UNIVERSITE JOSEPH KI-ZERBO

****

ECOLE DOCTORALE INFORMATIQUE ET CHANGEMENTS CLIMATIQUES

BURKINA FASO

Unité-Progrès-Justice

MASTER RESEARCH PROGRAM

SPECIALITY: INFORMATICS FOR CLIMATE CHANGE (ICC)

MASTER THESIS

Impacts of Climate Extremes on Crop Production: Modelling Yield Loss Using Machine Learning Algorithms in the Sudanese and North-Guinean climatic zones of Mali

Presented the (date) and by:
Mr. Issa KASSOGUE

Supervisors:

Dr. Seyni SALACK
Regional Thematic Coordinator (RTC) –
Risks and Vulnerability to Climate Extremes, WASCAL
Dr. Sarah Schönbrodt-Stitt
Department of Remote Sensing, Institute of Geography and Geology,
University of Würzburg, Germany

November 2021

Dedication

This work is dedicated to my wife (KASSOGUE Kani COULIBALY) and my daughter (Fatoumata IAO KASSOGUE), who always encouraged me for my studies.

Acknowledgments

This work was only made possible by the support and guidance of some. I would like to express from the bottom of my heart:

To the West African Scientific Service on Climate Change and Adapted Land Use (WASCAL) for providing this wonderful capacity-building framework and the German Ministry of Education and Research (BMBF) for funding this program;

To the Director of ED-ICC, Pr. Tanga Pierre ZOUNGRANA;

To the Deputy Director, Dr. Ousmane COULIBALY;

To Pr. Jean-Marie DIPAMA and Dr. Tizan DAHO, respectively the former Director and Deputy Directors of ED-ICC, and all the administrative staff of ED-ICC, my sincere thanks for the implementation of the educational and social activities.

I sincerely want to express my gratitude to my supervisors Dr. Seyni SALACK and Dr. Sarah Schönbrodt-Stitt, for all their precious advice, support, and teachings. I have always learned with you, your colleagues, and assistants (Dr. Belko Abdoul Aziz DIALLO, Dr. Kwame Oppong Hackman, Mr. Steven Hill, Mr. Koufanou HIEN, Mr. Elidaa Kossi DAKU, Mrs. Adjara DINDANE). I am eternally grateful to you.

Thank you to Mr. Amadou DIAKITE from Mali-Météo;

I address my sincere thanks to all the professors and all the people who, by their writings, advice, and criticism, have guided my reflections;

My sincere thanks to my TRAORE and KASSOGUE families for all the support to me and to that of my family. I am very grateful to you. I will never forget that.

My sincere thanks go to the COULIBALY Family place for all the support, and I am eternally grateful to you;

I thank Miss Andrée NENKAM and Mr. Sibiry TRAORE for your support, and I am grateful to you;

To my brother, cousin, friend, and imam, Mr. Makan Garba DIARISSO, I say thank you very much for everything you have done for me. I say a big thank you to all my student colleagues from the Malian community in Cape Coast, Ouaga, and Kamboissin. A big thank you to you, Mr. Mohamed Lamine SIDIBE; you have made the community a family;

A big thank you to the UMA community of 2iE-Ouaga. With you, I also learned a lot. I say thank you very much, may Allah accompany you in all your works;

My friends and brothers of ***"TEAM - OUAGA"*** (Ltd Mohamath DIANKHA "*from* "*Senegal*" and Mr. ELHADJ IBRAHIM MAMA GARBA), I am very grateful to you, thank you for your friendship and affection;

Thank you to my colleagues from the first batch of ICC-EDICC Ouaga. With you, I learned a lot and especially learned a lot in myself. Thank you to Miss Kokoye Lucien Glory for taking up the challenge and Mr. Issaka Abdou Razakou KIRIBOU for always being there for us; thank you, dear host. Thank you to all those from far and near who contributed directly or indirectly to the facilitation and realization of these studies.

Abstract

Climate change is a global phenomenon of climate transformation of the planet. Climate change causes environmental degradation and decreases in agricultural production and incomes and therefore poses a significant threat to food security and sustainable livelihoods, as well as to socio-economic stability. The approach to delineating impacts of climate extremes on crop production is complex and may involve the use of crop simulation models and, in some cases, the use of statistical techniques of equal complexity. The main objective of this study was to assess the impact of agroclimatic extremes on crop yields. In this research, two machine learning (ML) algorithms, namely logistic regression and Random Forest models, were used to assess the yield loss as a result of agroclimatic extremes. The input data included observed yields of cotton, maize, and millet and meteorological data (1990-2017) from the Sudanian and North Guinean zones of Mali. The growth of the digital age has made almost all human activities the source of ever-increasing amounts of information. machine learning techniques for data analysis can be understood as a problem of pattern recognition or, more informally, knowledge discovery and data mining. The agroclimatic extremes considered in this study are the late onset of the cropping season, early cessation of the cropping season, shorter duration of the rainy season, intra-seasonal heat waves, and seasonal rainfall deficit. While yield loss was the predictand, these agroclimatic extremes were considered as the predictors. When taken individually, a simple linear regression does not describe the relationship between the predictors and the predictand. When considered altogether in ML, such as random forest regression (RF) and logistic regression (LR) modelling, the relationship can be depicted as yield loss as a result of agroclimatic extremes. Our results showed that LCS and cropping season were dominant factors affecting yield. The LCS and cropping season were the most correlated indices. Predicting the occurrence of these agroclimatic extremes have the advantage of identifying suitable agricultural inputs and avoiding certain risks. However, RF showed much better performance compared to LR. Therefore, ML is useful and very robust tool to predict yields loss as a result of extreme climate events, especially when large-scale datasets are available.

Keywords: Agroclimatic extremes, Crop Yield, Machine Learning, Logistic Regression, Random Forest, Modelling, Mali.

Résumé

Le changement climatique est un phénomène global de transformation climatique de la planète. Il provoque une dégradation de l'environnement et une diminution de la production et des revenus agricoles et constitue donc une menace importante pour la sécurité alimentaire et les moyens de subsistance durables, ainsi que pour la stabilité socio-économique. L'approche pour délimiter les impacts des extrêmes climatiques sur la production agricole est complexe et peut impliquer l'utilisation de modèles de simulation de cultures et, dans certains cas, l'utilisation de techniques statistiques de complexité égale. L'objectif principal de cette étude était d'évaluer l'impact des extrêmes agroclimatiques sur les rendements des cultures. Dans cette recherche, deux algorithmes d'apprentissage automatique (en anglais : machine learning, ML), à savoir la régression logistique et les modèles de forêt aléatoire, ont été utilisés pour évaluer la perte de rendement due aux extrêmes agroclimatiques. Les données d'entrée comprenaient les rendements observés de coton, de maïs et de mil et des données météorologiques (1990-2017) des zones soudanienne et nord-guinéenne du Mali. La croissance de l'ère numérique a fait de presque toutes les activités humaines la source de quantités toujours croissantes d'informations. Les techniques d'apprentissage automatique pour l'analyse des données peuvent être comprises comme un problème de reconnaissance de formes ou, de manière plus informelle, de découverte de connaissances et d'exploration de données. Les extrêmes agroclimatiques considérés dans cette étude sont le début tardif de la saison agricole, l'arrêt précoce de la saison agricole, la durée plus courte de la saison des pluies, les vagues de chaleur intra-saisonnières et le déficit pluviométrique saisonnier.

Alors que la perte de rendement était le prédicteur, ces extrêmes agroclimatiques ont été considérés comme les prédicteurs. Prise individuellement, une régression linéaire simple ne décrit pas la relation entre les prédicteurs et le prédictant. Lorsqu'elle est considérée dans son ensemble dans la ML, telle que la modélisation par régression forestière aléatoire (RF) et par régression logistique (LR), la relation peut être décrite comme une perte de rendement résultant d'extrêmes agroclimatiques. Nos résultats ont montré que LCS et OCS étaient des facteurs dominants affectant le rendement. Le LCS et l'OCS étaient les indices les plus corrélés. Prévoir l'occurrence de ces extrêmes agroclimatiques a l'avantage d'identifier les intrants agricoles adaptés et d'éviter certains risques. Cependant, RF a montré de bien meilleures performances par rapport à LR. Par conséquent, ML est un outil utile et plus robuste pour prédire la perte de rendement à la suite d'événements climatiques extrêmes, en particulier lorsque des ensembles de données à grande échelle sont disponibles.

Mots clés : Extrêmes agroclimatiques, Rendement des cultures, Machine Learning, Régression logistique, Forêt aléatoire, Modélisation, Mali.

Liste of Content

Dedication i

Acknowledgments ii

Abstract iii

Résumé iv

Liste of Content vi

List of figures and maps viii

List of tables ix

List of equations x

Abbreviations xi

Introduction 1

1. Context 1

2. Justification of the study 4

3. Objectives and scope of the thesis 6

4. Hypotheses and research questions: 7

Chapter I: State-of-the-art 8

1.1. Definition of agroclimatic extremes 8

1.2. Modeling the effects of weather and climate extremes on crops 12

Chapter II: Data and Methods 13

2.1. Description of the study area 13

2.2. Crop yield and Climatic Data sets 18

2.3. Algorithmic computation of extreme Agroclimatic Indices 25

2.4. Logistic and Random Forest Multi-Linear Regression Models 27

Chapter III: Results and Discussion 31

3.1. Identification and quantification of agroclimatic extremes of strong impacts in Mali 31

3.2. Prediction of yield loss resulting from the synchronous occurrence of agroclimatic extremes........ 34

3.3. Evaluation of the estimation performance of ML modeling algorithms applied to diagnose the impact of agroclimatic extremes on crop yields in Mali........ 36

Conclusion 39

Recommendations........ 40

Bibliography xli

Appendix........ l

List of figures and maps

Figure 1: Distribution of the long-term (1990-2020) average annual rainfall (mm) over Mali. .. 14
Figure 2: The Global Land Cover map of MaliSource: Global Land Cover 2000 15
Figure 3: Pedological map of Mali. 17
Figure 4: Historical characteristics of cotton yields (Kg/ha) over time and space (ha) of Fana district 18
Figure 5: Historical characteristics of maize yields (Kg/ha) over time and space (ha) of Fana district 19
Figure 6: Historical characteristics of millet yields (Kg/ha) over time and space (ha) of Fana district 20
Figure 8: Historical characteristics of cotton yields (Kg/ha) over time and space (ha) of Kita district 20
Figure 9: Historical characteristics of maize yields (Kg/ha) over time and space (ha) of Kita district 20
Figure 10: Historical characteristics of millet yields (Kg/ha) over time and space (ha) of Kita district 21
Figure 12: Historical characteristics of cotton yields (Kg/ha) over time and space (ha) of Koutiala district 21
Figure 13: Historical characteristics of maize yields (Kg/ha) over time and space (ha) of Koutiala district 22
Figure 14: Historical characteristics of millet yields (Kg/ha) over time and space (ha) of Koutiala district 22
Figure 20: Location map of meteorological stations and agricultural data sources. Source: IER (Institut d'Economie Rural) 23
Figure 21: Random Forest model (Anomalous Predictors) and Logistic regression model (Anomalous Predictors) 34
Figure 22: Random Forest model (Dichotomic Predictors) and Logistic regression model (Dichotomic Predictors).. 36

List of tables

Table 1: Algorithmic computation of extreme Agroclimatic Indices 25
Table 2: Correlation coefficients between cotton yield and individual agroclimatic parameters. 32
Table 3: Correlation coefficients between maize yield and individual agroclimatic parameters 33
Table 4: Correlation coefficients between millet yield and individual agroclimatic parameters 33

List of equations

Equation 1: De – trending method 27
Equation 2: Logit link function 28
Equation 3: Relationship between the predictors 28
Equation 4: Mean Squared Error 30
Equation 5: Root-Mean-Squared Error 30
Equation 6: Coefficient of Determination (R^2) 30

Abbreviations

MLR	Multiple Regression Model
AETc	Actual Evapotranspiration
AGRHYMET	Agrometeorology and Operational Hydrology Training and Application Centre
CILSS	Comité permanent Inter-États de lutte contre la sécheresse dans le Sahel
CMDT	Compagnie Malienne pour le Développement des textiles
CPCS	Commission de Pédologie et de Cartographie des Sols
CSI	Critical Success Index
CYA	Crop yield anomaly
DHS	Daily heat stress
DS	Dry season
DTR	Daily range of temperature
EC	Early cessation
ECMWF	European Centre for Medium-Range Weather Forecasts
ECS	End of the cropping season
ED-ICC	Ecole Doctorale d'Informatique pour les Changements Climatiques
ERS	End of the rainy season
FAO	Food and Agriculture Organization of the United Nations
FER	Effective first rain
FO	False onset
FON	False Onset of the cropping season
GDP	Gross domestic product
HWN	Heat wave number
INRA	National Institute for Agricultural Research
ITCZ	Intertropical Convergence Zone
ITF	Convergence Intertropical Front
LCCC	Lin concordance correlation coefficient
LCS	Length of the Cropping Season
LRM (AP)	Logistic Regression model (Anomalous Predictors)
LRM (DP)	Logistic Regression model (Dichotomic Predictors)
MSE	Mean Squared Error

OCS	Onset of Cropping Season
ORS	Onset of the rainy season
OSS	Observatoire du Sahara et du Sahel
PETc	Potential evapotranspiration
PoD	Probability of Detection
R^2	Coefficient of determination
RFM (AP)	Random Forest model (Anomalous Predictors)
RFM (DP)	Random Forest model (Dichotomic Predictors).
RF	Random Forest
SR	Success Ratio
UNFPA	United Nations Population Fund
USAID	United States Agency for International Development
WASCAL	West African Science Service Centre for Climate Change and Adapted Land Use
WRSI	Water Requirement Satisfaction Index
WS	Wet season

Introduction

1. Context

Climate change is the global phenomenon of climate transformation of the planet that has multiple consequences, which constitute a major socio-economic issue (USAID, 2014). It causes environmental degradation and a decrease in agricultural production and income, therefore constitutes a significant threat to sustainable food security and livelihood, and socio-economic stability (USAID, 2017). According to the Climate Change Risk Profile for West Africa by the USAID (2017), West Africa is one of the globally most jeopardized regions affected by increasing climate variability. The region largely depends on rainfed agriculture to sustain food security and livelihood. Therefore, every change in rainfall pattern (e.g., rainfall amount, intensity and timing), temperatures, and the effect of strong winds may likely affect farming systems (Alhassane et al., 2013). According to Wilson & Minas (2017), the annual mean temperatures in West Africa, and particularly in the Sahel, have increased faster than the global trend, with increases of the annual mean temperature ranging from 0.2 °C to 0.8 °C per decade since the end of the 1970s in the Sahelo-Saharan, Sahelian and Sudanese areas.

According to AGRHYMET (2010a), after the drought seasons of the 1970s and 1980s in West Africa, AGRHYMET has taken up the challenges faced by the West African countries, particularly those in the Sahel (e.g., Mali and Niger), suffering the effects of heavy rains and devastating floods. AGRHYMET (2010b) estimates that the damages and losses linked to these extreme hydro-climatic events have caused costs of several million US$, which are higher than the costs linked to the implementation of adaptation strategies. For example, between 2000 and 2008, the costs of damage related to floods in the area governed by the Permanent Inter-State Committee against Drought in the Sahel (CILSS) were estimated to be between 39 and 80 billion US$; and for the minimum and maximum scenario for Mali to amount to 5,860,665 US$ (*1000) and to 12,029,353 US$ (*1000), respectively. In addition, flood and drought events have undermined human systems (human and material losses), agricultural systems (submerged crops), and economic infrastructure (e.g., roads, bridges, dams destroyed).

To ensure the accumulated adaptive capacity of smallholders and a transition of countries to climate-resilient development, the FAO (2016a) proposes a profound transformation of the global food and agricultural system. It underlines the importance of the factor "time" and recommends actions to be taken now to ensure sustainable food and agriculture for future climate extremes.

The Republic of Mali, like any other West African country, faces multiple pressures induced by climate (e.g., desertification, soil erosion, temperature rise, increase in floods) and demographic changes (Soumaré et al., 2020). Therefore, it must increase its capacity to develop agricultural strategies to ensure sustainable socio-economic balance, food security, and self-sufficiency. According to Ouedraogo (2013), beyond the socio-economic constraints (e.g., famine, the limitations on productivity and competitiveness) caused by frequent droughts, the rainy season generally results in a succession of wet and dry periods of varying duration, including false onset and early end of the cropping seasons.

Speaking of climate change, experts address more questions about its impacts on forest biomass, water availability, and fluctuating crop yields as the direct cause of the increased rainfall variability. The wide variability in the intensity, frequency, and timing of annual or seasonal precipitation creates significant challenges for farmers (King et al., 2014). Beyond the above challenges, Salack et al. (2020) indicate that in the Sahel, agroclimatic extremes such as the false onset of the season, floods, and early cessation can create restrictive conditions for a good harvest for farmers.

Agroclimatic extremes are extreme meteorological or climatic phenomena manifested by unexpected, unusual, severe meteorological conditions, i.e., beyond the threshold, and which weigh on agriculture. The impact of these "agroclimatic extremes" on crop yields in Mali has rarely been a topic of discussion. Understanding the relationship between these agroclimatic extremes and historical crop yields from 1990 to 2017 is essential to assess the sustainability of our agricultural production.

The Observatoire du Sahara et du Sahel, OSS (2013) defined an agricultural drought as a situation where soil water and water reserves become insufficient to meet crop needs in a given region.

The false onset is one of the factors that contribute to the drying out of the crops and soil (Ngoune Liliane & Shelton Charles, 2020). The dehydration can cause the soil impermeability, seeds desiccation , resulting in stagnation of germination, limitation of seedling root mobility, and the insolubility and/or unavailability of nutrients to plants by lack of water (Rawson & Macpherson, 2000; Fahad et al., 2017). Agricultural drought can also favor the proliferation of predators, such as insects and other microorganisms that live on juvenile and distressed plants whose eggs below the ground just needed a wet and dry situation to hatch (Skendžić et al., 2021).

Temperature variations have a significant impact on vegetative growth and grain yield depending on the type of crop.(Hatfield & Prueger, 2015). These effects are evident in a high rate of senescence (a physiological process that causes a slow degradation of cell functions), which reduces the ability of the crop to efficiently fill grains or fruits.

The early cessation of rains is the discontinuity of rainfall during the rainy season. During the rainy season, the early cessation deprives the plants of their water requirement, during their growth and development, when crops will be in a water deficit to survive. This difficulty causes wilting and dieback of the plants and their exposure to be attacked by predators. This adversely affects the growth and development of crops and their production performance.

The indicator of crop performance based on the availability of water to the crop during a growing season (Senay, 2004; Verdin & Klaver, 2002) denotes the balance between all the water resources that the plants need and those that are leaving with respect to a specific area (steep slope) and a defined period,(hot and high evaporation zone). The Water Requirement Satisfaction Index plays a major role in determining the quantity of water available and is necessary to satisfy the plant's water requirement, considering other factors (e.g., evapotranspiration, soil field capacity of water loss.

According to Dimitriadis & Goumopoulos (2008), the urgent need to increase agricultural production, especially on an increasingly small land suitable for agriculture, as well as the reduction of consumption of resources such as water and fertilizers vis-à-vis the environment, make the use of new techniques and methods a top priority.

The growth of the digital age has made almost all human activities the source of ever-increasing amounts of information. This information often takes the form of computable data, i.e., data available in a format that can be processed by machine and, ultimately, reasoned (Vellido et al., 2012). Advances in machine learning and simulation crop modelling have created new opportunities to improve prediction in agriculture (Baştanlar & Özuysal, 2014; Shakil Ahamed et al., 2015) These technologies have each provided unique capabilities and significant advancements in the prediction performance, however, they have been mainly assessed separately and there may be benefits integrating them to further increase prediction accuracy (Bengio, 2009).

This deluge of data is invading most scientific fields (Liakos et al., 2018). According to Liakos et al. (2018), agro-technology and precision agriculture, also known as digital agriculture, were born as new scientific fields that use data-intensive approaches to optimize agricultural productivity while minimizing environmental impact. The raw materials of complex data at different levels with increasing diversity of characteristics are used to attempt modelling using their wide range of methods and tools. (Vellido et al., 2012).

The data collected in modern agricultural activities is provided by a variety of sensors which allow a better understanding of the operating environment and the operation itself, being able to produce data with high precision and decision making in a better time.(Liakos et al., 2018). The obtained models are meant to be a synthetic representation of the available, observed data that captures some of their intrinsic regularities

or patterns. Therefore, the use of machine learning techniques for data analysis can be understood as a problem of pattern recognition or, more informally, of knowledge discovery and data mining Vellido et al. (2012) Making machine learning models interpretable. Precision agriculture is a suite of management strategies, technologies, and practices to solve the above problems. Precision agriculture applies technologies and principles that use the information to manage spatial and temporal variability to increase resource efficiency and minimize environmental degradation (Dimitriadis & Goumopoulos, 2008).

2. Justification of the study

Rainfed agriculture is susceptible to climate variability and extreme events, such as prolonged droughts and floods (Salack, 2016). According to Maiga et al. (2019)(Maiga et al., 2019), agriculture in Mali accounts for 36% of the gross domestic product (GDP) while employing about 80% of the population estimated to be 20.3 million inhabitants in 2020 (UNFPA, 2020), with the majority engaged in subsistence farming. Subsistence farming in the Sahel, particularly in Mali, highly depends on rainfall. However, rainfed agriculture is experiencing major challenges due to the climatic extremes that negatively impact crop productivity.

In Mali, crop productivity appreciation (consider good or bad in terms of the previous year's performance) is, in part, directly related to the frequency of agroclimatic extremes during critical growth phases of crop development.

According to the FAO (2021a), Agriculture continues to bear the brunt of the impacts of climate change, biological hazards, such as pests and epidemics, pose a serious risk to the life and health of humans, animals and animals. plants.The occurrences of pests and diseases often coincide with extreme weather events and abnormal weather conditions (Rosenzweig et al., 2000). Pests and diseases of plants and animals in general have always been a destabilizing factor for agriculture and a major threat to food security (FAO, 2021a).

In all their phenological stages, crops are sensitive to variations to temperature extremes. Temperature is the main factor that controls the speed of the development of the crop (Tshiabukole, 2018). In general, crop growth accelerates with increasing temperature, a phenomenon that is often described as a linear function of the daily mean temperature (FAO, 2016a). Any adverse effect of heat stress on membranes leads to disruption of cell activity or death of the crop (Bazzaz & Sombroek, 1997). According to Sarr et al. (2012), the increases in minimum and maximum temperatures, high rainfall variability, intense droughts, false onset, early cessations, and floods constitute climatic extremes, subsequently risks to the agricultural system. Likewise, the increase in temperature can translates into a decrease in humidity, a decrease in the number of cold days and nights, and an increase in hot days and nights (Loko et al., 2013). According to AGRHYMET (2010a),), the rise in temperature could decrease the duration of the phenological phases of cultures, as well as their cycles. Thus, the yields of crops such as millet and sorghum could be affected and cause a drop of more than 10% when the temperature increases by 2 ° C and of the order of 15 to 25% at 3 °C (AGRHYMET, 2009).

Temperatures in West Africa, particularly in the Sahel, have risen more quickly than the global average, ranging from 0.2 °C to 0.8 °C per decade since the beginning of the 20th century (AGRHYMET, 2010b).

Extremes such as false onset create inaccurate planning in the production cycle and increase farmers' number of working days. Subsequently, an additional purchase of seed and use of additional labor will be required. There is also the early cessation of the rains, which limits the productivity potential of crops due to water stress at critical stages of crop growth, development, and production.

According to the National Institute for Agricultural Research (INRA, 2006), an agricultural drought refers to any lack of water that does not allow cultivated plants to express the yield expected in a favorable situation. Thus, agricultural droughts affect the entire lifespan of plants, respectively, crop growing cycle and therefore the quality of crops and harvested products. According to Hatfield & Prueger (2015), the leaves of a plant allow it to absorb gases from the atmosphere through the stomata and capture the solar energy necessary for photosynthesis. An important consequence of reduced photosynthesis is the synthesis of toxic oxidizing compounds in cells. Drought also alters the nitrogen requirements of crops since these increase with the biomass produced. This results in a reduction in grains and, therefore, yield (Hatfield et al., 2014). On the other hand, floods can cause waterlogging of the soil and decrease the availability of oxygen necessary for the respiration of the plants. Anoxia leads to a slowing down or even a halt in the plant's metabolism (Trenberth, 2012). The roots stop functioning, which causes the stomata to close (i.e., blocking photosynthesis), resulting in the stop of nitrogen uptake. The germination rate can also be affected by waterlogging of the soil. (Pérez-Ramos & Marañón, 2009). In addition to this, the possible loss of topsoil caused by water erosion should be noted as a potential for long-term loss of water (Teixeira et al., 2013).

Farmers with limited financial resources and few technological opportunities experience significant upheaval and financial loss for proportionately abrupt shifts in crop yields and productivity (Pennsylvania, 2018). Subsequently, coping with the challenges of climatic extremes is constraining the livelihood of subsistence farmers. Therefore, it has become increasingly important to understand better, assess, and predict the impacts of climate on crop growth, development, and yield.

3. Objectives and scope of the thesis

The main objective of this study is to assess the impact of agroclimatic extremes on crop yields (i.e., cotton, maize, and millet) as observed during the period 1990-2017 in the Sudanian and North Guinean zones of Mali. The study used machine learning (ML) Algorithms to model the patterns relating yield loss to the synchronous occurrence of a combination of several agroclimatic extremes such as late onset of the cropping season, early cessation of the cropping season, shorter duration of the rainy season, intra-seasonal heat waves, and seasonal rainfall deficit.

The specific objectives are to:

(1) identify and quantify the agroclimatic factors of high impacts in Mali

(2) quantify the relationship between these extremes and cotton, maize and millet, fiber and grain yields observed for the period 1990-2017

(3) use ML algorithms to predict yield loss as a result of the synchronous occurrence of agroclimatic extremes;

(4) Assess the performance of ML algorithms to diagnose the impact of agroclimatic extremes on crop yields in Mali.

Following the introduction, this document is organized into three chapters, including state-of-the-art (Chapter 1), Data and Methods (Chapter 2), and results and discussions (Chapter 3). Finally, a conclusion and perspectives are provided at the end of the document.

4. Hypotheses and research questions:

Research questions

To test the hypotheses, the research answered the following questions:

- ✓ What are the agroclimatic extremes influencing crop yields in Mali?
- ✓ Are there separate and combined effects of agroclimatic extremes on yield?
- ✓ Can machine learning depict yield loss induced by agroclimatic extremes?

Hypotheses

To answer the research questions, following hypotheses will be tested:

i) Agroclimatic extremes affect crop yields in Mali,

ii) The synchronized collective influence of agroclimatic extremes outweighs their individual and separated effects on grain yield of staple corps, and

iii) RF and LR models can predict yield loss as a result of agroclimatic extremes.

Chapter I: State-of-the-art

1.1. Definition of agroclimatic extremes

1.1.1. False onset of the cropping season

Farmers usually aim to planting crops with the onset of the rainy season, but a clear demarcation of the onset can be difficult. Definitions often include a certain precipitation threshold, e.g., 20 mm rain over three days, and no seven-day drought period after that (Pashiardis & Michaelides, 2008). Sivakumar (1988) reported optimum soil moisture for planting. According to Soumaré (2008), the first rains in a new crop growing season are paramount in the crop growing cycle, thus also crop yields. However, the benefits of early planting may be negated in 'False onset' years, in which a wet period that encourages planting is followed by a drought that may reduce seedling density or necessitate replanting (Luna et al., 2011; Salack et al., 2020). Therefore, special attention should be paid to the role of the False onset of the cropping season in the quality of the cropping season.(Salack et al., 2020). The false onset of the cropping season event is a first rain followed by a period of drought (AGRHYMET, 2010a).

According to Salack et al. (2020), the false onset of a cropping season refers to the erratic most distribution at the beginning of the rainy season, which involves a heavy rain event followed by a long dry spell. Its manifestation during a rainy season creates water stress conditions in the whole process of agricultural production, especially from planning and management to carry out agricultural activities (Koufanou, 2019). The false onset, considered to be events that, can have negative impacts on agricultural production because it causes the topsoil to dry out, diverting germination or emergence of seedling Ati et al. (2002), or driving seedling abortion (Salack et al., 2020), and the exposure of seeds and seedlings to predators (Luna et al., 2011; Skendžić et al., 2021). This often forces farmers to resort to re-sowing, transplanting, or replanting to replace missing or lost seedlings. However, its identification and forecasting are of fundamental importance for the planning and management of agricultural activities.

1.1.2. Heavy rain events

Measures in the relative amount of annual rainfall delivered by significant, single-day precipitation events show change over time (FAO, 2002). Heavy rain events define as days with precipitation in the top 1 percent of all days. Heavy rain events are the most common causes of water stagnation, waterlogging of shallow soils, water erosion of arable land in high runoff areas, and fungal infestation of some crop leaves and roots (Salack et al., 2015).

Depending on its timing, severity, and previous environmental conditions, heavy rainfall events can provide much-needed relief from droughts and boost crop productivity, or it can exacerbate flooding on already saturated soils and decimate crops.

1.1.3. Onset of the cropping season

According to Stewart (1993), the onset of the cropping season is the most agriculturally relevant variable related to all the other seasonal variables. The amount of water available to plants depends on the onset of the cropping season, end of the cropping season, and length of the cropping season (King et al., 2014).

According to Salack et al. (2020), the Onset of the cropping season begins when the surface energy contrast between the ocean and the continent transforms the flow of tropical winds from the east and north-east (south of the equator) in a southwesterly flow, favoring the incursion of moisture from the ocean into the continent. Onset of the cropping season identification is based on daily analysis of the soil water balance over the initial growth stage (30 days) by identifying and quantifying the risk of failure of crop development (Mugalavai et al., 2008). The Onset of the cropping season determines the planting date, with planting too early possibly leading to crop failure and with planting too late leading to a reduced growing season and crop yield (Dodd & Jolliffe, 2001).

1.1.4. Early cessation of the cropping season

According to Mugalavai et al. (2008), the early cessation of the cropping season is based on the daily soil water balance analysis by identifying and quantifying the water stress in the root zone. The length of the growing season for a particular year is obtained from the difference between cessation and onset of that year. For Ed et al. (2013), the early cessation of the cropping season will result in the shortening of the growing season of crops; therefore, crops will not reach their stage of physiological maturity. Iortyom et al. (2017) reported that the early cessation of the rainy season has more effect on crop yield than the onset of the rainy season, especially when the harvest is approaching its fruiting stage, they require more water to the growth. Therefore, stopping early affects the development of the crop.

1.1.5. Water requirement satisfaction index

The Water Requirement Satisfaction Index is an indicator of crop performance based on the water availability for the crop during a growing season. FAO, (1977, 1986) studies have shown that the WRSI can be linked to crop production using a crop-specific linear yield reduction function. Permanent or temporary stress linked to water deficit or its excess limit the growth and distribution of vegetation and the performance of cultivated plants more than any other environmental factor (Senay, 2004).

1.1.6. Daily temperature range

Temperature variations can take several configurations: average temperature changes (monthly and annual); changes in high daytime temperatures and low nighttime temperatures; and changes in the timing, intensity and duration of extremely hot or cold weather(Hatfield et al., 2011). The daily temperature range defines as the difference between the minimum and maximum temperature at a given day.

In general, crops are most sensitive to high temperatures at their reproductive stage and the grain filling/fruit ripening stage (Hatfield et al., 2011). However, plants' responses to each type of temperature alteration are species-specific and mediated by photosynthetic activity for biomass accumulation. The latter is responsible for plant growth and changes of phenological and morphological characteristics occurring during plant development. Thus, each type of heat stress affects the growing time and overall plant productivity. Adapting to these effects will require different types of responses (Wahid et al., 2007).

1.1.7. Drought and wet stress

The wet season is the time of year when most of a region's average annual rainfall occurs. In the West African Sahel, the cumulative rainfall of extremely wet days and the maximum number of consecutive wet days have increased since the late 1980s, indicating that extreme rainfall events have become more frequent during the last decades (Salack et al., 2016). The soil is strongly affected by extreme precipitation (Priori et al., 2021). If it is too wet, it clogs the ground. The nutrients in the soil can be leached or drained, thus will not be available to the roots of the plants (Indoria et al., 2020). This leads to poor growth and overall poor health and can also lead to bacteria, fungi, and mold growth in the soil. Wet stress results from an imbalance between the supply provided by soil water and the amount needed by the plant as determined by the atmosphere, assuming complete plant cover (Allen et al., 1998). Wet stress arises from a lack of water supply concerning the water contained in the soil becoming insufficient for the needs of the plants, the rate of intense transpiration, which can contribute to a sharp reduction in the level of water contained in the cells of the plants (Rahman & Hasegawa, 2012). The dry season is a yearly period of low rainfall. Seasonal drought occurs in climates with well-defined annual rainfall (i.e., unimodal rainfall regime) and dry seasons like Mali (Sahelian climate).

It is recognized when temperatures induce high rates of evaporation and transpiration. Even frequent showers may not provide enough water to restore the amount lost; This leads to water deficiency, water deficit in the soil, and affects crop yields. According to Fortier (2021), a soil water deficit is a measurement index that makes it possible to differentiate between the field capacity and the actual soil moisture content.

1.1.8. Heat Stress

Temperature is a primary factor affecting the rate of plant development (Hatfield & Prueger, 2015). Therefore, heat stress has been recognized as a significant threat to food supply and security (Teixeira et al., 2013). Furthermore, crops and vegetation are among the most vulnerable systems to climate change, particularly climate extremes (Sun et al., 2019).

The rate of growth and development of plants depends on the surrounding temperature. Each species has a specific temperature range represented by a minimum, maximum, and optimum (Hatfield & Prueger, 2015). The responses to temperature differ among cultivated species throughout their life cycle and are mainly phenological responses, i.e., stages of plant development (Hatfield et al., 2014).

The stresses of low and high temperatures have a detrimental effect on plants. Temperature increases can lead to yield reductions of between 2.5% and 10% for several agronomic species throughout the 21st century (Hatfield et al., 2011). Decreasing or increasing temperatures above specific thresholds during the growing season triggers cold and heat stress for various crops, limiting their growth and metabolism and leading to significant crop losses (Wahid et al., 2007). In addition, it damages cell division and amyloplast replication in cereals, resulting in reduced crop yields.

Crops sensitive to the photoperiod would interact with temperature, causing a disruption of phenological development (Hatfield & Prueger, 2015). In general, extremely high temperatures during the breeding phase will affect pollen viability, fertilization, and kernel or fruit formation

(Hatfield et al., 2011). Chronic exposures to extreme temperatures are detrimental during the reproductive stages of development and reduce yield potential during pollination, the initial stage of grain or fruit set (Hatfield & Prueger, 2015).

1.2. Modeling the effects of weather and climate extremes on crops

According to Feng et al. (2018), the standard methods of exploring climate-yield relationships are crop (simulation) modeling and statistical analysis. Crop models that take into account - in addition to crop - multiple climatic factors, , soil, and management parameters, can promote a better understanding of the crop response to climate (Rosenzweig et al., 2014). The main advantage of using a crop model is that it completely characterizes the cropping system. If crop models are accurately calibrated with observed data, they can be applied to simulate possible interactions with management to better cope with predicted climate changes (Liu et al., 2009). However, most crop models perform poorly in dealing with the effects of extreme weather events on crop growth and development (Moriondo et al., 2011). This poor performance is related to the simplified description of specific processes, leading to inaccurate results. In addition, crop models require several years of experimental data to train and calibrate in the local environment (Chen et al., 2010), and recalibration should be performed when used in other regions.

Due to these limitations in crop models, some linear statistical models, such as multiple linear regression, have been widely used to characterize the relationship between yields and climate variables (Tebaldi & Lobell, 2008). Linear models are easy to handle and inexpensive to calculate (Feng et al., 2018). With the increasing availability and improving quality of observed data, linear models generally perform well. Innes et al. (2015) suggested a superior performance of linear models compared to crop models to identify climate-yield relationships. However, linear models are unable to detect nonlinear relationships or identify factors with multicollinearity. Multicollinearity occurs when two or more explanatory variables in a multiple regression model are strongly linearly correlated, resulting in incorrect coefficient estimates in the multiple regression (Siegel, 2016). Over the past decades, ML algorithms have gradually gained wide attention and are applied in many fields such as agriculture. ML methods can assess the nonlinear and hierarchical relationships between predictors and response using a whole learning approach (Shalev-Shwartz & Ben-David, 2014). They generally work well in prediction compared to the traditional linear regression model. Everingham et al. (2015) reported that ML is superior to temporary and time-consuming approaches.

Chapter II: Data and Methods

2.1. Description of the study area

Ensuring food security is the first step in achieving sustainable development for any nation (Soumare, 2004). Food security has always been at the forefront of the development objectives of successive governments in Mali (CSA-Mali, 2017). Agricultural production in Mali, as in most countries of the Sahel, takes place under complex and uncertain natural conditions (Soumaré et al., 2020). The first limiting factor for the development of crops is the climate and particularly agroclimatic extremes. The study area offers the most potential for Mali and is the leading agricultural region in Mali. The crop cultures concerned by this study are of paramount importance because of their multiple uses for the Malian population and the Malian government. All parts of these crops are used either for human food (grains) and livestock (grains, stems, leaves), the construction of sheds, fences, fuel (stems), canning bag (fiber of cotton stems), crafts (stems), and exports (cotton fibers). While the weight of climatic constraints on agricultural development was sufficiently emphasized in Sudano-Sahelian Africa, studies have rarely considered the impact of climatic extremes on crop yields (Roudier et al., 2011). These are the different factors and arguments that motivated the choice of this area in addition to the availability of agricultural and climatic data. Delimited by the borders of Guinea and Côte d'Ivoire to the south, Burkina Faso to the east, and the Niger River to the north, the Mali South zone is the breadbasket of the country. It feeds nearly a third of the Malian population (Warner, 2018).

This study is carried out in the south of Mali, in the Sudanese and Sudano-Guinean climatic zones. The study area is located between 10° and 14.5° North latitudes and 4° and 11° West longitudes (Soumaré et al., 2020). This zone covers an area of 134,518 km^2. In the study area, 3,346 administrative villages are located. Approximately 4,108,849 inhabitants live here. The villages are spread over 244 municipalities (CMDT, 2018). The study area covers the entire administrative region of Sikasso and part of the administrative regions of Kayes, Koulikoro, and Ségou (Figure 1).

The climate in Mali is influenced by the seasonal mobility of the continental, dry and warm air mass coming from the Sahara (Harmattan) and the humid air mass coming from the Saint Helena anticyclone to the south. west (maritime trade winds) (Blanchard, 2011). Their convergence gives rise to the Intertropical Convergence Zone (ITCZ), which moves from south to north following the sun's movement. In January, the ITCZ is in the south of the area, the Harmattan dominates, and the climate is dry and hot. The maritime trade winds become loaded with moisture during the hot season as they pass over the ocean and the equatorial zone. Their reinforcement moves the front towards the north of the country. During the rainy season, the humid air mass gives rise to rains when it cools by elevation (Soumaré, 2008).

The rainy season begins when the humid air masses associated with the monsoon reach the country. It runs from March to October. In this period, the ITCZ is in the north of the country towards 17° N latitude. The dry season lasts from November to March. Its duration is variable, like that of the rainy season, depending on the latitudinal position. It varies from 12 months in the Saharan part, which the humid air masses rarely reach, to 6 months in the pre-Guinean zone (Soumare, 2004). According to the FAO classification in 2011, Mali-South is located in the dry to sub-humid agroclimatic zone framed by a semi-arid fringe in the north and humid to sub-humid in the south (Soumare, 2004). It lies between isohyets of 600 mm in the north and can exceed 1,200 mm in the south (Figure 1).

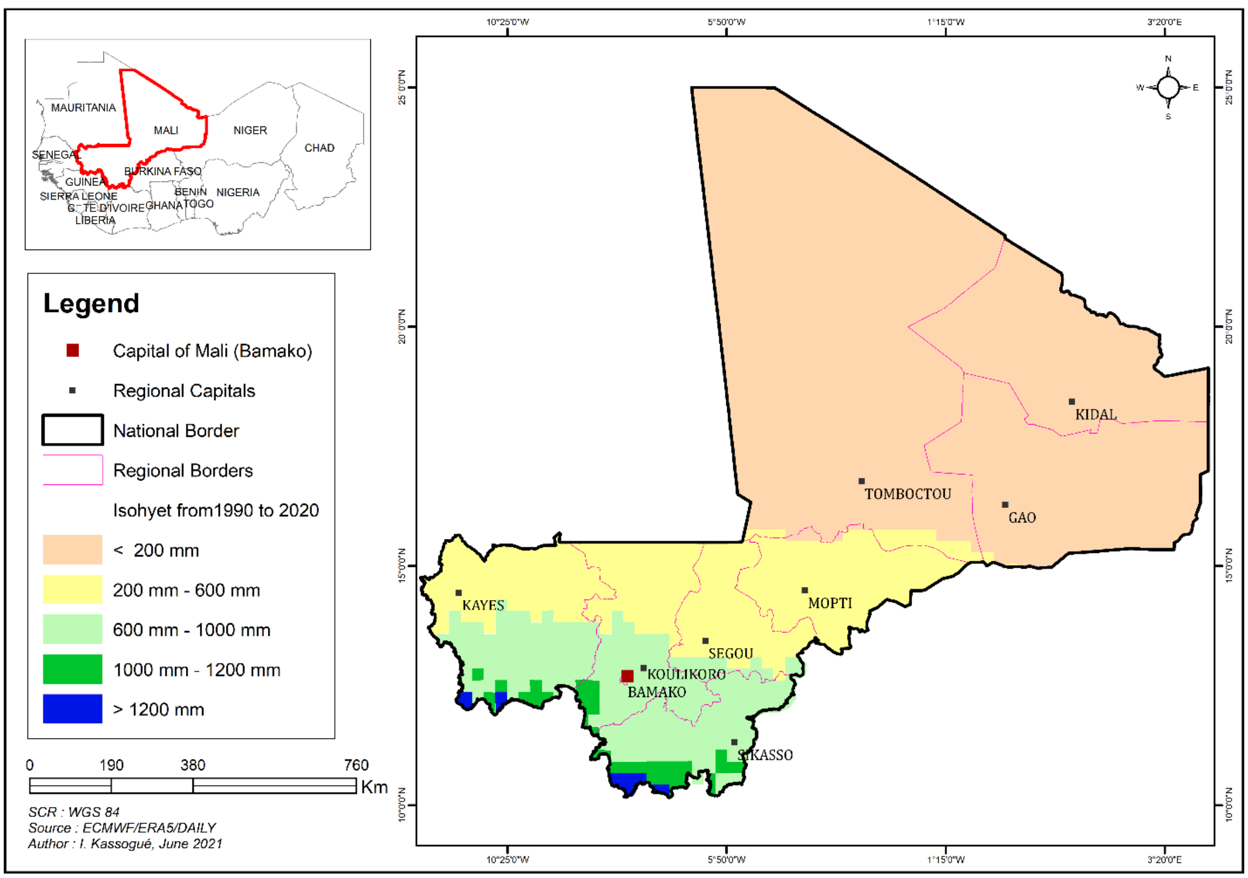

Figure 1: Distribution of the long-term (1990-2020) average annual rainfall (mm) over Mali.

Source: European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5 is the fifth generation ECMWF reanalysis for the global climate and weather for the past 4 to 7 decades

According to Blanchard (2011), the southern zone of Mali belongs to the Sudanese domain comprising two agroecological zones. The Sahel-Sudanian zone is characterized by the cultivation of millet, sorghum, cotton, and legumes. The natural vegetation is composed of wooded to shrub savannah. In the Sudano-Guinean zone (900 to 1,200 mm/year of rainfall during 80 days), the cultivation of cotton, maize, and legumes is accompanied by scattered distributed of dry cereals. The vegetation is characterized by a mosaic of shrub or tree savannas, and woodland forests (Figure 2), as illustrated by the ecological zones of Mali.

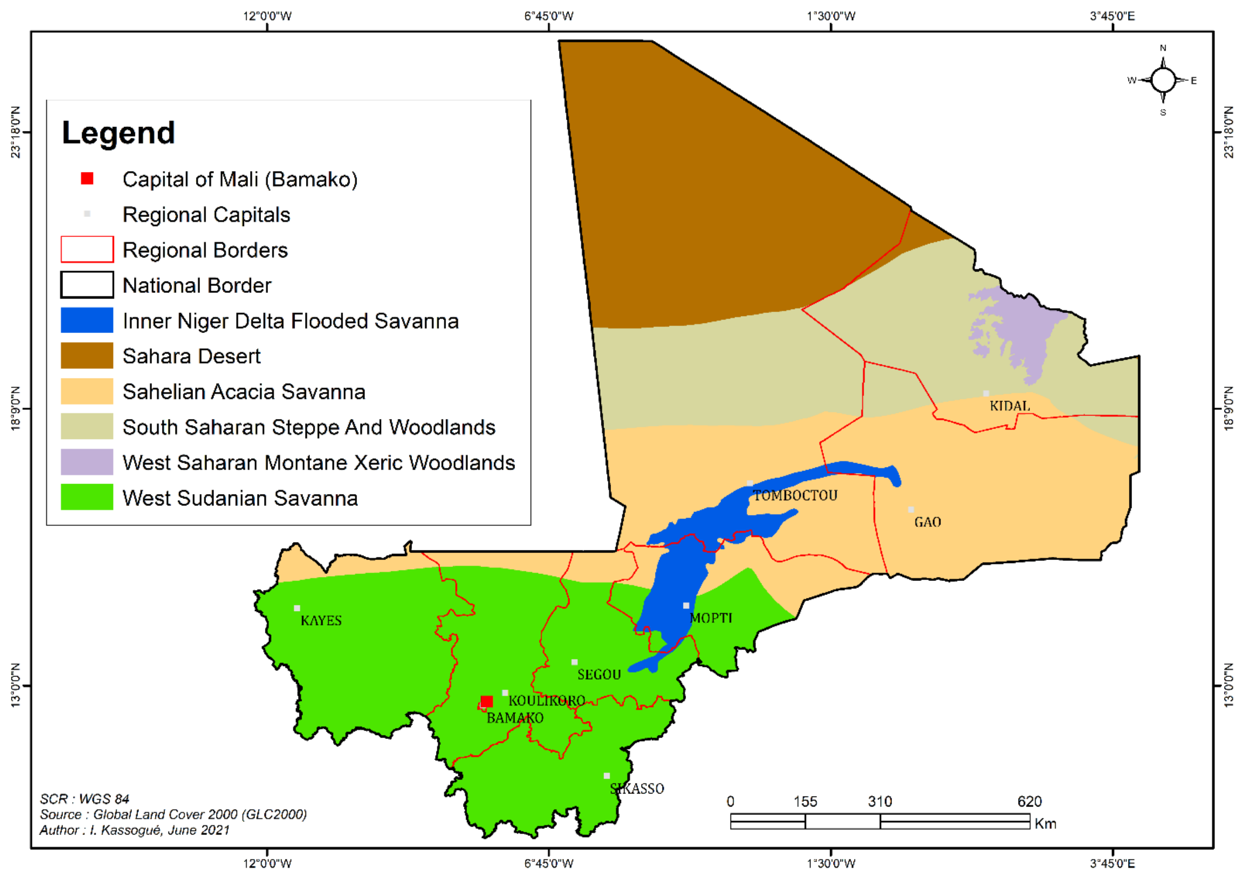

Figure 2: The Global Land Cover map of Mali; Source: Global Land Cover 2000

The relatively flat terrain of these zones (Sahelo-Sudanese and Sudano-Guinean) in Mali was affected by vertical movements, which gave rise to the Mandingo plateau (located between the Senegal river near Medina then bends towards the West until the vicinity of Farabana (around 14 12'lat. N.) 21km from "La Faleme"; from this point the boundary runs south-east, for more than 200km, to Dabia (12 45 'lat N; 13 30' long W.) (Chudeau, 1921)), and the Dogon plateau ((14 °

34 ’N), located in the Mopti region, is located between the central Niger delta to the west and the Seno plain to the south-east (Diallo, 2017)) (Keita, 2000). The sedimentary formations of the area are various. The Koutiala region is based on extensive, homogeneous, and thick formations of siliceous sandstone with rolled quartz seeds, characteristic of the Koutiala sandstones. To the south, the Sikasso region is based on fine-grained rounded quartz sandstones and ferruginous and clayey cement (Dakoure, 2011).

The soil formation results from the topography, lithology and climate action, time of evolution, and land use/cover and management). In this area of Mali, soils are formed from the underlying sandstones (with the exception of the western area in Mali-South). These soils result from the deep weathering of rocks under the action of the Quaternary period, characterized by an alternation of rainy and arid periods (Blanchard, 2011). The minerals of the basement rock have undergone profound reworking and were altered by mono-siallitization.

In the study area, the climate (heavy rain and heat) causes a separate migration of clays and iron hydroxides, which accumulate in separate layers (Dosso & Ruellan, 1993). Reduced vegetation protects the soil less, and soil erosion processes cause lateral transfer of material and large areas of colluvial deposits during the dry period of the season. Without insufficient drainage and a low slope, colluvium accumulates without being drawn into the hydrographic network (Traore, 2000). The soils of the Mali-south zone (Figure 3) correspond to the CPCS (Commission de Pédologie et de Cartographie des Sols) classification to tropical ferruginous soils (little or no leached and leached), to Lixisols according to the revised legend of the FAO (2015), or Alfisols according to the soil taxonomy (Keita, 2000).

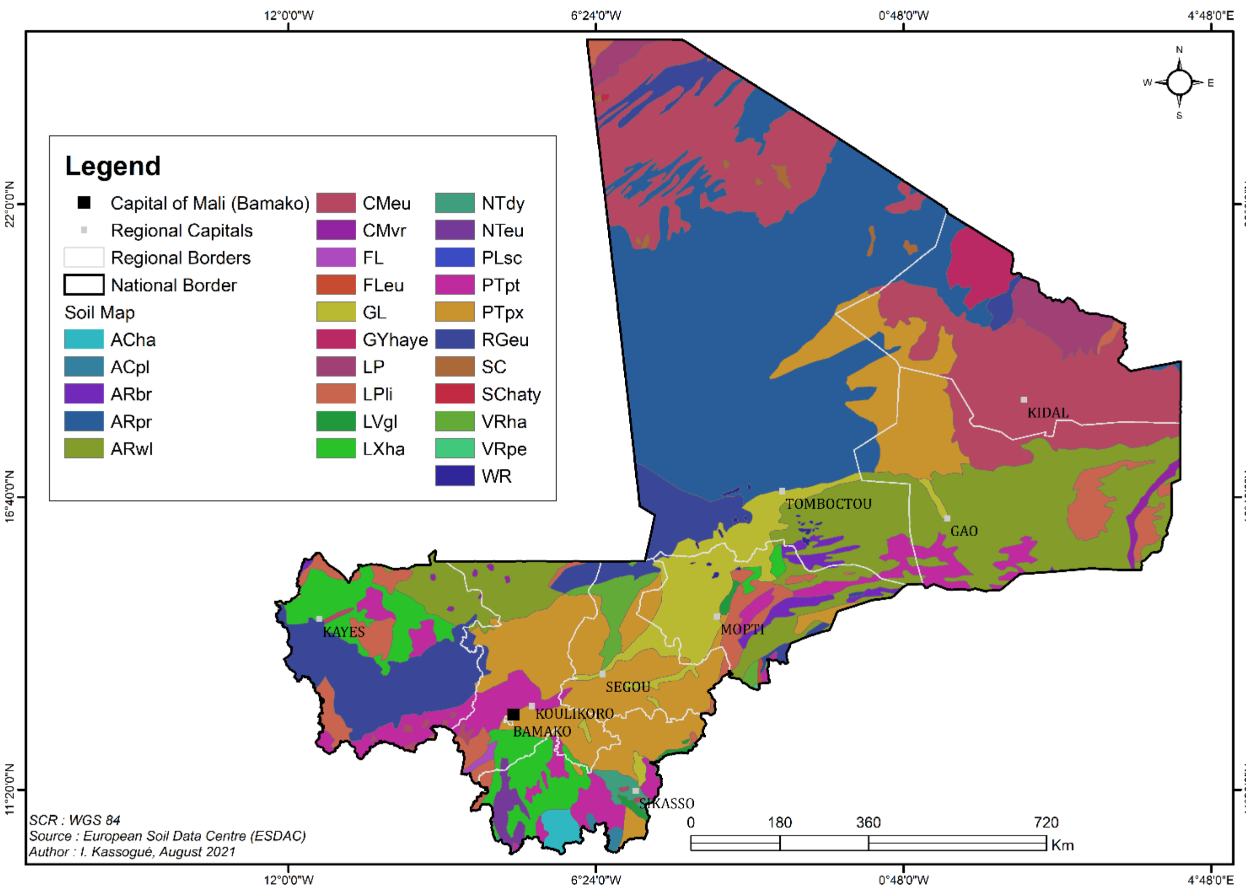

Figure 3: Pedological map of Mali.

Source: European Soil Data Centre (ESDAC).

*(**ACha** = Haplic Acrisols ; **ACpl**= Plinthic Acrisols; **ARbr**= Brunic Arenosols; **ARpr**= Protic Arenosols ; **ARwl**= Hypoluvic Arenosols; **CMeu**= Eutric Cambisols ; **CMvr**= Vetric Cambisols ; **FL**= Undifferentiated Fluvisols ; **FLeu** = Eutric Fluvisols ; **GL**= Undifferentiated Gleysol; **GYhaye**= Haplic Gypsisols; **LP**= Undifferentiated Leptosols ; **LPli**= Lithic Leptosols ; **LVgl**= **Gleyic** Luvisols; **LXha**= Haplic Luvisols; **NTdy**= Dystric Nitisols; **NTeu**= Eutric Nitisols; **PLsc**= Solodic Planosols; **PTpt**= Petric Plinthosols; **PTpx**= Pisoplinthic Plinthosols; **RGeu**= Eutric Regosols; **SC**= Undifferentiated Solonchak; **SChaty**= Haplic Solonchaks; **VRha**= Haplic Vertisols ; **VRpe**= Pellic Vertisols ; **WR**= Water body,* Source : Jones et al. (2013)*)*

2.2. Crop yield and Climatic Data sets

Agricultural data (cotton, maize, and millet) were collected from the different subsidiary (districts) over the last 27 years from 1990 to 2017 from the General Directorate of the Compagnie malienne pour le développement du textile (CMDT; www.cmdt-mali.net). Additional climatic datasets comprising including rainfall temperature and solar radiation were provided by the Competence Centre of the West African Science Service Center for Climate Change and Adapted Land Use (WASCAL; www.wascal.org) and Mali-Météo (www.malimeteo.ml). These data are from 1990 to 2017. Not having obtained temperature data (minimum and maximum) for some areas where yield data was collected, we had used the closest climate stations to fill this temperature data gap for analysis.

The different figures (4 to 12) below represent the correlation between the total surfaces and total production over a period of time for cotton, maize and millet in Fana, Kita and Koutiala respectively. This correlation does not stand for the causality, but only association between the two variable and the direction of variation of one compared to another (Chesneau, 2018). The plein line represent the regression line, which is to interpret the direction of the association, positive, negative or neutral. If the line is parallel to the x of axis, it means the neutral meaning that the two variables are not related in any way. Its shift clockwise or anticlockwise means the negative and positive association respectively.

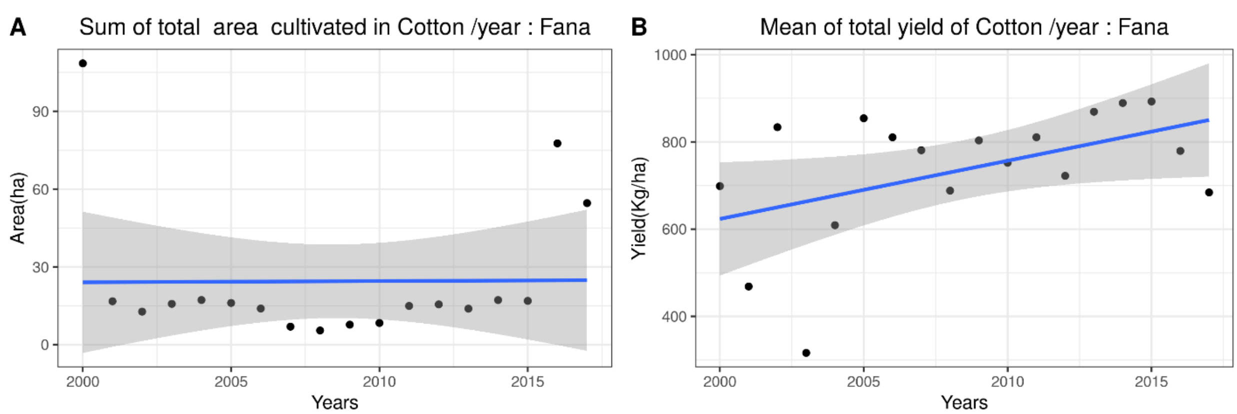

Figure 4: Historical characteristics of cotton yields (Kg/ha) over time and space (ha) of Fana district

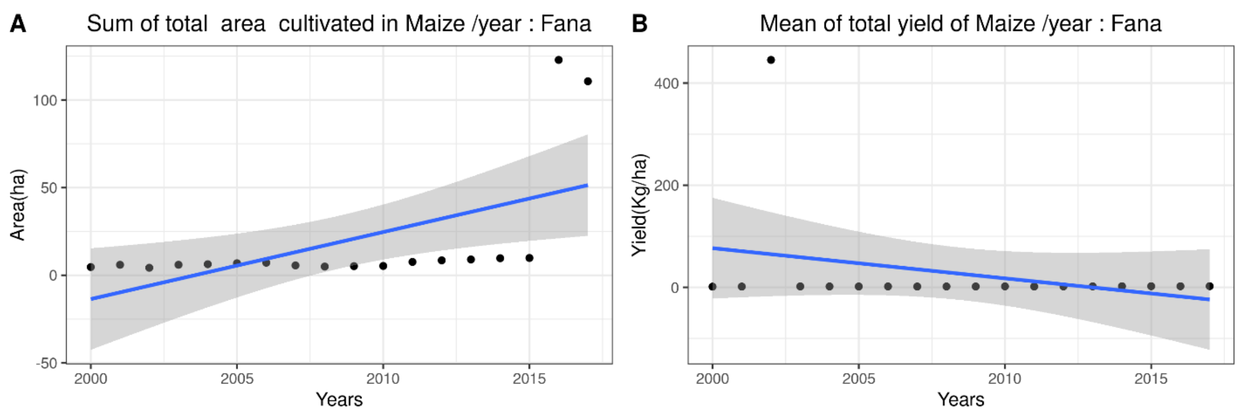

Figure 5: Historical characteristics of maize yields (Kg/ha) over time and space (ha) of Fana district

In Fana, the surfaces dedicated to cotton has been stable over the two decades while the total yield of the same crop has relatively been increased. This stability of surfaces of arable lands may be due to the urbanization and lack of surface for more arable land. This pressure on lands is because Fana is one of the city the closest to CMDT zones to the capital city of Mali, thus leading to occupation of lands for other activities such housing and business. The increase in total cotton yield may be explained by the improvement in cotton sector: research, extension services, motivation of farmers for the cash crop.

Both maize total surfaces and total yield decrease over the same period of time. This may be due to the abandonment of maize production for other crop that bring more incomes or requires less effort. At the same time the surface of millet as well as its total yield increase. The millet is less demanding in term of water, fertilization and production costs compared to maize, thus pushing more farmers to choose this crop over the maize. The millet is more appreciated than maize since it constitutes the most consumed cereals in this area. This cereal has also benefited more promotion as it tolerates the drawback of climate change than other cereals locally produced. Chauvin et al. have found the similar trend since 2012 between cash crops and staple foods in sub-Saharan Africa.

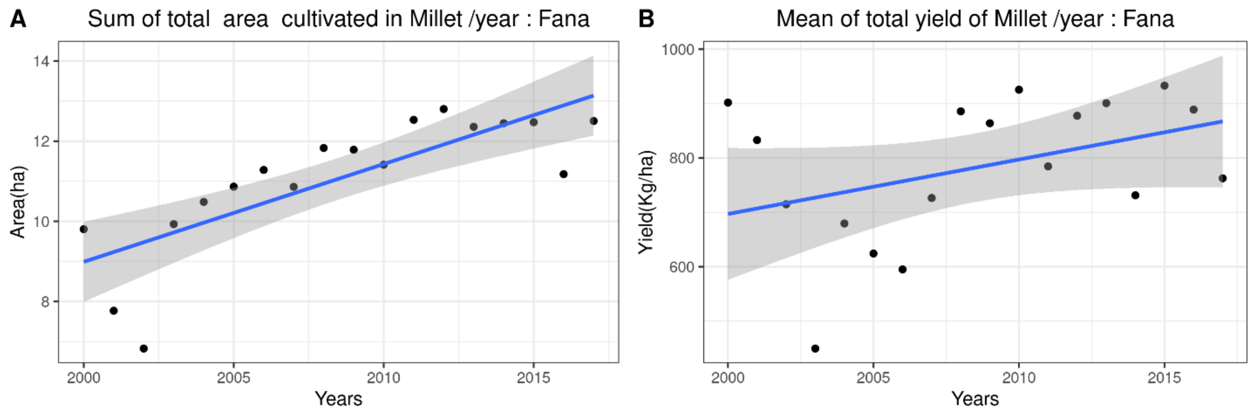

Figure 6: Historical characteristics of millet yields (Kg/ha) over time and space (ha) of Fana district

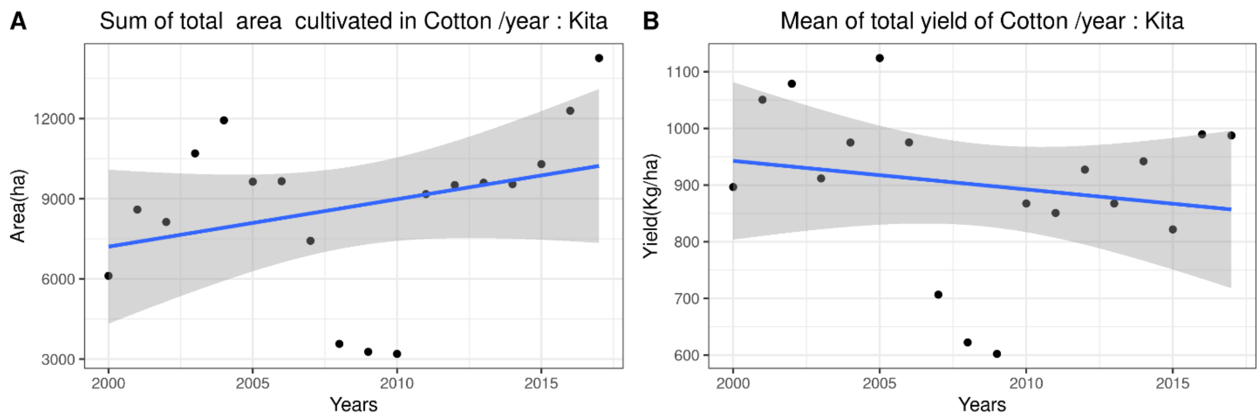

Figure 7: Historical characteristics of cotton yields (Kg/ha) over time and space (ha) of Kita district

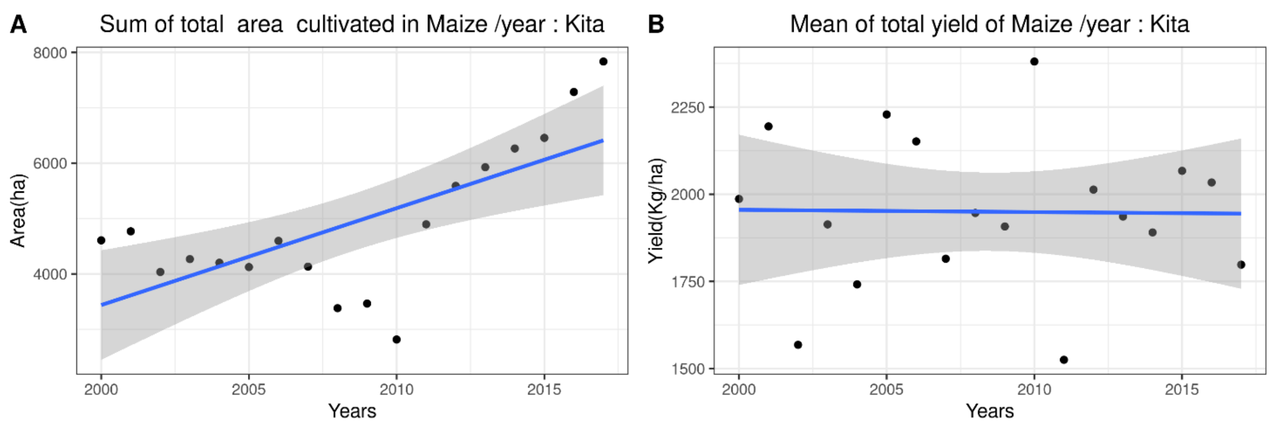

Figure 8: Historical characteristics of maize yields (Kg/ha) over time and space (ha) of Kita district

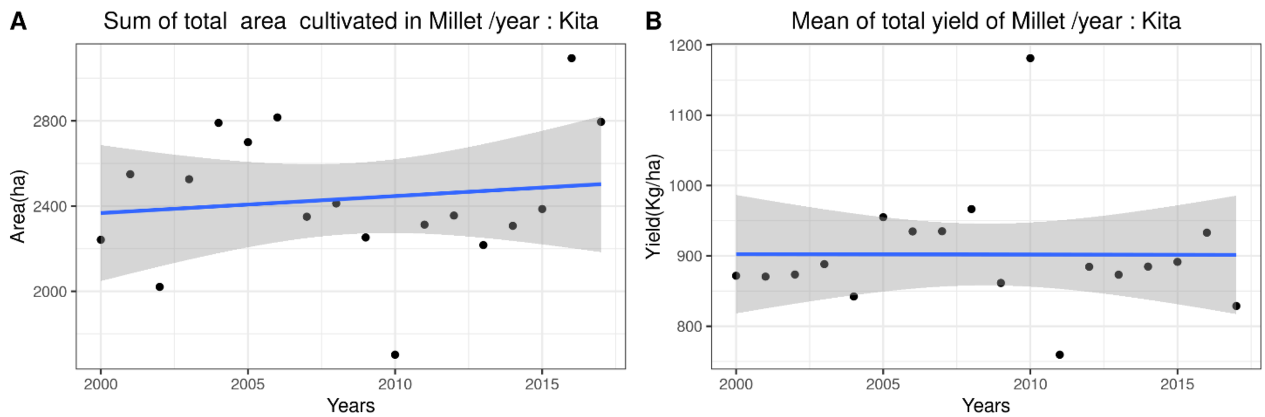

Figure 9: Historical characteristics of millet yields (Kg/ha) over time and space (ha) of Kita district

In Kita, the results of correlation analysis between the surfaces and total yields of the three crops are I contrast with the results observed in Fana. All three crops have decreasing yield along the considered time yet the cotton and maize have increased in surfaces. The extension of cotton and maize production may be imputable to the income benefit drawn from the cotton and promotion of maize as crop to ensure food security in the area. The letter crop is one of the most valued staple foods, thus taking over the millet though both, maize and millet are promoted. The decline in production in general may be caused by the pressure on agricultural lands since the same lands are always used for production. This lack of fallow and its negative subsequent results on agriculture in noticed by OECD & FAO (2016).

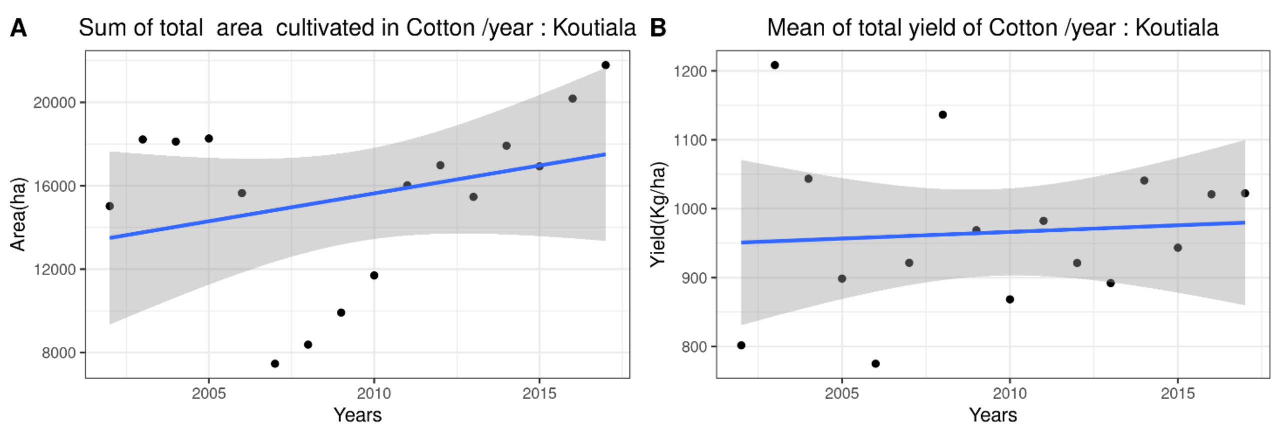

Figure 10: Historical characteristics of cotton yields (Kg/ha) over time and space (ha) of Koutiala district

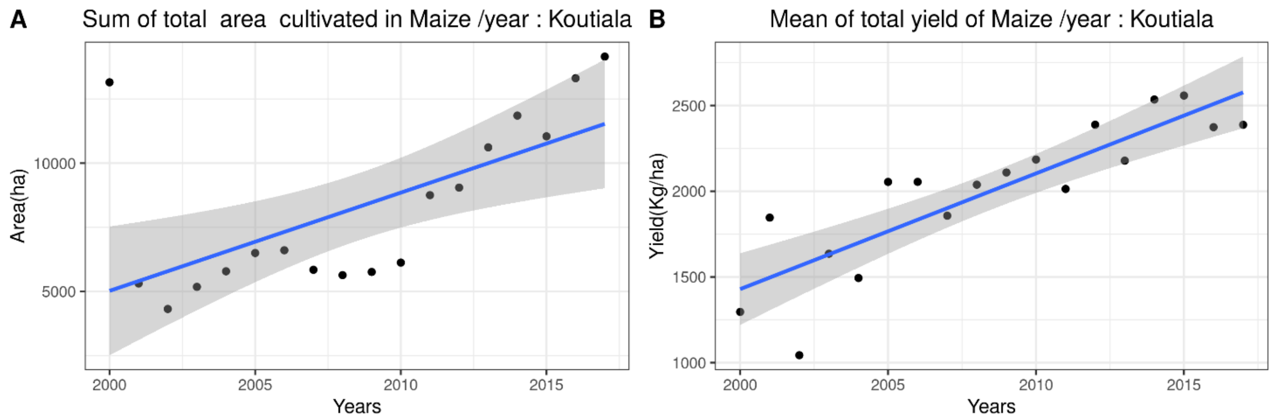

Figure 11: Historical characteristics of maize yields (Kg/ha) over time and space (ha) of Koutiala district

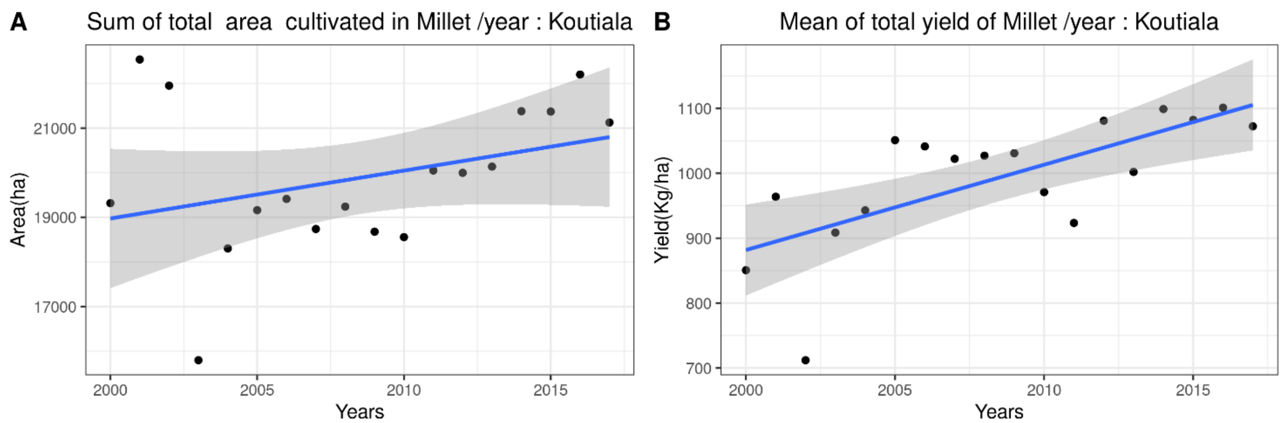

Figure 12: Historical characteristics of millet yields (Kg/ha) over time and space (ha) of Koutiala district

Regarding all crops, the case in Koutiala is an exception. The three crops have all increased in surfaces and total yields over the observation time. Nevertheless, the cotton has the smallest yield increase compared to the millet and maize. The latter has the biggest increase in the surfaces followed by the millet. This because Koutiala is the most industrialized zone after the capital city, Bamako. It also called as the "capital of white gold, the cotton" due to its experience and its place in national cotton production. Therefore, Koutiala has benefited all new techniques and technologies that are introduced in the country. Thanks to cotton, the other crops, millet and maize, benefit all needed inputs for crop production. It is also because of the diversification programs by CMDT, Government and other stakeholders in the area has always promoted cereal production. Though underfunded the agriculture (both cash crops and staple foods) has been improved thanks to allocation of high budget, extension services, infrastructures, technologies, laws, policies and business environment (FAO, 2021b).

Figure 20 represents the study area. It is located in the south of Mali. Initially, this study aimed to model all the main crops (i.e., cotton, maize, millet, and sorghum) produced in the study area. However, for reasons of availability of data for a long series, we had been obliged to use to consider the zones having data series between 1990 and 2017. As indicated in Figure 20, we had three types of data that come from: of each sector.

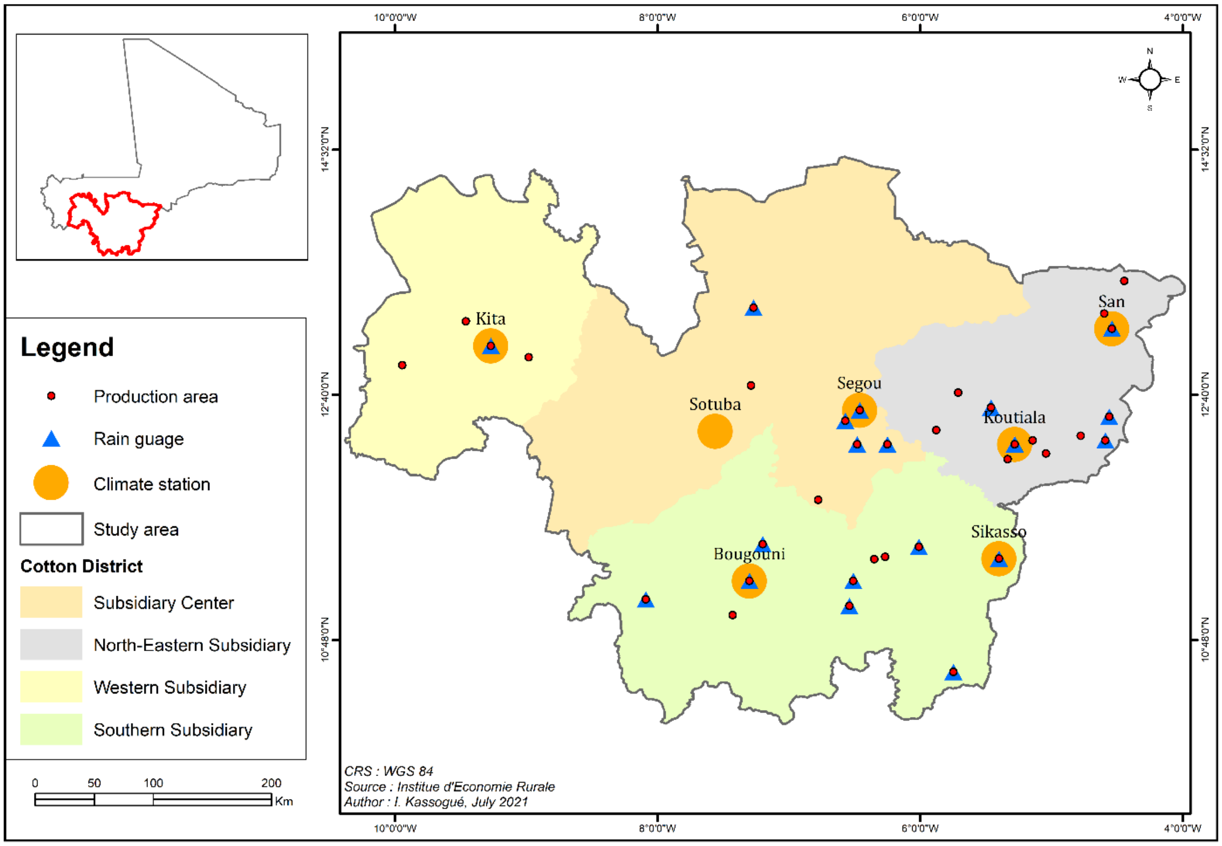

Figure 13: Location map of meteorological stations and agricultural data sources. Source: IER (Institut d'Economie Rural)

There are a total of 28 sectors which are alienated between 3 subsidiaries (districts) (subsidiary Center (Fana), Western subsidiary (Kita), North-eastern subsidiary (Koutiala)). of CMDT (Compagnie Malienne de Développement du Textile). At the level of the southern subsidiary (Sikasso), only the sectors of the Bougouni sub-subsidiary (called Bougouni cooridination). No sector of Sikasso subsidiary was concerned for lack of a series of data consistent with the period studied (1990 to 2017) for agricultural and climatic data. There was little area in which we could find all the data (agricultural, rainfall, temperatures, solar radiation).

A Python script with the NearestNeighbors and shapely.geometry packages has been developed to allow assignment of synoptic stations to the area without synoptic station data, at a distance of 100 km. NearestNeighbors arranges the unsupervised instruction of the nearest neighbors. It acts as a seamless interface with three distinct nearest neighbor algorithms: BallTree, KDTree, and a habit-based brute force algorithm in sklearn.metrics.pairwise (Goldberger et al., 2004). Shapely is a Python package for set-theoretic analysis and manipulation of planar features using (via Python's ctypes module) functions from the well-known and widely deployed GEOS library (Gillies, 2018). Since some agricultural production areas did not have climate data, use of data from stations near each agricultural production area was made to fill this climate data gap. This condition is motivated by the fact that OMM (2017) in its “WMO Guidelines for the Calculation of Climate Normals” says that It is possible to use a composite series of data obtained from a set of stations of this type for the calculation of climatic normals. The fundamental condition to be fulfilled is that the merged data set is homogeneous, either because the sites taken into account are sufficiently similar, or because the necessary adjustments have been made. These data consisted of temperatures (min and max) and solar radiation. All sectors had agricultural (yield) and rainfall data. These two have not been awarded.

2.3. Algorithmic computation of extreme Agroclimatic Indices

The indices that are the subject of this study are listed in Table 1. There are six of them. It is the calculation of these indices that allowed us to test and train the random forest (RF) models and logistic regression (LR) models. The Onset cropping season, end of cropping season are expressed in Julian day, the False onset of the cropping season in week number (Week of the year). The crop yield anomaly was defined according to a peasant perception which has the method of comparing the yield of the current season with the previous season. If the yield the current season is higher than that of the previous season, this means that the production is positive, in the contrary case the negative or deficit season. The Heat Wave number is expressed in number of times when the temperature threshold is exceeded during the season, if the number is less than or equal to 0, it is considered that there is no heat stress cropping season, and if it is greater than 0, it is considered that there is a Heat cropping season.

Table 1: Algorithmic computation of extreme Agroclimatic Indices

Indicators	Definition (mathematical expression)	Unit	Thresholds	Tag name	Variables	References
Onset of cropping season	**Planting date**: The date after May 1st when rainfall accumulated over 3 consecutive days is at least 20 mm and when no dry spell within the next 30 days exceeds 20 days.	Julian day	**Inter-seasonal OCS anomaly**: If OCS < 0 ➔ Late Onset If OCS > 0 ➔ Early Onset	OCS	Daily rainfall (mm)	Sivakumar (1988) modified by Salack et al. (2016)
End of the cropping season	**End date**: The date after September 1st when no rain occurs over a period of 20 consecutive days.	Julian day	**ECS anomaly**: If ECS < 0 ➔ Early cessation If ECS > 0 ➔ Late Cessation	ECS		Sivakumar (1988)
Length of the cropping season	Length of the cropping season is the result of subtracting the date of ECS in Julian day from the date of OCS in Julian day. $LCS = OCS - ECS$	Number (#)		LCS		
False onset of the cropping season	The false onset is the day after March 15th (before July 15th), when the first efficient rainfall (FER) is followed by a dry spell of at least 10 days (xDS). **Extraction algorithm** 1) between March 15th and July 15th, extract the date for which rainfall ≥ 9.75 mm (FER). and its corresponding week number. 2) From the date of the FER, extract the start date (STDATE) and the week number of the dry spell ≥ 10 days (xDS). If	Week number (Week-of-the-year)	FER = 9.75 mm/day xDS >= 10 days	FON	Daily rainfall (mm)	Koufanou (2019) ; Salack et al. 2020)

	STDATE is the FER date are in the same week or their week numbers differ by 1 or 2, then the FER date is a false onset. $FO = \begin{cases} 1\ if\ xDS - FER\ \leq 2;\quad then\ False\ Onset \\ 0\ if\ xDS - FER\ > 2; then\ \ No\ False\ Onset \end{cases}$					
Dry / Wet stress	**Inter-seasonal crop yield anomaly:** $\Delta X_i = X_{(i)} - X_{(i\text{-}1)}$ ΔX_i = first difference of Yield (X) at year i, $X_{(i)}$ = value of time series X at year $_i$ $X_{(i-1)}$ = value for the $(i-1)^{th}$ year	Not Available	**Inter-seasonal crop yield anomaly:** If $\Delta X \leq 1$ dry season ➔ Yield loss If $\Delta X \geq 0$ ➔ wet season		Daily rainfall (mm)	Katz and Glantz (1986); Lebel and Ali (2009) ; Salack et al. 2020)
Heat Wave Number	**1. Daily Heat Stress (DHS)** $DHS = \begin{cases} 0.0\ \text{ for } T_{mean}\ < T_{crit} \\ \dfrac{T_{day} - T_{crit}}{T_{lim} - T_{crit}}\ \text{ for }\ T_{crit}\ \leq\ T_{mean} < T_{lim} \\ 1.0\ \text{ for } T_{mean}\ \geq T_{lim} \end{cases}$ Where $\text{Tmean} = \dfrac{(\text{Tmax} + \text{Tmin})}{2}$ T_{crit} T_{lim} Cotton 30 °C 38 °C Maize 30 °C 35 °C Millet 35 °C 42 °C Sorghum 35 °C 40 °C **2. Seasonal Heat Wave Number (HWN)** $HWN_{season} = \sum_{ORS}^{ERS} DHS\ \ when\ DHS = 1$ HWN_{season} = cumulative number of DHS per season between Onset (ORS) and cessation (ERS) of the rainy season DHS = Daily Heat Stress only when $T_{mean} \geq T_{lim}$	Number (#)	**Inter-seasonal HWN anomaly**: If HWN ≤ 0 ➔ Little to no Heat Stress cropping season If HWN > 0 ➔ Heat stress cropping season	HWN	Daily minimum & maximum temperatures (ºC)	Teixeira et al. (2013); USAID (2014) USAID (2014); Fakhri Bazzaz et al. (1997)
Crop yield anomaly	**Inter-seasonal crop yield anomaly:** $\Delta X_i = X_{(i)} - X_{(i\text{-}1)}$ ΔX_i = first difference of Yield (X) at year i, $X_{(i)}$ = value of time series X at year $_i$ $X_{(i-1)}$ = value for the $(i-1)^{th}$ year	Kg	**Inter-seasonal crop yield anomaly:** If $\Delta X \leq$ -50 Kg ➔ Yield loss If $\Delta X \geq$ +50 Kg ➔ Yield gain I <= 1.5 wet season	CYA	Total Yield of individual crops (Cotton, Maize, Millet)	Own definition adapted based on field experiences

2.4. Logistic and Random Forest Multi-Linear Regression Models

2.4.1. Logistic Regression Model

Logistic regression analysis is a popular and widely used analysis similar to linear regression analysis, except that the result is dichotomous (Tatum et al., 2013). Logistic regression is one of the models of multivariate analysis. It measures the association between the occurrence of an event (qualitative explained variable or predictand) and the factors likely to influence it (explanatory variables or predictors) (El Sanharawi & Naudet, 2013). The contribution of each independent variable (Table 1) was assessed through relative importance measures calculated with Python. It is used to explain the relationship between a continuous dependent variable (crop yield) and two or more independent variables (agroclimatic extremes Table 1) (Feng et., 2018).

Crop yield is affected by climatic and non-climatic factors. To separately assess the effect of climate on yield variation, an increase in yield by factors other than climate should be excluded. In this study, a first difference method (Lobell & Asner, 2003), was used. This method is easy to implement and can minimize the influence of non-climatic factors, helping to explain climate-crop yield relationships. All-time series of crop yields anomaly (CYA) were calculated using the first differences approach using Equation 1:

Equation 1: De – trending method

$$\Delta X_{(t)} = X_{(t)} - X_{(t-1)}, \qquad t = 1990, \qquad 1991, \dots, \qquad 2017$$

where $\Delta X_{(t)}$ denotes the first difference of X at year t, $\Delta X_{(t)}$ denotes the values of times series X at year t and $X_{(t-1)}$ is the value for the (t-1)th year.

Logistic regression (LR) model explains the relationship between one continuous dependent variable and two or more independent variables. A variety of statistical techniques were used to develop crop-climate relationship, forecast models. The most common method is multiple linear regression, and random forest (RF) methods. However, when the predictand (crop yield) is "yes" or "no", binary logistic regression (BLR) often is employed (Shafer & Fuelberg, 2008). The BLR, also known as the binomial logit model, is an estimation technique for equations with dummy dependent variables that avoids the unboundedness problem of the linear probability model by using a variant of the cumulative logistic function (Wooldridge et al., 1997).

Yield loss is defined when CYA is less or equal to -50 kg (Table 1). Hence, we define binary predictands according to the nonlinear equation (Shafer & Fuelberg, 2008; Lawson, 2018; Salack et al., 2020).

Equation 2: Logit link function

$$\ln\left(\frac{P_i}{1\ P_i}\right) = b_0 + b_1 + x_1 + \cdots + b_k x_k$$

Equation 3: Relationship between the predictors

$$P_i = \frac{\exp\ (b_0 + b_1 + x_1 + \cdots + b_k x_k)}{1 + \exp\ (b_0 + b_1 + x_1 + \cdots + b_k x_k)}$$

The where *Pi* is the predicted probability resulting from the set of candidate predictors (x1, x2, . . ., xk), rainfall, rainy days, relative humidity, daily temperature range, wind speed, and solar radiation. The quantity on the left of equation (2) is the logit link function, which relates the log of the odds ratio (p/1-p) to a linear combination of predictors (Shafer & Fuelberg, 2008; Rajeevan et al., 2012). The parameters (b0, b1, . . . , bk) are estimated by maximizing a log-likelihood function using iterative methods (Wilks, 2006). Equation (3) guarantees that the probabilities are bounded within the interval (0, 1), and the relationship between the predictors and the response variable follow Bernoulli distributions (Lawson, 2018, Salack et al., 2020).

2.4.2. Random Forest Model

Random forest (RF) model is an ensemble learning algorithm based on classification and regression trees (Feng et al., 2018). Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently, and with the same distribution for all trees in the forest. The generalization error for forests converges to a limit as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correction between them (Breiman, 2001).

The RF consists of many independent trees, where each tree is generated by bootstrap samples, leaving a number of the aggregate sample for validation. Each tree split is determined using a random subset of the predictors for each node. The final result is the average of the results of all trees. RF can explore nonlinear and hierarchical relationships between predictors and response. It has been applied in agricultural studies, showing high precision and an ability to model complex interactions between variables (Feng et al., 2018). However, this method behaves like a "black box" since individual trees cannot be examined separately, and it does not calculate regression coefficients or confidence intervals (Cutler et al., 2007).

The Random Forest consists of a set of independent decision trees. Due to the double random selection, each tree has a fragmented view of the problem:

- ✓ Randomly select and replace observations (rows in the database) (called tree bagging),
- ✓ Random selection of variables (database columns) (called feature sampling).

All these independent decision trees come together. The random forest prediction on unknown data is the average of all trees (or votes, in the case of classification problems).

The basic idea of this algorithm is quite intuitive.

Random forest works on the same principle: Random Forest uses a few simple estimators (with lower individual quality) instead of a complex estimator that can do it all. Each estimator has a fragmented view of the problem. Then put all of these estimates together to get a big picture of the problem. It is the combination of all these estimators that makes the prediction very efficient.

2.4.3 Model Output Statistics

According to Roebber (2017), it is possible (in an approach conceptually similar to the Taylor diagram) to exploit the geometric relationship between four measures of dichotomous forecast performance: probability of detection (PoD), false alarm ratio, or its opposite, the success ratio (SR), bias and critical success index (CSI; also known as the threat score).

For good forecasts, PoD, SR, bias, and CSI approach unity, such that a perfect estimate lies in the upper right of the diagram. Deviations in a particular direction will indicate relative differences in PoD and SR, and consequently bias and CSI. Immediate visualization of differences in performance is thus obtained. Optimal increases in accuracy are obtained by moving at 45 degrees, that is, by maintaining unbiased forecasts through simultaneous increases in detection and reductions in false positives. Skill is assessed by plotting the forecast quality measure relative to a reference forecast (climatology, persistence, or any other desired baseline).

The influence of sampling variability is estimated using a form of resampling with replacement bootstrapping from the verification data. The 95th percentile range for SR and PoD is plotted as "cross-hairs" about the verification point, and simultaneously displayed variation in bias and CSI. Several new samples of the same size as the original can be created using the sampling frequencies of observed and forecast "yes" and "no" entries (i.e., the marginal frequencies), and the 25th and 95th accuracy measures are computed from these "climatological" samples to generate the 95th percentile range.

The mean absolute prediction error (MAE), the mean squared error (MSE), the coefficient of determination (R^2) and the Lin concordance correlation coefficient (LCCC) (Lin, 1989; Nickerson, 1997) . These indices were calculated according to the following formula. The LME measures the average prediction bias, and the MSE represents the sample standard deviation of the differences between the predicted and observed values. The LCCC represents the extent to which predicted and observed values follow the 45° line across the origin. The forecasts become more precise as MAE and RMSE approach 0 and R^2 and LCCC approach 1.

Equation 4: Mean Squared Error

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i|$$

Equation 5: Root-Mean-Squared Error

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_I)^2}$$

Equation 6: Coefficient of Determination (R^2)

$$R^2 = \left(\frac{\sum_{i=1}^{n}(O_i - \overline{O})\ (P_i - \overline{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \overline{O})^2}\ \sqrt{\sum_{i=1}^{n}(P_i - \overline{P})^2}}\right)^2$$

Where ***Pi*** and ***Oi*** are the predicted and observed values, respectively; ***O*** and ***P*** express the mean of the observed and predicted values, respectively; ***n*** is the number of samples; ***σP*** and ***σO*** are the variances of the predicted and observed values; and ***r*** is the Pearson correlation coefficient between the predicted and observed values.

Chapter III: Results and Discussion

3.1. Identification and quantification of agroclimatic extremes of strong impacts in Mali

A pairwise correlation analysis was performed on each crop for each village. The results of this analysis show variation from one village to another, but mainly two variables were correlated. The Seasonal rainfall was correlated with many variables such as Onset of cropping season, End of cropping season, and temperature. The Seasonal rainfall and Onset of cropping season were correlated at more than 50% level in village performing maize such as Bougouni, Dogo, Garalo (Bougouni sub-subsidiary), Fangaso, Karangaso, Kimparana (Koutiala subsidiary) while, Koumantou and Yanfolila (Bougouni sub-subsidiary). It was also correlated to End of cropping season in most of the villages dominated by diversification of maize, cotton, and millet. This the case in Djidian, Kita, Kokofata and Sebekoro. The Seasonal rainfall was correlated to Tp in maize production in only Kolondieba.

Using Logistic Regression analysis, every variable having a correlation of more than 40% was dropped since this value is considered to be high enough to affect yields us suggested by Chalil (2020). In this line, the maize production was positively affected the variables Crop yield anomaly (CYA) and Effective First Rain at 10% significant level in Bougouni. An additional unit of Effective First Rain will increase the CYA by a coefficient of 0.00551 in this village. The CYA was negatively affected by Onset of cropping season and Effective First Rain at 5% level and by the Tp at 1% in Karangana. A unit of increase in these three variables will decrease the CYA in Karangana by 0.01110, 0.0205 and 0.0184, respectively. The Effective First Rain that affected the CYA positively at 5% level in Koumantou by a coefficient of 0.00583.

There is no significant value for millet (Table 2) for its resistance to different climatic factors and its adaptability. This is confirmed by Vintrou (2012) and Kouressy et al. (2008), who explain that millet is well suited to this area because it is resistant and has a short growth cycle of around 90 days. Vintrou (2012) explains that the timing of a specific phenological stage of millet can vary from year to year due to variations in the start of the planting season.

The opposite results between maize villages and the villages of crop diversification can be explained by the role of cash crop and staple food. In the monoculture, maize benefits more attention from farmers, while in crop diversification zones more attention is given to cotton, which is the main cash crop in the area (FAO, 2017). Additionally, cotton production has historically required more attention in research related to climate adaptation, tolerance and resistance done by CMDT, the highest institution in charge of cotton in Mali, and its partners (Camara, 2016; Soumare & Havard, 2017). According to Robichaud (2009), cotton is produced for commercial purposes and regenerates significant financial resources. Furthermore, the acquisition of agricultural subsidies is conditional on cotton production (Vintrou, 2012). This motivates the attention given to its production and the reservation of arable land to produce this crop.

Surprisingly, neither in cotton production (Table 2) nor millet production (Table 3), the CYA was affected by any climate factors. The resistance of millet can explain this to the climate variabilities. As well as the cotton is concerned, the reason for the absence of CYA is that more research has been vulgarized in the cotton production system for more resistance and tolerance of this cash crop to climate change effects. Additionally, cotton production benefits more infrastructures and extension services since it is the highest contributor to the Malian agricultural products export (Bagayoko, 2014; IER et al., 1999). Many studies have seen crop diversification as the best way to counter the adverse effects of climate change by improving environmental aspects and socio-economic benefits (Kiani et al., 2021).

Table 2: Correlation coefficients between cotton yield and individual agroclimatic parameters.

Note: *, ** and *** means 10%, 5% and 1% respectively.

Stations	**OCS**	**ECS**	**FO**	**HW**	**Rtot**	**LCS**
Djidian	-0.11	-0.10	0.04	-0.06	-0.07	0.00
Kita	-0.18	0.02	-0.01	-0.13	-0.27	0.04
Kokofata	-0.19	0.00	0.10	-0.21	-0.21	0.15
Sebekoro	-0.19	-0.12	-0.11	-0.13	-0.17	0.02

Table 3: Correlation coefficients between maize yield and individual agroclimatic parameters

Stations	OCS	ECS	FO	HW	Rtot	LCS
Bla	0.04	0.20*	0.28*	0.14	0.10	0.47*
Bougouni	0.03	0.20*	0.12	0.08	0.06	-0.05
Djidian	-0.20*	0.24*	-0.10	-0.08	-0.03	0.01
Dogo	0.01	0.05	0.08	0.33*	0.27*	-0.15
Fangasso	0.03	0.23*	0.27*	0.18	0.05	0.09
Garalo	-0.04	0.00	-0.17	0.06	0.10	0.00
Karangana	-0.07	0.02	-0.03	0.24*	0.27*	0.09
Kimparana	-0.03	0.21*	0.24*	0.16	0.15	0.05
Kita	-0.08	0.19	-0.01	-0.10	-0.24	0.10
Kokofata	-0.04	0.05	-0.13	0.08	0.07	0.07
Kolondieba	-0.01	-0.05	-0.24*	0.41*	0.13	0.05
Konseguela	0.08	0.11	0.20*	0.21	0.01	0.01
Koumantou	0.04	0.07	-0.10	0.37*	0.19	0.09
Koutiala	0.02	0.15	0.29*	0.10	0.17	0.07
Molobala	-0.04	0.08	0.26*	0.24*	0.07	0.00
Mpessoba	0.08	0.13	0.27*	-0.04	0.15	0.02
Sebekoro	-0.21*	0.16	-0.17*	-0.02	-0.17	0.10
Yanfolila	0.08	0.09	-0.22*	0.26*	-0.12	0.08
Yorosso	-0.07	-0.08	0.04	0.04	0.19	0.04
Zebala	0.02	0.29*	0.30*	0.09	0.23*	0.11

Table 4: Correlation coefficients between millet yield and individual agroclimatic parameters

Stations	OCS	ECS	FO	HW	Rtot	LCS
Djidian	-0.12	-0.04	-0.11	0.19	0.09	-0.06
Kita	-0.14	-0.03	-0.07	-0.01	-0.06	-0.04
Kokofata	-0.08	0.14	-0.15	0.07	0.07	-0.03
Sebekoro	-0.09	0.06	-0.24	0.18	0.05	-0.07

3.2. Prediction of yield loss resulting from the synchronous occurrence of agroclimatic extremes

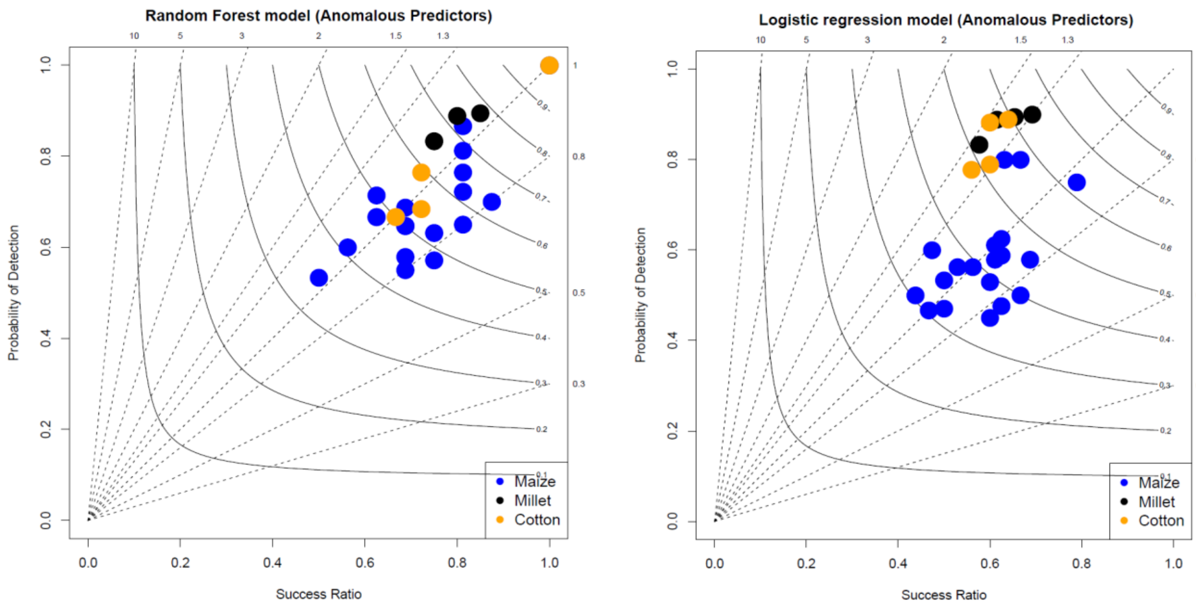

Figure 14: Random Forest model (Anomalous Predictors) and Logistic regression model (Anomalous Predictors)

In this performance diagram Figure 21a, the PoD values for cotton are almost perfect as three of the four points are between 0.62 and 0.77, and one point exactly on the perfect value which is (1). Its success Ratio (1-FAR), which represents the success rate is also important because three of the four points are between 0.66 and 0.75, and one point out of 1 (the perfect value) as its PoD. Its critical success index is of an important value and is between 0.50 and 1. For the Bias, two of the four points are exactly on the angle of 45°, one on 44.9 degrees is (0.91) and the last 45° either a little above 45° or (1.1).

As for millet in the same figure 21a of the performance diagram, the PoD for millet is estimated between 0.80 to 0.89, of which three of the four points are located between 0.85 and 0.89, including two points on the same value (0.89). Its success Ratio is between 0.77 and 0.89, with three of the four points between 0.83 and 0.8. The critical success index is between 0.65 and 0.78. All values are moderately above 45°, which should not be considered an overestimate.

The maize with sixteen points scattered between 0.50 and 0.83 for the PoD, the majority of which are between 0.55 and 0.70. Its Success Ration is between 0.50 and 0.90, including a large number of values between 0.70 and 0.80. The points are also scattered between 0.50 and 0.85 for its critical success index, including a large number of values located between 0.50 and 0.70. Most of the points are around the 45° angle, including two values on the 45° angle.

Figure 21b represents the logistic regression (LR) model (AP) graph on which the cotton values are distributed for the PoD between 0.75 and 87. They constitute two groups of the same number, including two values between 0.85 and 0.88 and the other two between 0.78 and 0.80. Its Success Ratio is between 0.55 and 0.65 of which all the values are almost grouped at the same level. Its critical success index is between 0.49 and 0.60, of which three of four points are located between 0.52 and 0.60. All values are grouped together above the appreciation angle (45°) in the overestimation zone, unlike the cotton values distribution of the random forest (RF) model (AP).

The values of the millet are distributed for the PoD between 0.81 to 0.88, which means a good probability of detection, on the four values of the millet the three are very close and aligned between 0.85 and 0.88. Its success ratio is between 0.55 and 0.70, the three out of four values of which are between 0.65 and 0.70. Its critical success index is between 0.52 and 0.65. The logistic regression (LR) model's millet values (AP) are a little off the 45° angle and are located in the overestimated area, unlike the random forest (RF) model (AP).

The maize has a bit scattered values but the majority are around the 45° angle with four values on the 45° angle line and two slightly below the appreciation zone. Its PoD is between 0.42 and 0.79 with the majority between 0.45 and 0.65 which is a little above the random forest (RF) model (AP). Its Success Ration is between 0.45 and 0.79, most of which are between 0.45 and 0.65, which is also lower than the random forest (RF) model (AP). The critical success index is between 0.30 and 0.63, the largest number of values being between 0.30 and 0.45.

In summary, the two models are all good and allow meaningful exploitation of the geometric relationship between four performance measures. Although both models are good, the random forest (RF) model (AP) performs better than the logistic regression (LR) model (AP) model, given the proximity of each performance measure closer to the best score for each of the performance measures and the distribution around the angle 45° from the different values.

3.3. Evaluation the performance of ML algorithms to diagnose the impact of agroclimatic extremes on crop yields in Mali

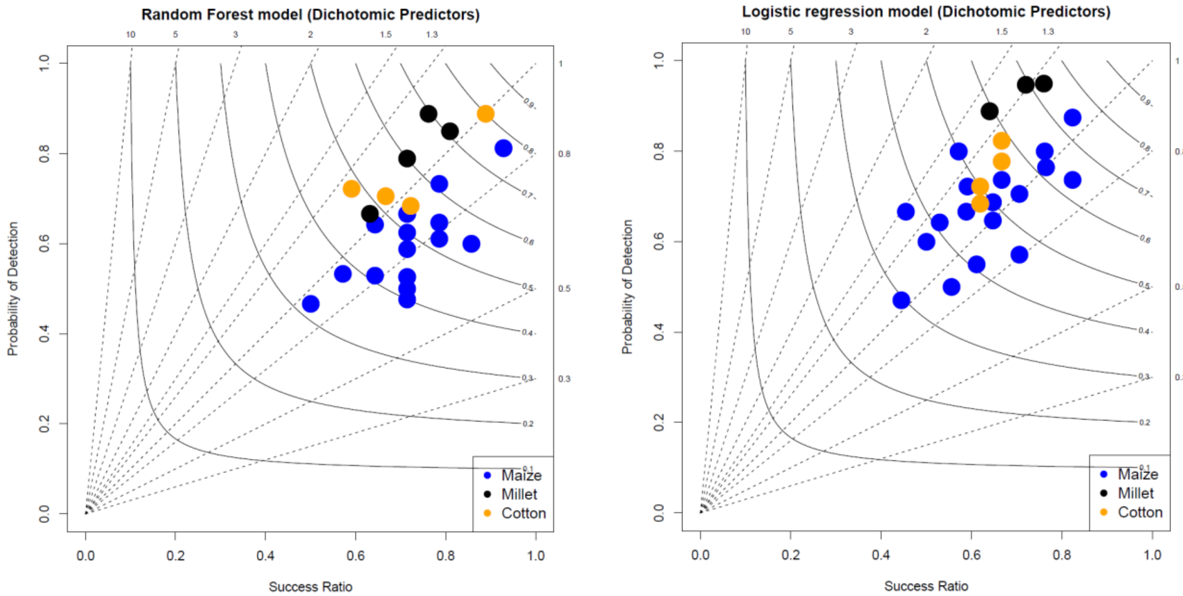

Figure 15: Random Forest model (Dichotomic Predictors) and Logistic regression model (Dichotomic Predictors)

Figure 15a shows the performance graph of the Random Forest model (Dichotomic Predictors). In this graph, cotton has its PoD values between 0.65 and 0.90 of which three of the four values are between 0.65 and 0.72, and the last value is the last with a value of 0.85. This indicates that the cotton's PoD is efficient because its value is greater than 0.50 and close to 1 (the best score). The values of its Success Ratio are between 0.59 and 0.85, of which three of the four values are distributed between 0.59 and 0.70 and the last value with a value of 0.85. Its Success Ratio is also good, and close to the value of the best score (1). Its critical success index is between 0.49 and 0.54 for the three values out of four and the last one a little closer to 1 or 0.8. Like the other performance measures (PoD and SR), the cotton critical success index is also important and is closer to the best score which 1.

The PoD of millet for Figure 15a, the values are scattered and are between 0.65 and 0.85 of which three of the four values are distributed over the values 0.77, 0.81 and 0.85, this distribution of the different values all have values greater than 0.5 then the PoD can be considered good because it is close to the best score 1. For its Success Ratio, its values are distributed between 0.63 and 0.78, including three of the four values between 0.71 and 0.78. This too can be considered good performance given the arrangement of the values of the different values beyond 0.50. The values of its critical success index are between 0.48 and 0.70, of which two of the four values are placed on 0.7 and one on 0.60. All four values are located more or less above the 45° angle but three of the four values are very close to the 45° angle but all are in the area where the performance is considered good.

For the maize, the performance measure of the PoD shows a distribution of values between 0.43 and 0.81 including a significant distribution of values between 0.43 and 0.65. Its Success Ratio values are divided into three groups, giving a group of values located between 0.50 and 0.60 and the second group of points numbering 7 out of 16 all out of 0.70 and the last group between 0.75 and 0.90. As for their positioning with respect to the 45° angle, they constitute two groups, most of which are close to the 45° angle, one value on the 45° angle, and four values in the area considered to be slightly underestimated.

In figure 22b, we have the random forest (RF) model (DP), where the cotton values are distributed between 0.65 and 0.82 of PoD, the four values of which are all located in this interval without a significant difference, this also explains a high probability of detection but does not reach the PoD of the random forest (RF) model (DP) which is between 0.65 and 0.90. The values of its Success Ratio are distributed between 0.63 and 0.70, like its PoD, its values almost form a cluster around these values. These SR values are also slightly lower than those of the random forest (RF) model (DP). The cotton critical success index for the logistic regression (LR) model (DP) is between 0.49 and 0.59, there is no great distance between the four values, still maintain their almost cluster formation but still slightly lower than the values of the random forest (RF) model values (DP). All values are slightly above the 45° angle but remain in the area deemed to be good performance.

Regarding millet, the values of its PoD located between 0.85 and 0.92 of which two of the four values are superimposed because they have the same values. These values are higher than those of the random forest (RF) model (DP) for millet. The values of his Success Ratio are distributed between 0.63 and 0.75 and three of the four values are in the range of 0.70 and 0.75. For the SR the two models are almost the same with a slight strong SR from the random forest (RF) model (DP). The values of the critical success index are between 0.49 and 0.59 and the greatest number of values are between 0.69 and 0.72. The values are distributed above the 45° angle and three of the four points are outside the assessment zone.

As for the maize, the PoD values are distributed between 0.42 and 0.85, most of which are between 0.57 and 0.77. The value of his SR points is also distributed between 0.45 and 0.80, a large number of values are between 0.55 and 0.75. The distribution of the value of the critical success index values is between 0.30 and 0.75 and most of the values are between 0.40 and 0.65. The values are near what all lie around the 45° angle.

The results obtained showed significant regression, Success Ratio and Probability of Detection PoD's between the observed yields and the impacts of agroclimatic extremes. In view of these results, it can be argued that increasing agroclimatic extremes are expected to trigger yield declines, and the associated impacts will likely lead to production losses and contribute to food insecurity and economic losses affecting production systems agricultural (Wu et al., 2015). This can be confirmed by the comments of Salehnia et al. (2020) and Bazzaz & Sombroek (1997) who report that extreme variability in agroclimatic extremes can impact yield either by lengthening the effective growing season in the case of a low, and in the opposite case of a low probably reducing the length of the effective growing season for the temperature case. And often even, impact agricultural productivity USAID, (2017) and devastate all production. As for the FAO (2016), an increase in these impacts would make it almost impossible for an adequate adaptation by the agricultural sectors in many places and would lead to drastic drops in productivity. It also emerges from Nassourou et al., (2018) research that the early end of the rains is cited as the most critical risk by 33% of farmers in western Niger and prolonged dry spells (20%).

Conclusion

Regarding climate variability, the exposure of agricultural yields to agroclimatic extremes contribute to declining production and productivity which can trigger risks of food insecurity and motivate immigration and unemployment. The LCS and OCS were the most correlated indices. Predicting the occurrence of these agroclimatic extremes have the advantage of identifying suitable agricultural inputs and avoiding certain risks such as false onset, identifying the optimal sowing periods, and anticipating sowing in the fields. swamps. And choose photosensitive cultures.

The introduction of machine learning has dramatically improved the accuracy of diagnoses of the relationship between agroclimatic extremes and yield, overcome the shortcomings of the linear model in processing correlated predictors, revealed new information on the different effects of similar climatic factors on crop yields. In addition, the comparison between machine learning and the linear model ensured the robustness of our results. Our results showed that LCS and OCS were dominant factors affecting yield. Overall, the variability of crop yields in the study area was mainly caused by dry season, while a wet season during growing seasons did not have a noticeable effect as they did not were not noticed frequently. The results of this study show that agroclimatic extremes impact crop yields on the basis of both. The performance of all models is good and very close, but random forest (RF) model (AP) and logistic regression (LR) model (DP) models are preferred. In addition, these forecasting models require a large amount of data for a more efficient assessment of performance quality. This is a relevant result as agriculture in Mali is heavily dependent on rain and extreme climatic factors can negatively affect crop yields which can lead to food insecurity.

Recommendations

Crop yield is well affected by climatic and non-climatic factors. In these studies, to separately assess the effect of climate on yield variation, an increase in yield by factors other than climate was excluded.

Therefore, it is necessary to extend the series of crops, other climatic factors (like WSRI) and other non-climatic ones like soil, fertilization… to better understand, the level of impact of each factor on the yield;

Data sharing between students and national structures via WACAL should be strongly encouraged and strengthened.

Bibliography

AGRHYMET. (2010a). *Bulletin mensuel d'information du Centre Régional AGRHYMET Septembre 2010*.

AGRHYMET. (2010b). Le Sahel face aux changements climatiques : Enjeux pour un dévelopement durable. In *Numéro spécial* (Issue Numéro spéciale). http://www.cilss.bf/fondsitalie/download/down/specialChC.pdf

Alhassane, A., Salack, S., Ly, M., Lona, I., Traoré, S. B., & Sarr, B. (2013). Évolution des risques agroclimatiques associés aux tendances récentes du régime pluviométrique en Afrique de l'Ouest soudano-sahélienne. *Science et Changements Planetaires - Secheresse*, *24*(4), 282–293. https://doi.org/10.1684/sec.2013.0400

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). *Crop evapotranspiration - Guidelines for computing crop water requirements*. http://www.fao.org/3/x0490e/x0490e04.htm#chapter 1 introduction to evapotranspiration

Ati, O. F., Stigter, C. J., & Oladipo, E. O. (2002). A comparison of methods to determine the onset of the growing season in Northern Nigeria. *International Journal of Climatology*, *22*(6), 731–742. https://doi.org/10.1002/joc.712

Bagayoko, K. (2014). *L ' importance et l ' avenir du coton en Afrique de l ' Ouest : cas du Mali To cite this version : HAL Id : tel-00959321 dans l ' École Doctorale de Sciences Economiques DU COTON EN AFRIQUE DE L ' OUEST : CAS DU MALI*.

Baştanlar, Y., & Özuysal, M. (2014). Introduction to Machine Learning. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 1107, pp. 105–128). https://doi.org/10.1007/978-1-62703-748-8_7

Bazzaz, F., & Sombroek, W. (1997). *Changements du climat et production agricole*. http://www.fao.org/3/w5183f/w5183f08.htm

Bengio, Y. (2009). Learning Deep Architectures for AI. *Foundations and Trends® in Machine Learning*, *2*(1), 1–127. https://doi.org/10.1561/2200000006

Blanchard, M. (2011). *Gestion de la fertilité des sols et rôle du troupeau dans les systèmes coton-céréales-élevage au Mali Sud*. UNIVERSITE PARIS-EST, CRETEIL VAL DE MARNE.

Breiman, L. (2001). *Random Forests*. https://link.springer.com/article/10.1023/A:1010933404324

Camara, M. (2016). *Atouts et limites de la filière coton au Mali*.

Chalil, K. (2020). *Statistical Methods for Development Correlation*. *III*(April), 1–21.

Chen, C., Wang, E., & Yu, Q. (2010). Modelling the effects of climate variability and water management on crop water productivity and water balance in the North China Plain. *Agricultural Water Management*, *97*(8), 1175–1184. https://doi.org/10.1016/j.agwat.2008.11.012

Chesneau, C. (2018). *Ajustement d'un nuage de points*.

Chudeau, R. (1921). Le plateau mandingue. *Annales de Géographie*, 362–373.

https://www.jstor.org/stable/23864648

CMDT. (2018). *Zones cotonnières*. https://www.cmdt-mali.net/index.php/activites/activites-agronomiques/zones-cotonieres.html

CSA-Mali. (2017). *Politique Nationale de Sécurité Alimentaire et Nutritionnelle (PolNSAN) tOME 3*.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783–2792. https://doi.org/10.1890/07-0539.1

Dakoure, D. (2011). *De la bordure sud-est du bassin sédimentaire de Taoudéni (Burkina Faso - Mali) - essai de modélisation*. Université Paris VI - Pierre et Marie Curie.

Diallo, D. (2017). Chapitre 36. Création de champs cultivés et gestion de l'eau et de la fertilité des sols sur le plateau Dogon (Mali). In É. Roose (Ed.), *Restauration de la productivité des sols tropicaux et méditerranéens*. IRD Éditions. https://doi.org/10.4000/books.irdeditions.24108

Dimitriadis, S., & Goumopoulos, C. (2008). Applying Machine Learning to Extract New Knowledge in Precision Agriculture Applications. *2008 Panhellenic Conference on Informatics*, 100–104. https://doi.org/10.1109/PCI.2008.30

Dodd, D. E. S., & Jolliffe, I. T. (2001). Early detection of the start of the wet season in semiarid tropical climates of western Africa. *International Journal of Climatology*, *21*(10), 1251–1262. https://doi.org/10.1002/joc.640

Dosso, M., & Ruellan, A. (1993). La couverture pédologique : histoire de sa découverte ; actualité de son exploration (Soil cover : an history of its discovery, present aspects of its exploration). *Bulletin de l'Association de Géographes Français*, *70*(2), 77–85. https://doi.org/10.3406/bagf.1993.1667

Ed, O., Ym, A., Mn, G., & Tukura, E. (2013). Effect of Rainfall Variability on Maize Yield in Gassol. *Journal of Environmental Protection*, *4*(May), 881–887.

El Sanharawi, M., & Naudet, F. (2013). Comprendre la régression logistique. *Journal Français d'Ophtalmologie*, *36*(8), 710–715. https://doi.org/10.1016/j.jfo.2013.05.008

Everingham, Y., Sexton, J., & Robson, A. (2015). A statistical approach for identifying important climatic influences on sugarcane yields. *Conference: Australian Society of Sugarcane TechnologistsAt: Bundaberg, Qld, Australia*. https://www.cabdirect.org/cabdirect/abstract/20163376293

Fahad, S., Bajwa, A. A., Nazir, U., Anjum, S. A., Farooq, A., Zohaib, A., Sadia, S., Nasim, W., Adkins, S., Saud, S., Ihsan, M. Z., Alharby, H., Wu, C., Wang, D., & Huang, J. (2017). Crop production under drought and heat stress: Plant responses and management options. *Frontiers in Plant Science*, *8*(June), 1–16. https://doi.org/10.3389/fpls.2017.01147

FAO. (1977). *Crop water requirements. Document de la FAO sur l'irrigation et le drainage n° 24, par Doorenbos J et WO Pruitt.*

FAO. (1986). Early Agrometeorological Crop Yield Assessment (Fao Plant Production &

Protection Paper). *FAO Plant Production and Protection Paper (FAO)*, *73*, 158. https://agris.fao.org/agris-search/search.do?recordID=XF876803988

FAO. (2002). *Eau et agriculture*. http://www.fao.org/3/Y3918F/y3918f00.htm#TopOfPage

FAO. (2016a). *Adapter l'agriculture au changement climatique*. www.fao.org/climate-change

FAO. (2016b). *La situation de l'alimentation et de l'agriculture : Changements climatiques, agriculture et sécurité alimentaire*.

FAO. (2016c). The state of food and agriculture - Climate Change, Agriculture and Food Security. In *FAO*.

FAO. (2017). *Mali Country fact sheet*. *3*(July), 1–6.

FAO. (2021a). The impact of disasters and crises on agriculture and food security: 2021. In *The impact of disasters and crises on agriculture and food security: 2021*. https://doi.org/10.4060/cb3673en

FAO. (2021b). Public expenditure on food and agriculture in sub-Saharan Africa. In *Public expenditure on food and agriculture in sub-Saharan Africa*. FAO. https://doi.org/10.4060/cb4492en

Feng, P., Wang, B., Liu, D. L., Xing, H., Ji, F., Macadam, I., Ruan, H., & Yu, Q. (2018). Impacts of rainfall extremes on wheat yield in semi-arid cropping systems in eastern Australia. *Climatic Change*, *147*(3–4), 555–569. https://doi.org/10.1007/s10584-018-2170-x

Fortier, J.-F., & AquaPortail. (2021). *Déficit hydrique*. https://www.aquaportail.com/definition-13134-deficit-hydrique.html

Gillies, S. (2018). *Shapely Documentation*.

Goldberger, J., Roweis, S., Geoff Hinton, & Salakhutdinov, R. (2004). *Neighbourhood Components Analysis*. https://papers.nips.cc/paper/2004/hash/42fe880812925e520249e808937738d2-Abstract.html

Hatfield, Boote, K. J., Kimball, B. A., Ziska, L. H., Izaurralde, R. C., Ort, D., Thomson, A. M., & Wolfe, D. (2011). Climate Impacts on Agriculture: Implications for Crop Production. In *Agronomy Journal* (Vol. 103, Issue 2). https://doi.org/10.2134/agronj2010.0303

Hatfield, J., Takle, G., Grotjahn, R., Holden, P., Izaurralde, R. C., Mader, T., Marshall, E., & Liverman, D. (2014). Climate Change Impacts in the United States. In *Climate Change Impacts in the United States: The Third National Climate Assessment* (pp. 150–174). http://nca2014.globalchange.gov/system/files_force/downloads/low/NCA3_Full_Report_06_Agriculture_LowRes.pdf?download=1

Hatfield, & Prueger. (2015). Temperature extremes: Effect on plant growth and development. *Weather and Climate Extremes*, *10*, 4–10. https://doi.org/10.1016/j.wace.2015.08.001

IER, IDS, & IIED. (1999). *Sustainability amidst diversity : Options for rural households in Mali*.

Indoria, A. K., Sharma, K. L., & Reddy, K. S. (2020). Hydraulic properties of soil under warming climate. In *Climate Change and Soil Interactions*. LTD. https://doi.org/10.1016/b978-0-12-818032-7.00018-7

Innes, P. J., Tan, D. K. Y., Van Ogtrop, F., & Amthor, J. S. (2015). Effects of high-temperature episodes on wheat yields in New South Wales, Australia. *Agricultural and Forest Meteorology*, *208*, 95–107. https://doi.org/10.1016/j.agrformet.2015.03.018

INRA. (2006). *Sécheresse et agriculture Sécheresse et agriculture Réduire la vulnérabilité de l'agriculture*.

Iortyom, E. T., Iorsamber, M. M., & Adelabu, O. A. (2017). The Effect of Onset and Cessation of Raining Season on Crops Yield in Lafia. *Journal of Human Ecology*, *59*(2–3), 117–122. https://doi.org/10.1080/09709274.2017.1379134

J-P. K. Tshiabukole. (2018). Evaluation de la sensibilité aux stress hydriques du maïs (Zea mays L.) Cultive dans la savane du Sud-Ouest de la RD Congo, cas de Mvuazi. In *Thèse*. https://hal.archives-ouvertes.fr/tel-02062632/

Jones, A., Breuning-Madsen, H., B., M., D., A., D., J., D., O., G., T., H., S., J., R., K., M., L. R., P., M., E., M., L., S., O., Thiombiano, L., Ranst, V., E., Y., … R, Z. (2013). *Soil Atlas of Africa*. https://doi.org/10.2788/52319

K. Warner, P. van de L. (2018). Climate Change PROFILE Climate Change. *Ministry of Foreign Affairs of the Netherlands*, *923*(July), 14. https://www.science.org/doi/10.1126/science.aan5360

Katz, R. W., & Glantz, M. H. (1986). Anatomy of a Rainfall Index. *Monthly Weather Review*, *114*(4), 764–771. https://doi.org/10.1175/1520-0493(1986)114<0764:AOARI>2.0.CO;2

Keita, B. (2000). *Les sols dominants du Mali*.

Kiani, A. K., Sardar, A., Khan, W. U., He, Y., Bilgic, A., Kuslu, Y., & Raja, M. A. Z. (2021). Role of agricultural diversification in improving resilience to climate change: An empirical analysis with gaussian paradigm. *Sustainability (Switzerland)*, *13*(17). https://doi.org/10.3390/su13179539

King, A. D., Klingaman, N. P., Alexander, L. V., Donat, M. G., Jourdain, N. C., & Maher, P. (2014). Extreme rainfall variability in Australia: Patterns, drivers, and predictability. *Journal of Climate*, *27*(15), 6035–6050. https://doi.org/10.1175/JCLI-D-13-00715.1

Koufanou, H. (2019). *Attribution et prévisibilité des faux départs de saisons agricoles au Burkina Faso*.

Kouressy, M., Traoré, S., Vaksmann, M., Grum, M., Maikano, I., Soumaré, M., Traoré, P. S., Bazile, D., Dingkuhn, M., & Sidibé, A. (2008). *Étude originale Caractérisation de l ' agrobiodiversité Adaptation des sorghos du Mali à la variabilité climatique*. *January*. https://doi.org/10.1684/agr.2008.0189

Lawson, K. N. (2018). *Thresholds for operational agro-climatic monitoring and early warning against high impact rainfall events in the Sudan-Sahel region , West Africa* .

Lebel, T., & Ali, A. (2009). Recent trends in the Central and Western Sahel rainfall regime (1990–

2007). *Journal of Hydrology*, *375*(1–2), 52–64. https://doi.org/10.1016/j.jhydrol.2008.11.030

Liakos, K., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine Learning in Agriculture: A Review. *Sensors*, *18*(8), 2674. https://doi.org/10.3390/s18082674

Lin, L. I.-K. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*, *45*(1), 255. https://doi.org/10.2307/2532051

Liu, Y., Wang, E., Yang, X., & Wang, J. (2009). Contributions of climatic and crop varietal changes to crop production in the North China Plain, since 1980s. *Global Change Biology*, *16*(8), 2287–2299. https://doi.org/10.1111/j.1365-2486.2009.02077.x

Lobell, D. B., & Asner, G. P. (2003). Climate and Management Contributions to Recent Trends in U.S. Agricultural Yields. *Science*, *299*(5609), 1032–1032. https://doi.org/10.1126/science.1078475

Loko, Y., Dansi, A., Agre, A., Akpa, N., Dossou-Aminon, I., Assogba, P., Dansi, M., Akpagana, K., & Sanni, A. (2013). Perceptions paysannes et impacts des changements climatiques sur la production et la diversité variétale de l'igname dans la zone aride du nord-ouest du Bénin. *International Journal of Biological and Chemical Sciences*, *7*(2), 672–695. https://doi.org/10.4314/ijbcs.v7i2.23

Luna, T., Wilkinson, K. M., & Kasten Dumroese, R. (2011). Seed germination and sowing options [Chapter 9]. *Tropical Nursery Manual - Seed Germination and Sowing Options*, 163–183. http://www.native

Maiga, O., Tounkara, M., Doumbia, S., & Sangho, H. (2019). Analyse de l'économie politique du Mali. In *Research Technical Assistance Center: Washington, DC, USA*.

Moriondo, M., Giannakopoulos, C., & Bindi, M. (2011). Climate change impact assessment: the role of climate extremes in crop yield simulation. *Climatic Change*, *104*(3–4), 679–701. https://doi.org/10.1007/s10584-010-9871-0

Mugalavai, E. M., Kipkorir, E. C., Raes, D., & Rao, M. S. (2008). Analysis of rainfall onset, cessation and length of growing season for western Kenya. *Agricultural and Forest Meteorology*, *148*(6–7), 1123–1135. https://doi.org/10.1016/j.agrformet.2008.02.013

Nassourou, L. M., Sarr, B., Alhassane, A., Traoré, S., & Abdourahamane, B. (2018). Perception et observation : les principaux risques agro-climatique de l'agriculture pluviale dans l'ouest du Niger. *VertigO*, *Volume 18 numéro 1*, 1–20. https://doi.org/10.4000/vertigo.20003

Ngoune Liliane, T., & Shelton Charles, M. (2020). Factors Affecting Yield of Crops. In *Agronomy - Climate Change and Food Security* (pp. 1–16). IntechOpen. https://doi.org/10.5772/intechopen.90672

Nickerson, C. A. E. (1997). A Note On "A Concordance Correlation Coefficient to Evaluate Reproducibility." *Biometrics*, *53*(4), 1503. https://doi.org/10.2307/2533516

Observatoire du Sahara et du Sahel. (2013). Vers un système d'alerte précoce à la sécheresse au maghreb. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9).

OECD, & FAO. (2016). *Agriculture in Sub-Saharan Africa: Prospects and challenges for the next* *decade* (Vol. 181, Issue November 1947, pp. 59–95). https://doi.org/10.1787/agr_outlook-2016-5-en

OMM. (2017). *Directives de l'OMM pour le calcul des normales climatiques*. https://library.wmo.int/opac/doc_num.php?explnum_id=4220

Ouedraogo, P. L. A. (2013). *Prédétermination des séquences sèches et intérêt de l'information climatique sur la production céréalière en zone Sahélienne*.

Pashiardis, S., & Michaelides, S. (2008). Implementation of the Standardized Precipitation Index (SPI) and the Reconnaissance Drought Index (RDI) for Regional Drought Assessment: A case study for Cyprus. *European Water*, *23*(November), 57–65.

Pennsylvania. (2018). *Pennsylvania Agriculture: A Look at the Economic Impacts and Future Trends*.

Pérez-Ramos, I. M., & Marañón, T. (2009). Effects of waterlogging on seed germination of three Mediterranean oak species: Ecological implications. *Acta Oecologica*, *35*(3), 422–428. https://doi.org/10.1016/j.actao.2009.01.007

Priori, S., Pellegrini, S., Vignozzi, N., & Costantini, E. A. C. (2021). Soil Physical-Hydrological Degradation in the Root-Zone of Tree Crops: Problems and Solutions. In *Agronomy* (Vol. 11, Issue 1). https://doi.org/10.3390/agronomy11010068

Rahman, I. M. M., & Hasegawa, H. (2012). Water Stress. In *Water Stress* (Issue May 2014). https://doi.org/10.5772/1419

Rajeevan, M., Unnikrishnan, C. K., Bhate, J., Niranjan Kumar, K., & Sreekala, P. P. (2012). Northeast monsoon over India: Variability and prediction. *Meteorological Applications*, *19*(2), 226–236. https://doi.org/10.1002/met.1322

Rawson, H. M., & Macpherson, H. G. (2000). *Irrigated wheat*. https://www.fao.org/3/x8234e/x8234e00.htm#Contents

Robichaud, A. (2009). *Réforme du secteur cotonnier dans le cadre des stratégies de réduction de la pauvreté au Mali : réflexions sur l'espace politique de décision*.

Roebber, P. (2017). *Forecast Verification methods Across Time and Space Scales : Performance Diagram*. https://www.cawcr.gov.au/projects/verification/Roebber/PerformanceDiagram.html

Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., Boote, K. J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T. A. M., Schmid, E., Stehfest, E., Yang, H., & Jones, J. W. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences*, *111*(9), 3268–3273. https://doi.org/10.1073/pnas.1222463110

Rosenzweig, C., Iglesius, A., Yang, X. B., Epstein, P. R., & Chivian, E. (2000). Climate change and extreme weather events - Implications for food production, plant diseases, and pests. *Annals of Operations Research*, *97*(1–4), 131–141. https://doi.org/http://dx.doi.org/10.1023/A:1015086831467

Roudier, P., Sultan, B., Quirion, P., & Berg, A. (2011). The impact of future climate change on West African crop yields: What does the recent literature say? *Global Environmental Change*, *21*(3), 1073–1083. https://doi.org/10.1016/j.gloenvcha.2011.04.007

Salack, S. (2016). *Global warming induced hybrid rainy seasons in the Sahel*. 1–15.

Salack, S., Hien, K., Lawson, N. K. Z., Saley, I. A., Paturel, J.-E., & Waongo, M. (2020). Prévisibilité des faux-départs de saison agricole au Sahel. In *Risques climatiques et agriculture en Afrique de l'Ouest* (Issue January, pp. 31–43). IRD Éditions. https://doi.org/10.4000/books.irdeditions.36114

Salack, S., Klein, C., Giannini, A., Sarr, B., Worou, O. N., Belko, N., Jan Bliefernicht, & Kunstman, H. (2016). Global warming induced hybrid rainy seasons in the Sahel. *Environmental Research Letters*, *11*(10), 104008. https://doi.org/10.1088/1748-9326/11/10/104008

Salack, S., Saley, I. A., Lawson, N. Z., Zabré, I., & Daku, E. K. (2018). Scales for rating heavy rainfall events in the West African Sahel. *Weather and Climate Extremes*, *21*(May), 36–42. https://doi.org/10.1016/j.wace.2018.05.004

Salack, S., Sarr, B., Sangare, S. K., Ly, M., Sanda, I. S., & Kunstmann, H. (2015). Crop-climate ensemble scenarios to improve risk assessment and resilience in the semi-arid regions of West Africa. *Climate Research*, *65*(September), 107–121. https://doi.org/10.3354/cr01282

Salehnia, N., Salehnia, N., Saradari Torshizi, A., & Kolsoumi, S. (2020). Rainfed wheat (Triticum aestivum L.) yield prediction using economical, meteorological, and drought indicators through pooled panel data and statistical downscaling. *Ecological Indicators*, *111*(May 2019), 105991. https://doi.org/10.1016/j.ecolind.2019.105991

Sarr, B., Atta, S., & Kafando, L. (2012). Review of climate indices used in agricultural index insurance systems in Africa. *Sécheresse*, *23*(4), 255–260. https://doi.org/10.1684/sec.2012.0348

Senay, G. (2004). Crop Water Requirement Satisfaction Index (WRSI) Model Description. *Journal of the Tennessee Medical Association*, *65*(9), 811–815.

Shafer, P. E., & Fuelberg, H. E. (2008). A perfect prognosis scheme for forecasting warm-season lightning over Florida. *Monthly Weather Review*, *136*(6), 1817–1846. https://doi.org/10.1175/2007MWR2222.1

Shakil Ahamed, A. T. M., Mahmood, N. T., Hossain, N., Kabir, M. T., Das, K., Rahman, F., & Rahman, R. M. (2015). Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh. *2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 1–6. https://doi.org/10.1109/SNPD.2015.7176185

Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding Machine Learning. In *Understanding Machine Learning: From Theory to Algorithms* (Vol. 9781107057). Cambridge University Press. https://doi.org/10.1017/CBO9781107298019

Siegel, A. F. (2016). Multiple Regression. In *Practical Business Statistics* (7th ed., pp. 355–418).

Elsevier. https://doi.org/10.1016/B978-0-12-804250-2.00012-2

Sivakumar, M. V. K. (1988). Predicting rainy season potential from the onset of rains in Southern Sahelian and Sudanian climatic zones of West Africa. *Agricultural and Forest Meteorology*, *42*(4), 295–305. https://doi.org/10.1016/0168-1923(88)90039-1

Skendžić, S., Zovko, M., Živković, I. P., Lešić, V., & Lemić, D. (2021). The Impact of Climate Change on Agricultural Insect Pests. *Insects*, *12*(5), 440. https://doi.org/10.3390/insects12050440

Soumare, M. (2004). *Contribution à la prévision de l ' aire de diffusion de variétés de sorgho au Mali.*

Soumaré, M. (2008). Dynamique et durabilite des systemes agraires a base de coton au Mali. In *These de doctorat en geographie humaine, economique et regionale thème / universite de paris X Nanterre*.

Soumare, M., & Havard, M. (2017). *Les zones cotonnières africaines dynamiques et durabilité*.

Soumaré, M., Traoré, S., & Havard, M. (2020). Croissance démographique, sécurité alimentaire et accès à la santé et à l'éducation en zone cotonnière du Mali. *Cahiers Agricultures*, *29*, 40. https://doi.org/10.1051/cagri/2020036

Stewart, J. I. (1991). *Principles and performance of response farming*. CAB International. https://agris.fao.org/agris-search/search.do?recordID=GB9114206

Sun, Q., Miao, C., Hanel, M., Borthwick, A. G. L., Duan, Q., Ji, D., & Li, H. (2019). Global heat stress on health, wildfires, and agricultural crops under different levels of climate warming. *Environment International*, *128*(March), 125–136. https://doi.org/10.1016/j.envint.2019.04.025

Tatum, P., Talebreza, S., Ross, J. S., & Widera, E. (2013). Multivariable Methods : Multiple Logistic Regression Analysis. In *Medecine et Hygiene* (Vol. 62, Issue 2505). https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_multivariable/

Tebaldi, C., & Lobell, D. B. (2008). Towards probabilistic projections of climate change impacts on global crop yields. *Geophysical Research Letters*, *35*(8), L08705. https://doi.org/10.1029/2008GL033423

Teixeira, E. I., Fischer, G., van Velthuizen, H., Walter, C., & Ewert, F. (2013). Global hot-spots of heat stress on agricultural crops due to climate change. *Agricultural and Forest Meteorology*, *170*, 206–215. https://doi.org/10.1016/j.agrformet.2011.09.002

Traore, K. (2000). *Le parc à Karite (Vitellaria paradoxa - Gaertner f.) et la fertilite des sols de la zone Mali-Sud : Cas d'une toposéquence de la région de Fana.* Université de Nancy I, 19 p.

Trenberth, K. E. (2012). Framing the way to relate climate extremes to climate change. *Climatic Change*, *115*(2), 283–290. https://doi.org/10.1007/s10584-012-0441-5

UNFPA, & WCARO. (2020). *Mali : Étude monographique sur la Démographie, la Paix et la Sécurité au Sahel*.

USAID. (2014). *A Review of Fifteen Crops Cultivated in the Sahel* (Issue August).

USAID. (2017). *Climate change risk profile in West Africa Sahel: Regional Fact Sheet* (Issue April). https://www.climatelinks.org/sites/default/files/asset/document/2017 April_USAID ATLAS_Climate Change Risk Profile - Sahel.pdf

Vanmaekelbergh, D. (2009). From droplets to devices. *Nature Nanotechnology*, *4*(8), 475–476. https://doi.org/10.1038/nnano.2009.216

Vellido, A., Martín-Guerrero, J. D., & Lisboa, P. J. G. (2012). Making machine learning models interpretable. In *ESANN 2012 proceedings, 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning* (Issue April). https://www.i6doc.com/en/error/?error=1&GCOI=28001100967420.Making

Verdin, J., & Klaver, R. (2002). Grid-cell-based crop water accounting for the famine early warning system. *Hydrological Processes*, *16*(8), 1617–1630. https://doi.org/10.1002/hyp.1025

Vintrou, E. (2012). *Cartographie et caractérisation des systèmes agricoles au Mali par télédétection à moyenne résolution spatiale Elodie Vintrou To cite this version : HAL Id : pastel-00781223 pour obtenir le grade de Docteur délivré par l ' Institut des Sciences et Industr*.

Wahid, A., Gelani, S., Ashraf, M., & Foolad, M. (2007). Heat tolerance in plants: An overview. *Environmental and Experimental Botany*, *61*(3), 199–223. https://doi.org/10.1016/j.envexpbot.2007.05.011

Wilks, D. S. (2006). Comparison of ensemble-MOS methods in the Lorenz '96 setting. *Meteorological Applications*, *13*(03), 243. https://doi.org/10.1017/S1350482706002192

Wilson, E., & Minas, P. (2017). Climate and climate change. *Methods of Environmental and Social Impact Assessment*, *January*, 134–163. https://doi.org/10.4324/9781315626932

Wooldridge, M. J., Street, C., Manchester, M., Jennings, N. R., Road, M. E., & London, E. (1997). *Formalizing the Cooperative Problem Solving Process 2 A Formal Framework*. 1–15.

Wu, G., Liu, X., Chen, T., Xu, G., Wang, W., Zeng, X., & Zhang, X. (2015). Elevation-dependent variations of tree growth and intrinsic water-use efficiency in Schrenk spruce (Picea schrenkiana) in the western Tianshan Mountains, China. *Frontiers in Plant Science*, *6*(MAY), 1–12. https://doi.org/10.3389/fpls.2015.00309

Appendix

Logit prediction code in R

```
################Model, anomalies des prediteurs################
  model1<-glm(data_train_rd~.,data=data_train_pred2,
              family = binomial(link = "logit"))
  exp(coef(model1)) # regression coefficients
  anova(model1, test = "Chisq") #significance level of each predictor
  ############End of model building############
  data_test_pred <- data_test[,-c(3,6)]
  data_test_rd <- ifelse(diff(data_test$Yield_kg,lag = 1)<=0,1,0) # anomalies des rdts
==> dichotomic
  fo_test <- data_test$FO # faux departs ==> dichotomic
  data_test_pred2 <- matrix(NA,nrow(data_test_pred)-1,ncol(data_test_pred))
  for(n in 1:ncol(data_test_pred)){
    data_test_pred2[,n] <- diff(data_test_pred[,n],lag = 1) #%>% as.data.frame()
  }
  #data_test_pred2 <-
as.data.frame(cbind(data_test_pred2,fo_test[length(data_test[,8])-1,8]))
  data_test_pred2 <- as.data.frame(cbind(data_test_pred2,fo_test[-length(fo_test)]))
  #### Prediction values
  predi1a<-predict(model1,data_test_pred2,type = "response")
  ## Performance
  pred<-round(predi1a,0) ### predicted response
  obs<- data_test_rd ### observed response
  ####contingency table
  ##number of hits
  a<-length(which(pred==1 & obs==1))
  ##number of false alarms
  b<-length(which(pred==1 & obs==0))
  ##number of misses
  c<-length(which(pred==0 & obs==1))
  ##number of correct negatives
  d<-length(which(pred==0 & obs==0))
  ###verification indices
  tab<-matrix(c(a,c,b,d),ncol = 2)
  ##generate verification indices
  ind<-table.stats(tab)
  #pdf(paste(workdir,"/Performance_logit_orig.pdf",sep=""),
  #   width=8,height=8,paper="special")
  ##performance diagram
  points(1-ind$FAR, ind$POD, pch=16,col= "orange", cex=3) # Bakel
  #performance.diagram(main="Ouahigouya")
  #points(1-ind$FAR, ind$POD, pch=16,col= "red", cex=3) # Ouahi
}
legend("bottomright",legend = c("Maize","Millet", "Cotton"),#,
       #  "Mango","Bolgatanga","Dano"),
       col=c("blue","black","orange"), #,"black","grey","brown"),
       pch = c(16,16,16),cex=1.2)
dev.off()
### plotting reliability diagram
#}
```

Logit prediction non anomalies code in R

```
###############Model, anomalies des prediteurs#####
model1<-glm(data_train_rd~.,data=data_train_pred2,
            family = binomial(link = "logit"))
exp(coef(model1)) # regression coefficients
anova(model1, test = "Chisq") #significance level of each predictor
###########End of model building###########
data_test_pred <- data_test[,-c(3,6)]
data_test_rd <- ifelse(diff(data_test$Yield_kg,lag = 1)<=0,1,0) # anomalies des rdts
==> dichotomic
fo_test <- data_test$FO # faux departs ==> dichotomic)
#  }
### identification of all the extremes & compound events
data_test_pred2 <- matrix(NA,nrow(data_test_pred)-1,ncol(data_test_pred))
data_test_pred2[,1] <- ifelse(diff((data_test_pred$ECS.Jld.-
data_test_pred$OCS.Jld.),lag = 1)<=0,1,0)
data_test_pred2[,2] <- ifelse(diff((data_test_pred$ECS.Jld.),lag = 1)<=0,1,0)
data_test_pred2[,3] <- ifelse(diff((data_test_pred$Tp),lag = 1)>0,1,0)
data_test_pred2[,4] <- ifelse(diff((data_test_pred$RR.Season),lag = 1)<=0,1,0)
#data_test_pred2 <- as.data.frame(cbind(data_test_pred2,fo_test[length(data_test[,8])-
1,8]))
data_test_pred2 <- as.data.frame(cbind(data_test_pred2,fo_test[-length(fo_test)]))
#### Prediction values
predi1a<-predict(model1,data_test_pred2,type = "response")
## Performance
pred<-round(predi1a,0) ### predicted response
obs<- data_test_rd ### observed response
####contingency table
##number of hits
a<-length(which(pred==1 & obs==1))
##number of false alarms
b<-length(which(pred==1 & obs==0))
##number of misses
c<-length(which(pred==0 & obs==1))
##number of correct negatives
d<-length(which(pred==0 & obs==0))
###verification indices
tab<-matrix(c(a,c,b,d),ncol = 2)
##generate verification indices
ind<-table.stats(tab)
##performance diagram
points(1-ind$FAR, ind$POD, pch=16,col= "black", cex=3) # Bakel
}
### training data
dd <- read.csv2(paste(workdir,"Djidian_Cotton.txt",sep="/"),
                head=T,dec=".",sep="",na.strings =c("","NA"))
### testing data
fdirect <- list.files(workdir,pattern = "_Cotton.txt")
fnames <- substr(fdirect,1,7)
for(ii in 1:length(fdirect)){
  dd2 <- read.csv2(paste(workdir,fdirect[ii],sep="/"),
                   head=T,dec=".",sep="",na.strings =c("","NA"))
  ### Split data into training & testing subset
  dd$Yield_kg <- ifelse(is.na(dd$Yield_kg), mean(dd$Yield_kg, na.rm =
TRUE),dd$Yield_kg)
  dd$Tp <- ifelse(is.na(dd$Tp), mean(dd$Tp, na.rm = TRUE),dd$Tp)
  dd$FO <- ifelse(is.na(dd$FO)|dd$FO>1,0,dd$FO)
  dd$RR.Season <- ifelse(is.na(dd$RR.Season), mean(dd$RR.Season, na.rm =
TRUE),dd$RR.Season)
  dd2$Yield_kg <- ifelse(is.na(dd2$Yield_kg), mean(dd2$Yield_kg, na.rm =
TRUE),dd2$Yield_kg)
  dd2$Tp <- ifelse(is.na(dd2$Tp), mean(dd2$Tp, na.rm = TRUE),dd2$Tp)
```

```
  dd2$FO <- ifelse(is.na(dd2$FO)|dd2$FO>1,0,dd2$FO)
  dd2$RR.Season <- ifelse(is.na(dd2$RR.Season), mean(dd2$RR.Season, na.rm =
TRUE),dd2$RR.Season)
  # Remove unwanted columns
  data_train <- as.data.frame(dd[,-c(1,3,5:7)])
  data_test <- as.data.frame(dd2[,-c(1,3,5:7)])
    ### extract the predictors from training data
  data_train_rd <- ifelse(diff(data_train$Yield_kg,lag = 1)<=0,1,0) # anomalies des
rdts ==> dichotomic
  fo_train <- data_train$FO # faux departs ==> dichotomic
  data_train_pred <- data_train[,-c(3,6)]
  data_train_pred2 <- matrix(NA,nrow(data_train_pred)-1,ncol(data_train_pred))
  data_train_pred2[,1] <- ifelse(diff((data_train_pred$ECS.Jld.-
data_train_pred$OCS.Jld.),lag = 1)<=0,1,0)
  data_train_pred2[,2] <- ifelse(diff((data_train_pred$ECS.Jld.),lag = 1)<=0,1,0)
  data_train_pred2[,3] <- ifelse(diff((data_train_pred$Tp),lag = 1)>0,1,0)
  data_train_pred2[,4] <- ifelse(diff((data_train_pred$RR.Season),lag = 1)<=0,1,0)

  #  }
  # data_train_pred2 <- data_train_pred
  data_train_pred2 <- as.data.frame(cbind(data_train_pred2,fo_train[-
length(fo_train)]))
  #Use the glm function with different link function to fit your predictant
  ################Model, anomalies des prediteurs#####
  model1<-glm(data_train_rd~.,data=data_train_pred2,
              family = binomial(link = "logit"))
  exp(coef(model1)) # regression coefficients
  anova(model1, test = "Chisq") #significance level of each predictor
  ############End of model building############
  data_test_pred <- data_test[,-c(3,6)]
  data_test_rd <- ifelse(diff(data_test$Yield_kg,lag = 1)<=0,1,0) # anomalies des rdts
==> dichotomic
  fo_test <- data_test$FO # faux departs ==> dichotomic
  ### identification of all the extremes & compound events
  data_test_pred2 <- matrix(NA,nrow(data_test_pred)-1,ncol(data_test_pred))
  data_test_pred2[,1] <- ifelse(diff((data_test_pred$ECS.Jld.-
data_test_pred$OCS.Jld.),lag = 1)<=0,1,0)
  data_test_pred2[,2] <- ifelse(diff((data_test_pred$ECS.Jld.),lag = 1)<=0,1,0)
  data_test_pred2[,3] <- ifelse(diff((data_test_pred$Tp),lag = 1)>0,1,0)
  data_test_pred2[,4] <- ifelse(diff((data_test_pred$RR.Season),lag = 1)<=0,1,0)
  #data_test_pred2 <-
as.data.frame(cbind(data_test_pred2,fo_test[length(data_test[,8])-1,8]))
  data_test_pred2 <- as.data.frame(cbind(data_test_pred2,fo_test[-length(fo_test)]))
  #### Prediction values
  predi1a<-predict(model1,data_test_pred2,type = "response")
  ## Performance
  pred<-round(predi1a,0) ### predicted response
  obs<- data_test_rd ### observed response
  ####contingency table
  ##number of hits
  a<-length(which(pred==1 & obs==1))
  ##number of false alarms
  b<-length(which(pred==1 & obs==0))
  ##number of misses
  c<-length(which(pred==0 & obs==1))
  ##number of correct negatives
  d<-length(which(pred==0 & obs==0))
  ###verification indices
  tab<-matrix(c(a,c,b,d),ncol = 2)
  ##generate verification indices
  ind<-table.stats(tab)
 }
legend("bottomright",legend = c("Maize","Millet", "Cotton"),#,
```

```
       pch = c(16,16,16),cex=1.2)
dev.off()
### plotting reliability diagram
#}
```

Random Forest prediction code in R

```
################Model, anomalies des prediteurs#####
attach(data_train_pred2)
#model1<-glm(data_train_rd~.,data=data_train_pred2,
#             family = binomial(link = "logit"))
model1<- train(rd ~ ocs + ecs + hw + rrtot + fo, #Pclass + Sex + SibSp +
               #Embarked + Parch + Fare, # Survived is a function of the variables we
decided to include
               data = data_train_pred2, # Use the train data frame as the training
data
               method = 'rf',# Use the 'random forest' algorithm
               trControl = trainControl(method = 'cv'), # Use cross-validation
               number = 5) # Use 5 folds for cross-validation
############End of model building############
data_test_pred <- data_test[,-c(3,6)]
data_test_rd <- ifelse(diff(data_test$Yield_kg,lag = 1)<=0,1,0) # anomalies des rdts
==> dichotomic
fo_test <- data_test$FO # faux departs ==> dichotomic
data_test_pred2 <- matrix(NA,nrow(data_test_pred)-1,ncol(data_test_pred))
for(n in 1:ncol(data_test_pred)){
  data_test_pred2[,n] <- diff(data_test_pred[,n],lag = 1) #%>% as.data.frame()
}
#data_test_pred2 <- as.data.frame(cbind(data_test_pred2,fo_test[length(data_test[,8])-
1,8]))
data_test_pred2 <- as.data.frame(cbind(data_test_pred2,fo_test[-length(fo_test)]))
#### Prediction values
#predi1a<-predict(model1,data_test_pred2,type = "response")
predi1a <- as.numeric(predict(model1, newdata = data_test_pred2))
## Performance
# pred<- ifelse(predi1a <1 | predi1a > 1,0,1) ### predicted response
pred<- ifelse(predi1a <1 | predi1a > 1,1,0) ### predicted response
obs<- data_test_rd ### observed response
####contingency table
##number of hits
a<-length(which(pred==1 & obs==1))
##number of false alarms
b<-length(which(pred==1 & obs==0))
##number of misses
c<-length(which(pred==0 & obs==1))
##number of correct negatives
d<-length(which(pred==0 & obs==0))
###verification indices
tab<-matrix(c(a,c,b,d),ncol = 2)
##generate verification indices
ind<-table.stats(tab)
##performance diagram
points(1-ind$FAR, ind$POD, pch=16,col= "black", cex=3) # Bakel
#performance.diagram(main="Ouahigouya")
#points(1-ind$FAR, ind$POD, pch=16,col= "red", cex=3) # Ouahi
}
### training data
dd <- read.csv2(paste(workdir,"Djidian_Cotton.txt",sep="/"),
                head=T,dec=".",sep="",na.strings =c("","NA"))
fdirect <- list.files(workdir,pattern = "_Cotton.txt")
fnames <- substr(fdirect,1,7)
for(ii in 1:length(fdirect)){
  dd2 <- read.csv2(paste(workdir,fdirect[ii],sep="/"),
                   head=T,dec=".",sep="",na.strings =c("","NA"))
  ### Split data into training & testing subset
  dd$Yield_kg <- ifelse(is.na(dd$Yield_kg), mean(dd$Yield_kg, na.rm =
TRUE),dd$Yield_kg)
  dd$Tp <- ifelse(is.na(dd$Tp), mean(dd$Tp, na.rm = TRUE),dd$Tp)
```

```
  dd$FO <- ifelse(is.na(dd$FO)|dd$FO>1,0,dd$FO)
  dd$RR.Season <- ifelse(is.na(dd$RR.Season), mean(dd$RR.Season, na.rm =
TRUE),dd$RR.Season)
  dd2$Yield_kg <- ifelse(is.na(dd2$Yield_kg), mean(dd2$Yield_kg, na.rm =
TRUE),dd2$Yield_kg)
  dd2$Tp <- ifelse(is.na(dd2$Tp), mean(dd2$Tp, na.rm = TRUE),dd2$Tp)
  dd2$FO <- ifelse(is.na(dd2$FO)|dd2$FO>1,0,dd2$FO)
  dd2$RR.Season <- ifelse(is.na(dd2$RR.Season), mean(dd2$RR.Season, na.rm =
TRUE),dd2$RR.Season)
  # Remove unwanted columns
  data_train <- as.data.frame(dd[,-c(1,3,5:7)])
  data_test <- as.data.frame(dd2[,-c(1,3,5:7)])
  data_train_rd <- ifelse(diff(data_train$Yield_kg,lag = 1)<=0,1,0) # anomalies des
rdts ==> dichotomic
  fo_train <- data_train$FO # faux departs ==> dichotomic
  data_train_pred <- data_train[,-c(3,6)]
  data_train_pred2 <- matrix(NA,nrow(data_train_pred)-1,ncol(data_train_pred))
  for(n in 1:ncol(data_train_pred)){
    data_train_pred2[,n] <- diff(data_train_pred[,n],lag = 1) #%>% as.data.frame()
  }
  data_train_pred2 <- as.data.frame(cbind(data_train_rd,data_train_pred2,fo_train[-
length(fo_train)]))
  colnames(data_train_pred2) <- c("rd","ocs","ecs","hw","rrtot","fo")
  # Converting ‘Survived’ to a factor
  data_train_pred2$rd <- factor(data_train_pred2$rd)
  #Use the glm function with different link function to fit your predictant
  ###############Model, anomalies des prediteurs#####
  attach(data_train_pred2)
  #model1<-glm(data_train_rd~.,data=data_train_pred2,
  #            family = binomial(link = "logit"))
  model1<- train(rd ~ ocs + ecs + hw + rrtot + fo, #Pclass + Sex + SibSp +
                 #Embarked + Parch + Fare, # Survived is a function of the variables
we decided to include
                 data = data_train_pred2, # Use the train data frame as the training
data
                 method = 'rf',# Use the 'random forest' algorithm
                 trControl = trainControl(method = 'cv'), # Use cross-validation
                 number = 5) # Use 5 folds for cross-validation
  ############End of model building############

  data_test_pred <- data_test[,-c(3,6)]

  data_test_rd <- ifelse(diff(data_test$Yield_kg,lag = 1)<=0,1,0) # anomalies des rdts
==> dichotomic
  fo_test <- data_test$FO # faux departs ==> dichotomic

  data_test_pred2 <- matrix(NA,nrow(data_test_pred)-1,ncol(data_test_pred))

  for(n in 1:ncol(data_test_pred)){
    data_test_pred2[,n] <- diff(data_test_pred[,n],lag = 1) #%>% as.data.frame()
  }
  #data_test_pred2 <-
as.data.frame(cbind(data_test_pred2,fo_test[length(data_test[,8])-1,8]))
  data_test_pred2 <- as.data.frame(cbind(data_test_pred2,fo_test[-length(fo_test)]))
  #### Prediction values
  #predi1a<-predict(model1,data_test_pred2,type = "response")
  predi1a <- as.numeric(predict(model1, newdata = data_test_pred2))
  ## Performance
  #pred<- ifelse(predi1a <1 | predi1a > 1,0,1) ### predicted response
  pred<- ifelse(predi1a <1 | predi1a > 1,1,0) ### predicted response
  obs<- data_test_rd ### observed response

  ####contingency table
```

```
  ##number of hits
  a<-length(which(pred==1 & obs==1))
  ##number of false alarms
  b<-length(which(pred==1 & obs==0))
  ##number of misses
  c<-length(which(pred==0 & obs==1))
  ##number of correct negatives
  d<-length(which(pred==0 & obs==0))
  ###verification indices
  tab<-matrix(c(a,c,b,d),ncol = 2)
  ##generate verification indices
  ind<-table.stats(tab)
  points(1-ind$FAR, ind$POD, pch=16,col= "orange", cex=3) # Bakel
}

legend("bottomright",legend = c("Maize","Millet", "Cotton"),#,
       col=c("blue","black","orange"),
       pch = c(16,16,16),cex=1.2)
dev.off()
### plotting reliability diagram
#}
```

Random Forest prediction no anomalie code in R

```
###############Model, anomalies des prediteurs#####
attach(data_train_pred2)
#model1<-glm(data_train_rd~.,data=data_train_pred2,
#            family = binomial(link = "logit"))
model1<- train(rd ~ ocs + ecs + hw + rrtot + fo, #Pclass + Sex + SibSp +
               #Embarked + Parch + Fare, # Survived is a function of the variables we
decided to include
               data = data_train_pred2, # Use the train data frame as the training
data
               method = 'rf',# Use the 'random forest' algorithm
               trControl = trainControl(method = 'cv'), # Use cross-validation
               number = 5) # Use 5 folds for cross-validation
############End of model building############
data_test_pred <- data_test[,-c(3,6)]
data_test_rd <- ifelse(diff(data_test$Yield_kg,lag = 1)<=0,1,0) # anomalies des rdts
==> dichotomic
fo_test <- data_test$FO # faux departs ==> dichotomic
data_test_pred2 <- matrix(NA,nrow(data_test_pred)-1,ncol(data_test_pred))
### identification of all the extremes & compound events
data_test_pred2 <- matrix(NA,nrow(data_test_pred)-1,ncol(data_test_pred))
data_test_pred2[,1] <- ifelse(diff((data_test_pred$ECS.Jld.-
data_test_pred$OCS.Jld.),lag = 1)<=0,1,0)
data_test_pred2[,2] <- ifelse(diff((data_test_pred$ECS.Jld.),lag = 1)<=0,1,0)
data_test_pred2[,3] <- ifelse(diff((data_test_pred$Tp),lag = 1)>0,1,0)
data_test_pred2[,4] <- ifelse(diff((data_test_pred$RR.Season),lag = 1)<=0,1,0)
#data_test_pred2 <- as.data.frame(cbind(data_test_pred2,fo_test[length(data_test[,8])-
1,8]))
data_test_pred2 <- as.data.frame(cbind(data_test_pred2,fo_test[-length(fo_test)]))
#### Prediction values
#predi1a<-predict(model1,data_test_pred2,type = "response")
predi1a <- as.numeric(predict(model1, newdata = data_test_pred2))
## Performance
pred<- ifelse(predi1a <1 | predi1a > 1,1,0) ### predicted response
obs<- data_test_rd ### observed response
####contingency table
##number of hits
a<-length(which(pred==1 & obs==1))
##number of false alarms
b<-length(which(pred==1 & obs==0))
##number of misses
c<-length(which(pred==0 & obs==1))
##number of correct negatives
d<-length(which(pred==0 & obs==0))
###verification indices
tab<-matrix(c(a,c,b,d),ncol = 2)
##generate verification indices
ind<-table.stats(tab)
##performance diagram
points(1-ind$FAR, ind$POD, pch=16,col= "blue", cex=3) # Bakel
}
### training data
dd <- read.csv2(paste(workdir,"Djidian_Millet.txt",sep="/"),
                head=T,dec=".",sep="",na.strings =c("","NA"))
### testing data
fdirect <- list.files(workdir,pattern = "_Millet.txt")
fnames <- substr(fdirect,1,7)
for(ii in 1:length(fdirect)){
  dd2 <- read.csv2(paste(workdir,fdirect[ii],sep="/"),
                   head=T,dec=".",sep="",na.strings =c("","NA"))
  ### Split data into training & testing subset
```

```
  dd$Yield_kg <- ifelse(is.na(dd$Yield_kg), mean(dd$Yield_kg, na.rm = 
TRUE),dd$Yield_kg)
  dd$Tp <- ifelse(is.na(dd$Tp), mean(dd$Tp, na.rm = TRUE),dd$Tp)
  dd$FO <- ifelse(is.na(dd$FO)|dd$FO>1,0,dd$FO)
  dd$RR.Season <- ifelse(is.na(dd$RR.Season), mean(dd$RR.Season, na.rm = 
TRUE),dd$RR.Season)
  dd2$Yield_kg <- ifelse(is.na(dd2$Yield_kg), mean(dd2$Yield_kg, na.rm = 
TRUE),dd2$Yield_kg)
  dd2$Tp <- ifelse(is.na(dd2$Tp), mean(dd2$Tp, na.rm = TRUE),dd2$Tp)
  dd2$FO <- ifelse(is.na(dd2$FO)|dd2$FO>1,0,dd2$FO)
  dd2$RR.Season <- ifelse(is.na(dd2$RR.Season), mean(dd2$RR.Season, na.rm = 
TRUE),dd2$RR.Season)
  # Remove unwanted columns
  data_train <- as.data.frame(dd[,-c(1,3,5:7)])
  data_test <- as.data.frame(dd2[,-c(1,3,5:7)])
  ### extract the predictors from training data
  data_train_rd <- ifelse(diff(data_train$Yield_kg,lag = 1)<=0,1,0) # anomalies des 
rdts ==> dichotomic
  fo_train <- data_train$FO # faux departs ==> dichotomic
  data_train_pred <- data_train[,-c(3,6)]
  #data_train_pred2 <- matrix(NA,nrow(data_train_pred)-1,ncol(data_train_pred))
  ### identification of all the extremes & compound events
  data_train_pred2 <- matrix(NA,nrow(data_train_pred)-1,ncol(data_train_pred))
  data_train_pred2[,1] <- ifelse(diff((data_train_pred$ECS.Jld.-
data_train_pred$OCS.Jld.),lag = 1)<=0,1,0)
  data_train_pred2[,2] <- ifelse(diff((data_train_pred$ECS.Jld.),lag = 1)<=0,1,0)
  data_train_pred2[,3] <- ifelse(diff((data_train_pred$Tp),lag = 1)>0,1,0)
  data_train_pred2[,4] <- ifelse(diff((data_train_pred$RR.Season),lag = 1)<=0,1,0)
  data_train_pred2 <- as.data.frame(cbind(data_train_rd,data_train_pred2,fo_train[-
length(fo_train)]))
  colnames(data_train_pred2) <- c("rd","ocs","ecs","hw","rrtot","fo")
  # Converting ‘Survived’ to a factor
  data_train_pred2$rd <- factor(data_train_pred2$rd)
  #Use the glm function with different link function to fit your predictant
  ###############Model, anomalies des prediteurs#####
  attach(data_train_pred2)
  #model1<-glm(data_train_rd~.,data=data_train_pred2,
  #             family = binomial(link = "logit"))
  model1<- train(rd ~ ocs + ecs + hw + rrtot + fo, #Pclass + Sex + SibSp +
                 #Embarked + Parch + Fare, # Survived is a function of the variables 
we decided to include
                 data = data_train_pred2, # Use the train data frame as the training 
data
                 method = 'rf',# Use the 'random forest' algorithm
                 trControl = trainControl(method = 'cv'), # Use cross-validation
                 number = 5) # Use 5 folds for cross-validation
   ###########End of model building###########
  data_test_pred <- data_test[,-c(3,6)]
  data_test_rd <- ifelse(diff(data_test$Yield_kg,lag = 1)<=0,1,0) # anomalies des rdts 
==> dichotomic
  fo_test <- data_test$FO # faux departs ==> dichotomic
  data_test_pred2 <- matrix(NA,nrow(data_test_pred)-1,ncol(data_test_pred))
  ### identification of all the extremes & compound events
  data_test_pred2 <- matrix(NA,nrow(data_test_pred)-1,ncol(data_test_pred))
  data_test_pred2[,1] <- ifelse(diff((data_test_pred$ECS.Jld.-
data_test_pred$OCS.Jld.),lag = 1)<=0,1,0)
  data_test_pred2[,2] <- ifelse(diff((data_test_pred$ECS.Jld.),lag = 1)<=0,1,0)
  data_test_pred2[,3] <- ifelse(diff((data_test_pred$Tp),lag = 1)>0,1,0)
  data_test_pred2[,4] <- ifelse(diff((data_test_pred$RR.Season),lag = 1)<=0,1,0)
  #data_test_pred2 <- 
as.data.frame(cbind(data_test_pred2,fo_test[length(data_test[,8])-1,8]))
  data_test_pred2 <- as.data.frame(cbind(data_test_pred2,fo_test[-length(fo_test)]))
  #### Prediction values
```

```
  #predi1a<-predict(model1,data_test_pred2,type = "response")
  predi1a <- as.numeric(predict(model1, newdata = data_test_pred2))
  ## Performance
  # pred<- ifelse(predi1a <1 | predi1a > 1,0,1) ### predicted response
  pred<- ifelse(predi1a <1 | predi1a > 1,1,0) ### predicted response
  obs<- data_test_rd ### observed response
  ####contingency table
  ##number of hits
  a<-length(which(pred==1 & obs==1))
  ##number of false alarms
  b<-length(which(pred==1 & obs==0))
  ##number of misses
  c<-length(which(pred==0 & obs==1))
  ##number of correct negatives
  d<-length(which(pred==0 & obs==0))
  ###verification indices
  tab<-matrix(c(a,c,b,d),ncol = 2)
  ##generate verification indices
  ind<-table.stats(tab)
  #pdf(paste(workdir,"/Performance_logit_orig.pdf",sep=""),
  #   width=8,height=8,paper="special")
  ##performance diagram
  points(1-ind$FAR, ind$POD, pch=16,col= "black", cex=3) # Bakel
  #performance.diagram(main="Ouahigouya")
  #points(1-ind$FAR, ind$POD, pch=16,col= "red", cex=3) # Ouahi
}
### training data
dd <- read.csv2(paste(workdir,"Djidian_Cotton.txt",sep="/"),
                head=T,dec=".",sep="",na.strings =c("","NA"))
### testing data
fdirect <- list.files(workdir,pattern = "_Cotton.txt")
fnames <- substr(fdirect,1,7)
for(ii in 1:length(fdirect)){
  dd2 <- read.csv2(paste(workdir,fdirect[ii],sep="/"),
                   head=T,dec=".",sep="",na.strings =c("","NA"))
  ### Split data into training & testing subset
  dd$Yield_kg <- ifelse(is.na(dd$Yield_kg), mean(dd$Yield_kg, na.rm =
TRUE),dd$Yield_kg)
  dd$Tp <- ifelse(is.na(dd$Tp), mean(dd$Tp, na.rm = TRUE),dd$Tp)
  dd$FO <- ifelse(is.na(dd$FO)|dd$FO>1,0,dd$FO)
  dd$RR.Season <- ifelse(is.na(dd$RR.Season), mean(dd$RR.Season, na.rm =
TRUE),dd$RR.Season)
  dd2$Yield_kg <- ifelse(is.na(dd2$Yield_kg), mean(dd2$Yield_kg, na.rm =
TRUE),dd2$Yield_kg)
  dd2$Tp <- ifelse(is.na(dd2$Tp), mean(dd2$Tp, na.rm = TRUE),dd2$Tp)
  dd2$FO <- ifelse(is.na(dd2$FO)|dd2$FO>1,0,dd2$FO)
  dd2$RR.Season <- ifelse(is.na(dd2$RR.Season), mean(dd2$RR.Season, na.rm =
TRUE),dd2$RR.Season)
  # Remove unwanted columns
  data_train <- as.data.frame(dd[,-c(1,3,5:7)])
  data_test <- as.data.frame(dd2[,-c(1,3,5:7)])
  ### extract the predictors from training data
  data_train_rd <- ifelse(diff(data_train$Yield_kg,lag = 1)<=0,1,0) # anomalies des
rdts ==> dichotomic
  fo_train <- data_train$FO # faux departs ==> dichotomic
  data_train_pred <- data_train[,-c(3,6)]
  #data_train_pred2 <- matrix(NA,nrow(data_train_pred)-1,ncol(data_train_pred))
  ### identification of all the extremes & compound events
  data_train_pred2 <- matrix(NA,nrow(data_train_pred)-1,ncol(data_train_pred))
  data_train_pred2[,1] <- ifelse(diff((data_train_pred$ECS.Jld.-
data_train_pred$OCS.Jld.),lag = 1)<=0,1,0)
  data_train_pred2[,2] <- ifelse(diff((data_train_pred$ECS.Jld.),lag = 1)<=0,1,0)
  data_train_pred2[,3] <- ifelse(diff((data_train_pred$Tp),lag = 1)>0,1,0)
```

```
  data_train_pred2[,4] <- ifelse(diff((data_train_pred$RR.Season),lag = 1)<=0,1,0)
  data_train_pred2 <- as.data.frame(cbind(data_train_rd,data_train_pred2,fo_train[-
length(fo_train)]))
  colnames(data_train_pred2) <- c("rd","ocs","ecs","hw","rrtot","fo")
  # Converting ‘Survived’ to a factor
  data_train_pred2$rd <- factor(data_train_pred2$rd)
  #Use the glm function with different link function to fit your predictant
  ################Model, anomalies des prediteurs#####
  attach(data_train_pred2)
  #model1<-glm(data_train_rd~.,data=data_train_pred2,
  #             family = binomial(link = "logit"))
  model1<- train(rd ~ ocs + ecs + hw + rrtot + fo, #Pclass + Sex + SibSp +
                #Embarked + Parch + Fare, # Survived is a function of the variables
we decided to include
                data = data_train_pred2, # Use the train data frame as the training
data
                method = 'rf',# Use the 'random forest' algorithm
                trControl = trainControl(method = 'cv'), # Use cross-validation
                number = 5) # Use 5 folds for cross-validation
  #exp(coef(model1)) # regression coefficients
  #anova(model1, test = "Chisq") #significance level of each predictor
  ############End of model building############
  data_test_pred <- data_test[,-c(3,6)]
  data_test_rd <- ifelse(diff(data_test$Yield_kg,lag = 1)<=0,1,0) # anomalies des rdts
==> dichotomic
  fo_test <- data_test$FO # faux departs ==> dichotomic
  data_test_pred2 <- matrix(NA,nrow(data_test_pred)-1,ncol(data_test_pred))
  ### identification of all the extremes & compound events
  data_test_pred2 <- matrix(NA,nrow(data_test_pred)-1,ncol(data_test_pred))
  data_test_pred2[,1] <- ifelse(diff((data_test_pred$ECS.Jld.-
data_test_pred$OCS.Jld.),lag = 1)<=0,1,0)
  data_test_pred2[,2] <- ifelse(diff((data_test_pred$ECS.Jld.),lag = 1)<=0,1,0)
  data_test_pred2[,3] <- ifelse(diff((data_test_pred$Tp),lag = 1)>0,1,0)
  data_test_pred2[,4] <- ifelse(diff((data_test_pred$RR.Season),lag = 1)<=0,1,0)
  #data_test_pred2 <-
as.data.frame(cbind(data_test_pred2,fo_test[length(data_test[,8])-1,8]))
  data_test_pred2 <- as.data.frame(cbind(data_test_pred2,fo_test[-length(fo_test)]))
  #### Prediction values
  #predi1a<-predict(model1,data_test_pred2,type = "response")
  predi1a <- as.numeric(predict(model1, newdata = data_test_pred2))
  ## Performance
  # pred<- ifelse(predi1a <1 | predi1a > 1,0,1) ### predicted response
  pred<- ifelse(predi1a <1 | predi1a > 1,1,0) ### predicted response
  obs<- data_test_rd ### observed response
  ####contingency table
  ##number of hits
  a<-length(which(pred==1 & obs==1))
  ##number of false alarms
  b<-length(which(pred==1 & obs==0))
  ##number of misses
  c<-length(which(pred==0 & obs==1))
  ##number of correct negatives
  d<-length(which(pred==0 & obs==0))
  ###verification indices
  tab<-matrix(c(a,c,b,d),ncol = 2)
  ##generate verification indices
  ind<-table.stats(tab)
  ##performance diagram
  points(1-ind$FAR, ind$POD, pch=16,col= "orange", cex=3) # Bakel
}
legend("bottomright",legend = c("Maize","Millet", "Cotton"),#,
       #  "Mango","Bolgatanga","Dano"),
       col=c("blue","black","orange"), #,"black","grey","brown"),
```

```
       pch = c(16,16,16),cex=1.2)
dev.off()
### plotting reliability diagram
#}
```