



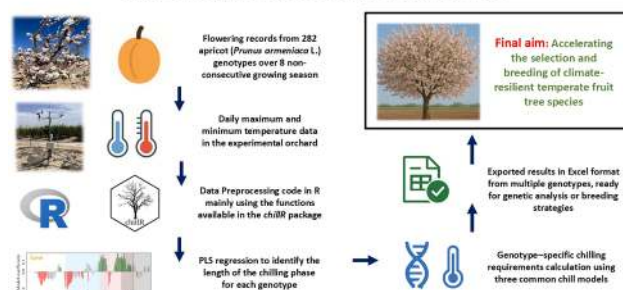
# R-based workflow to estimate chilling requirements in multiple fruit tree genotypes using Partial Least Squares regression: *Prunus armeniaca* L. case<sup>☆</sup>

Ana M. Muñoz-Morales, Germán Ortuño-Hernández, Juan Alfonso Salazar, Pedro Martínez-Gómez, Jose A. Egea, David Ruiz, Alvaro Delgado<sup>\*</sup>

Department of Plant Breeding, CEBAS-CSIC, P.O. Box 164, E-30100, Espinardo, Murcia, Spain

## GRAPHICAL ABSTRACT

R-based workflow to estimate chilling requirements in multiple fruit tree genotypes using Partial Least Squares regression: *Prunus armeniaca* L. case



## ARTICLE INFO

**Keywords:**  
Accelerated breeding  
Chill models  
Dormancy  
Progeny evaluation  
PLS  
Temperate fruit trees

## ABSTRACT

Accurate estimation of chilling requirements (CR) is essential for breeding and selecting temperate fruit trees adapted to specific agroclimatic conditions, particularly under global warming scenarios. Among the available methodologies to determine CR, the Partial Least Squares (PLS) regression procedure, based on long-term phenological and temperature records, offers a suitable approach to delineate the effective chill accumulation period. In this study, we present an R-based workflow developed using the agroclimatic functions integrated into the *chillR* package for R to determine the genotype-specific CR of 282 apricot (*Prunus armeniaca* L.)

### <sup>☆</sup> Related research article

None.

<sup>\*</sup> Corresponding author.

E-mail address: [adelgado@cebas.csic.es](mailto:adelgado@cebas.csic.es) (A. Delgado).

<https://doi.org/10.1016/j.mex.2025.103686>

Received 2 August 2025; Accepted 22 October 2025

Available online 26 October 2025

2215-0161/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

seedlings from two progenies grown in southwestern Spain. The pipeline generates standardized CR datasets suitable for downstream applications, including QTL mapping and the selection of promising genotypes for breeding purposes. This tool streamlines the estimation process, reducing the technical expertise and time required for CR estimation, thereby supporting efficient phenotypic selection and accelerating genetic research in temperate fruit trees. The complete code and associated datasets are freely available in a public repository (<https://github.com/CEBASFruitBreed/R-workflow-ChillPLS>), promoting the use across a range of temperate fruit species.

- Uses long-term flowering observations and temperature records to determine genotype-specific chilling requirements.
- Integrates PLS regression procedure within an R-based workflow to estimate chilling requirements from datasets comprising multiple genotypes.
- Generates standardized outputs suitable to support genetic analysis and informed breeding decisions.

| Specifications table                  |                                                                                                                             |
|---------------------------------------|-----------------------------------------------------------------------------------------------------------------------------|
| Subject area                          | Chilling requirement estimation in temperate fruit trees for large datasets                                                 |
| More specific subject area            | ChillPro: Evaluation of chilling requirements in large fruit trees progenies                                                |
| Name of your method                   | N/A                                                                                                                         |
| Name and reference of original method | Related R code is available in:                                                                                             |
| Resource availability                 | <a href="https://github.com/CEBASFruitBreed/R-workflow-ChillPLS">https://github.com/CEBASFruitBreed/R-workflow-ChillPLS</a> |

Background

Fruit breeders routinely generate large segregating progenies each year through controlled cross-pollinations. These progenies are monitored over multiple seasons for key phenotypic traits such as flowering time, to identify promising individuals for commercial release as new cultivars or use as potential parents in further breeding cycles. Additionally, biparental populations are evaluated to construct genetic maps and identify loci associated with important agronomic traits [1]. For temperate fruit trees, one of these traits is chilling requirements, especially important for the selection of new varieties in a context of climate change. The trees spend the winter months in a dormancy phase that allows them to survive unfavorable climatic conditions [2,3]. To break dormancy and ensure a good and uniform flowering in spring, each cultivar must meet its genotype-specific chilling requirement (CR). Consequently, the careful selection of cultivars well adapted to specific growing regions where it can fulfill its CR every year has become increasingly important for sustaining the profitability of fruit production [4,5]. Traditionally, CR in *Rosaceae* species such as apple, pear, almond, peach, sweet cherry, and apricot has been estimated using a labor-intensive and time-consuming forcing method [6,7]. This approach involves collecting shoots weekly and exposing them to controlled environmental conditions for a fixed period of time, limiting its use to a relatively small number of genotypes each year. As an alternative, statistical methods based on using long-term field observations of flowering dates and temperature records offer promising outcomes. Among them, Partial Least Squares (PLS) regression has been widely used to define the chilling period and estimate genotype-specific chilling requirements [8–10]. Although PLS regression can produce robust results depending on the species, location, and availability of phenological data [6], applying PLS regression to large datasets remains time-consuming since two separate steps are needed. First, determining genotype-specific chilling periods and second calculating CR values within those periods for numerous individuals. To address this challenge, an R script, using functions from the *chillR* package [11], and its corresponding workflow have been developed to automate this process. This code enables an efficient estimation of CR across large datasets of flowering records collected by breeders or researchers over recent decades. The final output is produced in a format ready for downstream molecular analyses, such as QTL mapping. This automated pipeline significantly accelerates the generation of CR data in breeding programs, supporting both phenotypic selection and genetic research. In this paper, we use as a case study the flowering records collected by the CEBAS-CSIC Fruit Breeding Team in Cieza (Murcia, southwestern Spain; 38° 16' N, 1° 16' W, 241 m a.s.l.) over eight non-consecutive growing seasons for two F1 apricot (*Prunus armeniaca* L.) progenies. These progenies resulted from the crosses ‘Bergeron’ × ‘Currot’ (‘B × C’) and ‘Goldrich’ × ‘Currot’ (‘G × C’), comprising 126 and 156 seedlings, respectively.

Method details

Following this introduction, we present the code and workflow used in our case study. The code, written in the R programming

**Table 1**

Example of the Excel file structure used for flowering date input in the 'G × C' population.

| Genotype | BD12 | BD13 | BD14 | BD21 | BD22 | BD23 | BD24 | BD25 |
|----------|------|------|------|------|------|------|------|------|
| 1_1      | *    | 75   | 69   | 71   | 81   | 83   | 80   | 85   |
| 1_2      | 70   | 53   | 53   | 60   | 63   | 72   | 59   | 69   |
| 1_3      | 72   | 63   | 62   | 64   | 72   | 80   | 75   | 76   |
| 1_4      | 72   | 63   | 63   | 64   | 67   | 72   | 68   | 75   |

language, outlines the essential steps required to generate the outputs and is accompanied by a brief explanation of the results and their implications. All dormancy-related metrics, including the PLS regression analysis and the calculation of CR using the three most common chill models, are implemented in the *chillR* package (version 0.76) [11]. The complete code and dataset used in the analyses described below are freely available in the following GitHub repository (<https://github.com/CEBASFruitBreed/R-workflow-ChillPLS>).

### R libraries

To run the functions used in this workflow, the following R libraries must be loaded prior to execute the script: *'tidyr'*, *'openxlsx'*, *'readxl'*, *'dplyr'*, *'chillR'*, *'stringr'*, and *'tools'*.

### Data preprocessing workflow

To conduct the analysis, flowering data, defined as full bloom (~50 % of flowers open) [12], must be expressed in Julian dates (day of the year; DOY) and organized in an Excel file. Each genotype should have a single DOY value per growing season (see Table 1). The dataset must be structured with columns arranged from left to right, starting with "Genotype," followed by columns labeled 'BD12' to 'BD25,' each representing a specific year in the 2000s for which flowering data were available in this case study. Each row corresponds to a unique genotype, identified using coded names.

The code below illustrates how to import and preprocess two datasets containing blooming dates, each corresponding to controlled hybridization crosses of apricot cultivars ('BxC' and 'GxC'). The data are read from an Excel file using the *read\_Excel* function, with each bi-parental population stored in a separate worksheet. To ensure data quality, any records marked with an asterisk (\*), which indicates very poor flowering and the inability to assign a reliable date under field conditions are treated as missing values and removed using the *na.omit* function, which eliminates rows with incomplete or invalid entries. While missing data can be handled according to user preference and potentially removed before starting the analysis, in this code the symbol "\*" is specifically used to exclude such observations. The same procedure would apply analogously for the GxC population.

```
Phenotype_Phenology_BxC_Definitive <- read_Excel("Phenotype_(Phenology)BxC_GxC_Definitive.xlsx",
  sheet = "BxC Phenotype Phenology", col_types = c("text", "numeric", "numeric", "numeric", "numeric",
  "numeric", "numeric", "numeric", "numeric"))
```

```
# Delete missing data; delete rows with * symbol
```

```
Phenotype_Phenology_BxC_filt<- na.omit(Phenotype_Phenology_BxC_Definitive)
```

The next step consists of reformatting the previously loaded dataset to match the structure used in the CEBAS-CSIC flowering records. This step can be skipped if the phenology dataset has already been prepared following the format shown in Table 2. In our example, we reformatted the dataset using the *process\_data\_format* function, specifically designed to prepare and filter the data for downstream analyses. This function uses the *pivot\_longer* function from the *tidyverse* package to convert the dataset from wide to long format, collapsing all columns starting with "BD" (bloom dates) into two new columns: "Year", which transforms the original column names by replacing the "BD" prefix with "20" to generate a four-digit year (e.g., "BD12" becomes "2012"), and "pheno", which contains the flowering date values for each year. Finally, the reformatted dataset is saved as an Excel file (Table 2) for subsequent analyses.

```
process_format_data<-function(data,new_name){
  # Transform the Excel format
  data_filt <- pivot_longer(data, cols = starts_with("BD"), names_to = "Year", values_to = "pheno")
  # Convert the year from alphanumeric format to numeric type.
  data_filt$Year <- as.numeric(sub("BD", "20", data_filt$Year))
  # Write an Excel file
  write.xlsx(data_filt,paste0(new_name,"_edit_filt.xlsx"))
  # The previously defined function is applied
  process_format_data(Phenotype_Phenology_BxC_filt,"BxC")
  process_format_data(Phenotype_Phenology_GxC_filt,"GxC")
}
```

The *load\_climatic\_file* function is designed to process temperature data for estimating cultivar-specific chilling requirements. The workflow begins by evaluating data completeness using the *check\_temperature\_record* function, which identifies any missing values

**Table 2**

Flowering records of the ‘BxC’ population formatted in Excel for analysis using the *chillR* package.

| Genotype | Year | pheno |
|----------|------|-------|
| 1_1      | 2012 | 74    |
| 1_1      | 2013 | 69    |
| 1_1      | 2014 | 62    |
| 1_1      | 2021 | 60    |
| 1_1      | 2022 | 67    |
| 1_1      | 2023 | 79    |
| 1_1      | 2024 | 75    |
| 1_1      | 2025 | 83    |
| 1_4      | 2012 | 78    |
| 1_4      | 2013 | 65    |
| 1_4      | 2014 | 63    |
| 1_4      | 2021 | 64    |
| 1_4      | 2022 | 65    |
| 1_4      | 2023 | 83    |
| 1_4      | 2024 | 71    |
| 1_4      | 2025 | 74    |

**Table 3**

Structure of the Excel file used for daily temperature data with columns representing the main variables used in the analysis: Day (day of the month), JDay (Julian day, indicating the day of the year), Month (calendar month), Year (calendar year), Tmax (daily maximum temperature in °C), and Tmin (daily minimum temperature in °C).

| Day | JDay | Month | Year | Tmax  | Tmin |
|-----|------|-------|------|-------|------|
| 1   | 1    | 1     | 2011 | 14.46 | 6.62 |
| 2   | 2    | 1     | 2011 | 15.72 | 6.76 |
| 3   | 3    | 1     | 2011 | 15.32 | 3.36 |
| 4   | 4    | 1     | 2011 | 15.17 | 0.42 |
| 5   | 5    | 1     | 2011 | 14.01 | 2.80 |
| 6   | 6    | 1     | 2011 | 20.34 | 3.85 |

(NAs). Temperature records are then filtered starting from one year prior to the earliest available phenological observation. For this period, temperature values are corrected using the *fix\_weather* function. The dataset is subsequently transformed into hourly temperatures using *stack\_hourly\_temps*. Based on the latitude of the experimental orchard, hourly values are reconstructed from daily minimum and maximum temperatures following the methods of [13] and [14]. Daily chill accumulation is then calculated using the *daily\_chill* function, which returns a list containing validation results, the number of missing values, the corrected temperature dataset, and the computed daily chill. This workflow is demonstrated using the temperature dataset “Cieza11–25.xlsx” (Table 3) which includes daily maximum and minimum temperature at latitude of 38.11°.

```
load_climatic_file<-function(dataset,Year,latitude){
  #The data type is checked
  dataset<-as.data.frame(dataset)
  dataset$Year<-as.integer(dataset$Year)
  dataset$Month<-as.integer(dataset$Month)
  dataset$Day<-as.integer(dataset$Day)
  dataset$Tmax<-as.numeric(dataset$Tmax)
  dataset$Tmin<-as.numeric(dataset$Tmin)
  ## Check the temperature Log
  check_temperatures<-check_temperature_record(dataset, hourly = FALSE, completeness_check = TRUE, no_variable_check = FALSE)
  #Calculate the number of missing values in the data set.
  result_na<-sum(is.na(dataset))
  #Correct the meteorological data, considering only the years greater than the one specified.
  weather<-fix_weather(dataset[which(dataset$Year>Year),])
  #Calculate the daily chill using the corrected data and the Latitude
  dc<-daily_chill(stack_hourly_temps(weather,latitude), 11)
  #Return a List with the results of checks, corrections and calculations.
  return(list(check_temperatures = check_temperatures,
    result_na = result_na,
    weather = weather,
    dc = dc))}
Cieza11_25_2 <- load_climatic_file(read_Excel("C:/working_path/Cieza11-25.xlsx"),2010,38.11)
weather<-Cieza11_25_2$weather
dc<-Cieza11_25_2$dc
```

The *process\_genotypes* function is designed to organize genotypic data into subsets and export each as a separate Excel file. From the Genotype column, two numerical groups are generated: group1, which corresponds to the number before the underscore symbol (), and group2, which identifies the number after the underscore. These identifiers are used to group the data with *group\_by*, and the records are then split into distinct subsets using *group\_split*. Each subset is saved as an individual Excel file, with filenames that include the provided identifier (id) along with the corresponding group1 and group2 values. This function simplifies data organization, facilitating downstream management and analysis. In this example, it was applied to the files BxC\_edit\_filt.xlsx and GxC\_edit\_filt.xlsx, resulting in multiple output files grouped by genotype combinations.

```
process_genotypes<-function(data,id){data %>%
# Divide the data into a list of subsets, one for each group.
mutate(group1 = as.numeric(sub("^(\\d+)_.*", "\\1", Genotype)),
group2 = as.numeric(sub("^\\d+_((\\d+).*)", "\\1", Genotype))) %>%
group_by(group1, group2) %>%
group_split() %>%
# Iterate over each group. For each .x subset, write an Excel file using write.xlsx.
purrr::walk(~write.xlsx(.x, paste0("Genotypes/",id,"_group_", unique(.x$group1), "_", unique(.x$group2)
, ".xlsx")))}
process_genotypes(read_Excel("BxC_edit_filt.xlsx"),"BxC")
process_genotypes(read_Excel("GxC_edit_filt.xlsx"),"GxC")
```

### PLS-based workflow for chilling requirement analysis

This section describes the automated processing of multiple Excel files containing flowering data to perform a PLS regression analysis. It begins by generating a list of files located in the "Genotypes/" directory that match a specific naming pattern. Each file is

|    | A           | B     | C              | D                | E              | F          | G          |
|----|-------------|-------|----------------|------------------|----------------|------------|------------|
| 1  | 1101        | Chill | -60            | 0,02941036       | 0,56729587     | 0,13605131 | 0,14917309 |
| 2  | 1102        | Chill | -59            | 0,03724155       | 0,50799305     | 0,14508694 | 0,16194175 |
| 3  | 1103        | Chill | -58            | 0,06196283       | 0,69066165     | 0,1653623  | 0,16002992 |
| 4  | 1104        | Chill | -57            | 0,05958861       | 0,66817957     | 0,16789043 | 0,16195138 |
| 5  | 1105        | Chill | -56            | 0,06226854       | 0,68101974     | 0,19300367 | 0,17827149 |
| 6  | 1106        | Chill | -55            | 0,06818096       | 0,7116051      | 0,18148601 | 0,18891662 |
| 7  | 1107        | Chill | -54            | 0,03308388       | 0,46656912     | 0,20769347 | 0,22098192 |
| 8  | 1108        | Chill | -53            | 0,03070482       | 0,43910415     | 0,19520572 | 0,20031521 |
| 9  | 1109        | Chill | -52            | 0,03316263       | 0,46745559     | 0,20820691 | 0,22189575 |
| 10 | 1110        | Chill | -51            | 0,03643605       | 0,53954832     | 0,20448682 | 0,20424094 |
| 11 | 1111        | Chill | -50            | 0,03354529       | 0,63952223     | 0,20716832 | 0,18613504 |
| 12 | 1112        | Chill | -49            | 0,03231541       | 0,72184609     | 0,22234701 | 0,16901895 |
| 13 | 1113        | Chill | -48            | 0,02532512       | 0,69829953     | 0,21839488 | 0,17547762 |
| 14 | 1114        | Chill | -47            | 0,0224821        | 0,51903936     | 0,22418573 | 0,16957337 |
| 15 | 1115        | Chill | -46            | 0,00383088       | 0,5677159      | 0,24760444 | 0,18252669 |
| 16 | 1116        | Chill | -45            | -0,02686752      | 0,71068362     | 0,25151003 | 0,14755812 |
| 17 | 1117        | Chill | -44            | -0,04768672      | 0,8879105      | 0,27649009 | 0,13330335 |
| 18 | 1118        | Chill | -43            | -0,02484827      | 0,49041096     | 0,31243466 | 0,16018448 |
| 19 | 1119        | Chill | -42            | -0,03319443      | 0,70185452     | 0,34452215 | 0,14661818 |
| 20 | 1120        | Chill | -41            | -0,0452012       | 0,80653141     | 0,3724142  | 0,17693842 |
| 21 | 1121        | Chill | -40            | -0,0471702       | 0,80933562     | 0,43814664 | 0,18664352 |
| 22 | 1122        | Chill | -39            | -0,02368821      | 0,45485502     | 0,4640653  | 0,14278493 |
| 23 | 1123        | Chill | -38            | -0,05718283      | 0,83775172     | 0,48489285 | 0,15319746 |
| 24 | 1124        | Chill | -37            | -0,05855717      | 0,80840668     | 0,53220004 | 0,16329455 |
| 25 | 1125        | Chill | -36            | -0,07832798      | 1,04938622     | 0,5627646  | 0,1756255  |
| 26 | 1126        | Chill | -35            | -0,06963138      | 0,84306398     | 0,56423173 | 0,16755309 |
| 27 | 1127        | Chill | -34            | -0,06377795      | 0,95268062     | 0,59170344 | 0,16297889 |
| 28 | 1128        | Chill | -33            | -0,04725734      | 0,75145192     | 0,63980833 | 0,16200456 |
| 29 | 1129        | Chill | -32            | -0,05718922      | 0,80549189     | 0,61078262 | 0,14357916 |
|    | object_type | pheno | Chilling_Hours | Utah_Chill_Units | Chill_Portions |            |            |

**Fig. 1.** Results of the chilling phase obtained by applying PLS regression. The third column indicates the day of the year (DOY) within the growing season, with negative values representing the number of days remaining until the end of the year. Columns 4 and 5 show the Variable Importance in the Projection (VIP) scores and the corresponding model coefficients, respectively. Columns 6 and 7 represent the daily mean and standard deviation of the selected chilling metric (Chill Portions in our study case).

read iteratively, extracting its content and base name while removing unnecessary columns (group1 and group2) and adjusting key data types (e.g., converting Year to integer and pheno to factor). The *PLS\_chill\_force* function is then applied to each dataset, using preprocessed climate data and setting the start of the dormant season to November (split\_month = 11). This starting month was selected based on previous regional studies showing that chill accumulation typically begins in November [15,16]. Across three consecutive seasons in these studies, the first chill portion was observed between November 4th and November 8th. The output of each analysis is saved as a new Excel file (see Fig. 1) in the “Genotypes/Results\_PLS\_Genotypes/” directory, with filenames corresponding to their source files to ensure organized tracking and easy retrieval. For each genotype, the function extracts group identifiers from the Cultivar column, splits the dataset into subsets based on unique group combinations, and exports each subset as a separate Excel file. This approach ensures that subsets are well-organized and readily accessible for further analysis. Chilling accumulation is calculated using three widely applied models in horticultural research: the Chilling Hours Model, the Utah Model, and the Dynamic Model. We strongly recommend using the Dynamic Model [17,18], as it has been shown to produce more reliable results in regions characterized by mild winter conditions [19,20].

```
files_genotypes<-list.files("Genotypes/",pattern = "\\..xlsx",full.names = TRUE)

for (i in files_genotypes){
  # Read the Excel file
  data<-read_Excel(i)
  name_data<-tools::file_path_sans_ext(basename(i))
  # Check the data type
  data$group1=NULL
  data$group2=NULL
  data<-as.data.frame(data)
  data$Year<-as.integer(data$Year)
  data$pheno<-as.factor(data$pheno)
  # Application of the PLS_chill_force function
  result <- PLS_chill_force(daily_chill_obj = dc, bio_data_frame = data, split_month = 10)
  #Save the results
  write.xlsx(result,paste0("Genotypes/Results_PLS_Genotypes/",name_data,".xlsx"))}
```

The *process\_PLS\_file* function is designed to handle multiple datasets containing daily chill values (e.g., Chill Portions) and extract key information on the chilling period associated with each genotype’s entrance and release from endodormancy. This function applies to the PLS regression approach, which determines the chill accumulation window based on two primary outputs: the Variable Importance in Projection (VIP) scores and the model coefficients.

Following the criteria established by [21], the endodormancy phase is identified by a series of consecutive days with negative model coefficients and VIP scores equal to or greater than 0.8. For each dataset, the genotype identifier is retrieved, and rows are filtered to retain only those that meet both conditions. The function then defines the chilling period for each genotype by identifying the first and last day within this filtered range. Because the PLS approach accounts for the overlap between chilling and heat accumulation phases, and the current consensus among dormancy-related researchers [22], the code restricts chill accumulation to end maximum one day before the genotype’s median bloom date. The resulting data is then compiled into a single data frame and exported

```

process_PLS_file <- function(data_PLS, data_pheno, name_file) {
  # Create an empty dataframe to store the results
  resultados_PLS <- data.frame(Genotypes = character(), first_day = numeric(), end_day = numeric(), stringsAsFactors = FALSE)
  # Iterate over each PLS data file
  for (file in data_PLS) {
    # Read the Excel file, specifying the chill portions sheet ("Chill_Portions")
    data <- read_Excel(file, col_names = FALSE, sheet = "Chill_Portions")
    # Extract the genotype identifier from the file name
    id_gen <- file_path_sans_ext(basename(file))
    pattern <- "_(\\d+(?:_\\d+)*)$"
    result_id_gen <- regmatches(id_gen, regexpr(pattern, id_gen))
    result_id_gen <- sub(" ", "", result_id_gen)
    # Rename the columns of the read data
    colnames(data) <- paste0("X", seq_len(ncol(data)))
    # Filter the dataset to retain rows corresponding to the type "Chill"
    data_filt <- data %>% filter(X2 == "Chill")
    # Identify values where both specific criteria are met
    indices <- which(data_filt$X4 < 0 & data_filt$X5 >= 0.8)
    # Check whether at least two rows meet these criteria
    if (length(indices) >= 2) {
      # Filter the corresponding data range
      data_filt_2 <- data_filt[indices[1]:indices[length(indices)], ]
      # Filter and calculate the average of the phenotypic value associated with the genotype
      data_pheno_filt <- data_pheno %>%
        filter(Genotype == result_id_gen) %>%
        summarise(pheno = mean(pheno, na.rm = TRUE))
      pheno_value <- data_pheno_filt$pheno
      # Validate that `pheno_value` is not null
      if (length(pheno_value) == 0 || is.na(pheno_value)) {
        print(paste("Warning: No `pheno` found for genotype", result_id_gen))
        next
      }
      # Get the first day of the filtered range
      first_day <- data_filt_2$X3[1]
      # Determine the last day within the chill accumulation window as one day before the genotype's median bloom date, consistent with the assumption that chill and heat accumulation phases may overlap
      possible_last_days <- data_filt_2$X3[data_filt_2$X3 < pheno_value]
      end_day <- if (length(possible_last_days) > 0) max(possible_last_days) else NA
      # Add results to the global dataframe
      resultados_PLS <- rbind(resultados_PLS, data.frame(
        Genotypes = result_id_gen,
        first_day = first_day,
        end_day = end_day,
        stringsAsFactors = FALSE
      ))
    } else {
      # Warning message if there are not enough rows that meet the conditions
      cat(paste("The file", file, "does not have enough rows that meet the conditions.\n"))
    }
  }
  # Save the results to an Excel file
  write.xlsx(resultados_PLS, paste0(name_file))
}

```

A)

|    | A     | B                 | C                 | D                 | E                 | F                 | G                 | H                 | I                 | J                   | K                   | L                   | M                   |
|----|-------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|---------------------|---------------------|---------------------|---------------------|
| 1  | Trait | Chilling_Hours_12 | Chilling_Hours_13 | Chilling_Hours_14 | Chilling_Hours_21 | Chilling_Hours_22 | Chilling_Hours_23 | Chilling_Hours_24 | Chilling_Hours_25 | Utah_Chill_Units_12 | Utah_Chill_Units_13 | Utah_Chill_Units_14 | Utah_Chill_Units_21 |
| 2  | 1_1   | 1020              | 848               | 918               | 853               | 837               | 855               | 674               | 662               | 1416                | 1325                | 1366                | 1077                |
| 3  | 1_11  | 814               | 635               | 571               | 616               | 618               | 746               | 522               | 561               | 1048                | 1002                | 932                 | 776                 |
| 4  | 1_12  | 987               | 856               | 909               | 849               | 918               | 872               | 648               | 656               | 1380                | 1387                | 1285                | 999                 |
| 5  | 1_13  | 1028              | 848               | 934               | 853               | 849               | 855               | 674               | 662               | 1427                | 1332                | 1382                | 1073                |
| 6  | 1_14  | 1061              | 879               | 979               | 881               | 866               | 860               | 705               | 689               | 1446                | 1438                | 1424                | 1085                |
| 7  | 1_17  | 975               | 855               | 897               | 849               | 912               | 864               | 632               | 656               | 1357                | 1376                | 1271                | 984                 |
| 8  | 1_18  | 990               | 842               | 889               | 835               | 831               | 855               | 644               | 662               | 1382                | 1297                | 1331                | 1065                |
| 9  | 1_4   | 1026              | 862               | 953               | 872               | 941               | 872               | 680               | 662               | 1427                | 1393                | 1319                | 1031                |
| 10 | 1_7   | 1045              | 866               | 962               | 864               | 858               | 855               | 691               | 670               | 1437                | 1410                | 1428                | 1071                |
| 11 | 1_8   | 1017              | 848               | 925               | 852               | 849               | 855               | 664               | 662               | 1421                | 1311                | 1363                | 1072                |
| 12 | 1_9   | 1045              | 866               | 949               | 864               | 849               | 855               | 691               | 670               | 1437                | 1361                | 1410                | 1084                |

B)

|                     | N                   | O                   | P                   | Q                 | R                 | S                 | T                 | U                 | V                 | W                 | X                 | Y |
|---------------------|---------------------|---------------------|---------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|---|
| Utah_Chill_Units_22 | Utah_Chill_Units_23 | Utah_Chill_Units_24 | Utah_Chill_Units_25 | Chill_Portions_12 | Chill_Portions_13 | Chill_Portions_14 | Chill_Portions_21 | Chill_Portions_22 | Chill_Portions_23 | Chill_Portions_24 | Chill_Portions_25 |   |
| 1189                | 1065                | 846                 | 901                 | 73.5              | 72.1              | 70.8              | 60.9              | 66.3              | 62.1              | 49.9              | 56.3              |   |
| 815                 | 836                 | 671                 | 819                 | 54.0              | 53.2              | 49.2              | 44.5              | 46.3              | 46.4              | 40.3              | 48.8              |   |
| 1248                | 1053                | 739                 | 776                 | 74.5              | 78.1              | 68.7              | 60.3              | 68.8              | 62.1              | 49.4              | 54.3              |   |
| 1200                | 1056                | 843                 | 899                 | 74.4              | 72.1              | 71.8              | 60.9              | 66.7              | 62.1              | 49.9              | 56.3              |   |
| 1216                | 1021                | 829                 | 950                 | 74.9              | 79.8              | 74.3              | 63.3              | 67.8              | 62.1              | 51.9              | 60.4              |   |
| 1211                | 1042                | 722                 | 763                 | 73.6              | 78.1              | 67.7              | 59.3              | 67.8              | 61.2              | 48.6              | 53.3              |   |
| 1172                | 1102                | 809                 | 889                 | 71.5              | 71.1              | 68.3              | 60.3              | 64.3              | 62.1              | 48.0              | 55.3              |   |
| 1286                | 1016                | 781                 | 798                 | 77.5              | 80.1              | 71.3              | 61.9              | 70.7              | 62.1              | 51.9              | 57.3              |   |
| 1216                | 1018                | 831                 | 929                 | 74.9              | 77.1              | 73.3              | 63.0              | 67.8              | 62.1              | 51.9              | 59.3              |   |
| 1198                | 1069                | 833                 | 893                 | 74.4              | 72.1              | 70.8              | 60.9              | 65.7              | 62.1              | 49.9              | 56.3              |   |
| 1197                | 1039                | 858                 | 927                 | 74.9              | 74.0              | 72.8              | 61.9              | 66.7              | 62.1              | 51.9              | 58.3              |   |

**Fig. 2.** Final output summarizing chilling accumulation for each apricot genotype in the BxC population across multiple seasons. The table presents values from three chilling models—Chill Hours, Chill Units, and Chill Portions—for each genotype and growing season. This standardized output facilitates the comparison of chilling requirements across years and genotypes, supporting downstream genetic and phenological analyses.

as an Excel file, facilitating the integration of climatic and phenological data for subsequent analyses.

The code below integrates phenological and climatic data for two apricot progenies, ‘BxC’ and ‘GxC’, using the *process\_PLS\_file* function. It begins by retrieving lists of Excel files containing Chill Portions data for each genotype from their respective directories. The function is then applied to both datasets, extracting the chilling period information for each genotype and saving the results into separate Excel files for streamlined analyses.

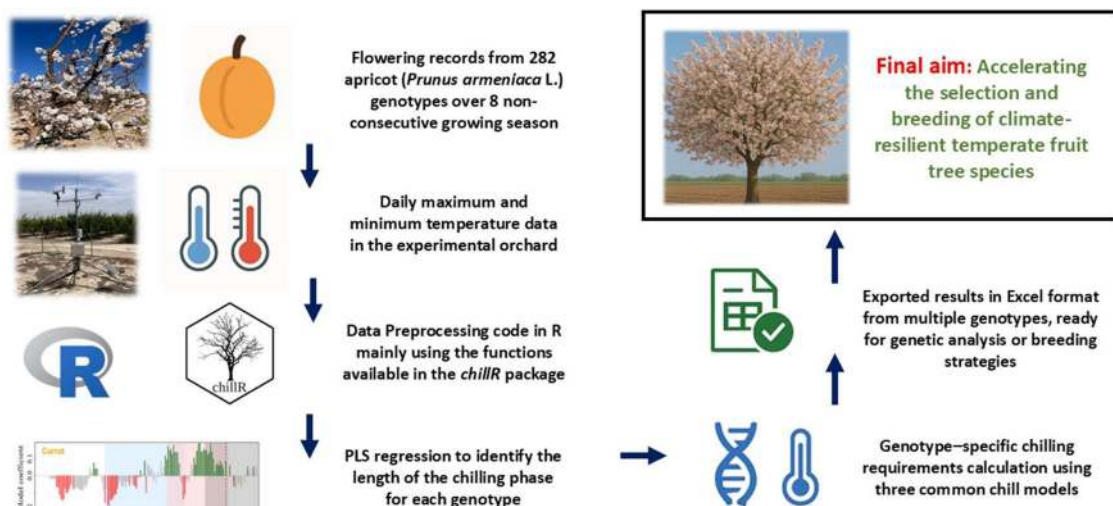
```
# BxC progeny
# List of Excel files to process
files_PLS_BxC <- list.files(path = "Genotypes/Results_PLS_Genotypes/BxC", pattern = "*.xlsx", full.names = TRUE)
# Read the phenology file corresponding to the genotypes
data_pheno_BxC<-read_Excel("BxC_edit_filt.xlsx")
# Run the function for the BxC genotypes
process_PLS_file(files_PLS_BxC,data_pheno_BxC,"Result_PLS_genotypes_BxC.xlsx")
# GxC progeny
files_PLS_GxC<- list.files(path = "Genotypes/Results_PLS_Genotypes/GxC", pattern = "*.xlsx", full.names = TRUE)
data_pheno_GxC<-read_Excel("GxC_edit_filt.xlsx")
# Run the function for the GxC progeny
process_PLS_file(files_PLS_GxC,data_pheno_GxC,"Result_PLS_genotypes_GxC.xlsx")
```

The *process\_Jday* function calculates temperature-based metrics for each genotype in a dataset, with a primary focus on chill accumulation. For every genotype, the function determines the starting Julian day (Start\_JDay) based on the provided daily values and processes the corresponding hourly temperature data using the *stack\_hourly\_temps* function. It then applies the *tempResponse* function to estimate chilling responses according to several chill models. The results for each genotype are stored in a structured list and exported as individual Excel files, named using the corresponding genotype and group ID. This approach enables a detailed, genotype-specific assessment of chilling requirements across multiple genotypes.

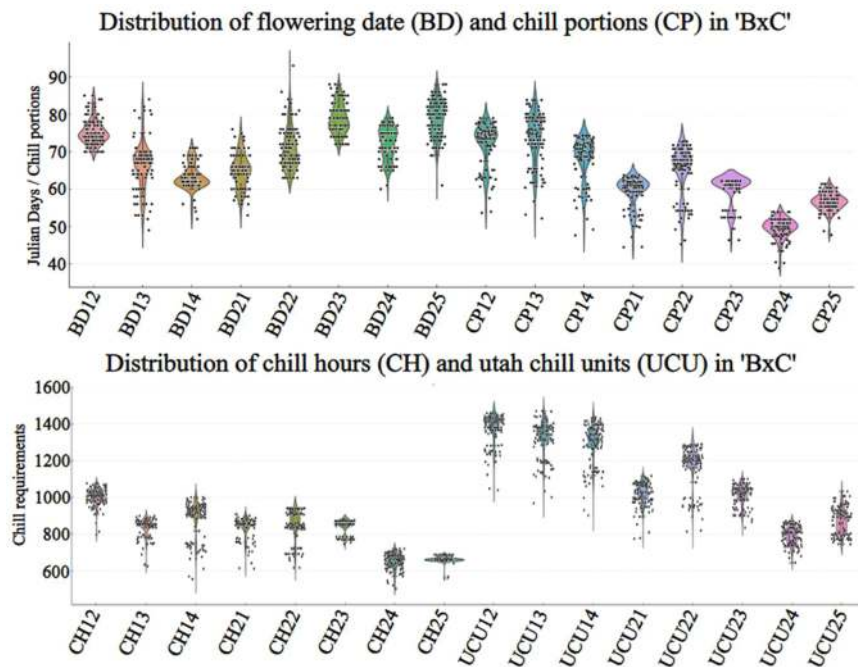
```

process_Jday <- function(data_PLS, group, path_output) {
  # Create an empty list to store the results for each genotype
  result_Jday <- list()
  # Iterate over each row (i.e., genotype) in the dataset
  for (i in 1:nrow(data_PLS)) {
    genotype <- data_PLS$Genotypes[i] # Extract the genotype name
    first_day <- data_PLS$first_day[i] # Extract the first day
    end_day <- data_PLS$end_day[i] # Extract the end day
    # Define the starting Julian day by adding 365 to the first day
    Start_JDay <- first_day + 365
    # Generate hourly temperature data for the location based on the Latitude
    hourtemps <- stack_hourly_temps(weather, latitude = 38.11)
    # Compute temperature responses using different chilling models
    result <- tempResponse(
      hourtemps,
      Start_JDay = Start_JDay,
      End_JDay = end_day, # Use the last valid day for the genotype
      models = list(
        Chilling_Hours = Chilling_Hours, # Chilling Hours model
        Utah_Chill_Units = Utah_Model, # Utah Chill Units model
        Chill Portions = Dynamic_Model, # Dynamic Chill Portions model
      ),
      misstolerance = 50, # Allow for missing values up to a specified tolerance
      whole_record = FALSE # Analyze only the specified time range
    )
    # Store the results in the list, indexed by genotype
    result_Jday[[genotype]] <- result
  }
  # Write the results for each genotype to Excel files
  for (genotype in names(result_Jday)) {
    name_file <- paste0(group, "_", genotype) # Generate a unique filename
    write.xlsx(
      result_Jday[[genotype]],
      paste0(path_output, name_file, "_CP_results.xlsx") # Save results with group and genotype info
    )
  }
}
process_Jday(read_Excel("Genotypes/Results_PLS_Genotypes/Result_PLS_genotypes_BxC.xlsx"), "BxC", "Genotypes/Results_PLS_Genotypes/Results_CP/BxC/")
process_Jday(read_Excel("Genotypes/Results_PLS_Genotypes/Result_PLS_genotypes_GxC.xlsx"), "GxC", "Genotypes/Results_PLS_Genotypes/Results_CP/GxC/")

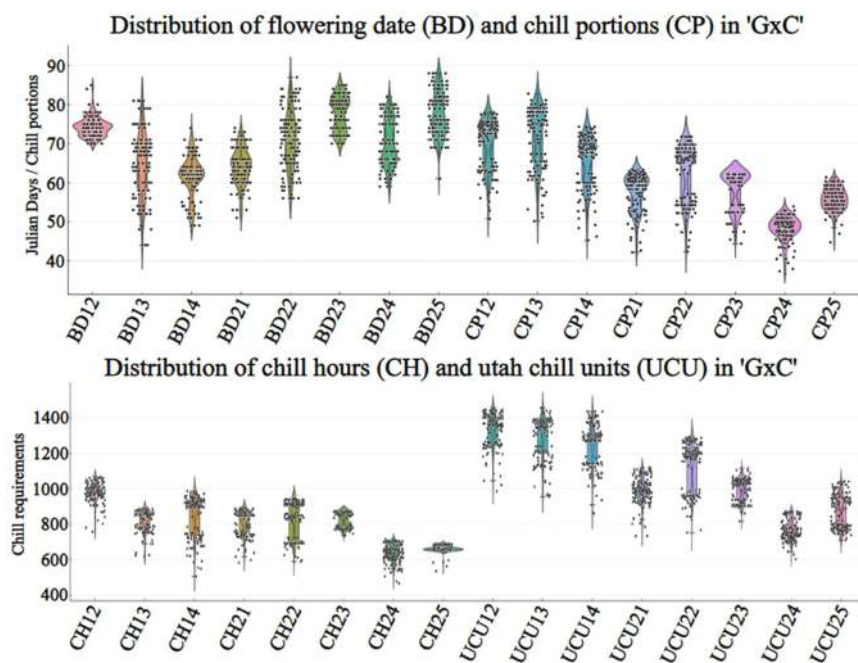
```



**Fig. 3.** Graphical summary of the workflow, showing input files (flowering and climate data in Excel files), intermediate steps (preprocessing and PLS regression analysis), and the final outputs (genotype-specific chilling requirement estimates).



**Fig. 4.** Distribution of flowering dates values and chilling requirement estimates (CH: Chill Hours, UCU: Utah Chill Units, and CP: Chill Portions) in the 'BxC' *P. armeniaca* population across multiple years. Each point represents an individual genotype, and the violin plots illustrate the density distribution for each season.



**Fig. 5.** Distribution of flowering dates values and chilling requirement estimates (CH: Chill Hours, UCU: Utah Chill Units, and CP: Chill Portions) in the 'GxC' *P. armeniaca* population across multiple years. Each point represents an individual genotype, and the violin plots illustrate the density distribution for each season.

The *process\_final\_file* function processes multiple Excel files containing phenological data and extracts chilling metrics for selected seasons. It begins by filtering the data to retain only the predefined seasons of interest. Next, a year-based suffix is then extracted to identify each season, and the corresponding genotype is assigned based on the file name. The function dynamically reshapes the data using pivot tables, generating separate columns for each season, each containing the three chill metrics used in the analysis: Chill Hours, Chill Units and Chill Portions. Finally, the results from all files are compiled and saved into a single Excel file, which serves as the final output of the workflow.

```
process_final_file <- function(files, name) {
  # Create an empty data frame to store the combined results
  combined_results <- data.frame()
  for (i in files) {
    # Load data from the Excel file
    data <- read_Excel(i)
    name_file <- tools::file_path_sans_ext(basename(i)) # Extract the file name without the extension
    # Filter data to include only relevant seasons (seasons with flowering records)
    seasons <- c("2011/2012", "2012/2013", "2013/2014", "2020/2021",
                 "2021/2022", "2022/2023", "2023/2024", "2024/2025")
    filtered_data <- data %>%
      filter(Season %in% seasons) # Select rows corresponding to specified seasons
    # Extract the year suffix to identify specific seasons
    filtered_data <- filtered_data %>%
      mutate(
        Year_suffix = str_sub(Season, 8, 9), # Extract the last two digits of the starting year
        Trait = str_extract(name_file, "\\d+\\.\\d+") # Identify the genotype from the file name
      )
    # Reshape the dataset to generate dynamic columns for each chilling metric and season
    pivoted_data <- filtered_data %>%
      select(Trait, Year_suffix, Chilling_Hours, Utah_Chill_Units, Chill_Portions) %>%
      pivot_wider(
        names_from = Year_suffix, # Use the year suffix as column names
        values_from = c(Chilling_Hours, Utah_Chill_Units, Chill_Portions) # Fill columns with respective
        values
      )
    # Merge the processed dataset with the cumulative results
    combined_results <- bind_rows(combined_results, pivoted_data)
  }
  # Export the combined results to an Excel file
  write_xlsx(combined_results, paste0(name, "_final_result.xlsx"))
}
```

In the final step of the workflow, all Excel files from the specified directories are read, relevant data are filtered, and seasonal chilling metrics are extracted. The processed results are then compiled to generate a final output file for each genotype. This step ensures a standardized, automated, and reproducible process for summarizing seasonal chill metrics across all genotypes (Fig. 2).

```
# Process results for BxC genotype
result_BxC<-process_final_file(list.files("Genotypes/Results_PLS_Genotypes/Results_CP/BxC/"),pattern = "\\
.xlsx",full.names = TRUE),"BxC")
# Process results for GxC genotype
result_GxC<-process_final_file(list.files("Genotypes/Results_PLS_Genotypes/Results_CP/GxC/"),pattern = "\\
.xlsx",full.names = TRUE),"GxC")
```

### Method validation

In fruit tree breeding, phenotyping is often a major bottleneck, especially for traits linked to dormancy release and flowering. Determining cultivar-specific CR is particularly important, since this trait defines cultivar adaptability to specific climates under current and future conditions. Traditionally, CRs are estimated using the forcing shoots method, which involves collecting dormant shoots during winter and placing them under controlled conditions (20–25 °C) to induce bud break. Although widely used, this approach is labor-intensive, time-consuming, and unsuitable for large breeding populations. The methodology presented here provides an efficient alternative, allowing breeders to rapidly process and analyze extensive phenological datasets and generate standardized CR values for downstream applications (Fig. 3).

The outputs generated in this study have several potential applications, including QTL mapping, which relies on precise phenotypic

data across segregating populations such as those evaluated in this study (BxC' and 'GxC'). Additionally, it facilitates the development of cultivar portfolios to guide growers in selecting varieties best suited to specific growing regions. As part of the validation, we present scatter plots showing the distribution of CR across numerous genotypes from both populations (Fig. 4 and Fig. 5). These visualizations provide a valuable reference for future studies investigating the genetic basis of CR through integrated phenotypic and genotypic analyses, ultimately facilitating marker-trait association studies aimed at developing climate-resilient cultivars.

### Limitations

The main contribution of this study is the development of an R-based workflow for estimating CR across a large number of temperate fruit tree genotypes using PLS regression. Reliable CR estimates are a valuable resource for agronomists, breeders, and researchers working on dormancy-related topics. The applicability of the code was demonstrated with a case study on apricot trees, and the CR results are well aligned with our current understanding of CR variation in a segregating population derived from a low- and high-chill parental crosses. Notably, the CR estimate for one of the parents in the 'GxC' progeny ('Currot') was consistent with previous values obtained using the forcing method at the same experimental site [6]. However, the relatively limited number of flowering seasons in our dataset may constrain the robustness of the results. Expanding the dataset with additional phenological records could improve accuracy. Additionally, this workflow improves the comparability of CR studies by establishing a standardized criterion for defining the chill accumulation period. While PLS regression has proven effective in estimating CR when experimental data are not available, some inconsistencies have been reported when comparing previous studies [23,24]. In this sense, future research should focus on refining methods to accurately determine the end of dormancy, as well as improving chill models that better represent how different temperature regimes influence phenology.

### Ethics statements

Not applicable.

### CRediT author statement

**Ana M. Muñoz-Morales:** Conceptualization, Methodology, Software. **Germán Ortuño-Hernández:** Validity tests, Data curation. **Juan Alfonso Salazar:** Visualization, Investigation, Data curation. **Pedro Martínez-Gómez:** Supervision. **Jose A. Egea:** Software, Validation. **David Ruiz:** Supervision, Data curation. **Álvaro Delgado:** Conceptualization, Methodology, Writing- Reviewing and Editing.

### Acknowledgments

The authors acknowledge the support of the Spanish Ministry of Science and Innovation through the project "Apricot Breeding (APRIBRED)" (PID2022–137392OB-I00). This research was also conducted within the framework of the AGROALNEXT programme, supported by the Ministry of Science and Innovation (MCIN), the European Union (NextGenerationEU, PRTR-C17.I1), and Fundación Séneca, funded by the Comunidad Autónoma de la Región de Murcia (CARM). Álvaro Delgado acknowledges support from the Spanish Ministry of Science and Innovation through a postdoctoral research grant (JDC2022–050,286-I), Juan Alfonso Salazar is supported by a 'Ramón y Cajal' postdoctoral contract (RYC2022–038,101-I), and Germán Ortuño-Hernández by a predoctoral fellowship (FPU21/03,563).

### Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Supplementary material and/or additional information [OPTIONAL]

*If you do submit supplementary files, you are welcome to provide supporting details in this OPTIONAL section.*

### Data availability

The complete code and associated datasets are freely available in a public repository (<https://github.com/CEBASFruitBreed/R-workflow-ChillPLS>)

### References

- [1] J.A. Salazar, D. Ruiz, J.A. Campoy, R. Sánchez-Pérez, C.H. Crisosto, P.J. Martínez-García, A. Blenda, S. Jung, D. Main, P. Martínez-Gómez, M. Rubio, Quantitative trait loci (QTL) and mendelian trait loci (MTL) analysis in Prunus: a breeding perspective and beyond, *Plant. Mol. Biol. Rep.* 32 (2014) 1–18, <https://doi.org/10.1007/s11105-013-0643-7>.
- [2] J.A. Campoy, D. Ruiz, J. Egea, Dormancy in temperate fruit trees in a global warming context: a review, *Sci. Hortic.* 130 (2011) 357–372, <https://doi.org/10.1016/j.scienta.2011.07.011>.

- [3] M. Faust, A. Erez, L.J. Rowland, S.Y. Wang, H.A. Norman, Bud dormancy in perennial fruit trees: physiological basis for dormancy induction, maintenance, and release, *HortScience* 32 (1997) 623–629, <https://doi.org/10.21273/HORTSCI.32.4.623>.
- [4] A. Delgado, J.A. Egea, J. López-Alcolea, J.A. Campoy, D. Ruiz, The timing and pattern of chill and heat exposure affect the estimation of agroclimatic requirements and adaptability potential in *Prunus* cultivars, *Sci. Hortic.* 339 (2025) 113910, <https://doi.org/10.1016/j.scienta.2024.113910>.
- [5] E. Fadón, E. Fernandez, H. Behn, E. Luedeling, A conceptual framework for winter dormancy in deciduous trees, *Agronomy* 10 (2020) 241, <https://doi.org/10.3390/agronomy10020241>.
- [6] A. Delgado, J.A. Egea, E. Fernandez, J.A. Campoy, J. Egea, F. Dicenta, D. Ruiz, Discrepancies in methodologies to determine chill requirements in temperate fruit trees constrain guidelines for future plantings in a global warming context, *Agric. For. Meteorol.* 349 (2024) 109970, <https://doi.org/10.1016/j.agrformet.2024.109970>.
- [7] D. Ruiz, J.A. Campoy, J. Egea, Chilling and heat requirements of apricot cultivars for flowering, *Environ. Exp. Bot.* 61 (2007) 254–263, <https://doi.org/10.1016/j.envexpbot.2007.06.008>.
- [8] A. Delgado, J.A. Egea, E. Luedeling, E. Dapena, Agroclimatic requirements and phenological responses to climate change of local apple cultivars in northwestern Spain, *Sci. Hortic.* 283 (2021) 110093, <https://doi.org/10.1016/j.scienta.2021.110093>.
- [9] E. Luedeling, A. Gassner, Partial least squares regression for analyzing walnut phenology in California, *Agric. For. Meteorol.* 158–159 (2012) 43–52, <https://doi.org/10.1016/j.agrformet.2011.10.020>.
- [10] J. Martínez-Lüscher, P. Hadley, M. Ordidge, X. Xu, E. Luedeling, Delayed chilling appears to counteract flowering advances of apricot in southern UK, *Agric. For. Meteorol.* 237–238 (2017) 209–218, <https://doi.org/10.1016/j.agrformet.2017.02.017>.
- [11] E. Luedeling, L. Caspersen, E. Fernandez, chillR: statistical methods for phenology analysis in temperate fruit trees. R Package Version 0.76 (2024). <http://cran.r-project.org/web/packages/chillR/>.
- [12] M. Baggiolini, Les stades repérés des arbres fruitiers à noyau, *Rev. Romande Agric. Vitic. Arboric* 8 (1952) 3–4.
- [13] J. Almorox, C. Hontoria, M. Benito, Statistical validation of daylength definitions for estimation of global solar radiation in Toledo, Spain, *Energy Conv. Manag.* 46 (2005) 1465–1471.
- [14] D.E. Linvill, Calculating chilling hours and chill units from daily maximum and minimum temperature observations, *HortSci* 25 (1990) 14–16, <https://doi.org/10.21273/HORTSCI.25.1.14>.
- [15] A. Delgado, J.A. Egea, J. López-Alcolea, J.A. Campoy, D. Ruiz, The timing and pattern of chill and heat exposure affect the estimation of agroclimatic requirements and adaptability potential in *Prunus* cultivars, *Sci. Hortic.* 339 (2025) 113910.
- [16] A. Delgado, D. Ruiz, J. Rodrigo, E. Luedeling, E. Fernandez, E. Fadón, Cultivar plasticity in temperate fruit trees across the Mediterranean region: different combinations of chill and heat lead to dormancy release, *Eur. J. Agron.* 173 (2025).
- [17] S. Fishman, A. Erez, G. Couvillon, The temperature dependence of dormancy breaking in plants: computer simulation of processes studied under controlled temperatures, *J. Theor. Biol.* 126 (1987) 309–321.
- [18] S. Fishman, A. Erez, G. Couvillon, The temperature dependence of dormancy breaking in plants: mathematical analysis of a two-step model involving a cooperative transition, *J. Theor. Biol.* 124 (1987) 473–483.
- [19] E. Fernandez, C. Whitney, E. Luedeling, The importance of chill model selection — A multi-site analysis, *Eur. J. Agron.* 119 (2020) 126103, <https://doi.org/10.1016/j.eja.2020.126103>.
- [20] E. Luedeling, E.H. Girvetz, M.A. Semenov, P.H. Brown, Climate change affects winter chill for temperate fruit and nut trees, *PLoS. One* 6 (2011) e20155.
- [21] E. Luedeling, A. Gassner, Partial least squares regression for analyzing walnut phenology in California, *Agric. For. Meteorol.* 158 (2012) 43–52.
- [22] K.S. Pope, D.Da Silva, P.H. Brown, T.M. DeJong, A biologically based approach to modeling spring phenology in temperate deciduous trees, *Agric. For. Meteorol.* 198–199 (2014) 15–23, <https://doi.org/10.1016/j.agrformet.2014.07.009>.
- [23] E. Fadón, S. Herrera, B.I. Guerrero, M.E. Guerra, J. Rodrigo, Chilling and heat requirements of temperate stone fruit trees (*Prunus* sp), *Agronomy* 10 (2020) 409.
- [24] C. Gonzalez-Noguer, Apple (*Malus × domestica* Borkh.) dormancy: a review of regulatory mechanisms and agroclimatic requirements, *Front. Horticul.* 2 (2023) 1217689.