



Ministério da
**Ciência, Tecnologia
e Inovação**



sid.inpe.br/mtc-m19/2013/08.02.13.06-RPQ

**ESTUDOS PRELIMINARES PARA MODELAMENTO
DA QUANTIDADE DE LÍQUIDO INTEGRADO
VERTICALMENTE UTILIZANDO SATÉLITE, E
VALIDADO POR RADAR DE SUPERFÍCIE**

Sylvio Luiz Mantelli Neto
Leandro Coser
Maximiliano Pessoa Junior
Adiel Mitmann
Mathias Henrique Weber
Eros Comunello
Aldo von Wangenheim
Mário A. Ribeiro Dantas
Izabelly Carvalho da Costa
Carlos Frederico Angelis

Relatório de Pesquisa.

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/3EJ78EL>>

INPE
São José dos Campos
2013

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE (RE/DIR-204):

Presidente:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Membros:

Dr. Antonio Fernando Bertachini de Almeida Prado - Coordenação Engenharia e Tecnologia Espacial (ETE)

Dr^a Inez Staciarini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Dr. Germano de Souza Kienbaum - Centro de Tecnologias Especiais (CTE)

Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Maria Tereza Smith de Brito - Serviço de Informação e Documentação (SID)

Luciana Manacero - Serviço de Informação e Documentação (SID)



Ministério da
**Ciência, Tecnologia
e Inovação**



sid.inpe.br/mtc-m19/2013/08.02.13.06-RPQ

**ESTUDOS PRELIMINARES PARA MODELAMENTO
DA QUANTIDADE DE LÍQUIDO INTEGRADO
VERTICALMENTE UTILIZANDO SATÉLITE, E
VALIDADO POR RADAR DE SUPERFÍCIE**

Sylvio Luiz Mantelli Neto
Leandro Coser
Maximiliano Pessoa Junior
Adiel Mitmann
Mathias Henrique Weber
Eros Comunello
Aldo von Wangenheim
Mário A. Ribeiro Dantas
Izabelly Carvalho da Costa
Carlos Frederico Angelis

Relatório de Pesquisa.

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/3EJ78EL>>

INPE
São José dos Campos
2013

Dados Internacionais de Catalogação na Publicação (CIP)

Marinke, Rogério.

Estudos preliminares para modelamento da quantidade de líquido integrado verticalmente utilizando satélite, e validado por radar de superfície / Sylvio Luiz Mantelli Neto

Leandro Coser

M339g Maximiliano Pessoa Junior

Adiel Mitmann

Mathias Henrique Weber

Eros Comunello

Aldo von Wangenheim

Mário A. Ribeiro Dantas

Izabelly Carvalho da Costa

Carlos Frederico Angelis . – São José dos Campos : INPE, 2013.

xxii + 73 p. ; (sid.inpe.br/mtc-m19/2013/08.02.13.06-RPQ)

Tese (Mestrado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2011.

Orientadores : Drs. Nandamudi Lankalapalli Vijaykumar, e Edson Luiz França Senne.

1. teste estrutural. 2. programa concorrente. 3. diagrama de Estado. I.Título.

CDU 004.415.532.2



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](https://creativecommons.org/licenses/by-nc/3.0/).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).

"The computer is not a competitor for the human brain. The computer is just a big, patient number cruncher. It doesn't get bored. It doesn't get sleepy and make mistake. It will do the same taxing calculation over and over and over and over - for several millions "overs" - again."

DAVID SALSBERG ON THE LADY TASTING TEA (2006,
PAG. 68).

AGRADECIMENTOS

Os autores agradecem ao grupo de pessoas envolvido nas atividades deste trabalho mas especialmente para as seguintes pessoas.

- a) Aos membros do grupo de pesquisas do UFSC-INE-LAPIX, que auxiliaram no desenvolvimento das técnicas de processamento de imagens, especialmente para M.Sc. Antonio Carlos Sobieranski.

Home page do LAPIX <http://www.lapix.ufsc.br/>

- b) Membros da equipe operacional do INPE-DSA que auxiliou na obtenção de imagens de radar e esclareceu dúvidas acerca de algumas técnicas afins, mas especialmente para Wagner Flauber Rocha.

Home page do INPE-DSA sigma.cptec.inpe.br/radar/

- c) Ao grupo operacional do DECEA, Departamento de Controle do Espaço Aéreo do Ministério da Aeronáutica.

Home page do DECEA <http://www.fab.mil.br/portal/capa/index.php?page=decea>

- d) Ao suporte financeiro do CNPQ no pagamento da bolsa de Adiel Mittman. Processo CNPQ: 380649/2012-8, durante o programa de incentivo aos INCTs relativo ao INCoD (Intituto Nacional de Ciência e Tecnologia para a Convergência Digital), e da Rede Clima.

Home page do programa INCT do CNPQ http://estatico.cnpq.br/portal/programas/inct/_apresentacao/

Home page da Rede Clima <http://redeclima.ccst.inpe.br/org.html>

- e) Ao suporte técnico operacional e científico fornecido pelo CEMADEN (Centro Nacional de Monitoramento e Alertas de Desastres Naturais).

Home page do CEMADEN <http://www.cemaden.gov.br/>

**PRELIMINARY STUDIES FOR MODELING VERTICALLY
INTEGRATED LIQUID USING SATELLITE, VALIDATED BY
SURFACE RADAR.**

ABSTRACT

This technical report describes the development of a model to estimate the Vertically Integrated Liquid (VIL) on the atmosphere using GOES-12 satellite, and 80 surface radar images (Z) obtained on April 25th 2012 from Morro da Igreja in Santa Catarina State, Brazil. Images from São Roque radar in State of São Paulo were used for model validation. Exploratory Data Analysis (EDA), Principal Component Analysis (PCA) and Design and Analysis of Experiments, were used to assess the domain and significance of variables and minimize uncontrolled factors on modeling. During this work it was developed routines for 3-D image reading and processing in C++ and resizing techniques to use images in the same spatial scale for statistical analysis. The results indicated the existence of several patterns (multi modal), related to different sources of intense precipitation. Linear regression modeling used described only 13.2% of observed phenomena variability and not very appropriate as preliminarily indicated. To improve this analysis it is recommended a new modeling paradigm and a new experiment, analyzing the patterns separately under the supervision of a meteorologist. According to the results, it is suggested the use of Gaussian Mixture Models (GMM) or equivalent ones instead of Linear Regression Models for VIL estimations.

RESUMO

Neste trabalho foi realizado o desenvolvimento de um modelo para a estimativa do conteúdo de água líquida integrada verticalmente (VIL) na atmosfera a partir dos dados obtidos das imagens do satélite GOES-12 e dos valores de Z de 80 imagens de superfície obtidas no dia 25 de abril de 2012 do Radar do Morro da Igreja em Santa Catarina. As Imagens do Radar de São Roque em São Paulo foram utilizadas na validação do modelo desenvolvido. Durante o modelamento foram utilizadas técnicas estatísticas de Análise Exploratória de Dados (EDA), de Componentes Principais (PCA) e de Projeto e Análise de Experimentos (PAE); para conhecer o domínio e a significância das variáveis independentes propostas e minimizar a influência de fatores incontroláveis no modelo. Durante o trabalho foram desenvolvidas rotinas de leitura e processamento em C^{++} de imagens 3-D de radar e estudadas técnicas de redimensionamento das imagens para serem utilizadas na mesma escala espacial durante a análise estatística. Os resultados indicaram a existência de vários padrões (multi modais), coincidentes à ocorrência de eventos de grande precipitação de diversas origens. O modelo de regressão linear proposto para todos os padrões em conjunto, descreve somente 13,2% da variabilidade dos fenômenos observados e não é muito adequado como preliminarmente indicado. Para melhorar esta valiação, recomenda-se um novo paradigma de modelamento em um novo experimento mais refinado analisando separadamente os padrões, supervisionado por um especialista em meteorologia. De acordo com os resultados obtidos, sugere-se utilizar um modelo de Misturas Gaussianas (GMM) ou similar ao invés do modelo de regressão linear para as estimativas de VIL.

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO	1
2 MATERIAIS E MÉTODOS	3
2.1 As imagens de radar do Morro da Igreja utilizadas no modelamento	4
2.2 As imagens de satélite do GOES-12 utilizadas no modelamento	5
2.3 Processamento preliminar de imagens e obtenção de parâmetros de análise	6
2.4 Análise Estatística	13
2.4.1 Análise Exploratória de dados	14
2.4.2 Análise dos componentes principais ou PCA	16
2.4.3 Modelo de regressão linear múltipla	18
2.4.4 Verificação da adequação do modelo de regressão	19
2.4.5 Teste de significância do modelo de regressão linear múltipla	19
2.4.5.1 Análise residual	20
2.4.5.2 Análise das superfícies de resposta	22
2.5 Validação	23
2.6 Comentários acerca da análise estatística	25
3 CONCLUSÕES	27
REFERÊNCIAS BIBLIOGRÁFICAS	29

LISTA DE FIGURAS

	<u>Pág.</u>
2.1 Ilustração acerca da visualização geral do presente método utilizado . . .	4
2.2 Ilustração acerca do pré-processamento utilizado para obter os parâmetros de modelamento.	7
2.3 Ilustração acerca da região de análise utilizada no entorno do radar realçada em azul, para a confecção do modelo. 14×14 e 38×38 em pixels correspondem a áreas de 56×56 e 152×152 em <i>km</i> . OBSERVAÇÃO. A figura encontra-se fora de proporção especialmente na dimensão vertical para ilustrar melhor o procedimento utilizado.	13
2.4 Gráficos dos dados obtidos simultaneamente a partir de 80 amostras de imagens dos canais do GOES, dos seus dados derivados e também do cálculo do VIL limitado à região de maior sensibilidade a partir das imagens do radar.	15
2.5 Gráfico do tipo <i>scree plot</i> , ilustrando a importância de cada regressor através dos autovalores na análise de componentes principais.	17
2.6 Figura indicando a análise dos resíduos do modelo de regressão utilizado.	21
2.7 Figura indicando os graficos de contorno do VIL em relação aos canais do GOES e seus derivados. As unidades de VIL estão em $kg.m^{-2}$, as de C3 e C4 em K e X3 e X4 são adimensionais.	22
2.8 Ilustração acerca da AED realizada nos dados do cálculo do VIL utilizados na validação.	23
2.9 Figura comparando estatisticamente os resultados do cálculo do VIL entre o modelo de regressão obtido pelo satélite GOES e pelo radar de São Roque.	24

LISTA DE TABELAS

	<u>Pág.</u>
2.1 Tabela indicando os detalhes do imageador do GOES. FONTE: http://noaasis.noaa.gov/NOAASIS/ml/imager.html	6
2.2 Tabela comparando a escolha da matriz básica de redimensionamento para a região de alcance do radar em relação à imagem do GOES.	10
2.3 Vantagens e desvantagens dos métodos de interpolação candidatos a <i>re-sampling</i> levantados.	12
2.4 Análise dos autovalores pela matriz de covariância	17
2.5 Análise dos autovalores pela matriz de correlação	17
2.6 matriz de covariância	18
2.7 matriz de correlação	18
2.8 Tabela dos coeficientes encontrados no estabelecimento da regressão linear múltipla.	19
2.9 Tabela da análise de variância do modelo de regressão linear múltipla. . .	19

LISTA DE ABREVIATURAS E SIGLAS

AED	–	Análise Exploratória de dados
alt	–	Altitude em metros
BIN	–	BINS do radar
CEMADEN	–	Centro Nacional de Monitoramento e Alerta de Desastres Naturais
CCST	–	Centro de Ciências do Sistema Terrestre do INPE
DN	–	Digital Number, número digital obtido a partir das imagens
GOES-12	–	Geostationary Operational Environment Satellite, número 12
HDF Viewer	–	Ferramenta de visualização de dados HDF-5 em código aberto
HDF-5	–	Hierarchical Data Format version 5
INCoD	–	Instituto Nacional de Ciência e Tecnologia para a Convergência Digital
INE	–	Departamento de Informática e Estatística da UFSC
INPE	–	Instituto Nacional de Pesquisas Espaciais
LAPIX	–	Laboratório de Processamento de Imagens e Computação Gráfica do INE
lat	–	Latitude em graus
lon	–	Longitude em graus
MCT	–	Ministerio da Ciência e Tecnologia
MVOL	–	Arquivo de saída do radar, referenciado a partir de sua extensão
OPENCV	–	Biblioteca em código aberto na linguagem C^{++} utilizada em processamento de imagens
OUTLIERS	–	Dados anormais com problemas ou fora da escala de observação
PCA	–	Análise de Componentes principais
R	–	Pacote estatístico em código aberto
SC	–	Estado de Santa Catarina, Brasil
SP	–	Estado de São Paulo, Brasil
UFSC	–	Universidade Federal de Santa Catarina Florianópolis SC Brasil
VCP	–	Padrões de cobertura de volume executada pelo radar ou <i>Volume Coverage Patterns</i>
VIL	–	Conteúdo de água líquida integrada verticalmente ou <i>Vertically Integrated Liquid</i>
2-D	–	Imagens ou dados apresentados em duas dimensões
3-D	–	Imagens ou dados apresentados em três dimensões

LISTA DE SÍMBOLOS

- a – é o raio médio da terra em metros
- C3 – Canal de vapor de água do GOES-12
- C4 – Canal de infra-vermelho do GOES-12
- C43 – Diferença entre os canais de infra-vermelho e de vapor de água do GOES-12 em graus kelvin
- dBZ – valor da refletividade do radar em decibéis ver equacao 2.3
- h – altura do feixe expressa em m
- r – é a linha de alcance do radar em ???
- VIL – Conteúdo de água líquida em kgm^{-2}
- X3 – Variância do canal de vapor de água do GOES-12 a partir de uma matriz de 9×9 pixels em graus kelvin
- X4 – Variância do canal de infra-vermelho de água do GOES-12 a partir de uma matriz de 9×9 pixels em graus kelvin
- Z – valor da refletividade obtido a partir da imagem do radar em mm^6m^{-3}
- θ – Altura expressa em metros do feixe do radar

1 INTRODUÇÃO

O presente documento é um relatório parcial de pesquisas, que estão sendo desenvolvidas como parte de uma cooperação científica entre o Instituto Nacional de Pesquisas Espaciais (INPE), o Centro de Monitoramento e Alerta de Desastres Naturais (CEMADEN) e o Instituto Nacional de Ciência e Tecnologia para a Convergência Digital (INCoD) da Universidade Federal de Santa Catarina (UFSC). O presente trabalho também faz parte das atividades desenvolvidas pela Rede Clima.

O objetivo deste relatório é descrever as atividades de pesquisas em andamento, para o desenvolvimento de um modelo de estimativa do conteúdo de água líquida integrada verticalmente (VIL) [Greene e Clark \(1972\)](#) presente na atmosfera utilizando simultaneamente imagens de satélite e radar. O modelo desenvolvido será utilizado para estimar o VIL fora das regiões de cobertura do radar a partir das imagens do satélite GOES-12. Serão utilizados os canais do satélite GOES-12 C3 (Vapor de água), C4 (Infra-vermelho), a variância destes dois canais, obtidos a partir de uma matriz com 3×3 pixels denominadas de X3 e X4 e a diferença entre eles C43. O VIL será calculado a partir das imagens 3-D do radar meteorológico do Morro da Igreja em SC. A pesquisa prevê também o desenvolvimento de rotinas de leitura e processamento das imagens 3-D do radar para o cálculo do VIL. Todos estes requisitos ou parâmetros foram previamente estabelecidos pelos pesquisadores do INPE e do CEMADEN. Foi definido também, utilizar o desenvolvimento de um modelo de regressão linear múltipla e de técnicas estatísticas multivariadas durante os trabalhos para estimar o VIL. Esta abordagem paramétrica fornecerá um modelamento matemático mais rigoroso para a estimativa do VIL. O modelo que estima o VIL será validado a partir dos dados do Radar de São Roque em São Paulo e os resultados desta comparação serão apresentados no decorrer do presente trabalho.

Para o desenvolvimento da presente pesquisa está sendo utilizado o método *Bayesiano* [Yuille e Kersten \(2006\)](#), [Charter et al. \(2006a\)](#), [Charter et al. \(2006b\)](#) em duas fases: análise supervisionada e classificação [NIST \(2009, cap. 1\)](#). Não estão sendo utilizadas redes neurais, por causa do aspecto caixa-preta, envolvidos no processo de modelamento [Johannet et al. \(2007\)](#), [Zhang \(2007\)](#). Depois que o modelo *Bayesiano* estiver implementado e consolidado, será possível obter as condições necessárias para a validação do funcionamento da rede neural, embora não seja este objetivo do presente trabalho de pesquisa. O presente documento está organizado de acordo com a seguinte estrutura.

Capítulo 1 Apresenta uma introdução e descrição sumária do presente relatório.

Capítulo 2 Apresenta os materias e métodos utilizados durante a pesquisa.

Capítulo 3 Descreve os resultados preliminares utilizados durante a análise do presente modelo em estudo, além da validação.

Capítulo 4 Conclusão.

2 MATERIAIS E MÉTODOS

Neste capítulo serão descritos os métodos utilizados no modelo estatístico de estimativa do VIL. Antes do modelamento propriamente dito foi necessário o desenvolvimento de um conjunto de rotinas para extrair os parâmetros das imagens de radar correspondentes à área de cobertura do satélite. Os dados do VIL (variável dependente do modelo) foram obtidos a partir de 80 imagens do radar meteorológico, localizado no Morro da Igreja em Santa Catarina (SC) (lat. = -28.1285956° , lon. = -49.47176413° alt. = 1822.0 m). A validação foi feita a partir de imagens do radar meteorológico localizado em São Roque em São Paulo (SP) (lat. = -23.6019° , lon. = -47.0943° , alt. = 1150.0 m) e das respectivas imagens simultâneas do satélite GOES-12 (variáveis independentes do modelo). A partir destas imagens foi estabelecida uma metodologia para a análise supervisionada, em que a partir de condições conhecidas pode-se definir o resultado da análise. Em seguida os modelos foram avaliados qualitativamente e finalmente estabelecidas as estimativas do VIL fora da área de cobertura do radar.

Este capítulo está estruturado nas seguintes seções.

Seção 2.1 : descreve preliminarmente como são obtidas as imagens de radar e a varredura utilizadas no presente trabalho.

Seção 2.2 : descreve sumariamente as imagens de satélite utilizadas no presente trabalho.

Seção 2.3 : descreve sumariamente o pré-processamento necessário para obter os parâmetros obtidos a partir das imagens que serão utilizadas no modelo. Esta seção também aborda o procedimento de redimensionamento e o volume de maior sensibilidade do radar utilizado para o modelamento.

Seção 2.4 : descreve a análise estatística dos dados.

Seção 2.5 : descreve a validação do modelo calculado na seção anterior utilizando os dados do Radar de São Roque.

Seção 2.6 : descreve uma análise dos dados e do modelo proposto com comentários acerca dos resultados.

Uma figura geral descrevendo a metodologia utilizada no método proposto, encontra-se ilustrado na figura 2.1. Conforme mencionado na introdução deste relatório será utilizado o método Bayesiano que possui duas fases: análise supervisionada e classificação NIST (2009, cap. 1).

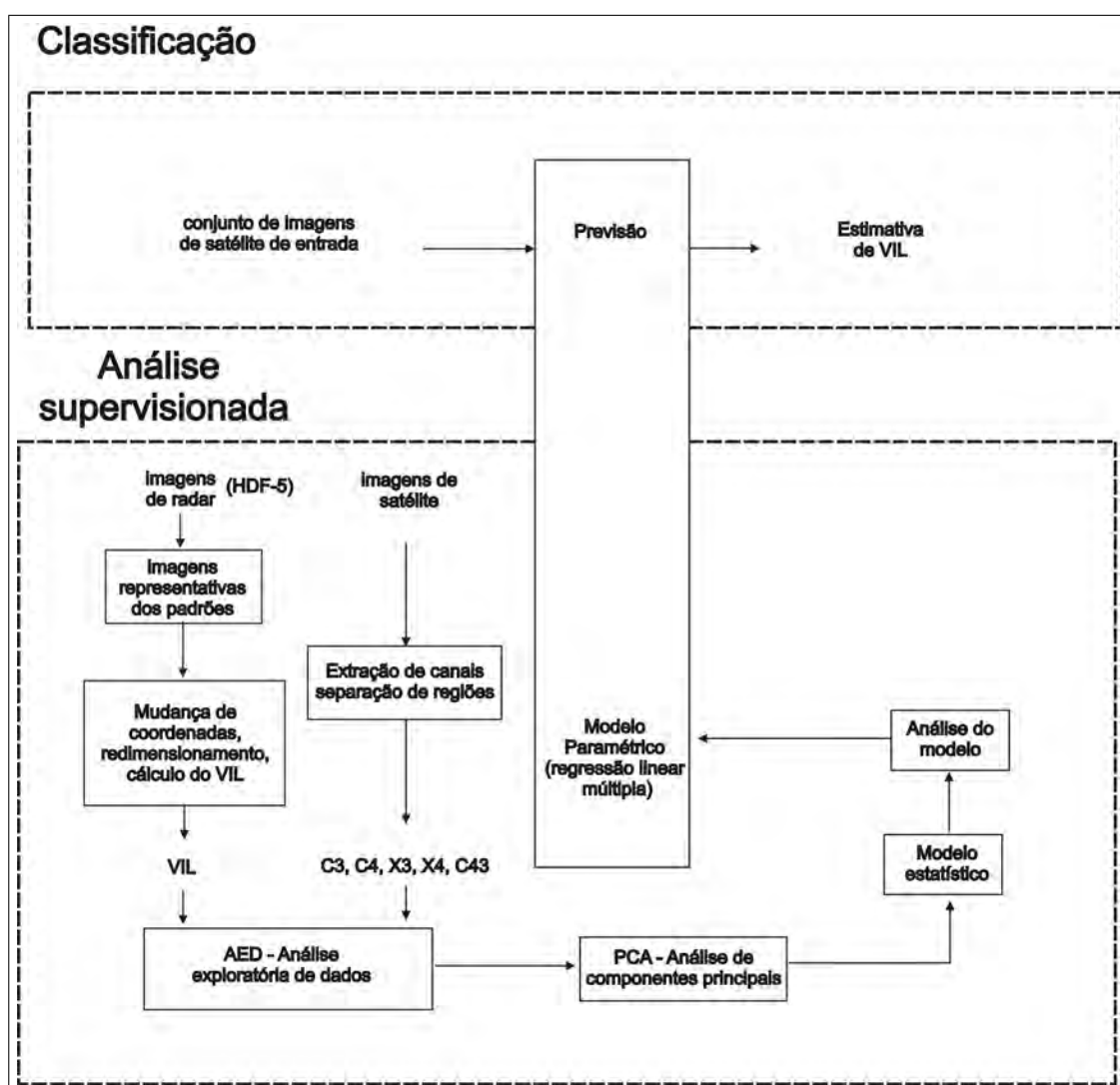


Figura 2.1 - Ilustração acerca da visualização geral do presente método utilizado

2.1 As imagens de radar do Morro da Igreja utilizadas no modelamento

O radar emite ondas a partir de um conjunto de trem de pulsos de alta potência (aproximadamente 450.000 W) em determinados ângulos de azimute e elevação e analisa o retorno destas reflexões. Estas ondas são emitidas através de um conjunto

de feixes cônicos com um ângulo de abertura de $1,98^\circ$, a cada grau de azimute horizontal (360 ângulos) para um determinado número de elevações chamados de padrões de cobertura de volume ou *Volume Coverage Patterns* (VCP). Estas elevações são estabelecidas de acordo com o *modus operandi* da equipe operacional do radar e também de acordo com as condições atmosféricas vigentes. Em geral para as condições de céu claro são monitoradas menos elevações do que nas condições de precipitação. Este procedimento é adotado pelas equipes operacionais para evitar armazenamento de informações desnecessárias, racionalizar o uso da energia, preservar a vida útil do radar, etc. Mas nas imagens disponibilizadas pela equipe operacional do Ministério da Aeronáutica, somente foram utilizadas imagens com somente um tipo de VCP com o maior número de elevações. O VCP configurado utiliza as seguintes elevações.

0°	$0,5^\circ$	$1,5^\circ$	$2,5^\circ$	$3,5^\circ$
$4,5^\circ$	$5,5^\circ$	$6,5^\circ$	$7,5^\circ$	$8,5^\circ$
$9,5^\circ$	$11,5^\circ$	$13,5^\circ$	$15,5^\circ$	$17,5^\circ$

Desta maneira o radar obtêm uma amostragem espacial ao seu redor em um determinado raio de alcance que pode chegar a 250 km. O resultado desta análise é um conjunto de imagens 3-D fornecidas no formato HDF-5 ¹ em coordenadas cilíndricas, além de outros parâmetros operacionais do radar. Os arquivos do radar possuem extensão MVOL, e foram visualizados através da ferramenta HDF Viewer ². O mesmo padrão de VCP foi utilizado para o conjunto de imagens do radar do Morro da Igreja utilizadas na análise supervisionada e na validação para o radar de São Roque, para maximizar a sensibilidade do modelo.

2.2 As imagens de satélite do GOES-12 utilizadas no modelamento

O Satélite GOES-12 ³ é utilizado basicamente para a previsão do tempo na América do Sul e monitorar o movimento de tempestades, ciclones e deslizamentos. As imagens do satélite GOES-12 são obtidas a partir de uma órbita aproximadamente a 36.370 km acima do solo na linha do equador. A tabela 2.1 descreve os sensores

¹HDF-5 Hierarchical Data Format, desenvolvido pelo National Center for Supercomputing Applications (NCSA) <http://www.ncsa.illinois.edu/>

²HDF Viewer é desenvolvido pelo *The HDF Group* em código aberto está disponível no link <http://www.hdfgroup.org/hdf-java-html/hdfview/>

³O GOES é administrado pela agência Americana NOAA (National Oceanic and Atmospheric Administration) http://www.noaaneews.noaa.gov/stories2010/20100525_satellite.html

correspondentes aos canais disponíveis neste satélite. O canal C1 visível (Vis), o C2 ondas curtas (OC), o C3 vapor de água ou umidade (Um) e dois C4 e C5 infra-vermelhos (Iv1 e Iv2). Conforme pode ser observado, o canal C3 possui resolução espacial distinta do C4.

Tabela 2.1 - Tabela indicando os detalhes do imageador do GOES. FONTE: <http://noaasis.noaa.gov/NOAASIS/ml/imager.html>

Canal	Sensor				
	1 (Vis)	2 (OC)	3 (Um)	4 (Iv1)	5 (Iv2)
Comp. de Onda em μm	0,55-0,75	3,80-4,00	6,50-7,00	10,20-11,20	11,50-12,50
resolução em km	1	4	8	4	4
incerteza	5% max	$Iv \leq 1K$			

A partir dos canais C3 vapor de água e C4 infra-vermelho, foram obtidas as imagens e dados, calculadas a variabilidade dos canais respectivamente denominados X3 e X4 e a sua diferença denominado C43. A variabilidade foi calculada a partir de uma matriz de 3×3 pixels ao redor do pixel central.

2.3 Processamento preliminar de imagens e obtenção de parâmetros de análise

Antes de utilizar os parâmetros necessários para o modelamento é necessário extraí-los das imagens. Dois tipos de imagens diferentes foram utilizadas: as do satélite GOES-12 e as de radar. Cada uma delas necessita de uma rotina específica de leitura. Foram desenvolvidos algoritmos de leitura destes arquivos em linguagem C++ e utilizados um conjunto de bibliotecas em código aberto específica para aplicações em Visão Computacional chamada OPENCV ⁴. Um esquema geral ilustrando o procedimento utilizado no pré-processamento pode ser observado na figura 2.2.

Para a confecção de um modelo preliminar, foram utilizadas 80 imagens do dia 25 de abril de 2012, obtidas a cada 15 minutos do radar do Morro da Igreja em Santa Catarina.

O volume ao redor do radar não é completamente amostrado, permitindo por esta razão, a ocorrência de um cone de silêncio e lacunas (ou pixels ausentes) de dados

⁴OPENCV está disponível no link <http://opencv.org/>

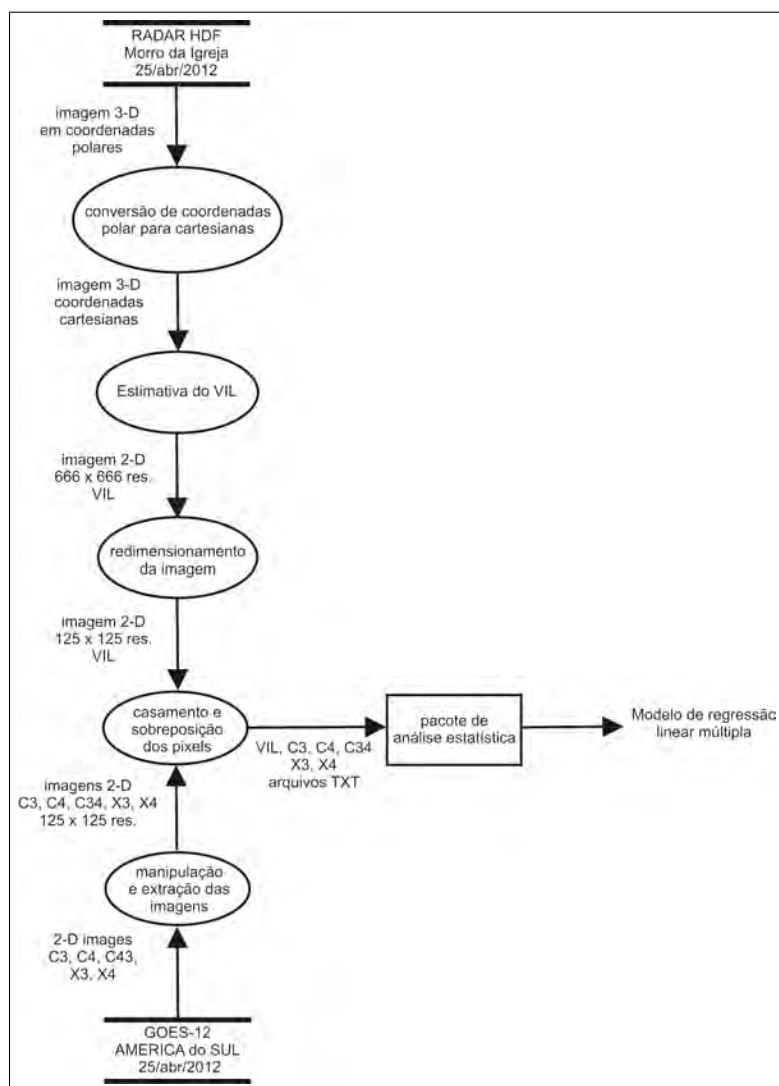


Figura 2.2 - Ilustração acerca do pré-processamento utilizado para obter os parâmetros de modelamento.

nas imagens. O cone de silêncio localiza-se na área imediatamente acima do radar, e é descartado do processamento de imagem durante o modelamento. As lacunas de dados ocorrem por dois motivos. Em primeiro lugar porque o radar não consegue varrer completamente a região ao seu redor e em segundo lugar por causa do processo de conversão de coordenadas esféricas para cartesianas. Estas lacunas poderão ser preenchidas completamente em um trabalho futuro, através de interpolação logarítmica Deserno (March 24, 2004), porque os dados de radar nas imagens MVOL se apresentam nesta escala logarítmica em decibéis. Atualmente a interpolação é aplicada somente no mesmo plano de elevação 2-D. A interpolação logarítmica 3-D

esta prevista em outra fase deste projeto de pesquisas. Logo em seguida, é efetuado o cálculo do conteúdo de água líquida, a partir de uma equação de conversão entre o valor da refletividade Z e o VIL fornecida pelos pesquisadores do CEMADEN. A equação 2.1 e baseada no método desenvolvido por [Greene e Clark \(1972\)](#), utilizada no presente trabalho encontra-se ilustrada a seguir.

$$VIL = 3.44 \times 10^{-6} \int_{base}^{topo} \left(\frac{Z_1 + Z_2}{2} \right)^{\frac{4}{7}} dh \quad (2.1)$$

em que:

Z_1 : é o valor da refletividade do radar em mm^6m^{-3} .

Z_2 : é o valor da refletividade do radar da camada superior em mm^6m^{-3} .

VIL : tem unidades em $kg.m^{-2}$ e representa a massa de gotas em um volume unitário

h : é a altura do feixe, obtida a partir da equação 2.2 expressa em metros.

$$h = \frac{r^2 \times \cos^2 \theta}{\frac{8}{3} \times a} + r \times \sin \theta \quad (2.2)$$

em que:

θ : é o ângulo de elevação em radiano.

r : é a linha de alcance do radar em metros.

a : é o raio da Terra (6.378.137) em metros ⁵ para análise dos dados.

Os valores de Z podem assumir valores em uma grande variabilidade de escala que pode ir desde 0,001 mm^6m^{-3} para neblina até valores da ordem de 40.000.000 mm^6m^{-3} para granizo de tamanho grande [WMO \(2005, pag. 13\)](#). Por esta razão

⁵O valor do raio da Terra foi obtido a partir de WGS 84 (World Geodetic System of 1984) http://home.online.no/~sigurdhu/WGS84_Eng.html

os valores de Z são expressos em escala logarítmica ou em dBZ em um intervalo aproximadamente entre -30 a $+76$ dBZ.

Os dados de radar em dBZ podem ser obtidos a partir da equação 2.3

$$dBZ = 10 \times \log_{10} Z \quad (2.3)$$

Quando os dados forem fornecidos em dBZ pode-se utilizar a formula descrita na equação 2.4 para o no cálculo do VIL a partir dos arquivos MVOL.

$$Z = 10^{\frac{dBZ}{10}} \quad (2.4)$$

Todas as imagens de radar com o VIL calculado, possuíam imagens correspondentes do satélite GOES-12 formando tuplas ⁶ de dados. O número de tuplas em cada imagem, pode variar de uma condição atmosférica de céu claro ou na presença de tempestades, e depende do tipo e da magnitude do fenômeno observado. No presente trabalho a título de modelamento, foram utilizadas condições fixas. A partir destas tuplas foram realizadas as análises estatísticas.

As imagens de radar e as de satélite não possuem a mesma resolução espacial. A resolução espacial do radar é de $0,75km/pixel$, enquanto que a de satélite é de $4km/pixel$. Antes de utilizá-las é necessário convertê-las para a mesma resolução através de um procedimento conhecido como redimensionamento. O método sugerido preliminarmente pelo grupo de pesquisas do CEMADEN foi pela média dos pixels, para reduzir a dispersão dos dados. Mas a média dos dados sem o redimensionamento não permite o casamento exato na geometria das imagens causando fracionamento (por exemplo encaixar 5,3 pixels de radar em 1 pixel de satélite). Nestas condições o uso do valor médio poderia ocasionar erros de aproximação na atribuição de valores por descasamento geométrico.

Para resolver este problema foi escolhida uma matriz básica quadrada de valores médio, máximo e mínimo dos pixels do VIL, de maneira a obter uma resolução mais próxima possível entre as duas imagens preservando a geometria quadrada.

⁶Tupla é um termo técnico derivado do inglês, comumente utilizado na área de computação. Representa um conjunto ordenado de dados. Fonte <http://www.merriam-webster.com/dictionary/-tuple>

Por exemplo se utilizarmos uma matriz básica de 5×5 pixels obteremos um múltiplo de pixels mais próximo da resolução da imagem do satélite para a região de análise. A tabela 2.2 indica o tamanho da matriz básica e o tamanho da imagem obtida em relação a imagem do GOES.

Tabela 2.2 - Tabela comparando a escolha da matriz básica de redimensionamento para a região de alcance do radar em relação à imagem do GOES.

Matriz básica	Resolução final	Diferença
GOES	125	0
4×4	166	41
5×5	133	8
6×6	166	14

Desta maneira o redimensionamento por interpolação fica restrita à menor diferença possível de 8 pixels entre as imagens. Para o redimensionamento de uma imagem de 133×133 para uma de 125×125 foram utilizadas rotinas de redimensionamento por interpolação disponíveis na biblioteca OPENCV. Vale a pena destacar que a interpolação por redimensionamento é diferente da interpolação logarítmica mencionada anteriormente, para o preenchimento das lacunas dos dados da varredura do radar. Cinco métodos de redimensionamento por interpolação para uma resolução maior ou menor estão disponíveis na biblioteca OPENCV, conforme descrito a seguir.

- **NEAREST**: nearest neighbor.
- **LINEAR**: bilinear (default).
- **AREA**: resampling.
- **CUBIC**: bicubic 4×4 pixel neighborhood.
- **LANCZOS4**: Lanczos over 8×8 pixel neighborhood.

Vale a pena salientar que durante o redimensionamento da imagem, não podem haver lacunas de dados, pois estas causam erros semânticos durante a interpolação utilizada no redimensionamento. Esta é outra razão pela qual as imagens são interpoladas somente em 2-D. Por exemplo, a interpolação CUBIC utiliza a vizinhança

4×4 dos pixels. Se houver uma ou mais lacunas, a interpolação apresentará resultados matematicamente válidos porém sinteticamente gerados. As lacunas ocorrem predominantemente entre os níveis de elevações. No presente trabalho o redimensionamento é aplicado somente nas imagens planas (ou que se encontram no mesmo plano) e não em imagens volumétricas em diferentes altitudes.

Cada um dos métodos de redimensionamento mencionados foi utilizado, e todos eles apresentaram resultados diferentes nas imagens. Por isto uma análise mais detalhada precisou ser realizada, para saber quais eram as diferenças entre eles. Lillesand e Kiefer (1994, cap. 7.2), comparam os métodos de redimensionamento aplicados ao processamento de imagens em sensoriamento remoto e os chama de *resampling*. Em geral o *resampling* é utilizado para aumentar o tamanho das imagens *upsampling* ou reduzir *downsampling* como no presente trabalho. Entretanto o mais importante a ser analisado, são os efeitos causados nas imagens de acordo com a aplicação. No próximo parágrafo serão analisadas algumas destas características para saber qual dos métodos é o mais apropriado para a presente pesquisa.

Lillesand e Kiefer (1994, sect. 7.2), Schowengerdt (2007, sect. 7.2.3), Dodgson (1992) e Parker et al. (1983), Richards (1995, cap. 2) mencionam os mesmos métodos de interpolação utilizados no redimensionamento (nearest neighbor, bilinear, cubic) disponíveis na biblioteca OPENCV.

Lillesand e Kiefer (1994, sect. 7.2) comentam que a interpolação **CUBIC** possui vantagens sobre os métodos **NEAREST** e **LINEAR** na aparência final da imagem. Lillesand e Kiefer (1994) não comentam acerca dos métodos **AREA** e **LANCZOS4**. Parker et al. (1983) também comparam 5 métodos de interpolação utilizados em imagens médicas, com três variações do método **CUBIC** (cubic- B-Spline, high-resolution cubic spline with edge enhancement e high-resolution cubic spline). Parker et al. (1983) concluíram que os métodos **CUBIC** possuem uma melhor resposta de frequência do que o **NEAREST** e o **LINEAR**. Apesar das vantagens dos métodos **LINEAR** e **CUBIC** apresentando uma imagem mais nítida, estes métodos acrescentam pixels na imagem e não preservam os valores originais. Por esta razão estes métodos são aplicados depois do reconhecimento de padrões, classificação Richards (1995, cap. 11) ou não são utilizados antes da análise estatística das imagens.

A tabela 2.3, apresenta um sumário das vantagens e desvantagens dos métodos

avaliados. Um estudo mais profundo acerca da definição do melhor método de interpolação excede ao escopo do presente trabalho, mas pode ser encontrado em [Dodgson \(1992\)](#). Baseado na indicação de [Schowengerdt \(2007, pag. 306\)](#) e [Richards \(1995, cap. 2\)](#) foi utilizado o redimensionamento **NEAREST**, para preservar os valores originais da imagem durante a análise estatística.

Tabela 2.3 - Vantagens e desvantagens dos métodos de interpolação candidatos a *resampling* levantados.

Método	Desvantagem	Vantagem	Uso
NEAREST	<i>aliasing</i> <i>offset</i> -	simplicidade não altera original não introduz novos pixels	técnicas estatísticas
LINEAR	sintético altera original problemas pós análise e executado depois do reconhecimento de padrões introduz novos pixels	<i>smoothing</i> reduz <i>aliasing</i> <i>blurring</i>	<i>upsampling</i>
CUBIC	aumenta um pouco produz <i>undershooting</i> problemas da LINEAR <i>halos</i>	melhor <i>aliasing</i> introduz <i>halos</i> precisão e rapidez mais precisa que LINEAR	<i>upsampling</i> <i>downsampling</i>

As imagens do radar foram redimensionadas de (0,75 x 0,75 km) para a resolução espacial do satélite (4 x 4 km). Foi utilizado na análise uma região de formato cúbica oca, onde a varredura do radar é mais intensa, conforme ilustrada na figura 2.3. A projeção desta região está indicada na área em azul na figura. Foram retirados das imagens a região relativa ao cone de silêncio indicado em cinza na figura, centralizada em uma área de aproximadamente 14 x 14 pixels (ou 56 x 56 km). Foram retirados também todos os pixels além de uma área de 38 x 38 pixels (ou 152 x 152 km), por que apresentavam ruídos sistemáticos e aleatórios. Estes procedimentos são recomendados para eliminar fatores incontroláveis do processo de modelamento estatístico [Montgomery \(2005, cap. 1\)](#).

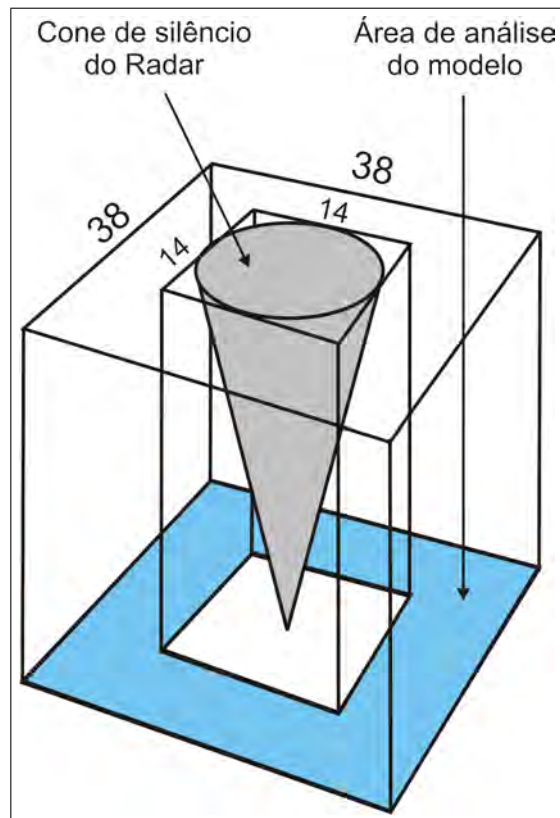


Figura 2.3 - Ilustração acerca da região de análise utilizada no entorno do radar realçada em azul, para a confecção do modelo. 14×14 e 38×38 em pixels correspondem a áreas de 56×56 e 152×152 em *km*. OBSERVAÇÃO. A figura encontra-se fora de proporção especialmente na dimensão vertical para ilustrar melhor o procedimento utilizado.

Depois de concluído o redimensionamento, foram extraídos das imagens os dados em formato texto geograficamente coincidentes necessários para o desenvolvimento do modelo estatístico, conforme descrito na próxima seção.

2.4 Análise Estatística

Nesta seção será descrito como foi construído o modelo de regressão linear estatístico [Montgomery e Peck \(1992, sec. 1.2\)](#), utilizado para a estimativa do conteúdo de água líquida. Os dados obtidos a partir do pré-processamento descrito na seção anterior são múltiplos, por isto serão utilizadas técnicas de regressão linear múltipla conforme sugerido por [Wilks \(2005, cap. 6\)](#). O método experimental utilizado para o desenvolvimento do modelo estatístico será o recomendado por [Montgomery](#)

(2005). Será utilizado também um pacote estatístico de arquitetura aberta o R ⁷ e o MINITAB ⁸ para análise dos dados.

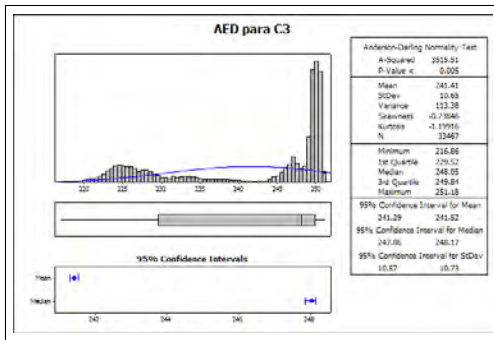
Montgomery (2005) recomenda que os fatores incontroláveis existentes em um experimento estatístico devem ser eliminados. De uma certa forma a interpolação das lacunas, o redimensionamento das imagens e a utilização da região maior sensibilidade de amostragem e confiabilidade dos dados monitorada pelo radar, são parte na mitigação destes fatores. Vale a pena salientar que as imagens e dados do radar fornecem uma amostragem incompleta da região monitorada verticalmente resultando na ocorrência de lacunas. Como o VIL é integrado verticalmente, o modelo calculado neste relatório pode subestimar os valores de VIL calculados, por não integrar completamente todos os dados possíveis de Z existentes na elevação por causa destas lacunas. Esta amostragem incompleta possibilita a ocorrência de erros sistemáticos, aumentando a incerteza do modelo. Este tipo de erro poderia ser reduzido através de uma interpolação vertical. Entretanto no presente trabalho não será utilizada a interpolação vertical para o estabelecimento do presente modelo, por tratar-se de processamento de imagens 3-D, exigindo recursos de computação gráfica, extrapolando muito o escopo do presente trabalho. Este tipo de erro sistemático pode ser mitigado também, através de uma calibração refinada e ou ajustes utilizando séries temporais mais longas nas estimativas dos dados.

2.4.1 Análise Exploratória de dados

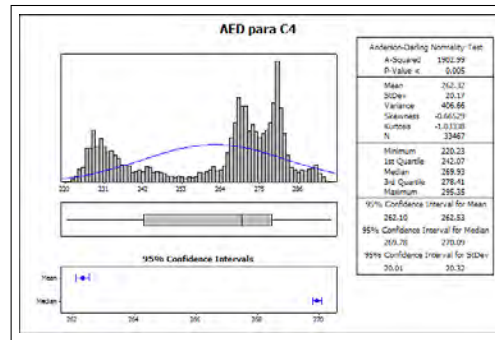
A análise exploratória de dados tem como objetivo conhecer mais detalhadamente o comportamento dos parâmetros utilizados no modelo. A partir de uma série de imagens obtidas da região de interesse ilustrada na figura 2.3, procedeu-se a uma análise exploratória (AED) nas tuplas de dados obtidos a partir das imagens. Foi utilizado um programa estatístico para gerar os gráficos para o canal do vapor de água do satélite C3 ilustrado na figura 2.4(a), para o canal infra-vermelho do satélite C4 ilustrado na figura 2.4(b), para as diferenças entre estes dois canais ou C43 ilustrado na figura 2.4(c), para a variância do canal do vapor d'água ou X3 ilustrado na figura 2.4(d), para a variância do canal infravermelho ou X4 ilustrado na figura 2.4(e) e por último do cálculo do VIL a partir do radar ilustrado na figura 2.4(f). Uma curva de distribuição normal foi sobreposta aos gráficos das figuras para facilitar a análise.

⁷The R Project for Statistical Computing, disponível no link <http://www.r-project.org/>

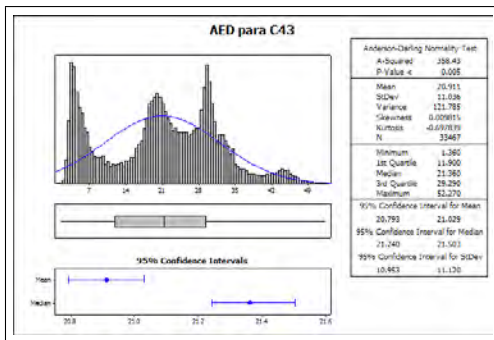
⁸MINITAB <http://http://www.minitab.com/>



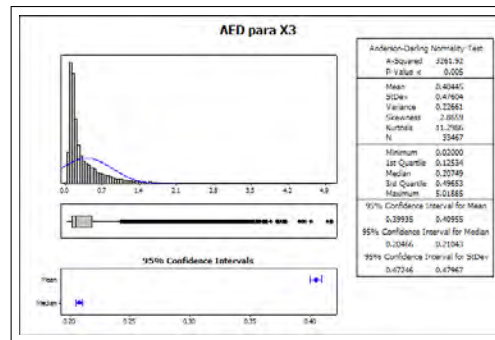
(a) C3 vapor de água



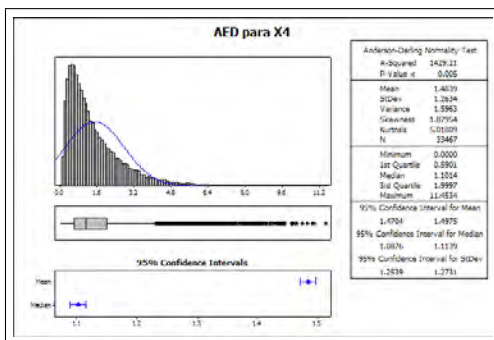
(b) C4 infravermelho



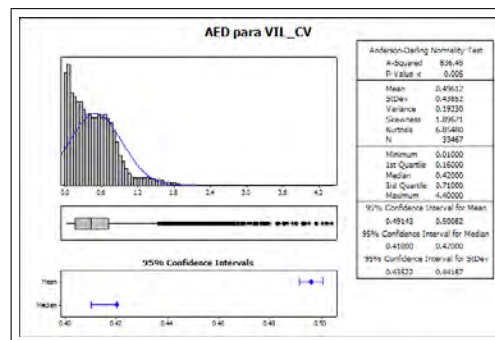
(c) diferença C4 - C3



(d) variabilidade do canal C3



(e) variabilidade do canal C4



(f) cálculo do VIL em kg/m^2

Figura 2.4 - Gráficos dos dados obtidos simultaneamente a partir de 80 amostras de imagens dos canais do GOES, dos seus dados derivados e também do cálculo do VIL limitado à região de maior sensibilidade a partir das imagens do radar.

Conforme pode ser observado na figura 2.4(a), pode-se notar dois padrões de dados típicos no canal de vapor d'água do GOES. O primeiro entre aproximadamente 220 a 240 K e um segundo de amplitude maior acima destes valores. Fisicamente estes dados são correspondentes a dois tipos diferentes de nuvens com temperaturas frias e quentes. Também nota-se a presença em menor quantidade de dados intermediários entre os dois padrões. Apesar de individualmente os dois padrões apresentarem um

perfil de distribuição normal, o perfil dos dois padrões em conjunto não se aproxima de uma distribuição normal.

Conforme pode ser observado na figura 2.4(b) no canal de infra-vermelho, pode-se notar também pelo menos dois padrões de dados típicos no canal C4 de infravermelho do GOES. O primeiro entre aproximadamente 220 a 253 K e um segundo acima destes valores. O segundo padrão maior, aparentemente possui dois padrões de dados menores misturados, mas como não foi encontrado ainda um significado físico ou evento justificável para separá-los, serão considerados um só padrão. Ainda não foi determinado se estes dois padrões maiores possuem o mesmo significado físico indicado na análise do C3. A distribuição dos padrões em conjunto não se aproxima da distribuição normal.

Conforme pode ser observado na figura 2.4(c), pode-se notar pelo menos três padrões típicos de dados sobrepostos, na diferença entre os canais C43 do GOES. Da mesma forma que na análise do C4, também não foi determinado ainda o significado físico destes padrões. A distribuição dos padrões em conjunto não se aproxima da distribuição normal. Fisicamente valores negativos nestes dados, são relativos a nuvens que apresentam grande movimento convectivo vertical, granizo ou a tempestades e descargas elétricas. De acordo com Walck (2007, Cap. 12) dados com este perfil podem ser descritos a partir de uma distribuição do tipo *Doubly Non-Central t-Distribution*.

Conforme pode ser observado nas figuras 2.4(d), 2.4(e) e 2.4(f), pode-se observar que os dados de variabilidade de X3, X4 e VIL não possuem o comportamento da distribuição normal. Aparentemente os dados são todos acima de zero e podem ser descritos a partir de uma distribuição gama específicas a cada caso.

2.4.2 Análise dos componentes principais ou PCA

Deseja-se verificar a contribuição de cada uma das cinco variáveis sugeridas pelos pesquisadores do CEMADEN C3, C4, C43, X3 e X4 como regressores e o de VIL como resposta. Mas antes é necessário investigar qual a possível contribuição de cada um deles na resposta, através da análise de componentes principais ou PCA. Existem dois tipos de análise de componentes principais a partir da matriz de covariância e da matriz de correlação. Graficamente os componentes são ilustrados através de gráficos chamados de *scree plot*, conforme indicado na figura 2.5.

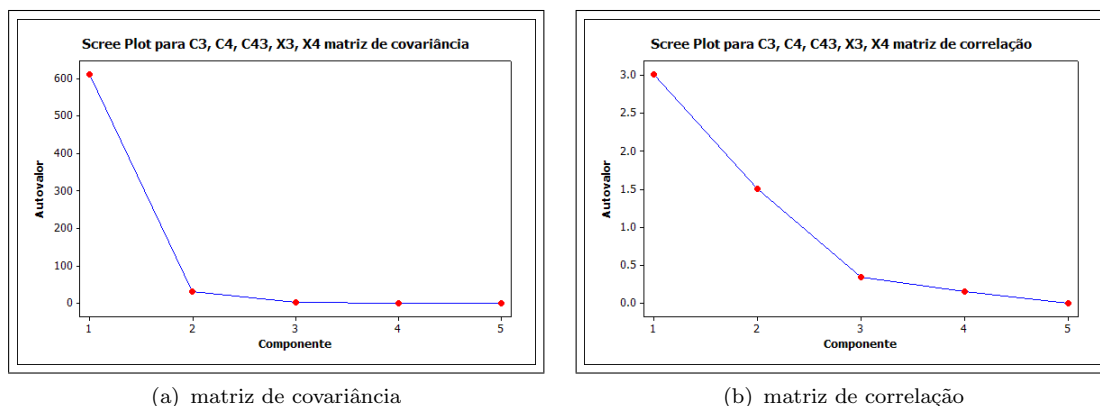


Figura 2.5 - Gráfico do tipo *scree plot*, ilustrando a importância de cada regressor através dos autovalores na análise de componentes principais.

A análise dos componentes principais também indica a importância dos componentes principais, obtidos a partir dos autovalores, a sua proporção em relação à variabilidade total e o acumulado, conforme pode ser observado nas tabelas 2.4 e 2.5. Nestas tabelas estão descritos os autovalores (Aut), a sua proporção individual do total da variância da população amostrada devido ao componente analisado (Prop) e a sua contribuição cumulativa na variabilidade (Acum).

Tabela 2.4 - Análise dos autovalores pela matriz de covariância

An	C3	C4	C43	X3	X4
Aut	610.12	31.88	1.57	0.07	0
Prop	0.948	0.050	0.002	0	0
Acum	0.948	0.997	1	1	1

Tabela 2.5 - Análise dos autovalores pela matriz de correlação

An	C3	C4	C43	X3	X4
Aut	3.0087	1.5035	0.3389	0.1489	0
Prop	0.602	0.301	0.068	0.030	0
Acum	0.602	0.902	0.970	1	1

As tabelas 2.6 e 2.7 indicam as matrizes de covariância e correlação obtidas nos cálculos respectivamente a partir da análise da matriz de covariância e de correlação.

A partir dos resultados obtidos na análise dos componentes principais, pode-se estabelecer quais são as principais variáveis dentre as propostas que contribuem para a confecção do modelo de regressão linear múltipla na próxima seção. De acordo com o indicado na tabela 2.6, C3 e C4 respondem cumulativamente com 99,7% da variabilidade do fenômeno modelado.

Tabela 2.6 - matriz de covariância

Var	PC1	PC2	PC3	PC4	PC5
C3	-0.399	0.710	-0.044	-0.022	0.577
C4	-0.816	0.010	-0.001	-0.007	-0.577
C43	-0.417	-0.700	0.043	0.015	0.577
X3	0.008	-0.038	-0.211	-0.977	0
X4	0.001	-0.055	-0.976	0.213	0

Tabela 2.7 - matriz de correlação

Var	PC1	PC2	PC3	PC4	PC5
C3	-0.545	-0.044	-0.358	-0.629	0.420
C4	-0.560	-0.185	0.101	-0.091	-0.796
C43	-0.497	-0.296	0.530	0.441	0.436
X3	0.364	-0.560	0.493	-0.577	0
X4	0.094	-0.750	-0.581	0.302	0

O regressor C43 foi responsável por 6,8% da variabilidade segundo a matriz de correlação, e 0,2% pela matriz de covariância. C43 não foi incluído na equação de regressão porque cancelava os termos C3 e C4 e também por sua contribuição não ser significativa de acordo com a análise da matriz de covariância.

2.4.3 Modelo de regressão linear múltipla

Apesar dos inconvenientes levantados nas subseções anteriores com relação à normalidade dos dados, será utilizado um modelo de regressão linear múltipla com a finalidade de estabelecer um parâmetro de comparação para a análise de previsão dos dados. O modelo de regressão obtido na análise está descrito na equação 2.5.

$$VIL = 3.13 - 0.005132 \times C3 - 0.00533 \times C4 + \varepsilon \quad (2.5)$$

Segundo Montgomery e Runger (2003, Sec. 10.4), “modelos de regressão não são necessariamente válidos para finalidade de extrapolação” e que “as relações de regressão são válidas somente para a faixa de valores dos dados originais”. Portanto no presente caso para um bom modelo de regressão recomenda-se um conjunto de amostragem mais representativo de toda a faixa de variabilidade dos dados. Vale a pena salientar que no presente relatório, está sendo realizada uma análise preliminar para verificar a viabilidade de utilização do método.

Os coeficientes obtidos no modelo de regressão estão descritos na tabela 2.8. A análise de variância obtida no modelo de regressão encontra-se descrita na tabela 2.9.

É importante observar o resultado obtido do coeficiente de determinação $R^2 = 13.2\%$ ou R-Sq na tabela 2.8. Este coeficiente indica a quantidade da variabilidade dos

Tabela 2.8 - Tabela dos coeficientes encontrados no estabelecimento da regressão linear múltipla.

Predictor	Coef	SE Coef	T	P
Constant	3.13305	0.06969	0.06969	0.000
C3	-0.0051308	0.0005604	-9.16	0.000
C4	-0.0053306	0.0002959	-18.01	0.000
erros	S = 0.408505	R-Sq = 13.2%	R-Sq(adj) = 13.2%	

Tabela 2.9 - Tabela da análise de variância do modelo de regressão linear múltipla.

Fonte de variabilidade	DF	Soma dos quadrados	MQ	F_0	P
Regressão	2	851.11	425.56	2550.13	0.000
Residual	33464	5584.35	0.17		
Total	33466	6435.47			

dados, considerada no modelo de regressão [Montgomery e Runger \(2003, Sec. 10.8.2\)](#). Em geral este coeficiente varia entre $0 \leq R^2 \leq 100\%$.

2.4.4 Verificação da adequação do modelo de regressão

Nesta seção será verificado se o modelo está se comportando de maneira linear de primeira ordem, através do teste de significância, análise dos resíduos, ao coeficiente de regressão e falta de ajuste. Para isto supomos que os erros a componente do erro do modelo tenham distribuição normal com média zero e variância σ^2 ou $N(0, \sigma^2)$.

2.4.5 Teste de significância do modelo de regressão linear múltipla

O teste é realizado para verificar se existe uma relação linear entre a variável de resposta (VIL) e o conjunto de regressores C3 e C4. As hipóteses apropriadas são as seguintes:

$$\begin{aligned}
 H_0 : C3 = C4 = 0 \\
 H_1 : C_j \neq 0
 \end{aligned}
 \tag{2.6}$$

onde a rejeição de H_0 implica que pelo menos um dos regressores contribui significativamente para o modelo. A partir da tabela 2.9 ANOVA do coeficiente de regressão,

obtemos os valores $\nu_1 = 2$, $\nu_2 = 43276 = \infty$ ⁹ graus de liberdade respectivamente para o numerador e denominador e $F_0 = 5341,24$. A partir de uma tabela estatística (MONTGOMERY; RUNGER, 2003, tab. v) , obtemos os seguintes pontos percentuais da distribuição F:

$$f_{\alpha=0,25;\nu_1=2;\nu_2=\infty} = 1,39$$

$$f_{\alpha=0,10;\nu_1=2;\nu_2=\infty} = 2,30$$

$$f_{\alpha=0,05;\nu_1=2;\nu_2=\infty} = 3,00$$

$$f_{\alpha=0,025;\nu_1=2;\nu_2=\infty} = 3,69$$

$$f_{\alpha=0,01;\nu_1=2;\nu_2=\infty} = 4,61$$

Como $F_0 > f_{\alpha;\nu_1;\nu_2=\infty}$ para quaisquer α indicado acima, rejeitamos H_0 e com isto concluímos que no mínimo um dos regressores (C3 ou C4) contribui significativamente para o modelo.

2.4.5.1 Análise residual

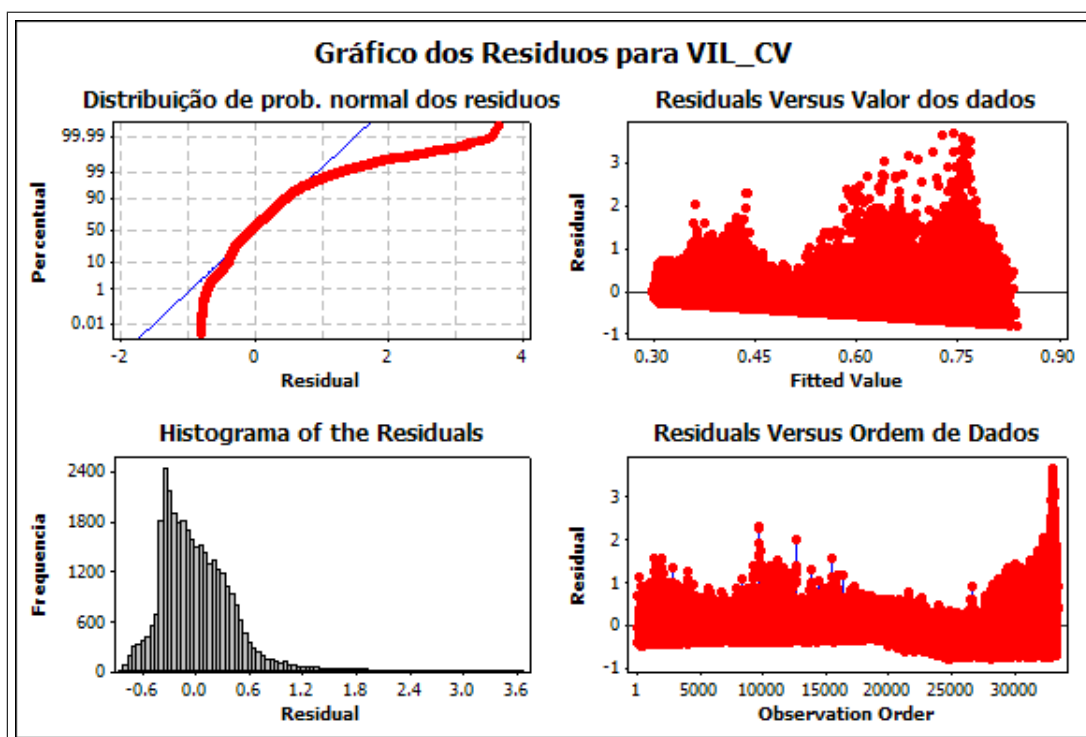
Conforme pode ser observado na análise dos resíduos ilustrado na figura 2.6(a), esta pode ser considerada boa suposição.

A partir dos dados da figura 2.6(a) (probabilidade normal dos resíduos padronizados). 95 % dos valores residuais devem se encontrar no intervalo $[-2, +2]$ (MONTGOMERY; RUNGER, 2003, sec 10.8.1). conforme pode ser observado no gráfico existem alguns *outliers* entre os dados.

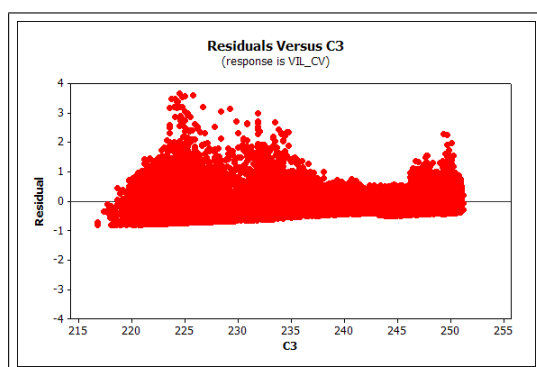
Outra observação importante obtida a partir do gráfico resíduos versus valores dos dados VIL, é que a variância das observações cresce com o valor da magnitude (MONTGOMERY; RUNGER, 2003, sec 10.8.1). A variância pode ser estabilizada com o uso de uma das seguintes transformações \sqrt{VIL} , $\ln VIL$ ou $1/VIL$. O perfil de distribuição dos resíduos indicam a presença de mais de um padrão, de acordo com os gráficos das figuras 2.4(a) e 2.4(b).

A resposta dos resíduos em relação ao regressor C3 (vapor de água) na figura 2.6(b), indica que o modelo subestima os valores de VIL para temperaturas entre 220 e 240

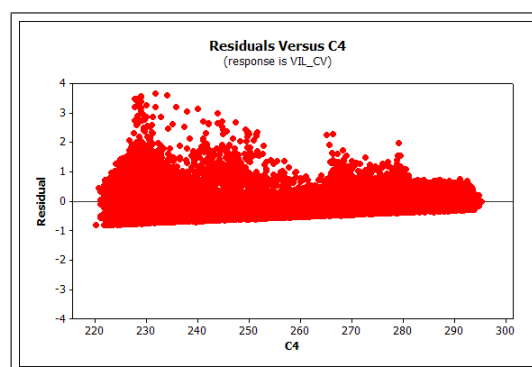
⁹Observação: $\nu_2 = 43276$?. Nas tabelas de distribuição F, valores acima de 120 são considerados ∞ graus de liberdade.



(a) Análise dos resíduos em relação ao VIL



(b) Análise dos resíduos em relação ao regressor C3



(c) Análise dos resíduos em relação ao regressor C4

Figura 2.6 - Figura indicando a análise dos resíduos do modelo de regressão utilizado.

e também acima de 245. Aparentemente esta relação não é linear e pode requerer um termo $C3^2$. Outro fato a ser observado é que a variância dos dados decresce com a magnitude de C3 e também que a variância não é constante. O Valor dos resíduos aumentam na proximidade da ocorrência de nuvens do tipo convectiva e estratiforme e é menor para a ocorrência de valores intermediários entre os dois tipos de nuvens.

A resposta dos resíduos em relação ao regressor C4 (infravermelho) 2.6(c), indica também que o modelo subestima os dados aproximadamente em 230. Como no

caso de C3, a variância também decresce, mas de uma forma mais constante com o aumento da magnitude da temperatura.

2.4.5.2 Análise das superfícies de resposta

A partir da análise das figuras de contorno ilustradas na figura 2.7, as seguintes observações podem ser elicitadas. Na figura 2.7(a), observa-se duas ocorrências distintas de VIL acima de 4 kg.m^{-2} , podendo resultar em duas categorias distintas causado por valores diferentes de C3 e C4. Duas categorias também foram observadas acima de 4 kg.m^{-2} quando analisamos C43 e C3 no gráfico 2.7(c) e quatro categorias distintas acima de 3 kg.m^{-2} . Quando analisamos C43 e C4 no gráfico 2.7(d) duas categorias de dados distintas são observadas acima de 3 kg.m^{-2} . O gráfico de contorno 2.7(b) de X3 e X4 que apresentou três categorias para valores acima de 3 kg.m^{-2} .

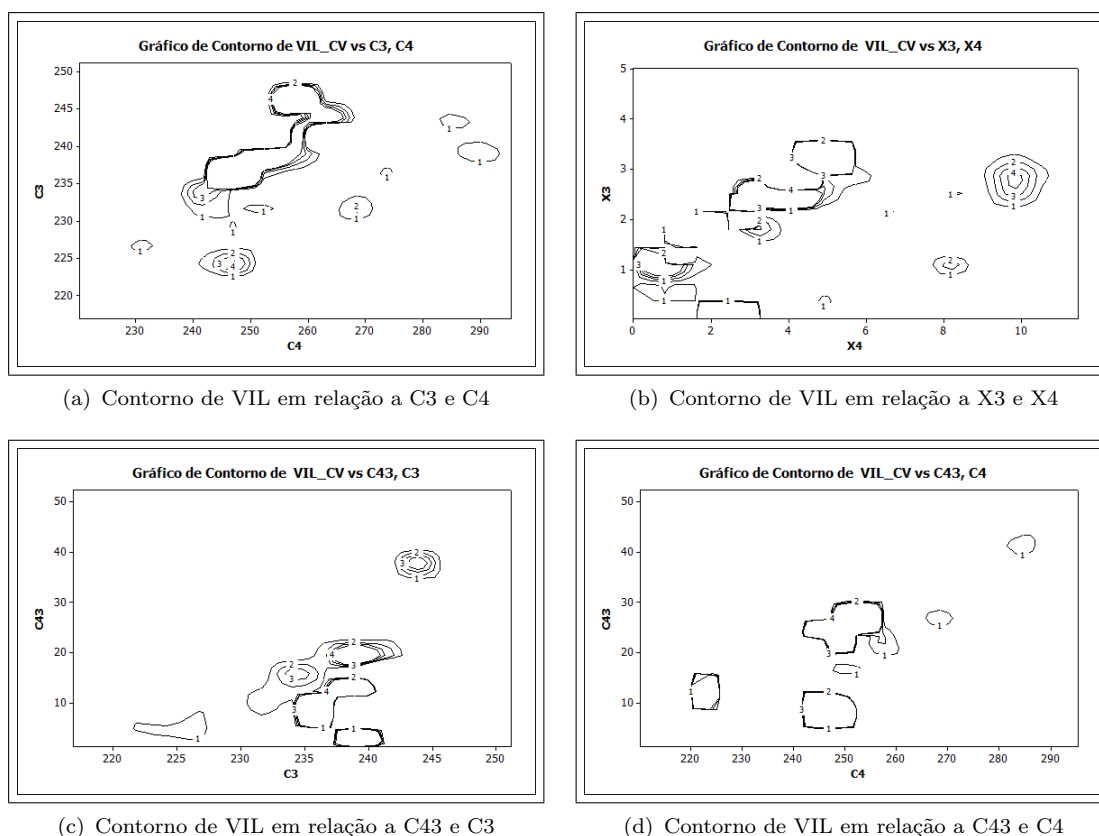


Figura 2.7 - Figura indicando os graficos de contorno do VIL em relação aos canais do GOES e seus derivados. As unidades de VIL estão em kg.m^{-2} , as de C3 e C4 em K e X3 e X4 são adimensionais.

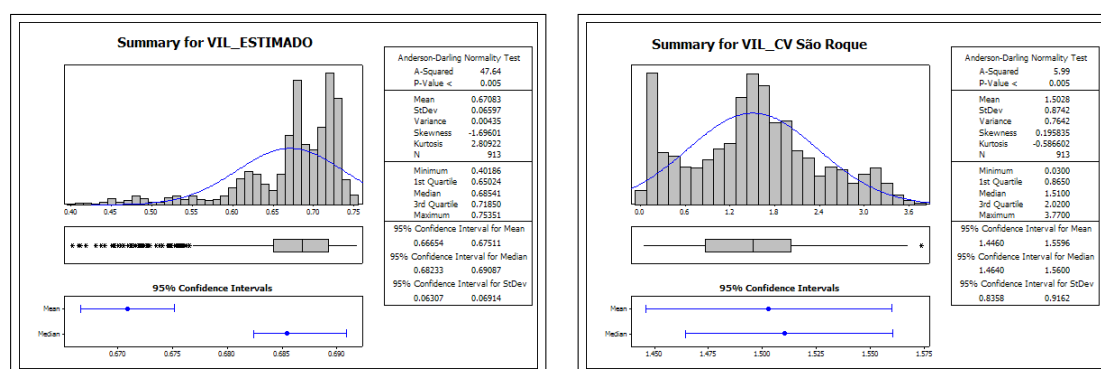
Conforme ilustrado na figura 2.7, pode-se observar que existem vários picos de grande descontinuidade nos gráficos. Estas descontinuidades dificultam o estabelecimento de um modelo de regressão, que em geral é utilizado para modelar fenômenos com variabilidade mais suave. A partir deste ponto é indispensável a participação de um especialista da área para auxiliar no estabelecimento de um domínio apropriado de análise dos fenômenos.

2.5 Validação

A validação do modelo foi realizada, comparando os dados do modelo de VIL estimados a partir do satélite com dados obtidos a partir do radar de São Roque. Foram utilizados dados de 47 imagens do dia 21 de abril de 2012, na validação. Vale a pena salientar o seguinte:

- os dados analisados são da mesma região e estão na mesma resolução espacial;
- somente os bins da região de maior sensibilidade do radar estão sendo considerados;
- 913 bins de radar de São Roque foram utilizados na validação.

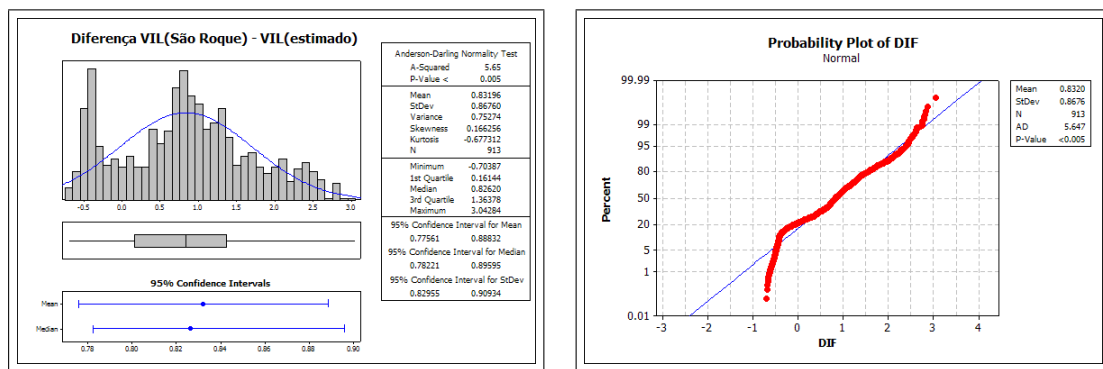
A AED do VIL estimado está ilustrado na figura 2.8(a) e o do VIL obtido a partir do radar de São Roque está ilustrado na figura 2.8(b).



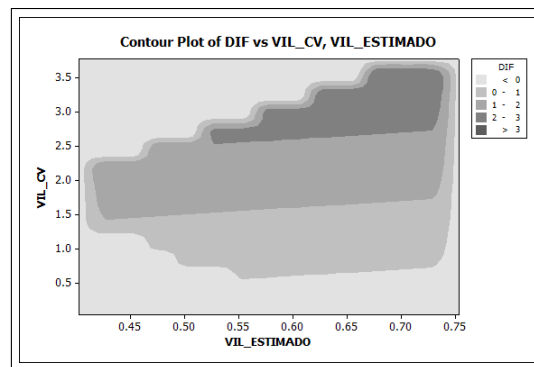
(a) Dados do VIL estimados obtidos a partir do modelo de regressão (b) Dados do VIL estimados obtidos a partir do radar de São Roque

Figura 2.8 - Ilustração acerca da AED realizada nos dados do cálculo do VIL utilizados na validação.

As diferenças entre os dois resultados para o mesmo pixel podem ser observadas na figura 2.9. Conforme pode ser observado nesta figura, as diferenças entre os dois dados tem uma distribuição aproximadamente normal conforme indicado nas figuras 2.9(a) e 2.9(b). A média e a mediana estão muito próximas, mas não possuem um valor próximo de zero indicando um *bias* de aproximadamente $0,83 \text{ kg.m}^{-2}$. A estimativa do modelo de regressão obtido a partir de satélite está subestimando os dados de VIL, em relação ao dado obtido a partir do radar de São Roque. São indicativos de erros sistemáticos que podem significar diferenças entre os entre os radares ou problemas no modelo. Mas este fato precisa ser investigado mais detalhadamente para ser conclusivo.



(a) Diferenças entre os valores estimados e calculados (b) Distribuição de probabilidade entre as diferenças



(c) Gráfico de contorno ilustrando a incidência das maiores diferenças nos valores de VIL entre os dois modelos

Figura 2.9 - Figura comparando estatisticamente os resultados do cálculo do VIL entre o modelo de regressão obtido pelo satélite GOES e pelo radar de São Roque.

Outra observação importante a ser notada é a grande variabilidade nas diferenças entre os valores acima de $2,50 \text{ kg.m}^{-2}$ de acordo com o ilustrado na figura 2.9(c).

Isto pode ser explicado pelo fato do modelo de regressão ter sido estabelecido em uma faixa de valores inferior aos dados observados pelo radar de São Roque. Isto é de acordo com a figura 2.4(f) os valores dos dados do coeficiente de regressão estavam em sua maioria abaixo de $1,6 \text{ kg.m}^{-2}$ com um máximo de $2,85 \text{ kg.m}^{-2}$. Os dados do radar de São Roque estavam monitorando um fenômeno mais intenso com valores bem acima dos utilizados para estabelecer o modelo de regressão com um máximo na ordem de $3,77 \text{ kg.m}^{-2}$. Conforme recomendado por Montgomery e Peck (1992, cap. 1) os modelos de regressão são validos somente dentro da faixa de variabilidade dos dados utilizados para a confecção do modelo. A figura 2.9(c) demonstra claramente este problema, indicando que as maiores diferenças existem para valores de VIL estimados para o Radar de São Roque aproximadamente acima de $2,50 \text{ kg.m}^{-2}$.

2.6 Comentários acerca da análise estatística

Vale a pena salientar que a análise realizada aqui é um estudo de caso. As 80 imagens de radar e satélite utilizadas são relativas a um só dia, para verificar a viabilidade do modelo proposto. Para um estudo mais refinado é necessário estabelecer um experimento estatístico que considere eventos distintos além da variabilidade diária e sazonal. A quantidade de dados analisada utilizada também foi muito grande. Para uma análise mais refinada recomenda-se o planejamento de um experimento e um conjunto de dados mais racionalizado, para reduzir a grande quantidade de fatores não controlados a ser considerados no modelo .

Inicialmente acreditava-se que todos os fenômenos pudessem ser modelados simultaneamente através de distribuições do tipo normal. Mas, conforme pode ser observado na análise dos gráficos na seção 2.4.1, o modelo de distribuição normal não é muito bom para representar simultaneamente os dois tipos de precipitação (estratiforme e convectiva) que aparecem em destaque nos dados. Estes fenômenos devem ser modelados separadamente. Sob o ponto de vista estatístico, utilizar um só modelo de regressão linear para abranger os dois padrões em um só, teria os seguintes inconvenientes.

- O modelamento assumiria que o comportamento dos dados teria distribuição normal, de acordo com o indicado na linha azul dos gráficos da figura 2.4.

- Os sistemas do tipo estratiforme e convectivo seriam subestimadas.
- As nuvens intermediárias seriam superestimadas.
- Os erros envolvidos no modelamento seriam maiores e a análise de consistência dos resíduos indicaria uma variância não constante.

Para tentar reduzir estes problemas um modelo em separado para cada tipo de sistema deve ser estabelecido. Recomenda-se também um conjunto de amostragem mais representativo de toda a faixa de variabilidade dos dados.

3 CONCLUSÕES

Neste relatório estão descritas as atividades relativas ao desenvolvimento de um modelo para a estimativa do conteúdo de água líquida (VIL), a partir de um estudo de caso, utilizando imagens de satélite, validado por imagens de radar meteorológico (Z). Foram utilizados os canais C3 e C4 do satélite GOES-12, além da diferença entre os canais C43 e suas variâncias X3 e X4. Para a construção do modelo foram utilizadas 80 imagens do dia 25 de Abril de 2012 obtidas a cada 15 minutos do radar do Morro da Igreja em Santa Catarina validado com as imagens do radar de São Roque no Estado de São Paulo. Durante o modelamento, foi utilizada uma análise exploratória de dados (EDA) para determinar as distribuições de dados presentes e estabelecer um modelo de regressão linear múltipla. A análise de componentes principais (PCA) foi utilizada para eliminar os parâmetros redundantes da análise estatística. Regiões como o cone de silêncio e o ruído foram removidas das imagens para melhorar a qualidade do modelo. Somente a região com monitoramento mais detalhado do radar foi utilizada no modelo. Devido a diferente resolução espacial existente entre as imagens de radar e satélite, foram utilizados algoritmos de redimensionamento para reduzir as imagens a mesma resolução. A interpolação Nearest Neighbour(NN) foi utilizada no algoritmo de redimensionamento para não inserir pixels artificiais nas imagens.

O modelo deve apresentar uma ligeira sub estimativa do VIL, por causa da amostragem incompleta devido a lacunas existentes nas imagens de radar utilizadas. Este problema poderá ser resolvido utilizando interpolação 3-D.

A AED indicou a presença de dados multimodais nos diferentes canais, o modelo de regressão foi mantido para verificar o comportamento dos dados. Pelo menos uma separação dos padrões convectivos e estratiformes deveriam ser utilizados para um modelo mais preciso.

PCA indicou que os canais C3 e C4 eram 99 % significantes do modelo de variância. O coeficiente de regressão indicou que o modelo representava somente 13.2 % da variabilidade dos dados. A grande quantidade de *outliers* e dados intermediários, influenciaram tendenciosamente os resultados devido a natureza média quadrática do modelo de regressão linear. Acredita-se que estes *outliers* nos dados representem uma grande quantidade de ruídos nas imagens que precisariam ser removidos para uma análise mais refinada do modelo

O gráfico de contorno dos canais C3 e C4 pelo VIL apresentou vários picos ou padrões confirmando a natureza multimodal de padrões, indicando que existem múltiplas causas de precipitações intensas. Durante a validação dos dados as diferenças entre o modelo e os dados do radar de São Roque apresentaram uma distribuição normal, mas com uma tendência de estimativa maior pelo radar de São Roque. Estas diferenças ocorreram por que o intervalo de dados do modelo de regressão utilizado no radar do Morro da Igreja era menor que os dados utilizados no radar de São Roque. De acordo com a literatura especializada a maior escala de dados possível deve ser utilizada para construir o modelo de regressão. As causas físicas das distribuições multi modais devem ser investigadas em um experimento controlado mais rigorosamente.

Durante as atividades, foram desenvolvidas rotinas para leitura e processamento de imagens utilizando bibliotecas de código em aberto OPENCV para extrair parâmetros das imagens 3-D de radar e de satélite, para redimensioná-las e compará-las simultaneamente. Estas rotinas facilitarão os trabalhos futuros a serem desenvolvidos nesta área.

Finalmente recomenda-se a realização de um novo experimento mais controlado sob a supervisão de um especialista em meteorologia que utilizando o seu conhecimento tácito poderá direcionar o experimento de maneira mais controlada utilizando um novo conjunto de dados.

REFERÊNCIAS BIBLIOGRÁFICAS

CHARTER, N.; TENENBAUN, J. B.; A., Y. Probabilistic models of cognition: Conceptual foundations. **Trends in cognitive sciences**, v. 10, p. 287–291, 2006. 1

_____. Probabilistic models of cognition: where next? **Trends in cognitive sciences**, v. 10, p. 292–293, 2006. 1

DESERNO, M. **Linear and Logarithmic Interpolation**. [s.n.], March 24, 2004. Disponível em: <www.cmu.edu/biolphys/deserno/pdf/log_interp.pdf>. 7

DODGSON, N. A. **Image resampling**. Dissertação (Mestrado) — University of Cambridge, 1992. 11, 12

GREENE, D. R.; CLARK, R. A. Vertically integrated liquid water—a new analysis tool. *mon. wea. rev.*, 100, 548–552. **Monthly Weather Review**, v. 100, p. 548–552, 1972. 1, 8

JOHANNET, A.; VAYSSADE, B.; BERTIN, D. Neural networks: From black box towards transparent box application to evapotranspiration modeling. **World Academy of Science, Engineering and Technology**, v. 30, p. 162–169, 2007. Disponível em: <<http://www.waset.org/journals/waset/v30.php>>. 1

LILLESAND, T. M.; KIEFER, R. W. **Remote Sensing and Image Interpretation**. [S.l.]: John Wiley & Sons, 1994. 750 p. 11

MONTGOMERY, D. C. **Design and Analysis of Experiments**. [S.l.]: John Wiley and Sons Inc., 2005. 643 p. 12, 14

MONTGOMERY, D. C.; PECK, E. A. **Introduction to Linear Regression Analysis**. 2nd. ed. [S.l.]: John Wiley & Sons, 1992. 13, 25

MONTGOMERY, D. C.; RUNGER, R. C. **Applied Statistics and Probability for Engineers**. 3 rd. ed. [S.l.: s.n.], 2003. 976 p. 18, 19, 20

NIST. **Engineering Statistics Handbook**. NIST - National Institute of Standards, 2009. Disponível em: <<http://www.itl.nist.gov/div898/handbook/eda/eda.htm>>. 1, 4

PARKER, J. A.; KENYON, R. V.; TROXEL, D. E. Comparison of interpolating methods for image restoration. **IEEE TRANSACTIONS ON MEDICAL IMAGING**, MI-2, p. 31–39, 1983. 11

RICHARDS, J. **Remote Sensing Digital Image Analysis: An Introduction**. 2nd. ed. Springer, 1995. ISBN 9783540251286. Disponível em: <<http://books.google.com.br/books?id=4PB5vhPBdJ4C>>. 11, 12

SCHOWENGERDT, R. A. **Remote Sensing Models and Methods for Image Processing**. 3rd. ed. [S.l.]: Academic Press, 2007. 515 p. 11, 12

WALCK, C. **Handbook on Statistical Distributions for Experimentalists**. [s.n.], 2007. Disponível em: <[www.physto.se/~walck\(0\)](http://www.physto.se/~walck(0))>. 16

WILKS, D. S. **Statistical methods in the atmospheric sciences**. [S.l.]: Academic Press, 2005. 627 p. (International Geophysics Series). 13

WMO, W. M. O. **Training Course on Weather Radar Systems**. ALANYA FACILITIES, ANTALYA, TURKEY: [s.n.], September 2005. 8

YUILLE, A.; KERSTEN, D. Vision as bayesian inference: analysis by synthesis? **Trends in Cognitive Sciences**, v. 10, n. 7, p. 301 – 308, 2006. ISSN 1364-6613. Special issue: Probabilistic models of cognition. Disponível em: <<http://www.sciencedirect.com/science/article/B6VH9-4K717XN-1/2/f574d61f93f8906ccc3f9457d3dba11d>>. 1

ZHANG, P. G. Avoiding pitfalls in neural network research. **IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS PART C APPLICATIONS AND REVIEWS.**, v. 37, p. 3–16, January 2007. 1