Cotton micronaire drivers are seed imbibition date, latitude and in-crop temperatures.

Weaver T[1*], Gardner B[2], Stevens L[2], Bange M[3], Teague C[3], and Gordon S[4]

[1]CSIRO, 21888 Kamilaroi Highway, Myall Vale, NSW 2390. tim.weaver@csiro.au
[2]CSIRO, 36 Gardiner Road, Clayton, Vic 3168.
[3]CSD, 'Shenstone'. 2952 Culgoora Road, Wee Waa, NSW 2388.
[4]CSIRO, 671 Sneydes Road, Werribee, Vic 3030.

Abstract
Environment and crop management can play an important role in determining upland cotton fibre quality. One of the important quality parameters is fibre micronaire, which is an indirect measure of fibre linear density (fineness) and maturity. Predicting micronaire in-season would allow growers to make informed tactical and strategic decisions in their management to maintain/improve fibre quality, for example, at harvest time (picking). Appropriately timing harvest aid applications is one such example. These predictions also have the potential to assist growers with marketing their cotton. A cotton yield prediction model, BARRY (Biometric Agronomy for Realising Representative Yield), was developed with the R programming language by Cotton Seed Distributors (CSD) and CSIRO using XGBoost (eXtreme Gradient Boosting) and crop data collected from CSD ambassador grower sites. BARRY was developed to equip growers with a tool to make informed in-crop decisions to maintain/optimise their yield (bales/ha). The same process was utilised to develop a predictive micronaire tool (i.e. an R Shiny web application, like BARRY) using the XGBoost package and achieved an r^2 of 0.842 and an RMSE of 0.18. The main factors that influenced the micronaire predictions were seed imbibition date, latitude, number of hot days and average temperature from seed imbibition to defoliation. These variables support previous findings of temperature impacting micronaire, and how latitude reflected local practices and cultivar choices.

Keywords
Cotton, Micronaire, Machine Learning, Fibre Quality.

Introduction
Micronaire (µg/inch) is an important fibre quality when it comes to spinning yarn for garments and the ideal range is between 3.8 to 4.5 (Bange et al. 2018). Australian upland cotton varieties are high yielding and produce exceptional quality, thus in high demand internationally by spinners. Cotton lint is sold into future contracts (prior to harvest) based on quality. Premiums and discounts are earned based on the quality of the ginned lint. This can be risky for growers, if they are unsure of their cotton quality (i.e. matching the desired micronaire for spinners). Previous research has been undertaken to predict micronaire from in-crop measurements, however, the analysis utilised mechanistic models and the dataset did not include crops from a diversity of climates, wide geographical range or differing management strategies (Bange et al. 2010; Bange et al. 2021). The research reported here utilised a dataset consisting of 273 variables collected by the CSD extension team over a period of 10 years (2015 to 2024). Data was collected from both irrigated and dryland cotton ambassador trials on sites ranging from the Northern Territory (NT) in Western Australia to Southern NSW. Only the irrigated model is reported in this research.

The aim of this research was to interrogate this large dataset that consists of a diversity in climate, management and genetics (including new Bollgard® 3 varieties) using machine learning techniques to identify key drivers for micronaire prediction. This knowledge will be used to deliver a tool to the cotton industry that will assist growers in making better in-crop management decisions to optimise their cotton fibre quality to meet market demand.

Methods

Cotton Seed Distributers Ambassador Dataset
The CSD ambassador dataset consists of 273 variables across all aspects of the cotton crop with over 2000 rows of captured data up to 2022 (now up to 2024). The data is collected throughout the season covering: general information, treatment information, pre-plant and planting operations, establishment methods, planting uniformity, mid-squaring snapshot, first flower snapshot, flowering progressions, cut-out snapshot, end of

season snapshot, quality data, irrigation information, management information and defoliation information. The dataset is captured by CSD extension agronomists through a bespoke in-house app using an iPad and collated by CSD, which includes varieties currently sown across the industry; i.e., Sicot 714 B3F, Sicot 746 B3F and Sicot 748 B3F.

Data interrogation

The previous published research by Bange et al. 2010 & 2021 indicated that one of the key variables from the dataset to be included in this study was the average temperature during boll fill (i.e. during fibre development; 430 day degrees after first flower to 870 day degrees using base 12; see equation 1 for Base 12 calculation. The first flower (white) occurs about 8-10 weeks after emergence, or at about 777 day degrees (Base 12) (Constable and Shaw 1988). The average temperature was calculated using The Long Paddock application programming interface (API) (https://www.longpaddock.qld.gov.au/) for the boll filling phase for each data point and added to the other 273 variables. The XGBoost analysis (Friedman et al. 2000; Friedman 2001) included all existing 273 variables plus additional climate data. The output from the analysis ranked the variables in importance. The next steps were to remove those with less than 5% importance and keep dropping variables out to develop the best relationship with micronaire. The best initial model was shown to contain 14 variables and achieved a relationship of r^2=0.7 with an RMSE of 0.137. The variables in decreasing importance were: latitude, defoliation hot days > 35°C (ie. days above 35ºC from seed imbibition to first defoliation date), day degree at cut-out (ie. day degrees from seed imbibition to cut-out), cut-out days after planting, average daily temperature during boll fill, first flower NAWF (nodes above first white flower), mid-squaring nodes to the first fruiting branch, first flower days after planting, first flower day degrees Base 12, cut-out total bolls, day degrees at cut-out, first flower first position retention, region and variety.

$$day\ degrees\ (Base\ 12) = \frac{(Tmax - 12) + (Tmin - 12)}{2}$$

Equation 1. Equation for calculating Base 12 day degrees in cotton. Day degrees (DD) are determined by subtracting 12ºC from both the daily maximum (Tmax) and daily minimum (Tmin). (Note: if the minimum is less than 12ºC Tmin is set to 12.) Note: new day degrees for Australian cotton production is calculated on a base temperature of 15.6ºC and an optimum of 32ºC thus called 1532.

Next steps in Modelling

Following the initial modelling, further investigations were made to reduce the number of variables and change the range of climate data. Seed imbibition date was added to the model and average temperature (average daily temperature was calculated then an average calculated over time) was calculated from seed imbibition date to the first defoliation date, previously was calculated only during boll fill. Utilising a high-performance computing cluster, a recursive feature elimination strategy of running multiple scenarios and dropping the weakest features each time was used for feature selection. Final model performance was measured using leave-one-out cross-validation. The four variables that were shown to have the best relationship with micronaire were: average temperature (from seed imbibition date to the first defoliation), latitude, the number of days above 35ºC (from seed imbibition date to the first defoliation date) and seed imbibition date.

Results

The final model built using XGBoost in R used four variables from the initial 273 from the CSD Ambassador dataset. The average temperature from seed imbibition to defoliation contributed to 52% of the variance in micronaire followed by Latitude (24%), then the number of days above 35ºC from seed imbibition to defoliation (13%) and finally the day of year when seed imbibition occurred (11%). The model was shown to have a relationship with micronaire with an r2=0.842 and RMSE = 0.18 (see figure 1).

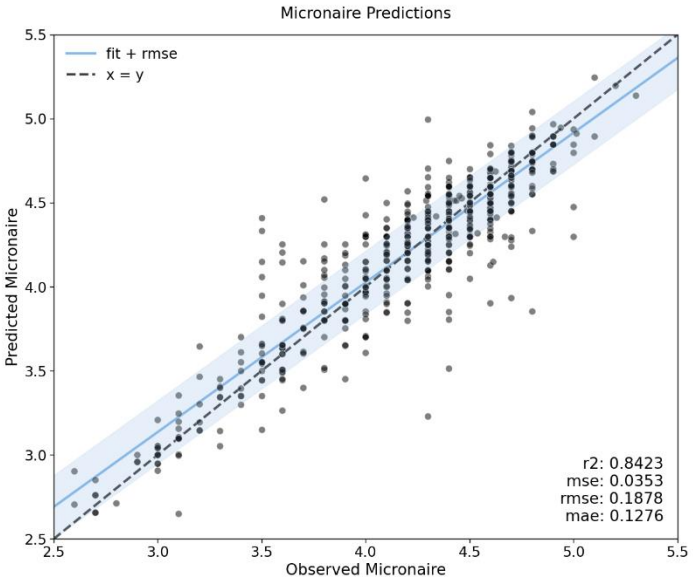

Figure 1. Observed versus predicted micronaire (µg/inch) generated from the XGBoost machine learning model using Latitude, Seed imbibition date, average temperature from seed imbibition to defoliation and number of hot days above 35ºC from seed imbibition to the first defoliation application. (shading indicates RMSE)

R Shiny app development

After the final model was built, an R Shiny app (Figure 2) was developed to empower growers with an interactive user interface to input their in-crop assessments. Growers can add their seed imbibition date, average daily temperature, number of hot days above 35ºC and latitude to estimate their micronaire.

The Shiny app also has a summary table with previous season values for all regions allowing growers to generate data for inputs that are not immediately known in the current season (e.g. number of hot days above 35ºC).

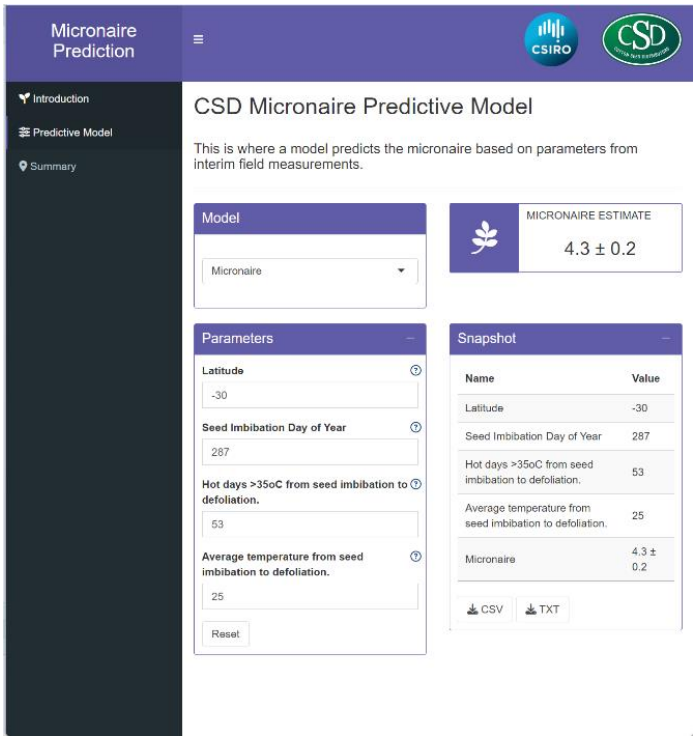

Figure 2. The R Studio Shiny app user interface for growers and consultants to predict micronaire in-season.

Conclusion

The interrogation of the CSD Ambassador dataset using XGBoost in R to determine key drivers of micronaire (i.e. climate, location and sowing date) were shown to support previous published research by Bange et al. 2021. The key drivers for micronaire found in the ambassador dataset were seed imbibition date, latitude, the

average daily temperature (from seed imbibition to defoliation) and the number of days above 35ºC (from seed imbibition to defoliation). An R Shiny tool (Mic) was developed using the XGBoost model to allow growers to input in-crop variables to predict their micronaire. This allows growers to make in-crop decisions, like harvest aid application timing, to optimise micronaire. It also allows growers to be better informed when selling into futures markets targeting specific cotton quality like micronaire.

Acknowledgements

This research was funded by Cotton Seed Distributers. A large thank you is given to the CSD team of extension agronomists for the collection of the ambassador dataset.

References

Bange MP, Long RL, Caton SJ and Finger N (2021). Prediction of upland cotton micronaire accounting for the effects of environment and crop demand from fruit growth. Crop Science 62, 397-409. (https://doi.org/10.1002/csc2.20679).

Bange MP, Constable GA, Johnston DA and Kelly D (2010). A method to estimate the effects of temperature on cotton micronaire. Journal of Cotton Science, 14, 164–172.

Bange MP, Constable GA, Gordon SG, Naylor, GRS and Van der Sluijs MHJ (2018). Importance of fibre quality. In MP Bange, GA Constable, SG Gordon, GRS Naylor and MHJ Van der Sluijs (Eds), FIBREpak: A guide to improving Australian cotton fibre quality (2nd ed., pp. 30–42) CSIRO and the Cotton Research and Development Corporation.

Constable, GA and Shaw AJ (1988). Temperature requirements for cotton (Agfact P5.3.5). Division of Plant Industries, New South Wales Department of Agriculture.

Friedman JH, Hastie T and Tibshirani R (2000). Additive Logistic Regression: A Statistical View of Boosting (with Discussion and a Rejoinder by the Authors). The Annals of Statistics, 28 (2). Institute of Mathematical Statistics: 337–407.

Friedman JH (2001). Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics. JSTOR, 1189–1232.