Processing of high-resolution temporal climate data for daily simulations of a complex agro-ecosystem

Maria Catarina Paz
catarina.paz@estbarreiro.ips.pt 0000-0001-8411-0896
CIQuiBio, Barreiro School of Technology, Polytechnic Institute of Setúbal,
Rua Américo da Silva Marinho. 2839-001 Lavradio, Portugal

Sónia A.P. Santos
sonia.santos@estbarreiro.ips.pt 0000-0003-1500-6360
CIQuiBio, Barreiro School of Technology, Polytechnic Institute of Setúbal,
Rua Américo da Silva Marinho. 2839-001 Lavradio, Portugal LEAF,
Instituto Superior de Agronomia, Tapada da Ajuda. 1349-017 Lisboa, Portugal

Raquel Barreira
raquel.barreira@estbarreiro.ips.pt 0000-0002-8326-1593
INCITE, Barreiro School of Technology, Polytechnic Institute of Setúbal,
Rua Américo da Silva Marinho. 2839-001 Lavradio, Portugal
CMAFcIO – Centro de Matemática, Aplicações Fundamentais e Investigação Operacional,
Faculdade de Ciências da Universidade de Lisboa, Campo Grande. 1749-016 Lisboa, Portugal

247

Natural pest control is an ecosystem service that consists on the reduction of pests by their natural control agents, which includes predators. Computer models can be used to simulate the interaction between pests and predators, and therefore to understand how landscape, animals, and agricultural management are related. Moreover, these models allow us to predict how human activity and climate may affect the ecosystem service itself.

Complex agro-ecosystem models are being developed for the olive grove in Portugal. In particular, a model constituted by (1) the model of one of the key pests of the olive grove, the olive fly *Bactrocera oleae* (2) the model of a ground spider, *Haplodrassus rufipes*, a generalist predator, and (3) the model of a landscape, in this case representing a section located in Mirandela, NE Portugal, a region mainly devoted to olive production, where traditional farming practices are still predominant. These models are simulated and articulated using the Animal, Landscape and Man Simulation System (ALMaSS) through climate data series and farm management events daily inputs. Whereas farm management events have a non-regular distribution in time, climate data series must have a regular distribution in time in order for this system to run. However, many times climate data series are not complete, mostly because of malfunctions of the measuring instruments, occasional interruptions of automatic stations and/or reorganizations of network, which demands

reliable methodologies for filling the gaps. A common way to do this is to replace missing data with reasonable values, a process that, in statistics, is called imputation. On the other hand, the time step of a measured climate data series may be different from the one needed for the system to run, which implies the conversion to the desired time step. In the case of the conversion from hourly data series to daily data series, there is a reduction of temporal resolution and, therefore, a loss of information in what concerns specific periods of the day. This can be a problem when trying to simulate, for instance, an animal behaviour that occurs during a specific period of the day. One way to segregate that specific period would be to create additional daily variables of interest that are calculated using only the hourly data comprehended in that period.

The objectives of this work are (1) to process climate data series using the open source programming language R (R core team, 2020), for further application in studies comprising simulations of *H. rufipes* and *B. oleae* in the landscape of Mirandela, (2) to qualitatively verify if the completed data are acceptable to feed to the system, and (3) to demonstrate how to create daily variables that represent specific periods of the day.

The time-step used for ALMaSS simulations of *H. rufipes* and *B. oleae* in the landscape of Mirandela is 1 day. In this case, the system needs the following daily climate variables inputs: precipitation (mm), wind speed (m/s), air temperature (°C), soil temperature (°C), relative humidity (%), and soil temperature during twilight (°C). This means that for each day of the simulation, ALMaSS requires one input value for each of these climate variables or, in other words, univariate (one attribute observed over time), evenly spaced (equal increments between successive data points), numeric (measurable quantities that are expressed as a number) time series for each variable. Because climate data series are therefore time series, they can be filled, when incomplete, through imputation, which involves inter-time correlations, in the case of univariate time series.

Climate data series referring to precipitation, wind speed, air and soil temperature, and relative humidity were measured hourly, from 2010 to 2020, at the Mirandela climate station, at an altitude of 250 m. This station is part of the network of climate stations of Instituto do Mar e da Atmosfera (IPMA), the Portuguese state agency for climate. Air temperature was measured at 1.5 m above the soil, and soil temperature was measured at a depth of 0.05 m.

The raw data contained time series, for each variable, of length 96360. However, a considerable amount of missing values was present for each measured variable. Processing of data to create complete daily data series was performed using R language, and followed the sequence displayed in figure 1. Each variable was treated separately. Most variables, such as air temperature and soil temperature (and even relative humidity and wind speed, although not so strongly), present daily and yearly seasonality. Precipitation does not fall in this category so it was treated using a slightly different approach. For completing the task of imputation, we have used some of the functions of the R package imputeTS, a package built specifically for univariate time series imputation of missing values. This package uses a process for imputation that starts by analysing the distribution of the missing values, then performs the imputation based on one of the available algorithms, and finally visualizes the imputations in the time series.

To infer if the data series obtained with our methodology were acceptable for using in the ALMaSS simulation, we calculated the monthly inter-annual means for the period 2010–2020, using the same method as the one used for calculating the climate normals, and compared the obtained annual pattern with that of the climate normals for the period 1971–2000 registered at the Mirandela climate station.

The applied methodology allowed to produce well blended imputations, creating acceptable completed hourly time series. The daily climate series also show acceptable quality, which could be concluded by (1) the comparison between monthly inter-annual means calculated for the period 2010–2020 and the climate normals for the period 1971–2000 registered at the Mirandela climate station, and (2) the change of patterns from the period 1971–2000, to the period 2010–2020, which is in accord with the global and regional climate projections for the Portuguese territory, which predict an increase in air temperature, and in precipitation during a shorter season, and are expected to have more impact in summer and autumn.

Our methodology also allowed to segregate the temperature during twilight as a daily measure. We verified that mean soil temperature during twilight is in general higher than daily mean soil temperature. The use of the more general soil temperature variable instead of the specific soil temperature during twilight

variable, would lead to incorrect results from the simulation of the spider movement, which is known to happen during the twilight period in real life". This way, it was possible to modulate an aspect that otherwise would not have impact on the simulations, meaning that we can detect and filter important attributes when using high-resolution temporal climate data, as suggested by Afrifa-Yamoah *et al.* (2020). Therefore, it was possible to reduce the error associated with the use of a daily climate value to express functions that occur during specific periods of the day, but respecting the time step of the system. These results may also give some more information to understand why the spider prefers this specific period of the day to hunt and move.

The specific methodology proposed in this paper will be further developed to process climate data for ALMaSS simulations of other pests and predators of the olive grove in the Mirandela study site. As ALMaSS is being implemented in other study sites including olive groves across Europe, the proposed methodology can also be used wherever there is available hourly climate data. We also intend to test our methodology with a more formal approach in the future.

With this work, we expect to have contributed with an expedite methodology and ideas for climate data series processing and further use in agro-ecosystem modelling. This way, we also wish to contribute to a better understanding of the mechanisms behind the functioning of ecosystems services, in particular the functioning of natural pest control, which is an important alternative to pesticide application as one of the answers to the great challenge of contemporary agriculture: to produce healthy and enough food, maintaining environmental quality and social dignity, conserving biodiversity, consciously managing finite resources, while adapting to climate change.
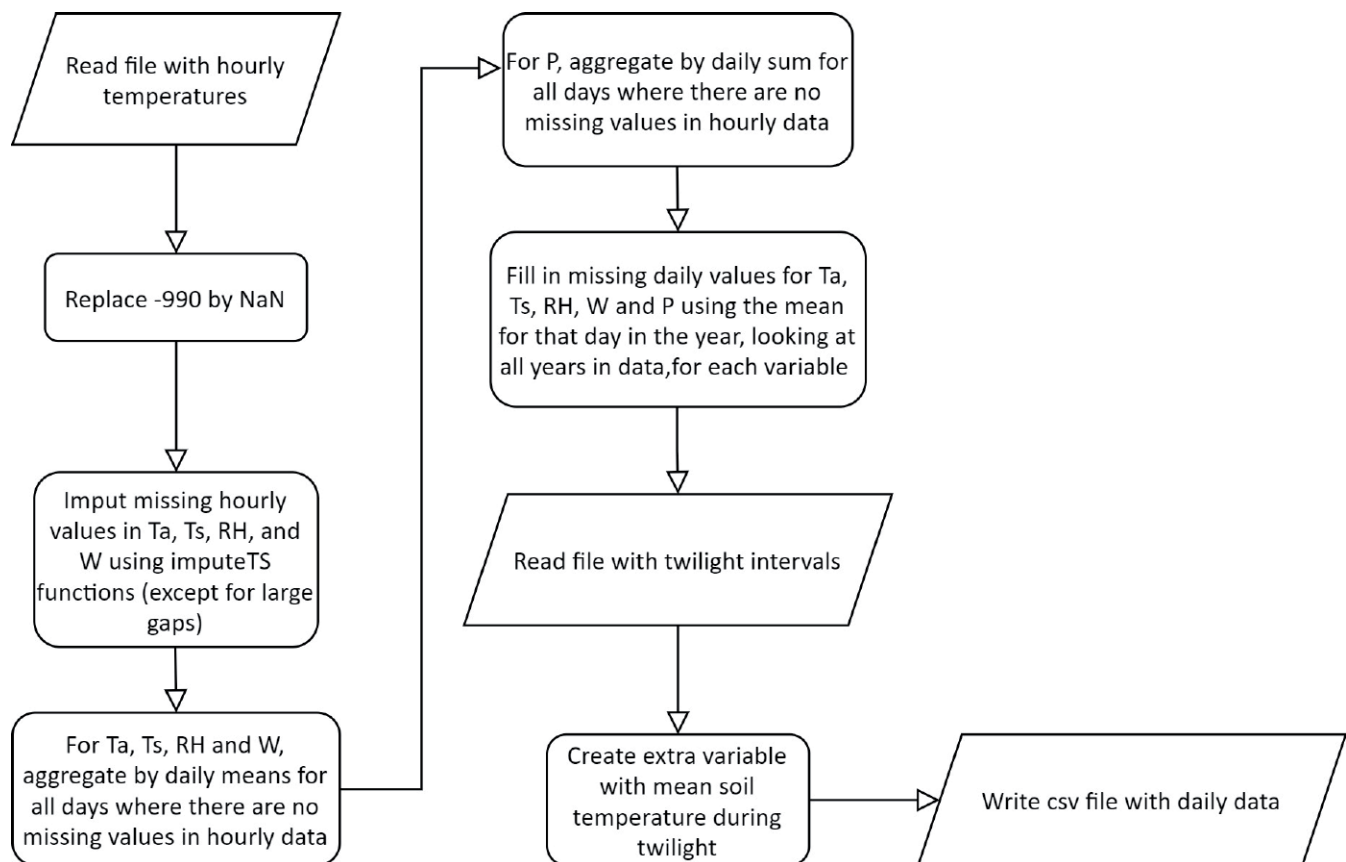
Figure 1. Sequence of the processing of climate data using R language.