doi:10.20944/preprints202211.0502.v1

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 [1*], Claudia Arcidiacono [1]

[1] Department of Agriculture, Food and Environment (Di3A), Building and Land Engineering Section, University of Catania, Catania, Italy; provvidenza.durso@phd.unict.it; carcidi@unict.it;

* Correspondence provvidenza.durso@phd.unict.it,

Abstract: South Italy is characterised by a semi-arid climate with scarce rain and high evaporative demand, therefore the need to optimise water resources in this area is crucial, and climate change could worsen this condition. In citrus cultivation, which is 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 constitutes valuable information for the stakeholders. Therefore, the objective of this study was to determine the probability of presence of citrus in Sicily at changing of the percentage of water deficit application, in order to analyse change in the surface area and localisation of the crop. The methodology was based on the application of Species Distribution Models and Geographic Information Systems to the case study of the Province of Syracuse in Sicily. Different geostatistical and machine learning models were applied, based on 3-decades bioclimatic variables, DTM and irrigation; assessment of the outcomes was carried by using classification evaluation metrics. The analysis of the outcomes showed that uncorrelated predictor layers mainly included water input that affected most the probability of presence. Moreover, GIS analyses showed that deficit irrigation strategies would generate an overall reduction of cultivation surfaces in the territory and a decrease of citrus presence in southern areas of the considered territory, where climate conditions are less favourable in terms of temperature and precipitations, thus providing 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

1. Introduction

The sustainable use of the natural resources is one of the most important targets of Agenda 2030. This target become even more crucial in agriculture considering that global warming, by producing temperature increase and modifying weather patterns, would determine considerable effects on those 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 at the basic prices of the citrus sector and the agricultural sector in South Italy. Sicily holds the primacy of national production, having considerable extensions such as that, for example, of the Plain of Catania, equal to 43,000 hectares [1].

In the Syracuse province, especially in the territories of Carlentini, Lentini and Francofonte, there is an excellent quality of citrus fruits recognised at European level with the recognition of the 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 the citrus groves,

 doi:10.20944/preprints202211.0502.v1

while enhancing their adaptation to water shortage conditions. For instance, this is the case of drought management in Morocco that has a high production (about 2.2 million tons of citrus in 2014), and an increase of citrus cultivated area (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 ecologo-climatic and geographical divergence of plant species [3], to potential biomass exploitation [4], risk assessments and prediction of future potential establishment of invasive species [5], and climate adaptation planning of protected areas [6].

Species distribution models (SDM), are frequently used to forecast shifts in species geographic distributions under climate change. When associations between species ranges and environmental factors can be reliably used to estimate the 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 presence and on input and output production.

Therefore, the main aim of the present study was to produce valuable information for resource optimization by pursuing the following objectives: (1) investigate feasibility of SDMs application to citrus in Mediterranean climate; 2) analyse the main factors that influence the presence of 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 Species distribution models (SDM) and Geographic Information System (GIS) tools. Different geostatistical and machine learning models were applied and assessment of the outcomes were compared by using appropriate metrics. In detail, VISTRAILS:SAHM software allowed utilisation of the SDMs 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 methodology defined in this study was developed in three phases. The first phase concerns the acquisition and processing of the input data set, with the support of GIS tools to manage spatial data. In the second phase, the modules of the software *VISTRAILS: SAHM* 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^2 and represents the southernmost geographical part of Italy (Fig. 1). From a geological point of view, the Syracuse area is characterised by a mountain range called Monti Iblei [8].

Based on data availability, described in the following Sections, the period of simulation was the year 2000. The methodology could be further applied to different time series.

 doi:10.20944/preprints202211.0502.v1

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

2.2. Input data acquisition and processing

The model requires presence data of the species (.csv) and other predictor layers in raster format (.tiff).

Citrus presence data was obtained by overlaying the Sicilian TRC (Technical Regional Cartography) with the IT2000 orthophotos available in the Sicilian Land Information System (SITR) by using GIS software, specifically ArcGIS® for Desktop 10.3 and QGIS 3.10.0. In detail, about 10000 presence points were used as input data in the *VISTRAILS:SAHM*.

By using the GIS tools, 19 bioclimatic variables provided by the *WorldClim* database for three decades, from 1970 to 2000, were represented in raster format. WorldClim data is routinely used for cropland suitability studies because they provide a comprehensive picture of monthly, quarterly, and annual bioclimatic conditions [9]. Furthermore, the set of covariates was enriched by the *Digital Terrain Model* (DTM) that brings valuable information about altitude for 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 (*'Agrumicultura Consapevole della Qualità e Uso dell'Acqua'* – *'Awareness of quality and use of water in Citrus cultivation'*), 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). 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, named Sir_Irr hereafter, suitable for the models input.

2.3. Modules selection

In the second phase of the methodology, the modules within *VISTRAILS: SAHM* were selected based on the scheme defined by Morisette et al. [10]. It establishes which modules are suitable for each phase of the model, in 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.

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 is 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 amount of presence data is critical and the training ratio of 70% with testing ratio equal to 30% produces a more robust model [6, 11, 12].

 doi:10.20944/preprints202211.0502.v1

Finally, the *CovariateCorrelationSelector* module provides a breakpoint in the modelling workflow to allow the user to evaluate how each variable explains the distribution of the sampled data points and allows you to remove any variables that may be highly correlated with others [13]; 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 ± 0.8 [4, 14, 15]. Therefore, since values of the coefficients 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 the percentage that the predictor provide to the probability for the specific model [16].

2.4. Accuracy measures

In this study, the measurement of the accuracy of the SDM classifications was conducted through the 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 [11, 12]. Moreover, ΔAUC values, computed as the difference between the AUC of the training and the AUC of the testing, were considered for assessing *overfitting*, according to [17]. In detail, when the ΔAUC value exceeds 0.05, *overfitting* occurs [18].

Moreover, *True Skills Stat* (TSS) [19] was considered in this study to compare the different models, by applying the following relations:

$$TSS = Sensitivity + Specificity - 1$$

in which *Sensitivity* (or *True Positive Rate* - TPR) and *Specificity* (or *True Negative Rate* - TNR) are defined as:

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

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

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 a random ranking. This has 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 [20]. TSS ranges from −1 to +1, where +1 indicates perfect agreement and zero or negative values indicate a performance no better than random [19]. To analyse TSS, according to some authors [21, 22], the difference ΔTSS between training e testing was computed.

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:

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

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. In detail, 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 (municipalities of Syracuse, Noto and Avola) while the central and southern areas were considered unsuitable by all models (Figure 2). In Figure 2, according to Akpoti [16] 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 output. 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 province level was carried out (Figure 2). Presence areas (in green colour) have a continuous aspect in GLM, MARS and MaxEnt models (figs. 2b, 2c, and 2d), whereas in RF and BRT models (figs. 2a and 2e) those areas are composed of polygons with holes. The highest differences among the models' output 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^2) 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, RF underestimated most compared to the average value of the model predictions, whereas MaxEnt overestimated. The differences among the simulations of the presence areas obtained by BRT, GLM, and MARS were not high (with a maximum of about 35 km^2), 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 3). The surfaces of each class were calculated for each model by using the QGIS software (Table 1). From the comparison of class 10 obtained by RF and BRT models, the localisation of the areas was similar (in the municipalities of Lentini, Augusta, Carlentini, Francofonte, Melilli, Sortino, Siracusa, Avola e Noto) and quite spread, though the surfaces are wider in RF than in BRT. In GLM model, class 10 is found in a lower number of areas (in the municipalities of Francofonte and Lentini) and even less for MARS model (in Carlentini municipality). The greatest surface areas for class 1 (1178.5 km^2) and class 10 (116.1 km^2) were confirmed for RF.

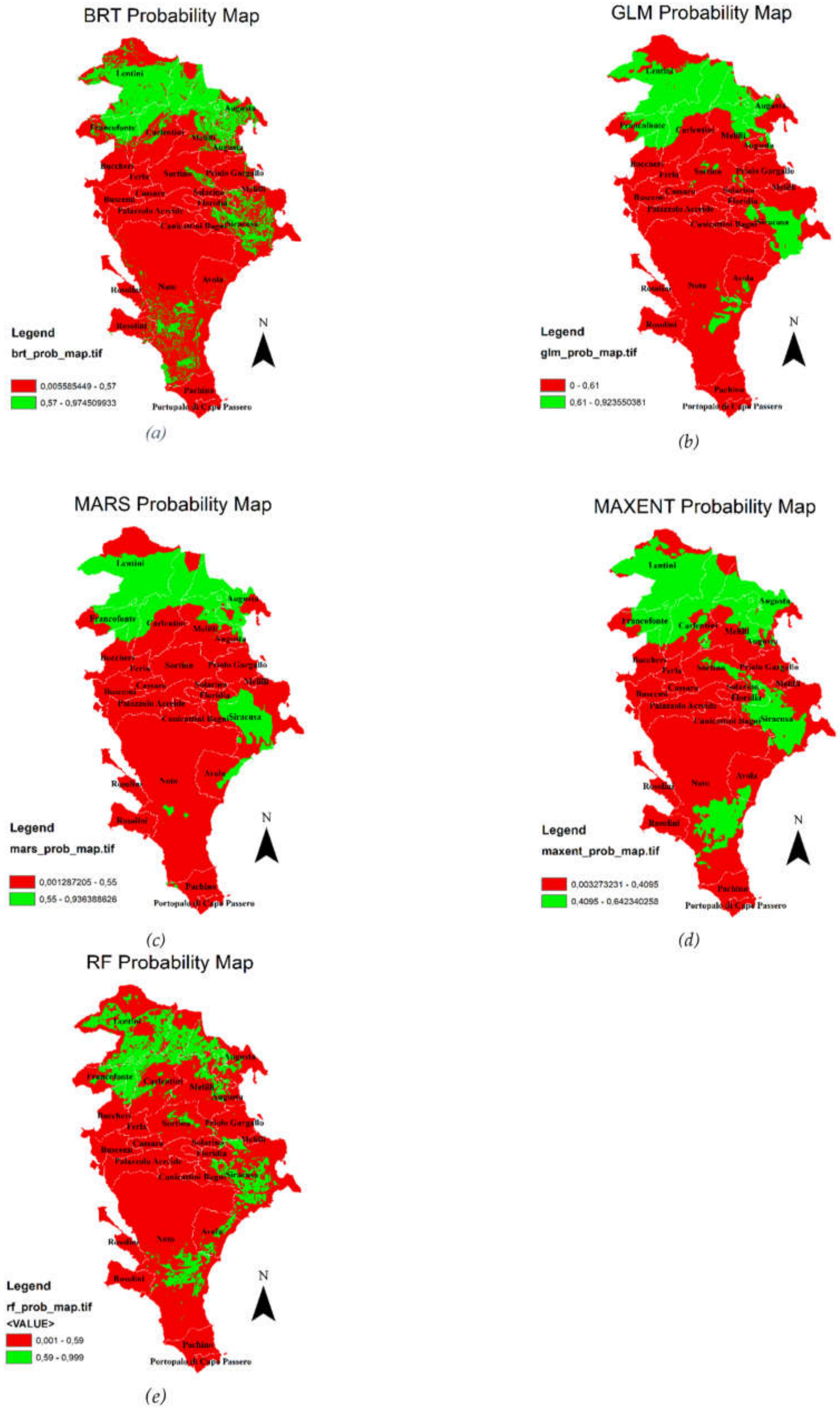


Figure 2. - Probability Maps, obtained by applying the threshold for probability, for each model in the province of Syracuse: a) BRT; b) GLM; c) MARS; d) MaxEnt; and e) RF.

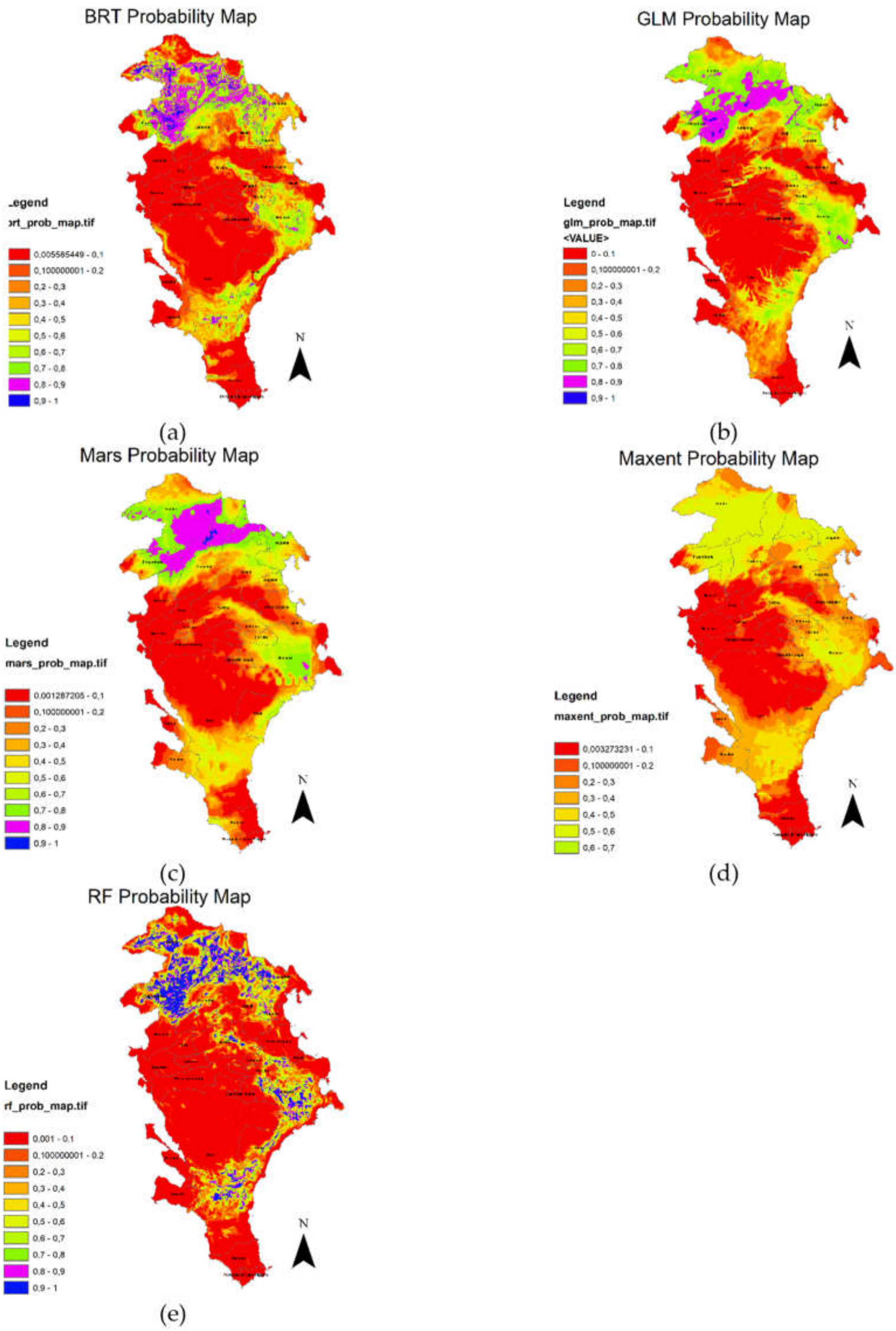


Figure 3. - Probability Maps, obtained by applying the 10-classes subdivision for probability, for each model in the province of Syracuse: a) BRT; b) GLM; c) MARS; d) MaxEnt; and e) RF.

Moreover, the sum of the surfaces from class 7 to class 10 for each model except for MaxEnt, that required summing the values from class 5 to 10, confirmed the outcomes obtained by the application of the thresholds, i.e., MaxEnt had the highest area (703.1 km^2) whereas RF had the lowest one (393.3 km^2). GLM was close to the mean value of the overall surface computed on all the models output (496.46 km^2), BRT and MARS slightly overestimated the overall surface area (444.8 and 442.7 km^2, respectively) whereas MaxEnt highly overestimated and RF moderately underestimated compared to the average value.

Table 1. Values of the surface areas (km²) 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	**315.5**	120.1	232.2
6	115.6	80.3	**385.4**	130.1	136.1
7	**122.50**	**82.3**	**2.2**	**176.6**	**122**
8	**147.90**	**88.1**	**0.0**	**190.4**	**152.8**
9	**153.40**	**106.8**	**0.0**	**127.1**	**164.3**
10	**21.00**	**116.1**	**0.0**	**4.3**	**3.6**

To assess which, among the models, had a higher capability to estimate the probability of presence of citrus, the metrics, produced by the SDMs 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 TSS was higher than 0.50.

Based on the results reported in Table 2, the BRT 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 ΔAUC 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 ΔAUC equal to 0 was found for RF model, while the values for MARS, GLM and MaxEnt showed a ΔAUC in the order of hundredths, thus also suitable because of the low *overfitting* associated.

Moreover, for ΔTSS, table 3 shows that BRT model had a high value (about 0.15) compared to those of the other SDMs, having values in the range 0 ÷ 0.005.

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

Table 2. - 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 3. – Values of ΔAUC e ΔTSS for the different SDMs.

	BRT	GLM	MARS	MAXENT	RF
ΔAUC	0.0819	0.002	-0.001	0.006	0
ΔTSS	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 presence though the localisation of the areas was similar (Figure 4). Moreover, the more the reduction of irrigation volume the more is the reduction of probability. The considerations deriving from the maps are confirmed by the numerical data processed (Tables 4, 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^2 (at a deficit irrigation equal to 50% of the actual) and a minimum one of 75.99 km^2 (at a deficit irrigation equal to 90% of the actual value) with an average value of 122.32 km^2 (Table 4).

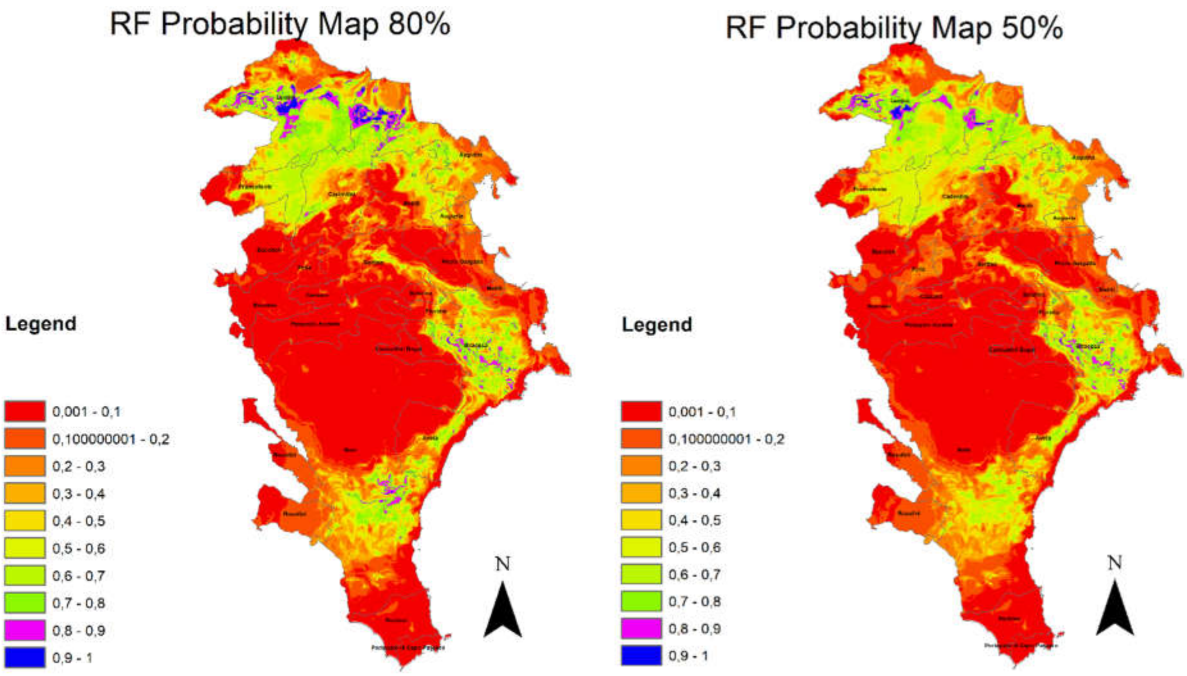
The reduction of irrigation would highly affect the probability of presence, especially for 9 and 10 classes. For instance, the surface of class 10, for a 90% irrigation, would reduce by 85.36%, compared to 100% irrigation, and from 5.51% to 0.81% of the whole surface of the province (Tables 5 and 6); whereas, for a deficit irrigation equal to 80% of the actual value, the surface area for class 10 is 9.20 km^2 (-92.07%) and 37.56 km^2 (-64.82%) for class 9 (Tab. 4 and 5). However, an increase of the surface of intermediate classes, mainly for 6 and 7 would occur. For instance, the surface of class 7 would increase by 65.29 km^2 (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 the lower one (albeit always considering the classes 7 to 10, that have a higher probability than about 0.6, according to the thresholds of presence/absence), however the surface loss would range from 19.3% (at 90% deficit irrigation) to 44.09% (at 50% deficit irrigation).

In table 5, 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 at increasing of irrigation reduction in the territory. This is confirmed in the probability maps where, as the irrigation contribution decreases, there is progressively a reduction in the extent of the presence of the species in the areas until it 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 to allow 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.

 doi:10.20944/preprints202211.0502.v1



(a) *(b)*

Figure 4. 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 4. Values of the surface areas S_i (km^2) 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^2] 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 5. 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.

Dsup [%]

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 November 2022 doi:10.20944/preprints202211.0502.v1

Table 6. 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.

	S_i / S_{tot} [%]									
Sir_Irr [%]	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 RF model (Figure 5) provided useful information about the range of the predictors most contributing to the probability values.

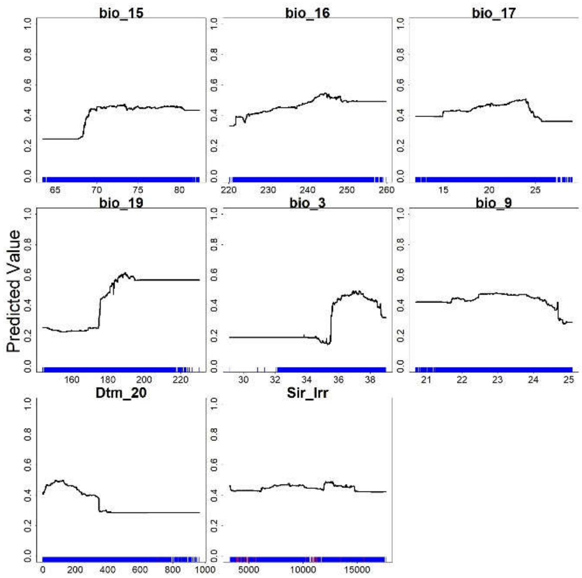


Figure 5. *ResponseCurves* of the predictors for the RF model, for the case study analysed.

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 presence of citrus, in the case study analysed.

The analysis of the *ResponseCurve* graphs for DTM_20 in the province of Syracuse showed that the altitudes suitable for citrus production were correctly identified by RF model, i.e., points with a probability greater than the threshold were found for altitudes lower than 400 m a.s.l. [23, 24]. The Response curves of predictors showed the following ranges where the predicted value for citrus probability of 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 precipitations during the coldest quarter of the year; Bio_15 greater than 70%, showing the variation of monthly precipitations 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 precipitations of the driest quarter. The contribute 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 influencing bioclimatic variables are related to precipitations.

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, as well as 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 Species distribution modelling (SDMs) to predict citrus distribution in the territory in order to derive information on SDMs application in 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 application to the case study of the Province of Syracuse, Italy.

This study constitutes 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 as major predictors suitable for this knowledge. General uniformities in the models' predictions suggest that the multi-model approach contributes to increase consistency of the outcomes. Modeling showed that BRT and RF produced higher evaluation metrics compared to the other models, however BRT suffered from *overfitting*. GIS contributed to the analysis of the outcomes by showing and quantifying the spatial distribution of citrus presence as well as allowing the comparison among the simulations of different levels of irrigation.

Investigation on model parameters would be object of further studies aimed at fine-tuning the performance of the predictions, and research is in progress to investigate on input data by involving computation of the bioclimatic data from local weather stations. Further analysis could regard predictions of future probability of 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, "Miglioramento delle produzioni agroalimentari mediterranee in condizioni di carenza di risorse idriche – WATER4AGRIFOOD", 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.

Acknowledgments: The authors wish to thank the Sicilian Region for SITR data (https://www.sitr.regione.sicilia.it/).

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

References

1. Del Bravo F.; Finizia A.; Lo Moriello M. S.; Ronga M. *La competitività della filiera agrumicola in Italia*. Rete Rurale Nazionale 2014-2020, 2020.
2. 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 at: https://openknowledge.worldbank.org/bitstream/handle/10986/30603/130404-WP-P159851-Morocco-WEB.pdf; accessed on 23-11-2022)
3. 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*. Contemporary Problems of Ecology, 2018, 11.6: 604-613. https://doi.org/10.1134/S1995425518060069.
4. 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. Bioproducts and Biorefining. 2022. 16.1: 54-67. https://doi.org/10.1002/bbb.2304.

5. 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 Analysis, 39(5), 1169-1191. 2019. https://doi.org/10.1111/risa.13230.
6. 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*. Ecological Informatics. 2015.30, 40-48. https://doi.org/10.1016/j.ecoinf.2015.08.010.
7. Diniz-Filho, JAF; Mauricio Bini, L; Fernando Rangel, T; Loyola, R.D., Hof, C., Nogués-Bravo, D. and Araújo, M.B. (2009), Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. Ecography, 32: 897-906. 2009. https://doi.org/10.1111/j.1600-0587.2009.06196.
8. 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 Atti del Terzo Congresso Nazionale di Selvicoltura (pp. 234-239). 2008. DOI: 10.4129/CNS2008.030.
9. Fitzgibbon, A.; Pisut, D.; Fleisher, D.; *Evaluation of Maximum Entropy (Maxent) Machine Learning Model to Assess Relationships between Climate and Corn Suitability*. Land, 11(9), 1382. 2022. https://doi.org/10.3390/land11091382.
10. 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, 36: 129-135. 2013. https://doi.org/10.1111/j.1600-0587.2012.07815.x
11. West A. M.; Evangelista P. H.; Jarnevich C. S; et al.; Integrating Remote Sensing with Species Distribution Models; Mapping Tamarisk Invasions Using the Software for Assisted Habitat Modeling (SAHM), Journal of Visualized Experiments, 116, 2016a. https://doi.org/10.3791/54578
12. 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. 2016b. https://doi.org/10.1016/j.ecoinf.2016.11.001.
13. Talbert C.; Talbert M.; *User Documentation for the Software for Assisted Habitat Modeling (SAHM) package in VisTrails*. USGS (U.S. Geological Survey). 2001.
14. 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. Ecological engineering, 51, 83-87. 2013. https://doi.org/10.1016/j.ecoleng.2012.12.004.
15. 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*. Ecological Engineering, 92, 260-269. 2016. https://doi.org/10.1016/j.ecoleng.2016.04.010.
16. 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*. Science of the total environment, 2020, 709: 136165. https://doi.org/10.1016/j.scitotenv.2019.136165.
17. Mukherjee T.; Vandana S.; Lalit K.S.; *et al. Landscape-level habitat management plan through geometric reserve design for critically endangered Hangul (Cervus hanglu hanglu)*. Science of The Total Environment. Volume 777. 2021,146031, https://doi.org/10.1016/j.scitotenv.2021.146031.
18. 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 Analysis, 39(5), 1169-1191. 2019. https://doi.org/10.1111/risa.13230.
19. Allouche, O; Tsoar, A and Kadmon, R; *Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS)*. Journal of Applied Ecology. 2006. 43: 1223-1232, https://doi.org/10.1111/j.1365-2664.2006.01214.x.
20. Phillips, S. J.; Dudík, M.; *Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation*. Ecography, 31(2), 161-175. 2008. https://doi.org/10.1111/j.0906-7590.2008.5203.x.
21. 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. Forest Ecology and Management, 518, 120249. 2022. https://doi.org/10.1016/j.foreco.2022.120249.
22. Brun, P.; Kiørboe, T.; Licandro, P. Payne, M.R.; *The predictive skill of species distribution models for plankton in a changing climate*. Glob Change Biol, 22: 3170-3181. 2016.https://doi.org/10.1111/gcb.13274.
23. Pignatti S., 2017-2019. "Flora d'Italia", 2017-2019 Citrus limon (L.) Burm. fil.Vol. 2 pg. 1090 key 4-682
24. Pignatti, S; *"La Flora d'Italia"*. Edagricole. 1982. Vol. 2 pg. 54.