Multisource Grapevine Phenology Dataset for Smart Farming and AI Modeling

Francisco José Lacueva Pérez
fjlacueva@unizar.es

Universidad de Zaragoza https://orcid.org/0000-0003-0998-2939

Rafael del Hoyo-Alonso
Instituto Tecnológico de Aragón https://orcid.org/0000-0003-2755-5500

Gorka Labata-Lezaún
Instituto Tecnológico de Aragón https://orcid.org/0000-0002-8634-4124

Juan José Barriuso-Vargas
Universidad de Zaragoza https://orcid.org/0000-0003-2980-5454

Sergio Ilarri-Artigas
Universidad de Zaragoza https://orcid.org/0000-0002-7073-219X

Data Note

Keywords: Smart Farming, Grapevine Phenology Dataset, Machine Learning, FAIR Principles

Posted Date: May 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9781395/v1

Additional Declarations: The authors declare no competing interests.

Multisource Grapevine Phenology Dataset for Smart Farming and AI Modeling

Francisco José Lacueva[1] [1,2], Rafael del Hoyo [2] [1], Gorka Labata [3] [1], Juan José Barriuso [4] [2], and Sergio Ilarri [¶] [2]

[1] Instituto Tecnológico de Aragón (Spain)
[2] I3A, Universidad de Zaragoza (Spain)

May 19, 2026

Abstract

Artificial Intelligence and Machine Learning rely on large, high-quality datasets for accurate and robust models, yet data scarcity remains a major challenge, especially in smart farming. Phenology modeling, a key application, studies how plant biological events relate to climate and seasons. Accurate phenology models improve crop quality, support climate adaptation, and guide decisions such as pesticide use and harvesting, enhancing environmental and economic sustainability. However, agricultural data are highly diverse and heterogeneous, complicating model development.

This study presents a proposed georeferenced dataset for Machine Learning-based grapevine phenology prediction across 3 Protected Designations of Origin in Aragón, Spain. Developed by a multidisciplinary team, the dataset combines 9 datasets from 8 sources (including meteorological time series, field phenology observations, and Copernicus Sentinel-2 multispectral imagery) covering the period 2016–2022. It supports both physical and Machine Learning-based phenology modeling and facilitates knowledge extraction in agronomy and plant biology. Its relevance lies in its comprehensive scope, the inclusion of 9 phenological stages, and a rigorous methodology ensuring reproducibility. This framework enables the creation of similar datasets for other regions or crops, advancing smart farming through scalable, data-driven solutions. The open publication of the code further supports this objective. We further anticipate its potential contribution to developing foundation models as well as to the creation of new knowledge in biology and agronomy.

Keywords: Smart Farming, Grapevine Phenology Dataset, Machine Learning, FAIR Principles.

1 Introduction

Artificial intelligence (AI), and especially machine learning (ML), relies on the availability of large and high-quality datasets. These data are essential for training models capable of performing tasks such as classification, prediction, and pattern detection. Without representative and well-structured data, models may suffer from overfitting, bias, or poor generalization, which limits their usefulness in real-

[1] fjlacueva@unizar.es, fjlacueva@ita.es

[2] rdelhoyo@ita.es

[3] glabata@ita.es

[4] jvargas@unizar.es [¶]silarri@unizar.es

world applications. Therefore, the quality of a dataset is as important as its size, since it directly affects the accuracy and robustness of the model [97].

From the point of view of the data, AI faces an important challenge, which is the scarcity of datasets to train very specific models in different fields such as health, industry, finance, etc. [4, 7, 46, 91]. Smart farming is not an exception, facing unique challenges due to the high diversity and heterogeneity of agricultural data [100]. Whilst Open Data initiatives offer a viable path to address this data scarcity [7, 47, 52], successfully integrating these heterogeneous sources requires domain expertise. Consequently, realizing the full potential of smart-farming solutions necessitates the formation of multidisciplinary teams capable of bridging data science and agronomy [58, 87].

The dataset proposed in this paper is based on grapevine data from 3 Protected Designations of Origin (PDOs) in Aragón, northeastern Spain. A multidisciplinary team was involved in its creation. This paper presents the finalized, fully integrated, and open-access multisource dataset. The main contributions of this dataset to existing phenological datasets are:

1. A rigorous methodology for dataset creation and refinement, enabling reproducibility in other grapevine regions and for different crops.

2. Integration of 9 heterogeneous datasets into a unified dataset, following the proposed methodology.

3. Inclusion of 9 phenological stages, significantly exceeding the 2 or 3 stages typically considered in previous studies.

4. A substantial number of independent variables for ML model training. The dataset combines raw climatic data with derived variables computed using formulas widely employed by biologists and agronomists to model plant phenology. These variables incorporate key parameters identified in literature, allowing ML algorithms to determine the most relevant factors and formulas whilst generating new insights in biology and agronomy.

5. This work adheres to the FAIR principles [98]. Beyond hosting the dataset on Zenodo [101] to ensure findability and accessibility, we achieve semantic interoperability and reusability by standardizing field observations into the international BBCH scale [12, 77]. Furthermore, the source code used for dataset creation and model training is publicly available on GitHub [57].

The structure of the rest of this paper is as follows. Section 2 introduces the context of this work, emphasizing the challenges of working with agricultural datasets, especially those concerning grapevine phenology, and includes a brief review of existing open-access datasets along with their limitations in spatial and temporal resolution. Section 3 outlines the experimental design, materials, and methods used to build the dataset, integrating field observations, meteorological records, and multispectral satellite imagery: it describes the data processing workflow, including phenological stage normalization, climate data curation, NDVI (Normalized Difference Vegetation Index) calculation from Sentinel-2 imagery, and the final dataset assembly for ML model training.
Section 4 briefly describes the dataset and its usefulness. Section 5 discusses the implications of the dataset, particularly in relation to data privacy and the risks associated with linking field observations to specific vineyard parcels. Finally, Section 6 presents the conclusions drawn from this study. The main text is complemented with two appendices, that include descriptive statistics of the dataset in

Appendix A and the naming conventions of the fields of the dataset along with a complete list of attributes in Appendix B.

2 Background

As in other domains where AI technologies are applied, one of the key challenges in smart farming is the lack of large high-quality open datasets that can be used to develop accurate and generalizable models [7, 44, 83]. There are several factors contributing to this scarcity of (high-quality) data:

- The variability inherent in natural systems, the diversity of farming practices among growers, the absence of standardized data recording methods, farmers' hesitancy to adopt new technologies, and concerns over data sharing (stemming from issues related to privacy, proprietary methods, and competitive advantage) pose significant challenges to agricultural data collection [7, 29, 31, 44, 83, 100].

- Agricultural data are also highly diverse and heterogeneous, originating from sources such as weather stations, satellites, drones, and sensors. These data vary in format, accuracy, update frequency, and reliability, with each type serving distinct purposes, such as crop yield prediction or pest detection [7, 100].

- Many existing datasets are biased and fragmented, often focusing primarily on large-scale industrial farms [29]. Even with precision agriculture techniques, data are often insufficient due to missing key information, such as weather conditions, irrigation schedules, and pest pressure—or due to a lack of temporal and spatial alignment across datasets [29, 31]. • The absence of a centralized system for aggregating data from the world's 570 million farms limits the ability of AI to leverage this vast pool of knowledge. Additionally, historical data are often scarce [83] and costly to obtain [31].

One major challenge is the difficulty of developing universal AI solutions that can be effectively applied across diverse agricultural regions, environments, and contexts [31, 83]. To harness the full potential of AI in agriculture, it is essential to ensure consistent and reliable data collection [31]. Key approaches include the creation of metadata standards and uniform data collection protocols, encouraging crowdsourced data gathering, and building open-access data platforms [7]. Technologies such as drones and IoT sensors have significantly enhanced data acquisition capabilities. Moreover, emerging methods like Generative Adversarial Networks (GANs) and self-supervised learning are opening new possibilities for data generation and labeling [100].

On the other hand, ongoing collaboration between data scientists and agricultural experts is essential to ensure that AI models are built on relevant information and deliver practical value [44]. This collaboration is particularly crucial for determining how to curate, transform, and accurately link data from diverse sources across both temporal and spatial dimensions, which is an essential step in building effective AI models [7, 44, 83]. Additionally, it plays a key role in selecting appropriate anonymization and encryption strategies, and in drafting confidentiality agreements that safeguard intellectual property whilst still enabling the extraction of valuable, fine-grained insights from the data [44].

In Section 2.1, we provide a brief overview of the data sources commonly used to develop ML models for phenology prediction. Subsequently, in Section 2.2, we present a short survey of existing open-access phenology datasets, with a particular focus on grapevine.

2.1 Woody Crops Phenology Data Sources

Phenology examines periodic biological events in relation to climatic and seasonal conditions [72]. In the case of plants, it focuses on the relationship between their annual life cycles and environmental factors such as climate variations and soil conditions [95]. These models play a crucial role in enhancing crop quality and productivity, supporting adaptation to global warming, and informing key agricultural decisions, such as optimal timing for pesticide application or harvesting [27, 94]. Ultimately, the ability to accurately predict phenological changes can significantly benefit both the environmental and financial sustainability of agricultural operations [90].

Big data analytics and ML have been increasingly applied to develop data-driven models for phenology prediction and monitoring [58]. Commonly used datasets include (crowdsourced) field observations, meteorological data (whether from station measurements [22] or estimations [67]), and multispectral imagery captured by drones and satellites [15]. These datasets may be used in their raw form, transformed, isolated, or combined.

In their studies, Canavera et al., Oses et al., and Dai et al. utilized temperature-based indices [15, 23, 67]. The Growing Degree Day (GDD) [13] and the Winkler index [5] are commonly used to determine the accumulated heat required to reach specific phenological stages. Equation 1 and Equation 2 show how these indexes are calculated, where *Tb* represents the temperature at which the metabolism of the plant is activated, T_m is the minimum temperature of the day, and T_M is the maximum temperature of the day; when any of these T_m or T_M temperatures is below 0°C, it is considered equal to 0°C (*effective temperature*) [3]. Similarly, indices such as cumulative chilling hours [96] and the Utah (Richardson) model [75] are used to quantify the chilling requirements necessary for bud burst.

$$GDD = \sum \begin{cases} \frac{T_M - T_m}{2} - T_b, & \text{if } \frac{T_M - T_m}{2} \geq T_b \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

$$WI = \sum \begin{cases} \frac{T_M + T_m}{2} - T_b, & \text{if } \frac{T_M + T_m}{2} \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Regarding the transformations applied to multispectral images for phenological monitoring, various vegetation indices are computed using relatively simple formulas based on the pixel values across different spectral bands. The most commonly used index is the Normalized Difference Vegetation Index (NDVI) [30], which effectively reflects the health of the plant canopy [87]. For each pixel of the multispectral image, the NDVI is calculated considering the values of the Near Infrared (NIR) and red bands (RED) using Equation 3. Azpiroz et al. [8] employed the NDVI to predict the phenology of olive trees, whilst Devaux et al. [26] and Tassopoulos et al. [88] used it to forecast the phenology of grapevines.

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{3}$$

Finally, some researchers, such as Li et al. [60] and Brook et al. [14], have integrated meteorological data, temperature-derived indices, and indices derived from multispectral satellite imagery into comprehensive datasets to enhance phenological predictions.

An important aspect to consider when developing ML models for phenology is the number of phenological stages included in the field observations datasets. Due to the labor-intensive process of collecting full-season datasets across multiple years, most studies focus only on a subset of these stages. The most commonly considered stages are bud burst [17, 71, 74, 76], flowering [17, 74, 76], berry development [17, 71, 74, 76], berry veraison [74, 76], and senescence [17, 76]. As a result, the models produced may lack accuracy when assessing the actual effectiveness of pest control treatments [42]. Due to this, our goal was to build a rich dataset considering a high number of phenological states and a wide variety of relevant attributes.

2.2 Phenology (Open) Datasets

A dataset is a structured collection of related observations, organized and formatted for a specific purpose [19]. Making a dataset openly available enables its reuse in various ways: to foster innovation and improve business processes, to enhance government transparency and empower citizens, and to support the reproducibility and dissemination of scientific research [19]. Open data are typically defined as digital information made available with the technical and legal conditions necessary for anyone to freely use, reuse, and redistribute it (anytime and anywhere) [79]. However, to fully unlock the value of a dataset, it must adhere to the FAIR principles: Findable, Accessible, Interoperable, and Reusable [19, 49]. Whilst open data and FAIR data share some conceptual foundations, FAIR introduces additional requirements, such as user authorization and verification [49]. Devaraju et al. established a comprehensive set of core metrics to operationalize the FAIR principles [24], alongside an automated tool designed to facilitate systematic evaluation based on these criteria, which is the F-UJI web application [25].

There are numerous initiatives aimed at promoting data sharing and reuse, ranging from scientific publishers requiring authors to submit datasets alongside their manuscripts, to the development of data marketplaces, open data portals, and specialized data communities [19]. Prominent examples of dataset repositories include Kaggle [50], Mendeley [32], Zenodo [101], Data.gov [1] or Copernicus Data Space Ecosystem (CDSE) [73].

Moreover, solutions that enable automated negotiation, authorization, and validation (ensuring datasets are used in accordance with the intentions of their creators) are still under active scientific development. The European Common Data Space initiative seeks to advance the creation of such systems by defining a shared architecture for data and services. This architecture supports diverse implementations of core functionalities, enables federated and cross-sectoral data spaces, and accommodates sector-specific requirements, such as those defined by the Agriculture Data Space and the European Open Science (EOSC) Data Space [16].

In a survey by Chamorro et al. performed in 2024 [18] an important insight was that most of the available agricultural (open) datasets are based on satellite images (probably, because they are made freely available) and other kinds of images, but that there is a lack of other kinds of information. This conclusion is aligned with the results of our survey of existing (open) datasets that can be used to create ML (grapevine) phenology models, that we performed during the creation of this paper: most of the phenology datasets consist mainly of images.

To conduct the survey, we consulted dataset repositories including Zenodo [101], Kaggle [50], Data.gov [1], and Mendeley [32]. We also leveraged the search capabilities of the Google Dataset Search Engine [39]. Using their respective user interfaces, we searched for datasets matching the following queries: “phenology”, “crop phenology”, “woody phenology”, and “grapevine phenology”. The number of datasets retrieved for each repository and query is summarized in Table 1. Table 1: Number of datasets in the analysed dataset repositories by search query.

Repository	**Total Datasets**	**Phenology**	**Crop Phenology**	**Woody Phenology**	**Grapevine Phenology**
Zenodo [101]	469506	1742	12	1	2
Kaggle [50]	527075	7	1	0	0
data.gov [1]	354462	373	37	2	0
Mendeley [32]	123748	56	743	839	87
Google [39]	N.A.	+100	+100	+100	3

During the search for phenology datasets, we observed several semantic inconsistencies in the results returned by major repositories. For instance, on Mendeley, the query “crop phenology” (743 results) yielded more entries than the broader term “phenology”. Additionally, among the 86 results retrieved for “grapevine phenology”, only a small fraction were actually related to vineyards, and none of the relevant datasets were comparable to the one presented in this study. These inconsistencies highlight a broader limitation of current data repositories: the lack of standardized metadata protocols and rigorous dataset descriptions, which significantly reduces the precision of semantic search. A summary of the most relevant phenology datasets found is provided in Table 2.

In summary, there are a few phenology datasets considering multiple data sources and combining them in an appropriate way to be used to predict phenology, and then the risk of pests. Given the scarcity of rich datasets in this area, in our work, a multisource dataset has been created and made available in a public repository [54]. The proposed dataset combines field observations of phenology, climatic data, and data derived from multispectral satellite images considering time and spatial resolution.

3 Experimental Design, Materials and Methods

We used multiple data sources to develop the dataset we describe in this paper, including field observations [61, 71], climate data (both observed [36] and estimated [76]), and Sentinel-2 multispectral imagery [36]. Additionally, we incorporated parcel location and grape variety information from the Spanish Cadastral Registry [85] and Aragón’s CAP (Common Agrarian Policy Registry) Open Data [40], applying relevant transformations to each dataset.

The datasets cover the geographical areas of three wine PDOs in Aragón (a region in northeastern Spain): Calatayud, Campo de Borja, and Cariñena. Figure 1 shows their location; more specifically, these PDOs lie within the WGS84 coordinates (41.98107, −2.177578) and (41.166320, −0.922575), spanning approximately 8500 km^2. Approximately, 18000 hectares of vineyards at elevations ranging

Table 2: Main related grapevine phenology datasets.

Repository	Dataset	Paper	Summary
Data.gov [1]	[68]	[69]	Budbreak, flowering, véraison, and maturity from 6 grapevine varieties (Cabernet Sauvignon, Chardonnay, Pinot Noir, Zinfandel, Pinot Gris, and Sauvignon Blanc) growing in 12 areas of California (USA) combined with historical climate data from gridMET (1991–2020) and climate projections from 20 global climate models (2040–2069, RCP 4.5). In addition, it includes 14 rel evant agroclimatic metrics (e.g., growing degree days, cold hardiness, chilling degree days, and frost damage days).
Google [39]	[28]	—	Budbreak, flowering and véraison dates for the grapevine variety Riesling were recorded since 1958 at the same location: the INRA experimental vineyard in Bergheim (Haut-rhin, France).

from 300 to 1000 meters above sea level conform these PDOs. Concerning climatic stations, 2 stations from the SIAR network were associated with Calatayud, 4 with Campo de Borja, and 3 with Cariñena. From the Grapevine network, 2 stations were linked to Calatayud, 5 to Campo de Borja, and 5 to Cariñena. Notably, one Grapevine station in Calatayud was relocated during the development of this work, and both positions are reported. Table 3 summarizes the spatial and temporal resolutions of the datasets, which were integrated into a multi-variable dataset for use with BDA (Big Data Analytics) and ML (Machine Learning) techniques.

Table 3: Characterization of the input datasets and their temporal and spatial granularity.

Dataset	Since	Temporal Resolution	Spatial Resolution
Phenological Observation	2009	Weekly (+/- 3 days)	km of distance to a parcel
Climatic Dataset	2010	30’	Tile 100 km x 100 km
Sentinel-2 derived data	2016	5 days	Tile 100 km x 100 km Image 10980 x 10980 pixels 100 m^2/pixel

Figure 2 outlines a Python-based data preparation workflow considered to build the integrated final dataset. The source code for implementing this workflow, as well as to train the ML models, can be accessed in a public repository [57]. Though presented linearly, the process was iterative, refining datasets, ML models, and predictions. The CRISP-DM methodology [99] guided collaboration between farmers, agronomists, and data scientists to understand the data and design transformation sub-workflows.

The data sources used to create the workflow are shown in Figure 2 (bottom): Red FARA field observations [41], cadastral data [85], Common Agrarian Policy (CAP) registry [40], Sentinel-2 imagery [37], and climatic data from SIAR [81] and GRAPEVINE [43], along with ECMWF IFS

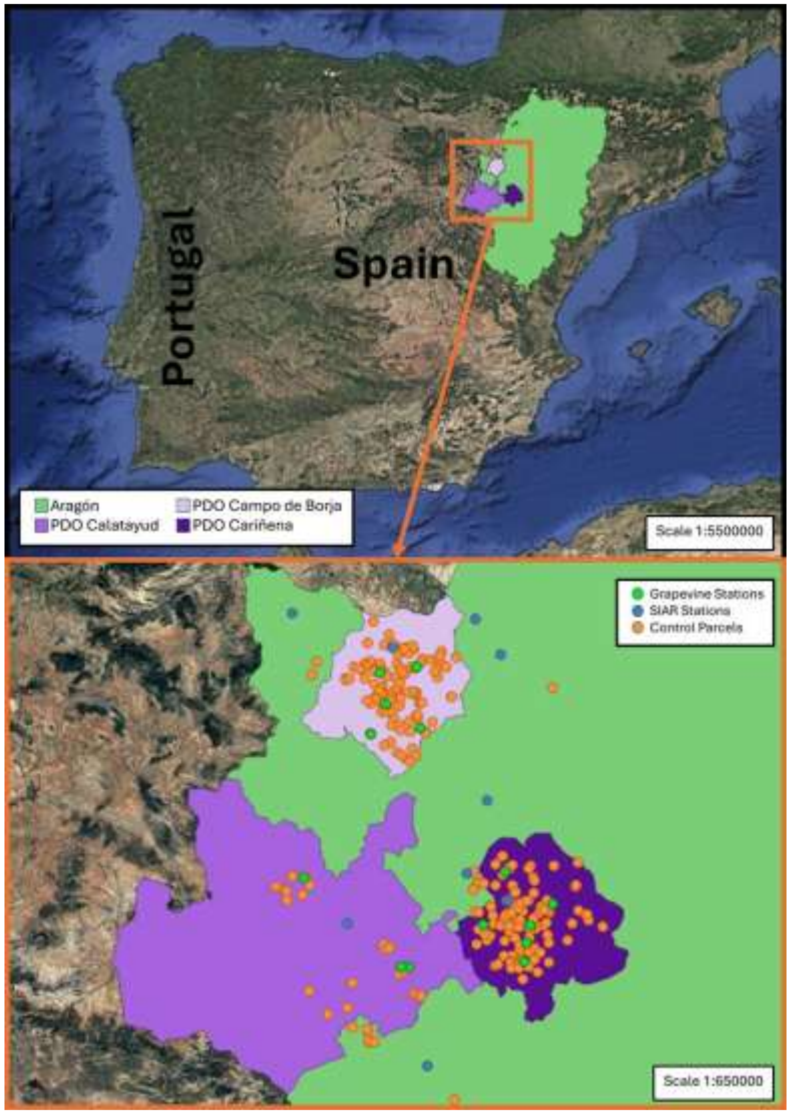

Figure 1: Distribution of the grapevine control parcels and of the climatic stations of the SIAR and Grapevine networks in the PDOs considered (Calatayud, Campo de Borja, and Cariñena). The map is provided by OpenStreetMap [66].

forecasts [34] and ERA5 estimations [20]. Data were accessed via public/private APIs, including Open Meteo for ERA5 [65]. Each dataset underwent specific transformations:

- Red FARA records were normalized to 9 phenology stages using the BBCH phenological scale to encode the states, with added states 00 and 99. Grape varieties were enriched using CAP data.

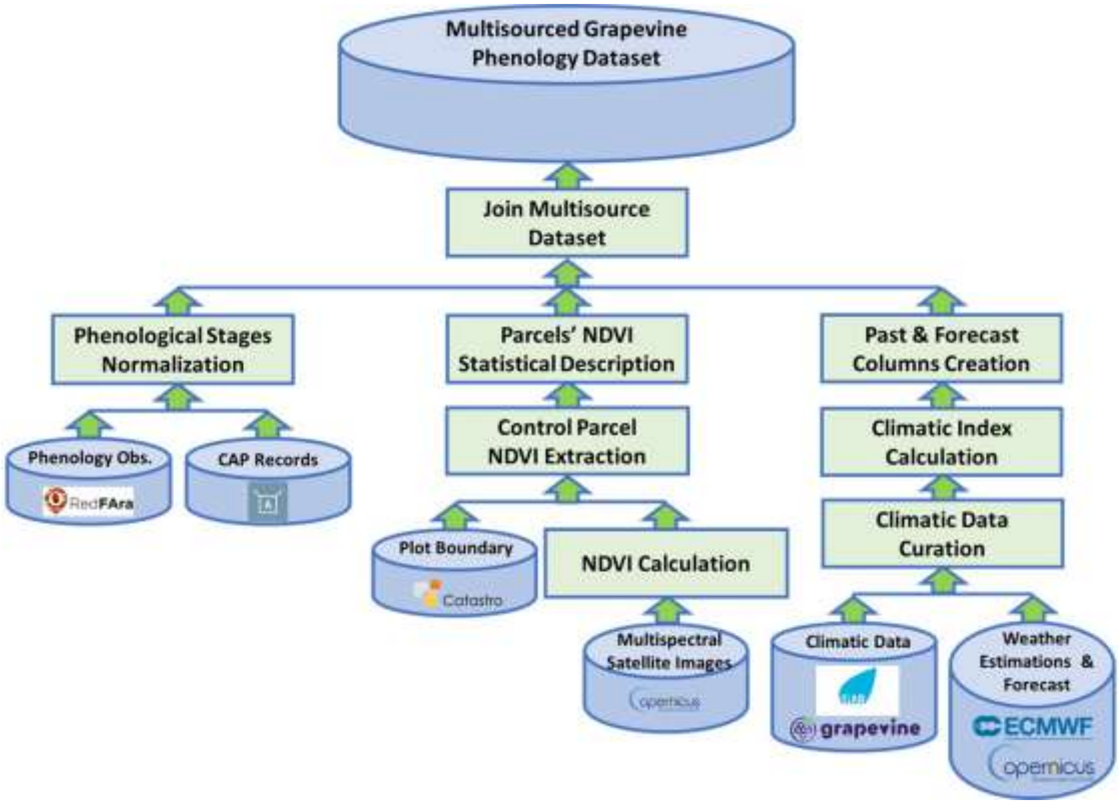

Figure 2: Workflow implemented to transform data used to create the Multisourced Grapevine Phenology Dataset.

- The Sentinel-2 NDVI (described in Section 2.1) was calculated for parcels using cadastral boundaries, with statistical descriptors extracted.
- ERA5 climatic data filled gaps in station data and supported index calculations (GDD, WI, chilling).
- The agronomic indices were calculated using various parameters: base temperatures (10 $^\circ$C, 4.5 $^\circ$C), heat thresholds (30 $^\circ$C, 40 $^\circ$C), accumulation start dates (Jan 1, February 1, dormancybased) and dormancy / growth season detection.
- The cold influence was modelled using the chilling hours [96] and the Utah model [75].
- Finally, all datasets were merged by timestamp and location to train ML models and generate phenology predictions, which are delivered via the Red FARA web interface.

The following subsections detail each transformation: normalization of phenology records from the Red FARA (Section 3.1), conversion of Sentinel-2 imagery to vineyard-level data (Section 3.2), processing of climatic data from SIAR stations, ERA5 forecasts and estimations (Section 3.3), and integration of all transformed data into the ML-ready dataset (Section 3.4).

3.1 Phenology Stages Normalization

The Phytosanitary Surveillance Network of Aragón, known as Red FARA [41], provides a platform for collecting data on crop phenology and pest presence. It follows a protocol based on the Baggiolini scale [10], defining weekly sampling, target phenological stages, pests, and field selection per PDO. Although data are available since 2010, only the period 2016–2022 was used for the modeling due to the availability of Sentinel-2 imagery, which started in 2016.

Although the sampling protocols are well established, the quality of the collected data remain inconsistent. To enhance standardization and ensure semantic interoperability, the phenological observations originally recorded using the Baggiolini scale were translated into the BBCH scale [12, 77]. This conversion aligns the dataset with the FAIR principles [78], promoting better data accessibility, transparency, and reproducibility. From the 17 grapevine phenological stages identified in the dataset, only 9 were selected due to their importance in predicting pest risks (specifically for Powdery Mildew, Mildew, Botrytis, and Lobesia Botrana). These selected stages and their relationship with the relevant pests are illustrated in Figure 3.

Each parcel is identified via its CAP code, which links to crop variety data (e.g., Cariñena, Grenache, Tempranillo, Chardonnay) from the Aragón Open Data Registry [40]. CAP codes were also converted to cadastral codes to geolocate parcels, allowing integration with climate and Sentinel-2 datasets for model training. Figure 1 shows the location of the final geo-referenced dataset.

Taking into account the considerations introduced in the previous paragraphs, for the study period 2016–2022, a total of 18,231 grapevine phenology samples were recorded in the Red FARA. Figures A.1 and A.2 (in Appendix A) provide a brief statistical overview of this data source: Figure A.1 shows the number of grapevine phenology samples per PDO and year as well as the overall total, and Figure A.2 illustrates the annual distribution of phenology samples per PDO and the total number of samples. Due to the distortion introduced by stage “99” (up to four times more frequent than stage “00”), this stage has been omitted from the plots. Although sample counts per year remain consistent after 2016, Figure A.1 reveals that stages “63”, “65”, and “68” are underrepresented, which may affect the accuracy of the model [23].

Additional issues were identified in the Red FARA dataset. Figures A.3 and A.4 (in Appendix A) highlight anomalies in the yearly phenological progression of a given parcel. Specifically, Figure A.3 shows cases where phenological stages in some parcels exhibit a “recession”, as certain recorded samples appear to jump backward between two successive observations. Conversely, Figure A.4 depicts parcels where the gap between two consecutive phenological stages exceeds two stages. In both figures, red circles mark unexpected jumps caused by a bug in the Red FARA app, which was later corrected. These issues were reported to agronomists and addressed to improve the overall quality of the final dataset.

3.2 Control Parcels' NDVI Calculation and Statistical Description

The Copernicus Sentinel-2 mission [37] consists of two twin satellites (Sentinel-2A and Sentinel-2B) that provide high-resolution multispectral imagery across Europe. Of the 13 spectral bands, the red, green, and NIR bands are the most relevant for calculating the NDVI used to monitor phenology [6, 33, 63, 82, 92]. Whilst implementing multispectral satellite imagery in non-arable crops like vineyards

Figure 3: BBCH Phenological states of interest in vineyards and the pests that can affect each of them. Photographs by Joaquín Balduque-Gil (ORCID: 0000-0001-5136-8639).

presents inherent challenges due to soil background interference [14], high-resolution sensors like Sentinel-2 have proven highly effective. Recent literature demonstrates that tracking the NDVI allows for robust canopy monitoring and reliable phenological forecasting in both olive trees and grapevines [8, 26, 27, 88].

Following Devaux et al. [26], each Sentinel-2 image was processed as follows. First, all images from tiles covering Aragón (IDs: T31TBG, T30TYM, T30TXM, T30TXL, T30TWM, T30TWL) from 2016–2022 were downloaded. In the process, images with more than 10% cloud cover were excluded to maintain data quality. Then, the NDVI was calculated using Equation 3. Control parcel areas were cropped using UTM/WGS84 polygons from the Spanish Cadastral Registry [85], resulting in images with 39 to 40000 pixels. Using the Pandas library [45], statistical descriptors (max, min, mean, median, std. dev.) of the NDVI were computed for each parcel. Level-2A products were downloaded using the Sentinelsat Python library [93], resulting in a dataset with dated NDVI records and descriptors for each control parcel.

3.3 Climatic Data Curation and Indexes Calculation

Grapevine phenology is strongly influenced by weather conditions [87], making accurate climate data essential for reliable modeling. In our work, we use data from two networks of automated weather stations: the Spanish Agroclimatic Information System for Irrigation (SIAR) [84] and stations deployed during the GRAPEVINE project [43]. The 9 SIAR stations are located within or near the PDO areas (within 15 km) and provide data since 2009, whilst the 14 GRAPEVINE stations, owned by the Government of Aragón, have been operational since 2021. Both networks record data every 30 minutes, including temperature, precipitation, humidity, wind speed, and direction.

Occasional data gaps were addressed using the ERA5 reanalysis dataset [20], accessed via the Open-Meteo API [65]. ERA5 offers global hourly weather data at 30 km resolution and is available from 1959 onward. Missing data were filled using ERA5 values matched by timestamp and location, with precipitation values halved to match a 30-minute resolution.

To enrich the dataset, several agronomic indicators were derived [80, 87], including heat accumulation indexes like GDD and Winkler. These were calculated using combinations of parameters: 2 base temperatures (T_b) of 10 $^\circ$C and 4.5 $^\circ$C were used; with/without a growth-inhibiting threshold of

40 $^\circ$C; using both daily and intra-day sampling contributions; and different accumulation start dates: January 1st, February 1st, and the onset of dormancy (the first day of autumn with a maximum temperature below 10°C). Cold-related indexes were also included, such as chilling hours [96] and the Utah model [75], calculated with varying start dates and time resolutions. Wind contribution was modeled using 8 directional sectors (cardinal points), computing weighted daily averages based on wind speed and direction.

The final climatic dataset comprises 110 variables, including timestamps, station metadata, raw climate data, derived indexes, and seasonal indicators (e.g., DOY or day of season). This dataset supports both model training and biological analysis [51, 87].

To improve prediction accuracy, time-shifted variables were added: for each raw and derived variable, values were included for each day two weeks before, one week before, and one week after the index date. This resulted in a final climatic dataset with 591 variables.

3.4 Creation of the Final Integrated Dataset for Training and Using Machine Learning Models

To train ML models, we created a unified dataset by merging the data sources introduced in Sections 3.1, 3.2, and 3.3. This integration involved combining datasets with different spatial and temporal resolutions [8], as summarized in Table 3. The goal was to improve upon classical models by leveraging a richer, multi-variable dataset [23, 89].

Control parcels were used as the spatial unit, since phenological data are recorded at that level. Their geographic descriptions, obtained from the Spanish Cadastral Registry [85], enabled linking phenological data to climate and remote sensing datasets. Specifically, the centroid of each parcel polygon was used to determine altitude and proximity to climate stations, and to associate Sentinel-2 imagery.

Climate station selection was based on proximity and data completeness: a station was linked to a parcel if it provided data for over 90% of the days in a season and had more than 90% sample coverage per day. This dynamic approach allows newer, closer stations to be used and excludes inactive ones.

Climate data, recorded every 30 minutes, was aggregated to daily values. To align with this frequency, phenological stages were extended from their first recorded date to the day before the next stage appeared. For example, if stage 15 was recorded on April 3rd and stage 63 on May 5th, all days in between were considered stage 15.

After merging phenological and climate data, the resulting dataset was further integrated with Copernicus Sentinel-2 data. Parcel polygons were used to crop NDVI images and compute statistical descriptors. Since Sentinel-2 imagery has lower temporal resolution, NDVI data were extended from the image date to the day before the next image was available, following the same approach used for phenological data.

4 Dataset Description and Usefulness

The completeness of the dataset presented in this document—in terms of phenological stages, meteorological variables (and their derivatives), and the inclusion of data derived from Copernicus Sentinel-2, makes it unique. It enables vineyard phenology studies based on classical approaches, as well as ML algorithms such as those conducted in our research [55, 56]. In addition, the availability of the code for creating the dataset supports replicating it in other locations, and consequently its potential extension to fit any intended purposes. Furthermore, given its temporal length, this dataset can be used to verify the feasibility of applying foundation models [53], liquid neural networks (LNN), or federated learning (FL) [59, 64] algorithms (e.g., it can be split by PDO) for predictive phenology modeling. Thus, this dataset can contribute to improve the knowledge of biologist, agronomist, and data scientist, as well as to demonstrate the highly valuable results that can be obtained from collaborative work. The dataset we created consists of 3 files:

- DIF Description.pdf”: provides a general overview of the dataset for users who access it independently of this paper.
- “DIF phenologicalstages.csv”: provides the mappings of the *phenologystageid* values in the main dataset to their corresponding BBCH scales.

- “DIF GrapevinePehologyDataset.csv”: contains the data of the dataset. A detailed description of its contents follows.

Each record in “DIF GrapevinePehologyDataset.csv” corresponds to the data associated with a specific vineyard parcel on a given date. Whilst Table B.1 in Appendix B.1 outlines the naming conventions adopted for the dataset fields, the full list of variable names is provided in Appendix B.2. Additionally, the dataset built is publicly available via Zenodo [54]. It is published under the Creative Commons Attribution 4.0 International license [21], ensuring open access and reuse. We anticipate that this resource will support the scientific community in advancing smart farming practices and contribute to broader research in agronomy and plant biology.

Once published in Zenodo, the dataset’s adherence to FAIR principles was evaluated using the F-UJI web application [25]. This application utilizes a web service that programmatically assesses the FAIRness of research data objects at the dataset level. This assessment is based on the FAIRsFAIR Data Object Assessment Metrics [24] and is performed using the DOI of the dataset. However, F-UJI has limitations, including sensitivity to metadata formats, reliance on machine-readable standards, and inconsistencies across evaluation frameworks [86]. Studies show that repositories may be penalized despite providing adequate information, due to mismatches between metadata exposure and tool expectations, an issue especially relevant for general-purpose repositories like Zenodo [9]. Nevertheless, Table 4 presents the evaluation results: the dataset obtained 22 out of 26 possible points (84%), with further improvements possible in the dimensions of accessibility, interoperability, and reusability. We manually checked the uncovered aspects indicated by F-UJI and no further improvement was deemed necessary (e.g., one of the uncovered points was that F-UJI was unable to detect the access level and access conditions of the data, which was already specified).

Table 4: Dataset evaluation using F-UJI.

Dataset	**Score earned**	**Fair level**
Findable	7 of 7	Advanced
Accessible	6 of 7	Moderate
Interoperable	4 of 6	Moderate
Reusable	5 of 6	Moderate

5 Discussion

One of the aspects that surprised us during our research on AI models for agriculture was the limited availability of phenological data repositories. Specifically, the number of datasets containing phenological information is remarkably low, and even more scarce for the case of grapevines. Additionally, existing datasets often lack standardized formatting, particularly at the metadata level. This poses a significant challenge for developing large-scale models for phenology prediction and smart farming applications, which rely not only on the volume of data but also on the availability of parallel datasets (datasets in several languages) and the diversity of data sources [48].

There are several reasons that make farmers and farming authorities reluctant to share their farming data. For example, farmers may be reluctant to share certain data because some practices (such as phytosanitary treatments) can be inferred and potentially affect subsidies. Moreover, such insights could benefit competitors at various scales. To build an effective global platform for agricultural data sharing, it must be neutral and governed by a diverse group of stakeholders (including governments, private entities, academia, civil society, and farmers' organizations) to ensure accountability and collective priorities. The platform should also support responsible sharing of research data, reduce duplication by leveraging existing infrastructure, enhance accessibility through digital tools, and explore crowdsourcing approaches [52]. The European common data spaces initiative seems to be a solution for the design and implementation of such platforms. In particular, the data spaces of the EOSC [35] and Agridataspace [2] are relevant for open science and agriculture.

Data spaces aim to ensure that data sharing and usage are governed by agreements among all stakeholders. On the technical side, Trusted Execution Environments (TEEs) and Federated Learning (FL) are two complementary approaches that enhance data privacy. TEEs protect sensitive data by isolating computations within secure hardware zones, even from privileged system software [38]. FL enables collaborative model training across decentralized devices without transferring local data, preserving privacy and security [70]. Although these approaches can be costly to implement, data spaces aim to offer anonymization services. Whilst precise geolocation of field observations is valuable for modeling, it also raises privacy concerns. Therefore, any dataset fields that could reveal the location of yields must be anonymized. Techniques such as random perturbation, affine transformation, and synthetic data generation can be applied to protect geospatial privacy [62].

In this paper, we have presented an open dataset that stands out when compared to existing open datasets identified in our literature review. Whilst the dataset was originally conceived to train phenology prediction models [55, 56], following our defined multidisciplinary methodology and utilizing our published code [57] it can easily be extended with new features. For instance, incorporating pest data would allow researchers to predict pest risks [11], and data from other crops could be integrated seamlessly. Consequently, we believe this dataset is highly valuable (on its own or by extending it, as required) for research in agronomy, biology, and the impacts of climate change on vegetation. Furthermore, it is of significant interest to data scientists. Beyond the opportunity to build Machine Learning models using well-established algorithms like Random Forest, XGBoost, or ANN (Artificial Neuronal Networks), it also opens the door to cutting-edge approaches such as LNN (Liquid Neural Networks) or time series foundation models, which show great promise for predicting time series such as phenology. Finally, the dataset can be split into 3 distinct subsets based on the parcels within each PDO, making it an excellent resource for developing and testing FL algorithms and infrastructure that usually is integrated on a data space.

6 Conclusions

In this paper, we have presented a rich grapevine phenology dataset. Its development involved a multidisciplinary team of agronomists, data scientists, and software engineers. Following the CRISP-DM methodology, the dataset was iteratively refined during the creation of ML-based models for grapevine phenology and pest risk prediction. The dataset integrates 9 georeferenced time-series datasets from 8 sources, covering the period 2016–2022. Although limited to three PDOs in Aragón, Spain, it encompasses more fields than any other phenology dataset identified, particularly for grapevines. It provides detailed annotations for 9 phenological stages (surpassing all other datasets analyzed) and incorporates raw climatic data and derived variables based on formulas widely used by biologists and agronomists. These variables reflect key parameters discussed in the literature, enabling ML algorithms to identify the most relevant factors whilst generating new insights in biology and agronomy. By publishing the dataset, the methodology used to construct it, and the source code used to generate it, in line with the FAIR principles, we aim to enable the reproduction of this process in other grapevine regions and for different crops.

Data Availability

The presented dataset is available in Zenodo, at https://zenodo.org/records/20124340 [54].
All the code used to generate the dataset, as well as examples of exploitation of the dataset for machine learning task is also available in Zenodo, at https://zenodo.org/doi/10.5281/zenodo.19953953 [57].

References

[1] U.S. General Services Administration. *Data.gov – U.S. Government Open Data Portal*. Last access: May 19, 2026. 2025. url: https://data.gov.

[2] AgriDataSpace Consortium. *AgriDataSpace Project*. Last access: May 19, 2026. 2025. url: https://agridataspace-csa.eu.

[3] F. A. Akyuz and Joel K. Ransom. "Growing degree day calculation method comparison between two methods in the northern edge of the US Corn Belt". In: *Journal of Service Climatology* 2015 (2015), pp. 1–9. doi: 10.46275/JoASC.2015.12.001.

[4] Laith Alzubaidi, Jinshuai Bai, Aiman Al-Sabaawi, Jose Santamar´ıa, A. S. Albahri, Bashar Sami Nayyef Al-dabbagh, Mohammed A. Fadhel, Mohamed Manoufali, Jinglan Zhang, Ali H. Al-Timemy, Ye Duan, Amjed Abdullah, Laith Farhan, Yi Lu, Ashish Gupta, Felix Albu, Amin Abbosh, and Yuantong Gu. "A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications". In: *Journal of Big Data* 10.1 (Apr. 2023). issn: 2196-1115. doi: 10.1186/s40537-023-00727-2.

[5] Maynard Amerine and Albert Winkler. "Composition and quality of musts and wines of California grapes". In: *Hilgardia* 15.6 (1944), pp. 493–675.

[6] Davoud Ashourloo, Hamed Nematollahi, Alfredo Huete, Hossein Aghighi, Mohsen Azadbakht, Hamid S. Shahrabi, and Salman Goodarzdashti. "A new phenology-based method for mapping wheat and barley using time-series of Sentinel-2 images". In: *Remote sensing of environment* 280 (Oct. 2022), 113206:1–113206:14. doi: 10.1016/j.rse.2022.113206.

[7] Anjana J. Atapattu, Lalith K. Perera, Tharindu D. Nuwarapaksha, Shashi S. Udumann, and Nuwandhya S. Dissanayaka. "Challenges in achieving Artificial Intelligence in agriculture". In: *Artificial Intelligence Techniques in Smart Agriculture*. Springer Nature Singapore, 2024, pp. 7–34. isbn: 9789819758784. doi: 10.1007/978-981-97-5878-4_2.

[8] Izar Azpiroz, Noelia Oses, Marco Quartulli, Igor G. Olaizola, Diego Guidotti, and Susanna Marchi. "Comparison of climate reanalysis and remote-sensing data for predicting olive

phenology through Machine-Learning methods". In: *Remote Sensing* 13.6 (2021), 1224:1–1224:16. issn: 2072-4292. doi: 10.3390/rs13061224.

[9] Janete Saldanha Bach, Claus-Peter Klas, Brigitte Mathiak, Yudong Zhang, and Peter Mutschke. *FAIRness assessment: A comparison of the RDA model and the F-UJI Automated tool report*. 2023. doi: 10.5281/zenodo.8308902.

[10] M. Baggiolini. "Les stades repérés dans le d´eveloppement annuel de la vigne et leur utilisation pratique". In: *Revue Romande d'Agriculture, de Viticulture et d'Arboriculture* 8.1 (1952), pp. 4–6.

[11] Joaquín Balduque-Gil, Francisco J. Lacueva-Pérez, Gorka Labata-Lezaun, Rafael del-HoyoAlonso, Sergio Ilarri, Eva S´anchez-Hern´andez, Pablo Martín-Ramos, and Juan J. BarriusoVargas. "Big Data and Machine Learning to improve European Grapevine Moth (*Lobesia botrana*) predictions". In: *Plants* 12.3 (2023), 633:1–633:16. issn: 2223-7747. doi: 10.3390/ plants12030633.

[12] Bernard Bloesch and Olivier Viret. "Stades phénologiques repères de la vigne". In: *Revue Suisse Vitic Arboric Hortic* 40.6 (2008), pp. I–IV. url: https://search.ebscohost.com/login.aspx?direct=true&AuthType=sso&db=edsdnp&AN=edsdnp.2790087ART&lang=es&site=eds-live&scope=site.

[13] Raymond Bonhomme. "Bases and limits to using degree day units". In: *European Journal of Agronomy* 13.1 (2000), pp. 1–10. doi: 10.1016/S1161-0301(00)00058-7.

[14] A. Brook, V. De Micco, G. Battipaglia, A. Erbaggio, G. Ludeno, I. Catapano, and A. Bonfante. "A smart multiple spatial and temporal resolution system to support precision agriculture from satellite images: Proof of concept on Aglianico vineyard". In: Remote Sensing of Environment 240 (2020), p. 111679. issn: 0034-4257. doi: 10.1016/j.rse.2020.111679.

[15] Ginevra Canavera, Eugenio Magnanini, Simone Lanzillotta, Claudio Malchiodi, Leonardo Cunial, and Stefano Poni. "A sensorless, Big Data based approach for phenology and meteorological drought forecasting in vineyards." In: *Scientific Reports* 13.1 (2023). issn: 2045-2322. doi: 10.1038/s41598-023-44019-4.

[16] European Commission. Joint Research Centre. *European Data Spaces: Scientific Insights into Data Sharing and Utilisation at Scale*. LU: Publications Office, 2023. doi: 10.2760/400188.

[17] P. B. Cerlini, M. Saraceni, F. Orlandi, L. Silvestri, and M. Fornaciari. "Phenological response to temperature variability and orography in Central Italy". In: *International Journal of Biometeorology* 66.1 (2022), pp. 71–86. doi: 10.1007/s00484-021-02190-1.

[18] Jorge Chamorro-Padial, Roberto Garc´ıa, and Rosa Gil. "A systematic review of open data in agriculture". In: *Computers and Electronics in Agriculture* 219 (Apr. 2024), p. 108775. issn: 0168-1699. doi: 10.1016/j.compag.2024.108775.

[19] Adriane Chapman, Elena Simperl, Laura Koesten, George Konstantinidis, Luis-Daniel Iba´n˜ez, Emilia Kacprzak, and Paul Groth. "Dataset search: a survey". In: *The VLDB Journal* 29.1 (Aug. 2019), pp. 251–272. issn: 0949-877X. doi: 10.1007/s00778-019-00564-x.

[20] Copernicus Climate Change Service (C3S). *ERA5 hourly data on single levels from 1959 to present*. Last access: May 19, 2026. 2023. url: https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview.

[21] *Creative Commons Attribution 4.0 International License*. Last access: May 19, 2026. 2026. url: https://creativecommons.org/licenses/by/4.0/deed.en.

[22] Wujun Dai, Huiying Jin, Yuhong Zhang, Tong Liu, and Zhiqiang Zhou. "Detecting temporal changes in the temperature sensitivity of spring phenology with global warming: Application of machine learning in phenological model". In: *Agricultural and Forest Meteorology* 279 (2019), 107702:1–107702:14. issn: 0168-1923. doi: 10.1016/j.agrformet.2019.107702.

[23] Wujun Dai, Huiying Jin, Lin Zhou, Tong Liu, Yuhong Zhang, Zhiqiang Zhou, Yongshuo H. Fu, and Guangze Jin. "Testing machine learning algorithms on a binary classification phenological model". In: Global Ecology and Biogeography 32.1 (2023), pp. 178–190. doi: 10.1111/geb.13612.

[24] Anusuriya Devaraju and Robert Huber. "An automated solution for measuring the progress toward FAIR research data". In: Patterns 2.11 (Nov. 2021), p. 100370. issn: 2666-3899. doi: 10.1016/j.patter.2021.100370. url: http://dx.doi.org/10.1016/j.patter.2021. 100370.

[25] Anusuriya Devaraju and Robert Huber. *F-UJI – Automated FAIR Data Assessment Tool*. Last access: May 19, 2026. 2020. doi: 10.5281/zenodo.6361400. url: https://www.f-uji.net.

[26] Nicolas Devaux, Thomas Crestey, Corentin Leroux, and Bruno Tisseyre. "Potential of Sentinel-2 satellite images to monitor vine fields grown at a territorial scale". In: OENO One 53.1 (Mar. 2019), pp. 52–59. doi: 10.20870/oeno-one.2019.53.1.2293.

[27] Lia Duarte, Ana C. Teodoro, Ant´onio T. Monteiro, M´ario Cunha, and Hernˆani Gon¸calves. "QPhenoMetrics: An open source software application to assess vegetation phenology metrics". In: Computers and Electronics in Agriculture 148 (2018), pp. 82–94. issn: 0168-1699. doi: 10.1016/j.compag.2018.03.007.

[28] Eric Duchêne. Vitis vinifera cv. Riesling developmental stages. Version V5. 2019. doi: 10.15454/GYSGNR.

[29] Florian Ehrlich-Sommer, Ferdinand Hoenigsberger, Christoph Gollob, Arne Nothdurft, Karl Stampfer, and Andreas Holzinger. "Sensors for Digital Transformation in Smart Forestry". In: Sensors 24.3 (Jan. 2024), p. 798. issn: 1424-8220. doi: 10.3390/s24030798.

[30] J. C. Eidenshink and J. L. Faundeen. "The 1 km AVHRR global land data set: first stages in implementation". In: International Journal of Remote Sensing 15.17 (Nov. 1994), pp. 3443–3462. issn: 1366-5901. doi: 10.1080/01431169408954339.

[31] Moussa El Jarroudi, Louis Kouadio, Philippe Delfosse, Clive H. Bock, Anne-Katrin Mahlein, Xavier Fettweis, Benoit Mercatoris, Frank Adams, Jillian M. Lenn´e, and Said Hamdioui."Leveraging edge artificial intelligence for sustainable agriculture". In: Nature Sustainability 7.7 (June 2024), pp. 846–854. issn: 2398-9629. doi: 10.1038/s41893-024-01352-4.

[32] Elsevier. *Mendeley Data Repository*. https://data.mendeley.com. Last access: May 19, 2026. 2025.

[33] Saeedeh Eskandari and Sajjad Ali Mahmoudi Sarab. "Mapping land cover and forest density in Zagros forests of Khuzestan province in Iran: A study based on Sentinel-2, Google Earth and field data". In: *Ecological Informatics* 70 (2022), 101727:1–101727:18. issn: 1574-9541. doi: 10.1016/j.ecoinf.2022.101727.

[34] European Centre for Medium-Range Weather Forecasts (ECMWF). *ECMWF Open Data*. Last access: May 19, 2026. 2026. url: https://www.ecmwf.int/en/forecasts/datasets/opendata.

[35] European Commission. *European Open Science Cloud - EU Node*. Last access: May 19, 2026. 2025. url: https://open-science-cloud.ec.europa.eu.

[36] Omar Garc´ıa-Tejera, Raúl Marcos-Matamoros, Boris Basile, Alessandro Mataffo, Pasquale Scognamiglio, Henar Prieto, Luis Mancha, Inˆes Cabral, Jorge Queiroz, Joana Valente, Fernando Alves, Nube González-Reviriego, Sara Hern´andez-Barrera, Mercè Mata, and Joan Girona. "A method for using monthly average temperatures in phenology models for grapevine (Vitis vinifera L.)" In: OENO One 56 (4 2022), pp. 173–182. doi: 10.20870/oeno-one.2022.56.4.5401.

[37] F. Gascon, E. G. Cadau, O. Colin, B. Hoersch, C. Isola, B. L´opez Fern´andez, and P. Martimort. "Copernicus Sentinel-2 mission: products, algorithms and Cal/Val". In: Earth Observing Systems XIX. Ed. by James J. Butler, Xiaoxiong (Jack) Xiong, and Xingfa Gu. SPIE, Sept. 2014, pp. 1–9. doi: 10.1117/12.2062260.

[38] Tim Geppert, Stefan Deml, David Sturzenegger, and Nico Ebert. "Trusted Execution Environments: Applications and organizational challenges". In: *Frontiers in Computer Science* 4 (July 2022). issn: 2624-9898. doi: 10.3389/fcomp.2022.930741.

[39] Google. *Google Dataset Search*. Last access: May 19, 2026. 2025. url: https://datasetsearch.research.google.com.

[40] Government of Aragón. *Aragón Open Data Home Page*. Last access: May 19, 2026. 2026. url: https://opendata.aragon.es.

[41] Government of Arag´on. Red FARA Home Page. Last access: May 19, 2026. 2026. url: http://web.redfara.es.

[42] W. P. Grant, D. Chandler, A. Bailey, J. Greaves, M. Tatchell, and G. Prince. Biopesticides: pest management and regulation. CABI Publishing, Aug. 2010, p. 232. isbn: 978-1845935597. doi: 10.1017/S0014479711000755.

[43] Grapevine Project Consortium. Grapevine Project Home Page. Grant agreement ID: 863463. https://grapevine-project.eu (Last access: August 8, 2024), https://web.archive.org/web/20230922054033/https://grapevine-project.eu (Last access: May 19, 2026), https://www.egi.eu/case-study/grapevine (Last access: May 19, 2026). 2022.

[44] Garima Gupta and Sudhir Kumar Pal. "Applications of AI in precision agriculture". In: Discover Agriculture 3.1 (Apr. 2025). issn: 2731-9598. doi: 10.1007/s44279-025-00220-9.

[45] Pramod Gupta and Anupam Bagchi. "Introduction to Pandas". In: *Essentials of Python for Artificial Intelligence and Machine Learning*. Springer, 2024, pp. 161–196. doi: 10.1007/9783-031-43725-0_5.

[46] Ali Hakami. “Strategies for overcoming data scarcity, imbalance & feature selection challenges in machine learning models for predictive maintenance”. In: Scientific Reports 14.1 (Apr. 2024). issn: 2045-2322. doi: 10.1038/s41598-024-59958-9.

[47] Jincai Huang, Yongjun Xu, Qi Wang, Qi (Cheems) Wang, Xingxing Liang, Fei Wang, Zhao Zhang, Wei Wei, Boxuan Zhang, Libo Huang, Jingru Chang, Liantao Ma, Ting Ma, Yuxuan Liang, Jie Zhang, Jian Guo, Xuhui Jiang, Xinxin Fan, Zhulin An, Tingting Li, Xuefei Li, Zezhi Shao, Tangwen Qian, Tao Sun, Boyu Diao, Chuanguang Yang, Chenqing Yu, Yiqing Wu, Mengxian Li, Haifeng Zhang, et al. “Foundation models and intelligent decision-making: Progress, challenges, and perspectives”. In: The Innovation 6.6 (June 2025), p. 100948. issn: 2666-6758. doi: 10.1016/j.xinn.2025.100948.

[48] Vivek Iyer, Bhavitvya Malik, Pavel Stepachev, Pinzhen Chen, Barry Haddow, and Alexandra Birch. *Quality or Quantity? On data scale and diversity in adapting Large Language Models for low-resource translation*. 2024. arXiv: 2408.12780[cs.CL]. url: https://arxiv.org/ abs/2408.12780.

[49] Putu Hadi Purnama Jati, Yi Lin, Sara Nodehi, Dwy Bagus Cahyono, and Mirjam van Reisen. “FAIR versus Open Data: A comparison of objectives and principles”. In: Data Intelligence 4.4 (2022), pp. 867–881. issn: 2641-435X. doi: 10.1162/dint_a_00176.

[50] Kaggle. *Kaggle dataset repository*. https://www.kaggle.com/datasets. Last access: May 19, 2026. 2026.

[51] Andreas Kamilaris, Andreas Kartakoullis, and Francesc X. Prenafeta-Boldú. “A review on the practice of Big Data analysis in agriculture”. In: *Computers and Electronics in Agriculture* 143 (2017), pp. 23–37. issn: 0168-1699. doi: 10.1016/j.compag.2017.09.037.

[52] Endalkachew Abebe Kebede, Hanan Abou Ali, Tyler Clavelle, Halley E. Froehlich, Jessica A. Gephart, Sarah Hartman, Mario Herrero, Hannah Kerner, Piyush Mehta, Catherine Nakalembe, Deepak K. Ray, Stefan Siebert, Philip Thornton, and Kyle Frankel Davis. “Assessing and addressing the global state of food production data scarcity”. In: Nature Reviews Earth & Environment 5.4 (Feb. 2024), pp. 295–311. issn: 2662-138X. doi: 10.1038/s43017024-00516-2.

[53] Siva Rama Krishna Kottapalli, Karthik Hubli, Sandeep Chandrashekhara, Garima Jain, Sunayana Hubli, Gayathri Botla, and Ramesh Doddaiah. Foundation Models for Time Series: A Survey. 2025. doi: 10.48550/arXiv.2504.04011. url: https://arxiv.org/abs/2504.04011.

[54] Francisco José Lacueva-Pérez, Rafael del Hoyo-Alonso, Gorka Labata-Lezaún, Juan José Barriuso-Vargas, and Sergio Ilarri-Artigas. *A multisource grapevine phenology dataset for smart farming and AI modeling*. Dataset on Zenodo. 2026. doi: 10.5281/zenodo.20124340.

[55] Francisco José Lacueva-Pérez, Rafael del Hoyo-Alonso, Gorka Labata-Lezaún, Juan José Barriuso-Vargas, and Sergio Ilarri-Artigas. “Developing machine learning models from multisourced real-world datasets to enhance smart-farming practices”. In: Computers and Electronics in Agriculture 231 (Apr. 2025), p. 110018. issn: 0168-1699. doi: 10.1016/j.compag. 2025.110018.

[56] Francisco José Lacueva-Pérez, Sergio Ilarri, Juan José Barriuso-Vargas, Joaquín Balduque, Gorka Labata-Lezaun, and Rafael del Hoyo-Alonso. “Grapevine phenology prediction: A comparison of physical and Machine Learning models”. In: *24th International Conference on*

Big Data Analytics and Knowledge Discovery (DaWaK). Vol. 13428. Lecture Notes in Computer Science (LNCS). Springer, ISSN 0302-9743, ISSN 1611-3349 (electronic), Print ISBN 978-3-031-12669-7, Online ISBN 978-3-031-12670-3, Aug. 2022, pp. 263–269. doi: 10.1007/978-3-031-12670-3_24.

[57] Francisco José Lacueva-Pérez, Gorka Labata Lezaun, Rafael del Hoyo-Alonso, Juan BarriusoVargas, and Sergio Ilarri-Artigas. *Grapevine Phenology Code: Creating a Dataset to Train Grapevine Phenology Models*. 2026. doi: 10.5281/zenodo.19953953. url: https://zenodo.org/doi/10.5281/zenodo.19953953.

[58] Francisco José Lacueva-Pérez., Sergio Ilarri., Juan José Barriuso Vargas., Gorka Labata Lezaun., and Rafael Del Hoyo Alonso. "Multifactorial evolutionary prediction of phenology and pests: Can Machine Learning help?" In: *16th International Conference on Web Information Systems and Technologies (WEBIST 2020)*. INSTICC. SciTePress, 2020, pp. 75–82. isbn: 978-989-758-478-7. doi: 10.5220/0010132900750082.

[59] Qinbin Li, Zeyi Wen, Zhaomin Wu, Sixu Hu, Naibo Wang, Yuan Li, Xu Liu, and Bingsheng He. "A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection". In: IEEE Transactions on Knowledge and Data Engineering 35.4 (Apr. 2023), pp. 3347–3366. issn: 2326-3865. doi: 10.1109/tkde.2021.3124599. url: http://dx.doi.org/10.1109/TKDE.2021.3124599.

[60] Wanjing Li, Qinchuan Xin, Xuewen Zhou, Zhicheng Zhang, and Yongjian Ruan. "Comparisons of numerical phenology models and machine learning methods on predicting the spring onset of natural vegetation across the Northern Hemisphere". In: *Ecological Indicators* 131 (2021), p. 108126. issn: 1470-160X. doi: 10.1016/j.ecolind.2021.108126.

[61] C. Llanaj and G. McGregor. "Climate Change, grape phenology, and frost risk in Southeast England". In: *Australian Journal of Grape and Wine Research* 2022.1 (2022), p. 9835317. doi: 10.1155/2022/9835317.

[62] Mahrokh Abdollahi Lorestani, Thilina Ranbaduge, and Thierry Rakotoarivelo. "Privacy risk in GeoData: A survey". In: *arXiv preprint arXiv:2402.03612* (2024). arXiv: 2402.03612 [cs.CR]. url: https://arxiv.org/abs/2402.03612.

[63] Sabelo Madonsela, Moses A. Cho, Laven Naidoo, Russell Main, and Nobuhle Majozi. "Exploring the utility of Sentinel-2 for estimating maize chlorophyll content and leaf area index across different growth stages". In: Journal of Spatial Science 68.2 (2023), pp. 339–351. doi: 10.1080/14498596.2021.2000898.

[64] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. "Communication-Efficient Learning of Deep Networks from Decentralized Data". In: (2016). doi: 10.48550/arxiv.1602.05629. url: https://arxiv.org/abs/1602.05629.

[65] Open Meteo. Open Meteo Historical Weather API. Last access: May 19, 2026. 2026. url: http://open-meteo.com/en/docs/historical-weather-api.

[66] OpenStreetMap contributors. *Open Street Maps Home Page*. Last access: May 19, 2026. 2026. url: http://openstreetmap.org.

[67] Noelia Oses, Izar Azpiroz, Susanna Marchi, Diego Guidotti, Marco Quartulli, and Igor G. Olaizola. “Analysis of Copernicus’ ERA5 climate reanalysis data as a replacement for weather station temperature measurements in Machine Learning models for olive phenology phase prediction”. In: *Sensors* 20.21 (2020). issn: 1424-8220. doi: 10.3390/s20216381.

[68] L.E. Parker, N. Zhang, and J.T. Abatzoglou. Data from: A variety-specific analysis of Climate Change effects on California winegrapes. DOI of the associated paper: 10.1007/s00484-02402684-8. 2024. url: https://catalog.data.gov/dataset/data-from-a-varietyspecific-analysis-of-climate-change-effects-on-california-winegrapes.

[69] Lauren E. Parker, Ning Zhang, John T. Abatzoglou, Isaya Kisekka, Andrew J. McElrone, and Steven M. Ostoja. “A variety-specific analysis of climate change effects on California winegrapes”. In: *International Journal of Biometeorology* 68.8 (Apr. 2024), pp. 1559–1571. issn: 1432-1254. doi: 10.1007/s00484-024-02684-8.

[70] Jiaming Pei, Wenxuan Liu, Jinhai Li, Lukun Wang, and Chao Liu. “A review of Federated Learning methods in heterogeneous scenarios”. In: *IEEE Transactions on Consumer Electronics* 70.3 (Aug. 2024), pp. 5983–5999. issn: 1558-4127. doi: 10.1109/tce.2024.3385440.

[71] Anne Pellegrino, Deidre Blackmore, Peter Clingeleffer, and Rob Walker. “Comparison of methods for determining budburst date in grapevine”. In: OENO One 56.1 (2022), pp. 73–86. doi: 10.20870/oeno-one.2022.56.1.4751.

[72] Phenology: An integrative environmental science. Springer Nature Switzerland, 2024. isbn: 9783031750274. doi: 10.1007/978-3-031-75027-4.

[73] European Space Agency - Copernicus Programme. *Copernicus data space ecosystem*. Last access: May 19, 2026. 2025. url: https://www.sentinel-hub.com/explore/copernicusdata-space-ecosystem.

[74] Samuel Reis, Helder Fraga, Cristina Carlos, José Silvestre, José Eiras-Dias, Pedro Rodrigues, and Joao A. Santos. “Grapevine phenology in four Portuguese wine regions: Modeling and predictions”. In: *Applied Sciences* 10.11 (2020). issn: 2076-3417. doi: 10.3390/app10113708.

[75] E. A Richardson. “A model for estimating the completion of rest for ‘Redhaven’ and ‘Elberta’ peach trees”. In: *HortScience* 9 (1974), pp. 331–332.

[76] Pedro Rodrigues, Vanda Pedroso, Samuel Reis, Chenyao Yang, and João A. Santos. “Climate Change impacts on phenology and ripening of cv. Touriga Nacional in the Da˜o wine region, Portugal”. In: International Journal of Climatology 42.14 (2022), pp. 7117–7132. doi: 10.1002/joc.7633.

[77] Catherine Roussey, Xavier Delpuech, Florence Amardeilh, Stephan Bernard, and Clement Jonquet. “Semantic description of plant phenological development stages, starting with grapevine”. In: *Metadata and Semantic Research*. Ed. by Emmanouel Garoufallou and María Antonia Ovalle-Perandones. Springer, 2021, pp. 257–268. doi: 10.1007/978-3-030-719036_25.

[78] Stefania Savoi, Panagiotis Arapitsas, Eric Duchêne, Maria Nikolantonaki, Ignacio Ontanón, Silvia Carlin, Florian Schwander, Régis D. Gougeon, Antonio C. S. Ferreira, Georgios Theodoridis, Reinhard To¨pfer, Urska Vrhovsek, Anne-Francoise Adam-Blondon, Mario Pezzotti, and Fulvio Mattivi. “Grapevine and wine metabolomics-based guidelines for FAIR

data and metadata management”. In: *Metabolites* 11.11 (Nov. 2021), p. 757. doi: 10.3390/metabo11110757.

[79] Ashraf Shaharudin, Bastiaan van Loenen, and Marijn Janssen. “Towards a common definition of Open Data intermediaries”. In: *Digital Government: Research and Practice* 4.2 (June 2023), pp. 1–21. issn: 2639-0175. doi: 10.1145/3585537.

[80] Vivek Sharma, Ashish Kumar Tripathi, and Himanshu Mittal. “Technological revolutions in smart farming: Current trends, challenges and future directions”. In: Computers and Electronics in Agriculture 201 (2022), p. 107217. issn: 0168-1699. doi: 10.1016/j.compag.2022.107217.

[81] Sociedad Aragonesa de Gestión Agroambiental (SARGA). *SARGA Home Page*. Last access: May 19, 2026. 2026. url: https://www.sarga.es.

[82] Jesús Soriano-González, Eduard Angelats, Maite Martínez-Eixarch, and Carles Alcaraz. “Monitoring rice crop and yield estimation with Sentinel-2 data”. In: *Field Crops Research* 281 (May 2022), 108507:1–108507:11. doi: 10.1016/j.fcr.2022.108507.

[83] Sumit Sow, Shivani Ranjan, Mahmoud F. Seleiman, Hiba M. Alkharabsheh, Mukesh Kumar, Navnit Kumar, Smruti Ranjan Padhan, Dhirendra Kumar Roy, Dibyajyoti Nath, Harun Gitari, and Daniel O. Wasonga. “Artificial Intelligence for maximizing agricultural input Use efficiency: Exploring nutrient, water and weed management strategies”. In: Phyton – International Journal of Experimental Botany 93.7 (2024), pp. 1569–1598. issn: 1851-5657. doi: 10.32604/phyton.2024.052241.

[84] Spanish Ministry of Agriculture, Fisheries and Food. *Agro-climatic Information System for Irrigation (SIAR) Home Page*. Last access: January 10, 2026. 2026. url: https://eportal.mapa.gob.es//websiar/Inicio.aspx.

[85] Spanish Treasury. *Spanish Cadastral Registry Electronic Home Page*. Last access: May 19, 2026. 2026. url: https://www.sedecatastro.gob.es.

[86] Chang Sun, Vincent Emonet, and Michel Dumontier. “A comprehensive comparison of automated FAIRness Evaluation Tools”. English. In: *Semantic Web Applications and Tools for Health Care and Life Sciences*. Vol. 3127. CEUR Workshop Proceedings. CEUR-WS, Jan. 2022.

[87] Jianwu Tang, Christian Ko¨rner, Hiroyuki Muraoka, Shilong Piao, Miaogen Shen, Stephen J. Thackeray, and Xi Yang. “Emerging opportunities and challenges in phenology: a review”. In: *Ecosphere* 7.8 (2016), e01436:1–e01436:17. doi: 10.1002/ecs2.1436.

[88] Dimitrios Tassopoulos, Dionissios Kalivas, Rigas Giovos, Nestor Lougkos, and Anastasia Priovolou. “Sentinel-2 imagery monitoring vine growth related to topography in a Protected Designation of Origin region”. In: *Agriculture* 11.8 (2021). issn: 2077-0472.

[89] Shawn D. Taylor and Ethan P. White. “Automated data-intensive forecasting of plant phenology throughout the United States”. In: *Ecological Applications* 30.1 (2020), e02025. doi: 10.1002/eap.2025.

[90] Henri E. Z. Tonnang, Ritter A. Guimapi, Anani Y. Bruce, Dan Makumbi, Bester T. Mudereri, Tesfaye Balemi, and Peter Craufurd. “PPMaP: Reproducible and extensible Open-Source software for Plant Phenological phase duration prediction and MaPping in Sub-Saharan Africa”. In: Agriculture 10.11 (2020). issn: 2077-0472. doi: 10.3390/agriculture10110515.

[91] S.E. Uribe, J. Issa, F. Sohrabniya, A. Denny, N.N. Kim, A.F. Dayo, A. Chaurasia, A. Sofi-Mahmudi, M. Bu¨ttner, and F. Schwendicke. "Publicly available dental image datasets for Artificial Intelligence". In: Journal of Dental Research 103.13 (Oct. 2024), pp. 1365–1374. issn: 1544-0591. doi: 10.1177/00220345241272052.

[92] Asier Uribeetxebarria, Ander Castellón, and Ana Aizpurua. "A first approach to determine if it is possible to delineate in-season N fertilization maps for wheat using NDVI derived from Sentinel-2". In: Remote Sensing 14.12 (2022), 2872:1–2872:19. issn: 2072-4292. doi: 10.3390/rs14122872.

[93] M. Valgur, M. Willer, Kernsten, T. Young-Audet, et al. *Sentinelsat Documentation*. Last access: May 19, 2026. 2023. url: https://sentinelsat.readthedocs.io/en/stable/.

[94] Daniel Wallach, Taru Palosuo, Peter Thorburn, Henrike Mielenz, Samuel Buis, Zvi Hochman,Emmanuelle Gourdain, Fety Andrianasolo, Benjamin Dumont, Roberto Ferrise, ThomasGaiser, Cecile Garcia, Sebastian Gayler, Matthew Harrison, Santosh Hiremath, Heidi Horan, Gerrit Hoogenboom, Per-Erik Jansson, Qi Jing, Eric Justes, Kurt-Christian Kersebaum, Marie Launay, Elisabet Lewan, Ke Liu, Fasil Mequanint, Marco Moriondo, Claas Nendel, Gloria Padovan, Budong Qian, Niels Schu¨tze, et al. "Proposal and extensive test of a calibration protocol for crop phenology models". In: Agronomy for Sustainable Development 43.4 (Aug. 2023), pp. 1–14. doi: 10.1007/s13593-023-00900-0.

[95] Yafeng Wang, Bradley Case, Sergio Rossi, Binod Dawadi, Eryuan Liang, and Aaron M. Ellison. "Frost controls spring phenology of juvenile Smith fir along elevational gradients on the Southeastern Tibetan Plateau". In: *International Journal of Biometeorology* 63.7 (Feb. 2019), pp. 963–972. doi: 10.1007/s00484-019-01710-4.

[96] John H. Weinberger. "Chilling requirements of peach varieties". In: *Proceedings of the American Society for Horticultural Science*. Vol. 56. 1950, pp. 122–128.

[97] Steven Euijong Whang, Yuji Roh, Hwanjun Song, and Jae-Gil Lee. "Data collection and quality challenges in deep learning: a data-centric AI perspective". In: The VLDB Journal 32.4 (Jan. 2023), pp. 791–813. issn: 0949-877X. doi: 10.1007/s00778-022-00775-9.

[98] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Merc`e Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, et al. "The FAIR guiding principles for scientific data management and stewardship". In: Scientific Data 3.1 (Mar. 2016). issn: 2052-4463. doi: 10.1038/sdata.2016.18.

[99] Rüdiger Wirth and Jochen Hipp. "CRISP-DM: Towards a standard process model for data mining". In: *Fourth International Conference on the Practical Applications of Knowledge Discovery and Data Mining*. Vol. 1. Manchester, 2000, pp. 29–40.

[100] Shuolei Yin, Yejing Xi, Xun Zhang, Chengnuo Sun, and Qirong Mao. "Foundation models in agriculture: A comprehensive review". In: *Agriculture* 15.8 (Apr. 2025), p. 847. issn: 2077-0472. doi: 10.3390/agriculture15080847.

[101] Zenodo. Zenodo research data repository. Last access: May 19, 2026. 2025. url: https://zenodo.org.

Author Contributions

F.L.: Writing – original draft, data curation, software, visualization, methodology; R.H.: project administration, funding acquisition; G.L.: software, data curation; J.B.: supervision; S.I.: Writing – original draft, Writing – review & editing, methodology, conceptualization, project administration, funding acquisition.

Competing interests

The authors declare no conflict of interest.

Acknowledgments

This research was partially funded by the TITAN-EOSC project, co-financed by the European Union's Horizon Europe research and innovation program under Grant Agreement No. 101129822. This publication is also part of the project PID2020-113037RB-I00 (NEAT-AMBIENCE project), funded by MICIU/AEI/10.13039/501100011033. Within the NEAT-AMBIENCE project, led by Sergio Ilarri, we consider as a use case the development of data management techniques for agriculture. Besides the NEAT-AMBIENCE project, we also thank the support of the Departamento de Ciencia, Universidad y Sociedad del Conocimiento del Gobierno de Aragón (Department of Science, University and Knowledge Society of the Government of Aragon), and more specifically to the following research groups: COSMOS, with reference T64 23R, and IODIDE, with reference T17 23R.

Appendix A Dataset Descriptive Statistics

In this appendix, we provide a brief statistical description of the data contained in the dataset presented. First, in Appendix A.1, we show the annual distribution of the considered samples and the agricultural plots where they were collected. Next, in Appendix A.2, the distribution of samples for each phenological stage is presented, both in total and for each of the three PDOs considered. Finally, in Appendix A.3, we highlight some of the errors detected in the data that affected their quality.

Appendix A.1 Distributions in the Red FARA Dataset

Figure A.1 presents the annual distribution by PDO, together with the total number of phenological observations and the number of agricultural plots for each of the campaigns (2016-2022) included in the dataset described in this article.

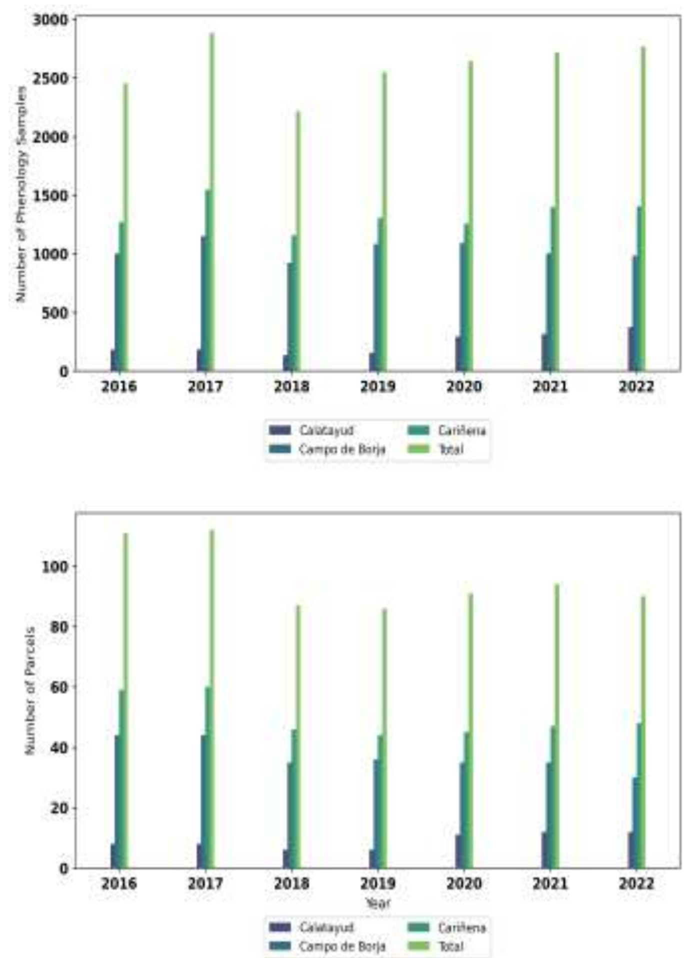

Figure A.1: Number of grapevine phenology samples and control parcels per year and PDO in the Red FARA.

Appendix A.2 Phenology Samples per Year and PDO

For each considered year (2016-2022), Figure A.2 shows, disaggregated by PDO and for the total aggregate, the distribution by phenological stage of the phenological observations included in the dataset described in this paper.

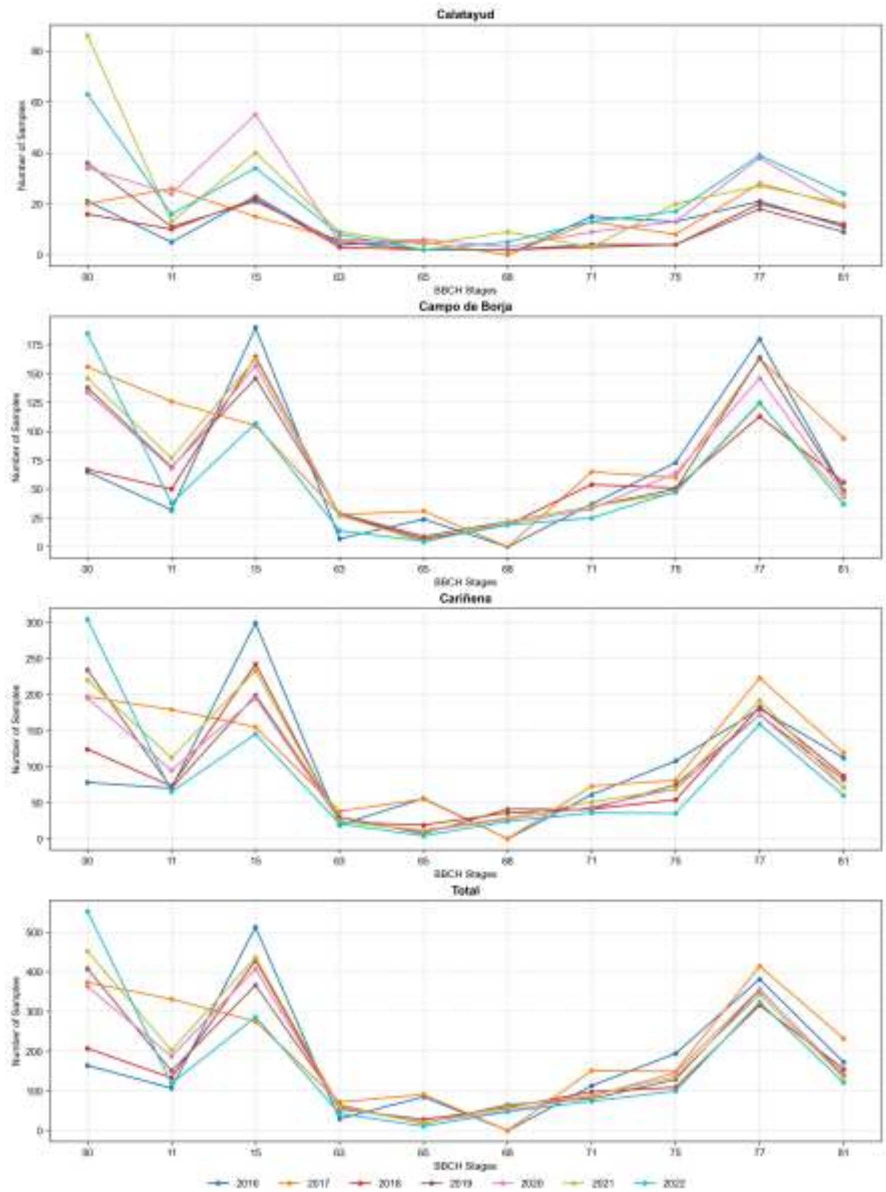

Figure A.2: Number of BBCH stage samples per year and PDO in the Red FARA.

Appendix A.3 Phenology Field Observations Quality Issues

During quality assessment, phenological "backward jumps" or "recessions" were detected in several plots (see Figure A.3), which means that certain recorded samples appear to jump backward in the phenological state between two successive observations. Additionally, Figure A.4 shows cases where gaps between consecutive stages exceed two levels. In both figures, red circles highlight unexpected jumps caused by a previous bug in the Red FARA app (now solved). These issues were reported to agronomists and addressed to refine the final dataset's quality.

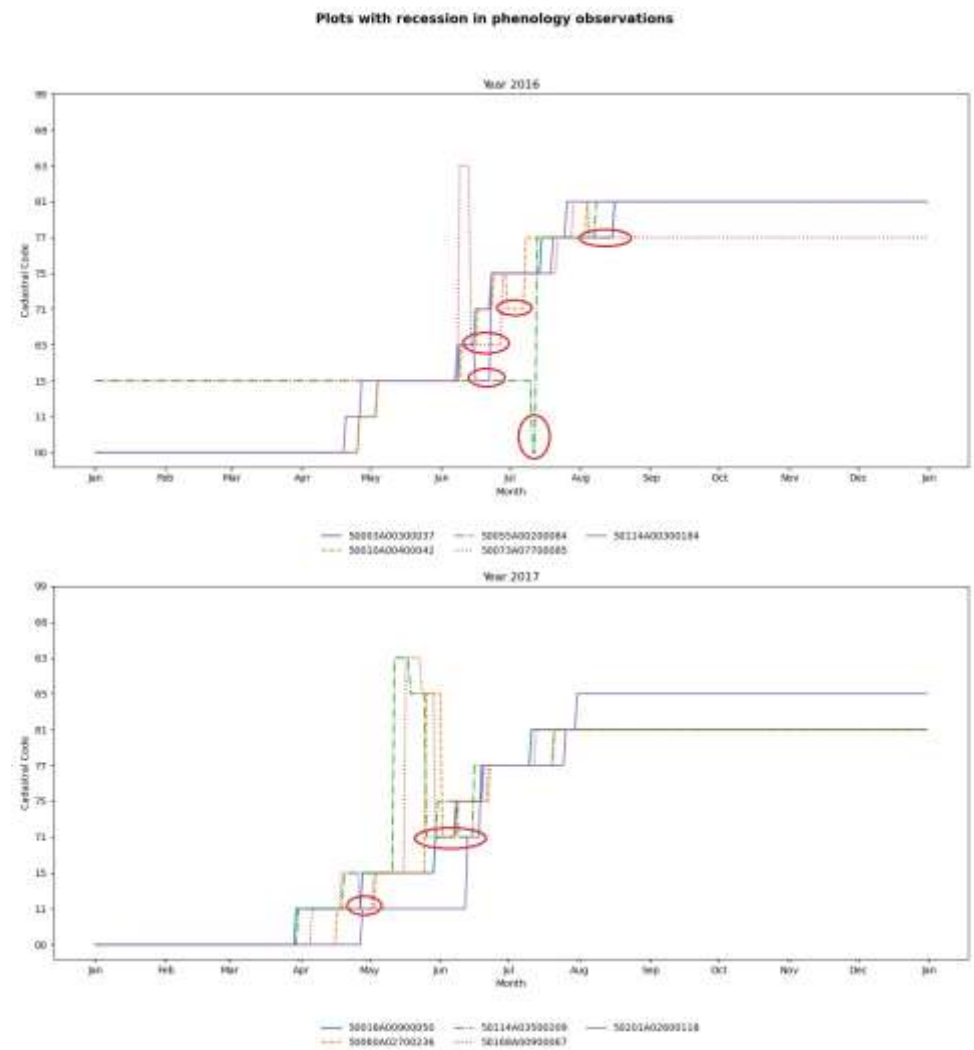

Figure A.3: Examples of parcels illustrating phenological stages exhibiting phenology "recession" (highlighted by red ellipses).

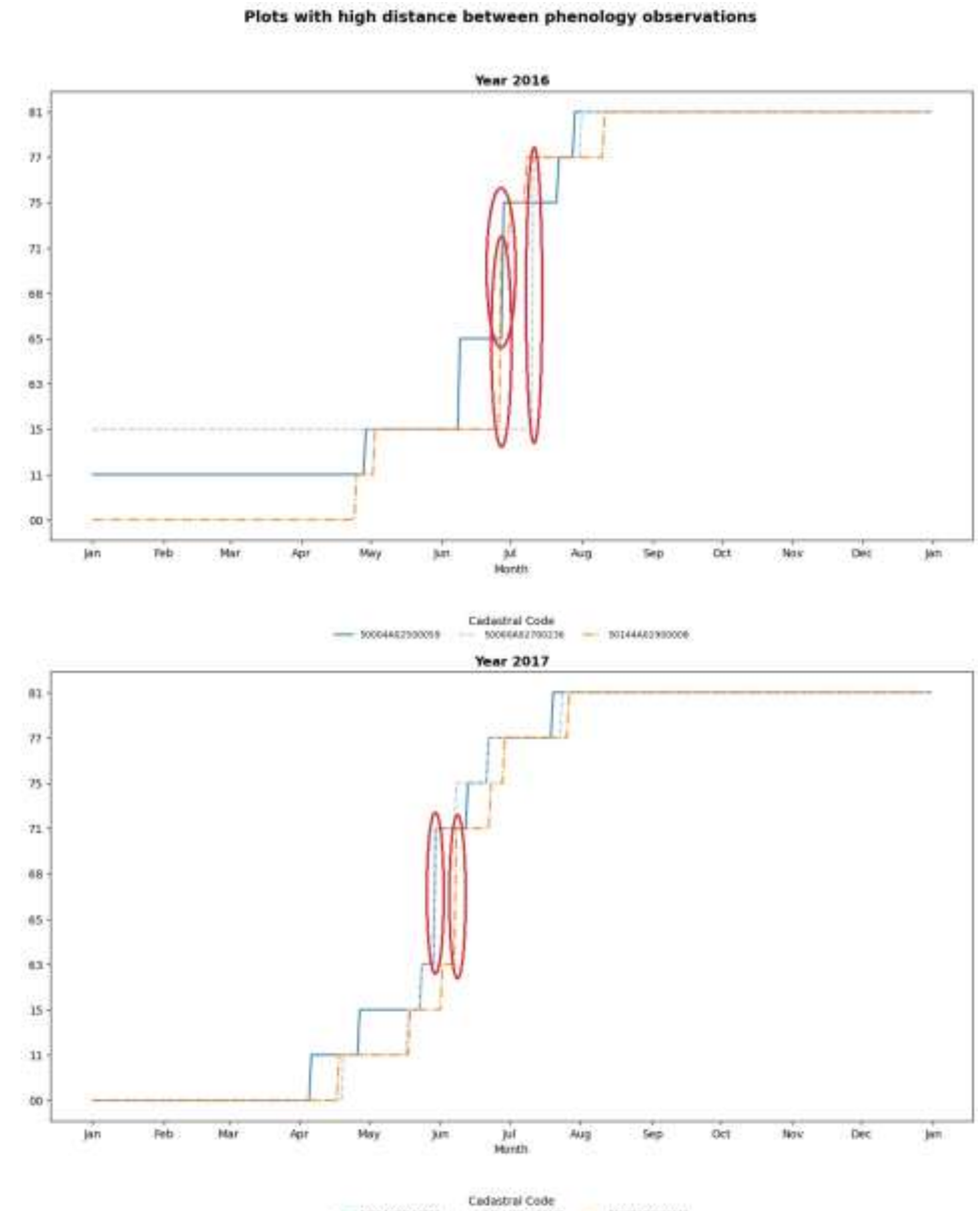

Figure A.4: Examples of parcels illustrating samples "jumping" more than 2 phenological stages (highlighted by red ellipses).

Appendix B Fields in the Dataset

In the following, we first describe some field naming conventions (Appendix B.1) and then the complete list of fields in the dataset (Appendix B.2).

Appendix B.1 Field Naming Conventions

Here, we provide a description of the field naming conventions we use in the file ''DIF_GrapevinePehologyDataset.csv'', which is part of the published dataset. Each of the records in this file corresponds to the data associated with a specific parcel (vineyard) on a given date. Table B.1 provides a description of the fields included in the dataset. For clarity, we have simplified the table by using an abbreviated notation for field names. Specifically, some field names include an asterisk ("*") followed by a range of integers in brackets ([...]), indicating that the dataset contains multiple fields where the asterisk is replaced by each integer in the range. This notation is used for variables that provide values for a number of days before ("*days before*") or after ("*days after*") the date of the record, as well as for a number of weeks before ("*weeks before*") or after ("*weeks after*"). Below are some examples:

- tmed min days before [1,13]: This notation indicates that the dataset includes the fields "*tmed min 1 days before*", "*tmed min 2 days before*", ..., "*tmed min 13 days before*", each representing the minimum temperature for each of the corresponding number of days prior to the record date.
- wind NE * days after [1,6]: This naming pattern indicates that the dataset includes the fields "*wind NE 1 days after*", "*wind NE 2 days after*", ..., "*wind NE 6 days after*", each representing the wind NE index for each of the corresponding number of days following the record date.
- gdd 4.5 t0 Tbase sum * weeks before [1,2]: This notation indicates that the dataset includes the fields '*gdd 4.5 t0 Tbase sum 1 weeks before*" and '*gdd 4.5 t0 Tbase sum 2 weeks before*", which represent the GDD values calculated using a base temperature of 4.5°C, starting from the beginning of the season.

Additionally, we use the symbol "—" to denote multiple options within a field name, enclosed in brackets ("*[...]*"). For example, "*rad_[min—MAX—mean]*" is a compact representation of three separate variables: "*rad_min*, *rad MAX*, and *rad mean*". Similar notations are used throughout the dataset. Table B.1 shows the summary of variables following the coding we defined in previous paragraphs. The complete list of variable names is provided in Appendix B.2.

Table B.1: Description of the patterns used to name the dataset columns.

Field Name (abbreviated notation) & Description
phenologystageid: Id of the phenological stage of the parcel on the given date. See file "DIF_phenologicalstages.csv".
variety: Grapevine variety: Cabernet Sauvignon, Chardonnay, Garnacha, Mazuela, Syrach, Tempranillo.
cadastralcode: Id of the parcel in the Spanish Cadastral Registry.

Field Name (abbreviated notation) & Description
longitude: Longitude of the centroid of the parcel.
latitude: Latitude of the centroid of the parcel.
altitudeASL: Altitude above the sea level of the centroid of the parcel.
Field Name (abbreviated notation) & Description
PDO id: Id of the Protected Designation of Origin (PDO): Calatayud, Carinena (Cariñena) and Campo de Borja.
date: The date of the record.
station: The name of the climatic station whose data are considered.
season: The season to which the record belongs.
day: The day of the year (DOY).
PDO _Borja, PDO _Calatayud, PDO Carinena, PDO Somontano: Boolean values; each value is true when the record corresponds to the given PDO.
variety _CABERNET SAUVIGNON, variety CHARDONNAY, variety GARNACHA, variety MAZUELA, variety SYRACH, variety TEMPRANILLO: Boolean values; each value is true when the record corresponds to a field with the given variety.
min, MAX, mean, std, median, diff : Values derived from the NDVI indexes calculated for each parcel from the Copernicus Sentinel 2 multispectral images. They represent the minimum, maximum, average, standard deviation, median and difference values.
tmed [min—MAX—mean]: [Minimum—Maximum—Mean] for the given date.
tmed [min—MAX—mean] * days before [1,13]: [Minimum—Maximum—Mean] temperatures for each of the 13 days before the given date.
rad _[min—MAX—mean]: [Minimum—Maximum—Mean] radiation for the given date (W/m^2).
rad [min—MAX—mean] * _days before [1,13]: [Minimum—Maximum—Mean] radiation for each of the 13 days before the given date.
rad [min—MAX—mean] * _days after [1,6]: [Minimum—Maximum—Mean] radiation for each of the 6 days following the given date.
hr _ mean: Average air relative humidity for the given date (%).
hr mean * days before [1,13]: Average air relative humidity radiation for each of the 13 days before the given date.
hr mean * days after [1,6]: Average air relative humidity radiation for each of the 6 days following the given date.
wind [N—NE—E—SE—S—SW—W—NW] * days before [1,13]: Wind index for the North, North-East, East, South-East, South, South-West, West, North-West area for each of the 13 days before the given date.
Wind _[N—NE—E—SE—S—SW—W—NW] * _days after [1,6]: Wind index for the North, North-East, East, South-East, South, South-West, West, North-West area for each of the 6 days following the given date.

Field Name (abbreviated notation) & Description
gdd _[4.5—10.0] _[t0—1—2] _[TBase—TbaseMAX] _sum: GDD heat accumulation index, calculated with a base temperature of 4.5 ◦C or 10.0 ◦C; accumulated since the beginning of the season (t_0), January the 1st of the date's year (t_1) or February the 1st (t_2); considering a maximum temperature threshold over which the heat accumulation stopped (T_{bM}, 35 ◦C) or not (T_b); of the given day (sum).
*gdd _[4.5—10.0] [t0—1—2] _[TBase—TbaseMAX] sum * weeks _before [1—2]*: GDD heat accumulation index, calculated with a base temperature of 4.5 ◦C or 10.0 ◦C; accumulated since the beginning of the season (t_0), January the 1st of the date's year (t_1) or February the 1st (t_2); considering a maximum temperature threshold over which the heat accumulation stopped (T_{bM}, 35 ◦C) or not (T_b); and for each of the 2 weeks previous to the given day.
*gdd [4.5—10.0] _[t0—1—2] _[TBase—TbaseMAX] sum * 1 weeks after*: GDD heat accumulation index, calculated with a base temperature of 4.5 ◦C or 10.0 ◦C; accumulated since the beginning of the season (t_0), January the 1st of the date's year (t_1) or February the 1st (t_2); considering a maximum temperature threshold over which the heat accumulation stopped (T_{bM}, 35 ◦C) or not (T_b); and considering the daily contribution of the given day (sum) for the next week to the given day.
*hr mean * days after [1,6]*: Average air relative humidity radiation for each of the 6 days following the given date.
ChillingDD 7.0 [t0—1—2] [TBase—Tbasemin] _sum: Richardson cold accumulation index, calculated with a base temperature of 7.0 ◦C; accumulated since the beginning of the season (t_0), January the 1st of the date's year (t_1) or February the 1st (t_2); considering a minimum temperature threshold above which the cold accumulation stopped (T_{bm}, -7 ◦C) or not (T_b); of the given day (sum).
ChillingDD 7.0 _[t0—1—2] _[TBase—Tbasemin] sum * weeks before [1—2]: Richardson cold accumulation index, calculated with a base temperature of 7.0 ◦C; accumulated since the beginning of the season (t_0), January the 1st of the date's year (t_1) or February the 1st (t_2); considering a minimum temperature threshold above which the cold accumulation stopped (T_{bm}, -7 ◦C) or not (T_b); and considering the daily contribution (sum) for each of the 2 weeks previous to the given day.
ChillingDD 7.0 _[t0—1—2] [TBase—Tbasemin]_ sum _ days after [1,6]*: Richardson cold accumulation index, calculated with a base temperature of 7.0 ◦C; accumulated since the beginning of the season (t_0), January the 1st of the date's year (t_1) or February the 1st (t_2); considering a minimum temperature threshold above which the cold accumulation stopped (T_{bm}, -7 ◦C) or not (T_b); and considering the daily contribution (sum). [Minimum—Maximum—Mean] temperatures for each of the 6 days following the given date.

Field Name (abbreviated notation) & Description
*ChillingDD 7.0 [t0—1—2] [TBase—Tbasemin] sum * weeks before [1—2]*: Richardson cold accumulation index, calculated with a base temperature of 7.0 ° C; accumulated since the beginning of the season (t_0), January the 1^{st} of the date's year (t_1) or February the 1^{st} (t_2); considering a minimum temperature threshold above which the cold accumulation stopped (T_{bm}, -7 ° C) or not (T_b); and considering the daily contribution (sum) for each of the 2 weeks previous to the given day.
*ChillingDD 7.0 [t0—1—2] _ [TBase—Tbasemin] sum * 1 weeks _ after*: Richardson cold accumulation index, calculated with a base temperature of 7.0 ° C; accumulated since the beginning of the season (t_0), January the 1^{st} of the date's year (t_1) or February the 1^{st} (t_2); considering a minimum temperature threshold above which the cold accumulation stopped (T_{bm}, -7 ° C) or not (T_b); and considering the daily contribution (sum) for the next week to the given day.
ChillingDD 7.0 _[t0—1—2] _ Utah sum: Utah cold accumulation index, calculated with a base temperature of 7.0 ° C; accumulated since the beginning of the season (t_0), January the 1^{st} of the date's year (t_1) or February the 1^{st} (t_2); considering a minimum temperature threshold above which the cold accumulation stopped (T_{bm}, -7 ° C) or not (T_b); and considering the daily contribution (sum).
*ChillingDD 7.0 [t0—1—2] _ Utah sum * weeks _before [1—2]*: Richardson cold accumulation index, calculated with a base temperature of 7.0º C; accumulated since the beginning of the season (t_0), January the 1^{st} of the date's year (t_1) or February the 1^{st} (t_2); considering a minimum temperature threshold above which the cold accumulation stopped (Tbasemin, -7ºC) or not (T_b); and considering the daily contribution (sum) for each of the 2 weeks previous to the given day.
*ChillingDD 7.0 _[t0—1—2] Utah sum * 1 weeks _after*: Richardson cold accumulation index, calculated with a base temperature of 7.0 ° C; accumulated since the beginning of the season (t_0), January the 1^{st} of the date's year (t_1) or February the 1^{st} (t_2); considering a minimum temperature threshold above which the cold accumulation stopped (T_{bm}, -7 ° C) or not (T_b); and considering the daily contribution (sum) for the next week to the given day.
rad sum: Accumulated radiation since the beginning of the season until the given date.
*rad sum * weeks before [1—2]*: Accumulated radiation since the beginning of the season until 1 or 2 weeks before the given date.
rad sum 1 weeks _after: Accumulated radiation since the beginning of the season until the next week after the given date.
precip _sum 1 weeks after: Accumulated precipitation since the beginning of the season until the next week after the given date.
*precip _sum * weeks _before [1—2]*: Accumulated precipitation since the beginning of the season until 1 or 2 weeks before the given date.

Field Name (abbreviated notation) & Description
winkler _[4.5—10.0] _[t0—1—2] _[TBase—TbaseMAX] sum: Winkler heat accumulation index, calculated with a base temperature of 4.5 $^\circ$C or 10.0 $^\circ$C; accumulated since the beginning of the season (t_0), January the 1st of the date's year (t_1) or February the 1st (t_2); considering a maximum temperature threshold over which the heat accumulation stopped (T_{bM}, 35 $^\circ$C) or not (T_b); and considering the daily contribution (sum).
*winkler _[4.5—10.0] _[t0—1—2] _[TBase—TbaseMAX] sum * weeks before [1—2]*: Winkler heat accumulation index, calculated with a base temperature of 4.5 $^\circ$C or 10.0 $^\circ$C; accumulated since the beginning of the season (t_0), January the 1st of the date's year (t_1) or February the 1st (t_2); considering a maximum temperature threshold over which the heat accumulation stopped (T_{bM}, 35 $^\circ$C) or not (T_b); and considering the daily contribution (sum) for each of the 2 weeks previous to the given day.
*inkler _[4.5—10.0] [t0—1—2] _[TBase—TbaseMAX] _sum * 1 weeks _after*: Winkler heat accumulation index, calculated with a base temperature of 4.5 $^\circ$C or 10.0; accumulated since the beginning of the season (t_0), January the 1st of the date's year (t_1) or February the 1st (t_2); considering a maximum temperature threshold over which the heat accumulation stopped (T_{bM}, 35 $^\circ$C) or not (T_b); and considering the daily contribution (sum) for the next week after the given day.

Appendix B.2 Complete List of Dataset Column Names in the File

Table B.2 provides the full list of field names included in the dataset. The naming conventions used to construct these names are detailed in Appendix B.1.

Table B.2: Complete list of dataset column names.

Field name	**Field name**
phenologystageid	variety
cadastralcode	longitude
latitude	altitudeASL
PDO id	date
station	season
min	MAX
mean	std
median	diff
day	PDO Borja
PDO Calatayud	PDO Carinena
PDO Somontano	variety CABERNET SAUVIGNON
variety CHARDONNAY	variety GARNACHA
variety MAZUELA	variety SYRACH
variety TEMPRANILLO	tmed min
tmed min 1 days before	tmed min 2 days before
tmed min 3 days before	tmed min 4 days before
tmed min 5 days before	tmed min 6 days before
tmed min 7 days before	tmed min 8 days before
tmed min 9 days before	tmed min 10 days before
tmed min 11 days before	tmed min 12 days before
tmed min 13 days before	tmed min 1 days after
tmed min 2 days after	tmed min 3 days after
tmed min 4 days after	tmed min 5 days after
tmed min 6 days after	rad min
rad _min 1 days _before	rad min 2 days before
rad _min 3 days _before	rad min 4 days before
rad _min 5 days _before	rad min 6 days before
rad _min 7 days _before	rad min 8 days before
rad _min 9 days _before	rad min 10 days before
rad _min 11 days before	rad min 12 days before
rad _min 13 days before	rad min 1 days after

Field name	Field name
rad _min 2 days _after	rad min 3 days after
rad _min 4 days _after	rad min 5 days after
rad _min 6 days _after	tmed MAX
tmed MAX 1 days before	tmed MAX 2 days before
tmed MAX 3 days before	tmed MAX 4 days before
tmed MAX 5 days before	tmed MAX 6 days before
tmed MAX 7 days before	tmed MAX 8 days before
tmed MAX 9 days before	tmed MAX 10 days before
tmed MAX 11 days before	tmed MAX 12 days before
tmed MAX 13 days before	tmed MAX 1 days after
tmed MAX 2 days after	tmed MAX 3 days after
tmed MAX 4 days after	tmed MAX 5 days after
tmed MAX 6 days after	rad MAX
rad _MAX 1 days _before	rad MAX 2 days before
rad _MAX 3 days _before	rad MAX 4 days before
rad _MAX 5 days _before	rad MAX 6 days before
rad _MAX 7 days _before	rad MAX 8 days before
rad _MAX 9 days _before	rad MAX 10 days before
rad _MAX 11 days before	rad MAX 12 days before
rad _MAX 13 days before	rad MAX 1 days after
rad _MAX 2 days _after	rad MAX 3 days after
rad _MAX 4 days _after	rad MAX 5 days after
rad _MAX 6 days _after	tmed mean
tmed mean 1 days before	tmed mean 2 days before
tmed mean 3 days before	tmed mean 4 days before
tmed mean 5 days before	tmed mean 6 days before
tmed mean 7 days before	tmed mean 8 days before
tmed mean 9 days before	tmed mean 10 days before
tmed mean 11 days before	tmed mean 12 days before
tmed mean 13 days before	tmed mean 1 days after
tmed mean 2 days after	tmed mean 3 days after
tmed mean 4 days after	tmed mean 5 days after
tmed mean 6 days after	rad mean
rad _mean 1 days before	rad mean 2 days before
rad _mean 3 days before	rad mean 4 days before

Field name	Field name
rad _mean 5 days before	rad mean 6 days before
rad _mean 7 days before	rad mean 8 days before
rad _mean 9 days before	rad mean 10 days before
rad _mean 11 days before	rad mean 12 days before
rad _mean 13 days before	rad mean 1 days after
rad _mean 2 days after	rad mean 3 days after
rad _mean 4 days after	rad mean 5 days after
rad _mean 6 days after	hr mean
hr mean 1 days before	hr mean 2 days before
hr mean 3 days before	hr mean 4 days before
hr mean 5 days before	hr mean 6 days before
hr mean 7 days before	hr mean 8 days before
hr mean 9 days before	hr mean 10 days before
hr mean 11 days before	hr mean 12 days before
hr mean 13 days before	hr mean 1 days after
hr mean 2 days after	hr mean 3 days after
hr mean 4 days after	hr mean 5 days after
hr mean 6 days after	wind N
wind N 1 days before	wind N 2 days before
wind N 3 days before	wind N 4 days before
wind N 5 days before	wind N 6 days before
wind N 7 days before	wind N 8 days before
wind N 9 days before	wind N 10 days before
wind N 11 days before	wind N 12 days before
wind N 13 days before	wind N 1 days after
wind N 2 days after	wind N 3 days after
wind N 4 days after	wind N 5 days after
wind N 6 days after	wind NE
wind NE 1 days before	wind NE 2 days before
wind NE 3 days before	wind NE 4 days before
wind NE 5 days before	wind NE 6 days before
wind NE 7 days before	wind NE 8 days before
wind NE 9 days before	wind NE 10 days before
wind NE 11 days before	wind NE 12 days before
wind NE 13 days before	wind NE 1 days after

Field name	**Field name**
wind NE 2 days after	wind NE 3 days after
wind NE 4 days after	wind NE 5 days after
wind NE 6 days after	wind E
wind E 1 days before	wind E 2 days before
wind E 3 days before	wind E 4 days before
wind E 5 days before	wind E 6 days before
wind E 7 days before	wind E 8 days before
wind E 9 days before	wind E 10 days before
wind E 11 days before	wind E 12 days before
wind E 13 days before	wind E 1 days after
wind E 2 days after	wind E 3 days after
wind E 4 days after	wind E 5 days after
wind E 6 days after	wind SE
wind SE 1 days before	wind SE 2 days before
wind SE 3 days before	wind SE 4 days before
wind SE 5 days before	wind SE 6 days before
wind SE 7 days before	wind SE 8 days before
wind SE 9 days before	wind SE 10 days before
wind SE 11 days before	wind SE 12 days before
wind SE 13 days before	wind SE 1 days after
wind SE 2 days after	wind SE 3 days after
wind SE 4 days after	wind SE 5 days after
wind SE 6 days after	wind S
wind S 1 days before	wind S 2 days before
wind S 3 days before	wind S 4 days before
wind S 5 days before	wind S 6 days before
wind S 7 days before	wind S 8 days before
wind S 9 days before	wind S 10 days before
wind S 11 days before	wind S 12 days before
wind S 13 days before	wind S 1 days after
wind S 2 days after	wind S 3 days after
wind S 4 days after	wind S 5 days after
wind S 6 days after	wind SW
wind SW 1 days before	wind SW 2 days before
wind SW 3 days before	wind SW 4 days before
wind SW 5 days before	wind SW 6 days before

Field name	**Field name**
wind SW 7 days before	wind SW 8 days before
wind SW 9 days before	wind SW 10 days before
wind SW 11 days before	wind SW 12 days before
wind SW 13 days before	wind SW 1 days after
wind SW 2 days after	wind SW 3 days after
wind SW 4 days after	wind SW 5 days after
wind SW 6 days after	wind W
wind W 1 days before	wind W 2 days before
wind W 3 days before	wind W 4 days before
wind W 5 days before	wind W 6 days before
wind W 7 days before	wind W 8 days before
wind W 9 days before	wind W 10 days before
wind W 11 days before	wind W 12 days before
wind W 13 days before	wind W 1 days after
wind W 2 days after	wind W 3 days after
wind W 4 days after	wind W 5 days after
wind W 6 days after	wind NW
wind NW 1 days before	wind NW 2 days before
wind NW 3 days before	wind NW 4 days before
wind NW 5 days before	wind NW 6 days before
wind NW 7 days before	wind NW 8 days before
wind NW 9 days before	wind NW 10 days before
wind NW 11 days before	wind NW 12 days before
wind NW 13 days before	wind NW 1 days after
wind NW 2 days after	wind NW 3 days after
wind NW 4 days after	wind NW 5 days after
wind NW 6 days after	gdd 4.5 t0 Tbase sum
gdd 4.5 t0 Tbase sum 1 weeks before	gdd 4.5 t0 Tbase sum 2 weeks before
gdd 4.5 t0 Tbase sum 1 weeks after	gdd 4.5 t0 TbaseMAX sum
gdd 4.5 t0 TbaseMAX sum 1 weeks before	gdd 4.5 t0 TbaseMAX sum 2 weeks before
gdd 4.5 t0 TbaseMAX sum 1 weeks after	gdd 4.5 1 Tbase sum
gdd 4.5 1 Tbase sum 1 weeks before	gdd 4.5 1 Tbase sum 2 weeks before
gdd 4.5 1 Tbase sum 1 weeks after	gdd 4.5 1 TbaseMAX sum
gdd 4.5 1 TbaseMAX sum 1 weeks before	gdd 4.5 1 TbaseMAX sum 2 weeks before
gdd 4.5 1 TbaseMAX sum 1 weeks after	gdd 4.5 2 Tbase sum
gdd 4.5 2 Tbase sum 1 weeks before	gdd 4.5 2 Tbase sum 2 weeks before

Field name	Field name
dd 4.5 2 Tbase _sum 1 weeks after	gdd 4.5 2 TbaseMAX sum
gdd 4.5 2 TbaseMAX sum 1 weeks before	gdd 4.5 2 TbaseMAX sum 2 weeks before
gdd 4.5 2 TbaseMAX sum 1 weeks after	gdd 10.0 t0 Tbase sum
gdd 10.0 t0 Tbase sum 1 weeks before	gdd 10.0 t0 Tbase sum 2 weeks before
gdd 10.0 t0 Tbase sum 1 weeks after	gdd 10.0 t0 TbaseMAX sum
gdd 10.0 t0 TbaseMAX sum 1 weeks before	gdd 10.0 t0 TbaseMAX sum 2 weeks before
gdd 10.0 t0 TbaseMAX sum 1 weeks after	gdd 10.0 1 Tbase sum
gdd 10.0 1 Tbase sum 1 weeks before	gdd 10.0 1 Tbase sum 2 weeks before
gdd 10.0 1 Tbase sum 1 weeks after	gdd 10.0 1 TbaseMAX sum
gdd 10.0 1 TbaseMAX sum 1 weeks before	gdd 10.0 1 TbaseMAX sum 2 weeks before
gdd 10.0 1 TbaseMAX sum 1 weeks after	gdd 10.0 2 Tbase sum
gdd 10.0 2 Tbase sum 1 weeks before	gdd 10.0 2 Tbase sum 2 weeks before
gdd 10.0 2 Tbase sum 1 weeks after	gdd 10.0 2 TbaseMAX sum
gdd 10.0 2 TbaseMAX sum 1 weeks before	gdd 10.0 2 TbaseMAX sum 2 weeks before
gdd 10.0 2 TbaseMAX sum 1 weeks after	chillingDD 7.0 t0 Tbase sum
chillingDD 7.0 t0 Tbase sum 1 weeks before	chillingDD 7.0 t0 Tbase sum 2 weeks before
chillingDD 7.0 t0 Tbase sum 1 weeks after	chillingDD 7.0 t0 Tbasemin sum
chillingDD 7.0 t0 Tbasemin sum 1 weeks before	chillingDD 7.0 t0 Tbasemin sum 2 weeks before
chillingDD 7.0 t0 Tbasemin sum 1 weeks after	chillingDD 7.0 t0 Utah sum
chillingDD 7.0 t0 Utah sum 1 _weeks before	chillingDD 7.0 t0 Utah sum 2 weeks before
chillingDD 7.0 t0 Utah sum 1 weeks after	chillingDD 7.0 1 Tbase sum
chillingDD 7.0 1 Tbase sum 1 weeks before	chillingDD 7.0 1 Tbase sum 2 weeks before
chillingDD 7.0 1 Tbase sum 1 weeks after	chillingDD 7.0 1 Tbasemin sum
chillingDD 7.0 1 Tbasemin sum 1 weeks before	chillingDD 7.0 1 Tbasemin sum 2 weeks before
chillingDD 7.0 1 Tbasemin sum 1 weeks after	chillingDD 7.0 1 Utah sum
chillingDD 7.0 1 Utah sum 1 weeks before	chillingDD 7.0 1 Utah sum 2 weeks before

Field name	Field name
chillingDD 7.0 1 Utah sum 1 weeks after	chillingDD 7.0 2 Tbase sum
chillingDD 7.0 2 Tbase sum 1 weeks before	chillingDD 7.0 2 Tbase sum 2 weeks before
chillingDD 7.0 2 Tbase sum 1 weeks after	chillingDD 7.0 2 Tbasemin sum
chillingDD 7.0 2 Tbasemin sum 1 weeks before	chillingDD 7.0 2 Tbasemin sum 2 weeks before
chillingDD 7.0 2 Tbasemin sum 1 weeks after	chillingDD 7.0 2 Utah sum
chillingDD 7.0 2 Utah sum 1 weeks before	chillingDD 7.0 2 Utah sum 2 weeks before
chillingDD 7.0 2 Utah sum 1 weeks after	rad sum
rad _sum 1 weeks before	rad sum 2 weeks before
rad _sum 1 weeks after	precip sum
precip sum 1 weeks before	precip sum 2 weeks before
precip sum 1 weeks after	winkler 4.5 Tbase
winkler 4.5 Tbase 1 weeks before	winkler 4.5 Tbase 2 weeks before
winkler 4.5 Tbase 1 weeks after	winkler 4.5 TbaseMAX
winkler 4.5 TbaseMAX 1 weeks before	winkler 4.5 TbaseMAX 2 weeks before
winkler 4.5 TbaseMAX 1 weeks after	winkler 10.0 Tbase
winkler 10.0 Tbase 1 weeks before	winkler 10.0 Tbase 2 weeks before
winkler 10.0 Tbase 1 weeks after	winkler 10.0 TbaseMAX
winkler 10.0 TbaseMAX 1 weeks before	winkler 10.0 TbaseMAX 2 weeks before
winkler 10.0 TbaseMAX 1 weeks after	gdd 4.5 t0 Tbase sum Cumm
gdd 4.5 t0 Tbase sum Cumm 1 weeks before	gdd 4.5 t0 Tbase sum Cumm 2 weeks before
gdd 4.5 t0 Tbase sum Cumm 1 weeks after	gdd 4.5 t0 TbaseMAX sum Cumm
gdd 4.5 t0 TbaseMAX sum Cumm 1 weeks before	gdd 4.5 t0 TbaseMAX sum Cumm 2 weeks before

Field name	**Field name**
gdd 4.5 t0 TbaseMAX sum Cumm 1 weeks after	gdd 4.5 1 Tbase sum Cumm
gdd 4.5 1 Tbase _sum Cumm 1 weeks before	gdd 4.5 1 Tbase sum Cumm 2 weeks before
gdd 4.5 1 Tbase _sum Cumm 1 weeks after	gdd 4.5 1 TbaseMAX sum Cumm
gdd 4.5 1 TbaseMAX sum Cumm 1 weeks before	gdd 4.5 1 TbaseMAX sum Cumm 2 weeks before
gdd 4.5 1 TbaseMAX sum Cumm 1 weeks after	gdd 4.5 2 Tbase sum Cumm
gdd 4.5 2 Tbase _sum Cumm 1 weeks before	gdd 4.5 2 Tbase sum Cumm 2 weeks before
gdd 4.5 2 Tbase _sum Cumm 1 weeks after	gdd 4.5 2 TbaseMAX sum Cumm
gdd 4.5 2 TbaseMAX sum Cumm 1 weeks before	gdd 4.5 2 TbaseMAX sum Cumm 2 weeks before
gdd 4.5 2 TbaseMAX sum Cumm 1 weeks after	gdd 10.0 t0 Tbase sum Cumm
gdd 10.0 t0 Tbase sum _Cumm 1 weeks before	gdd 10.0 t0 Tbase sum Cumm 2 weeks before
gdd 10.0 t0 Tbase sum _Cumm 1 weeks after	gdd 10.0 t0 TbaseMAX sum Cumm
gdd 10.0 t0 TbaseMAX sum Cumm 1 weeks before	gdd 10.0 t0 TbaseMAX sum Cumm 2 weeks before
gdd 10.0 t0 TbaseMAX sum Cumm 1 weeks after	gdd 10.0 1 Tbase sum Cumm
gdd 10.0 1 Tbase sum Cumm 1 weeks before	gdd 10.0 1 Tbase sum Cumm 2 weeks before
gdd 10.0 1 Tbase sum Cumm 1 weeks after	gdd 10.0 1 TbaseMAX sum Cumm
gdd 10.0 1 TbaseMAX sum Cumm 1 weeks before	gdd 10.0 1 TbaseMAX sum Cumm 2 weeks before
gdd 10.0 1 TbaseMAX sum Cumm 1 weeks after	gdd 10.0 2 Tbase sum Cumm
gdd 10.0 2 Tbase sum Cumm 1 weeks before	gdd 10.0 2 Tbase sum Cumm 2 weeks before
gdd 10.0 2 Tbase sum Cumm 1 weeks after	gdd 10.0 2 TbaseMAX sum Cumm

Field name	**Field name**
gdd 10.0 2 TbaseMAX sum Cumm 1 weeks before	gdd 10.0 2 TbaseMAX sum Cumm 2 weeks before
gdd 10.0 2 TbaseMAX sum Cumm 1 weeks after	chillingDD 7.0 t0 Tbase sum Cumm
chillingDD 7.0 t0 Tbase sum Cumm 1 weeks before	chillingDD 7.0 t0 Tbase sum Cumm 2 weeks before
chillingDD 7.0 t0 Tbase sum Cumm 1 weeks after	chillingDD 7.0 t0 Tbasemin sum Cumm
chillingDD 7.0 t0 Tbasemin sum Cumm 1 weeks before	chillingDD 7.0 t0 Tbasemin sum Cumm 2 weeks before
chillingDD 7.0 t0 Tbasemin sum Cumm 1 weeks after	chillingDD 7.0 t0 Utah sum Cumm
chillingDD 7.0 t0 Utah sum Cumm 1 weeks before	chillingDD 7.0 t0 Utah sum Cumm 2 weeks before
chillingDD 7.0 t0 Utah sum Cumm 1 weeks after	chillingDD 7.0 1 Tbase sum Cumm
chillingDD 7.0 1 Tbase sum Cumm 1 weeks before	chillingDD 7.0 1 Tbase sum Cumm 2 weeks before
chillingDD 7.0 1 Tbase sum Cumm 1 weeks after	chillingDD 7.0 1 Tbasemin sum Cumm
chillingDD 7.0 1 Tbasemin sum Cumm 1 weeks before	chillingDD 7.0 1 Tbasemin sum Cumm 2 weeks before
chillingDD 7.0 1 Tbasemin sum Cumm 1 weeks after	chillingDD 7.0 1 Utah sum Cumm
chillingDD 7.0 1 Utah sum Cumm 1 weeks before	chillingDD 7.0 1 Utah sum Cumm 2 weeks before
chillingDD 7.0 1 Utah sum Cumm 1 weeks after	chillingDD 7.0 2 Tbase sum Cumm
chillingDD 7.0 2 Tbase sum Cumm 1 weeks before	chillingDD 7.0 2 Tbase sum Cumm 2 weeks before
chillingDD 7.0 2 Tbase sum Cumm 1 weeks after	chillingDD 7.0 2 Tbasemin sum Cumm
chillingDD 7.0 2 Tbasemin sum Cumm 1 weeks before	chillingDD 7.0 2 Tbasemin sum Cumm 2 weeks before
chillingDD 7.0 2 Tbasemin sum Cumm 1 weeks after	chillingDD 7.0 2 Utah sum Cumm
chillingDD 7.0 2 Utah sum Cumm 1 weeks before	chillingDD 7.0 2 Utah sum Cumm 2 weeks before

Field name	Field name
chillingDD 7.0 2 Utah sum Cumm 1 weeks after	rad t0 Cumm
rad t0 Cumm 1 weeks before	rad t0 Cumm 2 weeks before
rad t0 Cumm 1 weeks after	rad 1 Cumm
rad 1 Cumm 1 weeks before	rad 1 Cumm 2 weeks before
rad 1 Cumm 1 weeks after	rad 2 Cumm
rad 2 Cumm 1 weeks before	rad 2 Cumm 2 weeks before
rad 2 Cumm 1 weeks after	precip t0 Cumm
precip t0 Cumm 1 weeks before	precip t0 Cumm 2 weeks before
precip t0 Cumm 1 weeks after	precip 1 Cumm
precip 1 Cumm 1 weeks before	precip 1 Cumm 2 weeks before
precip 1 Cumm 1 weeks after	precip 2 Cumm
precip 2 Cumm 1 weeks before	precip 2 Cumm 2 weeks before
precip 2 Cumm 1 weeks after	winkler 4.5 t0 Tbase Cumm
winkler 4.5 t0 Tbase Cumm 1 weeks before	winkler 4.5 t0 Tbase Cumm 2 weeks before
winkler 4.5 t0 Tbase Cumm 1 weeks after	winkler 4.5 t0 TbaseMAX Cumm
winkler 4.5 t0 TbaseMAX Cumm 1 weeks before	winkler 4.5 t0 TbaseMAX Cumm 2 weeks before
winkler 4.5 t0 TbaseMAX Cumm 1 weeks after	winkler 4.5 1 Tbase Cumm
winkler 4.5 1 Tbase Cumm 1 weeks before	winkler 4.5 1 Tbase Cumm 2 weeks before
winkler 4.5 1 Tbase Cumm 1 weeks after	winkler 4.5 1 TbaseMAX Cumm
winkler 4.5 1 TbaseMAX Cumm 1 weeks before	winkler 4.5 1 TbaseMAX Cumm 2 weeks before
winkler 4.5 1 TbaseMAX Cumm 1 weeks after	winkler 4.5 2 Tbase Cumm
winkler 4.5 2 Tbase Cumm 1 weeks before	winkler 4.5 2 Tbase Cumm 2 weeks before
winkler 4.5 2 Tbase Cumm 1 weeks after	winkler 4.5 2 TbaseMAX Cumm
winkler 4.5 2 TbaseMAX Cumm 1 weeks before	winkler 4.5 2 TbaseMAX Cumm 2 weeks before
winkler 4.5 2 TbaseMAX Cumm 1 weeks after	winkler 10.0 t0 Tbase Cumm
winkler 10.0 t0 Tbase Cumm 1 weeks before	winkler 10.0 t0 Tbase Cumm 2 weeks before
winkler 10.0 t0 Tbase Cumm 1 weeks after	winkler 10.0 t0 TbaseMAX Cumm

Field name	Field name
winkler 10.0 t0 _TbaseMAX Cumm 1 weeks before	winkler 10.0 t0 TbaseMAX Cumm 2 weeks before
winkler 10.0 t0 _TbaseMAX Cumm 1 weeks after	winkler 10.0 1 Tbase Cumm
winkler 10.0 1 Tbase Cumm 1 weeks before	winkler 10.0 1 Tbase Cumm 2 weeks before
winkler 10.0 1 Tbase Cumm 1 weeks after	winkler 10.0 1 TbaseMAX Cumm
winkler 10.0 1 TbaseMAX Cumm 1 weeks before	winkler 10.0 1 TbaseMAX Cumm 2 weeks before
winkler 10.0 1 TbaseMAX Cumm 1 weeks after	winkler 10.0 2 Tbase Cumm
winkler 10.0 2 Tbase Cumm 1 weeks before	winkler 10.0 2 Tbase Cumm 2 weeks before
winkler 10.0 2 Tbase Cumm 1 weeks after	winkler 10.0 2 TbaseMAX Cumm
winkler 10.0 2 TbaseMAX Cumm 1 weeks before	winkler 10.0 2 TbaseMAX Cumm 2 weeks before
winkler 10.0 2 TbaseMAX Cumm 1 weeks after	