

# **INTEGRATING ENVIRONMENTAL DATA WITH SATELLITE IMAGERY FOR LARGE-SCALE CROP GRAIN NUTRIENT MAPPING**

KWASI APPIAH-GYIMAH OFORI-KARIKARI

June, 2024

**SUPERVISORS:**


Dr. M.T. Marshall

Dr. M. Belgiu

**THESIS ASSESSMENT BOARD:**

Prof. A. Nelson (Chair)

Dr, M.A. Magham (External Examiner, Wageningen University and  
Research)



# **INTEGRATING ENVIRONMENTAL DATA WITH SATELLITE IMAGERY FOR LARGE-SCALE CROP GRAIN NUTRIENT MAPPING**

**KWASI APPIAH-GYIMAH OFORI-KARIKARI**  
Enschede, The Netherlands, June, 2024

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: MGEO - Natural Resource Management, 2022- 2024

**SUPERVISORS:**

Dr. M.T. Marshall

Dr. M. Belgiu

**THESIS ASSESSMENT BOARD:**

Prof. A. Nelson (Chair)

Dr, M.A. Magham (External Examiner, Wageningen University and Research)

#### DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

[this page left intentionally blank]

## ABSTRACT

Food Composition Tables (FCTs), which are widely used in food security analysis, offer country-level snapshots of food nutritional quality. FCTs are highly uncertain because crop nutrient concentrations vary over space and time. Little progress has been made to exploit satellite data for assessing the nutritional status of crop yield. Sentinel-1 and Sentinel-2 have emerged as invaluable data sources for agricultural monitoring from space, because they track crop growth and development at high (10-20m) spatial resolution through time. This study integrates Sentinel-1, Sentinel-2, and other geospatial information across Ethiopia in 2018 into a Random Forest to predict nutrient concentrations in grain yield for important global staple crops. The model was trained and evaluated with data from field measurements collected over the same period. Results were promising: Cu in Sorghum ( $R^2 = 0.73$ ), Mn in Teff ( $R^2 = 0.70$ ), Mg in Teff ( $R^2 = 0.64$ ), Cu in Maize ( $R^2 = 0.60$ ), Mg in Wheat ( $R^2 = 0.58$ ), Mn in Wheat ( $R^2 = 0.58$ ), Mg in Sorghum ( $R^2 = 0.57$ ), and Fe in Sorghum ( $R^2 = 0.54$ ). The findings underscore the effectiveness of Sentinel-2 narrowband vegetation indices and soil properties for nutrient analysis. The results obtained in this study have the potential to improve the availability of crop nutrient data across large spatial areas.

## ACKNOWLEDGEMENTS

I wish to extend my profound gratitude to Almighty God, whose divine guidance and blessings have been helpful throughout this journey.

My heartfelt appreciation also extends to my parents, Juliana and Alexander Ofori-Karikari and my siblings, whose unwavering faith in my abilities and constant prayers have been a source of strength and encouragement.

I want to specially acknowledge my supervisors, Dr. Michael Marshall and Dr. Mariana Belgiu. Their insightful explanations, guidance and motivation have been pivotal in enabling me to comprehend the methods used in this work and its significance.

I am deeply grateful to the chair of the thesis assessment board Prof. Andy Nelson, for his constructive feedback following both the proposal submission and the mid-term defence. Phases. His contributions and feedback got me thinking differently and enriched the quality of my research.

Furthermore, I express my deepest gratitude to Dr. Raymond Nijmeijer, the coordinator, for his arrangements and scheduling of proposal, mid-term and final defence. Without his invaluable assistance, the completion of this thesis would not have been possible.

The utilization of the Virtual Geospatial Computing Platform (CRIB) has been really helpful for this research. Therefore, I extend my sincere thanks to Dr. Serkan Girgin and his team at ITC for their dedication to developing and maintaining this essential tool.

Lastly, I would like to acknowledge my ITC colleague Swagatalaxmi Das for her support during my second year at ITC. Her assistance has been invaluable in navigating through various personal challenges and contributing to the success of my project.

# TABLE OF CONTENTS

---

1.	Introduction.....	7
1.1.	Background and motivation .....	7
1.2.	Main Research Objectives.....	11
2.	Study area .....	14
3.	Datasets and Methodology.....	16
3.1.	Crop Yield Grain Nutrient Composition of Cereals.....	16
3.2.	Sentinel-1 Image processing .....	17
3.3.	Sentinel-2 Image Processing.....	17
3.4.	Polarimetric and Sentinel-2 based Vegetation Indices.....	17
3.5.	Rainfall.....	20
3.6.	ERA5-Land Data and Processing.....	21
3.7.	Soil Data and processing.....	21
3.8.	Digital Elevation Model and processing .....	22
3.9.	Moderate Resolution Imaging Spectroradiometer (MODIS) product.....	22
3.10.	Random Forest Models and Optimization .....	23
4.	Results.....	29
4.1.	Descriptive statistics of Crop Grain Nutrient composition .....	29
4.2.	Findings.....	31
4.3.	Micronutrient Analysis for Teff.....	32
4.4.	Macronutrient Analysis for Teff.....	43
5.	Discussion.....	50
5.1.	Research Question for Objective 1:.....	50
5.2.	Research Question for Objective 2(a): .....	52
5.3.	Research Question for Objective 2(b):.....	53
5.4.	Research Question for Objective 3:.....	53
6.	Conclusion and Recommendations .....	57
6.1.	Conclusion.....	57
6.2.	Recommendations.....	57
6.3.	Societal impact .....	58
6.4.	Ethical Considerations.....	58
7.	APPENDICES .....	71
	Appendix 1. Data management.....	71
	Appendix 2. Results for barley.....	73
	Appendix 3. Results for maize .....	95
	Appendix 4. Results for sorghum.....	118
	Appendix 5. Results for teff .....	138
	Appendix 6. Results for wheat.....	161





## LIST OF FIGURES

---

Figure 1. Map of cropland areas in Ethiopia (Kumssa et al., 2022) .....	14
Figure 2. Crop Calendar of Ethiopia source: <a href="https://fews.net/east-africa/ethiopia">https://fews.net/east-africa/ethiopia</a> .....	15
Figure 3. Technical workflow proposed in this research to estimate various nutrients in staple crops using machine learning and a wide variety of co-variates .....	16
Figure 4. VSURF Result plots. Refer to section “Insights of VSURF plots” for explanation .....	24
Figure 5. ELI5 permutation importance. A: Permutation importance of variables from VSURF used in a model; B: Permutation importance of variables after manually discarding variables highlighted in red; C: Variables used in the final model after discarding variables in B.....	26
Figure 6. Violin plots showing Ca grain nutrient composition .....	29
Figure 7. Violin plots showing Cu grain nutrient composition.....	29
Figure 8. Violin plots showing Fe grain nutrient composition .....	30
Figure 9. Violin plots showing K grain nutrient composition.....	30
Figure 10. Violin plots showing S grain nutrient composition.....	30
Figure 11. Violin plots showing Mg grain nutrient composition.....	30
Figure 12. Violin plots showing Mn grain nutrient composition.....	30
Figure 13. Violin plots showing P grain nutrient composition .....	30
Figure 14. Violin plots showing Zn grain nutrient composition .....	31
Figure 15. Graphical representation of results for Teff grain and calcium nutrient composition.. .....	33
Figure 16. Graphical representation of results for Teff grain and Cu nutrient composition.....	35
Figure 17. Graphical representation of results for Teff grain and Fe nutrient composition.....	37
Figure 18. Graphical representation of results for Teff grain and Mn nutrient composition.....	38
Figure 19. Graphical representation of results for Teff grain and Zn nutrient composition.....	40
Figure 20. Graphical representation of results for Teff grain and Mg nutrient composition.. .....	42
Figure 21: Graphical representation of results for Teff grain and potassium nutrient composition.. .....	45
Figure 22: Graphical representation of results for Teff grain and P nutrient composition.. .....	46
Figure 23: Graphical representation of results for Teff grain and S nutrient composition.....	48
Figure 24: Tabular representation of statistical accuracy ( $R^2$ ) of models for all crops and all nutrients involved in this study. ....	49

## LIST OF TABLES

---

Table 1. Sentinel-1 Bands and specifications .....	18
Table 2. Sentinel-2 Bands and specifications .....	18
Table 3. Sentinel-1 Polarimetric indices .....	18
Table 4. Sentinel-2 vegetation indices .....	19
Table 5. Distribution of field crop nutrient measurements per regions. ....	24

# 1. INTRODUCTION

## 1.1. Background and motivation

The United Nations Sustainable Development Goal 2 (SDG 2) emphasizes achieving "Zero Hunger" by 2030. However, ensuring food security is a complex issue with multiple, interrelated dimensions. It goes beyond simply ensuring sufficient food production (availability) to encompass accessibility, utilization, and stability of the food supply (United Nations, 2015). While noteworthy progress has been made in increasing global food production, ensuring everyone has access to a safe, nutritious, and culturally appropriate diet remains a challenge (United Nations, 2023). Accessibility is influenced by factors like income levels, food distribution networks, and geographic location. Even with sufficient food available, individuals in remote areas or those living in poverty may struggle to afford or physically obtain the nutritious food they need. Utilization refers to the body's ability to convert food into energy and essential nutrients. This is influenced by factors like access to clean water, sanitation, and healthcare. Without these basic necessities, even a seemingly adequate diet may not provide the necessary building blocks for optimal health. Additionally, knowledge of proper nutrition and food preparation plays a crucial role in utilization. Finally, stability refers to the resilience of food systems to disruptions and shocks. Climate change, economic downturns, and political instability can all threaten food security. Building resilient food systems that can withstand these disruptions is essential for long-term food security (The State of Food Security and Nutrition in the World 2021, 2021).

### Hidden Hunger

While progress has been made in the areas mentioned above, a significant challenge remains, namely hidden hunger caused by micronutrient deficiencies. Individuals may have access to sufficient calories but lack the essential vitamins and minerals including zinc, iron, calcium, copper, magnesium, or manganese needed for optimal health. Diagnosing hidden hunger is difficult, as deficiencies often manifest only in later stages, leading to long-term health consequences. Pregnant women, young children, and individuals living in poverty are particularly vulnerable to hidden hunger.

Human health relies on a balanced intake of both macro and micronutrients (Ritchie, 2021). Macronutrients, such as nitrogen, phosphorus, and potassium, are crucial for basic bodily functions like energy production, cellular structure, and muscle development. Micronutrients, required in smaller quantities, are equally vital. Iron is essential for oxygen transport in the blood, zinc for immune function and DNA synthesis, and manganese for bone formation and metabolism. Deficiencies in these nutrients can lead to a cascade of severe health consequences, including anaemia, weakened immunity, impaired growth and development, and even increased risk of chronic diseases (Gombart et al., 2020).

Children experiencing chronic micronutrient deficiencies may suffer from stunting, a condition characterized by low height for age. This can lead to cognitive decline, reduced learning capacity, and lower future earning potential. Additionally, hidden hunger can increase susceptibility to infections, further compromising health and well-being (Ritchie & Roser, 2017).

## **The Challenge of Crop Nutrient Measurement**

Addressing nutrient deficiencies requires effective methods for measurement. However, traditional methods used in both public health and agriculture have limitations. Diagnosing hidden hunger through blood draws and surveys is expensive, time-consuming, and requires specialized equipment. This limits the frequency and comprehensiveness of data collection, making it difficult to identify regions most at risk of micronutrient deficiencies.

Traditional methods like wet chemical analysis are labour-intensive and time-consuming, hindering their use for macroscale or frequent crop nutrient content measuring and monitoring. Additionally, these methods prevent us from generating spatial information needed to capture the variability in soil properties and environmental conditions across different agricultural landscapes. This leads to imprecise nutrient management strategies that can negatively impact crop productivity and contribute to environmental degradation (Gibson, 2005).

## **Food Composition Tables (FCTs)**

Food Composition Tables (FCTs) are valuable tools for general dietary assessments as they provide averaged nutrient values for various foods. These tables are essential for evaluating the nutritional composition of diets, recipes, or commercially available food items, offering a comprehensive nutrient breakdown for each listed item (Ocké et al., 2021). However, FCTs have significant limitations, especially in the context of agricultural management and precision nutrition (Traka et al., 2020).

While FCTs are useful for broad dietary assessments, they do not account for variations in nutrient content within regions and countries. This lack of precision limits their effectiveness in informing site-specific nutrient management practices crucial for maximizing crop nutrient uptake and enhancing grain nutritional quality. For example, differences in soil properties, farming practices, and environmental conditions can lead to substantial variability in nutrient content, which FCTs cannot capture (FAO, 2018).

Additionally, the nutrient values in FCTs are often derived from a variety of estimation methods rather than direct chemical analysis. Common approaches for estimating nutrient values include utilizing values from foods that are similar, and computing values from various forms of the identical food and estimating from household recipes or commercial product formulations. Other methods involve converting values from nutrient labels of commercial products, calculating from product standards, or assuming zero values where data are missing. Quality assurance processes and programs for validating nutrient composition are necessary for ensuring the accuracy of these estimated values (Schakel et al., 1997). However, even with these measures, the reliance on estimated data introduces potential inaccuracies (FAO, 2018). These limitations emphasize the need for more precise and spatially explicit nutrient assessment methods.

## **Remote Sensing Data and Technologies for Agriculture Applications**

Despite the limitations, recent advancements in remote sensing and machine learning present promising avenues for addressing these issues. Remote sensing offers a powerful, non-invasive approach for collecting data on crop and soil nutrient content across vast areas, overcoming the limitations of traditional methods. It relies on the fundamental principle that different objects on Earth's surface interact with electromagnetic radiation in unique ways. Sensors mounted on satellites, airplanes, or drones capture the reflected or absorbed radiation across various wavelengths of the electromagnetic spectrum (Jensen, 2009). By analyzing this spectral information, scientists and farmers can extract valuable information about crop and soil properties.

The selection of the appropriate remote sensing sensor for a specific agriculture application depends on the desired level of detail along with spatial and temporal coverage. Satellites offer a synoptic view of vast regions, making them suitable for monitoring large-scale variations in crop and soil properties (Thenkabail et al., 2004). Once data is acquired, advanced machine learning algorithms can be employed to analyze and interpret the complex spectral information. These spectra and algorithms can identify patterns and relationships between the spectral signatures and biophysical parameters like fraction of photosynthetic active radiation, leaf area index, and crop biomass along with biochemical properties like canopy water content, leaf nitrogen and chlorophyll content (Kganyago et al., 2024).

By analyzing specific wavelengths sensitive to chlorophyll content, plant nitrogen status, and other nutrient-related factors, machine learning can estimate the spatial variability of soil and crop nutrients within a field (Benos et al., 2021). Spectral signatures can be used to infer soil properties like moisture content, organic matter, and potential nutrient availability (Castaldi et al., 2023). This information is crucial for developing targeted soil management practices. Remote sensing data can be integrated with weather data from satellites or ground stations to create detailed maps of temperature, precipitation, and other climate variables across a field. This allows for a holistic understanding of how environmental factors influence crop growth and nutrient uptake (Nieto et al., 2021).

### **Sentinel-1 and Sentinel-2 satellite images**

Incorporating Sentinel-1 and Sentinel-2 alongside other digital environmental data for crop and grain mapping and monitoring presents a complementary approach to enhancing agricultural monitoring and management. Sentinel-1 provides synthetic aperture radar (SAR) images, which are valuable for monitoring the surface of the Earth regardless of cloud cover or daylight conditions, offering insights into soil moisture, vegetation biomass, and surface roughness (Torres et al., 2012). On the other hand, Sentinel-2 offers optical imagery capable of capturing detailed information about vegetation indices, water content, and mineral content, which are crucial for assessing crop health and productivity (Drusch et al., 2012). Sentinel-2 imagery, although impacted by cloud cover, has the capability to detect photosynthetic pigments and the amount of chlorophyll absorbed by the crops (Darvishzadeh et al., 2019). Its red-edge bands are sensitive to leaf nitrogen, its near infrared bands are sensitive to leaf and canopy structures as well as leaf and canopy water content. The short-wave infra-red bands are sensitive to non-pigment compounds like cellulose, lignin, and proteins (Weiss et al., 2020). These bands can be used to derive various vegetation spectral indices relevant for the derivation of crop biophysical, biochemical and crop phenology parameters (Kganyago et al., 2021). The spatial (10 and 20m) and temporal resolutions (average 5 days) along with their spectral capacities make them suitable for small-scale farm research when compared with LANDSAT and Moderate Resolution Imaging Spectroradiometer (MODIS) products (Khan et al., 2023).

There is evident significant progress in the use of Sentinel-1 and 2 data for crop yield and nutrient mapping. Rao et al., (2021) highlighted the potential of Sentinel-1, and Sentinel-2 to map crop types on non-commercial small farms. This synergy leverages the strengths of each sensor type, namely Sentinel-1's ability to penetrate clouds and observe in all weather conditions and Sentinel-2's high spatial resolution for detailed crop discrimination, the frequent revisit times of both sensors for up-to-date information. The combination of Sentinel-1 and Sentinel-2 data has also been explored and shown to improve the accuracy of land use and land cover mapping (Steinhausen et al., 2018), as well as for tree species classification (Axelsson et al., 2021). For crop and soil nutrient studies Zhang et al., (2023) demonstrated that Sentinel-2 data can be used to accurately map the total nitrogen composition in soils. Sharifi, (2020) confirmed that the computation of vegetation indices from Near Infrared and Red-edge bands of Sentinel-2 leads to better predictions of nutrient uptake of the maize crop. Fernandes et al., (2024) and Mendes et al., (2023) also demonstrated the ability of Sentinel-2 data to predict and quantify nutrients in different crops and grasslands. These

applications demonstrate the versatility of Sentinel data in supporting a wide range of agricultural applications.

### **Geospatial Factors relevant for Crop Nutrient Mapping**

Geospatial factors, including topography, climate, and soil characteristics, are instrumental in shaping nutrient distribution and crop performance, thereby influencing agricultural productivity. The integration of these factors into nutrient mapping models significantly enhances their predictive power and accuracy (Khurana et al., 2022). Topographic features, such as slope and aspect, influence water drainage and soil erosion, which in turn affect soil fertility and crop health (Li et al., 2021). Climatic variables, including temperature and precipitation, modulate nutrient uptake and crop growth dynamics, further complicating nutrient management (Chen et al., 2022). Recognizing the connections between geospatial factors and crop nutrients is important for developing agricultural strategies that cater to the unique needs of different farm environments.

Recent research also highlighted the significance of integrating geospatial modelling in defining agricultural management zones, taking advantage of local topographical features and soil characteristics. Jena et al., (2022) demonstrated the use of geospatial modelling in identifying areas with similar soil and terrain characteristics to facilitate targeted crop management practices. This approach not only aids in optimizing resource allocation but also contributes to sustainable agricultural practices by aligning interventions with the inherent capabilities of the land. Moreover, the relationship between topography and soil nutrition content, as discussed by Karaca & Gülser, (2018), highlights the critical role of landform in determining soil fertility and crop yields.

### **Machine Learning Techniques in Remote Sensing for Agriculture**

Machine learning techniques, notably Random Forests (RF) developed by Breiman, (2001), have emerged as indispensable tools in the field of agriculture, particularly for predicting crop yields and tracking soil nutrients (Prasath et al., 2023). These techniques excel in handling large datasets, including remote sensing imagery, to show the importance of predictors thereby facilitating optimized agricultural practices and enhanced productivity. The robustness of RFs, coupled with their ability to navigate complex, nonlinear relationships inherent in agricultural data, positions them as highly effective models for predicting crop yields and identifying soil nutrient levels with remarkable accuracy (Hengl et al., 2017). Furthermore, the interpretability offered by RFs, despite their complexity, allows for clear insights into feature importance, guiding subsequent data collection and analysis efforts. This dual advantage of accuracy and interpretability has made RFs a preferred choice for numerous agricultural applications.

Many agricultural studies have harnessed the power of RFs and remote sensing data to tackle the challenges of crop yield prediction, crop nutrient estimation and soil nutrient monitoring. For instance, Priya et al., (2018) demonstrated the efficacy of Random Forests in accurately predicting crop yields. Jeong et al., (2016) highlighted the superior performance of Random Forests over traditional statistical models in predicting yields for wheat, potato, and maize crops. Additionally, the integration of machine learning techniques with remote sensing data, such as Sentinel-1 and Sentinel-2, has been explored for various agricultural applications, emphasizing the potential of these combined approaches to revolutionize agricultural practices. The studies mentioned above collectively highlight the potential of machine learning techniques, particularly Random Forests, in advancing agricultural productivity and sustainability through precise forecasting and informed decision-making.

## 1.2. Main Research Objectives

This main aim of this study is to build on the foundational work of Belgiu et al., (2023) which demonstrated the potential of PRISMA and Sentinel-2 images for predicting composition of macro and micronutrient in grains of cereals in Italy. The study highlighted the effectiveness of multispectral broadband and hyperspectral remote sensing in estimating grain nutrient composition, showing promising results for various crops such as soybean, wheat, corn, and rice. The study highlighted the feasibility of using remote sensing images to provide cost-effective, timely, and spatially explicit representations of crop grain nutritional quality. However, several areas require further exploration to enhance the practical application and accuracy of these methods. This study aims to capitalize on the initial research by addressing several critical aspects that remain unexplored.

First, while Belgiu et al. (2023) focused on a specific set of crops and a single farm in Italy, this study aims to broaden the scope by including a variety of crops (barley, teff, sorghum, maize, and wheat) and expanding the geographic focus to multiple agroecological zones in Ethiopia. The proposed study area consists of many different farms of sizes less than one hectare, with variable growing seasons and farm management practices (Headey et al., 2014). This expansion will allow for the assessment of the generalizability of remote sensing techniques across different environmental conditions and farming practices, thus providing a more comprehensive understanding of their applicability.

Secondly, to improve the accuracy and robustness of grain nutrient concentration predictions, this study will explore machine learning algorithms and integrate multi-temporal, multi-spectral, and multi-source remote sensing data. Belgiu et al. utilized Partial Least Squares Regression (PLSR) and two-band vegetation indices (TBVIs) along with random forests, achieving promising results with 40 observations of data. Building on this, this study will investigate the potential of RFs for prediction with variable datasets of more than 100 observations.

Third, understanding the interactions between soil properties and crop nutrient content is crucial for accurate nutrient mapping. This study will delve into these soil-nutrient dynamics by utilizing coarse-resolution soil maps and integrating geospatial factors like topography and climate for grain nutrient predictions. By incorporating these additional variables, a more holistic view can be provided of the factors influencing nutrient uptake and distribution in crops

By addressing these objectives, the study aims to investigate the applicability of remote sensing data for estimating crop grain nutrient status using RFs. The crops for study are barley, maize, sorghum, teff and wheat. The micronutrients to be analyzed include Calcium (Ca), Copper (Cu), Iron (Fe), Manganese (Mn), and Zinc (Zn). The macronutrients to be analyzed include Potassium (K), Magnesium (Mg), Phosphorus (P), Sulphur (S). The integration of various datasets with machine learning will lead to the creation of more detailed and location-specific crop nutrient information. This will provide spatially variable data for improving dietary assessments and agricultural management practices, supporting global efforts to achieve food security and nutritional health.

### **1.2.1. Sub-Objectives**

1. To identify the most important environmental factors including climatic, soil and topographic factors using RFs and explain their impacts on crop grain nutrient composition at different stages of crop growth using PDPs.
2. To identify the most important Sentinel-1 and Sentinel-2 derived data using RFs and explain their impact on estimating crop grain nutrient composition at different stages of crop growth using PDPs.
3. To evaluate the potential of remotely sensed derived biophysical properties such as leaf area index (LAI) and fraction of photosynthetic active radiation (FPAR) on crop grain nutrient composition using RFs at different stages of crop growth.

### **1.2.2. Research Questions**

#### **Research Question for Objective 1:**

- 1a. What are the important environmental factors (climatic, soil and topographic factors) for the estimation of nutrient composition of crop grains at different crop growth stages?
- 1b. What are the impacts of the important environmental factors on estimation of nutrients at different crop growth stages?

#### **Research Question for Objective 2(a):**

- 2a i). What are the important Sentinel-1 bands and polarimetric indices for the estimation of nutrient composition in crop grains across diverse agroecological zones at different crop growth stages?
- 2a ii). What are the impacts of the important Sentinel-1 bands and polarimetric indices on the prediction of nutrient composition in crop grains across diverse agroecological zones at different crop growth stages?

#### **Research Question for Objective 2(b):**

- 2b i). What are the important Sentinel-2 bands and spectral indices for the estimation of nutrient composition in crop grains across diverse agroecological zones at different crop growth stages?
- 2b ii). What are the impacts of the important Sentinel-2 bands and spectral indices on the prediction of nutrient composition in crop grains across diverse agroecological zones at different crop growth stages?

#### **Research Question for Objective 3:**

What is the impact of biophysical properties (FPAR AND LAI) on the prediction the nutrient composition of crop grains in diverse agroecological zones at different crop growth stages?



### **1.2.3. Expected outcomes.**

1. Variations in rainfall and temperature are expected to significantly influence the nutrient composition of crop grains at different stages of crop growth.
2. Soil properties, including composition, structure, and texture, are expected to directly affect the nutrient concentration in crop grains and lead to variations in crop nutrition.
3. Topographic factors such as elevation, slope, and the Topographic Wetness Index are expected to play an important role in determining the nutrient composition of crop grains.
4. Sentinel-1 derived data, including bands and polarimetric indices, are expected to play a minimal role in the prediction of nutrient composition in crop grains.
5. Sentinel-2 data, particularly the red-edge and short-wave infrared (SWIR) bands, as well as vegetation indices derived from the combination of both, are expected to impact the prediction of nutrient composition in crop grains at different stages of crop growth.
6. FPAR and Leaf Area Index LAI are expected to have a positive impact on the prediction of nutrient composition in crop grains.

## 2. STUDY AREA

Ethiopia is home to a multitude of crops, both staple and cash crops, reflecting the country's agricultural diversity. The agricultural landscape of Ethiopia serves as a rich and dynamic study area for agricultural research, offering a diverse range of agroecological zones and farming practices. Ethiopia is characterized by distinct climatic regions, including highland areas, lowlands, and arid zones, each influencing agricultural productivity differently. The varied topography, spanning from high plateaus to low-lying plains, contributes to diverse soil types and agricultural conditions. The seasonal rainfall patterns and temperature variations further impact crop growth cycles (Haileslassie et al., 2020).

Agricultural activities are intricately tied to the country's two distinct grain growing seasons: belg and meher. The belg season, which generally occurs within the between the second and fifth months of the year (from February to May), is shorter, while the main meher season extends from May to September. The success of cereal grain production, including crops like corn, wheat, sorghum, barley, and teff, hinges significantly on the reliability of rainfall patterns during the belg season (Ethiopia Climate and Agriculture, 2017). Variability in the onset and completion of these growing seasons in certain years directly impacts both crop production and quality as can be seen in Figure 2. Geographically, Ethiopia's landlocked status compounds challenges exacerbated by climate change. The nation encompasses two agriculturally significant regions: the lowlands and the highlands, each grappling with heightened temperatures, prolonged droughts, precipitation fluctuations, soil erosion, and desertification. The lowlands, where a significant portion of Ethiopia's livestock grazing land is situated, face distinct climate-related challenges compared to the highlands, which predominantly host subsistence farms (Dixon, 2018). The study can be seen in Figure 1.

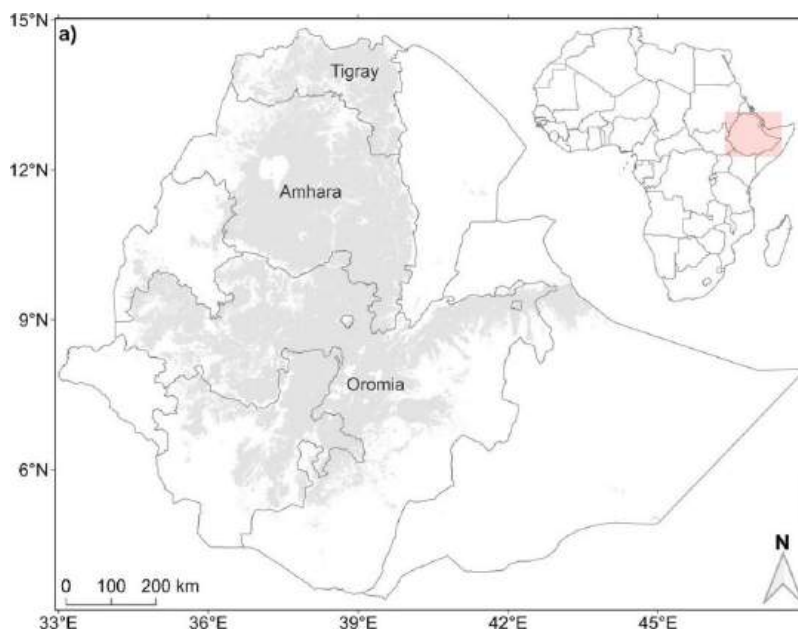


Figure 1. Map of cropland areas in Ethiopia (Kumssa et al., 2022)

Cereal grains and soil samples were only selected from identified crop areas (in grey shades) in the respective regions in Ethiopia. The map inset shows the location of Ethiopia on the African map (in pink) (Kumssa et al., 2022).

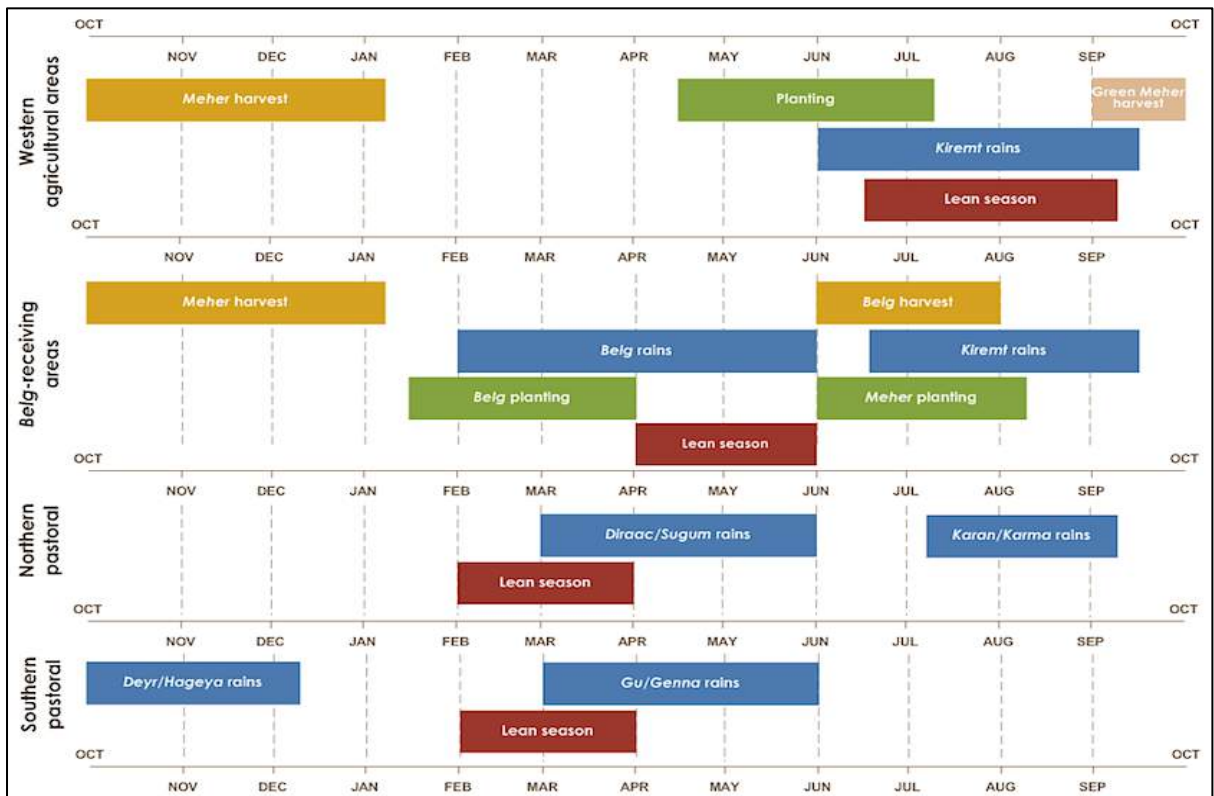


Figure 2. Crop Calendar of Ethiopia source: <https://fews.net/east-africa/ethiopia>

### 3. DATASETS AND METHODOLOGY

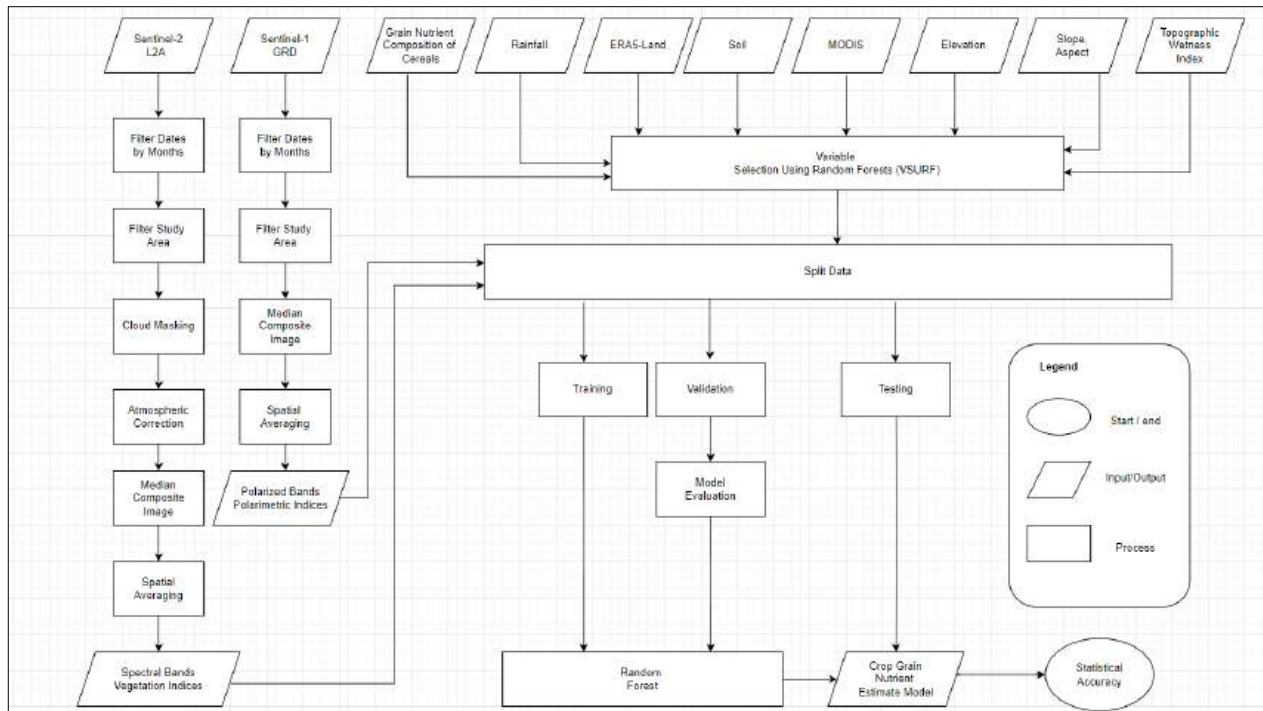


Figure 3. Technical workflow proposed in this research to estimate various nutrients in staple crops using machine learning and a wide variety of co-variables

#### 3.1. Crop Yield Grain Nutrient Composition of Cereals

The grain nutrient data used in this study is detailed by Kumssa et al., (2015), who described that samples were taken from locations where the probability of the crop lands under cultivation was greater than 85 percent ( $\geq 0.85$ ). Another criterion was that the croplands should be within 2.5 km of a road.

The collection of cereal grain samples from farms were completed in November 2017–February 2018 in Amhara and in November 2018–February 2019 in Oromiya, Tigray, and Amhara. The data collectors took grain samples from the fields only after farmers agreed it was okay to do so. Within a selected farm, samples were taken from a 100 m<sup>2</sup> plot. Five subsample points were located within the farm, and grain samples were collected at each of the five subsample points. These grain subsamples were stored together in a single envelope. Kumssa et al., (2015) state that the decision to not discard data from locations with positional uncertainties in their study was informed by the use of robust variogram estimators, which are known to be resilient to the effects of spatial outliers. The grain samples were then taken to the laboratory and further analysis were carried out on over 25 soil and crop nutrients samples. This study focuses only on the nutrients mentioned in the main objectives and the data collected in the 2018/2019 crop season

### **3.2. Sentinel-1 Image processing**

The Sentinel-1 dual polarisation C-band SAR data was acquired from Google Earth Engine. The image collection contains Ground Range Detected (GRD) scenes and has been processed with the Sentinel-1 Toolbox to create a calibrated ortho-corrected product. Processing involving GRD border noise removal, thermal noise removal, application of radiometric calibrated values and terrain correction were already completed. The relevant bands for this study are the VV Band (Single co-polarization, vertically transmit, vertically receive) and the VH Band (Dual-band cross-polarization, vertically transmit, horizontally receive) (Veci et al., 2014). These bands have successfully been used in crop mapping applications. According to the Google Earth Engine description, the image collection is updated daily.

Monthly median composites were generated based on calendar months corresponding to the planting schedules of the specified crops. Additionally, composites for the month preceding the planting period were included, as soil and environmental conditions prior to planting are crucial for crop development. The field data points were overlaid on the monthly composites. As suggested by (Congalton (, 2001), buffer with a kernel size three times the highest Sentinel-2 spatial resolution used in this study was applied to extract the average band values. The resulting data was then downloaded in tabular format. There were no data gaps. See Table 1 for detailed information on Sentinel-1.

### **3.3. Sentinel-2 Image Processing**

Mostly cloud-free Sentinel-2 Level 1C images, which are top-of-atmosphere images, were accessed from Google Earth Engine. The bands used in this study had spatial resolutions of 10 and 20 meters (see Table 2 for detailed Sentinel-2 band information). The Sentinel-2 images were processed to Level 2A products using the Bayesian atmospheric correction method developed by Yin et al., (2022). Cloud masking was performed using the Hollstein cloud mask (Hollstein et al., 2016). Monthly median composites were generated using calendar months throughout the crop growing season, including one month prior. The field data was overlaid on these composites, and average band values were extracted in tabular format using a kernel of the same size as that used for extracting Sentinel-1 data. After performing cloud masking, some data gaps remained due to certain months having all cloudy images. To avoid the loss of data, mean values were imputed (Rosenthal, 2017).

### **3.4. Polarimetric and Sentinel-2 based Vegetation Indices**

Polarimetric indices utilize the polarization properties of electromagnetic waves to offer detailed insights into the structural and dielectric characteristics of the Earth's surface, with an emphasis on vegetation. By analyzing the way polarized light interacts with vegetation, these indices help distinguish between various types of vegetation, assess soil moisture levels, and evaluate surface roughness. Derived from Sentinel-1 SAR data, polarimetric indices are invaluable for understanding the complex interactions between electromagnetic waves and different surface features (Mandal et al., 2019). This capability is particularly beneficial for agricultural monitoring and soil moisture assessment, where precise structural information is critical as in the case of this study.

Vegetation indices, on the other hand, are spectral transformations combining two or more spectral bands to assess the quantitative or qualitative properties of vegetation. These indices, derived from Sentinel-2 spectral bands, play a crucial role in monitoring various aspects of vegetation health and dynamics. They enable the assessment of biomass, crop water use, crop stress, crop structure, crop health, photosynthetic activity, and crop growth dynamics. By analyzing the differences in light absorption and reflection in specific spectral bands, vegetation indices provide reliable estimates of vegetation vigour and condition (Bannari et al., 1995). Combining polarimetric and vegetation indices creates a comprehensive approach to vegetation monitoring. While polarimetric indices provide structural and dielectric information, vegetation indices offer insights into the physiological state of the vegetation (Holtgrave et al., 2020). This integration will potentially

enhance the accuracy and depth of the analysis. These indices were calculated in earth engine and extracted in tabular format using the same method of extraction of Sentinel-1 and -2 data to ensure consistency. See Table 3 and Table 4 for sentinel 1 and sentinel 2 derived indices respectively.

Table 1. Sentinel-1 Bands and specifications

Satellite	Polarization	Frequency Range (GHz)	Centre frequency (GHz)	Band width (MHz)	Spatial resolution (m)
Sentinel-1	VV	4-8	5.405	0-100	10
	VH	4-8	5.405	0-100	10

Table 2. Sentinel-2 Bands and specifications

Satellite	Spectral band	Spectral Range(nm)	Central wavelength (nm)	Bandwidth (nm)	Spatial resolution (m)
Sentinel-2	Band 2 blue	458-523	490	65	10
	Band 3 green	543-578	560	35	10
	Band 4 red	650-680	665	30	10
	Band 5 vegetation red edge	698-713	705	15	20
	Band 6 vegetation red edge	733-748	740	15	20
	Band 7 vegetation red edge	773-793	783	20	20
	Band 8 NIR	785-900	842	115	10
	Band 8a narrow NIR	855-875	865	20	20
	Band 11 SWIR	1565-1655	1610	90	20
	Band 12 SWIR	2100-2280	2190	180	20

Table 3. Sentinel-1 Polarimetric indices

Abbreviation	S1 Polarimetric Index	Expression	Application/Characteristics	Citation
DPSVIm	Modified Dual Polarization SAR Vegetation Index	$(VV * VV + VV * VH) / 1.414213562$	Measures vegetation biomass and water content using dual polarization SAR data.	(dos Santos et al., 2021)

<b>RVIm</b>	<b>Radar Vegetation Index modified</b>	$(4 * VH) / (VV + VH)$	<b>Enhances the detection of vegetation cover and health by adjusting for terrain effects.</b>	(Nasirzadehdizaji et al., 2019)
<b>Pol</b>	<b>Normalized Index</b>	$(VH - VV) / (VH + VV)$	<b>Normalizes radar signal intensities to account for variations in incidence angle and surface roughness.</b>	(Hird et al., 2017)
<b>CR</b>	<b>Cross Ratio</b>	$VV / VH$	<b>Compares the vertical and horizontal polarizations to assess soil moisture levels and vegetation structure.</b>	(Frison et al., 2018)

Table 4. Sentinel-2 vegetation indices

<b>Abbreviation</b>	<b>S2 Vegetation Index</b>	<b>Expression</b>	<b>Application/Characteristics</b>	<b>Citation</b>
<b>ARI</b>	<b>Anthocyanin reflectance Index</b>	$(1 / B3) - (1 / B5)$	<b>Anthocyanin-physiological status of plant which are important indicators of different types of plant stress.</b>	(Gitelson et al., 2003)
<b>ARVI</b>	<b>Atmospherically Resistant Vegetation Index</b>	$(B8 - B6 - (B2 - B6)) / (B8 + B6 - (B2 - B6))$	<b>corrects for the atmospheric scattering effects using blue light reflectance</b>	(Kaufman & Tanré, 1992)
<b>CI_RE</b>	<b>Chlorophyll index at red edge</b>	$(B8 - B5) - 1$	<b>Chlorophyll</b>	(Gitelson et al., 2003)
<b>DSWI</b>	<b>Disease Water Stress Index</b>	$(B8 - B3) / (B11 + B4)$	<b>sensitive to stress due to water shortage and plant damage</b>	(Bochenek et al., 2018)
<b>EVI</b>	<b>Enhanced Vegetation Index</b>	$2.5 * ((B8 - B4) / (B8 + 6 * B4 - 7.5 * B2 + 1))$	<b>improved version of NDVI that reduces atmospheric influences</b>	(Matsushita et al., 2007)
<b>EVIredEdge</b>	<b>Red-Edge Enhanced Vegetation Index</b>	$2.5 * ((B8 - B6) / (B8 + 6 * B6 - 7.5 * B2 + 1))$	<b>estimate various biophysical parameters, such as LAI, chlorophyll content, and canopy water content,</b>	
<b>GCI</b>	<b>Green Chlorophyll Index</b>	$(B5 / B3) - 1$	<b>estimate chlorophyll content in vegetation</b>	(Gitelson et al., 2003)
<b>GNDVI</b>	<b>Green Normalized Difference Vegetation Index</b>	$(B8 - B3) / (B8 + B3)$	<b>more sensitive than NDVI to different concentration rates of chlorophyll, which is highly correlated at nitrogen.</b>	(A. A. Gitelson & Merzlyak, 1998)
<b>HMSSI</b>	<b>Heavy metal stress sensitive index</b>	$((B8 - B5) - 1) / ((B5 - B2) / B3)$	<b>Heavy metal</b>	(Z. Zhang et al., 2018)
<b>IRECI</b>	<b>Inverted Red-Edge Chlorophyll Index</b>	$(B7 - B4) / (B5 / B6)$	<b>canopy chlorophyll content</b>	(Jiang et al., 2023)
<b>MCARI</b>	<b>Modified Chlorophyll Absorption Ratio Index</b>	$((B5 - B4) - 0.2 * (B5 - B3)) * (B5 / B4)$	<b>responsive to leaf chlorophyll concentration and ground reflectance</b>	(Wu et al., 2008)
<b>MSR_RE</b>	<b>Modified simple ratio at red edge</b>	$((B8 / B4) - 1) / ((B8 / B4) + 1) ** 0.5$	<b>Chlorosis, increased sensitivity to vegetation biophysical parameters.</b>	(Wu et al., 2008)
<b>MTCI</b>	<b>MERIS Terrestrial Chlorophyll Index</b>	$(B6 - B5) / (B5 - B4)$	<b>chlorophyll content of vegetation canopies</b>	(Dash & Curran, 2004)
<b>NDTI</b>	<b>Normalized Difference Turbidity Index</b>	$(B11 - B3) / (B11 + B3)$	<b>assess water turbidity, which indicates the presence of suspended particles</b>	(Bid & Siddique, 2019)
<b>NDVI</b>	<b>Normalized difference vegetation</b>	$(B8 - B4) / (B8 + B4)$	<b>Green Biomass, leaf area index (LAI)</b>	(Sims & Gamon, 2002)

NDVI_RE	Normalized difference vegetation at the red edge	$(B8 - B5) / (B8 + B5), (B8 - B6) / (B8 + B6), (B8 - B7) / (B8 + B7)$	Chlorophyll	(A. Gitelson & Merzlyak, 1994)
NDWI	Normalized Difference Water Index	$(B8 - B12) / (B8 + B12)$	measures the presence and abundance of water.	(Gao, 1996)
NPCI	Normalized pigment chlorophyll index	$(B4 - B2) / (B4 + B2)$	Chlorophyll	(Huang et al., 2014)
NRI	Nitrogen reflectance index	$(B3 - B4) / (B3 + B4)$	Nitrogen	(Huang et al., 2014)
PhRI	Physiological reflectance index	$(B3 - B2) / (B3 + B2)$	Solar utilization efficiency during crop development. Determines disease and abiotic stress	(Huang et al., 2014)
PSRI	Plant senescence/reflectance index	$(B5 - B2) / B3$	Plant Senescence	(Yu et al., 2018; Z. Zhang et al., 2018)
PSSRa	Pigment Specific Simple Ratio Chlorophyll a Index	$B7 / B4$	chlorophyll index	(Psomiadis et al., 2017)
RERVI	Red Edge Ratio Vegetation Index	$B8 / B6$	estimating biomass and chlorophyll content	
RVI	Ratio Vegetation Index	$B8 / B4$	eliminate various effects such as: irradiance (topography), transmittance (atmospheric effects)	(Y. Tan et al., 2019)
RVSI	Red-Edge Vegetation Stress Index	$((B5 + B6) / 2) - B6$	assessing vegetative health and identifying stressors early	
S2REP	Sentinel-2 Red-Edge Position Index	$705 + 35 * ((B4 + B7) / 2 - B5) / (B6 - B5)$	both crop (chlorophyll content) N and growth status	(Eleveld et al., 2018)
SAVI	Soil Adjusted Vegetation Index	$((B8 - B4) / (B8 + B4 + 0.5)) * (1 + 0.5)$	reduces soil brightness effects from vegetation indices	(Huete, 1988)
SIPI	Structure insensitive pigment index	$(B8 - B2) / (B8 + B2)$	Pigment ratio between carotenoid and chlorophyll a. Canopy stress and LAI	(Yu et al., 2018)
TCARI	Transformed chlorophyll absorption and reflectance index	$3 * ((B5 - B4) - 0.2 * (B5 - B3)) * (B5 / B4)$	Chlorophyll, LAI	(Wu et al., 2008)
TVI	Triangular Vegetation Index	$0.5 * (120 * (B6 - B3) - 200 * (B4 - B3))$	Estimates green LAI Sensitive to increase in chlorophyll as canopy density increases.	(Qian et al., 2022)
WDRVI	Wide Dynamic Range Vegetation Index	$(0.2 * B8 - B4) / (0.2 * B8 + B4)$	Positive relation with vegetation fraction sensitivity to change in LAI	(A. A. Gitelson, 2004)

### 3.5. Rainfall

The Climate Hazard Group Infrared Precipitation with Station Data (Version 2.0 final) was accessed from Google Earth Engine. Monthly sum composites were generated for the entire growing season, including one month prior. This data has a resolution of approximately 5.6 km. CHIRPS combines satellite-measured precipitation with ground station data, resulting in low systematic bias. The decision to use CHIRPS rainfall data is supported by previous research demonstrating its effectiveness for hydrological forecasting and trend analysis in Ethiopia (Funk et al., 2015). The data was retrieved in tabular format.



### **3.6. ERA5-Land Data and Processing**

The ERA-5 land dataset, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) and the Copernicus Climate Change Service (C3S), provides a comprehensive climate reanalysis dataset. It provides hourly predictions for atmospheric, terrestrial, and marine climate variables, encompassing rainfall and temperature measurements, on a global scale. The data covers the Earth on an 11 km grid, with included information about uncertainties for all variables at reduced spatial and temporal resolutions (Muñoz-Sabater et al., 2021). The ERA-5 land dataset is available in google Earth Engine. The following datasets were used from ERA5 land: Temperature\_2m, soil temperature, volumetric soil water, evaporation from bare soil, and total evaporation.

#### **3.6.1. Temperature\_2m**

The 2-meter air temperature parameter represents the measured temperature of the atmosphere at a height of 2 meters from the land surface, oceans, or inland water systems. The measurement involves an interpolation process between the Earth's surface and the lowest model level, considering the atmospheric conditions through physical parameterization. The unit of measurement for this parameter is Kelvin (K). In Google Earth Engine, the data was extracted from the monthly aggregate asset of the hourly data. Temperature at 2 m was created by averaging all hourly data in a particular month over the growing season and a month prior. The data was extracted in tabular format from the field location points.

#### **3.6.2. Soil temperature**

This parameter measures the temperature within six layers of soil. The soil surface is defined at 0 cm. The soil temperature is measured at the central point of each layer, with calculations of heat transfer occurring at the boundaries between layers. In this study, soil temperature data was averaged for the first two layers (0 - 28 cm).

#### **3.6.3. Volumetric Soil Water**

The volumetric soil water content for the initial two layers (0 - 28 cm) was averaged in this study. This represents the volume of water present in the topsoil, influenced by factors such as texture and depth of soil, and the groundwater level. The surface is defined at 0 cm.

#### **3.6.4. Evaporation from Bare Soil**

This parameter quantifies the amount of water that evaporates from the bare soil surface. It represents the water loss occurring at the top of the land surface. Monitoring evaporation from bare soil is essential for understanding soil moisture dynamics and managing water resources in agricultural and natural ecosystems.

#### **3.6.5. Total Evaporation**

Total evaporation refers to the accumulated water that has transitioned into water vapor from Earth's surface plus the total transpiration of vegetation. This metric is expressed in meters of water above the surface. Downward fluxes are considered positive, so negative values relate to evaporation, while positive values represent condensation. Total evaporation is a critical parameter for assessing the water cycle, climate interactions, and hydrological processes.

### **3.7. Soil Data and processing**

The soil data for this study was obtained from the ISRIC - World Soil Information portal, which provides a comprehensive database of global soil information. This platform offers various soil properties at multiple spatial resolutions, which are crucial for agricultural and environmental research.

The primary source of soil data from ISRIC is the SoilGrids dataset, which provides global predictions for standard soil properties at six standard depths (0-5 cm, 5-15 cm, 15-30 cm, 30-60 cm, 60-100 cm, and 100-200 cm). These physical and chemical properties include soil organic carbon content, pH, texture fractions (sand, silt, and clay), bulk density, extractable nitrogen, extractable iron, and others. The data is generated through machine learning techniques applied to globally harmonized soil profile observations and environmental covariates, resulting in spatial predictions at 250-meter resolution (Hengl, De Jesus, et al., 2017). The soil data was downloaded from the ISRIC portal. For each soil property, values corresponding to the average crop root depth were extracted. This was accomplished using ArcGIS Pro, where the data was processed to obtain mean values across the root depth for each crop type. This integration of soil data provides a more accurate representation of the soil environment encountered by crop roots, which is critical for assessing nutrient availability and uptake (Rickson, 2023). The data was extracted using field location points.

### 3.8. Digital Elevation Model and processing

The Multi-Error-Removed Improved-Terrain (MERIT) Digital Elevation Model (DEM) is a product of the elimination of numerous error components from pre-existing spaceborne DEMs, inclusive of the Shuttle Radar Topography Mission 3-arcsecond (SRTM3) v2.1 and the Advanced Land Observing Satellite (ALOS) World 3D - 30m (AWD-30m) v1 (Yamazaki et al., 2017). This high-accuracy global DEM eliminates major error components from the SRTM3 and AWD-30m datasets, enhancing terrain representation and increasing vertical accuracy.

Topographic derivatives from the elevation band of the MERIT DEM include aspect and slope in degrees by calculating local gradient using the 4-connected neighbours of each pixel. These are described as primary derivatives since they are calculated directly from the digital elevation model. The primary derivatives are calculated using the Google Earth Engine “ee.Terrain.products” function. Secondary derivatives depend on a combination of primary surface derivatives or developed indices. The topographic wetness index used in this study is an example. The Topographic Wetness Index (TWI) quantifies water accumulation in areas with varying elevations. It is determined by both the slope of the terrain and the contributing area upstream (Schmidt & Persson, 2003).

$$TWI = \ln(a/\tan b) \quad \text{equation (1)}$$

where: a = upslope contributing area (m<sup>2</sup>)

B = slope in radians

These derivatives will be used to capture topographic variations and identify potential areas susceptible to specific agricultural challenges (Mukul et al., 2017). All products derived from the DEM were calculated in Google Earth Engine and extracted in a tabular format using the field location points.

### 3.9. Moderate Resolution Imaging Spectroradiometer (MODIS) product

The MODIS product used in this study provides essential data on the FPAR and LAI, which are both important for assessing vegetation health and productivity (Yan et al., 2016).

FPAR is a dimensionless metric indicating the fraction of incoming solar radiation within the photosynthetically active wavelength range (400-700 nanometers) absorbed by plant leaves. It is crucial for assessing photosynthetic efficiency and overall plant health by reflecting the canopy's light-capturing capability. LAI quantifies the total one-sided leaf area per unit ground area for a plant canopy, providing

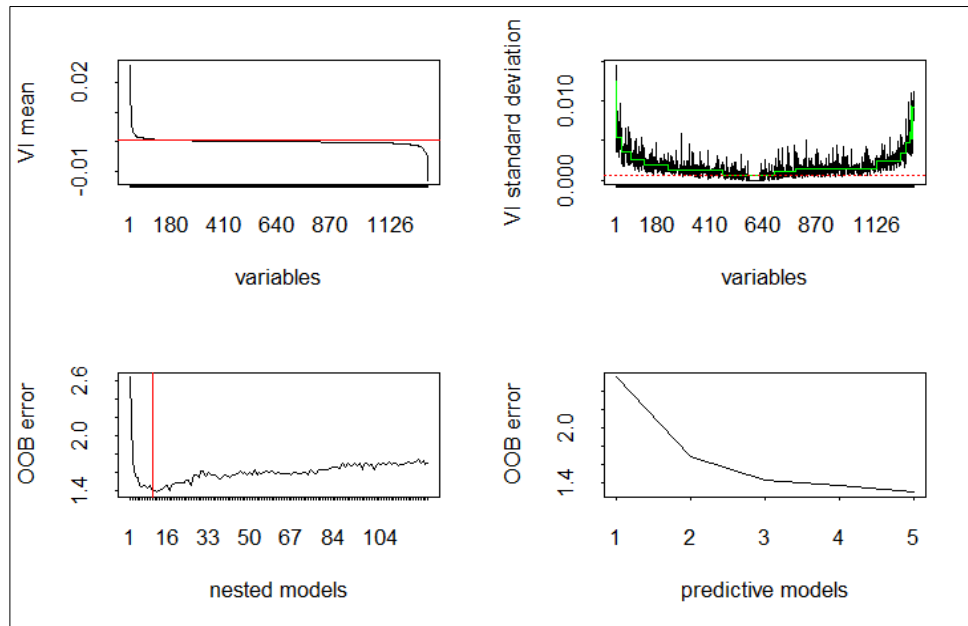
essential insights into the vegetation's capacity for photosynthesis, water regulation, and interaction with atmospheric and soil processes (Benjamin, 2017). This index is vital for understanding canopy density and light interception efficiency, which are critical for plant growth and development. The MODIS algorithm produces 8-day composite datasets at a 500-meter resolution by selecting the best pixel from multiple Terra sensor acquisitions, ensuring data accuracy and reliability.

To generate monthly composites, the sum of MODIS product was computed for each field location point, as the study focuses on quantifying the cumulative amount of FPAR absorbed by the crops within a given month. The data was exported in tabular format using field location points.

### **3.10. Random Forest Models and Optimization**

#### **3.10.1. Variable Selection using Random Forests**

In this study, managing a dataset with over five hundred potential predictor variables posed a significant challenge due to the Hughes phenomenon, also known as the curse of dimensionality (Alonso et al., 2011). This phenomenon states that the predictive power of a model decreases as the number of predictor variables increases beyond a certain point, especially when the sample size is limited. To address this, Variable Selection Using Random Forests (VSURF) developed by Genuer et al., (2010) was implemented as a preliminary step. This approach was instrumental in reducing the dimensionality of the dataset by eliminating irrelevant features, thereby mitigating the effects of the Hughes phenomenon. Consequently, it simplified the model-building process and enhanced the interpretability of subsequent analyses by focusing on the most influential variables. This strategic decision was crucial for balancing the depth of analysis with practical constraints, including computational resources and the risk of overfitting.



### Insights of VSURF plots

In Figure 4, the top left graph

Figure 4. VSURF Result plots. Refer to section “Insights of VSURF plots” for explanation illustrates the average variable importance (VI) in a descending sequence, denoted by a black curve in the thresholding phase. A red horizontal line demarcates the threshold beyond which variables are considered significant. This step eliminates variables of lesser importance, thereby simplifying the dataset. The top right graph presents the standard deviation of VI values, arranged according to their mean VI. A green line symbolizes the predictions of the Classification and Regression Trees (CART) model, and a dotted red line indicates the threshold based on the minimum CART prediction value. This procedure ensures examines carefully the variables with high variability in their importance scores. The bottom left graph describes the interpretation phase, plotting the average out-of-bag (OOB) error rate for nested RF models, ranging from a one variable to all retained variables. A vertical red line pinpoints the model with the minimum OOB error, thereby identifying the optimal subset of variables for the model. The bottom right graph corresponds to the prediction phase, depicting the average OOB error rate for models with variables added in a sequential manner. The final model with the lowest OOB error, ensures the selection of the most predictive set of variables (Genuer et al., 2010).

Table 5. Distribution of field crop nutrient measurements per regions. These measurements have been used to train and test the accuracy of the implemented predictive models

Crop				Region								
				Oromiya			Amhara			Tigray		
	Total	Train	Test	Total	Train	Test	Total	Train	Test	Total	Train	Test
Wheat	210	147	63	152	106	46	37	36	11	21	15	6
Maize	218	152	66	189	132	57	16	11	5	13	9	4
Teff	226	159	67	168	118	50	31	22	9	26	18	8
Sorghum	108	76	32	64	45	19	40	28	12	4	3	1
Barley	130	91	39	103	72	31	14	10	4	13	9	4

### 3.10.2. Hyperparameter Tuning and Random Forest Model

After the variable selection phase, the data was split into training and testing subsets, with 70% of the data from each administrative region designated for training the models and the remaining 30% from each administrative region reserved for independent validation (see Table 5). This allocation ensured that the models developed would demonstrate robustness against unseen data, thereby improving the generalizability of the findings. Following this preparation, a systematic approach to hyperparameter tuning was incorporated, beginning with Grid Search Cross-Validation (GridSearchCV). GridSearchCV systematically explored combinations of hyperparameters across a predefined range, employing cross-validation to evaluate each configuration and identify the optimal set (Malakouti et al., 2023). This method facilitated the fine-tuning of the RF parameters to better capture the complex patterns in the data.

Upon completion of the hyperparameter tuning processes, which encompassed both Grid Search Cross-Validation (GridSearchCV) and manual adjustments, the model was developed with the selected potential predictors identified during the VSURF phase. This initial model run served as a foundation, providing a baseline for assessing the predictive capabilities of the selected features. Following the execution of the model, the importance of individual predictors was critically evaluated using ELI5 Permutation Importance. This method, as outlined by Korobov & Lopuhin, 2017 in the ELI5 package documentation, measures how the model's score decreases when a feature is permuted, offering a quantitative measure of feature importance.

The ELI5 Permutation Importance analysis offered a comprehensive view of the contribution of each feature towards the model's predictions, enabling the identification and removal of less impactful predictors. This step was crucial for refining the model by focusing solely on the most critical features, thereby enhancing its predictive accuracy. The subsequent model reiteration, therefore, concentrated on the subset of features deemed most important by the ELI5 analysis. This refined approach not only improved the model's overall performance but also contributed to a more efficient and interpretable final model, which aligns with the goal of developing a robust and understandable predictive model.

Furthermore, the integration of ELI5 Permutation Importance into the model development process emphasized the importance of explainability in machine learning. By quantitatively assessing the influence of each feature on the model's predictions, this method provided insights into the dataset, aiding in debugging, and enhancing the model's accuracy.

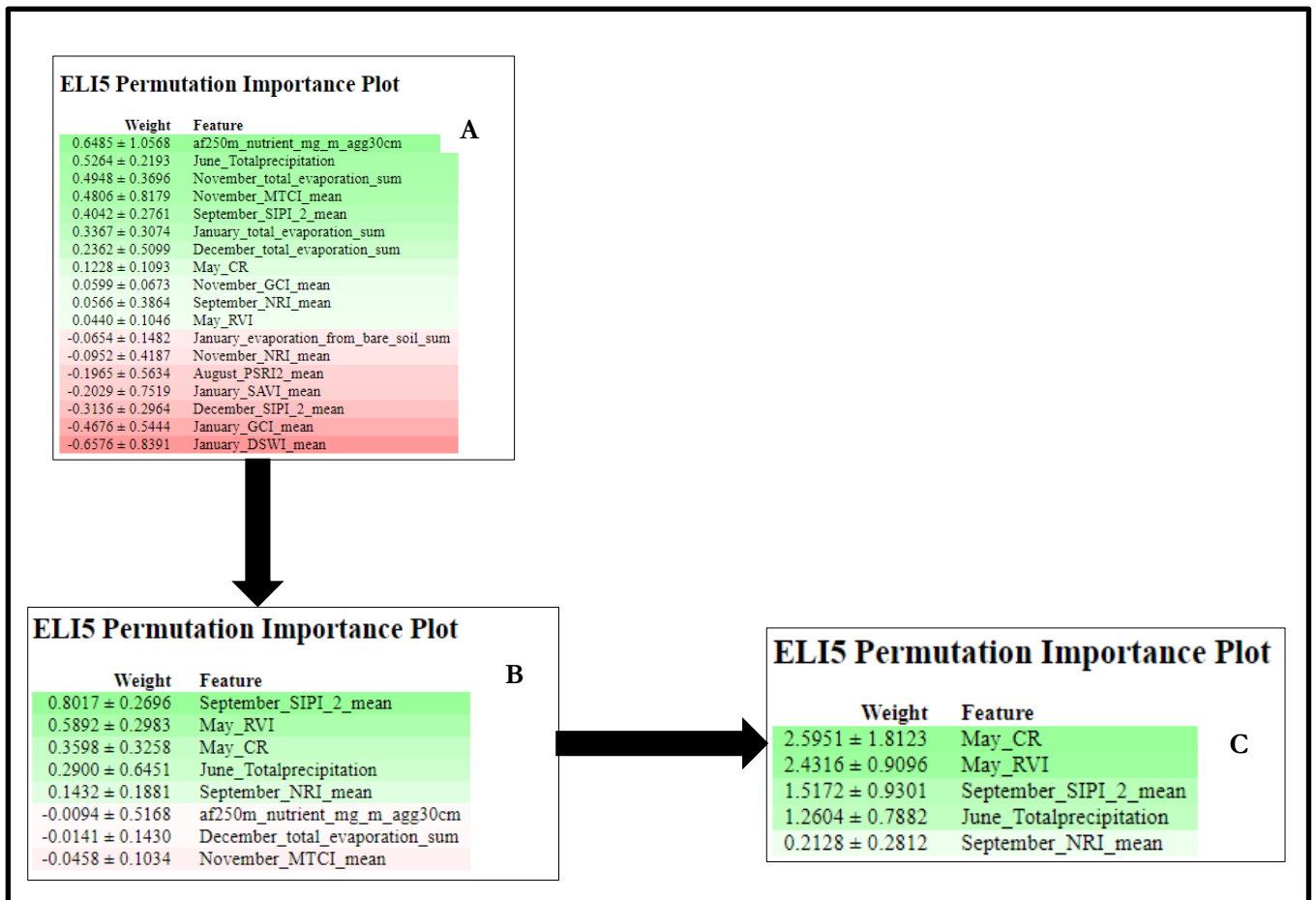


Figure 5. ELI5 permutation importance. A: Permutation importance of variables from VSURF used in a model; B: Permutation importance of variables after manually discarding variables highlighted in red; C: Variables used in the final model after discarding variables in B.

The visual representation in the Figure 5 shows a gradient of feature importance, delineated by A spectrum of colors ranging from vibrant green to faint green, transitioning into pink, and ultimately to red. This color gradient serves as a visual cue, indicating the relative significance of the features within the dataset. Notably, features highlighted in vibrant green are to hold the highest importance, gradually diminishing in significance towards those highlighted in red, which are deemed less important within the analytical context.

### 3.10.3. Assessing Model Performance

To assess the performance of the models, three statistical metrics were employed, including the coefficient of determination ( $R^2$ ), Root Mean Square Error (RMSE), and normalized Root Mean Square Error (nRMSE). These metrics provide a comprehensive evaluation of the model's accuracy and predictive capabilities and are widely recognized and used in the field of predictive modelling. The simultaneous use of these three metrics ensures a thorough examination of model performance.

The coefficient of determination, denoted as  $R^2$ , quantifies the fraction of the variance in the target variable that is explained by the explanatory variables (Draper & Smith, 2014). It is defined as equation 2:

$$R^2 = 1 - [\Sigma (y_i - \hat{y}_i)^2 / \Sigma (y_i - \bar{y})^2] \quad \text{(equation 2)}$$

where:

- $y_i$  represents the observed values,
- $\hat{y}_i$  denotes the predicted values,
- $\bar{y}$  represents the mean of the observed values.

$R^2$  values range from 0 to 1. An  $R^2$  value close to one indicates that a large proportion of the variance in the response variable is explained by the model, signifying high predictive accuracy. Conversely, an  $R^2$  value closer to zero suggests poor model performance. The coefficient of determination ( $R^2$ ) is a dimensionless metric.

The Root Mean Square Error (RMSE) quantifies the average magnitude of the prediction errors, providing an intuitive measure of model accuracy. It is calculated as equation 3:

$$RMSE = \sqrt{[\Sigma (y_i - \hat{y}_i)^2 / n]} \quad \text{(equation 3)}$$

where:

- $y_i$  is the observed values,
- $\hat{y}_i$  represents the predicted values,
- $n$  is the number of observations.

A smaller value of RMSE indicates better model performance, as it signifies less deviations between the observed and predicted values. A high RMSE might indicate that the model is making large errors on some predictions, suggesting a need for further tuning or additional features. The RMSE is typically measured in the same units as the dependent variable (Hyndman & Koehler, 2006).

The normalized Root Mean Square Error (nRMSE) is an extension of RMSE that accounts for the scale of the data, making it useful for comparing models across different datasets or scales (Stephen & Kazemi, 2014). It is defined as equation 4

$$nRMSE = RMSE / \bar{y} \quad \text{(equation 4)}$$

where  $\bar{y}$  is the mean of the observed data. This dimensionless metric facilitates direct comparisons, and a lower nRMSE value reflects higher accuracy and reliability of the model.

By utilizing these statistical metrics, the study ensured a rigorous assessment of model performance. The combination of  $R^2$ , RMSE, and nRMSE provided a robust framework for evaluating the predictive power

and accuracy of the models, guiding the refinement and optimization process to achieve reliable and interpretable results. These measures, being commonly used, allow for the comparison of model performance across different studies or research. This comparability is crucial for benchmarking and choosing the best model for a given task.

#### **3.10.4. Model Interpretation**

After the predictive models with the highest possible accuracies have been developed, PDPs which serve as a valuable tool interpreting machine learning model will be used to derive insights. PDPs provide a graphical representation of how individual features influence a target variable within RF models. By visualizing the relationship between predictors and the target nutrients, PDPs offer insights into positive or negative linear and nonlinear relations, facilitating a comprehensive evaluation of the model's relevant predictors (Friedman, 2001). A significant limitation of PDPs lies in their assumption that each predictor operates independently of others. This means that while analyzing the impact of a single feature on the target nutrient, PDPs ignore possible feature interactions. This approach simplifies the analysis but may potentially lead to incomplete interpretations of the model's behaviour.



## 4. RESULTS

### 4.1. Descriptive statistics of Crop Grain Nutrient composition

The descriptive statistics of nutrient composition in various crops, measured in  $\text{mgkg}^{-1}$ , reveal significant differences in nutrient densities based on median values (see Figure 6 to Figure 14). Teff consistently exhibits the highest median values across multiple nutrients: Ca ( $1530.95 \text{ mgkg}^{-1}$ ), Cu ( $6.88 \text{ mgkg}^{-1}$ ), Fe ( $146.31 \text{ mgkg}^{-1}$ ), Mg ( $1961.87 \text{ mgkg}^{-1}$ ), Mn ( $79.41 \text{ mgkg}^{-1}$ ), P ( $3937.78 \text{ mgkg}^{-1}$ ), S ( $2311.17 \text{ mgkg}^{-1}$ ), and Zn ( $30.55 \text{ mgkg}^{-1}$ ). Barley also shows high median values for several nutrients, including Ca ( $405.64 \text{ mgkg}^{-1}$ ), K ( $5172.37 \text{ mgkg}^{-1}$ ), Mg ( $1272.09 \text{ mgkg}^{-1}$ ), P ( $3602.89 \text{ mgkg}^{-1}$ ), S ( $1846.68 \text{ mgkg}^{-1}$ ), and Zn ( $29.27 \text{ mgkg}^{-1}$ ). Wheat demonstrates considerable nutrient densities with notable median values for Ca ( $436.40 \text{ mgkg}^{-1}$ ), K ( $4680.45 \text{ mgkg}^{-1}$ ), P ( $3252.93 \text{ mgkg}^{-1}$ ), and S ( $1814.81 \text{ mgkg}^{-1}$ ). In contrast, Maize has the lowest median values for most nutrients: Ca ( $68.06 \text{ mgkg}^{-1}$ ), Cu ( $1.62 \text{ mgkg}^{-1}$ ), Fe ( $22.75 \text{ mgkg}^{-1}$ ), Mg ( $1023.36 \text{ mgkg}^{-1}$ ), Mn ( $4.85 \text{ mgkg}^{-1}$ ), P ( $2684.34 \text{ mgkg}^{-1}$ ), S ( $1579.83 \text{ mgkg}^{-1}$ ), and Zn ( $19.67 \text{ mgkg}^{-1}$ ). Sorghum generally falls in the mid-range for most nutrients, with notable median values for Ca ( $183.18 \text{ mgkg}^{-1}$ ), K ( $4538.27 \text{ mgkg}^{-1}$ ), Mg ( $1485.87 \text{ mgkg}^{-1}$ ), and P ( $3210.56 \text{ mgkg}^{-1}$ ). These statistics underscore the nutritional prominence of Teff, Barley, and Wheat compared to Maize, which shows consistently lower nutrient concentrations.

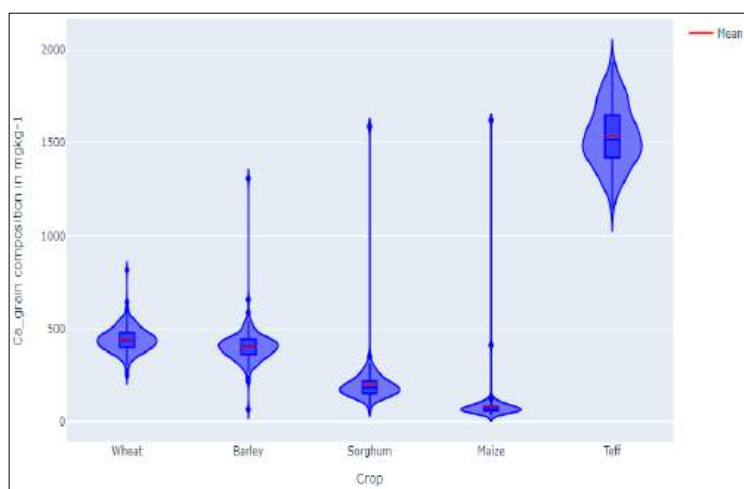


Figure 6. Violin plots showing Ca grain nutrient composition

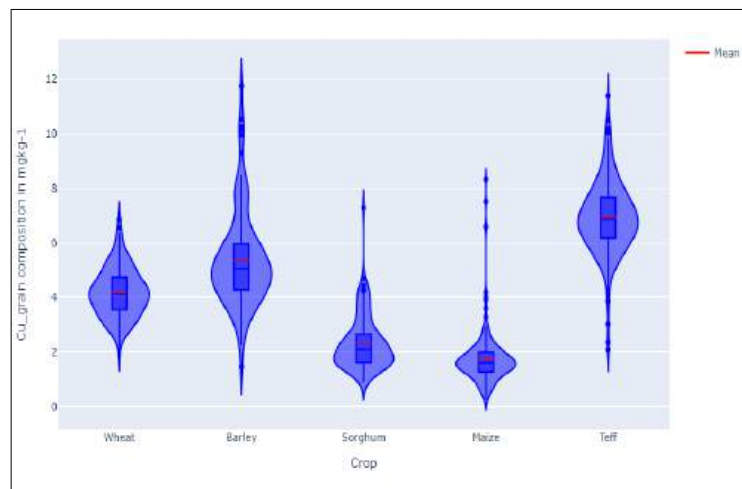


Figure 7. Violin plots showing Cu grain nutrient composition

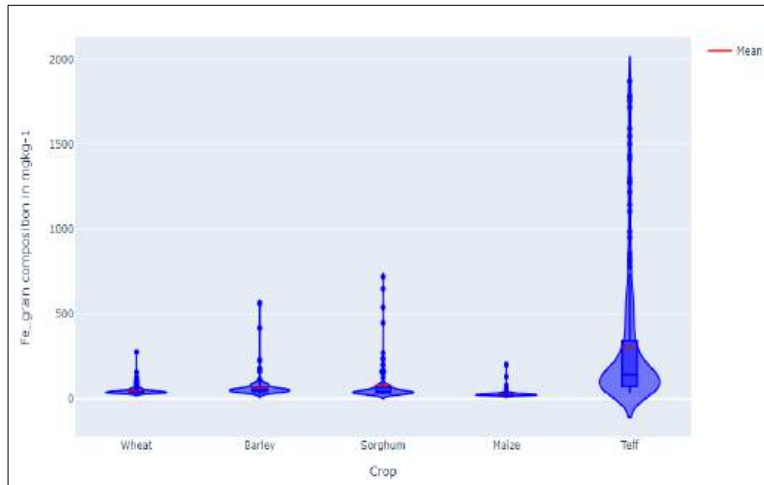


Figure 8. Violin plots showing Fe grain nutrient composition

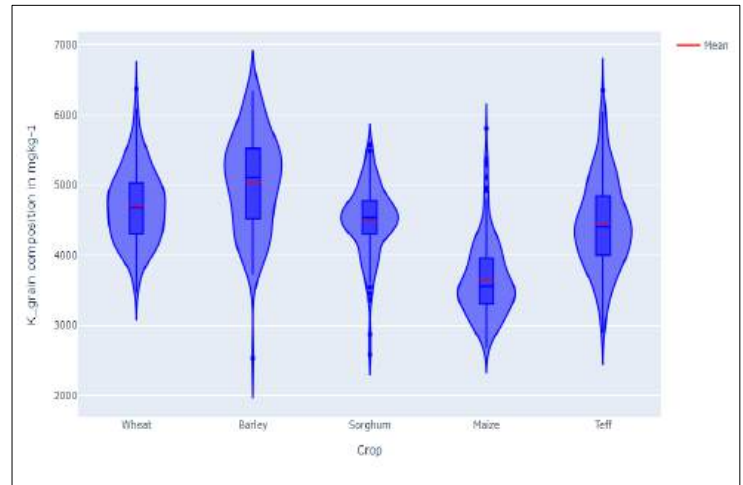


Figure 9. Violin plots showing K grain nutrient composition

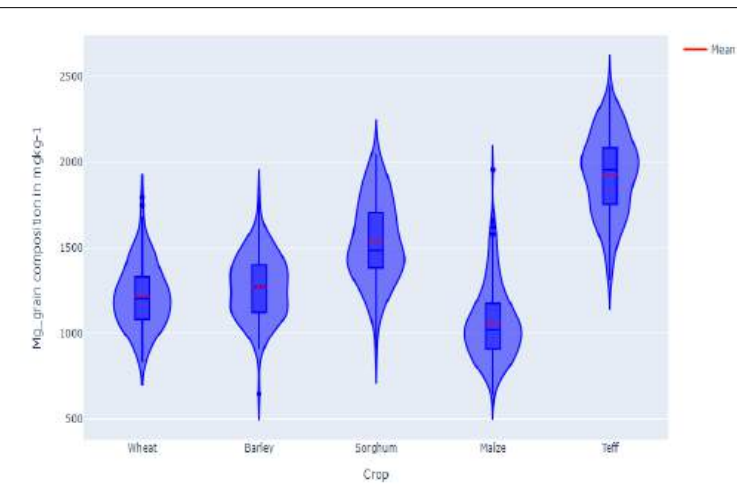


Figure 11. Violin plots showing Mg grain nutrient composition

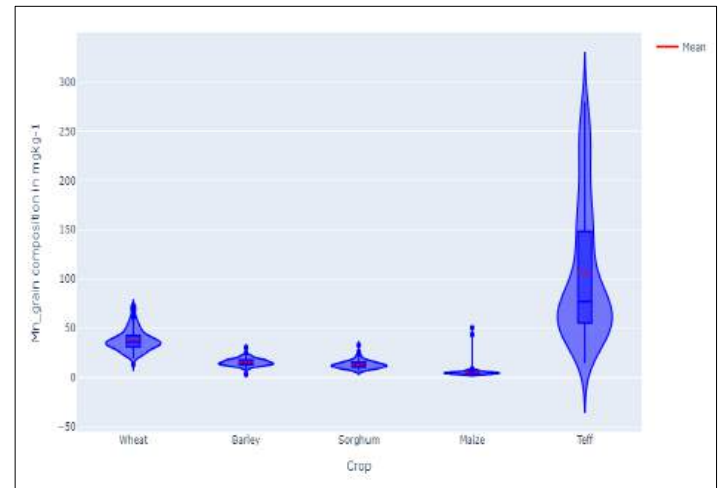


Figure 12. Violin plots showing Mn grain nutrient composition

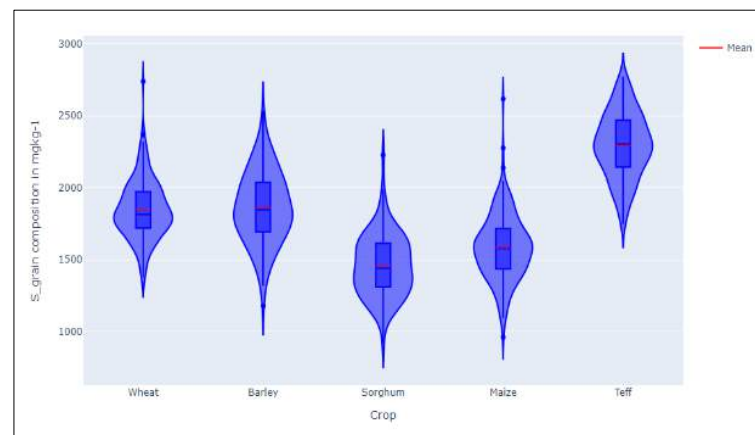


Figure 10. Violin plots showing S grain nutrient composition

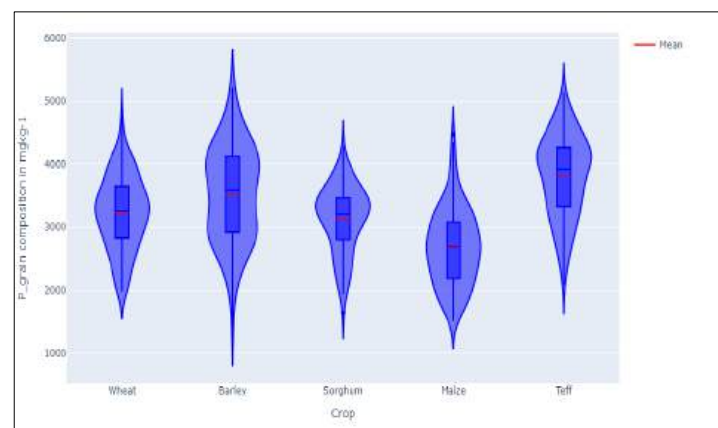


Figure 13. Violin plots showing P grain nutrient composition

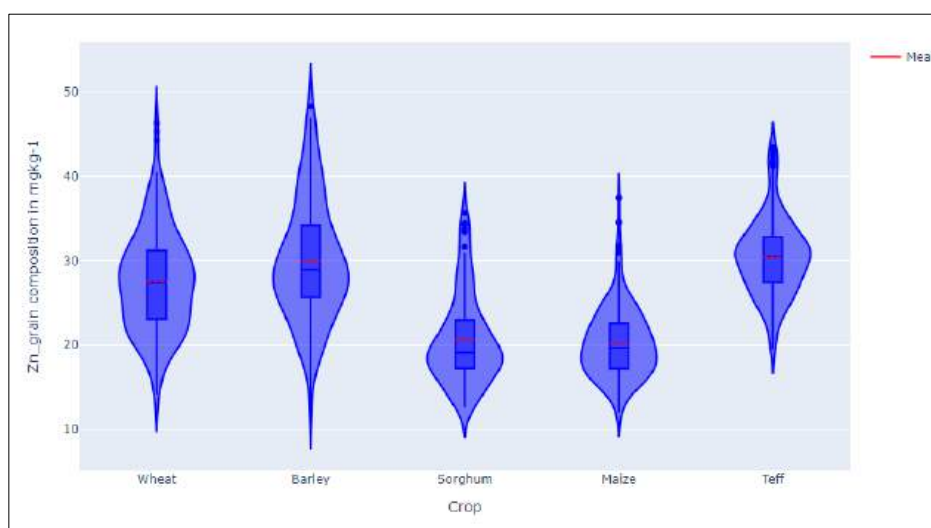


Figure 14. Violin plots showing Zn grain nutrient composition

## 4.2. Findings

Due to the large number of investigated nutrients (9) and crops (5) and total of 45 implemented predictive models, the study will focus on discussing in detail the results obtained for teff. In Ethiopia Teff is one of the most important staples and holds significant cultural and economic importance in the country, being a staple crop deeply ingrained in the country's culinary traditions and agricultural practices (Tadele & Hibistu, 2021). As one of the oldest domesticated grains globally, teff serves as a dietary staple for millions of Ethiopians, particularly in the form of Injera, a traditional sourdough flatbread. Moreover, teff plays a crucial role in Ethiopia's economy, contributing significantly to agricultural production and livelihoods across the country. Its resilience to adverse growing conditions, such as drought and poor soil fertility, makes it particularly valuable in regions where agricultural resources are limited. Additionally, teff's nutritional profile, rich in essential nutrients like Fe and Ca, further enhances its importance as a key component of the Ethiopian diet, particularly in combating malnutrition and addressing food security challenges (Kuyu et al., 2024).

Findings for the remaining crops will be presented with less detail, maintaining a broader perspective while highlighting significant trends and patterns in the discussion section. These statistical accuracies will also be presented in a tabular format at the end of this section. This approach ensures that the wealth of results generated from this study is shared, while maintaining the readability and coherence of the study.

While the discussion in the text focuses in detail on teff model in this section, it is essential to note that the results for all crops are equally significant and have been thoroughly analyzed. The figures of these results are in Appendix 2. Results for barley to Appendix 6. Results for wheat and can be referred to for an overview of the findings.

The results were interpreted in the context below:

May is the month before Teff is planted, with planting is occurring in June and July. August and September represent the mid-season, and October marks the end of the mid-season. December to February marks the end of the harvest season.

After the interpretation of results relating to Teff and each of the nutrients, a graphical representation of results will follow.

### **4.3. Micronutrient Analysis for Teff**

The micronutrient analysis for Teff includes an examination of Ca, Cu, Fe, Mn, Mg and Zn. These micronutrients are crucial for various bodily functions and maintaining overall health. These micronutrient deficiencies are widespread in the study area (Belay et al., 2022).

#### **4.3.1. Interpretation of Results for Teff grain - Ca**

The important features identified through the model include the amount of Ca nutrient in the soil, the Structure Insensitive Pigment Index (SIPI) for July, and the bulk density of the soil. The Ca content in Teff is significantly influenced by the concentration of Ca in the soil. The predictive model had an R-squared of 0.12 and RMSE of 741 mgkg<sup>-1</sup>.

From the relation in the partial dependence plot (see Figure 15C), generally, higher levels of Ca in the soil generally lead to higher Ca content in Teff. This indicates that the amount of soil Ca content is important in determining the nutritional composition of Teff.

SIPI serves as an indicator of plant health and stress levels (Tayade et al., 2022). The Partial dependence relation (see Figure 15D) generally shows that a higher SIPI value, indicative of better plant health and less stress at the start of the season, correlates with higher Ca content in Teff grains. This highlights the importance of environmental factors, particularly those affecting plant health, in influencing the nutritional composition of Teff.

Soil bulk density reflects soil compaction and porosity, which in turn affects root growth conditions and nutrient uptake (Indoria et al., 2020). Lower bulk density generally indicates better root growth conditions and potentially higher nutrient uptake, including Ca. The partial dependence relationship shows this (see Appendix 5. Results for **teff**). However, the impact of bulk density on Ca content in Teff appears to be less consistent compared to other factors, suggesting that it may be influenced by other interacting variables.

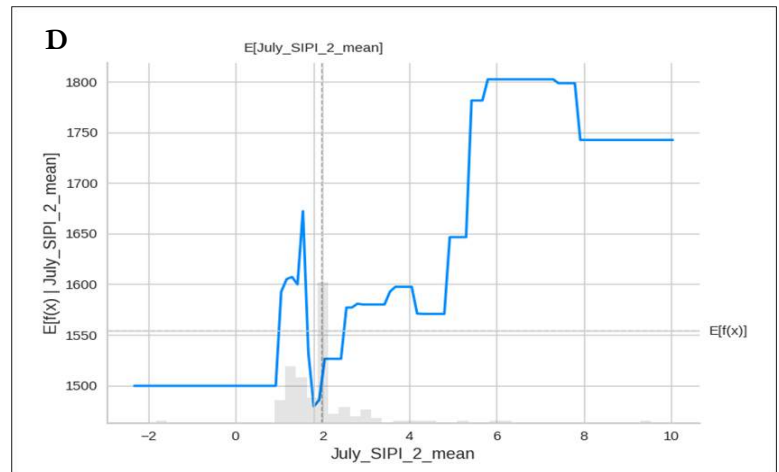
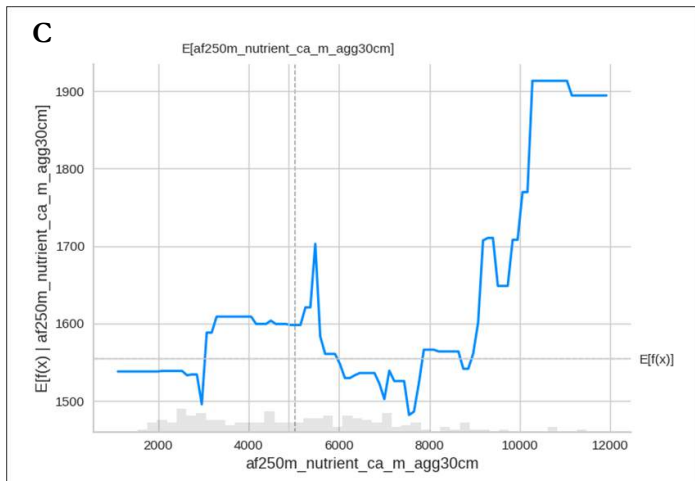
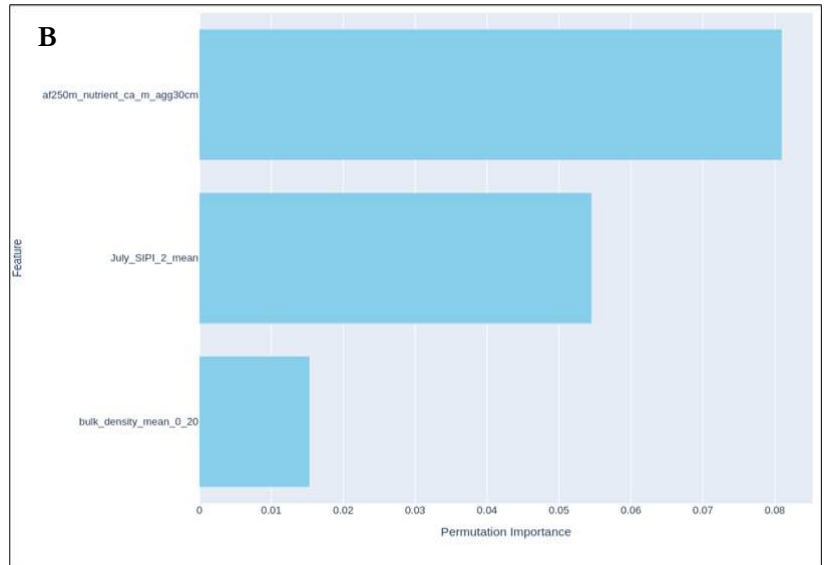
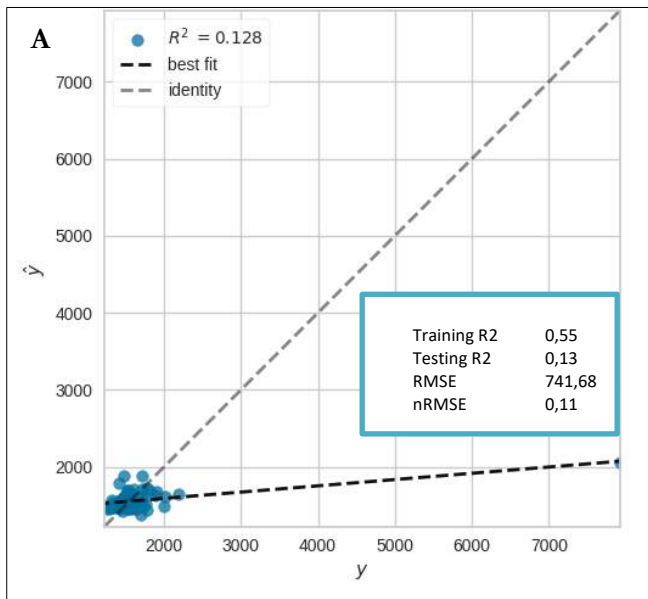


Figure 15. Graphical representation of results for Teff grain and Ca nutrient composition. Image A represents scatter plots of observed versus predicted values, offering a visual comparison between actual measurements and model predictions. Image B is the permutation importance plot, illustrating the significance of the predictors used in the final model. Images C and D are the partial dependence plots for the two most influential features, highlighting their effect on the grain nutrient composition.

#### 4.3.2. Interpretation of Results for Teff grain - Cu

The predictive model had an R-squared of 0.30 and RMSE of 1 mgkg<sup>-1</sup>.

The evaluation of vegetation indices in May, before the planting season, may seem unusual, yet they could provide valuable information about pre-existing soil conditions and residual vegetation cover. The FPAR during this period can have a lasting impact on the soil and environmental conditions. If there were any crops planted in the field (either cover crops or weeds), they could either improve the amount of nutrient in the soil or compete with the planted crops for nutrients (Scavo et al., 2022). Adequate sunlight exposure before planting can enhance soil temperature and moisture retention, creating favourable conditions for seed germination and early root development. These pre-planting conditions are crucial as they can influence the availability and uptake of Cu once the Teff begins to grow (Doane et al., 2019). Improved early soil conditions can lead to better root systems, which in turn can enhance the plant's ability to absorb Cu from the soil (Ngo & Cavagnaro, 2018). From the Partial dependence relationship (see Figure 16C), it is noted that there is a mixed impact of pre-season FPAR on Cu nutrient in Teff, which would need extra context to be interpreted.

The Anthocyanin Reflectance Index (ARI) in July measures the presence of anthocyanins, which are indicators of plant stress responses. During July, Teff plants are typically in a critical growth phase where they can experience various environmental stresses, such as high temperatures or drought conditions. Higher ARI values suggest that plants may be undergoing stress, which can affect physiological processes, including nutrient uptake (A. A. Gitelson et al., 2009). Stress conditions can alter the plant's demand and uptake for micronutrients like Cu, influencing its concentration in the plant tissues. From the Partial dependence plots (Figure 16D), the high ARI values in the early crop stage growth led to a reduction in Ca in the grain at the end of the season.

The SIPI in July is an indicator of plant health and photosynthetic activity. While it has a lower influence compared to the other features, SIPI provides valuable insights into the overall health of the Teff plants. Healthier plants with efficient photosynthetic processes are more likely to absorb and utilize micronutrients effectively, including Cu. The partial dependence relation shows that when the plant has high values of SIPI (refer to Appendix 5. Results for **teff**), Ca nutrient is high in the grain. This shows that efficient photosynthesis at the growth stage of Teff is important for Ca nutrient uptake. Monitoring SIPI values can help in understanding the plant's nutritional status and potential deficiencies.

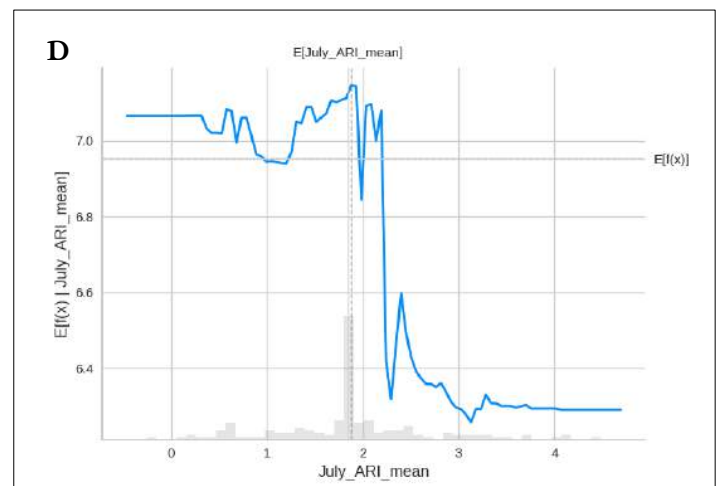
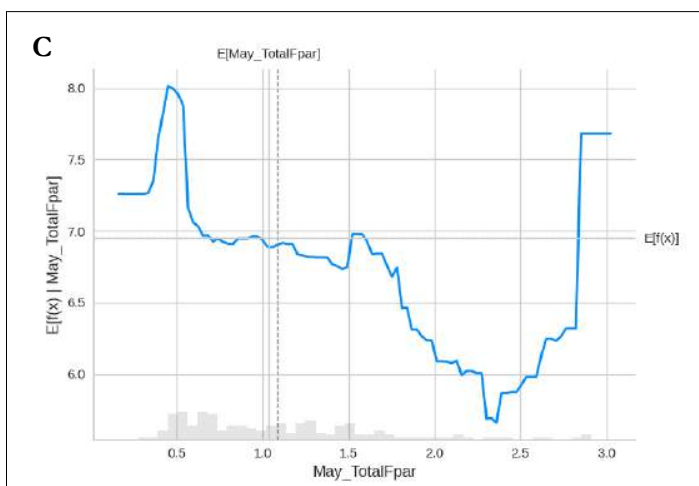
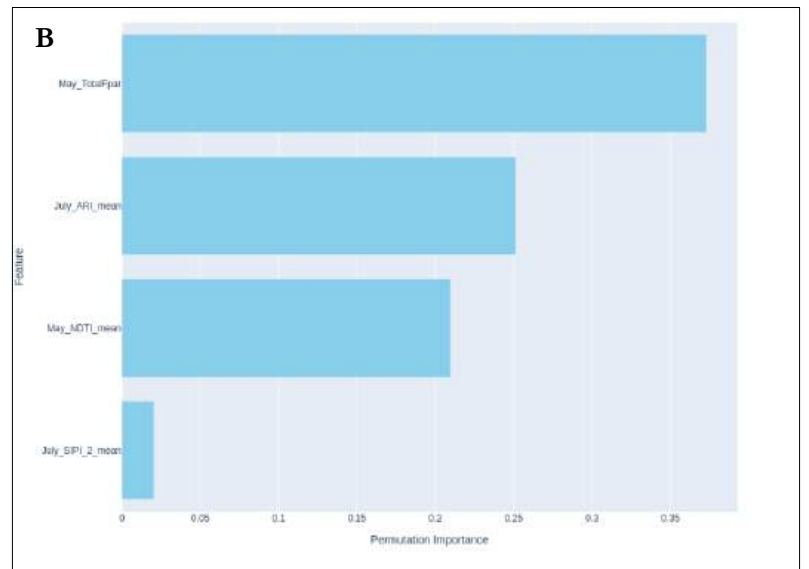
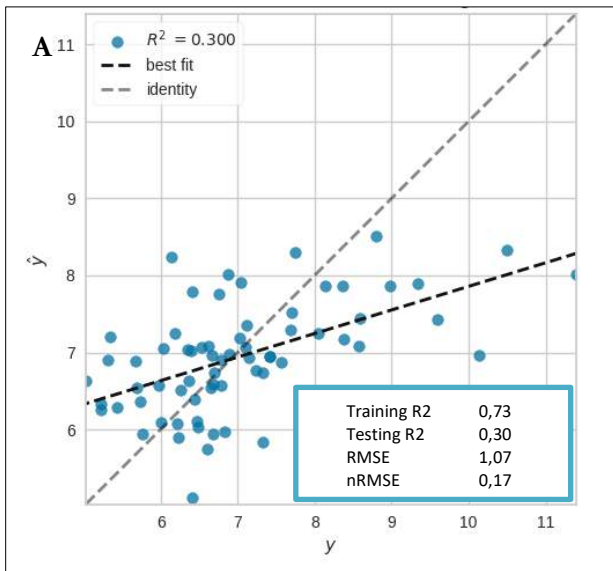


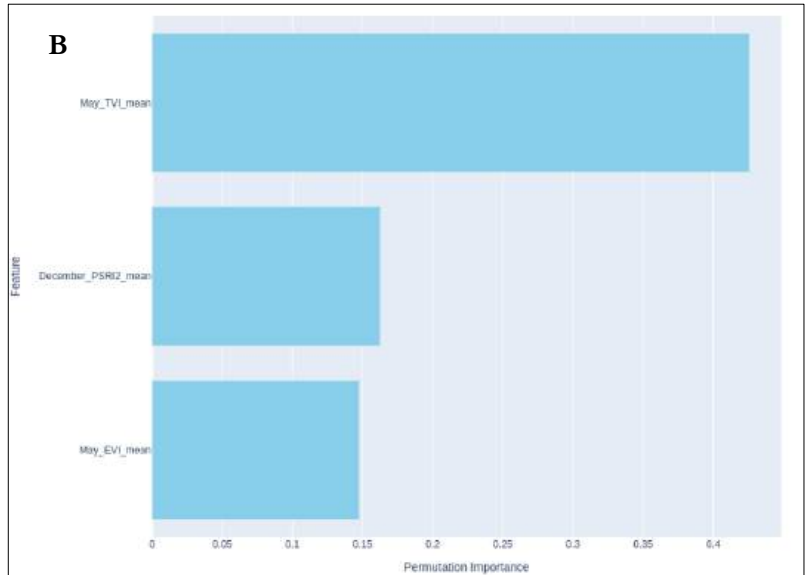
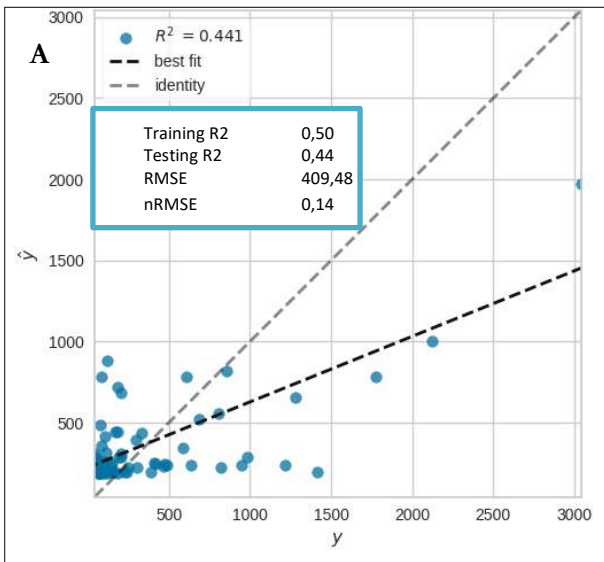
Figure 16. Graphical representation of results for Teff grain and Cu nutrient composition. Image A represents scatter plots of observed versus predicted values, offering a visual comparison between actual measurements and model predictions. Image B is the permutation importance plot, illustrating the significance of the predictors used in the final model. Images C and D are the partial dependence plots for the two most influential features, highlighting their effect on the grain nutrient composition.

### 4.3.3. Interpretation of Results for Teff - Fe

The predictive model had an R-squared of 0.44 and RMSE of 409.5 mgkg<sup>-1</sup>.

Low TVI and EVI values in May indicate less vigorous vegetation in the field before teff is planted. This could mean that the field is relatively free of other plants that could compete with teff for nutrients, including Fe once it is planted. Therefore, when teff is planted a month later, it might have access to a larger pool of available nutrients, leading to higher Ca content in the grain. High TVI and EVI values in May could indicate vegetation in the field before teff is planted. This could mean that the field is densely populated with other plants that could compete with teff for nutrients once it is planted. Therefore, when teff is planted a month later, it might face competition for available nutrients, leading to reduced Fe uptake in the grain. The plot (See Figure 17C) shows that when EVI increases in the pre-season, Fe in the grain is low

The PSRI reflects the level of plant senescence or aging just before harvest. Higher values indicate canopy stress, possibly due to factors like disease or environmental conditions, or senescence in vegetation leading to potentially reduced Fe uptake or translocation (Merzlyak et al., 1999). This measurement of PSRI at the end of the crop growth cycle, shows that increased crop senescence at the end of the growing season leads to a reduction of Fe nutrient concentration in the grain (refer to Figure 17D).





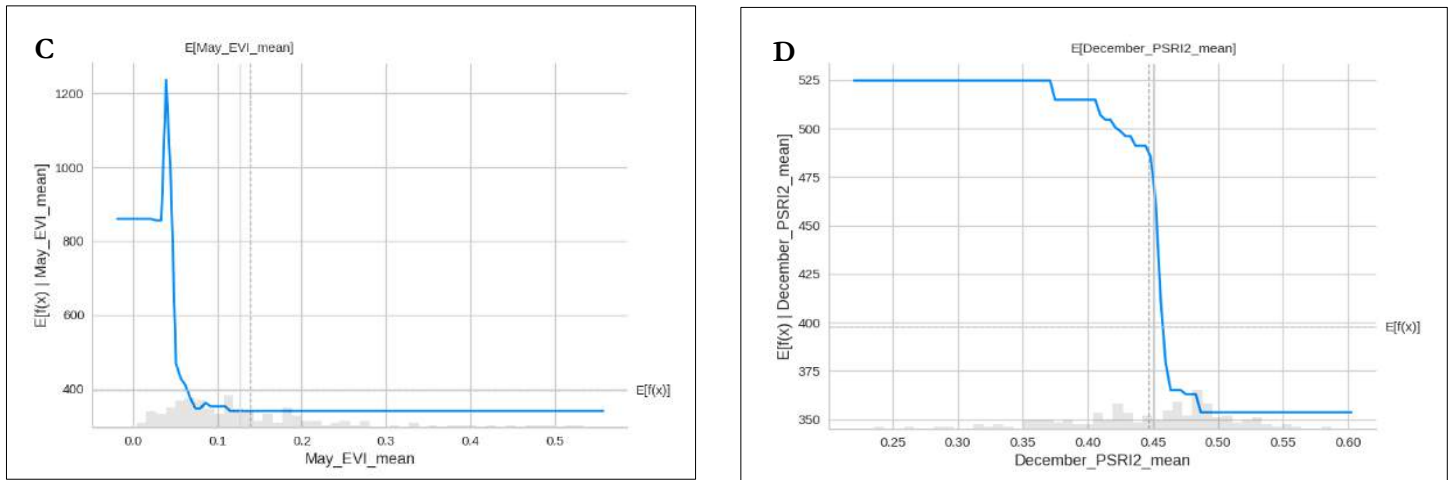


Figure 17. Graphical representation of results for Teff grain and Fe nutrient composition. Image A represents scatter plots of observed versus predicted values, offering a visual comparison between actual measurements and model predictions. Image B is the permutation importance plot, illustrating the significance of the predictors used in the final model. Images C and D are the partial dependence plots for the two most influential features, highlighting their effect on the grain nutrient composition.

#### 4.3.4. Interpretation of Results for Teff - Mn

The features considered include soil pH, evaporation from bare soil, and the normalized difference turbidity index (NDTI). The predictive model had an R-squared of 0.70 and RMSE of 46 mgkg<sup>-1</sup>.

Soil pH is a key determinant of nutrient availability. High pH levels (alkaline conditions) can reduce Mn availability because Mn forms insoluble compounds that plants cannot easily absorb. Conversely, low pH levels (acidic conditions) increase Mn availability, as the nutrient remains soluble and more accessible to Teff (Berbecea et al., 2011). Therefore, understanding the soil pH before planting and during the growing season is crucial for optimizing Mn uptake. This is evident in the PDP plot (see Figure 18D).

Evaporation rates in January, the end of the harvest season, influence soil moisture levels (See Figure 18C). High evaporation rates indicate drier soil condition, which can limit nutrient mobility and availability (Halli et al., 2016), including Mn. On the other hand, low evaporation rates suggest better soil moisture retention, supporting nutrient availability and uptake in the Teff grain.

Towards the end of a crop's life cycle, nutrients are often translocated from the leaves to the grain (Maillard et al., 2015). The NDTI which has been proved by (Bahrawi & Elhag, 2018) to estimate soil water content. At low NDTI values, the Mn content in teff grain is relatively low (See Appendix 5. Results for **teff**). As NDTI values increase, the Mn content in teff grain also increases. This could be mean that soil moisture at the end of the season is for transfer of Mn.

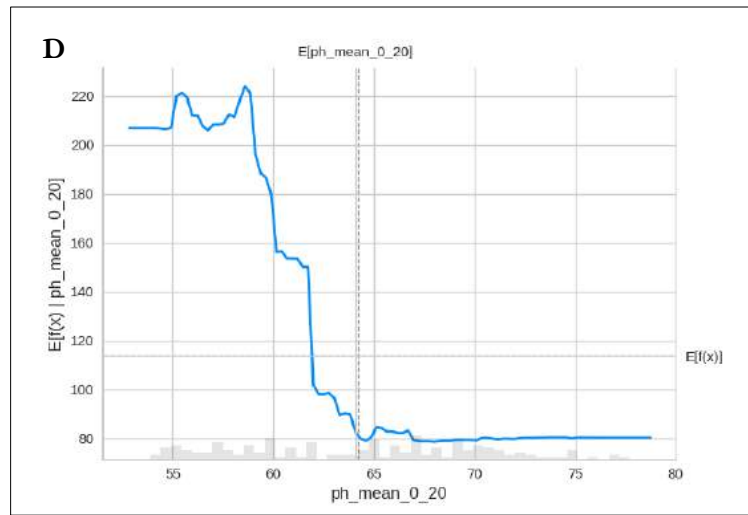
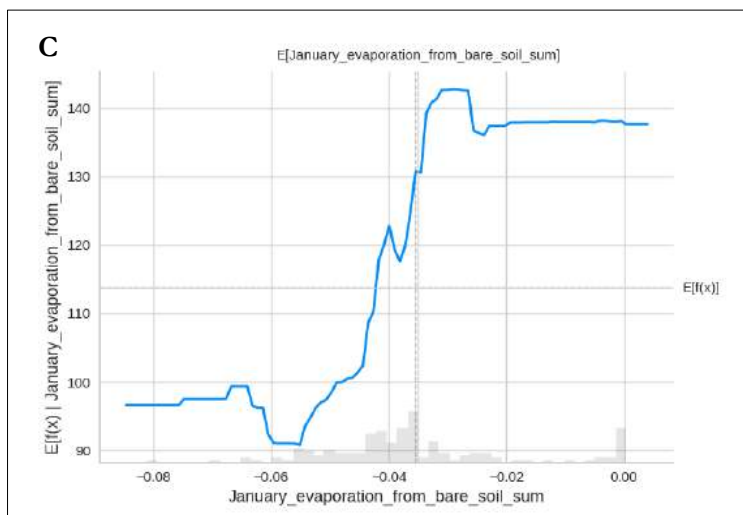
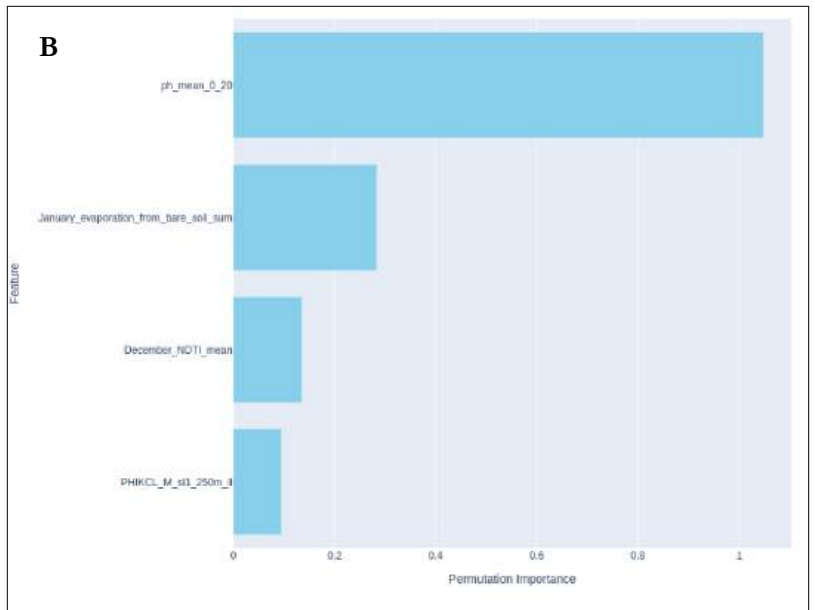
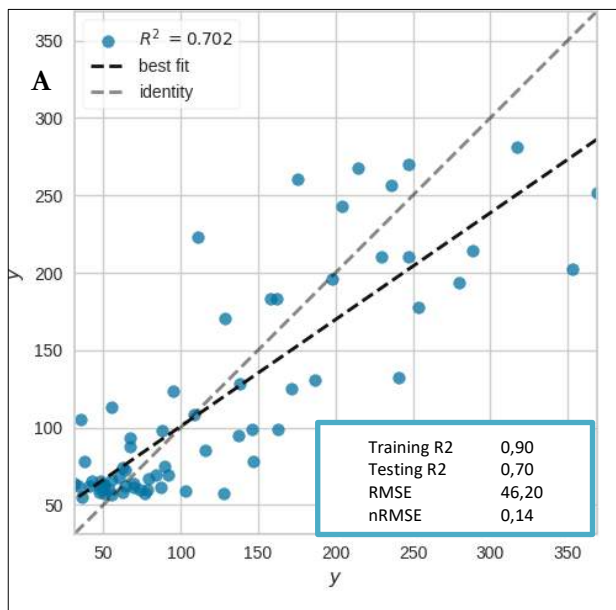


Figure 18. Graphical representation of results for Teff grain and Mn nutrient composition. Image A represents scatter plots of observed versus predicted values, offering a visual comparison between actual measurements and model predictions. Image B is the permutation importance plot, illustrating the significance of the predictors used in the final model. Images C and D are the partial dependence plots for the two most influential features, highlighting their effect on the grain nutrient composition.

#### 4.3.5. Interpretation of Results for Teff - Zn

The predictive model had an R-squared of 0.40 and RMSE of  $3.6 \text{ mgkg}^{-1}$ .

The features considered include the FPAR in July, the normalized difference vegetation index (NDVI) in July, the Ca nutrient content in the soil, the MODIS Terrestrial Chlorophyll Index (MTCI) in May, and total evaporation in September.

High values of July Total FPAR indicate a high fraction of incoming solar radiation being utilized for photosynthesis, which is critical during the planting period. This suggests healthy and vigorous plant growth at the onset of Teff development, enhancing nutrient uptake, including Zn. Conversely, low values imply less efficient photosynthesis, potentially leading to poorer plant growth and reduced Zn absorption. The Partial dependence relation does not show a uniformly positive or negative effect (See Appendix 5. Results for **teff**), The impact can vary significantly based on other contextual factors.

The PDP (See Figure 19C) shows that Zn content in Teff grains increases at low NDVI<sub>re</sub> values. This is difficult to interpret since the NDVI<sub>re</sub> values are negative which mean vegetation is absent and only the soil is visible (Sharma et al., 2022) . Further research is needed to understand this relationship.

Ca content in the soil can improve overall soil health and structure, facilitating better root development and nutrient uptake (Grzebisz et al., 2022). However, very high levels of Ca in the soil can cause a change in the photosynthetic activity of plants and hinder the absorption of nutrients (Weng et al., 2022). This can be seen in the PDP (See Appendix 5. Results for **teff**).

High values of the May MTCI suggest high chlorophyll content and healthy vegetation before the Teff planting season in June and July (Dash & Curran, 2007). The health and productivity of the previous crop or natural vegetation can serve as an indirect indicator of soil health. Healthy vegetation often suggests good soil fertility (Tahat et al., 2020), which influences the nutrient uptake of the subsequent crop. Low MTCI values suggest reduced vegetation health and nutrient status, potentially limiting Zn availability during the initial stages of Teff growth (See Figure 19D).

High total evaporation in September, indicates water loss, which can lead to drier soil or vegetation conditions. This can limit nutrient mobility and uptake, including Zn. Conversely, low evaporation rates suggest better soil and vegetation moisture retention (Hsu & Dirmeyer, 2023), supporting sustained nutrient availability and uptake by the crop (Végh, 1991a). The PDP (See Appendix 5. Results for **teff**) confirms this relation, but with slight mixed.

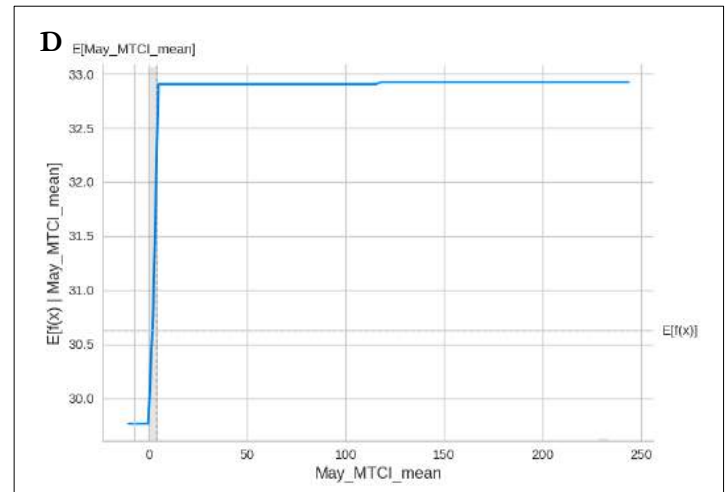
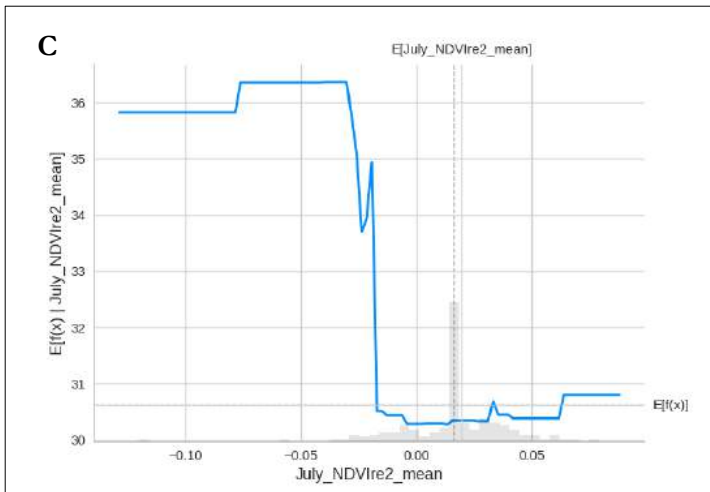
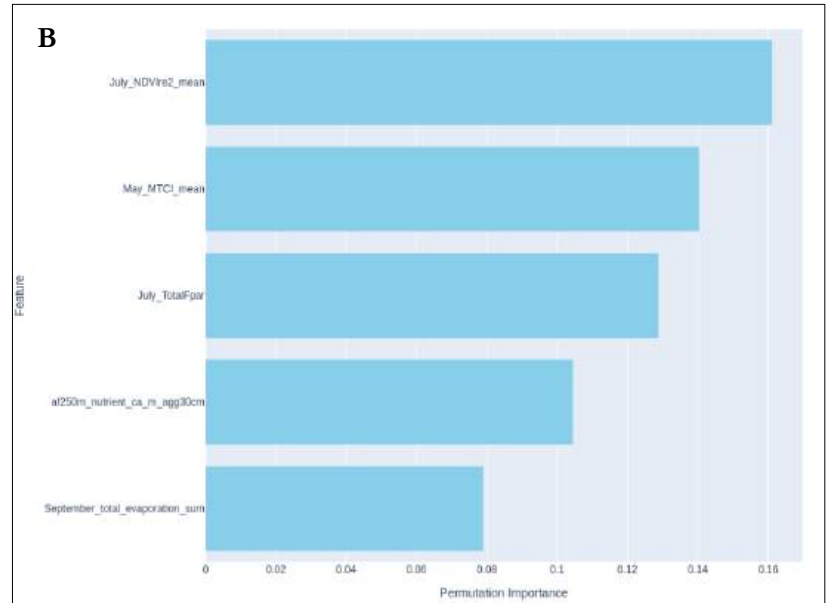
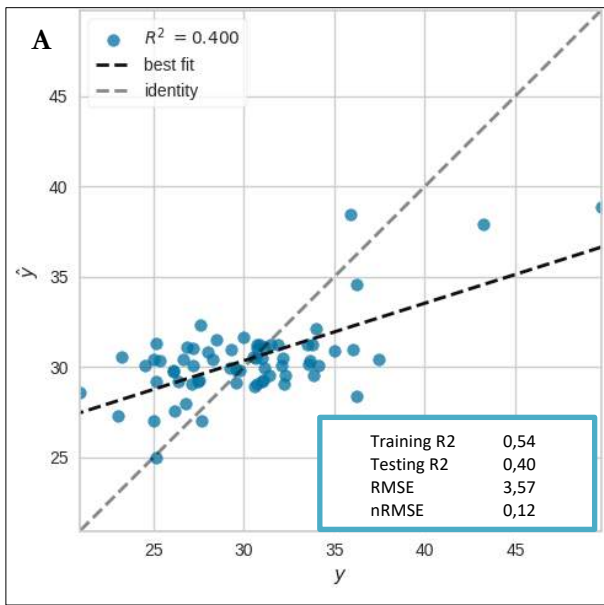


Figure 19. Graphical representation of results for Teff grain and Zn nutrient composition. Image A represents scatter plots of observed versus predicted values, offering a visual comparison between actual measurements and model predictions. Image B is the permutation importance plot, illustrating the significance of the predictors used in the final model. Images C and D are the partial dependence plots for the two most influential features, highlighting their effect on the grain nutrient composition.

#### 4.3.6. Interpretation of Results for Teff - Mg

The important features are soil K content, the enhanced vegetation index in December, the plant senescence/reflectance index in May, the red edge position index in June, and the normalized difference water index in December. The predictive model had an R-squared of 0.64 and RMSE of 152 mgkg<sup>-1</sup>.

From the plots (See Figure 20C), it can be observed that when the K nutrient level in the soil is low, the Mg content in Teff grain is relatively high. As the K nutrient level in the soil increases, there's a sharp decrease in the Mg levels in the Teff grain. This suggests a negative correlation between these two variables. The observed trend could be due to the competitive uptake of nutrients by the plant. In soils with high K levels, the plant might absorb more K at the expense of other nutrients. This phenomenon is known as "antagonistic interaction", where the presence of one nutrient inhibits the uptake or utilization of another (Xie et al., 2021).

When EVI values are low, it typically indicates less healthy or sparse vegetation (Z. Li et al., 2010). The relationship between EVI values and Mg content in Teff grain (See Figure 20D) shows that low EVI values, indicating less healthy or sparse vegetation at maturity, correspond to higher Mg content. Conversely, as EVI values rise, suggesting healthier and more abundant vegetation, Mg content begins to decrease. This phenomenon can be attributed to "nutrient dilution," where the increase in biomass associated with high EVI leads to a dispersion of the same amount of Mg across a larger plant mass, resulting in lower nutrient concentration in the grain (Jarrell & Beverly, 1981).

The relationship between NDWI values and Mg concentration in Teff grain (See Appendix 5. Results for **teff**) shows that low NDWI values, indicating diminished water content in vegetation, correspond to lower Mg concentration in the Teff grain. This suggests that reduced vegetation moisture at maturity stage of the Teff crop is accompanied by a decline in Mg concentration within the grain. Conversely, as NDWI values increase, which means increased moisture levels in the Teff crop during December, an increase in Mg concentration is observed in the grain. This positive correlation between NDWI and Mg concentration shows the pivotal role of vegetation and soil moisture in supporting optimal nutrient uptake (Végh, 1991b).

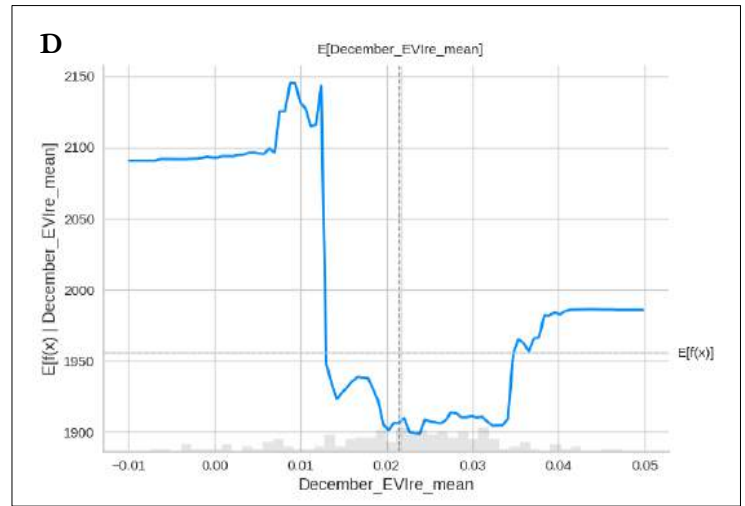
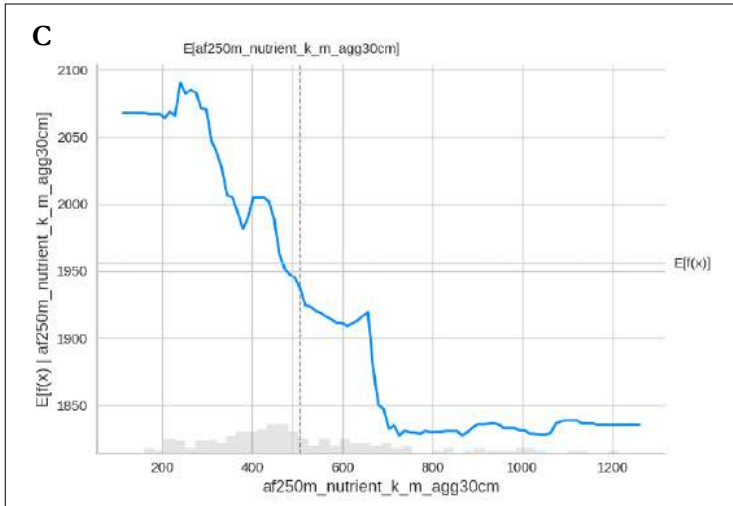
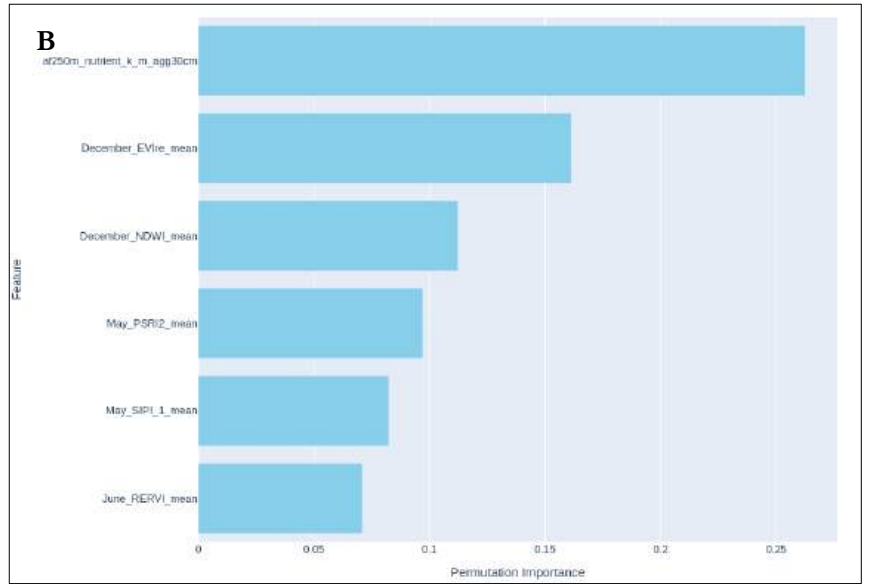
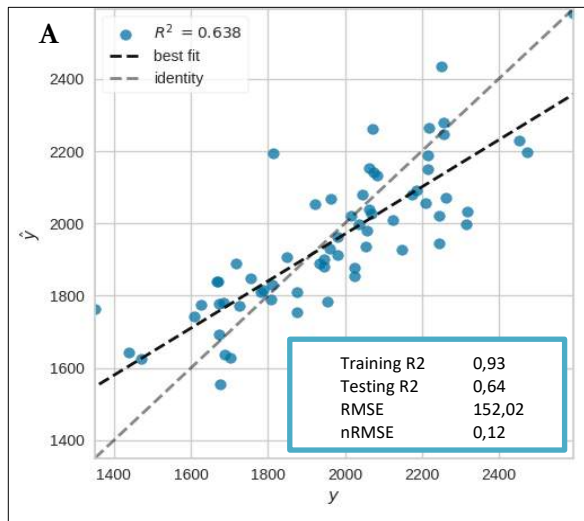


Figure 20: Graphical representation of results for Teff grain and Mg nutrient composition. Image A represents scatter plots of observed versus predicted values, offering a visual comparison between actual measurements and model predictions. Image B is the permutation importance plot, illustrating the significance of the predictors used in the final model. Images C and D are the partial dependence plots for the two most influential features, highlighting their effect on the grain nutrient composition.

#### 4.4. Macronutrient Analysis for Teff

The macronutrient analysis for Teff in this study includes an examination of potassium (K), phosphorus (P), and sulphur (S). These macronutrients are crucial for energy supply, body repair, and growth.

##### 4.4.1. Interpretation of Results for Teff - K

The features considered in this analysis include soil clay content, FPAR, soil pH, GCI in May, the TWI, and NDWI in June. The predictive model had an R-squared of 0.44 and RMSE of 435.7 mgkg<sup>-1</sup>.

Soil clay content profoundly impacts nutrient retention and water availability (Dias et al., 2023). High clay levels enhance nutrient retention, potentially improving K availability for Teff. However, excessive clay can cause waterlogging and poor aeration, hindering root growth and nutrient uptake (Alhaj Hamoud et al., 2019). Conversely, low clay content leads to rapid nutrient leaching, reducing K availability.

The relationship between clay content and K levels in Teff grains reveals a distinct pattern (See Figure 21C). Low clay content corresponds to low K in the grain. In the mid-range, K levels fluctuate unpredictably, indicating the influence of other soil factors. While higher clay content enhances K levels due to improved nutrient retention.

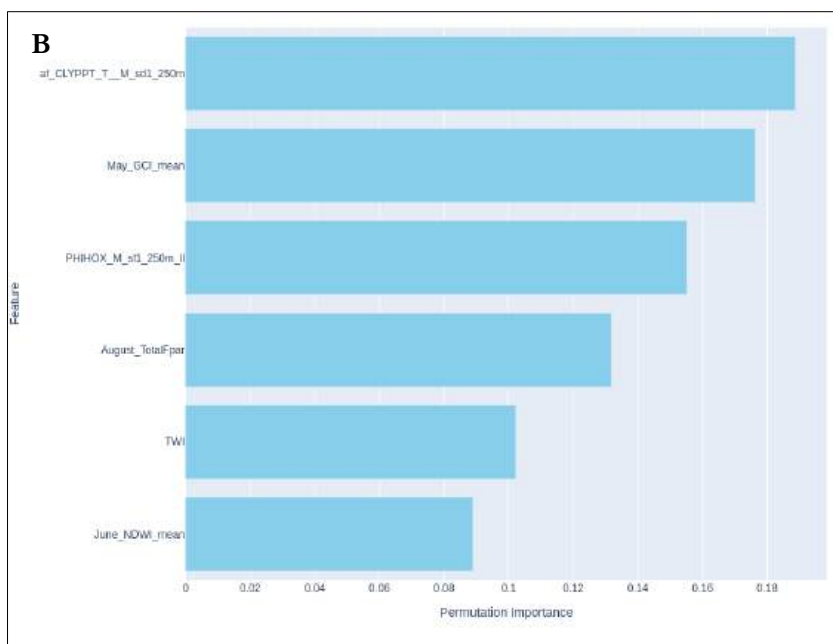
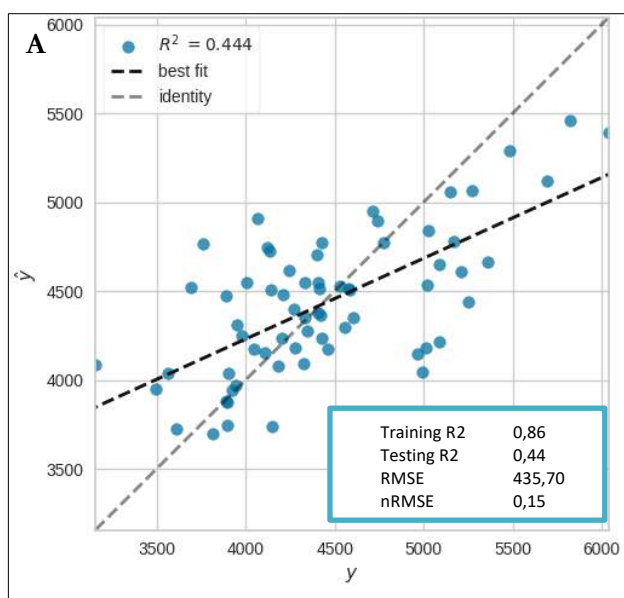
High values of August FPAR indicate a substantial fraction of incoming solar radiation utilized for photosynthesis during the mid-season, implying robust plant growth and heightened metabolic activity (C. Tan et al., 2018). Conversely, low values of FPAR suggest less photosynthetic crop activity due to low energy absorption, potentially resulting in K transfer into the grain during the mid-season. The observed trend in the plot (See Appendix 5. Results for **teff**) shows that relationship between FPAR and K content in Teff grains is negative. As FPAR increases, K levels tend to decrease. This can be due to dilution effect discussed by (Jarrell & Beverly, 1981), where the rapid growth of the crop, in this case caused by increased photosynthesis, leads to a decrease in the concentration of K in the grain.

The relationship (See Appendix 5. Results for **teff**) between soil pH and K levels in Teff grains reveals a pattern suggestive of the plant's preference and tolerance in slight acidic to alkaline soils (Berhanu et al., 2023). At low soil pH values, indicating high acidity, K levels in Teff grains are relatively low. However, as the soil pH increases, transitioning towards neutrality and then alkalinity, there is an increase in K content in the grains. This suggests a potential preference for slightly alkaline conditions for K uptake.

The PDP plot (See Figure 21D) for May GCI and K content in Teff grains shows the relationship between pre-season vegetation health and K availability. Low GCI values signify low chlorophyll levels, indicating less vegetation growth pre-season, which potentially allows Teff grains to access higher levels of K at that the start of the season. This suggests that less pre-season vegetation leads to reduced competition for soil nutrients, facilitating greater K uptake by Teff when it is planted. Conversely, elevated GCI values correspond to active and healthier vegetation (Wu et al., 2012), correlates with decreased K content in Teff grains. This inverse relationship implies competition for K among plants when Teff is planted, resulting in diminished availability of this nutrient for Teff

When TWI is low, indicating soils that are likely to be well-drained and less likely to accumulate water, the K content in Teff grain is relatively low. As the TWI increases (See Appendix 5. Results for **teff**), indicating areas that are more likely to accumulate water (Kopecký et al., 2021), the K content in Teff grain increases. This could be because in areas where water accumulates, there may be greater nutrient availability (S. Zhang et al., 2011a). However, when the TWI is increases further, the K content in Teff grain slightly decreases. This could be due to waterlogged conditions in areas with very high TWI values, which can lead to anaerobic conditions and potentially limit the availability or uptake of nutrients (Fitter & Hay, 2002).

The PDP (See Appendix 5. Results for **teff**) shows that when NDWI values are high, indicating ample water content in vegetation and soil, K levels in Teff grains are stable and moderately high. This suggests that adequate water availability during planting supports effective K uptake by the plants at the start of the season as was confirmed by (Xue et al., 2017). As NDWI values decrease to moderate levels, there is a slight decline in K content. At low NDWI values, indicating reduced water content, there is a noticeable drop in K levels, highlighting that less water availability limits nutrient absorption by teff crop. Further low NDWI values, reflective of dry conditions, correspond to the lowest K content, emphasizing that severe water stress during planting severely restricts K uptake by Teff plants.





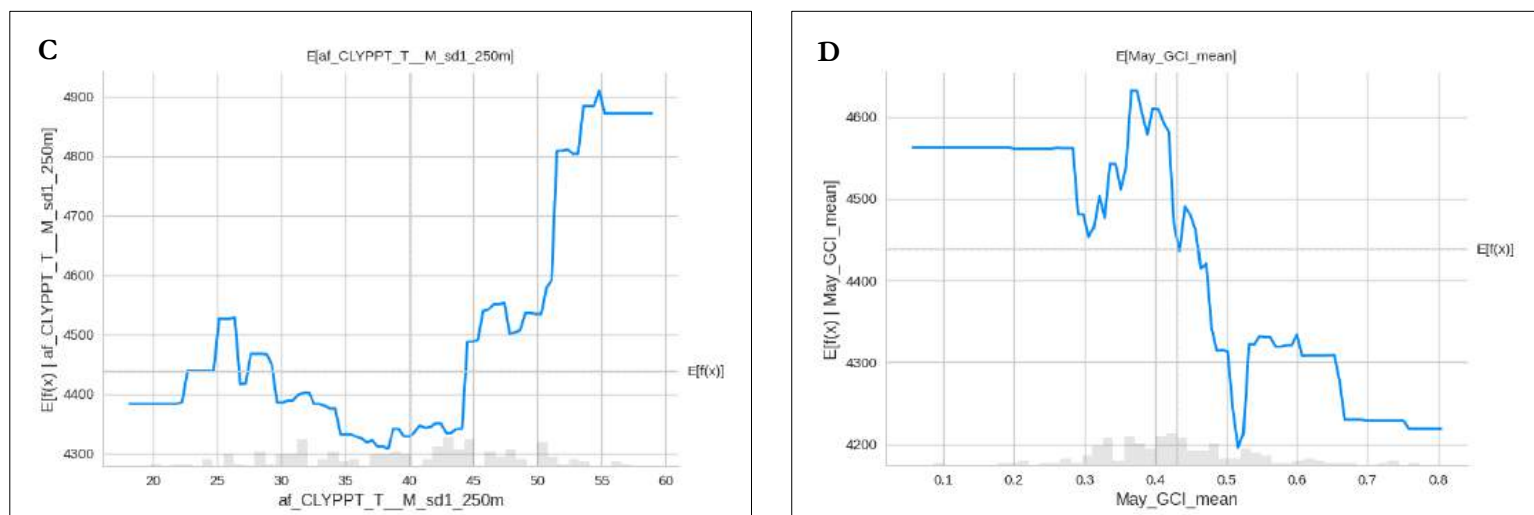


Figure 21: Graphical representation of results for Teff grain and potassium nutrient composition. Image A represents scatter plots of observed versus predicted values, offering a visual comparison between actual measurements and model predictions. Image B is the permutation importance plot, illustrating the significance of the predictors used in the final model. Images C and D are the partial dependence plots for the two most influential features, highlighting their effect on the grain nutrient composition.

#### 4.4.2. Interpretation of Results for Teff - P

The features considered include the MTCl in May, S2REP in September, and soil silt content. The predictive model had an R-squared of 0.37 and RMSE of 530 mgkg<sup>-1</sup>.

When MTCl values are low, the P concentration in the grain is also low. Conversely, as MTCl values increase, there is a rise in P concentration in the Teff grain, indicating vegetative activity by pre-season crops like cover crops (See Figure 22C). Higher chlorophyll content of vegetation is indicative of good soil health (Tahat et al., 2020). The vegetation cover could lead to improved soil conditions like regulation of soil temperature, structure and moisture which may be suitable for microbial activities, which lead to uptake of nutrients during the season (Philippot et al., 2013).

As the S2REP increases, indicating healthier and more vigorous vegetation, the P content in Teff grain generally decreases. This inverse relationship suggests that higher S2REP values is linked to reduced P uptake or utilization by the Teff crop during the mid-season (See Appendix 5. Results for teff).

The decline in P content with increasing S2REP values during mid-season could be due to the plants prioritizing other metabolic processes or the remobilization of P nutrient in the crop for vegetative growth (Bender et al., 2015).

The PDP (See Figure 22D) indicates that P content in Teff grain generally increases with higher silt content in the soil, suggesting that soils with high silt content may enhance P uptake. However, the plot also shows a non-linear pattern: P content initially increases, then decreases, and subsequently increases again as silt content rises. Specific research on silt content in soils and P is limited, but research has shown that on when clay content is high in soil, it has a negative relation with the amount of P in soil (Jian et al., 2022).

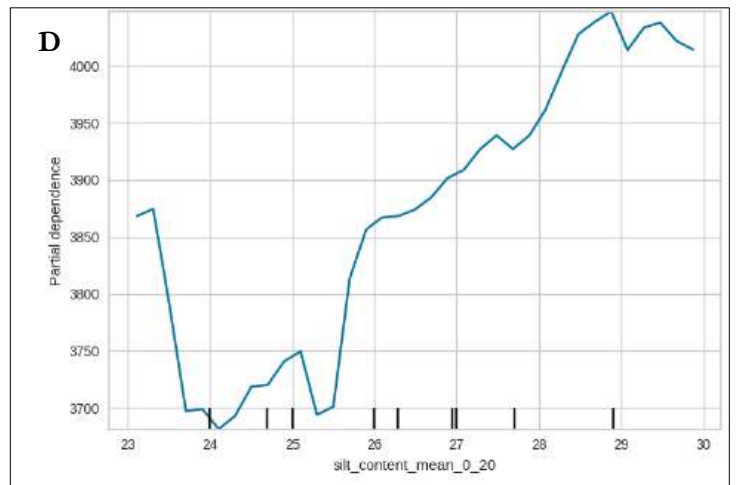
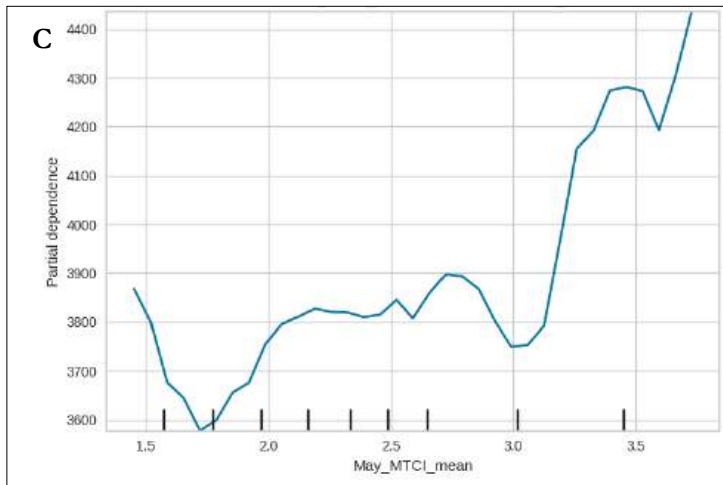
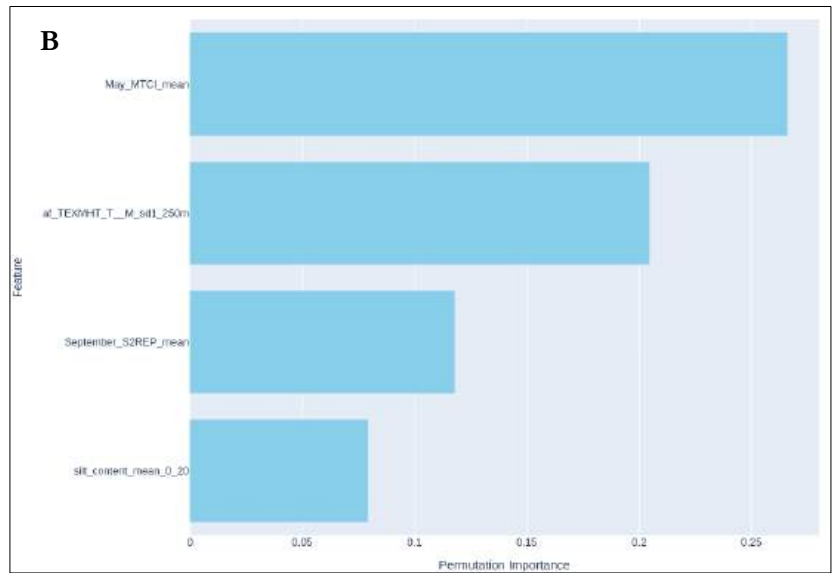
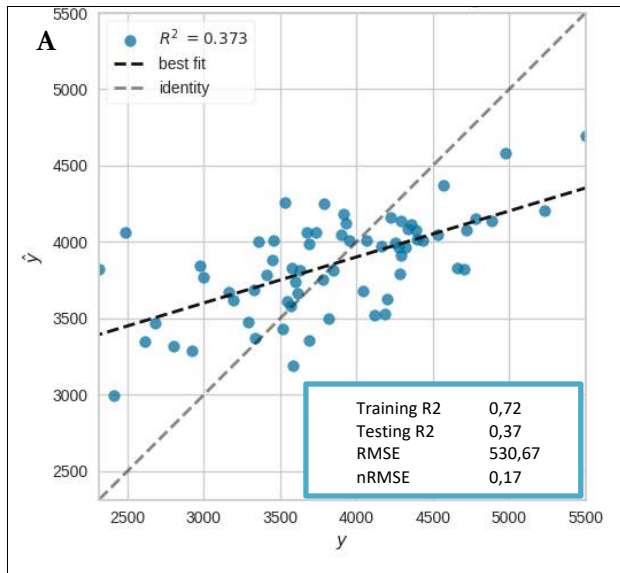


Figure 22: Graphical representation of results for Teff grain and P nutrient composition. Image A represents scatter plots of observed versus predicted values, offering a visual comparison between actual measurements and model predictions. Image B is the permutation importance plot, illustrating the significance of the predictors used in the final model. Images C and D are the partial dependence plots for the two most influential features, highlighting their effect on the grain nutrient composition.

#### 4.4.3. Interpretation of Results for Teff - S

The features considered include harvest month rainfall, the NDTI in August, the PSRI index in May, and the WDRVI in June. The predictive model had an R-squared of 0.36 and RMSE of 188 mgkg<sup>-1</sup>.

PSRI signifies canopy stress or the start of crop senescence (Merzlyak et al., 1999). The PDP (See Figure 23D) implies that while moderate increases in PSRI are beneficial for S accumulation in Teff grain, very high PSRI values lead to a decline in S in the grain. Pre-season crops could be undergoing senescence during this period, and their decomposition could lead to increased microbial activity add nutrients to the soil (Sievers & Cook, 2018). Although the Teff crop was not planted in May, PSRI measurements reflect the state of existing vegetation, which can significantly influence soil conditions and nutrient availability.

The relationship between the WDRVI in June and the S concentration in Teff grain (See Figure 23C), shows that early in the growing season, low WDRVI values correspond to low S concentrations. As WDRVI values increase, indicating improving plant health and Vigor (Gitelson, 2004), S concentration in the grain also rises. This phase aligns with the period shortly after planting when Teff plants are starting to uptake nutrients, including S, from the soil. However, beyond a certain threshold, as WDRVI continues to increase, S concentration in the grain decreases. This inverted U-shape relationship suggests that while moderate improvements in plant health support increased S uptake, excessively high WDRVI values may indicate that the plants are reallocating resources leading to reduced S accumulation in the grain.

Harvest month rainfall influences soil or vegetation moisture levels. Adequate rainfall supports soil moisture retention (Shahadha et al., 2021), facilitating nutrient uptake by Teff. However, excessive rainfall may lead to nutrient runoff (Skidmore et al., 2023), potentially reducing S availability. Monitoring rainfall patterns helps assess S availability for Teff during the harvest season. The PDP (See Appendix 5. Results for teff) shows an increase in S concentrations as rainfall increases in the month of harvest.

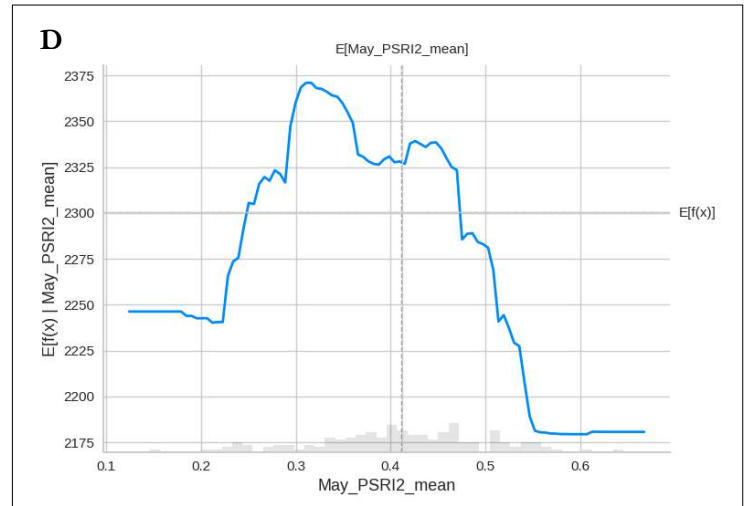
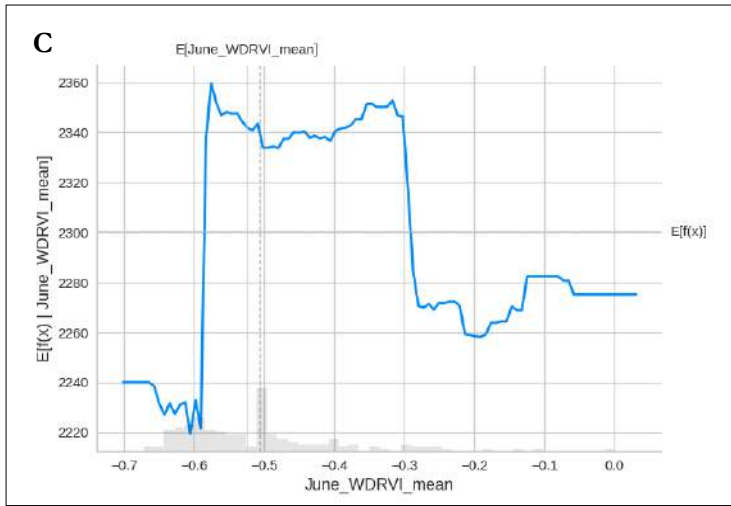
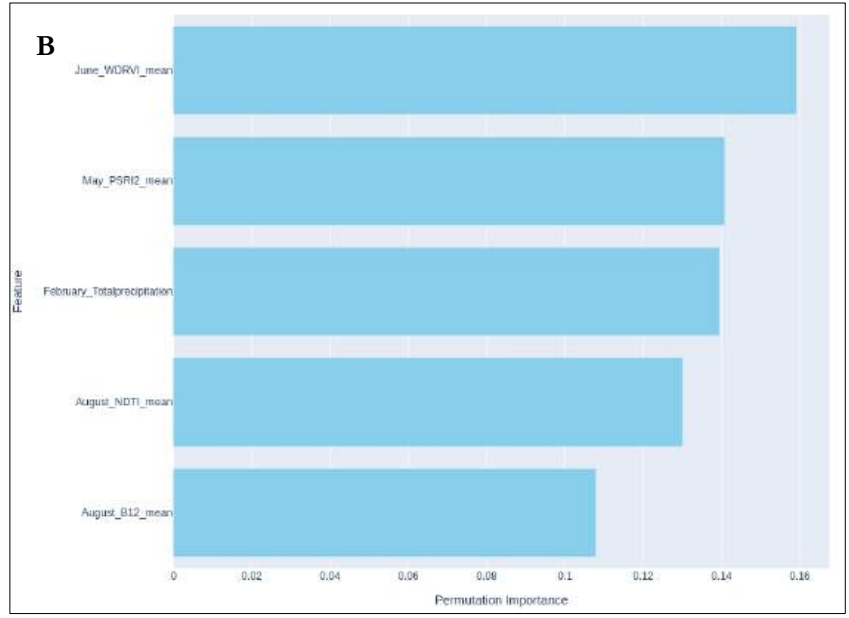
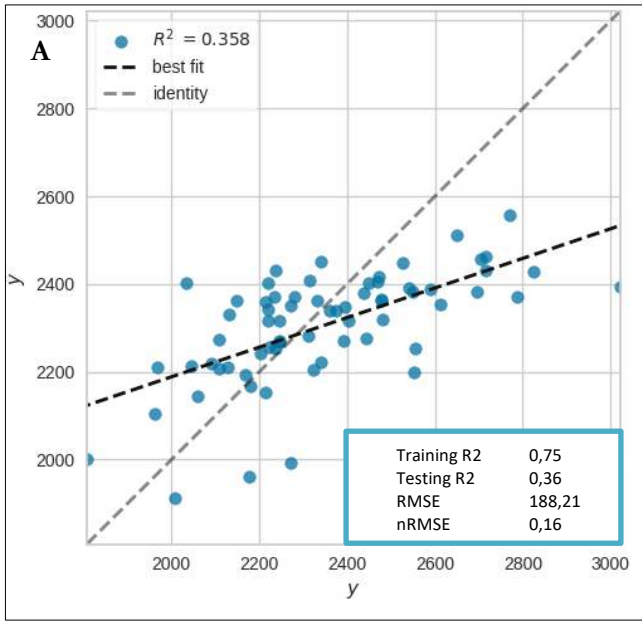


Figure 23: Graphical representation of results for Teff grain and S nutrient composition. Image A represents scatter plots of observed versus predicted values, offering a visual comparison between actual measurements and model predictions. Image B is the permutation importance plot, illustrating the significance of the predictors used in the final model. Images C and D are the partial dependence plots for the two most influential features, highlighting their effect on the grain nutrient composition.

WHEAT		MAIZE		TEFF		SORGHUM		BARLEY	
Nutrient	R-squared(%)	Nutrient	R-squared(%)	Nutrient	R-squared(%)	Nutrient	R-squared(%)	Nutrient	R-squared(%)
Ca	40	Ca		Ca	13	Ca	29	Ca	28
Cu	26	Cu	60	Cu	30	Cu	73	Cu	42
Fe	13	Fe		Fe	44	Fe	54	Fe	
K	39	K	26	K	44	K	46	K	40
Mg	58	Mg	33	Mg	64	Mg	55	Mg	33
Mn	58	Mn		Mn	70	Mn	34	Mn	32
P	32	P	24	P	37	P	33	P	46
S	19	S	30	S	35	S	30	S	41
Zn	21	Zn	13	Zn	42	Zn	46	Zn	26

Figure 24: Tabular representation of statistical accuracy (R<sup>2</sup>) of models for all crops and all nutrients involved in this study.

## 5. DISCUSSION

After presenting the results and discussing the impacts of different datasets on predicting nutrient grain composition in Teff, focus is shifted to the four other crops: Barley, Maize, Sorghum, and Wheat. This section will discuss first, the results for these crops and examine how various important geospatial datasets impact the estimation of their grain nutrients. The discussion will be structured around the research questions, which will be mentioned first, and followed by insights from the PDPs (See Appendix 2. Results for **barley** to Appendix 6. Results for **wheat**) and literature. The machine learning algorithm and the limitations of the study will follow.

### 5.1. Research Question for Objective 1:

#### 5.1.1. Climatic factors

This study found varying impacts of rainfall and temperature on the nutrient composition of grains for different crops. Generally, early-season rainfall was identified as a key predictor of nutrient composition, suggesting that adequate soil moisture at the start of the season positively impacts enhanced nutrient uptake during critical growth periods. However, in one scenario, late-season rainfall initially boosts S content in the grain but eventually compromises it due to excess moisture.

Barley exhibited a negative relationship between increasing rainfall and Ca, particularly at the mid-season, indicating that higher rainfall levels may lead to increased soil moisture and negatively impact Ca accumulation in the grain. This observation is similar to findings of (McKay Fletcher et al., 2022), who found that increased precipitation can lead to a reduction of nutrient use efficiency. Wheat benefited from mid-season rainfall, showing a positive association with both Ca and Cu content.

Similarly, macro-nutrient composition was influenced by climatic factors, with barley experiencing a negative relationship between rainfall and K, P, and Mg content at the early stages of the crop season. High pre-season temperatures negatively impacted S content in barley, confirming that high temperature negatively impacts nutrient absorption at the early stages of crop growth (Giri et al., 2017). Maize demonstrated mixed responses to increased rainfall. In sorghum, an increase in air temperature at the start of the season led to reduced K and P nutrient content of the grain. Wheat presented a unique scenario, with high pre-season rainfall correlating negatively with Mg content but positively influencing P content. During the mid-season an increase in rainfall generally led to the increase of P content in the wheat grain. These observations highlight the complex interplay between rainfall and macro-nutrient accumulation.

#### 5.1.2. Soil

Soil properties, including texture, pH, and organic matter content, play a fundamental role in determining the nutrient composition of crop grains. These characteristics influence the availability and uptake of nutrients by crops, impacting agricultural productivity across various agroecological zones (Tahat et al., 2020). The relationship between soil properties and nutrient concentrations in crops, showed the different ways through which these soil characteristics influence agricultural productivity.

Research has shown that fine-textured soils, rich in clay and silt, exhibit a high capacity for holding both water and nutrients, thereby supporting enhanced crop growth and nutrient accumulation (Matus, 2021). Specifically, a decrease in total evaporation at the start of the season was linked to a higher concentration of

Cu in barley grain, confirming the impact of moisture conservation on micronutrient uptake (Cavagnaro, 2016). Additionally, the interaction between soil pH and texture was highlighted, with slightly acidic soils reducing the nutrient uptake of certain nutrients, as was proved by (Zama et al., 2022). Slight alkaline soils were shown to increase nutrient solubility, leading to higher nutrient concentrations in certain crops.

In barley, high Mg content in the soil was associated with low Mn and Zn nutrient concentrations in the grain, indicating a direct trade-off between Mg abundance and the availability of these micronutrients. Similarly, maize exhibited a stable Fe grain content until a certain threshold of soil Zn composition was reached, beyond which Fe content sharply increased, suggesting a threshold effect where Zn influences Fe content in the grain. The nutrient composition of sorghum was influenced by bulk density and P content, with increases in bulk density leading to increased Cu nutrient. and increases in amount of P in the soil leading to reduced Cu and Zn compositions in the grain

For macro-nutrients, the K, P, and Mg grain nutrient content of barley was influenced by soil Mg content, with high evaporation rates and soil Mg levels correlating with increased K and Mg content in grain. This indicates a positive relationship between soil moisture and the accumulation of these macro-nutrients in the grain. The S content in the maize grain was negatively impacted by increasing Fe and clay content in soil, demonstrating the interconnectedness of soil properties and nutrient dynamics, where changes in soil composition directly affect S availability. Sorghum's K content in grain was positively correlated with soil Ca levels, confirming that soil K management can directly influence the uptake of other essential macro-nutrients (C. Wang et al., 2024). Wheat grain macronutrient composition in relation to soil properties revealed that an increase in soil Zn content was found to lead to an increase and stabilization of K nutrient content in wheat grain, despite further increasing Zn content in the soil. This indicates that subsequent addition of certain nutrients after the absorption of a required quantity of nutrients needed by plants for biochemical processes, will not increase plant nutrient use efficiency (Good & Beatty, 2011). Also, soil P content exhibits a direct positive relationship with P content in wheat grain, indicating that soil P availability is a key determinant of wheat's P nutrition.

In barley, a positive mid-season relationship between soil volumetric water content and grain Mn content is observed, suggesting that optimal moisture levels are essential for maximizing Mn uptake. Additionally, lower total evaporation at the start of the season correlates with higher Zn content in the grain, indicating that minimal evaporation supports higher nutrient concentrations. Sorghum presents a unique case where an increase in the start-of-season volumetric water content initially shows a low-stabilized relationship with grain Mn content. However, surpassing a certain threshold of increase, there is a significant increase in Mn content in the grain.

An increase in soil Mn content leads to an increase in S nutrient content of wheat grain, suggesting Mn might influence S metabolism or uptake mechanisms in wheat. Conversely, an increase in soil Mg content was associated with a decrease in S nutrient content of wheat grain, indicating that high Mg levels might interfere with S uptake or utilization in wheat.

In barley, high total evaporation towards the end of the season is associated with elevated P content in the grain, while mid-season S content remains low and stable until a certain threshold is surpassed, triggering a rapid increase. In maize, Mid-season total evaporation also demonstrates a general negative relationship with grain K content, highlighting the nuanced effects of evaporation on nutrient composition. The maize crops also showed a pattern where end-of-season volumetric water content initially correlates positively with grain K content, and then abruptly reverses after a certain threshold. In wheat the relationship between volumetric water content at the start of the season and grain K content shows no clear pattern and is most likely

influenced by other factors. Also in wheat, reduced soil evaporation pre-season and higher P content in the grain are positively correlated. These variable results demonstrate the varied impact of soil physical and chemical properties on micronutrient availability.

### **5.1.3. Topographic factors**

Topographic factors, including elevation, slope, and the Topographic Wetness Index (TWI), play a significant role in determining the nutrient composition of crop grains across different agroecological zones. These factors can influence both positively and negatively the concentration of essential nutrients in cereal crops (Zhang et al., 2011). In this study, an increase in elevation was found to correlate negatively with S content in maize grain, suggesting that higher altitudes might lead to reduced S availability due to changes in soil properties or atmospheric deposition rates. Conversely, as elevation increases, Fe nutrient content in wheat grain also increases, indicating a potential positive relationship between altitude and Fe availability in wheat. Also, as elevation rises, the S nutrient content in wheat grain decreases, further highlighting the complex nature of nutrient distribution influenced by topography.

The Topographic Wetness Index (TWI) was observed to have a specific effect on the K content in Teff grain, with increasing TWI values correlating with increased K content. Areas with higher TWI values, indicative of sites prone to waterlogging or saturation, may exhibit altered nutrient cycling processes (Kopecký et al., 2021), potentially leading to enhanced availability of certain nutrients like K in crops grown in these conditions.

## **5.2. Research Question for Objective 2(a):**

What is the impact of Sentinel-1 derived data, including bands, and polarimetric indices, on the prediction of nutrient composition in crop grains across diverse agroecological zones at different crop growth stages?

Utilizing Sentinel-1 data (bands and indices), presents a promising yet challenging approach to predict the nutrient composition of crop grains across various agroecological zones during different crop growth stages. Sentinel-1, with its C-band radar imaging capabilities, offers advantages in monitoring extensive agricultural landscapes regardless of weather conditions (Khabbazan et al., 2019). However, initial findings suggest that Sentinel-1 data does not significantly improve the predictions of nutrient compositions in crops such as Teff, Barley, Wheat, Sorghum, and Maize. This limited effectiveness needs to be studied to maximize Sentinel-1 data's utility in crop nutrient monitoring and management. Despite these limitations, Sentinel-1 data holds significant potential for agricultural applications.

Specifically, certain micronutrient relationships with Sentinel-1 derived parameters have been noted. For example, in barley, the Radar Vegetation Index (RVI) at the beginning of the season correlated positively with Mn grain content. In maize, the start of the season's vertical-vertical (VV) polarization and the mid-season's vertical-horizontal (VH) polarization had mixed impacts on Ca grain content. In sorghum, the mid-season RVI showed a negative relationship with Ca content in grain, while the end-of-season cross-ratio index indicated that Fe grain content remained low until a subsequent increase in the cross-ratio, after which it increased significantly.



### **5.3. Research Question for Objective 2(b):**

The utilization of Sentinel-2 indices, especially those involving red-edge and Short-Wave Infrared (SWIR) bands, or green bands, has advanced our ability to predict the nutrient composition of crop grains (Belgiu et al., 2023). These bands are uniquely capable of capturing detailed information about soil and vegetation health and structural properties, which are closely linked to nutrient status and crop growth stages (Liu et al., 2004). The red-edge bands, situated around 700 nm, are particularly sensitive to changes in leaf chlorophyll content and green Leaf Area Index (LAI), key indicators of plant health and vigor. These bands can detect subtle shifts in vegetation structure and pigment content, often associated with differences in nutrient availability and uptake. Similarly, the SWIR bands, operating in the 1600-2400 nm region, are effective in assessing soil moisture and organic matter content, which indirectly influence nutrient availability to plants (Liu et al., 2004). The combination of these bands enables a comprehensive assessment of both plant physiological status and soil conditions, providing a robust foundation for predicting nutrient composition in crops.

However, the timing of these measurements throughout the crop growth cycle is crucial. Indices derived from Sentinel-2 data in this study were found to be most informative during the months preceding planting and during the early stages of the season. Towards the end of the mid-season, as crops prepare for harvest, the relationships between Sentinel-2 indices and nutrient composition become less straightforward. This period could be marked by a shift in plant priorities towards reproductive development and senescence, which can obscure the direct links between remote sensing data and nutrient status (Colle et al., 2009).

In barley, the mid-season increase in the SIPI correlates negatively with Ca content, suggesting that increased stress levels, possibly due to suboptimal water or nutrient availability, can reduce the plant's capacity to absorb Ca efficiently. This finding aligns with the Disease Water Stress Index (DSWI)'s sensitivity to water stress, indicating that disease or stress conditions can impair nutrient absorption (Apan et al., 2004). At the end-of-mid-season, an increase in NDVI<sub>re</sub> correlates positively with Cu content, reflecting improved vegetation health and chlorophyll content, which enhances photosynthetic activity and nutrient utilization, thereby increasing Cu accumulation in grains. Maize shows a pre-season increase in the GCI correlating with increased Fe content in grain beyond a certain threshold. Sorghum shows a mid-season increase in SIPI correlating with low Ca content in the grains. Additionally, Wheat demonstrates an end-of-season increase in DSWI correlating with reduced Ca content, consistent with the index's sensitivity to water stress and plant damage and reduced plant functioning.

In barley, a positive relationship exists between mid-season NRI and S content, indicating that increased nitrogen content in crops can influence S uptake and assimilation. The study showed that with increasing mid-season NDWI, Mg nutrient in grain remains steadily low, then immediately increases after certain threshold. Sorghum exhibits a correlation between mid-season increases in the PSRI and decreased K content in grain, meaning that as plants are stressed K nutrient utilization in the crop is low. Wheat presents a negative correlation between pre-season high EVI and Mg content, suggesting that active vegetation in the pre-season may utilize Mg which may not be available when wheat is planted.

### **5.4. Research Question for Objective 3:**

In barley, the analysis shows that high LAI values during the mid-season led to a reduction in Mg grain content. In maize, the analysis indicates that an increase in total FPAR at the onset of the season results in a decrease in Mg and P grain content. This suggests that the plant's energy, primarily directed towards photosynthesis, could affect the uptake and allocation of these nutrients.

The nutrient composition in wheat is affected by both FPAR and LAI at different growth stages. The analysis shows that an increase in FPAR towards the end of the season results in higher Fe content in the grain, suggesting enhanced photosynthetic activity leads to increased Fe uptake. Higher FPAR at the start of the season resulted in increased Mn content in the grain, confirming the relation between Mn and photosynthesis and its uptake during this period (Messant et al., 2022). Higher mid-season LAI values were linked to lower Ca content in the grain, suggesting competition for Ca among the increased leaf area. Higher LAI values at the start of the season are associated with increased Cu content in the grain.

These findings collectively emphasize the importance of considering crop-specific responses and growth stages in optimizing nutrient composition.

## **Random Forests**

Random Forests are renowned for their resistance to overfitting, largely attributed to their ensemble nature, which combines the predictions of multiple decision trees to produce a final model (Belgiu & Drăgu, 2016). However, the study observed overfitting tendencies in most models, potentially influenced by factors such as the sample size of certain crops and the quality of the data. The approach to replace missing values with the mean of the available data, may not accurately represent the underlying distribution of the missing values, leading to inaccuracies in the model's predictions.

The study utilized a range of 10 to 150 trees for the Random Forest model, diverging from usual usage of 500 trees. This decision shows the importance of hyperparameter tuning and emphasizes that optimal configurations can vary significantly based on the dataset and problem at hand. The process of hyperparameter tuning revealed that increasing the number of trees improved model accuracy up to a certain number and then further increase in number of trees led to reduced accuracy. This highlights the need for careful consideration and validation of model parameters to achieve the best performance

Additionally, during the implementation of models, the significance of one parameter deemed not very important (not used in hyperparameter tuning) was observed. This parameter is the “random\_state” parameter in Scikit-learn's implementation of Random Forests. This parameter controls the seed for the random number generator, influencing the randomness in the model, thereby affecting its ability to generalize from the training data to unseen data (Pedregosa et al., 2011). Changing the random state parameter as part of the manual hyperparameter tuning steps, led to increase and decrease of model accuracy.

### **5.4.1. Limitations**

While the study presents promising advancements in the use of remote sensing and machine learning for agricultural applications, it is crucial to consider the limitations of the approach used in the study, which stem from both methodological constraints and broader challenges in remote sensing and agricultural modelling.

A significant limitation of this study is the omission of crop varietal differences. Crop variations can substantially influence nutrient uptake and translocation within plant tissues (Hossein et al., 2010; Lightfoot, 2018; Mir et al., 2022; Vinod, 2019). By not accounting for varietal differences, the models potentially overlook key determinants that affect nutrient composition, thereby reducing the accuracy of the predictions. Future research should endeavour to integrate data from crop varieties to enhance model precision and applicability across diverse agroecological contexts.

The study did not incorporate the heterogeneity of farm management practices, which are pivotal determinants of soil health and crop productivity. Variations in practices such as fertilizer application, weed management, and the use of cover crops introduce significant heterogeneity in soil conditions and crop responses at different stages of crop growth. These practices are known to modulate soil nutrient availability and microbial activity, thereby impacting crop nutrient composition (Gupta et al., 2022; Huffman et al., 2023; Khmelevtsova et al., 2022; L. Wang et al., 2021). The omission of these factors likely leads to an oversimplification of the agroecosystems under study, which could result in non-representative model outputs.

Another methodological limitation is the temporal limitation of the soil data, which was last updated in 2017. Soil properties are dynamic and subject to change due to factors such as climatic variations, land use changes, and continuous agricultural practices. The reliance on outdated soil data undermines the temporal relevance of the model predictions, as current soil conditions may differ significantly from those recorded several years ago. This temporal mismatch necessitates the acquisition of up-to-date soil data to improve the temporal accuracy of the model.

The spatial resolution of the datasets used also presents a significant challenge. For instance, ERA5-Land data, with a spatial resolution of 11 km, does not capture the variability within small-holder farms, which typically cover much smaller areas of less than 1 hectare in area. This discrepancy in spatial resolution can lead to a loss of critical detail, adversely affecting the model's ability to accurately predict nutrient composition at the farm level. Future studies should seek to utilize higher resolution datasets to better match the scale of small-holder farming operations.

The approach of averaging the spatial resolution of satellite images around farm point locations can introduce mixed pixel effects. In agricultural landscapes where farms are interspersed with other land uses such as residential areas, plantations, or different crop types, averaging can result in pixel values that do not accurately represent the specific farm in question. This mixed pixel problem is a well-recognized issue in remote sensing and underscores the need for different spatial analysis techniques that can remove these mixed signals.

Additionally, persistent cloud cover during the growing season poses a significant obstacle to obtaining clear Sentinel-2 imagery. Cloud masking techniques are essential but inevitably lead to data gaps. This issue is particularly acute in tropical and subtropical regions where cloud cover can be extensive. The resulting incomplete temporal coverage limits the ability to monitor crop development comprehensively throughout the season, which is critical for accurate nutrient modelling.

The limited size of the datasets used in this study also raises concerns about the robustness of the RF models. Inadequate dataset sizes can lead to overfitting, where the model performs exceptionally well on training data but fails to generalize to new, unseen data (Decuyper et al., 2019). This is evidenced by the disparity in  $R^2$  values between the training and test sets. Overfitting is a common problem in machine learning that can be mitigated by increasing the dataset size, thereby providing the model with a more representative sample of the variability in crop nutrient composition (Shalev-Shwartz & Ben-David, 2014).

The handling of missing data through mean imputation, while necessary given the small sample size, may introduce biases. Mean imputation assumes that missing values are random and that the mean is an adequate substitute, which is often not the case in complex systems like agriculture. This method can mask the true variability and relationships in the data, leading to potential misinterpretations and model inaccuracies. Other

machine and deep learning imputation techniques or strategies to mitigate data loss should be considered in future research to ensure the integrity of the dataset.

Another limitation of the study is the reliance on monthly composites. This approach presents a challenge when applying and comparing the results across different years. The root of this issue lies in the inherent variability of crop planting schedules. Crop planting does not consistently occur in the same month each year due to factors such as climatic variations and socio-economic influences. Therefore, the use of monthly composites could lead to a misrepresentation of the actual growth stage of the crops, resulting in potential inaccuracies in the predictions across different years or seasons. This problem will be solved when crop seasonal composites are used.

## 6. CONCLUSION AND RECOMMENDATIONS

### 6.1. Conclusion

The study used Random forest machine learning algorithm with satellite and other geospatial data to develop predictive models for five crops, barley, maize, sorghum, teff and wheat and 9 nutrients, Ca, Cu, Fe, Mn, Mg, Zn, P, K, and S.

The results showed that Sentinel-2 Vegetation Indices (VIs), a month before, and at the start of the season are the most effective in predicting nutrient composition in crop grains. The most important indices particularly those involving red-edge and SWIR bands, provide detailed information on vegetation health, capture subtle shifts in plant structure and pigment content related to nutrient availability. Soil properties, including texture, pH, and organic matter content, also play an important role in nutrient uptake, with clayey and silt textured soils and specific pH levels enhancing nutrient solubility and crop grain nutrient accumulation. Additionally, climatological factors such as rainfall and temperature are pivotal, with early-season rainfall and temperature variations significantly influencing grain nutrient concentration.

ERA5 data, providing detailed land surface variables like soil water content and evapotranspiration, enhances the reliability of agricultural models. MODIS-derived FPAR and LAI indices offer insights into photosynthetic activity and nutrient dynamics, while topographic factors such as elevation and the Topographic Wetness Index (TWI) significantly impact nutrient concentrations by affecting microclimates and soil properties. Although Sentinel-1's radar imaging has advantages in all-weather monitoring, its effectiveness in nutrient prediction in this study is limited, highlighting the need for further research to optimize its use in agricultural monitoring.

The results show that it is necessary to have tailored agricultural practices for optimal crop productivity and grain nutrient quality.

### 6.2. Recommendations

The collection of crop species-specific data is indeed crucial to discern which varieties exhibit superior nutrient uptake, even under unfavourable environmental conditions is important. This information can guide the selection of crop varieties that are not only high yielding but also nutrient-rich, thereby enhancing food security and nutritional quality. It is recommended that future research endeavours focus on generating comprehensive datasets that capture the variability in nutrient composition across different crop varieties.

Incorporating basic pre-season and on-season farm management practices into future research is also essential. These practices, ranging from soil preparation techniques, sowing times, and planting of cover crops to irrigation and pest management, significantly influence soil health, crop productivity, and consequently, nutrient composition.

Also, to avoid the mixed pixel problem, machine learning and deep learning techniques can be employed to detect farm boundaries, thereby ensuring that the data accurately represents the specific farm in question. Addressing missing data is another critical aspect that warrants attention. Instead of resorting to mean imputation, which can introduce biases, advanced deep learning techniques can be employed for imputation. These techniques can model the complex relationships in the data and provide more accurate estimates for missing values.

To ensure the interpretability and comparability of results across years, it is advisable to use crop growth stage composites instead of monthly composites. Crop growth stages are more meaningful units of time as

they directly correspond to the plant's development and physiological processes. These recommended composites can be generated using crop calendars, or phenological parameters derived from Satellite imagery. This approach can also help reduce missing data caused by clouds, as it allows for composites of periods.

The use of high-resolution datasets for certain predictors, as and when they become available, can significantly enhance the precision of future models. These datasets can better match the scale of small holder farming operations, capturing the heterogeneity within and between farms.

Lastly, funding organizations such as the Food and Agriculture Organization, the United Nations, and the World Health Organization should consider supporting projects that aim to collect more representative data across different years. This investment would enable the development of robust models for predicting crop grain nutrients, leading to more spatially variable food composition tables. Such initiatives would not only advance our understanding of crop nutrient composition but also facilitate the operationalization of research findings for the benefit of farmers, policymakers, and other stakeholders in the agricultural sector.

### **6.3. Societal impact**

The finding of this study can be utilized in public health and agricultural domains.

In the public health domain, the study provides grain nutrient insights which can be used for the creation of spatially and temporally variable food compositions databases, as well as nutritional and dietary planning. Diet recommendations or other targeted approaches can be made specifically for regions in a timely manner based on the early detection of potential nutrient deficiencies. These models can help to eliminate many diseases caused by lack of nutrients in women and children. The models will also help to eliminate the problem of late detection of nutrient deficiencies which happens when individuals are visibly unhealthy. Ensuring populations have access to nutrient-dense foods can lead to overall stabilized population health.

In the agriculture domain, the predictive models used in this study can be used as a guide for crop and grain quality improvement by improving nutrient and resource management. By using the insights of these models as a guide, farmers can enhance grain quality and potentially increase market value, providing economic benefits for their basic needs. These models can enable precise nutrient application decisions, considering climate patterns like high temperatures or excessive rainfall, and allow for strategic use of fertilizers to enhance soil health. Additionally, by understanding pre-season conditions, farmers can plant cover crops that enrich soil nutrients, leading to more sustainable farming practices and increased yields. In conclusion, the predictive models from this study offer a practical guide for farmers seeking to optimize their farm management practices for enhancement of yields and increased economic returns.

### **6.4. Ethical Considerations**

According to Kumssa et al., (2015) before sampling of grain and soil was initiated, both in farmers' fields and grain stores, consent was obtained from the participating farmers. The study received formal ethical approval from the University of Nottingham's School of Sociology and Social Policy Research Ethics Committee (REC), specifically under the identifiers BIO-1718-0004 for the research conducted in Ethiopia. These REC approvals were accepted by the Directors of Research at Addis Ababa University in Ethiopia.

## LIST OF REFERENCES

---

- Alhaj Hamoud, Y., Shaghaleh, H., Sheteiwy, M., Guo, X., Elshaikh, N. A., Ullah Khan, N., Oumarou, A., & Rahim, S. F. (2019). Impact of alternative wetting and soil drying and soil clay content on the morphological and physiological traits of rice roots and their relationships to yield and nutrient use-efficiency. *Agricultural Water Management*, 223, 105706.  
<https://doi.org/10.1016/J.AGWAT.2019.105706>
- Alonso, M., Malpica, J., Agirre, A., & Student, P. (2011). *CONSEQUENCES OF THE HUGHES PHENOMENON ON SOME CLASSIFICATION TECHNIQUES*.
- Apan, A., Held, A., Phinn, S., & Markley, J. (2004). Detecting sugarcane 'orange rust' disease using EO-1 Hyperion hyperspectral imagery. *International Journal of Remote Sensing*, 25(2), 489–498.  
<https://doi.org/10.1080/01431160310001618031>
- Axelsson, A., Lindberg, E., Reese, H., & Olsson, H. (2021). Tree species classification using Sentinel-2 imagery and Bayesian inference. *International Journal of Applied Earth Observation and Geoinformation*, 100, 102318. <https://doi.org/10.1016/J.JAG.2021.102318>
- Bahrawi, J., & Elhag, M. (2018). Consideration of seasonal variations on water radiometric indices for the estimation of soil moisture content in arid environment in Saudi Arabia-. *APPLIED ECOLOGY AND ENVIRONMENTAL RESEARCH*, 17(1), 285–303.  
[https://doi.org/10.15666/aeer/1701\\_285303](https://doi.org/10.15666/aeer/1701_285303)
- Bannari, A., Morin, D., Bonn, F., & Huete, A. R. (1995). A review of vegetation indices. *Remote Sensing Reviews*, 13(1–2), 95–120. <https://doi.org/10.1080/02757259509532298>
- Belay, A., Gashu, D., Joy, E. J. M., Lark, M. R., Chagumaira, C., Zerfu, D., Ander, L. E., Young, S. D., Bailey, E. H., & Broadley, M. R. (2022). Mineral micronutrient status and spatial distribution among the Ethiopian population. *The British Journal of Nutrition*, 128(11), 2170.  
<https://doi.org/10.1017/S0007114522000319>
- Belgiu, M., & Drăgu, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31.  
<https://doi.org/10.1016/J.ISPRSJPRS.2016.01.011>
- Belgiu, M., Marshall, M., Boschetti, M., Pepe, M., Stein, A., & Nelson, A. (2023). PRISMA and Sentinel-2 spectral response to the nutrient composition of grains. *Remote Sensing of Environment*, 292, 113567.  
<https://doi.org/10.1016/J.RSE.2023.113567>
- Bender, R. R., Haegele, J. W., & Below, F. E. (2015). Nutrient Uptake, Partitioning, and Remobilization in Modern Soybean Varieties. *Agronomy Journal*, 107(2), 563–573.  
<https://doi.org/10.2134/AGRONJ14.0435>
- Benjamin, L. R. (2017). Growth Analysis, Crops. *Encyclopedia of Applied Plant Sciences*, 3, 23–28.  
<https://doi.org/10.1016/B978-0-12-394807-6.00225-2>
- Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors* 2021, Vol. 21, Page 3758, 21(11), 3758.  
<https://doi.org/10.3390/S21113758>
- Berbercea, A., Radulov, I., Sala, F., Crista, F., & Lato, A. (2011). INTERRELATION BETWEEN METAL AVAILABILITY, SOIL pH AND MINERAL FERTILIZATION the following fertilization variants: N 0 P 0 K 0 , N 100 P 0 K 0 , N 200 P 0 K 0. *Research Journal of Agricultural Science*, 43(3).
- Berhanu, Z., Yifru, M., Geleta, D. S., & Liben, F. (2023). Comparison of Teff (*Eragrostis teff* (Zucc.) Trotter) Varieties in Response to Blended NPSZnB Fertilizer On-Farm and on Research Station. *Advances in Agriculture*, 2023(1), 2991419. <https://doi.org/10.1155/2023/2991419>

- Bid, S., & Siddique, G. (2019). Identification of seasonal variation of water turbidity using NDTI method in Panchet Hill Dam, India. *Modeling Earth Systems and Environment*, 5(4), 1179–1200. <https://doi.org/10.1007/S40808-019-00609-8/FIGURES/15>
- Bochenek, Z., Ziolkowski, D., Bartold, M., Orłowska, K., & Ochtyra, A. (2018). Monitoring forest biodiversity and the impact of climate on forest environment using high-resolution satellite images. *European Journal of Remote Sensing*, 51(1), 166–181. <https://doi.org/10.1080/22797254.2017.1414573>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324/METRICS>
- Castaldi, F., Halil Koparan, M., Wetterlind, J., Žydelis, R., Vinci, I., Özge Savaş, A., Kıvrak, C., Tunçay, T., Volungevičius, J., Obber, S., Ragazzi, F., Malo, D., & Vaudour, E. (2023). Assessing the capability of Sentinel-2 time-series to estimate soil organic carbon and clay content at local scale in croplands. *ISPRS Journal of Photogrammetry and Remote Sensing*, 199, 40–60. <https://doi.org/10.1016/J.ISPRSJPRS.2023.03.016>
- Cavagnaro, T. R. (2016). Soil moisture legacy effects: Impacts on soil nutrients, plants and mycorrhizal responsiveness. *Soil Biology and Biochemistry*, 95, 173–179. <https://doi.org/10.1016/J.SOILBIO.2015.12.016>
- Chen, Y., Zhang, S., Li, H., & Wang, Y. (2022). Drivers of nutrient content and spatial variability of soil multifunctionality in the topsoil of Kyrgyzstan. *Frontiers in Environmental Science*, 10, 1001984. <https://doi.org/10.3389/FENVS.2022.1001984/BIBTEX>
- Colle, C., Madoz-Escande, C., & Leclerc, E. (2009). Foliar transfer into the biosphere: review of translocation factors to cereal grains. *Journal of Environmental Radioactivity*, 100(9), 683–689. <https://doi.org/10.1016/J.JENVRAD.2008.10.002>
- Congalton, R. G. (2001). Accuracy assessment and validation of remotely sensed and other spatial information. *International Journal of Wildland Fire*, 10(3–4), 321–328. <https://doi.org/10.1071/WF01031>
- Darvishzadeh, R., Skidmore, A., Abdullah, H., Cherenet, E., Ali, A., Wang, T., Nieuwenhuis, W., Heurich, M., Vrieling, A., O’connor, B., & Paganini, M. (2019). Mapping leaf chlorophyll content from Sentinel-2 and RapidEye data in spruce stands using the invertible forest reflectance model. *Int J Appl Earth Obs Geoinformation*, 79, 58–70. <https://doi.org/10.1016/j.jag.2019.03.003>
- Dash, J., & Curran, P. J. (2004). The MERIS terrestrial chlorophyll index. *International Journal of Remote Sensing*, 25(23), 5403–5413. <https://doi.org/10.1080/0143116042000274015>
- Dash, J., & Curran, P. J. (2007). Evaluation of the MERIS terrestrial chlorophyll index (MTCI). *Advances in Space Research*, 39(1), 100–104. <https://doi.org/10.1016/J.ASR.2006.02.034>
- Decuyper, M., Stockhoff, M., Vandenberghe, S., -, al, & Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, 1168(2), 022022. <https://doi.org/10.1088/1742-6596/1168/2/022022>
- Dias, A. S., Hughes, P. N., Toll, D. G., & Glendinning, S. (2023). A simple method to determine soil–water retention curves of compacted active clays. *Transportation Geotechnics*, 43, 101138. <https://doi.org/10.1016/J.TRGEO.2023.101138>
- Dixon, A. B. (2018). Indigenous management of wetlands: Experiences in Ethiopia. *Indigenous Management of Wetlands: Experiences in Ethiopia*, 1–260. <https://doi.org/10.4324/9781315182018>
- Doane, T. A., Silva, L. C. R., & Horwath, W. R. (2019). Exposure to Light Elicits a Spectrum of Chemical Changes in Soil. *Journal of Geophysical Research: Earth Surface*, 124(8), 2288–2310. <https://doi.org/10.1029/2019JF005069>



- dos Santos, E. P., Da Silva, D. D., & do Amaral, C. H. (2021). Vegetation cover monitoring in tropical regions using SAR-C dual-polarization index: seasonal and spatial influences. *International Journal of Remote Sensing*, 42(19), 7581–7609. <https://doi.org/10.1080/01431161.2021.1959955>
- Draper, N. R., & Smith, H. (2014). Applied regression analysis. *Applied Regression Analysis*, 1–716. <https://doi.org/10.1002/9781118625590>
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., & Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 120, 25–36. <https://doi.org/10.1016/J.RSE.2011.11.026>
- Eleveld, B. & Bekkema, M. E., & Eleveld, M. (2018). *Mapping Grassland Management Intensity Using Sentinel-2 Satellite Data*. [https://doi.org/10.1553/giscience2018\\_01\\_s194](https://doi.org/10.1553/giscience2018_01_s194)
- Ethiopia Climate and Agriculture*. (2016). <https://www.prepdata.org/stories/ethiopia-climate-and-agriculture>
- Fernandes, M. H. M. da R., Fernandes Junior, J. de S., Adams, J. M., Lee, M., Reis, R. A., & Tedeschi, L. O. (2024). Using sentinel-2 satellite images and machine learning algorithms to predict tropical pasture forage mass, crude protein, and fiber content. *Scientific Reports 2024 14:1*, 14(1), 1–14. <https://doi.org/10.1038/s41598-024-59160-x>
- Fitter, A., & Hay, R. (2002). Toxicity. *Environmental Physiology of Plants*, 241–284. <https://doi.org/10.1016/B978-0-08-054981-1.50011-0>
- Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Source: The Annals of Statistics*, 29(5), 1189–1232. <http://www.jstor.orgURL:http://www.jstor.org/stable/2699986>
- Frison, P. L., Fruneau, B., Kmiha, S., Soudani, K., Dufrêne, E., Le Toan, T., Koleck, T., Villard, L., Mougin, E., & Rudant, J. P. (2018). Potential of Sentinel-1 Data for Monitoring Temperate Mixed Forest Phenology. *Remote Sensing 2018, Vol. 10, Page 2049*, 10(12), 2049. <https://doi.org/10.3390/RS10122049>
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data 2015 2:1*, 2(1), 1–21. <https://doi.org/10.1038/sdata.2015.66>
- Gao, B. C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3), 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
- Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31(14), 2225–2236. <https://doi.org/10.1016/J.PATREC.2010.03.014>
- Gibson, S. R. (2005). *Principles of Nutritional Assessment - Rosalind S. Gibson - Google Books*. [https://books.google.nl/books?hl=en&lr=&id=IBlu7UKI3aQC&oi=fnd&pg=PR11&ots=R YMBSV5rlz&sig=yBs3gGfeVDJoz7-jTVUUNgJnEqA&redir\\_esc=y#v=onepage&q&f=false](https://books.google.nl/books?hl=en&lr=&id=IBlu7UKI3aQC&oi=fnd&pg=PR11&ots=R YMBSV5rlz&sig=yBs3gGfeVDJoz7-jTVUUNgJnEqA&redir_esc=y#v=onepage&q&f=false)
- Giri, A., Heckathorn, S., Mishra, S., & Krause, C. (2017). Heat Stress Decreases Levels of Nutrient-Uptake and -Assimilation Proteins in Tomato Roots. *Plants*, 6(1), 443–448. <https://doi.org/10.3390/PLANTS6010006>
- Gitelson, A. A. (2004). Wide Dynamic Range Vegetation Index for Remote Quantification of Biophysical Characteristics of Vegetation. *Journal of Plant Physiology*, 161(2), 165–173. <https://doi.org/10.1078/0176-1617-01176>
- Gitelson, A. A., Chivkunova, O. B., & Merzlyak, M. N. (2009). Nondestructive estimation of anthocyanins and chlorophylls in anthocyanic leaves. *American Journal of Botany*, 96(10), 1861–1868. <https://doi.org/10.3732/ajb.0800395>

- Gitelson, A. A., Gritz, Y., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, *160*(3), 271–282. <https://doi.org/10.1078/0176-1617-00887>
- Gitelson, A. A., & Merzlyak, M. N. (1998). Remote sensing of chlorophyll concentration in higher plant leaves. *Advances in Space Research*, *22*(5), 689–692. [https://doi.org/10.1016/S0273-1177\(97\)01133-2](https://doi.org/10.1016/S0273-1177(97)01133-2)
- Gitelson, A., & Merzlyak, M. N. (1994). Spectral Reflectance Changes Associated with Autumn Senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. *Journal of Plant Physiology*, *143*(3), 286–292. [https://doi.org/10.1016/S0176-1617\(11\)81633-0](https://doi.org/10.1016/S0176-1617(11)81633-0)
- Gombart, A. F., Pierre, A., & Maggini, S. (2020). A Review of Micronutrients and the Immune System—Working in Harmony to Reduce the Risk of Infection. *Nutrients*, *12*(1). <https://doi.org/10.3390/NU12010236>
- Good, A. G., & Beatty, P. H. (2011). Fertilizing Nature: A Tragedy of Excess in the Commons. *PLOS Biology*, *9*(8), e1001124. <https://doi.org/10.1371/JOURNAL.PBIO.1001124>
- Grzebisz, W., Zielewicz, W., & Przygocka-Cyna, K. (2022). Deficiencies of Secondary Nutrients in Crop Plants—A Real Challenge to Improve Nitrogen Management. *Agronomy* *2023*, *Vol. 13*, Page 66, *13*(1), 66. <https://doi.org/10.3390/AGRONOMY13010066>
- Gupta, A., Singh, U. B., Sahu, P. K., Paul, S., Kumar, A., Malviya, D., Singh, S., Kuppusamy, P., Singh, P., Paul, D., Rai, J. P., Singh, H. V., Manna, M. C., Crusberg, T. C., Kumar, A., & Saxena, A. K. (2022). Linking Soil Microbial Diversity to Modern Agriculture Practices: A Review. *International Journal of Environmental Research and Public Health*, *19*(5), 3141. <https://doi.org/10.3390/IJERPH19053141>
- Haileslassie, A., Mekuria, W., Schmitter, P., Uhlenbrook, S., & Ludi, E. (2020). Changing Agricultural Landscapes in Ethiopia: Examining Application of Adaptive Management Approach. *Sustainability* *2020*, *Vol. 12*, Page 8939, *12*(21), 8939. <https://doi.org/10.3390/SU12218939>
- Halli, H. M., Angadi, S., Angadi, S. S., & Patil, R. H. (2016). Water and nutrient use efficiency in agriculture and the role of cereals-A review. *J. Farm Sci*, *29*(3), 299–306. <https://www.researchgate.net/publication/341607011>
- Headey, D., Dereje, M., & Taffesse, A. S. (2014). Land constraints and agricultural intensification in Ethiopia: A village-level analysis of high-potential areas. *Food Policy*, *48*, 129–141. <https://doi.org/10.1016/J.FOODPOL.2014.01.008>
- Hengl, T., De Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748. <https://doi.org/10.1371/JOURNAL.PONE.0169748>
- Hengl, T., Mendes De Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., Macmillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). *SoilGrids250m: Global Gridded Soil Information Based on Machine Learning*.
- Hird, J. N., DeLancey, E. R., McDermid, G. J., & Kariyeva, J. (2017). Google Earth Engine, Open-Access Satellite Data, and Machine Learning in Support of Large-Area Probabilistic Wetland Mapping. *Remote Sensing* *2017*, *Vol. 9*, Page 1315, *9*(12), 1315. <https://doi.org/10.3390/RS9121315>
- Hollstein, A., Segl, K., Guanter, L., Brell, M., & Enesco, M. (2016). Ready-to-Use Methods for the Detection of Clouds, Cirrus, Snow, Shadow, Water and Clear Sky Pixels in Sentinel-2 MSI Images. *Remote Sensing* *2016*, *Vol. 8*, Page 666, *8*(8), 666. <https://doi.org/10.3390/RS8080666>

- Holtgrave, A. K., Röder, N., Ackermann, A., Erasmi, S., & Kleinschmit, B. (2020). Comparing Sentinel-1 and -2 Data and Indices for Agricultural Land Use Monitoring. *Remote Sensing* 2020, Vol. 12, Page 2919, 12(18), 2919. <https://doi.org/10.3390/RS12182919>
- Hossein Khoshgofarmanesh, A., Schulin, R., Chaney, R. L., Daneshbakhsh, B., & Afyuni, M. (2010). Micronutrient-efficient genotypes for crop yield and nutritional quality in sustainable agriculture. A review Micronutrient-efficient genotypes for crop yield and nutritional quality in sustainable agriculture. A review. *Agronomy for Sustainable Development* Micronutrient-efficient genotypes for crop yield and nutritional quality in sustainable agriculture. A review. *Agron. Sustain. Dev.*, 30(1), 83–107. <https://doi.org/10.1051/agro/2009017i>
- Hsu, H., & Dirmeyer, P. A. (2023). Soil moisture-evaporation coupling shifts into new gears under increasing CO<sub>2</sub>. *Nature Communications* 2023 14:1, 14(1), 1–9. <https://doi.org/10.1038/s41467-023-36794-5>
- Huang, H., Liu, L., & Ngadi, M. O. (2014). Recent Developments in Hyperspectral Imaging for Assessment of Food Quality and Safety. *Sensors* 2014, Vol. 14, Pages 7248-7276, 14(4), 7248–7276. <https://doi.org/10.3390/S140407248>
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Huffman, J., Drouin, P., Renaud, J. B., Dunière, L., & LaPointe, G. (2023). Farm management practices and season dependent factors affect the microbial community and chemical profile of corn and grass-legume silages of farms in Ontario, Québec, and Northern New York. *Frontiers in Microbiology*, 14, 1214915. <https://doi.org/10.3389/FMICB.2023.1214915/BIBTEX>
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/J.IJFORECAST.2006.03.001>
- Indoria, A. K., Sharma, K. L., & Reddy, K. S. (2020). Hydraulic properties of soil under warming climate. *Climate Change and Soil Interactions*, 473–508. <https://doi.org/10.1016/B978-0-12-818032-7.00018-7>
- Jarrell, W. M., & Beverly, R. B. (1981). The Dilution Effect in Plant Nutrition Studies. *Advances in Agronomy*, 34(C), 197–224. [https://doi.org/10.1016/S0065-2113\(08\)60887-1](https://doi.org/10.1016/S0065-2113(08)60887-1)
- Jena, R. K., Bandyopadhyay, S., Pradhan, U. K., Moharana, P. C., Kumar, N., Sharma, G. K., Roy, P. D., Ghosh, D., Ray, P., Padua, S., Ramachandran, S., Das, B., Singh, S. K., Ray, S. K., Alsuhaibani, A. M., Gaber, A., & Hossain, A. (2022). Geospatial Modelling for Delineation of Crop Management Zones Using Local Terrain Attributes and Soil Properties. *Remote Sensing* 2022, Vol. 14, Page 2101, 14(9), 2101. <https://doi.org/10.3390/RS14092101>
- Jensen, J. R. (2009). *Remote Sensing of the Environment: An Earth Resource Perspective*, 2nd Ed. Prentice Hall, Upper Saddle River, NJ, 1–592.
- Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., Timlin, D. J., Shim, K. M., Gerber, J. S., Reddy, V. R., & Kim, S. H. (2016). Random Forests for Global and Regional Crop Yield Predictions. *PLOS ONE*, 11(6), e0156571. <https://doi.org/10.1371/JOURNAL.PONE.0156571>
- Jian, Z., Lei, L., Ni, Y., Xu, J., Xiao, W., & Zeng, L. (2022). Soil clay is a key factor affecting soil phosphorus availability in the distribution area of Masson pine plantations across subtropical China. *Ecological Indicators*, 144, 109482. <https://doi.org/10.1016/J.ECOLIND.2022.109482>
- Jiang, H., Liu, Z., Wang, J., Yang, P., Zhang, R., Zhang, X., & Zheng, P. (2023). Combining Chlorophyll Fluorescence and Vegetation Reflectance Indices to Estimate Non-Photochemical Quenching (NPQ) of Rice at the Leaf Scale. *Remote Sensing*, 15(17), 4222. <https://doi.org/10.3390/RS15174222/S1>

- Karaca, S., & Gülser, F. (2018). Relationships between soil properties, topography and land use in the Van Lake Basin, Turkey. *Eurasian J Soil Sci*, 7(2), 115–120. <https://doi.org/10.18393/ejss.348412>
- Kaufman, Y. J., & Tanré, D. (1992). Atmospherically Resistant Vegetation Index (ARVI) for EOS-MODIS. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 261–270. <https://doi.org/10.1109/36.134076>
- Kganyago, M., Adjorlolo, C., Mhangara, P., & Tsoeleng, L. (2024). Optical remote sensing of crop biophysical and biochemical parameters: An overview of advances in sensor technologies and machine learning algorithms for precision agriculture. *Computers and Electronics in Agriculture*, 218, 108730. <https://doi.org/10.1016/J.COMPAG.2024.108730>
- Kganyago, M., Mhangara, P., & Adjorlolo, C. (2021). Estimating Crop Biophysical Parameters Using Machine Learning Algorithms and Sentinel-2 Imagery. *Remote Sensing 2021, Vol. 13, Page 4314*, 13(21), 4314. <https://doi.org/10.3390/RS13214314>
- Khabbazan, S., Vermunt, P., Steele-Dunne, S., Arntz, L. R., Marinetti, C., van der Valk, D., Iannini, L., Molijn, R., Westerdijk, K., & van der Sande, C. (2019). Crop Monitoring Using Sentinel-1 Data: A Case Study from The Netherlands. *Remote Sensing 2019, Vol. 11, Page 1887*, 11(16), 1887. <https://doi.org/10.3390/RS11161887>
- Khan, H. R., Gillani, Z., Jamal, M. H., Athar, A., Chaudhry, M. T., Chao, H., He, Y., & Chen, M. (2023). Early Identification of Crop Type for Smallholder Farming Systems Using Deep Learning on Time-Series Sentinel-2 Imagery. *Sensors 2023, Vol. 23, Page 1779*, 23(4), 1779. <https://doi.org/10.3390/S23041779>
- Khmelevtsova, L. E., Sazykin, I. S., Azhogina, T. N., & Sazykina, M. A. (2022). Influence of Agricultural Practices on Bacterial Community of Cultivated Soils. *Agriculture 2022, Vol. 12, Page 371*, 12(3), 371. <https://doi.org/10.3390/AGRICULTURE12030371>
- Khurana, S., Heße, F., Hildebrandt, A., & Thullner, M. (2022). Predicting the impact of spatial heterogeneity on microbially mediated nutrient cycling in the subsurface. *Biogeosciences*, 19(3), 665–688. <https://doi.org/10.5194/BG-19-665-2022>
- Kopecký, M., Macek, M., & Wild, J. (2021). Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition. *Science of The Total Environment*, 757, 143785. <https://doi.org/10.1016/J.SCITOTENV.2020.143785>
- Korobov, M., & Lopuhin, K. (2017). *Permutation Importance — ELI5 0.11.0 documentation*. [https://eli5.readthedocs.io/en/latest/blackbox/permutation\\_importance.html](https://eli5.readthedocs.io/en/latest/blackbox/permutation_importance.html)
- Kumssa, D. B., Joy, E. J. M., Ander, E. L., Watts, M. J., Young, S. D., Walker, S., & Broadley, M. R. (2015). Dietary calcium and zinc deficiency risks are decreasing but remain prevalent. *Scientific Reports*, 5. <https://doi.org/10.1038/SREP10974>
- Kumssa, D. B., Mossa, A. W., Amede, T., Ander, E. L., Bailey, E. H., Botoman, L., Chagumaira, C., Chimungu, J. G., Davis, K., Gameda, S., Haefele, S. M., Hailu, K., Joy, E. J. M., Lark, R. M., Ligowe, I. S., McGrath, S. P., Milne, A., Muleya, P., Munthali, M., ... Nalivata, P. C. (2022). Cereal grain mineral micronutrient and soil chemistry data from GeoNutrition surveys in Ethiopia and Malawi. *Scientific Data 2022 9:1*, 9(1), 1–12. <https://doi.org/10.1038/s41597-022-01500-5>
- Kuyu, C. G., Abebe, A. H., Bereka, T. Y., Abdissa, Z. K., & Bekere, Y. B. (2024). Nutritional and Microbial Quality of teff Grain as Influenced by Economically Motivated Adulteration Along the Supply Chain. *Journal of Food Protection*, 87(2), 100216. <https://doi.org/10.1016/J.JFP.2023.100216>
- Li, C., Wang, X., & Qin, M. (2021). Spatial variability of soil nutrients in seasonal rivers: A case study from the Guo River Basin, China. *PLoS ONE*, 16(3). <https://doi.org/10.1371/JOURNAL.PONE.0248655>

- Li, Z., Li, X., Wei, D., Xu, X., & Wang, H. (2010). An assessment of correlation on MODIS-NDVI and EVI with natural vegetation coverage in Northern Hebei Province, China. *Procedia Environmental Sciences*, 2, 964–969. <https://doi.org/10.1016/J.PROENV.2010.10.108>
- Lightfoot, D. A. (2018). Developing crop varieties with improved nutrient-use efficiency. *Engineering Nitrogen Utilization in Crop Plants*, 1–11. [https://doi.org/10.1007/978-3-319-92958-3\\_1](https://doi.org/10.1007/978-3-319-92958-3_1)
- Liu, J., Miller, J. R., Haboudane, D., & Pattey, E. (2004). Exploring the relationship between red edge parameters and crop variables for precision agriculture. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2, 1276–1279. <https://doi.org/10.1109/IGARSS.2004.1368649>
- Maillard, A., Diquélou, S., Billard, V., Lainé, P., Garnica, M., Prudent, M., Garcia-Mina, J. M., Yvin, J. C., & Ourry, A. (2015). Leaf mineral nutrient remobilization during leaf senescence and modulation by nutrient deficiency. *Frontiers in Plant Science*, 6(MAY), 1–15. <https://doi.org/10.3389/FPLS.2015.00317/ABSTRACT>
- Malakouti, S. M., Menhaj, M. B., & Suratgar, A. A. (2023). The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction. *Cleaner Engineering and Technology*, 15, 100664. <https://doi.org/10.1016/J.CLET.2023.100664>
- Mandal, D., Vaka, D. S., Bhogapurapu, N. R., Vanama, V. S. K., Kumar, V., Rao, Y. S., & Bhattacharya, A. (2019). *Sentinel-1 SLC Preprocessing Workflow for Polarimetric Applications: A Generic Practice for Generating Dual-pol Covariance Matrix Elements in SNAP S-1 Toolbox*. <https://doi.org/10.20944/PREPRINTS201911.0393.V1>
- Matsushita, B., Yang, W., Chen, J., Onda, Y., & Qiu, G. (2007). Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-density Cypress Forest. *Sensors 2007, Vol. 7, Pages 2636-2651*, 7(11), 2636–2651. <https://doi.org/10.3390/S7112636>
- Matus, F. J. (2021). Fine silt and clay content is the main factor defining maximal C and N accumulations in soils: a meta-analysis. *Scientific Reports*, 11(1), 6438. <https://doi.org/10.1038/S41598-021-84821-6>
- McKay Fletcher, D., Ruiz, S., Williams, K., Petroselli, C., Walker, N., Chadwick, D., Jones, D. L., & Roose, T. (2022). Projected Increases in Precipitation Are Expected to Reduce Nitrogen Use Efficiency and Alter Optimal Fertilization Timings in Agriculture in the South East of England. *ACS ES and T Engineering*, 2(8), 1414–1424. [https://doi.org/10.1021/ACSESTENGG.1C00492/ASSET/IMAGES/LARGE/EE1C00492\\_0006.JPEG](https://doi.org/10.1021/ACSESTENGG.1C00492/ASSET/IMAGES/LARGE/EE1C00492_0006.JPEG)
- Mendes, M. B., Mota Barroso, J., & Rato, A. E. (2023). *The use of Sentinel 2 to quantify N, Ca, and K in walnut orchards*. <https://doi.org/10.21203/rs.3.rs-2603160/v1>
- Merzlyak, M. N., Gitelson, A. A., Chivkunova, O. B., & Rakitin, V. Y. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum*, 106(1), 135–141. <https://doi.org/10.1034/J.1399-3054.1999.106119.X>
- Messant, M., Hennebelle, T., Guérard, F., Gakière, B., Gall, A., Thomine, S., & Krieger-Liszky, A. (2022). Manganese excess and deficiency affects photosynthesis and metabolism in *Marchantia polymorpha*. *BioRxiv*, 2022.01.24.477552. <https://doi.org/10.1101/2022.01.24.477552>
- Mir, R. R., Rustgi, S., Zhang, Y. M., & Xu, C. (2022). Multi-faceted approaches for breeding nutrient-dense, disease-resistant, and climate-resilient crop varieties for food and nutritional security. *Heredity* 2022 128:6, 128(6), 387–390. <https://doi.org/10.1038/s41437-022-00542-0>
- Mukul, M., Srivastava, V., Jade, S., & Mukul, M. (2017). Uncertainties in the Shuttle Radar Topography Mission (SRTM) Heights: Insights from the Indian Himalaya and Peninsula. *Scientific Reports 2017 7:1*, 7(1), 1–10. <https://doi.org/10.1038/srep41672>

- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., & Thépaut, J. N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, *13*(9), 4349–4383. <https://doi.org/10.5194/ESSD-13-4349-2021>
- Nasirzadehdizaji, R., Sanli, F. B., Abdikan, S., Cakir, Z., Sekertekin, A., & Ustuner, M. (2019). Sensitivity Analysis of Multi-Temporal Sentinel-1 SAR Parameters to Crop Height and Canopy Coverage. *Applied Sciences* 2019, Vol. 9, Page 655, *9*(4), 655. <https://doi.org/10.3390/APP9040655>
- Ngo, H. T. T., & Cavagnaro, T. R. (2018). Interactive effects of compost and pre-planting soil moisture on plant biomass, nutrition and formation of mycorrhizas: a context dependent response. *Scientific Reports* 2018 *8*:1, *8*(1), 1–9. <https://doi.org/10.1038/s41598-017-18780-2>
- Nieto, L., Schwalbert, R., Prasad, P. V. V., Olson, B. J. S. C., & Ciampitti, I. A. (2021). An integrated approach of field, weather, and satellite data for monitoring maize phenology. *Scientific Reports* 2021 *11*:1, *11*(1), 1–10. <https://doi.org/10.1038/s41598-021-95253-7>
- Ocké, M. C., Westenbrink, S., van Rossum, C. T. M., Temme, E. H. M., van der Vossen-Wijmenga, W., & Verkaik-Kloosterman, J. (2021). The essential role of food composition databases for public health nutrition – Experiences from the Netherlands. *Journal of Food Composition and Analysis*, *101*, 103967. <https://doi.org/10.1016/J.JFCA.2021.103967>
- Pedregosa FABIANPEDREGOSA, F., Michel, V., Grisel OLIVIERGRISEL, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot and Édouardand, M., Duchesnay, and Édouard, & Duchesnay EDOUARDDUCHESNAY, Fré. (2011). Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830. <https://doi.org/10.5555/1953048.2078195>
- Philippot, L., Raaijmakers, J. M., Lemanceau, P., & Van Der Putten, W. H. (2013). Going back to the roots: the microbial ecology of the rhizosphere. *Nature Reviews Microbiology* 2013 *11*:11, *11*(11), 789–799. <https://doi.org/10.1038/nrmicro3109>
- Prasath, N., Sreemathy, J., Krishnaraj, N., & Vigneshwaran, P. (2023). Analysis of Crop Yield Prediction Using Random Forest Regression Model. *Smart Innovation, Systems and Technologies*, *324*, 239–249. [https://doi.org/10.1007/978-981-19-7447-2\\_22](https://doi.org/10.1007/978-981-19-7447-2_22)
- Priya, P., Muthaiah, U., & Balamurugan, & M. (2018). PREDICTING YIELD OF THE CROP USING MACHINE LEARNING ALGORITHM. *International Journal of Engineering Sciences & Research Technology*, *1*. <https://doi.org/10.5281/zenodo.1212821>
- Psomiadis, E., Dercas, N., Dalezios, N. R., & Spiropoulos, N. V. (2017). *Evaluation and cross-comparison of vegetation indices for crop monitoring from sentinel-2 and worldview-2 images*. 79. <https://doi.org/10.1117/12.2278217>
- Qian, B., Ye, H., Huang, W., Xie, Q., Pan, Y., Xing, N., Ren, Y., Guo, A., Jiao, Q., & Lan, Y. (2022). A sentinel-2-based triangular vegetation index for chlorophyll content estimation. *Agricultural and Forest Meteorology*, *322*, 109000. <https://doi.org/10.1016/J.AGRFORMET.2022.109000>
- Rao, P., Zhou, W., Bhattarai, N., Srivastava, A. K., Singh, B., Poonia, S., Lobell, D. B., & Jain, M. (2021). Using sentinel-1, sentinel-2, and planet imagery to map crop type of smallholder farms. *Remote Sensing*, *13*(10), 1870. <https://doi.org/10.3390/RS13101870/S1>
- Rickson, R. J. (2023). Water induced soil erosion. *Encyclopedia of Soils in the Environment, Second Edition*, 193–207. <https://doi.org/10.1016/B978-0-12-822974-3.00231-7>
- Ritchie, H. (2021). *Three billion people cannot afford a healthy diet*. Our World in Data. <https://ourworldindata.org/diet-affordability>

- Ritchie, H., & Roser, M. (2017). Micronutrient Deficiency. *Our World in Data*.  
<https://ourworldindata.org/micronutrient-deficiency>
- Rosenthal, S. (2017). Data Imputation. *The International Encyclopedia of Communication Research Methods*, 1–12.  
<https://doi.org/10.1002/9781118901731.IECRM0058>
- Scavo, A., Fontanazza, S., Restuccia, A., Pesce, G. R., Abbate, C., & Mauromicale, G. (2022). The role of cover crops in improving soil fertility and plant nutritional status in temperate climates. A review. *Agronomy for Sustainable Development* 2022 42:5, 42(5), 1–25. <https://doi.org/10.1007/S13593-022-00825-0>
- Schakel, S. F., Marilyn Buzzard, I., & Gebhardt, S. E. (1997). Procedures for Estimating Nutrient Values for Food Composition Databases. In *JOURNAL OF FOOD COMPOSITION AND ANALYSIS* (Vol. 10).
- Schmidt, F., & Persson, A. (2003). Comparison of DEM data capture and topographic wetness indices. *Precision Agriculture*, 4(2), 179–192. <https://doi.org/10.1023/A:1024509322709/METRICS>
- Shahadha, S. S., Wendroth, O., & Ding, D. (2021). Nitrogen and Rainfall Effects on Crop Growth—Experimental Results and Scenario Analyses. *Water* 2021, Vol. 13, Page 2219, 13(16), 2219.  
<https://doi.org/10.3390/W13162219>
- Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding Machine Learning: From Theory to Algorithms. *Understanding Machine Learning: From Theory to Algorithms*, 9781107057135, 1–397.  
<https://doi.org/10.1017/CBO9781107298019>
- Sharifi, A. (2020). Using Sentinel-2 Data to Predict Nitrogen Uptake in Maize Crop. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 2656–2662.  
<https://doi.org/10.1109/JSTARS.2020.2998638>
- Sharma, M., Bangotra, P., Gautam, A. S., & Gautam, S. (2022). Sensitivity of normalized difference vegetation index (NDVI) to land surface temperature, soil moisture and precipitation over district Gautam Buddh Nagar, UP, India. *Stochastic Environmental Research and Risk Assessment*, 36(6), 1779.  
<https://doi.org/10.1007/S00477-021-02066-1>
- Sievers, T., & Cook, R. L. (2018). Aboveground and Root Decomposition of Cereal Rye and Hairy Vetch Cover Crops. *Soil Science Society of America Journal*, 82(1), 147–155.  
<https://doi.org/10.2136/SSSAJ2017.05.0139>
- Sims, D. A., & Gamon, J. A. (2002). Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sensing of Environment*, 81(2–3), 337–354. [https://doi.org/10.1016/S0034-4257\(02\)00010-X](https://doi.org/10.1016/S0034-4257(02)00010-X)
- Skidmore, M., Andarge, T., & Foltz, | Jeremy. (2023). The impact of extreme precipitation on nutrient runoff. *Journal of the Agricultural and Applied Economics Association*, 2(4), 769–785.  
<https://doi.org/10.1002/JAA2.90>
- Steinhausen, M. J., Wagner, P. D., Narasimhan, B., & Waske, B. (2018). Combining Sentinel-1 and Sentinel-2 data for improved land use and land cover mapping of monsoon regions. *International Journal of Applied Earth Observation and Geoinformation*, 73, 595–604.  
<https://doi.org/10.1016/J.JAG.2018.08.011>
- Stephen, K. D., & Kazemi, A. (2014). Improved normalization of time-lapse seismic data using normalized root mean square repeatability data to improve automatic production and seismic history matching in the Nelson field. *Geophysical Prospecting*, 62(5), 1009–1027.  
<https://doi.org/10.1111/1365-2478.12109>
- Tadele, E., & Hibistu, T. (2021). Empirical review on the use dynamics and economics of teff in Ethiopia. *Agriculture and Food Security*, 10(1), 1–13. <https://doi.org/10.1186/S40066-021-00329-2/FIGURES/3>

- Tahat, M. M., Alananbeh, K. M., Othman, Y. A., & Leskovar, D. I. (2020). Soil Health and Sustainable Agriculture. *Sustainability* 2020, Vol. 12, Page 4859, 12(12), 4859. <https://doi.org/10.3390/SU12124859>
- Tan, C., Wang, D., Zhou, J., Du, Y., Luo, M., Zhang, Y., & Guo, W. (2018). Remotely assessing fraction of photosynthetically active radiation (FPAR) for wheat canopies based on hyperspectral vegetation indexes. *Frontiers in Plant Science*, 9, 378175. <https://doi.org/10.3389/FPLS.2018.00776/BIBTEX>
- Tan, Y., Sun, J. Y., Zhang, B., Chen, M., Liu, Y., & Liu, X. D. (2019). Sensitivity of a Ratio Vegetation Index Derived from Hyperspectral Remote Sensing to the Brown Planthopper Stress on Rice Plants. *Sensors (Basel, Switzerland)*, 19(2). <https://doi.org/10.3390/S19020375>
- Tayade, R., Yoon, J., Lay, L., Khan, A. L., Yoon, Y., & Kim, Y. (2022). Utilization of Spectral Indices for High-Throughput Phenotyping. *Plants*, 11(13). <https://doi.org/10.3390/PLANTS11131712>
- The State of Food Security and Nutrition in the World 2021. (2021). The State of Food Security and Nutrition in the World 2021. *The State of Food Security and Nutrition in the World 2021*. <https://doi.org/10.4060/CB4474EN>
- Thenkabail, P. S., Enclona, E. A., Ashton, M. S., & Van Der Meer, B. (2004). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment*, 91(3–4), 354–376. <https://doi.org/10.1016/J.RSE.2004.03.013>
- Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B. Ö., Floury, N., Brown, M., Traver, I. N., Deghaye, P., Duesmann, B., Rosich, B., Miranda, N., Bruno, C., L'Abbate, M., Croci, R., Pietropaolo, A., ... Rostan, F. (2012). GMES Sentinel-1 mission. *Remote Sensing of Environment*, 120, 9–24. <https://doi.org/10.1016/J.RSE.2011.05.028>
- Traka, M. H., Plumb, J., Berry, R., Pinchen, H., & Finglas, P. M. (2020). Maintaining and updating food composition datasets for multiple users and novel technologies: Current challenges from a UK perspective. *Nutrition Bulletin*, 45(2), 230–240. <https://doi.org/10.1111/NBU.12433>
- United Nations. (2015). *70/1. Transforming our world: the 2030 Agenda for Sustainable Development Transforming our world: the 2030 Agenda for Sustainable Development Preamble*.
- United Nations. (2023). *The-Sustainable-Development-Goals-Report-2023*.
- Veci, L., Prats-Iraola, P., Scheiber, R., Collard, F., Fomferra, N., & Engdahl, M. (2014). *The Sentinel-1 Toolbox*.
- Végh, K. R. (1991a). *Effect of Soil Water and Nutrient Supply on Root Characteristics and Nutrient Uptake of Plants*. 24, 143–148. <https://doi.org/10.1016/B978-0-444-89104-4.50024-4>
- Végh, K. R. (1991b). *Effect of Soil Water and Nutrient Supply on Root Characteristics and Nutrient Uptake of Plants*. 24, 143–148. <https://doi.org/10.1016/B978-0-444-89104-4.50024-4>
- Vinod, K. K. (2019). The Need for Nutrient Efficient Crop Varieties. *AgriRxiv*, 2019. <https://doi.org/10.31220/OSF.IO/T9Z3J>
- Wang, C., Xie, Y., & Tan, Z. (2024). Soil potassium depletion in global cereal croplands and its implications. *Science of The Total Environment*, 907, 167875. <https://doi.org/10.1016/J.SCITOTENV.2023.167875>
- Wang, L., Kaur, M., Zhang, P., Li, J., & Xu, M. (2021). Effect of different agricultural farming practices on microbial biomass and enzyme activities of celery growing field soil. *International Journal of Environmental Research and Public Health*, 18(23). <https://doi.org/10.3390/IJERPH182312862>
- Weiss, M., Jacob, F., & Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236, 111402. <https://doi.org/10.1016/j.rse.2019.111402>
- Weng, X., Li, H., Ren, C., Zhou, Y., Zhu, W., Zhang, S., & Liu, L. (2022). Calcium Regulates Growth and Nutrient Absorption in Poplar Seedlings. *Frontiers in Plant Science*, 13, 887098. <https://doi.org/10.3389/FPLS.2022.887098>



- Wu, C., Niu, Z., & Gao, S. (2012). The potential of the satellite derived green chlorophyll index for estimating midday light use efficiency in maize, coniferous forest and grassland. *Ecological Indicators*, 14(1), 66–73. <https://doi.org/10.1016/J.ECOLIND.2011.08.018>
- Wu, C., Niu, Z., Tang, Q., & Huang, W. (2008). Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest Meteorology*, 148(8–9), 1230–1241. <https://doi.org/10.1016/J.AGRFORMET.2008.03.005>
- Xie, K., Cakmak, I., Wang, S., Zhang, F., & Guo, S. (2021). Synergistic and antagonistic interactions between potassium and magnesium in higher plants. *The Crop Journal*, 9(2), 249–256. <https://doi.org/10.1016/J.CJ.2020.10.005>
- Xue, R., Shen, Y., & Marschner, P. (2017). Soil water content during and after plant growth influence nutrient availability and microbial biomass. *Journal of Soil Science and Plant Nutrition*, 17(3), 702–715. <https://doi.org/10.4067/S0718-95162017000300012>
- Yamazaki, D., Ikeshima, D., Neal, J. C., O’Loughlin, F., Sampson, C. C., Kanae, S., & Bates, P. D. (2017). MERIT DEM: A new high-accuracy global digital elevation model and its merit to global hydrodynamic modeling. *AGUFM*, 2017, H12C-04. <https://ui.adsabs.harvard.edu/abs/2017AGUFM.H12C..04Y/abstract>
- Yan, K., Park, T., Yan, G., Liu, Z., Yang, B., Chen, C., Nemani, R. R., Knyazikhin, Y., & Myneni, R. B. (2016). Evaluation of MODIS LAI/FPAR Product Collection 6. Part 2: Validation and Intercomparison. *Remote Sensing 2016*, Vol. 8, Page 460, 8(6), 460. <https://doi.org/10.3390/RS8060460>
- Yin, F., Lewis, P. E., & Gómez-Dans, J. L. (2022). Bayesian atmospheric correction over land: Sentinel-2/MSI and Landsat 8/OLI. *Geoscientific Model Development*, 15(21), 7933–7976. <https://doi.org/10.5194/GMD-15-7933-2022>
- Yu, K., Anderegg, J., Mikaberidze, A., Karisto, P., Mascher, F., McDonald, B. A., Walter, A., & Hund, A. (2018). Hyperspectral canopy sensing of wheat septoria tritici blotch disease. *Frontiers in Plant Science*, 9, 326021. <https://doi.org/10.3389/FPLS.2018.01195/BIBTEX>
- Zama, N., Kirkman, K., Mkhize, N., Tedder, M., & Magadla, A. (2022). Soil Acidification in Nutrient-Enriched Soils Reduces the Growth, Nutrient Concentrations, and Nitrogen-Use Efficiencies of *Vachellia sieberiana* (DC.) Kyal. & Boatwr Saplings. *Plants*, 11(24). <https://doi.org/10.3390/PLANTS11243564/S1>
- Zhang, S., Zhang, X., Huffman, T., Liu, X., & Yang, J. (2011a). Influence of topography and land management on soil nutrients variability in Northeast China. *Nutrient Cycling in Agroecosystems*, 89(3), 427–438. <https://doi.org/10.1007/S10705-010-9406-0/TABLES/6>
- Zhang, S., Zhang, X., Huffman, T., Liu, X., & Yang, J. (2011b). Influence of topography and land management on soil nutrients variability in Northeast China. *Nutrient Cycling in Agroecosystems*, 89(3), 427–438. <https://doi.org/10.1007/S10705-010-9406-0/TABLES/6>
- Zhang, W., Zhu, L., Zhuang, Q., Chen, D., & Sun, T. (2023). Mapping Cropland Soil Nutrients Contents Based on Multi-Spectral Remote Sensing and Machine Learning. *Agriculture (Switzerland)*, 13(8). <https://doi.org/10.3390/AGRICULTURE13081592>
- Zhang, Z., Liu, M., Liu, X., & Zhou, G. (2018). A new vegetation index based on multitemporal sentinel-2 images for discriminating heavy metal stress levels in rice. *Sensors (Switzerland)*, 18(7), 2172. <https://doi.org/10.3390/s18072172>



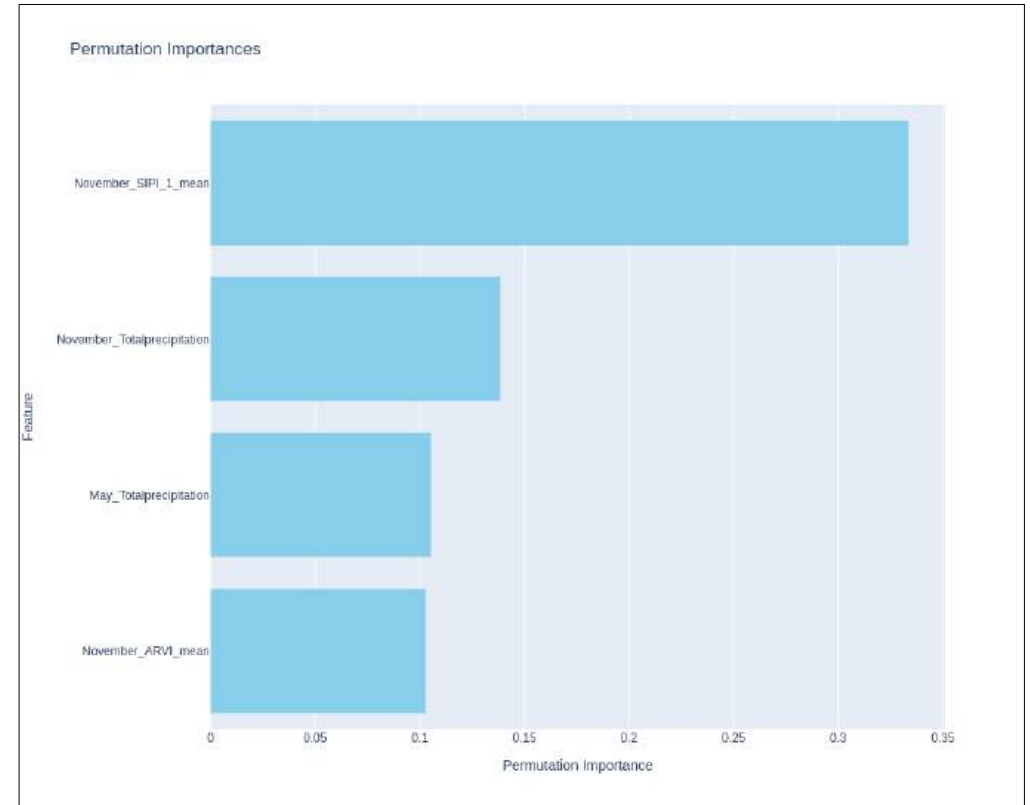
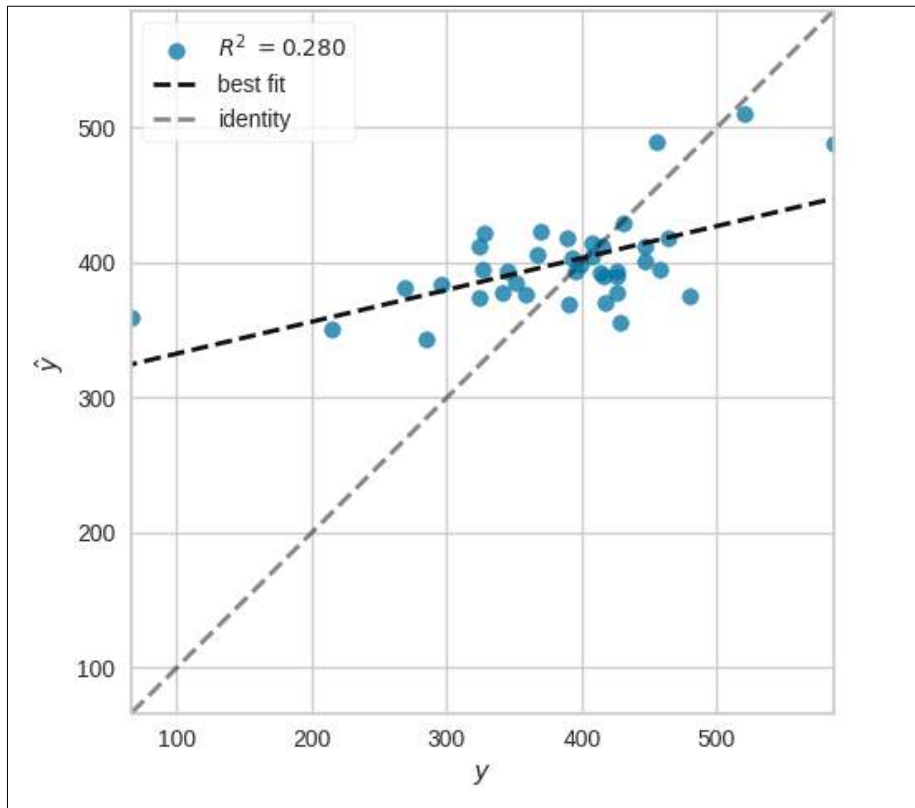
## 7. APPENDICES

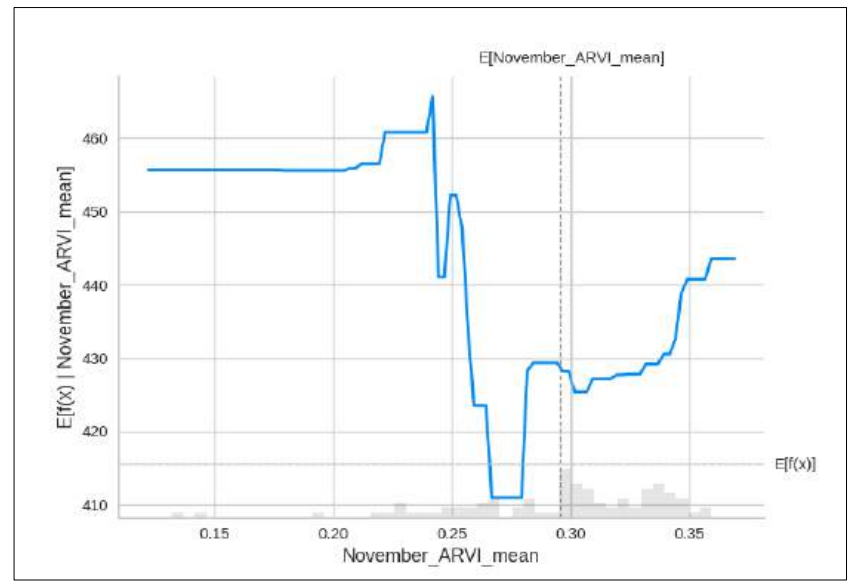
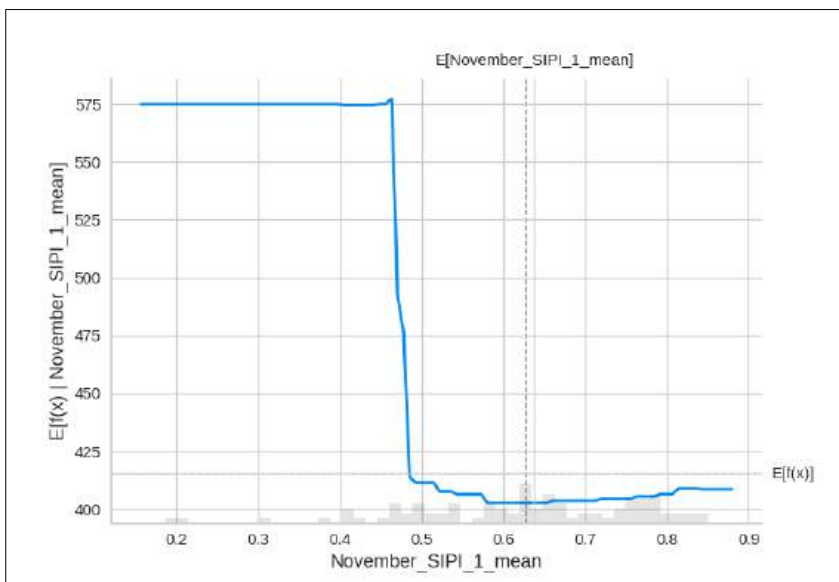
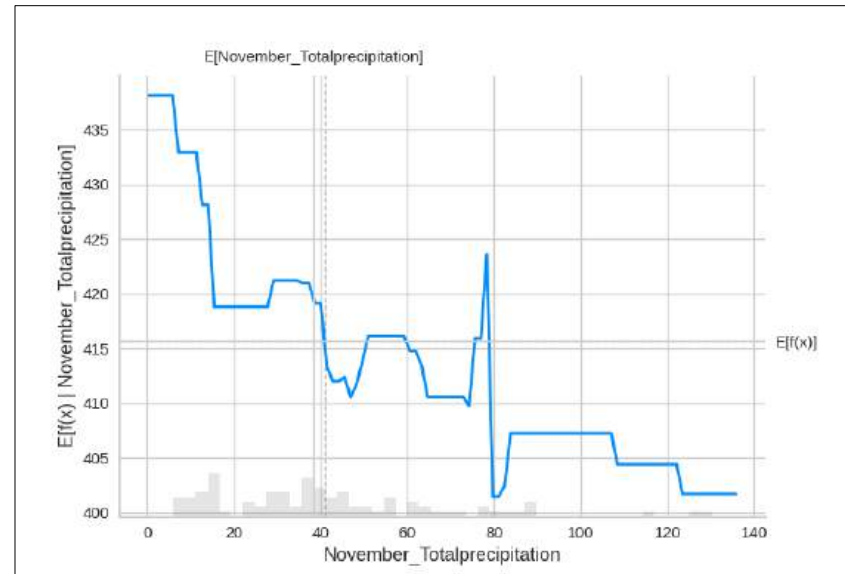
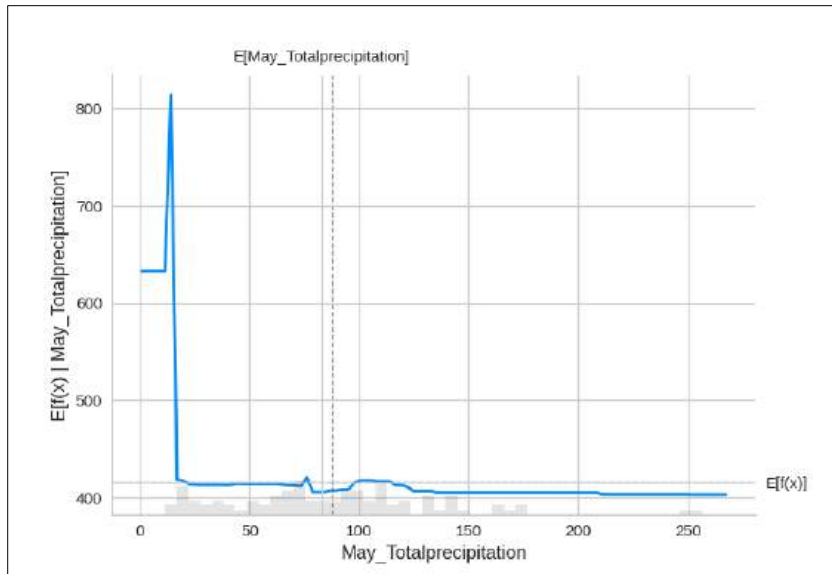
### Appendix 1. Data management

Dataset	Name of Data file	Source ( Primary or Secondary data	Owner, If Secondary	Restrictions and Licence	Data Form	Data Format	Year of Data	Contains Personal Data? Yes/No	Links
Grain nutrient composition	ETH_CropSoilChemData	Secondary	GeoNutrition project	Open data policy	Tabular	csv	2017/2018	No	<a href="https://doi.org/10.6084/m9.figshare.15911973">https://doi.org/10.6084/m9.figshare.15911973</a>
Sentinel-1	COPERNICUS/S1_GRD	Secondary	ESA/Copernicus	Open data policy	Raster Imagery	.tiff	2017/2018	No	<a href="https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S1_GRD">https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S1_GRD</a>
Sentinel-2	COPERNICUS/S2_SR	Secondary	ESA/Copernicus	Open data policy	Raster Imagery	.tiff	2017/2018	No	<a href="https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_HARMONIZED">https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_HARMONIZED</a>
Rainfall, Temperature, Environment Variables	ECMWF/ERA5/MONTHLY	Secondary	ECMWF / Copernicus Climate Change Service	Open data policy	Raster Imagery	.tiff	2017/2018	No	<a href="https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_MONTHLY">https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_MONTHLY</a>
SOIL	Soil_ISRIC	Secondary	ISRIC	Open data policy	Raster Imagery	.tiff	2016	No	<a href="https://data.isric.org/geonetwork/srv/en/g/catalog.search#/home">https://data.isric.org/geonetwork/srv/en/g/catalog.search#/home</a>
SOIL	Soil_ISDA	Secondary	Isda	Open data policy	Raster Imagery	.tiff	2017	No	<a href="https://developers.google.com/earth-engine/datasets/tags/isda">https://developers.google.com/earth-engine/datasets/tags/isda</a>
Terrain Products	MERIT DEM	Secondary	University of Tokyo	Open data policy	Raster Imagery	.tiff	2017	No	<a href="https://developers.google.com/earth-engine/datasets/catalog/MERIT_DEM_v1_0_3">https://developers.google.com/earth-engine/datasets/catalog/MERIT_DEM_v1_0_3</a>
MODIS_FPAR	MODIS_FPAR	Secondary	NASA Land Processes Distributed Active Archive Center (LP DAAC)	Open data policy	Raster Imagery	.tiff	2017/2018	No	<a href="https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD15A2H">https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MOD15A2H</a>

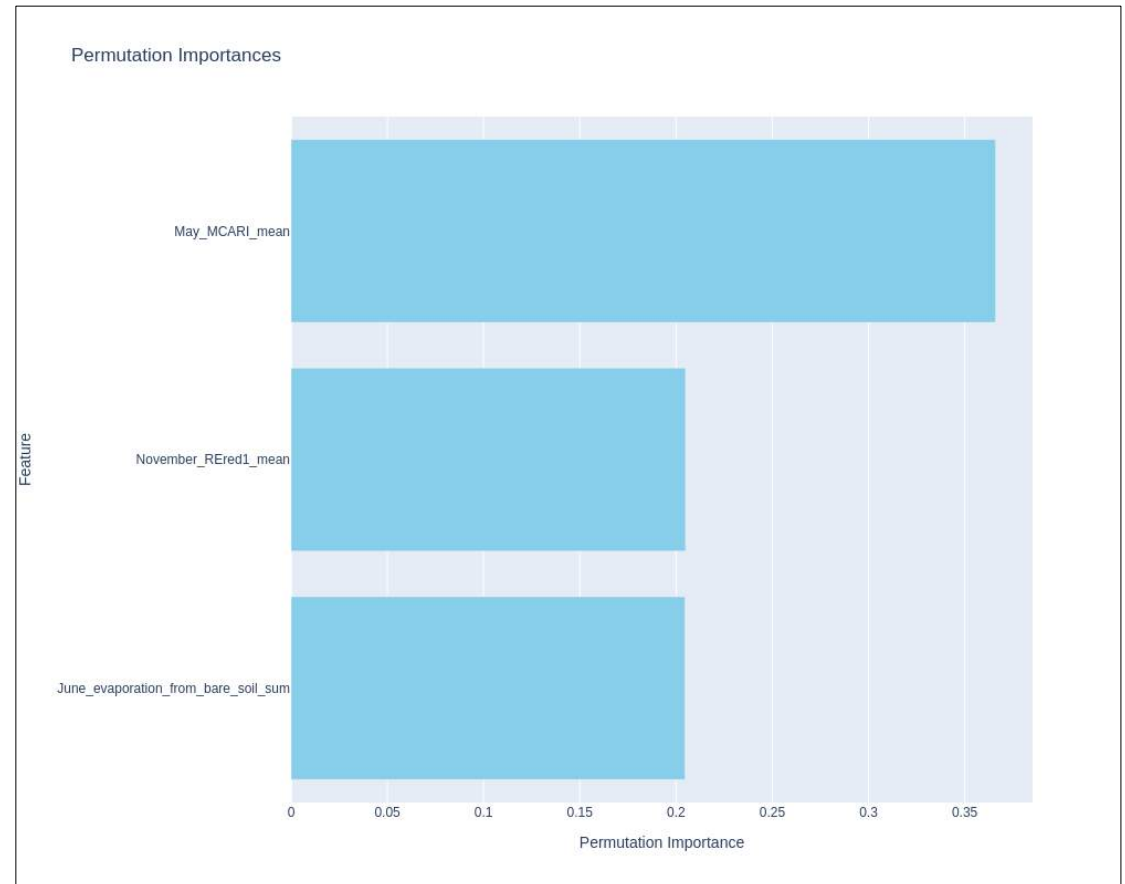
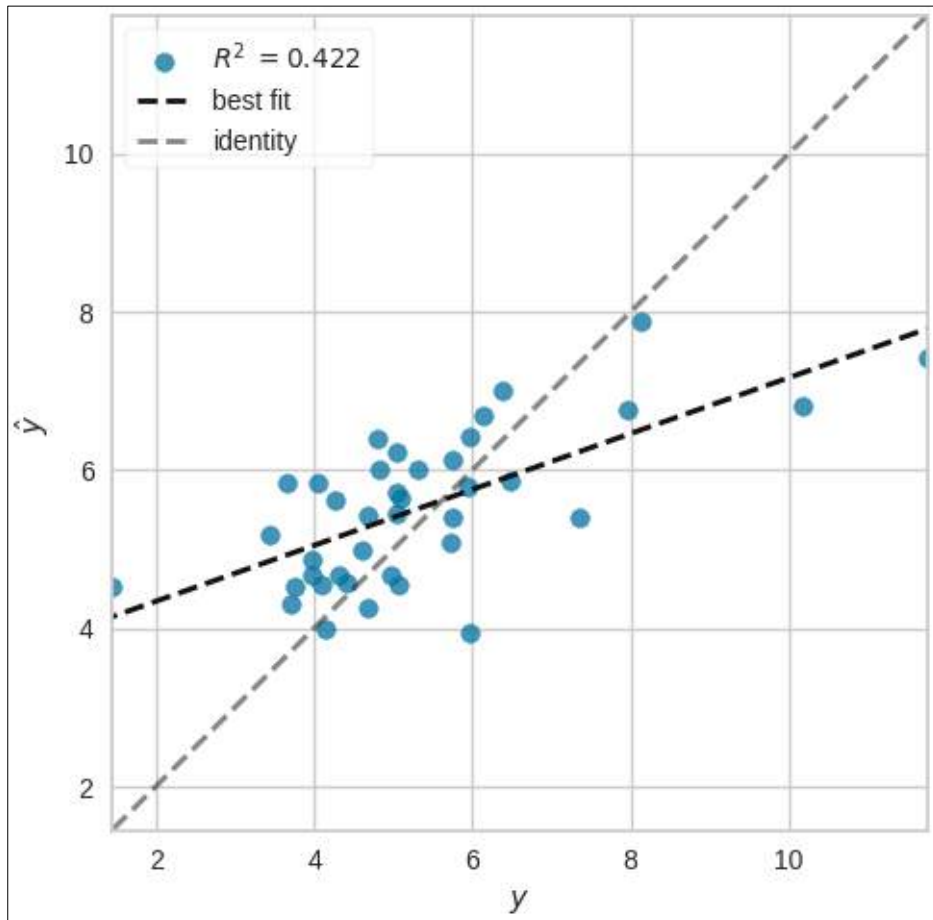
## Appendix 2. Results for barley

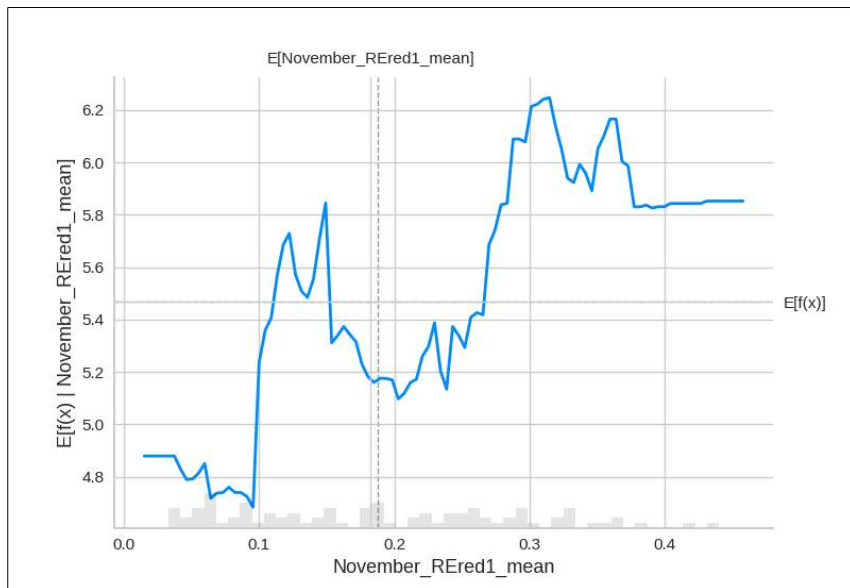
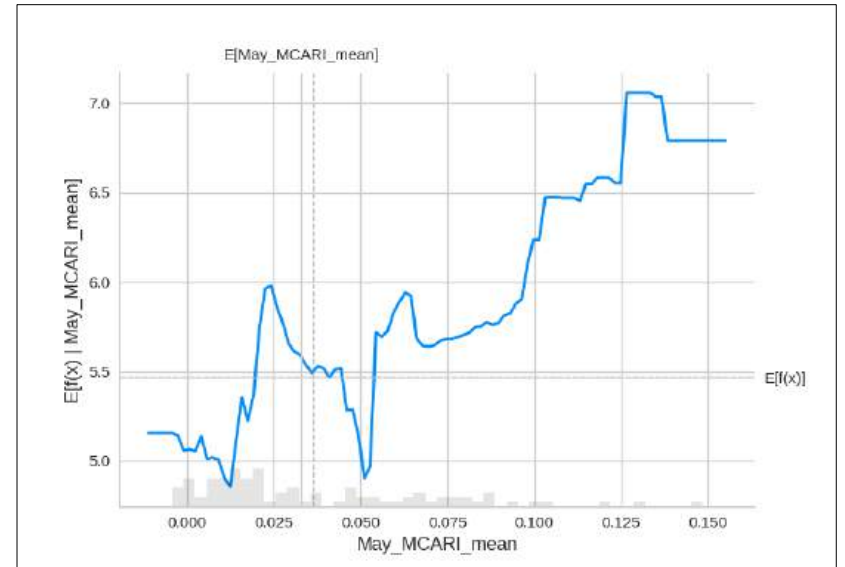
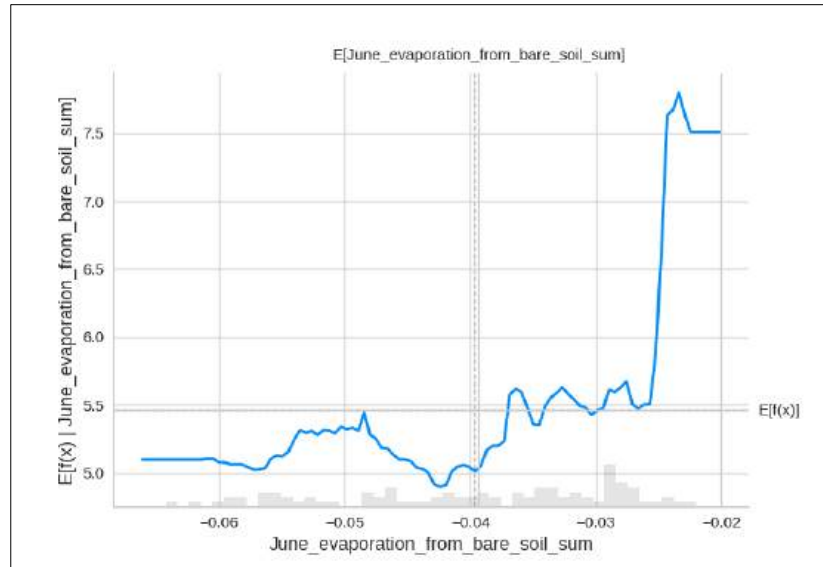
### a) BARLEY AND CALCIUM NUTRIENT COMPOSITION OF GRAINS.





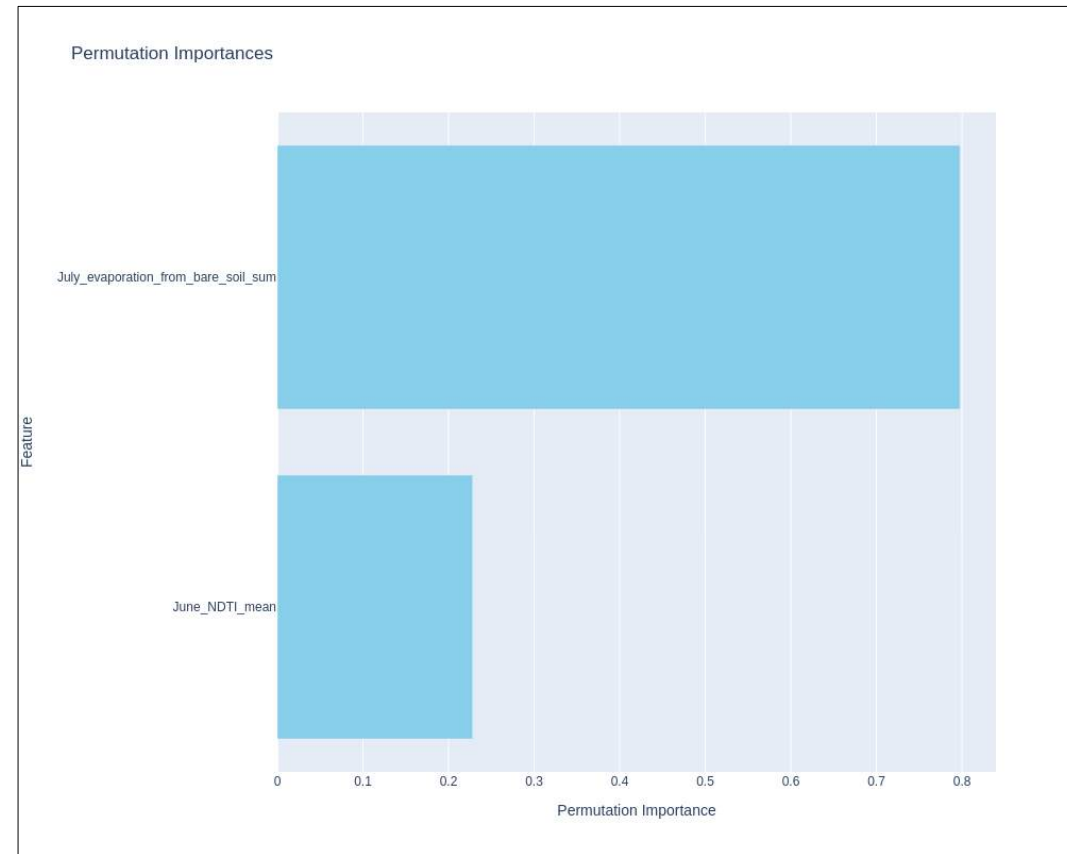
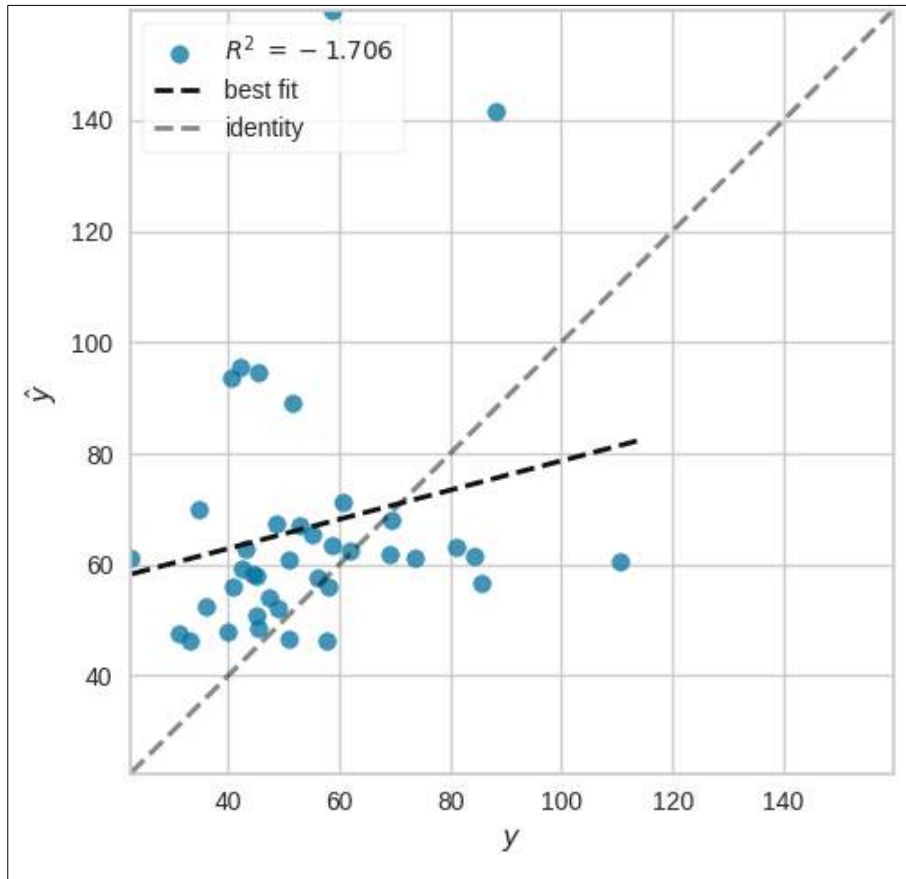
b) BARLEY AND COPPER NUTRIENT COMPOSITION OF GRAINS

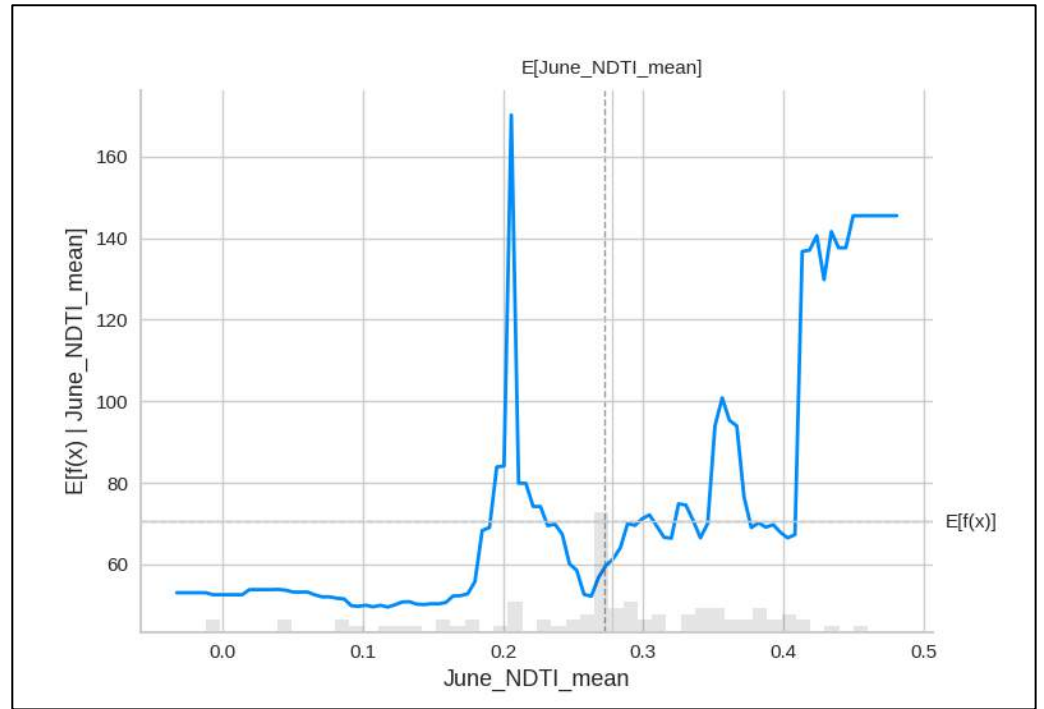
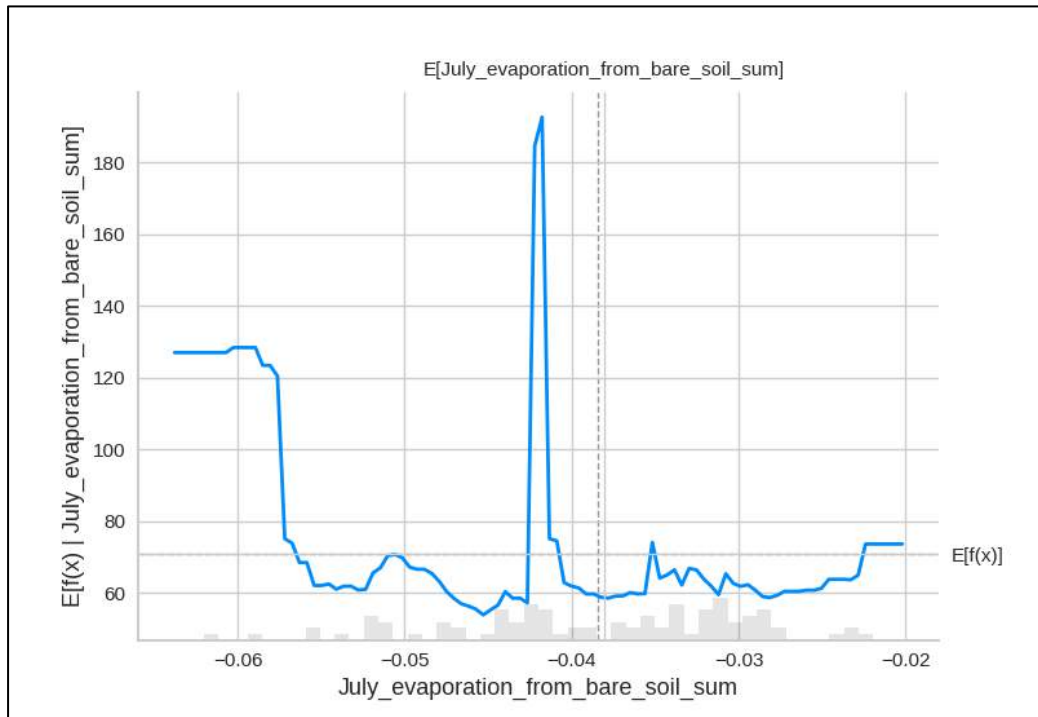




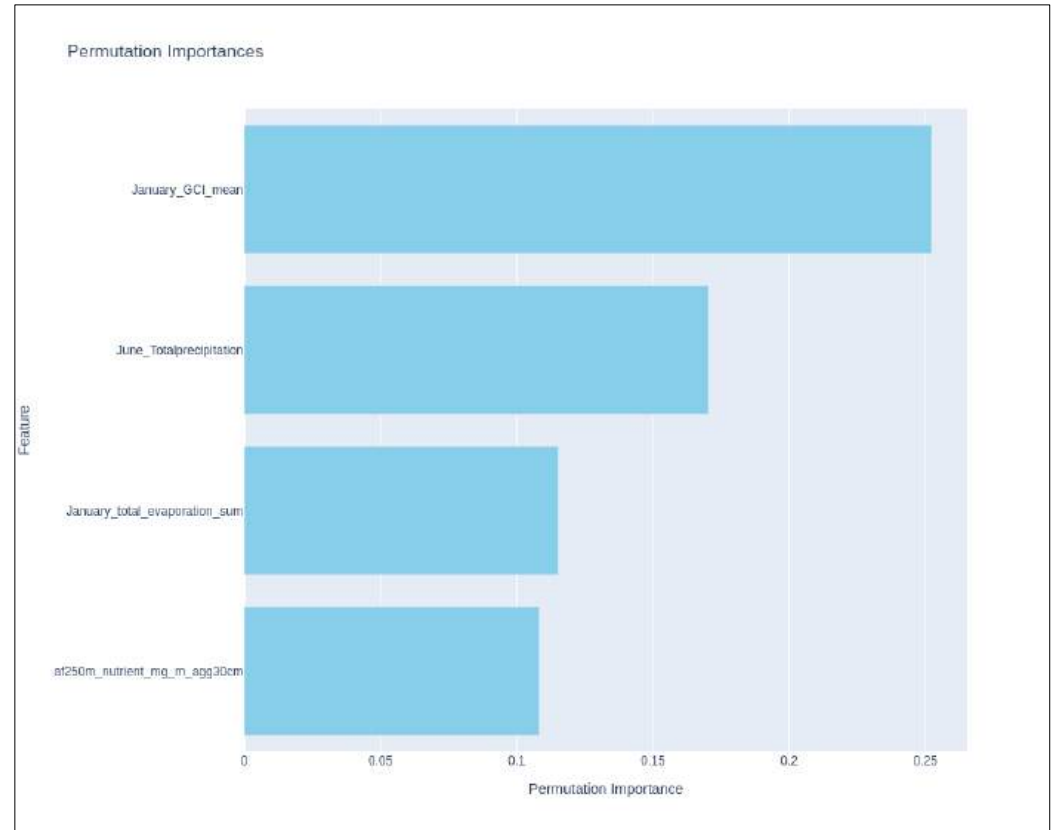
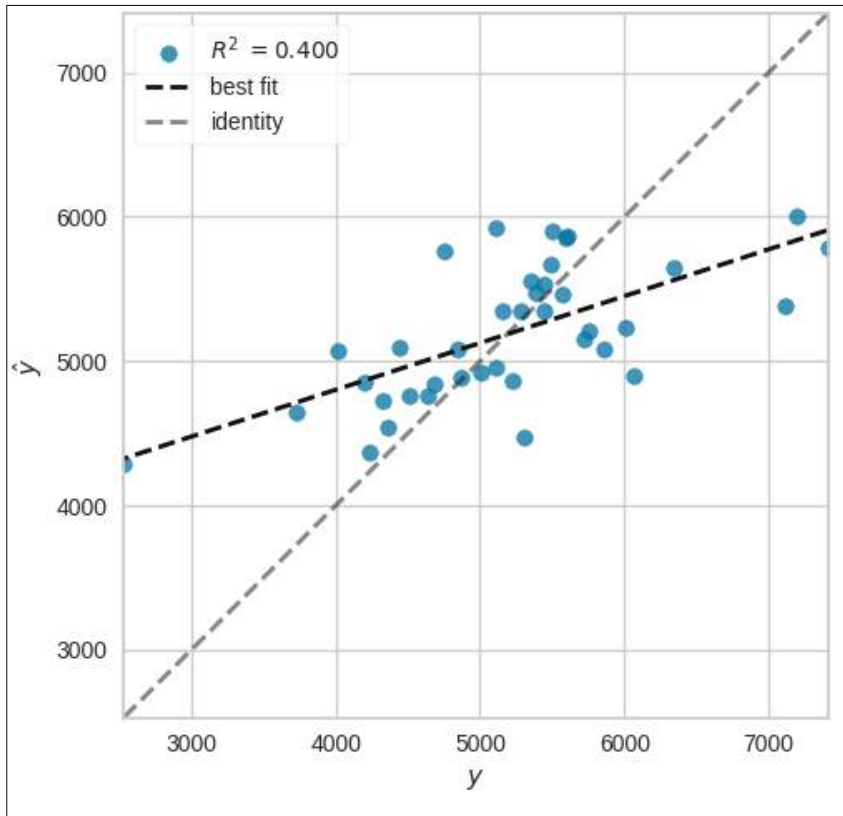


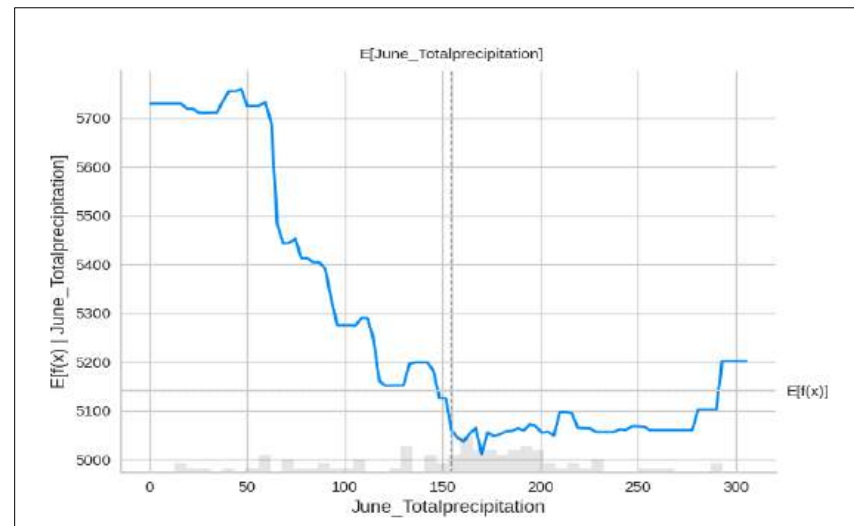
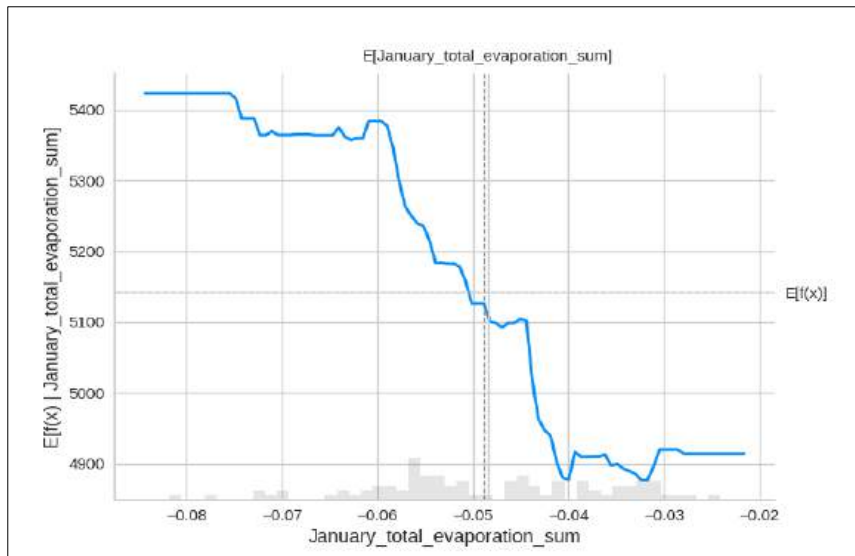
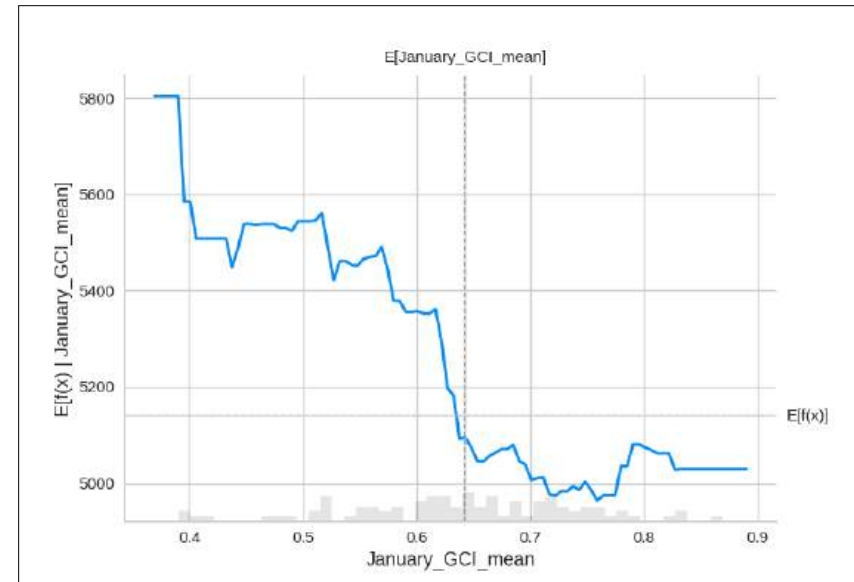
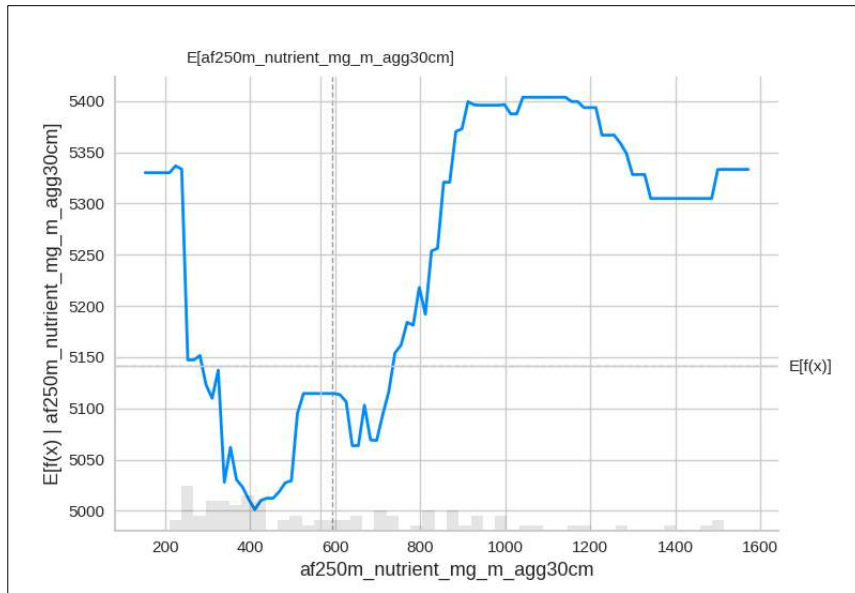
c) BARLEY AND IRON NUTRIENT COMPOSITION OF GRAINS



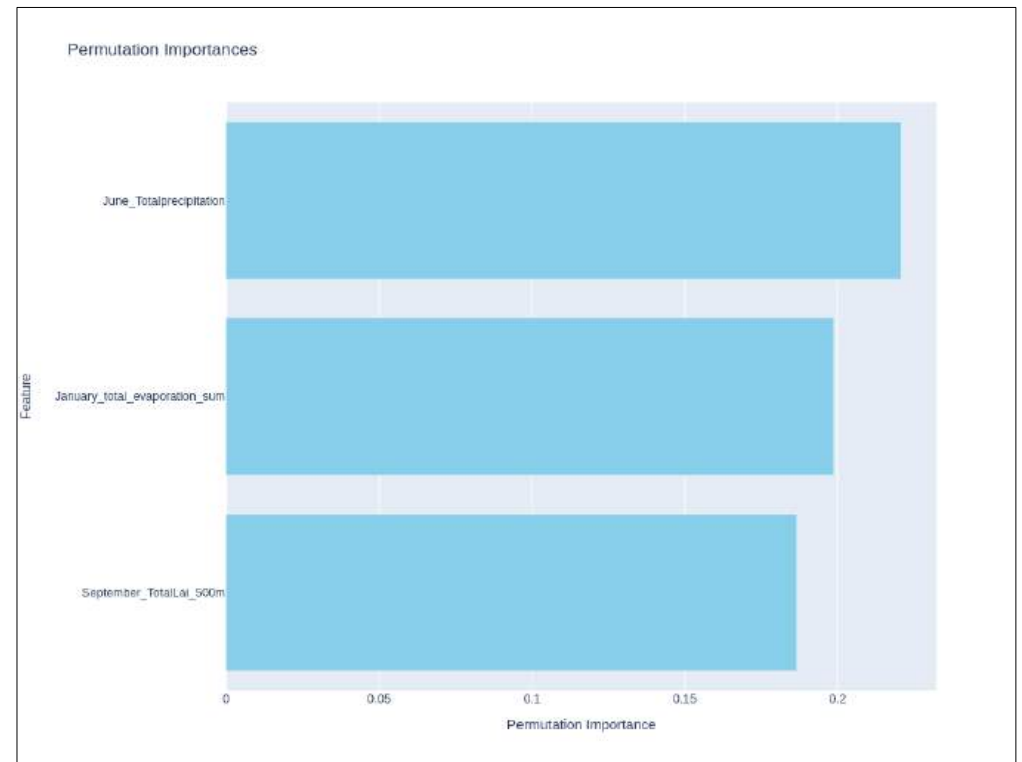
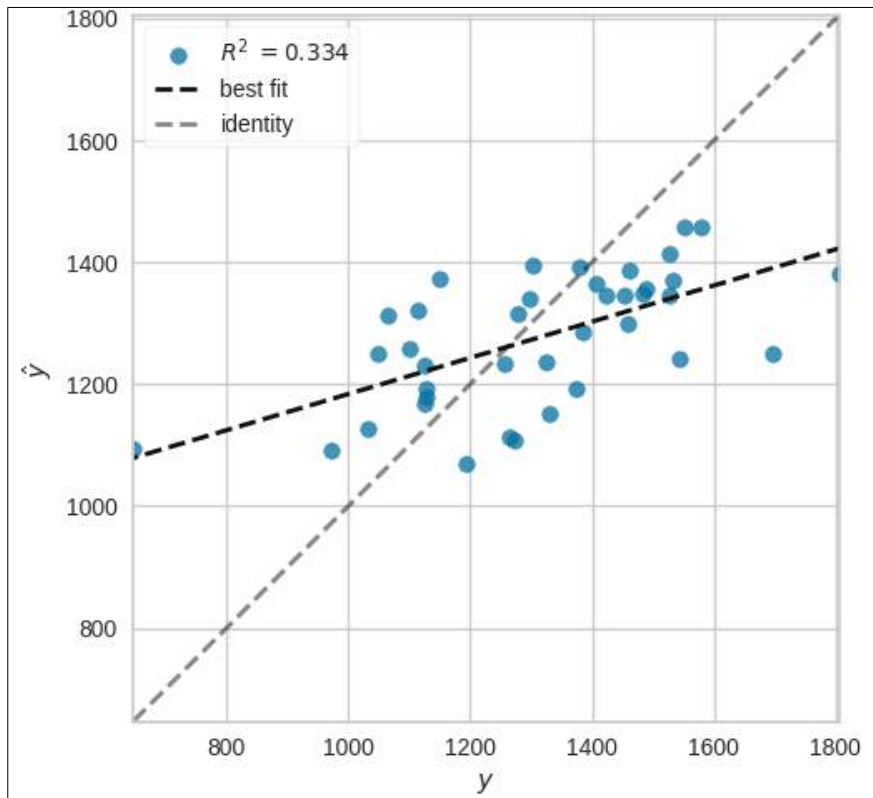


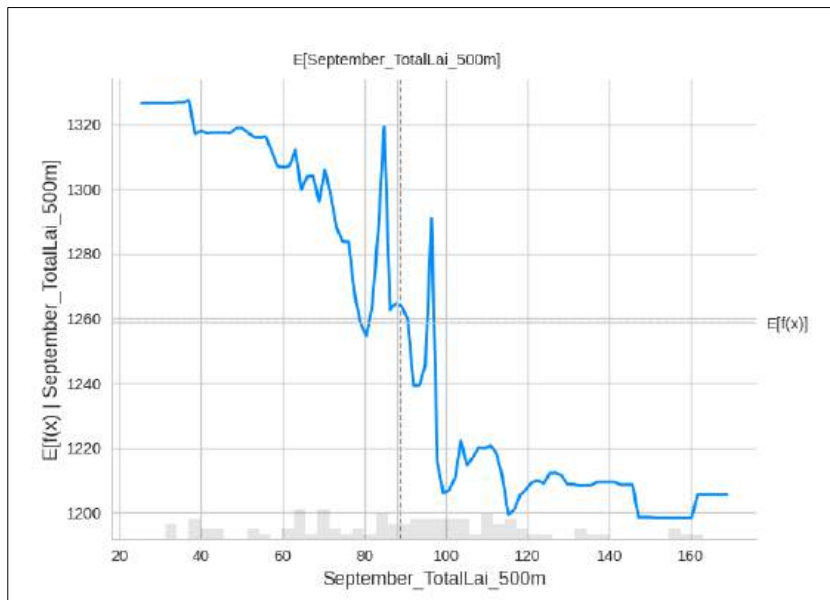
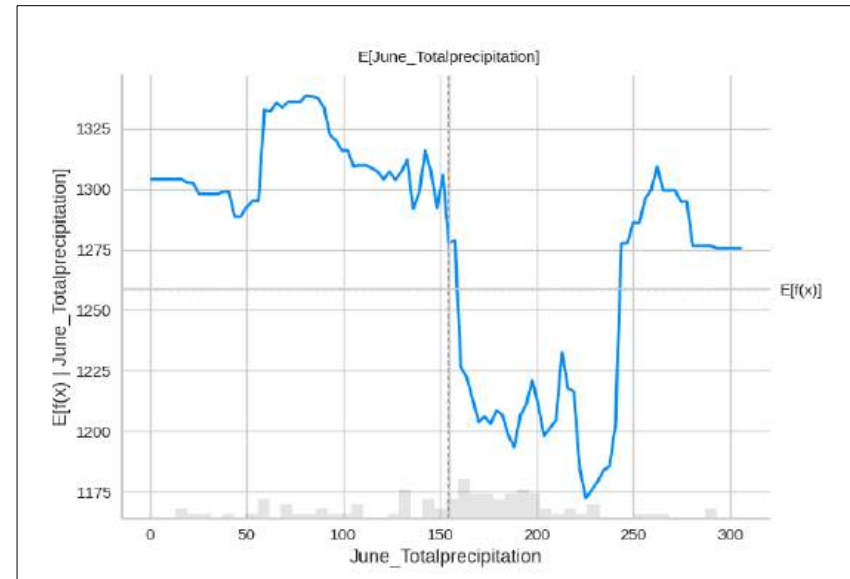
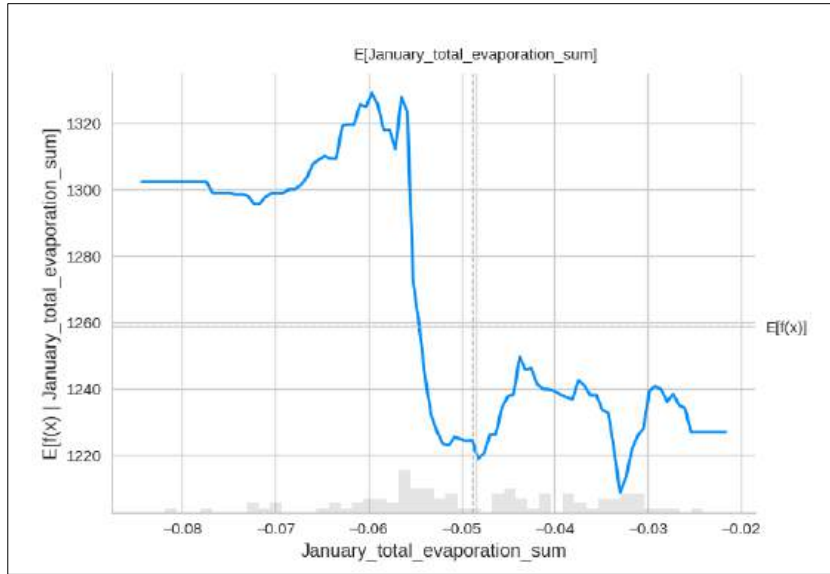
d) BARLEY AND POTASSIUM NUTRIENT COMPOSITION OF GRAINS



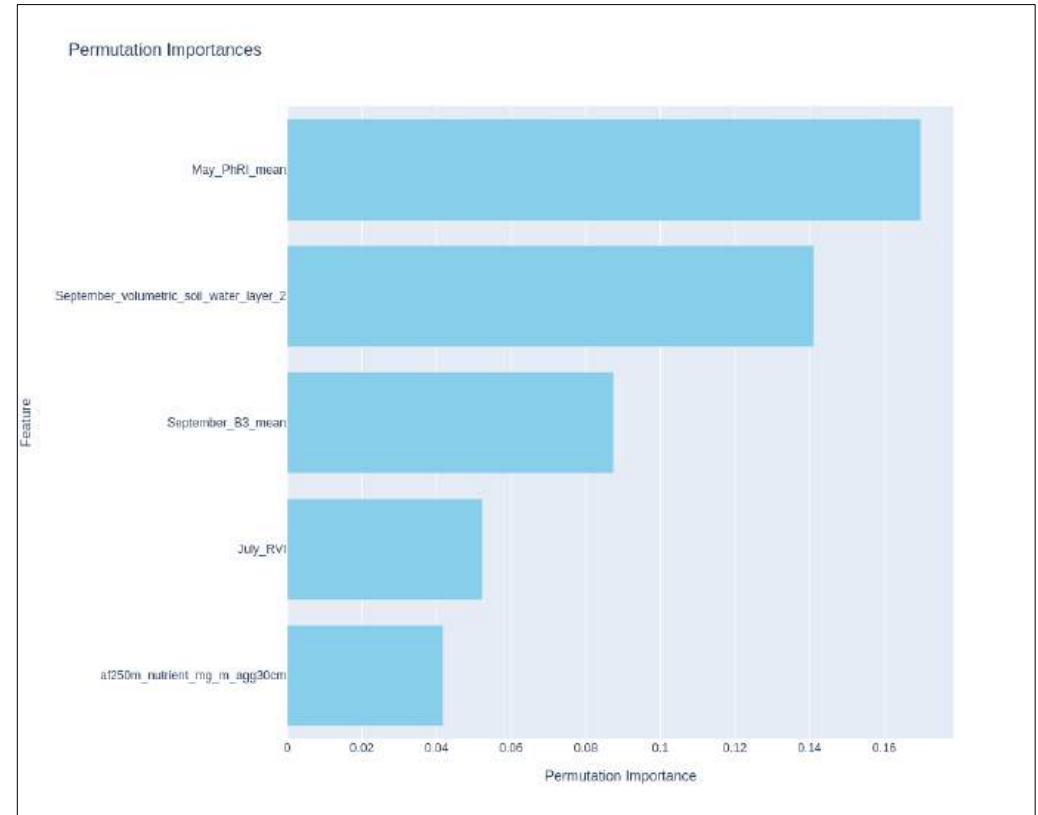
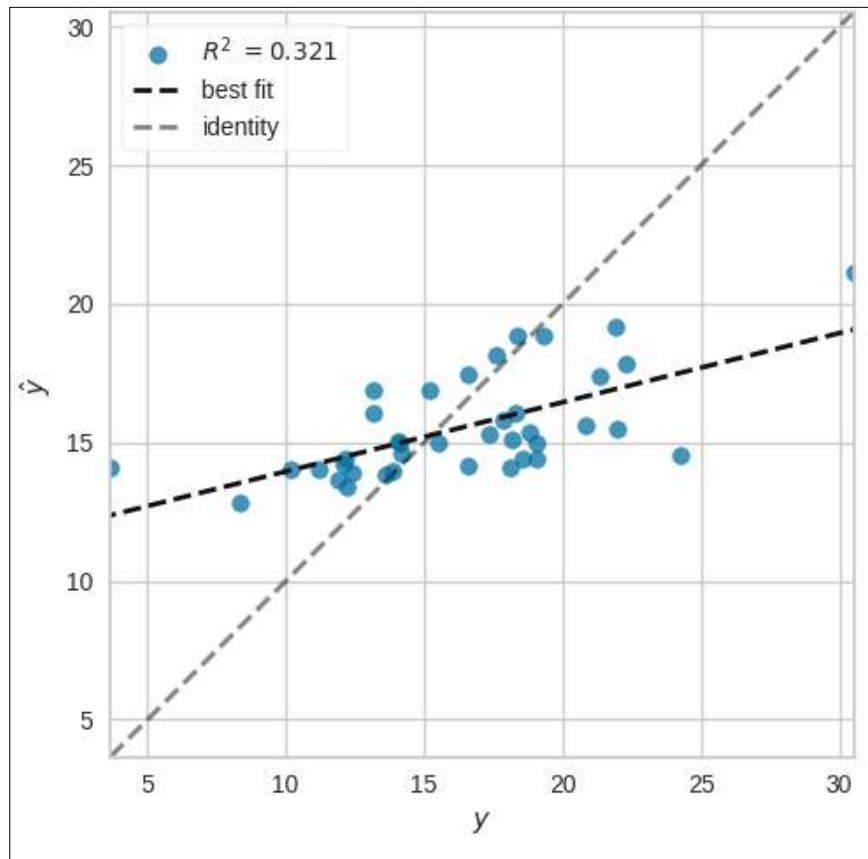


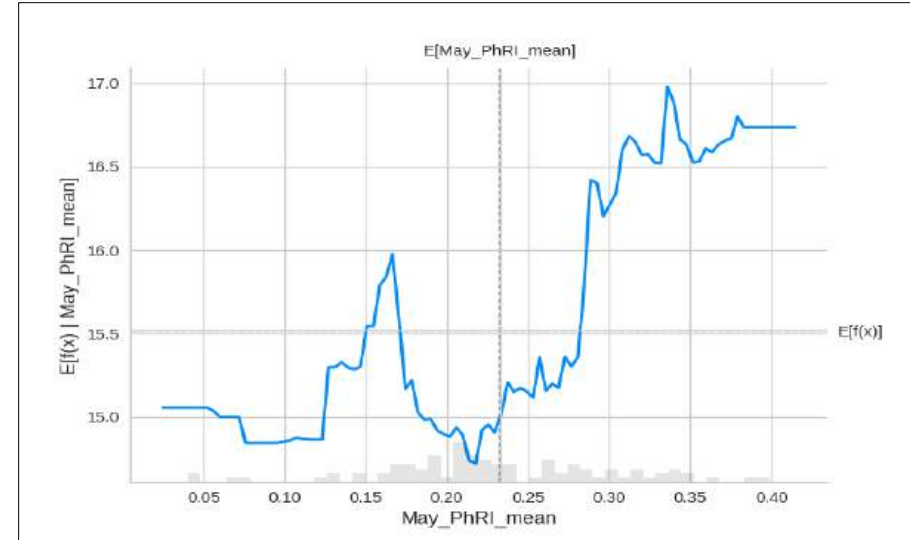
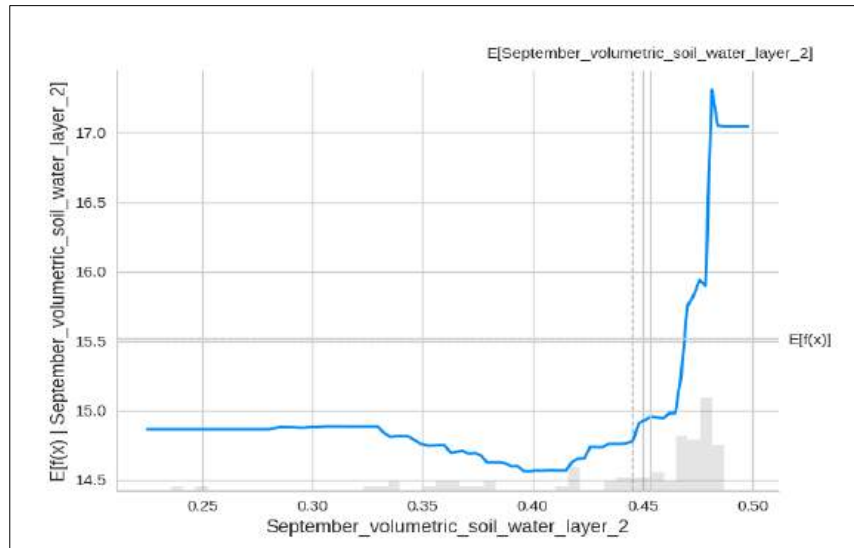
e) BARLEY AND MAGNESIUM NUTRIENT COMPOSITION OF GRAINS



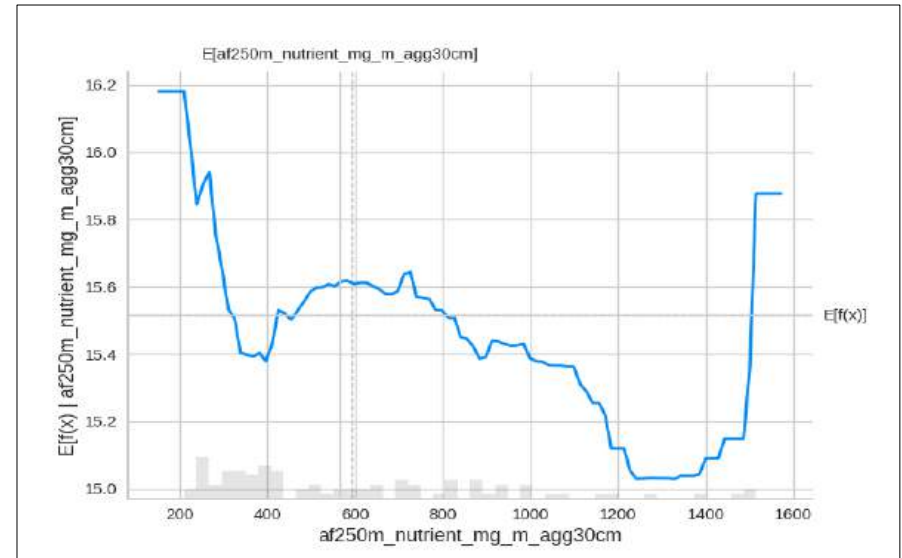
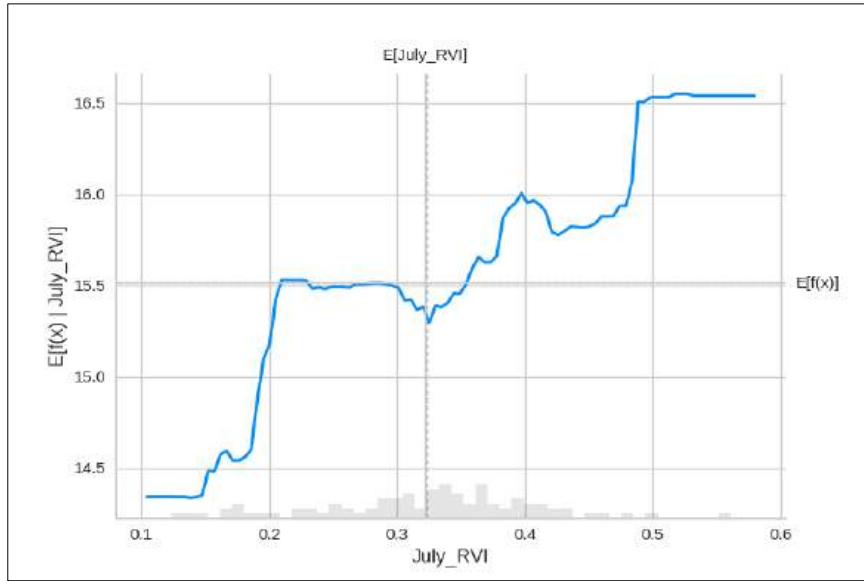


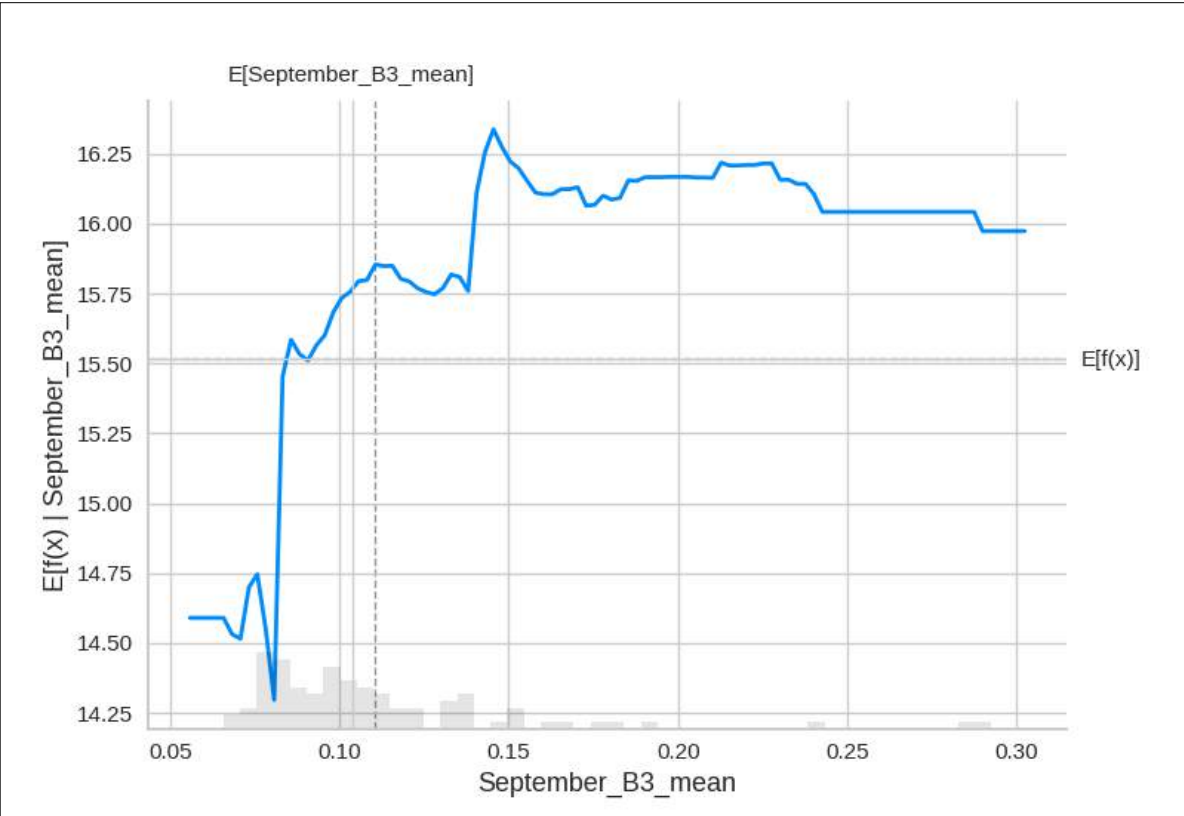
f) BARLEY AND MANGANESE NUTRIENT COMPOSITION OF GRAINS



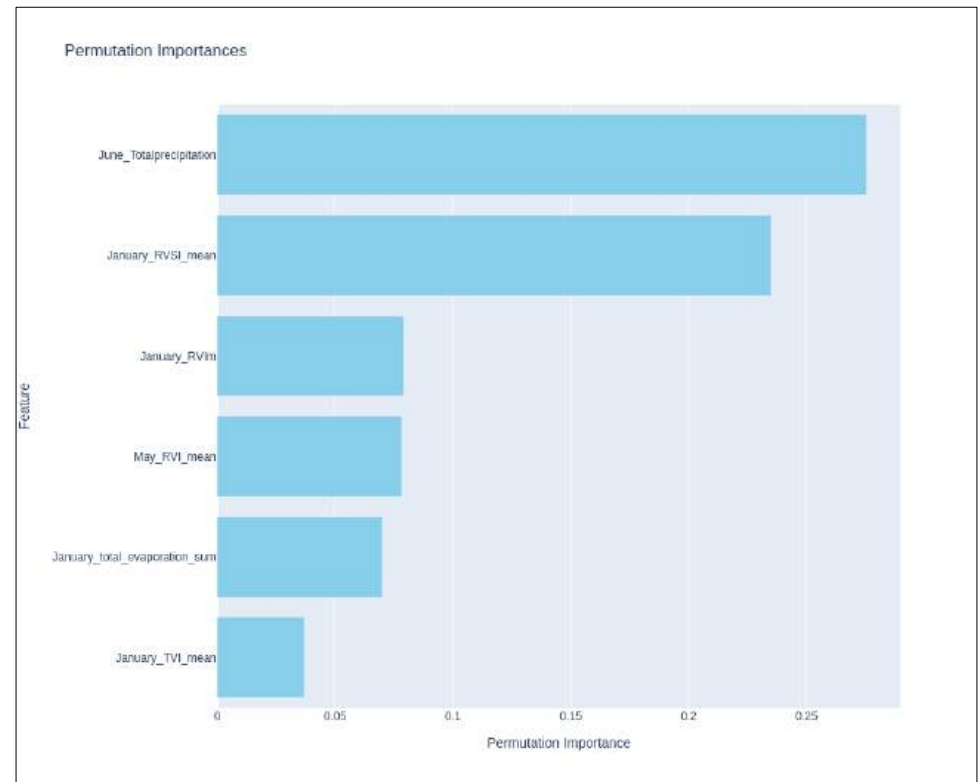
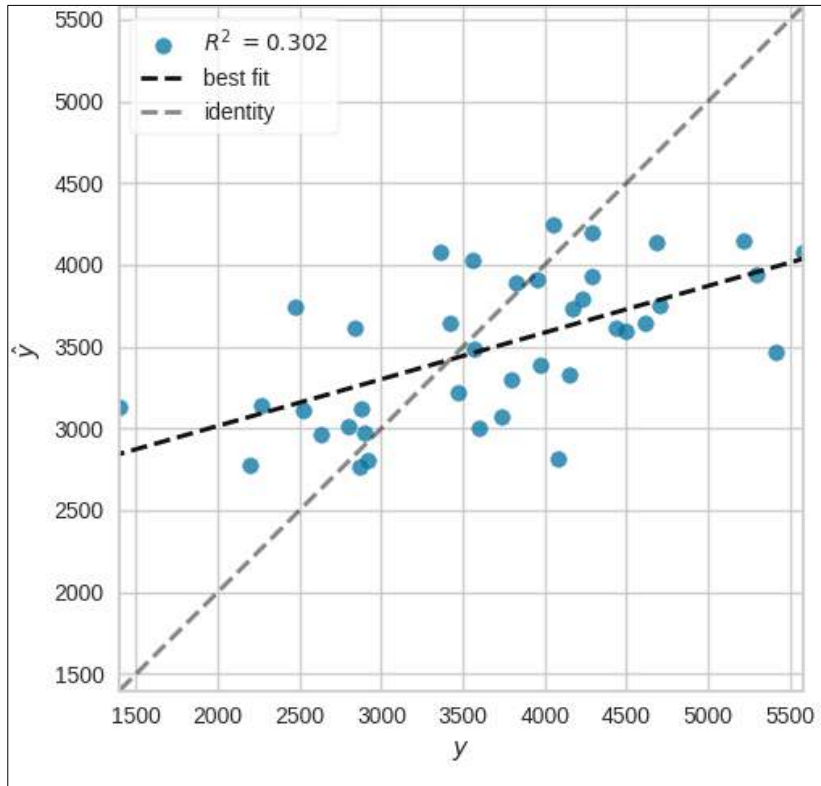


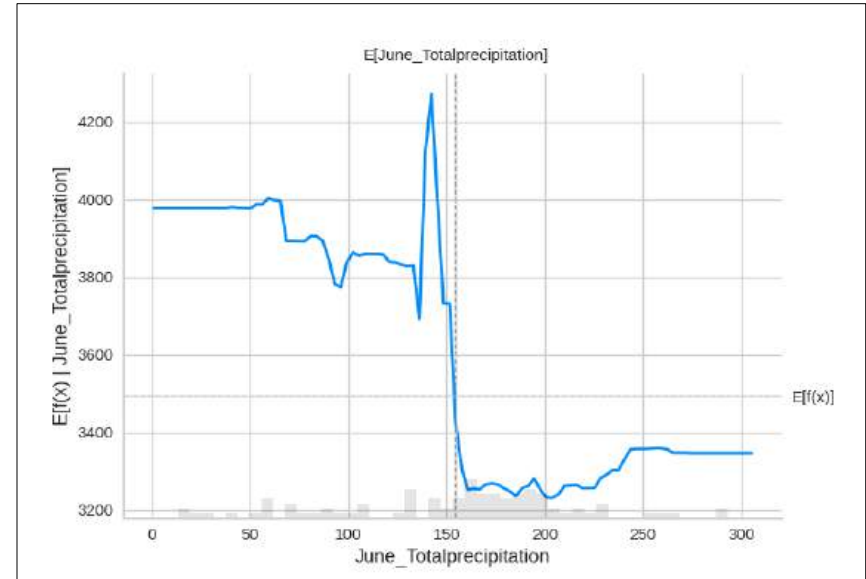
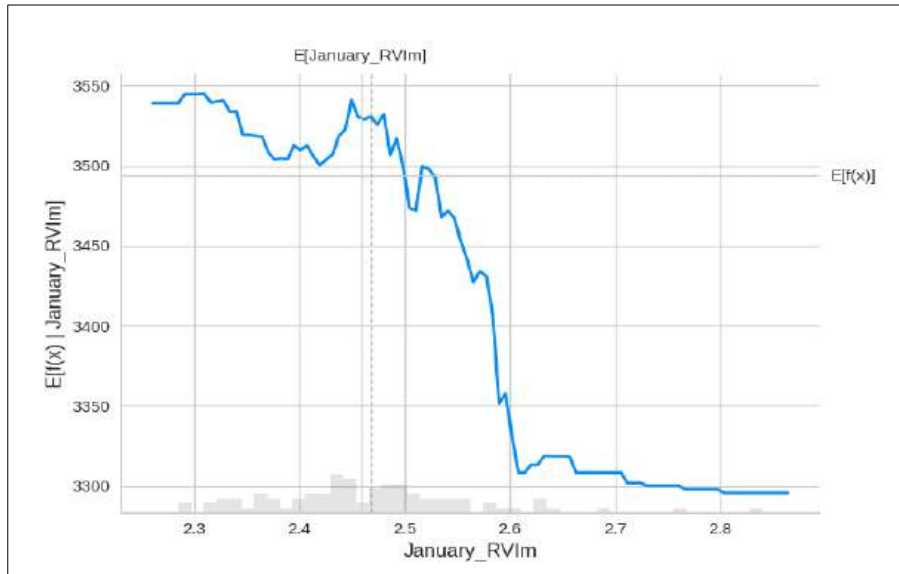


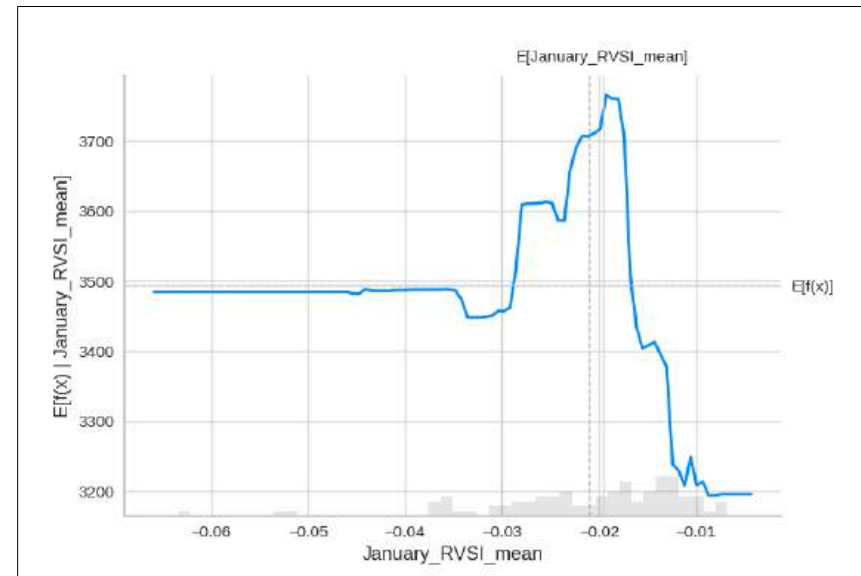
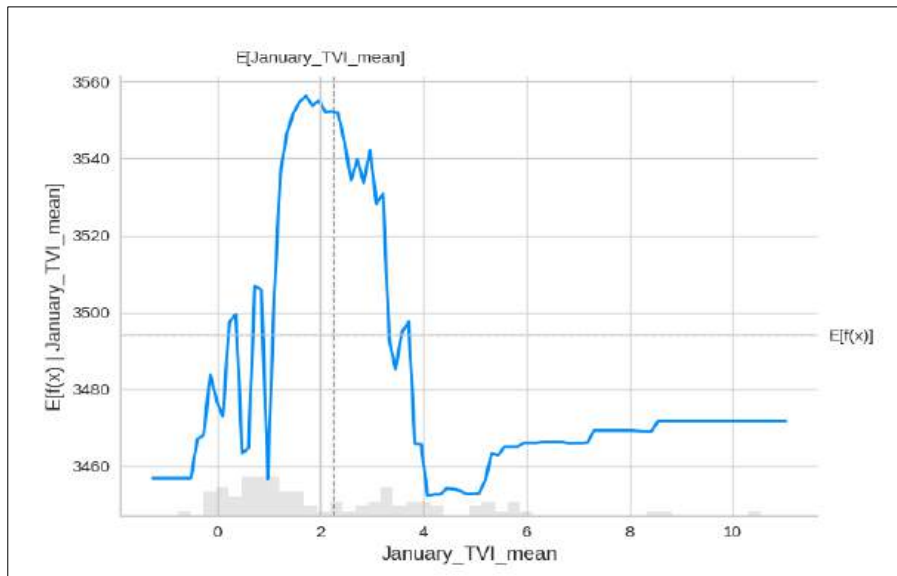


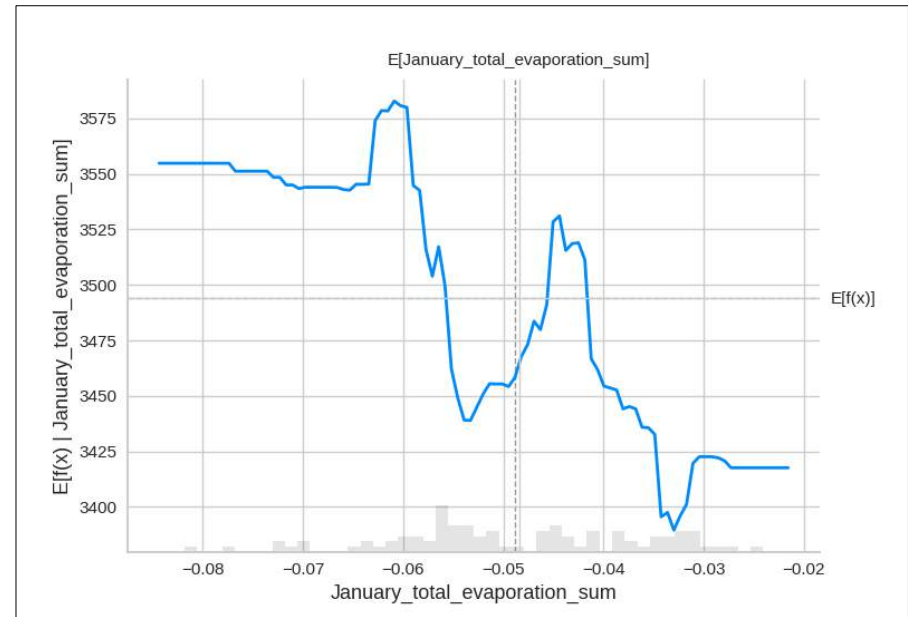
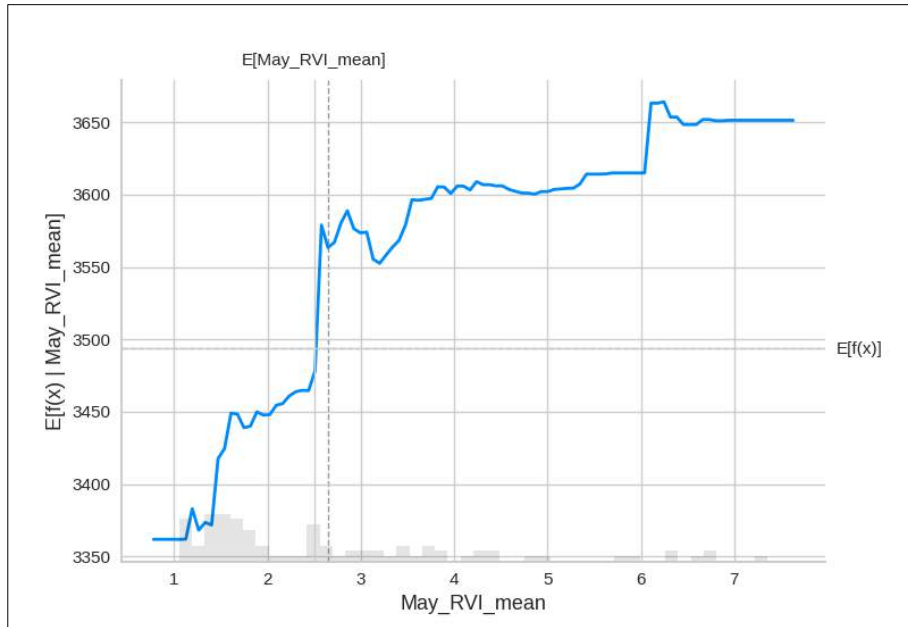


g) BARLEY AND PHOSPHORUS NUTRIENT COMPOSITION OF GRAINS

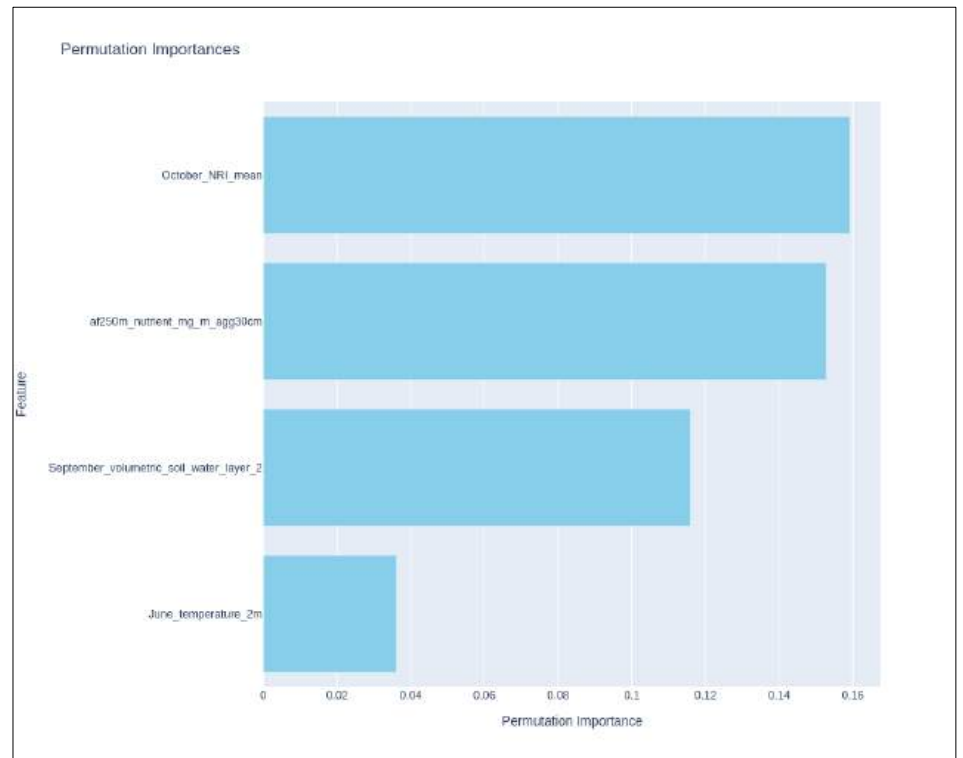
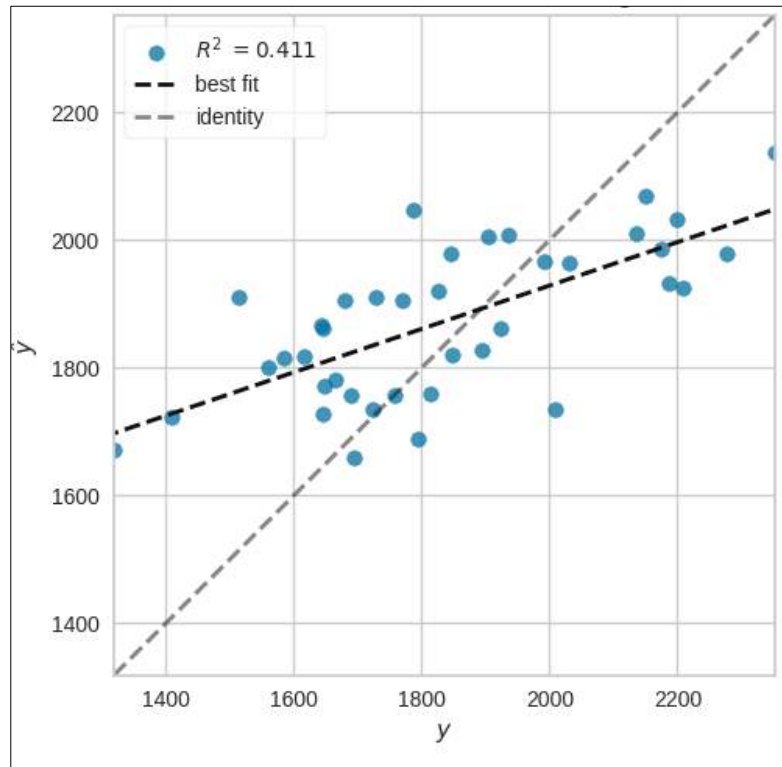


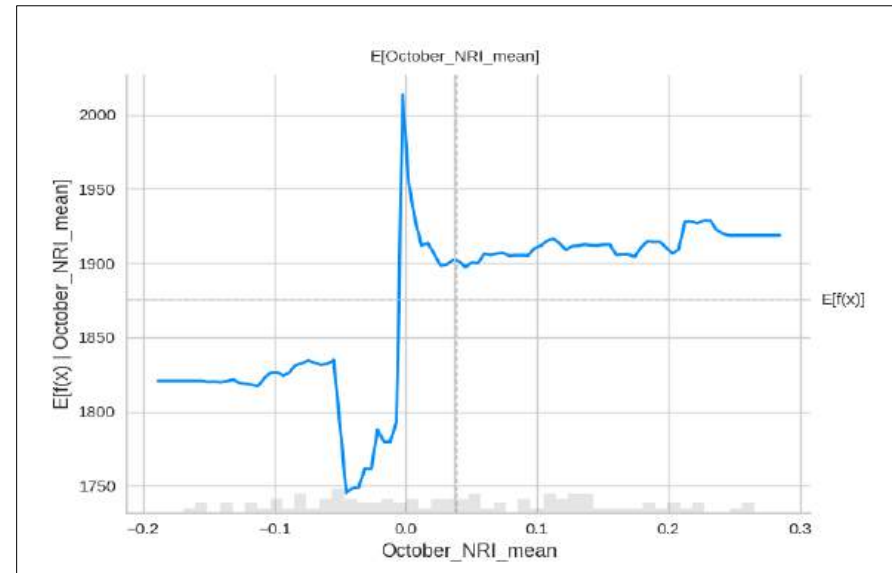
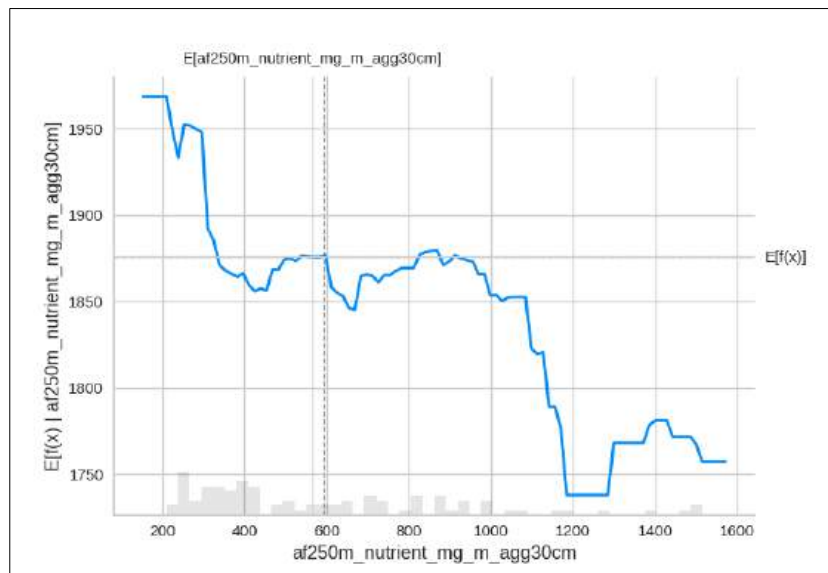
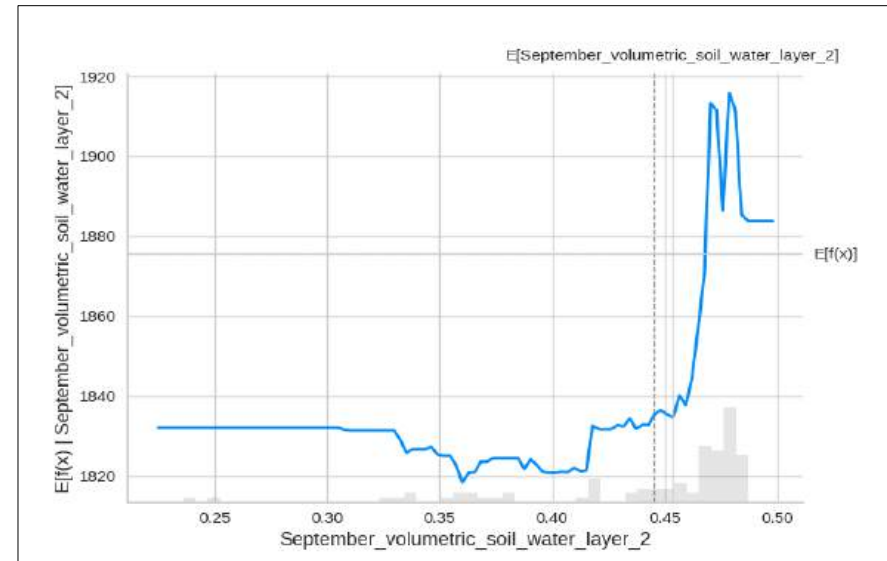
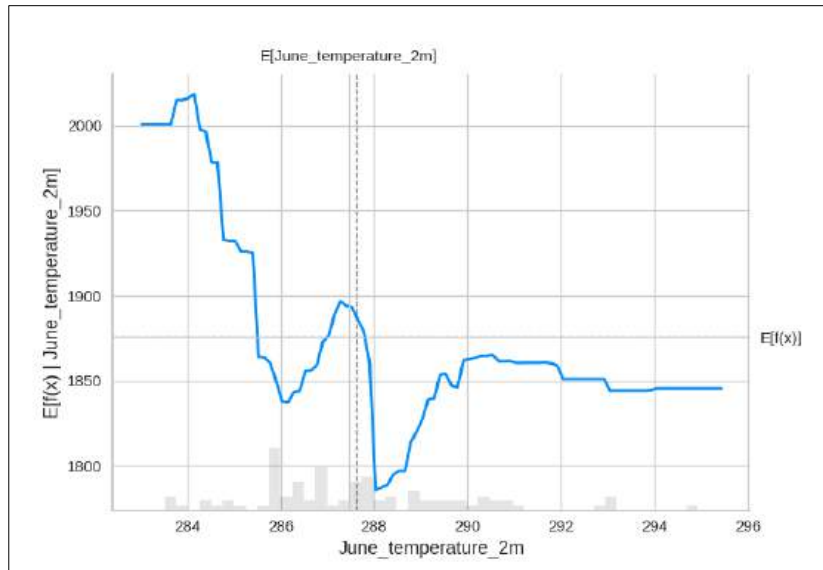






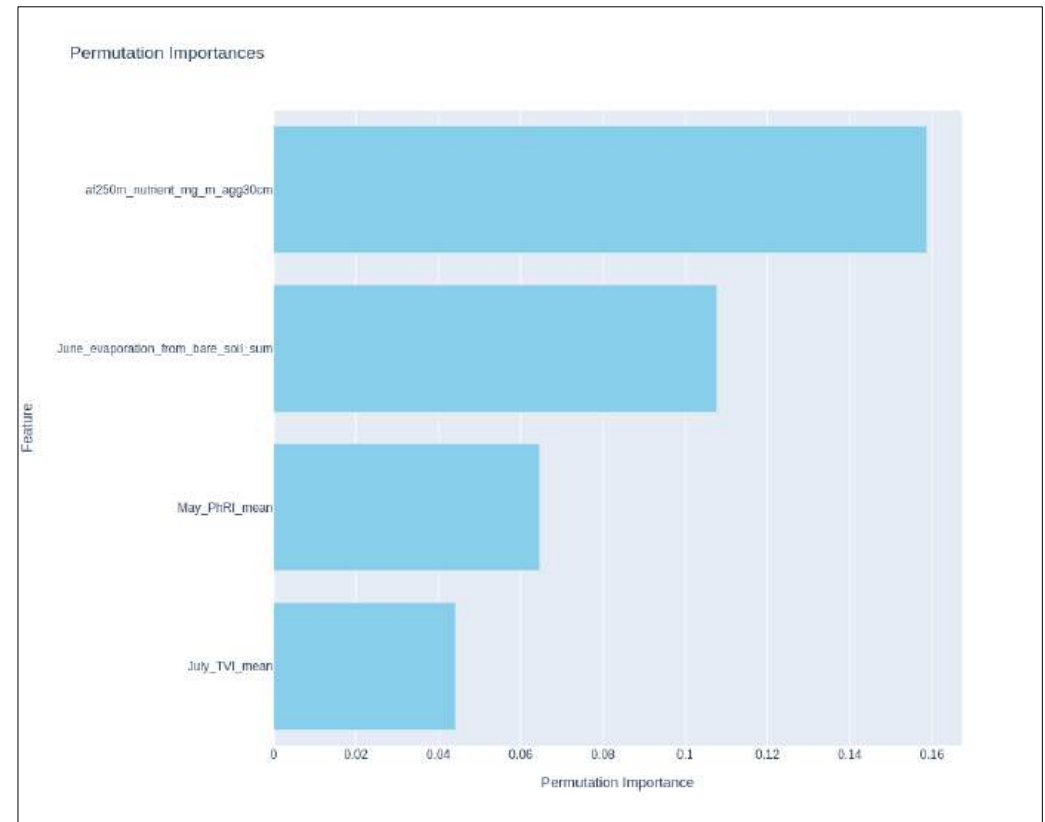
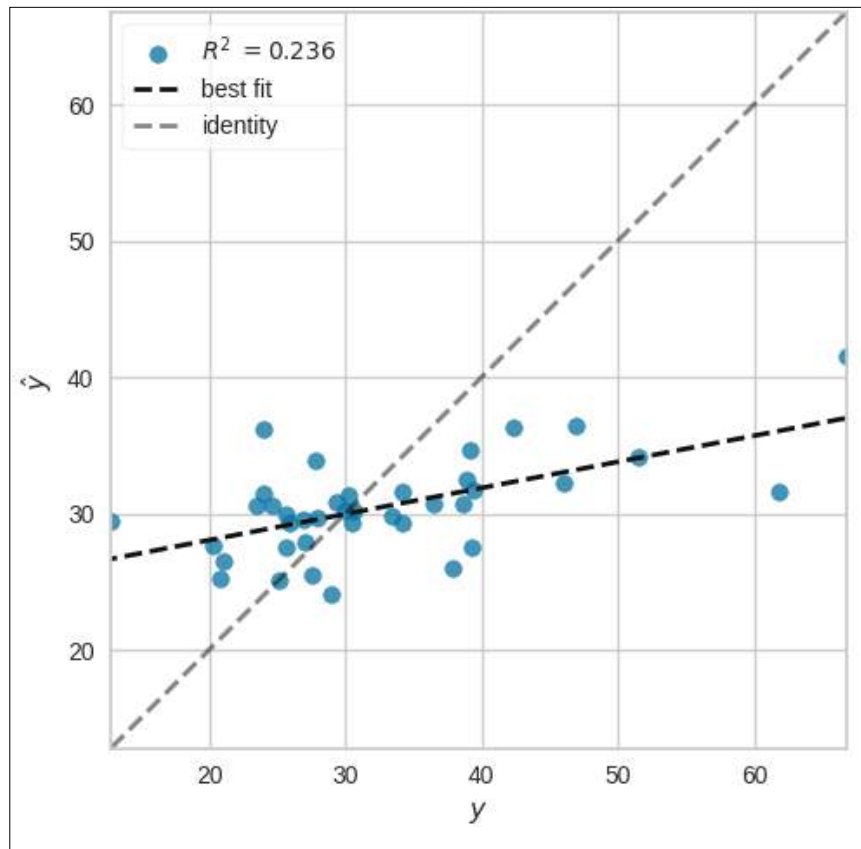
### h) BARLEY AND SULPHUR NUTRIENT COMPOSITION OF GRAINS

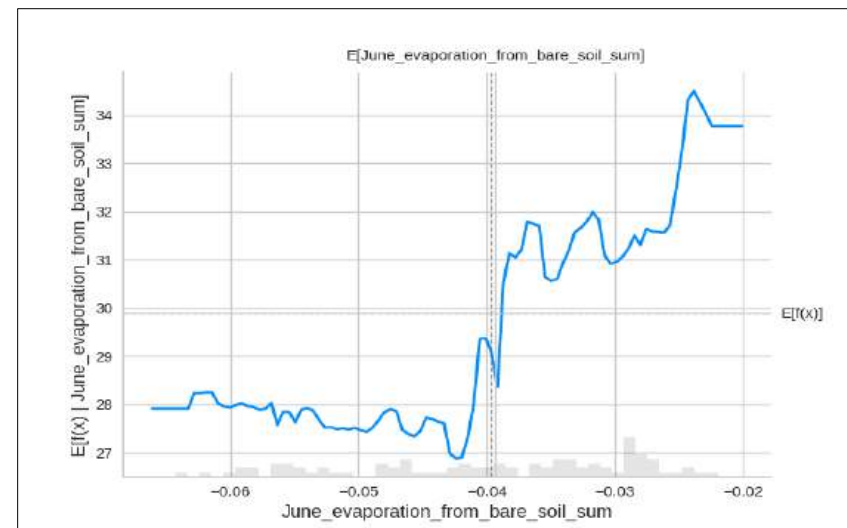
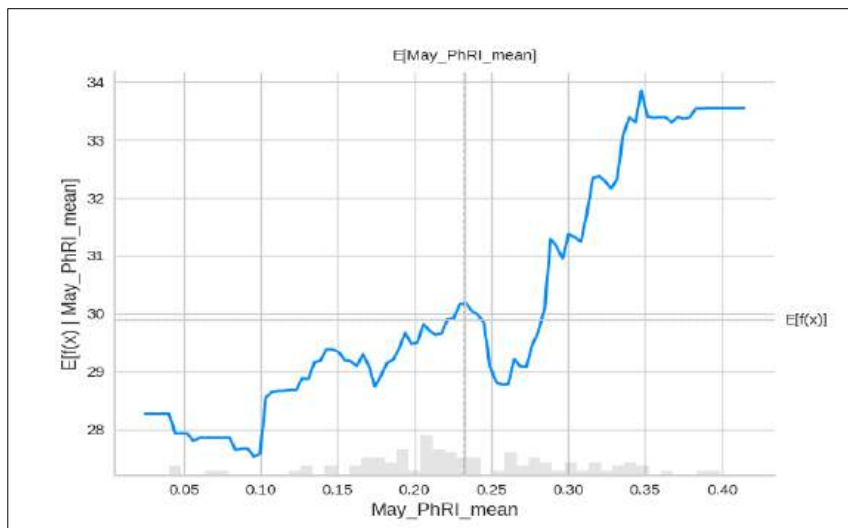
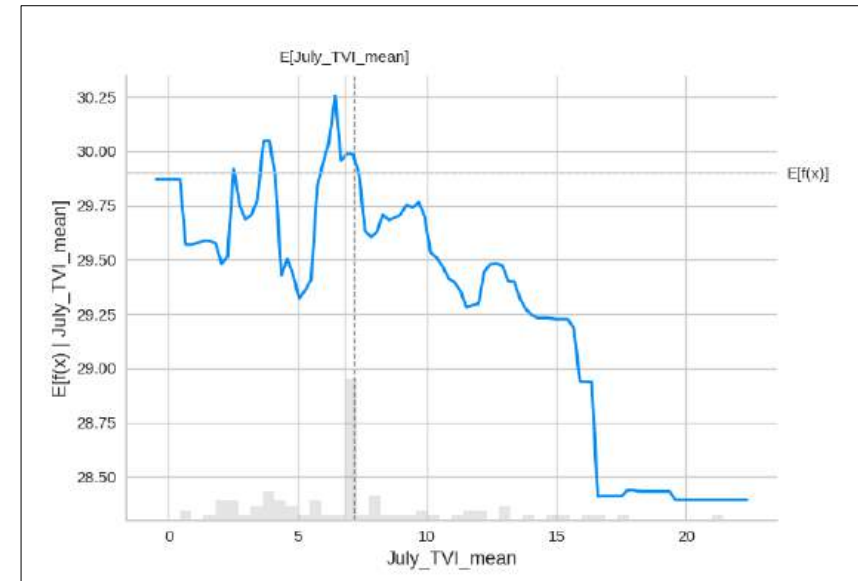
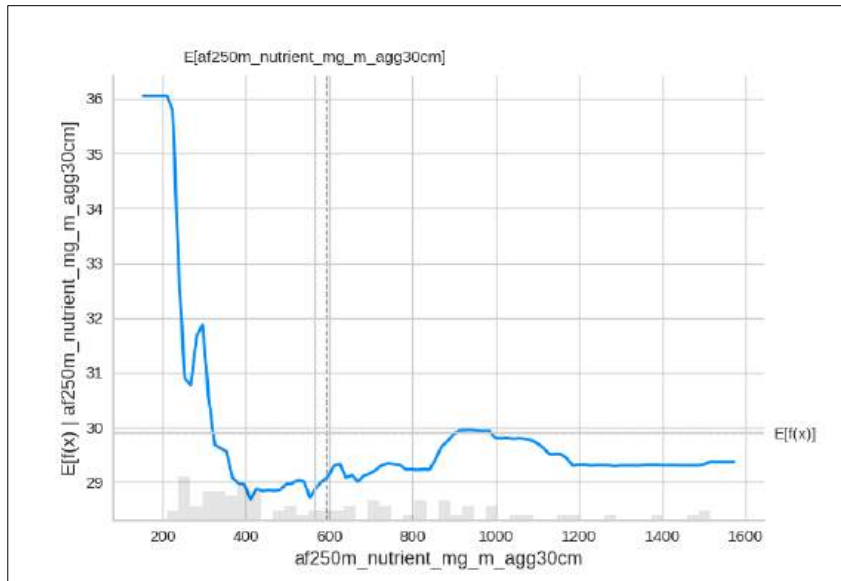






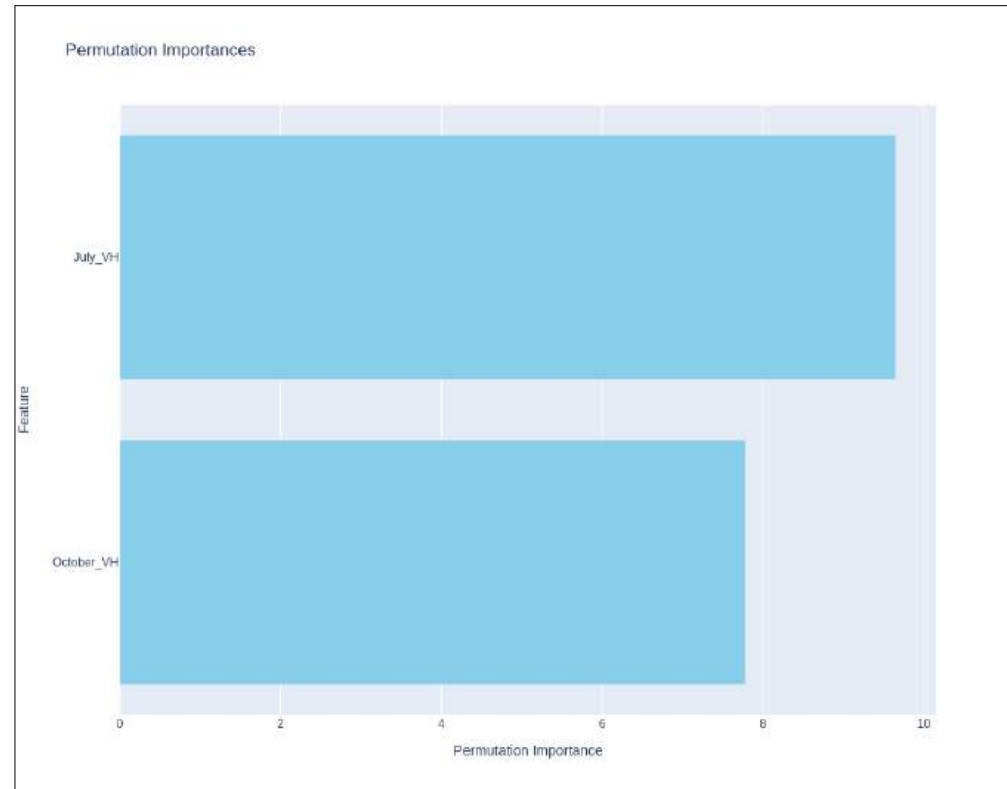
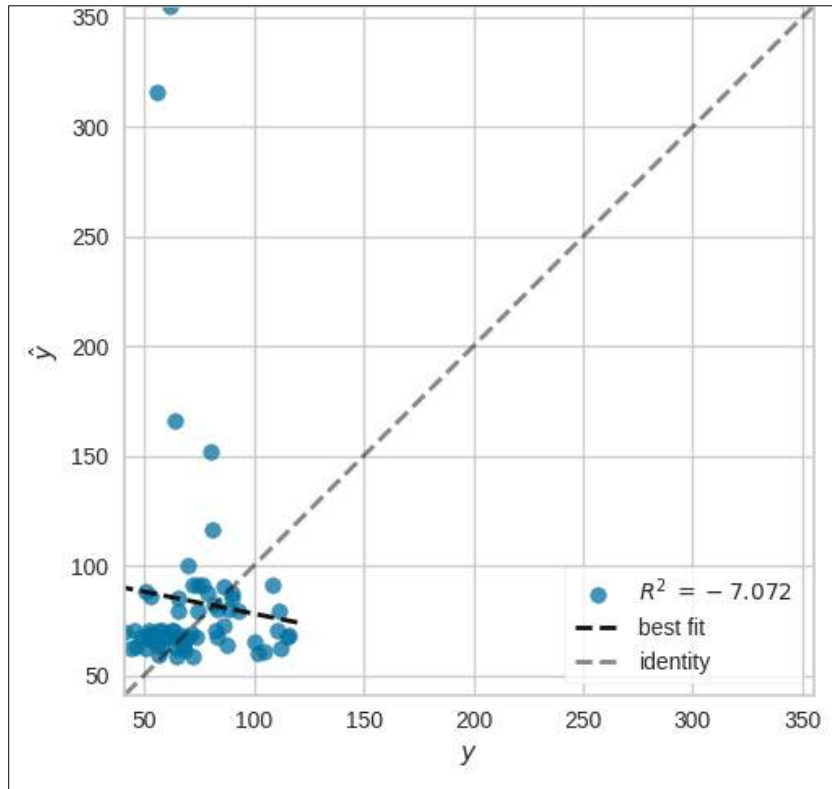
i) BARLEY AND ZINC NUTRIENT COMPOSITION OF GRAINS

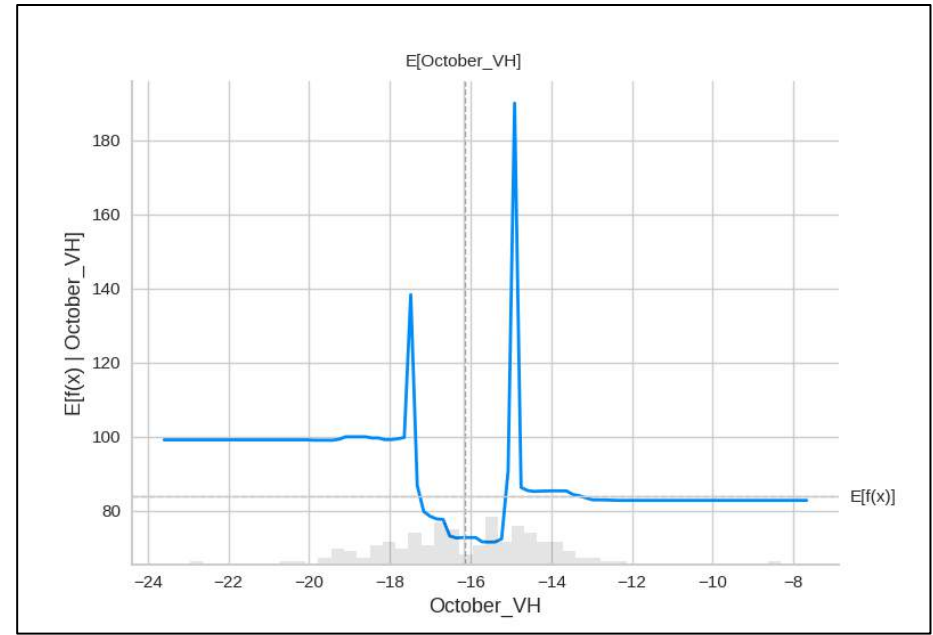
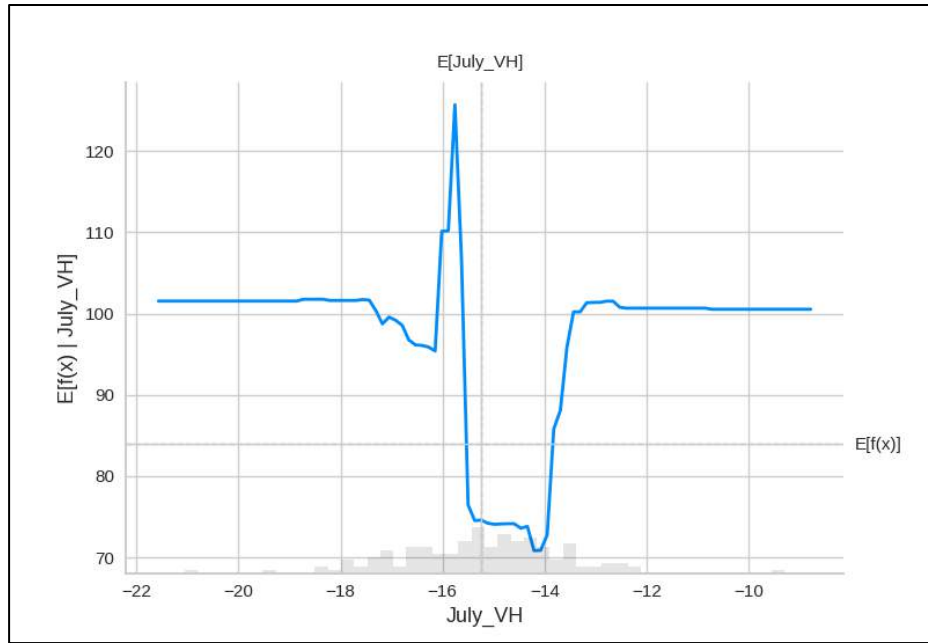




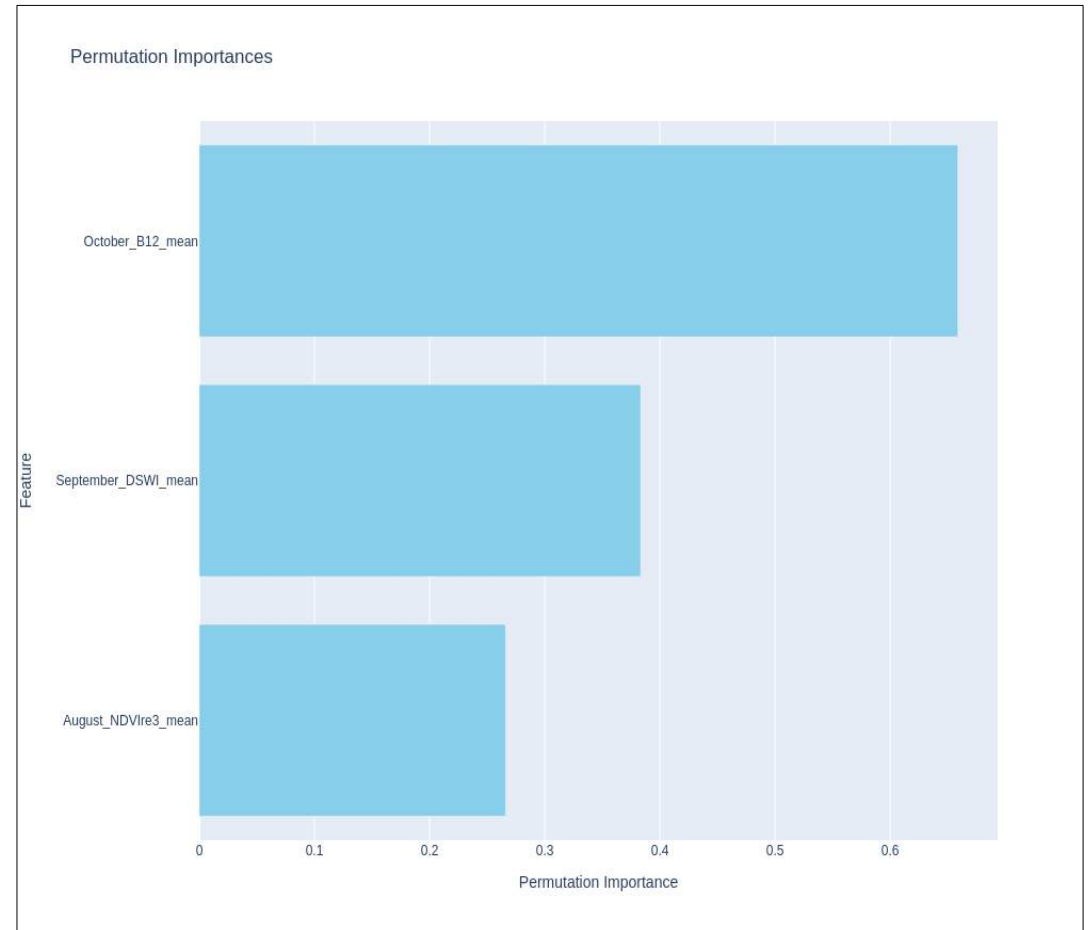
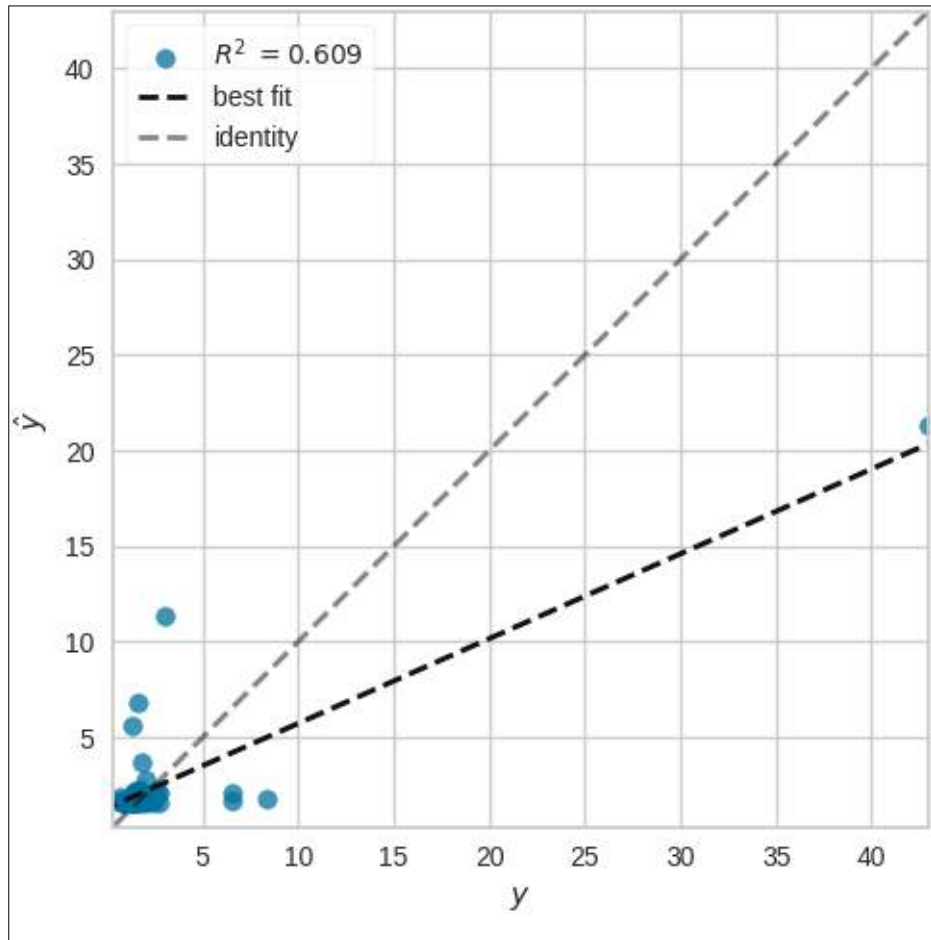
### Appendix 3. Results for maize

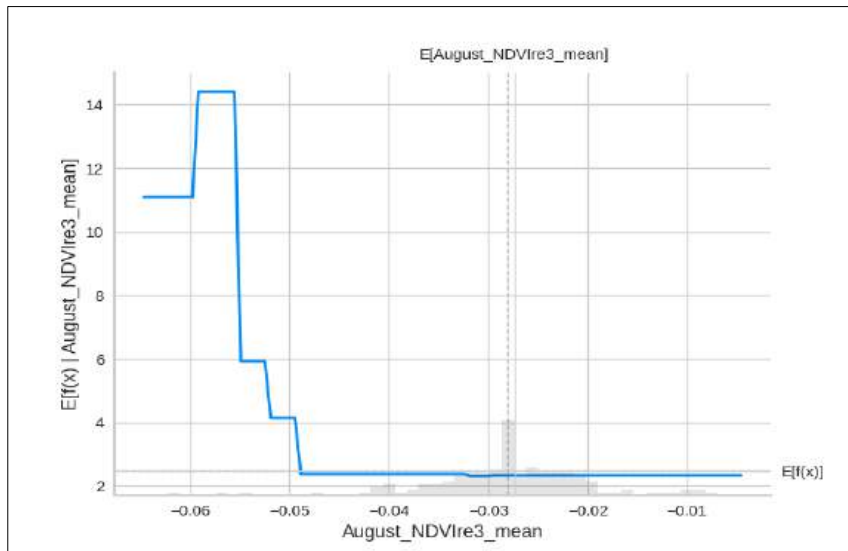
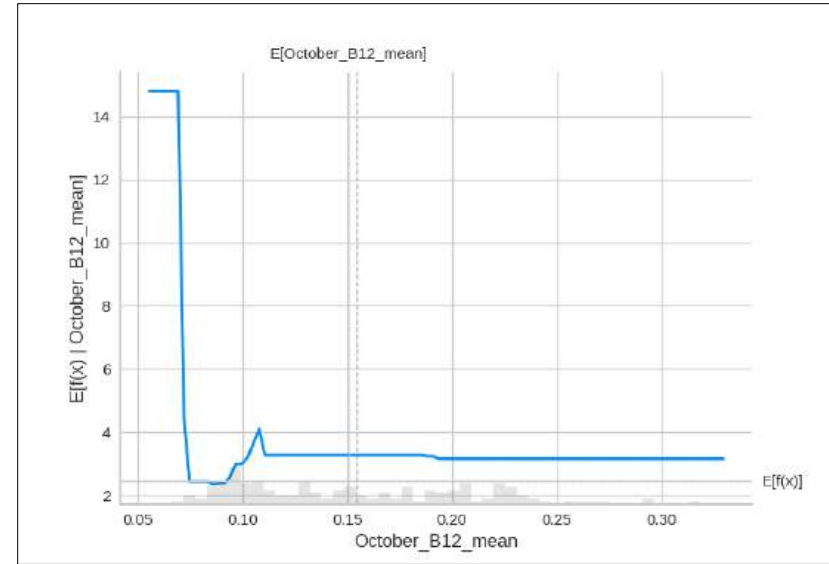
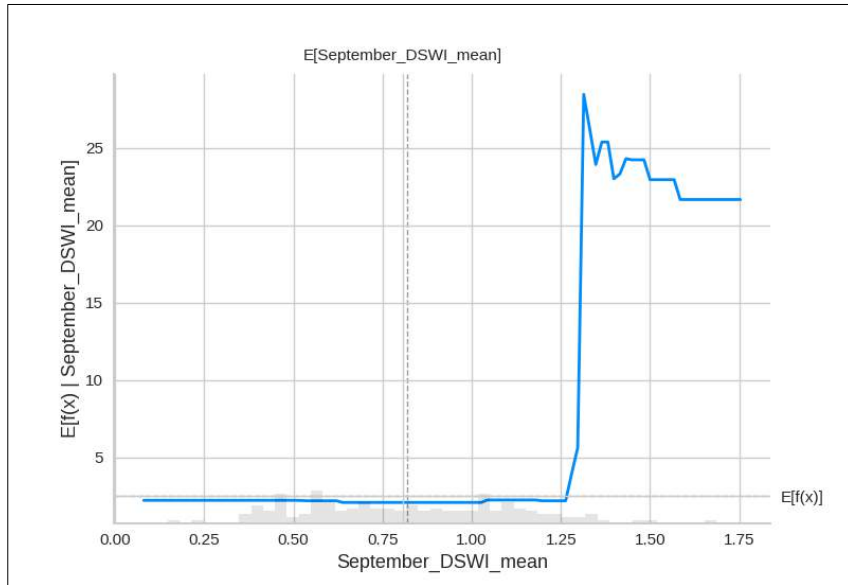
#### a) MAIZE AND CALCIUM NUTRIENT COMPOSITION OF GRAINS



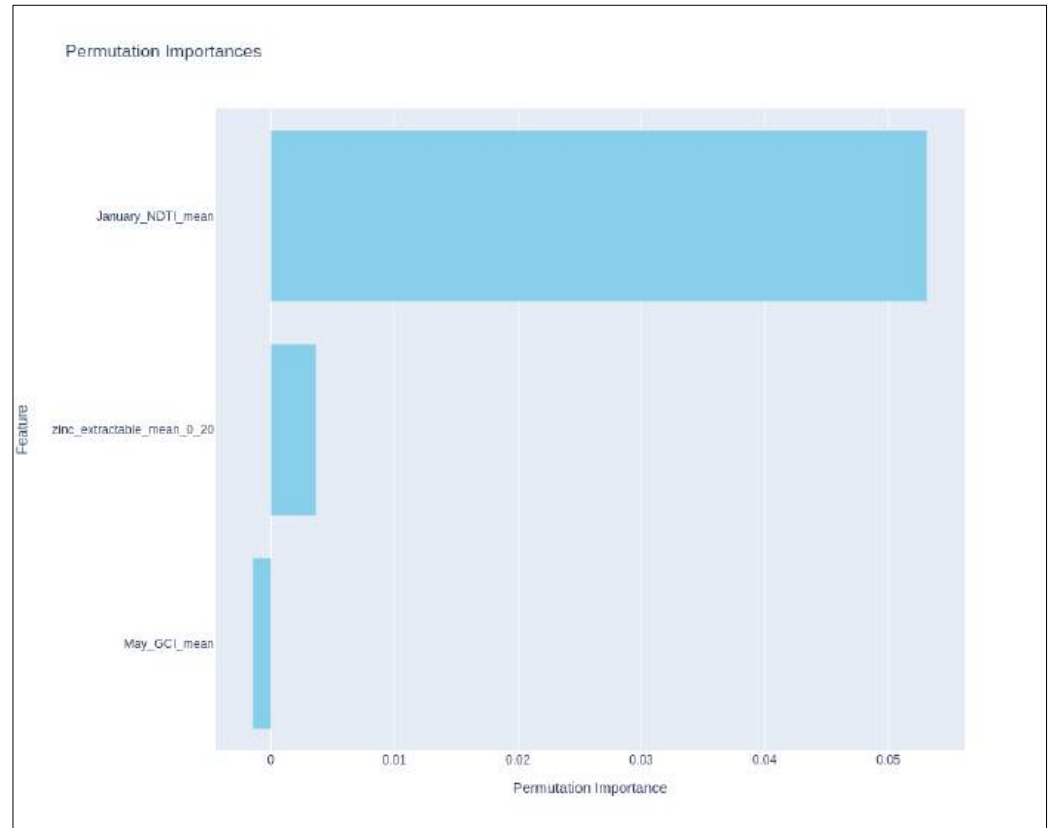
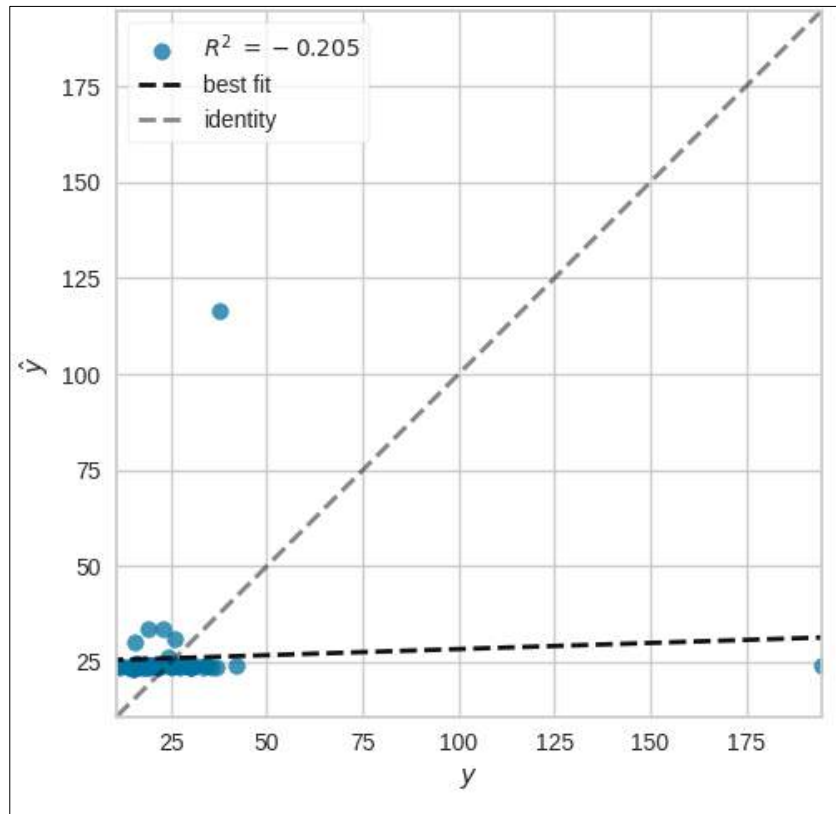


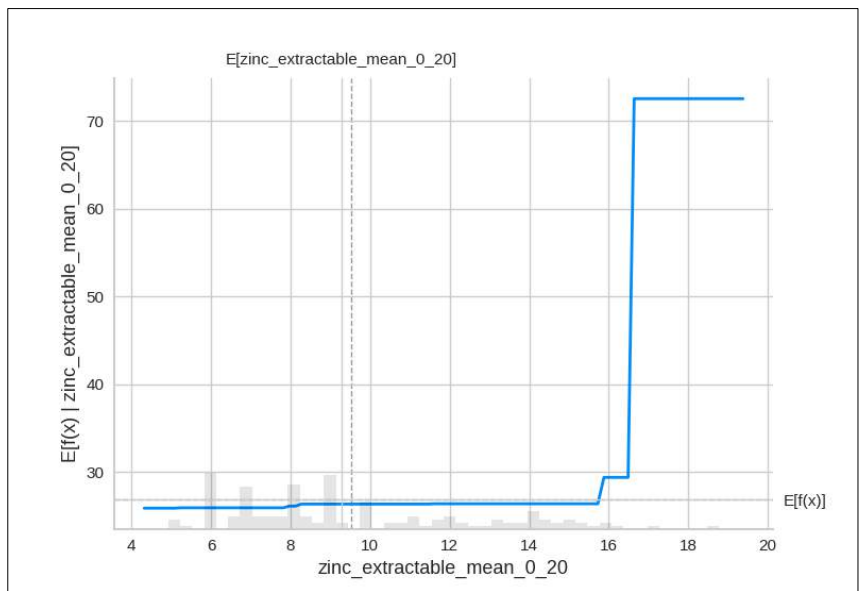
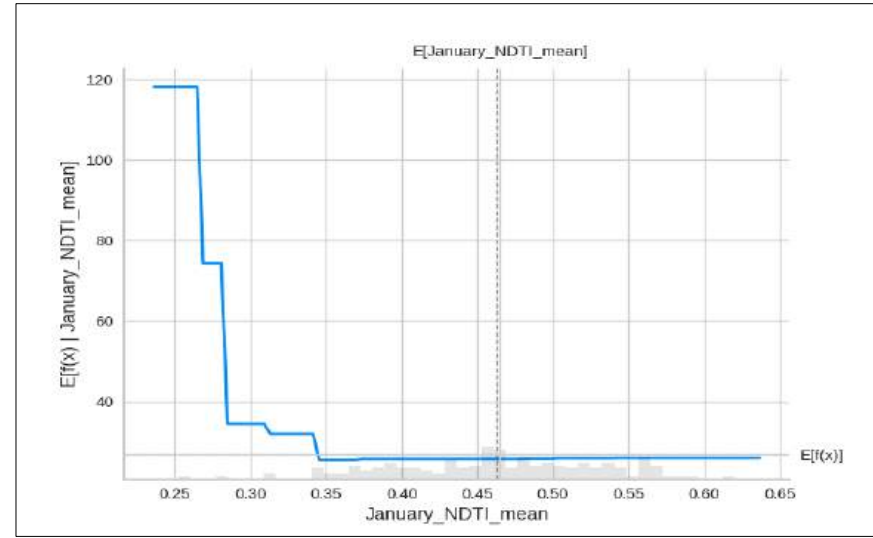
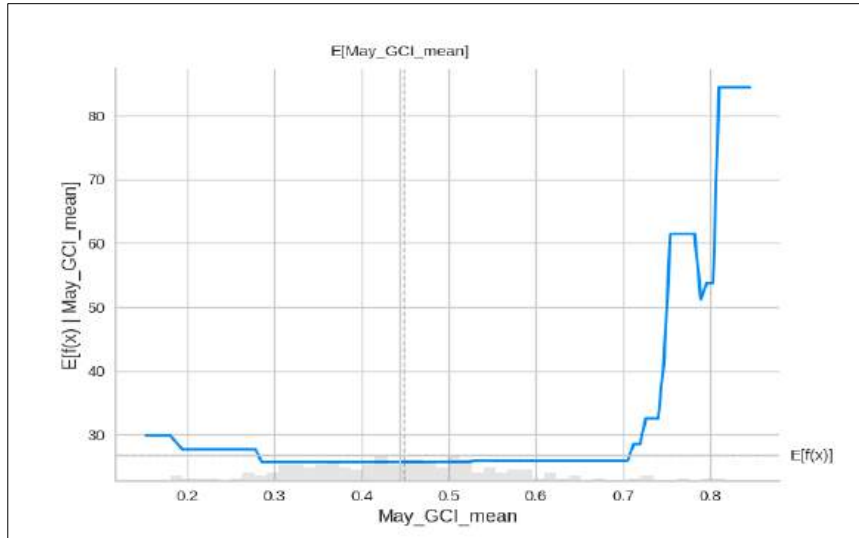
b) MAIZE AND COPPER NUTRIENT COMPOSITION OF GRAINS





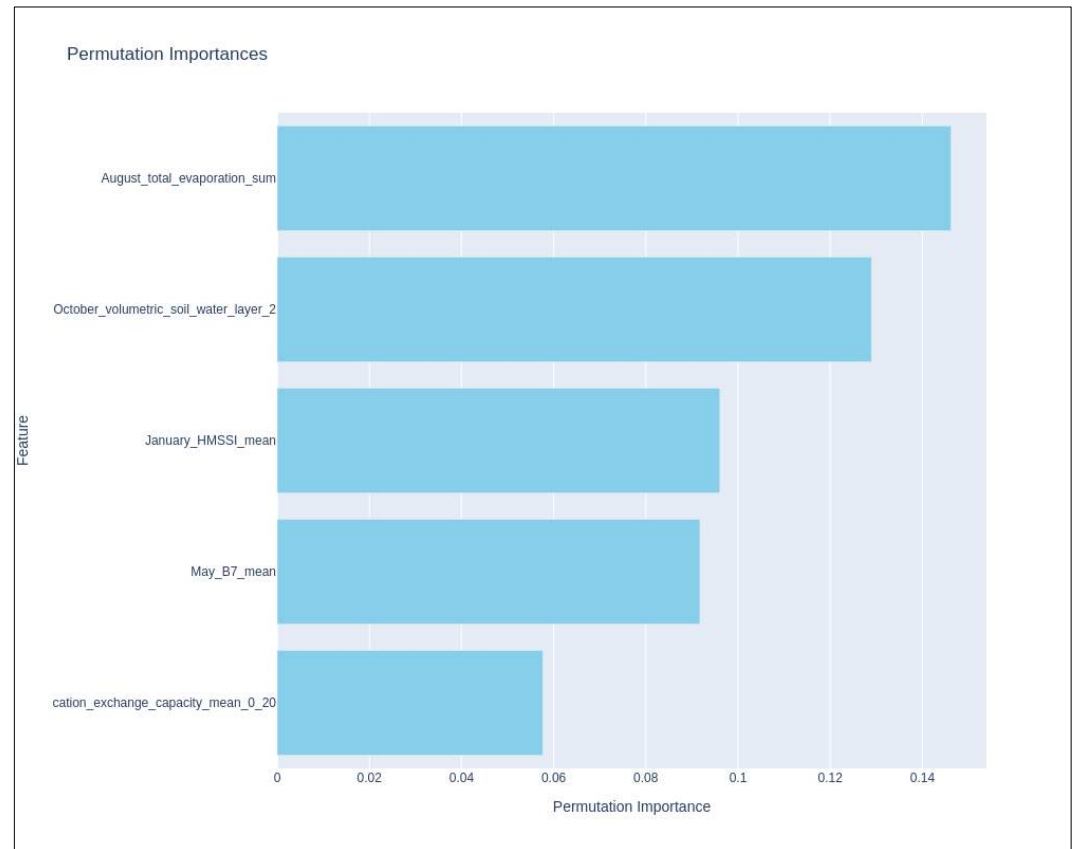
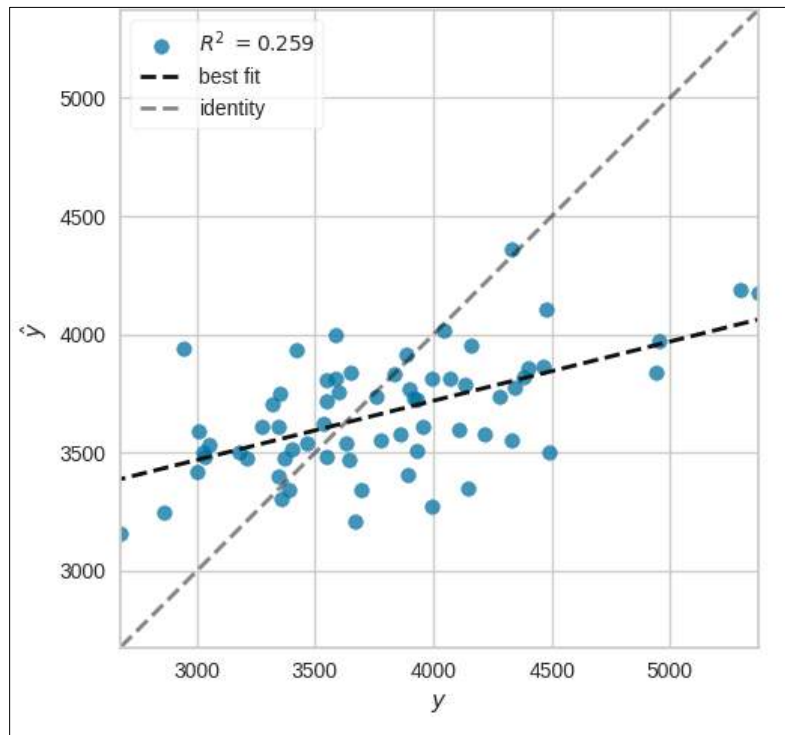
c) MAIZE AND IRON NUTRIENT COMPOSITION OF GRAINS

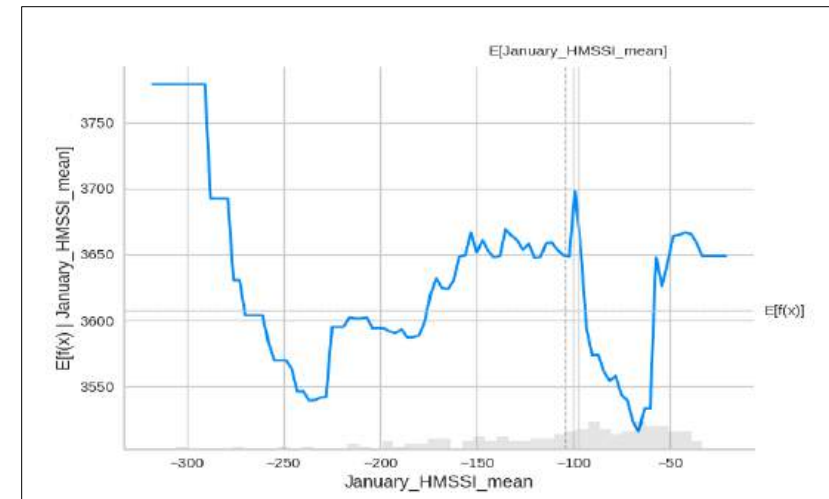
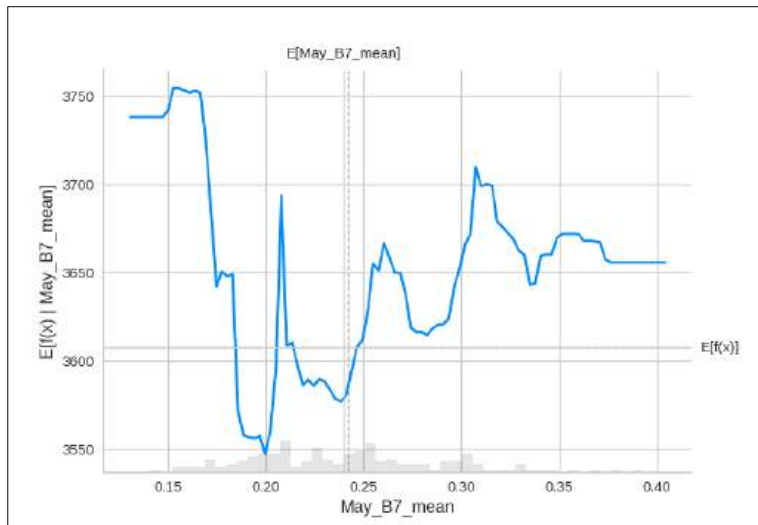
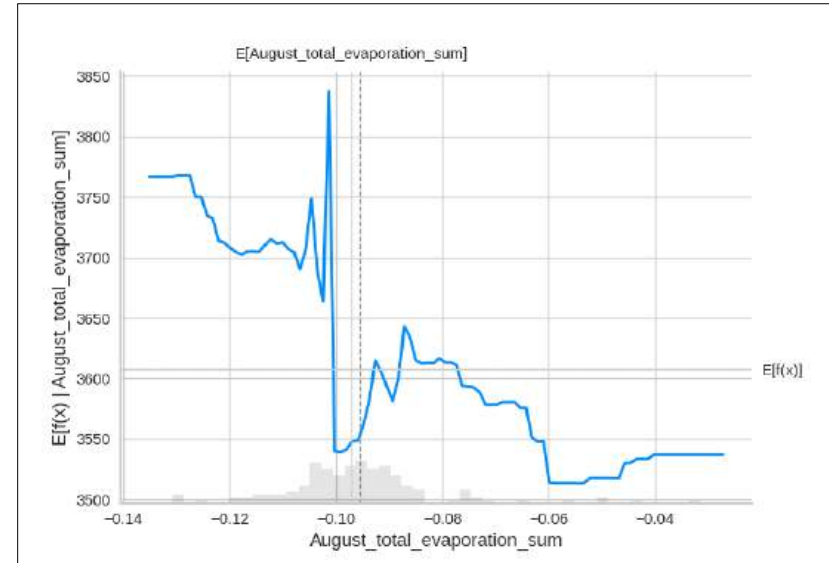
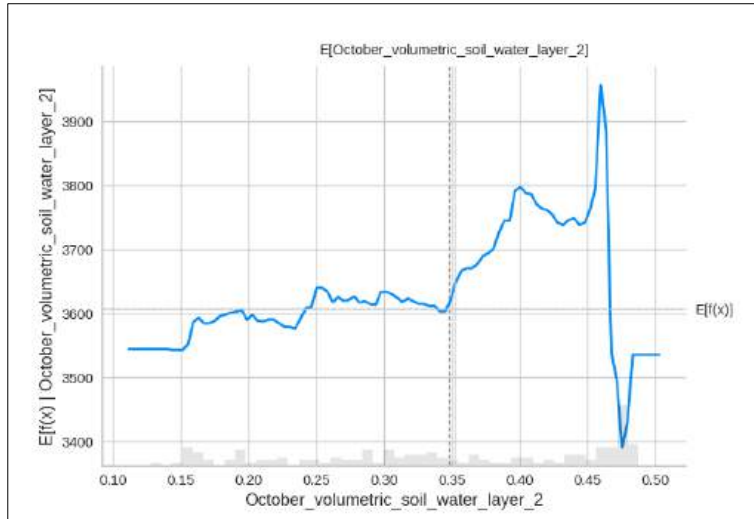


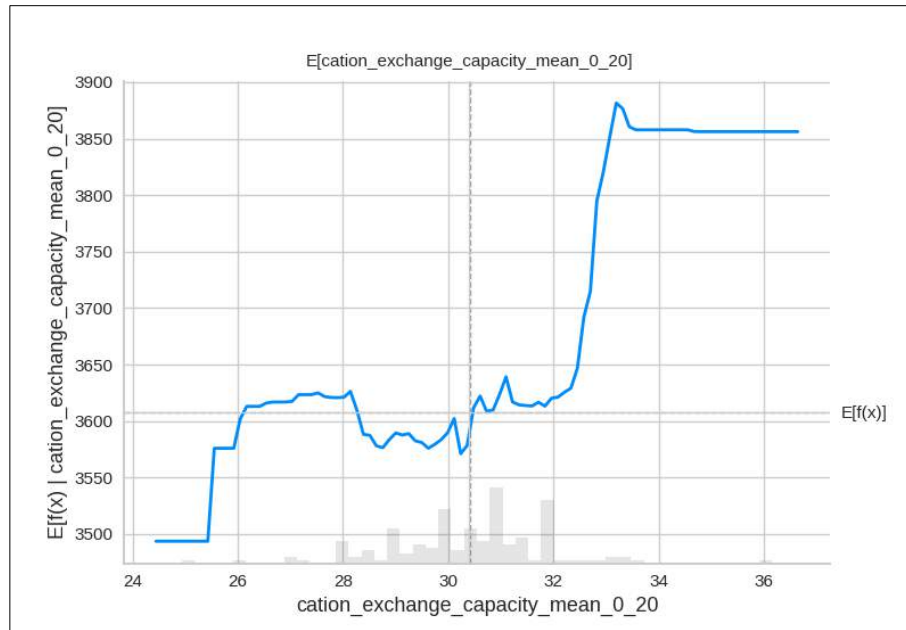




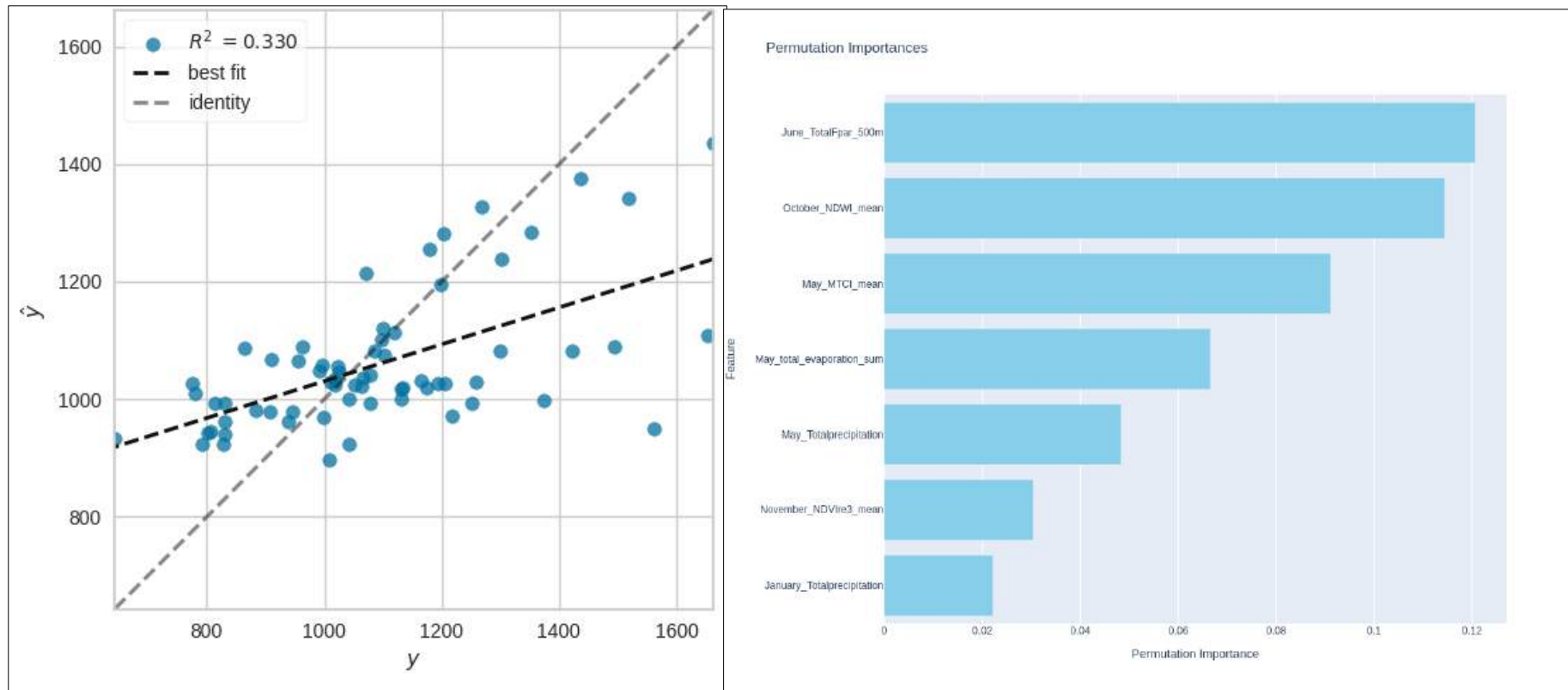
d) MAIZE AND POTASSIUM NUTRIENT COMPOSITION OF GRAINS

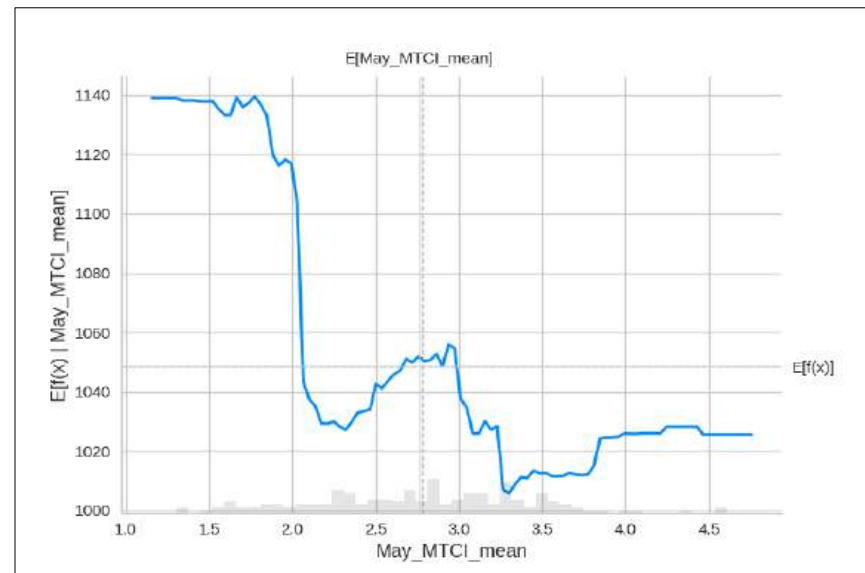
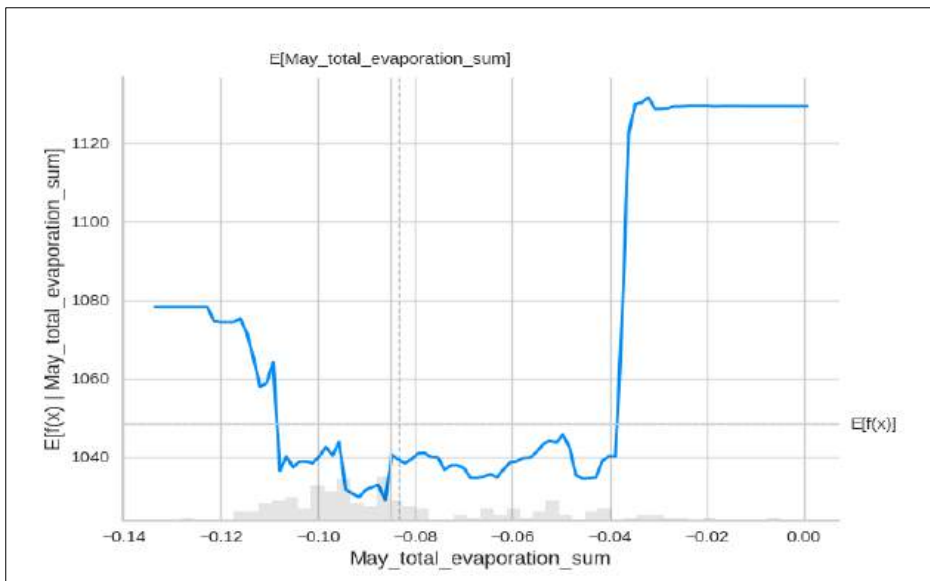
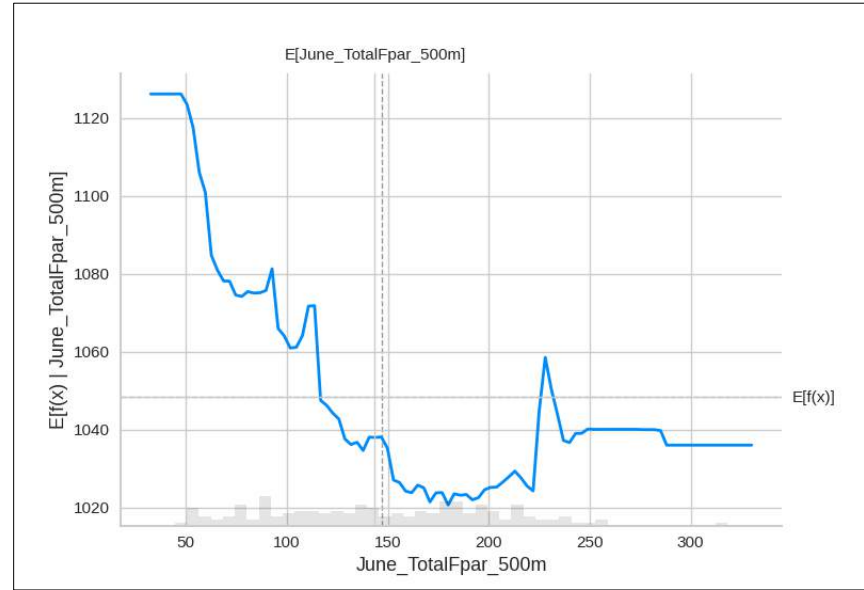
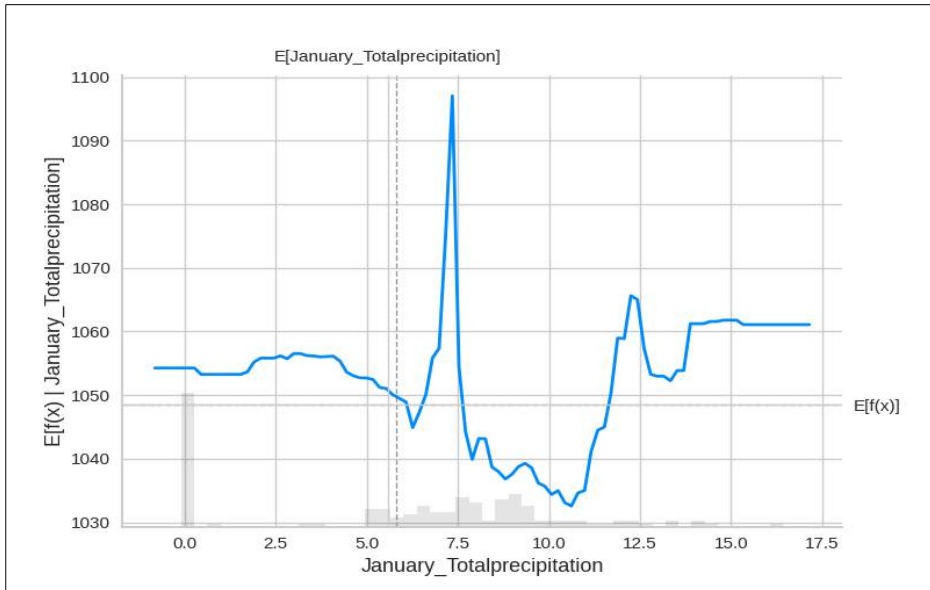


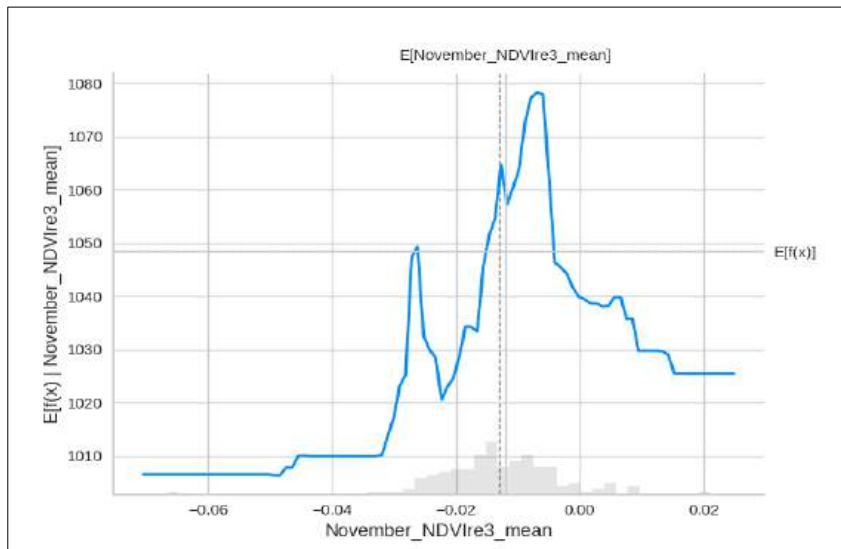
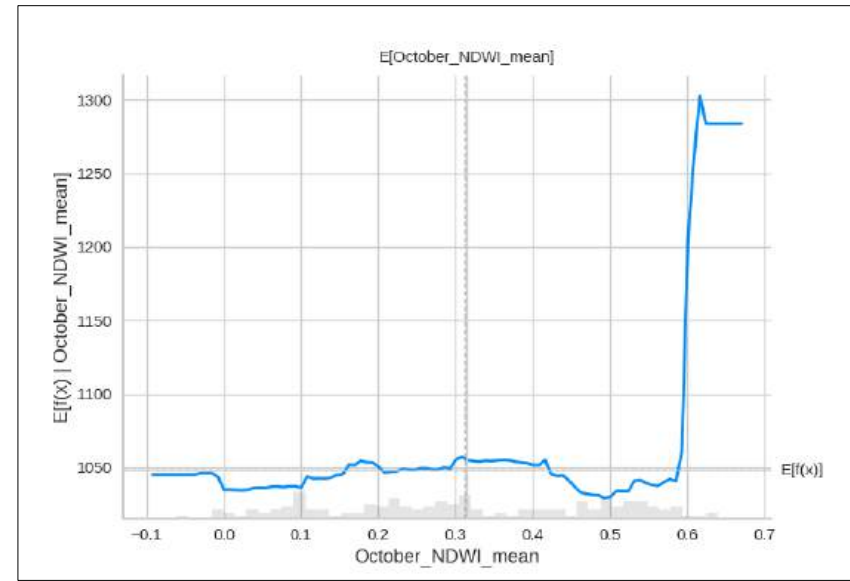
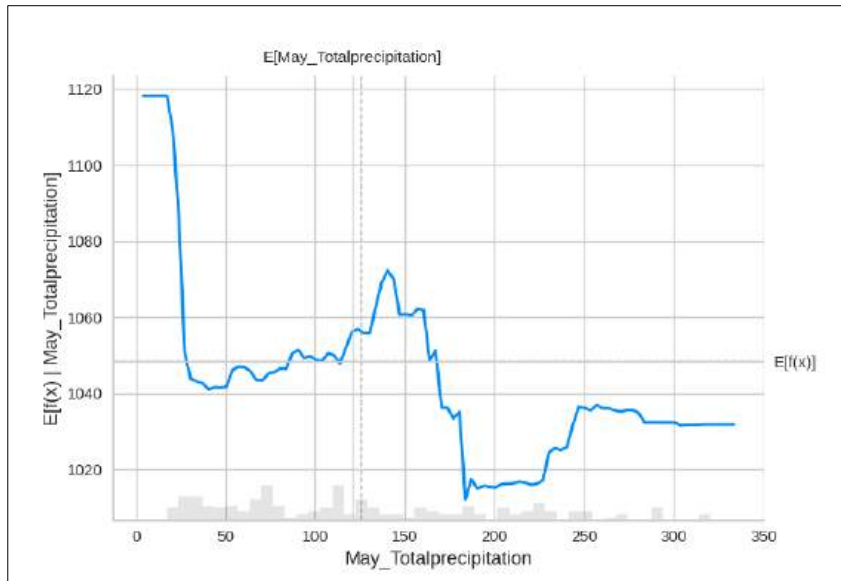




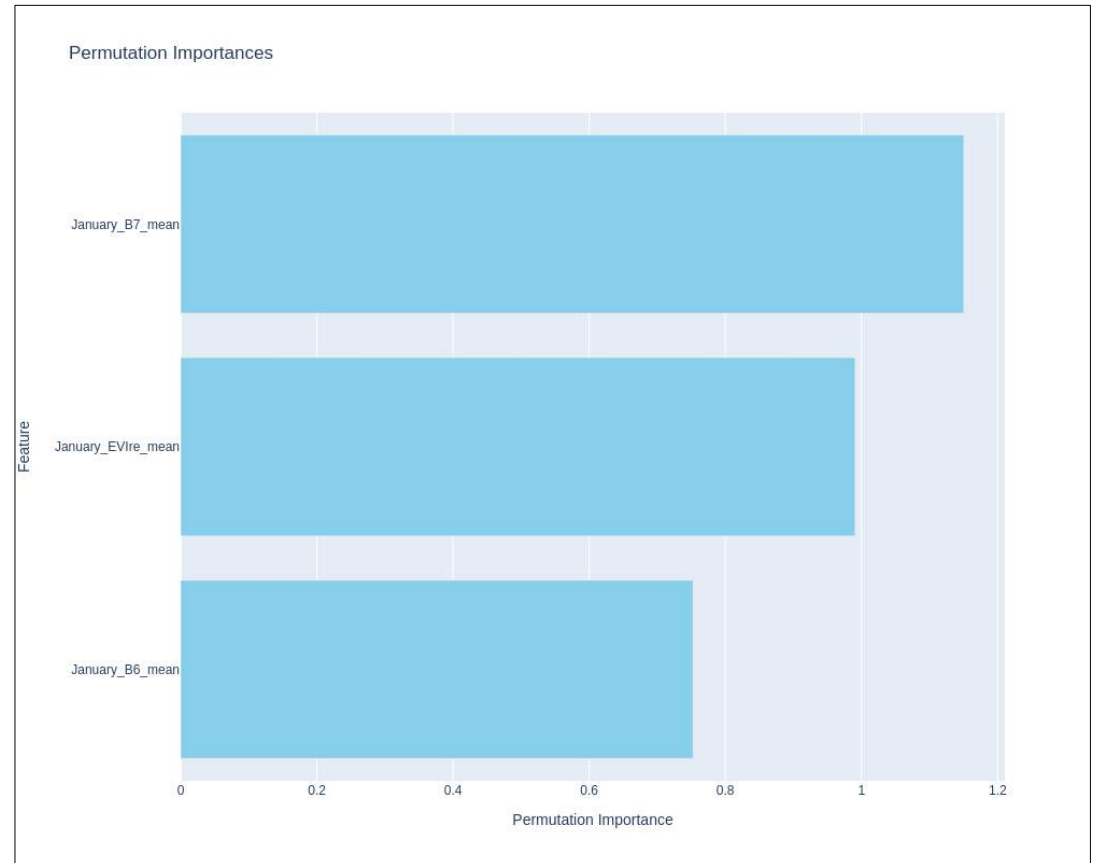
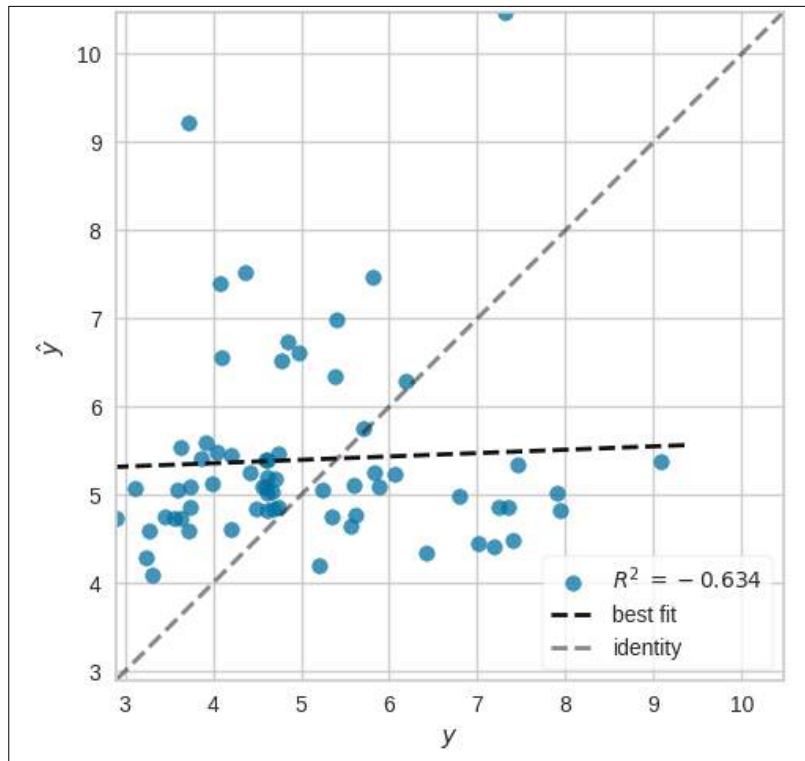
e) MAIZE AND MAGNESIUM NUTRIENT COMPOSITION OF GRAINS

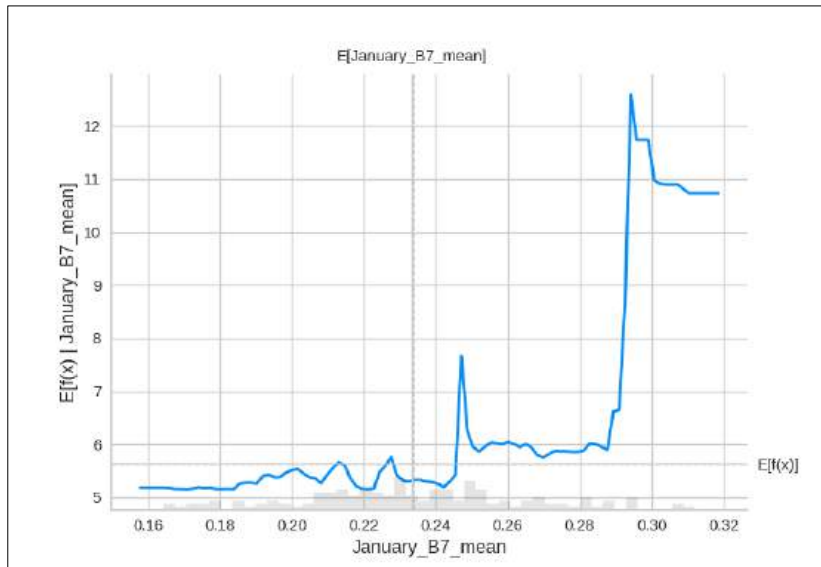
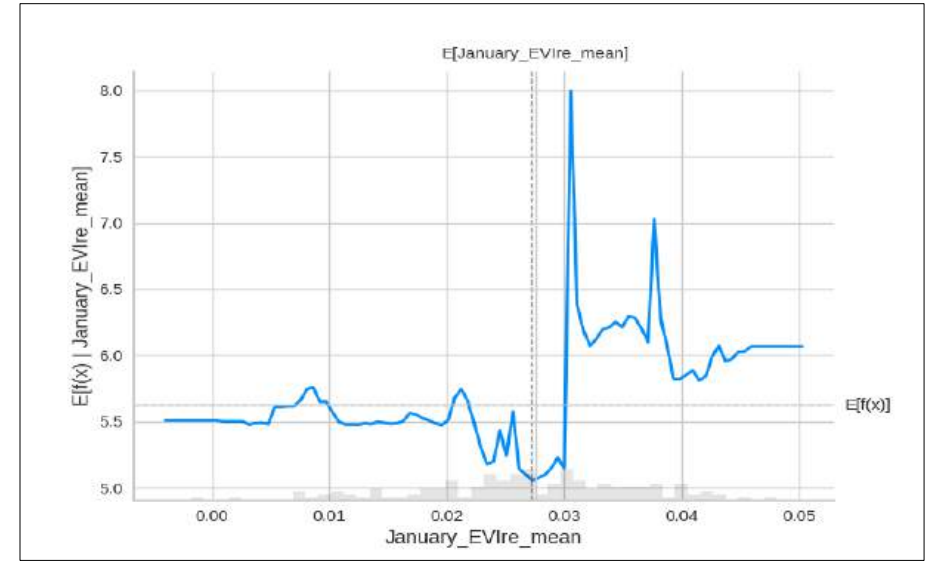
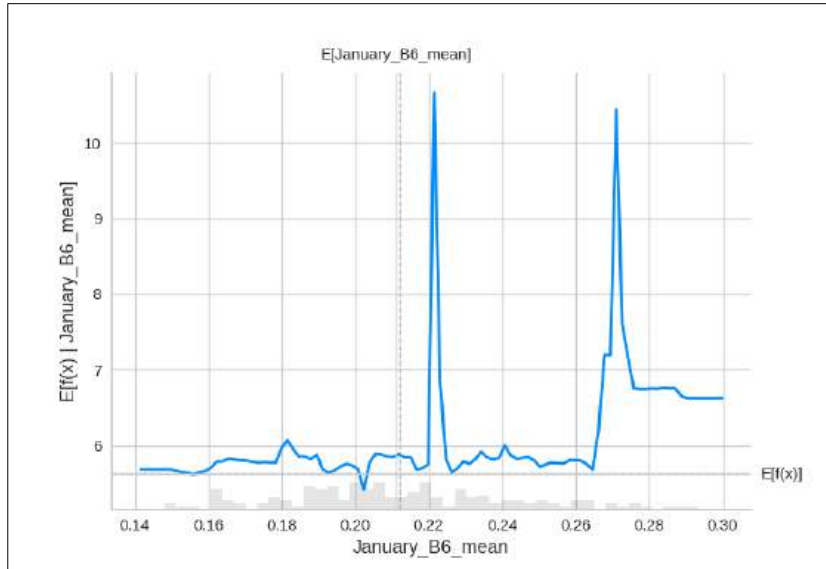






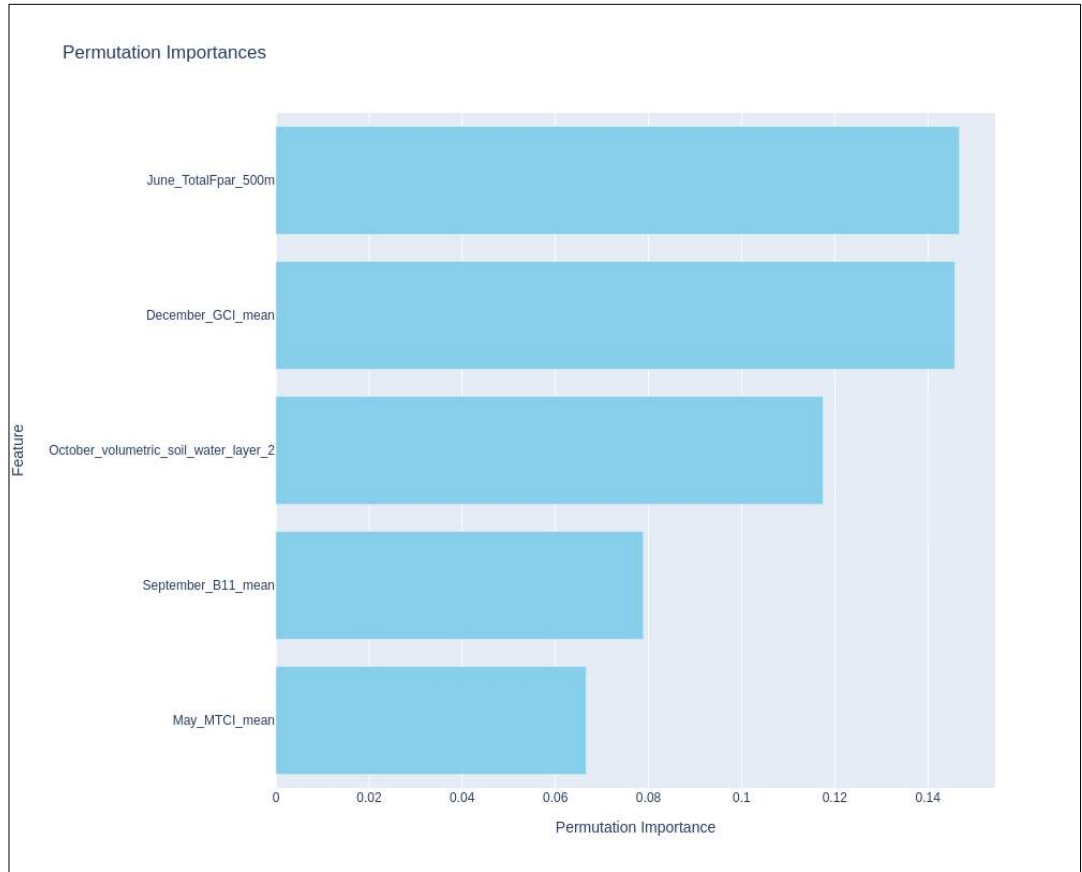
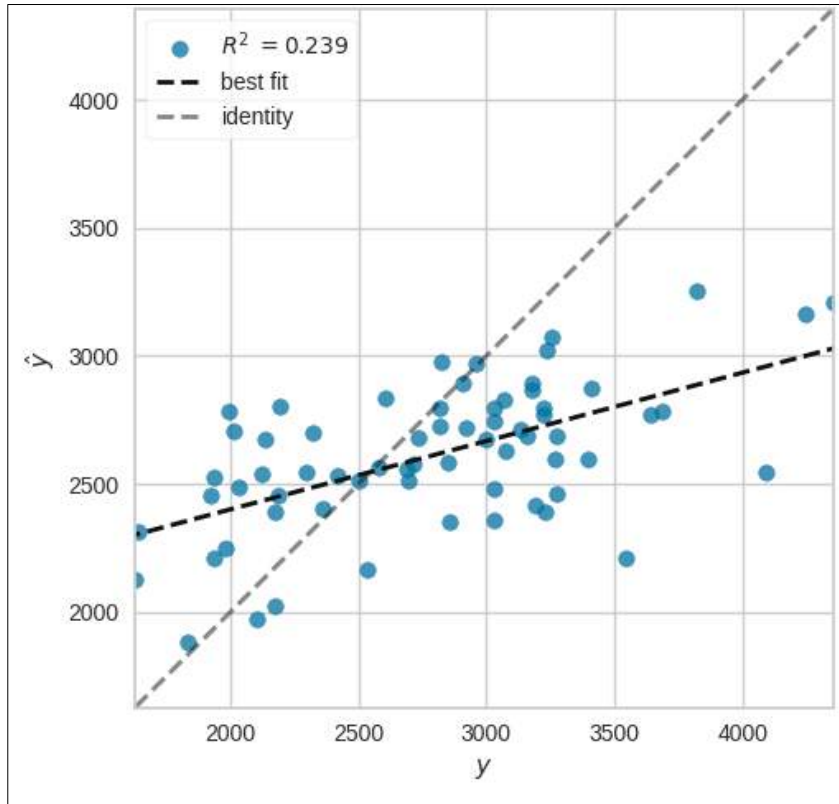
f) MAIZE AND MANGANESE NUTRIENT COMPOSITION OF GRAINS

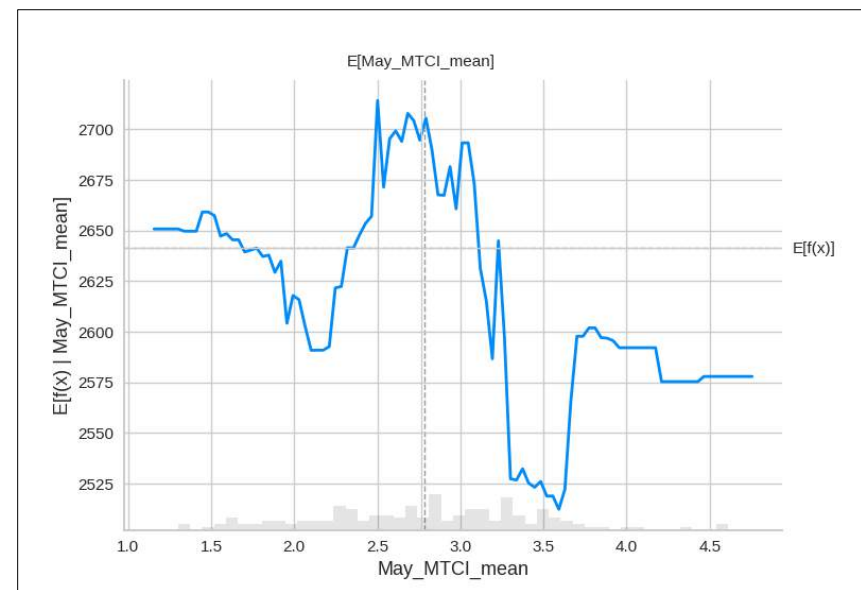
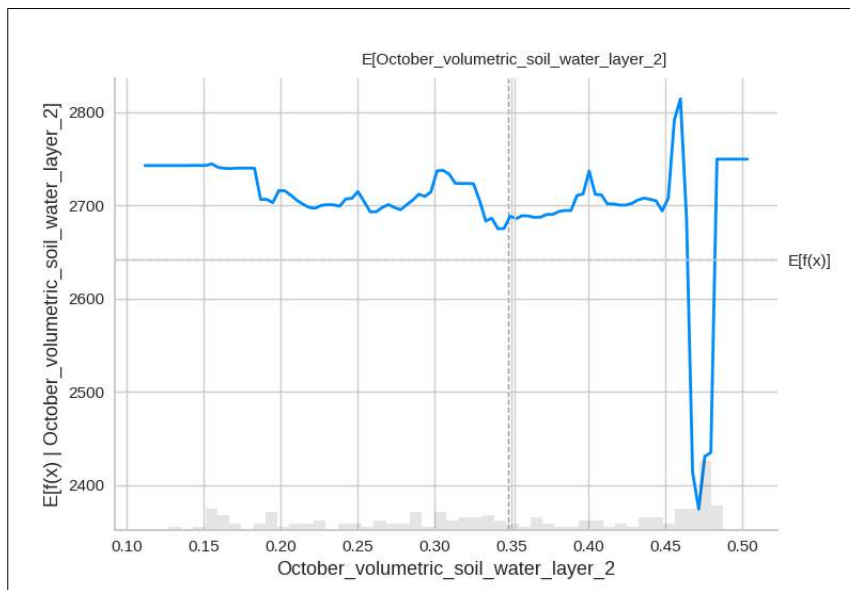
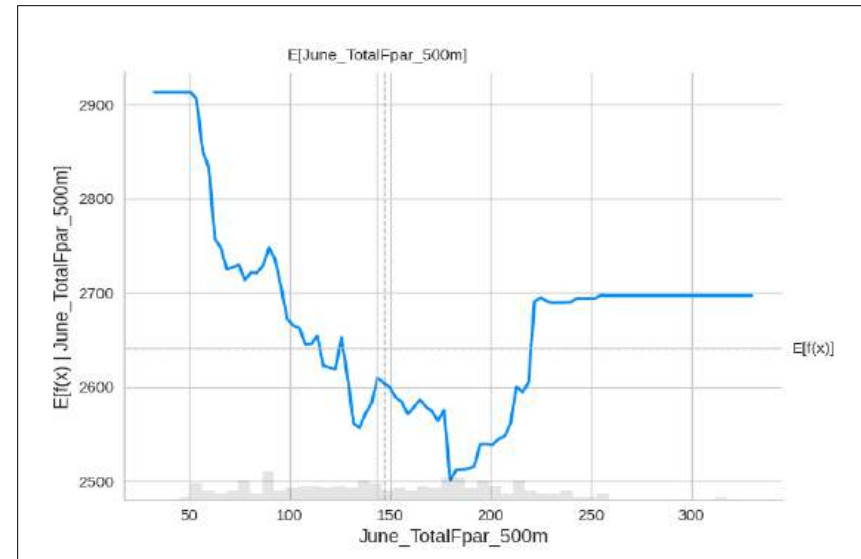
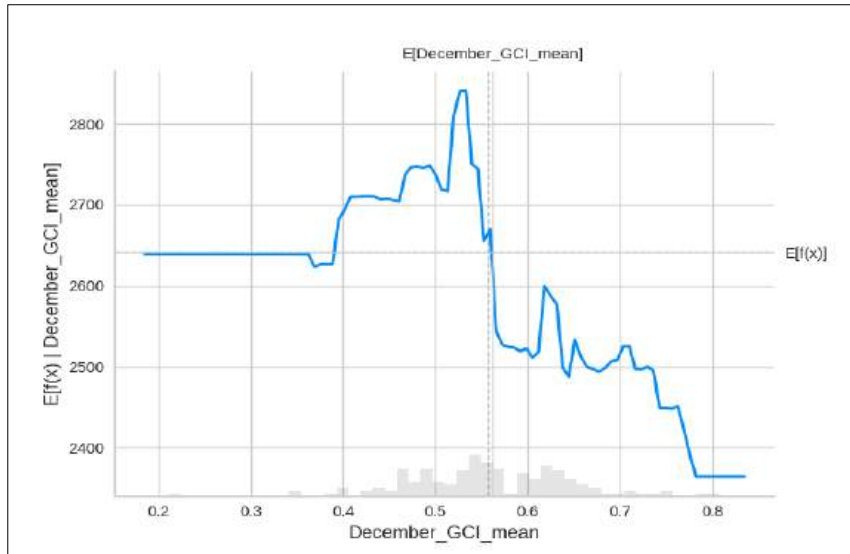


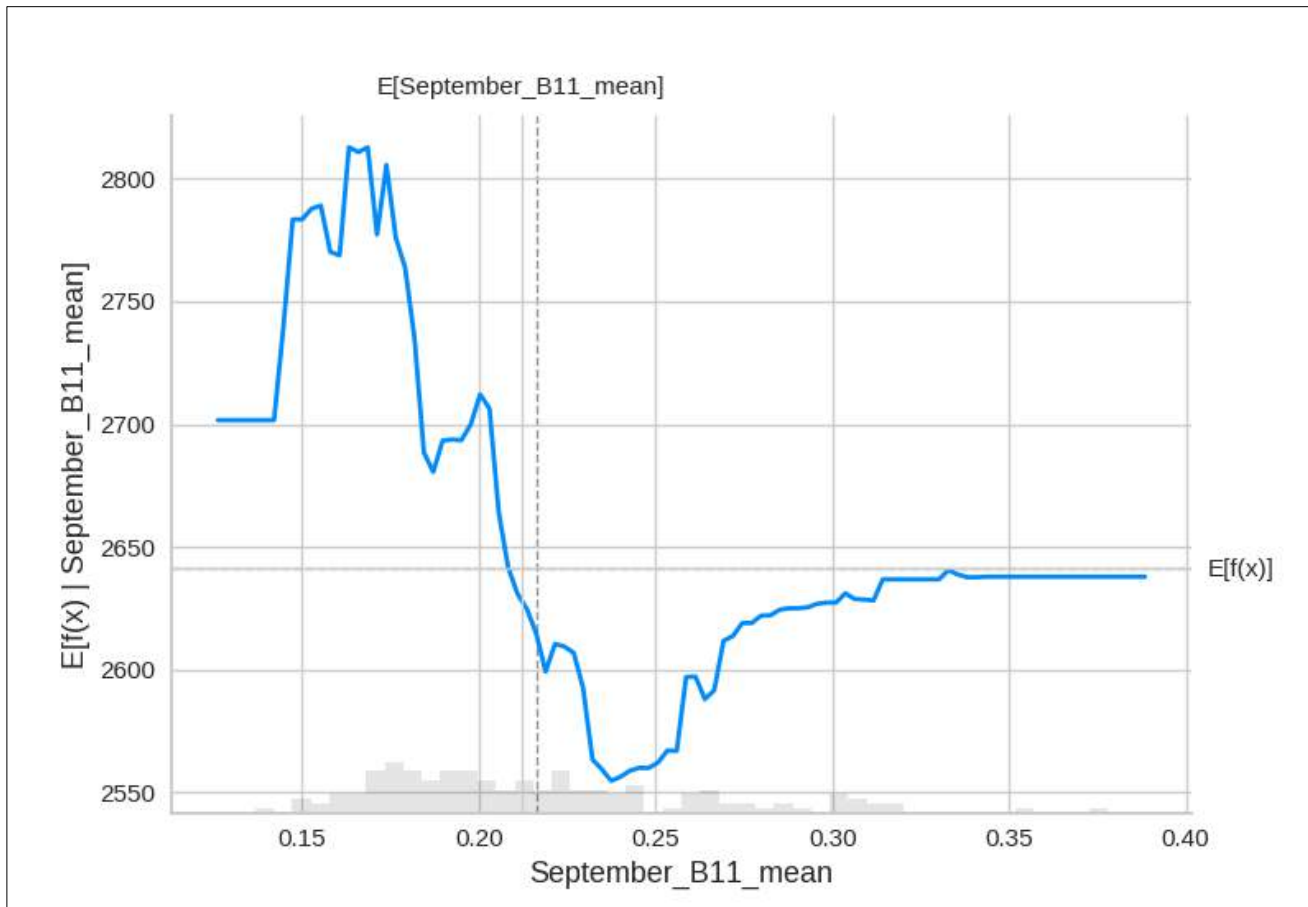




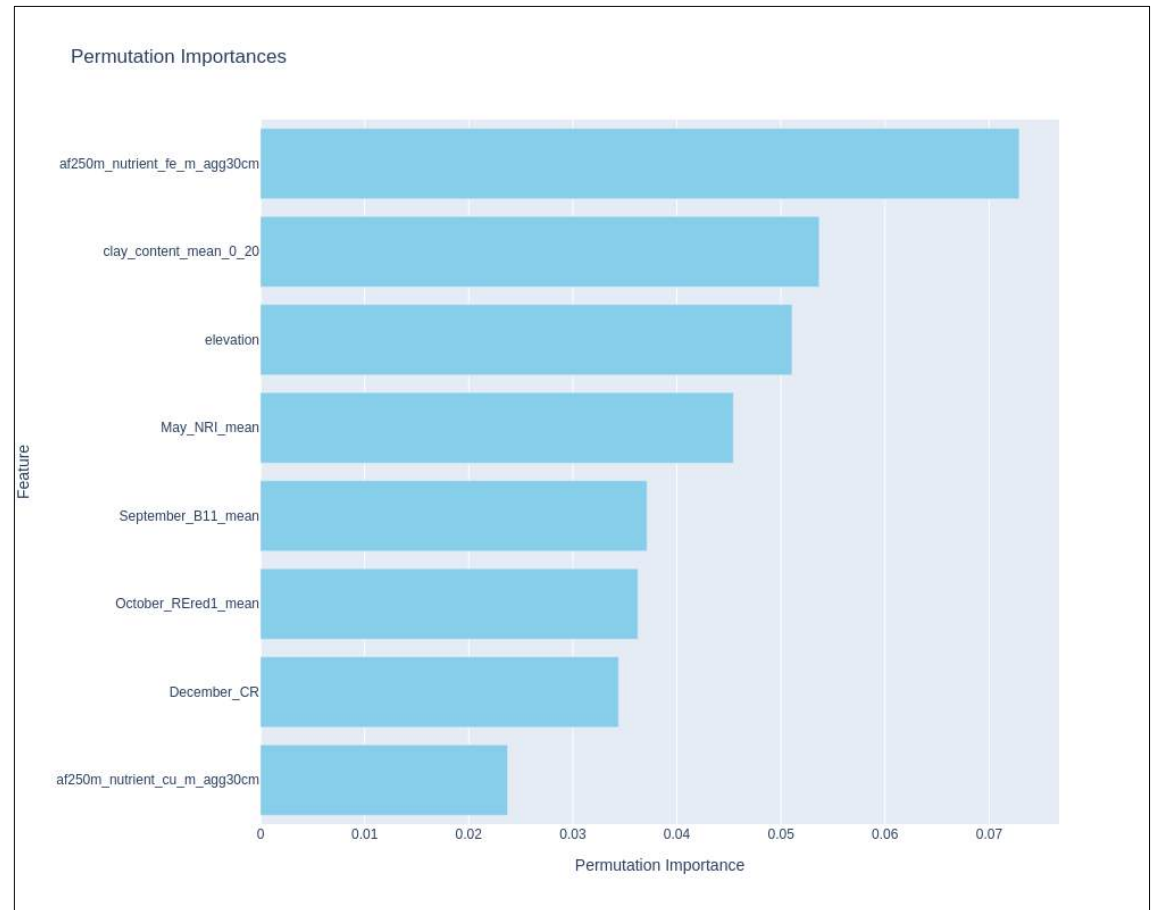
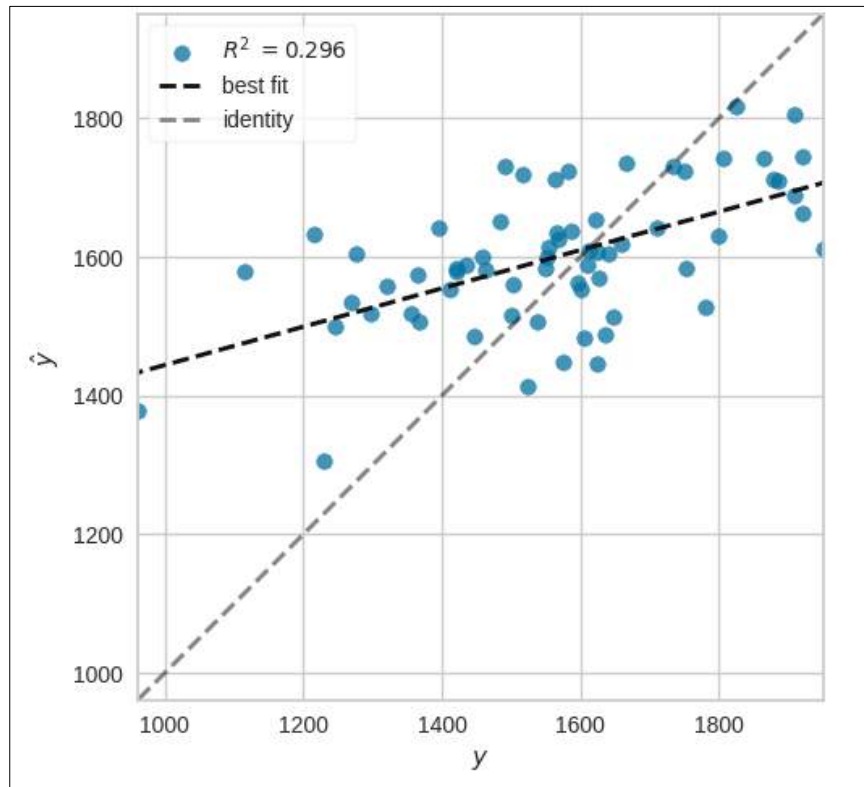
g) MAIZE AND PHOSPHORUS NUTRIENT COMPOSITION OF GRAINS

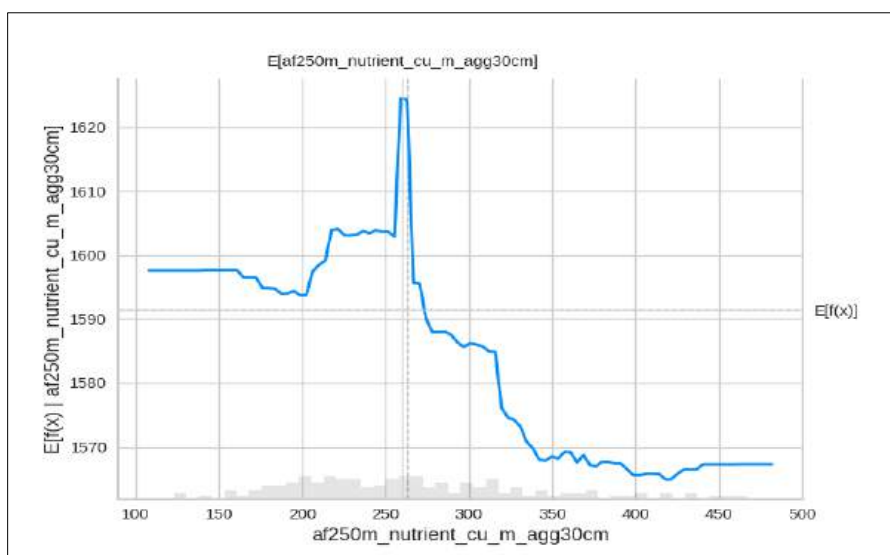
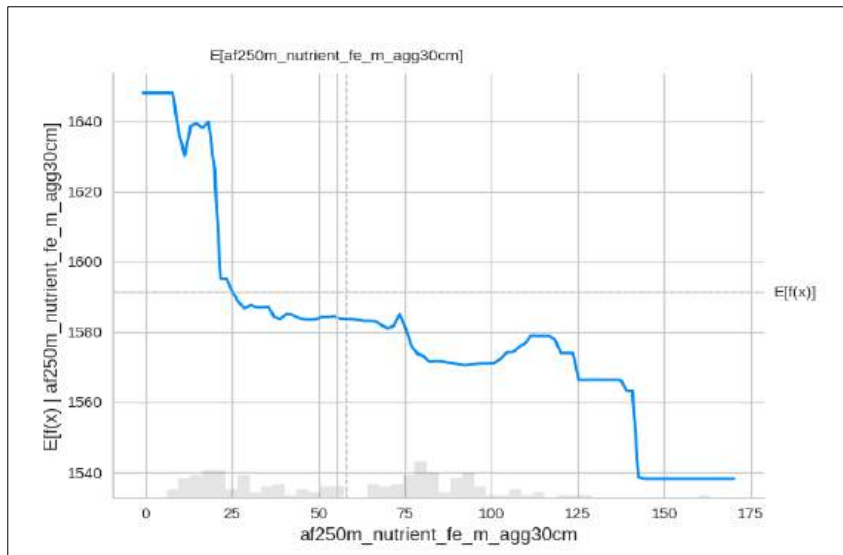
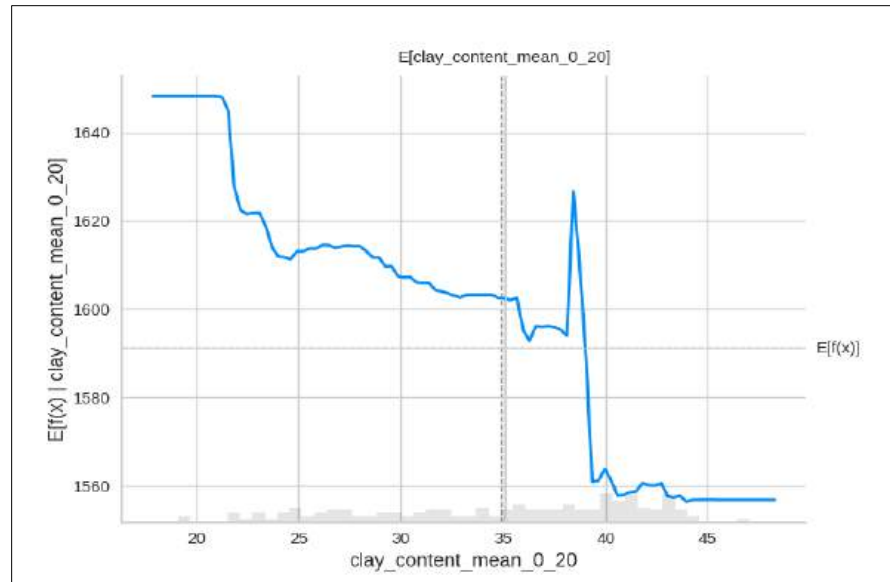
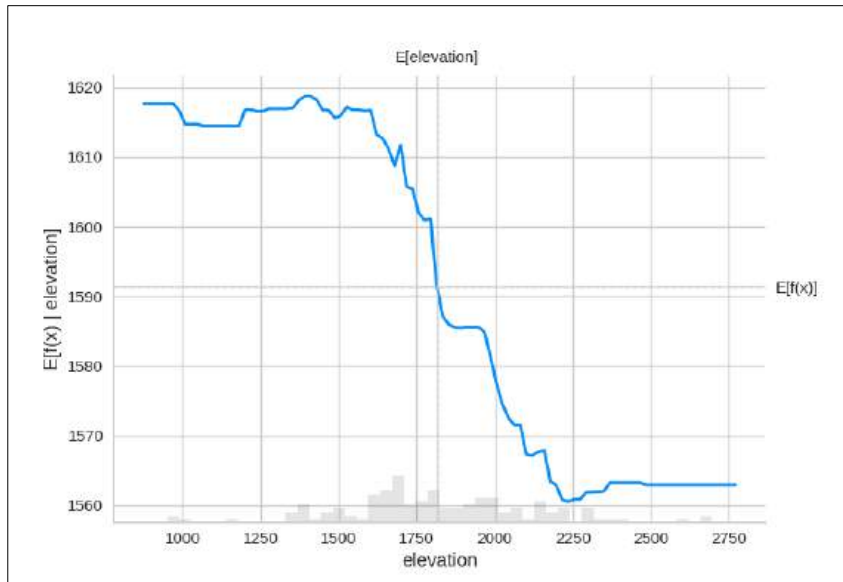


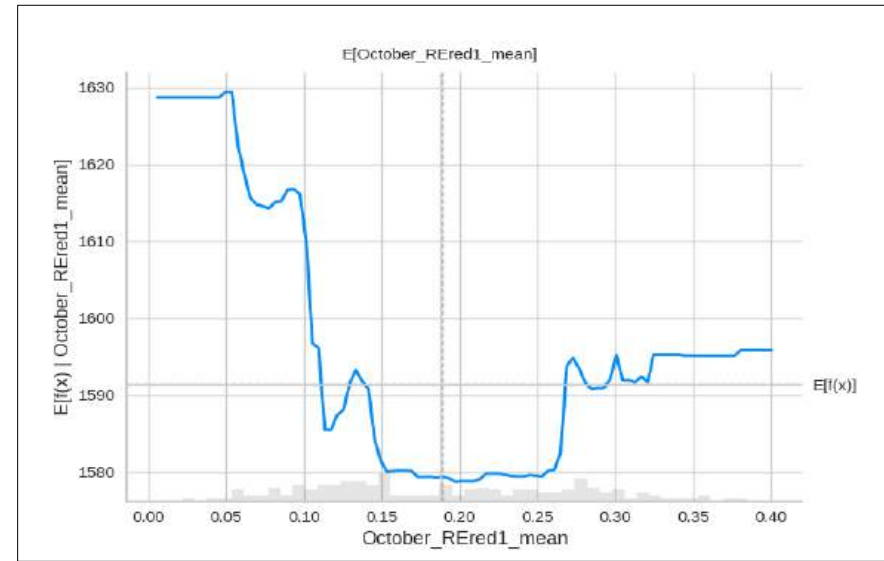
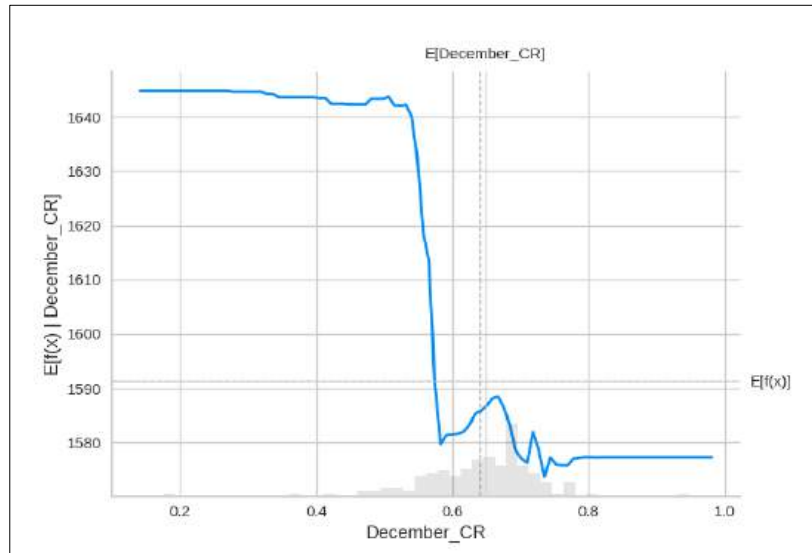


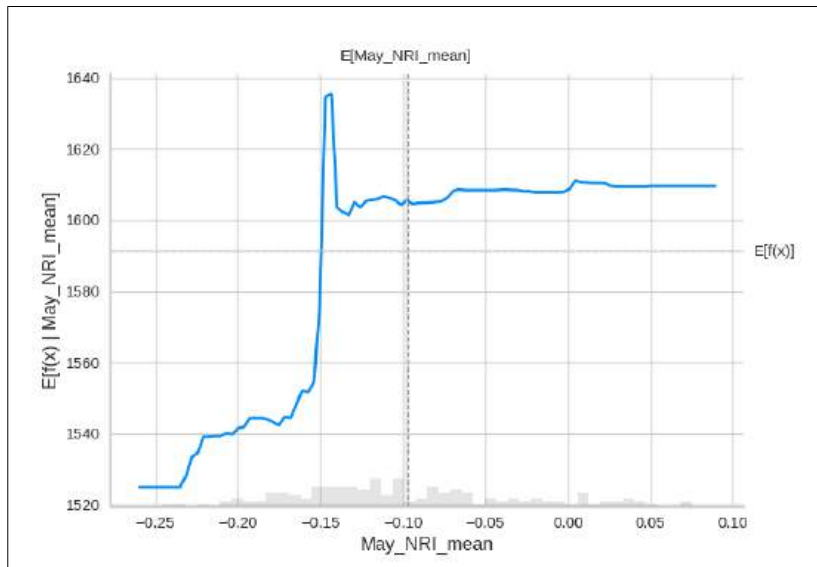


### h) MAIZE AND SULPHUR NUTRIENT COMPOSITION OF GRAINS

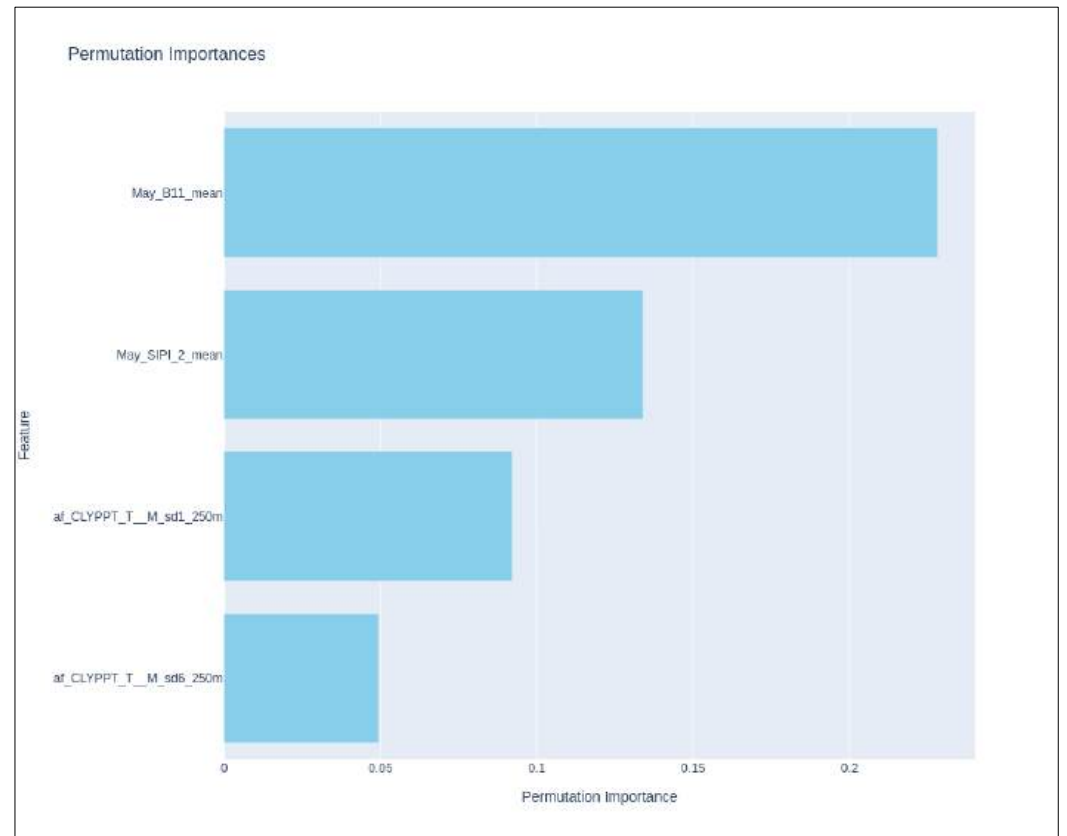
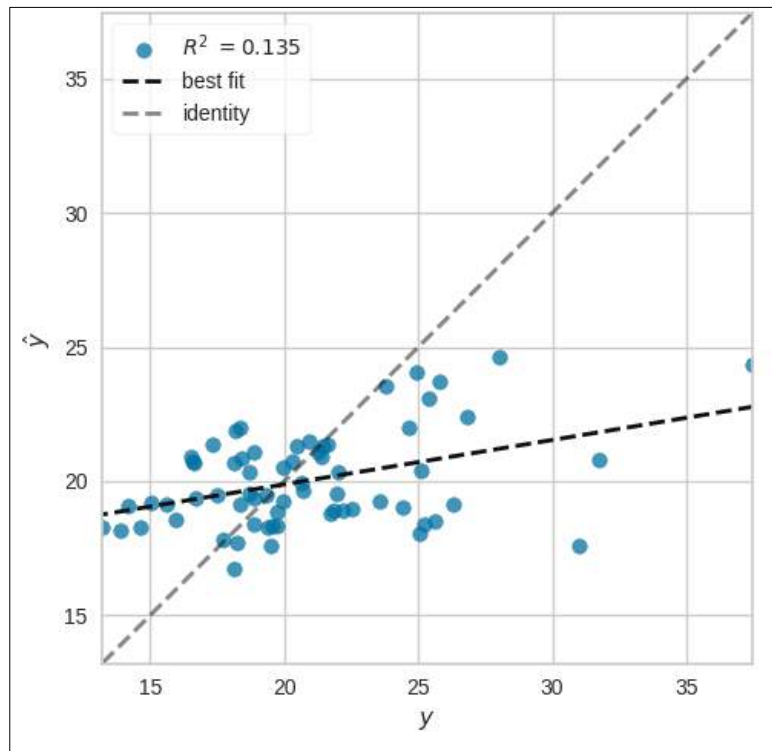




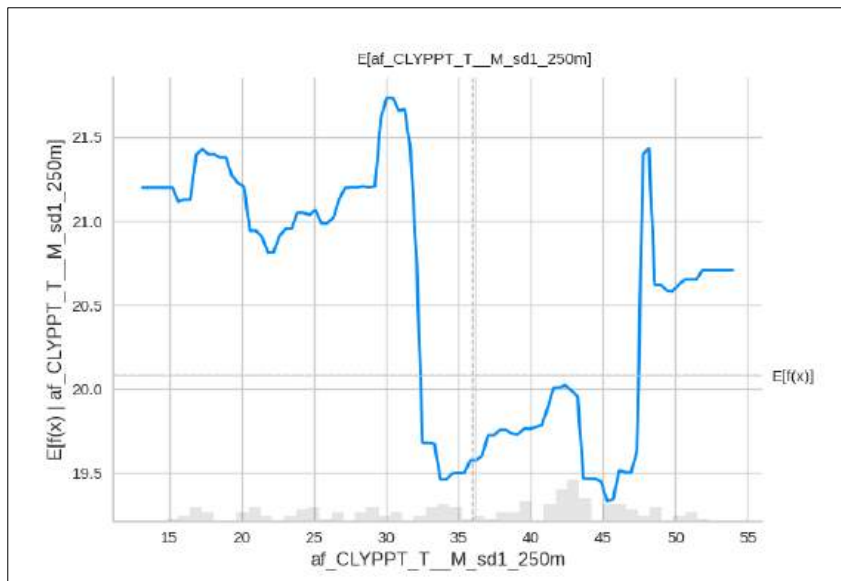
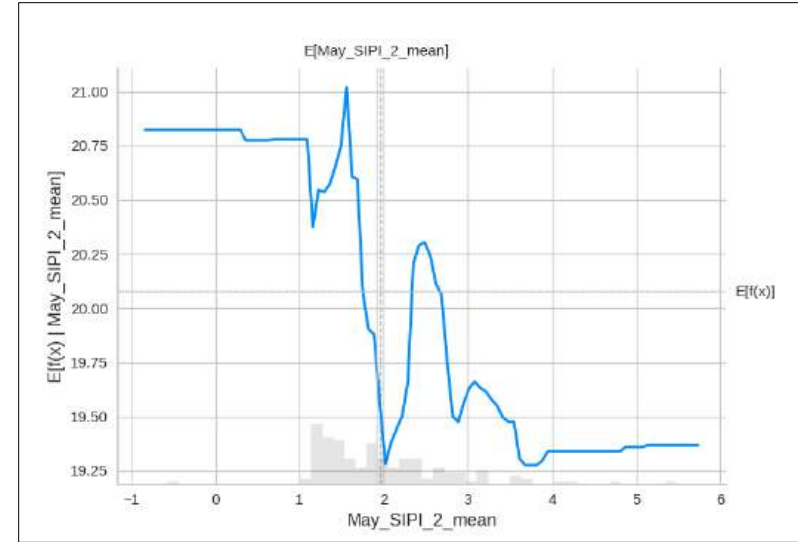
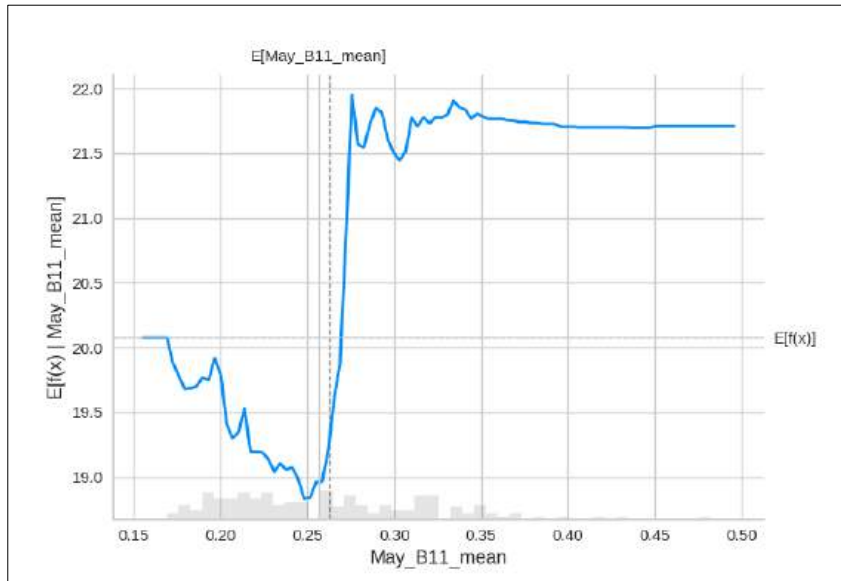




i) MAIZE AND ZINC NUTRIENT COMPOSITION OF GRAINS

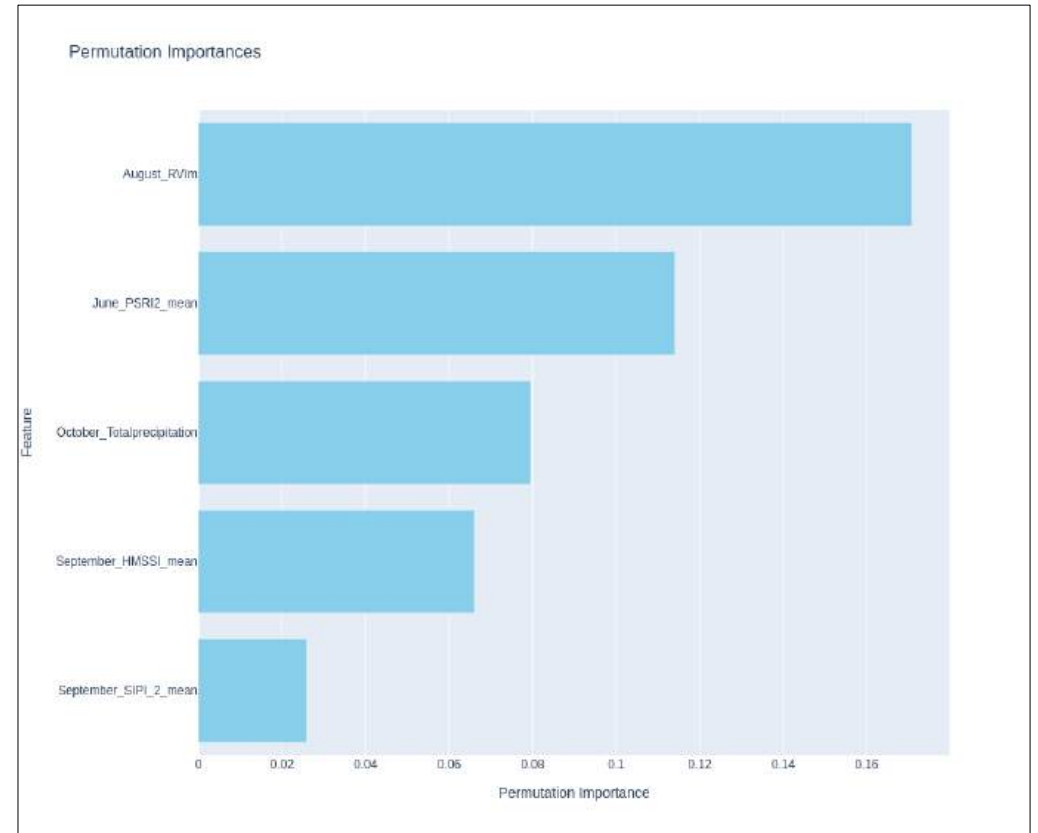
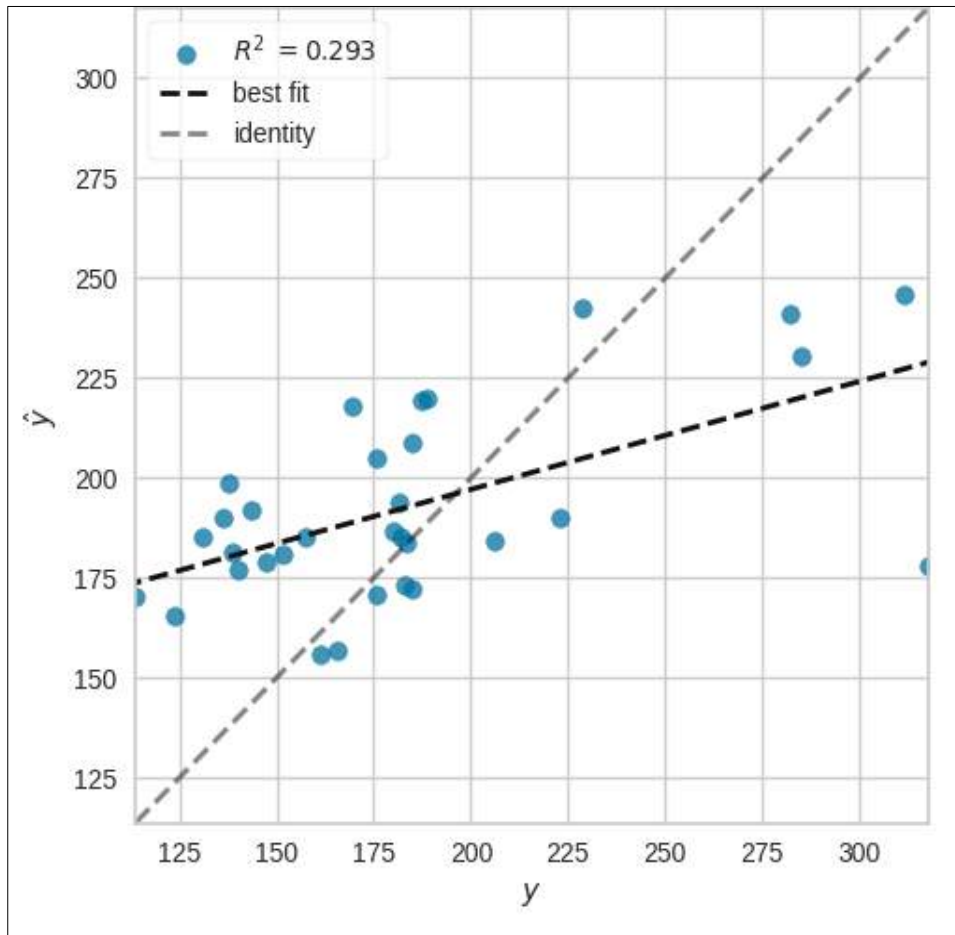


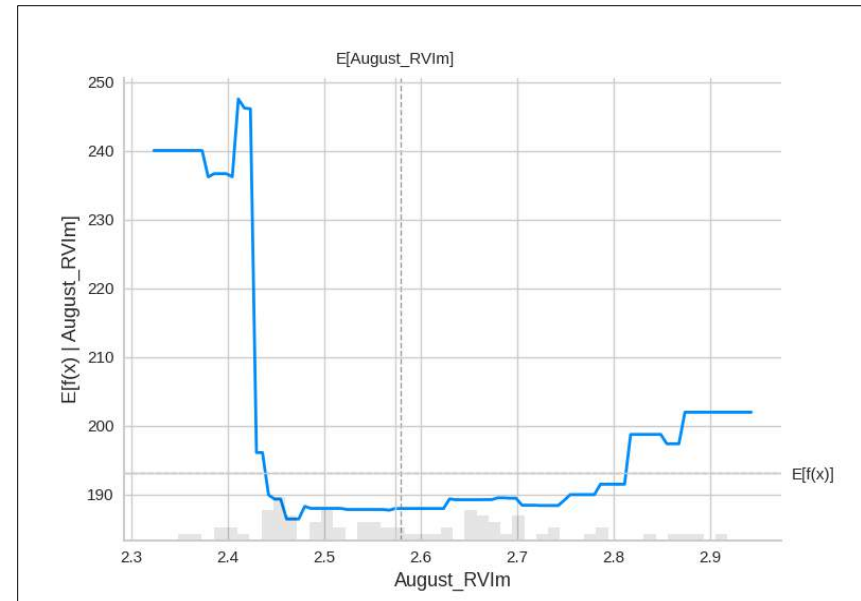
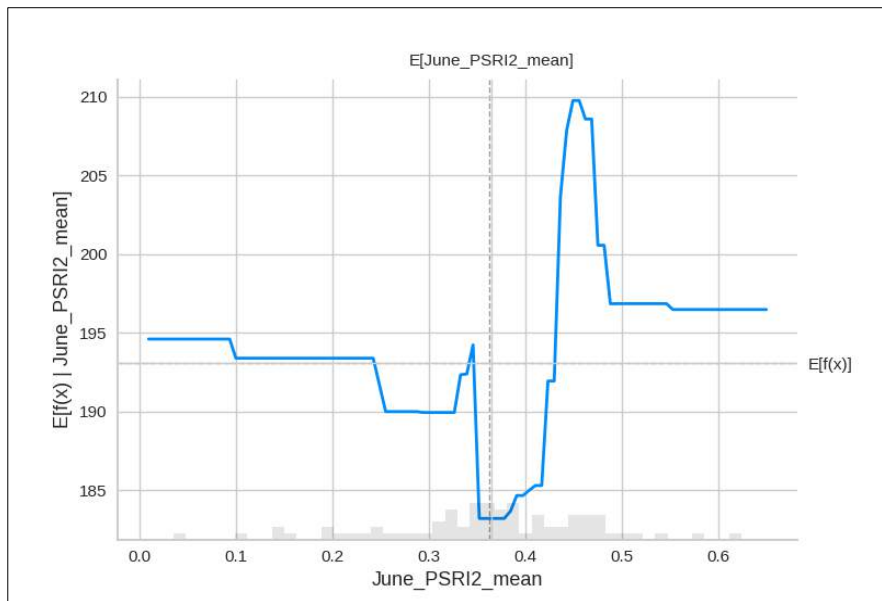
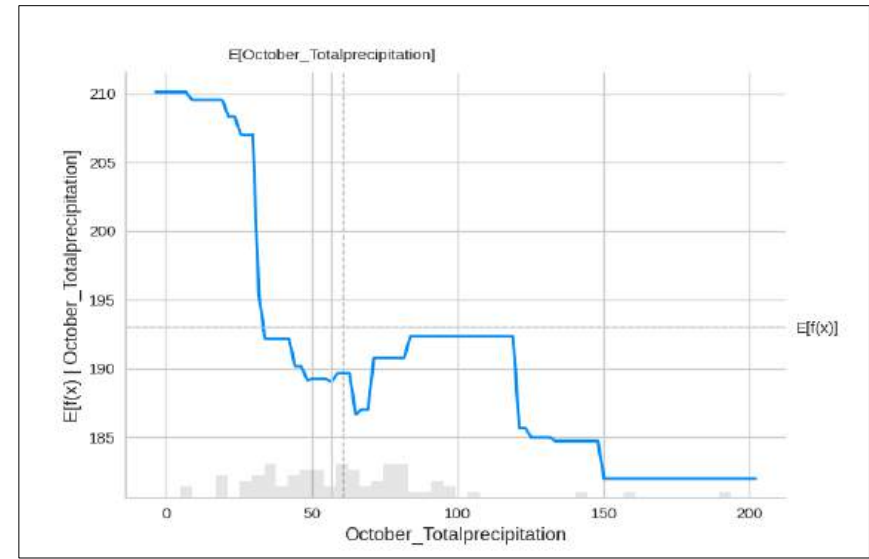
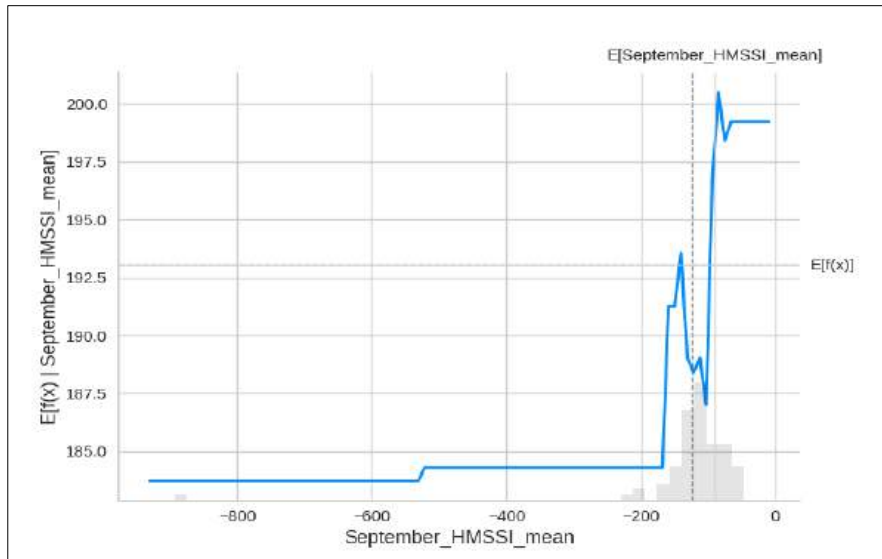


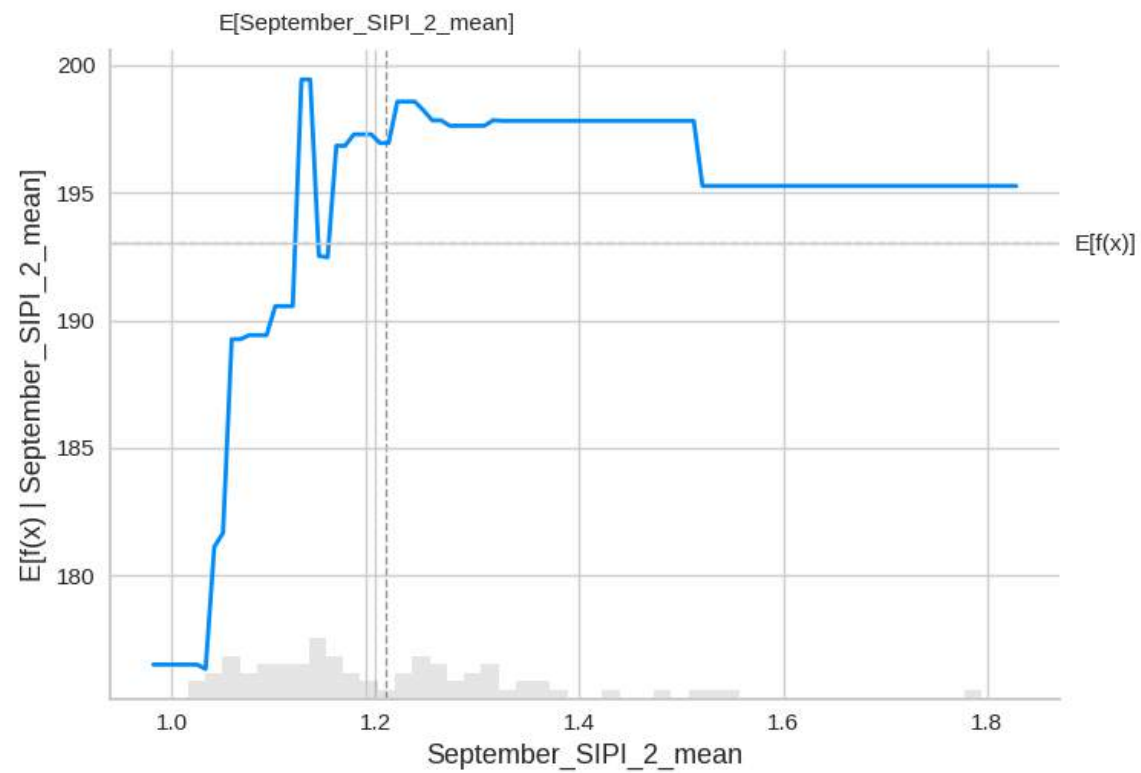


## Appendix 4. Results for sorghum

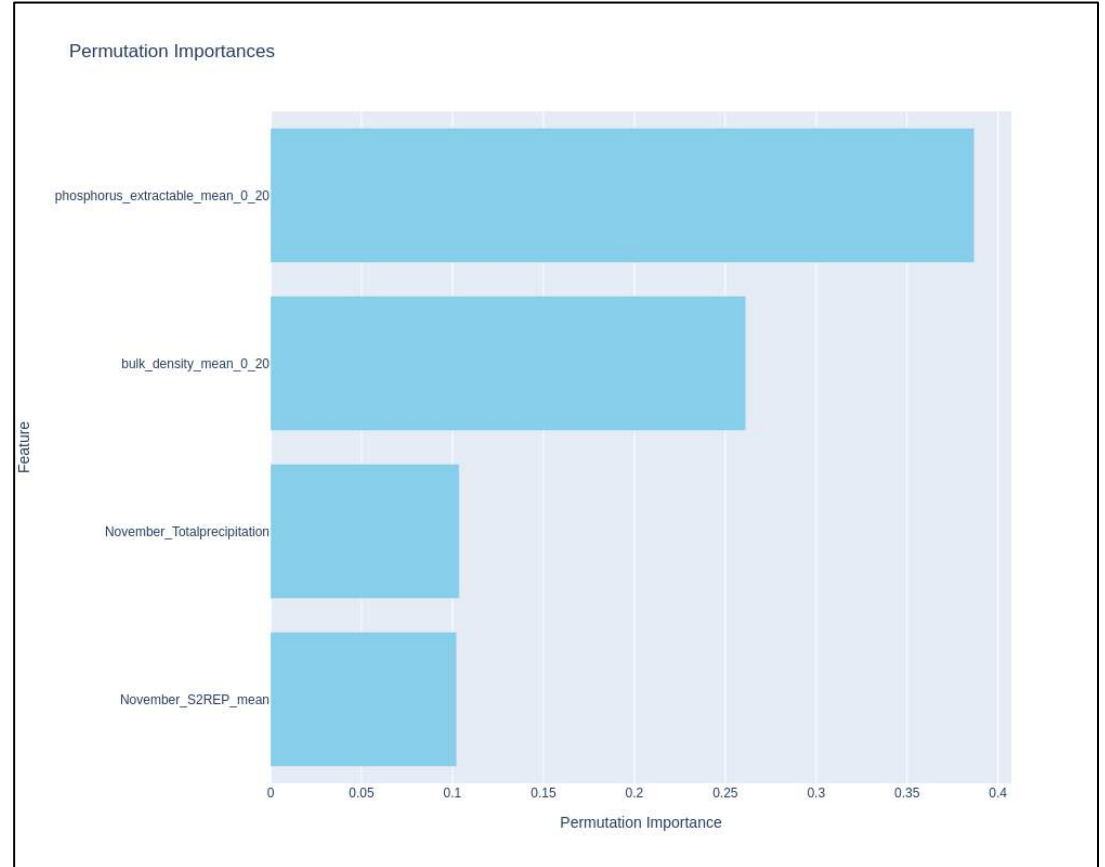
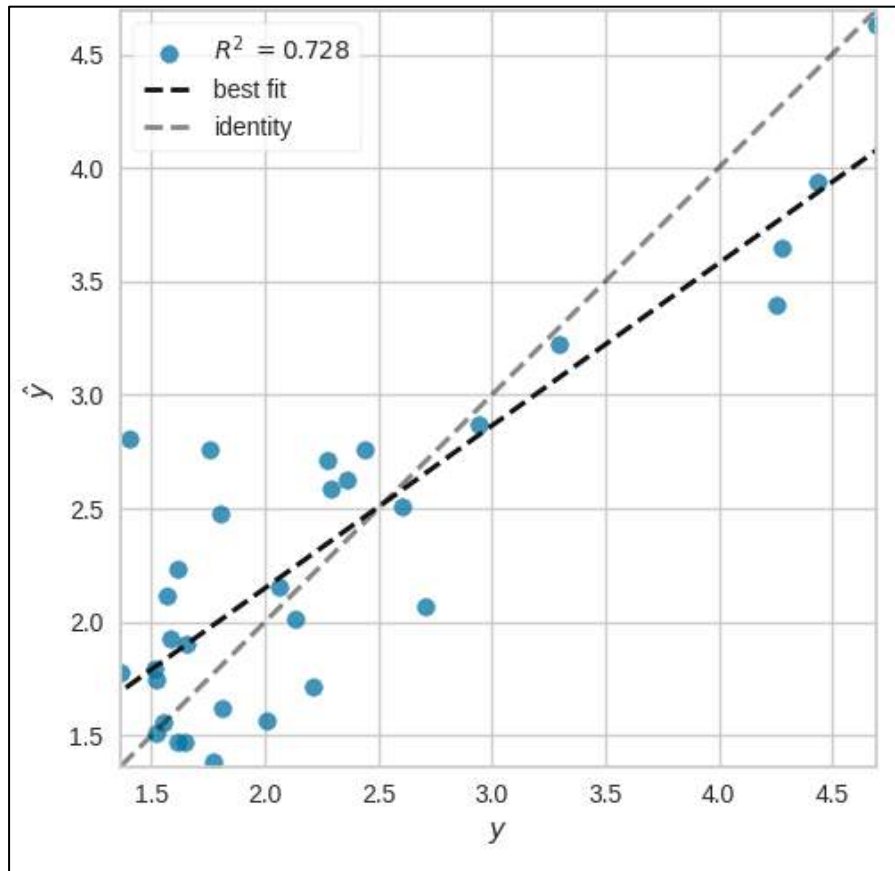
### a) SORGHUM AND CALCIUM NUTRIENT COMPOSITION OF GRAINS

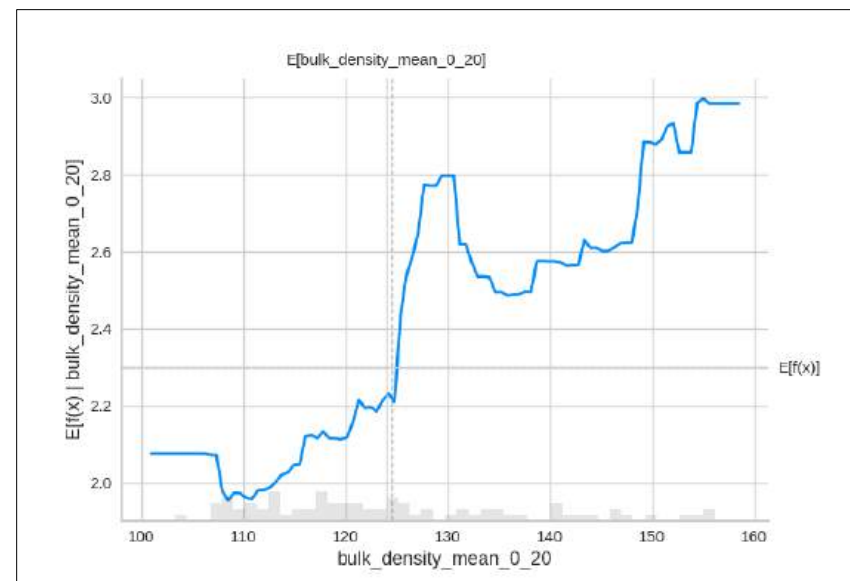
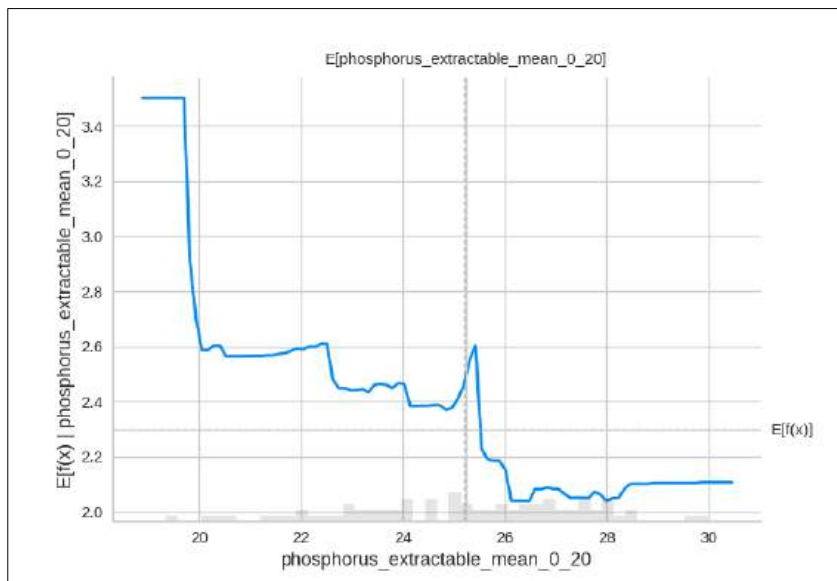
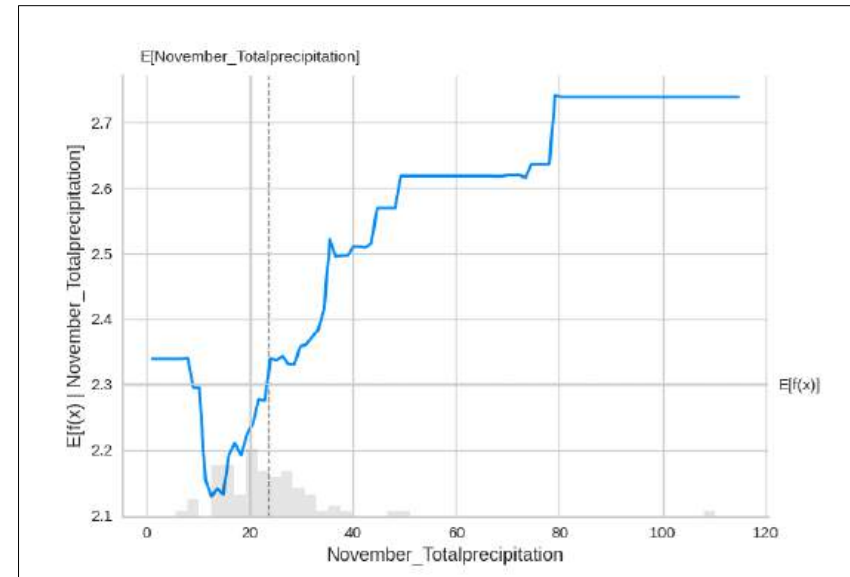
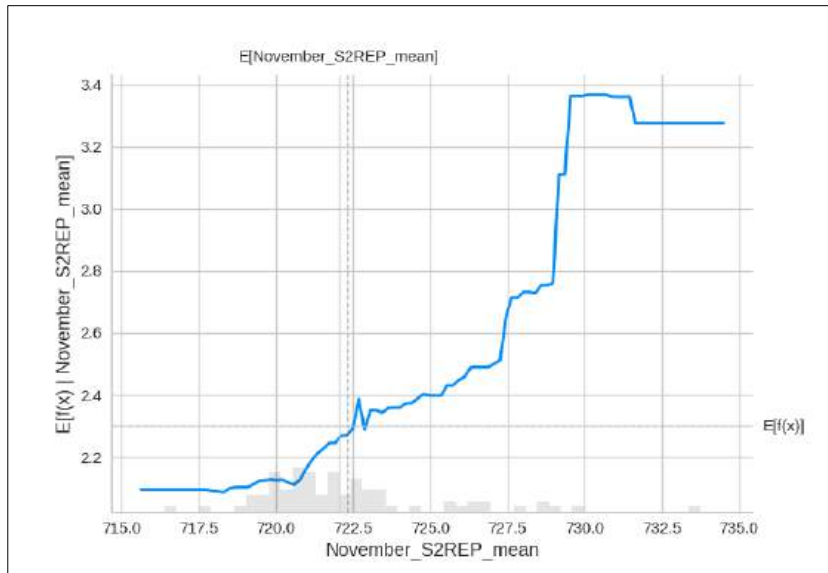




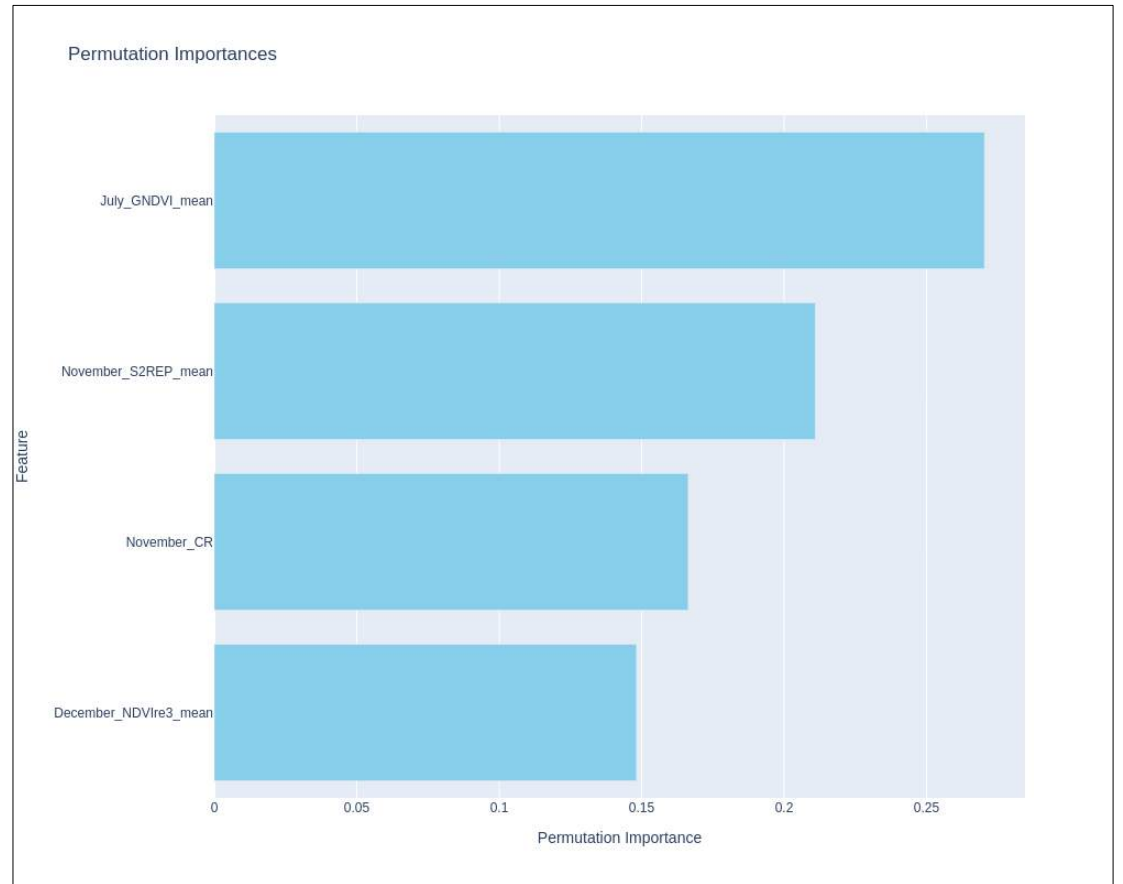
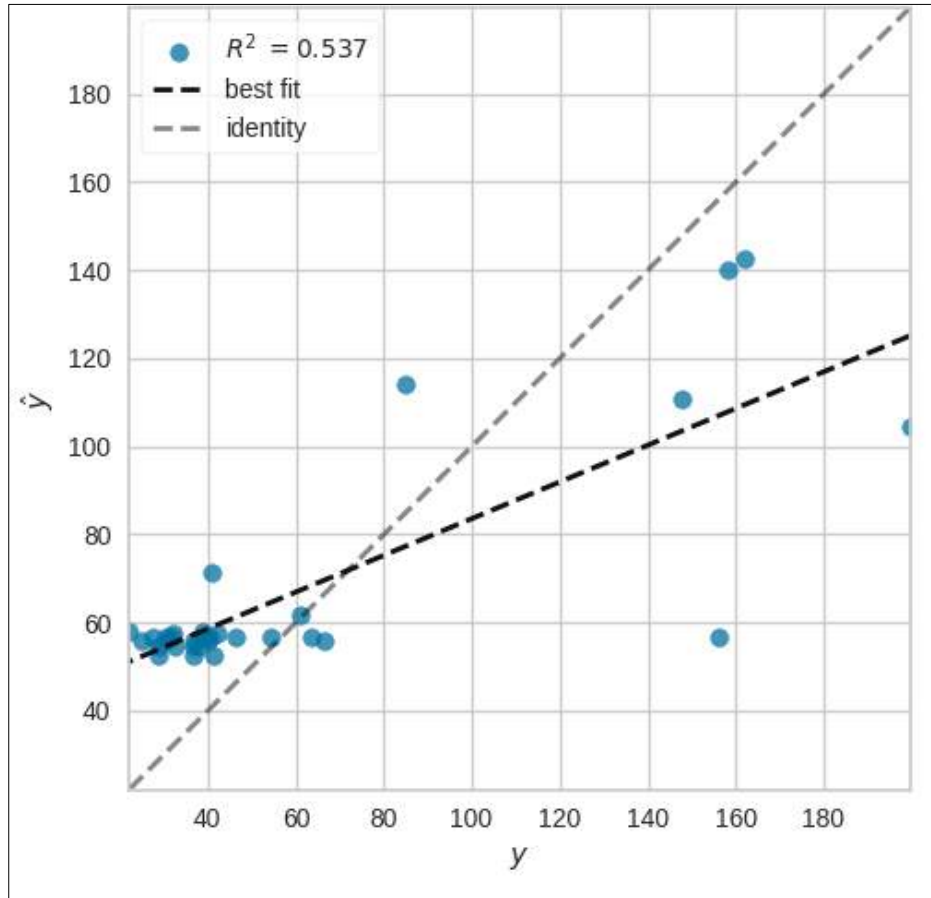


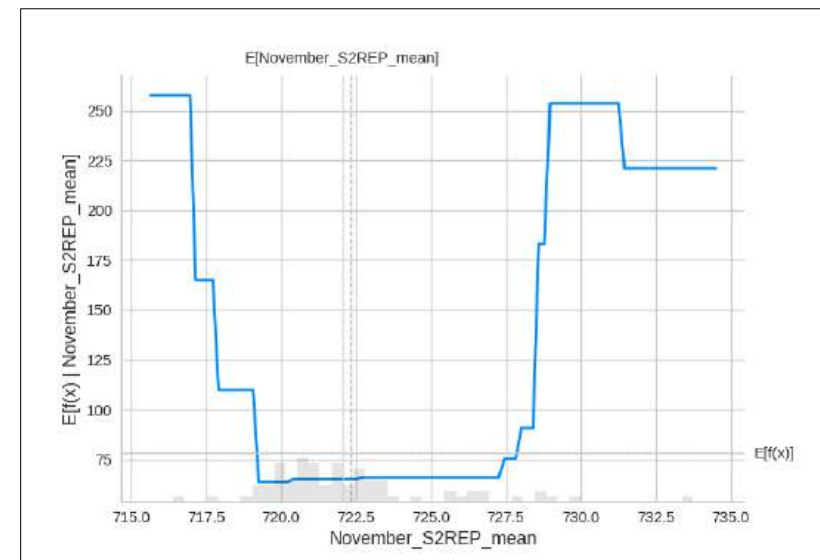
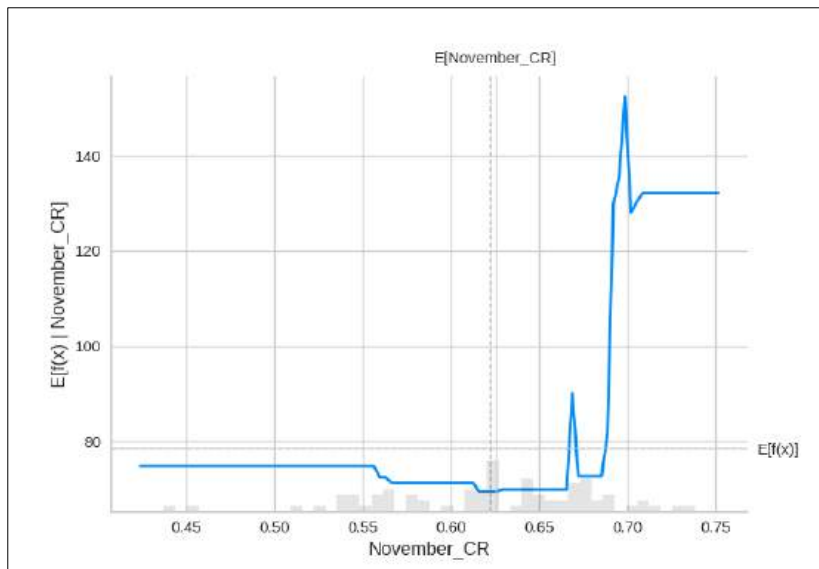
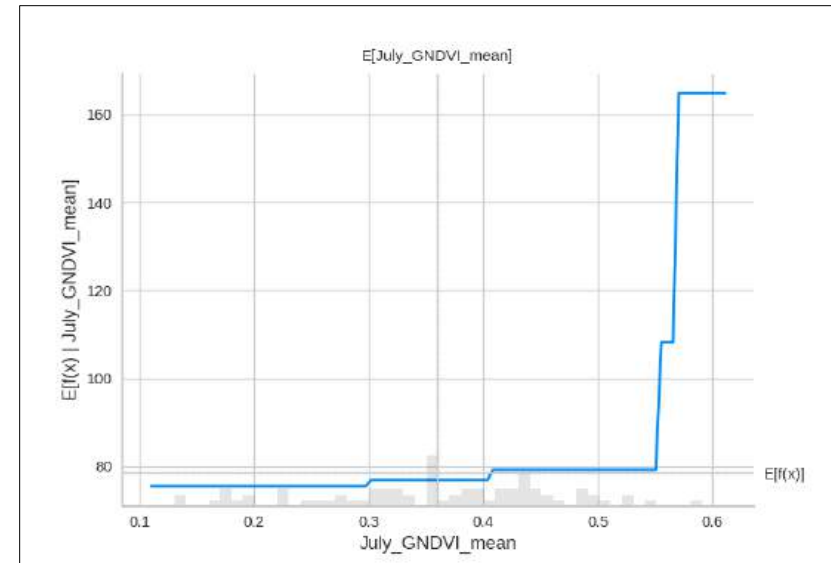
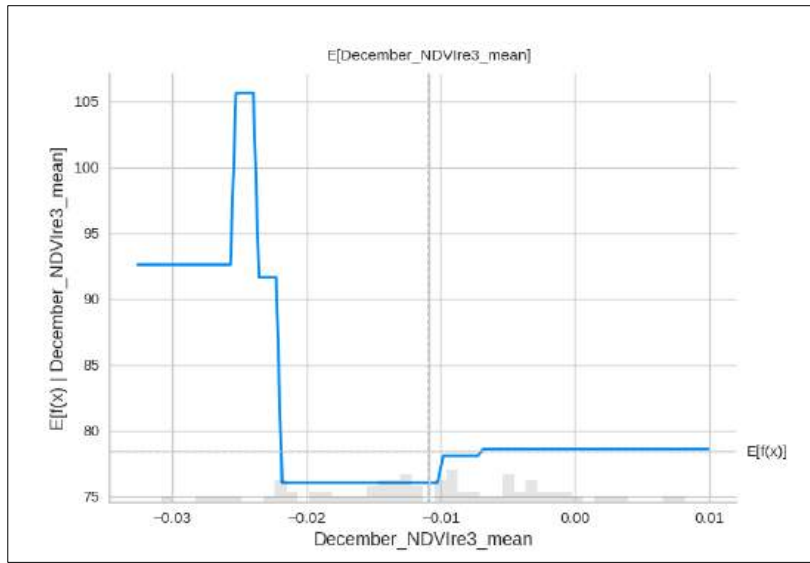
b) SORGHUM AND COPPER NUTRIENT COMPOSITION OF GRAINS





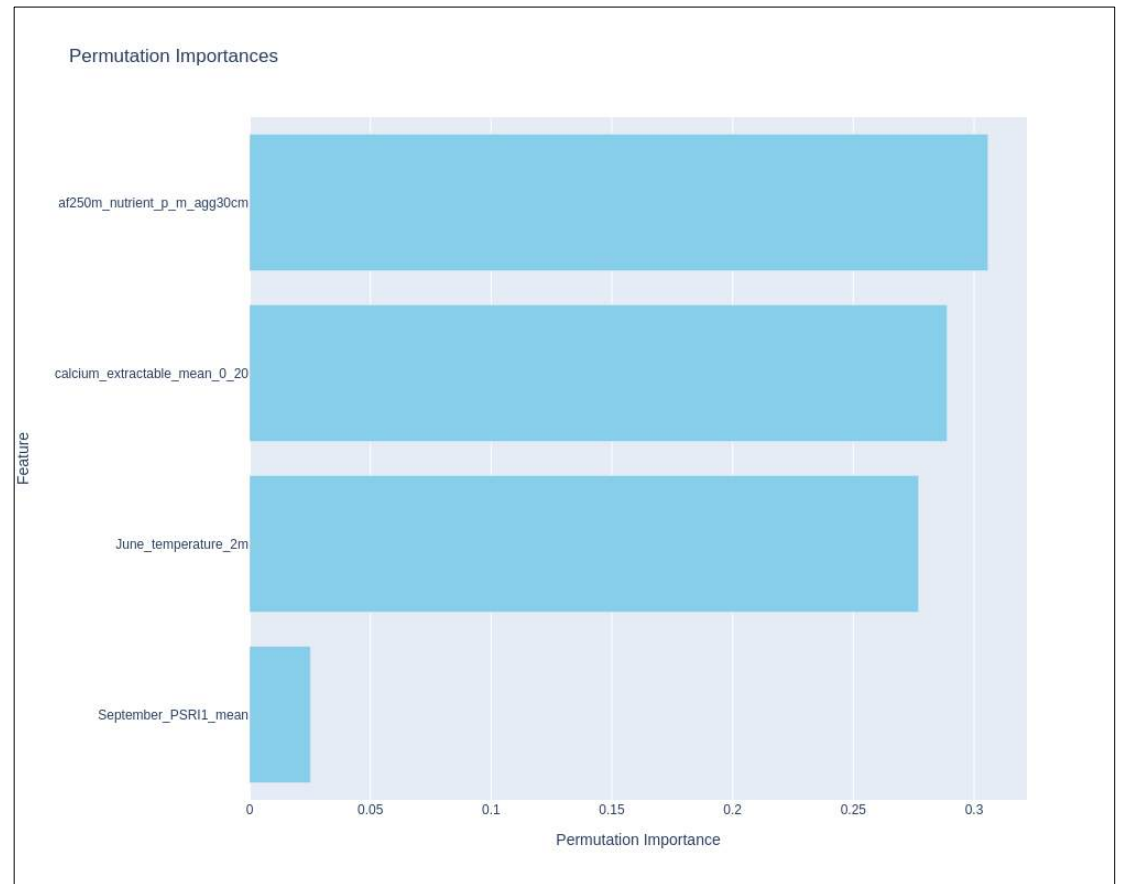
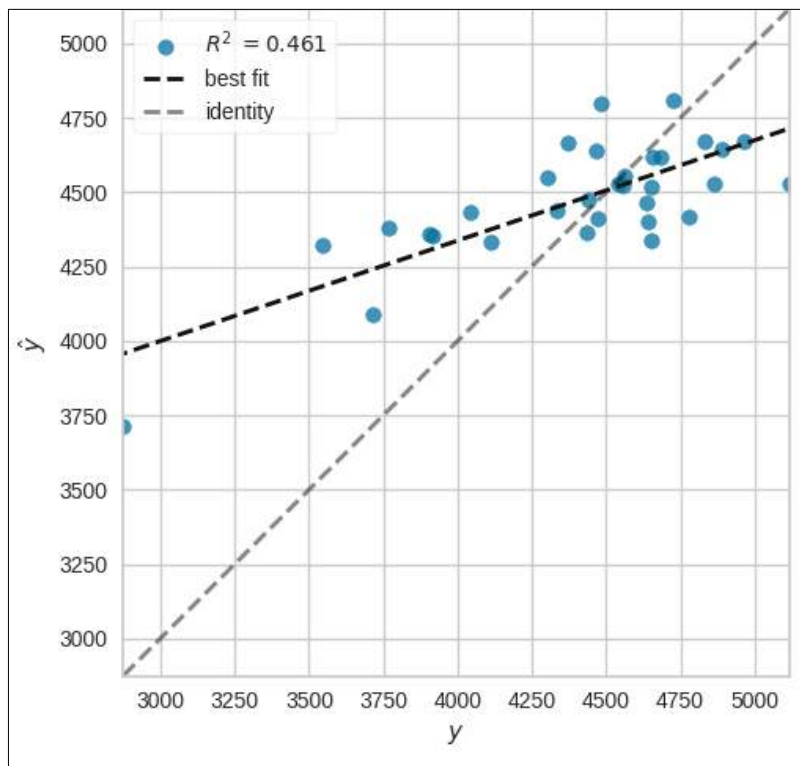
c) SORGHUM AND IRON NUTRIENT COMPOSITION OF GRAINS

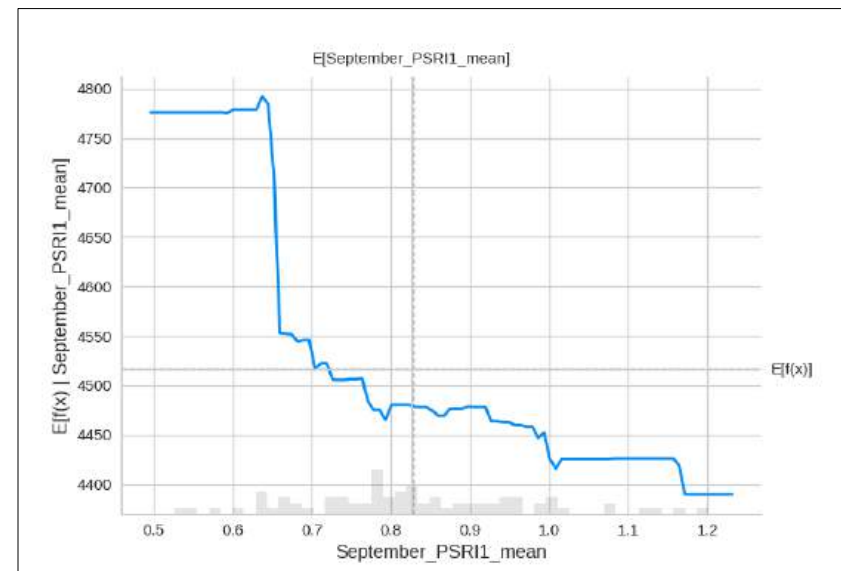
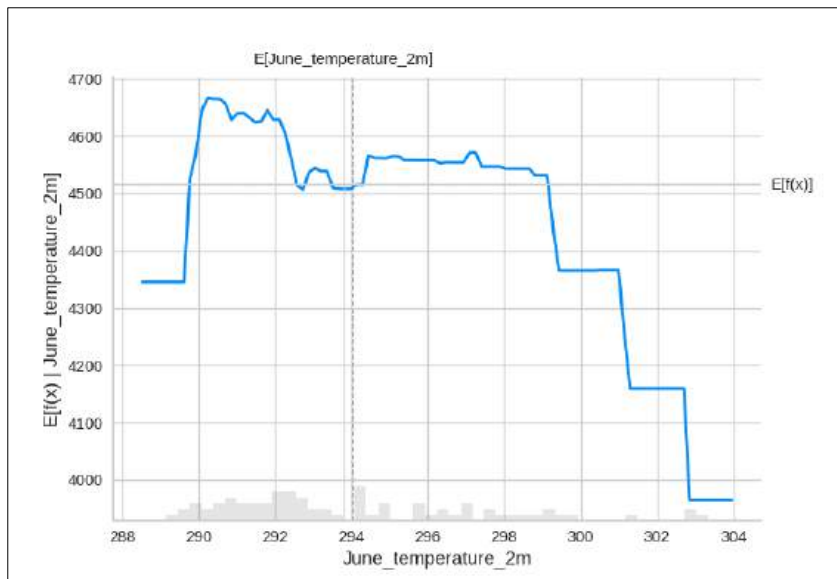
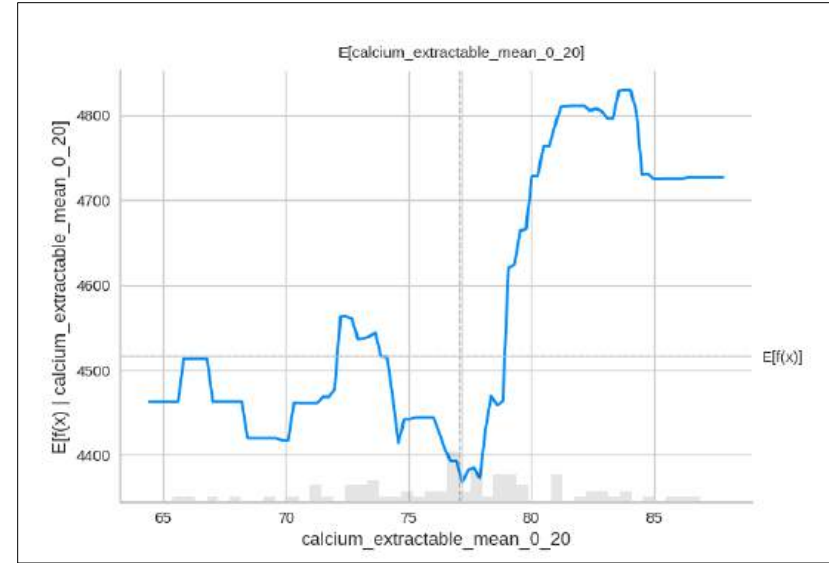
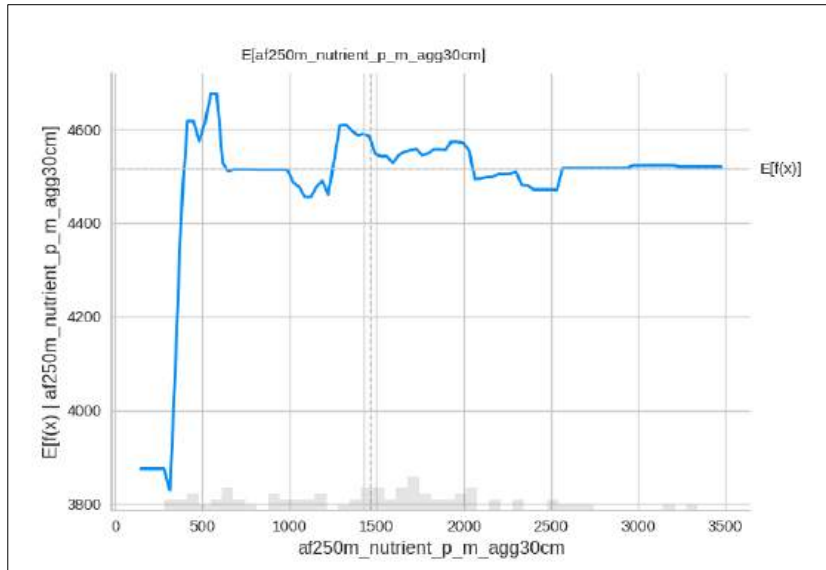




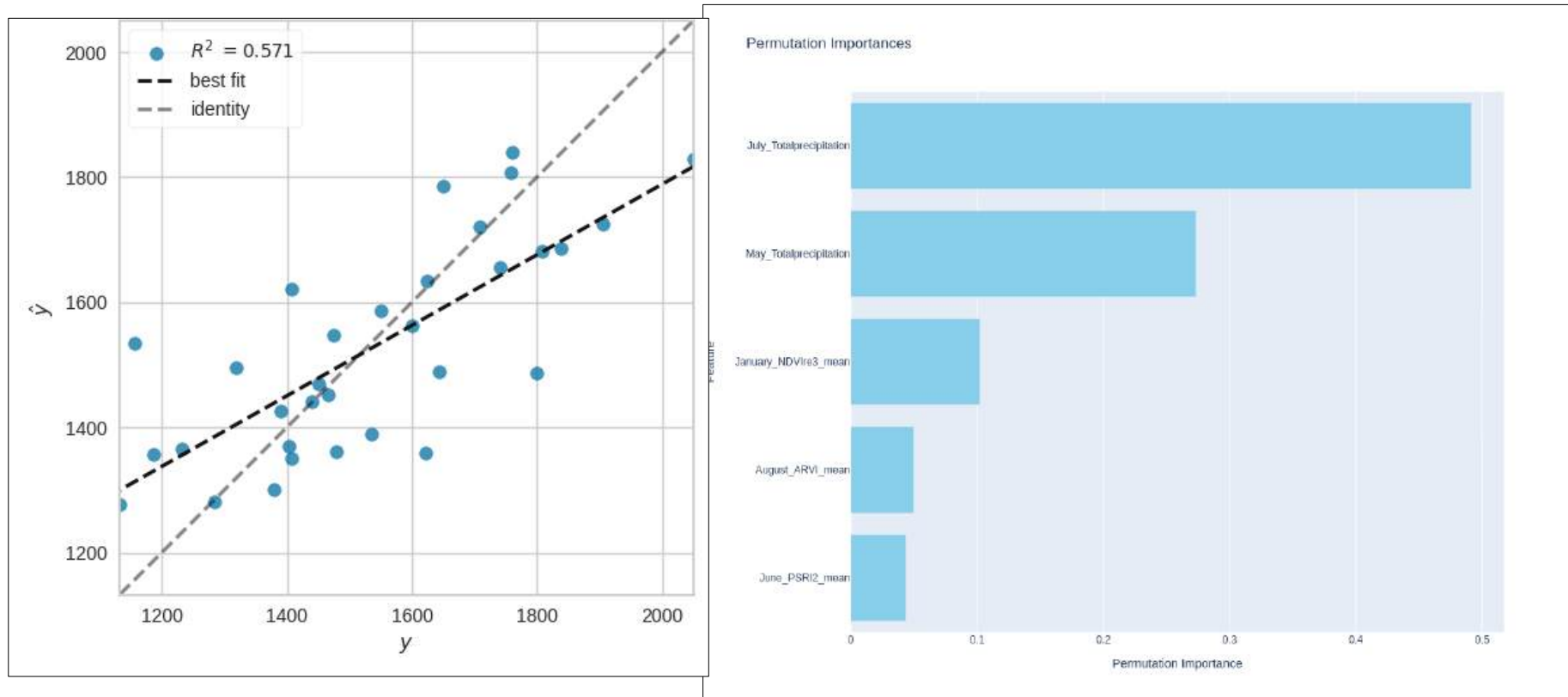


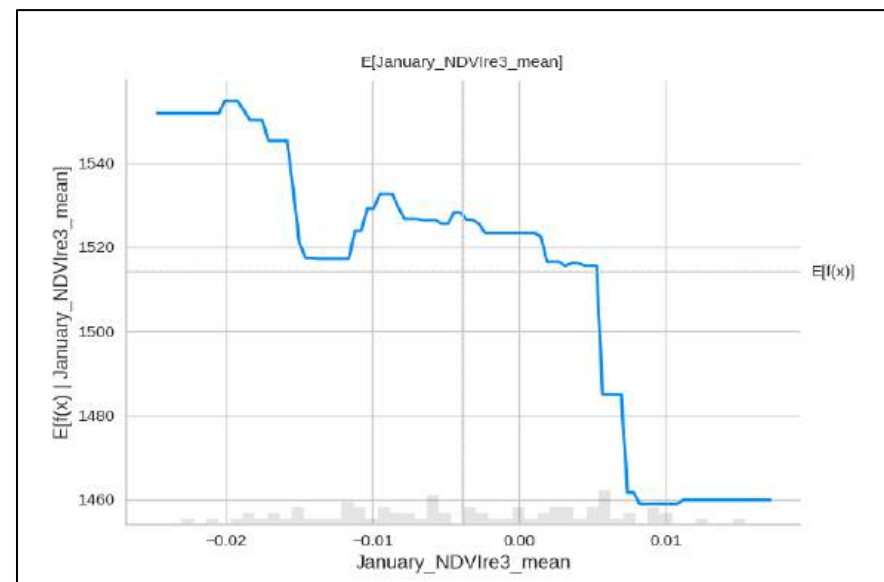
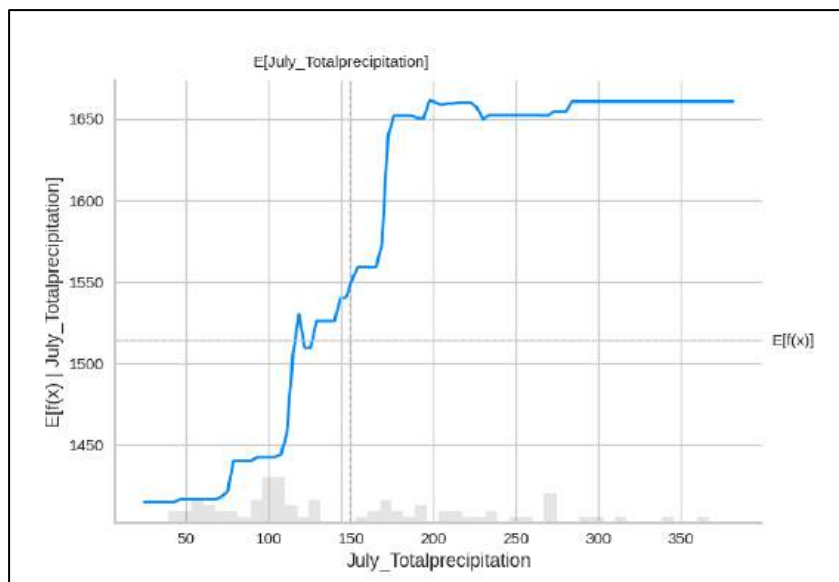
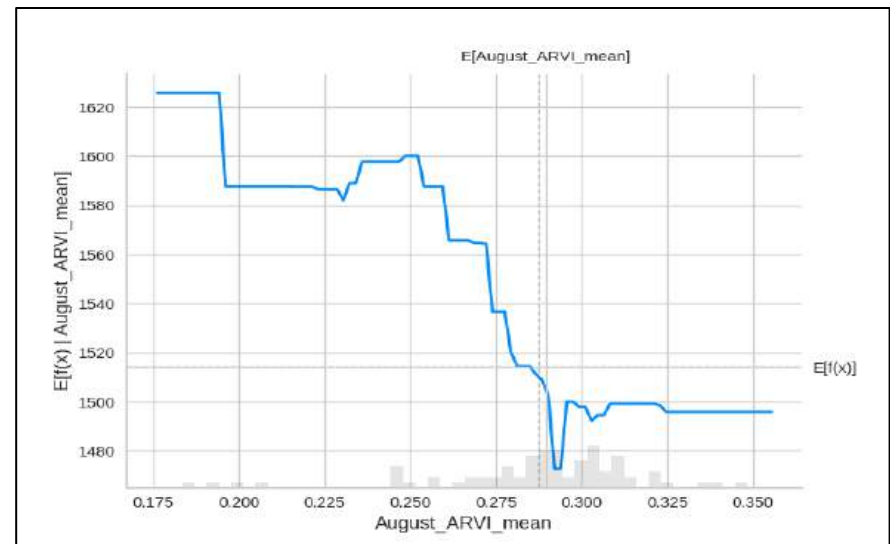
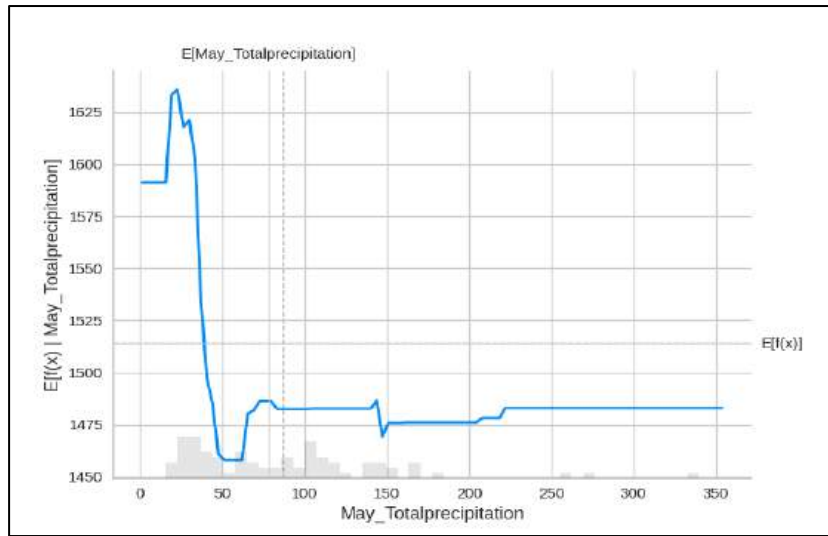
d) SORGHUM AND POTASSIUM NUTRIENT COMPOSITION OF GRAINS



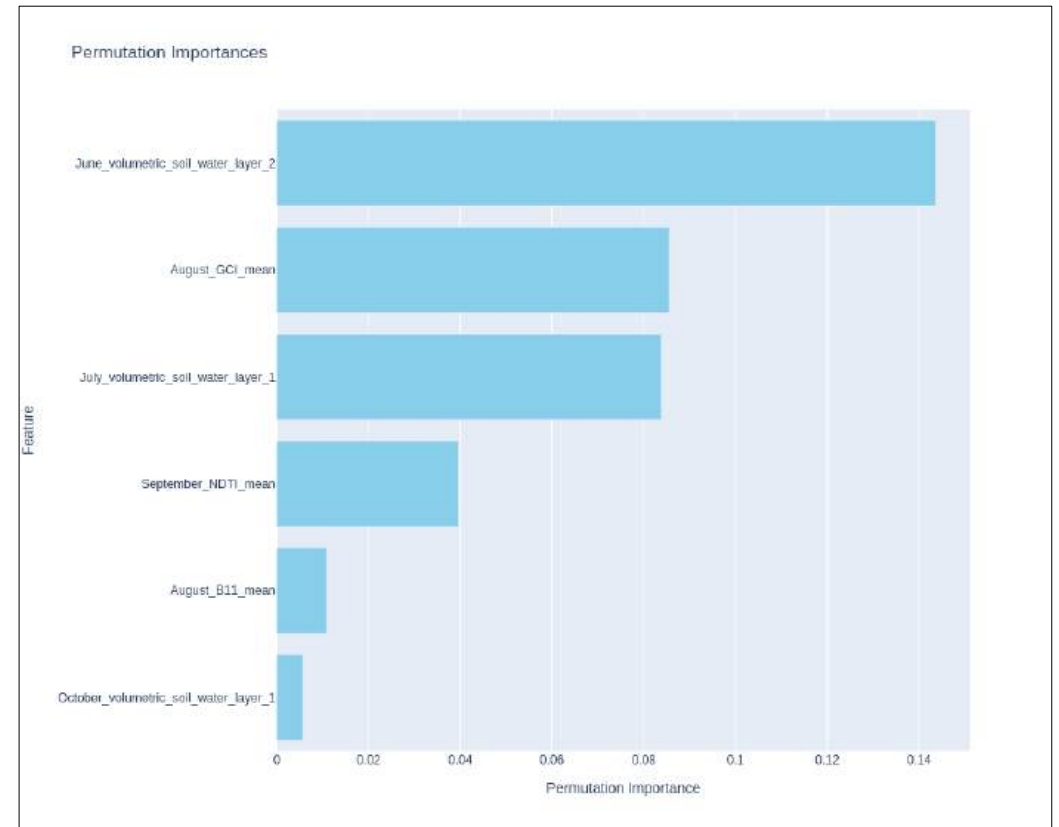
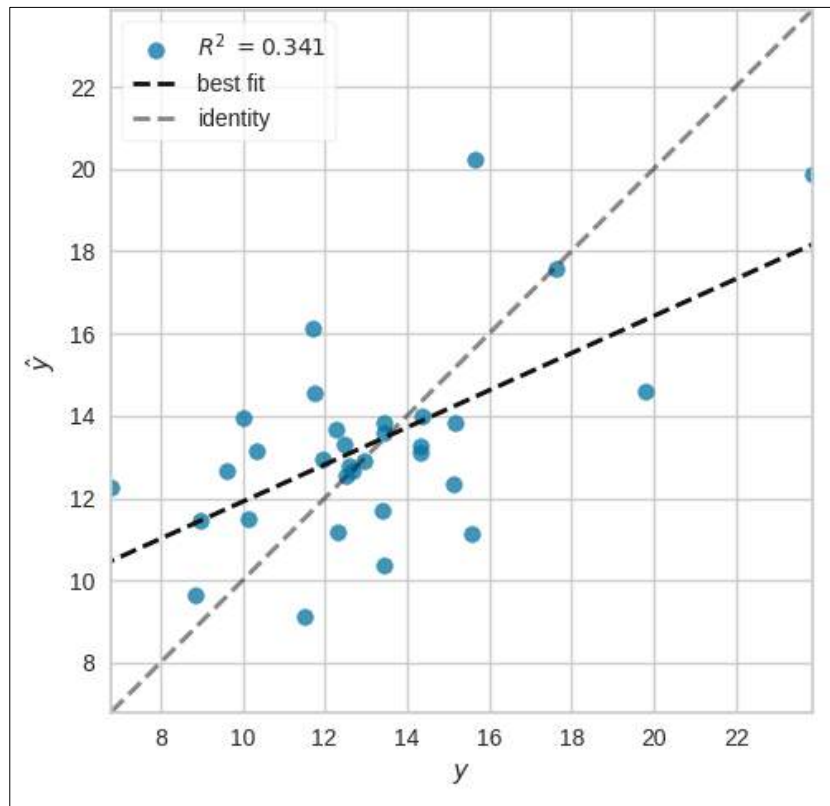


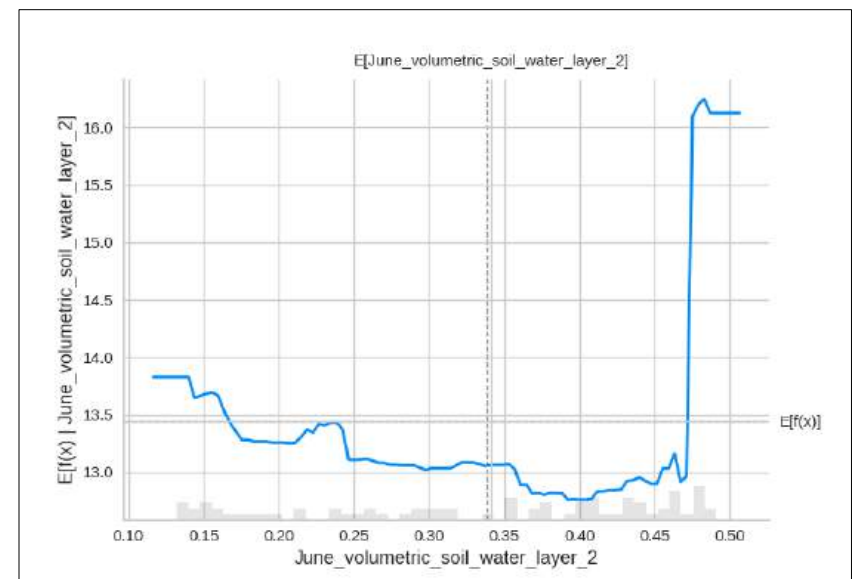
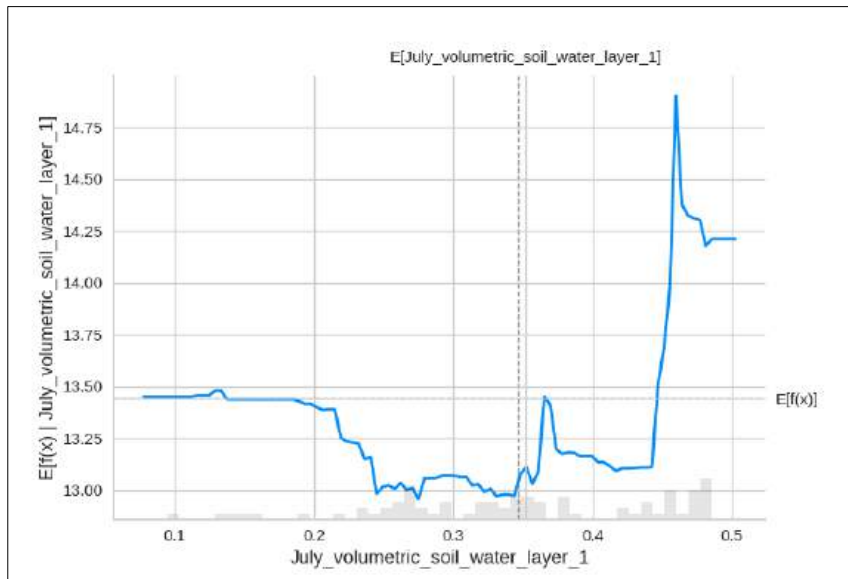
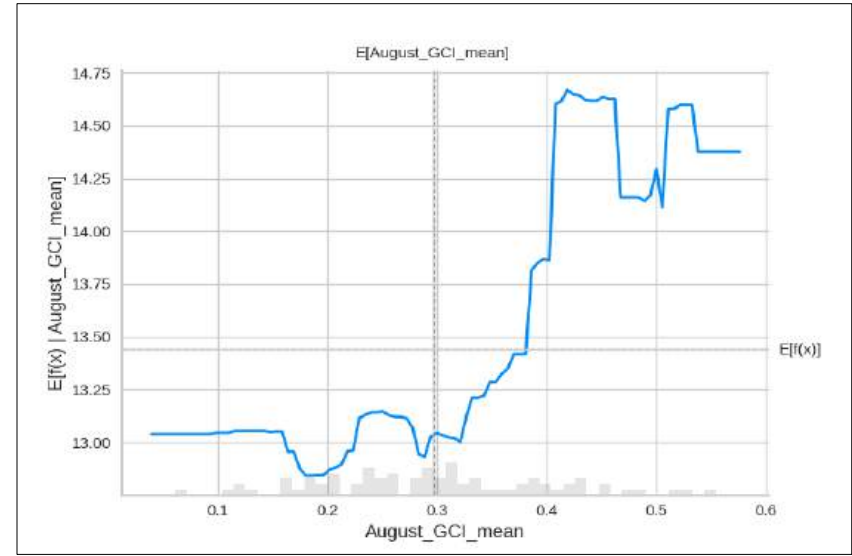
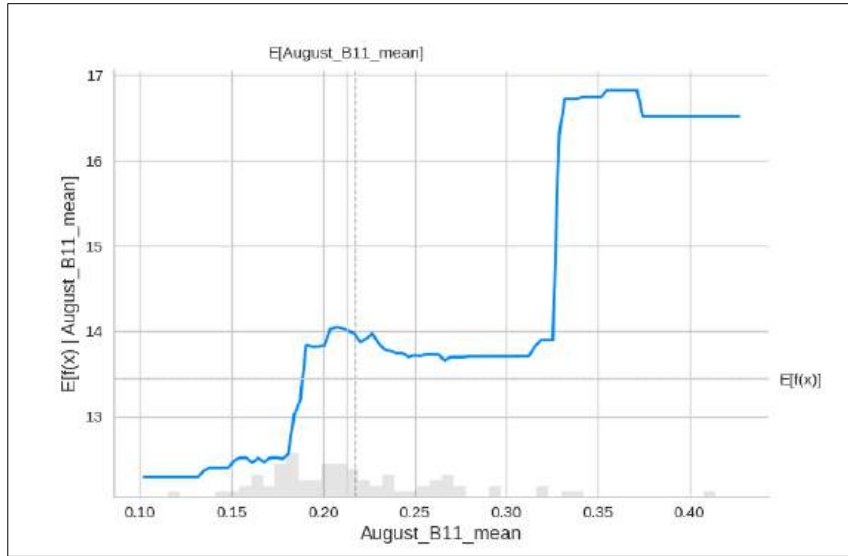
e) SORGHUM AND MAGNESIUM NUTRIENT COMPOSITION OF GRAINS

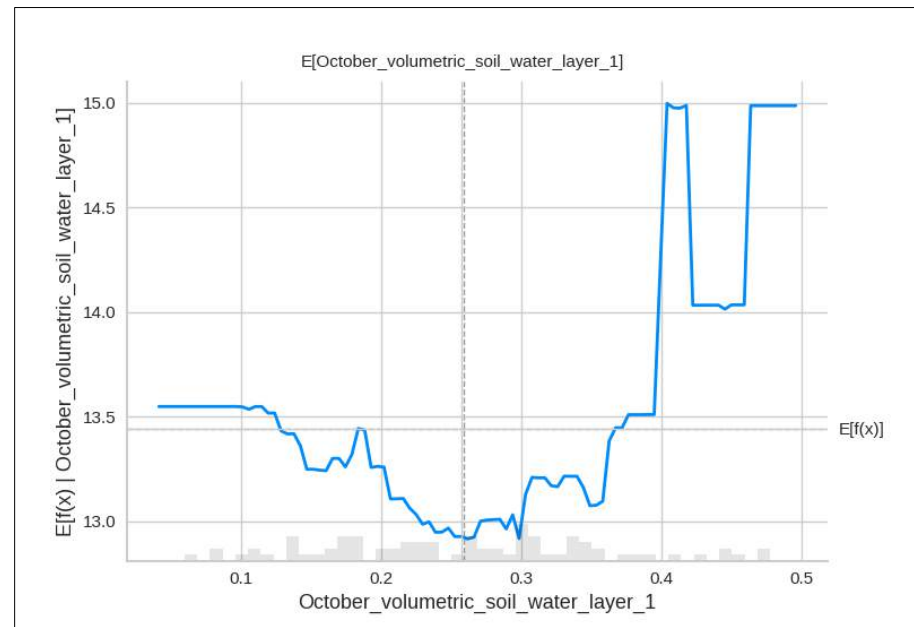
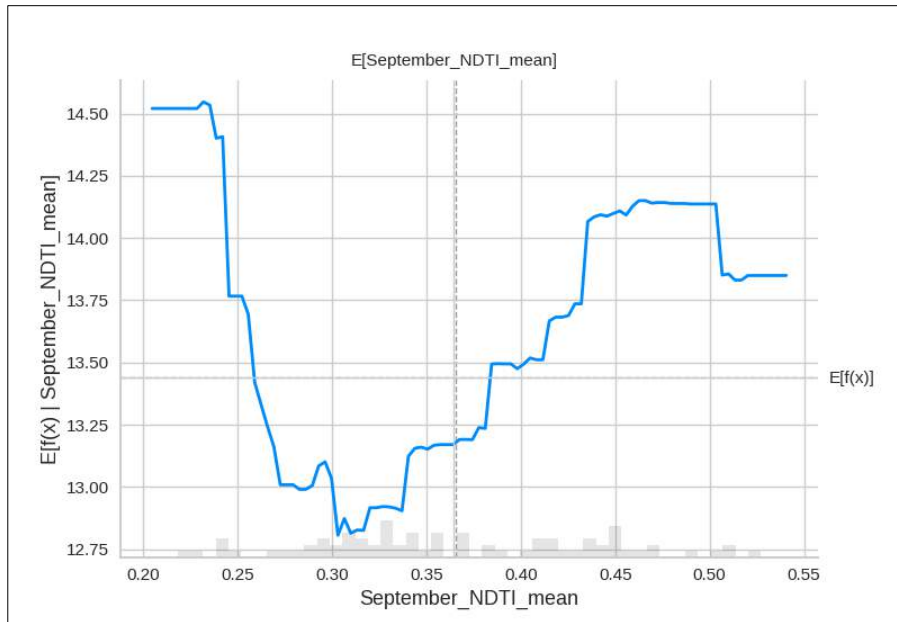




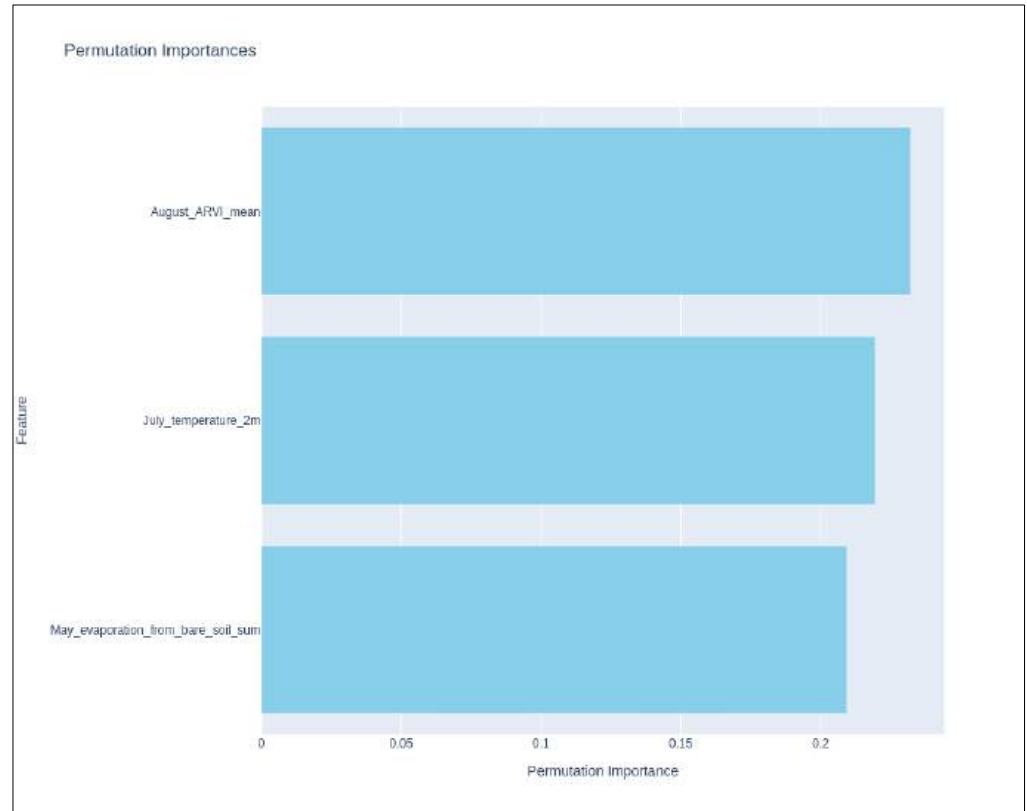
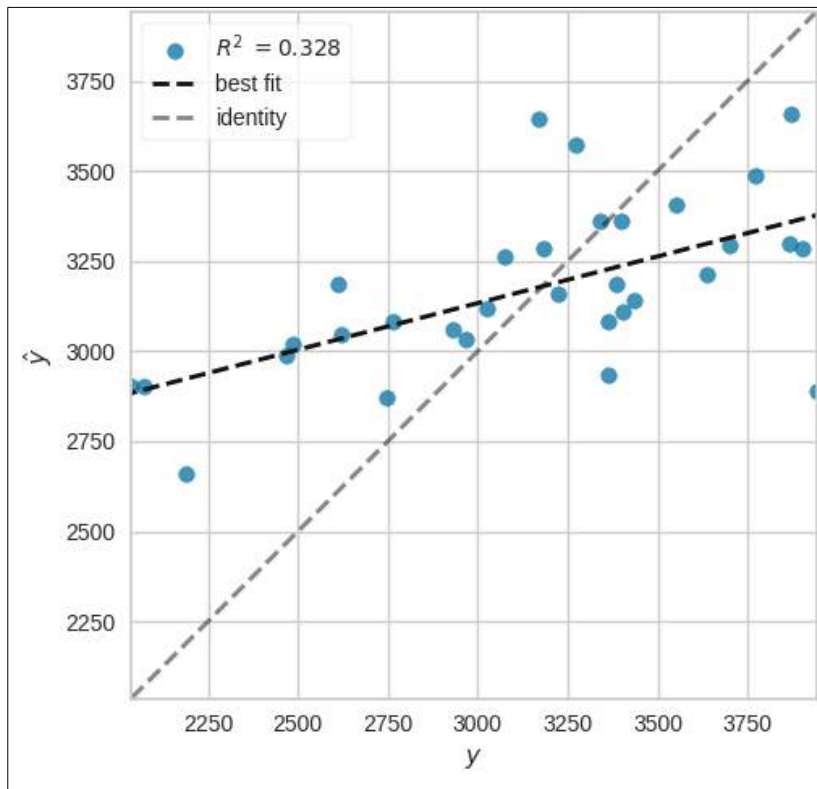
f) SORGHUM AND MANGANESE NUTRIENT COMPOSITION OF GRAINS



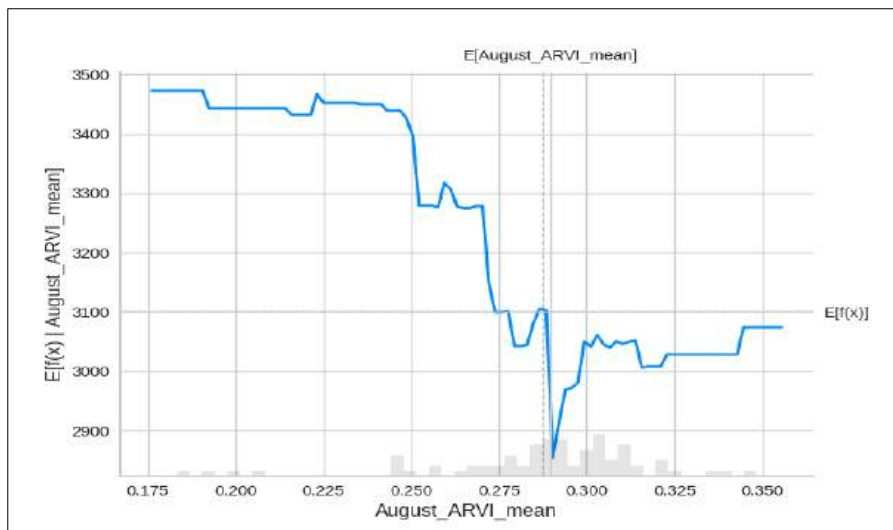
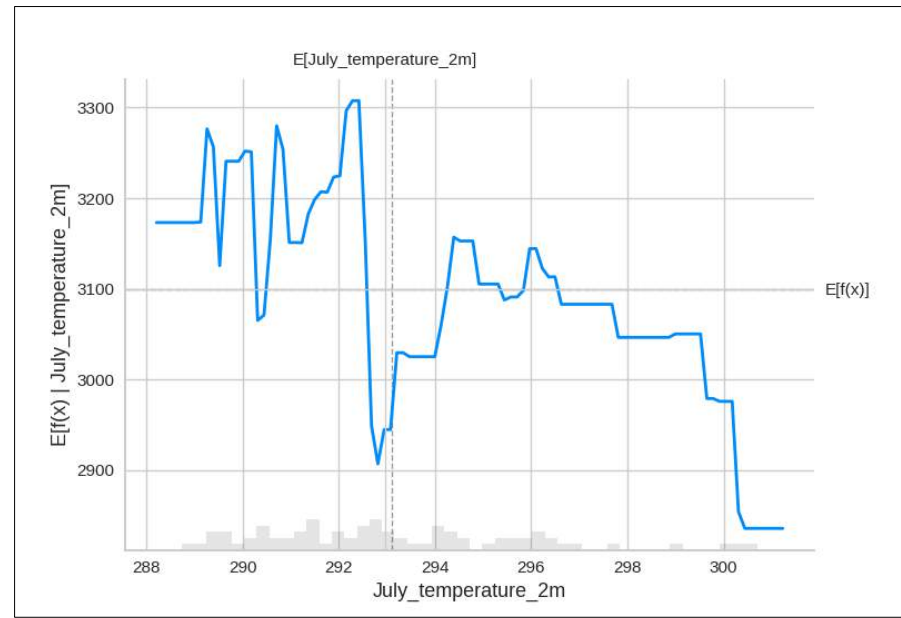
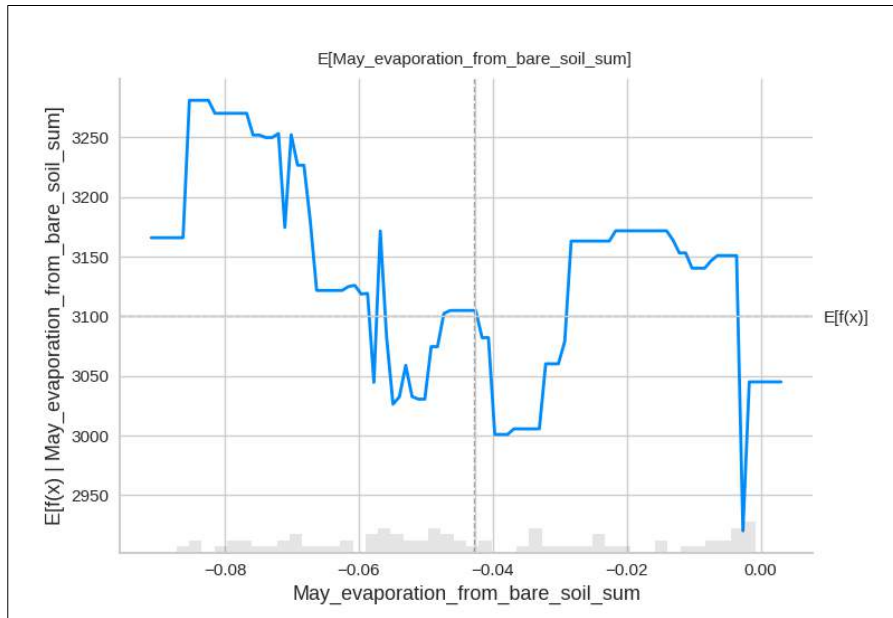




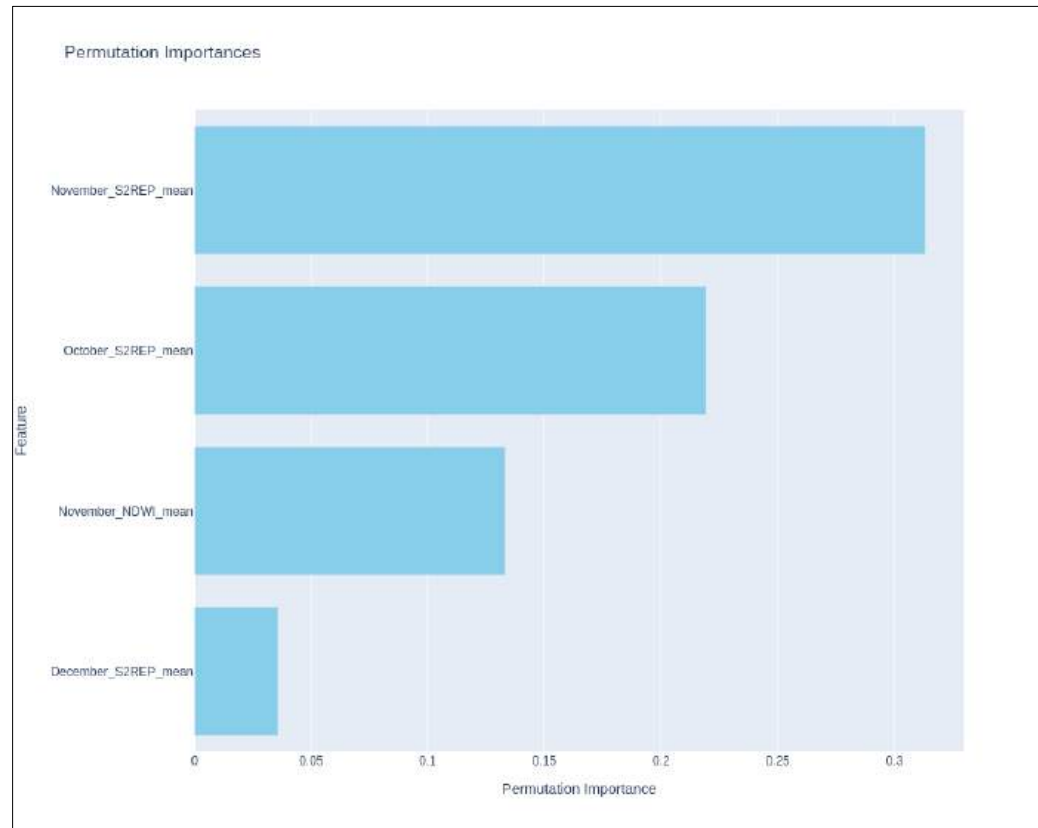
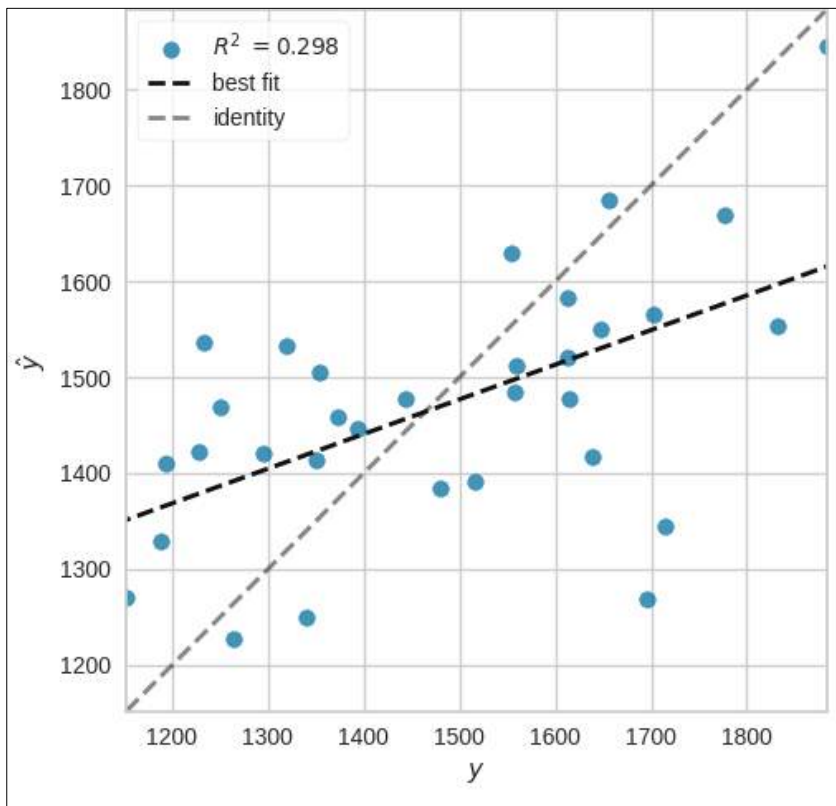
g) SORGHUM AND PHOSPHORUS NUTRIENT COMPOSITION OF GRAINS

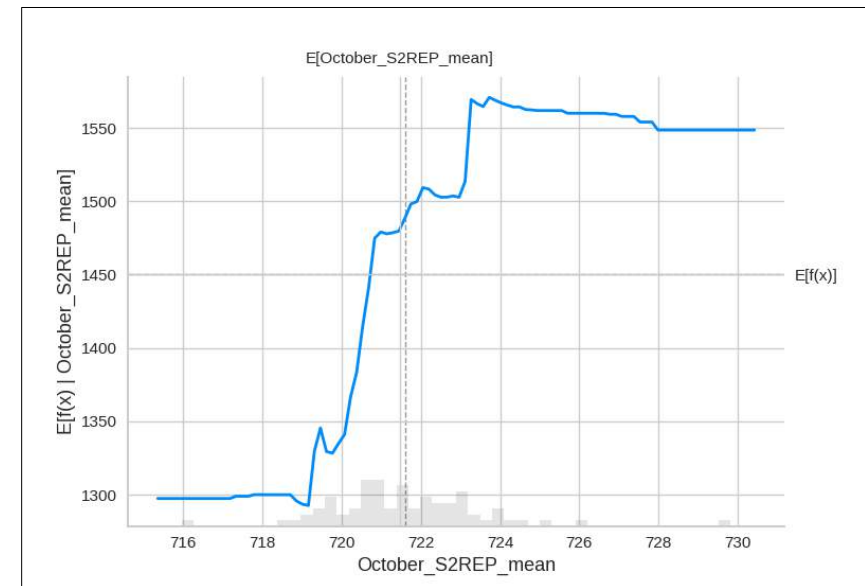
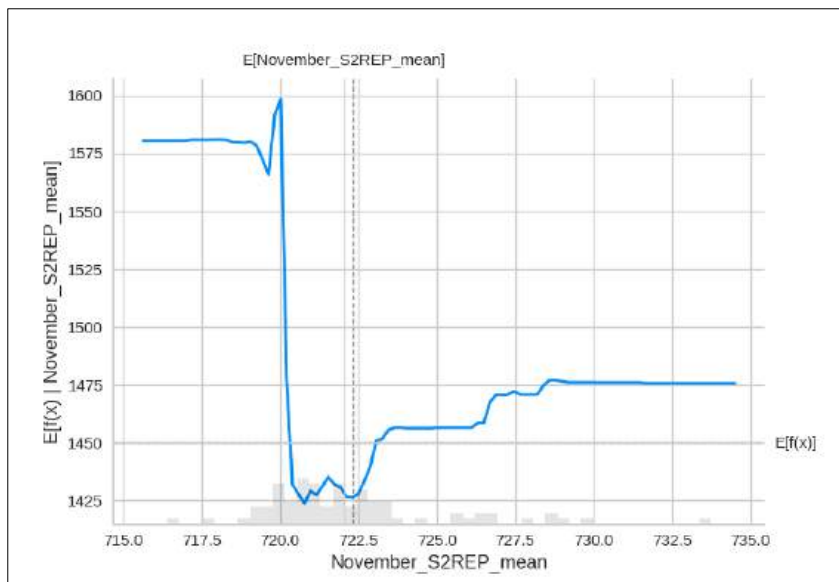
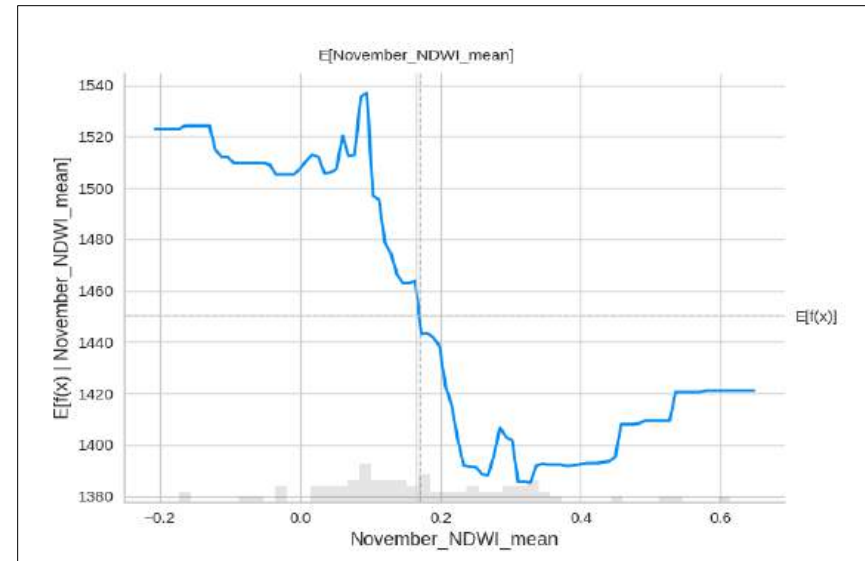
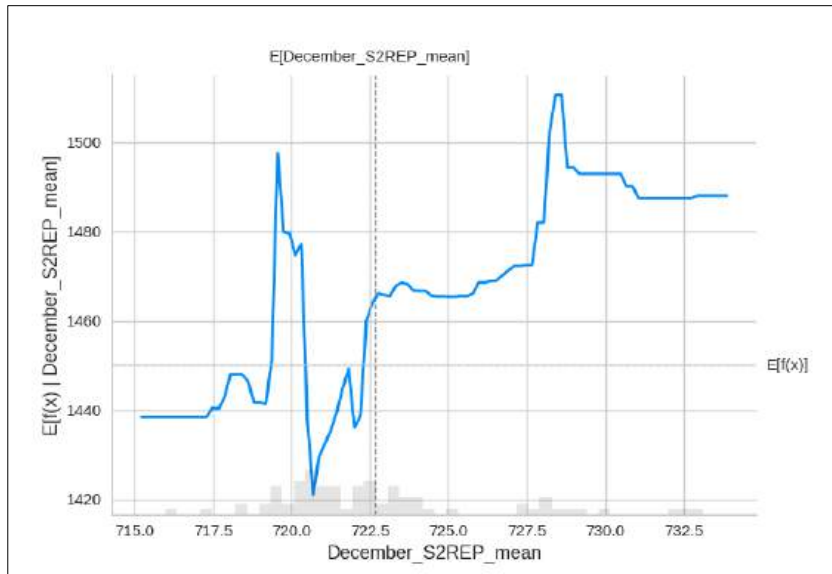




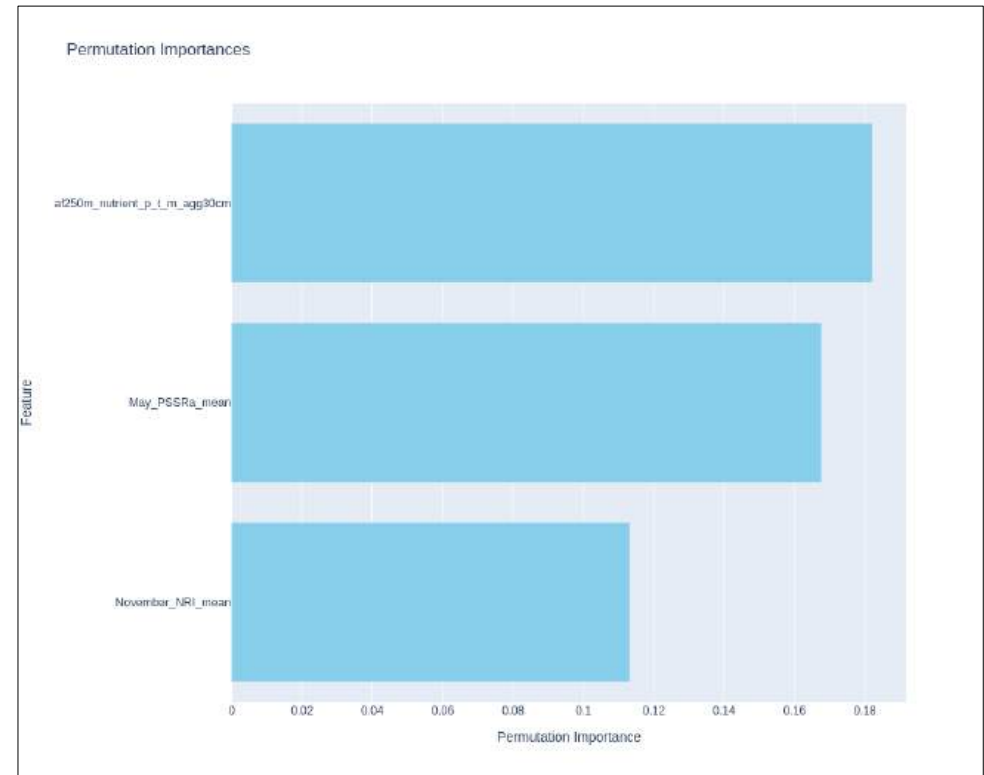
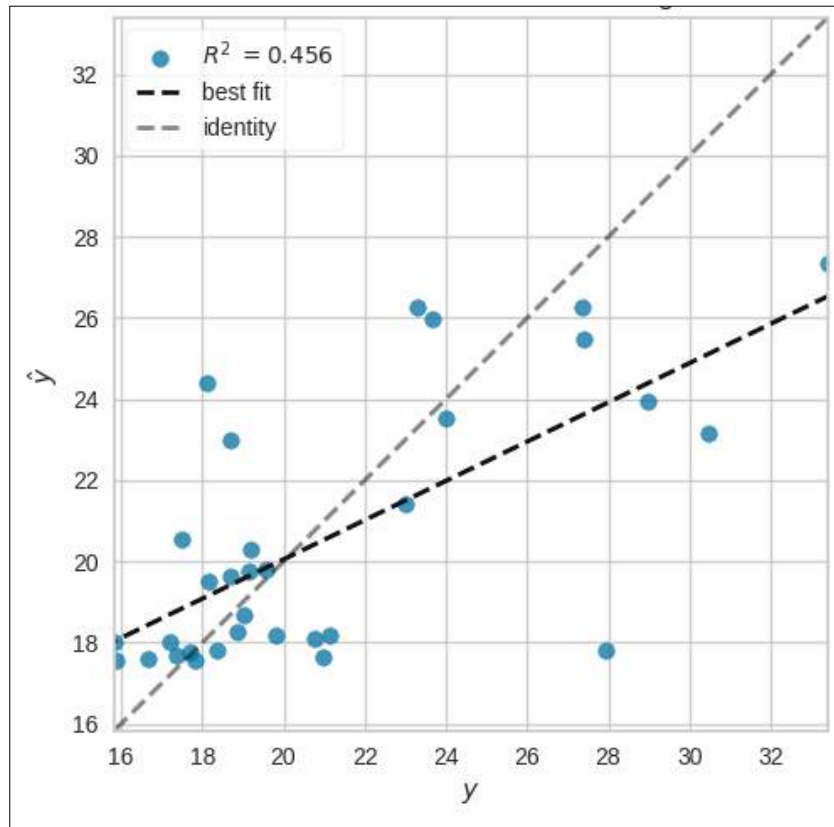


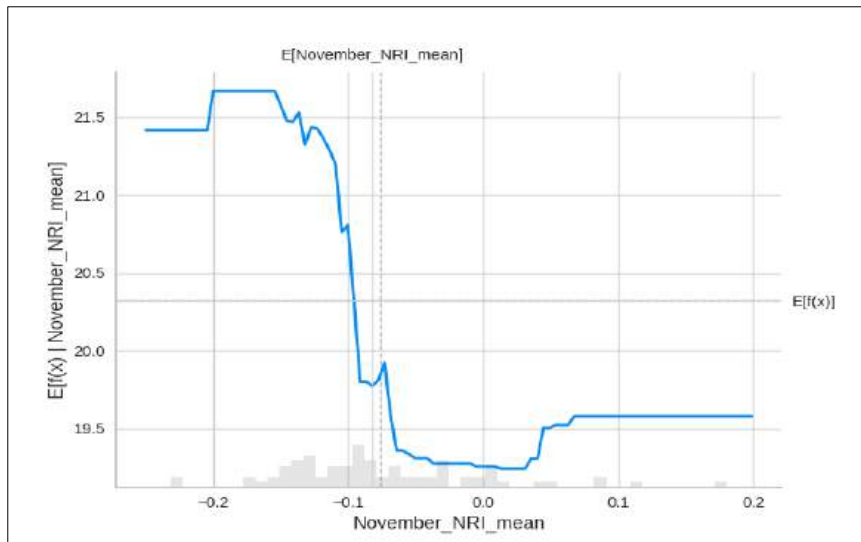
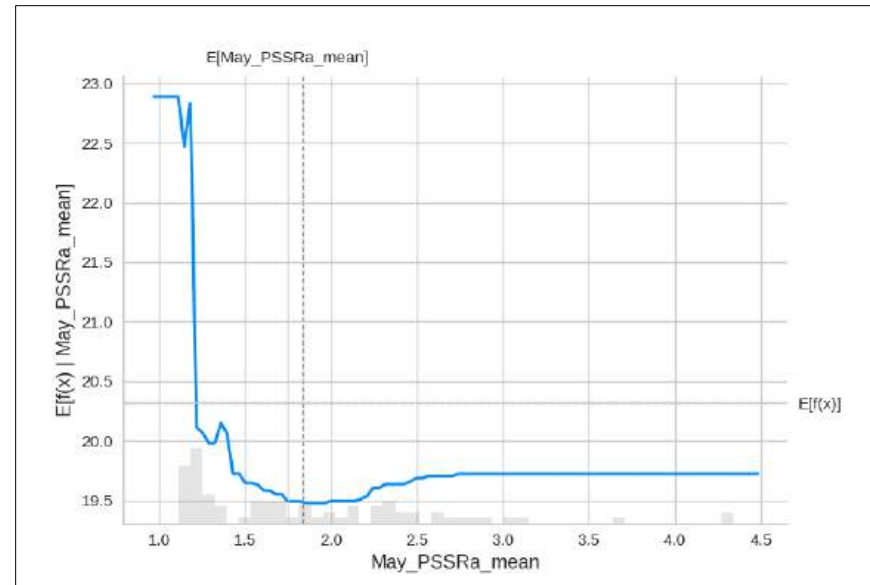
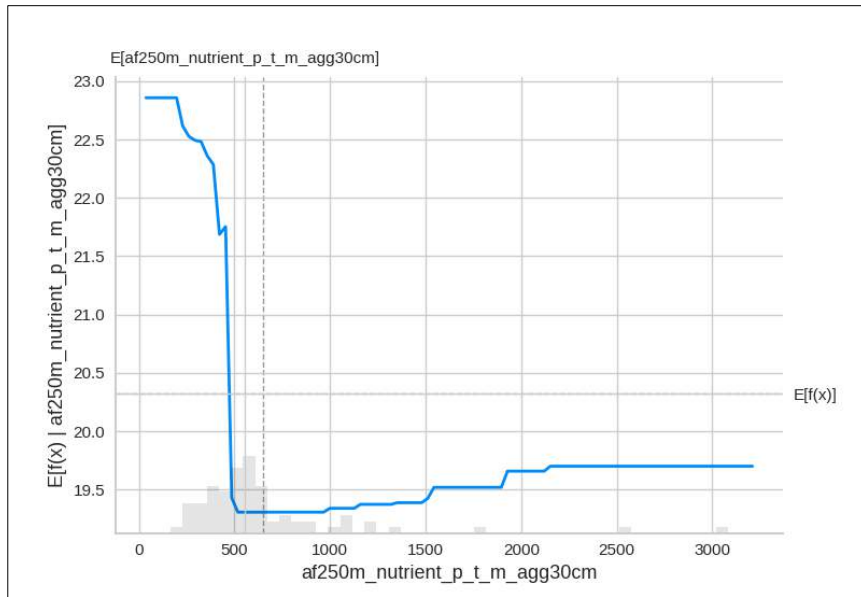
## h) SORGHUM AND SULPHUR NUTRIENT COMPOSITION OF GRAINS





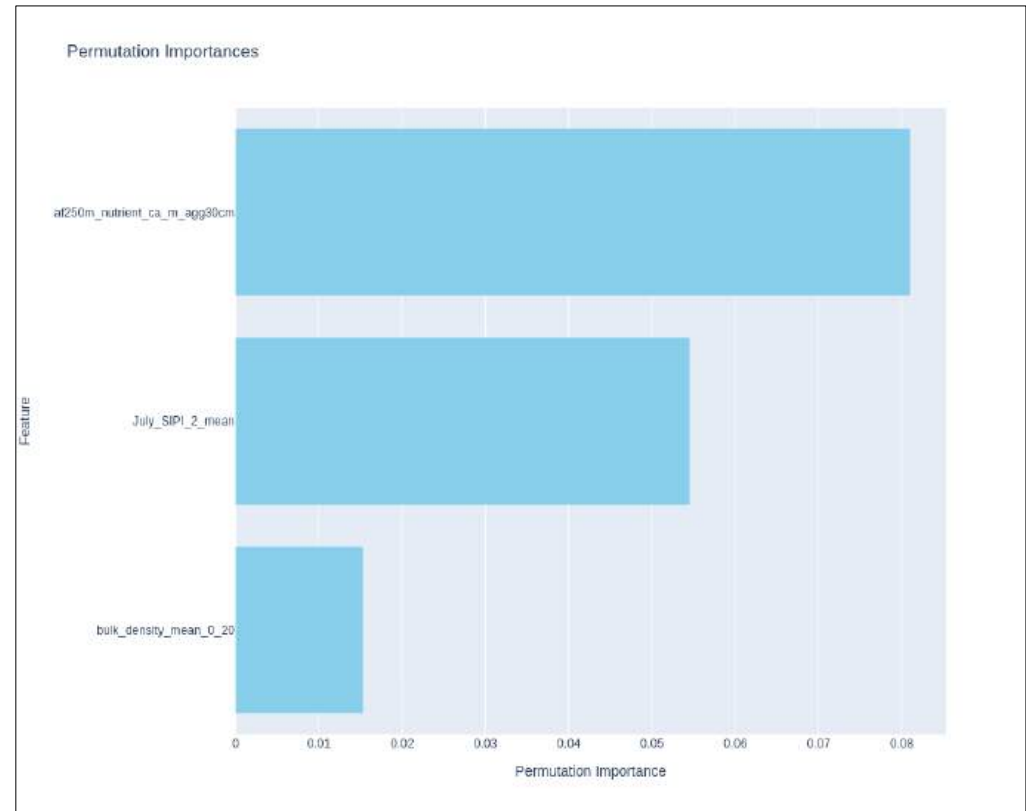
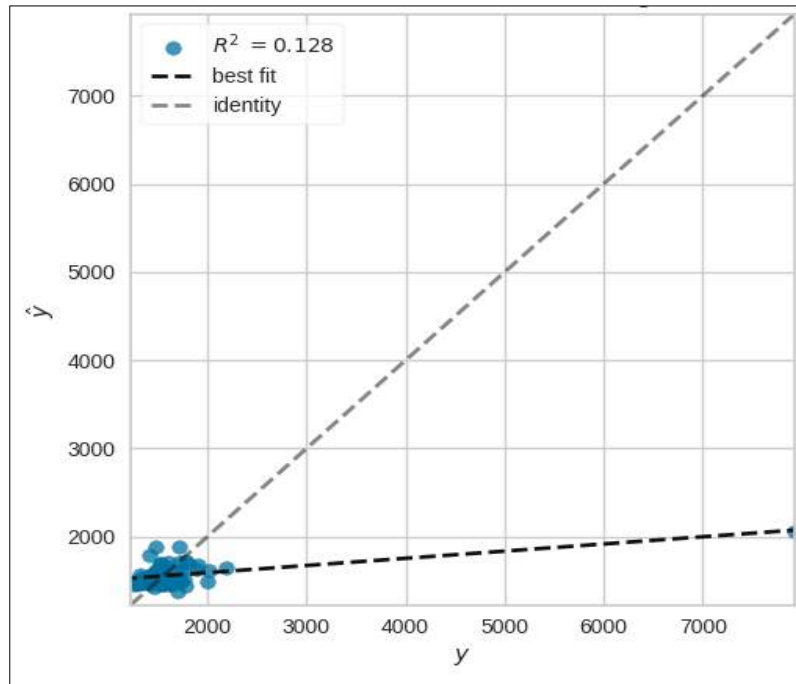
i) SORGHUM AND ZINC NUTRIENT COMPOSITION OF GRAINS

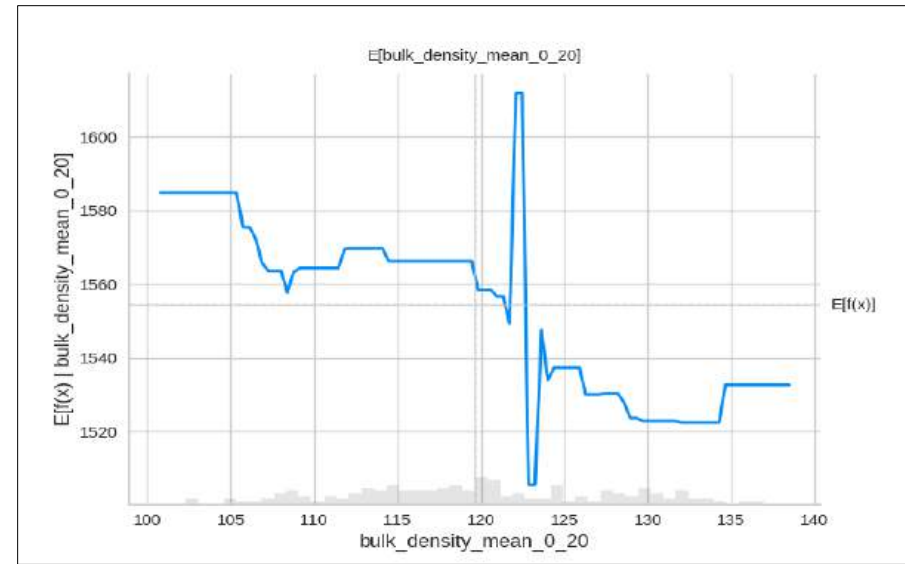
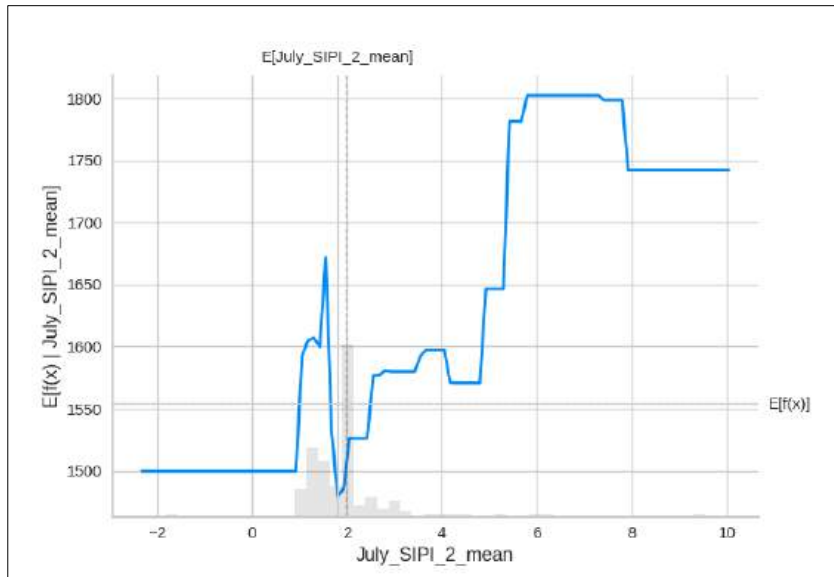


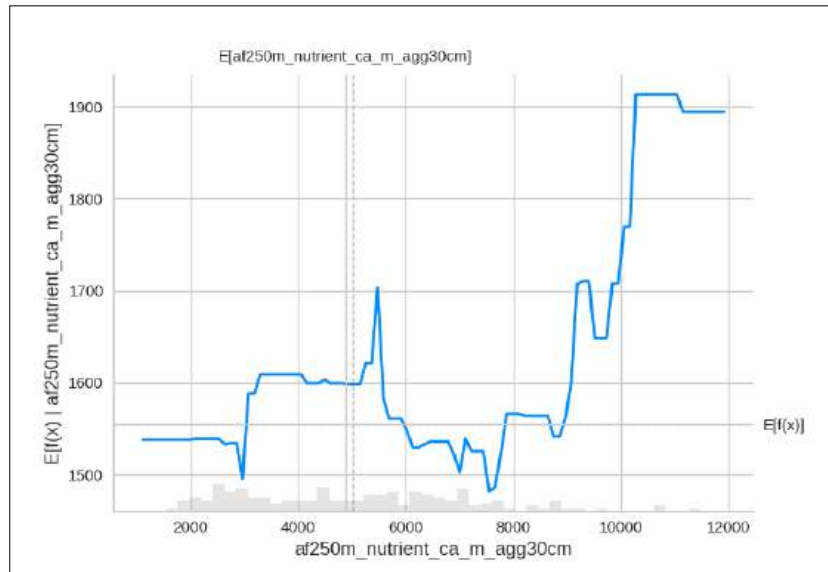


## Appendix 5. Results for teff

### a) TEFF AND CALCIUM NUTRIENT COMPOSITION OF GRAINS

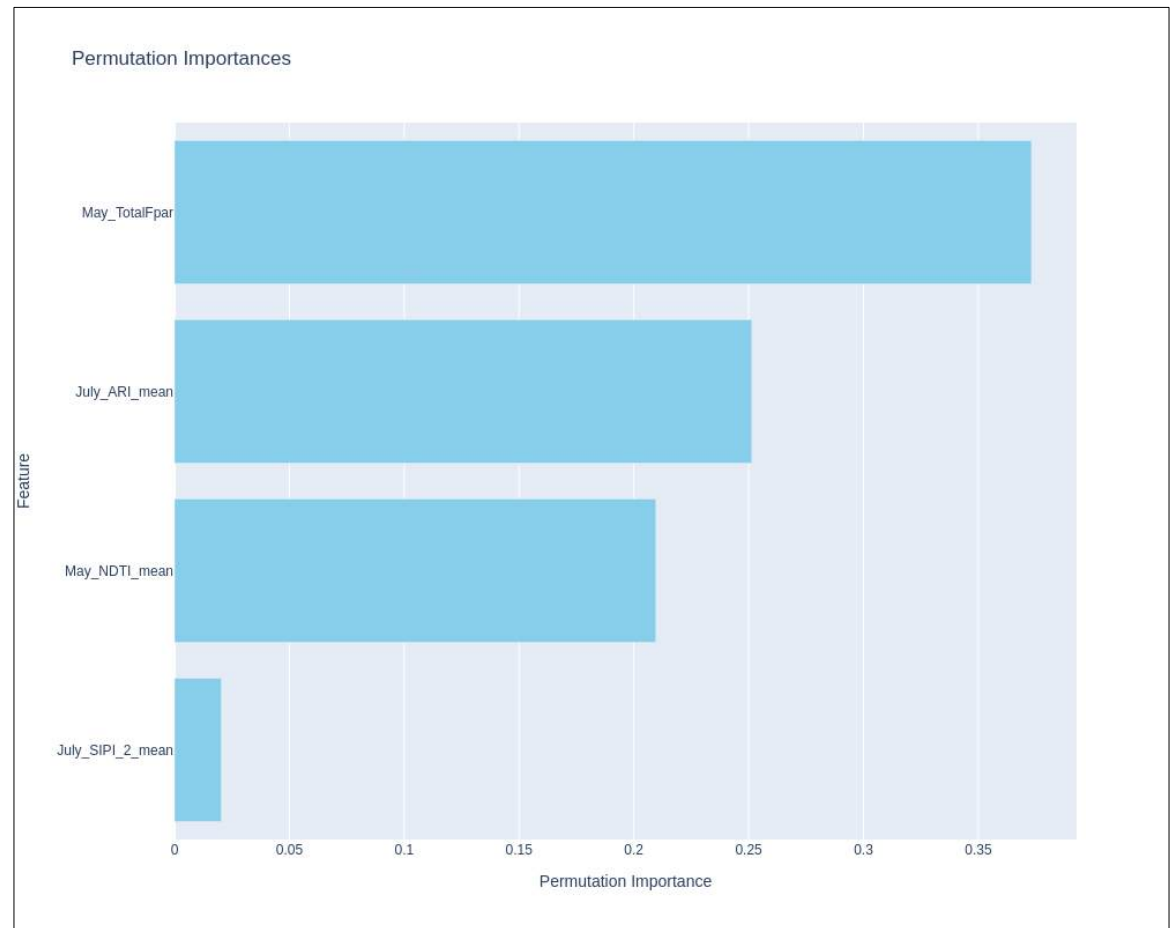
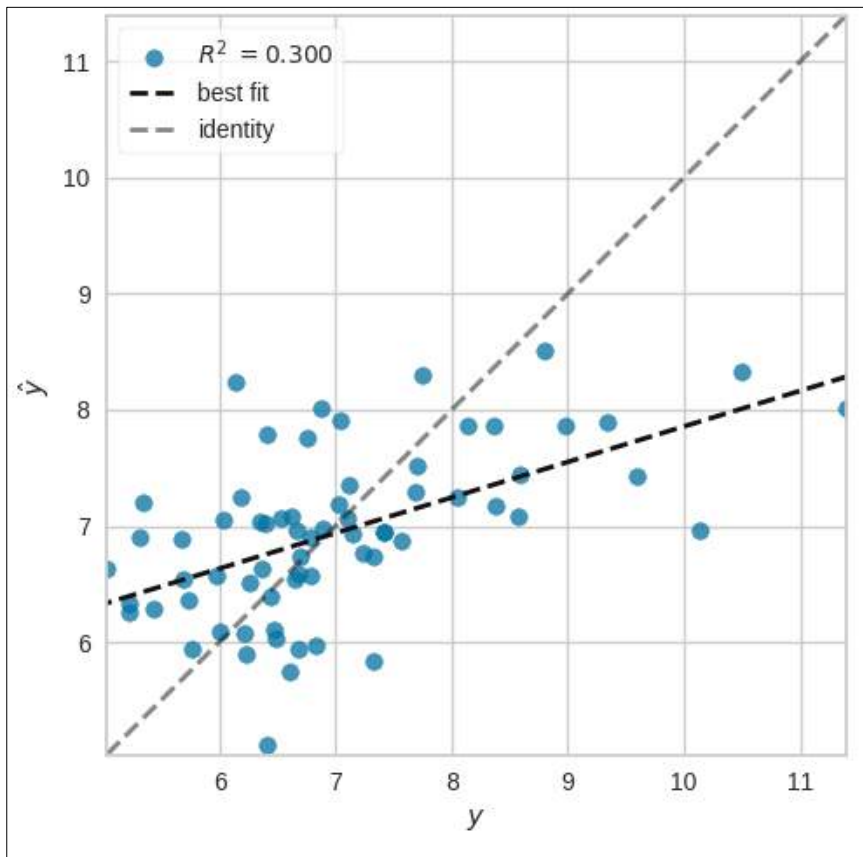


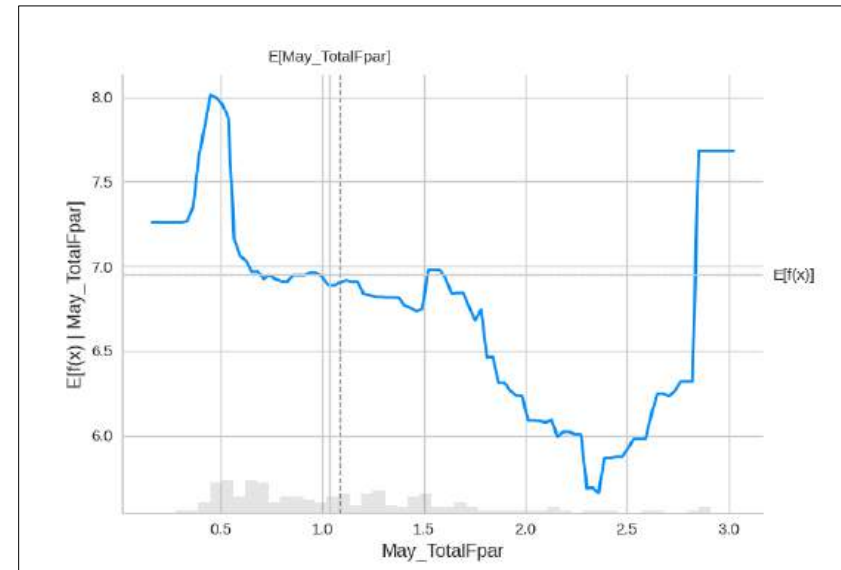
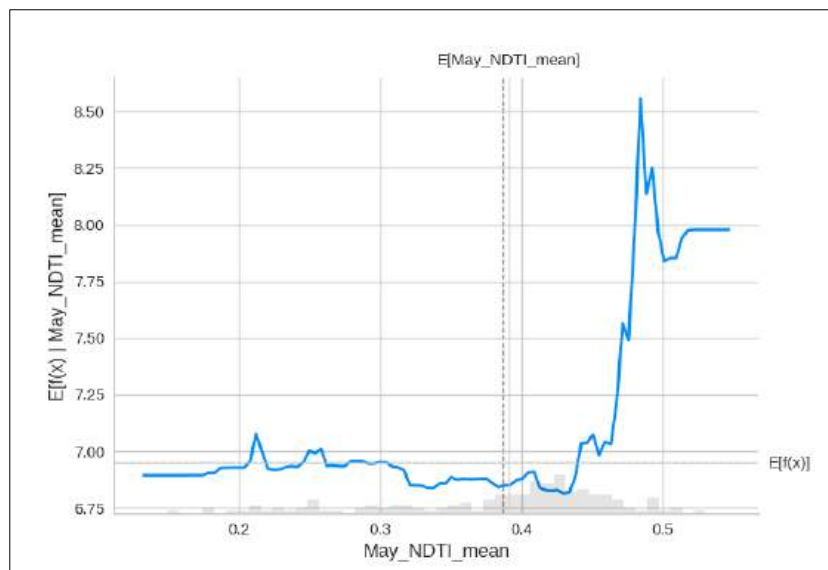
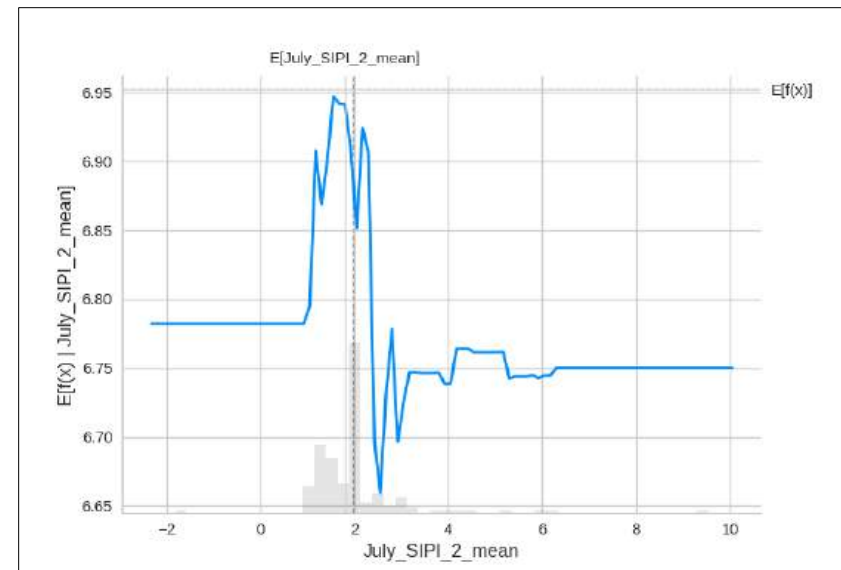
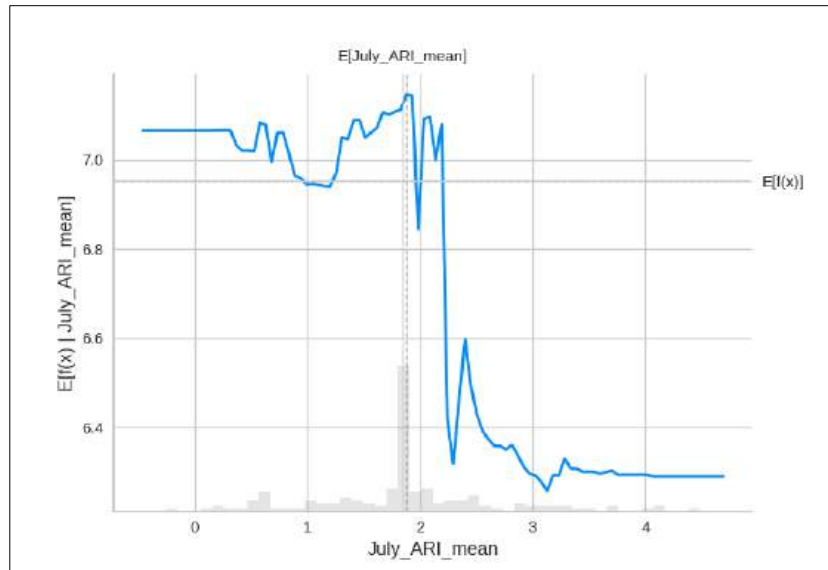




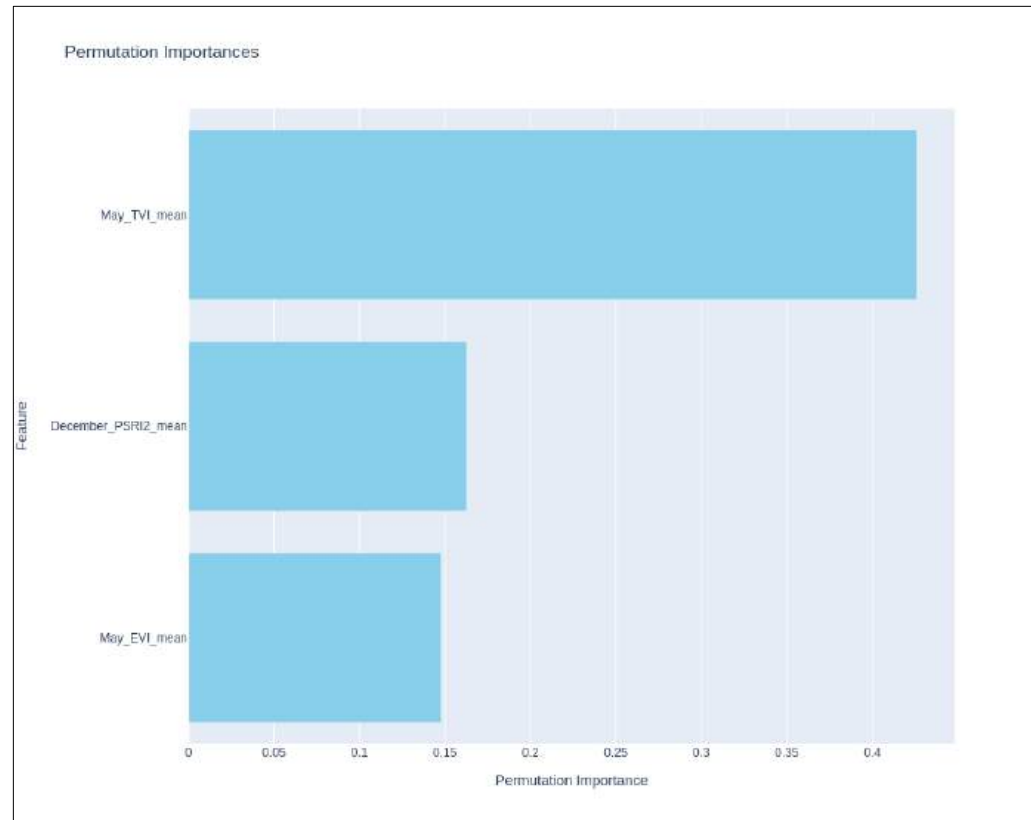
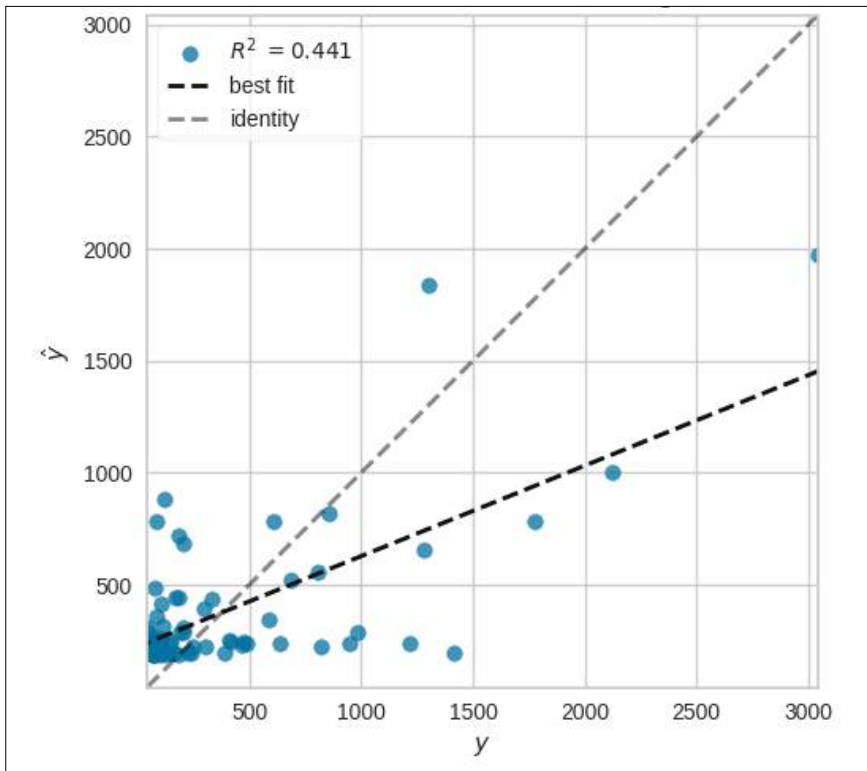


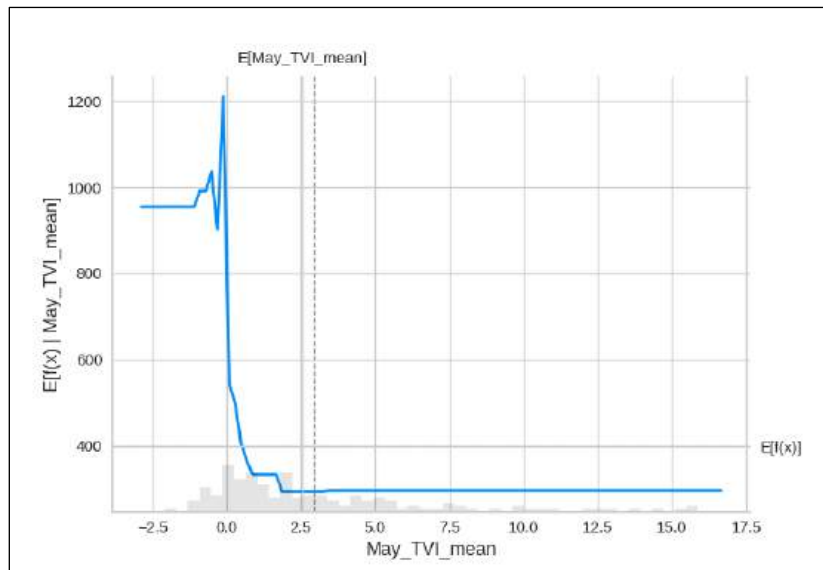
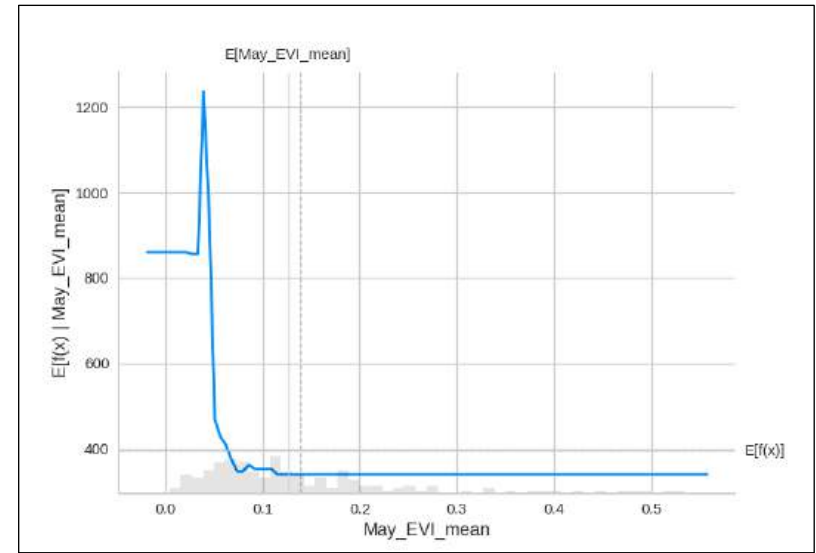
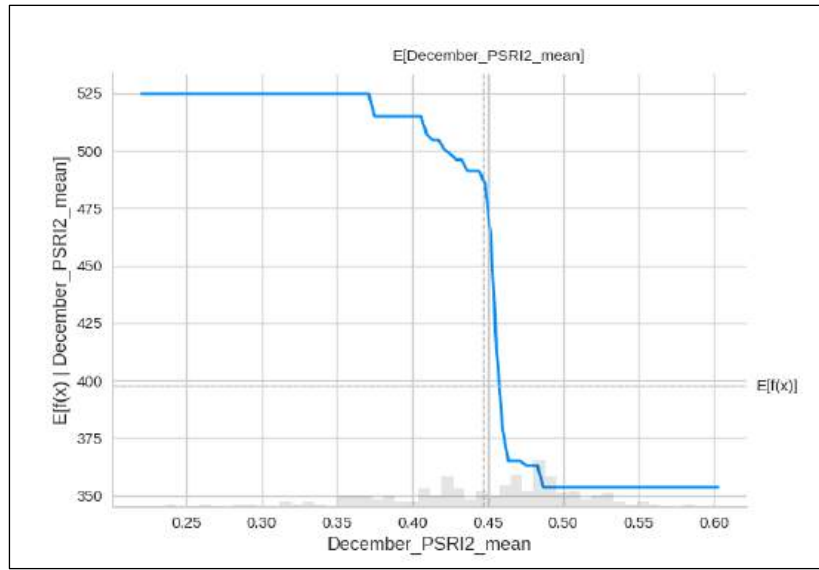
b) TEFF AND COPPER NUTRIENT COMPOSITION OF GRAINS



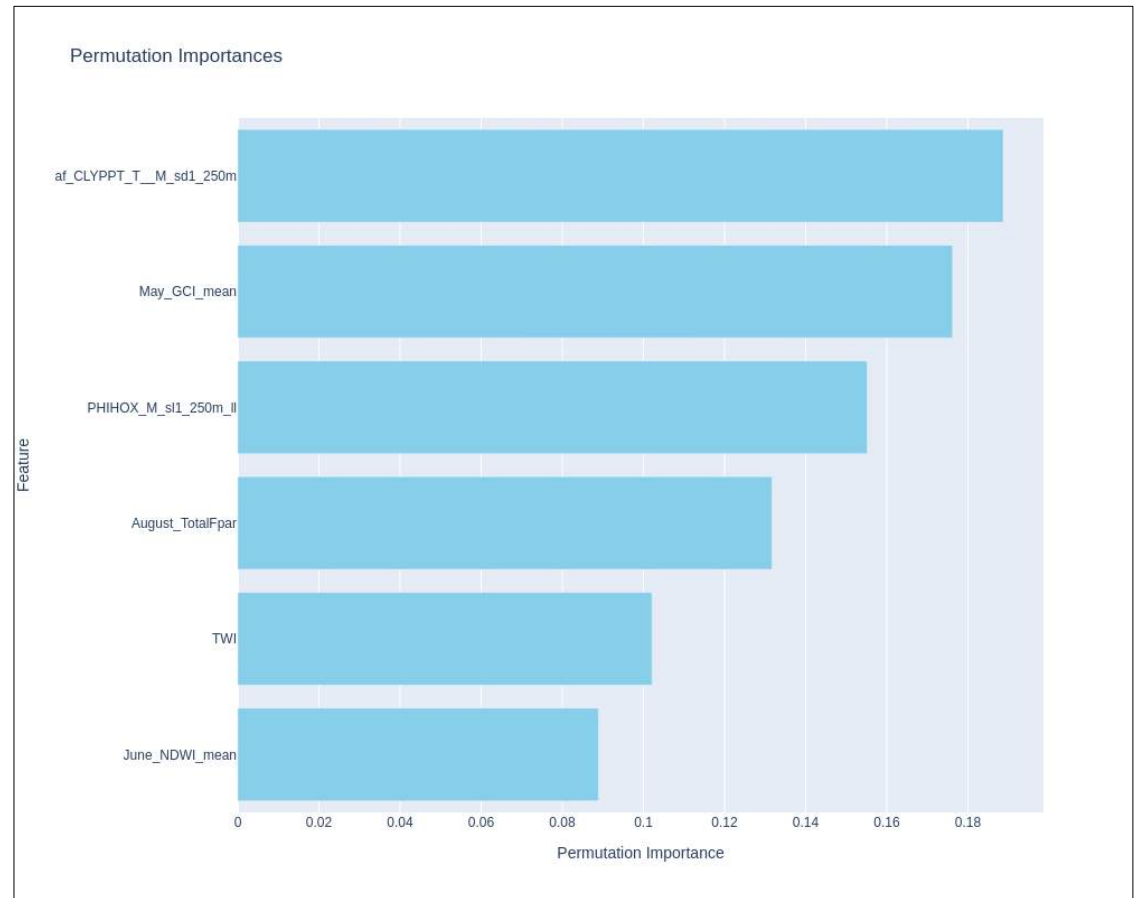
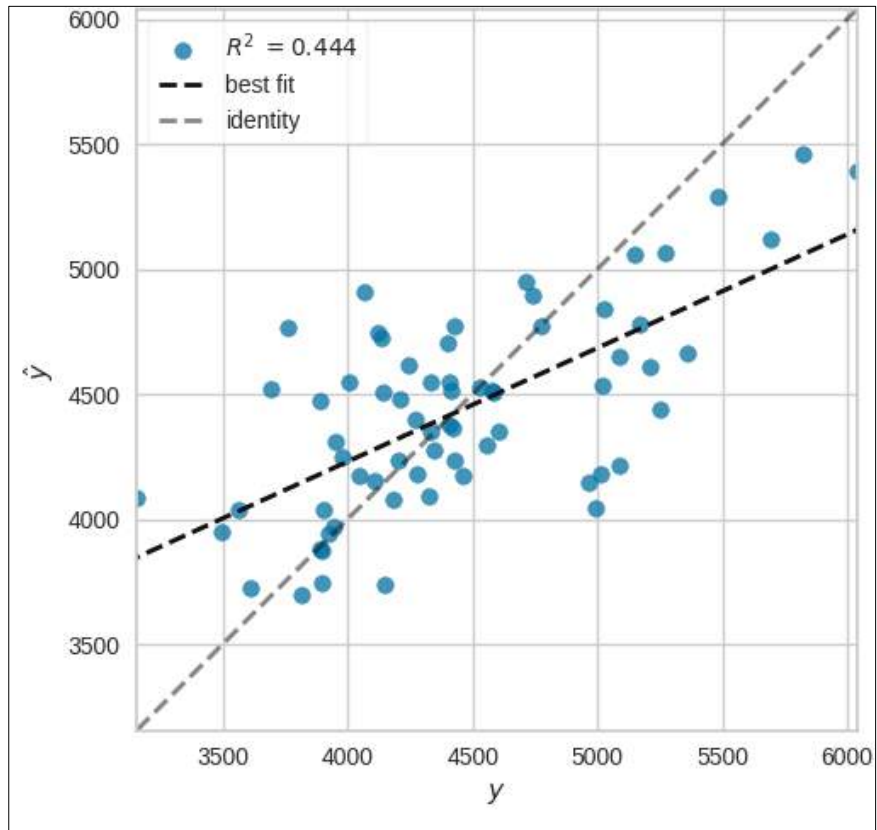


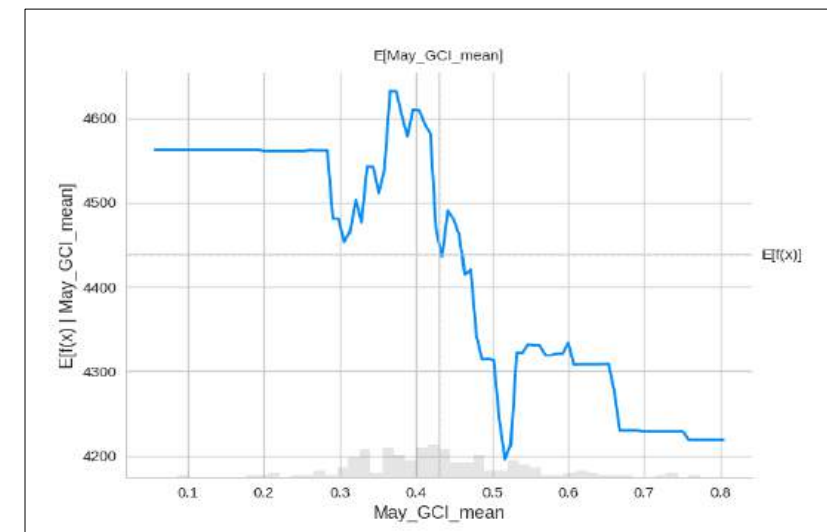
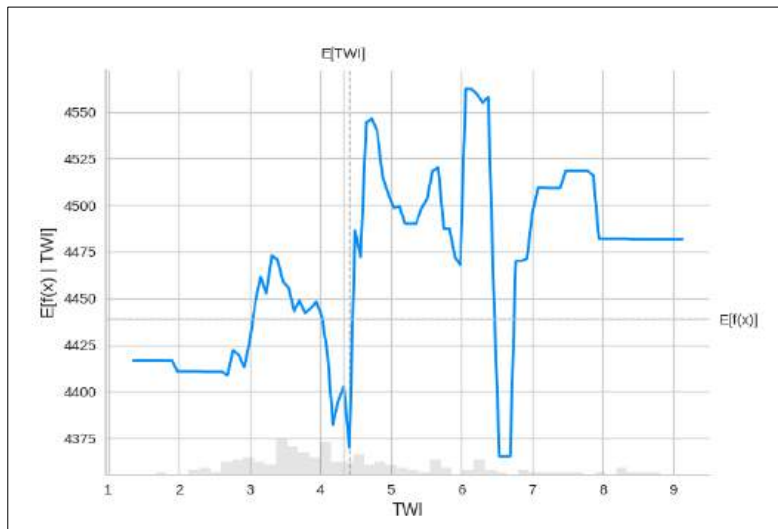
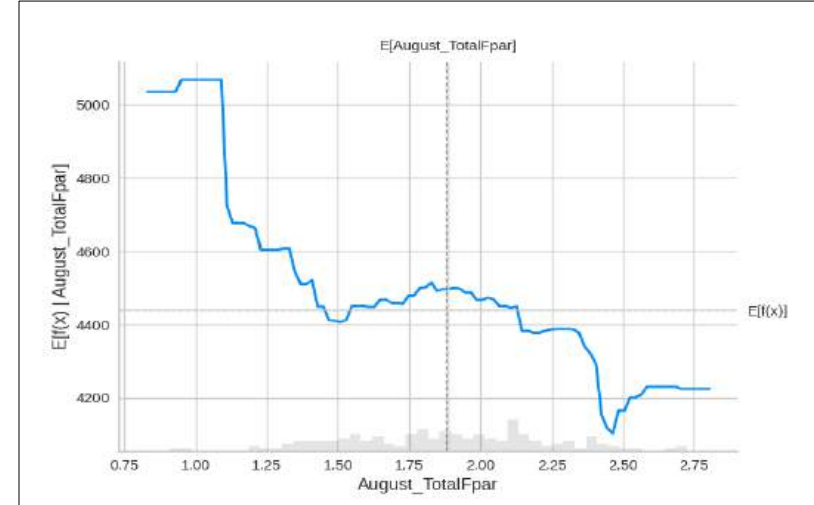
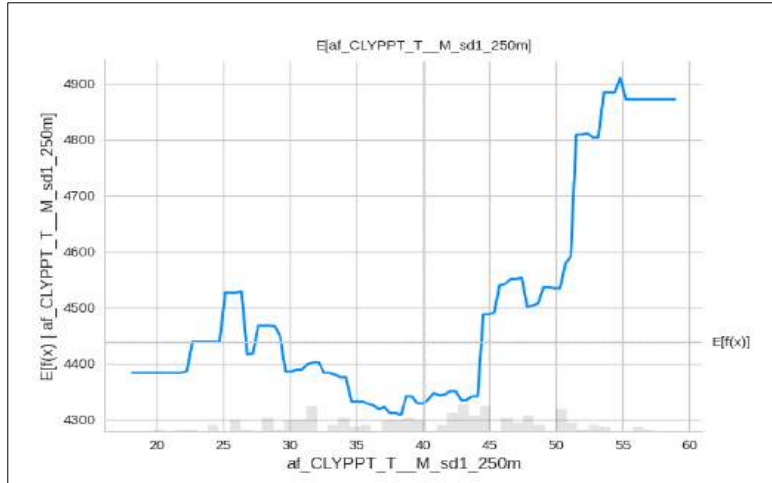
c) TEFF AND IRON NUTRIENT COMPOSITION OF GRAINS

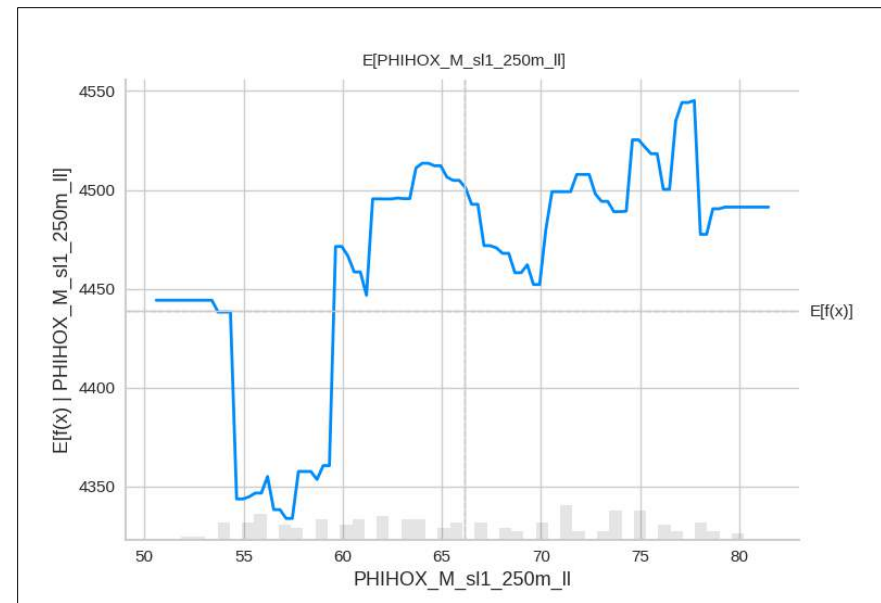
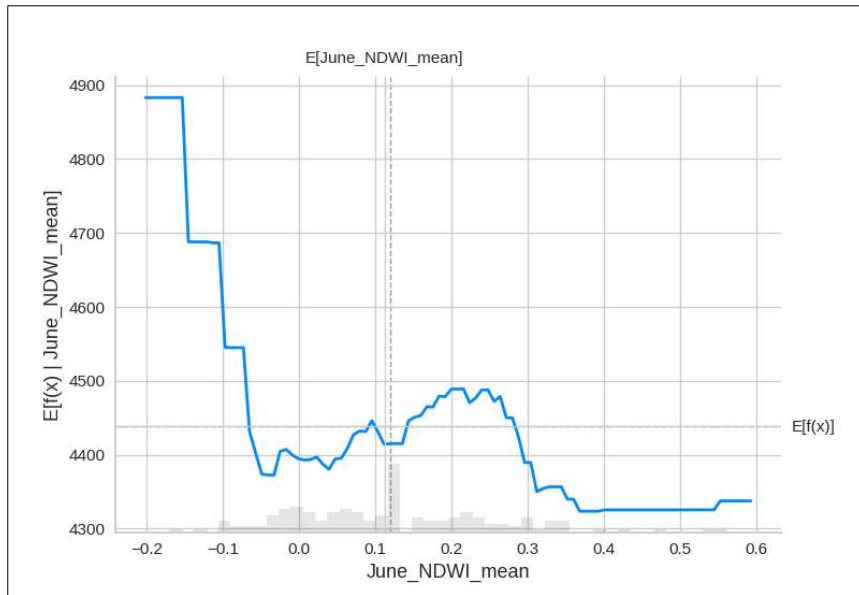




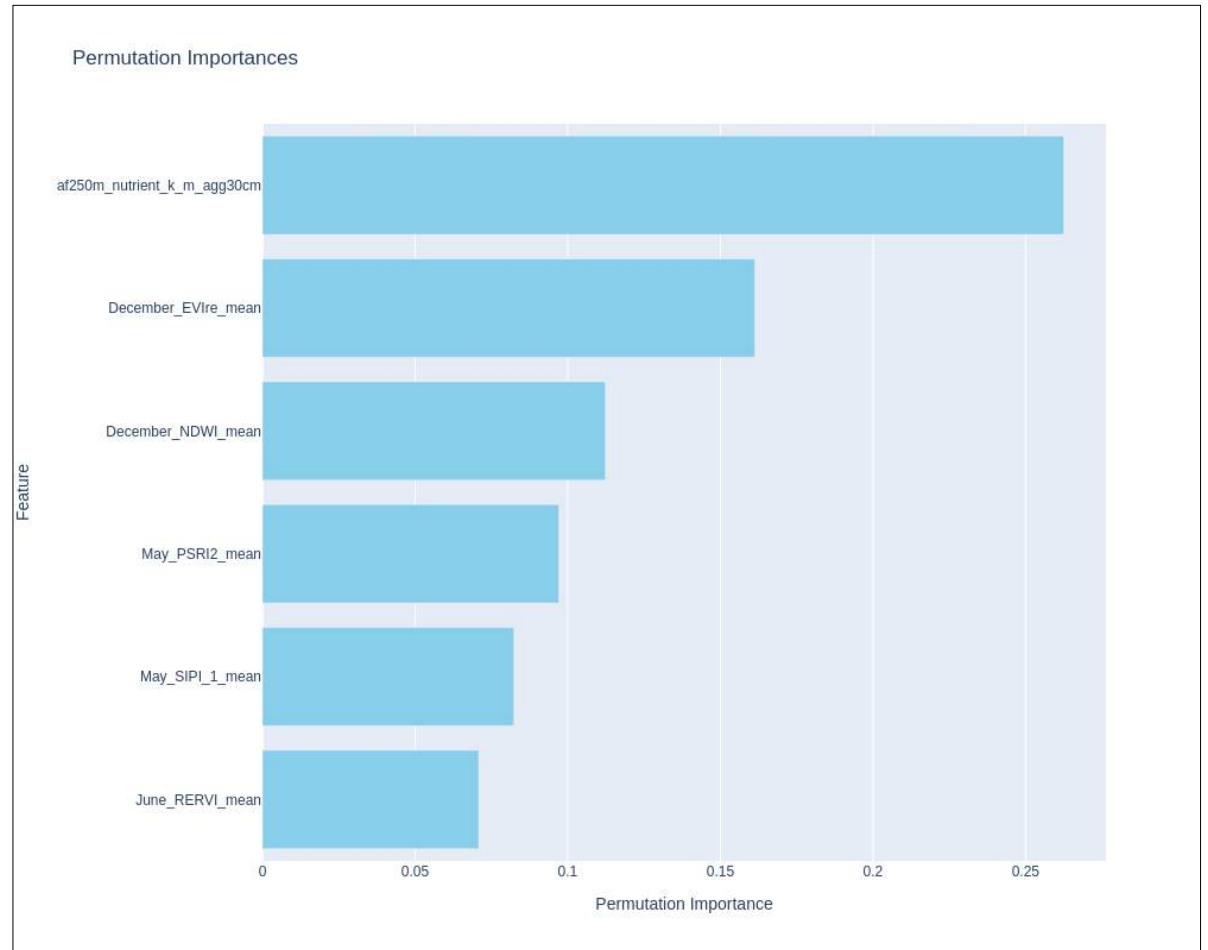
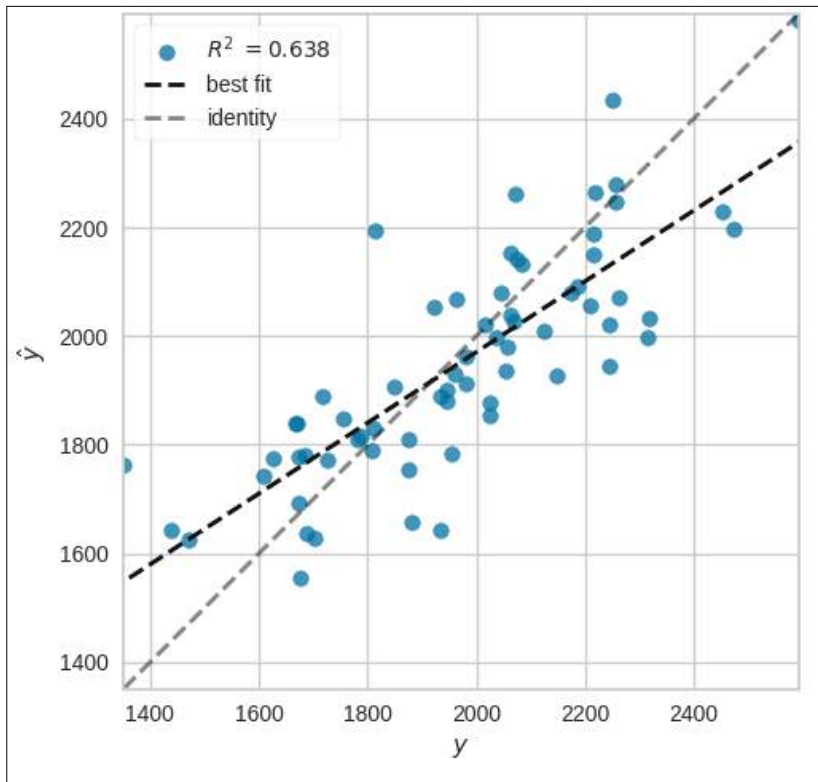
d) TEFF AND POTASSIUM NUTRIENT COMPOSITION OF GRAINS



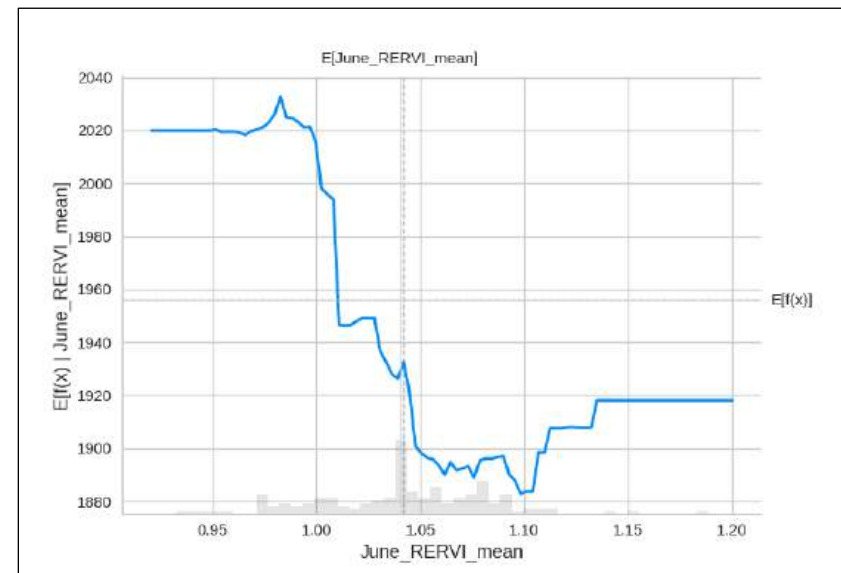
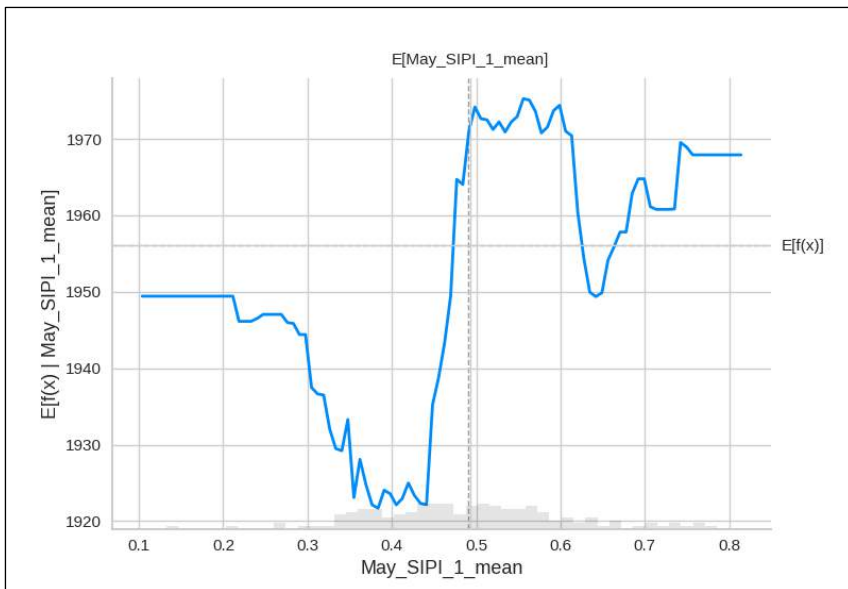
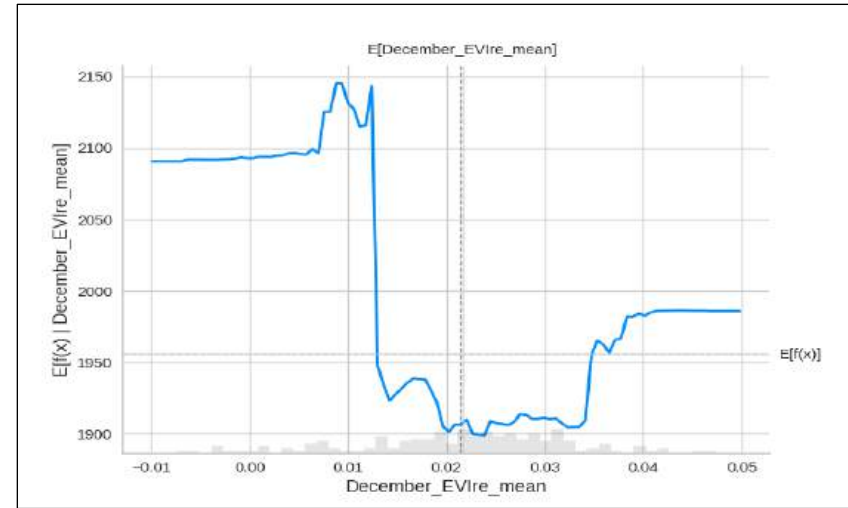
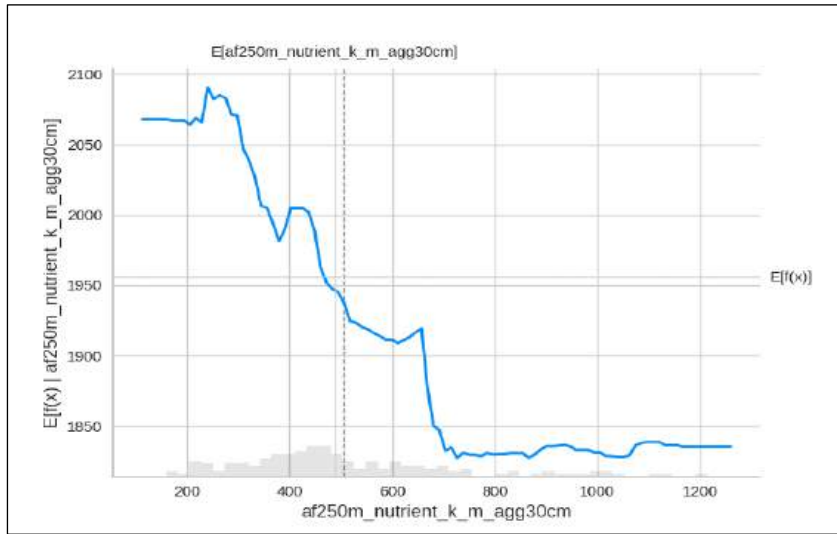


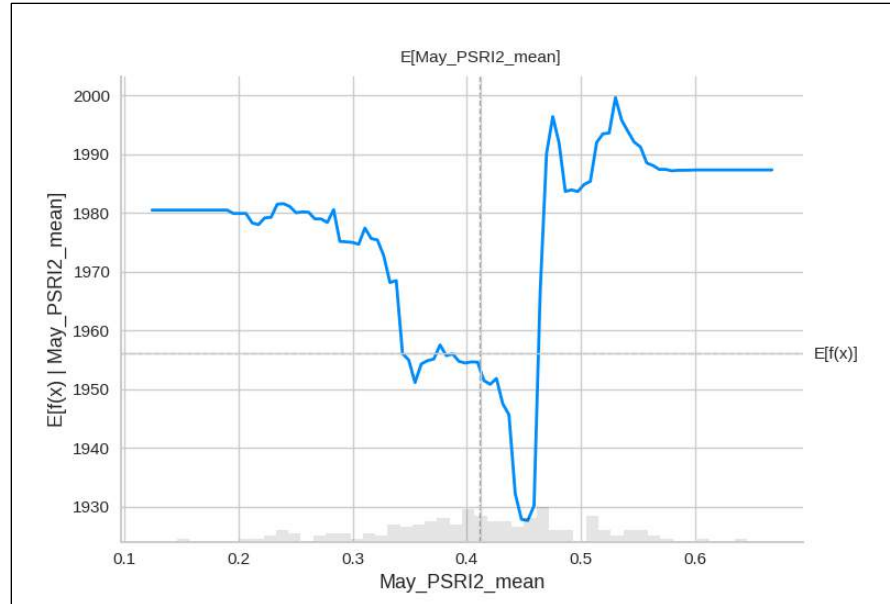
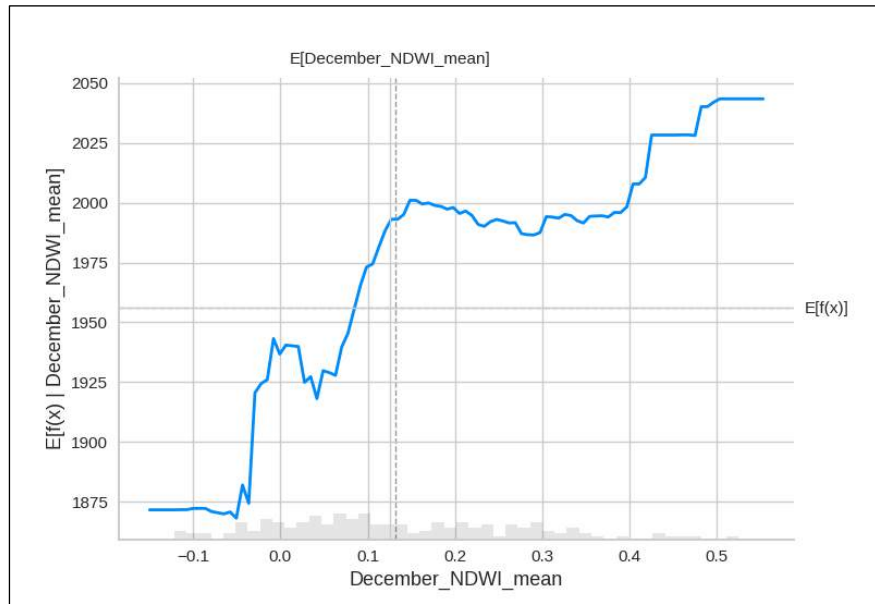


e) TEFF AND MAGNESIUM NUTRIENT COMPOSITION OF GRAINS

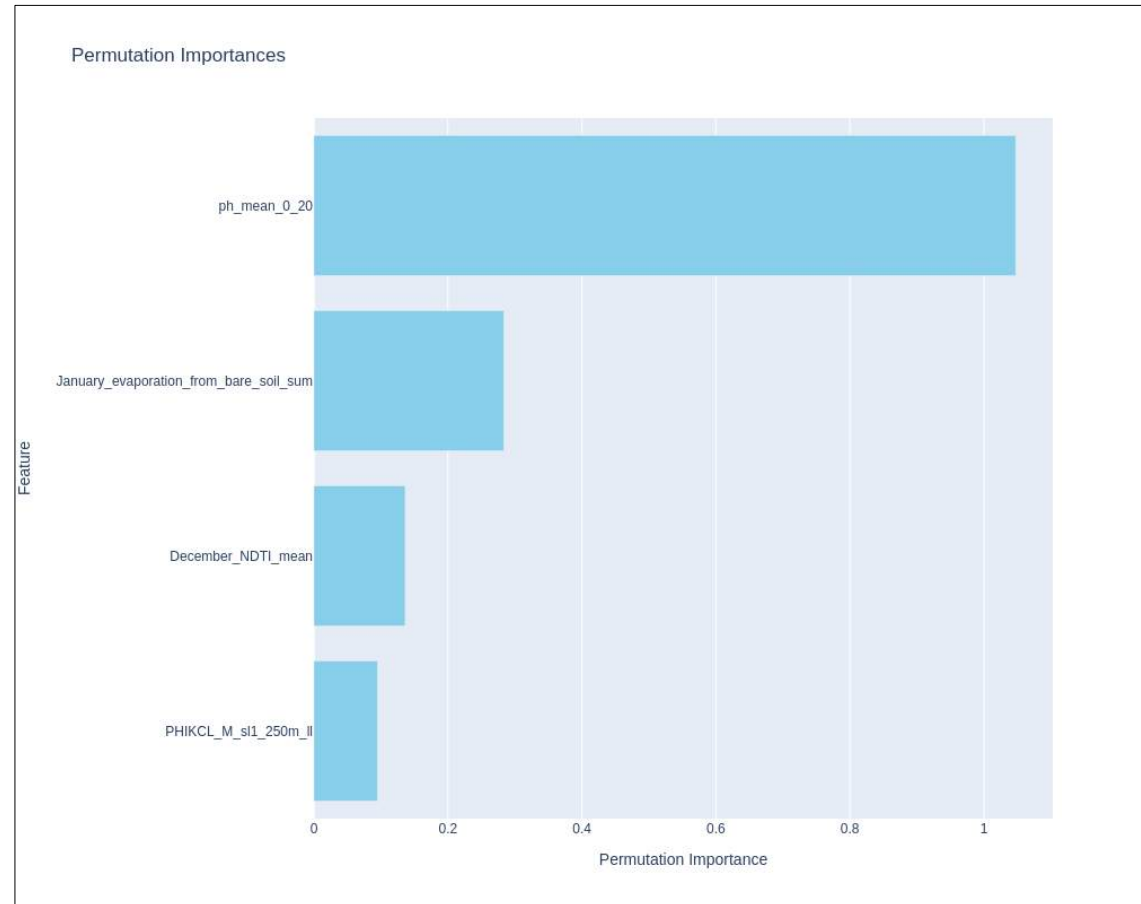
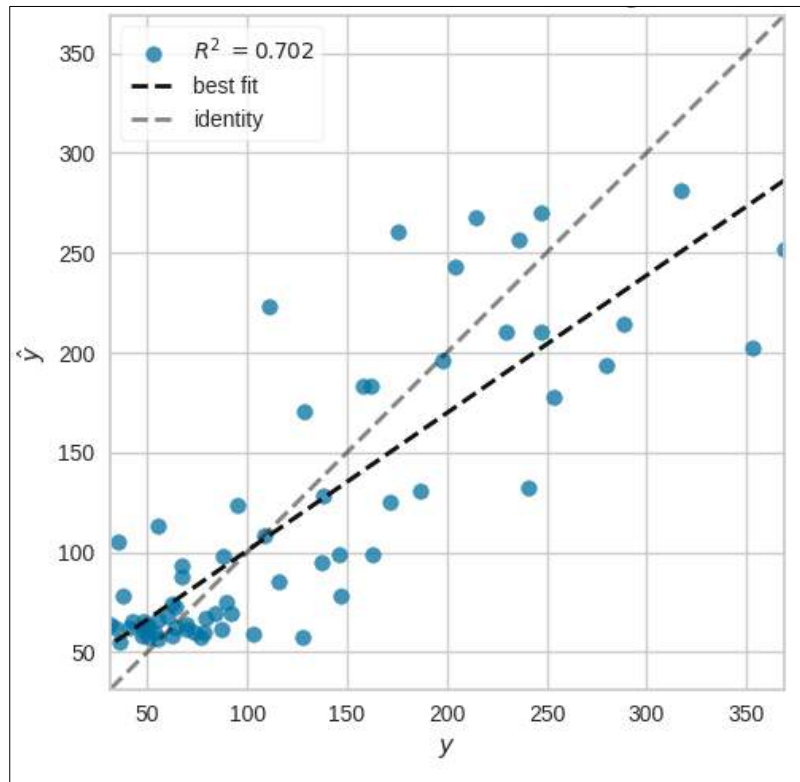


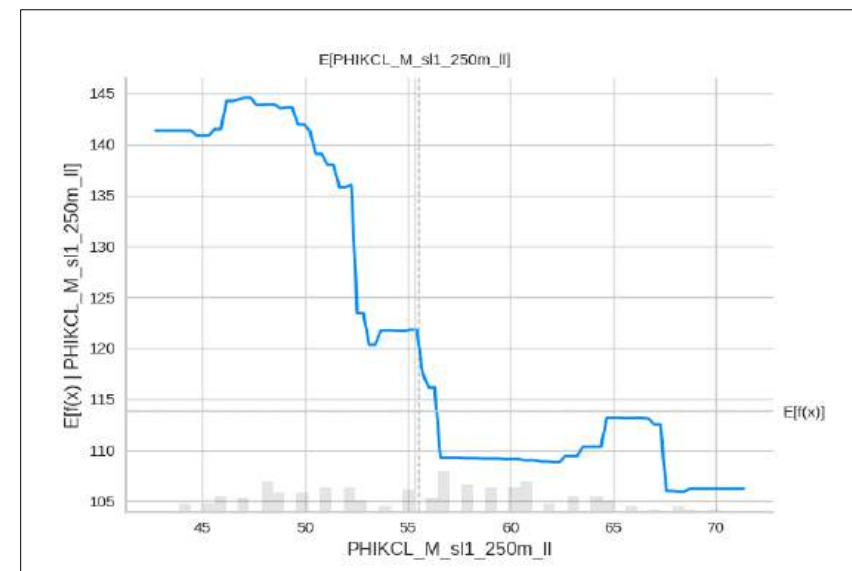
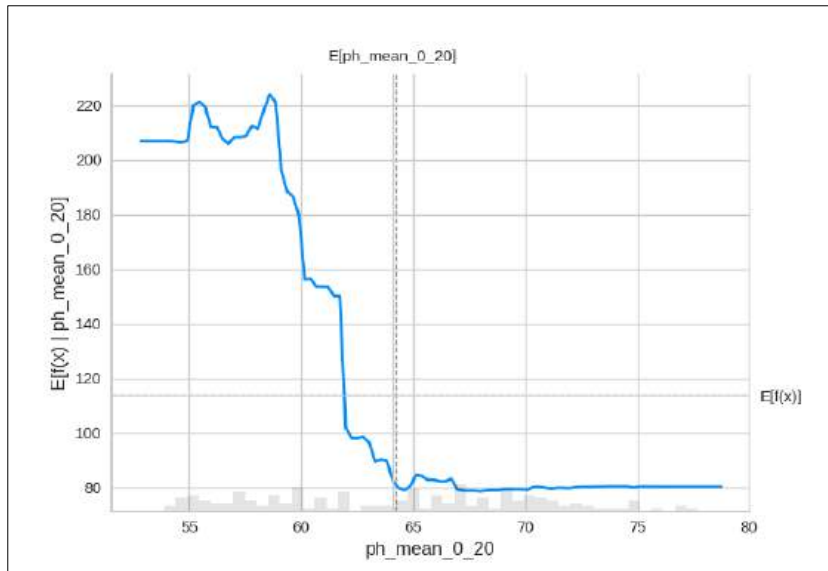
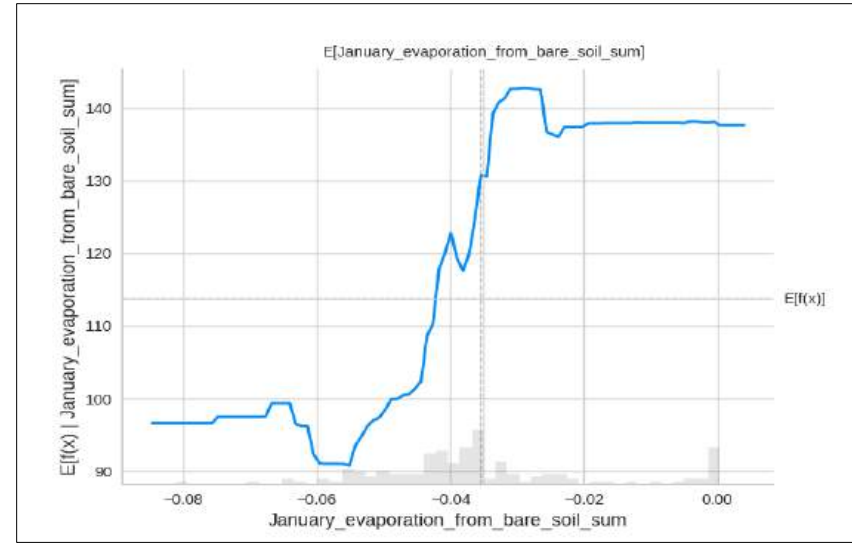
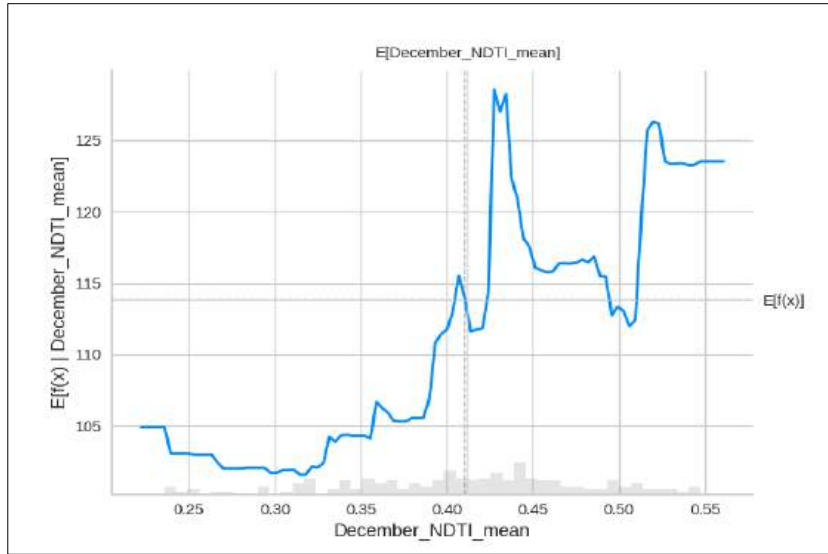




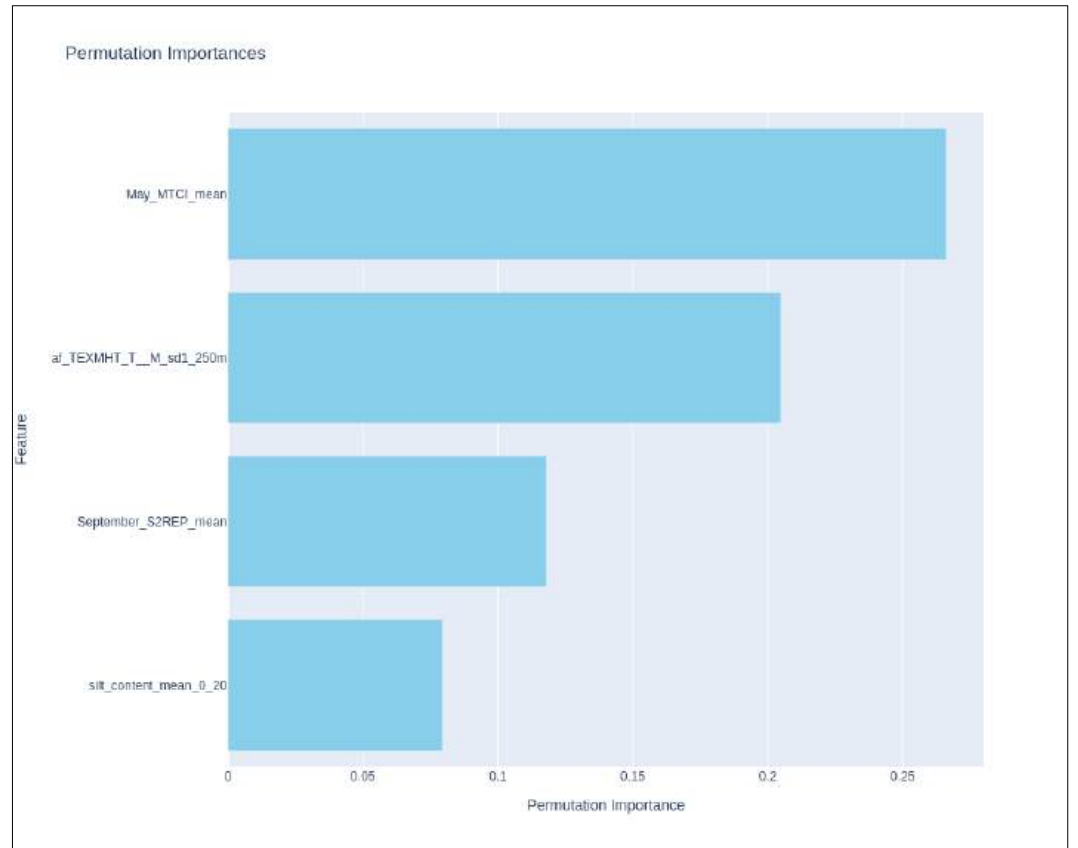
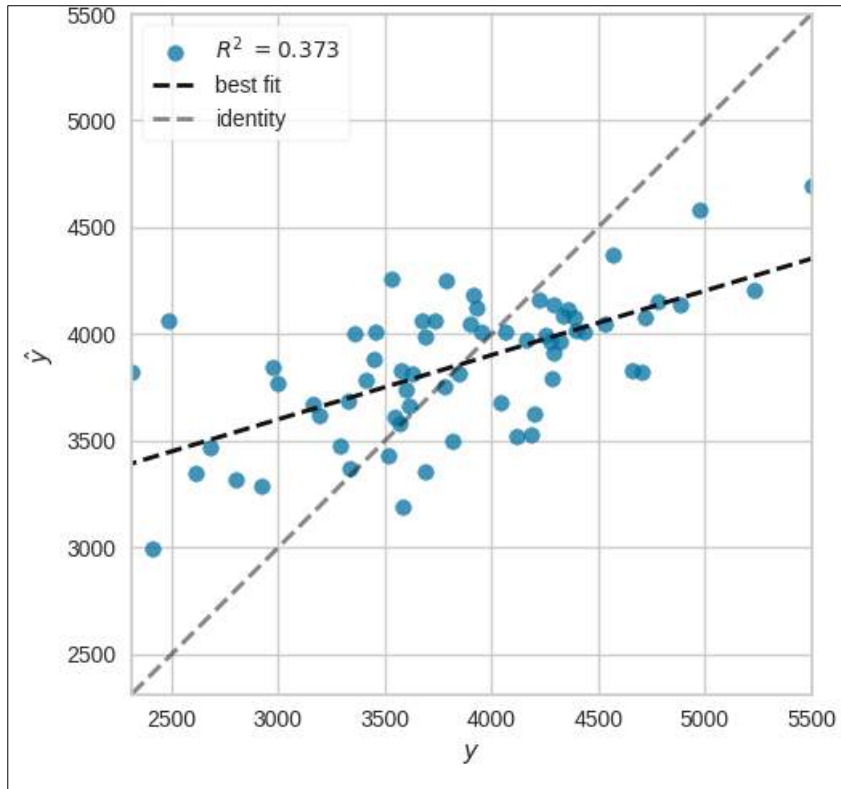


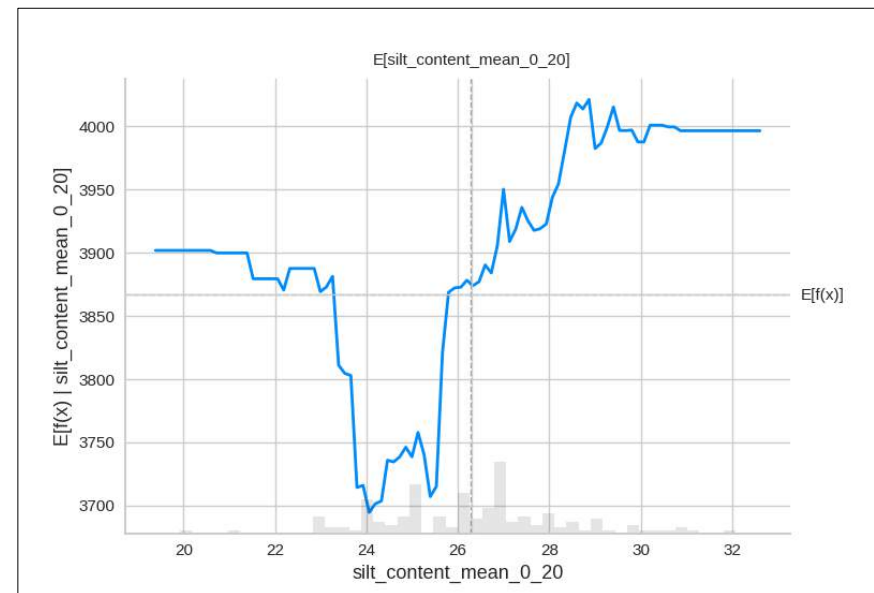
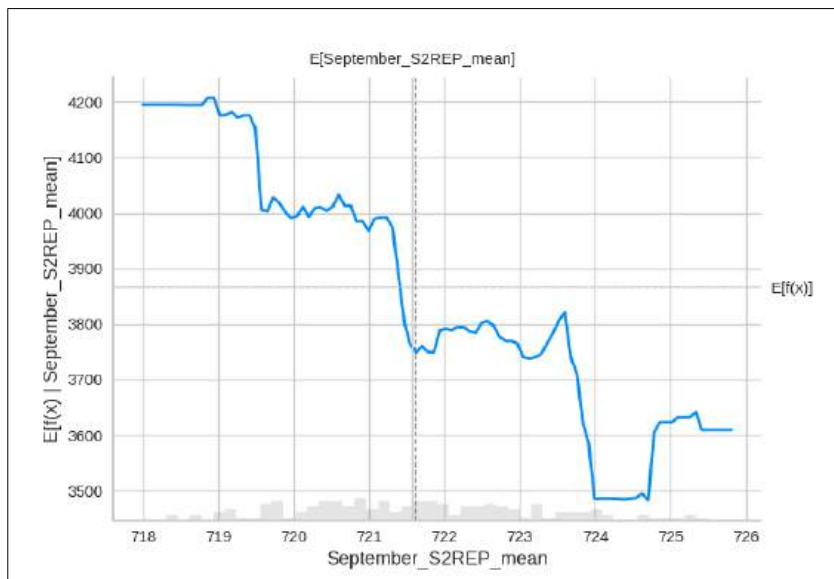
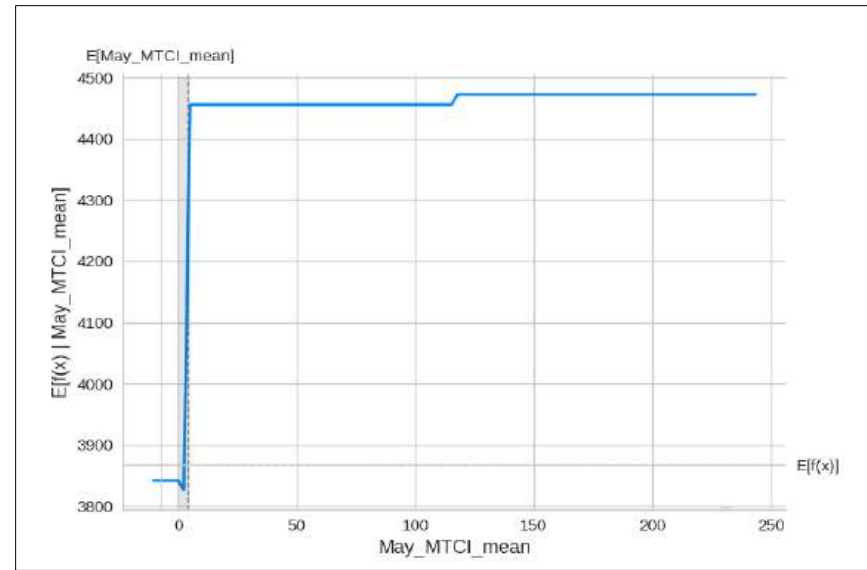
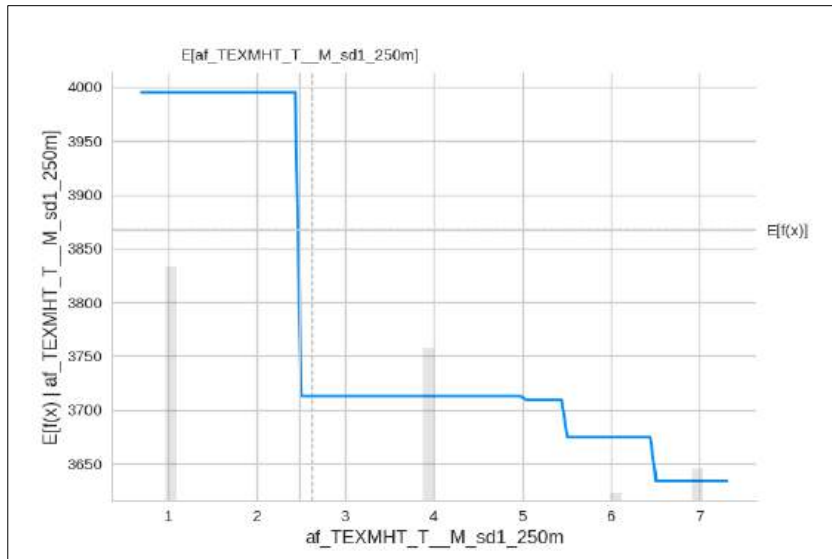
f) TEFF AND MANGANESE NUTRIENT COMPOSITION OF GRAINS



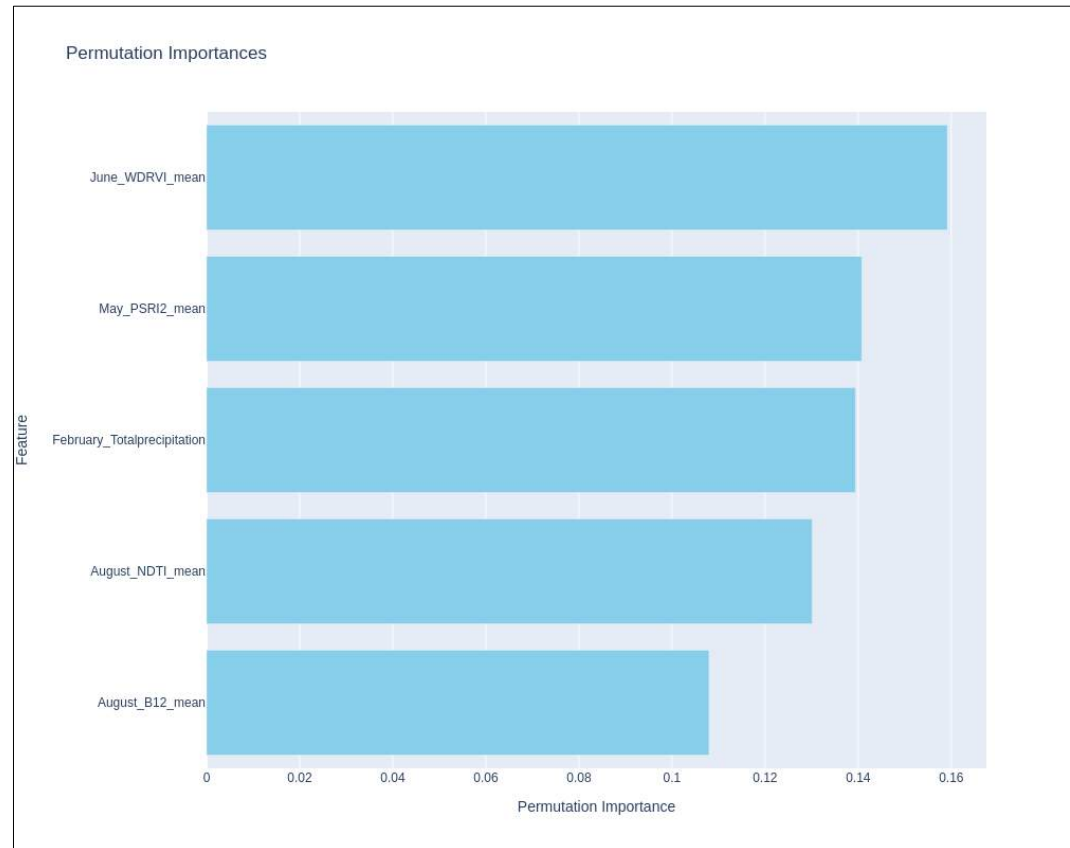
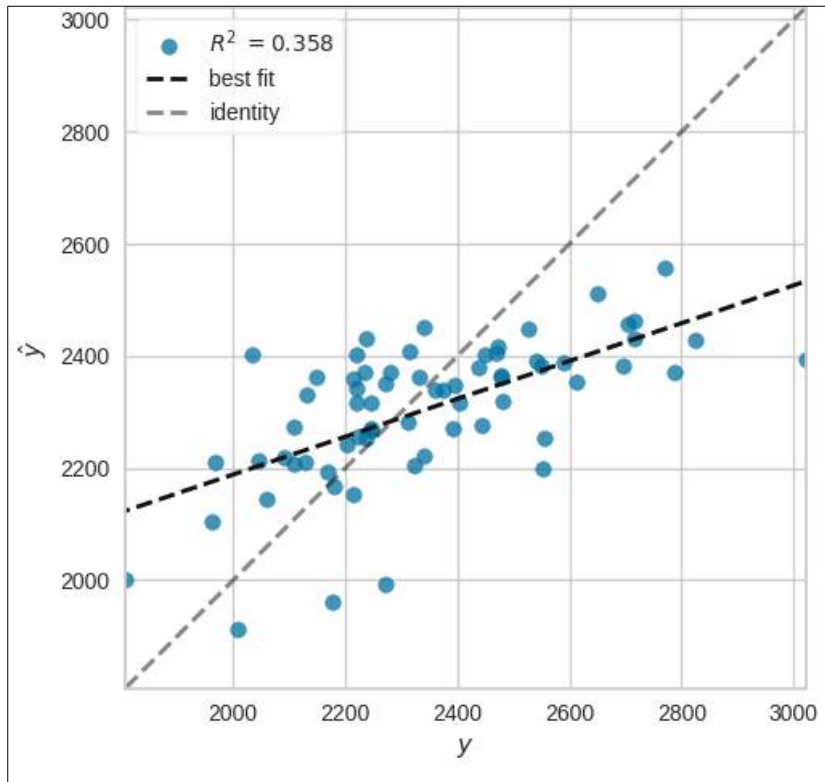


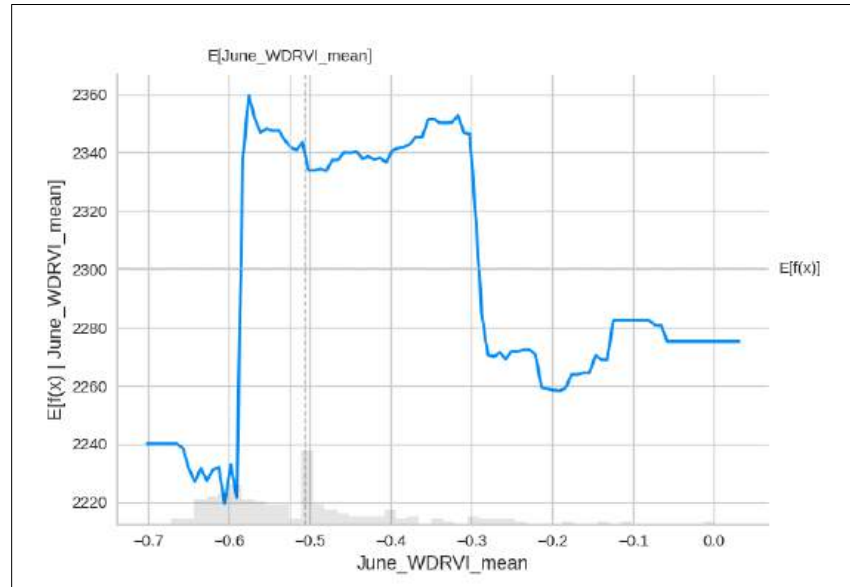
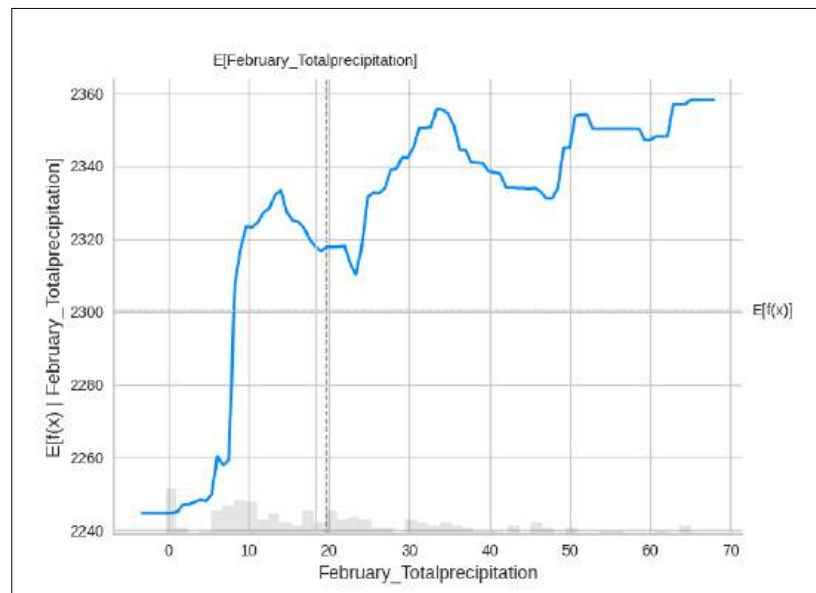
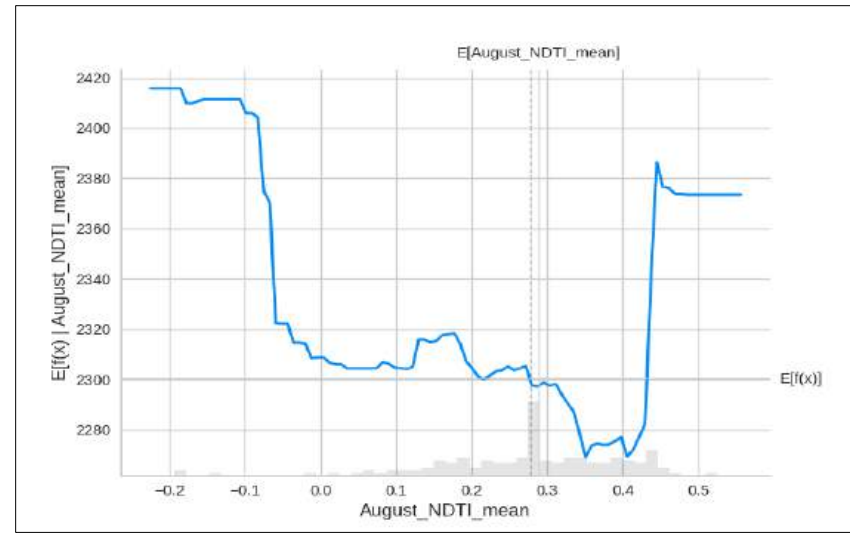
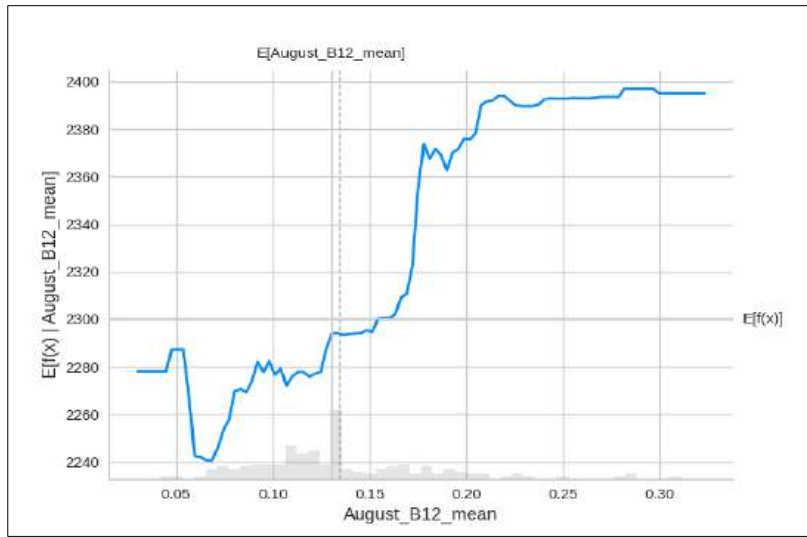
g) TEFF AND PHOSPHORUS NUTRIENT COMPOSITION OF GRAINS



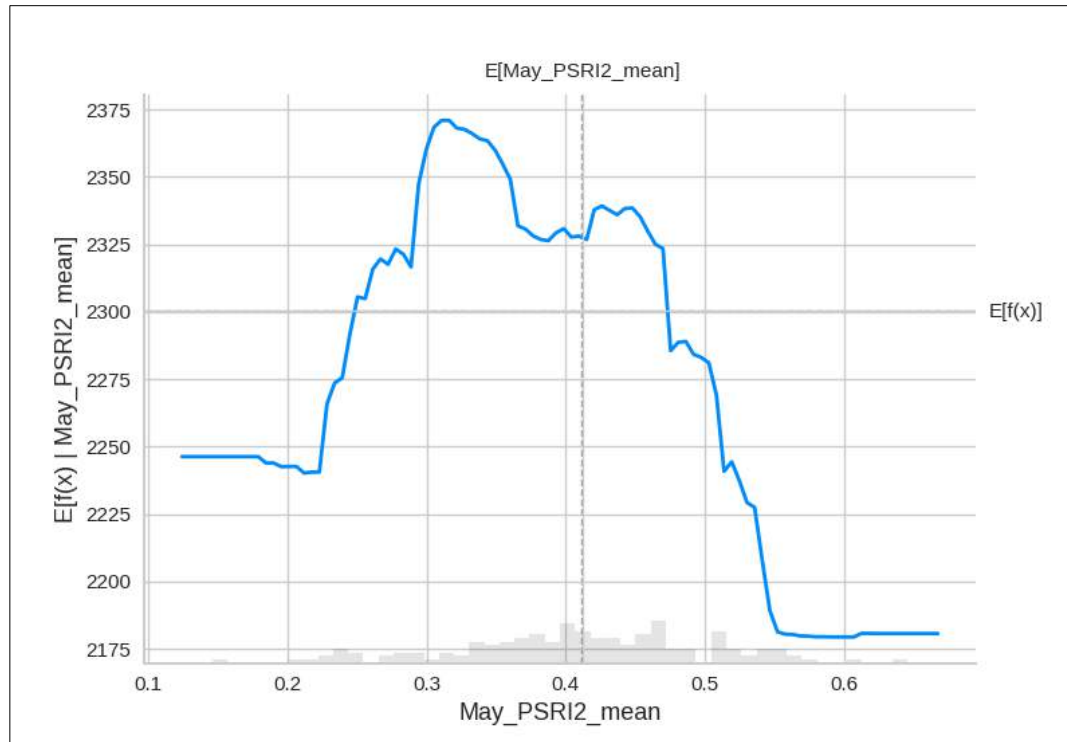


### h) TEFF AND SULPHUR NUTRIENT COMPOSITION OF GRAINS

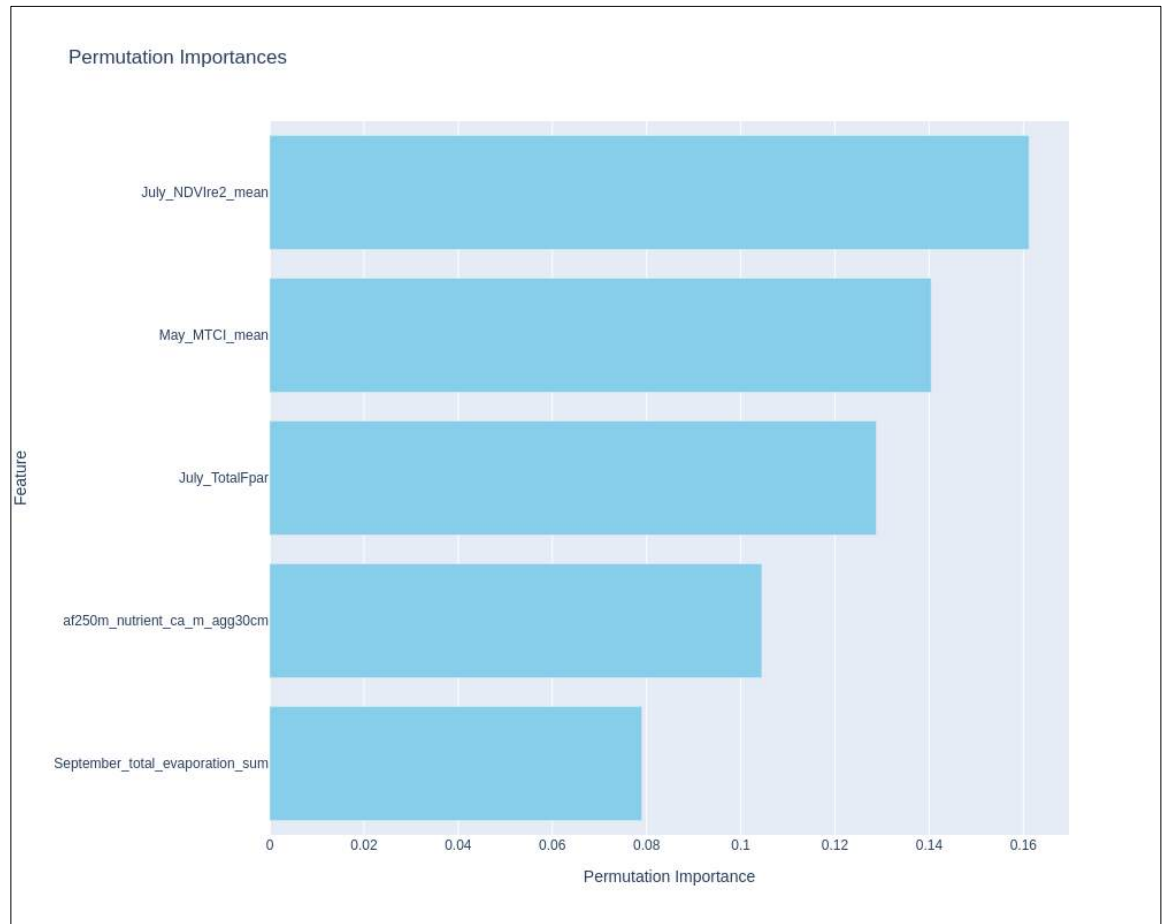
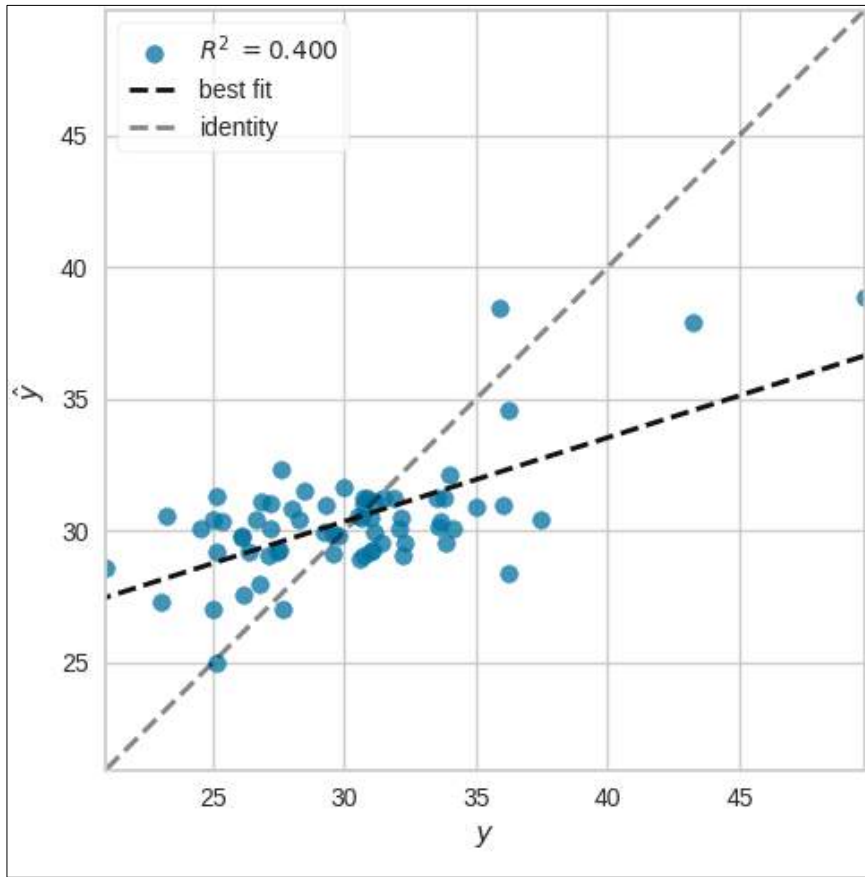


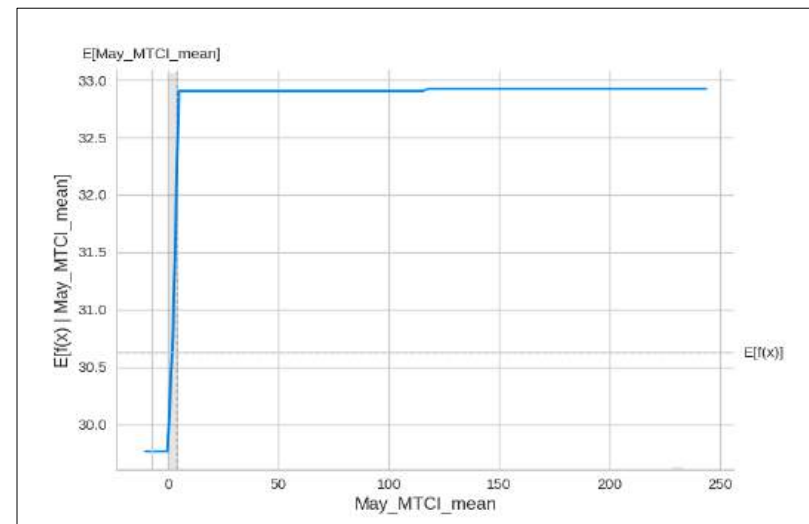
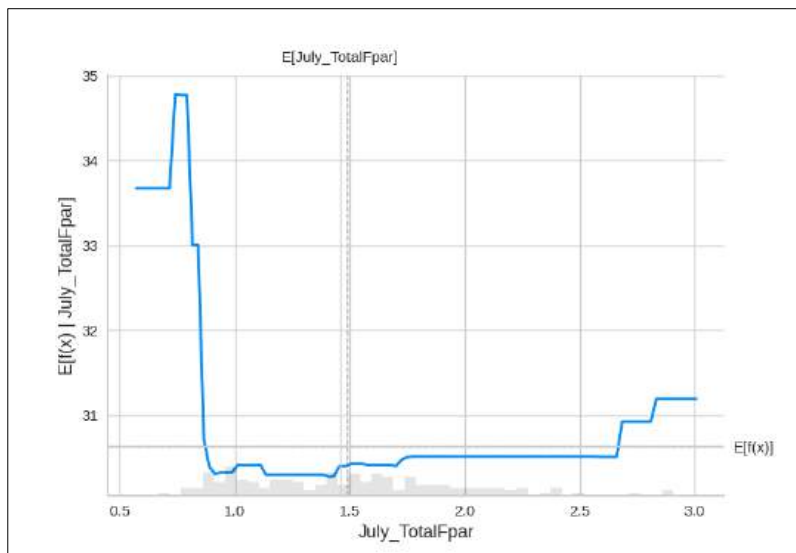
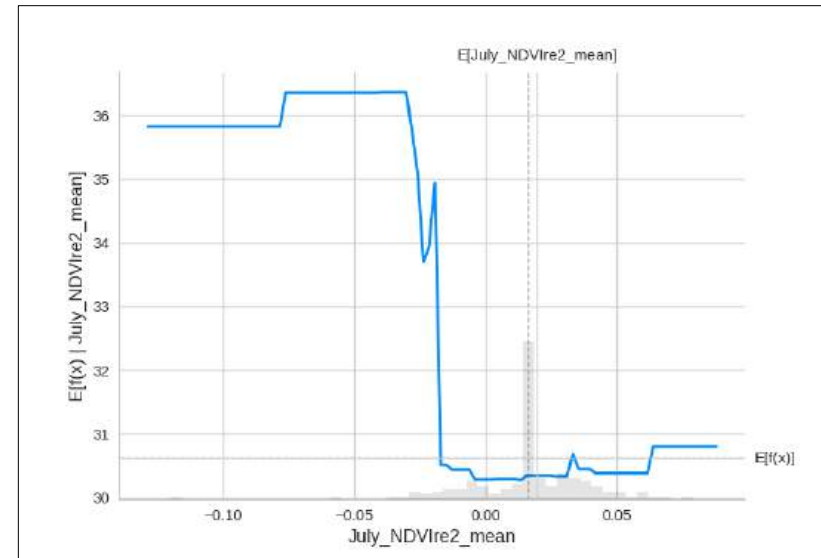
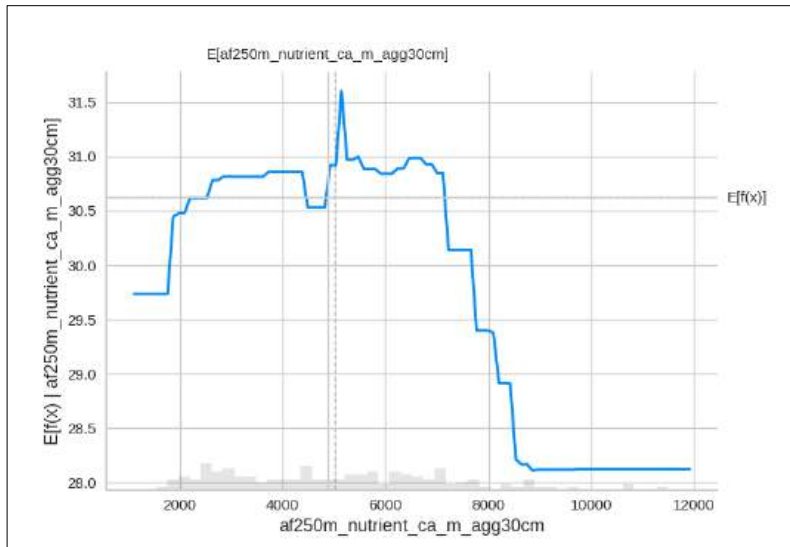


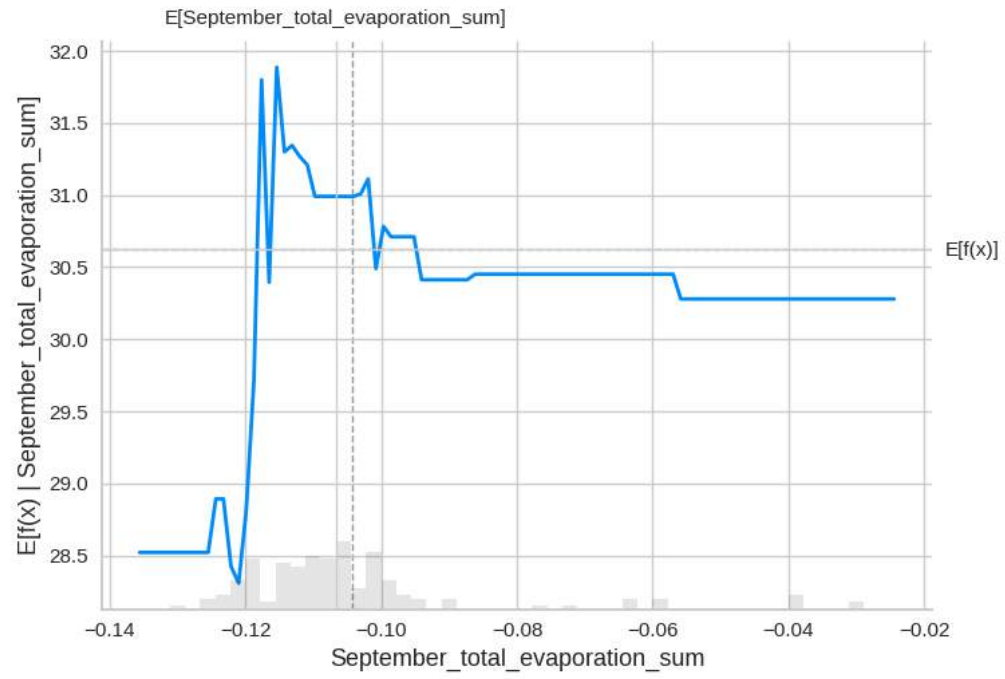




i) TEFF AND ZINC NUTRIENT COMPOSITION OF GRAINS

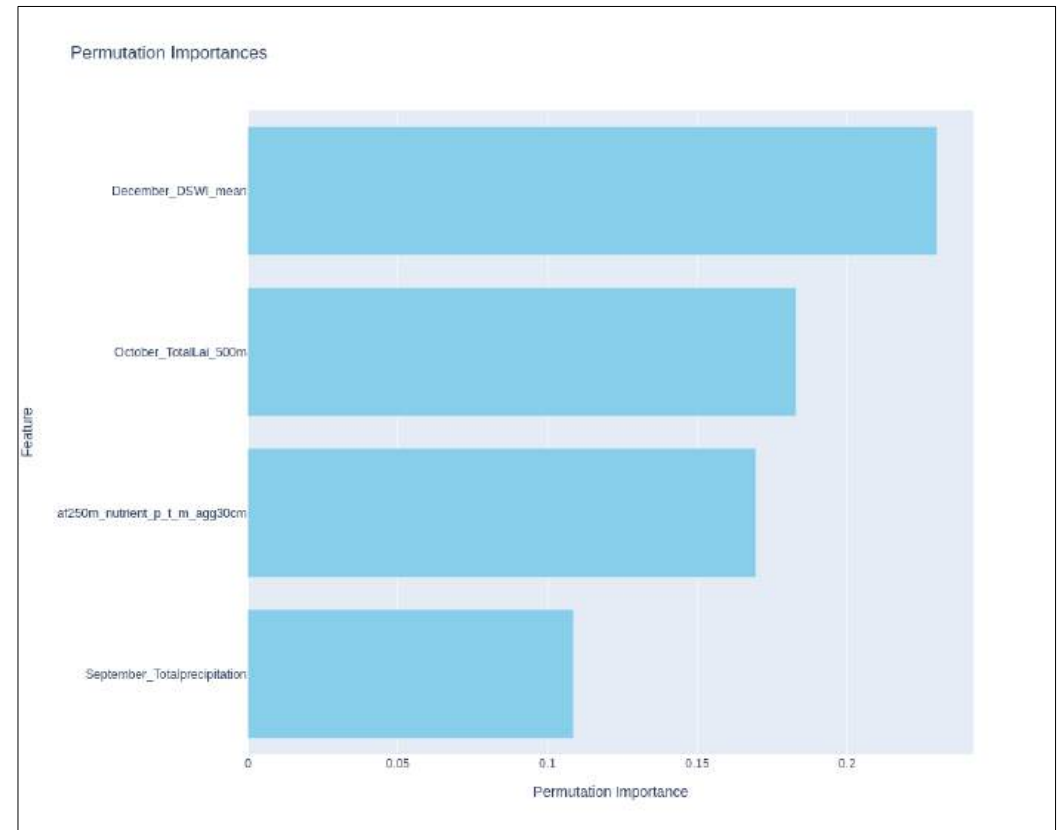
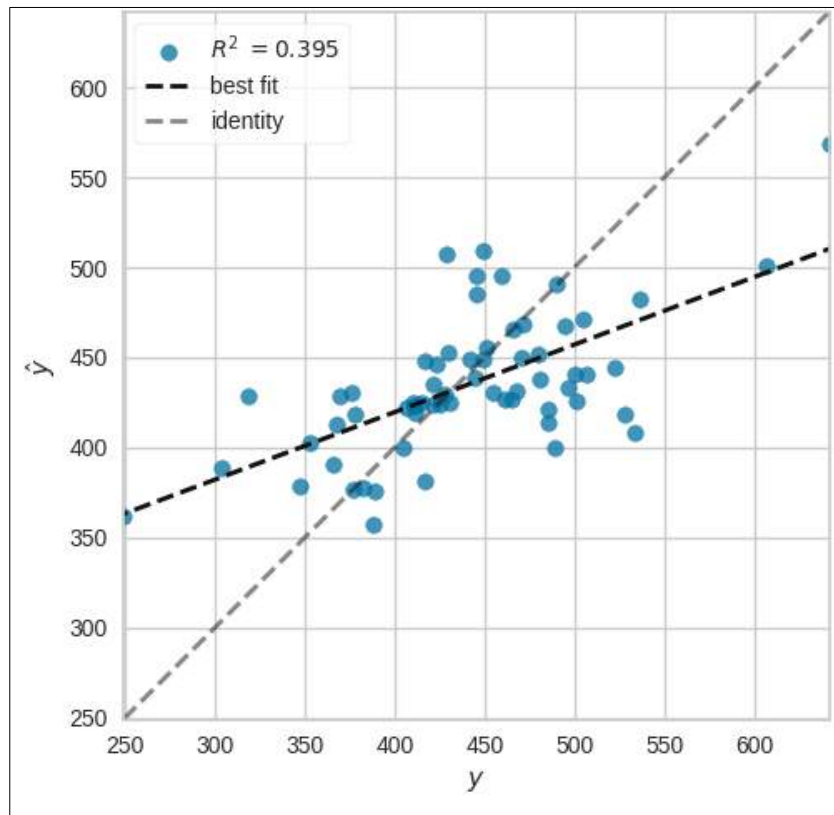


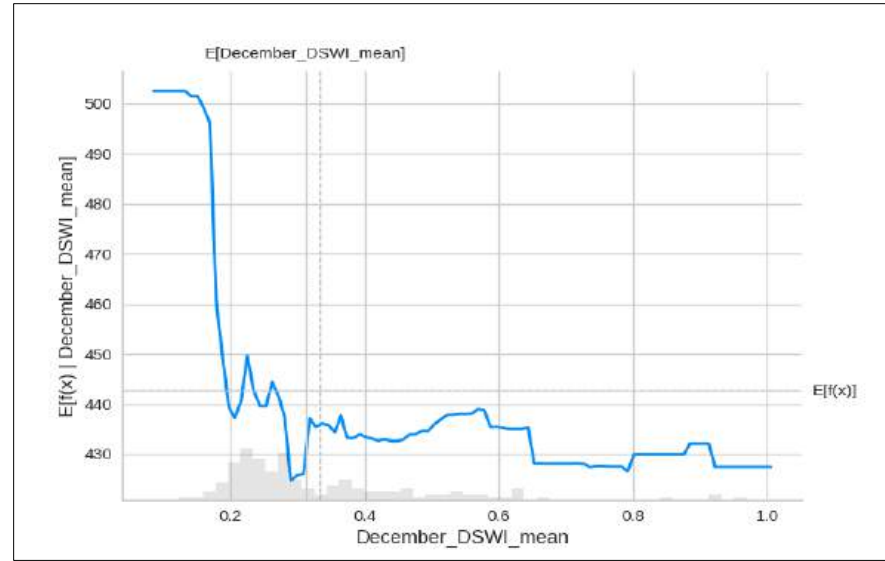
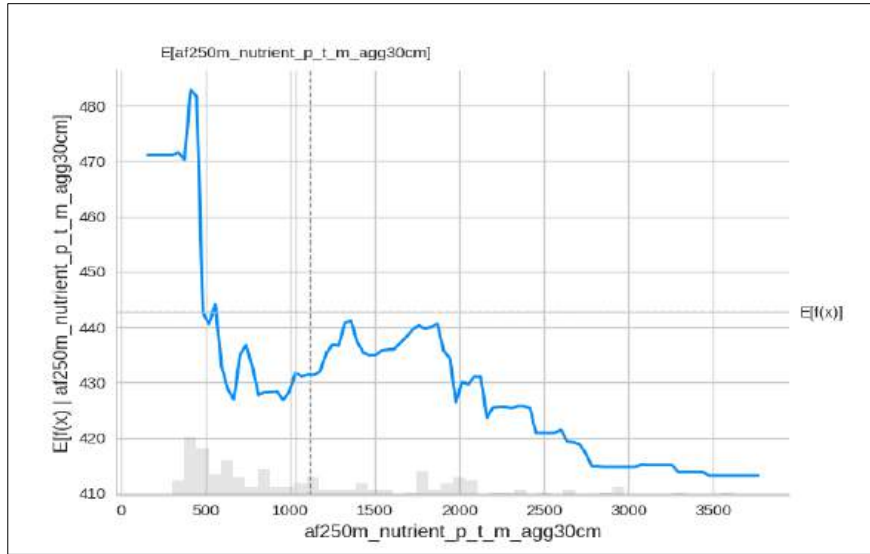


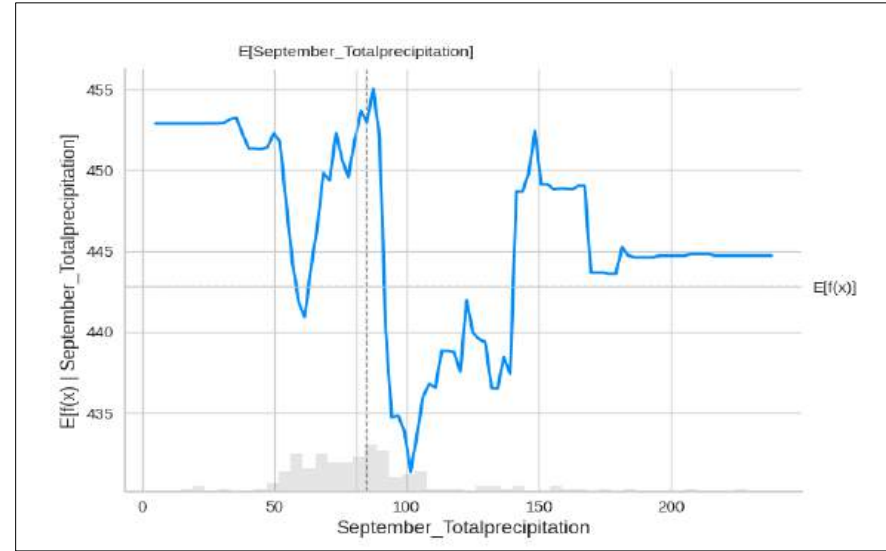
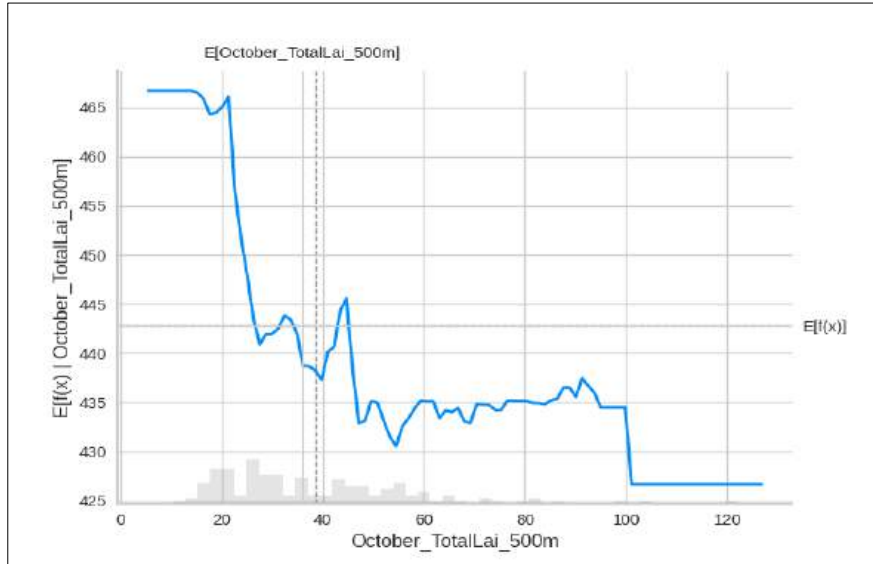


## Appendix 6. Results for wheat

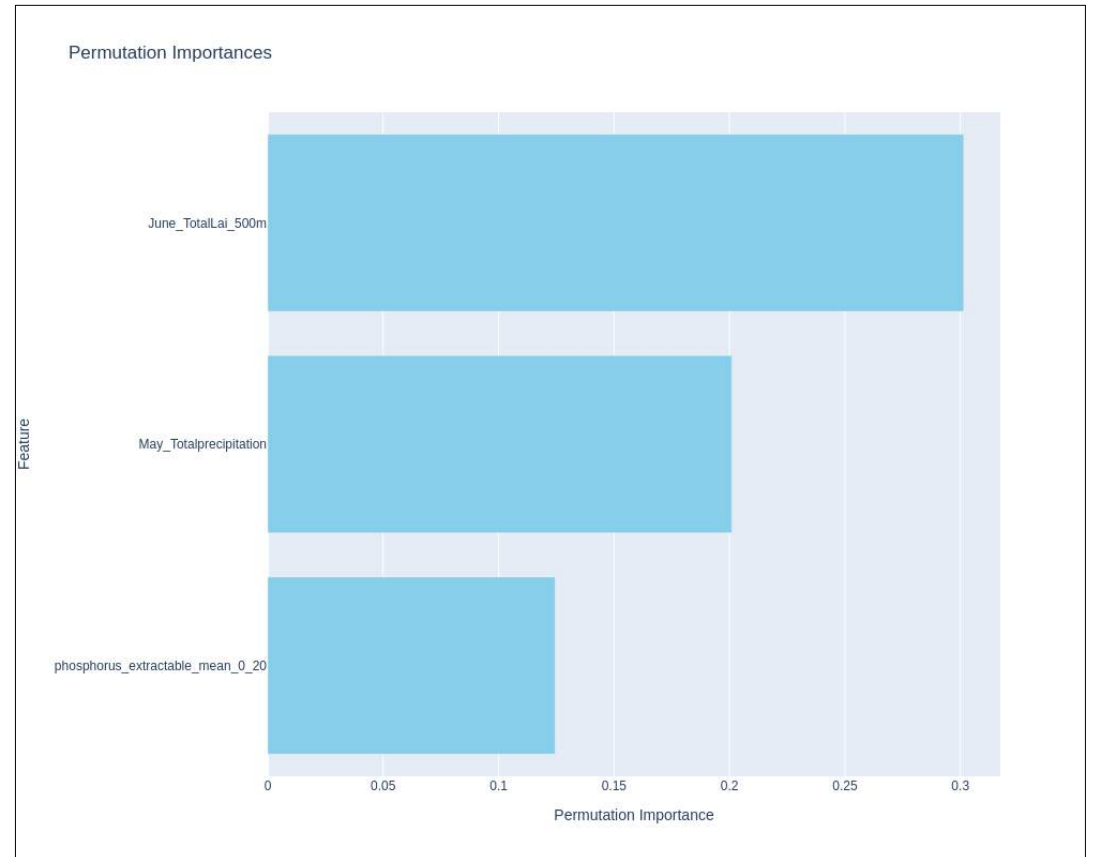
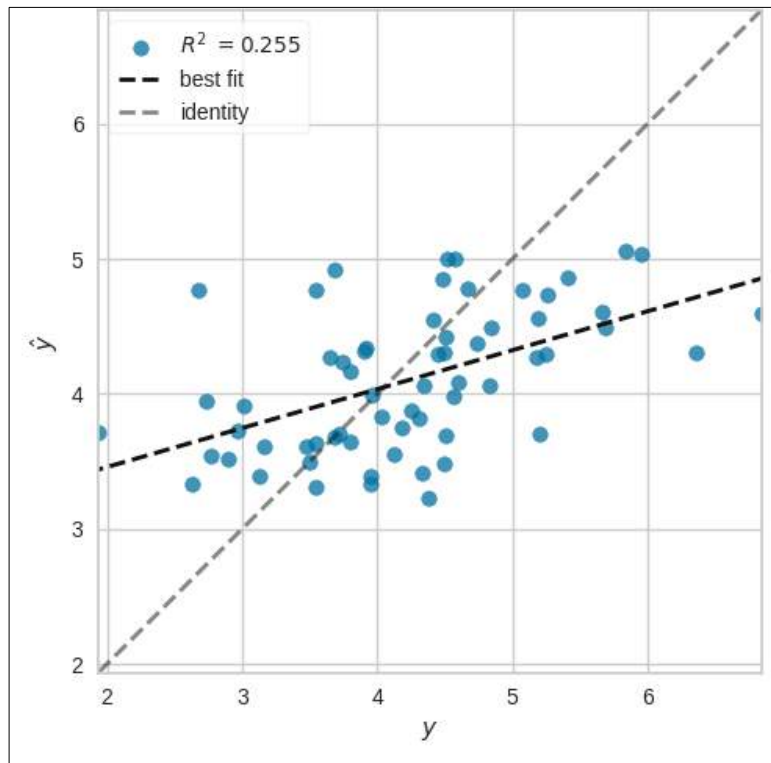
### a) WHEAT AND CALCIUM NUTRIENT COMPOSITION OF GRAINS



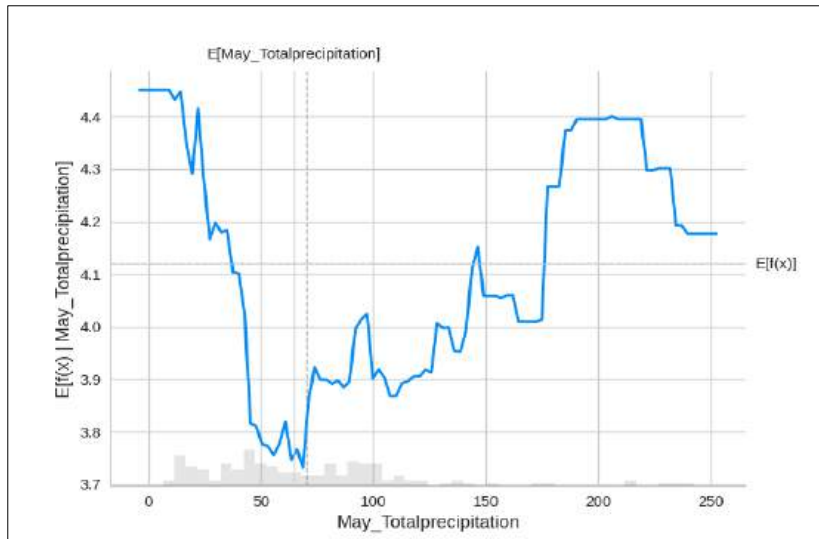
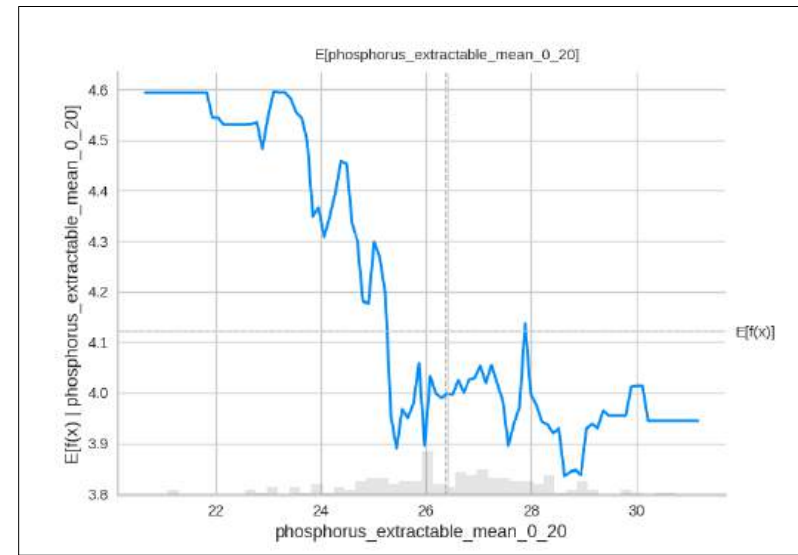
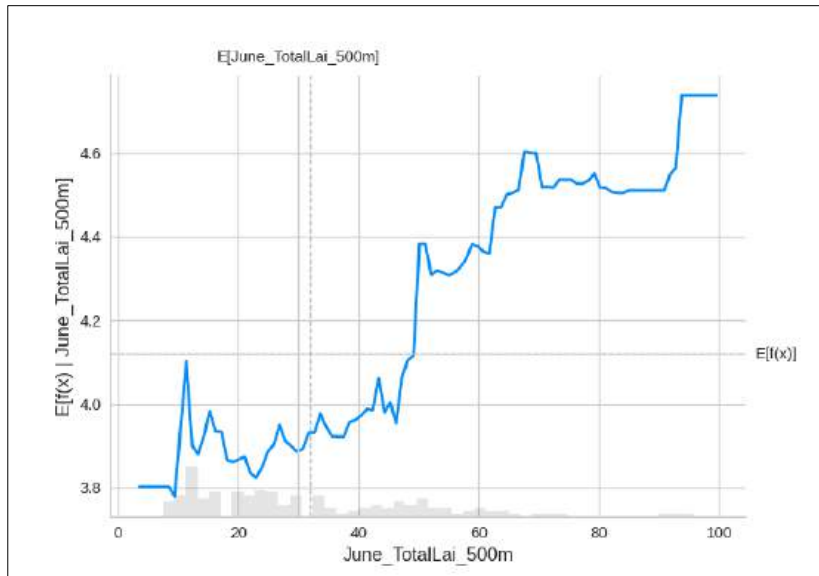




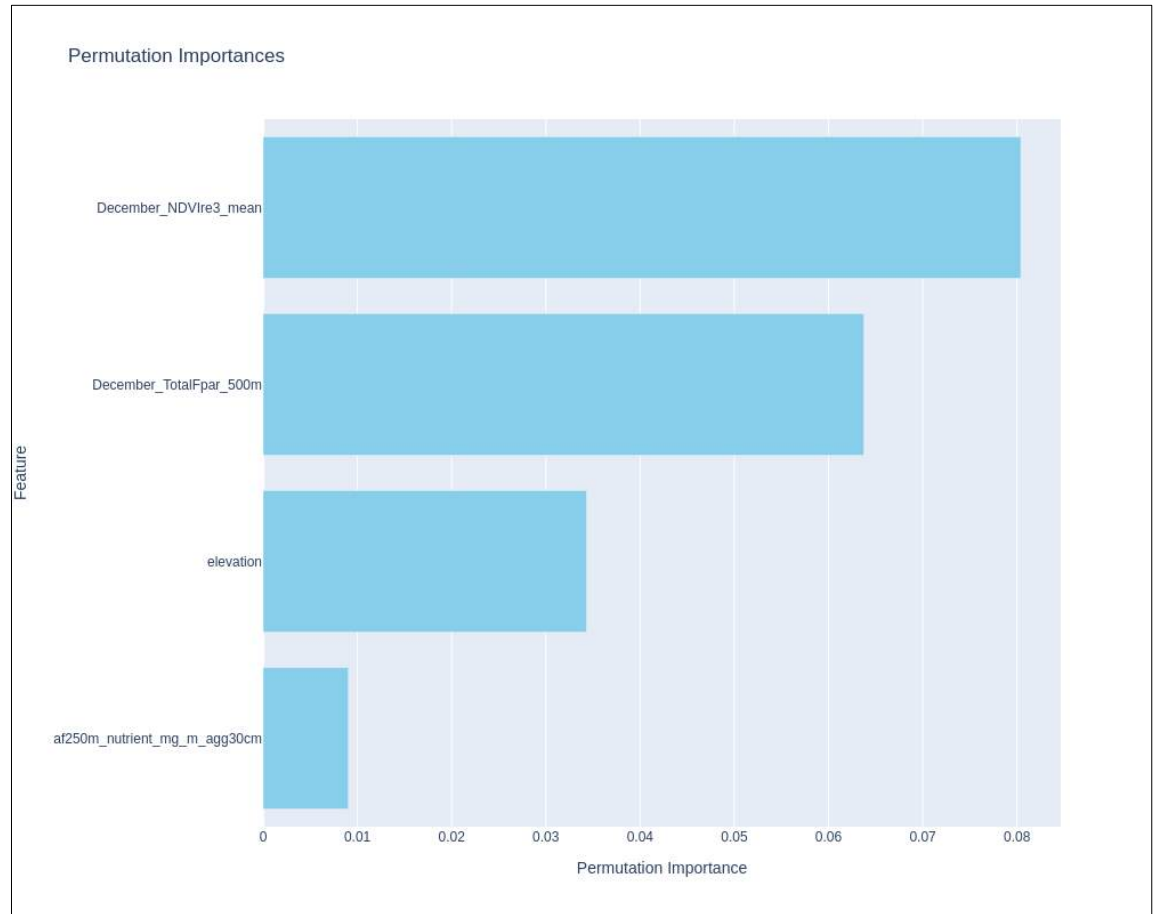
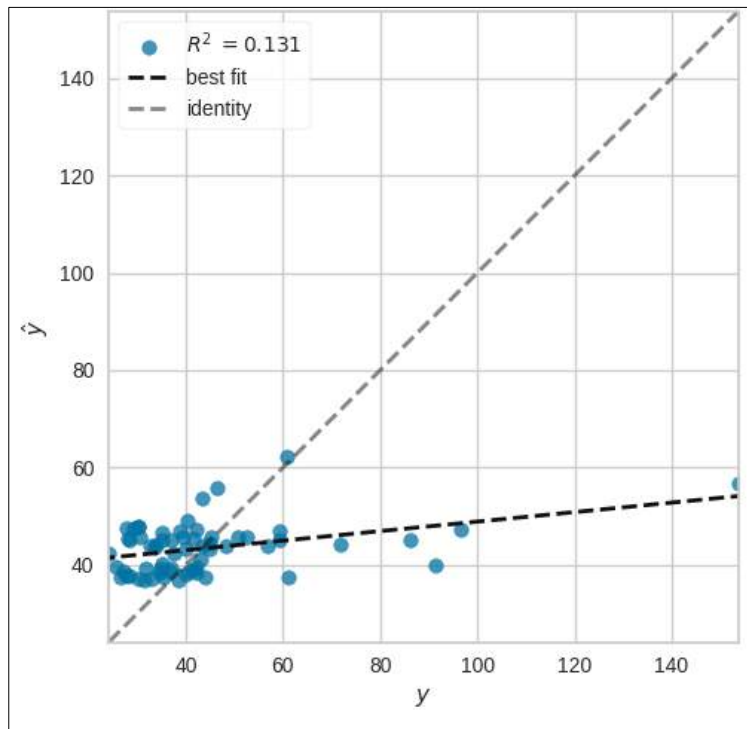
b) WHEAT AND COPPER NUTRIENT COMPOSITION OF GRAINS

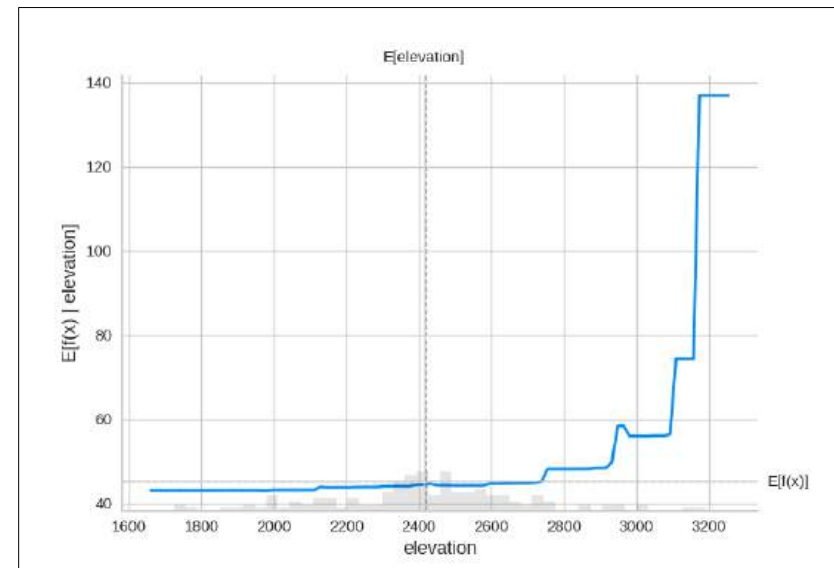
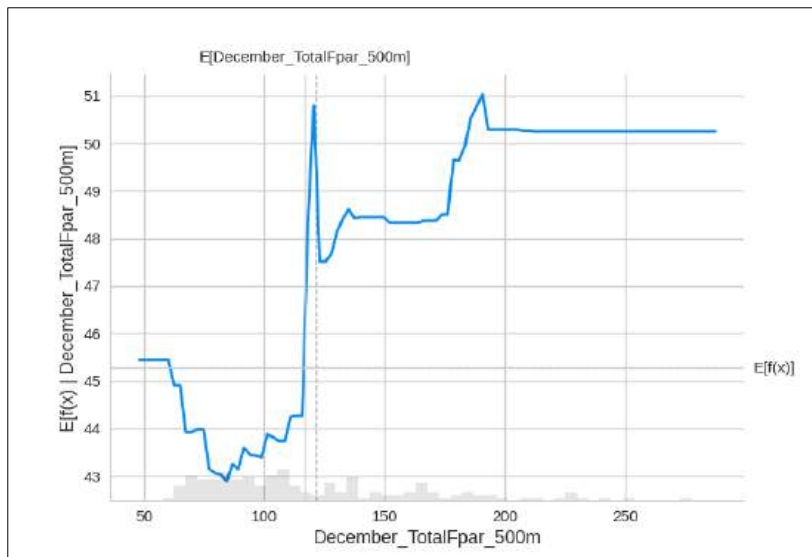
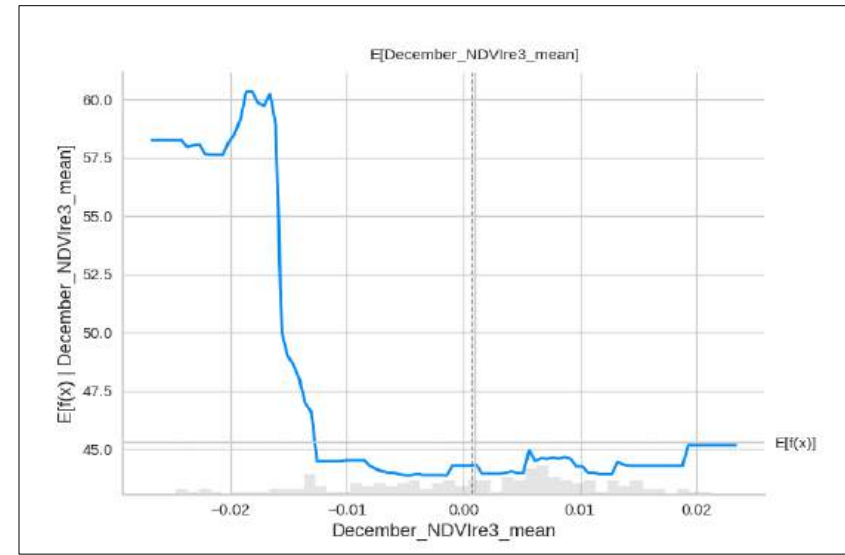
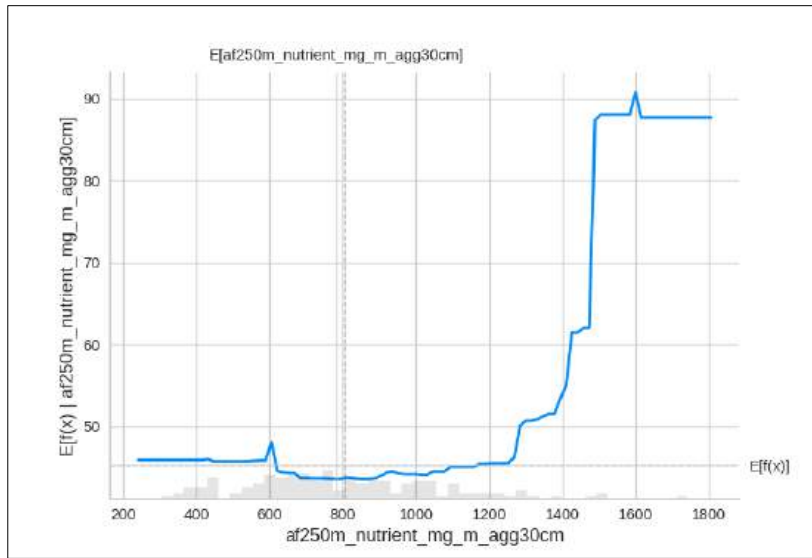




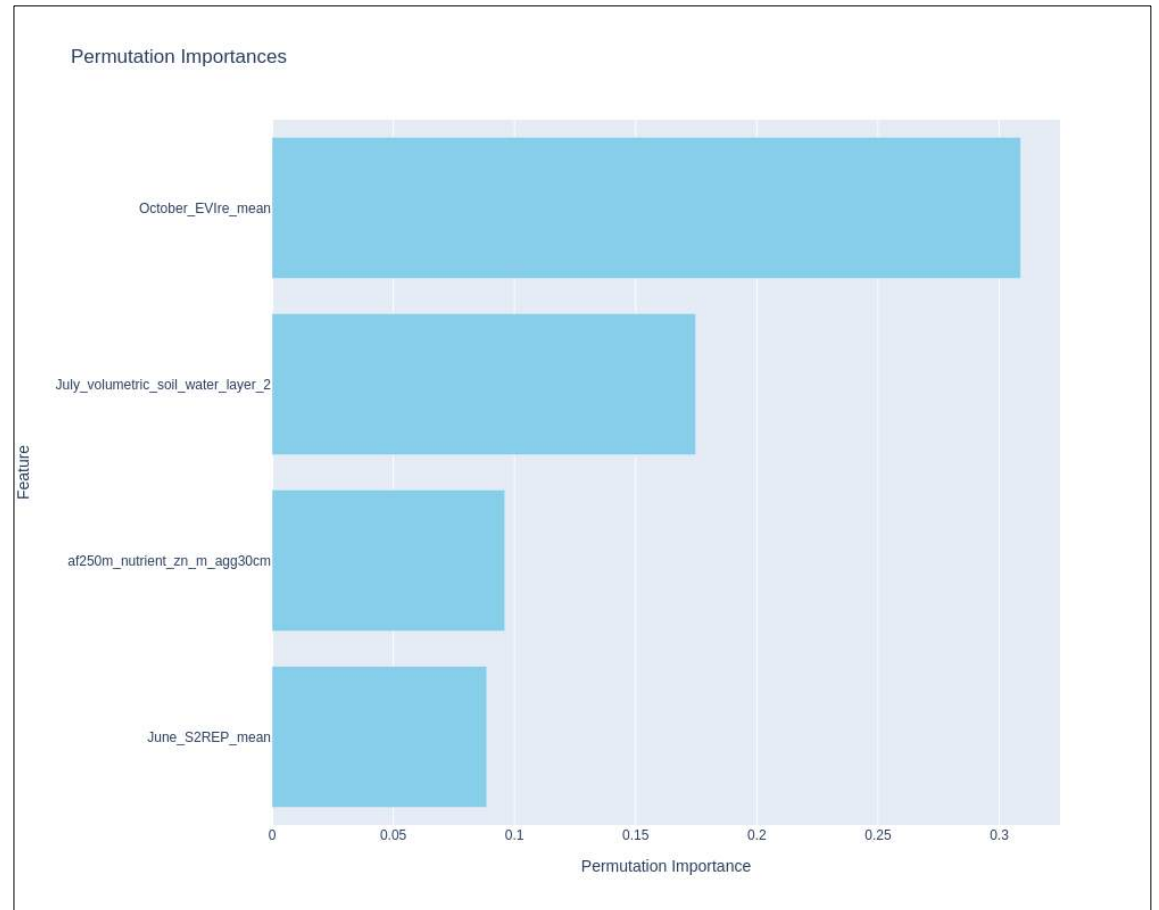
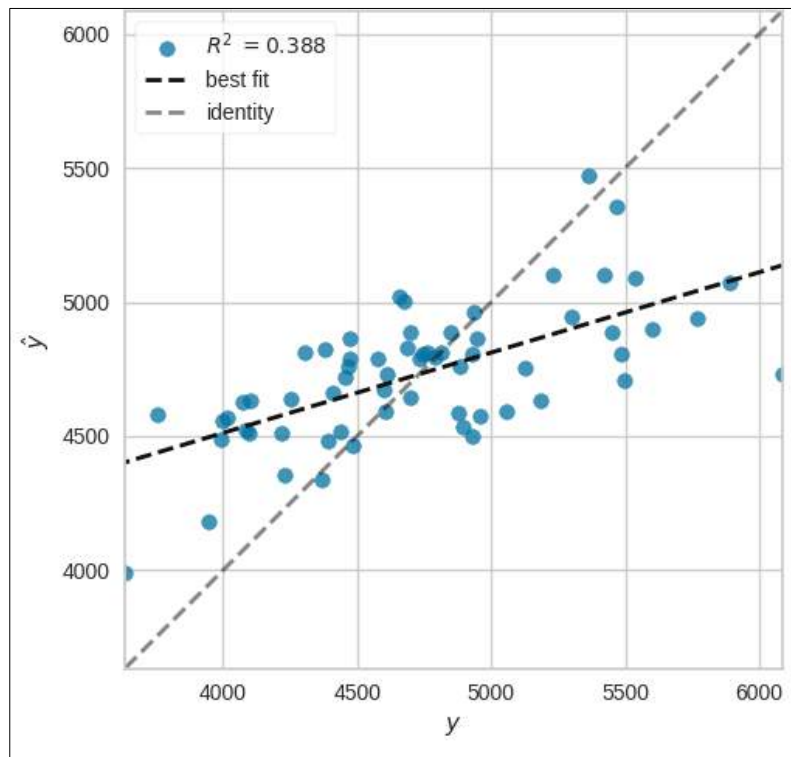


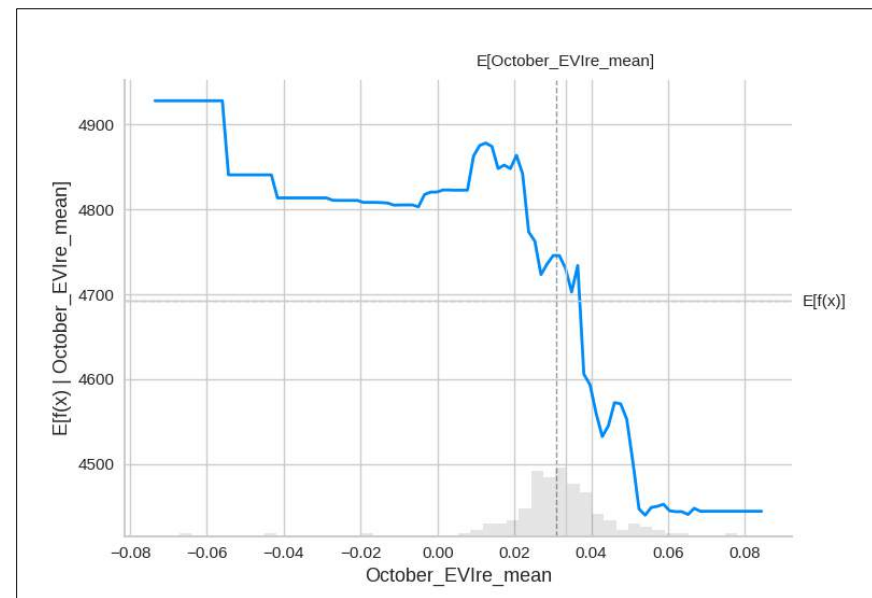
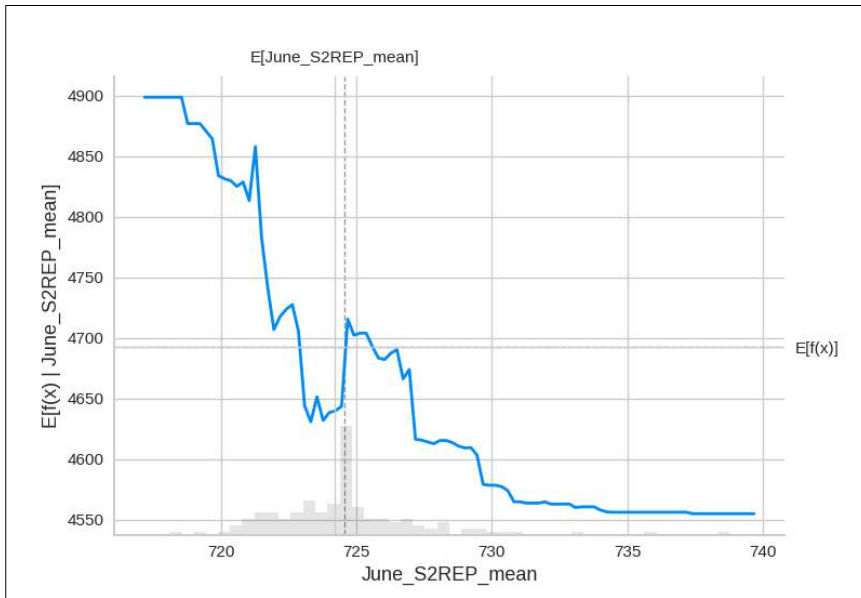
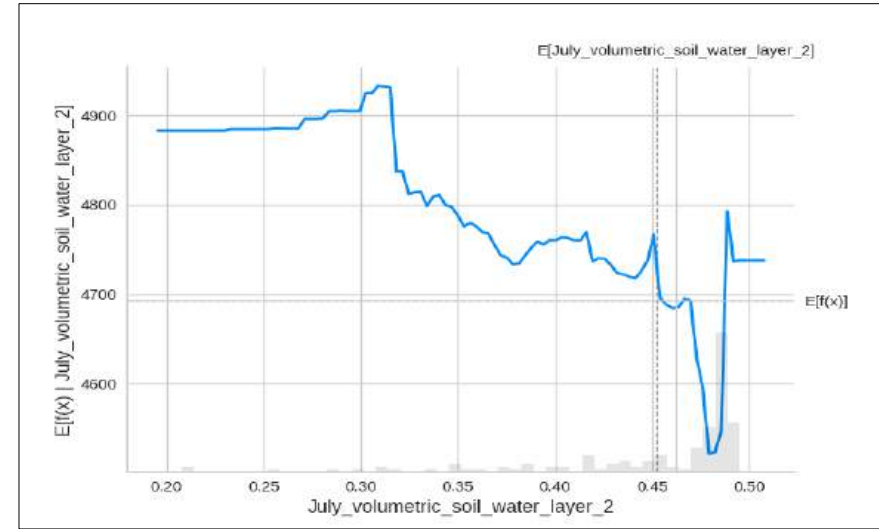
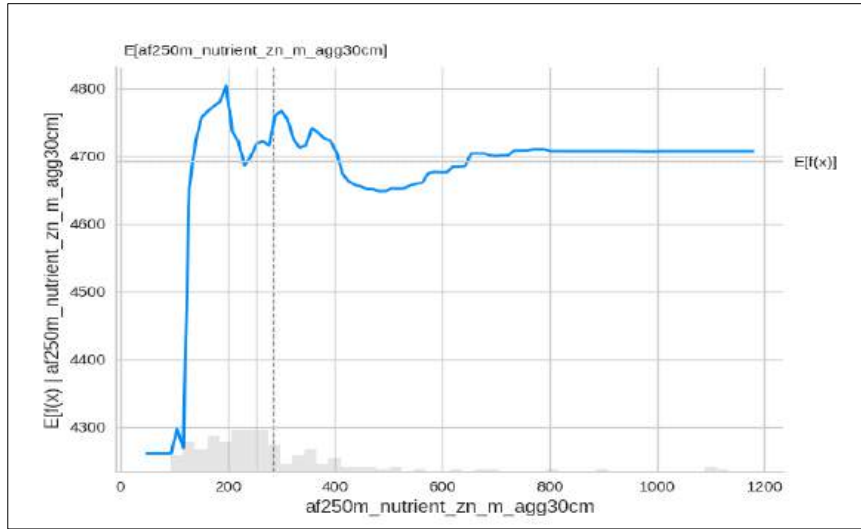
c) WHEAT AND IRON NUTRIENT COMPOSITION OF GRAINS



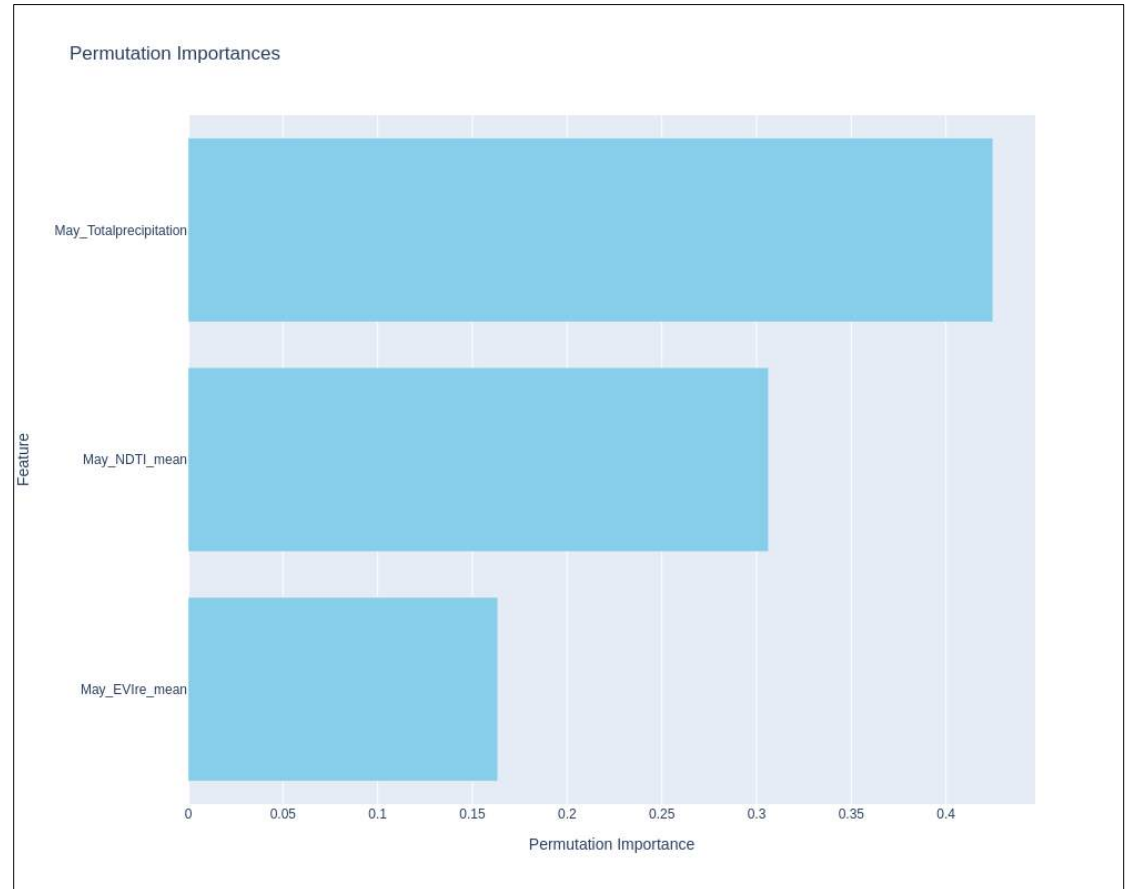
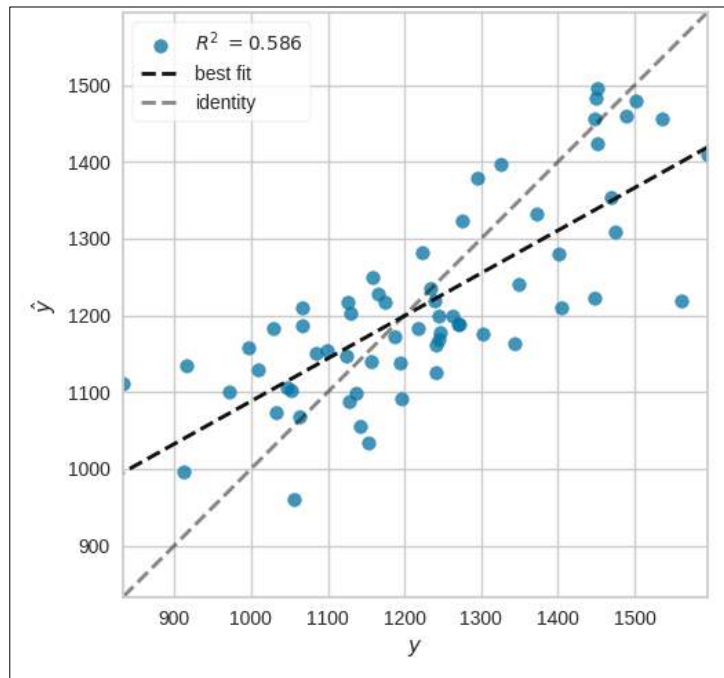


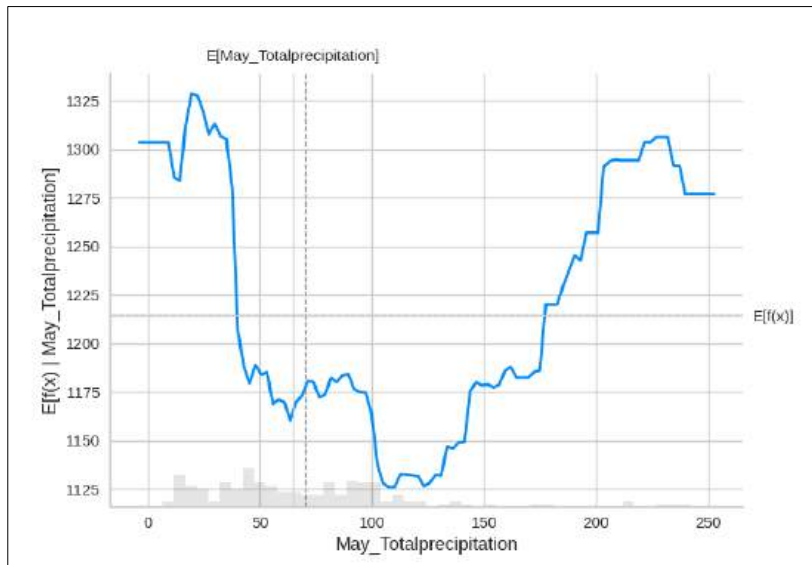
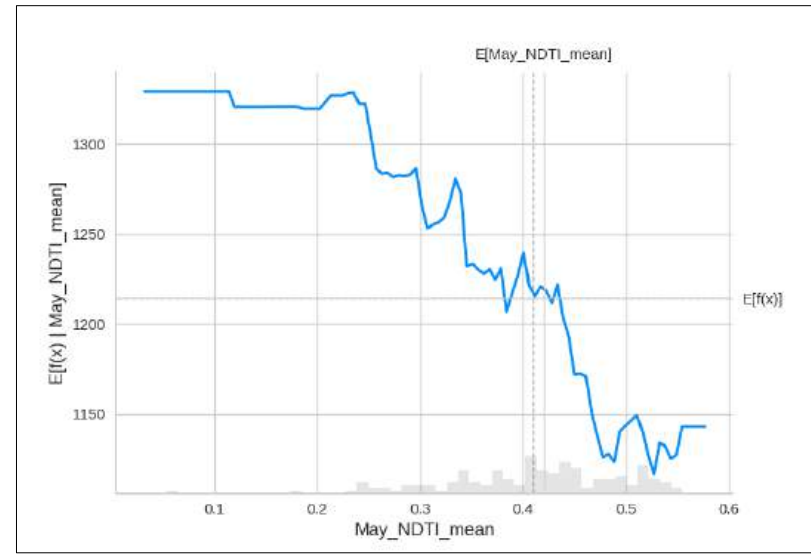
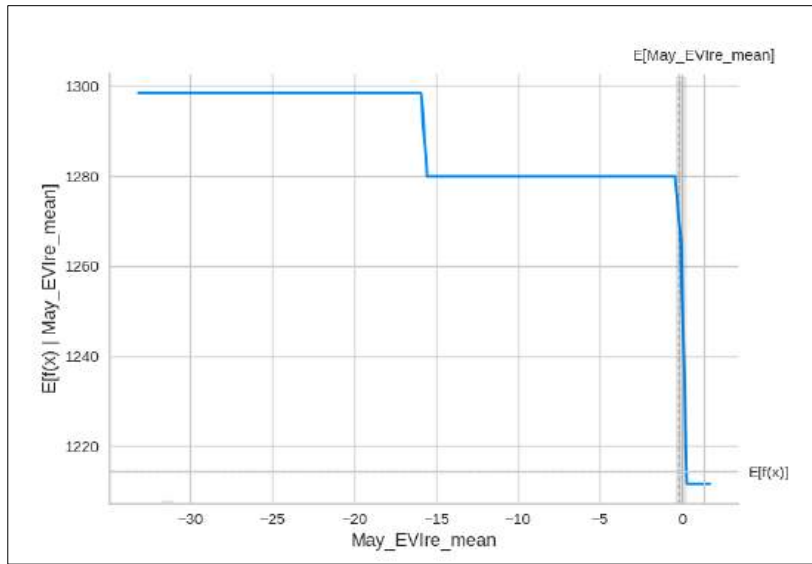
d) WHEAT AND POTASSIUM NUTRIENT COMPOSITION OF GRAINS



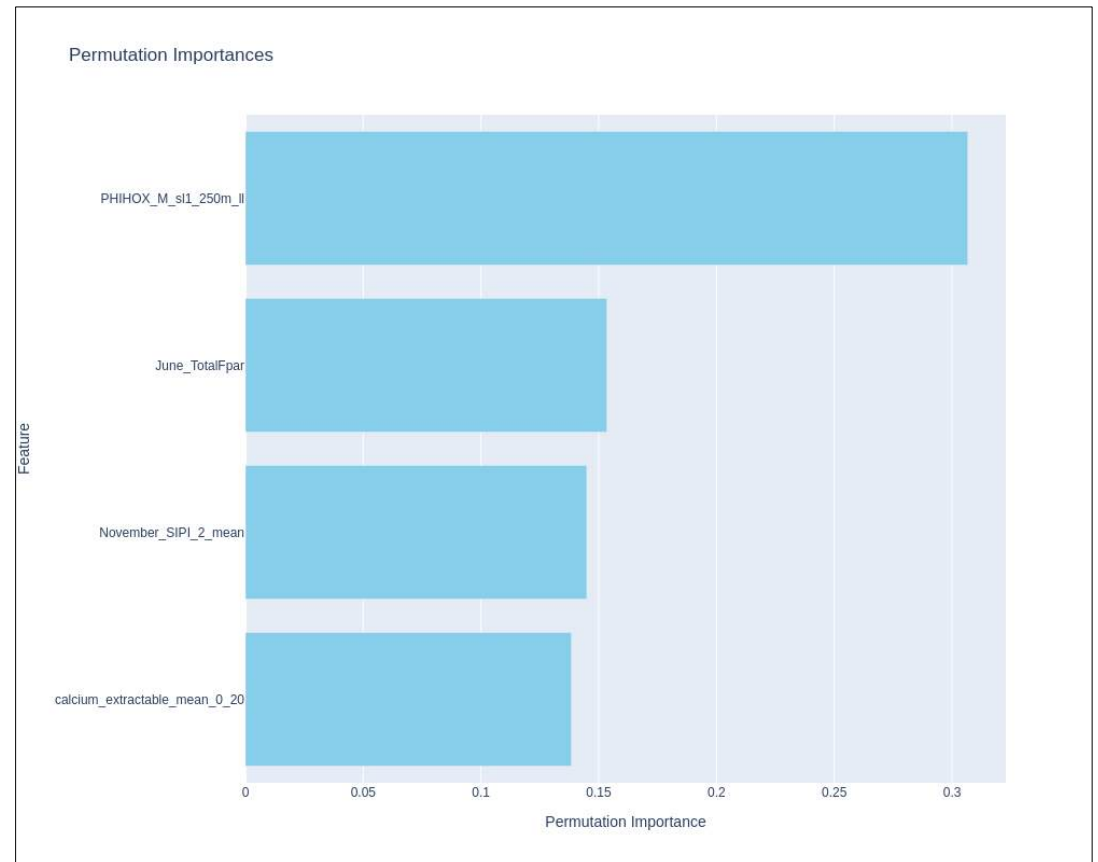
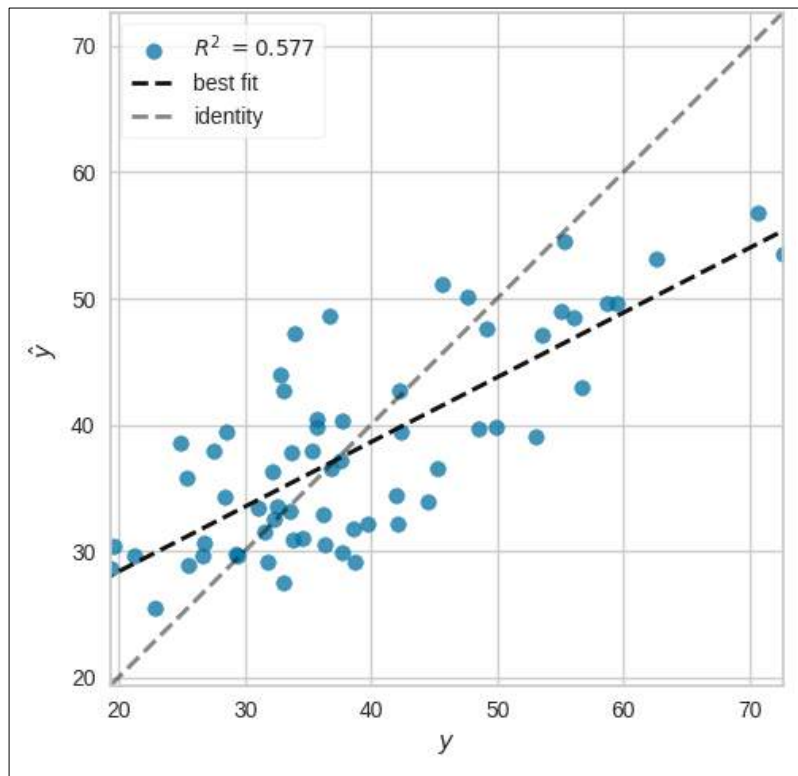


e) WHEAT AND MAGNESIUM NUTRIENT COMPOSITION OF GRAINS

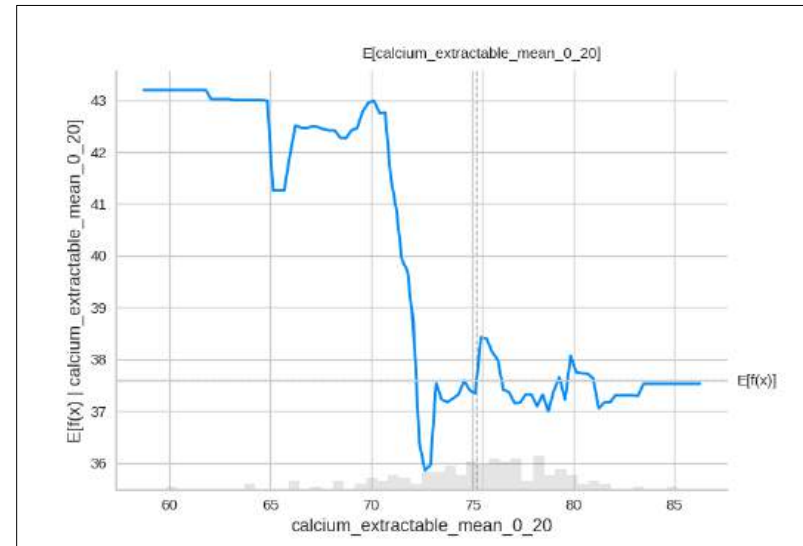
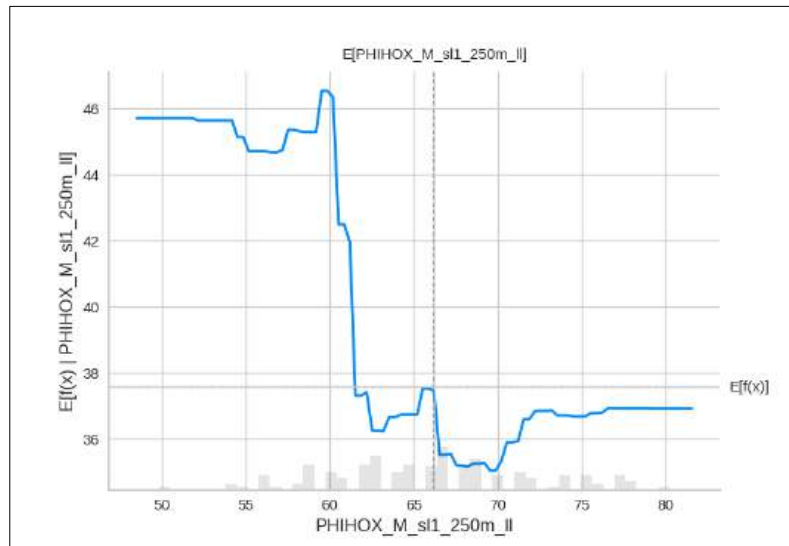
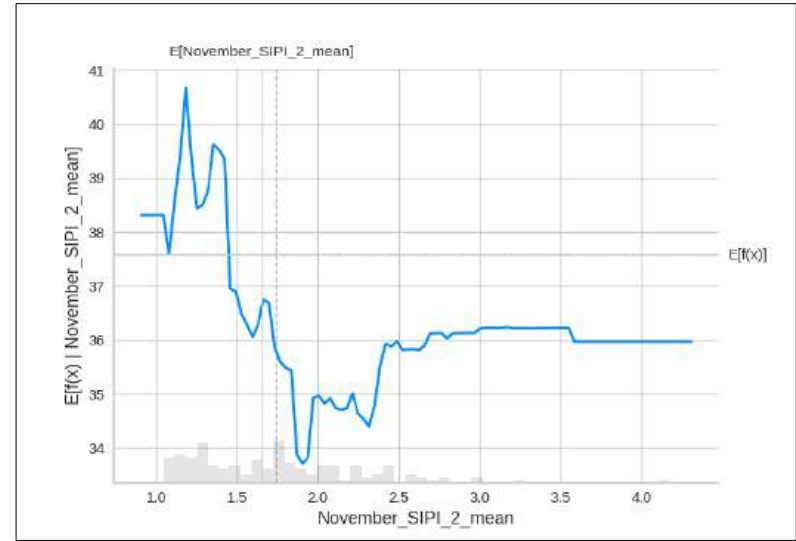
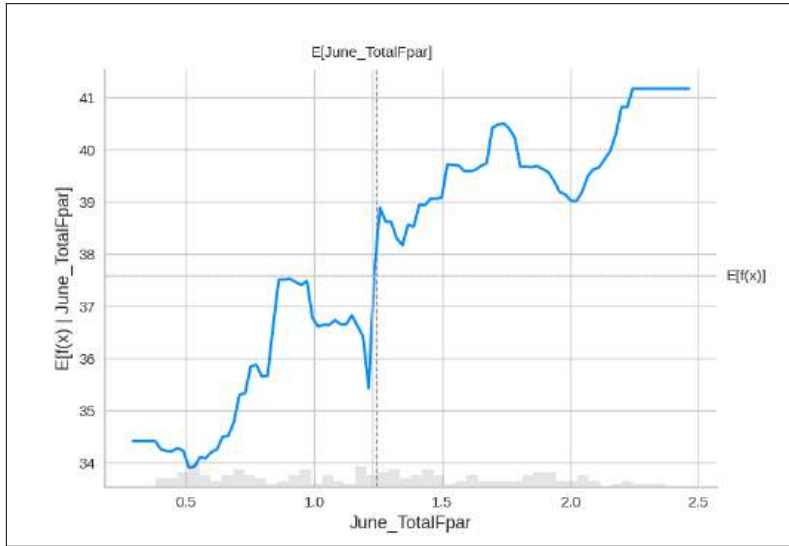




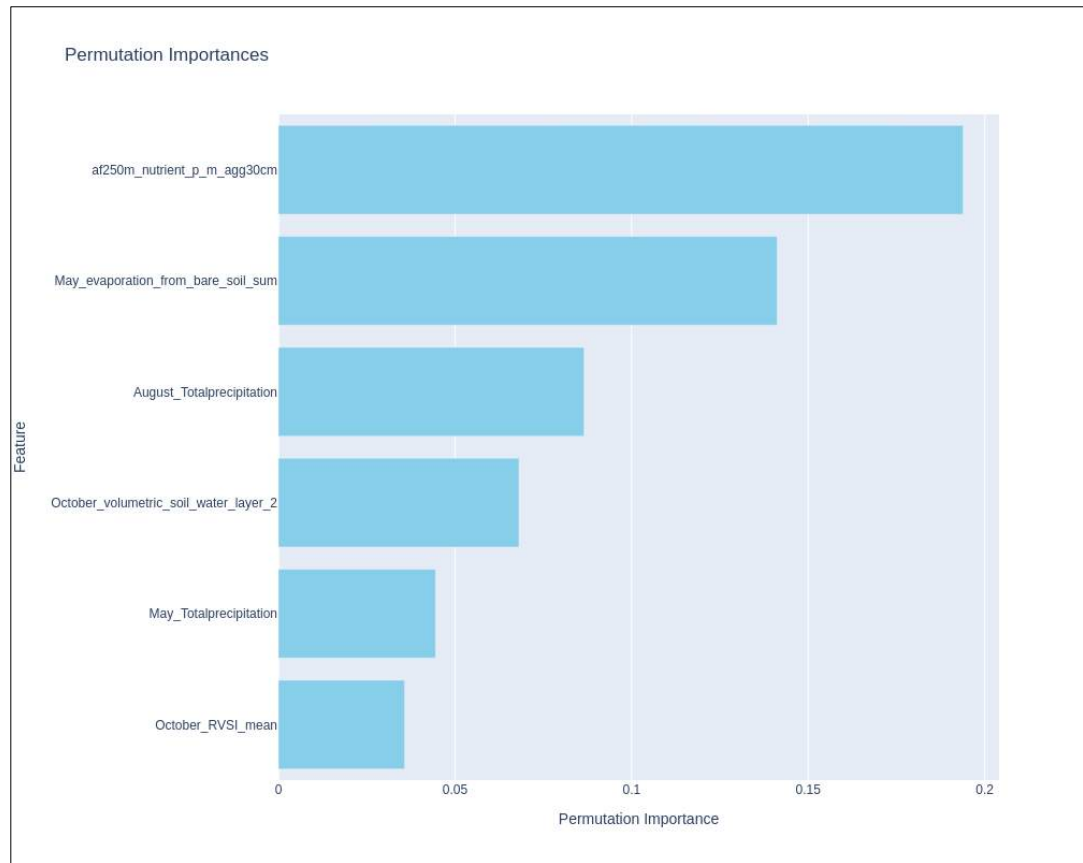
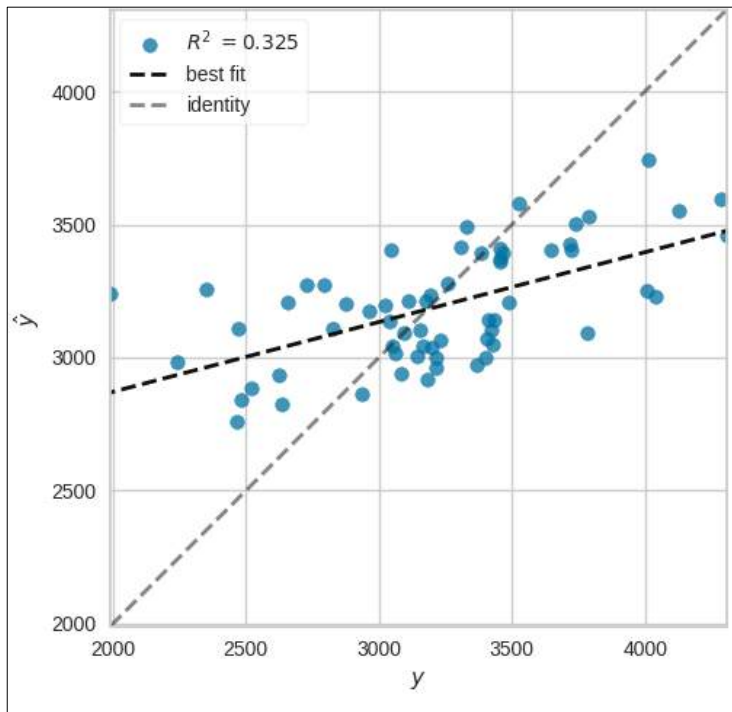
f) WHEAT AND MANGANESE NUTRIENT COMPOSITION OF GRAINS

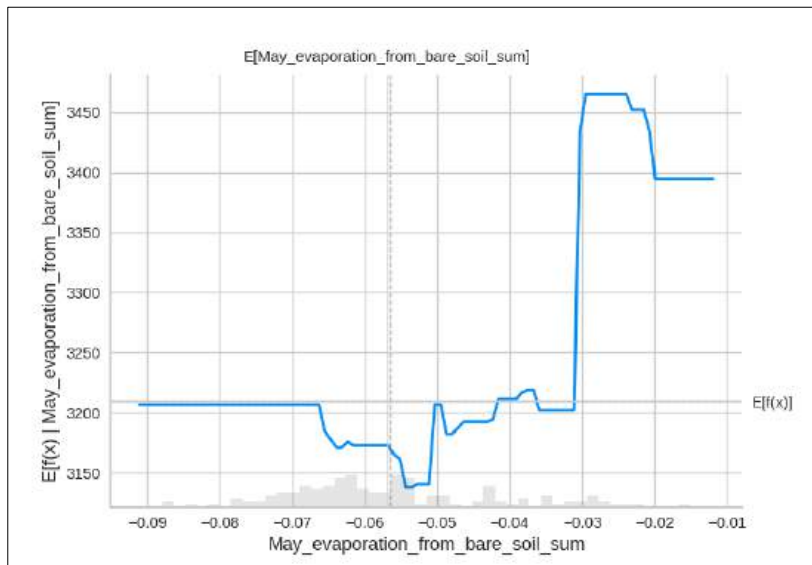
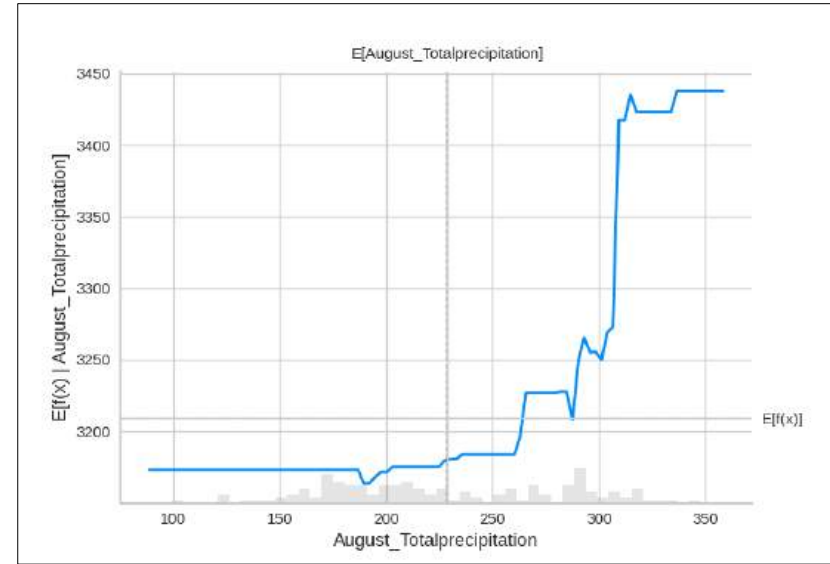
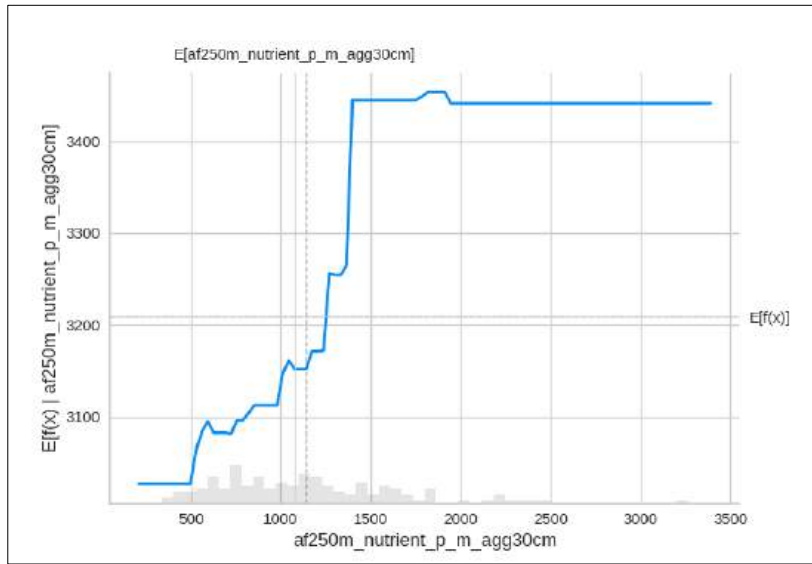


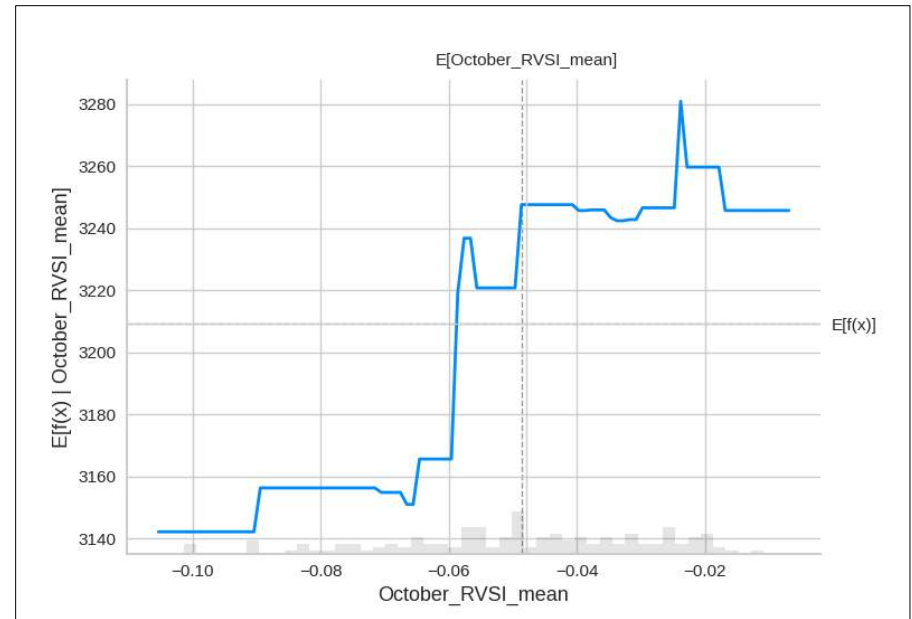
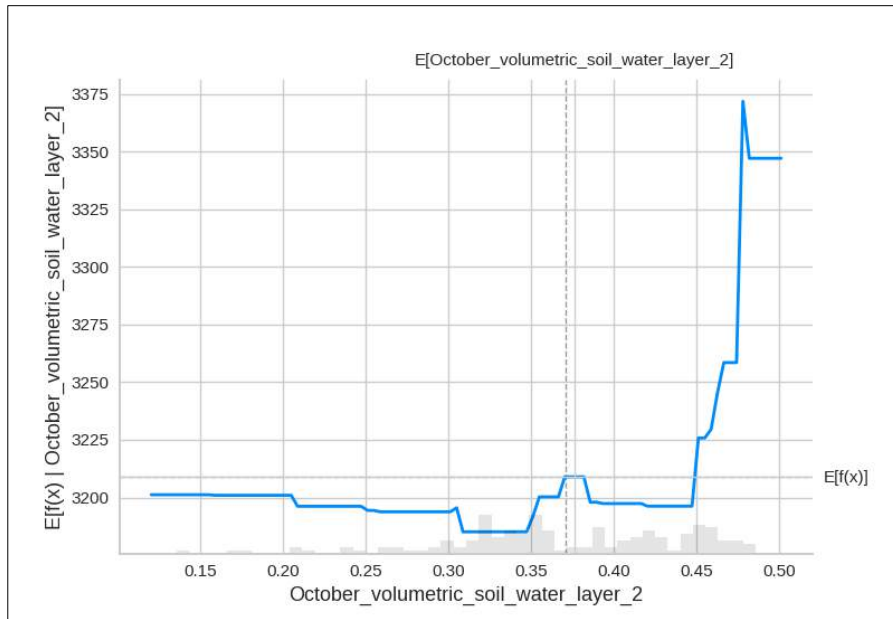




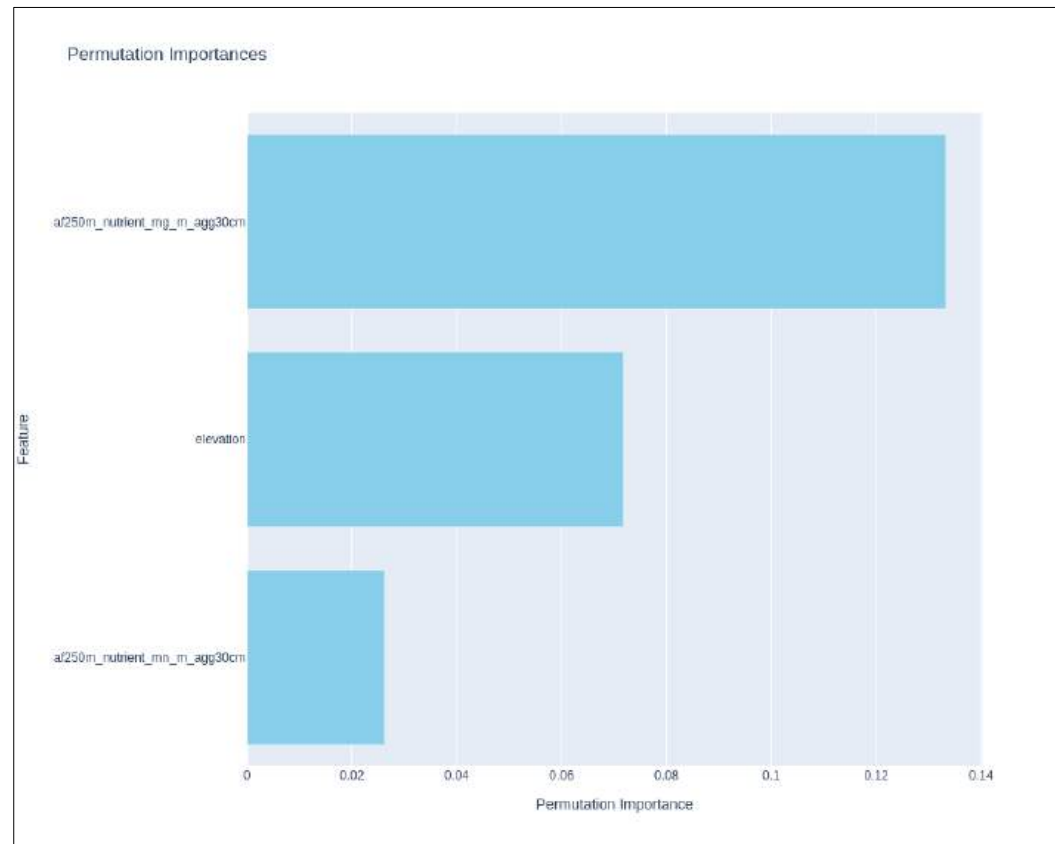
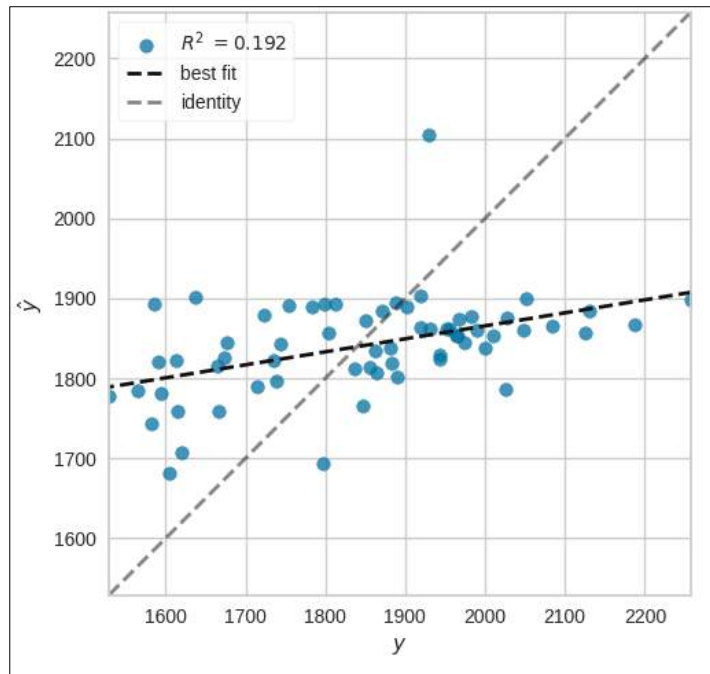
g) WHEAT AND PHOSPHORUS NUTRIENT COMPOSITION OF GRAINS

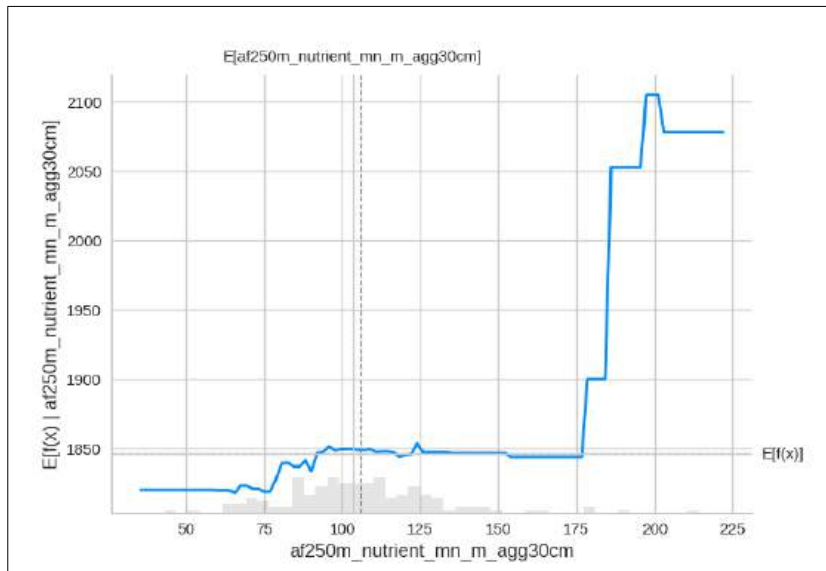
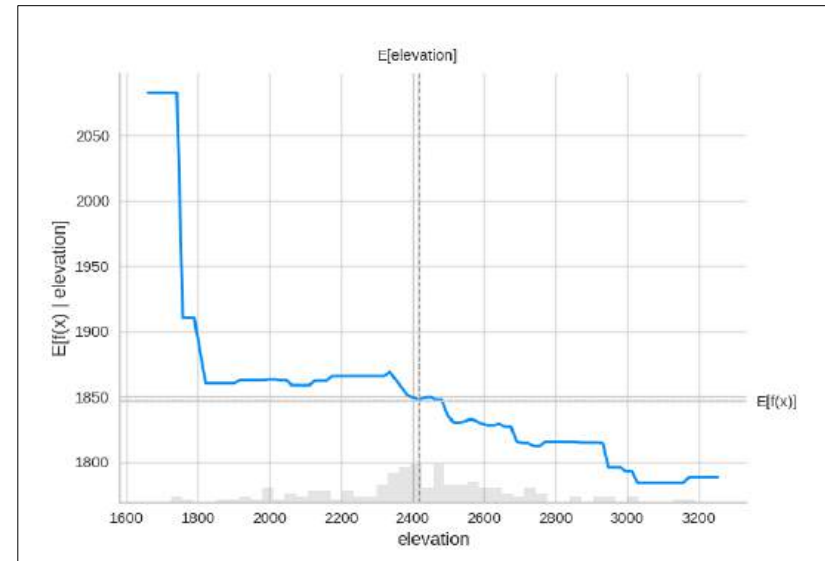
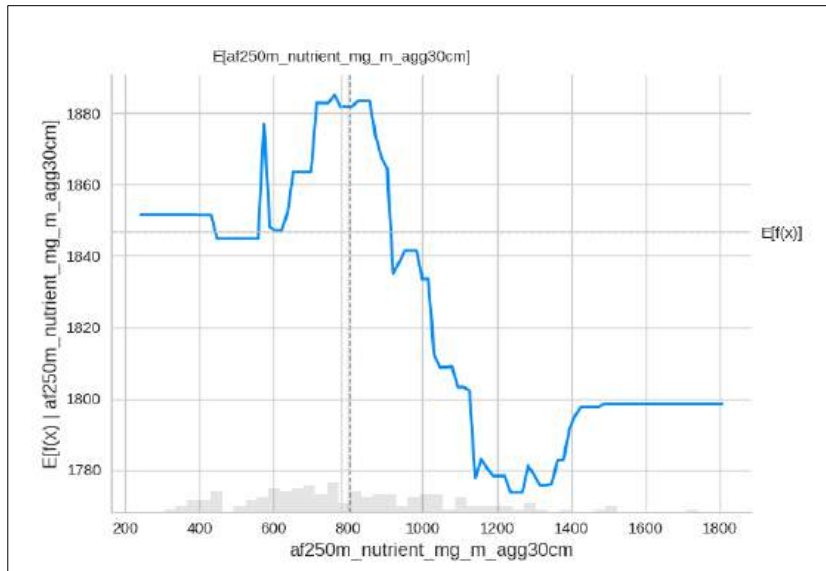






### h) WHEAT AND SULPHUR NUTRIENT COMPOSITION OF GRAINS





i) WHEAT AND ZINC NUTRIENT COMPOSITION OF GRAINS

