

## Article

# Calibration for an Ensemble of Grapevine Phenology Models under Different Optimization Algorithms

Chenyao Yang<sup>1,2,\*</sup> , Christoph Menz<sup>3</sup>, Samuel Reis<sup>1</sup> , Nelson Machado<sup>4</sup>, João A. Santos<sup>1</sup>   
and Jairo Arturo Torres-Matallana<sup>5</sup> 

<sup>1</sup> Centre for the Research and Technology of Agro-Environmental and Biological Sciences (CITAB), Institute for Innovation, Capacity Building and Sustainability of Agri-Food Production (Inov4Agro), University of Trás-os-Montes and Alto Douro (UTAD), 5000-801 Vila Real, Portugal

<sup>2</sup> College of Agronomy, Sichuan Agricultural University, Chengdu 611130, China

<sup>3</sup> Potsdam Institute for Climate Impact Research e. V. (PIK), Telegrafenberg A 31, 14473 Potsdam, Germany

<sup>4</sup> CoLAB VINES & WINES-National Collaborative Laboratory for the Portuguese Wine Sector, Associação para o Desenvolvimento da Viticultura Duriense (ADVID), Edifício Centro de Excelência da Vinha e do Vinho, Régia Douro Park, 5000-033 Vila Real, Portugal

<sup>5</sup> Environmental Sensing and Modelling Unit, Agro-Environmental Systems Group, Luxembourg Institute of Science and Technology (LIST), 41 Rue de Brill, L-4422 Belvaux, Luxembourg

\* Correspondence: cyang@utad.pt

**Abstract:** Vine phenology modelling is increasingly important for winegrowers and viticulturists. Model calibration is often required before practical applications. However, when multiple models and optimization methods are applied for different varieties, it is rarely known which factor tends to mostly affect the calibration results. We mainly aim to investigate the main source of the variability in the modelling errors for the flowering timings of two important varieties of vine in the Douro Demarcated Region (DDR) of Portugal; this is based on five phenology model simulations that use optimal parameters and that are estimated by three optimization algorithms (MLE, SA and SCE-UA). Our results indicate that the main source of the variability in calibration can be affected by the initially assumed parameter boundary. Restricting the initial parameter distribution to a narrow range impedes the algorithm from exploring the full parameter space and searching for optimal parameters. This can lead to the largest variation in different models. At an identified appropriate boundary, the difference between the two varieties represents the largest source of uncertainty, while the choice of algorithm for calibration contributes least to the overall uncertainty. The smaller variability among different models or algorithms (tools for analysis) compared to between different varieties could indicate the overall reliability of the calibration. All optimization algorithms show similar results in terms of the obtained goodness-of-fit: the RMSE (MAE) is 5–6 (4–5) days with a negligible mean bias and moderately good  $R^2$  (0.5–0.6) for the ensemble median predictor. Nevertheless, a similar predictive performance can result from differently estimated parameter values, due to the equifinality or multi-modal issue in which different parameter combinations give similar results. This mainly occurs for models with a non-linear structure compared to those with a near-linear one. Yet, the former models are found to outperform the latter ones in predicting the flowering timing of the two varieties in the DDR. Overall, our findings highlight the importance of carefully defining the initial parameter boundary and decomposing the total variance of prediction errors. This study is expected to bring new insights that will help to better inform users about the importance of choice when these factors are involved in calibration. Nonetheless, the importance of each factor can change depending on the specific situation. Details of how the optimization methods are applied and of the continuous model improvement are important.

**Keywords:** grapevine; phenology modelling; ensemble simulation; parameter estimation; parameter boundary; optimization approach



**Citation:** Yang, C.; Menz, C.; Reis, S.; Machado, N.; Santos, J.A.; Torres-Matallana, J.A. Calibration for an Ensemble of Grapevine Phenology Models under Different Optimization Algorithms. *Agronomy* **2023**, *13*, 679. <https://doi.org/10.3390/agronomy13030679>

Academic Editor: Carmen Galan

Received: 28 December 2022

Revised: 16 February 2023

Accepted: 23 February 2023

Published: 26 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Plant phenology, the annually recurring pattern of plant development, has been identified as one key biological indicator of climate change [1]. The relationship between the timing of phenological events and climatic factors, such as temperature, is fundamental for understanding the ecosystem's response to global warming [2]. Phenology prediction is primarily driven by temperature, which is a pivotal part of process-based crop modelling. It is typically implemented as a phenological module within crop models [3,4]. The accurate simulation of phenology development is a prerequisite for modelling other crucial physiological processes, such as transpiration, photosynthesis and yield formation, etc. [3,4]. Therefore, many crop modelling studies focus specifically on phenology modelling [5–8].

Viticulture phenology modelling has received much attention in the last decade, as many studies have shown the potential of using simple temperature-based statistical or mechanistic models to successfully predict and monitor the vine phenology of commercial grape varieties [9–12]. For instance, a simple Grapevine Flowering Veraison (GFV) model using the concept of growing degree days (GDD), was shown to be able to accurately (generally RMSE  $\leq$  7 days) reproduce the observed flowering and veraison timings of 95 and 104 common varieties, respectively, which spanned across 123 locations in Europe with a wide range of environmental conditions [9]. Similarly, Leolini et al. [12] demonstrate the applicability of six common phenology models for predicting the budbreak of main grapevine varieties. Although many of these models were satisfactorily calibrated and evaluated for simulating vine phenology, deciding which model to use in a new environment is often difficult, as no single individual model could consistently outperform the others [13]. Each model has its strengths and weaknesses, depending on the environmental conditions (e.g., climate and soil) and the type of cultivar being simulated [9–12].

A multi-model ensemble could be used to overcome the limitations of individual models, where the ensemble mean or median of the model outputs can be used to represent robust predictions [6,14]. The multi-model ensemble can be directly integrated into decision support systems. For example, they could be conveniently coupled with existing real-time climate forecasts to provide early-season phenology predictions [15,16]. This can support monitoring vine development, which is particularly necessary in the context of anticipating more frequent extreme weather conditions with the projected climate change [17–21]. Multi-model ensembles are also an integral part of the Agricultural Model Intercomparison and Improvement Project (AgMIP) [6,22,23].

Before any practical applications, phenology models need to be calibrated [24]. During calibration, model-specific parameters are adjusted (optimized), in order to simulate the observed conditions as close as possible. Different models have different parameters, but most of these parameters are considered to reflect the genetic differences between cultivars, i.e., cultivar-dependent characteristics of phenology development [5,7,9,25]. To estimate the optimal values of these parameters (i.e., parameter vector), a large diversity of calibration approaches are available, each having their own statistical assumptions, calibration steps, criteria of best-fit parameters and optimization algorithms [6–8,13,26]. Variations in calibration approaches, model structures and input data represent three main sources of uncertainty in crop model (including phenology model) simulations [5,26,27]. In particular, different optimization algorithms are available, but the algorithm selection, according to the calibration method, could considerably affect parameter estimations (including the associated uncertainties) and calibration results [7,8,13,26].

The best algorithm is often dependent on a specific context. For instance, Gao et al. [7] suggest that the importance of algorithms can vary depending on the priority of the study: if the priority is on the goodness-of-fit between the observations and simulations, the Ordinary Least Square (OLS) method is the most effective choice; meanwhile, in cases where an estimation of the parameter uncertainty and model prediction is more important, the Markov Chain Monte Carlo (MCMC)-based algorithms are more appropriate. Kawakita et al. [8] further report that uncertainties in the parameter estimations and the associated calibration performance are affected not only by parameter estimation methods

or optimization techniques, but also by the model structure. However, these studies do not yet address, when multiple models and algorithms are available, whether the application of different algorithms for a given model or the assessment of different models for a certain algorithm is more important. Available information on this matter is important. For instance, if different optimization algorithms tend to give similar results, users may just choose the simplest one for their models.

Some well-known examples of applied optimization algorithms include the Maximum Likelihood Estimation (MLE) [28], the Simulated Annealing (SA) [29] and the Shuffled Complex Evolution-University of Arizona (SCE-UA) [30]. These algorithms have different strategies to search along the parameter space. Nevertheless, it is essential to specify the parameter search space, e.g., by defining the lower and upper boundary in an optimal way. Yet, this is often defined empirically by researchers [7,8,31,32], and it is rarely investigated whether the pre-defined boundary of parameter distribution is sufficient to warrant finding the optimal parameters by the algorithm.

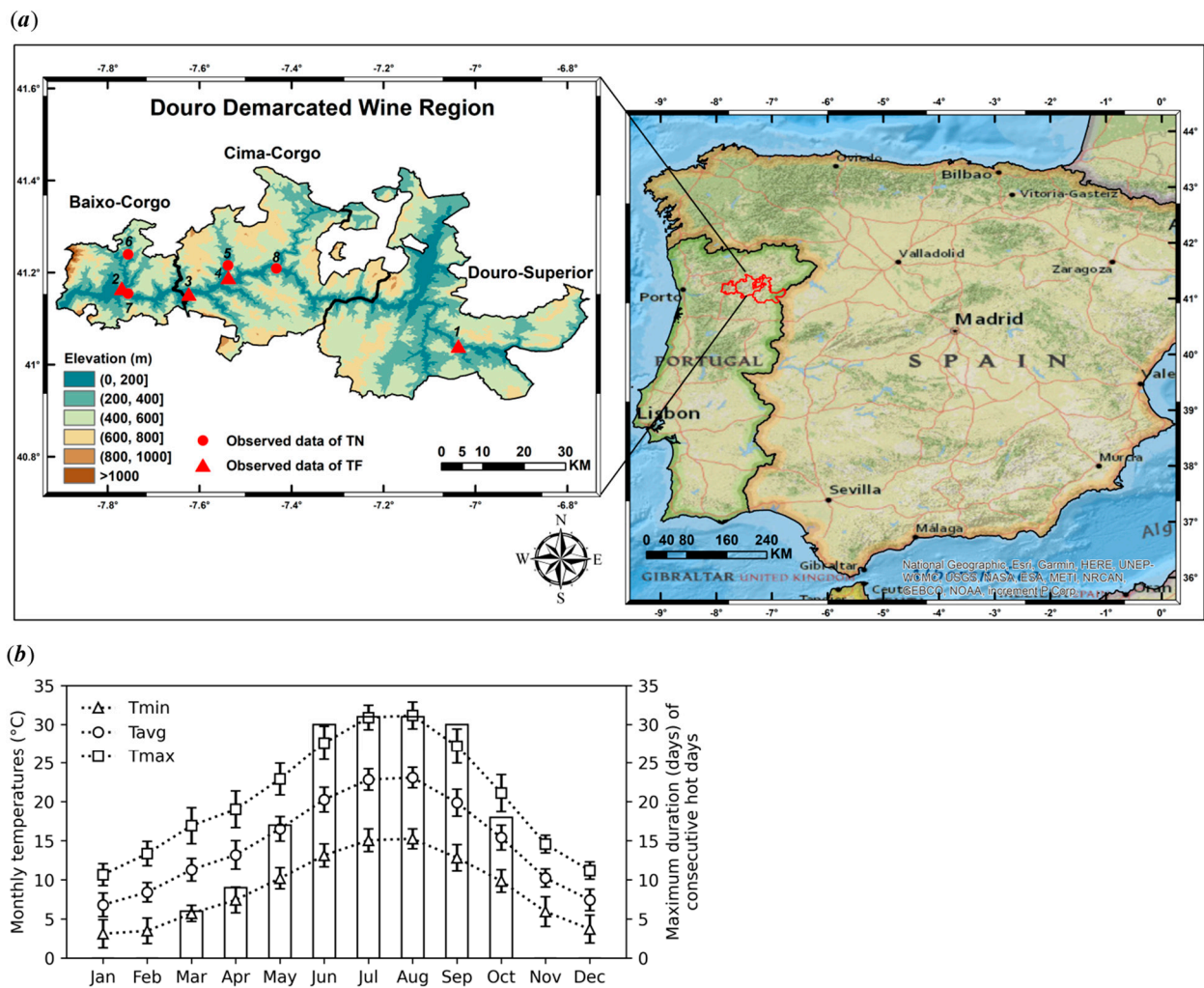
In this study, we utilized five phenology models with varying degrees of complexity, which were all previously shown to be able to adequately reproduce vine phenology (see Section 2.2). These models were calibrated for simulating the flowering timing of two main grapevine varieties in Portugal, i.e., Touriga Franca (TF) and Touriga Nacional (TN), using three optimization algorithms (MLE, SA and SCE-UA). In general, the main objective represents an attempt to provide one plausible answer to the research question: when common factors of parameter estimations are explicitly taken into account, including models, algorithms and varieties (predicting events), which one tends to have the greatest effect on the final calibrations? This is carried out using Analysis of Variance (ANOVA) to identify the main source of variability (a larger magnitude of variations) among 5 models  $\times$  3 algorithms  $\times$  2 varieties. Note that a preliminary analysis with different assumptions for the initial parameter boundary is performed to determine the most appropriate constraint for the prior parameter distribution during parameter optimization. Based on the estimated parameter values from each optimization algorithm, the calibrated multi-model ensemble is constructed. For each algorithm, we then evaluate the goodness-of-fit between the observations and individual model simulations using a number of conventional statistical measures. Consequently, the ensemble median of the multi-model simulations is used to represent our best predictions against the observations. The variability in the resulting estimated parameter values among the different algorithms is also presented.

## 2. Data and Methods

### 2.1. Study Region and Data Collection

#### 2.1.1. Study Region

The Douro Demarcated Region (DDR), known as the Douro/Port Wine Region, located in northeast Portugal, is one of the oldest winemaking regions in the world [33–37]. It is recognized and included in the world heritage list by UNESCO for its unique mountainous landscape and long history (nearly 2000 years) of a wine-producing culture [36]. DDR spreads over an area of ~250,000 ha and is divided into three subregions (Baixo Corgo, Cima Corgo and Douro Superior) that differ greatly from each other in their heterogeneous topography (Figure 1a). Viticulture is the main socioeconomic activity of the region and vineyards are traditionally grown on terraces over steep slopes along the Douro valley [36]. It comprises an area of about 45,000 ha, accounting for approximately 22% of the total area of Portugal [34–36]. DDR features a typical Mediterranean climate with warm dry summers and a mild wet autumn and winter. The low precipitation, high temperature and radiation amount in the summer months (June–August) frequently gives rise to situations of intense plant water stress [33].



**Figure 1.** The spatial distribution of (a) study sites (numbered per Table 1) where the phenology data of two grapevine varieties (TF: Touriga Franca, TN: Touriga Nacional) are collected within the Douro Demarcated Wine Region and (b) the monthly mean of daily minimum ( $T_{\min}$ , °C), average ( $T_{\text{avg}}$ , °C) and maximum ( $T_{\max}$ , °C) temperatures over 1991–2020 (error bars: standard deviation of monthly means), along with the maximum duration (days) of consecutive hot days ( $T_{\max} \geq 25$  °C for at least 5 days) for each month. The meteorological statistics are based on the spatial average of all study sites.

**Table 1.** Mean and standard deviation (sd) of the Day Of Year (DOY) for the observed flowering (BBCH65) of two main grapevine varieties (Touriga Franca: TF; Touriga Nacional: TN) within the Douro Demarcated Region (DDR) during the period of 2014–2021.

|                          | Touriga Franca (TF) |           |           |           | Touriga Nacional (TN) |           |           |           |
|--------------------------|---------------------|-----------|-----------|-----------|-----------------------|-----------|-----------|-----------|
|                          | Site1               | Site2     | Site3     | Site4     | Site5                 | Site6     | Site7     | Site8     |
| Longitude                | −7.037° W           | −7.769° W | −7.623° W | −7.537° W | −7.538° W             | −7.755° W | −7.755° W | −7.433° W |
| Latitude                 | 41.040° N           | 41.166° N | 41.153° N | 41.189° N | 41.215° N             | 41.240° N | 41.154° N | 41.209° N |
| Observations (Mean ± sd) | 134 ± 5             | 135 ± 4   | 139 ± 4   | 130 ± 6   | 139 ± 8               | 139 ± 10  | 131 ± 9   | 129 ± 8   |

### 2.1.2. Phenology Measurements

In the present study, we focused on the grapevine phenology stage of full flowering (BBCH65), i.e., the date when 50% flowering occurred, following the definition of the growth stages of mono- and dicotyledonous plants [38]. Data for two main Portuguese grapevine varieties (Touriga Franca: TF; Touriga Nacional: TN) were collected over 2014–2021 in

eight experimental vineyard plots within the DDR (Table 1), via a viticultural observatory network coordinated by the Association for the Development of Viticulture in the Douro Region (ADVID). The geographic locations of these plots are shown in Figure 1a. Across the study sites, there was a greater inter-annual variability in the observed flowering timing for TN (8–10) than for TF (4–6) (Table 1). To better represent the phenological characteristics of the variety, we pooled all site data (blend all) for a given variety; thus, each variety had 32 observations from 4 sites  $\times$  8 years for calibration. The amount of data was comparable to a similar study that used 27 observations of rice phenology from different year–site–season combinations in order to estimate the optimal model parameters [5].

### 2.1.3. Climate Observations

The recent E-OBS (v24.0e) observational gridded meteorological dataset, with a spatial resolution of  $0.1^\circ$  [39], was employed to extract daily temperature series for phenology modelling. The advantage of utilizing a gridded dataset, instead of data from site-specific weather stations, was its potential for large-scale applications in agro-environmental studies, e.g., producing regional maps [40,41], and its wide spatial coverage, which was especially important for data-scarce regions [42]. Moreover, the applicability and reliability of the E-OBS dataset, when coupled with phenology models, has already been successfully demonstrated to reproduce the observed vine phenology in many Portuguese wine regions (including the DDR) well [31,43,44]. The monthly temperature statistics for  $T_{\min}$ ,  $T_{\text{avg}}$  and  $T_{\max}$ , deriving from the E-OBS data, are shown in Figure 1b for the respective study sites. For the flowering stage, the critical months were mostly from April to June (end of spring to the beginning of summer), during which  $T_{\text{avg}}$  could increase from 13 to  $20^\circ\text{C}$  ( $T_{\max}$  from 18 to  $28^\circ\text{C}$ ), which was accompanied by a prolonged high-temperature event ( $T_{\max} \geq 25^\circ\text{C}$  for at least 5 days) (Figure 1b). Such high-temperature events could extend through to half of May and all of June (Figure 1b).

## 2.2. Phenology Models

The five models selected were GDD [45], GDD–Richardson (Richardson for short) [46], Sigmoid [47], Triangular [48] and Wang [49]. From an onset/starting date ( $t_0$ ), these thermal forcing/driving models simulated the occurrence of a given phenology stage at the day ( $t_s$ ), when the state of the forcing temperatures ( $S_f$ ) reached a threshold value of  $F^*$  (generally thought to be variety-dependent) [9,10]:

$$S_f(t_s) = \sum_{t_0}^{t_s} R_f(x_t) \geq F^* \quad (1)$$

where  $R_f(x_t)$  was calculated as a function of the daily mean temperature ( $x_t$ ). The phenology models differ depending on how  $R_f(x_t)$  are applied. They could be divided into two types: degree-day-based models (GDD and Richardson) that generally handled  $R_f(x_t)$  via the accumulation of  $x_t$  above a given threshold, and non-degree-day-based models (Sigmoid, Triangular and Wang) that computed  $R_f(x_t)$  based on the normalized function of temperature ( $R_f$  varied from 0 to 1) [48]. A graphic representation of each model with different settings of parameter values is shown in Figure S1. Detailed explanations of the model parameters are presented in Table 2. It is noteworthy that the Richardson, Triangular and Wang models share a similar definition of parameters ( $T_b$  and  $T_{\max}$ ) (Table 2). All models were initialized on January 1st, which was the initialization date adopted in other modelling studies for predicting the flowering timing of important varieties of vine in major Portuguese wine regions (including DDR) [10,50,51]. Suboptimal photoperiod and vernalization effects are not considered.

**Table 2.** Selected parameters of phenology models for calibration. The specified initial lower and upper boundary of uniform parameter distribution corresponds to the baseline setting. The “×” symbol indicates the parameter intended for the calibration of a given model. All model simulations are driven by the daily mean temperature from the first day of the year ( $t_0$ ).

| Parameter Abbreviation | Description  | Unit   | Initial Lower Bound | Initial Upper Bound | Model Name                 |            |         |            |      |
|------------------------|--|--|---------------------|---------------------|----------------------------|------------|---------|------------|------|
|                        |  |  |                     |                     | Growing Degree Day (GDD) * | Richardson | Sigmoid | Triangular | Wang |
| $F^*$                  | Critical state of thermal temperature              | degree.days <sup>-1</sup> *<br>or temperature ratios * | 40 * or<br>1100 *   | 80 * or<br>1300 *   | ×                          | ×          | ×       | ×          | ×    |
| $T_b$                  | Minimum development temperature (base temperature) | °C   | 0                   | 5                   |                            | ×          |         | ×          | ×    |
| $T_{opt}$              | Optimum development temperature                    | °C   | 22                  | 25                  |                            |            |         | ×          | ×    |
| $T_{max}$              | Maximum development temperature                    | °C   | 32                  | 38                  |                            | ×          |         | ×          | ×    |
| $d$                    | Fitted parameter of sharpness of response curve    | /  | −15                 | 0                   |                            |            | ×       |            |      |
| $e$                    | Fitted parameter of mid-response temperature       | °C   | 10                  | 20                  |                            |            | ×       |            |      |

\* GDD and Richardson models compute the  $F^*$  with the degree.days<sup>-1</sup>, while the other models compute  $F^*$  (explained in Equation (1)) with temperature ratios (between 0 and 1). Accordingly, the initial  $F^*$  search boundary is defined as 1100–1300 degree.days<sup>-1</sup> for GDD and Richardson, and 40–80 for other models. For the GDD model, the base temperature ( $T_b$ ) is constantly set at 0 °C.

### 2.2.1. GDD

The GDD model (Figure S1) is based on the classical thermal time concept [45]. It includes two parameters ( $T_b$  and  $F^*$ ) (Table 2), which generally assume a linear approach for the effective accumulation of daily thermal forcing when  $x_t > T_b$  (base temperature):

$$R_f(x_t) = \begin{cases} 0 & \text{if } x_t \leq T_b \\ x_t - T_b & \text{if } x_t > T_b \end{cases} \quad (2)$$

We set  $T_b$  at 0 °C (not calibrated) since this was successfully applied for the GDD in a simulation of the flowering timing of up to 95 grapevine varieties (over 123 locations) [9,11]. Hence, only a single parameter,  $F^*$ , was adjusted during parameter optimization (Table 2).

### 2.2.2. Richardson

The Richardson model (Figure S1) is a modified version of the GDD model. It includes three parameters ( $T_b$ ,  $T_{max}$  and  $F^*$ ) (Table 2), which additionally consider a plateau above the high threshold temperature ( $T_{max}$ ) [46]:

$$R_f(x_t) = \text{Max}(\text{Min}(x_t - T_b, T_{max} - T_b), 0) \quad (3)$$

As shown in previous studies, the Richardson model was able to accurately predict the flowering timing of both TF and TN for important wine regions in Portugal [50].

### 2.2.3. Sigmoid

The Sigmoid model (Figure S1) adopts the logistic function/sigmoid curve as the temperature–response curve [47]. It includes three parameters ( $d$ ,  $e$  and  $F^*$ ) (Table 2), which have a similar structure to the UniFORC model [10–12]:

$$R_f(x_t) = 1 / (1 + e^{(d*(x_t - e))}) \quad (4)$$

The sigmoid model has been calibrated and evaluated for simulating the flowering stage of numerous different grapevine varieties over multiple Portuguese wine regions [51].

#### 2.2.4. Triangular

The Triangular model (Figure S1) uses the piecewise linear function with three temperature thresholds ( $T_b$ ,  $T_{opt}$ ,  $T_{max}$ ) to define the temperature–response function [48]. It includes four parameters ( $T_b$ ,  $T_{opt}$ ,  $T_{max}$  and  $F^*$ ) (Table 2):

$$R_f(x_t) = \begin{cases} 0 & \text{if } x_t \leq T_b \\ \frac{x_t - T_b}{T_{opt} - T_b} & \text{if } T_b < x_t \leq T_{opt} \\ \frac{x_t - T_{max}}{T_{opt} - T_{max}} & \text{if } T_{opt} < x_t < T_{max} \\ 0 & \text{if } x_t \geq T_{max} \end{cases} \quad (5)$$

This model has been well adapted to simulate the flowering timing of many white and red varieties in Portugal [10].

#### 2.2.5. Wang

The Wang model (Figure S1) uses similar temperature thresholds ( $T_b$ ,  $T_{opt}$ ,  $T_{max}$ ) to the Triangular model, but integrates the temperature effects with a non-linear response function [49]. It includes four parameters ( $T_b$ ,  $T_{opt}$ ,  $T_{max}$  and  $F^*$ ) (Table 2):

$$R_f(x_t) = \begin{cases} \frac{2(x_t - T_b)^\alpha (T_{opt} - T_b)^\alpha - (x_t - T_b)^{2\alpha}}{(T_{opt} - T_b)^{2\alpha}} & \text{if } T_b \leq x_t \leq T_{max} \\ 0 & \text{if } x_t < T_b \text{ or } x_t > T_{max} \end{cases} \quad (6)$$

with

$$\alpha = \frac{\ln 2}{\ln \left( \frac{T_{max} - T_b}{T_{opt} - T_b} \right)}$$

This model has proven to be able to accurately represent the observed flowering timing of TF and TN in important Portuguese wine regions [50].

### 2.3. Applied Optimization Algorithms

#### 2.3.1. Maximum Likelihood Estimation (MLE)

The maximum likelihood estimation (MLE) is a simple optimization method intended to find the local maximum of the objective function [28]. Therefore, the objective function is regarded as a likelihood function with the parameter vector representing function variables. Hence, the optimum parameter vector maximizes the likelihood. For the present study, MLE is implemented as a Markov chain Monte Carlo process based on the Metropolis–Hastings algorithm [52]. In each iteration step, a random parameter vector is drawn from a Gaussian kernel in the neighborhood of the current best vector  $x_{old}$ . If the new parameter vector  $x_{new}$  outperforms the current one, the latter is exchanged and the next iteration cycle begins. MLE often only finds the local maxima of the objective function. Therefore, to improve the efficiency, the parameter space is scanned randomly to find the best initial parameter vector prior to the MLE iterations, i.e., burn-in samples [53].

#### 2.3.2. Simulated Annealing (SA)

Simulated annealing is a probabilistic optimization method that is able to find the global optimum of the objective function [29]. Its design resembles the annealing process in metallurgy. Simulated annealing utilizes the Metropolis algorithm [54]. The optimum parameter is found iteratively by randomly drawing a new parameter set  $x_{new}$  in the vicinity of the current optimum  $x_{old}$ . The new parameter set is accepted if the objective function  $f(x_{new})$

calculated from this set has increased from the previous optimum  $f(x_{old})$ . Otherwise, the non-optimal parameter set can still be accepted based on the Boltzmann distribution:

$$P(x_{new}) \propto \exp\left(-\frac{f(x_{old}) - f(x_{new})}{T}\right) \quad (7)$$

Here,  $T$  represents a numerical temperature, intended to be reduced during iteration. Reducing the temperature will effectively reduce the probability of accepting non-optimal parameter sets. Simulated annealing has three parameters (initial temperature, number of draws at a fixed temperature level and rate of temperature reduction) that need to be carefully tuned for each problem [29].

### 2.3.3. Shuffled Complex Evolution-University of Arizona (SCE-UA)

The Shuffled Complex Evolution method, developed at The University of Arizona (SCE-UA), is an efficient, flexible and robust algorithm that is able to find the global optimum of the calibrated model [30]. Following Duan et al. [30], SCE-UA embodies the properties that an optimization algorithm must possess: (1) global convergence when multiple local optimums can be found; (2) ability to avoid being trapped in non-optimum regions on the objective function surface; (3) robustness when the parameter sensitivities differ and parameter interdependencies exist; (4) non-reliance on the availability of an explicit expression for the objective function or the derivatives; and (5) high-parameter dimensionality handling. The method is based on four main concepts [30,55]: (1) a combination of deterministic and probabilistic methods; (2) systematic evolution of a 'complex' in a systematic search by drawing points spanning the parameter space, which is the global optimum; (3) competitive evolution; and (4) shuffling of the complex. Indeed, SCE-UA combines the algorithms of the simplex procedure [56], random search [57], competitive evolution [58] and complex shuffling [59]. The parametrization of SCE-UA needs to be carefully tuned for each problem, and has been one of the most widely used algorithms in hydrological applications [53].

## 2.4. Analysis Steps for Model Parameter Calibration

### 2.4.1. Application of Different Optimization Algorithms for Parameter Estimations

A brief summary of the important features and case-specific settings for the studied algorithms are presented in Table 3. The three algorithms differ in their approaches to deriving the initial best-guess value (and other initial settings), numeric methods and objective functions (Table 3). We have presented a diagram showing the general procedures to estimate the parameters for each model using any of these optimization algorithms (Figure 2). In brief, this could be described as follows:

1. Define the initial distribution of the parameters. Here, we choose the uniform probability distribution function (Table 3); thus, parameter values were drawn with an equal probability between a given lower and upper boundary.
2. Draw the initial parameter vector and choose the appropriate algorithm settings. In this step, we mostly adopt the default settings (Table 3) using the implementation software (Section 2.6) [53]. Note that the step size parameter was used for the following parameter sampling.
3. Sample the parameters around the current values, which vary mainly depending on the numeric method of the algorithm (Table 3).
4. Perform model simulations with the sampled parameter vector and obtain the objective function by comparing the model output to the observed values.
5. Evaluate whether the new parameter vector should be kept or not, which differs in the criterion for different algorithms (Table 3). In general, a better objective function could result in a new parameter position [53].
6. Repeat steps 3–5 in an iterative process unless the finishing criterion of the algorithm has been reached (Figure 2).

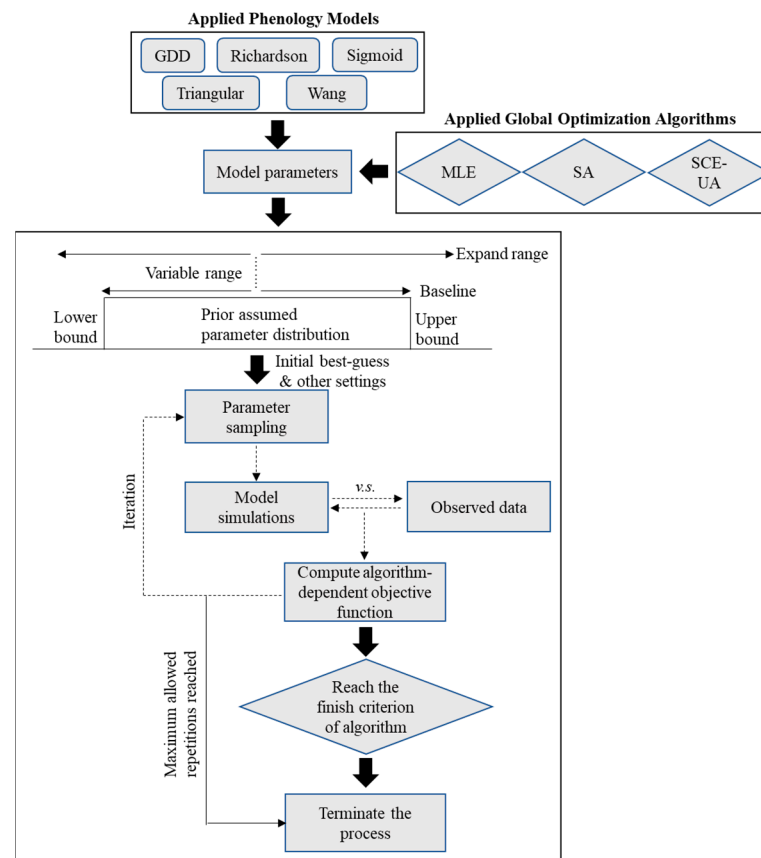


**Table 3.** A brief summary of the features and user-specified settings for the three employed optimization methods.

| Features   | Maximum Likelihood Estimation (MLE)  | Simulated Annealing (SA)   | Shuffled Complex Evolution-University of Arizona (SCE-UA)   |
|--|--|--|---|
| Optimization direction   | Adapting parameters in directions with an increasing likelihood  | Emulating the physical process whereby a solid is slowly cooled until the minimum possible energy                                    | Evolution based on a statistical “reproduction” using the “simplex” geo-metric shape towards an improvement direction |
| Algorithm  | Metropolis-Hastings  | Metropolis   | Combines the simplex procedure, random search, competitive evolution and complex shuffling                            |
| Risk to stuck in local optimum                                     | High   | Medium   | Low   |
| Supporting Reference   | [28]   | [29]   | [30]  |
| Case-study settings  |  |  |   |
| Prior assumed distribution of parameter values                     | Uniform distribution   | Uniform distribution   | Uniform distribution  |
| Maximum number of repetitions                                      | 30,000   | 30,000   | 30,000  |
| Initial best-guess of parameter values                             | Best parameter sets after running the first 10% of repetitions (a burn-in analysis)  | Median of randomly sampled values from the prior distribution (sample size = 1000)   | Randomly sampled from initial population  |
| Initial algorithm setting  | Step size: difference between 50th and 40th percentile over randomly sampled values from the prior distribution (sample size = 1000) | Step size: difference between 50th and 40th percentile over randomly sampled values from the prior distribution (sample size = 1000) | Number of complexes: 40;<br>Number of past evolution loops: 100   |
| Objective function calculated between simulations and observations | Maximize the logarithmic probability   | Maximize the negative RMSE   | Minimize the RMSE   |
| Finish criterion   | Maximum repetitions exhausted  | Maximum repetitions exhausted or complete cooling down of optimization   | Maximum repetitions exhausted or parameter convergence occur  |

The details of how the calibration approach was applied could have important impacts on the estimated parameters, e.g., the initial lower and upper boundary of parameter distribution [26]. Therefore, we have empirically defined the parameter boundary (baseline setting), shown in Table 2, according to the information collected from the literature data, modelling experts and local winegrowers or viticulturists. However, it could not be determined for certain that the specified boundary settings would guarantee that the algorithm would find optimal parameters. Therefore, we imposed in our experiment a variable range for the parameter boundary (Figure 2): we gradually expanded the baseline setting on each parameter range by a fixed percentage of 20%, 40%, 60%, 80% and 100%, respectively (i.e., 120–200% relative to baseline range). For each of these expanded range settings, we performed a full optimization run and computed the RMSE (modelling errors) between observations and simulations using the obtained optimized parameter vectors. Note that MLE and SA did not choose to minimize the RMSE as the objective function (Table 3). It was expected that a plateau/saturation in the modelling errors (RMSE)

would be found, which would possibly imply that the appropriate parameter search space was found.



**Figure 2.** A universal procedure to apply different optimization algorithms in order to estimate the parameters of phenology models. The iterative process is denoted by dash lines.

#### 2.4.2. Identification of the Main Source of Variability by ANOVA

ANOVA is a standard approach to quantifying different sources of uncertainty in crop modelling and climate impact assessment studies [5,21]. To quantify the contribution of the models (5 levels), algorithms (3 levels) and varieties (2 levels) to the total variance in the prediction performance (fit to calibration data), a 3-factor ANOVA with a fixed experiment design was conducted. To obtain a uniform structure for the ANOVA, the predictive performance of the calibrated model was quantified, as the RMSE explained in Section 2.4.1. Note that the ANOVA was conducted per boundary setting. This could be useful in investigating whether the main source of variability (i.e., uncertainty) during optimization/calibration could vary with different boundary settings.

#### 2.4.3. Parameter Variability between Algorithms

The parameter variability for each model was quantified as the variation (i.e., the coefficient of variation, CV) in the obtained optimal parameter values across the different optimization algorithms. The dimensionless nature of CV allows the comparison of parameters in different models. It also permits an assessment of whether a model has more unstable parameters than others when different algorithms are applied. Note that this was different from studies that computed the variability (uncertainty) in calibrated parameter values from either multiple folds of cross-validation [8], or from the obtained posterior distribution of parameters [7].

### 2.5. Evaluations of Goodness-of-Fit

A number of conventional goodness-of-fit measures were adopted to evaluate the agreement between the observation and simulation of the optimal parameters [5,7,8], including Mean Bias Error (MBE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and the coefficient of determination ( $R^2$ ). These metrics mainly evaluated how well the model simulations, using the obtained parameters for each algorithm, agreed with the observations. Note that the evaluation was based on the same data that were used for calibration, meaning that the results are generally not indicative of the predictive performance for new and different situations. This could be useful for comparison with other studies [5,13]. The ensemble median among different models was adopted as the best predictor, in accordance with a study with 27 different phenology modelling groups that showed a slightly better prediction performance for the ensemble median than that of the ensemble mean [6]. Moreover, since the selected models had varying degrees of complexity, Akaike's Information Criterion (AIC) [60] was used to rate the prediction accuracy over the model structures. It aimed to evaluate the trade-off between model parsimony and efficiency, as it allowed the combination of information on the model complexity in terms of the number of parameters and the prediction accuracy in terms of the mean squared error [60]. It has recently been recommended that its application is integrated into standard calibration guidelines [13]. The lower the AIC score is, the better trade-off the model has:

$$AIC = n \times \ln \left( \frac{\sum_{i=1}^n (S_i - O_i)^2}{n} \right) + 2k \quad (8)$$

where  $O_i$  and  $S_i$  were the observed and simulated values using the obtained best-fit parameters, respectively, and  $n$  and  $k$  corresponded to the sample size of the observation and the number of parameters, respectively.

### 2.6. Implementation Software

All the phenology models were written in python according to Equations (1)–(6). Numpy (v1.21.2), Panda (v1.3.4), Matplotlib (v3.4.3) and Statsmodels (v0.13.1) were involved in the data processing, statistical analysis and result visualization. For the parameter optimization procedures illustrated in Figure 2, the open source python package of Statistical Parameter Optimization Tool (SPOTPY, v1.5.14) was used [53]. SPOTPY provided a wide range of implementations for the parameter optimization algorithms for calibration, uncertainty and sensitivity analysis in a simple, independent, flexible and consistent environment [53].

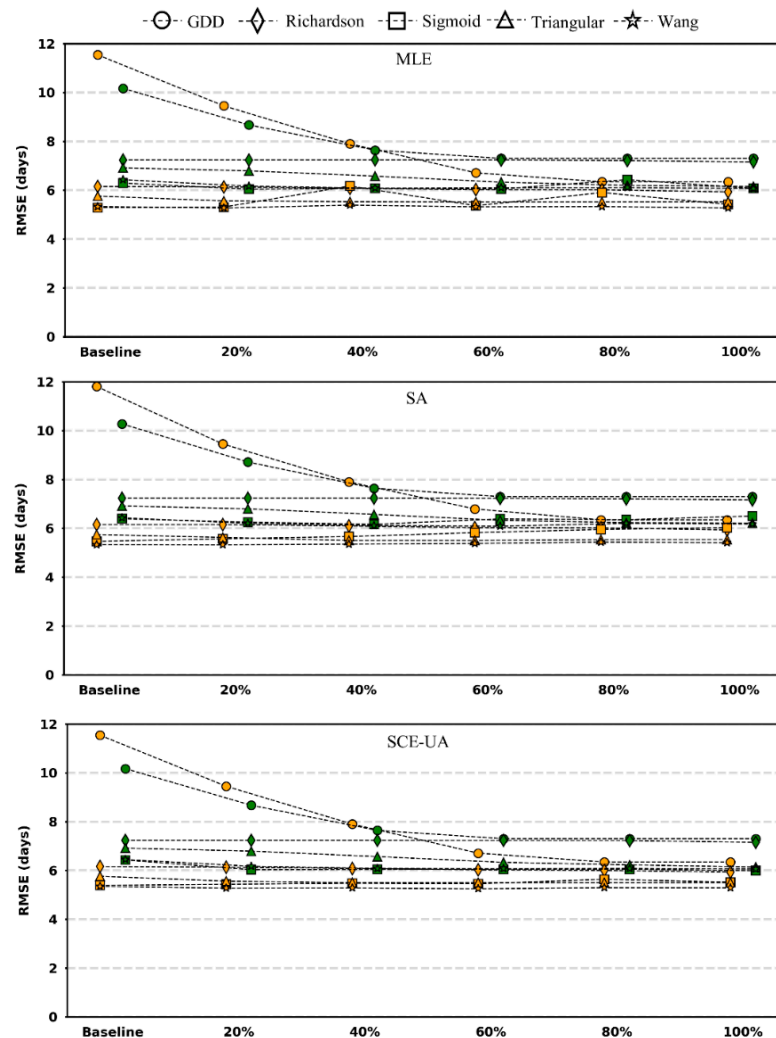
## 3. Results and Discussion

### 3.1. Overview of Optimization Performance

It is evident that all three algorithms tend to follow a very similar pattern, which consistently shows that the different boundaries of the initial parameter distribution have an impact on the optimization performance (Figure 3). However, it also depends on the phenology model in use. When the search space is expanded by 100% from the initial baseline setting (Table 2), the optimized final RMSE is considerably reduced from approximately 10–12 days (depending on the variety) to 6–7 days for the GDD model, while the remaining models have a relatively smaller variation of about 5–8 days (Figure 3). The plateau/saturation of the optimization performance for all the models and algorithms was when the expanded parameter range was between 80% and 100% (Figure 3).

A sufficient setting for the initial parameter distribution is important for the parameter optimization. The assumed uniform distribution is generally considered appropriate [7,32]. Therefore, our focus turns to the lower and upper parameter limits. The results indicate that the initially defined parameter boundary (baseline settings) (Table 2) is insufficient, as smaller optimized RMSE values are found when it is expanded. An appropriate boundary is found when the initial range is expanded by 100% (80% is not chosen to avoid random

variability between 60–80%, which is detected in MLE) (Figure 3). The corresponding updated parameter boundary is presented in Table S1, which should set the standard for the subsequent analysis. With this setting, a better fit to the calibration data is consistently found in TF rather in TN (Figure 3), which could be possibly due to the smaller temporal variability in the phenology observations for TF than for TN (Table 1).



**Figure 3.** Root Mean Squared Error (RMSE) computed between five phenology model simulations of the optimum parameters obtained from three global optimization algorithms (MLE: Maximum Likelihood Estimation; SA: Simulated Annealing; SCE-UA: Shuffled Complex Evolution) and the corresponding observed flowering data of two dominant Portuguese grapevine varieties (Touriga Franca: filled orange symbols; Touriga Nacional: filled green symbols). The orange and green symbols are slightly displaced in the  $x$ -axis to avoid the symbol overlap. The analysis is carried out by gradually expanding the initially defined range of distribution for each model parameter (baseline settings) by up to 100% (the percentage in the  $x$ -axis corresponds to each expanded percentage test category). An expanded range of 20–100% corresponds to 120–200%, relative to the initial baseline range. The observed phenology data correspond to blended multi-vineyard observations within the Douro region in the period of 2014–2021.

Many parameter estimation studies choose to define the initial parameter boundary using their empirical knowledge and information in the literature [7,8,31,32]. For instance, in a case study from arid Northwest China, Ran et al. [32] suggest that the initial lower and upper parameter boundary can be directly set as  $\pm 30\%$ , which is around the nominal value defined by the data in the literature; however, they found two key parameters with incorrect

posterior distributions, resulting from their inaccurate initial nominal values [32]. This implies that the empirical choice could sometimes be inaccurate. Hence, it is important to explicitly examine different boundaries. For all the algorithms, different boundary settings cause a larger variation in the optimized RMSE for the GDD model (for both varieties) than for the other models (Figure 3). GDD is a simple linear model and has only one parameter ( $F^*$ ); this is calibrated in our case (Table 2). Accordingly, more precise information on the initial parameter distribution is required. However, it also seems that the appropriate value of the thermal threshold ( $F^*$ ) for the flowering stage of the two varieties of vine is not covered by our initial baseline setting (Table 2).

### 3.2. Variations in the Main Source of Uncertainty

The ANOVA results quantify the different sources of variance to the total prediction errors (quantified using RMSE) of the calibrated models, which are presented in Table 4. The results indicate that the dominant source of variability can depend on the initial parameter boundary (Table 4). Variations among models explain most of the total variance when the initial boundary is expanded by up to 40% (Table 4). At a larger expanded boundary between 60% and 100%, the difference between the two varieties appears to account for most of the variance (Table 4). In comparison, the variations among the algorithms or the algorithm-related combinations (Algorithm  $\times$  Model or Algorithm  $\times$  Variety) contribute only a small fraction to the total variance throughout (Table 4). All factors, including their interactions, show statistical significance ( $p < 0.01$ ), except for the interaction between algorithm and variety (Table 4).

**Table 4.** Analysis of Variance (ANOVA) for RMSE computed between observations and simulations (with the optimal parameters) from different combinations of optimization algorithms, phenology models and grapevine varieties. ANOVA is conducted under different settings of parameter search space. Asterisks indicate the statistical significance of F statistics at  $p < 0.01$  (df: degree of freedom). The main source of variance is highlighted in bold for each experimental setting. The results for intercept are not shown. Note computation of all statistics (e.g., F value) is based on the full numeric precision, but numbers are all rounded to only three decimal digits for simplicity in this table.

| Experimental Settings on the Parameter Search Space | Source of Variability      | Sum of Squares | df       | Mean Squares  | F                   |
|---|----------------------------|----------------|----------|---------------|---------------------|
| Baseline  | Algorithm                  | 0.021          | 2        | 0.010         | 10.921 *            |
|   | <b>Model</b>               | <b>109.746</b> | <b>4</b> | <b>27.436</b> | <b>28,783.694 *</b> |
|   | Variety                    | 2.546          | 1        | 2.546         | 2671.396 *          |
|   | Algorithm $\times$ Model   | 0.050          | 8        | 0.006         | 6.502 *             |
|   | Model $\times$ Variety     | 7.561          | 4        | 1.890         | 1982.982 *          |
|   | Algorithm $\times$ Variety | 0.004          | 2        | 0.002         | 2.062               |
|   | Residual                   | 0.008          | 8        | 0.001         |                     |
| 20%   | Algorithm                  | 0.028          | 2        | 0.014         | 20.509 *            |
|   | <b>Model</b>               | <b>46.037</b>  | <b>4</b> | <b>11.509</b> | <b>16,937.038 *</b> |
|   | Variety                    | 2.891          | 1        | 2.891         | 4253.860 *          |
|   | Algorithm $\times$ Model   | 0.038          | 8        | 0.005         | 7.015 *             |
|   | Model $\times$ Variety     | 3.876          | 4        | 0.969         | 1425.862 *          |
|   | Algorithm $\times$ Variety | 0.001          | 2        | 0.000         | 0.630               |
|   | Residual                   | 0.005          | 8        | 0.001         |                     |
| 40%   | Algorithm                  | 0.039          | 2        | 0.019         | 1.547               |
|   | <b>Model</b>               | <b>16.555</b>  | <b>4</b> | <b>4.139</b>  | <b>329.154 *</b>    |
|   | Variety                    | 2.753          | 1        | 2.753         | 218.942 *           |
|   | Algorithm $\times$ Model   | 0.102          | 8        | 0.013         | 1.015               |
|   | Model $\times$ Variety     | 2.031          | 4        | 0.508         | 40.375 *            |
|   | Algorithm $\times$ Variety | 0.036          | 2        | 0.018         | 1.451               |
|   | Residual                   | 0.101          | 8        | 0.013         |                     |

Table 4. Cont.

| Experimental Settings on the Parameter Search Space | Source of Variability | Sum of Squares | df       | Mean Squares | F                 |
|---|-----------------------|----------------|----------|--------------|-------------------|
| 60%   | Algorithm             | 0.055          | 2        | 0.027        | 39.323 *          |
|   | Model                 | 7.792          | 4        | 1.948        | 2789.948 *        |
|   | <b>Variety</b>        | <b>4.749</b>   | <b>1</b> | <b>4.749</b> | <b>6802.155 *</b> |
|   | Algorithm × Model     | 0.139          | 8        | 0.017        | 24.871 *          |
|   | Model × Variety       | 0.382          | 4        | 0.095        | 136.731 *         |
|   | Algorithm × Variety   | 0.005          | 2        | 0.002        | 3.563             |
|   | Residual              | 0.006          | 8        | 0.001        |                   |
| 80%   | Algorithm             | 0.047          | 2        | 0.024        | 36.510 *          |
|   | Model                 | 5.399          | 4        | 1.350        | 2091.537 *        |
|   | <b>Variety</b>        | <b>4.989</b>   | <b>1</b> | <b>4.989</b> | <b>7731.389 *</b> |
|   | Algorithm × Model     | 0.106          | 8        | 0.013        | 20.564 *          |
|   | Model × Variety       | 0.478          | 4        | 0.120        | 185.283 *         |
|   | Algorithm × Variety   | 0.001          | 2        | 0.000        | 0.724             |
|   | Residual              | 0.005          | 8        | 0.001        |                   |
| 100%  | Algorithm             | 0.126          | 2        | 0.063        | 51.434 *          |
|   | Model                 | 5.658          | 4        | 1.415        | 1159.106 *        |
|   | <b>Variety</b>        | <b>5.139</b>   | <b>1</b> | <b>5.139</b> | <b>4210.815 *</b> |
|   | Algorithm × Model     | 0.249          | 8        | 0.031        | 25.455 *          |
|   | Model × Variety       | 0.446          | 4        | 0.111        | 91.354 *          |
|   | Algorithm × Variety   | 0.002          | 2        | 0.001        | 0.967             |
|   | Residual              | 0.010          | 8        | 0.001        |                   |

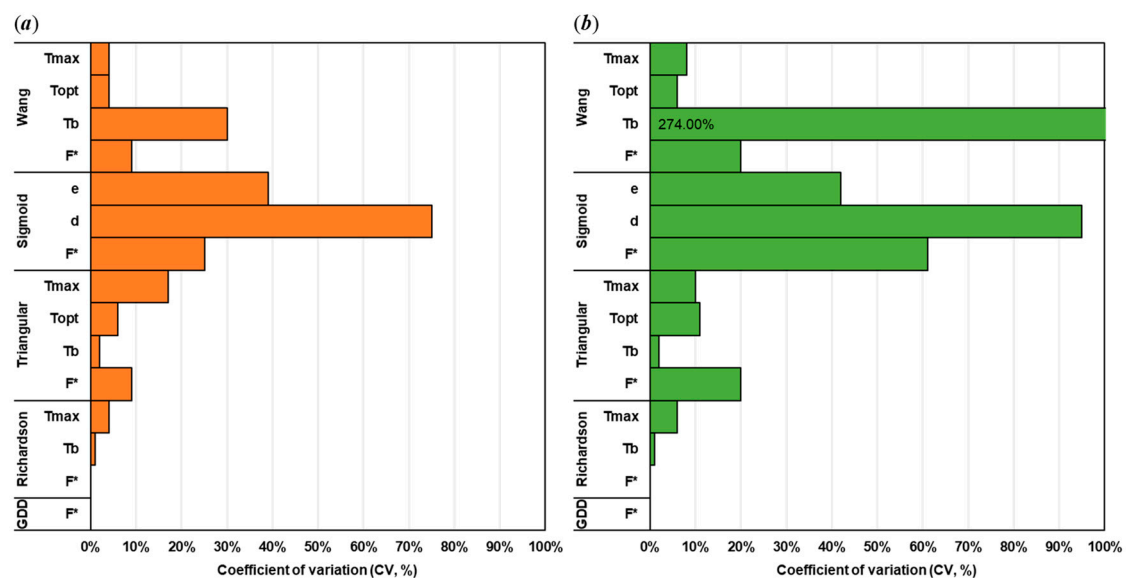
The ANOVA design is analogous to that of a study estimating the variance in model errors based on simulations with optimized parameters using three different calibration methods [7]. It is important to decompose the total variance in model errors to better understand the contributions of different sources. Similarly, Kawakita et al. [8] present a study involving 3 optimization methods × 3 phenology models × 3 Japanese wheat cultivars, concluding that all three factors impact the predictive performance (quantified with RMSE). However, they do not reveal which factor tends to play a major role [8]. In an attempt to fill this research gap, we demonstrate that the main source of uncertainty for the optimization results can vary with the assumed initial parameter boundary. The choice of model tends to play a major role when the initial parameter boundary is expanded by up to 40%, which can relate to the relatively high prediction errors of the GDD model (Figure 3). It appears that expanding the initial range by 40% (i.e., 140% relative to baseline setting) is still insufficient to optimize every model.

Given that the standard setting is to expand the initial baseline boundary by 100%, the “true” dominant source of uncertainty is identified as variety (Table 4). For a given variety, either the choice of the model or optimization method can be treated as a tool for analysis. Ideally, smaller uncertainties should be expected from them in order to better approximate observations of different varieties (predicting events), which can be considered as the external factor. This is especially important when there is a natural variability between observations, i.e., different variability in observed phenology between TF and TN (Table 1). Therefore, an overall reasonable calibration result is achieved in our study. Indeed, the difference between varieties represents the second largest source of uncertainty even when there is a still remarkable variability between models (boundary expansion ≤ 40%) (Table 4). In contrast, the selection of the optimization algorithm is identified as the smallest contribution to the overall uncertainty (Table 4), which can also be evidenced by a very similar pattern of evolution for the predictive performance shown in Figure 3. Gao et al. [7] come to similar findings using three different calibration approaches, namely Generalized Likelihood Uncertainty Estimation (GLUE), MCMC and OLS for the flowering timings of four different rice varieties in a multi-environmental trial in Madagascar and Senegal.

However, for another study simulating the heading date of Japanese winter wheat, the simulation errors of the calibrated models for new and different environments tend to vary considerably among different algorithms [8]. Therefore, we assume that whether the results among different algorithms agree or not may depend on the subject, region and focus of the study. In our case, for predicting the flowering timing of two vine varieties, the focus is on the predictive performance of the same data used for calibration. The identified small variability among the algorithms could prompt users to just choose the simplest one, i.e., MLE (relatively simple in terms of code implementation), to calibrate their models. However, it could be of high interest to evaluate how well the calibrated models that use these algorithms predict new and different situations using independent data sets [6]. Note that the spatial variability in the prediction performance can also contribute to the total prediction uncertainties; however, this is not separately quantified, as it is already included at the variety level.

### 3.3. Estimations of Parameter Variability

At the standard initial boundary setting (Figure 3), the estimated parameter values are considered optimal by a given algorithm, which is presented in Table S2. The resulting parameter variability among algorithms is shown to depend on the model structure (Figure 4). For both varieties, a relatively lower CV ( $\leq 20\%$ ) is found for models with near-linear functions (linear plateau for GDD and Richardson, and piecewise linear for Triangular), while the non-linear models (Sigmoid and Wang) present a much higher CV (up to 274%) (Figure 4). In particular, all the obtained optimal parameters diverge substantially among the algorithms for the Sigmoid model (CV is 25–75% for TF and 42–95% for TN), while for the Wang model, the unstable parameter among the algorithms mainly occurs for the  $T_b$  parameter (CV is 30% for TF and 274% for TN) (Figure 4). In contrast, stable (CV  $\leq 6\%$ ) parameters are obtained for the GDD and Richardson models, with a near-linear structure and a growing degree-day-based concept (Figure 4 and Table S2).



**Figure 4.** Coefficient of variation (CV, %) for obtained optimal parameter values among the three applied algorithms for (a) Touriga Franca (TF) and (b) Touriga Nacional (TN). These are obtained with the expanded lower and upper boundary of parameter distribution by 100%, relative to the baseline settings. The obtained individual parameter values are presented in Table S2. Note that the CV value for  $T_b$  parameter of the Wang model in TN is denoted as it exceeds 100%.

Selected optimization algorithms should ideally converge into the same or nearly the same parameter values for a given model. The variation in the final parameter values, if any, could be small and possibly result from inherently different numeric accuracies between

the algorithms or the intrinsic randomness of each algorithm [13,53]. In our case, similar parameter values occur for near-linear models (GDD, Richardson and Triangular). However, a larger parameter variability is found for Sigmoid and Wang. Parameter variability could be associated with the issue of equifinality or being multi-modal [61]: different parameter combinations (Figure 4 and Table S2) result in similar prediction errors among algorithms (Figure 3). This is a common phenomenon in phenology modelling [13,26,31]. The underlying cause can be the issue of parameter correlation [8,25]. Liu et al. [25] show that parameters with a higher uncertainty can have a stronger relationship with others, using data collected across major winter wheat production provinces in China. In most of the temperature–response functions of our models (Figure S1), a negative correlation is found between  $T_b$  and  $F^*$ , where a lower  $T_b$  value can be offset by a higher  $F^*$ . This means that an increased temperature accumulation rate (per day) due to a lower base temperature can be eventually offset by setting a higher thermal threshold (we confirm this point, but it is not shown). Therefore, calibration does not always lead to unique parameter values [8,13,26,27,31].

Moreover, our results illustrate that the variability in the estimated parameter values, using different optimization algorithms, may depend on the model structure. It was previously reported that models with fewer parameters can have smaller parameter uncertainties: the largest uncertainty mainly occurs for a sigmoidal and exponential function model (similar to our sigmoid model) with many mutually correlated parameters [8]. Besides the number of parameters, our study further reveals that models with a more complex (e.g., non-linear) structure tend to have more unstable parameters than those with a simpler structure (e.g., near-linear) (Figure 4). In other words, equifinality or being multi-modal is more prone to occur during the optimization of non-linear models than for near-linear ones. It is possible that complex models tend to have a more complex surface of objective functions, leading to multiple local optimums with different parameter vectors. Nevertheless, higher parameter variability is generally found for TN than for TF (Figure 4), which can be associated with the higher inter-annual variability in the observed phenology of TN than for TF (Table 1), i.e., it is more difficult to find best-fit parameters given the observed data.

### 3.4. Evaluations between Observations and Simulations of Optimized Parameters

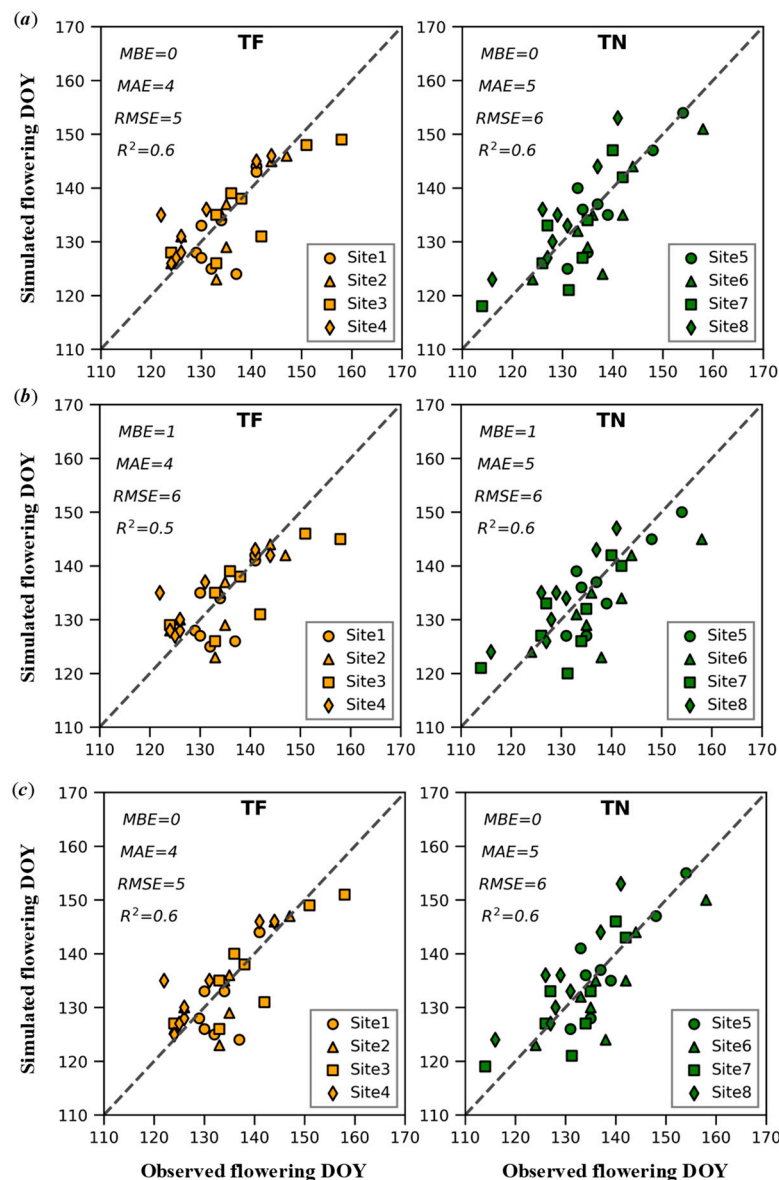
#### 3.4.1. Similarity among Algorithms

For individual models, the evaluation results generally show satisfactory agreement between the observations and simulations using optimized parameters, regardless of the algorithm (Figures S2–S4). Differences in the MBE, MAE and RMSE among the algorithms are negligible ( $<2$  days), while no changes are detected for  $R^2$  except for the Sigmoid model (Figures S2–S4). Specifically, the Sigmoid model optimized with SA (Figure S3) provides slightly less satisfactory predictions ( $R^2 = 0.5$ ) than those ( $R^2$  is 0.6–0.7) by coupling it with either MLE (Figure S2) or SCE-UA (Figure S4). Consequently, the ensemble median predictor results in a similar predictive performance over the algorithms, where the corresponding RMSE (MAE) is 5–6 (4–5) days, with minimized MBE (0–1 days) and moderately good  $R^2$  (0.5–0.6) (Figure 5).

The evaluation of how well the calibrated model (with the obtained optimized parameters) can reproduce the data is an integral part of model calibration studies [5,7,8,25]. Although these algorithms diverge in their obtained optimal parameter values (mostly for non-linear models) (Figure 4), they nearly converge in their prediction performance (Figures 5 and S2–S4). This is also consistent with the ANOVA results (Table 4). The algorithms are all demonstrated to be useful and effective for parameter estimations/calibrations for the multi-model ensemble, as evaluated by a range of goodness-of-fit metrics (Figures 5 and S2–S4). The results are generally comparable to those within the grapevine phenology modelling community [9–11,51]. For instance, the RMSE is 5.9–6.3 days for a differently parameterized Spring Warming model when predicting the flowering stage of 43 grapevine varieties based on 1092 observations across a wide range



of European wine regions [11].  $RMSE \leq 7$  days is generally regarded as the key criterion for an acceptable predictive performance [9–11,51]. This is obtained in the current study, together with almost zero mean bias and a small MAE ( $\leq 6$  days); this is the case for either the individual models or for the ensemble median predictor. The multi-model ensemble median predictions are comparable to or nearly as good as those of the best model (Figures 5 and S2–S4). This aspect is consistent with other studies [6,14].



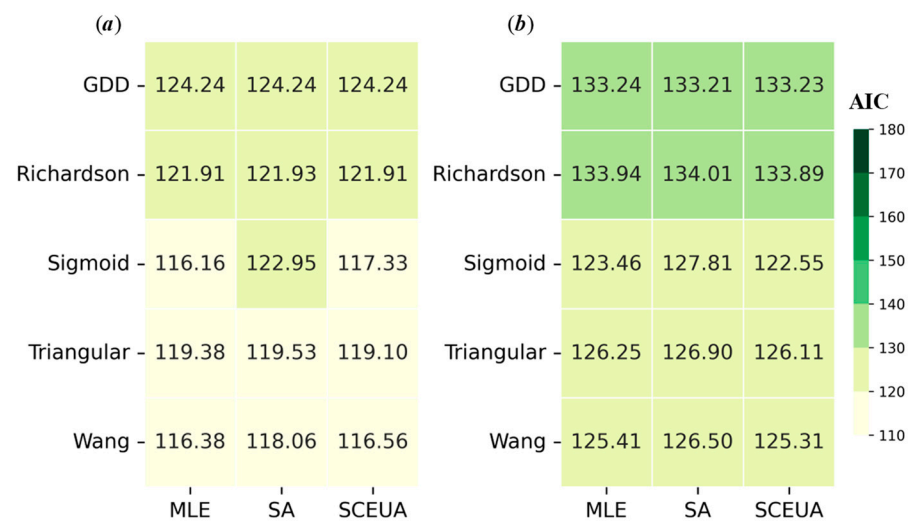
**Figure 5.** Comparison between observed flowering timing (DOY) and the median of the ensemble model simulations using optimized parameter values (Table S2) from (a) MLE, (b) SA and (c) SCE-UA. Touriga Franca (TF) and Touriga Nacional (TN) are denoted with orange and green symbols, respectively. Mean Bias Error (MBE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and the coefficient of determination ( $R^2$ ) are conventional statistics for evaluating the goodness-of-fit. The geographic locations of the study sites for each variety are presented in Figure 1 and Table 1. Evaluations for individual model simulations are shown in the supplementary material (Figures S2–S4).

Slightly poorer predictions for the Sigmoid model optimized by SA than those of the other two algorithms could be mainly due to the unrealistically small  $F^*$  value (7.7 for TN) obtained by the SA, accompanied by a high  $e$  value (Table S2), i.e., more difficulties

in obtaining effective temperature accumulations. SA, as a hyper-parameter algorithm, should be carefully tuned for each specific optimization task. Thus, it is advisable that the different parameter settings of SA are tested and explored in future studies. Moreover, the obtained  $R^2$  (0.4–0.7) is less satisfactory (Figures S2–S4), and the ensemble median predictor can only explain up to 60% of the total spatial–temporal variations in the observations (Figure 5). It should be noted that the current study utilized a gridded meteorological dataset ( $0.1^\circ$ ) [39] for phenology modelling. It is expected to have input errors during calibration. The overall moderate performance suggests the feasibility of the gridded dataset for modelling actual observed phenology, hinting at the potential for large-scale applications. Yet, to better capture the underlying spatial–temporal variability, models need to see continuous improvement (e.g., to incorporate the photoperiod and dormancy effects), along with better and high-quality field measurements (e.g., data from a target population) [6].

### 3.4.2. Difference between Modelling Types

AIC shows a similar pattern to that of previous evaluation results (Figures S2–S4), where a better performance is found for TF than for TN, and different algorithms result in very similar results but with variations among models (Figure 6). For TF, the AIC values are, on average, about 123 for degree-day (GDD, Richardson) and 118 for non-degree-day (Sigmoid, Triangular and Wang) models, respectively (Figure 6a). For TN, they are, respectively, 134 and 126 for degree-day and non-degree-day models (Figure 6b). Small variations are generally found within each model type (Figure 6).



**Figure 6.** Evaluations with the Akaike Information Criterion (AIC) using optimized parameters over different algorithms and models for (a) Touriga Franca (TF) and (b) Touriga Nacional (TN). The lower the AIC value, the better the trade-off between the model parsimony and prediction accuracy. The analysis is carried out under the expanded lower and upper boundary of the parameter distribution by 100%, relative to the baseline settings.

It is of particular interest to quantitatively evaluate individual model performances to understand the ensemble variability, which could indicate the ensemble quality, e.g., if some models outperform others. One of the somewhat commonly applied evaluation metrics for phenology modelling is AIC, which considers both the predictive performance and model complexity [11,12,50]. To predict the flowering stage in the DDR, Costa et al. [50] report AIC values of 159–172 for TF and 170–180 for TN, using four phenology models optimized by SA. In comparison, our AIC values are considerably smaller for both varieties (<135) (Figure 6), suggesting that a better trade-off (parsimony and efficiency) is obtained for the currently applied model and algorithm combinations. Moreover, our results indicate that the non-degree-day-based models mostly outperform the degree-day-based models

(Figure 6). Although GDD is preferentially selected over many other models as the best choice for simulating the flowering timing of 95 different grapevine varieties [9], it is not the favorite choice for TF and TN within the DDR, where high-temperature events are frequent (Figure 1b) [33]. It is known that models based on the concept of “degree-days” have limited applicability in phenology predictions under extreme climate conditions [2]. Meanwhile, for non-degree-day models, better performance is achieved with a non-linear structure (Sigmoid and Wang) (Figure 6). Although these models tend to have more unstable parameters when different algorithms are used (Figure 4), they seem to have more appropriate temperature–response functions for predicting the flowering stage of the two varieties in the DDR. Nevertheless, the best model and/or algorithm should be dependent on a specific situation, e.g., environment and/or variety.

#### 4. Conclusions

We have presented a study illustrating the influencing of common factors (model, optimization algorithm and variety) on the predictive performance of calibrated phenology models. The parameters of five models (GDD, Richardson, Sigmoid, Triangular and Wang) with varying degrees of complexity are optimized for simulating the flowering stage of two important grapevine varieties in a mountainous wine region (DDR), using three different optimization algorithms (MLE, SA and SCE-UA). We demonstrate that it is important to first define the appropriate boundary of the initial parameter distribution. A narrow range in the distribution could hinder the optimization algorithm from finding out the best-fit parameters, leading to a contrasting prediction performance between the model groups. Using the appropriate parameter boundaries, the variety in the simulation represents the largest contribution to the overall variability. Ideally, the variability among the models or optimization algorithms (considered as internal factors) should be smaller than that between varieties (considered as the external factor), in order to provide indicative results for a reasonable calibration. This is particularly necessary when there is natural variability between two varieties, as reflected by the different inter-annual variability in the observed phenology data. In contrast, the lowest contribution to the overall variability is the choice of the applied algorithm. All algorithms result in a similar magnitude of goodness-of-fit, with the calibrated multi-model ensemble median providing fairly good predictions. Within the multi-model ensemble, non-linear models (Sigmoid and Wang) outperform near-linear ones, especially for those with the degree-day-based concept (GDD and Richardson). However, the former models tend to have more variable parameters than the latter ones when different algorithms are used, suggesting the link between parameter variability and model structure.

Overall, our study could provide new insights in order to better inform users about the importance of their choice when different optimization methods are available and when estimating cultivar-dependent parameters for multiple phenology models. Consequently, this study might contribute to formulating guidelines on the appropriate calibration practices for phenology models or crop models in general.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy13030679/s1>, Figure S1: Graphic representation of examples of different parameter sets provided for each studied phenology model; Figures S2–S4: Comparison between observed and simulated flowering DOY (day of year) by individual phenology model simulations with optimized parameter values from the Maximum Likelihood Estimation (MLE), Simulated Annealing (SA) and Shuffled Complex Evolution-University of Arizona (SCE-UA) respectively for Touriga Franca (TF) and Touriga Nacional (TN). Conventional statistics such as Mean Bias Error (MBE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and the coefficient of determination (R<sup>2</sup>) are calculated; Table S1: Selected parameters of phenology models for calibration with the specified initial lower and upper boundary of parameter distribution corresponds to the expanded baseline setting of each parameter range in Table 2 by 100%; Table S2: The optimized parameter values of applied phenology models under different algorithms given the observed flowering data for the variety of Touriga Franca (TF) and Touriga Nacional (TN). These are

obtained with the expanded lower and upper boundary of parameter distribution by 100%, relative to the baseline settings.

**Author Contributions:** Conceptualization, C.Y., C.M. and J.A.T.-M.; Data curation, C.Y., C.M. and J.A.T.-M.; Formal analysis, C.Y., C.M., S.R., N.M., J.A.S. and J.A.T.-M.; Funding acquisition, J.A.S.; Investigation, C.Y., C.M. and J.A.T.-M.; Methodology, C.Y., C.M. and J.A.T.-M.; Project administration, J.A.S.; Resources, S.R., N.M. and J.A.S.; Software, C.Y., C.M. and J.A.T.-M.; Supervision, C.Y., C.M., J.A.S. and J.A.T.-M.; Validation, C.Y., C.M., S.R., N.M., J.A.S. and J.A.T.-M.; Visualization, C.Y. and C.M.; Writing—original draft, C.Y.; Writing—review and editing, C.Y., C.M., S.R., N.M., J.A.S. and J.A.T.-M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by Clim4Vitis project—“Climate change impact mitigation for European viticulture: knowledge transfer for an integrated approach”, funded by the European Union’s Horizon 2020 Research and Innovation Programme, under grant agreement no. 810176; it was also supported and funded by FCT—Portuguese Foundation for Science and Technology, under the project UIDB/04033/2020.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are mainly contained within the article or supplementary material. Specifically, the summary of the observed phenology data is presented in Table 1; The link to the utilized gridded climate data can be found at [https://surfobs.climate.copernicus.eu/dataaccess/access\\_eobs.php](https://surfobs.climate.copernicus.eu/dataaccess/access_eobs.php) (accessed on 15 June 2022). Figures S2–S4 show the comparison between the observed and simulated data on the flowering DOY (Day of Year) by individual phenology model simulations with optimized parameter values using MLE, SA and SCE-UA, respectively.

**Acknowledgments:** All authors are grateful for ADVID and their associates, for their collaboration in data collection and provision under the Viticulture Observatory of the DDR.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Menzel, A.; Sparks, T.; Estrella, N.; Koch, E.; Aasa, A.; Ahas, R.; Alm-Kübler, K.; Bissolli, P.; Braslavská, O.; Briede, A.; et al. European phenological response to climate change matches the warming pattern. *Glob. Chang. Biol.* **2006**, *12*, 1969–1976. [[CrossRef](#)]
2. Piao, S.; Liu, Q.; Chen, A.; Janssens, I.A.; Fu, Y.; Dai, J.; Liu, L.; Lian, X.; Shen, M.; Zhu, X. Plant phenology and global climate change: Current progresses and challenges. *Glob. Chang. Biol.* **2019**, *25*, 1922–1940. [[CrossRef](#)] [[PubMed](#)]
3. Brisson, N.; Gary, C.; Justes, E.; Roche, R.; Mary, B.; Ripoche, D.; Zimmer, D.; Sierra, J.; Bertuzzi, P.; Burger, P.; et al. An overview of the crop model stics. *Eur. J. Agron.* **2003**, *18*, 309–332. [[CrossRef](#)]
4. Holzworth, D.P.; Huth, N.I.; deVoil, P.G.; Zurcher, E.J.; Herrmann, N.I.; McLean, G.; Chenu, K.; van Oosterom, E.J.; Snow, V.; Murphy, C.; et al. APSIM—Evolution towards a new generation of agricultural systems simulation. *Environ. Model. Softw.* **2014**, *62*, 327–350. [[CrossRef](#)]
5. Wallach, D.; Nissanka, S.P.; Karunaratne, A.S.; Weerakoon, W.M.W.; Thorburn, P.J.; Boote, K.J.; Jones, J.W. Accounting for both parameter and model structure uncertainty in crop model predictions of phenology: A case study on rice. *Eur. J. Agron.* **2017**, *88*, 53–62. [[CrossRef](#)]
6. Wallach, D.; Palosuo, T.; Thorburn, P.; Gourdain, E.; Asseng, S.; Basso, B.; Buis, S.; Crout, N.; Dibari, C.; Dumont, B.; et al. How well do crop modeling groups predict wheat phenology, given calibration data from the target population? *Eur. J. Agron.* **2021**, *124*, 126195. [[CrossRef](#)]
7. Gao, Y.; Wallach, D.; Liu, B.; Dingkuhn, M.; Boote, K.J.; Singh, U.; Asseng, S.; Kahveci, T.; He, J.; Zhang, R.; et al. Comparison of three calibration methods for modeling rice phenology. *Agric. For. Meteorol.* **2020**, *280*, 107785. [[CrossRef](#)]
8. Kawakita, S.; Takahashi, H.; Moriya, K. Prediction and parameter uncertainty for winter wheat phenology models depend on model and parameterization method differences. *Agric. For. Meteorol.* **2020**, *290*, 107998. [[CrossRef](#)]
9. Parker, A.; de Cortázar-Atauri, I.G.; Chuine, I.; Barbeau, G.; Bois, B.; Boursiquot, J.-M.; Cahurel, J.-Y.; Claverie, M.; Dufourcq, T.; Génay, L.; et al. Classification of varieties for their timing of flowering and veraison using a modelling approach: A case study for the grapevine species *Vitis vinifera* L. *Agric. For. Meteorol.* **2013**, *180*, 249–264. [[CrossRef](#)]
10. Reis Pereira, M.; Ribeiro, H.; Abreu, I.; Eiras-Dias, J.; Mota, T.; Cunha, M. Predicting the flowering date of Portuguese grapevine varieties using temperature-based phenological models: A multi-site approach. *J. Agric. Sci.* **2018**, *156*, 865–876. [[CrossRef](#)]
11. Parker, A.K.; De Cortázar-Atauri, I.G.; Van Leeuwen, C.; Chuine, I. General phenological model to characterise the timing of flowering and veraison of *Vitis vinifera* L. *Aust. J. Grape Wine Res.* **2011**, *17*, 206–216. [[CrossRef](#)]

12. Leolini, L.; Costafreda-Aumedes, S.; Santos, J.A.; Menz, C.; Fraga, H.; Molitor, D.; Merante, P.; Junk, J.; Kartschall, T.; Destrac-Irvine, A.; et al. Phenological Model Intercomparison for Estimating Grapevine Budbreak Date (*Vitis vinifera* L.) in Europe. *Appl. Sci.* **2020**, *10*, 3800. [[CrossRef](#)]
13. Wallach, D.; Palosuo, T.; Thorburn, P.; Hochman, Z.; Gourdain, E.; Andrianasolo, F.; Asseng, S.; Basso, B.; Buis, S.; Crout, N.; et al. The chaos in calibrating crop models: Lessons learned from a multi-model calibration exercise. *Environ. Model. Softw.* **2021**, *145*, 105206. [[CrossRef](#)]
14. Yun, K.; Hsiao, J.; Jung, M.-P.; Choi, I.-T.; Glenn, D.M.; Shim, K.-M.; Kim, S.-H. Can a multi-model ensemble improve phenology predictions for climate change studies? *Ecol. Model.* **2017**, *362*, 54–64. [[CrossRef](#)]
15. Taylor, S.D.; White, E.P. Automated data-intensive forecasting of plant phenology throughout the United States. *Ecol. Appl.* **2020**, *30*, e02025. [[CrossRef](#)]
16. Yang, C.; Ceglar, A.; Menz, C.; Martins, J.; Fraga, H.; Santos, J.A. Performance of seasonal forecasts for the flowering and veraison of two major Portuguese grapevine varieties. *Agric. For. Meteorol.* **2023**, *331*, 109342. [[CrossRef](#)]
17. Santos, J.A.; Fraga, H.; Malheiro, A.C.; Moutinho-Pereira, J.; Dinis, L.T.; Correia, C.; Moriondo, M.; Leolini, L.; Dibari, C.; Costafreda-Aumedes, S.; et al. A review of the potential climate change impacts and adaptation options for European viticulture. *Appl. Sci.* **2020**, *10*, 3092. [[CrossRef](#)]
18. Yang, C.; Fraga, H.; van Ieperen, W.; Trindade, H.; Santos, J.A. Effects of climate change and adaptation options on winter wheat yield under rainfed Mediterranean conditions in southern Portugal. *Clim. Chang.* **2019**, *154*, 159–178. [[CrossRef](#)]
19. Fraga, H.; García de Cortázar Aauri, I.; Malheiro, A.C.; Santos, J.A. Modelling climate change impacts on viticultural yield, phenology and stress conditions in Europe. *Glob. Chang. Biol.* **2016**, *22*, 3774–3788. [[CrossRef](#)]
20. Yang, C.; Fraga, H.; Van Ieperen, W.; Santos, J.A. Modelling climate change impacts on early and late harvest grassland systems in Portugal. *Crop Pasture Sci.* **2018**, *69*, 821–836. [[CrossRef](#)]
21. Tao, F.; Rötter, R.P.; Palosuo, T.; Gregorio Hernández Díaz-Ambrona, C.; Mínguez, M.I.; Semenov, M.A.; Kersebaum, K.C.; Nendel, C.; Specka, X.; Hoffmann, H.; et al. Contribution of crop model structure, parameters and climate projections to uncertainty in climate change impact assessments. *Glob. Chang. Biol.* **2018**, *24*, 1291–1307. [[CrossRef](#)] [[PubMed](#)]
22. Rosenzweig, C.; Jones, J.W.; Hatfield, J.L.; Ruane, A.C.; Boote, K.J.; Thorburn, P.; Antle, J.M.; Nelson, G.C.; Porter, C.; Janssen, S.; et al. The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agric. For. Meteorol.* **2013**, *170*, 166–182. [[CrossRef](#)]
23. Asseng, S.; Ewert, F.; Rosenzweig, C.; Jones, J.W.; Hatfield, J.L.; Ruane, A.C.; Boote, K.J.; Thorburn, P.J.; Rötter, R.P.; Cammarano, D.; et al. Uncertainty in simulating wheat yields under climate change. *Nat. Clim. Chang.* **2013**, *3*, 827–832. [[CrossRef](#)]
24. Wallach, D. Crop Model Calibration: A Statistical Perspective. *Agron. J.* **2011**, *103*, 1144–1151. [[CrossRef](#)]
25. Liu, L.; Wallach, D.; Li, J.; Liu, B.; Zhang, L.; Tang, L.; Zhang, Y.; Qiu, X.; Cao, W.; Zhu, Y. Uncertainty in wheat phenology simulation induced by cultivar parameterization under climate warming. *Eur. J. Agron.* **2018**, *94*, 46–53. [[CrossRef](#)]
26. Seidel, S.J.; Palosuo, T.; Thorburn, P.; Wallach, D. Towards improved calibration of crop models—Where are we now and where should we go? *Eur. J. Agron.* **2018**, *94*, 25–35. [[CrossRef](#)]
27. Chapagain, R.; Remenyi, T.A.; Harris, R.M.B.; Mohammed, C.L.; Huth, N.; Wallach, D.; Rezaei, E.E.; Ojeda, J.J. Decomposing crop model uncertainty: A systematic review. *Field. Crops Res.* **2022**, *279*, 108448. [[CrossRef](#)]
28. Myung, I.J. Tutorial on maximum likelihood estimation. *J. Math. Psychol.* **2003**, *47*, 90–100. [[CrossRef](#)]
29. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. Optimization by Simulated Annealing. *Science* **1983**, *220*, 671–680. [[CrossRef](#)]
30. Duan, Q.; Sorooshian, S.; Gupta, V.K. Optimal use of the SCE-UA global optimization method for calibrating watershed models. *J. Hydrol.* **1994**, *158*, 265–284. [[CrossRef](#)]
31. Yang, C.; Menz, C.; Fraga, H.; Reis, S.; Machado, N.; Malheiro, A.C.; Santos, J.A. Simultaneous Calibration of Grapevine Phenology and Yield with a Soil–Plant–Atmosphere System Model Using the Frequentist Method. *Agronomy* **2021**, *11*, 1659. [[CrossRef](#)]
32. Ran, H.; Kang, S.; Hu, X.; Yao, N.; Li, S.; Wang, W.; Galdos, M.V.; Challinor, A.J. A framework to quantify uncertainty of crop model parameters and its application in arid Northwest China. *Agric. For. Meteorol.* **2022**, *316*, 108844. [[CrossRef](#)]
33. Jones, G.V.; Alves, F. Impact of climate change on wine production: A global overview and regional assessment in the Douro Valley of Portugal. *Int. J. Glob. Warm.* **2012**, *4*, 383–406. [[CrossRef](#)]
34. Fraga, H.; García de Cortázar Aauri, I.; Malheiro, A.C.; Moutinho-Pereira, J.; Santos, J.A. Viticulture in Portugal: A review of recent trends and climate change projections. *OENO One* **2017**, *51*, 61–69. [[CrossRef](#)]
35. Santos, J.A.; Ceglar, A.; Toreti, A.; Prodhomme, C. Performance of seasonal forecasts of Douro and Port wine production. *Agric. For. Meteorol.* **2020**, *291*, 108095. [[CrossRef](#)]
36. Rebelo, J.; Caldas, J.; Guedes, A. The Douro Region: Wine and Tourism. *Almatour.-J. Tour. Cult. Territ. Dev.* **2015**, *6*, 75–90. [[CrossRef](#)]
37. Fraga, H.; Costa, R.; Moutinho-Pereira, J.; Correia, C.M.; Dinis, L.T.; Gonçalves, I.; Silvestre, J.; Eiras-Dias, J.; Malheiro, A.C.; Santos, J.A. Modeling phenology, water status, and yield components of three Portuguese grapevines using the STICS crop model. *Am. J. Enol. Vitic.* **2015**, *66*, 482–491. [[CrossRef](#)]
38. Meier, U. *Growth Stages of Mono- and Dicotyledonous Plants*; Blackwell Wissenschafts-Verlag: Berlin, Germany, 2001; ISBN 3826331524.
39. Cornes, R.C.; van der Schrier, G.; van den Besselaar, E.J.M.; Jones, P.D. An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. *J. Geophys. Res. Atmos.* **2018**, *123*, 9391–9409. [[CrossRef](#)]

40. Yang, C.; Fraga, H.; van Ieperen, W.; Santos, J.A. Assessing the impacts of recent-past climatic constraints on potential wheat yield and adaptation options under Mediterranean climate in southern Portugal. *Agric. Syst.* **2020**, *182*, 102844. [[CrossRef](#)]
41. Yang, C.; Menz, C.; De Abreu Jaffe, M.S.; Costafreda-Aumedes, S.; Moriondo, M.; Leolini, L.; Torres-Matallana, A.; Molitor, D.; Junk, J.; Fraga, H.; et al. Projections of Climate Change Impacts on Flowering-Veraison Water Deficits for Riesling and Müller-Thurgau in Germany. *Remote Sens.* **2022**, *14*, 1519.
42. Yang, C.; Fraga, H.; Ieperen, W.V.; Santos, J.A. Assessment of irrigated maize yield response to climate change scenarios in Portugal. *Agric. Water Manag.* **2017**, *184*, 178–190. [[CrossRef](#)]
43. Rodrigues, P.; Pedroso, V.; Reis, S.; Yang, C.; Santos, J.A. Climate change impacts on phenology and ripening of cv. Touriga Nacional in the Dão wine region, Portugal. *Int. J. Climatol.* **2022**, *42*, 7117–7132. [[CrossRef](#)]
44. Yang, C.; Menz, C.; Fraga, H.; Costafreda-Aumedes, S.; Leolini, L.; Ramos, M.C.; Molitor, D.; van Leeuwen, C.; Santos, J.A. Assessing the grapevine crop water stress indicator over the flowering-veraison phase and the potential yield lose rate in important European wine regions. *Agric. Water Manag.* **2022**, *261*, 107349. [[CrossRef](#)]
45. Bonhomme, R. Bases and limits to using ‘degree.day’ units. *Eur. J. Agron.* **2000**, *13*, 1–10. [[CrossRef](#)]
46. Richardson, E.; Seeley, S.; Walker, D. A model for estimating the completion of rest for “Redhaven” and “Elberta” peach trees. *HortScience* **1974**, *9*, 331–332. [[CrossRef](#)]
47. Hänninen, H. Modelling bud dormancy release in trees from cool and temperate regions. *Acta For. Fenn.* **1990**, *213*, 1–47. [[CrossRef](#)]
48. Chuine, I.; de Cortazar-Atauri, I.G.; Kramer, K.; Hänninen, H. *Plant Development Models—Phenology: An Integrative Environmental Science*; Schwartz, M.D., Ed.; Springer: Dordrecht, The Netherlands, 2013; pp. 275–293; ISBN 978-94-007-6925-0.
49. Wang, E.; Engel, T. Simulation of phenological development of wheat crops. *Agric. Syst.* **1998**, *58*, 1–24. [[CrossRef](#)]
50. Costa, R.; Fraga, H.; Fonseca, A.; De Cortázar-Atauri, I.G.; Val, M.C.; Carlos, C.; Reis, S.; Santos, J.A. Grapevine phenology of cv. Touriga Franca and Touriga Nacional in the Douro wine region: Modelling and climate change projections. *Agronomy* **2019**, *9*, 210. [[CrossRef](#)]
51. Reis, S.; Fraga, H.; Carlos, C.; Silvestre, J.; Eiras-Dias, J.; Rodrigues, P.; Santos, J.A. Grapevine phenology in four portuguese wine regions: Modeling and predictions. *Appl. Sci.* **2020**, *10*, 3708. [[CrossRef](#)]
52. Hastings, W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97–109. [[CrossRef](#)]
53. Houska, T.; Kraft, P.; Chamorro-Chavez, A.; Breuer, L. SPOTting Model Parameters Using a Ready-Made Python Package. *PLoS ONE* **2015**, *10*, e0145180. [[CrossRef](#)] [[PubMed](#)]
54. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092. [[CrossRef](#)]
55. Duan, Q.Y.; Gupta, V.K.; Sorooshian, S. Shuffled complex evolution approach for effective and efficient global minimization. *J. Optim. Theory Appl.* **1993**, *76*, 501–521. [[CrossRef](#)]
56. Nelder, J.A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308–313. [[CrossRef](#)]
57. Price, W.L. Global optimization algorithms for a CAD workstation. *J. Optim. Theory Appl.* **1987**, *55*, 133–146. [[CrossRef](#)]
58. Holland, J.H. *Adaptation in Natural and Artificial Systems*; The University of Michigan Press: Ann Arbor, MI, USA, 1975; ISBN 9780262581110.
59. Duan, Q.; Sorooshian, S.; Gupta, V. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.* **1992**, *28*, 1015–1031. [[CrossRef](#)]
60. Akaike, H. *Information Theory and an Extension of the Maximum Likelihood Principle—Selected Papers of Hirotugu Akaike*; Parzen, E., Tanabe, K., Kitagawa, G., Eds.; Springer: New York, NY, USA, 1998; pp. 199–213; ISBN 978-1-4612-1694-0.
61. Beven, K. A manifesto for the equifinality thesis. *J. Hydrol.* **2006**, *320*, 18–36. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.