TO BURN OR NOT TO BURN? MACHINE LEARNING THE DRIVERS OF LARGE WILDFIRE OCCURRENCE IN SASKATCHEWAN'S BOREAL FOREST

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Master of Science
In the Department of Geography and Planning
University of Saskatchewan
Saskatoon

By

JEFFREY KYLE HARDER

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other uses of materials in this thesis in whole or part should be addressed to:

Head of the Department of Geography and Planning
117 Kirk Hall
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5C8 Canada

OR

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9 Canada

ABSTRACT

Previous attempts to characterize the drivers of large wildfire occurrence have considered the probability that an ignition event will happen in addition to the probability that it will grow into a large wildfire. Including the very low probability of an ignition event happening at a given time and place can obscure the understanding of what governs the transition from ignition to sustained combustion. The purpose of my thesis was to investigate the biotic and abiotic factors and conditions driving large wildfire occurrences in the boreal forest of Saskatchewan. My objectives were to a) analyze historical wildfire occurrences and patterns of what burned and when; b) identify and select variables that have been used to explain wildfire behavior; c) determine which variables were best suited to explaining wildfire occurrences by building a model; and d) apply the model as a lens through which to view historical wildfire occurrences to understand what will drive future occurrences. To investigate objective a, I analyzed historical large (> 200 ha) wildfire burn footprints occurring between 2008 and 2018 in Saskatchewan's boreal forest by associating each with fuel types and ecozone, and comparing area burned with area available of each respective fuel type and ecozone. This analysis showed that historical wildfire activity in the study area varied considerably by ecozone, fuel type, and time throughout the year. To investigate objectives b - d, I assembled a dataset of historical (1990 to 2018) wildfire ignition locations, associated them with weather and Fire Weather Index (FWI) component variables from the same day and location, and identified the Fire Behavior Prediction System (FPB) fuel type that each fire started in. I then built a random forest classifier using grid search hyperparameter tuning on the dataset to determine when conditions were suitable for a large wildfire (eventual fire size > 200 ha) to occur or not (eventual fire size < 0.5ha). Predicting unseen test data with the model resulted in a recall (a.k.a. sensitivity) of 0.788. The most important variables in the model were relative humidity and fuel type. Applying the model to average conditions in historical high and low wildfire activity months showed that suitable conditions for large wildfires can vary greatly within and among years, and among fuel types. Applying the model to historical daily weather observations demonstrated that different fuel types have different threshold conditions for large wildfires, and the threshold conditions for some fuel types are observed more frequently than for others. This study helps advance our understanding of wildfire activity, and it can be used to inform wildfire management decisions to reduce the risk wildfires pose.

ACKNOWLEDGEMENTS

I express my sincere gratitude to my supervisor Dr. Xulin Guo and committee member Dr. Thuan Chu for all of their guidance throughout my program. I acknowledge the funding support of the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Department of Geography and Planning, University of Saskatchewan. Thank you to the Saskatchewan Wildfire Management Branch (Saskatchewan Ministry of Environment) and the Canadian Forest Service (Natural Resources Canada) for providing data. I thank my partner and family for their encouragement and support, and my lab mates for all of their feedback, support, and friendship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS, ABBREVIATIONS, AND DEFINITIONS

AUC	Area Under the Curve
BUI	Buildup Index
bui	Buildup Index variable
C1	Spruce-Lichen Woodland fuel type
C2	Boreal Spruce fuel type
C3	Mature Jack Pine fuel type
C4	Immature Jack Pine fuel type
CaPA	Canadian Precipitation Analysis
CBI	Composite Burn Index
CFFDRS	Canadian Forest Fire Danger Rating System
CWFIS	Canadian Wildland Fire Information System
D1	Leafless Aspen fuel type
D2	Leafed Aspen fuel type
DC	Drought Code
dc	Drought Code variable
DMC	Duff Moisture Code
dmc	Duff Moisture Code variable
FBP	Fire Behavior Prediction
FFMC	Fine Fuel Moisture Code
ffmc	Fine Fuel Moisture Code variable
FN	False Negative
FP	False Positive
FWI	Fire Weather Index
fwi	Fire Weather Index variable
gstat	Spatial and Spatio-Temporal Geostatistical Modelling, Prediction, and Simulation
ha	hectares
IDW	Inverse Distance Weighted
ignition	For this study, an ignition event is the sudden addition of heat above ambient temperature from an external source (e.g. lightning strike, or burning match); it is required for initiation of a wildfire, but does not always lead to a wildfire

ISI	Initial Spread Index
isi	Initial Spread Index variable
Landsat	Land Satellite
M1	Leafless Boreal Mixedwood fuel type
M2	Leafed Boreal Mixedwood fuel type
O1	Grass fuel type
OK	Ordinary Kriging
precip	Precipitation variable
psill	Partial Sill
RF	Random Forest
rh	Relative Humidity variable
S1	Jack Pine Slash fuel type
SCWRAP	Saskatchewan Community Wildfire Risk Assessment Project
SK WFM	Saskatchewan Wildfire Management Branch
SPOT	Satellite Pour l'Observation de la Terre
temp	Temperature variable
TN	True Negative
TP	True Positive
wildfire	For this study, a wildfire occurs when fuel combustion results from an ignition and spreads away from the ignition point. It can grow to be varying size before becoming extinguished naturally, with human intervention, or a combination of both.
ws	Wind Speed variable

1 INTRODUCTION AND LITERATURE REVIEW

Wildfires are an important ecological process in many ecosystems, driving evolution, biodiversity, and carbon balance (Bond-Lamberty et al., 2007; Canada, 2020). At the same time, wildfires threaten human life, infrastructure, and economic opportunities (Taylor & Alexander, 2006). In ecosystems like the Canadian boreal forest, large wildfires (i.e. those growing to greater than 200ha) comprise a small fraction of the number of fires annually but are responsible for burning the greatest amount of area (Stocks et al., 2002). Large wildfires also pose the greatest risk to human and community safety (Johnson et al., 2008). Under a changing climate, wildfire risk is projected to increase in Saskatchewan (Kafka et al., 2001), and across most of western Canada due to increased temperatures and ignition events (Flannigan et al., 2009). Lightning strikes and human activity are the sources of wildfire ignitions, but the vast majority of ignitions do not lead to large wildfires (Stocks et al., 2002). Furthermore, the range of conditions determining large fire occurrence probability following an ignition event have yet to be identified (Figure 1. 1).

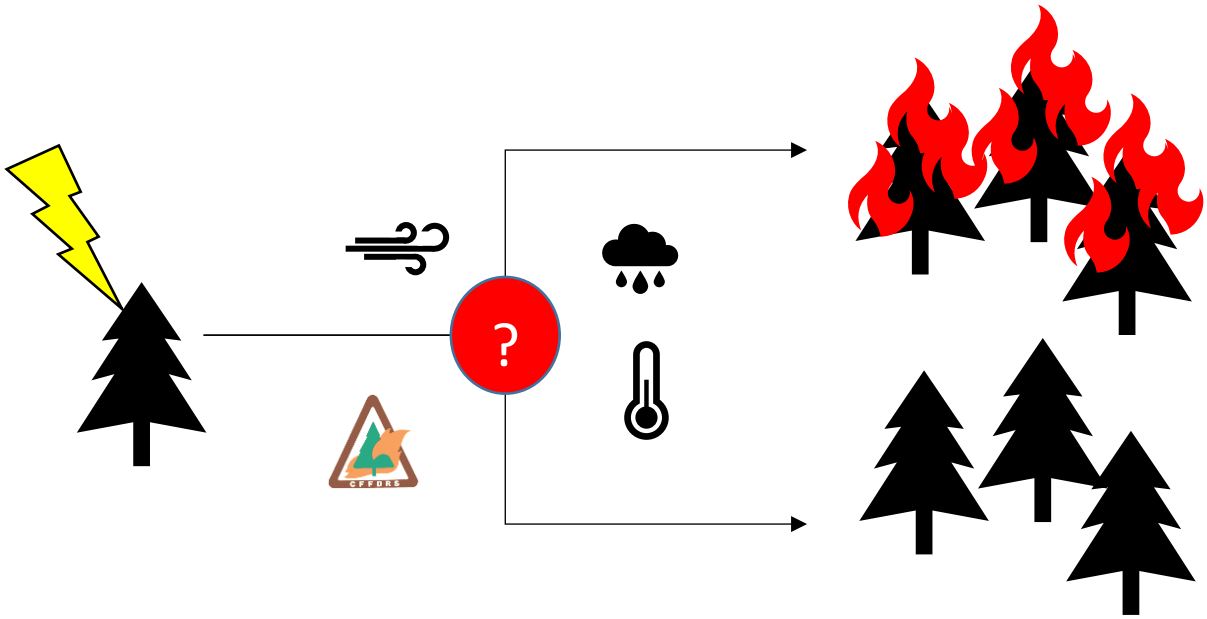


Figure 1. 1 Conceptual graphic illustrating the knowledge gap (white question mark) I will be addressing in my thesis. For simplification only a lightning strike is shown, however I considered human-caused ignitions as well. The four symbols surrounding the question mark represent important drivers in fire behavior, and are clockwise from top left wind, precipitation, temperature, and fuels (captured by the Canadian Forest Fire Danger Rating System). These elements should also play a role in large wildfire occurrences, however their individual contributions relative to each other are as yet unknown.

Conflicting evidence has been presented about whether climatic factors (Bessie & Johnson, 1995; Johnson et al., 1998) or fuels (Cumming, 2001; Renkin & Despain, 1992; Van Wagner, 1977) are the most important drivers of wildfire frequency. Results suggest that both climatic and fuel factors are determinants, in addition to topoedaphic and human factors. But the complex interactions of these factors and the spatially varying constraints on fire activity at the global scale (Krawchuk & Moritz, 2011) show a need for evaluating the drivers of wildfire at a local or regional scale in order to identify when conditions are suitable for large wildfires to occur in the future. Models have been developed to predict future wildfire occurrence using many of the same factors already identified as governing fire behavior (Adámek et al., 2018; Krawchuk et al., 2016). But these models tend to include the probability that an ignition event will occur, in addition to the probability that it will turn into a wildfire. Understanding under what climatic, fuel load, and fuel type conditions wildfires are likely to spread once ignition has occurred should help fire management agencies apply fuel treatments in high risk areas, and be prepared when conditions are favorable for large wildfires to occur (Lawson & Armitage, 1996). This will ultimately reduce the impact of wildfire on humans while allowing this important ecological process to persist.

Forest fire danger research has a rich history in Canada. Starting in the 1920's, James G. Wright and Herbert W. Beal used field and lab test fires to investigate the relationships between weather, fuel moisture, and fire behavior (Van Wagner, 1990). They identified threshold temperatures and levels of fuel moisture that would result in the combustion of fuels (particularly the duff layer), and recognized the importance of fuel moisture in governing fire behavior (e.g. that more energy was required to ignite fuels with higher moisture content, and that fuels with enough moisture would not ignite under natural conditions (Wright, 1932)). They also ascertained that different species of trees have different levels of flammability due to the resin content of the material and how much caloric energy it contains (resinous fuels contain more energy that can be released by fire, providing more heat to ignite adjacent fuels, leading to sustained combustion). These concepts were up to this point in time loosely understood but had not yet been quantified.

Many of the field methods Wright developed laid the foundation for future wildfire research. Importantly, he started relating fuel moisture content to weather observations (wind, temperature, precipitation, and relative humidity) recorded at a central weather station, as

opposed to having to measure fuel moisture content manually at each point of interest (Wright, 1932). This allowed for a more scalable estimation of fuel moisture, and thereby extended the spatial extent that fire danger could be estimated at. Because dead fuel material absorbs moisture during rainfall, and loses moisture to evaporation as it tends towards an equilibrium with the surrounding atmosphere, Wright developed indices (e.g. the tracer index (Wright, 1937)) to track the moisture content of fuels over time based on weather observations and the wetting and drying rates of various fuels. According to Van Wagner (1990), several of the indices Wright and Beal developed evolved into key components of the Canadian Forest Fire Weather Index (FWI) System.

Building on the idea that fuel moisture governs to a large extent the behavior of fire, the Canadian Forest Fire Weather Index (FWI) system (Van Wagner, 1987) was developed. It is comprised of a set of six indices that track the wetting and drying cycle of fuels based on daily weather station observations. Three indices estimate the moisture level of different fuel strata, two indices represent the rate of spread and amount of fuel available, and the over-arching Fire Weather Index combines all 5 sub-indices into one index that represents the intensity a fire would burn at.

At the same time as research into weather and fuel moisture as they relate to fire behavior was continuing to evolve, in the 1970's, researchers began to focus on investigating in greater depth the effects of fuel type (e.g. tree species) on various aspects of fire behavior. Aspects studied included fire intensity, fuel consumption, and rates of spread and behavior of head, flank, and back fires. This research culminated in the Canadian Forest Fire Behavior Prediction System (FBP) (Forestry Canada Fire Danger Group, 1992), which defines the effects on fire behavior of sixteen different fuel types spanning most of the boreal forest in Canada. The final version of the FBP system incorporates components of the FWI system (FFMC, ISI, and BUI), as well as fuel type, wind speed and direction, topography, leaf moisture content, and time and type of ignition to produce several outputs. The primary outputs include the rate at which fire is predicted to spread, how much fuel will be consumed, the intensity of head fire, and a description of the fire type (i.e. crown fire vs. surface fire). These values are specific to each fuel type and were designed to help forest fire management agencies understand and predict the potential behavior of fires and the relative difficulty of controlling them.

The FWI and FBP systems were developed out of the motivation to characterize forest fire danger on a daily basis (Forestry Canada Fire Danger Group, 1992; Van Wagner, 1987). The FWI system's focus is on generating a uniform fire danger index throughout Canada based solely on weather (Van Wagner, 1987), while the FBP system addresses fire behavior and how it varies among different fuel types (Forestry Canada Fire Danger Group, 1992). Around the 1990's, new areas of research on wildfires started to evolve. Researchers analyzed historical fire patterns to understand at larger spatial and temporal scales what physical processes were governing forest fire behavior. Conflicting evidence was provided about whether fuels (Cumming, 2001; Renkin & Despain, 1992; Van Wagner, 1977) or weather (Bessie & Johnson, 1995; Johnson et al., 1998) played a more important role in determining fire frequency, distribution, and activity. Later on, work by Krawchuk and Moritz (2011) showed that both fuels and weather play important roles in fire behavior, and the relative importance of each varies by fuel type and geographic region.

More recently, many researchers have been investigating how to assess the impact of fire on the landscape after it has burned. Burn severity studies have commonly used the Composite Burn Index (CBI; (Key & Benson, 2006)) to determine the impact (e.g. mortality rates of live trees, amount of combustion, and severity) of fire on all forest strata from soil to canopy. With the increased availability of high spatial and temporal resolution satellite-based remotely sensed data, and to address the limitation of resources available for collecting enough ground survey data to capture the variation of burn severity across the landscape, researchers have increasingly been using remotely sensed data to estimate burn severity and to determine its drivers (e.g. Birch et al., 2015; Boucher et al., 2017; Krawchuk et al., 2016; Whitman et al., 2018). Important drivers include fuel type, weather, topographic, and human factors, however the complexity of interactions between these factors and the fine spatial scale of burn severity variability (Whitman et al., 2018) limit the applicability of burn severity to wildfire risk assessment and prediction.

A key area of focus for wildfire researchers has been predicting the future locations of fires at a landscape scale (e.g. Adámek et al., 2018; Krawchuk et al., 2016; Sun & Zhang, 2018, see Abdollahi et al., 2018 for a list of others). A common approach is to classify small parcels of land within the area of interest into varying degrees of fire risk based on important variables (both remotely sensed and/or from ground observations) identified in the literature (such as fuel type, fuel moisture, weather, and ignition type), and to plot the number of historical fires falling into each category of risk. One potential issue with this approach is that researchers trying to

predict future fires often group the probability that an ignition will occur (i.e. that either lightning will strike or a human will light a fire at a certain place on the ground) with the probability that it will turn into a fire. Ignitions have been shown to be non-randomly occurring (Anderson et al., 2000), but their probabilities of occurrence at a given time in a given location tends to be extremely low. The inclusion of ignition probabilities into wildfire occurrence models could therefore confound the understanding of the mechanism behind the phenomenon. A need exists for identifying the range of biotic and abiotic conditions (e.g. weather, fuel type, and fuel moisture content) that will result in a wildfire occurrence should an ignition event take place, regardless of the probability an ignition event will occur.

Sustained ignition probabilities have been assessed in the past (e.g. Lawson et al., 1993, and Lynham & Martell, 1989; Rothermel, 1972 in Lawson et al., 1993) but these were small-scale studies using test fires, and their applicability to landscape scale fire conditions and behavior is unknown. By analyzing historical naturally occurring wildfires, the conditions conducive to wildfire development at a landscape scale can be identified. They can then be compared with test fire results, giving wildfire managers another tool in wildfire occurrence prediction. Some work has been done to identify the probability of a wildfire occurring given an ignition (see Wildfire Ignition Probability Predictor from Lawson & Armitage (1996)), but to my knowledge, this was never expanded outside of British Columbia or developed further.

The purpose of my thesis is to investigate the biotic and abiotic factors and conditions determining if an ignition will lead to a large wildfire occurrence in Saskatchewan's boreal forest. My objectives are to a) analyze historical wildfire occurrences and patterns of what burned and when; b) identify and select variables that have been used to explain wildfire behavior; c) determine which variables are best suited to explaining wildfire occurrences by building a model; and d) apply the model as a lens through which to view historical wildfire occurrences to understand what will drive future occurrences in the study area.

2 STUDY AREA

The study area for my thesis is the boreal forest of Saskatchewan, Canada. In Saskatchewan, the boreal forest is divided into three ecozones (Figure 2. 1) and seven ecoregions. The northernmost ecozone is the Taiga Shield, and it is further broken down into the Selwyn Lake Upland, and Tazin Lake Upland ecoregions. The Taiga Shield's climate transitions from subarctic in the south to arctic in the north. While a long photoperiod of over 18 hours in the summer provides much sunlight for vegetation growth, the low precipitation, cold temperatures, and short growing season lead to low biodiversity and biomass production in this ecozone. The dominant tree species are jack pine (especially in dryer areas) and black spruce (in wetter areas), with white spruce, balsam fir, tamarack larch, trembling aspen, and white birch also occurring but to lesser extents. Tree stands range from open to closed canopy. The fire cycle ranges from 100 – 200 years in this ecozone (Acton et al., 1998).

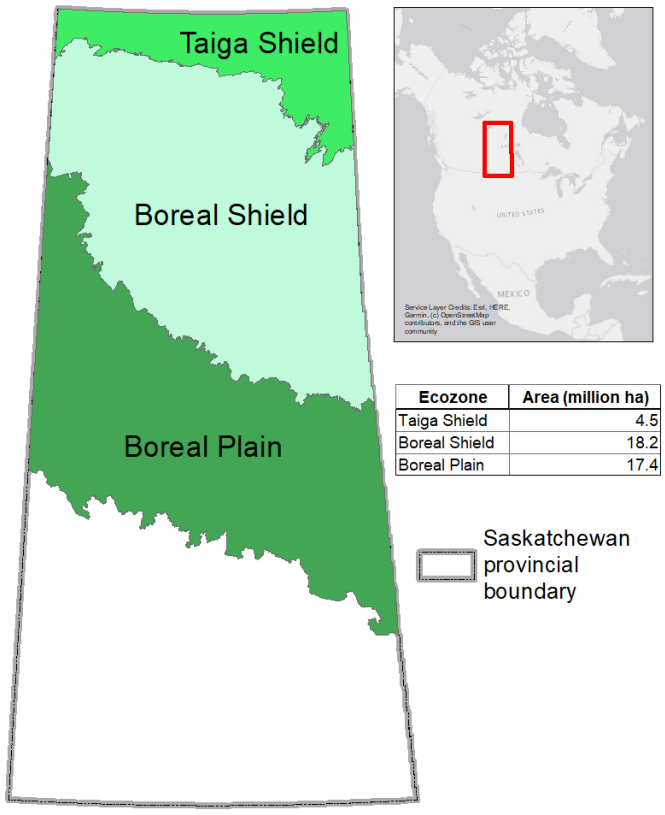


Figure 2. 1 Map of thesis study area – the boreal forest ecozones in Saskatchewan, Canada. Area calculations reported on the map are only for the portion of each ecozone that falls within Saskatchewan.

South of the Taiga Shield is the Boreal Shield ecozone, which is comprised of the Athabasca Plain (northwest) and Churchill River Upland (south east) ecoregions. In the Boreal Shield ecozone, exposed bedrock is interspersed with peatlands and glacial till deposits. As with the Taiga Shield, black spruce and jack pine are the dominant tree species, with closed-crown stands being common. White spruce, trembling aspen, balsam poplar, and white birch, as well as tamarack larch in wetter areas are also found in this ecozone. The climate in the Boreal Shield is warmer and wetter, especially moving to the southwest, compared with the Taiga Shield. It still has long (approximately 7 months), cold winters and a short growing season. The long photoperiod over the summer offsets the shortness of the growing season to some extent, resulting in greater biomass than further north (Acton et al., 1998).

The southern boundary of the boreal forest in Saskatchewan falls within the Boreal Plain ecozone. This ecozone is made up of the Mid-Boreal Upland (northwest, and isolated areas along the west and east edges), Mid-Boreal Lowland (northeast), and Boreal Transitions (south) ecoregions. It has much less exposed bedrock than the Boreal Shield, and is covered instead by thicker glacial deposits. It is warmer and wetter, and has a longer growing period than the Boreal Shield and Taiga Shield ecozones, enabling greater accumulation of biomass and biodiversity. Mixed-wood (e.g. trembling aspen and jack pine or white spruce) stands are more common here than further north, but large stands of conifers, especially white spruce, jack pine, and black spruce, are still prevalent. Peatlands and fens make up most of the wetlands in the Boreal Plain, compared with more bogs in the exposed bedrock dominated Boreal Shield. Farmland dominates the southern edge of the Boreal Plain ecozone with both livestock production and annual cropping being common (Acton et al., 1998).

Forest fires are a major natural disturbance in all three of these ecozones, with the fire cycle ranging from roughly 100 to nearly 300 years (Acton et al., 1998; M. A. Parisien et al., 2004). In the Boreal Plain, the fire regime is highly influenced by human activity through landscape alterations (including logging and agriculture), human-caused fires, and suppression of fires. In the Boreal Shield, pure coniferous forest stands are more prevalent, and human density is lower, leading to a greater proportion of lightning caused fires that also grow to be very large. Human density is also quite low in the Taiga Shield, where conifers also dominate. Forest stands are distributed more sparsely here, but the long photoperiod in the fire season allows for more drying time of fuels, making this ecozone still very prone to fires (M. A. Parisien et al., 2004).

3 METHODS

A flow chart showing the general methods, software, programming language, and primary package used in my thesis is shown in Figure 3. 1. Each step is described in greater detail in the sections that follow.

3.1 Dataset description

I created the dataset used to address my research purpose and objectives by combining four different datasets: 1) wildfire ignitions from the Canadian Wildland Fire Information System (CWFIS) (Canada, 2019); 2) weather and Fire Weather Index (FWI) data from federally operated weather stations (Canada, 2019); 3) weather and FWI data from provincially operated weather stations (WFM, 2019b); and 4) Canadian Fire Behavior Prediction (FBP) system fuel types (WFM, 2019a).

The wildfire ignitions dataset from CWFIS contains records of where and when wildfires started, when they were declared 'out,' how large they grew to be, whether they were caused by humans or lightning, and in some cases when suppression activities (if any) were first used to contain them. The dataset includes fires from across Canada, with the earliest record coming from British Columbia (1930). For Saskatchewan, the earliest year represented is 1959, although prior to 1981 only large fires (size > 200 ha) were recorded. For my thesis I included records from 1990 to 2018, and that started within the study area (Boreal Plain, Boreal Shield, and Taiga Shield ecozones). I was unable to locate or acquire a dataset of historical lightning strike locations to use as fire not started samples, so instead split the wildfire ignitions dataset into two classes for my analysis: 1) large fires (those which grew to be >= 200 ha); and 2) small fires (those which grew to be a total of less than 200 ha).

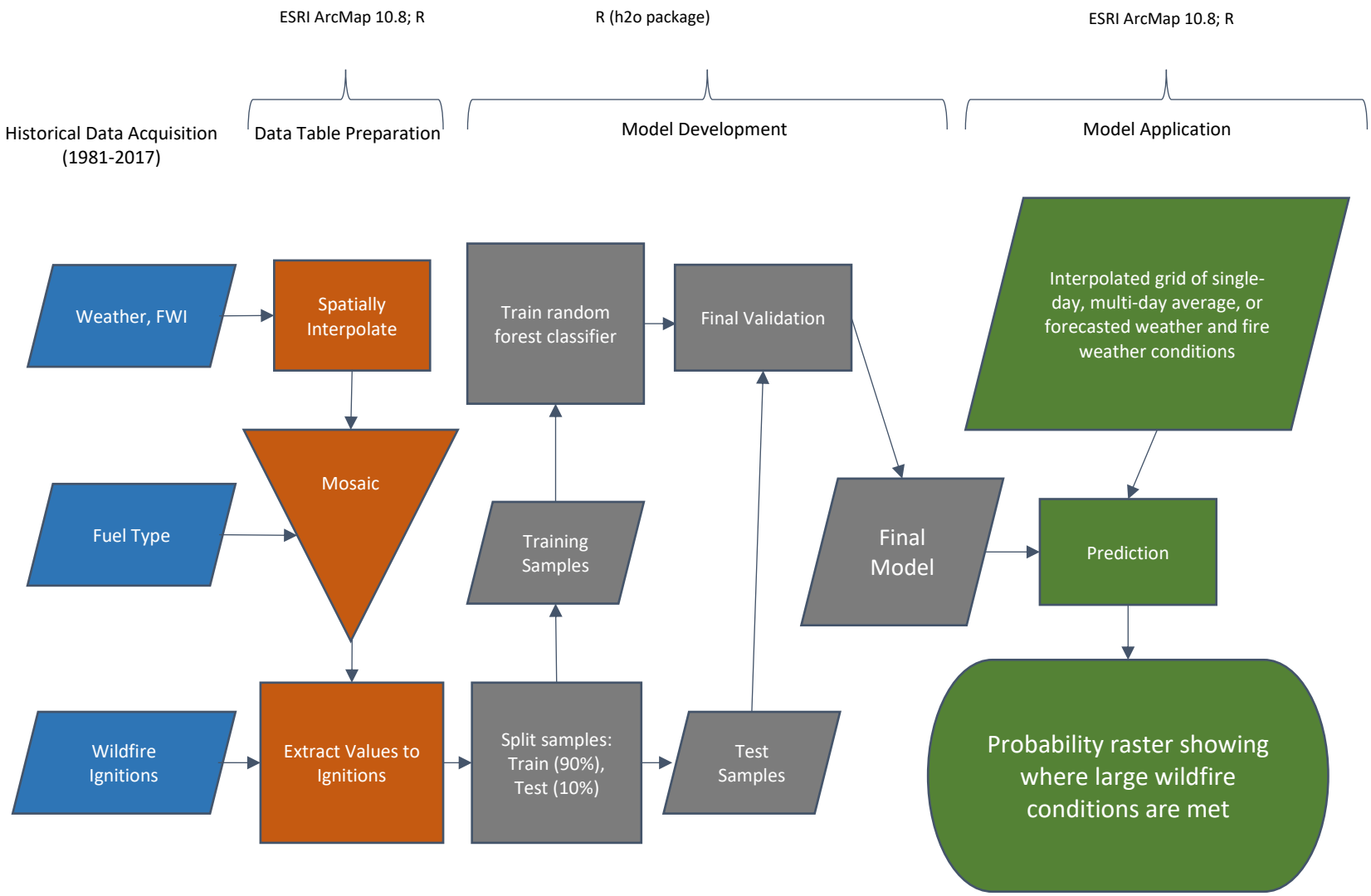


Figure 3. 1 Methods flow chart showing steps involved in the analysis for my thesis, including data acquisition (blue), data table preparation (red), model development (grey), and model application (green).

In the ignitions dataset, small fires outnumbered large fires by nearly 8.5 times (21169 small fires: 2491 large fires). For a classification problem, this creates what is known as an unbalanced dataset, and it can cause spurious results if not addressed. There are several ways to address an unbalanced dataset during analysis which will be discussed later. Having an unbalanced dataset also affords the opportunity to carefully select records of interest to aid in explaining the observed phenomenon prior to analysis (King & Zeng, 2003). In order to create more separation in observed values of my explanatory variables between the dependent variable values (ignitions that led to large fires and ignitions that did not), I removed from the dataset fires with sizes between 0.5 and 200 hectares. Fires that became greater than 200 ha were still classified as large fires, and small fires were re-classified as being those which did not grow larger than 0.5 ha. Fires of intermediate size (> 0.5 ha and < 200 ha) could still be predicted with the classifier, but the output would simply be whether each fire's conditions appear more like a small fire or a large fire.

To account for the potential influence of human intervention on eventual fire size, I filtered out small fires that occurred within full response zones (i.e. within 20 km of a community (K. Johnson et al., 2008)). In the full response zone, any wildfire receives a full suppression response as soon as it is reported (K. Johnson et al., 2008). I did not filter out large fires that started within a full response zone, since they still grew to be large fires despite any potential suppression activities. These fires may therefore have had very high large fire risk conditions.

After paring down the dataset, a total of 3858 records remained (2422 Small Fires; 1436 Large Fires). I then grouped these records by fire size and month, as well as by fuel type, to understand the distribution in my dataset of large and small fires by categorical variables by time throughout the fire season.

Fuel type rasters were provided by the Saskatchewan Wildfire Management Branch. They were created by classifying land cover using both Landsat and SPOT satellite imagery, and then mapping the land cover types to the Fire Behavior Prediction (FBP) (Forestry Canada Fire Danger Group, 1992) fuel types (S. Nicholson, personal communication, May 11, 2020). In Saskatchewan, only 8 of the fuel types exist (Table 3. 1), with aspen (D1/D2) and boreal mixed-wood (M1/M2) each having two phenological stages – leafed and leafless. The years for which fuel type rasters were created for Saskatchewan are 2007, 2008, 2014, 2015, 2018 and 2019.

Table 3. 1 Codes and descriptions of various physical aspects typical of each different FBP fuel type that is found in Saskatchewan. This table was adapted from Table 3 (pp 12 – 13), (Forestry Canada Fire Danger Group, 1992).

Fuel type code	**Description**	**Forest floor and organic layer**	**Surface and ladder fuels**	**Stand structure and composition**
C1	Spruce-Lichen Woodland	Continuous reindeer lichen; organic layer absent or shallow, uncompacted.	Very sparse herb/shrub cover and down woody fuels; tree crowns extend to ground.	Open black spruce with dense clumps; assoc. sp. Jack pine, white birch; well-drained upland sites.
C2	Boreal Spruce	Continuous feature moss and/or *Cladonia*; deep, compacted organic layer.	Continuous shrub (e.g., Labrador tea); low to moderate down woody fuels; tree crowns extend nearly to ground; arboreal lichens, flaky bark.	Moderately well-stocked black spruce stands on both upland and lowland sites; *Sphagnum* bogs excluded.
C3	Mature Jack Pine	Continuous feature moss; moderately deep, compacted organic layer.	Sparse conifer understory may be present; sparse down woody fuels; tree crowns separated from ground.	Fully stocked jack or lodgepole pine stands; mature.
C4	Immature Jack Pine	Continuous needle litter; moderately compacted organic layer.	Moderate shrub/herb cover; continuous vertical crown fuel continuity; heavy standing dead and down, dead woody fuel.	Dense jack or lodgepole pine stands; immature.
D1/D2	Aspen (leafed (D2)/leafless (D1))	Continuous leaf litter; shallow, uncompacted organic layer.	Moderate medium to tall shrubs and herb layers; absent conifer understory; sparse, dead, down woody fuels.	Moderately well-stocked trembling aspen stands; semimature; leafless (i.e. spring, fall, or diseased), or leafed (i.e. summer)
S1	Jack Pine Slash	Continuous feather moss; discontinuous needle litter; moderately deep, compacted organic layer.	Continuous slash, moderate loading and depth; high foliage retention; absent to sparse shrub and herb cover.	Slash from clearcut logging; mature jack or lodgepole pine stands.

O1	Grass	Continuous dead grass litter; organic layer absent to shallow and moderately compacted.	Continuous standing grass (current year crop). Standard loading is 0.3 kg/m^2, but other loading can be accommodated; percent cured or dead must be estimated. Sparse or scattered shrubs and down woody fuel. Subtypes for both early spring matted grass and late summer standing cured grass are included.	Scattered trees, if present, do not appreciably affect fire behavior.
M1/M2	Boreal Mixedwood (leafed (M2)/leafless (M1))	Continuous leaf litter in deciduous portions of stands; discontinuous feather moss and needle litter in conifer portions of stands; organic layers shallow, uncompacted to moderately compacted.	Moderate shrub and continuous herb layers; low to moderate dead, down woody fuels; conifer crowns extend nearly to ground; scattered to moderate conifer understory.	Moderately well-stocked mixed stand of boreal conifers (e.g., black/white spruce, balsam/subalpine fir) and deciduous species (e.g., trembling aspen, white birch). Fuel types are differentiated by season and percent conifer/deciduous sp. Composition.

Weather and fire weather index data are provided by two different organizations in Saskatchewan. The CWFIS stores data collected at weather stations operated by Environment and Climate Change Canada, while the SK WFM branch manages and provides data from their own weather stations. Both organizations follow the same standards for data collection and index calculations, as prescribed in (Lawson & Armitage, 2008). In brief, weather observations are recorded at 12 noon local time at each meteorological station. FWI values are calculated from those observations, as well as the previous day's index values from the same weather station. Occasionally a weather station may be inoperable for a period of time, and in that case, values for these days are interpolated from nearby weather stations by the data provider to ensure a complete dataset. The weather variables I used in my analysis are shown in Table 3. 2, and the FWI variables are shown in Table 3. 3.

Table 3. 2 Weather variables used in analysis, their abbreviations, and units.

Variable Name	Abbreviation	Units
Temperature	temp	^{o}C
Precipitation	precip	mm
Wind Speed	ws	km/h
Relative Humidity	rh	%

Table 3. 3 Fire Weather Index component variables, their abbreviations, time lags (where applicable), and ranges of possible values.

Variable Name	Abbreviation	Time Lag (days)	Value Range
Fire Weather Index	fwi		0-80*
Initial Spread Index	isi		0-60*
Buildup Index	bui		0-170*
Drought Code	dc	52	0-800*
Duff Moisture Code	dmc	12	0-150*
Fine Fuel Moisture Code	ffmc	2/3	0-99

* Range is open-ended, upper limit values represent highest probable value (De Groot, 1987).

The FWI system (Figure 3. 2) is comprised of three different codes that are calculated directly from weather observations and represent the moisture content of fuels. From these moisture codes, three fire behavior indices are calculated, which give a relative idea of how fast a fire should spread, how intensely it will burn, and how difficult it will be to control (De Groot, 1987). The fine fuel moisture code (FFMC) tracks the moisture content of cured small diameter fuels and litter on the forest floor, and is calculated from temperature, relative humidity, wind, and rain. It includes a time lag of 2/3 day, whereby any precipitation in the prior 16 hours will affect the moisture content of these fine fuels. The duff moisture code (DMC) represents the moisture content in the top organic soil layer which is loosely compacted. DMC is calculated from temperature, relative humidity, and precipitation. It includes a time lag of 12 days to account for the rate at which these fuels are able to dry once wet. The drought code (DC) is calculated from temperature and rain. It represents the moisture content in the deep organic layer which is compacted. It includes a time lag of 52 days to account for its drying rate. These three

codes are then combined into secondary indices. Initial spread index (ISI) combines wind and FFMC, and represents how fast a fire will spread independent of fuel load. The buildup index (BUI) is a combination of DMC and DC, and represents how much fuel is available to be burned. BUI and ISI are then combined to form the fire weather index (FWI), which represents how much energy would be released per unit length of the fire front, and is akin to the intensity of the fire (Van Wagner, 1987).

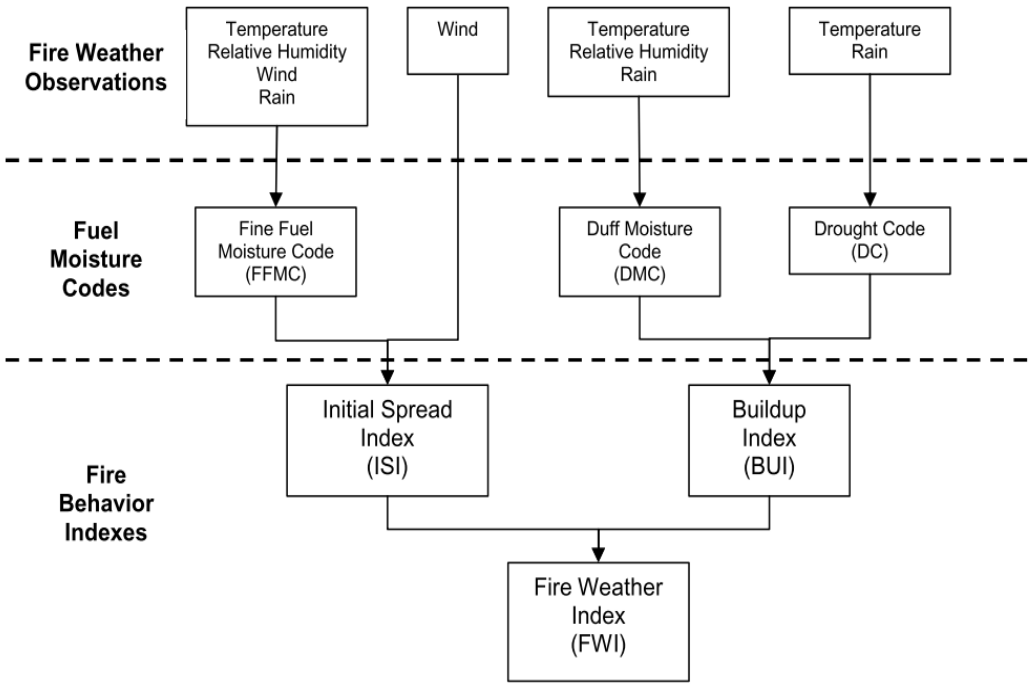


Figure 3. 2 Components of the FWI system and how they are used to calculate subsequent indices (De Groot, 1987).

The broad approach for assembling my dataset was to associate each wildfire record with fuel types, weather observations, and FWI values from the same time and location of the fire's ignition. Weather and FWI data are recorded at weather stations. Since these variables vary spatially, and not all fires are ignited right beside a weather station, the values at ignition locations had to be estimated from the weather station point locations. To accomplish this, I used spatial interpolation.

3.2 Interpolation of weather and fire weather data

In interpolation, unknown values are estimated from the values of nearby (e.g. in time, space, or more generally, feature space) known values. Three conditions must be met in order for spatial interpolation to be an applicable method: 1) there must be a measurable value at every location, 2) there can only be one value at each location at a given time, and 3) there must be non-zero auto-correlation in the variable of interest (i.e. locations that are close together are more similar than those that are farther apart, commonly known as Tobler's first law of geography (Tobler, 1970)).

Different methods exist for how to interpolate values, and in wildfire research in Canada, many of the methods have been compared for their performance in estimating weather and fire weather indices (see Flannigan & Wotton, 1989; Jain & Flannigan, 2017). For my project I used Ordinary Kriging (OK), which was the best performing method in (Jain & Flannigan, 2017). Ordinary Kriging, like other interpolation methods such as Inverse Distance Weighted (IDW), relies on estimating the values at unknown locations by calculating weighted means from nearby known values.

In OK interpolation, the weights used to estimate a value at an unknown location are calculated based on a mathematical model of the data representing to what degree the variable covaries over distance. The mathematical model is built by creating a variogram of the data, which is accomplished by calculating the distance between all pairs of known observations in distance bins, and the difference in values (semivariance) of all pairs to produce the weight (γ) in each bin by:

$$\gamma(h) = \frac{1}{2N_h}\sum_{i=1}^{N_h}[z(r_i) - z(r_i + h)]^2$$

Equation 1

where N is the number of pairs in each bin of distance h, and z is the observed value at location r_i or $r_i + h$. A model (or best fit line) minimizing the residual error is then fit to the entire variogram. This model is used to calculate weights for all known values in the weighted mean calculation for estimating unknown values.

Many different models can be fit to the variogram, but I chose the spherical model, which produced the best performance in (Jain & Flannigan, 2017). Kriging interpolations were performed using the gstat package (Pebesma, 2004) in R.

There are several model parameters which can be tuned in OK. In the same way slope and intercept are 'learned' from the data in a regression problem, the range, nugget, and sill can be learned while fitting a best fit line to a variogram. Range is the distance beyond which a point should no longer have an influence on another point. Lawson & Armitage (2008) note that fire weather index values are not reliable outside of a 160 km radius due to the spatial variation in weather variables used to calculate them, so I set the range value to 160 km for all variables. Nugget captures the possibility of measurement error and random variation in samples by allowing for semivariance in the lowest distance class. The starting nugget value I used was 1.0, and I set the algorithm to fit the nugget to the data. The partial sill (psill) is the semivariance value where distance equals the range, minus the nugget. I used an initial psill value of 1.0 and set the algorithm to learn this value as well.

Several different methods for fitting a line to the variogram are available in gstat (Pebesma (2014), see Cressie (1985) for theoretical background). The default is to use a weighted least squares fit. The weights of distance classes in the variogram are calculated as follows:

$$weight = \frac{N_j}{h_j^2}$$

Equation 2

where N_j is the number of pairs of points in distance class h_j. Occasionally using this method, a psill of 0 would be learned (i.e. a singular model produced), resulting in a horizontal line of best fit (i.e. suggesting zero auto-correlation in the data, and that all weather observations within 160km should be weighted equally in calculating the mean used for estimating the unknown point). While weather observations can sometimes vary very little over hundreds of kilometers, it is highly unlikely that all observations within this distance are exactly equal (except perhaps for precipitation if zero rainfall occurred at all stations). In order to find a weight function that resulted in a non-zero psill, I therefore implemented an iterative approach in method selection for kriging such that each method (Table 3. 4) would be used until one producing a psill <> 0 was

found. If all methods produced a psill of 0, the psill of 0.01 would be used, and the estimation of unknown values for that variable on that day for that fire ignition would proceed.

Table 3. 4 gstat fit method codes, their weights, and the order in which they were used in the variogram fitting algorithm. If the first four methods yielded a psill of 0, the fifth method (equivalent to fit code 0) was used and the psill was manually set to 0.01.

Order	Fit Code (gstat)	Weight
1	7	N_j/h_j^2
2	6	no weights
3	1	N_j
4	2	$N_j/\{\gamma(h_j)\}^2$
5	0	manually set psill to 0.01

Once a model was successfully fit to the variogram, I used it to predict the value for the variable at the ignition location. I used this same procedure to interpolate all variables for each fire ignition. I also followed the same Kriging interpolation procedures each time on a scaled (z-score) version of each variable for each ignition location to allow comparison of interpolation variance (i.e. estimation error due to interpolation) among variables and among ignitions. By scaling to z-score, the interpolation variance for different variables can be compared to see how accurately each variable can be interpolated relative to other variables, and the summed variance for all variables at an ignition location can be compared with other ignition locations. This enables the determination of what influence interpolation error has on variable importance and model performance.

3.3 Fuel type

To associate fire ignitions with the fuel type they started in, I selected the most recent fuel type year prior to the year of fire ignition (because for large fires, fuel types post fire should be burned and be classified as non-fuels). For ignitions from 2007 and earlier, I selected the 2019 layer to extract fuel types from because it would have allowed the most amount of time for a forest stand to return to its pre-fire conditions according to natural succession. This was a simplifying assumption, and its validity is considered in the discussion section of this thesis.

Several ignitions received non-fuel types (such as water). For small fires, I removed these data points because of the over-abundance of this class of data. For large fires, I used a

combination of satellite imagery interpretation and visual analysis of fuel type layers to determine what fuel type would be most appropriate for these ignitions. Several of these fires started on the shore of a waterbody, where most of the fuel type pixel contained water and was classified as such. For these cases I selected the nearest non-water fuel type from the appropriate fuel type layer.

The mixedwood (M) and deciduous (D) fuel types are each broken down into a leafed and leafless stage because of the impact green leaves have on the flammability of these fuels (Forestry Canada Fire Danger Group, 1992). The cutoff dates I used to distinguish between these stages were May 9 and September 28 (leafless (M1 and D1) is prior to May 9 and after September 28; leafed (M2 and D2) is between these dates). These dates are the average green up and end of senescence dates in the boreal plain ecozone of Saskatchewan from 1996 to 2003 (Barr et al., 2004). Average green up and end of senescence dates were not available for the other two ecozones in my study.

3.4 Fire history description

To understand the history of large wildfires in my study area, I downloaded the large fire polygons dataset from CWFIS online (Canadian Forest Service, 2019). This dataset has polygon delineated burn boundaries for all large fires in Canada since as early as 1917 (in British Columbia), and from 1945 to 2018 for Saskatchewan. Using this dataset, I determined the total area burned in each ecozone, and for each fuel type from all large fires that burned completely within Saskatchewan from 2008 to 2018. I also determined how much of each fuel type burned in each individual fire. I did this by filtering out large fires that occurred prior to 2008, since fuel type layers for the province are not available before 2007 (see section 5.3).

To see whether large fires burned preferentially in any fuel types or ecozones, I summed the total area burned by each fuel type and in each ecozone in each year, and compared that with the total area available to be burned in each category in the same year (burned minus available hereafter). This is a more basic version of the ‘compositions’ analysis of large fires performed by (Cumming, 2001). Since this analysis was not the focus of my thesis, I did not log-ratio transform the compositions or perform parametric statistical testing to determine whether the compositions of burned areas were significantly different from the available compositions. But

the visual comparison is quite compelling and useful for understanding wildfire dynamics in the study area nonetheless.

3.5 Classification

The classification algorithm I chose for investigating my objectives is random forest (Breiman, 2001). Random forest is an ensemble method based on the decision tree. Decision tree classifiers are grown by taking samples of the data (both records and variables) and iteratively creating binary splits (a.k.a. nodes) on variables where values greater than the split correspond with one class and values less than the split correspond with another class. Splits can also be created for non-numerical data types by grouping categorical variables into two distinct groups, each associated with one of the classes. At each node, both the variable selected and the threshold used are those out of all possible combinations that maximizes information gain. For classification problems, the maximum information gain is typically assessed by using one of two methods: Gini or Entropy (Bishop, 2006). For regression decision tree problems, the decision criteria for splitting nodes is based on which split (both variable and value) results in the lowest mean squared error. The h2o package (H2O.ai, 2020a) uses regression trees for all classification problems, so each split is based on minimum mean squared error.

Nodes are continually created until one of several stopping criteria is met (e.g. minimum improvement in mean squared error at a split; see parameter descriptions in Appendix A for more information). The maximum number of nodes in a single branch of a tree is called the tree depth. Having too great a depth can lead to overfitting due to increased complexity and fitting to the ‘seen’ data, so growing more trees in the random forest is preferred over growing trees to greater depths to avoid this problem and make the classifier more generalizable (Cook, 2016). With random forest, many decision trees are grown following these same steps but on different subsets of the training data. Each individual tree is seen as a weak predictor on its own, but when taken together, the forest produces a strong case for an observation belonging to a given class. The predicted class for a given record is the class receiving the majority of votes across all trees in the forest, based on the values of all of its variables. By repeatedly selecting different subsets of the data when growing the forest, each tree is identically distributed, and casts a vote of equal weight to all other trees. Once grown, the forest can be used to make predictions on unseen

observations, to determine what class they belong to based on how their attributes compare to what the model has learned from the training data.

3.5.1 Model development

To build and test my classifier, I used the h2o package (H2O.ai, 2020a) in R (R Core Team, 2020), and the Distributed Random Forest algorithm. This package was chosen specifically for several reasons. It provides flexibility in dealing with categorical variables that other packages do not. Other common machine learning packages require categorical variables be transformed to binary attributes (a.k.a. one-hot, or dummy variable encoding). One-hot encoding increases the number of variables in the dataset which has a computational and computer memory cost, and may exacerbate the 'absent levels' problem (Au, 2018). The h2o package, as mentioned before, uses regression trees for classification, producing probability values that each observation belongs in its predicted class. This provides more flexibility in setting probability thresholds for determining cutoff values when deciding what class a predicted record should belong to, in addition to having more options to deal with unbalanced datasets directly.

In order to help assess the generalization ability of my classifier, I first set aside 10% of the records (randomly selected without replacement) for final testing. These records were not seen by the classifier until it had been completely built. I then split the remaining 90% of records into a training dataset (80%) and validation dataset (20%) to build the classifier from. For building the classifier, I used random grid search (number of models = 500) and 10-fold cross-validation on each of the 500 models to find the optimal values for several tunable hyperparameters (see Table A1 for the list of hyperparameters and values used in the search). I chose the static values in Table A2 for other hyperparameters based on how they help deal with the imbalance in my dataset and to optimize accuracy while avoiding overfitting.

To determine which was the top model from the 500 that were built, I selected one with the highest average of the recall (a.k.a. sensitivity) metric calculated for both the cross-validation (training) and validation datasets in each model. The equation for recall is:

$$recall = \frac{TP}{TP + FN}$$

Equation 3

where *TP* is number of true positives, and *FN* is false negatives. *TP* and *FN* are calculated from a confusion matrix of predictions made from the model compared with their true classes.

Maximizing the recall minimizes false negatives which is desirable for this use case. When predicting whether conditions are suitable for a large fire to occur given an ignition, having false negatives is a greater threat than having false positives (i.e. saying the conditions are not suitable for a large fire when they really are could be more dangerous than saying conditions are suitable when they actually are not) (Taylor & Alexander, 2006). Given that the predicted results are probabilities between zero and one, by default, the probability threshold that maximizes recall will tend to be close to zero and will result in a recall of 1 (i.e. with a probability threshold of 0, all records will be predicted to be suitable for a large fire, and 0 records would be suitable for not large fire, resulting in recall of 1, simply by saying the conditions are always suitable for large fires). So maximizing recall on its own is not a suitable approach. To alleviate this issue, I used the probability threshold that maximized the mean per class accuracy, and calculated the recall using that threshold, then selected the model with the highest recall.

To assess model performance, I also reported the area under the ROC curve (AUC). AUC is used to indicate how well a binary classifier can distinguish true positives from false positives (H2O.ai, 2021). AUC is a commonly reported model performance metric and I included it to help compare how well the model performed relative to other models.

For evaluating the generalizability of the best model, I used the mean probability threshold produced from the cross-validation and validation parts of model training to predict the classes for the unseen test data (Cook, 2016), and compared that performance with the performance of the validation and cross-validation performance of the same model to see if over-fitting was an issue. If there was a large drop between performance during model training and performance in model testing, this could indicate over-fitting of the model and poor generalizability.

To evaluate whether the explanatory variables in the model can be used to distinguish with statistical significance between large and small fires, I used Press' Q statistic (Maroco et al., 2011). Press' Q uses a chi-square distribution to determine if the classification results on test data are significantly different than random association. The null hypothesis is that the classification results are completely random. The alternative hypothesis is that the classification results are not

random, suggesting the explanatory variables and model are capable of distinguishing between classes. For example, a confusion matrix with equal frequencies of true positives, true negatives, false positives, and false negatives would have a chi-square (1 df) of 0, which is less than the chi-square threshold for p = 0.05 of 3.84. The equation to calculate Press' Q is:

$$Q = \frac{(N - nk)^2}{N(k-1)} \sim \chi^2_{(1)}$$

Equation 4

where *N* is the total number of records predicted (sample size), *n* is the total number of records correctly classified (true positives plus true negatives), and *k* is the number of classes (2, in this case – large fire or small fire).

In order to understand the drivers of large wildfires, I analyzed the variable importance of the top 10 random forest models, to see which variables were most important in explaining whether conditions were suitable for a large wildfire to occur. In h2o, variable importance for a variable in decision tree-based algorithms is calculated as the difference between the squared error at each split using that variable, and the squared errors in its children nodes (H2O.ai, 2020b). Variables that consistently have lower mean squared errors than their children nodes throughout the forest of decision trees will have a higher importance, because their inclusion in a split provides greater information gain than other variables would.

3.5.2 Spatial heterogeneity of variables and observations in relation to model performance and variable importance

My dataset has inherent variation in spatial heterogeneity both among variables (e.g. precipitation and wind speed are difficult to accurately interpolate because of how much they vary over short distances compared with other variables like temperature (Jain & Flannigan, 2017)) and among observations (some observations of fire ignitions occurred farther from weather stations, making the interpolated weather and fire weather index values less accurate than observations near multiple weather stations). I therefore decided to test whether this variation impacted model performance and variable importance, to see whether: a) the model is indicating the most important drivers of large wildfire conditions or simply which variables are easiest to interpolate (i.e. have the least amount of spatial heterogeneity), and b) some

observations of large and small fires are more difficult to predict because of natural variation, or because they were observed farther from weather stations.

To determine the impact of variation in spatial heterogeneity among variables, I plotted the z-score scaled interpolated variances of all interpolated variables against their importance in the top 10 models. I also tested for significance of any influence of spatial heterogeneity on variable importance using a simple linear regression (lm function in R). To determine the impact of variation in spatial heterogeneity among fire ignition observations, I plotted the mean z-score scaled interpolated variances across interpolated variables for all observations in the test dataset against the distance of predicted class probabilities from the threshold determining class split (absolute value of predicted probability minus threshold). I considered the absolute distance from threshold probability value to be an indication of confidence of each prediction. Predicted probabilities near the threshold have less agreement in the model of which class they belong to, while probabilities farther from the threshold have greater agreement in the model, and can be considered more confident predictions. I also tested this relationship using a simple linear regression (lm function in R).

3.5.3 Model application

To help visualize how large fire conditions can vary temporally, I analyzed two different historical wildfire example types for each ecozone using the best model. For high wildfire activity examples, I selected the month which had the greatest number of large wildfire ignitions in each ecozone. For low wildfire activity examples, I selected the month with the highest number of small wildfire ignitions, considering only months that did not have any large wildfire ignitions. These months would have had the most ignition events (i.e. opportunities for a large wildfire to occur) without having the suitable conditions for a large wildfire to occur.

For each of the selected months, I calculated the mean daily weather and fire weather conditions for each weather station across the study area, and then interpolated the means to 100m x 100m gridded rasters (same grid as fuel types) across the respective ecozone using OK in ArcMap 10.8 (ESRI Inc., 2020). For fuel types, I selected the appropriate fuel layer for each month based on what year it occurred in and using the same fuel type determination algorithm defined in section 5.3. I imported the input rasters into R, and used the best model to predict the probability of large wildfire conditions being met at each raster cell. I then saved the predicted rasters to file and imported them into ArcMap 10.8 for visual inspection. I also calculated what

percentage of cells were above and below the probability threshold for large wildfire conditions for each ecozone in each month.

To understand how frequently large wildfire conditions are met for each of the fuel types, I used the best model to predict all weather observations from within the study area during the study period (1990 – 2018) once for each fuel type. I then calculated what percentage of the total number of weather observations (720,738) had large wildfire conditions for each fuel type, and plotted it on a bar graph.

4 RESULTS

4.1 Large fires, fuel types, and ecozones analysis

For large fires that were completely within the study area and that occurred between 2008 and 2018, the average annual difference between area burned and available for each fuel type and ecozone is shown in Figure 4. 1. C3 burned much more than expected based on its availability in the study area. During this time period, over 26,000 km^2 of C3 forest burned, nearly 3 times as much area as the fuel type which burned the second most – C4 (8,217 km^2 burned) (Figure 4. 3a). C4 also burned more than expected based on its availability, but to a lesser degree than C3 (

Figure 4. 1). Likewise, C1 and C2 also burned more than expected, but only slightly. The total area burned of these fuel types in this time period was still considerable (6,483 km^2, and 4,881 km^2, respectively).

The remaining fuel types burned less than expected given their prevalence in the study area. O1 had the second lowest (next to non-fuel) average burned minus expected, but it burned the third greatest total area during this period (7,161 km^2 in 511 fires). Non-fuels, which includes water, also burned a considerable amount (5,711 km^2 in 842 fires).

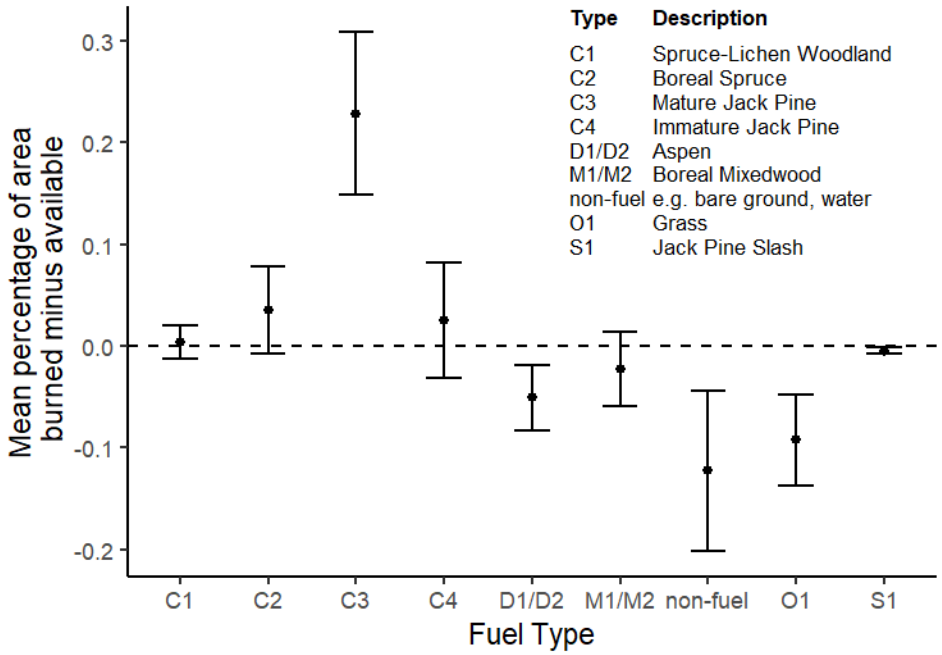


Figure 4. 1 Annual mean percentage of area burned in large fires minus available in the study area for each fuel type between 2008 and 2018. Bars represent +/- 95% confidence interval.

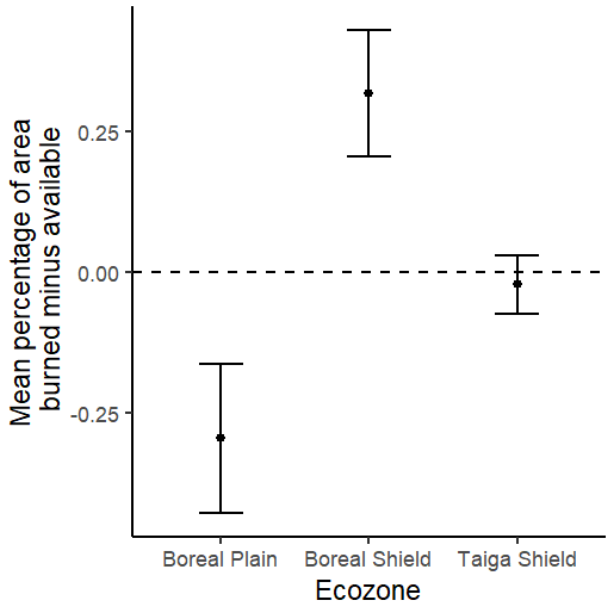
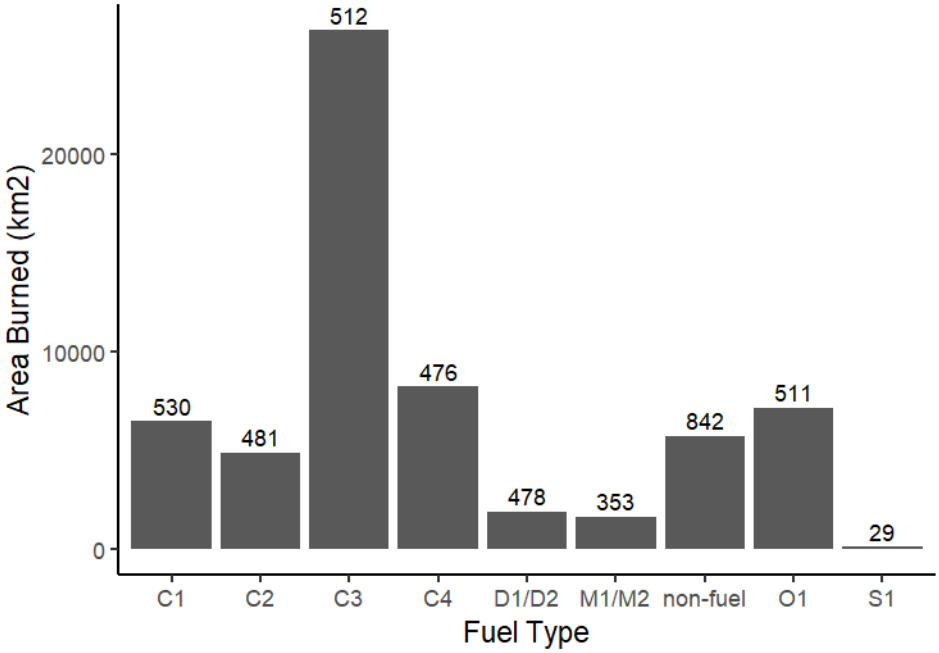


Figure 4. 2 Annual mean percentage of area burned in large fires minus available in the study area for each ecozone between 2008 and 2018. Bars represent +/- 95% confidence interval.

The Boreal Shield ecozone had the greatest amount of annual average area burned minus available for large fires (Figure 4. 2). It also had the greatest total area burned (75,341 km^2), and the greatest number of fires (676) occurring within it compared with the other two ecozones (Figure 4. 3b). Large fires in the Boreal Plain burned on average annually a much smaller area than expected given the size of the ecozone (total area burned = 21,256 km^2 in 274 fires). The Taiga Shield ecozone burned on average an area roughly proportional to what percentage of the study area it covers, with 12,034 km^2 burned in 183 fires during this time period.

a) by fuel type



b) by ecozone

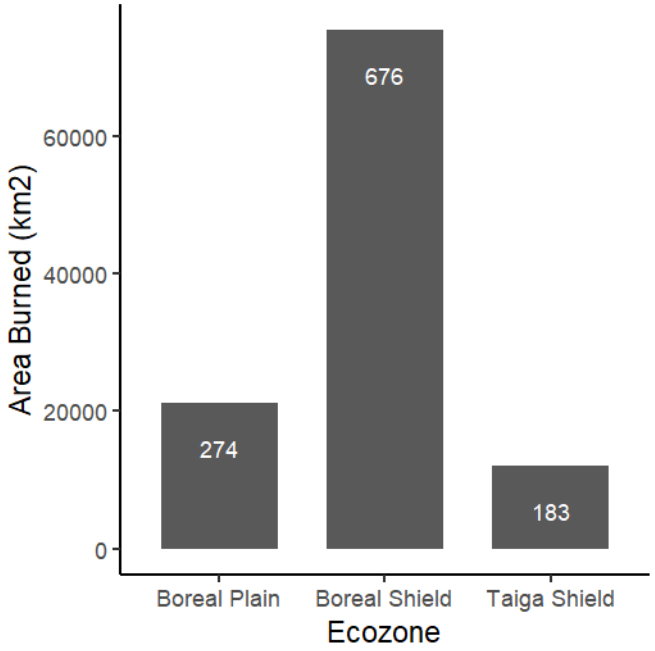


Figure 4. 3 Total area burned in large wildfires in the study area between 2008 and 2018, grouped by fuel type (a), and ecozone (b). Numbers in each bar indicate the number of large fires occurring in each category during this time.

For the fire ignitions dataset used in my classification (fires >= 200 ha, and fires <= 0.5 ha, from 1990 – 2018), the proportion of ignitions that became large fires varied throughout the year, and differed greatly among fuel types, and ecozones (Figure 4. 4). All fuel types had the most ignitions in June, July, and August. Quite a few fires also occurred in May for all fuel types, whereas September and October had fewer and fewer fire occurrences. The proportions of ignitions leading to large fires stayed fairly consistent throughout all months for coniferous fuel types (C1 – C4). For D and M fuel types, the proportion of ignitions leading to large fires appeared to taper off from July onward, likely because of the reduced flammability of aspen during the leaf-on stage that occurs at this time of year. Like with the large fires analysis above, C3 had the greatest amount of fire activity (i.e. ignitions), but is also the most prevalent fuel type in the study area. S1 was the least prevalent fuel type in the study area, and had the fewest ignitions. Most of the S1 ignitions resulted in small fires.

The Boreal Shield and Boreal Plain ecozones had similar numbers of ignitions (n = 1850, and n = 1714), although the proportion of ignitions that became large fires was much greater in the Boreal Shield (47.6% vs. 20.7%). The Taiga Shield had an even greater proportion of large fires (80.6%), however the total number of ignitions here was considerably lower (n = 294). It is important to remember that these results were calculated from the pared-down fire ignitions dataset and they do not represent the complete historical frequency of large and small fires that started in each ecozone or each fuel type.

The fire season appears to shift later in the year while moving further north (Figure 4. 4b). In the Boreal Plain, several fires started in April, and many started in May, whereas in the Boreal Shield, next to zero fires started in April, and far fewer started in May than in the Boreal Plain. In the Taiga Shield, no fires started in April, and almost no fires started in May, indicating a contraction of the fire season. The end of the fire season did not seem to follow this trend to the same extent, possibly due to the residual effect of longer photoperiods as one moves further north.

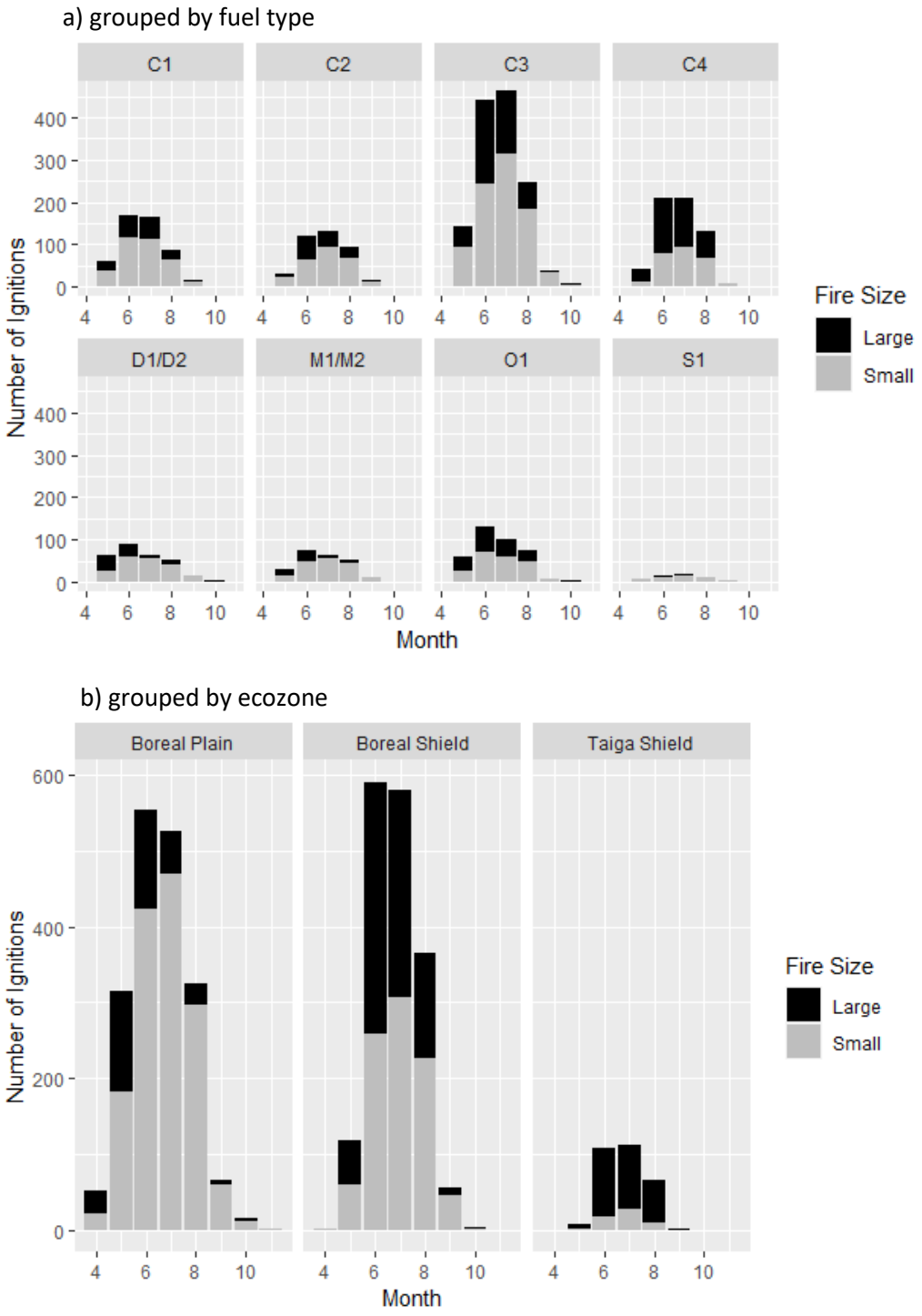


Figure 4. 4 Number of ignitions that started in each month, by fuel type (a) and ecozone (b), in the pared down dataset used for my analysis (Fires from 1990 – 2018). Bar shading indicates the proportion of ignitions that became large fires (>= 200 ha, black) and small fires (<= 0.5 ha, grey) for each grouping.

4.2 Classifications

4.2.1 Training performance metrics of top 10 models

Cross validation and validation performance metrics for the top ten models (out of the 500 that were built in the grid search hyperparameter tuning) are shown in Table 4. 1. The consistent values in recall and AUC between the cross validation and validation data indicates good generalizability of the models, and that the models were not over-fit during grid search hyperparameter tuning. There was also not a steep drop-off in performance among these models, suggesting that during model development, they all were quite good at distinguishing between when conditions were suitable for a large wildfire to occur or not.

Table 4. 1 Recall and AUC performance metrics for cross-validation, validation, and the mean of these two for the top 10 models from grid search model development. Models were sorted and ranked based on the mean recall from cross-validation and validation (see bold numbers), and the highest mean recall was the top performing model overall. Also shown is the mean accuracy.

Model Rank	Model ID	Cross-validation Recall	Cross-validation AUC	Validation Recall	Validation AUC	Mean Recall	Mean AUC	Mean Accuracy
1	RF_model_93	0.799	0.736	0.81	0.730	**0.805**	0.733	0.674
2	RF_model_128	0.754	0.735	0.836	0.727	**0.795**	0.731	0.672
3	RF_model_369	0.744	0.732	0.845	0.717	**0.794**	0.724	0.668
4	RF_model_424	0.744	0.732	0.845	0.717	**0.794**	0.724	0.668
5	RF_model_171	0.764	0.724	0.823	0.718	**0.793**	0.721	0.664
6	RF_model_354	0.759	0.732	0.823	0.718	**0.791**	0.725	0.668
7	RF_model_39	0.759	0.732	0.823	0.718	**0.791**	0.725	0.668
8	RF_model_344	0.733	0.724	0.845	0.730	**0.789**	0.727	0.672
9	RF_model_260	0.749	0.723	0.823	0.717	**0.786**	0.720	0.663
10	RF_model_311	0.725	0.733	0.845	0.728	**0.785**	0.731	0.673

The mean scaled variable importances for the top 10 models are shown in Figure 4. 5. For all 10 models, relative humidity (rh) was the most important variable. Fuel type was second most important, followed by ffmc and dc, which had overlapping 95% confidence intervals and were therefore quite similar in importance. Precipitation (precip) was by far the least important variable, followed by bui. The remaining variables (ws, temp, fwi, dmc, and isi) had overlapping 95% confidence intervals, and were moderately important (relative to the other variables) in determining whether an ignition would become a large wildfire.

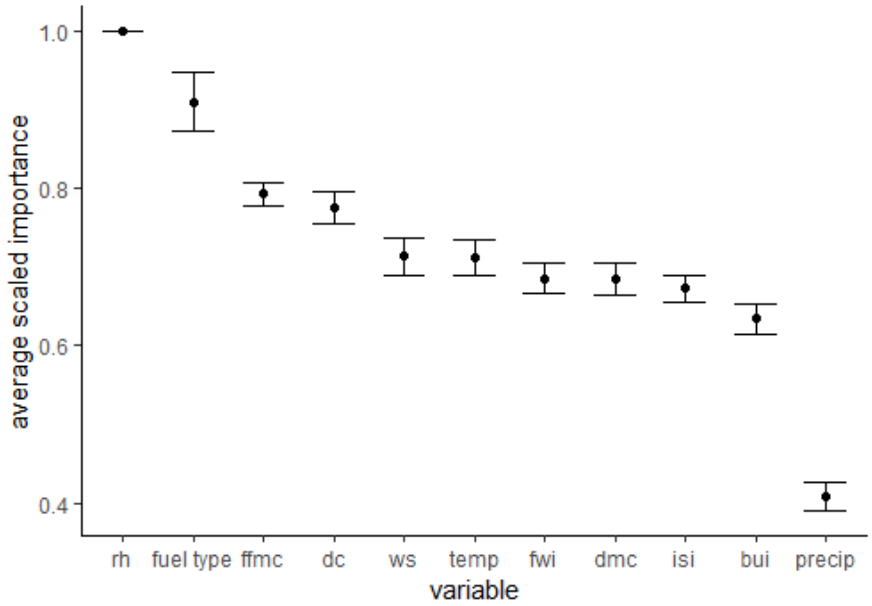


Figure 4. 5 Mean scaled importances of each variable for the top 10 random forest models. Bars indicate +/- 95% confidence intervals.

4.2.2 Performance metrics of top model on training data

The best model's performance in predicting whether the conditions in the test data records were suitable for large wildfire occurrence or not are shown in Table 4. 2, along with the thresholds used to determine which class each observation belonged in. The hyperparameters used to build this model are shown in Table A3.

Table 4. 2 Thresholds used in cross-validation, validation, and the optimal threshold (mean of cross-validation and validation thresholds) to use on test data. Observations with predicted class probability above or below the threshold were classified as having conditions suitable or not suitable for large wildfire occurrence, respectively. Performance metrics from using the best model to predict each test observation are also shown.

Model ID	Cross-validation Threshold	Validation Threshold	Optimal Threshold	Test Recall	Test AUC
RF_model_93	0.416	0.408	0.412	0.788	0.750

A confusion matrix of the results of passing the test dataset through the best model is shown in Figure 4. 6. This produced a recall of 0.788, slightly below the average recall (0.805) for the cross-validation and validation of the same model, indicating that over-fitting was not a large issue for this model. The recall results show that nearly 80% of the time when a large wildfire occurred, the model predicted the class accurately. The overall accuracy during testing

was 0.64. This may appear low, but is not unexpected given the tradeoff of selecting the best model based on recall which can lead to low accuracy in predicting the small fire class (greater number of false positives relative to true negatives). Press' Q test on the classification results of the test data resulted in a chi-squared value of 27.058 ($p < 0.05$, $df = 2$), indicating that the classifier performed significantly better than random chance alone.

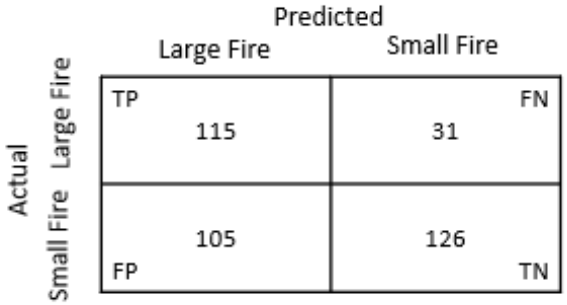


Figure 4. 6 Confusion matrix results of passing the test dataset through the best model.

Predictions on the test data were not completely uniform across ecozones. Table 4. 3 shows recall calculated for each ecozone. Boreal Plain had the highest recall, while Boreal Shield and Taiga Shield had lower but still reasonable values. The better performance in the Boreal Plain suggests that this model could be most useful in informing large wildfire occurrence risk accurately in the most populated ecozone of Saskatchewan's boreal forest.

Table 4. 3 Number of true positives (TP), false negatives (FN), and the recall for each ecozone on predictions of the test dataset using the best model.

Ecozone	TP	FN	recall
Boreal Plain	35	5	0.88
Boreal Shield	60	20	0.75
Taiga Shield	20	6	0.77

4.2.3 Analysis of the influence of spatial heterogeneity on classification results

4.2.3.1 Spatial heterogeneity among variables and variable importance

The mean interpolation variance for each variable in the test dataset observations are shown in Figure 4. 7. Precipitation had the highest mean variance, while Drought Code (dc) and wind speed (ws) had similarly high variance, with overlapping 95% confidence interval bars, showing that they were difficult to accurately interpolate because of how much they vary

spatially. The remaining FWI component variables had similar and intermediate variances, and relative humidity (rh) and temperature (temp) had the second least and least variance overall, indicating their interpolation estimates were the most reliable, and suggesting that these variables vary the least spatially.

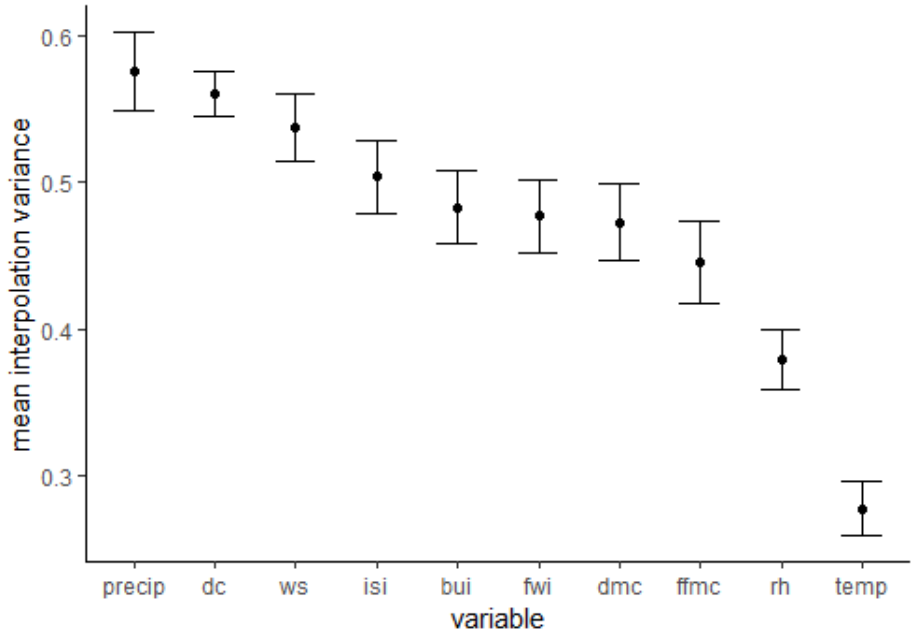


Figure 4. 7 Mean interpolation variance for each variable in the test dataset. The interpolation variance was calculated on a z-transformed version of each variable to allow for comparison among variables. Error bars represent 95% confidence intervals on the means.

Comparing the interpolation variance of the explanatory variables among themselves showed that there was a significant difference in how much the variables vary spatially (one-way ANOVA, $F = 55.29$, $df = 9$, $p < 0.05$). A post-hoc Tukey test showed that most pairwise variable comparisons were significant (Table 4. 4). The statistically significant differences and signs reflect quite closely the overlapping confidence interval bars in Figure 4. 7.

Table 4. 4 Pairwise comparisons of interpolation variance in each variable. Values are the mean of the row variable minus the mean of the column variable. A negative value indicates the column variable has greater spatial variance than the row variable. Asterisks (*) indicate statistically significant difference from the post-hoc Tukey test.

	bui	**dc**	**dmc**	**ffmc**	**fwi**	**isi**	**precip**	**rh**	**temp**	**ws**
bui	0	-0.077***	0.01	0.037	0.006	-0.021	-0.093***	0.103***	0.206***	-0.054*
dc	0.077***	0	0.088***	0.114***	0.083***	0.057*	-0.015	0.181***	0.283***	0.024
dmc	-0.01	-0.088***	0	0.027	-0.004	-0.031	-0.103***	0.093***	0.195***	-0.064**
ffmc	-0.037	-0.114***	-0.027	0	-0.031	-0.058*	-0.129***	0.066**	0.169***	-0.091***
fwi	-0.006	-0.083***	0.004	0.031	0	-0.027	-0.099***	0.097***	0.200***	-0.060*
isi	0.021	-0.057*	0.031	0.058*	0.027	0	-0.072***	0.124***	0.226***	-0.033
precip	0.093***	0.015	0.103***	0.129***	0.099***	0.072***	0	0.196***	0.298***	0.039
rh	-0.103***	-0.181***	-0.093***	-0.066**	-0.097***	-0.124***	-0.196***	0	0.102***	-0.157***
temp	-0.206***	-0.283***	-0.195***	-0.169***	-0.200***	-0.226***	-0.298***	-0.102***	0	-0.259***
ws	0.054*	-0.024	0.064**	0.091***	0.060*	0.033	-0.039	0.157***	0.259***	0

* $0.01 < p <= 0.5$
** $0.001 < p <= 0.01$
*** $p <= 0.001$

There is a slight inverse relationship between variable importance and interpolation variance (Figure 4. 8), suggesting that variable importance in the random forest models may be impacted by how much the variables vary spatially. However, a linear regression performed on these data showed that the relationship was not significant (*slope* = -0.799, $p = 0.158$, $R^2 = 0.23$). Most of the variables are clustered within a mean scaled importance between 0.6 – 0.8, but the variable with the lowest mean interpolation variance (temperature) is in the group of fourth most important variables. This suggests that variable importance is largely due to the nature of each variable and its influence on the suitability of conditions for wildfire activity, and not simply because of how reliably each variable can be interpolated.

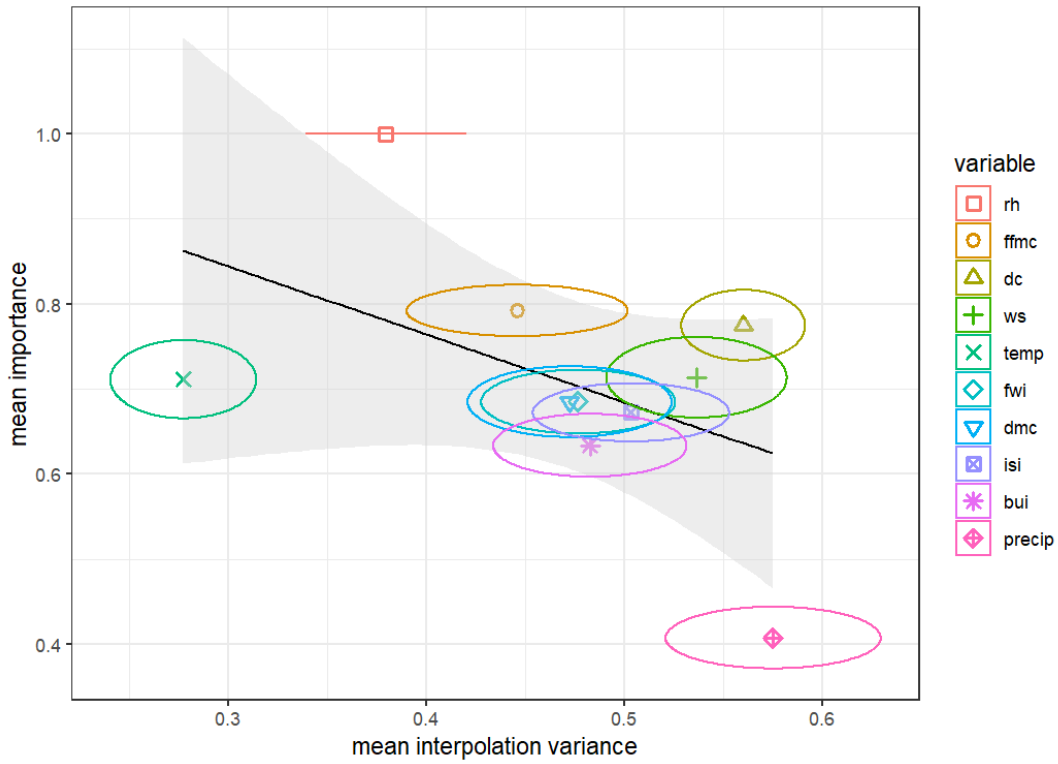


Figure 4. 8 Mean interpolation variance from the test data vs mean scaled importance in the top 10 random forest models for all 10 weather and FWI variables used. Ellipse height and width represent 95% confidence intervals for importance and interpolation variance of each variable, respectively. rh was the most important variable in all models so does not have any variation in the y axis. The black line is a linear regression line, while the grey ribbon indicates 95% confidence interval of the regression line.

4.2.3.2 Spatial heterogeneity among individual ignitions observations

To assess the influence of spatial heterogeneity on model performance and predictability, I plotted the total interpolation variance (for all interpolated variables) for each test observation

against the absolute difference between the prediction probability and prediction threshold (Figure 4. 9). My reasoning for investigating this was that observations which were nearer to the threshold would have been harder to discern what class they belong to at least in part because they occurred in locations further from weather stations, leading to greater interpolation variance. There was a weak decreasing pattern observable (Figure 4. 9) suggesting that observations with predictions nearer the threshold may have been those that were located farther from weather stations. However, a linear regression between summed variance and distance from probability threshold did not have a significant slope (*slope* = -0.945, $p = 0.148$, $R^2 = 0.006$) and therefore the pattern is not statistically significant.

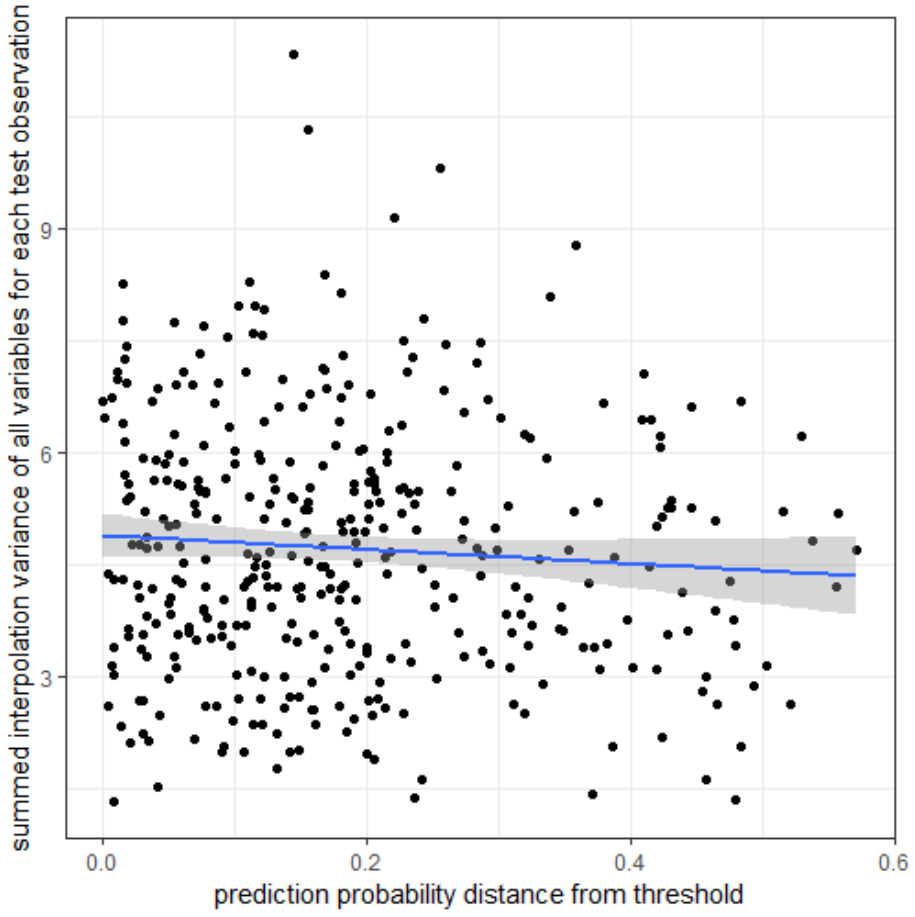


Figure 4. 9 Total interpolation variance of all variables for each observation in the test dataset vs how certain the prediction of the observation's class was (i.e. how far the prediction probability was from the threshold probability value). The blue line is a linear regression line, while the grey ribbon indicates 95% confidence interval of the regression line.

4.2.4 Predicting probability of large wildfire conditions for historical weather observations

The month with the greatest number of large fires started in each ecozone (high fire activity examples) is shown in Table 4. 5. Using interpolated (with ordinary kriging) mean (across all days) weather and fire weather values for those months, I created large fire conditions prediction maps using the top performing random forest model. For comparison, I also created large fire conditions prediction maps for each ecozone for the month with the most small fires (ignitions) for all months that did not have a single large fire start within them (low fire activity examples, Table 4. 5).

Table 4. 5 Month and year with the greatest number of large fires (high fire examples) and small fires (low fire examples) started for each ecozone. For low fire examples, only months where zero large fires were started were eligible to be selected. The timeframe for selecting high and low fire activity months was 1990 – 2018, inclusive.

Ecozone	Example type	Month	Year	n Large fires started	n Small fires started
Taiga Shield	High fire	July	2004	15	1
Boreal Shield	High fire	July	2004	43	10
Boreal Plain	High fire	June	2015	39	28
Taiga Shield	Low fire	July	2003	0	3
Boreal Shield	Low fire	September	1998	0	9
Boreal Plain	Low fire	August	1999	0	37

For the high fire activity months (Figure 4. 10a), 89.5% of predicted pixels across the entire study area met the conditions for large wildfires (i.e. were above the probability threshold of 0.412), while only 10.5% percent did not exhibit large wildfire conditions (i.e. were below the threshold). Comparatively, for the low fire activity months (Figure 4. 10b), only 20.5% of pixels had large wildfire conditions, while 79.5% did not. For the low fire activity months, large portions of land in the southern Boreal Plain ecozone were predicted to have large wildfire conditions. These areas are primarily classified as O1 (grass) fuel types, but are actually annual cropland. Annual crops could have still been green during this time of year, and therefore less flammable, which could explain why no large wildfires started here during this month.

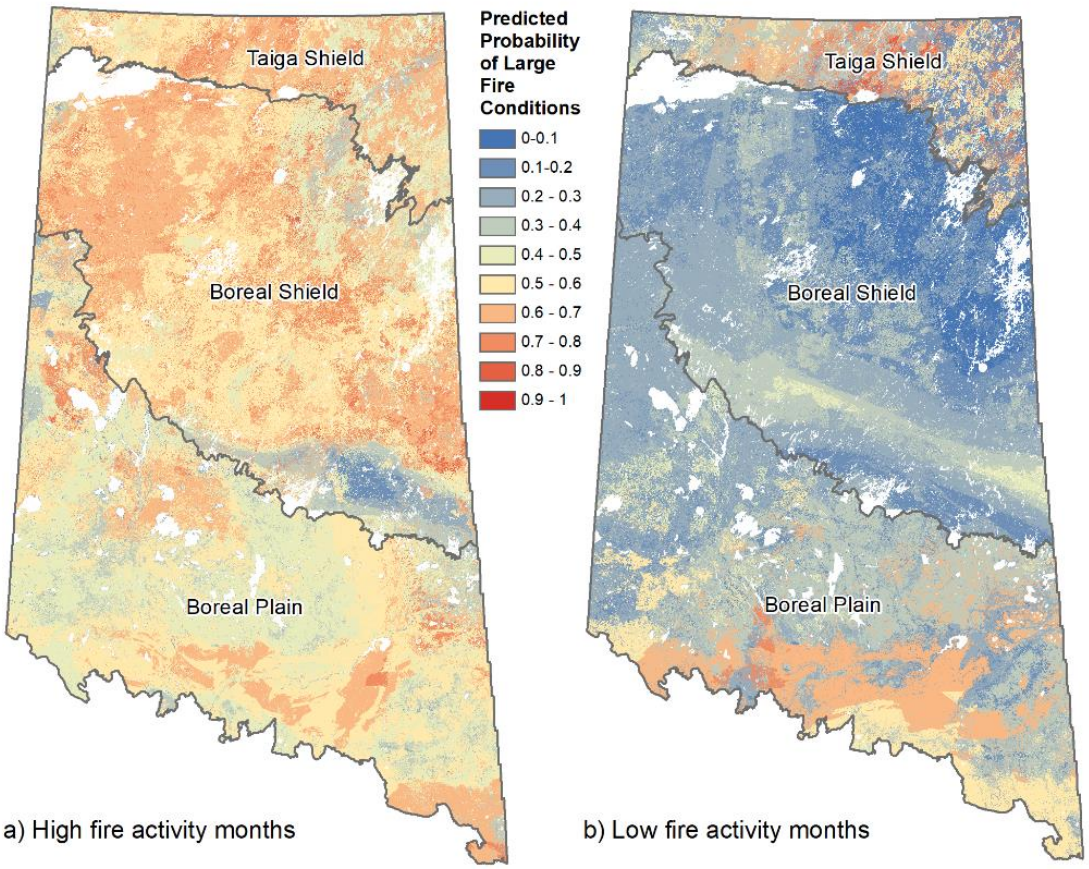


Figure 4. 10 Probabilities of large fire conditions predicted using the best model and average monthly values of input variables from high fire activity (a) and low fire activity (b) months for each ecozone. The month used for each ecozone and example type are shown in Table 4.5. Maps showing the locations of large fire burn footprints and small fire ignition locations during the high fire activity months, and small fire ignition locations during the low fire activity months are shown in Appendix B (Figures B 1 and B 2, respectively).

In the low fire activity prediction map (Figure 4. 10b), several areas (especially in the Taiga Shield and Boreal Plain ecozones) appear to have clusters of very high probability of large fire conditions adjacent to very low probability clusters. Upon closer investigation, these abrupt changes in large fire conditions probability are due to differences in fuel types. Figure 4. 11 shows a close-up example of four adjacent 100m x 100m pixels with three different fuel types and their respective predicted probabilities of large wildfire conditions. The differences in interpolated values of input variables at this spatial scale is very small, so the difference in probability values is almost entirely due to fuel type. As expected, C4 has the highest probability, while D2 has the lowest, and C1 has an intermediate probability, although still quite low. This pattern reinforces the importance that fuel type has in determining if the weather and fire weather

conditions will be suitable for a large fire to occur. It also illustrates the importance of having accurate fuel type data to use in model development and predictions.

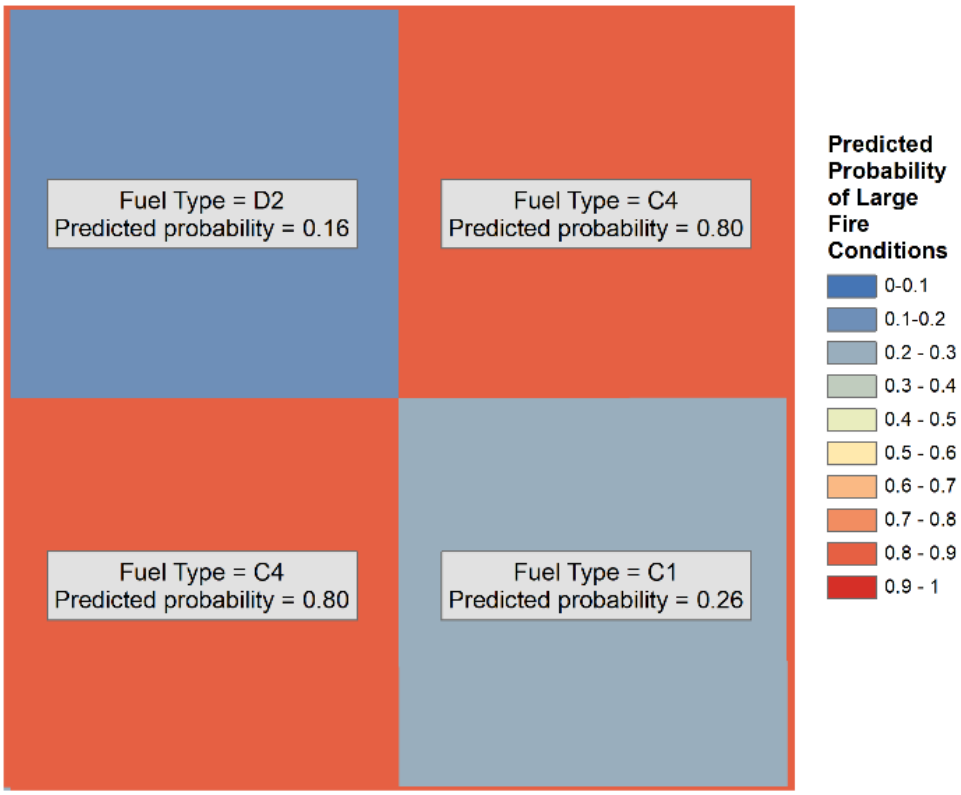


Figure 4. 11 Fuel types and predicted probabilities of large fire conditions for four adjacent pixels in the Taiga Shield. Input variables are interpolated mean daily values from July 2003.

The proportion of weather observations during the study period (1990 – 2018) where large wildfire threshold conditions for each fuel type were met is shown in Figure 4. 12. Nearly 60% of observations had conditions suitable for large wildfires to occur in immature jack pine (C4) forest. Over 50% of days also had large wildfire conditions for D1 and O1 fuel types. Large wildfire conditions for the remaining fuel types occurred less frequently. Large wildfire conditions for D2 were met the least frequently, at just over 20% of the observations.

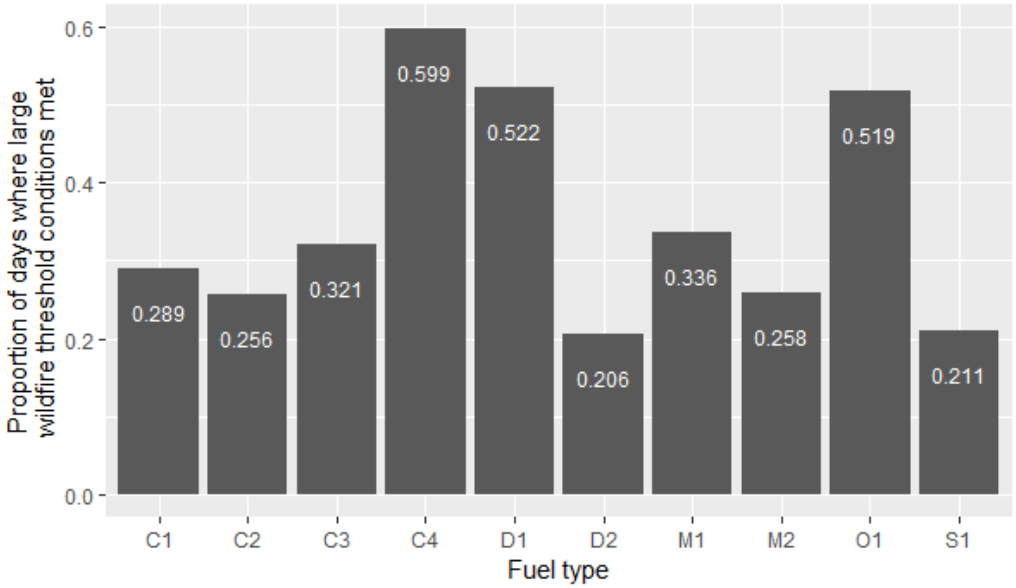


Figure 4. 12 Proportion of weather and fire weather observations during the study period (April – October, 1990 – 2018) predicted to have conditions suitable for a large wildfire to occur if an ignition event were to happen. Predictions were made using the best model, weather observations from all stations within the study area, and each fuel type assigned to each weather observation once.

5 DISCUSSION

5.1 Fire history

Compositions of large fires in the boreal forest of Saskatchewan between 2008 and 2018 showed that where and when large fires burn is not random. Mature jack pine was overwhelmingly the most burned compared to its availability, and immature jack pine was the second most burned. (Larsen, 1997) found that jack pine was the most frequently burned in a boreal forest in northern Alberta, adjacent to my study area. Jack pine's close link with fires is evidenced in its life history as well. Its seed cones require temperatures of roughly 50ºC to melt the resin that keeps them closed, allowing the seeds to disperse (Cameron, 1953, in Gauthier et al., 1996). Cumming (2001), in another study in Alberta, although on a forest further south than Larsen's, found that boreal spruce forest burned the most compared to its availability. Some differences in how the analysis was performed could explain the different results. Cumming (2001) modelled fuel types burned based on the fuel types in the surrounding unburned area, rather than determining what burned directly. While he showed these methods were statistically valid, they may be less accurate than determining the fuel types burned directly. Additionally, Cumming's study area was limited to boreal mixed-wood forest, whereas my study area included ecoregions comprised more dominantly of coniferous forest.

Unsurprisingly, Non Fuels (including water) burned far less than expected given their prevalence on the landscape in the study area, although a considerable amount of area burned was classified as Non Fuels. This could be the result of classification errors in the fuel types layer, as well as burn footprints encompassing primarily the outer most boundary fires and not capturing the patchy nature of burns in many cases. If Non Fuels exist patchily throughout a burn footprint, they would be considered burned and count towards the burned total. Visual inspection of several Non Fuel areas with satellite imagery revealed that some may still be vegetated (such as grass in a town park). This, along with the total amount of area burned in this class suggests revisiting or re-defining what is classified as Non Fuels in the FBP fuel type layers for Saskatchewan could be warranted

The proportion of large fires to small fires for each ecozone was vastly different. This could be caused by several factors. The Boreal Plain ecozone has the highest human density of these three ecozones, and human caused ignitions have been shown to be much more prevalent

closer to roads and communities (Adámek et al., 2018; K. Johnson et al., 2008). The increase in human activities in this ecozone can lead to greater detectability of even very small fires, as well as greater chance of intervention and preventing small fires from becoming large fires. I filtered out small fires within 20km of any communities to reduce the effect that human intervention might have, but human intervention likely still had at least some impact on preventing some small fires from becoming large. The Boreal Plain also has more mixed-wood forest (Acton et al., 1998), and large fire frequency in mixed-wood fuel types was shown to be lower than in other fuel types in my study.

Coniferous fuels are more prone to large fires, and the Boreal Shield has larger pure stands of these fuels than the Boreal Plain (Acton et al., 1998). This helps explain why large fires were more frequent and burned more area in the Boreal Shield than Boreal Plain. The human population is also lower in the Boreal Shield than further south, resulting in more difficulty of detecting small fires, and fewer interventions due to a lower priority for fire suppression. The Taiga Shield has an even smaller and less dense human population than the Boreal Shield, making small fires even less detectable and contributing to the unbalanced observations of large versus small fires in this ecozone.

Because of the great imbalance in fire size proportions by ecozone, I decided not to include ecozone as an explanatory variable in the random forest model. I wanted the model to be able to show what are the biotic and abiotic factors at a finer scale, and ecozones contain such a broad array of differences in both types of factors. Further study could focus on developing a separate model for each ecozone as well, however the fairly similar recall rates on test data by ecozone (Table 7) show that this may not be necessary, as the best model is quite robust in all ecozones.

There is conflicting evidence for ignition type having an influence on whether a large fire will occur from an ignition. For lightning-started fires, attributes of lightning strikes such as polarity, number of strokes per cloud to ground flash (a.k.a. multiplicity), and long continuing current have been investigated. But inconsistent results on the influence of these attributes have been reported, and low lightning detection accuracy has made it difficult to determine which exact strike was the cause of a given fire (Adámek et al., 2018; Pineda et al., 2014). For human-caused ignitions, Lawson et al. (1993) found that there was no significant difference in sustained ignition probability between test fires started with a match and test fires started from a

"campfire." This suggests that weather and fuel moisture conditions place a greater constraint on wildfire ignition probability than ignition energy.

The evidence of area burned minus available in large fires between 2008 and 2018 differing for both fuel types and ecozones supports the idea that fires have a preference in what and when they burn. Since wildfire occurrences therefore appear to be not random, pursuit of the question: "what are the drivers of large wildfire occurrence," the central objective of this thesis, is warranted.

5.2 Classification results

The best classification model performed quite well overall, and produced the most accurate predictions of large fire conditions for the Boreal Plain ecozone. The model did not do particularly well at distinguishing between true negatives and false positives, however. This illustrates that more work can be done on refining input data and developing explanatory models further.

Hyperparameter tuning is done to try maximize model performance, while minimizing model over-fitting, under the constraints of computing power and time. While hyperparameter tuning has been shown to improve model performance over default values, the process can be time consuming. Cook (2016) suggests that this time may be better spent on collecting more data or selecting features more carefully. These are areas I would like to explore more in the future, and they show that there is room for more research on my topic.

One prevailing dispute in wildfire research in the 1990s was between whether fuels or weather are more important in determining fire behavior (Cumming, 2001). However, my model helps illustrate a different way of viewing the phenomenon: namely that different fuel types have different thresholds of conditions for when a large fire could break out in them. For leafed aspen, these conditions would be much more extreme (e.g. severe drought, extremely high temperatures), and seldom encountered, unlike the conditions suitable for jack pine or spruce to have a large fire. The frequency of when large wildfire conditions are met varies by fuel type, which illustrates the importance of fuel type as a driver of wildfire occurrence. As Bessie & Johnson (1995) note, in extreme conditions, all fuels would be above thresholds for crown fire, although they did not state what these thresholds were.

Interestingly though, the fuel type that burned the largest area, and the greatest amount proportionate to its representation in the study area was mature jack pine (C3). But the fuel type that had the greatest number of weather observations above the threshold conditions for large wildfires was immature jack pine (C4). C3 had just over half as many large wildfire suitable weather observations as C4. While these are both the same species of tree, there is a fuel structure component to their designations. C3 tends to have a sparse understory covered by feature moss, will have fewer stems per unit area, and will have tree crowns not extending to the ground. C4 tends to have more needle litter and greater density of stems per unit area (Forestry Canada Fire Danger Group, 1992). Because of the differences in structure of these fuel types, C3 is much more prone to crown fire than C4. Crown fire intensity can be 10 – 100 times greater than a surface fire because of greater consumption of fuels and greater rate of spread (Van Wagner, 1980, found in Bessie & Johnson, 1995). Perhaps C4 has lower and more frequently observed thresholds for an ignition becoming a large fire than C3, but once above the threshold, and with ladder fuels to allow it, C3 can burn and spread much faster and therefore a much greater area due to crowning (Van Wagner, 1977).

Leafless aspen (D1) had the second greatest number of weather observations that were predicted to be above large wildfire condition thresholds. This matches what Larsen (1997) found in a forest adjacent to my study area in northern Alberta. This could be somewhat misleading however, because it is important to note the leafless stage of aspen (D1) only occurs in spring and fall. During the summer months when fire conditions tend to be greater, aspen has leaves and its moisture content is too great for it to burn (Alexander et al., 1984). The calculation of the proportion of days where large wildfire conditions are met could be modified for D and M fuel types to reflect this, but performing the calculations in this way helps demonstrate the influence that green aspen has on susceptibility to wildfires.

In addition to using the model to predict whether historical conditions were suitable for large wildfires to occur, by inputting forecasted future weather and fire weather conditions into the model, a map of future large fire condition probability could be produced with relative ease. This map could provide wildfire managers with an additional tool for evaluating fire danger across the landscape on a daily basis.

5.3 Variable importance

Using the random forest classifier, I was able to identify the most important drivers of large wildfire occurrence in Saskatchewan's Boreal forest, meeting one of the main objectives of my study. Interestingly, the most important driver was relative humidity, even though its contribution to wildfire activity tends to be understated in the literature. Fuel moisture and fuel flammability have been shown to be negatively correlated (Wright, 1932) because of the increased amount of energy required to combust fuels with more moisture, so the influence of relative humidity on fuel moisture at any given time must be significant.

It is interesting to note that the FWI variable was ranked in the lower half of variable importances, suggesting it is not the best suited to explaining when conditions are favorable for a large fire to occur. The interpretation of the FWI variable is that it represents the intensity of a fire (Van Wagner, 1987) or how much energy is released, and intuitively should be correlated to the probability of an ignition becoming a large wildfire. However, my current study does not support that intuition. Interestingly though, in a sensitivity analysis on FWI, relative humidity had the second most influence on changes in FWI (Dowdy et al., 2010), so the importance of relative humidity is captured to some extent in the underlying equations that make up the FWI.

From a fire danger prediction point of view, the CFFDRS/CWFIS produces daily forest fire danger and FWI maps for the entire country of Canada by grouping FWI values and other wildfire danger indices into classes of low, moderate, high, very high, and extreme values (Canadian Forest Service, 2020). Some of these maps consider only FWI values and do not incorporate fuel types. Large wildfires are responsible for burning the greatest amount of area, and are the greatest risk to human safety (K. Johnson et al., 2008). Given that FWI was not the most important variable in my study for explaining whether conditions are suitable for a large wildfire to occur, perhaps how the daily fire danger map is produced should be re-evaluated. In addition to the maps and indices already produced, a model which includes important drivers of large wildfire occurrences such as relative humidity and fuel types should be considered for use in informing public safety and wildfire management decisions.

Fuel type was the second most important variable in the classification model. The influence of fuel type on fire behavior has long been known and is the reason for incorporating fuel type into the Fire Behavior Prediction system in the first place (Forestry Canada Fire Danger Group, 1992). Observations from the large fires analysis component of my thesis, along with the

results of using the model to predict if large wildfire occurrence conditions were met during high and low fire activity months support the importance of fuel type in the model as well. It seems that all fuel types can burn, but the conditions required for certain fuel types to burn enough area to become a large wildfire are met more frequently for some fuel assemblages than for others.

FFMC and DC were roughly equally as important in determining whether conditions were suitable for a large wildfire to occur given an ignition. FFMC is analogous to the moisture content of the small thin pieces of wood (e.g. tinder or kindling) one would commonly use to start a campfire. If these are wet or effectively absent because of too high a moisture content, a campfire is extremely difficult to start. Even if they are dry and can be ignited, without larger dry fuels (analogous to DC), the fire will not be sustained, so it seems logical that these variables would also be important in this study.

Like DC, wind speed was also quite an important variable, however these two were in the top 3 most difficult to accurately interpolate. Wind speed has been shown to have a non-monotonic relationship with sustained ignition probability in test fires (see Fons & Stromberg, 1941 in Lawson et al., 1993). Some amount of wind can help sustain a flame by providing additional oxygen for combustion, but wind speeds above a certain threshold can have an opposite effect by withdrawing too much heat and preventing spread (see Countryman, 1983 in Lawson et al., 1993). Because of wind speed's importance and how difficult it is to interpolate, it would be a good candidate variable for further investigation on how to improve its estimation in areas distant from weather stations.

Surprisingly, the least important variable was precipitation. This could be a result of not only the difficulty by which it is interpolated, but also the amount of precipitation varying on a given day, with many days likely having zero or close to zero precipitation. Said differently, when it rains a lot, it could have a significant impact on fire behavior, but heavy rain days may be too infrequent in the dataset to make this variable very important overall. Furthermore, the fact that fuel moisture indices in the FWI system contain time lags to capture the rate at which different fuels can dry shows that the influence of precipitation stems much farther back than simply on the day of interest, so that may be why these variables better represent the influence of precipitation on large fire conditions than daily precipitation.

Investigating any of these variables on their own in a model that considers them together and can account for their interactions can be somewhat misleading and should be done with

caution. Methods exist for visualizing how input variables covary and interact in relation to a dependent variable (e.g. partial dependence plots, see Greenwell, 2017 for one example, and Friedman, 1991 for theoretical underpinnings), and would be worth investigating further. Furthermore, while random forests are generally accepted as being robust to correlated explanatory variables (multicollinearity) (Strobl et al., 2008) – which my weather variables and components of the FWI system are given how they are calculated (see Figure 3. 2) – importances can be biased by multicollinearity (Strobl et al., 2008). Using a modified version of variable importance such as conditional importance could help reduce the effects of multicollinearity on variable importance, and should be considered for refining random forest models in the future.

5.4 Spatial heterogeneity

Weather variables vary in differing amounts based not only on attributes inherent to each, but also due to weather patterns operating on different spatial and temporal scales than each variable measured daily in isolation. Cold fronts, warm fronts, pressure systems, and atmospheric stability can affect spatial auto-correlation of weather observations. Using an interpolation method that calculates a different weighting model for each variable and each day (i.e. OK) should help address some of these larger scale weather patterns (Jain & Flannigan, 2017). But it does not eliminate the variation in spatial heterogeneity among variables and observations.

The interpolated variables had differing amounts of spatial heterogeneity, as has been reported by others. Flannigan & Wotton (1989) suggest that highly variable rain events in summer are common, leading to precipitation being very difficult to accurately predict. This phenomenon has led to investigations into improving precipitation measurements for use in FWI calculations across the landscape. One way of accomplishing this is by combining Doppler radar data with weather station data such as in the Canadian Precipitation Analysis (CaPA). Hanes et al. (2017) found that incorporating CaPA into the FWI system improved estimates of precipitation significantly, resulting in improved accuracy of FFMC, ISI, and FWI as a result. Because radar coverage is not complete across Canada or Saskatchewan in particular, and CaPA has only been in operation since 2011, this approach was not suitable for my study. However, it should be considered for use where its temporal and spatial coverage is available for future studies.

Wind speed can be highly variable locally as well (Jain & Flannigan, 2017), and this variable was not significantly less heterogenous than precipitation. As already mentioned, its importance in the model suggest it would be a good candidate for trying to improve how well it can be interpolated. DC had similarly high variability spatially, which could be in part due to the long time lag that this variable includes (52 days) for drying rates, and how much precipitation can vary spatially and temporally. Temperature and rh had the lowest interpolation variance meaning they were easiest to interpolate accurately. Investigations into improving interpolation accuracy should therefore focus on other more spatially heterogenous variables, as mentioned above.

5.5 Fuel types

Selecting fuel types from the 2019 layer for fires starting prior to 2007 was a simplifying assumption, and its validity was not tested. A better approach may have been to build successional state and rates of state change models to predict previous fuel types in a given location similar to work completed in Minnesota by Hall et al. (1991). Because of the importance of fuel type in determining large wildfire occurrence, this would be a good candidate for future study.

Larsen (1997) found that proximity to water breaks was an important factor determining fire frequency in a northern Alberta boreal forest. Finding a way to characterize and quantify the surrounding fuel types of where ignitions took place could help with the classification performance and could improve the understanding of what is driving large wildfire occurrence.

5.6 Future study

In the process of completing my thesis, many different avenues for future study have become apparent. Improving on interpolation methods or ways of estimating weather and fire weather index values at unknown locations has already been a focus of study for wildfire researchers (e.g. Flannigan et al., 1998; Hanes et al., 2017; Jain & Flannigan, 2017). The importance of these variables in explaining large wildfire occurrence conditions, and the difficulty with which they are accurately interpolated as evidenced by my study demonstrate the need for improving how well these variables can be estimated at unknown locations. Alternatively, remote sensing methods could be explored further for their ability to augment daily weather and fire weather index values, or to identify important wavelengths or indices that

could be used to explain certain aspects of wildfire occurrence, like work done by Abdollahi et al. (2018) and Chowdhury & Hassan (2015).

The ignition data I used are only recorded for the day on which the ignition happened, however in many cases Conservation Officer reports on fires pinpoint the likely time of day of ignition, as well as more detailed information about the location of the fire. Analyzing these reports to determine what time each fire started, and projecting the weather and fire weather conditions to that same time using methods such as in Lawson et al. (1996) could help refine the understanding of when conditions are suitable for large wildfires to occur.

Training a classifier on a dataset is only one application of machine learning. Methods exist for generating altogether new features, and for selecting the most important features from a multi-dimensional dataset (e.g. Li et al., 2016). Even hyperparameter tuning, which I implemented, may not be as effective as careful feature selection (Cook, 2016). Many more weather, fire weather, and fuel moisture features or variables are used in understanding wildfire behavior, and they could be assessed and compared with the features I already used for their ability to explain large wildfire occurrences.

Fuel type was the second most important variable determining when conditions were suitable for a large wildfire to occur, however several aspects of the fuel type layers could be improved upon to provide a more accurate dataset. The spatial resolution of the fuel type layers is 100m x 100m, but this resulted in several fire ignitions on islands or shores of lakes being classified as water fuel types, because the majority of the pixel was made up of water. It also makes it more difficult to assess the fuel structure (e.g. how much downed woody material occurs, and what the understory looks like in a fuel type cell). This adds noise to the dataset and could obscure the analysis results if not carefully addressed. Improving the spatial resolution of fuel type layers should help alleviate this problem. Creating annual fuel type layers and using those to determine the fuel type of where each fire started could reduce noise in the dataset as well.

The fuel type component of my analysis considered only the fuel type that the ignition happened in, and not the surrounding fuels. Determining a way to characterize surrounding fuel types and their contiguity to the ignition location should help reduce erroneous classifications of observations. For instance, if an ignition occurred on a small island less than 0.5 ha in size, the weather and fire weather conditions could have been suitable for a large fire to occur. However,

because the surrounding fuels were water, there was no chance for the fire to become large. So there is a surrounding fuels component that is missing from my analysis that would help explain whether this ignition should have become large or not, and further study on how to accomplish this should be undertaken.

A critical assumption has been made in my study which has some supporting evidence for its validity, but is at the same time a somewhat arbitrary assertion. Choosing to determine when conditions are suitable for a large wildfire to occur (and saying that 200 ha is a large wildfire) could be focusing on the wrong point. If a goal of my research is to help understand wildfire behavior to minimize risk to humans, or to determine how best to allocate firefighting resources, then stronger evidence needs to be provided for what is the best metric to achieve that. For instance, if one fire slowly smolders and spreads across the ground over several weeks and burns an area of 300 ha, and another fire starts on the edge of a town and rapidly burns a very small area, including several houses, the small fire actually posed the greater threat to human life than the large fire. Projects such as the Saskatchewan Community Wildfire Risk Assessment Project (SCWRAP, (K. Johnson et al., 2008)) aim to mitigate wildfire risk around communities and human infrastructure, in part by encouraging adoption of FireSmart principles. These principles are intended to reduce the risk of wildfire damage whether the conditions are suitable for large fires or not. But being able to better predict when the conditions are suitable for a large wildfire to occur should help reduce the risk as well.

6 CONCLUSION

Wildfires have been researched extensively in Canada, especially in the boreal forest where wildfires occur frequently and are a natural ecosystem disturbance. Protecting human lives and interests has been the chief focus of wildfire managers (Parisien et al., 2020), and to accomplish that research has aimed to understand how fires behave, and why they occur where and when they do. Weather, fuel type, and fuel structure have been found to be critically important drivers of wildfire behavior, and they interact in complex ways to affect how fires burn (e.g. how fast they spread, how intensely they burn, how difficult they will be to control). The interactions between weather and fuel types are expressed primarily through moisture content of the fuels, which is affected by weather conditions impacting how fast fuels will dry out or become saturated with moisture. The moisture content of fuels governs how much energy will be required to ignite fuels and also how much energy will be released during combustion. The chemical makeup of different fuel types plays a role in this as well. Trees with higher resin content (such as jack pine and white spruce) have more energy to release during combustion, making it easier to ignite nearby fuels leading to sustained combustion and spreading of the fire. Sparsely arranged fuels make it more difficult for fire to spread since heat will dissipate more before reaching unburned fuels, so fuel structure and spatial arrangement also plays an important role.

While the physical processes governing wildfire occurrence and behavior have been investigated and are fairly well understood, a model considering and evaluating the drivers of these processes and the contributions of each element in determining whether a large wildfire could occur given an ignition (regardless of the probability that an ignition would occur) has prior to my thesis not been completed.

My research aimed to investigate the biotic and abiotic factors and conditions determining if an ignition will lead to a large wildfire in Saskatchewan's boreal forest. By first summarizing and analyzing the patterns of when, where, and in what fuel types large wildfires have historically burned I conclude that wildfire occurrences are not random. By building a quantitative model to classify under what weather, fuel moisture, and fuel type conditions an ignition is likely to become a large wildfire, I was able to determine that relative humidity and fuel type are the most important drivers of large wildfire occurrence. Because of the low density of weather stations in the study area, the spatial heterogeneity of many of the explanatory

variables I used, and the importance of several of the more difficult-to-interpolate variables, further research on improving interpolation accuracy should focus on wind speed and components of the FWI system. Using the model to predict the suitability of conditions to large wildfires further revealed the importance of fuel types. It also highlights the ideas that different fuel types have different threshold conditions for large wildfire occurrences, and the conditions suitable to large wildfires are observed more frequently for some fuel types (e.g. coniferous fuel types like jack pine, grass, leafless aspen, and boreal spruce) than for other fuel types (e.g. leafed aspen and leafed mixedwood).

My research demonstrates the importance of including a fuel type component in wildfire danger mapping in Canada in addition to weather and fuel moisture components, to capture the important drivers of large wildfire occurrence. By using the model I developed and putting forecasted weather and fire weather conditions into it, wildfire managers could create predicted maps of large wildfire conditions probability in the future, to help assess wildfire risk and act accordingly (e.g. put fire bans in place during high risk times, planning where to stage fire suppression resources, and deciding where to focus fuel treatments to minimize risk to humans). To help understand large wildfire occurrence risk further, where and when ignition events are likely to occur should be investigated, and that knowledge should be used in addition to the predictions of when conditions are suitable for an ignition to grow into a large wildfire.

The methods I used to investigate my research question could be improved in many ways. The fuels component could include other elements such as physical fuel loading and structure, and neighboring fuels composition, and should be available with higher spatial and temporal resolution to improve the underlying model and make predictions more accurate. Interpolation methods for weather and fuel moisture could also be improved by incorporating remotely sensed data where available, and increasing the density of weather stations. Different types of machine learning could be applied such as feature generation and feature selection as well. My approach for building a model to predict large wildfire conditions regardless of the probability of an ignition happening could also be used in other jurisdictions to get a locally specific picture of large wildfire occurrence drivers.

REFERENCES

Abdollahi, M., Islam, T., Gupta, A., & Hassan, Q. K. (2018). An advanced forest fire danger forecasting system: Integration of remote sensing and historical sources of ignition data. *Remote Sensing*, *10*(6). https://doi.org/10.3390/rs10060923

Acton, D. F., Padbury, G. A., & Stushnoff, C. T. (1998). *The Ecoregions of Saskatchewan*. Canadian Plains Research Center.

Adámek, M., Jankovská, Z., Hadincová, V., Kula, E., & Wild, J. (2018). Drivers of forest fire occurrence in the cultural landscape of Central Europe. *Landscape Ecology*, *2*, 2031–2045. https://doi.org/10.1007/s10980-018-0712-2

Alexander, M. E., Lawson, B. D., Stocks, B. J., & Van Wagner, C. E. (1984). User guide to the Canadian Forest Fire Behavior Prediction System. In *Canadian Forest Service*.

Anderson, K., Martell, D. L., Flannigan, M. D., & Wang, D. (2000). Modeling of fire occurrence in the boreal forest region of Canada. In E. Kasischke & B. J. Stocks (Eds.), *Fire, climate change and carbon cycling in the boreal forest* (pp. 357–367). Springer-Verlag. https://doi.org/10.1007/978-0-387-21629-4_19

Au, T. C. (2018). Random forests, decision trees, and categorical predictors: the "absent levels" problem. *Journal of Machine Learning Research*, *19*, 1–30.

Barr, A. G., Black, T. A., Hogg, E. H., Kljun, N., Morgenstern, K., & Nesic, Z. (2004). Inter-annual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production. *Agricultural and Forest Meteorology*, *126*(3–4), 237–255. https://doi.org/10.1029/2002JD003011

Bessie, W. C., & Johnson, E. A. (1995). The relative importance of fuels and weather on fire behavior in subalpine forests. *Ecology*, *76*(3), 747–762.

Birch, D. S., Morgan, P., Kolden, C. A., Abatzoglou, J. T., Dillon, G. K., Hudak, A. T., & Smith, A. M. S. (2015). Vegetation, topography and daily weather influenced burn severity in central Idaho and western Montana forests. *Ecosphere*, *6*(1), 1–23. https://doi.org/10.1890/ES14-00213.1

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer. https://doi.org/10.1021/jo01026a014

Bond-Lamberty, B., Peckham, S. D., Ahl, D. E., & Gower, S. T. (2007). Fire as the dominant driver of central Canadian boreal forest carbon balance. *Nature*, *450*(7166), 89–92.

https://doi.org/10.1038/nature06272

Boucher, J., Beaudoin, A., Hébert, C., Guindon, L., & Bauce, É. (2017). Assessing the potential of the differenced Normalized Burn Ratio (dNBR) for estimating burn severity in eastern Canadian boreal forests. *International Journal of Wildland Fire*, *26*, 32–45. https://doi.org/10.1071/WF15122

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Canada, N. R. (2019). *Canadian Wildland Fire Information System Datamart*. http://cwfis.cfs.nrcan.gc.ca/datamart

Canada, N. R. (2020). *Fire ecology*. https://www.nrcan.gc.ca/our-natural-resources/forests-forestry/wildland-fires-insects-disturban/forest-fires/fire-ecology/13149

Canadian Forest Service. (2019). *Canadian National Fire Database - Agency Fire Data*. http://cwfis.cfs.nrcan.gc.ca/en_CA/nfdb

Canadian Forest Service. (2020). *Fire Weather Maps*. https://cwfis.cfs.nrcan.gc.ca/maps/fw?type=fwi

Chowdhury, E., & Hassan, Q. K. (2015). Development of a new daily-scale forest fire danger forecasting system using remote sensing data. *Remote Sensing*, *7*(3), 2431–2448. https://doi.org/10.3390/rs70302431

Cook, D. (2016). *Practical Machine Learning with H2O*. O'Reilly Media, Inc. https://learning.oreilly.com/library/view/practical-machine-learning/9781491964590/

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology*, *17*(5), 563–586. https://doi.org/10.1007/BF01032109

Cumming, S. G. (2001). Forest type and wildfire in the Alberta boreal mixedwood: what do fires burn? *Ecological Applications*, *11*(1), 97. https://doi.org/10.2307/3061059

De Groot, W. J. (1987). Interpreting the Canadian forest fire weather index (FWI) system. *Proc. of the Fourth Central Regional Fire Weather Committee Scientific and Technical Seminar*, 1–9. https://doi.org/citeulike-article-id:14176512

Dowdy, A. J., Mills, G. A., Finkele, K., & De Groot, W. J. (2010). Index sensitivity analysis applied to the Canadian forest fire weather index and the McArthur forest fire danger index. *Meteorological Applications*, *17*(3), 298–312. https://doi.org/10.1002/met.170

ESRI Inc. (2020). *ArcMap (Version 10.8)*. https://desktop.arcgis.com/en/arcmap/

Flannigan, M. D., Krawchuk, M. A., De Groot, W. J., Wotton, B. M., & Gowman, L. M. (2009). Implications of changing climate for global wildland fire. *International Journal of Wildland Fire*, *18*(5), 483–507. https://doi.org/10.1071/WF08187

Flannigan, M. D., & Wotton, B. M. (1989). A study of interpolation methods for forest fire danger rating in Canada. *Canadian Journal of Forest Research*, *19*, 1059–1066.

Flannigan, M. D., Wotton, B. M., & Ziga, S. (1998). *A study on the interpolation of fire danger using radar precipitation estimates*.

Forestry Canada Fire Danger Group. (1992). Development and structure of the Canadian forest fire behavior prediction system. In *Information Report ST-X-3*. https://doi.org/ST-X-3

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, *19*(1), 1–141. https://doi.org/10.1214/aos/1176348654

Gauthier, S., Bergeron, Y., & Simon, J.-P. (1996). Effects of fire regime on the serotiny level of jack pine. *Journal of Ecology*, *84*(4), 539–548.

Greenwell, B. M. (2017). pdp: An R package for constructing partial dependence plots. *R Journal*, *9*(1), 421–436. https://doi.org/10.32614/rj-2017-016

H2O.ai. (2020a). *h2o: R Interface for H2O* (R package version 3.30.0.1). http://www.h2o.ai

H2O.ai. (2020b). *Variable Importance*. https://docs.h2o.ai/h2o/latest-stable/h2o-docs/variable-importance.html

H2O.ai. (2021). *Performance and Prediction*. http://docs.h2o.ai/h2o/latest-stable/h2o-docs/performance-and-prediction.html?highlight=AUC#auc-area-under-the-roc-curve

Hall, F. G., Botkin, D. B., Strebel, D. E., Woods, K. D., & Goetz, S. J. (1991). Large-scale patterns of forest succession as determined by remote sensing. *Ecology*, *72*(2), 628–640.

Hanes, C. C., Jain, P., Flannigan, M. D., Fortin, V., & Roy, G. (2017). Evaluation of the Canadian Precipitation Analysis (CaPA) to improve forest fire danger rating. *International Journal of Wildland Fire*, *26*(6), 509–522. https://doi.org/10.1071/WF16170

Jain, P., & Flannigan, M. D. (2017). Comparison of methods for spatial interpolation of fire weather in Alberta, Canada. *Canadian Journal of Forest Research*, *47*, 1646–1658.

Johnson, E. A., Miyanishi, K., & Weir, J. M. H. (1998). Wildfires in the western Canadian boreal forest: landscape patterns and ecosystem management. *Journal of Vegetation Science*, *9*(4), 603–610. https://doi.org/10.2307/3237276

Johnson, K., Maczek, P., & Fremont, L. (2008). *Saskatchewan Community Wildfire Risk Assessment Project*.

Kafka, V. G., Parisien, M. A., Hirsch, K., Flannigan, M. D., & Todd, J. B. (2001). Climate change in the prairie provinces: assessing lanscape fire behavior potential and evaluating fuel treatment as an adaptation treatment strategy. In *Prairie Adaptation Research Cooperative*. http://cfs.nrcan.gc.ca/pubwarehouse/pdfs/33457.pdf

Key, C. H., & Benson, N. C. (2006). Landscape assessment (LA): sampling and analysis methods. In *USDA Forest Service General Technical Report RMRS- GTR-164-CD. LA1–LA51. USDA Forest Service, Rocky Mountain Research Station*. https://doi.org/10.1002/app.1994.070541203

King, G., & Zeng, L. (2003). Logistic regression in rare events data. *Journal of Statistical Software*, *8*, 137–163. https://doi.org/10.18637/jss.v008.i02

Krawchuk, M. A., Haire, S. L., Coop, J., Parisien, M. A., Whitman, E., Chong, G., & Miller, C. (2016). Topographic and fire weather controls of fire refugia in forested ecosystems of northwestern North America. *Ecosphere*, *7*(12), e01632. https://doi.org/10.1002/ecs2.1632

Krawchuk, M. A., & Moritz, M. A. (2011). Constraints on global fire activity vary across a resource gradient. *Ecology*, *92*(1), 121–132. https://doi.org/10.1890/09-1843.1

Larsen, C. P. S. (1997). Spatial and temporal variations in boreal forest fire frequency in northern Alberta. *Journal of Biogeography*, *24*(5), 663–673. https://doi.org/10.1111/j.1365-2699.1997.tb00076.x

Lawson, B. D., & Armitage, O. B. (1996). Wildfire ignition probability predictor (WIPP). *Canadian Froest Service, Pacific Forestry Center, R&D Update. (Victoria, BC)*, *604*.

Lawson, B. D., & Armitage, O. B. (2008). Weather guide for the Canadian forest fire danger rating system. In *The Forestry Chronicle* (Vol. 65). https://doi.org/10.5558/tfc65258-4

Lawson, B. D., Armitage, O. B., & Dalrymple, G. N. (1993). Ignition probabilities for simulated people-caused fires in British Columbia's lodgepole pine and white spruce-subalpine fir forests. *Proceedings on the 12th Conference on Fire and Forest Meteorology*, 493–505.

Lawson, B. D., Armitage, O. B., & Hoskins, W. D. (1996). *Diurnal variation in the fine fuel moisture code: tables and computer source code*.

Li, J., Tran, M., & Siwabessy, J. (2016). Selecting optimal random forest predictive models: a case study on predicting the spatial distribution of seabed hardness. *PLoS ONE*, *11*(2).

https://doi.org/10.1371/journal.pone.0149089

Lynham, T. J., & Martell, D. L. (1989). Preliminary report on a national database of experimental fires in Canada. *Proceedings on the National Workshop on Fire Occurrence Prediction*, 41–44.

Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., & De Mendonça, A. (2011). Data mining methods in the prediction of dementia: a real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Research Notes*, *4*(299). https://doi.org/10.1186/1756-0500-4-299

Parisien, M. A., Hirsch, K. G., Lavoie, S. G., Todd, J. B., & Kafka, V. G. (2004). Saskatchewan fire regime analysis. In *Natural Resources Canada, Canadian Forest Service. Northern Forestry Centre. Edmonton AB* (Issue Information Report NOR-X-394). http://nofc.cfs.nrcan.gc.ca/bookstore_pdfs/24912.pdf

Parisien, Marc André, Barber, Q. E., Hirsch, K. G., Stockdale, C. A., Erni, S., Wang, X., Arseneault, D., & Parks, S. A. (2020). Fire deficit increases wildfire risk for many communities in the Canadian boreal forest. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-15961-y

Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. *Computers and Geosciences*, *30*(7), 683–691. https://doi.org/10.1016/j.cageo.2004.03.012

Pebesma, E. J. (2014). gstat user's manual. *Dept. of Physical Geography, Utrecht University, Utrecht, The Netherlands*.

Pineda, N., Montanyà, J., & Van der Velde, O. A. (2014). Characteristics of lightning related to wildfire ignitions in Catalonia. *Atmospheric Research*, *135–136*, 380–387. https://doi.org/10.1016/j.atmosres.2012.07.011

R Core Team. (2020). *R: a language environment for statistical computing.* R Foundation for Statistical Computing. https://www.r-project.org/

Renkin, R. A., & Despain, D. G. (1992). Fuel moisture, forest type, and lightning-caused fire in Yellowstone National Park. *Canadian Journal of Forest Research*, *22*, 37–45.

Rothermel, R. C. (1972). A mathematical model for predicting fire spread in wild land fuels (INT-115). In *USDA Forest Service Research Paper*.

Stocks, B. J., Mason, J. A., Todd, J. B., Bosch, E. M., Wotton, B. M., Amiro, B. D., Flannigan,

M. D., Hirsch, K. G., Logan, K. A., Martell, D. L., & Skinner, W. R. (2002). Large forest fires in Canada, 1959–1997. *Journal of Geophysical Research*, *108*(D1). https://doi.org/10.1029/2001JD000484

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*(307), 1–11. https://doi.org/10.1186/1471-2105-9-307

Sun, P., & Zhang, Y. (2018). A probabilistic method predicting forest fire occurrence combining firebrands and the weather-fuel complex in the northern part of the Daxinganling Region, China. *Forests*, *9*(7), 1–19. https://doi.org/10.3390/f9070428

Taylor, S. W., & Alexander, M. E. (2006). Science, technology, and human factors in fire danger rating: the Canadian experience. *International Journal of Wildland Fire*, *15*(1), 121–135. https://doi.org/10.1071/WF05021

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*(sup(1)), 234–240. https://doi.org/10.1126/science.11.277.620

Van Wagner, C. E. (1977). Conditions for the start and spread of crown fire. *Canadian Journal of Forest Research*, *7*, 23–24.

Van Wagner, C. E. (1980). Fire behavior in northern conifer forests and shrublands. In R. W. Wein & D. A. MacLean (Eds.), *Fire in northern circumpolar ecosystems* (pp. 45–80). Academic Press.

Van Wagner, C. E. (1987). *Development and structure of the Canadian forest fire weather index system (Forestry Technical Report 35)*. https://doi.org/19927

Van Wagner, C. E. (1990). Six decades of forest fire science in Canada. *The Forestry Chronicle*, *66*(2), 133–137. https://doi.org/10.5558/tfc66133-2

WFM. (2019a). *Fire Behavior Prediction Fuel Types Layers*. Saskatchewan Ministry of Environment.

WFM. (2019b). *SK WFM FWI Dataset*. Saskatchewan Ministry of Environment.

Whitman, E., Parisien, M. A., Thompson, D. K., Hall, R. J., Skakun, R. S., & Flannigan, M. D. (2018). Variability and drivers of burn severity in the northwestern Canadian boreal forest. *Ecosphere*, *9*(2), e02128. https://doi.org/10.1002/ecs2.2128

Wright, J. G. (1932). *Forest fire hazard research as developed and conducted at the Petawawa Forest Experiment Station (Information Report FF-X-5)*.

Wright, J. G. (1937). *Preliminary Improved Fire Hazard Index Tables for Pine Forests at Petawawa Forest Experiment Station*.

APPENDIX A

Distributed Random Forest algorithm parameters from h2o package for R.

Table A 1 Hyperparameter descriptions and values used in grid-search random forest model development. Parameter names and how they function are specific to the h2o Distributed Random Forest algorithm.

Hyperparameter	Description	Default Value	Impact	Trade-off	Values used
fold_assignment	method for assigning data to cross-validation folds	Random	Different methods better for different dataset sizes and number of classes	classification performance	'Random', 'Modulo', 'Stratified'
nbins_cats	number of bins used to build histograms for categorical variables	1024	using fewer bins can help avoid overfitting	generalizability of model	5, 10, 15, 20, 50
nbins_top_level	number of bins used to build histogram at top level of tree, at each subsequent level, number of histograms will be halved	1024	using fewer bins can help avoid overfitting	generalizability of model	32, 64, 128, 256, 512
nbins	minimum number of bins used to build histogram (number of bins will not be reduced to lower than this value)	20	using more bins can lead to overfitting	generalizability of model	10, 20, 30, 50
ntrees	number of trees to build	50	more trees will reduce variance of predictions	computational cost	100, 250, 500
max_depth	maximum depth of each tree (depth will not always be reached when growing trees though)	20	higher values can lead to overfitting	generalizability of model	10, 12, 15
min_split_improvement	minimum improvement of classification required to include a split; helps prevent overfitting	1e-5	higher values can help avoid overfitting	generalizability of model	1e-3, 1e-5, 1e-10, 1e-20
categorical_encoding	method for encoding categorical variables	Enum	Different methods better for variables with different numbers of categories, some methods better if the variable has an ordinal nature	classification performance	'Enum', 'EnumLimited', 'SortByResponse'

Table A 2 Static hyperparameters and values used in random forest model development. Parameter names and how they function are specific to the h2o Distributed Random Forest algorithm.

Static parameters	Description	Default Value	Impact	Trade-off	Value used
n_folds	number of folds to use for k-fold cross-validation each time a random forest is learned	n/a	higher values can help avoid overfitting	computational cost	10
balance_classes	whether to balance number of records for each response class when training model (to address unbalanced dataset)	n/a	using this can help address unbalanced dataset	minor computational cost	TRUE
class_sampling_factors	ratio by which to under/over-sample each class to achieve class balance (*sampling without replacement)	n/a	using this can help address unbalanced dataset	generalizability of model (some records not used)	undersample over-represented class until classes are equal size
sample_rate_per_class	ratio by which to under/over-sample each class to achieve class balance when building each tree (*sampling without replacement)	n/a	using this can help address unbalanced dataset	generalizability of model (some records not used)	undersample over-represented class until classes are equal size
histogram_type	if and how to group records to find splits	AUTO (steps of (max-min)/N)	computational cost	minor classification performance (potentially)	AUTO
mtries	column sampling rate				-1 (square root (p))

Table A 3 Grid search hyperparameters values of best model

Hyperparameter	Value
fold_assignment	Modulo
nbins_cats	20
nbins_top_level	128
nbins	50
ntrees	500
max_depth	15
min_split_improvement	1e-3
categorical_encoding	EnumLimited

APPENDIX B

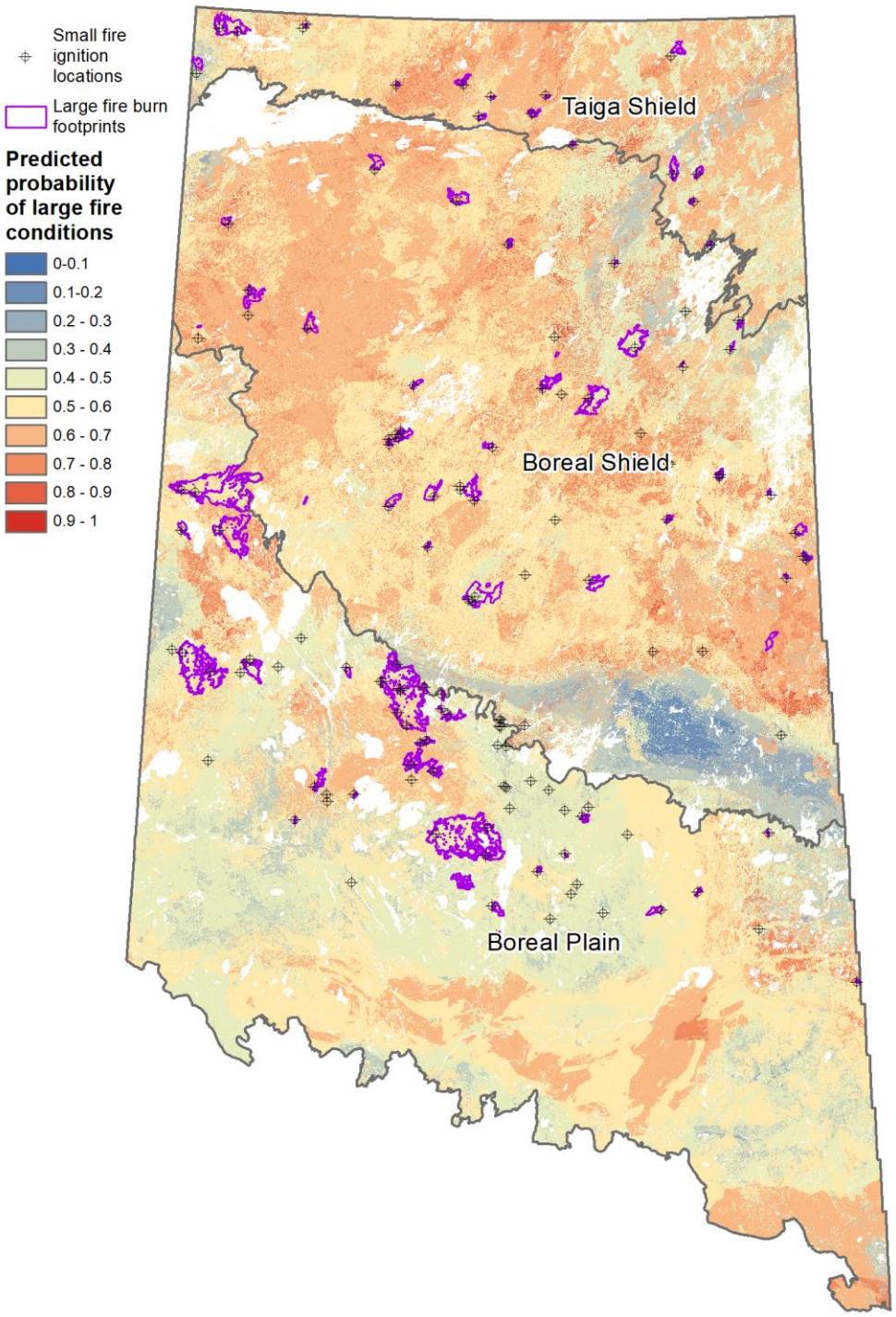


Figure B 1. Predicted large fire conditions probability for high fire activity months in each ecozone, with large fire burn footprints and small fire ignition locations that occurred in each month overlain. High fire activity months were: Taiga Shield - July 2004, Boreal Shield - July 2004, and Boreal Plain - June 2015.

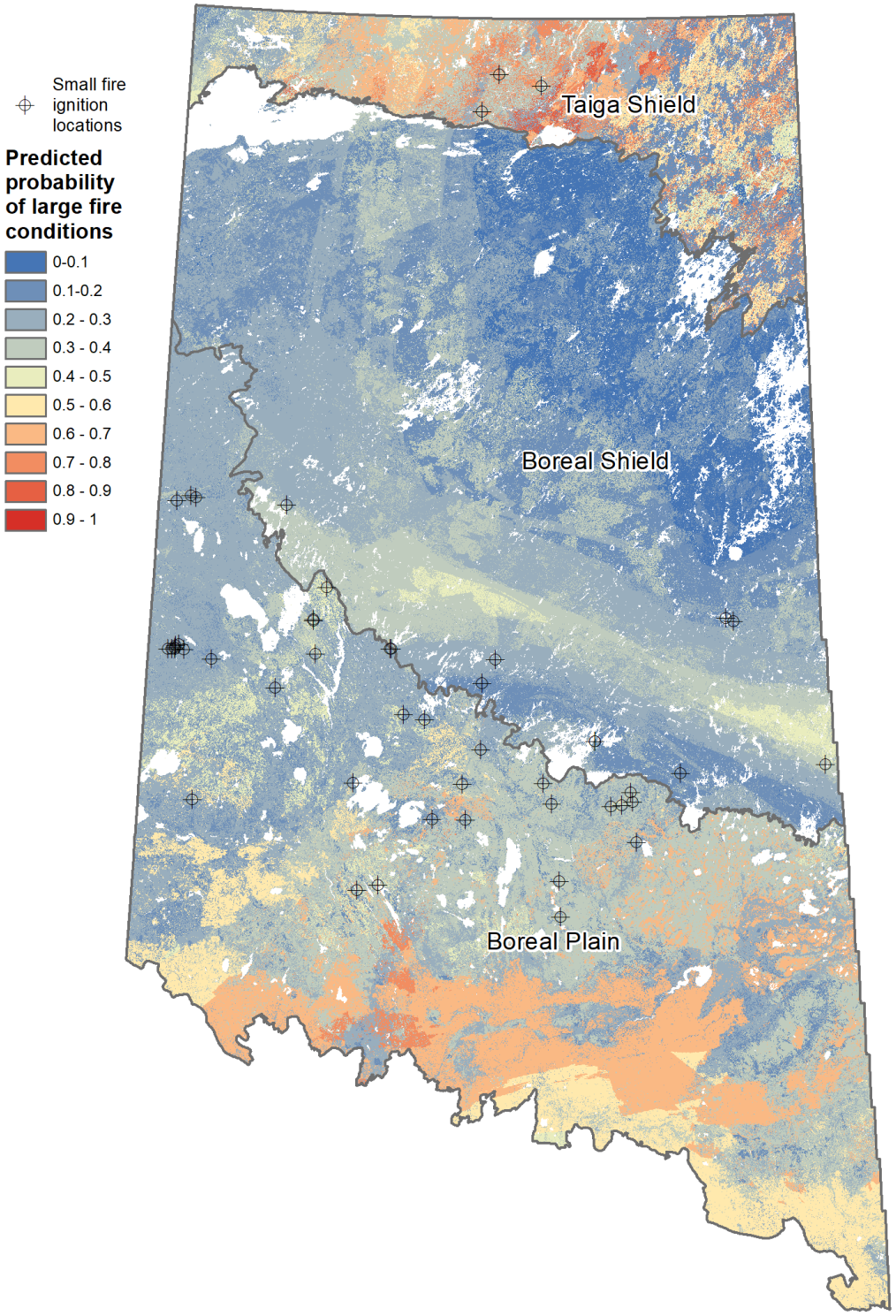


Figure B 2. Predicted large fire conditions probability for low fire activity months in each ecozone, with small fire ignition locations that occurred in each month overlain. Low fire activity months were: Taiga Shield - July 2003, Boreal Shield - September 1998, and Boreal Plain - August 1999.

APPENDIX C

Computer code for building, testing, and evaluating classifier models in R.

```
rm(list=ls(all=T))
library(h2o)
h2o.init(min_mem_size = '60G', max_mem_size = '60G')
library(tidyverse)
library(plotrix)


##########################################
###########    Build models    ###########
##########################################

#1. Read in data
ignitions_path <- '<file_path>.csv' # file path to ignitions dataset
(data table used to build model from)
d <- h2o.uploadFile(ignitions_path, destination_frame="ignitions.hex")
d$StartedFire <- as.factor(d$StartedFire)

#2. Prepare target encoding for categorical variables
d$fold <- h2o.kfold_column(d, nfolds=5, seed=33)
encoded_vars <- list('ZONNAME', 'fuel_type_adj_2')
targ_enc_map <- h2o.target_encode_fit(d, x=encoded_vars,
y='StartedFire', fold_column='fold')
encoded_d <- h2o.target_encode_transform(d, x=encoded_vars,
y='StartedFire',

target_encode_map=targ_enc_map,
                                                holdout_type='kfold',
                                                fold_column='fold',
                                                blended_avg=F,
                                                seed=33,
                                                noise=0)
#3. Declare variables
response <- 'StartedFire'

encode_fuel_type <- F
if(encode_fuel_type == T){
  predictors <- c('fwi', 'isi', 'bui', 'ffmc', 'dc', 'dmc',
                  'temp', 'rh', 'ws', 'precip3', 'fuel_type_adj_2_te')
} else{
  predictors <- c('fwi', 'isi', 'bui', 'ffmc', 'dc', 'dmc',
                  'temp', 'rh', 'ws', 'precip3', 'fuel_type_adj_2')
}

#4. Explore number of records by different categorical variable values
(started_fires_by_fuel_type <- as_tibble(h2o.group_by(encoded_d,
by=c('StartedFire', 'fuel_type_adj_2'), nrow('StartedFire'))))
(started_fires_by_ecozone <- as_tibble(h2o.group_by(encoded_d,
by=c('StartedFire', 'ZONNAME'), nrow('StartedFire'))))

#5. Split data into training (to build model), validation (to test
model), and testing (unseen data for final validation of model) sets
d_split <- h2o.splitFrame(encoded_d, ratios=0.90, seed=5,
destination_frames=c('train.hex', 'validation.hex'))
```

```
d_train <- d_split[[1]] # data that will be further split into
training and validation for model building
d_test <- d_split[[2]] # data that will not be seen until final
testing

d_train_split <- h2o.splitFrame(d_train, ratios=0.80, seed=28)
d_train <- d_train_split[[1]] # training data
d_validation <- d_train_split[[2]] # validation data


#6. Calculate undersampling rate for over-represented class
base_undersample <- 1.0
(samples_by_class <- h2o.group_by(d_train, by='StartedFire',
nrow('StartedFire')))
(undersample_rate <-
min(samples_by_class[,2])*base_undersample/max(samples_by_class[,2]))

#7. Set hyperparameters for grid search model building
rf_params <- list(fold_assignment = c('Random', 'Modulo',
'Stratified'),
                  nbins_cats = c(5, 10, 15, 20, 50),
                  nbins_top_level = c(32, 64, 128, 256, 512),
                  nbins = c(10, 20, 30, 50),
                  ntrees = c(100, 250, 500),
                  max_depth = c(10, 12, 15),
                  min_split_improvement = c(1e-3, 1e-5, 1e-10, 1e-20),
                  categorical_encoding = c("Enum", "EnumLimited",
"SortByResponse"))


search_criteria <- list(strategy="RandomDiscrete", max_models=500,
seed=4)

#8. Build classifier (clf) models using hyperparameter tuning and
cross-validation
#   or read it from file if already created and saved
file_path_to_output_grids_folder <- '<file_path>'
make_new_models <- TRUE
if(make_new_models == TRUE){
  clf_grid <- h2o.grid(algorithm="randomForest",
                       grid_id="RF_Grid_Search",
                       x=predictors, y=response,
                       training_frame=d_train,
                       validation_frame=d_validation,
                       seed=54321,
                       nfolds=10, # number of folds to use for cross-
validation
                       balance_classes=T,
                       class_sampling_factors=c(undersample_rate,
base_undersample),
                       sample_rate_per_class=c(undersample_rate,
base_undersample),
                       histogram_type="AUTO",
                       hyper_params=rf_params,
                       search_criteria=search_criteria)

  #9. Save models to file
```

```
  h2o.saveGrid(grid_directory = file_path_to_output_grids_folder,
"RF_Grid_Search")
} else {
  clf_grid <- h2o.loadGrid(sprintf("%s/RF_Grid_Search",
file_path_to_output_grids_folder))
}

#########################################################
###########    Evaluate and rank models    ###########
#########################################################

#1. Write model ids and desired metrics to a table
model_ids_metrics <- tibble(model_id=character(),
                            xval_threshold=numeric(),
                            xval_recall=numeric(),
                            xval_precision=numeric(),
                            xval_auc=numeric(),
                            val_threshold=numeric(),
                            val_recall=numeric(),
                            val_precision=numeric(),
                            val_auc=numeric(),
                            optimal_threshold_for_test=numeric(),
                            test_recall=numeric(),
                            test_precision=numeric(),
                            test_auc=numeric(),
                            avg_recall=numeric(),
                            avg_precision=numeric(),
                            avg_auc=numeric(),
                            accuracy_1=numeric(),
                            accuracy_0=numeric(),
                            overall_accuracy=numeric(),
                            nbins=numeric(),
                            tp=numeric(),
                            fn=numeric(),
                            tn=numeric(),
                            fp=numeric())

## use recall calculated on max mean_per_class_accuracy threshold (row
10 of max crit and metric scores table)
for(i in 1:length(clf_grid@model_ids)){
  model_id_ <- clf_grid@model_ids[[i]]
  model_obj <- h2o.getModel(model_id_)

  xval_threshold <-
model_obj@model$cross_validation_metrics@metrics$max_criteria_and_metr
ic_scores[10,2]
  xval_perf <- h2o.performance(model_obj, xval=T)
  xval_recall <- h2o.recall(xval_perf, xval_threshold)[[1]]
  xval_precision <- h2o.precision(xval_perf, xval_threshold)[[1]]
  xval_auc <- h2o.auc(model_obj, xval=T)
  xval_conf_mat <-
h2o.confusionMatrix(model_obj@model$cross_validation_metrics,
xval_threshold)
  xval_acc_1 <- 1 - xval_conf_mat[[2,3]]
  xval_acc_0 <- 1 - xval_conf_mat[[1,3]]
```

```
  val_threshold <-
model_obj@model$validation_metrics@metrics$max_criteria_and_metric_sco
res[10,2]
  val_perf <- h2o.performance(model_obj, valid=T)
  val_recall <- h2o.recall(val_perf, val_threshold)[[1]]
  val_precision <- h2o.precision(val_perf, val_threshold)[[1]]
  val_auc <- h2o.auc(model_obj, val=T)
  val_conf_mat <-
h2o.confusionMatrix(model_obj@model$validation_metrics, val_threshold)
  val_acc_1 <- 1 - val_conf_mat[[2,3]]
  val_acc_0 <- 1 - val_conf_mat[[1,3]]

  avg_acc_1 <- (xval_acc_1 + val_acc_1) / 2
  avg_acc_0 <- (xval_acc_0 + val_acc_0) / 2
  overall_accuracy <- (avg_acc_1 + avg_acc_0) / 2
  avg_recall <- (xval_recall + val_recall) / 2
  avg_precision <- (xval_precision + val_precision) / 2
  avg_auc <- (xval_auc + val_auc) / 2

  optimal_threshold_for_test <- (xval_threshold + val_threshold)/2
  test_performance <- h2o.performance(model_obj, d_test)
  test_recall <- h2o.recall(test_performance,
optimal_threshold_for_test)[[1]]
  test_precision <- h2o.precision(test_performance,
optimal_threshold_for_test)[[1]]
  test_auc <- h2o.auc(test_performance)
  conf_mat <- h2o.confusionMatrix(test_performance,
optimal_threshold_for_test)
  tp <- conf_mat[2,2]
  fn <- conf_mat[2,1]
  tn <- conf_mat[1,1]
  fp <- conf_mat[1,2]

  nbins <- model_obj@allparameters$nbins

  model_ids_metrics <- model_ids_metrics %>%
    add_row(model_id=model_id_,
            xval_threshold=xval_threshold,
            xval_recall=xval_recall,
            xval_precision=xval_precision,
            xval_auc=xval_auc,
            val_threshold=val_threshold,
            val_recall=val_recall,
            val_precision=val_precision,
            val_auc=val_auc,
            optimal_threshold_for_test=optimal_threshold_for_test,
            test_recall=test_recall,
            test_precision=test_precision,
            test_auc=test_auc,
            avg_recall=avg_recall,
            avg_precision=avg_precision,
            avg_auc=avg_auc,
            accuracy_1=avg_acc_1,
            accuracy_0=avg_acc_0,
            overall_accuracy=overall_accuracy,
            nbins=nbins,
            tp=tp,
            fn=fn,
```

```
            tn=tn,
            fp=fp)
}

#2. Sort models in descending order of desired metric (mean (of
validation and cross validation) recall)
(model_ids_metrics <- model_ids_metrics %>% arrange(desc(avg_recall)))
(best_id <- model_ids_metrics[[1,1]])
best_clf <- h2o.getModel(best_id)
best_clf@allparameters

#3. Determine thresholds to use to split between predicted large fire
and predicted small fire conditions for best model
#   and print confusion matrix for valudation and cross-validation
parts of best model
(cross_validation_threshold <-
best_clf@model$cross_validation_metrics@metrics$max_criteria_and_metri
c_scores[10,2])
h2o.confusionMatrix(best_clf@model$cross_validation_metrics,
cross_validation_threshold)
(validation_threshold <-
best_clf@model$validation_metrics@metrics$max_criteria_and_metric_scor
es[10,2])
h2o.confusionMatrix(best_clf@model$validation_metrics,
validation_threshold)
(optimal_threshold_for_test <- (validation_threshold +
cross_validation_threshold)/2)

#4. Evaluate performance of best model on test data, and create
confusion matrix
#   using threshold determined from validation and cross-validation
thresholds
(test_performance <- h2o.performance(best_clf, d_test))
(test_conf_mat <- h2o.confusionMatrix(test_performance,
optimal_threshold_for_test))


############################################################
############    Determine variable importances    ############
############################################################
get_importances <- function(model_n){
  model_obj <- h2o.getModel(model_ids_metrics[[model_n,1]])
  model_id <- model_obj@model_id
  model_varimp <- as_tibble(h2o.varimp(model_obj)) %>%
    rowid_to_column(.) %>%
    mutate(model_id=replicate(length(predictors), model_obj@model_id),
           model_rank=replicate(length(predictors), model_n))
  return(model_varimp)
}
n_importances = sequence(10) # number of top models to get variable
importance for
all_var_imps <- map_dfr(n_importances, get_importances)
path_to_varimps_table = '<file_path>.csv'
write_csv(all_var_imps, path_to_varimps_table)


############################################################
############    Analyze predictions on test data    ############
```

```
############    or other unseen (e.g. forecasted    ############
############    future data or historical means)    ############
#################################################################

#1. Make predictions on test data and join predictions to original
test dataset
#   and calculate various evaluation metrics (e.g. interpolation
variances)
pred <- as_tibble(h2o.predict(best_clf, d_test,
threshold=optimal_threshold_for_test))
preds_table <- inner_join(as_tibble(pred) %>% rowid_to_column(),
                          as_tibble(d_test) %>% rowid_to_column(),
                          by="rowid") %>%
  mutate(model_rank=1) %>%
  mutate(conf_mat_result=case_when(predict == 1 & StartedFire == 1 ~
'tp',
                                   predict == 0 & StartedFire == 0 ~
'tn',
                                   predict == 1 & StartedFire == 0 ~
'fp',
                                   predict == 0 & StartedFire == 1 ~
'fn')) %>%
  mutate(crossover=ifelse(temp>rh, 1, 0),
         ffmc_70=ifelse(ffmc>70, 1, 0),
         ffmc_87=ifelse(ffmc>87, 1, 0),
         dmc_34=ifelse(dmc>34, 1, 0),
         isi_4=ifelse(isi>4, 1, 0)) %>%
  mutate(total_var_z=fwi_var_z +
           isi_var_z +
           bui_var_z +
           ffmc_var_z +
           dc_var_z +
           dmc_var_z +
           temp_var_z +
           rh_var_z +
           ws_var_z +
           precip_cube_root_var_z) %>%
  mutate(dist_from_threshold=abs(p1-optimal_threshold_for_test))
path_to_predictions_results = '<file_path>.csv'
write_csv(preds_table, path_to_predictions_results)

#2. Analyze mis-classification rate by ecozone
test_obs_per_ecozone_and_test_result <- preds_table %>%
  group_by(ZONNAME, conf_mat_result) %>%
  filter(conf_mat_result %in% c('fn', 'tp')) %>%
  tally() %>%
  spread(key=conf_mat_result, value=n) %>%
  mutate(recall=tp/(tp+fn))

test_obs_per_ecozone_and_test_result %>%  write_csv('<file_path>.csv')


#3. Analyze mis-classification rate by fuel type
test_obs_per_fuel_type_and_test_result <- preds_table %>%
  group_by(fuel_type_adj_2, conf_mat_result) %>%
  filter(conf_mat_result %in% c('fn', 'tp')) %>%
  tally() %>%
  spread(key=conf_mat_result, value=n) %>%
```

```
  mutate(recall=tp/(tp+fn))

test_obs_per_fuel_type_and_test_result %>%
write_csv('<file_path>.csv')
```