



# Modeling plant phenology database: Blending near-surface remote phenology with on-the-ground observations



Greice C. Mariano<sup>a,\*</sup>, Leonor Patricia C. Morellato<sup>b</sup>, Jurandy Almeida<sup>a,c</sup>, Bruna Alberton<sup>b</sup>, Maria Gabriela G. de Camargo<sup>b</sup>, Ricardo da S. Torres<sup>a</sup>

<sup>a</sup> Institute of Computing, University of Campinas (UNICAMP), 13083-970 Campinas, SP, Brazil

<sup>b</sup> Institute of Biosciences, Universidade Estadual Paulista (UNESP), 13506-900 Rio Claro, SP, Brazil

<sup>c</sup> Institute of Science and Technology, Federal University of São Paulo (UNIFESP), 12247-014 São José dos Campos, SP, Brazil

## ARTICLE INFO

### Article history:

Received 16 June 2015

Received in revised form 7 December 2015

Accepted 3 March 2016

Available online 15 March 2016

### Keywords:

Database design

Leafing

Modeling

Digital images

Remote phenology

Image-based phenological indices

## ABSTRACT

Phenology research handles multifaceted information that needs to be organized and made promptly accessed by scientific community. We propose the conceptual design and implementation of a database to store, manage, and manipulate phenological time series and associated ecological information and environmental data. The database was developed in the context of the e-phenology project and integrates ground-based conventional plant phenology direct observations with near-surface remote phenology using repeated images from digital cameras. It also includes site-base information, sensor derived data from the study site weather station and plant ecological traits (e.g., pollination and dispersal syndrome, flower and fruit color, and leaf exchange strategy) at individual and species level. We validated the database design through the implementation of a Web application that generates the time series based on queries, exemplified in two case studies investigating: the relationship between flowering phenology and local weather; and the consistency between leafing patterns derived from ground-based phenology on leaf flush and from vegetation image indices (%Green). The database will store all the information produced in the e-phenology project, monitoring of 12 sites from cerrado savanna to rainforest, and will aggregate the legacy information of other studies developed in the Phenology Laboratory (UNESP, Rio Claro, Brazil) over the last 20 years. We demonstrate that our database is a powerful tool that can be widely used to manage complex temporal datasets, integrating legacy and live phenological information from diverse sources (e.g., conventional, digital cameras, seed traps) and temporal scales, improving our capability of producing scientific and applied information on tropical phenology.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

Plant phenology, the study of recurrent life cycles events and its relationship to climate, is a key discipline in climate change research (Schwartz, 2013). Plant phenology studies are based on a well-defined workflow, as shown in Fig. 1. These studies deal with a large amount of complex data and are more challenging in the tropics, where the species diversity is very high, and the phenological monitoring is usually continuous, involving a large number of marked trees, which are visited at regular intervals in order to observe changes on phenophases of flowers, fruits, and leaves (d'Eça Neves and Morellato, 2004; Morellato et al., 2000,

2010, 2013; Staggemeier and Morellato, 2011; Frankie et al., 1974; Williams et al., 1999; Borchert, 1983; Augspurger, 1983).

Also, the phenology data collection in the tropics is mostly individual-based, with several individuals observed per species over regular intervals (d'Eça Neves and Morellato, 2004; Morellato et al., 2010). In this scenario, the amount of information to be handled can be very high. For instance, in the short-term phenology study at the Atlantic rain forest conducted by Morellato et al. (2000), they worked on four sites, and observed 46–129 species and 277–422 trees per site, six phenophases (flower buds, open flowers, unripe fruits, ripe fruits, leaf flush, and leaf fall) over 17 months. Therefore, for one given month they can totalize about 7698 observation records and a total of about 130,866 phenology data cells for the complete study. Those records do not account for the ecological information (three seed dispersal classes and three leaf exchange functional groups) or climatic data.

\* Corresponding author.

E-mail address: [gmariano@ic.unicamp.br](mailto:gmariano@ic.unicamp.br) (G.C. Mariano).

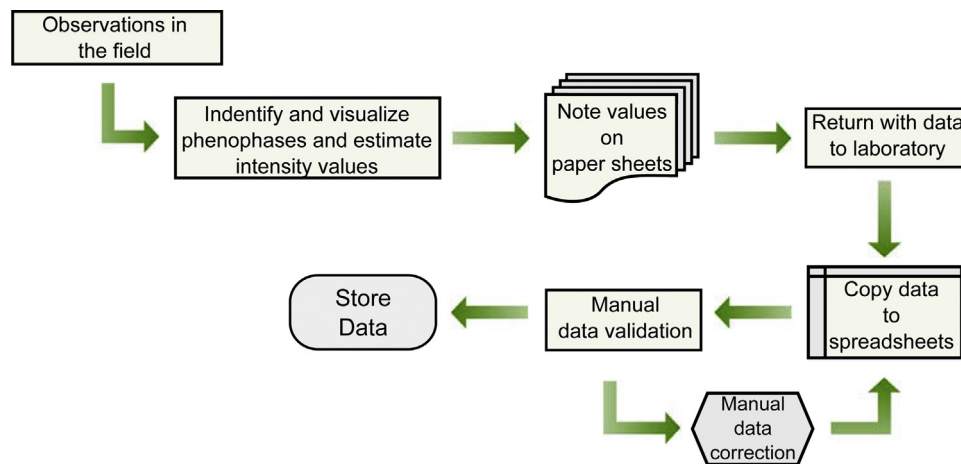


Fig. 1. Workflow of a typical on-the-ground phenology data acquisition task.

Plant phenology studies are concerned with the identification and understanding of temporal changes in phenological events (e.g., leafing, senescence, flowering, fruiting), known as phenophases (Morellato et al., 2000, 2010; d'Eça Neves and Morellato, 2004). Therefore, a special feature of any phenology information is that the data is temporal, i.e., continuously collected along a given period of time.

Usually, plants are observed directly in the field with the naked eye or with binoculars and the phenophases defined by the investigators are visually identified and registered on paper sheets (in the field) and the information is inserted on digital spreadsheets (in the laboratory) – see the workflow in Fig. 1. Later, data consistency analysis is performed to ensure data quality. For instance, in the case of temporal consistency, the plant fruiting phase must occur after a flowering event. Finally, it is performed a data analysis to test pre-defined hypothesis (as exemplified in the study cases, Section 3.3). All these steps require the use of multiple tools and are executed mainly manually. This, of course, brings about all kinds of interoperability problems due to system mismatch, data diversity, and variety of user profiles. In practical situations, those steps are time consuming and error prone. Therefore, extracting knowledge from such material becomes more and more difficult as additional data are collected.

A suitable alternative to address these challenges relies on the use of a conceptual design to model observation data and user needs with the aim of facilitating knowledge discovery. In phenology, these requirements may vary depending on the vegetation type and study goals. Some initiatives have been proposed for temperate vegetation zones. There are some phenology information systems and data sets available for download or review, especially for phenology networks dealing with legacy data sets and citizen science (Pan European Phenology Project PEP725<sup>1</sup>; National phenology network – USA NPN<sup>2</sup>; and Nature's Calendar.<sup>3</sup>) However, we were unable to find a information system for tropical phenology or any other database modeled to accommodate phenology patterns derived from diverse observations techniques and also considering ecological and environmental data.

An effective way to overcome such drawbacks is by adopting phenology information systems. In general, these systems can efficiently deal with a large amount of observational data by using

database management systems, which offer a large set of tools to support the indexing, processing, and retrieval of data (Almeida et al., 2014, 2015). However, just a few phenology database systems and scripts are available. While those systems manage large amounts of phenology data (Dierenbach et al., 2013), they do not handle the complexity of information we demand for our proposed database (see below). For instance, we found the Plant-Phenological On Line Database<sup>4</sup> offering access to phenological observations of plant species from Central Europe, largely Germany, from about 9000 stations, covering 130 years (1880–2009), encompassing more than 16 millions observations (Dierenbach et al., 2013), but no ecological information is considered or included in the database. There are some suggestions available on the USA National Phenology Network (USA NPN),<sup>5</sup> but again no scripts are available or published.

Motivated by the benefits of a database and by the lack of appropriate systems, we introduce here a phenological data model to guide the implementation of information systems dedicated to phenology studies. The possibility of using a single tool associated with an integrated data model in the management of different kinds of data may speed up the analysis process and, at the same time, allow for handling novel and more complex searches involving their combination that would not be possible using multiple tools. The goal of this article is to present a database conceptual model to be used in the construction of phenological information systems that support plant phenology studies, specially suited for tropical ecosystems.

The database model will be freely available for download from the e-phenology project home page. Published data and some data sets will be made freely available for the community in the e-phenology homepage. Also, any data series can be made available under request to the e-phenology project coordinators.

## 2. Materials and methods

This section presents the e-phenology Project, which serves as case study for the design and implementation of a plant phenology database to support integrative phenological studies focused in tropical systems.

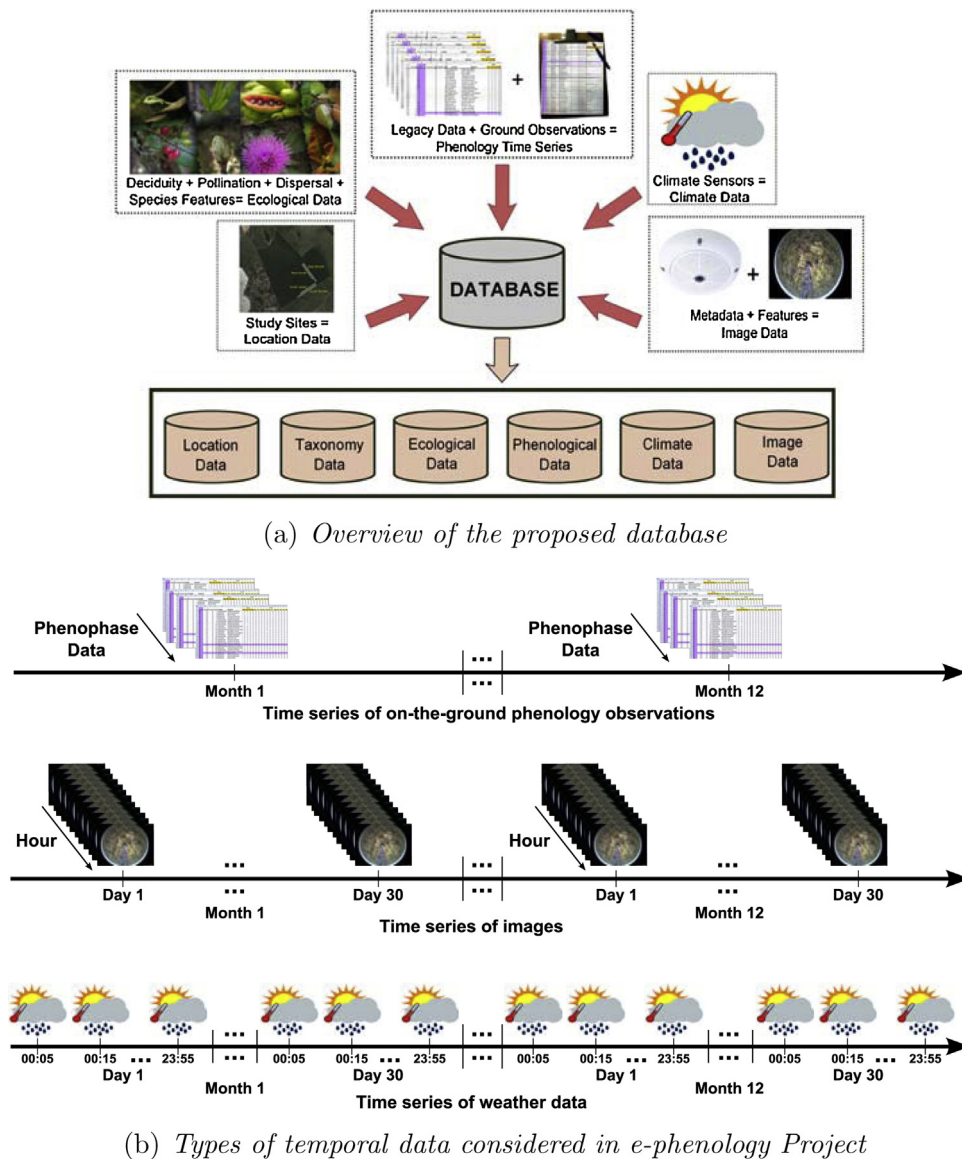
<sup>1</sup> <http://www.pep725.eu/> (as of December 7, 2015).

<sup>2</sup> <https://www.usanpn.org/> (as of December 7, 2015)

<sup>3</sup> <http://www.naturescalendar.org.uk/home.html> (as of December 7, 2015)

<sup>4</sup> <http://www.ppodb.de/> (as of December 7, 2015).

<sup>5</sup> <https://www.usanpn.org/> (as of December 7, 2015).



**Fig. 2.** Types of data considered in the database designed in the context of the e-phenology project. (a) We present an overview of the proposed database with non-temporal and temporal data; (b), we present the different time scales for temporal data.

### 2.1. The e-phenology project

The e-phenology<sup>6</sup> is a multidisciplinary project that combines research in computer science and phenology. Its main goal is to solve theoretical and practical problems involved in the use of new technologies for phenological observations, to detect local environmental changes, and to understand the effects of global warming in the tropics. The project involves:

- the use of new technologies for environment monitoring based on remote monitoring phenology systems;
- the creation of a protocol for a long-term phenology monitoring program in Brazil and the integration between different disciplines, advancing the current knowledge of seasonal responses within the tropics with respect to climate changes;

- providing models, methods, and algorithms to support the analysis, integration, and management of data from remote phenology systems.

Fig. 2 presents the types of data managed in the database designed in the context of the e-phenology project. As it can be observed at Fig. 2(a), these data include: *location* data that contains information about sites where phenological studies are conducted; *taxonomical* and *ecological* data about species and individuals; *legacy* data associated with on-the-ground observations, usually stored in non-conventional database systems using non-standard data insertion procedures; *weather* data obtained from sensors; and *image-related* phenological indices and metadata extracted from digital cameras (near-surface remote phenology).

An important feature of the e-phenology project is the acquisition of temporal data with different time scales. For example, in our tropical study, 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;

<sup>6</sup> <http://www.recod.ic.unicamp.br/ephenology/> (as of December 7, 2015).

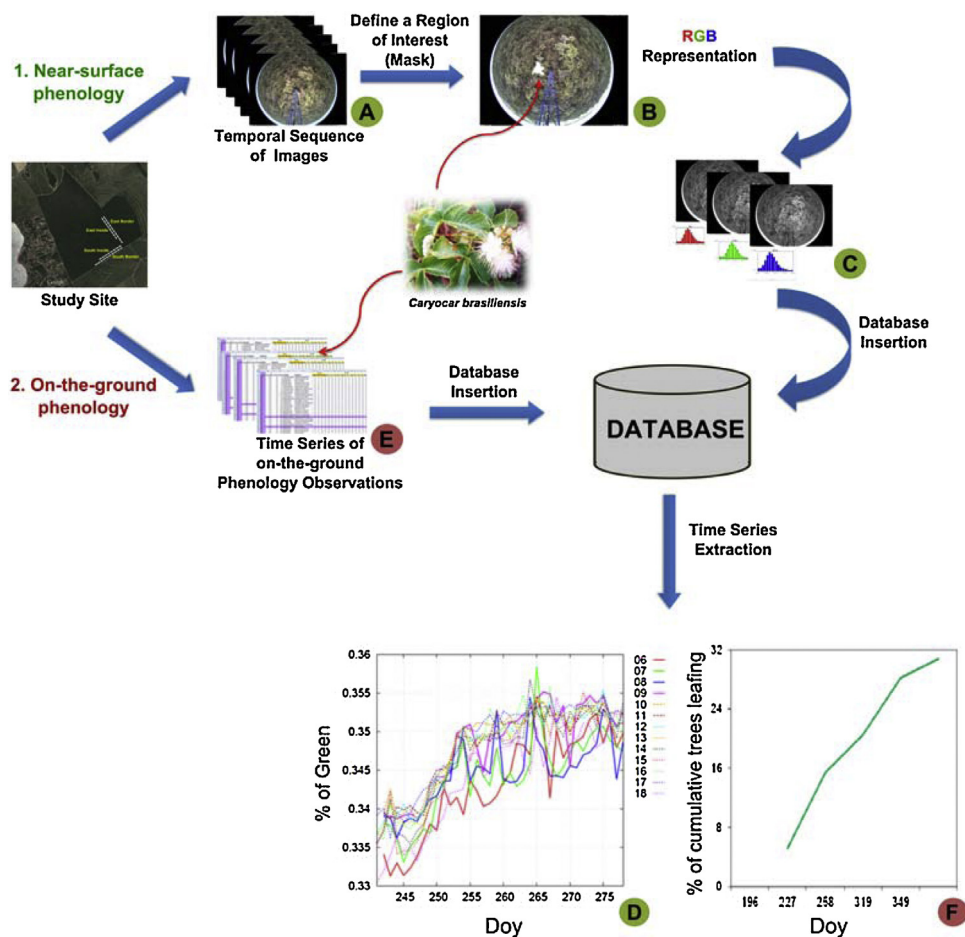


Fig. 3. Overview upon data integration and time series extraction from near-surface images and on-the-ground phenological observations.

and the weather data are obtained from different types of meteorological sensors every 10 min, daily, as shown Fig. 2(b). Therefore, the e-phenology project needs a database able to manage, store, and perform queries that integrate these heterogeneous data. Fig. 3 illustrates the integration of results between the data collected by digital images on near-surface remote phenology observations with the on-the-ground observations.

## 2.2. Study area

The core study area of the e-phenology project is a Cerrado *sensu stricto*, a savanna vegetation located at Itirapina (22° 10' 49.18" S/47° 52' 16.54" W), São Paulo State, Brazil (see Fig. S1 at Electronic Supplementary Material 1 – ESM1). The cerrado stricto *sensu* (Coutinho, 1978) is dominated by a discontinuous woody component reaching six to seven meters high and a continuous herbaceous layer, at some parts, the vegetation is denser, with some trees reaching up to 12 m high (Camargo et al., 2013; Reys et al., 2013). The cerrado savanna study site is about 260 ha, 610 m altitude and the regional climate is Cwa type (i.e., humid subtropical climate) according to Köppen classification (Köppen, 1931), with a cold, dry season from April to September and a warm, wet season from October to March.

A complete meteorological station were installed to collect data on rain precipitation, humidity, wind speed, temperature, among others in a 18 m tower, named Phenology Tower. In mid-August 2011, a digital camera was set up at the top of the Phenology tower to take daily digital images of the cerrado vegetation (Alberton et al., 2014).

The digital hemispherical lens camera (Mobotix Q24) was configured to take a daily sequence of five JPEG images (at 1280 × 960 pixels of resolution) per hour, from 6:00 to 18:00 h (UTC-3). At the same area, on-the-ground observations have been conducted monthly since September 2004 on 2122 individual marked trees for changes in flowering (flower buds and open flowers or anthesis), fruiting (unripe and ripe) and leafing (leaf fall and leaf flush), a total of six phenophases.

Therefore, in our core study site, we collect in a regular basis all the environmental, ecological, and phenological data to be integrated in the proposed database.

## 2.3. Requirements

In this section, we present a set of requirements, which were defined in the context of the e-phenology project, that guided the database modeling process.

### 2.3.1. Legacy data

In practical situations, phenology researchers often do not use conventional databases for storing their data, which usually are stored into spreadsheets. Legacy data sets are phenological data usually collected during a long period of time for a determined species, population, or community and may include from a few hundred thousands to more the millions of entries (Dierenbach et al., 2013). Retrieve information from those data sets to summarize temporal patterns is time consuming and error prone.

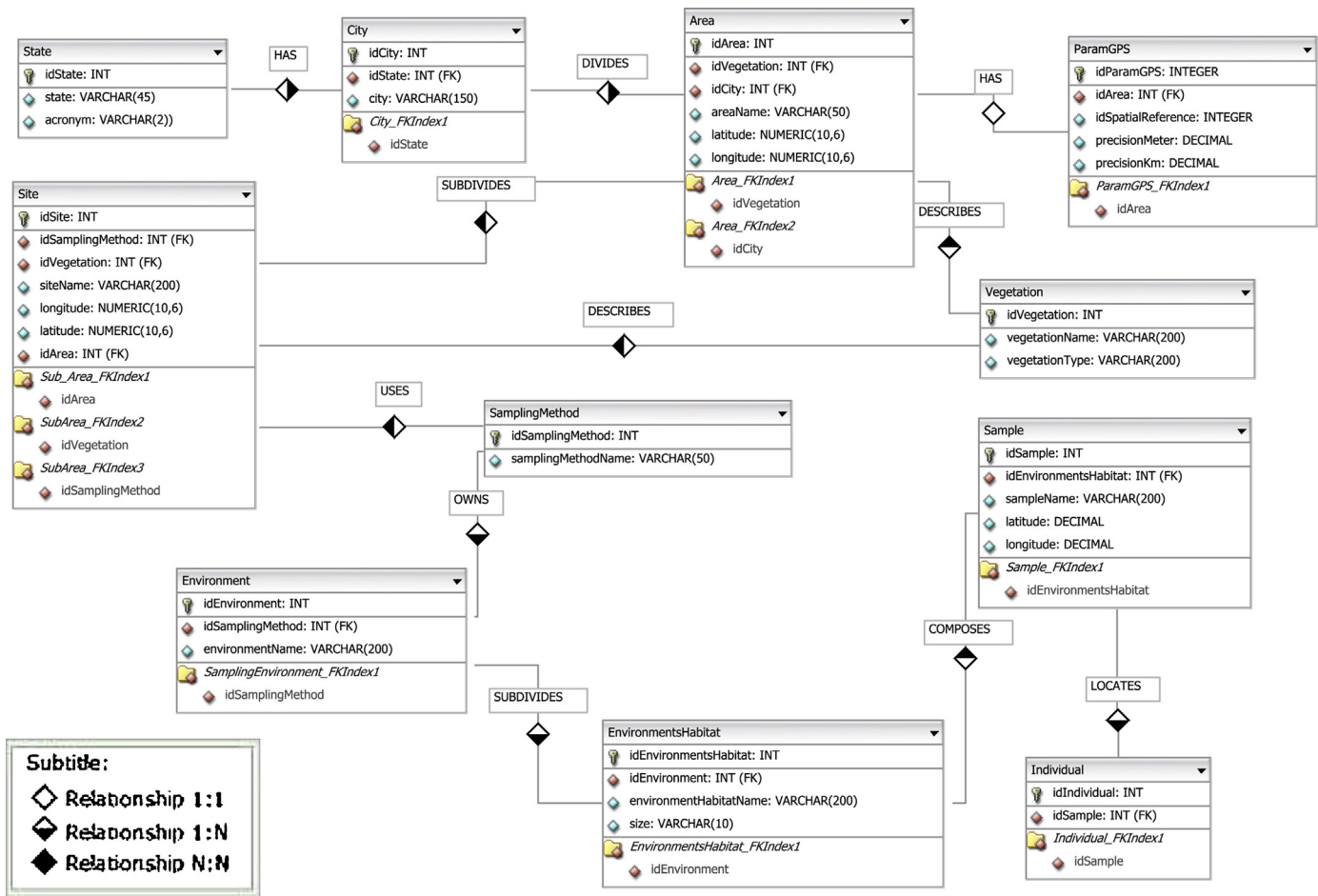


Fig. 4. Entity-relationship diagram related to Location data.

*Requirement 1: Database modeling should support the management of legacy data usually stored in non-conventional database systems using non-standard data insertion procedures.*

That demands the construction of domain-specific tools for data migration and insertion. Common problems faced include existence of missing values, and the ambiguity in recording phenology observations, that will be indicated during the input of data and easily corrected by the user. In the context of the e-phenology project, we have created a set of scripts for handling these data insertion problems.

### 2.3.2. Temporal data

Phenological data are necessarily temporal and cyclical. In fact, phenology studies are based on identifying changes/patterns on temporal/cyclical data. Basic phenology research aims to answer questions related to when (time of the year), for how long (duration), and at which amplitude (e.g., the intensity in number of species and individuals) a given event has occurred at that specific site or region (Morellato et al., 2000, 2010).

*Requirement 2: Database modeling should support the management in terms of storage and query processing of large volumes of temporal data.*

### 2.3.3. Data heterogeneity

Phenology studies usually rely on data of different types and formats. Besides the temporal data obtained on the basis of regular observations of plant phenophases (e.g., flower, fruit, and leaf changes), researchers relate that information to other temporal

(climate), ecological (e.g., fruit color or plant habit), taxonomic or evolutionary characteristic (phylogenetic relationship) of species.

*Requirement 3: Database modeling should support the management of ecological, taxonomic, climate, and phenological data.*

### 2.3.4. Support of complex data query

One common practice in phenology studies is the establishment of correlations among ecological and phenological data, and among climate variables and phenological indices (Morellato et al., 2000, 2010; Fournier, 1974; Newstrom et al., 1994). For instance, many researchers correlate flowering to precipitation, temperature, and day length with the aim of verifying which one would be the proximate cue for that phenophase.

*Requirement 4: Database systems should support the execution of complex queries, with the aim of extracting information that may support complex phenological data analysis.*

### 2.3.5. Image-related phenological indices

The near-surface remote phenology method applied in this manuscript relies on the use of cameras to obtain images of a cerado vegetation (see Fig. 3) as described by Alberton et al. (2014). This method has been efficiently applied as multi-channel imaging sensors to track vegetation changes over time (Richardson et al., 2007, 2009; Alberton et al., 2014).

A region of interest (ROI) is selected in the original image and is defined as a *mask*. A mask is a binary image created to delimitate the ROI within the image and usually is the individual crown of a particular species (Alberton et al., 2014) – see Fig. 3B. Color

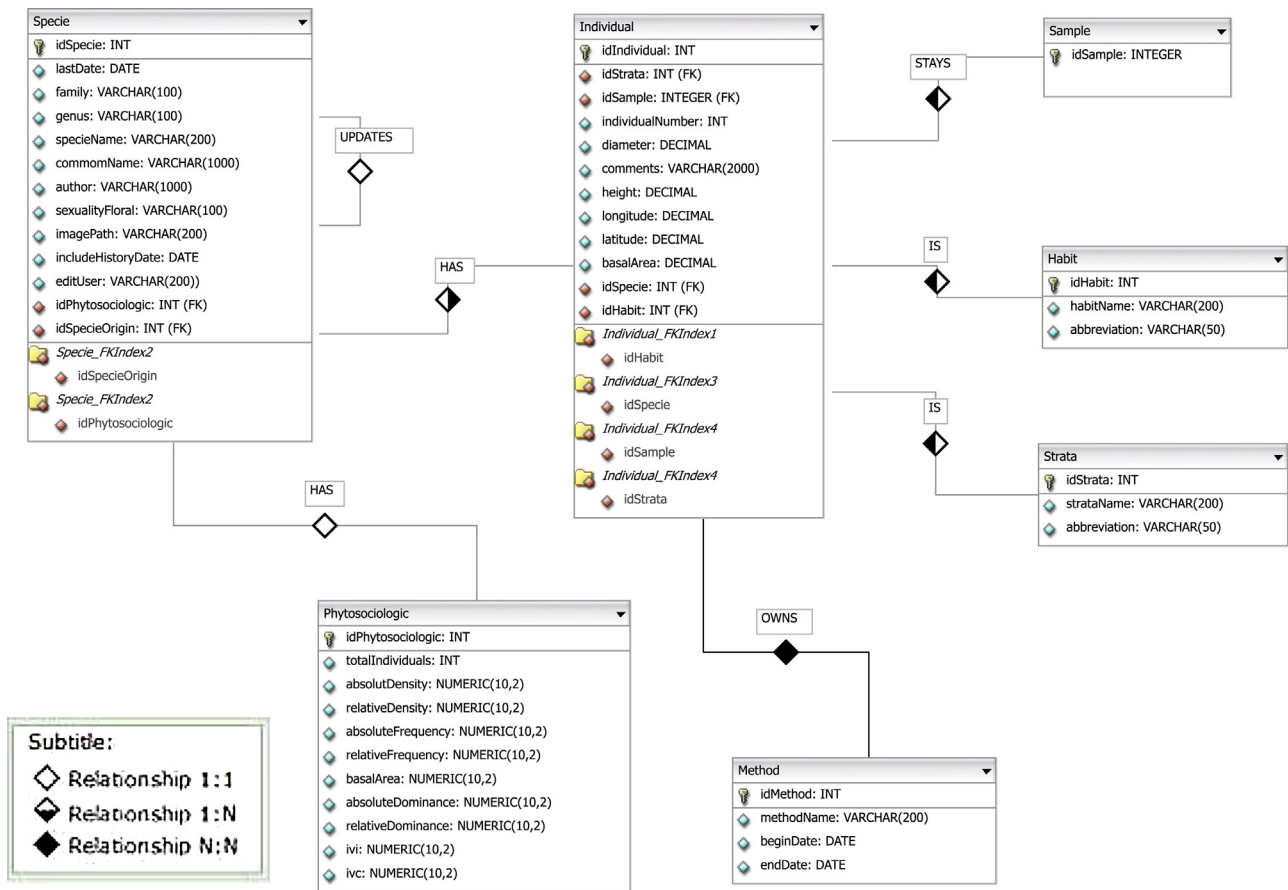


Fig. 5. Entity-relationship diagram related to Taxonomic data.

channels RGB (Red, Green, and Blue) are extracted from ROIs of digital images, and for instance, the changes in the green channel indicate changes over time on vegetation leaf color and therefore leafing phenology (Alberton et al., 2014).

Further, the vegetation indices are extracted from those images aiming to encode different phenological patterns over time (Fig. 3C and 3D). One example of index is the Green chromatic coordinate broadly used in phenological studies with cameras monitoring the vegetation (Richardson et al., 2007, 2009; Alberton et al., 2014).

In the case of the e-phenology project, a digital hemispherical lens camera (Mobotix Q24) was setup at the top of the phenology tower (Alberton et al., 2014). The camera was configured to take a daily sequence of five JPEG images (at 1280 × 960 pixels of resolution) per hour, from 6:00 to 18:00 h (UTC-3).

*Requirement 5: Database modeling should support the management of image-related phenological data, which are metadata and time series extracted from different color channels.*

### 3. Results

This section presents the proposed database model to handle phenological data, as well as implementation issues related to the implementation of an information system to manage these data.

#### 3.1. Database design

Given its complexity, the proposed data model was divided into six modules defined according to the kind of the data managed: Location Data, Ecological Data, Taxonomic Data, Phenological Data, Climate Data, and Image Data. These modules are presented below.

##### 3.1.1. Location data

The Location module refers to data associated with the position of plant individuals. Fig. 4 presents the entity-relationship diagram (Elmasri and Navathe, 1999) of this module. The main tables in this module are: *Site* that locates an individual within an area (*Area*), and *SamplingMethod* that determines which sampling method was used to select plant individuals in the phenological observation process.

According to the sampling method employed, the way to sample an individual changes (for example transects and parcels) (d'Eça Neves and Morellato, 2004). In some circumstances, it is necessary to divide the Site into environments (*Environment*), and the environments into habitats (*EnvironmentHabitat*). Finally, to limit the perimeter to survey each habitat, we use the transect method (*SamplingMethod*), which defines the exact local (*Sample*) where each sampled individual is located.

For example, in the case of the e-phenology project, the collected data refer to individuals (*Individual*) that are located within an area (*Area*) situated in the Itirapina county (*City*) in the São Paulo State (*State*) – see Fig. S1 at Electronic Supplementary Material 1 – ESM1. This Area is called Botelho and has one type of vegetation (*Vegetation*) and, for this reason, it has only one study site (*Site*) also named Botelho. Each site uses one sampling method (*SamplingMethod*), e.g., Transect. This method is applied across two environments (*Environment*): East and South; and further these environments are split into habitats (*EnvironmentHabitat*): border and interior. Thus, the Habitats are sampled using Transects that identified where the individuals are located (*Sample*). For instance, the East and South Borders were sampled by ten Transects each one, while the East Interior and South Interior by eight transects each, summing up are 36 transects or sample unities where

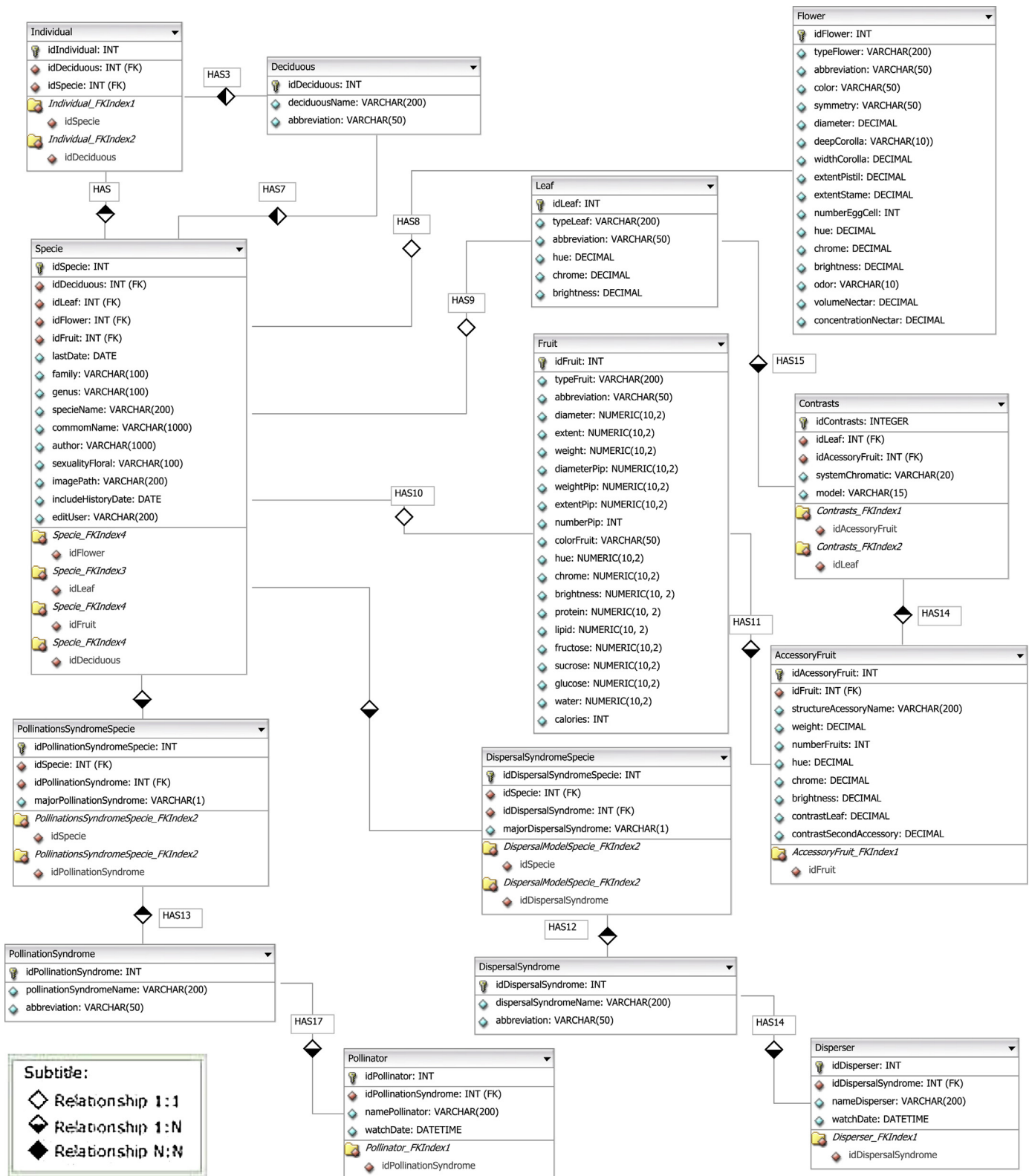


Fig. 6. Entity-relationship diagram related to Ecological data.

individuals can be found (Camargo et al., 2013; Alberton et al., 2014).

### 3.1.2. Taxonomical data

The Taxonomy module includes tables that describe individuals. Fig. 5 presents the entity-relationship diagram of this module. The

main tables of this module are: *Individual* that represents all plants within an area; and *Species* that represents groups of individuals of the same biological species and that share similar features.

Associated with one species, there are quantitative data that indicate the occurrences and the variation in space of specific features of this species, named phytosociological (Mueller-Dombois

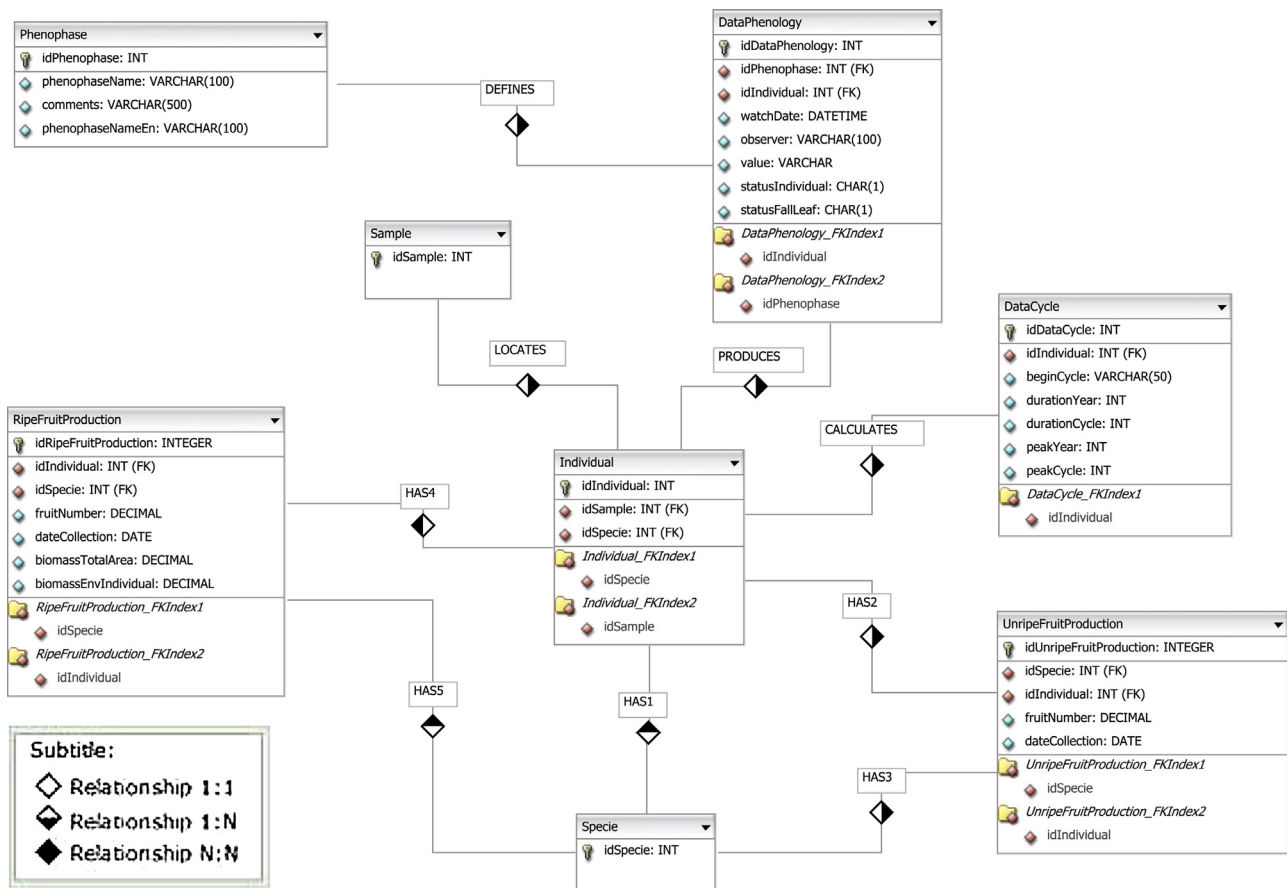


Fig. 7. Entity-relationship diagram related to *Phenological* data.

and Ellenberg, 2013) data (*Phytosociological*). One example of quantitative data is the number of individuals per sample unit (transect), environment (South or East) or by environmentHabitat (interior or border). These data are used to assess the structure of plant communities (Reys et al., 2013).

For the e-phenology project, as well as many other projects involving biodiversity data, changes in the species taxonomy are very important. Species may suffer taxonomic changes over time, which complicate the work of biologists when interested in group the individuals, for instance, by species or genera. Many times, biologists have to search up-to-date information on catalogs that record the taxonomy associated with species, indicating common names, variations over time (synonyms), and authors responsible for defining the species name. It is increasing the number of initiatives targeted to resolve the inconsistencies generated by these changes (Mira et al., 2013). In our database, we chose to track changes in taxonomic information over time. Both old and novel nomenclatures are recorded. This information is stored in table *Species*.

Individuals are characterized by both a habit (*Habit*) and by a strata (*Strata*). A habit defines the type of a plant, for instance, tree or shrub. The Strata represents the vertical position of the plant in the vegetation. For example, the strata of a creeping non-woody plant is understory and the habit is called herbaceous.

Table *Method* stores the kind of on-the-ground observations. For example, data can be collected using a qualitative method to assess the presence or absence of phenophase or the researcher may use a semi-quantitative method that assigns a different number for a phenophase, depending on its intensity (Fournier, 1974; Bencke and Morellato, 2002; Morellato et al., 2010; d'Eça Neves and Morellato, 2004) or a quantitative method storing the number of fruits or flowers (Camargo et al., 2013).

### 3.1.3. Ecological data

The Ecology module includes the tables that store features of both individuals and species (Fig. 6). The main tables of this module are: *Leaf*, *Flower*, and *Fruit*. These data are usually used to identify and classify species and individuals.

Other tables in this module describe the pollination (*Pollination*) and dispersal (*Dispersal*) modes or syndromes for different species. The pollination process requires different pollinator agents (*Pollinator*), such as animals or wind. For each pollinator agent, there is a correspondent designation, for example, for animals the pollination model is zoophily. Several pollination model (*PollinationModelSpecie*) can be associated with a single species.

Similarly, dispersal modes or syndromes occur in several ways. Usually, they also depend on disperse agents (*Disperser*). For example, the dispersal mode involving animals is called zoochory. Species can also have several dispersal agents (*DispersalModelSpecie*).

### 3.1.4. Phenological data

The Phenology module includes the tables that contain data related to on-the-ground plant phenology observations (Fig. 7). The main table is *DataPhenology*, which stores phenological data upon individuals.

The phenological data observations can vary over days, months, or years, which define time series. The stored information changes depending on the sampling method used in the phenological research. For the e-phenology project, data managed in this module refer to values of observations that measure the intensity of a phenophase (*Phenophase*) for an individual. With the measure of intensity some statistics values can be computed and stored in table *DataCycle* for future analysis.

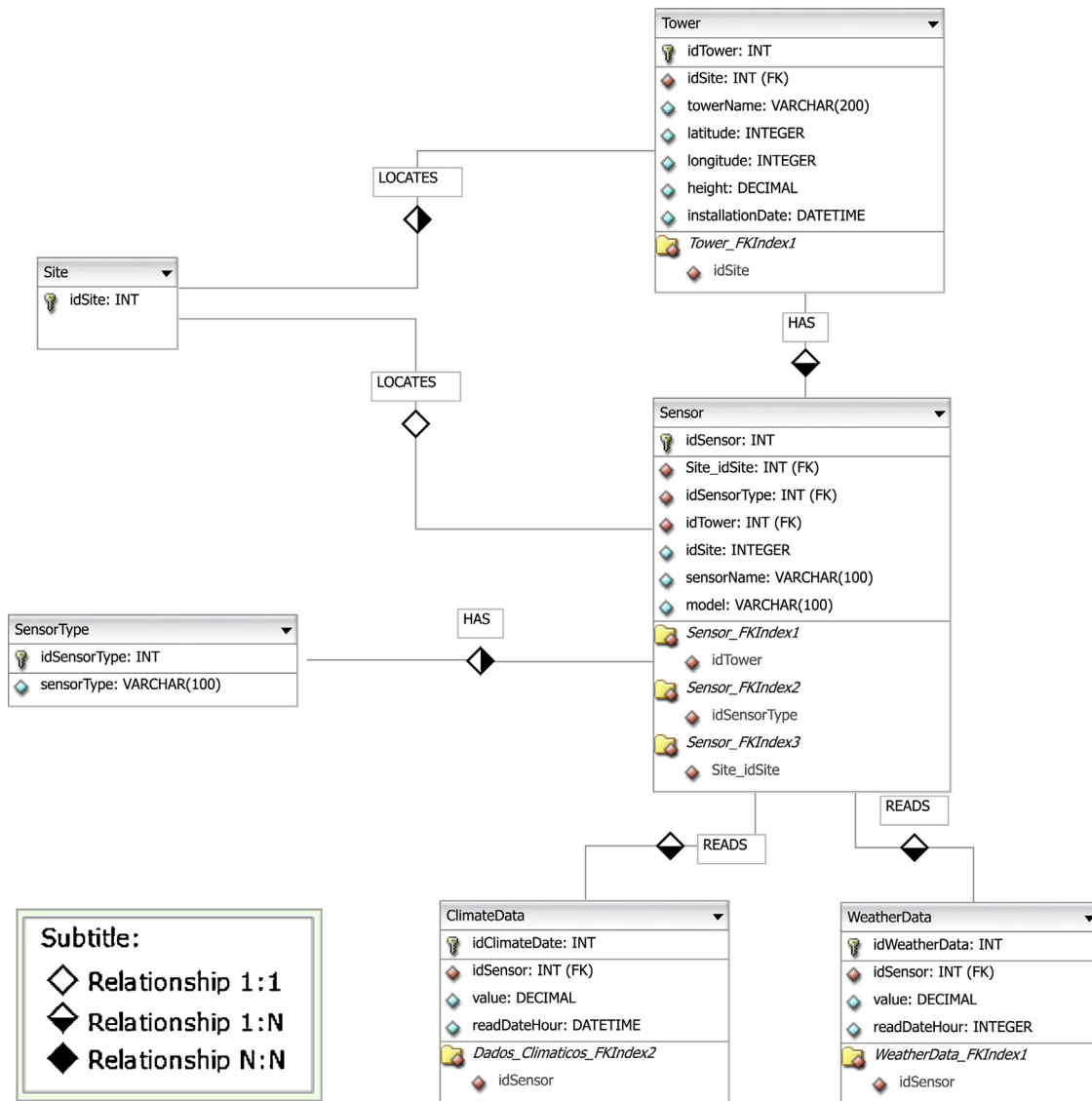


Fig. 8. Entity-relationship diagram related to Climate data.

### 3.1.5. Climate data

The Climate module includes tables related to the monitoring of temporal environmental data (Fig. 8). The main table is *WeatherData*, which contains data collected by various sensors installed within an area of study.

At the e-phenology core study area, the Phenology tower has a complete weather station with nine different sensors (*Sensor*) installed to collect weather data. Examples include moisture, air and soil temperature, and rain gauge sensors (*SensorType*).

### 3.1.6. Image-related data

This module includes the tables that store data related to images obtained by digital cameras. The main table is *Image* that records data about the camera, date, hour, and the path of an image (Fig. 9).

Near-surface phenology studies rely on the use of cameras (*Camera*) to obtain digital images of the vegetation. Later, features (*FeatureVector*) are extracted from those images aiming at encoding different phenological patterns over time (Alberton et al., 2014; Torres et al., 2013; Almeida et al., 2013, 2013). One example is the average green (*Descriptor*) associated with the regions of interest (ROIs), usually defined in masks (*ImageMaskSpecie*) – see Fig. 3 and (Alberton et al., 2014).

## 3.2. Implementation issues

This section provides an overview on typical queries, as well as employed technologies used to implement and validate the proposed data model.

### 3.2.1. Typical queries

Typical queries are classified into two groups: single and complex queries. Single queries are related to results obtained from a single table. In this case, the results are obtained by using common Relational Algebra operators (Elmasri and Navathe, 1999). One example is “Show all records related to species.” Such query requires only the use of a selection operation that retrieves all records of table *Species*. Complex queries, in turn, involve multiple tables. One example of complex query is “Show the properties of the flowers, fruits, and leaves of all species.” In this case, the result combines information from the *Species*, *Flower*, *Fruit*, and *Leaf* tables.

In the context of the e-phenology project, most of the queries involve to summarize temporal data into phenological variables. Data on phenophases, individuals, and species are then related to data that vary over time, such as climate data. One example of

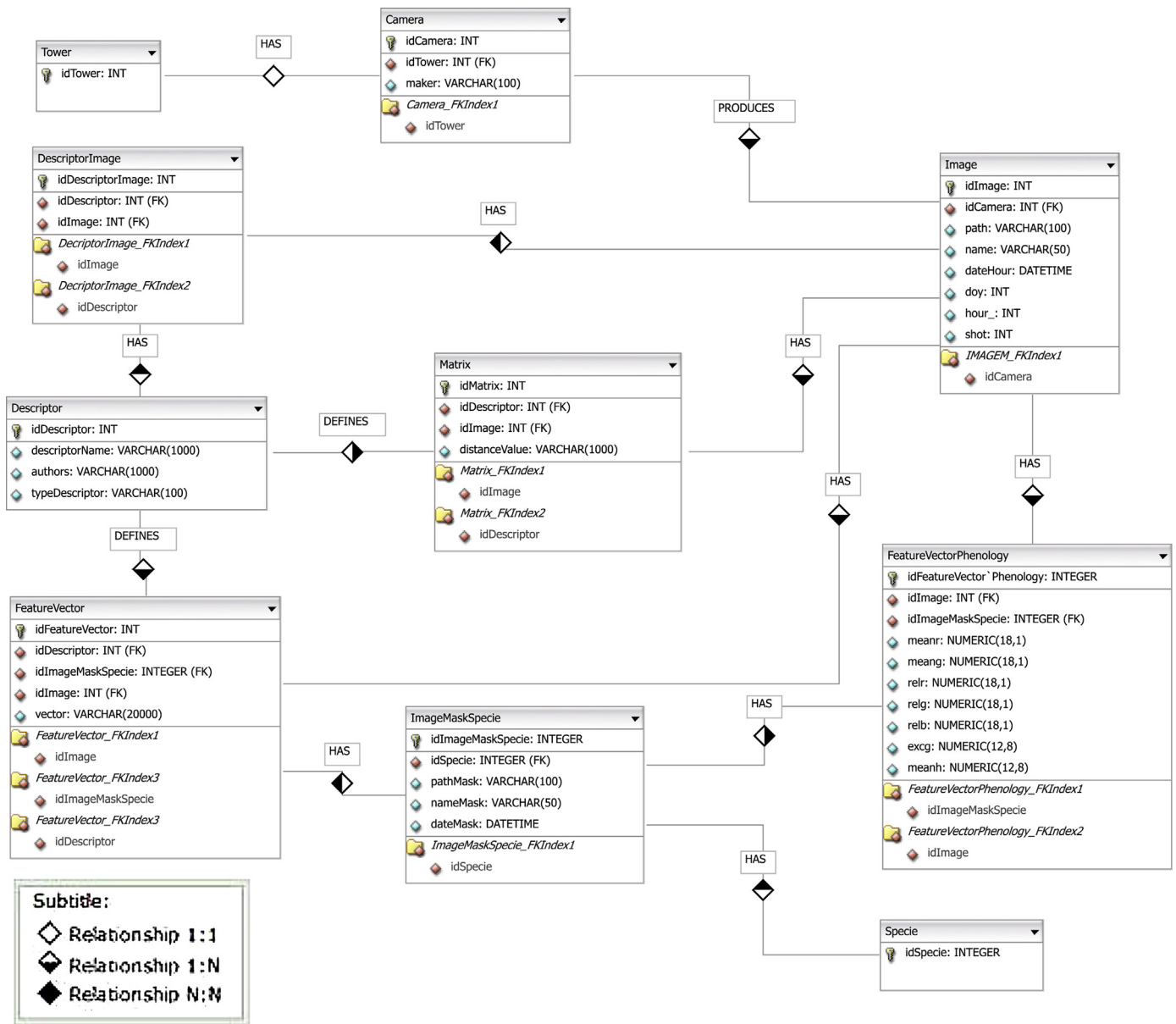


Fig. 9. Entity-relationship diagram related to Image-related data.

temporal query is “Which individuals have phenophase leaf flushing intensity greater than 0, from September of 2005 until September of 2007?” To process this query, it is necessary to perform join operations involving the tables *Individuals*, *Species*, *Phenophases*, and *DataPhenology*.

Some other complex queries that are of interest in the context of the e-phenology project are:

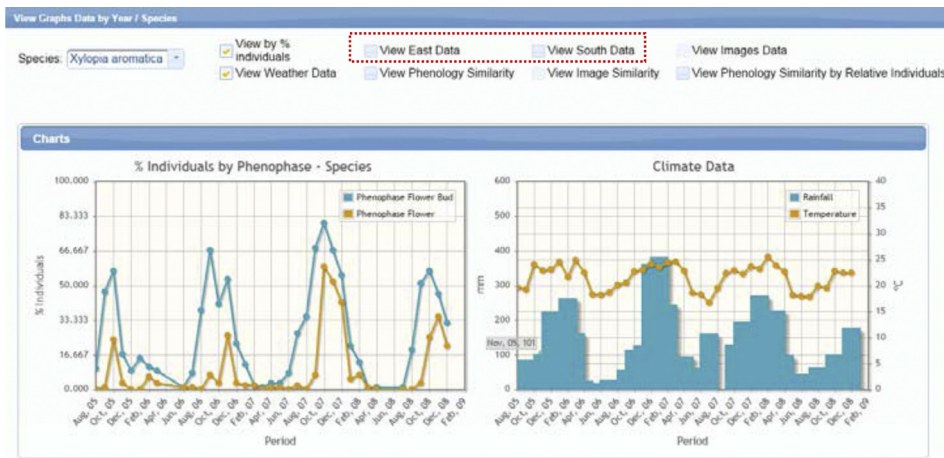
- “If there is a variation in the green channel of vegetation images, is this related to the period of species flushing new leaves on-the-ground?”;
- “If there is a variation of red channel (extracted from the RGB standard), can this be related to the period of species leaf fall?”;
- “When the leaf flushing phenophase begins for a particular species, also is there an increasing in the precipitation rate? What about the temperature variation?”;
- “Does the peak of leaf flushing coincide with the highest average values observed for the green channel?”

- “When is the peak of the a phenophase intensity for individuals of a particular species?”;
- “How long is a phenophase cycle for individuals of a particular species?”.

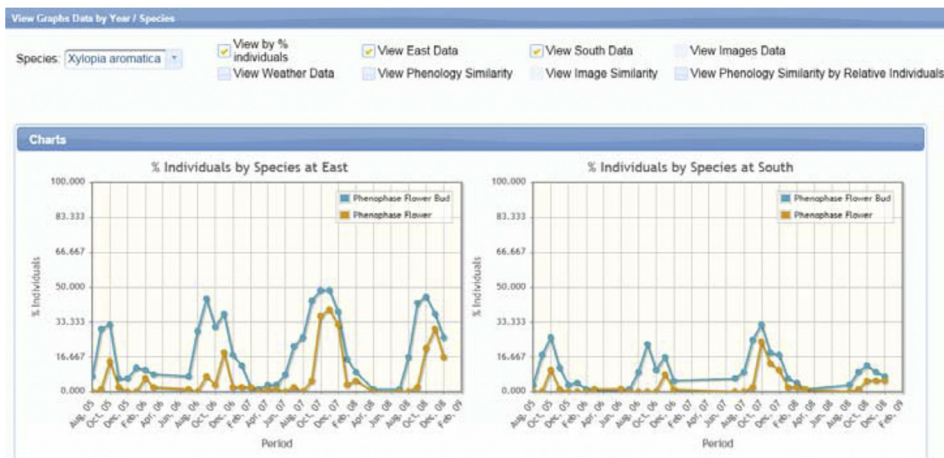
### 3.2.2. Technologies

Databases are managed, manipulated, and organized by a Database Management System (DBMS). A DBMS is a computer program responsible for managing one or more databases and handling users' requests regarding the maintenance of the data stored in these databases (Date, 2004). The DBMS chosen to implement the database for phenology studies was the PostgreSQL.

Regarding the implementation of the queries, we developed a Web application to facilitate the visualization and interaction of phenology experts with the database. We selected some queries of interest from e-phenology project and we implemented these queries using Script Query Language (SQL) and Java (application programming language). To implement the interface and display the results of queries, we used Java Server Faces (JSF), which



(a) Comparing on-the-ground phenological data with temperature and precipitation data



(b) Comparing on-the-ground phenological data between environments

**Fig. 10.** Visualizing on-the-ground phenological data. (a) The graphs show the percentage of individuals of species *Xylopia aromatica* with Flower and Flower Bud phenophases over months between 08/01/2005 and 12/31/2008 and highlights in red the available filters that allow for visualizing on-the-ground phenological data separated by environment (East and South Data) showed in (b). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

includes several interface components. SQL scripts for the phenology database creation and guidelines to install and perform the web application with the database can be found at <http://www.reco.ic.unicamp.br/ephenology/> (as of December 7, 2015) in menu item Resources. Recall that stored data can be exported to different formats (e.g., XLS, PDF, CSV, and XML) and manipulated in different tools such as information visualization systems.

### 3.3. Case studies

This section provides an overview on case studies conducted to validate the proposed database model. The case studies consider the use of a database containing about a million and a half records related to on-the-ground observations from 2004 up to December 2013, phenology series of camera images and different features of our data set. Both studies were conducted in the context of the e-phenology project at our core cerrado study area (see Section 2.2).

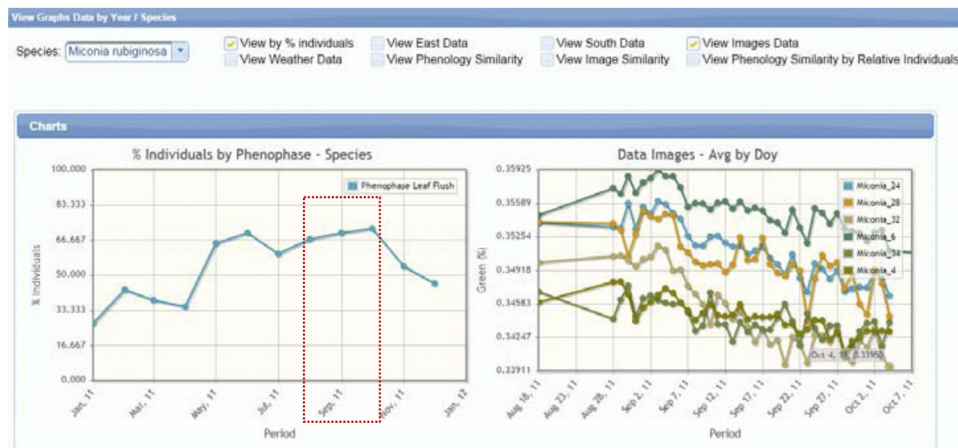
We have been developing a Web application that accesses the database, performs queries, and presents the results in user-friendly interfaces. More details regarding the presentation of

retrieved data in the case studies are described in the Electronic Supplementary Material (ESM2 and ESM3).

#### 3.3.1. Case study 1 – correlation among on-the-ground observations and climate data

Phenology studies seek to establish the correlation among on-the-ground observations and climate variables based mainly on long-term series analysis. We take as example the work by Camargo et al. (2011) and we demonstrate how the database can support and facilitate similar analysis. More specifically, we are interested here in exporting and visualizing data that allow to correlate climate data such as temperature and precipitation with the occurrence of individuals flowering in a defined period of time and in two different environments. A step-by-step description concerning the use of the tool (and performed queries) for this case study is available in the Electronic Supplementary Material ESM2.

Fig. 10(a) shows the screenshot with the percentage of individuals of species *Xylopia aromatica* with Flower and Flower Bud phenophases with the correspondent abiotic variables data for the period between 08/01/2005 and 12/31/2008 (Camargo et al., 2011). In order to compare on-the-ground data of the individuals

(a) Comparing data from species *Aspidosperma tomentosum*(b) Comparing data from species *Miconia rubiginosa*

**Fig. 11.** Comparing on-the-ground phenological data with vegetation image indices. One can see, on the left, the phenophase intensity over months for species *Aspidosperma tomentosum* in (a) and *Miconia rubiginosa* in (b), and on the right, the green scores from digital images for masks of species. The highlighted region (between August and October, 2011) on the left graph refers to the period where there are more individuals flushing. On the right, we can see the daily average green values for the same highlighted period for individuals (masks) of each species. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

situated at different environments (such as East and South), filters are available (highlighted in red) and the query result is presented in Fig. 10(b).

### 3.3.2. Case study 2 – correlation among on-the-ground observations and vegetation image indices

The objective of this case study is to demonstrate how the proposed database can be used to validate the green color channel time series with on-the-ground observations of leaf flushing of the cerrado using with example two species *Aspidosperma tomentosum* and *Miconia rubiginosa*. The first steps are described in the Electronic Supplementary Material ESM3.

The typical obtained result is illustrated in Fig. 11(a) where one can see, on the left, the amount of individuals of species *A. tomentosum* leafing per month for the year 2011 and, on the right, the daily average green scores for tree masks (e.g., *Aspidosperma\_7*, *Aspidosperma\_5*, *Aspidosperma\_20*) of the same species.

Another example is presented in Fig. 11(b), where on the left, the graph shows the amount of individuals of species *M. rubiginosa* leafing per month for the year 2011 and, on the right, the daily average green scores for six masks (e.g., *Miconia\_6*, *Miconia\_4*, *Miconia\_34*, *Miconia\_32*, *Miconia\_28*, *Miconia\_24*). Similar results have been reported in (Alberton et al., 2014).

## 4. Conclusions

Our study is the pioneer to offer a comprehensive database model to handle phenological data and associated ecological and abiotic information especially suited for tropical vegetation. The proposed database successfully integrates plant based phenological observation, ecological characteristics at individual and species level such as pollination or dispersal syndrome, flower and fruit color, leaf exchange strategy, and many other plants attributes. The database also incorporates site-base information of local environment and weather (soil type, temperature, precipitation, PAR, etc.), essential for the interpretation of phenological patterns. Investigators can easily store, organize, and manage all phenological, ecological, and environmental data derived from their research, and summarize results in a faster and more accurate way.

The two case studies presented confirm the expectations about the requirements to the database modeling process and demonstrate the wide potential to solve the e-phenology project questions and any other general query for phenology studies. In those examples the database integrated and allowed to retrieve data of different types and scales that can be read for further analyses. The validation based on our core study site, supported the application in the e-phenology project, comprising additionally four sites: two sites (Cerrado and Atlantic rainforest) with near-surface remote

phenology monitoring by digital cameras associated in complete carbon flux towers; one cerrado campo sujo site with a phenology tower; and one mountain range site with on-the-ground observations, and six phenology towers monitoring systems distributed in a altitudinal gradient (e.g., *campo rupestre*, *cerrado*, *altitudinal*, *grassland*).

The database has started with data from the e-phenology Project, but it was designed to be fed by phenology information from several study sites across the tropics. In a different way of current databases, the proposed one is dynamic, holding legacy datasets but allowing the input of new data with no storage limit and has advantages as: an easy and fast way to integrate and access phenological, climatic and ecological data; minimizing errors during data manipulation. Therefore, we proposed a powerful database model that can be widely used to manage complex temporal data sets, integrating legacy and live phenological information from diverse sources (e.g., conventional, digital cameras, seed traps) and temporal scales, improving our capability of producing scientific and applied information on tropical phenology.

Future work will consider the extension of the proposed database model to support data series derived from satellite imagery. We also plan to support series from digital images taken from Unmanned Aerial Vehicles (UAVs) in a similar way our database supports the entrance of data series derived from fixed cameras.

## Acknowledgements

This research was supported by the São Paulo Research Foundation (FAPESP), the FAPESP-Microsoft Research Virtual Institute (grants #2010/52113-5, #2011/51523-8 and #2013/50155-0). LPCM and RST receive a Productivity Research Fellowship from CNPq (grants #306243/2010-5 and #306587/2009-2). GM received a master fellowship (FAPESP grant #2011/51523-8) and BA receives a PhD fellowship (FAPESP grant #2014/00215-0). Also, we have been benefited from funds of CNPq, CAPES, and FAPESP (grants #2007/52015-0, #2007/59779-6, #2009/18438-7, and #2010/51307-0).

We thank all members of the Phenology Laboratory at UNESP for their support during several steps in the development of this database, with a special appreciation to I Mendoza and NC Soares.

## Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.ecoleng.2016.03.001>.

## References

- Alberton, B., Almeida, J., Helm, R., da, S., Torres, R., Menzel, A., Morellato, L.P.C., 2014. Using phenological cameras to track the green up in a cerrado savanna and its on-the-ground validation. *Ecol. Inform.* 19, 62–70, <http://dx.doi.org/10.1016/j.ecoinf.2013.12.011>.
- Almeida, J., Santos, J.A., Alberton, B., Morellato, L.P.C., Torres, R.S., 2013. *Plant species identification with phenological visual rhythms*. In: IEEE International Conference on eScience (eScience'13), pp. 148–154.
- Almeida, J., Santos, J.A., Alberton, B., Morellato, L.P.C., Torres, R.S., 2013. *Visual rhythm-based time series analysis for phenology studies*. In: IEEE International Conference on Image Processing (ICIP'13), pp. 4412–4416.
- Almeida, J., dos Santos, J.A., Alberton, B., da, S., Torres, R., Morellato, L.P.C., 2014. Applying machine learning based on multiscale classifiers to detect remote phenology patterns in cerrado savanna trees. *Ecol. Inform.* 23, 49–61, <http://dx.doi.org/10.1016/j.ecoinf.2013.06.011>, special Issue on Multimedia in Ecology and Environment.
- Almeida, J., Santos, J.A., Miranda, W.O., Alberton, B., Morellato, L.P.C., Torres, R.S., 2015. Deriving vegetation indices for phenology analysis using genetic programming. *Ecol. Inform.* 26, 61–69, <http://dx.doi.org/10.1016/j.ecoinf.2015.01.003>.
- Augsburger, C.K., 1983. *Phenology, flowering synchrony, and fruit set of six neotropical shrubs*. *Biotropica* 15, 257–267.
- Bencke, C.S.C., Morellato, L.P.C., 2002. Comparação de dois métodos de avaliação da fenologia de plantas, sua interpretação e representação. *Braz. J. Bot.* 25, 269–275, <http://dx.doi.org/10.1590/S0100-84042002000300003>.
- Borchert, R., 1983. *Phenology and control of flowering in tropical trees*. *Biotropica* 15, 81–89.
- Camargo, M.G.G., Cazetta, E., Schaefer, H.M., Morellato, L.P.C., 2013. Fruit color and contrast in seasonal habitats – a case study from a cerrado savanna. *Oikos* 122, 1335–1342, <http://dx.doi.org/10.1111/j.1600-0706.2013.00328.x>.
- Camargo, M.G.G., Souza, R.M., Reys, P., Morellato, L.P.C., 2011. Effects of environmental conditions associated to the cardinal orientation on the reproductive phenology of the cerrado savanna tree *Xylopia aromatica* (annonaceae). *An. Acad. Brasil. Ciênc.* 83, 1007–1020, <http://dx.doi.org/10.1590/S0001-37652011005000014>.
- Coutinho, L.M., 1978. *O conceito de cerrado*. *Braz. J. Bot.* 1, 17–23.
- Date, C.J., 2004. *An Introduction to Database Systems*, 8th ed. Pearson Education, Boston, London, Paris.
- Dierenbach, J., Badeck, F.-W., Schaber, J., 2013. *The plant phenological online database (ppodb): an online database for long-term phenological data*. *Int. J. Biometeorol.*
- Elmasri, R., Navathe, S.B., 1999. *Fundamentals of Database Systems*, 3rd ed. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
- Fournier, O., 1974. *Un metodo cuantitativo para la medición de características fenológicas en árboles*. *Turrialba* 24, 54–59.
- Frankie, G.W., Baker, H.G., Opler, P.A., 1974. Tropical plant phenology: applications for studies in community ecology. In: Lieth, H. (Ed.), *Phenology and Seasonality Modeling*. Springer, Berlin, Heidelberg, pp. 287–296, [http://dx.doi.org/10.1007/978-3-642-51863-8\\_24](http://dx.doi.org/10.1007/978-3-642-51863-8_24), Volume 8 of Ecological Studies.
- Köppen, W.P., 1931. *Grundriss der Klimakunde*. W. de Gruyter.
- Mira, C., ao, P.F., Meidanis, J., Duque-Estrada, T., Joly, C., 2013. *Tradução taxonômica: o caso do SinBiota. Relatório Técnico*. Instituto de Computação, Universidade de Campinas.
- Morellato, L.P.C., Camargo, M.G.G., Fernanda, F.D., Luiz, B.G., Mantovani, A., Hudson, I.L., 2010. *The Influence of Sampling Method, Sample Size, and Frequency of Observations on Plant Phenological Patterns and Interpretation in Tropical Forest Trees*. Springer.
- Morellato, L.P.C., Camargo, M.G.G., Gressler, E., 2013. *A Review of Plant Phenology in South and Central America*. Springer, Netherlands, pp. 91–113, [http://dx.doi.org/10.1007/978-94-007-6925-0\\_6](http://dx.doi.org/10.1007/978-94-007-6925-0_6).
- Morellato, L.P.C., Talora, D.C., Takahasi, A., Bencke, C.C., Romera, E.C., Zipparro, V.B., 2000. Phenology of Atlantic rain forest trees: a comparative study. *Biotropica* 32, 811–823, <http://dx.doi.org/10.1111/j.1744-7429.2000.tb00620.x>.
- Mueller-Dombois, D., Ellenberg, H., 2013. *Aims and Methods of Vegetation Ecology*. Geogr. Rev.
- d'Eça Neves, F.F., Morellato, L.P.C., 2004. *Methods applied for sampling and estimate tropical forest phenology*. *Acta Bot. Brasil.* 18, 99–108.
- Newstrom, L.E., Frankie, G.W., Baker, H.G., 1994. *A new classification for plant phenology based on flowering patterns in lowland tropical rain forest trees at La Selva, Costa Rica*. *Biotropica* 26, 141–159.
- Reys, P., Camargo, M.G.G., Grombone-Guaratini, M.T., de Pádua Teixeira, A., Assis, M.A., Morellato, L.P.C., 2013. *Estrutura e composição florística de um cerrado sensu stricto e sua importância para propostas de restauração ecológica*. *Hoehnea* 40, 449–464, <http://dx.doi.org/10.1590/S2236-89062013000300005>.
- Richardson, A.D., Braswell, B.H., Hollinger, D.Y., Jenkins, J.P., Ollinger, S.V., 2009. *Near-surface remote sensing of spatial and temporal variation in canopy phenology*. *Ecol. Appl.* 19, 1417–1428.
- Richardson, A.D., Jenkins, J.P., Braswell, B.H., Hollinger, D.Y., Ollinger, S.V., Smith, M.L., 2007. *Use of digital webcam images to track spring green-up in a deciduous broadleaf forest*. *Oecologia* 152, 323–334.
- Schwartz, M.D., 2013. *Phenology: An Integrative Environmental Science*. Springer, Netherlands.
- Staggemeier, V.G., Morellato, L.P.C., 2011. *Reproductive phenology of coastal plain Atlantic forest vegetation: comparisons from seashore to foothills*. *Int. J. Biometeorol.* 55, 843–854.
- Torres, R.S., Hasegawa, M., Tabbone, S., Almeida, J., Santos, J.A., Alberton, B., Morellato, L.P.C., 2013. *Shape-based time series analysis for remote phenology studies*. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS'13), pp. 3598–3601.
- Williams, R.J., Myers, B.A., Eamus, D., Duff, G.A., 1999. *Reproductive phenology of woody species in a north Australian tropical savanna*. *Biotropica* 31, 626–636.