INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS—INPE

DIVISÃO DE BIBLIOTECA—DIBIB

How to standardize T&D using INPE style

São José dos Campos
2023
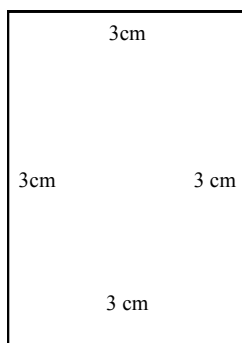
In this guide, you'll find the instructions to standardize your thesis or dissertation (T&D) accordingly to the style adopted by INPE

STARTING THE THESIS

Paper size: A4 (21,0 cm x 29,7 cm)

Margins: 3 cm (top, botton, left and right):

```
              3cm


3cm                      3 cm


              3 cm
```

Font: Times new roman 12, Arial or similar;

Spacing:

⇒ **1,5 cm** between lines and in the section titles that extend over two or more lines;

⇒ **Double space** between the section title and the first paragraph, and between paragraphs;

⇒ List of Figures, Table and Abbreviations: **Single spaced**.

Front Matter:

⇒ Use **roman numerals** starting with page two (ii) in the index card;

⇒ Use center alignment with **2.5cm** from the bottom;

⇒ Each section of the **Front matter** must start on an **odd page**. If necessary, leave a blank sheet.

Cataloging in Publication Data

Velame, Victória Maria Gomes.

Ve54o   Object detection from captive balloon imagery using deep learning / Victória Maria Gomes Velame. – São José dos Campos : INPE, 2020.

xxvi + 101 p. ; (sid.inpe.br/mtc-m21c/2020/03.30.16.14-TDI)

Dissertation (Master in Remote Sensing) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2020.

Guiding : Drs. José Claudio Mura, and Leonardo Sant'Anna Bins.

1. Object Detection. 2. Deep Learning. 3. Convolutional Neural Network. 4. Remote Sensing. 5. Captive Balloons. I.Title.

CDU 528.8:004.8

Text body:

⇒ From the **Introduction** section, enumerate with **arabic number** starting with page **1**

⇒ Use center alignment with **2.5cm** from the bottom.

1 INTRODUCTION

The increasing dependence of society on weather-sensitive technologies as well as the expansion of urban centers to risk areas are making meteorological modeling even more important in the last decades. Moreover, the development of powerful computational systems has made the implementation of new physical models capable of representing more precisely the atmosphere inducing several sectors of the economy to become even more dependent on weather forecasting.
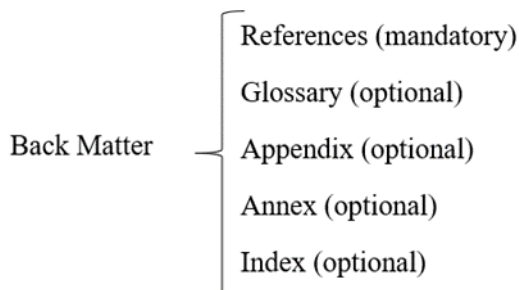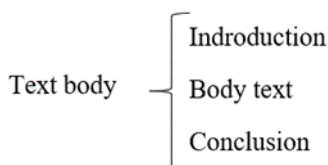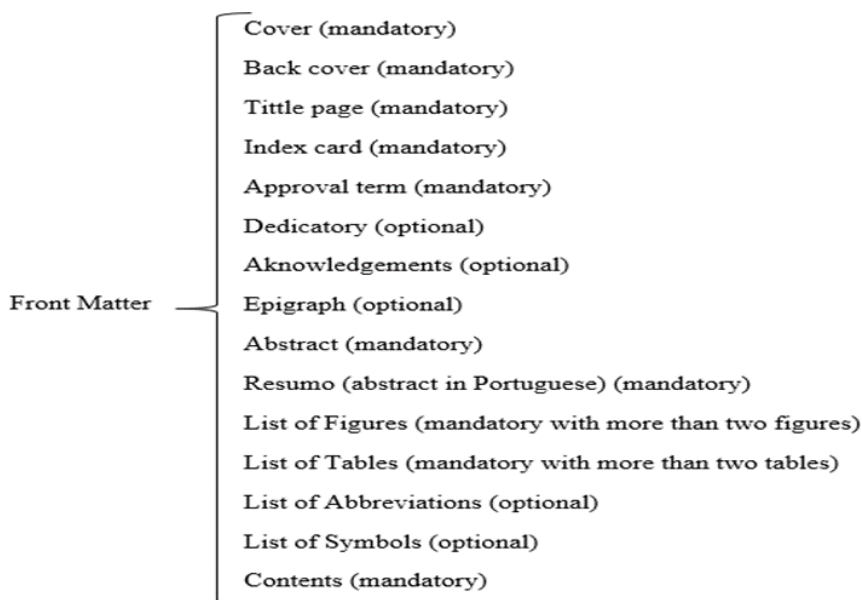
The prediction of the atmospheric conditions is usually made by using a set of equations that comprehend the laws of motion and principles like the conservation of mass and energy. Those equations do not have analytical solutions and it is necessary to solve them by applying numerical methods, that is why the weather forecasting procedures it is often referred to as Numerical Weather Prediction (NWP).

In order to improve the weather forecasts, it is possible to approach different modeling aspects. We can improve the forecasts (1) improving the model core and configurations like the spatial resolution, discretization methods, etc.; (2) Improving the physics, using different parameterizations or equations which describes the reality more precisely; (3) improving the data assimilation procedures; Or (4) applying ensemble techniques.

This work approaches the weather forecasts from the point of view of the data assimilation explaining how the insertion of a new data source can improve the initial conditions and subsequently the result of the simulation. Basically, the assimilation is the process of the combination of observational data from different sources with the simulated data, known as background or first guess, in order to obtain adjusted initial conditions, known as analysis, resulting in better weather forecasts (KALNAY, 2003; LORENZ, 1963).

The NWP is an initial value problem, i.e., the physical equations used in atmospheric modeling not only need initial conditions but also are highly sensitive to them. The sensibility of the models associated with the initial conditions as well as the high resolution of them make the simple interpolation

STRUCTURING YOUR THESIS

Front Matter
- Cover (mandatory)
- Back cover (mandatory)
- Tittle page (mandatory)
- Index card (mandatory)
- Approval term (mandatory)
- Dedicatory (optional)
- Aknowledgements (optional)
- Epigraph (optional)
- Abstract (mandatory)
- Resumo (abstract in Portuguese) (mandatory)
- List of Figures (mandatory with more than two figures)
- List of Tables (mandatory with more than two tables)
- List of Abbreviations (optional)
- List of Symbols (optional)
- Contents (mandatory)

Text body
- Indroduction
- Body text
- Conclusion

Back Matter
- References (mandatory)
- Glossary (optional)
- Appendix (optional)
- Annex (optional)
- Index (optional)

Cover elements:

  a) INPE logo;
  b) Identifier (included by DIBIB in the final publish);
  c) Títle (and subtitle);
  d) Author name;
  e) Description: Dissertation or Thesis, academic grade, Suppervisors names, and approval date;
  f) URL (included by DIBIB in the final publish);
  g) Institution, city and year.

MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÕES

PÁTRIA AMADA
BRASIL
GOVERNO FEDERAL

sid.inpe.br/mtc-m21d/2021/09.24.19.04-TDI

A STUDY ON THE GENERATION AND PROPAGATION OF CONCENTRIC GRAVITY WAVES OVER BRAZIL

Prosper Kwamla Nyassor

Doctorate Thesis of the Graduate Course in Space Geophysics, guided by Drs. Cristiano Max Wrasse, Delano Gobbi, and Igo Paulino da Silva, approved in october 25, 2021.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34T/45FSAG8>

INPE
São José dos Campos
2021

TITLE PAGE AND INDEX CARD

Title page: Mandatory element, with the same information from the cover.

Index card: The index card will be made by the INPE Library in the publication process.

ii

APPROVAL PAGE

Mandatory element, and it must appear after the Index card.
Approval page is provided by SEPGR.

ATTENTION: Leave a blank sheet to include the approval sheet in your thesis (which will be page **iii** of the publication).

MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÕES

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS
Serviço de Pós-Graduação - SEPGR

DEFESA FINAL DE DISSERTAÇÃO DE SABRINA PAES LEME PASSOS CORRÊA
BANCA Nº317/2021, REG 430777/2019

No dia 16 de dezembro de 2021, as 13h30min, por teleconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi APROVADO(A) pela Banca Examinadora, por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Mestra em Sensoriamento Remoto. O trabalho precisa da incorporação das correções sugeridas pela Banca e revisão final pelo(s) orientador (es).

Título: "Reference Sample Selection for supervised classification of a lower resolution image aided by a higher resolution image".

Observações da banca: O orientador é do exterior e vou encaminhar a declaração membro estrangeiro com as considerações dele (anexa ao processo).

Membros da Banca:

Dr. Thales Sehn Körting - Presidente - INPE
Dr. Luciano Vieira Dutra - Orientador - INPE
Dr. Derek D'Arcy Lichti - Orientador - Department of Geomatics Engineering - University of Calgary
Dr. Carlos Alberto Felgueiras - Membro Interno - INPE
Dra. Marinalva Dias Soares - Membro Externo - EMBRAER
Dr. Raian Varga Maretto - Membro Externo - Faculty of Geo-Information Science and Earth Observation, University of Twente

Declaração de aprovação do Dr. Derek D'Arcy Lichti anexa (8941442)

Documento assinado eletronicamente por **Thales Sehn Korting, Pesquisador**, em 03/02/2022, às 17:20 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Carlos Alberto Felgueiras, Tecnologista**, em 04/02/2022, às 10:03 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Raian Vargas maretto (E), Usuário Externo**, em 08/02/2022, às 14:10 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **LUCIANO VIEIRA DUTRA (E), Usuário Externo**, em 08/02/2022, às 22:40 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Epigraph: Optional. You can writte in English or in another language. You must cite the author and the content must be enclosed in quotation marks.

> "I fear the animals regard man as a being like themselves, seriously endangered by the loss of sound animal understanding; they regard him perhaps as the absurd animal, the laughing animal, the crying animal, the unfortunate animal."
>
> FRIEDRICH WILHELM NIETZSCHE
> "The Gay Science : Third Book, 224. Animal Criticism", 1882

v

Dedicatory: Optional

> To my parents **Eliomar** and **Marize** and my brother **Luan**, that have always supported me in this long journey.

vii

ACKNOWLEDGEMENTS

Optional.

ATTENTION: If you received any grant, the name of the funding agency must be included in the acknowledgements.

ACKNOWLEDGMENTS

To Professor Valeri Vlassov Vladimirovich, for his orientation and for his willingness to share his knowledge and advice, my eternal gratitude.

To the professors of National Institute for Space Research who with their knowledge, helped to develop this work.

To Dr. Rafael Lopes Costa who was friendly and helpful since my first day in INPE.

To Dra. Nadjara dos Santos and to Olga Viktorovna Kchoukina who helped to acoomplish the experiments of this work.

To my friends Olavo Mecias da Silva and André Caetano from the LVA – UFSC for their valuable contributions to this work.

To my friends at the UFSC, Mauricio Kubaski, Cristiano Engel, Pedro Thaines, Jonatan Feldkircher and Larissa Sirtuli who helped in some way for this work to be done.

To CAPES for the financial support.

vii

ABSTRACT

⇒ **Mantatory**. It must be clear, precise, and objective, highlighting the purposes, methodology, results, and conclusions of your thesis;

⇒ Structure your text in a single paragraph (single spaced).

⇒ In the abstract, avoid citations,formulas, abbreviations, symbols and equations;

⇒ Include **Keywords** after the Abstract.

ABSTRACT

Morphological classification is a key piece of information to define samples of galaxies aiming to study the large-scale structure of the universe. In essence, the challenge is to build up a robust methodology to perform a reliable morphological estimate from galaxy images. Here, I investigate how to substantially improve the galaxy classification within large datasets by mimicking human classification. I combine accurate visual classifications from the Galaxy Zoo project with machine and deep learning methodologies. I propose two distinct approaches for galaxy morphology: one based on non-parametric morphology and traditional machine learning algorithms; and another based on deep learning. To measure the input features for the traditional machine learning methodology, I and my collaborators have developed a system called CyMorph, with a novel non-parametric approach to study galaxy morphology. The main datasets employed comes from the Sloan Digital Sky Survey Data Release 7 (SDSS-DR7). I also discuss the class imbalance problem considering three classes. Performance of each model is mainly measured by overall accuracy (OA). A spectroscopic validation with astrophysical parameters is also provided for Decision Tree models to assess the quality of our morphological classification. In all of our samples, both Deep and Traditional Machine Learning approaches have over 94.5% OA to classify galaxies in two classes (elliptical and spiral). I compare our classification with state-of-the-art morphological classification from literature. Considering only two classes separation, I achieve 99% OA in average when using our deep learning models, and 82% when using three classes. I provide a catalog with 670,560 galaxies containing our best results, including morphological metrics and classification.

Keywords: Computational Astrophysics. Galaxy Morphology. Machine Learning. Deep Learning.

⇒ **Mandatory**, with the same characteristics of the abstract, but in Portuguese;

⇒ You must include at the begining of the page the thesis title in Portuguese;

⇒ **Single** spaced**;**

⇒ Include **Palavras-chave** after the *Resumo*.

APRENDIZADO DE MÁQUINA TRADICIONAL E PROFUNDO APLICADO A MORFOLOGIA DE GALÁXIAS

RESUMO

Classificação morfológica é peça chave de informação para definir amostras de galáxias com objetivo de estudar a estrutura do Universo em larga-escala. Em essência, o desafio é construir uma metodologia robusta para produzir uma estimativa morfológica confiável a partir de imagens de galáxias. Aqui, investigo como melhorar substancialmente a classificação automática de galáxias em grandes conjuntos de dados ao imitar a classificação fornecida por humanos. Combino classificações visuais do projeto Galaxy Zoo com metodologias de aprendizado de máquina tradicional e profundo. Proponho duas abordagens distintas para morfologias de galáxias: uma baseada em morfologia não-paramétrica e algoritmos de aprendizado de máquina tradicional; e outra baseada em aprendizado profundo. Para medir as características morfológicas de entrada para algoritmos de aprendizado de máquina tradicional, desenvolvi com meus colaboradores um sistema chamado CyMorph, com uma nova abordagem não-paramétrica para estudar morfologia de galáxias. O principal conjunto de dados explorado provém do Sloan Digital Sky Survey Data Release 7 (SDSS-DR7). Também discuto o problema de desbalanceamento de classes considerando o problema com três classes. A performance de cada modelo é medida principalmente por acurácia global. A validação espectroscópica com parâmetros astrofísicos também é fornecida para os modelos de Árvore de Decisão para avaliar a qualidade de nossa classificação morfológica. Em todas as nossas amostras, tanto com aprendizado de máquina profundo como tradicional, obtenho mais de 94.5% de acurácia global para classificar galáxias em duas classes (elíptica e espiral). Comparo minha classificação com classificações morfológicas do estado-da-arte da literatura. Considerando apenas duas classes, atingi 99% de acurácia global e média usando modelos de aprendizado profundo, e 82% usando três classes. Forneço uma catálogo com 670.560 galáxias contendo nossos melhores resultados, incluindo métricas morfológicas e classificações.

Palavras-chave: Astrofísica Computacional. Morfologia de Galáxias. Aprendizado de Máquina. Aprendizado de Máquina Profundo.

ix

LIST OF FIGURES

⇒ Mandatory when there is more than two figures in the text;

⇒ If possible, include the term "Figure" in each list item.

xiii

LIST OF TABLES

⇒   Mandatory when there is more than two tables in the text;

⇒   If possible, include the term "Table" in each list item.

LIST OF ABBREVIATIONS

⇒ Optional; abbreviations must appear alphabetically;

⇒ All abbreviations are to be written out.

LIST OF ABBREVIATIONS

AI	–	Artificial Intelligence
AMSR-E	–	Advanced Microwave Scanning Radiometer-Earth Observing System
AMSU	–	Advanced microwave Sound Unit
ANN	–	Artificial neural network
API	–	Application programming interface
CDRD	–	Cloud Dynamics and Radiation Database
CHUVA	–	Cloud processes of the main precipitation systems in Brazil: A contribution to cloud resolving modeling and to the Global Precipitation Measurement
CLASS	–	Comprehensive Large Array-data Stewardship System
CMB	–	Combined Radar-Radiometer Algorithm
CRM	–	Cloud resolving model
CSU	–	Colorado State University
DPR	–	Dual-frequency Precipitation Radar
DSD	–	Drop size distribution
GCOMW1	–	Global Change Observation Mission-Water 1
GFS	–	Global Forecast System
GMI	–	GPM Microwave Imager
GPM	–	Global Precipitation Measurement
GPM-CO	–	GPM Core Observatory
GPROF	–	Goddard Profiling Algorithm
H	–	Horizontal polarization
H-SAF	–	Satellite Application Facility on Support to Operational Hydrology and Water Management
ILW	–	Integrated liquid water
IMERG	–	Integrated Multi-satellite Retrievals for GPM
IR	–	Infrared
ITCZ	–	Intertropical Convergence Zone
IWP	–	Ice Water Path
JAXA	–	Japan Aerospace Exploration Agency
LWC	–	liquid water content
LWP	–	Liquid Water Path
MCC	–	Mesoscale Convective Complexes
METOP(A/B)	–	European Meteorological Operational A and B
MHS	–	Microwave Humidity Sounder
ML	–	Machine Learning
MRR	–	Micro Rain Radar
MT1	–	Megha-Tropiques 1
MW	–	Microwave
NN	–	Neural network

LIST OF SYMBOLS

⇒ Optional. You must include the symbols used in your thesis and their meanings written in full.

⇒ It is recommended to specify the units;

⇒ Sequence for List of the Symbols: Latin alphabet; Greek alphabet, other alphabet; higher index; lower index; special symbols.

LIST OF SYMBOLS

Symbol	Meaning
$c = \hbar = 1$	Natural units
t	Proper time
η	Conformal time
Greek letters (μ, ν)	space-time coordinates
Latin letters (a, b)	spatial coordinates
$\dot{}$	Derivative with respect to the proper time
$^{(3)}$	Three-dimensional space
∂_μ	Partial derivative with respect to the coordinates
∇_μ	Covariant derivative with respect to the coordinates
a	Scale Factor
$\gamma_{\mu\nu}$	Minkowski metric or background metric
$g_{\mu\nu}$	Space-time metric
q_{ab}	Hypersurface metric
N	Lapse function or e-fold number
N^α	Shift vector
δ^μ_ν	Kronecker delta
$\Gamma^\rho_{\mu\nu}$	Christoffel symbol
$R^\rho_{\mu\nu\sigma}$	Riemann tensor
$R_{\mu\nu}$	Ricci tensor
R	Ricci scalar
$G_{\mu\nu}$	Einstein tensor
G	Newtonian constant
m_{Pl}	Planck mass
M_{Pl}	Reduced Planck mass
ℓ_{Pl}	Planck length
t_{Pl}	Planck time
$T_{\mu\nu}$	Energy-momentum tensor
ρ	Total energy density
ρ_c	Critical energy density
ρ_{Pl}	Energy density at Planck scales
ρ_{kin}	Kinetic energy density
ρ_{eff}	Effective energy density
ϵ, η_{SR}	Slow-roll parameters
P	Pressure
w	State parameter
u^μ	Four-velocity
Ω	Energy density parameter or conformal transformation
Ω_k	Curvature parameter
ℓ	Loop length or multipole momentum
S	Action

xix

CONTENTS

Mandatory. It is the enumeration of the main topics of the thesis and their sections, with their respective page number starting in the Introduction.

CONTENTS

The text body is composed by the Introduction, the Main text, and the Con-
clusion.

INTRODUCTION

It contextualizes your research problem from a broadened to a specific pers-
pective where the research objectives and hypotheses are presented.

MAIN TEXT

It's divide into sections and subsections. Must contain:

a) **Literature review:** Critical analysis of the available knowledge about
the research theme. It discusses theoretical and/or practical limitations
and methodologies, pointing to new research perspectives;
b) **Materials and Methods:** It describes the methodology used for your
thesis: how, when and where the research was made;
c) **Results:** It contains all the relevant results necessary to answer your
hypothesis. Use tables and graphics to better expose them.
d) **Discussion:** reestate your main findings and discuss them based on the
current and/or classical studies. In this part, you can answer your pro-
posed questions and present future research directions.

CONCLUSION

Extract the meaning from your results in face of the discussion you have ma-
de in order to refute or corroborate your hypotheses, or to conclude about
your set goals.

Sequential numbering:

The thesis is structured in chapters, sections, subsections, etc. They are identified by progressive numbering using bold Arabic numerals left aligned.

Section titles:
Section titles (chapter, section, subsection, etc.) must be placed one space after their respective number.

Example:

2. THEORETICAL FOUNDATIONS

2.1 Galaxy morphology

2.1.1 Galaxy zoo

4 EXPERIMENTAL PLANNING

Once the theoretical idea is presented, in order to proceed with the development of the methodology, the experiment is done in a controlled situation to study the effect of samples quality when the image to be classified is perfectly registered to the baseline image which is to be obtained by a simulation.

In a nutshell, the algorithm uses the higher-resolution image (HR) together with the lower-resolution image(LR) and, for each LR pixel, it evaluates the number of HR cells inside a LR cell and determines whereas that LR cell will be selected or not as reference data. Figure 3.1 illustrates this approach.

Hence, by analysing a classified HR, the algorithm can determine which LR cells have an acceptable proportion of the modal class. A LR cell is selected if and only if the defined modal class proportion (*prop*) of HR cells from the modal class within it, in equals to or greater than 50%. Once the algorithm is finished selecting candidate samples, it can proceed to the LR classification and validating it (i.e. assessing thematic accuracy and completeness), presented in Section 4.2.

In order to move on to to studying the quality of reference data, the used materials for the stages are presented in Section 4.1.

4.1 Materials

The materials used for this study were softwares and programming environments as well as satellite imagery. An important factor considered was the open-source policy so this study can be easily reproduced.

Another point is that the author considered a colourblind friendly policy, therefore the majority of images are presented in a manner that colourblind people can easily read; if not, images in black & white (BW) format are presented in the Appendix C.

4.1.1 Study area

The study area lays in the Brazilian Amazon, located in Pará State, in the mesoregion of Sudoeste Paraense and microregion of Altamira. Its extents are: $3°36'48''S$ to $4°36'12''S$, and $53°18'40''W$ to $54°21'0''W$ in geographical coordinates, Datum SIRGAS 2000. The area is approximately $12,068 km^2$. Most of the study area belongs to the municipality of Altamira while the remaining area covers Placas and Uruará municipalities, with their respective urban areas, as shown in Figure 4.1.

When you list several subjects that do not have a section, you can subdivide them into points.

a) The last part of the text preceding the points ends in a colon;
b) The points are sorted alphabetically: a), b), c) etc.;
c) The letters that indicate the points start in the sixth space of the text;
d) Text begin under the first letter of the point text itself;
e) Points must be finished by a semicolon.

Subpoints: Points can be subdivided into subpoints:

– subpoints start in the ninth space of the text and must start with a hyphen,
– the following lines of the subpoint start under the first letter of the text itself,
– subpoints begin with a lowercase letter and are punctuated with a comma at the end.

Homogeneous nucleation applies to regions with temperatures below -40°C. In this case, the liquid particles freeze homogeneously and there is no need for the pre-existence of the freezing nucleus (STRAKA, 2009). However, gas-phase water still needs the presence of a freezing nucleus to freeze. It happens because it is statistically impossible that gas molecules could form the ice structure while colliding randomly.

In general, the ice particles are divided into ice crystals, snow, graupel, and hail:

a) ice Crystal: are small frozen particles that can be shaped like a column, plate, dendrite, or a combination of these shapes. They can be present in regions of the cloud with temperatures below -15°C. The ice particles formed in the homogeneous and heterogeneous nucleation process become ice crystals when they grow through the deposition process. Deposition in mixed clouds occurs due to the Bergeron Process[1] (BERGERON, 1935), as shown in Figure 2.3;
b) graupel: It is formed when the ice particles (usually ice crystals), generated by heterogeneous and homogeneous nucleation, grow by the process of accretion and /or deposition;
c) hail: In some cases, graupel growth continues until it becomes an ice rock known as hail. The hail refers to hail that grew by the accretion process and reached the criterion of 5 mm in diameter (criterion adopted by the United States National Weather Service). When the hail grows long enough that the upward current can no longer sustain it, it precipitates and, if it does not thaw on its descent, reaches the ground;
d) snow: When particles of ice crystals grow through the aggregation process, snow is formed. Snowfall occurs in nimbostratus clouds.

---

[1] When cloud droplets serve as a source to increase water vapor in the environment, they cause hail and / or ice crystals to grow by deposition more quickly. This occurs when the air is supersaturated with respect to ice and saturated with respect to water. This mechanism known as the Bergeron process. More information in Bergeron (1935) and Grimm (1999).

10

It is the attribution of authorship to a information that came from another source.

All in-text citations must appear in the **References**.

Direct quotes:
It is a textual transcription of part of the consulted work. With up to three lines, it must be enclosed in double quotation marks, and include the author(s), year and page.

The inclusion of the page number is **optional.**

When there is an interruption or omission of parts of the text transcription, use [. . .]. Example:

In Manning et al. (2009), some definitions are made as follows: "token is an instance of a sequence of characters [...] also referred to as dictionary or lexicon, is the set of terms.

Direct citation with more than three lines must be highlighted with an indentation of 4 cm from the left margin, with a smaller letter than the text main and without quotation marks.
Example:

> An observable change in coronal structures that (1) occurs on a timescale between a few minutes and several hours and (2) involves the appearance of a new, discrete, bright white-light feature in the coronagraph field of view (HOWARD et al., 1997, p. 18).

Indirect quotes:
It is the information or idea of another author, expressed in your own words.

Citation from another citation:
It is the citation of a text from which the original was not accessed. It should be avoided. When necessary, it is recommended to use "cited by". Examples:

For Sander (1920) cited by Phillips (1984) is the study of expressions.

CITATIONS

Author-date

Cite the author's surname or the name of the institution followed by the year of publication of the document.
The citation can be included in the sentence or enclosed in parentheses. Examples:

One author:

Cintra (2018) has applied gradient pattern analysis (GPA) on the type

Two authors:

The unipolar inductor model was suggested by Piddington and Drake (1968)

Three authors:

Kuhn, Smith, and Johnson (2013) point out that the number of repetitions

More than three authors

Use the expression **et al:** the analysis made by Torres et al. (2017)

Author-date

If both author's surname and year of publication coincide:

Include the authors' first name initials. Example:

Santos, J. (2019) and Santos, M. (2019) proposes the use of SOM

If author's surnames, first names, and year of publication coincide:

Write the author's names in full. Example:

Wiley, James (1985) and Wiley, John (1985) proposed the use of an antenna

Citing of more than two works of the same author in the same year:
I
nclude the letters a, b, c to differentiate the works. Example:

Barbera (2010a, b) studying a sample of 40000 ETGs

Citing of works of the same author published in different years:

Cite the works chronologicaly. Example:

Walter (2008, 2011) classification system were based on in situ observations

Citing several works of different authors:

Include the citations in alphabetical order by surnames. Example:

The NPhardness of the cluster editing problem was proved, independently, by Bansal et al (2007); Delvaux (2004) and Shamir et al. (2008).

When the citation is included in parentheses, all letters of the author's surname must be in uppercase.

Examples:

One author:

...(KURKDJAN, 1993).

Two authors:

...(LOCH; KURCHNER, 1988)

Three authors

... be unreliable (THEODORIDIS; SMITH; KOUTROUMBAS, 2009) **.**

More than three Authors:

Use the expression **et al:** (BINS et al., 1996);

Citing several works of different authors:

Include the authors in alphabetical order by surnames. Example:
...(FOSTER, 1985; JENSEN, 1983;WELCH, 1982).

Institution as the Author:

You can cite its acronym. Example:

... in various geomorphometric characterizations (INPE, 1987).

Works without indication of authorship or institution:

Use the first word of the title followed by an ellipsis and the year of publication of the document. Example:

(THE SUPERCOMPUTERS. . . , 2004, p. 20).

Citation of works in preparation:

Include the expression "in preparation" to replace the year. Example:

... (LEE et al., in preparation).

Citation of works submitted or accepted for publication:

Include the expression. "submitted" or "accepted". Example:

...(SABA, submitted in 2020).

CITATIONS

Highlighting cited texts:

If you want to highlight the text you have cited, use bold, italics, and add the expression "highlighted by the author" in parentheses. Example:

Smith (1982, p. 48, highlighted by the author).

Translation of the cited text:
When the citation includes text translated by the author, include the expression "our translation" in parentheses. Example:

...speckle filterin (TORRES et al, 2004 author's translation).

Numerical System

Use Arabic numbering to identify each work cited in the text in the order of their apprearance, maintaining their specific number along the text.
Example:

The MSU [. . .] theory given in (17).

ACRONYMS AND ABBREVIATIONS

Acronyms can be used after their appear in full in the text followed by the acronym in parentheses.

Examples:

Instituto Nacional de Pesquisas Espaciais (INPE);

United Nations Educational, Scientific and Cultural Organization (UNESCO).

⇒   Indications, observations, or additions to the text made by the author.

⇒   The second line must be aligned with the footnote number.

CHAPTER 2 - ESTIMATES OF ANNUAL SOIL LOSS RATES IN THE BRAZILIAN CERRADO

2.1 INTRODUCTION[1]

Soil erosion is a serious environmental problem that has adversely affected food production throughout the world by reduction of land productivity and water availability (PIMENTEL et al., 1987). Erosion is a natural geomorphic process that results from topsoil removal by wind and water (GARES et al., 1994), and can be influenced by several factors such as climate variables, slope steepness, soil physical parameters, vegetation and land use patterns (PIMENTEL et al., 1995). Moreover, erosion processes may be intensified by human intervention through inappropriate land use and land cover changes. Severe soil erosion has occurred in the world's major agricultural regions and worsened with growing agricultural activities in forest fringe areas (PIMENTEL et al., 1987). Agricultural extensification and intensification that did not take into account the bearing capacity of soils have accelerated the erosion processes in some tropical regions (GRECCHI et al., 2014). World food projections have pointed out that these areas are particularly important in the global agricultural scenario as potential sites of farmland expansion that will ensure food security to an additional 2.3 billion people by 2050 (RADA, 2013; TILMAN et al., 2011; ALEXANDRATOS & BRUINSMA, 2012). Nonetheless, many of these regions have been classified as hotspots of land degradation due to water erosion (SCHERR & YADAV, 1996).

Brazil has become the second-largest exporter of agricultural products in the world, particularly due to production expansion into soils of its tropical savanna, known as *Cerrado* (RADA, 2013). Between 2009 and 2010, the Cerrado accounted for 70% of Brazil's food production (WICKRAMASINGHE et al., 2012) growing 95% of its cotton, 54% of soybeans, 55% of meat and 43% of its sugarcane (IBGE, 2010a). It is the country's most important agricultural region and has been considered one of the world's great breadbaskets (THE ECONOMIST, 2010). However, due to landscape fragility and

---

[1] This chapter is an adapted version of the paper:
GOMES, L.; SIMOES, S.J.C.; FORTI, M.C.; OMETTO, J.P.H.B.; DALLA-NORA, E.L. Using geotechnology to estimate annual soil loss rate in the Brazilian Cerrado. **Journal of Geographic Information System**, v.9, p. 420-439, 2017.

FIGURES

a) The figure **title** must appear at the **top** of the figure;

b) Figures must be numbered sequentially in **Arabic numerals** followig the chapter number. Example: Figure 1 in Chaper 1 = Figure 1.1

c) Figures included in **appendices and annexes** must be numbered sequentially according to the letter of the appendix or annex: Figure A.1, Figure A.2; Figure B.1, Figure B.2;

d) Figures should be **centered** on the page

e) **Don't break** a paragraph with a Figure.;

f) Use a **10 or 11** font size for the title, caption and figure source;

g) If the figure title has a **single line**, the alignment of the title is **centered**;

h) If the figure title has **more than one line**, the alignment of the title is **justified.** The second line of the figure's title must start below the first letter of the title;

i) The figure **source** must be included **below** the figure:

   – The first letter of the word Source must be in uppercase and followed by a colon.

   – After the colon, include the author's surname and the year of publication in parentheses:

   Source: Smith (2015).

For tables and figures examples, access:

http://urlib.net/rep/8JMKD3MGP3W34R/45523QH

Figure examples:

The figure title with **only one line** (centered);

Figure 2.2 - Basic (a) and modified drainage patterns (b).
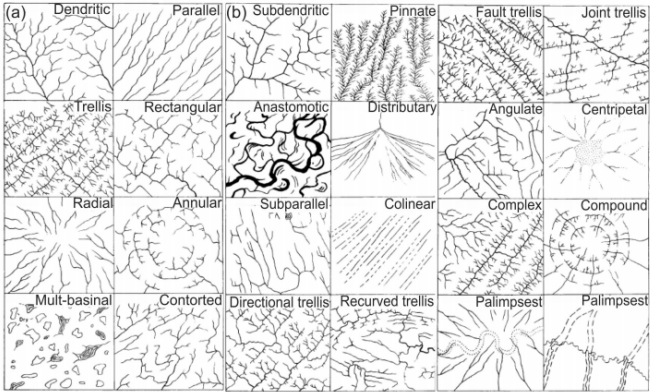
SOURCE: Adapted from Howard (1967).

–   The figure title with **two lines** (justified text with the second line aligned
the beginning of the first line—not with the Figure number.

Figure 2.1 - General methodology: For each year, we characterized the forest degraded areas
according to three indicators: i) water deficit anomaly indicator, ii) historical clear
cut deforestation indicator, and iii) protected areas indicator.
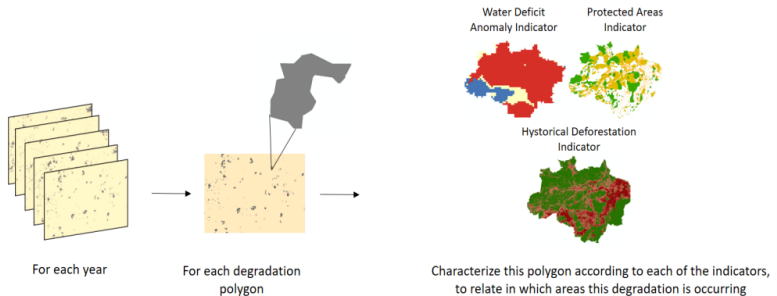
Source: Prepared by the author.

FIGURES

Figure with caption:

The figure caption must be below the image;

Figure 1.1 - Mars and its moons, Phobos and Deimos.
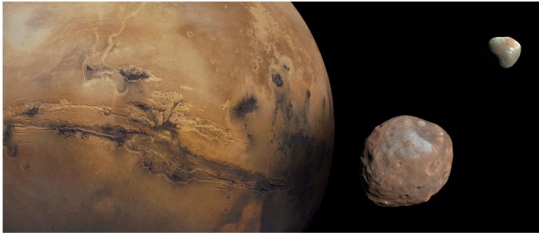
This is an assembly of pictures of Mars, Phobos, and Deimos, in scale. The mosaic of Mars is made from a compilation of images captured by the Viking Orbiter 1. The pictures of Phobos and Deimos are enhanced-color images from the Mars Reconnaissance Orbiter mission.

SOURCE: Courtesy NASA/JPL-Caltech (2008).

Caption and surce: both with justified alignment.

Figure 3.2. Associations among the size of a municipality's Region of Interest (ROI), the old-growth forest area (km²), total deforested area (km²) from 2017-2019 and the annual deforestation rate (%).
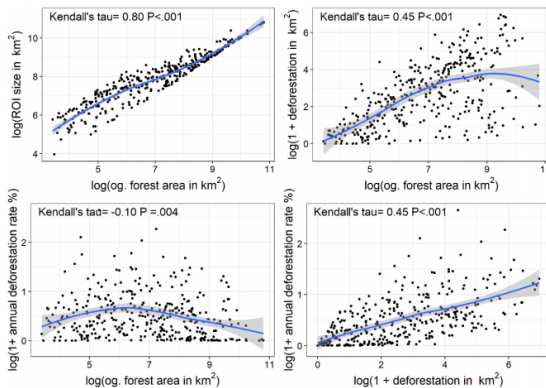
Values were log transformed for visualization purposes. Blue lines represent the estimated trend of the relationship based on a locally weighted smoothing (LOWESS), and shade represents the confidence interval for a 95% confidence level.

Source: Produced by this author.

Figures that occupy more than one page in the text:

a) At the end of the figure on each page, add the term "**to be** continued" or "**continue**";
b) At the beginning of the folowwing pages, include the figure number and replace the expression. "**Continuation**";
c) On the last page of the figure, include the figure number and replace the titlge by the expression. "**Conclusion**";
d) Caption: if the figure has caption, add it ate the end of the figure in the last page following the previous orientations;
e) Source: add it at the end of the figure in the last page following the previous orientations..

Example:

Figure 4.12: Percentage change between corrected and uncorrected data for BRDF by MAIAC.
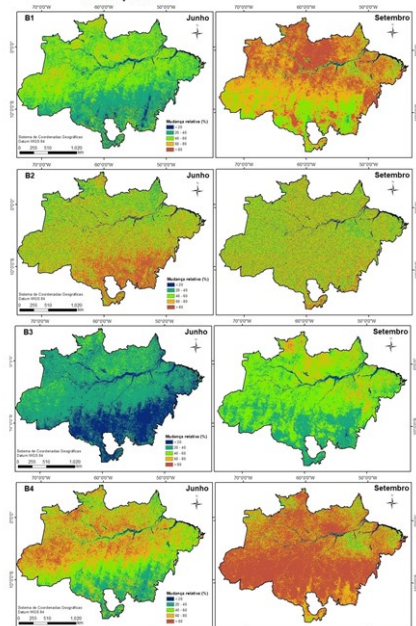
to be continued

Figure 4.12: Continuation.

to be continued

60

Figure 4.12: Conclusion.
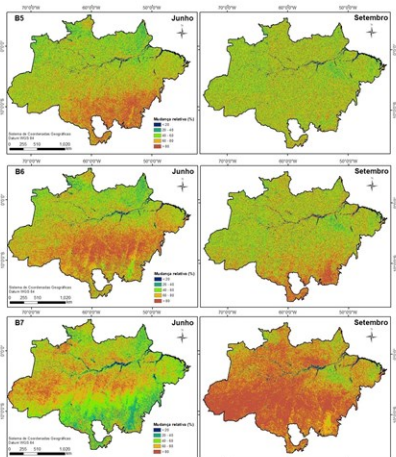
Source: Author.

TABLES

a) The **title** must appear at the top of the table;

b) Tables must be numbered sequentially in Arabic numerals followig the chapter number. Example: Table 1 in Chaper 2 = Table 2.1;

c) Tables included in **appendices and annexess**, must be numbered sequentially according to the letter of the appendix or annex: Table A.1, Table A.2; Table B.1, TableB.2;

d) Tables should be **centered** on the page;

e) **Don't break** a paragraph with a Table;

f) Use a **10 or 11** font size for the title, caption and table source;

g) If the tables title has a **single line**, the alignment of the title is **centered**;

h) If the table title has **more than on line**, the alignment of the title is **justified.** The second line of the table's title must start bellow the first letter of the title;

i) The table **source** must be included **bellow** the figure:

   – The first letter of the word Source must be in uppercase and followed by a colon:

   Source:

   – After the colon, include the author's surname and the year of publication in parentheses:

   Source: Smith (2015).

For tables and figures examples, access:

http://urlib.net/rep/8JMKD3MGP3W34R/45523QH

Table examples:

The table title with only one line (center aligned);

Table 5.1 - SIR-C images parameters.

Year of acquisition	1994
Image Size	407 x 370 pixels
Frequency	L band $(1.25 GHz)$ / C band $(5.304 GHz)$
Polarization	hh, hv, vv
Product Type	Complex Multilook
Number of nominal looks	4.79

SOURCE: Correia (1998).

The table title with two lines (justified text with the second line aligned with the beginning of the first line—not with the table number.

Table 4.1 - Aerosol parameters considered for the sensitivity tests using bin and bulk PSDs: intervals for values and steps between them.

Parameter	Set 1: bin PSDs		Set 2: bulk PSDs	
	Interval	Step	Interval	Step
N_a (cm^{-3})	$800 - 3600$	400	$200 - 900$	100
$\bar{r}_a$ (μm)	$0.05 - 0.11$	0.01	$0.02 - 0.08$	0.01
σ_a (-)	$1.6 - 2.2$	0.1	$1.1 - 1.9$	0.2
κ (-)	$0.1 - 0.5$	0.1	$0.1 - 0.5$	0.1

Table with caption:

Caption and Source: Both with justified alignment

Table 5.3 - Results with the unpaired two-samples T and Mann-Whitney U statistical tests comparing attributes derived from the northern and southern knickpoints of the study area.

Unpaired two-samples T-test							
Knickpoint attribute	Sector	Average	Standard deviation	Shapiro-Wilk test: p-value	Levene test: p-value	Two sample t-test: t-statistic	Two sample t-test: p-value
Elevation	N	339.1	200.0	0.337	0.004	0.119	0.907
	S	331.2	85.5	0.168			
* Relative relief	N	139.1	41.9	0.471	0.003	-1.640	0.124
	S	186.7	82.5	0.098			
** Relative relief	N	208.5	66.8	0.575	0.013	-2.610	0.021
	S	326.3	128.0	0.223			
*** Relative relief	N	277.6	106.0	0.122	0.591	-3.060	0.006
	S	421.7	109.0	0.074			
$*k_{sn}$	N	28.7	18.1	0.099	0.036	-1.630	0.124
	S	46.1	29.1	0.362			
$**k_{sn}$	N	24.5	14.8	0.090	0.383	-2.010	0.059
	S	38.7	17.7	0.506			
$***k_{sn}$	N	23.0	13.3	0.113	0.570	-1.747	0.097
	S	32.3	10.9	0.348			
Mann-Whitney U test							
Knickpoint attribute	Sector	Median	Interquartile range			Mann Whitney test: W-statistic	Mann Whitney test: p-value
Slope	N	2.71	1.9	-	-	47	0.5965
	S	2.74	0.9	-	-		
Incision	N	100	102	-	-	37	0.218
	S	118	122	-	-		
Drainage basin area	N	199	257	-	-	48	0.6468
	S	145	306	-	-		

Relative relief and k_{sn} tested using three different sizes of moving windows: 1000(*), 3000(**) and 5000 m (***).

Table 3.2 - Equations and references of the five extracted Vegetation Indices.

Vegetation Index	Equation	Reference
Normalized Difference Vegetation Index (NDVI)	$\dfrac{\rho NIR - \rho R}{\rho NIR + \rho R}$	Rouse et al. (1974)
Enhanced Vegetation Index (EVI)	$2.5\dfrac{\rho NIR - \rho R}{\rho NIR + C_1 * \rho R - C_2 * \rho B + L}$	Huete et al. (1997)
2-band Enhanced Vegetation Index (EVI2)	$2.5\dfrac{\rho NIR - \rho R}{\rho NIR + 2.4\rho R + 1}$	Jiang et al. (2008)
Soil-Adjusted Vegetation Index (SAVI)	$\dfrac{\rho NIR - \rho R}{\rho NIR + \rho R + L_1} * (1 + L_1)$	Huete (1988)
Modified Soil -Adjusted Vegetation Index 2 (MSAVI2)	$\dfrac{2 * \rho NIR + 1 - \sqrt{(2 * \rho NIR + 1)^2 - 8 * (\rho NIR - \rho R)}}{2}$	Qi et al. (1994)

where ρNIR is the Near Infrared reflectance, ρR is the reflectance in the Red band, ρB is the reflectance in the Blue band, L is the soil line adjustment parameter (1), C_1 and C_2 are the aerossol resistence terms (6 and 7.5) and L_1 is a correction factor that depends on vegetation cover (0.5).
SOURCE: Author's production.

TABLES

Tables that occupy more than one page in the text:

a) At the end of the table on each page, add the term "**to be continued**" or "**continue**".

b) At the beginning of the following pages, include the table number and replace the title with by the expression. "**Continuation**."

c) On the last page of the table, include the figure number and replace the title by the expression. "**Conclusion**."

d) Source: add it at the end of the table in the last page following the previous orientations.

e) Caption: If the table has caption, add it at the end of the table in the last page following the previous orientations nclude the table title after numerical identification

Example:

Table 2.2- Summary of studies exploring tree species classification using hyperspectral data. Works developed in tropical or subtropical forests are highlighted in gray. Those combining hyperspectral + LiDAR data contain the point density information at the 'Spatial resolution' column.

Study	Sensor	Spatial resolution (m)	Spectral resolution	Forest/ Country	Classifier	Number of species	Best accuracy (%)
Clark et al. (2005)	HYDICE	1.6	VNIR-SWIR (400–2500 nm; reduced to 30 bands selected)	Tropical Forest, Costa Rica	LDA, MLC, SAM	7	92
Jones et al. (2010)	AISA Dual	2 (0.4 points/m²)	VNIR-SWIR (429–2400 nm, reduced to 40 spectral bands)	Boreal Forest, Canada	SVM	11	72
Clark; Roberts (2012)	HYDICE	1.6	VNIR-SWIR (400–2500 nm; 210 bands)	Tropical Forest, Costa Rica	RF	7	87
Cho et al. (2012)	CAO Alpha	1.1	VNIR (384–1054 nm; 72 bands)	Savanna, South Africa	MLC	6	65
Dalponte et al. (2012)	AISA Eagle, GeoEye and ALS Optech ALTM	1 and 0.5 (8.6 and 0.48 points/m²)	VNIR (400–990 nm; 126 bands)	Temperate Forest, Italy	SVM e RF	7 species + non forest class	74
Naidoo et al. (2012)	CAO Alpha System	1.1 (1.3 point/m²)	VNIR (348-1054 nm, 72 bands)	Savanna, South Africa	RF	8	87.7

TABLES

Table 2.2- Continuation.

Study	Sensor	Spatial resolution (m)	Spectral resolution	Forest/ Country	Classifier	Number of species	Best accuracy (%)
Ferét; Asner (2013)	CAO Alpha	0.56	VNIR (390–1044 nm; 24 bands)	Tropical Forest, Hawaii	LDA, QDA, RDA, KNN, ANN and SVM	17	73
Dalponte et al. (2014)	HySpex VNIR-160 and ALS Optech ALTM	0.4 (7.4 point/m²)	VNIR (410-990 nm, 160 bands)	Boreal Forest, Norway	SVM	2 species + mixed class	93
Ghosh et al. (2014)	HyMap and Hyperion	4, 8 and 30 (12 point density)	VNIR-SWIR (450-2500 nm, 128 bands)	Temperate Forest, Germany	SVM and RF	5	86
Baldeck and Asner (2015)	CAO AToMS	1.12	VNIR-SWIR (380–2512 nm, 167 bands)	Tropical Forest, Panama	Single-class SVM	3	94
Ferreira et al. (2016)	ProSpecTIR-VS	1	VNIR-SWIR (450-2400 nm, 357 bands)	Subtropical Forest, Brazil	LDA, SVM and RF	8	84.9
Ballanti et al. (2016)	AISA Eagle and Leica ALS60	2 (4 and 8 points/m²)	VNIR (397.78–997.96 nm, 128 bands)	Temperate Forest, USA	SVM and RF	8	90
Richter et al. (2016)	AISA dual	2	VNIR-SWIR (400- 2497 nm, 267 bands)	Temperate Forest, Germany	SVM, RF and PLSDA	10	78.4
Graves et al. (2016)	CAO AToMS	2	VNIR-SWIR (380–2512 nm)	Tropical Forest, Panama	SVM imbalanced	20 species + 1 mixed class	63
Nevalainen et al. (2017)	UAV-FPI	0.086	VNIR (507-819 nm, 33 bands)	Boreal Forest, Finland	RF, MLP, C4.5, KNN	4	95
Piiroinen et al. (2017)	AISA Eagle and Optech ALTM 300	1 (9.6 points/m²)	VNIR (400-1000 nm, 129 bands)	Agroforestry, Kenya	SVM and RF	31	57.1
Raczko; Zagajewski (2017)	APEX	3.35	VNIR-SWIR (413-2447 nm, 288 bands)	Temperate Forest, Poland	ANN, SVM and RF	5	77
Shen; Cao (2017)	AISA Eagle and RIEGL LMS-Q680i	0.6 (10 points/m²)	VNIR (398.55– 994.44 nm, 64 bands)	Subtropical Forest, China	RF (object approach)	5	85.4

continue

Table 2.2- Conclusion.

Study	Sensor	Spatial resolution (m)	Spectral resolution	Forest/ Country	Classifier	Number of species	Best accuracy (%)
Tuominen et al. (2018)	UAV-FPI	0.08	VNIR-SWIR (409-1578 nm, 60 bands)	Arboretum, Finland	KNN+GA and RF	26	82.3
Maschler et al. (2018)	Hyspex VNIR 1600 (160SB)	0.4	VNIR (415-991 nm, 80 bands)	Temperate Forest, Austria	RF (object approach)	13	91.7
Dabiri; Lang (2018)	APEX	2.5	VNIR-SWIR (413-2451 nm, 288 bands)	Temperate Forest, Austria	RF	6	90
Marrs; Ni-Meister (2019)	G-LiHT imager	1 m (6 points/m²)	VNIR (418-918 nm, 114 bands)	Temperate Forest, USA	SVM, CN2 rules, ANN	10 and 15	67 and 59
Sothe et al. (2019a)	UAV-FPI	0.11 (35 points/m²)	VNIR (506-819 nm, 25 bands)	Subtropical Forest, Brazil	SVM	12	72.4
Fricker et al. (2019)	NEON AOP	1 m	VNIR-SWIR (280- 2510, 426 bands)	Temperate Forest, USA	CNN	7	87

Note: ANN= Artificial Neural Network; CNN= Convolutional Neural Network; GA= Genetic Algorithm; KNN= K-nearest neighbor; LDA= Linear Discriminant Analysis; MDA= Multiple Discriminant Analysis; MLC= Maximum Likelihood Classifier; MLP= Multilayer Perceptron; PLSDA= Discriminant Analysis based on Partial Least Square; QDA= Quadratic Discriminant Analysis; RDA= Regularized Discriminant Analysis; RF= Random Forest; SVM= Support Vector Machine; SWIR=short-wave infrared.
Source: Author's production.

EQUATIONS AND FORMULAS

a) Equations should appear centered in the text to make them easier to read;

b) If you need to break the equations into more than one line due to lack of space, you must interrupt them before the equal sign or after the plus, minus, multiplication, or division signs;

c) Equations and formulas must be sequentially numbered in Arabic numerals following the chapter number. Example: Equation 1 in Chapter = (3.1)

d) The number must be in parentheses, at the far end of the right margin;

e) If you include equations in appendices and annexes, they must be numbered sequentially according to the letter of the appendix or annex: Equation A.1, Formula A.1; Equation B.2, Formula B.2;

f) The first letter of the words "equation" and "formula" must be in uppercase when included in the text: Equatin, Formula.

Examples:

$$\tau_{i,j} = 2\pi[(I/k_{i,j})^{1/2} + \Delta_{i,j} \qquad (2.1)$$

$$\Delta_{i,j} = R.k_{i,j} \qquad (2.2)$$

$$(x+y)(x-y) = x^2 - xy + xy - y^2 = x^2 - y^2 \qquad (2.3)$$

REFERENCES

Standardized set of descriptive elements that idenfity a spectific document (ABNT NBR-6023, 2018).

The references are the identification of all the documents cited in the text.

If you want to include bibliographic material not mentioned in the text, make another list under the title **Consulted Bibliography**.

To write the references according to ABNT style, access: Writing the references using tha ABNT style.
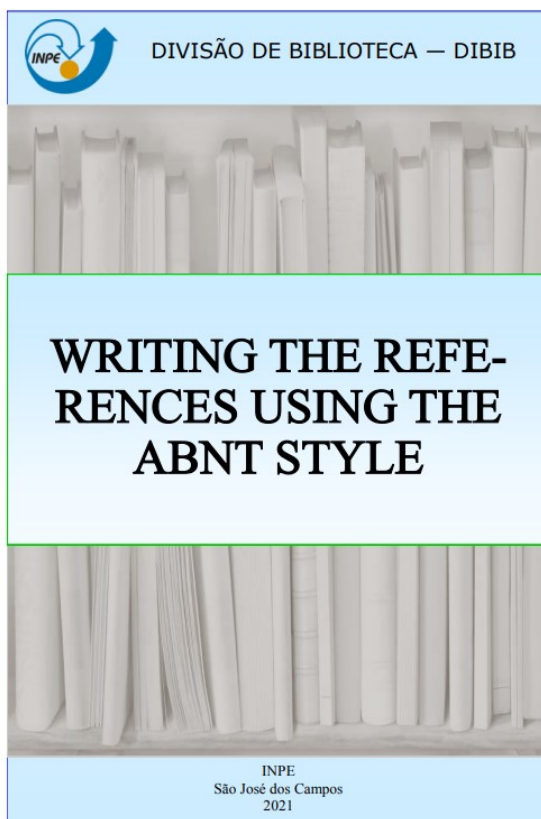
GLOSSARY

A glossary is a vocabulary in which the meaning of words or expressions referring to a particular technical, scientific, etc. specialty is given. The words/expressions must appear in alphabetical order.

GLOSSARY

Blog	A personal website that provides updated headlines and news articles of other sites that are of interest to the user, also may include journal entries, commentaries and recommendations compiled by the user; also written web log, Weblog.
Browser	A browser is an application that provides a way to navigate and interact with all the information on the Web. Technically, a Web browser is a client program that uses the Hypertext Transfer Protocol (HTTP) to make requests of Web servers throughout the Internet on behalf of the browser user.
Channel	In sales and marketing, a communication medium to reach customers with a specialized message. For a distribution channel, a way of selling a company's product either directly or via distributors; possible distribution channels are wholesalers or small retailers or retail chains or direct mailers or your own stores.
Click-through or Clickthru Advertisement	The selection of an advertisement or link on a web page which is counted as part of an effectiveness rating for these features; also written clickthrough.
Google PageRank	PageRank also known as PR is one of the methods Google uses to determine the relevance or importance of a Web page. PageRank is a vote, by all the other web pages on the internet, about how important a web page is. A link to a web page counts as a vote of support. If there are no incoming links to a web page then there is no support.
Microbrowser	A microbrowser is an internet or web browser designed for use on a handheld device such as a PDA or mobile phone. Microbrowsers have small file sizes to accommodate the low memory capacity and low-bandwidth of wireless handheld devices. Essentially it is a stripped-down web browser. Sometimes it is referred to as mobile browser, micro-browser, mini-browser or minibrowser.
Question-and-answer Service	A question-and-answer service is usually an internet-based online information service use by one user to ask a question and a different user or system to answer the question. The system answering the question may be a search engine or database.
QWERTY Keyboard	QWERTY is the modern-day layout of letters on most English language computer and typewriter keyboards. It takes its name from the first six letters shown on the keyboard's top row of letters.

APPENDIX

⇒ Optional element;

⇒ These are texts **created by the thesis author** to complement his/her arguments;

⇒ Appendices are included after the references.

APPENDIX A - BSC FOR IMAGES FOLLOWING THE GAUSSIAN DISTRIBUTION.

Single-look SAR image from homogeneous region, in-phase or quadrature, can follow the Gaussian distribution. Similarly, multi-look SAR images in amplitude from homogeneous region also can follow the Gaussian distribution. Although optical images are not accurately modeled by Gaussian distributions, this is one of the most applied models when dealing with this kind of images.

Let $x = \{\vec{x}_1, ..., \vec{x}_N\}$ be the dataset of unknown samples to be clustered, where $\vec{x}_n$ is represented as a q-tuple of real numbers, q is the dataset number of bands, and N is the number of samples. If $q > 1$, the dataset follows the multivariate Gaussian distribution denoted as $x \sim \mathcal{N}(\vec{\mu}, \Sigma)$, where $\vec{\mu}$ is the vector of q mean values, and Σ is a covariance matrix. The multivariate Gaussian distribution is represented as:

$$f(x; \vec{\mu}, \Sigma) = \frac{1}{(\sqrt{2\pi})^q}|\Sigma|^{-\frac{1}{2}}exp\left(-\frac{1}{2}(x-\vec{\mu})^T\Sigma^{-1}(x-\vec{\mu})\right) \qquad (A.1)$$

where $\mu = \frac{1}{N}\sum_{n=1}^{N}\vec{x}_n$, and $\Sigma = \frac{1}{N}\sum_{n=1}^{N}(\vec{x}_n - \vec{\mu})^T(\vec{x}_n - \vec{\mu})$.

The BSC algorithm can be easily adjusted to handle a dataset following the Gaussian distribution. Three adjusts must be done: The first one concerns on the initial parameter determination, where the PDDP should use the regular PCA approach, while the the EM should use the Gaussian Mixture model; the second adjust is about the stochastic distance, which shall be between two Gaussian distribution; and the third, is regarding the entropy calculation.

A.1 Initial parameter determination

A.1.1 PDDP

The PDDP algorithm is based on the PCA routing, which is a method often used to reduce the dimensionality of large data sets. However, in here it is used to get the principal component, which will be later used to divide the dataset into two subsets.

The PCA computes the mean ($\vec{\mu}$) and covariance matrix (Σ) of a given dataset, and decomposes the covariance matrix into eigenvectors (U) and eigenvalues (Λ) matrices. The q principal components of the observed dataset x are given by the vectors $\vec{\delta} = U^T(\vec{x}_n - \vec{\mu})$.

ANNEX

⇒   Optional element;

⇒   These are documents **created by others** and used by the thesis author;

⇒   Annexes are included aftet the Appendices.

ANNEX A - EVALUATING DISTANCE MEASURES FOR IMAGE TIME SERIES CLUSTERING IN LAND USE AND COVER MONITORING

This annex presents a paper published in the European Conference on Machine Learning and Principles and Pratice of Knowledge Discovery in Databases (ECML PKDD) in Würzburg, Germany on September 2019 (FERREIRA et al., 2019):

Evaluating distance measures for image times series clustering in land use and cover monitoring

Karine Reis Ferreira Lorena Santos and Michelle C. A. Picoli

Brazilian National Institute for Space Research (INPE)
São José dos Campos, São Paulo, Brazil

97

TEMPLATES FOR EDITING YOUR THESIS

In the website: www.inpe.br/biblioteca, it is possible:

⇒ Consult the complete Manual for preparing, formatting, and the submission of theses, dissertations, and other INPE publications;

⇒ Download the templates in MSWord 2010 or LaTeX

For editing and compiling using LaTeX on **Overleaf**, access:

https://www.overleaf.com/latex/templates/modeloinpe-2022/wgvyqymtcnrk

To publish other types of documents, access:

http://urlib.net/rep/8JMKD3MGP3W34P/458MFNP

INPE

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

DIVISÃO DE BIBLIOTECA — DIBIB
MEMÓRIA TÉCNICO-CIENTÍFICA

Av. dos Astronautas, 1.758
Jd. Granja - CEP 12227-010
São José dos Campos - SP
Brasil
Tel: 55 (12) 3208-6923 / 7348

pubtc@inpe.br
simone.delducca@inpe.br

www.inpe.br/biblioteca