MASTER THESIS

(SODECOTON, 2022a)

Exploration of the statistical relationships between rainfall indices and cotton yields in northern Cameroon, to strengthen the resilience of farmers to climate change.

By Clara KNOPS

Institut Agro Montpellier, AgroParisTech, Université de Montpellier

Thesis presented the: 03/09/2024

Master thesis
presented for the attribution of the Master 2 Degree
Major: Water
Specialization: Water and Agriculture

Exploration of the statistical relationships between rainfall indices and cotton yields in northern Cameroon, to strengthen the resilience of farmers to climate change.

By Clara KNOPS

Thesis presented the: 03/09/2024

Members of the jury:
Flavie CERNESSON
Gilles BELAUD
Jean-Stéphane BAILLY
Jérémy LAVARENNE
Frédéric ROSSEL

Host organization: UMR-TETIS, CIRAD
Thesis advisor: Jean-Stéphane BAILLY
Internship supervisor: Jérémy LAVARENNE

ACKNOWLEDGMENTS

First, I want to thank the INNOVACC project with all its members and contributing organisms – CIFOR-ICRAF, CIRAD, Energie pour le monde, the European Union and IRAD - for not only creating this amazing project, but also for financing my internship and essentially my thesis.

I would like to acknowledge as well, the great help brought by SODECOTON, giving access to their data without which this study probably couldn’t have been conducted. In this sense I also want to thank Antoine Leblois, for taking time out of his busy schedule and sharing, as well as explaining his datasets.

Most of all, I would like to acknowledge the great contribution of my internship supervisor Jérémy Lavarenne. Never too tired to answer my countless questions, look at my codes and read my draftiest drafts. I want to thank him for being exactly the kind of supervisor one would wish for, for her last internship experience. I would also like to thank the other project members - Ibrahim Njouenwet and Victor Nenwala - for their helpful contribution and sharing of amazing knowledge.

As well, all my other colleagues at the “Maison de la Télédétection” deserve to be acknowledged. Thank you for welcoming me so warmly along your midst, brightening up every day of work. A special thanks goes to Giulia Faustini, probably PhD by now, who on most days played the role of a second supervisor for free, to accompany me through the maze of research.

I would like to also thank my professors, who guided me through the last two years. My trusted thesis advisor Jean-Stéphane Bailly, who has taught me much about coding, statistics and much more. Gilles Belaud and Flavie Cernesson who gave it their all to manage the Master of Water and Agriculture, while it might have not always been easy to take care of 15 students at once.

I want to thank my family, always supporting me, even from far. My grandparents with their abandoned belief in me, my parents with their helpful advice, my siblings with their encouraging words, my brother-in-law with his great comfort food and my nephews with their bright smiles.

Of course I cannot miss thanking all my friends, especially Alice Levalet, Don Carlos Daniel Cortes, Beatrice Mayorca and Karen Diaz-Thomas for always showing me their support and listening to my successes as well as my difficulties during this time.

Lastly, even if it means that the acknowledgments will take up a second page, I want to give myself a pad on the shoulder for a job that was not always easy, but to my judgment, well done at the end.

RÉSUMÉ

Ce mémoire examine les relations statistiques entre les indices de pluviométrie et les rendements du coton dans le nord du Cameroun, une région fortement dépendante du coton et vulnérable au changement climatique en raison de sa forte variabilité des pluies. Les données quotidiennes de pluies provenant du jeu de données NoCORA ont été interpolées à l'aide du krigeage ordinaire pour calculer des cartes annuelles d'indices de pluviométrie pour un total de 25 indices. Les données de rendement du coton à deux niveaux géographiques différents ont également été fournies par SODECOTON. En appliquant des régressions linéaires simples et multiples, l'impact des indices de pluviométrie sur les rendements du coton a été analysé. Les indices les plus fortement liés de manière statistiquement significative étaient la date de début et de cessation des pluies ainsi que la longueur de la saison, le nombre de jours secs, les périodes secs 10 et 15, la quantité de pluies saisonnières, les jours de pluie, les jours humides 20 et 30, ainsi que les jours des fortes pluies. Nos résultats permettront de poursuivre les recherches sur ce sujet, en vue d'analyses prédictives utilisant des données de projection climatique.

Mots clés: Pluviométrie, coton, changement climatique, Cameroun, statistiques, interpolation

ABSTRACT

This thesis investigates the statistical relationships between rainfall indices and cotton yields in northern Cameroon, a region heavily dependent on cotton and vulnerable to climate change due to its high rainfall variability. Daily rainfall data from the NoCORA rainfall dataset was interpolated using Ordinary Kriging to calculate yearly rainfall indices maps for a total of 25 indices. Cotton yield data on two different geographical levels was additionally provided by SODECOTON. Applying simple and multiple linear regression, the impact of the rainfall indices on cotton yields were analyzed. The onset and cessation day of the rainy season as well as the season length, dry days, dry spell consecutive 10 and 15, seasonal rainfall amount, rain days, wet days 20 and 30, as well as heavy rain days were found to be the indices with the strongest, statistically significant relationships. Our findings will allow further research into the topic, serving for prediction-analysis using climate projection data.

Keywords: Rainfall, cotton, climate change, Cameroon, statistics, interpolation

TABLE OF CONTENTS

Acknowledgments ... 1

Résumé/Abstract ... 3

Table of Contents ... 4

Foreword ... 6

List of Notations ... 7

List of Abbreviations ... 7

List of Figures ... 9

List of Tables ... 10

1. Introduction ... 11

2. State of the Art ... 12

3. Objectives and Hypotheses ... 14

4. Material and Methods ... 14

4.1. Overall Methodology ... 14

4.2. Study Area ... 15

4.3. Data ... 16

4.3.1. Rainfall Data ... 16

4.3.2. Cotton Data ... 17

4.4. Methods ... 19

4.4.1. Generating daily rainfall maps from daily rain gauge data using Kriging interpolation ... 19

4.4.2. Rainfall Indices ... 21

4.4.3. Cotton data preparation ... 23

4.4.4. Exploratory data analysis of the relationships between rainfall indices and cotton yield, and Model fitting ... 24

5. Results ... 31

5.1. Interpolated rainfall maps ... 31

5.2. Average spatial distribution and variability of rainfall indices ... 33

5.3. Spatial distribution of average cotton yields ... 35

5.4. Statistical relationships between rainfall indices and cotton yields ... 37

5.4.1. Simple linear relationships ... 37

5.4.2. Multiple linear relationships 43
5.4.3. Evaluation of metrics 43
6. Discussion 44
7. Limitations and Perspectives 48
8. Conclusion 49
References 51
Appendices 56

FOREWORD

“The correlation coefficient shows that a master’s student’s happiness will increase the earlier the end of his or her thesis arrives.”

This quote was thought of by the great second year “Water and Agriculture” master’s student Clara Knops, as she wrote her final lines of this thesis paper.

The following pages contain said master thesis on the “Exploration of the statistical relationships between rainfall indices and cotton yields in northern Cameroon, to strengthen the resilience of farmers to climate change. “. It was written to fulfill the graduation requirement of the master’s in “Water” at the Institut Agro, AgroParisTech and the Université de Montpellier. The research and writing of my thesis took place between February and August 2024 during an internship at the CIRAD - UMR Tetis financed by the INNOVACC project (cf. Appendix 20 and 21).

Before making the first step into the career world, for one last time I wanted to push my know-how while being a student and acquire new skills in the areas of agriculture, climate change and international development. This study allowed me to get to know the beautiful country of Cameroon, its people, climate conditions and cotton production, even if working from France. I gained new knowledge in data preparation and processing, interpolation techniques and statistical methods. For the first time, I worked with and became proficient in Python. In addition, I was able to consolidate my previous skills with other computer programs, as well as my comprehension about climate change and its interplay with water and agriculture.

Once more, I learned about the struggles faced in research which has taught me valuable lessons both professionally and personally. I have the upmost respect for every researcher and thesis writer and wish them good luck. The end will arrive sooner than you think, and you will be very happy about your accomplishments.

LIST OF NOTATIONS

c: Covariance

ε: Residual error

m: Intercept

MAE: Mean Absolute Error

ME: Mean Error

Nan: Not a Number value

o: Observed rainfall value

p: P-value

R: Rainfall value

r: Pearson correlation coefficient

r^2: Coefficient of determination

RMSE: Root Mean Square Error

ry: Slope coefficient

μ: Lagrange multiplier

w: Kriging weights

x, y: Longitude and Latitude coordinates

Y: Yield value

z: Predicted rainfall value

∑: Summation symbol

LIST OF ABBREVIATIONS

CFDT: Compagnie Française de Développement des Textiles

CIFOR-ICRAF: Center for International Forestry Research and World Agroforestry

CIRAD: Centre de coopération Internationale en Recherche Agronomique pour le Développement

CROPGRO: CROP GROwth

DEM: Digital Elevation Model

DS: Dry Spells

GHCN: Hydroelectric Power Station, the Global Historical Climatology Network

IDW: Inverse Distance Weighting

INNOVACC: Innovation for Adaptation to Climate Change

IRAD: Institut de Recherche Agricole pour le Développement

IRD: Institut de Recherche pour le Développement

LOOCV: Leave-One-Out Cross-Validation

LULC: Land Use/ Land Cover

Nbr: Number

NoCORA: Northern Cameroon Observed Rainfall Archive

OK: Ordinary Kriging

OLS: Ordinary Least Square

SEMRY: Société d'Expansion et de Modernisation de la Riziculture dans la ville de Yagoua

SODECOTON: Société de Développement du Coton du Cameroun

Std: Standard deviation

TAHMO: Trans-African Hydro-Meteorological Observatory

UK: Universal Kriging

UMR TETIS: Unité Mixte de Recherche Territoires, Environnement, Télédétection et Information Spatiale

WS: Wet Spells

LIST OF FIGURES

Fig. 1: Overall methodology framework 15

Fig. 2: Digital Elevation Model of northern Cameroon... 16

Source figures on the left: Victor Nenwala, personal communication 16

Fig. 3: Localization of NoCORA rain gauge stations between 1927 and 2022. 17

Fig. 4: Localization of SODECOTON collection points between 2007 and 2010. 18

Fig. 5: Localization of SODECOTON sectors between 1991 and 2010. 18

Fig. 6: Dimensions applied for model fitting.. 26

Fig. 7: Model fitting framework 30

Fig. 8: Localization of rain gauge stations between 1991 and 2010... 31

Fig. 9: Zoom on interpolation performances for different methods and favorable variograms, evaluated through LOOCV for the error metrics: (a) r2 (original y-scale -1000 - 1), (b) ME (original y-scale -5 - 5), (c) MAE (original y-scale 1 - 100), (d) RMSE (original y-scale 1 - 70)32

Fig. 10: Localization of rain gauge stations between 1991 and 2010.. 33

Fig. 11: Spatial distribution of seasonal rainfall indices based on 20-year (1991-2010) averages: (a) Onset, (b) Cessation, (c) Seasonal rainfall amount, (d) Rain days, (e) Wet days 20, (f) Wet days 50, (g) Dry days, (h) DSC15... 34

Fig. 12: Spatial distribution of seasonal rainfall indices based on 20-year (1991-2010) standard deviations (Std): (a) Onset, (b) Cessation, (c) Seasonal rainfall amount, (d) Rain days, (e) Wet days 20, (f) Wet days 50, (g) Dry days, (h) DSC15. 35

Fig. 13: Average cotton yield of SODECOTON collection points between 2007 and 2010. 36

Fig. 14: Average cotton yield of SODECOTON sectors between 1991 and 2010. 36

Fig. 15: Cotton yields (kg/ha) vs (a) mean WS1 for the sector dataset (b) median WS1 for the sector dataset, (c) Seasonal rainfall amount for the collection point dataset. 37

Fig. 16: (a–i) Spatial relationships between cotton yields (kg/ha) and cessation, DSC15, onset, rain days, season length, seasonal rainfall amount, wet days 20, wet days 30, WS1 for the sector dataset based on 20-year median values, using OLS 40

Fig. 17: (a–g) Temporal relationships between cotton yields (kg/ha) and cessation, dry days, DSC10, rain days, relative dry days, relative rain days, seasonal rainfall amount for the sector dataset based spatial mean values, using OLS 41

Fig. 18: (a) Trend relationships between cotton yields (kg/ha) and Cessation 43

LIST OF TABLES

Table 1: Statistical criteria used for the accuracy assessment.. 20

Table 2: Seasonal rainfall indices based on daily rainfall during the rainy season....................... 22

Table 3: Interpolated rainfall datasets 33

1. INTRODUCTION

The Sudano-Sahelian region area of Cameroon with a relatively high population density and high growth rate, relies heavily on rainfed agriculture. This significantly shapes the local socio-economic landscape (E. Molua & Lami, 2009), yet only 43% of farmers achieve food self-sufficiency in this way (Mbétid-Bessane et al., 2006). In the 1980s northern Cameroon already faced significant food shortages because of severe droughts (Penlap et al., 2004). Those facing food deficits often rely on the production of cotton to meet their financial and food needs, since it presents 60% of agricultural revenues, standing out as the primary cash crop and vital income source for many. (Mbétid-Bessane et al., 2006).

Cotton is a heliophilous plant (significant demand for sunlight and warmth) that throughout its growth cycle needs a minimum of around 700 mm of water. Typically a cotton plant's growth cycle ranges from 150 to 170 days with a daily evapotranspiration demand between 1 to 2,5 mm in the early stages and 6 to 10 mm during the blossoming stage (Ezan et al., 1998).

In Cameroon, the cotton industry was introduced in 1950 by the Compagnie Française de Développement des Textiles (CFDT) (Folefack et al., 2011). In northern Cameroon since 1974 it is overseen by the Société de Développement du Coton du Cameroun (SODECOTON), ultimately representing 202,000 producers by 2014 (SODECOTON, 2022b). The industry has seen significant growth, reaching a peak production of 300,000 tons in 2004. Historically, the success of Cameroon's cotton sector has been attributed to mutual commitment and contractual agreements between cotton companies and producers, along with stable management and a long-term vision that avoided the restructuring seen in other African countries (Folefack et al., 2011).

However, climate change is likely to negatively affect cotton productivity (Sultan et al., 2009) and modify cotton growth conditions (Gérardeaux E. et al., 2018; Gérardeaux et al., 2013), as well as regional resilience, due to increased year to year climate variability, unpredictable seasons, and more frequent heavy rains and droughts (Field et al., 2012; E. L. Molua, 2006). Extreme weather events, such as heavy rains and droughts, exacerbate the socio-economic challenges, impacting human communities, the environment, and the economy (Tamoffo et al., 2023). The variability of wet and dry periods can even lead to human migration due to changes in long-term rainfall patterns, significantly impacting economic and demographic aspects (Beauvilain A., 1996).

This vulnerability to climatic hazards and the regions dependance on cotton, has made northern Cameroon a critical area for scientific study.

To understand how rainfall affects the cotton yield in northern Cameroon, we will analyze the statistical links between rainfall indices and yield values. For this we will first look at previous works around the topic, before creating daily interpolated rainfall maps used to calculate yearly rainfall indices that will then be compared to cotton yields via Exploratory Data Analysis and model fitting. Finally, we will discuss our results, as well as their limits and possible perspectives, before concluding on the subject.

2. STATE OF THE ART

The Sahelian region, already recognized for having the highest interannual rainfall variability globally over the last century, is experiencing increasingly strong interannual rainfall variability (Joël et al., 2015; Nicholson, 2000). These changes affect soil moisture, vegetation cover, and albedo, altering large-scale atmospheric patterns and reinforcing irregular rainfall anomalies (Nicholson, 2000). Bouba L. et al. (2017) and Vondou et al. (2021) analyzed the trends in rainfall and both discovered an increased tendency of mean annual rainfall in northern Cameroon. Further research by Njouenwet et al. (2022), on the spatiotemporal variability and trends of extreme rainfall events, using data from fifteen stations, revealed a decrease in the annual number of rainy days from the North to the Far North and a slight delay in the onset of the rainy season, but a rising intensity of rainfall, hinting towards an increase of rainfall as well. Collectively, these studies suggest a general trend of increasing annual average rainfall in the Sudano-Sahelian zone of Cameroon.

The agricultural productivity in northern Cameroon highly depends on weather and climate conditions, rendering the region particularly vulnerable to climate change (E. L. Molua, 2006). It closely depends on factors such as rainfall availability, onset and retreat date, as well as the duration of the rainy season, even if the role of rainfall variability is strongly reduced in farmers' exploitations where other non-climatic factors such as human management, biotic stresses, pests, etc., impact crop productivity (Sultan et al., 2009). Sultan et al. (2009) demonstrated in a study on the influence of rainfall on cotton yields in northern Cameroon, that cotton productivity significantly decreases with early (May–June) and late (September–October) season rainfall deficits, which shorten the rainy season length. Across the semi-arid Sahel too, vegetation is notably affected by variations in rainy day frequency, as well as the onset and retreat date of the rainy season. Especially this frequency of rainy days, as well as the occurrences of heavy rainfall events influence the relationship between growing season vegetation productivity and climate factors (W. Zhang et al., 2018). Cotton yields are particularly sensitive to both the overall seasonal rainfall amounts and heavy rainfall events (Njouenwet et al., 2021). In addition to those events, the impact of gradual changes in consecutive dry days has been notable, the two often leading to water supply challenges and soil degradation (M'Biandoun & Olina, 2009).

Sultan et al. (2009) suggest that a potential decrease in mean annual rainfall in the northern part of the cotton production area of Cameroon could reduce productivity and increase climate-related risks. In addition, climate change is expected to intensify extreme temperature and precipitation events in Cameroon (Tamoffo et al., 2023). The higher temperatures are projected to expedite crop maturation times without necessarily causing yield losses to a certain extent, after which productivity is expected to decrease because of physiological considerations. Nevertheless, to cope with these accelerated phenological cycles of cotton, farmers may need to consider adjustments in planting dates and cultivar phenology (Gérardeaux E. et al., 2018; Gérardeaux et al., 2013).

Despite these challenges, the cotton production region of Cameroon requires more studies on regional variability and trends in rainfall, as well as extreme rainfall and drought. The issue here was not really the lack of data discussed in Field et al. (2012), but a poor remobilization effort of a wealth of data available at SODECOTON. Furthermore, the significant local variations in rainfall and cotton yield highlight the necessity for spatialization of data and for a better understanding of rainfall trends in the region.

To this aim, Geostatistical estimation methods, such as Kriging, have been found to provide more accurate interpolations than deterministic techniques like Inverse Distance Weighting (IDW), making them valuable for predicting climatic impacts on agriculture in the Sudano-Sahelian area (Dassou et al., 2016; Moral, 2010).

Northern Cameroon, with its historical challenges of drought and food shortages, remains a crucial area for understanding the impacts of climate change. The interplay of rainfall trends, agricultural practices, and socio-economic conditions illustrates a complex and dynamic system that requires continuous study and adaptive strategies to mitigate future risks.

3. OBJECTIVES AND HYPOTHESES

The primary objective of this study is to analyze the cotton yield based on rainfall data, addressing the pressing need to understand and anticipate the impacts of climate change on cotton production.

For this, several sub-objectives were defined:

I. Ensure the integrity and reliability of the rainfall data.
II. Enhance the understanding of spatial and temporal rainfall distribution through spatial interpolation techniques and rainfall indices calculations.
III. Ensure the integrity and reliability of the cotton yield data.
IV. Provide valuable insights into the relationships between rainfall indices and cotton productivity.

Leveraging two extensive databases - one capturing daily rainfall observations across several hundred rain gauge stations, and the other detailing cotton yield data from numerous collection points and sectors - the study aims to equip farmers and scientists with valuable insights for effective adaptation strategies.

Since the study focuses on cotton specifically, the rainfall indices were tied to the rainy season, representing a critical period for cotton planting and growth. Rainfall is a major factor in the construction of cotton yields and was the only climatic variable considered in this study. Other meteorological and climate variables were not taken into consideration in this study due to data availability, with historical data not being available and satellite data being of poor resolution, given the extent of the region of interest, as well as being unreliable with only few synoptic stations present in this area of Africa. Furthermore, at the large scale offered by these products, it would be difficult to compare to the obtained cotton yield data at a close-up scale and give questionable results to the objective of the study.

4. MATERIAL AND METHODS

4.1. Overall Methodology

The framework displayed in Fig. 1 was designed to accomplish the objectives of the study.

The method was built on two data sources: the NoCORA rainfall dataset (mainly composed of SODECOTON rain gauge data) and the cotton yield data given by SODECOTON. Since the geographic coordinates of the rain gauges, as well as the cotton production locations, did not necessarily match, the decision was made to first interpolate the rainfall data. It is processed to create a dataset of daily interpolated rainfall maps used to calculate yearly index maps covering the whole study area. The interpolation method is selected by implementing Leave-One-Out Cross-Validation (LOOCV) and evaluating different error metrics calculated from the LOOCV results, as well as by executing a paired t-test with these error-metrics. By these means, the cotton yield locations can be colocalized to the corresponding index coordinates to enable efficient exploratory data analysis and model fitting.

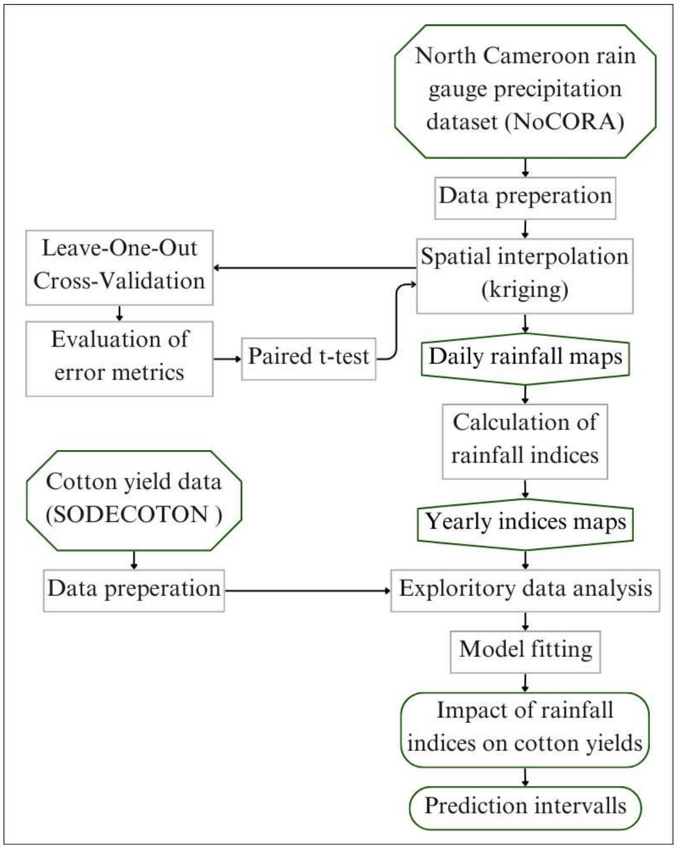
North Cameroon rain gauge precipitation dataset (NoCORA)
Data preperation
Leave-One-Out Cross-Validation
Spatial interpolation (kriging)
Evaluation of error metrics
Paired t-test
Daily rainfall maps
Calculation of rainfall indices
Cotton yield data (SODECOTON)
Yearly indices maps
Data preperation
Exploritory data analysis
Model fitting
Impact of rainfall indices on cotton yields
Prediction intervalls

Fig. 1: Overall methodology framework

4.2. Study Area

The North of Cameroon comprises two administration regions - North and Far North - lying in the Sudano-Sahelian zone of the country, between 7◦ N and 13◦ N latitude and 11◦30’ E and 16◦ E longitude. The topographically flat regions (cf. Fig. 2), with a surface area of 100 km^2, accommodate a population of around 6.4M inhabitants as of 2015, at a density of around 64/km^2 (Brinkhoff, 2020).

The climate of this zone can be divided into two: north of 10◦ N the Sahelian area with warm semi-arid climate and south of 10◦ N the Sudanian area with tropical savanna climate (Fick & Hijmans, 2017; National Geographic Society, 2024). Both areas follow a pattern of dry and wet season, influenced by the thermodynamic properties of the African Monsoon, though the Sahel-type climate brings rain from May to October, while the Sudan-type climate rainfall arrives in June and ceases in September. Following this pattern, the rivers underly a tropical regime with high water during the wet season and low water during the dry season. Characterized by a monomodal regime, the annual average rainfall increases from north to south with elevation, ranging from around 900mm in the north to around 1300mm in the south. The temperature during the growing season reaches from an average minimum of 21.8◦C to an average maximum of 34.8◦C.

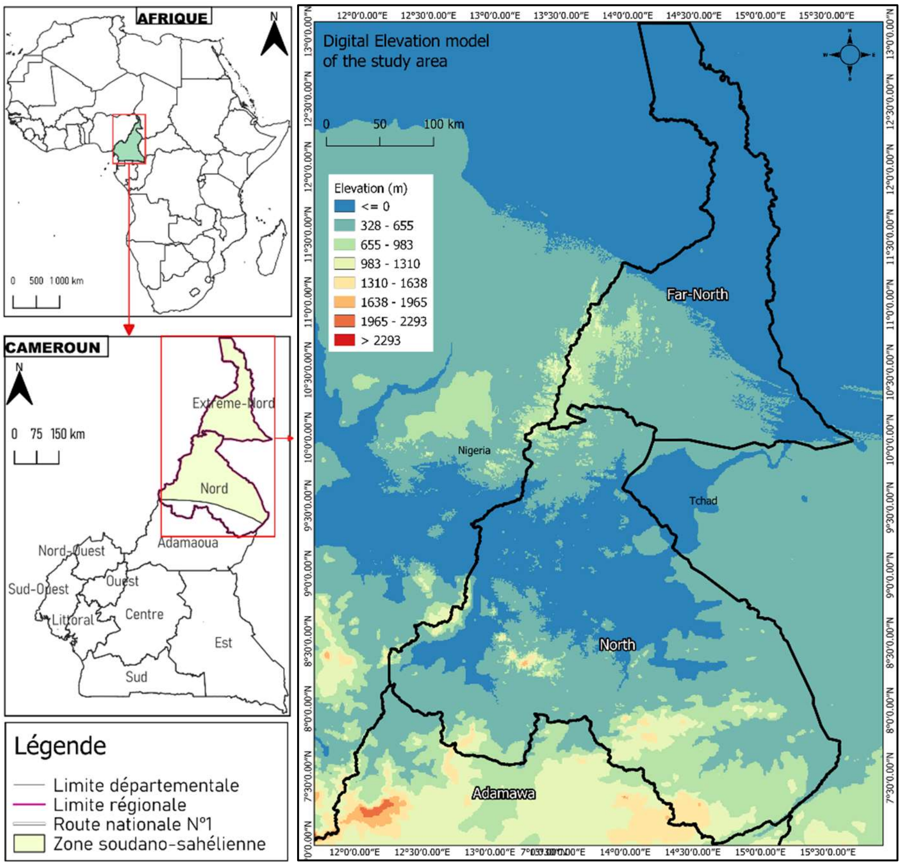

Fig. 2: Digital Elevation Model of northern Cameroon
Source figures on the left: Victor Nenwala, personal communication

Agriculture is one of the regions' main activities, overviewed by organizations such as the SODECOTON and the Société d'Expansion et de Modernisation de la Riziculture dans la ville de Yagoua (SEMRY). Though the primarily raised crop may vary from area to area, the most common ones are cotton, millet, sorghum, maize, rice, groundnuts and onions.

4.3. Data

4.3.1. Rainfall Data

The rainfall data used for the study was published as the Northern Cameroon Observed Rainfall Archive (NoCORA, doi: 10.5281/zenodo.10156437) by Lavarenne et al., (2023). The data was collected from 418 rainfall stations dispersed in the North and Far North regions of Cameroon, through rain gauge instruments, for the period of 1927 to 2022 (cf. Fig. 3). Several sources contributed to the records, including SODECOTON, Robert Morel (IRD), the Lagdo Hydroelectric Power Station, the Global Historical Climatology Network (GHCN) and the Trans-African Hydro-Meteorological Observatory (TAHMO). Every data entry is accompanied by the corresponding geographic coordinates of the rain gauge. Some stations from outside of the borders of Cameroon were also included in this dataset to serve as "anchor points" for the interpolation process, reducing the surfaces concerned only by extrapolation near the borders.

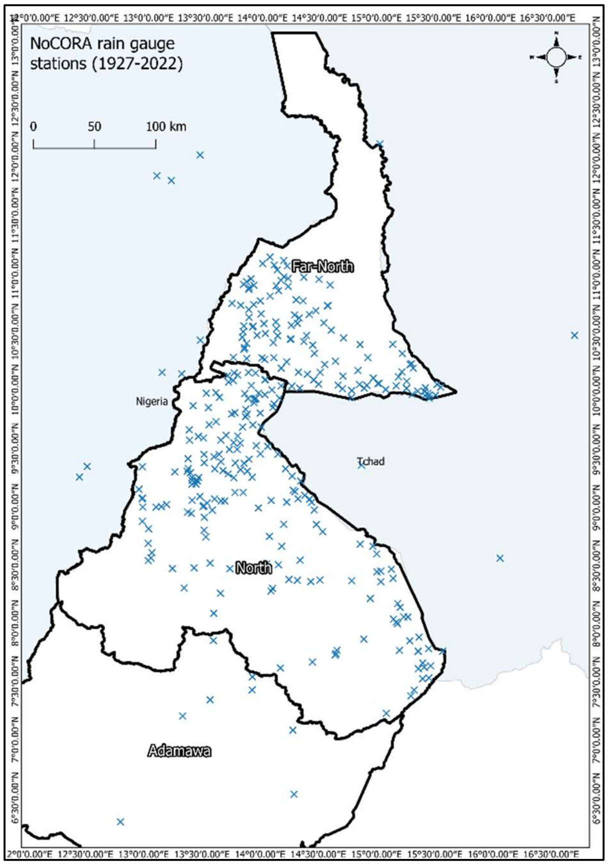
NoCORA rain gauge stations (1927-2022)
0
50
100 km
Far-North
Nigeria
Tchad
North
Adamawa
12°0'0.00"E 12°30'0.00"E 13°0'0.00"E 13°30'0.00"E 14°0'0.00"E 14°30'0.00"E 15°0'0.00"E 15°30'0.00"E 16°0'0.00"E 16°30'0.00"E

Fig. 3: Localization of NoCORA rain gauge stations between 1927 and 2022.

Given that the data was provided by different sources, an extensive preparation part was carried out by Lavarenne et al. (2023) for the construction of a complete dataset. Records were provided in either numeric or paper format, requiring a preliminary compilation before further data cleaning. Inconsistencies between the different records were reduced through standardization of station names and verification of coordinate accuracy, eliminating duplicates at the same time. In addition, missing coordinates were estimated employing platforms such as Google Maps and MapCarta, using the station name as point of reference.

The observations exhibit a strong variability in space and time, with the number of stations changing depending on the day. The records provided by SODECOTON only consider the rainfall for the period of March until October, during the rainy season, while rainfall during the dry season is not included.

4.3.2. Cotton Data

The cotton data used in this study was provided by SODECOTON, which collected the data at two different geographical levels:

1. Collection points: Points where trucks collect cotton seed, representing the cotton market.
2. Sectors: Cotton producer regrouped by larger scale areas

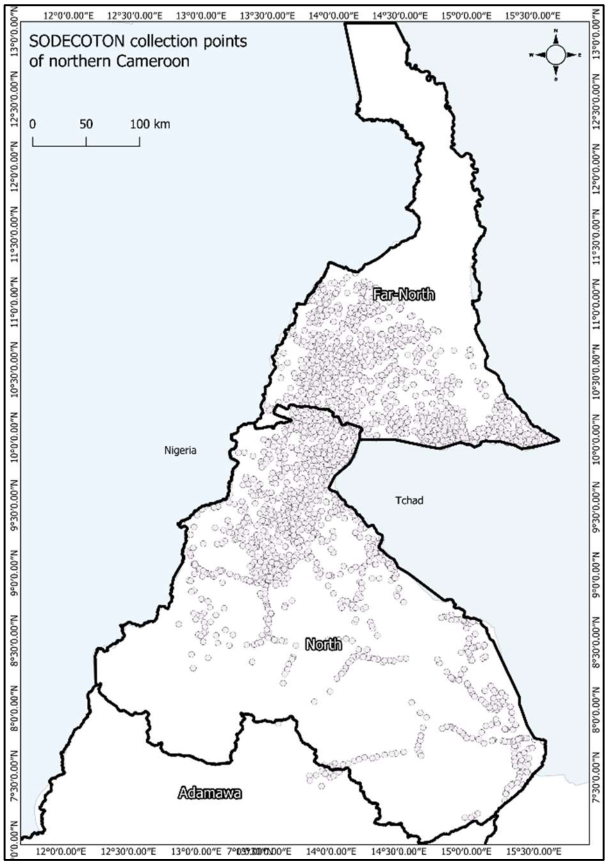

Fig. 4: Localization of SODECOTON collection points between 2007 and 2010.

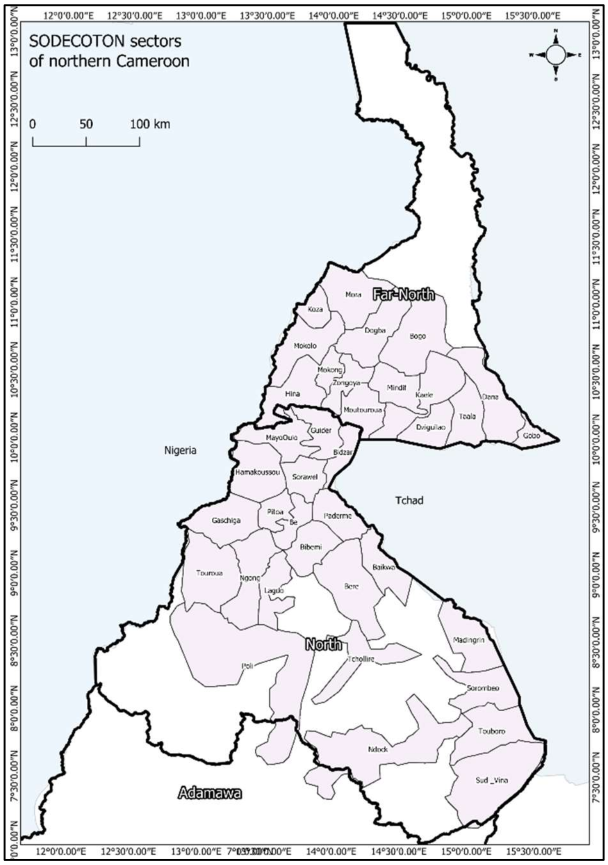

Fig. 5: Localization of SODECOTON sectors between 1991 and 2010.

A collection point dataset, made from data gathered by SODECOTON, was assembled by Antoine Leblois as part of his 2013 study about the potential of weather index-based insurance to mitigate risk for cotton farmers (Leblois et al., 2014). This original dataset included collection point names, as well as cotton yield values in kg/ha, determined from surface areas and cotton production values for 1883 collection points over a period of 4 years, from 2007 to 2010. A separate file contained the geographic location of every collection point with the related name.

In addition, Antoine Leblois assembled a sector dataset, once more with SODECOTON's observed cotton yield data in kg/ha. This original sector dataset contained several files with mean annual yields for 43 sectors, spanning 28 years from 1983 to 2010.

4.4. Methods

4.4.1. Generating daily rainfall maps from daily rain gauge data using Kriging interpolation

For the preparation of the rainfall data, days with data for only one station were filtered out, since they would produce poor, non-representative interpolation results. Therefore, all data entries before 1948 were excluded and only the subsequent period until 2021, englobing 395 stations, was interpolated.

Previous studies have shown that in the Sudano-Sahelian area of Cameroun, for daily map interpolation, Kriging gives better global predictions than Inverse Distance Weighing (IDW) (Dassou et al., 2016), with several studies relying on Kriging (Djoufack et al., 2012; Njouenwet et al., 2022; Njouenwet et al., 2021). Therefore, to generate high resolution daily rainfall maps, two kriging methods – Ordinary Kriging (OK) as well as Universal Kriging (UK) – were tested. Interpolation itself was performed under Python 3.11 using GSTools version 1.5.1 (Müller et al., 2022).

The objective of kriging is to use observed values (o_i) at fixed data points (x_i) to derive the value (o_0) of a field at a grid point (x_0) by using a weighted linear combination of the observed values:

$$o_0 = \sum_{i=1}^{n} w_i \times o_i$$

The weights (w) can change according to the location of x_0, as well as the variogram model applied and are calculated differently depending on the kriging method.

OK is a linear estimation method that assumes a constant mean. It is theoretically unbiased since it pursues to have a mean residual error equal to zero and in addition, OK aims to minimize the error variance. The equation for w resulting from the OK system can be expressed as follows:

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{pmatrix} = \begin{pmatrix} c(x_1, x_1) & \cdots & c(x_1, x_n) & 1 \\ \vdots & \ddots & \vdots & 1 \\ c(x_n, x_1) & \cdots & c(x_n, x_n) & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} c(x_1, x_0) \\ \vdots \\ c(x_n, x_0) \\ 1 \end{pmatrix}$$

where:

$c(x_i,x_j)$ = covariance of the given observation μ = Lagrange multiplier

For UK a deterministic trend is added to the method, allowing the mean to vary in different locations. In this study, a linear trend was applied, which assumes that the mean changes linearly with the spatial coordinates. The linear trend function of the x, y coordinates ($f(x_i)$ and $f(y_i)$) is added to the previous equation:

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \\ \mu_1 \\ \mu_1 \\ \mu_1 \end{pmatrix} = \begin{pmatrix} c(x_1, x_1) & \cdots & c(x_1, x_n) & f(x_1) & f(y_1) & 1 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ c(x_n, x_1) & \cdots & c(x_n, x_n) & f(x_n) & f(y_n) & 1 \\ f(x_1) & \cdots & f(x_n) & 0 & 0 & 0 \\ f(y_1) & \cdots & f(y_n) & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}^{-1} \begin{pmatrix} c(x_1, x_0) \\ \vdots \\ c(x_n, x_0) \\ f(x_0) \\ f(y_0) \\ 1 \end{pmatrix}$$

As kriging is a technique that uses variogram models to interpolate data, the two methods (OK and UK) were implemented using nine different variogram models: Circular, Exponential, Gaussian, Matern, JBessel, Rational, Spherical, Stable and SuperSpherical.

Table 1: Statistical criteria used for the accuracy assessment

Statistical Criteria	Definition
Coefficient of determination (r^2)	$r^2 = 1 - \frac{(\sum_{i=1}^{n}(o_i - \hat{z}))^2}{(\sum_{i=1}^{n}(z_i - \hat{z}))^2}$
Mean Error (ME)	$ME = \frac{1}{n} \sum_{i=1}^{n} (z_i - o_i)$
Mean Absolute Error (MAE)	$MAE = \frac{1}{n} \sum_{i=1}^{n} \lvert z_i - o_i \rvert$
Root Mean Square Error (RMSE)	$RMSE = \left[\frac{1}{n} \sum_{i=1}^{n} (z_i - o_i)^2\right]^{1/2}$

We evaluated the obtained results by applying a Leave One Out Cross Validation (LOOCV) approach, where one data point is consecutively left out of the interpolation procedure, while the interpolated value at the missing point coordinates is logged to be compared with the missing point value, this procedure being replicated as many times as there are validation points (Longman et al., 2019). For the number of observations (n), the error between the observed value (o) versus predicted value (z) is then assessed by employing a series of error metrics as described by Willmott, 1982 and Isaaks & Srivastava, 1989.

The coefficient of determination (r^2, with $\hat{p}$ being the mean of p-values) serves as a first indicator of the reliability of the model (Willmott, 1982). The bias of the model and its degree is described by the Mean Error (ME) (Isaaks & Srivastava, 1989). The Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are both "among the 'best' overall measures of model performance, as they summarize the mean difference in the unit of o and p" (Willmott, 1982), with the difference that RMSE emphasizes extreme values while MAE is less sensitive to those (Willmott, 1982).

For the best performing models and interpolation method, a paired t-test based on interpolation error metric values was carried out additionally, to identify which models present significantly different scores. The null hypothesis states that the mean difference between paired observations is zero. If the p-value is less than 0.05, the null hypothesis is rejected, indicating a statistically significant difference (Xu et al., 2017). As missing days were present in all datasets, due to certain days having less than two data-entries and some days not being able to be interpolated, a cross-filling method was used to assemble a composite of interpolated datasets with no statistically significant difference. In practice, this means that we completed missing days of the best interpolation method with available days of the second-best interpolation method to reconstruct incomplete time-series. In addition, in maps that produced negative interpolated values, these negative values were replaced with zeros.

Using the best performing variogram models, daily rainfall maps were interpolated for the period 1948-2022 at a resolution of 0.01°, or pixels of 1,11 x [1,08 ; 1,10] = [1,19 ; 1,22] km² at the latitude [9 ; 13] °N.

4.4.2. Rainfall Indices

Rainfall indices are parameters used to describe the amount, frequency, intensity and distribution of rainfall over a certain period and area, tracking events and patterns.

Regarding the available years of observed cotton yield data, the daily interpolated rainfall maps were processed for a period of 20 years, from 1991 to 2010, with a total of 85 stations included in the initial interpolation during this time span. The rainfall data was mainly collected from March to October during the rainy season, and the data for other months was sparse and only captured occasional rain events, therefore exclusively the interpolated maps for the rainy season were used since the dry season produced poor and unreliable results.

Table 2: Seasonal rainfall indices based on daily rainfall during the rainy season

Index name	Definition	Unit
Onset	Starting date of the rainy season	Day of year
Cessation	Retreat date of the rainy season	Day of year
Season length	Number of days between onset and cessation	Days
Seasonal rainfall amount	Rainfall amount during the rainy season	mm
Rainy days	Number of days with rainfall >1 mm	Days
Relative rainy days	Percentage of rainy days during the rainy season	%
Dry days	Number of days with rainfall <1 mm	Days
Relative dry days	Percentage of dry days during the rainy season	%
Wet days 20/30/40/50	Number of days with rainfall >20/30/40/50 mm	Days
Relative wet days 20/30/40/50	Percentage of wet days 20/30/40/50 during the rainy season	%
Heavy rainfall days (WS1 90P)	1 day with rainfall >90thpercentile of daily rainfall	Days
Wet spells cumulative 10/15/20 (WSC10/15/20 90P)	10/15/20-days rainfall >90thpercentile of 10/15/20-day cumulative rainfall	Nbr of events
Long dry spells (DSl)	8 - 14 consecutive dry days	Nbr of events
Extreme long dry spells (DSxl)	Consecutive dry days exceeding 15 days	Nbr of events
Dry spells cumulative 10/15/20 (DSC10/15/20)	10/15/20 days with less than 10/15/20 mm of rainfall	Nbr of events

Tied to the rainy season, simultaneously representing the growing season of cotton and a critical period for crop planting and growth, a total of 25 rainfall indices were calculated for the days of the rainy season (cf. Table 2). These indices were computed pixel-wise, based on the interpolated daily rainfall maps, per year. Days exhibiting >1mm of rainfall per day mark rainy days and <1mm of rainfall per day mark dry days. Four further thresholds of rainfall were defined, with >20mm, >30mm, > 40mm and >50mm of rainfall, indicating wet days as defined by Maidment et al. (2017).

Wet spells (WS) and dry spell (DS) indices were derived from (Fall et al., 2019). Dry spells depend on duration while wet spells also depend on intensity.

The methodology used to define the onset and retreat date of the rainy season was developed by Liebmann et al. (2012), where for each grid point the sum of the daily rainfall minus the climatological annual daily average is calculated. The day after the absolute minimum marks the onset date, indicating the start of consistent above-average precipitation. The maximum in this value establish the retreat date, indicating the transition to below-average precipitation. .

The output of calculation is one map per index and per year.

Also, to be able to analyze the spatial distribution of the rainfall indices and their variability in space, the average index value and standard deviation for the analyzed period were calculated for each index respectively.

4.4.3. Cotton data preparation

For the collection point dataset (cf. Fig. 4), quite detailed in space, the data was refined with the help of Ibrahim Njouenwet for the use of this study. The names of the different collection points in the two files did not consistently match due to different spelling, therefore they were standardized by uniformizing their spelling, before being attributed with the corresponding geographic coordinates. Non-attributed-number (Nan) values were then filtered, and duplicate lines dropped, creating a complete collection point dataset in csv format with 1883 collection points over 4 years, from 2007 to 2010.

For the operational use of the sector dataset (cf. Fig. 5), more detailed in time, data preparation was carried out as well. Since the years 1984, 1986-87 and 1989-90 were missing, the choice was made to only keep the data between 1991 and 2010, to insure accurate and reliable results. In addition, as shown in Fig. 5 only 37 sectors were presented, the entries of 6 sectors had to be removed, due to a restructuring of the SODECOTON sector limits. Names needed to be standardized for each sector due to differences in spelling, before the related cotton yield could be attributed cartographically to the corresponding polygon with the updated limits. Nan values and duplicate lines were removed, thus creating a sector dataset in shape format with 37 sectors for a time interval of 20 years, from 1991 to 2010.

4.4.4. Exploratory data analysis of the relationships between rainfall indices and cotton yield, and Model fitting

To observe the possible statistical relationships between the cotton yield dataset and our 25 rainfall indices, we first established scatterplots for an exploratory data analysis of the yield and rainfall index variables.

The cotton yield dataset (1883 collection points for a period of 4 years) contained several outliers restricting a proper first interpretation of the results. Therefore, the choice was made that data points over the 95th percentile and under the 5th percentile of yield values were filtered out. 25 scatterplots could then be produced, one for each index-yield combination.

Since the cotton yield values for the sector dataset (37 sectors for a period of 20 years) were not attributed to a specific pixel coordinate, but to a polygon area, mean and median aggregation functions were applied to the rainfall index maps to obtain aggregate rainfall index values for each sector. Hence, we produced 50 scatterplots for this dataset, 25 mean index-yield combinations, as well as 25 median index-yield combinations.

For each index-yield combination of the two datasets (75 in total – 25 Collection point dataset, 50 sector dataset) we then calculated the Pearson coefficient and the p-values:

The Pearson correlation coefficient (r) is a measure of the linear relationship between two sets of data (here R and Y), quantifying the degree to which pairs of data points deviate from their respective means in the same direction. The coefficient is calculated as follows:

$$r = \frac{\sum(R_i - \bar{R})\,(Y_i - \bar{Y})}{\sqrt{\sum(R_i - R)^2}\,\sqrt{\sum(Y_i - \bar{Y})^2}}$$

It ranges from -1 to +1, where +1 indicates a perfect positive linear relationship, -1 indicates a perfect negative linear relationship, and 0 indicates no linear relationship (Nettleton, 2014). A p-value can be associated to it, which helps assess the statistical significance of the observed correlation coefficient. As described in 4.4.2 Rainfall Indices, the interpretation of the p-value is based on the null-hypothesis. In this case a p-value <0.01 indicates a statistically significant relationship, while a p-value >0.01 indicates statistically non-significant relationships.

Different model fitting approaches were then applied to both the collection point and sector datasets (cf. Fig. 7), to further analyze the statistical relationship between cotton yield and indices, as well as to enable the possibility of predicting cotton yields through projected rainfall data.

Simple linear regression - Ordinary Least Square (OLS)

The Ordinary Least Squares (OLS) is a commonly applied linear regression method, estimating the parameters of a linear relationship by minimizing the sum of the squared residuals, where residuals are the differences between observed and predicted values.

Given a linear model:

$$Y = m + ry.R + \varepsilon$$

where in this case cotton (Y) is the dependent variable and the rainfall index (R) is the independent variable. The variable of ry is the slope coefficient, representing the yield response to seasonal rainfall trends, ε is the residual error, accounting for unexplained variance in the relationship and m is the intercept, representing the average yield change due to factors other than seasonal rainfall

The objective is to find the values of m and ry that minimize the sum of squared residuals:

$$\sum_{i=1}^{N} \varepsilon_i^2$$

where ε_i is the residual for the ith observation, defined as the difference between the observed value and the value predicted by the model. The minimization of this sum ensures the best linear fit to the data.

To achieve this, the residuals ε_i must be calculated for each observation, before each residual is squared to prevent positive and negative differences from canceling each other out. The squared residuals are then summed and the parameters m and ry optimized to minimize this sum (Dismuke & Lindrooth, 2005).

We performed OLS regression between the 25 rainfall indices and our yield datasets, using the statsmodels version 0.14.2 library (Waskom, 2021), applying three dimensions: solely spatial, solely temporal and spatiotemporal (cf. Fig. 6).

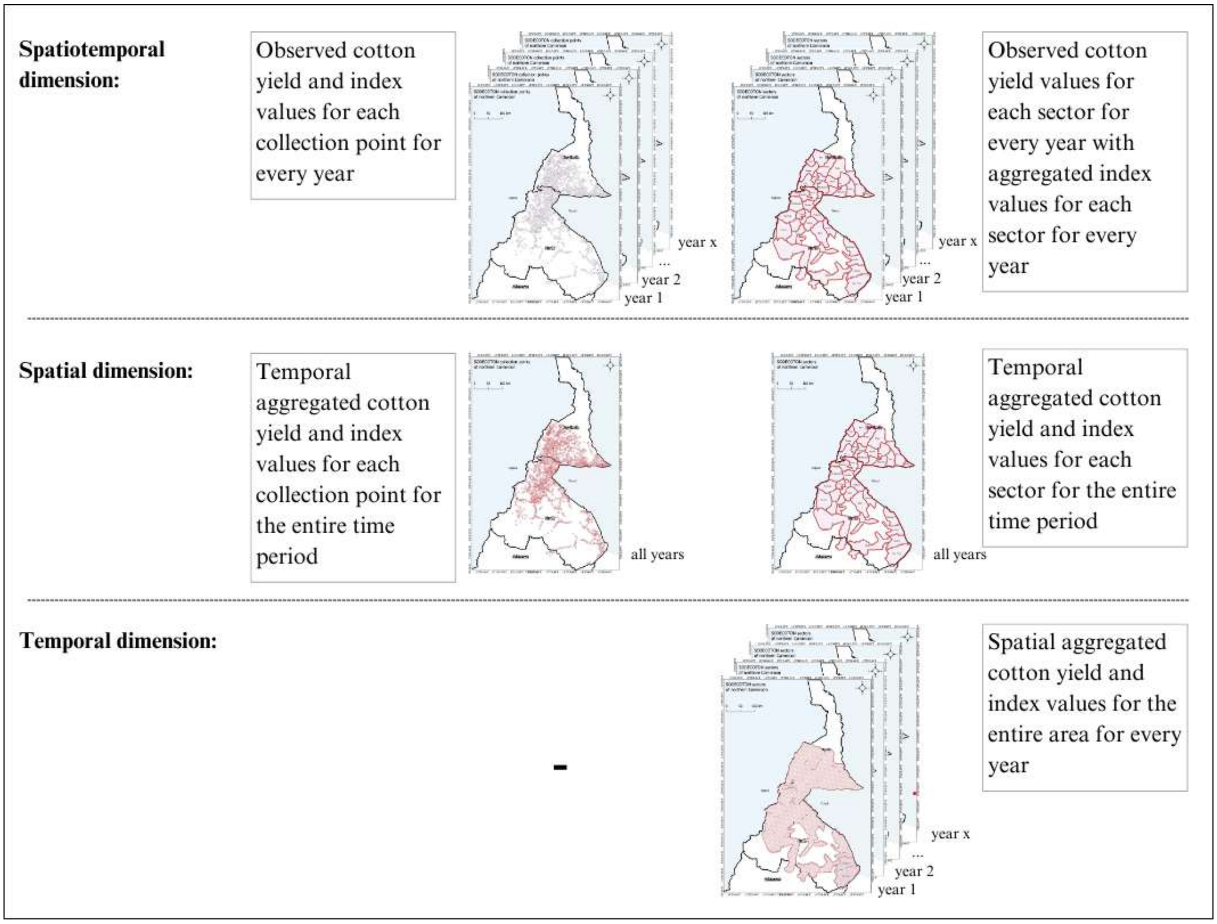
Spatiotemporal dimension:
Observed cotton yield and index values for each collection point for every year
year x
...
year 2
year 1
year x
...
year 2
year 1
Observed cotton yield values for each sector for every year with aggregated index values for each sector for every year
Spatial dimension:
Temporal aggregated cotton yield and index values for each collection point for the entire time period
all years
all years
Temporal aggregated cotton yield and index values for each sector for the entire time period
Temporal dimension:
-
year x
...
year 2
year 1
Spatial aggregated cotton yield and index values for the entire area for every year

Fig. 6: Dimensions applied for model fitting

A spatial dimension emphasizes the differences across space, helping to understand how different spatial data points perform over a specific time span, highlighting the production capacities for every sector while smoothing out year-to-year weather variations. For every data entry in space (1883 for the collection point dataset and 37 for the sector dataset), all values over time were taken and mean, as well as median aggregation functions were applied to obtain spatial aggregate yield and index values. Therefore, each dataset produced 100 results: 25 mean index-mean yield combinations, 25 mean index-median yield combinations, 25 median index-mean yield combinations and 25 median index-median yield combinations.

The temporal dimension reduces the influence of local characteristics and variability, helping to understand the influence of rainfall indices on cotton yield for the whole region over a longer period, highlighting temporal trends. Here, for each year, all values in space are taken into mean and median aggregation functions to obtain temporal aggregate yield and index values. Due to the short time span of the collection point dataset (4 years), the temporal dimension was only applied to the sector dataset (20 years), producing another 100 outputs with the same combinations as for the spatial dimension.

In a spatiotemporal dimension, both spatial and temporal dimensions are integrated, providing a comprehensive view of how cotton yield and its relationship with rainfall indices vary across space over time. This approach emphasizes both the geographical differences in the performance of spatial data points and the temporal trends in cotton yield influenced by rainfall patterns. Both datasets were implemented with each data entry in space for every year as described for the Exploratory Data Analysis, with no supplementary mean or median aggregation functions. Furthermore, for the sector dataset which depends on values attributed to polygons and not specific pixel coordinates, we also tried another strategy for the spatiotemporal dimension, where a cropland mask (Karra et al., 2021) was applied. Only the values of the pixels falling into a crop category were retained, to better estimate the areas of cotton production.

Various adaptations of the linear regression equation were then implemented for both datasets, apart from the slope/slope method which could only be applied to the sector dataset, given the short time interval of the collection point dataset. Moreover, the linear-log model, the relative approach and the slope/slope approach were only performed for the spatiotemporal dimension, while the First Difference method was implemented for the temporal dimension as well.

- **First difference**

The purpose of the first difference method is to consider the influence of omitted values which are attributed to the effect of rainfall indices, instead of other factors, such as the influence of CO_2.

For each data point in space the observed yield value on the day j is subtracted from the yield value of the previous day, j-1, thus only being applicable to the temporal and spatiotemporal dimension. The same procedure is repeated for the interpolated rainfall index values as follows:

$$\Delta Y = Y_{j+1} - Y_j$$

and

$$\Delta R = R_{j+1} - R_j$$

where ΔY describes the difference in yield and ΔR the difference in index values.

The differences ΔR and ΔY are implemented in OLS regression, to analyze how the changes in rainfall indices influence the changes in cotton yield:

$$\Delta Y = m + ry.\Delta R + \varepsilon$$

where in this case, m represents the yield change due to other factors than R.

- **Linear-log model**

In a log-linear model, the relationship between the rainfall indices and the cotton yield is expressed with R in logarithmic form, adjusting the linear regression equation as described underneath:

$$Y = m + ry.\log{(R)} + \varepsilon$$

This implies that multiplying R by the natural exponent e will result in a corresponding change in Y by ry units:

$$\log(R) + 1 = \log(R) + \log(e) = \log{(eR)}$$

The logarithmic transformation of R helps to address nonlinear relationships between variables. This transformation is particularly useful for normalizing heavily biased variables, making it more suitable for analysis. A log-transformed variable can better approximate a normal distribution, allowing for more accurate and reliable statistical conclusions (Benoit, 2011).

- **Relative values**

The relative values approach standardizes data across different sectors, making it easier to compare and detect long-term trends and anomalies, detaching from the average cotton yield.

For the entire time span of each dataset, the average value of the cotton yield (Y_b) is aggregated. This average value presents the baseline, equaling 100%. For each Y data entry, the relative values regarding the baseline value are then calculated as follows:

$$Y_r = Y/Y_b$$

where Y_r represents the relative cotton yield value.

Linear regression is then applied, using Y_r and R as dependent and independent variables respectively, giving the adjusted linear regression equation:

$$Y_r = m + ry.R + \varepsilon$$

- **Slope/ slope**

The slope/ slope method, first used by Zhang et al., 2015 and adapted to a Cameroonian context in 2021 by Njouenwet, serves to highlight the weak implications of index trends in cotton yields over time, focusing on long-term trends.

For each eligible sector, the slope of the cotton yield by year, as well as the index by year, are calculated via OLS, essentially representing cotton yield trends and rainfall indices trends. Sectors which have less than one data entry for the analyzed period are not considered in this method, since there are not enough points to implement a first linear regression analysis. Sectors with more than one data point will have an index slope value and a yield slope value attributed to them. To quantify how the trends in R impact trends in Y over time, OLS regression is then applied a second time, using the yield slope values and index slope values of all sectors as dependent and independent variable, giving the adapted equation:

$$Y_t = m + ry.R_t + \varepsilon$$

where Y_t represents the cotton yield trend and R_t the rainfall indices trends.

Multiple linear regression

Multiple linear regression is an extension of simple linear regression that incorporates multiple independent variables, in this case multiple indices, to predict the dependent variable, the yield, helping to understand their association to each other. The relationship between the variables is assumed to be linear, meaning the yield is modeled as a linear combination of the indices (Tranmer et al., 2020).

Simple linear regression was applied to all inter-index combinations (25 indices equaling 300 combinations of 2 indices), to find index-pairs with no statistically significant relationship between them as determined by the Pearson coefficient and corresponding p-value.

For the retained index combinations multiple linear regression was applied using the general equation:

$$Y = m + ry_1 . R_1 + ry_2 . R_2 + \varepsilon$$

where m represents the intercept and ry_1, ry_2 are the coefficients for the index variables R_1, R_2.

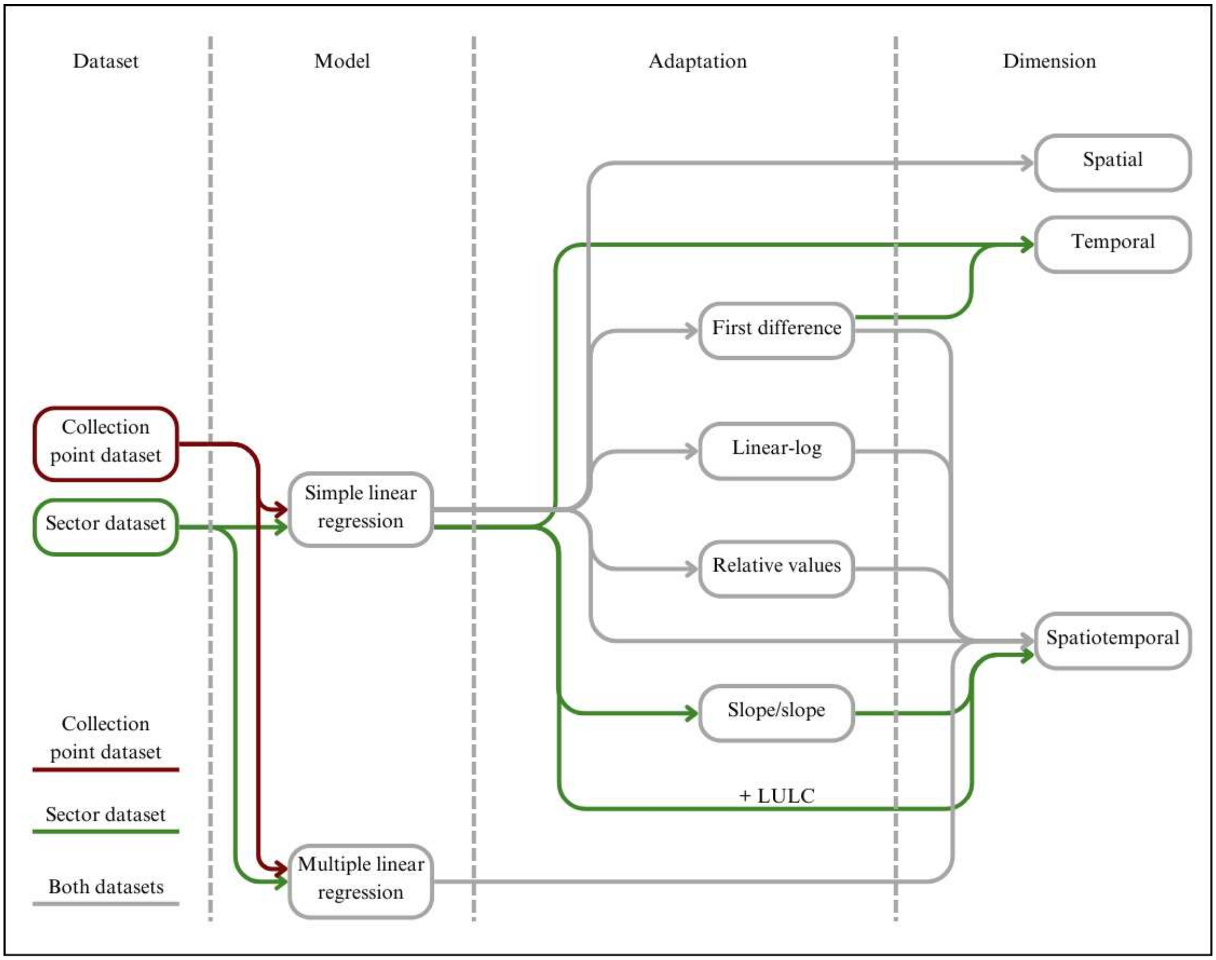

Fig. 7: Model fitting framework

Calculation of metrics

For every approach the Pearson correlation coefficient with their corresponding p-values were calculated once more.

In addition, the coefficient of determination (r^2), as well as the slope were given by the statsmodel output. Here, r^2 quantifies the proportion of the variance in the dependent variable Y that is predictable from the independent variables, the rainfall indices. Essentially, it is the squared

Pearson coefficient, measuring the goodness of fit of the model by indicating how well the rainfall indices explain the variation in the cotton yield. The slope quantifies the expected change in the dependent variable Y for a one-unit increase in the independent variables R. Through the integration of the p-value, statistically significant slopes ($p < 0.01$) can be identified, indicating a strong relationship between the variables (Burton, 2021).

5. RESULTS

5.1. Interpolated rainfall maps

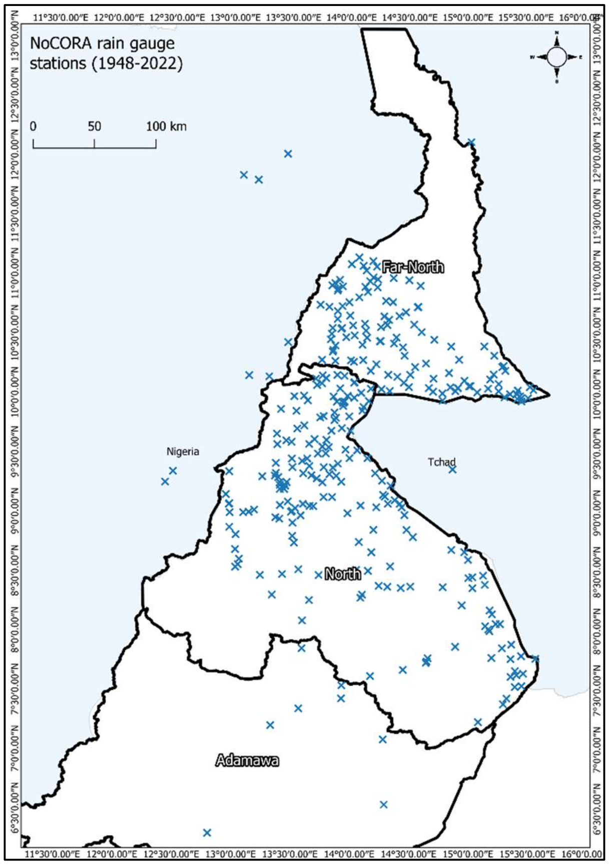

Fig. 8: Localization of rain gauge stations between 1991 and 2010.

The preparation of the rainfall data left us with the stations shown in Fig. 8, to compute LOOCV and the interpolation of daily rainfall maps.

The interpolation performance metrics resulting from the LOOCV of the two kriging methods are presented in Figure 9. They revealed that Circular, Exponential, Spherical and SuperSpherical variogram models for OK produce the best results. The figures show that OK presents as more reliable than UK and that all four models exhibit a low bias.

The paired t-test, which was carried out in addition, showed that there was no significant difference between OK Circular and OK Spherical regarding all error metrics (cf. Appendix 1).

The Spherical dataset using OK, which exhibited the least missing interpolation maps, was used as basis for the combined dataset, and completed by the Circular dataset using OK. Data access is presented in Table 3.

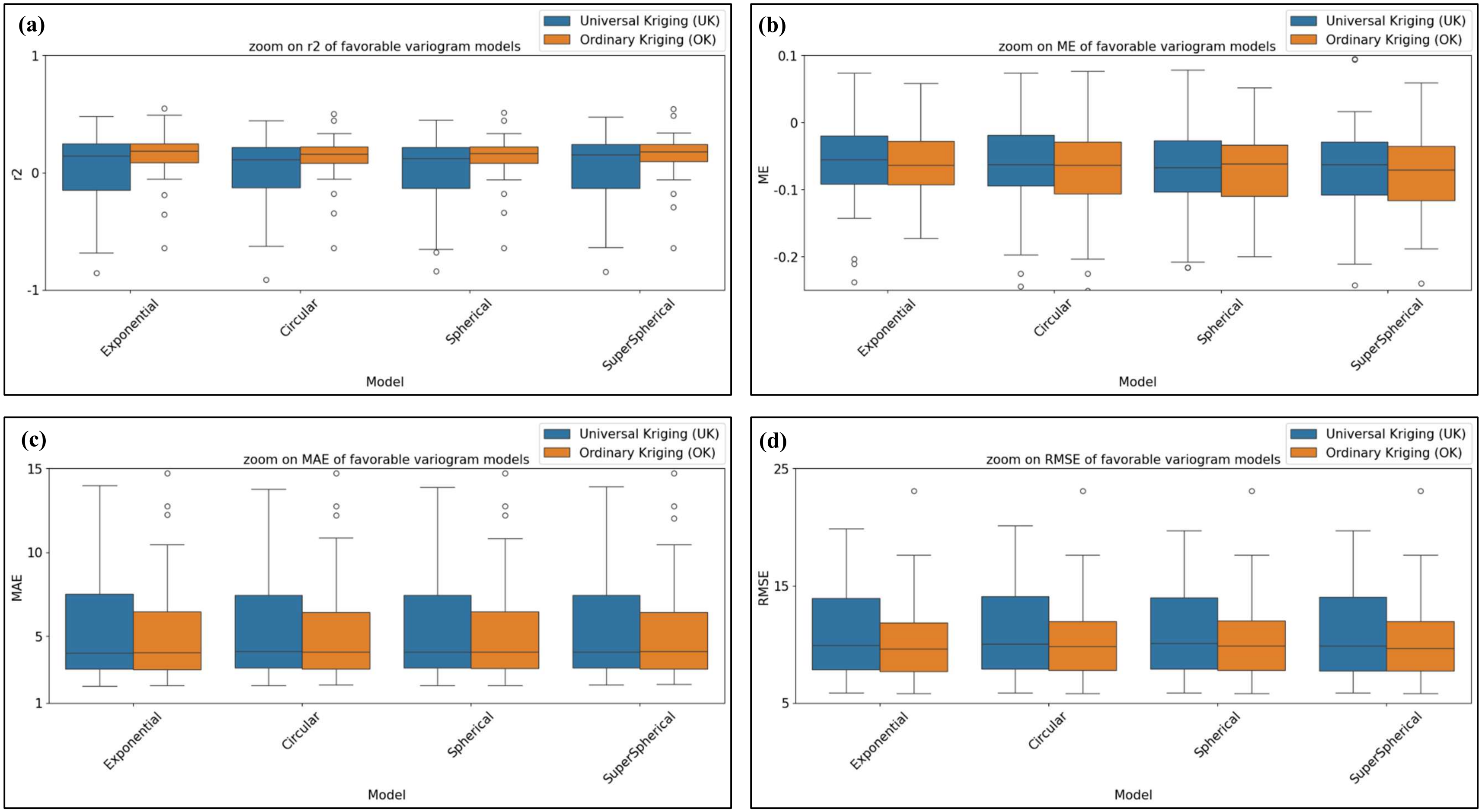

Fig. 9: Zoom on interpolation performances for different methods and favorable variograms, evaluated through LOOCV for the error metrics: (a) r2 (original y-scale -1000 - 1), (b) ME (original y-scale -5 - 5), (c) MAE (original y-scale 1 - 100), (d) RMSE (original y-scale 1 - 70)

Table 3: Interpolated rainfall datasets

Dataset name	Description	DOI
OK Circular	Produced with OK using Circular variogram model	https://zenodo.org/doi/10.5281/zenodo.10997276
OK Spherical	Produced with OK using Spherical variogram model	https://zenodo.org/doi/10.5281/zenodo.11045583
Combined dataset	Produced by completing the missing data from Spherical using Circular model results	https://zenodo.org/doi/10.5281/zenodo.11067784

5.2. Average spatial distribution and variability of rainfall indices

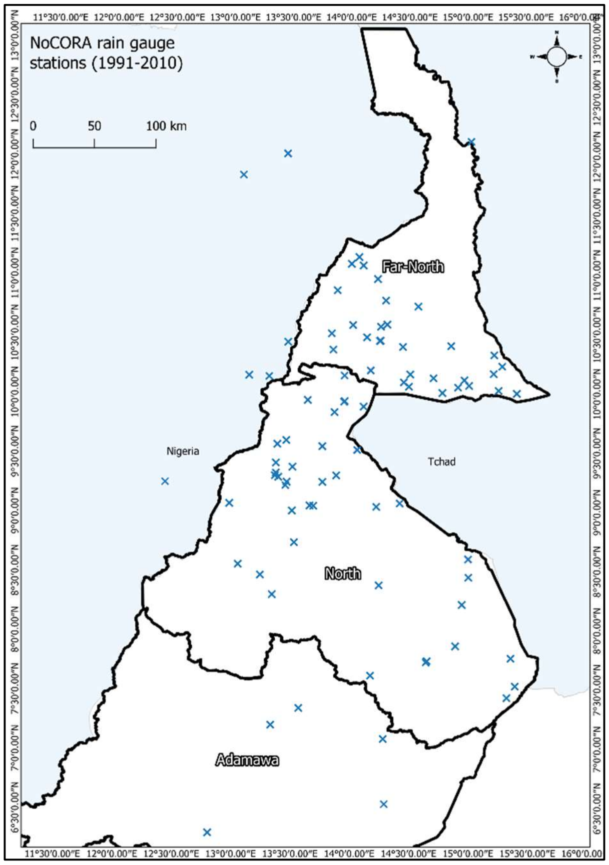

Fig. 10: Localization of rain gauge stations between 1991 and 2010.

We computed the yearly 25 rainfall indices for the period 1991-2010, using the daily interpolated rainfall data created with the stations shown in Fig. 10, to have an outlook of their spatial repartition and sense of variability, we present their average (cf. Fig. 11 and Appendix 2 for supplementary plots) and standard deviation (cf. Fig. 12 and Appendix 3 for supplementary plots) maps.

The maps (Fig. 11 a and b) show that the southern regions experience an early start (before March 31) of the rainy season and a retreat between October 12 and 17. The onset is getting progressively later as you move north, with the beginning of the rainy season being April 20 and after. For the cessation on the other hand, the latest end of the rainy season is observable in the central region, at around October 27, before getting earlier in the Far North (October 7 and before). The season length increases from north to south, extending from 140 to 230 days.

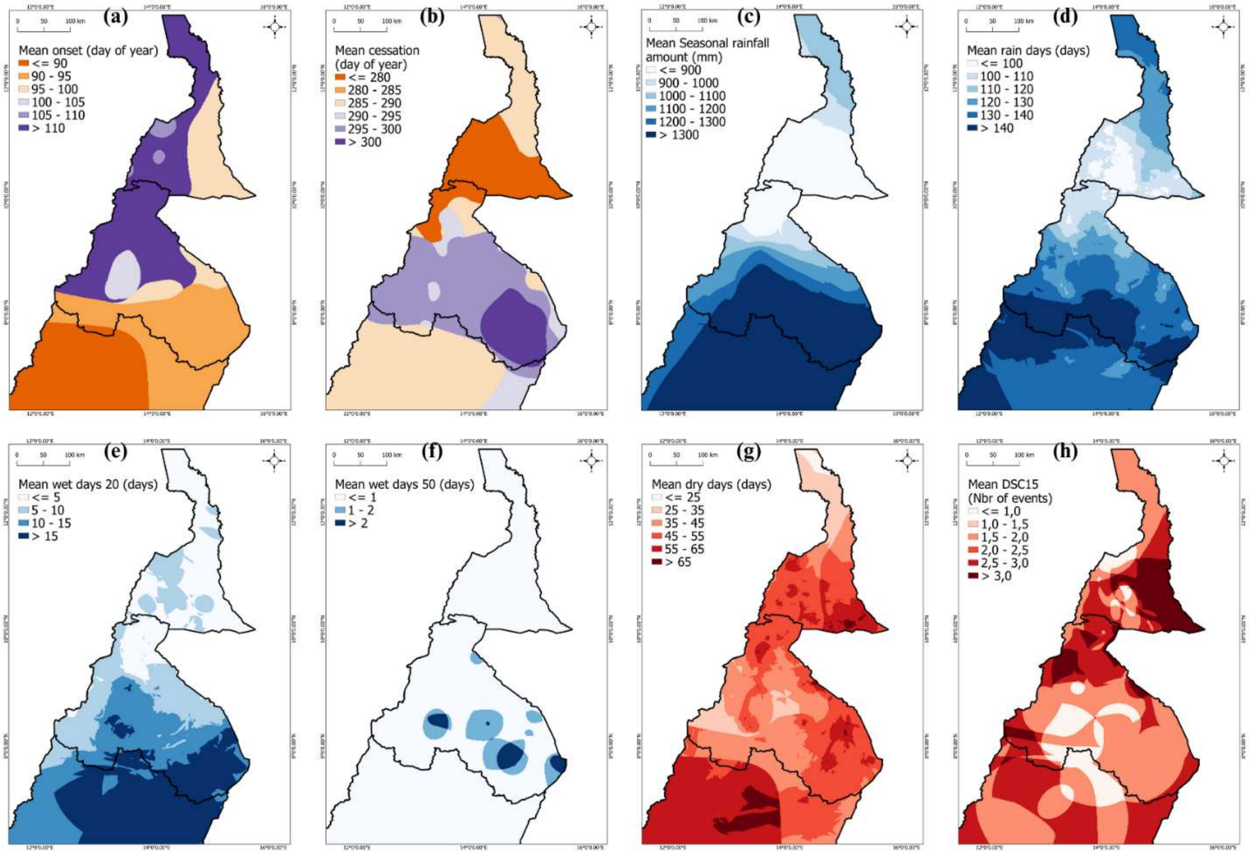

Fig. 11: Spatial distribution of seasonal rainfall indices based on 20-year (1991-2010) averages: (a) Onset, (b) Cessation, (c) Seasonal rainfall amount, (d) Rain days, (e) Wet days 20, (f) Wet days 50, (g) Dry days, (h) DSC15. Supplementary plots for indices not shown here, available in Appendix 2.

Following the north-south gradient, all wet days indices exhibit a clear augmentation of the number of days, or of their relative percentage, towards the south. Even though, the higher the given yield for the wet days indices is, the smaller the area of high values in the south, with wet days 20 reaching up to 20 wet days and wet days 50 only reaching up to 5.

DS events and dry day indices show a complex pattern across the region. Apart from DSC10, the indices show generally low numbers of events (1-3) and dry days (25 to 65) throughout the region with the central areas exhibiting the least. In addition, there are notable pockets with high values exceeding 4 or more events in the north for DSC10 and DSC15, as well as in the south for DSC10, implying more extreme drought conditions in these parts of the study area. Otherwise, DSl events are below 1 in almost all northern Cameroon, except for a very small area in the south, where the number of events goes up to 4.

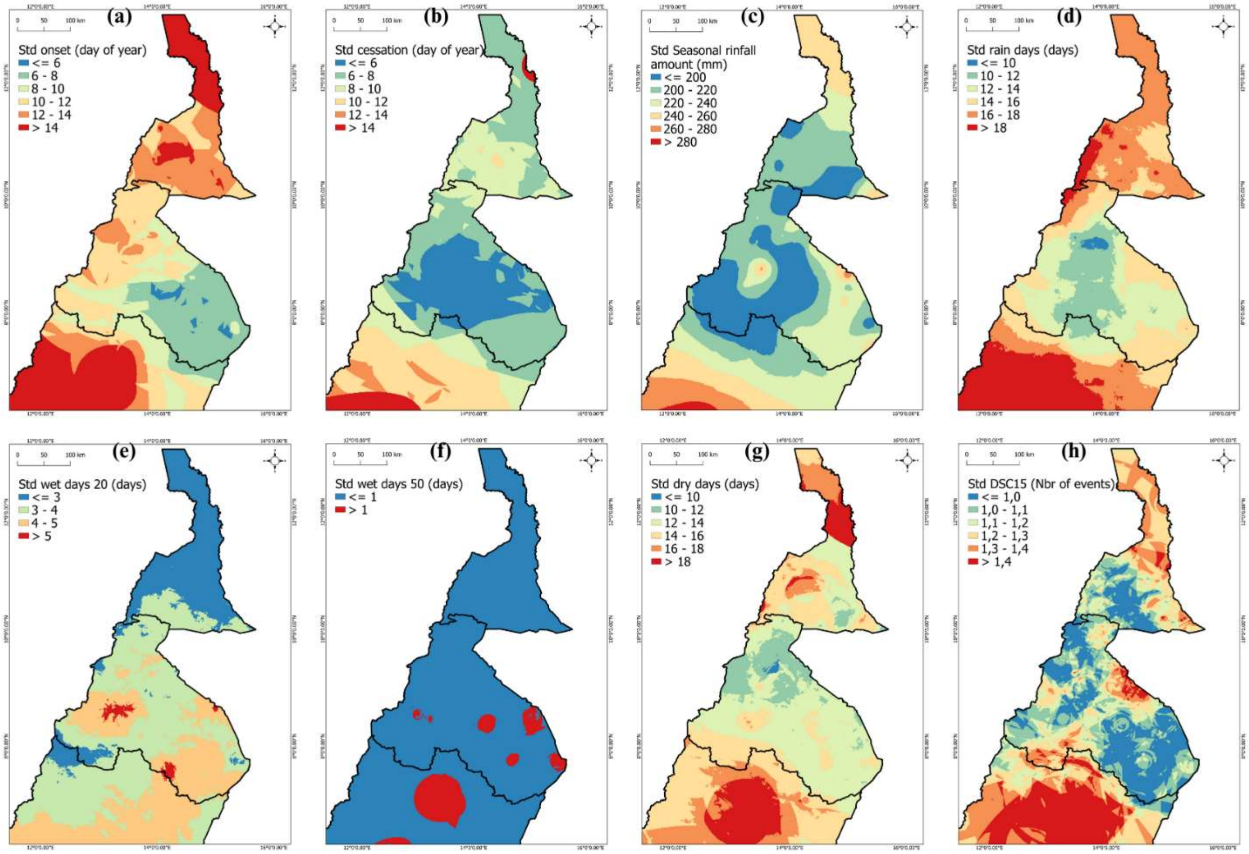

Fig. 12: Spatial distribution of seasonal rainfall indices based on 20-year (1991-2010) standard deviations (Std): (a) Onset, (b) Cessation, (c) Seasonal rainfall amount, (d) Rain days, (e) Wet days 20, (f) Wet days 50, (g) Dry days, (h) DSC15. Supplementary plots for indices not shown here, available in Appendix 3.

For wet day indices, WS1 and relative rain days exhibit a lower variability in comparison with other indices, being inferior to 0.2 for relative indices and inferior to 6 for indices by the number of days. Apart from relative rainy days, the variability follows a clear north-south gradient, with the variability increasing towards the south. For the relative rainy days, as well as most other indices, the variability is lower in the central area of the study zone while the northern and southern parts present a higher variability. A very complex variability with high values can be observed with DSC indices, ranging between 0.8 and 1.8 throughout most of the territory.

5.3. Spatial distribution of average cotton yields

The following maps show the average cotton yields in northern Cameroon for a period of 4 years for the collection point dataset (cf. Fig. 13) and a period of 20 years for the sector datasets (cf. Fig. 14).

Looking at both maps, we can observe that the very northern tip, as well as larger clusters in the south of the study area, either don’t produce any cotton or don’t fall under the supervision of SODECOTON.

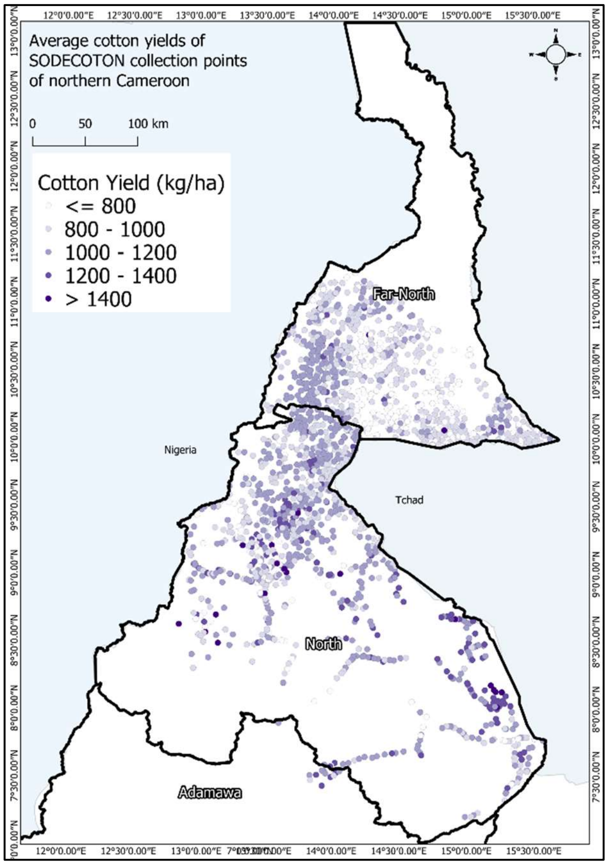

Fig. 13: Average cotton yield of SODECOTON collection points between 2007 and 2010.

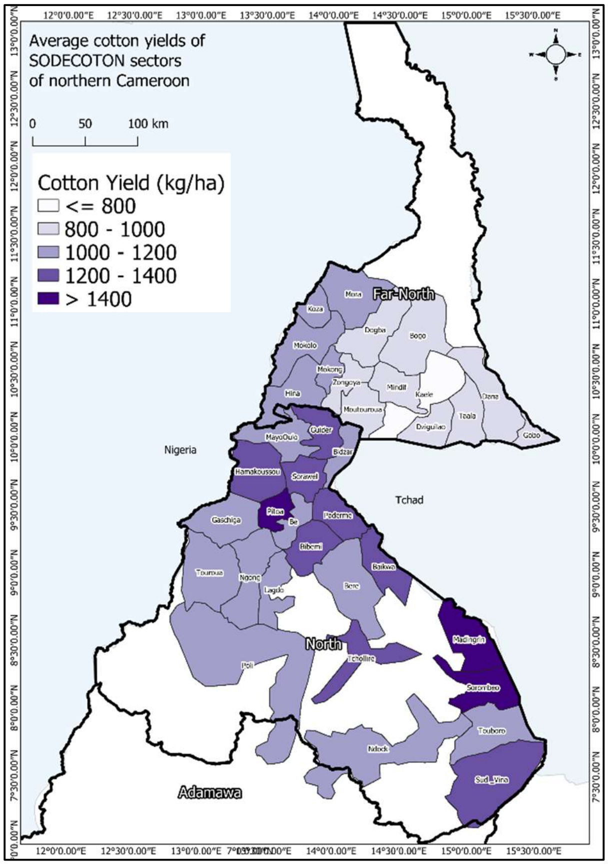

Fig. 14: Average cotton yield of SODECOTON sectors between 1991 and 2010.

The average cotton yield distribution of the collection point dataset reveals a dense concentration of cotton collection points with some exceeding a cotton production of over 1400 kg/ha in the center of northern Cameroon, around the border of the North and Far North, indicating a significant cotton production activity there. The eastern part of the North region as well, exhibits high cotton yield values, while the southern part shows rather average values, mostly round 800 to 1200 kg/ha, with few collection points present.

The sector dataset shows a similar distribution of average cotton yields with high values in the North in areas around Guider, Padame and Pitoa with over 1200 kg/ha, as well as some sectors in the east of the study region, including Madingrin and Sorombeo with over 1400 kg/ha. The Far-

North region generally exhibits lower yields, particularly in the sector of Kaele with average yields below 800 kg/ha.

5.4. Statistical relationships between rainfall indices and cotton yields

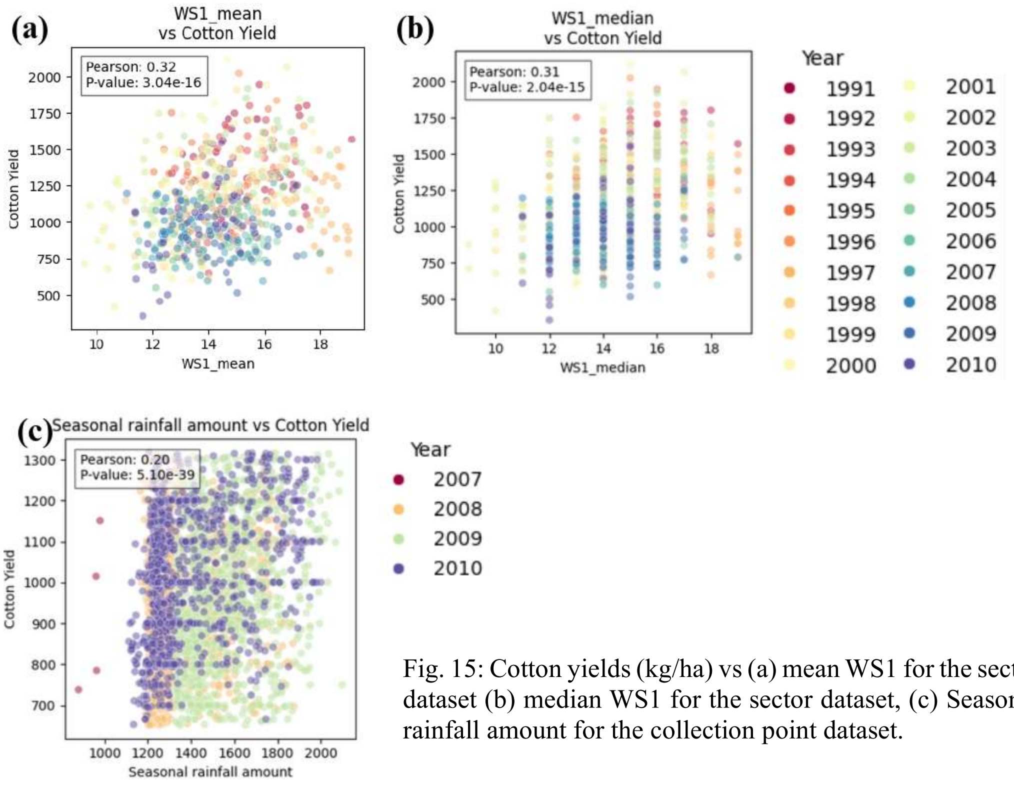

Fig. 15: Cotton yields (kg/ha) vs (a) mean WS1 for the sector dataset (b) median WS1 for the sector dataset, (c) Seasonal rainfall amount for the collection point dataset.

A first visual analysis of the scatterplots displaying the index values vs cotton yields (cf. Appendix 8 to 10), exhibited no clear tendencies in the graphic distribution of the values. Rather, we could observe large clusters of the values with low and high index values equally presenting low and high cotton yield values.

For the collection point dataset, the strongest Pearson coefficient was calculated for the seasonal rainfall amount (0.23, cf. Fig. 15), whereas for the sector dataset WS1 exhibited the highest correlation (0.31-0.32, cf. Fig. 15).

5.4.1. Simple linear relationships

Simple linear regression served to estimate the linear relationship between two variables, here the rainfall indices and the cotton yield. This regression method was applied to both datasets, using spatiotemporal (observed values in time and space, with mean and median aggregated values for the sector dataset index values), spatial (mean and median aggregations of yield and index values of temporal values at a geographic location) and temporal dimensions (mean and median aggregation of yield and index values of all spatial values in a year).

Four additional variations were tested. The first difference method, which takes omitted values into account, was implemented using the temporal dimension for the sector dataset, as well as the spatiotemporal dimension for both datasets. Possible non-linear relationships were addressed by applying a log-linear model for the spatiotemporal dimension of both datasets. The relative values approach, which standardizes the observed values in relation to an average yield and index baseline value, was applied to both datasets for the spatiotemporal dimension. The slope/slope method could only be applied to the sector dataset, due to the short time interval of the collection point dataset, highlighting the relationships between yield and index trends.

Collection point dataset

Looking at the spatiotemporal dimension of the collection point dataset, the strongest linear relationships with Pearson coefficients between 0.2 and 0.23, can be observed for the wet days 20 indices, seasonal rainfall amount and relative rain days, all having a positive impact on cotton yield (cf. Appendix 11). The index most associated with a decline of the cotton yield is the number of dry days, as well as its relative, with Pearson values of -0.18 and -0.19 respectively (cf. Appendix 11). A statistically significant relationship (p-value < 0.05) can be observed for these relationships, as well as all other indices, the exception being Onset and WSC indices (cf. Appendix 5). Nonetheless, the Pearson correlation indicates that the relationship for those indices is quite weak, with several indices exhibiting a correlation coefficient of 0.1 and under (cf. Appendix 4).

When calculating the spatial dimension, correlation coefficients still show weak relationships. Within, DSl, DSxl and DSC15 measure the strongest, with correlation coefficients between -0.22 and -0.27 (cf. Appendix 4). Ranging from 0.22 to 0.28, an increase in cotton yields can be observed with higher seasonal rainfall amount, as well as wet days 20 and 30 indices - apart from wet days 30 indices - when taken the median yield and median index value (cf. Appendix 4).

- **Statistical relationships considering omitted values**

The strongest correlation coefficients displayed for this approach are -0.04 and 0.04 for onset day of the year and dry days indices respectively (cf. Appendix 4). Other than that, only very few statistically significant relationships can be observed for all of the indices (cf. Appendix 5), and they all show a very weak correlation (cf. Appendix 4).

- **Statistical relationships considering non-linear relationships**

Fitting a log function using the OLS method, we can observe weak correlations throughout all indices. The strongest positive influence is displayed with wet days 20 indices, seasonal rainfall amount and relative rain days (0.15 to 0.2, cf. Appendix 5). A negative influence of logarithmic index values on the cotton yield is exhibited by dry days, as well as relative dry days (-0.13 to -0.12, cf. Appendix 5).

- **Statistical relationships using standardized values**

The results exhibit the same Pearson coefficient, slope and p-value as for the simple linear regression, with the strongest positive linear relationships for this method observed between cotton yield with the wet days 20 indices, seasonal rainfall amount and relative rain days respectively (cf. Appendix 4). The strongest negative relationships are with the number of dry days, as well as relative dry days (cf. Appendix 4).

Sector Dataset

For the sector dataset, WS1 indicates the strongest, but overall, still a moderate, relationship for the sector dataset, with a Pearson coefficient of 0.31 or 0.32 for median and mean index respectively (cf. Appendix 12 and 13). In addition, the wet days 30 and 40 indices show some of the higher correlations, between 0.18 and 0.2, indicating rather weak relationships with cotton yields. Otherwise also wet days 50 indices, DSC15 and 20, as well as dry days indices and relative rain days manifest statistically significant, but still quite weak, relationships.

After applying a LULC mask on the sector polygons, the tendencies observed stay the same for all statistically significant relationships (cf. Appendix 4 and 5). For the correlation coefficient of WS1, still displaying the strongest relationship, a minor deviation of the Pearson value (0.28) can be observed (cf. Appendix 4).

Exploring the datasets using the spatial dimension, only minor deviations arise between mean and median analysis. Cessation, season length, seasonal rainfall amount, rain days, heavy rainfall days and wet days 20 indices and wet days 30, manifest strong positive statistically significant relationships with correlation coefficients over 0.5, cessation reached the highest value between 0.62 and 0.67 (cf. Fig. 16). DSC15 and onset on the other hand have a strong negative impact on cotton yields, with the Pearson coefficient ranging from -0.47 to -0.61 and from -0.54 to -0.57 respectively (cf. Fig. 16). DSl and DSxl could not produce linear regression results when calculating median aggregations, due to the low number of occurrences of these dry spell indices, creating medians of zero for all sectors.

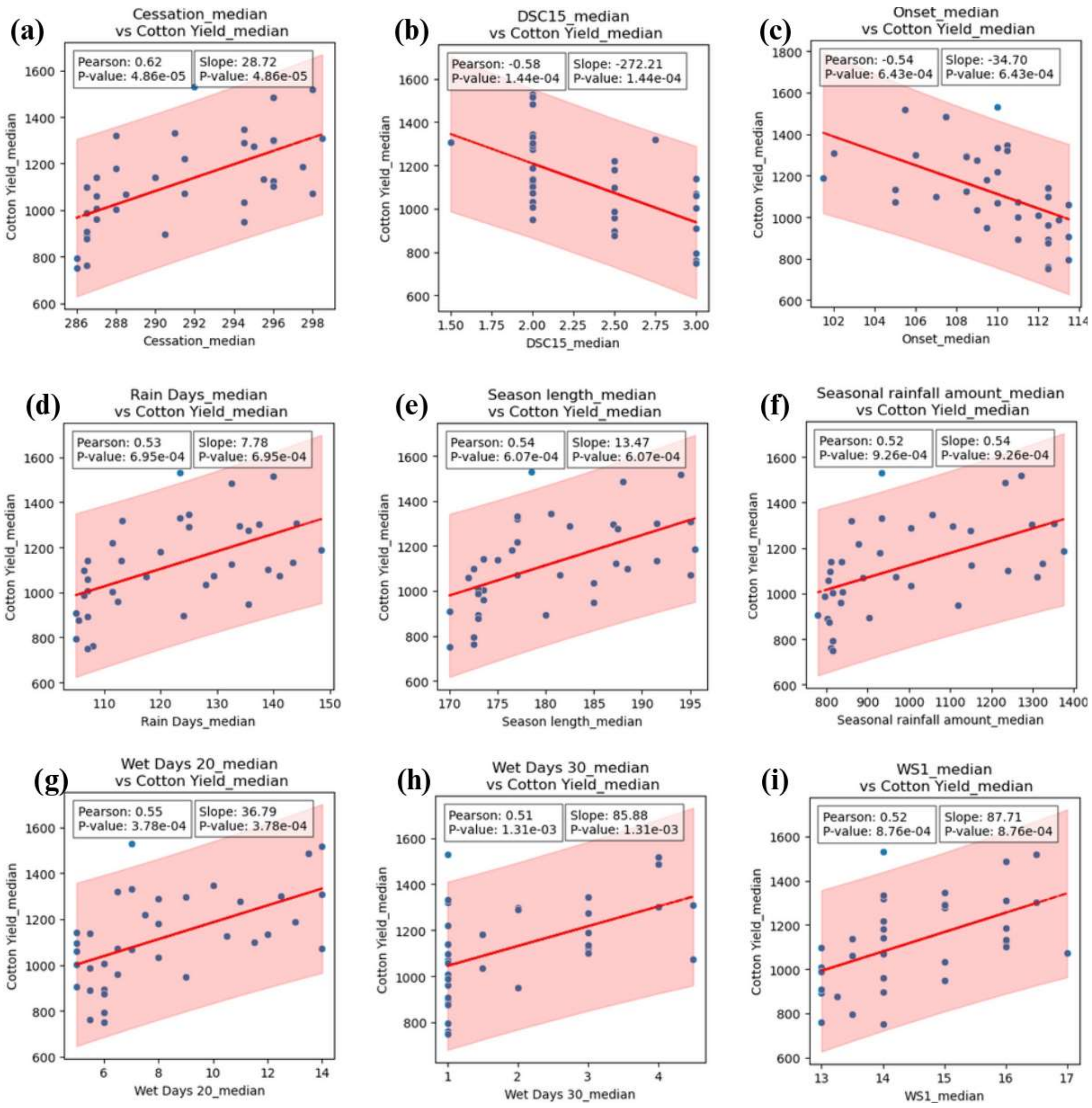

Fig. 16: (a–i) Spatial relationships between cotton yields (kg/ha) and cessation, DSC15, onset, rain days, season length, seasonal rainfall amount, wet days 20, wet days 30, WS1 for the sector dataset based on 20-year median values, using OLS

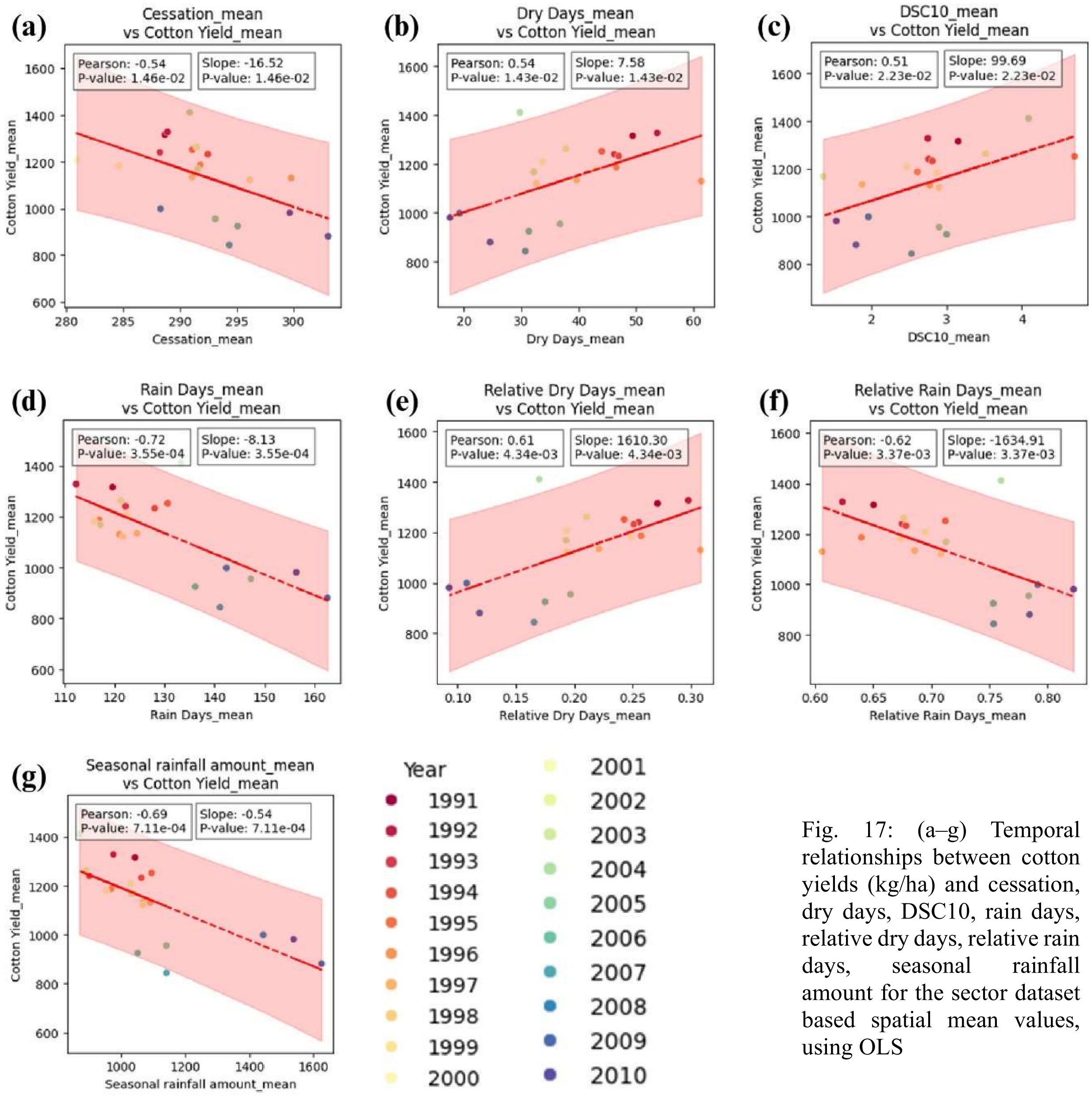

Fig. 17: (a–g) Temporal relationships between cotton yields (kg/ha) and cessation, dry days, DSC10, rain days, relative dry days, relative rain days, seasonal rainfall amount for the sector dataset based spatial mean values, using OLS

Due to the short time span of the collection point dataset, the temporal dimension could only be calculated for the sector dataset. There is a noticeable difference to the spatial dimension, with many linear relationships that showed a positive influence, showing a negative influence, as well as the contrary. Thus, dry days indices display a strong positive correlation coefficient between 0.52 and 0.62 (cf. Fig. 17), while the correlation of the seasonal rainfall amount lies between -0.63 and -0.69 (cf. Fig. 17). Similar values can be observed for DSC10 (0.51 to 0.53, cf. Fig. 17) and rain day indices (-0.58 to -0.72, cf. Fig. 17). Furthermore, the cessation day of the year exhibited a strong Pearson coefficient between -0.47 and -0.54 (cf. Fig. 17).

- **Statistical relationships considering omitted values**

In comparison with the simple linear regression, still all indices exhibit a low correlation degree. Although, for the spatiotemporal dimension of the sector dataset applying first difference, particularly onset (-0.18, cf. Appendix 4), cessation (0.22 to 0.24, Appendix 4) and season length (0.25, Appendix 4) exhibit higher Pearson values, as well as seasonal rainfall amount (0.22 to 0.23, Appendix 4) and rain days (0.23 to 0.24, Appendix 4). Wet days 30 indices reveal a correlation coefficient of 0.21 all the same (cf. Appendix 4), in addition to wet days 20 indices which present a higher Pearson value now as well (cf. Appendix 4), whilst it has diminished for wet days 40 indices and WS1 (cf. Appendix 4).

Furthermore, regarding the temporal dimension, at a correlation between 0.51 and 0.63, wet days 20 indices have a strong, more influential relationship (cf. Appendix 14 and 15), as well as cessation (-0.53 to -0.55, cf. Appendix 14 and 15) if taking the temporal index median. In comparison to the simple linear regression, only DSC10 still indicates a strong significant relationship, with slightly lower Pearson values between 0.49 and 0.52 (cf. Appendix 4 and 5).

- **Statistical relationships considering non-linear relationships**

When adding log fitting to the linear regression equation, WS1 exhibits the strongest Pearson coefficient with a moderate correlation degree (0.31 to 0.32, cf. Appendix 4). Wet days 30 indices show a rather weak Pearson coefficient of around 0.25 to 0.26, though only when taken the median index values (cf. Appendix 4). For DSl as well, a difference can be observed between mean and median index values. The median values indicate a negative impact on the cotton yield with a moderate correlation coefficient of -0.34 (cf. Appendix 4), while the mean values for the same index show no statistically significant relationship.

- **Statistical relationships using standardized values**

Anew, the results correspond to those of the simple linear regression, with WS1 indicating the strongest relationship for the sector dataset for this method (cf. Appendix 4) and the wet days 30 and 40 indices show some of the higher correlations observed (cf. Appendix 4).

- **Statistical relationships between yield and index trends**

Looking at the cessation index, the slope/slope method shows a strong positive correlation between 0.57 and 0.59 (cf. Fig. 18). The WS1 index displays the strongest negative correlation (-0.46 to -0.47, cf. Appendix 4), conversely to the findings of the spatial dimension of the sector dataset (cf. Fig. 16).

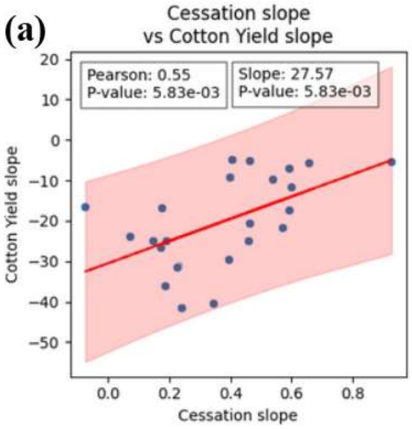

Fig. 18: (a) Trend relationships between cotton yields (kg/ha) and Cessation

5.4.2. *Multiple linear relationships*

The multiple linear regression method is used to assess the relationship between two rainfall indices and the cotton yield. It was applied to both datasets, using the spatiotemporal dimension.

Collection point dataset

The correlation analysis between index pairs for the collection point dataset revealed only one pair without statistically significant relationship: wet days 50 - rain days (cf. Appendix 16). The relationship between those indices and the cotton yield is also nonsignificant, with a weak Pearson coefficient at -0.01 (cf. Appendix 18).

Sector dataset

For the sector dataset several index combinations displayed nonsignificant relationships with no correlation, including cessation – DSC15, DSC15 – relative rain days, DSC20 – onset, DSC20 – WS1, DSl – WS1, DSxl – rain days and onset – relative dry days (cf. Appendix 17).

Still, none of the explored combinations implemented in multiple linear regression with the cotton yield show any statistically significant relationships, and all show very weak correlations (cf. Appendix 19). The strongest correlations can be observed with the cessation and DSC15 combination, with a correlation coefficient of 0.04 (cf. Appendix 18).

5.4.3. *Evaluation of metrics*

Examining the r^2 values for the different methods and indices (cf. Appendix 7), using the spatiotemporal dimension for both datasets, as well as the spatial dimension for the collection point dataset, indices explain only a negligeable part of the change in cotton yield. The only exception is the slope/slope approach, where the results show that the retreat of the rain season explains around 27 to 30% in the change of cotton yield. For the temporal and spatial dimensions of the sector dataset on the other hand, there is a much stronger influence to be observed. When applying simple linear regression in the spatial dimension, seasonal length and cessation account for up to 40 and 45% of the variation in cotton yield respectively. In the temporal dimension, it is wet days 20 and 30 indices that generate the highest variations, between 34 and 38%. Adjoining the first difference method to the linear regression of the sector dataset, the wet days 20 index still accounts for the highest change in cotton yield with 36 to 40% for the temporal dimension.

The slope values (cf. Appendix 6) highlight these influences, as well as the positive and negative impacts displayed by the Pearson correlation coefficient. For the slope/slope method, a delay by one day of the retreat date accounts for 27 to 30 kg/ha more of cotton. For the spatial dimension, one day of delay decreases the cotton yield by around 26 to 35 kg/ha of cotton and one a prolonged rainy season by one day adds around 12 to 15 kg/ha. On the other side, every added dry day decreases the cotton yield by around 10 kg/ha. For the temporal dimension, the cotton yield decreases by 22 to 43 kg/ha for every additional wet day over 20mm or 30 mm of rainfall.

6. DISCUSSION

When looking at the relationships found between rainfall indices and cotton yields, we can observe two larger groups that stand out. On one hand, we have dry days indices and DSC indices. On the other hand, we have seasonal rainfall amount, rain days indices, wet days 20 indices and wet days 30, as well as WS1. This applies to both datasets, although the correlations stand out much stronger for the sector dataset. This is likely because the aggregation of rainfall indices for each sector which creates a scaling effect, smoothing out local variabilities and capturing more consistent trends, whereas the finer-scale collection point dataset might show more variability and noise due to its higher spatial resolution and diverse conditions including non-climatic factors such as human management, biotic stresses, pests, etc.

The two groups of indices, differently impact cotton yields. In the spatial dimension (meaning independently of time, which is aggregated) for the sector dataset, cotton yields are highly positively affected by the seasonal rainfall amount, rain days indices, wet days 20 indices and wet days 30, as well as WS1. This coincides with findings of Njouenwet et al. (2021), for cotton in northern Cameroon and W. Zhang et al. (2018), for all vegetation productivity in the Sahel, who state that in space the number of heavy rainfall events and seasonal rainfall amount positively correlate with yield. On the other side, we found that DSC15 has a highly negative effect.

When analyzing the relationships from a temporal dimension (meaning independently of space, all northern Cameroon sector values are aggregated) for the sector dataset, the effects are opposite to those in the spatial dimension. DSC10 and dry days indices show a strong positive effect with yield, while Njouenwet et al. (2021) discovered a negative influence of consecutive dry days trends on the cotton yield in northern Cameroon. However, they only analyzed data from 16 stations in northern Cameroon, focusing on trends not only from a temporal point of view but also spatially, possibly leading to the differences in results. The seasonal rainfall amount, rain days indices, and wet days 20 indices exhibit a strong negative relationship in the temporal dimension. This coincides with findings of Gérardeaux E. et al. (2018), who explained that certain climate models predict lower cotton yields with rising rainfalls, although it is important to note that he is prognosing this impact based on climate prediction data and not historical data.

Given these two opposing results, it appears that the spatiotemporal dimension (which considers the observed and interpolated values in time and space), exhibits weak correlations possibly because the spatial and temporal components tend to counterbalance each other.

The correlations found for spatial dimension might be because of local characteristics, with certain sectors profiting from higher rainfalls for optimal cotton growth. Geographically, we observed that seasonal rainfall amount, rain days, wet days 20 indices and wet days 30, as well as WS1 follow a north-south gradient (cf. 5.2. Average spatial distribution and variability of rainfall indices), with higher index values in the south of the study area. It is in this region that farmers use primarily long cycle cotton varieties (Dessauw et al., 2010), taking full advantage of increased water availability. For sectors in the Far North, who receive much less rainfall and plant short cycle cotton, yields are lower due to quick maturation, not profiting as much from elevated rainfalls during the rainy season. In addition, in the Far North of Cameroon, an area used for cotton farming since long time, soils are heavily degraded (Tsozué et al., 2014), which could lead to lower cotton yields. Lastly, prolonged dry spells (such as DSC15) for short cotton cycles, especially during critical growing stages, could lead to detrimental losses as cotton crops require a certain amount of soil moisture for optimal growth (Datta et al., 2019).

However, when spatially aggregating yield and index value for the whole study area, seasonal rainfall amount, wet days 20 indices and rainy days indices could display a negative effect on cotton yields, due to floods. Gérardeaux E. et al. (2018) suggests that this negative influence stems from nitrogen leaching, presenting a bigger constraint to cotton yields than droughts. Additionally, northern Cameroon is known to be prawn to floods, in years of excessive rainfall, they can cover up entire parcels (Bouba L. et al., 2017; Tchotsoua, 2007). A consequence of these floods can be waterlogging towards which cotton exhibits poor tolerance, influencing its growth and development, as well as nutrient intake (Hocking et al., 1987; Hodgson, 1982). Moreover, it is important to recall the definitions of the chosen dry days and DSC indices, since the dry days with less than 1mm of rainfall per day are not analyzed in a consecutive count of days and consecutive dry spells have a higher rainfall threshold. W. Zhang et al. (2018) showed that vegetation productivity in the area only starts to be negatively impacted after more than 14 consecutive dry days. In the temporal dimension, the absence of a negative effect of DSC10 (exactly 10 days with each day <10mm of rainfall) on yields, could be attributed to the short duration of the DSC10 index. We can hypothesize that the relatively high rainfall threshold for dry spells indices of 10 mm for the DSC10 index could depict the positive influence of lighter rainfalls, as it is close to the optimal rainfall intensity for cotton productivity, which is supposed to be around 12.5 to 13 mm (Njouenwet et al., 2021; W. Zhang et al., 2018). Furthermore, the daily evapotranspiration demand of cotton is only between 1 to 10 mm depending on the growth stage (Ezan et al., 1998).

In addition to the previously mentioned indices, one further group stands out: the onset day of the year, the cessation day of the year and the season length. In the temporal dimension for the sector dataset, the later cessation date is associated with lower cotton yields. We can hypothesize that this could be due to an extended exposure of cotton crops to risks such as droughts and floods. In contrast, when analyzing trends using the slope/slope method, which helps disambiguate the effects of interannual variability from the effect of trends of both predictor and predicted variables, a later cessation date correlates with improved cotton yields, a correlation underlined by the findings of Gérardeaux E. et al. (2018). The difference in results could be explained by the fact that the analysis in the temporal dimension may vary widely due to short-term factors, while the slop/slope method shows broader patterns and more reliable results. This suggests that over time a later retreat date can be beneficial to cotton, maybe indicating the successful implication of new cotton varieties. The results of the spatial dimension support this theory, showing that cotton yields can profit from an earlier onset, later cessation and therefore longer season length, allowing a prolonged growing cycle for long cotton cycle varieties (Dessauw et al., 2010; Gérardeaux E. et al., 2018). The importance of these indices was also underlined by Sultan et al. (2009), who found a close spatial relationship between the three indices and cotton yields, with an early onset, late cessation and longer season length giving the highest yields.

Even though in this study only rainfall was considered, it is important to discuss other factors that influence the cotton yield. As shown in the state of the art (cf. 2. State of the Art), several studies found that the temperature plays a role in the productivity of cotton yields. It can reduce the duration of the cotton growth period, accelerate the phenological cycle, and increase the evapotranspiration demand because of higher water stress (Gérardeaux et al., 2013; Roudier et al., 2011). Another influence stems from CO2 acting as a fertilizer for cotton, which could benefit cotton yields within certain limits (Gérardeaux et al., 2013).

The results of the first difference method consider omitted values, revealing the impact of other independent variables that are otherwise attributed to rainfall. Although, we cannot determine which independent variable influences the yield to which degree. Two indices stand out particularly when looking at the temporal dimension for the sector dataset. The number of dry days, which had a strong positive effect, shows a weak negative correlation with cotton yields when considering omitted values. The strong correlation with the relative percentage of rainy days, which was negative when applying simple linear regression, now exhibits a weak positive relationship. This undermines the theory that other factors have an influence on cotton yields. The initial positive effect of dry days may have been due to effective irrigation practices mitigating the lack of rainfall. For another example, crop yields could also increase under tillage and conservation agriculture systems, helping to retain water (Gérardeaux et al., 2013). Conversely, the positive relationship with the relative percentage of rain days undermines that, in the presence of other favorable conditions, moderate rainfall can indeed benefit cotton yields.

The findings of our study clearly show that changes in rainfall provoke changes in cotton yields. The identified relationships between rainfall indices and cotton yields suggest an urgent need for the development and implementation of adaptive strategies to local climatic conditions. For instance, policies promoting adapted cotton varieties, and the adjustment of planting schedules based on more accurate seasonal forecasts could greatly enhance the resilience of cotton farmers. Furthermore, programs to improve farmers' knowledge on the climate change implications and adaptation strategies could empower them to better manage risks associated with unpredictable rainfall patterns. By integrating these points, northern Cameroon could improve the livelihoods of farmers and strengthen their resilience to climate change.

7. LIMITATIONS AND PERSPECTIVES

Several limits were encountered during the study, related to data and methodology.

Even so the data provided by the NoCORA dataset presented a very complete collection of rainfall measurements in time in space, many dates presented very few data entries with around less than 5 stations. This leads to the predicted rainfall values of these dates being far from the observed ones due to poor interpolation capacity. In addition, while the chosen kriging interpolation method was concluded to produce the best results, the interpolation method and variogram models couldn't interpolate the rainfall data of certain dates. Since these missing dates appeared during the same month at the same frequency for every year, the decision was made not to apply gap-filling, therefore our results have to be considered relative and not absolute.

Furthermore, we have to consider that our rainfall indices were calculated for the whole rainy season, which does not necessarily correspond to the planting and harvesting of cotton in the different regions. Therefore, heavy rainfalls, an elevated number of dry days or other index events that occur during the rainy season, but not necessarily during the cotton cycle, could influence our results. Moreover, it would have been interesting to analyze the relationships between rainfall indices and cotton yields during specific stages of the cotton cycle, to determine how the impact can differ depending on the level of cotton growth.

The cotton data must be handled with care as well. Since our dataset consists of yield values reported at the collection site and not measured at the field, the cotton yields registered can differ from the actual yield.

Lastly, if adapted data for other climatic and meteorological variables, such as the temperature and evapotranspirational needs, would be available, it would be an important measure to include those variables in the study. These variables affect the cotton yield as well and, in this study, could be attributed to the impacts of rainfall instead.

Nonetheless, the results of this study are intended to be a cornerstone for agroclimatic forecasting for cotton production, facilitating the evaluation of yield impacts over temporal and spatial scales using rainfall projections from sources like CORDEX (Coordinated Regional Climate Downscaling Experiment).

The continuance of our study involves downscaling high-resolution climate simulations (MCRs) to a 0.01° × 0.01° grid and correcting biases in simulated rainfall using the historical data fitted in this study. Four bias correction methods - Power Transformation (PT), Scaling (SCL), Generalized Quantile Mapping (GEQM), and Gamma Quantile Mapping (GAQM) - can then be applied to address systematic biases. In this context, as described by Sultan et al. (2009), the uncertainties in climate change projections must be reduced as well to produce reliable future scenarios of agricultural productivity. Therefore, the most suitable bias correction method would be selected based on a set of statistical indices, and the resulting correction coefficients could be used to adjust future climate projections for the RCP 2.6, 4.5, and 8.5 scenarios for the period 2010-2100. Finally, a multi-model ensemble would be generated for downscaled MCR projections.

Gérardeaux E. et al. (2018) put in place a similar project to determine the cotton ideotype for Cameroon, applying the CROPGRO-cotton model under the RCP 8.5 scenario. He noted that climate models can produce inconsistent results for rainfall distributions and amounts, highlighting the need for the continuance of this study.

By exploring future cotton productivity under diverse climatic conditions, this study is a pillar for providing critical foresight into the potential scenarios of cotton yield variability, ultimately supporting strategic planning and resilience-building for famers in the cotton sector of northern Cameroon.

8. CONCLUSION

The focus of this study was to analyze the statistical relationships between rainfall indices and cotton yields in northern Cameroon, to strengthen the resilience of farmers to climate change.

To calculate our rainfall indices, we found that Ordinary Kriging (OK) using Circular and Spherical variogram models were the best method to interpolate the rainfall data and produce annual rainfall indices maps for 25 indices in total.

During the study, after implementing different linear regression methods, we discovered statistically significant relationships between cotton yields and several rainfall indices. The following showed a strong degree of correlation for at least one linear regression approach tested: onset and cessation date as well as season length, dry days and relative dry days, DSC10 and 15, seasonal rainfall amount, rain days and relative rain days, wet days 20 and 30, as well as relative wet days 20, and WS1.

The correlations calculated were much higher for the cotton yields provided at the sector level rather than at the collection point. In addition, our results demonstrated that depending on the dimension (spatial, temporal or spatiotemporal), the influence of these indices may vary, not only in their degree of impact, but also in terms of positive or negative relationships. These contrasted impacts may be partly explained by cultivation improvements/degradation in time, counterbalanced by changes in rainfall patterns and indices.

The spatial and temporal dimension for the sector dataset exhibited the strongest metrics. We revealed that moderate dry periods, such as DSC10, and dry days are beneficial for cotton yields when considering the mean and median yield for the entire Sudano-Sahelian zone of Cameroon. Consistent stronger dry spells, such as DSC15, remain a limiting factor for cotton productivity for an extent of sectors. Conclusively, when aggregating all yearly values for each sector, seasonal rainfall amount, rain days, wet days 20 indices and wet days 30, as well as heavy rainfall days are contributing factors to cotton productivity. In addition, cotton yields profit from a longer season length, with earlier onset and later cessation. When considering the mean and median yield for the whole area, rising rainfall in addition to more rainy days and a later cessation date can be detrimental to cotton yields.

This study highlights the need for farmers and policymakers to consider changes in rainfall patterns when adjusting agricultural methods and shaping water management policies. The produced maps and historical data fitted will help with this, serving for the bias-correction and adjustment of future climate projections, used to model the impacts on cotton yields.

REFERENCES

Beauvilain A. (1996). La pluviométrie dans le Bassin du Lac Tchad. In *Atlas d'élevage du bassin du Lac Tchad* (pp. 11–21). CIRAD-EMVT-Service Infographie-Cartographie (FRA). Wageningen : CTA,.

Benoit, K. (2011, March 17). Linear Regression Models with Logarithmic Transformations. Methodology Institute London School of Economics.

Bouba L., Sauvagnargues S., Gonne B., Ayral P.A, & Ombolo A. (2017). Tendances pluviométriques et aléa inondation à l'Extrême-Nord Cameroun. *Geo-Eco-Trop.*, *3*(41), 339–358.

Brinkhoff, T. (2020, March 2). Cameroon. Retrieved from https://www.citypopulation.de/en/cameroon/cities/

Burton, A. L. (2021). OLS (Linear) Regression. In J. C. Barnes & D. R. Forde (Eds.), *The Encyclopedia of Research Methods in Criminology and Criminal Justice* (1st ed., pp. 509–514). Wiley. https://doi.org/10.1002/9781119111931.ch104

Dassou, E. F., Ombolo, A., Chouto, S., Mboudou, G. E., Essi, J. M. A., & Bineli, E. (2016). Trends and Geostatistical Interpolation of Spatio-Temporal Variability of Precipitation in Northern Cameroon. *American Journal of Climate Change*, *05*(02), 229–244. https://doi.org/10.4236/ajcc.2016.52020

Datta, A., Ullah, H., Ferdous, Z., Santiago-Arenas, R., & Attia, A. (2019). Water Management in Cotton. In K. Jabran & B. S. Chauhan (Eds.), *Cotton Production* (1st ed., pp. 47–59). Wiley. https://doi.org/10.1002/9781119385523.ch3

Dessauw, D., Oumarou, P., & Latrille-Debat, S. (2010). *Sélection cottonière rapport annuel campagne 2009/10* (Sélection du cotonnier) (p. 71). Cameroun: IRAD-CIRAD.

Dismuke, C. E., & Lindrooth, R. (2005). Ordinary Least Squares. In *Methods and Designs for Outcomes Research*. American Society of Health-System Pharmacists.

Ezan, M., Hala, N., Kesse, F., Koto, E., Kouaido, N., Kouassi, A., et al. (1998). *Culture du Coton: Manuel Technique*. Bouaké: CIRAD, IDESSA.

Fall, C. M. N., Lavaysse, C., Drame, M. S., Panthou, G., & Gaye, A. T. (2019, July 22). Wet and dry spells in Senegal: Evaluation of satellite-based and model re-analysis rainfall estimates. https://doi.org/10.5194/nhess-2019-185

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315.

Field, C. B., Barros, V., Stocker, T. F., & Dahe, Q. (Eds.). (2012). *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139177245

Folefack, D. P., Kaminski, J., & Enam, J. (2011). Note sur le contexte historique et gestion de la filière cotonnière au Cameroun. *Afrique pouvoir et politique*, *11*.

Gérardeaux, E., Sultan, B., Palaï, O., Guiziou, C., Oettli, P., & Naudin, K. (2013). Positive effect of climate change on cotton in 2050 by CO2 enrichment and conservation agriculture in Cameroon. *Agronomy for Sustainable Development*, *33*(3), 485–495. https://doi.org/10.1007/s13593-012-0119-4

Gérardeaux E., Loison R., Palaï O., & Sultan B. (2018). Adaptation strategies to climate change using cotton (Gossypium hirsutum L.) ideotypes in rainfed tropical cropping systems in Sub-Saharan Africa. A modeling approach. *Field Crops Research*, *226*, 38–47. https://doi.org/doi.org/10.1016/j.fcr.2018.07.007

Hocking, P. J., Reicosky, D. C., & Meyer, W. S. (1987). Effects of intermittent waterlogging on the mineral nutrition of cotton. *Plant and Soil*, *101*(2), 211–221. https://doi.org/10.1007/BF02370647

Hodgson, A. (1982). The effects of duration, timing and chemical amelioration of short-term waterlogging during furrow irrigation of cotton in a cracking grey clay. *Australian Journal of Agricultural Research*, *33*(6), 1019–1028.

Isaaks, E. H., & Srivastava, R. M. (1989). *Applied Geostatistics*. New York: Oxford University Press.

Joël, F. N. E., Ludovic, T., & Anselme, W. (2015). From time uncertainties to climate-smart agriculture in the Sudano-Sahelian zone of Cameroon. In *Building tomorrow's research agenda and bridging the science-policy gap*. Montpellier, France.

Karra, K., & et al. (2021). Global land use/land cover with Sentinel-2 and deep learning. [Data set]. IGARSS 2021-2021 IEEE International Geoscience and Remote Sensing Symposium. IEEE.

Lavarenne, J., Nenwala, V. H., & Foulna Tcheobe, C. (2023). NoCORA - Northern Cameroon Observed Rainfall Archive [Data set]. Zenodo. https://doi.org/10.5281/ZENODO.10204362

Leblois, A., Quirion, P., & Sultan, B. (2014). Price vs. weather shock hedging for cash crops: Ex ante evaluation for cotton producers in Cameroon. *Ecological Economics*, *101*, 67–80. https://doi.org/10.1016/j.ecolecon.2014.02.021

Liebmann, B., Bladé, I., Kiladis, G. N., Carvalho, L. M. V., B. Senay, G., Allured, D., et al. (2012). Seasonality of African Precipitation from 1996 to 2009. *Journal of Climate*, *25*(12), 4304–4322. https://doi.org/10.1175/JCLI-D-11-00157.1

Longman, R. J., Frazier, A. G., Newman, A. J., Giambelluca, T. W., Schanzenbach, D., Kagawa-Viviani, A., et al. (2019). High-Resolution Gridded Daily Rainfall and Temperature for the Hawaiian Islands (1990–2014). *Journal of Hydrometeorology*, *20*(3), 489–508. https://doi.org/10.1175/JHM-D-18-0112.1

Maidment, R. I., Grimes, D., Black, E., Tarnavsky, E., Young, M., Greatrex, H., et al. (2017). A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa. *Scientific Data*, *4*(1), 170063. https://doi.org/10.1038/sdata.2017.63

Mbétid-Bessane, E., Havard, M., & Djondang, K. (2006). Évolution des pratiques de gestion dans les exploitations agricoles familiales des savanes cotonnières d’Afrique centrale. *Cahiers Agricultures*, *15*(6), 555–561. https://doi.org/10.1684/agr.2006.0038

M’Biandoun, M., & Olina, J. (2009). Pluviosité en région soudano-sahélienne au Nord du Cameroun : conséquences sur l\’agriculture. *Agronomie Africaine*, *18*(2), 95–104. https://doi.org/10.4314/aga.v18i2.1683

Molua, E., & Lami, C. (2009). The economic impact of climate change on agriculture in Cameroon. *IOP Conference Series: Earth and Environmental Science*, *6*(9), 092017. https://doi.org/10.1088/1755-1307/6/9/092017

Molua, E. L. (2006). Climatic trends in Cameroon: implications for agricultural management. *Climate Research, 30*, 255–262.

Moral, F. J. (2010). Comparison of different geostatistical approaches to map climate variables: application to precipitation. *International Journal of Climatology*, *30*(4), 620–631. https://doi.org/10.1002/joc.1913

Müller, S., Schüler, L., Zech, A., & Heße, F. (2022). GSTools v1.3: a toolbox for geostatistical modelling in Python. *Geoscientific Model Development*, *15*(7), 3161–3182. https://doi.org/10.5194/gmd-15-3161-2022

National Geographic Society. (2024, July 19). Köppen Climate Classification System. In *National Geographic*. Retrieved from https://education.nationalgeographic.org/resource/koppen-climate-classification-system/

Nettleton, D. (2014). Chapter 6 - Selection of Variables and Factor Derivation. In D. Nettleton (Ed.), *Commercial Data Mining* (pp. 79–104). Boston: Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-416602-8.00006-6

Nicholson, S. (2000). The nature of rainfall variability over Africa on time scales of decades to millenia. *Global and Planetary Change*, *26*(1–3), 137–158. https://doi.org/10.1016/S0921-8181(00)00040-0

Njouenwet, I., Vondou, D. A., Ashu, S. V. N., & Nouayou, R. (2021). Contributions of Seasonal Rainfall to Recent Trends in Cameroon's Cotton Yields. *Sustainability*, *13*(21), 12086. https://doi.org/10.3390/su132112086

Njouenwet, I., Tchotchou, L. A. D., Ayugi, B. O., Guenang, G. M., Vondou, D. A., & Nouayou, R. (2022). Spatiotemporal Variability, Trends, and Potential Impacts of Extreme Rainfall Events in the Sudano-Sahelian Region of Cameroon. *Atmosphere*, *13*(10), 1599. https://doi.org/10.3390/atmos13101599

Penlap, E., Matulla, C., von Storch, H., & Kamga, F. (2004). Downscaling of GCM scenarios to assess precipitation changes in the little rainy season (March-June) in Cameroon. *Climate Research*, *26*, 85–96. https://doi.org/10.3354/cr026085

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

SODECOTON. (2022a). Campagne cotonnière 2020/2021. Retrieved June 8, 2024, from https://sodecoton.cm/tout-savoir-sur-la-campagne-cotonniere-2020-2021/

SODECOTON. (2022b). Historique. Retrieved May 8, 2024, from https://sodecoton.cm/historique/

Sultan, B., Bella-Medjo, M., Berg, A., Quirion, P., & Janicot, S. (2009). Multi-scales and multi-sites analyses of the role of rainfall in cotton yields in West Africa. *International Journal of Climatology*, *30*(1), 58–71. https://doi.org/10.1002/joc.1872

Tamoffo, A. T., Weber, T., Akinsanola, A. A., & Vondou, D. A. (2023). Projected changes in extreme rainfall and temperature events and possible implications for Cameroon's socio-economic sectors. *Meteorological Applications*, *30*(2), e2119. https://doi.org/10.1002/met.2119

Tchotsoua, M. (2007). Evaluation des risques d'inondation dans la vallée de la Bénoué en aval du barrage de Lagdo (Cameroun). *Actes des JSIRAUF*.

Tranmer, M., Murphy, J., Elliot, M., & Pampaka, M. (2020). Multiple Linear Regression (2nd Edition). *Cathie Marsh Institute Working Paper*.

Tsozué, D., Haiwe, B. R., Louleo, J., & Nghonda, J. P. (2014). Local Initiatives of Land Rehabilitation in the Sudano-Sahelian Region: Case of Hardé Soils in the Far North Region of Cameroon. *Open Journal of Soil Science*, *04*(01), 6–15. https://doi.org/10.4236/ojss.2014.41002

Vondou, D. A., Guenang, G. M., Djiotang, T. L. A., & Kamsu-Tamo, P. H. (2021). Trends and Interannual Variability of Extreme Rainfall Indices over Cameroon. *Sustainability*, *13*(12), 6803. https://doi.org/10.3390/su13126803

Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Willmott, C. J. (1982). Some Comments on the Evaluation of Model Performance. *Bulletin of the American Meteorological Society*, *63*(11), 1309–1313. https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2

Xu, M., Fralick, D., Zheng, J. Z., Wang, B., Tu, X. M., & Feng, C. (2017). The differences and similarities between two-sample t-test and paired t-test, *29*(3).

Zhang, Q., Zhang, J., Guo, E., Yan, D., & Sun, Z. (2015). The impacts of long-term and year-to-year temperature change on corn yield in China. *Theoretical and Applied Climatology*, *119*(1–2), 77–82. https://doi.org/10.1007/s00704-014-1093-3

Zhang, W., Brandt, M., Tong, X., Tian, Q., & Fensholt, R. (2018). Impacts of the seasonal distribution of rainfall on vegetation productivity across the Sahel. *Biogeosciences*, *15*(1), 319–330. https://doi.org/10.5194/bg-15-319-2018

APPENDICES

Please refer to the supplementary document: Knops, Clara, 2024. Exploration of the statistical relationships between rainfall indices and cotton yields in northern Cameroon, to strengthen the resilience of farmers to climate change. Appendices, Master Water, Specialization Water and Agriculture, AgroParisTech/Institut Agro Montpellier/Univ. Montpellier. 22.

To cite this work: Knops, Clara, 2024. Exploration of the statistical relationships between rainfall indices and cotton yields in northern Cameroon, to strengthen the resilience of farmers to climate change. Thesis of the Master 2 internship, Master Water, Specialization Water and Agriculture, AgroParisTech/Institut Agro Montpellier/Univ. Montpellier. 56.

2 Place Pierre Viala Campus La Gaillarde, 34000 Montpellier, France

RÉSUMÉ

Ce mémoire examine les relations statistiques entre les indices de pluviométrie et les rendements du coton dans le nord du Cameroun, une région fortement dépendante du coton et vulnérable au changement climatique en raison de sa forte variabilité des pluies. Les données quotidiennes de pluies provenant du jeu de données NoCORA ont été interpolées à l'aide du krigeage ordinaire pour calculer des cartes annuelles d'indices de pluviométrie pour un total de 25 indices. Les données de rendement du coton à deux niveaux géographiques différents ont également été fournies par SODECOTON. En appliquant des régressions linéaires simples et multiples, l'impact des indices de pluviométrie sur les rendements du coton a été analysé. Les indices les plus fortement liés de manière statistiquement significative étaient la date de début et de cessation des pluies ainsi que la longueur de la saison, le nombre de jours secs, les périodes secs 10 et 15, la quantité de pluies saisonnières, les jours de pluie, les jours humides 20 et 30, ainsi que les jours des fortes pluies Nos résultats permettront de poursuivre les recherches sur ce sujet, en vue d'analyses prédictives utilisant des données de projection climatique.

Mots clés: Pluviométrie, coton, changement climatique, Cameroun, statistiques, interpolation

ABSTRACT

This thesis investigates the statistical relationships between rainfall indices and cotton yields in northern Cameroon, a region heavily dependent on cotton and vulnerable to climate change due to its high rainfall variability. Daily rainfall data from the NoCORA rainfall dataset was interpolated using Ordinary Kriging to calculate yearly rainfall indices maps for a total of 25 indices. Cotton yield data on two different geographical levels was additionally provided by SODECOTON. Applying simple and multiple linear regression, the impact of the rainfall indices on cotton yields were analyzed. The onset and cessation day of the rainy season as well as the season length, dry days, dry spell consecutive 10 and 15, seasonal rainfall amount, rain days, wet days 20 and 30, as well as heavy rain days were found to be the indices with the strongest, statistically significant relationships. Our findings will allow further research into the topic, serving for prediction-analysis using climate projection data.

Keywords: Rainfall, cotton, climate change, Cameroon, statistics, interpolation