The Short-term Prediction of Grassland Curing in Victoria using Machine Learning

A thesis submitted to the Faculty of Science, Monash University in partial fulfilment of the requirements for the degree of Bachelor of Science Advanced (Global Challenges) – Honours

Gabrielle Malley 29755727

Kate Howard 29715539

Phuoc Gia Hy Duong 30510597

Monash University

July 2022

Supervisors:

Dr. Ailie Gallant

Monash University, School of Earth, Atmosphere and Environment

Dr. Caroline Poulsen

The Bureau of Meteorology

Dr. Danielle Wright

The Country Fire Authority

Table of Contents

List of Tables

List of Figures

Acknowledgements

We want to begin by acknowledging and paying our respects to the Traditional Owners and custodians of country throughout Australia on which this Honours project encompasses and Monash University operates. In light of the project topic, our consideration extends to the practical and spiritual importance of fire and country in the world's oldest continuous culture.

To our supervisors, Dr. Caroline Poulsen, Dr. Danielle Wright, and Dr. Ailie Gallant, thank you so much for your continued support and committed supervision of this project. From beginning to project fruition, your guidance, expertise, and patience have made this process a profound and memorable learning experience. An additional thank you to Danielle, whose in-depth understanding of the CFA and its operational practices was pertinent for this project's real-world impact.

To Dr Thomas Duff, thank you for your considerable contributions to our machine learning understanding. It took your input for us to understand the depths of this topic truly.

To Dr Leon Majewski, thanks for helping us climb the steep learning curve of coding and everything that it entailed. Your limitless patience and optimism made the trek more endurable and fun than it had any right to be.

To the Global Challenges course coordinators, Michael Brown and Djuke Veldhuis, thank you both for your dedication to delivering a course with students at the forefront of consideration. We have immensely enjoyed the last four years and have learnt so much from the experience.

Finally, thank you to all our friends and families who have supported us along the way; your unwavering support, empathy, and love have gotten us through more stressful periods than we'd like to admit.

Executive Summary

Grassland curing describes the transition of grass from live fuel to dead fuel. It is an essential input for calculating grassland fire danger in Australia. At present, grassland curing modelling is based solely on current conditions with no short-term forecasting model used in operation. The lack of a predictive tool reduces the accuracy of fire danger predictions, limiting fire management agencies' ability to make informed risk management decisions. The project aimed to use the eXtreme Gradient Boosting (XGBoost) algorithm to build a 4-day forecast model for the degree of curing and investigate the predictability of the grassland curing process in Victoria. XGBoost was chosen because of its high efficiency and scalability, suitable for operational purposes.

This project used reanalysis data to build and compare three models: one with curing and meteorological variables, one with only meteorological variables, and one with only curing variables. The model with just curing variables struggled to predict a decrease in curing, whilst the two meteorological-based models did not have a tendency to under or over-predict. The higher accuracy and lower variability in error of the model with both curing and meteorological variables made it the best-performing model in this research. Further validation of this model revealed that it performed well with new data and that it predicted fully cured grass better than partially cured grass. However, during weeks of consecutive cloud cover, there are operational benefits in having a model that does not include curing.

Overall, the project demonstrated that a machine learning model could accurately predict curing in Victoria; however, more research into the selection of variables is needed, and the model needs to be further validated on a larger dataset. Additionally, to be operationally viable, the model needs to be built from forecast data, not reanalysis data. Once operational, a curing forecast model should lead to more accurate grass fire danger predictions and fire preparedness.

1. Introduction

1.1 Fire in Australia

Fire has shaped the land and life histories of all its inhabitants across the Australian continent for millennia (Abram et al., 2021, Battistel et al., 2018). In recent history, the frequency, intensity, and scale at which fire has shaped the land has grown to become a substantial threat to the social, economic, and environmental stability of Australian communities (Adeleye et al., 2021, Collins et al., 2021). In the last two decades, the southeastern region of Australia has experienced destructively record-breaking fire seasons (Adams et al., 2020). Between the 2009 Black Saturday bushfires in Victoria and the Black Summer bushfires of 2019-20, 203 human lives have been lost, estimates of over 3 billion animals killed, and approximately 24 to 34 million hectares burned (Squire et al., 2021, Mark B. et al., 2020, Peg, 2019, Parrott et al., 2021).

In 2019-20, Australia's agriculture sector was estimated to have lost $4 billion to $5 billion to bushfires. Among the costs are damage to farm buildings and equipment, loss of crops, livestock deaths, and reductions in farmland value, as well as health impacts related to smoke inhalation (Bishopa et al., 2021). In southeastern Australia, grass fire events disproportionately impact rural and agricultural communities (Cowled et al., 2022). The majority of fire-based research in Australia focuses on forest fires; however, it is estimated that 70-85% of the Australian landscape is grass (New, 2019). Despite the prevalence of grassland and the impact of grass fires on rural and agricultural communities, research into understanding the risk of grass fires is lacking in southeastern Australia (Canadell et al., 2021, Deb et al., 2020, Nolan et al., 2020).

Operationally in Australia, the categorisation of fire danger is stratified by fuel type – grass fire danger and forest fire danger (AFAC, 2022). While grass fires produce fewer embers and burn less intensely than forest fires, they can generate substantial amounts of radiant heat, start easily, burn unpredictably, spread rapidly, and reduce air quality through smoke production (Cheney et al., 2012, Lindenmayer et al., 2003). An important factor in measuring fire danger is the moisture content of the fuel. In grasslands, this is represented by grass curing, which encapsulates the lifecycle of grass. The curing cycle can change over short time frames in response to multiple environmental factors, which adds complexity to predicting grass fire danger (Cheney and Sullivan, 2009).

Curing is measured as a percentage range, from 0% representative of green grass to 100% indicative of completely dried grass (Duff et al., 2020). Curing is monitored in near real-time, meaning operational four-day fire danger forecasts and seasonal bushfire outlooks rely on the current curing conditions. A seasonal curing forecast has been developed by the Bureau of Meteorology (referred to as the Bureau) that models curing three months ahead; however, no short-term forecast model exists (Martin et al., 2016). Given the accessibility of weather observations and forecasts, there is the potential to reconstruct curing levels from historical weather data and predict curing in the future (Ferreira et al., 2020). A short-term curing forecast has the potential to improve the accuracy of grass fire danger predictions, which would be beneficial for fire management and preparedness (D Wright 2022, personal communication). Recent advancements in machine learning have shown it to be a powerful predictive tool (Jain et al., 2020) and may be a possible method of creating such a forecast (Cruz et al., 2018).

1.2 Grass Fire and Fire Danger Indices

The rate that grass fires spread makes them particularly dangerous and challenging to maintain during firefighting operations (Cruz et al. 2015), especially given firefighting resources are typically not requested until the fire has escaped the early suppression efforts of local resources (Plucinski, 2013). To reduce the likelihood of fires getting out of control, implementing robust evidenced-based systems for monitoring and predicting its activity has been a priority area of research for critical stakeholders responsible for keeping Australian communities safe (BoM, 2021, CFA, 2021). Estimates of grass fire danger have been established using simple metrics that describe how likely a grass fire is to spread, the difficulty of fire suppression based on the vegetation fuel load and overlying meteorological conditions (Hines et al., 2010).

Operationally in Australia, there are two broad categories of fire danger; grass fire danger and forest fire danger, each with an associated Fire Danger Index (FDI) (AFAC, 2022). The Grassland Fire Danger Index (GFDI) measures grass fire risk and fuel behaviour for grass fuels. GFDI inputs include temperature, relative humidity, average wind speed at 10 m, fuel loads, and curing (Khastagir et al., 2018). A constant fuel load of 4.5 t/ha is assumed for indices computed for Victoria, while other states use variable fuel loads (Gould and Cruz, 2012). GFDI is calculated using equation 1, as follows:

$$GFDI = Q^{1.027} \times 0.2180567exp(-0.009432 \times (100 - C)^{1.536} + 0.02764T - 0.2205\sqrt{RH} + 0.6422\sqrt{V})$$

(Equation 1)

where Q is fuel load (t.ha^{-1}), C is the degree of curing (%), T is air temperature (^{o}C), V is 10-m wind speed (km h^{-1}), RH is relative humidity (%) (BoM, 2022).

The 2022 to 2023 fire season will see the introduction of the new Australian Fire Danger Rating System (AFDRS). The AFDRS incorporates a new measure of fire danger: the

Fire Behaviour Index, which considers eight fuel types instead of two as for the previous system. Despite the implementation of the AFDRS, the CFA continues to use the GFDI as part of their readiness response operations. The new grassland fire danger component of the Fire Behaviour Index is very similar to the GFDI and will still use curing as its main indication of fuel moisture in grasslands. This means that a curing forecast may benefit the CFA and other fire agencies using the AFDRS (Matthews et al., 2019).

1.3. Grass Curing

The process of grass curing describes the transition of grass from live fuel to dead fuel (Cheney and Sullivan, 2009). Quantitatively, the grass curing degree describes a proportion of dead grass in a grassland fuel bed. Grass species exhibit one of two growth cycles; annual or perennial growth. During a single season, annual grasses flower, seed, and then senesce. Conversely, perennial grass species survive multiple seasons, typically undergoing curing after seed production (CFA, 2014, Duff et al., 2020). Grassland curing and the degree of grassland curing are critical factors in the prediction of fire behaviour and grassland fire danger (Cruz et al., 2016, Kidnie et al., 2015). Fire propagation can occur in grasslands with curing levels as low as 20% (Cruz et al., 2016). Past studies have demonstrated a positive relationship between curing and the rate of grassfire spread. Given how dramatically dry grass can affect fire dynamics, grass curing remains the best predictor of whether a fire will propagate and sustain in a grassland fuel bed (Cruz et al., 2015).

1.3.1 Curing Dynamics

Curing dynamics can have a variable yet relatively predictable seasonal cycle that can vary from year to year (Malley, 2022, Sullivan et al., 2012). The rate of curing is affected by fuel moisture content and the growth and development life cycle of grass (Kidnie et al., 2015, Duff et al., 2020). Similarly, curing dynamics can fluctuate in response to a range of environmental factors, but vary predominately in response to water availability and temperature (Wittich, 2010). Temperature also modulates plant water availability through its effect on vapour pressure deficit (VPD), soil moisture, and relative humidity, as high temperatures exacerbate drought conditions which can restrict terrestrial plant productivity. A reduction in plant productivity can trigger early seeding, accelerating the curing process (Fahad et al., 2017, Will et al., 2013).

For grasslands in the temperate zone, spring brings a period of growth, increasing the fuel load of the grass bed while maintaining a high live FMC. In late spring to early summer, curing accelerates, as most grasses have produced a fully mature seed head and begin to lose their live moisture content (Cheney and Sullivan, 2009). The rapid increase in curing heightens the potential for a grass fire to ignite, spread, and sustain across the landscape (Cruz et al., 2015). As summer progresses, grasslands will continue to cure and eventually die or become dormant. At this stage, the dead grass moisture content is a direct function of atmospheric conditions (Cruz et al., 2016). However, perennial grass species can recover and regain live moisture content in high rainfall events, in a process called secondary growth. During a secondary growth cycle, grass recovers both with and without germination, dependent on species (Nerlekar and Veldman, 2020, Rayment and French, 2021).

1.3.2 Modelling and Monitoring Grass Curing Victoria

During the dynamic process of curing, grasses change physiologically, transitioning from a green to yellowish-brown in colour due to the gradual loss of chlorophyll and moisture content (Dilley et al., 2004). This physiological change can be monitored via satellite remote sensing at various spatial scales using spectral bands in the visible, near infrared and mid-infrared regions of the electromagnetic spectrum in combination with vegetation indices (Appendix i.) (Martin et al., 2009). In Victoria, the Victorian Improved Satellite Curing Algorithm (VISCA) is the current system used to monitor curing. VISCA maps of grassland curing with a spatial resolution of 3 km across Victoria are produced once a week by combining ground observations with the MapVictoria satellite-based grassland curing model, processed by the Bureau every day (Martin et al., 2016, Martin et al., 2015). This is shown in equation 2,

$$VISCA = C_{sat} - \frac{(C_{sat} - C_{ground}) \times GBI}{100} \quad \text{(Equation 2)}$$

where C_{sat} represents the near-real-time satellite curing data, C_{ground} represents the weekly ground-based curing observations and GBI is the Ground-Based Influence coefficients. The GBI takes into account elevation, distance from observation site and age of satellite data (Appendix ii. for GBI calculation) (Martin et al., 2015).

Equation 3 below, describes the vegetation indices used in the MapVictoria model to determine the degree of curing; the Normalised Difference Vegetation Index (NDVI) and the Global Vegetation Monitoring Index (GVMI) (Rouse Jr et al., 1973, Ceccato et al., 2002).

$$C_{sat} = (NDVI \times -88.41) + (GVMI \times -67.71) + 113.80 \quad \text{(Equation 3)}$$

where C_{sat} represents the near-real-time satellite curing data, NDVI is the Normalised Difference Vegetation Index and GVMI is Global Vegetation Monitoring Index (Appendix iii. for indices calculation) (Martin et al., 2015).

Due to limitations in the MapVictoria data, both the underestimation and overestimation of curing can occur. Cloud cover can impede the selection of a satellite image on a particular day; or can result in missing data where cloud cover is persistent over multiple sampling periods. Additionally, land surfaces can introduce biases, for example, sand dunes, bare soil, urban areas, undetected water bodies, and natural landscapes or cropland dominated by yellow flowers can result in an overestimation of curing. Likewise, the model can also underestimate curing in tree-contaminated pixels (Newnham et al., 2010, Martin et al., 2009). Although recently, tree contamination has been minimised through the employment of a Tree Density – VicMap Vegetation Dataset (White et al., 2020). As previously described, ground truthing for the VISCA model through surface observations can help to minimise the biases described. However, given VISCA maps are only produced operationally on a weekly basis this poses its own limitation given MapVictoria data is generated daily (Martin et al., 2015, Wright et al., 2016).

Despite its advantages, when used in fire management planning, near real time remote sensing has a fundamental limitation since grassland curing estimates are dependent on observed grassland conditions (Martin et al., 2015) and are not currently forecast in advance. The rapid fluctuations in grass curing over short time periods and the lack of a forecasting model to capture this change is a concern for fire weather forecasters and fire management authorities (D Wright 2022, personal communication). However, the impact of fire on Australia's people, economy, and environment is expected to reduce as new technologies in fire research are adapted (Twidwell et al., 2022).

1.4. Machine learning

1.4.1. What is machine learning?

Machine learning (ML) is a form of artificial intelligence that uses and develops computational algorithms to learn patterns and relationships between data (Mueller, 2021). It then uses those algorithms to build predictive, descriptive or actionable models. Because ML methods learn directly from the data they do not need physical models or rules to operate – instead developing their own internal model. This means they do not need to be explicitly provided with the physical parameters, a significant advantage in fields where physical models need to consider many parameters or complex relationships, or where no physical model yet exists (Murphy, 2012). This is the case for modelling grassland curing, and ML therefore may provide a novel approach to its prediction.

A review by Jain et al. (2020) described ML being used for diagnostics like fire detection and mapping; and predictive and prescriptive analytics such as fire management and forecasting. They also found that the availability of remote sensing data has improved the accuracy of the ML application in fire research. To ensure that appropriate data is used for the model, they noted that researchers should work with fire or domain experts. Therefore, this project was conducted in consultation with the Bureau and the CFA.

1.4.2. eXtreme Gradient Boosting

A well-established and popular learning algorithm is the eXtreme Gradient Boosting (XGBoost) algorithm. XGBoost models consistently achieve high accuracy scores and have fast running time, making them suitable for operational deployment. They are capable of handling missing values which is a unique advantage in the ML field (Chen and Guestrin 2016).

XGBoost, like any ML algorithms, has the "black box" limitation, meaning that the decision-making process of the algorithm is inaccessible (Wang et al. 2021). In other words, while the model predictions can be accessed, the quantitative relationships between the input and output variables cannot be extracted. The lack of interpretability can raise concerns about the research implications and reduce the end user's trust in the results (Jain et al. 2020). Improvements in 'explainable AI' have helped the issues, with sets of tools and frameworks that aid researchers in understanding and interpreting predictions made by ML models (Wang et al. 2021). There are multiple tools currently available for different coding languages and compatible for different ML algorithms that can rank the contribution of the input variables and identify if a predictor positively or negatively influences the output (Lundberg and Lee, 2017).

Considering the performance advantages, the applicability in fire science research and the existing solutions to address the interpretability issues, XGBoost is a suitable algorithm for this research's interest in predicting grassland curing.

1.5. Aims and Objectives

The project undertaken was supervised by two leading academic supervisors from The Bureau of Meteorology (BoM) and the Country Fire Authority (CFA). The overarching motivation of the research project is to improve the accuracy of fire danger predictions to enhance community safety, a motivation closely aligned with the institutional values of both the CFA and BoM.

Grass fire danger is currently predicted up to four-days in advance, however this prediction assumes curing is constant, as there is no operational curing forecast model. Given grass curing can fluctuate over short time periods, a model that accurately captures this change would have the potential to increase the accuracy of the fire danger predictions. Thus, the aim of this research project is to use the eXtreme Gradient Boosting machine learning algorithm to build a 4-day forecast model for the degree of curing. The overall intention of the project is to investigate the predictability of the grassland curing process. This would assist the CFA and Bureau in better understanding the potential to forecast grassland curing. There are four main objectives to the project:

(1) To build, optimise, and compare different compositions of a four-day grass curing forecast model and determine which model best predicts the curing process.

(2) To evaluate the importance of the input features on the model learning and explore their influence on the grass curing process.

(3) To identify the spatial and temporal variation in the observed and predicted curing value estimates and assess the spatial and temporal error of the best-performing curing model.

(4) To validate the best-performing curing model using a case study analysis of a grassfire incident in Victoria.

2. Methodology

This section is divided into five subsections. Subsection 2.1 describes the focus area of the research, as well as any project scope restrictions that affected the data included. Subsection 2.2 describes the datasets used to meet the study aim. Subsection 2.3 explains the machine learning and XGBoost fundamentals. Subsection 2.4 details the considerations and steps taken to prepare and process the data. Subsection 2.5 outlines the specific approach taken to achieve the four project objectives.

2.1 Site Description and Project Scope

The study site was the state of Victoria (140.96°E - 149.98°E, 33.98°S - 39.16°S), southeastern Australia (GA, 2004). The time frame of analysis included the seasonal curing ramp-up periods in Victoria, as established by prior research, from the spring to summer months of September to February from 2015 to 2018 (Malley, 2022). By limiting the modelled data to the curing ramp-up period, the model was simplified and trained to recognise the relationship and variables that lead to an increase in curing, which is generally aligned with an increase in fire danger (Howard, 2022, D Wright 2022, personal communication).

Literature on plant physiology and grass curing was used to determine which variables should be included, however curing research has primarily focussed on the impact of a particular variable, and no comprehensive set of variables was found. Further investigation, as well as hypotheses from academics and supervisors, resulted in a priority list of variables with likely influence on the grass curing process (D Wright and C Poulsen, personal communication). Each variable needed an accessible gridded dataset of moderate to high quality to be included in the project (see Section 2.2 for variables).

2.2. Dataset description

2.2.1. Curing data

The curing data used in this project was calculated using the MapVictoria curing model and the daily MODerate resolution Imaging Spectroradiometer (MODIS) satellite data (Equation 3). Results from the VISCA model were excluded as the VISCA maps are only produced on a weekly basis, hence were not suitable for daily forecasting.

There were a few temporal gaps of up to seven days within the MapVictoria curing dataset. Since these gaps were randomly distributed and did not account for a substantial portion of total curing data available, they were assumed to not have a substantial impact on the model development and analysis.

MapVictoria satellite curing data are gridded data of 0.005° (500 m) spatial resolution. Raw curing data can range between 0 and 100, with 0 representing green grass and no curing, while 100 represents fully cured grass. The MapVictoria satellite curing dataset also includes values between 101 and 115 as an operational margin which were adjusted to a value of 100 here, clouds and surface water can register with a curing value of 255 to indicate a contaminated pixel, and so any instances were removed from the dataset.

MapVictoria curing data are accompanied by an age for each observation, in ‘number of days’ (Martin et al., 2015). For example, an observation from the current day has an age value of 0, yesterday, an age value of 1, etc. Only the observations with age 0 were included to avoid repetition of curing observations.

2.2.2 Meteorological data

Precipitation data were obtained from the Bureau’s Australian Gridded Climate Data (AGCD)/AWAP v1.0.0 (Jones et al., 2009). AGCD is a gridded dataset of 0.05° (5km) spatial resolution. Each grid box is allocated with the daily total precipitation, measured in millimetres (mm). Temperature (^{o}C), relative humidity (%), wind speed (m s^{-1}), incoming shortwave and longwave radiation (W m^{-2}) were obtained from the Bureau of Meteorology Atmospheric high-resolution Regional Reanalysis for Australia (BARRA) datasets. The project used the BARRA-R model which covers the entire Australian continent. BARRA-R gridded data have a spatial resolution of 0.1° (approximately 12 km) and extend over 70 levels up to 80 km into the atmosphere.

All five variables were from BARRA's 'analysis' products rather than 'forecast'. For every variable, there are four forecasts generated at 00UTC (Coordinated Universal Time), 06UTC, 12UTC and 18UTC for the next 6-hour period every day. The analysis process happens three hours after each forecast starts and covers the next six hour period, meaning there are four corresponding analyses every day (Su et al., 2019). The analysis products were used for this reproject because they have included any corrections needed for the forecasts, meaning they are closer representations of the real conditions. For each variable, a daily minimum and maximum observation were extracted from the four analyses and used in this research.

Wind speed was calculated using the 10-metre zonal and meridional velocities (Equation 4), as follows

$ws = \sqrt{u^2 + v^2}$ (Equation 4)

where ws is wind speed, u is the zonal velocity (u-wind component) and v is the meridional velocity (v-wind component).

Wind observations have historically been hard and complex to model accurately (Jakob, 2010). It should be noted that BARRA experiences issues of overestimating low wind speeds and underestimating high wind speeds (Su et al., 2019).

2.2.3. Vapour pressure deficit

Vapour pressure deficit (VPD) describes the difference between the water vapour pressure at saturation and the actual water vapour pressure for a given temperature (unit: kPa). VPD was calculated following equation 5.

$VPD = e_s - e_a$ (Equation 5)

where e_s is saturation vapour pressure and e_a is actual vapour pressure.

Saturation vapour pressure e_s was calculated following Equation 6

$e_s = 6.1078 \times exp(\frac{17.269 \times T}{237.3+T})$ (Equation 6)

where T is temperature

Actual vapour pressure e_a was calculated following equation 7

$e_a = \frac{(e_s \times RH)}{100}$ (Equation 7)

where RH is relative humidity

Temperature *T* in Equation 6 was the daily maximum temperature and relative humidity RH in Equation 7 was the daily minimum relative humidity (A Gallant 2022, personal communication), both of which were products of BARRA.

2.2.4. Soil moisture data

Soil moisture data were obtained from the European Space Agency Climate Change Initiative Soil Moisture (ESA CCI SM) dataset, representative of the upper-most (~ 0-5cm) solid layer (unit: %) (Dorigo et al., 2017, Gruber et al., 2019). The project used the 'combined' product which takes in consideration various single-sensor active scatterometer and passive radiometer soil moisture products (Dorigo et al., 2017).

2.2.5. Supporting data

Due to the nature of the MapVictoria model, curing data provided for this project were available for the entire state, regardless of the actual fuel type. To identify the pixels of the datasets that are actually grassland, additional fuel type data were provided by the CFA. The fuel type data provided the new AFDRS fuel classifications, which includes eight fuels types, instead of just grass and forest, allowing for a more selective subsetting of fuel and ensuring only true grassland is included. The fuel type data should help build a more accurate model, as the previous work on this project found that curing measurements are not as accurate in other fuels that are currently considered grass by the fire danger index (Howard, 2022).

Previous work on this project also established that grass at different locations across the state of Victoria cures at different rates (Malley, 2022). To enhance the value of the curing data, variables containing spatial and temporal information such as longitude, latitude, day of year and day of cycle were extracted and attached to each curing observation.

2.3. Machine Learning

2.3.1. Supervised Learning

ML models typically fall into two fundamental categories: supervised and unsupervised learning. Supervised learning focuses on using the relationships that it learns about the input variables and the variable to be predicted, so that it can make a prediction. On the other hand, unsupervised learning focuses primarily on understanding the relationships between all variables. The supervised learning approach was chosen given the aim of this research is to build a forecast model and that curing is a generally well-understood process.

Any supervised learning framework contains two general principles: (1) the model and coefficients and (2) the objective function. In the first principle, the model is “the mathematical structure by which the prediction y is made from the input x” and the coefficients describe how each input x relates to prediction y. The coefficient is determined by the algorithm as it learns about the data (Cho et al. 2021).

The second principle of supervised learning is the objective function, consisting of a training loss function and a regularisation term. The training loss function measures how predictive the model is with respect to the actual output, with a common one for regression problems is *mean squared error* (Equation 8) (Cho et al. 2021).

$L(\theta) = \sum_i (\hat{y}_i - y_i)^2$ (Equation 8)

where L is the mean squared error loss function, $\hat{y}$ is the actual output and y is the predicted output

The regularisation term controls the model complexity and is important to prevent model overfitting (Cho et al. 2021). Overfitting occurs when the model is only able to predict very well during training but not in long-term practice. The common causes of overfitting are overexposure to information during training or inappropriate model settings that encourage complex model structure (James et al., 2013).

2.3.2 eXtreme Gradient Boosting (XGBoost)

XGBoost is a supervised learning algorithm built on the basis of *gradient boosted trees* which is a combination of the *decision tree* concept and the *gradient boosting* technique (Cho et al. 2021). A decision tree starts with a *root node* which contains a randomly selected variable. Based on the different thresholds of that variable, the tree then splits into *branches* which can have two possible outcomes. The first outcome is a *leaf node* which is a predicted output. The second outcome is a *decision node* containing another variable, from which the tree continues to split into branches. The process of growing a decision tree is complete when all the leaf nodes cover all the possible objects to be classified or the entire range of the discrete output values (James et al., 2013).

Each decision tree can be considered a model itself, however a single tree often has limited predictive capabilities and is often referred to as a *weak learner*. As a result, a functioning ML model is rarely built from just one weak learner but rather a combination of them to form a *strong learner*. One method of combining weak learners is *boosting*, which allows each successive weak learner to improve upon the errors of the previous one, resulting in the final strong learner being the most optimal weak learner (James et al., 2013).

Gradient boosting is a subdivision of the general boosting method where one weak learner is optimised before the next one is built. Optimising a weak learner refers to the optimisation of its loss function. In the gradient boosting technique, the loss function is optimised by taking its gradient, or derivative, to zero. When using the mean squared error loss function, the outcome of each gradient boosting round is a correction term to the current weak learner that helps it better fit to the actual observations. That correction term is carried onto the next weak learner (Cho et al. 2021).

2.4 Input data preparation

The main technique adopted to process the input data was collocation. Collocation is a procedure commonly used in remote sensing to match measurements derived from multiple sources. In this study, since degree of curing is the variable of prediction, other gridded datasets including BARRA, AGCD, ESA CCI SM and the fuel type classification were collocated on the MapVictoria gridded satellite curing dataset, and satellite age dataset, to find the corresponding ground conditions.

In order to train the 4-day forecast model to predict the current-day (t_0) curing values, observations 4 days prior (t_0-4) or more must be obtained. For every variable, this research used daily observations of the week leading up to day t_0-4 and the averages 4-day and 7-day prior to day t_0-4. In the context of this research, each daily observation was attached with the "_*x*day" suffix, where *x* is the number of days prior to the current day (t_0). Each average was attached with the "_avg_*y*day" suffix, where *y* is the number of days prior to the day of prediction (t_0-4). The combination of a variable and a suffix is referred to as a "feature" hereafter. There were up to 133 total features (Appendix iv).

Whilst it is common practice in machine learning to standardise data, XGBoost does not require standardisation as each branch of the decision tree only splits from one single feature (Raschka, 2014). Therefore, this study did not standardise the input data.

XGBoost includes a sparsity-aware split function which aids the algorithm's ability to handle missing values within the input variables (Chen and Guestrin 2016). Therefore, missing data of non-curing variables were retained in the input dataset. However, missing values in the output variable, curing, need to be handled manually. Days with no curing values due to cloud cover or non-current satellite age were therefore excluded.

2.3.4. Training-Testing-Validating datasets

The XGBoost algorithm requires input data to be divided into training, testing and validating datasets. For each of the datasets, the variable that the machine learning tries to predict is called the *target*, which is kept separate from the features. The training dataset is used by the machine learning algorithm during its training stage to understand the relationship between the features and the target. After the machine had finished training, it used the features of the independent testing dataset to make predictions which were compared against the testing target to assess the model performance.

While not always necessary, validating data are often real-life observations that are fed to the machine learning model to understand its predictive capabilities when processing unseen data. Validation of the curing model used a case study of a fire event in Scotsburn

(Ballarat, Victoria) in November and December 2015. The validating dataset was isolated from both the training and testing datasets.

It is important to provide sufficient information to adequately train the model, and still allow for independent testing. Here, two months were removed for the Scotsburn case study, and then 80% of the remaining data was used to train the model and 20% to test it. The 80:20 ratio is commonly used in the ML field (Joseph, 2022).

Two other important considerations for the training and testing datasets were their spatial and temporal representation of the research scope and the distribution of the target. Figures 1 to 4 illustrate the spatial and temporal coverage of the training and testing datasets. Figure 5 to 6 showed the distribution of the curing values in both datasets.

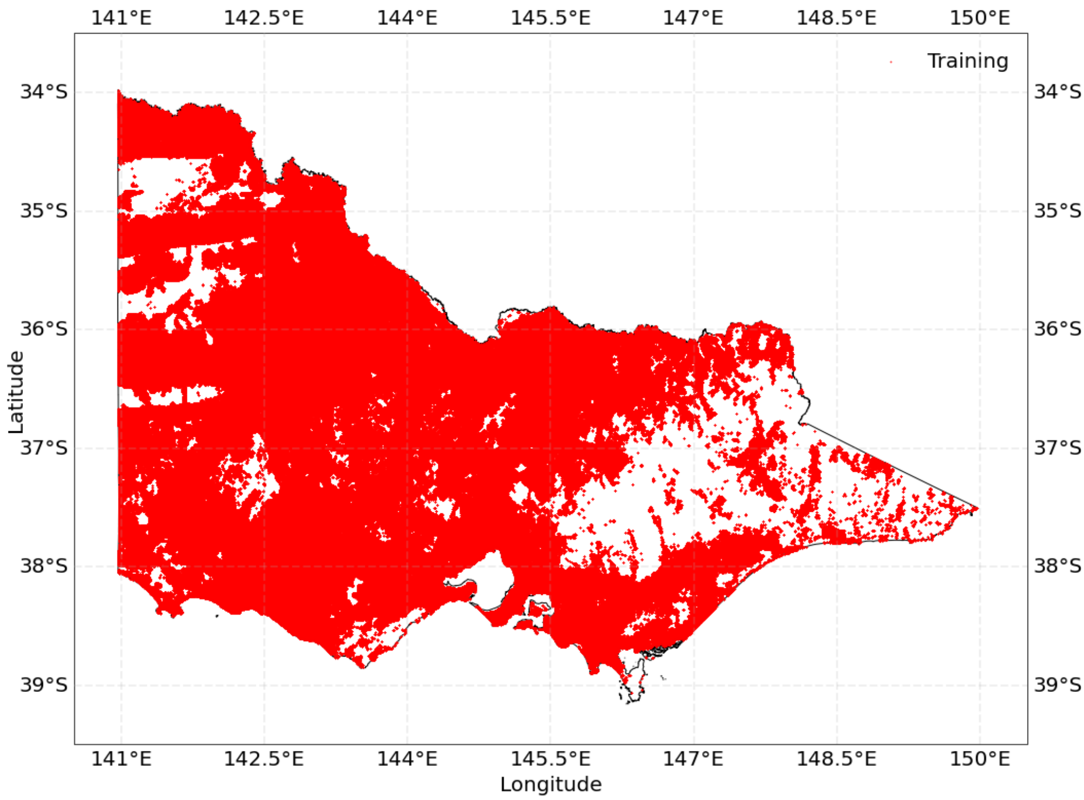


Figure 1: Spatial coverage of the training dataset across Victoria for the study time frame.

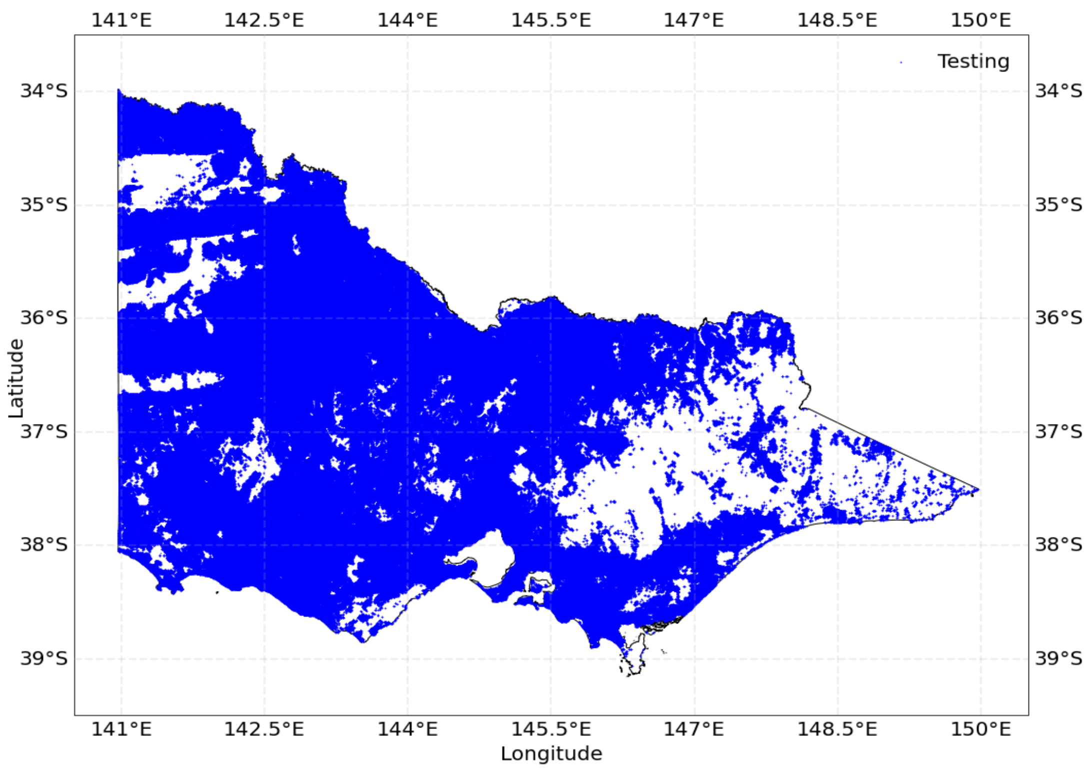


Figure 2: Spatial coverage of the testing dataset across Victoria for the study time frame.

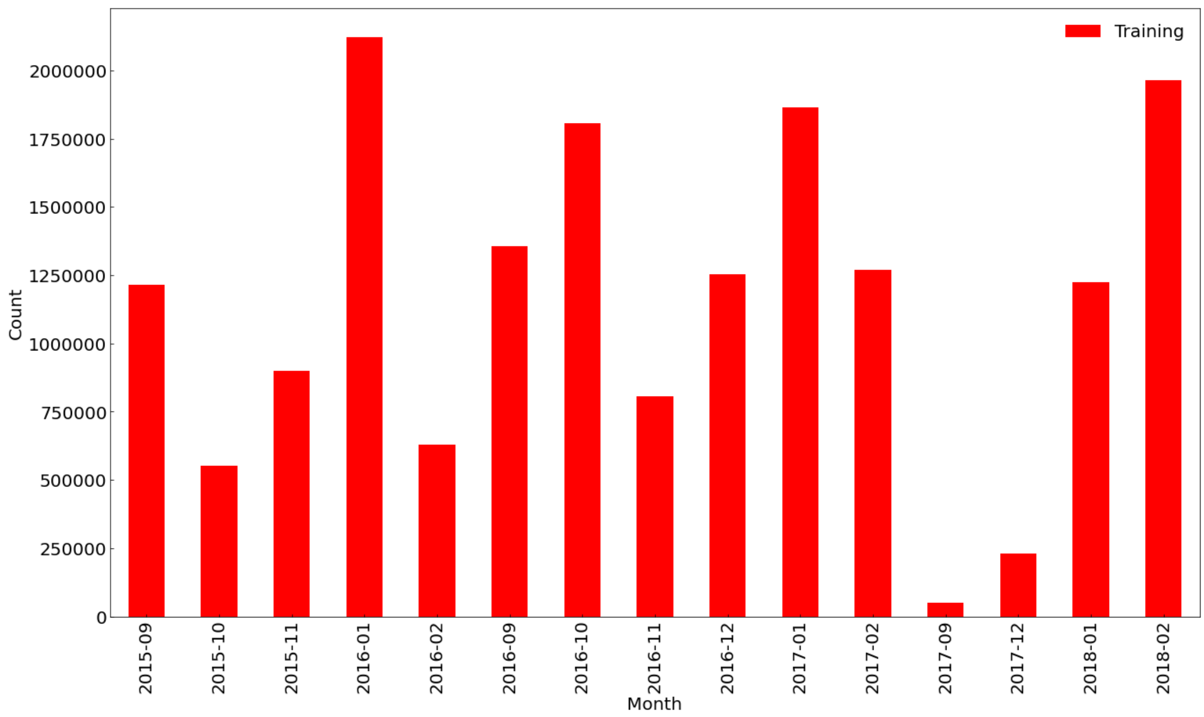


Figure 3: Temporal coverage of the training dataset over the study time frame.

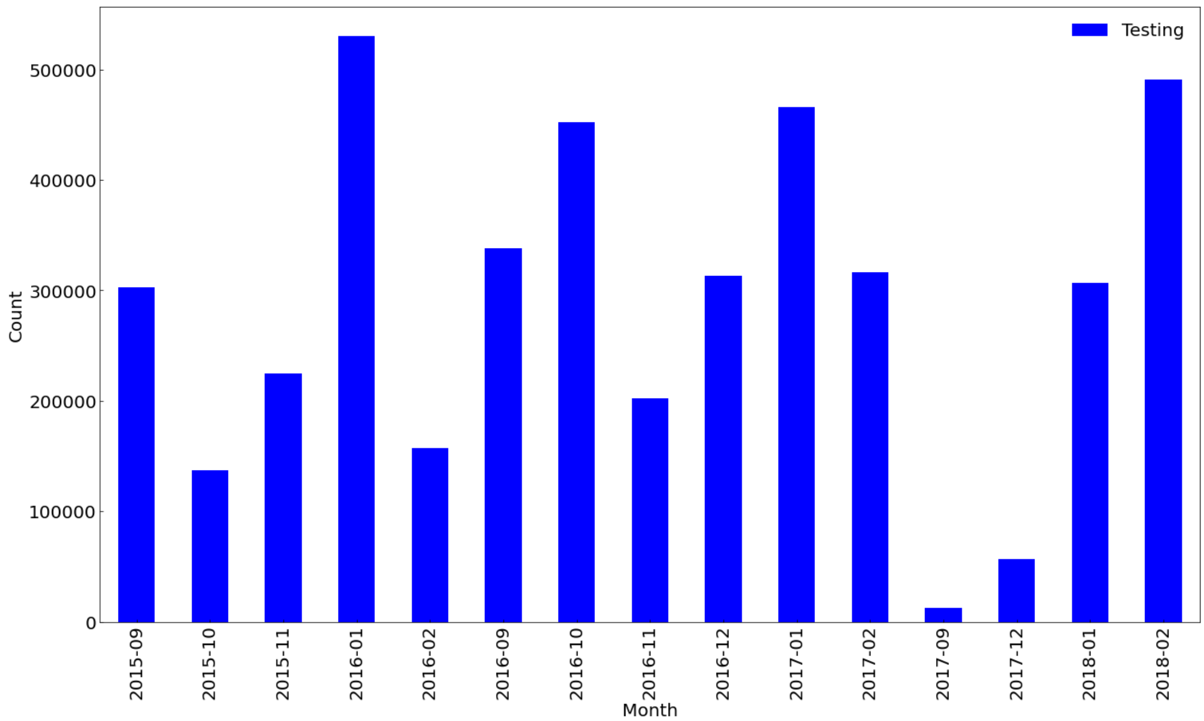


Figure 4: Temporal coverage of the testing dataset over the study time frame.

Both training and testing datasets had comprehensive spatial coverage of all the grass pixels in Victoria over the four years of research (Figures 1 and 2). The white space represents non-grass areas and is true to the AFDRS fuel map (Appendix v). Some months did include more data than others, however this is due to limitations within the dataset. The training and testing data still had the same distribution between the months (Figures 3 and 4). December 2015, and October and November 2017 had no data, which was attributed to a technical error within the satellite age file causing inaccurate assignments of non-current labels to current observations (D Wright 2022, personal communication). Nonetheless, since date was not an input variable for the model, the monthly gap in one year can be made up by the monthly data available in the other years, represented through the day of year and day of cycle variables. Figures 5 and 6 showed that there were disproportionately higher curing observations of 100 percent than any other values below, likely caused by the higher proportion of data in summer months (December - February) compared to spring months (September - November) (Figures 3 and 4). Again, whilst the skewed distribution is not ideal, it is represented in both the training and testing data, meaning both datasets are representative.

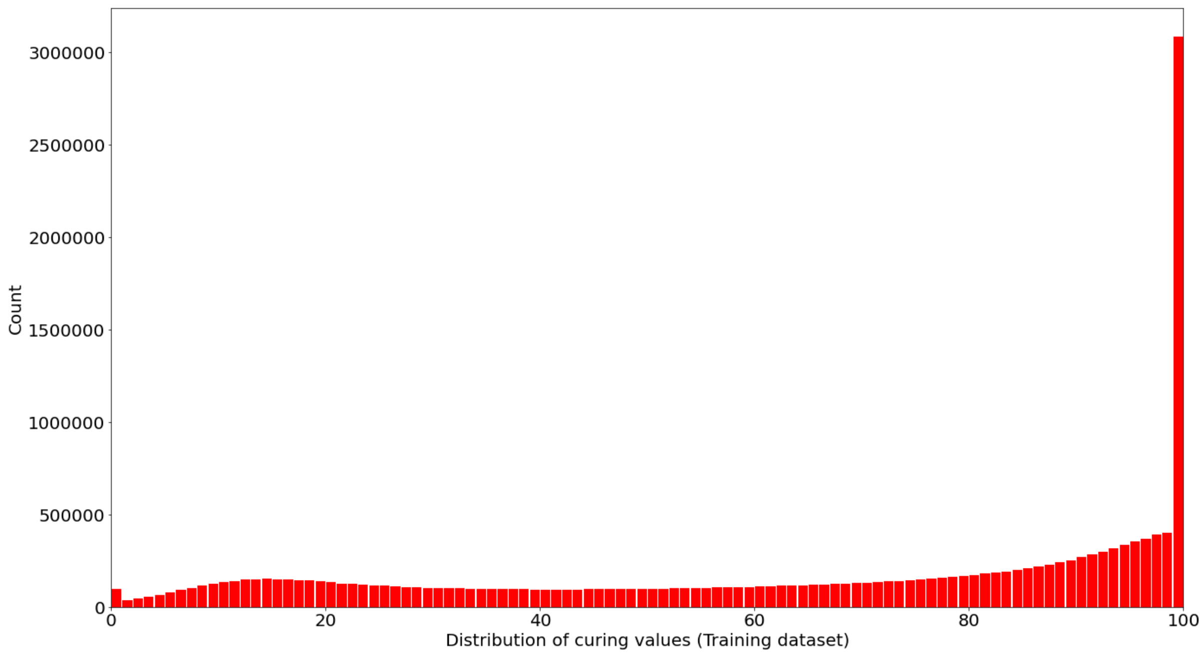


Figure 5: Distribution of curing values in the training dataset over the study time frame.

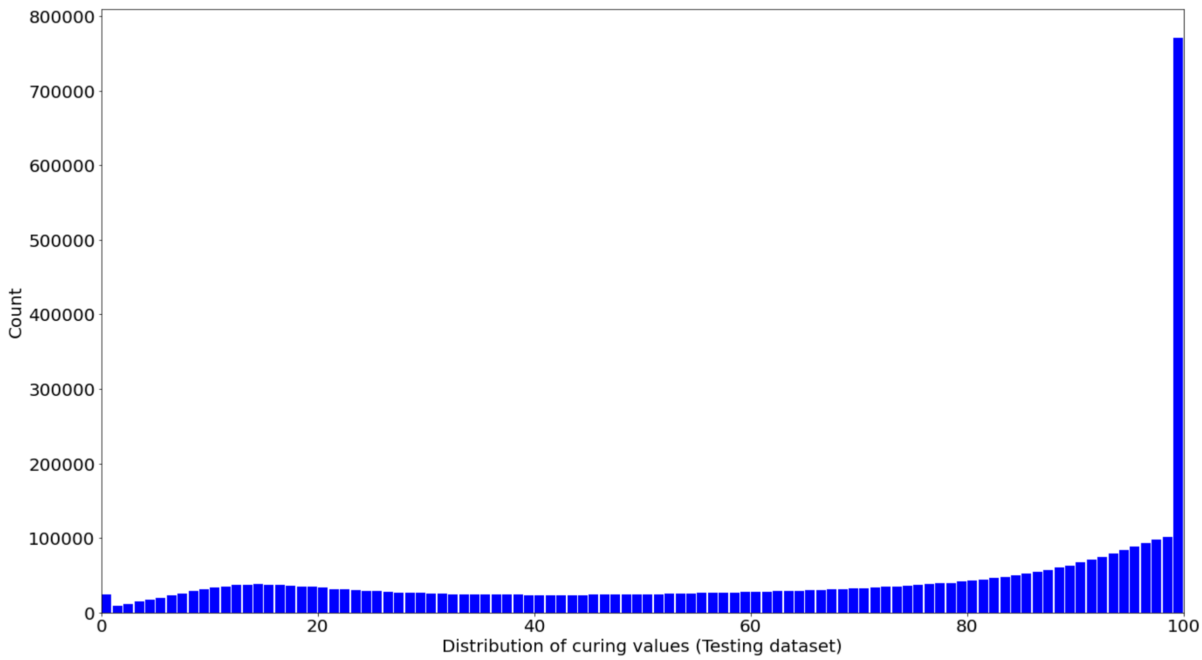


Figure 6: Distribution of curing values in the testing dataset over the study time frame.

2.4. Research Methodology

2.4.1. Objective 1: Build, Optimise & Compare

Objective one aimed to build, optimise, and compare different compositions of a four-day grass curing forecast model and determine which model best predicts the curing process. To do so, Python's existing XGBoost packages were used to build three models, each with a different variable composition. The first model contained both meteorological and curing variables, the second one contained only meteorological variables and the third one contained just curing variables.

While it is possible to simply build a model with all variables (including curing) to predict curing because of its strong and relatively consistent growth cycle, additional information can be obtained by testing the types of variables individually. By excluding curing-related factors, more can be understood about the power of other meteorological and environmental variables in representing the grassland curing process. In contrast, if the model that only uses short-term historical curing values can accurately predict current curing values, the simplicity of the input features is advantageous for its deployment for operation.

The optimisation of the model was done by tuning the hyperparameters. Hyperparameters affect the decision-making of the algorithm and are required to be specified prior to each training instance (Tran et al., 2020). This process is outlined below.

2.4.1.1 Hyperparameters selection

While there are numerous XGBoost hyperparameters that can be optimised, only a subset of them, those which help prevent overfitting the model, were selected for this project (Table 1).

Table 1: Description, default setting and tested range of XGBoost hyperparameters.

Hyperparameter	Description (Cho et al. 2021)	Default	Tested range
learning_rate	- Step size shrinkage used in update to prevents overfitting - Smaller values mean more trees must be built for the model	0.3	[0.018, 1]
min_child_weight	- Minimum sum of instance weight (hessian) needed in a child - Higher values make the algorithm more conservative	1	[0, 10]
max_depth	- Maximum depth of a tree - Increasing this value increases the chance of overfitting	6	[5, 65]
gamma	- Minimum loss reduction required to make a further partition on a leaf node of the tree - If the difference between the gain and gamma is negative, the branch is removed and vice versa - Higher values make the algorithm more conservative	0	[1, 20]
subsample	- The fraction of observations to be randomly samples for each tree. - Lower values make the algorithm more conservative and prevents overfitting but too small values might lead to under-fitting	1	[0.01, 0.99]
colsample_bytree	- The fraction of columns (or the number of features) to be randomly samples for each tree.	1	[0, 1]
reg_lambda	- L2 regularisation term on weights (analogous to Ridge regression) - Higher values make the algorithm more conservative	1	[0, 100]
reg_alpha	- L1 regularisation term on weight (analogous to Lasso regression) - Higher values make the algorithm more conservative	0	[0, 100]

The top six of these hyperparameters were chosen based on standard practice in XGBoost model building (Yan et al., 2022, Carmona et al., 2022, Shi et al., 2021, Srinivas and Katarya, 2022). In addition, two hyperparameters related to regularisation, reg_alpha and reg_lambda (explained further in Appendix vi), were included specifically to address the risk of introducing noise by having a large number of features. Regularisation shrinks the regression coefficient of each feature, or in other words, the influence each feature has on the final prediction, simplifying the model. Tuning the regularisation hyperparameters improves the overall model stability and consistency (James et al. 2013). Since the model with just curing only contained nine features, these regularisation hyperparameters were not tuned because it can lead to oversimplification.

2.4.1.2 Hyperparameter tuning method

The hyperparameters were tuned using Hyperopt, a Python library based on the fundamentals of Bayesian optimisation (Bergstra et al., 2013). The advantages of using Bayesian optimisation were, first, it allowed a discrete or continuous distribution of values for each hyperparameter and second, starting with some random combinations of values, it iteratively updated and restricted the search based on the "expected improvement" of the previous combination (Wu et al., 2019, Feurer and Hutter, 2019). As a result, it is more efficient and effective at improving model accuracy compared to traditional tuning methods (Wu et al., 2019). The set-up of Hyperopt is detailed in Table 2.

Table 2: The Bayesian optimisation setup as per Python's Hyperopt library.

Component	**Description**	**Values**
Objective function (`fn`)	The metric to be minimised at every iteration	Median Absolute Error (Section 2.4.1.3)
Search space (`space`)	Range of values for each hyperparameter	Outlined in Table 1
Tuning algorithm (`algo`)	Two main hyperparameter search algorithm: ● Random search ● Tree of Parzen Estimators	Tree of Parzen Estimators
Evaluations (`max_evals`)	Number of hyperparameter settings to try (the number of models to fit)	45

The values in Table 2 were determined predominantly by common practice (Databricks, 2022). For all three tested models, the optimised value for each hyperparameter was presented to discuss any particular setting that could have potentially led to overfitting the model. The optimised value for each hyperparameter is presented in Table 3, and further discussed in Appendix vii.

Table 3: Summary of the optimised hyperparameters for each grass curing model.

Hyperparameter	**Model with curing and meteorological features**	**Model with only meteorological features**	**Model with only curing features**
colsample_bytree	0.437	0.218	0.304
gamma	5	5	11
learning rate	0.055	0.679	0.671
max_depth	25	20	30
min_child_weight	0	5	5
reg_alpha	25.820	77.475	N/A
reg_lambda	34.082	76.885	N/A
sub_sample	0.697	0.635	0.581

2.4.1.3. Model evaluation

To evaluate the model accuracy, three standard statistical metrics were used: Mean Bias Error (MBE), Root Mean Squared Error (RMSE) and Median Absolute Error (MAE). MBE and RMSE respectively represent the mean and standard deviation of the differences between the actual and predicted output, making them prone to skewness if outliers are present. On the other hand, because MAE takes the median of all absolute differences between the target and prediction, it is insensitive to outliers (Scikit-Learn, 2021) and gives a more robust understanding of the model error (Willmott and Matsuura, 2005). MBE, RMSE and MAE are calculated using equation 9 to11:

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(predicted_i - actual_i) \text{ (Equation 9)}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(predicted_i - actual_i)^2} \text{ (Equation 10)}$$

$$MAE = median(|predicted_i - actual_i|) \text{ (Equation 11)}$$

where n is the number of instances, $actual_i$ is the actual curing value at the i^{th} instance and $predicted_i$ is the corresponding predicted curing value.

The results of each model were compared to determine which is predicting curing most accurately, and is therefore the best performing model.

2.4.2 Objective 2: Feature Importance

This objective evaluated the importance of the input features on the model learning and explored their influence on the grass curing process by extracting the internal feature importance scores, or the "gain" scores. Derived from the work of Friedman & Meulman (2003) and explained by Elith et al. (2008), the "gain" score is "based on the number of times a feature is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees". The sum of all the features' "gain" scores adds up to 100%, implying that for each feature, the higher the "gain", the more important it is for generating a prediction.

The results presented for objective two included the feature importance summary for each of the three tested models.

2.4.3 Objective 3: Spatial and Temporal Variation

The focus of objective three was to investigate the spatial and temporal variation in the observed and predicted curing across Victoria and to assess the best performing models ability to accurately capture the progression of the curing. By analysing plots of observed curing, predicted curing and MBE across Victoria, this approach highlighted the particular periods of the year and of the curing season or particular regions in Victoria that the model was not able to predict as accurately.

2.4.4 Objective 4: Case Study

A case study of the fire incident in the Scotsburn region (Ballarat) between November and December 2015 was conducted to validate the best performing curing model, as determined by objective one. For further details on how the case study was conducted see Section 6.2.

3-6 Results and Discussion

3. Objective 1: Build, optimise and compare models

The first objective of this project was to compare models of different compositions to determine which version would be most suitable for a four day curing forecast. Models with both curing and meteorological features, with only meteorological features, and with only curing features were build and optimised and analysed. The results are discussed in terms of both the statistical accuracy of the model, as well as the suitability of the model for operation within the CFA.

3.1 Metrics

Figure 7 shows the statistical metrics (as described in Section 2.4.1.3), along with the distribution of the difference between the predicted results and the observed curing for each model. The model that includes all variables has both the lowest MAE (2), which suggests it has the smallest average difference between its predicted results and the observed curing, and the lowest RMSE (5.48), suggesting it has the least variation between these differences (Figure 7a).

The negative MBE (-0.29) also indicates that the model tended to slightly underpredict (Figure 7a). However, all models exhibited a small MBE, ranging from -0.29 to 0.0036, indicating that all models predict curing with little bias (Figure 7).

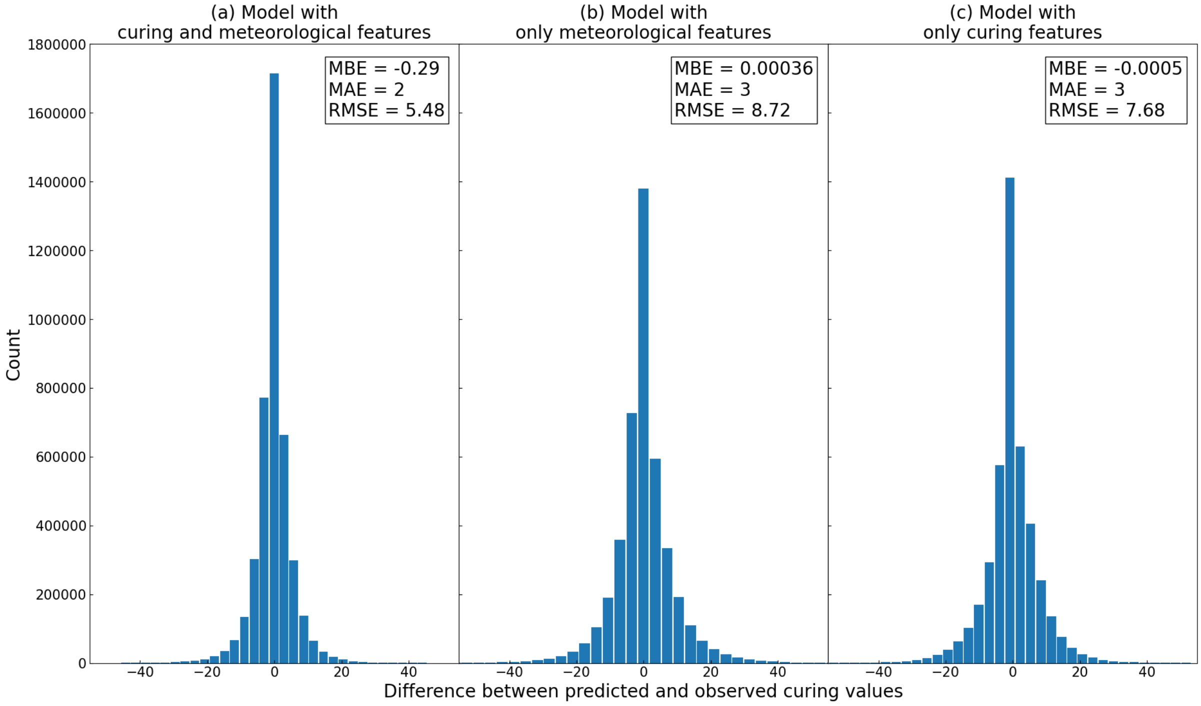


Figure 7: Histogram of the model bias and summary statistics for a) model with curing and meteorological features, b) model with only meteorological features, and c) model with only curing features.

The model with only curing features (Figure 7c) has the smallest bias (MBE of -0.0005) and a similar MAE (3), however its RMSE (7.68) is larger than the model with all variables. This means that whilst the (only curing) model does not under- or over- predict curing, it does not predict curing as accurately as the model with all variables. The model with all variables excluding curing (Figure 7b) had the highest RMSE (8.72) and the equal highest MAE (3), which suggested both a high variability and high average error between predicted values and observed values.

The statistics presented here suggest that all models have very similar average errors, however the model with all variables included has the least variation in that error (smallest RMSE), followed by the model with just curing, and then the model with all variables excluding curing. However, given how similar the results of the statistical testing are, further analysis into the differences between the predicted and observed values may highlight specific areas where each model excels or struggles.

3.2 Model results

The key predictions from the three models were graphed as scatter plots of (a) the predicted curing against the observed curing, and (b) the predicted 4-day curing change and the observed 4-day curing change. The 4-day change in curing represents the difference between the predicted or observed curing value (t0) and the curing value on the day of prediction (t0-4).

3.2.1 Model with both curing and meteorological features

Figure 8a shows that the majority of the data is centred evenly around the y=x line, which illustrates where the predicted values are exactly equal to the observed values. It is also important to consider the log scale on these graphs. This means that whilst there is a fairly large variation in the data overall, the majority is central to the y=x line in both graphs. This distribution of results suggests that the model predicted results close to the observed curing values.

Results that fall above the y=x line are the result of the model underpredicting, whereas results under the line are due to overprediction. The results are evenly distributed above and below the line (Figure 8a) which suggested that the model is not biased to under- or over-predicting. This is supported by the normal distribution of the error and the close-to-zero MBE error (Figure 7a). Additionally, the high density of data that is close to the y=x line at 80-100%, evident in the darker colour, shows that the model is predicting very well at these values.

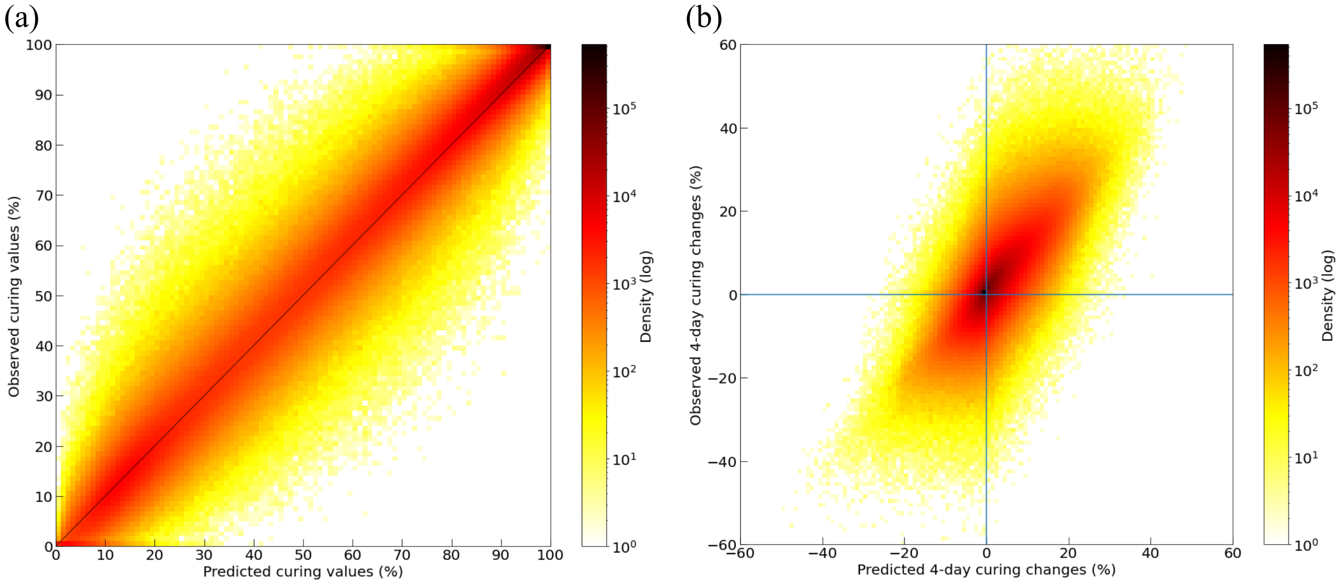


Figure 8: The predicted curing values from the model with curing and meteorological features, plotted as a) a direct comparison of curing values to the observed curing values, and b) a comparison of the predicted 4-day curing change and the observed 4-day curing

Figure 8b depicts the accuracy of the model during differing periods of curing change. Results in the top right quadrant are true positive results – curing increased and the model predicted an increase, whilst results in the bottom right are false positives – the model predicted a curing increase when there was in fact a decrease. Similarly, results in the bottom left are true negatives – both an observed and predicted decrease, and the top left results are a false negative – where the model has predicted a decrease in curing when there was an increase. The results lie largely in the true positive and negative quadrants, suggesting the model has done a good job of predicting how curing changed over the four-day prediction period.

Figure 8b also shows the model predicts best when the degree of curing only changes a small amount in the four day prediction period. Specifically, if there is a 5% decrease to a 10% increase in curing the model performs quite well. The model is cautious in predicting large changes in curing, no matter which direction that change is in. This is evident when comparing the range of predicted and observed curing changes. The predicted results only cover a curing change of -40% to 40%, whilst the observed results show actual changes of almost -60% to 60%. The conservative nature of the model may be due to the fact that there are significantly less data available at these times of significant change, and therefore the model could not learn the patterns that lead to these results as effectively as it learnt about smaller changes.

3.2.2 Model with only meteorological features

When curing features are excluded from the model, there is an increase in the spread of results (Figure 9a and b), consistent with the RMSE increased from 5.48 when all variables are included to 8.72 (Figure 7b). This is likely because without any curing features the model can predict an increase or decrease in curing, but does not have any knowledge of the curing values and the rate of change during that time period. When the curing features are included the model can predict not only that curing has increased, but that it has increased from a given curing value four days prior.

(a) (b)

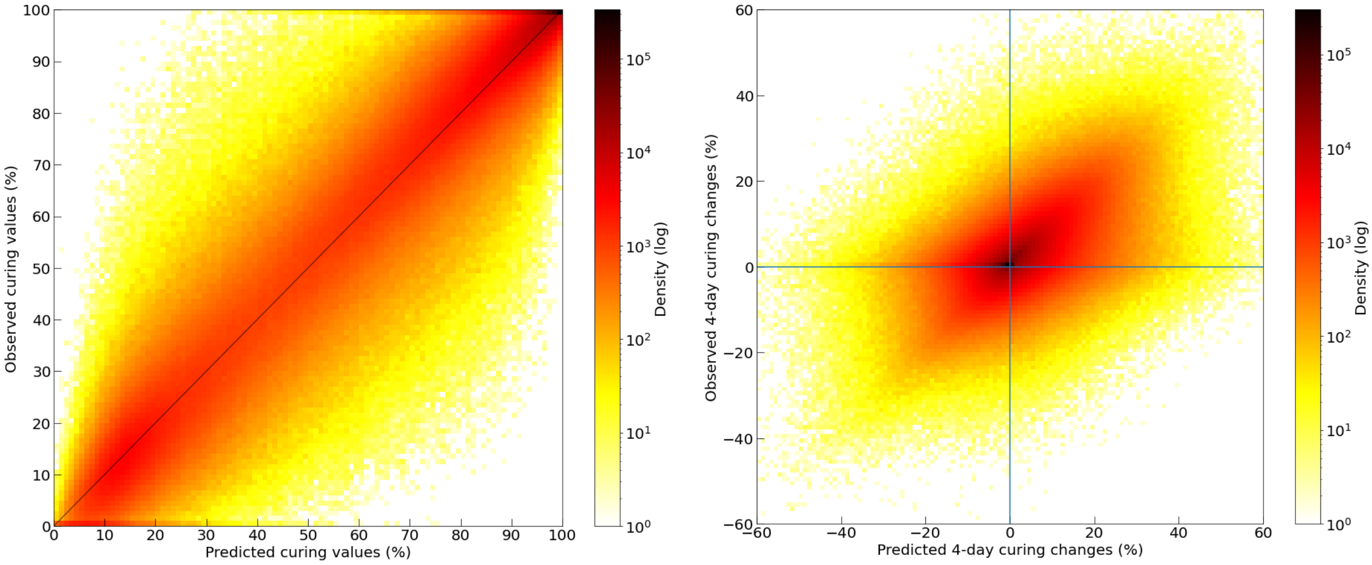


Figure 9: The predicted curing values from the model with all variables apart from curing presented as a) direct curing values compared to the observed curing values, and b) predicted 4-day curing change compared to the observed 4-day curing change.

Overall the statistics for this model do show it predicts well, and so whilst there is more variation in how the model predicts, it still has the potential operational value. A major issue currently faced by the CFA and fire management organisations is the inability to monitor curing through cloud cover. Currently when this occurs it can lead to multiple days of using the same curing value, which does not provide the CFA with any information as to how curing might be changing at the time. A model that does not use curing at all could be used at these times to fill the gaps that occur. It would likely not be as accurate as the normal satellite product, however, would be better than not having new information.

3.2.3 Model with only curing features

When only curing features are used, the model is much more likely to incorrectly predict an increase in curing, as seen by the overwhelming amount of data in the false positive quadrant (bottom right) (Figure 10b). This is likely because a decrease in curing during the ramp up period is a deviation from the natural curing cycle, caused primarily from meteorological conditions (Sullivan et al., 2012). Given none of these variables are included, the model had to try to learn when they would occur purely from the curing trend in the week leading up to the prediction. The model's bias towards predicting a positive curing change was also evident in the predicted curing values themselves, with Figure 10a showing a complete lack of predictions for curing under 5%.

The lower accuracy in predicting a decrease in curing was a surprising result, given that the RMSE and MBE suggested this was the second-best performing model (Section 3.1). However, curing does not commonly decrease throughout the ramp up period, therefore the model's decreased ability to accurately predict during these times does not have a major effect on the statistical overview of the models ability.

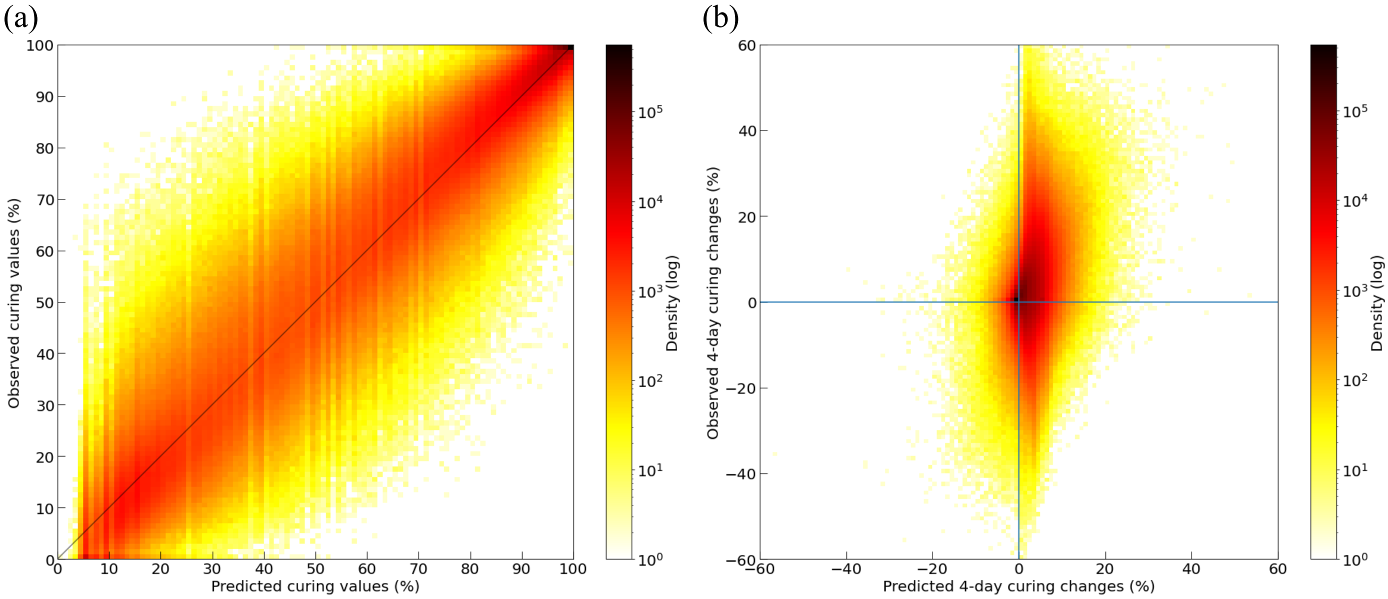


Figure 10: The predicted curing values from the model with only curing presented as a) direct curing values compared to the observed curing values, and b) predicted 4-day curing change compared to the observed 4-day curing change.

Additionally, this model is the most conservative in predicting changes in curing, with predicted changes running from -20% to 30%. This is likely because the majority of 4-day curing changes are relatively small, and without other variables to indicate a larger change will occur, the model will base its prediction on the trend in curing changes it has seen.

An usual feature in the scatter plot of predicted vs observed curing values is the vertical stripes which implies that some specific curing values were less frequently predicted compared to its close neighbouring values. While an exact reason is unknown, a possible explanation could be that given that curing was the only variable influencing the decision, the model is more likely to predict a curing value that it has seen before.

3.3 Model Selection

Both the MBE and MAE of the models suggested that all three generally predicted curing with high accuracy. The model with both curing and meteorological features had the lowest RMSE indicating that this model also had the least variation in the difference between the models predicted curing and the observed curing values. This is supported by the data distribution in Figures 7, which shows that the model with both types of variables (Figure 8a) has the least amount of data spread and is most centred on the y=x line, suggesting it is the most accurate and most precise of the models.

This model was conservative in its estimation of curing change, often predicting a smaller 4-day curing change than what was observed, especially during times of significant change (Figure 8b), however this conservativeness was also seen across all three models. All three models showing the same conservativeness suggests it is a result of how they are learning, not which variables are included.

When only curing variables are used, the model struggles to predict when curing decreases, as in the ramping up period that is being predicted curing decreases are usually due to environmental conditions. Not including curing at all however also leads to issues, as without a baseline indication of where curing is at, the model may know what effect the environment is having, but struggles to determine the actual curing level the change begins and ends at. The model using only curing features does statistically predict better than the one that does not use it all, likely because the frequency of events where curing decreased, and therefore the model predicted less accurately, were low. This highlights that whilst analysing the models through their plotted results is helpful in understanding where the models do struggle, they can often skew the perception of the model, and should not be used in isolation.

Overall, the statistical metrics and the distribution of the predicted results compared to the observed curing values suggest that the model with both meteorological and curing variables is most suitable of the three tested, and the hyperparameters suggest it is likely not under or over fitted to the training data. Hence, it has been determined to be the best performing model of the three, and is referred to as such throughout the report. However, these tests were only an initial analysis, and further investigation was required for a deeper understanding of the model performance. The following three objectives aimed to provide that understanding.

3.4 Objective 1 limitations

Whilst the models were predicting well, there were some limitations to how they were built and analysed. Machine learning models are generally dependent on the quantity and quality of data that they are provided. Given the pre-existing limitations to how curing is currently monitored, the curing dataset provided both potential quantity and quality issues. A significant amount of data had to be removed from the MapVictoria dataset as the curing value recorded was from previous days, leading to gaps in the dataset that the model could not learn from. Additionally, the MapVictoria model is not perfect. There is ~10% error in its curing values, and it underpredicts curing in areas of undetected woody vegetation and overpredicts it in areas with bare soil or water bodies, as discussed in the Introduction (Section 1.3.2) (Martin et al 2015). Using the AFDRS data to subset what is fed to the model has helped reduce the effect that underpredicting in woody areas may have, as the AFDRS fuel definitions have a stricter definition of grassland that excludes more woody vegetation. However, overall, the errors in the MapVictoria model are still propagated by the model, as it considers the curing value given as truth, and so will predict curing with a similar level of error.

Additionally, all the datasets were of different spatial resolutions, with the largest being the soil moisture ESA CCI SM dataset, which has a spatial resolution of 0.25°. While this is not particularly coarse, having datasets that match the curing resolution (0.005°) would reduce the potential for collocation errors and provide more accurate variable information.

In terms of analysing and assessing the best-performing model, machine learning itself has some limitations, primarily that all ML models are "black boxes". Whilst statistical metrics and graphs of the results can be used to determine how the model performs, why a model does or does not perform well can only be hypothesised – not tested or clarified. The rest of the objectives of this report are all aimed at mitigating this risk and do help determine where or when the model may be faltering, however, there is still no way of explaining the cause of the error.

Furthermore, each model was only tested once, and therefore it is unclear whether they are stable and would produce consistent results. The risk of only running the models once was partially mitigated by tuning the hyperparameters and having a separate validation dataset to conduct a case study. However, the lack of deep understanding of interactions between hyperparameters made it challenging to draw solid conclusions about the stability of the models.

3.5 Objective 1 Recommendations

Given these limitations, there are certain steps future researchers could take to improve upon this work. The most important of these is to adopt cross-validation techniques at multiple stages throughout the building and optimising of the machine learning algorithm. Cross-validation runs the step multiple times, using different sections of data each time the data needs to be sectioned, and then averaging the result. For example, during the training and testing stage, the data would be split into 'k' number of subsections and the model run 'k' number of times, each time using a different one of the subsections as the validation dataset. The statistical metrics of each run would be average to determine the overall measure of the model. This allows the model to both train and test on all of the data in the dataset, whilst also testing the stability of the model across different data.

Currently, the MapVictoria dataset is the only daily curing dataset available in Victoria, so the limitations of the data cannot be mitigated. The CFA has however been using this dataset operationally, and so is both aware of these limitations and willing to accept them (D Wright 2022, personal communication). However, datasets tracking water bodies do exist, and therefore could be included in the project design to remove these areas from the data given to the model.

4. Objective 2: Feature Importance

4.1 Feature Importance Summary

The focus of the second objective was to evaluate the contribution of each feature in each of the three curing models, quantified by the “gain” score, which indicates how important a feature is to the model's predictive process, as described in Section 2.4.2.

Figure 11a and 11bshows the descending gain score of the top 20 features of the model with all features including and excluding curing, respectively. Figure 11c illustrates similar results for the model with just curing features. When both meteorological and curing variables are included in the model, the curing features were most important for predicting current curing (Figure 11a). This is not surprising, as they provide a good indication of roughly where the predicted curing value should be (as discussed in Section 3.2.2). Interestingly, the model with both curing and meteorological variables had three curing features with a feature importance over 10% (curing_4day, curing_avg_4day, curing_avg_7day) whilst the model with only curing features had one observation that was overwhelmingly important (97%) (curing_6day). The reliance on just one curing feature when the other model used a combination is unclear.

Aside from the curing features, the weekly average of VPD and maximum temperature were considered the most important in the model with both curing and meteorological features (Figure 11a). In the model with only meteorological features, the weekly average maximum temperature and soil moisture were most important (Figure 11b).

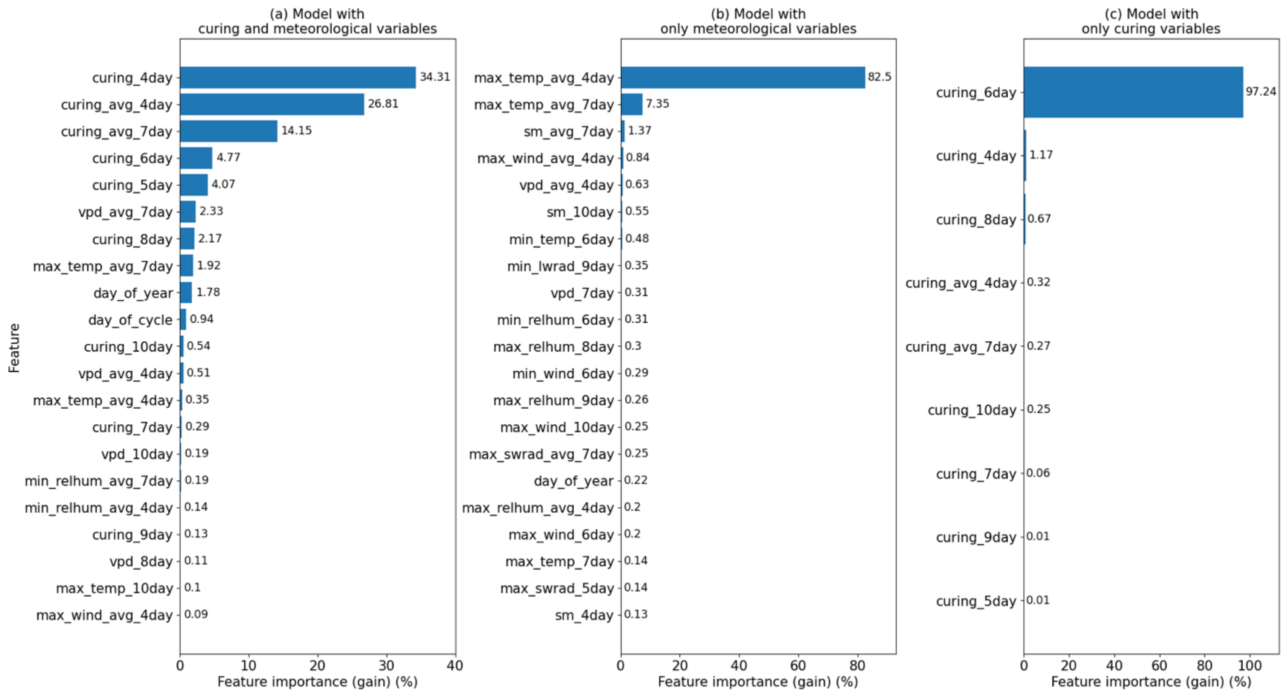


Figure 11: Feature importance of a) model with all curing and meteorological features, b) model with just meteorological features, and c) model with just curing features. Each daily observation was attatched with the "_xday" suffix, where x is the number of days prior to the current day (t_0). Each average wa attached with the "_avg_yday" suffix, where y is the number of the days prior to prediction (t_0-4).

It is commonly assumed that highly correlated variables would not both be considered important by the model, as they would likely show a similar relationship to curing. Given that VPD is calculated using maximum temperature and minimum relative humidity, it is highly correlated to both variables (see Appendix viii), and therefore it was surprising that both VPD and maximum temperature were deemed important by the model. This suggests that they may be providing slightly different information, or, it also may be because the model could sample either and therefore a combination was used. Further testing as to whether the model can predict as accurately with each variable independent of others would provide more in depth understanding to the importance of each one.

Interestingly, soil moisture was the second most important variable for the model with only meteorological variables (Figure 11b), but was not in the top 20 for the model with both curing and meteorological variables (Figure11b). This may be because the two variables are closely correlated (Appendix viii), and therefore the information provided by including soil moisture is already captured in the trends of the curing data. As such the model can use curing to make the same decision that it would use soil moisture for. However, given the model is a “black box”, it is not clear if this is the case.

In both of the models with meteorological features a temporal indicator, such as day of year, was included in the top 20 features (Figure 11a and b). This suggests that the model relies on being told what time of year it is to make an accurate prediction, and this information is not being conveyed by the other variables. Conversely, the coordinates of each pixel were not deemed important by the model. This indicated that any spatial trends are being adequately covered by the variables themselves.

Given no short-term curing forecast existed yet, there was little operational evidence to guide how far back in time should be considered for each variable. A weekly average, four day average, and daily observations for the week leading up to the prediction were chosen based on data availability, however it is highly likely that seven daily observations and two averages are not needed for each variable. These results show that for the meteorological variables, the seven day average is almost always the most important feature. This is likely because meteorological variables tend to fluctuate quickly between days, and the seven day average is more representative of the variable than the four day average or daily observations. Additionally, it is unknown whether a longer average, such as 10 days or two weeks, would be more beneficial to the model. Further research should aim to include and test the importance of a longer average to see if there is a particular scale with peak importance.

Precipitation was originally expected to be fairly important to the model, as it has a substantial impact on water availability and therefore the curing cycle (Fahad et al., 2017, Will et al., 2013). However, it was not deemed important by either model that included meteorological variables. This may be because water availability has been suitably encompassed by the other variables, but more likely, using the daily rainfall amount did not adequately portray the effect rainfall has on the curing process. By using the daily precipitation amount, prolonged periods of no rain are all just signified by 0 mm rainfall - the ongoing effect is not considered. If instead precipitation was measured using days since last rainfall, longer stretches of time without rain would be differentiated from one or two days without it. It is likely that this kind of precipitation measurement would be deemed more important to the model, as there would most likely be an identifiable relationship between it and curing.

Overall, these results show that there are many features that are not considered important to the model. The most important features to include are; average curing, a daily curing observation, average meteorological variables, and a temporal indicator, such as day of year. Additionally, variables that are highly correlated may not both be needed, however they should be tested systematically before being removed.

4.2 Objective 2 Limitations

The major limitation inherent to objective two is that the 'gain' score used to determine feature importance only considers how often that feature was used to make a decision in one of the model's decision trees. It does not tell us how or when the feature is used to make that decision, only that it was used. There is currently no way to determine how that feature was used, or if that feature is important to the actual curing process using ML or ML analysis tools.

4.3 Objective 2 Recommendations

To build upon objective two's findings, it is recommended that at least one other measure of feature importance is used and the results compared. This would either provide further support to the conclusions drawn or highlight any differences in why particular features were deemed important.

More importantly, the results of this section inspired many potential improvements to the methodology of this project. It highlighted that many of the included features had very little impact on the model, and therefore there are possible computational benefits from reducing the number of features included. It is recommended that further optimisation of the models is done by removing features deemed less important and investigating the overall effect on the model, or by researching methods such as the tuning of the hyperparameters to determine if there is an existing method that would lead to better results. When removing features, the first step should consider how many temporal scales of each variable is needed, as an average observation and potentially one daily observation should suffice. Further testing may however be needed to determine which daily observation is most beneficial. Specifically, it is also recommended that either VPD or maximum temperature and minimum relative humidity are systematically removed and the model retested, due to their high correlation to each other.

Additionally, it was hypothesised that including precipitation in ‘days since last rainfall’ instead of the number of millimetres that day would have a greater impact on the model. The model should be rerun with this new precipitation measure included to see whether it improves the accuracy of the model, or changes the feature importance of precipitation.

5. Objective 3: Spatial and Temporal Variation

The third objective investigated the spatial and temporal error of the best-performing model, as determined in objective one. This identified potential locations and months of high disagreement between the modelled and observed curing conditions.

5.1. Curing values

Figure 12 and 13 illustrated the monthly average of the observed and predicted curing values (%) over the study timeframe across Victoria. Figures 12 and 13 were visually indistinguishable which indicated that the predicted curing values were close to the observed curing values. Furthermore, the results demonstrated that the model was capable of capturing the year-to-year differences in the curing progression, especially during the spring months (September, October, November) of 2015 and 2016. In September 2016, both figures show the similar beginning of the curing process in the northwest and the gradual southward spread in October and November 2016. Even in 2015 when there was less of a spatial trend to the curing onset, the model accurately predicted the grassland areas that cured first, which were in the northwestern, southwestern and central parts of the state (Figure 12). In addition, the predicted areas of fully cured grass in summer months (January, February) were visually indistinguishable from the observed curing values.

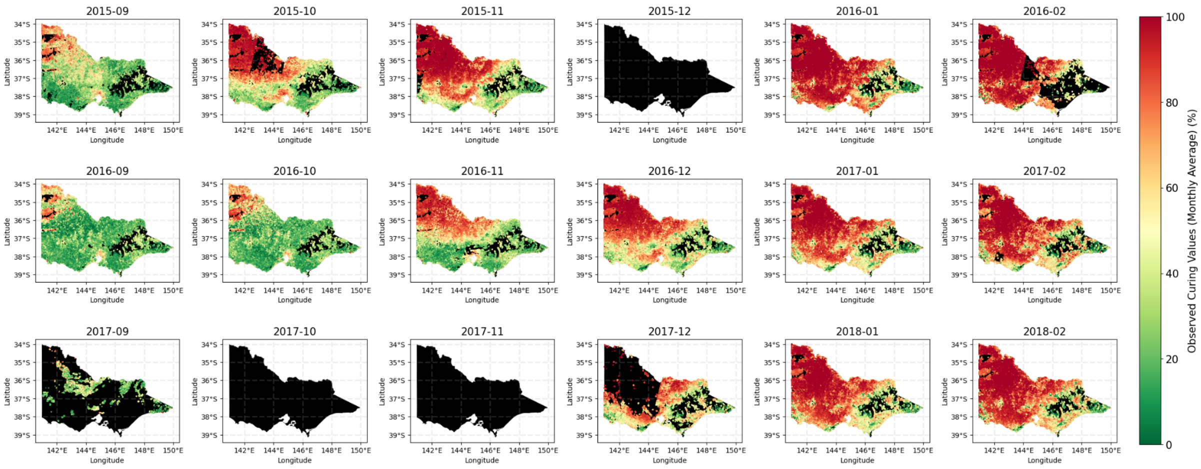


Figure 12: The monthly average of the observed curing values (%) during Spring and Summer months across Victoria from 2015 to 2018. Black portions of the map indicate missing data.

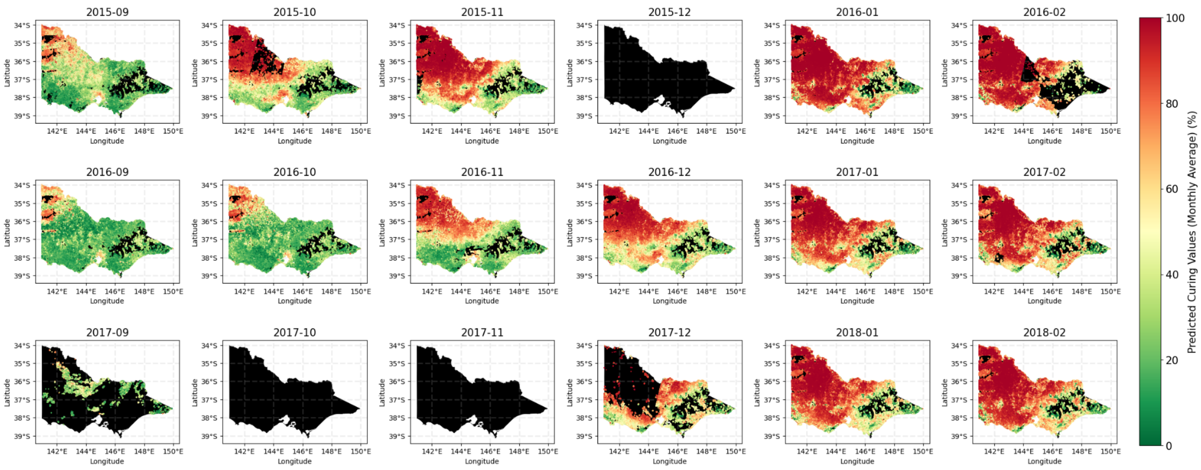


Figure 13: The monthly average of the predicted curing values (%) during spring and summer months across Victoria from 2015 to 2018. Black portions of the map indicate missing data.

Considering the high level of similarity between the predicted and actual observations, the model demonstrated its ability to accurately capture the different temporal and spatial variation and predict the areas of fully cured grass. Its capacity to predict the detailed process can be related to the use of the curing observations and averages of days before the day of prediction. If the curing process started early or accelerated more quickly in one year, the short-term curing variables would inform the model about such variations. Therefore, including short-term past curing measurements provided the benchmark and is important for predicting the current curing values.

Although the ramp up period was defined to be from September to February, Figure 12 shows that this did not capture the true beginning of the period. This is particularly evident in September 2015, where the northwest region was generally already over 50% cured. Therefore, the exact moment at which curing began was not included in the temporal research scope. When the curing ramp up period begins has an effect on how quickly grass across the entire state can cure, as seen by the faster state-wide rate of curing in 2015 compared to 2016. These results show that future research should conduct further analysis into when the curing ramp up period begins, to ensure it captures the entire process. This would likely involve extending the temporal scope for a few more months, and including late winter.

5.2 4-day curing value change

Figure 14 and 15 illustrate the monthly average of the observed and predicted 4-day curing value change over the study timeframe across Victoria. To the eye these figures are very similar, however slight differences are visible, meaning that the model predicted the day curing relatively close to the actual observations.

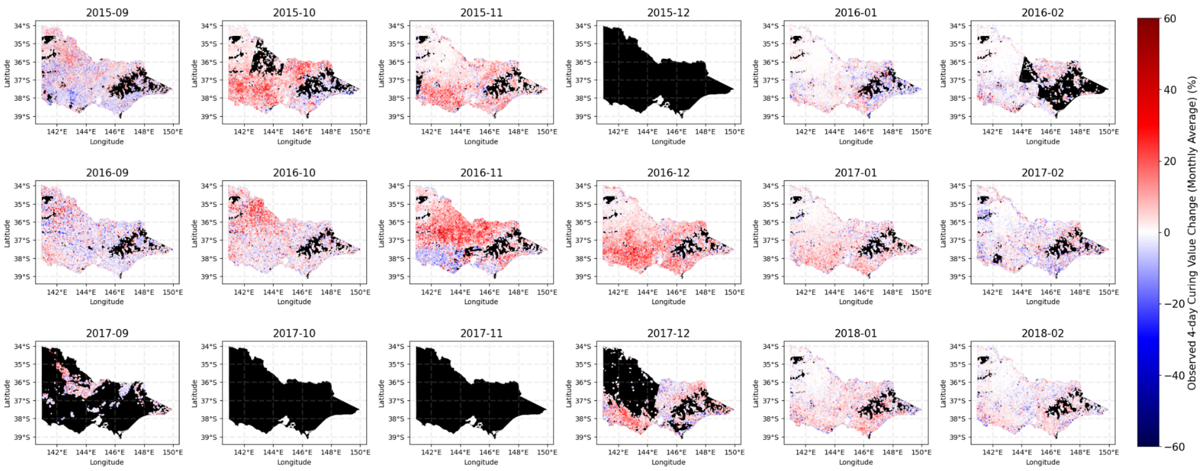


Figure 14: The monthly average of the observed 4-day curing value change (%) during spring and summer months across Victoria from 2015 to 2018. Black portions of the map indicate missing data.

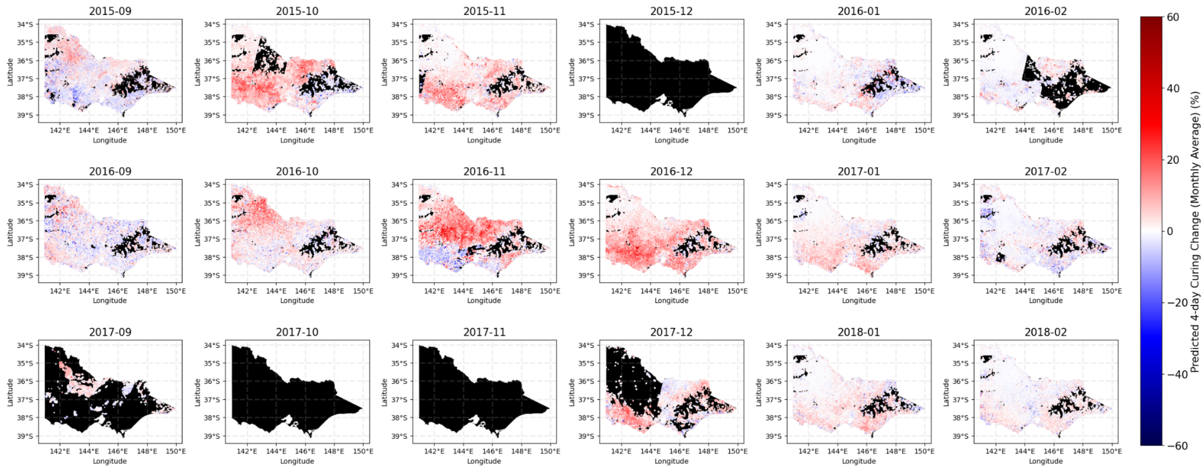


Figure 15: The monthly average of the predicted 4-day curing value change (%) during spring and summer months across Victoria from 2015 to 2018. Black portions of the map indicate missing data.

The model accurately predicted all the locations across the state experiencing the curing value increase, illustrated in the similar red areas, and the curing value decrease, illustrated in the similar blue areas of both figures. The value of the predicted 4-day curing change was consistently slightly lower than the obseved 4-day change, illustrated visually by the lighter shading of all maps in Figure 14 compared to the maps in Figure 15. This suggested that the model underestimated the true magnitude of the four day curing change across all of Victoria, which was similar to the finding from objective one (Section 3.2.1). It should be noted that the minimal colour indifference between Figure 14 and 15 implied that the true extent of the model underprediction of 4-day curing value change was not substantial. However, further investigation is warranted to evaluate the implication of a model that consistently underestimated the observed 4-day curing change in an operational context.

5.3 Model's Mean Bias Error (MBE)

Figure 16 presented the monthly average MBE for the study time frame. The MBE plot (Figure 16) reflects the difference between the predicted and observed curing values (Figure 12 and 13). This therefore also reflects the difference between the predicted and actual 4-day curing value change (Figure 14 and 15).

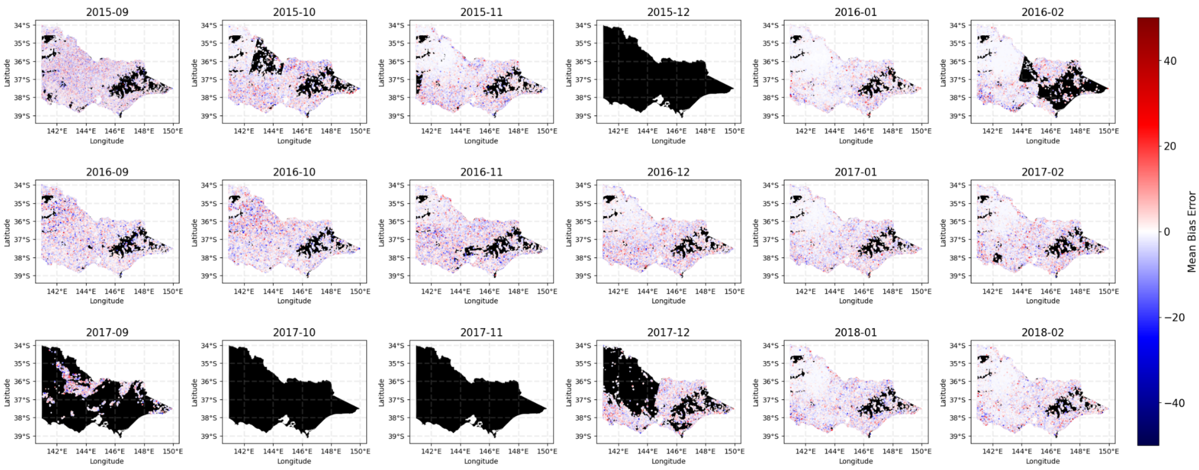


Figure 16: The monthly average Mean Bias Error (MBE) for the spring and summer months across Victoria from 2015 to 2018. Black portions of the map indicate missing data.

The MBE results were consistently close to zero in the northwest region of the January and February of all three years. These areas correlated with the locations of fully cured grass (Figure 12), which usually had minimal average 4-day curing value changes (Figure 14). Therefore, the model possessed good predictive skills for very high (~ above 90%) curing values. The enhanced skill for predicting extremely high curing values could stem from the fact that the distribution of input curing values was skewed towards 100% (Figure 5 and 6 from methods), supported further by the missing data of the spring months containing lower curing values. Alternatively, since 1005 is the maximum value the model can predict, once a region in Victoria reached that maximum level, the model could almost exclusively and easily predict 100%.

For areas with green and partially cured grass, there was a mix of negative and positive MBE values with no particular spatial trend, as seen in September 2015 and

September and October 2016 (Figure 12 and 16). The 10-20 MBE results in these areas suggested that the model is not as skilled when predicting curing values below 100%. However, the evenly mixed distribution of both positive and negative MBE suggested that there was no specific spatial bias in these areas. Considering the objective two finding that longitude and latitude features were not highly regarded by the model, the spatially comprehensiveness of the model could come from the inclusion of short-term past curing measurements which acted as a direct guideline for the current prediction.

Overall, the visualisation of MBE showed that the model had sufficient predictive skills, with higher consistency and accuracy for very high (~above 90) curing percentages than lower percentages, assumed to be partly due to the skewed distribution of the input curing values.

5.4 Limitations

Three important limitations of the study design were accentuated by the results and discussion of the third objective. The first limitation was the missing data within the September to December intervals, predominantly from months with lower curing values. The second limitation was the restricted data timeframe from the start of spring to the end of summer, as this time frame did not capture the beginning of the curing ramp-up period. Consequently, the first two limitations exposed the best-performing machine learning model to more data containing higher curing values than low values. This exposure likely explained the model's ability to better predict curing values at 100% and close to 100%.

Finally, the third limitation was only assessing the spatial and temporal bias of the best-performing model using the MBE. Given the findings of objective one indicates there is variation in the model's predictive error, the next step would be to identify where and when the variation occurred. This insight would allow for the determination of if further action is necessary to adjust the model if needed to ensure it makes the best possible predictions when given new data.

5.5 Recommendations

There are two actionable recommendations, given the findings and limitations of objective three. The first recommendation is to include more data for lower curing ranges. Unfortunately, due to the limitation of the MapVictoria model (as discussed 1.3.2) there will always be missing data, however, the temporal scope of future research could be extended. By beginning the analysis earlier in the year, it would ensure the onset of the curing ramp-up period is captured. This action would allow the model to learn the conditions that trigger the onset of curing. Additionally, analysing these months may provide additional operational benefits as it may determine if the onset of curing has shifted spatially across Victoria over time.

The second recommendation is to measure the spatial and temporal variance of the best-performing model by mapping MAE and RMSE in the same way MBE was mapped (Figure 16). Although the overall MAE and RMSE were evaluated for the best-performing model across Victoria as a whole, the evaluation could have been enhanced by temporally mapping MAE and RMSE to gain a better idea of where the greatest variation occurred in the model predictions.

6. Objective 4: Case Study

6.1. Background

The Scotsburn grass fire incident was selected for the case study due to the area's grassland-dominated landscape. Scotsburn is located northwest of central Victoria, approximately 16 km from the city of Ballarat. The dominant fuel type across Scotsburn is grassland, with a small area of dry eucalypt forest bordering the southern and western borders of the Scotsburn locality (Figure 17 and Figure 18).

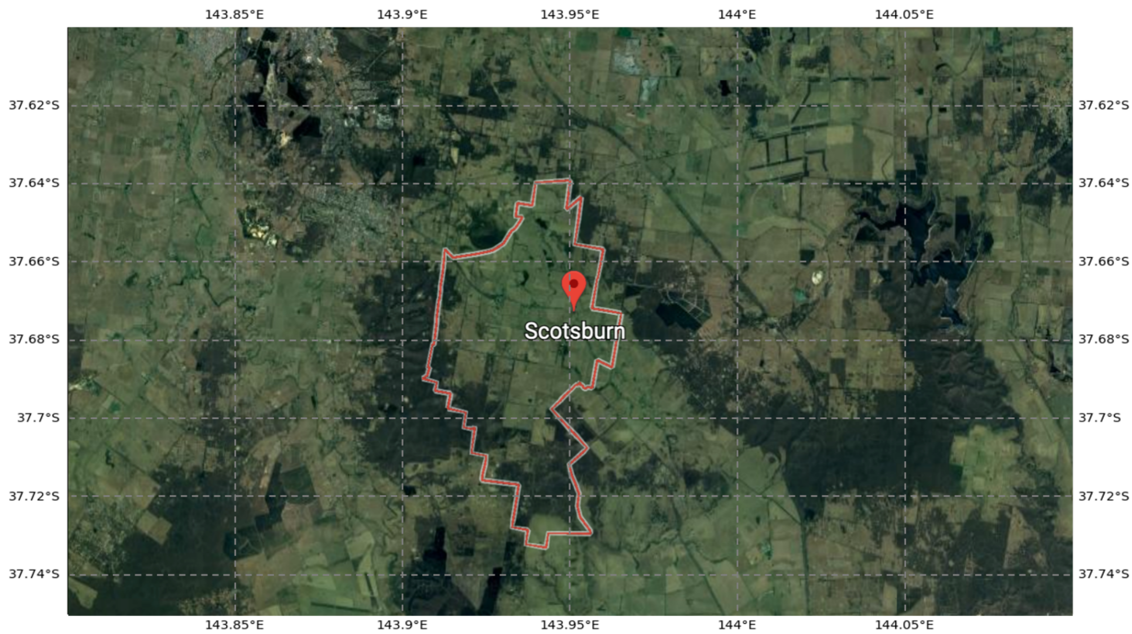
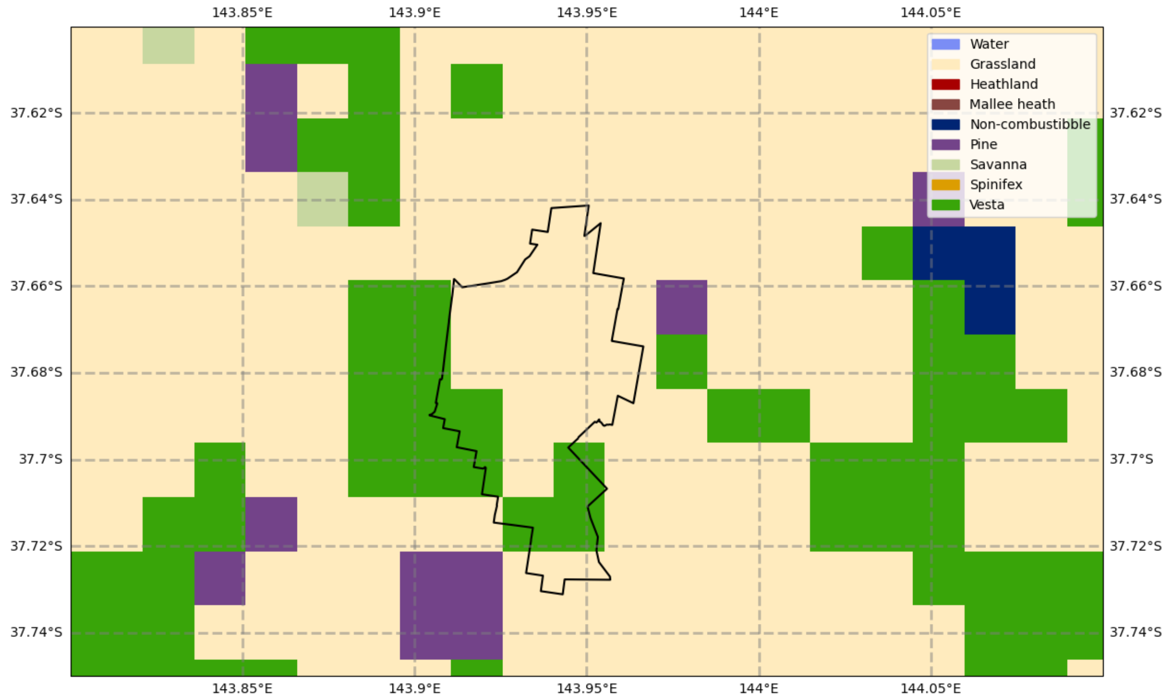


Figure 17: A 2019 Google Earth satellite image of the Scotsburn area outlined in red.



.

Figure 18: The AFDRS fuel type map for the Scotsburn area bounded in black.

Scotsburn grassfire ignited on the 19th of December in 2015 at Finns Road on a day of extreme fire danger. The Fire Danger Index was at the second highest level, Extreme, with values between 75 to 99. The ignition source of the fire was sparks created by farm machinery operation (EMV, 2016). The Scotsburn local who witnessed the grassfire described it as "the quickest thing I have ever seen," referring to the ignition and rate that the fire spread (Wrigley, 2019). The rapid spread of the grassfire was attributed to the 2015-16 summer being at the time the hottest on record in combination with dry fuel loads, consistently high temperatures, and strong wind speeds. A total of 50 fire crews and four air tankers responded to the incident, which lasted eight days and burned approximately 46 km^2 of agricultural land in Scotsburn and parts of neighbouring townships (Calligeros et al., 2015). The fire was declared ‘All Clear’ on the 1st of January 2016.

The purpose of the case study was to generate an in-depth, multi-faceted understanding of the best-performing curing model in the context of an authentic grass fire incident. It provided the opportunity to validate the model on a dataset it had not seen and allowed for a contextual understanding of the features most important to the model. Overall, it helped to understand if by forecasting grassland curing, communities like Scotsburn, and fire authorities like the CFA, could be better prepared in the event of extreme grass fire danger (EMV, 2016).

6.2 Methodology

The case study included the days between November 1st to December 31st, 2015, for the Scotsburn area and captured the conditions leading up to the grassfire incident. The validating dataset was created using the Methodology detailed in Section 2.3.4. Non-grass pixels and cloud or non-current curing observations were excluded from the validating dataset. The elimination criteria reduced the case study timeframe within November and December in 2015 to seven days of eligible curing observations. The seven day are; the 16th, 28th and 30th of November and the 4th, 13th, 16th and 18th of December. The validating dataset was isolated from the training and testing datasets, meaning the model had not seen this data before. The best-performing model that contained curing and meteorological features was presented with the validating dataset to predict a set of curing values.

6.3 Results and Discussion

The figures from objectives one and three were recreated using the validation dataset, to evaluate the performance of the model given new data. A time series of the observed average curing values and variables important to either the model or the Scotsburn grassfire were analysed to provide potential correlations between the variables themselves as well as the model feature importance.

6.3.1 Curing values

The predicted and observed curing values both ranged from approximately 25% to 90% (Figure 19). Figure 19 shows that the model predicted the observed conditions fairly accurately, as most data were close to the diagonal line y=x, however, the majority sat just above it. This indicated a slight tendency to underpredict. Despite this slight predictive underestimation, the findings suggest that the model was able to make reasonable predictions using validating data, given the data was new to the model.

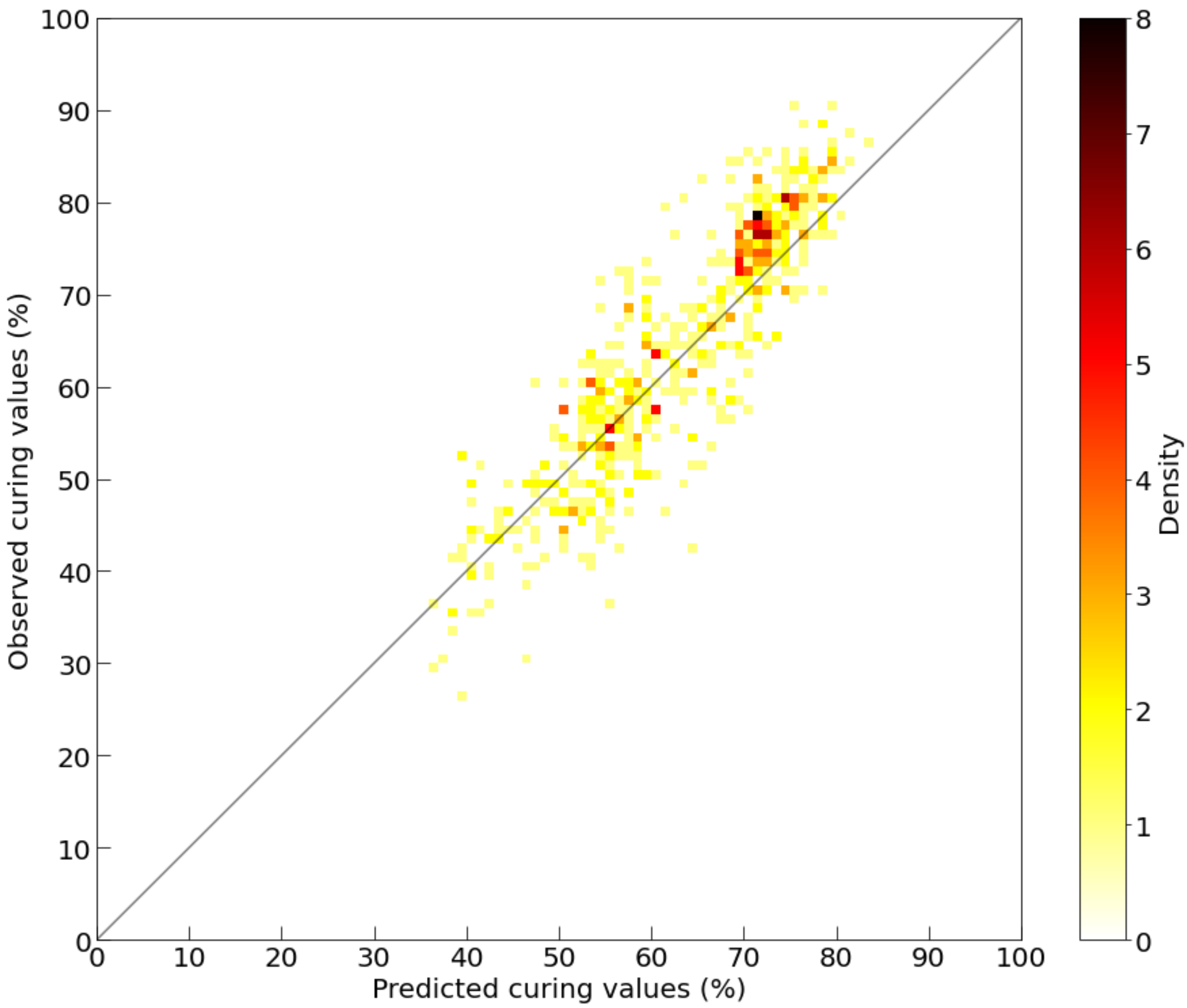


Figure 19: Predicted curing values (%) versus observed curing values (%) over seven days during November and December in the year 2015 across Scotsburn. The scale bar is indicative of the data point density with low density in yellow to high density in black.

As shown by the daily spatial plots, the model underpredicted the daily average observed curing values in the majority of cases by 1% to 5% (December 4th was disregarded due to low data) (Figure 20 and 21). The two days when the model overpredicted the actual curing values were November 16th and 28th, both with some curing values below 50%. This suggested that the model can slightly overpredict low curing values but underpredict high curing values. If the model was implemented in operation, fire authorities may find this slight inaccuracy acceptable given the small difference ranging between 1% to 5%. Overall, this analysis provided some evidence to suggest that the chosen model did have predictive skills when faced with data it had not seen before, however, it should be tested using more validating data to determine if its tendencies are consistent.

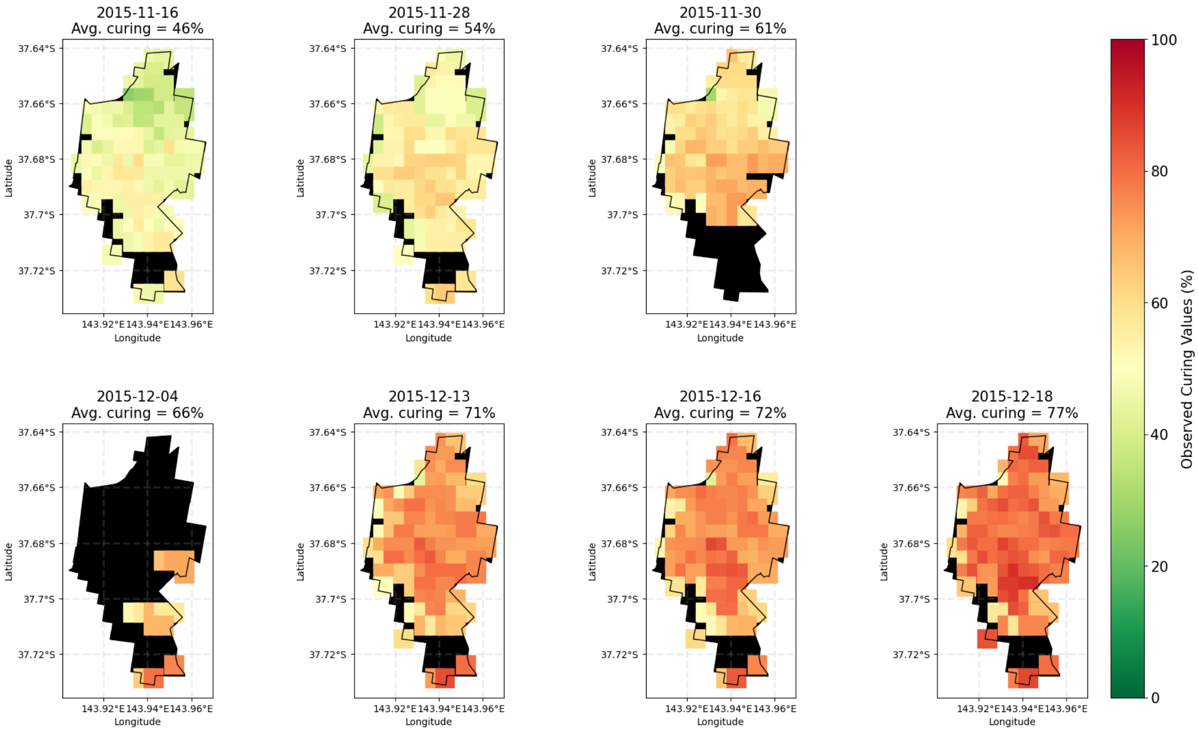


Figure 20: The observed curing values (%) of the best-performing model for seven days of available data across Scotsburn during November and December in the year 2015. The scale bar indicative of the curing percentage with low values illustrated in green and high values in red.

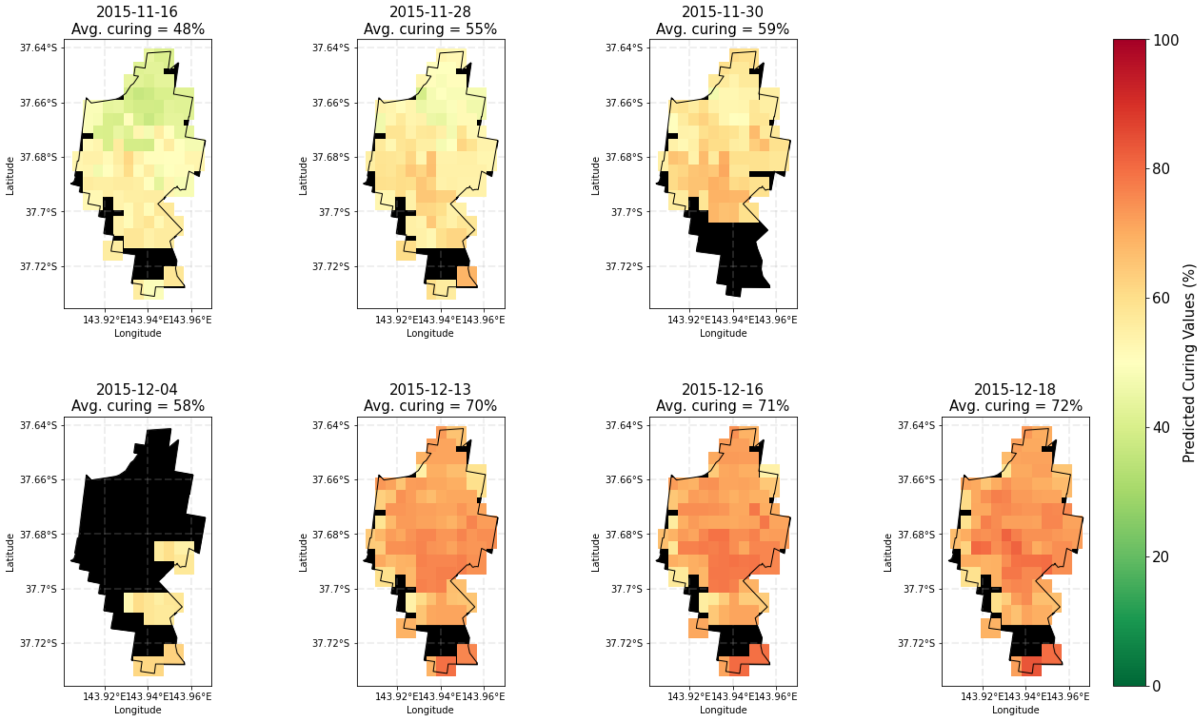


Figure 21: The predicted curing values (%) of the best performing model for seven days of available data across Scotsburn during November and December in the year 2015. The scale bar indicative of the curing percentage with low values illustrated in green and high values in red.

6.3.2 Curing value change

The range of the curing change across the four days was approximately 0% to 20% (Figure 22). Figure 22 showed that most of the data were within the true positive quadrant (top right) and close to zero for the predicted vs observed 4-day curing change. This indicated that the predicted 4-day curing change accurately captured the small-scale (0% to 10%) positive change of the observed curing values across Scotsburn. Notably, although it was not the bulk of the data density, there was a scatter of curing values within the false positive quadrant (bottom right), meaning that a positive change was predicted when in actuality, the curing change was negative. However, the false positive data range was not substantial, with most of the data within an approximate range of 0% to 5%.

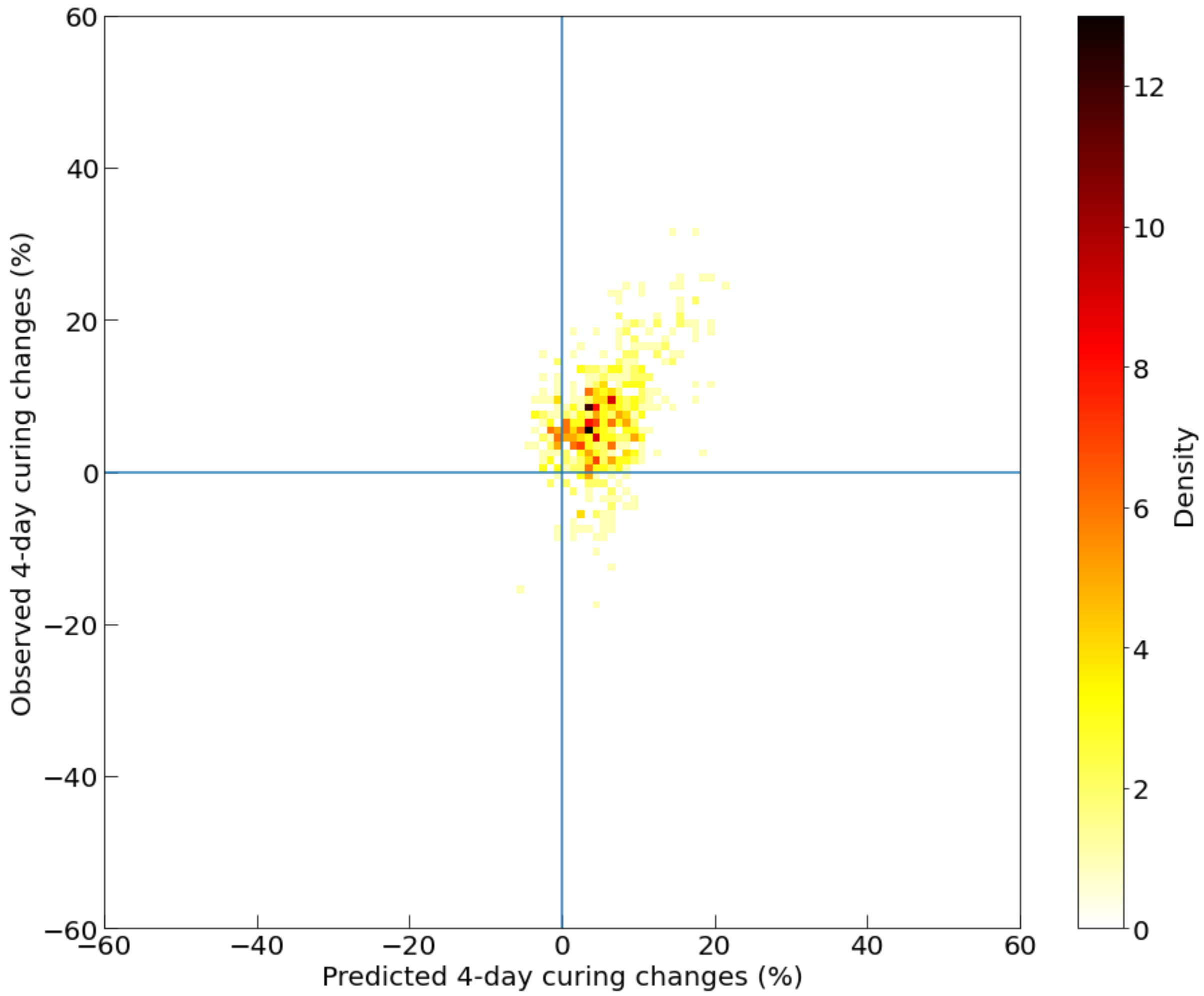


Figure 22: Predicted curing values (%) versus observed curing values (%) over seven days during November and December in the year 2015 across Scotsburn. The scale bar is indicative of the data point density with low density in yellow to high density in black.

In relation to Figure 23, Figure 24 showed that the model tended to underpredict the 4-day change when the observed 4-day change was positive and overpredict the 4-day change when the observed 4-day change was negative. This behaviour was consistent with the finding in the previous Section 6.3.1that the model had the tendency to overpredict low curing values and underpredict high curing values. The difference between the predicted and observed 4-day curing change was quite insignificant overall. This suggested that the chosen model had adequate skills to predict both the positive and negative change of curing values when presented with data it had not seen before.

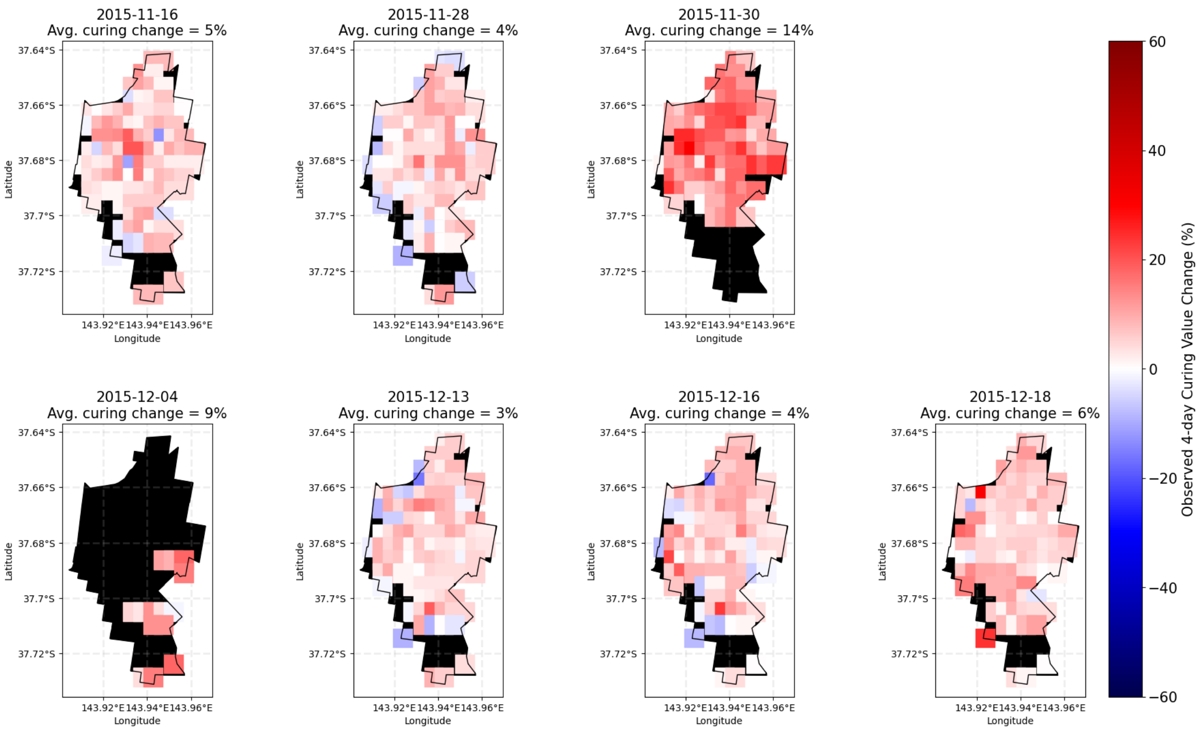


Figure 23: The observed 4-day curing value change (%) of the best performing model for seven days of available data across Scotsburn during November and December in the year 2015. The black space non-avalable data.

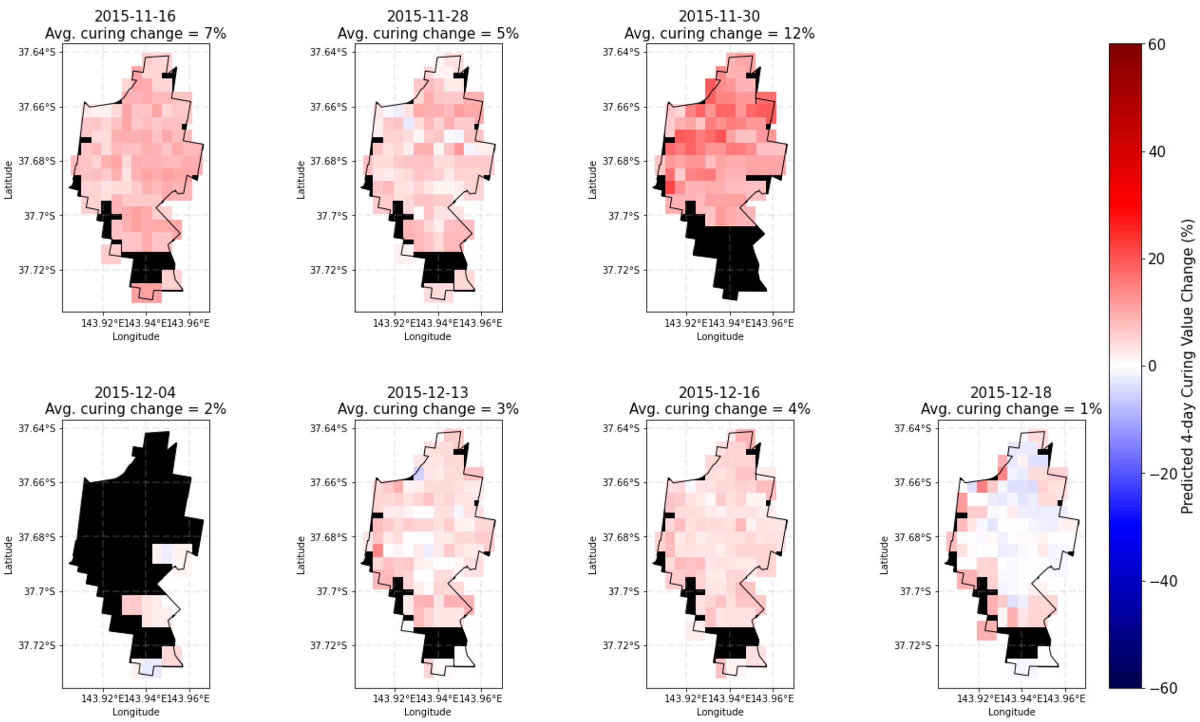


Figure 24: The predicted 4-day curing value change (%) of the best performing model for seven days of available data across Scotsburn during November and December in the year 2015. The black space non-avalable data.

6.3.3. Model Error

The persistent positive error values, shown in red in Figure 25, indicated the model was over-predicting curing. These values occurred along the outer edge of Scotsburn to the west, southwest, and south. This was a key finding, especially when the error results were compared to the distribution of vegetation in the Scotsburn area (Figure 17).

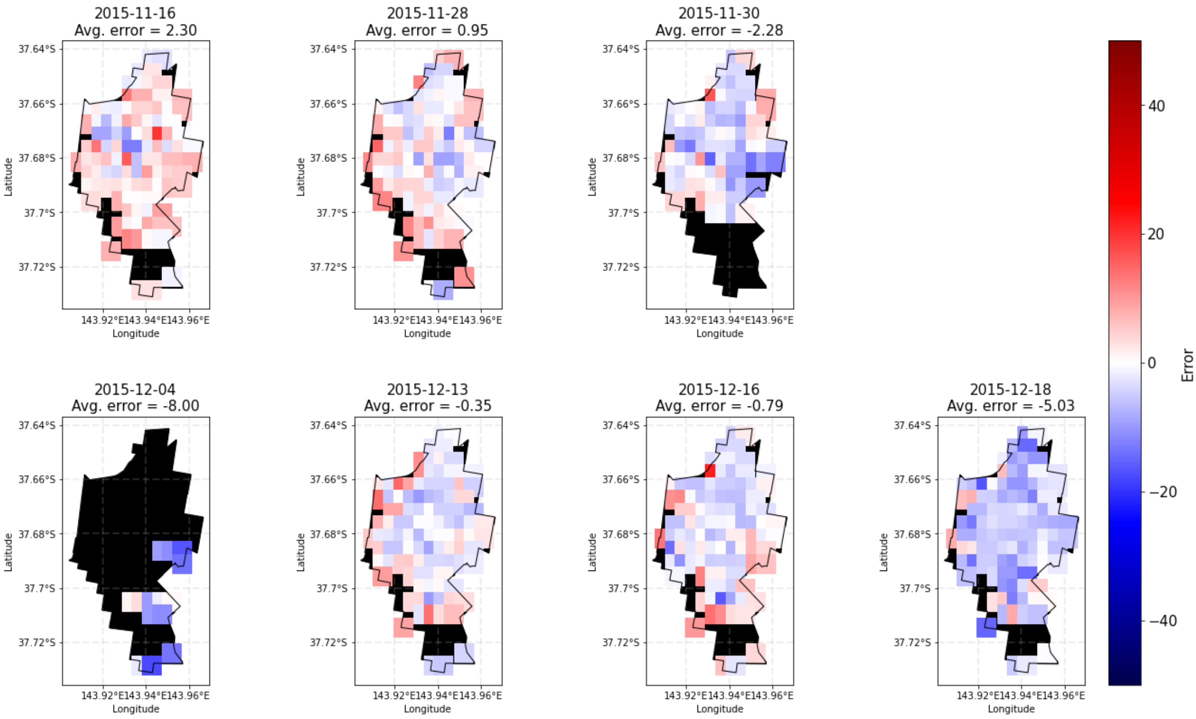


Figure 25: The difference between the predicted and observed curing change (%) of the best performing model for seven days of available data across Scotsburn during November and December in the year 2015. The black space non-avalable data.

Over the case study time frame, the pixels of grassland with persistent positive error values along the outer edge of Scotsburn were within regions where the grassland fuel type bordered the dry eucalypt fuel type (Figure 18). This finding suggested that areas of overpredicted curing were likely associated with tree-contaminated grassland pixels. This highlights a potential limitation of the resolution of the AFDRS fuel type data source used in the study design to extract curing values. If the AFDRS fuel type data was a higher resolution to account for vegetation change across the landscape, this might have reduced the tree-contaminated grassland pixels as suggested by the error.

6.3.4 Contextualising variables important for the prediction

Figure 26 shows the observed curing, precipitation, maximum temperature, VPD, relative humidity, soil moisture and maximum wind speed for the two months of the case study. These were included because they either had a high feature importance score in objective two, or were attributed to being a contributing factor to the Scotsburn fire. To capture the broader context of the incident, non-current curing observations were also extracted from the MapVictoria dataset, and all days in the two months included. Therefore, the whole two months is considered in this section, even though only seven days are run through the model. The inclusion of the non-current curing is likely the cause of the consistent curing plateaus throughout the time series.

In the week leading up to the fire precipitation was at 0 mm and had been low during most of late November and early December; the maximum temperature was increasing daily, and so too was VPD, while relative humidity fluctuated around 40%, and soil moisture was persistently low (Figure 26). In these conditions, the average curing percentage continued to increase (Figure 26). On the day of the Scotsburn grassfire incident, temperatures reached the high 40 degrees Celsius, VPD was at approximately 60 kPA, relative humidity was sitting at 20%, soil moisture was 0.18%, and the daily wind speeds were at a monthly high at around 12 m/s (Figure 26). These hot, dry, windy conditions were perfect for a grass fire to spread and be difficult to suppress (Cruz et al. 2015), as indicated by the eight days it took the Scotsburn grass fire to be given the all-clear.

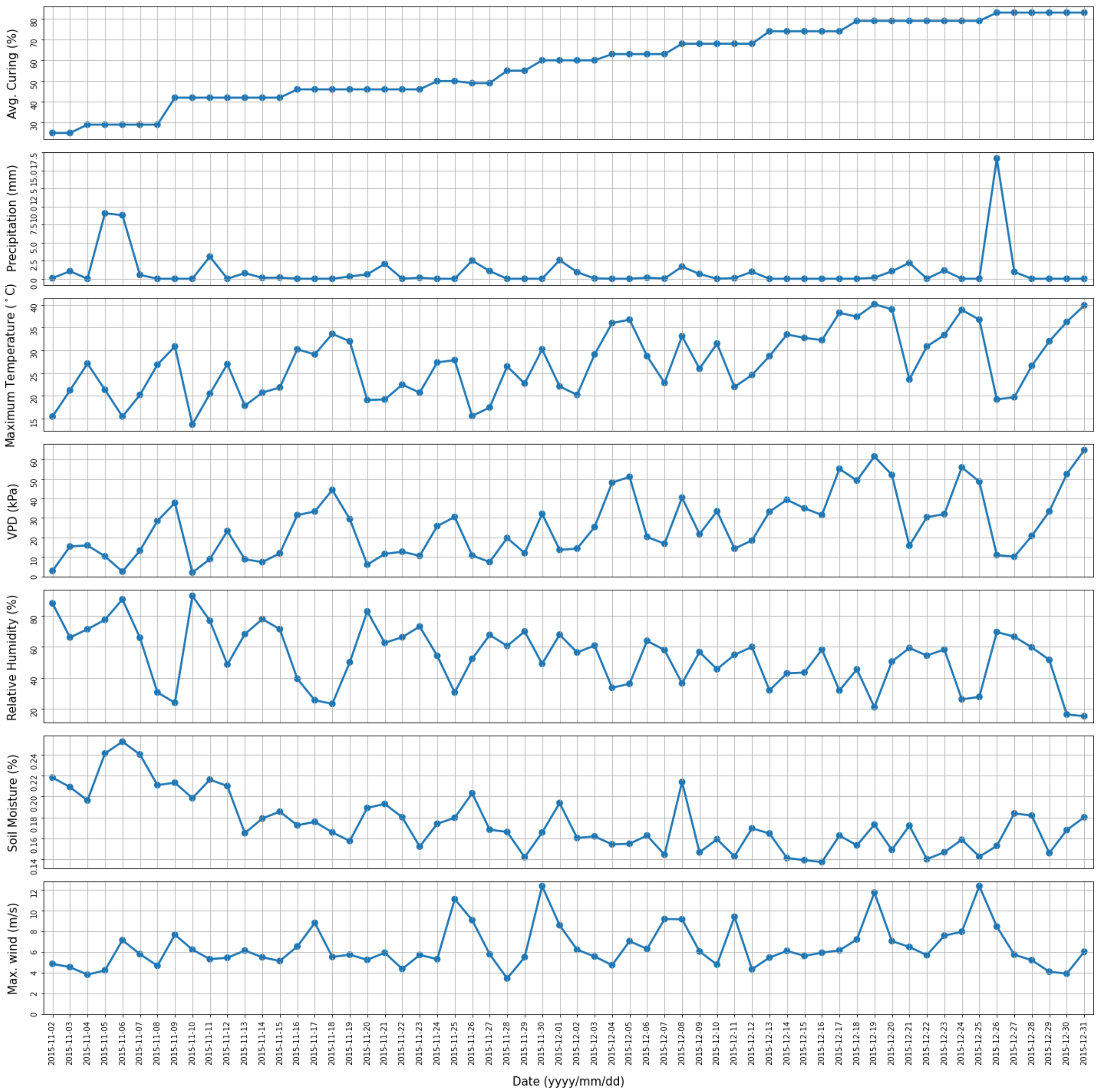


Figure 26: The top features in the machine learning models inclusive of average curing (%), precipitation (mm), maximum temperature (°C), vapour pressure deficit (VPD), relative humidity, soil moisture (%), and maximum wind (m/s) during November and December Scotsburn case study in the year 2015.

There is a gradual increase in the curing average (%), which is roughly correlated with an increase in the maximum temperature and VPD (Figure 26). The general trend in temperature and VPD are mirrored in the relative humidity time series, which often decreases as the other two increase. Additionally, soil moisture also gradually decreases over the two months. These variables combine to highlight an overall reduction in water availability. This reduction is likely a major cause of curing increasing, as plant productivity is reliant on water availability (Fahad et al., 2017, Will et al., 2013), and therefore as the water availability decreases it is more likely to seed and die. This supports the model having a good understanding of the curing process, as both maximum temperature and VPD had high feature importance scores.

The lack of trend in the precipitation time series highlights the potential inadequacy of using daily precipitation as a measure of rainfall's effect on curing. As discussed in objective two, the persistent lack of rain is currently not represented by the data, as each day is considered separately. Additionally, past a certain threshold, the amount of rain in a day does not make a difference to grass growth, as the soil reaches saturation. If instead, precipitation was measured in days since last rain, with a certain threshold as to what days are considered 'having rain', the 'precipitation' variable would likely have had an increased impact on the model and a higher feature importance score. If future research was to measure precipitation in this way, it is also likely that it would have correlated trends to the other variables.

6.4 Case Study Implications

The case study results demonstrated that the best-performing can sufficiently predict curing even at a small spatial and temporal scale. This is in addition to its adequate predictive capabilities at the statewide level, as shown in previous objectives. The results also showed that the model performs well when given data it has not seen before. The model therefore can support the Bureau in generating weekly and daily curing outlook for the entire state, which can be downscaled and used by the CFA and local fire agencies to prepare and plan for potentially dangerous curing conditions in a particular region.

Additionally the case study showed that curing can change rapidly across four days, and this is not represented in the current fire danger forecasting methods. If this model had been used to forecast the curing, it likely would have provided a more accurate fire danger rating for the Scotsburn region, and potentially improved the fire agencies preparedness for the fire. In saying this, the results also indicated that the model occasionally struggles to predict when curing decreases and can slightly underpredict curing at times of rapid curing change. Whilst it is likely that this still provides more operational benefit than assuming curing does not change at all, more testing into the operational implications is needed.

Overall, the case study results provide evidence that the model can consistently predict curing with high accuracy even when presented with unseen data. However, further testing of the model with new and more extensive data is also required before any conclusions can be made about the models overall accuracy or stability when provided new data., which is an important step in preparation for its deployment into operation.

7. Bringing it all together

7.1 Links between objectives, and major conclusions

The aim of this project was to use machine learning to build a 4-day forecast model for the degree of grassland curing. To do so, three models with; (1) curing and meteorological variables, (2) only meteorological variables and (3) only curing variables were built, optimised and tested. Based on statistical metrics as well as the comparison of predicted and observed curing values, the model that combined curing variables with meteorological variables performed best. The best-performing model was conservative in its estimation of curing change, often predicting a smaller 4-day curing change than what was observed, especially during times of significant change, however this conservativeness was also seen across all three models. All three models showing the same conservativeness suggests it is a result of how the models are learning, not which variables are included in the model.

The model with only curing features tended to predict an increase in curing when the observed curing decreases, likely because decreases in curing during the ramp up period are caused by meteorological variables. The model with meteorological features but not curing features did not show a bias towards under or over predicting, but had greater variation in the error of its predictions. Nonetheless, all three models achieved low and operationally acceptable error and variance, and could potentially provide operational benefits.

Each importance of the features in each model were then analysed in objective two, which concluded that prior curing estimates were the most important features for predicting grassland curing if they are included, followed by maximum temperature. It also suggested that it is only necessary to include an average measurement for each variable, however more testing needs to be done to determine what temporal scale is most beneficial. Additionally further reduction of the number of features considered could be done by removing either

VPD or maximum temperature and minimum relative humidity, given that VPD is calculated using the other two. However, all three had high gain scores for the model with all variables, and therefore further testing of the effect of removing them would be necessary. Overall, it is suggested that a 4-day predictive model for curing should include a temporal indicator (such as day of year), an average and daily curing observation, and a smaller list of meteorological variables, which would be dependent on the results of further testing. By reducing the number of features included the model would be more computationally efficient, and easier to implement into operational practice.

The purpose of the third objective was to identify any spatial and temporal variation in the predictions made by the best-performing model - which included both curing and meteorological variables - and the error associated with its predictions compared to observed curing estimates. The three main conclusions drawn from the third objective were; first, the model was good at predicting the seasonal cycle of curing over the study time frame with a slight tendency to underpredict the magnitude of the curing change, which supports the objective one conclusion. Second, there was no substantial spatial or temporal variation across Victoria where the model performed better or worse. And third, the model better predicts high curing values than it did lower curing values, likely due to its exposure to more testing and training data at 100%, and the stability of curing once it reaches 100%. This suggests that the model would likely be improved if more data at lower curing level were included.

The fourth objective validated the best-performing curing model using a case study analysis of a grassfire incident in Victoria. The new dataset, that included the Scotsburn data from November-December of 2015, was run through the model, with the results showing that the model also did a good job at predicting curing when provided with new data. The small scale of the case study results also highlighted some areas where the model struggled that were not apparent when looking at the results for the whole of Victoria, specifically that the model struggled in areas close to the boundary of grassland and other fuel types. This is likely because in these areas there is more woody vegetation or other mixed fuel types in the grassland, and therefore the relationship between the variables is different to what was learnt from the rest of the data. The model should therefore be further validated on a larger unseen dataset to test if these slight errors consistently occur.

Overall, the project results show that machine learning can be used to build an accurate 4-day curing forecast. The next step for further research is to test how this model could be applied in operation. Ideally it would be tested as an input to the AFDRS to determine if the model would end up significantly improving the fire danger forecast.

7.2 Project limitations and recommendations

The overall limitations of the project can be divided into two areas: data, and machine learning.

7.2.1 Data

While the data available greatly supported the algorithm's learning, the models built were limited by its training on reanalysis datasets. This was done as it was important to use the most comprehensive data available when testing the proof of concept, however, to be used in operation forecast data would be needed for all variables. Forecast data can be less accurate though, and the model would need to be retrained with the new datasets, and analysis repeated to ensure a similar level of accuracy.

Furthermore, the MapVictoria data used for training the model were imperfect. In particular, the MapVictoria model itself has a 10% margin of error and underpredicts curing in areas of undetected woody vegetation and overpredicts it in areas with bare soil and water bodies. These biases were likely propagated into this research's models and should be taken into account when the model is implemented by future research or deployed into operation.

Another data-related limitation of this research was the skewed distribution of the curing percentages given to the model for training and testing. There were disproportionately higher observations at 100% compared to any other curing values which potentially resulted in the model's enhanced ability to predict observed curing values at the high end but weaker predictive skill for values in the low to moderate range. The bias towards 100% curing values was exacerbated by the missing data of the spring months when curing values are low to moderate.

Recommendations:

The most important step that needs to be taken before a curing forecast model could be made operational is ensuring the model would be similarly effective when using near real time data. Accessing forecast datasets for each variable may be challenging, and therefore a determining factor in which variables are included in the model should be whether it is monitored in near real time. Further testing would need to be done to retrain the model on forecast data and to ensure the removal of any variables that do not significantly decrease its predictive ability.

In general, a skewed distribution of the target is not recommended when training a ML model. For the distribution of curing values specifically, the excessive 100% curing values could be restricted simply by not considering the February months, given that most grassland areas of the state are fully cured by January. A more advanced method is to identify the first day each grass pixel in Victoria reaches 100% and stop adding data for the remaining period. However, the potential effect of removing so many values at 100% should be considered. Right now the model predicts 100% curing very well, and removing the data the model has learnt from may reduce its ability to do so. Conversely, to obtain more low range curing values, future research could instead extend the research period to winter, ensuring the beginning of the curing ramp up period is captured. To enhance the model's practicality, future research should train the model with cloud-covered and non-current curing observations, as these are common in operation.

7.2.2 Machine learning

The biggest limitation of this research's methodology was the lack of strong, comprehensive cross-validations. Two of the three models tested were run and tested once, and the other one got additionally validated with a small (7-day) validating dataset. As a consequence, even when the model achieved high accuracy in this research, it cannot be guaranteed to achieve the same level of accuracy when presented with a large amount of new data in the future. This means that while the models demonstrated high operational potential, they are not ready for deployment and would require further testing to stabilise and improve the long-term consistency.

Another challenge of ML was its "black box" mechanics. While the relationships between the features (past curing and meteorological variables) and the target (current curing) were generally well understood, the prioritisation of one feature over the other cannot be manually controlled or fully explained. Furthermore, the model decision on which features to use was partially dependent on the hyperparameter settings in which there was limited understanding and tuning expertise. The "black box" nature of ML also decreases the interpretability of the model. While the importance ranking of the features was successfully extracted, it remained unknown if a feature influenced positively or negatively to the prediction and to what extent.

Recommendations:

Cross-validation is critically required for this research's models as well as any future updated ones prior to deploying into operation. Cross-validation can be conducted by splitting the training and testing datasets into small portions and allowing the model to iteratively run and generate results (as described in Section 3.5). Comparing the metrics extracted from each iteration can reveal the model stability and consistency. The size of the validating dataset should also be extended to allow for a more comprehensive and conclusive evaluation of the model.

Given the operational ambition of the model, its configurations should be thoroughly investigated and experimented with to find the optimal combination that can consistently generate accurately acceptable predictions. Additionally, instead of relying purely on the model's feature selection function, the features should be carefully predetermined, allowing the model to learn the feature-target relationships more directly and effectively. Alternatively, exploring different ML models and explainable ML approaches may also provide additional detail about the influence of each feature on the prediction.

7.3 Industry Implications

The most important finding of the project was that curing can be accurately predicted across Victoria using machine learning. Therefore, if a four-day forecast model were to become operational it should lead to more accurate grass fire danger predictions. Currently fire danger is forecasted four days out, however given curing is only monitored in real time, the grass fire danger predictions are forced to use current conditions (Martin et al., 2015). This model would allow the predicted curing value to be used in the calculation, theoretically improving the fire danger prediction.

MapVictoria is used to model curing at a near national scale, and so if shown to be operationally beneficial in Victoria, a curing forecast model could also be extended to other jurisdictions that use MapVictoria. Western Australia and the Northern Territory are the only states that use a different curing model, owing to the differing landscapes (D Wright 2022, personal communication). In order for the curing forecast to be implemented in the other states/territories a significant retraining process using their jurisdictions MapVictoria dataset would be needed, as a model trained on Victorian data would likely not perform well in other states. This project, as well as any following research, should provide a strong basis for such retraining to follow.

Additionally, this research showed that previous curing values do not need to be included for the model to predict, even though it increases the accuracy of the model. This means that a model without curing features could be implemented at times of consistent cloud cover when the MapVictoria model cannot produce new curing values. Currently MapVictoria will use an old curing value for up to 28 days of cloud cover, and after that is filled spatially from the closest current values, which at times of extreme cloud systems, could be over 100 kms away (D Wright 2022, personal communication). Having a predictive model that does not need curing will likely provide more accurate data than values from almost a month ago, or from large distances away, so whilst the model without curing is not as accurate as other models, it still would provide an incremental improvement in the operational feeds of curing data. However, in depth analysis of the model without curing was not included in this research, and should be conducted, along with the recommendations for the model with both curing and meteorological variables, before it is used operationally.

The CFA's main goal when introducing new operational practices is to have incremental improvements in their overall performance (D Wright 2022, personal communication). The organisation has the capability to trial and benefit from a new research product and resolve any issues as they arise, instead of waiting for a perfected system. Despite its limitations, the short-term curing forecast product developed during the project is an example of such a product. It has the potential to provide operational benefit to the CFA.

4. Conclusion

Grassland curing has a significant impact on grassland fires, and is an essential input in calculating grassland fire danger in Australia. It is currently only monitored in near real time, and fire danger forecasts have to use current curing conditions in their predictions. The aim of this research was therefore to use the eXtreme Gradient Boosting machine learning algorithm to build a 4-day forecast model for the degree of curing. Models with both meteorological and curing variables, only meteorological variables, and only curing variables were tested, and the results showed that all three models predicted with a high level of accuracy. However, all models also were conservative in their predictions during periods of rapid curing change. This is an aspect that needs further investigation, and recommendations were provided that may improve the models accuracy during these periods.

Of the three models, the model with both curing and meteorological variables performed best, and it is recommended that any operational models should include both kinds of variables. However, there are also operational benefits from a model that does not include curing, as it can still be used in times where curing cannot be monitored due to consecutive weeks of cloud cover. Whilst the results showed that meteorological variables are important to the predictive process, the number of features provided to the model can be reduced - particularly the amount of varying temporal observations for each variable. This will reduce the noise the algorithm must contend with and improve the computational efficiency of the model.

The model that performed best was also validated on a case study of a fire incident in Scotsburn in 2015. The results showed that the model also predicted accurately when provided with new data, however this was a small dataset and further validation using more data is necessary. Overall, the project findings showed that machine learning can be used to create a short-term curing forecast, and with some small improvements and further testing it could provide substantial operational value.

References

ABRAM, N. J., HENLEY, B. J., GUPTA, A. S., LIPPMANN, T. J. R., CLARKE, H., DOWDY, A. J., SHARPLES, J. J., NOLAN, R. H., ZHANG, T., WOOSTER, M. J., WURTZEL, J. B., MEISSNER, K. J., PITMAN, A. J., UKKOLA, A. M., MURPHY, B. P., TAPPER, N. J. & BOER, M. M. 2021. Connections of climate change and variability to large and extreme forest fires in southeast Australia. *Communications earth & environment,* 2**,** 1-17.

ADAMS, M. A., SHADMANROODPOSHTI, M. & NEUMANN, M. 2020. Causes and consequences of Eastern Australia's 2019–20 season of mega-fires: A broader perspective. *Glob Chang Biol,* 26**,** 3756-3758.

ADELEYE, M. A., CONNOR, S. E., HABERLE, S. G., HERBERT, A. & BROWN, J. 2021. European colonization and the emergence of novel fire regimes in southeast Australia. *The anthropocene review***,** 205301962110446.

AFAC 2022. Australian Fire Danger Rating Sytem Guidlines. East Melbourne, Victoria Australasian Fire and Emergency Service Authorities Council Ltd.

AUSTRALIA, G. 2004. *Area of Australia - States and Territories* [Online]. Canberra, ACT: Australian Government Available: https://www.ga.gov.au/scientific-topics/national-location-information/dimensions/area-of-australia-states-and-territories [Accessed 03/07/2022 2022].

BATTISTEL, D., KEHRWALD, N. M., ZENNARO, P., PELLEGRINO, G., BARBARO, E., ZANGRANDO, R., PEDELI, X. X., VARIN, C., SPOLAOR, A., VALLELONGA, P. T., GAMBARO, A. & BARBANTE, C. 2018. High-latitude Southern Hemisphere fire history during the mid- to late Holocene (6000-750BP). *Climate of the past,* 14**,** 871-886.

BERGSTRA, J., YAMINS, D. & COX, D. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. International conference on machine learning, 2013. PMLR, 115-123.

BISHOPA, J., BELL, T., HUANG, C., WARD, M. & BAILEY, C. G. 2021. Fire On The Farm: Assessing the Impacts of the 2019-2020 Bushfires on Food and Agriculture in Australia. Ultimo NSW: WWF.

BOM 2021. Bureau of Meteorology Annual Report 2020-21. Melbourne, Victoria: Bureau of Meteorology.

BOM 2022. Fire Weather Service Level Specification. Melbourne, VIC.

CALLIGEROS, M., GRAY, D. & VEDELAGO, C. 2015. More than 300 fires break out across the state as temperatures soar. *The Age*.

CANADELL, J. G., MEYER, C. P., COOK, G. D., DOWDY, A., BRIGGS, P. R., KNAUER, J., PEPLER, A. & HAVERD, V. 2021. Multi-decadal increase of forest burned area in Australia is linked to climate change. *Nat Commun,* 12**,** 6921-11.

CARMONA, P., DWEKAT, A. & MARDAWI, Z. 2022. No more black boxes! Explaining the predictions of a machine learning XGBoost classifier algorithm in business failure. *Research in International Business and Finance,* 61**,** 101649.

CECCATO, P., FLASSE, S. & GRÉGOIRE, J.-M. 2002. Designing a spectral index to estimate vegetation water content from remote sensing data: Part 2. Validation and applications. *Remote sensing of environment,* 82**,** 198-207.

CFA 2014. CFA Grassland Curing Guide. *In:* AUTHORITY, C. F. (ed.). Mount Waverley, Victoria: Victoria State Government

CFA 2021. CFA Annual Report 2020–21. Melbourne, Victoria: Country Fire Authority.

CHEN, T. & GUESTRIN, C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, 2016. 785-794.

CHENEY, N. P., GOULD, J. S., MCCAW, W. L. & ANDERSON, W. R. 2012. Predicting fire behaviour in dry eucalypt forest in southern Australia. *Forest ecology and management,* 280**,** 120-131.

COLLINS, L., BRADSTOCK, R. A., CLARKE, H., CLARKE, M. F., NOLAN, R. H. & PENMAN, T. D. 2021. The 2019/2020 mega-fires exposed Australian ecosystems to an unprecedented extent of high-severity fire. *Environmental research letters,* 16**,** 44029.

COWLED, B. D., HILLMAN, A., WARD, M. P., CLUTTERBUCK, H., DOYLE, M., WEBB WARE, J., THOMAS, M., PLAIN, K., BARWELL, R., LAURENCE, M. & PFEIFFER, C. 2022. The black summer bushfires: impacts and risk factors for livestock bushfire injury in south-eastern Australia. *Australian veterinary journal,* 100**,** 306-317.

CRUZ, M. G., GOULD, J. S., KIDNIE, S., BESSELL, R., NICHOLS, D. & SLIJEPCEVIC, A. 2015. Effects of curing on grassfires: II. Effect of grass senescence on the rate of fire spread. *International journal of wildland fire,* 24**,** 838-848.

CRUZ, M. G., SULLIVAN, A., KIDNIE, S., HURLEY, R. & NICHOLLS, D. 2016. The effect of grass curing and fuel structure on fire behaviour: Final report. Canberra, ACT: CSIRO.

CRUZ, M. G., SULLIVAN, A. L., GOULD, J. S., HURLEY, R. J. & PLUCINSKI, M. P. 2018. Got to burn to learn: the effect of fuel load on grassland fire behaviour and its management implications. *International journal of wildland fire,* 27**,** 727.

DATABRICKS. 2022. *Hyperopt concepts* [Online]. Apache Software Foundation. Available: https://docs.databricks.com/machine-learning/automl-hyperparam-tuning/hyperopt-concepts.html [Accessed 21/10/2022 2022].

DEB, P., MORADKHANI, H., ABBASZADEH, P., KIEM, A. S., ENGSTRÖM, J., KEELLINGS, D. & SHARMA, A. 2020. Causes of the Widespread 2019–2020 Australian Bushfire Season. *Earth's future,* 8**,** n/a.

DEVELOPERS, X. 2021. *XGBoost Documentation* [Online]. XGBoost Developers. Available: https://xgboost.readthedocs.io/en/stable/index.html# [Accessed 21/10/2022 2021].

DILLEY, A. C., MILLIE, S., O'BRIEN, D. M. & EDWARDS, M. 2004. The relation between Normalized Difference Vegetation Index and vegetation moisture content at three grassland locations in Victoria, Australia. *International journal of remote sensing,* 25**,** 3913-3930.

DORIGO, W., WAGNER, W., ALBERGEL, C., ALBRECHT, F., BALSAMO, G., BROCCA, L., CHUNG, D., ERTL, M., FORKEL, M., GRUBER, A., HAAS, E., HAMER, P. D., HIRSCHI, M., IKONEN, J., DE JEU, R., KIDD, R., LAHOZ, W., LIU, Y. Y., MIRALLES, D., MISTELBAUER, T., NICOLAI-SHAW, N., PARINUSSA, R., PRATOLA, C., REIMER, C., VAN DER SCHALIE, R., SENEVIRATNE, S. I., SMOLANDER, T. & LECOMTE, P. 2017. ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment,* 203**,** 185-215.

DUFF, T. J., BESSELL, R. & CRUZ, M. G. 2020. Grass Curing/Cured Fuels. *In:* MANZELLO, S. L. (ed.) *Encyclopedia of Wildfires and Wildland-Urban Interface (WUI) Fires.* Cham: Springer International Publishing.

ELITH, J., LEATHWICK, J. R. & HASTIE, T. 2008. A working guide to boosted regression trees. *Journal of animal ecology,* 77**,** 802-813.

EMV 2016. Emergency Management Operational Review 2015-16. Brunswick, VIC: Emergency Management Victoria.

FAHAD, S., BAJWA, A. A., NAZIR, U., ANJUM, S. A., FAROOQ, A., ZOHAIB, A., SADIA, S., NASIM, W., ADKINS, S., SAUD, S., IHSAN, M. Z., ALHARBY, H., WU, C., WANG, D. & HUANG, J. 2017. Crop Production under Drought and Heat Stress: Plant Responses and Management Options. *Frontiers in Plant Science,* 8.

FERREIRA, L. N., VEGA-OLIVEROS, D. A., ZHAO, L., CARDOSO, M. F. & MACAU, E. E. N. 2020. Global fire season severity analysis and forecasting. *Computers & geosciences,* 134**,** 104339.

FEURER, M. & HUTTER, F. 2019. Hyperparameter optimization. *Automated machine learning.* Springer, Cham.

FRIEDMAN, J. H. & MEULMAN, J. J. 2003. Multiple additive regression trees with application in epidemiology. *Statistics in medicine,* 22**,** 1365-1381.

GOULD, J. G. & CRUZ, M. 2012. Australian Fuel Classification: Stage II East Melbourne, Victoria: CSIRO.

GRUBER, A., SCANLON, T., VAN DER SCHALIE, R., WAGNER, W. & DORIGO, W. 2019. Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology. *Earth system science data,* 11**,** 717-739.

HINES, F., G. TOLHURST, K., AG. WILSON, A. & J. MCCARTHY, G. 2010. Overall Fuel Hazard Assessment Guide. *In:* DEPARTMENT OF ENVIRONMENT, L., WATER AND PLANNING (ed.). East Melbourne, Victoria: Victoria State Government.

HOWARD, K. 2022. A Comparative Analysis between the Grassland Fire Danger Index and Fire Weather Index for Five Fuel Types in the Mallee and Wimmera Fire Weather Districts. Melbourne, VIC: Monash University.

JAIN, P., COOGAN, S. C. P., SUBRAMANIAN, S. G., CROWLEY, M., TAYLOR, S. & FLANNIGAN, M. D. 2020. A review of machine learning applications in wildfire science and management. *Environmental Reviews,* 28**,** 478-505.

JAKOB, D. 2010. Challenges in developing a high-quality surface wind-speed data-set for Australia. *Australian Meteorological and Oceanographic Journal,* 60**,** 227-236.

JAMES, G., WITTEN, D., HASTIE, T. & TIBSHIRANI, R. 2013. *An introduction to statistical learning: With applications in R,* New York, New York: Springer.

JONES, D. A., WANG, W. & FAWCETT, R. 2009. High-quality spatial climate data-sets for Australia. *Australian meteorological and oceanographic journal,* 58**,** 233-248.

JOSEPH, V. R. 2022. Optimal Ratio for Data Splitting. *arXiv.org*.

KHASTAGIR, A., JAYASURIYA, N. & BHUYIAN, M. A. 2018. Assessment of fire danger vulnerability using McArthur's forest and grass fire danger indices. *Natural hazards (Dordrecht),* 94**,** 1277-1291.

KIDNIE, S., CRUZ, M. G., GOULD, J., NICHOLS, D., ANDERSON, W. & BESSELL, R. 2015. Effects of curing on grassfires: I. Fuel dynamics in a senescing grassland. *International journal of wildland fire,* 24**,** 828-837.

LINDENMAYER, D., DOVERS, S. & CARY, G. J. 2003. *Australia burning : fire ecology, policy and management issues,* Collingwood, Vic., Collingwood, Vic. : CSIRO Publishing.

LUNDBERG, S. M. & LEE, S.-I. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems,* 30.

MALLEY, G. 2022. Investigating the Spatiotemporal Variability of Grassland Curing Across the Nine Fire Weather Districts of Victoria, Australia. Monash University.

MARK B., M., BENNETTE, A. & MACINTOSH, A. 2020. Royal Commission into National Natural Disaster Arrangements Canberra, ACT: Commonwealth of Australia

MARTIN, D., CHEN, A., BESSELL, R., KIDNIE, S., NICHOLS, D. & SLIJEPCEVIC, A. 2016. Reducing Inconsistencies of Grassland Curing Assessment Between Multiple Jurisdictions Across Australia *5th Fuels and Fire Behaviour Conference.* Melbourne, Victoria International Association of Wildland Fire.

MARTIN, D., CHEN, T., NICHOLS, D., BESSELL, R., KIDNIE, S. & ALEXANDER, J. 2015. Integrating ground and satellite-based observations to determine the degree of grassland curing. *International journal of wildland fire,* 24**,** 329-339.

MARTIN, D., GRANT, I., JONES, S. & ANDERSON, S. 2009. Development of Satellite Vegetation Indices to Assess Grassland Curing Across Australia and New Zealand. Berlin, Heidelberg: Berlin, Heidelberg: Springer Berlin Heidelberg.

MATTHEWS, S., FOX-HUGHES, P., GROOTEMAAT, S., HOLLIS, J. J., KENNY, B. J. & SAUVAGE, S. 2019. Australian Fire Danger Rating System: Research Prototype. Lidcombe, NSW: NSW Rural Fire Service.

MCARTHUR, A. 1966. Weather and grassland fire behaviour. Canberra, ACT: Forestry and Timber Bureau.

MUELLER, J. 2021. *Machine learning*, Hoboken, New Jersey : John Wiley & Sons.

MURPHY, K. P. 2012. *Machine learning: a probabilistic perspective*, MIT press.

NERLEKAR, A. N. & VELDMAN, J. W. 2020. High plant diversity and slow assembly of old-growth grasslands. *Proc Natl Acad Sci U S A,* 117**,** 18550-18556.

NEW, T. R. 2019. Australian Grasslands – Variety and Extent. *In:* NEW, T. R. (ed.) *Insect Conservation and Australia's Grasslands.* Cham: Springer International Publishing.

NEWNHAM, G. J., GRANT, I. F., MARTIN, D. N. & ANDERSON, S. A. 2010. Improved methods for assessment and prediction of grassland curing. *Bushfire Cooperative Research Centre, Melbourne, VIC.*

NIE, Z. & NORTON, M. R. 2009. Stress Tolerance and Persistence of Perennial Grasses: The Role of the Summer Dormancy Trait in Temperate Australia. *Crop science,* 49**,** 2405-2411.

NOLAN, R. H., BOER, M. M., COLLINS, L., RESCO DE DIOS, V., CLARKE, H., JENKINS, M., KENNY, B. & BRADSTOCK, R. A. 2020. Causes and consequences of eastern Australia's 2019–20 season of mega-fires. *Glob Chang Biol,* 26**,** 1039-1041.

PARR, T. 2022. *A Visual Explanation for Regularization of Linear Models* [Online]. University of San Francisco. Available: https://explained.ai/regularization/index.html [Accessed 21/10/2022 2022].

PARROTT, M. L., WICKER, L. V., LAMONT, A., BANKS, C., LANG, M., LYNCH, M., MCMEEKIN, B., MILLER, K. A., RYAN, F., SELWOOD, K. E., SHERWEN, S. L. & WHITEFORD, C. 2021. Emergency Response to Australia's Black Summer 2019-2020: The Role of a Zoo-Based Conservation Organisation in Wildlife Triage, Rescue, and Resilience for the Future. *Animals (Basel),* 11**,** 1515.

PEG, F. 2019. Messy history. *Cultural studies review,* 25**,** 262-264.

PLUCINSKI, M. P. 2013. Modelling the probability of Australian grassfires escaping initial attack to aid deployment decisions. *International journal of wildland fire,* 22**,** 459.

RASCHKA, S. 2014. *About Feature Scaling and Normalization – and the effect of standardization for machine learning algorithms* [Online]. Available: https://sebastianraschka.com/Articles/2014_about_feature_scaling.html [Accessed 21/10/2022 2022].

RAYMENT, J. T. & FRENCH, K. 2021. Uncertainty in research about key invasion characteristics limits the evaluation of exotic perennial grasses in natural systems in New South Wales, Australia. *Ecological management & restoration,* 22**,** 53-63.

ROUSE JR, J. W., HAAS, R. H., SCHELL, J. & DEERING, D. 1973. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation.

SCIKIT-LEARN. 2021. *Metrics and scoring: quantifying the quality of predictions* [Online]. Scikit-learn developers Available: https://scikit-learn.org/stable/modules/model_evaluation.html#median-absolute-error [Accessed 21/10/2022 2022].

SHI, R., XU, X., LI, J. & LI, Y. 2021. Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization. *Applied Soft Computing,* 109**,** 107538.

SQUIRE, D. T., RICHARDSON, D., RISBEY, J. S., BLACK, A. S., KITSIOS, V., MATEAR, R. J., MONSELESAN, D., MOORE, T. S. & TOZER, C. R. 2021. Likelihood of unprecedented drought and fire weather during Australia's 2019 megafires. *NPJ climate and atmospheric science,* 4**,** 1-12.

SRINIVAS, P. & KATARYA, R. 2022. hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost. *Biomedical Signal Processing and Control,* 73**,** 103456.

SU, C.-H., EIZENBERG, N., STEINLE, P., JAKOB, D., FOX-HUGHES, P., WHITE, C. J., RENNIE, S., FRANKLIN, C., DHARSSI, I. & ZHU, H. 2019. BARRA v1.0: The Bureau of Meteorology Atmospheric high-resolution Regional Reanalysis for Australia. *Geoscientific Model Development,* 12**,** 2049-2068.

SULLIVAN, A. L., MCCAW, W. L., CRUZ, M. G., MATTHEWS, S. & ELLIS, P. F. 2012. Fuel, fire weather and fire behaviour in Australian ecosystems. *Flammable Australia: fire regimes, biodiversity and ecosystems in a changing world***,** 51-77.

TRAN, N., SCHNEIDER, J.-G., WEBER, I. & QIN, A. K. 2020. Hyper-parameter optimization in classification: To-do or not-to-do. *Pattern Recognition,* 103**,** 107245.

TWIDWELL, D., SCHOLTZ, R. & DONOVAN, V. 2022. Prescribed fires of the future: technology and fire. *Global Application of Prescribed Fire***,** 258.

WANG, S. S. C., QIAN, Y., LEUNG, L. R. & ZHANG, Y. 2021. Identifying key drivers of wildfires in the contiguous US using machine learning and game theory interpretation. *Earth's future,* 9**,** e2020EF001910.

WHITE, M., GRIFFIOEN, P. & NEWELL, G. 2020. Multi-temporal Land Cover and Native Vegetation Extent for Victoria Heidelberg, Victoria: The Victoria State Government.

WILL, R. E., WILSON, S. M., ZOU, C. B. & HENNESSEY, T. C. 2013. Increased vapor pressure deficit due to higher temperature leads to greater transpiration and faster mortality during drought for tree seedlings common to the forest–grassland ecotone. *New Phytol,* 200**,** 366-374.

WILLMOTT, C. J. & MATSUURA, K. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research,* 30**,** 79-82.

WRIGLEY, B. 2019. ‘I just wanted to die’: man responsible for Scotsburn fire speaks out. *The Courier*.

WU, J., CHEN, X.-Y., ZHANG, H., XIONG, L.-D., LEI, H. & DENG, S.-H. 2019. Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimizationb. *Journal of Electronic Science and Technology,* 17**,** 26-40.

YAN, H., YAN, K. & JI, G. 2022. Optimization and prediction in the early design stage of office buildings using genetic and XGBoost algorithms. *Building and Environment,* 218**,** 109081.

APPENDIX

Appendix i. MODIS spectral bands

Table 1: MODIS Spectral bands. Bands 1, 2, and 6 (unshaded) are used to determine grass curing degree.

MODIS Bands	Wavelength (µm)	Spectral region	Spatial resolution (m)
1	0.620 – 0.670	Red	250/500
2	0.841 – 0.876	Near Infrared	250/500
3	0.459 – 0.479	Blue	500
4	0.545 – 0.565	Green	500
5	1.230 – 1.250	Near Infrared	500
6	1.628 – 1.652	Mid Infrared	500
7	2.105 – 2.155	Mid Infrared	500

Appendix ii. Ground-Based Influence (GBI) calculation

GBI= -mx+c (Martin et al. 2015)

Where m represents a gradient coefficient derived from elevation, x represents the distance (km/100) to the closest field site location, and c represents the maximum influence (%) determined by the age of satellite data (Table D1) (Martin et al. 2015).

Table D1: Ground-Based Influence (GBI) coefficients determined by their corresponding attributes (Martin et al. 2015)

Elevation (m)	*m*	Distance (km)	*x*	Age (days)	*c*
≤25	131.02	1	0.01	0	90
25–75	149.65	3	0.03	1	91
75–125	170.94	6	0.06	2	92
125–175	195.26	9	0.09	3	93
175–225	223.03	12	0.12	4	94
225–275	254.76	15	0.15	5	95
275–325	291.00	18	0.18	6	96
325–375	332.39	21	0.21	7	97
375–425	379.68	24	0.24	8	98
425–475	433.69	27	0.27	≥9	99
475–525	495.38	30	0.30	Null	99
525–575	565.84	33	0.33		
575–625	646.34	36	0.36		
625–675	738.28	39	0.39		
675–725	843.30	–	–		
≥725	963.26	75	0.75		

Appendix iii. Vegetation indices calculation

$NDVI=R_2-R_1R_2+R_1$ (Rouse et al. 1973)

$GVMI=(R_2 + 0.1) - (R_6 + 0.02)(R_2 + 0.1) +(R_6 + 0.02)$ (Ceccato et al. 2002)

where R_1, R_2 and R_6 are MODIS surface reflectance bands 1, 2 and 6 respectively (Appendix i.)

Appendix iv. List of features

'y', 'x', 'age', 'date', 'day_of_year', 'day_of_cycle', 'curing_4day', 'sm_4day', 'precip_4day', 'max_temp_4day', 'min_temp_4day', 'max_relhum_4day', 'min_relhum_4day', 'max_wind_4day', 'min_wind_4day', 'max_swrad_4day', 'min_swrad_4day', 'max_lwrad_4day', 'min_lwrad_4day', 'vpd_4day', 'curing_5day', 'sm_5day', 'precip_5day', 'max_temp_5day', 'min_temp_5day', 'max_relhum_5day', 'min_relhum_5day', 'max_wind_5day', 'min_wind_5day', 'max_swrad_5day', 'min_swrad_5day', 'max_lwrad_5day', 'min_lwrad_5day', 'vpd_5day', 'curing_6day', 'sm_6day', 'precip_6day', 'max_temp_6day', 'min_temp_6day', 'max_relhum_6day', 'min_relhum_6day', 'max_wind_6day', 'min_wind_6day', 'max_swrad_6day', 'min_swrad_6day', 'max_lwrad_6day', 'min_lwrad_6day', 'vpd_6day', 'curing_7day', 'sm_7day', 'precip_7day', 'max_temp_7day', 'min_temp_7day', 'max_relhum_7day', 'min_relhum_7day', 'max_wind_7day', 'min_wind_7day', 'max_swrad_7day', 'min_swrad_7day', 'max_lwrad_7day', 'min_lwrad_7day', 'vpd_7day', 'curing_8day', 'sm_8day', 'precip_8day', 'max_temp_8day', 'min_temp_8day', 'max_relhum_8day', 'min_relhum_8day', 'max_wind_8day', 'min_wind_8day', 'max_swrad_8day', 'min_swrad_8day', 'max_lwrad_8day', 'min_lwrad_8day', 'vpd_8day', 'curing_9day', 'sm_9day', 'precip_9day', 'max_temp_9day', 'min_temp_9day', 'max_relhum_9day', 'min_relhum_9day', 'max_wind_9day', 'min_wind_9day', 'max_swrad_9day', 'min_swrad_9day', 'max_lwrad_9day', 'min_lwrad_9day', 'vpd_9day', 'curing_10day', 'sm_10day', 'precip_10day', 'max_temp_10day', 'min_temp_10day', 'max_relhum_10day', 'min_relhum_10day', 'max_wind_10day', 'min_wind_10day', 'max_swrad_10day', 'min_swrad_10day', 'max_lwrad_10day', 'min_lwrad_10day', 'vpd_10day', 'curing_avg_7day', 'sm_avg_7day', 'precip_avg_7day', 'max_temp_avg_7day', 'min_temp_avg_7day', 'min_relhum_avg_7day', 'max_relhum_avg_7day', 'max_wind_avg_7day', 'min_wind_avg_7day', 'min_swrad_avg_7day', 'max_swrad_avg_7day', 'max_lwrad_avg_7day', 'min_lwrad_avg_7day', 'vpd_avg_7day', 'curing_avg_4day', 'sm_avg_4day', 'precip_avg_4day', 'max_temp_avg_4day', 'min_temp_avg_4day', 'min_relhum_avg_4day', 'max_relhum_avg_4day', 'max_wind_avg_4day', 'min_wind_avg_4day', 'min_swrad_avg_4day', 'max_swrad_avg_4day', 'max_lwrad_avg_4day', 'min_lwrad_avg_4day', 'vpd_avg_4day'

Appendix v. AFDRS fuel map

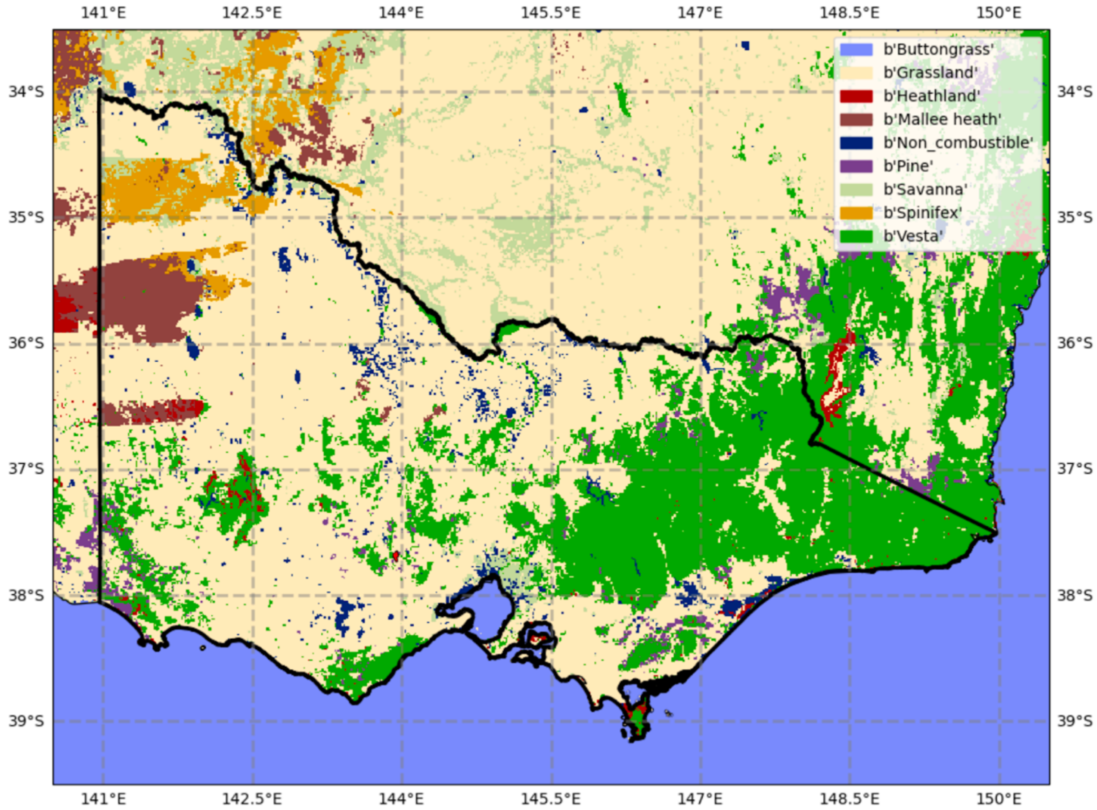


Figure A1: AFDRS fuel map (Matthews S., 2019)

Appendix vi. Regularisation hyperparameters

The hyperparameters reg_alpha and reg_lambda represent the two best-known regularisation techniques, Lasso and Ridge regression respectively. The key difference between the two techniques is that Lasso regression can set any regression coefficient to zero, whereas Ridge regression only forces the regression coefficient towards but not equal to zero (James et al. 2013). As a result, a model with Ridge regression operates with all the input features, whereas a model with Lasso regression only requires a subset of input features to make predictions. In practice, Lasso regression is useful when there are a large number of input features and Ridge regression is useful when the input features are highly correlated with each other or with the target (Parr, 2022).

Appendix vii. Optimised hyperparameters discussion

The model with all features (including curing) had a relatively small learning_rate value (0.05), which suggests that many decision trees were required in order for the model to make a good prediction. Whilst this is not inherently a bad thing, a small learning_rate can increase the computational resources needed to make predictions and risks overfitting the model to the training and testing data.

The max_depth hyperparameter determines how many decisions each tree can make before making a prediction. Given each decision can only relate to one variable, it also determines the number of variables each tree can consider. The other hyperparameter related to the number of features used was colsample_bytree, which determines what proportion of features can be accessed by each tree. The max_depth and colsample_bytree for this model imply that approximately 20 to 40 percent of the total number of features were used to make a prediction in this tree. Including this many features means it is unlikely the model is being underfit, without overwhelming the model and risking overfitting. Similarly, the two regularisation hyperparameters, reg_alpha and reg_lambda, were not too different nor high relative to the range tested. Considering that 20 to 40 percent of the total features were considered for each tree, the effect of regularisation appeared to be quite reasonable in retaining the most relevant features but not excluding too many which can pose the risk of underfitting.

Having a min_child_weight of 0 was surprising, as this implies that each tree was encouraged to split indefinitely. In reality, a min_child_weight value of 0 can only happen for classification problems and not regression problems of this research. However, due to limited knowledge of XGBoost hyperparameters before this project, it is possible that an impractical range for this hyperparameter was used. Regardless, the implication of this hyperparameter remained the same - there was no restriction on when a tree should stop splitting. This can result in an overfitted model, however this was mitigated in some way thanks to the gamma hyperparameter. Gamma dictates the level of improvement a decision needs to provide before it is included in a tree, and therefore ensures each decision a tree makes is bringing value to the prediction. Additionally, a 0.7 value for the sub_sample hyperparameter implied that on average, 70 percent of the observations were sampled to make a prediction. This suggests that enough data was used by each tree that it is unlikely that critical spatial or temporal information was lost, especially given different data would be used by each sequential tree. Considering the function of all these hyperparameters altogether, the model with all variables (including curing) had some risk of overfitting.

The model with all variables apart from curing had some hyperparameters that are very similar to that of the model with all variables. The gamma value was the same (5) and sub_sample value within 7% of each other (0.63 and 0.70 respectively), indicating that this model also both requires a decent improvement in the model for a decision to be included and sampled a reasonable number of observations when making each prediction. This suggests it also is unlikely to be underfitted to the data. However, the hyperparameters related to feature selection displayed noticeable differences between the two models. Specifically, the combination of max_depth and colsample_bytree implied that only 16 to 22 percent of the total number of features were used by each tree, and the relatively high regularisation terms, reg_alpha and reg_lambda, meant that features were strictly selected for the predicting process. It was inconclusive whether or not this high level of restriction was good, as it may have helped retain the most relevant features and prevented overfitting, or bad, as it could also have overlooked too many features and undertrained the model.

Another key difference between these models is that the model without curing has a higher value for learning_rate (0.68 compared to 0.05). This implies that this model included less decision trees, making the composition of the model simpler. This may be part of the reason it resulted in higher RMSE and MAE metrics. However, this one-time trade off in bias and variance might support a more stable and consistent performance in the long term. Each model was only tested once, and on one set of data, so it is unclear as to whether or not the models have under or fitted to the training or testing data. This is something that should be explored more in future research.

Appendix viii. Variables correlation

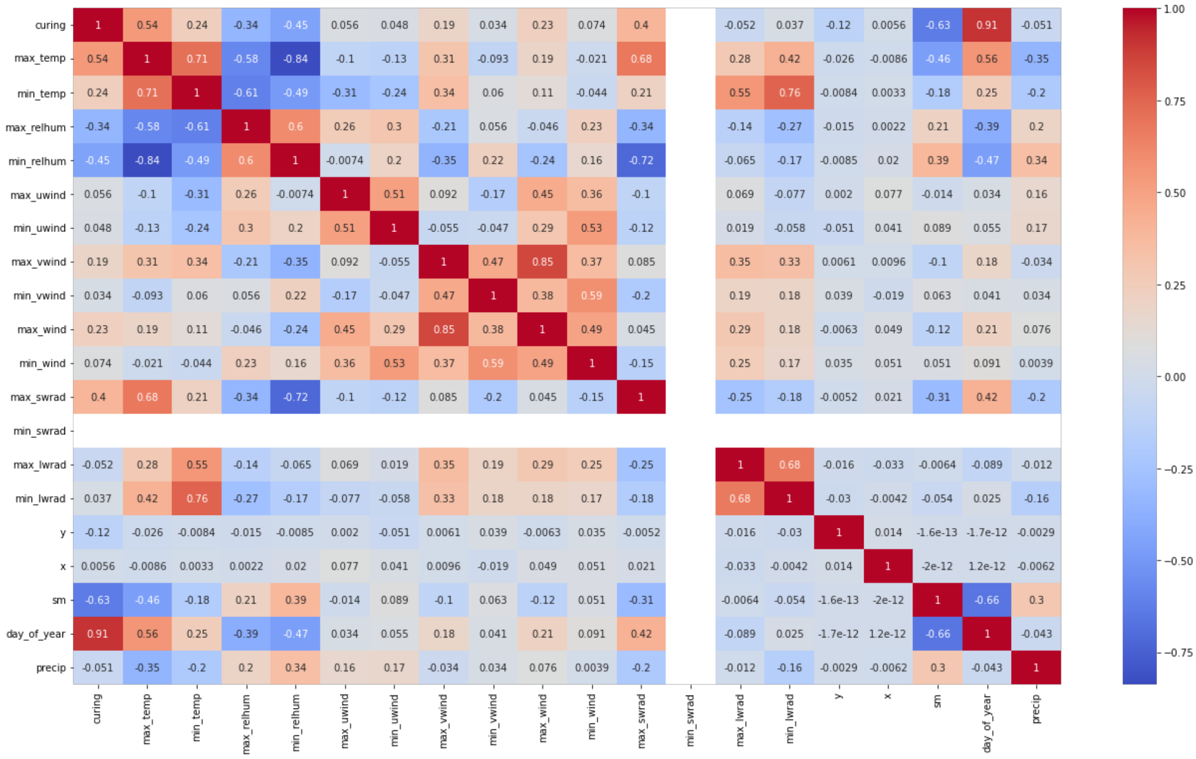


Figure A2: Correlation between variables