Studentship Project: Annual Progress Report
October 2021 to September 2023

Student Name:	Anisa Aubin	**AHDB Project Number:**	21130077
Project Title:	An empirical, data-driven, model for wheat cultivars and optimisation for future climate scenarios		
Lead Partner:	University of Sussex		
Supervisor:	Seb Oliver		
Start Date:	01 October 2020	**End Date:**	30 April 2025

1. Project aims and objectives

The goal of this project is to match data from the AHDB Recommended Lists (RL) UK wheat trials with temperature and precipitation data for trial site locations, to better understand how weather patterns explain variation in wheat yield. We assume that management practices across the trial sites are consistent. From this, we can better understand how wheat yield might be expected to change under predicted future climate scenarios and how this will vary across the UK. A similar study was conducted looking at maize in the US and the intent was to adapt the methodology for studying wheat grown in the UK. Additional aims and objectives include understanding how variety selection can enhance adaptation of UK wheat production to climate change.

2. Key messages emerging from the project

It was not possible to replicate the success of the US maize modelling work with the UK AHDB RL wheat trial data. We believe we have some understanding of why that is.

Data modelling for wheat in the UK is quite different from maize in the US, and there is not a clear visible relationship between climate and yield using the existing methodology. We believe that this is for three main reasons:

- The intrinsic variability of the trial sites in the UK is higher than the sites used in the US (i.e. there are more secondary factors influencing yield)
- Climate variations in the US extend further from the optimum growing conditions of maize, than the same for UK climate variations and wheat, i.e. the correlation signal we could measure is weaker
- The nature of the data used, i.e. small trial sites have a larger noise q.v. commercial scale sites and data in the US study are aggregated at county scale

Additionally, adding new feature variables such as sunshine or humidity data to the model does not appear to lead to any improvements. Adding these additional parameters adds complexity to the model but the relationship is no better defined. Using the annual accumulation/average of temperature and precipitation is an oversimplification and masks the variability that may influence the growth of the crop.

3. Summary of results from the reporting years

The primary objective was to model yield as a function of climate variables. We focused on yield anomalies as our target variable. Yield anomilies were calculated as the difference between a specific year/county compared to the average across all years. This way, it was easier to see if the yield was particularly good for that year or just that county is always good in terms of yield.

The original Gaussian model that had been used for maize in USA did not produce robust results that had been anticipated. So, most of the work in this period was devoted to exploratory data analysis of the feature variables and the target variables. The aim was to understand the data better and gain insights as to why the methods were less successful.

The following graphs illustrate some of the investigations that were carried out. The yield anomalies were then compared with the temperature and rainfall data to see if the anomalies are driven by weather. Figure 1 below presents yield against annual accumulation of temperature (growing degree days) while Figure 6 shows the relationship between yield anomalies by county and year and rainfall.

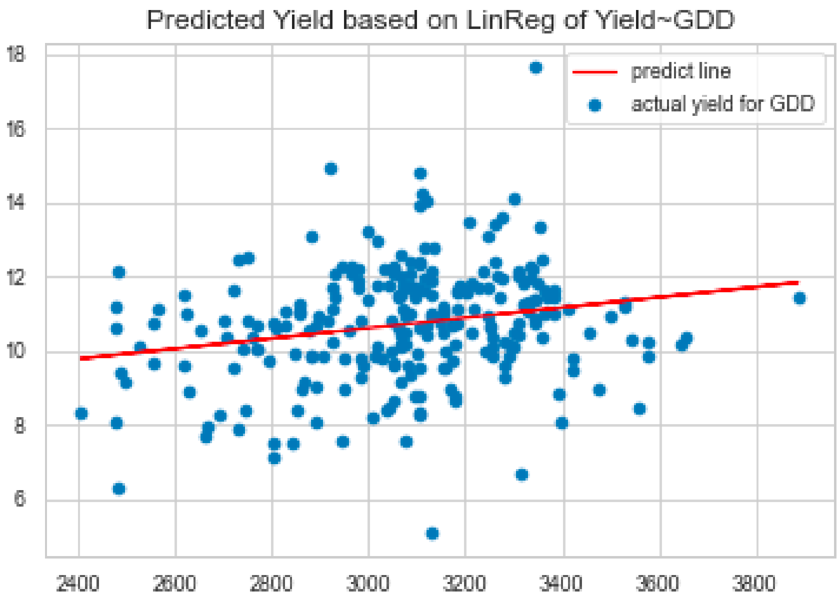

Figure 1. Using a linear regression to predict the yield with the annual accumulation of GDD units.

Since the Gaussian model had not worked, we explored a simpler model. Figure 1 illustrates a linear regression model fit to the growing degree day data. This shows a very weak trend with a large scatter. This large scatter helps to illustrate why it was challenging to fit a more sophisticated model.

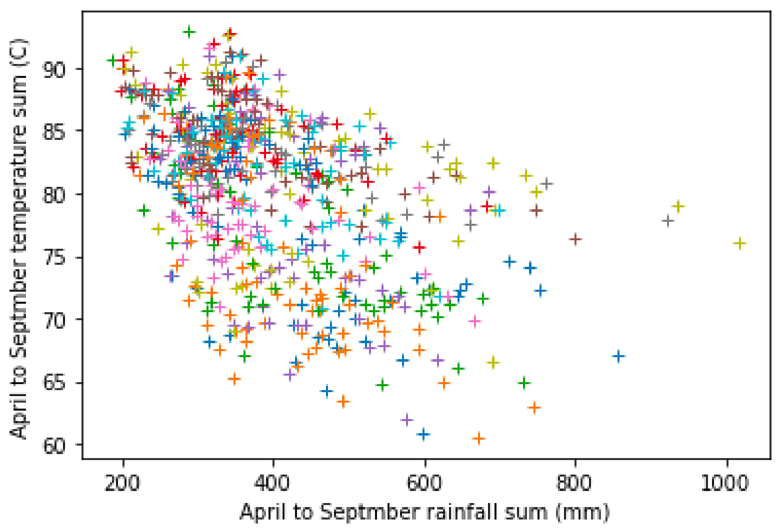

Figure 2. Correlation between summer temperature and rainfall.

Figure 2 illustrates the distribution of temperature and rainfall, colour-coded by different regions within which the Recommended List wheat trial sites are located.

Figure 3. Yield anomolies in each county as a function of date.

Figure 3 gives another view of the yield-anomaly data, showing the yield anomalies in each county over time. This also shows no systematic effects that would be easily modelled.

The temperature values do not vary much from year to year when looking at the whole year profile from county to county, but the rainfall has much variability (which is why the total value was originally used).

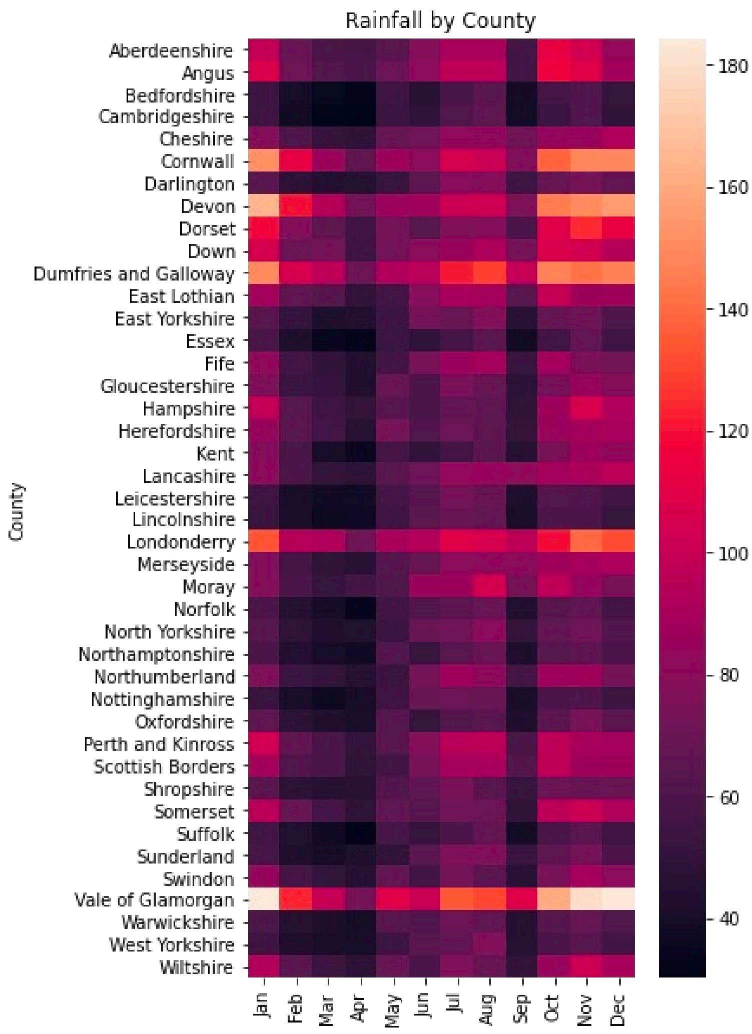

Figure 4. Monthly rainfall values by county

Figure 4 shows the monthly rainfall by county. This indicates a strong correlation which is explored further in Figure 5 which shows the variability of rainfall over one year and the strong correlation across different counties. This links with the points made in section 2 above – within a given year, there is little variation in (average) weather conditions across many of the trial locations in the UK, indicating that other factors are likely to be driving the variation in yield (Figure 3). These could include within-year variation in weather conditions (e.g. the importance of temperature or rainfall at specific stages of crop growth) or could be less directly related to weather, e.g. crop pest incidence, which would require more complex modelling.

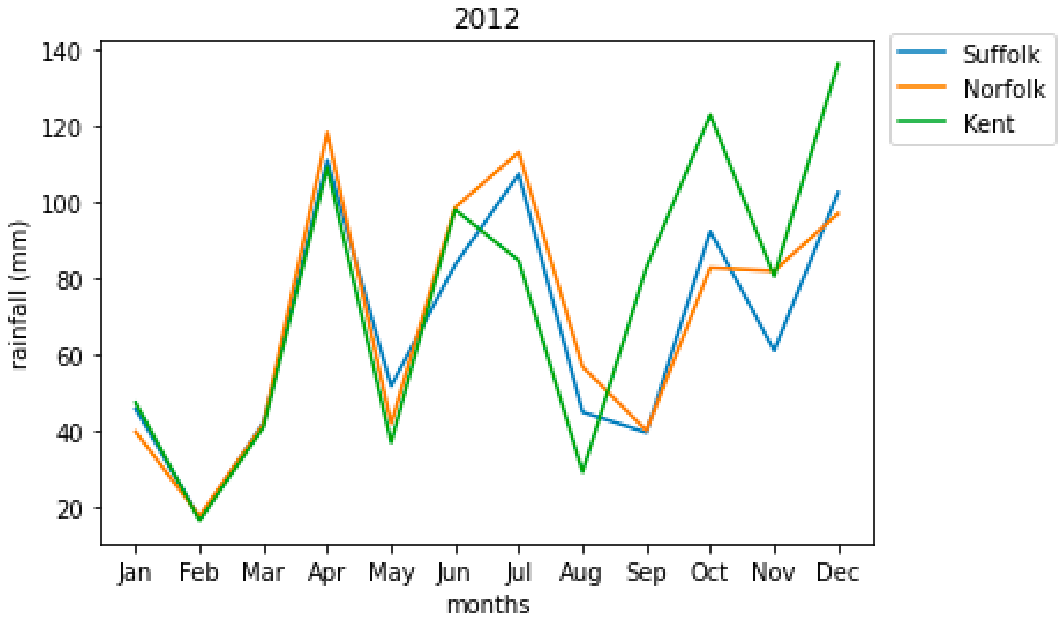

Figure 5. Rainfall for a single year by month for a few selcted counties.

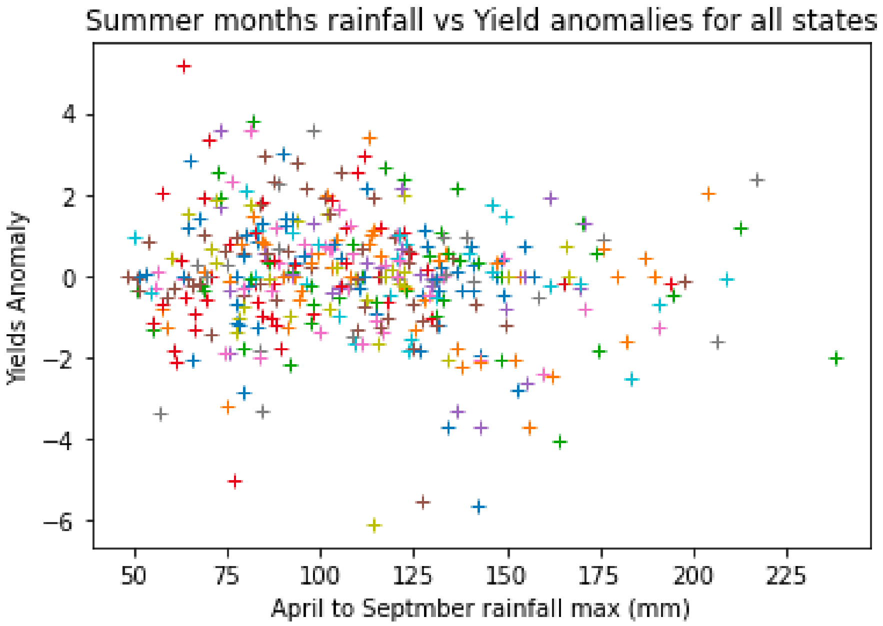

Figure 6. The raltionship or lack of it for the yield anomilies over the summer months rainfall.

4. Key issues to be addressed in the next year

The primary aim will be to identify a new investigation approach that can be successfully carried out with the available data to answer similar questions to those originally posed. We are exploring other datasets available from AHDB (e.g. Monitor Farm network) and other sources (e.g. breeder trial site information across larger spatial scales, and DEFRA wheat and oilseeds survey) that can be used either in addition or in replacement of the RL data.

5. Outputs relating to the project

(events, press articles, conference posters or presentations, scientific papers):

Output	Detail
National Confenerce	Peterborough – Dec 2021
Regional Conferences	South East (Maidstone), South West (Swindon) and East (Newmarket) 2022
MonoGram Poster Presentation	Reading, April 2023

6. Partners (if applicable)

Scientific partners	University of Reading
Industry partners	Quant Foundry and the Met Office
Government sponsor	