

**JOSIP JURAJ STROSSMAYER UNIVERSITY OF OSIJEK  
RUĐER BOŠKOVIĆ INSTITUTE ZAGREB**

**University postgraduate interdisciplinary doctoral study  
Environmental protection and nature conservation**

**Dorijan Radočaj, M. Sc.**

**AGROECOLOGICAL AND PEDOLOGICAL MODELING OF CROPLAND  
SUITABILITY FOR SOYBEAN CULTIVATION BY INTEGRATING SATELLITE  
REMOTE SENSING DATA AND MACHINE LEARNING**

**Ph. D. Thesis**

**OSIJEK, 2022.**

## **BASIC DOCUMENTATION CARD**

---

**Josip Juraj Strossmayer University of Osijek  
Ruđer Bošković Institute, Zagreb  
Postgraduate Interdisciplinary University Doctoral Study of  
Environmental Protection and Nature Conservation**

**PhD thesis**

**Scientific Area: Biotechnical Sciences  
Scientific Field: Agronomy**

### **AGROECOLOGICAL AND PEDOLOGICAL MODELING OF CROPLAND SUITABILITY FOR SOYBEAN CULTIVATION BY INTEGRATING SATELLITE REMOTE SENSING DATA AND MACHINE LEARNING**

**Dorijan Radočaj, M. Sc.**

**Thesis performed at the Faculty of Agrobiotechnical Sciences Osijek, Josip Juraj Strossmayer University of Osijek**

**Supervisors: Prof. Dr. Mladen Jurišić, Prof. Dr. Oleg Antonić**

#### **Summary**

Increased global food demand pressured farmers into producing higher crop yields to keep track of the population growth and increased life standard. The standard method of calculating cropland suitability in previous studies was GIS-based multicriteria analysis, most often combined with the Analytic Hierarchy Process (AHP). While this is a flexible and widely accepted procedure, it has major fundamental disadvantages in a cropland suitability determination, including lack of accuracy assessment, high human subjectivity and computational inefficiency in case of complex input data. To improve these disadvantages, a reliable, objective and computationally efficient procedure for determining the cropland suitability, consisting of: 1) a computationally efficient validation method of cropland suitability using global satellite missions in high (Sentinel-2) and medium spatial resolution (PROBA-V); 2) an automatization method of spatial modeling of abiotic criteria for the example of soil texture according to a globally accepted standard; 3) suitability prediction method based on machine learning algorithms and globally available spatial data, which allows high reliability of prediction with reduced user subjectivity compared to GIS-based multicriteria analysis. The proposed methods add to an important paradigm shift in a cropland suitability determination, which focuses on the objective, automated and computationally efficient prediction method, as well as ensuring reliable validation data from open data sources on a global scale.

**Number of pages: 86**

**Number of figures: 29**

**Number of tables: 24**

**Number of references: 283**

**Original in: English**

**Keywords: geographic information system, vegetation index, biophysical variables, automation, Sentinel-2**

**Date of the thesis defense: 21.03.2022.**

#### **Reviewers:**

1. Ante Šiljeg, PhD, Assistant professor, Department of Geography, University of Zadar
2. Neven Cukrov, PhD, Senior research associate, Ruđer Bošković Institute, Zagreb
3. Irena Rapčan, PhD, Full professor, Faculty of Agrobiotechnical Sciences Osijek, Josip Juraj Strossmayer University of Osijek
4. Branimir Hackenberger Kutuzović, PhD, Full professor, Department of Biology, Josip Juraj Strossmayer University of Osijek (substitute)

**Thesis deposited in:** National and University Library in Zagreb, Ul. Hrvatske bratske zajednice 4, Zagreb; City and University Library of Osijek, Europska avenija 24, Osijek; Josip Juraj Strossmayer University of Osijek, Trg sv. Trojstva 3, Osijek

## TEMELJNA DOKUMENTACIJSKA KARTICA

---

Sveučilište Josipa Jurja Strossmayera u Osijeku  
Institut Ruđer Bošković, Zagreb  
Poslijediplomski interdisciplinarni sveučilišni studij  
Zaštita prirode i okoliša

Doktorska disertacija

Znanstveno područje: Biotehničke znanosti  
Znanstveno polje: Poljoprivreda

### AGROEKOLOŠKO I PEDOLOŠKO MODELIRANJE POGODNOSTI POLJOPRIVREDNOG ZEMLJIŠTA ZA UZGOJ SOJE INTEGRACIJOM PODATAKA DOBIVENIH IZ SATELITSKIH DALJINSKIH ISTRAŽIVANJA I STROJNOG UČENJA

Dorijan Radočaj, mag. ing. geod. et geoinf.

Doktorska disertacija je izrađena u: Fakultet agrobiotehničkih znanosti Osijek, Sveučilište Josipa Jurja Strossmayera u Osijeku

Mentor/komentor: prof. dr. sc. Mladen Jurišić (mentor), prof. dr. sc. Oleg Antonić (komentor)

#### Sažetak doktorske disertacije:

Povećana globalna potražnja za hranom uzrokovala je potražnju za većim prinosima usjeva uslijed porasta broja stanovnika i povećanog životnog standarda. Standardna metoda izračuna pogodnosti poljoprivrednog zemljišta u prethodnim studijama je GIS multikriterijska analiza temeljena na GIS-u, najčešće u kombinaciji s analitičkim hijerarhijskim procesom (AHP). Iako je ovo fleksibilan i široko prihvaćen postupak, sadrži temeljne nedostatke u određivanju pogodnosti poljoprivrednog zemljišta, uključujući nedostatak validacije rezultata, visoku ljudsku subjektivnost i računsku neučinkovitost u slučaju složenih ulaznih podataka. Kako bi se poboljšali ovi nedostaci, razvijen je pouzdan, objektivan i računski učinkovit postupak za određivanje pogodnosti poljoprivrednog zemljišta, koji se sastoji od: 1) računalno učinkovite metode validacije pogodnosti korištenjem globalnih satelitskih misija visoke (Sentinel-2) i srednje prostorne rezolucije (PROBA-V); 2) automatizirane metode prostornog modeliranja abiotičkih kriterija na primjeru teksture tla prema globalno prihvaćenom standardu; 3) metode predviđanja pogodnosti koja se temelji na algoritmima strojnog učenja i globalno dostupnim prostornim podacima, što omogućuje visoku pouzdanost predviđanja uz smanjenu subjektivnost korisnika u usporedbi s GIS multikriterijskom analizom. Predložene metode doprinose važnoj promjeni paradigme pri određivanju pogodnosti poljoprivrednog zemljišta, koja se temelji na objektivnoj, automatiziranoj i računski učinkovitoj metodi, kao i na dostupnost pouzdanih validacijskih podataka iz otvorenih izvora podataka na globalnoj razini.

Broj stranica: 86

Broj slika: 29

Broj tablica: 24

Broj literaturnih navoda: 283

Jezik izvornika: engleski

Ključne riječi: GIS, vegetacijski indeks, biofizičke varijable, automatizacija, Sentinel-2

Datum obrane: 21.03.2022.

#### Povjerenstvo za obranu:

1. izv. prof. dr. sc. Ante Šiljeg, izvanredni profesor, Odjel za geografiju Sveučilišta u Zadru
2. dr. sc. Neven Cukrov, viši znanstveni suradnik, Institut Ruđer Bošković, Zagreb
3. prof. dr. sc. Irena Rapčan, redovita profesorica, Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet agrobiotehničkih znanosti Osijek
4. prof. dr. sc. Branimir Hackenberger Kutuzović, redoviti profesor, Sveučilište Josipa Jurja Strossmayera u Osijeku, Odjel za biologiju (zamjena)

Rad je pohranjen u: Nacionalnoj i sveučilišnoj knjižnici Zagreb, Ul. Hrvatske bratske zajednice 4, Zagreb; Gradskoj i sveučilišnoj knjižnici Osijek, Europska avenija 24, Osijek; Sveučilištu Josipa Jurja Strossmayera u Osijeku, Trg sv. Trojstva 3, Osijek

The thesis topic titled “Agroecological and pedological modeling of cropland suitability for soybean cultivation by integrating satellite remote sensing data and machine learning” is accepted on the session of the University council for postgraduate interdisciplinary (doctoral) studies on 9<sup>th</sup> December 2021.



# ACKNOWLEDGEMENTS

---



I am deeply grateful to my mentor, Prof. Dr. Mladen Jurišić, for his positive impact during the entire process of my Ph.D. study. Thank you for your constant guidance in my path of becoming a better individual. I would also like to thank my co-mentor, Prof. Dr. Oleg Antičić, for being dependable, straightforward and a man of his word.

My sincere thanks to the reviewers of this dissertation, Assoc. Prof. Ante Šiljeg, Neven Cukrov, PhD and Prof. Dr. Irena Rapčan, who invested their precious time and effort to evaluate this dissertation in a highly constructive manner. They have been professional and helpful during the entire reviewing process.

Big thanks for Assist. Prof. Mateo Gašparović, who had a highly influential role in my scientific development, even before starting my Ph.D. study. Thank you for always being ready to give your positive impact. My sincere gratitude also goes to Assoc. Prof. Ivan Plaščak, for his selfless help in many of my professional and non-professional activities.

I greatly appreciate the support of the dean of the Faculty of Agrobiotechnical Sciences Osijek, Prof. Dr. Krunoslav Zmaić, as well as its Scientific fund, for the continuous support of the Chair for Geoinformation Technology and GIS, as well as my scientific work in this dissertation. Many thanks to the Head of doctoral study, Prof. Dr. Enrih Merdić, and all professors which I encountered during the study. I will remember this study as a pleasant and resourceful experience.

Finally, I am deeply thankful to my family for continuous and unconditional support in my life, as well as to anyone with a positive intention who gave a contribution to this dissertation.

# CONTENTS

---

◆

<b>GENERAL INTRODUCTION</b> .....	1
<b>CHAPTER 1: Optimal Soybean (<i>Glycine max</i> L.) Land Suitability Using GIS-based Multicriteria Analysis and Sentinel-2 Multitemporal Images</b> .....	3
<b>CHAPTER 2: Delineation of Soil Texture Suitability Zones for Soybean Cultivation: A Case Study in Continental Croatia</b> .....	28
<b>CHAPTER 3: Cropland Suitability Assessment Using Satellite-Based Biophysical Vegetation Properties and Machine Learning</b> .....	51
<b>GENERAL DISCUSSION</b> .....	73
<b>CONCLUSIONS</b> .....	79
<b>REFERENCES</b> .....	81
<b>CURRICULUM VITAE</b> .....	83

# GENERAL INTRODUCTION

---

The accelerated population growth globally necessitates the production of increasing amounts of food (Taghizadeh-Mehrjardi et al., 2020). Short-term meeting of the food demand is achieved by conventional intensive agricultural production, but often at the cost of long-term sustainability and soil degradation (Mandal et al., 2020). An additional challenges climate change and environmental contamination, undermining the effectiveness of conventional cropland management (Chemura et al., 2020). The two most commonly used approaches within conventional agricultural production systems to increase crop yield are: 1) formation of new agricultural areas by transforming the land cover; and 2) yield increase on existing cropland by adapting agrotechnical operations, including increased use of fertilizers and pesticides.

Land use transformation has a greater potential to increase overall yield compared to improving agricultural practices on existing cropland (Layomi Jayasinghe et al., 2019; Song and Zhang, 2021). The most common transformation of new agricultural land is from forest and wetland areas, which consequently affects the destruction of natural habitats and poses a threat to biodiversity. Habitat destruction is the most common cause of flora endangerment in the Republic of Croatia, according to the Red Book of Vascular Flora (Nikolić and Topić, 2005). The application of fertilizers and pesticides is necessary for the continuous production of high and stable yields in conventional intensive agricultural production (Møller et al., 2021). Their quantity further increases in case of non-compliance with crop rotation, inadequate agrotechnical operations and cultivation of certain agricultural crops in a naturally unsuitable location (Song and Zhang, 2021). This approach results in contaminating the environment by imputing heavy metals (primarily copper) and persistent organic pollutants, per the Ordinance on the protection of cropland from contamination (Official Gazette, 71/2019). Environmental contamination thus has a direct impact on human health, flora and fauna by spreading through surface water and groundwater and by accumulation in living organisms.

In response to the need for the long-term sustainability of increasing agricultural production, alternative approaches of increasing crop yields have been developed in recent decades. Among them, suitability studies and precision agriculture, based on the integration of geoinformation technologies within the geographic information system (GIS) stand out. By cultivating crops in naturally suitable areas, it is possible to achieve equal or higher yields while reducing the need for fertilizers and pesticides on existing cropland with a conventional approach (Dedeoğlu and Dengiz, 2019). Given that relevant abiotic factors are difficult or impossible to influence, it is necessary to adjust the agricultural management plans and crop rotation in order to create conditions for sustainable and environmentally-friendly agriculture (Song and Zhang, 2021). Due to the temporal variability of abiotic factors, primarily caused by climate change, regular monitoring and analysis of suitability level changes are necessary (Chemura et al., 2020). The inventory of the current suitability levels has a direct practical significance for agriculture, serving as a basis for issuing recommendations to adjust agrotechnical operations and establish irrigation systems in vulnerable areas (Taghizadeh-Mehrjardi et al., 2020). To this end, methods of spatial analysis in the GIS environment have

been developed, of which GIS-based multicriteria analysis in combination with analytic hierarchy process (AHP) is currently leading as a standard for determining cropland suitability. While it is recognized as a standard for suitability determination, this approach is susceptible to subjectivity, complexity and inability of integrating big data (Progênio et al., 2020), which negatively impacts the reliability and repeatability of such methods.

The application of machine learning classification methods has successfully resolved similar shortcomings in the prediction of biological, chemical and physical soil properties, upgrading the conventional approach. Hengl et al. (2017) integrated soil sampling data and freely available spatial data representing abiotic factors into a modern machine-based prediction system in the SoilGrids project. This represents the beginning of a paradigm shift in the prediction of spatial variables in the environment, which allows for the exceptional potential for application in disciplines outside of pedology. However, this potential was only recognized last year by the scientific community for the purpose of determining the cropland suitability and is at a very early stage of development. Of the few studies conducted, Taghizadeh-Mehrjardi et al. (2020) compared the suitability prediction accuracy of wheat and barley cultivation by comparing traditional parametric methods and machine learning. The overall precision of the machine learning method resulted in 29% higher overall accuracy for wheat and 26% higher for barley compared to the traditional approach. This approach also quantified the relative importance of individual input abiotic factors, which represents an objective and inverse approach to the weighting process in the AHP method. Møller et al. (2021) recognized a similar potential of machine learning in the suitability prediction, stating the possibility of accurately determining the individual components of suitability, emphasizing the socio-economic and ecological components.

While the exceptional potential of machine learning and remote sensing data in suitability determination was noted, there are still no comprehensive and straightforward solutions. The aim of this dissertation is to upgrade these methods regarding the three currently most significant shortcomings of the conventional GIS-based multicriteria analysis. Cropland suitability determination methods were developed for any major crop type, while soybean was selected as a representative crop type in the dissertation due to its increasing importance in global food production. Soybean was cultivated at 838.5 km<sup>2</sup> in 2020, covering about 13.7% of the country's arable cropland. Its importance in human and livestock nutrition is steadily increasing, as one of the main sources of edible oils and proteins. Soybean cultivation enhances soil fertility and enriches it with nutrients, especially nitrogen, which makes it a very useful and flexible crop in the formation of crop rotation. Three research hypotheses have been defined:

1. Suitability results can be validated using a computationally efficient method using globally available satellite multispectral images.
2. Automation of spatial GIS modeling of abiotic suitability factors for the example of soil texture is feasible according to a globally accepted classification standard and regardless of the range of input values.
3. Suitability determination using machine learning can be performed with high reliability by reducing subjectivity in the selection of abiotic factors and their weighting in relation to the conventional GIS-based multicriteria analysis.

# CHAPTER 1

---

Optimal Soybean (*Glycine max* L.) Land Suitability Using GIS-based  
Multicriteria Analysis and Sentinel-2 Multitemporal Images

Article

# Optimal Soybean (*Glycine max* L.) Land Suitability Using GIS-Based Multicriteria Analysis and Sentinel-2 Multitemporal Images

Dorijan Radočaj <sup>1,\*</sup> , Mladen Jurišić <sup>1</sup>, Mateo Gašparović <sup>2</sup>  and Ivan Plaščak <sup>1</sup> 

<sup>1</sup> Faculty of Agrobiotechnical Sciences Osijek, Josip Juraj Strossmayer University of Osijek, Vladimira Preloga 1, 31000 Osijek, Croatia; mjurisic@fazos.hr (M.J.); iplascak@fazos.hr (I.P.)

<sup>2</sup> Faculty of Geodesy, University of Zagreb, Kačićeva 26, 10000 Zagreb, Croatia; mgasparovic@geof.unizg.hr

\* Correspondence: dradocaj@fazos.hr; Tel.: +385-31-554-879

Received: 24 March 2020; Accepted: 4 May 2020; Published: 5 May 2020



**Abstract:** Soybean is regarded as one of the most produced crops in the world, presenting a source of high-quality protein for human and animal diets. The general objective of the study was to determine the optimal soybean land suitability and conduct its mapping based on the multicriteria analysis. The multicriteria analysis was based on Geographic Information System (GIS) and Analytic Hierarchy Process (AHP) integration, using Sentinel-2 multitemporal images for suitability validation. The study area covered Osijek-Baranja County, a 4155 km<sup>2</sup> area located in eastern Croatia. Three criteria standardization methods (fuzzy, stepwise and linear) were evaluated for soybean land suitability calculation. The delineation of soybean land suitability classes was performed by k-means unsupervised classification. An independent accuracy assessment of calculated suitability values was performed by a novel approach with peak Normalized Difference Vegetation Index (NDVI) values, derived from four Sentinel-2 multispectral satellite images. Fuzzy standardization with the combination of soil and climate criteria produced the most accurate suitability values, having the top coefficient of determination of 0.8438. A total of 14.5% of the study area (602 km<sup>2</sup>) was determined as the most suitable class for soybean cultivation based on k-means classification results, while 64.3% resulted in some degree of suitability.

**Keywords:** AHP; standardization; NDVI; fuzzy algorithms; k-means; FAO suitability classification

## 1. Introduction

The increasing demand for food and bioenergy in the world stimulated the development of land suitability calculation methods, as a basis for effective agricultural land management and environmental sustainability [1,2]. Conventional agriculture is characterized by a very high input of fertilizer, pesticides and herbicides, having a negative long-term impact on sustainability [3]. The selection of naturally suitable areas for the particular crop type cultivation reduces the overall application of inputs, creating an optimal environment for crop growth [4]. Agricultural land management plans based on inappropriate evaluation of natural resources limit crop yields and increase production costs [5]. Chen et al. [6] recommended the development of crop-specific evaluation indices through land suitability determination to ensure the sustainability of agricultural production. Delineation of suitability classes with homogenous characteristics is a fundamental segment of land evaluation, allowing effective implementation of land use planning in the field [7]. Mapping of such suitability classes for particular crop cultivation is essential for the transfer of knowledge to the end-users, whether to land management experts or individual farmers and farming companies [8].

Multicriteria analysis is widely recognized as the method for the selection of the most suitable (optimal) location and its alternatives in various areas, such as agriculture [9–11], forestry [12], land

management [13] and environmental planning [14]. With the integration of spatial components from Geographic Information System (GIS), an approach of GIS-based multicriteria analysis enables a suitability modeling for any entity related to space [15]. The benefit of GIS-based multicriteria analysis in agriculture is its universal applicability regarding crop types, area size and location in the world. Its potential is conditioned only by the crop expert's knowledge of the optimal agroecological conditions of the selected crop type and the quality of the input spatial data [16,17]. The core procedures of multicriteria analysis are based on the establishment of the relationship between relevant criteria [18]. Many methods have been developed to determine the relative relationship between criteria for the suitability calculation, one of the most notable being the Analytic Hierarchy Process (AHP) [19]. Standardization of criteria values is a necessary procedure in the calculation of suitability models [20]. Input values of all criteria are commonly transformed into a unique number interval during the standardization process for further processing. Many authors noted an impact of the selection of standardization methods in multicriteria analysis on land suitability values [21–24]. Standardization using simple linear scale transformation was usually the selected method in these studies. The hybridization of GIS-based multicriteria analysis with unsupervised classification presents a novel approach in the management of calculated suitability values, enabling the effective creation of land suitability classes [25]. The delineation of suitability classes is regarded as a necessary procedure in suitability analyses and precision agriculture, as it significantly facilitates the application of GIS-based multicriteria analysis results in the field [26]. Van Niekerk noted the superiority of computer algorithms over traditional manual mapping techniques for the delineation of suitability classes, as they allow objective and time-efficient classification [27].

Soybean is a fundamental component of agricultural land management plans worldwide, presenting a major source of protein for humans and a high-quality animal feed [28]. According to the Food and Agriculture Organization of the United Nations (FAO) publication [29], soybean accounted for about one-third of the total harvested area devoted to annual and perennial crops, while its share in global oil-crop output was 44%. According to the projection by the European Commission for the period between 2019 and 2030, the production of soybean food products will continue to grow due to demand for locally produced plant-protein food [30]. The same source stated that the soybean area would show significant land-use change, resulting in a 5% harvested area increase by 2030. The long-term projection from the United States Department of Agriculture to 2029 predicts continuous global demand for soybean oil for biodiesel production [31]. The competitive position of soybean among arable crops has steadily improved due to consistent improvements in yield and reductions in production costs [29]. However, additional investment for soybean yield improving research was urged to land policymakers and managers [32].

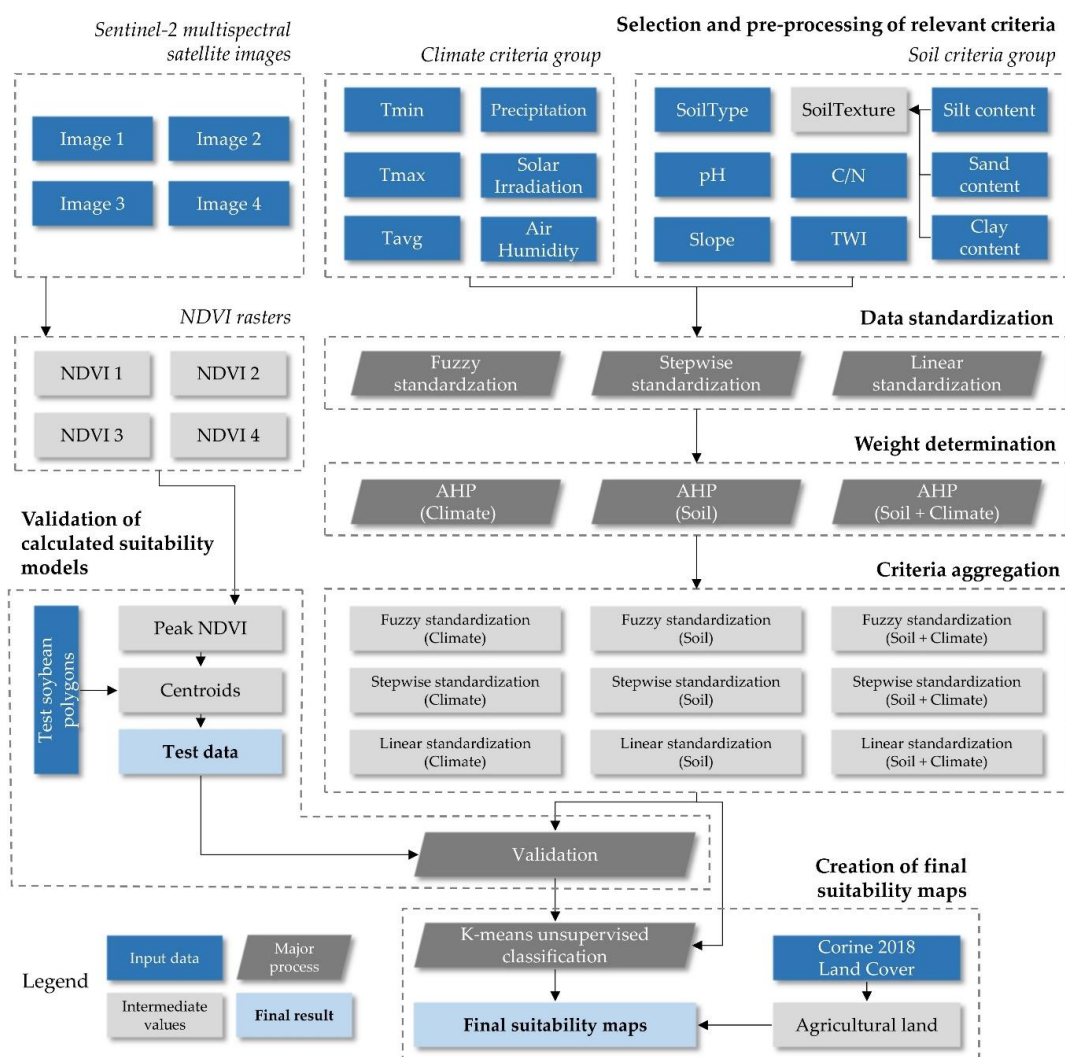
Medium-resolution multispectral satellite imagery presents an important data source for the observation of crop characteristics in yield improving research and mapping for agricultural land management [33,34]. Sentinel-2 is a multispectral satellite mission from the European Space Agency Copernicus program, started in 2015 [35]. The constellation of Sentinel-2 mission is based on two satellites, Sentinel-2A (S2A) and Sentinel-2B (S2B), orbiting 180° apart [36]. Sentinel-2 provides the possibility of an effective crop monitoring, having a 290 km swath width, high imaging resolution (up to 10 m) and revisit time of two to three days at mid-latitudes [36]. Crop parameters with the application of remote sensing are commonly monitored using vegetation indices [37]. The most used vegetation index is the Normalized Difference Vegetation Index (NDVI), which enables the determination of the relationship between photosynthetic and optical properties of crops [38]. NDVI is the most widely used indicator in studies regarding crop biomass and chlorophyll content [39]. Satellite-derived vegetation indices have been successfully used as a part of various crop models, such as the estimation of maximum evapotranspiration and irrigation requirements [40] and multitemporal crop monitoring [38].

The general objective of the study was the determination of optimal soybean land suitability and its mapping based on GIS-based multicriteria analysis. For this purpose, the evaluation of the three

most commonly used standardization methods in recent GIS-based multicriteria analysis studies was conducted. The validation of calculated suitability values was performed using a novel peak NDVI method derived from Sentinel-2 multitemporal images as an independent accuracy assessment.

## 2. Materials and Methods

The proposed method of GIS-based multicriteria analysis for soybean suitability calculation and its validation consists of six major steps (Figure 1). These are the selection and preprocessing of relevant criteria, data standardization, weight determination, criteria aggregation, validation of calculated suitability models and creation of final suitability maps. Open-source GIS software was used in this research: SAGA GIS v7.4.0 (Hamburg, Germany) [41] for data preprocessing and calculations, QGIS v3.8.3 (Grüt, Switzerland) [42] for data visualization and GRASS GIS v7.8.2 (Bonn, Germany) [43] for the calculation of solar irradiation.



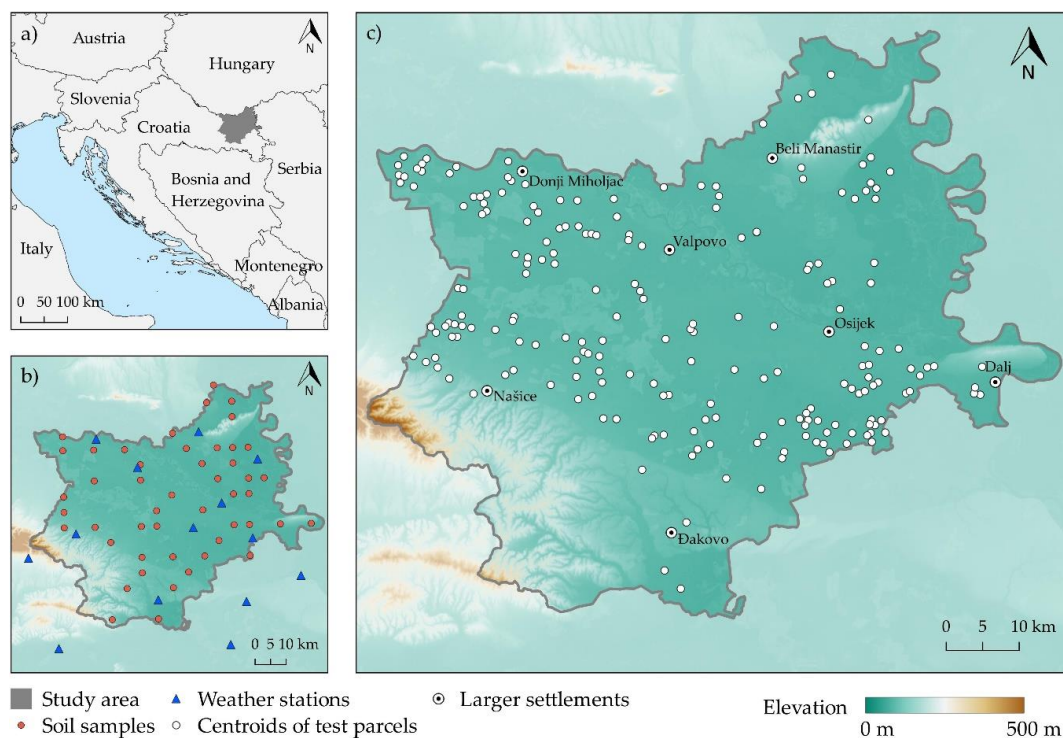
**Figure 1.** The workflow of soybean land suitability calculation and evaluation.

### 2.1. Study Area

The study area covers Osijek-Baranja County, a 4155 km<sup>2</sup> area located in eastern Croatia (Figure 2). According to the Croatian Bureau of Statistics [44], Osijek-Baranja County has the largest utilized agricultural area in Croatia, covering 13.6% of total country agricultural land in 2016. Corine 2018 Land Cover data showed that agricultural areas are the dominant land cover class in the county, covering 64.5% of the county's area. Forests are the second-largest class covering 26.7% of the county's area,



followed by artificial surfaces (3.8%), wetlands (2.7%) and water bodies (2.3%). Croatia follows a world trend of an increase in soybean cultivation. According to [45], a total of 111,316 t was produced in 2013 and 244,075 t in 2016, a 119% soybean production increase during four years. Soybean is the fourth highest cultivated crop in the study area behind maize, wheat and sunflower, covering 8.54% of the total agricultural land in the county during 2019 per Croatian Paying Agency for Agriculture, Fisheries and Rural Development data. Osijek-Baranja County has a moderately warm, rainy climate on a Köppen scale, according to the same source. Early soybean variants were commonly sown in eastern Croatia in recent years [46]. Sowing of these variants was conducted in late April, while harvest typically occurred mid-October. The Croatian agricultural and forestry advisory service recommends the application of 30 kg/ha of nitrogen (N), 60 kg/ha of phosphorus pentoxide ( $P_2O_5$ ) and 90 kg/ha of potassium oxide ( $K_2O$ ) during basic fertilization in autumn and minor adjustments before sowing [47]. Irrigation of soybean fields has been extremely overlooked in the study area, with only 64 ha employing some sort of irrigation system, according to Croatian Paying Agency for Agriculture, Fisheries and Rural Development data for 2019.



**Figure 2.** Study area: (a) location of Osijek-Baranja County, (b) location of used weather stations and soil samples, (c) location of centroids of test soybean parcels.

## 2.2. Selection and Preprocessing of Relevant Criteria

Climate and soil criteria were considered to have the most significant impact on soybean cultivation, based on previous research regarding soybean land suitability [48–52]. Relevant climate and soil criteria selected in this study are shown in Table 1. The selected number of criteria in each criteria group is six, which meets the specifications for the application of the AHP method for the consistency of information between criteria [53]. The main data sources for the modeling of selected criteria were: Croatian Meteorological and Hydrological Service (DHMZ) data during the soybean growth period between April and October for years 2015–2019; Croatian Agency for the Environment and Nature (CAEN) soil samples collected during 2016; a basic pedologic map of Croatia; Shuttle Radar Topography Mission (SRTM) and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) global digital elevation models.

**Table 1.** Selected criteria for soybean suitability calculation.

Criteria Group	Criterion Name	Unit	Description	Source	Native Format
Climate	T <sub>min</sub>	°C	Mean minimum air temperature	DHMZ	tabular
	T <sub>avg</sub>	°C	Mean average air temperature	DHMZ	tabular
	T <sub>max</sub>	°C	Mean maximum air temperature	DHMZ	tabular
	Precipitation	mm	Total precipitation amount	DHMZ	tabular
	AirHumidity	%	Mean relative air humidity	DHMZ	tabular
	Solar Irradiation	kWh/m <sup>2</sup>	Mean daily global horizontal irradiation	ASTER	raster
Soil	SoilType	/	Soil type classes	Basic pedologic map of Croatia	vector (polygon)
	pH	/	Soil pH values	CAEN	vector (point)
	SoilTexture	/	Soil texture classes	CAEN	vector (point)
	C/N	/	Carbon-to-nitrogen soil ratio	CAEN	vector (point)
	Slope	°	Terrain slope	SRTM	raster
	TWI	/	Topographic wetness index	SRTM	raster

Climate criteria consisting of T<sub>min</sub>, T<sub>avg</sub>, T<sub>max</sub>, Precipitation and AirHumidity were interpolated from 15 DHMZ weather station data. Soil samples for modeling of soil criteria were downloaded from HAOP Web Feature Service. Samples were collected in the field by Croatian Geological Survey and Croatian Agency for Agricultural Land according to standards ISO 10390:2005 for soil pH, ISO 11277:2009 for soil texture and ISO 11466:1995 for soil C/N. From 72 regularly distributed samples in the study area, 48 were detected as agricultural area land cover and filtered for further processing. The solar irradiation criterion was defined as Global Horizontal Irradiation (GHI), according to previous research [54]. GHI was calculated using the novel method developed by Gašparović et al. [55], based on the ASTER digital elevation model, Linke turbidity factor [56] and effective cloud albedo acquired from Meteosat Second Generation satellites SARAH Edition 2 [57]. According to Gašparović et al. [55], GHI produced 3% higher accuracy than commercial solar irradiation data, compared to the data measured from ground stations. The same approach for GHI calculation was successfully applied in similar research [14]. The final GHI used in this research was an average daily mean value (kWh/m<sup>2</sup>) between April and October in 2013 to 2015. A three-year period was used to reduce the influence of weather conditions. The time frame between April and October covers the vegetation period of all major soybean varieties cultivated in the study area. T<sub>min</sub> represents a mean minimum air temperature for April and May, as soybean is most susceptible to frost at early vegetation stages that occur during April and May. T<sub>max</sub> represents mean maximum air temperatures in June, July and August and refers to the drought risk. Soil type classes were derived from a basic pedologic map of Croatia. Soil pH was selected as a criterion due to soybean demands for neutral soil, with neutral and mildly acidic soils having optimal values. Soybean does not have strict demands regarding soil C/N [58], but it has a major impact on sustainable planning of overall agricultural production in the study area. Soil C/N deficiency presents a major issue in the study area [59], so crop types with larger demands for C/N would benefit more for being cultivated on soil with higher soil organic content. Soil texture classification was conducted in twelve classes from the soil texture triangle, according to silt, sand and clay soil content [60]. Loam was considered as the most suitable soil texture, allowing the optimal root system development for soybean [61]. The process of soil texture raster generation from silt, sand and clay soil content rasters was automatized using Python v3.7.4 (Wilmington, Delaware, United States of America) [62]. The slope was determined using SRTM 1-arc second global digital elevation model. The slope is associated with the exploitation of air humidity in soybean cultivation, which is important for its development in reproductive growth stages. Terrain slope values were considered as hilly terrain restricts the adequate exploitation of agricultural machinery. Topographic wetness index (TWI) was

calculated in SAGA GIS for the determination of specific catchment areas, representing the effect of inclined slopes on the soil water content [63].

Preprocessing of input criteria consisted of the conversion of polygon vector criterion to a raster format, georeferencing of weather station data and conversion to point vector format, evaluation of interpolation methods for climate and soil points, interpolation of these values using the most accurate method and clipping of rasters to study area extents. The selected projection coordinate system was the Croatian Terrestrial Reference System (HTRS96/TM, EPSG: 3765) with 250 m spatial resolution (ground sample distance), so all rasters were reprojected and resampled accordingly. Interpolation methods selected for interpolation accuracy assessment were Ordinary Kriging (OK) [64], Inverse Distance Weighting (IDW) [65] and Angular Distance Weighting (ADW) [66]. OK and IDW were selected as they resulted as the most accurate interpolation methods for the interpolation of soil parameters in studies by Yao et al. and Qiao et al. [67,68]. IDW and ADW produced the highest accuracy of tested deterministic interpolation methods for climate data in a study by Xavier et al. [69]. Descriptive statistics (Table 2) were calculated for the determination of input samples normality and stationarity with the aim of selection of optimal interpolation method [70]. The mathematical model and fitting range for OK interpolation were selected on the highest possible fitting coefficient of determination to variogram. Both IDW and ADW were performed with a weight parameter of 2. Accuracy assessment was conducted based on cross-validation with the leave-one-out procedure, with the coefficient of determination ( $R^2$ ), root-mean-square error (RMSE) and normalized RMSE (NRMSE) as statistical indicators. NRMSE was calculated by dividing RMSE with a respective average of measured values from input climate or soil samples.

### 2.3. Data Standardization

Three standardization methods were evaluated in this research: linear standardization, stepwise standardization and fuzzy standardization. The defined number intervals for standardization were closed intervals containing values from zero to one (0.00, 1.00) or one to zero (1.00, 0.00), where 0.00 designates least favorable and 1.00 designates the most favorable impact on suitability values. The integration of qualitative and quantitative criteria values was conducted in this procedure. Values from the selected number interval during standardization were designated to either thematic classes in case of qualitative criteria or input numeric values for quantitative criteria, according to the level of suitability. Linear, stepwise and fuzzy standardization were applied to all selected criteria, except for the qualitative criteria SoilType and SoilTexture. For the two qualitative criteria, stepwise standardization was selected as a standardization method, being the only method that could operate with qualitative values. Consequentially, these values were used in combination with all three applied standardization methods for quantitative criteria.

The linear standardization method, based on linear scale transformation, is the most frequently used deterministic method for standardization in GIS-based multicriteria analysis [71]. The deterministic nature of the method ensures completely objective standardization, with no crop expert's impact on the standardization result. Linear standardization produced good results for the standardization of distances to infrastructure objects [72] and climate data for crop suitability calculation [73]. The calculation of linear standardization values was based on the score range procedure [20].

Stepwise standardization is also a commonly used standardization method in GIS-based multicriteria analysis [74]. It is based on the distinctive selection of the input values in number intervals defined by the crop expert for the standardization in new classes. An arbitrary value in the standardization interval has been assigned to each new class, allowing an extensive subjective crop expert's impact on the result. Four to six classes were used for stepwise standardization, depending on the heterogeneity of input criteria values. Stepwise standardization was successfully applied for the standardization of soil chemical properties [75] and climate criteria [76].

Fuzzy standardization was based on the application of fuzzy membership functions. Similarly to stepwise standardization, fuzzy membership functions allowed extensive impact on the standardization result by the subjective crop expert. Three membership functions were used in this research: linear, S-shaped and J-shaped [77]. Linear membership function (2) was a base for the calculation of S-shaped (3) and J-shaped (4) membership function values. Topographic indicators, such as slope, were successfully standardized using fuzzy membership functions [78]. The same method was also applied for the standardization of soil criteria [79] and the combination of soil and climate criteria [80] in agriculture land suitability calculations. Membership function parameters were marked as  $a$  and  $d$  for the definition of support, alongside  $b$  and  $c$  for the definition of the core of the membership function [81]. The variety of selected membership functions allowed the subjective crop expert's impact on the intensity of ascending and descending of standardized values. Intermediate values between parameter values were gradually calculated, allowing the continuous representation of the criteria effect on suitability. Equations for the calculation of the standardized values using fuzzy membership functions were derived from Novák [77].

#### 2.4. Weight Determination

Weight determination was conducted using the AHP method in three approaches for climate criteria group, soil criteria group and combined criteria group consisting of both climate and soil criteria. Two pairwise comparison matrices were created to quantify criteria weights for climate and soil criteria, as a base of the AHP method [82]. Individual weights per criteria group were calculated using separate pairwise comparison tables. Climate and soil criteria groups were considered as equally influential in suitability result. The primary motive of using three approaches was the estimation of criteria groups' impact on the suitability result in the validation process. This procedure allowed the criteria modifications for further application, through the elimination of redundant criteria and the addition of new criteria. AHP is regarded as one of the most advantageous weight determination methods in the multicriteria analysis by Musakwa [83], primarily due to being flexible, straightforward and comprehensive. It was successfully applied in GIS-based multicriteria analyses for crop cultivation suitability [10,84,85], irrigation suitability [76,86] and environmental analyses [14].

Each combination of two criteria per criteria group was evaluated in pairwise comparison matrices by assigning the preferred criteria value ranging from 1 (equal preference) to 9 (extreme preference) [82]. The determined values were associated with preferred criteria in the pairwise comparison matrix, while the reciprocal value was assigned to the less-preferred criteria. The consistencies of pairwise comparisons per matrix were evaluated by Consistency Ratio (CR) (1), using the Consistency Index (CI) (2) and the Random Consistency Index (RI) [87]:

$$CR = \frac{CI}{RI}, \quad (1)$$

$$CI = \frac{\lambda - n}{n - 1}, \quad (2)$$

where  $\lambda$  is the average value of consistency vector and  $n$  is the number of parameters.  $RI$  values were predefined by Saaty [87], depending on the number of compared criteria in the pairwise comparison matrix. The acceptable value of  $CR$  is 0.10, while higher values indicate that the revision and modification of the pairwise matrix should be conducted [88].

#### 2.5. Criteria Aggregation

The weighted linear combination was the selected method for criteria aggregation in the process of soybean suitability calculation. It is the most commonly used aggregation method in the GIS-based multicriteria analyses [89]. The advantages of the selected method were simplicity and time-efficiency in the criteria aggregation, as well as the objective selection of the most suitable location and its alternatives, ranked by the suitability values [90].

The nine suitability combinations of three standardization methods and three weight determination methods were calculated, resulting in nine suitability maps. Soybean suitability index (*SSI*) was selected as soybean land suitability value, ranging from 0 to 1. According to Eastman [20], the equation for *SSI* calculation using weighted linear combination was (3):

$$SSI = \sum w_i X_i \times C \quad (3)$$

where  $w_i$  denotes the weight of factor  $i$ ,  $X_i$  denotes the standardized values of factor  $i$  and  $C$  denotes a Boolean layer of constraints.

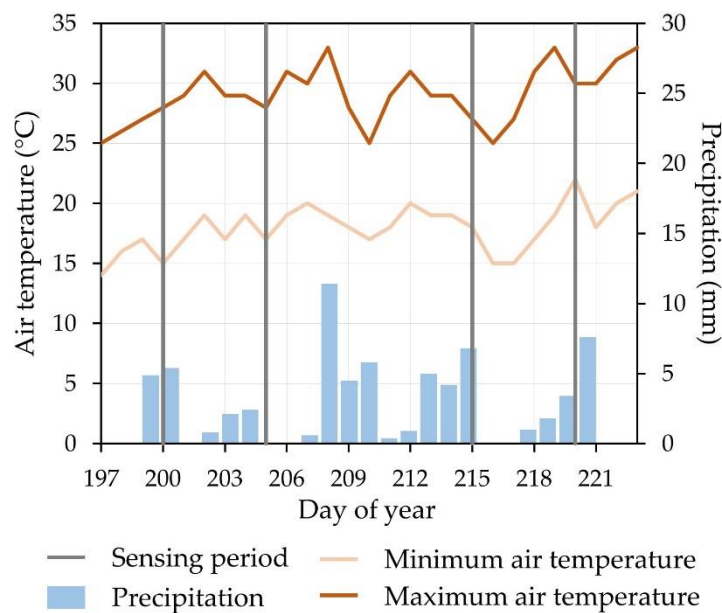
### 2.6. Validation of Calculated Suitability Models

Calculated suitability values were validated by NDVI values derived from Sentinel-2 Level 2A images. No official database or reliable records about yield exist in Croatia, so NDVI was selected as a basis for validation. Since it was derived from Sentinel-2 images, nearly global applicability is ensured. Four images were downloaded, reprojected to HTRS96/TM and clipped to study area extents to cover the soybean seed filling growth stage (R6). This growth stage is associated with the highest total above-ground biomass during the soybean phenology cycle [91]. NDVI derived from Sentinel-2 images produced a high correlation with biomass and chlorophyll content of crops in previous research [92]. Soybean total above-ground biomass also resulted in a high correlation with yield [93]. These factors were selected as representative for validation with NDVI, as higher biomass and yield are expected on more suitable land for soybean cultivation. Not all soybean variants cultivated in the study area reach the R6 stage at the same time due to different periods of sowing, so the necessity for multitemporal images emerged. The time frame starting from 15 July to 10 August was determined for all soybean varieties in the study area to reach the R6 growth stage, based on the empirical knowledge of soybean cultivation in the study area. Four Sentinel-2 images were selected for validation from this time frame based on their availability and lowest possible cloud cover percentage (Table 3). The used images were temporally evenly distributed in the time frame of the R6 soybean growth stage and sensed in very similar meteorological conditions (Figure 3). All four NDVI images were used for the determination of peak NDVI of soybean test parcels, with slightly more than two-thirds having its peak NDVI during late July.

A total of 204 soybean test parcels, evenly distributed on the study area, were selected as the ground truth data. The total cumulative area of soybean parcels was 849.3 ha, with an average individual area of 4.16 ha. Shapefile vector data representing soybean parcels were collected from Croatian Paying Agency for Agriculture, Fisheries and Rural Development official spatial database, filtered by the soybean crop type. Four NDVI rasters derived from downloaded Sentinel-2 images were overlaid with the centroids of soybean parcels, so all NDVI values within a soybean parcel were averaged and designated to that centroid. Since spatial resolutions of suitability rasters and Sentinel-2 images were different (250 m and 10 m, respectively), centroids of reference soybean parcels with peak NDVI values were used to perform validation of suitability rasters. Centroids of soybean test parcels were forced inside polygons during creation to represent geometrically irregular parcels accurately. The comparison of these data sets was achieved with fast processing time, as point vector was overlaid with raster data. Peak NDVI values per centroid were determined as the maximum NDVI value per centroid, based on four NDVI rasters derived from multitemporal Sentinel-2 images.

The use of peak NDVI allowed the detection of the R6 growth stage for all ground truth polygons, which represented all soybean variants in the study area. Such a procedure proved to be robust regarding the atmospheric effects like clouds and haze on satellite images since validation does not depend on the individual images. Identification data of Sentinel-2 images used for the validation of suitability models are shown in Table 2.





**Figure 3.** Weather data at the sensing time of four used Sentinel-2 images from weather station Osijek provided by the Croatian Meteorological and Hydrological Service (DHMZ).

**Table 2.** Identification of Sentinel 2 images, date and the number of peak Normalized Difference Vegetation Index (NDVI) values per image.

Tile ID	Satellite	Sensing Date	Day of Year	Peak NDVI Values
T34TCR	S2B	19th July 2019	200	38
T33TYL	S2A	24th July 2019	205	102
T33TYL	S2A	3rd August 2019	215	49
T33TYL	S2B	8th August 2019	220	15

Four regression functions were selected for the calculation of  $R^2$  between suitability values and reference peak NDVI values. Selected functions were linear ( $R^2_{lin}$ ), logarithmic ( $R^2_{log}$ ), exponential ( $R^2_{exp}$ ), second-order polynomial ( $R^2_{poly2}$ ) and third-order polynomial ( $R^2_{poly3}$ ). Different regression functions were evaluated to determine the most accurate relationship model between suitability values and peak NDVI. Validation of each suitability model was based on the highest  $R^2$  calculated using these functions. RMSE and Cohen's d index were calculated as a complementary statistical values to  $R^2$  for accuracy assessment. Cohen's d values were interpreted according to [94]. Peak NDVI and suitability values were normalized in the (0.00, 1.00) range prior to RMSE and Cohen's d calculation to describe suitability results while referring to the same number interval.

### 2.7. Creation of Final Suitability Maps

Unsupervised classification using the k-means algorithm was applied for the delineation of five suitability classes for each calculated suitability model. Created suitability classes were classified according to FAO standards in classes representing very suitable (S1), moderately suitable (S2), marginally suitable (S3), currently unsuitable (N1) and permanently unsuitable area (N2) [95]. Unsupervised classification was therefore conducted in five classes to meet the FAO specification. Ranking of class suitability values was conducted using mean suitability values of all pixels per class. The selected method approached the suitability values on a relative basis using a computer-automated classification. This procedure enabled the creation of suitability classes independently of selected criteria or study area properties. Such an approach reduced the effect of subjectivity in the pairwise comparisons of criteria in AHP, which is its major disadvantage [96]. For the application of this method in the regionalization of agricultural production, the integration of suitability values for

different crop types, and consequentially from different crop experts, is necessary. By managing the suitability values using unsupervised classification, the integration was possible since the relative criteria relationship was retained. The crop experts' judgment levels have been balanced in the process, preventing overestimation or underestimation of suitability. The applied approach also ensured the objectiveness in the creation of suitability classes, as the training data creation was not necessary, in contrast to supervised classification. Training data creation presents a common source of human error in classification due to the operator's subjectivity in the selection of training polygons [97]. Input data in the k-means algorithm was raster representing suitability values from the most accurate suitability model. Algorithms based on unsupervised classification were successfully applied for suitability classification with a single raster input [27]. K-means was successfully applied in various combinations with GIS-based multicriteria analysis in crop suitability analyses [25,98,99] and delineation of suitability classes in precision agriculture [100]. Final suitability maps were created by the removal of constrained areas from the unsupervised classification results. CORINE 2018 Land Cover data was applied to exclude agricultural land cover, as no other land cover was available for efficient soybean cultivation. The goal of this step was to exclude permanent natural (forests, wetlands, water bodies) and built-up objects from classification in a 250 m spatial resolution, so CORINE 2018 Land Cover data was selected as the most fitting data source due to its high thematic accuracy (85%) [101]. All calculated suitability models were mutually compared using Pearson's correlation coefficient in the correlation matrix.

### 3. Results

The descriptive statistics of input samples for evaluation of interpolation methods are shown in Table 3. Moderate skewness and high kurtosis deviations were observed from optimal values of zero and three, respectively. While low coefficients of variation indicated high data stationarity, low data normality was observed for both climate and soil criteria. Sand soil content was the only exception from this observation, as it produced both low normality and stationarity.

**Table 3.** Descriptive statistics of climate and soil samples for interpolation.

Criteria name	n	mean	CV	SK	KT
T <sub>min</sub>	15	13.8	0.029	0.075	−1.302
T <sub>avg</sub>	15	18.2	0.033	0.284	−1.365
T <sub>max</sub>	15	22.9	0.046	−0.469	−1.048
Precipitation	15	457.5	0.100	1.275	0.818
AirHumidity	15	71.9	0.052	1.077	1.689
pH	48	6.7	0.159	0.061	1.400
SoilTexture (Clay)	48	31.5	0.332	0.525	0.176
SoilTexture (Silt)	48	57.8	0.215	0.600	0.563
SoilTexture (Sand)	48	10.7	1.190	1.945	4.348
C/N	48	7.6	0.251	0.747	1.537

CV: coefficient of variation, SK: skewness, KT: kurtosis.

Accuracy assessment results of tested interpolation methods are shown in Table 4. Both deterministic methods outperformed OK in a case of climate data with a lower number of samples. IDW was selected as an optimal method for the interpolation of climate data, producing higher accuracy values than ADW for T<sub>min</sub>, Precipitation and AirHumidity. Accuracy of interpolation for T<sub>avg</sub> and T<sub>max</sub> produced mixed results, as ADW produced higher R<sup>2</sup> for T<sub>avg</sub> and higher RMSE and NRMSE for T<sub>max</sub> than IDW, but lower RMSE, NRMSE and R<sup>2</sup>, respectively. All tested methods produced lower mean accuracy for soil data compared to climate data. IDW produced the highest accuracy values for the interpolation of moderately variable data: Clay, Silt and C/N. ADW produced the highest R<sup>2</sup> for pH interpolation and second-best RMSE and NRMSE values behind IDW. OK produced slightly higher RMSE and NRMSE for interpolation of sand, but 0.0346 lower R<sup>2</sup> than IDW. IDW was therefore also selected as an optimal interpolation method for soil data.

**Table 4.** Accuracy assessment of tested interpolation methods.

Input Data	OK			IDW			ADW		
	$R^2$	RMSE	NRMSE	$R^2$	RMSE	NRMSE	$R^2$	RMSE	NRMSE
T <sub>min</sub>	0.6992	3.479	0.252	<b>0.8371</b>	<b>0.778</b>	<b>0.056</b>	0.7598	0.817	0.059
T <sub>avg</sub>	0.7491	1.965	0.108	0.8150	<b>0.797</b>	<b>0.043</b>	<b>0.8198</b>	0.831	0.045
T <sub>max</sub>	0.6060	4.091	0.178	<b>0.8029</b>	1.155	0.050	0.7248	<b>1.050</b>	<b>0.045</b>
Precipitation	0.6787	72.451	0.158	<b>0.8165</b>	<b>25.915</b>	<b>0.056</b>	0.7806	26.269	0.057
Air Humidity	0.6589	8.106	0.112	<b>0.7311</b>	<b>5.049</b>	<b>0.070</b>	0.6842	5.983	0.083
pH	0.5991	0.643	0.096	0.7313	<b>0.526</b>	<b>0.078</b>	<b>0.7407</b>	0.603	0.090
Clay	0.5512	6.570	0.208	<b>0.7022</b>	<b>5.751</b>	<b>0.182</b>	0.6467	6.604	0.209
Silt	0.5930	10.299	0.178	<b>0.6695</b>	<b>9.247</b>	<b>0.160</b>	0.6487	10.458	0.180
Sand	0.6183	<b>1.472</b>	<b>0.137</b>	<b>0.6529</b>	1.477	0.138	0.6458	1.531	0.143
C/N	0.5872	1.873	0.246	<b>0.7726</b>	<b>0.654</b>	<b>0.086</b>	0.6937	0.744	0.097

The most accurate statistical values per interpolation result were bolded.

Soil texture classes resulting after classification were silty clay, silty clay loam, clay loam, loam and silt loam. The dominant soil texture class in the study area was silty clay loam, covering 72.1% of the total area. The next two largest classes, silty clay and silt loam, had significant silt content, covering 21.6% of the study area combined. Clay loam and loam mostly covered the transitional areas between two larger classes, resulting in 3.2% and 3.1% of the study area, respectively. All preprocessed input criteria rasters are shown in Figure 4.

Parameters for each standardization method were selected as shown in Figure 5. The presence of a few extreme input values prevented the linear standardization method to expand the full (0.00, 1.00) number interval to the majority of pixels, resulting in an inaccurate representation of these criteria. This primarily refers to slope and C/N, where the number interval covers most of the pixels in (0.85, 1.00) and (0.00, 0.25) number interval, respectively. Intermediate values for stepwise standardization were selected based on standardization classes count, alongside a minimum value of 0.00 and a maximum of 1.00. The J-shaped and linear methods were the most commonly used fuzzy membership function for standardization, both being applied for 4 out of 10 quantitative criteria. Precipitation, SolarIrradiation and AirHumidity resulted in a high variability, while T<sub>avg</sub>, T<sub>min</sub> and T<sub>max</sub> produced moderate variability of the climate conditions in the study area. All soil criteria resulted in high variability in the study area.

Pairwise comparison matrices were created for climate (Table 5) and soil criteria (Table 6). Consistency tests for both matrices resulted in the allowed tolerance. The weight sums of interpolated criteria were 0.877 for climate and 0.431 for soil criteria group.

Calculated  $R^2$ , RMSE and Cohen's  $d$  values for the validation of suitability models are displayed in Table 7. Suitability calculated using the climate criteria group constantly produced the lowest  $R^2$ , with a maximum for fuzzy standardization using exponential regression. Stepwise and fuzzy standardization models both produced the highest accuracies for soil and combined criteria groups, with combined variants being slightly more accurate. Overall, fuzzy standardization with combined criteria produced the highest  $R^2$  and RMSE, presenting the optimal model for soybean land suitability mapping. RMSE values were based on normalized values, so they resulted in (0.00, 1.00) number interval and therefore enabled objective accuracy assessment of suitability values. Cohen's  $d$  values indicated the strongest relationship between fuzzy standardization and soil criteria with peak NDVI values. The fuzzy standardization with combined criteria also resulted in a small effect size, having a large value gap between the third-lowest  $d$  value of fuzzy standardization with climate criteria.



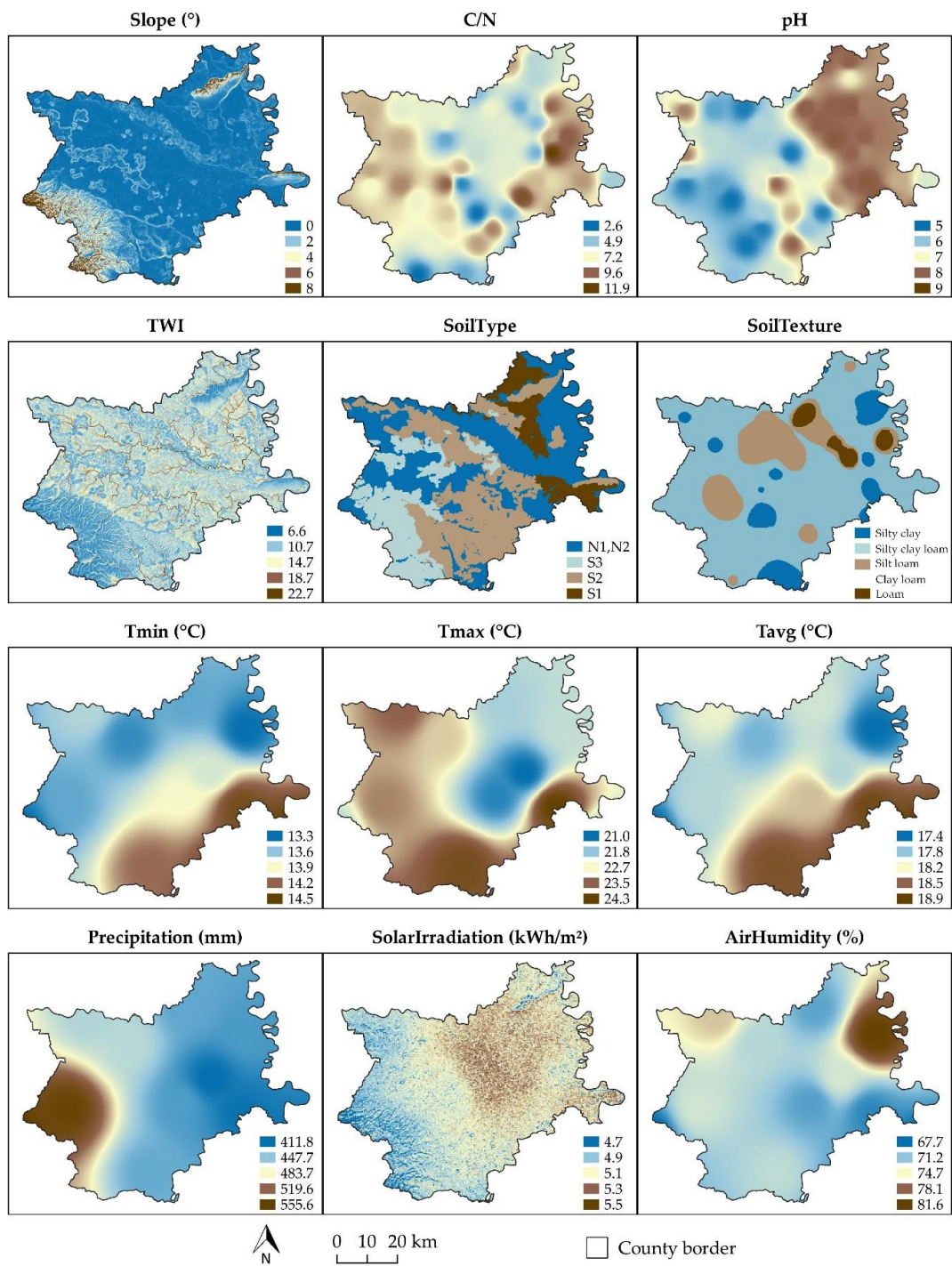


Figure 4. Preprocessed criteria rasters.

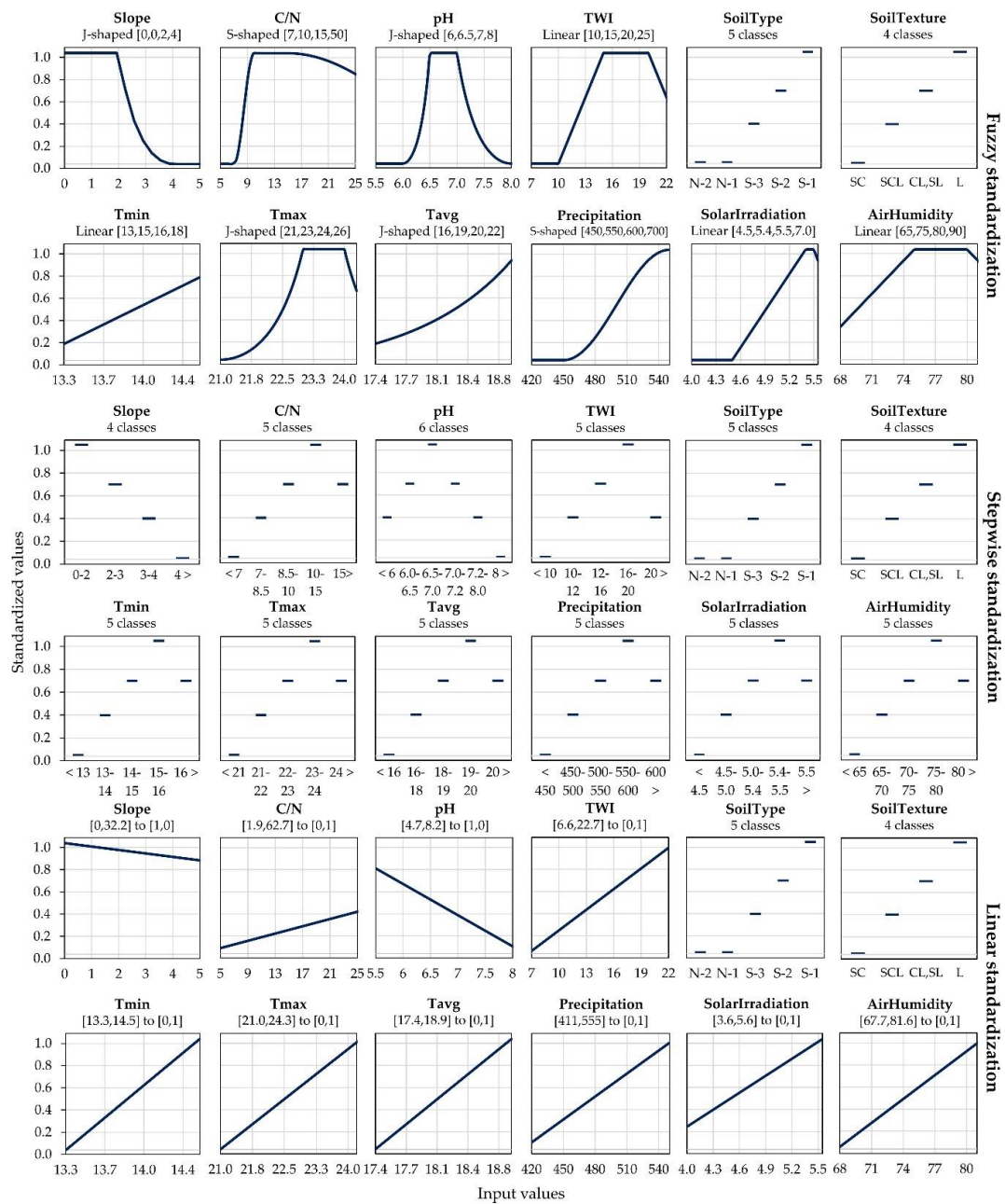


Figure 5. Graphs of standardization parameters per criterion for evaluated standardization methods.

Table 5. Pairwise comparison matrix of climate criteria for the weight determination in AHP.

	Tmin	Precipitation	Solar Irradiation	Tavg	Tmax	AirHumidity	Weight
T <sub>min</sub>	1	2	3	3	4	7	0.388
Precipitation	1/2	1	2	3	3	6	0.261
SolarIrradiation	1/3	1/2	1	2	3	4	0.156
T <sub>avg</sub>	1/3	1/3	1/2	1	2	3	0.104
T <sub>max</sub>	1/4	1/3	1/3	1/2	1	3	0.082
AirHumidity	1/7	1/6	1/4	1/3	1/3	1	0.042

$$n = 6, CI = 0.050, RI = 1.240, CR = 0.040.$$

**Table 6.** Pairwise comparison matrix of soil criteria for the weight determination in Analytic Hierarchy Process (AHP).

	SoilType	pH	Slope	SoilTexture	C/N	TWI	Weight
SoilType	1	2	3	3	4	7	0.383
pH	1/2	1	2	3	4	6	0.254
Slope	1/3	1/2	1	3	4	5	0.178
SoilTexture	1/3	1/3	1/3	1	2	4	0.104
C/N	1/4	1/4	1/4	1/2	1	3	0.073
TWI	1/7	1/6	1/5	1/4	1/3	1	0.039

$$n = 6, CI = 0.093, RI = 1.240, CR = 0.075.$$

**Table 7.** Accuracy assessment values of calculated suitability models using reference peak NDVI values.

Value	Fuzzy standardization			Stepwise standardization			Linear standardization		
	Climate	Soil	Climate + Soil	Climate	Soil	Climate + Soil	Climate	Soil	Climate + Soil
$R^2_{lin}$	0.4056	0.6839	0.8416	0.3016	0.6116	0.6947	0.3290	0.4855	0.6337
$R^2_{log}$	0.4143	0.6644	0.8273	0.3175	0.6046	0.6944	0.3249	0.4819	0.6279
$R^2_{exp}$	<b>0.4191</b>	0.6672	0.8417	0.2844	0.5760	0.6771	<b>0.3422</b>	0.4755	0.6289
$R^2_{poly2}$	0.4161	0.6907	0.8429	0.3319	0.6117	0.6975	0.3293	0.4858	0.6337
$R^2_{poly3}$	0.4162	<b>0.6923</b>	<b>0.8438</b>	<b>0.3326</b>	<b>0.6291</b>	<b>0.7095</b>	0.3310	<b>0.4929</b>	<b>0.6357</b>
RMSE	<b>0.1874</b>	<b>0.1435</b>	<b>0.0962</b>	<b>0.2070</b>	<b>0.1891</b>	<b>0.1926</b>	<b>0.2156</b>	<b>0.2089</b>	<b>0.1925</b>
$d$	<b>0.2717</b>	<b>0.0049</b>	<b>0.0147</b>	<b>0.2987</b>	<b>0.3720</b>	<b>0.5990</b>	<b>0.2780</b>	<b>0.3434</b>	<b>0.4738</b>

The most accurate statistical values per suitability result are bolded.  $R^2_{lin}$ :  $R^2$  from linear regression;  $R^2_{log}$ :  $R^2$  from logarithmic regression;  $R^2_{exp}$ :  $R^2$  from exponential regression;  $R^2_{poly2}$ :  $R^2$  from second-order polynomial regression;  $R^2_{poly3}$ :  $R^2$  from third-order polynomial regression; RMSE: root mean square error,  $d$ : Cohen's  $d$  index.

Visual representation of regression functions that produced the highest coefficient of determination per model is presented in Figure 6. The similarity of stepwise and fuzzy standardization with soil and combined criteria results can be observed, with fuzzy models resulting in slightly smaller dispersion. RMSE values showed similar results, as fuzzy methods produced lower RMSE than stepwise standardization for both soil (difference 0.0456) and combined criteria group (difference 0.0964). Climate criteria resulted in the lowest sensitivity to the selected standardization method, having closely distributed values for all applied standardization methods. The third-order polynomial regression function resulted in the highest  $R^2$  in the validation of all soil and combined criteria. The exponential function most accurately represented the relationship of suitability and peak NDVI values for climate criteria with fuzzy and linear standardization methods.

Shares of suitability classes calculated using k-means unsupervised classification regarding the covered area are shown in Table 8. According to the most accurate suitability values calculated using fuzzy standardization and combined criteria, 64.3% of the study area is suitable for soybean cultivation to some degree. The S1 class covers 602 km<sup>2</sup> of the county area. The area determined as very suitable covers three major parts: the eastern part near Dalj, mainly characterized with high average minimal air temperatures; the northeastern part by the Baranja hill with optimal C/N values, pH and SoilType values; and the central-south part by the city of Đakovo with suitable SoilTexture and SoilType values. The final suitability maps after the removal of constrained areas from the unsupervised classification results are shown in Figure 7. The most accurate suitability model was exported to vector Shapefile and raster GeoTIFF format.

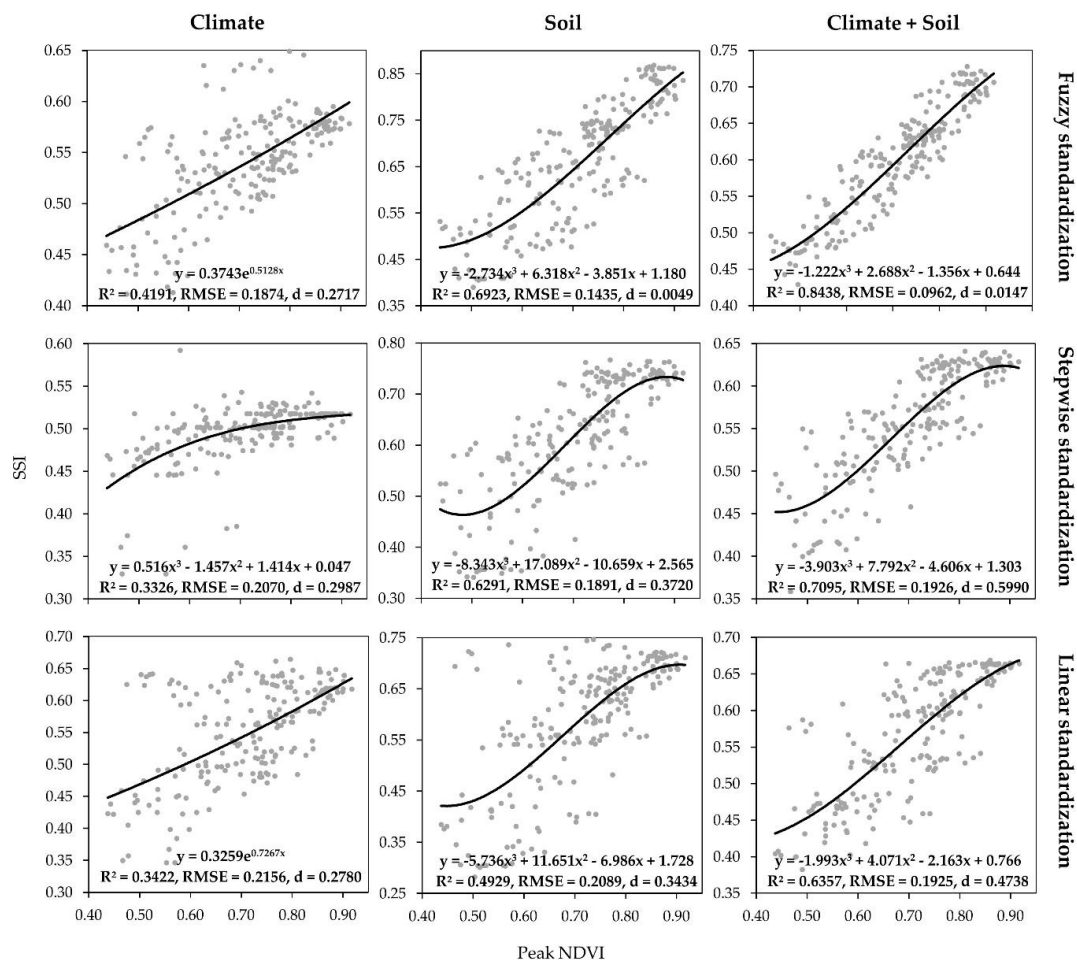


Figure 6. Regression graphs for best fitting functions per the suitability model.

Table 8. Shares of suitability classes for calculated models from unsupervised classification.

Model		S1 (%)	S2 (%)	S3 (%)	N1 (%)	N2 (%)
Fuzzy standardization	Climate	13.1	25.5	26.2	23.2	12.0
	Soil	23.6	14.7	30.5	21.7	9.5
	Climate + Soil	14.5	22.2	27.6	22.5	13.2
Stepwise standardization	Climate	33.4	26.9	3.6	27.5	8.6
	Soil	19.1	23.0	15.9	18.5	23.5
	Climate + Soil	9.2	26.3	25.0	20.9	18.6
Linear standardization	Climate	16.8	23.5	24.6	24.9	10.2
	Soil	19.2	20.5	21.6	15.4	23.3
	Climate + Soil	18.6	25.5	23.9	17.1	14.9

Suitability classes: very suitable (S1), moderately suitable (S2), marginally suitable (S3), currently unsuitable (N1) and permanently unsuitable (N2).



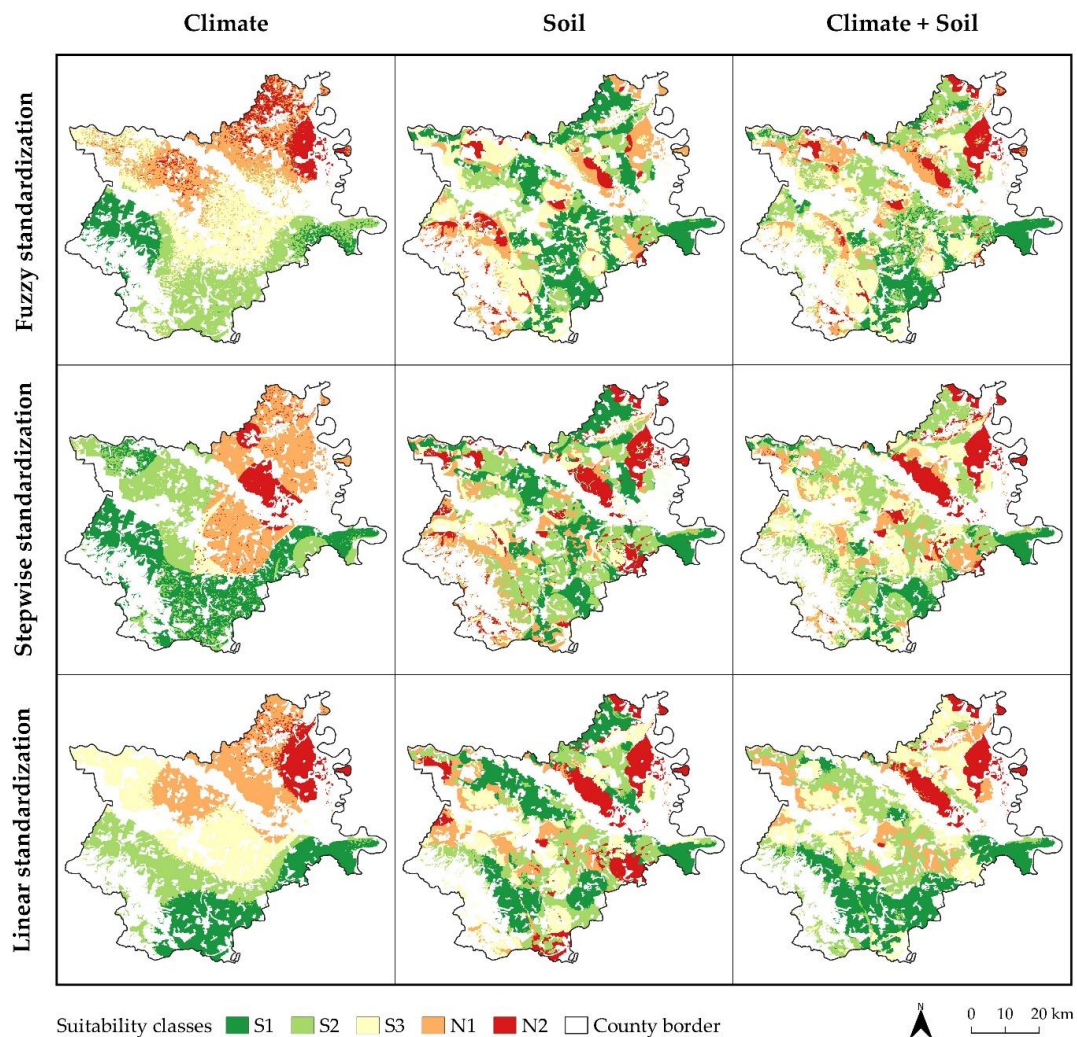


Figure 7. Final suitability maps for soybean cultivation for calculated models.

The correlation matrix of all combinations of calculated suitability models (Table 9) showed a constantly high correlation of suitability values calculated using soil and combined criteria for all three standardization methods. Suitability values using climate criteria produced a moderate correlation with combined criteria values. A high correlation was achieved for fuzzy and stepwise standardization with combined criteria, which produced two most accurate suitability values.

Table 9. Correlation matrix between calculated suitability models.

	FC	FS	FCS	SC	SS	SCS	LC	LS	LCS
FC	1.000	0.285	0.578	0.578	0.278	0.396	0.718	0.242	0.520
FS		1.000	0.947	0.343	0.909	0.892	0.313	0.779	0.736
FCS			1.000	0.486	0.866	0.891	0.506	0.743	0.799
SC				1.000	0.333	0.555	0.493	0.292	0.452
SS					1.000	0.969	0.326	0.901	0.834
SCS						1.000	0.415	0.870	0.853
LC							1.000	0.294	0.689
LS								1.000	0.895
LCS									1.000

FC: fuzzy standardization-climate criteria, FS: fuzzy standardization-soil criteria, FCS: fuzzy standardization-combined criteria, SC: stepwise standardization-climate criteria, SS: stepwise standardization-soil criteria, SCS: stepwise standardization-combined criteria, LC: linear standardization-climate criteria, LS: linear standardization-soil criteria, LCS: linear standardization-combined criteria.

#### 4. Discussion

Linear standardization was proven to be less effective in adjusting to standardization specifications for nearly all quantitative criteria. This occurred due to the linear standardization's property of monotonous ascending or descending of standardized values in the specified number interval. Therefore, nonlinear varying standardization requirements of Precipitation, SolarIrradiation,  $T_{avg}$ ,  $T_{max}$ , AirHumidity, pH and TWI could not be fully adjusted. However, linear standardization performed as the second-best standardization method for climate criteria, as few extreme values were present in these criteria. Stepwise standardization allowed high subjective impact on standardized results due to the ability to join the arbitrary input value ranges to standardization classes. These classes were discretely defined, with no intermediate values between classes. The most important property of the stepwise standardization method was the ability of the standardization of qualitative criteria, which presents a necessity for the integration with quantitative criteria. Fuzzy standardization methods allowed the best adjustment of standardized values to the crop expert's specifications for quantitative criteria, consequentially resulting in the highest accuracies of suitability models. Aside from setting the intervals of least and most favorable input values, the crop expert also had supervision over the intensity of increasing and decreasing the standardized values by selecting the optimal fuzzy model. Sigmoidal and J-shaped membership functions were also recognized as the most convenient for the standardization of similar criteria in a study by Aydi et al. [89]. Investigated properties of fuzzy standardization proved that it could be used as a universal standardization method, possibly upgrading the calculation of suitability values where linear and stepwise standardization were previously combined [74]. The FAO guidelines offered a standard for the delineation of suitability classes, which enables a comparison between similar suitability models. These standards were applied in many studies regarding suitability analyses recently [10,15,78]. The creation of suitability classes is considered as a necessary procedure in effective land management in similar studies. They were previously used in suitability analyses regarding crop cultivation and irrigation [99].

Suitability models calculated with the combined criteria group produced the highest accuracy, showing the importance of climate and soil criteria groups in crop-related suitability analysis. A low correlation between models using only climate and soil criteria was observed for all standardization methods. Climate and soil criteria groups were both necessary for reliable suitability calculation [7,102]. The possible reason for the lower accuracy of suitability models that used climate criteria compared to soil or combined criteria is a relatively small study area (4,155 km<sup>2</sup>), causing low variability for selected climate criteria. Precipitation produced the highest variability of all data from weather stations, with a CV of only 0.100. However, climate data was considered as mandatory criteria group especially in the case of the larger study area, having an impact on high suitability values for combined criteria. Many recent studies successfully integrated GIS-based multicriteria analysis with AHP as a weight determination method for crop suitability calculation [11,25,80,83,84]. Besides weight determination, three modifications of AHP in this study enabled the determination of the reliability of input data used for the modeling of criteria per criteria groups. The combined criteria group produced the highest accuracy for all standardization methods. Fuzzy and stepwise standardization methods also produced moderate accuracy using soil criteria, while climate criteria produced the lowest accuracy. Linear standardization produced similar accuracies for all AHP modifications. Fuzzy standardization resulted in higher accuracy using the combined criteria group than the other two methods, due to better adjustment of values to the crop expert's specifications in the standardization process. All interpolation methods performed better in case of less variable climate data compared to soil data. OK usually outperformed IDW in studies comparing interpolation methods for soil and climate data but has some limitations regarding normality and stationarity of input samples [103]. IDW outperformed OK in case of low normality and high variability of sample data [104]. Kriging was unable to produce quality results in when its prerequisites of sample data normality and stationarity were not met in a different study, where IDW resulted in higher accuracy [25]. This emphasizes the importance of the selection of optimal interpolation method in similar studies as it depends on the properties of input samples [105].

No standard methodology of crop suitability values validation calculated by GIS-based multicriteria analysis was noted during the literature review. Moreover, the majority of these studies did not employ any validation procedure for crop suitability results [11,25,51,83,84,89]. Multitemporal NDVI offers a potential solution in the validation of crop suitability, due to its global accessibility from global multispectral satellite missions (Sentinel-2, Landsat 8) and being a reliable predictor of various crop properties [92]. Moreover, Sentinel-2 derived NDVI was a reliable predictor in a study of crop yield and was superior to other individual tested vegetation indices [106]. The applicability of NDVI was expanded in this study as it enabled an effective validation of suitability models, primarily due to the objective assessment of all major soybean varieties using peak NDVI values. An application of NDVI in the validation of wheat suitability combined with yield data was successfully made in a study by Dedeoğlu and Dengiz [107]. However, the application of peak NDVI values that were used for direct validation of suitability values before their classification in suitability classes represents a significant improvement for its applicability in suitability validation. Data about conducted agrotechnical operations (sowing, fertilization, spraying) for every test soybean parcel remain an ambiguity and could affect the reliability of validation using peak NDVI. Fertilization is traditionally balanced in the entire study area, but its exact effect on NDVI values for validation remain to be tested. Soybean satisfies around 60% of its nitrogen requirements by nitrogen fixation [108], which is one of the reasons for farmers in the study area to apply conventional fertilization without conducting detailed soil analysis commonly. Therefore, nitrogen fertilization was not expected to have a significant impact during validation using NDVI. The effect of pests, diseases and weather trend on crop status prevent NDVI from being fully reliable validation data, as vegetation depends on numerous individual factors [109]. A relatively small study area partially reduced the effects of weather conditions, having low variability. The averaging of NDVI values from soybean parcels to centroids allowed the lower impact of potentially affected areas by pests or disease. A possible solution for the mentioned restrictions is the integration of multiple complementary vegetation indices for validation, as the spectral resolution of Sentinel-2 allows for the calculation of most known vegetation indices [110]. The Soil-Adjusted Vegetation Index (SAVI) is considered as one of the complementary indices to NDVI, as it minimized the soil brightness effect in areas with sparser vegetation in a comparable study [111]. Of the four Sentinel-2 images used for the calculation of peak NDVI, the dataset sensed on 8 August 2019 produced the lowest amount of peak NDVI values due to the lack of precipitation at the beginning of the month. This case implies the importance of considering meteorological conditions for adjustment of the time frame for validation. The validation time frames should be slightly calibrated for each growing season, as drought could accelerate the ripening process of soybean, therefore lowering the vegetation indices values.

While the simplicity and effectiveness of the weighted linear combination for criteria aggregation were previously mentioned and successfully applied, the Ordered Weighted Averaging (OWA) method offers a potential upgrade to the applied methods. Its potential in GIS-based multicriteria analysis is primarily notable through the integration of fuzzy measures in the criteria aggregation process [112]. OWA was successfully applied for suitability analysis of crops in combination with AHP [113] and multicriteria analysis regarding soil chemical properties in combination with IDW as the interpolation method [114]. The majority of applied methods for GIS-based multicriteria analysis in this study support the automatization of the calculation process, which will be a subject of future research. Inputs consisting of criteria selection, pairwise matrix comparison and standardization specifications should be assigned at the beginning of the algorithm. The main benefits of automatization in this research were noticeable during the process of soil texture map calculation, which are time-efficiency and easier applicability of the process to future calculations [78]. Automatization is considered as a major property for making the GIS-based methods and algorithms considerably more applicable in regular practice [115]. Since the user does not have to conduct each process in the GIS environment manually, the accessibility of automated methods expands to users with limited knowledge of geospatial data.

The addition of the ecological criteria group to the multicriteria analysis presents a possibility for the improvement of the present study. Aside from the ecological criteria group, the proposed method

supports the addition of multiple other criteria groups through the creation of additional pairwise comparison matrices, as long as the number of criteria per group satisfies specifications by Saaty and Ozdemir [53].

## 5. Conclusions

Soybean suitability values were calculated based on agronomist specifications in this study, resulting in maps of soybean suitability classes, as a base for land policy managers to improve soybean quality and reduce production costs. This was achieved by the determination of soybean land suitability using GIS-based multicriteria analysis and AHP. During the standardization process, it was observed that the fuzzy standardization method allowed the best adjustment of standardized values to soybean land suitability specifications. The separation of criteria for soybean suitability calculation to climate and soil groups allowed weight determination according to specifications of AHP. This approach also reduced the redundancy of criteria in the process of pairwise comparison matrices creation. The authors recommend the evaluation and selection of optimal interpolation methods for the preprocessing of input criteria when applicable, as was the case with the most influential criteria in this study. The validation method using Sentinel-2 derived peak NDVI values enabled the determination of the most accurate suitability model. The calculation of peak NDVI values based on four independent NDVI values allowed the objective determination of peak biomass for all major soybean varieties. This procedure was also proven to be robust regarding the cloud cover on satellite images since validation does not depend on the individual images. Fuzzy standardization resulted in the most accurate standardization method combined with all three criteria groups. The suitability model using fuzzy standardization with a combined criteria group produced the highest  $R^2$  and RMSE from nine calculated suitability values. According to suitability classes created by unsupervised classification of the most accurate suitability values, 64.3% of the study area was determined as suitable for soybean cultivation in some degree. The most suitable class for soybean cultivation covered 14.5% of the study area, an equivalent of 602 km<sup>2</sup> in the field. Three regions of highest suitability for soybean cultivation were observed in the suitability map: the eastern part near Dalj, the northeast part by the Baranja hill and the central-south part by the city of Đakovo. The export of final suitability maps to common raster and vector formats were performed for effective knowledge sharing to agricultural land policy managers.

**Author Contributions:** Conceptualization: D.R. and M.G.; methodology: D.R.; validation: D.R., M.J. and M.G.; formal analysis: M.J. and I.P.; investigation: D.R.; resources: D.R. and M.G.; data curation: D.R.; writing—original draft preparation: D.R., M.G.; writing—review and editing: D.R., M.J., M.G. and I.P.; visualization: D.R.; supervision: M.J. and M.G.; funding acquisition: M.J. and I.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** This work was supported by the Faculty of Agrobiotechnical Sciences Osijek as a part of the scientific project ‘AgroGIT—technical and technological crop production systems, GIS and environment protection’.

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

## References

1. Bajželj, B.; Richards, K.S.; Allwood, J.M.; Smith, P.; Dennis, J.S.; Curmi, E.; Gilligan, C.A. Importance of food-demand management for climate mitigation. *Nat. Clim. Chang.* **2014**, *4*, 924. [[CrossRef](#)]
2. Pulighe, G.; Bonati, G.; Fabiani, S.; Barsali, T.; Lupia, F.; Vanino, S.; Nino, P.; Arca, P.; Roggero, P.P. Assessment of the Agronomic Feasibility of Bioenergy Crop Cultivation on Marginal and Polluted Land: A GIS-Based Suitability Study from the Sulcis Area, Italy. *Energies* **2016**, *9*, 895. [[CrossRef](#)]
3. von Arb, C.; Bünemann, E.K.; Schmalz, H.; Portmann, M.; Adamtey, N.; Musyoka, M.W.; Frossard, E.; Fliessbach, A. Soil quality and phosphorus status after nine years of organic and conventional farming at two input levels in the Central Highlands of Kenya. *Geoderma* **2020**, *362*, 114112. [[CrossRef](#)]



4. Davis, K.F.; Rulli, M.C.; Seveso, A.; D'Odorico, P. Increased food production and reduced water use through optimized crop distribution. *Nat. Geosci.* **2017**, *10*, 919. [[CrossRef](#)]
5. Purnamasari, R.A.; Noguchi, R.; Ahamed, T. Land suitability assessments for yield prediction of cassava using geospatial fuzzy expert systems and remote sensing. *Comput. Electron. Agric.* **2019**, *166*, 105018. [[CrossRef](#)]
6. Chen, J.; Yu, J.; Zhang, Y. Multivariate video analysis and Gaussian process regression model based soft sensor for online estimation and prediction of nickel pellet size distributions. *Comput. Chem. Eng.* **2014**, *64*, 13–23. [[CrossRef](#)]
7. FAO Soils bulletin—A framework for land evaluation. Available online: <http://www.fao.org/3/X5310E/x5310e00.htm> (accessed on 17 April 2020).
8. You, L.; Wood, S.; Wood-Sichra, U.; Wu, W. Generating global crop distribution maps: From census to grid. *Agric. Syst.* **2014**, *127*, 53–60. [[CrossRef](#)]
9. Wang, R.; Jiang, Y.; Su, P.; Wang, J. Global Spatial Distributions of and Trends in Rice Exposure to High Temperature. *Sustainability* **2019**, *11*, 6271. [[CrossRef](#)]
10. Layomi Jayasinghe, S.; Kumar, L.; Sandamali, J. Assessment of Potential Land Suitability for Tea (*Camellia sinensis* (L.) O. Kuntze) in Sri Lanka Using a GIS-Based Multi-Criteria Approach. *Agriculture* **2019**, *9*, 148. [[CrossRef](#)]
11. Plaščak, I.; Jurišić, M.; Radočaj, D.; Barač, Ž.; Glavaš, J. Hazel plantation planning using GIS and multicriteria decision analysis. *Poljoprivreda* **2019**, *25*, 79–85. [[CrossRef](#)]
12. Yun, H.J.; Kang, D.J.; Kim, D.-K.; Kang, Y. A GIS-Assisted Assessment and Attribute-Based Clustering of Forest Wetland Utility in South Korea. *Sustainability* **2019**, *11*, 4632. [[CrossRef](#)]
13. Herzberg, R.; Pham, T.G.; Kappas, M.; Wyss, D.; Tran, C.T.M. Multi-Criteria Decision Analysis for the Land Evaluation of Potential Agricultural Land Use Types in a Hilly Area of Central Vietnam. *Land* **2019**, *8*, 90. [[CrossRef](#)]
14. Gašparović, I.; Gašparović, M. Determining Optimal Solar Power Plant Locations Based on Remote Sensing and GIS Methods: A Case Study from Croatia. *Remote Sens.* **2019**, *11*, 1481. [[CrossRef](#)]
15. Chen, Y.; Yu, J.; Khan, S. Spatial sensitivity analysis of multi-criteria weights in GIS-based land suitability evaluation. *Environ. Modell. Softw.* **2010**, *25*, 1582–1591. [[CrossRef](#)]
16. Chakhar, S.; Mousseau, V. Spatial multicriteria decision making. In *Encyclopedia of Geographic Information Science*; Shekhar, S., Xiong, H., Eds.; Springer: New York, NY, USA, 2008; pp. 747–753. [[CrossRef](#)]
17. Ishizaka, A.; Labib, A. Analytic hierarchy process and expert choice: Benefits and limitations. *Or Insight* **2009**, *22*, 201–220. [[CrossRef](#)]
18. Yu, R.; Tzeng, G.H. A soft computing method for multi-criteria decision making with dependence and feedback. *Appl. Math. Comput.* **2006**, *180*, 63–75. [[CrossRef](#)]
19. Aragonés-Beltrán, P.; Chaparro-González, F.; Pastor-Ferrando, J.P.; Pla-Rubio, A. An AHP (Analytic Hierarchy Process)/ANP (Analytic Network Process)-based multi-criteria decision approach for the selection of solar-thermal power plant investment projects. *Energy* **2014**, *66*, 222–238. [[CrossRef](#)]
20. Eastman, J.R. Multi-criteria evaluation and GIS. In *Geographical Informational Systems*; Goodchild, M.F., Maguire, D.J., Rhind, D.W., Eds.; John Wiley and Sons: New York, NY, USA, 1999; pp. 493–502.
21. Malczewski, J.; Chapman, T.; Flegel, C.; Walters, D.; Shrubsole, D.; Healy, M.A. GIS—Multicriteria evaluation with ordered weighted averaging (OWA): Case study of developing watershed management strategies. *Environ. Plan. A* **2003**, *35*, 1769–1784. [[CrossRef](#)]
22. Tang, Z.; Yi, S.; Wang, C.; Xiao, Y. Incorporating probabilistic approach into local multi-criteria decision analysis for flood susceptibility assessment. *Stoch. Environ. Res. Risk Assess.* **2017**, *32*, 701–714. [[CrossRef](#)]
23. Nguyen, T.T.; Verdoodt, A.; Van, Y.T.; Delbecque, N.; Tran, T.C.; Van Ranst, E. Design of a GIS and multi-criteria based land evaluation procedure for sustainable land-use planning at the regional level. *Agric. Ecosyst. Environ.* **2015**, *200*, 1–11. [[CrossRef](#)]
24. Sallwey, J.; Bonilla Valverde, J.P.; Vásquez López, F.; Junghanns, R.; Stefan, C. Suitability maps for managed aquifer recharge: A review of multi-criteria decision analysis studies. *Environ. Rev.* **2019**, *27*, 138–150. [[CrossRef](#)]
25. Jurišić, M.; Plaščak, I.; Antonić, O.; Radočaj, D. Suitability Calculation for Red Spicy Pepper Cultivation (*Capsicum annum* L.) Using Hybrid GIS-Based Multicriteria Analysis. *Agronomy* **2020**, *10*, 3. [[CrossRef](#)]
26. Koch, B.; Khosla, R.; Frasier, W.M.; Westfall, D.G.; Inman, D. Economic feasibility of variable-rate nitrogen application utilizing site-specific suitability classes. *Agron. J.* **2004**, *96*, 1572–1580. [[CrossRef](#)]

27. Van Niekerk, A. A comparison of land unit delineation techniques for land evaluation in the Western Cape, South Africa. *Land Use Policy* **2010**, *27*, 937–945. [CrossRef]
28. Pagano, M.C.; Miransari, M. The importance of soybean production worldwide. In *Abiotic and Biotic Stresses in Soybean Production*; Miransari, M., Ed.; Academic Print: Cambridge, MA, USA, 2016; pp. 1–26.
29. The Role of Soybean in Fighting World Hunger. Available online: <http://www.fao.org/3/a-bs958e.pdf> (accessed on 21 December 2019).
30. EU Agricultural Outlook—European Commission. Available online: [https://ec.europa.eu/info/sites/info/files/food-farming-fisheries/farming/documents/agricultural-outlook-2019-report\\_en.pdf](https://ec.europa.eu/info/sites/info/files/food-farming-fisheries/farming/documents/agricultural-outlook-2019-report_en.pdf) (accessed on 17 April 2020).
31. USDA Agricultural Projections to 2029. Available online: <https://www.ers.usda.gov/webdocs/publications/95912/oce-2020-1.pdf?v=8056.6> (accessed on 17 April 2020).
32. Masuda, T.; Goldsmith, P.D. World soybean production: Area harvested, yield, and long-term projections. *Int. Food Agribus. Man.* **2009**, *12*, 1–20.
33. Maas, S.J.; Rajan, N. Estimating ground cover of field crops using medium-resolution multispectral satellite imagery. *Agron. J.* **2008**, *100*, 320–327. [CrossRef]
34. Immitzer, M.; Vuolo, F.; Atzberger, C. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sens.* **2016**, *8*, 166. [CrossRef]
35. Sentinel-2A Launch. Available online: [http://marine.copernicus.eu/wp-content/uploads/2016/06/r2495\\_9\\_sentinel\\_2a.pdf](http://marine.copernicus.eu/wp-content/uploads/2016/06/r2495_9_sentinel_2a.pdf) (accessed on 21 December 2019).
36. Sentinel-2 User Handbook. Available online: [https://sentinel.esa.int/documents/247904/685211/Sentinel-2\\_User\\_Handbook](https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook) (accessed on 13 April 2020).
37. Dong, T.; Meng, J.; Shang, J.; Liu, J.; Wu, B. Evaluation of chlorophyll-related vegetation indices using simulated Sentinel-2 data for estimation of crop fraction of absorbed photosynthetically active radiation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 4049–4059. [CrossRef]
38. Veloso, A.; Mermoz, S.; Bouvet, A.; Le Toan, T.; Planells, M.; Dejoux, J.F.; Ceschia, E. Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications. *Remote Sens. Environ.* **2017**, *199*, 415–426. [CrossRef]
39. Walz, Y.; Wegmann, M.; Dech, S.; Vounatsou, P.; Poda, J.N.; N’Goran, E.K.; Utzinger, J.; Raso, G. Modeling and validation of environmental suitability for schistosomiasis transmission using remote sensing. *PLoS Negl. Trop. Dis.* **2015**, *9*, e0004217. [CrossRef]
40. Vanino, S.; Nino, P.; De Michele, C.; Bolognesi, S.F.; D’Urso, G.; Di Bene, C.; Pennelli, B.; Vuolo, F.; Farina, R.; Napoli, R. Capability of Sentinel-2 data for estimating maximum evapotranspiration and irrigation requirements for tomato crop in Central Italy. *Remote Sens. Environ.* **2018**, *215*, 452–470. [CrossRef]
41. Conrad, O.; Bechtel, B.; Bock, M.; Dietrich, H.; Fischer, E.; Gerlitz, L.; Wehberg, J.; Wichmann, V.; Böhner, J. System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geosci. Model. Dev. Discuss.* **2015**, *8*, 1991–2015. [CrossRef]
42. QGIS Development Team 2019. QGIS Geographic Information System. Open Source Geospatial Foundation Project. Available online: <http://qgis.osgeo.org> (accessed on 17 April 2020).
43. GRASS Development Team 2019. Geographic Resources Analysis Support System (GRASS) Software, Version 7.8. Open Source Geospatial Foundation. Available online: <http://grass.osgeo.org> (accessed on 17 April 2020).
44. Croatian Bureau of Statistics—Structure of Agricultural Holdings. Available online: [https://www.dzs.hr/Hrv\\_Eng/Pokazatelj/Poljoprivreda%20-%20pregled%20po%20zupanijama.XLSX](https://www.dzs.hr/Hrv_Eng/Pokazatelj/Poljoprivreda%20-%20pregled%20po%20zupanijama.XLSX) (accessed on 20 December 2019).
45. Statistical Yearbook of the Republic of Croatia 2018. Available online: [https://www.dzs.hr/Hrv\\_Eng/ljetopis/2018/sljh2018.pdf](https://www.dzs.hr/Hrv_Eng/ljetopis/2018/sljh2018.pdf) (accessed on 21 December 2019).
46. GalićSubašić, D. Influence of Irrigation, Nitrogen Fertilization and Genotype on the Yield and Quality of Soybean (*Glycine max* (L.) Merr.). Ph.D. Thesis, Josip Juraj Strossmayer University of Osijek, Osijek, Croatia, 12 December 2019.
47. Agrotechnics of Soybean Cultivation (*Glycine max* L.). Available online: <https://www.savjetodavna.hr/wp-content/uploads/publikacije/AgrotehnikaSojeWeb102018.pdf> (accessed on 17 April 2020).
48. Gibson, K.E.; Yang, H.S.; Franz, T.; Eisenhauer, D.; Gates, J.B.; Nasta, P.; Farmaha, B.; Grassini, P. Assessing explanatory factors for variation in on-farm irrigation in US maize-soybean systems. *Agric. Water Manag.* **2018**, *197*, 34–40. [CrossRef]

49. Bagherzadeh, A.; Ghadiri, E.; Darban, A.R.S.; Gholizadeh, A. Land suitability modeling by parametric-based neural networks and fuzzy methods for soybean production in a semi-arid region. *Model. Earth Syst. Environ.* **2016**, *2*, 104. [CrossRef]
50. Munene, P.; Chabala, L.; Mweetwa, M. Land Suitability Assessment for Soybean (*Glycine max* (L.) Merr.) Production in Kabwe District, Central Zambia. *J. Agric. Sci.* **2017**, *9*, 1–16. [CrossRef]
51. He, Y.; Liu, D.; Yao, Y.; Huang, Q.; Li, J.; Chen, Y.; Shi, S.; Wan, L.; Yu, S.; Wang, D. Spatializing growth suitability for spring soybean cultivation in northeast China. *J. Appl. Meteorol. Climatol.* **2013**, *52*, 773–783. [CrossRef]
52. Seyedmohammadi, J.; Sarmadian, F.; Jafarzadeh, A.A.; Ghorbani, M.A.; Shahbazi, F. Application of SAW, TOPSIS and fuzzy TOPSIS models in cultivation priority planning for maize, rapeseed and soybean crops. *Geoderma* **2018**, *310*, 178–190. [CrossRef]
53. Saaty, T.L.; Ozdemir, M.S. Why the magic number seven plus or minus two. *Math. Comput. Model.* **2003**, *38*, 233–244. [CrossRef]
54. Huld, T. PVMAPS: Software tools and data for the estimation of solar radiation and photovoltaic module performance over large geographical areas. *Sol. Energy* **2017**, *142*, 171–181. [CrossRef]
55. Gašparović, I.; Gašparović, M.; Medak, D. Determining and analysing solar irradiation based on freely available data: A case study from Croatia. *Environ. Dev.* **2018**, *26*, 55–67. [CrossRef]
56. Kasten, F. The Linke turbidity factor based on improved values of the integral Rayleigh optical thickness. *Sol. Energy* **1996**, *56*, 239–244. [CrossRef]
57. Meteosat Solar Surface Radiation and Effective Cloud Albedo Climate Data Record SARAH-2, Validation Report. Available online: [https://www.cmsaf.eu/SharedDocs/Literatur/document/2016/saf\\_cm\\_dwd\\_val\\_meteosat\\_hel\\_2\\_1\\_pdf.pdf](https://www.cmsaf.eu/SharedDocs/Literatur/document/2016/saf_cm_dwd_val_meteosat_hel_2_1_pdf.pdf) (accessed on 22 March 2020).
58. Jantalia, C.P.; Resck, D.V.; Alves, B.J.; Zotarelli, L.; Urquiaga, S.; Boddey, R.M. Tillage effect on C stocks of a clayey Oxisol under a soybean-based crop rotation in the Brazilian Cerrado region. *Soil Till Res.* **2007**, *95*, 97–109. [CrossRef]
59. Mesic, M.; Birkás, M.; Zgorelec, Z.; Kisic, I.; Sestak, I.; Jurisic, A.; Husnjak, S. Soil Carbon Variability in Some Hungarian and Croatian Soils. In *Soil Carbon*; Hartemink, A., McSweeney, K., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 419–426. [CrossRef]
60. United States Department of Agriculture—Soil Survey Manual Handbook No. 18. Available online: <https://www.iec.cat/mapasols/DocuInteres/PDF/Llibre50.pdf> (accessed on 21 December 2019).
61. North Dakota State University—Soybean production. Available online: <https://library.ndsu.edu/ir/bitstream/handle/10365/5450/a250.pdf?sequence=1> (accessed on 22 March 2020).
62. Python Software Foundation. Python Language Reference, Version 3.7. Available online: <http://www.python.org> (accessed on 17 April 2020).
63. Böhner, J.; Selige, T. Spatial prediction of soil attributes using terrain analysis and climate regionalisation. In *SAGA—Analysis and Modelling Applications*; Böhner, J., McCloy, K.R., Strobl, J., Eds.; Verlag Erich Goltze: Goltze, Germany, 2006; pp. 13–27.
64. Oliver, M.A.; Webster, R. Geostatistical prediction: Kriging. In *Basic Steps in Geostatistics: The Variogram and Kriging*; Oliver, M.A., Webster, R., Eds.; Springer International Publishing: Berlin, Germany, 2015; pp. 43–69. [CrossRef]
65. Lu, G.Y.; Wong, D.W. An adaptive inverse-distance weighting spatial interpolation technique. *Comput. Geosci.* **2008**, *34*, 1044–1055. [CrossRef]
66. Hofstra, N.; New, M. Spatial variability in correlation decay distance and influence on angular-distance weighting interpolation of daily precipitation over Europe. *Int. J. Climatol.* **2009**, *29*, 1872–1880. [CrossRef]
67. Yao, X.; Yu, K.; Deng, Y.; Zeng, Q.; Lai, Z.; Liu, J. Spatial distribution of soil organic carbon stocks in Masson pine (*Pinus massoniana*) forests in subtropical China. *Catena* **2019**, *178*, 189–198. [CrossRef]
68. Qiao, P.; Lei, M.; Yang, S.; Yang, J.; Guo, G.; Zhou, X. Comparing ordinary kriging and inverse distance weighting for soil as pollution in Beijing. *Environ. Sci. Pollut. Res.* **2018**, *25*, 15597–15608. [CrossRef]
69. Xavier, A.C.; King, C.W.; Scanlon, B.R. Daily gridded meteorological variables in Brazil (1980–2013). *Int. J. Climatol.* **2016**, *36*, 2644–2659. [CrossRef]
70. Cruz-Cárdenas, G.; López-Mata, L.; Ortiz-Solorio, C.A.; Villaseñor, J.L.; Ortiz, E.; Silva, J.T.; Estrada-Godoy, F. Interpolation of Mexican soil properties at a scale of 1: 1,000,000. *Geoderma* **2014**, *213*, 29–35. [CrossRef]
71. Malczewski, J. *GIS and Multicriteria Decision Analysis*; John Wiley & Sons: New York, NY, USA, 1999; p. 408.
72. Kaykhosravi, S.; Abogadil, K.; Khan, U.T.; Jadidi, M.A. The Low-Impact Development Demand Index: A New Approach to Identifying Locations for LID. *Water* **2019**, *11*, 2341. [CrossRef]

73. Alamgir, M.; Mohsenipour, M.; Homsy, R.; Wang, X.; Shahid, S.; Shiru, M.S.; Alias, N.E.; Yuzir, A. Parametric Assessment of Seasonal Drought Risk to Crop Production in Bangladesh. *Sustainability* **2019**, *11*, 1442. [[CrossRef](#)]
74. Bonilla Valverde, J.P.; Blank, C.; Roidt, M.; Schneider, L.; Stefan, C. Application of a GIS Multi-Criteria Decision Analysis for the Identification of Intrinsic Suitable Sites in Costa Rica for the Application of Managed Aquifer Recharge (MAR) through Spreading Methods. *Water* **2016**, *8*, 391. [[CrossRef](#)]
75. Kumari, M.; Sakai, K.; Kimura, S.; Yuge, K.; Gunarathna, M. Classification of Groundwater Suitability for Irrigation in the Ulagalla Tank Cascade Landscape by GIS and the Analytic Hierarchy Process. *Agronomy* **2019**, *9*, 351. [[CrossRef](#)]
76. Bezdan, A.; Blagojevic, B.; Vranesevic, M.; Benka, P.; Savic, R.; Bezdan, J. Defining Spatial Priorities for Irrigation Development Using the Soil Conservation and Water Use Efficiency Criteria. *Agronomy* **2019**, *9*, 324. [[CrossRef](#)]
77. Novák, V. *Fuzzy Sets and Their Applications*; Taylor & Francis: London, UK, 1989; p. 190.
78. Domazetović, F.; Šiljeg, A.; Lončar, N.; Marić, I. Development of automated multicriteria GIS analysis of gully erosion susceptibility. *Appl. Geogr.* **2019**, *112*, 102083. [[CrossRef](#)]
79. Hamzeh, S.; Mokarram, M.; Haratian, A.; Bartholomeus, H.; Ligtenberg, A.; Bregt, A.K. Feature Selection as a Time and Cost-Saving Approach for Land Suitability Classification (Case Study of Shavur Plain, Iran). *Agriculture* **2016**, *6*, 52. [[CrossRef](#)]
80. Zhang, J.; Su, Y.; Wu, J.; Liang, H. GIS based land suitability assessment for tobacco production using AHP and fuzzy set in Shandong province of China. *Comput. Electron. Agric.* **2015**, *114*, 202–211. [[CrossRef](#)]
81. Klir, G.J. Fuzzy set theory. In *Uncertainty and Information: Foundations of Generalized Information Theory*; Klir, G.J., Ed.; Wiley-IEEE Press: Hoboken, NJ, USA, 2006; pp. 260–314.
82. Saaty, T.L. A scaling method for priorities in hierarchical structures. *J. Math. Psychol.* **1977**, *15*, 234–281. [[CrossRef](#)]
83. Musakwa, W. Identifying land suitable for agricultural land reform using GIS-MCDA in South Africa. *Environ. Dev. Sustain.* **2018**, *20*, 2281–2299. [[CrossRef](#)]
84. Islam, M.M.; Ahamed, T.; Noguchi, R. Land Suitability and Insurance Premiums: A GIS-based Multicriteria Analysis Approach for Sustainable Rice Production. *Sustainability* **2018**, *10*, 1759. [[CrossRef](#)]
85. Aldababseh, A.; Temimi, M.; Maghelal, P.; Branch, O.; Wulfmeyer, V. Multi-Criteria Evaluation of Irrigated Agriculture Suitability to Achieve Food Security in an Arid Environment. *Sustainability* **2018**, *10*, 803. [[CrossRef](#)]
86. Nigussie, G.; Moges, M.A.; Moges, M.M.; Steenhuis, T.S. Assessment of Suitable Land for Surface Irrigation in Ungauged Catchments: Blue Nile Basin, Ethiopia. *Water* **2019**, *11*, 1465. [[CrossRef](#)]
87. Saaty, T.L. *Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process*; RWS Publications: Pittsburgh, PA, USA, 2000; p. 478.
88. Chhetri, S.; Kayastha, P. Manifestation of an analytic hierarchy process (AHP) model on fire potential zonation mapping in Kathmandu Metropolitan City, Nepal. *ISPRS Int. J. Geo Inf.* **2015**, *4*, 400–417. [[CrossRef](#)]
89. Aydi, A.; Abichou, T.; Nasr, I.H.; Louati, M.; Zairi, M. Assessment of land suitability for olive mill wastewater disposal site selection by integrating fuzzy logic, AHP, and WLC in a GIS. *Environ. Monit. Assess.* **2016**, *188*, 59. [[CrossRef](#)] [[PubMed](#)]
90. Shahabi, H.; Keihanfard, S.; Ahmad, B.B.; Amiri, M.J.T. Evaluating Boolean, AHP and WLC methods for the selection of waste landfill sites using GIS and satellite images. *Environ. Earth sci.* **2014**, *71*, 4221–4233. [[CrossRef](#)]
91. Cui, X.; Dong, Y.; Gi, P.; Wang, H.; Xu, K.; Zhang, Z. Relationship between root vigour, photosynthesis and biomass in soybean cultivars during 87 years of genetic improvement in the northern China. *Photosynthetica* **2016**, *54*, 81–86. [[CrossRef](#)]
92. Vizzari, M.; Santaga, F.; Benincasa, P. Sentinel 2-Based Nitrogen VRT Fertilization in Wheat: Comparison between Traditional and Simple Precision Practices. *Agronomy* **2019**, *9*, 278. [[CrossRef](#)]
93. Bishop, K.A.; Betzelberger, A.M.; Long, S.P.; Ainsworth, E.A. Is there potential to adapt soybean (*Glycine max* Merr.) to future [CO<sub>2</sub>]? An analysis of the yield response of 18 genotypes in free-air CO<sub>2</sub> enrichment. *Plant. Cell Environ.* **2015**, *38*, 1765–1774. [[CrossRef](#)]
94. Sawilowsky, S.S. New effect size rules of thumb. *J. Mod. Appl. Stat. Methods* **2009**, *8*, 26. [[CrossRef](#)]



95. The Structure of the FAO Framework Classification. Available online: <http://www.fao.org/3/x5648e/x5648e0j.htm> (accessed on 22 March 2020).
96. Lolli, F.; Ishizaka, A.; Gamberini, R. New AHP-based approaches for multi-criteria inventory classification. *Int. J. Prod. Econ.* **2014**, *156*, 62–74. [[CrossRef](#)]
97. Foody, G.M.; Pal, M.; Rocchini, D.; Garzon-Lopez, C.X.; Bastin, L. The Sensitivity of Mapping Methods to Reference Data Quality: Training Supervised Image Classifications with Imperfect Reference Data. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 199. [[CrossRef](#)]
98. Bo, L.I.; Zhang, F.; Zhang, L.W.; Huang, J.F.; Zhi-Feng, J.I.N.; Gupta, D.K. Comprehensive suitability evaluation of tea crops using GIS and a modified land ecological suitability evaluation model. *Pedosphere* **2012**, *22*, 122–130. [[CrossRef](#)]
99. Neji, H.B.B.; Turki, S.Y. GIS—Based multicriteria decision analysis for the delimitation of an agricultural perimeter irrigated with treated wastewater. *Agric. Water Manag.* **2015**, *162*, 78–86. [[CrossRef](#)]
100. Gašparović, M.; Zrinjski, M.; Barković, Đ.; Radočaj, D. An automatic method for weed mapping in oat fields based on UAV imagery. *Comput. Electron. Agric.* **2020**, *173*, 105385. [[CrossRef](#)]
101. CLC2018 Technical Guidelines. Available online: [https://forum.eionet.europa.eu/nrc\\_land\\_covers/library/copernicus-2014-2020/pan-european-component/corine-land-cover-clc-2018/technical-guidelines/clc2018-technical-guidelines/download/en/2/CLC2018TechnicalGuidelines\\_final.pdf](https://forum.eionet.europa.eu/nrc_land_covers/library/copernicus-2014-2020/pan-european-component/corine-land-cover-clc-2018/technical-guidelines/clc2018-technical-guidelines/download/en/2/CLC2018TechnicalGuidelines_final.pdf) (accessed on 17 April 2020).
102. Rhebergen, T.; Fairhurst, T.; Zingore, S.; Fisher, M.; Oberthür, T.; Whitbread, A. Climate, soil and land-use based land suitability evaluation for oil palm production in Ghana. *Eur. J. Agron.* **2016**, *81*, 1–14. [[CrossRef](#)]
103. Gunnink, J.L.; Burrough, P.A. Interactive spatial analysis of soil attribute patterns using exploratory data analysis (EDA) and GIS. In *Spatial Analytical Perspectives on GIS*, 1st ed.; Ficher, M., Scholten, H., Unwin, D., Eds.; Taylor & Francis Ltd.: London, UK, 2019; pp. 87–100.
104. Gong, G.; Mattevada, S.; O’Bryant, S.E. Comparison of the accuracy of kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas. *Environ. Res.* **2014**, *130*, 59–69. [[CrossRef](#)]
105. Heuvelink, G.B.; Pebesma, E.J. Spatial aggregation and soil process modelling. *Geoderma* **1999**, *89*, 47–65. [[CrossRef](#)]
106. Vrieling, A.; de Beurs, K.M.; Brown, M.E. Variability of African farming systems from phenological analysis of NDVI time series. *Clim. Chang.* **2011**, *109*, 455–477. [[CrossRef](#)]
107. Dedeoğlu, M.; Dengiz, O. Generating of land suitability index for wheat with hybrid system approach using AHP and GIS. *Comput. Electron. Agric.* **2019**, *167*, 105062. [[CrossRef](#)]
108. Schipanski, M.E.; Drinkwater, L.E.; Russelle, M.P. Understanding the variability in soybean nitrogen fixation across agroecosystems. *Plant. Soil* **2010**, *329*, 379–397. [[CrossRef](#)]
109. Schultz, P.A.; Halpert, M.S. Global correlation of temperature, NDVI and precipitation. *Adv. Space Res.* **1993**, *13*, 277–280. [[CrossRef](#)]
110. Frampton, W.J.; Dash, J.; Watmough, G.; Milton, E.J. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS J. Photogramm. Remote Sens.* **2013**, *82*, 83–92. [[CrossRef](#)]
111. Ren, H.; Zhou, G.; Zhang, F. Using negative soil adjustment factor in soil-adjusted vegetation index (SAVI) for aboveground living biomass estimation in arid grasslands. *Remote Sens. Environ.* **2018**, *209*, 439–445. [[CrossRef](#)]
112. Jiang, H.; Eastman, J.R. Application of fuzzy measures in multi-criteria evaluation in GIS. *Int. J. Geogr. Inf. Sci.* **2000**, *14*, 173–184. [[CrossRef](#)]
113. Zabihi, H.; Alizadeh, M.; Kibet Langat, P.; Karami, M.; Shahabi, H.; Ahmad, A.; Nor Said, M.; Lee, S. GIS Multi-Criteria Analysis by Ordered Weighted Averaging (OWA): Toward an Integrated Citrus Management Strategy. *Sustainability* **2019**, *11*, 1009. [[CrossRef](#)]
114. Mokarram, M.; Hojati, M. Using ordered weight averaging (OWA) aggregation for multi-criteria soil fertility evaluation by GIS (case study: Southeast Iran). *Comput. Electron. Agric.* **2017**, *132*, 1–13. [[CrossRef](#)]
115. Gašparović, M.; Zrinjski, M.; Gudelj, M. Automatic cost-effective method for land cover classification (ALCC). *Comput. Environ. Urban Syst.* **2019**, *76*, 1–10. [[CrossRef](#)]





## **CHAPTER 2**

---

Delineation of Soil Texture Suitability Zones for Soybean Cultivation: A Case  
Study in Continental Croatia

Article

# Delineation of Soil Texture Suitability Zones for Soybean Cultivation: A Case Study in Continental Croatia

Dorijan Radočaj , Mladen Jurišić, Vladimir Zebec and Ivan Plaščak 

Faculty of Agrobiotechnical Sciences Osijek, Josip Juraj Strossmayer University of Osijek, Vladimira Preloga 1, 31000 Osijek, Croatia; mjurisic@fazos.hr (M.J.); vladimir.zebec@fazos.hr (V.Z.); iplascak@fazos.hr (I.P.)

\* Correspondence: dradocaj@fazos.hr; Tel.: +385-31-554-879

Received: 10 May 2020; Accepted: 8 June 2020; Published: 10 June 2020



**Abstract:** Soil texture is a vital criterion in most cropland suitability analyses, so an accurate method for the delineation of soil texture suitability zones is necessary. In this study, an automated method was developed and evaluated for the delineation of these zones for soybean cultivation. A total of 255 soil samples were collected in the Continental biogeoregion of Croatia. Three methods for interpolation of clay, silt and sand soil content were evaluated using the split-sample method in five independent random repetitions. An automated algorithm for soil texture classification based on the United States Department of Agriculture (USDA) in 12 classes was performed using Python script. Suitability classes for soybean cultivation per soil texture class were determined according to previous agronomic and soybean land suitability studies. Ordinary kriging produced the highest accuracy of tested interpolation methods for clay, silt and sand. Highly suitable soil texture classes for soybean cultivation, loam and clay loam, were detected in the northern part of the study area, covering 5.73% of the study area. The analysis of classification results per interpolation method indicated a necessity of the evaluation of interpolation methods as their performance depended on the normality and stationarity of input samples.

**Keywords:** automation; classification; interpolation; GIS; python script; kriging; split-sample validation

## 1. Introduction

Knowledge of naturally suitable locations for the cultivation of a particular crop type is a basis for effective natural resource management and sustainable agriculture [1]. The development of spatial models of cropland suitability analyses integrated with geographic information system (GIS) enabled decision-makers to develop a crop management system that increases yield and land productivity [2]. The selection of the relevant criteria is based on crop and site-specific requirements, typically compiled from previous studies or crop expert surveys in the study area. Land suitability analysis commonly integrates climate and soil criteria groups of specific land use [3]. Soil texture presents a vital criterion in the GIS-based soybean suitability analyses for the determination of locations with the highest potential yield increase [4,5]. Many soil physical and chemical properties are directly influenced by soil texture, affecting soil fertility and productivity [6,7]. Recent research proved the direct influence of soil texture on the important segments of the ecosystem, such as agricultural yield variability [8], soil microbial and structural recovery [9] and richness and diversity of soil bacterial communities [10]. Automation of classification algorithms allows more time-efficient and robust classification in comparison to the conventional manual approach [11]. Recently developed algorithms for soil classification cover a wide range of applications, most notably measurement of soil surface roughness [12], prediction of soil fertility [13] and estimation of degraded soils [14]. In addition, mapping of soil texture variability

in both horizontal and vertical dimensions has been a subject of interest in previous studies [15,16]. The mapping of the GIS analysis results presents an effective tool in their distribution to the end-users, both for farmers on a large scale, as well as for land policy managers on a small (regional) scale. Such data representation does not require extensive knowledge of spatial data and therefore allows simple interpretation, especially when integrated into web GIS maps.

Interpolation overcomes the limited representation of the agricultural soil properties by discrete soil sampling. Instead, it incorporates these ground-truth data to predict soil properties and enable a continuous representation of the entire agricultural field [17]. Interpolation methods are generally divided into geostatistical and deterministic methods. According to [18], geostatistical interpolation methods model the uncertainty associated with a spatial estimation based on random function theory, while the deterministic methods use only the input sample values and mathematical functions for interpolation. Geostatistical methods also apply statistical functions to quantify the predictions' uncertainty, in contrast to deterministic interpolation methods. The basic step in the implementation of the geostatistical interpolation methods is the variogram analysis for the fitting of the most suitable theoretical mathematical model to an empirical variogram [19]. The evaluation of the interpolation methods enables the selection of the optimal, most accurate interpolation method for a particular soil sample [20]. The interpolation accuracy of soil samples has a direct impact on the creation of land-use management plans and variable-rate applications in precision agriculture [21].

Soybean (*Glycine max* L.) is the main oilseed crop in the world, presenting a high-quality source of oil and protein [22,23]. The long-term projection by [24] predicts a 1.3% global annual increase of soybean yield, which does not meet the food demands in the world. The development of soybean suitability analyses is therefore necessary for the improvement of soybean yield, especially to land policy managers [25]. A study by [26] noted the importance of soybean introduction in the cropping system, allowing the effective exploitation of soil nutrients and the overall reduction of pesticides and herbicides in precision agriculture. Soil texture is proven to have a significant impact on soybean cultivation and yield [27,28]. Improvement of research regarding soybean yield enhancement and sustainability is essential to satisfy growing soybean global demands [29].

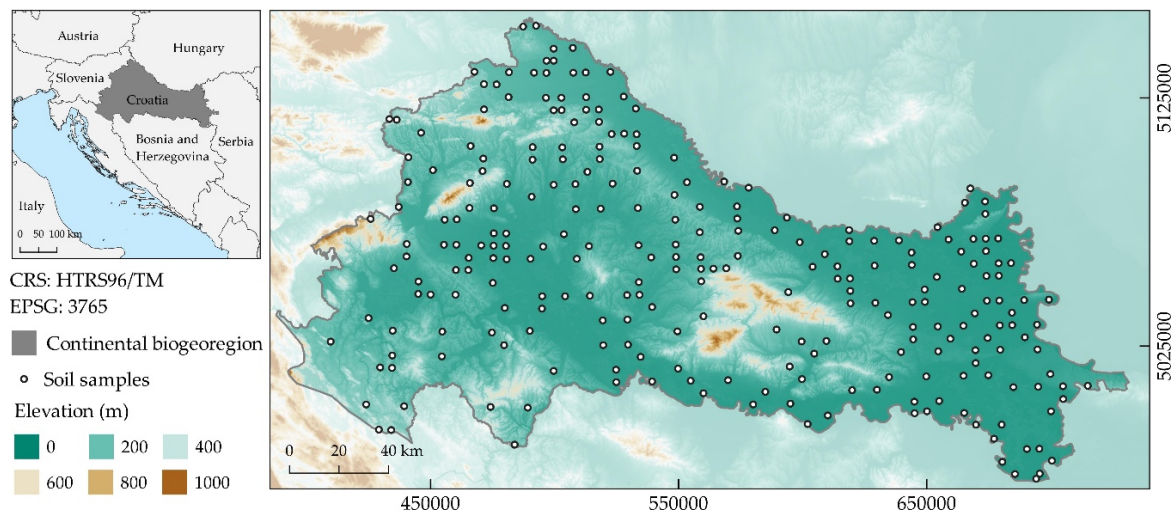
The purpose of this study was to delineate soil texture suitability zones for soybean cultivation using an automated soil texture classification algorithm in a GIS environment. The specific objectives of the study were to: (1) evaluate and select the most accurate method for the interpolation of clay, silt and sand from input samples, (2) automate the process of soil texture classification in a GIS environment and (3) provide a map of soil texture suitability zones for soybean cultivation at a regional scale.

## 2. Materials and Methods

### 2.1. Study Area

The study area is the Continental biogeoregion of Croatia, covering 30,863 km<sup>2</sup> and being one of the three national biogeoregions defined by the European Environment Agency [30] (Figure 1). The other two biogeoregions present in Croatia are alpine and Mediterranean biogeoregions. The study area is characteristically a moderately warm, rainy climate, as per Köppen's classification. Soybean is gradually becoming one of the most produced crops in Croatia, with a 119% yield production increase between 2013 and 2016 as per Croatian Bureau of Statistics data. The terrain is dominantly lowland, at lower elevations. According to CORINE 2018 land cover data, agricultural areas represent the major land cover class with 54.0% of the study area. The second-largest land cover classes are forests and seminatural areas with 40.7% of the study area, followed by artificial surfaces (3.5%), water bodies (1.4%) and wetlands (0.5%).



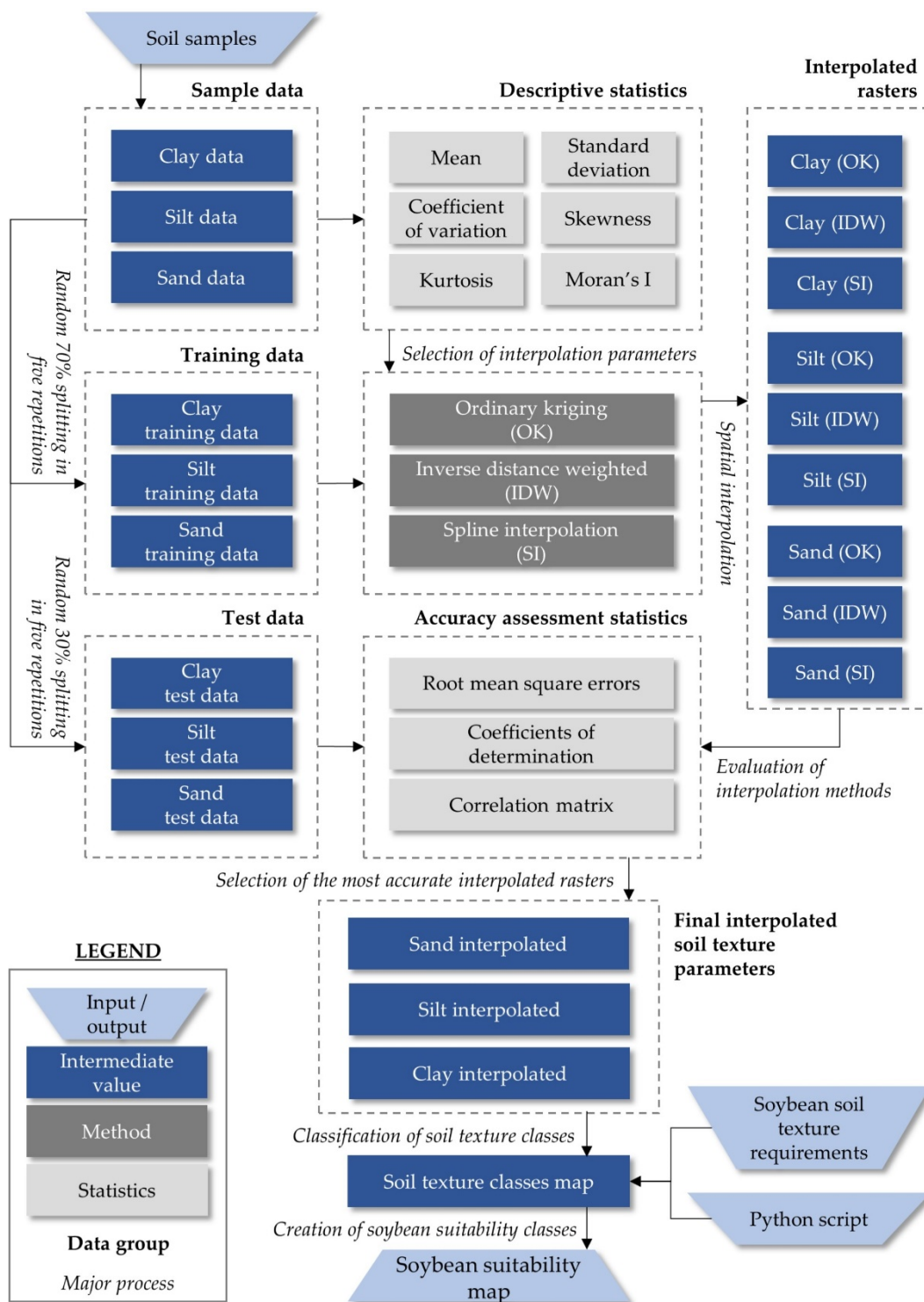


**Figure 1.** Continental biogeoregion of Croatia.

## 2.2. Automated Classification for Delineation of Soil Texture Suitability Zones for Soybean Cultivation in a GIS Environment

The selected interpolation methods for evaluation in this research were ordinary kriging (OK), inverse distance weighted (IDW) and spline interpolation (SI). These methods were recommended for the interpolation of soil properties by [31]. The workflow of this research is shown in Figure 2. All spatial calculations in this study were conducted using open-source GIS software: SAGA GIS v7.4.0 for the samples pre-processing and spatial interpolation and QGIS v3.8.3 for interpolation results and soil texture map visualization. Python v3.7.4 was used for scripting as a base for the automation of soil texture classification. The selected coordinate reference system (CRS) was the Croatian Terrestrial Reference System (HTRS96/TM). The spatial resolution of interpolated rasters was set to 250 m.

Input soil sample data was downloaded using the Croatian Agency for the Environment and Nature soil data WFS service. Soil samples were collected in the field according to ISO 11272:2017 and were pre-treated according to ISO 11464:2006 [32]. A total of 255 samples were collected in the study area during the year 2016. The soil texture content was determined by sieving and the sedimentation method prescribed in ISO 11277:2009. Soil fractions (<2 mm) were reclassified to three classes regarding particle size: clay (<0.002 mm), silt (0.002–0.063 mm) and sand (>0.063 mm). Each soil sample collected in the field represented a set of clay, silt and sand data on the same location. These values were filtered by attributes from input samples into a separate vector point layer in a GIS environment. The split-sample method was applied for the creation of training and test datasets, randomly splitting input samples to training data (70%) and test data (30%). From the total of 255 collected soil samples in the study area, 179 training samples and 76 test samples were randomly split from the original set individually for each repetition. The split-sample method was performed in five independent repetitions, resulting in five training and five corresponding test sample sets. Descriptive statistics consisting of mean, standard deviation (SD), coefficient of variation (CV), skewness (SK) and kurtosis (KT) were calculated to quantify normality and stationarity of the training sets. These properties are regarded as necessary for the determination of interpolation parameters for each method [33]. Normality designated the distribution of sample values, quantifying its deviation to normal distribution. Stationarity represented the number of local variations in the sample values, where low local variations indicate higher stationarity. Input data normality and stationarity are regarded as necessary indicators for the selection of optimal interpolation parameters [34].



**Figure 2.** Research workflow for the delineation of soil texture suitability zones for soybean cultivation.

OK is the most commonly used geostatistical interpolation method and is thoroughly described in [19,35]. It is also considered as the best linear unbiased spatial predictor [36]. OK incorporates a

variogram as a tool for the modelling of spatial dependence between the input sample values and the mutual distance of samples [37]. The variogram was modelled based on (1):

$$\gamma(h) = \frac{1}{2N} \sum_{x=1}^N [Z(x) - Z(x+h)]^2, \quad (1)$$

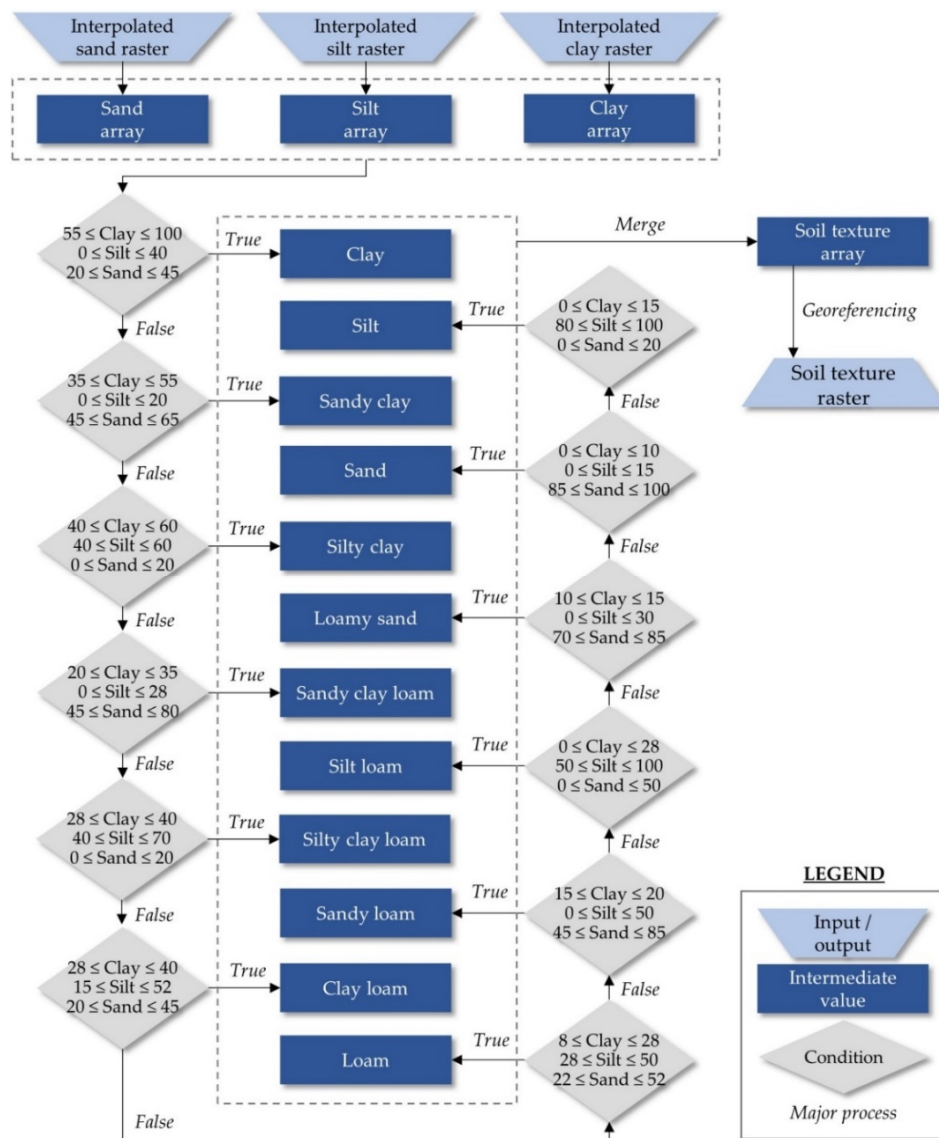
where  $\gamma(h)$  is the semivariance,  $N$  is the number of sample pairs and  $Z(x)$  and  $Z(x+h)$  are the observation values at point  $x$  and at a point separated by  $h$ .

IDW is a deterministic interpolation method that predicts the values based on a weighted average of the input samples. The weights were assigned to each sample in an inversely proportional way to the distance from the predicted location [38]. The prediction principle of IDW is outlined in detail in [39].

SI is the deterministic interpolation method that performs the spatial interpolation by using the polynomial functions for describing the components of the interpolated area. Mutual fitting of those functions is performed, so that they join smoothly, consequentially resulting in the smooth representation of the interpolated area [17]. SI performs best in the case of gently varying surfaces, which is commonly associated with samples with high stationarity [31].

The three selected interpolation methods were applied to five randomly created training sets, resulting in 15 interpolation variations separately for clay, silt and sand soil content. The selection of interpolation parameters for all three interpolation methods was performed according to the evaluated normality and stationarity of the training sets. The existence and the range of the spatial variability of training samples and full samples were observed using Moran scatterplots, Moran's  $I$  values and autocorrelograms [40]. All three tested interpolation methods were implemented in SAGA GIS software. OK was conducted with logarithmic transformation in the pre-processing due to lack of normal distribution of sample values for all soil parameters, as proposed by [41]. The starting parameters used for OK interpolation of training sets and final interpolation using full sample sets were a range of 170,000 m with 25 lags, with each lag covering a distance of 6800 m. Tested theoretical mathematical models for fitting to the empirical variogram were linear, square root, logarithmic, Gaussian and spherical. The selection of these models and corresponding ranges was performed based on the highest fitting coefficient of determination to the empirical variogram in ranges with existing spatial autocorrelation. The interpolation by the IDW method was performed by using the power parameter of 3 to prevent the emphasis of local variations in the interpolation results, due to moderate stationarity of training sets. SI was conducted by the thin plate spline process to ensure the smoothing of the interpolated surface to decrease the impact of higher local variabilities of the input samples [42]. The ranges of deterministic interpolation methods were selected based on observed spatial autocorrelation, amounting to maximum distances with continuous positive autocorrelation values.

Final clay, silt and sand soil content raster were interpolated with the full input sample data, using the most accurate interpolation method determined during accuracy assessment. Soil texture classification was performed in 12 classes according to the specifications of the United States Department of Agriculture (USDA) [43]. The classification process was automated using the algorithm realized in Python script (Figure 3, Appendix A). The automation in the GIS environment was developed to perform classification robustly and independently of the CRS, location extent, or spatial resolution of input rasters.



**Figure 3.** An automatic algorithm for soil texture classification in the GIS environment according to USDA specifications.

### 2.3. Accuracy Assessment of Interpolated Rasters

Accuracy assessment of the selected interpolation methods was conducted by the calculation of coefficients of determination ( $R^2$ ) and root-mean-square error ( $RMSE$ ) for all interpolated rasters. These statistical values are considered as complementary for the evaluation of interpolation results for soil parameters [44]. The  $R^2$  quantified fitting of interpolated values to the regression line, while  $RMSE$  was used to assess variation in data based on absolute fit. The  $R^2$  values were calculated using a linear regression model. The higher  $R^2$  and lower  $RMSE$  indicate a better model fitting. Formulas for the calculation of  $R^2$  (2) and  $RMSE$  (3) were [45]:

$$R^2 = 1 - \frac{\sum_1^n (y_i - \hat{y}_i)^2}{\sum_1^n (y_i - \bar{y}_i)^2} \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_1^n (y_i - \hat{y}_i)^2} \tag{3}$$

where  $y_i$  is the measured (sampled) value,  $\hat{y}_i$  is the predicted value,  $\bar{y}_i$  is the average of the measured value  $y$ , and  $n$  is the number of observations.

#### 2.4. Determination of Soybean Suitability Zones According to Soil Texture Classes

Loamy soil is the most suitable for the majority of crops, generally containing more organic soil matter and retaining moisture and nutrients better than other soil textures [46]. Levels of the suitability of 12 soil texture classes for soybean cultivation were evaluated based on previous soybean-specific agronomic and land suitability studies. Soybean-specific agronomic studies that evaluated the effects of soil texture for various applications in soybean cultivation are shown in Table 1. All evaluated studies noted a higher level of suitability for loamy textures compared to other classes present in the study area. Loamy soil textures with higher silt or clay content were regarded as more suitable compared to higher sand content, such as sandy loam or loamy sand. Recent studies based on GIS-based multicriteria land suitability for soybean cultivation containing soil texture as an influential criterion are shown in Table 2. The suitability of soil texture classes was natively represented according to Food and Agriculture Organization (FAO) specifications [47] or adjusted to this specification from original classes. FAO specifications for land suitability consist of five suitability zones: highly suitable (S1), moderately suitable (S2), marginally suitable (S3), marginally not suitable (N1) and permanently not suitable (N2). Similar to agronomic studies, loamy textures were considered as the most suitable, with clay loam and loam being regarded as highly suitable. Suitability levels generally decreased with the increase of clay, silt or sand soil contents. The final classification of soil texture suitability zones for soybean cultivation was performed according to FAO specifications and is presented in Figure 4.

**Table 1.** Soil texture suitability for various applications from previous soybean-specific studies.

Source	Soil Texture Classes in the Study	Application
[48]	loamy sand, sandy loam, <b>silt loam</b>	Nitrous oxide emission in agricultural fields
[49]	loamy sand, <b>sandy clay loam</b>	Cyst nematode population density
[50]	sandy loam, <b>silty clay loam</b>	Yield response to salinity
[51]	<b>loamy textures</b> , sandy textures, clay textures	Effect of root zone temperatures on interorganismal signal molecules
[52]	<b>loamy textures</b> , clay textures, sandy textures	Carbon and nitrogen mineralization
[53]	clay, <b>sandy clay loam</b>	Canopy dry matter production
[54]	<b>sandy loam, silt loam, loam</b> , sandy clay loam, clay loam, silty clay loam, clay	Incidence of brown stem rot in conservation-till fields

The most suitable soil texture classes per study were bolded.

**Table 2.** Soil texture suitability for soybean cultivation from previous GIS-based multicriteria analysis studies.

Source	Suitability Level for Soybean Cultivation				
	S1	S2	S3	N1	N2
[55]	loam	sandy loam, clay loam, silt loam	sandy clay, silty clay	other classes	sand
[56]	loam, clay loam, silty clay loam, silt loam	sandy loam, clay, silty clay, silty clay loam	loamy sand, silt	sand	
[57]	silt loam, silty clay loam, loam	sandy clay loam	sandy loam	/	/
[58]	silty clay loam, sandy clay loam, clay loam	loam, silt, silty loam, sandy clay, silty clay	sandy loam, clay	loamy sand	sand
[59]	clay loam	sandy clay loam	loamy sand	sandy loam	/



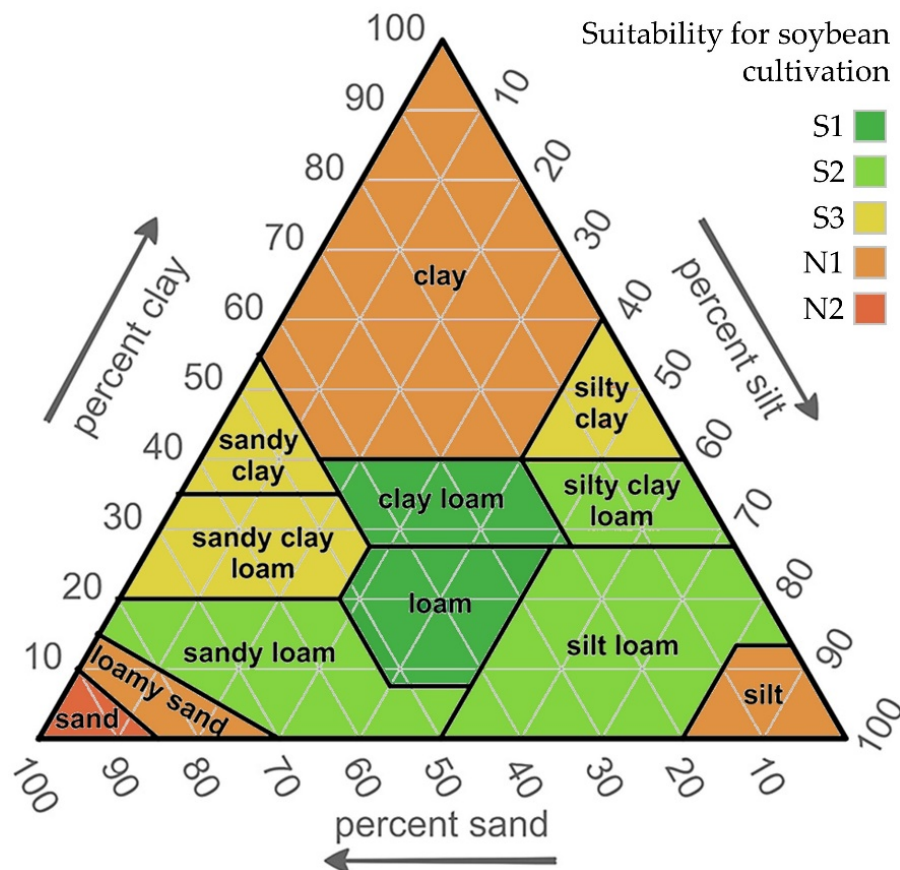


Figure 4. Soil texture suitability zones for soybean cultivation according to USDA soil texture classification.

### 3. Results

Descriptive statistics used for the estimation of training sample sets' normality and stationarity are shown in Table 3. Silt soil content was observed as the dominant soil texture parameter in the study area, followed by clay. Clay and silt soil content data resulted in similar CV values through all five repetitions, representing low-variance data. Sand soil content produced CV values higher than 1.100 in all repetitions, indicating a high-variance data. SK remained consistent for all three soil parameters through repetitions. Clay and silt soil content resulted in right-tailed and left-tailed moderate skewness, respectively. Sand soil content consistently produces the highest SK values of all soil parameters, indicating a highly skewed data. KT resulted in the most variability of the calculated descriptive statistics for all soil parameters. However, values in all repetitions indicated a considerable platykurtic distribution for clay and silt soil contents and a considerable leptokurtic distribution for sand soil content. Sand soil content data values indicate a highly skewed distribution and moderate data stationarity.

The Moran scatterplot per training set is shown in Figure 5. Positive spatial autocorrelation was observed for all three soil parameters, with mean Moran's I resulting in 0.396, 0.259 and 0.333 for clay, silt and sand, respectively. Interpolation ranges for IDW and SI per training sample set are presented in Appendix B. Mean distances until negative autocorrelation values were 81,600 m for clay, 29,920 m for silt and 54,400 m for sand. Variograms and interpolation parameters using OK for five training sample sets are presented in Appendix B. Square root and Gaussian theoretical mathematical models allowed the best fitting on most occasions, being applied 6 times each.

Table 3. Descriptive statistics of random training sample sets.

Set No.	Soil Parameter	Mean (%)	CV	SK	KT
Set 1	Clay	30.997	0.368	0.678	0.228
	Silt	58.141	0.223	−0.760	0.833
	Sand	10.862	1.207	2.322	6.515
Set 2	Clay	31.355	0.332	0.584	0.336
	Silt	57.945	0.211	−0.599	0.391
	Sand	10.701	1.150	2.065	5.093
Set 3	Clay	29.889	0.343	0.581	0.159
	Silt	57.958	0.225	−0.800	0.874
	Sand	12.153	1.141	2.060	4.861
Set 4	Clay	29.899	0.354	0.633	0.334
	Silt	58.276	0.225	−0.805	0.860
	Sand	11.826	1.173	2.116	5.000
Set 5	Clay	30.795	0.353	0.594	0.241
	Silt	57.305	0.228	−0.688	0.674
	Sand	11.900	1.151	2.082	5.050
Mean	Clay	30.587	0.350	0.614	0.260
	Silt	57.925	0.222	−0.730	0.726
	Sand	11.488	1.164	2.129	5.304

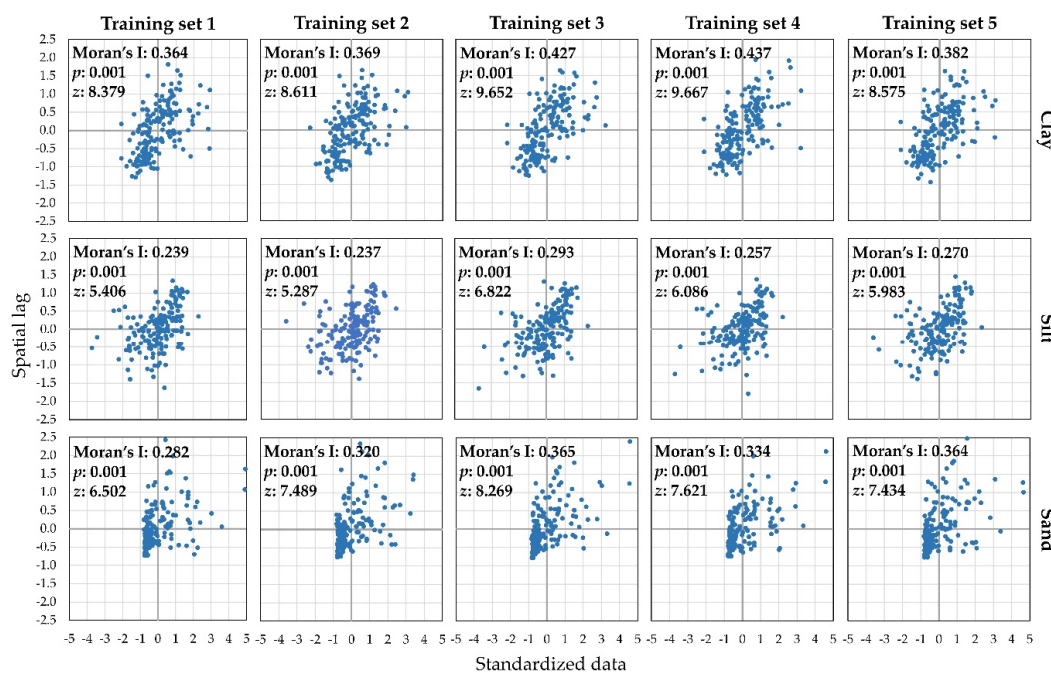


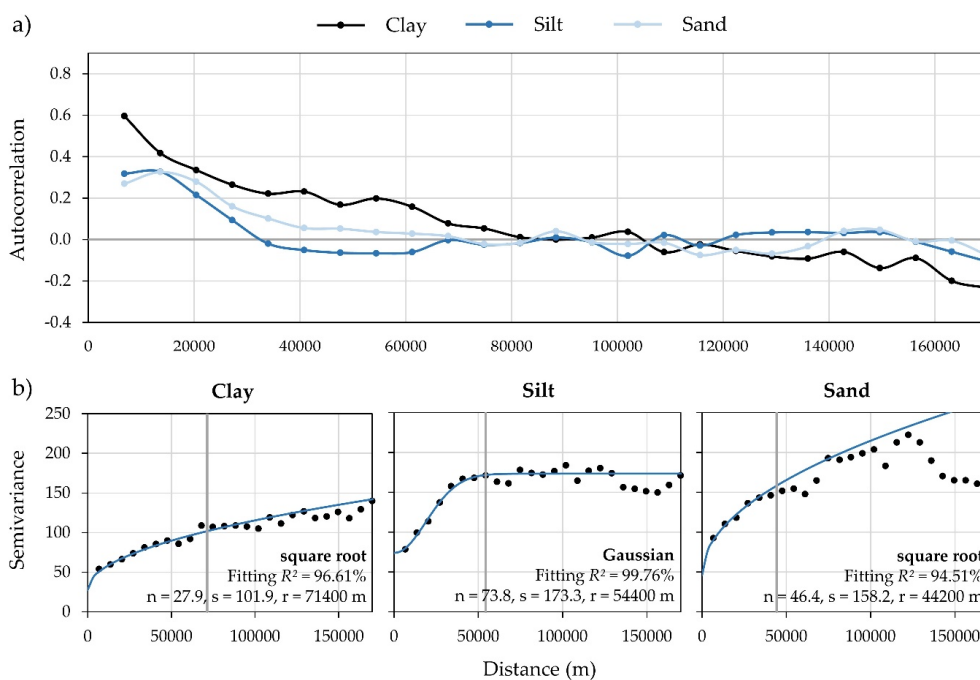
Figure 5. Moran scatterplots for five training sample sets for clay, silt and sand.

OK resulted in the highest mean *RMSE* and  $R^2$  values for all soil parameters and was determined as the most accurate interpolation method in this study (Table 4). It was followed by IDW as the second most accurate method according to both statistical values, while SI was the least accurate interpolation method. The interpolation results within the same soil parameter and interpolation methods generally produced high variability in accuracy values. Most notably, the difference between maximum and minimum values for *RMSE* was 2.26 for clay interpolation using SI, while  $R^2$  was 0.353 for the interpolation of clay using IDW. All three interpolation methods produced the highest accuracy for the interpolation of sand, followed by clay and silt soil contents.

**Table 4.** Accuracy assessment of the evaluated interpolation methods per soil parameter.

Set No.	Stat. Value	OK			IDW			SI		
		Clay	Silt	Sand	Clay	Silt	Sand	Clay	Silt	Sand
Set 1	RMSE	1.85	3.04	1.79	3.13	2.91	2.51	3.50	4.80	3.79
	R <sup>2</sup>	0.786	0.632	0.740	0.456	0.560	0.645	0.490	0.328	0.444
Set 2	RMSE	2.01	1.80	1.52	2.51	2.06	2.68	3.19	4.15	2.90
	R <sup>2</sup>	0.733	0.727	0.846	0.720	0.703	0.680	0.475	0.437	0.597
Set 3	RMSE	2.70	3.59	1.60	3.00	3.28	3.28	5.50	3.85	3.77
	R <sup>2</sup>	0.655	0.469	0.842	0.598	0.422	0.537	0.313	0.385	0.497
Set 4	RMSE	2.78	3.08	2.45	2.75	2.58	2.62	3.24	3.27	2.61
	R <sup>2</sup>	0.538	0.630	0.673	0.552	0.632	0.637	0.441	0.472	0.528
Set 5	RMSE	2.13	3.98	1.70	1.81	3.14	2.42	4.03	3.66	2.40
	R <sup>2</sup>	0.788	0.503	0.820	0.809	0.528	0.710	0.446	0.392	0.580
Mean	RMSE	2.29	3.10	1.81	2.64	2.79	2.70	3.89	3.94	3.09
	R <sup>2</sup>	0.700	0.592	0.784	0.627	0.569	0.642	0.433	0.403	0.529

Autocorrelogram and OK interpolation parameters of full sample sets are shown in Figure 6. Deterministic interpolation methods were performed with maximum ranges of 81,600 m for clay, 27,200 m for silt and 68,000 m for sand, matching maximum distances with positive autocorrelation. The specific local variabilities of the interpolation results were analysed according to Figure 7. OK produced interpolation results with low extreme values, compared to IDW and SI results. Consequentially, local variability in all soil fractions was minimized in interpolation results. Generalization of the local variabilities for OK is particularly noticeable in the south-west part of the study area for sand soil content interpolation. IDW results displayed the effects of all local variabilities to some degree but restricted their impact on interpolated results on local surroundings of the training samples. SI was the most balanced interpolation method considering the effects of local variabilities. The effects of local variations were retained in the interpolation results, having higher impacts on the surroundings of the training samples than IDW.



**Figure 6.** Properties of full sample sets for clay, silt and sand: (a) autocorrelogram, (b) variograms and OK interpolation parameters (n: nugget, s: sill, r: range).



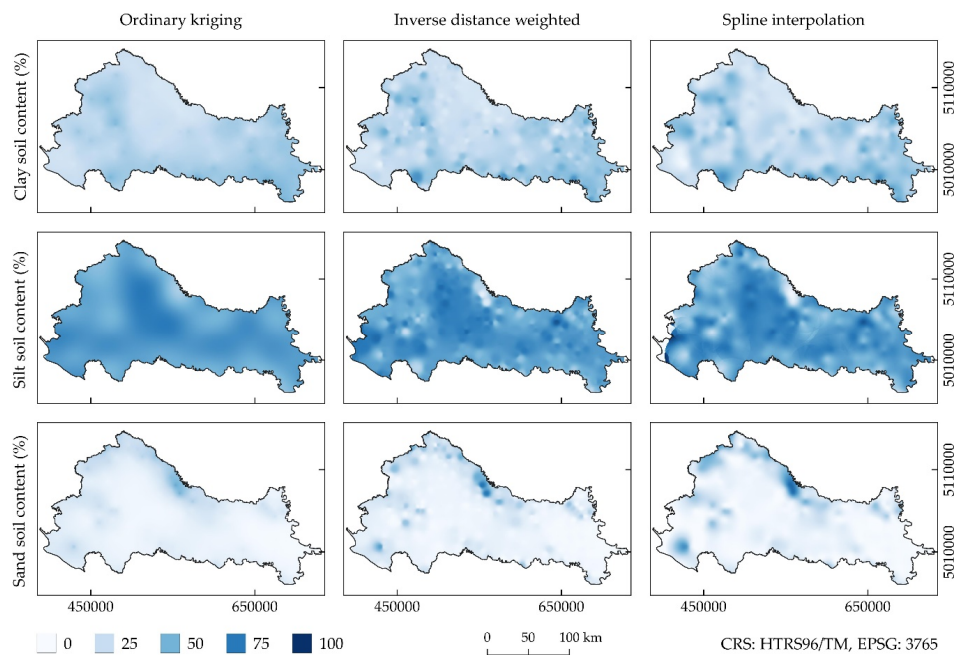


Figure 7. Interpolation results for evaluated interpolation methods using full sample sets.

Histograms of interpolation results enabled quantification of the effect of local variabilities in the interpolation results (Figure 8). IDW resulted in the narrowest value interval of the tested interpolation method while having the top frequency value for every soil parameter. In contrast, SI produced the most variability regarding the interpolation values interval. Standard deviations of sand interpolation resulted in the greatest variability, with 7.91 for OK, 9.90 for IDW and 12.80 for SI.

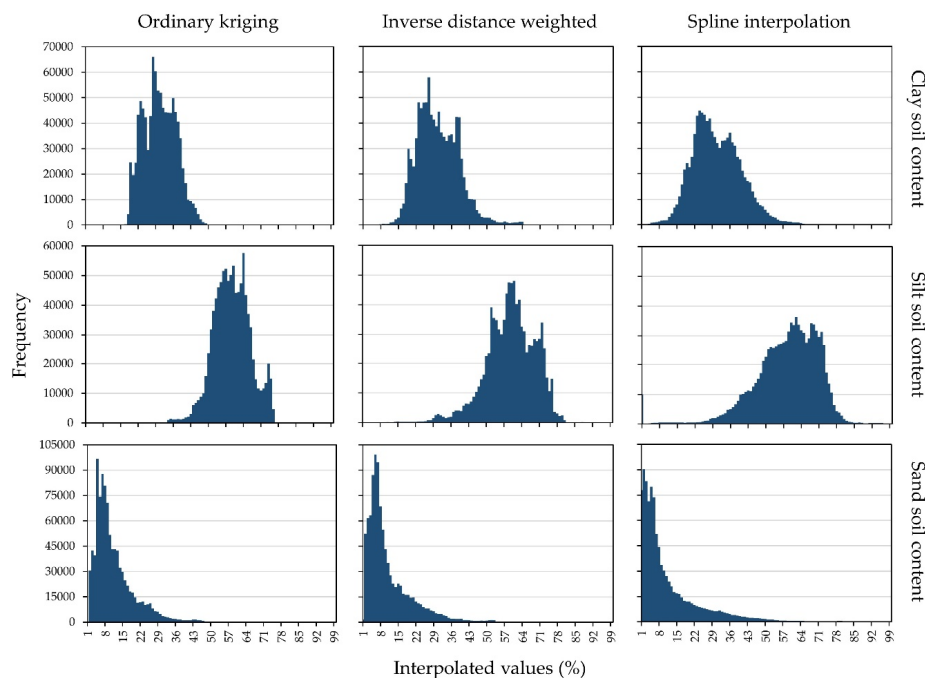


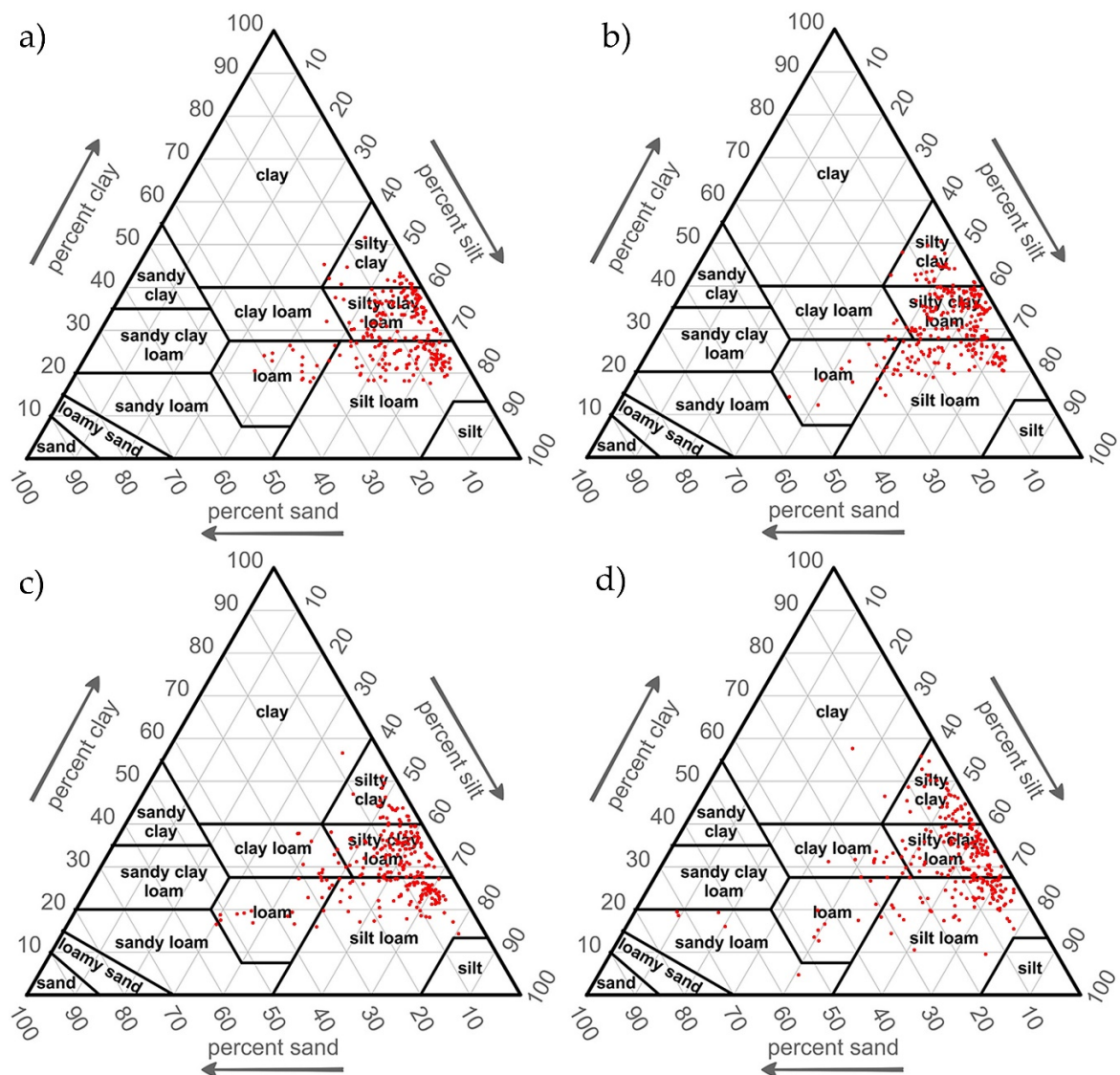
Figure 8. Histograms of interpolation results per soil texture parameter using full sample sets.

The correlation matrix of tested interpolation methods for the respective soil fractions is displayed in Table 5. The maximum correlation was achieved for silt interpolation between IDW and SI. The lowest correlation was observed between OK and SI for both clay and sand interpolation. The visualization

of classified soil texture in soil texture triangles is performed in Figure 9. OK produced the lowest dispersion of soil texture classes, dominantly covering silty clay loam and silt loam. IDW and SI covered a slightly wider range of soil texture classes. This observation particularly refers to loam, which is present in ground truth data (255 soil samples), but under-classified by OK. Both deterministic methods showed a notable presence of results in silty clay and clay loam classes, however, these classes had a low presence in ground truth data.

**Table 5.** Correlation matrix of interpolation results using Pearson’s coefficient of correlation.

	Clay			Silt			Sand		
	OK	IDW	SI	OK	IDW	SI	OK	IDW	SI
OK	1.000			1.000			1.000		
IDW	0.869	1.000		0.826	1.000		0.818	1.000	
SI	0.775	0.868	1.000	0.803	0.914	1.000	0.726	0.869	1.000

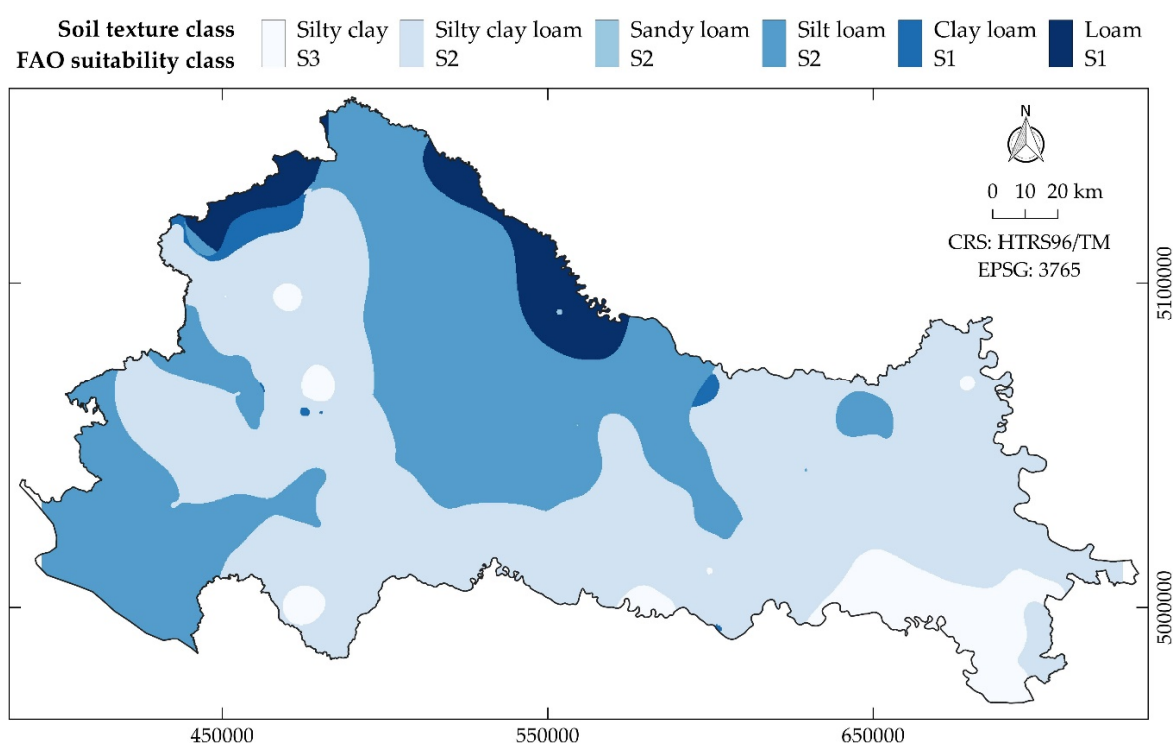


**Figure 9.** Soil texture triangles classified from: (a) ground truth data, (b) OK interpolation, (c) IDW interpolation, (d) SI interpolation.

The interpolation of final clay, silt and sand soil fractions was performed by OK, as the most accurate interpolation method, with all 255 input samples. The results of the soil texture classification using Python script are displayed by coverage per soil texture class (Table 6). Loam and clay loam, as the two most suitable soil texture zones for soybean cultivation, were detected in the northern part of the study area, mostly covering the border area near the river Drava. Silt loam and silty clay loam, the two moderately suitable zones, covered 88.22% of the study area. Silty clay, as the only marginally suitable zone, was mostly present in the south-east on the study area. The soybean suitability thematic map according to soil texture classes is shown in Figure 10.

**Table 6.** Coverage per soil texture class in the study area.

Coverage	Silty Clay	Silty Clay Loam	Clay Loam	Loam	Sandy Loam	Silt Loam
Area (%)	6.04	53.26	0.99	4.74	0.01	34.96
Area (km <sup>2</sup> )	1861	16,412	306	1460	2	10,773



**Figure 10.** Thematic map of soybean suitability according to soil texture classes.

#### 4. Discussion

The first objective of the study regarding the selection of the most accurate interpolation method was fulfilled by accuracy assessment based on five independent random split samples. Clay and silt possessed values moderately deviated from a normal distribution and moderate data stationarity, while sand possessed highly skewed values. Previous studies [60–63] suggested that these cases are common in analyses of soil properties, so the transformation of skewed data was determined as a necessary procedure in the algorithm. The application of logarithmic transformation in the pre-processing to OK minimized the effect of skewed distribution, as suggested by [64]. The descriptive statistics of clay, silt and sand in this study suggest that the thresholds for performing logarithmic transformation could be  $CV > 0.500$  as the first condition, and the second condition that SK and KT deviate 0.500 or more from values of the normal distribution sample of zero and three, respectively. However, it is necessary to test this assumption on the different soil samples to evaluate its reliability. Similar studies imply the applicability of statistical tests like the Shapiro–Wilk [65] or

Kolmogorov–Smirnov test [66] for the evaluation of data normality and could present an upgrade to the soil classification algorithm in future research. Sparse sampling density of one sample per 118 km<sup>2</sup> used in this study is expected to have an impact on overall interpolation accuracy results. Denser soil sampling is expected to result in the lower CV of soil parameter values [67] and more accurate interpolation results [68,69]. However, the mapping of soil properties on a large scale was successfully performed using low-density samples, using one sample per 199 km<sup>2</sup> in [70] and even one sample per 443 km<sup>2</sup> in [71]. Due to the structure of the workflow and automatic soil texture classification algorithm in a GIS environment, a similar performance is expected by adding more samples in the existing area or using the same number of samples on a smaller area. Slightly more time-expensive computation in these cases is the only notable difference expected, as more input data should be processed.

The highest fitting  $R^2$  values for OK were achieved on shorter distances with full sample sets (71,400 m, 54,400 m and 44,200 m, compared to 84,320 m, 69,360 m and 119,680 m) in contrast to mean values of five training sample sets, which correspond better to observed spatial autocorrelation. OK was determined as the most accurate interpolation method for all three soil fractions, with  $R^2$  ranging from 0.592 to 0.784 and *RMSE* from 1.81 to 3.10. IDW was the second most accurate method, resulting in the most uniform accuracies through the interpolation repetitions. Moderate interpolation accuracies imply the necessity for denser soil samples in soil texture classification on a national level [72]. Sampling using regular grids proved in most cases to be the optimal suitable field sampling method [73]. The evaluation of interpolation methods displayed the importance of testing multiple methods. Mean  $R^2$  differences between OK as the most accurate and SI as the least accurate methods were 0.267 for clay, 0.189 for silt and 0.255 for sand. Mean *RMSE* differences showed the same trend, amounting to 1.60 for clay, 0.84 for silt and 1.28 for sand. The correlation matrix supported the previous statement, as the lowest correlation was observed between OK and deterministic interpolation methods, especially SI. OK also produced the most accurate interpolation results in a similar study, compared to deterministic interpolation methods [31]. However, a study by [34] proved that in the case of low sample data normality and stationarity, deterministic methods like IDW produce more accurate results than OK. The accuracy assessment in five independent repetitions was proven to be effective, as it allowed testing of interpolation performances in variable conditions. The interpolation results of IDW and SI were suitable for maintaining the effects of local soil variabilities, however the same can be achieved by the modifications of OK parameters if necessary [74].

The fulfilment of the second objective of the study regarding the automation of soil texture classification enabled the operator to avoid manual reclassification of classes. Automation of the classification process naturally resulted in shorter processing time than by manual tools execution [75]. Lower human interference in the data processing also provided a uniform classification result regardless of the input data, as well as its universal applicability. Recent advances in machine learning algorithms development allowed their implementation and automation in soil texture mapping, also presenting a new possibility in future research [76]. The current limitation of this study is the full automation of the interpolation and classification algorithm, which could be performed by R or Python programming languages.

The applied soil texture classification algorithm allowed the delineation of soil texture suitability zones, which is often used as a criterion in various GIS-based multicriteria analysis studies. Soil texture was an impactful criterion in many natural resource management studies: grassland conservation [77], irrigation management [78], soil salinity analyses [79] and drought risk assessment [80]. Future directions of suitability analyses using soil texture zones are planned for vegetable suitability multicriteria analyses, which are highly beneficial in sustainable planning [81].

## 5. Conclusions

The automation of soil texture classification enabled time-efficient delineation of suitability zones for soybean cultivation. Clay, silt and sand soil contents were evaluated using a random split-sample method in five repetitions, which ensured the objective determination of training data normality and stationarity. The descriptive statistics of evaluated training data indicated moderate data normality and stationarity for clay and silt, while sand soil content resulted in highly skewed distribution and moderate stationarity. The application of logarithmic transformation as a pre-processing to OK reduced the effect of non-normal distribution, particularly for the sand soil fraction. Tested deterministic methods, IDW and SI, produced lower interpolation accuracy values than OK, but were considered as viable methods in the case of low data normality and stationarity. The authors recommend the evaluation of interpolation methods for suitability calculations, as their performance is variable and depends on the properties of input samples. It was also observed that these methods were effective in retaining the local variation data in the interpolation result. Loam and clay loam were determined as the most suitable soil texture zones for soybean cultivation, covering the northern part of the study area near the river Drava. Classified soil texture classes with the highest presence in the study area were moderately suitable silty clay loam (53.26%) and silt loam (34.96%). According to the proposed methods, the study area is dominantly moderately suitable for soybean cultivation, based on the soil texture data. The integration of Sentinel-2 multispectral satellite images and GIS-based multicriteria analysis with current results presents the foundation of future research.

**Author Contributions:** Conceptualization, D.R.; methodology, D.R.; software, D.R.; validation, M.J., V.Z. and I.P.; formal analysis, D.R.; investigation, D.R.; resources, M.J., V.Z. and I.P.; data curation, D.R.; writing—original draft preparation, D.R.; writing—review and editing, D.R., M.J., V.Z. and I.P.; visualization, D.R.; supervision, M.J. and I.P.; funding acquisition, M.J. and I.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** This work was supported by the Faculty of Agrobiotechnical Sciences Osijek as a part of the scientific project “AgroGIT—Technical and Technological Crop Production Systems, GIS and Environment Protection”.

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

## Appendix A. Automatic Soil Texture Classification Algorithm Scripted in Python

```
#import of the required libraries
import gdal
import numpy
from osgeo import ogr
import os
#import of interpolated clay, silt and sand soil content rasters (marked in italic)
clay_all = gdal.Open('clay.tif')
clay_band = clay_all.GetRasterBand(1)
clay = clay_band.ReadAsArray()
silt_all = gdal.Open('silt.tif')
silt_band = silt_all.GetRasterBand(1)
silt = silt_band.ReadAsArray()
sand_all = gdal.Open('sand.tif')
sand_band = sand_all.GetRasterBand(1)
sand = sand_band.ReadAsArray()
#assignment of new soil texture raster
soil_texture = clay
[cols,rows] = soil_texture.shape
#automatic classification function for soil texture
def classification (clay,silt,sand,x,y):
```

```

soil_texture_class = 0
if 55 <= clay[x][y] <= 100 and 0 <= silt[x][y] <= 40 and 20 <= sand[x][y] <= 45:
    soil_texture_class = 1 #clay
elif 35 <= clay[x][y] <= 55 and 0 <= silt[x][y] <= 20 and 45 <= sand[x][y] <= 65:
    soil_texture_class = 2 #sandy clay
elif 40 <= clay[x][y] <= 60 and 40 <= silt[x][y] <= 60 and 0 <= sand[x][y] <= 20:
    soil_texture_class = 3 #silty clay
elif 20 <= clay[x][y] <= 35 and 0 <= silt[x][y] <= 28 and 45 <= sand[x][y] <= 80:
    soil_texture_class = 4 #sandy clay loam
elif 28 <= clay[x][y] <= 40 and 40 <= silt[x][y] <= 70 and 0 <= sand[x][y] <= 20:
    soil_texture_class = 5 #silty clay loam
elif 28 <= clay[x][y] <= 40 and 15 <= silt[x][y] <= 52 and 20 <= sand[x][y] <= 45:
    soil_texture_class = 6 #clay loam
elif 8 <= clay[x][y] <= 28 and 28 <= silt[x][y] <= 50 and 22 <= sand[x][y] <= 52:
    soil_texture_class = 7 #loam
elif 15 <= clay[x][y] <= 20 and 0 <= silt[x][y] <= 50 and 45 <= sand[x][y] <= 85:
    soil_texture_class = 8 #sandy loam
elif 0 <= clay[x][y] <= 28 and 50 <= silt[x][y] <= 100 and 0 <= sand[x][y] <= 50:
    soil_texture_class = 9 #silt loam
elif 10 <= clay[x][y] <= 15 and 0 <= silt[x][y] <= 30 and 70 <= sand[x][y] <= 85:
    soil_texture_class = 10 #loamy sand
elif 0 <= clay[x][y] <= 10 and 0 <= silt[x][y] <= 15 and 85 <= sand[x][y] <= 100:
    soil_texture_class = 11 #sand
elif 0 <= clay[x][y] <= 15 and 80 <= silt[x][y] <= 100 and 0 <= sand[x][y] <= 20:
    soil_texture_class = 12 #silt
return soil_texture_class
#automatic classification of each pixel in new soil texture raster
for x in range(len(soil_texture)):
    for y in range(len(soil_texture[x])):
        soil_texture[x][y] = classification(clay,silt,sand,x,y)
#export of classified soil texture raster (marked in italic)
driver = gdal.GetDriverByName("GTiff")
outdata = driver.Create('soil_texture.tif', rows, cols, 1, gdal.GDT_UInt16)
outdata.SetGeoTransform(clay_all.GetGeoTransform())
outdata.SetProjection(clay_all.GetProjection())
outdata.GetRasterBand(1).WriteArray(soil_texture)
outdata.GetRasterBand(1).SetNoDataValue(-99999)
outdata.FlushCache()
outdata = None
band = None
soil_texture = None

```



Appendix B. Interpolation Parameters of Training Sample Sets

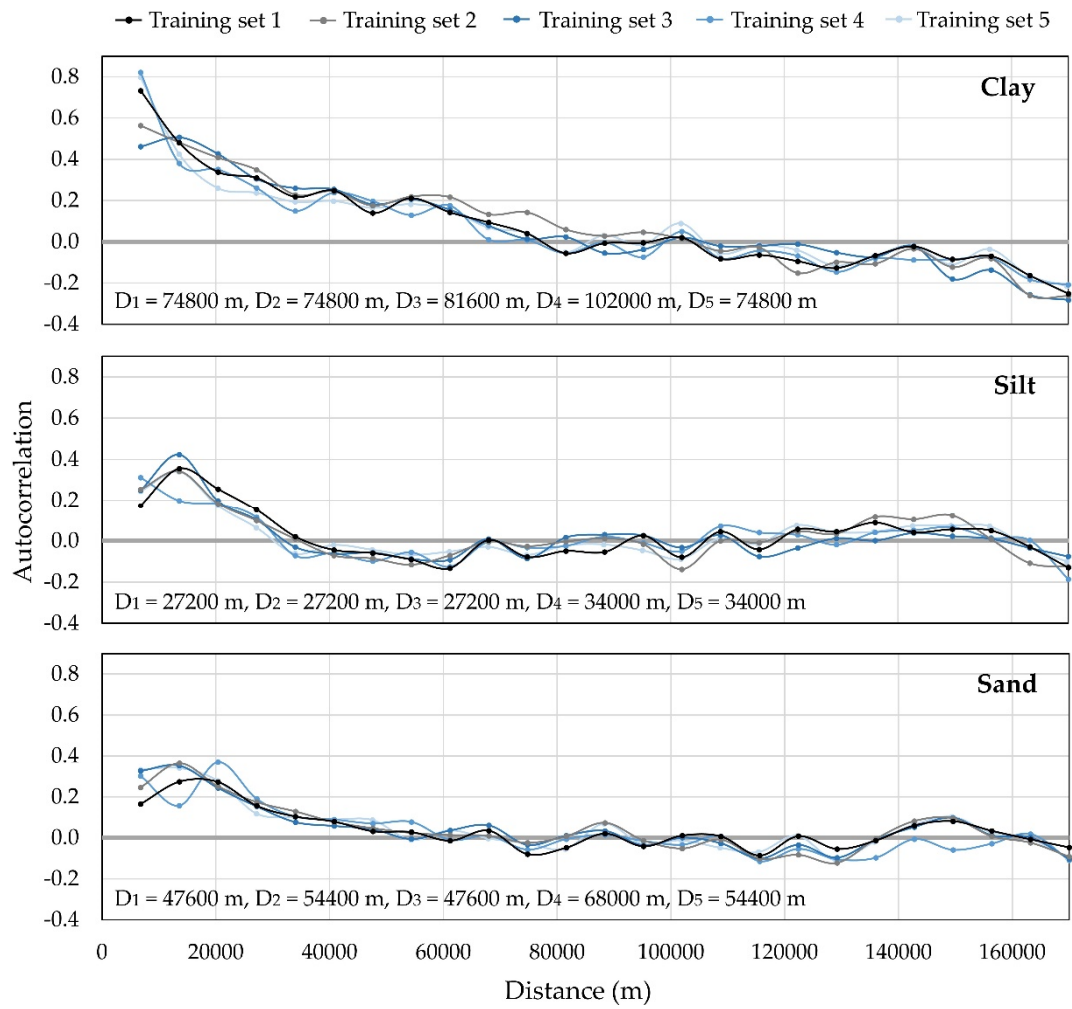
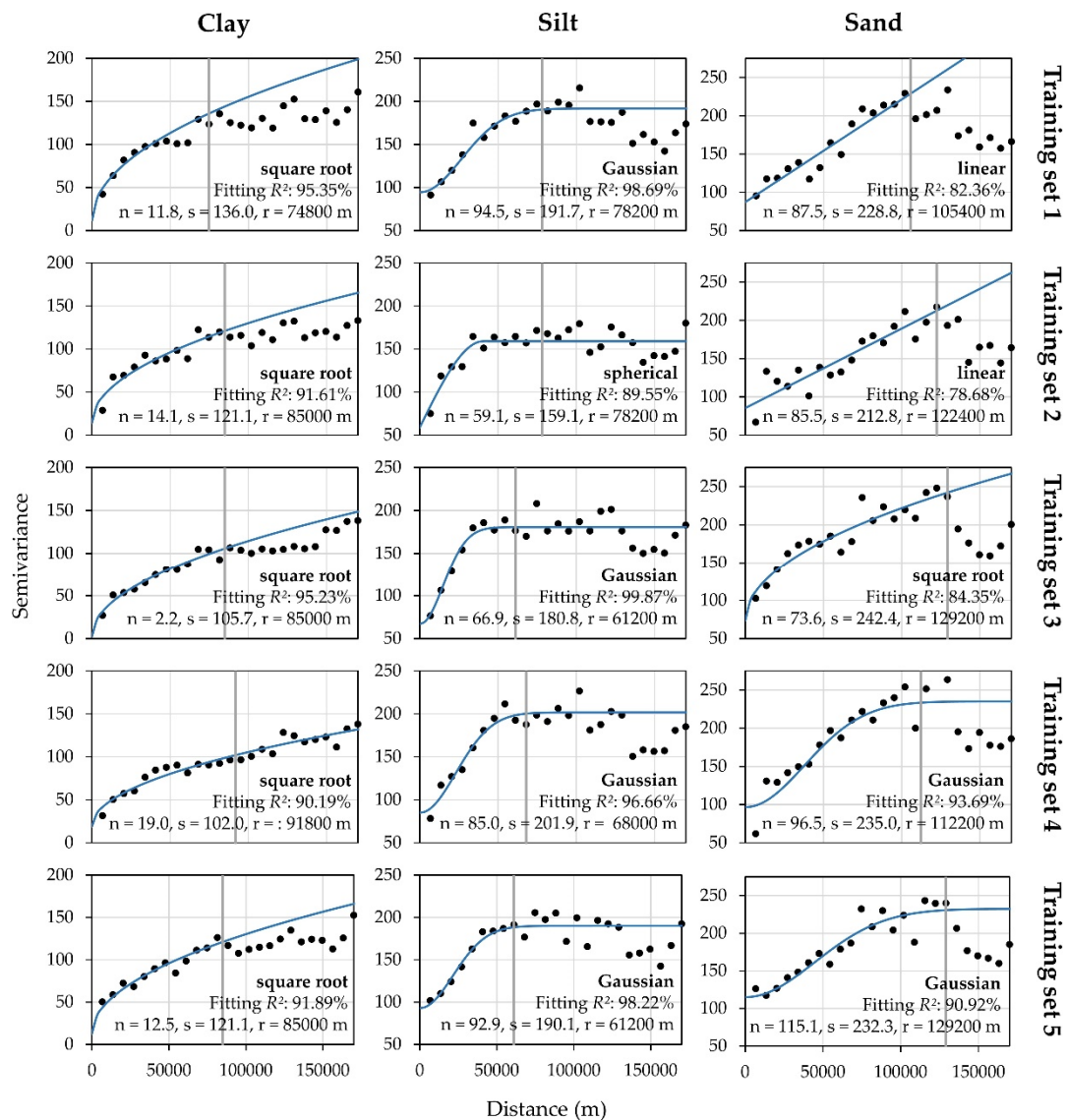


Figure A1. Autocorrelograms of five training sample sets for clay, silt and sand.  $D_n$  denominates maximum distance with continuous positive autocorrelation for  $n$  training set.



**Figure A2.** Variograms and OK interpolation parameters of five training sample sets (n: nugget, s: sill, r: range).

## References

1. Plaščak, I.; Jurišić, M.; Radočaj, D.; Barač, Ž.; Glavaš, J. Hazel plantation planning using GIS and multicriteria decision analysis. *Poljoprivreda* **2019**, *25*, 79–85. [[CrossRef](#)]
2. El Baroudy, A.A. Mapping and evaluating land suitability using a GIS-based model. *Catena* **2016**, *140*, 96–104. [[CrossRef](#)]
3. Radočaj, D.; Jurišić, M.; Gašparović, M.; Plaščak, I. Optimal Soybean (*Glycine max* L.) Land Suitability Using GIS-Based Multicriteria Analysis and Sentinel-2 Multitemporal Images. *Remote Sens.* **2020**, *12*, 1463. [[CrossRef](#)]
4. Kazemi, H.; Akinci, H. A land use suitability model for rainfed farming by Multi-criteria Decision-making Analysis (MCDA) and Geographic Information System (GIS). *Ecol. Eng.* **2018**, *116*, 1–6. [[CrossRef](#)]
5. Greve, M.H.; Kheir, R.B.; Greve, M.B.; Bøcher, P.K. Quantifying the ability of environmental parameters to predict soil texture fractions using regression-tree model with GIS and LIDAR data: The case study of Denmark. *Ecol. Indic.* **2012**, *18*, 1–10. [[CrossRef](#)]
6. García-Tomillo, A.; Mirás-Avalos, J.M.; Dafonte-Dafonte, J.; Paz-González, A. Mapping Soil Texture Using Geostatistical Interpolation Combined with Electromagnetic Induction Measurements. *Soil Sci.* **2017**, *182*, 278–284. [[CrossRef](#)]



7. Zebec, V.; Semialjac, Z.; Marković, M.; Tadić, V.; Radić, D.; Rastija, D. Influence of physical and chemical properties of different soil types on optimal soil moisture for tillage. *Poljoprivreda* **2017**, *23*, 10–18. [[CrossRef](#)]
8. Zipper, S.C.; Soylu, M.E.; Booth, E.G.; Loheide, S.P. Untangling the effects of shallow groundwater and soil texture as drivers of subfield-scale yield variability. *Water Resour. Res.* **2015**, *51*, 6338–6358. [[CrossRef](#)]
9. Bach, E.M.; Baer, S.G.; Meyer, C.K.; Six, J. Soil texture affects soil microbial and structural recovery during grassland restoration. *Soil Biol. Biochem.* **2010**, *42*, 2182–2191. [[CrossRef](#)]
10. Chau, J.F.; Bagtzoglou, A.C.; Willig, M.R. The effect of soil texture on richness and diversity of bacterial communities. *Environ. Forensics* **2011**, *12*, 333–341. [[CrossRef](#)]
11. Gašparović, M.; Zrinjski, M.; Gudelj, M. Automatic cost-effective method for land cover classification (ALCC). *Comput. Environ. Urban Syst.* **2019**, *76*, 1–10. [[CrossRef](#)]
12. Gilliot, J.M.; Vaudour, E.; Michelin, J. Soil surface roughness measurement: A new fully automatic photogrammetric approach applied to agricultural bare fields. *Comput. Electron. Agric.* **2017**, *134*, 63–78. [[CrossRef](#)]
13. Sirsat, M.S.; Cernadas, E.; Fernández-Delgado, M.; Barro, S. Automatic prediction of village-wise soil fertility for several nutrients in India using a wide range of regression methods. *Comput. Electron. Agric.* **2018**, *154*, 120–133. [[CrossRef](#)]
14. Bonini Neto, A.; Bonini, C.S.B.; Reis, A.R.; Piazzentin, J.C.; Coletta, L.F.S.; Putti, F.F.; Heinrichs, R.; Moreira, A. Automatic recovery estimation of degraded soils by artificial neural networks in function of chemical and physical attributes in Brazilian Savannah soil. *Commun. Soil Sci. Plant Anal.* **2019**, *50*, 1785–1798. [[CrossRef](#)]
15. Adhikari, K.; Kheir, R.B.; Greve, M.B.; Bøcher, P.K.; Malone, B.P.; Minasny, B.; McBratney, A.; Greve, M.H. High-resolution 3-D mapping of soil texture in Denmark. *Soil Sci. Soc. Am. J.* **2013**, *77*, 860–876. [[CrossRef](#)]
16. Gomez, C.; Dharumarajan, S.; Féret, J.B.; Lagacherie, P.; Ruiz, L.; Sekhar, M. Use of sentinel-2 time-series images for classification and uncertainty analysis of inherent biophysical property: Case of soil texture mapping. *Remote Sens.* **2019**, *11*, 565. [[CrossRef](#)]
17. Webster, R.; Oliver, M.A. *Geostatistics for Environmental Scientists*, 2nd ed.; John Wiley & Sons: Chichester, UK, 2007; p. 330.
18. Augusto Filho, O.; Soares, W.; Irigaray, C. Mapping of compactness by depth in a quaternary geological formation using deterministic and geostatistical interpolation models. *Environ. Earth Sci.* **2017**, *76*, 607. [[CrossRef](#)]
19. Oliver, M.A. The variogram and kriging. In *Handbook of Applied Spatial Analysis*; Ficher, M., Getis, A., Eds.; Springer: Berlin, Germany, 2010; pp. 319–352. [[CrossRef](#)]
20. Bhunia, G.S.; Shit, P.K.; Maiti, R. Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC). *J. Saudi Soc. Agric. Sci.* **2018**, *17*, 114–126. [[CrossRef](#)]
21. Long, J.; Liu, Y.; Xing, S.; Qiu, L.; Huang, Q.; Zhou, B.; Shen, J.; Zhang, L. Effects of sampling density on interpolation accuracy for farmland soil organic matter concentration in a large region of complex topography. *Ecol. Indic.* **2018**, *93*, 562–571. [[CrossRef](#)]
22. Du, Y.; Zhao, Q.; Chen, L.; Yao, X.; Xie, F. Effect of Drought Stress at Reproductive Stages on Growth and Nitrogen Metabolism in Soybean. *Agronomy* **2020**, *10*, 302. [[CrossRef](#)]
23. Kumagai, E.; Takahashi, T. Soybean (*Glycine max* (L.) Merr.) Yield Reduction due to Late Sowing as a Function of Radiation Interception and Use in a Cool Region of Northern Japan. *Agronomy* **2020**, *10*, 66. [[CrossRef](#)]
24. Ray, D.K.; Mueller, N.D.; West, P.C.; Foley, J.A. Yield trends are insufficient to double global crop production by 2050. *PLoS ONE* **2013**, *8*, e66428. [[CrossRef](#)] [[PubMed](#)]
25. Masuda, T.; Goldsmith, P.D. World soybean production: Area harvested, yield, and long-term projections. *Int. Food Agribus. Manag. Rev.* **2009**, *12*, 1–20. [[CrossRef](#)]
26. Toleikienė, M.; Brophy, C.; Arlauskienė, A.; Rasmussen, J.; Geçaitė, V.; Kadžiulienė, Ž. The introduction of soybean in an organic crop rotation in the Nemoral zone: The impact on subsequent spring wheat productivity. *Zemdirbyste* **2019**, *106*, 321–328. [[CrossRef](#)]
27. Melakeberhan, H.; Avendaño, F.; Pierce, F. Spatial analysis of soybean yield in relation to soil texture, soil fertility and soybean cyst nematode. *Nematology* **2004**, *6*, 527–545. [[CrossRef](#)]
28. Arora, V.K.; Singh, C.B.; Sidhu, A.S.; Thind, S.S. Irrigation, tillage and mulching effects on soybean yield and water productivity in relation to soil texture. *Agric. Water Manag.* **2011**, *98*, 563–568. [[CrossRef](#)]
29. Pagano, M.C.; Miransari, M. The importance of soybean production worldwide. In *Abiotic and Biotic Stresses in Soybean Production*; Miransari, M., Ed.; Academic Press: Cambridge, MA, USA, 2016; pp. 1–26. [[CrossRef](#)]

30. European Environment Agency—Biogeographical Regions. Available online: <https://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-3> (accessed on 9 May 2020).
31. Robinson, T.P.; Metternicht, G. Testing the performance of spatial interpolation techniques for mapping soil properties. *Comput. Electron. Agric.* **2006**, *50*, 97–108. [[CrossRef](#)]
32. Croatian Geological Institute—Description and Explanation of Applied Methods. Available online: [http://www.haop.hr/sites/default/files/uploads/news/2017-12/Opis\\_i\\_obrazlozenje\\_koristenih\\_metoda\\_istrazivanja\\_HGI.pdf](http://www.haop.hr/sites/default/files/uploads/news/2017-12/Opis_i_obrazlozenje_koristenih_metoda_istrazivanja_HGI.pdf) (accessed on 9 May 2020).
33. Hengl, T.; Heuvelink, G.B.; Stein, A. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* **2004**, *120*, 75–93. [[CrossRef](#)]
34. Jurišić, M.; Plaščak, I.; Antonić, O.; Radočaj, D. Suitability Calculation for Red Spicy Pepper Cultivation (*Capsicum annum* L.) Using Hybrid GIS-Based Multicriteria Analysis. *Agronomy* **2020**, *10*, 3. [[CrossRef](#)]
35. Oliver, M.A.; Webster, R. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena* **2014**, *113*, 56–69. [[CrossRef](#)]
36. Negreiros, J.; Painho, M.; Aguilar, F.; Aguilar, M. Geographical information systems principles of ordinary kriging interpolator. *J. Appl. Sci.* **2010**, *10*, 852–867. [[CrossRef](#)]
37. Ciotoli, G.; Lombardi, S.; Annunziatellis, A. Geostatistical analysis of soil gas data in a high seismic intermontane basin: Fucino Plain, central Italy. *J. Geophys. Res. Solid Earth* **2007**, *112*, B05407. [[CrossRef](#)]
38. Pesquer, L.; Cortés, A.; Pons, X. Parallel ordinary kriging interpolation incorporating automatic variogram fitting. *Comput. Geosci.* **2011**, *37*, 464–473. [[CrossRef](#)]
39. De Mesnard, L. Pollution models and inverse distance weighting: Some critical remarks. *Comput. Geosci.* **2013**, *52*, 459–469. [[CrossRef](#)]
40. Fortin, M.J.; Drapeau, P.; Legendre, P. Spatial autocorrelation and sampling design in plant ecology. In *Progress in Theoretical Vegetation Science*; Grabherr, G., Mucina, L., Dale, M.B., Ter Braak, C.J.F., Eds.; Springer: Dordrecht, The Netherlands, 1990; pp. 209–222.
41. Gribov, A.; Krivoruchko, K. New flexible non-parametric data transformation for trans-gaussian kriging. In *Geostatistics Oslo 2012*; Abrahamson, P., Hauge, R., Kolbjørnsen, O., Eds.; Springer: Dordrecht, The Netherlands, 2012; pp. 51–65. [[CrossRef](#)]
42. Chen, C.; Li, Y. A robust method of thin plate spline and its application to DEM construction. *Comput. Geosci.* **2012**, *48*, 9–16. [[CrossRef](#)]
43. Ditzler, C.; Scheffe, K.; Monger, H.C. *Soil Survey Manual*; Government Printing Office: Washington, DC, USA, 2017; p. 639.
44. An, D.; Zhao, G.; Chang, C.; Wang, Z.; Li, P.; Zhang, T.; Jia, J. Hyperspectral field estimation and remote-sensing inversion of salt content in coastal saline soils of the Yellow River Delta. *Int. J. Remote Sens.* **2016**, *37*, 455–470. [[CrossRef](#)]
45. Merdun, H.; Çınar, Ö.; Meral, R.; Apan, M. Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. *Soil Tillage Res.* **2006**, *90*, 108–116. [[CrossRef](#)]
46. University of Florida—Soil Texture. Available online: <https://ufdcimages.uflib.ufl.edu/IR/00/00/31/07/00001/SS16900.pdf> (accessed on 9 May 2020).
47. Food and Agriculture Organization—The Structure of the FAO Framework Classification. Available online: <http://www.fao.org/3/x5648e/x5648e0j.htm> (accessed on 9 May 2020).
48. Gaillard, R.; Duval, B.D.; Osterholz, W.R.; Kucharik, C.J. Simulated effects of soil texture on nitrous oxide emission factors from corn and soybean agroecosystems in Wisconsin. *J. Environ. Qual.* **2016**, *45*, 1540–1548. [[CrossRef](#)] [[PubMed](#)]
49. Avendaño, F.; Pierce, F.J.; Schabenberger, O.; Melakeberhan, H. The spatial distribution of soybean cyst nematode in relation to soil texture and soil map unit. *Agron. J.* **2004**, *96*, 181–194. [[CrossRef](#)]
50. Butcher, K.; Wick, A.F.; DeSutter, T.; Chatterjee, A.; Harmon, J. Corn and soybean yield response to salinity influenced by soil texture. *Agron. J.* **2018**, *110*, 1243–1253. [[CrossRef](#)]
51. Miransari, M.; Smith, D. Using signal molecule genistein to alleviate the stress of suboptimal root zone temperature on soybean-Bradyrhizobium symbiosis under different soil textures. *J. Plant Interact.* **2008**, *3*, 287–295. [[CrossRef](#)]
52. Hassink, J. Effects of soil texture and structure on carbon and nitrogen mineralization in grassland soils. *Biol. Fertil. Soils* **1992**, *14*, 126–134. [[CrossRef](#)]

53. Rosolem, C.A.; Sgariboldi, T.; Garcia, R.A.; Calonego, J.C. Potassium leaching as affected by soil texture and residual fertilization in tropical soils. *Commun. Soil Sci. Plant Anal.* **2010**, *41*, 1934–1943. [[CrossRef](#)]
54. Workneh, F.; Yang, X.B.; Tylka, G.L. Soybean brown stem rot, *Phytophthora sojae*, and *Heterodera glycines* affected by soil texture and tillage relations. *Phytopathology* **1999**, *89*, 844–850. [[CrossRef](#)] [[PubMed](#)]
55. He, W.Y.; Luo, X.L.; Sun, G.J. The Trend of GIS-Based suitable planting areas for Chinese soybean under the future climate scenario. In *Ecosystem Assessment and Fuzzy Systems Management*; Cao, B.Y., Ma, S.Q., Cao, H.H., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 325–338.
56. Subiyanto, H.; Arief, U.M.; Nafi, A.Y. An accurate assessment tool based on intelligent technique for suitability of soybean cropland: Case study in Kebumen Regency, Indonesia. *Heliyon* **2018**, *4*, e00684. [[CrossRef](#)] [[PubMed](#)]
57. Boix, L.R.; Zinck, J.A. Land-use planning in the Chaco plain (Burrucacú, Argentina). Part 1: Evaluating land-use options to support crop diversification in an agricultural frontier area using physical land evaluation. *Environ. Manag.* **2008**, *42*, 1043–1063. [[CrossRef](#)] [[PubMed](#)]
58. Seyedmohammadi, J.; Sarmadian, F.; Jafarzadeh, A.A.; Ghorbani, M.A.; Shahbazi, F. Application of SAW, TOPSIS and fuzzy TOPSIS models in cultivation priority planning for maize, rapeseed and soybean crops. *Geoderma* **2018**, *310*, 178–190. [[CrossRef](#)]
59. Bandyopadhyay, S.; Jaiswal, R.K.; Hegde, V.S.; Jayaraman, V. Assessment of land suitability potentials for agriculture using a remote sensing and GIS based approach. *Int. J. Remote Sens.* **2009**, *30*, 879–895. [[CrossRef](#)]
60. Song, Y.Q.; Zhao, X.; Su, H.Y.; Li, B.; Hu, Y.M.; Cui, X.S. Predicting Spatial Variations in Soil Nutrients with Hyperspectral Remote Sensing at Regional Scale. *Sensors* **2018**, *18*, 3086. [[CrossRef](#)] [[PubMed](#)]
61. Metwally, M.S.; Shaddad, S.M.; Liu, M.; Yao, R.-J.; Abdo, A.I.; Li, P.; Jiao, J.; Chen, X. Soil Properties Spatial Variability and Delineation of Site-Specific Management Zones Based on Soil Fertility Using Fuzzy Clustering in a Hilly Field in Jianyang, Sichuan, China. *Sustainability* **2019**, *11*, 7084. [[CrossRef](#)]
62. Delbari, M.; Afrasiab, P.; Loiskandl, W. Geostatistical analysis of soil texture fractions on the field scale. *Soil Water Res.* **2011**, *6*, 173–189. [[CrossRef](#)]
63. Zhang, Y.; Guo, L.; Chen, Y.; Shi, T.; Luo, M.; Ju, Q.; Zhang, H.; Wang, S. Prediction of Soil Organic Carbon based on Landsat 8 Monthly NDVI Data for the Jiangnan Plain in Hubei Province, China. *Remote Sens.* **2019**, *11*, 1683. [[CrossRef](#)]
64. Zhang, C.; Fay, D.; McGrath, D.; Grennan, E.; Carton, O.T. Use of trans-Gaussian kriging for national soil geochemical mapping in Ireland. *Geochem. Explor. Environ. A* **2008**, *8*, 255–265. [[CrossRef](#)]
65. Bartóg, P.; Hlisnikovský, L.; Kunzová, E. Effect of Digestate on Soil Organic Carbon and Plant-Available Nutrient Content Compared to Cattle Slurry and Mineral Fertilization. *Agronomy* **2020**, *10*, 379. [[CrossRef](#)]
66. Bogunović, I.; Telak, J.L.; Pereira, P. Agriculture Management Impacts on Soil Properties and Hydrological Response in Istria (Croatia). *Agronomy* **2020**, *10*, 282. [[CrossRef](#)]
67. Dong-Sheng, Y.; Zhang, Z.; Hao, Y.; Xue-Zheng, S.; Man-Zhi, T.; Wei-Xia, S.; Hong-Jie, W. Effect of soil sampling density on detected spatial variability of soil organic carbon in a red soil region of China. *Pedosphere* **2011**, *21*, 207–213. [[CrossRef](#)]
68. Nanni, M.R.; Povh, F.P.; Dematté, J.A.M.; Oliveira, R.B.D.; Chicati, M.L.; Cezar, E. Optimum size in grid soil sampling for variable rate application in site-specific management. *Sci. Agric.* **2011**, *68*, 386–392. [[CrossRef](#)]
69. Schmidt, J.P.; Taylor, R.K.; Milliken, G.A. Evaluating the potential for site-specific phosphorus applications without high-density soil sampling. *Soil Sci. Soc. Am. J.* **2002**, *66*, 276–283. [[CrossRef](#)]
70. Ballabio, C.; Panagos, P.; Monatanarella, L. Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma* **2016**, *261*, 110–123. [[CrossRef](#)]
71. Cruz-Cárdenas, G.; López-Mata, L.; Ortiz-Solorio, C.A.; Villaseñor, J.L.; Ortiz, E.; Silva, J.T.; Estrada-Godoy, F. Interpolation of Mexican soil properties at a scale of 1:1,000,000. *Geoderma* **2014**, *213*, 29–35. [[CrossRef](#)]
72. Zebec, V.; Rastija, D.; Lončarić, Z.; Bensa, A.; Popović, B.; Ivezić, V. Comparison of chemical extraction methods for determination of soil potassium in different soil types. *Eurasian Soil Sci.* **2017**, *50*, 1420–1427. [[CrossRef](#)]
73. Pennock, D.; Yates, T.; Braidek, J. Soil sampling designs. In *Soil Sampling and Methods of Analysis*; Carter, M.R., Gregorich, E.G., Eds.; CRC Press: Boca Raton, FL, USA, 2007; pp. 1–14.
74. Lloyd, C.D.; Atkinson, P.M. Assessing uncertainty in estimates with ordinary and indicator kriging. *Comput. Geosci.* **2001**, *27*, 929–937. [[CrossRef](#)]

75. AlShahrani, A.M.; Al-Abadi, M.A.; Al-Malki, A.S.; Ashour, A.S.; Dey, N. Automated system for crops recognition and classification. In *Computer Vision: Concepts, Methodologies, Tools, and Applications*; Khosrow-Pour, M., Ed.; IGI Global: Hershey, PA, USA, 2018; pp. 1208–1223.
76. Hengl, T.; Mendes de Jesus, J.; Heuvelink, G.B.; Ruiperez Gonzalez, M.; Kilibarda, M.; Blagotić, A.; Shangquan, W.; Wright, M.N.; Geng, X.; Bauer-Marschallinger, B.; et al. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **2017**, *12*, e0169748. [[CrossRef](#)] [[PubMed](#)]
77. Vázquez-Quintero, G.; Prieto-Amparán, J.A.; Pinedo-Alvarez, A.; Valles-Aragón, M.C.; Morales-Nieto, C.R.; Villarreal-Guerrero, F. GIS-Based Multicriteria Evaluation of Land Suitability for Grasslands Conservation in Chihuahua, Mexico. *Sustainability* **2020**, *12*, 185. [[CrossRef](#)]
78. Rodríguez Sousa, A.A.; Barandica, J.M.; Rescia, A. Ecological and Economic Sustainability in Olive Groves with Different Irrigation Management and Levels of Erosion: A Case Study. *Sustainability* **2019**, *11*, 4681. [[CrossRef](#)]
79. Hamdi, L.; Suleiman, A.; Hoogenboom, G.; Shelia, V. Response of the Durum Wheat Cultivar Um Qais (*Triticum turgidum* subsp. *durum*) to Salinity. *Agriculture* **2019**, *9*, 135. [[CrossRef](#)]
80. Wijitkosum, S.; Sriburi, T. Fuzzy AHP Integrated with GIS Analyses for Drought Risk Assessment: A Case Study from Upper Phetchaburi River Basin, Thailand. *Water* **2019**, *11*, 939. [[CrossRef](#)]
81. Hayashi, K.; Kawashima, H. Integrated evaluation of greenhouse vegetable production: Toward sustainable management. In *ISHS Acta Horticulturae 655, Proceedings of the XV International Symposium on Horticultural Economics and Management, Berlin, Germany, 1 September 2004*; Bokelmann, W., Ed.; ISHS: Leuven, Belgium, 2004; pp. 489–496. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).

## **CHAPTER 3**



Cropland Suitability Assessment Using Satellite-Based Biophysical Vegetation  
Properties and Machine Learning



## Article

# Cropland Suitability Assessment Using Satellite-Based Biophysical Vegetation Properties and Machine Learning

Dorijan Radočaj <sup>1</sup>, Mladen Jurišić <sup>1</sup>, Mateo Gašparović <sup>2</sup>, Ivan Plaščak <sup>1</sup> and Oleg Antonić <sup>3,\*</sup>

<sup>1</sup> Faculty of Agrobiotechnical Sciences Osijek, Josip Juraj Strossmayer University of Osijek, Vladimira Preloga 1, 31000 Osijek, Croatia; dradocaj@fazos.hr (D.R.); mjurisic@fazos.hr (M.J.); iplascak@fazos.hr (I.P.)

<sup>2</sup> Faculty of Geodesy, University of Zagreb, Kačićeva 26, 10000 Zagreb, Croatia; mgasparovic@geof.unizg.hr

<sup>3</sup> Department of Biology, Josip Juraj Strossmayer University of Osijek, Cara Hadrijana 8/A, 31000 Osijek, Croatia

\* Correspondence: oleg.antonc@biologija.unios.hr; Tel.: +385-31-399-917

**Abstract:** The determination of cropland suitability is a major step for adapting to the increased food demands caused by population growth, climate change and environmental contamination. This study presents a novel cropland suitability assessment approach based on machine learning, which overcomes the limitations of the conventional GIS-based multicriteria analysis by increasing computational efficiency, accuracy and objectivity of the prediction. The suitability assessment method was developed and evaluated for soybean cultivation within two 50 × 50 km subsets located in the continental biogeoregion of Croatia, in the four-year period during 2017–2020. Two biophysical vegetation properties, leaf area index (LAI) and a fraction of absorbed photosynthetically active radiation (FAPAR), were utilized to train and test machine learning models. The data derived from a medium-resolution satellite mission PROBA-V were prime indicators of cropland suitability, having a high correlation to crop health, yield and biomass in previous studies. A variety of climate, soil, topography and vegetation covariates were used to establish a relationship with the training samples, with a total of 119 covariates being utilized per yearly suitability assessment. Random forest (RF) produced a superior prediction accuracy compared to support vector machine (SVM), having the mean overall accuracy of 76.6% to 68.1% for Subset A and 80.6% to 79.5% for Subset B. The 6.1% of the highly suitable FAO suitability class for soybean cultivation was determined on the sparsely utilized Subset A, while the intensively cultivated agricultural land produced only 1.5% of the same suitability class in Subset B. The applicability of the proposed method for other crop types adjusted by their respective vegetation periods, as well as the upgrade to high-resolution Sentinel-2 images, will be a subject of future research.

**Keywords:** leaf area index (LAI); fraction of absorbed photosynthetically active radiation (FAPAR); random forest (RF); support vector machine (SVM); soybean; GIS-based multicriteria analysis; covariates



**Citation:** Radočaj, D.; Jurišić, M.; Gašparović, M.; Plaščak, I.; Antonić, O. Cropland Suitability Assessment Using Satellite-Based Biophysical Vegetation Properties and Machine Learning. *Agronomy* **2021**, *11*, 1620. <https://doi.org/10.3390/agronomy11081620>

Academic Editor:  
Belen Gallego-Elvira

Received: 1 July 2021

Accepted: 13 August 2021

Published: 16 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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 sustainability of present agricultural production faces severe global challenges in the form of rapid population growth [1], climate change [2] and increasing environmental contamination [3]. These factors are projected to cause serious global food nutrient deficiency by 2050 [4], thus urging for more efficient utilization of the current agricultural land. Current agricultural land management plans are frequently based on obsolete environmental conditions and monetary priorities [5], so their upgrade should be a first step in improving agricultural production systems. With the selection of suboptimal locations to cultivate crops, farmers often turn to using excessive mineral fertilizers and pesticides to achieve desired yields, damaging the ecosystem in the process [6]. Determining the cropland suitability for major crop types is the mandatory process for efficient agricultural land management planning [7]. This procedure is a key basis of globally sustainable agriculture and food security, meeting the Sustainability Development Goals of the United



Nations [8]. Soybean has a particularly increasing importance within crop rotation systems on a global scale, with a constant increase of yield and harvested land in all world regions from 1979 to projections in 2030 [9]. According to the most recent World Agricultural Supply and Demand Estimates report, its use for food, oil and biofuel production is high and is further expected to grow in the forthcoming years [10]. This indicates a high priority for solving the problematics of cropland suitability determination limitations and more efficient soybean cultivation systems globally.

Present state-of-the-art methods of cropland suitability determination are most commonly based on the geographic information system (GIS)-based multicriteria analysis, combined with the advanced criteria weighing procedures, like the Analytic Hierarchy Process (AHP). Numerous cropland suitability determination studies based on the GIS-based multicriteria analysis of both various major [11,12] and obscure crop types [13] were successfully performed. Remote sensing data from global open data satellite missions were among the fundamental data sources in these analyses [14,15]. A high degree of flexibility in the suitability determination process is one of the main advantages of GIS-based multicriteria analysis being widely applied globally [16]. However, this method has some distinctive disadvantages, which were only partially solved so far. The most obvious one is the overreliance on the user's subjective assessment of criteria selection and importance, especially within the AHP process. AHP is limited to five to nine criteria or criteria groups as per the recommendations of Saaty and Ozdemir [17], so the inclusion of additional important covariates results in more complex processing. The entire method is consequentially more susceptible to human-made blunders in pairwise comparison and criteria weight determination [18]. At the same time, the inclusion of a limited number of environmental factors results in the incomplete representation of cropland suitability. The accuracy assessment of the conventional GIS-based multicriteria analysis results is often non-existent, with some successfully performed approaches using ground truth yield data [12] or satellite-derived vegetation indices [11], which include only a segment of cropland suitability in the validation process. The possibility of an objective and easily accessible validation procedure for cropland suitability results would ensure a straightforward comparison between the prediction models and suitability results of multiple crop types [11]. This would also ensure the integration of various cropland suitability results into a unique agricultural land management foundation.

Machine learning algorithms present a possible solution to the abovementioned limitations of the GIS-based multicriteria analysis in cropland suitability assessment. They provided more efficient modelling of non-linear relationships of various environmental features and covariate data, compared to the parametric methods in recent studies [19]. Its efficiency is primarily caused due to the ability to integrate complex climate, soil and topography factors into a prediction model, unlike conventional statistical methods [20]. At the same time, the user is not expected to establish the relationships between these data. The user's main task in the machine learning prediction is the determination of covariates that are relevant to the study aim to avoid redundancy and possible bias due to the inaccurate or irrelevant covariate selection. So far, machine learning has been widely utilized with satellite-derived vegetation indices for the detection of crop rotation systems [21], crop health status [22], crop type distribution [23] and yield prediction [24]. Over the past few years, some initiatives of the machine learning application for cropland suitability assessment have achieved promising but limited results. Taghizadeh-Mehrjardi et al. [25] proved the superiority of machine learning methods compared to traditional cropland suitability determination procedures. They determined cropland suitability using empirically calculated potential yield for wheat and barley, following the Food and Agriculture Organization of the United Nations (FAO) specifications. The application of FAO standardized suitability classes is widely recognized as a stable procedure of the cropland suitability assessment, regardless of the crop type and geographical location [26]. The implementation of standardized cropland suitability classes enables effective integration with existing agricultural land management plans [11]. It also has the advantage of the



suitability comparison with other crop types to determine the best possible alternatives for the optimal agricultural subsidizing and adjustment of crop rotations. Akpoti et al. [27] successfully determined cropland suitability for rice cultivation using niche ground truth data, which required a considerable time for the preparation of the machine learning prediction. However, the potential of machine learning predictions in cropland suitability determination is still largely unutilized. With the existence of reliable and globally available training data, machine learning could represent a novel and superior approach to conventional cropland suitability determination using GIS-based multicriteria analysis.

While machine learning allows higher computational efficiency and accuracy compared to the conventional methods, there is a challenge to provide indicators that reliably specify cropland suitability levels. Many researchers related the cropland suitability with the increased crop yield and biomass [28–30]. The majority of these studies also indicated the large potential of biophysical vegetation properties in cropland suitability assessment. Leaf area index (LAI) and the fraction of absorbed photosynthetically active radiation (FAPAR) are regarded as complementary biophysical properties for crop yield estimations, frequently used as the essential variables in crop productivity assessment [31–33]. Recent studies successfully integrated LAI and FAPAR with a conventional GIS-based multicriteria analysis, producing superior suitability and yield prediction accuracy of various crop types [28,34]. These biophysical vegetation properties also showed considerable potential when used individually in crop suitability studies. LAI derived from remote sensing products was highly correlated with the crop biomass, yield and overall crop status, especially in early growth stages [35]. FAPAR showed a strong correlation with the total crop biomass production [28], its predictive modelling [36], as well as with its temporal variation during the crop vegetative period [37]. Biophysical properties derived from satellite observations produced a very high correlation with the in-situ measurements, resulting in a coefficient of determination up to 0.96 for LAI and up to 0.98 for FAPAR [31]. These biophysical vegetation properties have a long-term availability at 300 m spatial resolution from the PROBA-V mission, seamlessly upgraded to the Sentinel-3 products for global and stable use in the future [38]. By implementing a cropland suitability indicator based on multitemporal LAI and FAPAR data in the machine learning algorithms, there is considerable potential in forming a computationally efficient and globally available cropland suitability assessment method.

The aim of this study was to propose a novel cropland suitability assessment and accuracy assessment approach based on machine learning. This approach is designed to simplify the calculation of cropland suitability on a global scale and to increase the objectivity of prediction compared to the conventional GIS-based multicriteria analysis approach. The method was evaluated for the soybean cropland suitability determination, with the potentially universal applicability for other crop types.

## 2. Materials and Methods

The generalized major components of the proposed approach of cropland suitability assessment are presented in Figure 1. The cropland suitability assessment was performed using solely open-source GIS software. SAGA-GIS v7.9.0 (Hamburg, Germany) was used for input data preprocessing, machine learning prediction and accuracy assessment, while QGIS v3.14 (Grüt, Switzerland) was used for map creation. All input spatial data and suitability assessments were georeferenced to the Croatian Terrestrial Reference System (HTRS96/TM). The complete computational process of the study was performed using a desktop personal computer, which is standard equipment for agricultural land management users in the majority of the world.

The workflow of the proposed cropland suitability assessment method contains two primary steps (Figure 2): (1) spatial data acquisition and preprocessing; and (2) machine learning prediction of cropland suitability. The cropland suitability classes were determined for soybean cultivation, while potentially supporting its universal applicability with the adjustments related to the vegetation period of the selected crop type.

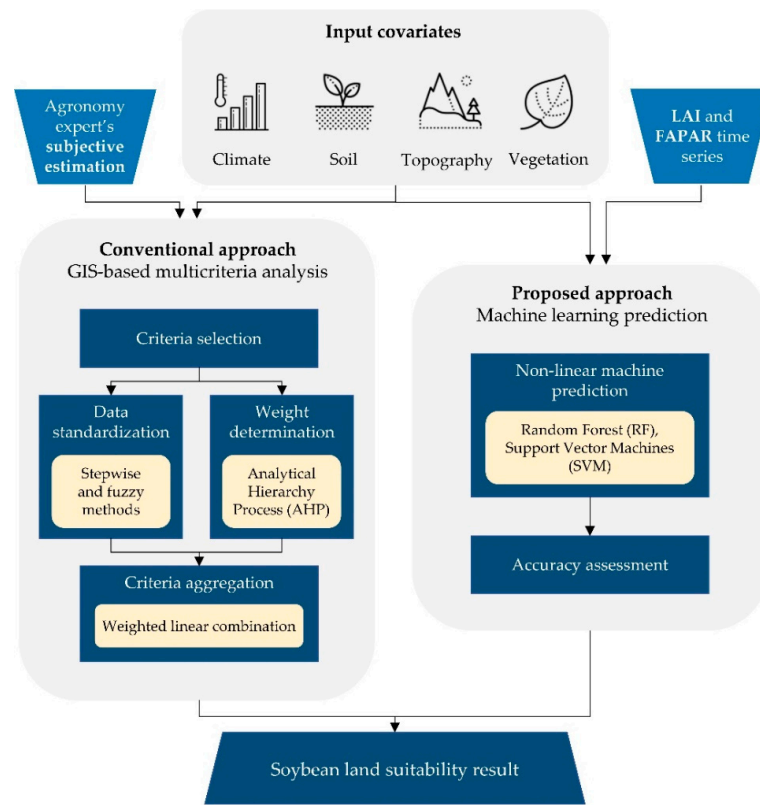


Figure 1. Major components of the conventional and proposed cropland suitability determination methods.

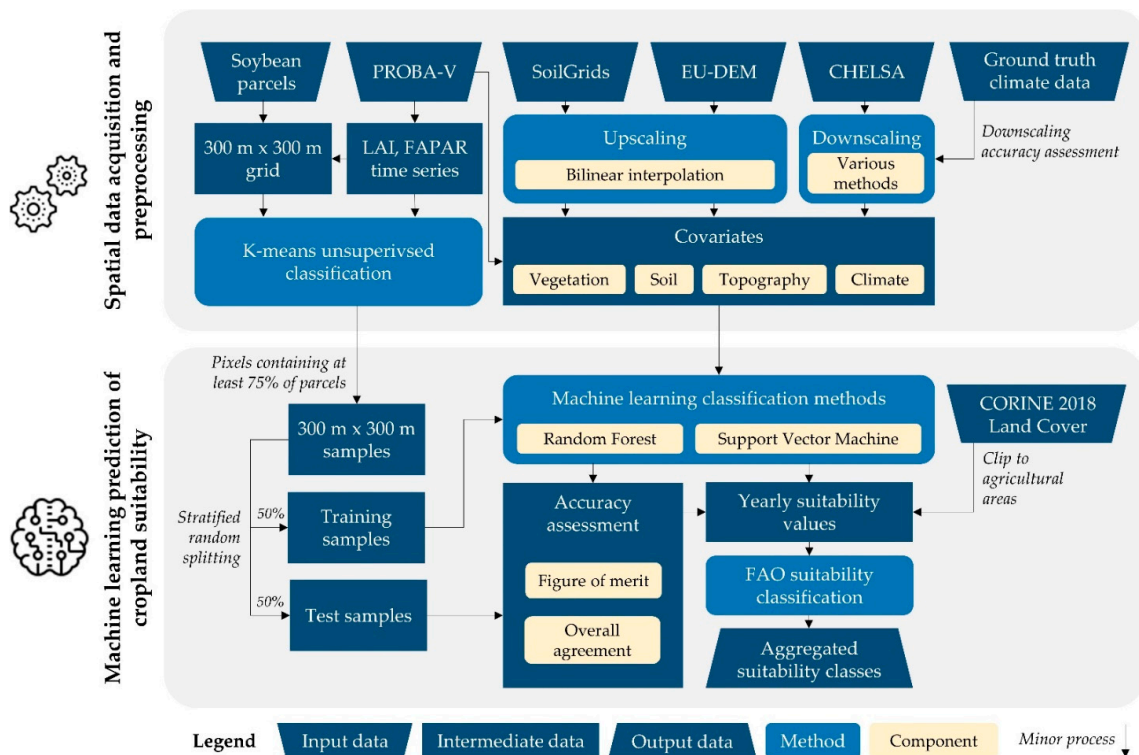
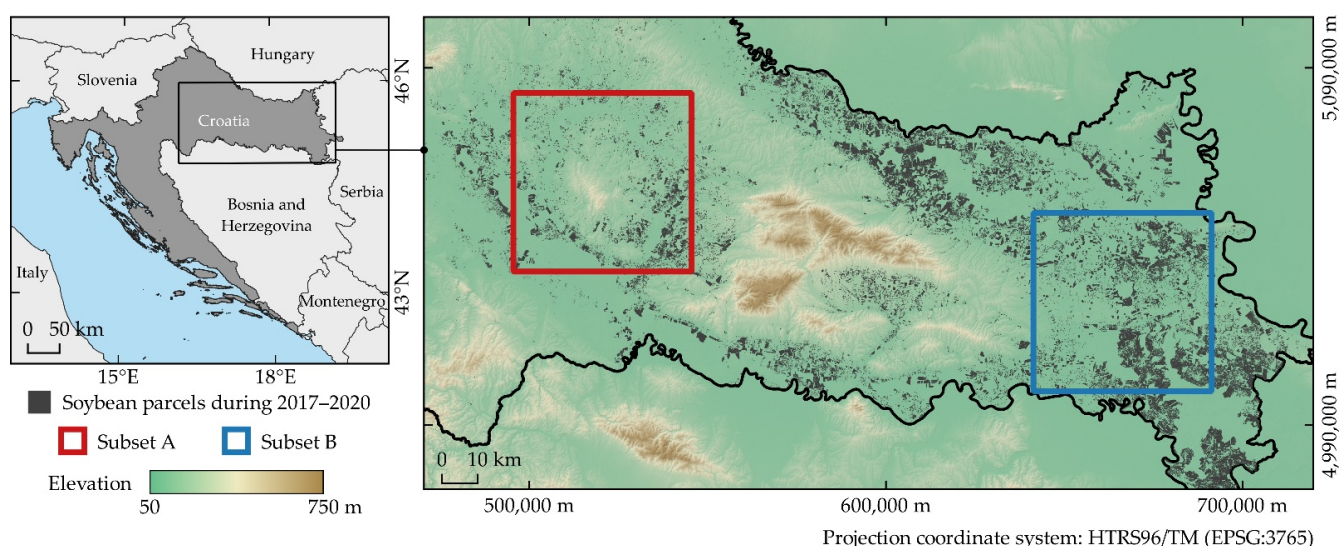


Figure 2. Workflow of the proposed cropland suitability assessment method.

### 2.1. Study Area

The study area covered two  $50 \times 50$  km subsets located in the continental biogeoregion of Croatia (Figure 3). Agriculture is one of the major activities in Continental Croatia, with agricultural areas covering 52.9% of its total area per CORINE 2018 Land Cover data. Multiple recent studies noted the considerable variability of soybean cropland suitability in the study area, urging for more efficient agricultural land management planning [11,39–41]. Subset A is characterized by hilly terrain and sparsely located agricultural parcels, often in the proximity of forests. Subset B is situated in the lowland area in eastern Croatia, being traditionally used for intensive agricultural production. Soybean is cultivated conservatively by specific land owners in both subsets, with the union of soybean parcels during 2017–2020 covering only 11.9% and 19.0% of agricultural area in Subset A and B, respectively. The general properties of these subsets are presented in Table 1. The major deviation from mean climate data occurred in 2018, which was extremely hot and dry in the soybean vegetation period. The other notable deviation was a relatively high precipitation in 2019 for both study subsets. All yearly air temperature and precipitation data in study period during 2017–2020, between April and October, are shown in Appendix A, Table A1.



**Figure 3.** The study area with two  $50 \times 50$  km subsets.

**Table 1.** General properties of subset areas.

Properties	Subset		Data Source
	A	B	
Longitude/Latitude	16°45' E, 45°41' N	18°38' E, 45°20' N	/
Major land cover classes	Agricultural areas (55.5%), Forests (39.9%), Urban areas (2.9%)	Agricultural areas (75.7%), Forests (17.8%), Urban areas (5.7%)	CORINE 2018
Total country soybean area in 2020	10.1%	22.8%	APPRRR
Mean annual air temperature	11.0 °C ± 0.2 °C	11.1 °C ± 0.1 °C	CHELSEA
Mean air temperature (April–October)	17.5 °C ± 0.3 °C	17.9 °C ± 0.1 °C	CHELSEA
Total annual precipitation	859.1 mm ± 34.7 mm	685.9 mm ± 24.9 mm	CHELSEA
Total precipitation (April–October)	547.6 mm ± 28.3 mm	449.2 mm ± 14.2 mm	CHELSEA
Mean elevation	134.8 m ± 41.0 m	91.1 m ± 9.7 m	EU-DEM
Mean slope	1.5°	0.4°	EU-DEM
Major soil types per FAO85 classification	Dystric Gleysol (Gd), Stagno-Gleyic Luvisol (Lgs)	Eutric Gleysol (Ge), Mollic Gleysol (Gm), Orthic Luvisol (Lo)	ESDC

APPRRR: Paying Agency for Agriculture, Fisheries and Rural Development of Croatia, ESDC: European Soil Data Centre.

Soybean production has increasing importance in Croatia, ranking second in a cultivated agricultural area with 83 thousand ha, behind maize, with the average yield of  $3.2 \text{ t ha}^{-1}$  [42]. According to the same source, the overall production of soybean in Croatia increased by 14.0% in 2020 compared to the year prior, with the prospect of further growth. The most common soybean variety in the study area is the mid-early maturity group 0, with an average vegetative period of 115–125 days [43]. The early maturity group 00 in subset A and mid-late maturity group I in subset B are periodically cultivated. The usual vegetative period of these soybean maturity groups ranges from late April to mid-September, covering days of the year (DOY) from 120 to 245. The duration of vegetative growth stages is in the range of 35 to 45 days after sowing. Full bloom (R2), beginning seed (R5) and full seed (R6) are regarded as the most important soybean growth stages for stable yield [44]. These stages commonly cover DOY ranges of 170–180, 190–200 and 200–220 in the study area, respectively. According to the common annual anomalies of soybean growth, the study period was determined from 1 April to 31 October. This approach included vegetation periods of all soybean parcels in the study area, regardless of their maturity group and agrotechnical operations performed by farmers.

## 2.2. Spatial Data Acquisition and Preprocessing

The machine learning prediction and accuracy assessment of cropland suitability for soybean cultivation were performed using open remote sensing and GIS data. LAI and FAPAR biophysical properties were used for the training of supervised classification machine learning models, as complementary and reliable indicators of crop yield [31]. The 10-day LAI and FAPAR products from a PROBA-V satellite with 300 m spatial resolution were downloaded from the Copernicus Global Land Service website for the period between April and October in 2017–2020. PROBA-V enables highly accurate and consistent determination of biophysical vegetation properties, on par with similar missions and observations from the ground [38]. Training and test data for suitability assessment were created according to the  $300 \text{ m} \times 300 \text{ m}$  regular grid, which corresponds to LAI and FAPAR raster grids derived from PROBA-V. The spatial resolution of 300 m was determined as suitable for various monitoring, and land management uses in agriculture at the macro level, representing medium-sized and larger agricultural parcels [45]. The pixels from this grid were filtered based on the coverage of ground truth soybean parcels within the pixels, designated separately for each year during the 2017–2020 period.

Reference soybean parcels were obtained from the official Paying Agency for Agriculture, Fisheries and Rural Development (APPRRR) of Croatia, being applied and controlled for agricultural incentive distribution. These data were additionally visually inspected and verified using the 0.5 m spatial resolution digital orthophoto provided by the State Geodetic Administration of Croatia. At least 75% soybean parcel coverage was determined as a filtering threshold to reduce the spectral mixing near the boundary of neighboring land cover classes [46]. Training and test data were created separately for each individual year in the 2017–2020 period, using data sensed during a soybean vegetative period of major soybean varieties in the study area. This approach ensures the robustness of the prediction by considering the entire vegetation period of all soybean varieties present in the study area. This is reflected in the resistance in temporal variabilities of sowing periods and particular soybean growth stage duration. These components are commonly affected by the numerous abiotic factors and farmer decision making, including annual weather trends, land cultivation systems, fertilization and irrigation systems.

Various complementary covariates were used for the establishment of the relationship between the soybean cropland suitability represented by LAI and FAPAR with the environmental conditions in the study area. The three primary environmental factors that condition the cropland suitability are climate, soil and topography [20]. Raster covariates representing these environmental requirements of soybean cultivation and the auxiliary vegetation covariates are presented in Table 2. The selection of particular covariates was performed based on the various environmental effects on the quality and quantity of



soybean production from previous studies [11,37,43,44,47,48]. A total of 119 covariates were used per the individual prediction of yearly cropland suitability classes for soybean cultivation, consisting of 47 climate, 24 soil, 6 topographic and 42 vegetation covariates.

**Table 2.** A generalized description of covariates used in the study.

Covariate Group	Covariate	Measurement Unit	Native Spatial Resolution (m)	Data Source
Climate	Mean monthly air temperature	°C	1000	CHELSA [49]
	Minimum monthly air temperature	°C		
	Maximum monthly air temperature	°C		
	Total monthly precipitation	mm		
	Bioclimatic variables	varying		
Soil	Nitrogen	cg kg <sup>-1</sup>	250	SoilGrids [50]
	Soil organic carbon	dg kg <sup>-1</sup>		
	pH	/		
	Cation exchange capacity	mmol(c) kg <sup>-1</sup>		
	Clay content	g kg <sup>-1</sup>		
	Silt content	g kg <sup>-1</sup>		
	Sand content	g kg <sup>-1</sup>		
Bulk density	cg cm <sup>-3</sup>			
Topographic	Digital elevation model	m	25	EU-DEM [51]
	Slope	°		derived from EU-DEM
	Aspect	°		
	Total potential solar radiation	kWh m <sup>-2</sup>		
	Topographic wetness index	/		
Wind exposition index	/			
Vegetation	Dry matter productivity	kg ha <sup>-1</sup> day <sup>-1</sup>	300	PROBA-V [52]
	Fraction of vegetation cover	/		

Climate has the dominant effect on the duration of soybean vegetative and reproductive growth stages, emerging efficiency after sowing and its overall requirements of sunshine and water [53]. Climate data was represented using the CHELSA dataset [49], containing the most recent global climate data at the 1 km spatial resolution during 1979–2013. Air temperature and precipitation covariates were filtered from April to October. The 19 bioclimatic variables were derived from CHELSA historical monthly data, representing air temperature and precipitation quarterly extremes and their value ranges for ecological modelling [49]. Soil chemical and physical properties have a major impact on soybean protein and oil quantity, while their variability is associated with the anomalies in soybean yield [47]. Per European Soil Data Centre, Gleysol and Luvisol soils are dominant in both subset areas, with moderate variability of their subtypes. These soil properties were represented by SoilGrids data at 0–5 cm, 5–15 cm and 15–30 cm soil depths [50], which dominantly affect soybean growth and produced yield [48]. Topography has an important role in representing the interaction of the elevation and terrain configuration with climate and soil effects on soybean cultivation [54]. Various theoretical topography indicators were used to model the micro variations of climate and soil conditions, especially regarding solar, wind and water drainage effects. The topographic wetness index was determined using the Multiple Flow Direction procedure. Total potential solar radiation and the wind exposition index were calculated according to the topo-climatology models by Böhner and Antonić [55]. Vegetation covariates derived from PROBA-V products, which are not directly related to crop yield, were added as supplementary biophysical properties to LAI and FAPAR. Dry matter productivity (DMP) indirectly represented the efficiency of solar radiation and air temperature on the dry biomass increase, while the fraction of vegetation cover (FCOVER) assessed the percentage of ground coverage by vegetation, without dependency on the crop optical properties [33]. These data produced a low to moderate correlation with LAI and FAPAR, preventing the suitability assessment bias.

Resampling of covariate rasters was performed to match the spatial resolution of LAI and FAPAR rasters of 300 m. The upscaling allows a straightforward and accurate creation of lower spatial resolution data, while the downscaling represents a more limited

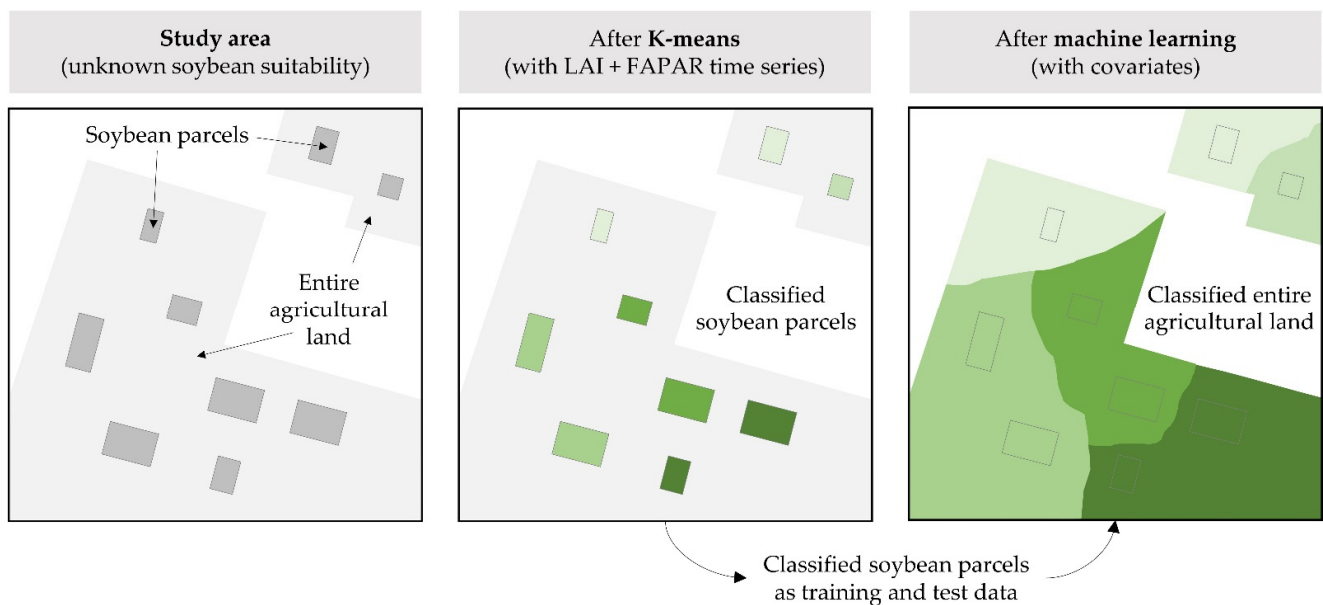
and generally less accurate process [56]. Soil and topography covariates were upscaled to 300 m spatial resolution using a bilinear interpolation method, which achieved higher accuracy compared to similar resampling methods in a recent study by Liu and Weng [57]. Various downscaling methods of the CHELSA climate data with the 1 km native spatial resolution were evaluated to ensure optimal downscaling accuracy. Nearest neighbour (NN), bilinear interpolation (BI) and B-spline interpolation (BSI) were included in the process, producing high accuracy for the downscaling of similar spatial data [58]. Accuracy assessment of the downscaled rasters was performed according to the ground truth climate data from 34 stations in the spatial coverage of study area subsets or their close proximity. Mean air temperature and total monthly precipitation were obtained from the Croatian Meteorological and Hydrological Service (DHMZ), representing the most recent official climate data (1971–2000) in Croatia. The individual monthly data between April and October was used for the accuracy assessment for air temperature, while the sum of precipitation in the same period was used to reduce bias caused by its high inter-annual variability. The coefficient of determination ( $R^2$ ) and root mean square error (RMSE) were used for the downscaling accuracy assessment, which increased with higher  $R^2$  and lower RMSE.

### 2.3. Machine Learning Prediction of Cropland Suitability

The cropland suitability for soybean cultivation was assessed following a two-step classification principle: (1) determination of suitability levels in reference soybean parcels based on K-means classification of multitemporal LAI and FAPAR; and (2) machine learning prediction of cropland suitability for soybean cultivation in the entire agricultural land in the study area, using covariates to establish a relationship between the suitability levels and environmental conditions (Figure 4). The cropland suitability prediction was performed individually for each year within the 2017–2020 period. The primary reason for that procedure was the presence of crop rotation systems, as soybean should not be cultivated in the same location in the two- or three-year consecutive span. This approach prevented interference with the spectral information of other crop types. Additionally, inter-annual weather conditions and diseases are highly variable, which significantly affect soybean biomass and yield [54]. The proposed method avoids the bias caused by integrating these conditions over multiple years by assessing cropland suitability individually for each year, preventing the impact of extremely beneficial or non-beneficial events for a particular year. The proposed method instead considers the relative suitability values in subset areas, which are almost equally affected by the weather events or diseases in the  $50 \times 50$  km areas. Therefore, K-means classification evaluated the relative soybean cropland suitability levels per year, while machine learning models were used for the absolute cropland suitability assessment, expanding the evaluation on the entire agricultural area in subset areas besides reference soybean parcels. This approach ensured objective assessment of soybean cropland suitability in the 88.1% and 81.0% of the agricultural area which was not utilized for the soybean cultivation in the 2017–2020 period for Subset A and B, respectively. The suitability assessment over the entire available area enables expansion and regionalization of soybean cultivation in new locations, supporting the increasing need for high quality and quantity of produced soybean.

LAI and FAPAR annual biophysical properties in the  $300 \times 300$  m grid were classified into five suitability values using the K-means unsupervised classification method for their determination prior to machine learning model training. The suitability values in the 1–5 range were ranked according to mean LAI and FAPAR, where higher LAI and FAPAR values indicate higher cropland suitability for soybean cultivation. A relative approach of training and test data creation using LAI and FAPAR using an unsupervised classification ensured the possibility of multi-year suitability comparison, despite annual weather variability and extremes.





**Figure 4.** Two-step classification principle using K-means and machine learning for soybean cropland suitability assessment.

Training and test data were separated from the unique classified dataset using the stratified random splitting in the 50:50 ratio. The same procedure was successfully applied with the machine learning supervised classification methods in a recent study [59]. This approach met the recommendations of Hengl et al. [50] and Colditz [60], who noted the importance of a sufficient amount of training data for machine learning prediction. Random forest (RF) and support vector machine (SVM) were applied for soybean cropland suitability assessment, being the most often applied machine learning methods due to their computational efficiency and straightforwardness [25]. They also achieved superior accuracy compared to other machine learning and conventional supervised classification algorithms in previous environmental studies [59,61]. Determining the parameters for RF and SVM prediction was based on the iterative procedure, using the parameters that ensured the highest prediction accuracy. Soybean cropland suitability assessment was performed individually for each year in the 2017–2020 period. Yearly suitability classes were assessed separately to reduce prediction bias caused by annual weather extreme events, which represent rare occurrences in the perspective of agricultural land management. These rasters were clipped to the agricultural areas land cover class from CORINE 2018, extracting the possible area for soybean cultivation. The relative importance of input covariates on the predicted soybean cropland suitability results using RF was performed by the Gini decrease measure, being a frequently used and stable measure of importance [61]. It proportionally quantifies the purity of model performance during node splits for a particular covariate, meaning that the higher Gini decrease indicates higher importance of a covariate in the prediction model.

Machine learning prediction accuracy was assessed using the figure of merit ( $F$ ), which was developed by Pontius and Millones [62] as an upgrade to kappa coefficients in remote sensing studies. It is expressed per suitability value according to the formula:

$$F = \frac{a}{o + a + c} \cdot 100\%, \quad (1)$$

where  $a$  (agreement) represents correctly predicted suitability values,  $o$  (omission) represents falsely predicted suitability values in other suitability classes and  $c$  (commission) represents falsely predicted suitability values of the particular suitability class. The overall performance of soybean cropland suitability assessment was determined using an overall agreement (OA) value, calculated as the ratio of total agreements and total classified values.

Four yearly suitability rasters were averaged and aggregated based on the FAO methodology for land suitability assessment in five classes [63]. The ranking of soybean cropland suitability values was performed according to the FAO suitability classes, including highly suitable (S1), moderately suitable (S2), marginally suitable (S3), currently not suitable (N1) and permanently not suitable (N2) classes [26]. These classes were associated with the suitability values according to the percentage of maximum suitability, per previously referenced FAO specifications (Table 3). Permanently non suitable areas in the N2 class contained all non-agricultural areas from CORINE 2018 which did not support soybean cultivation without major ecosystem disturbance.

**Table 3.** Designation of FAO suitability classes according to suitability values obtained after machine learning classification.

FAO Suitability Class	Percentage of Maximum Suitability per FAO Specifications [63]	Range of Suitability Values
S1	80–100%	4–5
S2	60–80%	3–4
S3	40–60%	2–3
N1	20–40%	1–2
N2	0–20%	non-agricultural

### 3. Results

Mean air temperature and precipitation original CHELSA 1000 m data showed a high correlation with the ground truth climate data from DHMZ stations (Appendix A, Table A2). All three evaluated downscaling interpolation methods produced high accuracy values, preserving climate values from the original data at a high degree. The B-spline interpolation method produced the highest downscaling accuracy, achieving a higher correlation with the ground truth data for precipitation compared with the original CHELSA climate data. Figure 5 displays the correlation between these datasets with the ground truth DHMZ climate data. A slightly higher correlation was observed for April and May air temperatures compared to summertime values for both subsets. Lower precipitation values in Subset B were also more accurately represented by CHELSA data compared to the higher precipitation in Subset A.

Seasonal trends of mean LAI and FAPAR values in soybean parcels during its vegetative period between April and October from 2017 to 2020 are represented in Figure 6. Both LAI and FAPAR generally reached their peak in late July or early August, which corresponds to the usual periods of soybean varieties in the study area entering the R6 growth stage. These values reached higher peaks in Subset A than in Subset B in all observed years. A slightly later vegetative period of soybean in Subset A compared to Subset B is also noted, which is characteristic for the early soybean maturity groups commonly present in this area. Meanwhile, the soybean vegetative period in Subset B matched the duration of mid-early and mid-late maturity groups, which confirms their presence in the study area from previous studies. Minor sudden changes in LAI and FAPAR trends for the year 2017 in Subset A and year 2020 in Subset B implied the susceptibility of LAI and FAPAR to annual extreme weather conditions during the early reproductive soybean growth stages. These anomalies are a common occurrence caused by drought in the study area, and were almost fully equalized in the latter soybean growth stages.

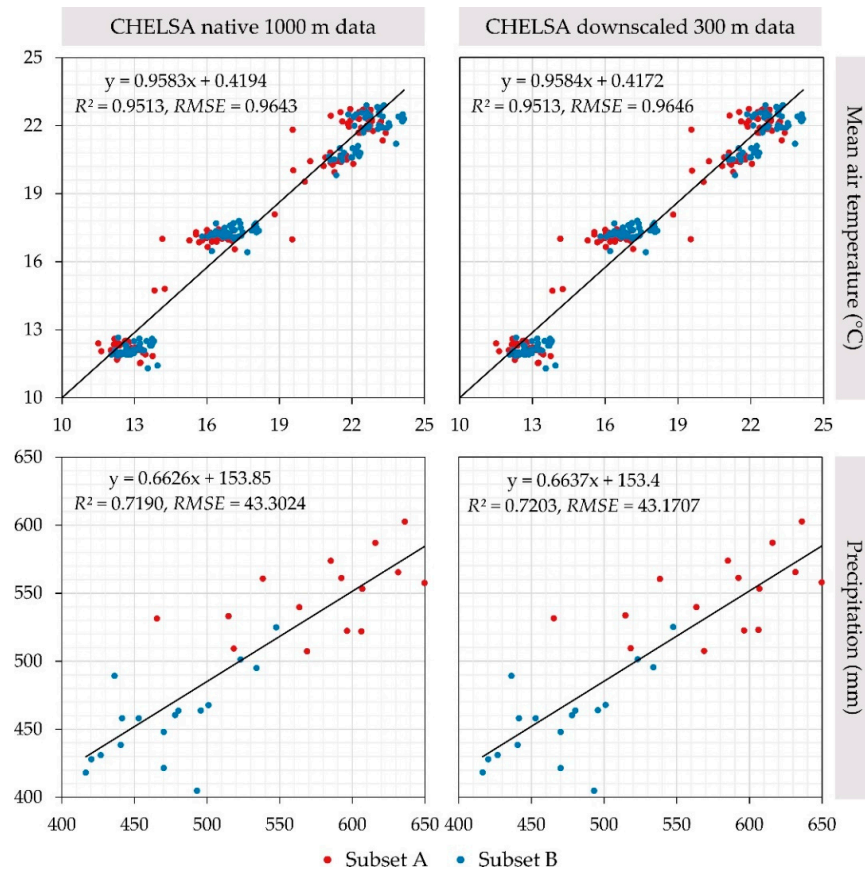


Figure 5. Scatterplot between ground truth climate data from DHMZ and the most accurate downscaling B-spline interpolation method.

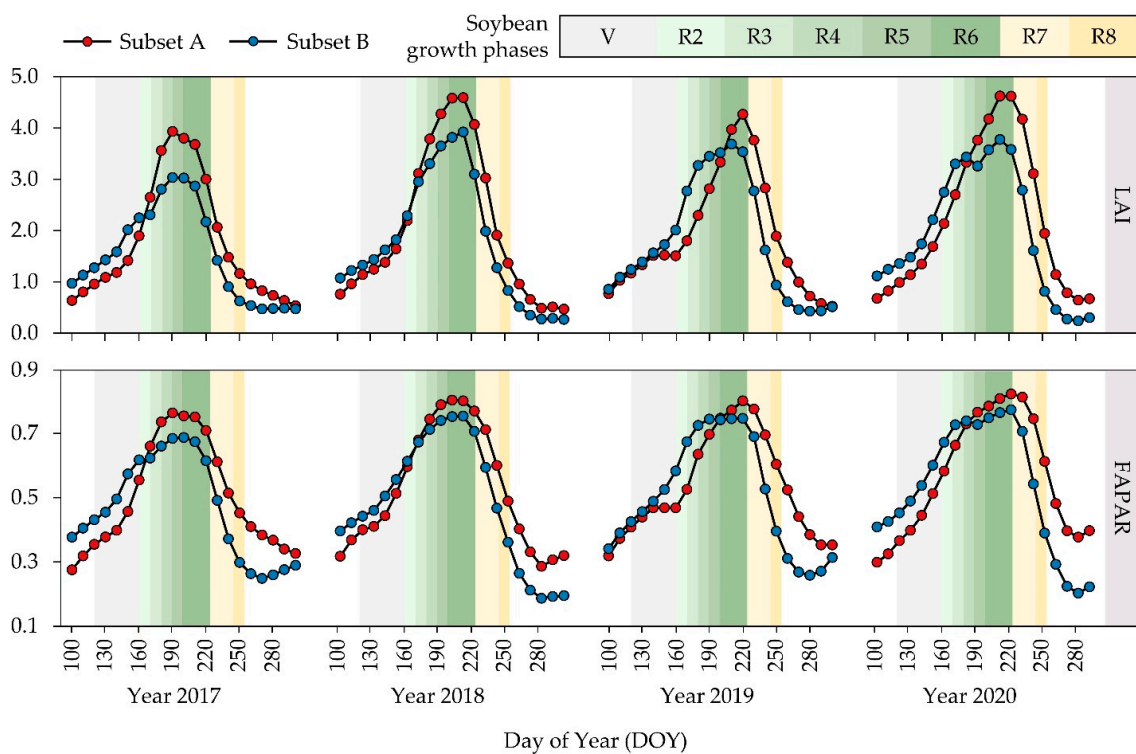


Figure 6. Mean LAI and FAPAR values in the soybean vegetative periods during 2017–2020.

The total count and coverage of the soybean samples used for the determination of training and test data for soybean cropland suitability assessment are displayed in Table 4. Both subsets produced a relatively stable sample count for each year within the 2017–2020 period. Subset B produced a higher sample count and relative coverage in the subset area due to the significantly higher amount of enlarged soybean parcels in lowland areas than Subset A. A relative percentage of sample coverage within the subset agricultural areas is in accordance with the specifications by Colditz [60], who proposed at least 0.25% from the classified area being designated as training data.

**Table 4.** Training and test sample count and area covered per subset agricultural land.

Subset/Year	Total Sample Count	Area (ha)	Percentage of Subset Agricultural Land (%)
A/2020	236	2124	1.53
A/2019	206	1854	1.34
A/2018	304	2736	1.97
A/2017	299	2691	1.94
B/2020	560	5040	2.67
B/2019	618	5562	2.94
B/2018	667	6003	3.18
B/2017	668	6012	3.18

The mean LAI and FAPAR values after suitability classification using K-means as the preprocessing to training data creation are displayed in Appendix A, Table A3. RF produced a superior classification accuracy for soybean cropland suitability in seven of eight yearly suitability values compared to SVM (Table 5). It produced higher mean OA values than SVM in both subsets, resulting in 76.6% to 68.1% for Subset A and 80.6% to 79.5% for Subset B. Cropland suitability assessment accuracy was slightly lower in Subset A, with RF producing significantly higher accuracy in conditions of more limited training data. Both machine learning methods produced a high prediction accuracy in Subset B. The correlation of the higher soybean samples with higher prediction accuracy is also observed for yearly predictions within the subsets. RF and SVM predictions for 2017 and 2018 produced a higher mean OA of 6.5% for Subset A and 5.2% for Subset B, compared to the 2019 and 2020 predictions. A general trend of higher prediction accuracy represented by the figure of merit was observed for three more suitable values for soybean cultivation (5, 4, 3) in relation to the less suitable values.

The relative importance of individual covariates divided into the abiotic (climate, soil and topography) and vegetation covariates are presented in Figure 7. Vegetation covariates produced higher individual Gini decrease values compared to abiotic covariates from the same prediction period. However, their total count of 42 compared to the 77 abiotic covariates per prediction indicated similar overall importance of these covariate groups. FCOVER and DMP sensed during June and July had the dominant importance out of the top-five importance vegetation covariates per prediction. These covariates contained 77.5% of the most impactful vegetation covariates considering both subsets. The importance of abiotic covariates varied between the subsets. SoilGrids covariates were the most represented in the top-five most impactful abiotic covariates in Subset A, covering one half of the group. It is closely followed by CHELSA climate data, with 45.0% of the top-five most important abiotic covariates. Precipitation values over the entire soybean vegetative period were the most represented climate data. SoilGrids data were dominantly included within the most impactful covariates in Subset B, representing a 75.0% share. Soil nitrogen was the most frequent soil covariate, especially at the 5–15 cm soil depth. Topographic covariates derived from EU-DEM produced 20.0% of the most impactful abiotic covariates in Subset B.



**Figure 7.** Relative importance of top five (out of 119) abiotic and vegetation covariates using Gini decrease values per yearly cropland suitability result.



Yearly and aggregated cropland suitability classes for soybean cultivation per subset are displayed in Figure 8. The average aggregated suitability values were 2.376 and 2.406 for Subsets A and B, respectively. Yearly suitability values in 2018 and 2020 produced up to 24.4% lower values compared to the mean yearly suitability in Subset A (Table 6). Traditionally intensively utilized agricultural areas in subset B produced a slightly higher percentage of suitable classes for soybean cultivation (S1–S3) with 49.5% of the subset area, compared to the 49.3% in Subset A. However, multiple locations in Subset A reached a higher suitability peak, especially regarding the most suitable S1 class. The most suitable aggregated suitability classes in Subset A were observed in the central part of the subset, containing soybean parcels dominantly surrounded by forests. The most suitable areas for soybean cultivation in Subset B were largely dispersed over the subset, while the currently non-suitable land was dominantly concentrated by the larger settlements.

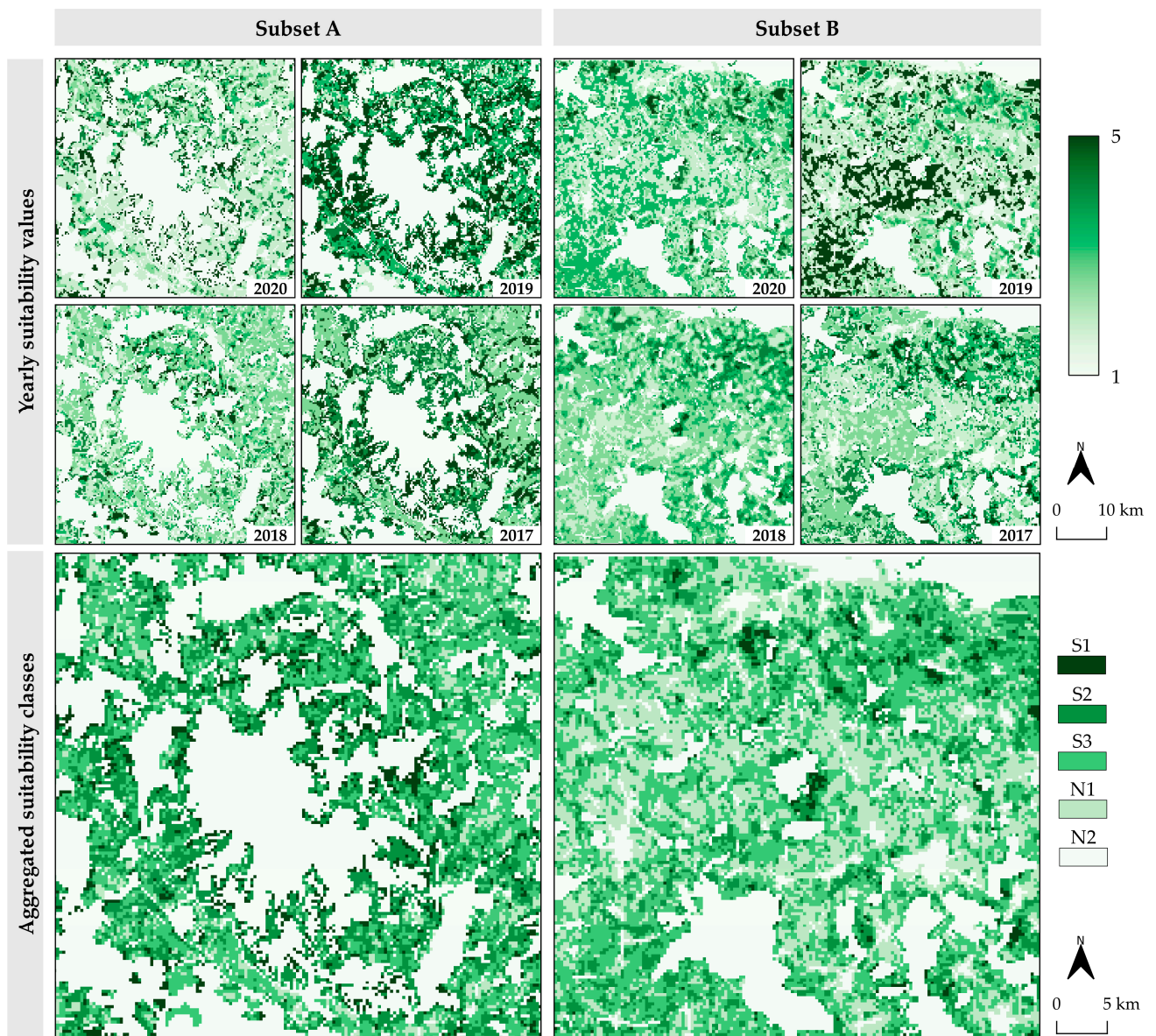


Figure 8. Yearly and aggregated cropland suitability classes for soybean cultivation.



**Table 5.** Accuracy assessment of cropland suitability values prediction using machine learning methods.

Subset/ Year	Method	Suitability Values															OA
		5 (Very High)			4 (High)			3 (Moderate)			2 (Low)			1 (Very Low)			
		<i>F</i>	<i>o</i>	<i>c</i>	<i>F</i>	<i>o</i>	<i>c</i>	<i>F</i>	<i>o</i>	<i>c</i>	<i>F</i>	<i>o</i>	<i>c</i>	<i>F</i>	<i>o</i>	<i>c</i>	
A/2020	RF	46.2	5.0	0.8	38.7	7.6	8.4	62.5	5.0	5.0	57.6	8.4	3.4	71.4	0.8	9.2	<b>73.1</b>
	SVM	43.8	4.2	3.4	40.6	6.7	9.2	51.7	9.2	2.5	36.2	10.1	15.1	59.0	6.7	6.7	63.0
A/2019	RF	66.7	3.8	1.9	77.8	2.9	2.9	65.7	3.8	7.7	46.2	7.7	5.8	41.7	6.7	6.7	<b>75.0</b>
	SVM	64.7	4.8	1.0	38.5	13.5	1.9	53.2	1.9	19.2	42.9	7.7	7.7	45.5	6.7	4.8	65.4
A/2018	RF	73.2	4.6	2.6	60.0	4.6	0.7	69.4	5.2	2.0	54.5	5.2	7.8	64.4	2.0	8.5	<b>78.4</b>
	SVM	51.0	7.8	7.8	52.4	5.2	1.3	55.9	9.2	0.7	44.7	7.2	9.8	58.0	2.0	11.8	68.6
A/2017	RF	76.9	1.3	0.7	50.0	6.0	3.3	69.0	6.0	0.0	66.1	3.3	9.3	72.2	3.3	6.7	<b>80.0</b>
	SVM	69.2	2.0	0.7	48.3	6.0	4.0	57.6	6.7	2.7	66.7	5.3	6.0	60.7	4.7	11.3	75.3
B/2020	RF	60.7	2.9	1.1	64.8	5.4	6.1	62.8	3.2	2.5	61.5	6.1	9.0	56.9	6.1	5.1	<b>76.2</b>
	SVM	44.0	5.1	0.0	55.8	7.6	7.6	60.0	4.3	1.4	62.2	3.6	12.6	59.2	5.8	4.7	73.6
B/2019	RF	67.9	1.6	1.3	73.2	1.6	1.9	69.2	5.2	3.9	62.7	7.5	5.8	65.0	4.2	7.1	79.9
	SVM	70.8	2.3	0.0	83.8	1.3	0.6	67.7	5.2	4.5	60.0	6.5	9.1	67.4	4.5	5.5	<b>80.2</b>
B/2018	RF	82.1	1.8	0.3	62.5	5.7	4.2	65.9	5.4	7.2	74.5	3.3	4.5	78.9	1.2	1.2	<b>82.5</b>
	SVM	99.9	0.0	0.0	66.3	6.3	1.8	61.5	4.8	10.8	64.8	5.7	5.4	77.8	1.8	0.6	81.3
B/2017	RF	73.0	2.7	3.3	65.1	3.0	1.5	69.2	6.4	4.5	76.2	1.8	1.2	76.6	2.1	5.5	<b>83.9</b>
	SVM	78.3	2.7	1.8	52.3	4.5	1.8	72.4	3.9	6.4	80.0	1.8	0.6	69.1	3.9	6.4	83.0

*F*: figure of merit (%); *o*: omission (%), *c*: commission (%); OA: overall agreement (%); highest OA values per subset/year were bolded.

**Table 6.** Class coverage area per aggregated soybean suitability class.

Subset	Class Coverage per Aggregated Suitability Class (%)				
	S1	S2	S3	N1	N2
A	6.1	21.0	22.2	5.7	45.0
B	1.5	13.4	34.6	25.1	25.3

#### 4. Discussion

The advantages of the proposed novel cropland suitability assessment method consist of a straightforward and computationally efficient machine learning application and the global availability of open climate, soil, topographic and vegetation data. These are some of the main factors which could ensure its extensive application in the future [64]. The proposed approach provides a stable basis for cropland suitability determination as a potentially superior long-term alternative to GIS-based multicriteria analysis. This approach allows the user to overcome the two most impactful disadvantages of the conventional GIS-based multicriteria approach, allowing the inclusion of complex input data [50] and avoiding subjectivity in suitability assessment [18]. The additional advantage to the conventional approach is the accuracy assessment of the easily accessible LAI and FAPAR data [38]. This process provides a step forward in the objective assessment of model performance and the comparison of FAO suitability classes between two individual datasets. The proposed method was successfully performed using a commonly available desktop personal computer, which cuts down the need for expensive hardware. However, there are still several limitations yet to be resolved. Based on the results obtained in this study and the extensive review of the literature, the improvements of the proposed cropland suitability assessment approach can be performed in four general directions, namely by:

- adopting performance evaluation for multiple crop types with the aim of determining the multidimensional cropland suitability dataset for a particular study area, presenting a complete solution for the agricultural land management;
- modification of the suitability assessment approach using high-resolution Sentinel-2 satellite images for the cropland suitability assessment at micro-locations;
- improvement of the present suitability assessment method considering the optimization of training samples and input covariates;
- implementation of the predicted soybean cropland suitability in practice considering present agricultural practices in the study area.

Cropland suitability classes predicted using the proposed method primarily reflect suitability consistency throughout multiple years. While the advantage of the relative

unsupervised classification of LAI and FAPAR values for the creation of training samples makes it resistant to the annual extremes caused by weather events, at the same time there is some ambiguity in the absolute suitability levels of these values. This could be addressed with the integration of yield data as the measure of absolute suitability levels with the proposed approach based on satellite-derived biophysical crop properties. LAI and FAPAR showed the sensitivity to vegetation properties of various crop types in previous studies, possibly presenting universal crop suitability indicators [65]. Gitelson [37] noted the FAPAR sensitivity to canopy structures and photosynthetic specificities of various crop types, including soybean and maize. The accurate determination of biophysical crop properties was maintained using the high spatial resolution Sentinel-2 and moderate spatial resolution Sentinel-3 products. This strongly indicates the applicability of the proposed method through multiple scales of interest for agricultural land management for major crop types. Moreover, the global coverage and open data availability of both Sentinel-2 and Sentinel-3 ensure widespread applicability of this method in the future [15]. Present methodology focuses on the larger agricultural parcels (above 10 ha) due to the restrictions of spatial resolution of PROBA-V products. Sentinel-2 images require additional processing but would enable expansion of the proposed methods on much smaller agricultural parcels. LAI and FAPAR derived from the 10 m spatial resolution Sentinel-2 images were successfully implemented in crop suitability studies [28], presenting a basis of the proposed method upgrade for the micro-scale analysis. The possible limitation during the creation of training and test samples using LAI and FAPAR as reference values might be the availability of soybean or other crop types ground truth agricultural parcels on the national level. Since these data are usually collected and distributed by national agencies, these data might not be easily accessible in some less developed parts of the world. However, this can be overcome by implementing crop type classification algorithms based on LAI and FAPAR using machine learning, which enabled the extraction of a particular crop type with 80% or higher accuracy in a recent study by Waldner et al. [65].

The yearly prediction accuracy trends from this study imply that the inclusion of the larger sample count over the larger area would result in higher prediction accuracy. Such an approach would also ensure the applicability of the proposed cropland suitability assessment method for minor crop types. The optimal training sample count in the restricted subset areas within Continental Croatia could be ensured for major crop types like soybean, wheat, maize, sunflower and rapeseed at the present time [40]. Additionally, the inclusion of additional covariates in the proposed cropland suitability assessment method could benefit the prediction accuracy, especially using RF [50]. The observations from sensitivity analysis considering the proximity to major land cover classes and soil types indicate that these covariates would be an important addition to future studies. Heterogeneity of suitability values of proximity zones to urban areas indicated a possible significant impact of socio-economic covariates in cropland suitability assessment, like population density [66]. Potential and actual evapotranspiration [8] and actual solar irradiation [67] were successfully derived using the free remote sensing data sources. These covariates would also likely improve the presently used theoretical values calculated from DEM for the majority of crop types besides soybean. Hengl et al. [50] noted the sensitivity of machine learning methods to inaccuracies in covariate data, which could have a significant impact on model performance. This indicates a necessity of accurate harmonization of input data during the resampling process, especially considering a more sensitive downscaling process, which should be considered during the addition of new covariates. The implementation of deep learning methods could present a viable option for the improvement of cropland suitability assessment accuracy in the future. At the present time, these methods generally lack computational efficiency of prediction due to the presence of large and complex training and covariate data [68]. Since this approach requires considerable and expensive hardware resources, it presently impairs the global and low-cost character of the proposed method. With the further improvement of deep

learning, it is expected that it will enable an upgrade to conventional machine learning, presenting an even more effective basis of cropland suitability assessment.

Another possible improvement of prediction accuracy of the machine learning methods is through implementing the most recent covariate data. The likely reasons for the lower prediction accuracy for the years 2019 and 2020 compared to the years prior were minor temporal disagreements of input covariates with the study period. Low mean cropland suitability values during the dry and hot year of 2018, compared to the years 2017 and 2019, further reinforces the need for accurate recent data due to sensitivity of suitability values to climate input. The application of the SoilGrids data referenced to the year 2017 was slightly obsolete for the prediction forthcoming years, since soil chemical properties like N and SOC are susceptible to temporal variations [69]. The ambiguities related to the inability to accurately track and model crop and soil management systems by farmers at the local scale indicate the necessity of including multiple data sources for the multi-year cropland suitability assessment. A possible solution is the integration of presently used SoilGrids data with the updated SoilGrids version 2.0 [70], which should establish a reliable global soil dataset for future studies. Similar observations were made about the climate data, which are subjected to the recent impact of climate change [4]. Even with the application of the most recent global climate data using the CHELSA dataset (1979–2013) [49], the effects of climate change within the past decade remain largely uncovered by previous cropland suitability assessment studies. Therefore, the comparison of multitemporal cropland suitability results, periodically updated with new climate data, would likely reflect climate change effects and allow farmers to make necessary adjustments. A possible solution for including the most recent climate change in the prediction could be the integration of present global climate datasets with the historic weather data within the study period [11]. This approach could enhance the proposed method by including the most recent climate trends using the freely accessible global weather data from a variety of online weather portals or national meteorological agencies.

## 5. Conclusions

The proposed cropland suitability assessment method based on machine learning represents a potential alternative and upgrade to conventional cropland suitability determination using a conventional GIS-based multicriteria analysis. Its advantages are primarily reflected in its computational efficiency, objectivity during the prediction and the ability to integrate complex input covariates. The proposed method is based on open remote sensing and GIS data and software, which makes it widely available worldwide. RF produced superior suitability assessment results to SVM in cases of moderate sample count and a high amount of complex input covariates. Its accuracy is expected to further grow with the inclusion of the additional covariates, including socio-economic covariates, evapotranspiration, solar irradiation and proximity to land cover classes. The creation of the study area larger than  $50 \times 50 \text{ km}^2$  is also expected to increase suitability assessment accuracy, due to the increased training sample count and better model fitting. The presence of the highly suitable S1 class per FAO classification was noted in the 6.1% of Subset A and 1.5% of Subset B. This observation encourages the re-evaluation of present agricultural land management plans, as the agricultural land in Subset A is presently not adequately utilized for soybean cultivation, contrary to the intensively cultivated agricultural land in Subset B.

The accurate and straightforward cropland suitability determination method is necessary to ensure a widely available solution for effective agricultural land management for the sustainability of agricultural production. The proposed method overcomes the limitations of the conventionally used GIS-based multicriteria analysis, and could turn the attention to machine learning in future cropland suitability determination studies. Future studies will be directed in its adjustment to various crop types and the scaling to micro-locations by implementing high-resolution Sentinel-2 images.

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

**Funding:** This research received no external funding.

**Acknowledgments:** This work was supported by the Faculty of Agrobiotechnical Sciences Osijek as a part of the scientific project: ‘AgroGIT—technical and technological crop production systems, GIS and environment protection’. This work was supported by the University of Zagreb as a part of the scientific project: “Advanced photogrammetry and remote sensing methods for environmental change monitoring” (Grant No. RS4ENVIRO).

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

## Appendix A

**Table A1.** Climate properties of subset areas per year in the 2017–2020 period between April and October.

Subset/Year	Air Temperature (April–October)			Precipitation (April–October)		
	Mean CHELSA	Annual	Difference from Mean	Mean CHELSA	Annual	Difference from Mean
A/2020	17.5 °C	17.1°C	−2.1%	547.6 mm	640.1 mm	+16.9%
A/2019		17.4°C	−0.3%		689.5 mm	+25.9%
A/2018		18.4°C	+5.2%		479.6 mm	−12.4%
A/2017		17.4°C	−0.5%		546.5 mm	−0.2%
B/2020	17.9 °C	17.8°C	−0.3%	449.2 mm	462.3 mm	+2.9%
B/2019		18.1°C	+1.3%		558.7 mm	+24.4%
B/2018		19.2°C	+7.2%		422.0 mm	−6.1%
B/2017		18.0°C	+0.8%		449.5 mm	+0.1%

**Table A2.** The accuracy assessment of downscaling methods of CHELSA mean air temperature and precipitation data compared to ground truth data from DHMZ stations.

CHELSA Dataset	Mean Air Temperature (°C)		Precipitation (mm)	
	R <sup>2</sup>	RMSE	R <sup>2</sup>	RMSE
Native (1000 m)	0.9513	0.9643	0.7190	43.3024
NN (300 m)	0.9507	0.9659	0.7137	43.9376
BI (300 m)	0.9512	0.9631	0.7128	43.9296
BSI (300 m)	0.9513	0.9646	0.7203	43.1707

NN: nearest neighbour, BI: bilinear interpolation, BSI: B-spline interpolation.

**Table A3.** Mean LAI and FAPAR values per suitability class after K-means unsupervised classification.

Year	Suitability Class	Subset A			Subset B		
		Elements	Mean LAI	Mean FAPAR	Elements	Mean LAI	Mean FAPAR
2020	S1	23	3.058	0.647	52	2.787	0.551
	S2	42	2.488	0.535	148	2.355	0.545
	S3	52	2.376	0.571	74	1.990	0.545
	N1	57	2.126	0.552	171	1.988	0.534
	N2	62	1.924	0.561	115	1.556	0.495
2019	S1	32	2.122	0.582	48	1.969	0.541
	S2	47	2.280	0.533	69	2.366	0.493
	S3	54	1.912	0.544	157	2.166	0.507
	N1	39	1.967	0.506	186	1.833	0.507
	N2	34	1.808	0.520	158	1.651	0.508

Table A3. Cont.

Year	Suitability Class	Subset A			Subset B		
		Elements	Mean LAI	Mean FAPAR	Elements	Mean LAI	Mean FAPAR
2018	S1	74	2.524	0.538	75	2.496	0.498
	S2	37	2.291	0.521	153	2.183	0.500
	S3	66	1.954	0.560	197	1.848	0.490
	N1	63	2.072	0.526	174	1.552	0.475
	N2	64	1.788	0.511	68	1.566	0.455
2017	S1	23	2.203	0.588	128	2.076	0.495
	S2	48	2.017	0.554	78	1.721	0.507
	S3	57	2.131	0.488	203	1.646	0.461
	N1	84	1.571	0.495	78	1.465	0.450
	N2	87	1.801	0.477	181	1.333	0.441

## References

- Bengochea Paz, D.; Henderson, K.; Loreau, M. Agricultural Land Use and the Sustainability of Social-Ecological Systems. *Ecol. Model.* **2020**, *437*, 109312. [CrossRef] [PubMed]
- Food and Agriculture Organization of the United Nations (Ed.) *The Future of Food and Agriculture: Trends and Challenges*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2017; ISBN 978-92-5-109551-5.
- Yu, J.; Wu, J. The Sustainability of Agricultural Development in China: The Agriculture–Environment Nexus. *Sustainability* **2018**, *10*, 1776. [CrossRef]
- Nelson, G.; Bogard, J.; Lividini, K.; Arsenault, J.; Riley, M.; Sulser, T.B.; Mason-D’Croz, D.; Power, B.; Gustafson, D.; Herrero, M.; et al. Income Growth and Climate Change Effects on Global Nutrition Security to Mid-Century. *Nat. Sustain.* **2018**, *1*, 773–781. [CrossRef]
- Tang, L.; Hayashi, K.; Kohyama, K.; Leon, A. Reconciling Life Cycle Environmental Impacts with Ecosystem Services: A Management Perspective on Agricultural Land Use. *Sustainability* **2018**, *10*, 630. [CrossRef]
- Jurišić, M.; Radočaj, D.; Šiljeg, A.; Antonić, O.; Živić, T. Current Status and Perspective of Remote Sensing Application in Crop Management. *J. Cent. Eur. Agric.* **2021**, *22*, 156–166. [CrossRef]
- Song, G.; Zhang, H. Cultivated Land Use Layout Adjustment Based on Crop Planting Suitability: A Case Study of Typical Counties in Northeast China. *Land* **2021**, *10*, 107. [CrossRef]
- Akpoti, K.; Kabo-bah, A.T.; Zwart, S.J. Agricultural Land Suitability Analysis: State-of-the-Art and Outlooks for Integration of Climate Change Analysis. *Agric. Syst.* **2019**, *173*, 172–208. [CrossRef]
- Harrison, P. (Ed.) *World Agriculture: Towards 2015/2030: Summary Report*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2002; ISBN 978-92-5-104761-3.
- United States Department of Agriculture. World Agricultural Supply and Demand Estimates. Available online: <https://www.usda.gov/oce/commodity/wasde/wasde0421.pdf> (accessed on 13 May 2021).
- Radočaj, D.; Jurišić, M.; Gašparović, M.; Plaščak, I. Optimal Soybean (*Glycine Max* L.) Land Suitability Using GIS-Based Multicriteria Analysis and Sentinel-2 Multitemporal Images. *Remote Sens.* **2020**, *12*, 1463. [CrossRef]
- Dedeoğlu, M.; Dengiz, O. Generating of Land Suitability Index for Wheat with Hybrid System Approach Using AHP and GIS. *Comput. Electron. Agric.* **2019**, *167*, 105062. [CrossRef]
- Jurišić, M.; Plaščak, I.; Antonić, O.; Radočaj, D. Suitability Calculation for Red Spicy Pepper Cultivation (*Capsicum Annum* L.) Using Hybrid GIS-Based Multicriteria Analysis. *Agronomy* **2020**, *10*, 3. [CrossRef]
- Binte Mostafiz, R.; Noguchi, R.; Ahamed, T. Agricultural Land Suitability Assessment Using Satellite Remote Sensing-Derived Soil-Vegetation Indices. *Land* **2021**, *10*, 223. [CrossRef]
- Radočaj, D.; Obhodaš, J.; Jurišić, M.; Gašparović, M. Global Open Data Remote Sensing Satellite Missions for Land Monitoring and Conservation: A Review. *Land* **2020**, *9*, 402. [CrossRef]
- Seyedmohammadi, J.; Sarmadian, F.; Jafarzadeh, A.A.; McDowell, R.W. Development of a Model Using Matter Element, AHP and GIS Techniques to Assess the Suitability of Land for Agriculture. *Geoderma* **2019**, *352*, 80–95. [CrossRef]
- Saaty, T.L.; Ozdemir, M.S. Why the Magic Number Seven plus or Minus Two. *Math. Comput. Model.* **2003**, *38*, 233–244. [CrossRef]
- Li, Z.; Fan, Z.; Shen, S. Urban Green Space Suitability Evaluation Based on the AHP-CV Combined Weight Method: A Case Study of Fuping County, China. *Sustainability* **2018**, *10*, 2656. [CrossRef]
- Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [CrossRef]
- Roell, Y.E.; Beucher, A.; Møller, P.G.; Greve, M.B.; Greve, M.H. Comparing a Random Forest Based Prediction of Winter Wheat Yield to Historical Yield Potential. *Agronomy* **2020**, *10*, 395. [CrossRef]
- Feyisa, G.L.; Palao, L.K.; Nelson, A.; Gumma, M.K.; Paliwal, A.; Win, K.T.; Nge, K.H.; Johnson, D.E. Characterizing and Mapping Cropping Patterns in a Complex Agro-Ecosystem: An Iterative Participatory Mapping Procedure Using Machine Learning Algorithms and MODIS Vegetation Indices. *Comput. Electron. Agric.* **2020**, *175*, 105595. [CrossRef]



22. Chemura, A.; Mutanga, O.; Dube, T. Separability of Coffee Leaf Rust Infection Levels with Machine Learning Methods at Sentinel-2 MSI Spectral Resolutions. *Precis. Agric.* **2017**, *18*, 859–881. [CrossRef]
23. Jiang, D.; Ma, T.; Ding, F.; Fu, J.; Hao, M.; Wang, Q.; Chen, S. Mapping Global Environmental Suitability for Sorghum Bicolor (L.) Moench. *Energies* **2019**, *12*, 1928. [CrossRef]
24. Gómez, D.; Salvador, P.; Sanz, J.; Casanova, J.L. Potato Yield Prediction Using Machine Learning Techniques and Sentinel 2 Data. *Remote. Sens.* **2019**, *11*, 1745. [CrossRef]
25. Taghizadeh-Mehrjardi, R.; Nabiollahi, K.; Rasoli, L.; Kerry, R.; Scholten, T. Land Suitability Assessment and Agricultural Production Sustainability Using Machine Learning Models. *Agronomy* **2020**, *10*, 573. [CrossRef]
26. Food and Agriculture Organization of the United Nations (FAO). A Framework for Land Evaluation, Chapter 3: Land Suitability Classifications. Available online: <http://www.fao.org/3/x5310e/x5310e04.htm> (accessed on 2 May 2021).
27. 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]
28. Ayu Purnamasari, R.; Noguchi, R.; Ahamed, T. Land Suitability Assessments for Yield Prediction of Cassava Using Geospatial Fuzzy Expert Systems and Remote Sensing. *Comput. Electron. Agric.* **2019**, *166*, 105018. [CrossRef]
29. Wannasek, L.; Ortner, M.; Amon, B.; Amon, T. Sorghum, a Sustainable Feedstock for Biogas Production? Impact of Climate, Variety and Harvesting Time on Maturity and Biomass Yield. *Biomass Bioenergy* **2017**, *106*, 137–145. [CrossRef]
30. Baldini, M.; Ferfuaia, C.; Zuliani, F.; Danuso, F. Suitability Assessment of Different Hemp (Cannabis Sativa L.) Varieties to the Cultivation Environment. *Ind. Crops Prod.* **2020**, *143*, 111860. [CrossRef]
31. Fensholt, R.; Sandholt, I.; Rasmussen, M.S. Evaluation of MODIS LAI, FAPAR and the Relation between FAPAR and NDVI in a Semi-Arid Environment Using in Situ Measurements. *Remote Sens. Environ.* **2004**, *91*, 490–507. [CrossRef]
32. Gitelson, A.A.; Peng, Y.; Huemmrich, K.F. Relationship between Fraction of Radiation Absorbed by Photosynthesizing Maize and Soybean Canopies and NDVI from Remotely Sensed Data Taken at Close Range and from MODIS 250m Resolution Data. *Remote Sens. Environ.* **2014**, *147*, 108–120. [CrossRef]
33. Khamala, E. *Review of the Available Remote Sensing Tools, Products, Methodologies and Data to Improve Crop Production Forecasts*; Food and Agriculture Organization of the United Nations: Rome, Italy, 2017.
34. Marshall, M.; Thenkabail, P. Developing in Situ Non-Destructive Estimates of Crop Biomass to Address Issues of Scale in Remote Sensing. *Remote Sens.* **2015**, *7*, 808–835. [CrossRef]
35. Casa, R.; Varella, H.; Buis, S.; Guérif, M.; De Solan, B.; Baret, F. Forcing a Wheat Crop Model with LAI Data to Access Agronomic Variables: Evaluation of the Impact of Model and LAI Uncertainties and Comparison with an Empirical Approach. *Eur. J. Agron.* **2012**, *37*, 1–10. [CrossRef]
36. Habyarimana, E.; Piccard, I.; Catellani, M.; De Franceschi, P.; Dall'Agata, M. Towards Predictive Modeling of Sorghum Biomass Yields Using Fraction of Absorbed Photosynthetically Active Radiation Derived from Sentinel-2 Satellite Imagery and Supervised Machine Learning Techniques. *Agronomy* **2019**, *9*, 203. [CrossRef]
37. Gitelson, A.A. Remote Estimation of Fraction of Radiation Absorbed by Photosynthetically Active Vegetation: Generic Algorithm for Maize and Soybean. *Remote Sens. Lett.* **2019**, *10*, 283–291. [CrossRef]
38. Fuster, B.; Sánchez-Zapero, J.; Camacho, F.; García-Santos, V.; Verger, A.; Lacaze, R.; Weiss, M.; Baret, F.; Smets, B. Quality Assessment of PROBA-V LAI, FAPAR and FCOVER Collection 300 m Products of Copernicus Global Land Service. *Remote Sens.* **2020**, *12*, 1017. [CrossRef]
39. Radočaj, D.; Jurišić, M.; Zebec, V.; Plaščak, I. Delineation of Soil Texture Suitability Zones for Soybean Cultivation: A Case Study in Continental Croatia. *Agronomy* **2020**, *10*, 823. [CrossRef]
40. Jurišić, M.; Radočaj, D.; Krčmar, S.; Plaščak, I.; Gašparović, M. Geostatistical Analysis of Soil C/N Deficiency and Its Effect on Agricultural Land Management of Major Crops in Eastern Croatia. *Agronomy* **2020**, *10*, 1996. [CrossRef]
41. Bogunovic, I.; Trevisani, S.; Seput, M.; Juzbasic, D.; Durdevic, B. Short-Range and Regional Spatial Variability of Soil Chemical Properties in an Agro-Ecosystem in Eastern Croatia. *Catena* **2017**, *154*, 50–62. [CrossRef]
42. Croatian Bureau of Statistics, Areas and Production of Cereals and Other Crops. 2020. Available online: [https://www.dzs.hr/Hrv\\_Eng/publication/2020/01-01-18\\_01\\_2020.htm](https://www.dzs.hr/Hrv_Eng/publication/2020/01-01-18_01_2020.htm) (accessed on 15 May 2021).
43. Galić Subašić, D. Influence of Irrigation, Nitrogen Fertilization and Genotype on the Yield and Quality of Soybean (*Glycine max* (L.) Merr.). Ph.D. Thesis, Josip Juraj Strossmayer University of Osijek, Faculty of Agrobiotechnical Sciences Osijek, Osijek, Croatia, 2018.
44. Liu, X.; Jin, J.; Herbert, S.J.; Zhang, Q.; Wang, G. Yield Components, Dry Matter, LAI and LAD of Soybeans in Northeast China. *Field Crops Res.* **2005**, *93*, 85–93. [CrossRef]
45. Yadav, K.; Congalton, R.G. Accuracy Assessment of Global Food Security-Support Analysis Data (GFSAD) Cropland Extent Maps Produced at Three Different Spatial Resolutions. *Remote Sens.* **2018**, *10*, 1800. [CrossRef]
46. Hsieh, P.F.; Lee, L.C.; Chen, N.Y. Effect of Spatial Resolution on Classification Errors of Pure and Mixed Pixels in Remote Sensing. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 2657–2663. [CrossRef]
47. Anthony, P.; Malzer, G.; Sparrow, S.; Zhang, M. Soybean Yield and Quality in Relation to Soil Properties. *Agron. J.* **2012**, *104*, 1443–1458. [CrossRef]
48. Müller, M.; Schneider, J.R.; Klein, V.A.; da Silva, E.; da Silva Júnior, J.P.; Souza, A.M.; Chavarria, G. Soybean Root Growth in Response to Chemical, Physical, and Biological Soil Variations. *Front. Plant Sci.* **2021**, *12*, 272. [CrossRef] [PubMed]



49. Karger, D.N.; Conrad, O.; Böhrner, J.; Kawohl, T.; Kreft, H.; Soria-Auza, R.W.; Zimmermann, N.E.; Linder, H.P.; Kessler, M. Climatologies at High Resolution for the Earth's Land Surface Areas. *Sci. Data* **2017**, *4*, 170122. [[CrossRef](#)] [[PubMed](#)]
50. Hengl, T.; de Jesus, J.M.; Heuvelink, G.B.M.; Gonzalez, M.R.; Kilibarda, M.; Blagotić, A.; Shangguan, W.; Wright, M.N.; Geng, X.; Bauer-Marschallinger, B.; et al. SoilGrids250m: Global Gridded Soil Information Based on Machine Learning. *PLoS ONE* **2017**, *12*, e0169748. [[CrossRef](#)] [[PubMed](#)]
51. EU-DEM v1.1—Copernicus Land Monitoring Service. Available online: <https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1> (accessed on 21 April 2021).
52. PROBA-V Products User Manual v3.01. Available online: [https://proba-v.vgt.vito.be/sites/proba-v.vgt.vito.be/files/products\\_user\\_manual.pdf](https://proba-v.vgt.vito.be/sites/proba-v.vgt.vito.be/files/products_user_manual.pdf) (accessed on 21 April 2021).
53. Liu, Y.; Dai, L. Modelling the Impacts of Climate Change and Crop Management Measures on Soybean Phenology in China. *J. Clean. Prod.* **2020**, *262*, 121271. [[CrossRef](#)]
54. Liu, S.; Zhang, P.; Marley, B.; Liu, W. The Factors Affecting Farmers' Soybean Planting Behavior in Heilongjiang Province, China. *Agriculture* **2019**, *9*, 188. [[CrossRef](#)]
55. Böhrner, J.; Antonić, O. Land-Surface Parameters Specific to Topo-Climatology. In *Developments in Soil Science*; Hengl, T., Reuter, H.I., Eds.; Geomorphometry; Elsevier: Amsterdam, The Netherlands, 2009; Chapter 8; Volume 33, pp. 195–226.
56. Stein, A.; Riley, J.; Halberg, N. Issues of Scale for Environmental Indicators. *Agric. Ecosyst. Environ.* **2001**, *87*, 215–232. [[CrossRef](#)]
57. Liu, H.; Weng, Q. Scaling Effect of Fused ASTER-MODIS Land Surface Temperature in an Urban Environment. *Sensors* **2018**, *18*, 4058. [[CrossRef](#)]
58. Peng, S.; Ding, Y.; Liu, W.; Li, Z. 1 Km Monthly Temperature and Precipitation Dataset for China from 1901 to 2017. *Earth Syst. Sci. Data* **2019**, *11*, 1931–1946. [[CrossRef](#)]
59. Dabija, A.; Kluczek, M.; Zagajewski, B.; Raczko, E.; Kycko, M.; Al-Sulttani, A.H.; Tardà, A.; Pineda, L.; Corbera, J. Comparison of Support Vector Machines and Random Forests for Corine Land Cover Mapping. *Remote Sens.* **2021**, *13*, 777. [[CrossRef](#)]
60. Colditz, R.R. An Evaluation of Different Training Sample Allocation Schemes for Discrete and Continuous Land Cover Classification Using Decision Tree-Based Algorithms. *Remote Sens.* **2015**, *7*, 9655–9681. [[CrossRef](#)]
61. Grabska, E.; Hostert, P.; Pflugmacher, D.; Ostapowicz, K. Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sens.* **2019**, *11*, 1197. [[CrossRef](#)]
62. Pontius, R.G.; Millones, M. Death to Kappa: Birth of Quantity Disagreement and Allocation Disagreement for Accuracy Assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429. [[CrossRef](#)]
63. Food and Agriculture Organization of the United Nations (FAO). A Framework for Land Evaluation, Chapter 7: Land Suitability Assessment. Available online: <http://www.fao.org/3/t0741e/T0741E10.htm> (accessed on 16 May 2021).
64. Wolanin, A.; Camps-Valls, G.; Gómez-Chova, L.; Mateo-García, G.; van der Tol, C.; Zhang, Y.; Guanter, L. Estimating Crop Primary Productivity with Sentinel-2 and Landsat 8 Using Machine Learning Methods Trained with Radiative Transfer Simulations. *Remote Sens. Environ.* **2019**, *225*, 441–457. [[CrossRef](#)]
65. Waldner, F.; Lambert, M.-J.; Li, W.; Weiss, M.; Demarez, V.; Morin, D.; Marais-Sicre, C.; Hagolle, O.; Baret, F.; Defourny, P. Land Cover and Crop Type Classification along the Season Based on Biophysical Variables Retrieved from Multi-Sensor High-Resolution Time Series. *Remote Sens.* **2015**, *7*, 10400–10424. [[CrossRef](#)]
66. Møller, A.B.; Mulder, V.L.; Heuvelink, G.B.M.; Jacobsen, N.M.; Greve, M.H. Can We Use Machine Learning for Agricultural Land Suitability Assessment? *Agronomy* **2021**, *11*, 703. [[CrossRef](#)]
67. Gašparović, I.; Gašparović, M.; Medak, D. Determining and Analysing Solar Irradiation Based on Freely Available Data: A Case Study from Croatia. *Environ. Dev.* **2018**, *26*, 55–67. [[CrossRef](#)]
68. Darwin, B.; Dharmaraj, P.; Prince, S.; Popescu, D.E.; Hemanth, D.J. Recognition of Bloom/Yield in Crop Images Using Deep Learning Models for Smart Agriculture: A Review. *Agronomy* **2021**, *11*, 646. [[CrossRef](#)]
69. Pu, X.; Xie, J.; Cheng, H.; Yang, S. Temporal Trends of Soil Organic Carbon and Total Nitrogen Losses in Seasonally Frozen Zones of Northeast China: Responses to Long-Term Conventional Cultivation (1965–2010). *Environ. Process.* **2014**, *1*, 415–429. [[CrossRef](#)]
70. Poggio, L.; de Sousa, L.M.; Batjes, N.H.; Heuvelink, G.; 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](#)]

# GENERAL DISCUSSION

---

Previous studies dominantly regarded GIS-based multicriteria analysis as the current standard of quantifying cropland suitability (Dedeoğlu and Dengiz, 2019; Layomi Jayasinghe et al., 2019; Mandal et al., 2020; Jurišić et al., 2020). Its flexibility and global applicability made it an indispensable method in suitability studies in a wide field of scientific disciplines. In addition to agriculture, it has frequent applications in forestry, ecology, healthcare and economics (Jurišić et al., 2021). However, fundamental flaws of this approach reduce its effectiveness with the increasing demand for reliability and global applicability. Their three major disadvantages were addressed in this dissertation, with their impact on the cropland suitability being analyzed according to major processing steps in this chapter. The standard procedure of GIS multicriteria analysis consists of six basic steps (Šiljeg et al., 2020):

- determining the aim of the analysis,
- selection of suitability criteria,
- standardization of input values,
- weighting of criteria,
- suitability calculation,
- sensitivity analysis.

The aim of GIS-based multicriteria analysis in previous studies was set to determine the suitability of agricultural crops in a particular geographical area. An important component of the analyzed studies is the temporal definition of the analysis, which is conditioned by the availability of data from one or more consecutive sowing seasons (Feyisa et al., 2020). Knowledge of the time span of the analysis is also necessary due to the temporal variability of the basic abiotic factors, primarily climatic and pedological (Hengl et al., 2017). The proposed methods, which integrate machine learning and open data satellite imagery, were founded on the same assumptions. This property enables a multitemporal comparison of suitability results from the two approaches, allowing historical accuracy assessment and update of cropland management plans. With the development of satellite missions with open data access, the adaptation of cropland management to the classification of abiotic factors (climate, soil and topography) became increasingly accessible (Radočaj et al., 2020), which was utilized in the dissertation. Their availability has a strong perspective for future decades, given the longevity of existing satellite missions and their continuous development and upgrade. These facts allow regular updating of the cropland suitability results for cultivating certain crops, which is necessary due to the temporal variability of climate and soil properties caused by climate change and inadequate cropland management (Chemura et al., 2020). Previous cropland suitability studies in continental Croatia indicated high variability in the suitability levels for cultivating certain crops (Jurišić et al., 2020; Šiljeg et al., 2020). Farmers' target yields based on relative comparisons with neighboring agricultural parcels are an additional factor that may result in increased application of fertilizers and pesticides in naturally unsuitable locations. Therefore, computationally efficient suitability determination methods based on machine learning should be increasingly in demand.

The selection of abiotic criteria for agroecological suitability modeling using GIS-based multicriteria analysis in previous research was conducted by combining the analysis of previous studies and expert opinion (Dedeoğlu and Dengiz, 2019). This procedure was based on the assumption that each micro location for the cultivation of an agricultural crop is agroecologically specific (Mandal et al., 2020). The selection of type and quantity of abiotic criteria for a particular location and crop type was based on the knowledge of one or more agronomic experts. Abiotic criteria were regularly divided into climate, soil and topography criteria groups, representing three major impacts on cropland suitability (Layomi Jayasinghe et al., 2019; Dedeoğlu and Dengiz, 2019; Møller et al., 2021). Although most of these studies focused on these criteria groups, their quantity differed significantly depending on the crop type and geographical location (Table 1). The exceptions to the application of climate, soil and topography criteria were constraints, representing suitable land cover classes and irrigation system categories. The selection of climatic criteria directly depends on the extent of the study area, whereas for smaller areas (like municipalities) climate homogeneity is assumed (Jurišić et al., 2020).

Table 1. Distribution of three major criteria groups in previous studies based on cropland suitability determination using GIS-based multicriteria analysis.

Crop type	Criteria group content (%)			Total criteria	Reference
	climate	soil	topography		
tea	22%	22%	56%	9	Layomi Jayasinghe et al. (2019)
maize, rice	17%	33%	50%	8	Mandal et al. (2020)
soybean, maize	27%	18%	55%	16	Song and Zhang (2021)
wheat	0%	20%	80%	10	Dedeoğlu and Dengiz (2019)
maize	88%	0%	13%	8	Chemura et al. (2020)
pepper	0%	14%	86%	7	Jurišić et al. (2020)
barley	20%	50%	30%	10	Šiljeg et al. (2020)

The variability in the number of abiotic suitability criteria used and their distribution among the criteria groups indicate a high influence of human subjectivity in their selection. Although this approach allows for effective benefit modeling based on expert knowledge, it is potentially unreliable and biased (Taghizadeh-Mehrjardi et al., 2020). The computational inefficiency of this approach is expressed by the deviation from seven criteria as the optimal number in AHP, optionally ranging from five to nine criteria per Saaty and Ozdemir (2003). In line with these recommendations, the application of fewer than four criteria allow only a limited representation of cropland suitability. Ten or more criteria represent a wider range of abiotic criteria but also increase the possibility of error and computational complexity in subjective comparison of the relative importance of criteria. Both subjectivity and inability of integrating environmental data in the GIS environment were effectively resolved with the application of machine learning. It supported the processing of big data, as well as the integration of their various types, establishing complex nonlinear relationships between training data and independent predictors (covariates) (Hengl et al., 2017). Contrary to manual and subjective

weight calculation of individual abiotic criteria on suitability result (Layomi Jayasinghe et al., 2019; Jurišić et al., 2020), machine learning approach in this dissertation allowed fully automated and objective feature importance determination.

Spatial modeling of selected abiotic criteria in the GIS environment is most often performed in a raster data model (Jurišić et al., 2020). The input data are commonly distributed in numerous combinations of data types, according to various institutional and scientific sources. Spatial modeling of complex abiotic criteria can also cause human-based errors during their preprocessing and transfer them to the final suitability result. Soil texture, as a frequently used criterion in cropland suitability studies, was partially represented in previous studies by global standards with subjective modifications (Layomi Jayasinghe et al., 2019) or application of different standards with the same goal (Taghizadeh-Mehrjardi et al., 2020; Møller et al., 2021). Since one of the most common approaches to selecting criteria is the analysis of previous studies, such cases cause potential inaccuracy in the selection of value ranges in further standardization and weighting procedures. The input data for soil texture is point vector soil sampling data, containing the value of clay, silt and sand content. To convert this data into a raster form, it is necessary to perform a prediction of soil values at unsampled locations by spatial interpolation (Song and Zhang, 2021). The selection of the optimal method and parameters of spatial interpolation was a necessity for reliable modeling of input criteria, which decreases significantly if these are not adjusted to the properties of input values (Jurišić et al., 2020). The relative complexity of soil texture modeling, as well as subjectivity in the selection of spatial interpolation method, parameters and classification standards, indicates a potential reduction in human error by automating the process (Hengl et al., 2017). The same approach in this dissertation increased the time efficiency since it does not require individual tool processing in GIS and facilitates data distribution using a globally accepted standard. By the same principle, the developed processing framework is easily adaptable to various abiotic criteria in cropland suitability studies.

In the standardization process, heterogeneous input value ranges of spatially modeled abiotic criteria were transformed to a uniform numerical standardization interval (Layomi Jayasinghe et al., 2019). Numerical intervals such as  $[0,1]$ , or more often  $[1,5]$  are commonly used, which allows straightforward representation of suitability with the five classes defined by the Food and Agriculture Organization of the United Nations (FAO) standard. Aside from combining values expressed in various measurement units, it also integrates both quantitative and qualitative data, which is often necessary for determining cropland suitability (Dedeoğlu and Dengiz, 2019). Three basic standardization methods were used in previous studies: linear stretching, stepwise standardization and fuzzy standardization. The minimum and maximum input values in linear stretching correspond to the limit values of the defined standardization interval. Although very simple and completely objective, the linear stretching method produces unreliable standardization if the input data contain extreme values, which is often the case in suitability studies. On the contrary, the stepwise standardization method is a completely subjective method, based on discrete ranges of input values for an individual standardized value. The suitability level is thus quantified by generalized and approximated numerical values, mostly of equal coverage (Jurišić et al., 2020). Due to the simplicity and flexibility of the procedure, this method has found the most frequent application in previous cropland

suitability studies (Layomi Jayasinghe et al., 2019; Šiljeg et al., 2020). Standardization by the fuzzy method combines the advantages of the previous two methods with continuous standardization and relative objectivity using mathematical models, as well as implementing standardization thresholds based on a subjective approach (Mandal et al., 2020). Alternatives in choosing mathematical models of fuzzy logic (linear, S-shaped, J-shaped, and G-shaped) allow for additional standardization flexibility. While the majority of previous studies implemented a simple stepwise standardization (Mandal et al., 2020; Šiljeg et al., 2020), fuzzy method allowed greater flexibility and produced more accurate suitability results in this research. Nevertheless, fuzzy logic methods are much less commonly used in suitability studies compared to the stepwise standardization method. The exact impact of standardization methods on the accuracy of suitability results was not found in the existing literature, leaving users a subjective choice of standardization method without a scientific basis. The comparative assessment of these standardization methods in this dissertation proved that the variety of available methods in the complex GIS-based multicriteria analysis should be more thoroughly evaluated in future studies.

In contrast to standardization, which evaluates suitability within a single criterion according to its value range, criteria weights quantify the relative importance of all selected suitability criteria (Saaty and Ozdemir, 2003). Since input abiotic factors affect the cropland suitability to a varying degree, weights are assigned to all criteria to proportionally express their impact on final suitability. Sensitivity analyses in previous studies found the criteria weighting had the highest impact on cropland suitability results in the GIS-based multicriteria analysis (Dedeoğlu and Dengiz, 2019). There is a number of alternative methods for the weighting of abiotic criteria, ranging from simple estimation methods to advanced methods such as AHP, TOPSIS, ELECTRE, and PROMETHEE (Layomi Jayasinghe et al., 2019). A common feature of all weighting methods is that the sum of all weights equals 1, denoting 100% of the influence of selected input abiotic criteria on suitability. Previous studies noted the advantages of AHP in terms of flexibility and simplicity (Šiljeg et al., 2020; Song and Zhang, 2021), causing its dominant selection in cropland suitability studies (Table 2).

Table 2. The application of criteria weighting methods in cropland suitability papers indexed in Web of Science Core Collection during 2000–2020.

Method	Published papers indexed in Web of Science Core Collection	
	2000–2020	2010–2020
AHP	160	152
TOPSIS	7	7
PROMETHEE	4	4
ELECTRE	3	3
machine learning	20	20

The principle of AHP is based on the relative pairwise comparison of all combinations of input criteria, denoting a proportionally more influential abiotic factor with an integer in the interval [1,9]. Although each result of the pairwise comparison is checked by consistency index, in the case of more than nine criteria, the weighting process becomes too complex and prone to inconsistency. Previous suitability studies have identified the difficulties of comprehensive

suitability modeling with the recommended number of abiotic factors (Dedeoğlu and Dengiz, 2019; Song et al., 2021). The susceptibility of weighting to human subjective assessments in a high number of pairwise comparisons and the inability to select an arbitrary number of criteria are currently the two biggest shortcomings of the application of AHP in GIS-based multicriteria analysis. The application of machine learning in the method proposed in this study resolved these issues, removing the most impactful source of subjective sources on the cropland suitability results.

The suitability calculation based on standardized values of input abiotic factors and their weights is the simplest and most uniform step in GIS-based multicriteria analysis (Dedeoğlu and Dengiz, 2019; Jurišić et al., 2020). The weighted linear combination is a conventional choice for calculating suitability, multiplying the standardized values and their respective weights. The range of suitability values corresponds to an arbitrarily selected numerical standardization interval. Many authors emphasized the importance of applying globally accepted standards in classifying the cropland suitability for cultivating a particular crop type. The FAO classification of cropland suitability into five classes is considered the standard in numerous studies (Layomi Jayasinghe et al., 2019; Šiljeg et al., 2020). Its application facilitated the comparison of suitability values between crops and for different locations for the same crop.

While it is a necessary procedure in all similar forms of spatial analysis, accuracy assessment is often omitted from cropland suitability studies using GIS-based multicriteria analysis. There is a very limited set of data that can be used to validate the complex concept of cropland suitability (Taghizadeh-Mehrjardi et al., 2020). In the majority of previous studies in which the accuracy assessment has been applied, crop yield data have been used as a real measure of suitability (Dedeoğlu and Dengiz, 2019). It is also affected by components that cannot be modeled in a GIS environment, such as the implementation of agrotechnical operations at the micro level, which makes it an incomplete indicator of suitability. The availability of yield data per agricultural parcel is very limited and uneven globally. In Croatia, there is currently no official and reliable database of yield data for individual agricultural parcels in a continuous area. Therefore, the implementation of conventional cropland suitability studies is very limited, subjected to the subjectivity of individual experts and without an external and objective accuracy assessment of the suitability results. Although previous suitability studies based on machine learning indicated superior performance for suitability calculation over GIS-based multicriteria analysis, present development only partially addressed its fundamental shortcomings. This primarily refers to the lack of reference parameters for the validation of suitability results, which is based on the same assumptions as for the conventional approach. Frampton et al. (2013) reported the possibility of calculating various biophysical variables from satellite multispectral data, such as leaf area index (LAI), the fraction of absorbed photosynthetically absorbed radiation (FAPAR), and the canopy chlorophyll content. A very important feature of these data is the high correlation with crop yields in individual growth stages, representing the possibility of using these data for training and validation of suitability models (Habyarimana et al., 2019). The importance of LAI and FAPAR is emphasized in determining the impact of climate change on vegetation according to United Nations standards, indicating high potential in suitability studies (Frampton et al., 2013). These possibilities were further enhanced by the launch of Sentinel-2 and Sentinel-3 satellite



multispectral missions, which enable the modelling of biophysical vegetation variables in high and medium spatial resolution. The application of remote sensing data additionally avoids expensive and time-consuming collection of ground-truth data by terrestrial methods, especially in the case of larger areas and less developed transport infrastructure, which is characteristic of most agricultural parcels (Feyisa et al., 2020).

# CONCLUSIONS

---

Two of the most commonly used approaches to increase crop yield production include expanding cropland by transforming land cover and increased application of fertilizers and pesticides, which is not sustainable in the long term and threatens biodiversity. An alternative approach to increase crop yield while ensuring agricultural sustainability focuses on improving cropland management, by cultivating each crop in a naturally suitable location, which also decreases the need for fertilizers and pesticides. The expected scientific contribution of this dissertation is a more reliable, objective and computationally efficient procedure for determining the cropland suitability compared to conventional GIS-based multicriteria analysis. Three methods of spatial data analysis were developed:

1. A computationally efficient method of suitability validation using global satellite missions of high (Sentinel-2) and medium spatial resolution (PROBA-V);
2. An automated method of spatial modeling of abiotic criteria on the example of soil texture according to a globally accepted standard;
3. Suitability prediction method based on machine learning algorithms and globally available spatial data which allow high reliability of prediction with reduced subjectivity of users in relation to GIS-based multicriteria analysis.

The upgrade of the conventional GIS-based multicriteria analysis was performed primarily by developing and analyzing the method of validation of suitability results using the NDVI vegetation index from Sentinel-2 multispectral images. During the full ripening (R6) growth stage, the peak NDVI values of each soybean parcel were determined. Based on previous research, the highest correlation of NDVI with soybean grain yield was determined at this growth stage, which has a high potential for an effective and widely available alternative to the conventional validation method. In addition, the impact of the standardization method on the accuracy of cropland suitability determination was analyzed, which was not yet performed in previous research. By choosing the optimal spatial interpolation method and parameters for climate and soil criteria, suitability for smaller datasets was modelled, which is characteristic for local administrative units. The cropland suitability for soybean cultivation was classified by the K-means method of unsupervised classification for compliance with the FAO standard of suitability classification. This approach is applicable for any crop type which produces a high correlation between any satellite-derived vegetation index and ground-truth yield components during one or more growth stages. Due to the open data availability of Sentinel-2 and similar missions, the possibilities of cropland suitability accuracy assessment became accessible for the majority of future studies.

Automating the process of spatial modeling of soil texture criterion was performed based on 255 soil samples on agricultural land in continental Croatia. By determining the optimal method and parameters of spatial interpolation according to five randomly created training and test data sets, the transformation of input data into a raster form was performed. Based on interpolated rasters of clay, silt and sand, a Python script for spatial georeferencing and automated classification of soil texture according to the USDA classification in 12 classes was developed. Testing of the procedure for the purpose of determining the cropland suitability

for soybean cultivation was performed by analyzing the suitability of individual classes of soil texture from previous studies, resulting in standardized suitability values in accordance with the FAO standard. The proposed procedure represents an example of the freely available algorithm for the automation of the component of suitability determination, which might be a foundation of an open-access library for more reliable and time-efficient spatial processing.

LAI and FAPAR from PROBA-V satellite mission data were selected to form reference data in a novel approach of machine learning cropland suitability prediction due to the high correlation with biomass and crop yield. The suitability classes of the reference data were determined based on their values on historical soybean agricultural parcels by the unsupervised K-means classification. Random division of this data formed training and test samples, which also satisfied the need to validate the suitability results. Based on training data and covariates representing climate, soil and topography criteria, RF and SVM machine learning methods were used to determine the suitability level on the entire agricultural land in the study area. The relative importance of input abiotic criteria was assessed after classification by machine learning methods, which represent an objectively determined equivalent of AHP weights from the conventional approach of GIS-based multicriteria analysis. The accuracy assessment of individual annual suitability results was performed on the basis of the figure of merit and overall accuracy, with rasters determined by the optimal machine learning method being singled out for further processing. Their values were aggregated within the unsupervised classification by the K-means method, resulting in the final suitability classes. This approach resolved most major disadvantages of the GIS-based multicriteria analysis, producing a completely objective results, allowing the integration of big and complex spatial data and enabling cropland suitability accuracy assessment using the freely available satellite data. Due to the requirement of training and test data, this method is applicable for all major crop types but can lack for less common crops since these are cultivated on a lower number of smaller agricultural parcels. Future studies will be directed on improving the spatial resolution of the proposed method by implementing biophysical variables from Sentinel-2, allowing up to 10 m spatial resolution. This would be a complimentary analysis to one presented in the dissertation, being applicable on a much larger scale and possibly representing a base for variable rate application in precision agriculture for agricultural parcels larger than 10 ha. While proposed methods were fundamentally developed for the application for all major crop types, future studies will also be directed on their evaluation for other major crops, as well as their accuracy under different agricultural land management systems.

## REFERENCES

---

- Chemura A, Schauburger B, Gornott C. 2020. Impacts of climate change on agro-climatic suitability of major food crops in Ghana. *Plos One* 15(6):e0229881.
- Dedeoğlu M, Dengiz O. 2019. Generating of land suitability index for wheat with hybrid system approach using AHP and GIS. *Comput Electron Agric* 167:105062.
- Feyisa GL, Palao LK, Nelson A, Gumma MK, Paliwal A, Win KT, Nge KH, Johnson DE. 2020. Characterizing and mapping cropping patterns in a complex agro-ecosystem: An iterative participatory mapping procedure using machine learning algorithms and MODIS vegetation indices. *Comput Electron Agric* 175:105595.
- Frampton WJ, Dash J, Watmough G, Milton EJ. 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS J Photogramm Remote Sens* 82:83-92.
- Habyarimana E, Piccard I, Catellani M, De Franceschi P, Dall'Agata M. 2019. Towards predictive modeling of sorghum biomass yields using fraction of absorbed photosynthetically active radiation derived from Sentinel-2 satellite imagery and supervised machine learning techniques. *Agronomy* 9(4):203.
- Hengl T, Mendes de Jesus J, Heuvelink GB, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, Shangquan W, Wright MN, Geng X, Bauer-Marschallinger B, Guevara MA, Vargas R, MacMillan RA, Batjes NH, Leenaars JGB, Ribeiro E, Wheeler I, Mantel S, Kempen, B. 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One* 12(2):e0169748.
- Jurišić M, Plaščak I, Antonić O, Radočaj D. 2020. Suitability Calculation for Red Spicy Pepper Cultivation (*Capsicum annum* L.) Using Hybrid GIS-Based Multicriteria Analysis. *Agronomy* 10(1):3.
- Jurišić M, Glavaš J, Plaščak I, Antonić O, Radočaj D. 2021. Geoinformacijske tehnologije GIS u ekonomiji. Ekonomski fakultet u Osijeku, Osijek, 279 pp.
- Layomi Jayasinghe S, Kumar L, Sandamali J. 2019. Assessment of potential land suitability for tea (*Camellia sinensis* (L.) O. Kuntze) in Sri Lanka using a GIS-based multi-criteria approach. *Agriculture* 9(7):148.
- Mandal VP, Rehman S, Ahmed R, Masroor M, Kumar P, Sajjad H. 2020. Land suitability assessment for optimal cropping sequences in Katihar district of Bihar, India using GIS and AHP. *Spat Inf Res* 28(5):589-599.
- Møller AB, Mulder VL, Heuvelink GB, Jacobsen NM, Greve MH. 2021. Can We Use Machine Learning for Cropland Suitability Assessment?. *Agronomy* 11(4):703.
- Nikolić T, Topić J. 2005. The red book of Croatian vascular flora. Ministry of Culture, State Institute for Nature Protection, Zagreb, 693 pp.
- Official Gazette. 71/2019. Ordinance on the protection of cropland from contamination: [https://narodne-novine.nn.hr/clanci/sluzbeni/2019\\_07\\_71\\_1507.html](https://narodne-novine.nn.hr/clanci/sluzbeni/2019_07_71_1507.html)

Progênio MF, da Costa Filho FAM, Crispim DL, Raiol Souza MJ, Pimentel da Silva GD, Fernandes LL. 2020. Ranking sustainable areas for the development of tidal power plants: A case study in the northern coastline of Brazil. *Int J Energy Res* 44(12):9772-9786.

Radočaj D, Obhodaš J, Jurišić M, Gašparović M. 2020. Global Open Data Remote Sensing Satellite Missions for Land Monitoring and Conservation: A Review. *Land* 9(11):402.

Saaty TL, Ozdemir MS. 2003. Why the magic number seven plus or minus two. *Math comput model* 38(3-4):233-244.

Song G, Zhang H. 2021. Cultivated Land Use Layout Adjustment Based on Crop Planting Suitability: A Case Study of Typical Counties in Northeast China. *Land* 10(2):107.

Šiljeg A, Jurišić M, Radočaj D, Videković M. 2020. Modeliranje pogodnosti poljoprivrednog zemljišta za uzgoj ječma uporabom višekriterijske GIS analize. *Poljoprivreda* 26(1):40-47.

Taghizadeh-Mehrjardi R, Nabiollahi K, Rasoli L, Kerry R, Scholten T. 2020. Land suitability assessment and agricultural production sustainability using machine learning models. *Agronomy* 10(4):573.

# CURRICULUM VITAE

---

Dorijan Radočaj was born on March 2, 1994 in Osijek. He enrolled at the Faculty of Geodesy, the University of Zagreb in 2012 and graduated in 2017, earning the title of Master of Engineering in Geodesy and Geoinformatics, course Geoinformatics. During his undergraduate and graduate study, he was the scholarship recipient by the Republic of Croatia and the University of Zagreb in the category of excellence (category A). He achieved a grade point average of 5,000 during the graduate study, with three Faculty of Geodesy awards for excellence as a student with the highest grade point average in a generation. As a student demonstrator, he did four semesters at the Department of Surveying.

He was hired in January 2018 as a research and teaching assistant at the Department of Surveying in the Faculty of Geodesy, University of Zagreb. During 2018, he cooperated with the Ministry of Defense of the Republic of Croatia on software development and field testing of ammunition. In December 2018, he was employed as a research and teaching assistant at the Faculty of Agrobiotechnical Sciences Osijek, Department of Geoinformation Technologies and GIS, where he is currently employed. He enrolled in the postgraduate university study of Environmental Protection and Nature Conservation in 2020.

He participates in teaching in ten undergraduate and graduate courses and has mentored three bachelor theses. He held two courses related to the implementation of geoinformation technologies in agriculture and environmental protection in Croatia, as well as multiple field demonstration activities with the Ministry of Agriculture. He was awarded by the Faculty of Agrobiotechnical Sciences Osijek as the first author of a scientific paper with the largest impact factor in the academic year 2019/2020 and for successful young scientist for scientific papers published in Q1 and Q2 in the academic year 2020/2021. He is an authorized operator of the unmanned aerial vehicle system in the C1 category by the Croatian Civil Aviation Agency. He became a certified Microsoft database associate in 2015. He is a collaborator on several professional projects in Osijek-Baranja County on the topic of digitalization of agriculture and the use of unmanned aerial vehicles for family farms. He has a certificate of knowledge of English at the B2 level and actively uses it.

He is the author and co-author of scientific and professional publications as follows:

University books and manuals (2):

1. **Radočaj D**, Jurišić M, Plaščak I. 2021. Geoinformacijske tehnologije GIS u poljoprivredi i zaštiti okoliša – Praktikum. Fakultet agrobiotehničkih znanosti Osijek, Osijek, 124 pp.



2. Jurišić M, Glavaš J, Plaščak I, Antonić O, **Radočaj D**. 2021. Geoinformacijske tehnologije GIS u ekonomiji. Ekonomski fakultet u Osijeku, Osijek, 279 pp.

Papers indexed in Web of Science – Current Contents (CC) and Science Citation Index Expanded (SCIE) (12):

1. **Radočaj D**, Jurišić M, Gašparović M. 2022. The Role of Remote Sensing Data and Methods in a Modern Approach to Fertilization in Precision Agriculture. *Remote Sens* 14(3):778.
2. **Radočaj D**, Jurišić M, Gašparović M. 2022. A wildfire growth prediction and evaluation approach using Landsat and MODIS data. *J Environ Manage* 304:114351.
3. **Radočaj D**, Plaščak I, Heffer G, Jurišić M. 2022. A Low-Cost Global Navigation Satellite System Positioning Accuracy Assessment Method for Agricultural Machinery. *Appl Sci* 12(2):693.
4. **Radočaj D**, Jug I, Vukadinović V, Jurišić M, Gašparović M. 2021. The Effect of Soil Sampling Density and Spatial Autocorrelation on Interpolation Accuracy of Chemical Soil Properties in Arable Cropland. *Agronomy* 11(12):2430.
5. **Radočaj D**, Jurišić M, Gašparović M, Plaščak I, Antonić O. 2021. Cropland Suitability Assessment Using Satellite-Based Biophysical Vegetation Properties and Machine Learning. *Agronomy* 11(8):1620.
6. **Radočaj D**, Jurišić M, Antonić O. 2021. Determination of soil C: N suitability zones for organic farming using an unsupervised classification in eastern Croatia. *Ecol Indic* 123:107382.
7. Jurišić M, **Radočaj D**, Krčmar S, Plaščak I, Gašparović M. 2020. Geostatistical analysis of soil C/N deficiency and its effect on agricultural land management of major crops in eastern Croatia. *Agronomy* 10(12):1996.
8. **Radočaj D**, Obhodaš J, Jurišić M, Gašparović M. 2020. Global Open Data Remote Sensing Satellite Missions for Land Monitoring and Conservation: A Review. *Land* 9(11):402.
9. **Radočaj D**, Jurišić M, Zebec V, Plaščak I. 2020. Delineation of soil texture suitability zones for soybean cultivation: a case study in continental Croatia. *Agronomy* 10(6):823.
10. **Radočaj D**, Jurišić M, Gašparović M, Plaščak I. 2020. Optimal soybean (*Glycine max* L.) land suitability using GIS-based multicriteria analysis and Sentinel-2 multitemporal images. *Remote Sens* 12(9):1463.
11. Gašparović M, Zrinjski M, Barković Đ, **Radočaj D**. 2020. An automatic method for weed mapping in oat fields based on UAV imagery. *Comput Electron Agric* 173:105385.
12. Jurišić M, Plaščak I, Antonić O, **Radočaj D**. 2020. Suitability calculation for red spicy pepper cultivation (*Capsicum annum* L.) using hybrid GIS-based multicriteria analysis. *Agronomy* 10(1):3.

Papers indexed in Web of Science – Emerging Sources Citation Index (ESCI) (10):

1. **Radočaj D**, Jurišić M, Plaščak I, Barač Ž, Zimmer D. 2021. Cropland suitability determination for maize (*Zea mays* L.) using multilevel GIS-based multicriteria analysis in continental Croatia. *J Cent Eur Agric* 22(3):636-645.
2. Jurišić M, **Radočaj D**, Plaščak I, Rapčan I. 2021. A comparison of precise fertilization prescription rates to a conventional approach based on the open source GIS software. *Poljoprivreda* 27(1):52-59.
3. Jurišić M, **Radočaj D**, Šiljeg A, Antonić O, Živić T. 2021. Current status and perspective of remote sensing application in crop management. *J Cent Eur Agric* 22(1):156-166.
4. Plaščak I, Jurišić M, **Radočaj D**, Vujić M. 2021. An Overview of Precision Irrigation Systems Used in Agriculture. *Teh Glas* 15(4):546-553.
5. Jurišić M, Plaščak I, Barač Ž, **Radočaj D**, Zimmer D. 2021. Sensors and Their Application in Precision Agriculture. *Teh Glas* 15(4):529-533.
6. Zimmer D, Plaščak I, Barač Ž, Jurišić M, **Radočaj D**. 2021. Application of Robots and Robotic Systems in Agriculture. *Teh Glas* 15(3):435-442.
7. **Radočaj D**, Jurišić M, Župan R, Antonić O. 2020. Spatial Prediction of Heavy Metal Soil Contents in Continental Croatia Comparing Machine Learning and Spatial Interpolation Methods. *Geod List* 74(4):357-372.
8. **Radočaj D**, Velić N, Jurišić M, Merdić E. 2020. The remediation of agricultural land contaminated by heavy metals. *Poljoprivreda* 26(2):30-42.
9. Šiljeg A, Jurišić M, **Radočaj D**, Videković M. 2020. Modeliranje pogodnosti poljoprivrednog zemljišta za uzgoj ječma uporabom višekriterijske GIS analize. *Poljoprivreda* 26(1):40-47.
10. Plaščak I, Jurišić M, **Radočaj D**, Barač Ž, Glavaš J. 2019. Hazel plantation planning using GIS and multicriteria decision analysis. *Poljoprivreda* 25(2):79-85.

Scientific conference papers with international peer-review (11):

1. Jurišić M, Glavaš J, Plaščak I, Antonić O, **Radočaj D**. 2021. New trends of implementation of geoinformation technologies in economics. *Interdisciplinary management research XVII*: 1004-1017.
2. Plaščak I, Jurišić M, **Radočaj D**, Zimmer D, Gjajić L. 2020. Inventory and management of vineyards in GIS environment. *13th international scientific/professional conference Agriculture in nature and environment protection*: 287-292.
3. Zimmer D, Jurišić M, Plaščak I, **Radočaj D**. 2020. Application of robots and robotic systems in agricultural practice. *13th international scientific/professional conference Agriculture in nature and environment protection*: 356-361.
4. Jurišić M, **Radočaj D**, Rapčan I, Zimmer D, Lončar A. 2020. Modelling of precision fertilization map using geostatistics. *13th international scientific/professional conference Agriculture in nature and environment protection*: 54-59.

5. **Radočaj D**, Jurišić M, Plaščak I, Barač Ž, Sigurnjak E. 2020. Monitoring of soybean growth stages using an unmanned aerial vehicle (UAV). *13th international scientific/professional conference Agriculture in nature and environment protection*: 60-65.
6. Barač Ž, Plaščak I, Jurić T, Jurišić M, Heffer G, Zimmer D, Vidaković I, **Radočaj D**, Kujundžija P. 2020. Produced noise level of the crusher in relation to individual exploitation factors. *13th international scientific/professional conference Agriculture in nature and environment protection*: 350-355.
7. Jurišić M, Plaščak I, **Radočaj D**. 2020. GIS-based multicriteria analysis methods in natural resources management: A review. *Interdisciplinary management research XVI*: 1001-1017.
8. Zimmer D, Jurišić M, Barač Ž, **Radočaj D**, Marić P. 2019. Technical Fertilizer Factors Using GIS Technology - BogBalle. *12th international scientific/professional conference Agriculture in nature and environment protection*: 300-304.
9. Jurišić M, Plaščak I, **Radočaj D**, Barač Ž, Zdravac A, Ramić M. 2019. Vizualization and methodology of management of urban vegetation in GIS environment. *12th international scientific/professional conference Agriculture in nature and environment protection*: 278-282.
10. Jurišić M, Plaščak I, **Radočaj D**, Barač Ž, Rapčan I. 2019. Precision fertilization of crops using nitrogen (N) sensors. *12th international scientific/professional conference Agriculture in nature and environment protection*: 289-294.
11. Barač Ž, Plaščak I, Jurić T, Jurišić M, Heffer G, Zimmer D, Vidaković I, **Radočaj D**, Majstorović S. 2019. Impact of vibrations on the hand-arm system during the exploitation of IMT 539 tractor. *12th international scientific/professional conference Agriculture in nature and environment protection*: 295-299.

Scientific conference papers with domestic peer-review (1):

1. Barković Đ, **Radočaj D**, Zrinjski M, Gašparović M. 2018. Analiza mogućnosti primjene bespilotnih letjelica u preciznoj poljoprivredi. *11. simpozij ovlaštenih inženjera geodezije*: 110-115.