land

Article

SDM- and GIS-Based Prediction of Citrus Suitability in Southern Italy: Evaluating the Influence of Local Versus Global Climate Datasets

Giuseppe Antonio Catalano [1], Provvidenza Rita D'Urso [2,*] and Claudia Arcidiacono [2]

[1] Department of Veterinary Sciences, University of Messina, Polo Universitario dell'Annunziata Viale Giovanni Palatucci n. 23, 98168 Messina, Italy; giuseppe.catalano@studenti.unime.it
[2] Department of Agriculture, Food and Environment, University of Catania, Via S. Sofia n. 100, 95123 Catania, Italy; carcidi@unict.it
* Correspondence: provvidenza.durso@unict.it

Abstract

This study investigated the application of Species Distribution Models (SDMs), based on Boosted Regression Tree (BRT) and Random Forest (RF), to predict the distribution of citrus crops in a Mediterranean climate by comparing climate data from WorldClim with those from the Regional Territorial Information System of Sicily (S.I.T.R.). To this aim, 19 bioclimatic variables were calculated from monthly temperature and precipitation data in the period 2003–2021 by using the biovars package in R software version 2023.12.0+369. Soil properties, terrain elevation, slope, and soil water retention capacity were considered to adequately simulate pedoclimatic conditions in the Syracuse area in Sicily (Italy). The SDM algorithms performed well (AUC: 0.84–0.93; TSS: 0.51–0.69), and Random Forest was selected to compare global and local outcomes. Using data from local meteorological stations increased the model's reliability, resulting in a difference of approximately ~800 ha in the predicted citrus distribution compared to WorldClim data. This approach also provided a more accurate representation of precipitation patterns, for instance, in the municipality of Augusta, where WorldClim underestimated the average annual rainfall by 284 mm. These findings emphasise the importance of incorporating local environmental data into SDMs to improve prediction accuracy and inform future hybrid approaches to enhance model robustness in the context of climate change. Finally, the results contribute to expanding knowledge of citrus soil and climate conditions, with potential implications for land-use planning.

Keywords: precision agriculture; local climate data; citrus distribution; Geographic Information Systems; predictors' territorial analysis; regional climate models

Academic Editor: Nir Krakauer

Received: 29 August 2025
Revised: 5 November 2025
Accepted: 6 November 2025
Published: 10 November 2025

Citation: Catalano, G.A.; D'Urso, P.R.; Arcidiacono, C. SDM- and GIS-Based Prediction of Citrus Suitability in Southern Italy: Evaluating the Influence of Local Versus Global Climate Datasets. *Land* **2025**, *14*, 2223. https://doi.org/10.3390/land14112223

1. Introduction

Climate change is having a significant impact on various human activities, particularly agriculture, influencing crop production and sustainability. The latest report by the Intergovernmental Panel on Climate Change [1] highlights the likelihood of climate imbalances in the context of climate change. The increase in greenhouse gases (GHGs) in the atmosphere, particularly carbon dioxide (up by about 47%) and methane (up by about 156%) since 1750, has significantly amplified extreme weather events, such as heat waves and heavy rainfall. This has caused a substantial increase in mean annual temperatures by 1.4 °C in the Mediterranean region, from 1880 to 2018 [2]. In Sicily, annual precipitation is characterised by a decreasing trend, with a reduction from −5.2 to −1.5 mm/year from 1921 to 2012 [3]. In the Mediterranean area, increasing average temperature and rainfall concentrations, as well as drought events and the decrease in water availability in the Sicilian territory, have become common conditions.

Within the territory of Syracuse, citrus orchards, mainly consisting of *Citrus limon* (L.) Burm. f., *Citrus* × *clementina* and *Citrus sinensis* (L.) Osbeck, constitute the main type of agricultural production, covering 23,250 hectares and producing 528,500 tonnes in 2025, according to ISTAT data [4]. However, in recent years climatic conditions have become extreme. In 2024, SIAS weather bulletins [5] reported that annual precipitation in Sicily averaged 453 mm, falling below the critical threshold of 500 mm for the first time since the severe 2002 drought, when the regional average was 415 mm. Areas in central-eastern and central-southern Sicily were particularly affected, with annual rainfall dropping below 300 mm and deficits exceeding 60%. In the province of Syracuse, total precipitation amounted to around 400 mm, which consequently reduced annual production.

Therefore, targeted research is essential to evaluate the climatic suitability of different geographical areas for citrus species under changing climatic conditions. Such studies are fundamental for optimising resource use and ensuring that crop requirements are met. Within this context, the present study aims at exploring the potential implications of climate change for citrus cultivation and at enhancing the availability of reliable information to support agricultural planning.

In the literature, Species Distribution Models (SDM) have become increasingly popular in recent decades [6–9] for their capability of simulating potential species distribution in current scenarios, by combining data on species presence with environmental characteristics including weather data. In this field of research, VisTrails:SAHM software (Version VisTrails 2.2.3 SAHM 2.0.1) has been initially applied by Talbert et al. [10] and Morisette et al. [11] and subsequently widely considered in research studies in the literature [12,13]. VisTrails:SAHM is an open-source management and scientific workflow system designed to integrate both scientific workflow and scientific visualisation systems. The Software for Assisted Habitat Modelling (SAHM) was developed to build the habitat modelling process into modules and provide a comprehensive record of the operations performed. The key advantage of SAHM is its integration of multiple machine learning and geostatistical algorithms, allowing for the identification of the best-performing models for the dataset. SAHM facilitates the management of input data, the pre-processing and post-processing phases, and the modelling options used in the creation of SDMs, while VisTrails supports visualisation of the results and folder management.

Meteorological data models, such as WorldClim (https://www.worldclim.org/, accessed on 16 July 2020) or CMCC Data Delivery System (https://dds.cmcc.it/#/, accessed on 20 July 2020), are often used in species distribution modelling and related ecological modelling techniques [8,14]. However, in the literature, the SDM is often limited by the availability of input data, and the spatial resolution of the available data is often too coarse to depict environmental processes at relevant scales [6]. The 19 bioclimatic variables available in those databases are the most commonly used set of climate variables for creating SDM and ecological niche models (ENM). In the literature, the applications of climatic biovariables are numerous. Several authors have used them to analyse the distribution of plants and animals. Furthermore, the "biovars" function in the "dismo" package has been widely used to create bioclimatic variables based on the needs of the specific study.

In detail, Akpoti et al. [15] conducted a study in West Africa to identify the most suitable areas for rice production to achieve self-sufficiency for the needs of Togo and Benin. The analysis of the WorldClim 1970–2000 bioclimatic variables has identified several predictors having a significant impact. These predictors included altitude, isotherm, rainfall in the wettest quarter, annual rainfall, total phosphorus, soil water retention capacity, bulk density, and the proximity of inland valleys to roads and urban centres. In a recent study, Pinilla-Buitrago [16] predicted a change in the distribution limits of *Cryptotis mexicanus*, a montane shrew, in response to climate change. The study computed the bioclimatic variables for the period 1981–2010 by using the 'biovars' function (dismo R package) with monthly data from the CHELSA v2.1 database. Finally, variation in biovariables at different global and local scales was the focus of a recent study by Kyrgyzbay et al. [17], while no application of SDM was carried out. Open-source global climate data (WCLIM) were considered and high-resolution data on temperature and precipitation were found essential for understanding and predicting the impact of climate on the structure and functions of terrestrial ecosystems. Interest in these applications has risen, in specific projects under PNRR funding, such as Spokes 2 and 3 of AGRITECH National Center [18–20], with the aim to improve resource management and reduce pesticide use. In previous research of the authors [18,19], the feasibility of using Species Distribution Models (SDMs), based on WorldClim data for bioclimatic variables, to calculate the probability of citrus presence in a Mediterranean area was evaluated. The models' parameters were fine-tuned to optimise algorithm performance, and the best models were selected based on metric assessments. In this context, the use of local data could be of paramount importance for refining model input data and verify how the models' output would change.

Based on the analysis of the literature, it was acknowledged how several studies have investigated the application of global datasets for predicting species distribution, whereas few studies have focused on the application of local ones by using data acquired by several weather stations.

Therefore, the aim of this study was to compare the application of SDMs based on either local or global meteorological data to predict the potential distribution of species in a specific territory. In detail, this study answers the following research questions:

- RQ1: What is the performance of different SDM algorithms based on local and global climate?
- RQ2: How does the use of different climate data sources (i.e., global and local data) affect the probability maps?
- RQ3: Which covariates have the greatest effect on the prediction?
- RQ4: What kind of differences in spatial distribution of bioclimatic variables can be identified when comparing the different approaches (global vs. local data)?

Specifically, RQ1 and RQ2 are mainly addressed in Section 3.1, while RQ3's answers are described in Sections 3.2 and 3.3, and RQ4 is considered in Section 3.4.

2. Materials and Methods

The study was developed using a four-phase methodology (Figure 1) where climate suitability maps were obtained and correlation and contribution indices were utilised to address the research questions. In the first phase, spatial data acquisition and processing were carried out by using Geographic Information System (GIS) tools and R software version 2023.12.0+369. The second step included the definition of a specific pipeline of VisTrails:SAHM software (version VisTrails 2.2.3 SAHM 2.0.1) modules to produce predictions of citrus potential distribution in the considered territory, as well as the SDM assessment through accuracy measures [20]. In the third step, SDM selection was carried out, and the SDM results based on the two different climate models were evaluated by using specific metrics, both for correlated and uncorrelated predictors. Finally, species distribution results of the two simulations, as well as biovariables, were mapped and overlaid by using GIS tools to facilitate spatial comparisons and quantification of the surfaces. To this end, a combination of different software was profitable since the elaborations of

VisTrails:SAHM results by GIS tools facilitated the computation of the predictors' impact on simulation outcomes.

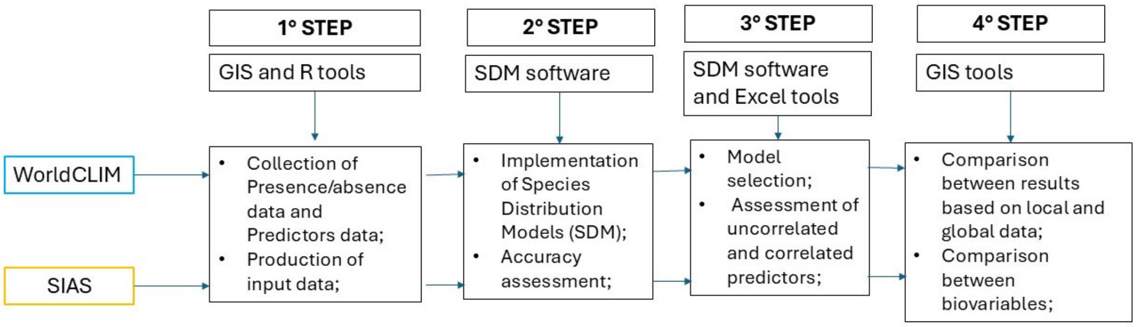


Figure 1. Methodology framework.

2.1. Study Area and Citrus Species Presence Data

The study was applied to the province of Syracuse (2124 km^2), located in Sicily, Italy (Figure 2). Situated in the south-eastern areas of Sicily, the province is characterised by small plains in the coastal areas to the East and the hilly and mountainous system constituting the Iblean Mountains, which includes the highest point, Monte Lauro, at an altitude of 986 m above sea level (m a.s.l.). The territory extends to the North into the Catania plain, encompassing an area of approximately 43,000 hectares, which originated from alluvional deposits [21]. According to the Köppen classification, climate classification system, Sicily has a temperate humid climate (type C), with an average temperature below 18 °C in the coldest month but above −3 °C. More specifically, the region has a humid subtropical mesothermal climate with dry summers (type Csa) [22]. This is a typical Mediterranean climate, characterised by an average temperature of the warmest month above 22 °C and a distinctive precipitation pattern [22].

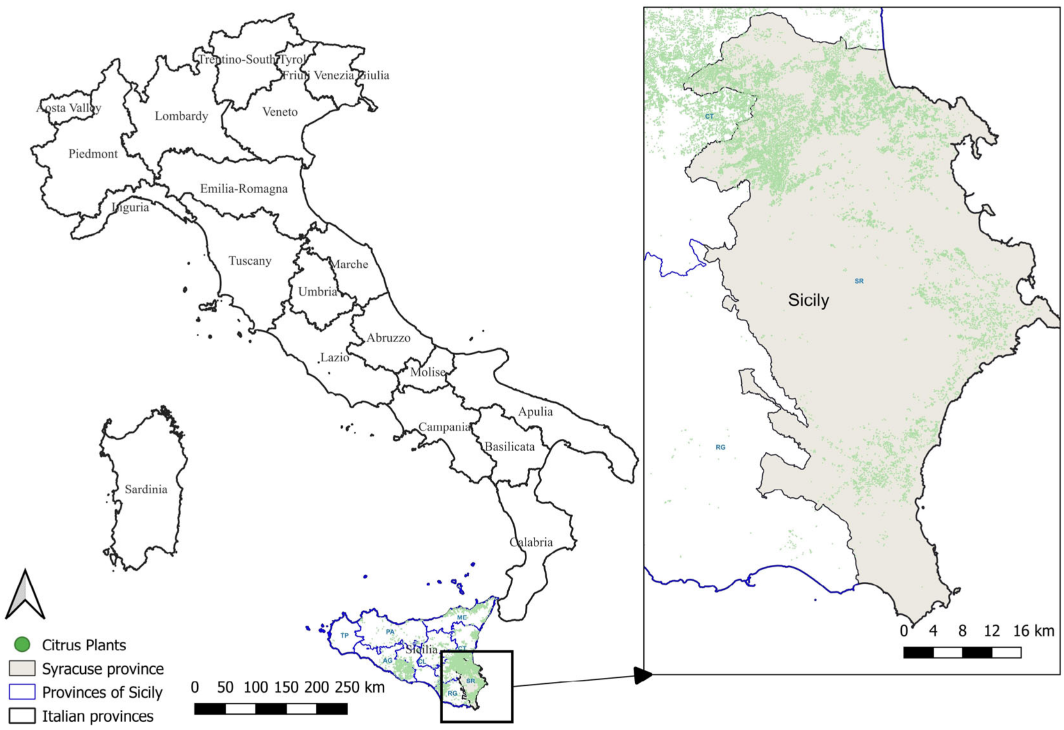


Figure 2. Study area localisation within Sicily (Italy) and citrus presence points from SITR (in green).

The number of species presence points is a crucial factor for achieving research objectives, as noted in previous studies [18,23–25]. In this study, the presence points were identified by using the vector file available in the Regional Territorial Information System of Sicily (SITR) platform [26]. The presence points were generated from the shapefile of the Regional Technical Map (2012–2013), available for download in the SITR platform. In this study, the shapefile was clipped by using the Syracuse province boundary shapefile layer, and 12,299 points were identified and represented in UTM WGS84 coordinates (Figure 2). This dataset was used as input for VisTrails:SAHM. Since historical data suitable for generating pseudoabsence points were unavailable, they were not included in the dataset. Moreover, the pseudoabsence points were considered unreliable due to possible human activities, such as the removal of citrus plants for reasons unrelated to unsuitability for cultivation, such as phytopathologies [27]. Therefore, background points were used for the calibration of the models [18].

2.2. Production of Bioclimatic Variables

The climatic variables were acquired from two different databases, a global one and a local one. The global database WorldClim [28], named "WCLIM" hereafter, provides data acquired from monthly averaged temperature and precipitation records at 1 km spatial resolution [25]. The local database was built based on the data acquired by the Regional Territorial Information System of Sicily [5], named "SIAS" hereafter. The SIAS data network encompasses 96 stations distributed across Sicily, yet only 94 of the stations were considered, as 2 of them were not operational at the time of deployment or were discontinued shortly thereafter, thus being inadequate for representing the entire period under study. Hence, 16 meteorological stations located in the Syracuse province and in its neighbouring areas were selected; in particular they were integrated in the area of the Catania plain, which is known to be an area suitable for the cultivation of citrus fruits. Following the guidelines of O'Donnel [29], the daily climate data from each weather station were aggregated to create a monthly dataset. The climate data were then interpolated by using the thin-plate spline method [30], with a spatial resolution of 20 m.

Based on the acquisition of WCLIM and SIAS climate data, 19 bioclimatic variables (Table S1) have been considered in this study. In detail, the bioclimatic variables consisted of two sets of 19 global grid data fields derived from average monthly temperature and precipitation measurements collected by the weather stations in the study area at global and local levels (WCLIM and SIAS, respectively). Bioclimatic variables capturing aspects of temperature, precipitation and seasonality were then derived from the interpolated monthly values of temperature and precipitation measurements. In detail, bioclimatic variables from Bio 1 to Bio 7 are related to temperature, while bioclimatic variables from Bio 12 to Bio 15 are rainfall-dependent. The remaining variables, i.e., from Bio 8 to Bio 11, are quarterly indices as they are based on 3-month intervals. To produce the indices related to the end of the year, the quarterly period was defined by using the months at the beginning of the following year. For example, if December is the target month, January and February of the same average time period were included in the quarterly period [29]. Finally, the variables Bio 8, Bio 9, Bio 18 and Bio 19 are also referred to as 'combined' as they are the only ones that depend on both temperature and precipitation parameters.

The application of predictive algorithms for the simulation of species occurrence in the study area involved the use of established tools for generating predictive variables. In this study, the bioclimatic variables were computed in the time interval 2003–2021 by using the function biovars from the 'dismo' R package. The efficacy of the 'dismo' package for generating bioclimatic variables has previously been demonstrated in studies, such as that conducted by Schwager and Berg [7] on the distribution of Alpine plant species.

In the remainder of the text, 'SIAS_{Sim}' represents the results obtained by using SIAS data in SDM simulations and, similarly, '$\text{WCLIM}_{\text{Sim}}$' indicates the results obtained by using WorldClim data in the simulations.

2.3. Soil-Related Input Data

The edaphic conditions of the study area were added and analysed based on the importance that some authors [31,32] attributed to them. Among these, the following soil variables, were taken into account:

- Digital Terrain Model (DTM): The DTM at a 20 m resolution was obtained from the SITR geodatabase through the Web Feature Service (WFS). This digital terrain model provides detailed information on the elevation and topography of the study area, allowing a better understanding of the terrain morphology [33].
- Terrain Slope (Slope_{T}): The slope of the terrain was calculated by using the "Slope" tool available in QGIS 3.22.1 applied to the considered DTM. This parameter indicates the slope of the terrain and facilitates acquisition of knowledge on how topography influences the distribution of plant species [34].
- Volume of water in soil at a −60 cm depth, at soil saturation of −10 kPa (named $\text{V}_{\text{w-10}}$ hereafter), at field capacity of −33 kPa (named $\text{V}_{\text{w-33}}$ hereafter), and at permanent wilting point of −1500 kPa (named $\text{V}_{\text{w-1500}}$ hereafter), obtained from Soilgrid250 (https://soilgrids.org/, accessed on 2 October 2023) [35]. In general, SoilGrids produces maps of soil properties for the entire globe at medium spatial resolution (250 m cell size) by using machine learning methods to generate the simulations (Poggio et al. [36]). According to Turek et al. [37], soil water retention controls multiple processes related to mass and energy cycling in the soil–plant–atmosphere system. Therefore, soil water retention data were considered suitable for building accurate SDM predictions.

2.4. Application of VisTrails:Sahm Modelling to the Case Study

SAHM's main work steps include data input, pre-processing, preliminary model analysis and decision, models, and output procedures.

In this study, a specific pipeline was designed to implement the proposed methodology (Figure 3).

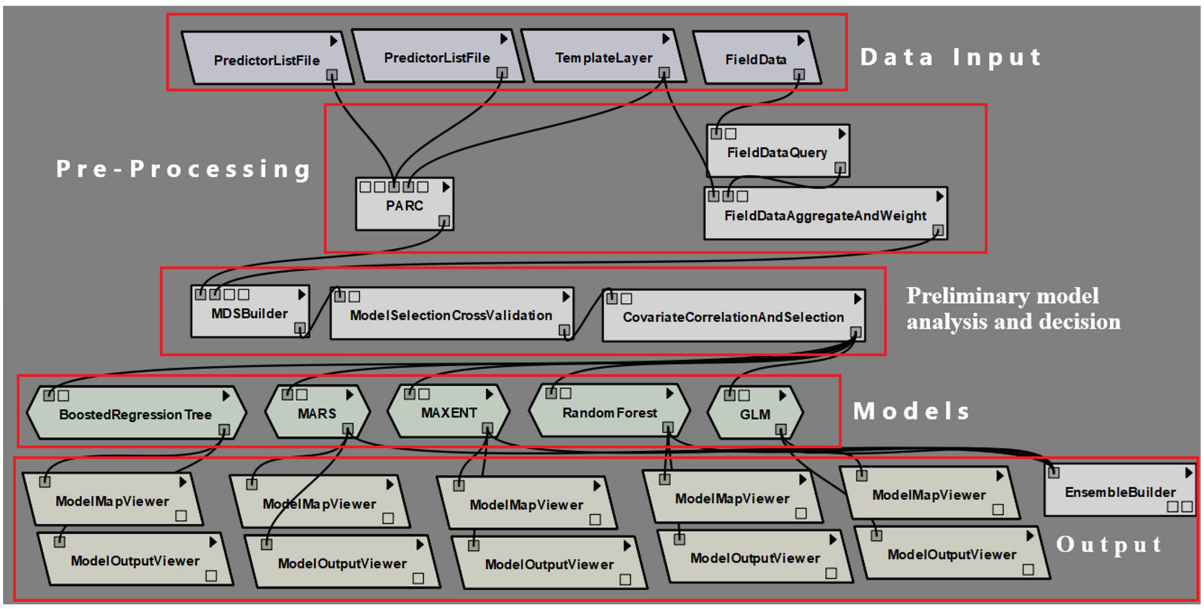


Figure 3. Pipeline specifically designed in VisTrails:SAHM for the proposed methodology.

Three main modules were used in the 'Input Data' phase of the process: 'PredictorListFile', 'TemplateLayer' and 'FieldData'. The 'PredictorListFile' and 'FieldData' allow the handling of predictor data in raster format and presence data in .csv format, respectively. In the 'TemplateLayer' module, a specially defined raster in .tiff format was loaded, with the following key characteristics to be applied to all rasters: reference system (WGS 84/UTM 33N, EPSG: 32633), resolution (20 m), and area of interest limited to the territory of Syracuse province.

In the pre-processing phase, presence data in .csv were checked by 'FieldDataAggregateandWeight' to assess whether they were within the considered study area and whether there was any overfitting due to overlapping occurrences in the same pixel. In addition, a key module utilised for optimising the input data processing was the PARC module. This tool makes it possible to perform a series of operations such as projection, aggregation, resampling and clipping on the input geospatial data in order to align it with the 'TemplateLayer' of the model.

In the preliminary model analysis and decision phase, the 'Merged Data Set Builder' (MDSBuilder) module was used at this stage to merge the presence and background data into a .csv file. In detail, 10,000 background data points were randomly collected within the 'TemplateLayer' boundaries. Then, the 'ModelSelectionCrossValidation' module estimates the error rate of input data by omitting some of the training data in the calibration process and then testing the model on the excluded observations [38]. In the remainder of the text, 'CV' represents the cross-validation method. The 'CovariateCorrelationSelector' module then allows the influence of each variable on the distribution of the sampled data to be assessed. This module makes it possible to identify and display additional information that helps remove variables that are highly correlated with others, allowing the user to examine the spatial relationship between the predictor and the response [10]. To this end, the maximum value of the PSK matrix coefficients, calculated for pairs of variables, was used. Specifically, in this study, the threshold for the three coefficients was set to ± 0.75 [39] for the input data (i.e., climate and soil variables). Since a high value of the coefficients means that there is a strong association between the two variables, in this case one of the two variables was considered and the other was discarded. This selection was also based on the contribution (%) of the variables to the simulation as produced by the 'CovariateCorrelationSelector' module [40]. Uncorrelated predictors were then used by the models to perform the predictions. This module enabled us to analyse the differences in the correlation of variables between crops and climate conditions. Section 3.2 provides an in-depth analysis of the differences in predictor selection in the two simulations. The SAHM software, version 2.0.1, includes the use of 5 SDM algorithms: MaxEnt, Boosted Regression Tree (BRT), Multivariate Adaptive Regression Splines (MARS), Generalised Linear Model (GLM), and Random Forest (RF).

In VisTrails:SAHM, output visualisation is provided by 'ModelMapViewer' and 'ModelOutputViewer' modules. These modules provide a suitable means for viewing the numerous spatial and various non-spatial and diagnostic outputs produced by each individual model. In this study, elaborations in GIS were carried out on the visualisation output of VisTrails. In detail, map algebra-based analyses were performed to extract information regarding differences in surface areas and their distribution in the territory, comparing the various model simulations.

2.5. Assessment of Simulations Carried out at Local and Global Levels

2.5.1. Assessment Through Metrics

The reliability of each SDM algorithm was evaluated by using suitable metrics such as True Skills Statistics (TSS) metrics, and area under the ROC curve (AUC) [8,41,42].

The Receiver Operating Characteristic (ROC) curve was constructed by evaluating all possible test values, calculating the proportion of true positives and diagnostics for each, according to D'arrigo [43]. The AUC, which can take values between 0.5 and 1.0, represents the area under the ROC curve. To interpret the values of the area under the ROC curve, refer to the following classification:

AUC = 0.5: the test is not informative since the value indicates that the test is no better than chance;

At $0.5 < \text{AUC} \leq 0.7$, the test is inaccurate;

At $0.7 < \text{AUC} \leq 0.9$, the test is moderately accurate;

At $0.9 < \text{AUC} < 1.0$, the test is highly accurate;

AUC = 1 signifies a perfect test.

The overfitted models lack generality and thus the model application in different contexts, such as to another region [44] or time period [45], or in research studies that aim at comparing models [46], could produce incorrect probability maps. Therefore, model overfitting should be taken into account when models are assessed. As stated by West et al. [12], overfitting occurs when a model demonstrates an excessive degree of fit to the dataset utilised for training, resulting in a notable decline in predictive capacity. If the discrepancy exceeds a specified threshold, e.g., the difference between AUC for training and CV ($\Delta\text{AUC} > 0.05$), the model can be described as possessing characteristics of overfitting [13].

The TSS metric is calculated on the elements of the confusion matrix, consisting in the number of correct and incorrect predictions for the presence and absence regions, defined as (sensitivity + specificity − 1), where sensitivity is the proportion of presence accurately predicted, whereas specificity is the proportion of absences accurately predicted [47].

Based on the analysis of the accuracy measures (AUC, ΔAUC, and TSS) for both $\text{WCLIM}_{\text{Sim}}$ and SIAS_{Sim}, the suitable SDM algorithm for this case study was selected (see Section 3.1). The results of the two simulations (i.e., $\text{WCLIM}_{\text{Sim}}$ and SIAS_{Sim}) were assessed in terms of different distribution of the species presence in the territory by using GIS tools. To this end, a combination of different software was profitable since the elaborations of VisTrails:SAHM results by GIS tools facilitated the computation of the predictors' impact on simulation outcomes.

In detail, maps related to the two simulations were overlaid in the software QGIS to identify differences in predicted species distribution. Specifically, the overlapping alternatives among datasets were analysed: areas consistent between the two simulations, areas predicted only by the SIAS_{Sim}, and areas predicted only by the $\text{WCLIM}_{\text{Sim}}$ (Figure 4).

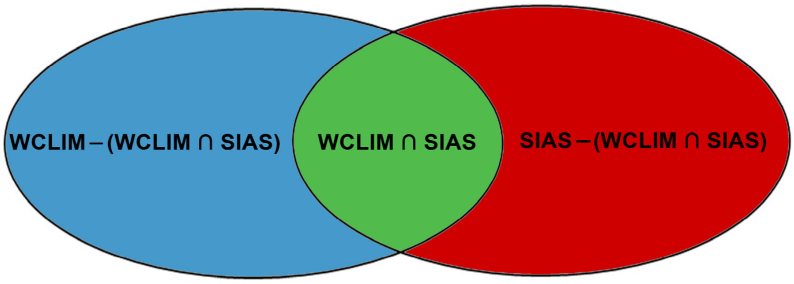


Figure 4. Overlapping alternatives among simulation's output areas based on datasets of WCLIM and SIAS: intersection (green area), and symmetric difference of the two output areas, highlighting the overestimation of the predicted area in the two simulations, i.e., blue area for $\text{WCLIM}_{\text{sim}}$ and red area for SIAS_{sim}.

Geospatial analyses were carried out for each municipality to quantify the differences in surface areas and to better understand how variations in predicted potential presence based on the two simulations may affect specific geographical areas.

In addition, a comparison between covariates selected by the 'CovariateCorrelationSelector' module was carried out to identify the main predictors for each simulation based on global and local data (i.e., WCLIM and SIAS). Results of this analysis are presented in Section 3.2.

2.5.2. Assessment Through Indices Based on Predictor Variables

This study investigated the contribution of the predictors to the SDM results based on the two simulations at global and local levels. In detail, since environmental and geographical variables can have an important influence on the presence or absence of a species or phenomenon in a given environment, it is crucial to assess which of these predictors are the most important in determining correlations within models. The PSK correlation index matrix and the MDA index provided by RF were utilised to this end. The use of the PSK matrix and the MDA index is corroborated by specific studies in the field [15,48] where they have been applied to identify the predictors that most influence the potential species presence. With regard to the first approach, a specific module of the VisTrails:SAHM software, namely the 'CovariateCorrelationSelector', produces the PSK correlation matrix and allows the analysis of correlations among the different predictor variables. In the second approach, the MDA assessment index generated by the RF model was considered to have information on the marginal importance of correlated variables. The MDA index calculates the importance of predictors by permuting 'out-of-bag' (OOB) samples. The OOB sample contains observations that were not used to build the current tree. It is used to estimate prediction error and assess variable importance [49]. Moreover, to evaluate the performance of RF algorithms, refs. [48,50] used the OOB error rate alongside the classic AUC and TSS metrics as an additional internal validation index. In particular, the OOB error rate, calculated on the observations excluded in each bootstrap, allowed them to estimate the model's predictive capacity without bias, eliminating the need for external test sets.

In this context, Section 3.2 illustrates how the selection of predictors varies according to the standard methodology, while Section 3.3 analyses how all predictors influence the PSK and MDA indices.

Results of the first and second approaches, based on PSK and MDA, are reported in Sections 3.3.1 and 3.3.2, respectively.

3. Results and Discussion

The Results and Discussion Section is organised into four subsections. In Section 3.1, the outcomes of the two simulations, based on global and local biovariables, are described and discussed in terms of SDM performance and assessment metrics. Based on the best SDM algorithm, predictors computed at local and global levels are analysed and compared in Section 3.2 to assess their adequacy in describing the species presence in the territory. Then, Section 3.3 quantifies how much predictors influenced species distribution by using the PSK matrix generated by the CovariateCorrelationSelector (Section 3.3.1) and the MDA index generated by the RF model (Section 3.3.2). In Section 3.4, the impact of predictors on simulation outcomes is analysed from the agronomic point of view at the territorial level to verify the coherence of the predictions with agronomic practices and pedoclimatic conditions.

3.1. SDM's Simulation Outcomes and Related Metrics

Table 1 reports the predictive performance of the models for both training and CV, showing high performance for AUC (0.84–0.93) and TSS (0.51–0.69), according to the threshold defined by D'Arrigo et al. [43]. Only the BRT model showed overfitting, since differences in AUC values for training and CV exceeded the threshold (ΔAUC > 0.05) in $WCLIM_{Sim}$. Furthermore, response curves identifying the contribution of biovariables within the model [19] were used to validate agronomic predictions for citrus distribution in the territory. Figure S3 shows the response curves for biovariables 5 and 7, which were obtained by using the RF model.

In this regard, numerous studies in the scientific literature on the performance of SDMs have been carried out to identify the optimal model. In many of these analyses, the RF, BRT and MAXENT models emerged as the best performing models [8,15,51,52]. Differences in model performance are often related to model complexity, as highlighted by [15], which is connected to the computational time required to fit the models. Models with longer run times tend to produce higher evaluation metrics. In our study, MARS, GLM, MAXENT, BRT and RF models showed increasing computational times, which were consistent with the observed differences in model performance (Table 1).

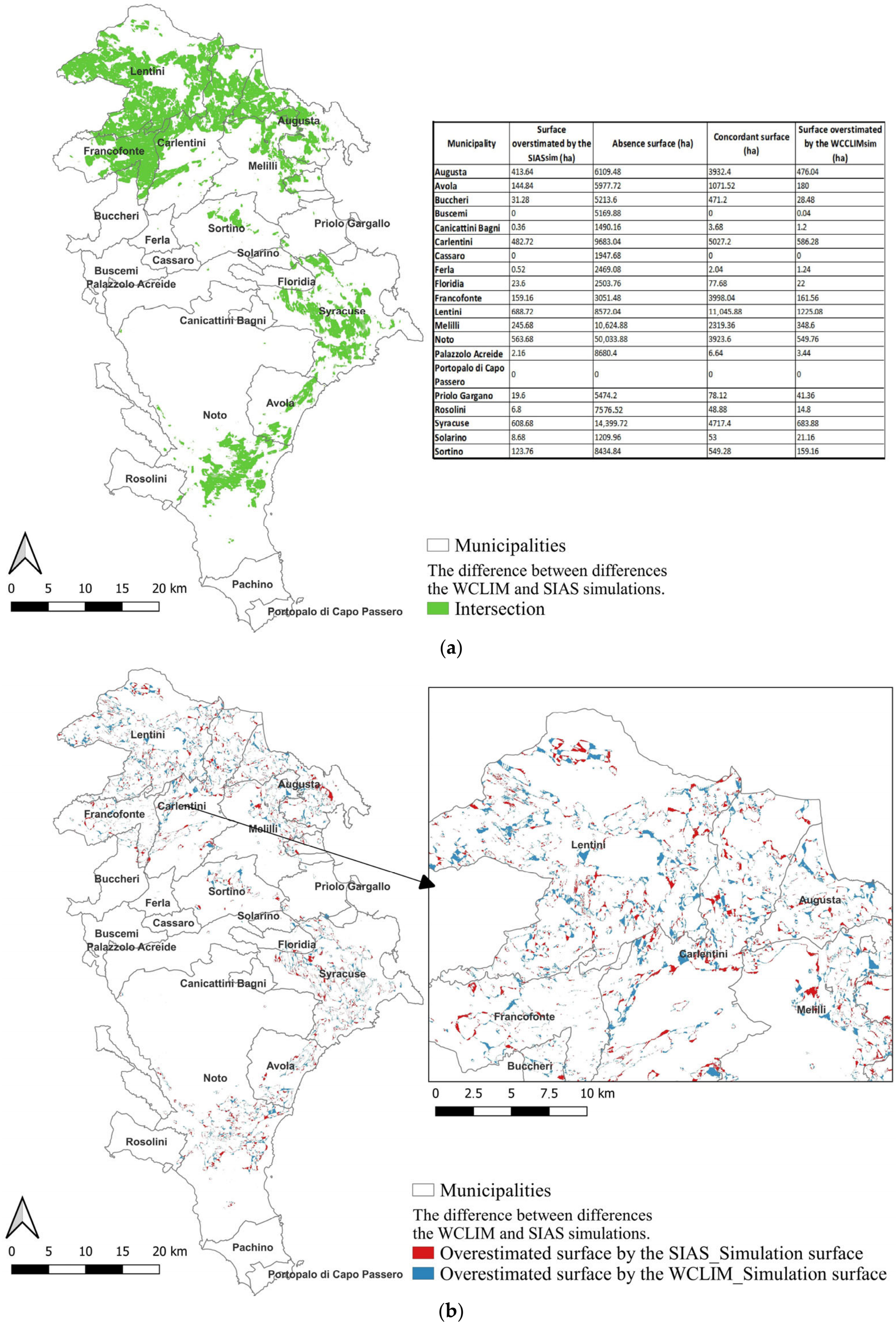
In this study, the BRT algorithm showed overfitting and produced anomalies in spatial prediction, as found in previous studies [18,19,53], and therefore it was excluded in the further analyses. Conversely, the RF model, though not achieving a very high performance (i.e., AUC did not exceed 0.90), was selected for its high metrics and could therefore be expected to be more reliable than the other models in predicting spatial species distribution. Furthermore, Roberts [54] found that the use of accuracy measures to evaluate model performance can sometimes lead to erroneous or difficult to interpret results. These errors can be caused by known, detectable factors, such as residual autocorrelation, or by factors that are more complex or more difficult to detect. This analysis highlights the importance of not relying solely on the results of model evaluation metrics. Indeed, it is critical to consider other factors, such as understanding the context and interpreting the results at the territorial level, to make a complete and accurate assessment of model performance. Another study by Huang [55] found that the change between the two datasets did not significantly affect model performance in terms of AUC and TSS. However, differences in species distribution predictions in the territory were found; therefore, the type of climate dataset used influenced the models in plant distribution. To assess reliability of territorial prediction, the predictors' geospatial distribution results were analysed and compared to the plant's physiological characteristics in order to answer research questions 2, 3, and 4.

Based on these outcomes, the RF model was selected to perform a deeper analysis on predictor selection, and we discuss the results of the spatial distribution of predicted plant presence. In detail, Figure 5 was produced to quantify the differences in the spatial distribution of the two simulations ($WCLIM_{sim}$ and $SIAS_{sim}$) by using the RF model. In Figure 5a, the green areas indicate the areas of intersection between the WCLIM and SIAS simulations, where both simulations agree on the distribution. Moreover, in Figure 5b the red and blue areas highlight the differences between the two simulations: the red areas represent the areas overestimated by the $SIAS_{Sim}$, while the blue areas show the areas overestimated by the $WCLIM_{sim}$. This analysis made it possible to spatially identify the differences in the prediction of the distribution areas between the two models in each municipality. The analysis conducted on the 'WCLIM' and 'SIAS' simulations revealed information regarding the agricultural potential in the province of Syracuse, with specific regard to the cultivation of citrus fruits. From the results obtained (Figure 5a), some areas of the municipalities of Lentini, Carlentini, Melilli, Syracuse, Avola and Noto were identified as particularly suitable for citrus growing due to their favourable pedoclimatic conditions. In the northern part of the Syracuse province, there is more land available for citrus cultivation than what is currently being used; i.e., there is potential for expanding agricultural crops in this area to meet the demands of citrus fruit production and support the local agricultural industry.

Table 1. Accuracy measures of the two simulations ($SIAS_{Sim}$ and $WCLIM_{Sim}$) for the different SDM algorithms and results of Δ AUC and TSS.

$SIAS_{Sim}$																			
BRT				*MARS*				*MAXENT*				*GLM*				*RF*			
AUC		*TSS*		*AUC*		*TSS*		*AUC*		*TSS*		*AUC*		*TSS*		*AUC*		*TSS*	
Train	CV	Train	CV	Train	CV	Train	CV	Train	CV	Train	CV	Train	CV	Train	CV	Train	CV	Train	CV
0.9	0.87	0.63	0.58	0.84	0.84	0.51	0.51	0.85	0.84	0.54	0.53	0.86	0.85	0.55	0.54	0.87	0.87	0.59	0.62
0.04		0.05		0		0		0.01		0.01		0		0.01		0		−0.03	

$WCLIM_{Sim}$																			
BRT				*MARS*				*MAXENT*				*GLM*				*RF*			
AUC		*TSS*		*AUC*		*TSS*		*AUC*		*TSS*		*AUC*		*TSS*		*AUC*		*TSS*	
Train	CV	Train	CV	Train	CV	Train	CV	Train	CV	Train	CV	Train	CV	Train	CV	Train	CV	Train	CV
0.93	0.87	0.69	0.59	0.84	0.84	0.51	0.51	0.85	0.84	0.54	0.53	0.84	0.84	0.53	0.53	0.87	0.87	0.59	0.63
0.06		0.1		0		0		0.01		0.01		0		0		0		−0.04	



Municipality	Surface overestimated by the SIASsim (ha)	Absence surface (ha)	Concordant surface (ha)	Surface overestimated by the WCCLIMsim (ha)
Augusta	413.64	6109.48	3932.4	476.04
Avola	144.84	5977.72	1071.52	180
Buccheri	31.28	5213.6	471.2	28.48
Buscemi	0	5169.88	0	0.04
Canicattini Bagni	0.36	1490.16	3.68	1.2
Carlentini	482.72	9683.04	5027.2	586.28
Cassaro	0	1947.68	0	0
Ferla	0.52	2469.08	2.04	1.24
Floridia	23.6	2503.76	77.68	22
Francofonte	159.16	3051.48	3998.04	161.56
Lentini	688.72	8572.04	11,045.88	1225.08
Melilli	245.68	10,624.88	2319.36	348.6
Noto	563.68	50,033.88	3923.6	549.76
Palazzolo Acreide	2.16	8680.4	6.64	3.44
Portopalo di Capo Passero	0	0	0	0
Priolo Gargano	19.6	5474.2	78.12	41.36
Rosolini	6.8	7576.52	48.88	14.8
Syracuse	608.68	14,399.72	4717.4	683.88
Solarino	8.68	1209.96	53	21.16
Sortino	123.76	8434.84	549.28	159.16

Figure 5. (**a**) The green area indicates the intersection area between the WCLIM and SIAS simulations; (**b**) the red and blue areas highlight the symmetric differences between the two simulations; i.e., the red areas represent the areas overestimated by the $SIAS_{sim}$, while the blue areas show the areas overestimated by the $WCLIM_{sim}$.

Furthermore, the spatial analysis of the differences between the $WCLIM_{Sim}$ and $SIAS_{Sim}$ provided further details on the variation in potential presence forecasts at the territorial level. A good overall consistency between the two simulations (see Figure 5a,b) was found at the spatial level, though slight discrepancies in surface distribution emerged. In detail, the area of intersection between $SIAS_{Sim}$ and $WCLIM_{Sim}$ amounted to 37,325.92 ha, while a discrepancy of 980.2 ha was found between the surface areas of the two symmetric differences (Figure 5b), considering the 3523.88 ha overestimation by $SIAS_{Sim}$ and the 4504.08 ha overestimation by $WCLIM_{Sim}$ (Figure 5b). Therefore, the ratio between the overestimation (either by $WCLIM_{Sim}$ or $SIAS_{Sim}$) and the intersection area is within the range 9.44–12.06%. Spatial differences between the simulations were distributed in clusters, ranging from 0.03 ha to 38.35 ha (Figure 5a), with the spatial relations described in Figure 5b.

Figure 5b shows the clusters of spatial disagreement between WCLIM_Sim and SIAS_Sim, revealing minimal differences across the region. The spatial distribution of these clusters shows that discrepancies are generally small and scattered, with no large-scale or consistent bias across the study area. Overall, the figure shows that, while both models perform similarly in predicting areas suitable for citrus cultivation, using local-scale climatic data slightly improves the spatial definition of suitability zones, reducing the risk of overestimation or underestimation in specific microclimatic contexts.

Investigating the models' predictions at the spatial level is crucial to carefully evaluate the modelling methodologies and the sources of uncertainty in the results, in order to ensure the robustness of the predictions from the agricultural point of view. Some authors [56,57], in fact, have not analysed the outcomes from this point of view, but only based on the assessment metrics (e.g., AUC, and TSS). Based on the probability maps, the area estimated by $WCLIM_{Sim}$ was equal to 42,273.96 ha, while the area by $SIAS_{Sim}$ was 41,988.76 ha. Compared to the currently cultivated area of 25,250, as reported by the latest ISTAT survey (2021), these results suggest that there is potential for further expansion of citrus fruit production in the Syracuse area, from a pedoclimatic perspective. However, since Figure 5b does not allow a synthetic analysis on the differences between the simulations, a specific analysis was carried out at the municipal level (Figure 6).

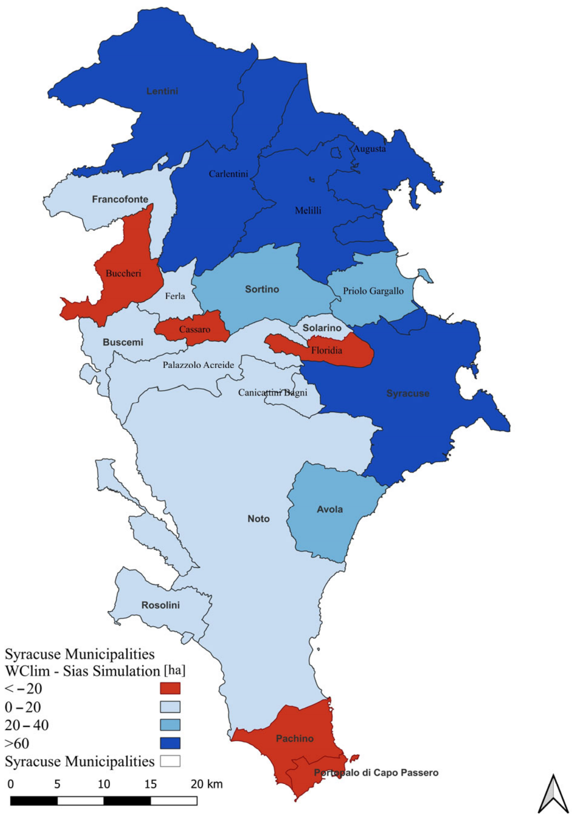


Figure 6. WCLIM-SIAS simulation difference in predicted potential area of citrus presence, in ha, for each municipality of the study area. In detail, $WCLIM_{Sim}$ overestimated areas in blue and $SIAS_{Sim}$ overestimated areas in red.

Therefore, Figure 6 describes the difference in the surface areas (A_{WCLIM}-A_{SIAS}) obtained from the model simulations, i.e., blue and red areas of Figure 4. It shows how the climate conditions provided by the $WCLIM_{Sim}$ data determined an overestimation of crop growth potential surface in the northern territory of the province compared to $SIAS_{Sim}$ in the southern part of the province, and they agree in the central areas (Figure 6). In detail, $WCLIM_{Sim}$ overestimated the surface areas by 536.36 ha in Lentini, 103.56 ha in Carlentini, 102.92 ha in Melilli, and 75.2 in Syracuse, whereas, the $SIAS_{Sim}$ overestimated in Noto (−13.92 ha), in Florida (−1.6 ha), and in Buccheri (−2.8 ha).

3.2. Predictors' Selection

In this study, the CovariateCorrelationSelector module made it possible to examine the differences in the selection of variables, allowing analysis of the relationships between the crop and the climate and providing further insight into the expected changes, as confirmed by some authors [58].

The variables selected as predictors for each simulation were as follows:

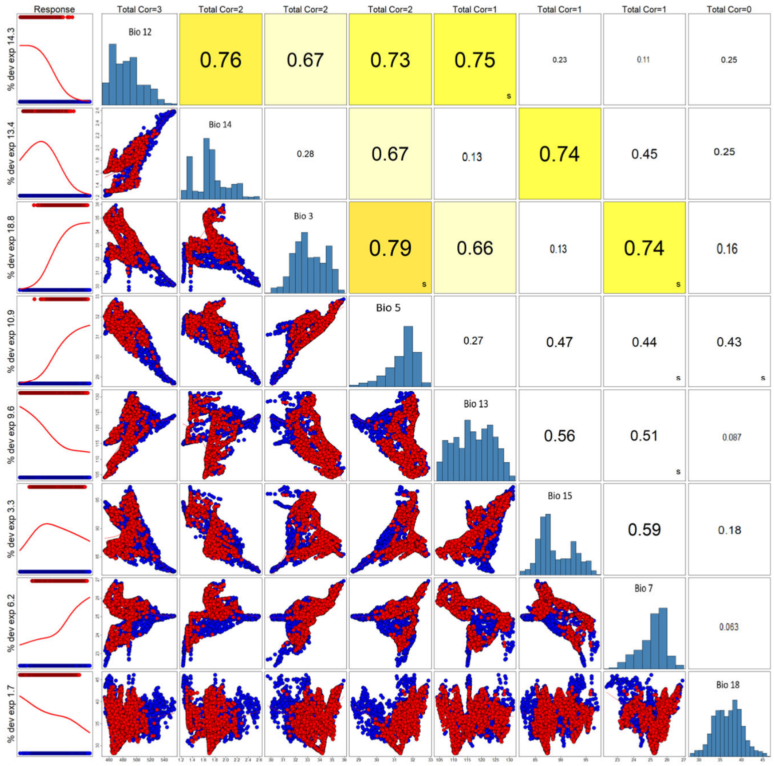
- $WCLIM_{Sim}$: Bio 3 (Isothermality), Bio 5 (Max Temperature of Warmest Month), Bio 7 (Temperature Annual Range), Bio 12 (Annual Precipitation), Bio 13 (Precipitation of Wettest Month), Bio 14 (Precipitation of Driest Month), Bio 15 (Precipitation Seasonality (Coefficient of Variation)), Bio 18 (Precipitation of Warmest Quarter), Slope, and V_{w-33}.
- $SIAS_{Sim}$: Bio 1 (Annual Mean Temperature), Bio 2 (Mean Diurnal Range), Bio 5 (Max Temperature of Warmest Month), Bio 7 (Temperature Annual Range), Bio 8 (Mean Temperature of Wettest Quarter), Bio 9 (Mean Temperature of Driest Quarter), Bio 13 (Precipitation of Wettest Month), Bio 15 (Precipitation Seasonality (Coefficient of Variation)), Bio 18 (Precipitation of Warmest Quarter), DTM, Slope, and V_{w-33}.

The initial set of 24 pedoclimatic variables was reduced to 12 predictors for $SIAS_{Sim}$ and 10 for $WCLIM_{Sim}$, based on the PSK correlation matrix threshold and the contribution (%) of the variables. In $WCLIM_{Sim}$, three out of the eight biovariables depended on temperature compared to six out of nine in $SIAS_{Sim}$. Therefore, it was found that $WCLIM_{Sim}$ was more affected by precipitation compared to $SIAS_{Sim}$.

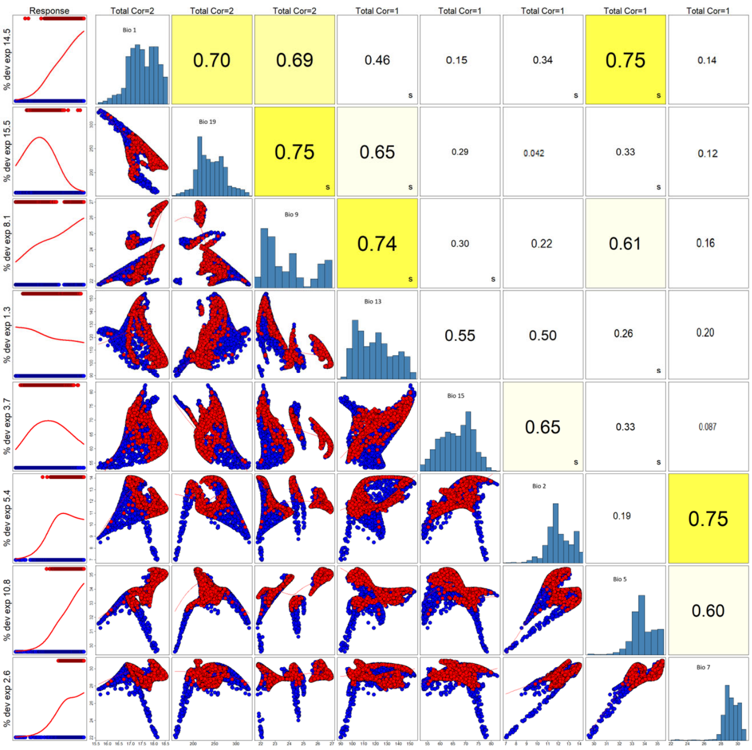
The first column of the correlation matrices in Figure 7a,b illustrates the relationship between the response and each predictor. The headings over each variable column indicate the number of other variables that the environmental predictor is correlated with, based on the user-defined threshold. The remaining plots form a matrix, where histograms for each variable are displayed along the diagonal.

Predictors' selection needs to be investigated when two correlated covariates show the same contribution (%) to the prediction as provided by the 'CovariateCorrelationSelector' module.

For instance, this was the case for Bio 5 (mean temperature of the warmest month) and Bio 17 (total precipitation during the driest three months of the year) in $WCLIM_{Sim}$ by using the information in Figures S1 and S2. The four graphs in Figures S1 and S2 report the output of the 'PredictorInspections' module for the two simulations. In detail, the block diagram indicates the frequency of presence (red) and absence (blue) data, which corresponds to the diagonal in Figure 7a,b. The map in the bottom-left corner shows data distribution in the study area; the map in the top-right corner describes the distribution of the predictor in the study area (at 1 km resolution); and the curve at the bottom-right corner is that reported in the first column of Figure 7a,b. In Figure S1, the maximum contribution occurred between 32 and 33 °C for Bio 5, while Bio 17 showed a greater contribution between 15 and 20 mm of precipitation. Since the contribution of water during the driest quarter comes from both rainfall and irrigation, the latter varies between 10,000 and 15,000 m^3 per hectare to meet the needs of the plants according to [59,60]. Therefore, the contribution of the Bio 17 curve (Figure S2) is not fully representative in this context. Consequently, since citrus productivity and quality decline at temperatures above 30 °C, as also found by [60], the Bio 5 variable was considered capable of best describing this plant biology parameter in the territory and, thus, it was selected, while Bio 17 was discarded.



(**a**)



(**b**)

Figure 7. (**a**) Pearson–Spearman–Kendall matrix in $\mathrm{WCLIM_{Sim}}$; (**b**) PSK matrix in SIAS simulation. The scatter plots with the smoothing curve are below the diagonal: red for presence, blue for absence and yellow for background data. The correlation coefficient between the predictor variables is shown above the diagonal.

Indeed, plant biology parameters are crucial for covariate selection, based on the PSK matrix. The minimum temperature required for citrus cultivation (i.e., biological/physiological zero or base temperature) is 13 °C (which can be considered a threshold value for vegetative and reproductive growth) [61]. In this study, Bio 6 (minimum temperature of the coldest month) in the cultivation areas had minimum values of 5 °C for $WCLIM_{Sim}$ and 3.5 °C for $SIAS_{Sim}$, and Bio 11 (mean temperature of the coldest quarter) in the cultivation areas had minimum values of 10 °C for $WCLIM_{Sim}$ and 9 °C for $SIAS_{Sim}$. Therefore, there are some periods of the year when the temperature falls below the base temperature; thus, yield would be affected in those areas.

In addition, citrus yield is negatively correlated with the number of days with a temperature above 30 °C, and an average temperature below 24 °C is necessary to induce flowering [62,63]. In this study, Bio 10 (mean temperature of the warmest quarter) in the cultivation areas had minimum values of 27 °C for $WCLIM_{Sim}$ and 26 °C for $SIAS_{Sim}$, and Bio 5 (mean temperature of the warmest month) had minimum values of 35 °C for $WCLIM_{Sim}$ and 33 °C for $SIAS_{Sim}$ in the cultivation areas, the latter being one of the variables chosen in both simulations. Therefore, in the warmest months the average temperature was always higher than 30 °C.

In this regard, the study by Vitasse et al. [63] analysed the effects of climate change on temperate European trees and confirmed that plants adjust their phenology in response to climatic variations. Specifically, the phenology of species, such as *Quercus* and *Fagus*, has been shown to be affected by altitude, with consequences on the duration of canopy growth. For instance, at lower altitudes, the growing season may start earlier as a result of rising temperatures.

As climate parameters are strictly connected to plant biology, future research should focus on long-term climate projections for citrus cultivation based on the local scale, with a particular focus on adaptation measures (e.g., cultivar selection for hot climate areas, and shade nets). Furthermore, the predicted socio-economic impacts on citrus-dependent sectors should also be investigated in order to develop effective adaptation strategies.

3.3. Assessment of Predictors' Contributions

The comparison between the contributions of the predictors in the two simulations was carried out by using the PSK matrix and the MDA index. The PSK matrix provided information on the contribution of each predictor variable to the predicted potential presence of the species based on the two simulations (Section 3.3.1). Assessment through the MDA index provided information on the marginal importance of correlated variables (Section 3.3.2). In Sections 3.3.1 and 3.3.2 analyses on PSK- and MDA-based assessments were reported, respectively.

3.3.1. Assessment Based on Pearson–Spearman–Kendall Matrix

The contribution of predictors to the output, derived from the PSK matrix, are reported in Figure 8. The variables with the greatest contribution were Bio 3 (18.78%), Bio 16 (16.01%), Bio 12 (14.26%), DTM_20 (14.1%), and Bio 14 (13.42%) in $WCLIM_{Sim}$, whereas they were Bio 19 (15.5%), Bio 1 (14.54%), Bio 11 (15.1%), DTM (14.12%), Bio8 (14.07%), and Bio 14 (13.42%) in $SIAS_{Sim}$.

The variable 'Isothermality' (Bio 3) showed a great influence (about 20%) on the distribution of the species in the $WCLIM_{Sim}$, while it did not in the $SIAS_{Sim}$ where the contribution was equal to 5.1%. The SIAS data showed a significant decrease (up to 13.7%) for the Bio 3 variable, particularly along the east coast, indicating a reduction in the diurnal temperature range compared to the annual temperature range.

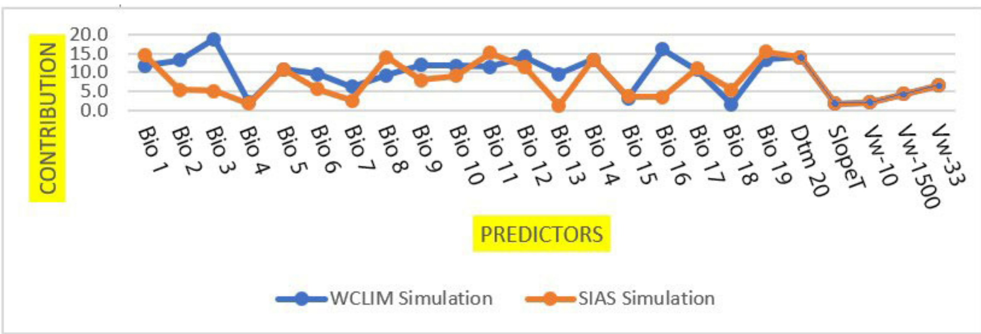


Figure 8. Covariates' importance (%) in PSK matrix, for each simulation.

The difference between the contributions of the variable Bio 18 in the two simulations is 3.8%. This difference is due to the inclusion of storm events in the SIAS database, which influence the predicted values, with variations between 103 mm and 131 mm. This result highlights how rainfall during the warmer months favours plant presence despite the high temperatures recorded. In addition, it supports the capacity of the PSK matrix to link the environmental variables to crop predicted presence. Moreover, since low rainfall during the warmest period implies a lower value from the variable, the predicted rainfall from the WCLIM database was less adequate than the SIAS one, due to the high variations.

The variables with the lowest difference between them in the two simulations were Bio 14 (0%), Bio 5 (0.1%), Bio 17 and 4 (−0.3%), and Bio 15 (−0.5%). Although both simulations attributed a similar contribution to precipitations of the driest quarter (Bio 17), equal to about 11%, the differences in the observed precipitation values, especially in the Iblei Mountains area (mainly in the municipalities of Noto and Rosolini), indicated a variation of 45 mm, while this variation is less than 5 mm in the areas of citrus presence. The difference in Bio 17 values between the two simulations was higher in mountainous areas, as also demonstrated by Kyrgyzbay et al. [17]. In detail, variations in predictors' values in relation to altitude were found by Kyrgyzbay et al. [17], who averaged data from open climate databases (WorldClim and CHIRPS), and predictors were determined over a 10-year period in natural and high-altitude areas. Their results showed large differences in predictors' values in mountainous regions, whereas moderate differences were observed in lowland areas. These outcomes were confirmed in this study (Figure 5a,b), where the major differences between the data collected by the SIAS platform and the global climate model (WCLIM) are concentrated in the Iblei mountain plateau, within the municipalities of Palazzo Acreide, Cassaro, Solarino, Ferla, and Sortino. However, since in this study the analysis of predictors was developed on the areas where the citrus plant is present, the results of the two simulations were not affected by discrepancy in this variable.

3.3.2. Assessment Based on Mean Decrease Accuracy Index

In Figure 9, the impact of the most significant variables on the simulations, as determined by the MDA index generated by the RF model, is reported. The MDA index values were normalised to their respective maximum values to allow comparisons.

In Figure 9, the highest contributions of the predictors in $WCLIM_{sim}$, as described by MDA index, were by Bio 14, DTM_20, $V_{w\text{-}10}$, $V_{w\text{-}33}$, and $V_{w\text{-}1500}$, with MDA values higher than 80, whereas in $SIAS_{sim}$ the MDA values of the biovariables Bio 15, Bio 17, Bio 2, Bio 7, DTM_20, $V_{w\text{-}10}$, $V_{w\text{-}33}$, and $V_{w\text{-}1500}$ were the highest.

For both datasets (WCLIM and SIAS), the terrain and water retention variables were highly important, while the bioclimatic ones had the greatest effect on $SIAS_{sim}$. In both simulations, the variable with the greatest contribution was $V_{w\text{-}10}$, indicating the importance of water volume at saturation in soil for the accuracy of the model simulation. This confirms the model's ability to identify the main factors influencing the presence of the plant. In fact, it is well known that the effects of water stagnation and salinity excess reduce vigour, thus reducing the potential citrus presence.

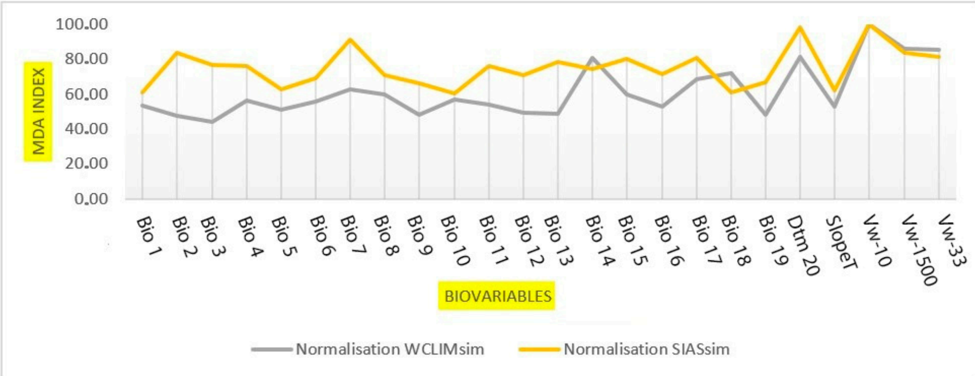


Figure 9. Predictors' importance by Mean Decrease Accuracy index.

Among the bioclimatic variables, temperature had a considerable impact in the $SIAS_{Sim}$, specifically Bio 7 and Bio 2. This indicates that the temperature biovariables, in general, had more influence on the model based on SIAS data. In $SIAS_{Sim}$, MDA for Bio 2, equal to 83.67, showed an increase of 43% compared to the biovariable's MDA in $WCLIM_{Sim}$ (Figure 9). A similar increase (42%) was found for Bio 3, whereas MDA for Bio 7 increased by 31%.

Conversely, in the $WCLIM_{Sim}$, rainfall-related variables, such as Bio 14 and Bio 19, and variables related to the volume of water in soil, such as $V_{w\text{-}10}$, had a significant impact on citrus potential presence, reflecting the importance of rainfall and soil moisture for citrus crops. MDA for Bio 14, equal to 80.57, showed a decrease of 9% compared to the biovariable's MDA in $SIAS_{Sim}$ (Figure 8). In this regard, crop productivity is influenced by rainfall and soil moisture, with water requirements estimated between 10,000 m^3 and 15,000 m^3 per hectare [58], emphasising the need to maintain adequate soil moisture for optimal growth.

In both simulations, MDA for the topographic variable DTM showed high values, indicating its significant role in the model. Based on these results, the distribution of the plant seems to be more concentrated in areas with altitudes below 400 m a.s.l., which is a result in line with values reported by [64]. Furthermore, it was found that at increasing altitude, citrus predicted presence decreases, in line with studies conducted by [60], who highlighted the risk of frost in citrus species and cultivars at increasing altitude. Although there is no precise value for the maximum altitude for citrus cultivation, the FAO indicates an altitude below 700 m (https://www.fao.org/home/en/ accessed on 13 May 2025). This is an important altitude value to be considered when assessing the agricultural potential of mountainous areas.

In the literature, there is no consensus on the use of uncorrelated variables in PSK method versus all variables used in MDA assessment. Therefore, in this study, the methodology considered the PSK matrix to select the uncorrelated predictors following the approaches described by Akpoti et al. [15] and West et al. [12], whereas the method for assessing the influence of all predictors in the two simulations follows the recommendations of Breiman [65] and Strobl et al. [66], i.e., utilising all the predictors in the MDA, to ensure a careful assessment of variable importance and interactions. Actually, RF models typically evaluate the marginal importance of variables by using the MDA index. According to Breiman [65], the inclusion of a large number of predictors may only marginally increase the prediction error, suggesting that the robustness of the model is relatively unaffected by the addition of more predictors. Strobl et al. [66] support this remark by observing that the use of all variables, even correlated ones, can capture complex interactions that may not be detected when only uncorrelated predictors are considered. The capability of considering these interactions is one of the reasons why RF models are often preferred for handling high-dimensional datasets. Conversely, Biau et al. [67] pointed out that the empirical properties of the MDA can be affected by the inclusion of highly correlated predictors, which can lead to misleading importance scores. Since there is no consensus in the literature on using selected uncorrelated variables or all the predictors to assess variable importance for the outcomes, further studies are needed to investigate this issue.

3.4. Comparison Between WCLIM$_{sim}$ and SIAS$_{sim}$ Biovariables

It is well known how the accuracy of agricultural predictions is influenced by the quality of the climate data used as input to the SDM. Therefore, a further comparison was made between the biovariables of the two approaches based on global and local climate data. Among the most relevant bioclimatic variables, Bio 1 and Bio 12 were analysed since they represent annual averages that are crucial for the distribution of crops. In addition, the variables Bio 3 and Bio 16 were considered, as they show a greater variation in their influence on the output of the two simulations, highlighting significant differences in the response of the model to specific climatic conditions.

The assessment of the main variables, shown in Table 2, indicates differences in both the surface values and the predictor values between the simulations. Furthermore, numerical values alone are insufficient for analysing the distribution of observed variations. Therefore, additional analyses were performed to provide a comprehensive assessment of the discrepancies between the models.

Table 2. A comparison of the values of the climatic biovariables derived from the two modelling approaches (WCLIM$_{Sim}$ and SIAS$_{Sim}$).

Biovariable	WCLIM$_{Sim}$		SIAS$_{Sim}$	
	Class	Surface [ha]	Class	Surface [ha]
Bio 1	14.72–15.50	1.48	15.90–16.50	19.40
	15.50–16.50	103.40	16.50–17	3115.88
	16.50–17.50	4512.92	17–17.50	12,066.80
	17.50–18.50	28,380.64	17.50–18	11,062.48
	18.50–19.16	9275.52	18–18.55	15,724.20
Bio 3	29.75–30.95	21.48	29.09–32.50	21.48
	30.95–32.15	2001.12	32.50–36	2001.12
	32.15–33.35	13,216.52	36–39.55	13,216.52
	33.15–34.55	14,183.20	39.55–43	14,183.20
	33.55–35.75	12,851.64	43–46.52	12,851.64
Bio 12	444.40–460.50	5702.16	580–644.10	14,555.68
	460.50–476.50	15,243.36	644.10–708.30	12,515.72
	476.50–492.50	15,062.76	708.30–772.40	7993.52
	492.50–508.50	5395.12	772.40–836.55	6893.92
	508.50–547.10	870.56	836.55–903	29.92
Bio 16	224.903–231.50	3040.20	236.8–266.45	1,645.12
	231.50–238.50	11,810.64	266.45–296.10	15,196.12
	238.50–245.50	11,585.48	296.10–325.75	9051.52
	245.50–252.00	7758.04	325.75–355.4	5169.72
	252.00–258.92	8079.60	355.40–384.97	10,926.28

The analysis of the Bio 1 differences (Figure 10a) between the two simulations showed that the average temperature difference in Sicily is 0.80 °C. However, in the province of Syracuse, the SIAS data showed a greater value, up to 1.80, compared to the simulation based on WCLIM$_{Sim}$, particularly in the areas of Pachino, Lentini, Avola, Syracuse and Noto.

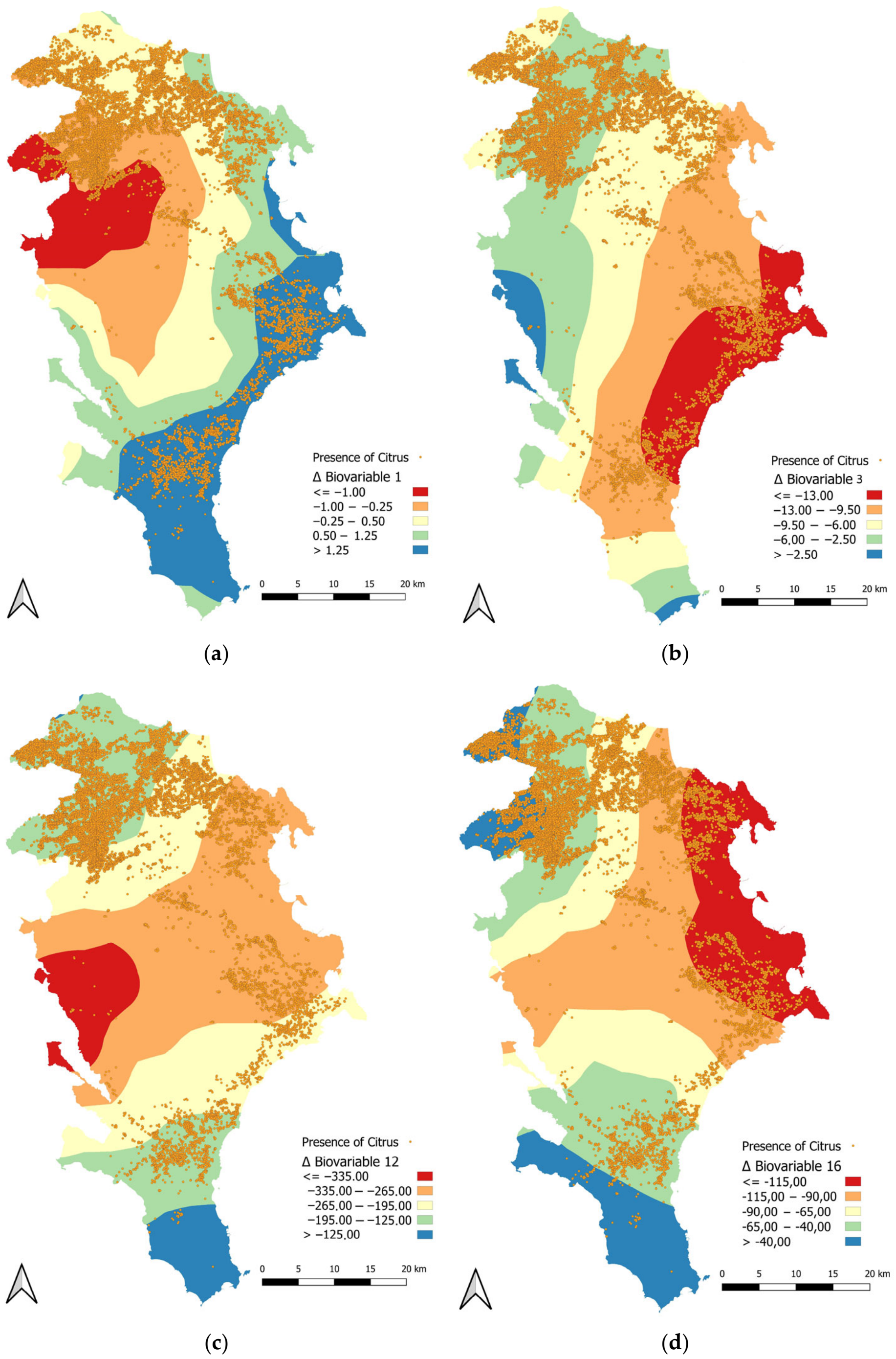


Figure 10. Spatial distribution of the differences between the following biovariables in the two simulations ($WCLIM_{Sim}$–$SIAS_{Sim}$): (**a**) Bio 1; (**b**) Bio 3; (**c**) Bio 12; (**d**) Bio 16.

As for Bio 3 (Figure 10b), there was a significant increase in the temperature oscillations between day and night, with variations between −16.70% and 12% compared to the annual oscillations, especially in the south-eastern areas. Bio 12 (Figure 10c), which reflects the mean annual precipitation, showed significant differences between the two simulations. In particular, the south-eastern areas between the municipalities of Syracuse, Priolo Gargallo and Augusta received more precipitation according to the SIAS data, on average 334 mm more than in the WCLIM simulation.

Bio 16 (Figure 10d), which represents the rainfall of the wettest quarter, showed significant differences in the coastal areas of Priolo Gargallo, Augusta and Syracuse, with a higher level of rainfall according to the SIAS data, by an average of 140 mm more than the

WCLIM data. In the central inland municipalities, such as Palazzolo Acreide, Canicattini Bagni, Floridia and Solarino, the increase was smaller, about 114 mm on average.

Overall, SIAS data showed that on an annual basis, inland areas have higher rainfall rates, whereas, during the wettest quarter, a more pronounced increase in precipitation was observed in coastal areas compared to WCLIM data. Moreover, the higher values of precipitation in the internal areas (Sortino, Canicattini Bagni, Palazzolo Acreide, and Floridia) recorded by SIAS, compared to those in WCLIM, are noteworthy for time series. Over the 18-year analysis period, spanning from 2003 to 2021, monthly precipitation levels exceeding 245 mm were documented on 12 distinct occurrences; these events were primarily concentrated in the autumn/winter months (September, November, January, February and March). In the municipality of Augusta, for instance, the total amount of rainfall in WCLIM, averaged over the 2003–2021 period, was 489.8 mm, while that acquired by SIAS was 773.73 mm, with a difference of 284 mm.

In detail, there are evident discrepancies in September, October, November, December, February, and April (Figure 11). In general, inconsistencies in precipitation values between the two simulations were found in the whole province. Indeed, precise knowledge of rainfall levels is crucial to simulate the effect on species distribution, adapt resource management, and investigate long-term impacts.

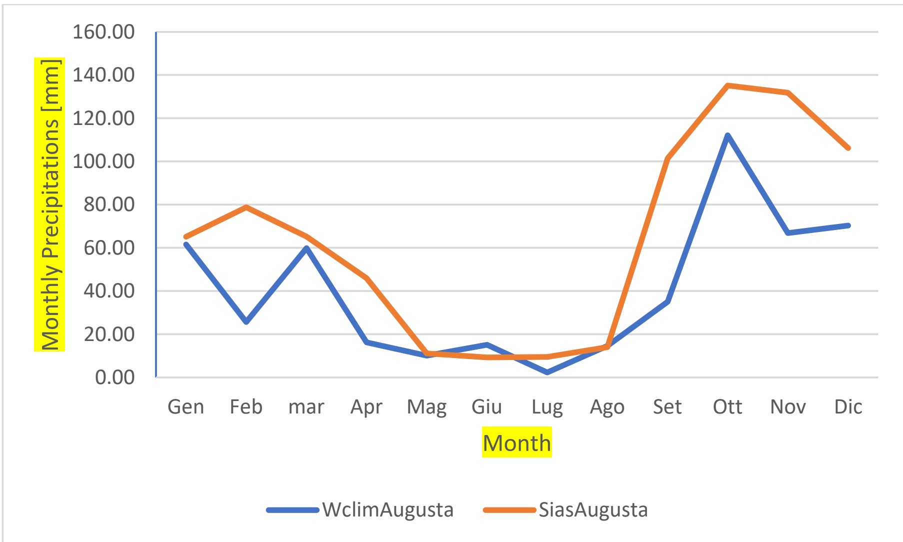


Figure 11. Monthly data of precipitation [mm] in the municipality of Augusta, averaged over the two-decade period, for the two simulations ($WCLIM_{Sim}$ and $SIAS_{Sim}$).

The above-described discrepancies in precipitation data align with observations of Kyrgyzbay et al. [17] who highlighted inconsistencies in data agreement between open climate datasets (such as WorldClim and CHIRPS) and ground-based meteorological stations.

Therefore, the use of local ground-based meteorological monitoring stations represents a strategy to accurately quantify the actual amount of precipitation in the territory, thus providing greater reliability and truthfulness of the data obtained. In the literature, indeed, local data are generally indicated as more preferable, if available, compared to global data. The utilisation of monthly meteorological data and temperatures from stations of the Spanish Meteorological Agency, as highlighted by Pérez et al. [39], allowed the generation of climate predictors during a drought event. The utilisation of data from ground-based meteorological stations offers a number of advantages, including superior spatial and temporal resolution in comparison to data obtained from global sources, particularly where conditions may vary due to, for example, terrain variations. The results of this study indicated that the species under analysis was mainly influenced by average temperatures, particularly during the hot months, and by seasonality and the quantity of rainfall. These factors, as highlighted by Zouabi [68], show a correlation with farming practices since the increase in temperature and the concentration of rainfall make it necessary for farmers to use irrigation. In contexts where water scarcity is a reality, the adoption of deficit irrigation practices can help minimise water waste, thus optimising the use of this resource [51,52,69], and specific simulations should be carried out (Catalano et al. [15]).

Indeed, both the analysed simulations ($SIAS_{Sim}$ and $WCLIM_{Sim}$) agreed on the greater potential of species presence in the northern and central regions of the province; however, dependent on precipitation, human intervention would be required to maintain a suitable environment for cultivation.

Considering the previous remarks, models based on global data may not produce accurate results due to strong discrepancies in input values compared to models based on local data, resulting in incorrect predictions by the SDM. To improve accuracy, metadata should be included both in local and global data to facilitate comparisons among biovariables computed from these data.

The integration of local environmental data, combined with GIS tools, could significantly improve the accuracy of models, capturing spatial and temporal variations in environmental conditions. In fact, evaluating the predictive capacities of the SDMs based on the use of local data would be of interest for local authorities and decision makers who could base their strategic choices on simulations performed by using data representing the actual climatic conditions at a lower territorial level (e.g., provincial and municipal instead of supranational, national or regional) and on the knowledge of precise potential spatial distribution of the crop.

4. Conclusions

Species Distribution Models (SDMs) are widely used in the scientific community to analyse the distribution of vegetation, organisms and animals. In this study, the use of VisTrails:SAHM software optimised the habitat modelling process and facilitated the management of input data.

The aim of this study investigating predictor selection was pursued by analysing and assessing models' performances through approaches based on the PSK matrix and MDA index. The PSK matrix assessment helped identify uncorrelated variables, reducing redundancy, while the MDA index provided insight into the marginal contribution of each variable to model accuracy.

Furthermore, the MDA index highlighted the importance of terrain and water retention variables in both datasets, with variations in the relevance of climatic and topographic variables between the two simulations. These results underline the importance of considering a variety of environmental factors when assessing plant distributions, especially in mountainous areas where climatic and altimetric variations can significantly influence agricultural productivity.

This study has indicated that the choice of utilising data from local meteorological stations instead of global data can influence the predicted spatial distribution of citrus crops by about 10%, in the case study analysed.

The different distributions were due to the fact that the bioclimatic variables selected for the simulations differed between models based on global and local data, which in turn depends on the different biovariables' patterns. In the $WCLIMS_{sim}$, the isotherm, annual precipitation and precipitation of the driest month were the most important variables, whereas in the SIAS models, the mean annual temperature, the digital terrain model (DTM) and the mean temperature of the wettest quarter were the most important variables. The predictors differed significantly between global and local data, especially precipitation. As for rainfall, SIAS data showed that on an annual basis, inland areas have higher values compared to WCLIM data, whereas, during the wettest quarter, a more pronounced increase in precipitation was observed in coastal areas. However, despite these differences in the values of the variables, the predictions of the two simulations differed by less than 800 ha (~8 km^2).

A methodology that integrates local meteorological data, such that proposed in this study, constitutes a fundamental step for precision agriculture, contributing to the improvement of resource management, facilitating the application of climate change mitigation strategies and supporting local authority decision making to develop more effective adaptation strategies. Detailed knowledge of the impact of other predictors on citrus distribution, such as anthropic burden or cultivar, could lay the groundwork for further scientific advancements. Exploiting the potential of SDM applications, the next step would be to assess the impact of future climate scenarios based on different datasets and develop adaptation strategies taking into account the most influential environmental variables for the considered species.

The use of local meteorological data and the assessments proposed in this research study contributed to refining the knowledge base on the prediction of crop distribution compared to global data; it allowed for analysing specific local climate variables influencing the presence of citrus.

Supplementary Materials: The supporting information can be downloaded at https://www.mdpi.com/article/10.3390/land14112223/s1. Table S1. Bioclimatic variables. Figure S1. Predictor inspections for Bio5 in $WCLIM_{Sim}$. Figure S2. Predictor inspections for Bio17 in $WCLIM_{Sim}$. Figure S3. Response curves for Bio5 and Bio7 from the two simulations ($WCLIM_{Sim}$ and $SIAS_{Sim}$).

Author Contributions: Conceptualization, P.R.D.; methodology, G.A.C. and C.A.; software, G.A.C., P.R.D.; validation, P.R.D.; formal analysis, G.A.C. and C.A.; investigation, G.A.C. and C.A.; resources, C.A.; data curation, G.A.C.; writing—original draft preparation, G.A.C. and P.R.D.; writing—review and editing, G.A.C., P.R.D. and C.A.; visualisation, G.A.C. and P.R.D.; supervision, C.A.; project administration, C.A.; funding acquisition, C.A. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR)—MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4—D.D. 1032 17/06/2022, CN00000022—CUP: E63C22000960006). The work of prof. Claudia Arcidiacono was carried out within the same project. This manuscript reflects only the authors' views and opinions; neither the European Union nor the European Commission can be considered responsible for them. This research is also in line with the 'Piano incentivi per la ricerca di Ateneo 2024–2026 (DR 2306 del 03/06/2024)—Pia.ce.ri, Linea 2′ project on 'Innovative solutions and strategies for sustainability of Cities, Land and Society' (SIA3) by the University of Catania, coordinated by Claudia Arcidiacono.

Data Availability Statement: The data presented in this study are available at the Sicilian Region for Land Information System (SITR) (https://www.sitr.regione.sicilia.it/ accessed on 19 September 2025), the Regional Territorial Information System of Sicily (SIAS) (http://www.sias.regione.sicilia.it/ accessed on 19 September 2025), the global database WorldClim (https://www.worldclim.org/ accessed on 16 July 2020), and Soilgrid250 (https://soilgrids.org/ accessed on 2 October 2023).

Acknowledgments: The authors wish to thank the Sicilian Region for Land Information System (SITR) for data (https://www.sitr.regione.sicilia.it/) and the Regional Territorial Information System of Sicily (SIAS) (http://www.sias.regione.sicilia.it/).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

A_{SIAS}	Area predicted by $SIAS_{Sim}$
AUC	Area Under the Curve
A_{WCLIM}	Area predicted by $WCLIM_{Sim}$
Bio 1	Annual Mean Temperature
Bio 2	Mean Diurnal Range (Mean of monthly (max temp–min temp))
Bio 3	Isothermality (BIO2/BIO7) (×100)
Bio 4	Temperature Seasonality (standard deviation × 100)
Bio 5	Max Temperature of Warmest Month
Bio 6	Min Temperature of Coldest Month
Bio 7	Temperature Annual Range (Bio 5–Bio 6)
Bio 8	Mean Temperature of Wettest Quarter
Bio 9	Mean Temperature of Driest Quarter
Bio 10	Mean Temperature of Warmest Quarter
Bio 11	Mean Temperature of Coldest Quarter
Bio 12	Annual Precipitation
Bio 13	Precipitation of Wettest Month (Coefficient of Variation)
Bio 14	Precipitation of Driest Month
Bio 15	Precipitation Seasonality
Bio 16	Precipitation of Wettest Quarter
Bio 17	Precipitation of Driest Quarter
Bio 18	Precipitation of Warmest Quarter
Bio 19	Precipitation of Coldest Quarter
BRT	Boosted Regression Tree
CMCC	Euro-Mediterranean Center on Climate Change
DTM	Digital Terrain Model
ENM	Ecological Niche Models
GIS	Geographic Information System
IPCC	Intergovernmental Panel on Climate Change
MARS	Multivariate Adaptive Regression Splines
MAXENT	Maximum Entropy
MDA	Mean Decrease Accuracy
MDSBuilder	Merged Data Set Builder
OOB	Out-of-bag
PSK	Pearson–Spearman–Kendall
RF	Random Forest
SAHM	Software for Assisted Habitat Modelling
SDM	Species Distribution Models
SIAS	Regional Territorial Information System of Sicily
$SIAS_{Sim}$	Regional Territorial Information System of Sicily Simulation
SITR	Regional Land Information System
$Slope_T$	Terrain Slope
TSS	True Skills Statistics
$V_{w\text{-}10}$	Volume of water in soil at a -60 cm depth, at soil saturation of -10 kPa
$V_{w\text{-}1500}$	Volume of water in soil at a -60 cm depth, at soil saturation of -1500 kPa
$V_{w\text{-}33}$	Volume of water in soil at a -60 cm depth, at soil saturation of -33 kPa
WCLIM	WorldClim
$WCLIM_{Sim}$	WorldClim Simulation
WFS	Web Feature Service

References

1. IPCC. *Climate Change 2021 The Physical Science Basis*; IPCC: Geneva, Switzerland, 2021.
2. Cramer, W.; Guiot, J.; Fader, M.; Garrabou, J.; Gattuso, J.-P.; Iglesias, A.; Lange, M.A.; Lionello, P.; Llasat, M.C.; Paz, S.; et al. Climate change and interconnected risks to sustainable development in the Mediterranean. *Nat. Clim. Change* **2018**, *8*, 972–980. [CrossRef]
3. Liuzzo, L.; Bono, E.; Sammartano, V.; Freni, G. Analysis of spatial and temporal rainfall trends in Sicily during the 1921–2012 period. *Theor. Appl. Clim.* **2016**, *126*, 113–129. [CrossRef]
4. Istituto Nazionale di Statistica "ISTAT". Available online: https://esploradati.istat.it/ (accessed on 22 September 2025).
5. Servizio Informativo Agrometeorologico Siciliano "SIAS". Available online: http://www.sias.regione.sicilia.it/ (accessed on 19 September 2025).
6. Elith, J.; Leathwick, J.R. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annu. Rev. Ecol. Evol. Syst.* **2009**, *40*, 677–697. [CrossRef]
7. Schwager, P.; Berg, C. Remote sensing variables improve species distribution models for alpine plant species. *Basic Appl. Ecol.* **2021**, *54*, 1–13. [CrossRef]
8. Song, Y.; Xu, X.; Zhang, S.; Chi, X. Uncertainty Assessment of Species Distribution Prediction Using Multiple Global Climate Models on the Tibetan Plateau: A Case Study of *Gentiana yunnanensis* and *Gentiana siphonantha*. *Land* **2024**, *13*, 1376. [CrossRef]
9. Chauvier, Y.; Descombes, P.; Guéguen, M.; Boulangeat, L.; Thuiller, W.; Zimmermann, N.E. Resolution in species distribution models shapes spatial patterns of plant multifaceted diversity. *Ecography* **2022**, *2022*, e05973. [CrossRef]
10. Talbert, C.; Talbert, M. *User Documentation for the Software for Assisted Habitat Modeling (SAHM) Package in VisTrails*; USGS (U.S. Geological Survey): Reston, VA, USA, 2001.
11. Morisette, J.T.; Jarnevich, C.S.; Holcombe, T.R.; Talbert, C.B.; Ignizio, D.; Talbert, M.K.; Silva, C.; Koop, D.; Swanson, A.; Young, N.E. VisTrails SAHM: Visualization and workflow management for species habitat modeling. *Ecography* **2013**, *36*, 129–135. [CrossRef]
12. West, A.M.; Jarnevich, C.S.; Young, N.E.; Fuller, P.L. Evaluating Potential Distribution of High-Risk Aquatic Invasive Species in the Water Garden and Aquarium Trade at a Global Scale Based on Current Established Populations. *Risk Anal.* **2019**, *39*, 1169–1191. [CrossRef]
13. Mukherjee, T.; Sharma, V.; Sharma, L.K.; Thakur, M.; Joshi, B.D.; Sharief, A.; Thapa, A.; Dutta, R.; Dolker, S.; Tripathy, B.; et al. Landscape-level habitat management plan through geometric reserve design for critically endangered Hangul (*Cervus hanglu hanglu*). *Sci. Total Environ.* **2021**, *777*, 146031. [CrossRef]
14. Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA). *Il Clima Futuro in Italia: Analisi Delle Proiezioni dei Modelli Regionali*; Stato Dell'Ambiente 58/2015; ISPRA: Rome, Italy, 2015; ISBN 978-88-448-0723-8.
15. Akpoti, K.; Kabo-Bah, A.T.; Dossou-Yovo, E.R.; Groen, T.A.; Zwart, S.J. Mapping suitability for rice production in inland valley landscapes in Benin and Togo using environmental niche modeling. *Sci. Total Environ.* **2020**, *709*, 136165. [CrossRef]
16. Pinilla-Buitrago, G.E. Predicting potential range shifts using climatic time series and niche models: A Neotropical montane shrew's case. *Ecol. Inform.* **2023**, *77*, 102212. [CrossRef]
17. Kyrgyzbay, K.; Kakimzhanov, Y.; Sagin, J. Climate data verification for assessing climate change in Almaty region of the Republic of Kazakhstan. *Clim. Serv.* **2023**, *32*, 100423. [CrossRef]
18. Catalano, G.A.; D'urso, P.R.; Maci, F.; Arcidiacono, C. Influence of Parameters in SDM Application on Citrus Presence in Mediterranean Area. *Sustainability* **2023**, *15*, 7656. [CrossRef]
19. Catalano, G.A.; Maci, F.; D'urso, P.R.; Arcidiacono, C. GIS and SDM-Based Methodology for Resource Optimisation: Feasibility Study for Citrus in Mediterranean Area. *Agronomy* **2023**, *13*, 549. [CrossRef]
20. Catalano, G.; D'urso, P.; Arcidiacono, C. Predicting potential biomass production by geospatial modelling: The case study of citrus in a Mediterranean area. *Ecol. Inform.* **2024**, *83*, 102848. [CrossRef]
21. Pavone, P.; Spampinato, G.; Costa, R.; Minissale, P.; Ronsisvalle, F.; Sciandrello, S.; Tomaselli, V. La vegetazione forestale dei Monti Iblei (Sicilia sud-orientale): I querceti. In Proceedings of the Terzo Congresso Nazionale di Selvicoltura, Florence, Italy, 16–19 October 2009; pp. 234–239. [CrossRef]
22. Cartabellotta, D.; Drago, A.; Lo Bianco, B.; Lombardo, M. *Climatologia della Sicilia*; Regione Siciliana, Assessorato Agricoltura e Foreste: Palermo, Italy, 1998.
23. Guisan, A.; Graham, C.H.; Elith, J.; Huettmann, F. NCEAS Species Distribution Modelling Group Sensitivity of predictive species distribution models to change in grain size. *Divers. Distrib.* **2007**, *13*, 332–340. [CrossRef]
24. Valavi, R.; Guillera-Arroita, G.; Lahoz-Monfort, J.J.; Elith, J. Predictive performance of presence-only species distribution models: A benchmark study with reproducible code. *Ecol. Monogr.* **2022**, *92*, e1486. [CrossRef]
25. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **2017**, *37*, 4302–4315. [CrossRef]

26. Regional Territorial Information System of Sicily "SITR". Available online: https://www.sitr.regione.sicilia.it/portal/home/item.html?id=06b441f103024aa4b1b9f966b1e4e3f9 (accessed on 19 September 2025).
27. Young, N.E.; Jarnevich, C.S.; Sofaer, H.R.; Pearse, I.; Sullivan, J.; Engelstad, P.; Stohlgren, T.J. A modeling workflow that balances automation and human intervention to inform invasive plant management decisions at multiple spatial scales. *PLoS ONE* **2020**, *15*, e0229253. [CrossRef]
28. WorldClim, Maps, Graphs, Tables, and DATA of the Globla Climate. Available online: https://www.worldclim.org/ (accessed on 16 July 2020).
29. O'Donnel, M.S.; Ignizio, D.A. *Bioclimatic Predictors for Supporting Ecological Applications in the Conterminous United States (No. 691)*; US Geological Survey: Reston, VA, USA, 2012.
30. Roy, D.; Das, B.; Singh, P.; Santra, P.; Deb, S.; Bhattacharya, B.K.; Govind, A.; Jatav, R.; Sethi, D.; Ghosh, T.; et al. Assessing the accuracy of multi-model approaches for downscaling land surface temperature across diverse agroclimatic zones. *Sci. Rep.* **2025**, *15*, 10824. [CrossRef]
31. Schlaepfer, D.R.; Bradford, J.B.; Lauenroth, W.K.; Shriver, R.K. Understanding the future of big sagebrush regeneration: Challenges of projecting complex ecological processes. *Ecosphere* **2021**, *12*, e03695. [CrossRef]
32. Garcia, L.; Celette, F.; Gary, C.; Ripoche, A.; Valdés-Gómez, H.; Metay, A. Management of service crops for the provision of ecosystem services in vineyards: A review. *Agric. Ecosyst. Environ.* **2018**, *251*, 158–170. [CrossRef]
33. Regional Territorial Information System of Sicily "SITR", Digital Terrain Model (DTM). Available online: https://map.sitr.regione.sicilia.it/gis/services/modelli_digitali/mdt_2013/ImageServer/WMSServer?service=WMS&version=1.3.0&request=GetCapabilities (accessed on 19 September 2025).
34. Sphinx, QGIS Manual User. Available online: https://docs.qgis.org/3.40/it/docs/training_manual/index.html (accessed on 7 April 2025).
35. Soilgrid250. Available online: https://soilgrids.org/ (accessed on 2 October 2023).
36. Poggio, L.; De Sousa, L.M.; Batjes, N.H.; Heuvelink, G.B.; Kempen, B.; Ribeiro, E.; Rossiter, D. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL* **2021**, *7*, 217–240. [CrossRef]
37. Turek, M.E.; Poggio, L.; Batjes, N.H.; Armindo, R.A.; van Lier, Q.D.J.; de Sousa, L.; Heuvelink, G.B. Global mapping of volumetric water retention at 100, 330 and 15 000 cm suction using the WoSIS database. *Int. Soil Water Conserv. Res.* **2023**, *11*, 225–239. [CrossRef]
38. Mushagalusa, C.A.; Fandohan, A.B.; Kakaï, R.G. Random forest and spatial cross-validation performance in predicting species abundance distributions. *Environ. Syst. Res.* **2024**, *13*, 23. [CrossRef]
39. Pérez Navarro, M.Á.; Sapes, G.; Batllori, E.; Serra-Diaz, J.M.; Esteve, M.A.; Lloret, F. Climatic Suitability Derived from Species Distribution Models Captures Community Responses to an Extreme Drought Episode. *Ecosystems* **2019**, *22*, 77–90. [CrossRef]
40. Ficetola, G.F.; Bonardi, A.; Mairota, P.; Leronni, V.; Padoa-Schioppa, E. Predicting wild boar damages to croplands in a mosaic of agricultural and natural areas. *Curr. Zool.* **2014**, *60*, 170–179. [CrossRef]
41. Allouche, O.; Tsoar, A.; Kadmon, R. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* **2006**, *43*, 1223–1232. [CrossRef]
42. Gülçin, D.; Velázquez, J.; Rincón, V.; Mongil-Manso, J.; Tonyaloğlu, E.E.; Özcan, A.U.; Ar, B.; Çiçek, K. Identification of Restoration Pathways for the Climate Adaptation of Wych Elm (*Ulmus glabra* Huds.) in Türkiye. *Land* **2025**, *14*, 1391. [CrossRef]
43. D'Arrigo, G.; Provenzano, F.; Torino, C.; Zoccali, C.; Tripepi, G. I test diagnostici e l'analisi della curva ROC. *G. Ital. Di Nefrol.* **2011**, *28*, 642–647.
44. Phillips, S.J.; Dudík, M. Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography* **2008**, *31*, 161–175. [CrossRef]
45. Hijmans, R.J.; Graham, C.H. The ability of climate envelope models to predict the effect of climate change on species distributions. *Glob. Change Biol.* **2006**, *12*, 2272–2281. [CrossRef]
46. Araújo, M.B.; Whittaker, R.J.; Ladle, R.J.; Erhard, M. Reducing uncertainty in projections of extinction risk from climate change. *Glob. Ecol. Biogeogr.* **2005**, *14*, 529–538. [CrossRef]
47. Yoon, S.; Lee, W.-H. Application of true skill statistics as a practical method for quantitatively assessing CLIMEX performance. *Ecol. Indic.* **2023**, *146*, 109830. [CrossRef]
48. Han, H.; Guo, X.; Yu, H. Variable selection using mean decrease accuracy and mean decrease gini based on random forest. In Proceedings of the 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 26–28 August 2016; IEEE: New York, NY, USA; pp. 219–224. [CrossRef]
49. Andrews, C.; Foster, J.R.; Weiskittel, A.; D'amato, A.W.; Simons-Legaard, E. Integrating historical observations alters projections of eastern North American spruce–fir habitat under climate change. *Ecosphere* **2022**, *13*, e4016. [CrossRef]
50. Khanum, R.; Mumtaz, A.S.; Kumar, S. Predicting impacts of climate change on medicinal asclepiads of Pakistan using Maxent modeling. *Acta Oecologica* **2013**, *49*, 23–31. [CrossRef]

51. Chen, Y.; Zhang, J.-H.; Chen, M.-X.; Zhu, F.-Y.; Song, T. Optimizing water conservation and utilization with a regulated deficit irrigation strategy in woody crops: A review. *Agric. Water Manag.* **2023**, *289*, 108523. [CrossRef]
52. Suleiman, A.; Tight, M.R.; Quinn, A.D. Hybrid Neural Networks and Boosted Regression Tree Models for Predicting Roadside Particulate Matter. *Environ. Model. Assess.* **2016**, *21*, 731–750. [CrossRef]
53. Anselmetto, N.; Morresi, D.; Barbarino, S.; Loglisci, N.; Betts, M.G.; Garbarino, M. Species distribution models built with local species data perform better for current time, but suffer from niche truncation. *Agric. For. Meteorol.* **2025**, *362*, 110361. [CrossRef]
54. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]
55. Huang, J.; Zhang, S.; Zhang, J.; Zheng, X.; Meng, X.; Yang, S.; Bai, Y. Integrating Meteorological and Remote Sensing Data to Simulate Cropland Nocturnal Evapotranspiration Using Machine Learning. *Sustainability* **2024**, *16*, 1987. [CrossRef]
56. Challinor, A.J.; Wheeler, T.R.; Craufurd, P.Q.; Slingo, J.M.; Grimes, D.I.F. Design and optimisation of a large-area process-based model for annual crops. *Agric. For. Meteorol.* **2004**, *124*, 99–120. [CrossRef]
57. Chen, B.; Zhao, G.; Tian, Q.; Yao, L.; Wu, G.; Wang, J.; Yu, Q. Climate-driven shifts in suitable areas of Alternaria leaf blotch (Alternaria mali Roberts) on apples: Projections and uncertainty analysis in China. *Agric. For. Meteorol.* **2025**, *364*, 110464. [CrossRef]
58. Germanà, C. Coltivazione, Gli Agrumi—Coltura & Cultura Bayer CropScience S.R.L. 2009. Available online: https://issuu.com/colturaecultura (accessed on 13 January 2025).
59. Zabihi, H.; Ahmad, A.; Vogeler, I.; Said, M.N.; Golmohammadi, M.; Golein, B.; Nilashi, M. Land suitability procedure for sustainable citrus planning using the application of the analytical network process approach and GIS. *Comput. Electron. Agric.* **2015**, *117*, 114–126. [CrossRef]
60. Bevington, K.B.; Castle, W.S. Annual root growth pattern of young citrus trees in relation to shoot growth, soil temperature, and soil water content. *J. Am. Soc. Hortic. Sci.* **1985**, *110*, 840–845. [CrossRef]
61. Agustí, M.; Reig, C.; Martínez-Fuentes, A.; Mesejo, C. Advances in Citrus Flowering: A Review. *Front. Plant Sci.* **2022**, *13*, 868831. [CrossRef]
62. Hall, A.E.; Khairi, M.M.A.; Asbell, C.W. Air and Soil Temperature Effects on Flowering of Citrus1. *J. Am. Soc. Hortic. Sci.* **1977**, *102*, 261–263. [CrossRef]
63. Vitasse, Y.; François, C.; Delpierre, N.; Dufrêne, E.; Kremer, A.; Chuine, I.; Delzon, S. Assessing the effects of climate change on the phenology of European temperate trees. *Agric. For. Meteorol.* **2011**, *151*, 969–980. [CrossRef]
64. Pignatti, S.; Guarino, R.; La Rosa, M. *Flora D'italia*; Edagricole-New Business Media: Milan, Italy, 2017; Volume 1.
65. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
66. Strobl, C.; Boulesteix, A.-L.; Kneib, T.; Augustin, T.; Zeileis, A. Conditional variable importance for random forests. *BMC Bioinform.* **2008**, *9*, 307. [CrossRef]
67. Biau, G.; Scornet, E. A random forest guided tour. *TEST* **2016**, *25*, 197–227. [CrossRef]
68. Zouabi, O.; Kadria, M. The direct and indirect effect of climate change on citrus production in Tunisia: A macro and micro spatial analysis. *Clim. Change* **2016**, *139*, 307–324. [CrossRef]
69. Appiah, S.A.; Li, J.; Lan, Y.; Alordzinu, K.E.; Al Aasmi, A.; Wang, H.; Qiao, S.; Owusu-Sekyere, E.; Afful, E.A.; Asenso, E.; et al. Regulated deficit irrigation (RDI) under citrus species production: A review. *J. Hortic. For.* **2021**, *13*, 81–95. [CrossRef]