Actuaries Climate Index and Multivariate Wheat Yield Modelling: A Comparison of VAR and LSTM

Final Year Project

Matthew Farant Andreson — TP053983

Asia Pacific University of Technology and Innovation

August 5, 2022

Outline

Section 1

Introduction

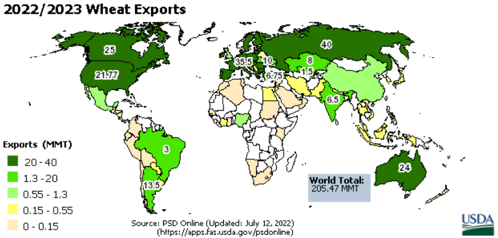
Raging Climate Crisis in Australia

Why Australia?

- The increasing frequency of extreme climatic events in Australia has been alarming over the last six decades (CSIRO, 2020)
- Its impact extents to various fields, including the agriculture sector that significantly contributes to the national food security (Ojumu et al., 2020)

Wheat Production at Risk

Figure 1 2022/2023 Global Wheat Export



- Wheat production needs more focus amid the climate uncertainty
- Understanding how essential wheat production and the potential impact that climate change can cause has motivated this study to examine the impact of climate risk to Australian wheat yield

Climate Risk Measured by AACI

- This study used the **Australian Actuaries Climate Index** (AACI), a retrospective measure of climate change in Australia and its risk towards numerous fields
- The index consists of six components: high temperatures, low temperatures, rainfall/precipitation, drought, wind, and sea level
- Up to this day, there are no initiatives of using the AACI to examine the movement of wheat yield

To Sum Up The Problems

- **Climate change** is a major issue in Australia that manifests itself through the occurrence of extreme weather events
- **Wheat production** in Australia has experienced more disruptions due to frequent extreme weather incidents
- Therefore, it is important, especially for agriculture and insurance companies, to have the ability to **forecast** the potential future loss of wheat yield using the best performing forecasting model

Why VAR and LSTM?

- The strong relationship between wheat production and climate condition in Australia (Hochman, Gobbett, & Horan, 2017) suggested a multivariate forecasting approach
- The Vector Autoregression (VAR) model is known for its flexibility and proven success in dealing with dynamic multivariate time series (Zivot & Wang, 2006)
- The Long Short-Term Memory (LSTM) model is known for its ability to learn complex relationship and solve the vanishing gradient issue (Gonzalez & Yu, 2018)

Research Objective & Question

Formulate both Vector Autoregression (VAR) and Long Short-Term Memory (LSTM) models to forecast Australian wheat yield using the Australian Actuaries Climate Index (AACI) and determine the best performing model

Research Questions

1. What is the VAR model that is used to forecast the Australian wheat yield?
2. What is the LSTM model that is used to forecast the Australian wheat yield?
3. Based on the performance comparison between VAR and LSTM models, which model is the best performing model?

Section 2

Literature Review

The Actuaries Climate Index

The Actuaries Climate Index (ACI) provides a helpful monitoring tool for the frequency of extreme weather and the extent of changes in sea level

Table 1 Components of ACI

Component	Notation	Measurement
Drought	$MaxCDD_{std}(j,k)$	Maximum consecutive dry days (<1 mm) in year
Sea Level	$S_{std}(j,k)$	Sea level
Precipitation	$MaxP^{(5-day)}{}_{std}(j,k)$	Maximum five-day precipitation in month
Warm Temperatures	$FT : warm_{std}(j,k)$	Frequency of temperatures above the 90^{th} percentile
Cool Temperatures	$FT : cool_{std}(j,k)$	Frequency of temperatures below the 10^{th} percentile
Wind Power	$FWP_{std}(j,k)$	Frequency of strongest wind power

(where "j" indicates the month and "k" indicates the year)

The Actuaries Climate Index (Cont'd)

- Each component measures the change in standardized form with regard to the 30-years reference period of 1961–1990 (American Academy of Actuaries, 2021)
- The standardized anomaly for each component shall be calculated as the deviation between the current period and the mean of reference period, which will be scaled later by dividing it with the standard deviation of reference period

$$X_{std}(j,k) = \frac{X(j,k) - \mu_{ref}}{\sigma_{ref}} \tag{1}$$

The Actuaries Climate Index (Cont'd)

- The index is defined by the average of the standardized components, such that it measures an average departure from the mean in terms of the number of standard deviations
- Higher value of ACI implies higher climate risk

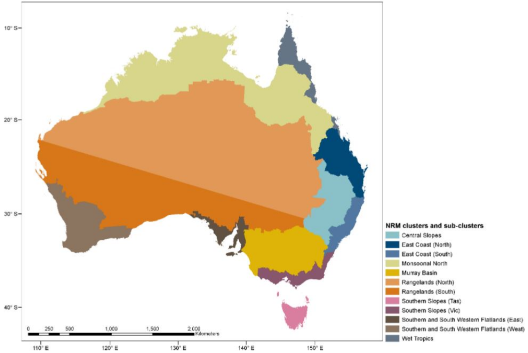
$$ACI(j,k) = \frac{1}{6}\sum_{i=1}^{6} C_i \tag{2}$$

where

C=$\{MaxCDD_{std}(j,k),\ S_{std}(j,k),\ MaxP^{(5-day)}{}_{std}(j,k),\ F\,T : warm_{std}(j,k),\ \text{-}F\,T : cool_{std}(j,k),\ F\,WP_{std}(j,k)\}$

The Australian Actuaries Climate Index

Figure 2 Regions of AACI



ACI as a measure of climate risk

- Unlike AACI, the ACI (ACI for North America) has been used in several studies as one of the measures for climate risk
- A study by **Pan (2021)** suggested that the ACI can provide some insights in crop yield modelling. However, more research is needed as the model changes depending on the **temporal** and spatial resolution of the data

VAR and LSTM in Multivariate Time Series Forecasting

- LSTM is well-known in outperforming traditional statistical methods in univariate time series forecasting (Azari, Papapetrou, Denic, & Peters, 2019; Siami-Namini & Namin, 2018)
- How about multivariate time series forecasting?
  - VAR performed better in the condition of small dataset and linear relationship between multiple time series (Dissanayake et al., 2021; Kaur et al., 2021)
  - LSTM performed better in capturing non-linear relationship (Kuhnert et al., 2021; Zhang et al., 2020)
  - A study by Ouhame Hadi (2019) also explores the possibility of combining both models as an ensemble model

Section 3

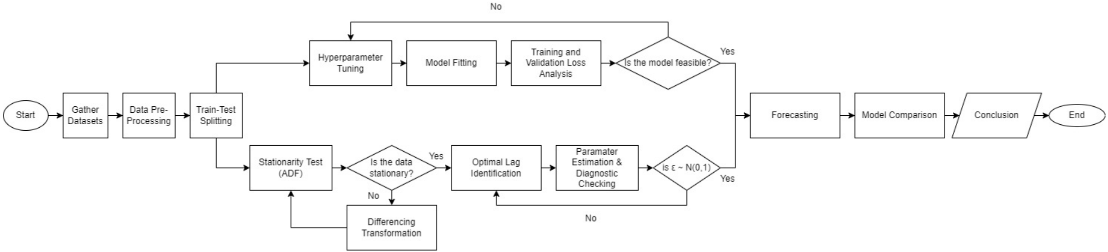
Research Methodology

The Data Used for This Study

- In this study, there were two main datasets: AACI and wheat yield, which are both classified as a secondary data
- The quarterly **AACI dataset** is publicly available in the Institute of Actuaries, Australia website in xlsx. extension
- The **wheat yield dataset** is collected from the Department of Agriculture, Water, and Environment from the year 1981-2020

Research Flow

Figure 3 Research Flow



Data Preprocessing & Analysis

- To gain a sense of the data, an Exploratory Data Analysis (EDA) was conducted in a form of data visualization
- To ensure the quality of the data processed before modelling, feature engineering was conducted for both AACI and wheat yield data
- After the feature engineering process was conducted, the wheat yield data was split into training and testing data (30 years for training + 10 years for testing)

Vector Autoregression

- The first multivariate forecasting process was using the Vector Autoregression (VAR) model that is derived from the Autoregressive (AR) model (Zivot & Wang, 2006)
- An example of a $VAR(1)$ model with two variables/time series, x and y

$$\underbrace{\begin{pmatrix} x_t \\ y_t \end{pmatrix}}_{variables} = \underbrace{\begin{pmatrix} \alpha_x \\ \alpha_y \end{pmatrix}}_{constants} + \underbrace{\begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}}_{estimates} \underbrace{\begin{pmatrix} x_{t-1} \\ y_{t-1} \end{pmatrix}}_{lags} + \underbrace{\begin{pmatrix} \epsilon_{xt} \\ \epsilon_{yt} \end{pmatrix}}_{errors} \tag{3}$$

The General Form of VAR(p)

Hence, a $VAR(p)$ model can be expressed as follows

$$z_t = \alpha + \sum_{n=1}^{p} \phi_n z_{n-1} + \epsilon_t \tag{4}$$

where

z = vector of variables
α = vector of constants
ϕ = matrix of estimates
ϵ = vector of errors

Model Assumptions

There are two main assumptions to be considered in the time series data for building a VAR model:

1. Stationarity (Augmented Dickey Fuller test)
2. Normal and Independent Errors (Durbin-Watson d test and Jarque-Bera test)

Before estimating the parameters in the VAR model, it is also necessary to determine the VAR order or the maximum length of lag.

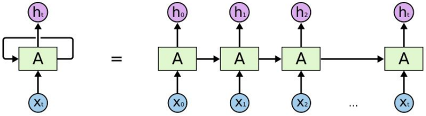
Theoretical VAR Model

After the model assumptions are fulfilled, the theoretical VAR model for this particular study will be expressed as follows. Recall that the Principal Component Analysis (PCA) was conducted beforehand to reduce the number of variables used into n Principal Components

$$\begin{pmatrix} Y_t \\ PC1_t \\ PC2_t \\ \vdots \\ PCn_t \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} + \sum_{i=1}^{p} \begin{pmatrix} \beta_{11,i} & \cdots & \beta_{1(n+1),i} \\ \vdots & \ddots & \vdots \\ \beta_{(n+1)1,i} & \cdots & \beta_{(n+1)(n+1),i} \end{pmatrix} \begin{pmatrix} Y_{t-i} \\ PC1_{t-i} \\ PC2_{t-i} \\ \vdots \\ PCn_{t-i} \end{pmatrix} + \begin{pmatrix} \epsilon_0 \\ \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \tag{5}$$

Recurrent Neural Network

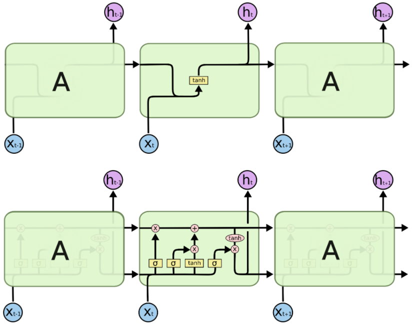
Figure 4 Recurrent Neural Network



- The LSTM model that this study used is a type of Recurrent Neural Network (RNN) that commonly used in text processing, time series forecasting, and several other fields
- RNN loops past information as "memories", hence the current state in RNN is a function of its previous steps

Long Short-Term Memory

Figure 5 LSTM Process



- In practice, RNN can't really handle long-term dependencies. Thankfully, LSTMs don't have this problem
- It is designed specifically to remember information for long periods of time, hence the name "Long Short-Term Memory"

Hyperparameter Tuning

- One of the essential procedures to enhance NN model performance and minimize manual parameter tuning is called as a hyperparameter tuning (Cho et al., 2020)
- The forecasting performance of a NN layers can be described as a black-box function f. The goal of hyperparameter tuning is to maximize the objective function and find the global optimum x^*, such that $x^* = arg\,\underset{x \in \chi}{max} f(x)$
- Two of the commonly used methods are Random Search (RS) and Bayesian Optimization (BO)

Performance Metrics

In this study, the Root Mean Squared Error (RMSE) was used due to its interpretability

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}} \tag{6}$$

Additionally, the Mean Absolute Percentage Error (MAPE) was also used due to its ability to assess model performance with a benchmark

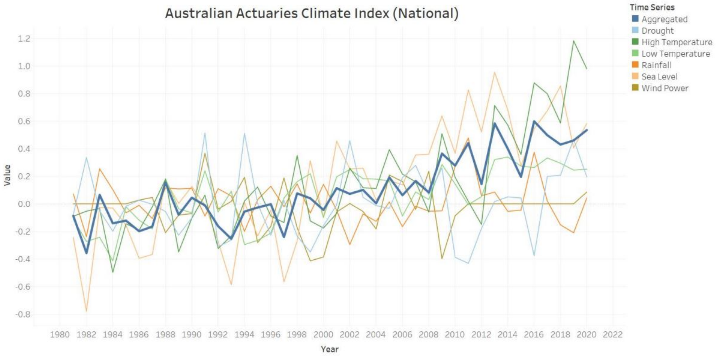
$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \tag{7}$$

Section 4

Result and Discussion

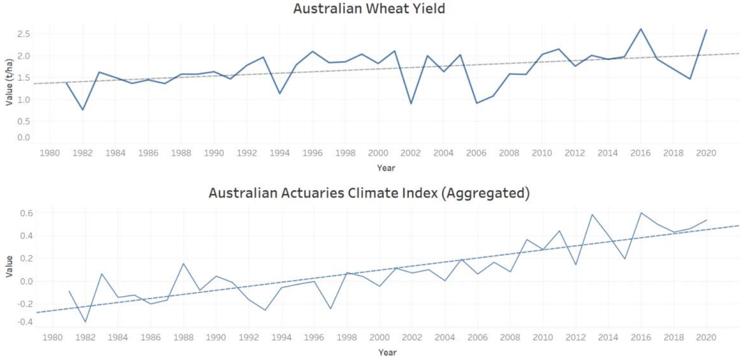
Exploratory Data Analysis of AACI

Figure 6 Time Series of National AACI



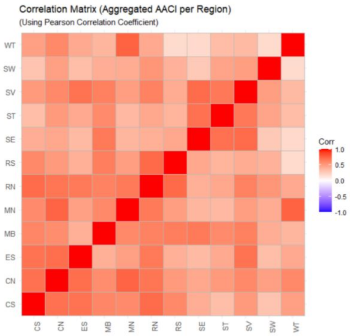
Comparing AACI with Wheat Yield Data

Figure 7 Time Series of Wheat Yield and AACI



Multicollinearity Issue

Figure 8 Correlation Matrix



- Like a non-stationary data, multicollinearity is a phenomenon that needs to be addressed to avoid biased result for the VAR modelling
- LSTM didn't require a dimensionality reduction due to the neural network's natural ability to put weights in each of the variables (Boehmke & Greenwell, 2020)

Principal Components Analysis

Figure 9 Scree Plot of PCA

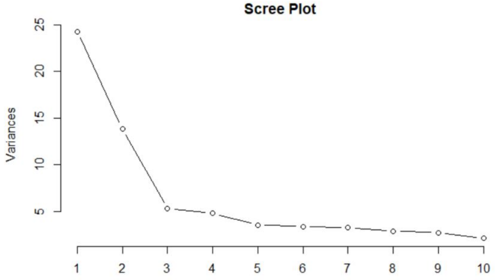


Table 2 ADF Test Result

Variable	No. of Differencing	p-value
Yield	1	0.01
PC1	2	0.01
PC2	2	0.01
PC3	1	0.01475
PC4	2	0.02362
PC5	1	0.0233

VAR Estimates & Forecasting Result

Figure 10 Actual vs. Forecast Plot (VAR)

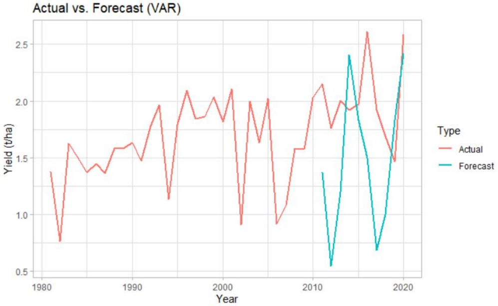


Table 3 VAR Estimates

	term	estimate	std.error	statistic	p.value
1	yield.l1	-0.42	0.52	-0.81	0.44
2	pc1.l1	0.07	0.05	1.24	0.25
3	pc2.l1	0.02	0.05	0.43	0.68
4	pc3.l1	0.02	0.05	0.36	0.73
5	pc4.l1	-0.01	0.03	-0.42	0.69
6	pc5.l1	0.05	0.05	1.04	0.33
7	yield.l2	0.14	0.42	0.34	0.74
8	pc1.l2	0.09	0.05	1.77	0.12
9	pc2.l2	-0.04	0.05	-0.82	0.44
10	pc3.l2	0.01	0.05	0.12	0.91
11	pc4.l2	0.00	0.04	0.12	0.90
12	pc5.l2	-0.01	0.05	-0.13	0.90
13	yield.l3	-0.95	0.44	-2.19	0.06
14	pc1.l3	0.08	0.04	2.13	0.07
15	pc2.l3	0.02	0.03	0.46	0.66
16	pc3.l3	-0.04	0.06	-0.73	0.49
17	pc4.l3	-0.00	0.04	-0.04	0.97
18	pc5.l3	0.02	0.07	0.28	0.78
19	const	0.00	0.07	0.02	0.98

Data Preparation

- Like the VAR model, the datasets were firstly gathered and transformed into a tidy form
- It is important to reshape the dataset first into an appropriate format by scaling it
    - In this case, a standard scaler was used to scale both the wheat yield and AACI data
- 5-year lags of each variable were included in the dataset, considering the possibility of correlation between present and past values of each variable

Hyperparameter Tuning Result

- As mentioned previously, the Bayesian Optimization was used to do the hyperarameter tuning
- BO has achieved its optimized level after three iterations (MSE: 0.00034)

Table 4 Best Parameters by Hyperparameter Tuning

Parameters	Hyperparameter Space	Best Value
Activation Function	{relu,tanh,linear,selu,elu}	relu
Number of RNN Layers	[0,20]	12
Recurrent Dropout	[0,0.99]	0.5
Number of Units	[0,100]	64
Learning Rate	[1e-10,1e-2]	0.01

LSTM Forecasting Result

Figure 11 Training and Validation Loss

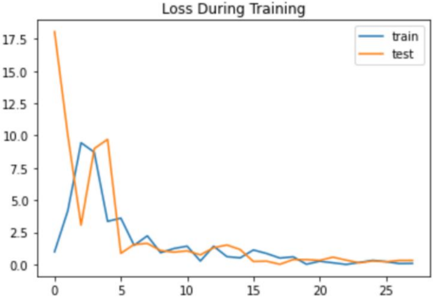
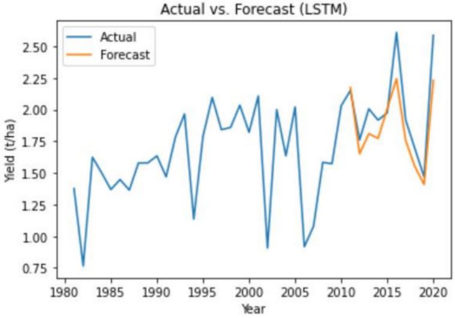


Figure 12 Forecasting Result



Summary of Performance

- Finally, after doing both VAR and LSTM modelling, the RMSE and MAPE values were compared between the two models to determine which model is the best to forecast wheat yield using AACI in this study

Table 5 Summary of Performance

Model	RMSE (t/ha)	MAPE
VAR	0.9591347	38.28%
LSTM	0.1961116	8.32%

Section 5

Conclusion

Conclusion

- As an effort to face climate change in Australia, this study was intended to forecast the Australian wheat yield by using the Australian Actuaries Climate Index (AACI)
- The process of hyperparameter tuning has suggested a VAR(3) model and LSTM model of 12 layers and 64 units (neurons)
- This study has implied the usability of AACI to forecast Australian wheat yield and opened the possibility of using the index to formulate a weather-based crop insurance products in Australia

Limitations of the Study

- As the use of AACI to forecast Australian wheat yield was examined in this study, non-climatic events such as socio-political events and the Covid-19 pandemic was not focused on this study
- Several technical difficulties such as long training time in developing the LSTM model due to hardware limitations and sudden notebook reset that caused memory loss was encountered during the study

Recommendation for Future Studies

- Future studies may look for other sources of crop datasets with quarter granularity to be used for the dependent variable, so that the number of observations can be increased
- To complement the AACI, other types of datasets can also be used as the independent variable for crop modelling
- Studies with more financial and technical advantages may also utilize AutoML in cloud machines to experiment with multiple forecasting models

Thank you!

Matthew Farant Andreson
Actuarial Studies
Asia Pacific University of Technology and Innovation
TP053983@mail.apu.edu.my

Special thanks to:
Supervisor — Dr. Ho Ming Kang
Second Marker — Mohamed Ubaidullah