

## Article

# GIS and SDM-Based Methodology for Resource Optimisation: Feasibility Study for Citrus in Mediterranean Area

Giuseppe Antonio Catalano, Federico Maci, Provvidenza Rita D'Urso \* and Claudia Arcidiacono

Department of Agriculture, Food and Environment (Di3A), Building and Land Engineering Section, University of Catania, 95123 Catania, Italy

\* Correspondence: provvidenza.durso@phd.unict.it

**Abstract:** South Italy is characterised by a semi-arid climate with scarce rain and high evaporative demand. Since climate change could worsen this condition, the need to optimise water resources in this area is crucial. In citrus cultivation, which involves one of the most important crops bred in Southern Italy, and more generally in Mediterranean regions, deficit irrigation strategies are implemented in order to cope with limited resource availability. On this basis, knowledge on how the territorial distribution of citrus would change in relation to these strategies represents valuable information for stakeholders. Therefore, the objective of this study was to determine the probability of the presence of citrus in Sicily based on changes in the percentage of water deficit in order to identify and analyse change in the surface area as well as the location of the crop. The methodology was based on the application of species distribution models (SDM) and Geographic Information Systems (GIS) to the case study of the province of Syracuse in Sicily. Different geostatistical and machine learning models were applied based on bioclimatic variables measured over three decades, a *Digital Terrain Model* and irrigation. Assessment of the outcomes was carried out using classification evaluation metrics. The analysis of the outcomes showed that uncorrelated predictor layers mainly included water input that most affected the probability of the presence of citrus fruits. Moreover, GIS analyses showed that deficit irrigation strategies would generate an overall reduction of cultivation surfaces in the territory (e.g., for the *Random Forest* model the surface reduction was equal to 41.15%) and a decrease of citrus presence in southern areas of the considered territory. In this area, climate conditions are less favourable in terms of temperature and precipitation; thus, these analyses provide useful information for decision support tools in agriculture and land use policy.

**Keywords:** VisTrails:SAHM software; citrus; spatial distribution; probability of presence; Mediterranean climate; predictor layers

**Citation:** 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. <https://doi.org/10.3390/agronomy13020549>

Academic Editor:  
Enrico Borgogno-Mondino

Received: 25 November 2022

Revised: 7 February 2023

Accepted: 9 February 2023

Published: 14 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

The sustainable use of natural resources is one of the most important targets of Agenda 2030. This target becomes even more crucial in agriculture considering that global warming, by producing temperatures that increase and modify weather patterns, could have considerable effects on resources vital for agriculture, such as water availability.

Sicily is a region highly suited to agriculture. In 2018, the production of citrus fruits in Sicily reached a value of around 600 million euros, calculated based on the basic prices of the citrus sector and the agricultural sector in South Italy. Sicily is the primary national region for citrus production, incorporating extended areas such as, for example, the Plain of Catania, which covers 43,000 hectares [1].

In the Syracuse province, especially in the territories of Carlentini, Lentini and Francfonte, there is an excellent quality of citrus fruits recognised at the European level with the label of PGI (Protected Geographical Indications).

In the Mediterranean semi-arid environment, irrigation plays a crucial role in the success of citrus production. It is therefore essential to manage the available water resources in a sustainable way in order to optimise the productivity of citrus groves as well as enhance their adaptation to conditions of water shortage. For instance, this has been the case in drought management in Morocco, which has a high level of production (about 2.2 million tons of citrus in 2014) and an increasing area of cultivated citrus (24 percent between 2008 and 2014) [2].

Species distribution models (SDM) and Geographic Information Systems (GIS) have been applied for different types of geospatial studies at different territorial levels (global, regional, provincial and municipal). Examples of those applications range from ecologico-climatic and geographical divergence of plant species [3] to potential biomass exploitation [4], risk assessments, prediction of future potential establishment of invasive species [5] and climate adaptation planning for protected areas [6].

SDMs are frequently used to forecast shifts in the geographic distribution of species under conditions of climate change. When associations between species ranges and environmental factors can be reliably used to estimate ecological requirements, these associations can be utilised to forecast species range shifts under climate change scenarios [7]. In this field, the use of GIS tools provides an added value to the analysis of spatial distribution of the probability of species presence and to the analysis of input and output production.

The use of satellite remote sensing to map the distribution of invasive plants has been complemented by SDMs. In particular, some authors [8] tested the response of the software and five models on the distribution of an invasive species (*Tamarix* spp.) along the Arkansas River in Southeastern Colorado, and a previous study [9] analysed the distribution of *Bromus tectorum* L. in Rocky Mountain National Park (CO, USA).

A key issue limiting the use of SDMs is the development of sound and replicable models; therefore, reliable processes should be proposed and easily replicable outputs such as response curves for expert review should be considered.

In SDMs, the application of algorithms to a specific species and specific conditions requires in-depth investigation. In fact, based on the state-of-the-art, there is a lack of investigation of the feasibility of employing these promising tools in relation to citrus production in the Mediterranean area. Specific studies are required to improve the application of these tools in order to provide useful information for decision support tools in agriculture and land use policy in specific climatic conditions.

Therefore, the main aim of the present study was to produce valuable information for resource optimisation by pursuing the following objectives: (1) investigate the feasibility of SDM application to citrus in the Mediterranean climate; (2) analyse the main factors influencing the presence of the citrus plant; (3) simulate the effects of deficit irrigation on the spatial distribution of citrus in the territory.

## 2. Materials and Methods

The methodology was based on the application of SDM and GIS tools. Different geostatistical and machine learning models were applied and the outcomes were compared using appropriate metrics. In particular, *VisTrails:SAHM* software was utilised. *VisTrails:SAHM* is an open-source provenance management and scientific workflow system designed to integrate the best of both scientific workflow and scientific visualisation systems. It combines a provenance-enabled workflow system with powerful visualisation techniques [10]. The software allowed for utilisation of the SDM algorithms (i.e., MaxEnt, Boosted Regression Tree (BRT), Multivariate Adaptive Regression Splines (MARS), Generalized Linear Model (GLM) and Random Forest (RF)) in order to predict the distribution of citrus across geographic space.

The strength of this software is the use of several algorithms that facilitate identification of the best representative model for the data.

The methodology defined in this study was developed in three phases. The first phase concerns the acquisition and processing of the input dataset, with the support of GIS tools to manage spatial data. In the second phase, the modules of the software *VisTrails:SAHM* (*VisTrails* v.2.2.3 and *SAHM* v.2.0.1) were analysed and applied. Finally, the results obtained were assessed through specific metrics and mapped by using GIS tools.

### 2.1. Study Area and Period of Simulations

The methodology was applied to the case study of the province of Syracuse in Sicily (Italy). The province of Syracuse extends for about 2100 km<sup>2</sup> and represents the southernmost geographical part of Italy (Figure 1). From a geological point of view, the Syracuse area is characterised by a mountain range called Monti Iblei [11].

Based on data availability, described in the following sections, the year 2000 was used for the simulation period, although the methodology could also be applied to different time series.



**Figure 1.** Location of Syracuse province within Sicily (Italy).

### 2.2. Input Data Acquisition and Processing

The *VisTrails:SAHM* requires data regarding the presence of the species in .csv format and other predictor layers in raster format (i.e., .tiff).

Citrus presence (Figure 1) data were obtained by overlaying the Sicilian Technical Regional Cartography (TRC) with the IT2000 orthophotos available in the Sicilian Land Information System (SITR). The orthophotos are derived from aerial acquisition, are geometrically corrected and are georeferenced within the ATA2000 project, which is commissioned by the Sicilian Region. The overlay was carried out by using GIS software (specifically, ArcGIS® for Desktop 10.3 and QGIS 3.10.0). In detail, about 10,000 georeferenced presence points were used as input data in the *VisTrails:SAHM*.

By using the GIS tools, 19 bioclimatic variables (Table 1) provided by the *WorldClim* database for three decades, from 1970 to 2000, were represented in raster format. *WorldClim* data are routinely used for cropland suitability studies because they provide a comprehensive picture of monthly, quarterly and annual bioclimatic conditions [12].

**Table 1.** Explanation of the WorldClim bioclimatic variables.

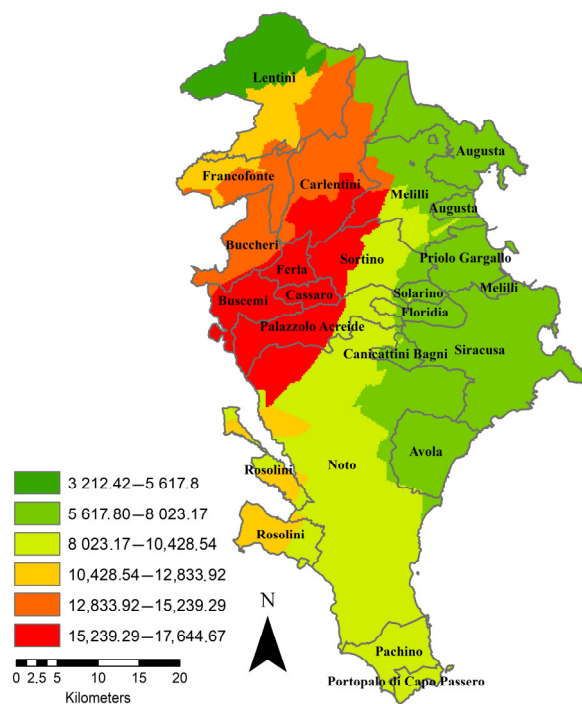
Variables	Description	Unit of Measure
BIO1	Annual Mean Temperature	Degrees Celsius
BIO2	Mean Diurnal Range	Degrees Celsius
BIO3	Isothermality	Percent
BIO4	Temperature Seasonality	Temperature (degrees Celsius)
BIO5	Max Temperature of Warmest Month	Degrees Celsius
BIO6	Min Temperature of Coldest Month	Degrees Celsius

BIO7	Temperature Annual Range	Degrees Celsius
BIO8	Mean Temperature of Wettest Quarter	Degrees Celsius
BIO9	Mean Temperature of Driest Quarter	Degrees Celsius
BIO10	Mean Temperature of Warmest Quarter	Degrees Celsius
BIO11	Mean Temperature of Coldest Quarter	Degrees Celsius
BIO12	Annual Precipitation	Millimeters
BIO13	Precipitation of Wettest Month	Millimeters
BIO14	Precipitation of Driest Month	Millimeters
BIO15	Precipitation Seasonality (Coefficient of Variation)	Percent
BIO16	Precipitation of Wettest Quarter	Millimeters
BIO17	Precipitation of Driest Quarter	Millimeters
BIO18	Precipitation of Warmest Quarter	Millimeters
BIO19	Precipitation of Coldest Quarter	Millimeters

Furthermore, the set of covariates was enriched by the *Digital Terrain Model (DTM)*, in metres, which brings valuable information about the effects of altitude on plant presence. This layer is related to a 20 m resolution and DTM\_20 was the related predictor.

To simulate the effects of deficit irrigation, the watering volume was gradually reduced by 10% from the 100% values acquired from the A.C.Q.U.A project (*‘Agrumicoltura Consapevole della Qualità e Uso dell’Acqua’ – ‘Awareness of quality and use of water in Citrus cultivation’*) [13], which surveyed the actual irrigation volumes in the area under study. These data can be acquired from the WebGIS of the project ([http://www.distrettoagrumidisicilia.it/wp-content/web\\_gis/index.html](http://www.distrettoagrumidisicilia.it/wp-content/web_gis/index.html) accessed on 5 January 2023). To transform irrigation data from point data to continuous data, the *‘Kriging Ordinary’* interpolation method was applied with default settings to the irrigation data to produce a map in raster format, hereafter named *Sir\_Irr* ( $m^3 ha^{-1}$ ), suitable for the models’ input (Figure 2).

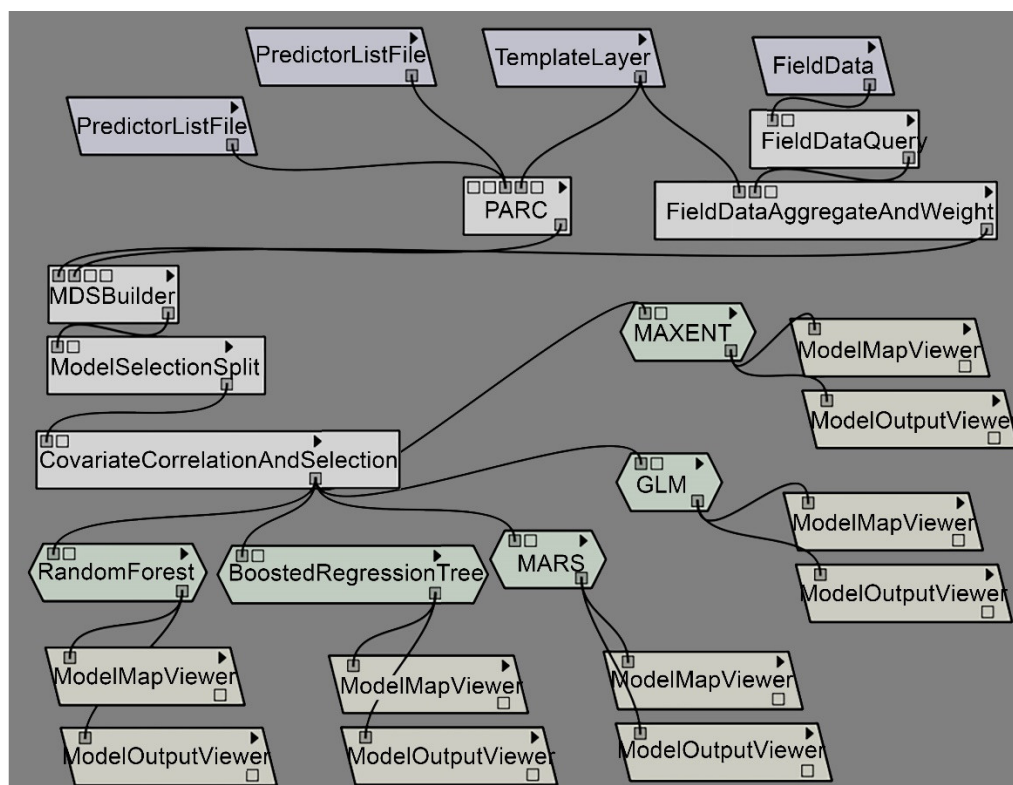
**Irrigation volumes map ( $m^3 ha^{-1}$ )**



**Figure 2.** Irrigation volumes map with data from the A.C.Q.U.A. project.

### 2.3. Modules Selection

In the second phase of the methodology, the modules within *VisTrails:SAHM* were selected based on the scheme defined by Morissette et al. [14] (Figure 3). The scheme establishes which modules are suitable for each phase of the model in relation to problems of the type examined in this study. In addition to that scheme, the *ApplyModel* module was added to allow comparison among various simulations of increasing deficit irrigation.



**Figure 3.** VisTrails v.2.2.3 pipeline with the SAHM v.2.0.1 (Software for Assisted Habitat Modeling) for this specific study.

The most important software modules, at each stage, were the following:

- in the 'Preprocessing phase', the key module to speed up the input processing phase was the *PARC* module, which allows for projection, aggregation, resampling and clipping of input geospatial data to match the *TemplateLayer*;
- in the 'Preliminary model analysis and decision' phase, the *ModelSelectionSplit* and *CovariateCorrelationAndSelection* modules were used. The first module reserves some of the data from the model training process for testing the model and reports evaluation metrics on all models. Based on the literature in this field, the training ratio of 70% with a testing ratio equal to 30% for presence data was considered in this study. In fact, the amount of presence data proved to be critical for the prediction, and those ratios produced a more robust model [6,8,9].

Finally, the *CovariateCorrelationSelector* module provides a breakpoint in the modeling workflow to allow the user to evaluate how each variable explains the distribution of the sampled data points and allows the user to remove any variables that may be highly correlated with others [15]; in fact, collinearity can lead to large model prediction errors [6]. To select only predictors that were less strongly correlated, the maximum value between the *Pearson*, *Spearman* and *Kendall coefficients* calculated for the pairs of variables was used; specifically, the threshold value for the three coefficients was set to  $\pm 0.8$  [4,16,17]. Therefore, since coefficient values higher than the thresholds infer 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.

The description of the influence of predictors in the model was analysed through the *responseCurve* graphs provided by the software for each model. They describe the values of each predictor in relation to the probability of presence in order to select the range where the probability is higher, and describe the contribution of the predictor to the probability for the specific model [18].

#### 2.4. Accuracy Measures

In this study, the measurement of the accuracy of the SDM classifications was conducted through calculation of the Receiver Operating Characteristic (ROC) curve and the related metric Area Under the Curve (AUC), a threshold-independent metric that evaluates the ability of a model to discriminate the presence from the background [8,9].

For the interpretation of the values of the area under the ROC curve it is possible to refer to the classification reported in the study of D'Arrigo et al. [19]:

- 1)  $AUC = 0.5$  the test is not informative;
- 2)  $0.5 < AUC \leq 0.7$  the test is inaccurate;
- 3)  $0.7 < AUC \leq 0.9$  the test is moderately accurate;
- 4)  $0.9 < AUC < 1.0$  the test is highly accurate;
- 5)  $AUC = 1$  perfect test.

Moreover,  $\Delta AUC$  values, computed as the difference between the AUC of the training and the AUC of the testing, were considered for assessing *overfitting*, according to Mukherjee et al. [20]. Specifically, when the  $\Delta AUC$  value exceeds 0.05, *overfitting* occurs [21].

Moreover, *True Skills Stat (TSS)* [22] was considered in this study to compare the different models by applying the following relations (Equation (1)):

$$TSS = Sensitivity + Specificity - 1 \quad (1)$$

in which *Sensitivity* (Equation (2)) (or *True Positive Rate—TPR*) and *Specificity* (Equation (3)) (or *True Negative Rate—TNR*) are defined as:

$$Sensitivity (TPR) = \frac{TP}{TP + FN} \quad (2)$$

$$Specificity (TNR) = \frac{TN}{TN + FP} \quad (3)$$

where *TP* is the number of True Positives, *FN* is the number of false negatives, *TN* is the number of True Negatives, and *FP* is the number of False Positives.

The thresholds for significance of these metrics are referred to as random rankings. These have on average an AUC value of 0.5, whereas a perfect ranking achieves the best possible AUC value equal to 1.0; models with values above 0.75 are considered potentially useful [23]. TSS ranges from  $-1$  to  $+1$ , where  $+1$  indicates perfect agreement and zero or negative values indicate a performance no better than random [19]. In this study, according to some authors [24,25], the difference  $\Delta TSS$  between training and testing was computed in order to analyse TSS.

To divide the continuous model predictions into binary presence/absence predictions, the maps were produced by using thresholds specifically computed by the software for each model [26].

Elaborations on output surfaces at a 10% step of probability (classes 1 to 10) were carried out by computing the weighted variation in percentage between the surface  $Sup_i$  related to the considered deficit irrigation level and the surface  $Sup_{100}$  related to 100% irrigation, according to the following relation (Equation (4)):

$$Dsup = \frac{(Sup_i - Sup_{100})}{Sup_{100}} \quad (4)$$

Moreover, other elaborations were carried out to relate the surface area of the class to the overall surface of the province.

### 3. Results and Discussion

All models showed a quite stable output: there was wide consensus among the different models on the location of areas with the highest probability of presence for the species. Specifically, the results showed that the highest probability of presence of citrus trees in the study area was found in the northern areas of the province and in eastern ones (i.e., in the municipalities of Syracuse, Noto and Avola). The central and southern areas were considered unsuitable by all models (Figure 4). General results were consistent with data showing that reliable processes have been proposed.

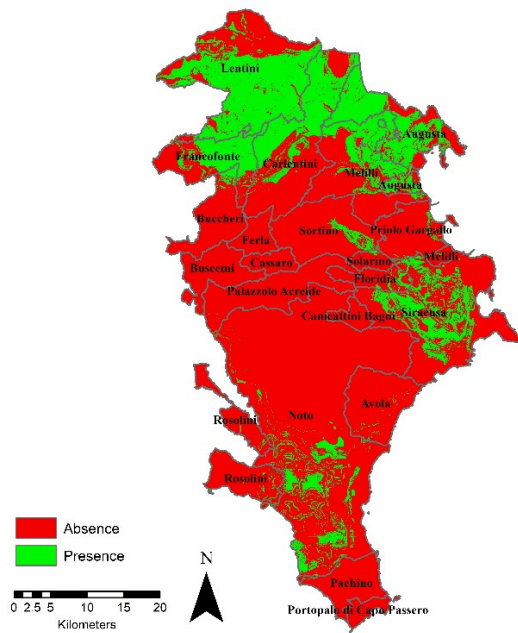
In Figure 4, according to Akpoti [18], the probability was subdivided into presence/absence through the utilisation of the threshold values computed by each SDM in order to allow comparisons among the different models' outputs. The threshold values obtained for this case study were: 0.59 for RF; 0.55 for Mars; 0.4095 for MaxEnt; 0.61 for GLM; and 0.57 for BRT.

Based on these thresholds and the use of GIS tools, a detailed territorial analysis at the provincial level was carried out (Figure 4). Presence areas (in green colour) have a continuous aspect in GLM, MARS and MaxEnt models (Figure 4b–d), whereas in RF and BRT models (Figure 4a,e) those areas are composed of polygons with holes. The greatest differences among the models' outputs were found in the northern part of the province (especially in the municipalities of Francofonte, Carlentini and Augusta) and in the inner part of the territory (municipalities of Sortino and Floridia), where, for instance, the MARS model showed no presence of the species.

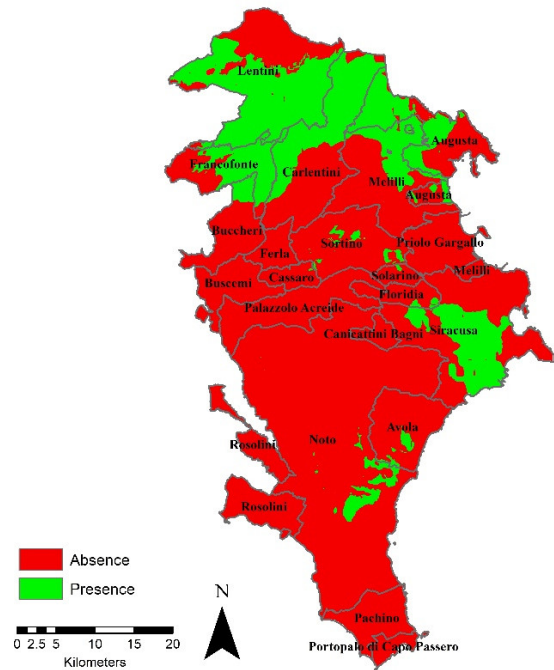
The overall surface area (km<sup>2</sup>) where the models predict a probability of presence above the threshold (in green) are the following: 519.59 for BRT; 484.35 for GLM; 505.17 for MARS; 676.30 for MaxEnt; and 401.45 for RF. Therefore, the RF model underestimated most compared to the average value of the model predictions, whereas the MaxEnt model overestimated. The differences among the simulations of the presence areas obtained by the BRT, GLM and MARS models were not great (with a maximum of about 35 km<sup>2</sup>), and the information of the territorial distribution of probability of presence acquired by GIS representation was extremely valuable.

For a more in-depth spatial analysis of the results, the probability was subdivided into 10 classes, at 10% intervals of probability, in order to refine comparison among the different areas (Figure 5). The surfaces of each class were calculated for each model by using the QGIS software (Table 2). From the comparison of Class 10 obtained by the RF and BRT models, the localisation of the areas was similar (in the municipalities of Lentini, Augusta, Carlentini, Francofonte, Melilli, Sortino, Siracusa and Avola e Noto) and quite spread out, though the surfaces were wider in the RF model than in the BRT model. In the GLM model, Class 10 was found in a lower number of areas (in the municipalities of Francofonte and Lentini) and in even fewer for the MARS model (in Carlentini municipality). The greatest surface areas for Class 1 (1178.5 km<sup>2</sup>) and Class 10 (116.1 km<sup>2</sup>) were confirmed for the RF model.

**BRT Probability Map**



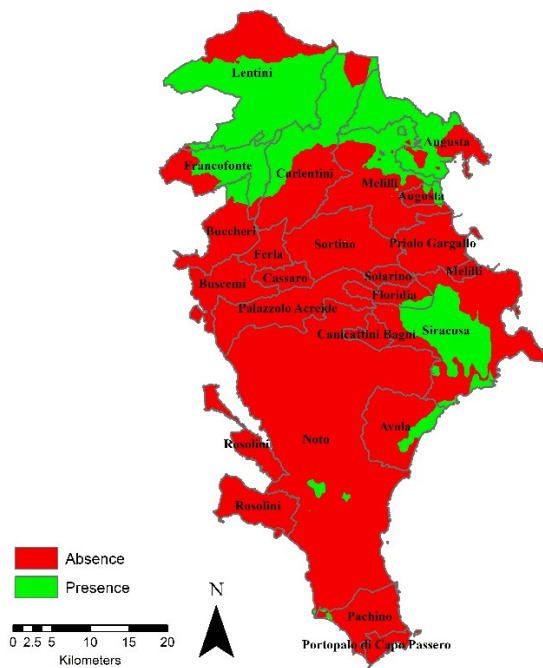
**GLM Probability Map**



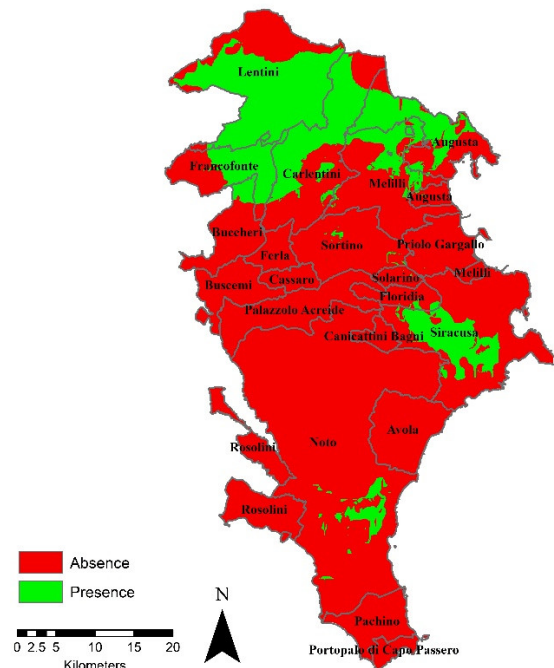
(a)

(b)

**MARS Probability Map**



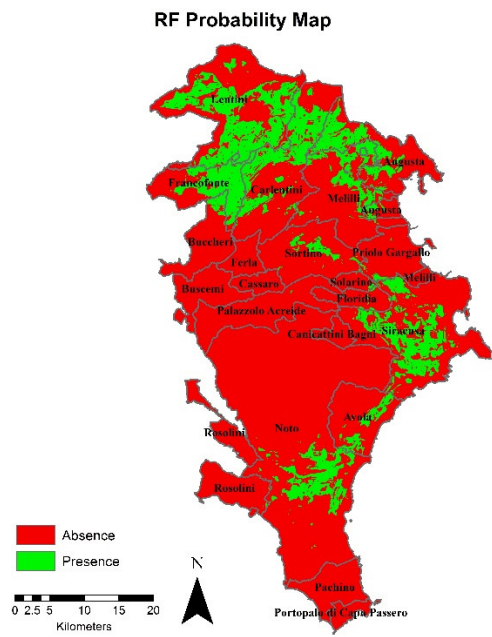
**MAXENT Probability Map**



(c)

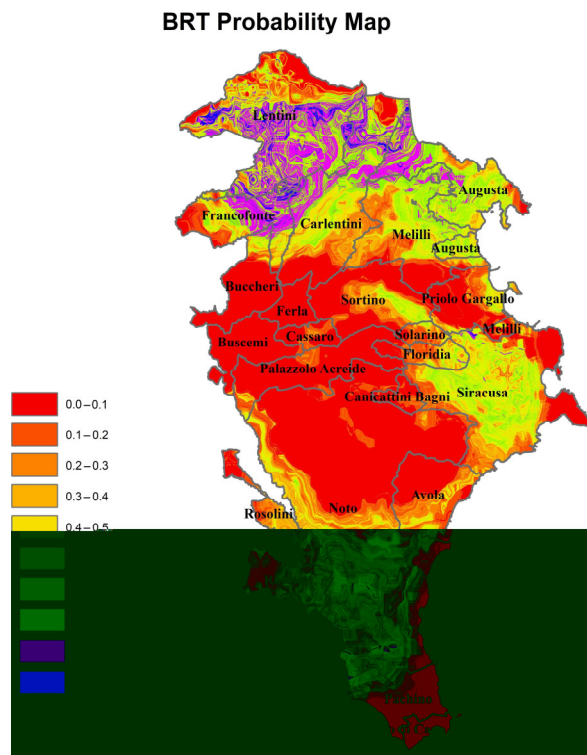
(d)



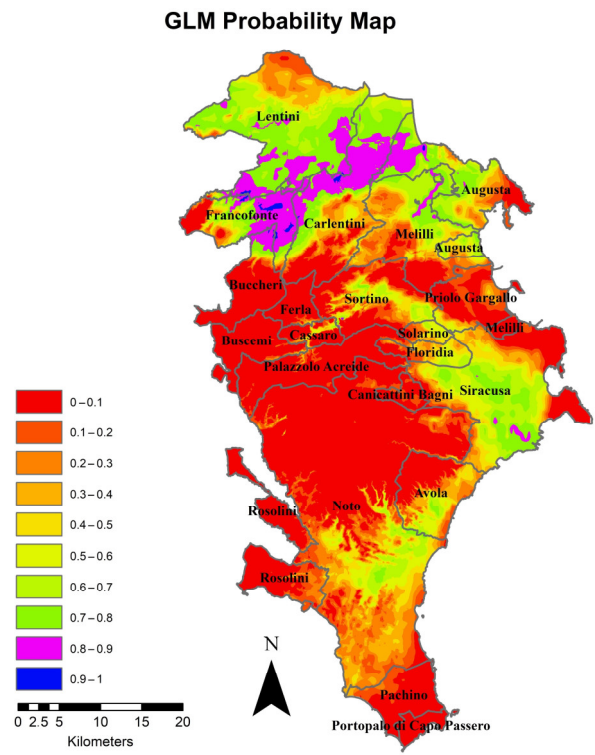


(e)

Figure 4. Probability Maps, obtained by applying the threshold for probability for each model, in the province of Syracuse: (a) BRT; (b) GLM; (c) MARS; (d) MaxEnt; and (e) RF.



(a)



(b)

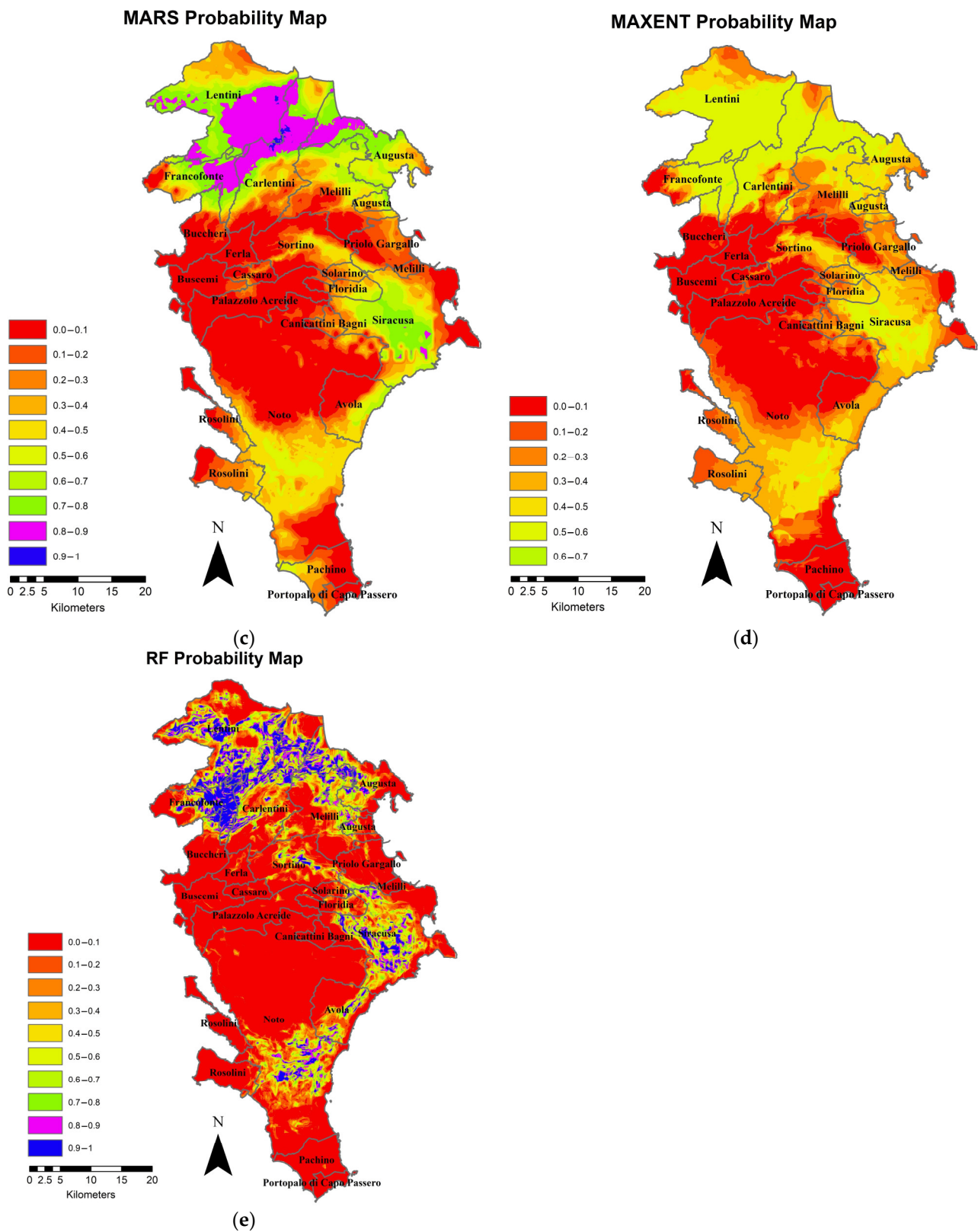


Figure 5. Probability Maps, obtained by applying the 10-classes subdivision for probability, for each model in the province of Syracuse: (a) BRT; (b) GLM; (c) MARS; (d) MaxEnt; and (e) RF.

Moreover, the sum of the surfaces from Class 7 to Class 10 for each model that required summing the values from Class 5 to 10, except for the MaxEnt model, confirmed the outcomes obtained by the application of the thresholds, i.e., MaxEnt had the highest area (703.1 km<sup>2</sup>) whereas RF had the lowest one (393.3 km<sup>2</sup>). GLM was close to the mean

value of the overall surface computed on all of the models' outputs (496.46 km<sup>2</sup>) and BRT and MARS slightly overestimated the overall surface area (444.8 and 442.7 km<sup>2</sup>, respectively), whereas MaxEnt highly overestimated and RF moderately underestimated compared to the average value.

**Table 2.** Values of the surface areas (km<sup>2</sup>) for each class and model: in bold green the presence data, while in red the absence data.

Class	BRT	RF	MAXENT	GLM	MARS
1	961.8	1178.5	688.7	826.0	715.9
2	187.9	175.3	197.3	189.6	210.9
3	140.3	111.2	235.4	175.0	173.4
4	133.3	84.5	278.4	163.9	191.8
5	125.4	79.8	<b>315.5</b>	120.1	232.2
6	115.6	80.3	<b>385.4</b>	130.1	136.1
7	<b>122.50</b>	<b>82.3</b>	<b>2.2</b>	<b>176.6</b>	<b>122</b>
8	<b>147.90</b>	<b>88.1</b>	<b>0.0</b>	<b>190.4</b>	<b>152.8</b>
9	<b>153.40</b>	<b>106.8</b>	<b>0.0</b>	<b>127.1</b>	<b>164.3</b>
10	<b>21.00</b>	<b>116.1</b>	<b>0.0</b>	<b>4.3</b>	<b>3.6</b>

To assess which model had a higher capability to estimate the probability of the presence of citrus, the metrics produced by the SDM in VisTrails:SAHM were considered and analysed.

The analysis of the metrics for the classifications (Tables 2 and 3) highlighted that the AUC of all the models exceeded 0.83; therefore, the classifications were assessed as 'very good', and the TSS was higher than 0.50.

Based on the results reported in Table 2, the BRT model had the highest metrics for training among the models whereas the MARS model showed the lowest metrics.

In Table 3, the BRT model shows a  $\Delta$ AUC equal to 0.08, thus highlighting that model *overfitting* is high, and therefore this SDM was not considered as adequate. In detail, *overfitting* would make the algorithm produce very different predictions for similar data (low bias and high variance). The lowest  $\Delta$ AUC equal to 0 was found for the RF model, while the values for MARS, GLM and MaxEnt showed a  $\Delta$ AUC in the order of hundredths, thus also suitable because of the low *overfitting* associated.

Moreover, for  $\Delta$ TSS, Table 4 shows that the BRT model had a high value (about 0.15) compared to those of the other SDMs, which had values in the range 0 ÷ 0.005.

Based on all the considerations described above, RF can be considered the model with the highest ability to predict citrus coverage. Therefore, the subsequent simulations on deficit irrigation were carried out by applying the RF model.

**Table 3.** Values of the metrics for training and testing and for each model.

	Model Metrics for Training					Model Metrics for Testing				
	BRT	GLM	MARS	MAXENT	RF	BRT	GLM	MARS	MAXENT	RF
AUC	0.91	0.85	0.83	0.86	0.88	0.83	0.85	0.83	0.85	0.88
PCC	82.60	76.30	75.70	77.60	81.10	75.15	76.33	75.40	77.70	80.82
TPR	0.82	0.76	0.76	0.77	0.81	0.75	0.76	0.76	0.79	0.81
TNR	0.83	0.77	0.76	0.78	0.81	0.75	0.77	0.75	0.77	0.81
TSS	0.65	0.53	0.51	0.55	0.62	0.50	0.53	0.51	0.55	0.62

**Table 4.** Values of  $\Delta$ AUC e  $\Delta$ TSS for the different SDMs.

	<b>BRT</b>	<b>GLM</b>	<b>MARS</b>	<b>MAXENT</b>	<b>RF</b>
$\Delta$ AUC	0.0819	0.002	- 0.001	0.006	0
$\Delta$ TSS	0.148941	- 0.001	0	- 0.003	0.005

### 3.1. Deficit Irrigation

To simulate the effects of deficit irrigation, the watering volume was gradually reduced by 10%, from 100% of the actual irrigation volume to 50%.

Based on the results of the elaborations, the probability maps showed that the increasing reduction of irrigation volume produced a reduction of the probability values of species presence, though the localisation of the areas was similar (Figure 6). Moreover, the greater the reduction of irrigation volume the greater the reduction of probability. The considerations derived from the maps are confirmed by the numerical data processed (Tables 5 and 6).

For RF, the sum of the surface areas from Class 7 to 10 proved that the deficit irrigation simulations would cause a maximum surface reduction of 173.41 km<sup>2</sup> (at a deficit irrigation equal to 50% of the actual) and a minimum one of 75.99 km<sup>2</sup> (at a deficit irrigation equal to 90% of the actual value) with an average value of 122.32 km<sup>2</sup> (Table 5).

The reduction of irrigation would highly affect the probability of species presence, especially for Classes 9 and 10. For instance, for a 90% irrigation the surface of Class 10 would decrease by 85.36% (compared to 100% irrigation), and would decrease from 5.51% to 0.81% in relation to the whole surface of the province (Tables 6 and 7). In contrast, for a deficit irrigation equal to 80% of the actual value, the surface area for Class 10 is 9.20 km<sup>2</sup> (- 92.07%) and 37.56 km<sup>2</sup> (- 64.82%) for Class 9 (Tables 5 and 6). However, an increase of the surface of intermediate classes, mainly 6 and 7, would occur. For instance, the surface of Class 7 would increase by 65.29 km<sup>2</sup> (mainly in the municipality of Lentini) compared to 100% irrigation, under the hypothesis of a reduction of irrigation to 60%.

Although the loss of probability in one class could be compensated by the area in a lower class (albeit always considering the Classes 7 to 10, that have a higher probability than about 0.6, according to the thresholds of presence/absence), the surface loss would range from 19.3% (at 90% deficit irrigation) to 44.09% (at 50% deficit irrigation).

In Table 6, the weighted variation in percentage between the deficit irrigation level and the 100% irrigation was reported;  $D_{sup}$  shows that the probability of presence drastically reduced for Classes 8 to 10 during increased irrigation reduction in the territory. This is confirmed in the probability maps, where as the irrigation contribution decreases there is a progressive reduction in the extent of the presence of the species in the areas until the presence remains only in the north and east of the provincial territory.

Therefore, the spatial analysis outcomes show that eastern and northern areas of the province would be the most suitable for deficit irrigation for this kind of species, whereas for the southern citrus producing areas of the province it would not be advisable to perform deficit irrigation.

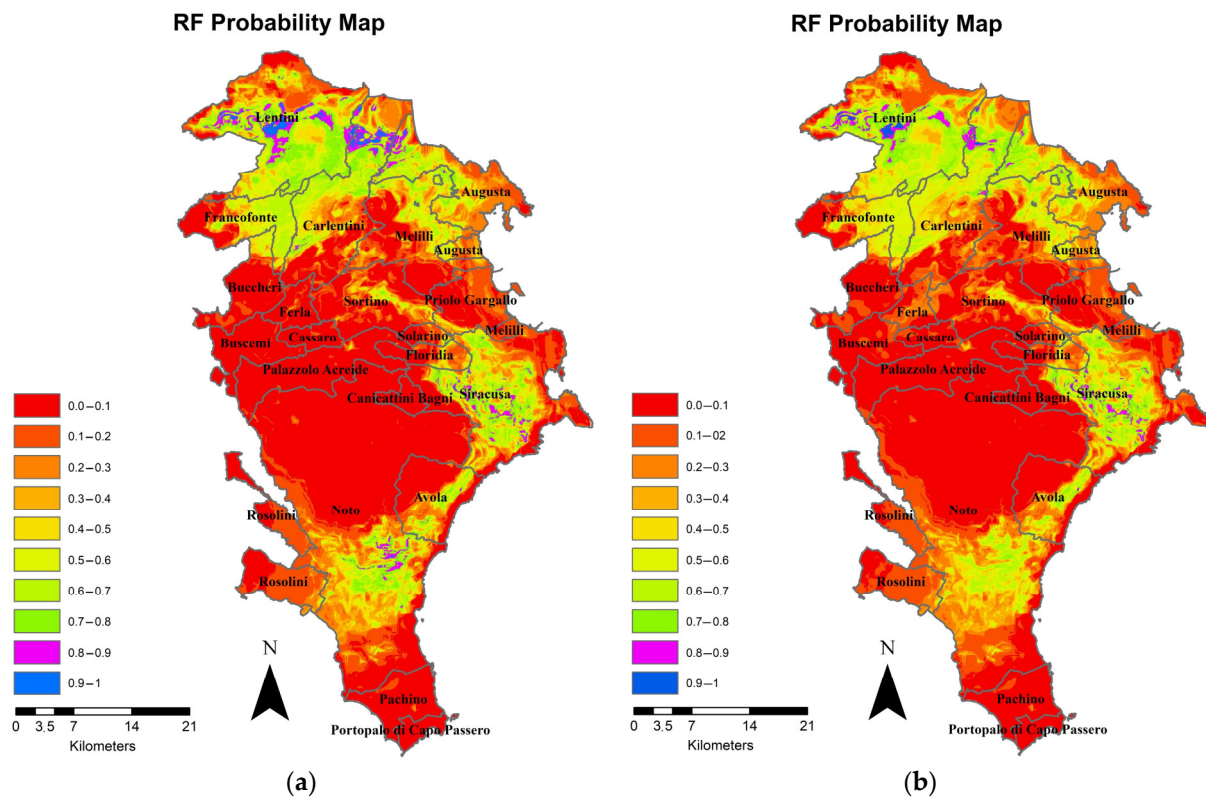


Figure 6. Probability Maps for RF model, reporting a deficit irrigation of 80% (a) and 50% (b) of the actual value, in the province of Syracuse.

Table 5. Values of the surface areas  $S_i$  (km<sup>2</sup>) for the deficit irrigation from 100% to 50% of the actual value and the ten classes, for the RF model.

Surface Areas $S_i$ [km <sup>2</sup> ] for RF										
Sir_Irr [%]	Class 1 (0–0.1)	Class 2 (0.1–0.2)	Class 3 (0.2–0.3)	Class 4 (0.3–0.4)	Class 5 (0.4–0.5)	Class 6 (0.5–0.6)	Class 7 (0.6–0.7)	Class 8 (0.7–0.8)	Class 9 (0.8–0.9)	Class 10 (0.9–1)
100	1178.51	175.31	111.24	84.47	79.82	80.29	82.30	88.11	106.78	116.12
90	1037.24	258.91	135.18	114.69	110.68	128.96	133.06	104.70	62.56	17.00
80	938.60	297.99	173.17	132.91	126.14	150.23	146.69	90.47	37.56	9.20
70	897.24	323.55	190.02	136.93	126.88	151.10	148.63	88.85	33.04	6.73
60	878.25	338.28	195.36	145.18	130.22	159.13	147.59	75.55	29.10	4.30
50	873.12	347.13	191.05	147.93	142.12	181.72	130.23	66.01	20.48	3.18

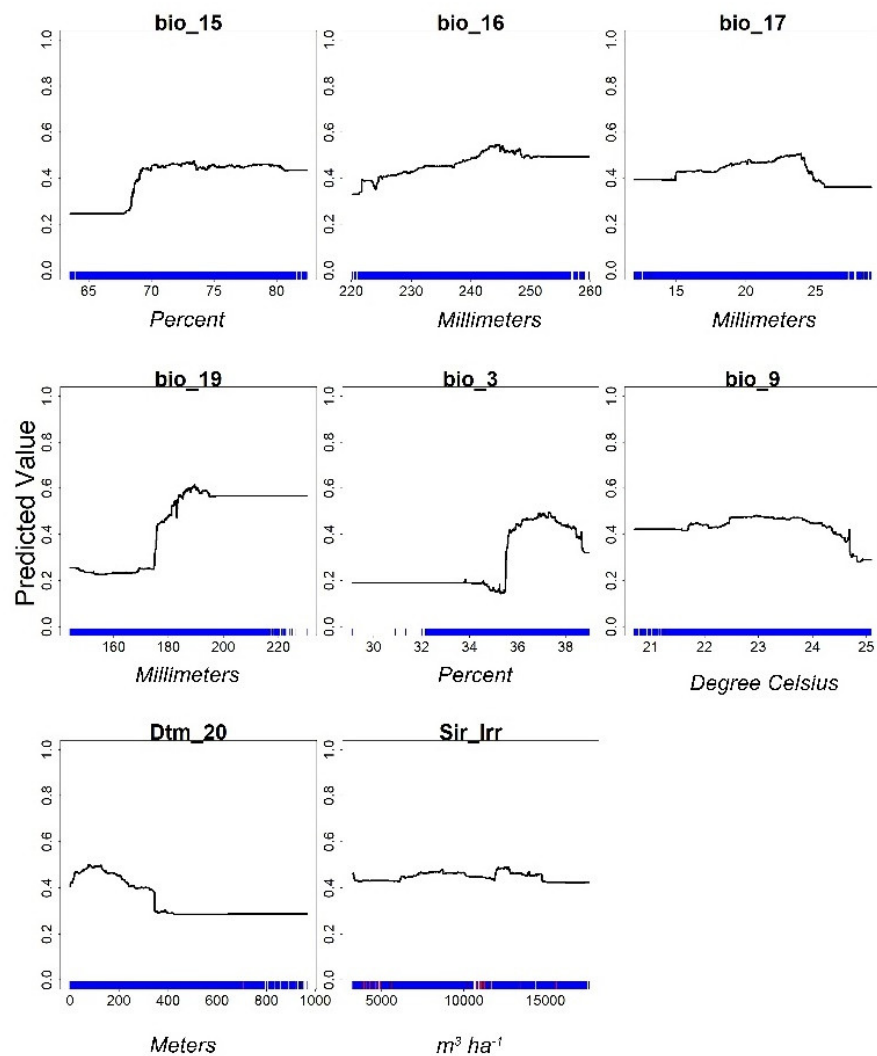
Table 6. Values of the weighted difference in percentage  $D_{sup}$ , for the deficit irrigation from 100% to 50% of the actual value and the ten classes, for the RF model.

$D_{sup}$ [%]										
Sir_Irr [%]	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10
90	- 11.99	47.69	21.51	35.77	38.65	60.62	61.67	18.82	- 41.42	- 85.36
80	- 20.36	69.98	55.67	57.34	58.02	87.10	78.24	2.68	- 64.82	- 92.07
70	- 23.87	84.55	70.81	62.09	58.96	88.19	80.59	0.84	- 69.06	- 94.20
60	- 25.48	92.96	75.61	71.87	63.14	98.19	79.32	- 14.26	- 72.75	- 96.29
50	- 25.91	98.00	71.74	75.13	78.04	126.33	58.23	- 25.08	- 80.82	- 97.26

**Table 7.** Values of the ratio between the surface areas  $S_i$  and the total area of the province  $S_{tot}$  for the deficit irrigation from 100% to 50% of the actual value and the ten classes, for the RF model.

Sir_Irr [%]	Si/Stot [%]									
	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10
100	55.88	8.31	5.27	4.01	3.78	3.81	3.90	4.18	5.06	5.51
90	49.18	12.28	6.41	5.44	5.25	6.11	6.31	4.96	2.97	0.81
80	49.18	12.28	6.41	5.44	5.25	6.11	6.31	4.96	2.97	0.44
70	44.50	14.13	8.21	6.30	5.98	7.12	6.96	4.29	1.78	0.32
60	42.54	15.34	9.01	6.49	6.02	7.16	7.05	4.21	1.57	0.20
50	41.40	16.46	9.06	7.01	6.74	8.62	6.17	3.13	0.97	0.15

The analysis of the *ResponseCurve* produced for the RF model (Figure 7) provided useful information about the range of the predictors that most contribute to the probability values.



**Figure 7.** ResponseCurves of the predictors for the RF model, for the case study analysed.

The predictors Bio\_15 (Precipitation Seasonality (Coefficient of Variation)), Bio\_16 (Precipitation of Wettest Quarter), Bio\_19 (Precipitation of Coldest Quarter), Bio\_9 (Mean

Temperature of Driest Quarter), Bio\_17 (Precipitation of Driest Quarter), Bio\_3 (Isothermality), DTM\_20 and Sir\_Irr were considered by the models as those affecting the probability of the presence of citrus in the case study analysed.

The analysis of the *ResponseCurve* graphs produced valuable information on each predictor range associated with the fitted values while holding all other predictors constant at their means. These graphs can be studied to assess whether the relationships agree with the biological meaning of the species under study [15].

Specifically, the graph for DTM\_20 showed that the altitudes suitable for citrus production were correctly identified by the RF model, i.e., points with a probability greater than the threshold were found for altitudes lower than 400 m a.s.l. [27,28]. The response curves of predictors showed the following ranges where the predicted value for the probability of citrus presence is higher: Bio\_3 between 35% and 39%, showing the diurnal temperature range lower than annual one; Bio\_19 greater than 180 mm, representing the level of overall precipitation during the coldest quarter of the year; Bio\_15 greater than 70%, showing the variation of monthly precipitation during the year; Bio\_16 above 220 mm of rain, showing the precipitation amount in the wettest quarter; Bio\_17 above 15 mm, indicating the minimum precipitation of the driest quarter. The contribution to prediction, equal to about 0.45, deriving from the values of Sir\_Irr confirms the importance of water input for the citrus crop; in fact, most of the influential bioclimatic variables are related to precipitation.

It is well known, in fact, that citrus trees require large volumes of water compared to other tree crops, especially when precipitation scarcity is recurrent, and require suitable temperature levels and other growing conditions beneficial for achieving high quality productions. Therefore, water consumption is one of the most demanding issues for the citrus sector, especially in times of climate variability and change.

#### 4. Conclusions

The research study described in this paper investigated the feasibility of applying algorithms for SDMs to predict citrus distribution in a territory in order to derive information on SDM application in the Mediterranean climate and analyse the main factors that influence the presence of the citrus plant. The aim of providing improved knowledge on spatial distribution of the species was also achieved by analysing the effects of deficit irrigation in the case study of the province of Syracuse, Italy.

This study represents the first step toward an in-depth spatial knowledge of citrus in Mediterranean areas in relation to bioclimatic variables and other driving factors. Climate covariates and terrain elevation as well as irrigation were analysed in this study as major predictors suitable for this knowledge. General uniformities in the models' predictions suggest that the multi-model approach contributes to an increased consistency of outcomes. Modeling showed that the BRT and RF models produced higher evaluation metrics compared to the other models; however, the BRT model suffered from *overfitting*. GIS contributed to the analysis of the outcomes by showing and quantifying the spatial distribution of citrus presence as well as by allowing comparison among the simulations of different levels of irrigation.

This study was limited to the analysis of outcomes related to the default settings of the models' parameters. Investigation on model parameters is the object of on-going studies aimed at fine-tuning the performance of the predictions, and research is in progress to investigate input data by involving computation of the bioclimatic data from local weather stations. Further analysis could consider predictions of future probability of species presence based on climate models for the determination of future bioclimatic predictors.

**Author Contributions:** Conceptualization, C.A. and G.A.C.; methodology, C.A. and G.A.C.; software, G.A.C.; validation, P.R.D. and C.A.; formal analysis, P.R.D. and F.M.; investigation, P.R.D. and C.A.; resources, C.A.; data curation, G.A.C. and F.M.; writing—original draft preparation, C.A.

and G.A.C.; writing—review and editing, C.A., F.M. and P.R.D.; visualization, 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 research study was funded by the University of Catania through the project ‘PON “RICERCA E INNOVAZIONE” 2014–2020, Azione II – Obiettivo Specifico 1b – Progetto “Miglioramento delle produzioni agroalimentari mediterranee in condizioni di carenza di risorse idriche—WATER4AGRI FOOD”, Cod. progetto: ARS01\_00825, CUP: B64I20000160005 and it is in line with ‘Piano incentivi per la ricerca di Ateneo 2020–2022-Linea 2’ project on ‘Engineering solutions for sustainable development of agricultural buildings and land—LANDSUS’ (ID: 5A722192152) coordinated by Prof. Claudia Arcidiacono.

**Data Availability Statement:** Data available upon request.

**Acknowledgments:** The authors wish to thank the Sicilian Region for Sitr data (<https://www.sitr.regione.sicilia.it/> accessed on 13th June 2022).

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

## References

- Del Bravo, F.; Finizia, A.; Lo Moriello, M.S.; Ronga, M. *La Competitività Della Filiera Agrumicola in Italia*; Rete Rurale Nazionale 2014–2020: 2020, ISMEA, Rome, Italy.
- Verner, D.; Tréguer, D.; Redwood, J.; Christensen, J.; McDonnell, R.; Elbert, C.; Konishi, Y.; Belghazi, S. Climate Variability, Drought, and Drought Management in Morocco’s Agricultural Sector—World Bank Report 2017. Available online: <https://openknowledge.worldbank.org/bitstream/handle/10986/30603/130404-WP-P159851-Morocco-WEB.pdf> (accessed on 23 November 2022).
- Olonova, M.V.; Vysokikh, T.S.; Mezina, N.S. Structure of Ecologo-Climatic Niches of *Poa palustris* L. and *P. nemoralis* L. (Poaceae) in Asian Russia. *Contemp. Probl. Ecol.* **2018**, *11*, 604–613. <https://doi.org/10.1134/S1995425518060069>.
- Leanza, P.M.; Valenti, F.; D’Urso, P.R.; Arcidiacono, C. A combined MaxEnt and GIS-based methodology to estimate cactus pear biomass distribution: Application to an area of southern Italy. *Biofuels. Bioprod. Biorefining* **2022**, *16*, 54–67. <https://doi.org/10.1002/bbb.2304>.
- 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. <https://doi.org/10.1111/risa.13230>.
- Piekielek, N.B.; Hansen, A.J.; Chang, T. Using custom scientific workflow software and GIS to inform protected area climate adaptation planning in the Greater Yellowstone Ecosystem. *Ecol. Inform.* **2015**, *30*, 40–48. <https://doi.org/10.1016/j.ecoinf.2015.08.010>.
- Diniz-Filho, J.A.F.; Mauricio Bini, L.; Fernando Rangel, T.; Loyola, R.D.; Hof, C.; Nogués-Bravo, D.; Araújo, M.B. Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography* **2009**, *32*, 897–906. 2009. <https://doi.org/10.1111/j.1600-0587.2009.06196>.
- West, A.M.; Evangelista, P.H.; Jarnevich, C.S.; Young, N.E.; Stohlgren, T.J.; Talbert, C.; Talbert, M.; Morisette, J.; Anderson, R. Integrating Remote Sensing with Species Distribution Models; Mapping Tamarisk Invasions Using the Software for Assisted Habitat Modeling (SAHM). *J. Vis. Exp.* **2016a**, *116*, e54578. <https://doi.org/10.3791/54578>.
- West, A.M.; Kumar, S.; Brown, C.S.; Stohlgren, T.J.; Bromberg, J. Field validation of an invasive species Maxent model. *Ecol. Inform.* **2016b**, *36*, 126–134. <https://doi.org/10.1016/j.ecoinf.2016.11.001>.
- Freire, J.; Silva, C.T.; Callahan, S.P.; Santos, E.; Scheidegger, C.E.; Vo, H.T. Managing Rapidly-Evolving Scientific Workflows. In *Provenance and Annotation of Data*; Lecture Notes in Computer Science; Moreau, L., Foster, I., Eds.; IPAW 2006: Chicago, IL, USA; Springer: Berlin, Heidelberg, 2006; volume 4145. [https://doi.org/10.1007/11890850\\_2](https://doi.org/10.1007/11890850_2).
- 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 third national silviculture congress held on 16–19 October in Taormina, Sicily, Italy, volume 1, p.234–239 <https://doi.org/10.4129/CNS2008.030>.
- Fitzgibbon, A.; Pisut, D.; Fleisher, D. Evaluation of Maximum Entropy (Maxent) Machine Learning Model to Assess Relationships between Climate and Corn Suitability. *Land* **2022**, *11*, 1382. <https://doi.org/10.3390/land11091382>.
- Università Degli Studi di Catania; CREA; Distretto Agrumi Sicilia; Coca Cola Foundation; A.C.Q.U.A. Project Results, 2020. Available online: <https://www.distrettoagruidisicilia.it/wp-content/uploads/Dossier-Acqua5.pdf> (accessed on 6 September 2022).
- 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. <https://doi.org/10.1111/j.1600-0587.2012.07815.x>.
- 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.



16. Yang, X.Q.; Kushwaha, S.P.S.; Saran, S.; Xu, J.; Roy, P.S. Maxent modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in Lesser Himalayan foothills. *Ecol. Eng.* **2013**, *51*, 83–87. <https://doi.org/10.1016/j.ecoleng.2012.12.004>.
17. Yi, Y.J.; Cheng, X.; Yang, Z.F.; Zhang, S.H. Maxent modeling for predicting the potential distribution of endangered medicinal plant (*H. riparia* Lour) in Yunnan, China. *Ecol. Eng.* **2016**, *92*, 260–269. <https://doi.org/10.1016/j.ecoleng.2016.04.010>.
18. 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. <https://doi.org/10.1016/j.scitotenv.2019.136165>.
19. D'Arrigo, G.; Provenzano, F.; Torino, C.; Zoccali, C.; Tripepi, G. I test diagnostici e l'analisi della curva ROC. *G Ital Nefrol* **2011**, *28*, 642–647.
20. 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. <https://doi.org/10.1016/j.scitotenv.2021.146031>.
21. West, A. M.; Kumar, S.; Brown, C. S.; Stohlgren, T. J.; Bromberg, J. Field validation of an invasive species Maxent model. *Ecological Informatics*, *36*, 126–134. 2016. <https://doi.org/10.1016/j.ecoinf.2016.11.001>.
22. 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. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>.
23. Phillips, S.J.; Dudík, M. Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography* **2008**, *31*, 161–175. <https://doi.org/10.1111/j.0906-7590.2008.5203.x>.
24. Baer, K.C.; Gray, A.N. Biotic predictors improve species distribution models for invasive plants in Western US Forests at high but not low spatial resolutions. *For. Ecol. Manag.* **2022**, *518*, 120249. <https://doi.org/10.1016/j.foreco.2022.120249>.
25. Brun, P.; Kjørboe, T.; Licandro, P.; Payne, M.R. The predictive skill of species distribution models for plankton in a changing climate. *Glob. Chang. Biol.* **2016**, *22*, 3170–3181. <https://doi.org/10.1111/gcb.13274>.
26. Brun, P.; Payne, M.R.; Kjørboe, T. Trait biogeography of marine copepods—An analysis across scales. *Ecol. Lett.* **2016**, *19*, 1403–1413. <https://doi.org/10.1111/ele.12688>.
27. Pignatti, S. 2017–2019. “Flora d’Italia”, 2017–2019 *Citrus limon* (L.) Burm. fil. Volume 2, p.1090. key 4-682.
28. Pignatti, S. La Flora d’Italia. *Edagricole* **1982**, *2*, 54.

**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.