Functional Data Analysis for Assessing the Economic Impact of Soilborne Diseases

Braden Thorne[1], Niloofar Vaghefi[2], Samuel Blanch[3], Lisa Kelly[4], Adam Sparks[1], Christopher Little[5], Natalie Moore[3]

[1]Curtin University | [2]University of Melbourne | [3]NSW Department of Primary Industries and Regional Development | [4]QLD Department of Primary Industries | [5]Kansas State University

Centre for Crop and Disease Management

Curtin Biometry and Agricultural Data Analytics

Curtin University

ABSTRACT

The timing and magnitude of weather events is a critical factor in determining the incidence of a soilborne disease outbreak. As such the use of publicly available weather data to quantify the probability of disease epidemics can play a critical role in determining the potential economic impact of soilborne diseases in different conditions. We present a collection of methods being developed within AAGI that leverages functional data analysis (FDA) to connect a time series of weather data to the probability of a disease epidemic and its associated economic losses. This is part of an ongoing AAGI service and support project applying these methods to assess the economic impact of charcoal rot in soybean across the eastern and northern growing regions.

Application Context

- Soybean is a critical summer crop, with 429 million tonnes of grain produced globally in 2025/26[2].
- Soybean production in Australia is relatively small (52 thousand tonnes in 2025/26[2]) but with noticeable growth (5% 10-year compound average growth[2]).
- Charcoal rot, caused by *Macrophomina phaseolina*, is a significant disease of soybean with losses of up to 2.5 tonnes per hectare reported during conducive years.
- Charcoal rot is known to favour warm, drought-prone environments.
- NSW DPIRD has collected 16 years of trial data with corresponding charcoal rot ratings from a site in Grafton, NSW as part of the Australian Soybean Breeding Program.

Methods

FDA[3] was used to identify critical weather variables using methods detailed by Shah et al. (2019)[4] and codified in the R *fda* package[5]. Weather data was obtained from the Scientific Information for Land Owners database[6]. Smooth weather time series $x(t)$ were analysed using a function-on-scalar regression model defined as

$$x(t) = \alpha(t) + (-1)^{Epi}\beta(t) + \epsilon(t)$$

where $\alpha(t)$ is the mean weather series across all seasons and $\beta(t)$ is the deviance from this mean in non-epidemic ($Epi = 1$) and epidemic ($Epi = 2$) seasons. The significance of $\beta(t)$ was assessed using functional t-tests, with the time of maximum functional test statistic t^* recorded.

The effect of epidemic seasons and soybean variety on yield was assessed using linear mixed effect models implemented in the *ASReml-r* package[7], with the final model represented by the following *ASReml-r* syntax:

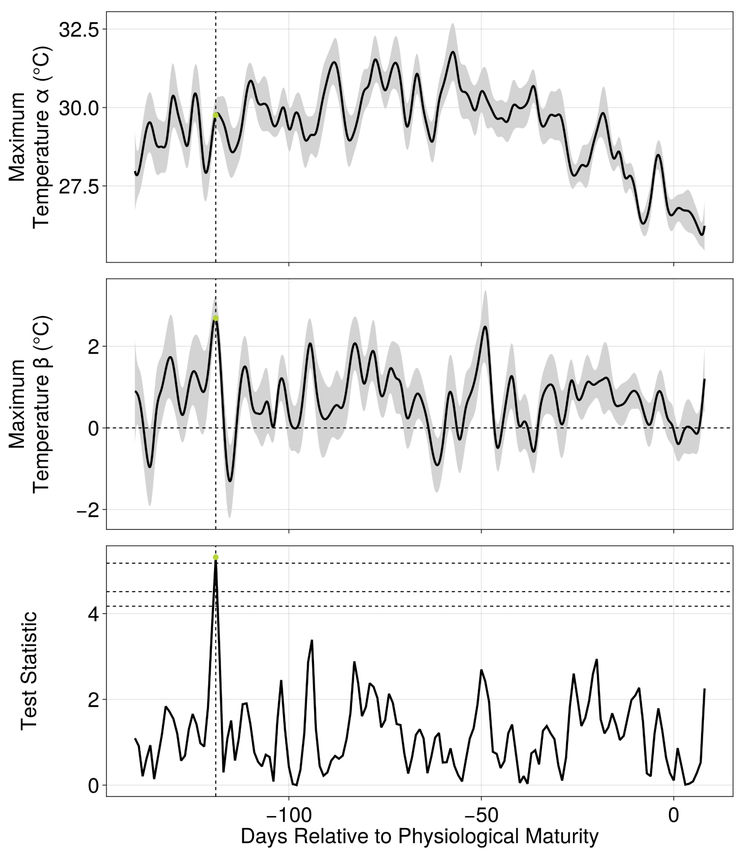
```
› Fixed = Yield ~ Variety*Epidemic
› Random = ~ Season
› Residual = ~ id(units)
```

Current work is exploring how to apply the FDA model to novel weather time series to establish disease epidemic risk. Thresholds for epidemic and non-epidemic seasons are calculated at the critical time t^* using the formula

$$\alpha(t^*) + (-1)^{Epi}\beta(t^*)\left(1 - 1.96 \cdot \frac{SD_{\hat{x}}(t^*)}{|\beta|}\right).$$

Analysis of Weather Data

Maximum temperature is significantly different during charcoal rot epidemics.

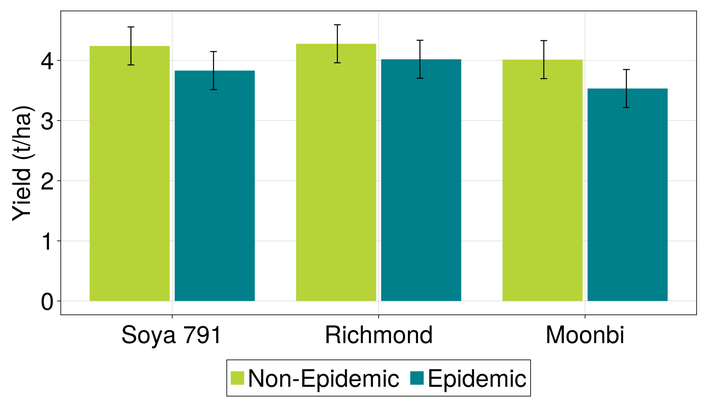


FDA of maximum daily temperature data for Grafton, NSW. The test statistic is maximised at 119 days to maturity, with a p-value of 0.008. Results adapted from Vaghefi et al. (2026)[1].

Economic Impact of Epidemics

Epidemics have a significant economic impact through reduced soybean yield.

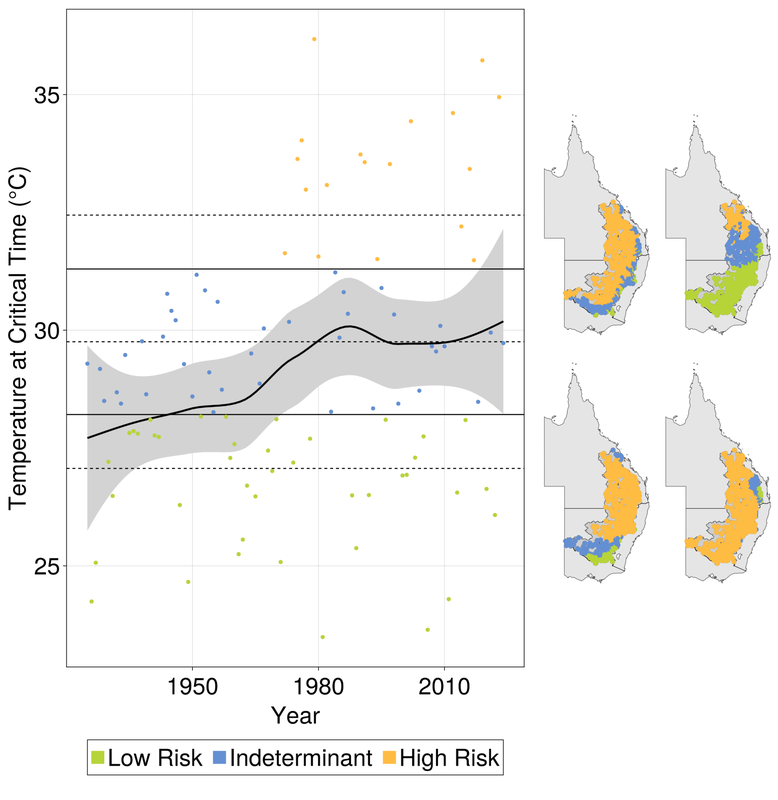
Source of variation	F-statistic	p-value
Yield		
Variety	13.37	<0.001
Epidemic	5.18	0.039
Variety:Epidemic	1.20	0.303
Seed Size		
Variety	89.28	<0.001
Epidemic	0.05	0.821
Variety:Epidemic	1.87	0.158



Top: Test results assessing the impact of variety and epidemic occurrence on yield and seed size.
Bottom: Predicted soybean yields across varieties and epidemic occurrence.

Epidemic Risk *(Ongoing)*

Next steps: how do we combine results to predict future economic impacts?



Left: Estimation of epidemic risk for Grafton NSW over the past 100 years using maximum daily temperature.
Right: Estimation of epidemic risk across eastern Australia for 2022-25 using maximum daily temperature.

Acknowledgements

This work is in support of GRDC investment UOM2401-002RTX Understanding the economic impact of *Macrophomina* (Charcoal rot)

Lead: The University of Melbourne

Partners: Queensland Government; NSW Government; Kansas State University

AAGI Analytics for the Australian Grains Industry

GRDC Grains Research & Development Corporation

Adelaide University

Curtin University

The University Of Queensland Australia

The support of the Grains Research and Development Corporation (GRDC) provided through the Analytics for the Australian Grains Industry (AAGI) project is acknowledged.

References

[1]Vaghefi, N., Thorne, B. J., Blanch, S. G., Kelly, L. A., Sparks, A. H., Little, C. R., & Moore, N. Y. (2026). Economic Impact, Epidemiology, and Environmental Drivers of Charcoal Rot of Soybean. Phytopathology®.
[2]USDA Foreign Agricultural Services, Production – Soybean. https://www.fas.usda.gov/data/production/2222000
[3]Ramsay, J. O., & Silverman, B. W. (1997). Functional data analysis. New York, NY: Springer New York.
[4]Shah, D. A., De Wolf, E. D., Paul, P. A., & Madden, L. V. (2019). Functional data analysis of weather variables linked to Fusarium head blight epidemics in the United States. Phytopathology®, 109(1), 96-110.
[5]Ramsay, J. O., Hooker, G., & Graves, S. (2009). Functional Data Analysis with R and MATLAB. New York, NY: Springer New York.
[6]Queensland Government SILO climate database. https://www.longpaddock.qld.gov.au/silo/
[7]Butler, D., Cullis, B., Gilmour, A., Gogel, B. & Thompson, R. (2017) ASReml-r reference manual version 4. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.