadas na irradiância - também denominadas rampas - em intervalos sub-horários, o que inclui uma escala temporal inferior a 1 minuto. Há, porém, inconvenientes encontrados nesses dispositivos. A banda de sombra assim como o braço que segura a câmera tornam parte da imagem obscura, ocasionando perda de informações.

O *sky imager inteiro* (WSI) foi desenvolvido e aperfeiçoado desde os idos de 1950 (SHIELDS *et al.*, 1998). O sistema de predição desse equipamento consiste em 4 componentes (PELLAND *et al.*, 2013):

- Aquisição de imagens nas vizinhanças da região onde há o estudo de predição;
- Análise da imagem do céu para identificar nuvens;
- Estimativa do vetor de deslocamento de nuvens usando imagens sucessivas;
- Uso da localização das nuvens e dados do vetor de deslocamento para avaliar a nebulosidade, irradiância e efetuar predições de potência captada.

Figura 12 – Exemplo de dispositivo *sky imager*



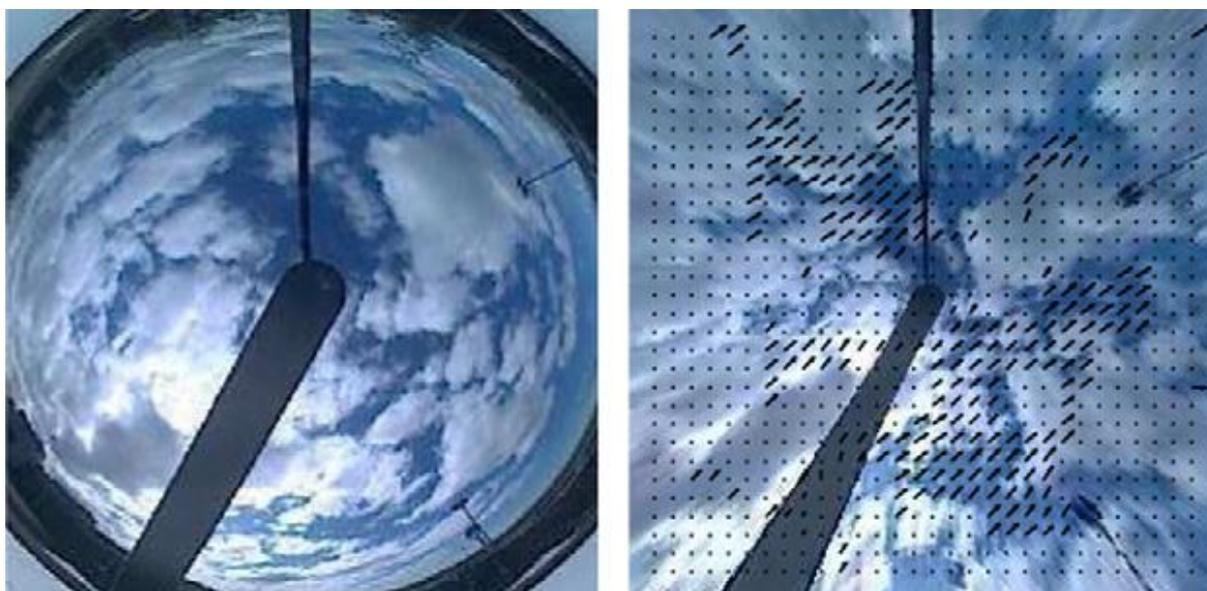
Fonte: (CHOW *et al.*, 2011)

Sobre o horizonte temporal de predição através de um dispositivo *sky imager*, é útil lembrar que ele depende da altura das nuvens assim como de suas velocidades, pois tem a limitação relativa ao tempo que a imagem monitorada atravessa a área de interesse. Nos estudos desenvolvidos por Chow (CHOW *et al.*, 2011) esse tempo foi determinado na faixa entre 5 e 25 minutos. Por outro lado, um trabalho conduzido por Marques e Coimbra (MARQUEZ e COIMBRA, 2013) para predição da irradiância normal direta usou um horizonte variando entre 3 e 15 minutos. Para nuvens baixas e rápidas, o horizonte de predição pode ser de apenas 3 minutos, ao passo que para nuvens altas e lentas esse horizonte pode se estender além dos 30 minutos, mas geralmente o valor típico é de 5 a 20 minutos (PELLAND *et al.*, 2013). É importante observar que mesmo se o tamanho da nuvem e sua velocidade puderem ser determinados precisamente, a acurácia da predição depende da taxa em que as nuvens estejam evoluindo ao longo do seu vetor de movimento.

Diretamente associada à identificação de nuvens por imagens, está a determinação de seus vetores de movimento, o que representa uma etapa seguinte ao estudo. Métodos variados para extrair o vetor de deslocamento a partir de imagens estão presentes em diversos trabalhos. Há um exemplo de obtenção do vetor de movimento através de uma conhecida biblioteca de

tratamento de imagens, OpenCV, (WOOD-BRADLEY, ZAPATA e PYE, 2012) que tem interfaces que permitem seu uso nas linguagens C/C++, Python ou Java. Em outra técnica, através de correlação cruzada entre pixels retirados de imagens sucessivas, Chow e colaboradores (CHOW *et al.*, 2011) calcularam o deslocamento médio assumindo que a nuvem seja formada por uma só camada. Foi observado que eventos como a formação, dissipação e deformação de nuvens poderiam resultar em uma diminuição da correlação. Em mais um exemplo de obtenção de vetores de movimento (MARQUEZ e COIMBRA, 2013), através de um *software* desenvolvido no ambiente Matlab (de nome MPIV, cujo *download* pode ser feito no endereço “www.oceanwave.jp/software/mpiv/”), foi aplicado um método chamado de MQD, que significa mínimo quadrado da diferença. Assim foram obtidos campos de vetores por meio de imagens sucessivas afastadas no tempo por 1 minuto. Depois de submetidos a operações de filtragem (por meio de um caixa de ferramentas incluídas no software), é obtido um resultado final para os vetores. A Figura 13 ilustra um resultado.

Figura 13 – Imagem de nuvens à esquerda e vetores de movimento à direita

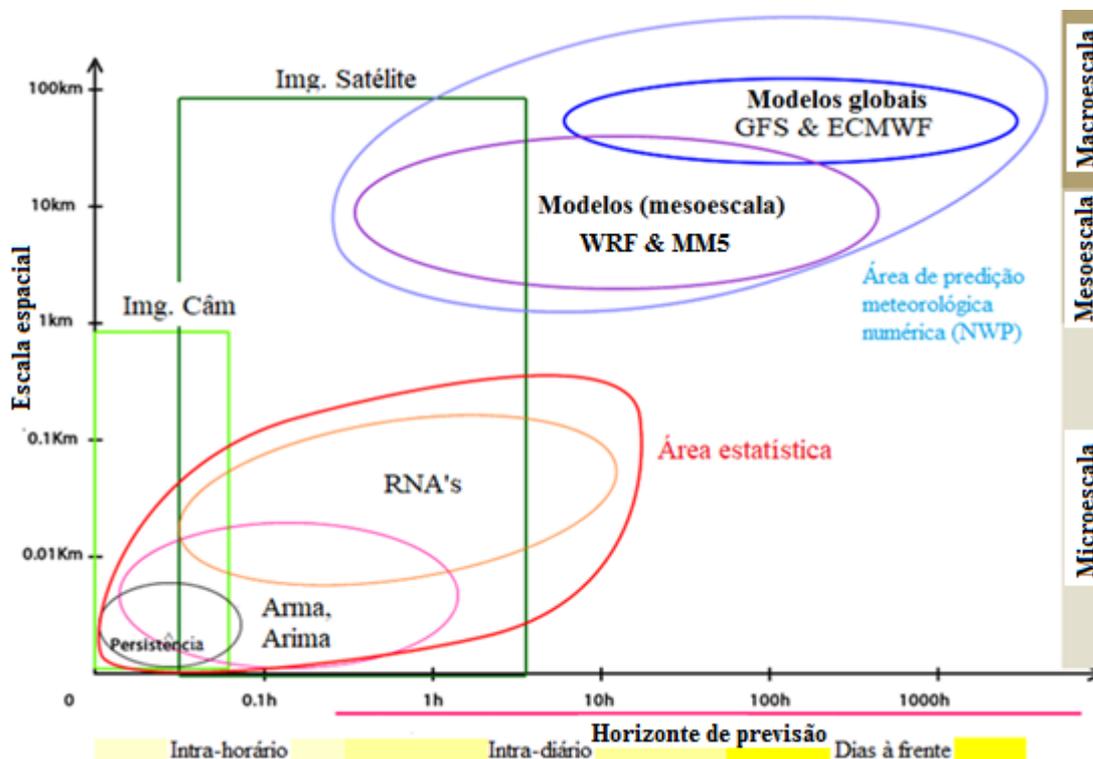


Fonte: adaptado de (MARQUEZ e COIMBRA, 2013)

A Figura 14, resultante do trabalho de Diagne (DIAGNE *et al.*, 2013), apresenta-se praticamente como uma síntese de boa parte do que foi exposto nesse e nos últimos tópicos. Encontram-se relacionados vários elementos: os horizontes espaciais, os horizontes temporais e os sistemas de aquisição de imagens (câmera ou satélite). Por fim, como um guia de metodologias preditivas, são apresentadas diferentes abordagens para cada família de modelos

de predição em função das interseções dos horizontes e dos dispositivos de aquisição de dados, sejam imagens ou dados meteorológicos/atmosféricos.

Figura 14 – Horizontes temporais, espaciais, imagens e modelos de predição



Fonte: adaptado de (DIAGNE *et al.*, 2013)

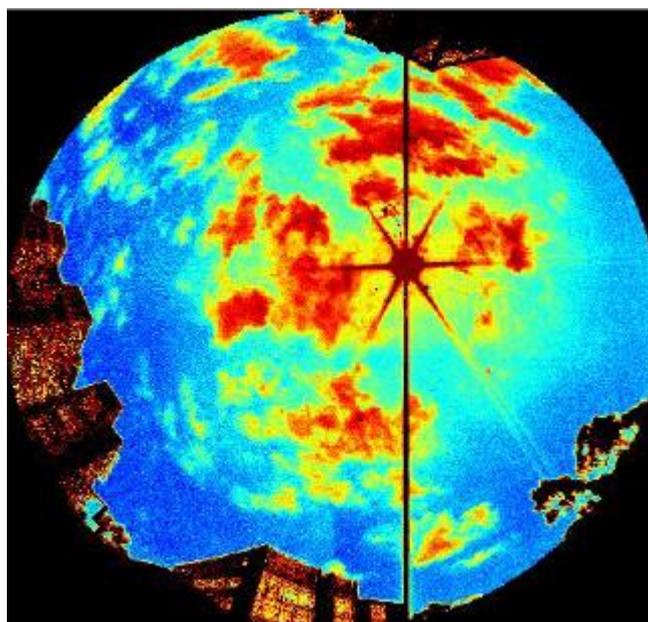
2.3 Pixels, espaços de cores e segmentação

Dada a quantidade de trabalhos que exploraram espaços de cores dos pixels de imagens, faz-se necessário estudar e avaliar quais informações foram mais utilizadas e de que forma foram empregadas em cada pesquisa. Assim, ficou evidente durante revisões de literatura os esforços direcionados em diversas abordagens para extrair com maior eficiência possível o conjunto de pixels de uma imagem de modo a melhor identificar as formações de nuvens, segmentando-as em meio aos outros elementos dessa imagem. Suas formas - extremamente indefinidas e com variabilidade de cores que dependem de fatores como a posição relativa do Sol, da hora do dia e de quanto são espessas/densas – têm representado um grande desafio que se reflete em métricas e taxas de acertos em processos de classificação.

O trabalho de Chow (CHOW *et al.*, 2011) explorou a importância da razão R/B (considerando o espaço de cores RGB) e o uso de valores limiares para construção de um

algoritmo de decisão sobre a presença de nuvens. O valor médio de R/B também assumiu importância como parâmetro classificatório na medida em que é usado para identificar se há sombreamento do Sol. A figura 15, colorida artificialmente em função do parâmetro RBR (outra denominação da razão de R para B) de cada pixel, exemplifica essa técnica que, no caso particular desta Figura, consiste em classificar os pixels como pertencentes a uma nuvem quando $RBR \approx 1$ ou como céu quando $RBR < 1$ (KURTZ *et al.*, 2015)

Figura 15 – Imagem com aplicação da razão vermelho para azul (RBR)



Fonte: (PELLAND *et al.*, 2013)

Liu e colaboradores (LIU *et al.*, 2015), por outro lado, desenvolveram um método de detecção de nuvens com base em um algoritmo denominado superpixel segmentation (SPS) (REN e MALIK, 2003), uma técnica popular na área de segmentação de imagens. O nome decorre do fato de a imagem ser dividida em regiões irregulares denominados superpixels. Essa divisão irregular é construída com base em similaridades entre texturas, similaridades de brilho e continuidade de contorno em uma imagem. Para cada região de imagem, ou superpixel, é obtido um limiar. Tomando todas as regiões, é construída uma matriz de limiares por interpolação bilinear dos limiares locais. Finalmente, a detecção de nuvens é alcançada por comparação pixel a pixel entre uma característica derivada da expressão $(R - B)$ com a matriz de limiares. O resultado apresentou uma acurácia (F-Score, Precision e Recall) acima de, respectivamente, 82%, 84% e 81% para uma base de dados e de 93%, 94% e 92% para outra base de dados. Segmentação de nuvens em imagens noturnas (DEV *et al.*, 2017) também já usaram o conceito de superpixels bem como parâmetros que são usuais na segmentação diurna.

O trabalho de Narain (NARAIN e SOHI, 2015) trouxe uma compilação de 10 algoritmos desenvolvidos para segmentação de nuvens. O algoritmo mais simples utiliza para segmentação limiares pré-estabelecidos a partir das expressões “R/B” e “(R-B)”. Outro algoritmo apresentado realiza um método mais sofisticado por meio de uma transformada rápida de Fourier (FFT) de imagem para analisar a homogeneidade de pixels simétricos, sabendo-se que na ausência de nuvens há maior homogeneidade nos elementos de cores do céu sob análise. A partir de então são realizadas comparações com base na razão “R/B” e em histogramas relativos a esse parâmetro. Na etapa final, ainda há um mecanismo de busca de contornos das nuvens. Outras técnicas enumeradas nesse texto fazem uso da razão “(B-R)/(B+R)” e usam redes neurais do tipo K-Means Clustering e Fuzzy Clustering como mecanismos de classificação. Segundo o texto, a última dessas técnicas, denominada “Graph Cut Algorithm”, conseguiu 94.7% de acurácia na classificação. É oportuno observar aqui que outros trabalhos, como o de Marquez (MARQUEZ e COIMBRA, 2013), também utilizam a expressão “(B-R)/(B+R)” como uma alternativa normalizada para o parâmetro “R/B”.

Dois espaços de cores, no caso RGB e HSV, foram utilizados para classificação de nuvens no artigo de West (WEST *et al.*, 2014). Os parâmetros R/B e (R-B) também foram incluídos pelo fato de conseguirem lidar com uma alta faixa dinâmica presente em várias diferentes condições de céu. Contudo, os parâmetros anteriores podem produzir erros de classificação quando a faixa dinâmica total da cena não for capturada, o que é comum quando o Sol e nuvens densas são visíveis simultaneamente. Erros ocorreram frequentemente em certas situações: no meio de nuvens muito escuras e espessas, nas áreas ao redor do sol devido ao brilho intenso e no início/final do dia quando as cores do céu claro mudam substancialmente. Para superar essas dificuldades cogitou-se sobre a possibilidade de usar uma biblioteca de imagens que pudesse fornecer informações sobre todas as condições de céu claro ao longo do dia, mas exigiria uma grande quantidade de dados e poderia tornar-se de implementação muito difícil, considerando-se a especificidade de cada região e a quantidade delas. Além das características de cores, esse trabalho incluiu mais outras quatro: uma estimativa do movimento do pixel em relação à imagem anterior, a distância do pixel ao Sol e os ângulos azimutais e zenitais do Sol. A acurácia desse modelo atingiu 97%. De acordo com os autores, o teste de validação cruzada para 10 folders foi usado.

No artigo de Dev (DEV, LEE e WINKLER, 2014), os pesquisadores trabalharam na segmentação de nuvens a partir do estudo da bimodalidade (KNAPP, 2007) aplicado às distribuições de alguns parâmetros baseados em cores, como também pela análise de

componentes (PCA – Principal Component Analysis) voltada a esses parâmetros. Posteriormente, aplicou-se uma ferramenta de classificação de pixels através da rede neural artificial do tipo Fuzzy Clustering. Vários espaços de cores foram analisados na procura pelos melhores parâmetros que produzissem distribuições bimodais: RGB, HSV, YIQ, CIE e Lab. Dos canais existentes nesses espaços, os autores selecionaram três parâmetros como os mais importantes do ponto de vista da bimodalidade: “S”, “R/B” e “(B-R)/(B+R)”. A análise através do método PCA também evidenciou esses mesmos três parâmetros pela sua maior significância dentro da primeira componente principal. Novos parâmetros também foram gerados para uma abordagem bidimensional. Uma avaliação posterior indicou que a diferença entre a componente I (do espaço de cores YIQ) e as três componentes anteriores resultaram em três parâmetros ainda mais significativos: “S – I”, “R/B – I” e “(B-R)/(B+R) – I”. A estratégia final levou a um aumento da precisão, inicialmente em 84%, para algo em torno de 90%.

A técnica de separação dos canais de cores e regiões de pixels é também uma abordagem comum utilizada em outras áreas de estudo que usam recursos de processamento/tratamento de informações digitais capturadas em cenas (CHAITANYA e VIKRAM, 2015). Trata-se, na realidade, de método geral e bem difundido de segmentação de imagem através da divisão da mesma em diversas sub-regiões (conjuntos de pixels) (NARAIN e SOHI, 2015) com o objetivo de mudar e/ou simplificar sua análise (GONZALEZ e WOODS, 2001). A Figura 16, por exemplo, exibe as etapas usadas para segmentação da imagem adquirida de um *sky imager* pelo método RBR (SLATER, LONG e TOOMAN, 2001). A partir da imagem original são extraídas as componentes em vermelho e azul. Então, a imagem final é obtida numericamente através da razão R/B e um valor limiar, realçando e destacando a presença de nuvens. Fica clara a obstrução de parte da imagem pela presença da estrutura do *sky imager*, ocultando muitos pixels e informações. Esse inconveniente gera a necessidade de colocar filtros (ou máscaras) para retirar da análise da imagem a estrutura do equipamento.

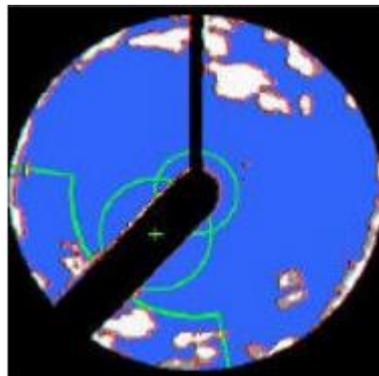
Figura 16 – Etapas para segmentação da imagem segundo a razão vermelho/azul



Fonte: (SLATER, LONG e TOOMAN, 2001)

A Figura 17 apresenta a transformação da imagem original de acordo com o critério de decisão adotado. Abaixo de um limiar da razão RBR os pixels receberam a cor azul. Acima desse valor a região foi classificada como nublada e recebeu a cor branca. Segundo os autores do artigo, as cores cinza e vermelha ainda foram sobrepostas na imagem para representar regiões opacas ou de nuvens finas.

Figura 17 – Imagem para critérios de decisão



Fonte: (SLATER, LONG e TOOMAN, 2001)

Ainda é possível utilizar uma biblioteca de imagens de céu claro, o que auxilia especialmente no caso em que as nuvens se apresentem na imagem como um halo fino. Contudo, isso requer a criação de uma base de dados significativamente grande de imagens de céu claro para serem analisadas, o que pode ser impraticável para aplicações rápidas e em várias localidades distintas (WEST *et al.*, 2014). Outros métodos, além do RBR, foram concebidos para classificar os pixels de nuvens na imagem, como, por exemplo, através do espaço de cores

HSV que usa uma métrica para o cálculo da saturação (S) da cor a partir das cores RGB de um pixel, conforme as expressões (Eq. 7) e (Eq. 8).

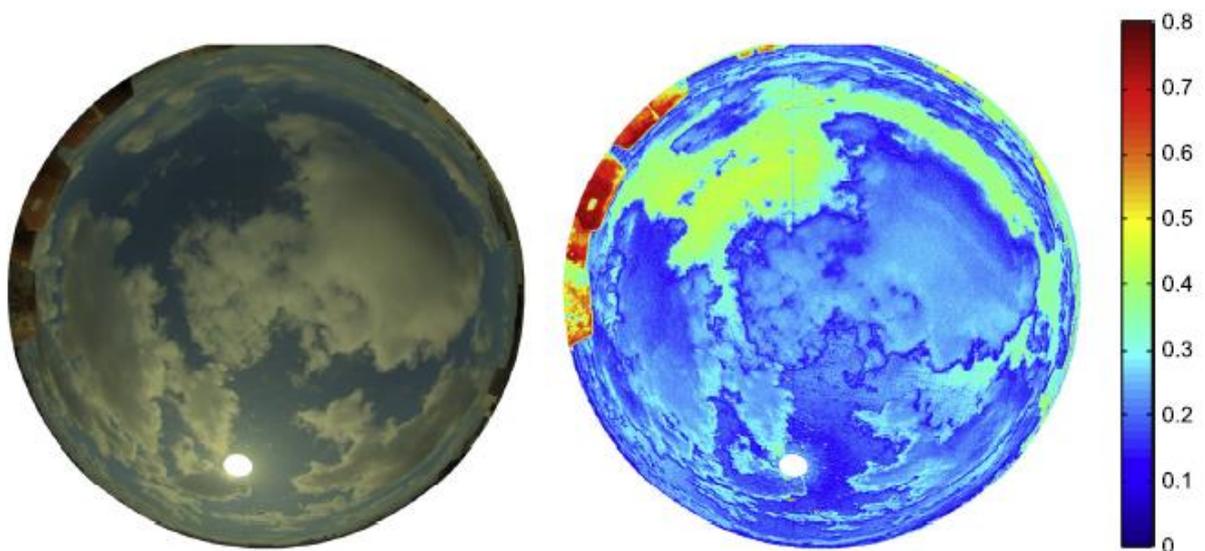
$$S = 0, \text{ se } \text{Max}(R, G, B) = 0 \quad \text{Eq. 7}$$

$$S = \frac{\text{Max}(R,G,B) - \text{Min}(R,G,B)}{\text{Max}(R,G,B)}, \text{ se } \text{Max}(R, G, B) > 0 \quad \text{Eq. 8}$$

A Figura 18 mostra outro exemplo de segmentação de uma imagem, dessa vez usando esse valor de saturação para determinar as cores artificiais e destacar as nuvens. A correspondência numérica estabelecida entre essas cores e o valor da saturação é apresentada em uma coluna à direita nesta imagem.

Além dessas técnicas, existe a possibilidade de usar esses parâmetros funcionando simultaneamente em conjunto como entradas de uma rede neural artificial para classificação, de modo a identificar a eventual presença e localização de nuvens na imagem.

Figura 18 – Exemplo da aplicação da saturação (do espaço HSV) a uma imagem



Fonte: (NGUYEN e KLEISSL, 2014)

Entre tantos trabalhos existentes, não poderiam deixar de serem mencionados aqueles que envolvem a aplicação de redes neurais do tipo convolucional (CNN) ou suas variantes, considerando que em outras áreas de estudo têm sido parte de soluções para segmentação de objetos de um modo geral presentes em uma cena. No caso específico de segmentação de

nuvens, há trabalhos como o de (DRÖNNER *et al.*, 2018) em que foram tomadas imagens de um conjunto de 11 satélites para classificação. Essas imagens foram obtidas com resolução espacial de 3km x 3km e resolução temporal de 15 minutos. Considerando os diversos cenários das imagens, céu-claro, céu sobrecarregado de nuvens, regiões de solo e neve, os autores obtiveram taxa de acertos superior a 94%. Também a explorar imagens de satélite e a técnica das redes neurais do tipo CNN, Xie e colaboradores (XIE, *et al.*, 2017) desenvolveram seu algoritmo de segmentação. Segundo o *site* que representa o satélite utilizado (GLCF, 2019), a área da cena corresponde a um quadrado de dimensões 16,5km x 16,5km. Foram utilizadas 81 imagens, sendo 38 para treinamento e 43 para testes. Além disso, utilizaram blocos de imagens (patches) de tamanhos 55 x 55 e 111 x 111, havendo um total de 204.000 pares de patches para treinamento dos quais foram reservados, respectivamente, 26.000, 22.000 e 156.000 para nuvens, nuvens finas e não-nuvens. Conceitos como superpixels foram utilizados juntamente com os espaços de cores RGB e HSV como elementos internos dentro do processamento da RNA convolucional. Ao final da implementação do algoritmo, os pesquisadores conseguiram acurácia de 97,8% e 84,67% para as classificações como nuvens e nuvens finas, respectivamente.

A revisão de literatura apresentou um amplo espectro de soluções extraíndo informações meteorológicas e de imagens por recursos matemáticos, estatísticos e computacionais das mais variadas formas ao mesmo tempo em que estavam a aperfeiçoar e/ou desenvolver novas estratégias. O estudo relatado neste texto optou também por trilhar alguns caminhos diferentes em busca de soluções inovadoras e com potencial de apresentar resultados acurados. No que está associado, por exemplo, aos espaços de cores, houve uma busca intensa para encontrar novos parâmetros derivados de alguns dos espaços já mencionados. Para segmentação de nuvens, optou-se por adotar uma máquina de aprendizado do tipo SVM (SVC) por meio de uma biblioteca de software bem-conceituada para uma abordagem dirigida aos pixels. Some-se a essa abordagem um espaço vetorial inovador beneficiado por novos atributos desenvolvidos. O cálculo (modelo) das irradiâncias por imagens do céu, por sua vez, demandou a introdução de algoritmos e parâmetros de natureza estatística classificatória juntamente com os resultados de segmentação obtidos em etapa anterior. Na etapa final todos os dados e soluções inovadoras obtidas foram agrupados na construção de um modelo preditivo de irradiâncias.

3 MATERIAL E MÉTODOS

Existe uma grande diversidade de abordagens com equipamentos muitas vezes de alto custo ou de difícil acesso para predição das flutuações da irradiância solar. Seguindo uma orientação voltada ao desempenho dos objetivos dentro da melhor relação custo/benefício, a metodologia aqui adotada procurou associar a escolha, quando possível, de equipamentos de precisão a custos de aquisição menores que aqueles usados em outras abordagens. Dessa forma, ao invés de um equipamento *sky imager* de preço elevado, como elemento de aquisição de dados de imagem, foi adquirida uma câmera com grande campo de visão (*fisheye*) e com várias alternativas de resolução para registrar o céu capturando todo o hemisfério. Outra fonte substancial de dados foi proveniente da estação solarimétrica do Grupo de Estudos FAE: estação rastreadora Solys 2, pireliômetro sNIP e piranômetro CMP-21. Em adição, um equipamento para medição da altura das nuvens, denominado ceilômetro/tetômetro, foi conseguido através de doação do órgão CINDACTA III de Recife e foi utilizado apenas inicialmente para estimativas da altura base de nuvens. Ainda fizeram parte outros elementos para registro/armazenamento e processamento de dados como computadores, *softwares* (linguagens de programação, bibliotecas de programação), etc.

3.1 Câmera tipo *fisheye*

A câmera selecionada foi o modelo FE8174V da Vivotek (Figura 19). Seguem algumas especificações da câmera:

- Sensor CMOS progressivo (1/2,5") com resolução 2560 x 1920;
- Resolução máxima: 1920 x 1920;
- Campos de visão: 180° x 180° x 180° (horiz. x vert. x diag.);
- IP, compatível com o padrão ONVIF;
- Aceita alimentação via PoE (*power over ethernet*);
- Fluxo de imagens via rede de dados padrão Ethernet;
- Dimensões: Ø: 145 mm x 47 mm;
- Peso: 545g.

Figura 19 – Câmera *fisheye* Vivotek modelo FE8174V



Uma imagem obtida por ela é apresentada na Figura. 20, porém com a devida observação de que havia alguns parâmetros da câmera, como filtros de imagem (WDR, BLC, etc) que foram posteriormente desativados para ser possível adquirir fotografias mais naturais, sem que tenham sido modificadas por algoritmos internos cujo funcionamento seja desconhecido por tratar-se de segredos industriais de produção do dispositivo de videomonitoramento. A essa imagem foi aplicado o método RBR para detecção de nuvens, cujo resultado é exemplificado por meio da Figura 21, uma imagem segmentada de forma extremamente simples e exclusivamente através de limiares entre as razões dos canais R e B do espaço de cores RGB. Neste exemplo, ainda rudimentar comparado ao que poderia ser feito, percebe-se um clarão em volta do Sol que dificulta sobremaneira a distinção entre os pixels que representam nuvem, céu e Sol. A saturação da imagem com a forte exposição à luz também está associada às lentes usadas que são atingidas por esse excesso de luz direta e ofuscante. Seguem exemplos que podem ser vistos na Figura 22.

Figura 20 – Imagem obtida pela câmera no teto do laboratório



Figura 21 – Método simplificado da aplicação do método RBR para detecção de nuvens

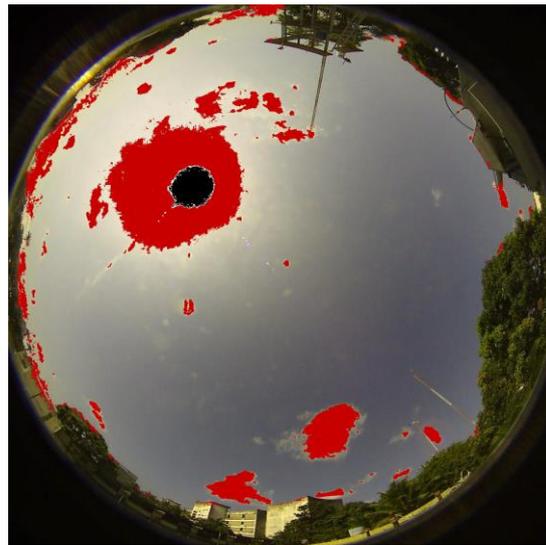
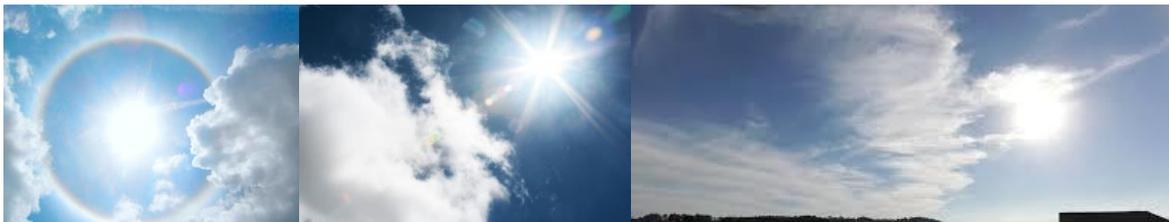


Figura 22 – Ofuscamento, saturação, "flare" e outros efeitos em imagens



Entre as muitas razões para a escolha da câmera de vídeo FE8174V, uma delas foi a grande flexibilidade em poder selecionar suas múltiplas resoluções além de outras configurações via

protocolo TCP/IP. É sabido que uma resolução alta não significa necessariamente um aumento proporcional da precisão obtida em reconhecimento de imagens. Ainda mais quando esses dados de imagem servirão de entrada para a máquina de aprendizado (ML) que terá o papel apenas de identificar se os pixels da cena pertencem a uma nuvem. É possível, conforme já foi relatado (WEST *et al.*, 2014) que uma diminuição conveniente da resolução da imagem com ganho em velocidade de processamento não tenha uma contrapartida negativa ou proporcionalmente desfavorável em relação à diminuição da capacidade de reconhecimento da imagem e predição do resultado. Tomando como referência o estudo de West e colegas (WEST *et al.*, 2014) que escolheram um intervalo de 10s entre as imagens, uma imagem a cada 5 s foi a taxa escolhida de amostragem para permitir uma maior resolução temporal, granularidade da sucessão de cenas, além de possibilitar uma mudança eventual para uma taxa de amostragem menor para efeito de testes, ou seja, uma imagem a cada 10 s, por exemplo. É importante mencionar que a capacidade de armazenamento das imagens no computador está intimamente ligada a essas escolhas, tendo sido necessário o dimensionamento apropriado do seu disco rígido (HD) para acumular a quantidade de imagens e dados em geral necessários ao pós-processamento.

3.2 Estação solarimétrica e *datalogger*

Os valores das irradiâncias global, direta e difusa (GHI, DNI e DHI, respectivamente) foram coletados através da estação solarimétrica mostrada na Figura 23. Mas antes dos valores numéricos seguirem para a base de dados local, foram digitalizados e armazenados provisoriamente no *datalogger* CR-3000 (Figura 27), uma vez que os sinais de saída das irradiâncias são inicialmente enviados na forma analógica. Depois de armazenados no *datalogger*, esses dados foram transferidos via comunicação serial para o computador e tomaram a forma de arquivo texto onde, finalmente, o software desenvolvido fez a leitura para o processamento dos valores medidos. A importância da acurácia das leituras dos dispositivos se mostrou essencial para o êxito do experimento, pois através desse conjunto de instrumentos de precisão foi possível acompanhar melhor as flutuações das irradiâncias diante da presença de nuvens. Dessa forma, a discussão dos equipamentos de medida em termos de acurácia é um aspecto que merece ênfase. Nesse contexto, o padrão ISO 9060:1990 (*Solar energy - Specification and classification of instruments for measuring hemispherical solar and direct solar radiation*) estabelece dentro das especificações sobre piranômetros, sob a referência 3c,

que equipamentos classificados como de primeira ou segunda classe têm uma faixa de erros de $\pm 20 \text{ W/m}^2$. A classificação designada como padrão secundário, caso do piranômetro CMP-21 (Figura 25), fabricado pela *Kipp & Zonnen* e incorporado à estação solarimétrica do Grupo FAE, encontra-se dentro de uma acurácia de $\pm 10 \text{ W/m}^2$. É importante salientar que sua faixa espectral compreende desde o infravermelho até parte do ultravioleta, percorrendo o intervalo de 285 a 2800nm – e designado como classe A espectralmente plana, segundo o fabricante. De forma semelhante, o pireliômetro sNIP, Figura 24, fabricado pela *Eppley Laboratory, Inc*, também está classificado de acordo com o padrão secundário, tendo assim a mesma acurácia. Sua faixa espectral compreende o intervalo de 250 a 3000nm. Esse dispositivo (pireliômetro) está montado sobre o sistema de rastreamento solar Solys 2 (da *Kipp & Zonnen*) que se pode ver na Figura 23. Quanto ao espectro eletromagnético e sua correspondência com as cores foi incluída a Figura 26 para exemplificar qual a largura visível do espectro de cores para esses equipamentos de medida.

Figura 23 – Estação solarimétrica do grupo FAE



Figura 24 – Pireliômetro sNIP

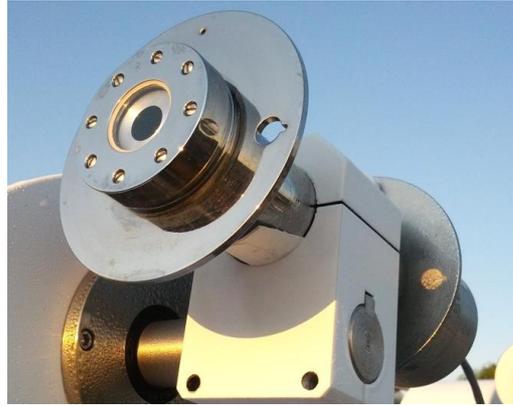


Figura 25 – Piranômetro CMP-21



Figura 26 – Espectro de frequências eletromagnéticas e cores associadas

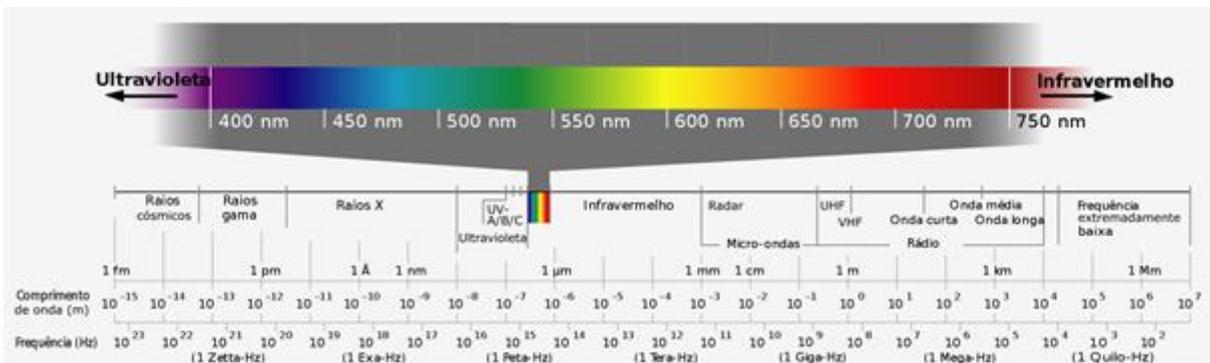


Figura 27 – Datalogger CR3000

3.3 Captura de imagens e coleção de dados de irradiâncias

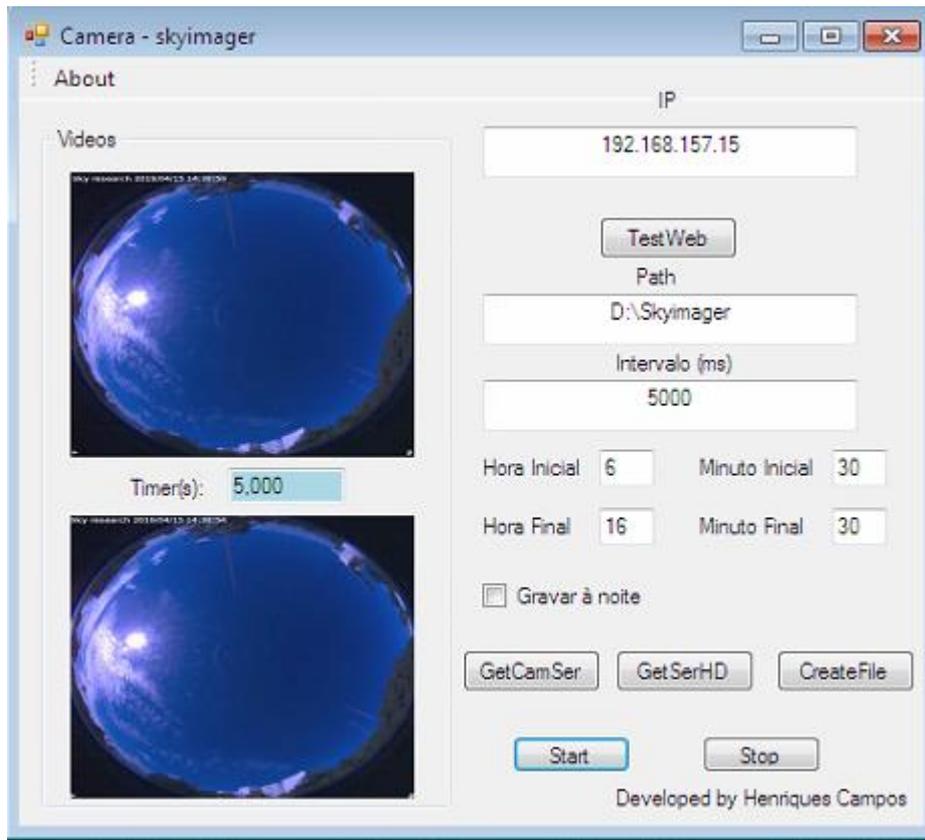
O fluxograma que ilustra o percurso dos dados coletados pode ser visto na Figura 28 que oferece uma visão global do processo de aquisição. No fluxograma encontra-se a parte inicial do experimento, a captura de imagens. Foi realizada através de um programa desenvolvido em C#/C++ especificamente com esse propósito. O armazenamento de imagens foi programado no aplicativo para ser efetuado em intervalos regulares próximos a 5000ms, como se pode ver na Figura 29. Pode-se ver na figura também um instante em que o fluxo de imagens é capturado e uma fotografia é selecionada e mostrada na tela após ser armazenada em uma base de dados diária. Para tornar possível esse fluxo de imagens fez-se necessário configurar uma rede de dados LAN para acesso do computador ao fluxo de imagens proveniente dessa câmera que usa o protocolo TCP/IP para comunicar-se. Em relação ao período diário de aquisição de imagens, foi atribuído um intervalo horário de captura das 6:30 da manhã às 16:30 da tarde. As imagens foram selecionadas no aplicativo da câmera com resolução de 1056 x 1056 pixels e posteriormente reduzidas para 528 x 528. Quanto à profundidade das imagens, esse é um atributo fixo, não selecionável, sendo de 8 bits para cada canal de cores RGB de acordo com a especificação da câmera.

Figura 28 – Fluxograma de entrada de dados no computador do laboratório



Além das imagens, os valores das irradiâncias total, direta e difusa foram armazenados em arquivos distintos, dia após dia, em intervalos de um segundo entre cada medição diária. Propositalmente, o intervalo diário de medição das irradiâncias engloba o intervalo de armazenamento de imagens para ser possível realizar análises e inferências. O *datalogger* CR3000 realizou as medições digitais das irradiâncias da estação solarimétrica através de um conversor analógico-digital (16 bits, resolução de uma parte em 65536) e enviou através de uma interface serial RS-232 esses dados em tempo real para a base de dados do mesmo computador que também estava a coletar as imagens.

Figura 29 – Programa desenvolvido para captura de imagens



3.4 Fluxo ótico e velocidade das nuvens

As nuvens estão associadas às variações mais acentuadas na irradiância - também denominadas rampas. A frequência e a duração do fenômeno, por sua vez, naturalmente são influenciados pela velocidade de passagem das nuvens, o que inclui a direção do movimento. Por essa razão e para aumentar o conhecimento sobre o deslocamento das nuvens em imagens, o estudo da velocidade de nuvens foi incluído neste trabalho.

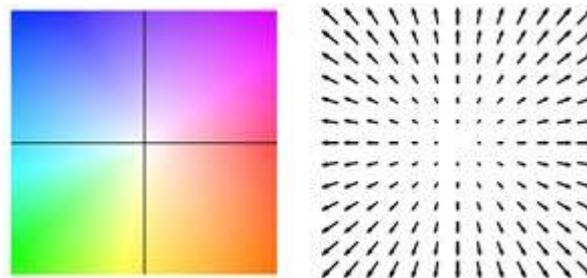
O estudo do deslocamento de objetos e pixels em imagens é, sob o ponto de vista de Visão Computacional, conhecido como fluxo ótico. Há vários estudos sobre esse assunto cujas abordagens ainda o distinguem em duas classificações: como fluxo ótico denso ou esparsos.

3.4.1 Fluxo ótico denso

Inicialmente optou-se por avaliar através do fluxo ótico denso as velocidades (em pixels/por unidade de tempo) através de um algoritmo largamente utilizado e conhecido por Gunnar Farneback (FARNEBACK, 2003), incluído na biblioteca OpenCV. Para melhor

visualização da velocidade nas fotografias, uma das técnicas é representar sobre a imagem pequenos vetores em intervalos regulares indicando módulo (normalizado) e direção do deslocamento de pixels. Uma outra alternativa também utilizada é a representação através de mapas de cores, alguns chamados de “*wheels of colors*”, mas há o inconveniente de ocultarem a imagem original. Na Figura 30 abaixo vemos um exemplo dessas duas representações.

Figura 30 – Mapeamento de velocidades por cores (à esquerda) e vetores (à direita)



A imagem retratada na Figura 30 do lado direito exemplifica o uso de vetores na caracterização do fluxo óptico denso.

3.4.2 Fluxo óptico esparsos

Várias alternativas encontram-se disponíveis quando se trata do cálculo do fluxo óptico esparsos. Um algoritmo frequentemente utilizado por aqueles que dispõem da biblioteca OpenCV é o de Shi (SHI e TOMASI, 1994) que procura por características (pixels) em duas imagens que apresentem maior similaridade. Esse e outros detectores de características (em imagens) muito utilizados em Visão Computacional e reconhecimento de padrões foram testados com sucesso, tais como SURF (BAY *et al.*, 2008) e SIFT (LOWE, 1999), mas optou - se por usar o detector FAST (ROSTEN, PORTER e DRUMMOND, 2010), não apenas pelos bons resultados que o método apresenta e por isso utilizado em grande quantidade de trabalhos, mas também pela rapidez de processamento do algoritmo, qualidade importante. Abaixo, segue uma fotografia colocada no *site* da biblioteca OpenCV como resultado da aplicação de um algoritmo para fluxo óptico esparsos (Figura 31).

Figura 31 – Fluxo óptico esparso aplicado ao monitoramento de veículos

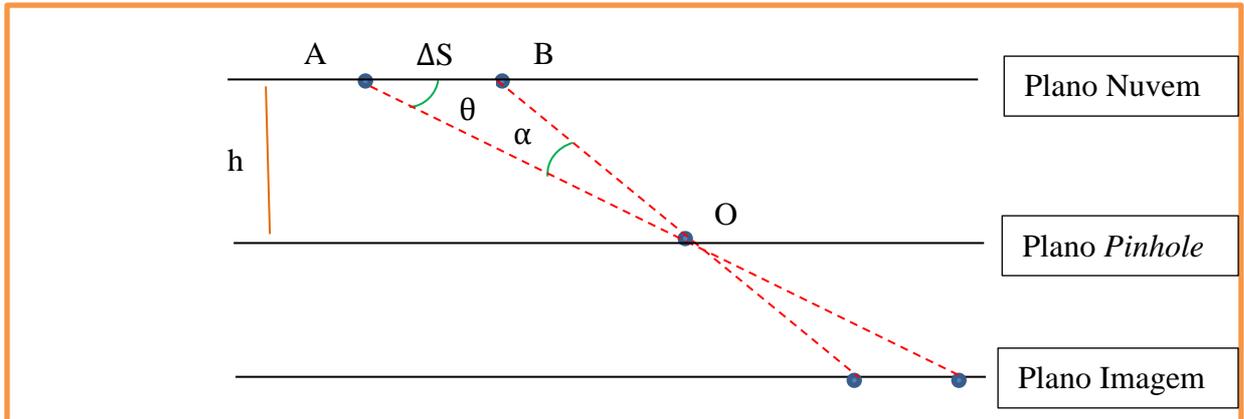


Fonte: biblioteca OpenCV

Apesar da natureza aleatória que envolve o movimento das nuvens, para efeito de avaliação e simplificação de cálculos, será feita a suposição de que um determinado ponto (de uma nuvem) percorre uma trajetória retilínea com velocidade constante durante um certo intervalo de tempo (no plano da nuvem) e é captado em um plano paralelo, o plano da imagem, na forma de um pixel. Então dois casos diferentes serão estudados. No primeiro, imagine-se uma trajetória radial que passaria pelo centro da imagem; no segundo caso uma trajetória tangencial em determinado ponto.

Na Figura 38 mais à frente, que compõe o modelo utilizado para a câmera *fisheye*, é possível notar que a posição de um pixel na imagem é determinada pelos ângulos formados pelo ponto da imagem real em relação à abertura pela qual passam os raios luminosos. Assim, na primeira situação, ilustrada pela Figura 32, são consideradas três retas paralelas em planos paralelos, de forma tal a retratar o caso em que a trajetória do pixel passa pelo centro da imagem. O ponto A pertencente a uma nuvem dirige-se ao ponto B movendo-se segundo uma velocidade constante “ v ” durante o deslocamento da nuvem. O ponto “O” é o orifício que representa a abertura da câmera.

Figura 32 – Deslocamento radial de um ponto na nuvem e sua imagem em forma de pixel



Sejam considerados “ θ ” o ângulo de elevação e “ h ” a distância do plano (horizontal, por hipótese) da nuvem ao plano da abertura por onde entram na câmera os raios luminosos. Então seguem as relações matemáticas que descrevem a variação do deslocamento de um ponto da nuvem (ΔS) (dentro de um determinado intervalo de tempo (Δt)) em correspondência com a variação do ângulo da câmera (α) que representa esse ponto na forma de um pixel de imagem. Uma vez que foi selecionada a resolução final da imagem em 528 x 528 pixels para 180° da câmera *fisheye*, a distância percorrida por um pixel poderá ser associada à variação angular (α) por uma regra de três simples. Finalmente, após alguns cálculos pode-se estabelecer que:

$$\frac{\Delta S}{h} = \frac{\sin \alpha}{\sin \theta \sin(\theta + \alpha)} \quad \text{Eq. 9}$$

A função seno por sua vez, expandida em série de Taylor, apresenta-se como:

$$\text{sen } x = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} \dots + (-1)^k \frac{x^{2k+1}}{(2k+1)!} + \dots \quad \text{Eq. 10}$$

Antes das argumentações que seguem, seria útil lembrar nesse caso que a velocidade do ponto no plano da nuvem é radial e constante. Se o primeiro membro da Eq. 9 for considerado constante em intervalos regulares de tempo e o ângulo “ α ” suficientemente pequeno, quanto maior o ângulo “ θ ”, maior será o “ $\sin \alpha$ ”, cujo máximo é atingido nas vizinhanças de “ $\theta = 90^\circ$ ”. Com auxílio de um equipamento denominado Ceilômetro (ou Lidar), no início do trabalho de pesquisa foi possível fazer uma leitura das alturas de nuvens que passavam sobre o

laboratório (na vertical diretamente acima do feixe de laser propagado pelo equipamento) ao lado de onde foi instalada a câmera e a estação solarimétrica. Pode-se dizer que as alturas mais baixas se situavam, aproximadamente, em torno de 1000m de altura. Um cálculo simples a partir desse dado oferece uma estimativa razoável do ângulo “ α ” considerando uma nuvem a se deslocar a 63km/h, com a ressalva de que essa velocidade é considerada o limite inferior de uma tempestade tropical de acordo com a escala de classificação de Saffir–Simpson (NOAA, 2019). Em 5 segundos, intervalo entre cada fotografia, essa velocidade e altura significariam para a medida do ângulo “ α ” o valor de 0,0873 radianos, aproximadamente. A simplificação da função seno para pequenos valores de entrada, tomando-se apenas o primeiro termo “ x ” da série de Taylor em comparação com um valor mais preciso resultaria nesse ponto em uma diferença em termos absolutos de $\sim 1,11 \times 10^{-4}$ rad, desprezível para essa estimativa particular, justificando assim a aproximação para apenas o primeiro termo “ x ” dessa série. Caso fosse considerado o intervalo de 10s, a aproximação ainda seria razoável. Desse modo, para pequenos valores de “ α ” a equação 9 pode ser reescrita da seguinte forma:

$$\frac{\Delta S}{h} \approx \alpha / (\sin \theta \sin(\theta + \alpha)) \quad \text{Eq. 11}$$

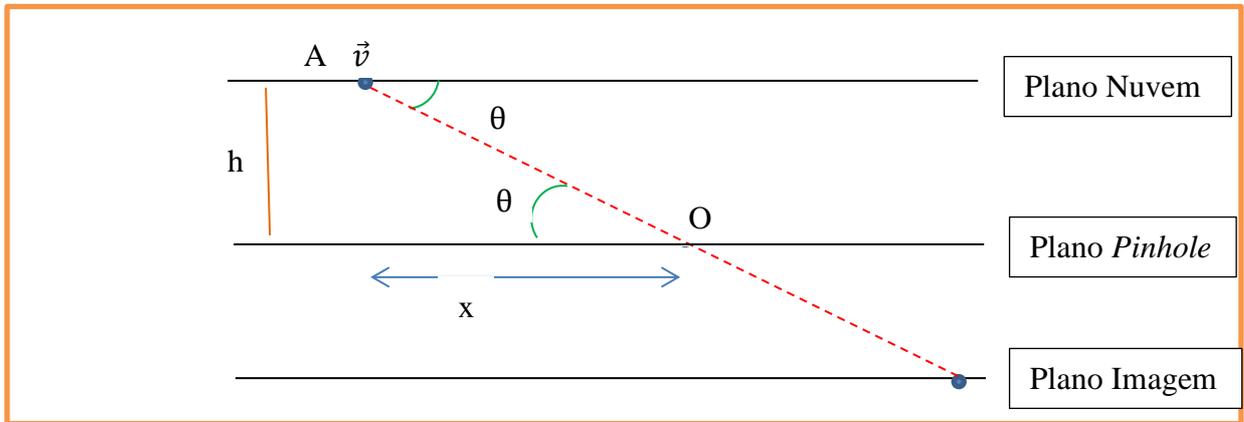
Isolando o parâmetro “ α ”, tem-se que:

$$\alpha \approx \frac{\Delta S}{h} \sin \theta \sin(\theta + \alpha) \quad \text{Eq. 12}$$

A interpretação desta última expressão é útil na medida em que revela o seguinte fato: quanto menor for o ângulo de elevação (horizontal) “ θ ”, menor será o deslocamento angular “ α ”, ao que fará o ponto da nuvem aparentar pela imagem estar parado no limite do horizonte ($\theta \approx 0$). Também é possível inferir que nuvens, mesmo deslocando-se em velocidades constantes sobre um plano e em direção ao centro da imagem, aparentarão deslocar-se com velocidade variável.

Se, ao invés do deslocamento “ ΔS ”, cálculos com o mesmo propósito forem realizados em termos da velocidade pelo uso da ferramenta de cálculo diferencial, os resultados serão concordantes. A próxima ilustração, Figura 33, serve como base ilustrativa aos cálculos desenvolvidos para essa abordagem.

Figura 33 – Velocidade radial de um ponto na nuvem e sua imagem em forma de pixel



É possível estabelecer a seguinte relação matemática:

$$\tan \theta = \frac{h}{x} \quad \text{Eq. 13}$$

Em que “h” é a altura, ou seja, a distância do plano da nuvem ao plano (*Pinhole*) da abertura, “ θ ” o ângulo de elevação e “x” a distância da projeção ortogonal do ponto “A” da nuvem (sobre o plano *Pinhole*) ao ponto “O”.

A expressão anterior resulta na seguinte igualdade:

$$x = \frac{h}{\tan \theta} = h \cot \theta \quad \text{Eq. 14}$$

Através de derivadas parciais (em relação ao tempo), pode-se reescrever assim:

$$\frac{\partial}{\partial t}(x) = \frac{\partial}{\partial t}(h \cot \theta) \quad \text{Eq. 15}$$

Considerando que foi postulada a altura constante da nuvem em um certo intervalo de tempo, tem-se que:

$$\dot{x} = - \frac{h}{(\sin \theta)^2} \dot{\theta} \quad \text{Eq. 16}$$

Então obtém-se a próxima relação que lembra a equação 11 considerando naquela expressão “ α ” muito pequeno e a divisão de ambos os membros pelo intervalo de tempo:

$$\frac{\dot{x}}{h} = - \frac{1}{(\sin \theta)^2} \dot{\theta} \quad \text{Eq. 17}$$

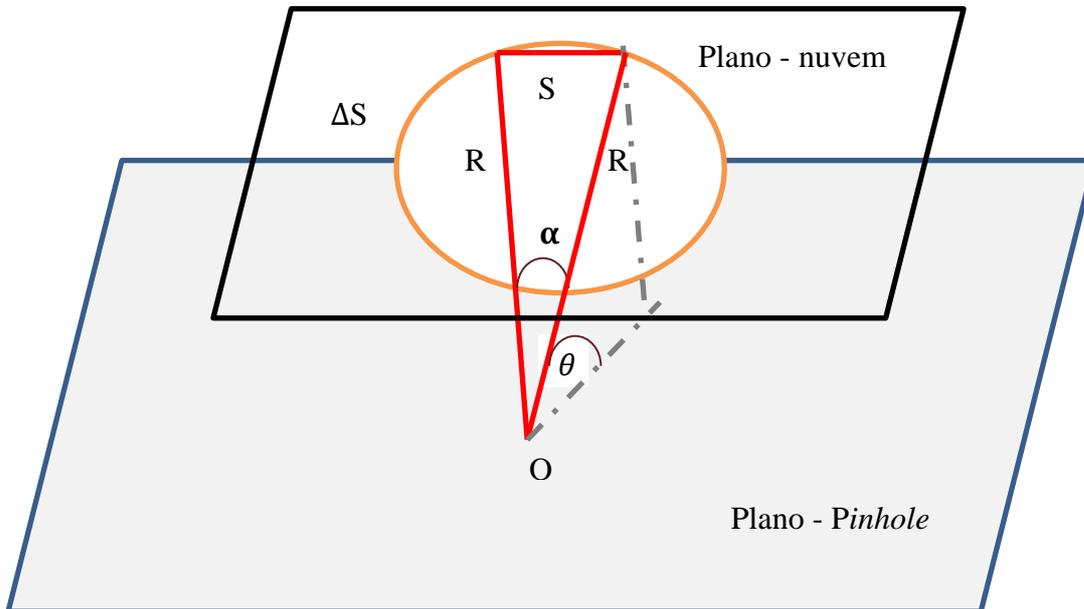
O sinal negativo indica na expressão anterior que a velocidade do pixel “ $\dot{\theta}$ ” na imagem é positiva na medida em que “ \dot{x} ” é negativa, ou seja, quando o espaço “ x ” diminui. Indica que o pixel se aproxima da origem e o ponto correspondente da nuvem se aproxima da vertical sobre o centro da imagem com aumento da magnitude de “ θ ”. Equivale a dizer que esse sinal representa o crescimento do ângulo “ θ ” com a aproximação da nuvem e a consequente diminuição de “ x ”.

Considerando a premissa inicial de que a velocidade da nuvem é constante (\dot{x} constante) em certo intervalo temporal, isolar a variável “ $\dot{\theta}$ ” é útil, pois significa que a velocidade dos pixels diminui mesmo que a nuvem esteja em velocidade, causando aparente paralisação em baixas elevações ($\theta \approx 0$). A expressão a seguir pode ser usada para melhor compreender o efeito (radial) do afastamento dos pixels do centro da imagem, observado que, por hipótese, o termo \dot{x}/h é considerado uma constante.

$$\dot{\theta} = - \frac{\dot{x}}{h} (\sin \theta)^2 \quad \text{Eq. 18}$$

Por outro lado, se o vetor velocidade tiver componente tangencial (em relação ao centro da imagem projetado sobre o plano da nuvem), os resultados serão diferentes. Seja considerada a Figura 34 que ilustra esta situação. Podemos ver um cone reto cujo vértice encontra-se no orifício do plano *Pinhole* de entrada dos raios luminosos. O comprimento da geratriz assume o valor “ R ” e a altura do cone “ h ” é a própria distância ao plano onde encontra-se a nuvem. Nesse caso, será importante calcular qual seria a velocidade de um pixel nessa circunstância em que o respectivo ponto sobre a nuvem descreve uma trajetória tangencial em relação à projeção do ponto “ O ” (*Pinhole*) sobre o plano da nuvem.

Figura 34 – Representação da imagem do deslocamento tangencial de um ponto na nuvem



Pela lei dos cossenos, pode-se escrever que:

$$S^2 = R^2 + R^2 - 2R^2 \cos \alpha \quad \text{Eq. 19}$$

Ou seja,

$$S^2 = 2R^2 (1 - \cos \alpha) \quad \text{Eq. 20}$$

Tomando-se as derivadas parciais em relação ao tempo:

$$\frac{\partial}{\partial t}(S^2) = \frac{\partial}{\partial t}(2R^2 (1 - \cos \alpha)) \quad \text{Eq. 21}$$

O que resulta em:

$$S\dot{S} = R^2 (\sin \alpha) \dot{\alpha} \quad \text{Eq. 22}$$

De outro lado, a lei dos senos aplicada para o triângulo em vermelho da figura estabelece que:

$$\frac{S}{\sin \alpha} = \frac{R}{\sin \gamma} \quad \text{Eq. 23}$$

Em que γ é um dos ângulos da base do triângulo isósceles em vermelho. Visto que $\gamma = 90^\circ - \frac{\alpha}{2}$, tem-se, após substituições, a seguinte relação:

$$\frac{\dot{S}}{\cos(\frac{\alpha}{2})} = R\dot{\alpha} \quad \text{Eq. 24}$$

Mas $\sin \theta = h/R$, de modo que a expressão anterior torna-se:

$$\frac{\dot{S}}{\cos(\frac{\alpha}{2})} = \frac{h}{\sin \theta} \dot{\alpha} \quad \text{Eq. 25}$$

Ou

$$\frac{\dot{S}}{h} = \frac{\cos(\frac{\alpha}{2})}{\sin \theta} \dot{\alpha} \quad \text{Eq. 26}$$

Em que o fator $(1/h)$ volta a surgir como um fator normalizante da velocidade, nesse caso da velocidade linear tangencial de um ponto da nuvem.

E a velocidade (tangencial) dos pixels da imagem pode ser escrita segundo a expressão:

$$\dot{\alpha} = \frac{\dot{S}}{h} \frac{\sin \theta}{\cos(\frac{\alpha}{2})} \quad \text{Eq. 27}$$

A expressão anterior apresenta também, ainda que não obedeça a uma variação quadrática como no caso radial, uma diminuição da velocidade tangencial dos pixels na medida em que há um afastamento do centro da imagem. Também é oportuno observar que o termo \dot{S}/h foi assumido representar uma constante dentro de um intervalo de tempo.

O fator $\cos(\alpha/2)$ que aparece acima poderá ser negligenciado – por assumir um valor unitário com $\alpha/2$ suficientemente pequeno - desde que a altura, a velocidade e o intervalo de tempo entre as medições permitam a aproximação.

A partir das expressões matemáticas anteriores, é oportuno observar que em uma situação real onde as nuvens não ficam confinadas em planos e, portanto, há movimentação vertical, onde há variação das velocidades lineares e onde, além de tudo, há modificação de suas formas – incluindo tanto a dissipação quanto a formação de nebulosidade –, o problema ganha, sem dúvida, ares bem maiores de complexidade. De toda forma, para intervalos de tempo curtos o suficiente, as aproximações ou suposições realizadas podem vir a representar soluções dentro desse limite temporal e dentro de uma expectativa de desvio dos valores reais. No entanto, desconsiderar o fato de as velocidades dos pixels - e suas respectivas componentes radiais e tangenciais - sofrer alterações com a posição dos mesmos em relação ao centro da imagem pode levar a medições bastante imprecisas ou elevar significativamente a relação sinal/ruído na medida em que as medições se aproximam das bordas de uma imagem, onde ambas as componentes assumem valores mínimos. É importante observar também que em ambas as situações de deslocamento, radial ou tangencial, a velocidade aparente em pixels diminui com o aumento da altura da nuvem. Uma nuvem alta, mesmo que esteja eventualmente mais rápida que outras nuvens em planos mais baixos, pode aparentar deslocar-se mais lentamente.

A Tabela 3 foi calculada para avaliação do que representam as atenuações em termos das velocidades aparentes no deslocamento dos pixels segundo uma câmera olhando para cima na vertical.

Tabela 3 – Atenuação das velocidades radiais e tangenciais de pixels

Elevação $\theta(^{\circ})$	<i>Vel. Radial</i> Atenuação(%)	<i>Vel. Tangencial</i> Atenuação(%)
80	3,02	1,52
70	11,70	6,03
60	25,00	13,39
50	41,32	23,40
40	58,68	35,72
30	75,00	50,00
15	93,30	74,12

Uma das consequências das diferentes atenuações radial/tangencial nas velocidades dos pixels é que, se um vetor velocidade possuir as mesmas intensidades em ambas as direções (radial e tangencial), a 15° de elevação, como exemplo, a velocidade tangencial será percebida como 3,86 vezes maior que a radial.

Outro aspecto que merece ser comentado é o fato de que a maior visibilidade da imagem pela escolha de uma câmera que aumente o campo de visão “FOV” – denominação da língua inglesa – tem uma contrapartida negativa: a maior distorção das velocidades dos pixels nas periferias da imagem.

Pelo exposto, nota-se que para um vetor velocidade ser medido na imagem sem as distorções provocadas pela elevação, faz-se necessário corrigir suas componentes (radial e tangencial) de acordo com as expressões já obtidas. Então uma componente radial precisaria ser multiplicada pelo fator:

$$F_{adj_R} = \frac{1}{(\sin \theta)^2} \quad \text{Eq. 28}$$

E a componente tangencial seria corrigida pelo fator:

$$F_{adj_T} = \frac{\cos(\frac{\alpha}{2})}{\sin \theta} \xrightarrow{\alpha/2 \cong 0} \frac{1}{\sin \theta} \quad \text{Eq. 29}$$

Porém, como as componentes dos vetores de velocidade medidas na imagem não se encontram necessariamente para cada ponto nas direções radial e tangencial do movimento, é indispensável uma mudança de coordenadas para colocar a origem dos eixos no centro da imagem e os eixos ortogonais paralelos às componentes tangencial e radial. Isso para cada ponto da imagem em que se pretenda medir a velocidade (em pixels por unidade de tempo) de forma corrigida. Após a operação de translação, em que as coordenadas do centro da imagem (x_c, y_c) são subtraídas de todos os pontos da imagem de interesse (x_i, y_i), é realizada uma rotação para cada ponto em que se pretenda corrigir a velocidade em pixels. O ângulo de rotação beta é determinado pelas novas coordenadas desses pontos após a translação, de modo que:

$$\beta_i = \text{arc tan} \frac{y_i - y_c}{x_i - x_c} \quad \text{Eq. 30}$$

Após o cálculo do ângulo de rotação para cada ponto, as componentes ortogonais dos vetores velocidade, v_x e v_y , originalmente calculadas pelo algoritmo de fluxo ótico esparsos serão convertidas para as novas coordenadas de modo a serem encontradas as componentes radial e tangencial ao ponto de interesse cuja rotação de eixos associada é determinada pelo produto matricial:

$$\begin{bmatrix} v'_x \\ v'_y \end{bmatrix} = \begin{bmatrix} \cos \beta_i & \sin \beta_i \\ -\sin \beta_i & \cos \beta_i \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} \quad \text{Eq. 31}$$

Em seguida foram corrigidos os vetores radial e tangencial de acordo com a operação abaixo:

$$\begin{bmatrix} v''_x \\ v''_y \end{bmatrix} = \left(\frac{1}{\cos \theta_i} \right) \begin{bmatrix} \frac{1}{\cos \theta_i} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v'_x \\ v'_y \end{bmatrix} \quad \text{Eq. 32}$$

Em que agora θ_i é o ângulo zenital correspondente – em função do ângulo máximo de abertura da câmera e a resolução da imagem em número de pixels - à distância do centro da imagem ao pixel cujo vetor de velocidade está a ser corrigido. Em termos matemáticos, tem-se para a câmera escolhida com campo de visão de 180° que

$$\theta_i = \pi \left(\frac{\text{DistEuclid}((x_c, y_c), (x_i, y_i))}{\text{Largura da imagem}} \right) \quad \text{Eq. 33}$$

E então realizando a operação inversa de rotação através da matriz quadrada transposta à matriz da Eq. 31, obtém-se as novas componentes v_x e v_y corrigidas a partir das componentes v''_x e v''_y .

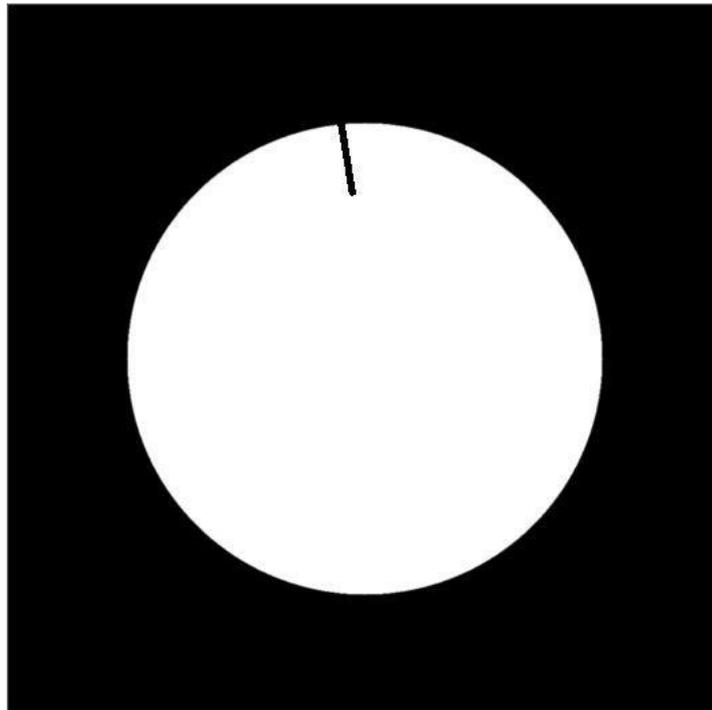
Após análise do conjunto das equações anteriores nota-se, de acordo com os postulados iniciais estabelecidos, que o cálculo das posições subsequentes de um determinado pixel pode ser realizado através de um algoritmo iterativo. A cada iteração um novo (e suficientemente pequeno) deslocamento deverá ser computado baseado na velocidade aparente do último pixel levando-se em consideração a distância desse pixel ao centro da imagem. Isso será feito sucessivamente durante o intervalo total de tempo ou variação total de ângulo de interesse.

O fato do fluxo ótico denso (pelo algoritmo de Farnback (Farneback, 2003)) subdimensionar as magnitudes das velocidades em regiões muito homogêneas que ocorrem normalmente no interior de nuvens, o que será discutido posteriormente, além do tempo de

processamento gasto para calcular a velocidade de todos os pixels de uma imagem, levou este trabalho por escolher realizar os cálculos de velocidade por meio do fluxo esparso. No entanto, os cálculos que envolvem as correções das velocidades dos pixels continuam válidos. Foi justamente a comparação das velocidades nas bordas periféricas da imagem que chamou a atenção para a sistemática diminuição das velocidades nesses arredores, o que suscitou os cálculos anteriores sobre velocidade aparente. Então, sem dúvida, essa foi uma das contribuições do estudo via fluxo ótico denso através do algoritmo selecionado de Farneback.

Os resultados anteriores mostraram que, quanto mais longe do centro da imagem, mais as velocidades são atenuadas, de modo que foi criada uma máscara, Figura 35, para limitar na análise de fluxo ótico esparso a seleção dos pontos dentro da área branca da próxima imagem cuja elevação não é menor que 30° , de modo a não permitir que o ruído (de quantização/detecção) seja significativo quando comparado a medidas atenuadas.

Figura 35 – Máscara para aplicação do algoritmo FAST

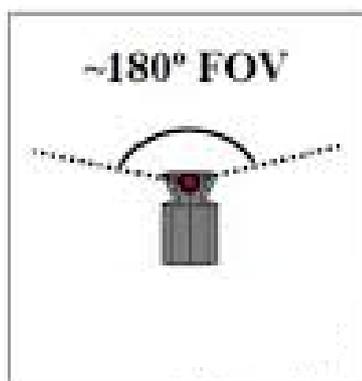


3.5 Calibragem da câmera *fisheye*

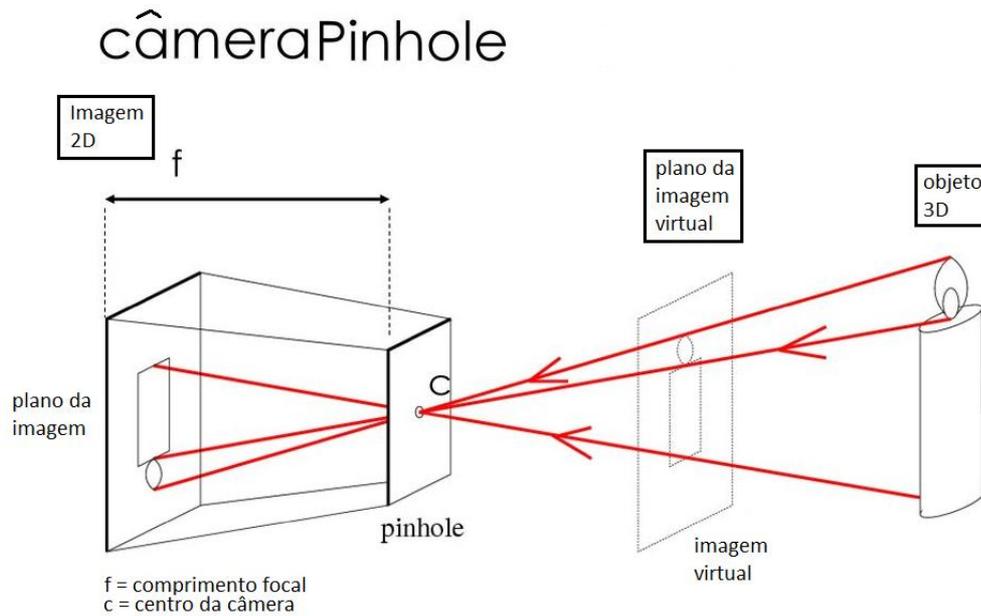
A câmera escolhida tem lente do tipo olho de peixe (*fisheye*) e apesar de exibir uma ampla imagem do hemisfério quando considerada a totalidade de sua abertura ou campo de visão (FOV) de 180° , Figura 36, apresenta distorções típicas desse tipo de lente. Esse

inconveniente representa erros pelo deslocamento de pixels da imagem real quando submetida à projeção no plano dos pixels dessa câmera. Isso significa, por exemplo, distorcer a posição real do Sol, modificando os ângulos que caracterizam a localização do ponto central do disco solar: azimute e elevação. Em outras palavras, em função das coordenadas (u,v) do Sol na imagem, sua posição na câmera poderá estar significativamente afastada da posição determinada pelo modelo matemático escolhido que descreve seus ângulos de localização. Na realidade, a posição dos pixels relativos ao Sol e demais objetos presentes na cena contida na imagem será função dos assim chamados parâmetros intrínsecos e extrínsecos da câmera. Os parâmetros intrínsecos, em termos gerais, são aqueles que se relacionam ao modelo ótico da câmera e à sua construção física que compreende características como lentes, espelhos, comprimento focal, centro da imagem, etc. E são os parâmetros intrínsecos aqueles aqui que detêm maior interesse, dado que fornecerão meios para serem realizadas as correções necessárias na imagem quando se fizer necessário.

Figura 36 – Campo de visão da câmera olho de peixe ou *fisheye*

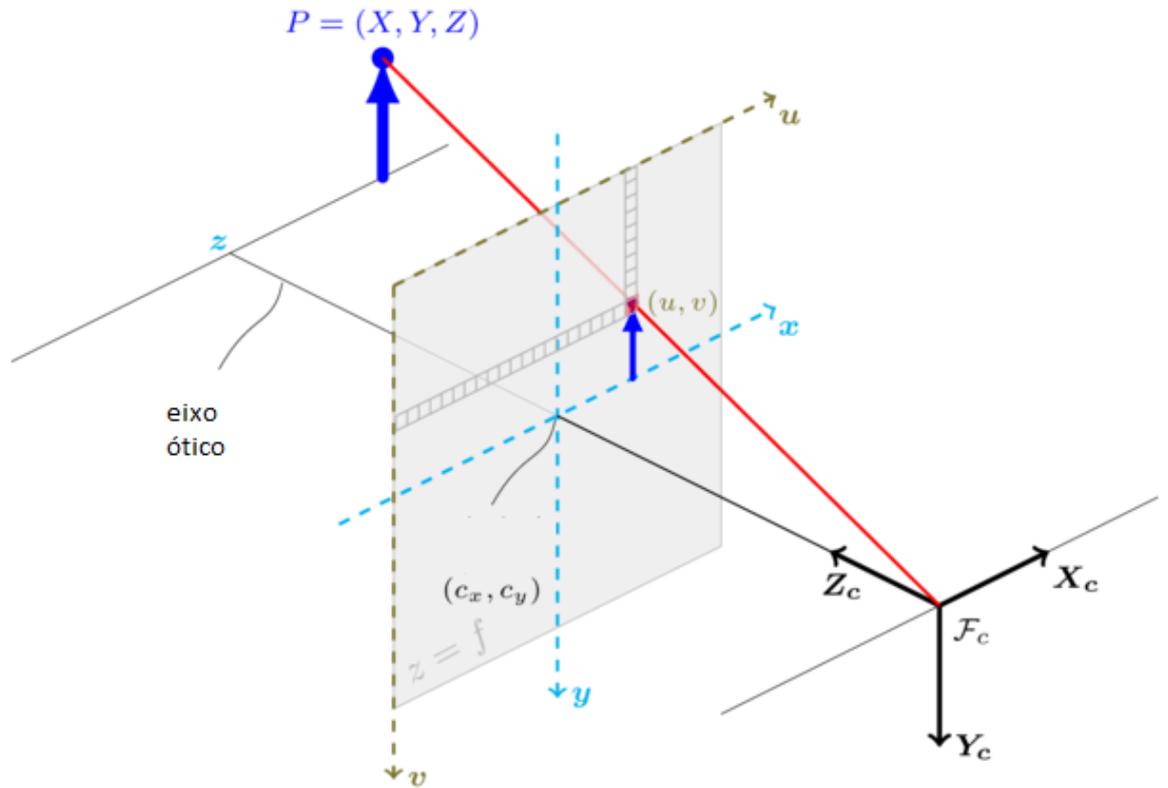


Existem vários modelos desenvolvidos para calibragem das imagens geradas por meio de lentes *fisheye*. Alguns se destacam, dentre eles o modelo desenvolvido por Scaramuzza (SCARAMUZZA, MARTINELLI e SIEGWART, 2006) e o modelo para calibragem de câmeras *fisheye* encontrado na biblioteca de algoritmos para visão computacional OpenCV (BRADSKI, 2000), fundamentado nos trabalhos de Zhang (ZHANG, 2000) e Bouguet (BOUGUET, 2015), amplamente utilizado em trabalhos científicos. Esse último modelo, escolhido para o trabalho, baseia-se na conhecida concepção de câmeras *Pinhole* (ou também chamada de estenopeica), conforme o exemplo da Figura 37

Figura 37 – Câmera *Pinhole*

De acordo com o modelo usado na biblioteca OpenCV, o processo de calibragem para correção das distorções extrai parâmetros intrínsecos e extrínsecos. Entre os primeiros está a assim chamada matriz câmera - formada pelo centro estimado da imagem (c_x, c_y) assim como pelos comprimentos focais f_x e f_y (componentes da projeção focal no plano xy) - e os coeficientes k_i e p_j , respectivamente denominados de coeficientes de distorção radial e tangencial. A próxima imagem, Figura 38, auxilia a compreensão geométrica das grandezas em 3D/2D.

Figura 38 – Representação geométrica em 3D do modelo adotado



Fonte: biblioteca OpenCV

Em termos matemáticos é usual escrever esse modelo através da seguinte notação matricial:

$$s \cdot \mathbf{m}' = \mathbf{A} \cdot [\mathbf{R} | \mathbf{t}] \cdot \mathbf{M}' \quad \text{Eq. 34}$$

Ou, de outra forma,

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} r_{12} r_{13} t_1 \\ r_{21} r_{22} r_{23} t_2 \\ r_{31} r_{32} r_{33} t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad \text{Eq. 35}$$

Em que:

- (X,Y,Z) são as coordenadas de um ponto em 3D;
- (u,v) são as coordenadas da projeção em pixels;

- A é a matriz câmera, ou matriz dos parâmetros intrínsecos;
- (c_x, c_y) é o ponto principal associado ao centro da imagem;
- “ f_x ” e “ f_y ” são os comprimentos focais expressos em unidades de pixels;
- “s” é um fator escalar.

Se uma imagem passa por um fator de escala “s”, isso significa que todos os parâmetros devem ser escalonados (multiplicados ou divididos) pelo mesmo fator. Quanto à matriz “A” de parâmetros intrínsecos, essa poderá ser reutilizada diversas vezes, desde que o comprimento focal seja fixo, considerando eventualmente lentes com *zoom*, o que não é o caso presente. A matriz composta $[R|t]$ é chamada de matriz dos parâmetros extrínsecos, pois envolve tanto uma rotação quanto uma translação da cena em relação à câmera (ou vice-versa).

Pode-se ainda estabelecer as seguinte relações para as coordenadas do ponto (X,Y,Z) , com a condição de que a coordenada z seja diferente de zero:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t \quad \text{Eq. 36}$$

$$x' = x/z \quad \text{Eq. 37}$$

$$y' = y/z \quad \text{Eq. 38}$$

$$u = f_x \cdot x' + C_x \quad \text{Eq. 39}$$

$$v = f_y \cdot y' + C_y \quad \text{Eq. 40}$$

Considerando que as lentes reais normalmente têm algum tipo de distorção, a maioria uma distorção radial e uma leve distorção tangencial, o modelo acima pode ser reescrito assim:

$$x'' = x' \frac{1+k_1r^2+k_2r^4+k_3r^6}{1+k_4r^2+k_5r^4+k_6r^6} + 2p_1x'y' + p_2(r^2 + 2x'^2) \quad \text{Eq. 41}$$

$$y'' = y' \frac{1+k_1r^2+k_2r^4+k_3r^6}{1+k_4r^2+k_5r^4+k_6r^6} + 2p_2x'y' + p_1(r^2 + 2y'^2) \quad \text{Eq. 42}$$

Em que:

$$r^2 = x'^2 + y'^2 \quad \text{Eq. 43}$$

Então, as equações 39 e 40 para correção das distorções assumem agora as expressões:

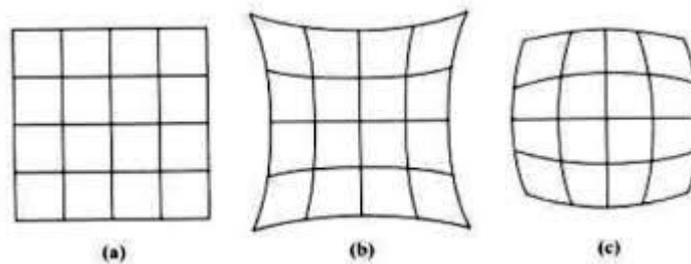
$$u = f_x \cdot x'' + C_x \quad \text{Eq. 44}$$

$$v = f_y \cdot y'' + C_y \quad \text{Eq. 45}$$

Os parâmetros $k_1, k_2, k_3, k_4, k_5, k_6$ são coeficientes de distorção radial ao passo que p_1, p_2 são coeficientes de distorção tangencial.

Coefficientes de ordem mais altas não são considerados na biblioteca OpenCV. As próximas imagens exemplificam dois tipos comuns de distorção: do tipo barril e almofada. Quando barril, o valor típico é tal que $k_1 > 0$. E quando almofada, $k_1 < 0$. A Figura 39 ilustra essas distorções.

Figura 39 – Imagem sem distorção (a), com distorção tipo almofada (b) e barril (c)



3.6 Determinação da posição do Sol

Há vários algoritmos bem estabelecidos para o cálculo da posição solar como o descrito por Rabl (RABL, 1985), porém o algoritmo de Reda (REDA e ANDREAS, 2004) que se utiliza de dados astronômicos, apresentou-se como mais promissor uma vez que oferece considerável precisão por apresentar erros da magnitude de 0,0003 graus.

Com a implementação do algoritmo em linguagem C++ e a colocação dos dados de entrada relativos às variáveis locais, entre as quais latitude e longitude, tornou-se possível determinar seu ângulo azimutal bem como o ângulo zenital (ângulo complementar da elevação).

3.7 Máquina de aprendizado supervisionado SVM

As Máquinas de Vetores de Suporte, ou SVM's - sigla derivada de *Support Vector Machine(s)* - constituem uma técnica de aprendizado supervisionado que se estabeleceu dentro do conceito de aprendizado de máquina (ou ML de *Machine Learning*). É uma ferramenta que tem sido usada largamente e com sucesso em problemas que requerem classificação ou ainda regressão. Quando aplicada a soluções de classificação, essa ML costuma ser associada à sigla SVC, ao passo que é chamada SVR quando se trata de regressão. Tem suas bases fundamentadas na teoria do aprendizado estatístico desenvolvida por Vapnik (VAPNIK, 1999) (VAPNIK, 1998). A aplicação dessa técnica tem apresentado resultados que são comparáveis e muitas vezes superiores aos obtidos por outros algoritmos de aprendizado, como as Redes Neurais Artificiais (RNA's).

As SVM's possuem algumas características que merecem ser ressaltadas:

- Teoria bem definida. As SVM's possuem, dentro das áreas da Matemática e da Estatística, uma formulação teórica bem estabelecida;
- Convexidade da função objetivo. A utilização da SVM significa otimizar uma função quadrática que apresenta apenas um mínimo global. Esta é uma vantagem, por exemplo, sobre redes neurais artificiais, RNA's onde é possível a existência de mínimos locais na função objetivo, alvo da minimização;
- Boa capacidade de generalização. Os classificadores resultantes das SVM's possuem em geral boa capacidade de generalização, considerando que a eficiência de um preditor é medida através dos dados que não pertencem ao conjunto de treinamento. Desse modo evita-se o *overfitting*, situação em que o preditor torna-se bastante especializado no conjunto de treinamento, porém apresenta baixo desempenho diante de novos padrões;
- Robustez em grandes dimensões. As SVM's são robustas quando aplicadas a objetos de grandes dimensões como imagens, por exemplo. Outros tipos de

classificadores comumente geram *overfitting* na presença de dados dessa natureza.

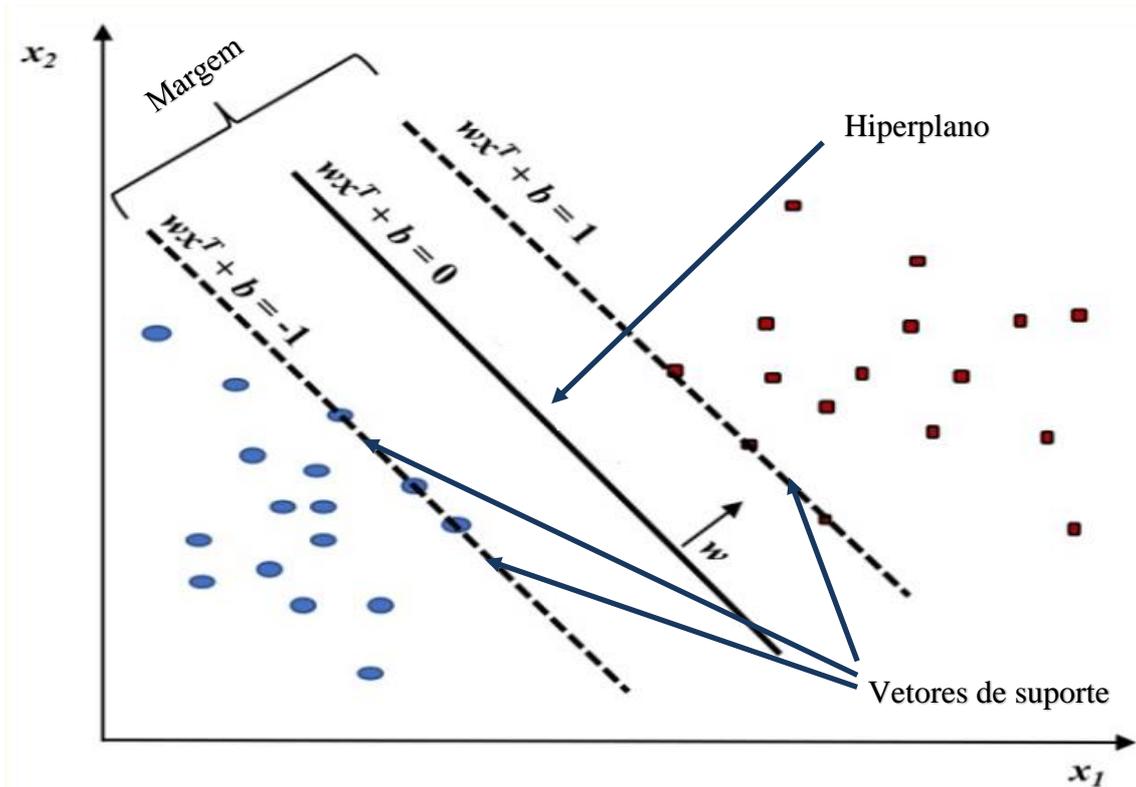
Basicamente a SVM é uma máquina linear (HAYKIN, 2019). SVM's lineares são aplicáveis na classificação de dados linearmente separáveis, como se pode ver na Figura 40. Nesse caso o algoritmo busca um hiperplano que em duas dimensões é representado pela reta em azul da Figura 40. Então por meio do hiperplano obtém-se uma separação que será função dos parâmetros de entrada do algoritmo associado à penalidade relativa às classificações errôneas. As linhas pontilhadas estão dentro de uma margem (de distância “d”) configurável do hiperplano através da parametrização de entrada da SVM de modo a ser possível estabelecer um afastamento que assegure a melhor taxa de acertos na separação. Essas linhas pontilhadas representam especialmente os vetores de suporte de forma tal que maximizam a margem “d” entre o hiperplano e as duas classes representadas em vermelho e azul. Enquanto os círculos vermelhos e azuis significam os exemplos da fase de treinamento, as cruces (vermelhas e azuis) representam a fase de testes. O ajuste dos parâmetros de entrada é feito normalmente através de métodos heurísticos ou guias para otimização do modelo SVM (C.W., C.C., & C.J, 2019).

Porém, há muitos casos em que não é possível dividir satisfatoriamente os dados de treinamento por um hiperplano, quando uma fronteira ou superfície curva passaria a ser uma solução mais adequada. Nesse caso as SVM's lidam com esse problema através do mapeamento do conjunto de treinamento em seu espaço original, dito espaço de entradas, para um espaço de dimensão maior, denominado espaço de características. A Figura 41 ilustra essa estratégia realizada por meio de uma função “ Φ ” em que é possível, após o mapeamento, realizar a separação através de um hiperplano. Essa técnica é fundamentada no teorema de Cover (HAYKIN, 2019). O teorema estabelece que, considerando um conjunto “X” no espaço de entradas, “X” poderá ser transformado no espaço de características “ Ω ” em que com alta probabilidade os dados serão linearmente separáveis, desde que duas condições sejam satisfeitas. A primeira é que a transformação seja não-linear. E a segunda é que o espaço de características tenha uma dimensão suficientemente alta.

Dessa técnica de mapeamento decorre o que vem a ser chamado de *kernel* (ou núcleo) K de uma transformação. Um *kernel* K é uma função que recebe dois pontos x_i e x_j do espaço de entrada e efetua o produto interno no espaço de características: $K(x_i ; x_j) = \Phi(x_i) \cdot \Phi(x_j)$. Entre os núcleos mais utilizados estão o polinomial, o sigmoidal e o gaussiano, também conhecido como de base radial (ou RBF de *radial basis function*). Para este trabalho o núcleo

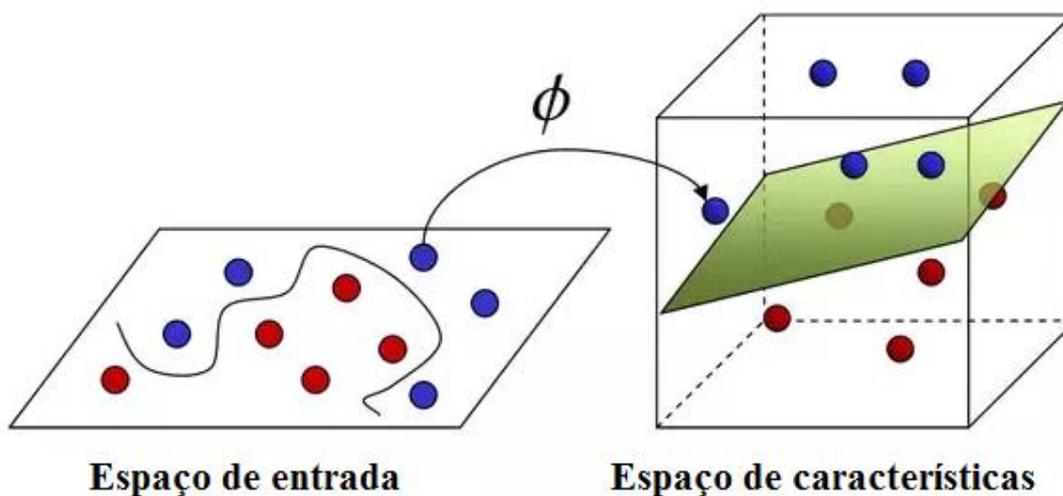
RBF foi escolhido pelos resultados satisfatórios que apresentou. A biblioteca utilizada para implementar a SVM, LIBSVM (CHANG e LIN, 2011), também incorpora um algoritmo (FAN, CHEN e LIN, 2005) do tipo SMO (sequential minimal optimization) (PLATT, 1998) para melhor eficiência.

Figura 40 – SVM linear com dados linearmente separáveis



Fonte: adaptado de (HUANG et al, 2018)

Figura 41 – Mapeamento do espaço de entrada no espaço de características



Fonte: adaptado de (NALINI e PALANIVEL, 2015)

3.8 Segmentação de nuvens

Segmentar as nuvens em uma imagem é um processo que desempenha um papel de extrema relevância para o cálculo das irradiâncias, visto que tanto a oclusão (ou não) do Sol quanto a fração de cobertura das nuvens pode alterar significativamente a radiação solar que chega ao solo.

A abordagem para segmentação nesse caso, porém, difere totalmente do que foi realizado para o Sol, que segue um caminho no céu de acordo com modelos que têm graus de incerteza muito reduzidos para os propósitos desse trabalho. As nuvens, por outro lado, possuem trajetórias cujo comportamento se destaca pela natureza aleatória e estocástica, dificultando sobremaneira a determinação de suas posições e as regiões por elas ocupadas em uma imagem. Além disso, é preciso observar que as formações nebulosas não apenas variam em forma e tamanho, mas também nas cores desde o amanhecer ao pôr do Sol. Sem esquecer ainda que seus vetores de velocidade também sofrem alterações ao longo do dia. E, como se toda essa variabilidade já não constituísse um forte obstáculo, o céu, mesmo na ausência de nuvens, também é captado por câmeras em tonalidades diferentes entre os primeiros raios de Sol até os últimos instantes de luz solar. Nesse cenário de incertezas, dentre as soluções propostas por diversos trabalhos de pesquisa que obtiveram maior êxito em segmentação, destacam-se aquelas que recorreram às máquinas de aprendizado (ML's), entre elas as RNA's – redes neurais artificiais.

Assim como em outros trabalhos, decidiu-se separar os pixels do céu em dois grupos ou classes: céu ou nuvem. Para essa tarefa há um leque extenso de ferramentas de classificação em termos de máquinas de aprendizado/redes neurais. Foram testados alguns algoritmos de aprendizado dentre os quais MLP, LDA, DBN - *Deep Belief Network* e SVC, sendo o último aquele que apresentou mais vantagens em termos de implementação e integração ao programa já existente escrito em C#/C++, além das características já destacadas no tópico anterior. Em particular, foi usada a biblioteca LIBSVM (CHANG, 2011), seleção que levou em consideração o grande número de usuários e o suporte disponível em *sites* para obter-se o melhor e mais rápido funcionamento dessa biblioteca.

A revisão de literatura apresentou alguns atributos ou características que têm sido reutilizados em vários trabalhos. Em especial, considerados os espaços de cores RGB e HSV, as expressões “(B-R)”, “R/B”, “(B-R)/(B+R)” e “S” têm demonstrado obter em conjunto ou associadas a algumas técnicas, resultados bem sucedidos. Essas variáveis representaram o ponto de partida e também o início de um processo de busca por outras variáveis que pudessem

caracterizar tão bem ou melhor, se possível, as classes nuvens e céu. O objetivo, na realidade, é poder diferenciar cada classe do ponto de vista matemático/numérico tanto quanto houver possibilidade. É algo como garimpar dados provenientes dos canais dos espaços de cores dos pixels para poder separar as classes com maior facilidade. Essa busca demandou o uso de métricas para selecionar os parâmetros (ou características) mais promissores. Três métricas assumiram importante papel na avaliação das características nuvens/céu, a saber: o coeficiente de correlação de Pearson ($\rho_{X,Y}$), a avaliação (por meio da entropia) do ganho de informação de um atributo/característica da classe (*Ginfo*) e a avaliação da razão de ganho (relativa à entropia) de um atributo/característica da classe (*RGinfo*). O coeficiente de correlação de Pearson, que foi utilizado várias vezes ao longo deste trabalho, pode ser interpretado como uma medida da correlação linear entre duas variáveis. Assume valores dentro do intervalo fechado $[-1;1]$, o que significa que no máximo seria igual a “+1”, quando se diz haver correlação positiva perfeita ou que no mínimo seria igual a “-1”, caso em que haveria correlação negativa perfeita. Um valor nulo significaria não haver correlação linear entre as duas variáveis. No que diz respeito ao coeficiente de Pearson aplicado aos atributos de classificação por meio da Eq. 46, a variável (X) representa o valor do atributo em análise e a outra variável (Y) representa o valor numérico atribuído à classe a que pertence o pixel, sendo “1” para nuvem e “-1” para não-nuvem. Formalmente, as expressões que seguem definem essas métricas na ordem em que se encontram, respectivamente:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \quad \text{Eq. 46}$$

$$Ginfo(Classe, Atributo) = H(Classe) - H(Classe|Atributo) \quad \text{Eq. 47}$$

$$RGinfo(Classe, Atributo) = \frac{H(Classe) - H(Classe|Atributo)}{H(Atributo)} \quad \text{Eq. 48}$$

Em que:

- “Cov” representa a covariância entre duas variáveis aleatórias;
- “ σ ” é o desvio padrão da variável aleatória;
- $H(C)$ e $H(C|A)$ são, respectivamente, a entropia e entropia condicional, expressões cujas definições seguem abaixo nessa ordem:

$$H(X) = - \sum_{x \in X} p(x) \cdot \log(p(x)) \quad \text{Eq. 49}$$

$$H(Y|X) = - \sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \cdot \log(p(y|x)) \quad \text{Eq. 50}$$

Muitas expressões foram concebidas para efeito de testes e avaliadas utilizando as três métricas acima como potenciais candidatas a tornarem-se características/atributos em um vetor de classificação de pixels (como nuvens ou céu). Após sucessivas filtragens, sempre buscando um conjunto com métrica mais favorável de modo a obter melhor desempenho da SVC, foram escolhidos 7 atributos:

$$f_0 = \frac{(Cb1-G1)}{(Cr1+G1+0,5)} - \frac{(Cr1-Cb1)}{(Cr1+Cb1+0,5)} \quad \text{Eq. 51}$$

$$f_1 = e^{\left[\frac{(Cb1-R1)}{(Cr1+B1+0,5)} + e^{\frac{(Cb1-a1)}{(Cr1+R1+0,5)}} \right]} \quad \text{Eq. 52}$$

$$f_2 = \text{Dist}(\text{PtoSol}(x_s; y_s), \text{PtoPixel}(x; y)) \quad \text{Eq. 53}$$

$$f_3 = \frac{(Cb-G)}{(Cr+G+0,5)} - \frac{(Cr-Cb)}{(Cr+Cb+0,5)} \quad \text{Eq. 54}$$

$$f_4 = \frac{(Cb-G)}{(Cr+0,5)} - \frac{(Cr-Cb)}{(G+0,5)} \quad \text{Eq. 55}$$

$$f_5 = e^{\left[\frac{(Cb-R)}{(Cr+B+0,5)} + e^{\frac{(Cb-a)}{(Cr+R+0,5)}} \right]} \quad \text{Eq. 56}$$

$$f_6 = S \quad \text{Eq. 57}$$

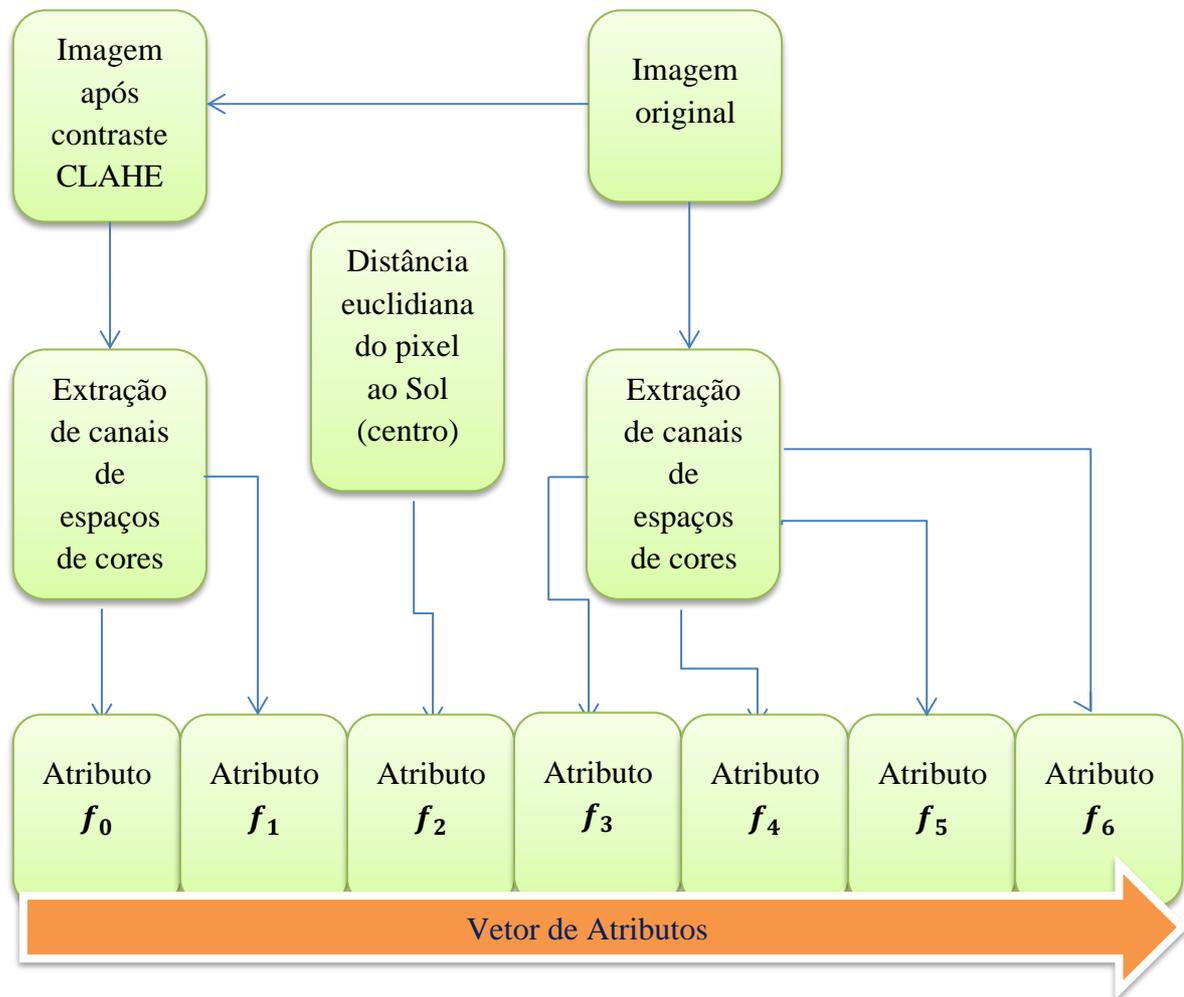
Em que:

- Os parâmetros “R”, “B” e “G” são canais do espaço de cores RGB;
- Os parâmetros “Cr” e “Cb” são canais do espaço de cores YCrCb;
- O parâmetro “a” é um canal do espaço de cores Lab;
- O parâmetro “S” é um canal do espaço de cores HSV;
- O atributo “ f_2 ” representa a distância do pixel ao centro do disco solar;

- O sufixo “1” nas expressões que definem os parâmetros “ f_0 ” e “ f_1 ” indica que foi utilizada a saída da imagem após ser processada pelo algoritmo CLAHE (ZUIDERVELD, 1994), que tem a função de aumentar o contraste da imagem original.

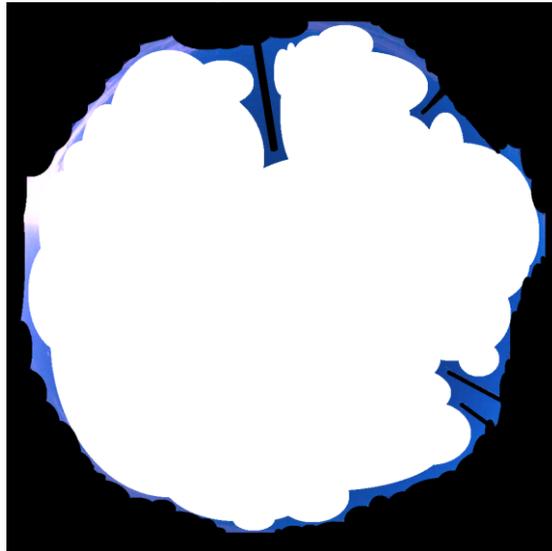
Para ilustrar a construção acima do vetor de atributos em termos de um fluxograma segue a Figura 42:

Figura 42 – Fluxograma para criação do espaço vetorial de segmentação



Para a tarefa de segmentação, objetos da cena como árvores e prédios foram preliminarmente excluídos através de uma máscara de imagem, em que os pontos escuros (pixels de cor preta) têm a função de eliminar os pixels indesejáveis da imagem original. Abaixo, segue a máscara utilizada apresentada através da Figura 43.

Figura 43 – Máscara utilizada para segmentação



3.9 Modelo para estimar as irradiâncias a partir das imagens

Inicialmente, sabe-se por definição que:

$$Irrad_{Total} = (Irrad_{Direta}) \cos \gamma + Irrad_{Difusa} \quad \text{Eq. 58}$$

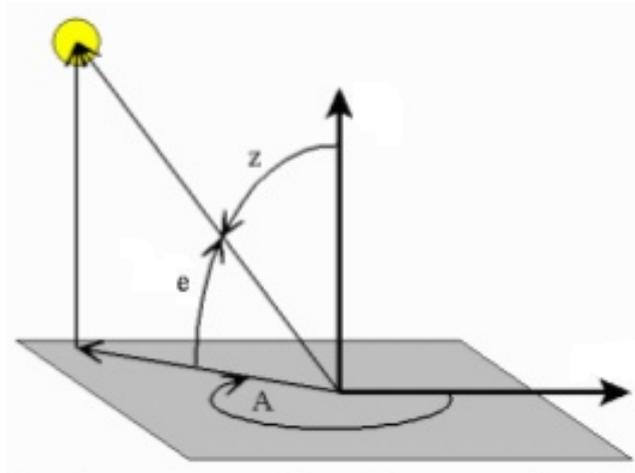
Outra notação equivalente e mais comumente encontrada nos livros-texto para a expressão anterior é:

$$GHI = DNI \cos z + DHI \quad \text{Eq. 59}$$

Em que “ γ ” e “ z ” representam a mesma medida, o ângulo zenital, GHI é a irradiância global total do hemisfério, DNI, a irradiância direta e DHI a irradiância difusa do hemisfério.

No caso deste trabalho, a irradiância difusa poderia ser estimada indiretamente a partir das duas primeiras estimativas. E o ângulo zenital algumas vezes poderá ser devidamente substituído (em uma expressão) pelo ângulo complementar, chamado de elevação solar. A Figura 44, a seguir, ilustra a geometria dos raios solares sob o ponto de vista da irradiância direta, DNI, representada por um vetor com origem no sistema de coordenadas a apontar para o Sol. No cálculo da irradiância total esse termo surge como uma parcela previamente multiplicada pelo fator “ $\cos z$ ” e representa sua componente normal relacionada à projeção sobre a vertical local.

Figura 44 – A radiação direta e sua projeção sobre a vertical local. “A” é o azimute



A segmentação das nuvens em uma imagem desempenhou um caminho essencial para o cálculo das irradiâncias, visto que tanto a oclusão do Sol, parcial ou efetiva, quanto à fração de cobertura das nuvens podem alterar significativamente a radiação solar que chega ao solo. Mas durante a etapa de segmentação, ficou claro que outras medidas que surgiram em meio ao desenvolvimento do algoritmo poderiam trazer informações adicionais que seriam correlacionadas às grandezas que o estudo se propôs a estimar: as irradiâncias total e direta. Criou-se então, para cada imagem, em intervalos regulares de 5s, um espaço amostral de vetores cujos parâmetros de entrada reuniram características na forma de coordenadas segundo a ordem abaixo, a saber:

- Cosseno do ângulo zenital;
- Seno do ângulo zenital;
- Fração de cobertura de nuvens;
- Valor médio dos tons de cinza dos pixels da região sem nuvens;
- Valor médio dos tons de cinza dos pixels da região com nuvens;
- Valor médio dos tons de cinza sobre o disco solar;
- Valor médio do parâmetro f_4 para os pixels da região sem nuvens;
- Valor médio do parâmetro f_4 para os pixels da região com nuvens;
- Valor médio do parâmetro f_4 para os pixels sobre o disco solar;
- Valor médio do parâmetro f_6 para os pixels da região sem nuvens;
- Valor médio do parâmetro f_6 para os pixels da região com nuvens;
- Valor médio do parâmetro f_6 para os pixels sobre o disco solar;
- Valor médio da banda inferior do método de Otsu sobre o disco solar (OT1);

- Valor médio da banda superior do método de Otsu sobre o disco solar (OT2);
- Valor percentual do limite entre classes do método de Otsu sobre o disco solar (p);
- $OT1 \cdot p$;
- $OT2 \cdot (1 - p)$;
- $OT1 \cdot p + OT2 \cdot (1 - p)$;

Há de serem feitas uma série de considerações sobre as características anteriores selecionadas. O cosseno de cada ângulo zenital está diretamente correlacionado com as irradiâncias total e direta. Por outro lado, sabemos que as derivadas das funções cosseno e seno assumem seus valores máximos com diferença de fase de 90° , ou seja, a função cosseno será mais sensível nas proximidades de 90° , o que significa reagir mais rapidamente às variações angulares. Por analogia (complementar), a função seno percebe melhor essas alterações nas proximidades de 0° . Dessa forma ambas foram escolhidas, seno e cosseno do ângulo zenital.

Quanto à fração de cobertura de nuvens, sempre houve através de observações práticas a percepção de que a irradiância difusa, que por sua vez colabora com a irradiância total, sofre grande influência das nuvens, especialmente nuvens brancas que agem como refletores naturais.

A gradação de tons de cinza que difere céu, nuvens e Sol em imagens P&B também pareceu uma escolha natural que viesse a trazer informações complementares sobre as irradiâncias. E uma vez que os tons de cinza das imagens P&B podem carregar informações adicionais, os parâmetros f_4 e f_6 que (anteriormente nesse texto) tão bem caracterizaram a diferença entre céu e nuvem para efeito de classificação dessas regiões, trazem informações suplementares sobre as irradiâncias daquele instante em particular.

E finalmente foi utilizado o método de Otsu (OTSU, 1979) como base de cálculo de diversas características. Trata-se de um algoritmo largamente aplicado – e com diversas vertentes e aperfeiçoamentos - em trabalhos que envolvem segmentação de imagem e que por sua utilidade encontra-se inclusive presente na biblioteca OpenCV. De forma resumida, considerado um dado histograma, o método de Otsu destina-se a encontrar um valor limite que minimiza a variância intra-classe (variância dentro da classe) ao mesmo tempo em que maximiza a variância inter-classe. Suponha-se haver a partir de uma imagem P&B um histograma (Figura 45) extraído com base na frequência de tons de cinza, tons esses exibidos ao longo do eixo horizontal do histograma.

O que o método de Otsu busca é um valor limiar (da coordenada horizontal) que consiga separar o histograma em duas classes, minimizando a variância dentro de cada uma das classes que surgem a partir dessa separação e de acordo com as definições estabelecidas.

Assim, sejam agora as expressões abaixo, onde a equação 60 representa a variância intra-classe definida como a soma ponderada da variância de duas classes:

$$\sigma_w^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t) \quad \text{Eq. 60}$$

Em que

$$\omega_0(t) = \sum_{i=0}^{t-1} p(i) \quad \text{Eq. 61}$$

e

$$\omega_1(t) = \sum_{i=t}^{L-1} p(i) \quad \text{Eq. 62}$$

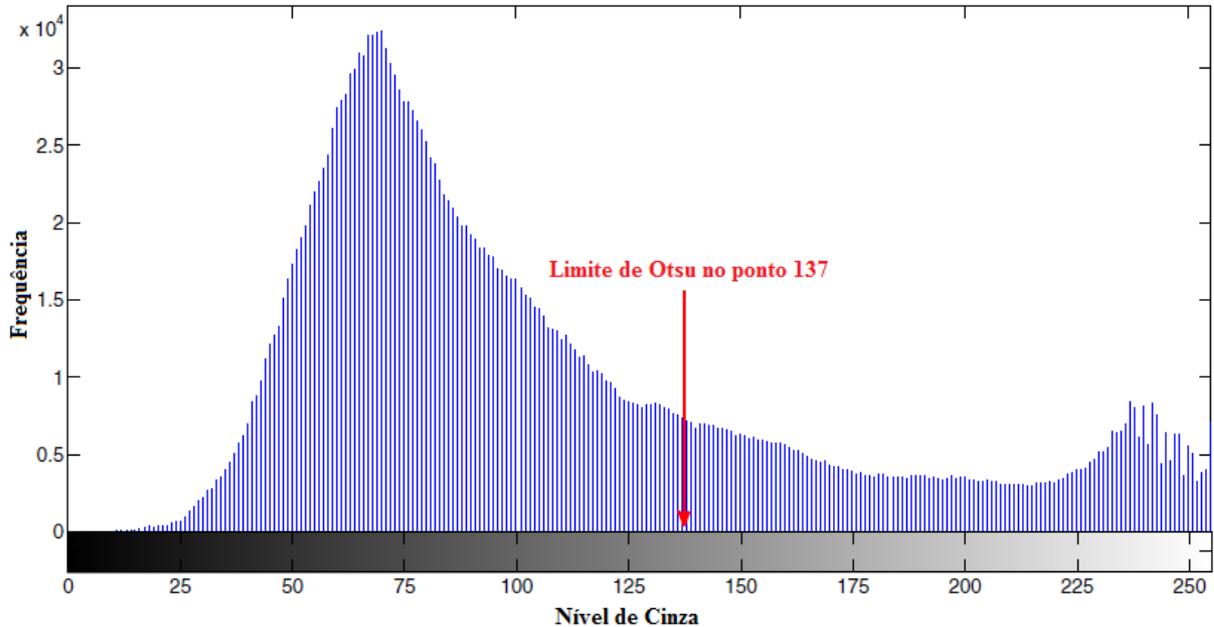
Em que ω_0 e ω_1 são calculados sobre as janelas (*bins*) do histograma dividido através de um ponto limiar “t” sobre o eixo horizontal e $p(i)$ representa a probabilidade do “bin” de índice “i”. Para um determinado índice k, então haverá um valor t_k que será o limiar obtido iterativamente no algoritmo para minimizar “ $\sigma_w^2(t_k)$ ”. Pelo método de Otsu, minimizar a variância intra-classe é equivalente também a maximizar a variância inter-classe.

Entre vários artigos que empregam esse algoritmo, é ilustrativo citar um deles que apresenta uma aplicação prática do método de Otsu como ferramenta para classificação e separação de pixels de um histograma. Percebe-se na Figura 45 do trabalho de HAGGERTY (HAGGERTY *et al.*, 2014) a divisão do histograma em duas partes (inferior e superior em termos do eixo da abscissa) e a localização desse limiar que representa o ponto de máximo da variância inter-classe.

No entanto, é oportuno comentar que neste trabalho as amostras sobre o disco solar são inicialmente ordenadas de forma crescente, pois o objetivo é separar as irradiâncias por

intensidade, ou seja, aquelas supostamente relacionadas à irradiância direta, de intensidade maior, daquelas outras menores que representam algum tipo de cobertura sobre os raios solares.

Figura 45 – Ilustração gráfica do método de Otsu



Fonte: (HAGGERTY *et al.*, 2014)

Desse modo, o parâmetro OT1 representa a média da distribuição sobre o intervalo cujos valores individuais são menores que o limite das regiões obtido por Otsu. O parâmetro OT2 representa a média sobre o segundo intervalo, aquele igual ou superior ao limite de Otsu. Fica claro que $\widehat{OT1} \leq \widehat{OT2}$.

O valor “p” incluído no vetor de características é um quociente calculado através da razão entre quantos pontos do histograma são menores que o limite de Otsu e o total de pontos (da base) do histograma.

As três últimas características, “ $OT1 \cdot p$ ”, “ $OT2 \cdot (1 - p)$ ” e “ $OT1 \cdot p + OT2 \cdot (1 - p)$ ” derivam das características anteriormente definidas. Em resumo, foi concebido um espaço vetorial cujos vetores são formados pelas 18 características acima extraídas de cada imagem.

3.10 Modelo para predições de irradiâncias

Foram encontrados alguns trabalhos durante a revisão bibliográfica que utilizaram em suas metodologias estimativas da velocidade das nuvens com objetivo de predição de irradiâncias. Alguns métodos utilizaram a detecção do sombreamento por meio de sensores na superfície; outros por técnicas de análise da velocidade de deslocamento de pixels. Este trabalho estudou o segundo caso, mais especificamente no que diz respeito à capacidade e complexidade de predição das posições de um pixel representativo de um ponto ou região da nuvem durante sua trajetória em imagens sucessivas. Para o caso ideal não foi considerada a aleatoriedade da direção e velocidade dos ventos, nem a eventual componente da velocidade das nuvens no eixo vertical (componente ascendentes ou descendente) e nem mesmo o fenômeno de formação ou dissipação de nuvens. A importância deste estudo simplificado e de caráter determinístico - no qual a velocidade das nuvens seria constante sobre um plano paralelo ao plano da imagem - mostrou sua importância na medida em que as equações resultantes revelaram uma trajetória complexa na imagem em que a velocidade de um pixel a todo instante é função também da própria posição nesse instante. Como consequência, por exemplo, na medida em que o pixel se afasta do centro da imagem as medições da velocidade tornam-se menos precisas com a desaceleração que ocorre até a velocidade ser virtualmente nula no limite do horizonte. Além disso, revelou a existência de duas componentes diferentes de velocidade (radial e tangencial) na imagem.

Em se tratando do caso não ideal, cujos efeitos foram negligenciados nesse estudo, a complexidade torna-se bem maior, o que justificou uma abordagem estatística de tratamento do problema de predição, ao contrário de uma abordagem com predominância da natureza determinística. Em suma, esta análise orientou a pesquisa de modo a buscar modelos preditivos apoiados em sólidos alicerces estatísticos, o que remeteu ao uso da ML do tipo SVR (SVM aplicada a problemas de regressão) - cujas bases são fundamentadas na teoria do aprendizado estatístico desenvolvida por Vapnik (VAPNIK, 1999) (VAPNIK, 1998) – para elaboração dos modelos preditivos para as irradiâncias total e direta.

Através de alguns dos atributos desenvolvidos no modelo anterior e dos dados das irradiâncias (total e direta) foi criado o vetor de características do modelo de predição. Levando-se em consideração a evolução das características com o passar do tempo, foi construída uma série temporal contendo as irradiâncias e atributos numéricos relativos às fotografias obtidas a

cada 5s. A estrutura do vetor de entrada para a SVR foi então criada com base em uma sequência de parâmetros em 30 intervalos regulares de 1 minuto no passado mais um valor de saída do modelo, GHI ou DNI, em um intervalo futuro (do horizonte de predição) com um passo ou janela de deslocamento de também 1 minuto. Para cada intervalo de 1 minuto na linha do tempo das observações foi calculado o valor médio de cada parâmetro nesse período para representá-lo. Considerando que para cada minuto foram obtidas 12 imagens, resulta que cada parâmetro ficou representado por uma média sobre seu respectivo conjunto de 12 valores. Então cada um dos 30 intervalos foi formatado com os atributos abaixo e ao final dos 30 intervalos foi incluída uma das irradiâncias (GHI ou DNI) no horizonte escolhido de predição como saída, conforme a relação e descrição a seguir:

- Cosseno do ângulo zenital;
- Fração da cobertura de nuvens;
- Valor médio do parâmetro f_4 para os pixels correspondentes às nuvens;
- Valor médio do parâmetro f_4 para os pixels correspondentes ao céu;
- Valor médio do parâmetro f_4 para os pixels correspondentes ao disco solar;
- Valor “ $OT1 \cdot p$ ” definido na seção anterior (3.9);
- Valor “ $OT2 \cdot (1 - p)$ ” definido na seção anterior (3.9);
- Valor “ $OT1 \cdot p + OT2 \cdot (1 - p)$ ” definido na seção anterior (3.9);
- Valor medido da irradiância total
- Valor medido da irradiância direta.
- Inclusão da saída (GHI ou DNI) ao final dos 30 intervalos (com os 10 atributos acima). As irradiâncias, GHI ou DNI, foram colocadas para os instantes futuros de 1, 2, 3, 4, 5, 10, 15 e 30 minutos à frente no tempo relativamente ao mais recente (na linha do tempo) dos 30 intervalos. Ou seja, para cada um dos horizontes de predição, entre 1 e 30 minutos, foi construído um modelo SVR particular considerando o mesmo conjunto no passado desde $t = t_0 - 29$ até $t = t_0$ para cada t_0 ;
- Como resultado, cada vetor de treinamento foi composto por 30 x 10 valores de entrada mais um valor de saída.

Os modelos de regressão foram desenvolvidos por meio da biblioteca “*sklearn.svm.SVR*” existente na linguagem Python. O processo de avaliação dos resultados empregou uma metodologia para treinamento/validação/teste diferente do método de validação

cruzada do modelo anterior (seção 3.9), uma vez que esse modelo reflete uma série temporal. Assim, foi substituída a metodologia de avaliação anterior da ML SVR em que amostras foram escolhidas de forma aleatória. Dessa vez a metodologia de avaliação seguiu a ordem cronológica em que as amostras foram obtidas, de modo que a etapa de testes foi realizada com 1/11 das últimas amostras e o restante foi usado para treinamento/validação. Amostras de imagens que, eventualmente, não encontraram suas respectivas irradiâncias naquele instante foram descartadas para benefício do sincronismo necessário. Os elementos dos vetores de entrada também foram testados diante da eventual presença de valores numéricos falhos tais como NAN (*not a number*).

Para cada um dos modelos de predição via regressão SVR foram extraídos – dentro do conjunto reservado para teste do modelo - os valores das suas respectivas métricas RMSE, MAE e MBE. A tempo, é importante observar que tanto os atributos quanto os valores de saída, foram normalizados para o intervalo [-1;1] antes de serem submetidos ao modelo SVR. Após o treinamento de cada modelo suas respostas de predição foram “de-normalizadas”, ou seja, receberam o processamento inverso ao da normalização inicialmente realizada.

4 RESULTADOS E DISCUSSÃO

Uma série de ferramentas de software através de algoritmos específicos foram utilizados para obter soluções particulares que, reunidas, pudessem contribuir com resultados e soluções em cada meta para alcançar o objetivo final. Assim, desde a coleta de fotografias do céu e dados de irradiância até modelos para máquinas de aprendizado ou então para modelagem da irradiância solar por imagem foram todos implementados ou incorporados por meio de bibliotecas de programação para, por fim, serem finalmente aplicados ao sistema de predição desenvolvido.

4.1 Calibragem da câmera *fisheye*

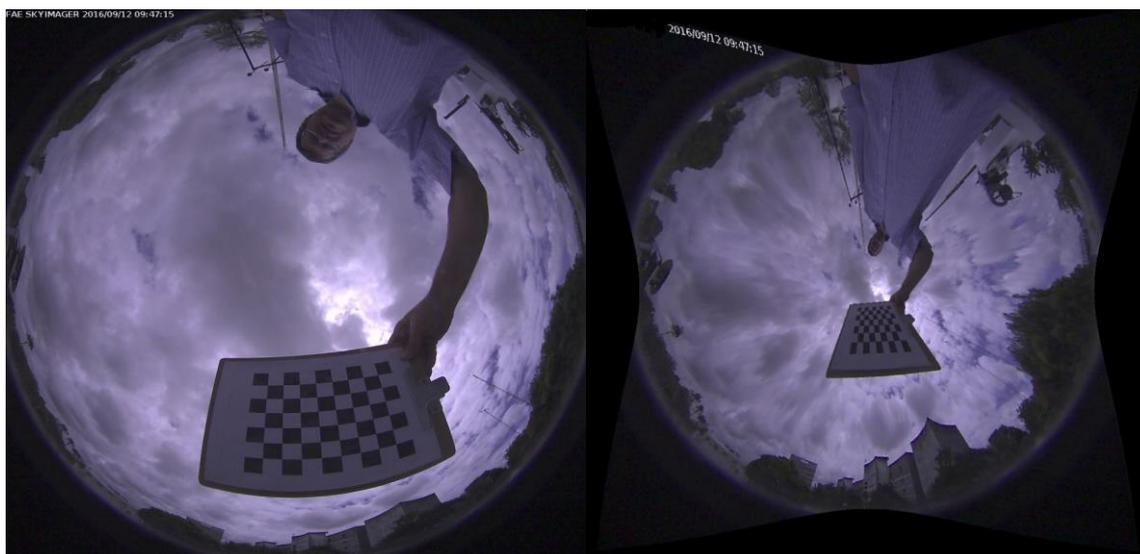
Para determinar os coeficientes do modelo (OpenCV) e ter a capacidade de corrigir as distorções foi necessário realizar o processo de calibragem da câmera *fisheye* em conformidade com as necessidades de dados de entrada do algoritmo que realiza esse procedimento. Dessa forma, imagens na forma de tabuleiro de xadrez como a da Figura 46 foram apresentadas à câmera em diversos ângulos e posições diferentes gerando fotografias que depois foram armazenadas e posteriormente submetidas ao algoritmo apropriado com a formatação necessária através de um software de gerenciamento. Após ter sido reunida uma quantidade suficiente de imagens em um total de 266, foi identificado um subgrupo de 20 imagens em posições variadas que apresentou o menor erro médio quadrático ou MSE igual a 1,04, valor esse obtido pela função do algoritmo (OpenCV) responsável pela calibragem. A partir de então, para uso posterior, foram armazenados em arquivo os valores dos coeficientes de distorção radial e tangencial bem como a matriz de parâmetros intrínsecos.

Figura 46 – Imagem usada para calibrar e extrair parâmetros intrínsecos da câmera



Com essa técnica, imagens puderam ser corrigidas em termos da distorção provocada pela lente *fish-eye*, porém, como se percebe abaixo na Figura 47 após o processamento, há um custo de redução da área útil da imagem, aproximando ou mesmo suprimindo pixels representativos das condições de nebulosidade do céu. Em algumas imagens, por exemplo, foi notada a redução drástica de pequenas áreas de céu rodeadas por nuvens em virtude da aplicação do algoritmo de correção bem como do algoritmo de interpolação de pontos da biblioteca OpenCV. Em virtude do fato de que esse processo como um todo é não linear, parâmetros como a fração de cobertura de nuvens poderiam ser alterados. Por essa razão e também para não diminuir a área útil das imagens, o processo de correção foi suprimido a menos do algoritmo para cálculo da posição do Sol, essencial ao estudo e cujo cálculo não exigiu essa interpolação de ajuste.

Figura 47 – Calibragem da imagem *fisheye* à direita, original à esquerda

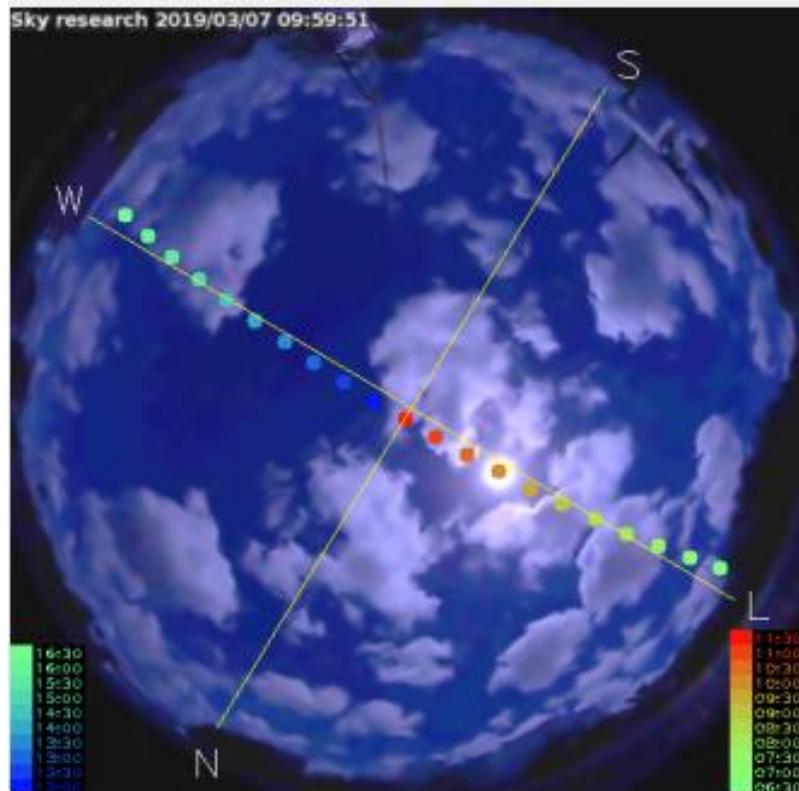


4.2 Determinação da posição do Sol

Uma vez extraídos os parâmetros intrínsecos e coeficientes da câmera, a próxima etapa consistiu em combinar o modelo da câmera com o modelo de Reda (REDA e ANDREAS, 2004) para então obter a posição solar no hemisfério.

Como consequência, conhecidos os ângulos solares e solucionado o problema de correção das distorções da câmera *fisheye*, passou a ser possível encontrar de forma determinística a posição do centro do disco solar bem como projetar a trajetória do Sol ao longo de um dia, conforme ilustra a Figura 48. Esse algoritmo teve importância fundamental para a segmentação do Sol na imagem e dessa forma proporcionar estimativas das irradiâncias.

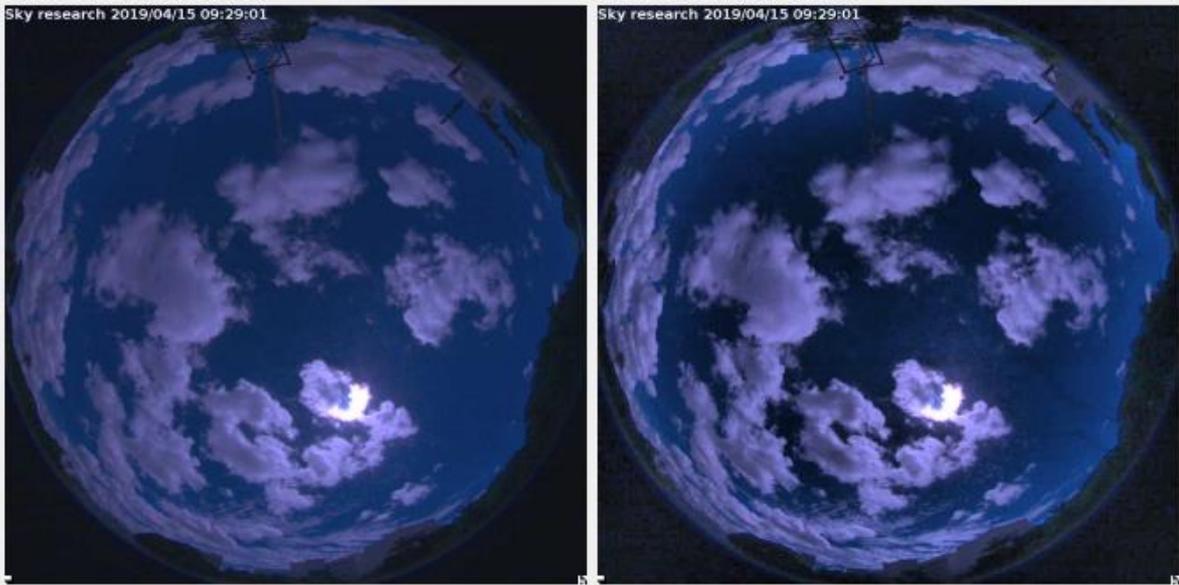
Figura 48 – Determinação da posição e da trajetória solar ao longo de um dia



4.3 Segmentação de nuvens

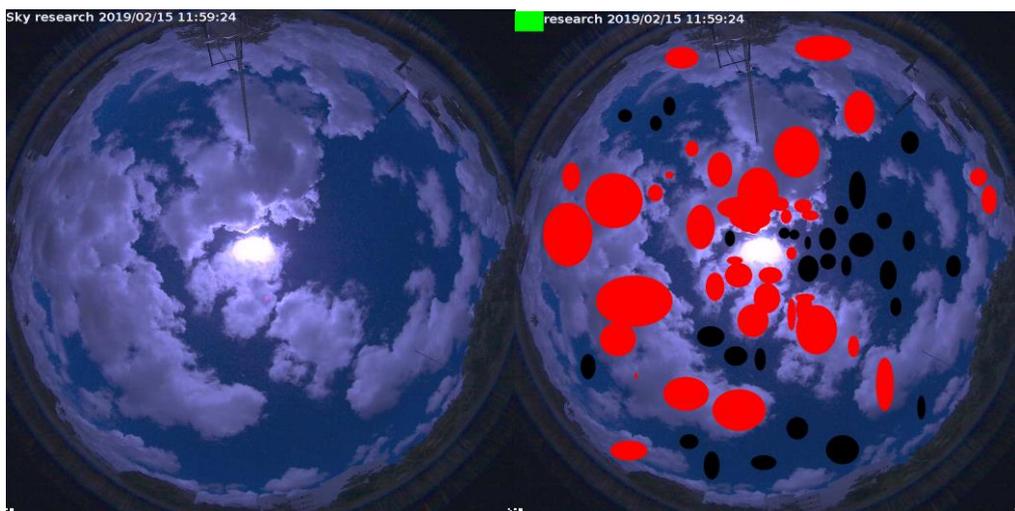
A segmentação teve como uma de suas etapas intermediárias o processo de realce do contraste das imagens através de um algoritmo selecionado para esse fim. No caso optou-se pelo uso do assim denominado algoritmo CLAHE (ZUIDERVELD, 1994), abreviatura para “*Contrast Limited Adaptive Histogram Equalization*”. Para ilustrar o exemplo da aplicação dessa técnica, encontra-se mais à frente (Figura 49) uma imagem do dia 15/04/19 processada através desse algoritmo. Nota-se o aumento do contraste das nuvens diante do fundo representado pelo céu. A ideia de incluir esse algoritmo acrescentou mais informações dentro dos vetores de características sobre os pixels a serem classificados. Essas informações foram acrescentadas nas formas dos atributos “ f_0 ” e “ f_1 ”.

Figura 49 – Aplicação do algoritmo CLAHE à imagem à esquerda e resultado à direita



Para treinamento de uma máquina de aprendizado (ML) com supervisão, foi selecionado um conjunto de amostras de imagens do hemisfério com exemplos de céu e nuvens contendo grande variabilidade de cores bem como de formatos de nuvens. Escolheu-se a cor vermelha ($R=255, G=B=0$) do espaço RGB para marcar os pixels característicos de nuvens e a cor negra ($R=G=B=0$) para representar os pixels do céu. A próxima ilustração (Figura 50) demonstra essa técnica de rotular os pixels para depois serem submetidos à análise de suas características.

Figura 50 – Método usado para classificar pixels como nuvem ou céu



Com o processamento de todos os pixels rotulados como vermelho ou preto criou-se uma base de dados de vetores, cada um contendo sete coordenadas. Esse procedimento totalizou mais de 813 mil vetores. E as coordenadas de cada vetor foram ordenadas de acordo com as expressões selecionadas (“ f_0 ” até “ f_6 ”). Através da tabela de classificação (Tabela 4) tornou - se possível comparar os atributos entre si de acordo com o valor assumido para cada métrica, representada na tabela pela sua respectiva coluna. A utilização de três métricas distintas demonstrou através da variação de desempenho de cada atributo, a importância dos classificadores de atributos como guia para seleção de acordo com critérios matemáticos que captam diferentes características para cada um dos sete parâmetros avaliados.

Tabela 4 – Resultados das métricas aplicadas a cada atributo

Atributos \ Métricas	Coef. de Pearson	<i>Ginfo</i>	<i>RGinfo</i>
f_0	0,864	0,7715	0,1473
f_1	0,877	0,7452	0,1453
f_2	0,151	0,0414	0,0100
f_3	0,851	0,7917	0,1230
f_4	0,867	0,7955	0,1242
f_5	0,864	0,7645	0,1239
f_6	0,908	0,7561	0,2034

Para complementar as informações da tabela anterior, gráficos dos atributos representados sob a forma de histogramas demonstram a importância da seleção de características que tenham a capacidade de dividir numericamente o máximo possível as regiões que representam nuvens e céu. Assim, estão colocados mais à frente desde a Figura 51 até a Figura 57 os mais significativos atributos anteriores em formato de histogramas. É possível perceber graficamente e de imediato a ideia de separação entre cores. Nesse caso particular, as regiões azuis representam as nuvens e as vermelhas o céu. À máquina de aprendizado, ML, coube a tarefa de classificar os vetores separando-os como céu ou nuvem em função dos exemplos apresentados durante a etapa de treinamento da rede. É interessante observar que, apesar da característica “ f_2 ” não ter obtido índices significativos em relação aos demais atributos escolhidos, sua seleção deve-se ao fato de a mesma ter contribuído de forma

perceptível no treinamento da SVC auxiliando na obtenção do percentual de acertos atingido. Quando foi testada sua eliminação dentro dos vetores apresentados à SVC, esse algoritmo manifestou um decréscimo indesejável na taxa de acertos, supostamente pela ausência das informações contidas nesse atributo. Tal comportamento revelou a necessidade de mantê-lo incluído dentro dos vetores de características.

Figura 51 – Histograma do atributo f_0

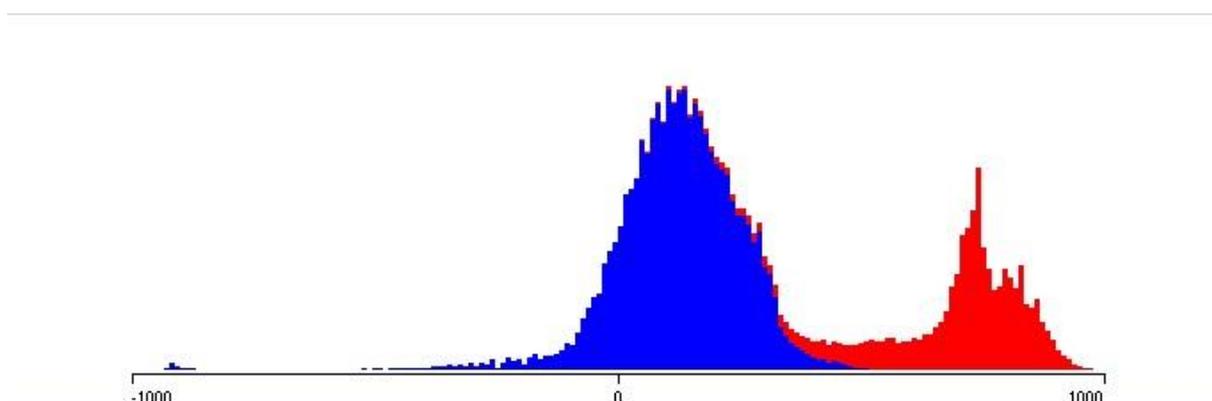


Figura 52 – Histograma do atributo f_1

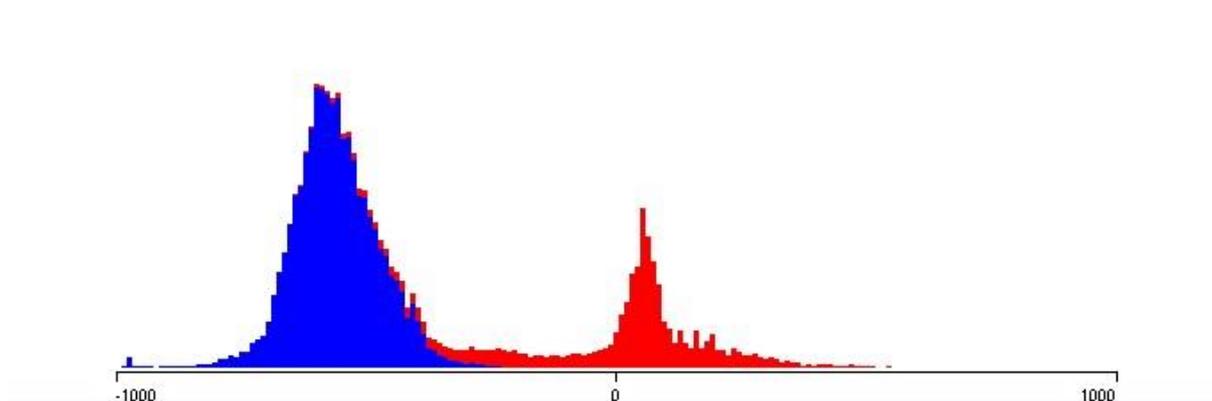


Figura 53 – Histograma do atributo f_2

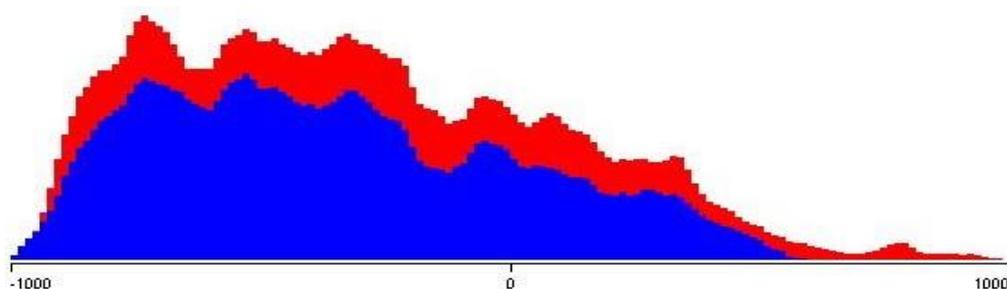


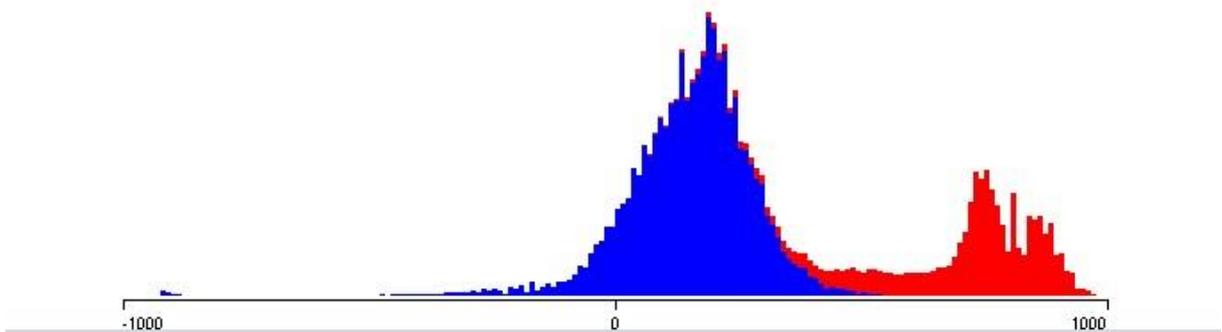
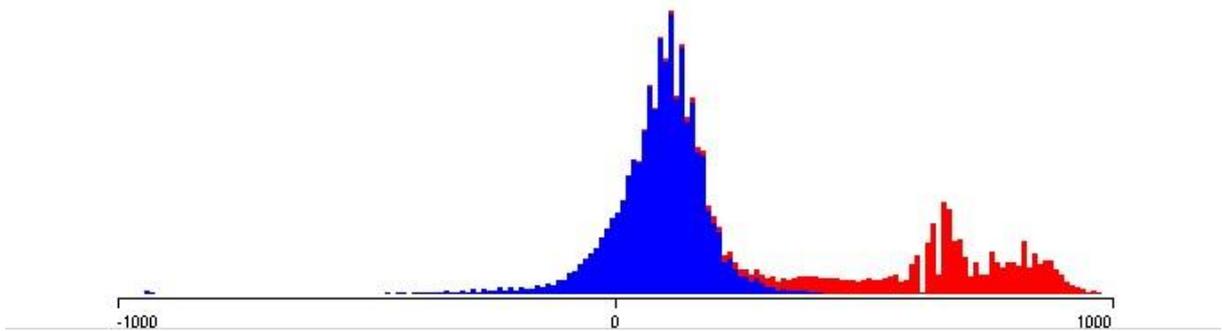
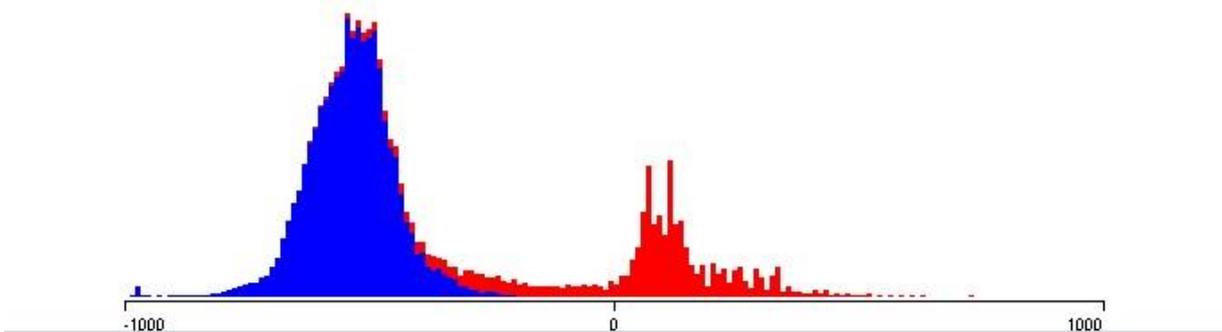
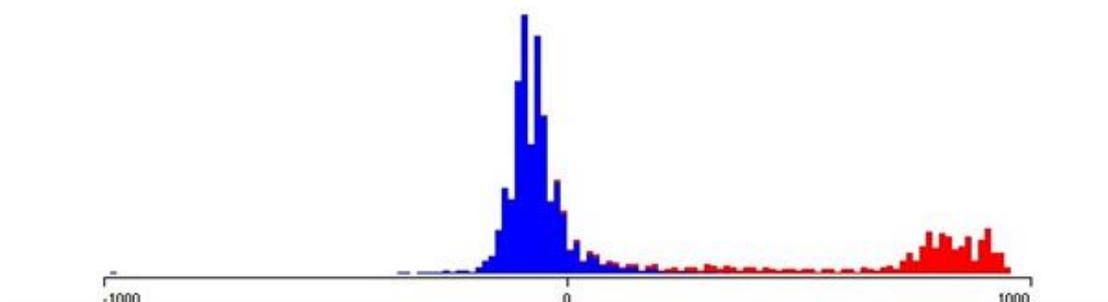
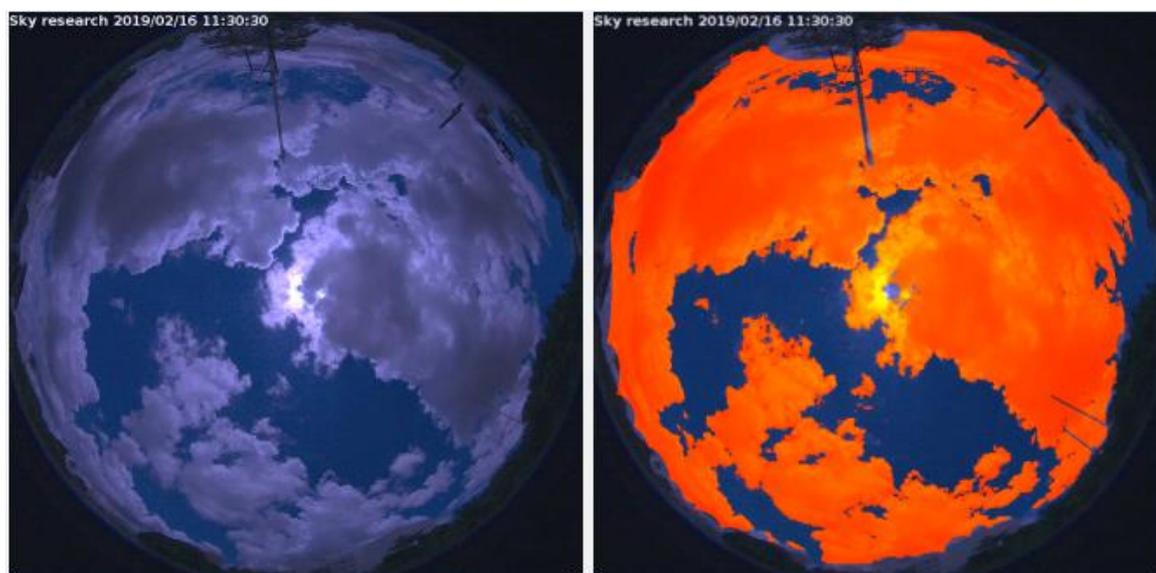
Figura 54 – Histograma do atributo f_3 **Figura 55 – Histograma do atributo f_4** **Figura 56 – Histograma do atributo f_5** 

Figura 57 – Histograma do atributo f_6 

Um total de mais de 813 mil vetores foi armazenado, cada qual com os valores das 7 características acima normalizadas para o intervalo $[-1;1]$ para serem submetidos à ML escolhida (do tipo SVC). Vetores supostamente em duplicidade foram testados e, em caso afirmativo, previamente removidos do conjunto de amostras, de modo a deixar apenas um exemplar único de cada ocorrência. Desse modo a SVC foi treinada e validada pelo método de validação cruzada. Assim, foi calculada a média percentual de acertos através do método de 10-*folders*, onde 10 subconjuntos disjuntos tiveram suas taxas de acertos calculadas em sequência após o treinamento no algoritmo LIBSVM com os respectivos 90% restantes das amostras. Como resultado, foi obtida uma acurácia de 99,2%, cuja Figura 58 (à direita) é resultante do treinamento/validação e finalmente aplicação desse algoritmo classificatório.

Figura 58 – Imagem segmentada (em tons de laranja) à direita

A título de observação, na imagem segmentada anterior no lado direito há uma pequena circunferência, próxima ao centro da imagem e cercada dos tons laranja. Essa circunferência representa a posição do Sol, nesse momento coberto por nuvens.

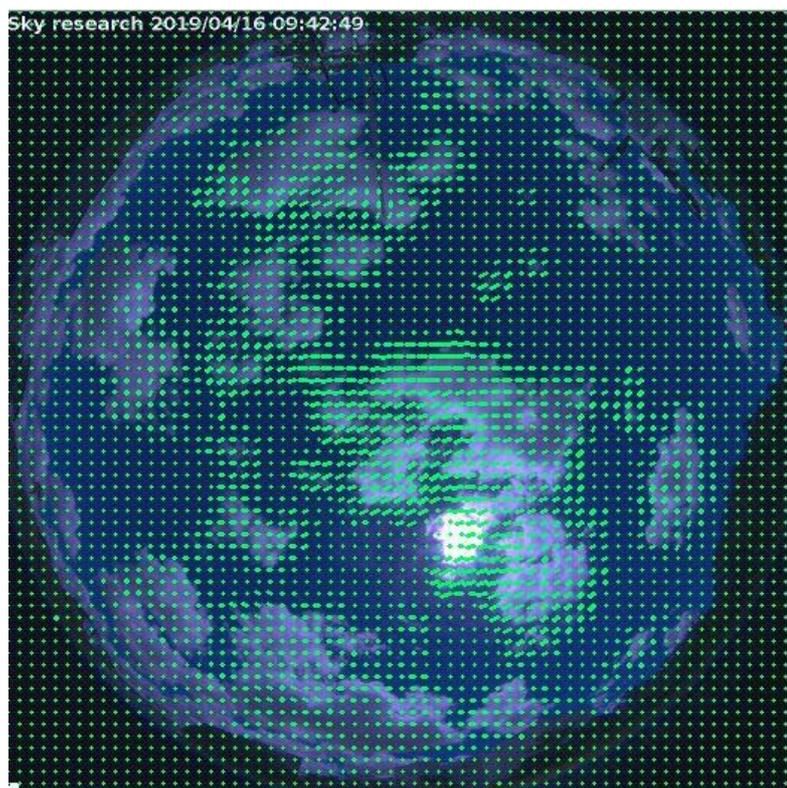
4.4 Fluxo ótico e velocidade das nuvens

Os sub-tópicos a seguir correspondem aos resultados obtidos para as duas alternativas de pesquisa desenvolvidas, ora por meio de fluxos óticos densos, a primeira a ser implementada, e depois através de fluxos esparsos, o caminho que foi efetivamente escolhido pelas razões que serão apresentadas a seguir.

4.4.1 Fluxo ótico denso

Na imagem da Figura 59, exemplo prático em que essa técnica foi empregada, pode-se ver as direções e os módulos (normalizados) da velocidade. Veem-se ainda pequenos pontos que indicam que não houve mudança de pixels.

Figura 59 – Mapeamento do deslocamento de pixels após algoritmo Gunnar Farneback

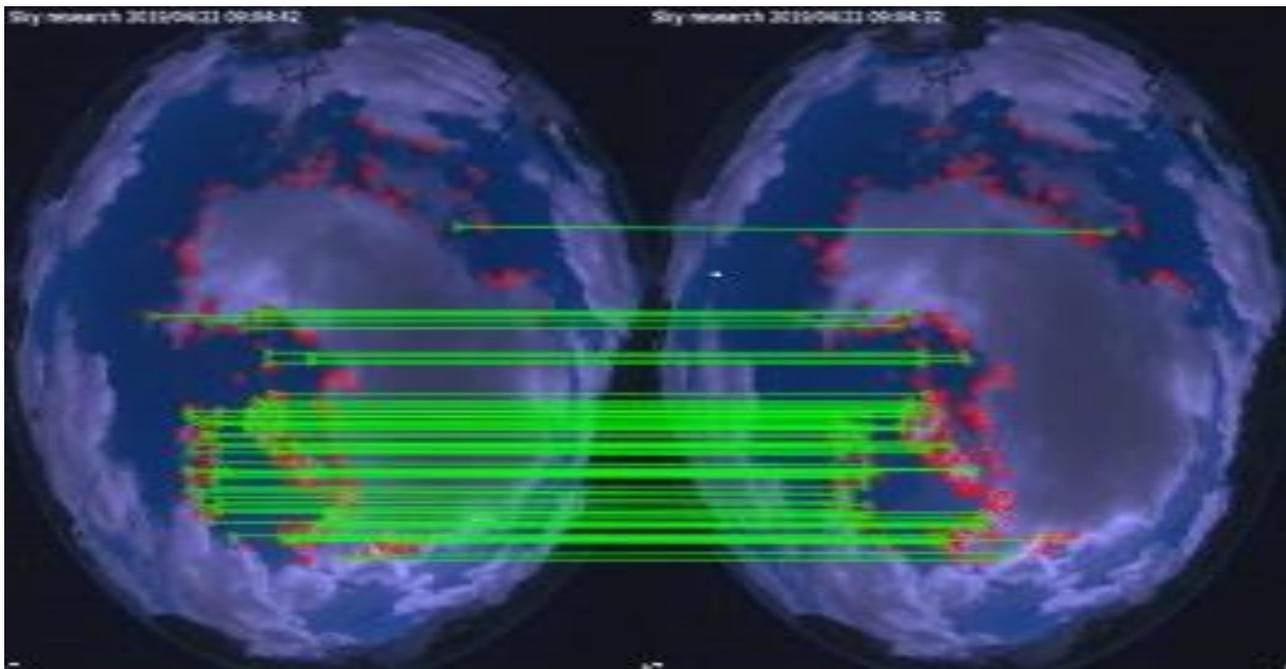


Esta análise mostrou-se bastante útil sob diversos aspectos. Além de indicar a direção e amplitude (previamente normalizada) de deslocamento, outras características se fizeram notar:

- Dado que as formações nebulosas são objetos com tamanhos e formatos muito variados, o tamanho da janela de pixels para análise - parâmetro de entrada desse algoritmo – não pode ser ajustado a todos os objetos encontrados (nuvens);
- O algoritmo encontra maior dificuldade em calcular a velocidade em regiões acentuadamente homogêneas em nuvens pela similaridade entre os pixels. Em outras palavras, em nuvens de tamanho maior a semelhança entre seus pixels internos fará com que esses apresentem menor velocidade que nas bordas;
- Essa homogeneidade resulta em uma velocidade média menor de uma nuvem com essas características considerado o somatório sobre todos os pontos;
- Quanto mais distante do centro a imagem, menor será o módulo da velocidade (dos pixels) das nuvens. É um fato tendencioso que foi observado, o que mereceu uma análise mais cuidadosa, dada a importância de conhecer melhor o tipo de movimento descrito pelos pixels em função de sua posição.

4.4.2 Fluxo ótico esperso

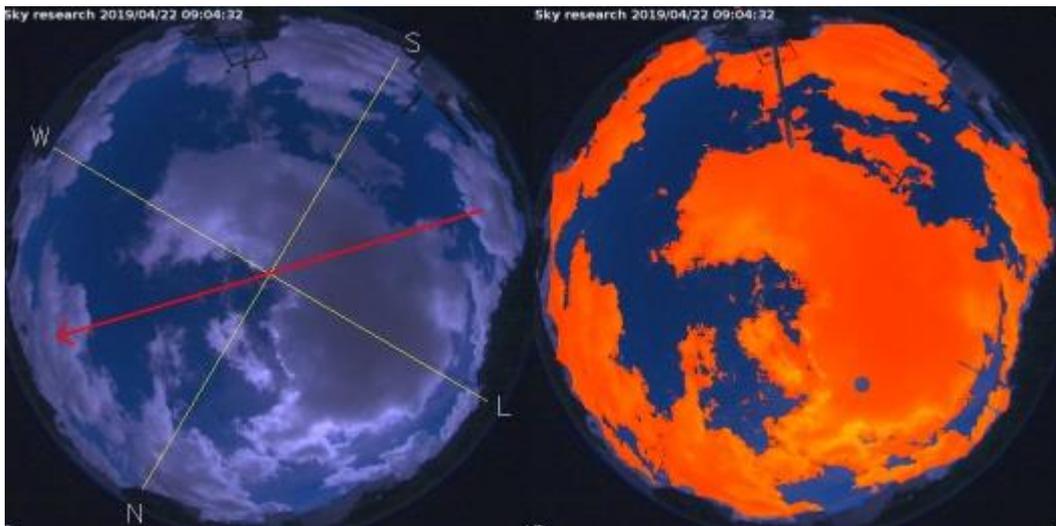
De acordo com a metodologia usada quando adotou-se o algoritmo FAST (ROSTEN, PORTER e DRUMMOND, 2010), segue através da Figura 60 um exemplo da utilização desse algoritmo para rastrear pixels de nuvens de uma fotografia em relação à outra obtida em sequência.

Figura 60 – Algoritmo FAST para encontrar pontos coincidentes em imagens seguidas

Eventualmente, pontos sobre o Sol ou mesmo pontos resultantes de alguma mancha sobre o vidro da câmera poderiam ser correlacionados e em função disso foi estabelecido um valor de limiar mínimo de deslocamento para excluir ou reduzir a ocorrência desses pontos.

Além disso, após serem aplicados os fatores de ajuste de velocidade mencionados no item anterior, foram escolhidos os pontos cujo deslocamento naquele intervalo de tempo apresentassem valores angulares dentro do intervalo $[\mu - 2\sigma ; \mu + 2\sigma]$, considerando que “ μ ” e “ σ ” sejam, respectivamente, a média e desvio padrão da distribuição angular. Para acumular uma quantidade mais expressiva de amostras dos valores de ângulos (como também de módulos), foi escolhida uma janela (móvel) de 10 minutos de intervalo de tempo de coleta entre sucessivas imagens. A combinação das técnicas e algoritmos escolhidos resultou na determinação de vetor velocidade (média) dentro de cada janela de 10 minutos. A Figura 61 abaixo, à esquerda, ilustra a colocação da direção média e sentido do vetor velocidade associado às imagens anteriores.

Figura 61 – Vetor velocidade média à esquerda e imagem segmentada à direita



A propósito da figura acima, o pequeno círculo acinzentado na imagem à direita, segmentada, representa a posição do Sol na fotografia.

4.5 Modelo para estimar as irradiâncias a partir das imagens

Para o modelo de estimativa das irradiâncias, foi selecionado um dia em que houve grande variabilidade das condições entre céu claro e nublado, de modo a que o modelo pudesse ser abrangente diante de uma diversidade de situações. Para exemplificar melhor, seguem fotografias do dia 18/04/2019 como amostras visuais dessa variabilidade (figuras 62 e 63). Mais à frente (Figura 64), gráficos com as irradiâncias total e direta ao longo do dia constataam essa variabilidade. Os dados foram analisados em um intervalo compreendendo desde 07:47 da manhã até às 16:29 da tarde.

Figura 62 – Nebulosidade e trajetória do Sol no dia 18/04/2019 às 14h00min

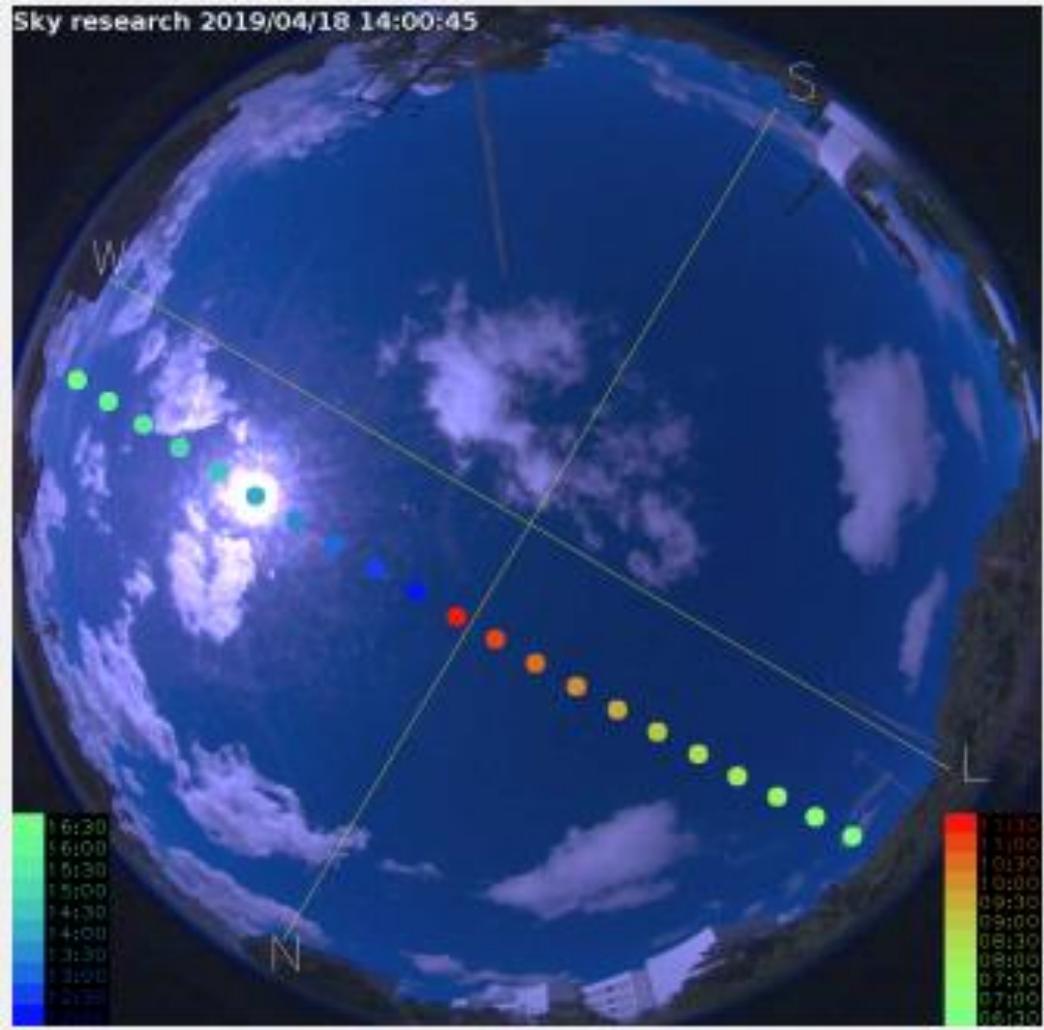
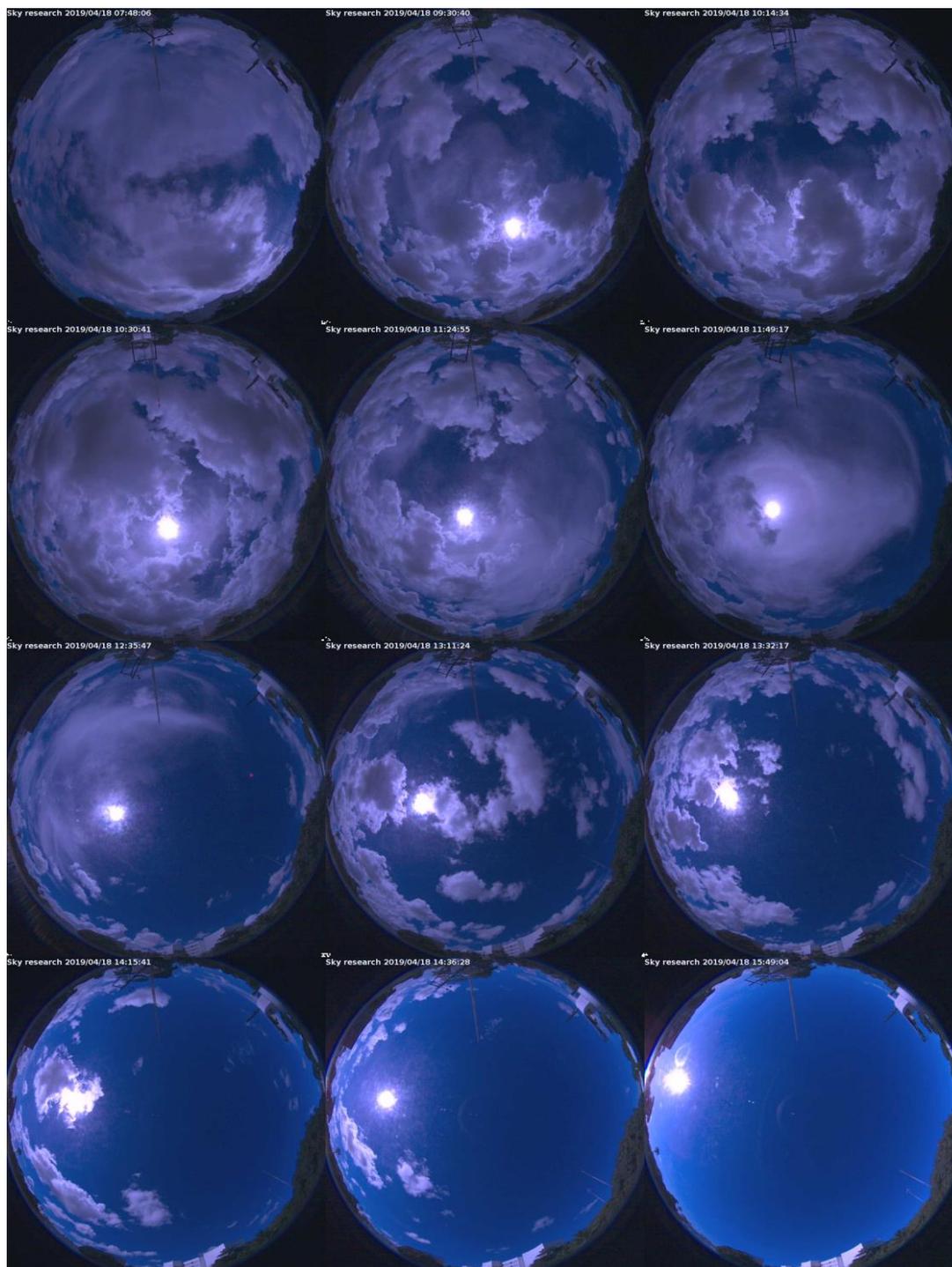


Figura 63 – Sequência de fotografias do hemisfério no dia 18/04 em ordem cronológica



A análise das imagens resultou em dois conjuntos destinados respectivamente a cada modelo de irradiância (total e direta). Cada conjunto reuniu 6236 vetores únicos e colocados em ordem aleatória para serem submetidos à ML SVR através da biblioteca LIBSVM. Nesta etapa do estudo a SVM foi usada não mais para classificação, mas para o cálculo de regressão. Novamente o método de validação cruzada de 10 *folders* foi aplicado para treinar e validar o

modelo. Ambos os valores de entrada de irradiância, seja total ou direta, foram normalizados para o intervalo $[-1;1]$ – ver Eq. 66. Os modelos usaram como entrada de dados apenas aqueles oriundos das fotografias da câmera conforme foi descrito no tópico 3.9. Os valores estimados para as irradiâncias total e direta foram calculados na etapa de validação cruzada.

A Tabela 5 a seguir é um resumo dos resultados dos modelos segundo as métricas utilizadas para avaliação dos desvios.

Tabela 5 – Resultados dos modelos para as irradiâncias total e direta

Irradiâncias	(RMSE) (W/m^2)	(MAE) (W/m^2)	(MBE/Viés) (W/m^2)	Coef. de Pearson(%)	Núm. de amostras
Total	71,8 ±9,3	30,2 ±3,0	6,5 ±2,6	98,0 ±0,6	6236
Direta	61,5 ±8,1	26,1 ±2,5	5,7 ±2,0	98,0 ±0,5	6236

Nas expressões a seguir estão as definições usadas para as métricas RMSE (Eq.63), MAE (Eq. 64) e MBE (Eq. 65) em que “ x_i ” representa o valor normalizado por meio da equação 66 (onde “ m_i ” representa o valor medido). Ao mesmo tempo “ x_p ” significa o valor previsto na mesma escala de normalização do modelo. A “de-normalização” para cálculo dos valores previstos pelo modelo encontra-se explicitada na equação 67 através dos valores medidos máximo e mínimo do conjunto de amostras.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - x_p)^2} \quad \text{Eq. 63}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i - x_p| \quad \text{Eq. 64}$$

$$MBE = \frac{1}{N} \sum_{i=1}^N (x_i - x_p) \quad \text{Eq. 65}$$

$$x_i = -1 + 2 \left(\frac{m_i - m_{min}}{m_{max} - m_{min}} \right) \quad \text{Eq. 66}$$

Da equação anterior deduz-se a relação inversa que fornece o valor estimado de-normalizado “ m_p ” relativo ao valor previsto “ x_p ”:

$$m_p = \frac{m_{max} + m_{min}}{2} + \left(\frac{m_{max} - m_{min}}{2}\right)x_p \quad \text{Eq. 67}$$

O coeficiente de “ x_p ” relativo aos valores medidos máximo e mínimo, “ $\left(\frac{m_{max} - m_{min}}{2}\right)$ ”, foi responsável por um fator multiplicativo sobre os valores das métricas RMSE, MAE e MBE para de-normalizar os resultados apresentados na tabela anterior.

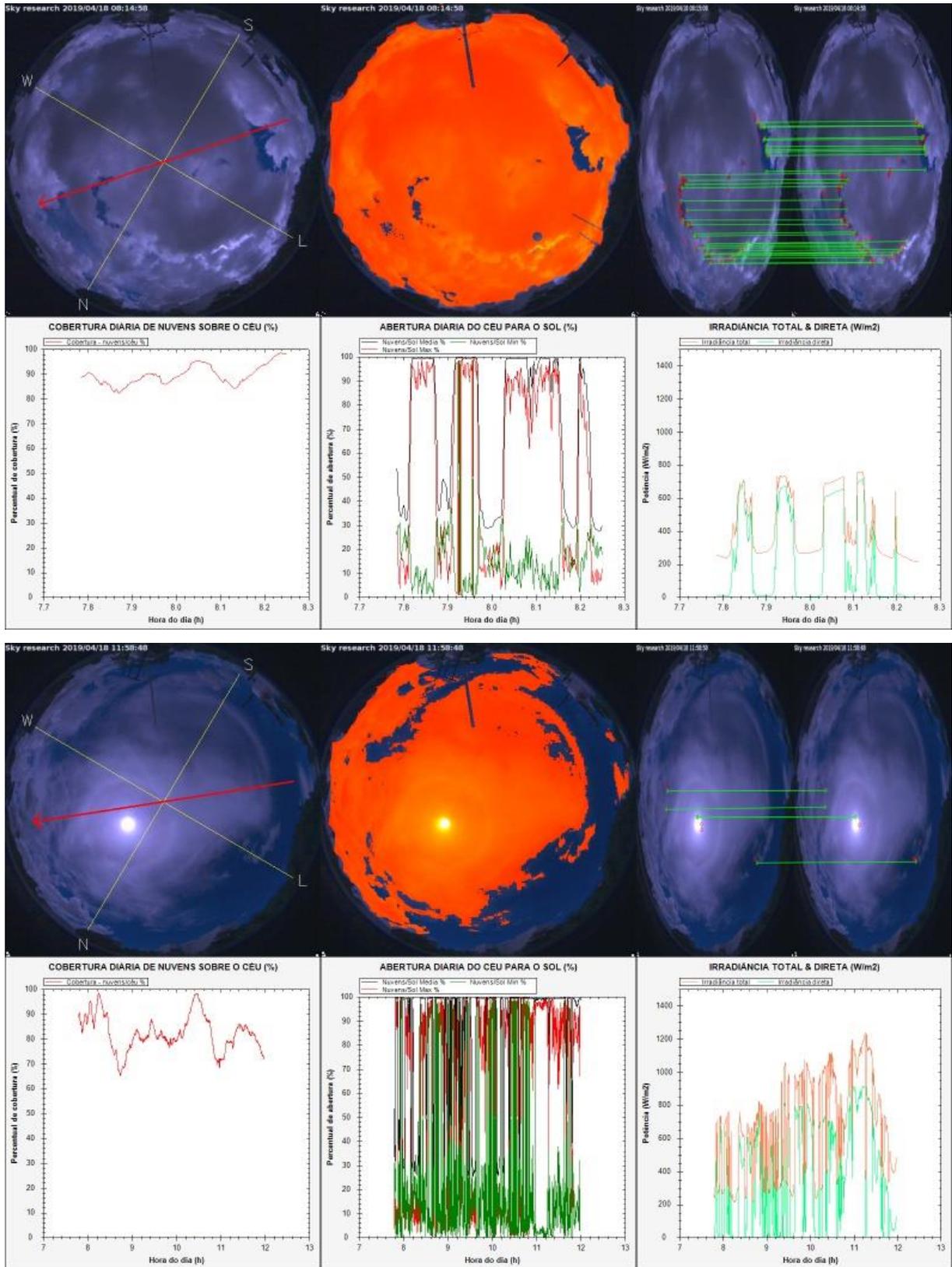
O trabalho de Faber (FABER, 1999), apresenta um resultado útil associado à incerteza sobre o estimador RMSE, cuja notação adotada foi “ \widehat{RMSE} ”. É mostrado que:

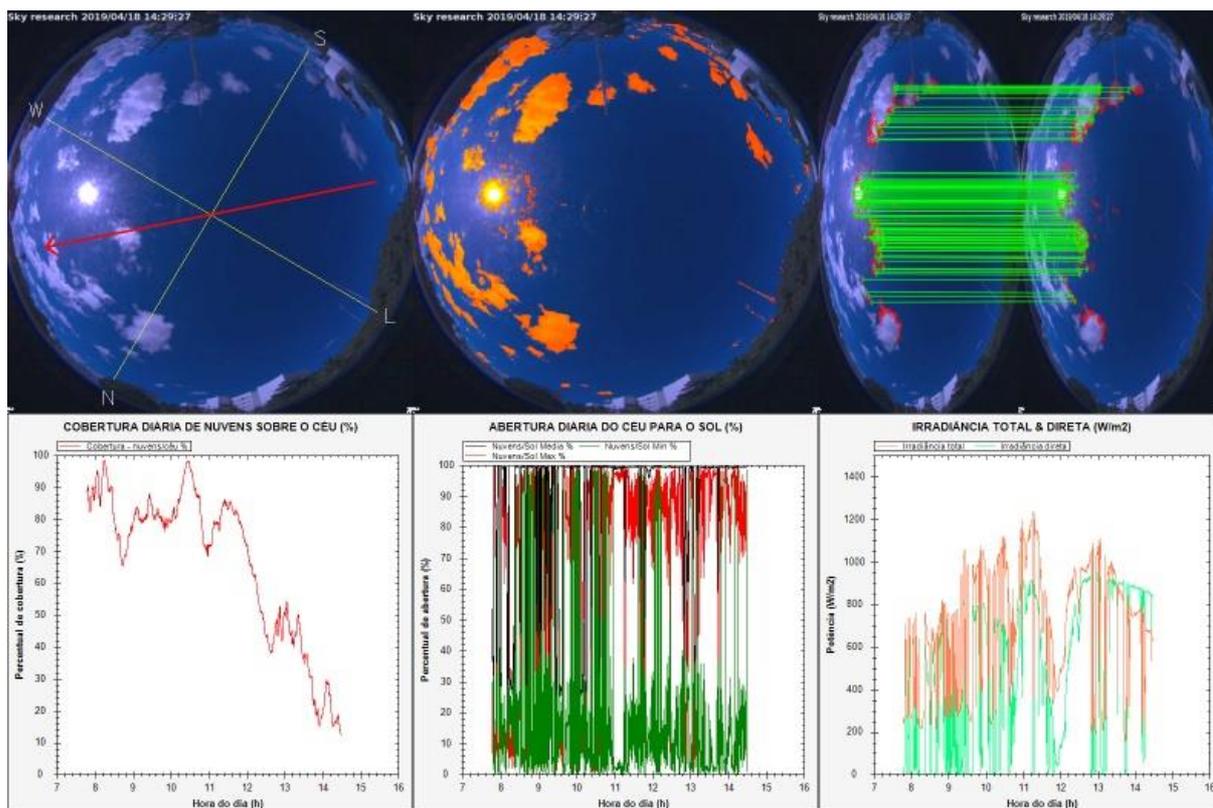
$$\frac{\sigma(\widehat{RMSE})}{RMSE} \approx \sqrt{\frac{1}{2n}} \quad \text{Eq. 68}$$

Em que “ n ” é o número de amostras - 10% sobre 6236 considerando a divisão em 10 *folders* - e “ σ ” o desvio padrão sobre o estimador. Substituindo o valor de “ n ” na expressão anterior, tem-se que o quociente (do primeiro termo) da equação (Eq. 68) é aproximadamente 0,028. Dessa forma, através do número de amostras usadas para cada folder do processo de validação, é possível concluir que em cada uma das etapas de validação o desvio padrão para o estimador dos valores RMSE das irradiâncias assume um valor baixo quando comparado ao próprio RMSE que se pretende estimar.

Paralelamente ao processamento dos dados para geração dos dois modelos anteriores, gráficos da cobertura de nuvens, gráficos relativos aos três últimos atributos dos vetores de características (que deram origem ao modelo) e também gráficos das irradiâncias total e direta foram acumulados em conjunto com a imagem do céu em cada instante. A sequência de imagens a seguir (Figura 64) exemplifica alguns desses instantes do cálculo computacional realizado durante a extração dos dados a partir dos quais foram gerados os modelos. Na Figura 64, à esquerda, na parte inferior das fotografias do céu, estão gráficos relativos à fração de cobertura das nuvens com o tempo até aquele instante. À direita, gráficos da irradiância total (em vermelho) e direta (em verde). E ao centro 3 gráficos relativos ao método de Otsu para tons de cinza: em verde o valor “ $OT1 \cdot p$ ”, em vermelho o valor “ $OT2 \cdot (1 - p)$ ” e por fim, em preto, o valor “ $OT1 \cdot p + OT2 \cdot (1 - p)$ ”. Todas as 3 expressões anteriores foram definidas na seção 3.9.

Figura 64 – Imagens e gráficos durante a modelagem das irradiâncias total e direta





Na sequência da esquerda para direita da figura anterior, Figura 64, tem-se a cada instante a imagem original contendo os pontos cardeais e a orientação do vetor de velocidade média, ao centro a imagem com a segmentação de nuvens e ao final a última imagem exibindo os pontos correspondentes entre a fotografia atual e a próxima pelo algoritmo FAST. A tempo, é importante observar que nas imagens anteriores estão fotografias do céu para três instantes diferentes, porém do mesmo dia 18/04/2019.

4.6 Modelo para previsões em curtos intervalos de tempo

O modelo utilizou 8 dias com grande variabilidade: dados do dia 18/04 visto no tópico anterior adicionados aos dados dos dias 15/04, 16/04, 19/04, 22/04, 23/04, 03/05 e 21/05 de 2019, todos marcados por pronunciada variabilidade das condições de nebulosidade, o que representou uma situação mais difícil para previsão. Foram colocados abaixo, Figuras 65 a 69, gráficos das irradiâncias dos dias 18/04, 19/04, 22/04, 23/04 e 21/05 para exemplificar a variabilidade encontrada. O conjunto total de amostras foi coletado de modo a incluir o intervalo entre, aproximadamente, 08h00m às 16h00m, com fotografias do hemisfério capturadas a cada 5 segundos. O treinamento/validação e teste dos modelos de previsão para

GHI e DNI usou a base de dados construída com 2939 amostras com intervalos de 1 minuto de observação, considerando que esses períodos de 1 minuto foram representados pelos valores médios calculados. A avaliação final do modelo baseou-se no subconjunto de testes - compreendendo 4,45h de predição ou 267 amostras - não utilizado no treinamento da ML SVR e cronologicamente o último à frente.

Figura 65 – Gráficos das irradiâncias total (vermelho) e direta (verde) para o dia 18/04

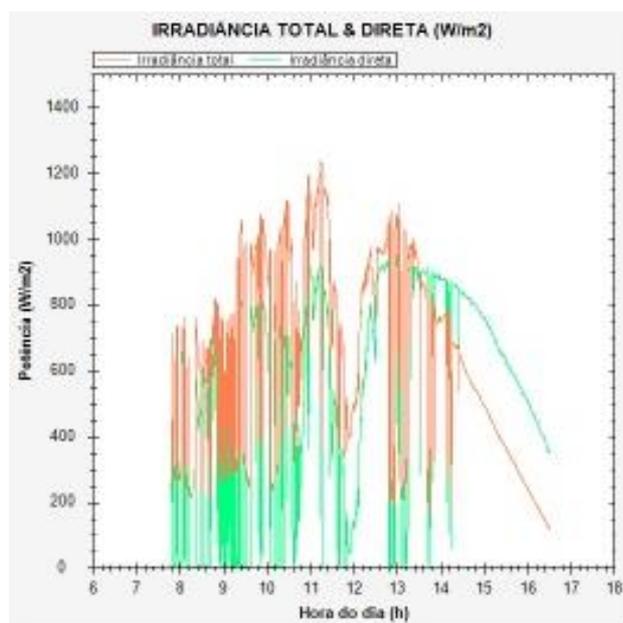


Figura 66 – Gráficos das irradiâncias total (vermelho) e direta (verde) para o dia 19/04

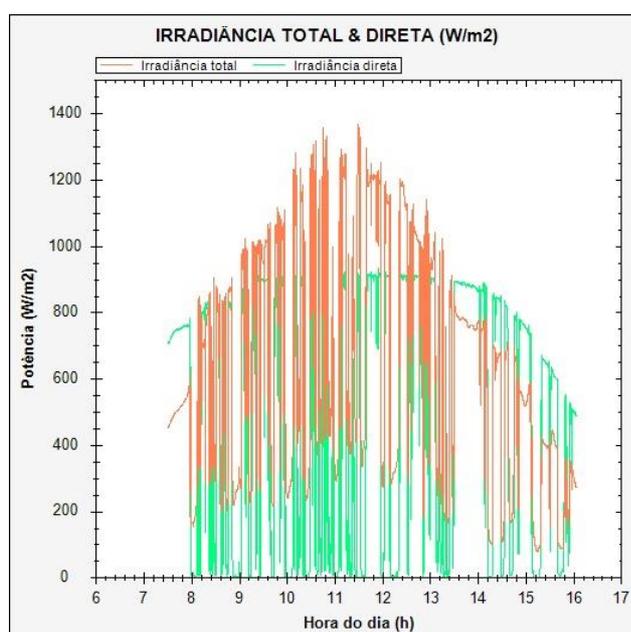


Figura 67 – Gráficos das irradiâncias total (vermelho) e direta (verde) para o dia 22/04

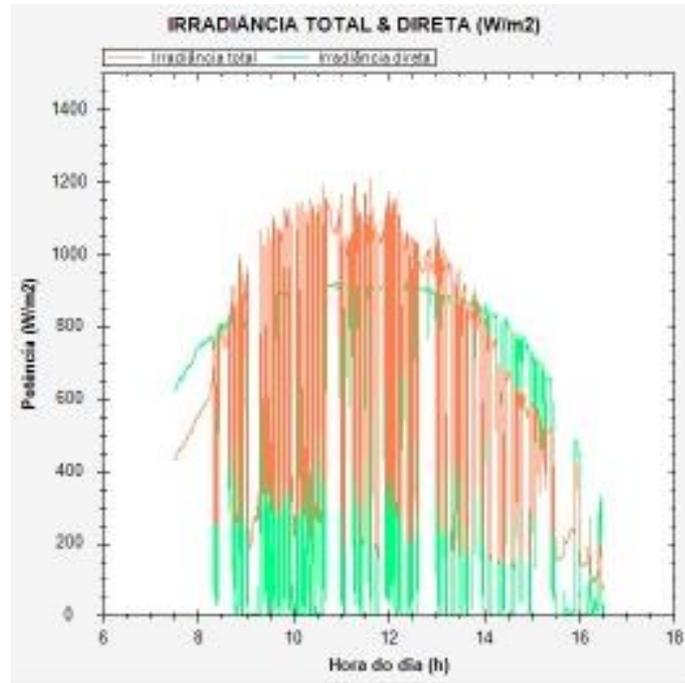


Figura 68 – Gráficos das irradiâncias total (vermelho) e direta (verde) para o dia 23/04

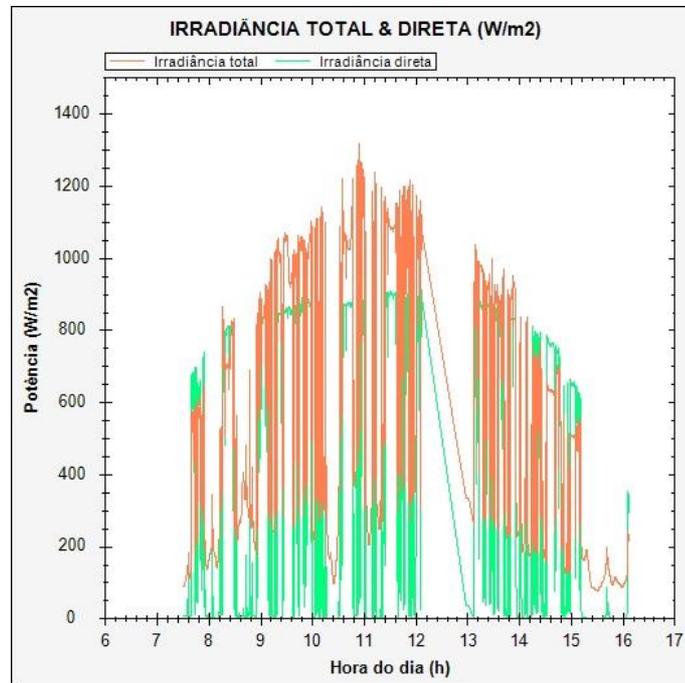
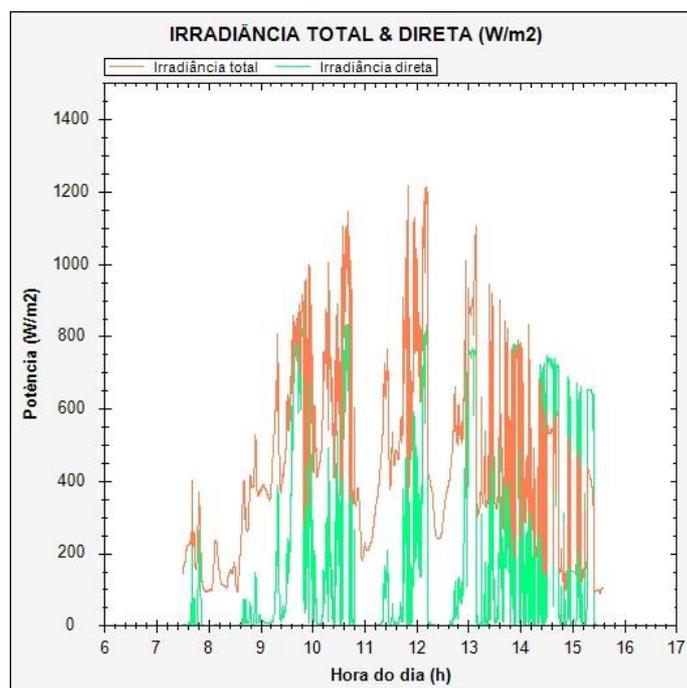


Figura 69 – Gráficos das irradiâncias total (vermelho) e direta (verde) para o dia 21/05



A seguir podem ser vistas as Tabela 6 e 7 que contém os resultados para as métricas aplicadas à predição da irradiância total (GHI) e direta (DNI).

Tabela 6 – Resultados dos modelos de predição para a irradiância total (GHI)

Intervalos de predição (média em 1 min)	(RMSE) (W/m^2)	(MAE) (W/m^2)	(MBE/Viés) (W/m^2)	Núm. de minutos
01	179,4	122,6	11,4	267
02	228,7	173,4	32,1	267
03	248,1	195,5	52,8	267
04	257,3	206,8	71,8	267
05	265,8	217,8	87,5	267
10	291,2	246,3	106,6	267
15	299,4	255,2	114,8	267
30	304,6	251,2	89,3	267

Tabela 7 – Resultados dos modelos de predição para a irradiância direta (DNI)

Intervalos de predição (média em 1 min)	(RMSE) (W/m^2)	(MAE) (W/m^2)	(MBE/Viés) (W/m^2)	Núm. de minutos
01	214,7	139,8	18,5	267
02	269,6	203,8	42,3	267
03	287,6	235,7	69,3	267
04	300,7	257,5	85,7	267
05	313,9	275,3	99,8	267
10	367,1	328,8	120,4	267
15	387,4	346,0	128,2	267
30	401,5	356,8	120,3	267

Em termos gráficos, será possível visualizar os resultados RMSE das tabelas anteriores por meio das Figuras 70 e 71.

Figura 70 – Gráfico RSME(GHI) prev. x Horizonte de predição

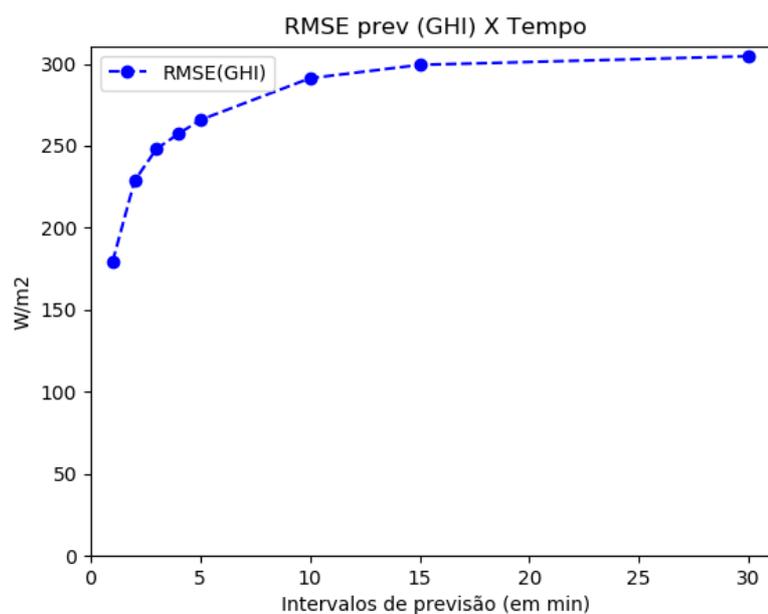
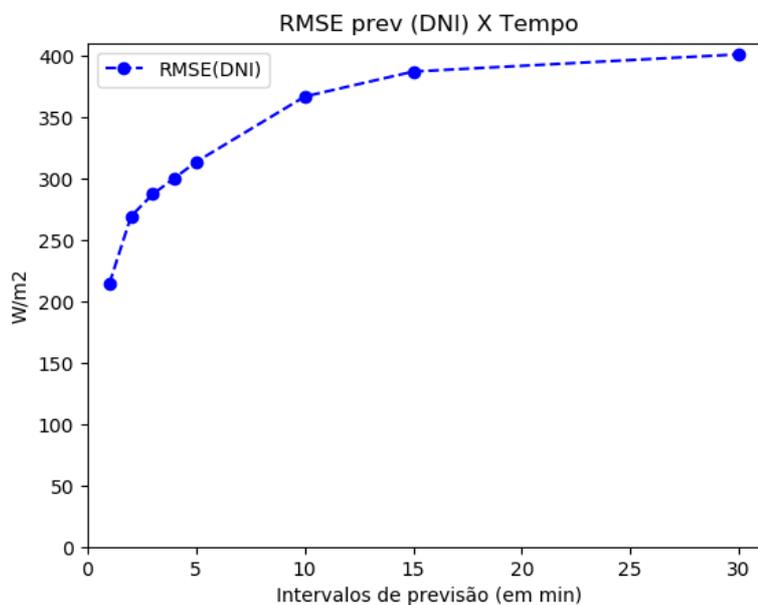
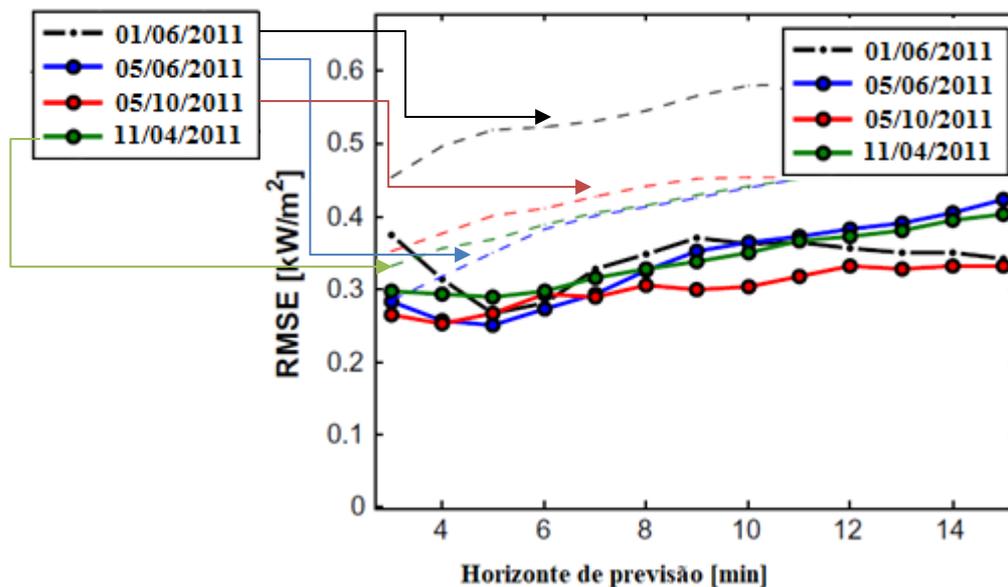


Figura 71 – Gráfico RSME(DNI) prev. x Horizonte de predição

Para melhor avaliação dos resultados é útil situá-los em relação a trabalhos que têm sido referenciados na literatura em termos de modelos preditivos para GHI e DNI voltados a intervalos de curta duração.

O gráfico da Figura 72 refere-se ao trabalho de Marquez e Coimbra (MARQUEZ e COIMBRA, 2013) para predição da irradiância direta em que foram utilizadas imagens tomadas a partir do solo em intervalos de 1 minuto através do sistema TSI (*Total Sky Imager*). O estudo dos autores acima conseguiu resultados preditivos satisfatórios medidos através da redução da métrica RMSE em relação aos modelos por persistência simbolizados pelas linhas pontilhadas (da Figura 72) para cada um dos 4 dias analisados: 01/06, 05/06, 05/10 e 11/04/2011.

Figura 72 – Modelos (DNI): persistência e novo método (linhas pontilhadas/cheias)



A Figura 73 mostra o gráfico para a DNI medida e prevista considerando o dia 05/06 da figura anterior, um dos que apresentou melhor resultado após 15 minutos. Apesar da variabilidade da irradiância direta durante esse dia, nota-se períodos com, aproximadamente, até 40 minutos caracterizados por discretas flutuações, o que contribui na redução dos desvios gerados em algoritmos de predição. Ao final de 15 minutos, em função das condições atmosféricas de cada dia analisado, nota-se que o desvio RMSE ficou, aproximadamente, na faixa entre 300 e 400 kW/m². É importante ressaltar que para cada um dos dias acima e para cada intervalo (de 3 a 15 minutos) foi selecionada uma variável de predição com o menor RMSE entre seis possíveis escolhas. Além disso, o intervalo horário analisado foi limitado entre 10h00min e 14h00min, conforme apresenta a Figura 73, o que contribuiu para reduzir os valores RMSE. Caso houvesse sido tomado um período diário semelhante ao deste trabalho esse desvio teria sido maior.

Os dados utilizados neste trabalho relativamente à irradiância direta conseguiram desempenho similar, contudo houve um intervalo diário de coleta de dados mais extenso que cobriu assim uma faixa mais ampla de irradiâncias em função da maior abertura do ângulo zenital. Os gráficos das Figuras 74 e 75, relativamente a 1 e 5 minutos de predição, revelaram intensa variabilidade em comparação com o da Figura 73 e aos demais do trabalho de Marquez e Coimbra Coimbra (MARQUEZ e COIMBRA, 2013).

Figura 73 – Gráfico de predição para o dia 05/06 (horizonte de 5 minutos)

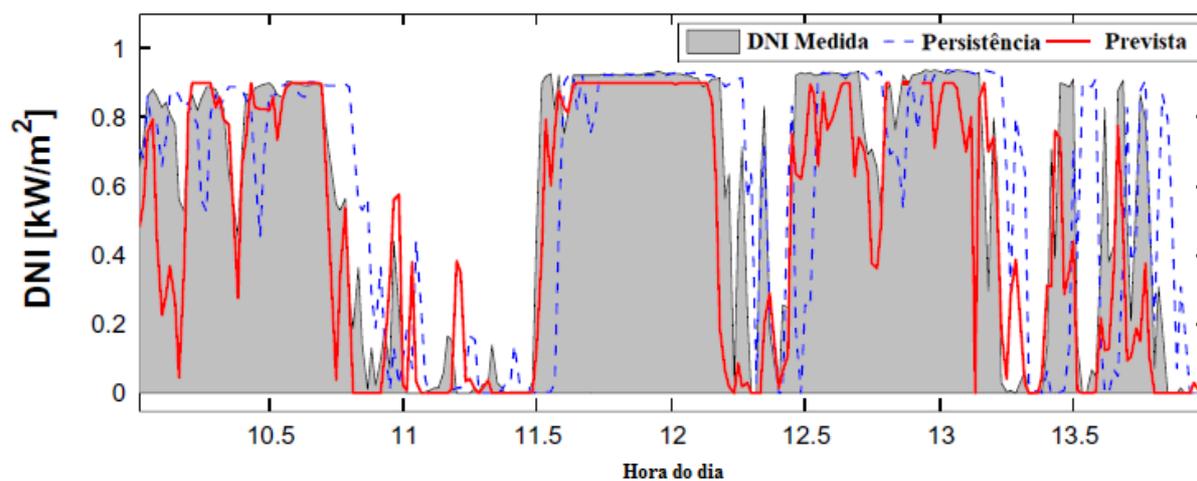


Figura 74 – Gráfico do modelo SVR de predição para o horizonte de 1 minuto

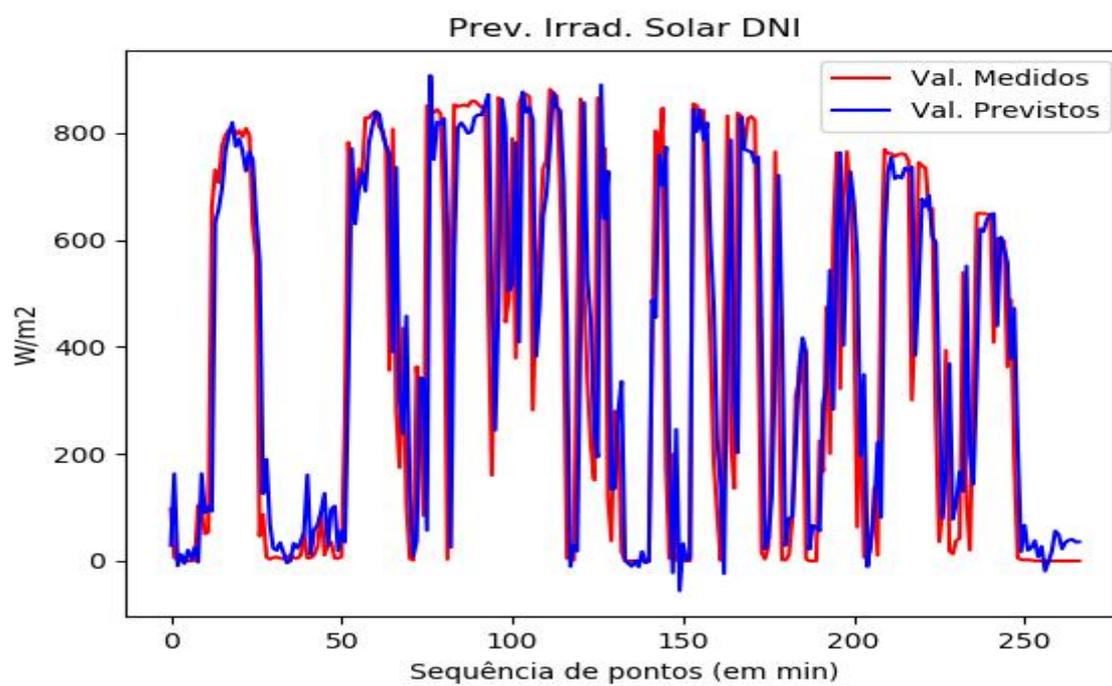
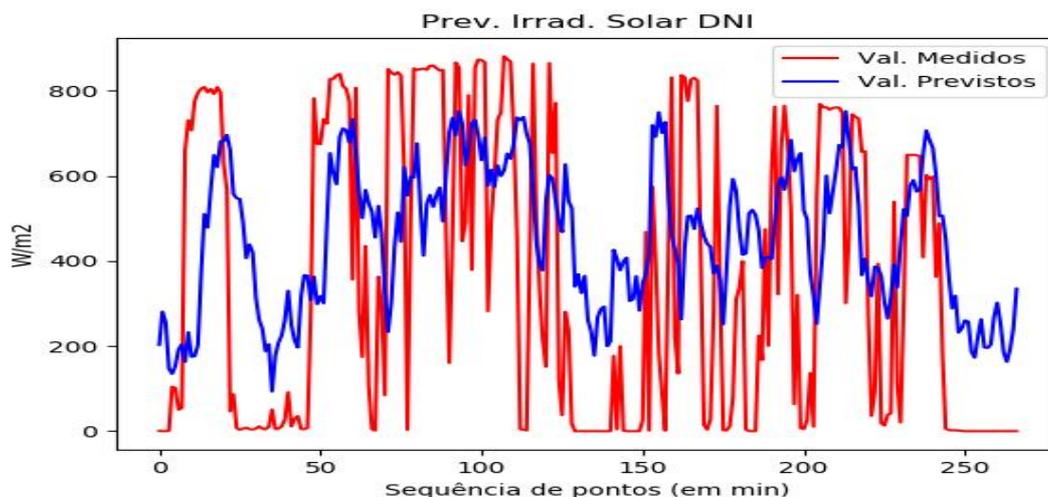


Figura 75 – Gráfico do modelo SVR de predição (DNI) para o horizonte de 5 minutos



Por meio da Figura 76 encontra-se graficamente representado um método desenvolvido por Quesada-Ruiz (QUESADA-RUIZ *et al.*, 2014) para predição da DNI e intitulado “setor em escada-setor”. É útil observar que o gráfico foi obtido com medições apenas durante o período de 10h00min às 14h00min, conforme ilustrado na Figura 77. O método denominado escada-setor obteve melhor desempenho conforme é visto na próxima Figura 76. Esse trabalho levou em conta dados acumulados de 14 dias. No entanto, o valor RMSE resultante do método de persistência – para fins de comparação - apresentou um valor máximo de 62,3 W/m² e um valor médio de 18,4 W/m² para um subconjunto de 6 dos 14 dias utilizados. Uma vez que os modelos de persistência usualmente não detêm os melhores índices RMSE quando comparados a outros modelos, é possível cogitar sobre o quanto foi significativa a variabilidade durante parte ou todo o período de amostragem de dados. Além disso, selecionar 6 dias com tempo significativamente nublado ou de céu claro em 14 dias, totaliza 42,8% dos dias escolhidos, o que significa uma escolha que reduzirá os desvios RMSE no cálculo final. A Figura 78, por meio da mesma legenda da Figura 77, anterior, contém exemplos de dois dias: um de céu claro (a) e outro intensamente nublado (b) fornecidos pelos autores do trabalho. É imediata a percepção de que algoritmos de modelagem seriam menos exigidos em relação a esses dois dias e produziriam resultados muito melhores do que nos casos de grande flutuação de irradiâncias.

Figura 76 – Método setor em escada ajustável para predição da DNI

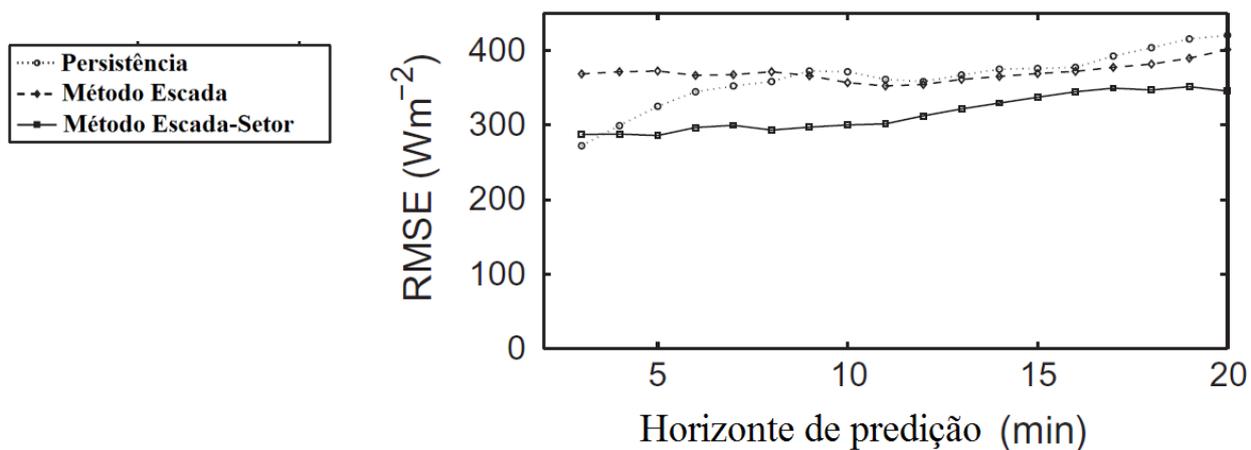


Figura 77 – Gráfico da variação da DNI relativo ao da figura anterior

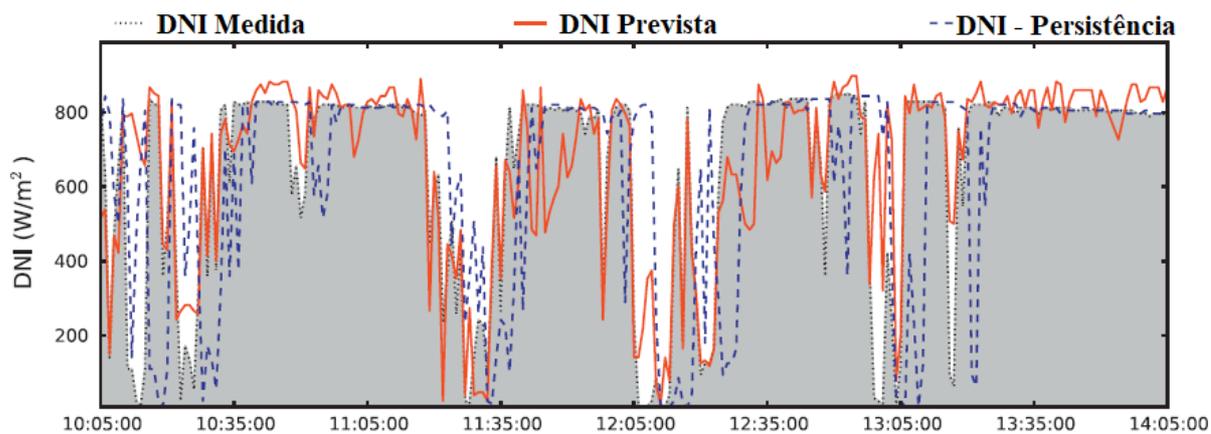
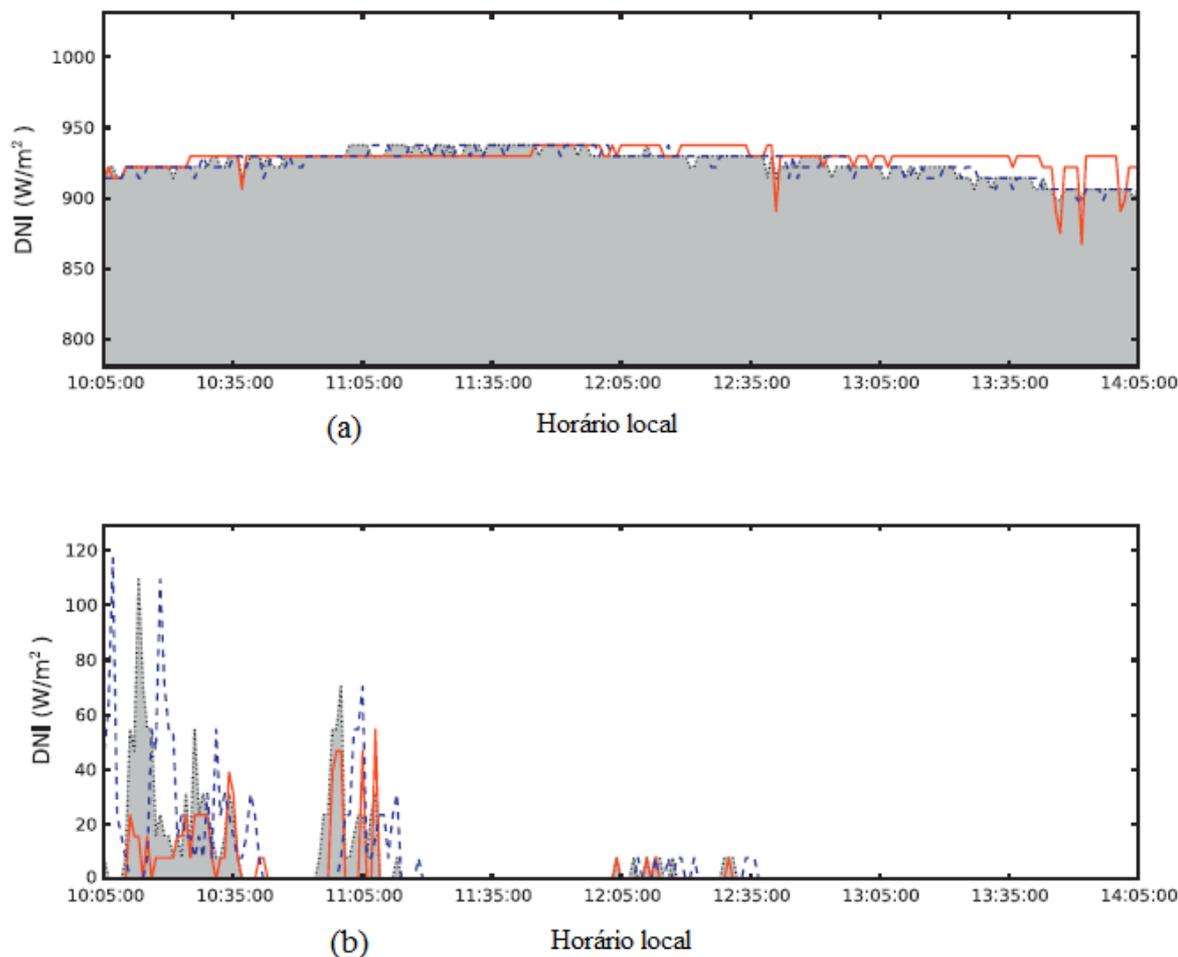


Figura 78 – Dias com pequena variabilidade: (a) céu claro e (b) nublado



A respeito da predição da radiação total, Kumler (KUMLER e ZHANG, 2019) obteve bons resultados por meio de um método denominado PSPI, sigla em alusão a um modelo inteligente de persistência baseado na Física. A Tabela 8, a seguir, contém resultados desse estudo para até 30min de predição. O método de Kumler, no entanto, diferentemente dos dois trabalhos citados anteriormente e da metodologia usada neste texto, utilizou valores médios em intervalos de 5 minutos com um passo ou deslocamento de 5 minutos ao longo das amostras. Evidentemente essa técnica causa maior impacto sobre o efeito de suavização de uma curva, especialmente nas regiões que existem mudanças abruptas na forma de picos positivos ou negativos. E com isso o algoritmo de predição será mais favorecido no que diz respeito a acompanhar a curva de irradiância GHI do que no caso de médias tomadas a cada minuto cujas curvas resultantes serão menos suavizadas. Desse modo, a comparação entre uma metodologia que utiliza médias em intervalos 5 vezes superiores (a outras metodologias) fica comprometida

não apenas por essa desigualdade, mas também, naturalmente, em termos de diferentes bases de dados de diferentes regiões com seus climas e condições atmosféricas peculiares.

Tabela 8 – Modelo PSPI para predição da GHI

Tempo Predição (min.)	RMSE W/m^2	MAE W/m^2	MBE (viés) W/m^2
5	107,6	47,5	-1,63
15	140,0	68,8	0
30	163,6	85,8	2,74

Os gráficos apresentados nesse artigo relativamente ao modelo PSPI podem ser vistos nas Figuras 79 e 80. Apresentaram um seguimento satisfatório dos valores medidos. Entretanto, os dias escolhidos para demonstrar graficamente a eficiência referem-se, respectivamente, a um dia claro e um dia nublado com variações da irradiância (GHI) visualmente menores do que as variâncias mostradas neste trabalho (Figuras 65 a 69) e também visualmente menores em relação aos valores medidos nos gráficos de predição (GHI) para 1 minuto (Figura 81) e 5 minutos (Figura 82) originados deste trabalho. É oportuno enfatizar que 5 minutos representam o intervalo menor para os valores médios do algoritmo PSPI (Figuras 79 e 80), ao passo que nesse trabalho utilizou-se 1 minuto. Isso significa que 5 min e 1 min traduzem, respectivamente, os melhores resultados (gráficos) para cada método.

Figura 79– Dias com pequena variabilidade: céu claro

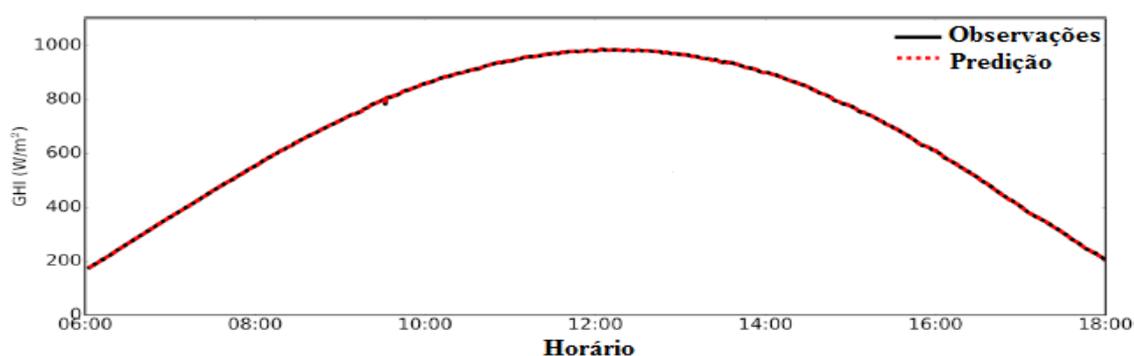


Figura 80 – Dias com pequena variabilidade: nublado

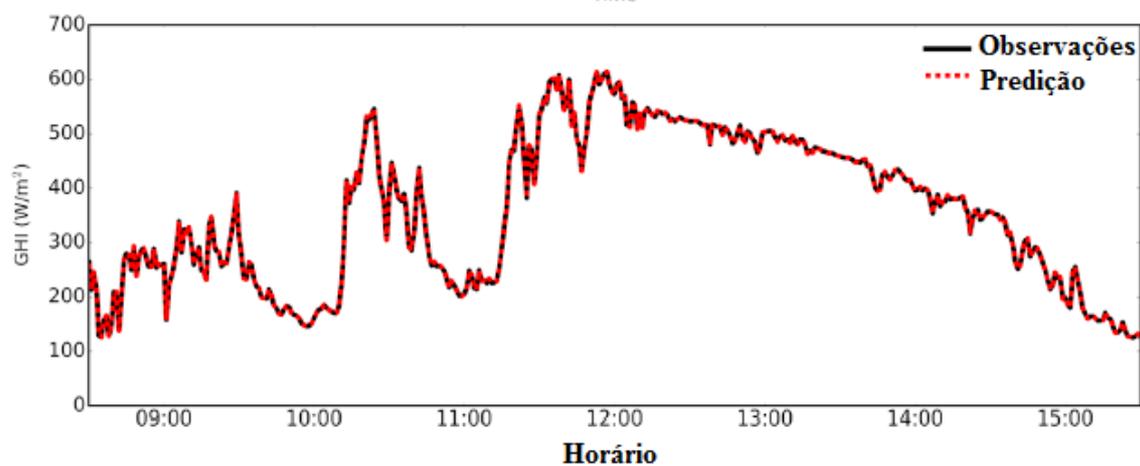


Figura 81 – Gráfico do modelo SVR de predição (GHI) para o horizonte de 1 minuto

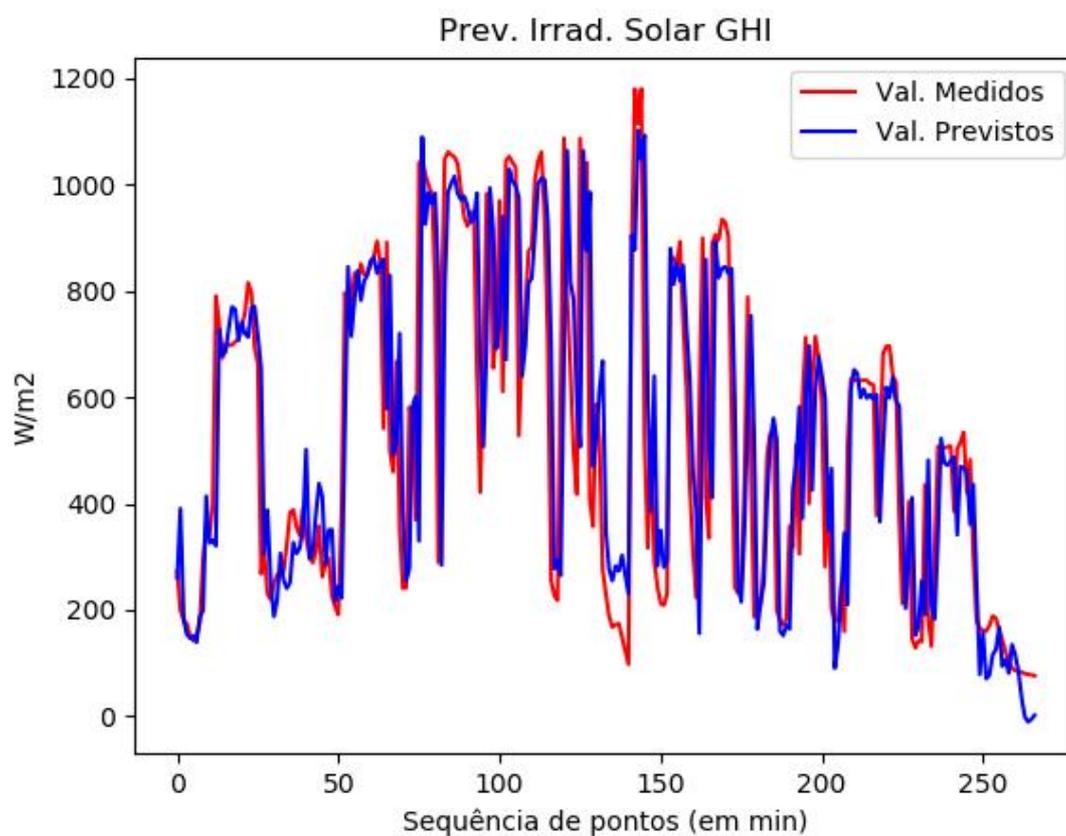
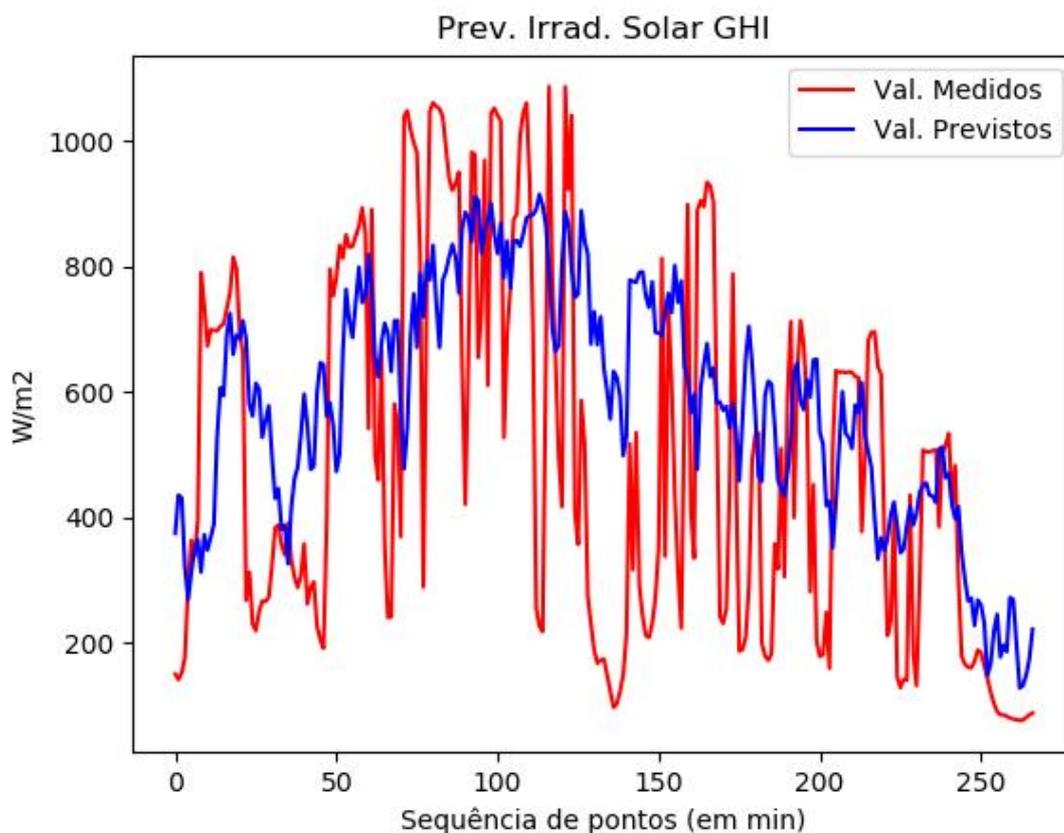


Figura 82 – Gráfico do modelo SVR de predição (GHI) para o horizonte de 5 minutos



Diante do vasto conjunto de modelos existentes para predição da irradiação solar e da possibilidade de grande número de comparações possíveis, há um artigo recente que analisa diversos aspectos relacionados à comparação entre trabalhos que se apresentam como aqueles que alcançam os melhores resultados em termos de predição (YAGLI, YANG e SRINIVASAN, 2019). Esse artigo então trouxe para discussão a necessidade de um trabalho demonstrar o aperfeiçoamento e acurácia de seu método diante de revisores e editores de uma publicação. E como consequência, virtualmente toda publicação reivindica superioridade sobre outras. Segundo YAGLI, YANG e SRINIVASAN, tal fato não seria surpreendente, visto que há uma variedade de formas de se obter a conclusão desejada. Por exemplo:

- Rodar os algoritmos usando uma base de dados cuidadosamente selecionada e manipulada;
- Realizar comparações injustas, tais como: comparar modelos espaço-temporais com modelos de uma variável, modelos híbridos com modelos não híbridos ou modelos bem ajustados com modelos sem ajustes específicos;

- Comparar apenas com modelos básicos como aqueles de persistência simples.

Assim, é abordado o mito da existência do melhor modelo preditivo. Finalmente, os autores sugerem mudar a questão sobre “qual modelo usando ML seria o mais acurado” por “em quais condições um determinado modelo seria apropriado”. Ou “em quais bases a comparação entre modelos seria feita”. Esses problemas não seriam exclusivos à questão de predição solar, mas de Engenharia Solar em geral.

5 CONCLUSÕES E CONSIDERAÇÕES FINAIS

É indiscutível o papel desempenhado por modelos de predição para o desenvolvimento da geração de energia solar em larga escala. Dentro desse cenário, abordagens com diversos graus de complexidade e variado uso de ferramentas matemáticas, estatísticas e computacionais foram vistas no decorrer da revisão de artigos, método que forneceu ideias novas ao mesmo tempo em que foram apresentadas soluções que seguiram linhas de pesquisa bem-sucedidas.

Inicialmente neste trabalho foram definidos os objetivos principais, equipamentos utilizados e em seguida foi concebida a arquitetura de hardware/software para coleta e armazenamento de dados. A câmera *fisheye* exigiu a criação de um programa voltado à calibragem de suas imagens e consequente determinação de seus parâmetros intrínsecos de acordo com o modelo então escolhido, o que teve aplicação direta no momento em que foi necessário construir algoritmos para segmentar o Sol localizando sua posição nas imagens. Essa etapa possibilitou extrair importantes características de pixels relativos à imagem do Sol que foram exploradas na construção de modelos.

Uma vez segmentado o Sol, foi desenvolvido o primeiro modelo criado com o propósito de segmentar as nuvens. Para tanto houve a necessidade de realçar o contraste de imagens e a criação inovadora de algumas funções características ou atributos (f_0 , f_1 e f_4) que apresentaram métricas bastante satisfatórias. Esses procedimentos obtiveram êxito mensurável após a criação do modelo de classificação por ML do tipo SVC, atingindo 99,2% de taxa de acertos.

A partir da segmentação das nuvens foi possível obter outros parâmetros e outras características, como, por exemplo, a fração de cobertura das nuvens. Houve também a criação de um índice estatístico inovador – não encontrado nos artigos consultados - para ser atribuído à cobertura do Sol através da incorporação do método estatístico de Otsu aplicado ao histograma da segmentação do disco solar. Como resultado desta etapa pós-segmentação, outro modelo foi desenvolvido e dessa vez por SVR via regressão para estimar as irradiâncias total e direta associadas a uma imagem. Essas estimativas alcançaram 98% segundo o coeficiente de correlação de Pearson, considerando toda a faixa de incursão dos sinais de irradiância dentro das 6236 amostras do período.

Os modelos preditivos para GHI e DNI desenvolvidos neste estudo para os intervalos de 1, 2, 3, 4, 5, 15 e 30 minutos apresentaram resultados comparáveis a alguns trabalhos publicados, apesar de terem sido realizados em locais diferentes e assim sujeitos às condições meteorológicas típicas de cada local. Além dessa diferença, existem as metodologias particulares usadas para criar cada base de dados. A comparação entre modelos preditivos é um

assunto complexo e envolve diversos aspectos, alguns mais evidentes e outros encobertos por sutilezas, discussão incluída no tema do trabalho de Yagli (YAGLI, YANG e SRINIVASAN, 2019). É mais razoável supor que a comparação entre modelos seria mais justa se, no mínimo, a base de dados fosse a mesma para todos submetidos à análise de desempenho, no entanto não se pode desprezar o fato de que cada região contém dados atmosféricos característicos e peculiares à sua localização. Uma vez que a maioria dos modelos é construída com dados coletados em uma região, um modelo para ser genericamente melhor que tantos outros precisaria ser testado dentro das mesmas condições solares e atmosféricas dos demais, compartilhar a mesma base de dados e as mesmas características espaço-temporais de saída, além de uma mesma natureza não híbrida (ou híbrida). E, ao final, demonstrar maior acurácia preditiva considerando as métricas adequadas.

Quanto ao estudo do movimento das nuvens em termos do deslocamento dos pixels que as representam nas imagens, foram obtidas expressões matemáticas que revelaram a complexidade preditiva desse deslocamento, apesar da hipótese formulada de as nuvens estarem a percorrer trajetórias retilíneas em velocidade constante sobre um plano. Mesmo que essa hipótese seja razoável dentro de um intervalo de tempo curto, ainda há os fenômenos de dissipação ou formação de nebulosidade a serem considerados, o que reforça o emprego das abordagens e modelos de características estatísticas. Outro aspecto a destacar e que foi detectado pelas imagens obtidas é que a velocidade de algumas nuvens poderá ser substancialmente diferente das demais e isso pode decorrer (também) em função da altitude em que se encontram. Assim, a introdução da velocidade das nuvens - em um problema que tem sido majoritariamente tratado de forma estatística - seria mais proveitoso se acompanhado da segmentação individual das nuvens. E para tanto, assumiria grande importância o cálculo de suas alturas por meio de uma visão estereoscópica proporcionada por mais de uma câmera.

Esta pesquisa procurou contribuir cientificamente ao longo das várias etapas de seu desenvolvimento. Em cada cenário particular, estudos e modelos foram criados juntamente com novas metodologias visando a incorporação das informações de imagens aos dados solarimétricos coletados. Parte significativa foi desenvolvida através dos modelos por SVC e SVR, mas é necessário observar que outras ML's, principalmente RNA's, também foram testadas na busca por melhores resultados. Mesmo assim, as configurações atuais de modelagem ainda têm espaço para otimizações. Alternativamente outras metodologias com máquinas de aprendizado que possuam arquiteturas e concepções diferentes podem apresentar melhores resultados, seja no caso de classificação ou de regressão, uma vez que há grande e crescente número de soluções aplicáveis a esse tipo de situação. O conjunto de equipamentos

de reconhecida precisão responsável pela medição das irradiâncias representou não apenas vantagem pela acurácia na coleta de dados a serem modelados, mas também tornou o trabalho mais desafiador e instigante uma vez que os resultados dos modelos seriam comparados nesse mesmo patamar.

Finalmente, resta observar que cada um dos modelos desenvolvidos poderia ser aperfeiçoado através de outras opções de escolha tanto dentro do aspecto de *hardware* como de *software*. Em termos de *hardware*, por exemplo, sugere-se uma câmera com maior profundidade de pixels, ou seja, maior resolução em cada canal de cor para ser testada/comparada diante dos resultados alcançados. Existem câmeras com 10, 12 ou 16 bits de resolução por canal (R,G,B), o que eventualmente resultaria no aumento do contraste e na maior diferenciação entre os pixels. Outro caminho seria implementar uma rede de câmeras distribuídas segundo uma geometria favorável à integração entre suas imagens, compondo-as em uma única fotografia. Essa solução proporcionaria uma visão estereoscópica das nuvens, permitindo medir as suas alturas. Como reflexo desse aperfeiçoamento, no âmbito de *software* essa distribuição de câmeras forneceria pela visualização estereoscópica um atributo adicional para segmentação de nuvens entre si através da medida de suas altitudes. Há, enfim, também um leque de algoritmos detectores de características de pixels e de segmentação sendo desenvolvidos continuamente que, associados aos atributos para classificação ou regressão desenvolvidos nessa pesquisa, teriam condições de levar mais à frente os resultados obtidos neste trabalho.

REFERÊNCIAS

- ALFADDA, A.; RAHMAN, S.; PIPATTANASOMPORN, M. Solar irradiance forecast using aerosols measurements: A data driven approach. *Solar Energy*, v. 170, p. 924-939, 2018.
- ALZHRANI, A.; SHAMSIA P.; DAGLIB, C.; FERDOWSIA, M. Solar Irradiance Forecasting Using Deep Neural Networks. *Procedia Computer Science*, v. 114, p. 304-313, 2017.
- ANEEL. Sala de Imprensa. Leilão de geração “A-4” termina com deságio de 59,07%. Agência Nacional de Energia Elétrica (ANEEL), 2018. Disponível em: <http://www.aneel.gov.br/sala-de-imprensa-exibicao/-/asset_publisher/XGPXSqdMFHrE/content/leilao-de-geracao-a-4-termina-com-desagio-de-59-07-656877?inheritRedirect=false>. Acesso em: 25 fev. 2019.
- BACHER, P.; MADSEN, H.; NIELSEN, H. A. Online short-term solar power forecasting. *Sol Energy*, v. 83(10), p. 1772–1783, 2009.
- BALDWIN, R. R.; COLLINS, K. Cloud Tracking. U.S. Patent US 20110060475, 2011.
- BAY, H.; ESS, A.; TUYTELAARS, T.; GOOL, L. V. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, v.110, n. 3, p. 346-359, 2008.
- BEYER, H. G.; MARTINEZ, J. P.; SURI, M.; TORRES, J. L.; LORENZ, E.; MÜLLER, S. C.; HOYER-KLICK, C.; INEICHEN, P. Report on Benchmarking of Radiation Products. ResearchGate - Management and Exploitation of Solar Resource Knowledge, 2009. Disponível em: <http://www.researchgate.net/publication/265362324_Report_on_Benchmarking_of_Radiation_Products>. Acesso em: 03 out. 2015.
- BOSCH, J. L.; ZHENG, Y.; KLEISSL, J. Deriving cloud velocity from an array of solar radiation measurements. *Solar Energy*, v. 87, p. 196–203, 2013.
- BOUGUET, J. Camera Calibration Toolbox for Matlab. Computational Vision at CALTECH, 2015. Disponível em: <http://www.vision.caltech.edu/bouguetj/calib_doc/>. Acesso em: 25 fev. 2019.
- BRADSKI, G. The OpenCV Library. Dr. Dobb's Journal of Software Tools, 2000.
- BRAGA, A. D. P.; CARVALHO, A. P. D. L.; LUDERMIR, T. B. Redes Neurais Artificiais: Teoria e Aplicações. 2. ed. [S.l.]: LTC, 2014.
- CAPIZZI, G.; NAPOLI, C.; BONANNO, F. Innovative Second-Generation Wavelets Construction with Recurrent Neural Networks for Solar Radiation Forecasting. *Neural Networks and Learning Systems*, IEEE Transactions on, v. 23, p. 1805-1815, 2012. ISSN 2162-237X.
- CHAITANYA, G.; VIKRAM, B. R. Different Type of Segmentation Algorithms of Cloud Detection Using Superpixel Segmentation. *International Journal of Professional Engineering Studies*, v. 5 (5), 2015.

- CHANG, C. C.; LIN, C. J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, v. 2 (3), n. 27, 2011.
- CHOW, C.; URQUHART, B.; LAVE, M.; DOMINGUEZ, A.; KLEISSL, J.; SHIELDS, J.; WASHOM, B. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Solar Energy*, v. 85 (11), p. 2881–2893, 2011.
- DEV, S.; SAVOY, F. M.; LEE, Y. H.; WINKLER, S. Nighttime sky/cloud image segmentation. *In: IEEE International Conference on Image Processing (ICIP)*, Beijing, p. 345-349, 2017.
- DEV, S.; LEE, Y. H.; WINKLER, S. Systematic study of color spaces and components for the segmentation of sky/cloud images. *In: IEEE International Conference on Image Processing (ICIP)*, Paris, p. 5102-5106, 2014.
- DIAGNE, M.; DAVID, M.; LAURET, P.; BOLAND, J.; SCHMUTZ, N. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews*, v. 27, p. 65-76, 2013.
- DRÖNNER, J.; KORFHAGE, N.; EGLI, S.; MÜHLING, M.; THIES, B.; BENDIX, J.; FREISLEBEN, B.; SEEGER, B. Fast Cloud Segmentation Using Convolutional Neural Networks. *Remote Sensing*, v. 10 (11), 1782, p. 1-24, 2018.
- DURRANI, S. P.; BALLUFF, S.; WURZER, L.; KRAUTER, S. Photovoltaic yield prediction using an irradiance forecast model based on multiple neural networks. *Journal of Modern Power Systems and Clean Energy*, v. 6 (2), p. 255-267, 2018. ISSN 2196-5420.
- FABER, N. Estimating the uncertainty in estimates of root mean square error of prediction: Application to determining the size of an adequate test set in multivariate calibration. *Chemometrics and Intelligent Laboratory Systems*, v. 49 (1), p. 79-89, 1999. DOI: 10.1016/S0169-7439(99)00027-1.
- FAN, R.E.; CHEN, P. H.; LIN, C.J. Working set selection using second order information for training SVM. *Journal of Machine Learning Research*, v. 6, p. 1889-1918, 2005.
- FARNEBACK, G. Two-frame motion estimation based on polynomial expansion. *Lecture Notes in Computer Science*, p. 363-370, 2003.
- GLCF, Global Land Cover Facility, 2019. Disponível em: <<http://glcf.umd.edu/data/quickbird/>>. Acesso em: 01 maio 2019.
- GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. 2. ed. [S.l.]: Prentice Hall, 2001.
- GUERRINI, F. SR Schwartz Report, Forbes, 2015. <http://www.schwartzreport.net/>. Disponível em: <<http://www.schwartzreport.net/solar-power-to-become-cheapest-source-of-energy-in-many-regions-by-2025-german-experts-say/>>. Acesso em: 2 nov. 2015.
- HAGGERTY, J.; WANG, X. N.; DICKINSON, A.; O'MALLEY, C.; MARTIN, E. Segmentation of epidermal tissue with histopathological damage in images of haematoxylin and eosin stained human skin. *BMC medical imaging*, v. 14 (7), 2014. DOI: 10.1186/1471-2342-14-7.

- HAMMER, A.; HEINEMANN, D.; LORENZ, E.; LÜCKEHE, B. Short-term forecasting of solar radiation: a statistical approach using satellite data. *Solar Energy*, v. 67, p. 139–150, 1999.
- HAMMER, A.; STOLZENBURG, K. Analyse kurzfristiger Fluktuationen der Solarstrahlung unter Berücksichtigung von Wolkenfeldstrukturen. Carl von Ossietzky University of Oldenburg, Oldenburg, Germany, 1993, 112 p.
- HARPER, J. P. Statistical analysis of solar irradiation in a distributed microgrid. – Universidade da California, San Diego, 2010, 65 p.
- HARVEY, A. C. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, 1989. ISBN 052132196.
- HAYKIN, S. Redes Neurais: Princípios e Prática. 2. ed. [S.l.]: Bookman, 2002.
- HINKELMAN, L. Spatial and temporal variability of incoming solar irradiance at a measurement site in Hawai'i. *In: 91st American Meteorological Society Annual Meeting*, Seattle, WA, United States, p. 23-27, 2011.
- HSU, C. W.; CHANG, C. C.; LIN, C. J. A Practical Guide to Support Vector Machine. Department of Computer Science and Information Engineering of National Taiwan University, 2019. Disponível em: <<https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>>. Acesso em: 25 fev. 2019.
- HUANG, SHUJUN; CAI, NIANGUANG; PACHECO, PEDRO; NARRANDES, SHAVIRA; WANG, YANG; XU, WAYNE. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. *Cancer genomics & proteomics*, v. 15 (1), p. 41-51, 2018. DOI: 21873/cgp.20063.
- IEA. A Snapshot of Global PV Markets. IEA PVPS Snapshots, 2019. Disponível em: <http://www.iea-pvps.org/fileadmin/dam/public/report/statistics/IEA-PVPS_T1_35_Snapshot2019-Report.pdf>. Acesso em: 01 maio 2019.
- IEA. A Snapshot of Global PV. International Energy Agency, Photovoltaic Power Systems Programme, 2015. ISBN 978-3-906043-32-9. Disponível em: <http://www.iea-pvps.org/fileadmin/dam/public/report/technical/PVPS_report_-_A_Snapshot_of_Global_PV_-_1992-2014.pdf>. Acesso em: 03 nov. 2018.
- INMAN, R. H.; PEDRO, H. T. C.; COIMBRA, C. F. M. Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, v. 39 (6), p. 535-576, 2013.
- JAMALY, M.; KLEISSL. Spatiotemporal Interpolation and Forecast of Irradiance Data Using Kriging. *Solar Energy*, v. 158, p. 407-423, 2017. DOI: 10.1016/j.solener.2017.09.057.
- JIANG, H.; DONG, Y. Forecast of hourly global horizontal irradiance based on structured Kernel Support Vector Machine: A case study of Tibet area in China. *Energy Conversion and Management*, v. 142, p. 307-321, 2017. DOI: 10.1016/j.enconman.2017.03.054.
- KNAPP, T. R. Bimodality Revisited. *Journal of Modern Applied Statistical Methods*, v. 6 (1), p. 8-20, 2007. DOI: 10.22237/jmasm/1177992120.

KUMLER, A.; XIE, Y.; ZHANG, Y. A Physics-based Smart Persistence model for Intra-hour forecasting of solar radiation (PSPI) using GHI measurements and a cloud retrieval technique. *Solar Energy*, v. 177, p. 494-500, 2019. DOI: 10.1016/j.solener.2018.11.046.

KURTZ, B. CM2E 2014 (Solar Forecasting with Ground Imagery). Disponível em: <http://cm2e.energie.mq/wp-content/uploads/2014/06/martinique_bkurtz.pdf>. Acesso em: 26 set. 2015.

LAPPALAINEN, K.; VALKEALAHTI, S. Effects of PV array layout, electrical configuration and geographic orientation on mismatch losses caused by moving clouds. *Solar Energy*, v. 144, p. 548-555, 2017. DOI: 10.1016/j.solener.2017.01.066

LAVE, M.; KLEISSL, J.; STEIN, J. S. A Wavelet-Based Variability Model (WVM) for Solar PV Power Plants. *IEEE Trans. Sustain Energy*, v. 4, p. 501–509, 2013.

LIMA, F. J. L.; MARTINS, F. R.; PEREIRA, E. B.; LORENZ, E.; HEINEMANN, D. Forecast for surface solar irradiance at the Brazilian Northeastern region using NWP model and artificial neural networks. *Renewable Energy*, v. 87 (1), p. 807-818, 2016. DOI: 10.1016/j.renene.2015.11.005.

LINGFORS, D.; KILLINGER, S.; ENGERER, N. A.; WIDÉN, J.; BRIGHT, J. M. Identification of PV system shading using a LiDAR-based solar resource assessment model: An evaluation and cross-validation. *Solar Energy*, v. 159, p. 157-172, 2018. DOI: 10.1016/j.solener.2017.10.061.

LIU, S.; ZHANG, L.; ZHANG, Z.; WANG, C.; XIAO, B. Automatic Cloud Detection for All-Sky Images Using Superpixel Segmentation. *IEEE Geoscience and Remote Sensing Letters*, v. 12 (2), p. 354-358, 2015.

LORENZ, E.; REMUND, J.; MÜLLER, S.; TRAUNMÜLLER, W.; STEINMAURER, G.; POZO-VAZQUEZ, D.; RUIZ-ARIAS, J.; LARA FANEGO, V.; SANTIGOSA, L.; GASTÓN, MARTÍN; KURZ, C.; MARTÍN POMARES, L.; GEIJO, C. Benchmarking of different approaches to forecast solar irradiance. *In: 24th European Photovoltaic Solar Energy Conference, Hamburg, Germany*, p. 21-25, 2009.

LOWE, D. Object Recognition from Local Scale-Invariant Features. *In: Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece*, v. 2, p. 1150-1157, 1999.

MARCOS, J.; STORKËL, O.; MARROYO, L.; GARCÍA, M.; PIGUEIRAS, E. Storage requirements for PV power ramp-rate control. *Solar Energy*, v. 99, p. 28-35, 2014.

MARQUEZ, R.; COIMBRA, C. F. M. Intra-hour DNI forecasting based on cloud tracking image analysis. *Solar Energy*, v. 91, p.327–336, 2013.

MARQUEZ, R.; PEDRO, H. T. C.; COIMBRA, C. F. M. Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs. *Solar Energy*, v. 92, p. 176–188, 2013.

MATHEW, JOSEPH. Large Eddy Simulation. *Defence Science Journal* [online], v. 60, n. 6, p. 598-605, 2010. ISSN 0011748X. DOI: 598-605. 10.14429/dsj.60.602.

MATHIESEN, P.; KLEISSL, J. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. *Solar Energy* v. 85 (5), p. 967–977, 2011.

- MELLIT, A.; BENGHANEM, M.; KALOGIROU, S. A. An adaptive wavelet-network model for forecasting daily total solar-radiation. *Applied Energy*, v. 83, p. 705–722, 2006.
- MILLS, A.; WISER, R. Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power. Ernest Orlando Lawrence Berkeley National Laboratory, 2010. LBNL-3884E.
- NALINI, N.J.; PALANIVEL, S. Music emotion recognition: The combined evidence of MFCC and residual phase. *Egyptian Informatics Journal*, v. 17 (1), p. 1-10, 2016. DOI: 10.1016/j.eij.2015.05.004.
- NARAIN, H. K.; SOHI, N. Review: Segmentation Algorithms for Cloud Detection. *International Journal of Engineering Sciences & Research Technology (IJESRT)*, 2015. ISSN: 2277-9655.
- NGUYEN, D.; KLEISSL, J. Stereographic methods for cloud base height determination using two sky imagers. *Solar Energy*, v.107, p. 495–509, 2014.
- NOAA. National Hurricane Center and Central Pacific Hurricane Center, US Dept of Commerce - National Oceanic and Atmospheric Administration, 2019. Disponível em: <<https://www.nhc.noaa.gov/aboutgloss.shtml>>. Acesso em: 18 fev. 2019.
- OTSU, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, v. 9, p. 62–66, 1979.
- PEDRO, H. T. C.; COIMBRA, C. F. M. Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, v. 86 (7), p. 2017–2028, 2012.
- PELLAND, S.; REMUND, J.; KLEISSL, J.; OZEKI, T.; DE BRABANDERE, K. Photovoltaic and Solar Forecasting: State of the Art. International Energy Agency (IEA). Report IEA PVPS T14-01:2013, [S.l.], 2013. ISBN 978-3-906042-13-8. Disponível em: <<http://iea-pvps.org/index.php?id=278>>. Acesso em: 03 nov. 2018
- PELLAND, S.; GALANIS, G.; KALLOS, G. Solar and photovoltaic forecasting through post-processing of the global environmental multiscale numerical weather prediction model. *Progress in Photovoltaics: Research and Applications*, v. 21 (3), p. 284-296, 2013. DOI: 10.1002/pip.1180.
- PLATT, J. C. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. Microsoft Research. Technical Report MSR-TR-98-14, 1998.
- QUESADA-RUIZ, S.; CHU, Y.; TOVAR-PESCADOR, J.; PEDRO, H.T.C.; COIMBRA, C.F.M. Cloud-tracking methodology for intra-hour DNI forecasting. *Solar Energy*, v. 102, p. 267-275, 2014.
- RABL, A. Active solar collectors and their applications. New York: Oxford University Press, 1985, 502 p.
- RAMALHO, R. D. Estudo da variação da potência em um conjunto de sistemas solares distribuídos causada pela passagem de nuvens. Orientador: Prof. Dr. Chiguera Tiba. 2015. 132 p. Dissertação (Mestrado Tecnologias Energéticas e Nucleares), Universidade Federal de Pernambuco, UFPE, DEN, Recife, 2015.
- REDA, I.; ANDREAS, A. Solar position algorithm for solar radiation applications. *Solar Energy*, v. 76(5), p. 557-589, 2004.

- REIKARD, G. Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, v. 83(3), p. 342–349, 2009.
- REINDL, T.; WALSH, W.; YANQIN, Z.; BIERI, M. Energy meteorology for accurate forecasting of PV power output on different time horizons. *Energy Procedia*, v. 130, p. 130-138, 2017. DOI: 10.1016/j.egypro.2017.09.415.
- REN, X.; MALIK, J. Learning a classification model for segmentation. *In: Proceedings Ninth IEEE International Conference on Computer Vision (ICCV)*, Nice, France, v. 1, p.10-17, 2003.
- ROSTEN, E.; PORTER, R.; DRUMMOND, T. Faster and Better: A Machine Learning Approach to Corner Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 32(1), p. 105-119, 2010.
- SAUAIA, R. L. Operador Nacional do Sistema Elétrico. ONS, 2019. Disponível em: <<http://ons.org.br/AcervoDigitalDocumentosEPublicacoes/02-Setor-Dr.RodrigoLopesSauaia-Absolar.pdf>>. Acesso em: 28 abr. 2019.
- SCARAMUZZA, D.; MARTINELLI, A.; SIEGWART, R. A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure from Motion. *In: Fourth IEEE International Conference on Computer Vision Systems (ICVS'06)*, 2006.
- SHEDD, S. HODGE, B. M., FLORITA, A.; ORWIG, K. A Statistical Characterization of Solar Photovoltaic Power Variability at Small Timescales. *In: 2nd Annual International Workshop on Integration of Solar Power into Power Systems*, Lisboa, Portugal, 2012.
- SHI, J.; TOMASI, C. Good Features to Track. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, p. 593-600, 1994.
- SHIELDS, J.; KARR, M. E.; TOOMAN, T. P.; SOWLE, D. H.; MOORE, S. T. The Whole Sky Imager - A Year of Progress. [S.l.]: [s.n.]. 1998. p. 23–27. Presented at the Eighth Atmospheric Radiation Measurement (ARM) Science Team Meeting, Tucson, Arizona.
- SLATER, D. W.; LONG, C. N.; TOOMAN, T. P. Total Sky Imager/Whole Sky Imager Cloud Fraction Comparison. *In: Eleventh ARM Science Team Meeting Proceedings*, Atlanta, Georgia, p. 19-23, 2001.
- TROCCOLI, A.; DUBUS, L.; HAUPT, S. E. *Weather Matters for Energy*. New York, NY: Ed. Springer, 2014, 548 p.
- VAPNIK, V. N. *Statistical Learning Theory*. Wiley-Interscience; 1. ed. 1998, 768 p.
- VAPNIK, V. N. An overview of statistical learning theory. *IEEE Transactions Neural Networks*, v. 10, p. 988-999, 1999.
- WEST, S. R.; ROWE, D.; SAYEEF, S.; BERRY, A. Short-term irradiance forecasting using skycams: Motivation and development. *Solar Energy*, v. 110, p. 188–207, 2014.
- WOOD-BRADLEY, P.; ZAPATA, J.; PYE, J. Cloud tracking with optical flow for short-term solar forecasting. *In: 50th Conference of the Australian Solar Energy Society*, Melbourne, dez. 2012.

WU, Y. K.; CHEN, C. R.; RAHMAN, H. A. A Novel Hybrid Model for Short-Term Forecasting in PV Power Generation. *International Journal of Photoenergy*, v. 2014, ID 569249, 2014, 9 p.

XIE, F.; SHI, M.; SHI, Z.; YIN, J.; ZHAO, D. Multilevel Cloud Detection in Remote Sensing Images Based on Deep Learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, v. 10(8), p. 3631 – 3640, 2017.
DOI: 10.1109/JSTARS.2017.2686488.

YAGLI, G.; YANG, D.; SRINIVASAN, D. Automatic hourly solar forecasting using machine learning models. *Renewable and Sustainable Energy Reviews*, v. 105, p. 487-498, 2019. DOI: 10.1016/j.rser.2019.02.006.

YANG, H.; KURTZ, BEN; NGUYEN, D.; URQUHART, BRYAN; CHOW, C. W.; GHONIMA, M.; KLEISSL, J. Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego. *Solar Energy*, v.103, p. 502–524, 2014.

ZHANG, Z. A Flexible New Technique for Camera Calibration. *Transactions on Pattern Analysis and Machine Intelligence*, v. 22(11), p. 1330-1334, 2000.

ZUIDERVELD, K. Contrast Limited Adaptive Histogram Equalization. *Graphics Gems IV*, 1994.