Universidade Estadual de Campinas
Instituto de Computação

Greice Cristina Mariano

Visualization of Cyclical Temporal Patterns in Phenology Studies

Visualização de Padrões Temporais Cíclicos em Estudos de Fenologia

CAMPINAS
2018

Greice Cristina Mariano

Visualization of Cyclical Temporal Patterns in Phenology Studies

Visualização de Padrões Temporais Cíclicos em Estudos de Fenologia

Tese apresentada ao Instituto de Computação da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação.

Dissertation presented to the Institute of Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Computer Science.

Supervisor/Orientador: Prof. Dr. Ricardo da Silva Torres
Co-supervisor/Coorientadora: Profa. Dra. Leonor Patrícia Cerdeira Morellato

Este exemplar corresponde à versão final da Tese defendida por Greice Cristina Mariano e orientada pelo Prof. Dr. Ricardo da Silva Torres.

CAMPINAS
2018

Agência(s) de fomento e nº(s) de processo(s): CNPq, 162312/2015-6; CAPES; FAPESP, 2013/501550-0

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

M337v

Mariano, Greice Cristina, 1986-
Visualization of cyclical temporal patterns in phenology studies / Greice Cristina Mariano. – Campinas, SP : [s.n.], 2018.

Orientador: Ricardo da Silva Torres.
Coorientador: Leonor Patrícia Cerdeira Morellato.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Computação.

1. Visualização de informação. 2. Fenologia vegetal. 3. Ritmo visual. I. Torres, Ricardo da Silva, 1977-. II. Morellato, Leonor Patrícia Cerdeira. III. Universidade Estadual de Campinas. Instituto de Computação. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Visualização de padrões temporais cíclicos em estudos de fenologia
Palavras-chave em inglês:
Information visualization
Plant phenology
Visual rhythm
Área de concentração: Ciência da Computação
Titulação: Doutora em Ciência da Computação
Banca examinadora:
Leonor Patrícia Cerdeira Morellato [Coorientador]
Debora Pignatari Drucker
Julio Cesar dos Reis
Thiago Sanna Freire Silva
Hélio Pedrini
Data de defesa: 08-06-2018
Programa de Pós-Graduação: Ciência da Computação

Universidade Estadual de Campinas
Instituto de Computação

Greice Cristina Mariano

Visualization of Cyclical Temporal Patterns in Phenology Studies

Visualização de Padrões Temporais Cíclicos em Estudos de Fenologia

Banca Examinadora:

- Prof. Dra. Leonor Patrícia Cerdeira Morellato
  IC - UNICAMP
- Dra. Debora Pignatari Drucker
  EMBRAPA
- Prof. Dr. Julio Cesar dos Reis
  IC - UNICAMP
- Prof. Dr. Thiago Sanna Freire Silva
  UNESP
- Prof. Dr. Hélio Pedrini
  IC - UNICAMP

A ata da defesa com as respectivas assinaturas dos membros da banca encontra-se no processo de vida acadêmica do aluno.

Campinas, 08 de junho de 2018

"time goes past us, from front to back."
(Lakoff and Johnsen)

Acknowledgements

First of all, my gratitude to God for allowing me to conclude this stage, in the face of so many challenges that have arisen during the PhD.

I thank my supervisor professor Ricardo da S. Torres for the opportunity and for believing that I would be able to carry out this work. I appreciate the friendship, patience, and especially the teachings and for always being so willing to develop any idea or new work.

I also thank professor Patrícia Morellato from UNESP of Rio Claro for once again being my co-supervisor and enabling me to collaborate with the e-phenology project. Thank you for all the support during the research development; for allowing me to travel with your group to the Serra do Cipó, Minas Gerais, and to know how phenology is done in practice; for receiving me in your research laboratory several times and helping me with the development of the work and for accepting to preside over my defense.

I would like to thank all the people and institutions that contributed to the development and completeness of this work. This thesis was only possible thanks to the partnership and the support of these institutions. In this way, I thank the Institute of Computing for the structure and opportunity. To the friends of the laboratory RECOD of the Institute of Computing and the friends of the Phenology Laboratory of UNESP of Rio Claro, thanks for the collaboration and friendship. In particular, I thank Prof. Dr. Sara Diamond from OCAD University, located at Toronto, Canada, for agreeing to be my supervisor during my exchange period and to cheerfully received me into her laboratory, the Visual Analytics Lab (VAL). To the friends I made in Canada, especially Marcus and Hamster, thank you very much for accepting the idea and helping me to develop a conceptual design model for the problem of this work.

I am very grateful to my parents, Édina Kirali Mariano and José Mariano Filho, who never stopped motivating me and encouraging me to study and evolve as a person and professional. During this journey, they suffered some health problems, even so, they always understood my yearnings and gave me freedom so that I could reach them. I know that my mood was often not the best during this period, but they were always there to received me to listen and counsel. They are my greatest source of strength and inspiration. I also thank my brothers Jádson and Gláucia for the partnership and the union we have as a family.

I also thank my roller skates friends, in particular to the friends of the group *Patins Piracicaba*. Thank you for the tours and joyful moments that allowed me to relax and to be prepared for the next steps of the work. Also, thank you to my personal friends Maine, Paulo e Eline, for the patience and for all the assistance during this time.

To Fernando Novello, my gratitude for the company's travel time to Campinas, for the patience and understanding in the moments I have been absent, especially during the six months of my exchange. Thank you for always supporting me and trusting (even more than me) that everything would work out.

This research was supported by FAPESP, the São Paulo Research Foundation (grant #2013/50155-0 FAPESP-Microsoft Research Virtual Institute). We have also been benefited from funds from Emerging Leaders in the Americas Program (ELAP), CNPq (grant #162312/2015-6), Coordination for the Improvement of Higher Education Personnel (CAPES), and FAPESP (grants #2007/59779-6, #2009/54208-6, #2010/51307-0, #2014/12236-1, #2015 /24494-8, #2016/50250-1, and #2017/20945-0).

Resumo

Em diversas aplicações, grandes volumes de dados multidimensionais têm sido gerados continuamente ao longo do tempo. Uma abordagem adequada para lidar com estas coleções consiste no uso de métodos de visualização de informação, a partir dos quais padrões de interesse podem ser identificados, possibilitando o entendimento de fenômenos temporais complexos. De fato, em diversos domínios, o desenvolvimento de ferramentas adequadas para apoiar análises complexas, por exemplo, aquelas baseadas na identificação de padrões de mudanças ou correlações existentes entre múltiplas variáveis ao longo do tempo é de suma importância. Em estudos de fenologia, por exemplo, especialistas observam as mudanças que ocorrem ao longo da vida de plantas e animais e investigam qual é a relação entre essas mudanças com variáveis ambientais. Neste cenário, especialistas em fenologia cada vez mais precisam de ferramentas para, adequadamente, visualizar séries temporais longas, com muitas variáveis e de diferentes tipos (por exemplo, texto e imagem), assim como identificar padrões temporais cíclicos. Embora diversas abordagens tenham sido propostas para visualizar dados que variam ao longo do tempo, muitas não são apropriadas ou aplicáveis para dados de fenologia, porque não são capazes de: (i) lidar com séries temporais longas, com muitas variáveis de diferentes tipos de dados e de uma ou mais dimensões; e (ii) permitir a identificação de padrões temporais cíclicos e *drivers* ambientais associados.

Este trabalho aborda essas questões a partir da proposta de duas novas abordagens para apoiar a análise e visualização de dados temporais multidimensionais. Nossa primeira proposta combina estruturas visuais radiais com ritmos visuais. As estruturas radiais são usadas para fornecer informação contextual sobre fenômenos cíclicos, enquanto que o ritmo visual é usado para sumarizar séries temporais longas em representações compactas. Nós desenvolvemos, avaliamos e validamos nossa proposta com especialistas em fenologia em tarefas relacionadas à visualização de dados de observação direta da fenologia de plantas em nível tanto de indivíduos quanto de espécies. Nós também validamos a proposta usando dados temporais relacionados a imagens obtidas de sistemas de monitoramento de vegetação próxima à superfície. Nossa segunda abordagem é uma nova representação baseada em imagem, chamada *Change Frequency Heatmap* (CFH), usada para codificar mudanças temporais de dados numéricos multivariados. O método calcula histogramas de padrões de mudanças observados em sucessivos instantes de tempo. Nós validamos o uso do CFH a partir da criação de uma ferramenta de caracterização de mudanças no ciclo de vida de plantas de múltiplos indivíduos e espécies ao longo do tempo. Nós demonstramos o potencial do CFH para ajudar na identificação visual de padrões de mudanças temporais complexas, especialmente na identificação de variações entre indivíduos em estudos relacionados à fenologia de plantas.

Abstract

In several applications, large volumes of multidimensional data have been generated continuously over time. One suitable approach for handling those collections in a meaningful way consists in the use of information visualization methods, based on which patterns of interest can be identified, triggering the understanding of complex temporal phenomena. In fact, in several domains, the development of appropriate tools for supporting complex analysis based, for example, on the identification of change patterns in temporal data or existing correlations, over time, among multiple variables, is of paramount importance. In phenology studies, for instance, phenologists observe changes in the development of plants and animals throughout their lives and investigate what is the relationship between these changes with environmental changes. Therefore, phenologists increasingly need tools for visualizing appropriately long-term series with many variables of different data types, as well as for identifying cyclical temporal patterns. Although several approaches have been proposed to visualize data varying over time, most of them are not appropriate or applicable to phenology data, because they are not able: (i) to handle long-term series with many variables of different data types and one or more dimensions and (ii) to support the identification of cyclical temporal patterns and associated environmental drivers.

This work addresses these shortcomings by presenting two new approaches to support the analysis and visualization of multidimensional temporal data. Our first proposal to visualize phenological data combines radial visual structures along with visual rhythms. Radial visual structures are used to provide contextual insights regarding cyclical phenomena, while the visual rhythm encoding is used to summarize long-term time series into compact representations. We developed, evaluated, and validated our proposal with phenology experts using plant phenology direct observational data both at individuals and species levels. Also we validated the proposal using image-related temporal data obtained from near-surface vegetation monitoring systems. Our second approach is a novel image-based representation, named Change Frequency Heatmap (CFH), used to encode temporal changes of multivariate numerical data. The method computes histograms of change patterns observed at successive timestamps. We validated the use of CFHs through the creation of a temporal change characterization tool to support complex plant phenology analysis, concerning the characterization of plant life cycle changes of multiple individuals and species over time. We demonstrated the potential of CFH to support visual identification of complex temporal change patterns, especially to decipher interindividual variations in plant phenology.

List of Figures

Contents

Chapter 1

Introduction

Time series is one of the most common way to represent real-world events based on observations made over time and can be found in several applications and domains. Usually, time-related studies handle a huge amount of collected observations that must be ordered in meaningful arrangements to facilitate the interpretation and speed up analyzes. In addition to the data volume, those studies may also involve multidimensional data and multiple variables observed at the same time. Since the domain experts are interested in identifying temporal changes and patterns, in the case of multidimensional and multivariate temporal data, this analysis becomes a challenging task.

In phenology studies, for instance, phenology experts – also known as phenologists – observe changes in the development of plants and animals throughout their lives and investigate what is the relationship between these changes with environmental conditions. In the case of plant phenology studies, phenologists are interested, for example, in the timing of leafing, flowering, and fruiting of plant species and the relation to its ecological interactions [82, 95, 122]. Besides that, recent plant phenology studies are not only concerned with the analysis of data from direct plant observations in the field [85] (usually known as on-the-ground phenology), but also consider near-remote data derived from sequential vegetation images taken by digital cameras [4, 5] (usually referred to as near-surface phenology).

One suitable approach for handling those collections in a meaningful way consists in the use of data visualization methods, based on which patterns of interest can be identified, triggering the understanding of complex temporal phenomena. The term data visualization refers to methods used to represent data and communicate the significance information using sophisticated visual ways, for example, graphics with interactive capabilities. According to Telea [118], the data visualization field has three main subfields: (i) scientific visualization, also known as scivis, (ii) information visualization, also known as infovis, and (iii) visual analytics. In general, according to [27, 30, 118], the differences between them are:

- **Scientific Visualization:** uses interactive visualization techniques to communicate scientific information or phenomena to be understood, mainly, by researchers; and deals with spatial data or data with geometry structure well defined;

- **Information Visualization:** uses interactive visualization techniques to communicate information and results for an easily understanding; deals with abstract and non-spatial data;

- **Visual Analytics:** is a consequence of information and scientific visualization subfields and combines visualization methods with data analysis and data mining, dealing with large and complex datasets.

In the case of phenology studies, phenologists increasingly need tools for visualizing appropriately long-term series with many variables of different data types, as well as to identify cyclical temporal patterns. Although there are many temporal visualization techniques, such as the ones summarized in the book *Visualization of Time-Oriented Data* [3], and techniques for visualizing multidimensional and multivariate data [128] for large datasets [94]; visualizing cyclic temporal data with multiple variables and multidimensional ones, as in the case of several applications (e.g., phenology studies), is still a challenge. This is mainly due to three factors:

- (i) the volume of data to be analyzed (characterized by long time series);
- (ii) the heterogeneity of the data (which are obtained from different sources – field collection, sensors, cameras, etc.), and
- (iii) the heterogeneity in terms of the time scales of the series (e.g., daily, monthly, or yearly).

This work addresses these shortcomings by presenting two new visual analytics approaches to support the analysis and visualization of multidimensional and multivariate temporal data. We used the phenology as the target domain to specify, implement, and validate information visualization techniques to facilitate the recognition of cyclical temporal patterns. For this purpose, we used as a case study the e-phenology Project,[1] whose data refer to plant observations in the field – both at individuals and species levels – and sequences of vegetation remote sensing images. Our first proposal combines radial structures along with visual rhythms. Radial visual structures are used to provide contextual insights regarding cyclical phenomena, while the visual rhythm encoding is used to summarize long-term time series into compact representations. We also validate this proposal using image-related temporal data obtained from near-surface vegetation monitoring systems. Our second approach is a novel image-based representation, named as Change Frequency Heatmap (CFH), used to encode temporal changes of multivariate numerical data. The CFH method computes histograms of change patterns observed at successive timestamps.

This chapter presents in Section 1.1 some scenarios that motivate this work. In Section 1.2, in turn, we discuss the hypotheses and the main research questions addressed. Finally, Section 1.3 outlines the organization of this document.

[1] `www.recod.ic.unicamp.br/ephenology` (As of July 25, 2018).

1.1 Motivating Scenarios

Environmental changes have become an important issue on the global agenda. In order to support the formulation of policies for environmental management and to maintain the balance of the ecosystem, it is necessary to have an accurate view of the existing conditions and to understand the complex changes that occur at all levels of the planet [120].

An essential step to foster studies about those changes is to collect relevant data on the environment and living beings and to develop computer systems to manage these data and to support complex data analysis targeting knowledge discovery. These systems must also combine newly collected data with historical information and legacy data (for example, from different data types and encoded in different file formats), using a unified management perspective. Therefore, scientists involved with environmental issues should seek support from a large set of systems, a task typically associated with several interoperability issues, due to system incompatibility, data diversity, and a variety of user profiles [120].

In this section, we present the main motivating scenarios related to the use of multidimensional and multivariate temporal data and its visualization in the context of phenology studies. The first scenario refers to the questions and importance of phenology studies as a whole (Section 1.1.1). The second scenario refers to our study case, the e-phenology project and its demands for data analysis tools (Section 1.1.2).

1.1.1 Phenology studies

A representative example of the problems related to the development of systems to support complex data analysis arises in the context of phenology studies. Phenology is a traditional science that observes the development of plants and animals and their effects in relation to the environmental changes [108]. Currently, phenology is a simple, and the most reliable indicator of the effects of climate change on life cycles of plants and animals [97, 107, 124, 125]. The main interest of phenologists is the timing of biological events, such as, flowering, migration, reproduction, hibernation, etc., in relation to changes in season and climate. In plant penology study, for example, the different dates when plants bloom and have fruits can indicate that changes in the seasons occurred.

The data collection for plant phenology studies can be performed using different methods, as follows:

- **On-the-ground phenology:** refers to the direct observations made by experts in the field over time and can be quantified in terms of observed phenological patterns of individuals, populations, and comunities.

- **Remote sensing land surface phenology:** refers to the use of satellites imagery to track phenological events (seasonal changes) in vegetation indexes at regional to global scales [130]. Remote sensing land surface phenology supports, for example, the identification of broad-scale phenological trends that would be difficult or impossible using only on-the-ground observations, such as assessing crop conditions, drought severity, and wildfire risks,[2] among others. Many different methods have been developed to determine phenological patterns using remote sensing phenology and most of them use time series of vegetation index (VI) extract from images, such as normalized difference vegetation index (NDVI) [1] and green leaf area index (LAI). These indices can be used, for example, to detect the vegetation greenup and senescence [15, 116, 130].

- **Near-surface remote phenology:** refers to the use of repeated potographs taken by digital cameras set up near to the ground to document plant changes. Due to the high frequency collection (hour-daily photos) allows phenologists to refine the temporal gaps of the remote sensing and traditional on-the-ground phenology methods [4, 5].

In the case of traditional on-the-ground phenology studies (main focus of this thesis), the phenologists usually monitor several marked individuals – from the same or different species – periodically at well defined intervals [83, 85], and some measurements include distribution statistics (e.g., mean, range, and variance) of determined parameters such as, time of occurrence, duration, intensity or amplitude and synchrony [90]. The measurements can also be quantified (e.g., number of reproductive structures) [16, 80, 100]. These parameters are independently calculated for each sampled individual or for population or species, as usually done in community studies, therefore, their analysis demands extensive time and special attention [71].

Typical studies related to plant phenology studies involve the analysis of :

- all phenophases of individuals from the same species over time;
- a single phenophase of individuals from the same species but from different sites;
- a single individual with its all phenophases over time (e.g., years).
- a single phenophase of an individual over time;
- the correlation of the variation of phenology changes of species or individuals over time with climate data; or
- the variation of leaf color from species or individuals over time, based on vegetation images.

Furthermore, in phenology studies, experts handle well-defined expected temporal patterns both at individual level, as well as species and community levels, such as *when* flower buds, flowers, immature and mature fruits or new leaves began to appear; *how long* a particular phenomenon lasts, and *how is the pattern* observed for individuals of the same species located within a same area or in different regions [72].[3]

A critical problem faced relies on the lack of appropriate tools to support these analyses. Although there are many visualization techniques for exploring time series, there

[2] `https://phenology.cr.usgs.gov/overview.php` (As of July 25, 2018).

[3] These are some examples of patterns explored in the context of this thesis, for specific species. The investigation of each possible patterns can be used for tropical vegetation is left for future work.

is no standard tool with which phenology experts may explore their data. Furthermore, comparative studies involving multiple variables and dimensions are even more difficult to be performed [81].

For instance, to investigate the climate drivers for plant phenology changes, experts often have to produce a graphic joint visualization that integrates both, phenology and climate data. Some examples of typical graphics used in phenology studies can be found in [4, 25, 114]. In summary, phenology experts spend much time to organize useful data and to create appropriate graphical representations that might be used to identify existing temporal relations. According to plant phenology experts, generally, the main problems related to the most commonly used data visualization approaches are:

1. There is no standard approach or tool with which they can visualize and analyze temporal data.

2. The graphics used are static and commonly do not support multiple variables and/or multidimensional analysis.

3. The graphics used are created by using more than one software in several cases, and several times may combine manually using graph editors to reach the visualization goal. This is a costly and time-consuming task as it demands a lot of efforts and training in many software at the same time.

In order to properly perform those analyses, phenology experts need access to tools to support the analysis of multidimensional temporal data. Several initiatives and collaborations between phenologists and computer science professionals have emerged in recent years in an attempt to create more robust applications and integrating the use of new technologies to monitor phenology remotely. In Brazil, for example, some efforts in that direction have been performed in the e-phenology Project. In this context, this work proposes visualization techniques to be used in such tools and uses the e-phenology Project demands to define motivating case studies for the proposed approaches.

1.1.2 The e-phenology Project

Overview

Studies on plant phenology enable researchers to analyze the conditions of plant life cycles in relation to their environment and climate changes [101]. In this way, it is possible to evaluate how climate changes affect the phenology of several species, and as consequence, biodiversity [78, 79]. Plant phenology studies involve periodic surveys, in which observations of species and climate are made over days, months, years, and even decades. Plant observations are made based on the life cycles of the plant, called phenophases [117] and depend on the experience of the specialist who goes to the field to carry out the study.

In order to assist phenology specialists and minimize the errors found in field data sets, in 2011 an innovative project called e-phenology was created. This is a multidisciplinary project developed by the phenologists team from Laboratory of Phenology of the

Universidade Estadual Paulista (UNESP), located in Rio Claro, São Paulo State, Brazil together with a team of computer science professionals from the Institute of Computing of the University of Campinas (UNICAMP), Campinas, São Paulo State, Brazil.

This project uses new technologies to integrate the traditional field plant phenology observations with near-surface remote sensing, and with climatic data obtained by meteorological sensors. The main objective is to provide information about the behavior of vegetation in tropical regions – as in the case of vegetation in most of Brazil and South America – to detect and understand the environmental changes and the warm global effects on plant species and biomes [78].

During the first stage of development (from August 2011 to August 2013), the project aimed to apply the use of digital cameras to carry out remote phenological monitoring, install climate sensors to obtain local data and to develop a database to store field phenology collection information (previously stored in electronic spreadsheets) and also new data produced by the use of new installed technologies [71]. From 2013 on, the focus has been to address theoretical and practical problems involving the combination of two remote phenology monitoring systems: digital and hyperspectral cameras at three scales: on-the-ground, phenology tower, and near-space using Unmanned Aerial Vehicle (UAV). Since the beginning, large volumes of data have been generated by the project and new methodologies are demanded to be developed to perform the visualization and interpretation of temporal data.

On-the-ground versus near-surface phenology

An important kind of phenological data handled comes from direct on-the-ground observations of plants' phenophases (illustrated in Figure 1.1). These observations refer to biological events of plants, such as flowering, fruiting and leafing, and can be collected daily, fortnightly, or monthly, over years [4, 25, 123]. One example of typical phenophases observed in the context of the e-phenology are shown in the Figure 1.1. For each observed phenophase, phenologists assign scores representing its intensity: 0 (absence), 1 (between 1 to 50% of low activity) and 2 (above 50% up to 100% or peak of activity).

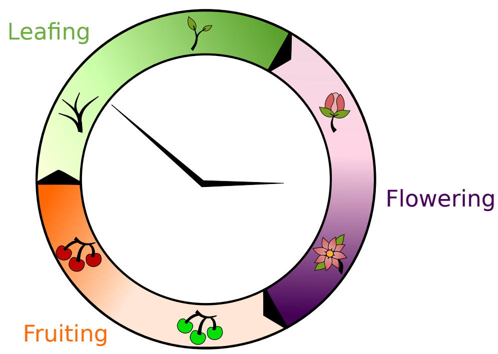
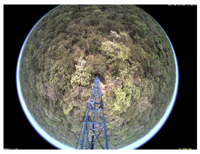


Figure 1.1: Typical phenophases considered in plant phenology studies

Near-surface phenology refers to remote phenology observations using images. For example, data are obtained by digital cameras installed at observation towers in study sites of interest to phenologists. The cameras are programmed to take images throughout the day in a scheduled time interval, during all year.

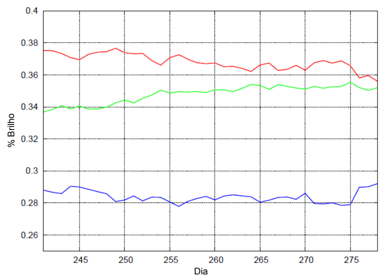
The first tower, relative to the first case study of the project, is presented in Figure 1.2(a). The tower is installed in an area of Cerrado, located in the city of Itirapina, State of São Paulo, Brazil. The tower is 18 meters high and has an arm positioned to the side where the digital camera is installed. Currently, the project has another six cameras installed in different sites with different kinds of vegetation in Brazil: Atlantic Rain Forest, Cerrado PEG, Amazon Forest, Caatinga, EE Itirapina and Serra Cipó. Besides that, the Serra Cipó site presents a set of towers arranged in different areas of this region.[4] In the case of Itirapina, the installed digital camera was programmed to take five pictures in each full hour of the day, from six in the morning to six in the afternoon. These pictures are similar to the image presented in Figure 1.2(b), and due to the variety of photos throughout the day, it has been possible to process them to obtain the temporal variation of characteristics of Cerrado plants, based on color and texture visual properties [7, 8].

(a) *Tower and digital camera installed at Cerrado, Itirapina-SP.*



(b) *Image taken in Itirapina-SP.*

(c) *Mask for Caryocar brasiliensis species.*



(d) *Image processing for Caryocar brasiliensis species*

Figure 1.2: View of tower, digital camera and images taken in a cerrado vegetation area: (a) Tower and camera installed in the Cerrado field in Itirapina-SP, (b) Example of typical image of the camera in Itirapina-SP, (c) Mask of *Caryocar brasiliensis* species applied to the image, and (d) Resulting time series obtained by using the mask of species *Caryocar brasiliensis* for images at different time stamps. This chart presents the channel Red, Green, and Blue (RGB) over the course of days [8].

Figure 1.2(c) presents the mask for the species *Caryocar brasiliensis* and Figure 1.2(d) presents an example of images processing using that mask. It was used color as main feature. The color channel patterns RGB was extracted from each pixel for each region of the mask.

[4]The phenocam network can be found at `www.recod.ic.unicamp.br/ephenology/client/index.html#/phenocamNetwork` (As of July 25, 2018).

In addition to the camera, a weather station was also installed in the same tower. Nine sensors were installed at this station. They measure soil moisture, rainfall, wind speed, and solar radiation. Data obtained from these sensors contribute to assess the impact of the climatic changes on the plant phenology, as well as to compare with the changes identified when processing the images in the vegetation of the Cerrado.

However, there is still a lack of methodologies to correlate and visualize time series related to data from phenology observations made in the field and climatic data with the pixel-based indices obtained from vegetation images.

Architectural View

Figure 1.3 shows the simplified architecture defined to the e-phenology information system and highlight in orange the components involved in the development of this work. Three main activities should be supported by this architecture: (i) data insertion, (ii) query processing, and (iii) data visualization. Research activities related to data insertion and query processing were covered in a previous work [71] with the specification and implementation of a database for the e-phenology project. The database was designed to support different types of data, such as: data about the location, taxonomy, and ecological information about individuals and species, and data collected periodically, as in the case of the on-the-ground observations (phenology data); global and local climate changes collected by weather sensors (weather data), and information and features extracted from images (image data).

Regardless the type of activity (data insertion, data query, or data visualization) user wants to perform, the system organization is the same. First, the user interacts with a interface, where one defines which activity will be performed. Later, the system processes the defined activity. For example, if the user wants to perform a data insertion, first one is required to inform the data to be recorded in the interface and the system will process and store the data into the database (arrow A). Suppose now that the user wants to retrieve data of interest. Using the model specified in the previous work [71], the first thing the user must do is to define a query pattern through the interface and the query processing module will perform the required tasks to fetch the data of interest from the database. The retrieved records will then be displayed in the preview interface. These fluxes are illustrated by arrows B, C, and D, respectively.

Since complex temporal data analysis is difficult to be performed through visualization approaches based on tables or simple graphs when there are a lot of variables involved, this work is concerned with the development of the third activity: *visualization* and its related components. The objective is to provide visualization mechanisms and tools for two types of data supported: time series related to textual data (phenology and weather) and image data. The main objective is to transform the conventional data stored into the database in an interactive visualization. To do that, we are working with the concepts of information visualization process, where the system needs to transform the data resulting from the original query to data tables that can be presented through visual structures. We refer to this module as *data processing*, and it is discussed in Chapter 5.

This thesis intend to specify and create adequate information visualization techniques to represent the scientific data of the e-phenology project. Also, we expect that changes on data on species and phenology obtained from vegetation images can be correlated with patterns found on data related to in-the-field observations.

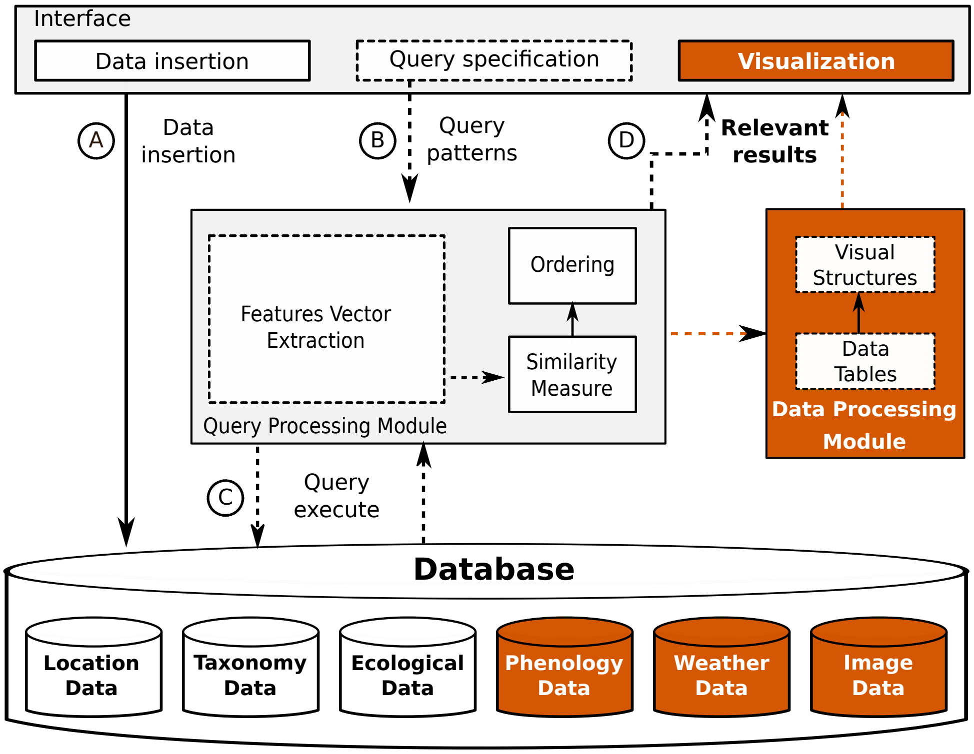


Figure 1.3: Architecture schematic view of the e-phenology information system.

Handled Data types

The e-phenology database integrates different data types, classified as temporal and non-temporal data. As it can be observed in Figure 1.3, non-temporal data include: *location* data, which describe sites where phenological studies are conducted; and *taxonomy* and *ecological* data, which classify species and individuals. These non-temporal data are out of the scope of this work.

Temporal data, in turn, include: *phenology* data associated with on-the-ground observations; *weather* data, which contains data collected by metereological sensors; and *image-related* data, which are collected by the near-surface remote phenology process. Properly visualizing these data will be the main focus of this work.

Figure 1.4 shows examples of time series managed in the context of e-phenology project. The highlighted area refers to a time series with one phenophase varying over time: *flower bud*. In this case, we refer to this data as a univariate time series. In the case of analyzing all the three variables (*flower bud*, *flower*, and *leaf flush*) varying over time, we refer to this data as a multivariate time series. In both cases, the set of phenophases compose a set of variables V, which values are integer numbers (0, 1, or 2) varying over time instants t_i in T, where T is a set of timestamps.

Climate data, in turns, comes from weather sensors installed in the site where the phenology study is performed. The main objective is to monitor the local environment

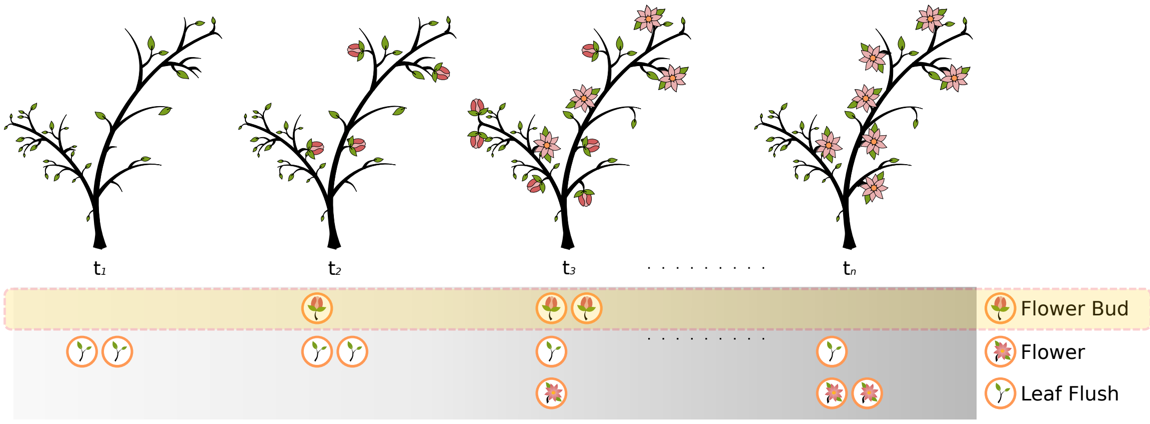


Figure 1.4: Example of time series using on-the-ground observations with three phenophases: *flower bud*, *flower*, *leaf flush*. For each phenophase, the values 0, 1, or 2 are presented by the icons and they are associated with each time instant.

information about amount of rainfall, wind speed, solar radiation, and relative humidity. The data typically are collected daily in predefined time intervals. In the case of e-phenology, sensor data are collected every five minutes, during 24 hours a day, for every day of year.

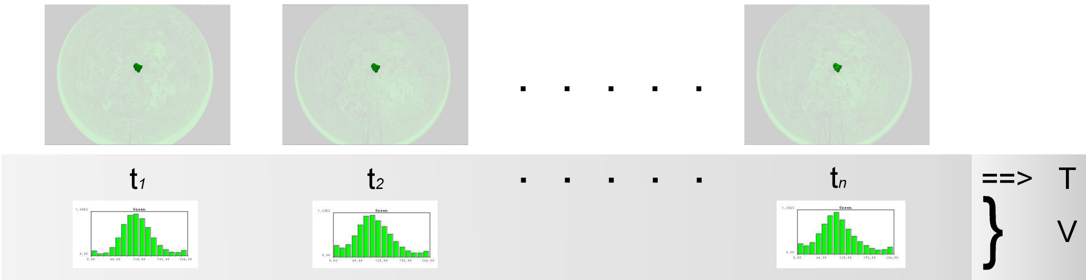
Finally, image data are obtained by processing sequential vegetation images taken by digital cameras [4]. Vegetation indices are extracted from the images taken by digital cameras periodically. One usual example refers to the detection of Red, Green, and Blue (RGB) channels (see Figure 1.5(b)). From the green channel, we can extract an index of green (e.g., %Green or Gcc[5]), that corresponds to the leaf flushing, and verify variations in the phenophases for individuals or the community over time [4].

Figure 1.5 shows two examples of multidimensional time series. The first example uses a multidimensional time series with one variable, defined as the histogram of Gcc scores for different timestamps. The second example encodes a multivariate multidimensional time series. In this case, three histograms (one for Gcc, one for Bcc, and another for Rcc) for each timestamp are illustrated.

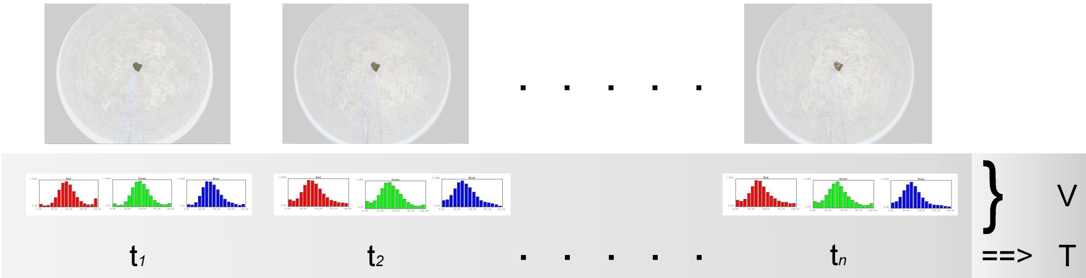
An important feature of the e-phenology project is the acquisition of temporal data with different time scales. For example, in our tropical core study area, conventional phenology observation data are usually collected once a month; the phenological indices are derived from images taken by digital cameras every hour, on a daily basis; and the weather data are obtained from different types of meteorological sensors every ten minutes, daily.

In summary, both traditional direct on-the-ground observations and the near-surface data acquisition procedures have resulted on large volumes of multidimensional temporal data, whose analysis demands the development of appropriate visualization tools.

[5]Similar reasoning may be employed for the %Blue (Bcc) and %Red (Rcc). For more details, the reader may refer to [4, 104].



(a) *Example of univariate multidimensional time series. For each time instant* t_i*, there is a histogram of the green values observed in the target region (in the middle of the circle).*



(b) *Example of multivariate multidimensional time series. For each time instant* t_i*, there are three histograms (for the R, G, and B color channels) observed in the target region.*

Figure 1.5: Example of multidimensional time series with one (a) and more (b) variables.

1.2 Hypothesis & Research Questions

This work focuses on how to visualize multivariate time series related to phenology studies in order to support phenology experts in their tasks related to data analyzes. Since we have different scenarios of visualizations, our proposition is to investigate which information visualization techniques could be applied and combined to visualize multivariate temporal data aiming to support the identification of cyclical temporal patterns; and, at the same time, could be used to develop an information visualization system to support phenology experts.

In summary, we worked in two propositions: (i) the use of radial structures combined with visual rhythm approaches to visualize both numerical and image-related temporal data and (ii) the use of a motion encoding technique combined with heatmaps to identify and compare phenophase changes in the life cycles of individuals.

Our main hypothesis, therefore, can be enunciated as follows:

> ***Visual structures combined with data summarization approaches are effective to create a scientific visualization system to identify cyclical patterns in multivariate multidimensional temporal data and potentially improve phenology analyzes.***

Given this hypothesis, the following research questions are addressed in this work:

1. *Which visual structures can display time series satisfactorily covering the different types of data and multiple variables involved in phenology analyzes? Which interaction mechanisms should be supported?*

2. *Which data summarization approach can be used to visualize long-term series in a compact way? How to encode both numeric and image-related multivariate temporal data in a single visual structure to make them comparable? Can visual rhythms be used for that?*

3. *Which visualization technique can be used to encode cyclical temporal changes? Which data summarization approach can be used to encode those changes?*

4. *Which phenology research questions/activities can benefit from the developed visual structures? How to validate developed data structures with phenology experts?*

1.3 Text Organization

In Chapter 2, we present the fundamental concepts related to time series visualization, and phenology analysis. We also present an overview about existing research challenges in the phenology visualization area. Next, in Chapter 3, we present our first proposed information visualization technique, which combines a radial structure with a visual rhythm representation to visualize numerical data related to both on-the-ground observations and vegetation images, both in the context of phenology studies. In Chapter 4, we describe the Change Frequency Heatmap method, proposed to visualize changes in the life cycles of different individuals for the same species. In Chapter 5, we present some mockups and details regarding the partial implementation of an information visualization tool, which exploits the proposed techniques in this work. Finally, we concludes this work in Chapter 6, presenting a summary of the main contributions of this work and possible future work.

Chapter 2

Basic Concepts & Related Work

The development of this thesis required research mainly in three different areas: information visualization, computer vision for visual analytic, and phenology. In this chapter, we present the main concepts, used in the context of this thesis, related to these areas. We discuss the main research venues concerning the use and validation of information visualization techniques (Section 2.1), and visual analytic approaches to support the identification and analysis of changes associated with multiple variables observed over time (Section 2.2). We also discuss the use of visualization approaches in the context of phenology studies (Section 2.3).

2.1 Information Visualization

Information visualization is a term often used by the scientific community to refer to the process of transforming data to knowledge through a visual representation, which can be understood given the natural visual capabilities of humans beings [27, 47]. Its main objective is to foster knowledge extraction based on the analysis of a dataset [111] using visual displays to communicating information in a meaningful way [75]. The process combines aspects of Computer Graphics, Human-Computer Interfaces, and Data Mining to represent data into a visual form [46] with human perceptual and cognitive capabilities to enable the user to observe the information [47].

The benefits of its applications are more clearly perceived in the context of the management of large volumes of data. In those scenarios, when a visualization approach is used, data interpretation and analysis may become faster and more efficient. Furthermore, hidden features, patterns, or trends may be evidenced, leading to insights about complex phenomena [47, 126]. Handling large volumes of data has demanded the creation of novel approaches for information visualization [47, 111]. In this case, one key aspect in the visualization process consists in the definition of which technique should be employed for a specific target application, which depends on the kind of data being manipulated and the user tasks, which are expected to be performed using them [46]. Some authors categorize such techniques, depending on the objectives they are devised for, in the support of the decision-making process.

According to [110, 128], existing techniques may be divided into categories depending on: (1) the kind of data (unidimensional, multidimensional, temporal, hierarchical or graph-based); or (2) tasks supported for in-depth data analysis: *overview*,[1] *zoom*, filtering, detailing, identification of relationships, awareness of contextual information, and information extraction. The visualization techniques, therefore, should use resources that allow the data interaction and representation graphically, in such a way that they facilitate and even optimize data analysis and information comprehension. Information visualization techniques have been used in many different domains, such as, distance learning, diagnosis based on medical data analysis, and even in initiatives related to fraud detection [111]. In engineering and natural science studies, for example, the main challenge is to visualize time series in a meaningful way [69], i.e., the visualization must be effective and support the identification of temporal patterns, making easily the understanding of complex temporal phenomena.

For this thesis, the main related work are those associated with time series visualization (Section 2.2) and information visualization techniques developed in the context of phenology studies (Section 2.3). They are discussed in more details in the next sections.

2.1.1 Validation in Information Visualization

According to Card et al. [27], information visualization aims to provide tools and useful techniques to detect patterns and to support comprehension about datasets, or broadly, to increase cognition. As the information visualization area is influenced by different domains, such as psychology, semiotics, computer graphics, and arts, it becomes more difficult to make evaluations with quantitative studies, as in the case of user evaluations [132].

In this context, visualization evaluation should not only be concerned with the visual aspects or usability of visualizations techniques, but also with the processes for which the tool or technique was developed [66]. Various challenges faced when planning, conducting, and executing the evaluation of a visualization tool [66] have been investigated. According to Plaisant [98], for example, to carry out information visualization evaluations, the researchers need to choose and prepare compelling examples using real data and within the context of target users. This will allow another potential users, with similar data, to feel more confident to use the implemented visualization technique.

In 2000, a special edition of *International Journal of Human-Computer Studies* [31] called attention to the relevance of making assessments of information visualization techniques with users. Since then, many papers have been developing specific methods of evaluations to the area [29, 43, 115]. According to a research by Lam [66], most of the published papers (at three major conferences and one important journal of the visualization area) do not include any kind of evaluation, not even case studies.

In order to consolidate a visualization tool for multidimensional and multivariate temporal data, this thesis uses validation methods centered on potential users and their expertise. We consider in our validations not only aspects of computer-human interface, but also subjective aspects of information visualization. For example, we consider the

[1]The *overview* approach allows the observation of the whole object collection, allowing for a more detailed view based on the use of *zoom* operations [110].

most common tasks involved in phenology studies to generate data representation and to validate the use of the proposed visualization approaches based on these tasks. In summary, the proposed methods were validated in the context of phenology studies with phenology experts.

2.2 Time Series Visualization

A time series is an ordered sequence of observations collected at periodic intervals and can be taken for any variable that changes over time, for example, meteorological, social, financial, and medical data. The area concerning the proposal of methods aiming to determine correlations among different time series and/or understanding complex phenomena over time is commonly referred to as time series analysis. Time series analysis always involves the use of statistical methods to analyze one or more time series whose data vary over the same period of time. The objective of time series analysis is to extract meaningful statistics, patterns, and trends about the data collected over time [2]. According to Frank (1998) and Goralwalla et al. (1998) [45, 50], there are many aspects to be considered when we want to analyze time series. In general, these aspects are related to two main challenges: (i) how to model time and the associated data and (ii) how to represent and visualize temporal data graphically, so that insights can be drawn from the graphical representation.

This section introduces basic concepts related to time series visualization (Section 2.2.1), as well as discusses some of the existing initiatives in the context of temporal data representations (Section 2.2.2).

2.2.1 Background

Given a set of values defining a set of time instants T, where $T = \{t_1, t_2, \ldots, t_n\}$, one time series S can be represented in two ways: (i) using a variable V whose values vary according to each time instant in T, $S = \{V(t), \mid t \in T\}$; or (ii) using the notation v_i to indicate each value associated with each time instant in T, $S = \{v_{t_1}, v_{t_2}, \ldots, v_{t_n}\}$. In both representations, the values varying over time can be from any domain, for instance, may belong to the set of real or integer numbers or even be a textual data.

A simple example of time series is shown in Figure 2.1. The values that vary over time from a variable V are represented in the y axis and are associated with the time instants in the x axis. In the case of example, variable V represents the leaf flush phenophase and belongs to the domain of the integer numbers. The values varying over time T represents the amount of individuals trees of the species *Aspidosperma Tomentosum* which presented this phenophase from Jan. 2011 to Feb. 2012.

Time series may be associated with one single variable (in this case, known as univariate time series) or multiple variables (multivariate time series). In another taxonomy, time series data may be divided into two categories depending on whether it refers to a variable with either one single dimension (unidimensional) or multiple dimensions (multidimensional time series).

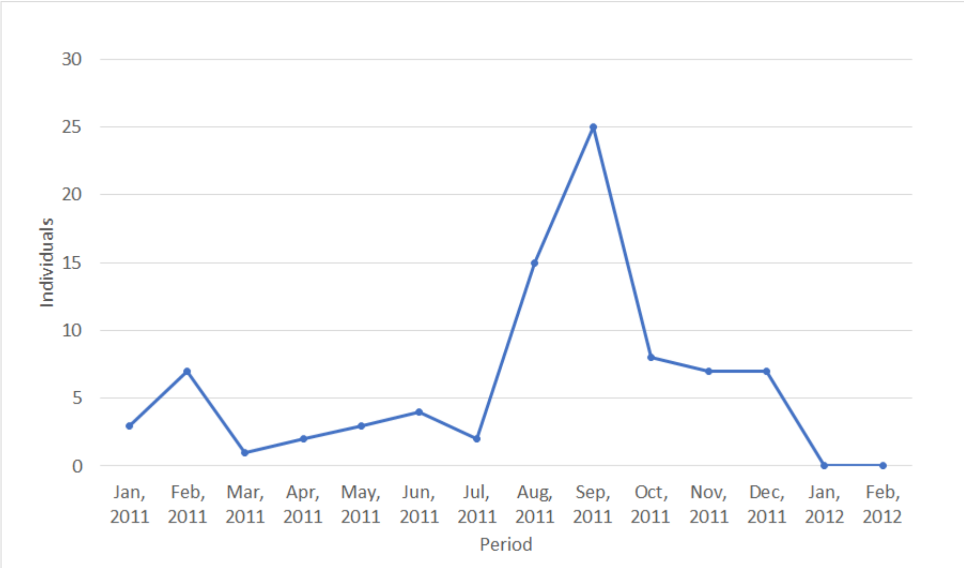


Figure 2.1: Time series example. For each value v_i in y axis, there is a time instant t_i associated. The values varying over time T represents the number of individuals of the tree species *Aspidosperma tomentosum* presenting this phenophase from January 2011 (Jan 2011) to February 2012 (Feb 2012).

- **Multivariate:** Let $\mathcal{V} = \{V_1, V_2, \ldots, V_m\}$ be a set with m variables. Each variable V_i varies over time, i.e., there is a set with n ordered values for each variable V_i: $V_1 = \{v_{1_1}, v_{1_2}, \ldots, v_{1_n}\}$; $V_2 = \{v_{2_1}, v_{2_2}, \ldots, v_{2_n}\}$; $\ldots$; $V_m = \{v_{m_1}, v_{m_2}, \ldots, v_{m_n}\}$. Thus, there are m combinations with n values of variables and for whose there is a time instant t associated, forming the time series $\mathcal{V} = \{V_1(t), V_2(t), \ldots, V_m(t) \mid t \in T\}$. When $m = 1$, we have a univariate time series.

- **Multidimensional:** Let $V \in \mathbb{R}^n$ be a variable which is represented by a point in the $\mathbb{R}^n$ space, i.e., V can be seen as a vector containing n different dimensions. If variable V varies over time, we define a multidimensional time series. If multiple multidimensional variables vary over time, the associated time series is said to be **multidimensional** and **multivariate**.

There are several approaches to visualize time series as shown in the seminal book written by Aigner et al. [3]. Most of those initiatives, however, have limitations to visualize long time series of different types and domains, i.e., they are not robust enough to visualize multidimensional temporal data. In this sense, researchers have developed visualization techniques for specific purposes. In [14, 41, 56, 67], for example, we found different biodiversity systems with some features to visualize data varying over time in their context. None of them combines radial structures with visual rhythm representations at the same time. The work [67], however, takes advantage of visual rhythms and special color mapping strategies to depict changes of image-related phenological features over time.

2.2.2 Temporal Multivariate Data Representations

The analysis of temporal data attained increasing importance in various domains, such as biology, engineering, and medical sciences. The main goal is to detect trends and patterns for gain insights and understanding the behavior of data over time [2]. Besides the variation of the data in the time, another important aspect for many researchers is the correlation and analysis of multivariate data, which consider the analysis of many attributes (variables) simultaneously.

In this context, many techniques to visualize multivariate data were proposed, such as Parallel Coordinates [58] and its radial version Star Coordinates [60]), pixel-oriented techniques, such as Recursive Pattern [63] and Pixel Bar Charts [62], among others as shown in the survey by Kehrer and Hauser [61]. None of these techniques consider the time aspect from multivariate data in an explicit way. Müller and Schumman [87] present an overview of visualization techniques for multivariate temporal data, such as *ThemeRiver*, *Spiral Graph*, and *Calendar View*, and so on. All of them are techniques used to visualize how attributes change over time, but not the correlations between variables. Many other visualization techniques have been proposed and published for temporal data (see [3] for more details), but each one of them is applied in a specific context.

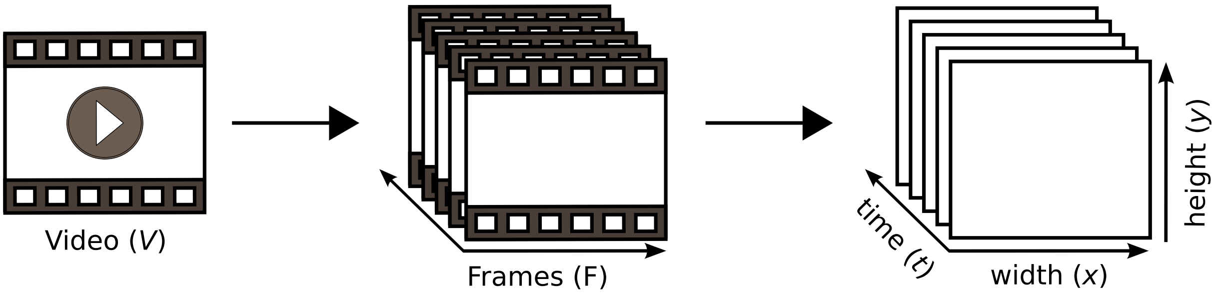
In more recent studies, we can find techniques that explore the correlation of multivariable and time-varying data aspects using cluster algorithms [129], graph analysis techniques [51], similarity algorithms [17], and dimensionality reduction approaches [59]. In our case, we want to detect correlations between variables and to identify change patterns over time, not only for a specific timestamp. Therefore we can not cluster samples or use similarity algorithms, because we want to know how is the behavior of a time series both at each time instant and for all time intervals.

Different from all previous initiatives, our method is validated in the context of phenological studies using plant phenology data. Although there are different biodiversity systems with features to visualize data varying over time [3], our methods are innovative ways to detect temporal patterns from a individual and populational perspective (intraspecific), a very important and still under explored aspect in phenological research [16, 82, 90].

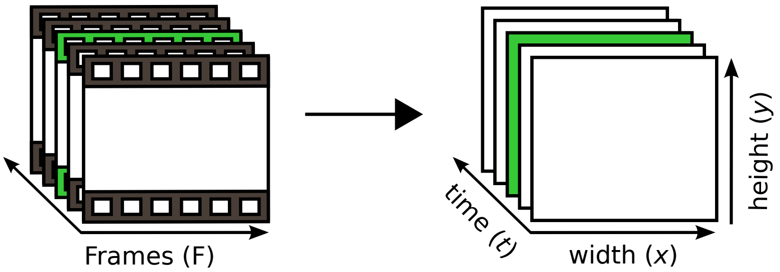
2.2.3 Visual Rhythm

In general, a video can be physically viewed as a set of images (usually referred to as frames) sequentially ordered over time [91]. Each image that composes a video can be sampled using different criteria (e.g., based on a vertical, horizontal, or a diagonal line). Features may be extracted from those samples and later combined into an image in order to encode video changes over time. The image generated by this methodology is called Visual Rhythm [32, 91]. The main objective is to facilitate video analysis by reducing the storage needs associated with features extracted from images and speeding up video processing algorithms [52].

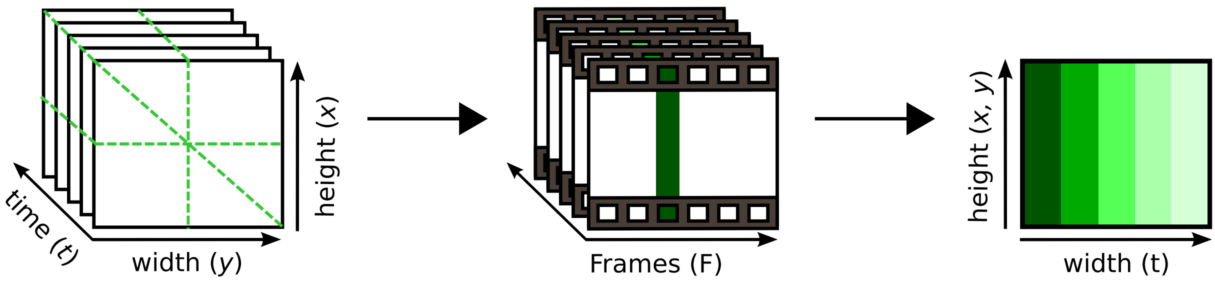
Formally, a visual rhythm is the simplification of one video V at domain 2D + t to domain 1D + t, as presented in Figure 2.2. Each video V has T frames, where each frame F has the same height and width, therefore $V = \{F_t\}$, $t \in [1, T]$ (Figure 2.2(a)). Each frame F_t is characterized by a set of pixels – samples – (x, y) associated with a time instant t, then, $F_t = \{(x,\, y,\, t)\}$. F_t is also known as a video cut [32] (Figure 2.2(b)).



(a) *Representation of a video in the 2D + t format. In this format, a video is seen as a set of images (frames) with the same width and height. These frames are sequentially ordered over time.*



(b) *Representation of a video cut on a 2D + t video. A cut represents an image at a timestamp t.*



(c) *Representation of a video in the 1D + t format using the visual rhythm concept. This example shows a sample of pixels defined by a vertical line of each frame F_t.*

Figure 2.2: **Visual rhythm example.** Examples of a video, a video cut, and a visual rhythm representation.

In order to analyze the video from the temporal point of view, new cuts can be defined in $\{F_t\}$ using different sampling strategies (e.g., based on vertical, horizontal, or diagonal lines). Figure 2.2(c) illustrates the use of a vertical sampling approach. In this example, each F_t is transformed into a column on a new image, which represents the video visual rhythm (Figure 2.2(c)). Therefore, the visual rhythm (VR) of a video is given by $VR = \{F_{VR}(i,\, t)\} = \{F_{sampling}(x(z),\, y(z),\, t)\}$, $i \in \{0, 1, 2, \ldots\}$ and $t \in T$, where $F_{sampling}(x,\, y,\, t)$ is a spacial reduced version from $F_t(x,\, y,\, t)$, and $x(z)$ and $y(z)$ are functions of one dimension of independent variable z, where z and t refer to vertical and horizontal axis, respectively [65].

The visual rhythm approach has been used with success in different applications, such as: video cut detection, identification of transitions and flashes [52]; content-based video retrieval [92,121]; social analysis [55]. More recently, some studies have been using visual rhythms in Biology [70, 74], including remote phenolgy [9–11] and hyperspectral remote sensing image analysis [36]. In this thesis, we construct a visual rhythm representation for time series associated with numerical data stored in relational tables derived from both traditional direct phenological observation and vegetation images.

2.2.4 Radial Visualization

Radial visualization is the term used to describe a visualization system that uses circular or elliptical layout to display data. This visualization layout has been increasingly used for many purposes [37]. For temporal data, for example, the radial display has been widely used to build visualizations, such as spiral graph [127], circle view [64], circle segments [12], TimeRadarTrees [24], and axes-based visualization [119]. In another research venue, this approach is combined with other techniques. Some examples include the use of the radial visualization for geoscience observation data [68] and of the radial projection for geo-information [38]. Finally, phenology has used radial representations instead of the linear representations thorough the circular statistics analyses [81, 84]. The circular representations is a common way to represent temporal data with no true zero or starting point [42]. Many other applications of radial visualization are introduced by Draper et al. [37].

Inspired by these works, we also use the radial layout to represent data from phenology studies. Our approach encodes time dimension using concentric circles [33] and shows multivariate data using a similar concept presented in the circle view [64], and circle segments [12] approaches. Furthermore, we also use visual rhythm to summarize a large volume of data, building an visualization that enables the identification of cyclical patterns related to phenology data.

According to Draper et al. [37], many types of radial visualization exist and they can be combined to produce novel visualizations, such as the case of [73], which combines radial structure with pie chart to visualize the proportions of total energy used by one type of device across a measured period of time. Both [28] and [127] proposed the use of spirals structures to show the seasonality of data. In our proposal, we also use the concept of radial structure to organize time-varying phenological data.

2.2.5 Motion Representations

The proposed CFH follows the principles of methods proposed for human motion characterization in computer vision applications. Bobick and Davis [21] introduced the idea of projecting a temporal pattern of motion into a single image-based representation named as *temporal template* [19–21, 34, 77]. They proposed a representation and recognition theory that is based on *motion-energy image* (MEI) and *motion-history image* (MHI) [20, 21, 34]. MEI is a binary cumulative image that describes *where* a motion has occurred in an image sequence. MHI, in turn, is an improvement of MEI, in which pixel intensity is a function of the motion history at that pixel and describes *how* motion is changing [21, 77]. They used video sequences of aerobics exercises for evaluating the power of the proposed temporal template representation [34].

Bradski and Davis [23] proposed a timed MHI (tMHI) for motion segmentation that allows to determine the normal optical flow. Other descriptive representations of motion based on MHI context were developed by Moeslund et. al [77]. In their study, they present a survey of several techniques proposed to compact human motion sequence into a single image.

More recently, Meng et al. [76] proposed the *motion history histogram* (MHH) also for human action recognition. The MHH basically encodes motion patterns found in videos, but differently from MHI, the MHH encodes all temporal changes into a single image, while in the MHI only the last change is represented. An important feature of MHH is that the representation is based on a collection of histogram-based grayscale images [76]. Thus, each histogram is computed from a sequence of video frames where a binary image is generated. This image is expected to encode the most significant motion patterns. A motion pattern is defined by a sequence of 0's and 1's connected (e.g., 010, 0110, 01110, 011110). Therefore, histograms associated with the binary representations are used to detect the most relevant motion patterns. Those histograms are used along with classification apparatus to find events of interest in videos.

Different from those initiatives, we generalize the motion pattern characterization solutions to detect and represent temporal changes in arbitrary temporal multivariate numerical data. Our algorithm is similar to the one proposed in [76]; but instead of using a video clip as entry data, we use numerical matrices as input and our output is a set of heatmaps representations that encode the frequency of temporal change patterns. In this work, we also exploit different binary encoding approaches to characterize phenological events of interest. To the best of our knowledge, this is the first work using this strategy to characterize complex phenological phenomena.

2.3 Information Visualization in Phenology

This section introduces common approaches for visualizing phenological data.

2.3.1 Common Practices

Plant phenology studies are based on the observation of individuals and/or species and the correlation of data related to their life cycle with climate variables [85]. They are not only concerned with the analysis of data related to observations in the field, but also, more recently, phenological studies are associated with vegetation images taken by digital cameras remotely [4, 5, 103, 104]. This scenario demands the creation of appropriate tools to support the discovery process associated with the identification of patterns observed in multidimensional and multivariate time series.

2.3.2 Visualizing On-the-ground Phenological Data

On-the-ground phenological data are periodically collected from individuals located in an observation area named site. The data values of observation vary according to the sampling method used by experts, but usually is related to the identification of the intensity of particular phenophases [71, 81].

The most used approach for visualizing the behavior of phenological variables over time relies on the use of linear charts. Figure 2.1 shows a typical example, illustrating the variation of the number of individuals of the species *Aspidosperma Tomentosum* over time, from Jan. 2011 to Feb 2012.

Phenology experts have been seeking for computational and technological options that support them in data analysis, organization, and storage. One example towards this direction is the *Plant Phenological Online Database*[2] (PPODB) [35], *USA national phenology network*[3] [109], *Nature's Notebook*[4] [106], among others. However, in those studies, it is still common the use of simple visual structures such as tables and linear charts.

In another research venue, several researchers represent temporal data associated with plant life cycles (phenology) in radial graphs [81]. Those structures are useful for understating cyclical events, such as the phenological ones, and allow the use of specific circular statistics, which are suitable for the analysis of such kind of data [81, 84, 114]. Figure 2.3, for example, illustrates the use of a radial structure to encode the intensity of the leaf flush and flowering for different species, over the months of a year.

The analysis of existing relations among multiple variables is also very important in phenology studies. For example, the relation between two successive phenophses (e.g. leaf flush and leaf fall) over time or between a climate variable (e.g., rainfall) with particular phenophases. In this last case, researchers may be interested in understanding how a particular variable drives the plant behavior. Figure 2.4 shows a typical chart used to represent multiple variables. In the example, a one-year multivariate representation is used: a bar plot is employed to represent the rainfall variable, while lines encode the number of individuals (in percentage) for which the phenophases leaf flush and leaf fall were observed. In Figure 2.4(a), we have the monthly total precipitation and the phenophases of leaf flush and flowering of the species *Xylopia aromatica*, while, in Figure 2.4(b), we have the monthly total rainfall and the phenophases of leaf flush and flowering of the species *Pouteria torta*. The data considered refer to phenological observations that took place from January to December 2011 for a cerrado *sensu stricto* vegetation, located at Itirapina, Southeastern Brazil. Rainfall data are from Centro de Recursos Hídricos e Estudos Ambientais (CHREA), located 3 km from the vegetation site.

Figure 2.5, in turn, refers to a long-term analysis of the same species (*Xylopia aromatica* and *Pouteria torta*). In the example, phenology experts may be interested in analyzing correlations over multiple variables over long periods of time (from January 2008 to December 2011).

2.3.3 Visualizing Near-Surface Phenological Data

The "Near surface remote phenology" consists in using digital cameras and other sensors installed close to ground level or at the top of towers for the monitoring of species

[2] `http://www.ppodb.de/` (As of July 25, 2018).
[3] `https://www.usanpn.org/` (As of July 25, 2018).
[4] `https://www.usanpn.org/natures_notebook#` (As of July 25, 2018).

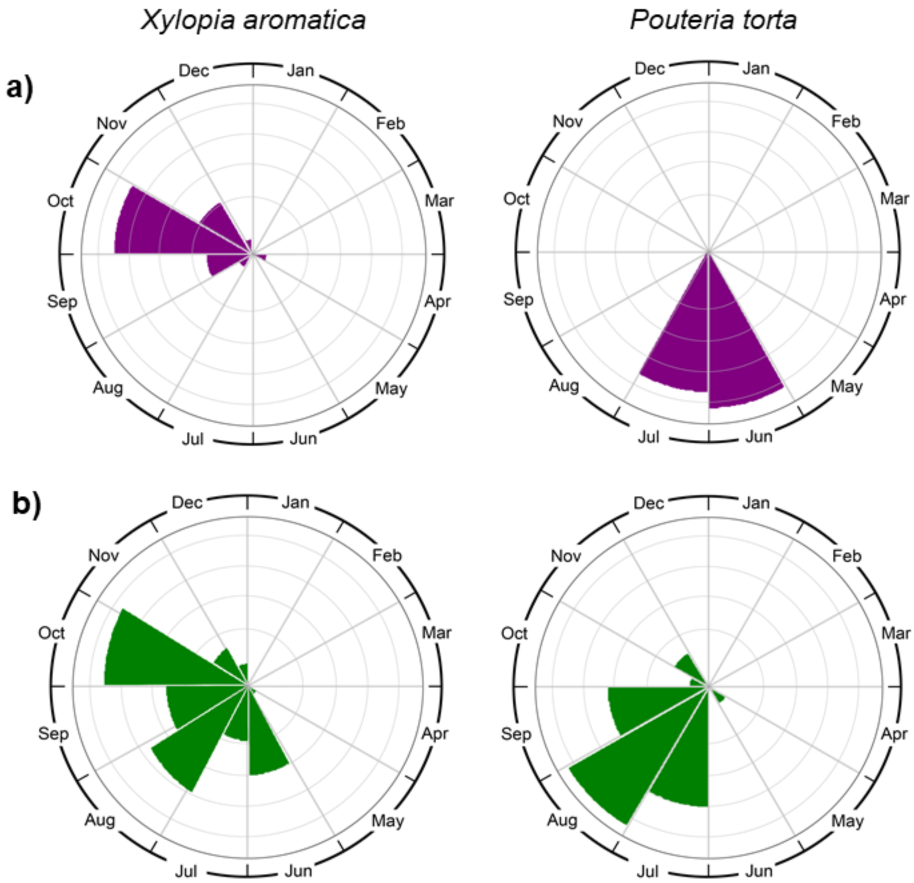


Figure 2.3: Radial structure representing the frequency distribution of individuals initial dates of (a) flowering (purple bars) and (b) leaf flush (green bars), for the species *Xylopia aromatica* (left) and *Pouteria torta* (rigth). Phenological data observations were collected from January to December 2011 in a cerrado *sensu stricto* vegetation, Itirapina, São Paulo state, Southeastern Brazil. The scale of the bars values are presented in the primary Y axis of the correspondent species in Figure 2.4.

up to ecosystem-scale vegetation changes. In comparison with the on-the-ground direct traditional phenology, the use of digital cameras to track leaf exchange phenophases allows long-term monitoring of high-frequency data (hourly, daily), simultaneously across multi-sites, with reduced human labor, and present an important role by filing the "gap of observation" among scales from ground to the land surface monitoring using satellite images [5, 86].

The process of digital images analysis depends on the definition of Regions of Interest (ROI). A ROI represents a region within the sample image defined for analysis (see [4, 5]). The ROI can represent crowns of species, populations, a community profile or even a vegetation type within the landscape [5]. From each ROI, the RGB (red, green, and blue) information of the image dataset is extracted and vegetation indices are enabled to be calculated (e.g., RGB chromatic coordinates). For instance, the Gcc index (Green chromatic coordinate, see [48] can represent the amount of leaves being putting out (leaf flushing) by a single tree or a whole vegetation community (Figure 2.6). Daily Gcc values

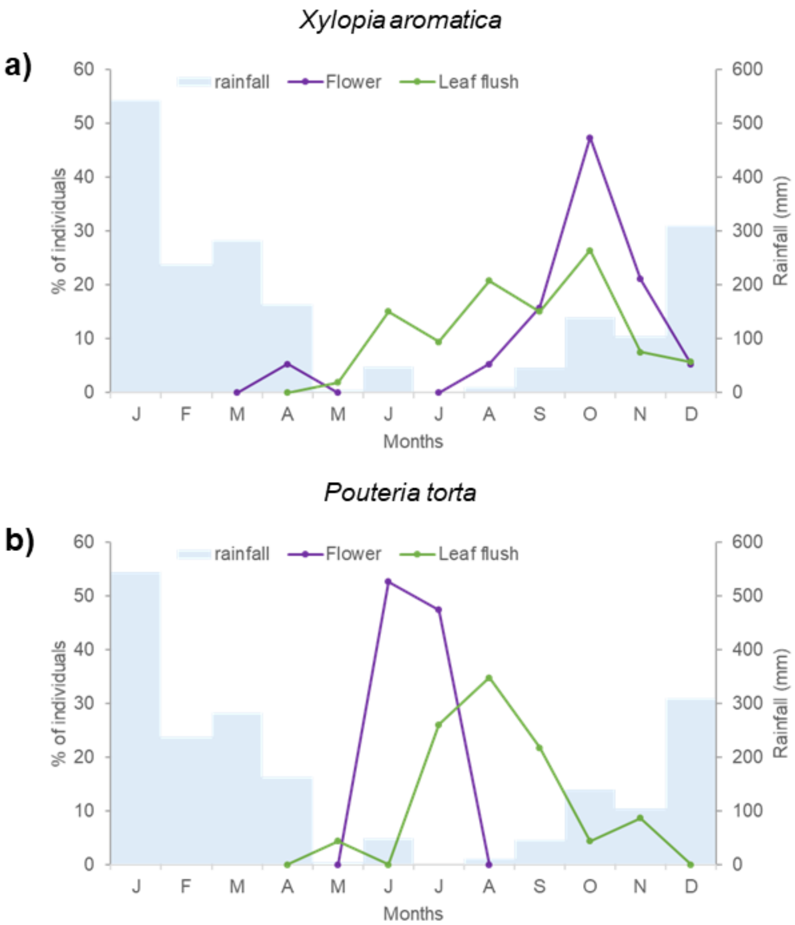


Figure 2.4: Linear chart showing monthly phenophases and climate variables. Monthly cumulative precipitation (bars) and the start of the phenophases leaf flush (green line) and flowering (purple line) of the species (a) *Xylopia aromatica*, and (b) *Pouteria torta*. Phenological observations from January to December 2011 for a cerrado *sensu stricto* vegetation at Itirapina, São Paulo state, Southeastern Brazil. Rainfall data are from CHREA (Centro de Recursos Hídricos e Estudos Ambientais), located 3 km from the study site.

are plotted against a time line representing the day of the year in a X-Y plot (Fig. 2.6 b-e). Gcc curves swings can be analyzed and used to provide information from the seasonality of leaf exchanges.

2.4 Conclusions

Many questions arise when we need to analyze multidimensional and multivariate temporal data. For example, how can many sets of temporal data be analyzed at the same time? How can scientists interact with data? How can cyclical patterns be detected? One way to address those questions relies on the use of information visualization techniques.

In this context, this chapter introduced the basic concepts and the main related work associated with the development of this thesis. We introduced the main concepts related

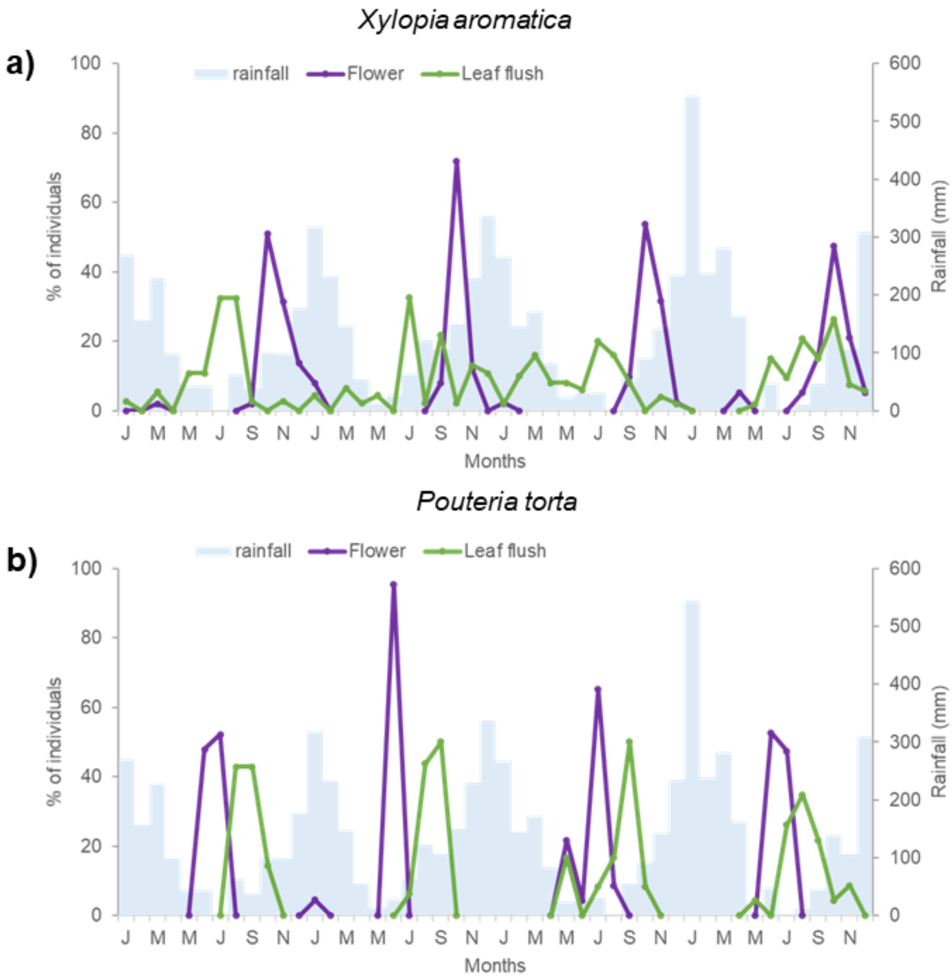


Figure 2.5: Linear chart showing monthly phenophases and climate variables over several years. Monthly total precipitation (bars) and the leaf flush and flowering (lines) of the species (a) *Xylopia aromatica*, and (b) *Pouteria torta*. Phenological observations from January 2008 to December 2011 for a cerrado *sensu stricto* vegetation at Itirapina, São Paulo state, Southeastern Brazil. Rainfall data are from Centro de Recursos Hídricos e Estudos Ambientais (CHREA), located 3 km from the study site.

to the information visualization area and the importance of validation of the developed techniques targeting users' needs. We also introduced the concept of time series and described some existing visualization techniques developed to visualize temporal data focusing mainly on radial visualization, method chosen in this work. Besides that, in the context of time series visualization, we also discussed the state of the art related to the visual rythm and motion recognition methods. Both are used in the context of this thesis to support the data visualization by phenology experts. Finally, we introduced some traditional visualization methods used by phenologists on their daily tasks.

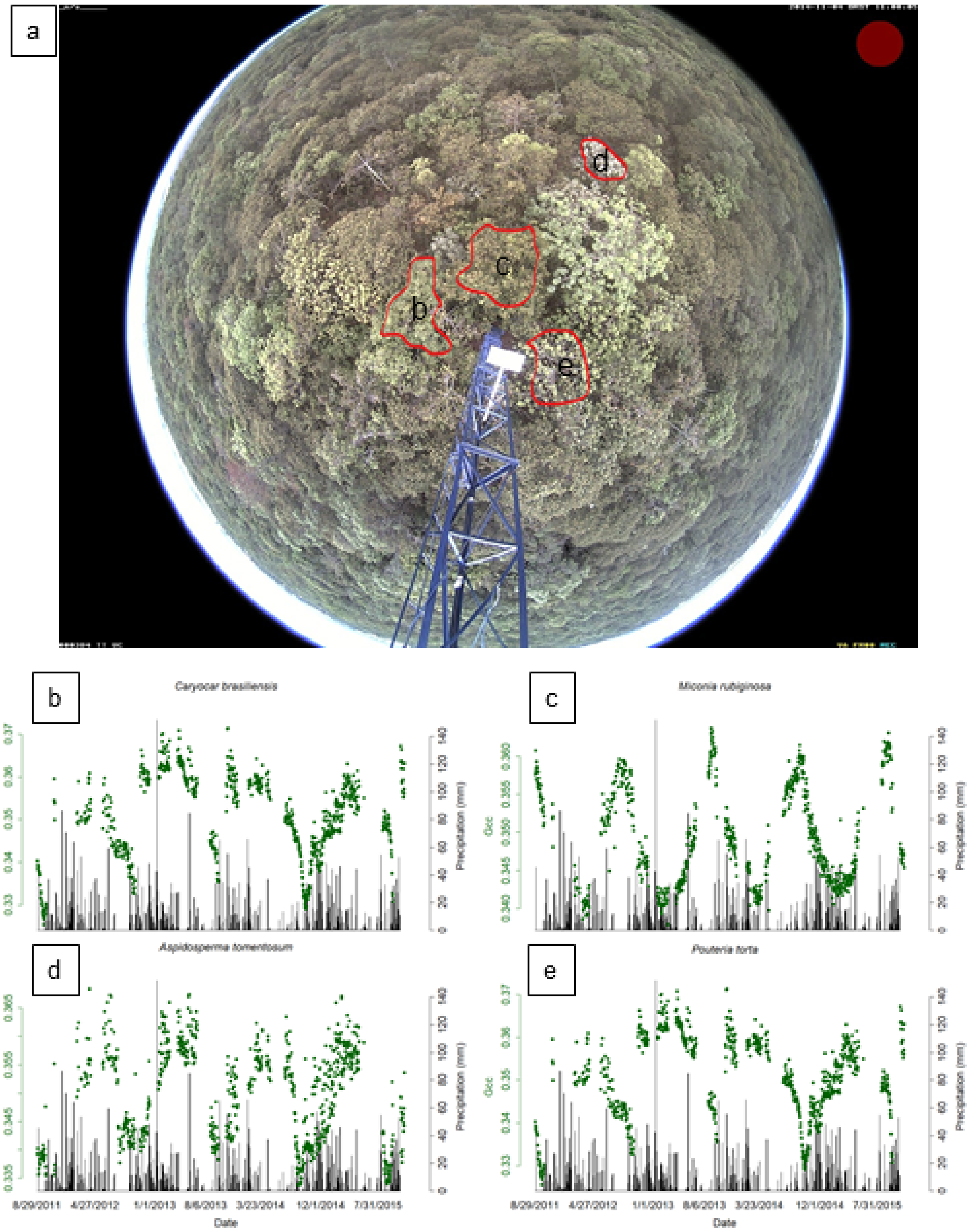


Figure 2.6: Example of (a) an image from a fisheye camera showing four regions of interest (ROIs) marked in red; each ROI represents one species' crown. Linear charts of daily Gcc values (green chromatic coordinates – green dots) representing the leaf flusing phenophase for ROIs of the species: (b) *Caryocar brasiliensis*, (c) *Myrcia rubiginosa*, (d) *Aspidosperma tomentosum* and (e) *Pouteria torta*. The representation also encodes information about precipitation (black bars).

Chapter 3

Radial Structure combined with Visual Rhythm Approach

This chapter presents the proposed approach to encode on-the-ground observations and image data into radial structures in a summarized way using the visual rhythm approach. First, we introduce the problem (Section 3.1) and present an overview of our approach (Section 3.1.1) and the usage scenario (Section 3.1.2). We then discuss some implementation details (Section 3.2) and the user-centered evaluations performed with phenology experts (Section 3.3). In that section, we describe target research question (Section 3.3.1), the evaluated prototypes (Section 3.3.2), the experimental setup (Section 3.3.3), the results (Section 3.3.4) and a discussion about the validation (Section 3.3.5).

3.1 Introduction

A common approach used for data organization relies on the use of database-oriented solutions, such as relational databases, to store plant phenology data [35, 71]. Despite their success, the need of visual analytics tools still persists. Although several approaches have been proposed to visualize time series [3], most of them are not appropriate or applicable to phenology data, as they are not able: (i) to handle long-term series with many variables of different data types and one or more dimensions and (ii) to identify cyclical temporal patterns and the associated environmental drivers. Furthermore, visualization tools should support data analysis under different perspectives: for example, a detailed view focusing on individual, and at large scales, a summarized view considering data about species and communities.

In this chapter, we address these issues by introducing one suitable alternative for visualizing data related to phenology studies. Our approach is based on using a radial layout along with visual rhythms (VR). The use of a radial structure aims to provide contextual insights regarding cyclical temporal phenomena. The use of visual rhythms, in turn, aims to summarize data at species and community levels that will be later represented using the radial layout.

The use of radial representation has been successfully applied on phenology studies instead of the linear representations through the circular statistics analyses [81, 84]. The circular analyses are a particular way to represent temporal data with no true zero or starting point [42]. Traditionally, VR representations were proposed to encode temporal change from video data using linear pixel sampling [65, 91]. In more recent studies, VR representations have been used to visualize temporal properties from digital images [9, 67]. The main advantages of this approach rely on efficiency aspects, as the compact representations generated are usually less costly in terms of storage and processing. With this in mind, the main novelty here is on exploiting the use of VR representations to encode conventional phenological numerical data. To the best of our knowledge, this is the first initiative dedicated to the use of VR representations to encode conventional numerical data.

In this context, the key contribution of this work is the specification and the implementation of a visualization approach using radial layouts, along with visual rhythm representations, to explore large volumes of multidimensional temporal data, usually associated with different types of variables. The objective is to support the detection of cyclical patterns. We demonstrate the potential use of our tool in the context of phenology studies using data from the e-phenology project as a case study [4, 71].

We developed ten prototypes that were separated and validated in two groups with phenology experts (potential users of our tool), such as: (i) six prototypes were created to represent phenological data about individuals – which we call *data visualization at detailed level* and (ii) four prototypes were created to represent phenological data about species – which we call *data visualization at summarization level.* We also included two prototypes based on linear bar graphs (one for each level), which represent the typical visual charts usually created by phenology experts in their temporal analysis studies. All prototypes encode, visually, data associated with research questions related to the correlation among phenology phenomena and climatic conditions.

Lastly, another novelty of this work refers to the description of user-centered evaluation procedures aiming to validate different prototypes in the context of real-world phenology studies. To the best of our knowledge, this is one of the first works in the literature to describe experiments with phenology experts aiming to validate information visualization tools in a real scenario setting.

3.1.1 Proposed Approach Overview

The main objective of this work is to propose a visualization approach with the objective of supporting the identification of cyclical temporal patterns from long time series. More specifically, we want to investigate, for the first time, the use of visual rhythms in the analysis of conventional numerical, on-the-ground direct observational data used in phenology studies. In this context, Figure 3.1 presents an overview upon our proposal.

In A, we have a representation of temporal data used in phenology studies and explored in our present proposal. We consider the analysis of cyclical data associated with numerical (data related to both direct on-the-ground observations and climatic information). Once the data are obtained, they are processed and stored into conventional database (component B in the figure). Next, visual rhythm-based representations may be extracted from both numerical and image data. This step is illustrated by component C

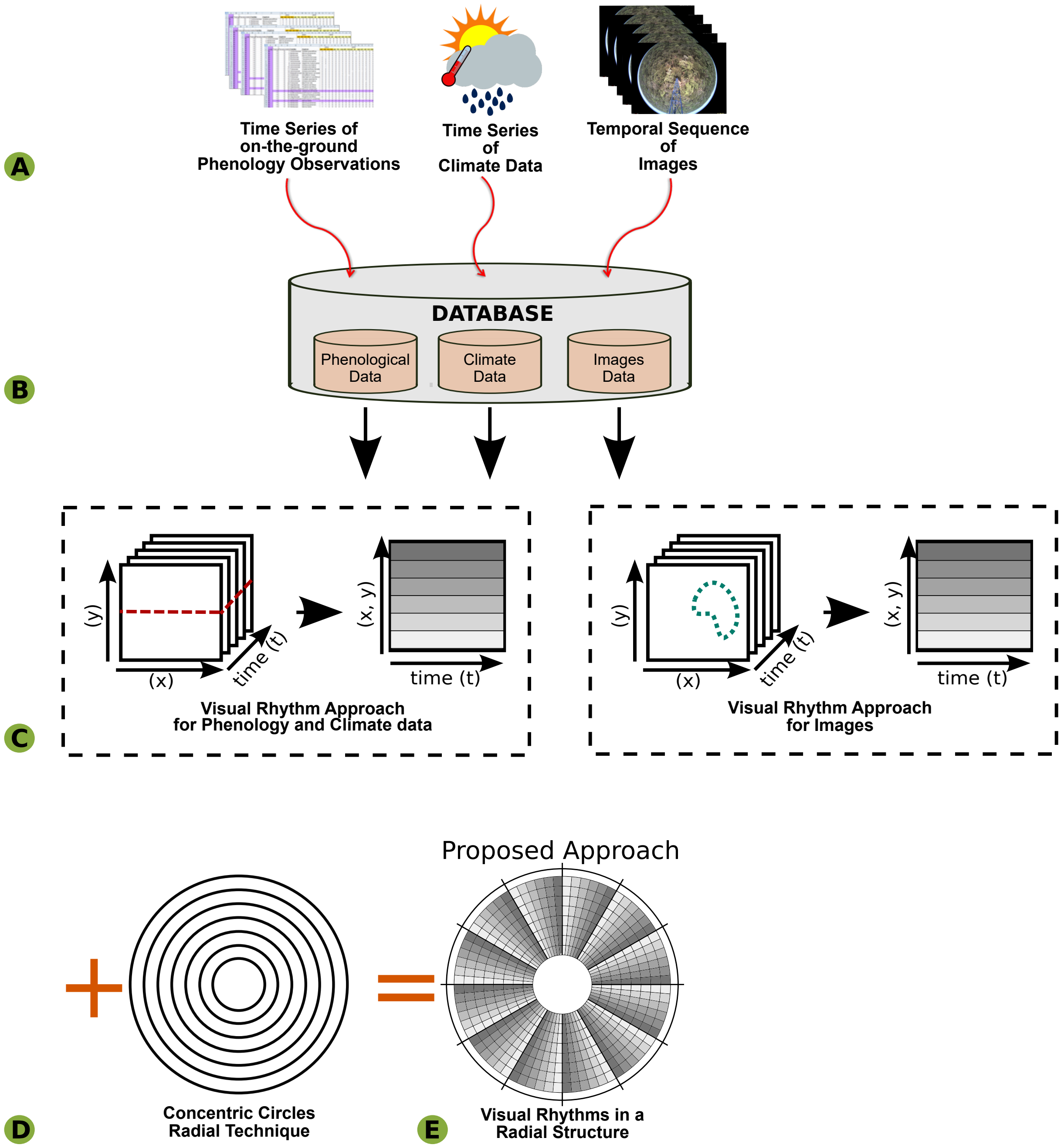


Figure 3.1: Overview of the visualization technique proposed for textual and climate data.

in the figure (see next section for more details). The visual rhythm representation is based on a two-dimensional image, which is expected to summarize the most important properties of the multidimensional data. Since our interest is to support the identification of cyclical temporal patterns, we use a radial visualization structure (component D) to represent the temporal data. Finally, as shown in E, we encode the visual rhythm representation into a radial structure.

3.1.2 Visual Rhythms for Phenological Data

In our proposal, we want to explore the use of visual rhythms for conventional numerical data. In the following, we detail how to represent these data using visual rhythms.

Phenology studies typically are based on the analysis of large volumes of temporal and cyclical data. For example, those studies often consider the observation of the life cycles of thousands of plants over time [35]. Common research questions addressed in those studies refer to the analysis of numerical data related, for example, to when (time of the year), for how long (duration), and at which intensity a given event (e.g., leafing or flowering) has occurred at a specific region [85, 117]. The most typical approach to store conventional numerical data associated with phenology studies relies on the use of relational databases [35, 71].

In this work, we propose the use of visual rhythm representations to encode numerical temporal data stored in relational tables. The objective is to support the identification of cyclical patterns in numerical data related to phenological studies. In order to represent temporal changes of numerical data, we use the same idea of original visual rhythm approach wherein a video is decomposed into multiple images. Without loss of generality, we assume that it is possible to encode phenological temporal data into a set of tables ordered in time.

Let $R(A_0, A_1, \ldots, A_m, T)$ be a relation (table), where $A_i (0 \leq i \leq m)$ is a numeric attribute and $T \in \mathbb{N}$ is an attribute associated with time instants. Without loss of generality, the relation R can be decomposed into several smaller tables by performing selections based on the values of the time attribute. For example, Figure 3.2 illustrates the decomposition of relation R based on the selection of tuples (rows of the table) associated with different timestamps t $(1 \leq t \leq n)$. In the example, relations R_t refer to the set of tuples whose the value of attribute T is equal to t. Using the relational algebra notation, $R_t = \sigma^R_{T=t}$, where σ is the *selection* operator.[1]

Each R_t has its width equal to the number of numerical attributes (m) and height equal to the number of tuples associated with the same time instant t. In phenology studies, for example, A_i may refer to the phenophase intensity (e.g., for flower bud or anthesis) observed for different individuals over time. A specific R_t, in turn, may refer to the set of the phenophase intensities observed for all individuals at timestamp t.

The decomposition of relation based on time information leads to a set of matrices that can be used to generate visual rhythm representations. For example, Figure 3.3(a) shows the creation of a visual rhythm representation based on column information. Following our previous example, supposed we are interested in studying changes of a particular phenophase (defined by attribute A_i) over time. In this case, $R_v(t, z) = \{\pi_{A_i}(R_t)\}$, $t \in [1, n]$ and $z \in [1, H]$, where n and H are its width and height, respectively, and π stands for the relational algebra *projection* operator that returns a subset of R_t, restricted to the set of attributes defined by A_i.

Suppose now that we are interested in observing changes on intensities of all phenophases for a particular individual $< id >$. In order to support this analysis, we can create a visual rhythm representation based on multiple horizontal selections (one for each relation R_t).

[1] For more details upon the typical relational algebra operators, readers may refer to [39].

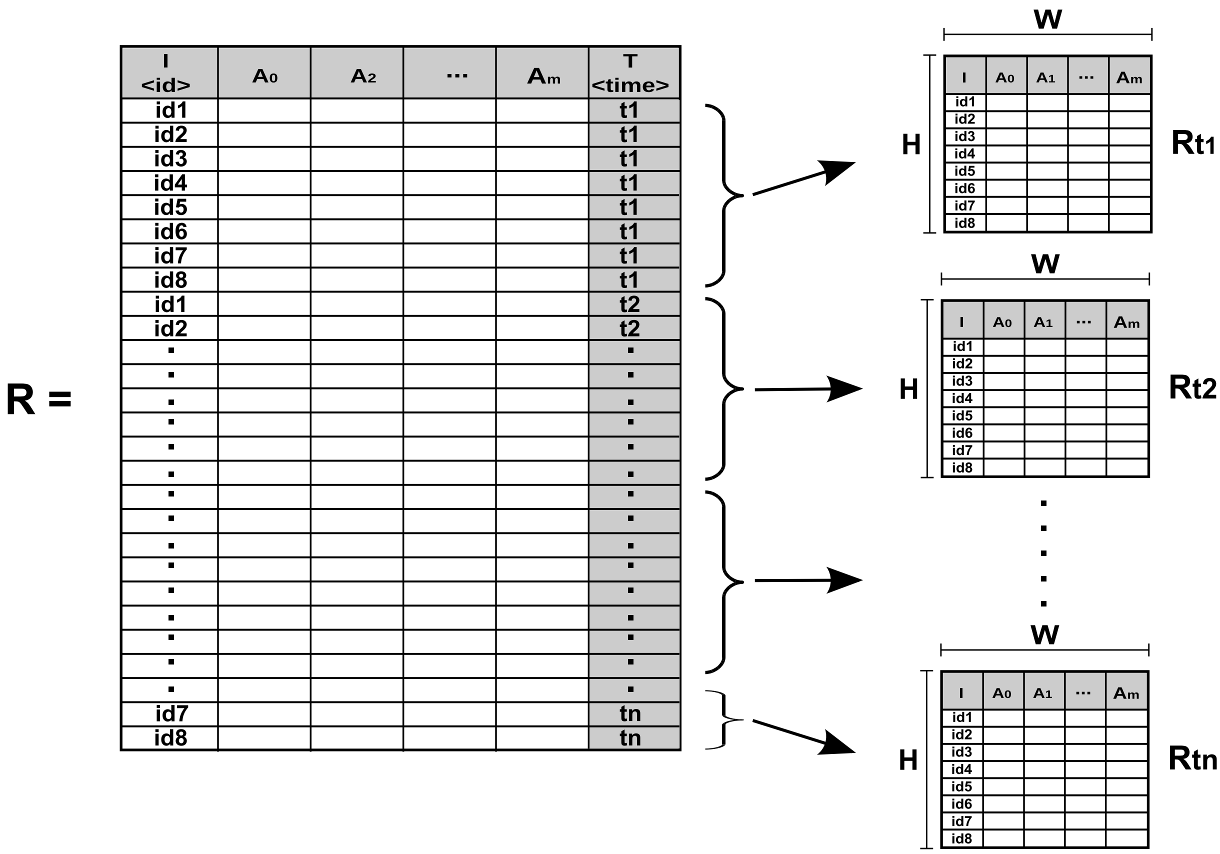


Figure 3.2: Decomposition of relation R into several relations R_t, where t refers to a timestamp.

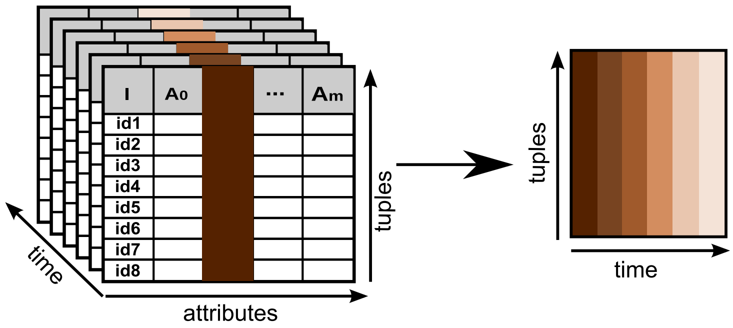
Formally, $R_v(t,z) = \{\pi_{\mathbb{A}}(\sigma_{I=<id>}(R_t))\}$, $t \in [1,n]$ and $z \in [1,W]$, where $\mathbb{A}$ is a set of attributes associated with different phenophases. In this case, the numerical and time attributes will form y-axis and x-axis, respectively, into visual rhythm image as shown Figure 3.3(b).

Another way explored in this thesis to build visual rhythms from relational tables relies on the selection of more than one value for each attribute and more than one attribute. Thus, in addition to projection and selection operations we also have to use a aggregation function to summarize data, such that $R_v(t,z) = {}_{\{A_i\}}F_{agregation}(\pi_{\{A_i\}}(\sigma_{condition}(R_t)))$, $t \in [1,W]$ and $z \in [1,H]$, where z on this case is a multidimensional variable. Typical aggregation functions used include *count* and *average*. Figure 3.4 illustrates this example.

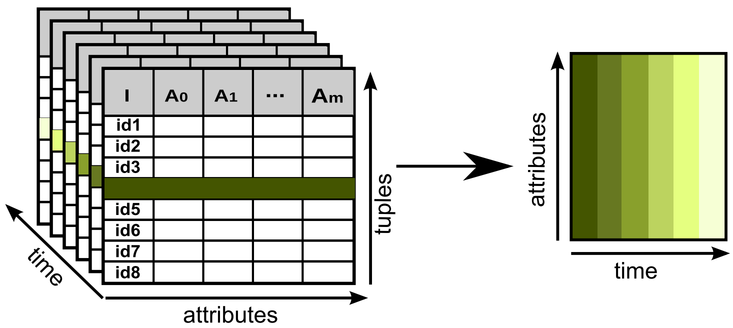
3.2 Implementation Details

Once we have used the visual rhythm approach to encode conventional data (e.g, data stored as a set of tables ordered in time as explained in the previous section), we have a bi-dimensional table, which can be represented using the radial layout. Figure 3.5 shows the general idea behind our visualization approach for phenology data. The main objective is to develop a tool that supports the joint visualization of all variables typically involved in phenology studies.

In our representation, we use the concept of concentric circles divided into segments



(a) *Example of visual rhythm creation based on vertical projections.*



(b) *Example of visual rhythm creation based on horizontal selections.*

Figure 3.3: Examples of visual rhythm creation based on selections (horizontal cuts) and projections (vertical cuts).

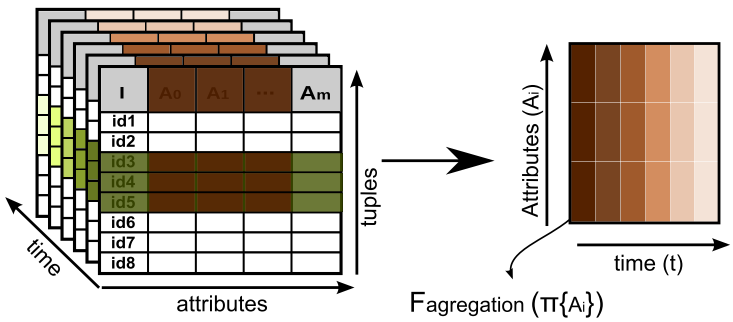


Figure 3.4: Example of the use of the visual rhythm approach to summarize tabular data using aggregation operations.

to show the attributes related to phenological, climate, or image-related data. Thus, each segment s, for example, can represents a time instant $t \in T$ $(1 \leq t \leq n)$ and may be associated with one or more attributes, A_i $(1 \leq i \leq m)$.

As an example, suppose we want to visualize the visual rhythm obtained for m phenophases (A_i) for a particular group of three individuals $< id_i >$ over one year (n months). Using the idea of horizontal cuts (see Figure 3.3(b)), the resulting visual rhythms comprise three tables (one visual rhythm for each individual) composed of values of *phenophases* attributes over time (attribute *time*) each one. In this case, circles refer to *individuals* (one for each circle), segments are related to *months*. Each segment can be further divided into six other sub-segments used to represent *phenophases*. Thus,

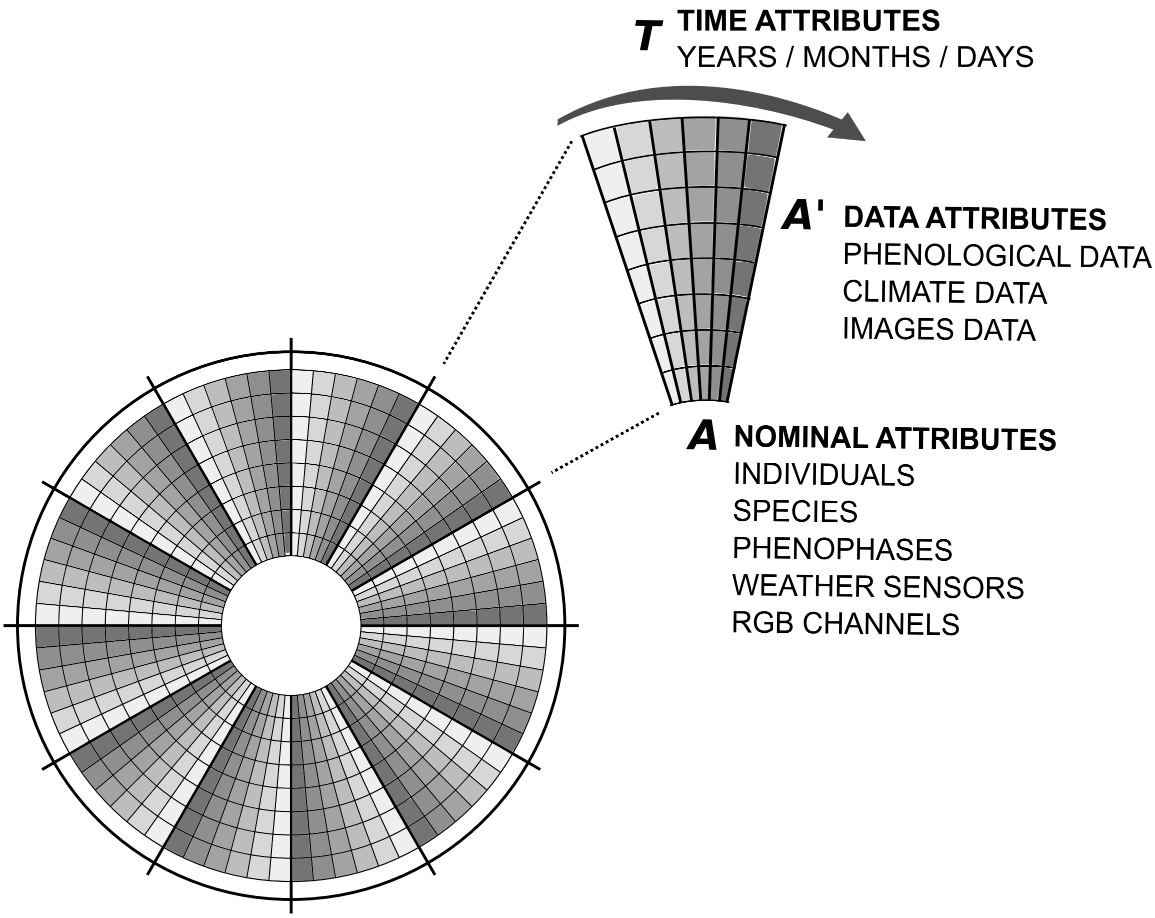


Figure 3.5: Overview of the visualization proposed for numerical data.

the amount of segments for one circle will be equal to $n \times m$. Note that this distribution and the visual rhythm to be represented may be different depending on the user goals.

For example, the data organization within the prototypes – developed to validate our approach – concerning time information works as follows: circles are associated with *years*, while segments are used to represent *months*. Therefore, we have 12 segments that can be subdivided according to the number of attributes considered. The numerical attributes are associated with phenological, climate, or image data.

3.3 Validation

The validation of the proposed visual representation relies on a user-centered evaluation of prototypes, which encode on-the-ground direct observation phenology data.

We carried out experiments with phenology experts aiming to validate the prototypes created to represent both, detailed and summarized data visualizations, on the context of phenology studies. The experiments aimed at evaluating the effectiveness of the proposed visualization prototypes using a task-oriented evaluation methodology.

3.3.1 Target Research Questions

The data considered in our case study refer to phenology observations conducted our core cerrado site of e-phenology project described above, in a period ranging from 2005 to 2007. For each individual, researchers observe six reproductive phases, as follows: (i) flower buds, (ii) open flowers or anthesis, (iii) unripe fruits, (iv) ripe fruits, and the vegetative phases of (v) leaf fall and (vi) leaf flush or new leaves, as defined in [123]. Recall that to each phenophase, in each observation, a score from 0 to 2 is assigned. The current implementation assumes that target users, phenology experts, are interested in comparing phenological and climate data only. The climate data considered in our study refer to the precipitation observed for the same time period.

Some examples of typical research questions related to phenology studies that experts try to address using these data are:

1. Is it possible to visualize phenological data (all phenophase intensities) of one specific individual for a specific period of time, spanning several years?

2. Is it possible to identify the date of peak for a specific phenophase intensity for one individual?

3. Is it possible to visualize phenological data (all phenophase intensities) of one specific individual combined with climate data for a period of time, spanning several years?

4. Is it possible to visualize phenological data (all phenophase intensities) for one species, spanning several years?

5. Is it possible to identify the date of peak of phenophase intensities for one species, spanning several years?

6. Is it possible to visualize phenological data (all phenophase intensities) of one species combined with climate data for a period of time, spanning several years?

These questions aim to support the understanding of phenological changes over time and are extremely important in the context of assessing triggers and the impact of climate changes on plants phenology.

3.3.2 Evaluated Prototypes

Based on the research questions presented in Section 3.3.1, we developed a generic visual representation using the concept of radial display and visual rhythm. The use of this representation is dependent on the kind of question being considered. Two scenarios are studied:

- **Detailed visualization:** questions that are related to the visualization of individual-related data (questions 1 to 3);

- **Summarized visualization:** questions that that are related to the visualization of species-related data (questions 4 to 6).

Since phenology studies are based on long-term time series and involve a complex process between the observation in the field and the data analysis, we opted for a set of published and consolidated data in the literature. In this context, the data represented in the prototypes both in detailed and summarized visualization levels refer to phenological observations related to the species *Myrcia guianensis* collected in the core study site of the e-phenology project, and which was presented in [123].

Detailed Visualization

In order to address questions from 1 to 3 presented in Section 3.3.1, we have implemented seven prototypes (see Figures 3.6 and 3.7). All the prototypes present the phenophases as attributes distributed within the segments. The phenophases are presented using values 0, 1, or 2, which represent the respective observed intensity.

The main difference between them is concerned with how the climate data is shown and, in the case of Prototype 7, how the phenophases and intensities are represented – in this case, in a linear way (Figure 3.7 – Prototype 7). This is the most common method used by experts and therefore considered here as our baseline. The prototypes whose evaluation is close to the evaluation of the baselines will be considered good candidates for being used in phenology studies.

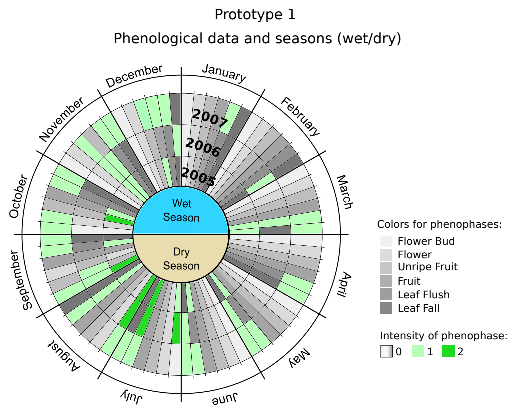
In the first prototype (Figure 3.6 – Prototype 1), the climate data are shown in the more inner circle in a generic fashion. In all other prototypes (Figure 3.6 – Prototypes 2, 3, 4, 5, and 6), in turn, climate data are placed along with the phenophase data (regions highlighted in blue).

In all prototypes, we consider for each month that the experts want to visualize all six phenophases. Furthermore, note that in the first, second, and fourth prototype, we represent the phenophase intensity values equal to 0 in gray, while in the other prototypes, in white. We also change the way a phenophase is represented: gray scale (Figure 3.6 – Prototype 2), icons (Figure 3.6 – Prototype 3), letters (Figure 3.6 – Prototypes 4 and 5), and numbers (Figure 3.6 – Prototype 6). With these seven representations, we want to support the visualization of cyclical data with the objective of allowing the identification of patterns and relations among the different variables. In the considered examples, relations among phenophase intensities and precipitation are considered. All prototypes also allow the identification of peaks in the intensity values (bright green cells).

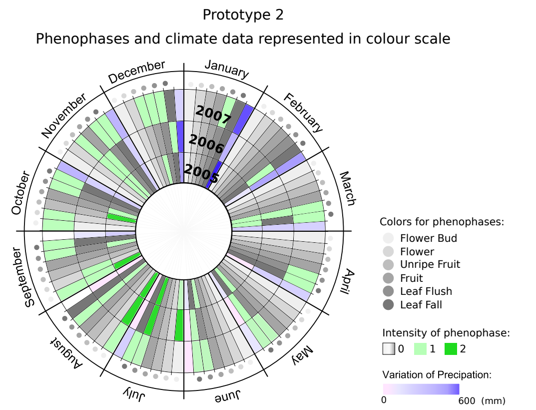
Summarized Visualization

With regard to questions from 4 to 6 presented in Section 3.3.1, five other prototypes were designed. All prototypes implemented for this step (see Figures 3.8 and 3.9) consider data about one single species.

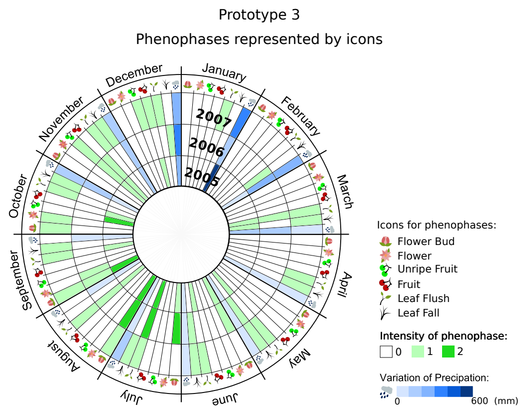
The main difference here is that data about one species include records associated with one or more individuals, which demands the use a summarization approach. In this case, we applied our approach based on the visual rhythm representation. To compute the phenophase data to be presented, we computed the visual rhythm using an aggregation function. Thus, instead of presenting values of intensity, here we calculate how many individuals of a particular species present the phenophase value different from zero. For



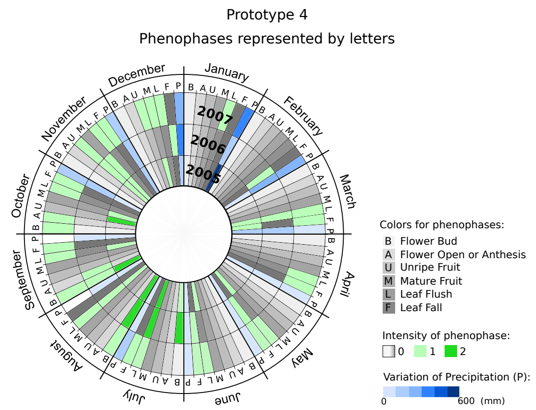
(a)



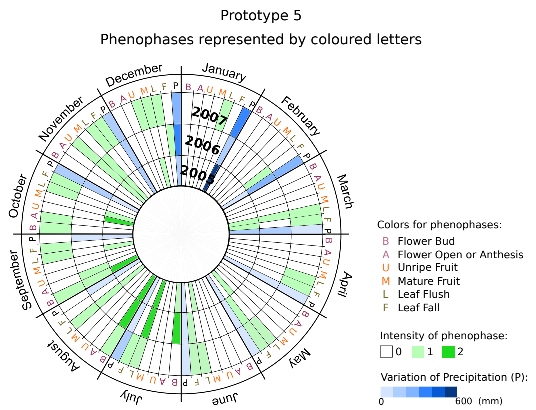
(b)



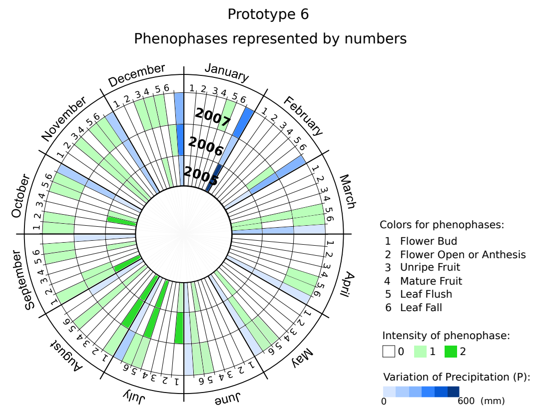
(c)



(d)



(e)



(f)

Figure 3.6: Generic prototypes developed to visualize phenological patterns from one individual and climate data using different graphical representations – detailed visualization level.

instance, for phenophase flower bud, we count *how many individuals* presented 1 or 2 in the time period considered in the study. Therefore, the values associated with phenophase segments that we present in our visualization refer to the relative percentage of individuals with phenophase intensity value different from zero or just the presence of phenological activity. Similarly to the prototypes of individuals, all prototypes have six phenophases

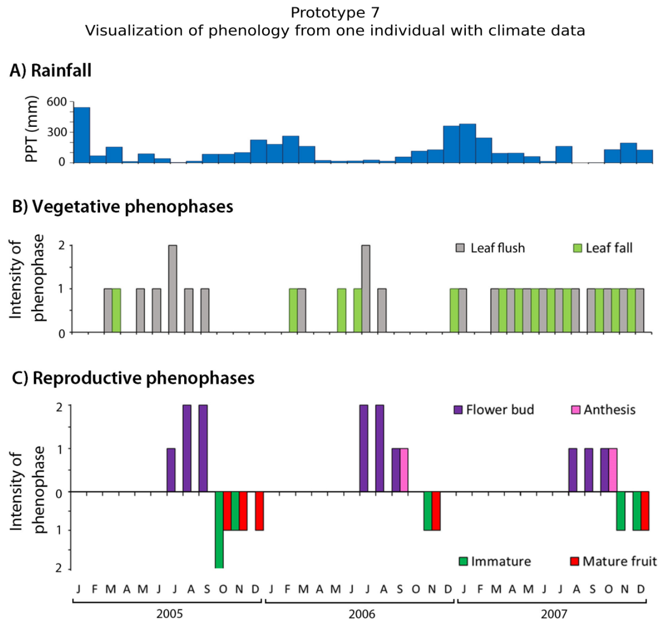


Figure 3.7: Generic prototypes developed to visualize phenological patterns from one individual and climate data using different graphical representations – detailed visualization level. In this example, bars are used to encode temporal data.

and the fifth prototype is our baseline linear visualization (Figure 3.8 – Prototype 5). Besides that, to identify each phenophase, we use different strategies: gray-level scale (Figure 3.8 – Prototype 1), icons (Figure 3.8 – Prototype 2), letters (Figure 3.8 – Prototype 3), and numbers (Figure 3.8 – Prototype 4).

3.3.3 Experimental Setup

User profiles: Eleven phenology experts were recruited by email to take part in the evaluation of the prototypes – six active researchers (Ph.D.), five Ph.D. candidates. One of the Ph.D. candidates did not fill out the evaluation form related to prototypes of species. The key selection requirement relied on their expertise in plant phenology domain. Subjects of any age (over 18 years) or gender were accepted in the study.

Task:. Before any procedures the research plan and forms were submitted for approval in the Unicamp humans ethical committee and the participants had to sign a consent form. For the evaluation, we created two types of forms – one used to evaluate prototypes representing individuals data and another to evaluate those that manage species data. Thinking in the target research questions presented previously, in both forms the evaluators were asked: (i) to indicate how easy is the identification of phenophases intensities over time and if they could identify a relation between production of flowering phenophases (flower bud and flower); (ii) to indicate if they could identify peaks of flowering and fruiting in specific dates; and (iii) if they could identify a relationship between leaf flush and precipitation. The results obtained for the questions related to item (i) had as objective to address the target research questions 1 and 4; the results related to item

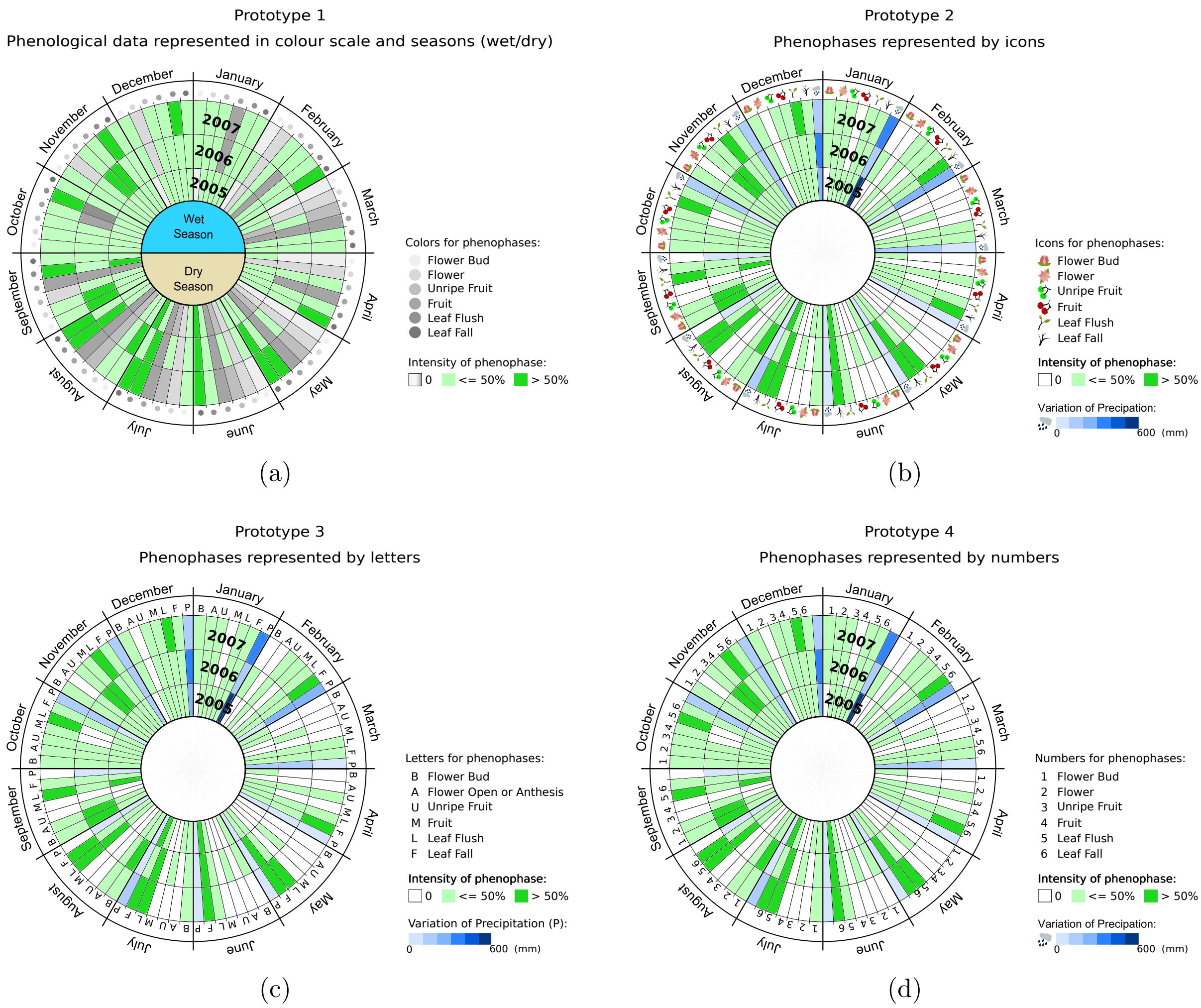


Figure 3.8: Generic prototypes developed to visualize phenological patterns from one species and climate data using different graphical representations – summarized visualization level.

(ii) had as objective to address the target research questions 2 and 5; finally, the results related to item (iii) had as objective to address the target research questions 3 and 6.

Procedure: All evaluators were asked to evaluate all prototypes. Each prototype was associated with a different evaluation form. All recruited evaluators filled out the form about individuals' prototypes and only ten filled out the form about species' prototypes. For each question of the forms, the user should indicate, in a Likert scale (from 1 to 5), whether it was easy to identify the required standards on issues. In the scale: 1 means that the user totally disagrees; 2, disagrees; 3, neither agrees or disagrees; 4, agrees; 5, totally agrees.

Opening questionnaire: Evaluators were asked to fill out a questionnaire concerning their familiarity with computers and visualization techniques, as well as, their expertise with tools to visualize phenological data. Only one participant indicated to know the information visualization area and reported to have used the package *ggplot2* from language R to represent data of interest.

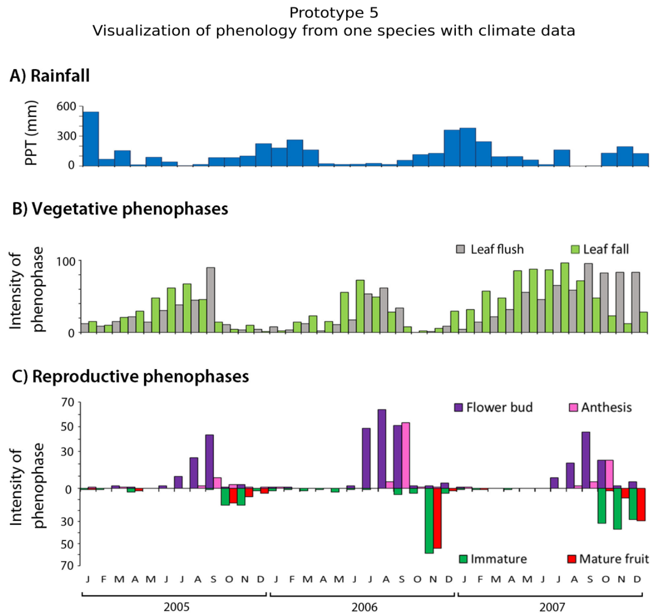


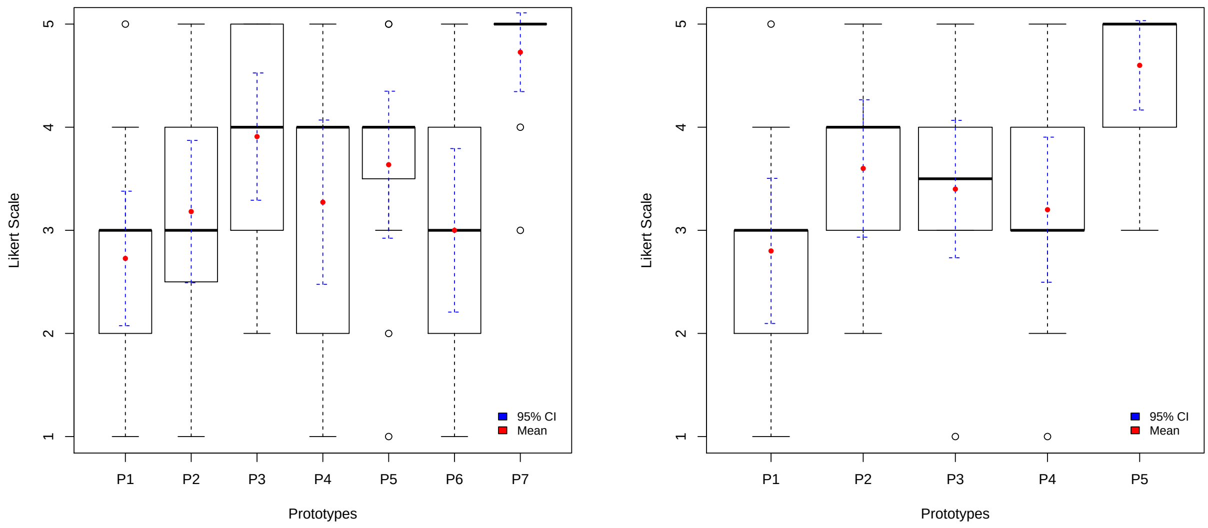
Figure 3.9: Generic prototypes developed to visualize phenological patterns from one species and climate data using different graphical representations – summarized visualization level.In this example, bars are used to encode temporal data.

3.3.4 Results

Figure 3.10 shows the results with regard to the satisfaction of users concerning how easy is the identification of phenophase intensities over years considering all prototypes evaluated. Figure 3.10(a) shows the results of prototypes that represent phenological data at detailed level (Prototypes 1, 2 ,3, 4, 5, 6, and 7 from Figures 3.6 and 3.7), i.e. using data observations about individuals, while Figure 3.10(b) shows the results of prototypes that represent phenological data in a summarized level (Prototypes 1, 2, 3, 4, and 5 from Figures 3.8 and 3.9), i.e., using data from species. In general, higher scores were observed for Prototype 3 related to data of individuals, i.e., this is the prototype whose evaluation is closer to the baseline (Prototype 7 for individuals). Regarding species data (Figure 3.10(b)), Prototypes 2, 3, and 4 have similar results. As it was expected, in all situations the baseline has the best scores, due to the familiarity of phenology experts in using the data representations employed by them.

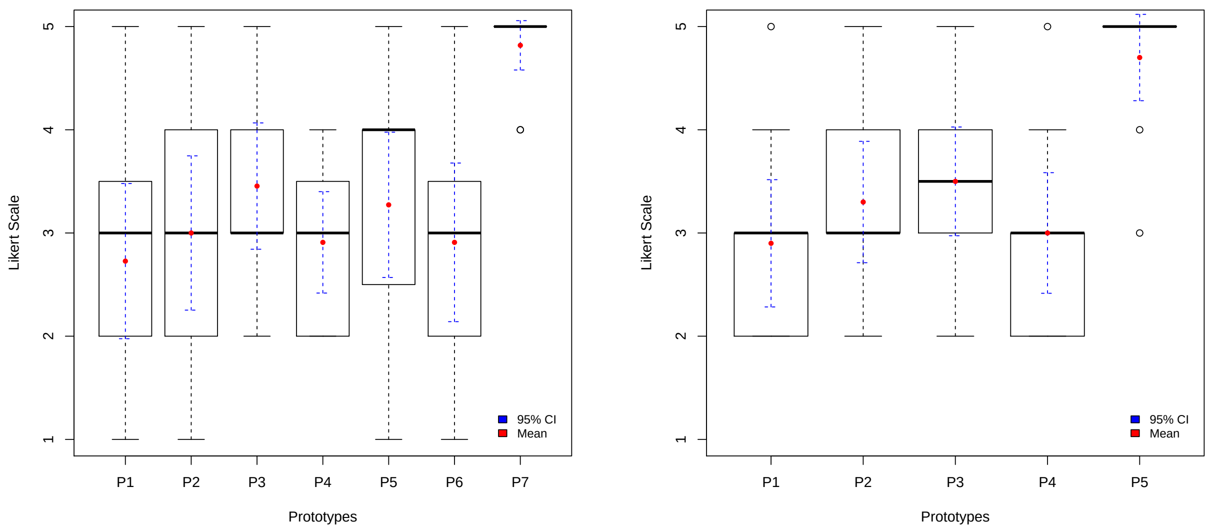
Figure 3.11 shows the results with regard to the satisfaction of users concerning how easy is the task of relating the production of flowers with the production of flower bud. In this case, we aim to assess whether it was understandable and easy for users to relate two phenophases. There is no clear winner with regard to prototypes for individuals (Figure 3.11(a)), but for the baseline. Prototype 3 (for individuals) has the the highest average score. All the prototypes for species data yielded similar results (Figure 3.11(b)), Prototype 3 with a slightly better mean.

Figure 3.12 shows the results with regard to the satisfaction of users concerning how easy is the identification of peaks of flowering for both individuals and species. Here, we



(a) Prototypes for individual-related data (b) Prototypes for species-related data

Figure 3.10: How easy is the identification of phenophase intensities over time.

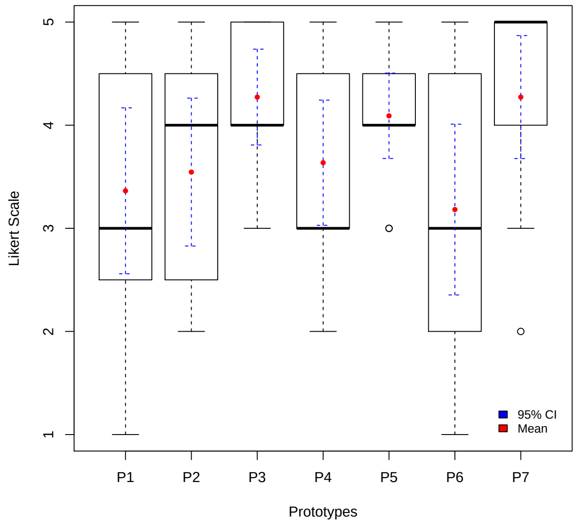


(a) Prototypes for individual-related data (b) Prototypes for species-related data

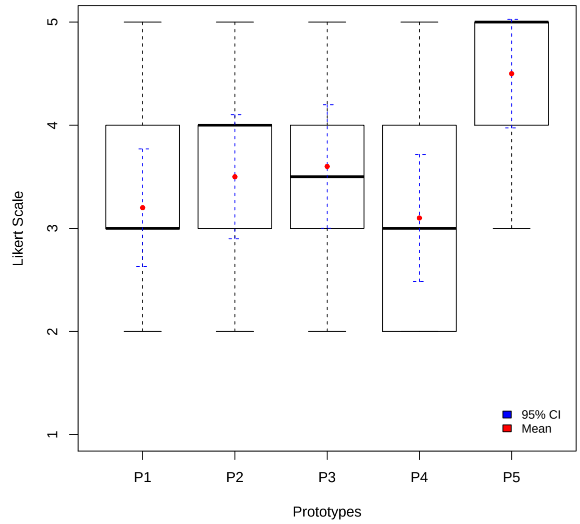
Figure 3.11: How easy is the identification of the relation between the production of flowers and the production of flower bud.

aim to assess whether the users are able to identify when flowering occurs. Again, Prototype 3 for individuals (Figure 3.12(a)) and Prototypes 2 and 3 for species (Figure 3.12(b)) were those which had the best scores.

Figure 3.13 shows the results with regard to the satisfaction of users concerning how easy is the identification of peaks of fruiting for both individuals and species. Here, we aim



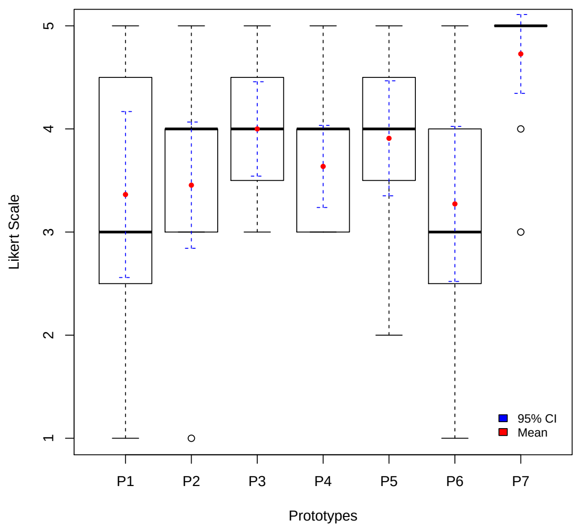
(a) Prototypes for individual-related data



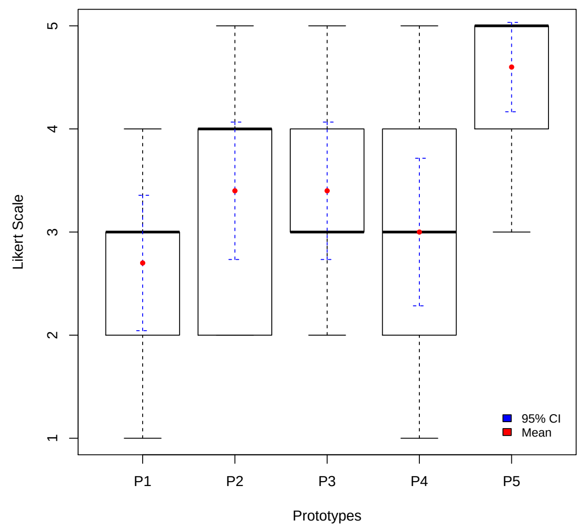
(b) Prototypes for species-related data

Figure 3.12: How easy is the identification of flowering.

to assess whether the users are able to identify when fruiting occurs. Again, Prototype 3 for individuals (Figure 3.13(a)) and Prototypes 2 and 3 (Figure 3.13(b)) for species were those which had the best scores.



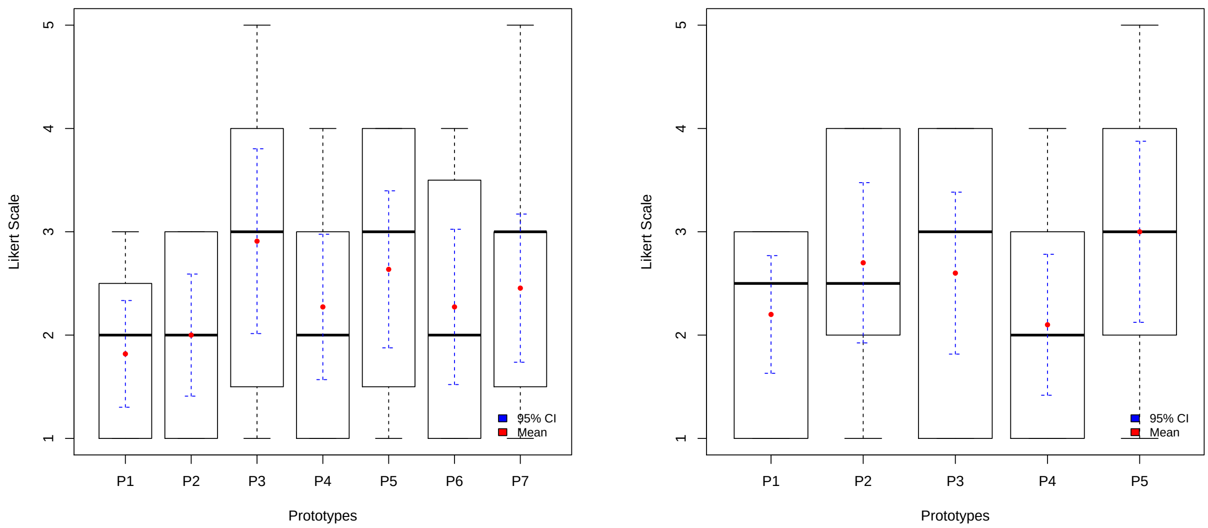
(a) Prototypes for individual-related data



(b) Prototypes for species-related data

Figure 3.13: How easy is the identification of fruiting.

Finally, we investigated how easy is for users to correlate climatic data (rainfall) with leaf flush phenophase. Figure 3.14 shows the results. In this task, no prototype achieved outstanding scores. Even the baseline for individual data (Prototype 7, Figure 3.14(a)) was not well evaluated. A slightly better mean is observed for Prototype 3 when dealing with individual's data. These results demonstrate how challenging is the task of providing appropriate visualization approaches for supporting the identification of correlation among complex variables and maybe a suitable approach is to visualize climate data separately from phenology data.



(a) Prototypes for individual-related data

(b) Prototypes for species-related data

Figure 3.14: How easy is the identification of impact of rainfall to leaf flush.

The evaluators also provided their feedback upon all prototypes in a text field. We have selected the most relevant positive and negative comments provided. In general, the evaluators commented that the use of a radial structure is a good strategy for displaying multivariate temporal data into a single representation. They believe, however, that at the same time, too much information is encoded in the representation, which hampers the analysis process. For example, in several studies, they are not interested in analyzing all phenophases altogether. Usually studies focus on specific stages of the plant life cycle (e.g., the reproductive phenology related to the flower bud and flower phenophases). One suitable alternative to address this issue would be the support of the use of filtering options (e.g., based on years, phenophases, or data types) with which users may select and analyze more easily data of interest.

The evaluators believe that the approach to encoding information about wet and dry seasons in the inner ring as used in Prototype 1 (for both species and individuals) is a suitable alternative to analyze the impact of environmental conditions on plants. Except for Prototype 1, climate data is encoded as a variable, which was welcome by the evaluators. On the other side, the evaluators commented that this strategy makes difficult to establish correlations among climate and phenological data. In some situations (e.g., Prototypes 4 and 5 for species), the provided color information made the correlation identification process even more difficult. Allowing users to change the type of chart utilized for represent climatic data based on their specific interests would address this issue.

The different prototypes use different strategies to support the identification of phenophases. For example, for individuals (see Fig. 3.6), it is used gray-level colors (Prototypes 1 and 2), icons (Prototype 3), letters (Prototypes 4 and 5), and numbers (Prototype 6). The evaluators, in general, preferred letters to numbers, and icons to letters. They believe, however, that the used icon-based representation should be avoided in scientific reports. To facilitate the identification of years or phenophases, customized controls for highlighting specific rings, segments, or cells could be provided.

Furthermore, prototypes (e.g., 3, 5, and 6 for individuals) with white background (usually used to encode phenophases with intensity equal to zero) are preferred to the ones with gray-level colors, because the evaluators were able to identify more easily the intensity and the peak of phenophases. We believe that the existence of customized controls to support the definition of different colors for specific data (segments) would be useful.

Another positive aspect of the radial structure emphasized by evaluators refers to the possibility of presenting data about multiple years into a single representation. In fact, this is the most commented drawback of the baselines (Prototype 7 of individuals and Prototype 5 for species). On the other side, according to the evaluators, the inner segments, which refer to data observed in year 2005, are too small. One alternative to address this issue would be the existence of customized controls for zooming in and zooming out (overview and detail levels). This approach would be useful also in the context of the use of the radial structures for representing long-term (multiple years) temporal data.

3.3.5 Discussion

Phenology studies rely on the analysis of the life cycle of living beings and its relationship with weather variables. Usually, performed analyses rely on the use of different kinds of data, typically handled though multiple tools. Therefore, scientists concerned with phenology studies usually seek support from a large set of information systems. This, of course, brings about all kinds of interoperability problems due to system mismatch, data diversity, and variety of user profiles [71]. Our approach is based on the possibility of using one single technique to visualize and manage different kinds of data, including multiple phenophases and environmental information in a comparative way, a especially important situation where a large number of temporal data has to be analyzed.

In the performed evaluations, we wanted to investigate what is the perception of users about visualizing multiple variables at the same time. In our evaluation, we opted for interface mockups associated with worst-case scenarios for the visualization. In this context, we presented data about all phenophases at the same time with the goal of assessing if, for the possible users, it would be possible to identify correlations between plant phenophases and climate data, and also to identify useful information regarding the whole life cycle of individuals or species. Note also that the considered scenarios, involving multiple phenophases, are useful for intraspecific comparison of individuals of the same species [72, 113, 123].

In summary, Prototype 3 for individuals and Prototypes 2 and 3 for species were the most promising ones to be used to support complex phenology studies, mainly related to the analysis of multidimensional data over several years. The use of icons as exploited in Prototype 3 (for individuals) and Prototype 2 (for species) seems more appropriate for data presentation in non-scientific purposes or in scenarios involving less experienced users (such those related to citizen science actions).

Also, the performed evaluation suggests the need of the implementation of interaction mechanisms. We noticed that the users would perform the proposed tasks more easily, possibly with the assistance of customization tools (e.g., interaction facilities such as selection or filtering features) to reduce and increase the amount of data presented in the proposed visual structure as her convenience. In particular, we noticed that these kinds of interactions are important mainly when phenology experts want to compare individuals from the same species or individuals from the same species but from different locations.

In fact, from the lessons learned from this study, we have conceived and have been implementing a complete visualization information system where the user can interact and control what the users will see and explore from the existing datasets. For example, users may select the number of phenophases and the color with which phenological data will be presented. Some screenshots are presented below.

Recall Prototype 4 (see Figure 3.8(d)), which was used to depict the pheno-phase intensity for one species during the years of 2005, 2006, and 2007. In Figure 3.15, we present the screenshot of the information system being conceived to represent the same data considered in Prototype 4. Suppose now that the user of the system is interested in studying only the reproductive cycle of individuals of this species. In this case, the user may select phenophases flower bud, flower, unripe fruit, and fruit, indicating the interest in visualizing their intensity scores. Figure 3.16(a) presents the radial structure layout after the selection of reproduction-related phenophases. Similarly, the user may be interested in comparing the intensity scores only for two phenophases. Figure 3.16(b) shows an example using the radial layout structure after the selection of only flowering-related phenophases (e.g., flower and flower bud).

In this tool, both naive and experienced users will be able to visualize collections of temporal multidimensional data. More experienced users may be interested in narrowing down data analysis, by taking advantage of selection and filtering options available in the tool interface. By using those features, the user may, for example, select specific phenophases (e.g., those related to reproductive cycle as illustrated before), individuals (e.g., those of a single region), and years (e.g., aiming to identify changes over specific time periods). Also, customization tools may be used, for example, to select and define different colors to represent the intensity of particular phenophase.

Another usage scenario refers to the possibility of using the proposed visualization approach for depicting monthly data related to vegetation indices extracted from sequences of images [4,5]. In this case, the daily changes can be observed and compared for different individuals (represented as regions within the image) of interest. Figures 3.17 presents an example of daily image taken by digital cameras associated with a region of interest. Using the region of interest, we can process the daily images to obtain the temporal features, such as the gcc variation [5], which can be visualized using the proposed visual structure.

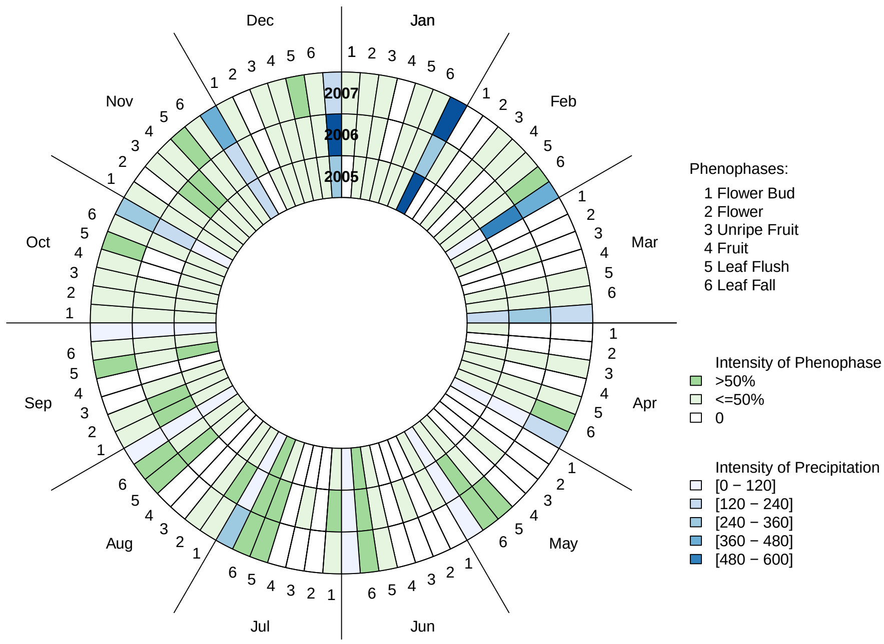


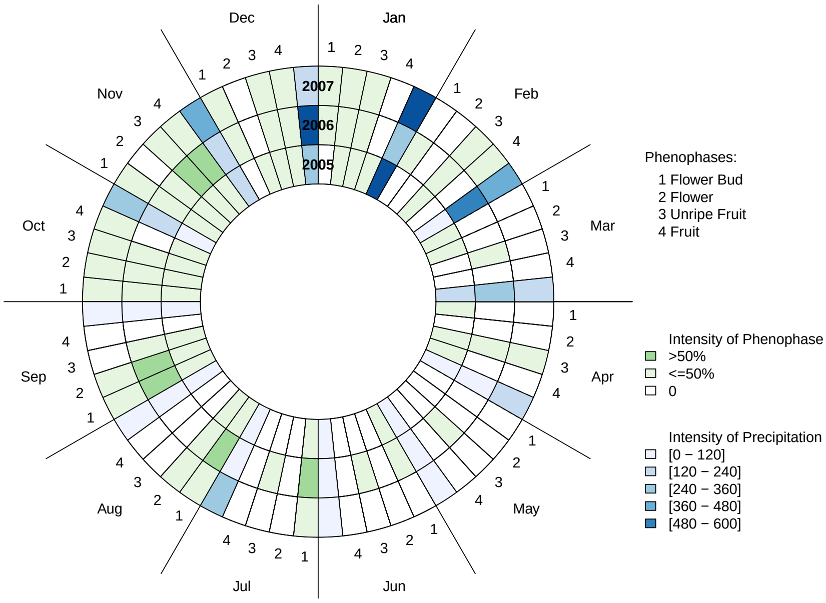
Figure 3.15: Screenshot of the information system developed based on the evaluations. In (a), all phenophases and the variation of precipitation are presented from 2004 to 2007 (same data showed for Prototype 4 at summarization level).

Figure 3.18 shows the results for the gcc variation over 2014, 2015, and 2016, associated with an individual of species *Caryocar brasiliensis*.

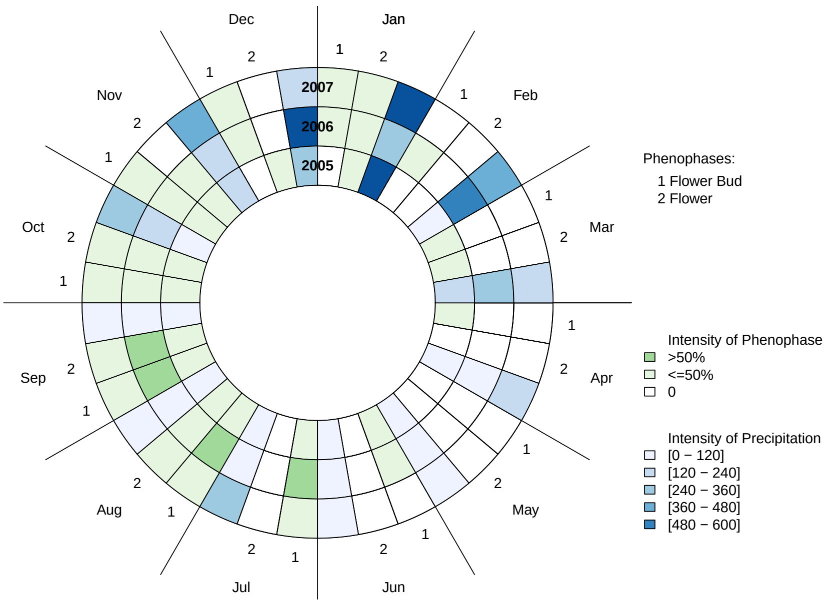
3.3.6 Conclusion

This chapter has introduced a novel visualization approach for presenting cyclical multidimensional temporal data associated with phenology studies. Our proposal combines a radial visualization with the visual rhythm approach, which are applied in numerical and climate data and extensible to images data. The first results obtained show that our proposal was able to support the analysis of phenological data, mainly when the experts are interested in understanding the relations among multiple variables associated with plant life cycle events. Furthermore, we were able to identify the need of creating interaction mechanisms to support phenology experts in tasks such as data filtering (i.e., by year, phenophase, and type of data) and data browsing. The objective is to improve their experience in the understanding of complex temporal change patterns.

In summary, we presented in this chapter a comprehensive formal description of the combination procedure, concerning the use of both phenological and climate data, from a database perspective and a possibility to use the same approach for data obtained from sequences of vegetation images. The objective is to guide researchers and developers in the creation of novel realizations and extensions of the proposed approach for managing multidimensional data in different applications.



(a)



(b)

Figure 3.16: Screenshots of the information system developed based on the evaluations. In (a), data from Figure 3.15 are filtered based on the selection of reproduction-related phenophases. In (b), data from Figure 3.15 are filtered based on the selection of two phenophases (flower bud and flower).

Figure 3.17: Example of daily vegetation image taken in the context of e-phenology project, and the mask (blue shadow) for a species crown, that represents a region of interest (ROI). Based on the processing of the ROI within the image, we can extract some temporal features for the vegetation, for instance, the green chromatic correlation Gcc variation [4, 5], that corresponds to the leafing phenology.

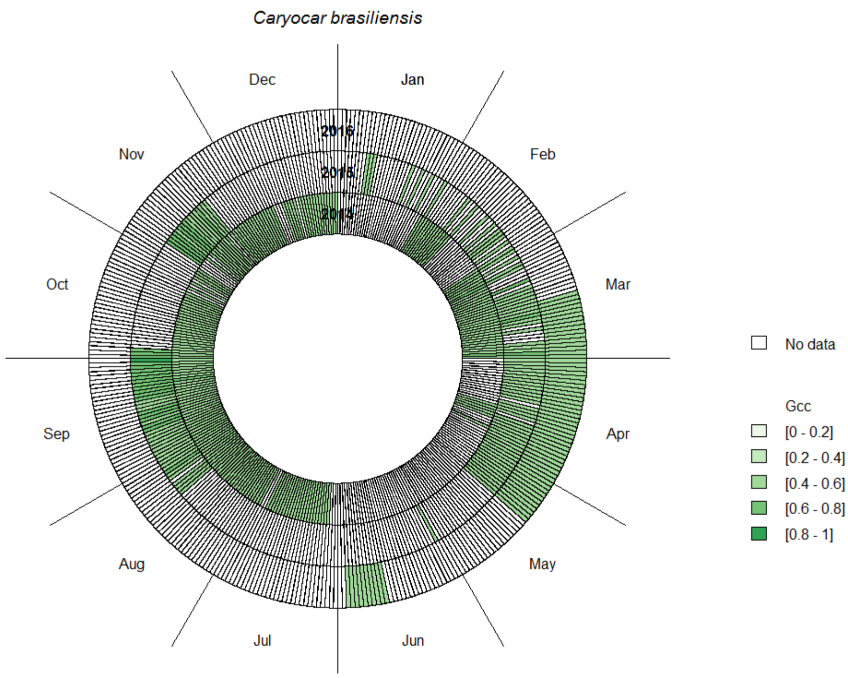


Figure 3.18: Screenshot of the use of the radial structure layout for encoding data about vegetation indices (e.g., Gcc) extracted from a sequence of images of a region of interest (ROI) that represents the species Cariocar brasilienses (see Figure 19 for details).

Chapter 4

Change Frequency Heatmap

This chapter presents the proposed approach to encode temporal change patterns. First, we introduce the method Change Frequency Heatmap (Section 4.1), where we describe the matrix representation used to encode temporal multivariate data (Section 4.1.1), how binary representations are created based on temporal multivariate data (Section 4.1.2), and how they are used to encode change patterns into an image (Section 4.1.3). In the following, we discuss about our case studies (Section 4.2) presenting the dataset (Section 4.2.1) and how the duration of cycles (Section 4.2.2) and phenophase onset date (Section 4.2.3) are encoding using heatmaps.

4.1 Introduction

Innovative technologies for sensing, storing, and sharing have enabled the accumulation of huge volumes of temporal multivariate data. This scenario has demanded the creation of appropriate tools for supporting the decision-making process based on data analysis. In special, the understanding of multivariate temporal data changes is of paramount importance in different applications, such as sport dynamic characterization [105], urban planing [131], social network evolution analysis [89], and phenology studies [82, 90, 93]. However, existing analysis approaches are usually not suitable for handling typical challenges faced, which are related to the data volume in terms of the number of records and variables considered and the length of time series.

This chapter addresses those challenges by introducing a novel image-based representation to encode complex changes in multivariate temporal data. The proposed approach, named Change Frequency Heatmap (CFH), encodes the histogram of change patterns, allowing effective and efficient representation of large volumes of multivariate temporal data into a single image. The definition of CFH generalizes the well-known Motion History Histogram (MHH) [76], usually employed in complex video analysis based on motion changes [21]. The generalization consists in allowing the computation of change patterns not only associated with motion activities in videos, but also with any temporal change observed in multivariate conventional numerical data.

We validated the use of CHFs in the context of phenology studies. Generally, in phenology studies, experts handle well-defined expected temporal patterns (see Figure 4.1),

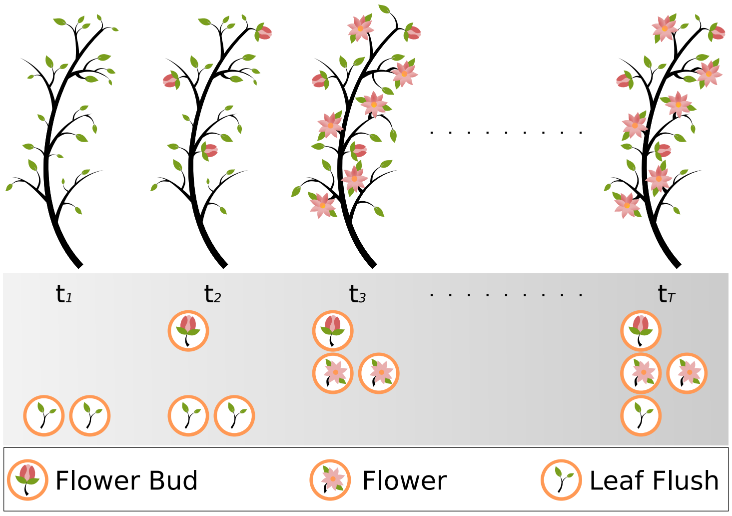


Figure 4.1: Intensities for three different phenophases (flowering: flower bud and flower; and leaf flush) for a particular individual over time, indicated by the number of icons of a giving phenophase.

such as *when* flower buds, flowers, immature and mature fruits or new leaves began to appear; *how long* a particular phenomenon lasts, and *how is the pattern* observed for individuals of the same species located within a same area or in different regions. In this chapter, we analyze the different temporal change patterns associated with reproductive phenophases (flowering and fruiting) of plant species individuals. We conducted two case studies concerning the use of CHFs in the identification of *when* a phenophase starts and for *how long* it lasts.

4.1.1 Temporal Multivariate Data Matrix Representation

Let $\mathcal{S} = \{X_1, X_2, \ldots X_n\}$ be a set with n elements $X_i =< x_{i,1}, x_{i,2}, \ldots, x_{i,m} >$ with m dimensions. For example, $\mathcal{S}$ may refer to a set of n plant individuals, each one characterized by the intensity observed of m particular phenophase activities. Figure 4.1 illustrates an example of observed intensities, over time, of three phenophases, flower bud, flower, and leaf flush for a plant individual.[1] For example, at timestamp t_1, only the leaf flush phenophase was observed with intensity equal to 2 (there are two icons). At timestamp t_2, in turn, the phenopase flower bud was observed with intensity 1, while phenophase leaf flush was observed with intensity equal to 2.

We represent set $\mathcal{S}$ as a temporal matrix $\mathcal{A}$, where $\mathcal{A} =< A_1, A_2, \ldots, A_T >$, where A_t (represented below) is an $n \times m$ numerical matrix at timestamp $t \in$[1,T] composed of n lines and m columns. For each instant of time, we say that we have a *instant matrix* A_t.

$$A_t = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{13} & \ldots & a_{1,m} \\ a_{2,1} & a_{2,2} & a_{23} & \ldots & a_{2,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & a_{n,3} & \ldots & a_{n,m} \end{bmatrix}$$

[1]Note that the proposed approach can be easily extended to handle data related to a multidimensional variables associated with several individuals.

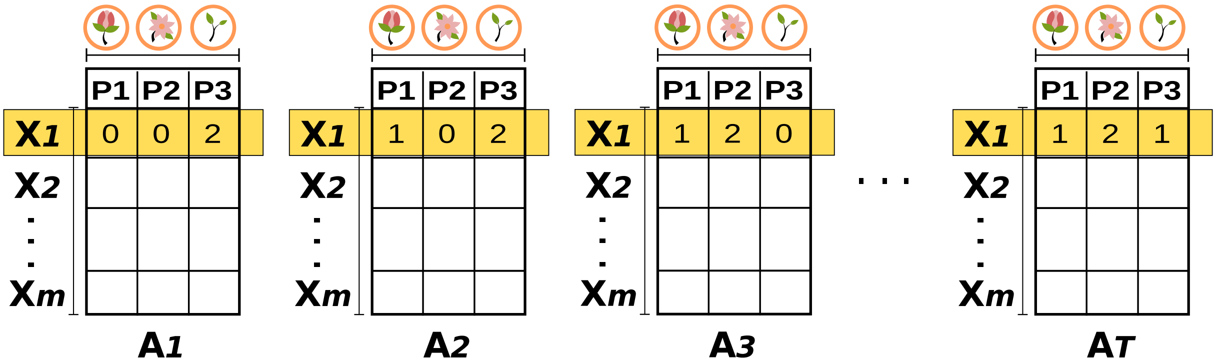


Figure 4.2: Temporal multivariate data representation in *temporal matrices*. The yellow highlights the phenophases intensities of the plant individual in Figure 4.1.

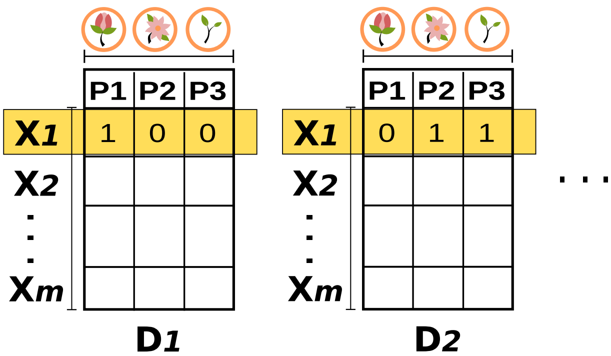


Figure 4.3: Binary pattern representation related to the of use of Eq. 4.2 on the first three temporal matrices of Figure 4.2.

Figure 4.2 illustrates the use of temporal matrices to encode temporal multivariate data. We highlight in yellow the phenophase intensities related to to the individual considered in the example showed in Figure 4.1.

4.1.2 Binary Pattern Representation

Our goal is to represent temporal changes in a sequence of numerical matrices. A change frequency heatmap is a summarized representation that allows to identify temporal change patterns associated with different temporal matrix positions (l, c), where $1 < l < n$ and $1 < c < m$.

The first step of the method consists in creating a binary image representation of the pattern of interest. Let D be an image representation with dimensions $n \times m \times T$. $D(l, c, t)$, with $1 < l < n$, $1 < c < m$, and $1 < t < (T - 1)$, represents the occurrence of a particular phenomenon of interest at position (l, c), at timestamp t, such that:

$$D(l, c, t) = (b_1, b_2, \ldots, b_{T-1}), b_i \in \{0, 1\} \tag{4.1}$$

Different functions may be used to properly define patterns of interest. For example, the function defined in Equation 4.2 creates a binary representation, which encodes changes over time in a sequence of temporal matrices. Figure 4.3 presents the binary pattern representation defined by the use of Equation 4.2 on the first three temporal matrices of Figure 4.2. The function defined in Equation 4.3, in turn, refers to a pattern associated with abrupt changes defined in terms of a threshold $\mathcal{T}$.

$$D_1(l,c,t) = \begin{cases} 1, & A_t(l,c) \neq A_{t+1}(l,c) \\ 0, & \text{otherwise} \end{cases} \tag{4.2}$$

$$D_2(l,c,t) = \begin{cases} 1, & A_{t+1}(l,c) - A_t(l,c) \geq \mathcal{T} \\ 0, & \text{otherwise} \end{cases} \tag{4.3}$$

4.1.3 Change Frequency Heatmap Computation

Depending on application, we define patterns of interest P_i in the $D(l,c,t)$ sequences, based on the number of '0's or '1's connected. Typical binary patterns include $P_1 = 010$, $P_2 = 0110$, $P_3 = 01110$, etc. The number of '1's in the pattern of interest is defined by u, $1 < u < U$.

Similarly to [76], we denote a subsequence $C_{I,k}$ by Equation 4.4, where I and k are the indices of starting and ending matrices and denote the set of all subsequences of $D(l,c,:)$. Thus, for each position (l,c), we can count the number of occurrences of each specific pattern P_i in the sequence $D(l,c,:)$ as proposed by [76] and shown in Equation 4.5.

$$C_{I,k} = (b_1, b_{1+2}, \ldots, b_n), \tag{4.4}$$

where $1 \leq I < k \leq T-1$.

$$CFH_{(l,c,u)} = \sum_{I,k} \mathbb{1}_{\{C_{I,k}=P_u | C_{I,k} \in \Omega\{D(l,c,:)\}\}}, \tag{4.5}$$

where $1 \leq I < k \leq T-1, 1 \leq u \leq U$, $\mathbb{1}$ is the indicator function, and $\Omega\{D(l,c,:)\}$ is the set of all subsequences of $D(l,c,:)$.

Finally, from each pattern of interest P_i, we can create a heatmap and the collection of heatmaps (for all patterns P_i defined) we named them as *Change Frequency Heatmaps* representation. Algorithm 1 outlines the main steps for computing Change Frequency Heatmaps. In the algorithm, $I(l,c)$ is matrix used to define at which timestamp t, the last pattern was found.

The next section describes conducted case studies concerning the use of CHFs in phenology.

4.2 Case Studies: Temporal Change Patterns in Phenology Studies

The case studies are concerned with the use of CFHs for the representation and identification – in a summarized way – of relevant change events on conventional temporal phenological data associated with individuals of two different species. CFHs are expected to encode change patterns related to phenological activity, allowing the analysis of *how* different individuals, which belong to the same species, behave over time.

Algorithm 1 Change Frequency Heatmaps

Input: file Matrix $M(l,c,t)$, $l = 1,\dots,n$, $c = 1,\dots,m$, $t \in T$, where $t = 0,1,\dots,T$
Initialization: pattern P, $HH_i(1:L\ ,1:C)$, $i = 1,\dots,P$
Compute: $D(:,:,t)$
for $t = 1$ to T **do**
    for $l = 1$ to n **do**
        for $c = 1$ to m **do**
            if subsequence $D(l,c,I(l,c)),\dots,D(l,c,t){=}P_i$ **then**
                Update: $CFH_i(l,c) \leftarrow CFH_i(l,c) + 1$
            end if
            Update: $I(l,c) \leftarrow t$
        end for
    end for
end for
Output: return P

Figure 4.4 illustrates how the CFHs are computed in a typical phenology study. One species may have one or more individuals observed over time. We define a sequence of individuals from one species by $\mathcal{S} = \{X_1, X_2, X_3, \dots, X_n\}$. For each individual X_i, one or more phenophases can be observed at the same instant of time. In the case of our data set, up to six phenophases are observed, then we define a set of phenophases as $P = \{P_1, P_2, P_3, P_4, P_5, P_6\}$. Therefore, each matrix defined at a particular timestamp t (say M_t) represents a set of phenophase intensity observations from one or more individuals of the same species.

Generally, in phenology studies, experts are interested in specific temporal patterns, such as *when* flower buds, flowers, immature and mature fruits or new leaves began to appear; for *how long* a particular plant life cycle phenomenon lasts; and *how is the typical behavior* of individuals of the same species located at both the same area and different places [16, 26, 90, 117, 123]. More specifically, we define three main time-based behavior patterns that are of interest:

1. *Change:* definition of *when* a phenophase changed its intensity of activity or for *how long* a particular phenophase lasted.

2. *Increasing:* definition of *when* a phenophase started or *when* a phenophase changed to a higher intensity of activity.

3. *Decreasing:* definition of *when* a phenophase finished or *when* a phenophase changed to a lower intensity of activity.

Based on these patterns of interest, we propose to compute the binary matrix D (see Section 4.1.2), by considering the relation among intensity values of phenophases for successive timestamps (for example, matrices M_t and M_{t+1}). For example, to represent a change of phenophase intensities, i.e., to identify if there is a *change* at instant t in the position (l,c), we may compare the intensity scores of two successive matrices (say M_t and M_{t+1}). In this case, $D(l,c,t) = 1$, if there is a change and $D(l,c,t) = 1$, otherwise (see

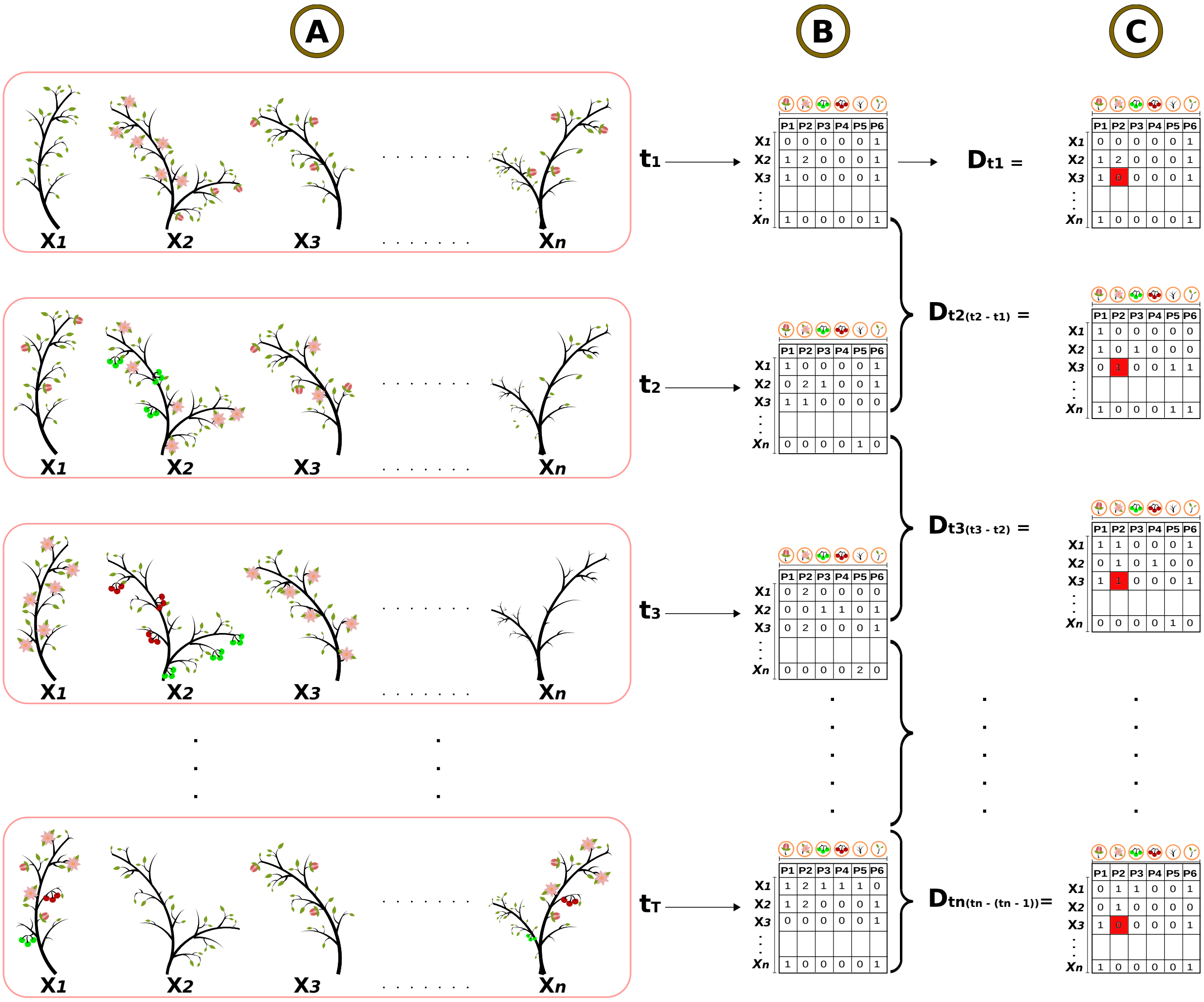


Figure 4.4: Change Pattern Heatmap computation in a typical multi-individual-based phenology study.

Equation 4.6). Equations 4.7 and 4.8 exemplify two functions to represent the *increasing* and the *decreasing* of phenophase intensities, respectively, at two consecutive timestamps.

$$D_{change}(l,c,t) = \begin{cases} 1 & \text{if } (M_t(l,c) \text{ - } M_{t+1}(l,c)) \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.6}$$

$$D_{increasing}(l,c,t) = \begin{cases} 1 & \text{if } (M_{t+1}(l,c) \text{ - } M_t(\text{l,c})) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.7}$$

$$D_{decreasing}(l,c,t) = \begin{cases} 1 & \text{if } (M_{t+1}(l,c) \text{ - } M_t(\text{l,c})) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.8}$$

In Figure 4.4 part C, we highlight in red the resulting values in matrix D (defined in Equation 4.6), considering individual X_3 and phenophase *flower* (P_2). The associated binary pattern is illustrated in Equation 4.9.

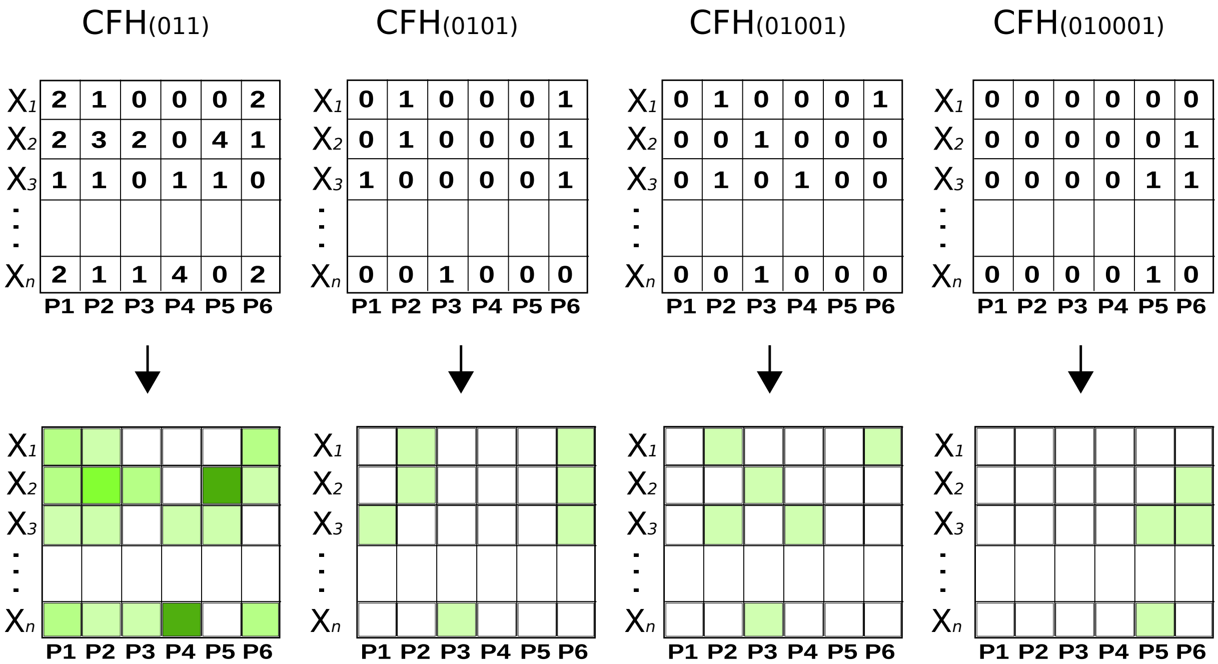


Figure 4.5: Change Pattern Heatmaps for different change patterns: 010, 0110, 01110, and 011110. Lines refer to individuals, while columns are associated with phenophases.

$$D(I_3, P_2, :) = 011001001001 \tag{4.9}$$

By using the sequence calculated in D and applying our algorithm, we can calculate the matrices history heatmaps, as shown in Figure 4.5 for the patterns 010, 0110, 01110, and 011110. Another binary patterns can be defined according to the goal of a study. For instance, in our study cases, we defined the pattern 010 to identify when a phenophase began and we defined the patterns 011, 0101, 01001, 0100 . . . 1 to identify for how long a phenophase ocurred.

4.2.1 Dataset

As in the case of the previous study, we opted for published data to validate our approach, mainly because, in this study, we did not perform evaluations with users. We validated our approach comparing our results with published results and based on reviews made by phenology experts. Thus, we are using on-the-ground monthly phenological observations of individuals from two plant species *Trembleya laniflora* Cogn. (Melastomataceae) (92 individuals) and *Myrcia guianensis* (Aubl.) Kuntze (Myrtaceae) (90 individuals) situated in two distinct natural areas. *T. laniflora* reproductive phenology observations were realized along 2014 in a rupestrian grassland area at *Serra do Cipó*, Minas Gerais State, Brazil [113], while *M. guianensis* observations were collected during the period from 2005 to 2011 in a *Cerrado* area in Itirapina, São Paulo State, Brazil [123].

Trembleya laniflora (Melastomataceae) and *Myrcia guianensis* (Myrtaceae) belong to two of the most important and representative families in rupestrian grasslands and *Cerrado*, respectively [49, 102, 112] and present annual flowering and fruiting patterns [113,

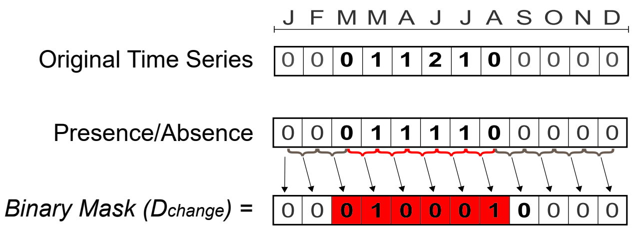


Figure 4.6: Example of the use Equation 4.6: indicating a cycle of duration of four months for one phenophase.

123]. *Trembleya laniflora* reproduces mainly during the dry and post-dry seasons, flowering from May to October and fruiting from June to December [113], while *M. guianensis* flowering season occurs during the transition between dry and wet seasons (from July to November) and fruiting during the wet season (from September to January) [123].

We aim to identify and compare the occurrence time (onset date and period) and duration of reproductive phenophases (flowering and fruiting) for *T. laniflora* and *M. guianensis* individuals. However, we are not interested in quantifying the phenophase intensity and therefore, in this study, quantitative and semi-quantitative phenological data (intensity indexes) [44,96] were converted in presence and absence data (0 and 1) [18]. We validated the CFHs results using on-the-ground direct phenology data upon both species and comparing the change patterns we found with the reproductive phenology described for *T. laniflora* by [113] and *M. guianensis* by [123].

4.2.2 Duration of cycles: *Trembleya laniflora* and *Myrcia guianensis*

In this study, we are interested in comparing the change patterns observed for individuals of those species over time, more specifically, to know for *how long* one phenophase lasted. In this case, we use Equation 4.6 to encode the binary patterns. Figure 4.6 illustrates how we calculate the final binary mask using this equation. Note that an intermediate step is conducted to identify the presence or absence of a phenological activity, given the original time series. This intermediate encoding is used as input for identifying temporal pattern of interest (change). We defined the patterns 011, 0101, 01001, 0100 . . . 1 to identify the duration in months of a particular phenophase.

Short-term analysis

Figures 4.7 and 4.9 present CFHs for pattern 011 for *Trembleya laniflora* and *Myrcia guianensis*, respectively. This pattern refers to the frequency of occurrence of one-month duration phenophases for different individuals of both species along one year. The one-month pattern is more frequently observed for individuals of *Myrcia guianensis*, especially for the Flower bud phenophase. Figures 4.8 and 4.10 illustrate the occurrence of a longer pattern (four-month phenophase duration) for *Trembleya laniflora* and *Myrcia guianensis*, respectively. The CFHs, in this case are very different. Individuals of *Trembleya laniflora*

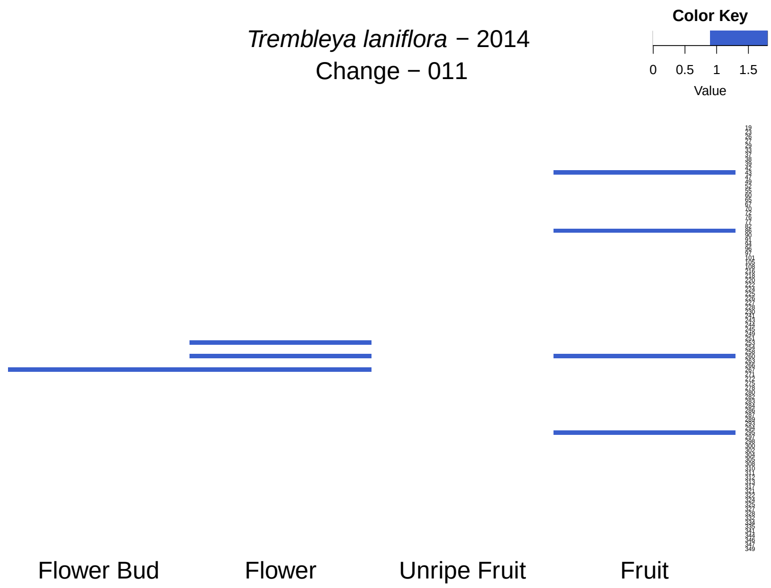


Figure 4.7: Pattern 011: Change pattern indicating a one-month duration of different phenophases of individuals of the species *Trembleya laniflora* Cogn. (Melastomataceae).

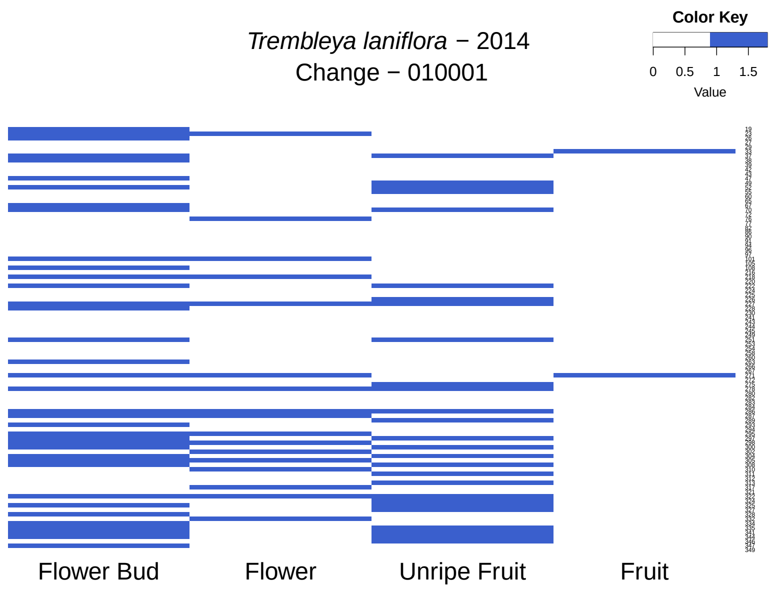


Figure 4.8: Pattern 010001: Change pattern indicating a four-month duration of different phenophases of individuals of the species *Trembleya laniflora* Cogn. (Melastomataceae).

presented the long four-month pattern more frequently than the individuals of *Myrcia guianensis*. In summary, we conclude that individuals of *Trembleya laniflora* have longer reproductive pattern than the ones of *Myrcia guianensis*.

Long-term Analysis

The same patterns 011, 0101, 01001, 0100...1 were used to identify the duration in months of phenophases for *M. guianensis* individuals during the period from 2005 to 2011 (Figures 4.11 and 4.12). In this case, we are interested in testing defined patterns

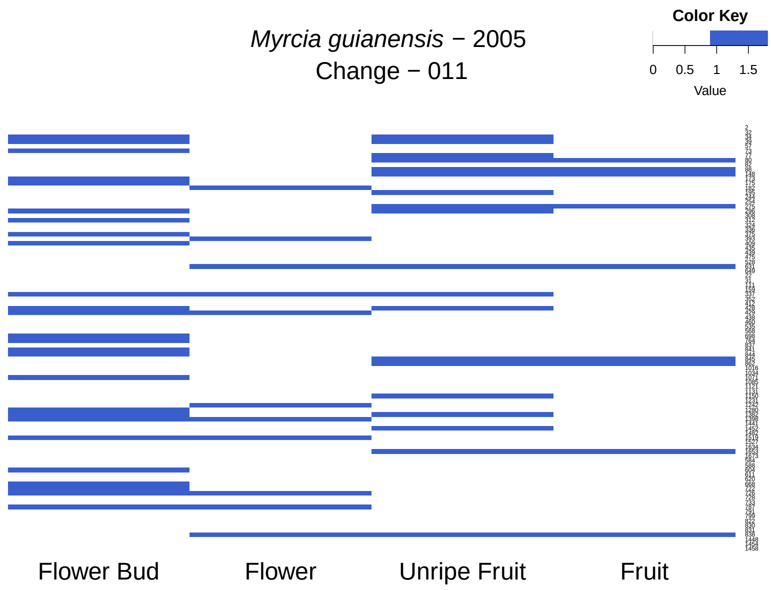


Figure 4.9: Pattern 011: Change pattern indicating a one month duration of different phenophases of individuals of the species *Myrcia guianensis* (Aubl.) Kuntze (Myrtaceae).

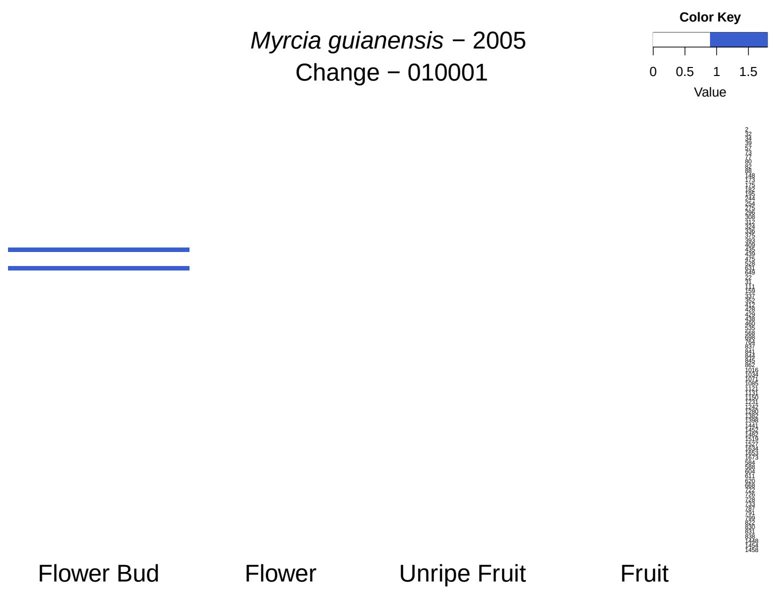


Figure 4.10: Pattern 010001: Change pattern indicating a four-month duration of different phenophases of individuals of the species *Myrcia guianensis* (Aubl.) Kuntze (Myrtaceae).

upon long-term data with the objective of identifying if the shorter reproductive period found for *M. guianensis* individuals is maintained over the years. Figure 4.11 illustrates a frequent occurrence of short events (one-month duration) of flowering (flower bud and flower) and fruiting (immature fruit and fruit) for *M. guianensis* individuals over time. Comparing Figures 4.11 and 4.12, we observe that short events are more common between the individuals than longer events. This is in accordance with previous studies concerning the reproductive phenology of those species [123]. The CFHs enable to identify the individuals that present short cycles more frequently over time and additionally allow to

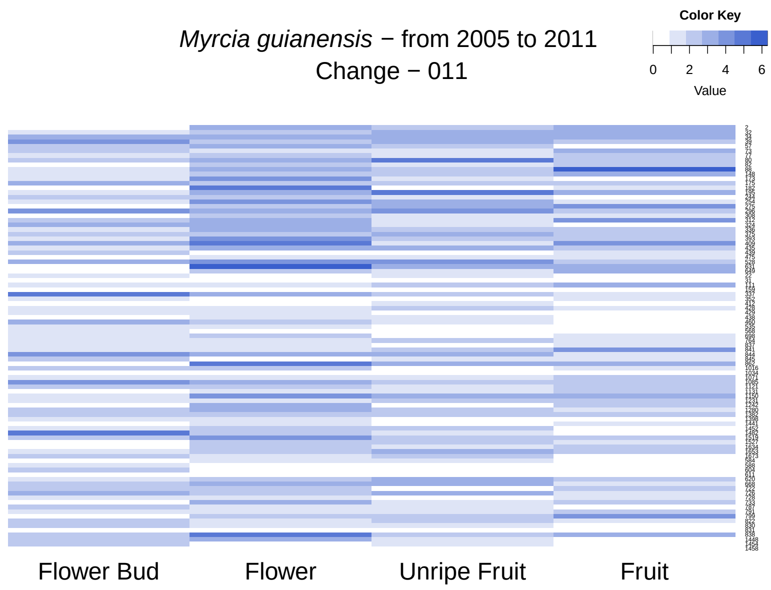


Figure 4.11: Pattern 011: Change pattern indicating a one month duration of a different phenophases of individuals of the species *Myrcia guianensis* Cogn. (Melastomataceae) from 2005 to 2011 period.

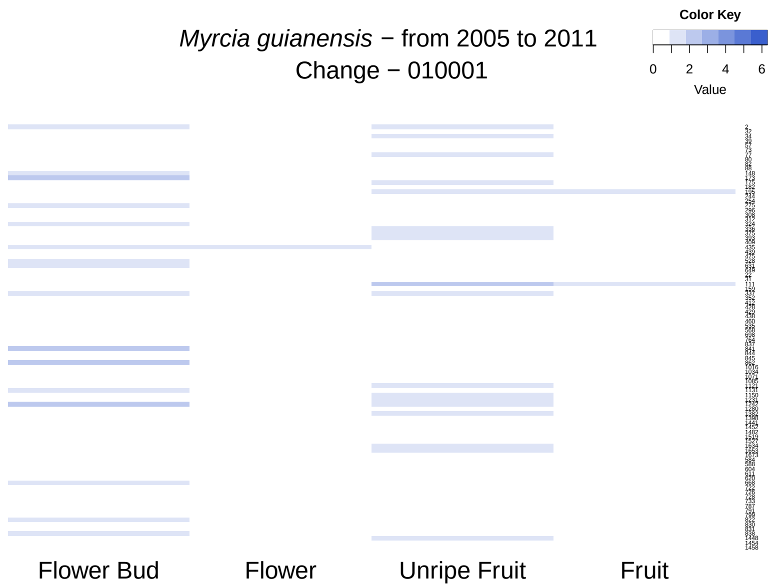


Figure 4.12: Pattern 010001: Change pattern indicating a four-month duration of a different phenophases of individuals of the species *Myrcia guianensis* Cogn. (Melastomataceae) from 2005 to 2011 period.

visualize the intra-populational differences.

4.2.3 Phenophase Onset Date

In this case study, we are interested in analyzing the patterns associated with the beginning of phenophases. We use again the same dataset of the previous section. However, now we compute the binary pattern encoded by $D_{increasing}$ (Equation 4.7) and we used the

Figure 4.13: Example of the use of Equation 4.7 for indicating that the onset of the cycle of phenophase occurred in April.

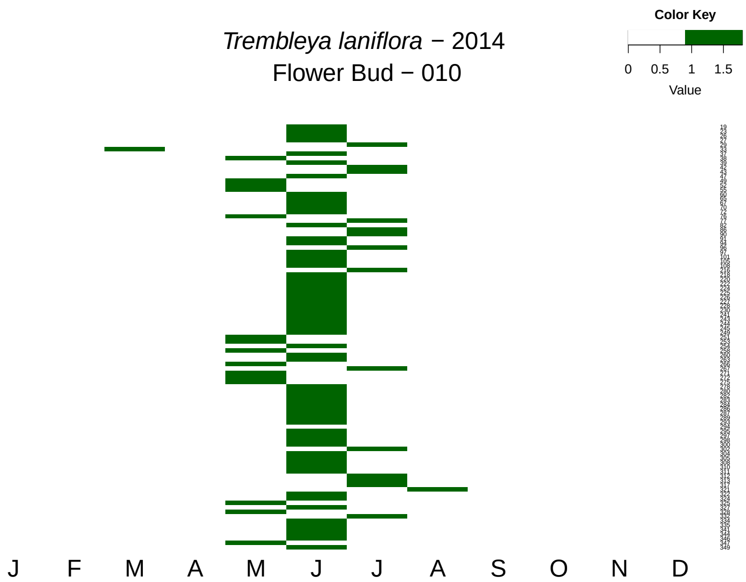


Figure 4.14: Pattern 010: Flowering onset date (flower bud) for *Trembleya laniflora* Cogn. (Melastomataceae) individuals.

pattern 010 to identify *when* the phenophase started or the reproductive onset. Figure 4.13 illustrates how we calculate the final binary encoding used to find when a phenophase cycle started. Again, we take advantage of a an intermediate representation, which encodes the presence or absence of phenological activity.

Short-term analysis

Figures 4.14, 4.15, 4.16, and 4.17 refer to the CFHs for flowering (flower bud and flower) and fruiting (immature fruit and fruit), respectively, of *T. laniflora* individuals. Figures 4.18, 4.19, 4.20, and 4.21 refer to the same reproductive phenophases, but now considering *M. guianensis* individuals. Note that the temporal multivariate data are organized in a different manner. Variables (CFH columns), in this case, refer to months of the year.

We can identify the phenophases onset dates of individuals, detect intraspecific variations in the populations' phenology, and infer the time of occurrence of the reproductive cycles species. For example, we can observe that the flowering (flower bud and flower) starts mainly in June and July for *T. laniflora* and in August and September for *M.*

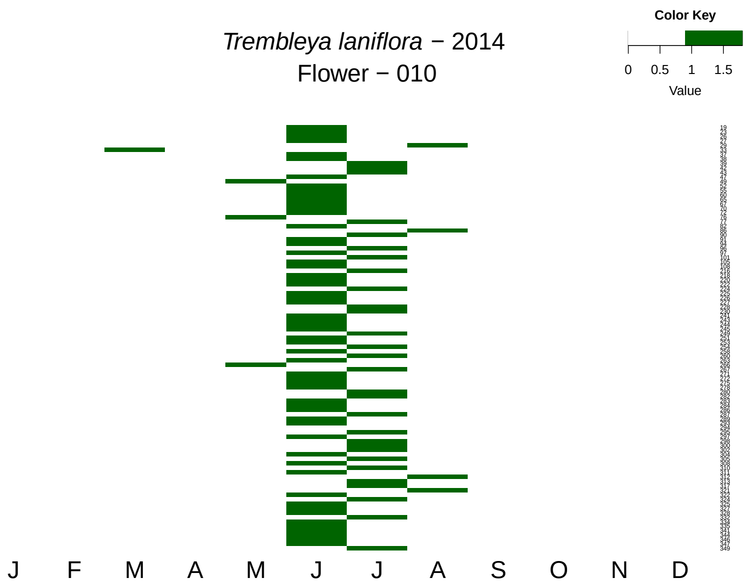


Figure 4.15: Pattern 010: Flowering onset date (flower) for *Trembleya laniflora* Cogn. (Melastomataceae) individuals.

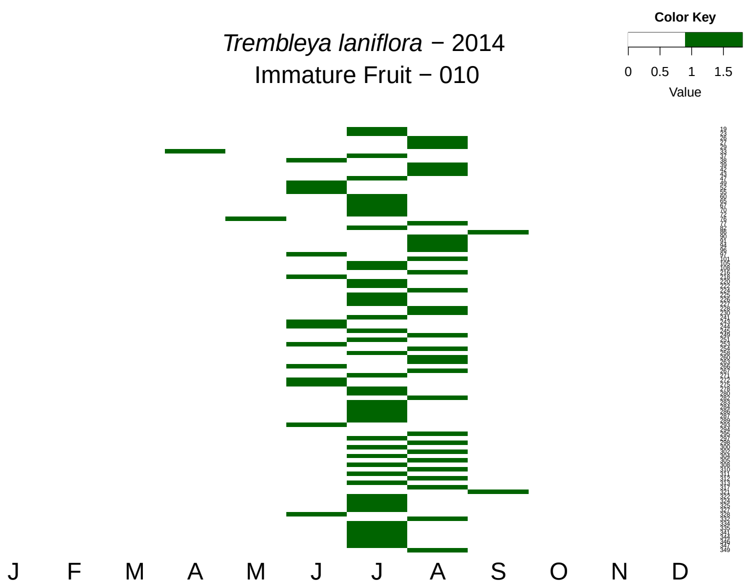


Figure 4.16: Pattern 010: Fruiting onset date (immature fruit) for *Trembleya laniflora* Cogn. (Melastomataceae) individuals.

guianensis as before described for the two species [113, 123]. We also observe that some individuals present advanced or delayed flowering and fruiting responses when compared to the population mean. Therefore, we can detect intraspecific variations in the phenology of *T. laniflora* and *M. guianensis* that would be difficult to evaluate with the typical mean-based population analysis (e.g., mean angle, mean date, activity index) usually employed [18, 80, 84].

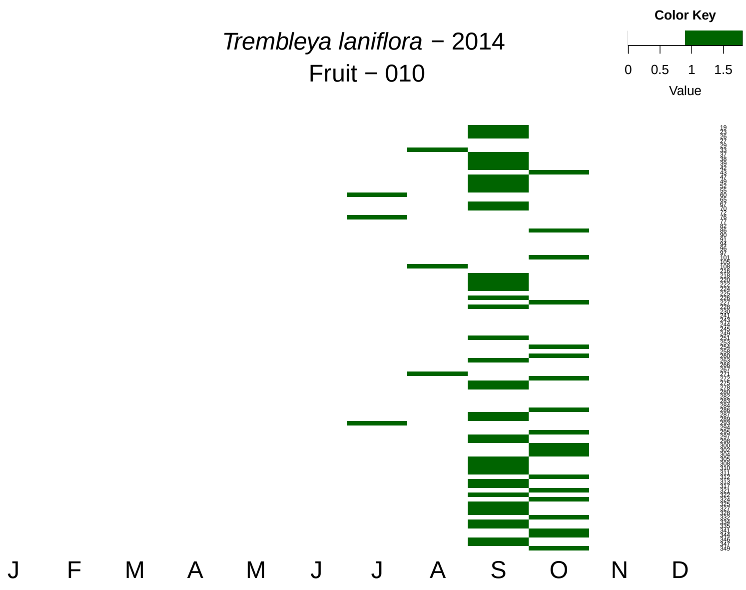


Figure 4.17: Pattern 010: Fruiting onset date (fruit) for *Trembleya laniflora* Cogn. (Melastomataceae) individuals.

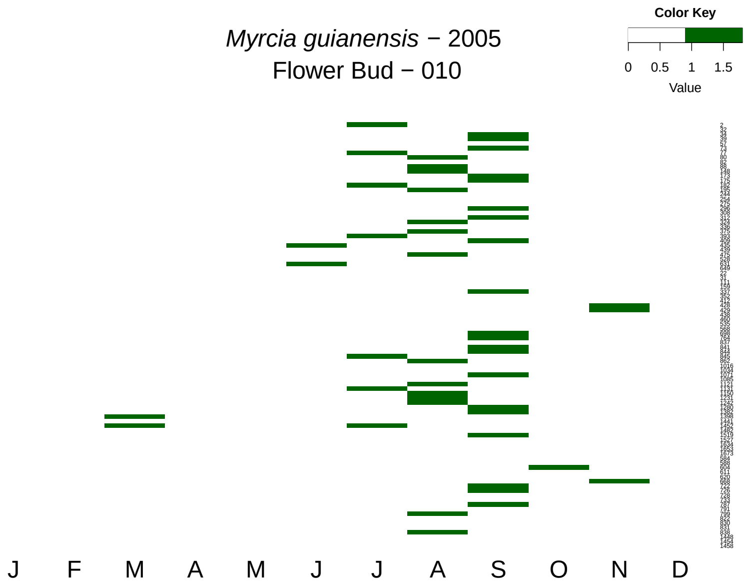


Figure 4.18: Pattern 010: Flowering onset date (flower bud) for for *Myrcia guianensis* (Aubl.) Kuntze (Myrtaceae) individuals.

Long-term Analysis

In this section, we are interested in analyzing the patterns associated with the beginning of phenophases to long-term observations at *M.guianensis* individuals. Figures 4.22, 4.23, 4.24, and 4.25 demonstrate phenophases onset dates (months) of individuals during the seven years of study (from 2005 to 2011). It is possible to identify the months presenting more individuals in flowering and fruiting, and visualize the species reproductive main period and the individual variation patterns over time. We can identify flowering responses mainly during July (flower bud), August and September (flower),

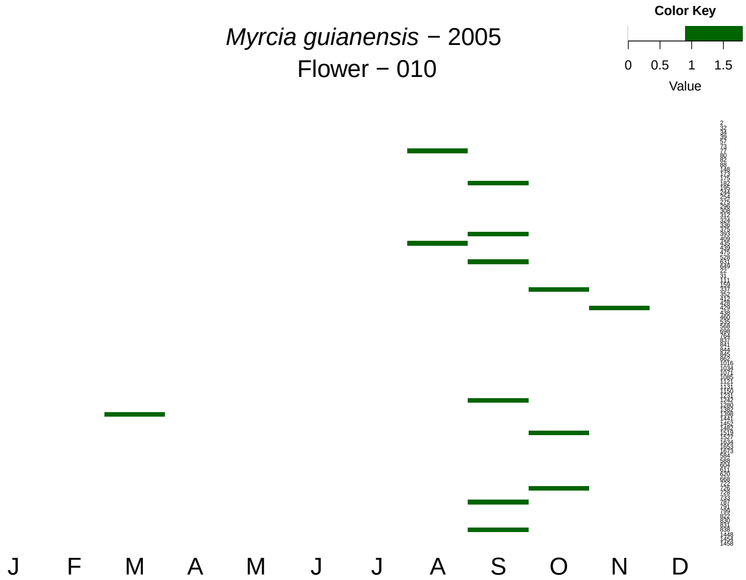


Figure 4.19: Pattern 010: Flowering onset date (flower) for *Myrcia guianensis* (Aubl.) Kuntze (Myrtaceae) individuals.

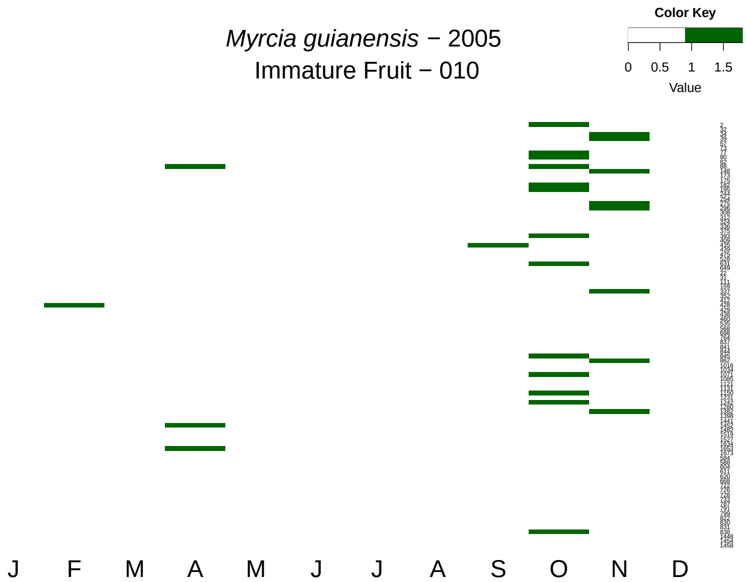


Figure 4.20: Pattern 010: Fruiting onset date (immature fruit) for *Myrcia guianensis* (Aubl.) Kuntze (Myrtaceae) individuals.

and fruiting along October and November in accordance with the previously tested pattern to the species in the same period (mean angles; Rayleigh test) [123]. Additionally, it was possible observed small phenological intrapopulational variations. Some individuals presented different phenophases onset dates over the years, that can be related, for example, to distinct responses the population to annual climatic variations [100], or temperature and light differential conditions [123]. In fact, with our new approach analyzing the phenophases, it was easier to understand how the whole reproductive cycle of each individual and species worked over time.

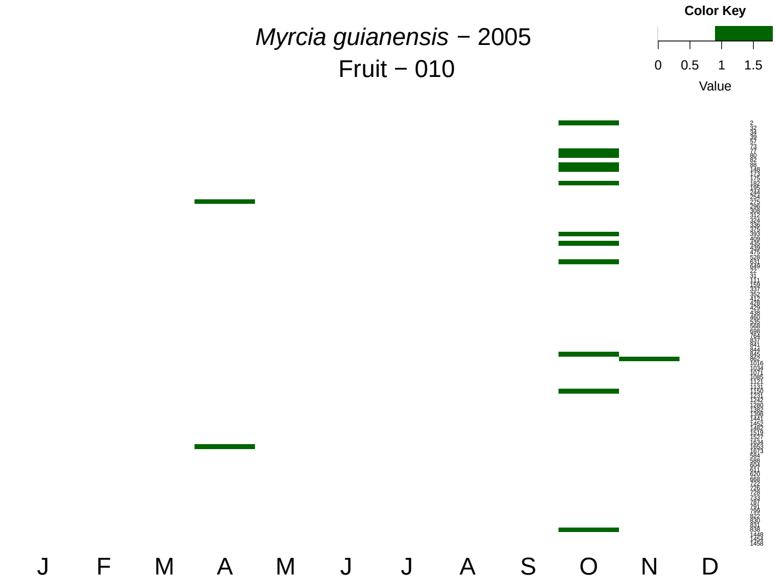


Figure 4.21: Pattern 010: Fruiting onset date (fruit) for *Myrcia guianensis* (Aubl.) Kuntze (Myrtaceae) individuals.

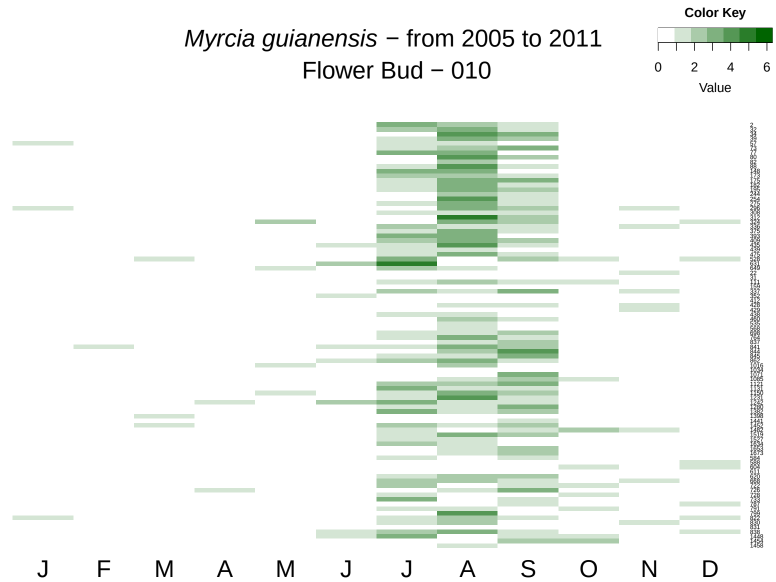


Figure 4.22: Pattern 010: Flowering onset date (flower bud) for *Myrcia guianensis* Cogn. (Melastomataceae) individuals from 2005 to 2011 period.

4.3 Discussion

The new temporal pattern characterization approach enabled to evaluate the duration and onset date of different phenophases for *T. laniflora* and *M. guianensis* individuals during short and long-term periods of observations. The patterns identified here were consistent to the patterns previously tested and described for both species [113, 123]. CFH provided a simple analysis of individual responses, facilitating the detection of flowering and fruiting phenological patterns at the species level, and the assessment of populational and individual variations. It was possible to visualize independently the parameters for

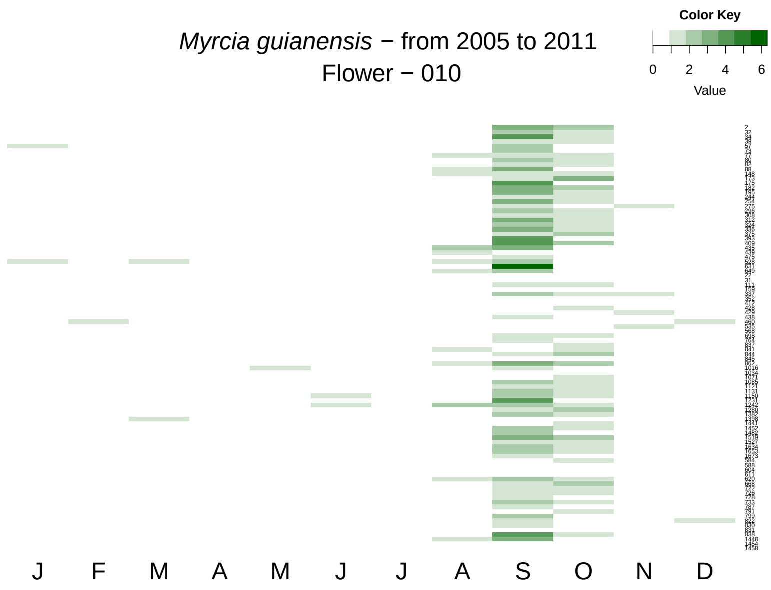


Figure 4.23: Pattern 010: Flowering onset date (flower) for *Myrcia guianensis* Cogn. (Melastomataceae) individuals from 2005 to 2011 period.

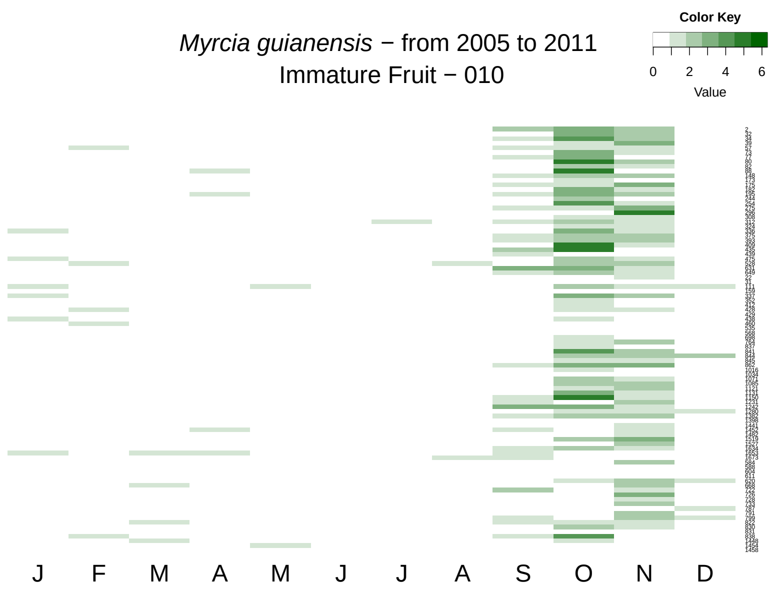


Figure 4.24: Pattern 010: Fruiting onset date (immature fruit) for *Myrcia guianensis* Cogn. (Melastomataceae) individuals from 2005 to 2011 period.

each sampled individual and evaluate the variations within the population over time. Populational phenological patterns, in general, are demonstrated by mean values such as Fournier's intensity index, activity index, mean angle, and mean date (e.g., [26, 80, 123]), which are not able to indicate the intrapopulational variability [95]. However, phenology patterns can differ among individuals of the same population and this variation must be considered when dealing with plant responses to environmental cues and plant-animal interactions [16, 40, 54, 99]. Thereby, the new temporal change representation proposed here makes easier the evaluation of intraspecific variations (mainly intrapopulational) in

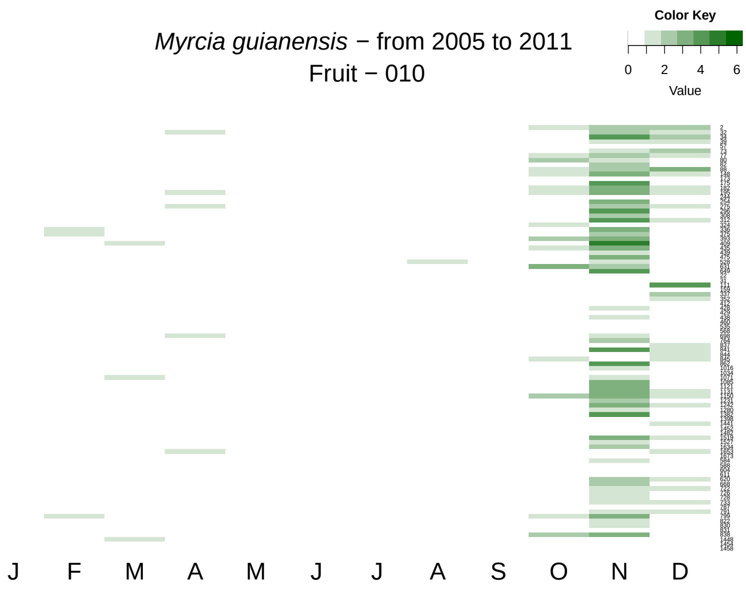


Figure 4.25: Pattern 010: Fruiting onset date (fruit) for *Myrcia guianensis* Cogn. (Melastomataceae) individuals from 2005 to 2011 period.

the species phenology and contributes to the analysis of the phenological patterns. One underexploited issue is the specialization within populations [13, 22, 54] and the possibility, for instance, of divergence of a few individuals in their reproductive niche (time of flowering) consistently over time, leading to a reproductive isolation. Also, CFH is a important tool to evaluate individual responses of plants to changing environment such as fragmentation and climate change [53].

4.4 Conclusions

This chapter introduced Change Frequency Heatmaps (CFHs), a novel image-based representation for encoding change patterns found in temporal multivariate data. The main advantages of the proposed temporal change representation relies on its flexibility and effectiveness in representing the most relevant features, in terms of the frequency of occurrences, of the patterns observed in large volumes of data, considering the number of objects handled in the study, the number of variables, and the length of time series. The case studies' analyses conducted based on real phenological data demonstrated the potential of the proposed image representation to support visual identification of complex temporal change patterns, especially suitable for deciphering intraspecific and interindividual variations in phenology patterns. Our research leverages novel opportunities for investigations related to the use of image processing algorithms to highlight important patterns in multivariate temporal data, represented in CFHs.

Chapter 5

Phenology Data Visualization Tool

This chapter introduces a phenological data visualization system, which has been developed based on the integration of the contributions of this work, presented in Chapters 3 and 4. Usage scenarios described in this chapter are based on common tasks related to pehnology studies conducted in the context of the e-phenology project (Section 1.1.2). Therefore, provided examples consider the information system and the databases devised for that project [72].

The remaining of this chapter is organized as follows: Section 5.1 presents an operational and architectural overview for the developed system, while Section 5.2 describes usage scenarios related to the use of the developed tool.

5.1 System Overview

Chapters 3 and 4 presented different approaches to deal with data visualization problems found in phenology studies. In both cases, we developed tools using the language R. Since phenologists are concerned with new systems to support in their data analyses, we propose to incorporate these tools into a user-friendly visualization system, named as phenology data visualization, integrated with a database, and with interaction mechanisms.

This section presents the proposed phenology data visualization tool. Section 5.1.1 presents the architecture used to create the visualization methods. Section 5.1.2 presents the operational view for the methods, emphasizing the importance of an interface through which users can interact with complex temporal data.

5.1.1 Architectural View

Figure 5.1 presents an architectural view of the developed visualization system. The architecture is based on the reference model for information visualization proposed by Card et al. [27]. The first component of the architecture is the *source data*, which is related to the user task. The data source can be a file or a database. In the case of the data source be a file, it must be formatted according to what the visualization system needs. In the case of a database, the system can connect with the database and retrieve data of interest in the desired format by performing queries [72]. Besides that, some kinds of data may need to be transformed to a final format, for example, formatted dates.

The process of preparing the data to the format required by the system is called as *Data Transformations*, while the resulting data, as *Data Tables*. At this point, the handled data are ready to be visualized.

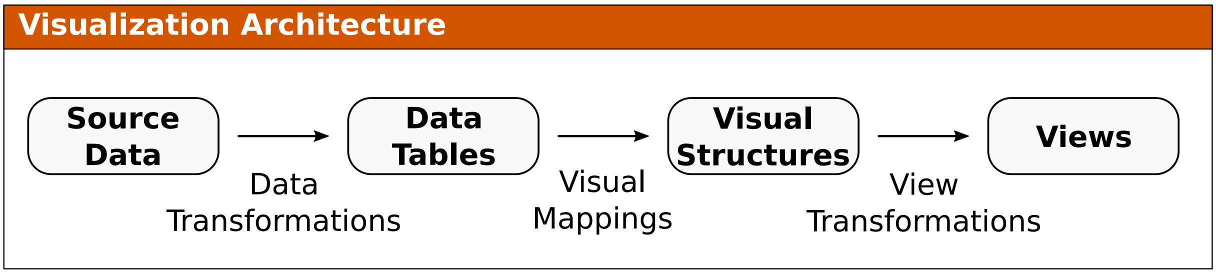


Figure 5.1: Visualization architecture based on reference model for visualization [27].

The next component presented in the architecture is the *Visual Structures*, which depend on which visualization method will be used. The variables from data tables need to be mapped to the visual elements of the chosen visualization method, such as, which variable will be represented and how (e.g., which color, and size). This process is named as *Visual Mappings*. In our tool, the visual mapping is in charge of encoding data tables in a radial layout (Chapter 3) or in a change frequency heatmap image representation (Chapter 4).

Finally, once the data are mapped, the last step is concerned with generating the *Views* for the user. In this phase, the chosen visualization method is presented with the mapped data for the user with some interaction mechanisms. The last two components are implemented in module *Visualization* of the *Interface* of Figure 1.3.

5.1.2 Functional View

Figure 5.2 presents a general functional overview for the proposed phenological data visualization system. The tool was initially developed as a web application using Java[1] language. Currently, it has been migrated to R (main components), with an interface constructed using the Shiny[2] framework, which allows to develop web applications and dashboards using R code underneath.

First, the user define which task will be executed based on some filters shown in the main interface of the system. The task defines the queries that must be performed in the *database*. Based on the user task, the system performs the *data processing*. For example, if the user wants to see on-the-ground phenology observations (considered as numerical data in the context of this thesis) and chooses the radial visualization, the data will be summarized using the visual rhythm approach; but if the user chooses the CFH method, the data will be converted into frequency matrices of occurrences.

Based on the data generated by the processing stage, the next step is to perform the process of *visualization*, as explained in the previous section. Thus, retrieved data need to be mapped to the selected visualization method and interactions mechanisms need to

[1] `www.java.com` (As of July 25, 2018).
[2] `www.shiny.rstudio.com` (As of July 25, 2018).

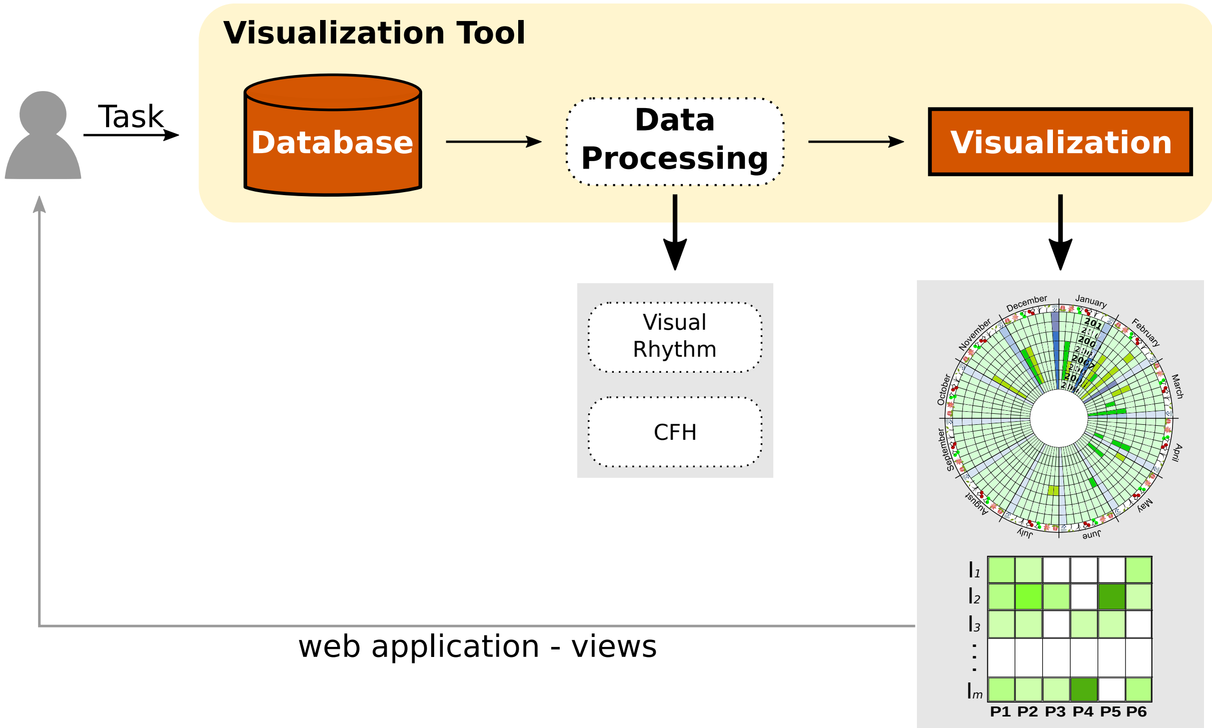


Figure 5.2: Functional overview of the proposed phenological data visualization system.

be defined. In summary, the resulting application is a web interface, integrating data filtering options, visualization approaches, and user interaction mechanisms.

5.2 Usage Scenarios

This section describes a case study, which illustrates the use of this system in a real-world scenario. Our case study refers to vegetation image analysis in the context of the e-phenology project. Provide examples use a dataset related to the images taken by digital camera installed in an area of Cerrado, located in the city of Itirapina, State of São Paulo, Brazil. The images are processed and some features are extracted, such as means of the values for the channels Red, Green, and Blue (RGB), as well as some vegetation indices, such as, Rcc, Gcc, and Bcc (see Sections 1.1.2 and 2.3.3). Often, in their time series analyses based on image data, phenologists choose a specific hour of the day, and use all images at this time, for all days of the year. In most cases, the main time is at noon. In this sense, we use images taken during the period between 2014 and 2016 at noon time.

Figure 5.3 presents the initial interface, from which users may have access to the main filters to define which data and visualization approach are of interest (defining a task). Users are also expected to define what kind of visualization (*radial* or *CFH*) and what kind of data (e.g., for *individuals*, *climatic*, *image* or by *site*) is of interest. According to these selections, extra filters may appear or change. For example, if a user opts to explore *individuals* data, it is necessary also to define, for example, which species, which phenophases, and which period are of interest.

Figure 5.4, in turn, illustrates available filters when a user wants to explore *image*

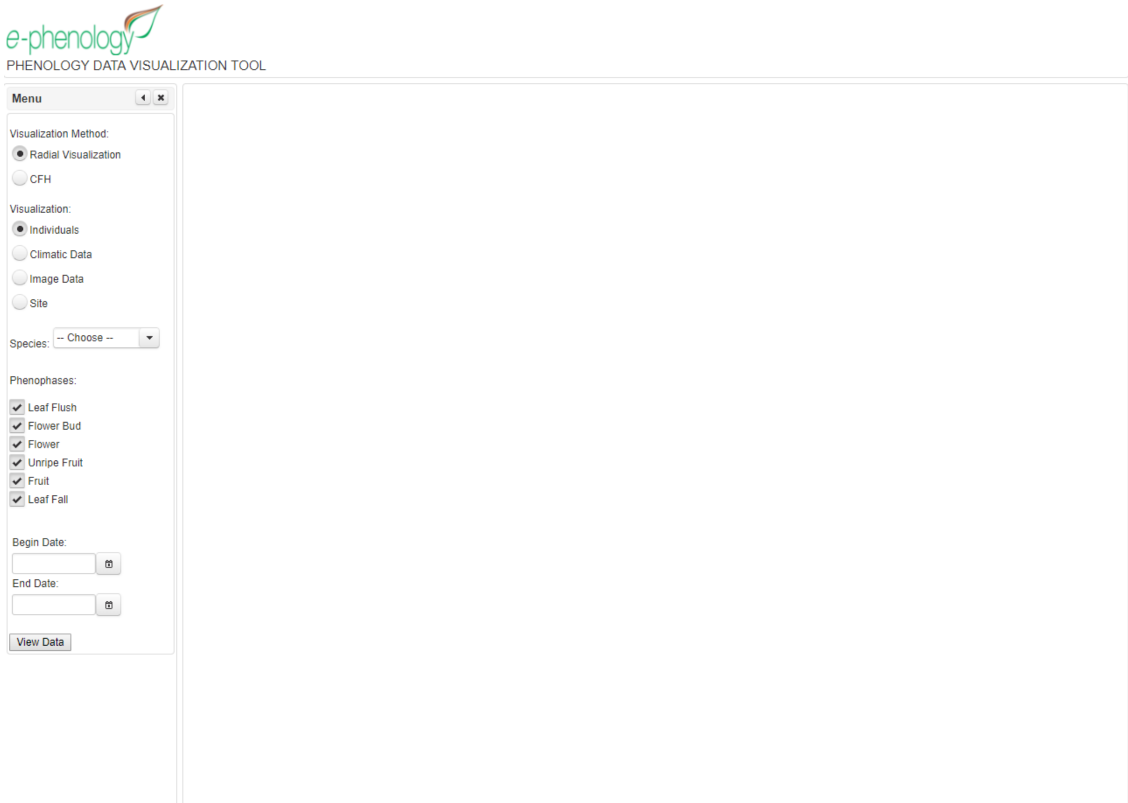


Figure 5.3: Initial interface of phenology data visualization system.

data. The main filters on this case are: the visualization method, species, vegetation indexes (based on channels), and period. After that, users are provided with the visual structure after clicking on the *view data* button.

Suppose, now, that a user wants to analyze how the Gcc vegetation index obtained from sequences of images using masks for the species *Aspidosperma tomentosum* evolves over time, in period from 2014 to 2016. Figure 5.5 shows the screen shot of the tool after filtering options are defined. After requesting the visualization, the user is able to see the visual structure showed in Figure 5.6. Once a visualization is presented, some interaction mechanisms are available to the user, such as, changes of colors, time period, and even on the selected masks. Figure 5.7 shows an example after a user chooses *all ROIs* in the menu *masks* located at right side of the interface.

Going back to the initial interface, the user may select all RGBcc indices to be visualized for one year. Figure 5.8 shows one example of selection of filters, while Figure 5.9 shows the results based on this selection. Other possible interactions, for example, related to the visualization of the Gcc index data for different species at the same time or related to the use of different colors for the RGBcc indices, are shown in the Figures 5.10 and 5.11, respectively. Resulting visual structures, in this case, may be useful in tasks related to the analysis of inter-species differences with concern with different phenophases (e.g., leafing through the Gcc index variation over time and senescence trough the Rcc variation).

Another usage scenario refers to the use of the CFH method to explore how individuals' patterns change over time and how they start to present the phenophases. In this context, returning to the initial interface, when a user selects the CFH method, the main filters

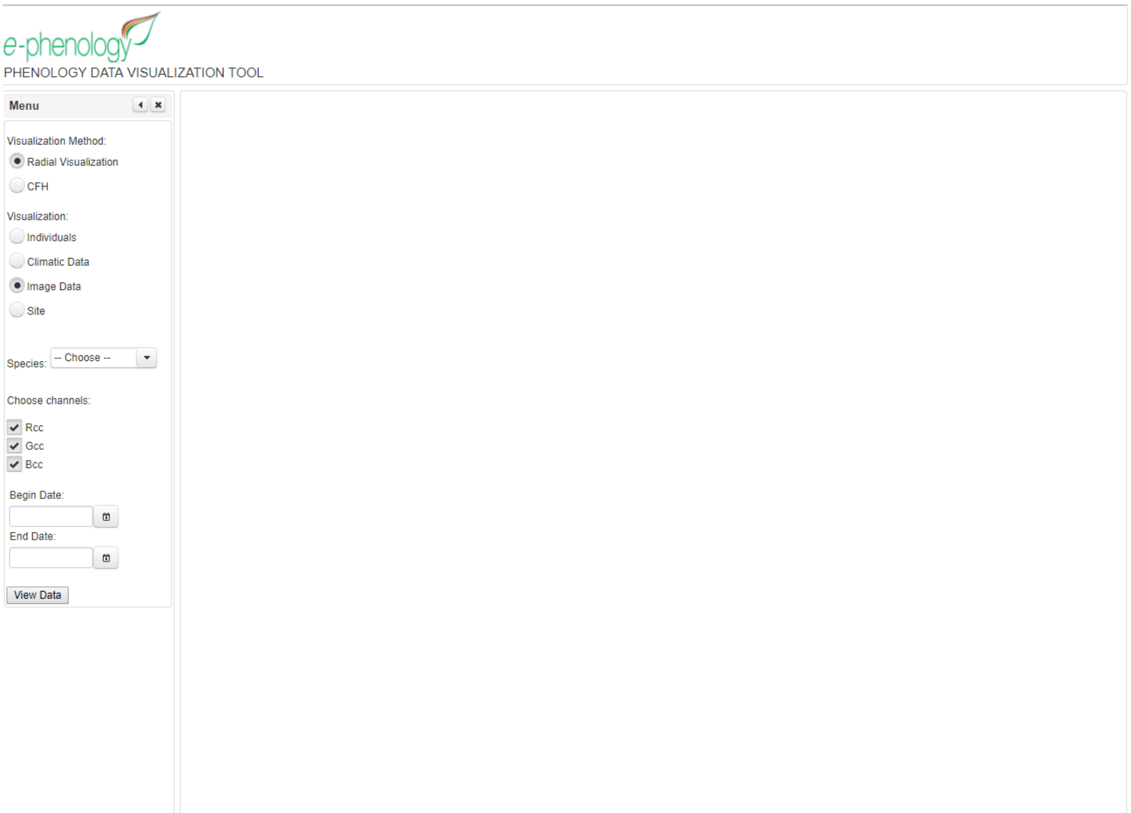


Figure 5.4: Interface with filters related to image data.

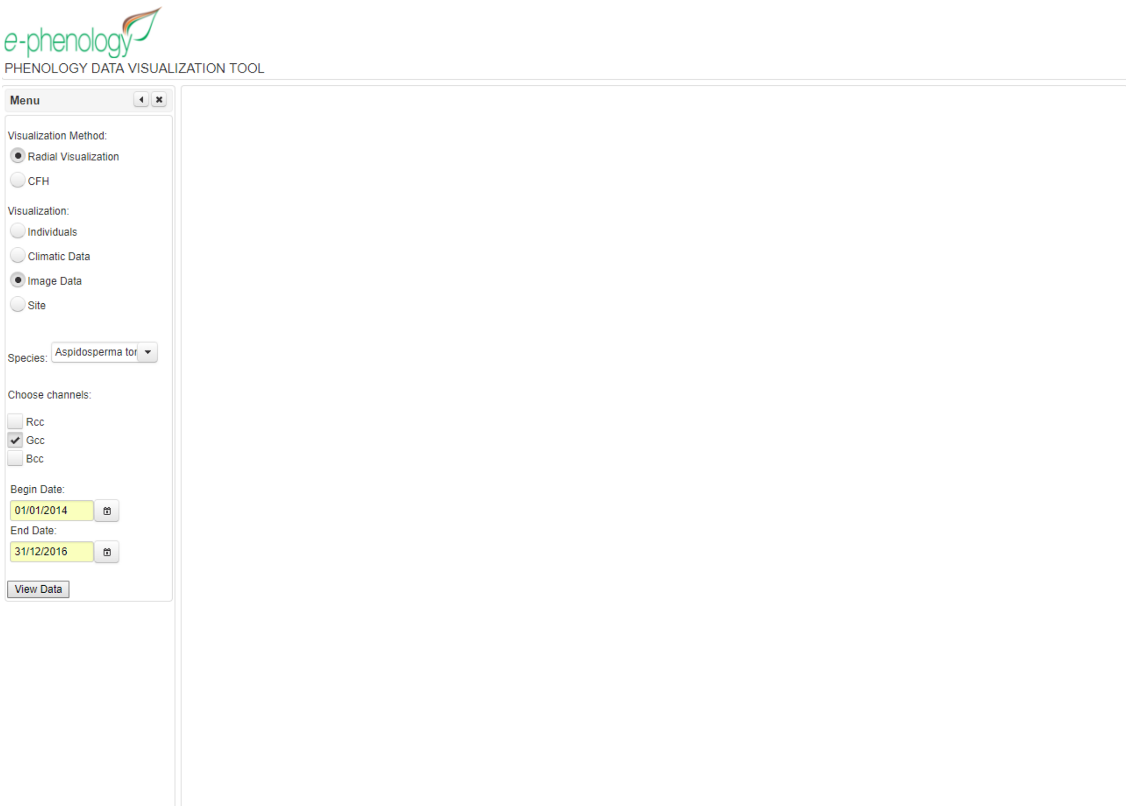


Figure 5.5: Interface with selected filters related with image data

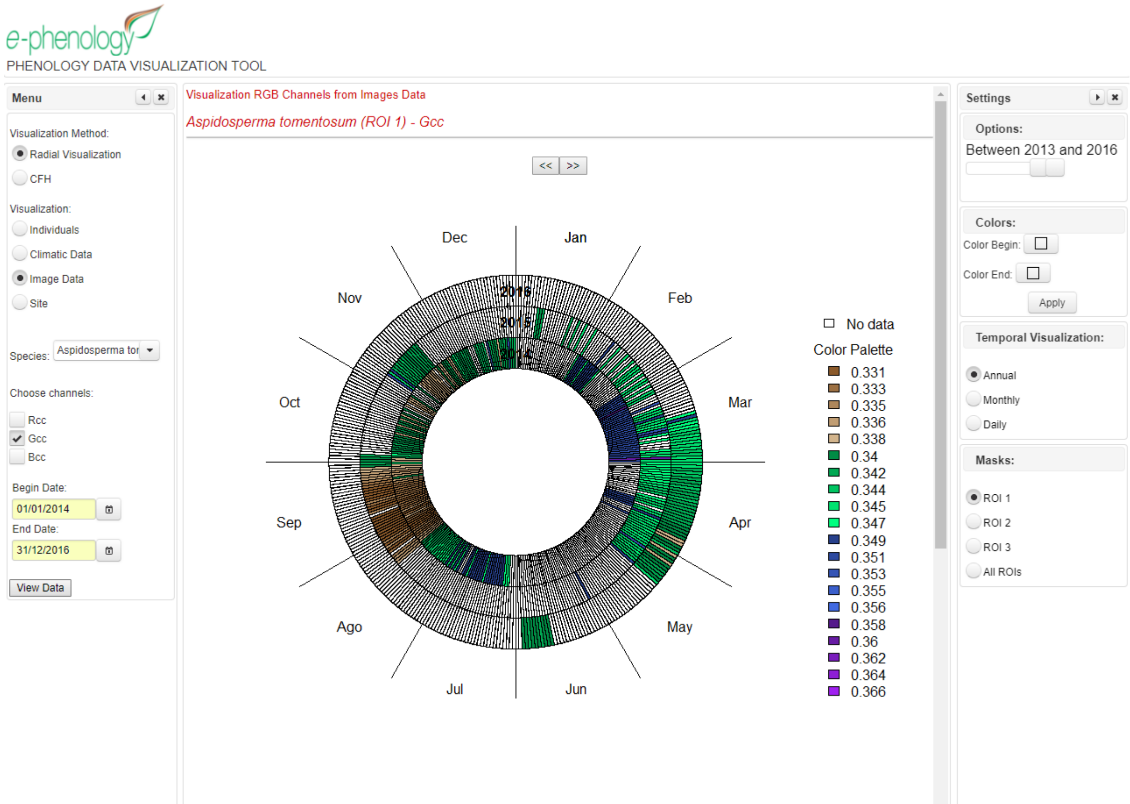


Figure 5.6: Interface with image-related data for the *Aspidosperma tomentosum* species.

are: which species, binary pattern, phenophases, and period are of interest. Figure 5.12 presents these filters. Suppose, for example, that a user selected: the *Myrcia guianensis* species, the *change* pattern, the *flower bud, flower, unripe fruit and fruit* phenophases for the period between 2005 and 2011 (these filters are the same used in the study case of Chapter 4). Figure 5.13 presents the resulting display in the interface. Another possibility is to explore the onset date for the phenophases. In this case a user can, instead of selecting the change pattern, choose the *greening* pattern in the initial interface. In this case, the resulting display is shown in Figure 5.14.

5.3 Conclusions

This chapter introduced the implementation and use of a visualization tool combining the main techniques presented and explored on this work. We designed and implemented a web application using the R language and shiny framework. We also discussed different usage scenarios of the proposed tool in the context of the e-phenology project, demonstrating its potential in the target application.

We also illustrated the proposed visualization tool using image-related data obtained in the context of the e-phenology project. We demonstrated its use in the visualization of leaf color variations using vegetation indices. The investigation of its use to support the analysis of correlations among image-related and in-the-field data is left for future work.

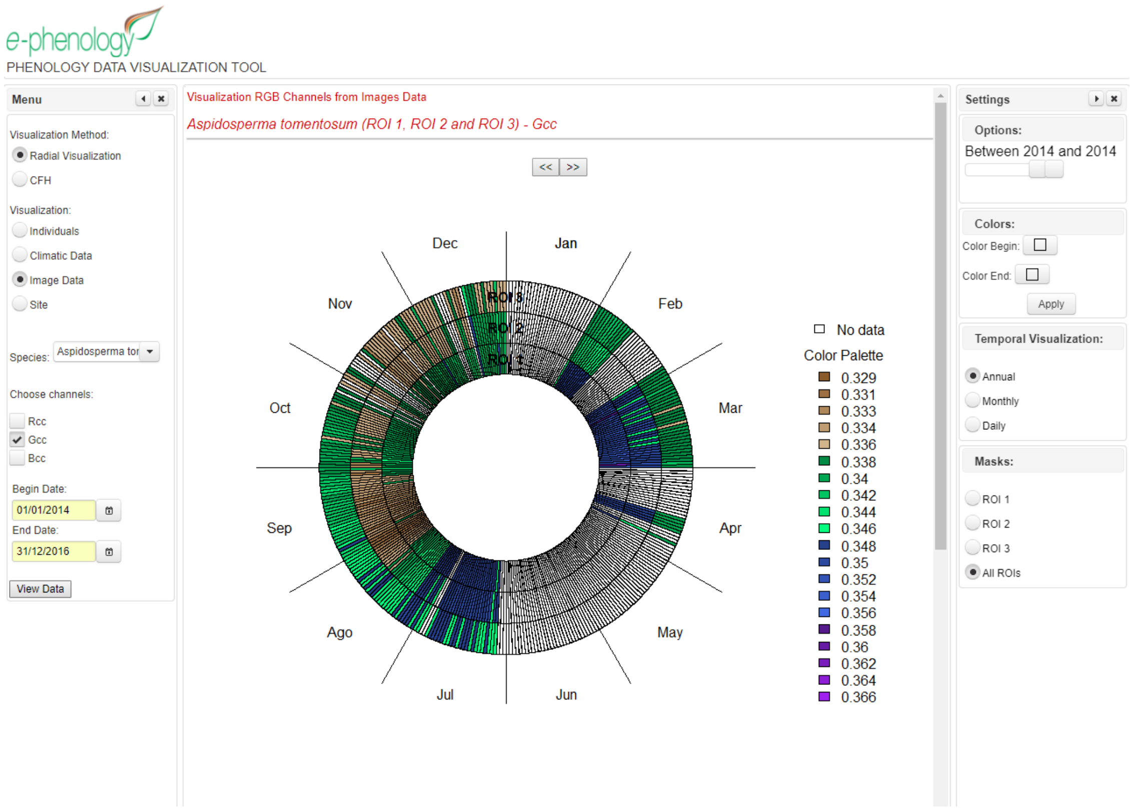


Figure 5.7: Interface with image-related data for the *Aspidosperma tomentosum* species - All ROIs.

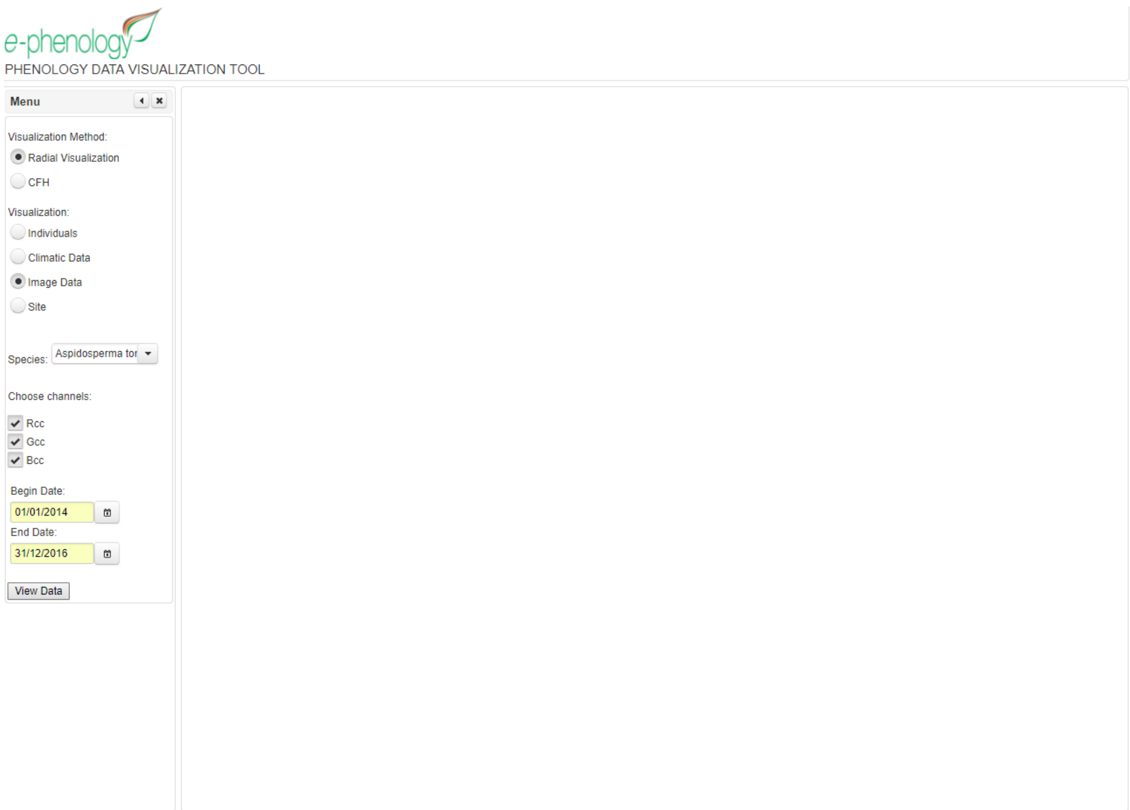


Figure 5.8: Interface with image-related data for the *Aspidosperma tomentosum* species – Rcc, Gcc, and Bcc filters.

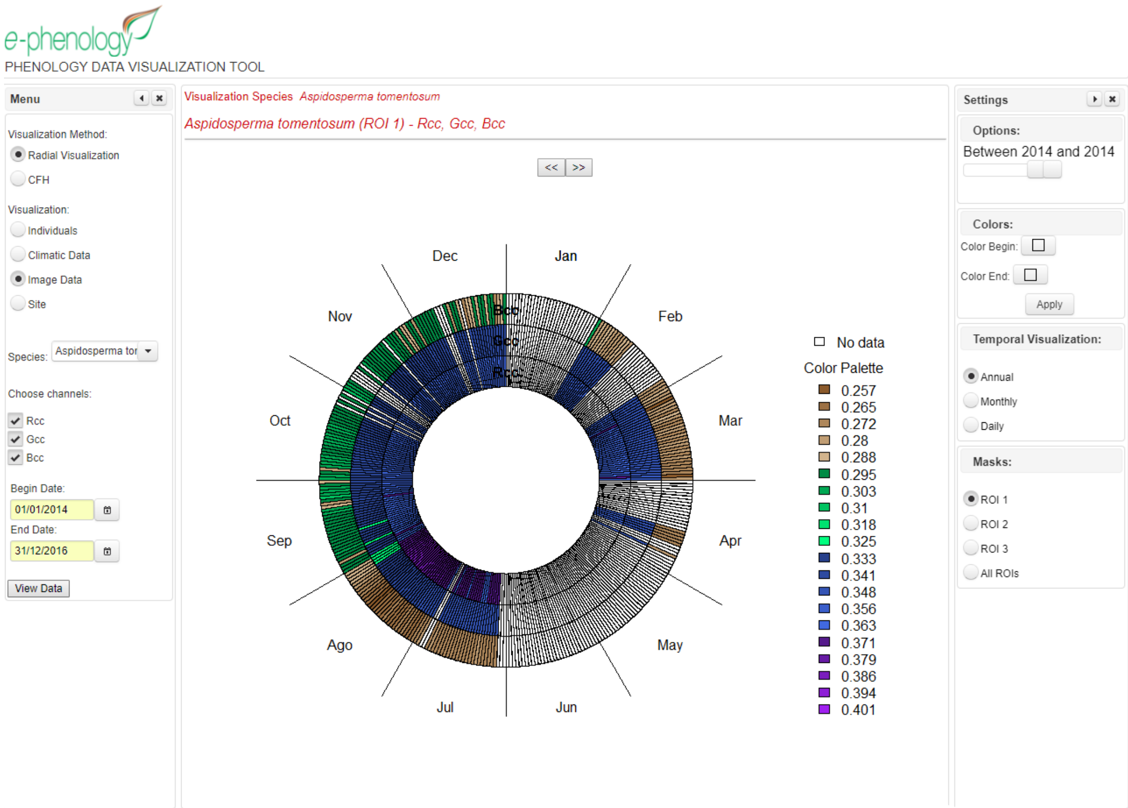


Figure 5.9: Interface with image-related data for the *Aspidosperma tomentosum* species – Rcc, Gcc, and Bcc filters applied.

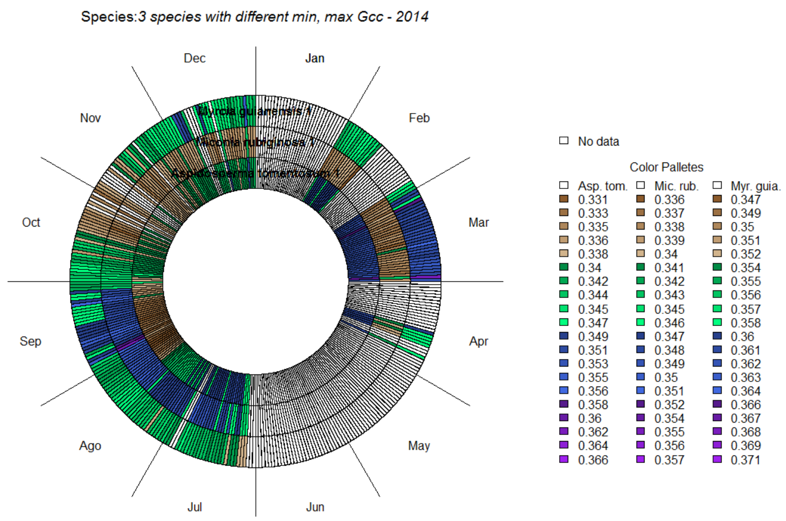


Figure 5.10: Interface with data related to three different species.

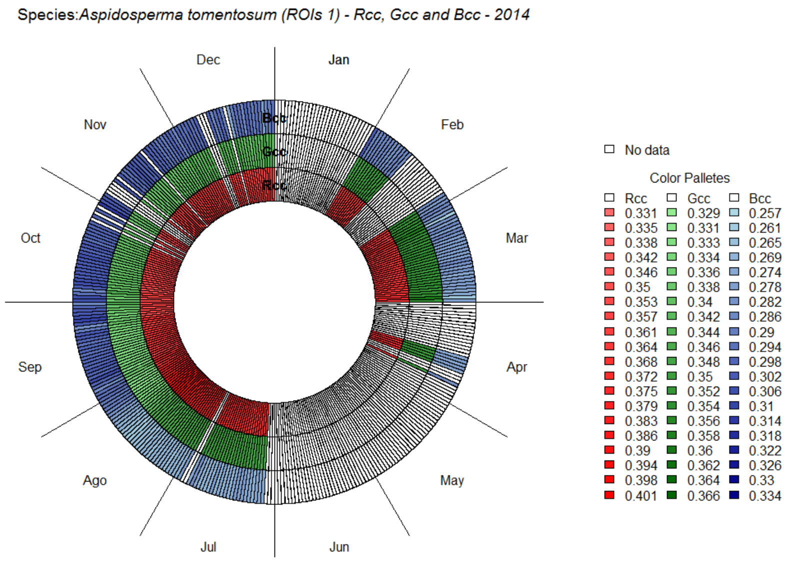


Figure 5.11: Interface with image-related data for the *Aspidosperma tomentosum* species for indices Rcc, Gcc, and Bcc, using different color palletes.

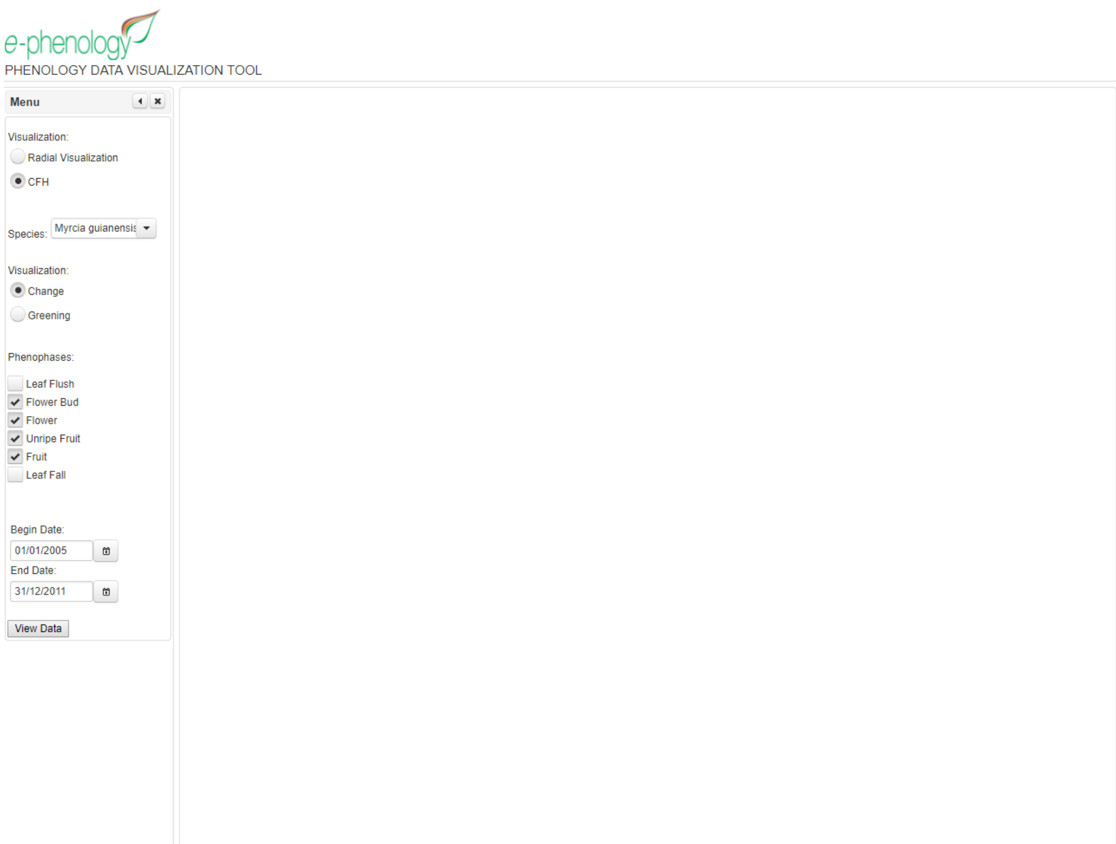


Figure 5.12: Interface with species-related data for the *Myrcia guianensis* using CFH – filtering selection.

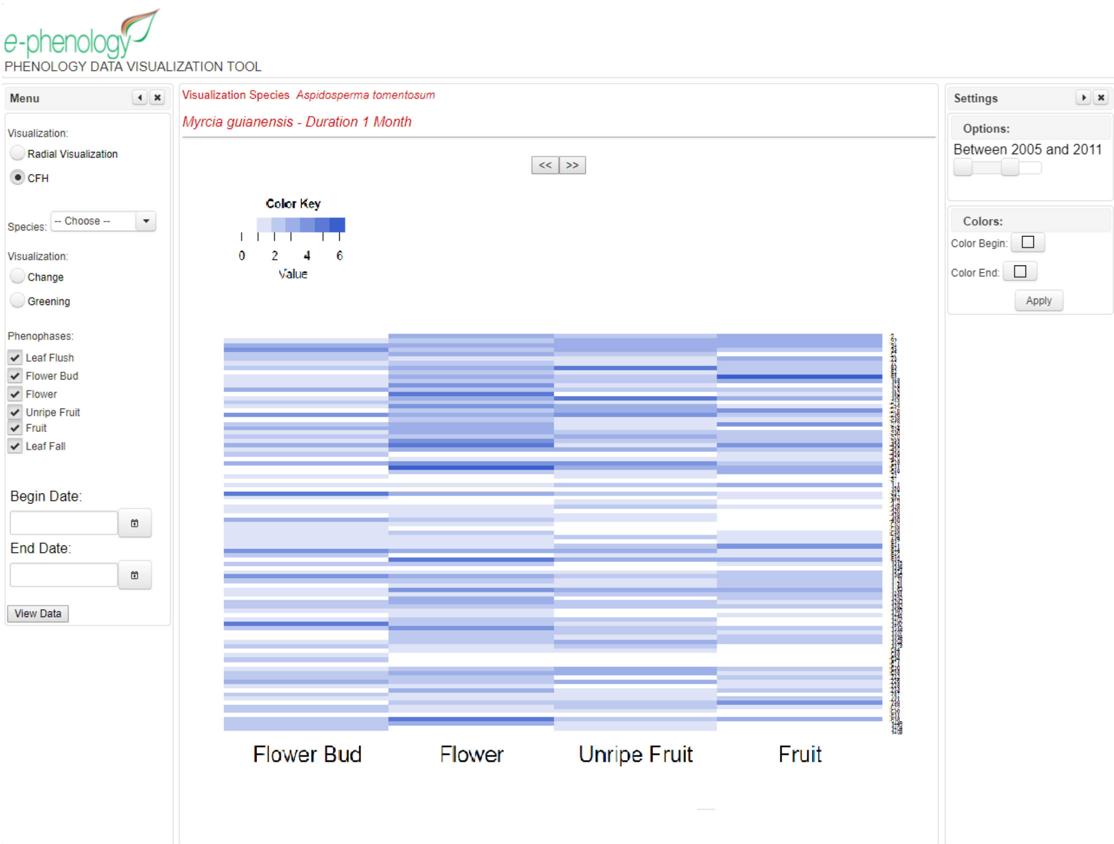


Figure 5.13: Interface with species-related data for the *Myrcia guianensis* using CFH – change pattern.

Figure 5.14: Interface with species-related data for the *Myrcia guianensis* using CFH – greening pattern.

Chapter 6

Conclusions

In several domains, the development of appropriate data visualization tools for supporting complex analysis based, for example, on the identification of change patterns in temporal data or existing correlations, over time, among multiple variables, is of paramount importance. Thus, several approaches to handle and visualize time series have been proposed, but most of them are not robust enough to deal with long time series and multidimensional and multivariate temporal data, as in the case of phenology studies.

Phenology is a discipline concerned with the understanding of life cycle events of living beings and their relationship with climate drives. Usually, phenology experts need to deal with huge volumes of heterogeneous data, in terms of formats, kinds, and scale. In traditional plant phenology studies, for example, collected data refer to phenophase observations made by experts in the field and usually the cyclical changes are correlated with climate data obtained by weather sensors [71, 82]. More recently, plant phenology studies also consider near-remote data derived from sequential vegetation images taken by digital cameras to remotely monitor phenology [5, 103, 104]. Since the phenology studies are based on short to long-term time series and different types of data, the experts often face the challenge of understanding relations among different data types over time.

One suitable approach to support phenology experts in their daily data analysis activities relies on development of appropriate visualization approaches, which may help them to understand evolution of cyclical phenomena over time, complex temporal change pattern changes, or even relations and correlations among multidimensional variables. In this context, this work has introduced novel two image-based visualizations approaches for presenting cyclical multidimensional temporal data associated with phenology studies, based on the typical issues involved in the phenology analyzes (see Section 1.1.1). The study was developed according to the research questions proposed in Section 1.2. In the following, we summarize how each raised research question was addressed:

- The first question was concerned with the investigation of **which visual structures can display time series satisfactorily covering the different types of data and multiple variables involved in phenology analysis**. We addressed this question by investigating visualization methods tipically used to accomplish phenology analysis. Since considered data are cyclical and the phenologists use circular statiscal analysis in their studies, we choose the radial structures as main visual structure to visualize plant phenology time series (see Chapter 3). The use of radial layouts aims at providing contextual information about multiple variables varying over time. We developed some prototypes and carried out evaluations with the phenology experts to validate our proposal, which enabled us to identify **which interaction mechanisms should be supported** by a visualization tool for phenology. Our first evaluations were concerned with the representation of traditional on-the-ground direct observation data in a detailed level, i.e., using data at plant individual level.

- Another research question addressed by the study presented in Chapter 3 was concerned with the representation of phenological data in a summarization level, which lead us to investigate **which data summarization approach could be used to visualize long-term times series in a compact way**. To address this question, we introduced the use of the visual rhythm approach. Usually, visual rhythms are used to encode video temporal changes and have been successfully used to visualize sequences of vegetation images in the e-phenology Project context [10, 11], but in our study we used it to encode tabular temporal data. We combined the radial structures with visual rhythm approach into prototypes and we carried out evaluations with phenology experts using data at species level, i.e., we summarized plant phenology data of individuals of the same species and represented these data in radial structures to be evaluated by experts. We also validated our proposal using image-related data to investigate if is possible to compare on-the-ground phenology observations with near-surface remote data.

- The third and fourth questions were concerned with the investigation of **which visualization technique can be used to encode cyclical temporal changes** and **which data summarization approach could be used to encode those changes**, respectively. We addressed these questions by proposing a new method named as Change Frequency Heatmap (CFH). This method combines heatmaps with motion recognition to summarize the temporal data as frequencies. The main advantages of the proposed temporal change representation relies on its flexibility and effectiveness in representing the most relevant features, in terms of the frequency of occurrences, of the patterns observed in large volumes of data, considering the number of objects handled in the study, the number of variables, and the length of time series.

- The last question was concerned with **which phenology research activities can benefit from the developed visual structures** and **how to validate the developed data structures with phenology experts**. We addressed these questions according to the scenarios presented in Section 1.1.1. Thus, in Chapter 3, we used radial visualization combined with visual rhythms to visualize data from individuals and species. We divided it in two levels of visualization: (i) detailed level, addressing the questions related to the visualization of one or more phenophases for one or more individuals over time, and (ii) summarization level, addressing the questions related to the visualization of one or more phenophases for one or more species over time. Both kinds of visualization were correlated with climate data. In Chapter 4, we used heatmaps combined with motion recognition approach to the visualization of temporal changes at individual levels. In this case, the phenology questions are those related to visualize and compare phenological temporal patterns between different individuals from the same species, such as, when a phenomenon starts and for how long a phenomenon lasts. The question about **how to validate developed data structures with phenology experts** was addressed in the studies, by interactions and formal evaluations with phenology experts using prototypes and forms (in the case of Chapter 3) and comparing our results with published data (in the case of Chapter 4).

The main hypothesis that guided this study was: *Visual structures combined with data summarization approaches are effective to create a scientific visualization system to identify cyclical patterns in multivariate multidimensional temporal data and potentially improve phenology analyzes.* We explored two different methods to encode and represent multivariate multidimensional temporal data. In the first one, we combined the radial structure with visual rhythm to represent phenological data both at individuals and species levels. We carried out evaluations with users, which demonstrated that the proposed prototypes were promising. Besides that, the method showed that is possible to summarize data for individuals and species over several years helping the phenologists to quickly identify phenological patterns. The second method, named as Change Frequency Heatmap was inspired by a computer vision method and was developed to encode frequencies of temporal changes. The objective was to support the visualization and identification of complex temporal phenomena investigated in the context of phenology studies, such as the greening of plants and the duration of the recurrent events. The combination of the proposed methods gave us the possibility to develop a powerful tool to support phenology analyzes. We believe, therefore, that the raised hypothesis was confirmed.

6.1 Summary of Main Contributions

This work provides contributions for two different areas: (i) Computer Science (CS) and (ii) Phenology. In the CS area, we advance the state of the art with regard to multidimensional and multivariate time series visualization. In this context, this work focuses on proposing new scientific visualization using novel visual structures to visualize temporal data, as well as to combine these techniques with summarizing methods to manage multidimensional data. We first propose a new method using radial layouts combined with visual rhythm approach to visualize numerical data and validated this technique with phenology experts. Furthermore, we propose the use of motion history histogram combined with heatmaps to visualize the phenophases onset dates and changes over time.

All proposed visual structures have been validated in the context of phenology studies. Therefore, our work contribute to the Phenology area by introducing novel approaches, algorithm, and tools to support phenology experts in their daily tasks aiming to address complex research questions related to the analysis of multidimensional variables over time. Although we developed and validated the proposed techniques using on-the-ground observations (phenological data) and precipitation (climate data), our first results indicate that the proposed visual structures are adequate and promising to support experts to detect recurrent temporal changes. Besides that, the technique was designed to be generic and flexible enough to be used, for example, to explore new drivers and detect climate changes. Also, the technique can be extended for other temporal studies with similar features, as in the case of football matches analysis and urban informatics. In summary, the main contributions of this work are:

- a proposal of an information visualization technique using radial layouts combined with visual rhythm to visualize on-the-ground phenology observations and image temporal properties in both summarization and detailed levels (Chapter 3);
- a proposal of an information visualization techniques using motion history histogram combined with heatmaps to identify *when* a phenophase starts and *how long* a phenophase lasts for short and long time series (Chapter 4);
- implementation of proposed approaches using R and validation of the proposed approaches with phenology experts (Chapters 3 and 4);
- initial design for a future implementation of a new information visualization system considering design and human-interface aspects (Chapter 5).

6.2 Future Work and Possible Extensions

The contributions presented in this work focus mainly on solving problems related to data visualization in the context of plant phenology studies. We investigated visualization methods and used some specific phenological questions to starting the development of a scientific tool to support phenology analyzes. However, there are many more scenarios in phenology to be explored aiming to better support phenology experts in their daily research tasks. Besides that, the proposed approaches may be validated or adapted to handle temporal data from other contexts and domains. In this way, this research opens novel opportunities for investigations. Some of them are listed in the following:

1. *New variables arrangement to correlate more time series.* In this work we validated our radial structure visualization proposal using on-the-ground phenology observations combined with rainfall measurement obtained by precipitation sensors. However, in phenology studies, phenologists can correlate data on individuals or species with other climatic measures, such as, soil moisture, wind speed and solar radiation. In this way, new variable arrangements can be tested to correlate on-the-ground phenology with these weather data and novel visualization techniques with this goal can be investigated in the future.

2. *Additional formal validations with experts.* For the radial-layout-based proposal, we performed validations with phenology experts using on-the-ground and climatic data. Formal validations also can be conducted with experts to validate the proposed visual structures in tasks associated with the investigation of the correlation between

data obtained from images with on-the-ground observations. In the same venue, formal evaluation protocols can be developed to assess the CFH method in daily image-informed research tasks.

3. *Validation of CFH in other domains.* We identified that the CFH is a potential method to discover temporal patterns based on functions of interest. Therefore, an important research venue consists in investigating the use of this representation in other applications, such as sport analysis. For example, soccer match analyzes based on the temporal networks [105] can benefit from the CFH encoding approach and can be considered another important application for validation. Besides that, novel approaches to encode visually binary patterns over time may be investigated in future work.

4. *Development of novel visual structures based on the combination of proposed methods.* The methods presented in this thesis were explored using the same case studies and databases, and they were included in a scientific tool for supporting phenology studies. One possible extension of this work, however, is to encode the data generated by the CFH method into radial structures. The resulting visual structures, defined in terms of the combination of both strategies, would allow, for example, to compare data for each individual with their on-the-ground phenology observations as it was explored in Chapter 3.

5. *Use of machine learning to detect individuals with similar patterns.* One important question in plant phenology is to know how the individual of the same species from different sites behave. For example, possible research questions to be addressed could be: Do individuals present the same period of flowering? Do they have peak at the same time for the flower phenophases? Do they present the same onset date for the phenophases? Do individuals from different species, but for the same family, have similar behavior? In this context, one possibility is to define functions of interest and compute the binary time series using CFH and use this as input data of machine learning methods, such as metric learning approaches [6] to detect individuals or species with similar patterns.

6. *Implementation and validation of the developed system.* Further development activities need to be performed in order to make the phenology visualization data tool completely operational. Also, its complete validation with phenology experts, based on well-defined tasks is left for future work. Another possible extension is to generalize the system to be used with data from different phenology-related projects.

7. *Integration with other phenology tools and statistical approaches.* Many approaches have been developed in the context of plant phenology analyzes. We plan to integrate the developed system with some of these approach, as is the case of the PHENOR model framework developed by Hufkens et al. [57]. The PHENOR is an R package that combines data from near-surface remote sensing through the Phenocam network using another packages to compose a phenology modelling framework, which covers

data preparation, model optimization and model visualization. Also, we plan to integrate the radial structures with circular statistical analyses.

8. *Development of interaction mechanisms.* Another important research venue relies on further developing the visualization tool so that it will be rich in interaction mechanisms, based on which users can manipulate data to have insights and to identify cyclical temporal patterns in a more customized and possibly easy way. The existence of simple and complex interaction mechanisms may suit the tool to existing needs of potential users with different profiles, ranging from naive to more experienced users.

9. *Investigation of novel designs based on phenological data.* Finally, new ways to visualize phenological data using similar approaches to the ones presented in this work can be explored. For example, Figure 6.1 shows four mockups (A, B, C, and D) developed based on design and human-interface principles [75, 88]. These examples illustrate new ways of integrating the visualization of on-the-ground observations with layouts produced using the CFH method. These interfaces provide interaction mechanisms, based on which users can explore and navigate through the data and even use the depicted visual layout as a report chart for scientific communication.

6.3 Published Contributions

This research is associated with the following publications:

- G. C. Mariano, V G. Staggemeier, L. P. C. Morellato and R. da S. Torres. Multivariate cyclical data visualization using radial visual rhythms: A case study in phenology analysis. *Ecological Informatics*, Volume 46, 2018, pages 19–35, ISSN 1574-9541, https://doi.org/10.1016/j.ecoinf.2018.05.003.

- G. C. Mariano, N. C. Soares, L. P. C. Morellato and R. da S. Torres. Change frequency heatmaps for temporal multivariate phenological data analysis. In *2017 IEEE 13th International Conference on e-Science (e-Science)*, pages 305–314, Oct 2017.

- G. C. Mariano, V G. Staggemeier, L. P. C. Morellato and R. da S. Torres. A visual rhythm approach to visualize multidimensional cyclical data: A case study on phenological data analysis. In *International Conference on Digital Libraries*, 2016.

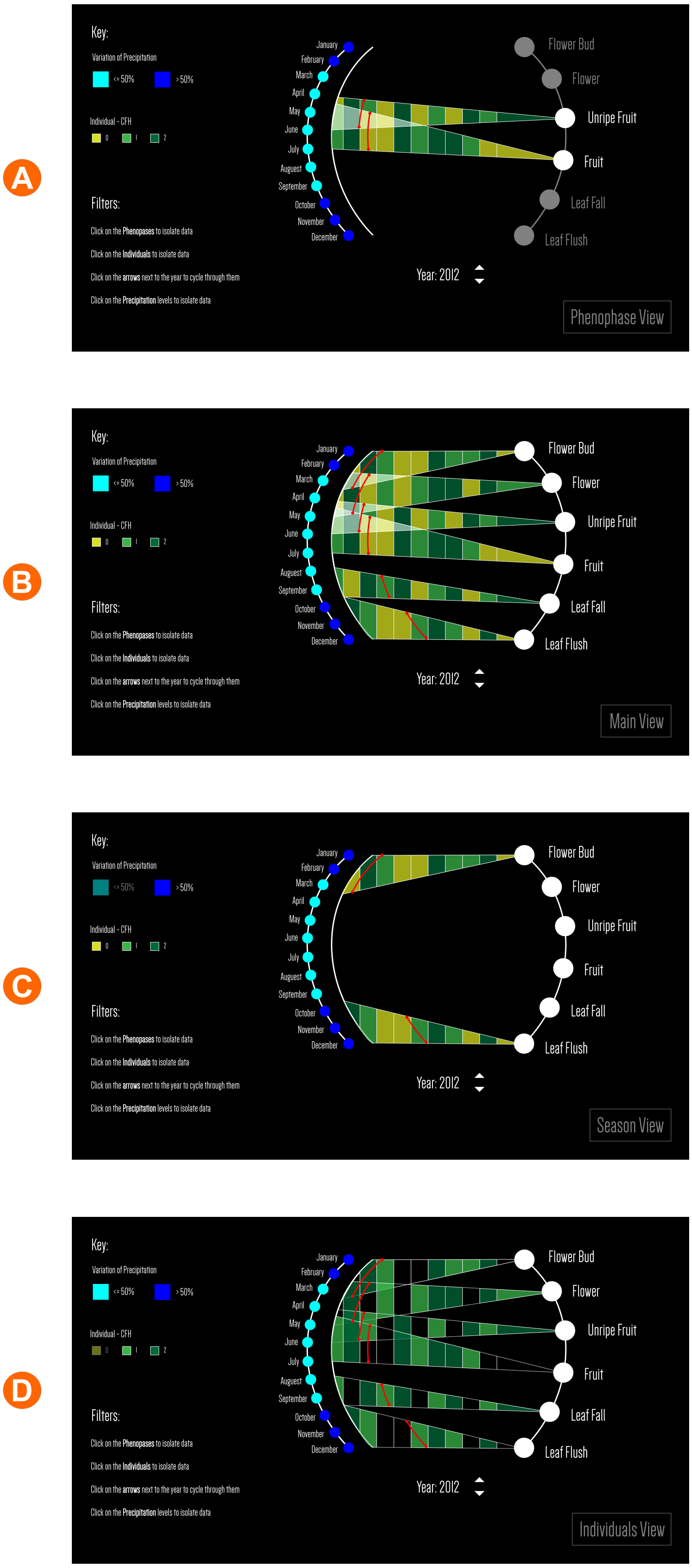


Figure 6.1: Mockups (A, B, C, and D) developed based on design and human-interface principles [75, 88]. These examples illustrate new ways of integrating the visualization of on-the-ground observations with layouts produced using the CFH method.

Bibliography

[1] White Michael A., Thornton Peter E., and Running Steven W. A continental phenology model for monitoring vegetation responses to interannual climatic variability. *Global Biogeochemical Cycles*, 11(2):217–234, 1997.

[2] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski. Visualizing time-oriented data—a systematic view. *Computers & Graphics*, 31(3):401 – 409, 2007.

[3] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of Time-Oriented Data*. Human-Computer Interaction Series. Springer, 2011.

[4] B. Alberton, J. Almeida, R. Helm, R. da S. Torres, A. Menzel, and L. P. C. Morellato. Using phenological cameras to track the green up in a cerrado savanna and its on-the-ground validation. *Ecological Informatics*, 19(0):62 – 70, January 2014.

[5] B. Alberton, R. da S. Torres, L. F. Cancian, B. D. Borges, J. Almeida, G. C. Mariano, J. dos Santos, and L. P. C. Morellato. Introducing digital cameras to monitor plant phenology in the tropics: applications for conservation. *Perspectives in Ecology and Conservation*, 15(2):82 – 90, 2017.

[6] A. E. Almeida and R. da S. Torres. Remote sensing image classification using genetic-programming-based time series similarity functions. *IEEE Geoscience and Remote Sensing Letters*, 14(9):1499–1503, September 2017.

[7] J. Almeida, J. A. dos Santos, B. Alberton, R. da S. Torres, and L. P. C. Morellato. Remote phenology: Applying machine learning to detect phenological patterns in a cerrado savanna. In *eScience*, pages 1–8. IEEE Computer Society, 2012.

[8] J. Almeida, J. A. dos Santos, B. Alberton, R. da S. Torres, and L. P. C. Morellato. Applying machine learning based on multiscale classifiers to detect remote phenology patterns in cerrado savanna trees. *Ecological Informatics*, pages –, 2013.

[9] J. Almeida, J. A. dos Santos, B. Alberton, L. P. C. Morellato, and R. da S. Torres. Phenological visual rhythms: Compact representations for fine-grained plant species identification. *Pattern Recognition Letters*, 81:90–100, 2016.

[10] J. Almeida, J. A. Santos, B. Alberton, L. P. C. Morellato, and R. S. Torres. Plant species identification with phenological visual rhythms. In *Proceedings of the 9th IEEE International Conference on eScience (eScience 2013)*. eScience, 2013.

[11] J. Almeida, J. A. Santos, B. Alberton, L. P. C. Morellato, and R. S. Torres. Visual rhythm-based time series analysis for phenology studies. In *Proceedings of the 20th IEEE International Conference on Image Processing*, pages 1–5. ICIP, 2013.

[12] M. Ankerst, D. A. Keim, and H p Kriegel. Circle segments : A technique for visually exploring large multidimensional data sets. In *Visualization '96, Hot Topic Session, San Francisco, CA, November, 1996*, 1996.

[13] M. S. Araújo, D. I. Bolnick, and C. A. Layman. The ecological causes of individual specialisation. *Ecology Letters*, 14(9):948–958, 2011.

[14] R. Arsenault, C. Ware, M. Plumlee, S. Martin, L.L. Whitcomb, D. Wiley, T. Gross, and A. Bilgili. A system for visualizing time varying oceanographic 3d data. In *OCEANS '04. MTTS/IEEE TECHNO-OCEAN '04*, volume 2, pages 743 – 747 Vol.2, nov. 2004.

[15] G. Asrar, E.T. Kanemasu, and M. Yoshida. Estimates of leaf area index from spectral reflectance of wheat under different cultural practices and solar angle. *Remote Sensing of Environment*, 17(1):1 – 11, 1985.

[16] C. K. Augspurger. Phenology, flowering synchrony, and fruit set of six neotropical shrubs. *Biotropica*, 15(4):257–267, 1983.

[17] B. Bach, C. Shi, N. Heulot, T. Madhyastha, T. Grabowski, and P. Dragicevic. Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):559–568, Jan 2016.

[18] C. S.C. Bencke and L. P. C. Morellato. Comparação de dois métodos de avaliação da fenologia de plantas, sua interpretação e representação. *Brazilian Journal of Botany*, 25:269 – 275, 09 2002.

[19] A. Bobick and J. Davis. An appearance-based representation of action. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 1, pages 307–312. IEEE, 1996.

[20] A. F. Bobick. Movement, activity and action: the role of knowledge in the perception of motion. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 352(1358):1257–1265, 1997.

[21] A. F. Bobick and J. W. Davis. The recognition of human movement using temporal templates. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(3):257–267, March 2001.

[22] D. I Bolnick, P. Amarasekare, M. S. Araújo, R. Bürger, J. M. Levine, M. Novak, V. H. W. Rudolf, S. J. Schreiber, M. C. Urban, and D. a Vasseur. Why intraspecific trait variation matters in community ecology. *Trends in ecology & evolution*, 26(4), February 2011.

[23] G. R. Bradski and J. Davis. Motion segmentation and pose recognition with motion history gradients. In *Proceedings Fifth IEEE Workshop on Applications of Computer Vision*, pages 238–244, 2000.

[24] M. Burch and S. Diehl. Timeradartrees: Visualizing dynamic compound digraphs. *Comput. Graph. Forum*, 27(3):823–830, 2008.

[25] M. G. G. Camargo, R. M. Souza, P. Reys, and L. P. C. Morellato. Effects of environmental condi- tions associated to the cardinal orientation on the reproductive phenology of the cerrado savanna tree *Xylopia aromatica* (annonaceae). *Annals of the Brazilian Academy of Sciences)*, 83:1007–1020, 2011.

[26] M. G. G. Camargo, R. M. Souza, P. Reys, and L. P.C. Morellato. Effects of environmental conditions associated to the cardinal orientation on the reproductive phenology of the cerrado savanna tree xylopia aromatica (annonaceae). *Anais da Academia Brasileira de Ciências*, 83:1007 – 1020, 09 2011.

[27] S. K. Card, J. D. Mackinlay, and B. Shneiderman, editors. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.

[28] J. V. Carlis and J. A. Konstan. Interactive visualization of serial periodic data. In *Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology*, UIST '98, pages 29–38, New York, NY, USA, 1998. ACM.

[29] Sheelagh Carpendale. *Evaluating Information Visualizations*, pages 19–45. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[30] C. Chaomei. Information visualization. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):387–403, 2010.

[31] CHAOMEI CHEN and MARY P. CZERWINSKI. Empirical evaluation of information visualizations: an introduction. *International Journal of Human-Computer Studies*, 53(5):631 – 635, 2000.

[32] M. G. Chung, H. Kim, and S. M-H Song. A scene boundary detection method. In *ICIP*, pages 933–936, 2000.

[33] C. Daassi, M. Dumas, M-C Fauvet, L. Nigay, and P-C Scholl. Visual exploration of temporal object databases, 2000.

[34] J. W. Davis and A. F. Bobick. The representation and recognition of human movement using temporal templates. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 928–934, Jun 1997.

[35] J. Dierenbach, F-W. Badeck, and J. Schaber. The plant phenological online database (ppodb): an online database for long-term phenological data. *Int J Biometeorol*, 2013.

[36] L. C. B. dos Santos, S. J. F. Guimarães, A. de A. Araújo, and J. A. dos Santos. Unsupervised hyperspectral band selection based on spectral rhythm analysis. In *27th SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI 2014, Rio de Janeiro, Brazil, August 27-30, 2014*, pages 157–164, 2014.

[37] G. M. Draper, Y. Livnat, and R. F. Riesenfeld. A survey of radial methods for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 15(5):759–776, 2009.

[38] Y. Drocourt, Rita Borgo, K. Scharrer, T. Murray, S. I. Bevan, and M. Chen. Temporal Visualization of Boundary-based Geo-information Using Radial Projection. *Computer Graphics Forum*, 2011.

[39] R. Elmasri and S. B. Navathe. *Fundamentals of Database Systems*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 3rd edition, 1999.

[40] J. A. Elzinga, A. Atlan, A. Biere, L. Gigord, A. E. Weis, and G. Bernasconi. Time after time: flowering phenology and biotic interactions. *Trends in Ecology & Evolution*, 22:432–439, 2007.

[41] N. Ferreira, L. D. Lins, D. Fink, S. Kelling, C. Wood, J. Freire, and C. T. Silva. Birdvis: Visualizing and understanding bird populations. *IEEE Trans. Vis. Comput. Graph.*, 17(12):2374–2383, 2011.

[42] N. I. Fisher, T. Lewis, and B. J. J. Embleton. *Statistical analysis of spherical data*. Cambridge University Press, 1993.

[43] C. Forsell. A guide to scientific evaluation in information visualization. In *2010 14th International Conference Information Visualisation*, pages 162–169, July 2010.

[44] L. A. Fournier. Un método cuantitativo para la medición de características fenológicas en árboles. *Turrialba*, 24(4):422–423, 1974.

[45] A. U. Frank. Different types of "times" in gis. *Spatial and temporal reasoning in geographic information systems*, pages 40–62, 1998.

[46] C. M. D. S. Freitas, O. M. Chubachi, P. R. G. Luzzardi, and R. A. Cava. Visualização de informações. *Revista de Informática Teórica e Aplicada*, 8(2):143–158, 2001.

[47] N. D. Gershon, S. G. Eick, and S. K. Card. Information visualization. *Interactions*, 5(2):9–15, 1998.

[48] A. R. Gillespie, A. B Kahle, and R. E. Walker. Color enhancement of highly correlated images. ii. channel ratio and "chromaticity" transformation techniques. *Remote Sensing of Environment*, 22(3):343–365, 1987.

[49] A. M. Giulietti, N. L. Menezes, J. R. Pirani, M. Meguro, and M. G. L. Wanderley. Flora da serra do cipó, minas gerais: Caracterização e lista das espécies. *Boletim de Botânica*, 9(0):1–151, 1987.

[50] I. A. Goralwalla, M. T. Özsu, and D. Szafron. *An object-oriented framework for temporal data models*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

[51] Y. Gu, C. Wang, T. Peterka, R. Jacob, and S. H. Kim. Mining graphs for understanding time-varying volumetric data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):965–974, Jan 2016.

[52] S. J. F. Guimarães, M. Couprie, A. de A. Araújo, and N. J. Leite. Video segmentation based on 2d image analysis. *Pattern Recognition Letters*, 24(7):947 – 957, 2003.

[53] M. Hagen, W. D. Kissling, C. Rasmussen, M. A. M. De Aguiar, L. E. Brown, D. W. Carstensen, I. Alves-Dos-Santos, Y. L. Dupont, F. K. Edwards, J. Genini, P. R. Guimarães Jr., G. B. Jenkins, P. Jordano, C. N. Kaiser-Bunbury, M. E. Ledger, K. P. Maia, F. M. D. Marquitti, Ó. Mclaughlin, L. P. C. Morellato, E. J. O'Gorman, K. Trøjelsgaard, J. M. Tylianakis, M. M. Vidal, G. Woodward, and J. M. Olesen. 2 - biodiversity, species interactions and ecological networks in a fragmented world. In Ute Jacob and Guy Woodward, editors, *Global Change in Multispecies Systems Part 1*, volume 46 of *Advances in Ecological Research*, pages 89 – 210. Academic Press, 2012.

[54] C.M. Herrera. *Multiplicity in Unity: Plant Subindividual Variation and Interactions with Animals*. Interspecific Interactions. University of Chicago Press, 2009.

[55] N. Hochman and R. Schwartz. Visualizing instagram: Tracing cultural visual rhythms. 2012.

[56] M. S. Horn, M. Tobiasz, and C. Shen. Visualizing biodiversity with voronoi treemaps. In *Proceedings of the 2009 Sixth International Symposium on Voronoi Diagrams*, ISVD '09, pages 265–270, Washington, DC, USA, 2009. IEEE Computer Society.

[57] K. Hufkens, D. Basler, T. Milliman, E. K. Melaas, and A. D. Richardson. An integrated phenology modelling framework in r. *Methods in Ecology & Evolution*, 01/2018 2018.

[58] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multidimensional geometry. In *Proceedings of the First IEEE Conference on Visualization: Visualization '90*, pages 361–378, Oct 1990.

[59] D. Jäckle, F. Fischer, T. Schreck, and D. A. Keim. Temporal mds plots for analysis of multivariate data. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):141–150, Jan 2016.

[60] E. Kandogan. Star coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions. In *In Proceedings of the IEEE Information Visualization Symposium, Late Breaking Hot Topics*, pages 9–12, 2000.

[61] J. Kehrer and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(3):495–513, March 2013.

[62] D. A. Keim, M. C. Hao, U. D., and M. Hsu. Pixel bar charts: a visualization technique for very large multi-attribute data sets. *Information Visualization*, 1:20–34, 2002.

[63] D. A. Keim, H. P. Kriegel, and M. Ankerst. Recursive pattern: a technique for visualizing very large amounts of data. In *Visualization, 1995. Visualization '95. Proceedings., IEEE Conference on*, pages 279–286, 463, Oct 1995.

[64] D. A. Keim, J. Schneidewind, and M. Sips. Circleview: A new approach for visualizing time-related multidimensional data sets. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '04, pages 179–182, New York, NY, USA, 2004. ACM.

[65] H. Kim, S-J. Park, J. Lee, W. M. Kim, and S. M. Song. Processing of partial video data for detection of wipes. In *Storage and Retrieval for Image and Video Databases VII (1998)*, volume 3656, pages 280–289, 1998.

[66] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, Sept 2012.

[67] R. A. Leite, L. M. Schnorr, J. Almeida, B. Alberton, L. P. C. Morellato, R. da S. Torres, and J. L.D. Comba. Phenovis – a tool for visual phenological analysis of digital camera images using chronological percentage maps. *Information Sciences*, 372:181 – 195, 2016.

[68] J. Li, Z-P. Meng, M-L. Huang, and K. Zhang. An interactive radial visualization of geoscience observation data. In *Proceedings of the 8th International Symposium on Visual Information Communication and Interaction*, VINCI '15, pages 93–102, New York, NY, USA, 2015. ACM.

[69] A. Lu and H-W. Shen. Interactive storyboard for overall time-varying data visualization. In *PacificVis*, pages 143–150. IEEE, 2008.

[70] D. J. M, R. M. Lopes, and R. M. Cesar. Visual rhythm-based method for continuous plankton monitoring. *2013 IEEE 9th International Conference on e-Science*, 0:204–211, 2013.

[71] G. C. Mariano, L. P. C. Morellato, J. Almeida, B. Alberton, M. G. G. de Camargo, and R. da S. Torres. Modeling plant phenology database: Blending near-surface remote phenology with on-the-ground observations. *Ecological Engineering*, 91:396 – 408, June 2016.

[72] G. C. Mariano, N. C. Soares, L. P. C. Morellato, and R. D. S. Torres. Change frequency heatmaps for temporal multivariate phenological data analysis. In *2017 IEEE 13th International Conference on e-Science (e-Science)*, pages 305–314, Oct 2017.

[73] M. Masoodian, B. Endraß, R. Bühling, P. Ermolin, and E. André. Time-pie visualization: Providing contextual information for energy consumption data. In Ebad Banissi, Hanane Azzag, Mark W. McK. Bannatyne, Stefan Bertschi, Fatma Bouali, Remo Burkhard, John Counsell, Alfredo Cuzzocrea, Martin J. Eppler, Barbara Hammer, Mustapha Lebbah, Francis T. Marchese, Muhammad Sarfraz, Anna Ursyn, Gilles Venturini, and Theodor G. Wyeld, editors, *IV*, pages 102–107. IEEE, 2013.

[74] D. J. Matuszewski, C. I. O. Martins, R. M. C. Jr., J. R. Strickler, and R. M. Lopes. Visual rhythm-based plankton detection method for ballast water quality assessment. In *19th IEEE International Conference on Image Processing, ICIP 2012, Lake Buena Vista, Orlando, FL, USA, September 30 - October 3, 2012*, pages 1009–1012, 2012.

[75] I. Meirelles. *Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations*. EBSCO ebook academic collection. Rockport Publishers, 2013.

[76] Hongying Meng, Nick Pears, Michael Freeman, and Chris Bailey. *Motion History Histograms for Human Action Recognition*, pages 139–162. Springer London, London, 2009.

[77] Thomas B. Moeslund, Adrian Hilton, and Volker Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2–3):90 – 126, 2006. Special Issue on Modeling People: Vision-based understanding of a person's shape, appearance, movement and behaviour.

[78] L. P. C. Morellato. *Phenological data, networks, and research: South America*, pages 75–92. Kluwer Academic Publishers, 2003.

[79] L. P. C. Morellato. *Fenologia de plantas e os efeitos das mudanças climáticas*, pages 181–191. Rima Editora, 2008.

[80] L. P. C. Morellato, L. F. Alberti, and I. L. Hudson. Applications of circular statistics in plant phenology: a case studies approach. In Irene L. Hudson and Marie R. Keatley, editors, *Phenological Research*, pages 339–359. Springer Netherlands, 2010.

[81] L. P. C. Morellato, L. F. Alberti, and I. L. Hudson. Applications of circular statistics in plant phenology: a case studies approach. In Irene L. Hudson and Marie R. Keatley, editors, *Phenological Research*, pages 339–359. Springer Netherlands, 2010.

[82] L. P. C. Morellato, B. Alberton, S. T. Alvarado, B. Borges, E. Buisson, M. G. G. Camargo, L. F. Cancian, D. W. Carstensen, D. F.E. Escobar, P. T.P. Leite, I. Mendoza, N. M. W. B. Rocha, N. C. Soares, T. S. Freire Silva, V. G. Staggemeier, A. S. Streher, B. C. Vargas, and C. A. Peres. Linking plant phenology to conservation biology. *Biological Conservation*, 195:60 – 72, 2016.

[83] L. P. C. Morellato, M. G. G. Camargo, and E. Gressler. *A Review of Plant Phenology in South and Central America*, pages 91–113. Springer Netherlands, 2013.

[84] L. P. C. Morellato, D. C. Talora, A. Takahasi, C. C. Bencke, E. C. Romera, and V. B. Zipparro. Phenology of atlantic rain forest trees: A comparative study. *Biotropica*, 32(4b):811–823, 2000.

[85] M. G. G. Morellato, L. P. C.and Camargo, F. F. D'Eça Neves, B. G. Luize, A. Mantovani, and I. L. Hudson. *The Influence of Sampling Method, Sample Size, and Frequency of Observations on Plant Phenological Patterns and Interpretation in Tropical Forest Trees*, pages 99–121. Springer Netherlands, Dordrecht, 2010.

[86] J. T. Morisette, A. D. Richardson, A. K. Knapp, J. I Fisher, E. A. Graham, J. Abatzoglou, B. E Wilson, D. D. Breshears, G. M. Henebry, J. M. Hanes, and L. Liang. Tracking the rhythm of the seasons in the face of global change: phenological research in the 21st century. *Frontiers in Ecology and the Environment*, 7(5):253–260, 2009.

[87] W. Muller and H. Schumann. Visualization methods for time-dependent data - an overview. In *Proceedings of the 2003 Winter Simulation Conference, 2003.*, volume 1, pages 737–745 Vol.1, Dec 2003.

[88] T. Munzner and E. Maguire. *Visualization analysis and design.* AK Peters visualization series. CRC Press, Boca Raton, FL, 2015.

[89] J. Nahman and D. Perić. Path-set based optimal planning of new urban distribution networks. *International Journal of Electrical Power & Energy Systems*, 85:42 – 49, 2017.

[90] L. E. Newstrom, G. W. Frankie, and H. G. Baker. A new classification for plant phenology based on flowering patterns in lowland tropical rain forest trees at la selva, costa rica. *Biotropica*, 26(2):141–159, 1994.

[91] C-W. Ngo, T-C Pong, and R. T. Chin. Detection of gradual transitions through temporal slice analysis. In *CVPR*, pages 1036–1041. IEEE Computer Society, 1999.

[92] C.-W. Ngo, T.-C. Pong, and H. Zhang. Motion analysis and segmentation through spatio-temporal slices processing. *IEEE Transactions on Image Processing*, 12(3):341–355, 2003.

[93] J. M. Olesen, J. Bascompte, H. Elberling, and P. Jordano. Temporal dynamics in a pollination network. *Ecology*, 89(6):1573–1582, 2008.

[94] G. Oliveira, J. Comba, R. Torchelsen, C. Silva, and M. Padilha. Visualizing running races through the multivariate time-series of multiple runners. In Nina Hirata, Luciana Nedel, Claudio Silva, and Kim Boyer, editors, *SIBGRAPI 2013 (XXV Conference on Graphics, Patterns and Images)*, Arequipa, Peru, august 2013.

[95] J. Ollerton and A.J. Lack. Flowering phenology: An example of relaxation of natural selection? *Trends in Ecology & Evolution*, 7(8):274 – 276, 1992.

[96] P. A. Opler, G. W. Frankie, and H. G. Baker. Rainfall as a factor in the release, timing, and synchronization of anthesis by tropical trees and shrubs. *Journal of Biogeography*, 3(3):231–236, 1976.

[97] C. Parmesan and G. Yohe. A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421:37–42, 01/2003 2003.

[98] Catherine Plaisant. The challenge of information visualization evaluation. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '04, pages 109–116, New York, NY, USA, 2004. ACM.

[99] R. B. Primack. Variation in the phenology of natural populations of montane shrubs in new zealand. *Journal of Ecology*, 68(3):849–862, 1980.

[100] B. Rathcke and E. P. Lacey. Phenological patterns of terrestrial plants. *Annual Review of Ecology and Systematics*, 16:179–214, 1985.

[101] B. C. Reed, J. F. Brown, D. Vanderzee, T. R. Loveland, J. W. Merchant, and D. O. Ohlen. Measuring phenological variability from satellite imagery. *Journal of Vegetation Science*, 5:703–714, 1994.

[102] P. Reys, Camargo, M. G. G. de, Grombone-Guaratini, M. T., A. de P. Teixeira, Assis, M. A., and L. P. C. Morellato. Estrutura e composição florística de um Cerrado sensu stricto e sua importância para propostas de restauração ecológica. *Hoehnea*, 40:449 – 464, 09 2013.

[103] A. D. Richardson, B. H. Braswell, D. Y. Hollinger, J. P. Jenkins, and S. V. Ollinger. Near-surface remote sensing of spatial and temporal variation in canopy phenology. *Ecological Applications*, pages 1417–1428., 2009.

[104] A. D. Richardson, J. P. Jenkins, B. H. Braswell, D. Y. Hollinger, S. V. Ollinger, and M-L. Smith. Use of digital webcam images to track spring green-up in a deciduous broadleaf forest. *Oecologia*, 152(2):323–334, 2007.

[105] D. C. U. M. Rodrigues, F. A. Moura, S. A. Cunha, and R. da S. Torres. Visualizing temporal graphs using visual rhythms: A case study in soccer match analysis. In *8th International Conference on Information Visualization Theory and Applications*, 2017.

[106] A. H. Rosemartin, T. M. Crimmins, C. A.F. Enquist, K. L. Gerst, J. L. Kellermann, E. E. Posthumus, E. G. Denny, P. Guertin, L. Marsh, and J. F. Weltzin. Organizing phenological data resources to inform natural resource conservation. *Biological Conservation*, 173(0):90 – 97, 2014.

[107] C. Rosenzweig, D. Karoly, M. Vicarelli, P. Neofotis, Q. Wu, G. Casassa, A. Menzel, T. L. Root, N. Estrella, B. Seguin, P. Tryjanowski, C. Liu, S. Rawlins, and A. Imeson. Attributing physical and biological impacts to anthropogenic climate change. *Nature*, 453:353–357, 2008.

[108] M. D. Schwartz. Phenoclimatic Measures Phenology: An Integrative Environmental Science. In Mark D. Schwartz, editor, *Phenology: An Integrative Environmental Science*, volume 39 of *Tasks for vegetation science 34*, chapter 21, pages 331–343. Springer Netherlands, Dordrecht, 2003.

[109] Mark D Schwartz, Julio L Betancourt, and Jake F Weltzin. From caprio's lilacs to the usa national phenology network. *Frontiers in Ecology and the Environment*, 10(6):324–327, 2012.

[110] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*, VL '96, pages 336–, Washington, DC, USA, 1996. IEEE Computer Society.

[111] C. G. Silva. Considerações sore o uso de visualização de informação no auxílio à gestão de informação. In *Anais do XXVII Congresso da SBC - XXXIV Seminário Integrado de Software e Hardware*, pages 2070–2084, Rio de Janeiro, RJ, Brasil, 2007.

[112] F. A. O. Silveira, D. Negreiros, N. P. U. Barbosa, E. Buisson, F. F. Carmo, D. W. Carstensen, A. A. Conceição, T. G. Cornelissen, L. Echternacht, G. W. Fernandes, Q. S. Garcia, T. J. Guerra, C. M. Jacobi, J. P. Lemos-Filho, S. Le Stradic, L. P. C. Morellato, F. S. Neves, R. S. Oliveira, C. E. Schaefer, P. L. Viana, and H. Lambers. Ecology and evolution of plant diversity in the endangered campo rupestre: a neglected conservation priority. *Plant and Soil*, 403(1):129–152, 2016.

[113] Natalia Costa Soares and Leonor Patrícia Cerdeira Morellato. Crepuscular pollination and reproductive ecology of trembleya laniflora (melastomataceae), an endemic species in mountain rupestrian grasslands. *Flora*, 238:138 – 147, 2018.

[114] V. Staggemeier and L. P. C. Morellato. Reproductive phenology of coastal plain atlantic forest vegetation: comparisons from seashore to foothills. *International Journal of Biometeorology*, 55(6):843–854, 2011.

[115] John Stasko. Value-driven evaluation of visualizations. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, BELIV '14, pages 46–53, New York, NY, USA, 2014. ACM.

[116] Annia Susin Streher, João Francisco Ferreira Sobreiro, Leonor Patrícia Cerdeira Morellato, and Thiago Sanna Freire Silva. Land surface phenology in the tropics: The role of climate and topography in a snow-free mountain. *Ecosystems*, 20(8):1436–1453, Dec 2017.

[117] D. C. Talora and L. P. C. Morellato. Fenologia de espécies arbóreas em floresta de planície litorânea do sudeste do Brasil 1. *Revta brasil. Bot*, 3(1966):13–26, 2000.

[118] A. C. Telea. *Data Visualization: Principles and Practice, Second Edition*. A. K. Peters, Ltd., Natick, MA, USA, 2nd edition, 2014.

[119] C. Tominski, J. Abello, and H. Schumann. Axes-based visualizations with radial layouts. In *Proceedings of the 2004 ACM Symposium on Applied Computing*, SAC '04, pages 1242–1247, New York, NY, USA, 2004. ACM.

[120] R. S. Torres, C. B. Medeiros, M. A. Gonçcalves, and E. A. Fox. A Digital Library Framework for Biodiversity Information Systems. *International Journal on Digital Libraries*, 6(1):3–17, February 2006.

[121] F. B. Valio, H. Pedrini, and N. J. Leite. Fast rotation-invariant video caption detection based on visual rhythm. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications - 16th Iberoamerican Congress, CIARP 2011, Pucón, Chile, November 15-18, 2011. Proceedings*, pages 157–164, 2011.

[122] C. P. van Schaik, J. W. Terborgh, and S. J. Wright. The phenology of tropical forests: Adaptive significance and consequences for primary consumers. *Annual Review of Ecology and Systematics*, 24:353–377, 1993.

[123] N. Oliveira Vogado, M. G. G. de Camargo, G. M. Locosselli, and L. P. C. Morellato. Edge effects on the phenology of the guamirim, myrcia guianensis (myrtaceae), a cerrado tree, brazil. *Tropical Conservation Science*, 9(1):291–312, 2016.

[124] G. R. Walther. Plants in a Warmer World. *Perspectives in Plant Ecology, Evolution and Systematics*, 6(3):169–185, 2003.

[125] G. R. Walther, E. Post, P. Convey, A. Menzel, C. Parmesan, T. J. C. Beebee, J-M. Fromentin, O. Hoegh-Guldberg, and F. Bairlein. Ecological responses to recent climate change. *Nature*, 416(6879):389–395, March 2002.

[126] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

[127] M. Weber, M. Alexa, and W. Ḿuller. Visualizing time-series on spirals. In *Proceedings of the IEEE Symposium on Information Visualization*, INFOVIS '01, page 7, Washington, DC, USA, 2001. IEEE Computer Society.

[128] P. Chung Wong and R. Daniel Bergeron. 30 years of multidimensional multivariate visualization. In *Scientific Visualization, Overviews, Methodologies, and Techniques*, pages 3–33, Washington, DC, USA, 1997. IEEE Computer Society.

[129] H. Zhang, Y. Hou, D. Qu, and Q. Liu. Correlation visualization of time-varying patterns for multi-variable data. *IEEE Access*, 4:4669–4677, 2016.

[130] Xiaoyang Zhang, Mark A. Friedl, Crystal B. Schaaf, Alan H. Strahler, John C.F. Hodges, Feng Gao, Bradley C. Reed, and Alfredo Huete. Monitoring vegetation phenology using modis. *Remote Sensing of Environment*, 84(3):471 – 475, 2003.

[131] F. Zhao and A. K. H. Tung. Large scale cohesive subgraphs discovery for social network visual analysis. *PVLDB*, 6(2):85–96, 2012.

[132] Torre Zuk, Lothar Schlesier, Petra Neumann, Mark S. Hancock, and Sheelagh Carpendale. Heuristics for information visualization evaluation. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, BELIV '06, pages 1–6, New York, NY, USA, 2006. ACM.