

Dissertação de Mestrado

Visualização de Matrizes de Covariância Complexas: uma aplicação em dados PolSAR

Antônio Marcos Larangeiras Lima
amlarangeiras@gmail.com

Orientador e Co-orientador:

Dr. Alejandro César Frery
Dr. Raydonal Ospina Martínez

Maceió, Março de 2017

Antônio Marcos Larangeiras Lima

Visualização de Matrizes de Covariância Complexas: uma aplicação em dados PolSAR

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Modelagem Computacional de Conhecimento do Instituto de Computação da Universidade Federal de Alagoas.

Orientador: Dr. Alejandro César Frery
Co-orientador: Dr. Raydonal Ospina Martínez

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Bibliotecária: Janaina Xisto de Barros Lima

L732v Lima, Antônio Marcos Larangeiras.

Visualização de Matrizes de Covariância Complexas: uma aplicação em dados PolSAR/ Antônio Marcos Larangeiras Lima. – 2017
68 f. : il.

Orientador: Alejandro César Frery.

Coorientador: Raydonal Ospina Martínez.

Dissertação (Mestrado em Modelagem Computacional de Conhecimento) – Universidade Federal de Alagoas. Instituto de Computação. Pós-Graduação em Modelagem Computacional de Conhecimento. Maceió, 2017.

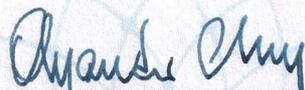
Bibliografia: f. 66–68.

1. Imagens PolSAR. 2. Estatística Multivariada – 3. Visualização.
I. Título.

CDU: 004.932:004.852

Membros da Comissão Julgadora da Dissertação de Mestrado de Antônio Marcos Larangeiras Lima, intitulada: “Visualização de Matrizes de Covariância Complexas: uma aplicação em dados PolSAR”, apresentada ao Programa de Pós-Graduação em Modelagem Computacional de Conhecimento da Universidade Federal de Alagoas, em 24 de março de 2017, às 9h00min, na sala 15 do Instituto de Computação da Ufal.

COMISSÃO JULGADORA



Prof. Dr. Alejandro Cesar Frery

Ufal – Instituto de Computação

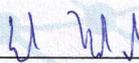
Orientador



Profa. Dra. Eliana Silva de Almeida

Ufal – Instituto de Computação

Examinadora



Prof. Dr. Leonardo Melo de Medeiros

Ifal – Instituto Federal de Alagoas

Examinador

RESUMO

O monitoramento do nosso planeta através de sensores imageadores pode ser empregado para inspecionar: desmatamentos e desertificação de florestas; o ciclo da água; crescimento urbano entre outras. Entre as tecnologias de sensoriamento remoto, PolSAR (*Polarimetric Synthetic Aperture Radar*) vem se destacando. O PolSAR possibilita o sensoriamento remoto em quase todas as condições meteorológicas. Cada célula de resolução na imagem PolSAR está associada a uma matriz de espalhamento complexa. Uma maneira, de caracterizar os dados PolSAR single-look é utilizar a matriz de covariância complexa. A matriz de covariância é de extrema importância em Análise Multivariada pois permite mensurar e avaliar o grau de dependência entre as variáveis que compõem o conjunto de dados.

Este trabalho apresenta uma abordagem de visualização de dados multivariados para visualizar informações contidas em matrizes de covariância complexa. As informações sobre que tipo de região homogênea está contida na imagem PolSAR a ser estudada, pode ser obtida analisando a abordagem de visualização de dados fornecida neste trabalho. Para a validação da abordagem, aplicamos a nossa ferramenta em imagens PolSAR - não simuladas - sobre amostras homogêneas, ou seja, amostras que seguem uma determinada família de distribuição de probabilidade. Em seguida, a proposta foi aplicada em três alvos distintos, à saber: região desértica, região urbana e região aquática. Portanto, a abordagem de visualização implementada no R permite discriminar os alvos.

Palavras-chave: Imagens PolSAR; Estatística Multivariada; Visualização; Dependência Efetiva; Variância Efetiva.

ABSTRACT

The monitoring of our planet through image sensors can be used to inspect: deforestation and desertification of forests; the water cycle; urban growth, among others. Among the technologies of remote sensing, PolSAR (*Polarimetric Synthetic Aperture Radar*) has been highlighting. The PolSAR enables remote sensing in almost all weather conditions. Each resolution cell in the PolSAR image is associated with a complex scattering matrix. One way to characterize the single-look PolSAR data is to use the complex covariance matrix. The covariance matrix is of extreme importance in Multivariate Analysis because it allows measure and evaluate the degree of dependence between the variables that compose the data set.

This work presents a multivariate data visualization approach to visualize information contained in complex covariance matrices. The information about what type of homogeneous region is contained in the PolSAR image to be studied can be obtained by analyzing the data visualization approach provided in this work. For the validation of the approach, we applied our tool in non-simulated PolSAR images based homogeneous samples, i. e., samples that follow a certain of probability distribution. After that, the proposal was applied in three distinct targets, namely: a desertic region, a urban region and an aquatic region. Therefore, the implemented visualization approach in the \mathbb{R} permits to discriminate the targets.

Keywords: PolSAR Images; Multivariate Statistics; Visualization; Effective Dependence; Effective Variance.

AGRADECIMENTOS

Agradeço primeiramente a Deus pela oportunidade de fazer esta Pós-graduação e por ter me dado força para terminar este curso; a Prof^a Anamelea de Campos Pinto pelo incentivo para fazer um mestrado durante o período que passei na CIED; a Cássia Monalisa pelo estímulo e por acreditar que seria capaz de fazer esse curso; ao Eduardo Santana por dividir a casa e as despesas comigo durante o mestrado, além do encorajamento diário para cursar um doutorado, futuramente.

Também gostaria de agradecer, imensamente, ao meu orientador Prof. Alejandro Frery e co-orientador Prof. Raydonal Ospina por mostrar o verdadeiro pensamento de um pesquisador e pela paciência durante o mestrado; a CAPES pela bolsa de estudos, pois sem ela seria complicado fazer o curso. Aos professores que contribuíram para minha formação, em especial: Heitor Soares, André Aquino e Rodrigo Paes. Aos momentos descontraídos com os colegas de curso em especial: Pedro Dantas, Vilker Tenório, Carlos Gama, Fábio, Hugo, Vitor Torres e Ana Maria.

Além disso, gostaria de agradecer aos colegas de laboratório pelas dúvidas de programação em C e alegrias proporcionadas no dia-a-dia, principalmente: Ricardo Lopes, Leonildo Melo, Rivo Sarmiento e Israel.

Por fim, gostaria de agradecer ao espaço fornecido pelo CPMAT para o desenvolvimento das minhas atividades.

Antônio Marcos Lorangeiras Lima

LISTA DE FIGURAS

2.1	Espectro Eletromagnético e bandas da região de micro-ondas.	15
2.2	Diagrama de Sensores.	16
2.3	Ilustração da transmissão do pulso eletromagnético e retorno dos sinais após atingir o objeto.	17
2.4	Esboço da geometria do imageamento SAR em uma visão tridimensional.	19
2.5	Esboço da geometria do imageamento SAR em uma visão frontal.	19
2.6	Esboço da geometria da Resolução.	20
2.7	Esboço da geometria da Polarização.	21
2.8	Fotos de um sistema aerotransportado.	22
2.9	Sistema espacial SIR-C/X-SAR.	23
2.10	Sistemas aerotransportados.	24
2.11	Tetraedro elíptico.	26
2.12	Matriz de correlações em forma de elipses.	28
2.13	Gráfico <i>corrgram</i> para matrizes de correlação.	29
3.1	Método de visualização em 4 camadas: <i>VisCov</i>	31
3.2	Ilustração de uma imagem PolSAR no formato enxuto.	33
3.3	Diagrama de Atividades para obtenção dos elementos visuais da ferramenta.	34
3.4	Resultado do modelo proposto aplicado na Imagem Niigata.	36
3.5	Resultado da aplicação da função <i>VisMatCovComplex.selecionada</i> nas imagens de Niigata, Vale da Morte, e Baía de São Francisco.	37
4.1	Visualização mais específica através da função <i>VisMatCovComplex.selecionada</i>	42
4.2	AIRSAR(NASA/JPL), Vale da Morte(CA) US, PolSAR.	43
4.3	Visualização de Matrizes de Covariância Complexa da Imagem de Vale da Morte região Amarela.	44
4.4	Visualização de Matrizes de Covariância Complexa da Imagem de Vale da Morte região Azul.	45
4.5	Visualização de Matrizes de Covariância Complexa da Imagem de Vale da Morte região Vermelha.	46
4.6	Visualização de Matrizes de Covariância Complexa da Imagem de Vale da Morte região Verde.	47
4.7	PISAR(NICT/JAXA), Niigata JP, PolSAR.	48
4.8	Visualização de Matrizes de Covariância Complexa da Imagem de Niigata região Amarela.	49
4.9	Visualização de Matrizes de Covariância Complexa da Imagem de Niigata região Azul.	50
4.10	Visualização de Matrizes de Covariância Complexa da Imagem de Niigata região Vermelha.	51
4.11	Visualização de Matrizes de Covariância Complexa da Imagem de Niigata região Verde.	52
4.12	AIRSAR(NASA/JPL), Baía de São Francisco(CA) US, PolSAR.	53
4.13	Visualização de Matrizes de Covariância Complexa da Imagem de São Francisco região Amarela.	54
4.14	Visualização de Matrizes de Covariância Complexa da Imagem de São Francisco região Azul.	55

4.15	Visualização de Matrizes de Covariância Complexa da Imagem de São Francisco região Vermelha.	56
4.16	Visualização de Matrizes de Covariância Complexa da Imagem de São Francisco região Verde.	57

LISTA DE TABELAS

2.1	Correlação entre 11 variáveis de performance automobilística.	27
4.1	Resumos estatístico das amostras retiradas da imagem de Niigata.	58
4.2	Resumos estatístico das amostras retiradas da imagem de Vale da Morte.	59
4.3	Resumos estatístico das amostras retiradas da imagem de São Francisco.	60

SUMÁRIO

1	INTRODUÇÃO	9
1.1	Definição do Problema	9
1.2	Contribuições	11
1.3	Objetivos	11
1.4	Organização da Dissertação	11
2	FUNDAMENTAÇÃO TEÓRICA	12
2.1	Processamento de Imagens	12
2.2	Radar	14
2.2.1	Radar de Abertura Sintética (SAR)	17
2.2.2	Radar de Abertura Sintética Polarizado (PolSAR)	21
2.3	Alguns Conceitos de Estatística Multivariada	23
2.4	Visualização de Dados Multivariados	26
3	VISUALIZAÇÃO DE MATRIZES DE COVARIÂNCIA COMPLEXAS EM DADOS POLSAR	30
3.1	Metodologia	30
3.1.1	A ferramenta desenvolvida para visualizar matrizes de covariância complexa	35
4	RESULTADOS E ANÁLISES	39
4.1	Dados PolSAR	39
5	CONCLUSÕES	62
5.1	Considerações Finais	62
A	AMBIENTE REPRODUTÍVEL E COMPUTACIONAL	64
A.1	Ambiente computacional utilizado	64
	REFERÊNCIAS	66

1

INTRODUÇÃO

1.1 Definição do Problema

NAS últimas quatro décadas, foi iniciado o monitoramento do nosso planeta através de sensores imageadores com o objetivo de obter informações de natureza diferente, em escala global. Este tipo de imageamento pode ser utilizado para diversas aplicações, como por exemplo: em agricultura, auxiliando na previsão de safras, identificação de culturas, monitoramento de umidade do solo; em florestas, monitorando desmatamento e desertificação; em neve e gelo, monitorando o ciclo d'água; em áreas urbanas, monitorando o crescimento das cidades e desastres naturais dentre outras (Meneses & Almeida, 2012; Lee & Pottier, 2009; Liu et al., 2013; Jagdhuber et al., 2013; Chen & Sato, 2013).

O Sensoriamento Remoto é uma ciência que visa a obtenção de imagens da superfície terrestre por meio da detecção e medição quantitativa das respostas das interações da radiação eletromagnética com os materiais terrestres. Entre as tecnologias de sensoriamento remoto, PolSAR (*Polarimetric Synthetic Aperture Radar*) vem se destacando. O PolSAR possibilita o sensoriamento remoto em quase todas as condições meteorológicas, o imageamento em escala global e a estimação de importantes parâmetros biológicos e geofísicos sobre a superfície da Terra (Meneses & Almeida, 2012; Moreira et al., 2013).

Os sensores imageadores orbitais e aerotransportados são divididos em três classes: sensores ópticos que utilizam fonte solar de radiação eletromagnética; sensores termiais que usam a Terra como a fonte de radiação; e sensores de radar que usam fontes artificiais de radiação construídas pelo homem. Neste trabalho iremos nos concentrar em sensores de radar. A sua principal vantagem em relação aos ópticos e termiais é que o comprimento de onda das micro-ondas não é barrado ou absorvido pelas micrométricas partículas de gases da atmosfera e o imageamento pode ser feito a qualquer hora do dia. A tecnologia empregada nos Radares de Abertura Sintética (SAR) simulam uma antena de centenas de metros em uma

antena pequena usando modificações nos registros dos dados e técnicas de processamento de sinais (Meneses & Almeida, 2012).

O modelo estatístico mais utilizado para descrever o sinal retroespalhado PolSAR é a distribuição Wishart complexa escalonada multivariada, denotada por $\mathbf{Z} \sim \mathcal{W}(C, L)$. Essa distribuição possui dois parâmetros: C , a matriz de covariância complexa, e $L \geq 1$, o número de looks. O domínio de C é o conjunto de matrizes hermitianas positivas definidas. A matriz de covariância C é o parâmetro que caracteriza os dados retornados sob análise e, portanto, diferencia o comportamento de diferentes alvos sobre a mesma imagem. A densidade dessa distribuição é dada por

$$f_{\mathbf{Z}}(\mathbf{Z}', C, L) = \frac{L^m |\mathbf{Z}'|^{L-m} \exp(-L \text{tr}(C^{-1} \mathbf{Z}'))}{|C|^L \Gamma_m(L)},$$

onde m é a dimensão de C , $m \leq L$, $\Gamma_m(L) = \pi^{m(m-1)/2} \prod_{k=0}^{m-1} \Gamma(L-k)$, $\Gamma(\cdot)$ é a função gamma, $E(\mathbf{Z}) = C$, $\text{tr}(\cdot)$ é o traço de uma matriz e $|\cdot|$ é o determinante de uma matriz. (Frery et al., 2013; Frery et al., 2014; Nascimento et al., 2014).

Nas imagens PolSAR o valor em cada coordenada é representado por uma matriz de valores complexos, denominada “matriz de covariância” definida da seguinte forma:

$$C = \begin{bmatrix} E(S_1 S_1^*) & E(S_1 S_2^*) & \cdots & E(S_1 S_m^*) \\ E(S_2 S_1^*) & E(S_2 S_2^*) & \cdots & E(S_2 S_m^*) \\ \vdots & \vdots & \ddots & \vdots \\ E(S_m S_1^*) & E(S_m S_2^*) & \cdots & E(S_m S_m^*) \end{bmatrix}.$$

onde S_{ij} representa o sinal de retroespalhamento para a i -ésima transmissão e j -ésima recepção de uma polarização linear, os elementos da diagonal principal são números reais positivos, e denominam-se “intensidades”, o “*” denota o conjugado de um número complexo, por fim $E(\cdot)$ denota o operador esperança estatística (Lee & Pottier, 2009; Nascimento et al., 2014).

Análise Estatística Multivariada é a análise estatística simultânea sobre um conjunto de variáveis. A matriz de covariância é de extrema importância em Análise Multivariada pois permite mensurar e avaliar o grau de dependência entre as variáveis que compõem o conjunto de dados. Na análise Multivariada de dados é extremamente relevante a utilização de gráficos e diagramas, pois facilita a organização e extração de informação dos dados (Anderson et al., 1958; Everitt & Hothorn, 2011).

Vários trabalhos de visualização sobre dados multivariados que assumem valores reais foram propostos na literatura. Por exemplo: Murdoch & Chow (1996) propuseram a exibição de correlações através de elipses que apresenta um gráfico intuitivo para matriz de correlação. Um outro método bastante interessante foi a utilização de uma mapa de calor (Friendly, 2002).

Devemos ressaltar que os trabalhos anteriores foram desenvolvidos para a visualização de uma única matriz de correlação. Tokuda et al. (2011) propuseram uma série de gráficos para visualizar matrizes de covariância.

Conforme o exposto anteriormente, neste trabalho propomos a comparação de matrizes de covariância complexa através de técnicas de visualização em imagens PolSAR.

1.2 Contribuições

Neste trabalho desenvolvemos uma abordagem de visualização de dados multivariados implementada em \mathbb{R} para visualizar matrizes de covariância complexa, em particular, dados PolSAR. A abordagem implementada foi aplicada em imagens PolSAR de característica bastante diferentes.

1.3 Objetivos

Este trabalho apresenta uma nova abordagem para visualizar propriedades de matrizes de covariância complexa, em particular, na avaliação da estrutura de covariância em imagens PolSAR.

Objetivos Específicos

- Criar uma função para visualização de informações em matrizes complexas;
- Aplicar essa função em dados PolSAR;
- Avaliar e comparar o conteúdo das informações obtidas em dados PolSAR reais, entendendo reais como dados não simulados.

1.4 Organização da Dissertação

Esse trabalho organiza-se da seguinte maneira: No capítulo 2 resumiremos os principais conceitos e técnicas que dão base para a construção da ferramenta de visualização. No capítulo 3 detalharemos o processo de criação da ferramenta para visualização de matrizes de covariância complexa. Discussão sobre os resultados obtidos são apresentados no capítulo 4. E, finalizamos com a conclusão no capítulo 5.

O foco desse trabalho foi propor uma nova abordagem de visualização para matrizes de covariância complexa, em dados PolSAR.

2

FUNDAMENTAÇÃO TEÓRICA

NESTE capítulo, apresentaremos os conceitos necessários para a construção da nossa ferramenta de visualização. Aqui discutiremos brevemente sobre: radares e sensores, imagens e análise estatística de dados multivariados.

2.1 Processamento de Imagens

Na literatura, não existe um consenso geral sobre o que é de fato processamento de imagens. Porém é útil levar em conta o seguinte paradigma: *processamento de nível baixo*, que contém operações primitivas, por exemplo, filtragem para reduzir o ruído e o melhorar o contraste. *Processamento de nível médio* que inclui a segmentação, isto é, separação de uma imagem em regiões ou objetos, e também a descrição desses objetos para reduzi-los a uma forma adequada para armazenamento no computador e classificação de objetos individuais. Finalmente, *processamento de alto nível*, que permite “dar significado cognitivo” a um conjunto de objetos reconhecidos (Gonzalez & Woods, 2010).

Iremos utilizar Velho et al. (2009), Gonzalez & Woods (2010) e Frery & Perciano (2013) para definir conceitos relevantes para nosso trabalho.

Definição 2.1.1 (Imagem) Considere $S = \{1, \dots, m\} \times \{1, \dots, n\} \subset \mathbb{N}^2$. Dizemos que uma imagem é uma função da forma

$$\begin{aligned} f : S &\longrightarrow \mathbb{k} \\ s &\longmapsto f(s), \end{aligned}$$

com $f \in S^{\mathbb{k}}$.

O par $(s, f(s))$ é denominado *pixel*. O conjunto S é chamado de suporte da imagem, que é uma região conexa. Os elementos de S são chamados de *sites*, *coordenadas* ou *posições*.

O conjunto \mathbb{k} é o conjunto dos valores possíveis em cada posição. Esse conjunto pode ser caracterizado de diversas formas, por exemplo, $\{0,1\}$, \mathbb{C} , \mathbb{C}^p etc. Quando \mathbb{k} é definido da forma \mathbb{k}^p o valor p representa a quantidade de bandas existentes na imagem, entendendo banda como $\underbrace{\mathbb{k} \times \cdots \times \mathbb{k}}_{p\text{-vezes}}$. O armazenamento de \mathbb{k} no computador chama-se *quantização*.

A imagem é um objeto matemático cuja a visualização não foi definida. Definiremos visualização de maneira formal:

Definição 2.1.2 (Visualização) Visualização é uma função

$$\begin{aligned} \nu : S^{\mathbb{k}} &\longrightarrow (\mathcal{M}, \mathcal{C}) \\ f &\longmapsto \nu(f), \end{aligned}$$

onde \mathcal{M} é a área física do monitor, \mathcal{C} é o conjunto de cores disponíveis no monitor para o software que está sendo empregado e $\nu(f)$ é o conjunto de pontos correspondentes a cada coordenada $s \in S$.

Definição 2.1.3 (Vizinhança) A vizinhança de qualquer posição $s \in S$ é qualquer conjunto de posições que não incluam s . Denotaremos por $\partial_s \subset S \setminus \{s\}$, onde “ \setminus ” é a diferença entre dois conjuntos. A definição mais geral de vizinhança envolve o conceito de distância, isto é,

$$\partial_s = \left\{ t \in S \setminus \{s\}; d(s, t) \leq k \right\},$$

em que $d(s, t)$ é uma distância entre s e t e k é não negativo.

Definição 2.1.4 (Sub- Imagem) Uma sub-imagem com respeito a janela $\overline{\partial_s}$ é uma função da forma

$$\begin{aligned} f_{\overline{\partial_s}} : \overline{\partial_s} &\longrightarrow \mathbb{R} \\ t &\longmapsto f_{\overline{\partial_s}}(t), \end{aligned}$$

onde $\overline{\partial_s} = \partial_s \cup \{s\}$ é o fecho de ∂_s .

Iremos exibir três propriedades básicas de imagens com apenas uma banda, isto é, quando $\mathbb{k} = \mathbb{R}$. Note que definindo \mathbb{k} dessa forma temos um caso particular do espaço de funções, entendendo espaços de funções como definido em Álgebra Linear.

- **Adição:** Sejam $f_1(s)$ e $f_2(s) \in S^{\mathbb{k}}$ duas imagens, então $g(s) = (f_1 + f_2)(s) = f_1(s) + f_2(s) \in S^{\mathbb{k}}$ ainda é uma imagem, para todo $s \in S$.
- **Multiplicação por escalar:** Sejam $f(s) \in S^{\mathbb{k}}$ e $\alpha \in \mathbb{R}$ uma imagem e um escalar, respectivamente. Então $g(s) = \alpha f(s)$ é uma imagem, para todo $s \in S$.
- **Elemento neutro aditivo :** Seja $\mathbf{0} \in S^{\mathbb{k}}$ definida como $\mathbf{0}(s) = 0$ para todo $s \in S$, então para todo $f(s) \in S^{\mathbb{k}}$ temos $\mathbf{0} + f(s) = f(s)$.

Baseada nas propriedades acima iremos definir outras operações importantes em imagens, são elas:

- **Negativo de uma imagem:** Seja $f(s) \in S^{\text{lk}}$ uma imagem. Se multiplicarmos essa imagem pelo escalar -1 ainda teremos uma imagem $g(s) = -f(s)$, para todo $s \in S$.
- **Diferença entre duas imagens:** Sejam $f_1(s)$ e $f_2(s) \in S^{\text{lk}}$ duas imagens, então $g(s) = (f_1 - f_2)(s) = f_1(s) - f_2(s) \in S^{\text{lk}}$ ainda é uma imagem, para todo $s \in S$.
- **Produto entre duas imagens:** Sejam $f_1(s)$ e $f_2(s) \in S^{\text{lk}}$ duas imagens, então $g(s) = (f_1 \cdot f_2)(s) = f_1(s) \cdot f_2(s) \in S^{\text{lk}}$ ainda é uma imagem, para todo $s \in S$.
- **Quociente entre duas imagens:** Sejam $f_1(s)$ e $f_2(s) \in S^{\text{lk}}$ duas imagens, então $g(s) = (f_1 / f_2)(s) = f_1(s) / f_2(s) \in S^{\text{lk}}$ ainda é uma imagem, para todo $s \in S$ e $f_2(s) \neq 0$.
- **Somatório de imagens:** Como S é finito é possível assumir a soma de uma imagem $f(s) \in S^{\text{lk}}$ como $\sum_{s \in S} f(s)$.
- **Média de uma imagem:** Seja $f(s) \in S^{\text{lk}}$ uma imagem. A média de uma imagem é dada por $\bar{f} = \sum_{s \in S} f(s) / (mn)$.
- **Produto interno entre duas imagens:** Sejam $f_1(s)$ e $f_2(s) \in S^{\text{lk}}$ duas imagens, então $g(s) = \langle f_1(s), f_2(s) \rangle = \overline{f_1(s)} \cdot f_2(s) \in S^{\text{lk}}$ ainda é uma imagem, para todo $s \in S$.
- **Norma de uma imagem:** Seja $f(s) \in S^{\text{lk}}$ uma imagem. A norma de uma imagem é dada por $\|f(s)\| = \sqrt{\langle f(s), f(s) \rangle}$.
- **Imagem constante:** É a imagem $\mathbf{1} \in S^{\text{lk}}$ definida como $\mathbf{1}(s) = 1$ para todo $s \in S$.

2.2 Radar

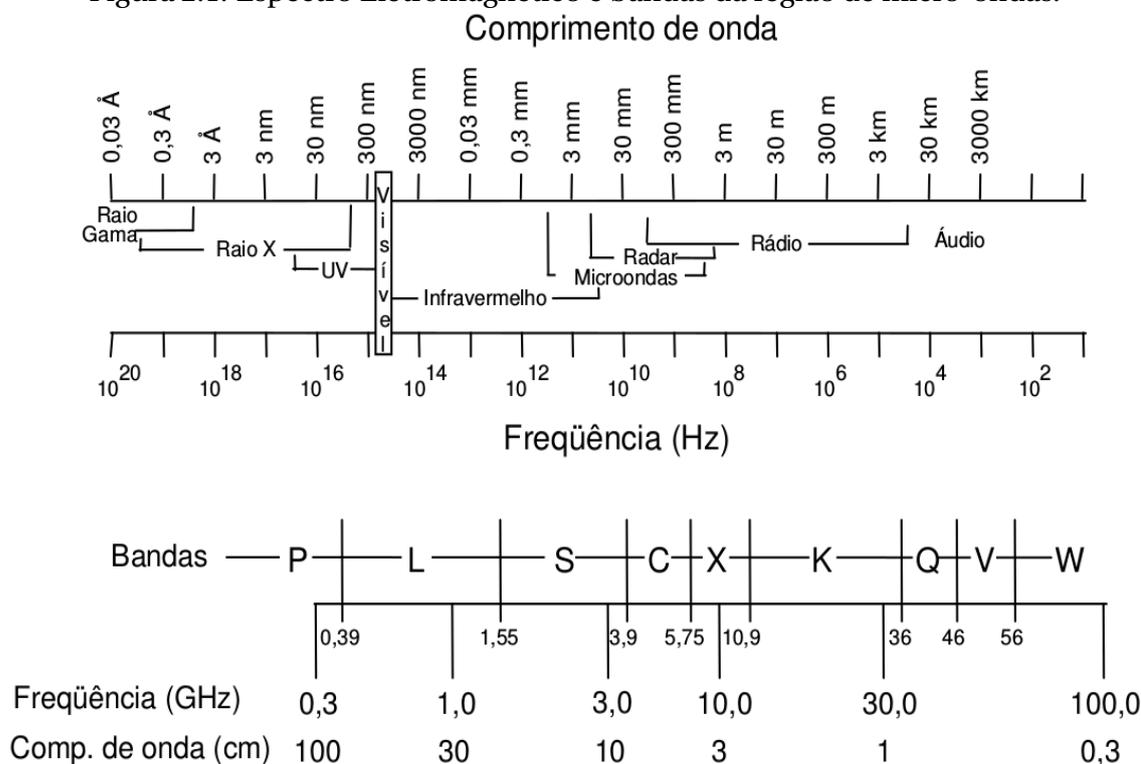
Segundo [Meneses & Almeida \(2012\)](#), radar é um sensor que se utiliza de uma fonte de radiação eletromagnética artificial. O termo RADAR significa “Radio Detection and Ranging”. Esse sensor é capaz de detectar um objeto (alvo), indicando sua distância (range) e posição ([Trevett, 1986](#)).

Os sensores podem ser classificados conforme a região do espectro eletromagnético que ele opera. Os sensores *ópticos* operam aproximadamente na faixa 0,30 a 15 μm e podem ser classificados em função do tipo de energia que detectam, por exemplo: *sensores termais* são aqueles que operam na região de 7,0 a 15 μm ; *sensores de energia refletida* são aqueles que operam na região de 0,38 a 3,0 μm . Algumas das desvantagens pertencentes a esses tipos

de sensores são: a necessidade de fonte externa de iluminação (*sensor passivo*) e a grande dependência das condições climáticas (Freitas et al., 2003).

Por outro lado, existem sensores que operam na faixa de micro-ondas do espectro eletromagnético, que é dividida em partes. Na década de 40, os militares tentaram manter segredo sobre as faixas espectrais as quais os primeiros sistemas de radar operavam. Nessa tentativa, letras maiúsculas foram utilizadas para identificar os comprimentos de onda das imagens de radar, essas regiões representadas por letras maiúsculas receberam o nome de bandas. A Figura 2.1 mostra valores de frequência bem como os de comprimento de onda.

Figura 2.1: Espectro Eletromagnético e bandas da região de micro-ondas.



Fonte: Imagem adaptada de Freitas et al., 2003.

Há autores que preferem usar a frequência, porque quando a radiação eletromagnética ultrapassa materiais com diferentes densidades, a frequência não é alterada, já o comprimento de onda e a velocidade se alteram (Bruder et al., 2003; Meneses & Almeida, 2012).

Os sensores que trabalham na faixa de micro-ondas tem as vantagens de não depender da iluminação solar e sofrerem pouca influência das condições climáticas. Os sensores de micro-ondas são divididos em dois grupos:

radiômetros sensores passivos que detectam a energia emitida pelo alvo analisado;

radares sensores ativos que possuem sua própria fonte de energia e captam a energia refletida pelo alvo.

Os sistemas de radar podem ser agrupados em imageadores e não imageadores. Os imageadores compreendem os sistemas de:

Antena rotatória ;

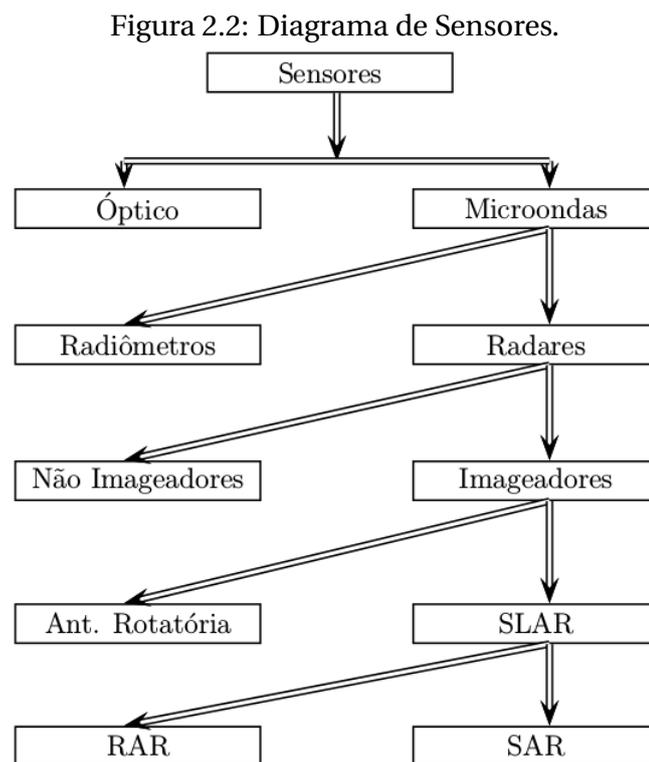
Visada lateral (SLAR) Side Looking Airborne Radar.

A definição de SLAR engloba dois tipos de radar, o de abertura:

Real (RAR) Real Aperture Radar;

Sintética (SAR) Synthetic Aperture Radar.

Para ilustrar a classificação de sensores descrita acima, construímos o diagrama exibido na Figura 2.2



Fonte: Imagem adaptada de Nascimento, 2012.

É interessante perceber que dependendo do alvo a ser identificado existem comprimentos de onda específicos a serem escolhidos. [Meneses & Almeida \(2012\)](#) mostra três exemplos desse tipo de escolha: Para identificar a variações texturais de solos, a melhor banda a ser utilizada é a banda de comprimento de onda X; para mapeamento geológico, a banda mais coerente é a L - devido as grandes dimensões das feições; e por fim se o interesse for uma grande penetração, por exemplo em regiões de mata a banda conveniente é a banda P.

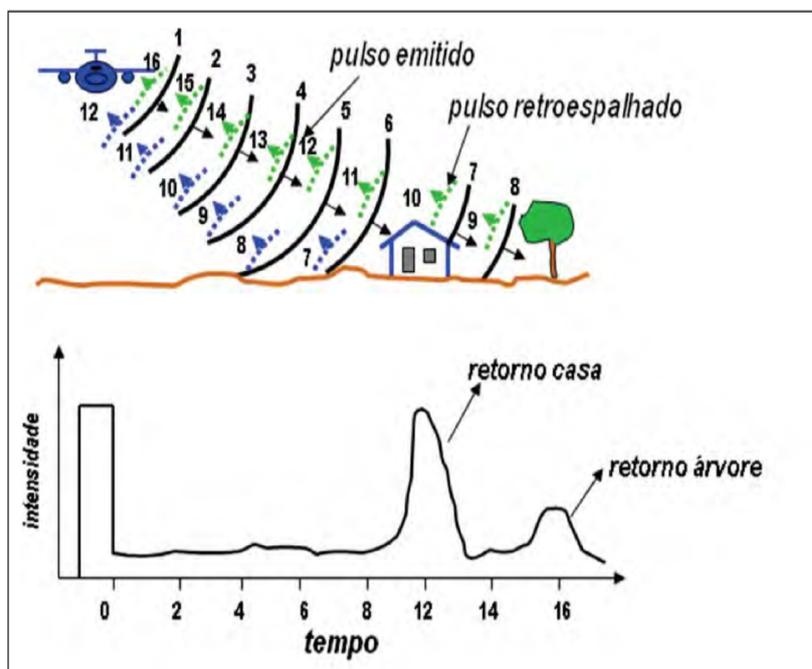
2.2.1 Radar de Abertura Sintética (SAR)

Em 1978, com o lançamento bem sucedido do satélite SEASAT que transportava a tecnologia SAR, os radares imageadores mostraram ser um instrumento capaz e indispensável para o sensoriamento remoto da superfície terrestre, permitindo de forma geral a discriminação do terreno e a detecção de alvos (Lee & Pottier, 2009).

A formação de uma imagem SAR requer o processamento coerente do sinal de retorno recebido após a emissão dos pulsos. As imagens SAR obtidas por esse processo podem ser representadas em intensidade ou amplitude (Freitas et al., 2003).

O radar opera enviando pulsos descontínuos de fótons que ao chegar ao terreno se espalham. Nesse processo da propagação no espaço o feixe se alarga na forma de um cone até atingir o alvo. Em seguida a energia se retroespalha e retorna para mesma antena, ou seja, esse tipo de antena utilizada é transmissora de pulsos eletromagnéticos de alta potência e receptora de sinais. Isso é feito ligando uma chave para emitir o pulso de alta potência e depois o desliga. Quando essa chave é ligada marca-se o tempo de transmissão do pulso e depois o desliga. Quando o sinal de retorno volta ao sensor é computado o tempo de chegada do sinal, o exposto acima é ilustrado pela Figura 2.3.

Figura 2.3: Ilustração da transmissão do pulso eletromagnético e retorno dos sinais após atingir o objeto.



Fonte: Imagem adaptada de Meneses & Almeida, 2012.

Os radares usados para aplicação de sensoriamento remoto são chamados de radares de visada lateral, pois emitem pulsos de radiação eletromagnética que utilizam uma antena

retangular pequena fixada na lateral da aeronave ou satélite. Além disso, o radar é apontado lateralmente em relação à direção da trajetória percorrida pela aeronave (Meneses & Almeida, 2012).

Na Figura 2.3, observamos a ilustração da transmissão do pulso eletromagnético e retorno dos sinais após atingir o objeto. A emissão de um pulso, no tempo, é representado pelos arcos de cor preta, os pulsos retroespalhados são representados por arcos menores nas cores azul e verde que retrata o sinal de retorno de uma casa e uma árvore, respectivamente. Os números representam o tempo que um pulso leva até chegar ao alvo e retornar ao sensor. Finalmente o gráfico ilustra a intensidade do pulso de cada objeto quando retornado para o sensor.

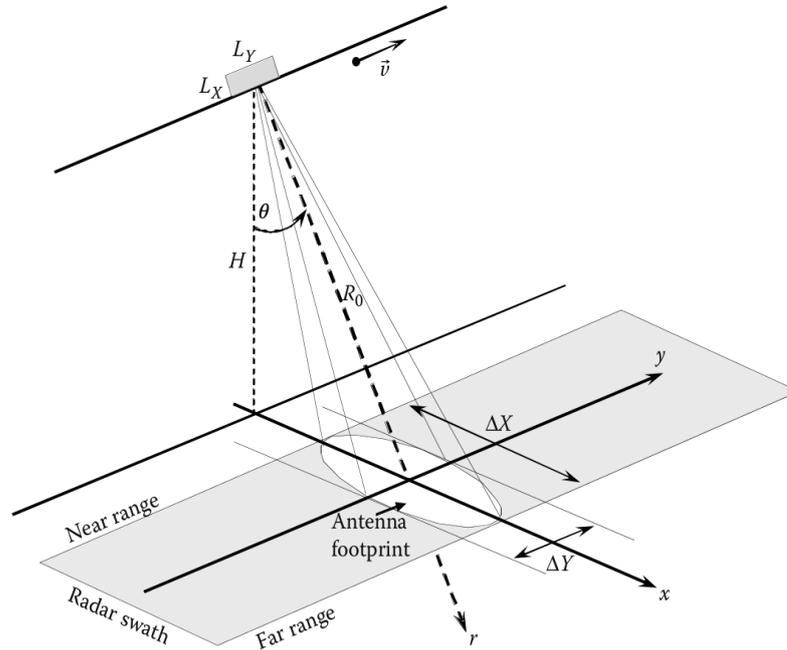
Imageamento Básico SAR

A mudança de tecnologia de radares imageadores de abertura real (RAR) para os de abertura sintética (SAR), foi devido a necessidade de se obter melhores resoluções azimutais. Para isso era necessário utilizar antenas longas (dezenas ou centenas de metros), ou seja, inviáveis de serem armazenadas em uma avião ou satélite. Outro problema do uso dessa tecnologia seria usar pulsos eletromagnéticos com pequenos comprimentos de onda forçando voos de baixa altura. A solução foi simular o comportamento de antenas grandes em antenas de pequeno porte. Essa modificação de tecnologia foi possível devido ao desenvolvimento de técnicas de processamento de sinais e modificações na forma de registrar os dados. Por esse motivo, atualmente SAR é uma das tecnologias mais utilizadas em sistemas de radares (Meneses & Almeida, 2012; Van Zyl & Kim, 2011).

Imagens da superfície geradas por radares aparentam ser idênticas às imagens produzidas por instrumentos que operam na faixa do visível ou infravermelho do espectro eletromagnético. Habitualmente, as imagens de radar são obtidas, conforme o avião ou satélite que transporta o sistema de radar move-se ao longo de sua trajetória. Imagens SAR tem resoluções que são independentes da distância entre o radar e o alvo. Por isso, um SAR transportado em um satélite tem vantagem sobre um RAR no que se refere a alta resolução. Os radares imageadores utilizam uma antena que ilumina a superfície por apenas um lado da pista de voo, ver Figura 2.4. Normalmente, a antena tem um feixe e ilumina uma área muito alongada, formando uma elipse na superfície (Van Zyl & Kim, 2011).

Podemos analisar na Figura 2.4 o esboço da geometria do imageamento SAR, que possui dimensões físicas L_X e L_Y , com uma altitude H e uma velocidade \vec{v} . Uma direção azimutal y e uma direção de alcance do terreno x . O feixe luminoso é direcionado para o plano com um ângulo de incidência θ . O eixo radial (*slant-range*) é denotado por r . Os alcances máximo (*Far range*) e mínimo (*Near range*) estão representados e também a faixa do radar (*Radar swath*). As faixas do radar e azimutal são denotadas por ΔX e ΔY , nessa ordem. Por fim, a distância entre o radar e o centro do *footprint* da antena é denotado por R_0 .

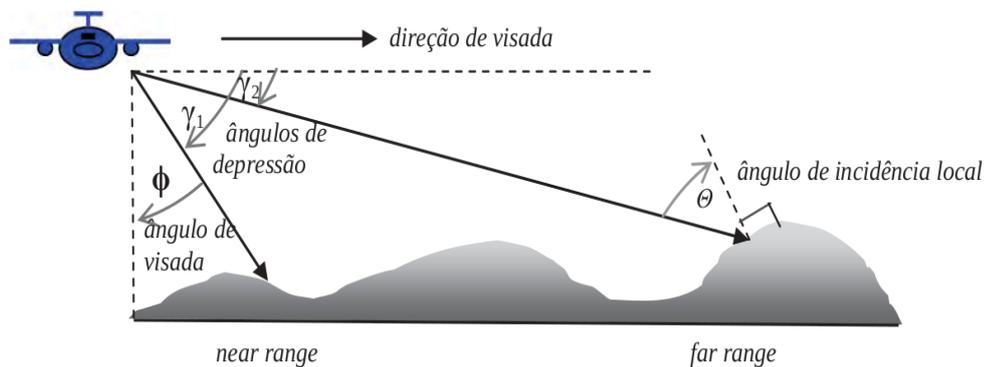
Figura 2.4: Esboço da geometria do imageamento SAR em uma visão tridimensional.



Fonte: Imagem adaptada de Lee & Pottier, 2009 e Van Zyl & Kim, 2011.

Para definir características da imagem parâmetros importantes são levados em consideração. São eles: polarização da onda, direção de azimute, direção de visada, ângulo de depressão, ângulo de visada e ângulo de incidência local. Iremos definir esses parâmetros com o auxílio da Figura 2.5.

Figura 2.5: Esboço da geometria do imageamento SAR em uma visão frontal.



Fonte: Imagem adaptada de Meneses & Almeida, 2012 e Van Zyl & Kim, 2011.

- **Direção de azimute:** É a direção de voo da plataforma.
- **Direção de visada:** É a direção de iluminação ao longo da qual o pulso de radar se propaga.

- **Ângulo de depressão:** É o ângulo formado entre a linha de horizonte de voo e a linha que define a direção de visada.
- **Ângulo de visada:** É o complemento do ângulo de depressão.
- **Ângulo de incidência local:** É o ângulo formado entre o pulso de radar e a vertical ao plano da superfície no ponto onde o pulso toca o terreno.

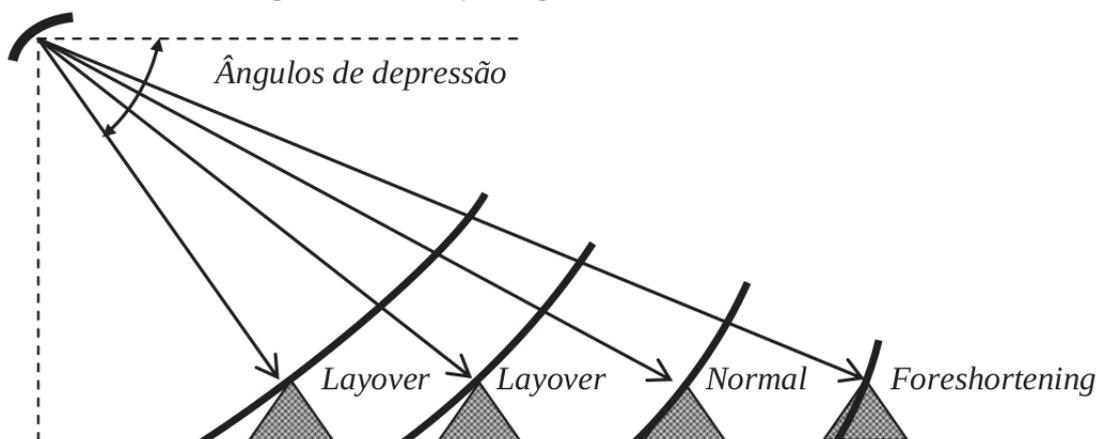
Resolução Espacial

Os parâmetros que controlam a célula de resolução espacial de um radar de visada lateral são independentes: comprimento do pulso e largura angular do feixe do pulso. Se essas medidas são independentes, então há duas medidas de resolução espacial:

- **Resolução em alcance ou longitudinal:** É a medida ao longo da direção de iluminação do pulso de onda.
- **Resolução azimutal ou transversal:** É a medida na direção paralela à direção da linha de voo.

As imagens dos objetos mais próximos da antena acabam sendo deformadas, pois a imagem é construída pelo registro de sinais medidos no plano inclinado. Dois tipos de distorções, de natureza mais complexas são: *Layover* e o *Foreshortening*.

Figura 2.6: Esboço da geometria da Resolução.



Fonte: Imagem adaptada de Meneses & Almeida, 2012 e Van Zyl & Kim, 2011.

Tomemos como exemplo quatro morros simétricos, como mostra a Figura 2.6. Observe que nos dois primeiros morros a onda atinge primeiro o topo do morro e depois a base. Assim, acontece uma inversão do relevo. Esse tipo de distorção recebe o nome de *Layover*. O encurtamento da rampa recebe o nome de *Foreshortening*, como ilustrado no último morro

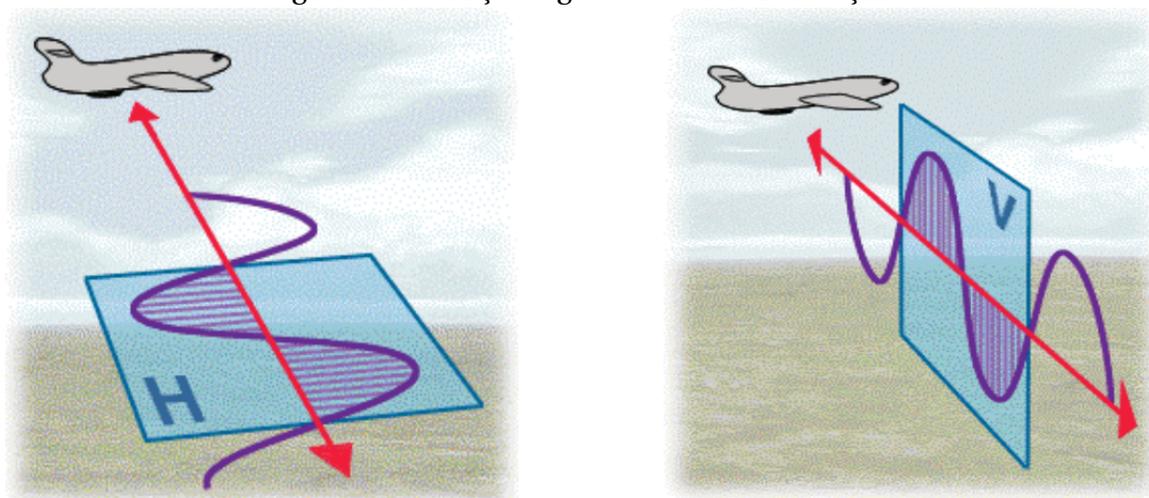
da Figura 2.6. Que faz com que a face da encosta de frente para a antena apresente-se na imagem em tons mais claros. Mesmo com essas distorções a visada lateral é a razão principal para a eficiência de imagens de radar no que se refere a analisar texturas, pois criam uma percepção visual tridimensional do relevo (Meneses & Almeida, 2012; Van Zyl & Kim, 2011).

2.2.2 Radar de Abertura Sintética Polarizado (PolSAR)

Iniciaremos essa seção com o objetivo de definir polarização. Três autores foram fundamentais para termos a compreensão sobre esse termo, são eles: Erasmus Bartolinus que publicou o primeiro trabalho quantitativo em observação em 1669; C. Huygens que propôs a natureza ondulatória da luz e descobriu a luz polarizada em 1677; e finalmente E.L. Malus provou a conjectura de Newton que dizia que “a polarização é uma propriedade intrínseca da luz” em 1808 (Lee & Pottier, 2009).

“Quando um pulso de energia é emitido pela antena de um radar e o vetor do seu campo elétrico é controlado por um cristal polarizador com eixo de transmissão situado na horizontal ou na vertical, para fazer vibrar a onda em uma direção horizontal (H) ou vertical (V) dizemos que a onda eletromagnética está polarizada (Meneses & Almeida, 2012).” Ver ilustração na Figura 2.7. Iremos definir de maneira formal conforme Freitas et al. (2003).

Figura 2.7: Esboço da geometria da Polarização.



Fonte: Imagem adaptada de Meneses & Almeida, 2012.

Definição 2.2.1 (Polarização) é a orientação segundo a qual oscila, no tempo, o vetor campo elétrico de onda eletromagnética.

Um pulso quando é transmitido pode ser emitido na direção horizontal ou vertical. Ao alcançar o alvo acontece uma despolarização, isso significa que o retorno do sinal pode ser na

mesma direção que foi emitida ou não. Por exemplo, se o sensor emite um pulso na direção V o sinal do retorno pode ser V ou H, ver Figura 2.7.

Assim, temos as possíveis polarizações: HH, HV, VH e VV. As polarizações HV e VH são chamadas de **polarizações cruzadas**. Já as polarizações HH é conhecida como **polarização plano-polarizada**, e por fim a polarização VV recebe o nome de **polarização plano-paralela**. Uma observação interessante sobre o imageamento em vegetações é a tendência que ela possui para mostrar diferentes retornos nas polarizações, devido a despolarização nos vários retroespalhamentos das folhas e galhos (Meneses & Almeida, 2012; Van Zyl & Kim, 2011).

O início da pesquisa em imagens de radar polarimétrico foi entre as décadas de 40 e 60, com o foco para utilização de radar polarimétrico para caracterizar alvos de aeronaves. Em 1985, com a implementação bem sucedida de AIRSAR as imagens de radar polarimétrico obtiveram um avanço, ver Figura 2.8.

Figura 2.8: Fotos de um sistema aerotransportado.



(a) AIRSAR



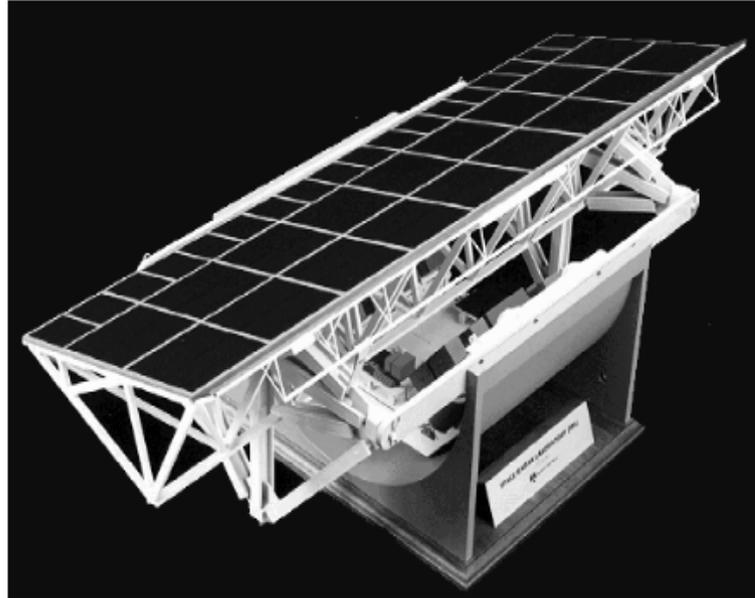
(b) Sensor SAR implantado no AIRSAR

Fonte: Disponível em <http://airsar.jpl.nasa.gov/index_detail.html> Acesso em jan. 2016.

Além disso, um outro avanço considerável para imagens PolSAR foi o lançamento com sucesso do SIR-C/X-SAR em 1994 adquirindo imagens SAR digitais da terra, ver figura 2.9. Em seguida, com o lançamento do RADARSAT-2 em 2007. Com estes satélites lançados nesse período foi possível obter informações, por exemplo, para monitoramento remoto de risco de inundações e terremotos, cobertura de neve, corrente de oceano etc (Lee & Pottier, 2009).

Satélites que utilizam as técnicas de sensoriamento PolSAR fornecem informações úteis para extração de parâmetros biofísicos e geofísicos da superfície terrestre. Destacam-se aplicações no monitoramento de: desmatamentos, mapeamentos de queimadas, umidade do solo, delimitação de áreas inundadas, detecção de vazamento de óleo, gestão de riscos de inundações e de terremotos, entre outras.

Figura 2.9: Sistema espacial SIR-C/X-SAR.



Fonte: Disponível em <<http://trmm.jpl.nasa.gov/SIRCdesc.html>> Acesso em jan. 2016.

Em suma, uma imagem SAR é um arranjo bidimensional cujos pixels estão associados com pequenas áreas da superfície da terra. Cada uma desses pixels fornece um número complexo associado a refletividade de todos espalhamentos contidos na célula de resolução de SAR (Lee & Pottier, 2009).

As Figuras 2.8, 2.10 (a), 2.10 (b) mostram alguns sensores de radar polarimétricos aero-transportados, para maiores detalhes consultar (Lee & Pottier, 2009).

2.3 Alguns Conceitos de Estatística Multivariada

Para definir os conceitos relevantes para nosso trabalho iremos nos basear em: Johnson & Wichern (2007), James (2013), Magalhães (2011) e Goodman (1963).

Definição 2.3.1 (Variável aleatória multivariada) Dizemos que uma variável aleatória multivariada é uma função da forma

$$\begin{aligned} \mathbf{X} : \Omega &\longrightarrow \mathbb{R}^n \\ \omega &\longmapsto \mathbf{X}(\omega) = (X_1(\omega), \dots, X_n(\omega)), \end{aligned}$$

em que, Ω é o espaço amostral, ω é um evento na σ -álgebra definida sobre Ω e cada $X_i(\omega)$, em que $i = 1, \dots, n$, é uma variável aleatória.

Definição 2.3.2 (Função distribuição acumulada multivariada) Se \mathbf{X} é um vetor aleatório,

Figura 2.10: Sistemas aerotransportados.



(a) Convair-580 C/X SAR



(b) E-SAR

Fonte: Disponível em <

<http://www.nrcan.gc.ca/earth-sciences/geomatics/satellite-imagery-air-photos/satellite-imagery-products/educational-resources/9335> e

<http://brandenburg.geoecology.uni-potsdam.de/projekte/opaque/air_radar.php>.

Acesso em fev. 2016.

sua função distribuição acumulada é definida por

$$F : \mathbb{R}^n \longrightarrow \mathbb{R}$$

$$\mathbf{X} \longmapsto F(\mathbf{X}) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

Definição 2.3.3 (Vetor média) O valor esperado de um vetor aleatório \mathbf{X} é dado por

$$E(\mathbf{X}) = (E(X_1), \dots, E(X_n)).$$

A matriz de covariância de um vetor aleatório é a matriz de dimensão $n \times n$ dada por $\Sigma = E(X_i X_j) - E(X_i)E(X_j)$, com $1 \leq i, j \leq n$. Os elementos da diagonal principal são as variâncias $\sigma_{ii} = \text{Var}(X_i)$ e os elementos fora da diagonal principal são as covariâncias $\sigma_{ij} = E(X_i X_j) - E(X_i)E(X_j)$. A matriz de correlação R é a matriz de dimensão $n \times n$ dada pelos coeficientes de correlação

$$\rho_{ij} = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_i)\text{Var}(X_j)}}.$$

É fácil ver que os elementos da diagonal principal são todos iguais a 1, isto é, $\rho_{ii} = 1$. Ainda sobre correlação é bastante conhecido que: se Σ é positiva definida (isto é, $v^t R v > 0 \forall$ vetor coluna $v \neq 0$) temos as seguintes conclusões $-1 < \rho_{ij} < 1$ e $\rho_{ij} = 0$ quando X_i e X_j são variáveis aleatórias independentes.

Definição 2.3.4 (Variável Aleatória Gaussiana Complexa) Dizemos que uma variável aleatória univariada $Z = X + iY$ é uma variável aleatória complexa se as partes real X e imaginária Y são variáveis aleatórias definidas no mesmo espaço de probabilidade. No caso em que

(X, Y) seguem uma distribuição Gaussiana bivariada a variável aleatória Z segue uma distribuição Gaussiana complexa univariada.

Definição 2.3.5 (Variável aleatória Gaussiana complexa multivariada) Sejam $\mathbf{X} = (X_1, \dots, X_n)$ e $\mathbf{Y} = (Y_1, \dots, Y_n)$ vetores aleatórios tal que o vetor $(X_1, Y_1, \dots, X_n, Y_n)$ de partes real e imaginária segue uma distribuição Gaussiana $2n$ -variada. Dizemos que um vetor aleatório da forma $\mathbf{Z} = (Z_1, \dots, Z_n)$, em que Z_1, \dots, Z_n são variáveis aleatórias complexas, é chamado de vetor aleatório Gaussiano complexo.

Definição 2.3.6 (Variância complexa) A variância complexa de uma variável aleatória complexa Z é $E(|Z|^2)$, em que $|\cdot|$ é o módulo de um número complexo.

Definição 2.3.7 (Covariância complexa) A covariância complexa entre duas variáveis aleatórias complexas Z_i e Z_j é dada por $E(Z_i Z_j^*)$, onde $E(\cdot)$ é o operador esperança estatística e $*$ é o conjugado de um número complexo.

Definição 2.3.8 (Vetor média complexo) O vetor média de uma variável aleatória complexa multivariada \mathbf{Z} é dado por

$$E(\mathbf{Z}) = (E(Z_1), \dots, E(Z_n)),$$

em que $E(Z) = E(X) + iE(Y) = \mu_X + \mu_Y$.

Definição 2.3.9 (Matriz de covariância complexa) A Matriz de covariância complexa $C_{n \times n}$ é dada por

$$C_{n \times n} = E(\mathbf{Z}\mathbf{Z}^*) = \begin{bmatrix} E(Z_1 Z_1^*) & \dots & E(Z_1 Z_n^*) \\ \vdots & \ddots & \vdots \\ E(Z_n Z_1^*) & \dots & E(Z_n Z_n^*) \end{bmatrix}.$$

Definição 2.3.10 (Matriz de correlação complexa) A Matriz de correlação complexa $R_{n \times n}$ é definida por

$$R_{n \times n} = \begin{bmatrix} \frac{E(Z_1 Z_1^*)}{\sqrt{E(|Z_1|^2)E(|Z_1|^2)}} & \dots & \frac{E(Z_1 Z_n^*)}{\sqrt{E(|Z_1|^2)E(|Z_n|^2)}} \\ \vdots & \ddots & \vdots \\ \frac{E(Z_n Z_1^*)}{\sqrt{E(|Z_n|^2)E(|Z_1|^2)}} & \dots & \frac{E(Z_n Z_n^*)}{\sqrt{E(|Z_n|^2)E(|Z_n|^2)}} \end{bmatrix}.$$

Teorema 2.3.1 O $\det(R) = |\det(C)|^2$, em que $\det(\cdot)$ é o determinante de uma matriz.

Demonstração. Para maiores detalhes sobre a demonstração ver [Goodman \(1963\)](#).

A *variância efetiva* e a *dependência efetiva* para vetores aleatórios reais são medidas descritivas para dados multivariados. Seja \mathbf{X} um vetor aleatório com variáveis aleatórias reais, a variância efetiva é dada por

$$V_e(\mathbf{X}) = |\Sigma|^{\frac{1}{p}} = (\lambda_1 \cdots \lambda_p)^{\frac{1}{p}},$$

em que Σ é a matriz de covariância, $|\cdot|$ é o determinante, p é a dimensão de Σ e os λ_i (com $i = 1, \dots, p$) são os autovalores da matriz de covariância. A dependência efetiva é definida por

$$D_e(\mathbf{X}) = 1 - |R|^{\frac{1}{p}},$$

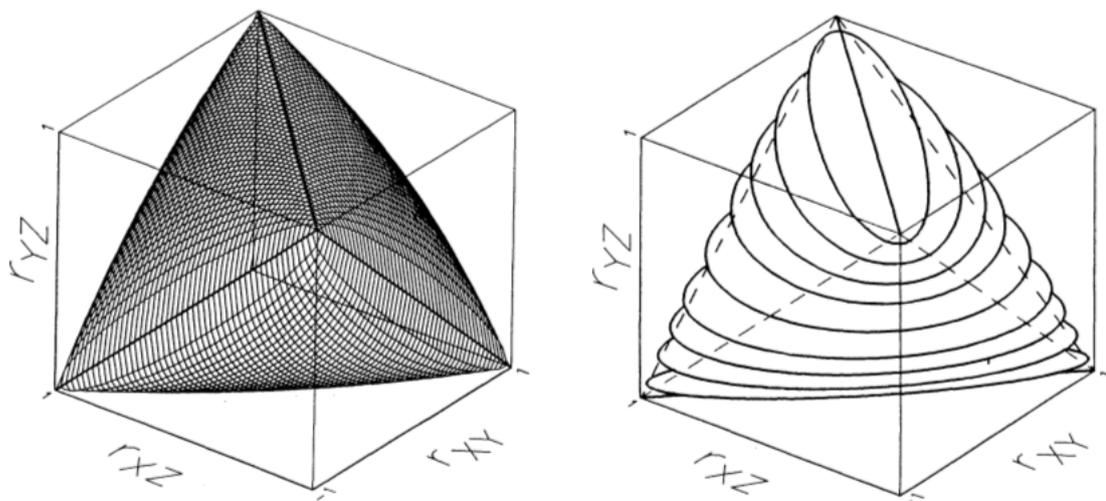
em que R é a matriz de correlação de Σ . Essas medidas tem a capacidade de permitir comparar variáveis aleatórias multivariadas de diferente dimensão (Peña & Rodríguez, 2003).

2.4 Visualização de Dados Multivariados

Nessa seção iremos fazer um breve resumo sobre algumas técnicas de visualização de dados multivariados. Normalmente, essas técnicas são desenvolvidas para visualizar uma única matriz de correlação, porém todas elas com entradas de números reais.

Rousseeuw & Molenberghs (1994) construíram um gráfico tridimensional de todas as combinações possíveis ($\rho_{XY}, \rho_{XZ}, \rho_{YZ}$), onde X, Y e Z são variáveis aleatórias. Este gráfico recebeu o nome de *tetraedro elíptico* o qual é um conjunto convexo, entendendo um conjunto convexo como um conjunto que: dados dois pontos pertencentes a ele e traçarmos um segmento de reta ligando esses dois pontos o segmento continua pertencendo ao conjunto, ver Figura 2.11.

Figura 2.11: Tetraedro elíptico.



(a) Conjunto de todas as correlações possíveis entre X e Y .

(b) Seccionando o conjunto mostrado ao lado com planos perpendiculares ao plano r_{YZ} .

Fonte: Rousseeuw & Molenberghs, 1994.

É uma superfície suave apenas acentuada nos vértices que contém as bordas de um

tetraedro. Todas as suas seções transversais perpendiculares aos eixos são elipses. Esses gráficos podem ajudar a desenvolver intuição para as restrições que existem entre correlações calculada a partir de três ou mais variáveis.

A exibição gráfica proposta por [Murdoch & Chow \(1996\)](#) apresenta correlações como elipses para expor a matriz de correlação de forma simples. Esse gráfico é bastante intuitivo para matrizes de correlações grandes. As elipses são contornos de uma distribuição normal bivariada, onde cada ponto é dado por $P = \left(\cos\left(\theta + \frac{d}{2}\right), \cos\left(\theta - \frac{d}{2}\right) \right)$, com $\theta \in [0, 2\pi]$ e $\cos(d) = \rho$. Como exemplo utilizamos o conjunto de dados *mtcars* (*Motor Trend Car Road Tests*), um conjunto de dados disponível na linguagem R. Esses dados foram retirados da revista *Motor Trend 1974* e estão formados por 32 observações em 11 variáveis. As 11 variáveis são: mpg (Milhas / (US) galão), cyl (Número de cilindros), disp (Deslocamento), hp (potência bruta), drat (relação do eixo traseiro), wt (Peso (1000 libras)), qsec (tempo de 1/4 de milha), vs (v/s), am (Transmissão (0 = automática, 1 = manual)), gear (Número de velocidades para a frente) e carb (Número de carburadores). A Tabela 2.1 mostra a matriz de correlação dos dados *mtcars*.

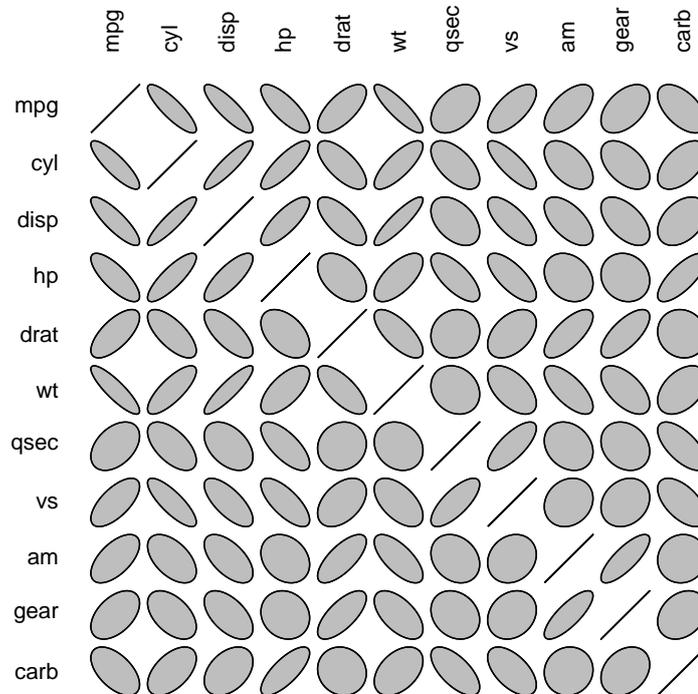
Tabela 2.1: Correlação entre 11 variáveis de performance automobilística.

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mpg	1.00	-0.85	-0.85	-0.78	0.68	-0.87	0.42	0.66	0.60	0.48	-0.55
cyl	-0.85	1.00	0.90	0.83	-0.70	0.78	-0.59	-0.81	-0.52	-0.49	0.53
disp	-0.85	0.90	1.00	0.79	-0.71	0.89	-0.43	-0.71	-0.59	-0.56	0.39
hp	-0.78	0.83	0.79	1.00	-0.45	0.66	-0.71	-0.72	-0.24	-0.13	0.75
drat	0.68	-0.70	-0.71	-0.45	1.00	-0.71	0.09	0.44	0.71	0.70	-0.09
wt	-0.87	0.78	0.89	0.66	-0.71	1.00	-0.17	-0.55	-0.69	-0.58	0.43
qsec	0.42	-0.59	-0.43	-0.71	0.09	-0.17	1.00	0.74	-0.23	-0.21	-0.66
vs	0.66	-0.81	-0.71	-0.72	0.44	-0.55	0.74	1.00	0.17	0.21	-0.57
am	0.60	-0.52	-0.59	-0.24	0.71	-0.69	-0.23	0.17	1.00	0.79	0.06
gear	0.48	-0.49	-0.56	-0.13	0.70	-0.58	-0.21	0.21	0.79	1.00	0.27
carb	-0.55	0.53	0.39	0.75	-0.09	0.43	-0.66	-0.57	0.06	0.27	1.00

A simplicidade na exibição gráfica proposta por [Murdoch & Chow \(1996\)](#) é mostrada na Figura 2.12. A inclinação da elipse para direita indica correlação positiva e para esquerda negativa, quando a elipse fica mais fina sinaliza que a correlação é alta entre as variáveis.

[Friendly \(2002\)](#) propôs um conjunto de técnicas, para exibir padrões nas correlações, que tem como base 2 regimes: o valor de renderização de uma correlação (*correlation rendering*) e a reordenação das variáveis (*correlation ordering*). O primeiro regime retrata o sinal e magnitude da correlação já o segundo coloca as variáveis semelhantes, no sentido de uma ordem angular entre os auto vetores da matriz de correlação, juntas para facilitar a percepção visual. Essa magnitude do sinal é feita colocando cores quando os valores se afastam de zero. Se o afastamento é positivo adota-se a cor azul e para o caso negativo a cor vermelha. Além disso, foram colocadas linhas brancas nas diagonais dos quadrados menores para indicar se a correlação é positiva ou negativa para o caso de uma visualização em preto e branco, como

Figura 2.12: Matriz de correlações em forma de elipses.



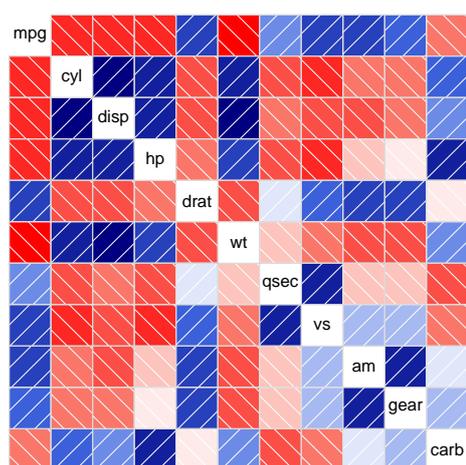
Fonte: processada pelo autor, 2016.

ilustra a Figura 2.13.

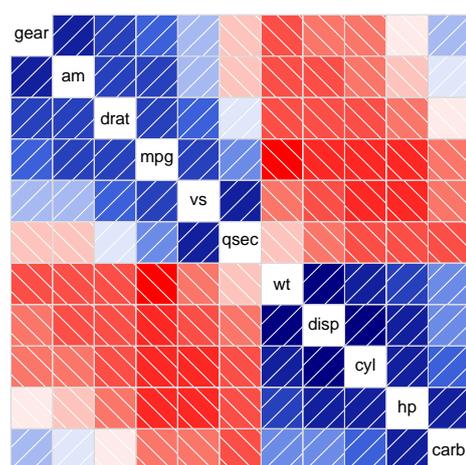
Diferentemente da Figura 2.13 (a) é possível perceber na Figura 2.13 (b) um agrupamento relativamente homogêneo com correlações negativas intermediárias - entendendo intermediárias como correlações não muito altas e nem muito baixas - entre o conjunto de variáveis wt, disp e cyl e o conjunto de variáveis gear, am e drat. Também é possível observar que a variável qsec tem correlação negativa mais baixa com o conjunto de variáveis wt, disp, cyl, hp e cab.

O presente capítulo apresentou alguns conceitos sobre processamento de imagens, radar e estatística multivariada. No final, expôs algumas técnicas sobre visualização de dados multivariados.

Figura 2.13: Gráfico *corrgram* para matrizes de correlação.



(a) Ordenamento em ordem alfabética das variáveis.



(b) Ordenamento com base nos ângulos dos auto vetores.

Fonte: Processada pelo autor, 2016.

3

VISUALIZAÇÃO DE MATRIZES DE COVARIÂNCIA COMPLEXAS EM DADOS POLARSAR

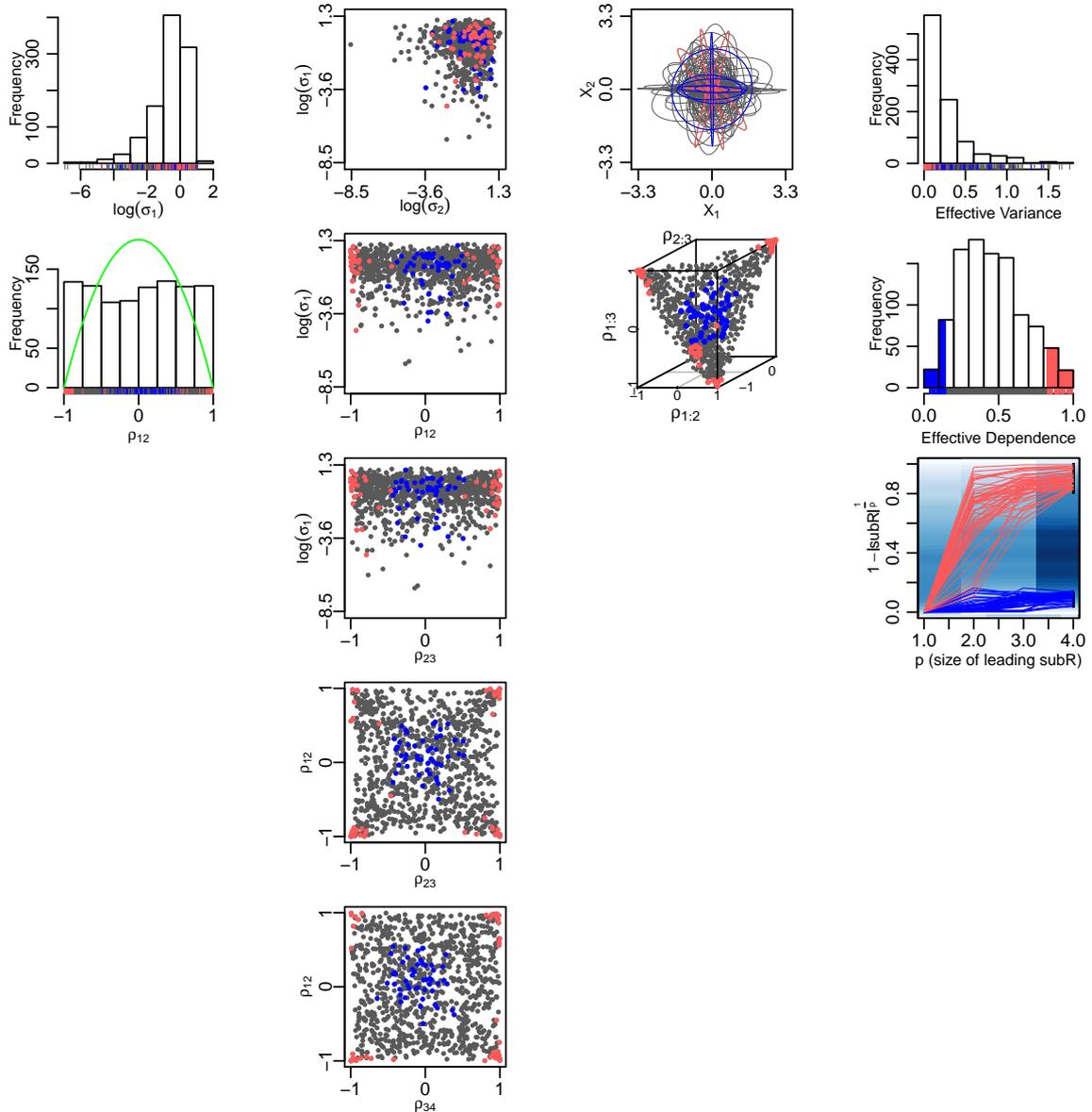
ESTE capítulo tem o objetivo de apresentar uma abordagem para a visualização de dados multivariados complexos, em particular, dados PolSAR com a finalidade de extrair novas informações sobre as imagens.

3.1 Metodologia

Existem várias ferramentas estatísticas para a análise de uma estrutura multivariada de dados, dentre elas podemos citar: análise de componentes principais, análise fatorial, análise de correlação canônica etc. Essas técnicas têm a finalidade de reduzir a estrutura multivariada de alta dimensão para uma dimensão menor, de modo que as relações entre as variáveis que compõem o dado possam ser melhor compreendidas (Friendly, 2002).

Tokuda et al. (2011) propuseram uma ferramenta de visualização em camadas. Esse método normalmente utiliza 1000 matrizes de covariância por amostragem. A exibição desse método é em quatro camadas: a primeira camada é composta por histogramas univariados: um do logaritmo dos desvios padrão $\log \sigma_1$ e o outro das correlações ρ_{12} . A segunda camada é um conjunto de gráficos de dispersão bivariados de desvios padrões ou correlações. Na terceira é construído um gráfico composto por várias elipses e um gráfico tridimensional de correlações. Por fim, a quarta camada tem o objetivo de reduzir cada matriz de covariância nas medidas escalares: variância efetiva e dependência efetiva, ver Figura 3.1.

Figura 3.1: Método de visualização em 4 camadas: VisCov.



Fonte: Processada pelo autor, adaptada de Tokuda, 2011.

A nossa proposta de pesquisa - aqui desenvolvida - é baseada na visualização de matrizes de covariância, porém as nossas matrizes são complexas, em particular, matrizes hermitianas. Entendendo matriz hermitiana como uma matriz complexa igual a sua transposta conjugada. Antes de detalhar o nosso método iremos discorrer de maneira geral como é construída a nossa abordagem de visualização de dados multivariados complexos. De forma análoga ao trabalho desenvolvido por Tokuda et al. (2011), a nossa abordagem é apresentada em camadas, entendendo camadas por colunas contendo vários gráficos de distribuição amostral de várias estatísticas.

Primeira Coluna

- é formada por um histograma univariado do logaritmo dos desvios padrão das matrizes

de covariância complexa C_i ($i = 1, \dots, k$),

- um histograma univariado das correlações das matrizes de correlação complexa R_i ($i = 1, \dots, k$) e
- um gráfico tridimensional das correlações das matrizes de correlação complexa R_i ($i = 1, \dots, k$).

Segunda Coluna

- exibe um gráfico de dispersão entre os $\log \sigma_1 \times \log \sigma_2$ das matrizes de covariância complexa C_i ($i = 1, \dots, k$) e
- o segundo um gráfico de dispersão entre as correlações ρ_{23} e ρ_{12} das matrizes de correlação complexa R_i ($i = 1, \dots, k$).

Terceira Coluna

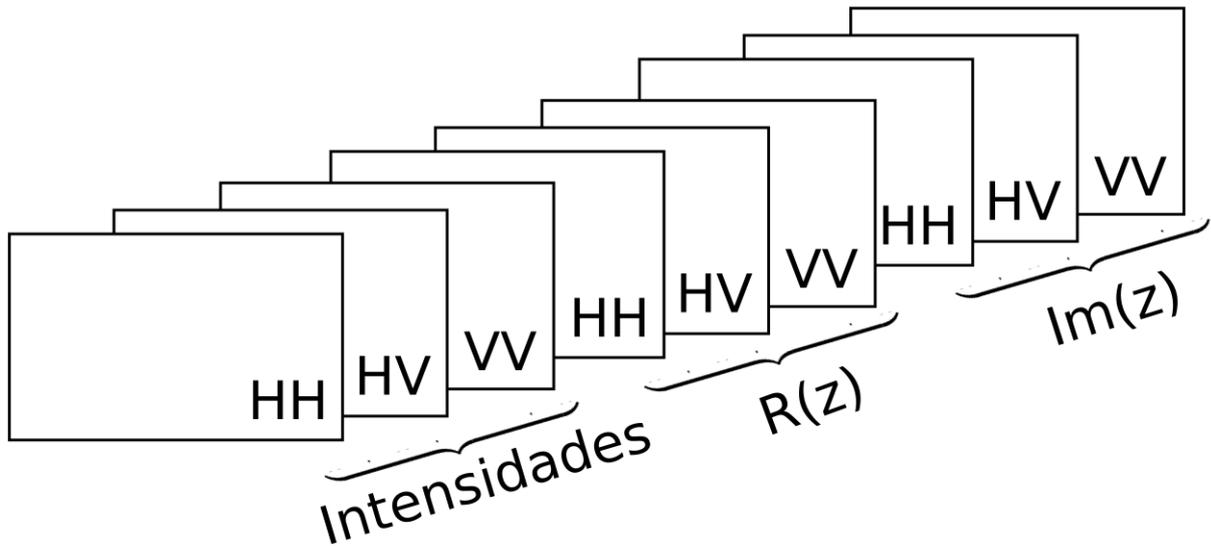
- mostra um histograma das variâncias efetiva das matrizes de covariância complexa C_i ($i = 1, \dots, k$) e
- um gráfico de dispersão entre as correlações ρ_{12} das matrizes de correlação complexa R_i ($i = 1, \dots, k$) e o logaritmo dos desvios padrão $\log \sigma_1$ das matrizes de covariância complexa C_i ($i = 1, \dots, k$).

Quarta Coluna

- apresenta um histograma das dependências efetiva das matrizes de correlação complexa R_i ($i = 1, \dots, k$) e
- um gráfico de dispersão entre as correlações ρ_{23} das matrizes de correlação complexa R_i ($i = 1, \dots, k$) e o logaritmo dos desvios padrão $\log \sigma_1$ das matrizes de covariância complexa C_i ($i = 1, \dots, k$).

Essas operações de correlação e desvio padrão são feitas com a parte real de um número complexo, denotada por $\Re(z)$. As imagens PolSAR utilizadas nesse trabalho estão no formato enxuto, entendendo por formato enxuto a organização de uma imagem PolSAR que possui três bandas em um arranjo de nove dimensões. As três primeiras dimensões do arranjo contém os dados que representam as intensidades definidas na página 10, as três dimensões seguintes contém a parte real do número complexo e finalmente as três últimas dimensões a parte imaginária de um número complexo como ilustra a Figura 3.2.

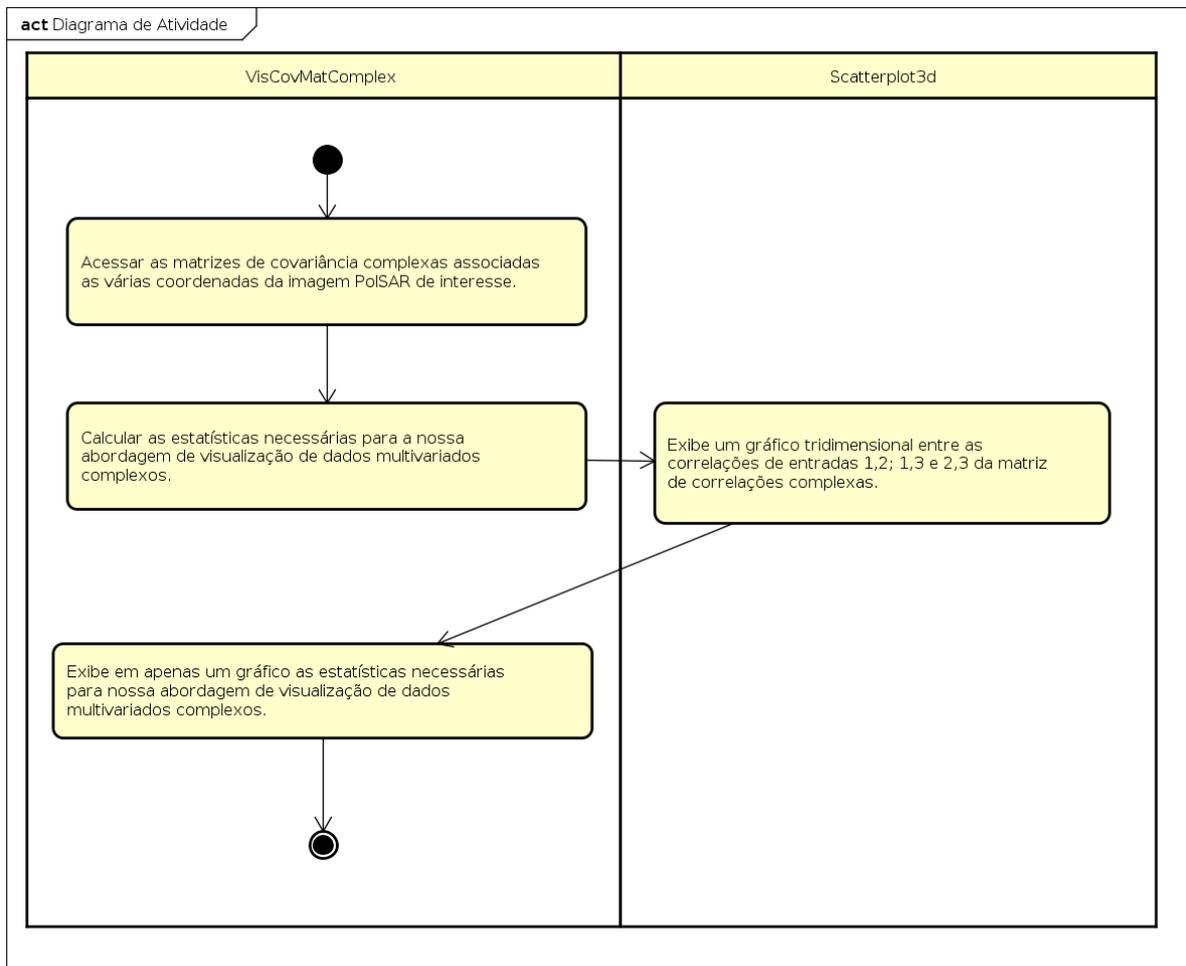
Figura 3.2: Ilustração de uma imagem PolSAR no formato enxuto.



Fonte: autor, 2016.

Dentre os vários diagramas oferecidos por uma UML (Unified Modeling Language) optamos em utilizar um diagrama de atividades (Ver Figura 3.3), pois através dele conduzimos o leitor ao que a nossa abordagem executa, além disso o diagrama de atividades é o mais adequado para uma biblioteca não orientada a objetos (Bell, 2016).

Figura 3.3: Diagrama de Atividades para obtenção dos elementos visuais da ferramenta.



powered by Astah

Fonte: autor, 2016.

3.1.1 A ferramenta desenvolvida para visualizar matrizes de covariância complexa

Nessa subseção, iremos explicar como utilizar a ferramenta desenvolvida para nossa abordagem de visualização de dados multivariados complexos. A implementação da nossa abordagem foi produzida na linguagem R. A função **VisMatCovComplex** deve ser aplicada em dados PolSAR no formato enxuto e necessita das seguintes variáveis:

- uma Imagem PolSAR no formato enxuto;
- a linha onde vai iniciar a região da imagem a qual se pretende analisar;
- a linha onde vai terminar a região da imagem a qual se pretende analisar;
- a coluna onde vai iniciar a região da imagem a qual se pretende analisar;
- a coluna onde vai terminar a região da imagem a qual se pretende analisar.

O nome **VisMatCovComplex** é derivado de “Visualização de Matrizes de Covariância Complexa”. O **VisMatCovComplex** é uma ferramenta composta por vários gráficos de distribuição amostral de várias estatísticas. A nossa ferramenta permite uma comparação fácil entre diferentes conjuntos de dados PolSAR.

O Algoritmo 3.1 mostra como usar a nossa ferramenta em R. A ferramenta é *Open Source* e está disponível em <https://sites.google.com/site/antoniomarcoslarangeiras/dissemtacao>.

Algoritmo 3.1: Ilustração da aplicação da função **VisMatCovComplex**

```
1 source("VisMatCovComplex")
2 VisMatCovComplex(Imagem, linha.inicial, linha.final, coluna.inicial, coluna.final)
```

Além disso, a nossa ferramenta oferece a possibilidade de aplicar uma das estatísticas disponíveis para nossa camada em até quatro imagens com o objetivo de comparar as formas exibidas nos gráficos de distribuição amostral da estatística escolhida para cada imagem.

Isso é possível através da função **VisMatCovComplex.selecionada**. Essa função precisa das variáveis (ver Algoritmo 3.2):

- o nome da estatística a ser aplicada;
- as imagens a qual deseja-se analisar.

As estatísticas definidas na página 32 são empregadas na função **VisMatCovComplex.selecionada** como:

Histograma.Variancia, Grafico2D.s1_s2, Histograma.VarianciaEfetiva, Corr3D, Grafico.s1_r23, Histograma.Corr, Grafico2D.r12_r23, Grafico.s1_r12 e Histograma.DependenciaEfetiva.

Algoritmo 3.2: Ilustração da aplicação da função `VisMatCovComplex.selecionada`

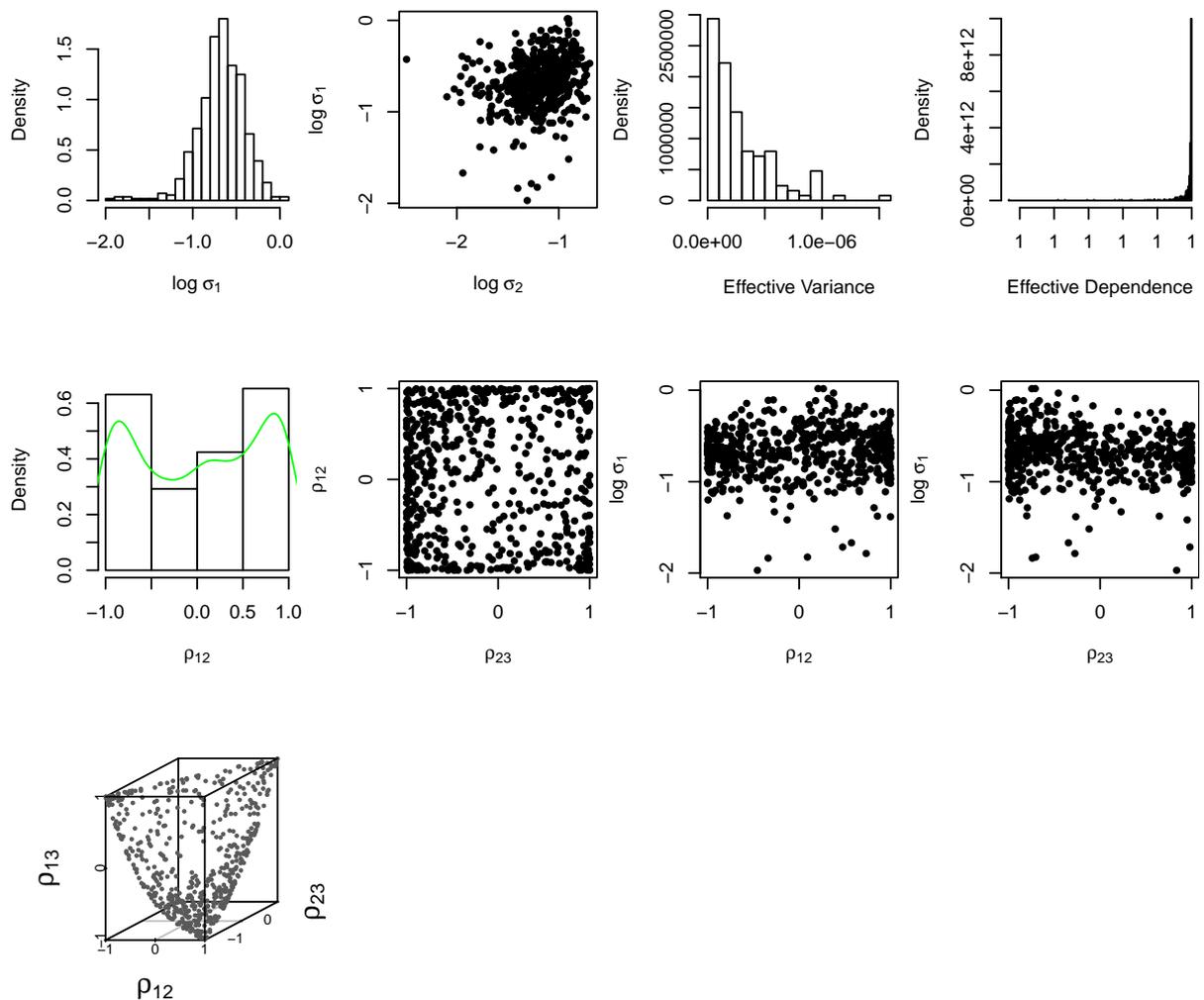
```

1 source("VisMatCovComplex")
2 VisCovMatComplex.selecionada(estatistica, lista.imagens)

```

Um exemplo do resultado obtido após a utilização da nossa ferramenta **VisMatCovComplex** na imagem de Niigata é mostrado na Figura 3.4.

Figura 3.4: Resultado do modelo proposto aplicado na Imagem Niigata.



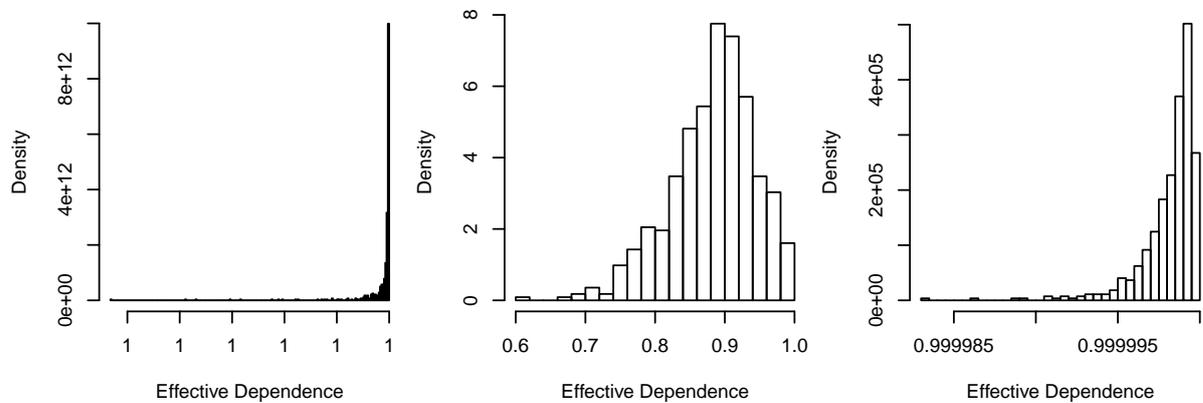
Fonte: autor, 2016.

Para analisar os resultados da nossa metodologia iremos concentrar-se nos gráficos: de dispersão bidimensional entre as variáveis (ρ_{23}, ρ_{12}) exibidos na segunda camada; de dispersão bidimensional entre as variáveis $(\rho_{12}, \log \sigma_1)$ e $(\rho_{23}, \log \sigma_1)$ contidos na terceira e quarta camada, respectivamente. Pois estes são os gráficos que despertam maior atenção

em um primeiro momento e facilitam uma comparação. Os demais gráficos associados aos citados acima revelam resultados mais minuciosos. A análise através desta função é feita observando as forma e valores assumidos pelas estatística.

O resultado da aplicação da função **VisMatCovComplex.selecionada**, por exemplo, na estatística **dependência efetiva** e nas imagens de Niigata, Vale da Morte e na Baía de São Francisco são mostrado na Figura 3.5, respectivamente. Esta função possibilita uma análise mais minuciosa dos gráficos para aprimorar a compreensão dos dados multivariados complexos.

Figura 3.5: Resultado da aplicação da função `VisMatCovComplex.selecionada` nas imagens de Niigata, Vale da Morte, e Baía de São Francisco.



Fonte: autor, 2016.

O presente capítulo tratou dos métodos que foram utilizados para aferir a nova abordagem sobre visualização de matrizes complexas em imagens PolSAR. A seguir, no quarto capítulo, apresentaremos os resultados e análises dos experimentos.

4

RESULTADOS E ANÁLISES

NESTE capítulo, será exposto o roteiro da elaboração do experimento para validar a metodologia apresentada. Os resultados exibidos a seguir foram feitos utilizando a plataforma R.

4.1 Dados PolSAR

As imagens da Baía de São Francisco, Vale da Morte e Niigata vistas nas Figuras 4.12, 4.2 e 4.7 (O RGB dessas imagens foram obtidos através da decomposição Sinclair), respectivamente, foram obtidas em <https://sites.google.com/site/antoniomarcoslarangeiras/dissertacao>. Elas foram escolhidas por terem características bastante distintas, isto é, apresenta regiões geográficas: aquática, desértica e urbana. A Imagem 4.12 e 4.2 foram obtidas pelo sensor AIRSAR que operou em modo polarimétrico considerando as bandas P-(0.45 GHz), L-(1.26 GHz) e C-(5.31 GHz) simultaneamente. Já a imagem 4.7 foi obtida pelo sensor Pi-SAR que operou com frequências duplas: banda L e banda X ambas com funções polarimétricas.

Aplicando nossa função **VisCovMatComplex** nas imagens: Vale da Morte, Niigata e São Francisco obtemos os seguintes resultados, respectivamente:

Vale da Morte (Ver Figuras 4.3, 4.4, 4.5 e 4.6)

- **Primeira camada:** um histograma simétrico do $\log \sigma_1$ com valores contidos no intervalo $(-0.8, 0)$. Um histograma simétrico da ρ_{12} com valores concentrados em torno do zero. Uma associação linear positiva muito forte exibida entre as variáveis $(\rho_{12}, \rho_{23}, \rho_{13})$ no tetraedro elíptico.
- **Segunda camada:** uma associação linear positiva concentrada em torno do intervalo $(-1, -0.4)$ entre as variáveis $(\log \sigma_2, \log \sigma_1)$. Uma associação linear positiva muito forte entre as variáveis (ρ_{23}, ρ_{12}) no gráfico de dispersão bidimensional.

- **Terceira camada:** um histograma simétrico da variância efetiva com valores contidos no intervalo $(0,0.8)$. Um gráfico de dispersão bidimensional bastante concentrado na vertical em torno do zero entre as variáveis $(\rho_{12}, \log \sigma_1)$.
- **Quarta camada:** um histograma com simetria negativa da dependência efetiva com valores em torno de $(-0.2,1)$. Um gráfico de dispersão bidimensional bastante concentrado na vertical em torno do zero entre as variáveis $(\rho_{23}, \log \sigma_1)$.

Niigata (Ver Figuras 4.8, 4.9, 4.10 e 4.11)

- **Primeira camada:** um histograma com simetria negativa do $\log \sigma_1$ com valores contidos no intervalo $(-2.5,0)$. Um histograma bimodal da correlação ρ_{12} . Valores muito dispersos entre as variáveis $(\rho_{12}, \rho_{23}, \rho_{13})$ no tetraedro elíptico.
- **Segunda camada:** uma associação linear positiva concentrada em torno do intervalo $(-2,0)$ das variáveis $(\log \sigma_2, \log \sigma_1)$. Valores muito dispersos entre as variáveis (ρ_{23}, ρ_{12}) no gráfico de dispersão bidimensional.
- **Terceira camada:** um histograma com simetria positiva da variância efetiva com valores contidos no intervalo $(0,1 \times 10^{-5})$. Um gráfico de dispersão bidimensional bastante disperso entre as variáveis $(\rho_{12}, \log \sigma_1)$.
- **Quarta camada:** um histograma com simetria negativa extremamente concentrado no valor 1 da dependência efetiva. Um gráfico de dispersão bidimensional bastante disperso entre as variáveis $(\rho_{23}, \log \sigma_1)$.

São Francisco (Ver Figuras 4.13, 4.14, 4.15 e 4.16)

- **Primeira camada:** um histograma simétrico do $\log \sigma_1$ com valores contidos no intervalo $(-1.8, -0.2)$. Um histograma simétrico da correlação ρ_{12} . Uma associação linear positiva forte exibida entre as variáveis $(\rho_{12}, \rho_{23}, \rho_{13})$ no tetraedro elíptico.
- **Segunda camada:** uma associação linear positiva forte das variáveis $(\log \sigma_2, \log \sigma_1)$ em torno do intervalo $(-2, -1)$. Uma associação linear positiva forte entre as variáveis (ρ_{23}, ρ_{12}) no gráfico de dispersão bidimensional.
- **Terceira camada:** um histograma com simetria positiva da variância efetiva com valores contidos no intervalo $(0,3 \times 10^{-2})$. Um gráfico de dispersão bidimensional concentrado em torno do zero entre as variáveis $(\rho_{12}, \log \sigma_1)$.
- **Quarta camada:** um histograma com simetria negativa tendendo a 1 da dependência efetiva. Um gráfico de dispersão bidimensional concentrado em torno do zero entre as variáveis $(\rho_{23}, \log \sigma_1)$.

Para analisar os resultados iremos seguir o método proposto na subseção 3.1.1, isto é, focar nos gráficos que facilitam a comparação. Os gráficos que despertam maior atenção em um primeiro momento são: de dispersão bidimensional entre as variáveis (ρ_{23}, ρ_{12}) exibidos na segunda camada; os gráficos de dispersão bidimensional entre as variáveis $(\rho_{12}, \log \sigma_1)$ e $(\rho_{23}, \log \sigma_1)$ contidos na terceira e quarta camada, respectivamente. Isso pode ser analisado - de forma específica - através da função **VisMatCovComplex.selecionada** aplicada nas imagens de Niigata, Vale da Morte e Baía de São Francisco, nessa ordem (Ver Figuras 4.2(a), 4.2(b) e 4.2(c)).

Os demais gráficos associados aos gráficos que facilitam uma comparação imediata, citados acima, revelam resultados mais minuciosos. Por exemplo, se compararmos as imagens do Vale da Morte e da Baía de São Francisco poderíamos ser induzidos ao erro de pensar que os dados vêm de uma mesma classe. Por outro lado, se agregarmos os outros gráficos em nossa análise teremos convicção que são de classes diferentes, ou seja, se observarmos o histograma de dependência efetiva exibido na quarta camada iremos perceber que na imagem da Baía de São Francisco os valores se aproximam, mas não atingem o valor 1 (Ver Figuras 4.13, 4.14, 4.15 e 4.16).

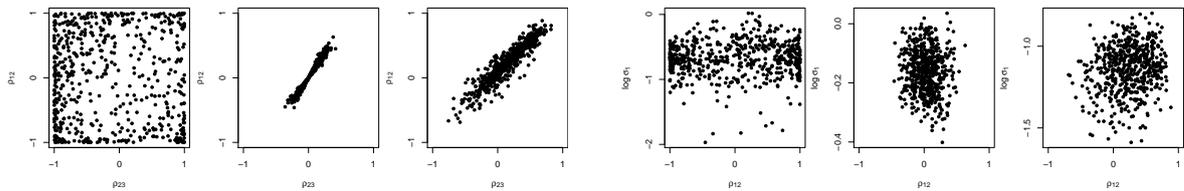
O histograma da correlação ρ_{12} contido na primeira camada não é tão concentrado em torno do zero quanto o da imagem do Vale da Morte. Além disso, os valores dos logaritmos dos desvios padrão contidos no histograma do $\log \sigma_1$ e no gráfico de dispersão bidimensional entre os $(\log \sigma_2, \log \sigma_1)$ exibidos na primeira e segunda camada, respectivamente, são menores que os da imagem do Vale da morte (Comparar as Figuras 4.3, 4.4, 4.5 e 4.6 com as Figuras 4.13, 4.14, 4.15 e 4.16).

Se tivermos interessados em comparar a imagem de Niigata (Ver Figuras 4.8, 4.9, 4.10 e 4.11) com a imagem do Vale da Morte (Ver Figuras 4.3, 4.4, 4.5 e 4.6) iremos adotar o mesmo raciocínio adotado na situação descrita acima. Observamos os gráficos de dispersão bidimensional entre as variáveis (ρ_{23}, ρ_{12}) exibidos na segunda camada; os gráficos de dispersão bidimensional entre as variáveis $(\rho_{12}, \log \sigma_1)$ e $(\rho_{23}, \log \sigma_1)$ contidos na terceira e quarta camada, respectivamente. Com isso percebemos que esses gráficos são bem dispersos na imagem de Niigata.

Observado o histograma de dependência efetiva na imagem do Vale da Morte verificamos que os valores são muito pequenos com relação as outras imagens analisadas. Também é possível notar os valores dos logaritmos dos desvios padrão contidos no histograma do $\log \sigma_1$ são muito pequenos. Além disso, no gráfico de dispersão bidimensional entre as variáveis $(\log \sigma_2, \log \sigma_1)$ exibido segunda camada são menores que os da imagem do Vale da morte. E por fim, o histograma de correlação ρ_{12} é bimodal na imagem de Niigata diferentemente da imagem do Vale da Morte e de São Francisco.

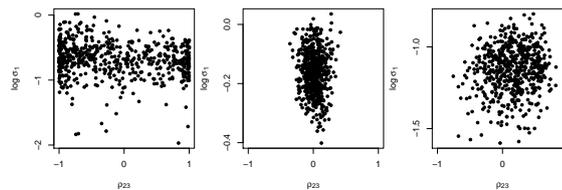
As tabelas 4.1, 4.2 e 4.3 apresentam informações quantitativas de resumo sobre o conjunto de dados que escolhemos como amostra em cada imagem. Através dessas informações

Figura 4.1: Visualização mais específica através da função **VisMatCovComplex.selecionada**.



(a) Visualização, mais específica, da forma do gráfico de dispersão bidimensional entre as variáveis (ρ_{23}, ρ_{12}) nas imagens de Niigata, Vale da Morte e Baía de São Francisco, nessa ordem.

(b) Visualização, mais específica, da forma do gráfico de dispersão bidimensional entre as variáveis ($\rho_{12}, \log \sigma_1$) nas imagens de Niigata, Vale da Morte e Baía de São Francisco, nessa ordem.



(c) Visualização, mais específica, da forma do gráfico de dispersão bidimensional entre as variáveis ($\rho_{23}, \log \sigma_1$) nas imagens de Niigata, Vale da Morte e Baía de São Francisco, nessa ordem.

Fonte: Autor, 2016.

quantitativas podemos perceber que:

- a região desértica tem os maiores valores para o $\log \sigma_1$.

- a região urbana tem os menores valores para V_e .
- na região urbana os valores da dependência efetiva (D_e) são os maiores e concentrados no número 1. Ao contrário da região desértica que possuem os menores valores de dependência efetiva.

Figura 4.2: AIRSAR(NASA/JPL), Vale da Morte(CA) US, PolSAR.

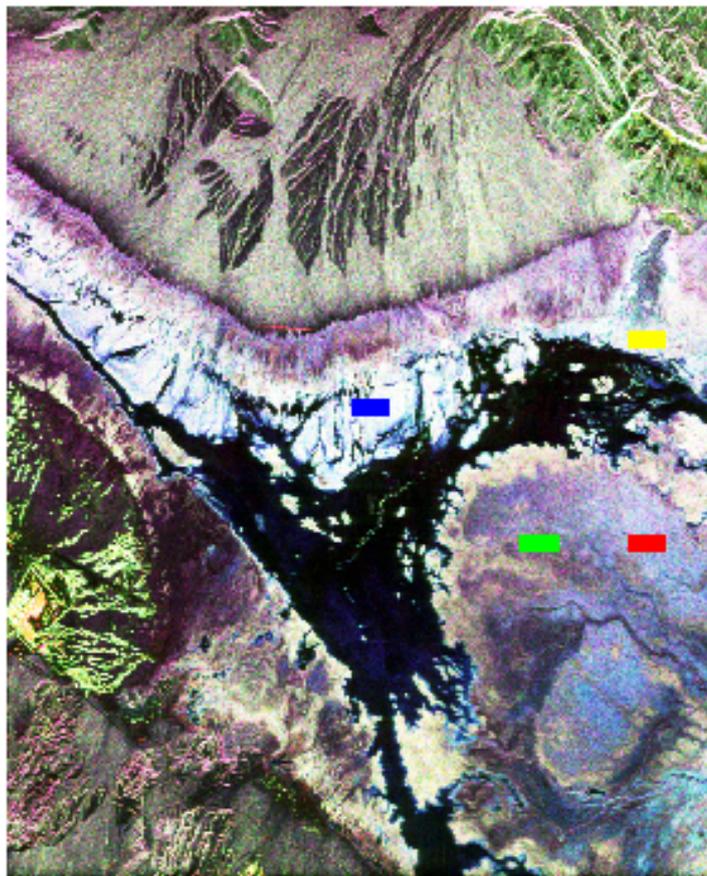
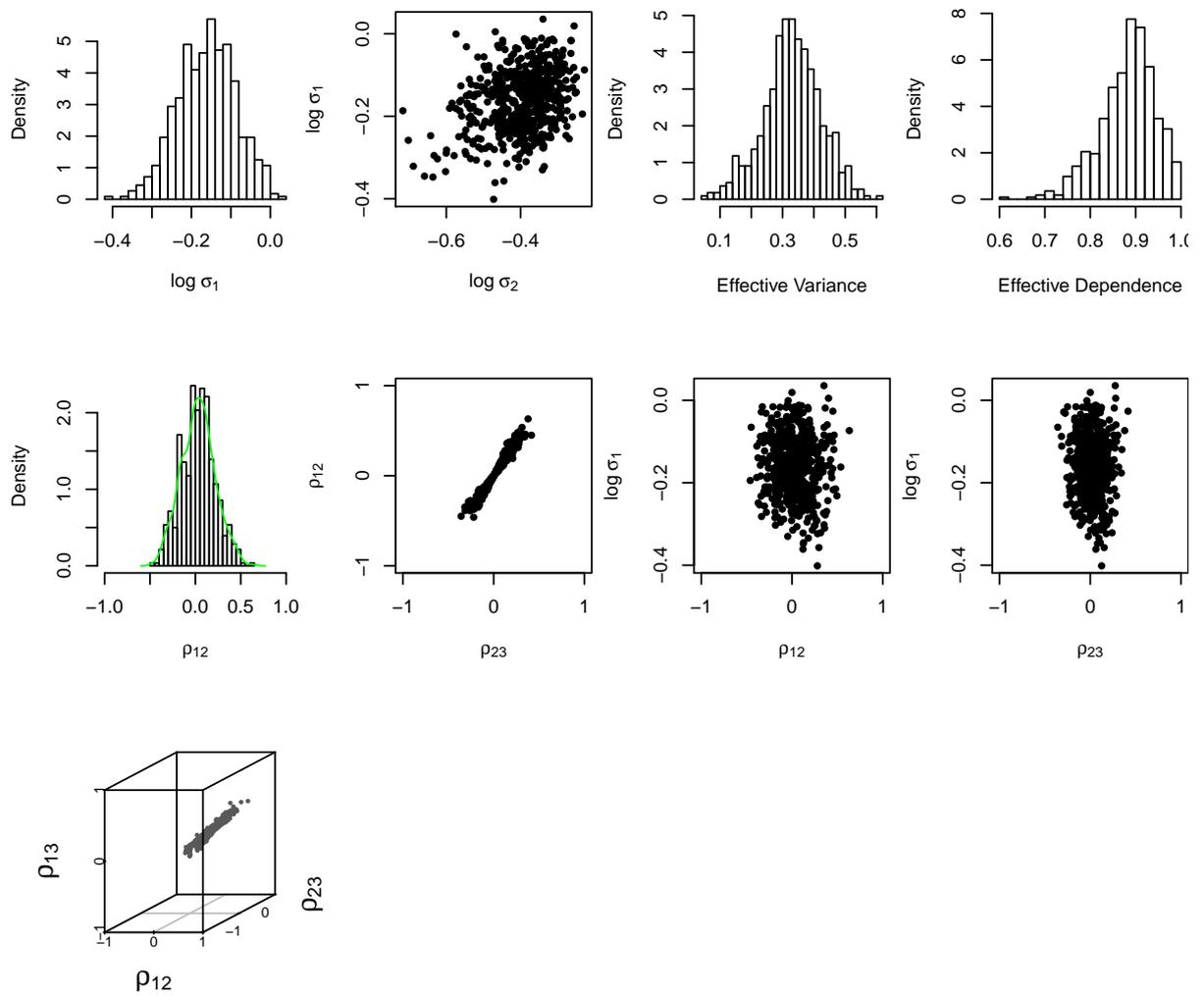
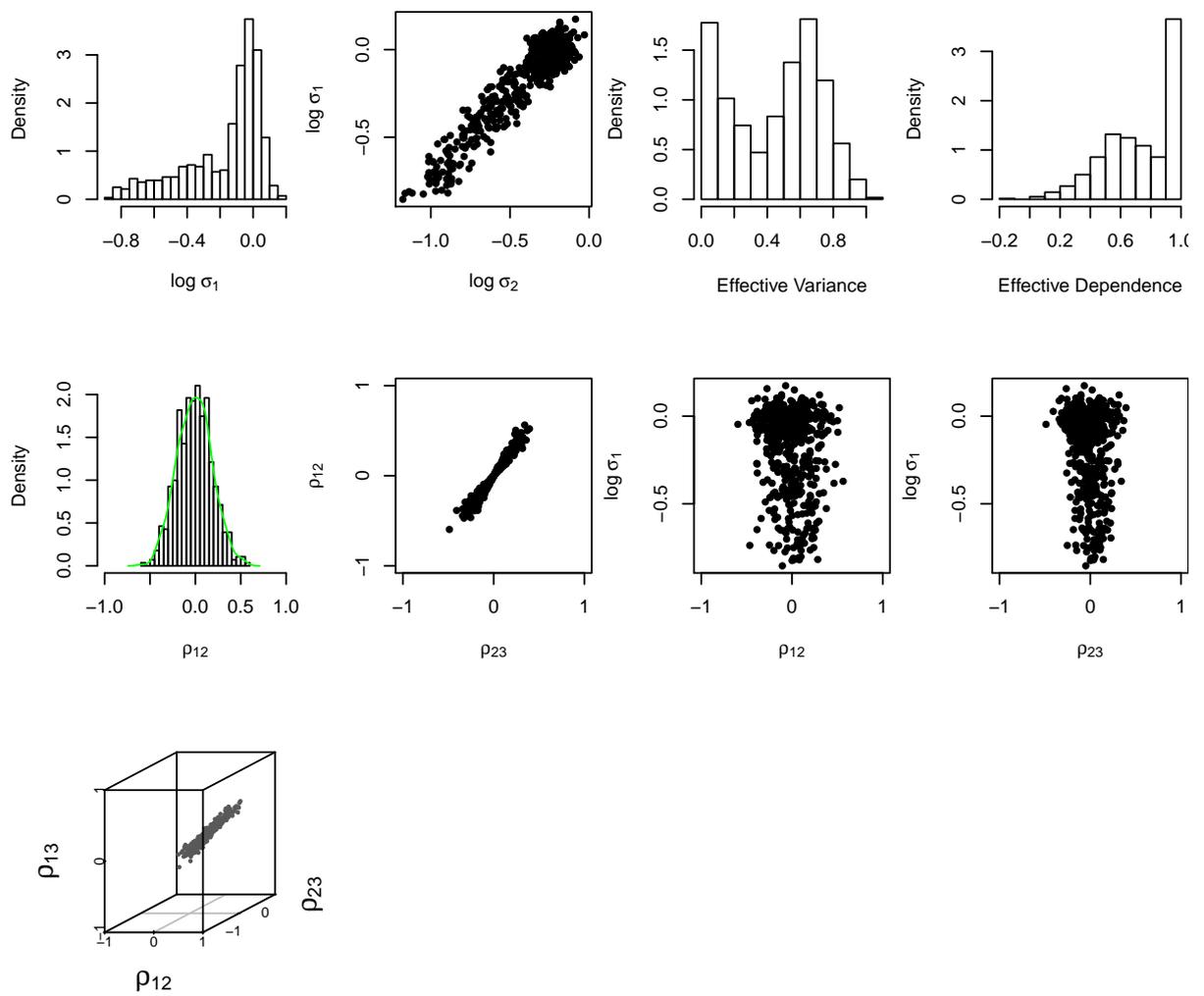


Figura 4.3: Visualização de Matrizes de Covariância Complexa da Imagem de Vale da Morte região Amarela.



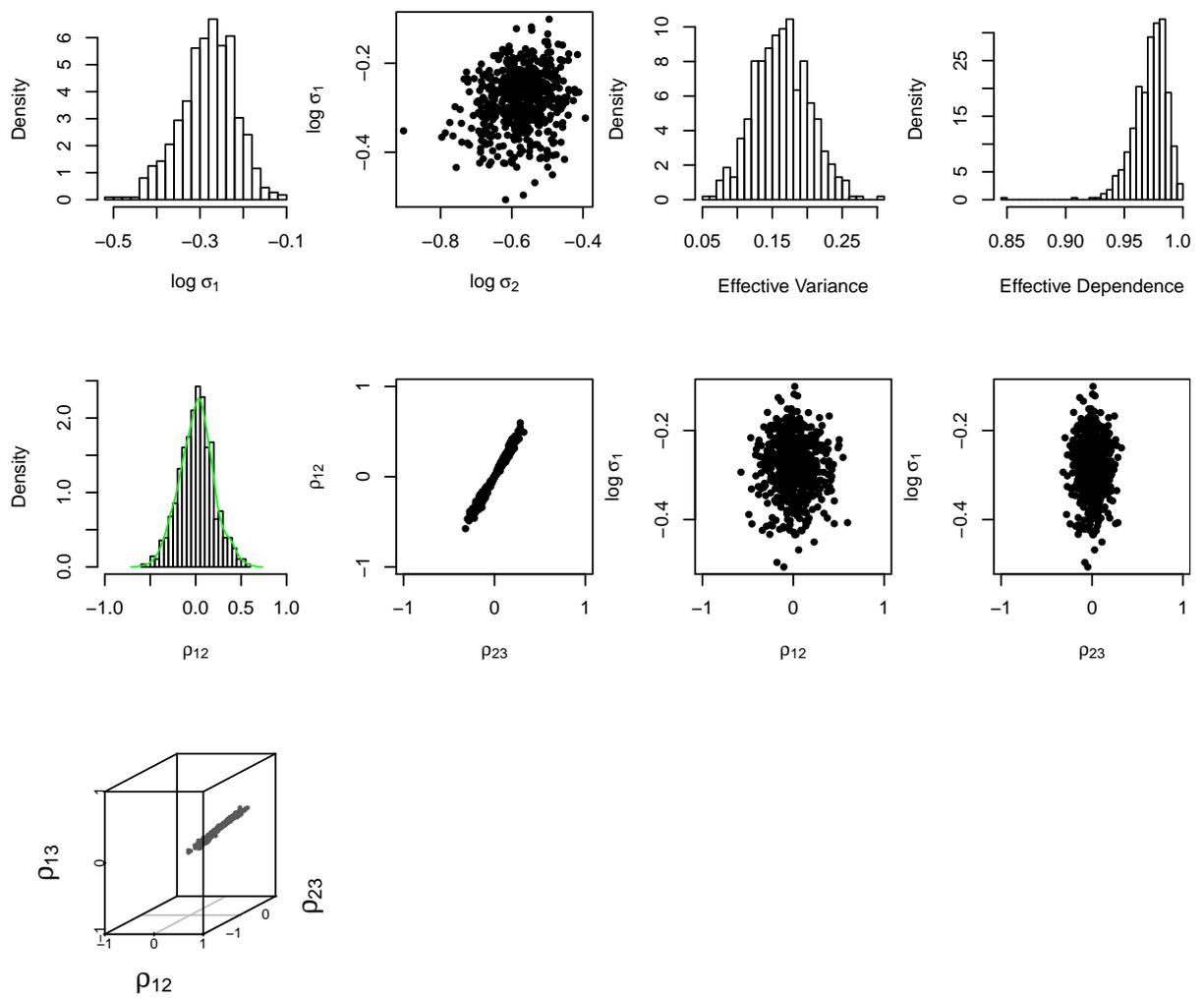
Fonte: autor, 2016.

Figura 4.4: Visualização de Matrizes de Covariância Complexa da Imagem de Vale da Morte região Azul.



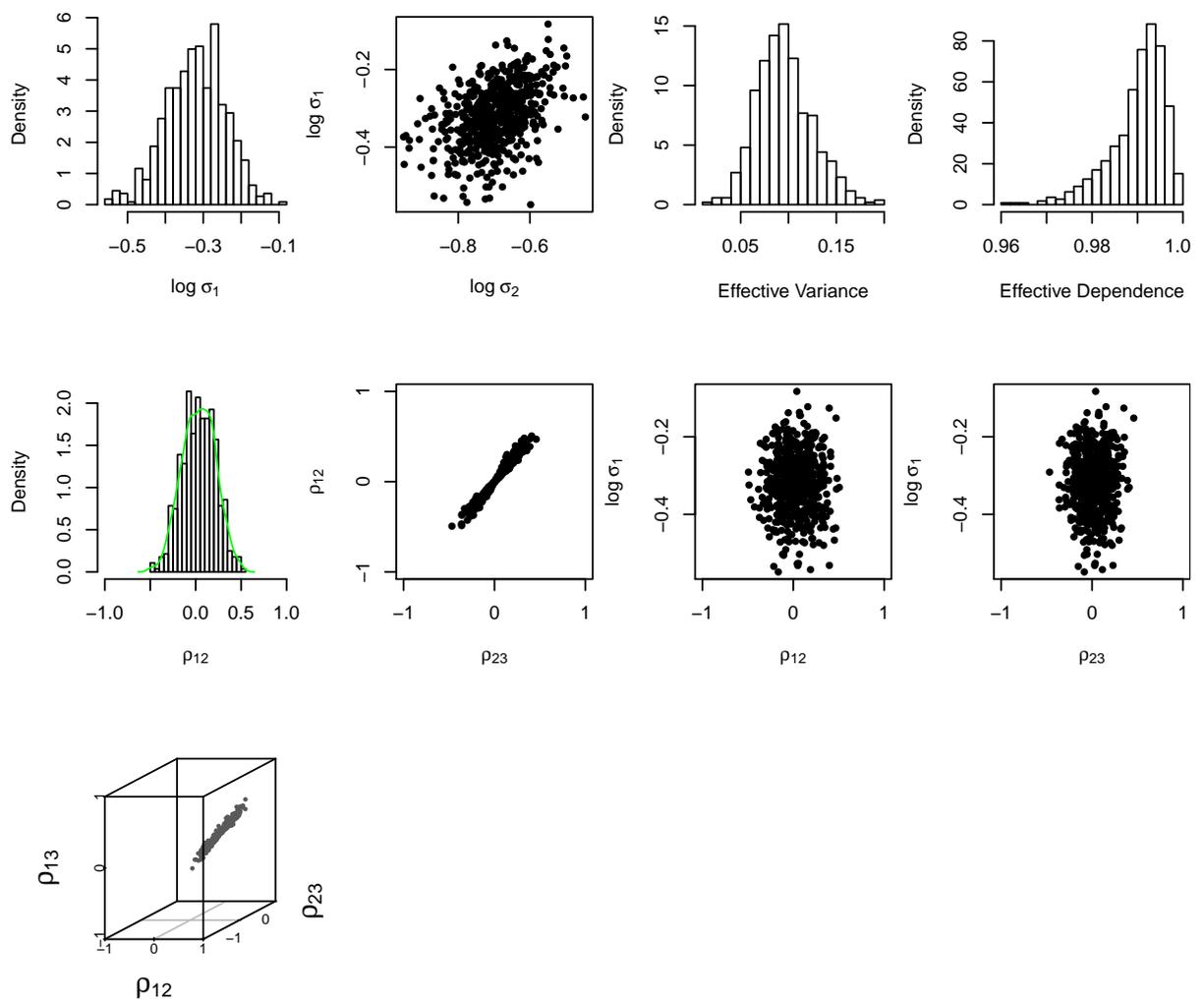
Fonte: autor, 2016.

Figura 4.5: Visualização de Matrizes de Covariância Complexa da Imagem de Vale da Morte região Vermelha.



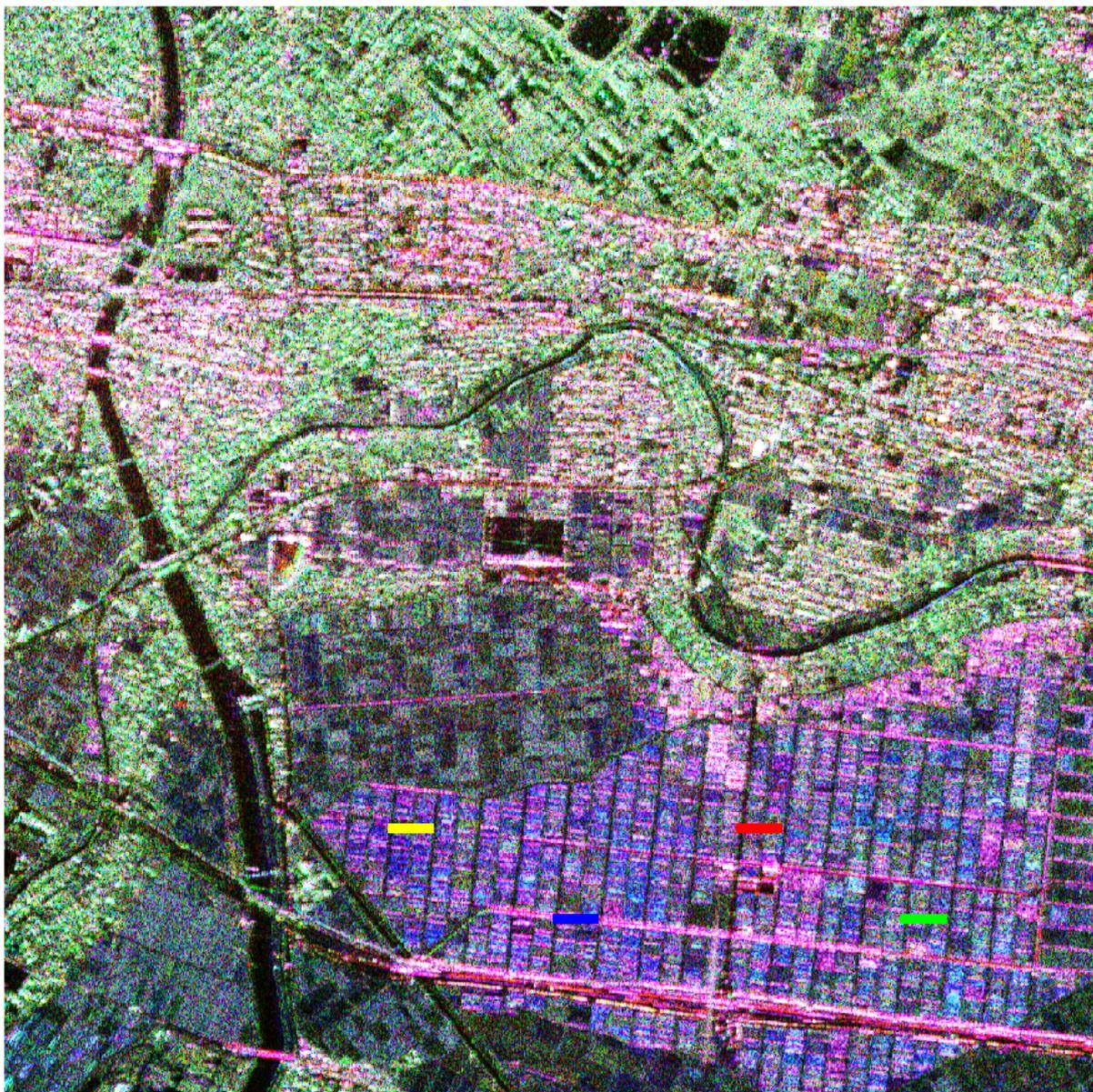
Fonte: autor, 2016.

Figura 4.6: Visualização de Matrizes de Covariância Complexa da Imagem de Vale da Morte região Verde.



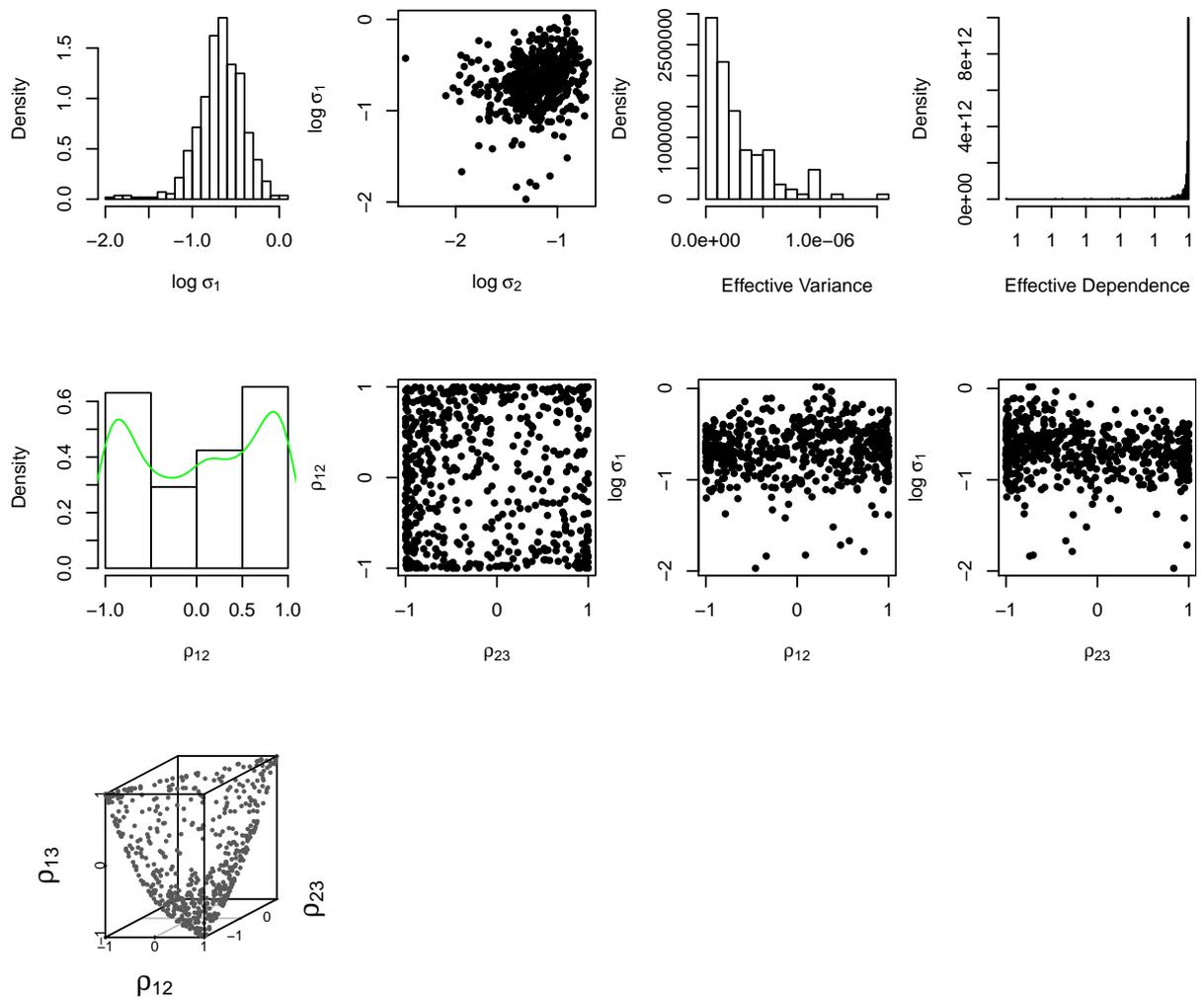
Fonte: autor, 2016.

Figura 4.7: PISAR(NICT/JAXA), Niigata JP, PolSAR.



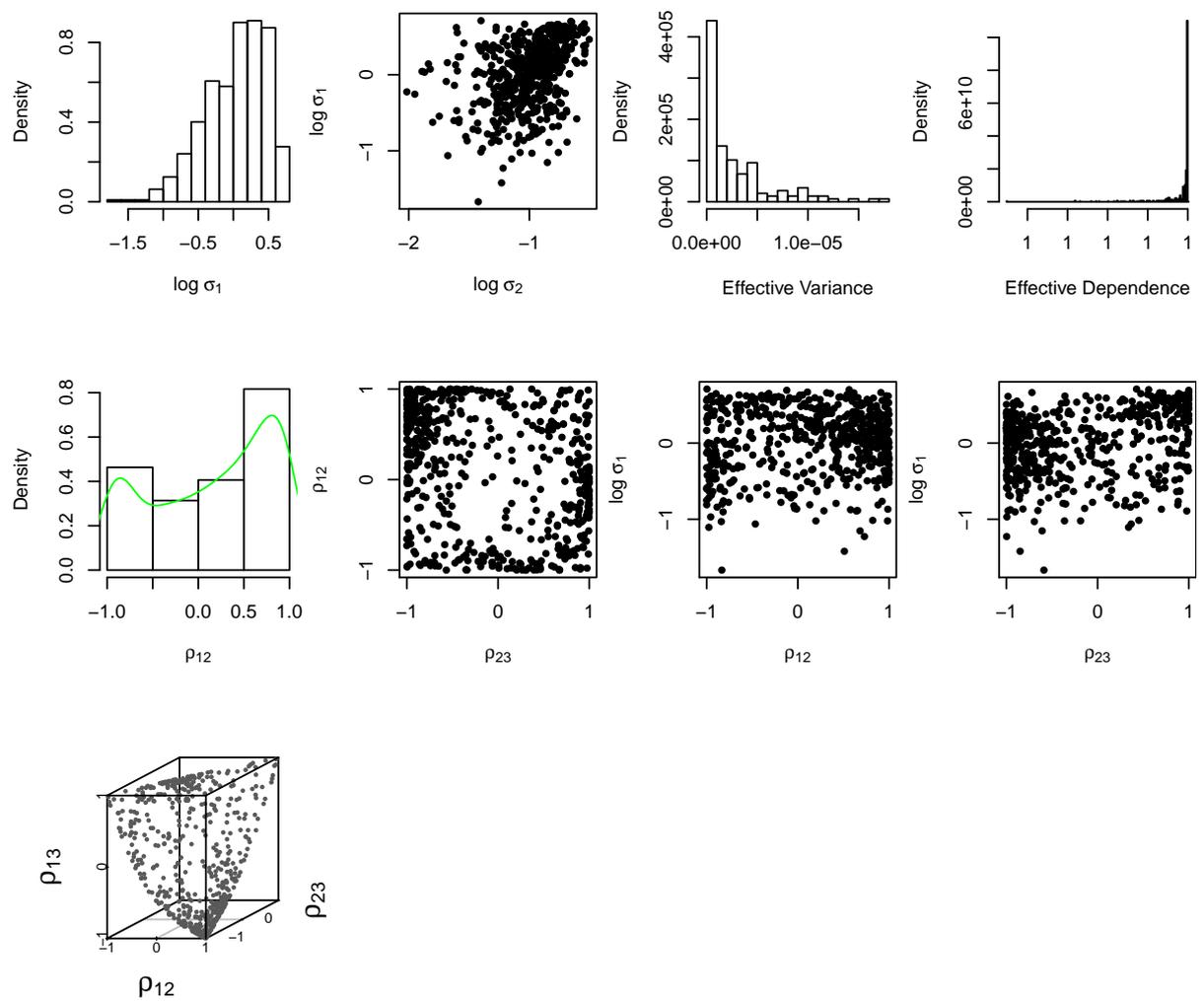
Fonte: processada pelo autor, 2016.

Figura 4.8: Visualização de Matrizes de Covariância Complexa da Imagem de Niigata região Amarela.



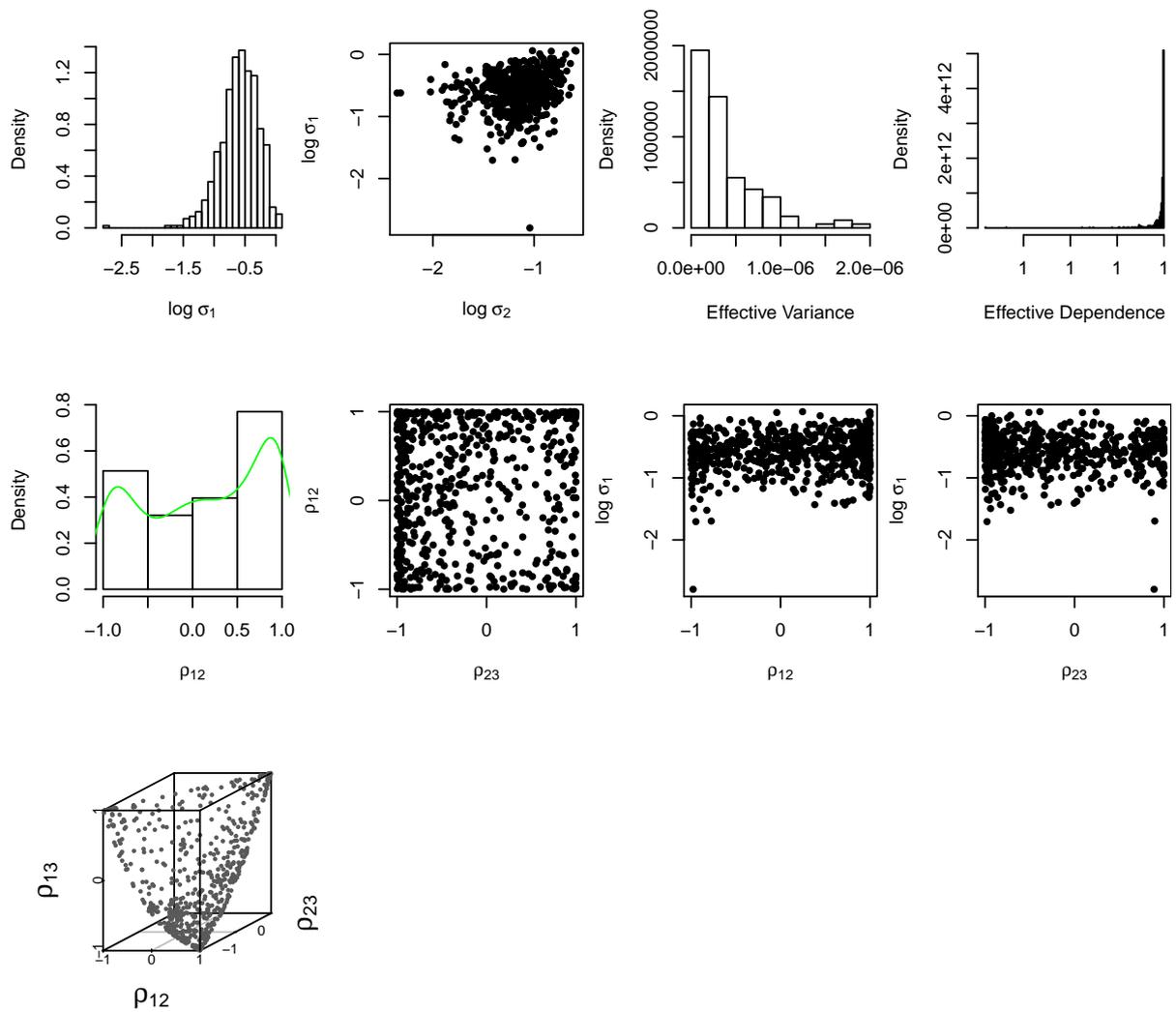
Fonte: autor, 2016.

Figura 4.9: Visualização de Matrizes de Covariância Complexa da Imagem de Niigata região Azul.



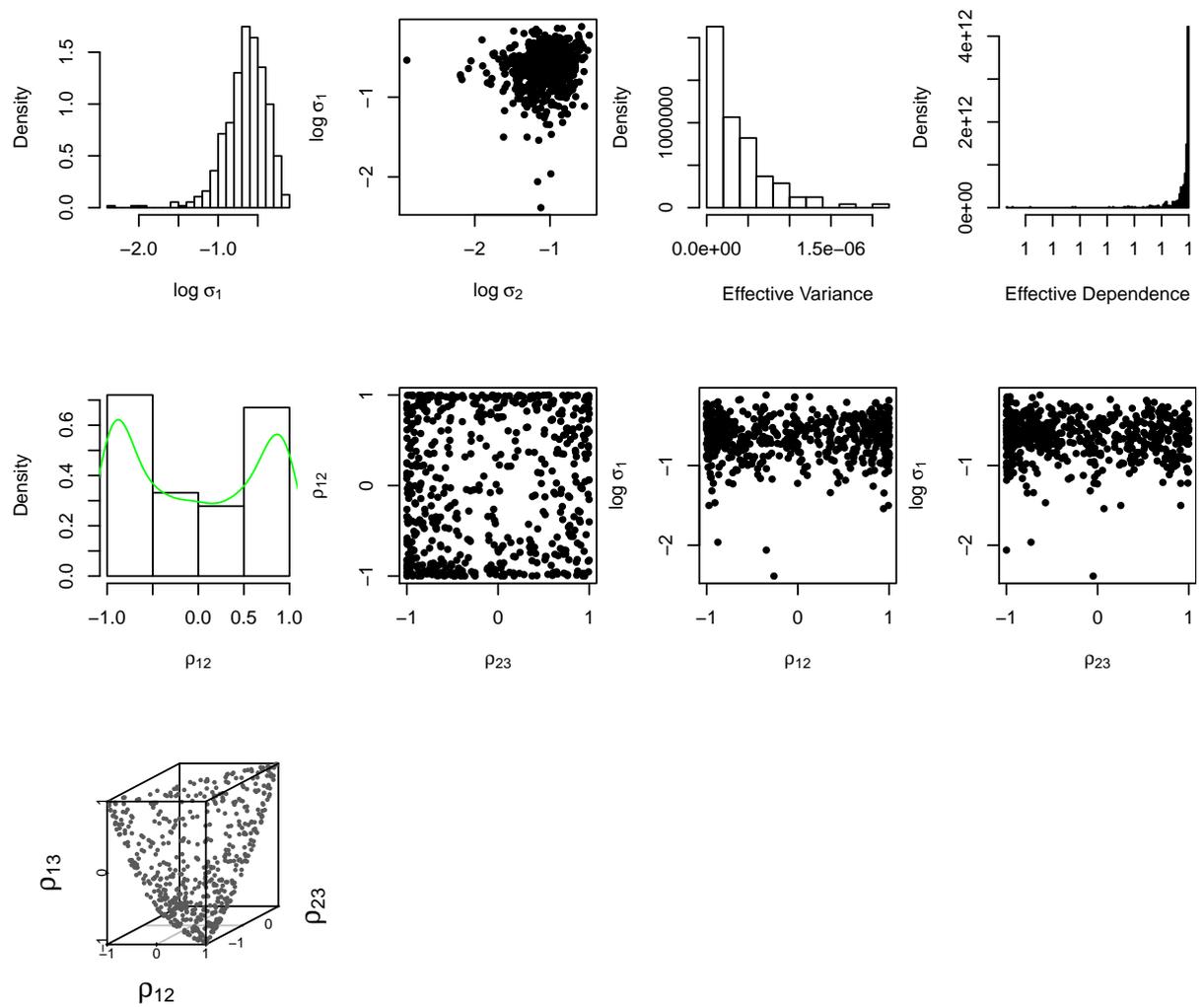
Fonte: autor, 2016.

Figura 4.10: Visualização de Matrizes de Covariância Complexa da Imagem de Niigata região Vermelha.



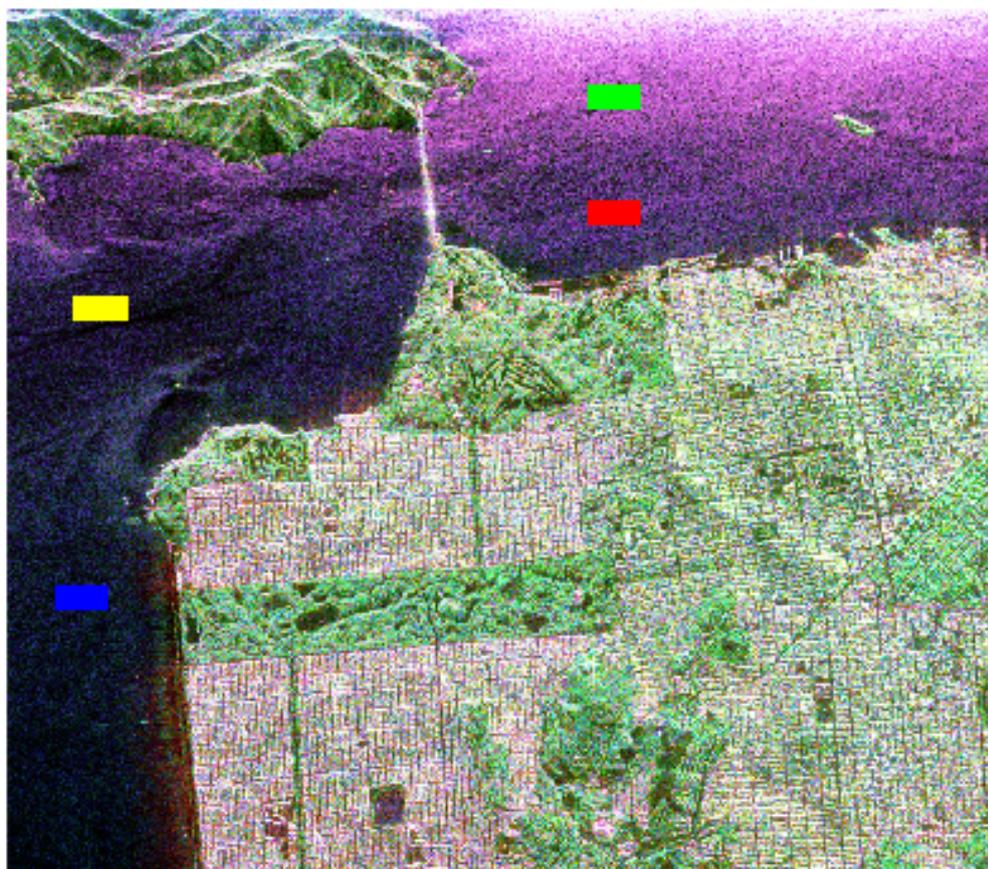
Fonte: autor, 2016.

Figura 4.11: Visualização de Matrizes de Covariância Complexa da Imagem de Niigata região Verde.



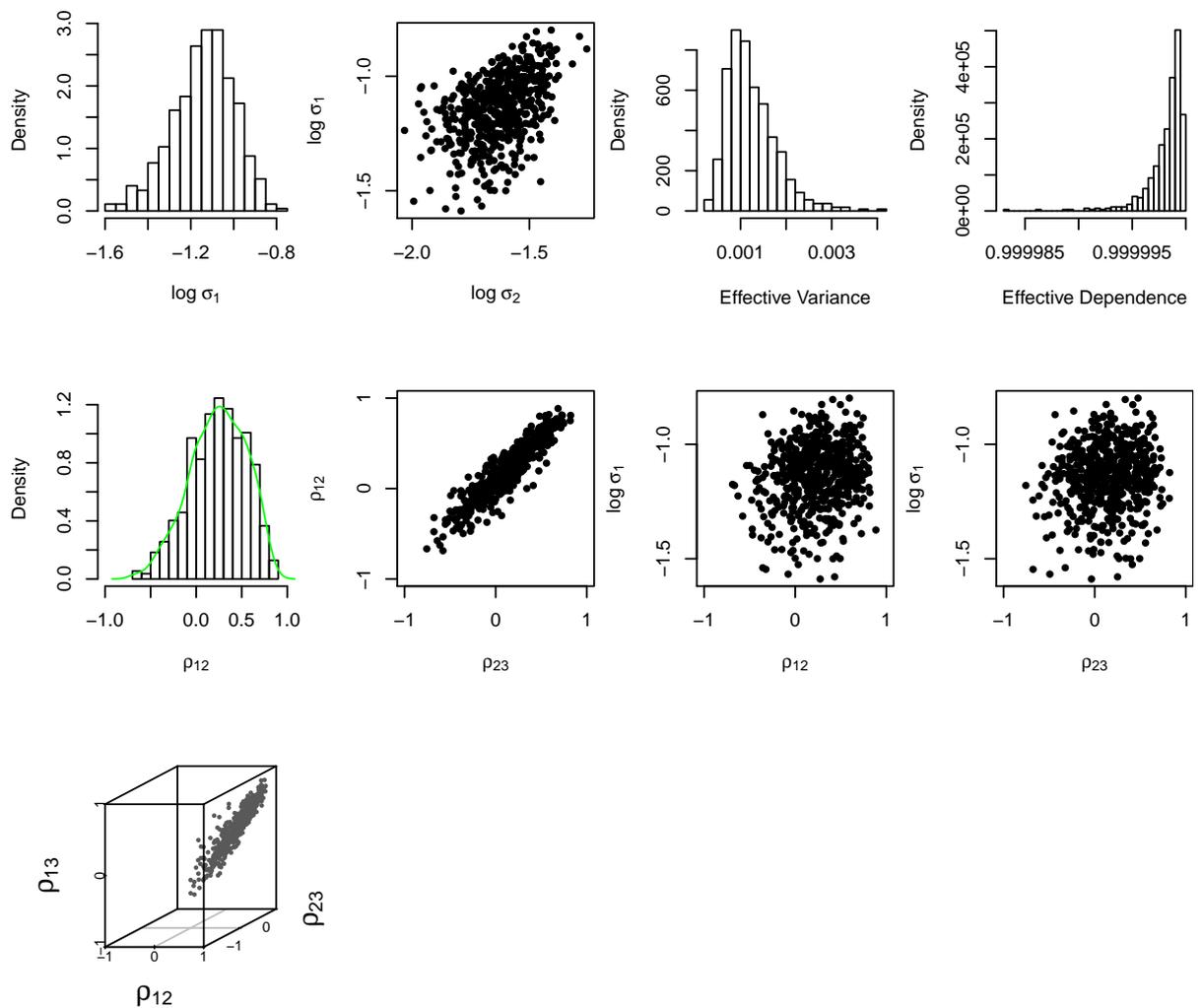
Fonte: autor, 2016.

Figura 4.12: AIRSAR(NASA/JPL), Baía de São Francisco(CA) US, PolSAR.



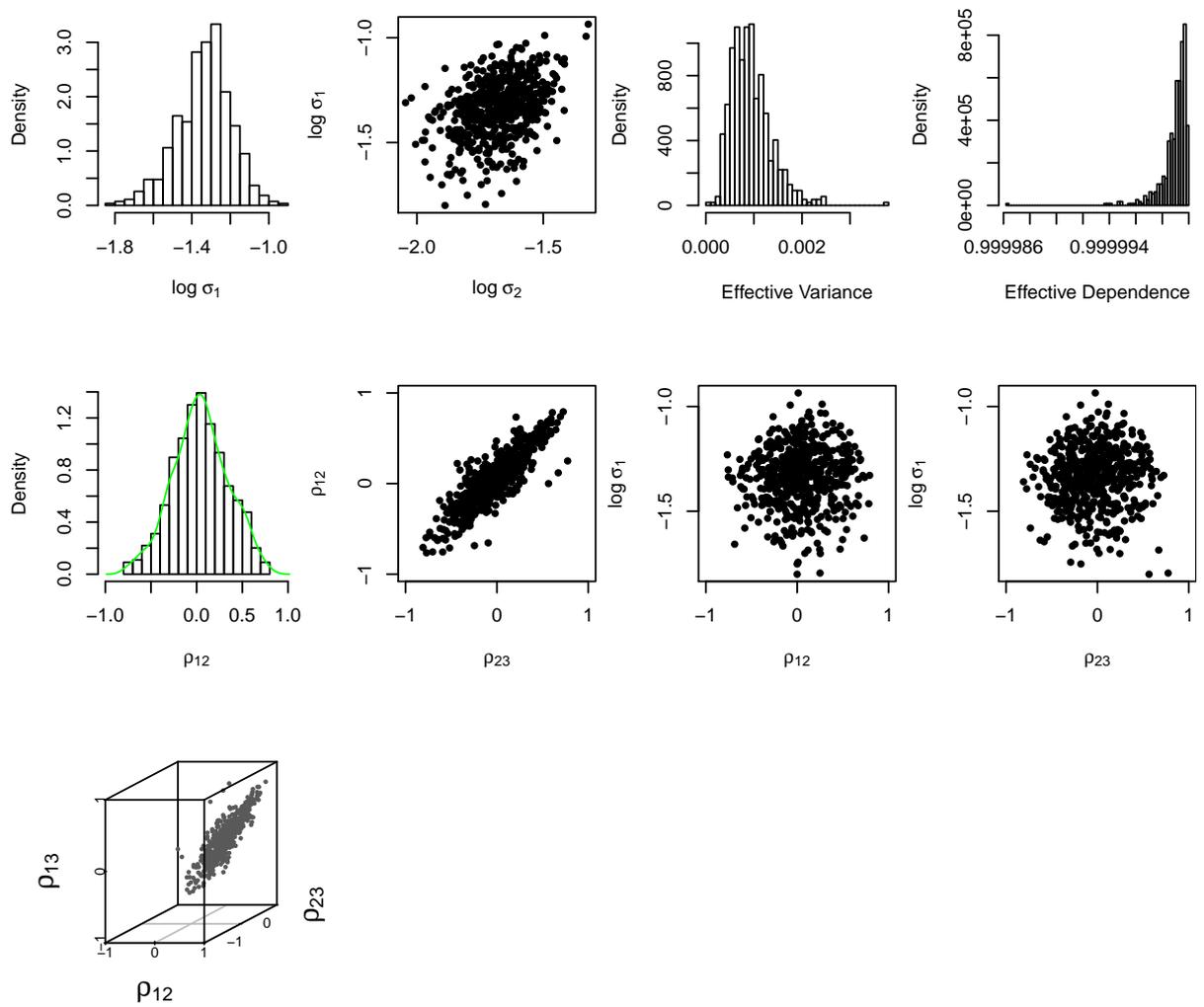
Fonte: processada pelo autor, 2016.

Figura 4.13: Visualização de Matrizes de Covariância Complexa da Imagem de São Francisco região Amarela.



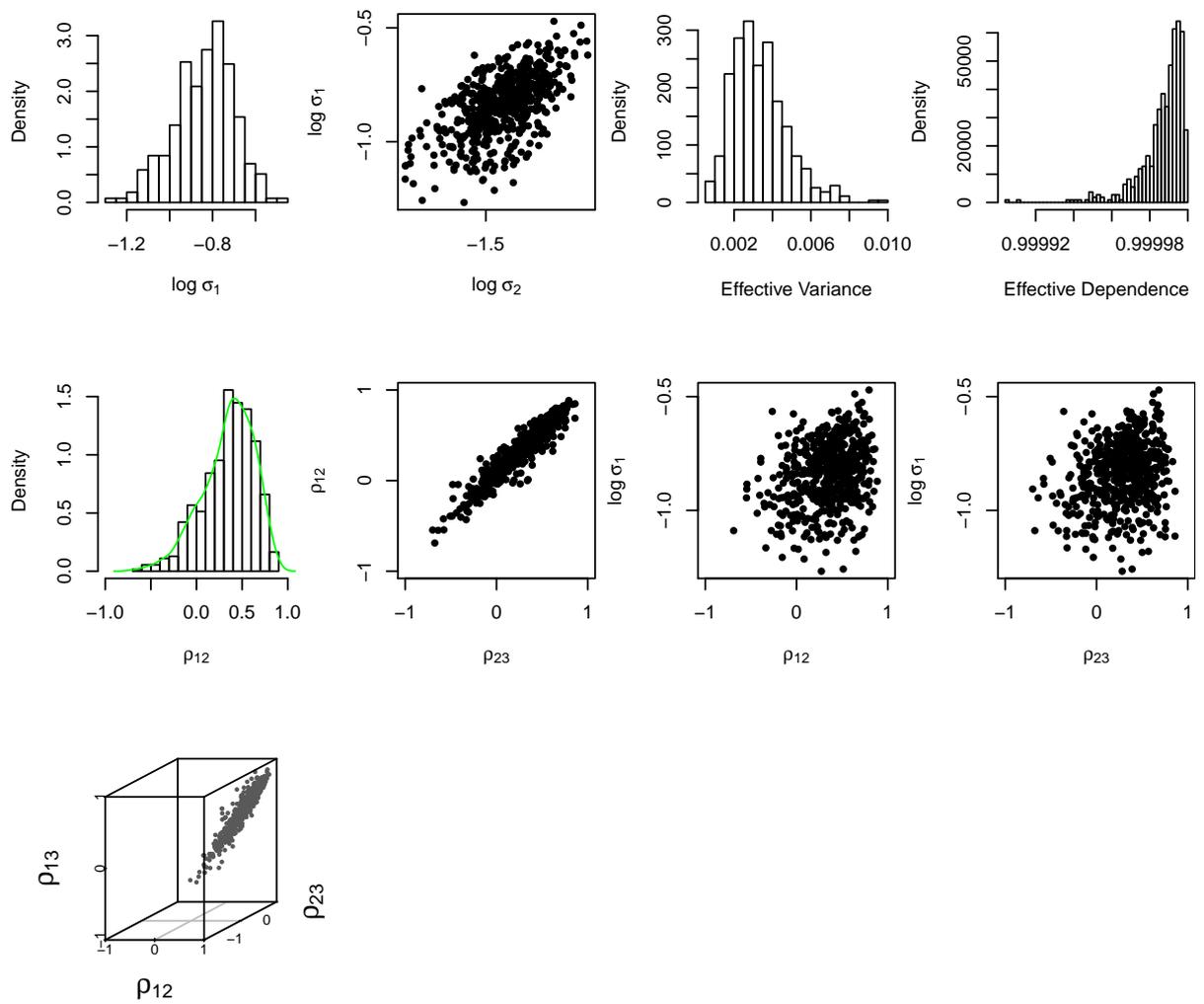
Fonte: autor, 2016.

Figura 4.14: Visualização de Matrizes de Covariância Complexa da Imagem de São Francisco região Azul.



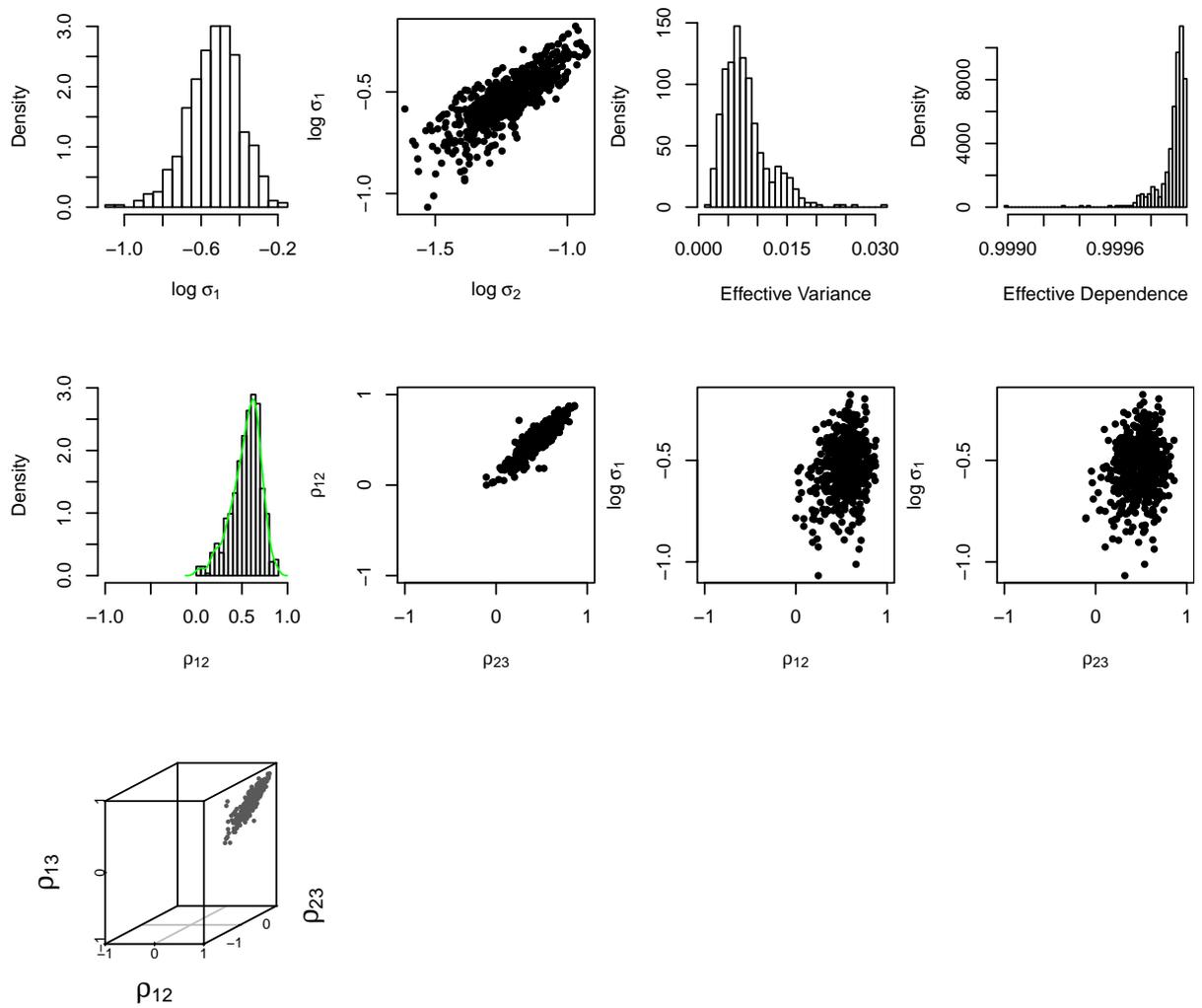
Fonte: autor, 2016.

Figura 4.15: Visualização de Matrizes de Covariância Complexa da Imagem de São Francisco região Vermelha.



Fonte: autor, 2016.

Figura 4.16: Visualização de Matrizes de Covariância Complexa da Imagem de São Francisco região Verde.



Fonte: autor, 2016.

Tabela 4.1: Resumos estatístico das amostras retiradas da imagem de Niigata.

Summaries														
Niigata N=500														
Amarela								Azul						
	$\log\sigma_1$	$\log\sigma_2$	V_e	D_e	ρ_{12}	ρ_{13}	ρ_{23}	$\log\sigma_1$	$\log\sigma_2$	V_e	D_e	ρ_{12}	ρ_{13}	ρ_{23}
Min	-1.96	-2.49	0.00	1.00	-0.99	-0.10	-0.99	-1.66	-2.01	0.00	1.00	-0.99	-0.48	-0.99
1st Q	-0.81	-1.34	8.08×10^{-8}	1.00	-0.65	-0.01	-0.81	-0.25	-1.14	2.55×10^{-7}	1.00	-0.43	-0.04	-0.84
Median	-0.67	-1.15	1.89×10^{-7}	1.00	0.06	0.00	-0.28	0.10	-0.94	1.27×10^{-6}	1.00	0.28	-8.00×10^{-3}	-0.35
3rd Q	-0.49	-1.00	4.04×10^{-7}	1.00	0.68	0.00	0.52	0.37	-0.81	4.12×10^{-6}	1.00	0.75	0.02	0.64
Max	0.01	-0.69	1.57×10^{-6}	1.00	0.99	0.03	0.99	0.70	-0.50	1.77×10^{-5}	1.00	0.99	0.47	0.99
Vermelha								Verde						
	$\log\sigma_1$	$\log\sigma_2$	V_e	D_e	ρ_{12}	ρ_{13}	ρ_{23}	$\log\sigma_1$	$\log\sigma_2$	V_e	D_e	ρ_{12}	ρ_{13}	ρ_{23}
Min	-2.79	-2.35	0.00	1.00	-0.99	-0.11	-0.99	-2.38	-2.90	0.00	1.00	-0.99	-0.14	-0.99
1st Q	-0.76	-1.26	9.46×10^{-8}	1.00	-0.54	-0.01	-0.86	-0.78	-1.22	1.19×10^{-7}	1.00	-0.79	-0.01	-0.78
Median	-0.56	-1.12	2.66×10^{-7}	1.00	0.18	-5.33×10^{-3}	-0.41	-0.61	-1.04	2.79×10^{-7}	1.00	-0.04	-3.33×10^{-3}	-0.23
3rd Q	-0.36	-0.97	4.76×10^{-7}	1.00	0.78	-2.73×10^{-3}	0.31	-0.46	-0.87	5.30×10^{-7}	1.00	0.73	7.89×10^{-3}	0.51
Max	0.06	-0.58	1.82×10^{-6}	1.00	0.99	0.05	0.99	-0.11	-0.47	2.03×10^{-6}	1.00	0.99	0.09	0.99

Tabela 4.2: Resumos estatístico das amostras retiradas da imagem de Vale da Morte.

Summaries														
Vale da Morte														
N=500														
	Amarela							Azul						
	$\log\sigma_1$	$\log\sigma_2$	V_e	D_e	ρ_{12}	ρ_{13}	ρ_{23}	$\log\sigma_1$	$\log\sigma_2$	V_e	D_e	ρ_{12}	ρ_{13}	ρ_{23}
Min	-0.40	-0.71	0.04	0.61	-0.46	-0.16	-0.35	-0.85	-1.17	0.01	-0.13	-0.59	-0.33	-0.48
1st Q	-0.21	-0.44	0.27	0.84	-0.10	-0.02	-0.06	-0.30	-0.54	0.16	0.54	-0.14	-0.05	-0.10
Median	-0.15	-0.38	0.33	0.89	0.03	0.00	0.02	-0.07	-0.29	0.50	0.74	-0.01	-9.60×10^{-4}	-8.30×10^{-3}
3rd Q	-0.11	-0.34	0.38	0.92	0.14	0.04	0.09	-3.46×10^{-3}	-0.21	0.67	0.97	0.12	0.03	0.09
Max	0.03	-0.22	0.61	0.99	0.63	0.22	0.41	0.17	-0.02	1.06	0.99	0.56	0.38	0.39
	Vermelha							Verde						
	$\log\sigma_1$	$\log\sigma_2$	V_e	D_e	ρ_{12}	ρ_{13}	ρ_{23}	$\log\sigma_1$	$\log\sigma_2$	V_e	D_e	ρ_{12}	ρ_{13}	ρ_{23}
Min	-0.50	-0.90	0.05	0.84	-0.57	-0.09	-0.31	-0.54	-0.95	0.01	0.96	-0.49	-0.05	-0.46
1st Q	-0.31	-0.61	0.13	0.96	-0.11	-0.01	-0.06	-0.37	-0.75	0.07	0.98	-0.09	-8.64×10^{-3}	-0.06
Median	-0.27	-0.56	0.16	0.97	0.01	1.88×10^{-3}	8.88×10^{-3}	-0.31	-0.70	0.09	0.99	0.03	3.65×10^{-3}	0.02
3rd Q	-0.23	-0.52	0.19	0.98	0.12	0.01	0.07	-0.26	-0.64	0.11	0.99	0.16	0.01	0.12
Max	-0.10	-0.39	0.30	0.99	0.59	0.09	0.32	-0.08	-0.44	0.19	0.99	0.50	0.09	0.45

Tabela 4.3: Resumos estatístico das amostras retiradas da imagem de São Francisco.

Summaries														
San Francisc N=500														
	Amarela							Azul						
	$\log\sigma_1$	$\log\sigma_2$	V_e	D_e	ρ_{12}	ρ_{13}	ρ_{23}	$\log\sigma_1$	$\log\sigma_2$	V_e	D_e	ρ_{12}	ρ_{13}	ρ_{23}
Min	-1.58	-2.03	3.35×10^{-4}	0.99	-0.68	-2.94×10^{-4}	-0.75	-1.80	-2.04	6.14×10^{-5}	0.99	-0.76	-3.61×10^{-3}	-0.80
1st Q	-1.23	-1.71	8.58×10^{-4}	0.99	0.02	-1.32×10^{-4}	-0.04	-1.42	-1.76	6.36×10^{-4}	0.99	-0.14	-6.09×10^{-4}	-0.26
Median	-1.12	-1.62	1.12×10^{-3}	0.99	0.25	-4.51×10^{-4}	0.16	-1.32	-1.67	8.79×10^{-4}	0.99	0.04	-1.00×10^{-4}	-0.04
3rd Q	-1.04	-1.54	1.53×10^{-3}	0.99	0.47	-1.38×10^{-3}	0.39	-1.24	-1.59	1.16×10^{-3}	0.99	0.24	4.05×10^{-4}	0.17
Max	-0.79	-1.25	4.10×10^{-3}	0.99	0.88	-6.33×10^{-3}	0.82	-0.93	-1.32	3.71×10^{-3}	1.00	0.79	3.12×10^{-3}	0.77
	Vermelha							Verde						
	$\log\sigma_1$	$\log\sigma_2$	V_e	D_e	ρ_{12}	ρ_{13}	ρ_{23}	$\log\sigma_1$	$\log\sigma_2$	V_e	D_e	ρ_{12}	ρ_{13}	ρ_{23}
Min	-1.26	6.31×10^{-4}	6.31×10^{-4}	0.99	-0.68	-6.91×10^{-3}	-0.69	-1.06	-1.61	1.24×10^{-3}	0.99	0.00	-9.73×10^{-4}	-0.10
1st Q	-0.93	2.27×10^{-3}	2.27×10^{-3}	0.99	0.18	6.25×10^{-4}	0.10	-0.61	-1.30	5.23×10^{-3}	0.99	0.45	6.21×10^{-3}	0.38
Median	-0.82	3.07×10^{-3}	3.07×10^{-3}	0.99	0.38	2.36×10^{-3}	0.31	-0.52	-1.22	7.07×10^{-3}	0.99	0.57	0.01	0.51
3rd Q	-0.74	4.06×10^{-3}	4.06×10^{-3}	0.99	0.56	4.47×10^{-3}	0.49	-0.43	-1.12	9.52×10^{-3}	0.99	0.65	0.016	0.60
Max	-0.47	9.76×10^{-3}	9.76×10^{-3}	0.99	0.88	0.01	0.86	-0.17	-0.92	3.16×10^{-2}	0.99	0.87	0.05	0.86

Esse capítulo expôs os resultados da aplicação da metodologia proposta apresentada em dados PolSAR.

5

CONCLUSÕES

NESTE capítulo, iremos evidenciar alguns comentários sobre o trabalho apresentado.

5.1 Considerações Finais

Este trabalho, apresentou uma nova abordagem para visualizar dados multivariados complexos, em particular, valores das coordenadas de imagens PolSAR a qual estão organizados no formato enxuto. O objetivo dessa técnica é obter informações - através dessas matrizes de covariância complexas - que discriminem regiões contidas na imagem antes de processar a imagem.

A ferramenta foi utilizada em quatro amostras de tamanho 500 em cada imagem. Após utilizar a ferramenta nas imagens somos induzidos a supor que as imagens de regiões desérticas estão associadas a matrizes de covariância complexas que possuem: uma associação linear positiva muito forte entre as variáveis (ρ_{23}, ρ_{12}) no gráfico de dispersão bidimensional da segunda camada; um gráfico de dispersão bidimensional entre as variáveis $(\rho_{12}, \log \sigma_1)$ e entre as variáveis $(\rho_{23}, \log \sigma_1)$ bastante concentrado na vertical em torno do zero e um histograma da correlação ρ_{12} contido na primeira camada é concentrado em torno do zero (Ver Figuras 4.3, 4.4, 4.5 e 4.6).

Usando o raciocínio análogo ao do parágrafo anterior intuímos que regiões com características urbanas estão associadas a: um gráfico de dispersão bidimensional entre as variáveis (ρ_{23}, ρ_{12}) bastante dispersos; a gráficos de dispersão bidimensional entre as variáveis $(\rho_{12}, \log \sigma_1)$ e $(\rho_{23}, \log \sigma_1)$, respectivamente, muito dispersos; um histograma de dependência efetiva com valores muito pequenos e um histograma de correlação ρ_{12} bimodal (Ver Figuras 4.8, 4.9, 4.10 e 4.11).

Por fim, uma região aquática está relacionada a: uma associação linear positiva forte entre as variáveis (ρ_{23}, ρ_{12}) no gráfico de dispersão bidimensional; um gráfico de dispersão

bidimensional entre as variáveis $(\rho_{12}, \log \sigma_1)$ e entre as variáveis $(\rho_{23}, \log \sigma_1)$ concentrado em torno do zero (Ver Figuras 4.13, 4.14, 4.15 e 4.16).

Dessa forma, concluímos que a nossa abordagem de visualização de matrizes de covariância complexa é capaz de discriminar regiões dos alvos antes de processar a imagem.

Apêndice A

AMBIENTE REPRODUTÍVEL E COMPUTACIONAL

NESTE capítulo, apresentaremos o ambiente computacional necessário para o desenvolvimento da ferramenta. E por fim comentamos sobre sistema tipográfico.

A.1 Ambiente computacional utilizado

O ambiente computacional foi implantado em um *desktop* com a seguinte configuração:

- Processador intel Core i7;
- 8 GB de Memória RAM;
- 1T de HD;
- Sistema Operacional Ubuntu 14.04 LTS;
- Linguagem R versão 3.2.2;
- IDE Rstudio Version 0.98.1062 – ©2009-2013;
- T_EX 3.1415926 (TeX Live 2013/Debian);
- IDE T_EXmaker 4.1 Copyright ©2003-2013 by Pascal Brachet;

Grande parte do estado-da-arte desse trabalho, no que se refere a malha de conhecimento, foi obtido com o auxílio do [ISI Web of Knowledge \(2017\)](#). O [ISI Web of Knowledge \(2017\)](#) é um site acadêmico que tem como função indexar citações e fornecer conteúdo bibliográfico. Ele mostra-se uma ferramenta muito importante para cientistas, alunos de graduação e de pós-graduação de várias áreas do conhecimento. O site fornece informações relevantes, por

exemplo, *impact factor*, que permite selecionar os melhores periódicos da grande área de estudo e da linha de pesquisa, mais também os melhores autores através de uma medida chamada *index-h*. Ele contém hoje 23 mil acadêmicos e revistas científicas, incluindo *web of science*, 23 milhões de patentes, 110 mil conferências e 9 mil sites. A cobertura destas informações abrange o período de 1900 até os dias de hoje.

Para obter os artigos foi empregado o [Periódico CAPES \(2017\)](#). O [Periódico CAPES \(2017\)](#) é uma biblioteca virtual brasileira que possibilita o acesso a melhor produção científica do mundo. Essa biblioteca contém mais de 38 mil títulos com texto completo, 123 bases referenciais etc.

A implementação da ferramenta proposta nesse trabalho foi construída em R. [R Core Team \(2013\)](#) é uma linguagem, *Open Source*, para computação estatística e gráficos. Possui várias técnicas estatísticas disponíveis e gráficos. Encontra-se acessível como código fonte pela GNU (*General Public License*) e compila nos principais sistemas operacionais: UNIX, Windows e MacOS entre outros.

O ambiente reprodutível web utilizado para realização desta pesquisa foi <https://lproject.lccv.ufal.br/>.

Adotamos a linguagem \LaTeX para escrever o nosso texto, ao invés do word, write entre outros. \TeX é um software tipográfico que produz publicação de textos de alta qualidade, em particular, para escrita de equações matemáticas. É disponível de forma gratuita. O \LaTeX possui um conjunto de comandos adicionais para o \TeX produzido por Leslie Lamport.

REFERÊNCIAS

- Anderson, T. W., Anderson, T. W., Anderson, T. W. & Anderson, T. W. (1958), *An introduction to multivariate statistical analysis*, Vol. 2, Wiley New York. ISBN: 978-0-471-36091-9.
- Bell, D. (2016), 'An introduction to the unified modeling language'. URL <http://www.ibm.com/developerworks/rational/library/769.html>, [Online; accessed 10-Ago-2016].
- Bruder, J., Carlo, J., Gurney, J. & Gorman, J. (2003), 'IEEE standard for letter designations for radar-frequency bands', *IEEE Aerospace & Electronic Systems Society* pp. 1–3. URL <http://dx.doi.org/10.1109/IEEESTD.2003.94224>.
- Chen, S.-W. & Sato, M. (2013), 'Tsunami Damage Investigation of Built-Up Areas Using Multitemporal Spaceborne Full Polarimetric SAR Images', *IEEE Transactions on Geoscience and Remote Sensing* **51**(4, 1, SI), 1985–1997. URL <http://dx.doi.org/10.1109/TGRS.2012.2210050>.
- Everitt, B. & Hothorn, T. (2011), *An introduction to applied multivariate analysis with R*, Springer Science & Business Media. ISBN:978-1-4419-9650-3.
- Freitas, C., Sant'Anna, S., Rennó, C. & Correia, A. (2003), *Utilização de imagens de radar de abertura sintética na classificação de uso e ocupação do solo*, Geomática: Técnicas e Aplicações Avançadas. URL www.dpi.inpe.br/menu/Capacitacao/livros.php.
- Frery, A. C. & Perciano, T. (2013), *Introduction to Image Processing Using R*, SpringerBriefs in Computer Science, Springer. ISBN:978-1-4471-4950-7.
- Frery, A. C., Cintra, R. J. & Nascimento, A. D. C. (2013), 'Entropy-based statistical analysis of polar data', *IEEE Transactions on Geoscience and Remote Sensing* **51**(6), 3733–3743. URL <http://dx.doi.org/10.1109/TGRS.2012.2222029>.
- Frery, A. C., Nascimento, A. D. C. & Cintra, R. J. (2014), 'Analytic Expressions for Stochastic Distances Between Relaxed Complex Wishart Distributions', *IEEE Transactions on Geoscience and Remote Sensing* **52**, 1213–1226. URL <http://dx.doi.org/10.1109/TGRS.2013.2248737>.
- Friendly, M. (2002), 'Corrgrams: Exploratory displays for correlation matrices', *American Statistician* **56**(4), 316–324. URL <http://dx.doi.org/10.1198/000313002533>.
- Gonzalez, R. C. & Woods, R. E. (2010), *Processamento Digital de Imagens. Tradução: Cristina Yamagami e Leonardo Piamonte*, Sao Paulo: Pearson Prentice Hall. ISBN-13: 978-0131687288.

- Goodman, N. (1963), 'Statistical-Analysis Based on a Certain Multivariate Complex Gaussian Distribution (An Introduction)', *Annals of Mathematical Statistics* **34**(1), 152–&. URL <http://dx.doi.org/10.1214/aoms/1177704250>.
- ISI Web of Knowledge (2017), 'web of science'. URL www.webofknowledge.com, [Online; accessed 07-Fevereiro-2017].
- Jagdhuber, T., Hajnsek, I., Bronstert, A. & Papathanassiou, K. P. (2013), 'Soil Moisture Estimation Under Low Vegetation Cover Using a Multi-Angular Polarimetric Decomposition', *IEEE Transactions on Geoscience and Remote Sensing* **51**(4, 2), 2201–2215. URL <http://dx.doi.org/10.1109/TGRS.2012.2209433>.
- James, B. R. (2013), *Probabilidade: um curso em nível intermediário*, Projeto Euclides. ISBN: 978-85-244-0101-5.
- Johnson, R. A. & Wichern, D. W. (2007), *Applied Multivariate Statistical Analysis*, Prentice hall Englewood Cliffs, NJ. ISBN: 0-13-187715-1.
- Lee, J.-S. & Pottier, E. (2009), *Polarimetric radar imaging: from basics to applications*, CRC press. ISBN:9781420054972.
- Liu, C., Shang, J., Vachon, P. W. & McNairn, H. (2013), 'Multiyear Crop Monitoring Using Polarimetric RADARSAT-2 Data', *IEEE Transactions on Geoscience and Remote Sensing* **51**(4, 2), 2227–2240. URL <http://dx.doi.org/10.1109/TGRS.2012.2208649>.
- Magalhães, M. N. (2011), *Probabilidade e Variáveis Aleatórias*, Edusp. ISBN: 978-85-314-0945-5.
- Meneses, P. & Almeida, T. (2012), *Introdução ao processamento de imagens digitais de satélites de sensoriamento remoto*, Editora Universidade de Brasília. URL <http://www.cnpq.br/documents/10157/56b578c4-0fd5-4b9f-b82a-e9693e4f69d8>.
- Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I. & Papathanassiou, K. P. (2013), 'A tutorial on synthetic aperture radar', *IEEE Geoscience and Remote Sensing Magazine* **1**(1), 6–43. URL <http://dx.doi.org/10.1109/MGRS.2013.2248301>.
- Murdoch, D. & Chow, E. (1996), 'A graphical display of large correlation matrices', *American Statistician* **50**(2), 178–180. URL <http://dx.doi.org/10.2307/2684435>.
- Nascimento, A. D. C. (2012), *Teoria Estatística da Informação para Dados de Radar de Abertura Sintética Univariados e Polarimétricos*, PhD thesis, Universidade Federal de Pernambuco.

- Nascimento, A. D. C., Frery, A. C. & Cintra, R. J. (2014), 'Bias Correction and Modified Profile Likelihood Under the Wishart Complex Distribution', *IEEE Transactions on Geoscience and Remote Sensing* **52**, 4932–4941. URL <http://dx.doi.org/10.1109/TGRS.2013.2285927>.
- Peña, D. & Rodríguez, J. (2003), 'Descriptive measures of multivariate scatter and linear dependence', *Journal of Multivariate Analysis* **85**(2), 361–374. URL [http://dx.doi.org/10.1016/S0047-259X\(02\)00061-1](http://dx.doi.org/10.1016/S0047-259X(02)00061-1).
- R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rousseeuw, P. & Molenberghs, G. (1994), 'The Shape Of Correlation-Matrices', *American Statistician* **48**(4), 276–279.
- Periódico CAPES (2017), 'Busca por assuntos'. URL <http://www.periodicos.capes.gov.br/>, [Online; acessado 07-Fevereiro-2017].
- Tokuda, T., Goodrich, B., Van Mechelen, I., Gelman, A. & Tuerlinckx, F. (2011), 'Visualizing distributions of covariance matrices', *Columbia Univ., New York, NY, USA, Tech. Rep.*
- Trevett, J. W. (1986), *Imaging radar for resources surveys*, New York, NY, USA, Chapman and Hall. ISBN: 978-94-009-4089-5.
- Van Zyl, J. & Kim, Y. (2011), *Synthetic aperture radar polarimetry*, Vol. 2, John Wiley & Sons. ISBN: 978-1-118-11511-4.
- Velho, L., Frery, A. C. & Gomes, J. (2009), *Image processing for computer graphics and vision*, Springer Science & Business Media. ISBN: 978-1-84800-193-0.

Este trabalho foi redigido em \LaTeX utilizando uma modificação do estilo IC-UFAL. As referências bibliográficas foram preparadas no JabRef e administradas pelo \BIBTeX com o estilo LaCCAN. O texto utiliza fonte Fourier-GUTenberg e os elementos matemáticos a família tipográfica Euler Virtual Math, ambas em corpo de 12 pontos.