POLITECNICO DI TORINO

Master's Programme in Ingegneria per l'Ambiente e il Territorio

Tutela Ambientale

Global analysis of relative maize productivity under three different regenerative practices

Emanuela De Giorgi

Prof. Francesco Laio, **Politecnico di Torino**
Prof. Matti Kummu, **Aalto University**

Master's thesis
2024

Author Emanuela De Giorgi

Title of thesis Global analysis of relative maize productivity under three different regenerative managements.

Programme Master's Programme in Ingegneria per l'ambiente e il territorio

Major Tutela ambientale

Thesis supervisor Prof. Francesco Laio & Prof. Matti Kummu

Thesis advisor(s) PH.D Ozias Hounkpatin

Date 15.08.2024 **Number of pages** 90 **Language** English

Abstract

Constant population growth leads to an increase in the demand for food. Until now, traditional agricultural techniques have been able to meet human demands, but at the expense of the environment. In recent years, new regenerative techniques such as no-tillage (NT), agroforestry (AF) and organic farming (OF) have emerged with the aim of safeguarding the environment. However, there is still no consensus on the outcome of their implementation as the interplay of many factors lead to different results. The aim of this analysis is to define under which environmental conditions maize yield may increase or decrease when regenerative management (NT, AF, OF) is implemented and the possible relationship between environmental factors and relative productivity. For that purpose, a global dataset of observations taken from different literature studies was used and several environmental factors were analysed to see if it was possible to define intervals in which homogeneous changes in productivity would occur. Subsequently, predictive models of effect size were obtained for each management using Random Forest model which also provides ranking of the most important variables affecting prediction. From the results, NT and AF appeared to have a higher potential for increasing maize productivity when the cultivated sites were in less favourable environmental conditions for the maize growth (Hyper-Arid and Arid climatic zone, Unsuitable and Hight Heat Stress GDD, Low SOC content ...) while OF recorded an average decrease in all the zones. The factor that mostly influence the effect size is different for each management with aridity index ranking first for NT while pH affected primarily AF and OF.

Considering the accuracy of the Random Forest model based on R^2 values, most reliable estimate turned out to be NT (R^2 = 0.36), followed by OF (R^2 = 0.13), and finally AF (R^2 = 0.07) which did not seem to return reliable values. Once AF is discarded considering the low accuracy of its model, the geographic area where NT seemed to lead to the highest increase in productivity were located in North-West America, North Africa, and India, while those where OF had the highest potential were South America and South Africa. Both practices, however, seems to have positive effect size in the North-East America. Moreover, the analysis shows that the relationship with the environmental component and the relative productivity is complex, and that it is not possible to define ranges of the variable in which the maize production increases or decreases with high accuracy.

Keywords Productivity, Effect Size, Regenerative Practices, No-Tillage, Agroforestry, Organic Farming, Maize, Random Forest.

Author Emanuela De Giorgi

Title of thesis Global analysis of relative maize productivity under three different regenerative managements.

Programme Master’s Programme in Ingegeria per l’ambiente e il territorio

Major Tutela ambientale

Thesis supervisor Prof. Francesco Laio & Prof. Matti Kummu

Thesis advisor(s) PH.D Ozias Hounkpatin

Date 15.08.2024 **Number of pages** 90 **Language** English

Abstract

La costante crescita della popolazione negli anni futuri porterà ad un aumento della domanda di cibo pari circa al 60% rispetto a quella attuale. Finora, le tecniche agricole tradizionali sono state in grado di soddisfare le richieste produttive dell’uomo, ma a spese dell'ambiente. Negli ultimi anni sono emerse nuove tecniche rigenerative come il no-tillage (NT), l'agroforestry (AF) e l'agricoltura biologica (OF) che non hanno più come unico obiettivo la produttività del suolo, ma anche la salvaguardia dell'ambiente. Tuttavia, non c'è ancora consenso sull'esito della loro applicazione per quanto riguarda la produttività delle culture. L'obiettivo di questa analisi è definire per ogni tecnica rigenerativa (NT, AF, OF) le condizioni ambientali in cui la resa del mais aumenta o diminuisce, e la possibile relazione tra i fattori ambientali e la produttività relativa. A tal fine, è stato utilizzato un dataset globale di osservazioni tratte da diversi studi di letteratura e sono stati analizzati diversi fattori ambientali per verificare se fosse possibile definire intervalli in cui si verifichino cambiamenti omogenei nella produttività. Successivamente, per ogni tecnica agricola, sono stati ottenuti modelli predittivi sulla produttività relativa (effect size) utilizzando il modello Random Forest. Questo modello è anche in grado di fornire una classifica delle variabili più importanti che influenzano la previsione. Dai risultati, NT e AF sembrano avere un potenziale maggiore per l'aumento della produttività del mais quando i siti coltivati si trovano in condizioni ambientali meno favorevoli per la sua crescita (zone climatiche ad alto stress idrico o di calore ...), mentre OF ha registrato una diminuzione media in tutte le zone. Il fattore che influenza maggiormente l’effect size è diverso per ogni pratica, l’aridity index è al primo posto per il NT, mentre il pH è la variabile principale per AF e OF. Considerando l'accuratezza del modello Random Forest in base ai valori di R^2, la stima più affidabile è risultata essere NT (R^2 = 0,36), seguita da OF (R^2 = 0,13) ed infine da AF (R^2 = 0,07) che non sembra restituire valori affidabili. Una volta scartata l’AF, il NT sembra registrare il maggiore aumento della produttività del mais in Nord America, in Nord Africa e in India, mentre quelle in cui OF ha il maggiore potenziale sono il Sud America e il Sudafrica. Entrambe le pratiche, tuttavia, sembrano avere un effetto positivo nell'America nord-orientale. Inoltre, l'analisi mostra che la relazione con la componente ambientale e la produttività relativa è complessa e che non è possibile definire con precisione degli intervalli in cui la variazione della produttività del mais sia omogenea.

Keywords Produttività, Effect Size, Pratiche rigenerative, No-Tillage, Agroforestry, Organic Farming, Maize, Random Forest.

Table of contents

Preface

This thesis was part of a project named SOS.aquaterra funded by European Research Council (ERC) under the European Union's Horizon 2020 and Aalto University whom I wish to thank for enabling my work within this project.

My sincerest thanks go to my thesis advisor Matti Kummu. Thank you for listening to my personal interests and for proposing topics that reflected my passions; in this way, you encouraged me to see my thesis not as an obstacle, but as a chance to deepen my interests. Thank you for having always been available to answer my questions and for taking the time to discuss my progress.

I also thank infinitely my advisor Ozias Hounkpatin who has always been close to me in both the planning and development of this work. I thank him for being supportive not only on a technical level but also on a moral level. His comments have allowed me to improve and develop a more discerning critical eye. I thank him for the many meetings and for his availability throughout the drafting of my project and for all the advices that helped me to enlarge my knowledge and my technical skills.

I would also like to thank the Water and Development Research Group and the professors who followed me in my final year at university for always being helpful and for making my little journey through Aalto University peaceful and knowledge-rich.

Espoo 10.06.2024

Emanuela De Giorgi

Aalto University
School of Engineering

Ringraziamenti

Ringrazio il Prof. Francesco Laio per avermi seguito nella fase finale del mio percorso di laurea magistrale, per avermi incoraggiata a credere nelle mie capacità e a mettermi in gioco nella mia futura carriera lavorativa.

Ringrazio la mia famiglia per il sostegno che mi ha sempre dimostrato anche in questi ultimi due anni di magistrale. In particolar modo ringrazio i miei genitori per avermi supportata in tutti questi anni, per avermi spronata a dare il massimo e per avermi spinta a costruire e rincorrere i miei sogni. Ringrazio i mei fratelli, Matteo, Federico ed Elisabetta che sono riusciti ad alleggerire le situazioni più stressanti che si sono presentate durante il mio percorso universitario e per aver sopportato i miei sbalzi d'umore durante le numerose sessioni d'esami. Un sincero ringraziamento anche alle mie nonne Adele e Colette che sono sempre state disposte ad ascoltarmi e pronte a consigliarmi.

Un sincero ringraziamento anche ai mei colleghi di università Alessia, Gabriele e Flavia con cui ho trascorso questi anni di studi e con cui ho avuto la possibilità di confrontarmi e di crescere. Grazie per essermi stati vicini e per il supporto che mi avete sempre dimostrato anche nei momenti più complicati e di sconforto. Grazie per esserci stati e per la serenità con cui mi avete permesso di affrontare anche le sfide più complesse.

Vorrei anche ringraziare Martina, Andrea, Edoardo, Marco ed Elena. È grazie a voi se al primo anno di università ho deciso di proseguire questo percorso. Vi ringrazio per le risate durante le esercitazioni di Analisi 1, la disperazione condivisa per i codici in c++ e per gli sguardi di confusione nelle ore della Grassini, non sarebbe stato lo stesso senza di voi. Abbiamo percorso solo parte del nostro viaggio insieme, ma vi sono grata per esserci stati.

Ringrazio Fabrizia e Thomas, siete stati due punti fondamentali per la mia crescita personale. Siamo cresciuti insieme tra confronti costruttivi e spensieratezza. Sia nel mio percorso di vita che in quello accademico siete stati un punto di riferimento e una spalla su cui potevo riposarmi quando sentivo di non essere in grado di gestire le situazioni. Grazie per i caffè improvvisati prima di una lezione o dopo un esame. Grazie per le piccole azioni quotidiane che ci hanno permesso di creare e di accrescere il nostro rapporto.

Ringrazio il gruppo scout Torino 6 e i ragazzi del reparto Atacama. È stato complicato proseguire il mio percorso universitario e quello scoutistico allo stesso tempo, ma questi due cammini mi hanno permesso di crescere sia professionalmente che personalmente. Sono convinta che non sarei la persona che sono oggi se avessi rinunciato a uno dei due. Grazie ragazzi per avermi ricordato quanto passi in fretta il tempo, la potenza di un sorriso nei

momenti più bui e la bellezza celata nelle esperienze inaspettate. Grazie per ogni passo che abbiamo condiviso.

Ringrazio anche Angelica, Alice ed Edoardo. Ognuno di voi mi ha ricordato la mia unicità e ha coltivato il mio lato irrazionale nonostante la razionalità richiesta dal mio percorso di studi. Siete stati coloro che mi hanno dimostrato per primi che il mio dolore e le mie difficoltà erano degne di essere ascoltate, mi avete insegnato che sono fallibile, ma che questa mia fragilità è un mio punto di forza perché mi permette di comprendere le difficoltà altrui. Grazie per aver custodito parte delle mie preoccupazioni e di avermi permesso di crescere per essere in grado di sostenerle da sola. Grazie per non avermi lasciata da sola e per avermi concesso di fermarmi quando ne avevo più bisogno.

Ringrazio infinitamente anche Umberto. Sei stato il primo a dirmi che la strada dell'ingegneria sarebbe stata la decisione giusta per me, io non ti ho creduto in quinta liceo e anzi, ho negato prontamente questa possibilità. Invece avevi ragione tu, e dopo cinque anni mi sono ritrovata a laurearmi in ingegneria ambientale, onestamente non ci credo ancora. Grazie per esserci stato quando mi stavo disperando per l'iscrizione all'università e per avermi sempre ricordato come ho superato le sfide che ho incontrato nel mio percorso e che mi hanno portata qui oggi.

Vorrei anche ringraziare Martina, Chenyue, Adam, Jana, Jan, Simon, Antonio, Thilo, Emanuele e Michele. Abbiamo trascorso solo un anno insieme, ma è stato fondamentale e unico per me, vi ringrazio perché mi avete supportata nei momenti di solitudine in Finlandia e avete reso l'esperienza Erasmus unica e serena. Sono cresciuta tanto in questi nove mesi passati insieme. Tra una scivolata sul ghiaccio invernale, una serata karaoke, le giornate di studio a business e preoccupazioni per il futuro ho iniziato a sentirmi a casa insieme a voi. Grazie davvero per tutte le piccole e grandi esperienze che abbiamo vissuto insieme.

I would like to thank Martina, Chenyue, Adam, Jana, Jan, Simon, Antonio, Thilo, Emanuele, and Michele. We barley spent a year together, but it was very important and unique for me. I am grateful to you for being there when I felt lonely in Finland, you made my Erasmu experience memorable and peaceful. Between the slippery winter ice, karaoke nights, afternoon spent in business and the worries about the future, I began to feel at home with you. Sincere thanks for the great and small experiences we have had together.

Infine ci tengo a ringraziare tutti i professori che mi hanno seguita sia in Italia che in Finlandia. Un particolare ringraziamento al Prof. Roberto Revelli. Ho avuto la possibilità di seguire con lui un unico corso, ma mi ha insegnato che la branca dell'ingegneria non è semplicemente tecnica, ma anche passione e ascolto. Ascolto per le esigenze dei cittadini, consapevolezza delle proprie responsabilità e fiducia nelle proprie capacità. È stata la prima figura educativa che mi ha permesso di comprendere l'importanza dei miei pensieri e delle mie idee. La ringrazio professore per avermi fatta sentire un'eccellenza e non una scatola da riempire.

Symbol and Abbreviations

Abbreviations

AF	Agroforestry
AI	Aridity index
BD	Bulk Density
CI	Confidence Interval
DfI	Diffuse Insolation
ES	Effect Size
ffs	Forward Feature Selection
GDD	Growing Degree Days
KDD	Killing Degree Days
NT	No-Tillage
OF	Organic Farming
P	Phosphorus
PE	Potential Evapotranspiration
QRF	Quantile Regression Forest
RF	Random Forest
RMSE	Root Mean Square Error
SOC	Soil Organic Carbon
RP	regenerative Practice
SV	Shapley Value

1 Introduction

1.1 Background and main objective of the project

Agricultural productivity will have to be increased by 60% to be able to meet the planet's demand for food as the global population is projected to go beyond 9 billion by 2050 *(Department of Economic and Social Affairs, 2017)*. Meanwhile, half of the habitable land is already used for this purpose *(Ritchie and Roser, 2019)* with humanity overexploiting the Earth's capacities with negative impacts on the environment. In fact, many consequences have been associated with the use of non-sustainable farming practices.

Conventional agriculture focuses on food productivity by relying heavily on external inputs (synthetic fertilisers and other agricultural chemicals). It has high production costs and leads to overproduction of land. These techniques also manipulate the physical properties of the soil and aim to control weeds and achieve monoculture fields. They have led to the reduction of biodiversity and the damage of ecosystems leading to crop resilience *(Choden and Ghaley, 2021)*. Additionally, overexploitation of the soil has led to soil degradation resulting in the reduction of the concentration of organic carbon as well as the decline of its productive capacity. It is reporting that the soil has been unable to store about 80 billion ton of CO_2, increasing thereby its concentration in the atmosphere *(Lal, 2004)*. Also, with reduction in soil production capacity, anthropogenic inputs are increased to maximise its productivity. The excessive use of fertilizers, pesticides, insecticides, and herbicides has led to the contamination of soil as well as water. For instance, 38% of the world's water bodies *(Choden and Ghaley, 2021)* are contaminated while about 60-75% of the European cultivated soils have unnecessary nutrient inputs *(European Environment Agency, 2023)*.

Conventional farming techniques have made it possible to satisfy the world population's demand for food, but in the light of the preceding observations, the damage caused to the environment is not negligible and compromises its productive capacity. In that context, these production techniques are more and more questioned leading to the consideration of more sustainable alternatives such as regenerative agriculture which target both current and future productivity, with the aim of safeguarding the environment, and particularly the soil.

Key among the regenerative agriculture techniques are no-tillage (NT), organic farming (OF) and agroforestry (AF) which are being promoted as sustainable farming practices. However, contrasting findings are reported across studies that compare the productivity of regenerative and traditional farming techniques *(Félix et al., 2018; Choden and Ghaley, 2021; Achankeng and Cornelis, 2023)*.

The ***NT*** focuses on minimizing the soil disturbance; this management does not allow the drilling of the soil; therefore, the seeds are planted into an unprepared terrain. Usually, the land is covered at least for the 30% with mulch, a protective layer deposed on the topsoil *(Triplett and Dick, 2008)*. Compared with the traditional tillage, NT was associated positive effect on soil conservation, water consumption, nutrient regulation, and crop diversity within Europe and North Africa *(Choden and Ghaley, 2021)*. However, a global NT analysis showed no consistent variation in crop productivity with outcome mainly depending on management practices *(Su, Gabrielle and Makowski, 2021a)*. A study considering the relation between crop relative productivity, soil texture, rotation, and climate in Europe found an average reduction of 8.5% in maize yield under NT while under particular conditions, crop yield increased also by 4% (*Achankeng and Cornelis, 2023*). Considering the situation in Asia, this management seems to do not have significant impact on crop productivity with studies reporting a general reduction around 2% in China while it leads to an increase in productivity varying from 0% to 10.2% in South Asia according to the crop class *(Zhao et al., 2017; Anantha et al., 2021; Hashimi, Kaneko and Komatsuzaki, 2023)*. In America this practice is already implemented in many croplands with a positive improvement for the environmental factors and the crop yield is related to the type of crop without any consistent changes in productivity *(Nunes et al., 2018; Awe, Reichert and Fontanela, 2020; Jacobs et al., 2022)*.

The ***OF*** is a management which avoids the introduction in the land of synthetic fertilizers or pesticides. In the OF land cover material is composed by organic scraps (leaves, grass, crop residues ...), but it can also be made by an inorganic material (stones, plastic ...) *(FAO, 2024)*. Compared to the conventional agriculture, it increases the carbon sequestration lowering the emission of greenhouses gas, increase the soil organic carbon (SOC) and the soil biodiversity, it also reduces the energy consumption and nutrient losses. A global meta-analysis showed that OF is generally 18.4% less productive than conventional farming *(de la Cruz et al., 2023)*. However, the available data did not allow an accurate analysis because they did not have a good distribution as about 86% of the observations came from studies conducted either in North America or Europe *(de la Cruz et al., 2023)*. This study also considers just three environmental factors for the analysis (pH, soil texture and climate). Other studies have shown that the yield with the implementation of OF is typically 20-50% lower than the conventional farming system, but the results are closely related to site characteristics *(Choden and Ghaley, 2021)*. However, OF seems to be more suitable for the arid zone because it has a strong resilience against the water stress with an increase of 16% in productivity especially for African countries *(Hine and Pretty, 2006; Niggli, 2015)*. The main challenge of this technique which results in lower yield compared to conventional farming is the control of weeds and pets, and the nutrients availability as nitrogen and phosphorus *(Hine and Pretty, 2006; Niggli, 2015; Choden and Ghaley, 2021)*.

AF is a practice that include woody autochthonal species into the productive system of agricultural crops. Their introduction augments the biodiversity of the site, and it is also favourable to modify the microclimate regulating the temperature thanks to the presence of the tree shading. A global meta-analysis shows that this management has an average increase of 7% in the maize yield, the best situation is recorded in subtropical and tropical zones, where the growth of productivity is equal to 16% *(Baier et al., 2023)*. The result is affected by the species of woody plant implemented and the region of the study. It is important to underline that, even if it is a global analysis, most of the available observation are located in Africa. Indeed, 61% of the studies were carried out in Africa, while only 3% are in Europe and Middle East *(Baier et al., 2023)*. An European study demonstrate that the crop productivity with the implementation of this management can have different results which vary according to the density of the trees and their age *(Ivezić, Yu and Werf, 2021)*. This study also reported a reduction around 2.6% every year due to the age of the plants.
In summary, it is possible to say that previous studies have showed that the regenerative practices (RP) have led to either increase or decrease in productivity compared to conventional techniques. However, the results are still not globally clear, since most of the studies are site specific, and the global analyses do not consider many environmental factors together. The central point is that there is no universal rule regarding the outcome of the RP as many components must be considered, including for example climate, the type of crop used, soil properties, the terrain etc. In addition, though these studies have analysed the variation in productivity between regenerative and conventional farming techniques, they only focus on one type without investigating their comparative potential across various factors

and which management could be potentially more beneficial for a specific location. Consequently, the main goal of this analysis is to locate potential areas suitable for regenerative management practices while also assessing the underlying environmental factors. Defining the geographical position where the implementation of one or more RP would lead to an increase in productivity will help inform management and provide insight to decision making towards smart farming.

Therefore, this study aims specifically at: (1) analysing major moderators explaining yield discrepancy among different RP (NT, OF, or AF) at a global scale, (2) mapping the spatial distribution of yield under different RP along with the main influencing factors, (3) and finally determining which RP could be the best option to adopt for specific locations, and where their implementation can be an advantage for the agricultural activity. The study is based on a predictive machine learning model which uses several biophysical factors and global experimental observations to quantify the relative productivity at unknown locations along with the related variable importance. In that regards observation collected through different studies were used to achieve these objectives.

1.2 Research questions

The research questions that are addressed include the following:

1. How does relative crop yield vary across several components of climatic and environmental factors under different RP?
2. How does the spatial distribution in relative crop yield and related local factors vary across different RP at global scale?
3. Which regenerative management technique (NT, OF, or AF) could be potentially more beneficial for specific location?

1.3 Scope of the study

Due to the limited time available, it was not possible to consider many crop types this study. Therefore, the study will focus only on maize productivity.

Another limitation is due to the distribution of the observation points in the dataset. The *Figure 1* shows their distribution according with the cropland density. Most of the observation points belong to areas where agricultural activity is heavily practised. However, not all the high-density croplands are considered. For instance, there are no observations belonging to the India region, as well as Russia. While most of the observations are in the United State. That could be a limitation for the study, because it will not consider geographical areas that could play a key role in determining maize productivity trends. Moreover, the observations are not well divided into the three RPs, *Figure 2* shows their distribution into the cropland.

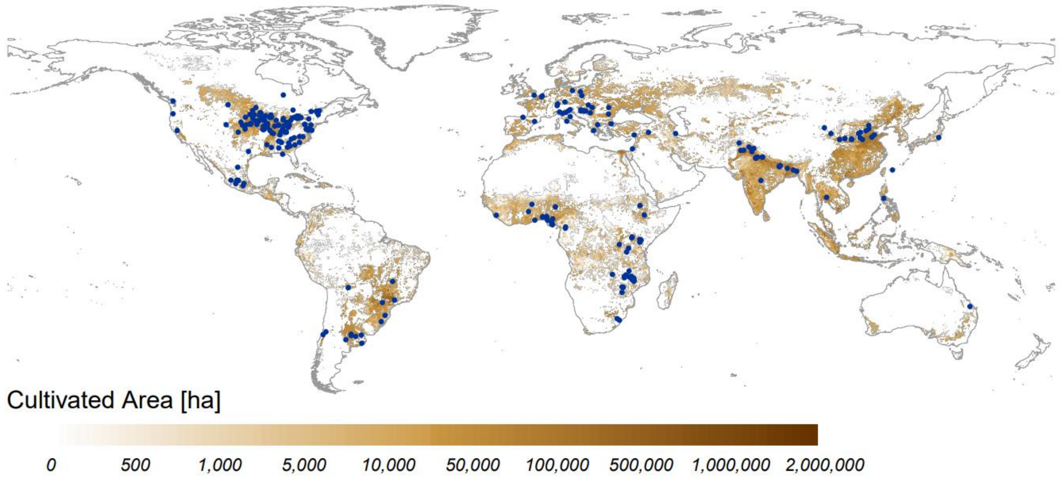


Figure 1, Observation points distribution across the cropland 2020 *(International Food Policy Research Institute, 2020)*.

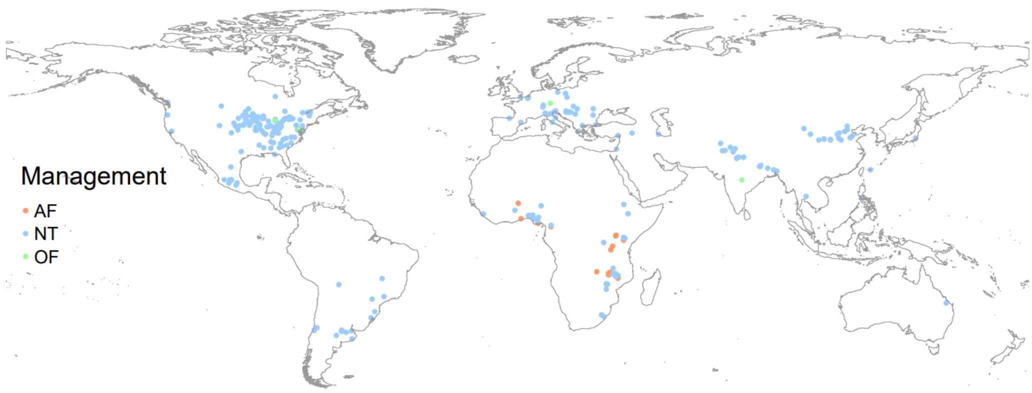


Figure 2, Observation distribution of the dataset divided per RPs.

1.4 Structure of the study

The study is divided into two main parts. The first one is related to the inspection of the dataset. Its aim is to analyse the data and understand if there is a linear relationship between the productivity and some variables according to the management techniques. The second one involves the implementation of a machine learning model to predict the maize productivity all over the world based on the observation points in the dataset and a set of environmental factors.

2 Research material and methods

2.1 Data source

The dataset used for the analysis combined various data with observations coming from different studies *(Pittelkow et al., 2015; Félix et al., 2018; Jian, Du and Stewart, 2020; Su, Gabrielle and Makowski, 2021b)*. For each point the coordinates and several environmental variables are available. Moreover, all the observations report the maize productivity of the land under traditional agricultural practice and under one of the three regenerative management that will be analysed. The sources used to create the dataset have been reported in the *Appendix 1*. The dataset contains 260 different coordinates unequally distributed across the globe. However, it contains more observations, this is due to the fact that, for some locations are available data from different years. The study considered all of them. Coordinates with a larger number of observations have greater reliability because the values of the variables do not represent an outlier caused by an extraordinary event that may have occurred during the sampling period, but a representative value for the analysed region. About 49% of the location has less than five observations and just 31 geographical positions have two observations (*Figure 3*).

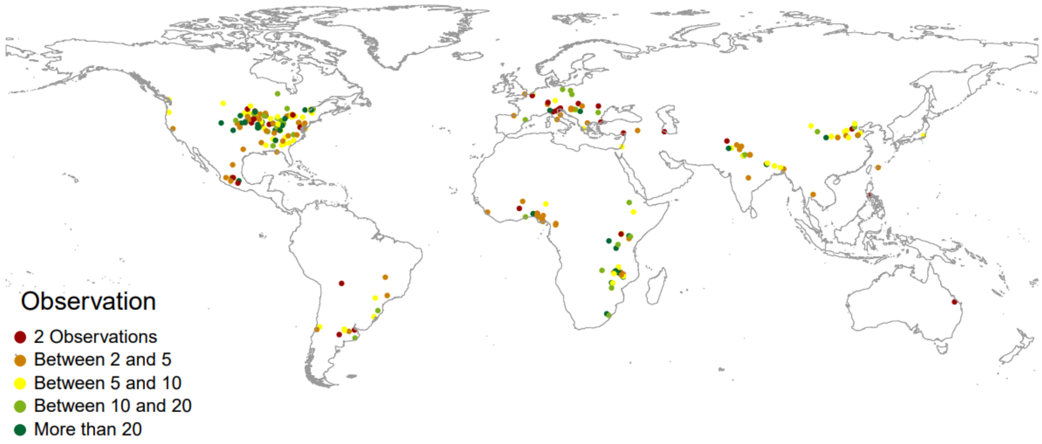


Figure 3, Number of observations per coordinates. The observation contains the productivity of the land (traditional and one RP), and the environmental variables.

2.2 Effect Size

The effect size (ES) is the main focus used in this study. It has been computed from the available data in the dataset (maize productivity under traditional techniques and under RP). It allows to quantify the changes of the subject comparing control and treatment groups (Xu *et al.*, 2021). In this study, the subject is the land productivity with crop yield considered as indicator, while the groups are plots submitted to different agricultural managements. Usually, management experiments are carried out with some plots with the traditional way of management (control) while other plots are submitted to a conservative management (the one whose effect is to be detected, the treatment plot). The ES can be expressed according to the following equation:

$$ES = ln\left(\frac{y_t}{y_c}\right). \qquad 1$$

The variables are:

- y_t = treatment yield. It is the productivity recorded in the treatment plot.
- y_c = control yield. Productivity of the control plot.

If the ES is negative, the productivity of the treatment is lower than the control, which means that the RP has a lower productivity compared with the traditional. On the other hands, if it is positive, the traditional technique has a lower productivity than the detected one. If it is equal to zero, there is no difference between the two yields. Therefore, the threshold between a positive result for the implementation of the RP and a negative output is *ES equal to 0*.

2.3 Potential factors affecting crop productivity

The potential factors affecting crop yield considered in this study (*Tables 1*) are mainly environmental especially its abiotic components including climate, soil properties and landform features (Liliane *et al.*, 2020). Consequently, the biotic components such as pests, insects and diseases that also affect crop production are not considered.

The process of defining *Table 1* is described in *Figure 4*. First, the potential environmental factors which could influence maize productivity were selected based on literature. These factors were classified according to different category and sub-category. However, if the factor in *Table 1* is accessible and the published data are reliable and of good quality, it is maintained otherwise it is not taken into consideration for the analysis. This step has been repeated for all the variables and the final factors that are retained are reported in bold.

Table 1: Factors affecting the crop productivity.

	Sub-Category	***Parameter***	***Factors***
Soil	Physical property	Soil colour	Minerals
		Soil structure	***BD*** Porosity Biological soil properties ***SOC*** ***Soil texture***
		Soil water availability	Water holding capacity Hydraulic conductivity Saga Wetness Index Water Table Depth
	Chemical property	Soil salinity	Electric conductivity
		Soil acidity	***pH***
		Nutrients	N, **P**, K, Mg, Ca, and S
			Fe, B, Cl, Zn, Cu, Mn, and Mo
	Biological property	Flora	Plants
		Fauna	Animal species
		Micro-organisms	Bacteria Protozoa Fungi
Climate	Climatic component	Sun exposition	***Insolation*** (direct, diffuse) Light (quality, quantity, duration)
		Temperature	Mean temperature ***GDD*** PE
		Rainfall	Precipitation
		Humidity	Relative humidity ***AI***
		Wind	Wind velocity
		Atmospheric gasses	CO_2 and O_2 concentration
	Climatic stress	Water	Drought (SPEI) Flood
		Temperature	Heat stress Cold stress
Landform	Terrain	Topography	***Slope*** ***Elevation*** Physical form
		Erosion	Distance to water Wind

Table 2: Zone ranges and studies.

Category	*Zones*	*Range*	*Study*
Bulk Density	Favourable BD	BD < 1.2 kg/dm^3	(Minnesota Pollution Control Agency, 2024) & (Chen and Weil, 2011)
	Transition BD	1.2 kg/dm^3 < BD < 1.47 kg/dm^3	
	Restrictive BD	BD > 1.47 kg/dm^3	
Soil Texture	Sandy	From class 1 to 3	(Allakonon *et al.*, 2022)
	Silty	From class 4 to 10	
	Clay	From class 11 to 12	
SOC	Low SOC	SOC < 5 g/kg	(Oldfield, Bradford and Wood, 2019)
	Moderate SOC	5 g/kg < SOC < 10 g/kg	
	High SOC	SOC > 10 g/kg	
pH	Acidic	pH < 6.3	(Sirisuntornlak *et al.*, 2021)
	Neutral	6.3 < pH < 7.4	
	Alkaline	pH > 7.4	
Olsen-P	Low P	P < 10.9 mg/kg	(Bai *et al.*, 2013) & (Fulton, Advisor and Counties, 2010)
	Moderate P	10.9 mg/kg < P < 21.4 mg/kg	
	High P	P > 21.4 mg/kg	
GDD	Unsuitable	GDD < 800°C/y	(Croitoru *et al.*, 2020) & (Dong *et al.*, 2021)
	Suitable	800°C/y < GDD < 2700°C/y	
	Low Heat Stress	2700°C/y < GDD < 4000°C/y	
	Moderate Heat Stress	4000°C/y < GDD < 6000°C/y	
	High Heat Stress	GDD > 6000°C/y	
Climate	Hyper-Arid	AI < 0.05	(Allakonon *et al.*, 2022)
	Arid	0.05 < AI < 0.2	
	Semiarid	0.2 < AI < 0.5	
	Sub-Humid	0.5 < AI < 0.65	
	Humid	AI > 0.65	
Insolation	Low DfI	DfI < 0.45 kWh/m^2	(Yang *et al.*, 2019) & (Campillo, Fortes and Del Henar Prieto, 2012)
	Moderate DfI	0.45 kWh/m^2 < DfI < 0.55 kWh/m^2	
	High DfI	DfI > 0.55 kWh/m^2	
Slope	Flat	Slope < 0.2%	(FAO, 2006) & (Neumann *et al.*, 2010)
	Level	0.2% < Slope < 1%	
	Gently	1% < Slope < 5%	
	Sloping	5% < Slope < 15%	
	Steep	Slope > 15%	
Elevation	Low Elevation	Elevation < 250 m	(Arshad, 2021)
	Moderate Elevation	250 m < Elevation < 1000 m	
	High Elevation	Elevation > 1000 m	

The variables retained were further described based on specific ranges of their values which in turn were labelled as zones (*Table 2*). Primarily, range definition was based on literature. However, when it was not possible to find any literature-based ranges, the zones were individuated through its distribution into the dataset. For that purpose, the distribution of the factor based on histograms helped define the zones as well as the spatial extension of each class. A distribution is accepted if the data is divided into the ranges equally as far as possible, but at the same time the geographical location and class size on the globe must be taken into account. Based on these two aspects, it was possible to determine whether the division could be appropriate or not. For cases where the distribution was not consistent, the ranges were defined again until the class extent and related observations were satisfactory.

Although many potential factors are presented in *Table 1*, only the selected (in bold) ones are further considered in the next sections.

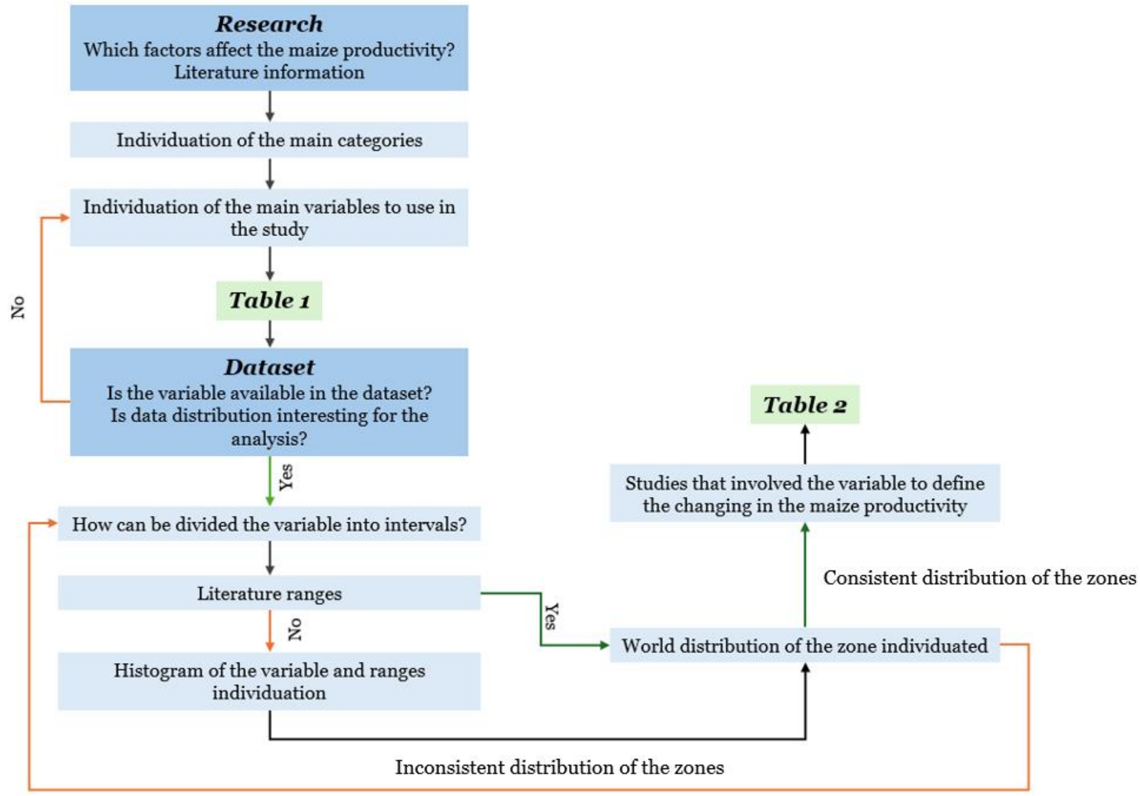


Figure 4, Diagram for the decision of the variables involved in the study and the division into zones.

2.3.1 Soil physical and chemical properties

The soil properties (bulk density, SOC, sand, silt, clay, pH) considered (*Table 1)* in this study were downloaded from *SoilGrids* at 250 m resolution *(Poggio et al., 2021)*. *SoilGrids* represents a global soil information platform which was generated using about 110,000 world soil profiles from all over the world with data available at different soil depths. Data were first downloaded for the following depths 0 - 5 cm, 5 - 15 cm, 15 - 30 cm and a depth averaging was carried out to get the soil properties for the 0 – 30 cm depth.

The bulk density (BD), texture (sand, silt, clay) and SOC define the ***soil structure*** which affects the fluid movement in the ground, the stability against erosion, the extent of carbon sequestration as well as soil fertility *(MSU Extension Service, 2005).*

✓ *Bulk Density (BD):*

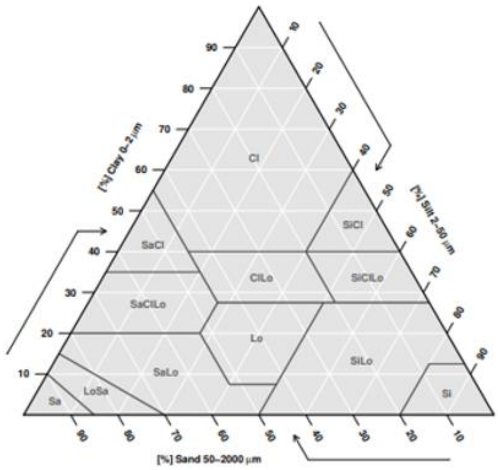
The ***BD*** is the ratio between the mass of the soil dry fraction and its volume (kg/dm3). It is an important indicator of the level of soil porosity and compaction. The soil compaction influences the root growth and therefore, the productivity of the crop *(Almendro-Candel et al., 2018).*

The zone ranges were defined for the BD (*Table 2*) based on the average of the threshold of each soil types *(Minnesota Pollution Control Agency, 2023)* and on the ranges used in a previous study *(Chen and Weil, 2011)*. When the BD has low values, for the plant is easier to reach the nutrient and the water, while with high values the terrain is compact and the roots growth is limited by the high mechanical impedance *(Vepraskas, 1988).*

✓ *Soil Texture:*

The ***soil texture*** represents the percentage of clay, silt and sand contained in a unit soil sample. It affects the physical properties of the soil, particularly the available water capacity *(Amsili, van Es and Schindelbeck, 2021).*

The USDA soil texture classification (*Figure 5*) was used to define three zones (*Table 2*). The clay represents the soil with a prevalence of fine particles (texture class 1,2 and 3 class). The silty, is the class which collects the soil with a medium size, and which has almost an equal division between the fine frictions of the hearth; therefore, there is not a particles size which strongly prevail on the others (from texture class 4 to 10). And finally, the sandy zone is the one which collect the soil with big particles size (texture class 11, and 12).



Texture class	Abb.	Name
1	Cl	clay
2	SiCl	silty clay
3	SaCl	sandy clay
4	ClLo	clay loam
5	SiClLo	silty clay loam
6	SaClLo	sandy clay loam
7	Lo	loam
8	SiLo	silty loam
9	SaLo	sandy loam
10	Si	silt
11	LoSa	loamy sand
12	Sa	sand

Figure 5, Soil texture triangle *(Moeys, 2018). USDA classification of the soil texture.*

✓ *Soil Organic Carbon (SOC)*

The ***SOC*** is the carbon component of the soil organic matter (SOM) which allows the storage of the nutrients, the water retention and gives habitat and energy to microorganisms. It is also the result of the biological activity in the soil and can indirectly describe the biological properties of soil *(FAO, 2017)*. The division of the zones (*Table 2*) was based on a global meta-analysis which demonstrated that, the highest productivity was recorded in a range of SOC between 5 g/kg and 20 g/kg with the yield being 1.2 times higher at 10 g/kg compared to 5 g/kg SOC *(Oldfield, Bradford and Wood, 2019)*.

✓ *pH*

pH is the measure of the hydrogen ions in the soil and defines the acidity and the alkalinity of soil water. It affects soil biochemical processes, and nutrient availability depending on the acidity level. For instance, when the pH has low values (acidic soil) the growth of the plant is limited to the presence of some toxic substances in the soil (aluminium and manganese) and to the absence of others (calcium, magnesium and phosphorous). High level of pH corresponds to deficiencies of zinc, copper, manganese, and boron and to high level of sodium. Usually a neutral pH is the optimal condition for the growth of the crops *(The State of Queesland, 2024)*.

✓ *Olsen-Phosphorus*

The ***phosphorus*** (P) is a fundamental macronutrient, it regulates cellular processes of the plants, their water content, and reduces the adverse effects of salts; therefore, it directly affects the productivity of the plants *(Kumar, Kumar and Mohapatra, 2021)*. The P data (1 km resolution) was acquired from the platform provided by the global study of *McDowell et al. (2023)*. This study used the bicarbonate-extractable Olsen P as the measure of plant available soil P with data (n = 574,375) from regional or global databases and published studies.

2.3.2 Climatic components

The Growing Degree Days (GDD), aridity and insolation were used as the climate variables considered in this study.

✓ *Growing Degree Days (GDD)*

GDD describes the amount of heat a crop needs to develop from one life-cycle phase to another (Ahmad *et al.*, 2017). It required daily maximum temperature, daily minimum temperature, and base temperature (T_b) for its computation. Considering maize, the T_b is around 5-10°C (Fatima *et al.*, 2020). The GDD data (1 km) was obtained from *Ahvo et al. (2023)* analysis.

Typically, the maize reaches its complete maturity when the GDD is in a range between 2000 and 3000 according with the T_b used and the maize species *(Emmalea, 2020; Vâtcă et al., 2021)*. Maize also needs a minimum GDD to reach its maturity according with the FAO classification it must be higher than 800°C/y; as well as a maximum value after which it is no more suitable resulting in heat stress. When the GDD exceeds this threshold, it changed its name into Killing Degree Days (KDD) *(Croitoru et al., 2020; Dong et al., 2021)*. On the base of these considerations, and on the data distribution, the zones have been defined.

✓ *Aridity Index (AI)*

AI is the ratio between the precipitation and the potential evapotranspiration (PE) based on study by *Zomer et al. (2022)*. The same study provided the AI map (1 km resolution) which was used in this study.

Based on it, five different climate zones were defined following the European classification *(European Union, 2019)*. The first three zones (*Hyper-Arid*, *Arid*, and *Semiarid*) have similar characteristics, they are region where water scarcity and high climate variability occur, they are also characterized by degradation phenomena. On the opposite, when the AI has higher value, the region belongs to the *Humid* zone, which are places where the precipitation are well distributed over the year and are higher than the evapotranspiration (*Henry, 2005*). In this study, the *'Cold Zone' (European Union, 2019)* considered by the European climate classification was left out because there are no observations belonging to this climate zone.

✓ *Insolation*

Solar light intercepted by crops affects their photosynthetic activities resulting in growth and productivity. This study considered the ***diffuse insolation*** (DfI) which was derived from a 1 km digital elevation model *(Amatulli et al., 2018)* using the Saga software *(Conrad et al., 2015)*. The DfI represents the part of total solar radiation which is scattered by the atmosphere, and which reaches the ground *(Szatten and Więcław, 2021)*.
The division of this variable into zone was based on the dataset distribution.

2.3.3 Terrain

The terrain features in this study include the ***elevation*** and the ***slope***. Elevation generally affects the climatic condition of the site *(Baker and Capel, 2011)* along with the slope which controls erosion processes *(Ma et al., 2019)*. They were freely accessible (1 km resolution) on the platform provided by the global study of Amatulli *(Amatulli et al., 2018)*.
The zone division for the elevation was based on the data distribution while the FAO analysis *(FAO, 2006)* was used for the slope.

2.4 Modelling with Random Forest

This study used the Random Forest (RF) algorithm to build a quantitative relationship between the factors used as predictors and the ES which represents the target variable. It has already been implemented in many studies related to the ecological problem (Marques Ramos et al., 2020; Burdett and Wellen, 2022; Ahvo et al., 2023) and it has been demonstrated that, considering the regression problem, the RF is one of the most reliable and accurate method that can be used (Su et al., 2022). Basically, (1) it creates randomly a bootstrap sample from the training data, (2) randomly select a subset of predictors (mtry), (3) train a tree based on the bootstrap sample and the subset of predictor, (4) repeat the same activities (random bootstrapping and selection of subset of predictor variables) several times, and finally (5) compute the average of all the predictions from all trees involved in the regression.

The NT data was split into training and testing set. For the OF and AF, the whole dataset was used for training the model because the number of observations was too small to further divide into two sets. A feature selection was carried out based firstly on a correlation threshold of 0.70 and secondly on the forward feature selection (*ffs*) procedure during model training as carried out by *Meyer et al. (2019)*. For the first approach, some predictors with correlation higher 0.70 were removed. The *ffs* removes iteratively the variables that reduce the accuracy of the model and finally retains those that contribute most accuracy *(Meyer et al., 2019)*.

Using the RF, it is possible that overfitting will occur. It is a negative aspect because it reduces the ability of the model in the prediction of the unknown values *(Meyer et al., 2019)* , this is the reason why the spatial cross-validation as internal validation of the model has been used to increase the accuracy of the predictive model.

A tuning of the hyperparameters was carried out for the number of predictors to use at each split (*mtry*) and the minimum number of observations in a leaf (*n*). The number of trees was set at 500 for all the models.

Figure 6 allows to visualize the steps followed to get the final results of the analysis. The first phase is the preparation of the data and the computation of the correlation between every predictor. The model requires the choice of a spatial variable to be used for cross-validation. Different spatial variable can produce models with different accuracy. Therefore, the coordinates, the geographical regions (Africa, America etc...), climate class were tested as spatial variables. For that purpose, the model is trained and the statistical values of the training dataset are computed. These steps are repeated for all the spatial variable in order to obtain the most accurate prediction. The model with the lowest root mean square error (RMSE) and the highest R^2 were chosen. If both previous conditions are fulfilled the highest relevance has been given to R^2. The last two steps require the computation of the Shapley values (SVs, see section below) which state which factors are the most important in the prediction. Finally, the models were used to produce the maps (5 km resolution) of the predicted ES, of the prediction uncertainty (see section below) as well as that of the predictors with the highest SV for each location at a global scale.

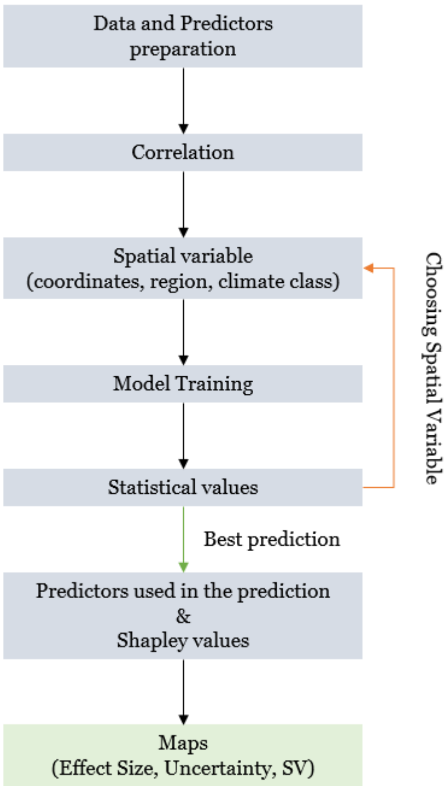


Figure 6, Machine learning diagram.

The assessment of the model accuracy was based on R^2 and RMSE. The R^2 and the RMSE are defined for both the training and the test data. The R statistical software language was used for all analysis in this study. For the different predictions, the ranger based RF was implemented using the R '*caret*' package *(Kuhn, Kuhn and Max, 2015)* using tenfold cross-validation with three repetitions.

2.4.1 Variable importance based on Shapley values (SVs)

The analysis of the variable importance was based on the SVs (i.e. Shapley values). It gives an average rank of the most relevant variables *(Corley, 2017)*. They were first computed for the training set of each model to get the partial dependency plots which shows the trend in the relationship between predictor and ES. At a second stage, the SV was computed at grid level for every raster cell in order to obtain a map which shows the most important predictor contributing to the spatial variability of the ES at a specific location.

2.4.2 Uncertainty map

To have a better idea of the accuracy of the model, uncertainty analysis was carried out that resulted an uncertainty map for each of the predictions. It allows to visualize in which area there is an accurate prediction and the area where the result does not have a good reliability. For that purpose, the Quantile Regression Forest method (QRF) was used *(Vaysse and Lagacherie, 2017)* with the hyperparameters obtained from the tuning of the model. It carries out the accuracy assessment by computing the upper and lower limit of the confidence interval (CI). It defines the interval within which 90% of the predicted values fall. The upper limit is defined by the 95^{th} quantile, while the lower one by the 5^{th} quantile *(Dharumarajan et al., 2024)*. The lower the length of the CI, the lower the uncertainty of the model.

3 Results

3.1 Descriptive analysis

The objective in this section is to define whether there is a relationship between ES and the factors analysed in the section *3.2*. It is divided into three parts. In the first, the distribution of observations in the different zones for each category is shown. In the second, the boxplots show the statistical values for each zone and RP. Finally, the graphical representation of the factors and the ES for each management makes it possible to detect whether there is a linear trend between the relative productivity and the variables.

3.1.1 Data distribution into the zones

Each variable in the dataset has been divided into ranges in order to understand if would have been possible to define some environmental boundaries in which the RP has a similar ES average value. The spatial extent of the defined zones for the different variables used in this study are specific for each factor with some zones having larger or lower extent (*Figure 7*, *Figure 8*, *Figure 9*). This is understandable since factors with more classes (> 3, e.g. climate zones, GDD, slopes) will present more spatial variability with lower extent with some classes compared to those with only three classes (e.g. BD, SOC etc.).

The distribution of the observations into each zone is presented in *Table 3*. The observations are 2956 in total with 82% of them belonging to the NT management, 17% to the AF, and 1% to the OF. Therefore, it is more probable to have observations for every class in the NT management compared to AF and OF. Indeed, looking at *Table 3* it is visible that only the *Hyper-Arid* aridity index zone does not have any data for the NT, while the number of zones without observations increase for the AF and even more for the OF.

For the NT, most of the variables tend to have two classes that have roughly the same number of points and then a third to which belongs less than 15% of the data (*Table 3*). It is the only management which has some observations for the *Alkaline* pH, the *Arid* climate, the *Unsuitable* and *Low Heat Stress* GDD, and *Steep* slope zones. For NT and AF, the factor with the worst distribution is the pH, where more than 90% of the data belonging to *Acidic pH* class. For these two groups, factors with less skewed distribution are DfI, AI, Olsen-P and elevation. Specifically, AF usually has one zone in which most of the observations belongs. The OF is the management where typically the less homogeneous division occurs, except for pH and SOC. The worst situation is recorded in the DfI where all the observation points belong to one class.

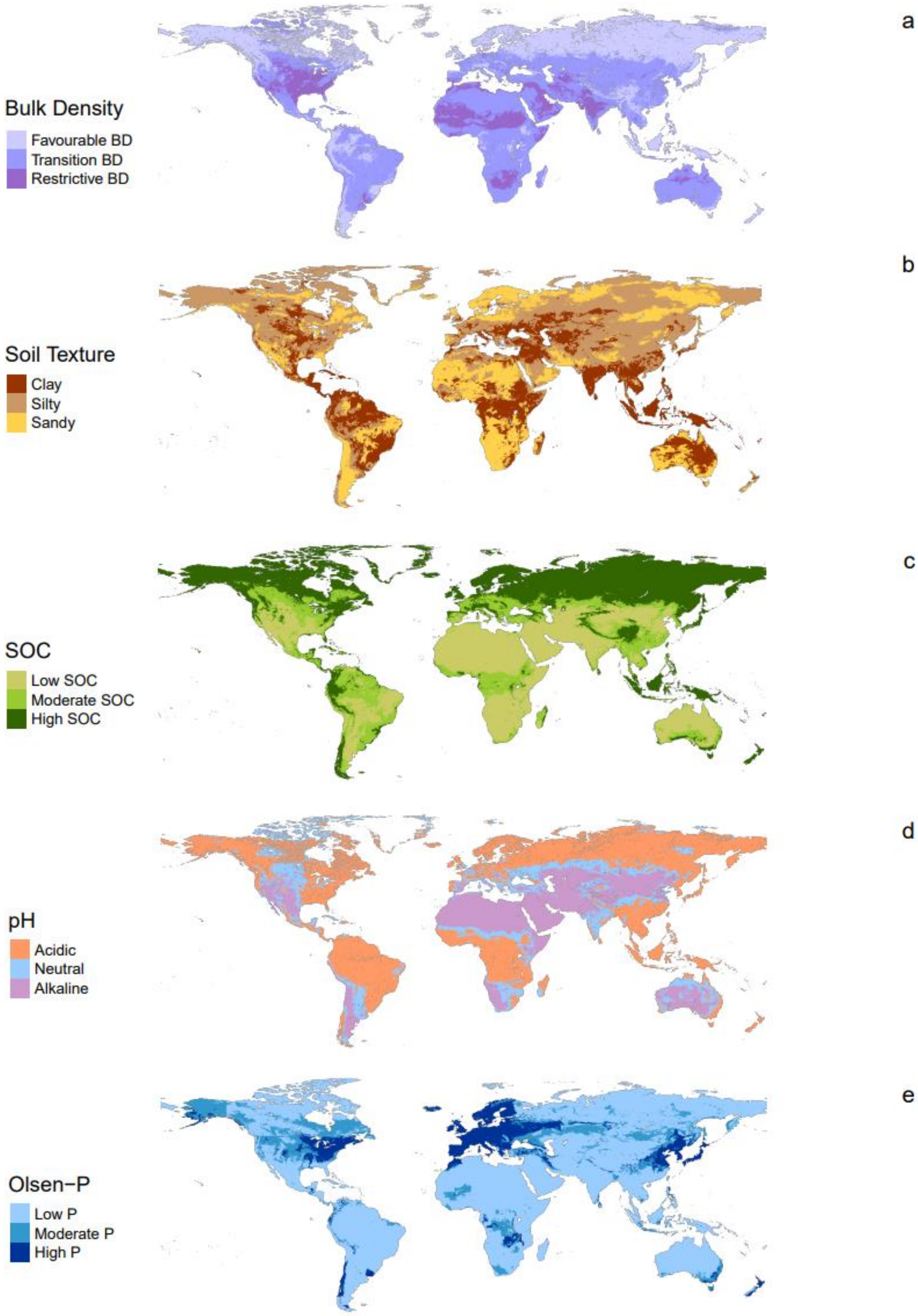


Figure 7, Global extension of the zones of the different factors belonging to the soil category.

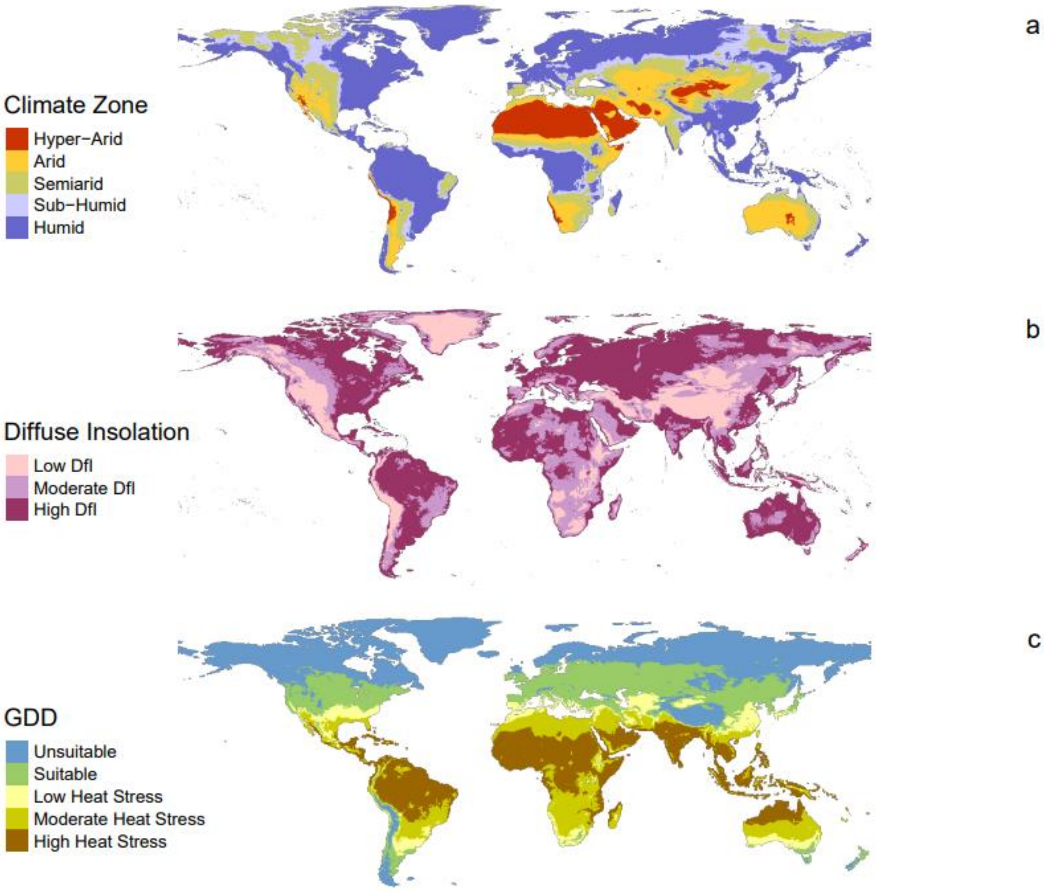


Figure 8, Global extension of the zones of the different factors belonging to the climatic category.

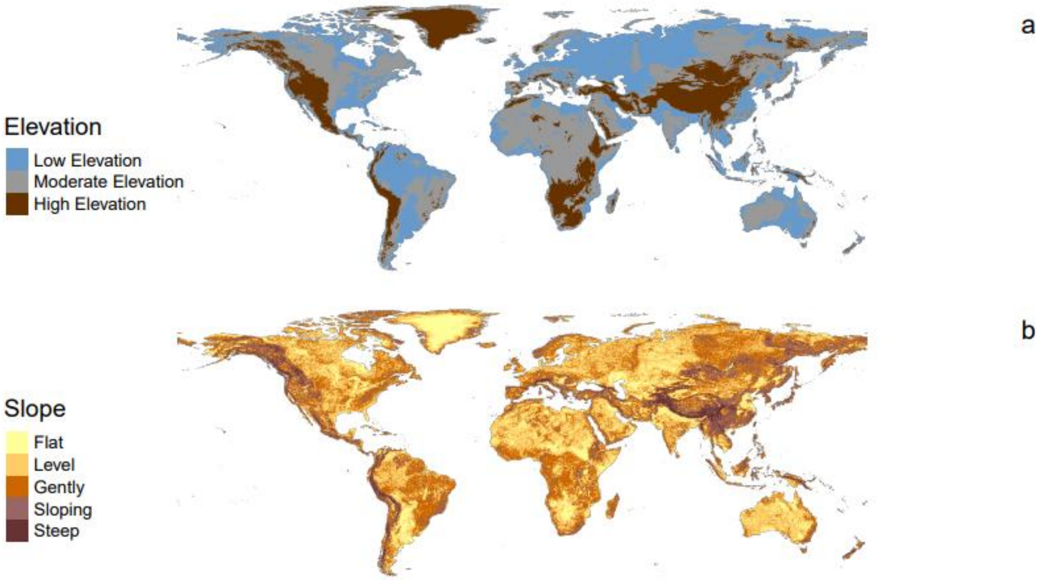


Figure 9, Global extension of the zones of the different factors belonging to the landform category.

Table 3: Observation distribution into the different zones.
The red colour represents the classes which have an overall number of observation lower than 50, while the blue are the classes with an overall number of observation higher than 1000.

	Zone	***NT***		***AF***		***OF***		***All managements***	
		Obs.	*%*	*Obs.*	*%*	*Obs.*	*%*	*Obs.*	*%*
All		2431	82	492	17	33	1	2956	100
BD	Favourable BD	94	3.87	13	2.64	-	0.00	107	3.62
	Transition BD	1129	46.44	399	81.10	27	81.82	1555	52.60
	Restrictive BD	1208	49.69	80	16.26	6	18.18	1294	43.78
Texture	Clay	902	37.10	81	16.46	2	6.06	985	33.32
	Silty	1298	53.39	86	17.48	27	81.82	1411	47.73
	Sandy	231	9.50	325	66.06	4	12.12	560	18.94
SOC	Low SOC	989	40.68	457	92.89	4	12.12	1450	49.05
	Moderate SOC	1169	48.09	33	6.71	24	72.73	1226	41.47
	High SOC	273	11.23	2	0.41	5	15.15	280	9.47
pH	Acidic	2242	92.23	482	97.97	23	69.70	2747	92.93
	Neutral	180	7.40	10	2.03	10	30.30	200	6.77
	Alkaline	9	0.37	-	0.00	-	0.00	9	0.30
Olsen-P	Low P	928	38.17	183	37.20	6	18.18	1117	37.79
	Moderate P	603	24.80	147	29.88	15	45.45	765	25.88
	High P	900	37.02	162	32.93	12	36.36	1074	36.33
AI	Hyper-Arid	-	0.00	-	0.00	-	0.00	0	0.00
	Arid	50	2.06	-	0.00	-	0.00	50	1.69
	Semiarid	573	23.57	176	35.77	-	0.00	749	25.34
	Sub-Humid	355	14.60	110	22.36	2	6.06	467	15.80
	Humid	1453	59.77	206	41.87	31	93.94	1690	57.17
DfI	Low DfI	326	13.41	67	13.62	-	0.00	393	13.29
	Moderate DfI	310	12.75	291	59.15	-	0.00	601	20.33
	High DfI	1795	73.84	134	27.24	33	100.00	1962	66.37
GDD	Unsuitable	18	0.74	-	0.00	-	0.00	18	0.61
	Suitable	1715	70.55	-	0.00	31	93.94	1746	59.07
	Low Heat Stress	276	11.35	-	0.00	-	0.00	276	9.34
	Moderate Heat Stress	249	10.24	360	73.17	-	0.00	609	20.60
	High Heat Stress	173	7.12	132	26.83	2	6.06	307	10.39
Slope	Flat	463	19.05	-	0.00	-	0.00	463	15.66
	Level	1174	48.29	218	44.31	6	18.18	1398	47.29
	Gently	638	26.24	272	55.28	27	81.82	937	31.70
	Sloping	155	6.38	2	0.41	-	0.00	157	5.31
	Steep	1	0.04	-	0.00	-	0.00	1	0.03
Elevation	Low Elevation	1046	43.03	130	26.42	21	63.64	1197	40.49
	Moderate Elevation	1009	41.51	124	25.20	12	36.36	1145	38.73
	High Elevation	376	15.47	238	48.37	-	0.00	614	20.77

For having more information about the cropland division into the different environmental variables and the observation distribution into the classes, it is possible to look at the *Appendix 3A*.

3.1.2 Effect Size (ES) variation for management with different variables

The distribution of the ES (i.e. effect size) of the different management within the defined zones of the variables belonging to the soil, climate, and landform category, is shown in *Figure 10*.

Looking at the different boxplots (*Figure 10*), it is evident that AF tends to have the highest positive impact on maize yield compared to NT and OF with an increase of about 30% in average (*Appendix 2*). Under AF, all the zones of the variables have a positive ES value except for *Favourable BD,* the only class in which about 8% reduction in productivity occurs (*Figure 10, Appendix 2*). The highest increase on maize yield under AF was observed under higher P (50%, *Appendix 2*), in *Sandy soil* (42.1%, *Appendix 2*) and in *Semiarid* areas (40.4%, *Appendix 2*) while no significant change can be recorded for *Low P* (8.6%, *Appendix 2*), *High Heat Stress* GDD (8.3%, *Appendix 2*), and *Low Elevation* (9.4%, *Appendix 2*).
Under NT, about 6% reduction in productivity occurred in average (*Appendix 2*) when considering all the factor zones. However, some increase in maize yield occurred in *Sandy soil* (almost 1%, *Appendix 2*), *Unsuitable GDD* (2.8%, *Appendix 2*) *High Heat Stress* (0.5%, *Appendix 2*) and *Arid climate areas (*6.8%, *Appendix 2)* where this management was the most impactful. The lowest ES was recorded for neutral pH class with a reduction of about 18% (*Figure 9, Appendix 2*). The average value suggests however that the implementation of this RP (i.e. regenerative practice) does not lead to excessive changes in productivity. In fact, about 88% of the identified zones show an ES mean value between -10% and +10% (*Appendix 2*).
OF appears to be the RP that presents the least benefit on maize yield in its implementation with no positive average ES values (*Appendix 2*). The only cases in which it presents a relative productivity greater than NT are: *Neutral* pH, *Low P*, *Clay* texture, and *Sub-humid* climate (*Figure 10*). Its best performance occurred with *High Heat Stress* GDD (> 6000°C/y, *Table 2*) with only 3% reduction while the *Acidic* soils recorded the highest reduction with 25% decrease in productivity (*Appendix 2*).

In order to have more information about the mean value and the standard deviation for each zone, it is possible to have a look at *Appendix 2*. While into the *Appendix 3B*, there are more information about the boxplot and the outliers.

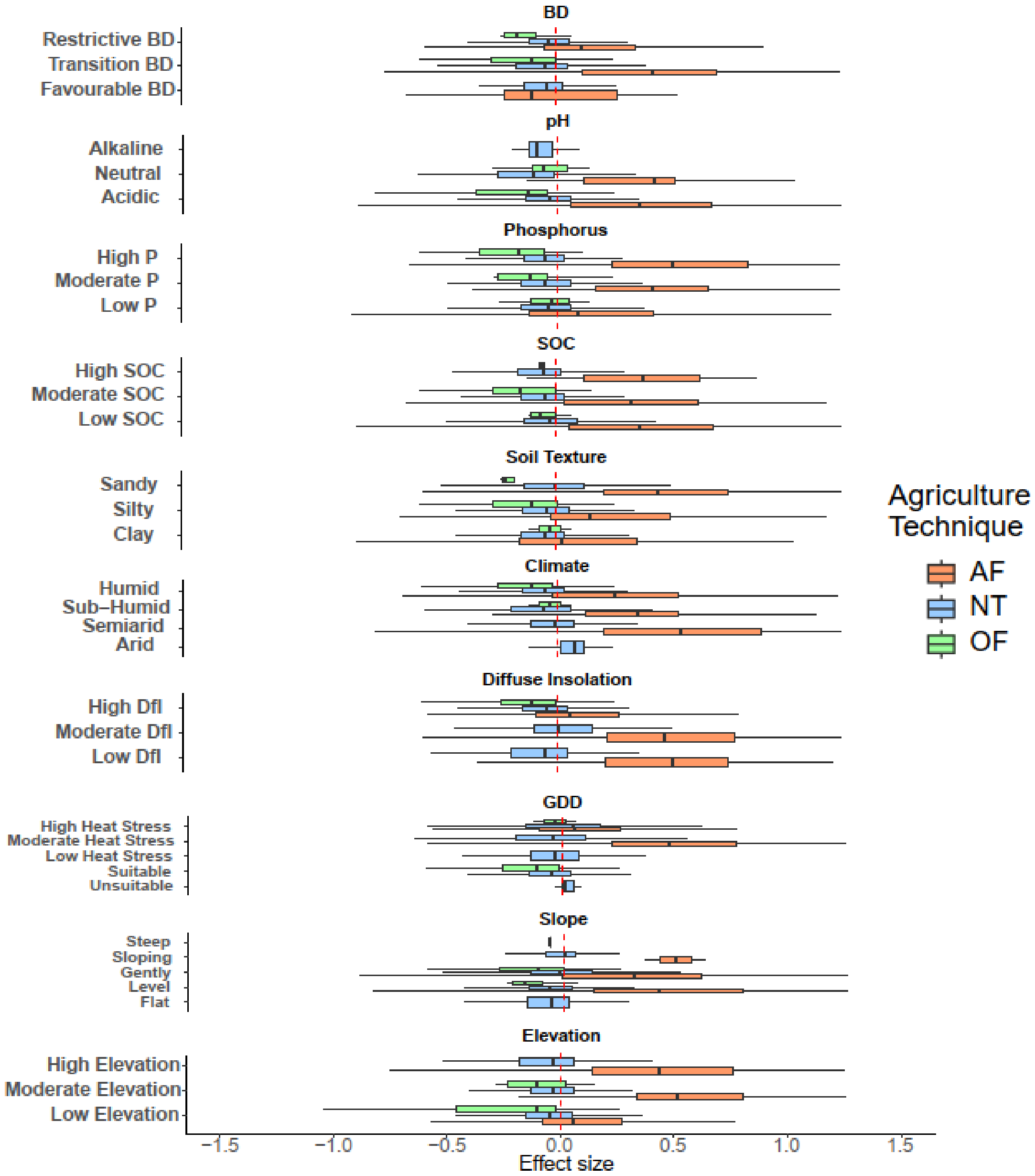


Figure 10, Boxplot of the effect size for the regenerative managements within the zones of soil, environmental and landform factors. The ends of the box represent the 25th and 75th quantiles and the median value is reported in the middle of the box. The horizontal black lines report the highest and the lowest values of the class excluding the outliers. They are also called whiskers, and their length is equal to 1.5 times the interquartile (75th quartile minus 25th quartile). The values which are not included in this range are called outliers.

It is important to emphasise that outliers were not reported in the boxplots. These are observations that have a value greater than the upper whisker (75^{th} quantile plus 1.5 times the interquartile), and lower than the lower whisker (25^{th} quartile minus 1.5 times the interquartile). *Table 4* shows both the number of outliers per management and the total number of observations

NT recorded the lowest percentages of outliers, while the highest were found with AF followed by OF (*Tabel 4*). The zones which do not have outliers are the ones with the lowest amount of observation. For the NT, the highest percentage of outliers (24.18%) occurred with *High SOC* (> 10 g/kg, *Table 2*), while under OF highest number of outliers (23.81%) were found in the *Low Elevation* zone. This allows us to understand that for these two farming techniques, the boxplots consider almost all the observations (around 95% for NT, and 90% for OF) while for AF about 30% of the data were considered as outliers. The minimum percentage values of outliers for NT, OF and AF are 2.13% (*Favourable BD*), 15.15% (*High diffuse insolation*) and 18.46% (*Low elevation*) respectively. The worst situation occurs in the *Semiarid* climatic zone for the AF, where outliers account for 62.5% (*Table 4*). This means that more than half of the observed values belonging to this zone are not represented by the boxplots.

From the description of the boxplots and the analysis of the outliers, it is interesting to note that the zoning of the different factors seems to have more homogeneous values for NT and OF, while the range of ES covered by AF seems to be wider in each zone. This may be due to three reasons. The first concerns the number of observations for classes. In fact, it is possible that by increasing their number, the value of the ES for each zone stabilises (e.g. *Moderate elevation* zone for the NT). On the other hand, having a small number of observations may mean that there is a low dispersion, but that the value obtained from the statistical analysis is not representative of the class, this occurs in several classes: *Unsuitable* GDD, *Steep* slope, and *Alkaline pH* for NT; *Sloping* slope for AF; and *Clay* texture, *Sub-humid* climate, and *High Heat Stress* GDD for OF (*Figure 10,11* and *12*; *Tabel 3*).
The second could be that the zone division is optimal for some RP and not for others. In fact, the NT management seems to fit better to the class division compared with the AF, particularly for the BD, SOC where NT presents its lower deviation standard for every zone, while it is the wider for AF (*Appendix 2*). The last possibility is that the ES does not strictly depend on that factor and therefore either the average values are the same for each zone (e.g. elevation for the NT, *Figure 12*), or there is high dispersion (P or SOC for AF, *Figure 11*).

Table 4: Outliers table.

	Zone	NT			AF			OF		
		Outliers	*Obs*	*%*	*Outliers*	*Obs*	*%*	*Outliers*	*Obs*	*%*
BD	Favourable BD	2	94	2.13	3	13	23.08	-	-	-
	Transition BD	108	1129	9.57	193	399	48.37	5	27	18.52
	Restrictive BD	69	1208	5.71	23	80	28.75	0	6	0.00
Texture	Clay	64	902	7.10	23	81	28.40	0	2	0.00
	Silty	85	1298	6.55	28	86	32.56	5	27	18.52
	Sandy	30	231	12.99	168	325	51.69	0	4	0.00
SOC	Low SOC	84	989	8.49	203	457	44.42	0	4	0.00
	Moderate SOC	29	1169	2.48	15	33	45.45	4	24	16.67
	High SOC	66	273	24.18	1	2	50.00	1	5	20.00
pH	Acidic	163	2242	7.27	215	482	44.61	5	23	21.74
	Neutral	16	180	8.89	4	10	40.00	0	10	0.00
	Alkaline	0	9	0.00	-	-	-	-	-	-
Olsen-P	Low P	84	928	9.05	58	183	31.69	0	6	0.00
	Moderate P	29	603	4.81	69	147	46.94	3	15	20.00
	High P	66	900	7.33	92	162	56.79	2	12	16.67
AI	Arid	2	50	4.00	-	-	-	-	-	-
	Semiarid	44	573	7.68	110	176	62.50	-	-	-
	Sub-Humid	47	355	13.24	40	110	36.36	0	2	0.00
	Humid	86	1453	5.92	69	206	33.50	5	31	16.13
DfI	Low DfI	28	326	8.59	38	67	56.72	-	-	-
	Moderate DfI	46	310	14.84	156	291	53.61	-	-	-
	High DfI	105	1795	5.85	25	134	18.66	5	33	15.15
GDD	Unsuitable	0	18	0.00	-	-	-	-	-	-
	Suitable	97	1715	5.66	-	-	-	5	31	16.13
	Low Heat Stress	21	276	7.61	-	-	-	-	-	-
	Moderate Heat Stress	46	249	18.47	194	360	53.89	-	-	-
	High Heat Stress	15	173	8.67	25	132	18.94	0	2	0.00
Slope	Flat	40	463	8.64	-	-	-	-	-	-
	Level	55	1174	4.68	106	218	48.62	0	6	0.00
	Gently	83	638	13.01	112	272	41.18	5	27	18.52
	Sloping	1	155	0.65	1	2	50.00	-	-	-
	Steep	0	1	0.00	-	-	-	-	-	-
Elevation	Low Elevation	71	1046	6.79	24	130	18.46	5	21	23.81
	Moderate Elevation	70	1009	6.94	75	124	60.48	0	12	0.00
	High Elevation	38	376	10.11	120	238	50.42	-	-	-

3.1.3 Relationships between regenerative practice (RP), effect size (ES), and environmental factors

It is interesting to notice that in many cases the observation points are distributed into vertical columns suggesting that many points with same variable value have different ES (*Figure 10*, *Figure 11*, *Figure 12*). The relationships between the ES and different factors are presented in *Figure 10*, *Figure 11*, and *Figure 12*. The R^2 values appeared to be very low no matter the management and variables (< 10%) except for pH under OF which recorded $R^2 = 0.3$. While no clear trend could be recorded under NT and OF, climatic factors and the SOC generally decreased with increasing ES with the opposite for pH and phosphorus under AF.

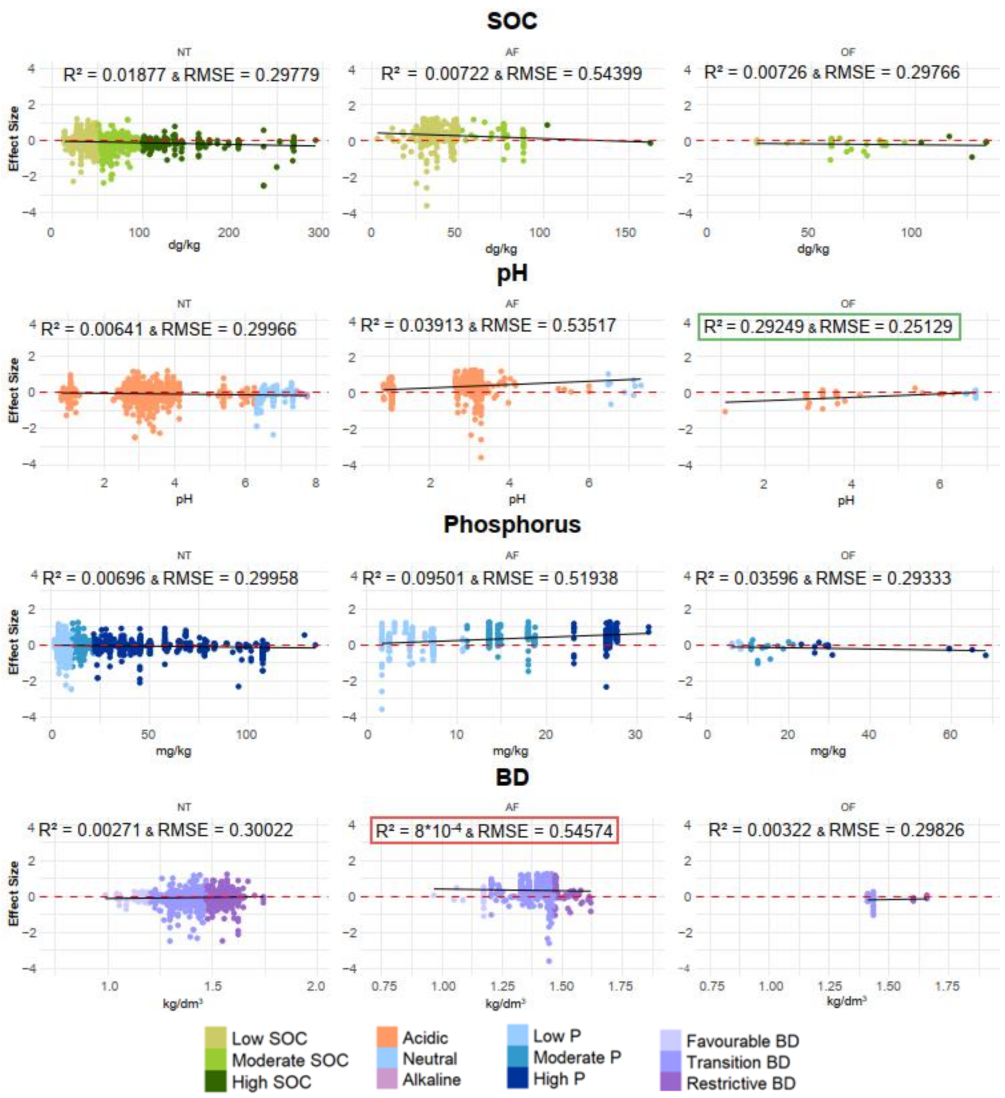


Figure 11, Relationship between the effect size and the soil factors. Black line: trend line, red dotted line: effect size equal 0, red rectangle: linear model with worst estimation between all the soil variables, green rectangle best estimation between all the soil variables.

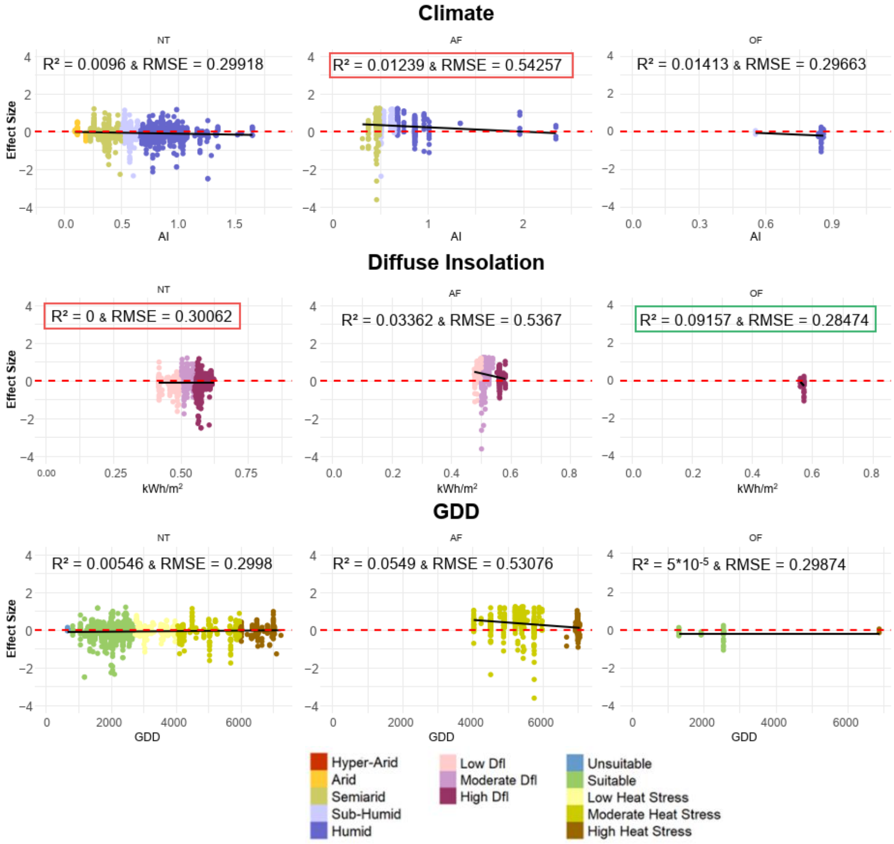


Figure 12, Relationship between the effect size and the climatic factors. Black line: trend line, red dotted line: effect size equal 0, red rectangle: linear model with worst estimation between all the soil variables, green rectangle best estimation between all the soil variables

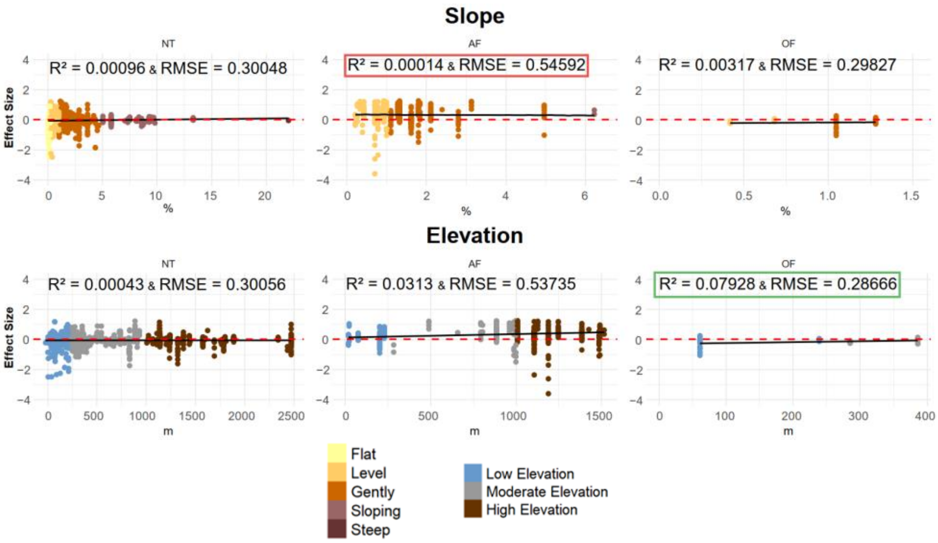


Figure 13, Relationship between the effect size and the landform factors. Black line: trend line, red dotted line: effect size equal 0, red rectangle: linear model with worst estimation between all the soil variables, green rectangle best estimation between all the soil variables.

3.2 Effect size (ES) prediction using Random Forest (RF) model

The following section will describe the result got from the RF (i.e. random forest) prediction model. The first part will present the model performance followed by the variable importance based on the SV.

The second part is related to the predictive maps including those displaying where the RP have a higher productivity than the conventional.
It will present the predictive maps, a statistical table which would quantify the average reliability of the model and the map of uncertainties that will describe the accuracy of the predicted value for each cell.

Finally, the last maps intend to show in which cropland the RP have a positive effect on the maize yield and which prediction has the highest ES.

3.2.1 Model performance

The performance of the RF model in predicting the ES is summarized in *Table 5* for both the training and the test set data. The average values of accuracy of the model are not optimal. The R^2 is always lower than 0.5 for both the training and the test data. The highest performance for testing the model occurred with modelling with all the dataset (R^2 = 0.25) and under NT (R^2 = 0.23). Performance after cross-validation resulted in low R^2 values for OF (R^2 = 0.12) and even lower for the AF (R^2 = 0.07).

Table 5: Performance statistical values of the random forest for predicting the effect size and spatial variable used for the spatial cross-validation.

		R^2	***RMSE***	***Spatial variable***
All	*Train data*	0.317	0.244	Coordinates
	Test data	0.258	0.353	
NT	*Train data*	0.359	0.214	Coordinates
	Test data	0.237	0.252	
AF	*Train data*	0.072	0.536	Climatic class
OF	*Train data*	0.126	0.208	Region

3.2.2 Variable importance based on Shapley value (SV)

Figure 13 reports the partial dependency plots which show the influence of the predictors on the prediction of the ES. The graphs are divided into the four different predictions and report the main three variables used from the machine learning model. While *Figure 15* exhibits the spatial distribution of the variables with the highest SV according with the prediction.

The first graphs in *Figure 14* are related to the *All managements* prediction. In this case, the main variable is GDD. The ES tents to be positively influenced when it belongs to the *Moderate Heat Stress* class (4000-6000°C/y, *Table 2*) and when the elevation has values between 500 and 1000 m (*Moderate Elevation*, *Table 2*). Indeed, there is a negative covariate contribution for higher clay content (>40 %). The GDD covers almost all the cropland (88.74%) followed by the clay (6.01%), and the elevation has the lowest percentage (*Figure 15a*).

For NT, the most important average contributor to the ES turns out to be AI (*Figure 14b*). It contributed best for values < 0.4 (*Arid* areas, *Table 2*) and a small range of humid area (0.8-1.0) compared to very humid areas when index values are beyond 1. Considering the DfI, it has a positive influence on the productivity in the second half of the *Moderate DfI* class (0.50-0.55, *Table 2*) while increasing values of BD seemed to be associated with either decreasing (up to 1.2) or neutral ES (> 1.2). Spatial coverage of these factors shows that Dfl ranked first covering about 83.26% of the cropland followed by AI (13.14%) and BD (< 4%) (*Figure 15b*).

For AF (*Figure 14c*) the most significant variable is the pH and it covers about 72.92% of the cropland in *Figure 15c*. When its value is between 1 and 2 there would seem to have the greatest negative impact on the value of productivity. In contrast, the range in which the covariate contribution seems to be positive is between 2.5 and 4. On the other hand, BD appeared to have positive effects when the observation belonged to the *Transition BD* zone (1.2 - 1.47 kg/dm^3, *Tabel 2*) while a negative influence is recorded when the observation belongs to the *Restrictive BD* class (> 1.47 kg/dm^3, *Table 2*). With regard to SOC, the SV trend is variable in the *Low SOC* class (< 5 g/kg, *Table 2*), while it appears to have a positive influence when the observations belong to *Moderate SOC* zones (5-10 g/kg, *Table 2*). However, only the pH associated the BD have the highest spatial influence in ES prediction for most cells in the AF grid prediction (*Figure 15c*).

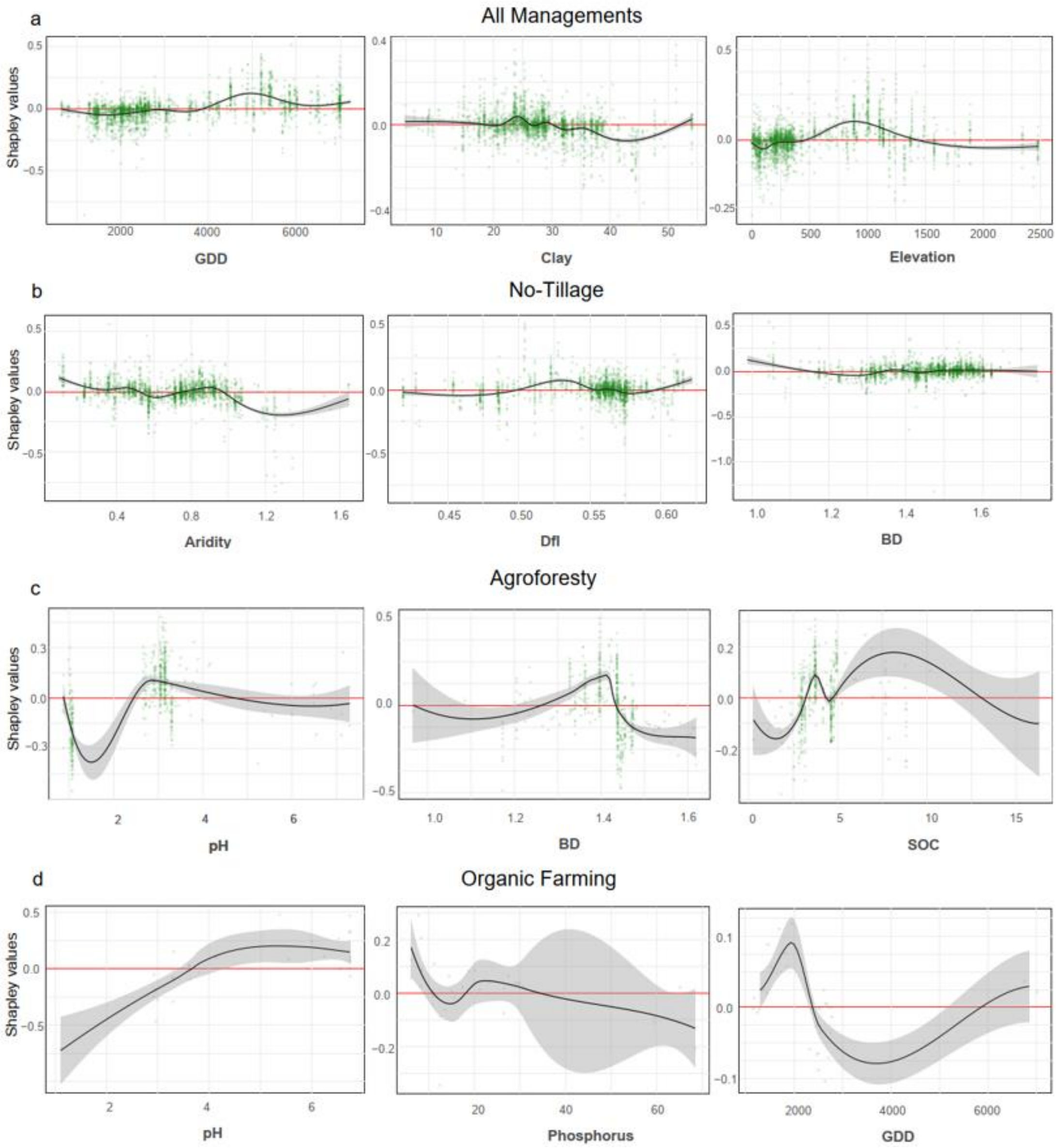


Figure 14, Partial dependency plot for the three more relevant variables for each prediction. The red line underlines the division between a positive implementation (below the red line) of the regenerative management for a negative result (under the red line). The green dots represent the known point from the dataset and the SV computed from the model. These graphs help to understand the relationship between the predicted variable (ES) and the predictor used in the model; therefore, the black line can be seen as the trend that describe the relationship between the two variables according with the RF model.

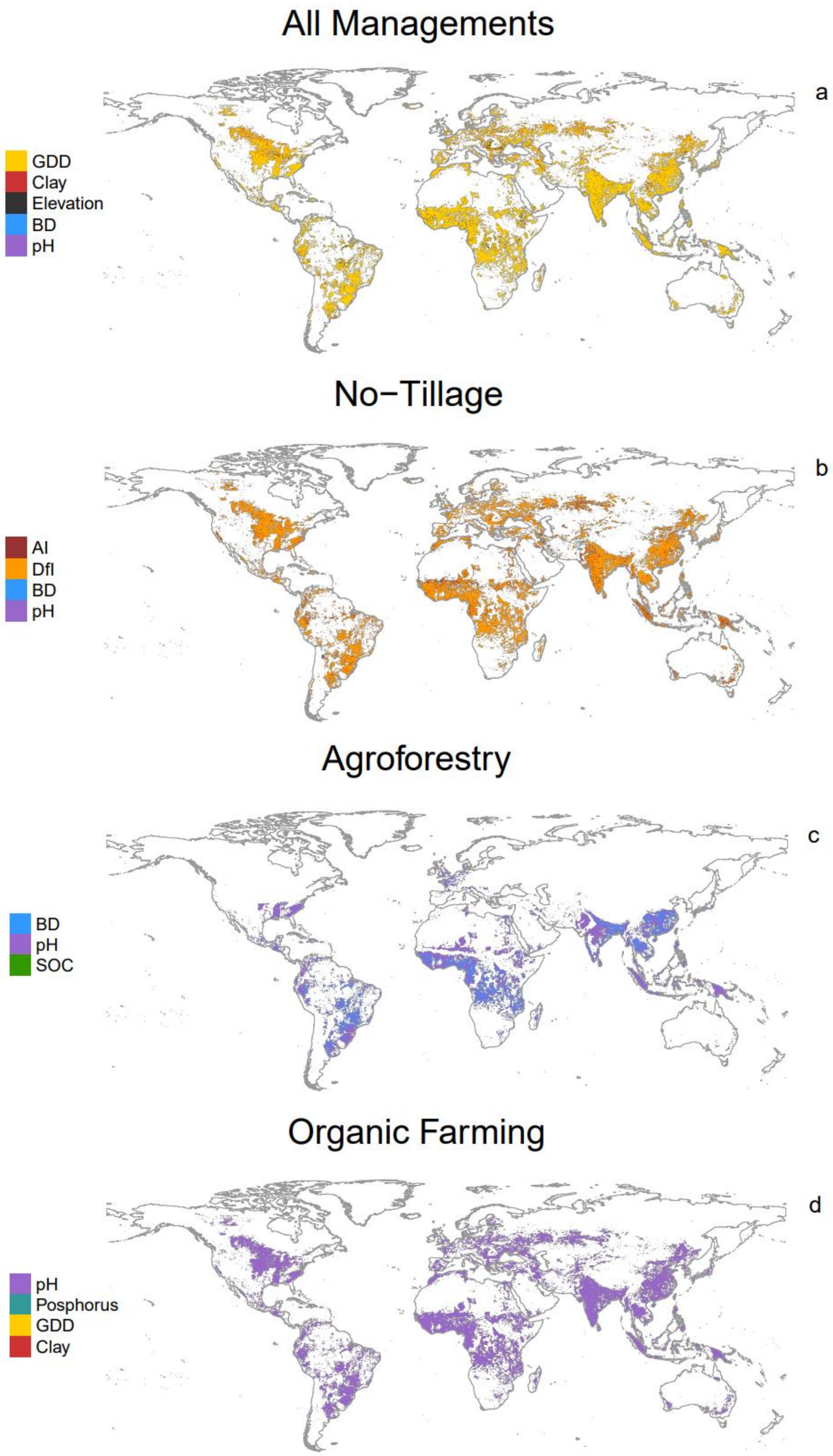


Figure 15, Shapley value maps. It shows the spatial distribution of the predictors with the highest SV. It allows to understand the variable with the highest relevance used form the predictive model. On the left side of the images are reported all the main predictors that have been used to obtain the maps.

The pH turned out to be also the key variable affecting the ES under OF (*Figure 14d*). In this case, it appeared to reduce the maize yield when values are below 4 (*Hyper-Acidic* pH) but a higher positive contribution when beyond (*Acidic* pH). It is the most relevant predictor for almost all the OF map (99.4%) (*Figure 15d*). For P, there seems to be a positive contribution when the *Low P* class occurs (< 10.9 mg/kg, *Table 2*), while high values lead to the decrease of the productivity (> 40 mg/kg). The GDD is the one with the lower SV range covered, but it is interesting to notice that it leads to a decrease in productivity for the *Low Heat Stress* and *Moderate Heat Stress* GDD (2700 - 6000°C/y, *Table 2*), while it has a positive influence within the *Suitable* GDD class (800 - 2700°C/y, *Table 2*) and *High Heat Stress* GDD (> 6000°C/y).

Considering all the predictions, it seems that the key variables across managements are the pH and BD, since they were involved in most of the models even if with different percentage of influence. Though P, elevation, GDD, and SOC were involved only in one predicted model, they still had main role in the predictions but just in the single management.

3.2.3 Spatial distribution of the Effect Size (ES) and uncertainty maps

Figure 16 shows the predictive and the uncertainty maps for all the dataset and for each management. They allow to understand how the maize yield varies across the globe and in which geographical regions the model is more accurate. While *Table 6* allows to understand the average ES values for every continent.

The implementation of NT (*Figure 16b*), seems to provoke a general decrease in the productivity of around -7.38% (*Table 6*). The area with a positive ES covered 26.35% of the cultivated land and are mostly located in India, South of Europe, North Africa, and in the eastern United States. The most frequent values are between -15% and -10% (*Figure 16b*). Compared with the *Figure 16a*, it usually has lower predicted values. Considering the uncertainty map (*Figure 15b'*), the CI is between 0.14 and 3.37. Most of the cropland has an uncertainty around 0.8 and 1 and that it has higher accuracy compared with the '*All management*' map (*Figure 16a'*). The areas with the highest uncertainty seem to be the United States and Central India.

All the cropland has a positive value of ES if the AF prediction is considered (*Figure 16c, Tabel 6*). The lowest relative productivity value is 3%, and the highest is 66%. The area which reaches an increase of yield higher than 20% covers 82.20% of the land, and the most frequent values in the map are around 40% and 45% (*Figure 16c*). The regions with the best result seem to be Central Africa and China. It is the management which has the most similar ES distribution compared with the '*All managements*' prediction. The uncertainty map (*Figure 16c'*) shows that the CI values are around 0.9 and 2.0. The places with the lowest accuracy are the North India and the United States. The most frequent interval is between 1.2 and 1.4. Therefore, it is the prediction with the highest uncertainty compared with the other three RP.

The OF prediction seems to have the highest reduction of productivity in Europe, while the best result is in Southern part of the world with an increase equal to 1.32% in South America (*Figure 16d, Table 6*). The values are between -27% and +12%. ES mostly is around -8% and -6%. The uncertainty map (*Figure 16d'*), in this case, seems to have more homogeneous CI values compared with the others as they are between 0.22 and 1.31. The highest accuracy seems to be the North America and the most frequent range for CI is around 0.8 and 0.9.

It is interesting to notice that the dataset prediction (*Figure 16a*) seems to not be related with the other managements. Since NT is considered, they have a different ES distribution in the space, while AF has a similar spatial trend but the variable range do not match with the one of the first prediction. OF seems to be the most similar (same ES value for Europe, and Africa), however the CI between the two estimations do not coincide. This means that the regenerative techniques have different effect on the land productivity and that it cannot be describe properly by the "*All management*" prediction.

Table 6: The table shows the average ES predicted value according with the world regions divided for each management.

Region	***NT***	***AF***	***OF***
North America	-5.37 %	31.19 %	-3.59 %
South America	-13.89 %	55.57 %	1.32 %
Europe	-8.36 %	36.45 %	-9.51 %
Africa	-6.08 %	35.20 %	-2.00 %
Asia	-6.57 %	29.98 %	-3.25 %
Australia	-1.73 %	29.78 %	-5.15 %
World	-7.38 %	32.76 %	-3.11 %

For more information about the correlation of the predictors for the different predictive models have a look at *Appendix 4A*. While, to have a better visualization of the predictive maps, more images are available in the *Appendix 4B*.

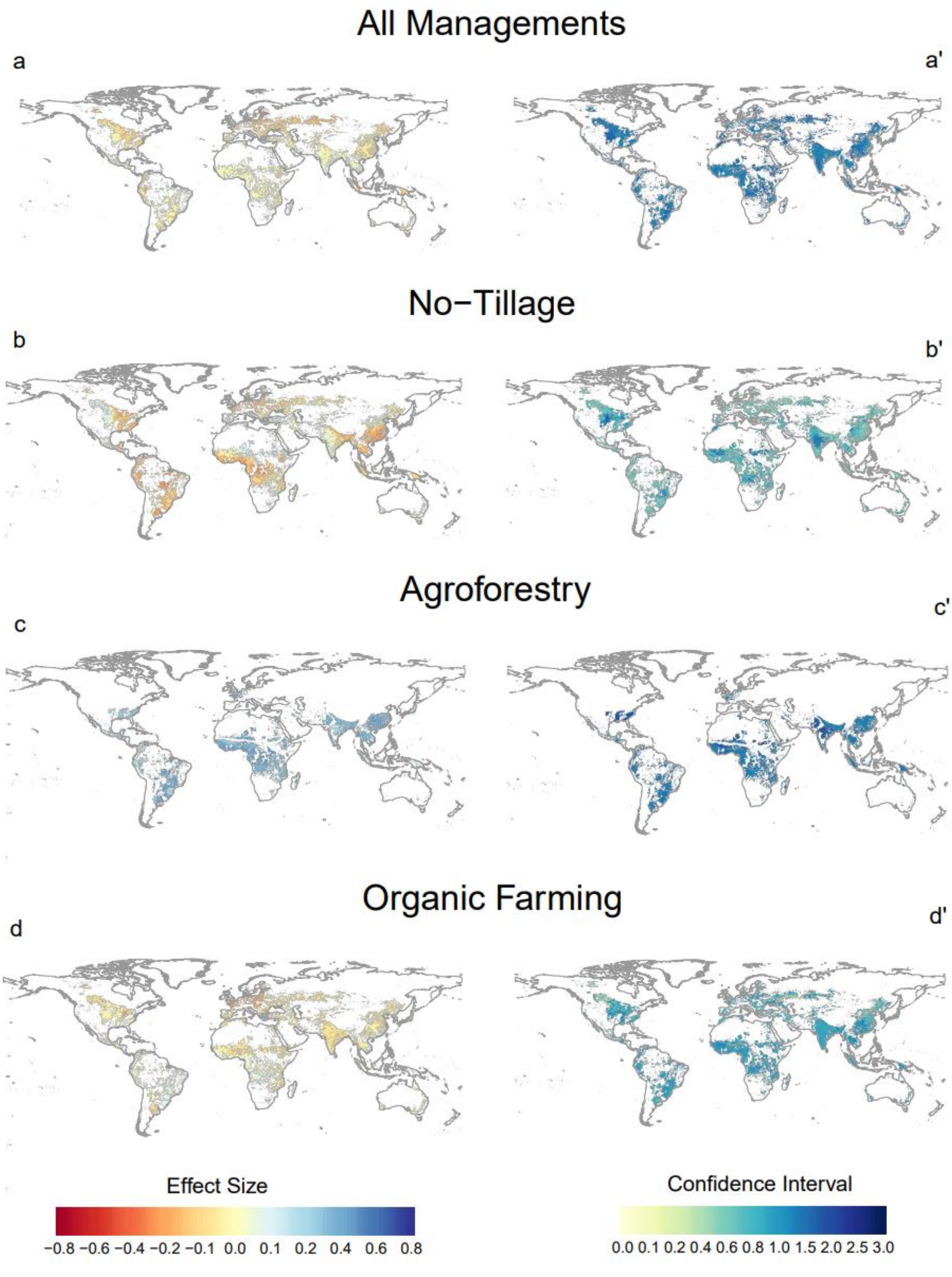


Figure 16, Prediction and uncertainty maps for all the models. On the left side are reported the effect size map. While, the right side shows the confidence interval of the predictions. On the bottom of the figure the legends are reported. For all the maps has been used the same scale.

3.2.4 Positive implementation of the regenerative practices (RPs)

The aim of this section is to define where the ES is positive and which technique has the highest increase in productivity.

Looking at *Figure 17a*, the area where the AF has the highest ES cover almost 60% of the cropland followed by the NT with a percentage around 17% and by the OF (around 6%). Considering the distribution of the managements, the AF seems to obtain the best result in tropical and equatorial zones mostly, and in North Europe. The NT and OF, instead, are mainly located in the United State, Central Europe, and Central Asia. It is interesting to notice that seems to do not be a location where all the managements have negative ES. On approximately 18% of the cropland both NT and OF lead to a reduction of the productivity, they are mostly located into the tropical and temperate climatic classes where the AF prediction is not considered. Its uncertainty map (*Figure 17b*), shows that the cropland with the best accuracy is the northernmost area of the globe. The highest CI is located in the coordinates where the AF has the highest increase in the productivity, while it has the lowest value for the NT cells. The confidence interval is between 0.20 and 2.77, and most of the cell has a value around 1.2 and 1.4.

AF has positive values for almost all the cropland in which it is defined (60%, *Figure 17c*). Consequently, it appears to be the management with the highest number of cells that have a growth in the maize yield with the highest ES value (*Figure 17a,c*). The percentage of the world where the NT appears to increase the productivity is around 26%, while for the OF is 35% (*Figure 17c*). Less than 1% of the cells present a positive ES for all the managements considered together and they are mostly located in Central Africa and in the Middel Easte countries. Most of the map present a positive implementation for both OF and AF (27%), usually this class is located in South America, Central Africa, and South-East Asia. NT and AF, indeed, covered 9% of the cropland and are mostly in North Africa. While NT and OF (2%) usually have positive effects in North-West America. The uncertainty map (*Figure 17d*) is similar to the previous one, both have lower values on the northern part of the globe and the lowest accuracy belong to the region where there is the AF.

The results in *Figure 17* seem to be too optimistic with the RP having the potential to augment the maize yield almost all over the globe. For this reason, another map is produced to show the pattern in productivity distribution if only the NT and OF managements are considered. The AF has been excluded because it seems to be the least accurate prediction (*Table 5*). The results are shown in *Figure 18*.

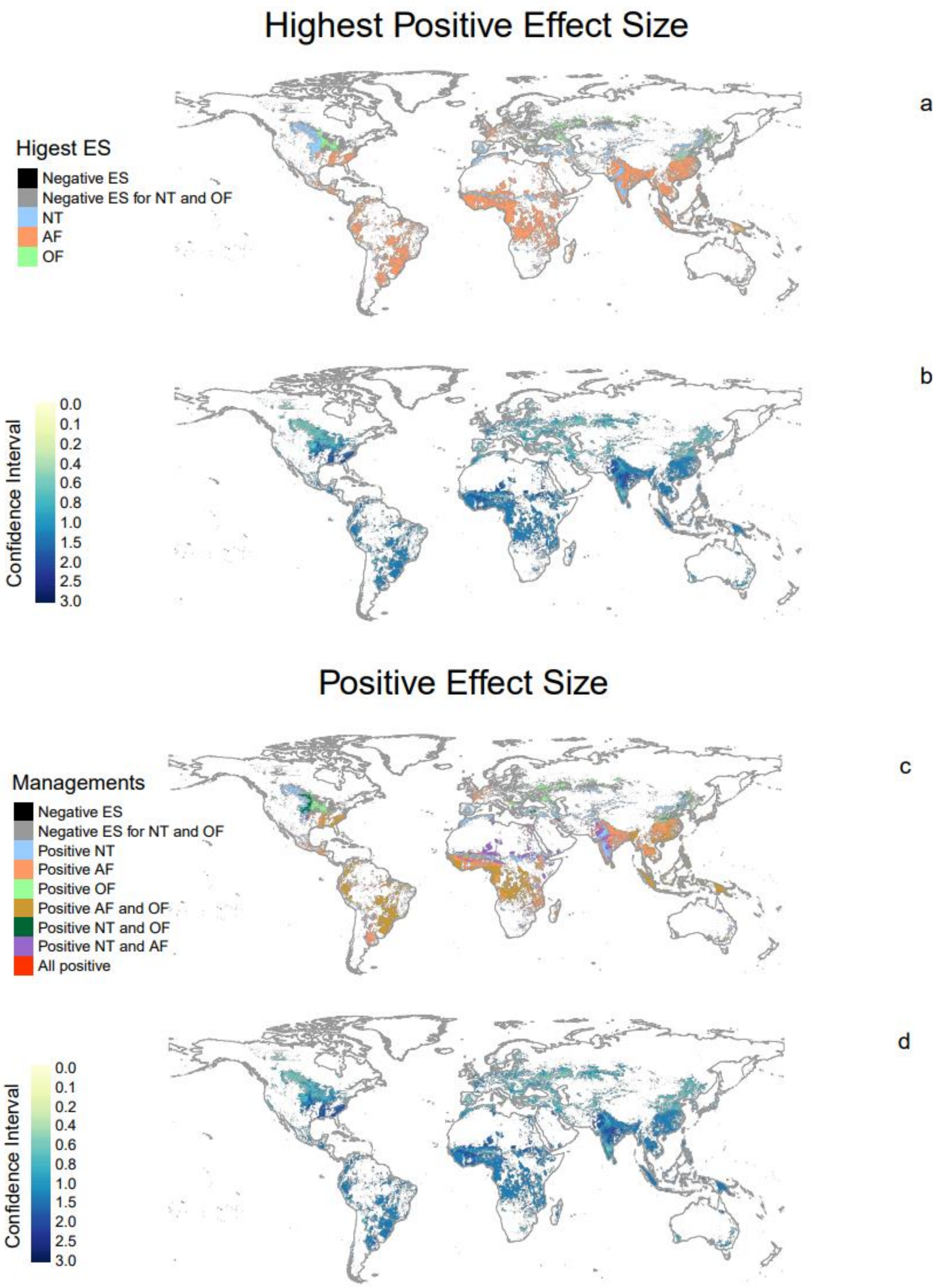


Figure 17, Location where the regenerative managements could lead to an increase in productivity. Map a: cropland with the highest positive management effect size; map b: uncertainty map for map a. It reports the CI of the management shown in map a; therefore, if in one location the AF has the highest relative productivity, in the same position, map b reports the CI of the AF. Map c: managements with positive ES, in this case more than one regenerative technique can be defined for each cell. Map d: uncertainty map of map c, it has been obtained considering in each location which RP has the highest CI compare to the other plotted in the same cell.

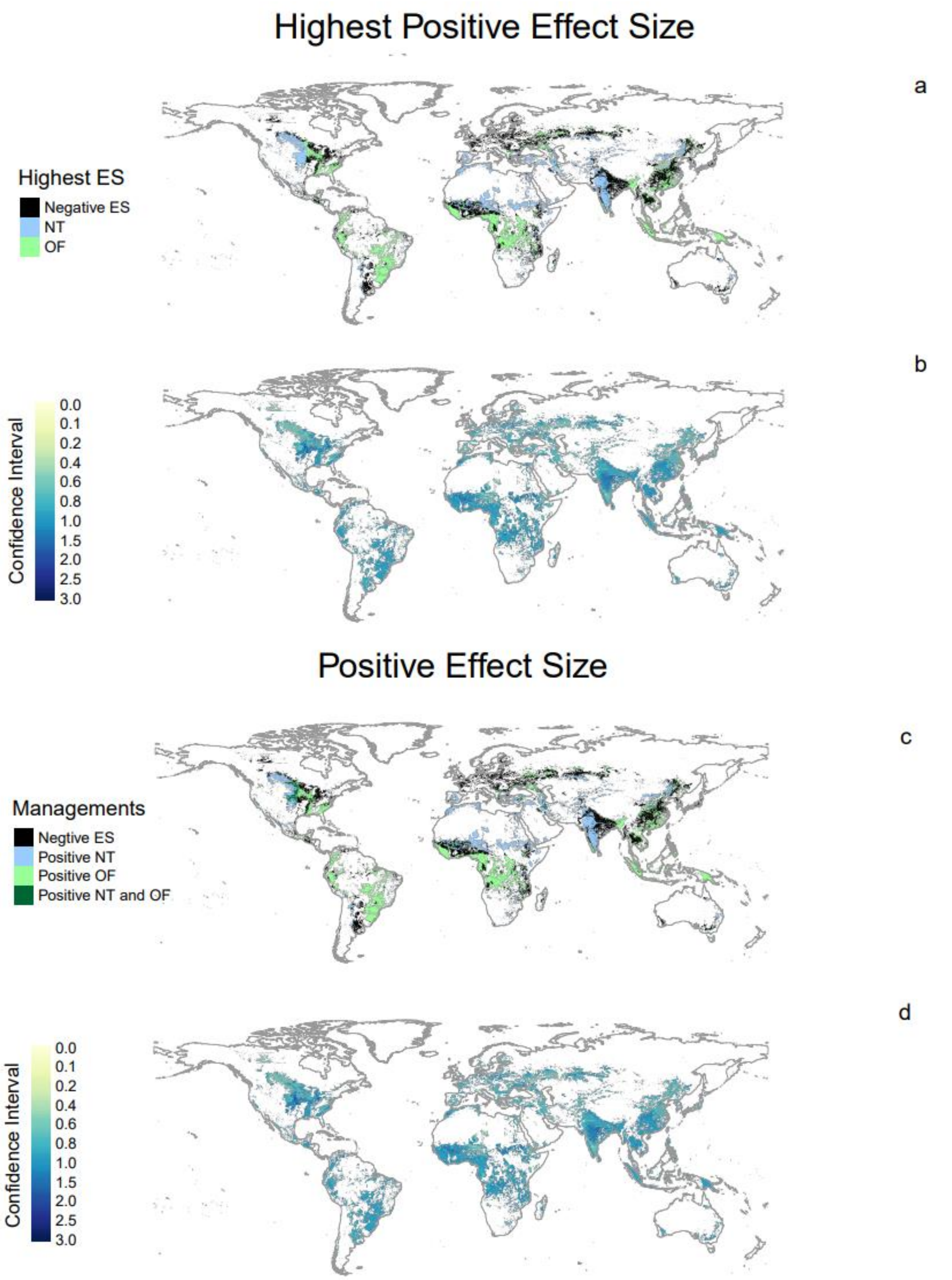


Figure 18, Location where the NT and/or OF could lead an increase in productivity. Map a: cropland which management has the highest positive effect size. Map b: uncertainty map for map a. It reports the CI of the management shows in the map a; therefore, if in one location the OF has the highest relative productivity, in the same position, map b reports the CI of the OF. Map c: managements with positive ES, in this case more than one agricultural strategy can be defined for each cell. Map d: uncertainty map of map c, it has been obtained considering in each location which RP has the highest CI compare to the other plotted in the same cell.

The two managements have a similar distribution compared with the previous image. However, there are 41.12 % of the cropland where the ES has negative values. The NT has the highest ES in the northern part of the globe, while the OF has it in the southern part (*Figure 18a*). Looking at *Figure 18c*, 23.75% of the cells have a positive ES only for the NT, and 32.53% for the OF, while both the managements are positive for 2.60% (*Figure 18c*).

The CI is between 0.21 and 3.37 with most of the values around 0.8 and 1 (*Figure 18b,d*). It is interesting to notice that the most accurate prediction occurs when the ES is negative. Even if the range of the CI is higher, the most frequent value interval is lower in the second prediction. It is therefore possible to state that the reliability of *Figure 18* is higher than the one in *Figure 17*.

To have a better visualization of the predictive maps, more images are available in the *Appendix 4*.

4 Discussion

In this section, the results of the analysis will be discussed and an attempt will be made to understand the limitations and strengths of the analysis. It will also compare the results obtained with literature data. It is divided into three parts. In the first, the accuracy of the model is analysed by comparing R^2 and RMSE with studies that used the same predictive method. In the second, the results related to the environmental factors will be commented on. And finally, the maps obtained through the machine learning process will be analysed.

4.1 Performance of the model

Results showed that prediction accuracy of the RF models were low ($R^2 < 0.4$) especially for AF ($R^2 = 0.07$) and OF ($R^2 = 0.13$). The analysis seems to present less reliable result compared with other studies which used the same predictive method. For instance, a global-scale study that predicted ES values for NT achieved an R^2 value equal to 0.52 (Su *et al.*, 2022). Remaining studies are usually site-specific focusing mainly on maize yield prediction, recording R^2 between 0.35 to 0.85 *(Jeong et al., 2016; Marques Ramos et al., 2020; Burdett and Wellen, 2022; Morales and Villalobos, 2023)*. However, the aim of some of these studies was to demonstrate the predictive ability of the RF model and usually they had a bigger dataset compared to the single RP of this study. Moreover, most of the analyses considered local and not global scale and the predictions are not related to the regenerative management but to detect the maize productivity over the time.

Compared to *Su et al. (2022)* findings, this study considered more observations introducing thereby more variability in the ES of NT. The low accuracy of the AF is probably due to the data distribution. The highest accuracy was gotten when the spatial variable used was '*climate class*'. However, its observations did not cover the whole climatic areas compared to the other two RP as they belong only to the tropical and temperate classes.

A possibility to get more reliable estimation would to be to increase the observation points for each management considering not only the total count but also the spatial distribution particularly for the AF. Additional consideration would be to consider a higher number of predictors such as multispectral images (e.g. Sentinel2 derived) with higher resolution (10 – 30 m) which were reported along with deep learning *(Sayago and Bocco, 2018; Sun et al., 2020; Desloires, Ienco and Botrel, 2023; Mohammad et al., 2023)*.

4.2 Relationship between Effect Size (ES) and environmental factors

The descriptive analysis allows to make some hypothesis about the relation between the ES and the environmental conditions through different managements.

The first aspect visible from the results (Section 3.1.2) is that the implementation of AF seems to lead to a general increase in productivity regardless of the values of the environmental factors considered, whereas OF appears to have almost exclusively negative ES values under all environmental conditions, and NT does not present huge variation of relative productivity. This could mean that AF would be more resilient to different environmental factors compared with the other managements. However, either the high dispersion (high number of outliers, *Table 4*) or the low number of observations belonging to each zone of the variables related to AF do not allow to define reliable average values for its classes. On the other hand, OF does not have enough data per zone. Despite the greater number of observations, the NT also does not return clear pattern for its results.

If we look at the average values for each management, those obtained for NT do not deviate from the results obtained in the study of *Achanken and Cornelis (2023)*. Whereas, when considering the overall analyses for AF, the average ES in this study seems to be optimistic. In fact, according to *Baier et al. (2023),* an increase of 16% occurred in the best cases while from *Appendixes 2* it is visible that the zones mean values are around 30%. For *de la Cruz et al. (2023)* study, OF leads on average to a decrease of 20-50% in productivity while for the present analysis the reduction is around 15%. This is slightly lower than the minimum reported with the previous result. The deviation can be due to the different size of the dataset considered, as the available one has fewer observation points compared with the previous study.

Finally, the linear graphs (*Section 3.1.3*) show that it is not possible to describe the relation between the environmental factors and the ES with a linear model suggesting more complex interactions requiring the implementation of more advanced data analysis models. For this reason, a predictive model was used that also allowed to define the covariate correlation through SV. This analysis showed that the factors that most influence the ES prediction are different for each management and that the prediction considering the entire dataset seems reductive as it is unable to describe the variability of the individual RPs. The most significant results for each management are discussed below.

AI is the most reliable predictor for the NT prediction. The zone which mostly affect positively the ES is the Hyper-Arid and Arid climate zones. This result seems to be coherent with a global study which sustains that the implementation of this management has a probability of around 0.5 to increase the maize productivity into the dry areas (Su, Gabrielle and Makowski, 2021c). Considering site specific studies, it is possible to state that the productivity variation changes according to the climate. For instance, *Galani et al. (2022)* analysis confirms the positive effect of NT practices within the Arid and Hyper-arid climate zones. While it seems to do not have either substantial or positive variation for the Humid climate zone according with the *Achankeng and Cornelis (2023).*

Both the AF and the OF are mostly influenced by the pH. The previous management is negatively affected by pH value lower than 3, while OF for values lower than 4. For AF, a global meta-analysis caried by *Baier (2023),* demonstrates that the most significant implementation for the maize productivity occurs in moderate acidic condition (5-6 pH) while negative effects are registered for hyper-acidic soil (pH lower than 4). Therefore, the two results seem to be coherent for the low pH range, while they give a different result for higher values. However, the available dataset does not have many observations for the neutral pH class to account accurately for such variability. For OF, the study of *de la Cruze et al. (2023)* seems to not detect significant variation for strong acidic pH for the maize yield.

Maize grows better in warmer temperate regions and humid subtropics, in well drained soils such as sandy loam soils, high SOC, flat areas (good soil fertility and sun exposure), and adequate nutrient supply (NPK) *(Otegui and Slafer, 2000; NSW Department of Primary Industries, 2009)*. However, it is interesting to notice that AF and NT, were able to still increase ES or reduce losses in productivity under unfavourable conditions such as arid, semi-arid areas (water stress high), unsuitable or high heat stress, low SOC soils, acidic soil, sloping and gently sloping areas, sandy soils. These results can be justified from the environmental benefits related to the implementation of these managements.

AF, in fact, allows to increase the SOM above and below the ground. Consequently, it enhances the soil health and the SOC availability *(Dollinger and Jose, 2018)*. Moreover, the tree roots increase the water storage and the microorganism presence in the terrain that augment the nutrients in the soil. They also reduce the typical erosion processes in the arid and semiarid climate *(Gayathri et al., 2024)*. On the other hand, NT has benefits on the environment with the coverage of the land surface which reduces its exposition to erosion processes and minimizes evapotranspiration. Moreover, by avoiding the soil disturbance, the microbiological community increases and improve the soil structure *(Choden and Ghaley, 2021)*. From these considerations, it is possible to say that these RP seem to advantage the maize productivity in many critical environmental conditions, thanks to the increase of soil quality (enhance of SOC, better soil structure, pH increase, etc.), soil humidity, and nutrients availability. The situation is different for the OF because the weed control is more challenging with more pest and disease pressures which are higher in this management *(Czarnecka et al., 2022)*. Though it also leads to the some improvement of soil properties (soil structure, SOM, water capacity, etc., *Choden (2021)*, it however requires time to obtain the optimal condition to get a proper productivity *(Durrer et al., 2021)*.

4.3 Predictive maps and positive implementation of regenerative practices (RPs)

The final part of the analysis detected the predictive models and defines in which geographical areas a positive increase in productivity occurs and under which management.

As far as the NT is concerned, there is on average a decrease of 7.38% (*Tale 6*) when it is implemented. The areas with the best results (ES around 0.2) are North-West America, North Africa and India (*Figure 17b*). It is interesting to note that the areas where ES is highest coincide with *Arid* or *Semiarid climate*, *Moderate* or *High DfI*, and *Restrictive BD* zones. A study carried out by *Liu and Basso* (2020) observed how the implementation of this technique in the Mid-West part of the United States led to a reduction in crop productivity loss when adverse climatic phenomena occurred. This technique in addition is used to cultivate 23% of US cropland *(Triplett and Dick, 2008)*. In India, on the other hand, a study by *Pradhan et al. (2016)* showed that maize productivity does not undergo significant changes but that it has substantial benefits to the environmental component. Another analysis in North-West of India location showed that it leads to a productivity increase of around 10% and that on average it reduces the water demand for maize cultivation *(Jat et al., 2013)*. Finally, in North Africa, a study reported that this practice leads to improvements in cultivation, but that productivity is compromised by the presence of weeds when they are not controlled *(Mrabet, 2011)*.

AF has a productivity increase around 32.76% (*Table 6*). The areas where it is most favourable (ES > 0.5) are found in Central-West Africa and China (*Figure 17c*). These areas coincided with the geographical areas belonging to the *Acidic pH, Transition BD,* and *Moderate SOC*. The results seem to be consistent with studies in the literature, as far as Africa is concerned. Numerous analyses have shown the positive effects on productivity linked to this management *(Félix et al., 2018; Choden and Ghaley, 2021)*. For example, a study carried out in Ethiopia showed that productivity can even increase by 50% under particular conditions *(Dilla et al., 2019)*. As far as China is concerned, however, not many croplands are subjected to this RP. In fact, about 1% of the cultivated land are reported to be under AF *(Hong et al., 2017)*. A recent analysis demonstrates that a reduction around 30% occurs when the agroforestry is implemented, this is probably due to the nutrients and water competition between the trees and the maize *(Yang et al., 2023)*. However, some analyses showed that it led to an increase in the soil properties, particularly it enhances the nutrient content in the soil and the SOC; moreover, even if the shadows of the trees reduce the amount of light reaching the crops and the photosynthetic rate, it also increases the light use efficiency raising the soil productivity compared with the sole tree stands and sole crops *(Guo et al., 2020; Qiao et al., 2020; Dong et al., 2021; Wan et al., 2024)*. The discrepancy of the result can be due to the low reliability of the AF predictive model.

Considering OF, it registered an average reduction of 3.11% (*Table 6*). It also tends to perform best in South America and South Africa (*Figure 18d*) with a productivity increase of 10%. The highest values fall in the *Acidic pH*, *Low P*, and *Moderate* or *High Heat stress GDD* zones. Considering previous analyses, there is no much implementation of this management in Africa compared to the other RPs. It might be due the need for more investments as systems with herbicide-mediated weed control were more successful *(Lotter, 2015)*. In fact, only 0.12% of croplands had implemented this RP in 2007 *(El and Scialabba, 2007)*. A global analysis registered a reduction in the productivity equal to 40% in Africa, and equal to 50% in South America *(de la Cruz et al., 2023)*. However, only 12 and 4 observations respectively belong to these two geographical zones limiting the generalization of current analysis.

There are also some locations in which more than one RP can increase the maize productivity (*Figure 17c, Figure 18c*). For instance, in South America, South Africa and in South-East Asia, both AF and OF showed positive ES while in Central Africa the positive managements seem to be NT and AF. Moreover, NT and OF seemed to be the best opportunities for the United States. All the RPs could lead to an augment of maize yield only in 0.33% of the cropland (Middel Easte and Central Africa).

The results obtained from this analysis is mainly a starting point, as the available observations are limited, but at the same time, in certain situations it seems to give more interesting results than previous literature studies.

4.4 Limitations a way forward

The results of the analysis do not have a high reliability on the basis of statistical values. However, the study can be considered as a starting point for future surveys on the productivity of RPs (i.e. regenerative practices).

The main limitations of this study were the low number of observations for OF and the poor spatial distribution of observation points for AF. Another limitation was the available predictors. Many environmental factors were used, but some key aspects such as soil water availability and climatic stress were not considered. Therefore, the dataset should be implemented in order to have a larger number of observations homogeneously distributed in the regions and more soil and climate factors to be sampled in each territory. Finally, the resolution of the multispectral maps used by the RF (i.e. random forest) model could have a higher resolution (10-30 m).

5 Conclusion

The aim of this study was to locate potential areas suitable for regenerative management practices while also assessing the underlying environmental factors. For such purpose, the effect size of three RP (NT, AF, OF) collected from existing global literature was subjected to both descriptive statistics and modelling using RF with the consideration of different environmental factors.

From this analysis it was possible to observe that the variation in productivity related to the distribution of the ES was not only related to the type of management used, but also to various environmental components. From the statistical analysis and the partial dependency graphs, it was possible to deduce that the implementation of NT and AF would seem to have better results when the cultivated sites are in less favourable environmental conditions for the maize growth (Hyper-Arid and Arid climatic zone, Unsuitable and Hight Heat Stress GDD, Low SOC content, Acidic soil). Thanks to the SV rank average of the predictive models, it was also possible to observe that the factors that most influence ES are different for each management. In fact, for NT, it was AI, while pH ranked first for AF and OF. In particular, Hyper-Arid and Arid zones (AI < 0.2) seemed to have a positive covariate correlation for the NT productivity while Arid pH (4-6 pH) positively influenced the maize yield when the OF technique was implemented. Also, AF seemed to be negatively affected by very low pH values (< 2).

Predictive models made it possible to observe how ES varied with each management. The best performing RP is AF, where mostly occurs an increase in productivity of over 20%. While for NT and OF there tends to be a reduction of 10-15% and 6-8% respectively. When considering the spatial distribution, AF and OF had higher values in the southern part of the globe (Central Africa, South America, and South-East Asia), while NT in the equatorial zone. Considering the accuracy of the RF model based on R^2 values, most reliable estimate turned out to be NT ($R^2 = 0.36$), followed by OF ($R^2 = 0.13$), and finally AF ($R^2 = 0.07$) which did not seem to return reliable values. Once OF is discarded considering the low accuracy of its model, the geographic area where NT seemed to lead to the highest increase in productivity were located in North-West America, North Africa, and India, while those where OF had the highest potential were South America and South Africa. Both RP, however, presented positive ES value in North-West America.

This study could be used as a preliminary analysis to understand whether the implementation of an RP might lead to benefits in maize production. Furthermore, it showed that potential for positive influence would be highest in climatic zones of high stress, and therefore emphasised the importance of conducting more in-depth global analyses to obtain more reliable results.

6 References

Achankeng, E. and Cornelis, W. (2023) 'Conservation tillage effects on European crop yields: A meta-analysis', *Field Crops Research*, 298. Available at: https://doi.org/10.1016/j.fcr.2023.108967.

Ahmad, L. *et al.* (2017) 'Growing Degree Days to Forecast Crop Stages', in *Experimental Agrometeorology: A Practical Manual*. Springer, Cham, pp. 95–98. Available at: https://doi.org/10.1007/978-3-319-69185-5_14.

Ahvo, A. *et al.* (2023) 'Agricultural input shocks affect crop yields more in the high-yielding areas of the world', *Nature Food 2023 4:12*, 4(12), pp. 1037–1046. Available at: https://doi.org/10.1038/s43016-023-00873-z.

Allakonon, M.G.B. *et al.* (2022) 'Grain yield, actual evapotranspiration and water productivity responses of maize crop to deficit irrigation: A global meta-analysis', *Agricultural Water Management*, 270, p. 107746. Available at: https://doi.org/10.1016/J.AGWAT.2022.107746.

Almendro-Candel, M.B. *et al.* (2018) 'Physical Properties of Soils Affected by the Use of Agricultural Waste', in *Agricultural Waste and Residues*. IntechOpen. Available at: https://doi.org/10.5772/INTECHOPEN.77993.

Amatulli, G. *et al.* (2018) 'A suite of global, cross-scale topographic variables for environmental and biodiversity modeling', *Scientific Data 2018 5:1*, 5(1), pp. 1–15. Available at: https://doi.org/10.1038/sdata.2018.40.

Amsili, J.P., van Es, H.M. and Schindelbeck, R.R. (2021) 'Cropping system and soil texture shape soil health outcomes and scoring functions', *Soil Security*, 4. Available at: https://doi.org/10.1016/j.soisec.2021.100012.

Anantha, K.H. *et al.* (2021) 'Impact of best management practices on sustainable crop production and climate resilience in smallholder farming systems of South Asia', *Agricultural Systems*, 194, p. 103276. Available at: https://doi.org/10.1016/J.AGSY.2021.103276.

Arshad, M. (2021) 'Fortnightly dynamics and relationship of growth, dry matter partition and productivity of maize based sole and intercropping systems at different elevations', *European Journal of Agronomy*, 130(April 2020), p. 126377. Available at: https://doi.org/10.1016/j.eja.2021.126377.

Awe, G.O., Reichert, J.M. and Fontanela, E. (2020) 'Sugarcane production in the subtropics: Seasonal changes in soil properties and crop yield in no-tillage, inverting and minimum tillage', *Soil and Tillage Research*, 196, p. 104447. Available at: https://doi.org/10.1016/J.STILL.2019.104447.

Bai, Z. *et al.* (2013) 'The critical soil P levels for crop yield, soil fertility and environmental safety in different soil types', *Plant and Soil*, 372(1–2), pp. 27–37. Available at: https://doi.org/10.1007/S11104-013-1696-Y/METRICS.

Baier, C. *et al.* (2023) 'Effects of agroforestry on grain yield of maize (Zea mays L.)—A global meta-analysis', *Frontiers in Sustainable Food Systems*, 7, p. 1167686. Available at: https://doi.org/10.3389/FSUFS.2023.1167686/BIBTEX.

Baker, N.T. and Capel, P.D. (2011) *Environmental Factors That Influence the Location of Crop Agriculture in the Conterminous United States Scientific*, *pubs.usgs.gov*. Available at: http://photogallery.nrcs.usda.gov/. (Accessed: 17 March 2024).

Burdett, H. and Wellen, C. (2022) 'Statistical and machine learning methods for crop yield prediction in the context of precision agriculture', *Precision Agriculture*, 23(5), pp. 1553–1574. Available at: https://doi.org/10.1007/S11119-022-09897-0/METRICS.

Campillo, C., Fortes, R. and Del Henar Prieto, M. (2012) 'Solar Radiation Effect on Crop Production', in E.B. Babatunde (ed.) *Solar Radiation*. InTech. Available at: www.intechopen.com (Accessed: 15 March 2024).

Chen, G. and Weil, R.R. (2011) 'Root growth and yield of maize as affected by soil compaction and cover crops', *Soil and Tillage Research*, 117, pp. 17–27. Available at: https://doi.org/10.1016/J.STILL.2011.08.001.

Choden, T. and Ghaley, B.B. (2021) 'A portfolio of effective water and soil conservation practices for arable production systems in Europe and North Africa', *Sustainability (Switzerland)*, 13(5), pp. 1–18. Available at: https://doi.org/10.3390/su13052726.

Conrad, O. *et al.* (2015) 'System for Automated Geoscientific Analyses (SAGA) v. 2.1.4', *Geoscientific Model Development*, 8(7), pp. 1991–2007. Available at: https://doi.org/10.5194/GMD-8-1991-2015.

Corley, H.W. (2017) 'Normative Utility Models for Pareto Scalar Equilibria in n-Person, Semi-Cooperative Games in Strategic Form', *Theoretical Economics Letters*, 07(06), pp. 1667–1686. Available at: https://doi.org/10.4236/TEL.2017.76113.

Croitoru, A.E. *et al.* (2020) 'Refining the spatial scale for maize crop agro-climatological suitability conditions in a region with complex topography towards a smart and sustainable agriculture. case study: Central Romania (Cluj county)', *Sustainability (Switzerland)*, 12(7), pp. 1–17. Available at: https://doi.org/10.3390/su12072783.

Czarnecka, D. *et al.* (2022) 'The Occurrence of Fungal Diseases in Maize in Organic Farming Versus an Integrated Management System', *Agronomy 2022, Vol. 12, Page 558*, 12(3), p. 558. Available at: https://doi.org/10.3390/AGRONOMY12030558.

Department of Economic and Social Affairs (2017) *World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100 | United Nations*, *www.un.org*. Available at: https://www.un.org/en/desa/world-population-projected-reach-98-billion-2050-and-112-billion-2100 (Accessed: 12 January 2024).

Desloires, J., Ienco, D. and Botrel, A. (2023) 'Out-of-year corn yield prediction at field-scale using Sentinel-2 satellite imagery and machine learning methods', *Computers and Electronics in Agriculture*, 209, p. 107807. Available at: https://doi.org/10.1016/J.COMPAG.2023.107807.

Dharumarajan, S. *et al.* (2024) 'Quantification and mapping of the carbon sequestration potential of soils via a quantile regression forest model', *Earth Science Informatics*, pp. 1–13. Available at: https://doi.org/10.1007/S12145-024-01298-3/METRICS.

Dilla, A.M. *et al.* (2019) 'Tree pruning, zone and fertiliser interactions determine maize productivity in the Faidherbia albida (Delile) A. Chev parkland agroforestry system of Ethiopia', *Agroforestry Systems*, 93(5), pp. 1897–1907. Available at: https://doi.org/10.1007/S10457-018-0304-9/METRICS.

Dollinger, J. and Jose, S. (2018) 'Agroforestry for soil health', *Agroforestry Systems*, 92(2), pp. 213–219. Available at: https://doi.org/10.1007/S10457-018-0223-9/METRICS.

Dong, X. *et al.* (2021) 'Responses of maize with different growth periods to heat stress around flowering and early grain filling', *Agricultural and Forest Meteorology*, 303, p. 108378. Available at: https://doi.org/10.1016/J.AGRFORMET.2021.108378.

Durrer, A. *et al.* (2021) 'Organic farming practices change the soil bacteria community, improving soil quality and maize crop yields', *PeerJ*, 9, p. e11985. Available at: https://doi.org/10.7717/PEERJ.11985/SUPP-5.

El, N. and Scialabba, H. (2007) 'ORGANIC AGRICULTURE AND FOOD SECURITY IN AFRICA', *Food and Agriculture Organization of the United Nations* [Preprint].

Emmalea, G.E. (2020) *2019-2020 Corn Growing Degree Days | Weekly Crop Update*, *sites.udel.edu*. Available at: https://sites.udel.edu/weeklycropupdate/?p=15918 (Accessed: 16 March 2024).

European Environment Agency (2023) *Soil - Key Facts*, *www.eea.europa.eu*. Available at: https://www.eea.europa.eu/en/topics/in-depth/soil?activeTab=fa515f0c-9ab0-493c-b4cd-58a32dfaae0a (Accessed: 12 January 2024).

European Union (2019) *WAD | World Atlas of Desertification*, *wad.jrc.ec.europa.eu*. Available at: https://wad.jrc.ec.europa.eu/patternsaridity (Accessed: 18 January 2024).

FAO (2006) *GUIDELINES FOR SOIL DESCRIPTION*. 4th edn. Rome.

FAO (2017) 'SOIL ORGANIC CARBON - The hidden potential', *www.fao.org*.

Fatima, Z. *et al.* (2020) 'The fingerprints of climate warming on cereal crops phenology and adaptation options', *Scientific Reports*, 10(1). Available at: https://doi.org/10.1038/s41598-020-74740-3.

Félix, G.F. *et al.* (2018) 'Enhancing agroecosystem productivity with woody perennials in semi-arid West Africa. A meta-analysis', *Agronomy for Sustainable Development*. Springer-Verlag France. Available at: https://doi.org/10.1007/s13593-018-0533-3.

Fulton, A., Advisor, F. and Counties, S. (2010) *Primary Plant Nutrients: Nitrogen, Phosphorus, and Potassium*.

Galani, Y.J.H. *et al.* (2022) 'Conservation Agriculture Affects Grain and Nutrient Yields of Maize (Zea Mays L.) and Can Impact Food and Nutrition Security in Sub-Saharan Africa', *Frontiers in Nutrition*, 8, p. 804663. Available at: https://doi.org/10.3389/FNUT.2021.804663/BIBTEX.

Gayathri, P.. *et al.* (2024) 'The role of agroforestry in mitigating climate changes in arid regions of Rajasthan', *Agricultural and Biological Research*, 40(1), pp. 861–864. Available at: https://doi.org/10.35248/0970-1907.24.40.861-864.

Guo, J. *et al.* (2020) 'Afforestation and agroforestry enhance soil nutrient status and carbon sequestration capacity in eastern China', *Land Degradation & Development*, 31(3), pp. 392–403. Available at: https://doi.org/10.1002/LDR.3457.

Hashimi, R., Kaneko, N. and Komatsuzaki, M. (2023) 'Impact of no-tillage on soil quality and crop yield in Asia: A meta-analysis', *Land Degradation & Development*, 34(4), pp. 1004–1018. Available at: https://doi.org/10.1002/LDR.4512.

Henry, J.A. (2005) 'Humid climates', *Encyclopedia of Earth Sciences Series*, pp. 412–413. Available at: https://doi.org/10.1007/1-4020-3266-8_101/COVER.

Hine, R. and Pretty, J. (2006) 'Promoting Production and Trading Opportunities for Organic Agricultural Products in East Africa Capacity Building Study 3: Organic Agriculture and Food Security in East Africa'.

Hong, Y. *et al.* (2017) 'Intercropping and agroforestry in China – Current state and trends', *Agriculture, Ecosystems & Environment*, 244, pp. 52–61. Available at: https://doi.org/10.1016/J.AGEE.2017.04.019.

International Food Policy Research Institute (2020) 'Global Spatially-Disaggregated Crop Production Statistics Data for 2020 Version 1.0', *Harvard Dataverse* [Preprint]. Available at: https://doi.org/10.7910/DVN/SWPENT.

Ivezić, V., Yu, Y. and Werf, W. van der (2021) 'Crop Yields in European Agroforestry Systems: A Meta-Analysis', *Frontiers in Sustainable Food Systems*, 5, p. 606631. Available at: https://doi.org/10.3389/FSUFS.2021.606631/BIBTEX.

Jacobs, A.A. *et al.* (2022) 'Cover crops and no-tillage reduce crop production costs and soil loss, compensating for lack of short-term soil quality improvement in a maize and soybean production system', *Soil and Tillage Research*, 218, p. 105310. Available at: https://doi.org/10.1016/J.STILL.2021.105310.

Jat, M.L. *et al.* (2013) 'Double no-till and permanent raised beds in maize–wheat rotation of north-western Indo-Gangetic plains of India: Effects on crop yields, water productivity, profitability and soil physical properties', *Field Crops Research*, 149, pp. 291–299. Available at: https://doi.org/10.1016/J.FCR.2013.04.024.

Jeong, J.H. *et al.* (2016) 'Random Forests for Global and Regional Crop Yield Predictions', *PLOS ONE*, 11(6), p. e0156571. Available at: https://doi.org/10.1371/JOURNAL.PONE.0156571.

Jian, J., Du, X. and Stewart, R.D. (2020) 'A database for global soil health assessment', *Scientific Data*, 7(1), pp. 3–10. Available at: https://doi.org/10.1038/s41597-020-0356-3.

Kuhn, M., Kuhn and Max (2015) 'caret: Classification and Regression Training', *ascl*, p. ascl:1505.003. Available at: https://ui.adsabs.harvard.edu/abs/2015ascl.soft05003K/abstract (Accessed: 10 June 2024).

Kumar, Suresh, Kumar, Santosh and Mohapatra, T. (2021) 'Interaction Between Macro- and Micro-Nutrients in Plants', *Frontiers in Plant Science*, 12, p. 665583. Available at: https://doi.org/10.3389/FPLS.2021.665583/BIBTEX.

de la Cruz, V.Y. V. *et al.* (2023) 'Yield gap between organic and conventional farming systems across climate types and sub-types: A meta-analysis', *Agricultural Systems*, 211. Available at: https://doi.org/10.1016/j.agsy.2023.103732.

Lal, R. (2004) 'Soil carbon sequestration to mitigate climate change', *Geoderma*, 123(1–2), pp. 1–22. Available at: https://doi.org/10.1016/J.GEODERMA.2004.01.032.

Liliane, T.N. *et al.* (2020) 'Factors Affecting Yield of Crops', *Agronomy - Climate Change and Food Security* [Preprint]. Available at: https://doi.org/10.5772/INTECHOPEN.90672.

Lotter, D. (2015) 'Facing food insecurity in Africa: Why, after 30 years of work in organic agriculture, I am promoting the use of synthetic fertilizers and herbicides in small-scale staple crop production', *Agriculture and Human Values*, 32(1), pp. 111–118. Available at: https://doi.org/10.1007/S10460-014-9547-X/METRICS.

Ma, B. *et al.* (2019) 'Effects of crop-slope interaction on slope runoff and erosion in the Loess Plateau', *Acta Agriculturae Scandinavica Section B: Soil and Plant Science*, 69(1), pp. 12–25. Available at: https://doi.org/10.1080/09064710.2018.1488988.

Marques Ramos, A.P. *et al.* (2020) 'A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices', *Computers and Electronics in Agriculture*, 178, p. 105791. Available at: https://doi.org/10.1016/J.COMPAG.2020.105791.

McDowell, R.W. *et al.* (2023) 'A Global Database of Soil Plant Available Phosphorus', *Scientific Data 2023 10:1*, 10(1), pp. 1–10. Available at: https://doi.org/10.1038/s41597-023-02022-4.

Meyer, H. *et al.* (2019) 'Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction', *Ecological Modelling*, 411(September), p. 108815. Available at: https://doi.org/10.1016/j.ecolmodel.2019.108815.

Minnesota Pollution Control Agency (2023) *Soil health indicator sheet - Soil compaction (bulk density) - Minnesota Stormwater Manual*, *stormwater.pca.state.mn.us*. Available at: https://stormwater.pca.state.mn.us/index.php?title=Soil_health_indicator_sheet_-_Soil_compaction_(bulk_density) (Accessed: 17 January 2024).

Minnesota Pollution Control Agency (2024) *General relationship of soil bulk density to root growth based on soil texture - Minnesota Stormwater Manual*, *stormwater.pca.state.mn.us*. Available at: https://stormwater.pca.state.mn.us/index.php/General_relationship_of_soil_bulk_density_to_root_growth_based_on_soil_texture (Accessed: 11 March 2024).

Mohammad, N. *et al.* (2023) 'Yield Forecasting Model for Maize Using Satellite Multispectral Imagery Driven Vegetation Indices', *Qeios* [Preprint]. Available at: https://doi.org/10.32388/COEBSC.

Morales, A. and Villalobos, F.J. (2023) 'Using machine learning for crop yield prediction in the past or the future', *Frontiers in Plant Science*, 14, p. 1128388. Available at: https://doi.org/10.3389/FPLS.2023.1128388/BIBTEX.

Mrabet, R. (2011) 'No-Tillage Agriculture in West Asia and North Africa', *Rainfed Farming Systems*, pp. 1015–1042. Available at: https://doi.org/10.1007/978-1-4020-9132-2_40.

MSU Extension Service (2005) 'Basic Soil Properties', in *Soil and Water*.

Neumann, K. *et al.* (2010) 'The yield gap of global grain production: A spatial analysis', *Agricultural Systems*, 103(5), pp. 316–326. Available at: https://doi.org/10.1016/J.AGSY.2010.02.004.

Niggli, U. (2015) 'Sustainability of organic food production: challenges and innovations', *Proceedings of the Nutrition Society*, 74(1), pp. 83–88. Available at: https://doi.org/10.1017/S0029665114001438.

NSW Department of Primary Industries (2009) *Procrop Maize growth & development*. Edited by Jan Edwards.
Nunes, M.R. *et al.* (2018) 'No-till and cropping system diversification improve soil health and crop yield', *Geoderma*, 328, pp. 30–43. Available at: https://doi.org/10.1016/J.GEODERMA.2018.04.031.

Oldfield, E.E., Bradford, M.A. and Wood, S.A. (2019) 'Global meta-analysis of the relationship between soil organic matter and crop yields', *SOIL*, 5(1), pp. 15–32. Available at: https://doi.org/10.5194/SOIL-5-15-2019.

Otegui, M.E. and Slafer, G. (2000) *Physiological Bases for Maize Improvement*.
Pittelkow, C.M. *et al.* (2015) 'Productivity limits and potentials of the principles of conservation agriculture', *Nature*, 517(7534), pp. 365–368. Available at: https://doi.org/10.1038/nature13809.

Poggio, L. *et al.* (2021) 'SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty', *SOIL*, 7(1), pp. 217–240. Available at: https://doi.org/10.5194/SOIL-7-217-2021.

Pradhan, A., Idol, T. and Roul, P.K. (2016) 'Conservation agriculture practices in rainfed uplands of india improve maize-based system productivity and profitability', *Frontiers in Plant Science*, 7(2016JULY), p. 183534. Available at: https://doi.org/10.3389/FPLS.2016.01008/BIBTEX.

Qiao, X. *et al.* (2020) 'Apricot-based agroforestry system in Southern Xinjiang Province of China: influence on yield and quality of intercropping wheat', *Agroforestry Systems*, 94(2), pp. 477–485. Available at: https://doi.org/10.1007/S10457-019-00412-5/TABLES/4.

Ritchie, H. and Roser, M. (2019) *Land Use - Our World in Data*, *OurWorldInData.org*. Available at: https://ourworldindata.org/land-use (Accessed: 12 January 2024).

Sayago, S. and Bocco, M. (2018) 'Crop yield estimation using satellite images: Comparison of linear and non-linear models', *AgriScientia*, 35(1), pp. 1–9. Available at: https://doi.org/10.31047/1668.298X.V1.N35.20447.

Sirisuntornlak, N. *et al.* (2021) 'Interactive Effects of Silicon and Soil pH on Growth, Yield and Nutrient Uptake of Maize', *Silicon*, 13(2), pp. 289–299. Available at: https://doi.org/10.1007/S12633-020-00427-Z/METRICS.

Su, Y. *et al.* (2022) 'Performances of Machine Learning Algorithms in Predicting the Productivity of Conservation Agriculture at a Global Scale', *Frontiers in Environmental Science*, 10, p. 812648. Available at: https://doi.org/10.3389/FENVS.2022.812648/BIBTEX.

Su, Y., Gabrielle, B. and Makowski, D. (2021a) 'A global dataset for crop production under conventional tillage and no tillage systems', *Scientific Data 2021 8:1*, 8(1), pp. 1–17. Available at: https://doi.org/10.1038/s41597-021-00817-x.

Su, Y., Gabrielle, B. and Makowski, D. (2021b) 'A global dataset for crop production under conventional tillage and no tillage systems', *Scientific Data*, 8(1), pp. 1–17. Available at: https://doi.org/10.1038/s41597-021-00817-x.

Su, Y., Gabrielle, B. and Makowski, D. (2021c) 'The impact of climate change on the productivity of conservation agriculture', *Nature Climate Change*, 11(7), pp. 628–633. Available at: https://doi.org/10.1038/s41558-021-01075-w.

Sun, J. *et al.* (2020) 'Multilevel Deep Learning Network for County-Level Corn Yield Estimation in the U.S. Corn Belt', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, pp. 5048–5060. Available at: https://doi.org/10.1109/JSTARS.2020.3019046.

Szatten, D. and Więcław, M. (2021) 'Solar Climate Features Taking into Account the Morphometric Conditions of the Area and the Possibility of Using Them in Heliotherapy on the Example of the Cieplice and Kołobrzeg Health Resorts (Poland)', *Atmosphere 2021, Vol. 12, Page 383*, 12(3), p. 383. Available at: https://doi.org/10.3390/ATMOS12030383.

The State of Queesland (2024) *Soil pH | Environment, land and water | Queensland Government*, *www.qld.gov.au*. Available at: https://www.qld.gov.au/environment/land/management/soil/soil-properties/ph-levels (Accessed: 10 February 2024).

Triplett, G.B. and Dick, W.A. (2008) 'No-Tillage Crop Production: A Revolution in Agriculture!', *Agronomy Journal*, 100(3 SUPPL.), p. S-153. Available at: https://doi.org/10.2134/AGRONJ2007.0005C.

Vâtcă, S.D. *et al.* (2021) 'Agrometeorological Requirements of Maize Crop Phenology for Sustainable Cropping—A Historical Review for Romania', *Sustainability 2021, Vol. 13, Page 7719*, 13(14), p. 7719. Available at: https://doi.org/10.3390/SU13147719.

Vaysse, K. and Lagacherie, P. (2017) 'Using quantile regression forest to estimate uncertainty of digital soil mapping products', *Geoderma*, 291, pp. 55–64. Available at: https://doi.org/10.1016/J.GEODERMA.2016.12.017.

Vepraskas, M.J. (1988) 'Bulk Density Values Diagnostic of Restricted Root Growth in Coarse-textured Soils', *Soil Science Society of America Journal*, 52(4), pp. 1117–1121. Available at: https://doi.org/10.2136/SSSAJ1988.03615995005200040042X.

Wan, L ; *et al.* (2024) 'Effect of Agroforestry Systems on Soil NPK and C Improvements in Karst Graben Basin of Southwest China', *Agronomy 2024, Vol. 14, Page 1179*, 14(6), p. 1179. Available at: https://doi.org/10.3390/AGRONOMY14061179.

Xu, H. *et al.* (2021) 'Effects of biochar application on crop productivity, soil carbon sequestration, and global warming potential controlled by biochar C:N ratio and soil pH: A global meta-analysis', *Soil and Tillage Research*, 213. Available at: https://doi.org/10.1016/J.STILL.2021.105125.

Yang, T. *et al.* (2023) 'Understanding the effects of tree-crop intercropping systems on crop production in China by combining field experiments with a meta-analysis', *Agricultural Systems*, 210, p. 103705. Available at: https://doi.org/10.1016/J.AGSY.2023.103705.

Yang, Y. *et al.* (2019) 'Improving maize grain yield by matching maize growth and solar radiation', *Scientific Reports*, 9(1), pp. 1–11. Available at: https://doi.org/10.1038/s41598-019-40081-z.

Zhao, X. *et al.* (2017) 'Crop yields under no-till farming in China: A meta-analysis', *European Journal of Agronomy*, 84, pp. 67–75. Available at: https://doi.org/10.1016/J.EJA.2016.11.009.

Zomer, R.J., Xu, J. and Trabucco, A. (2022) 'Version 3 of the Global Aridity Index and Potential Evapotranspiration Database', *Scientific Data 2022 9:1*, 9(1), pp. 1–15. Available at: https://doi.org/10.1038/s41597-022-01493-1.

Appendix

Appendix 1

Table 7: Sources table.

Source	***Number of Observations***
Felix at al. (2018)	2
Jian et al. (2022)	584
Pitteölow et al. (2014)	1536
Su et al. (2021)	834

Appendix 2

Table 8: Statistical values table. The table reports the average and the standard deviation of the ES divided for each zone.

	Zone	*NT*		*AF*		*OF*	
		Average	*StDv*	*Average*	*StDv*	*Average*	*StDv*
BD	Favour BD	-0.067	0.154	-0.088	0.457	-	-
	Transition BD	-0.085	0.328	0.371	0.555	-0.199	0.331
	Restrict BD	-0.046	0.280	0.156	0.456	-0.139	0.119
Texture	Clay	-0.092	0.298	0.057	0.463	-0.030	0.134
	Silty	-0.059	0.295	0.206	0.434	-0.199	0.331
	Sandy	0.007	0.327	0.421	0.565	-0.194	0.076
SOC	Low SOC	-0.029	0.295	0.328	0.547	-0.050	0.087
	Moderate SOC	-0.080	0.276	0.258	0.551	-0.216	0.301
	High SOC	-0.132	0.394	0.379	0.717	-0.167	0.430
pH	Acidic	-0.056	0.295	0.324	0.548	-0.245	0.337
	Neutral	-0.181	0.350	0.321	0.469	-0.058	0.148
	Alkaline	-0.076	0.102	-	-	-	-
Olsen-P	Low P	-0.069	0.341	0.086	0.595	-0.032	0.145
	Moderate P	-0.046	0.251	0.427	0.429	-0.231	0.384
	High P	-0.073	0.287	0.499	0.488	-0.213	0.233
AI	Arid	0.068	0.188	-	-	-	-
	Semiarid	-0.032	0.293	0.404	0.695	-	-
	Sub-Humid	-0.088	0.365	0.334	0.442	-0.030	0.134
	Humid	-0.077	0.288	0.249	0.433	-0.199	0.310
DfI	Low DfI	-0.107	0.282	0.427	0.481	-	-
	Moderate DfI	0.041	0.350	0.408	0.600	-	-
	High DfI	-0.076	0.291	0.089	0.356	-0.188	0.303
GDD	Unsuitable	0.028	0.043	-	-	-	-
	Suitable	-0.078	0.295	-	-	-0.199	0.310
	Low Heat Stress	-0.029	0.231	-	-	-	-
	Moderate Heat Stress	-0.073	0.399	0.412	0.577	-	-
	High Heat Stress	0.005	0.292	0.083	0.356	-0.030	0.134
Slope	Flat	-0.120	0.346	-	-	-	-
	Level	-0.072	0.240	0.376	0.622	-0.139	0.119
	Gently	-0.018	0.342	0.281	0.477	-0.199	0.331
	Sloping	-0.047	0.353	0.493	0.192	-	-
	Steep	-0.057	-	-	-	-	-
Elevation	Low Elevation	-0.086	0.335	0.094	0.347	-0.237	0.357
	Moderate Elevation	-0.040	0.256	0.487	0.498	-0.103	0.155
	High Elevation	-0.074	0.304	0.364	0.617	-	-

Appendix 3

A. Division of the cropland and the observation points into the different environmental zones

✓ *Bulk density*

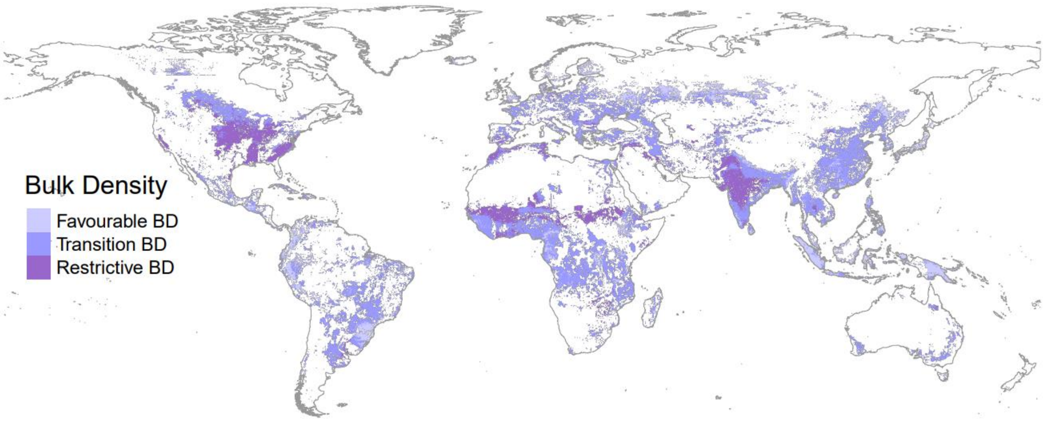


Figure 19, Division of the cropland into the BD zones.

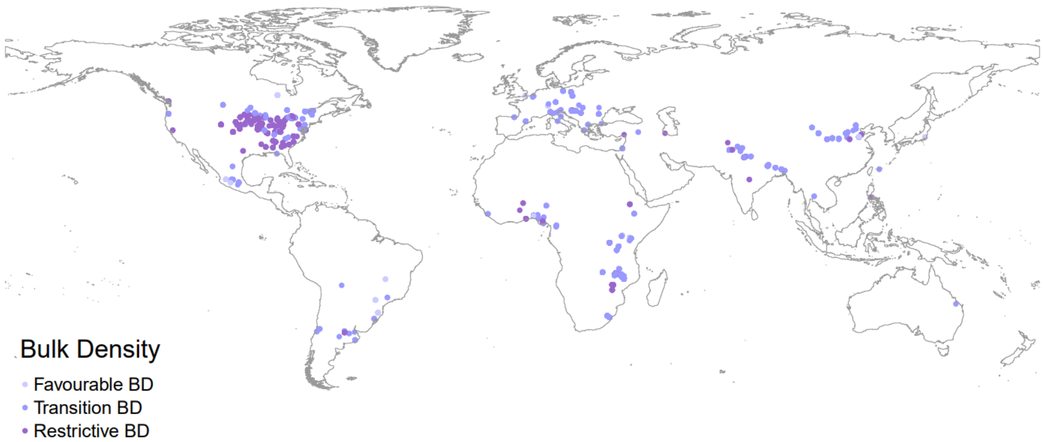


Figure 20, Observation points distribution into the different zones of BD, printed on the cropland map. In the figure, the brownish areas represent the cropland, while in grey are reported the land without cropland.

Table 9: The table reports the percentage of the cropland and number of observations belonging to one of the three BD zones.

Zones	*Range*	*Cropland percentage*	*Number of observations*
Favourable BD	Lower than 1.2 kg/dm^3	19.17%	107
Transition BD	Between 1.2 kg/dm^3 and 1.47 kg/dm^3	63.80%	1555
Restrictive BD	Higher than 1.47 kg/dm^3	17.03%	1294

✓ *Soil texture*

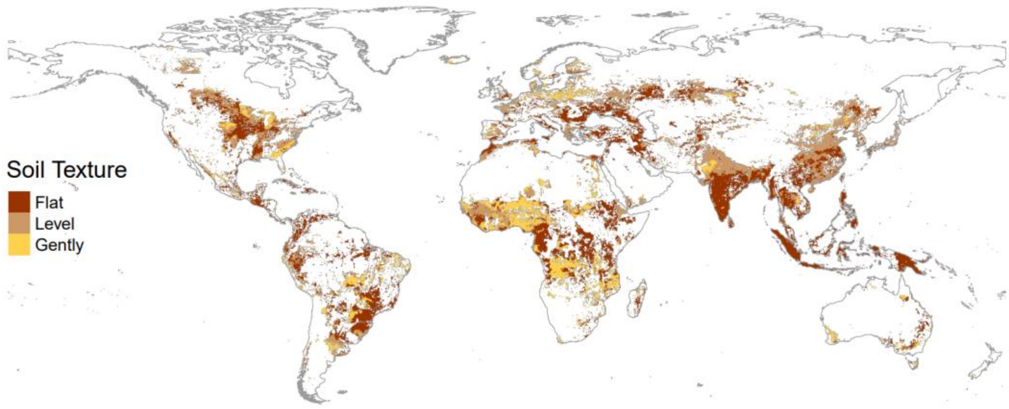


Figure 21, Division of the cropland into the soil texture zones.

Table 10: The table reports the percentage of cropland and number of observations belonging to the different soil texture zones.

Zones	***Cropland percentage***	***Number of observations***
Clay	43.97%	985
Silty	35.72%	1411
Sandy	20.31%	560

✓ *SOC*

Table 11: The table reports the percentage of the cropland and number of observations belonging to the different SOC zones.

Zones	***Range***	***Cropland percentage***	***Number of observations***
Low SOC	Lower than 5 g/kg	40.04%	1450
Moderate SOC	Between 5 g/kg and 10 g/kg	37.48%	1226
High SOC	Higher than 10 g/kg	22.48%	280

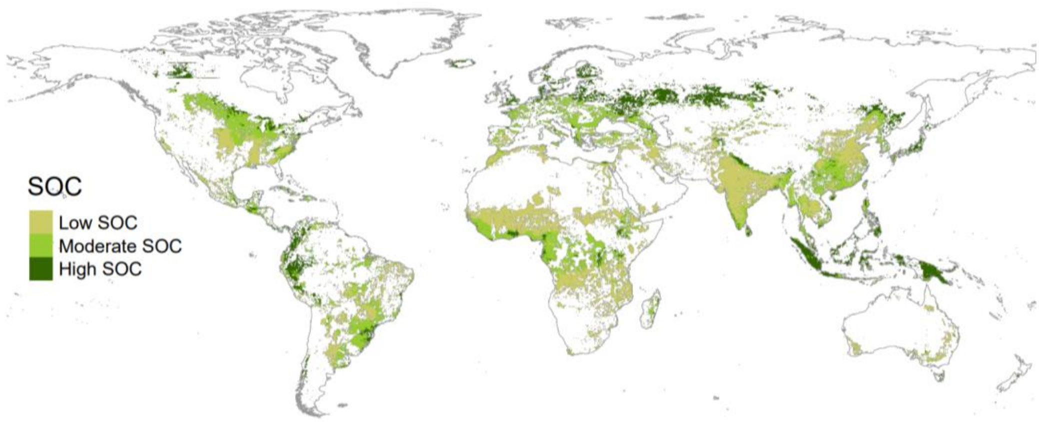


Figure 22, Division of the cropland into the SOC zones.

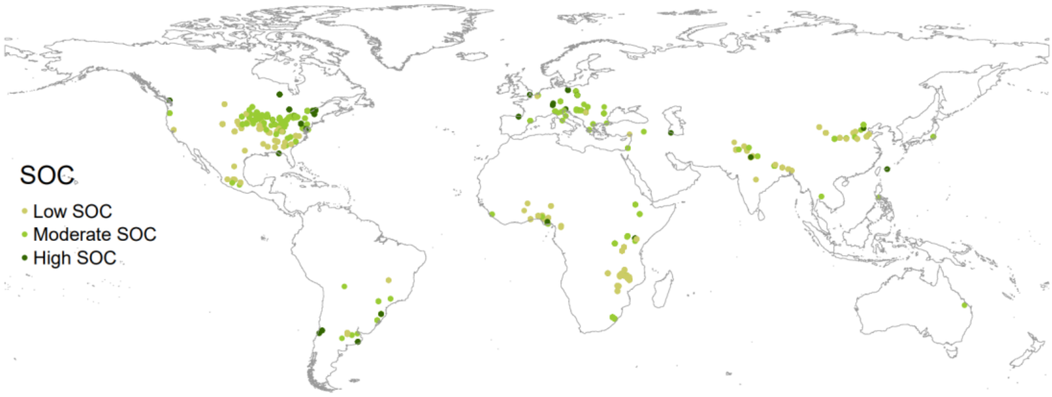


Figure 23, Observation points distribution into the different zones of SOC, printed on the cropland map. In the figure, the brownish areas represent the cropland, while in grey are reported the land without cropland.

✓ *pH*

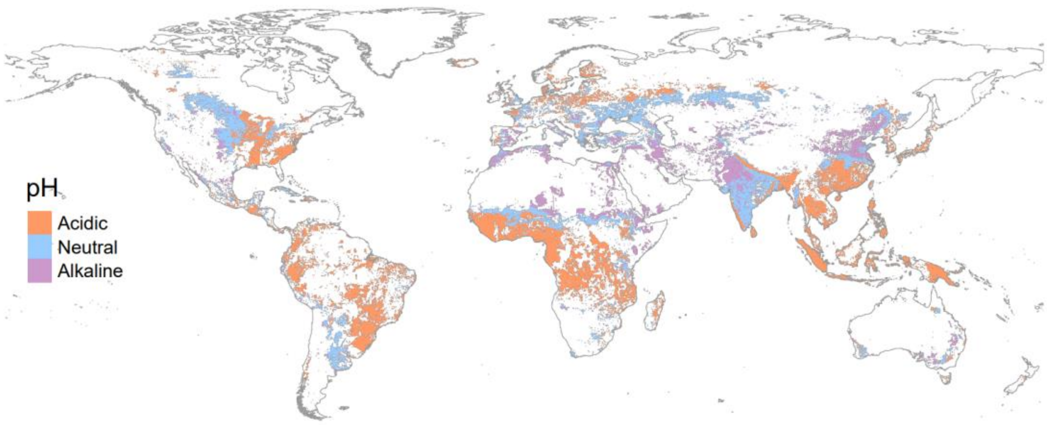


Figure 24, Division of the cropland into the different pH zones.

Table 12: The table shows the percentage of cropland and number of observations belonging to the different pH zones.

Zones	***Range***	***Cropland percentage***	***Number of observations***
Acidic	Lower than 6.3	51.56%	2747
Neutral	Between 6.3 and 7.4	30.10%	2000
Alkaline	Higher than 7.4	18.33%	9

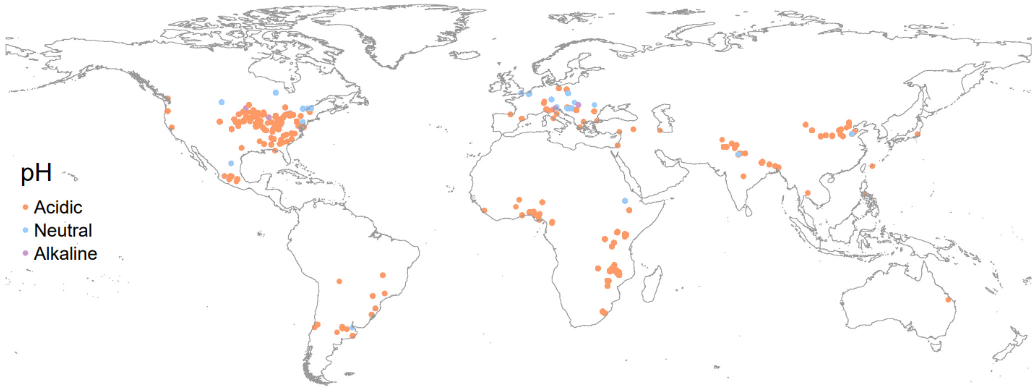


Figure 25, Observation points distribution into the different zones of pH, printed on the cropland map. In the figure, the brownish areas represent the cropland, while in grey are reported the land without cropland.

✓ ***Olsen-P***

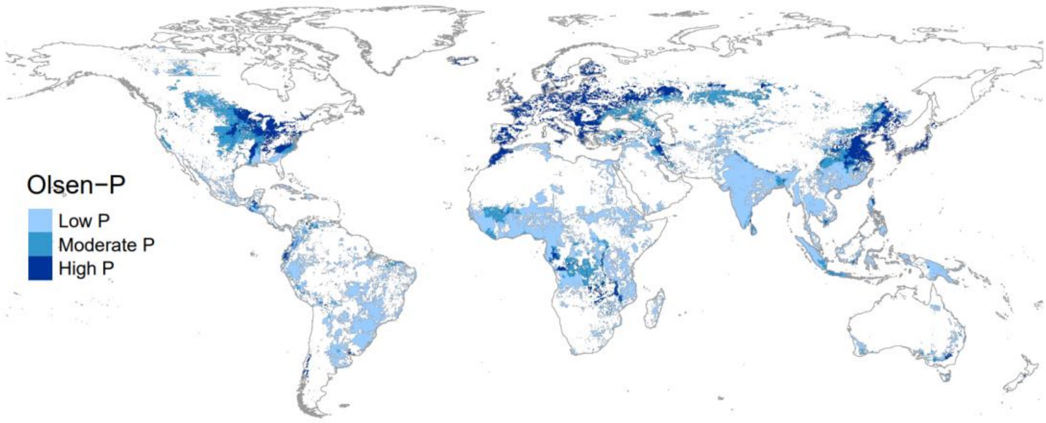


Figure 26, Cropland division into the different Olsen-P zones.

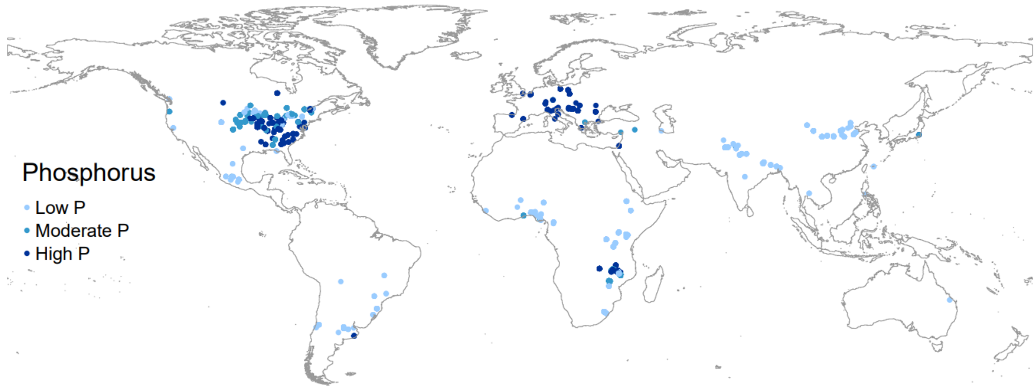


Figure 27, Observation points distribution into the different zones of Olsen-P, printed on the cropland map. In the figure, the brownish areas represent the cropland, while in grey are reported the land without cropland.

Table 13: The table shows the percentage of cropland belonging to the different Olsen-P zones.

Zones	***Range***	***Cropland percentage***	***Number of observations***
Low P	Lower than 10.9 mg/kg	58.24%	1117
Moderate P	Between 10.9 mg/kg and 21.4 mg/kg	20.14%	765
High P	Higher than 21.4 mg/kg	21.62%	1074

✓ *AI*

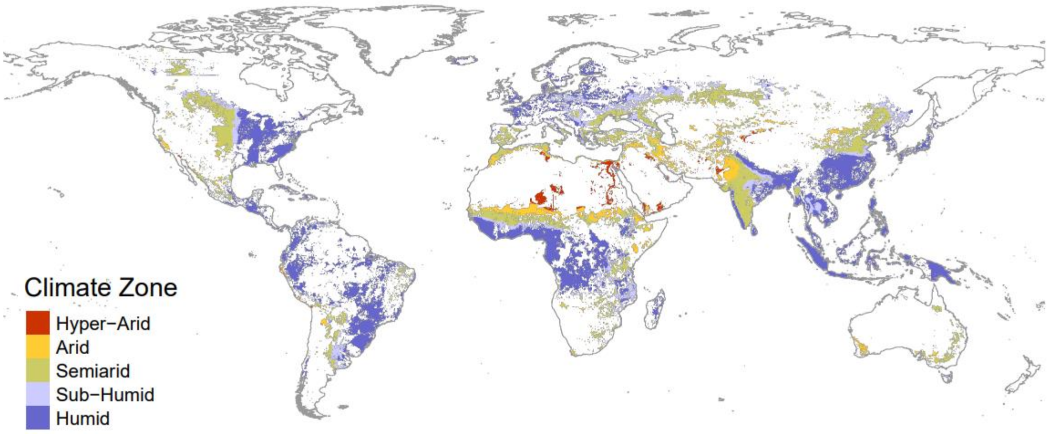


Figure 28, Cropland division into the different climate zones.

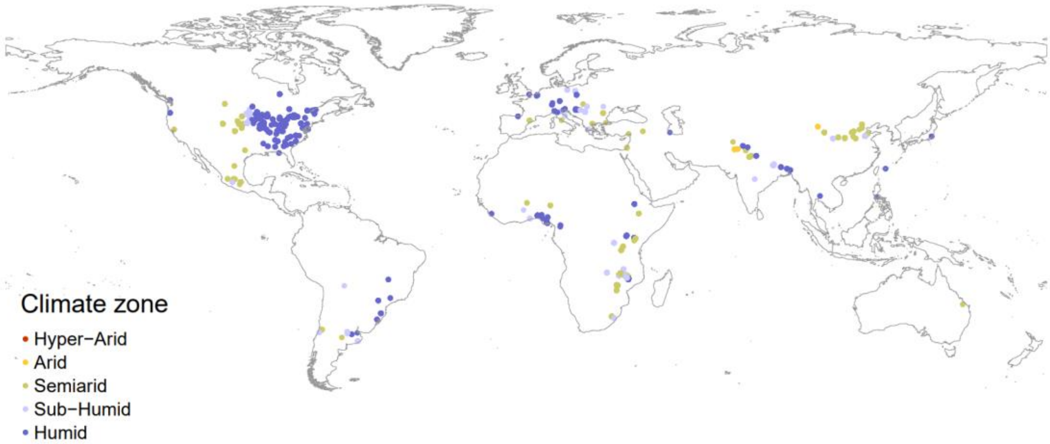


Figure 29, Observation points distribution into the different climatic zones according with the AI value, printed on the cropland map. In the figure, the brownish areas represent the cropland, while in grey are reported the land without cropland.

Table 14: The table reports the percentage of the cropland and number of observations belonging to the different climatic zones.

Zones	*Range*	*Cropland percentage*	*Number of observations*
Hyper-Arid	AI lower than 0.05	2.75%	0
Arid	AI between 0.05 and 0.20	9.18%	50
Semiarid	AI between 0.020 and 0.50	27.16%	749
Sub-Humid	AI between 0.50 and 0.65	13.16%	467
Humid	AI higher than 0.65	47.75%	1690

✓ *DfI*

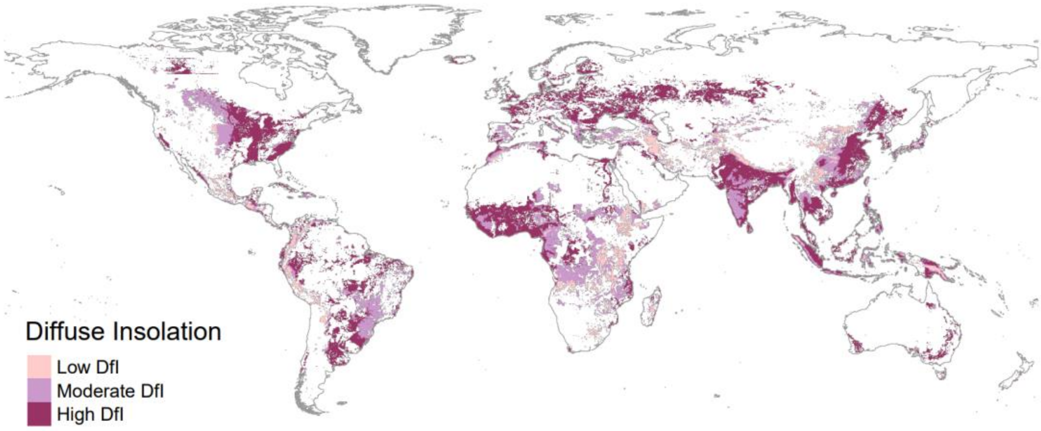


Figure 30, Cropland division into the different DfI zones.

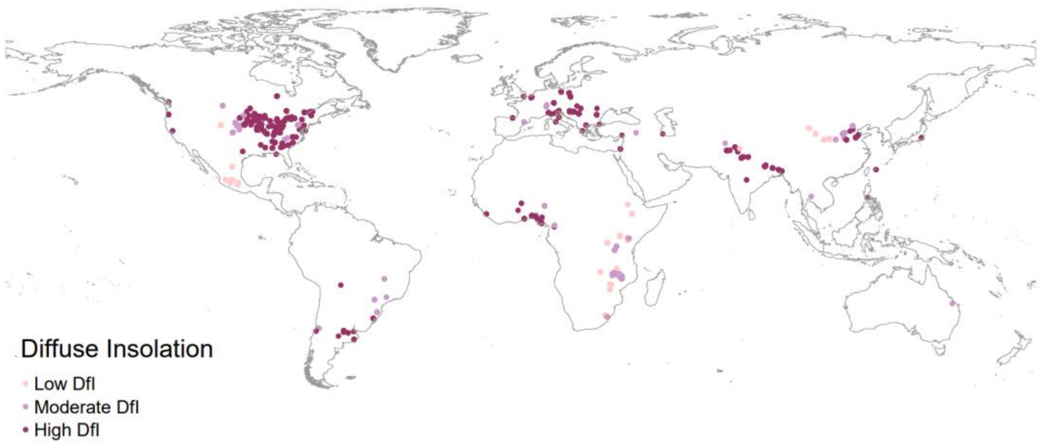


Figure 31, Observation points distribution into the different zones of DfI, printed on the cropland map. In the figure, the brownish areas represent the cropland, while in grey are reported the land without cropland.

Table 15: The table shows the percentage of cropland and the number of observations belonging to the different zones.

Zones	***Range***	***Cropland percentage***	***Number of observations***
Low DfI	Lower than 0.50 kWh/m²	12.67%	393
Moderate DfI	Between 0.50 kWh/m² and 0.55 kWh/m²	30.16%	601
High DfI	Higher than 0.55 kWh/m²	57.17%	1962

✓ <u>*GDD*</u>

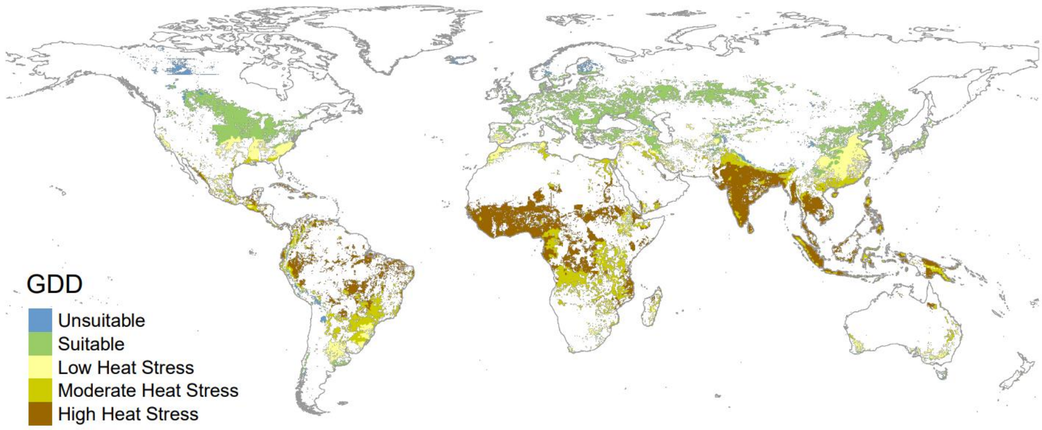


Figure 32, Cropland division into the different GDD zones.

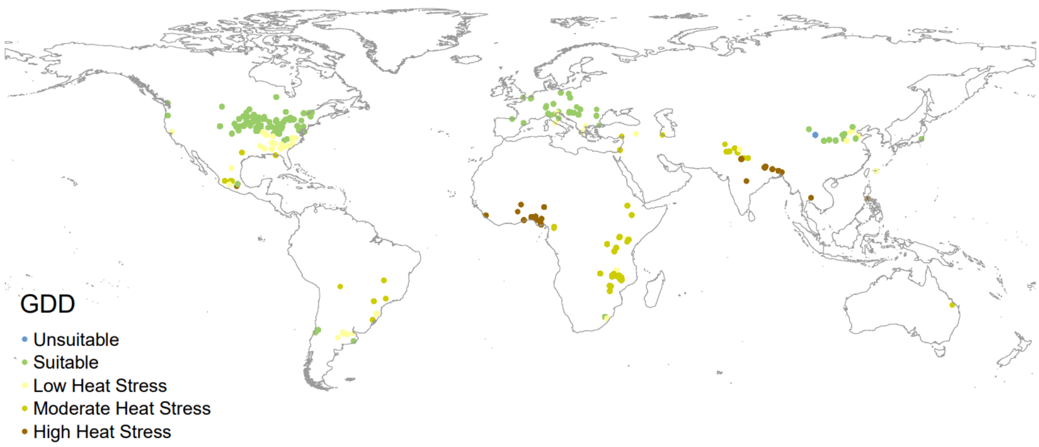


Figure 33, Observation points distribution into the different zones of GDD, printed on the cropland map. In the figure, the brownish areas represent the cropland, while in grey are reported the land without cropland.

Table 16: The table reports the percentage of the cropland and the number of observations belonging to the different GDD zones.

Zones	***Range***	***Cropland percentage***	***Number of observations***
Unsuitable	GDD lower than 800°C/y	2.48%	18
Suitable	GDD between 800°C/y and 2700°C/y	32.35%	1746
Low Heat Stress	GDD between 2700°C/y and 4000°C/y	14.06%	276
Moderate Heat Stress	GDD between 4000°C/y and 6000°C/y	19.54%	609
High Heat Stress	GDD higher than 6000°C/y	31.57%	307

✓ <u>*Elevation*</u>

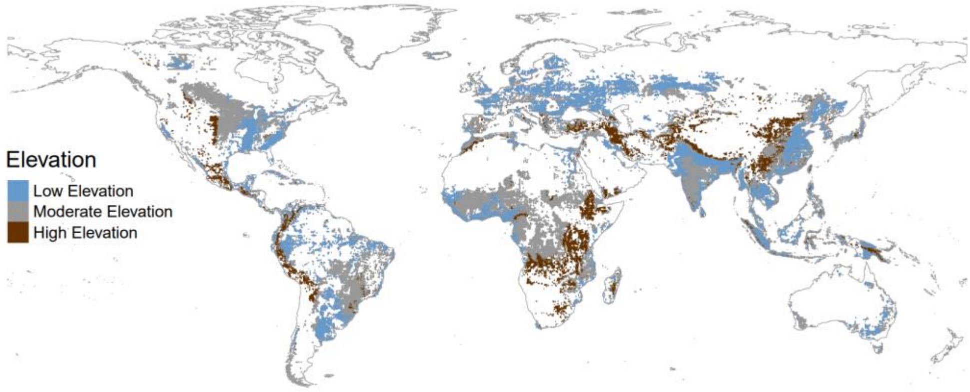


Figure 34, Cropland division into the different elevation zones.

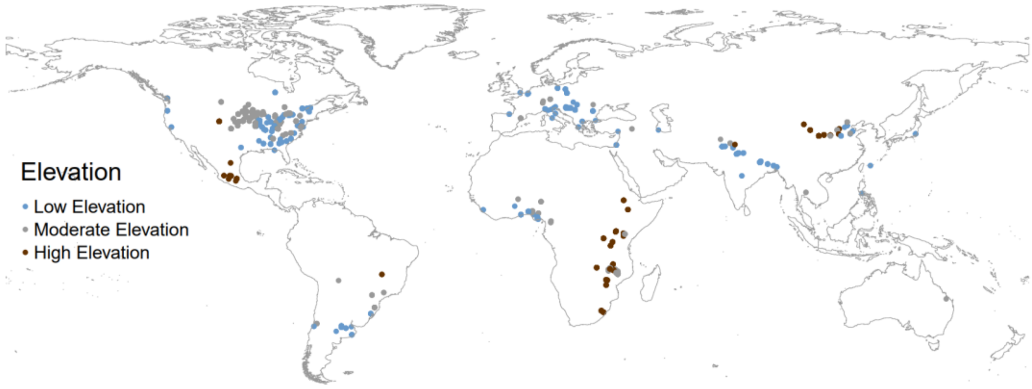


Figure 35, Observation points distribution into the different zones of the elevation, printed on the cropland map. In the figure, the brownish areas represent the cropland, while in grey are reported the land without cropland.

Table 17: The table shows the percentage of cropland and the number of observations belonging to the different zones.

Zones	***Range***	***Cropland percentage***	***Number of observations***
Low Elevation	Lower than 250 m	40.92%	1197
Moderate Elevation	Between 250 m and 1000 m	42.03%	1145
High Elevation	Higher than 1000 m	17.05%	641

✓ *Slope*

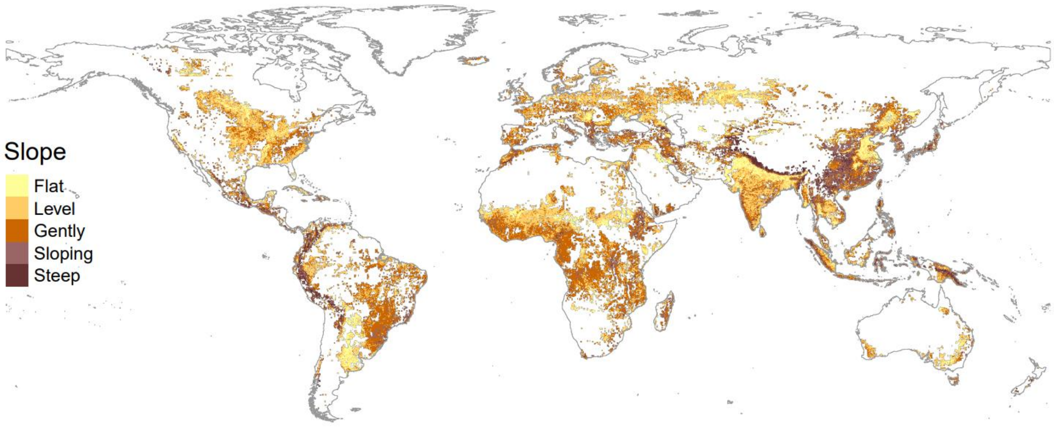


Figure 36, Cropland division into the different slope zones.

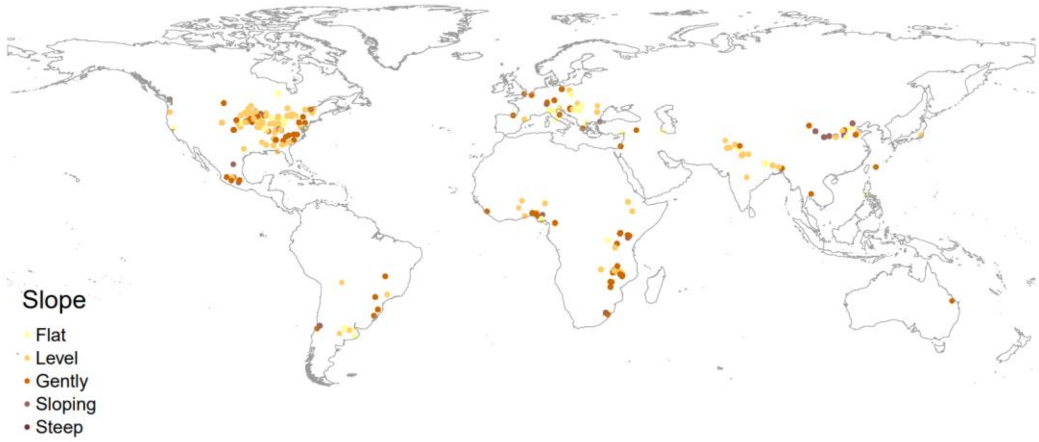


Figure 37, Observation points distribution into the different zones of the slope, printed on the cropland map. In the figure, the brownish areas represent the cropland, while in grey are reported the land without cropland.

Table 18: The table reports the percentage of the cropland and the number of observations belonging to the different Slope zones.

Zones	***Range***	***Cropland percentage***	***Number of observations***
Flat	Slope lower than 0.2%	13.03%	463
Level	Slope between 0.2% and 1%	35.32%	1398
Gently	Slope between 1% and 5%	33.45%	937
Sloping	Slope between 5% and 15%	13.64%	157
Steep	Slope higher than 15%	4.56%	1

B. Boxplot with outliers and number of observations per environmental zones divided for management

✓ *Climatic factors*

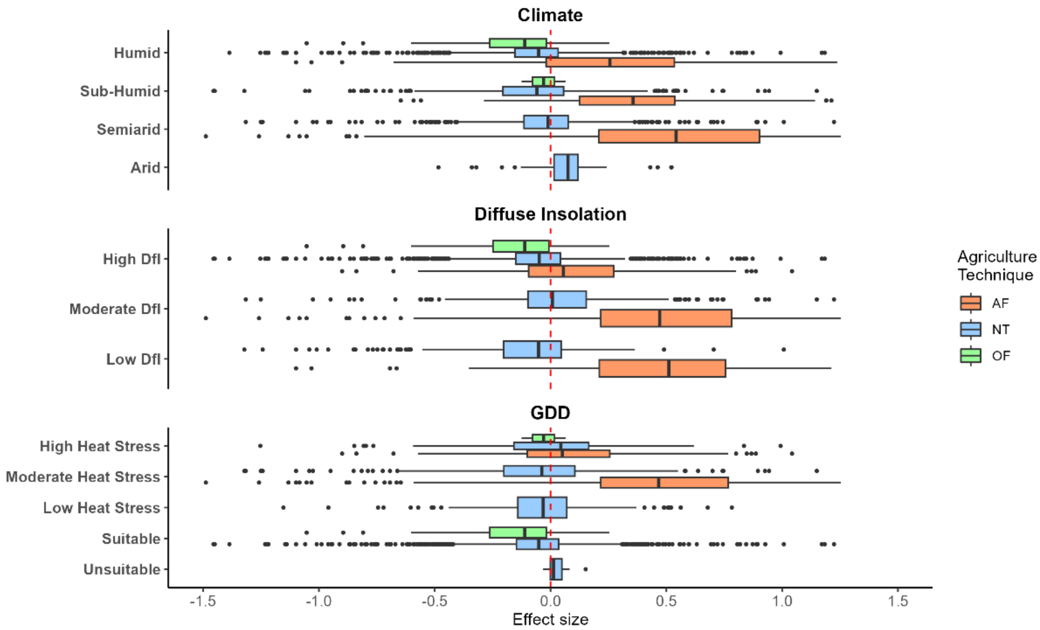


Figure 38, Boxplot of the effect size for the RPs within the zones of climatic factors. The ends of the box represent the 25th and 75th quantiles and the median value is reported in the middle of the box. The horizontal black lines report the highest and the lowest values of the class excluding the outliers which are represented by the black dots. They are also called whiskers, and their length is equal to 1.5 times the interquartile (75th quartile minus 25th quartile). The values which are not included in this range are called outliers.

Table 19: Number of observations in every climatic factor zone per management.

	Zone	*Number of observations*		
		NT	*AF*	*OF*
Climate	Hyper-Arid	0	0	0
	Arid	50	0	0
	Semiarid	573	176	0
	Sub-Humid	355	110	2
	Humid	1453	206	31
DfI	Low DfI	326	67	0
	Moderate DfI	310	291	0
	High DfI	1795	134	33
GDD	Unsuitable	18	0	0
	Suitable	1715	0	31
	Low Heat Stress	276	0	0
	Moderate Heat Stress	249	360	0
	High Heat Stress	173	132	2

✓ *Soil factors*

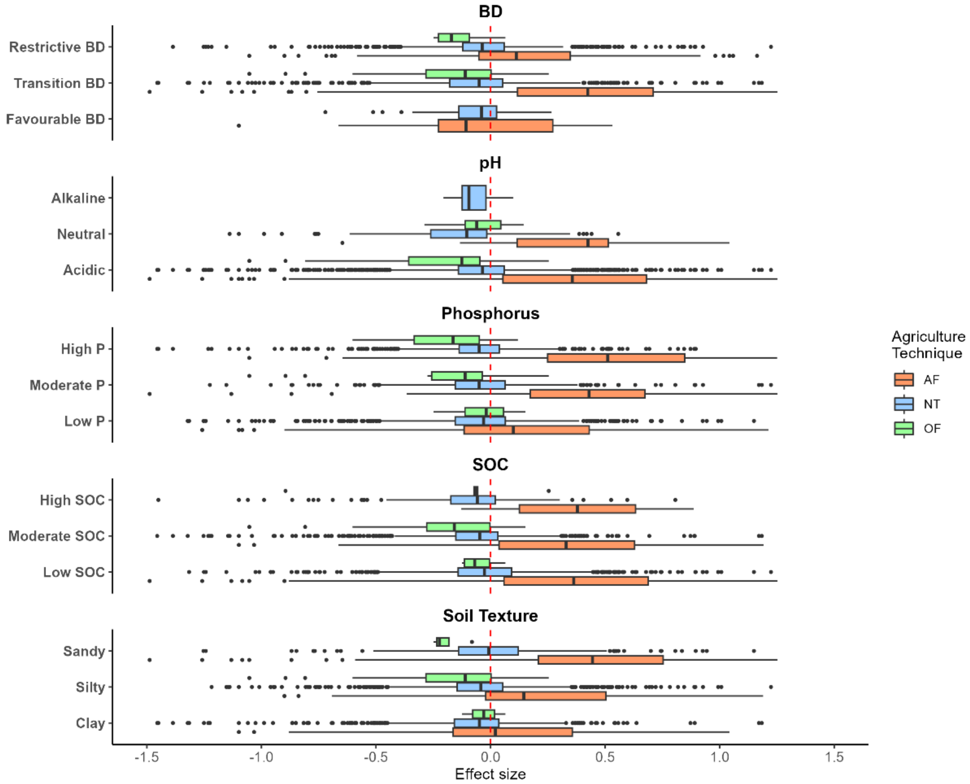


Figure 39, Boxplot of the effect size for the RPs within the zones of soil factors. The ends of the box represent the 25th and 75th quantiles and the median value is reported in the middle of the box. The horizontal black lines report the highest and the lowest values of the class excluding the outliers which are represented by the black dots. They are also called whiskers, and their length is equal to 1.5 times the interquartile (75th quartile minus 25th quartile). The values which are not included in this range are called outliers.

Table 20: Number of observations of every soil factor zone divided for the RP.

	Zone	*Number of observations*		
		NT	*AF*	*OF*
BD	Favourable BD	94	13	0
	Transition BD	1129	399	27
	Restrictive BD	1208	80	6
pH	Acidic	2242	482	23
	Neutral	180	10	10
	Alkaline	9	0	0
Phosphorus	Low P	928	183	6
	Moderate P	603	147	15
	High P	900	162	12
SOC	Low SOC	989	457	4
	Moderate SOC	1169	33	24
	High SOC	273	2	5
Soil texture	Clay	902	81	2
	Silty	1298	86	27
	Sandy	231	325	4

✓ <u>*Landform factors*</u>

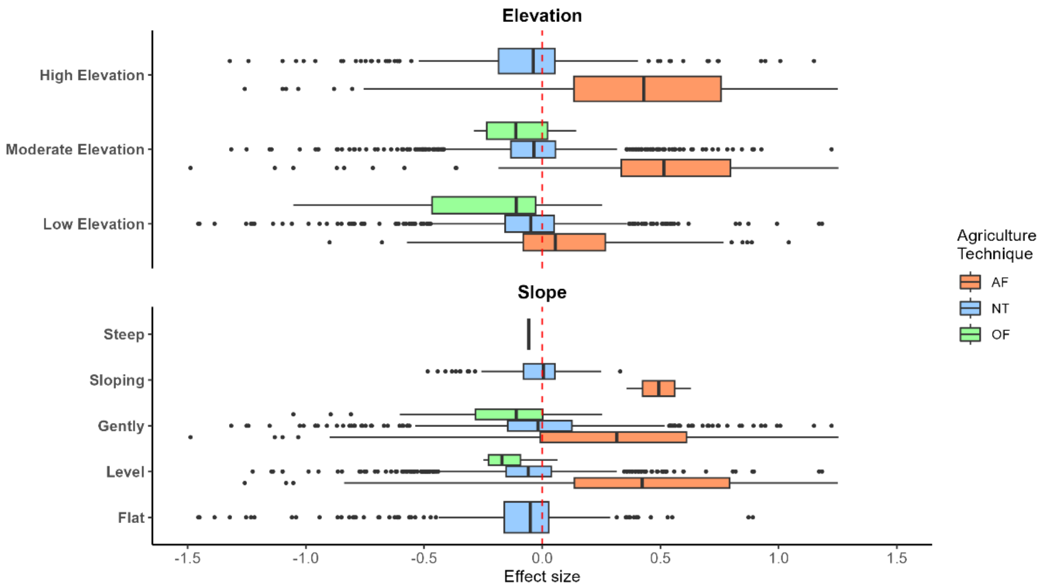


Figure 40, Boxplot of the effect size for the RPs within the zones of landform factors. The ends of the box represent the 25th and 75th quantiles and the median value is reported in the middle of the box. The horizontal black lines report the highest and the lowest values of the class excluding the outliers which are represented by the black dots. They are also called whiskers, and their length is equal to 1.5 times the interquartile (75th quartile minus 25th quartile). The values which are not included in this range are called outliers.

Table 21: Number of observations for every landform factor zone per management.

	Zone	***Number of observations***		
		NT	*AF*	*OF*
Elevation	Low elevation	1046	130	21
	Moderate elevation	1009	124	12
	High elevation	376	238	0
Slope	Flat	463	0	0
	Level	1174	218	6
	Gently	638	272	27
	Sloping	155	2	0
	Steep	1	0	0

Appendix 4

A. Correlation of the predictors

✓ *All managements*

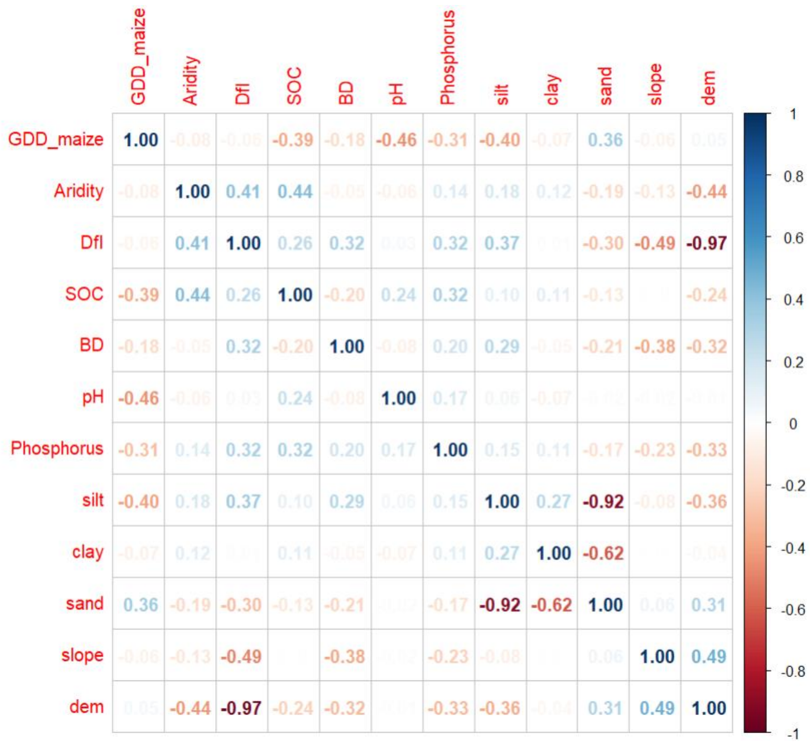


Figure 41, Correlation matrix of the predictors for the prediction of the ES map, when the RF model consider all the different managements together.

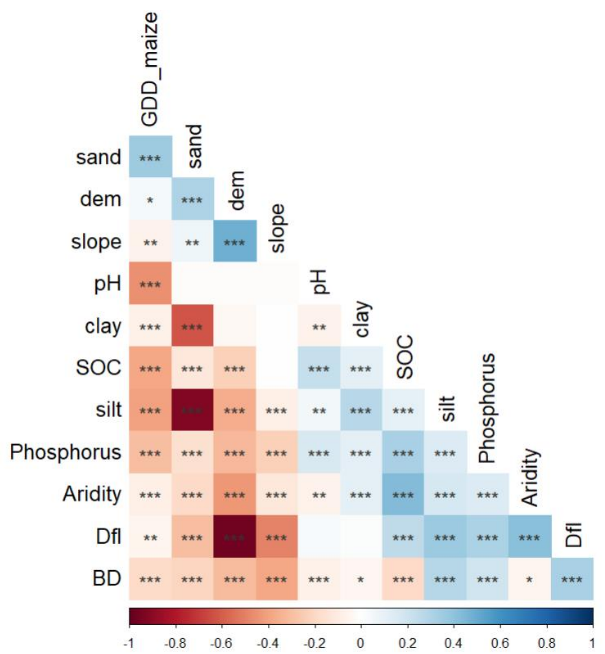


Figure 42, Correlation pyramid for the 'All management' prediction

✓ ***<u>No-Tillage</u>***

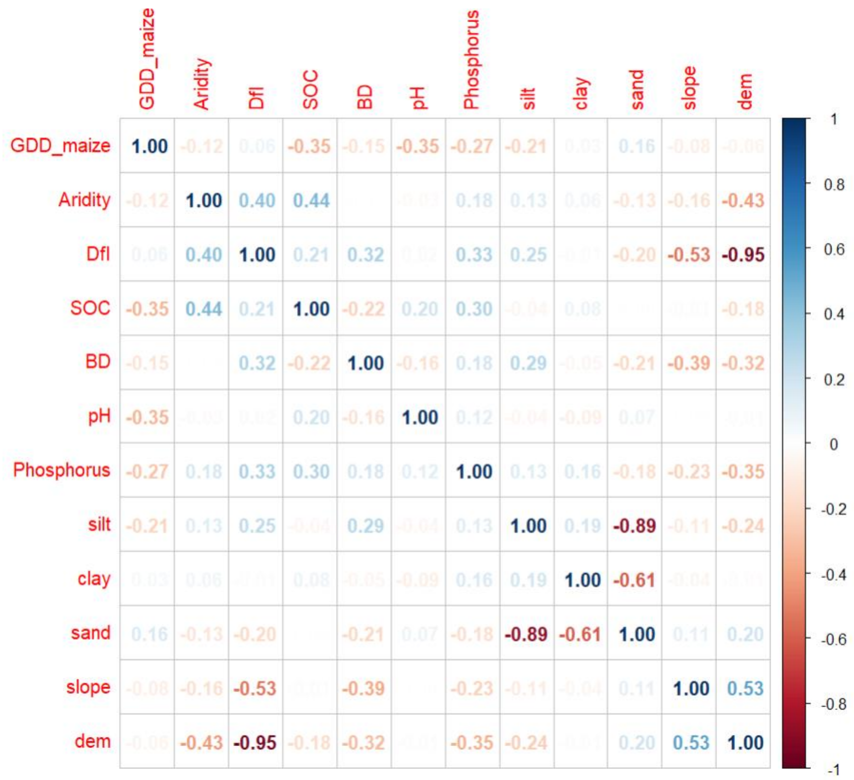


Figure 43, Correlation matrix of the predictors for the prediction of the ES map, when the RF model consider NT as RP.

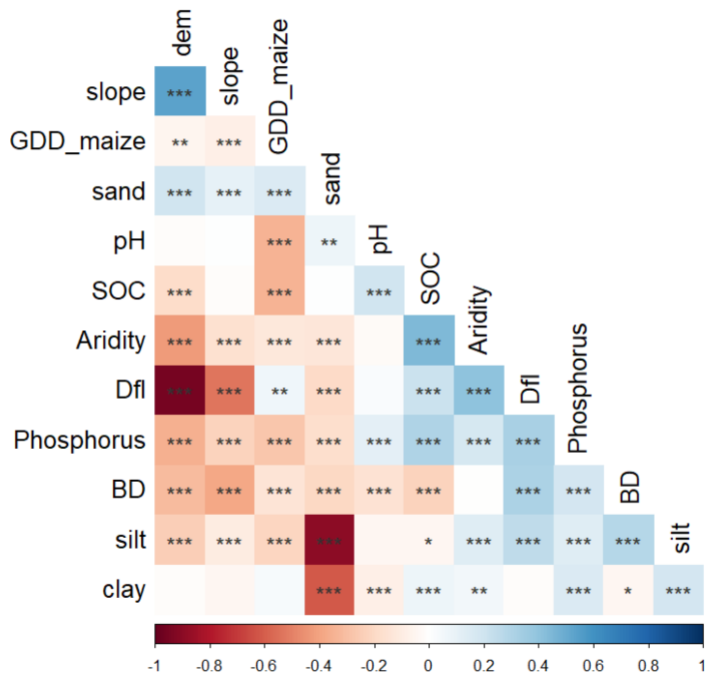


Figure 44, Correlation pyramid for the 'NT management' prediction

✓ *Agroforestry*

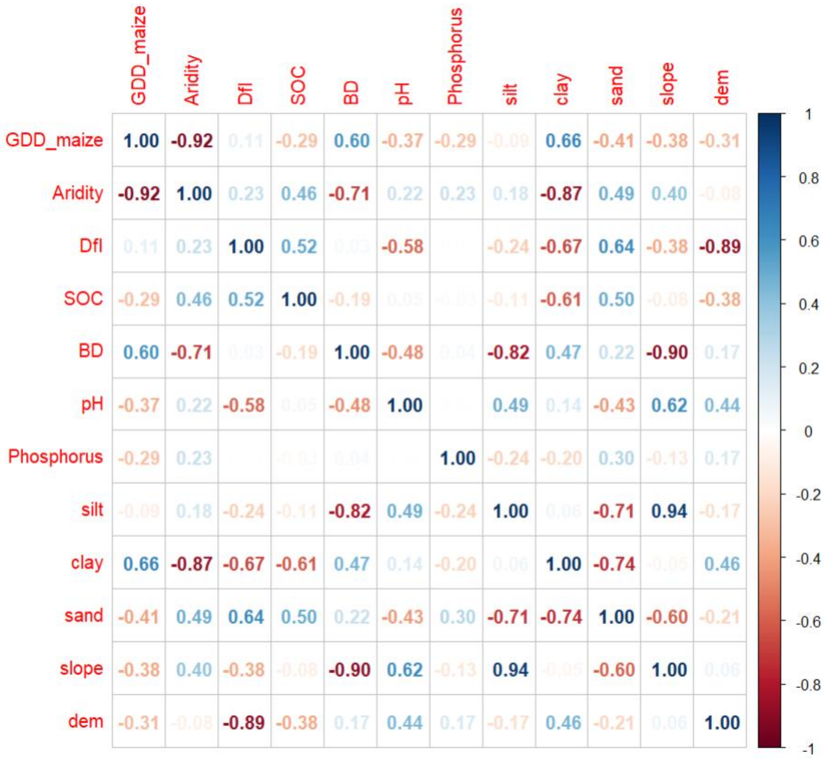


Figure 45, Correlation matrix of the predictors for the prediction of the ES map, when the RF model consider AF as RP.

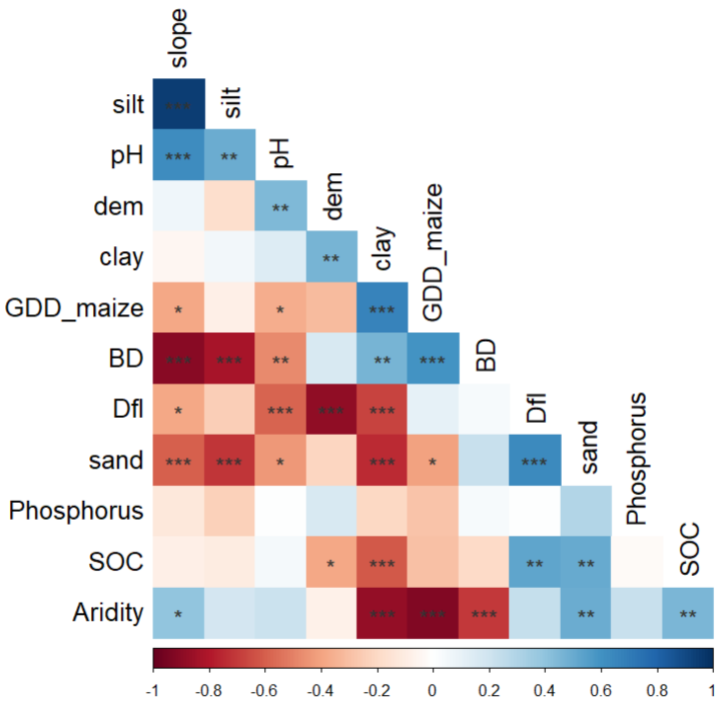


Figure 46, Correlation pyramid for the 'AF management' prediction

✓ ***Organic farming***

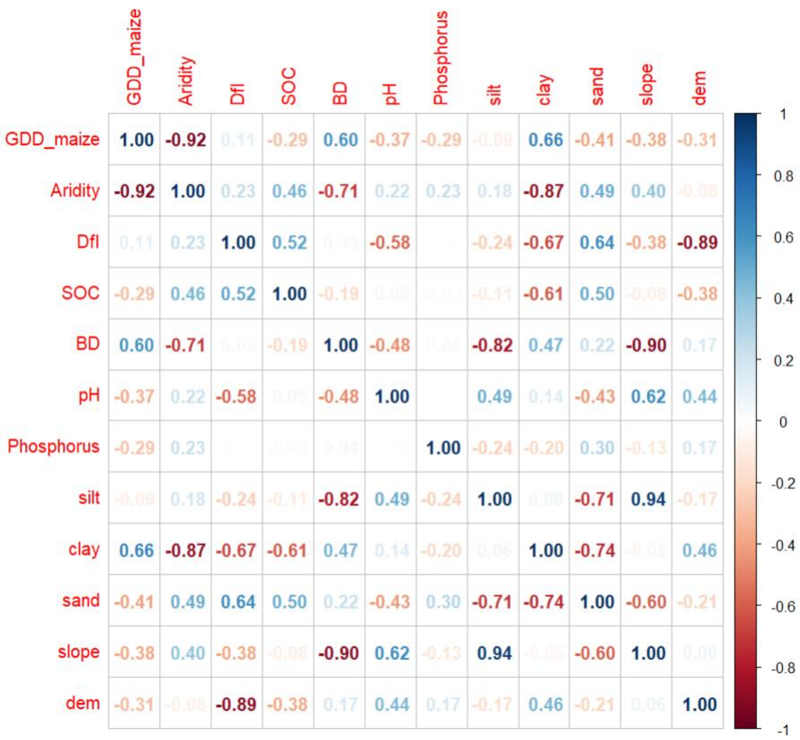


Figure 47, Correlation matrix of the predictors for the prediction of the ES map, when the RF model consider OF as RP.

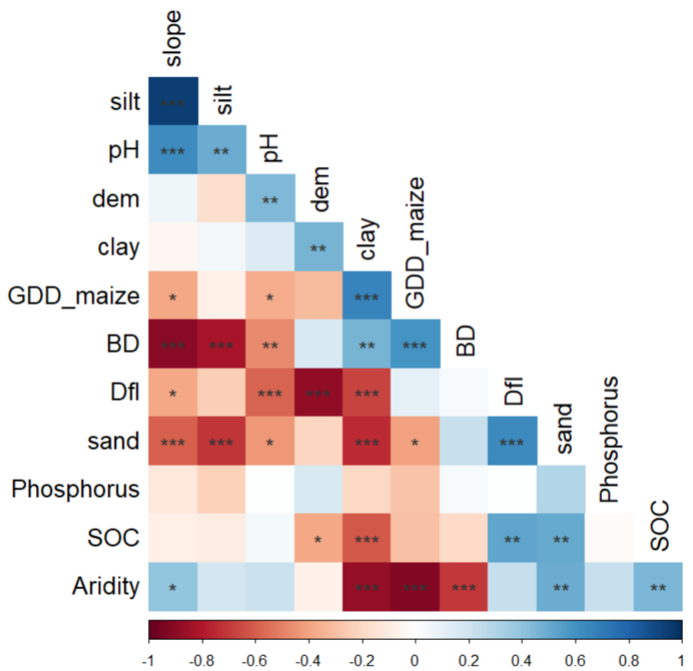


Figure 48, Correlation pyramid for the 'OF management' prediction

B. Prediction and uncertainty maps

✓ *All managements*

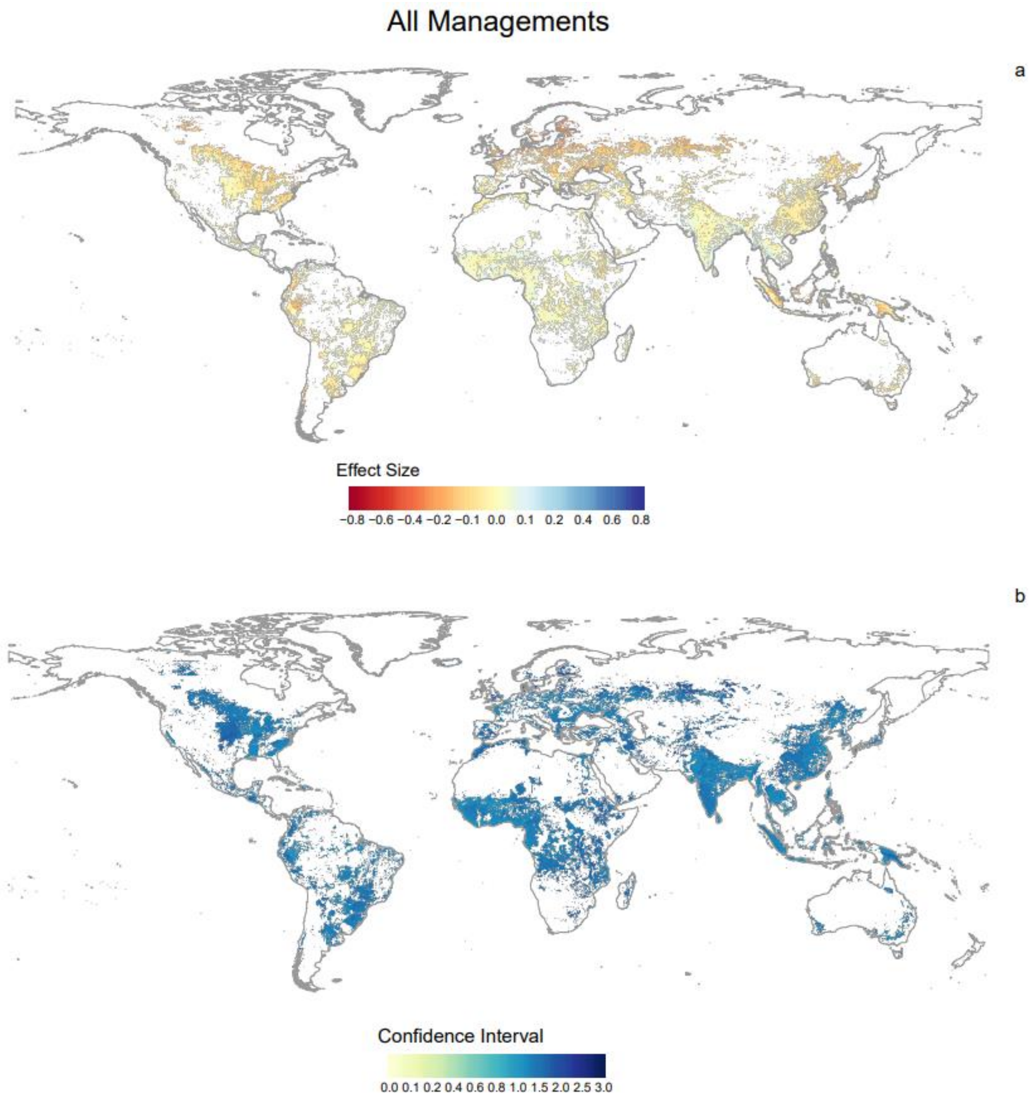


Figure 49, Prediction (a) and uncertainty (b) map of the '*All managements*' prediction.

✓ *No-Tillage*

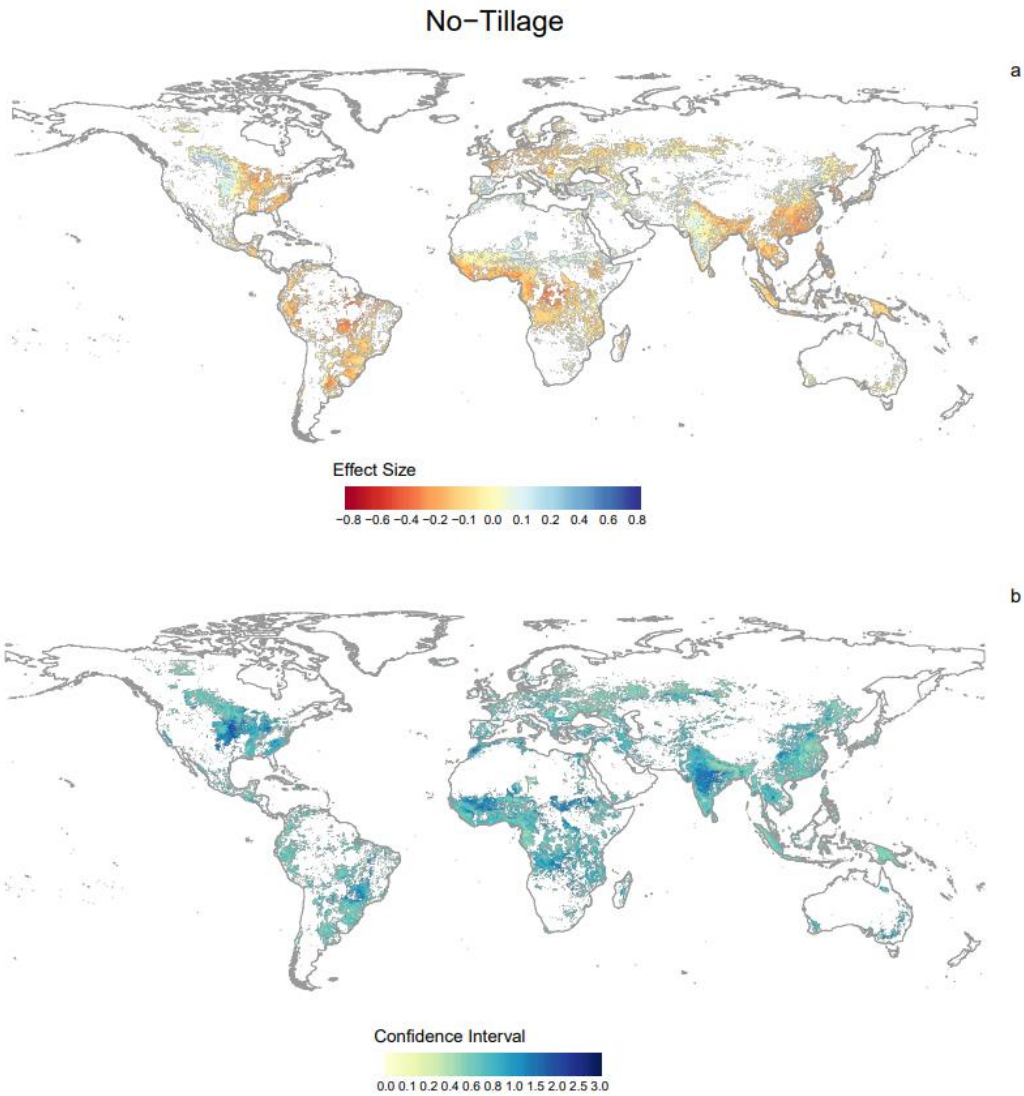


Figure 50, Prediction (a) and uncertainty (b) map of the ES for the NT management.

✓ *Agroforestry*

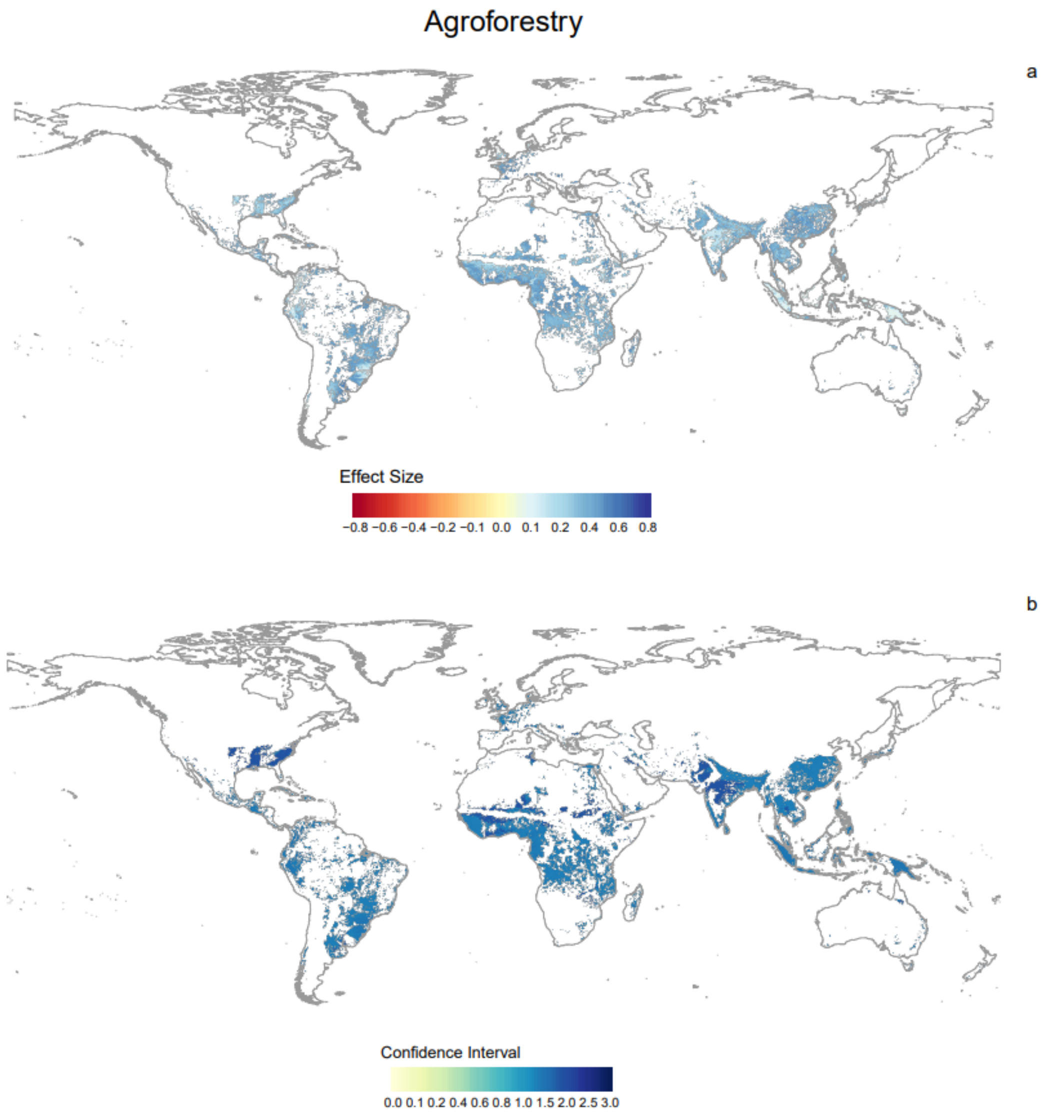


Figure 51, Prediction (a) and uncertainty (b) map for the AF management.

✓ ***Organic Farming***

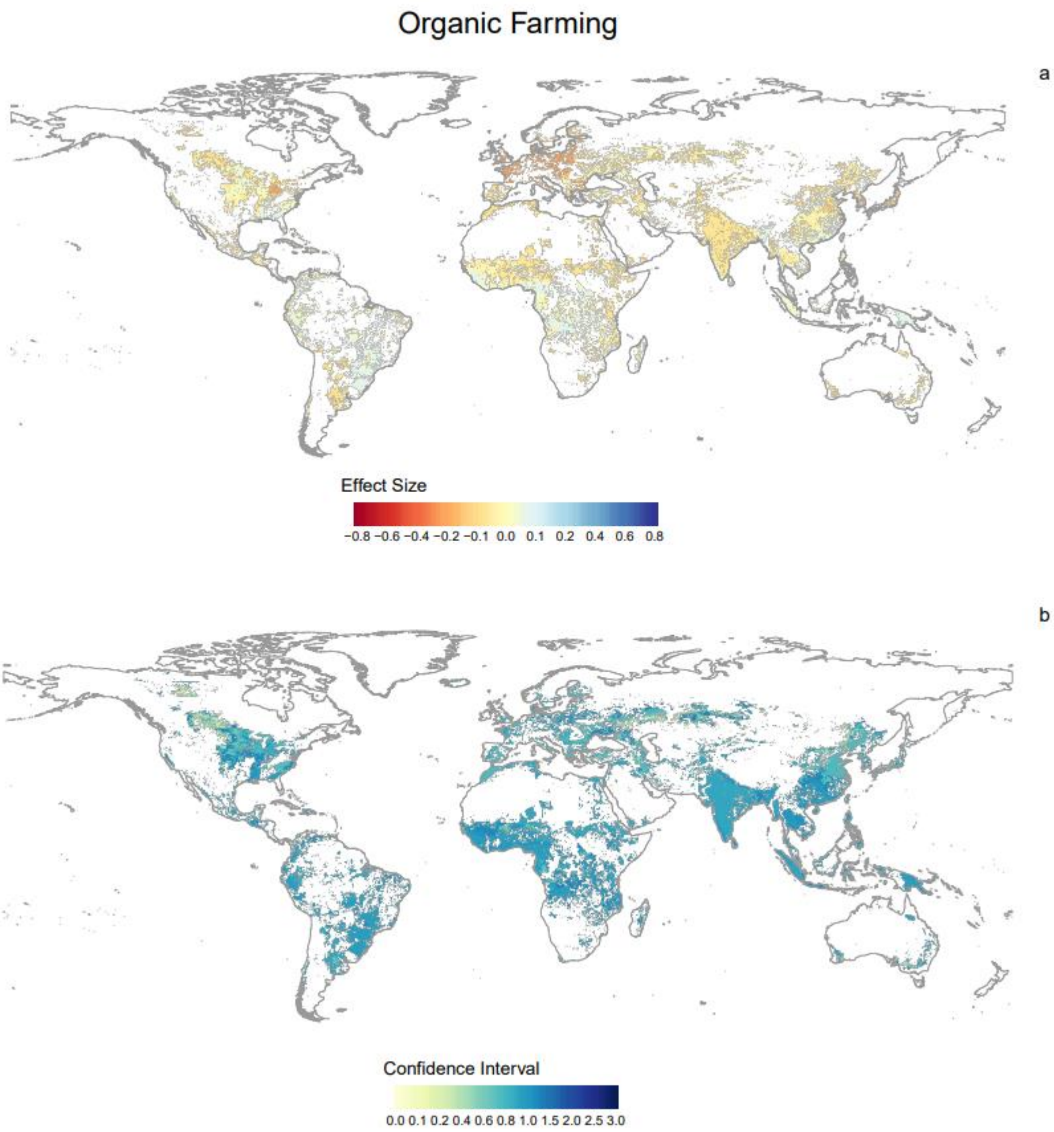


Figure 52, Prediction (a) and uncertainty (b) map of the OF management.

C. Positive and highest effect size (ES) maps

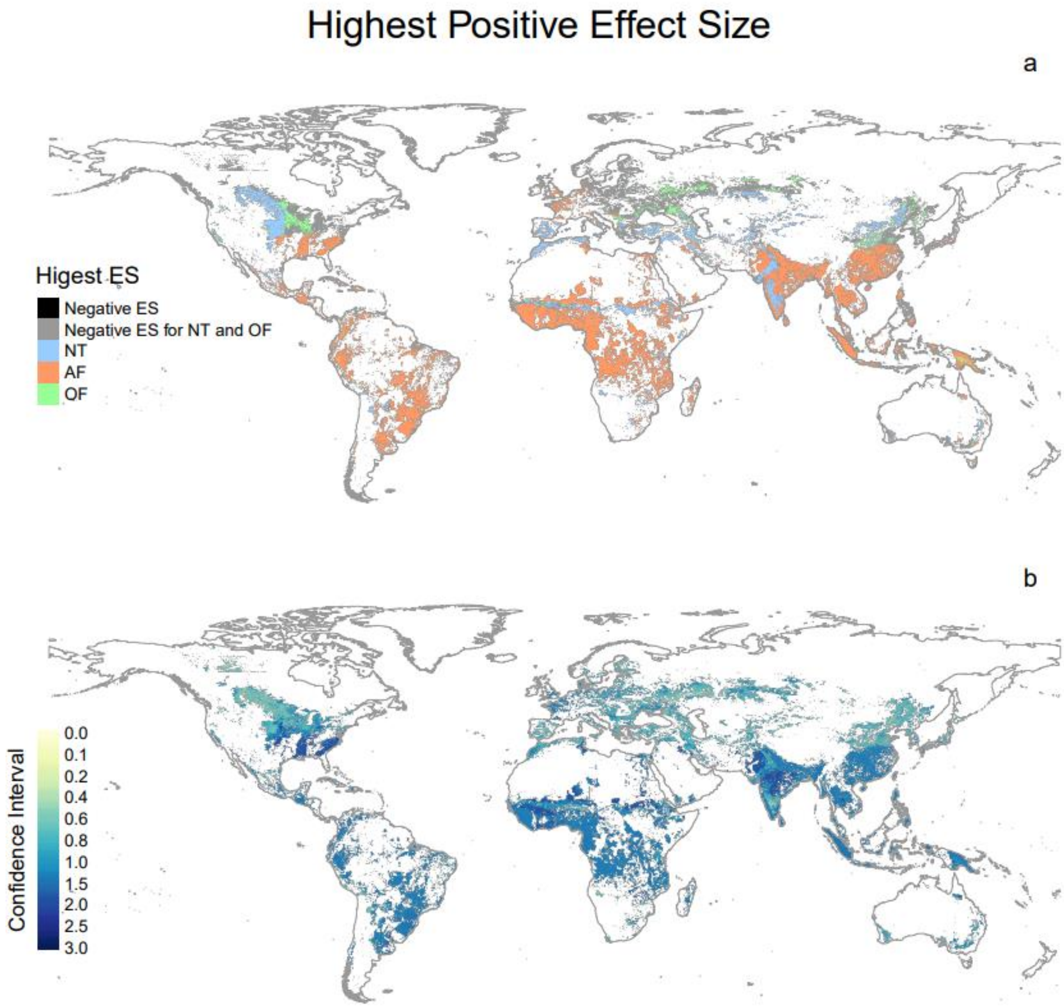


Figure 53, Location where the regenerative managements (NT, OF, AF) could lead to an increase in productivity. Map a: cropland with the highest positive management effect size; map b: uncertainty map for map a. It reports the CI of the management shown in map a; therefore, if in one location the AF has the highest relative productivity, in the same position, map b reports the CI of the AF. While, when there are both the management in the cell the highest CI is plotted in the same cell.

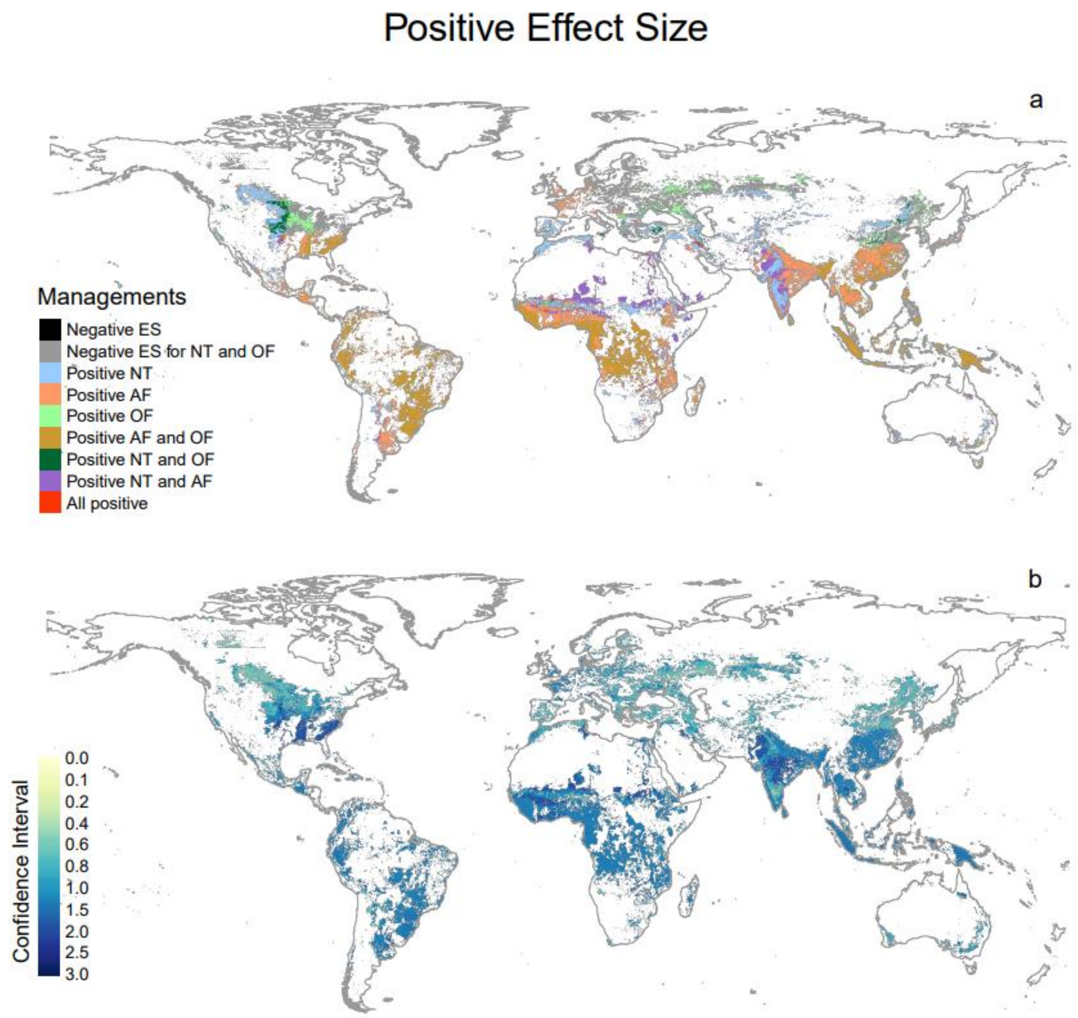


Figure 54, Location where the regenerative managements (NT, OF, AF) could lead to an increase in productivity. Map a: managements with positive ES, in this case more than one regenerative technique can be defined for each cell. Map b: uncertainty map of map a, it has been obtained considering in each location which RP has the highest CI compare to the other plotted in the same cell.

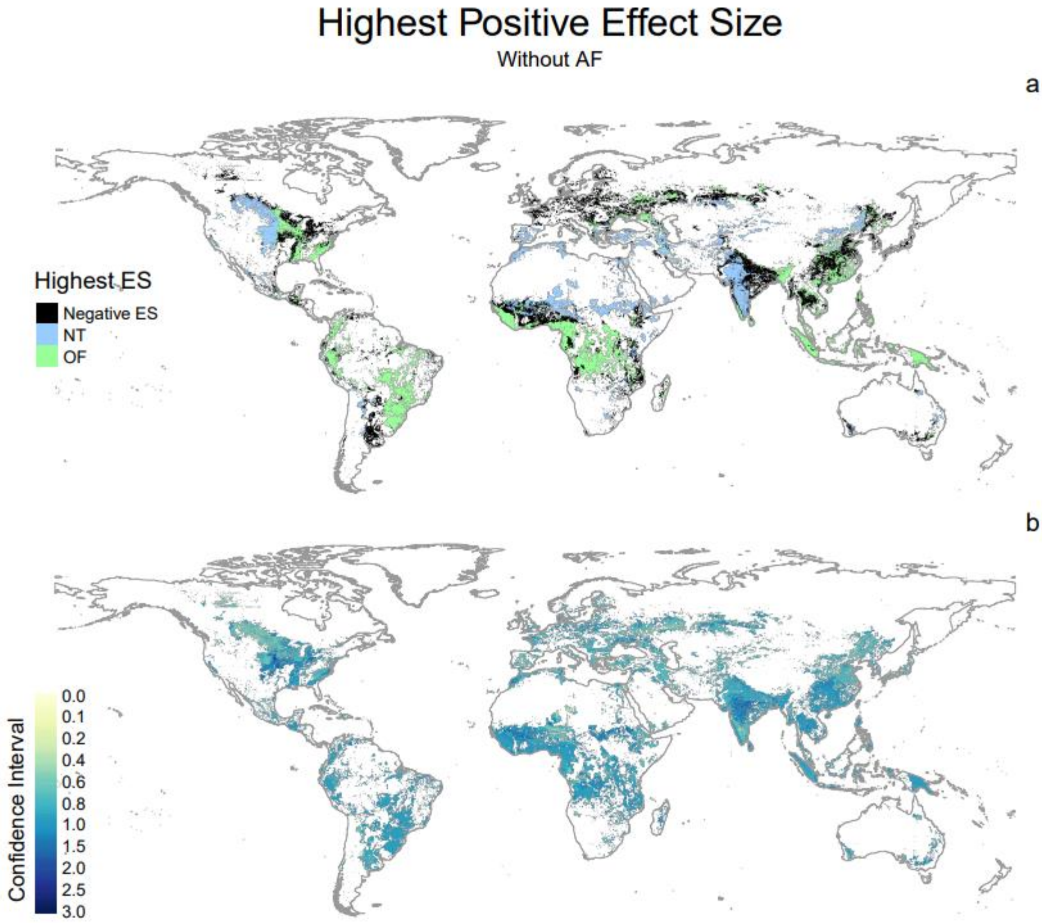


Figure 55, Location where the NT and/or OF could lead an increase in productivity. Map a: cropland which management has the highest positive effect size. Map b: uncertainty map for map a. It reports the CI of the management shows in the map a; therefore, if in one location the OF has the highest relative productivity, in the same position, map b reports the CI of the OF. While, when there are both the management in the cell ("Negative ES") the highest CI is plotted in the same cell.

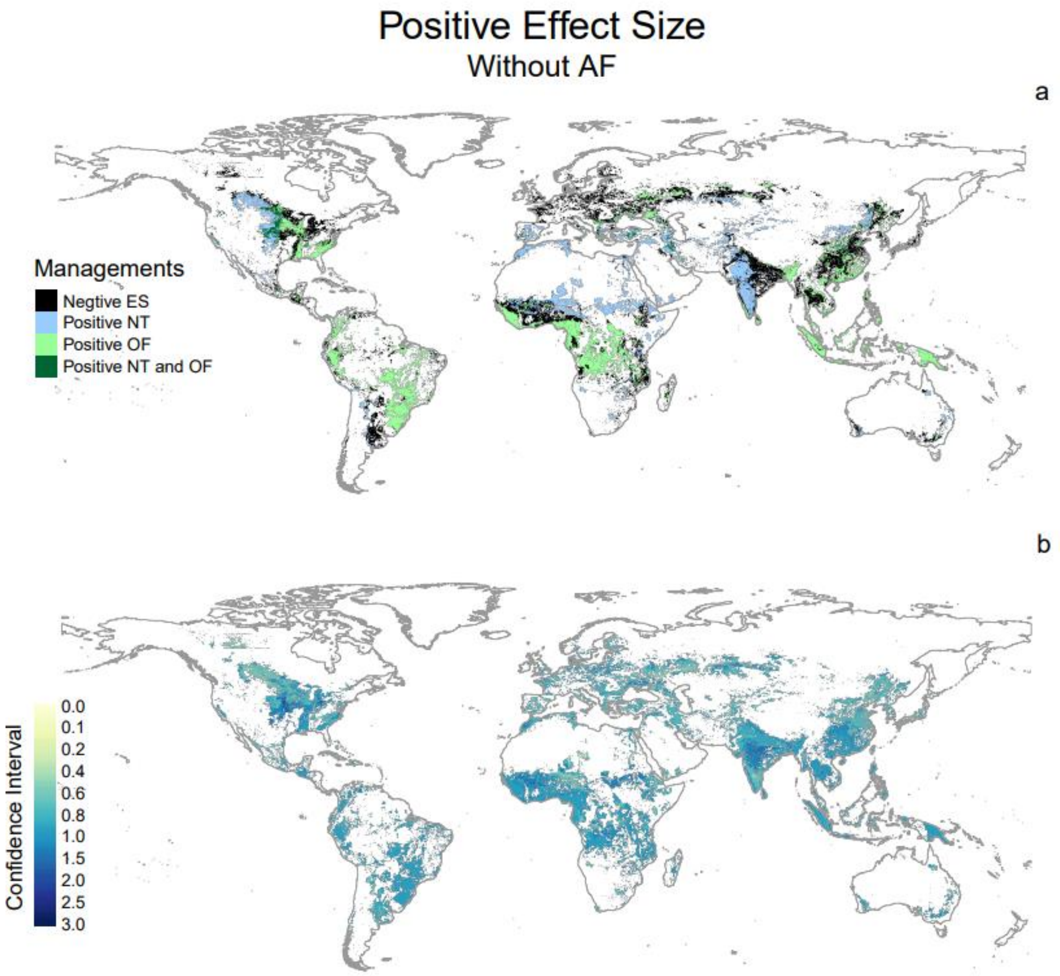


Figure 56, Map a: managements (NT, OF) with positive ES, more than one agricultural strategy can be defined for each cell. Map b: uncertainty map of map a, it has been obtained considering in each location which RP has the highest CI compare to the other plotted in the same cell.