Improved Forest Fire Danger Rating Using Regression Kriging with the Canadian Precipitation Analysis (CaPA) System in Alberta

by

Xinli Cai

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in
Forest Biology and Management

Department of Renewable Resources

University of Alberta

Abstract

Estimating precipitation is currently one of the key challenges of accurate Fire Danger Rating. New gridded precipitation analysis systems such as the Canadian Precipitation Analysis System (CaPA) may be superior to the current analytical interpolation strategies, thin plate spline (TPS), and inverse distance weighting (IDW). To compare the performance of CaPA and interpolation methods in the forested area of Alberta, I evaluated precipitation estimates from CaPA and 17 algorithms of five interpolation methods, including IDW, smoothed TPS, non-smoothed TPS, ordinary kriging, and regression kriging. Precipitation estimates were generated using station observations through leave-one-out cross-validation and were evaluated using a range of skill scores (MAE, Bias, ETS, and FBI). I also assessed impacts of these precipitation estimates on the Canadian Forest Fire Weather Index (FWI) System and examined the impacts of weather station density on model performance.

Results show that CaPA was only a mid-tiered method (13^{th} of 18), except in Doppler radar covered areas, where CaPA performed second best. Regression kriging (with CaPA as a covariate) was the best method and produced precipitation estimates with 19.6% lower MAE compared with IDW. I found that the best method shifted with station density; CaPA was the best method when fire weather station density fell below 0.6 stations per 10 000km^2 while regression kriging was the best method above this threshold. Additionally, this study showed that the FWI System responded to precipitation estimates differently due to their varying drying time lag of the indexes. Quick drying indexes (FFMC, ISI, FWI) preferred methods with lower MAE (e.g., regression kriging with smoothing), while slow drying indexes (DMC, DC, BUI) preferred methods with lower Bias (e.g., regression kriging without smoothing). Overall, I recommend the use of regression kriging with CaPA as a covariate to estimate fire danger across landscapes.

Acknowledgments

The completion of this research would not have been possible without many supporters! The greatest gratitude is to my supervisor, Dr Mike Flannigan., who provided consistent support and encouragement throughout this study. I feel very lucky to have you guide me through this journey and your knowledge in fire - weather interactions greatly helped in deepening this study! I appreciate the insightful questions and suggestions made by my examing committee, Dr Jen Beverly (Chair), Dr Nadir Erbilgin, and Dr Xianli Wang.

Thank you Dr Xianli Wang, who was always available to help. Your thoughtful advice in organising the ideas, coding, and writing have greatly helped me in my academic development. Very special thanks to Dr Piyush Jain, who greatly helped in the statistical portions of this research. I really appreciated our coffee break conversations and hope some of your amazing technical skills have rubbed off on me! I am also thankful for the great feedback provided by Bo Lu from the CWFIS group!

To the students in the fire lab group, thank you for your love and support. The highs and lows that we shared together are some of my best memories in Canada! Finally, I would like to thank my family, especially my parents, who always support me and encourage me to chase my dreams!

Thank you to the University of Alberta and staff in the Department of the Renewable Resources in providing support through every stage of my studies. The Western Partnership for Wildland Fire Science, Alberta Agriculture and Forestry, and Chinese Scholarship Council provided the funding for this research. Fire weather station data were provided by the Alberta Agriculture and Forestry with help from Brett Moore and CaPA System data were provided by Environment and Climate Change Canada with help from Chelene Hanes.

Table of Content

List of Tables

List of Figures

List of Abbreviations

AAF	Alberta Agriculture and Forestry
ANOVA	Analysis of Variance
Bias	Mean Error
BUI	Built Up Index
CaPA	Canadian Precipitation Analysis System
cbrt	Cubic Root Transformation
CFFDRS	Canadian Forest Fire Danger Rating System
CWFIS	Canadian Wildland Fire Information System
DC	Drought Code
DMC	Duff Moisture Code
ECCC	Environment and Climate Change Canada
ETS	Equitable Threat Score
FBI	Frequency Bias Index
FBP	Fire Behavior Prediction
FFMC	Fine Fuel Moisture Code
FWI	Fire Weather Index
GCV	Generalized Cross Validation
GEM	Global Environmental Multiscale
GPM	Global Precipitation Measurement
IDEW	Inverse Distance Elevation Weighting
IDW	Inverse Distance Weighing
ISI	Initial Spread Index
ln	Natural Logarithm Transformation
LO	Lookout Towers
LOOCV	Leave-one-out Cross-validation
MAE	Mean Absolute Error
MET	Environment Canada METARS
METAR	Aviation routine weather report
NWP	Numerical Weather Prediction
OK	Ordinary Kriging
PCP	Precipitation
QPE	Quantitative Precipitation Estimate
RAWS	Remote Automatic Weather Stations
RDPA	Regional Deterministic Prediction System
RH	Relative Humidity
RK	Regression Kriging
rk(capa)	Regression Kriging with CaPA estimates
RMCQ	Réseau Météorologique Coopératif du Québec

RMSE	Root Mean Square Error
ROS	Rate of Spread
RS	Ranger Stations
RZ	Contract Stations
SBB	Stationary Block Bootstrapping
SHEF	American Meteorological cooperative
sqrt	Square Root Transformation
SYNOP	Automatic synoptic station
Temp	Temperature
TPS	Thin Plate Spline
tps_ns	Thin Plate Spline Non Smoothing
tps_s	Thin Plate Spline Smoothing
WS	Wind Speed
WUI	Wildland-Urban Interface
wstn	Weather Station

Chapter 1 Introduction

1.1 Context

Wildland fires, ranging from stand to landscape levels, frequently occur in the Canadian boreal forest ecosystem. Fire is a dominant disturbance in the boreal forest; it drives the physical and ecological dynamics of forest composition, density, productivity, as well as carbon cycling and storage (Rowe 1983). Fire also threatens communities in the boreal forest and causes damages measured in billions of dollars (e.g., the 2011 Slave Lake fire and the 2016 Fort McMurray fire). In Canada, fire is suppressed when it threatens human life, infrastructure, and valuable resources. Accurate prediction of fire activity is challenging but is needed for effective suppression. Fire activity is governed by factors like weather, fuel, topography, ignition sources, and humans (Johnson 1972; Flannigan *et al.,* 2005). Of these factors, weather varies the most in space and time, but weather is the best predictor of daily fire activity (Flannigan *et al.,* 2005; Parisien *et al.,* 2011).

To quantify the impacts of daily weather on fire danger, Canadian scientists developed the Canadian Forest Fire Weather Index (FWI) System (Van Wagner 1987). The FWI System is a strictly weather-based system and relies on noon LST weather observations of temperature, precipitation, wind speed, and relative humidity. Weather inputs of the FWI System are recorded on a station-by- station basis and outputs of the FWI System are typically accurate within a 40 km radius (Turner and Lawson 1978). Thus, the usefulness of the FWI System for predicting conditions over large areas is dependent on the density and distribution of weather stations. However, weather stations are sparsely distributed in the boreal forest, making FWI System less accurate in areas without nearby stations. To obtain accurate FWI in these areas, fire management agencies interpolate observations from surrounding weather stations using algorithms such as inverse distance weighting (Englefield *et al.,* 2000) and thin-plate splines (Flannigan *et al.,* 1998). Other more sophisticated interpolation algorithms may lead to better predictions of FWI System. Alternative real-time spatial weather products have recently become available and may also result in more accurate fire danger rating in data scarce areas.

One of the biggest problems in estimating fire danger is the accuracy of summer precipitation (Flannigan and Wotton 1989). This is because most precipitation in the summer results from convective storms and thunderstorms, which is highly localized and varies over short distances. Other fire weather inputs (i.e., temperature, relative humidity, and wind speed) are spatially less variable compared with the precipitation field, making the usefulness of current interpolation algorithms constrained by the variability of precipitation. This thesis addresses the challenges of obtaining accurate precipitation estimates in data scarce areas by evaluating a gridded precipitation product and multiple advanced interpolation algorithms in the forested area of Alberta.

1.2 Wildland fire in Canada

Wildland fire is the dominant natural disturbance in the Canadian boreal forest; it alters the structure and function of the boreal ecosystem by reshaping vegetation composition, density, and biogeochemical cycling (Rowe 1983). From 2003 to 2013, an average of 7 000 wildland fires occurred each year burning over 2 million hectares in Canada (CIFFC 2013). The size of fires in Canada vary widely; with only 3% of fires growing larger than 200 hectares but representing 97% of the area burned (Stocks *et al.,* 2003). Wildland fires in Canada are ignited by two common sources: human and lightning. Compared to human-caused fires, lightning-caused fires only represent one-third of the total fires but account for 85% of the area burned (Krezek-Hanes *et al.,* 2011). This is because lightning-caused fires can occur in remote areas of the boreal forest, where detection and suppression are often more challenging. Wildland fires are classified into three types: ground, surface, and crown. Ground fires burn in the organic matter below the litter layer where they can burn deep into the peat and smoulder over winter. Surface fires burn the litter and duff by flaming combustion with low intensity. Crown fires burn canopy or tree crowns at high intensity. Under favourable burning conditions and the presence of ladder fuel, ground fire and surface fires can turn into crown fires. In Canada, most of the areas burned are attributed to the stand-renewing, high-intensity crown fires (Stocks *et al.,* 2003).

In the past decade, an average of $800 million was spent annually on fire management in Canada (Stocks and Martell 2016). Fire management involves the activities concerned with protecting people, property, and forest areas from wildland fires and the use of prescribed

burning for the attainment of forest management and other land use objectives (Merrill and Alexander 1987). Fire management in boreal forest is the responsibility of each province and territory, whereas the federal government manages and suppresses fires within national parks and military bases. Not all fires are suppressed as the ecological impacts of fires have been recognized by the fire managers. In Canada, the levels of fire suppression range from complete extinguishment to "let it burn", determined by a hierarchy of priorities (i.e., environmental, social, and economic criteria) set by the local fire management agency (Martell 2001).

Attitudes towards wildland fire suppression in Canada have shifted over the years. As described by Taylor and Alexander (2006), "fire management" in Canada is a term that was originally used to describe the unlimited use of wildland fire by rural and Aboriginal people as a part of traditional land management practices, such as to clear land for farming. In the late 19th century, suppression of wildland fire was formalized by the government and fire laws were instituted to restrict traditional wildland fire practices. In the early 20th century, fire management agencies attempted to suppress all fires and restricted the use of prescribed burns for fuel management. Fire managers eventually realized that the unnatural suppression of all fires was undesirable to the boreal ecosystem and consequently led to an increase of homogenous and ageing stands. These stands were more likely to escape the initial attack and carried a greater risk of causing catastrophic fires. By the 1970s, the recognition of the need for fires on the landscape had grown. Therefore, fire management strategies in Canada shifted from strict suppression of all fires to allow fires burn in remote areas when lives or industry were not threatened.

Currently, one of the biggest challenges in wildland fire management is responding to the recent increase in fire activity, most of which could be attributed to human-introduced climate change (Flannigan *et al.,* 2009; Beverly *et al.*, 2011; Moritz *et al,.* 2012). In Canada, warming temperatures have resulted in an increased fire season length and area burned compared to the past four decades (Gillett *et al.,* 2004), all of which will continue with rise according to future projections (Wotton and Flannigan 1993; Flannigan *et al.,* 2005). Although an increase of precipitation could compensate for the drying caused by warmer temperatures, Flannigan *et al* (2016) suggested that for each degree of warming, precipitation needs to increase by 5%-15% from the baseline (1971-2000) to compensate. Due to this sensitivity of fuel moisture to temperature and precipitation, Flannigan *et al* (2016) further suggested that temperature

increases will result in a future with more frequent extreme fire weather conditions. The increase of these extreme fire weather conditions could lead to a substantial increase in the days with active fire spread (Wang *et al.,* 2015) and the occurrence of large fires (de Groot *et al.,* 2013), all of which will challenge fire management capacity.

The expansion of Wildland-Urban Interfaces (WUI) in recent decades also presents significant challenges to fire management (Johnston 2016). Fire suppression in the WUI is difficult because the WUI can create an environment in which fires can move rapidly between structural and vegetation fuels. A failure of initial attack in WUI often leads to the development of catastrophic fires that threaten human lives and cause millions of dollars' worth of damages. Two recent examples are the 2011 Slave Lake fire, which has cost over $1 billion for suppression, evacuation, insurance, and recovery (Flat Top Complex Wildfire Review Committee 2012) and the 2016 Fort McMurray fire, which destroyed over 2 400 structures and led to the evacuation of 88 000 people (Benfield 2016).

As fire management in Canada is challenged by the current and future uncertainties, it is necessary to improve current information systems to support the fire management decision-making process (Lee *et al.,* 2002). To accomplish this goal, we must have a better understanding of expected daily fire activity, which is largely influenced by the interactions between the surrounding environmental conditions, ignition sources, and human activities (Johnson 1972; Flannigan *et al.,* 2005). In particular, one of the major areas for improvement is fire danger rating, which is largely impacted by the surrounding fire environment.

1.3 Fire environment

Fire environment is the surrounding conditions of fuel, topography, and weather that influence or modify the behaviour of fire (Countryman 1972). Predicting fire behaviour is challenging due to the complex interactions between fire environment components, where changes in one environmental component will influence the rest of the environmental component and resultant fire behaviour. Among the three components of fire environment, topography is relatively constant through time, while weather conditions change significantly in space and time. McArthur (1968) showed that head fire spread rates double on 10-degree slopes compared with

spread rates on flat terrain. These rates increase further when the combined effect of slope and wind are considered. Fuel (including load, type, structure, continuity, and moisture content) is critical to fire occurrence and spread. Fuel load and fuel continuity are necessary for fires to start and spread, while fuel structure is important for the interplay of fire dynamics. For example, understory shrubs on the forest floor can act as ladder fuels that may spread surface fires vertically and result in high-intensity crown fires (Flannigan *et al..,* 2009). Fuel moisture largely determines fire behaviour, suppression effectiveness, and has been correlated with area burned (Flannigan *et al.,* 2005).

The relative importance of topography, fuel, and weather in determining fire behaviour varies across the scale of landscape (Heyerdahl *et al.,* 2011). At the stand level, fire behaviour varies mainly as a function of site-specific factors (e.g., fuel structure and slope effect). For example, study has shown that the burn probabilities modelled by Burn-P3 (a spatial fire simulation model, see Parisien *et al.,* 2005) in Wood Buffalo National Park were attributed by fuels, weather, and ignition in 67.4%, 29.2%, and 3.4%, respectively (Parisien *et al.,* 2011). On the other hand, weather-climate is the dominant factor influencing fire behaviour at the ecosystem level. This is because weather varies the most at space and time.

Weather, including temperature, precipitation, wind speed, and atmospheric moisture, is arguably the best predictor of daily fire activity in the Canadian boreal forest (Flannigan and Harrington 1988; Flannigan *et al.,* 2005; Cary *et al.,* 2006). The impacts of weather on fire activity are two-fold: first, weather influences the fuel flammability and sources of ignition (lightning). Long term weather determines the climate of a region, which will affect the type and amount of fuel at a given location. Second, weather directly impacts fire activity. Studies have shown that temperature is the most significant factor in determining the annual wildland fire activity, as areas with warmer temperatures burn more (Gillett *et al.,* 2004; Wang *et al.,* 2014). Also, the variability of precipitation (i.e., timing and location) also impacts fire activity, where consecutive days without significant precipitation events may lead to extreme conditions and catastrophic fires (Holden *et al.,* 2007, Jen and Martell 2005). Studies have shown that hot, dry, and windy conditions were of critical importance in promoting the occurrence of large fires in areas with sufficient fuels (Balshi *et al.,* 2009). In Western Canada, these critical burning conditions are results of the summer blocking high-pressure patterns, which creates a long period

of dry days that is correlated with the area burned (Flannigan and Harrington 1988). A recent example of fire under these hot, dry, and windy conditions is the 2016 Fort McMurray fire. Therefore, a better understanding of the interplay between weather conditions and fire activity is necessary for efficient fire management.

1.4 Fire danger-weather interactions

Fire danger rating is the assessment of fire environment factors (i.e., fuel, topography. and weather) that determine the ease of ignition, rate of spread, difficulty of control, and the impacts of fires (Merrill and Alexander 1987). In Canada, scientists have conducted research on fire danger rating since 1925, which led to the development of the Canadian Forest Fire Danger Rating System (CFFDRS). The CFFDRS (Stocks *et al.,*1989) is a semi-empirical modular system that assists fire managers in optimizing the management actions (e.g., preparedness planning, resources pre-positioning, alert levels) through four linked components. These components include Fire Weather Index (FWI) System, Fire Behavior Prediction (FBP) System, Fire Occurrence Prediction (FOP) System, and the Accessory Fuel Moisture (AFM) System. The CFFDRS is used by all Canadian fire management agencies during the fire season and has been adopted by other countries (e.g., New Zealand, Mexico, and Portugal). A detailed description of the CFFDRS components, inputs, and usage considerations could be found in Taylor and Alexander (2006), Lawson and Armitage (2008), and Wotton (2009).

The FWI System is one of the core components of CFFDRS and is used as the fundamental unit of fire danger rating in Canada. The FWI System (Van Wagner 1987) is a weather-based system that relies on solar noon daily weather observations of screen-level temperature and relative humidity, 10 m open space wind speed, and 24-hour precipitation. The FWI System produces three indicators of fuel moisture content at different forest floor depths and three indicators of the maximum daily fire behaviour potential in a standard pine forest type. Fuel moisture content of the surface fine dead fuel is tracked by Fine Fuel Moisture Code (FFMC), which is used to inform the sustainability of surface fire spread (Beverly and Wotton 2007) and is linked to the occurrence of human-caused fires (Wotton *et al.,* 2003). Fuel moisture content of the loosely compacted organic material is tracked by Duff Moisture Code (DMC), which is important to the sustainability of shouldering and is linked to the occurrence of lighting-

cause fires (Wotton and Martell 2005). Fuel moisture in the deep compacted organic soil layer is tracked by Drought Code (DC), which is a useful indicator of drought conditions. FWI System fuel moisture codes are also directly related to three fire behaviour indexes. FFMC combined with wind speed produces an indicator of the potential spread of a fire without consideration of fuel types (Initial Spread Index, ISI); DMC and DC are combined to produce the potential total fuel consumption (Buildup Index, BUI). Finally, BUI and ISI are combined to generate the Fire Weather Index (FWI), which is used as an overall indicator of the intensity of a spreading fire and is used as the main indicator of fire danger in a specific area to inform public.

Outputs of the FWI System are also used as inputs to the FBP System. The FBP is another core component of the CFFDRS and takes the outputs of the FWI System along with site-specific inputs (e.g., slope and moisture content of live fuels) to provide quantitative estimates of fire growth and fire intensity for 16 benchmark fuel types (Forestry Canada Fire Danger Group 1992). These two core components of the CFFDRS have been used to support the predictions of fire growth and the mapping of burn probability. For example, two widely used fire management tools; the Prometheus Fire Growth Simulation Model (Tymstra *et al.,* 2010) and the Burn-P3 model (Parisien *et al.,* 2005) both use the FWI and FBP System models. The FWI system has also been used extensively in training fire management crews and understanding the interactions between wildland fires and climate change (Flannigan *et al.,* 2009).

1.5 Previous studies on fire weather interpolation

As mentioned in Section 1.1, Turner and Lawson (1978) indicated that the FWI System was only representative of a 40 km radius surrounding the weather station or where they were calculated. This radius varies according to topography and proximity to large water bodies. Since weather stations in the boreal forest are sparse, there is a need to interpolate the available station data to build reliable spatial FWI System products. The goal of interpolation is always the same: each method tries to estimate a weather variable at an unmeasured location by applying weights to observations from the surrounding areas. Canadian fire management agencies use either inverse distance elevation weighting or more sophisticated strategies such as cubic splines for the interpolation of fire danger across the landscape. According to Flannigan and Wotton (1989), both of these methods may not address the spatial variability of fire weather variables, especially accumulated daily precipitation.

Overview of various interpolation algorithms in environmental sciences can be found in Daly (2006) and Li and Heap (2011), while basic concepts of interpolation can be found in Cressie (1993). Recent algorithms developed using machine learning and random forests can be found in Sanabria *et al* (2013). Among the many types of the interpolation algorithms as reviewed by Li and Heap (2014), ordinary kriging, thin-plate spline, and universal kriging were the most frequently compared algorithms. Consensus suggests that in regions without topographical features, inverse distance weighting and ordinary kriging are sufficient if the weather stations can capture the major climate patterns (Daly 2006; Yang *et al.,* 2015). Regions characterized by significant topographical features can be reasonably addressed by interpolation algorithms that account for elevation, such as trivariate thin-plate spline smoothing (Hutchinson 1995) and regression kriging with elevation as an auxiliary variable (Ly *et al.,* 2011). Overall, the best performing interpolation algorithm varies as a function of the spatial structure of interpolated variables, station density, and terrain features.

Most studies evaluate interpolation algorithms using various climate variables (e.g., daily min/max temperature, monthly temperature, and precipitation), while few studies focus primarily on interpolating daily precipitation. Summer Precipitation is difficult to interpolate because precipitation in summer is primarily convective in nature, which could vary over short distances (Flannigan and Wotton 1989). Thin-plate spline was found to be superior to the distance weighting algorithm in interpolating monthly precipitation across Canada (Price *et al.,* 2000). However, thin-plate spline may have difficulty in simulating the daily precipitation data over a mountainous region (Shen *et al*., 2001, Hutchinson *et al.,* 2009). There are some studies that have evaluated using precipitation from radar or gridded weather products to estimate FWI Indexes, such as Flannigan *et al* (1998) and Horel *et al* (2014). These two studies both reported that gridded weather information resulted in improvement of FWI System other than indexes like DMC and DC, which is sensitive to seasonal accumulated precipitation. Recently, the geostatistical interpolation algorithm (i.e., regression kriging) was found to be an improvement in estimating hourly precipitation and could be used to combine station observations with radar data (Haberlandt 2007). To date, advanced geostatistical interpolation algorithms such as regression kriging have not been tested in the FWI System.

1.6 Canadian Precipitation Analysis (CaPA) System

Gridded precipitation analysis systems have been recently developed to improve the estimates of precipitation in data scarce regions. Three precipitation analysis systems cover Canada, including the Canadian Precipitation Analysis (CaPA) System (Mahfouf *et al.,* 2007) and two satellite-based precipitation estimates, the Global Precipitation Measurement (GPM) (Smith *et al.,* 2007) and Precipitation Estimation from Remote Sensing Information using Artificial Neural Network (PERSIANN) (Sorooshian *et al.,* 2005). Currently, none of these precipitation estimates has been used in support of fire danger rating in Canada. In this study, the CaPA System was chosen to compare with multiple interpolation methods in estimating accumulated daily precipitation in Alberta.

Environment and Climate Change Canada (ECCC) developed the CaPA System to improve precipitation estimates with sparse station network. The CaPA System is a data assimilation system that produces near real-time gridded precipitation analysis at a 10km resolution over North America (Mahfouf *et al.,*2007). The CaPA System applies an optimal interpolation technique (Daley 1991) to merge station observations, radar quantitative precipitation estimates (QPEs), and Global Environment Multiscale (GEM) model forecasts (Fortin *et al.,*2015).

The development of CaPA System started in 2003, and its first operational applications were released in 2011. Mahfouf *et al* (2007) first documented described introduction in algorithms of CaPA System.A bias-correction method was later recommended to reduce the negative bias caused by a cubic root transformation of the CaPA system inputs (Evans 2013). Lespinas *et al* (2015) compared estimates of CaPA System with the raw GEM forecasts over Canada; this study showed that the CaPA System was significantly superior to GEM forests when ECCC station density exceeded 1.17 weather stations / 10 000km^2. Lespinas *et al* (2015) also suggested that the CaPA System performed worse in summer as the GEM models can not accurately predict the convective precipitation. To improve the performance of CaPA System, Fortin *et al* (2015) successfully calibrated the radar Quantitative Precipitation Estimates (QPEs) into CaPA System, all of which has proved to be a significant improvement in prediction accuracy.

1.7 Research objectives

In this study, I address the challenges of estimating precipitation between weather stations and assess the implications of improved precipitation estimates on fire danger rating. To accomplish these objectives, I first compare the performance of the CaPA System with the current operationally-used interpolation algorithms (IDW and TPS based algorithms) and kriging based interpolation algorithms in estimating precipitation. I also compare the performance of the CaPA System and multiple interpolation algorithms in regions with or without Doppler radar coverage. Second, I examine how precipitation estimation errors will impact the fuel moisture codes and the fire behaviour indexes of the FWI System. Lastly, I study how the performance of interpolation algorithms changes with different scenarios of weather station densities. In this study, the forested area of Alberta, Canada was chose as study area.

Chapter 2 Data and methods

2.1 Study area

This study was performed in the forested portion of the province of Alberta, Canada, which lies between 49° - 60° N and 110° - 120° W (Figure 2.1a). With an area of ~661,000 km^2, Alberta is classified into six natural regions due to its diverse geography (Figure 2.1b). The northern half of the province is primarily covered by the Boreal forest with a small portion of the Canadian Shield in the northeast corner. The Foothills transition between the Boreal forest in the north and the Rocky Mountain along the south-west border. Much of south-east Alberta is covered by Aspen Parkland and Grassland (Natural Regions Committee 2006). The study areas fall primarily within the Boreal, Foothills, and parts of the Rocky Mountain natural regions, where wildland fire activity and suppression mainly occur.

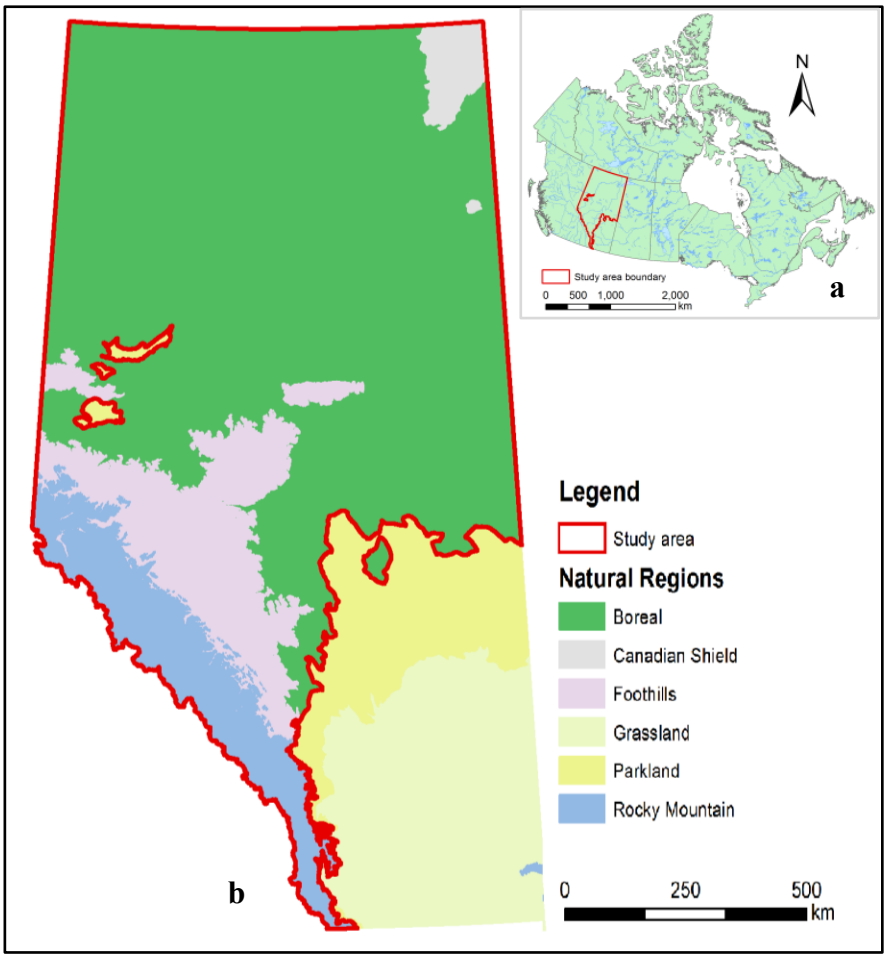


Figure 2.1 (a) Location of the study area and (b) the natural region types in the study area.
Data obtained from Alberta Agriculture and Forestry.

The climate in the study area is characterized by cold, dry winter, and short warm summers (Natural Regions Committee 2006). Historical mean January temperatures range from below -24°C in the Boreal to -12°C in the Foothills; mean July temperatures range from below 13°C in the Rocky Mountains to 15°C in the Boreal (Government of Alberta 2015). Annual precipitation in the study area ranges from a minimum of 400 mm in the northeast to more than 600 mm in the south-west (Figure 2.2). In general, the Foothills natural region is wetter than the Boreal natural region. In Alberta, about 60% of the annual total precipitation falls in the summer months; most of which is a result of highly localized convective storms and thunderstorms (Strong and Leggar 1992). Due to the convective nature of summer precipitation (e.g., rain of 40mm at one location while no rain 15 km away), interpolating fire danger is difficult in areas without nearby weather station.

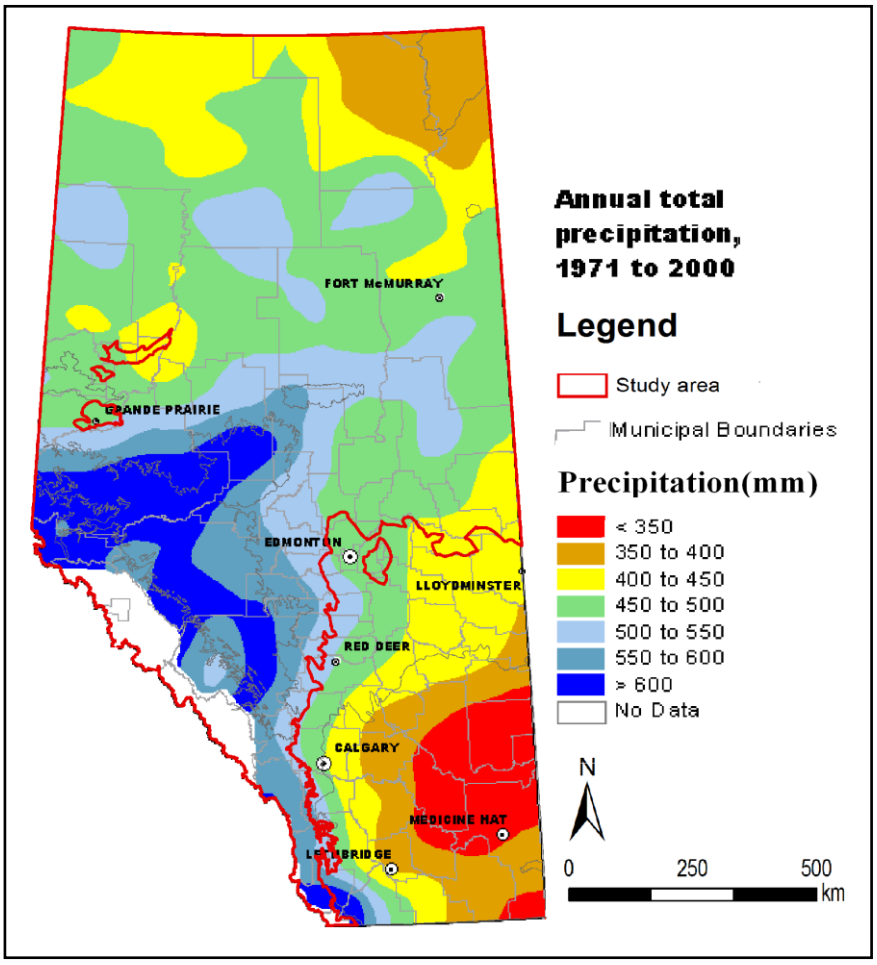


Figure 2.2 Annual precipitation (mm) from 1971 to 2000 in study area
Map displayed on generalized townships. Data obtained from Environment and Climate Chance Canada.

Elevations in the study area range from ~200 m in the north to ~3500 m in the south-west (Figure 2.3). Much of the study area is relatively flat with an elevation below 700 m, while the mountainous areas in the south-west exceed 3000 m (Figure 2.3). Elevation influences the weather, for example, annual total precipitation in our study area decreases with elevation, and so does annual mean temperature for areas within similar latitudes.

Fuel types in the study area are common tree species in the Boreal forest, including trembling Aspen (*Populus tremuloides*), balsam poplar (*Populus balsamifera*), jack pine (*Pinus banksiana*), white sprue (*Picea glauca*), black spruce (*Picea mariana*), lodgepole pine (*Pinus contorta*), and tamarack (*Larix laricina*). These species are fire prone and some of them rely on wildland fires to regenerate.

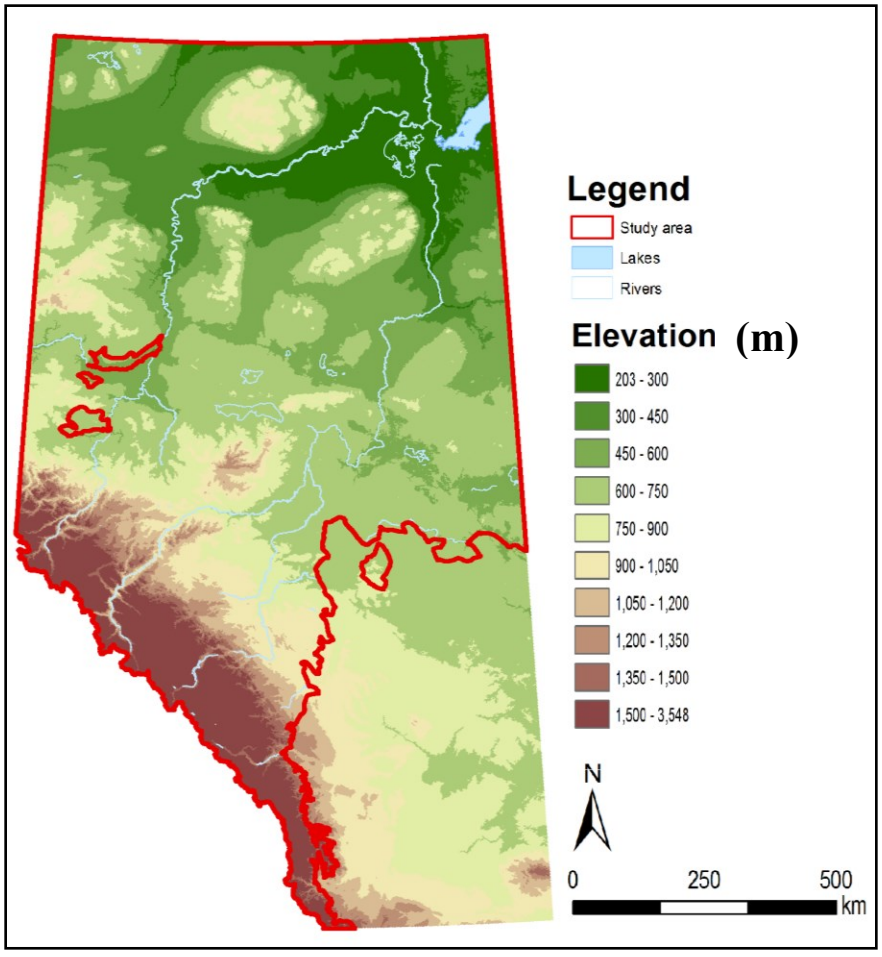


Figure 2.3 Elevation and hydrological networks in study area
Data obtained from Alberta Agriculture and Forestry.

2.2 Data

Fire weather station observations and Canadian Precipitation Analysis (CaPA) System outputs were used in this study. These data were collected from July 14th to August 31st, 2014 (N=49 days) and from May 1st to August 31st, 2015 (N=124 days) in the study areas. The 2014 study period is shorter because the newest CaPA System that incorporated radar precipitation began to produce reliable outputs on July 14, 2014 (Fortin *et al.,* 2015).

To reduce the edge effect caused by the lack of stations beyond the Alberta border, we buffered the study area inwards by 75 km and divided the study area into a validation area and a buffer area (Figure 2.4). Weather stations within the validation area were used as validation points, where the precipitation was estimated with a variety of methods. The validation area was also partitioned in two different ways: (1) radar and non-radar sub-areas using the 120 km Doppler radar range and (2) Boreal and Foothills sub-regions (Figure 2.4).

2.2.1 Fire weather station data

Alberta Agriculture and Forestry (AAF) operates a weather station network (Figure 2.4), recording daily observations of surface temperature (Temp), relative humidity (RH), 10 m wind speed (WS), and 24h precipitation (PCP) at 1200 hour Local Standard Time. These observations are used to calculate fuel moisture codes and fire behaviour indexes using the FWI System (see Section 2.4 for details). In this study, weather observations of AAF stations and indexes of the FWI System calculated by these observations were used as true values.

AAF station network is unevenly distributed in the study area with fewer stations in Northern Alberta (Figure 2.4). The station network consists of five types of stations: Lookout Towers (LO), Ranger Stations (RS), Contract Stations (RZ), Remote Automatic Weather Stations (RAWS), and Environment Canada METARS (MET). We excluded MET observations from the analysis because it is already calibrated into the CaPA System station networks (see Section 2.2.2. for details). In 2014 and 2015, 100 and 81 weather stations, respectively remained in validation areas after data cleaning (Table 2.1). In areas of radar and non-radar, the station density is similar; while Foothills had ~150% higher station density than the Boreal (Table 2.1). AAF station network details are presented in Appendix 1.

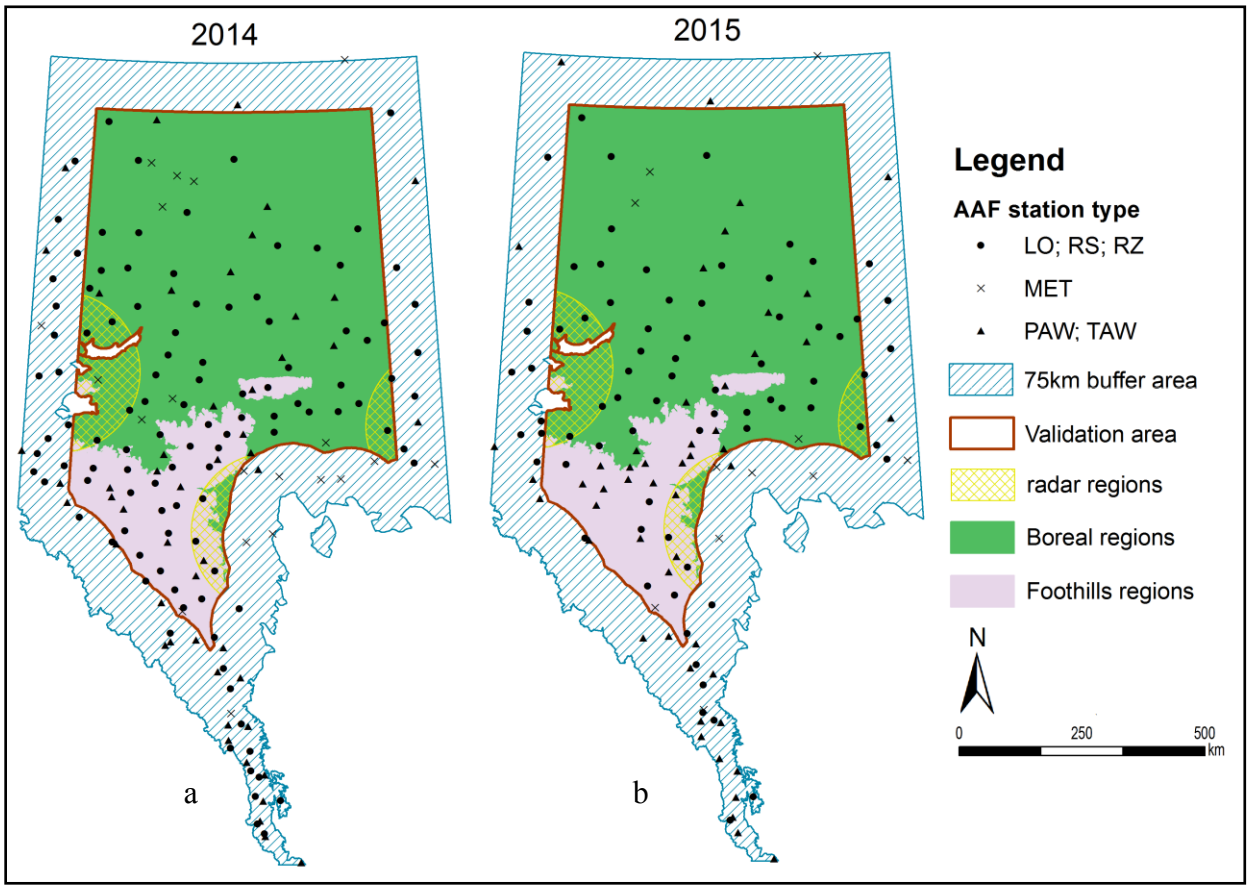


Figure 2.4 Alberta Agriculture Forest station network in (a) 2014 and (b) 2015
Study areas are divided into validation area and 75km buffer area. Validation area is divided into radar and non-radar sub-regions using the 120km Doppler radar range; validation area is also divided into Boreal and Foothills sub-regions. Stations are displayed by type.

Table 2.1 Summaries of weather stations in study areas in 2014 and 2015

Year	Region	Area (km^2)	No. of stations	No. of stations per 10 000 km^2
2014	Study area	505 964	179	3.54
	Validation area	280 679	100	3.56
	radar	28 010	9	3.21
	non-radar	252 669	91	3.60
	Boreal	229 248	64	2.79
	Foothills	49 128	36	7.33
2015	Study area	505 964	136	2.69
	Validation area	280 679	81	2.89
	radar	28 010	8	2.86
	non-radar	252 669	73	2.89
	Boreal	229 248	53	2.31
	Foothills	49 128	28	5.70

We cleaned the AAF station data in two steps: 1) we excluded extreme values such as Temp >45^{o}C or <-45^{o}C; 24h PCP >200 mm; RH >100%; and WS >150 km/h; 2) we estimated missing data to produce continuous records for FWI System calculations. For stations with less than three consecutive days of missing data, the missing data were interpolated from the surrounding stations using thin plate spline interpolation method; for stations that had more than 3 consecutive days of missing data, the stations were dropped from the analysis to ensure the reliability of the dataset.

In summary, the summer months of 2014 and 2015 were both drier than the 30 years historical normal (i.e., 1984-2013) (Figure 2.5). In 2014 and 2015, the distribution of rain events was uneven across the validation areas and was highly skewed toward zero (Table 2.2). In 2015, the mean daily precipitation in the Foothills region was 31% larger than the Boreal region; but radar and non-radar areas received similar amounts of mean daily precipitation (Table 2.2). The probability of 24h precipitation >0.5 mm, >1.5 mm, and >2.8 mm, (i.e., the minimal precipitation amounts needed to change fuel moisture codes of the three fuel layers as described in the FWI System, see Section 2.4 for details) was low (Table 2.2).

Table 2.2 Distribution of daily precipitation in study areas

Year	Region	24h precipitation distribution and means (mm)[1]					Probability of effective PCP for fuel moisture codes[2]		
		mean	50th	75th	90th	max	>0.5 mm	>1.5 mm	>2.8 mm
2014	Validation area	1.7	0	0	4.5	94.4	27.6%	19.9%	14.5%
	radar	2.0	0	0.6	5.4	94.4	27.6%	20.0%	16.0%
	non-radar	1.7	0	0.8	4.8	75.8	27.3%	19.6%	14.3%
	Boreal	1.3	0	0.4	3.3	49.7	22.5%	15.8%	11.1%
	Foothills	2.4	0	1.9	6.8	94.4	36.5%	27.2%	20.5%
2015	Validation area	1.8	0	1.2	5.33	64.2	31.6%	22.8%	16.0%
	radar	2.0	0	1.4	6.8	44.1	34.3%	24.8%	18.1%
	non-radar	1.7	0	1.2	5.2	64.2	31.3%	22.6%	15.8%
	Boreal	1.6	0	1.1	4.8	52.6	31.2%	22.3%	15.3%
	Foothills	2.1	0	1.3	6.8	64.2	32.2%	23.7%	17.3%

[1] There are 4900 (49 days x100 weather stations) and 10044 (124 days x 81 weather stations) observations in 2014 and 2015 respectively.
[2]PCP amount starts to change the value of fuel moisture codes FFMC, DMC, and DC (Van Wagner 1987).

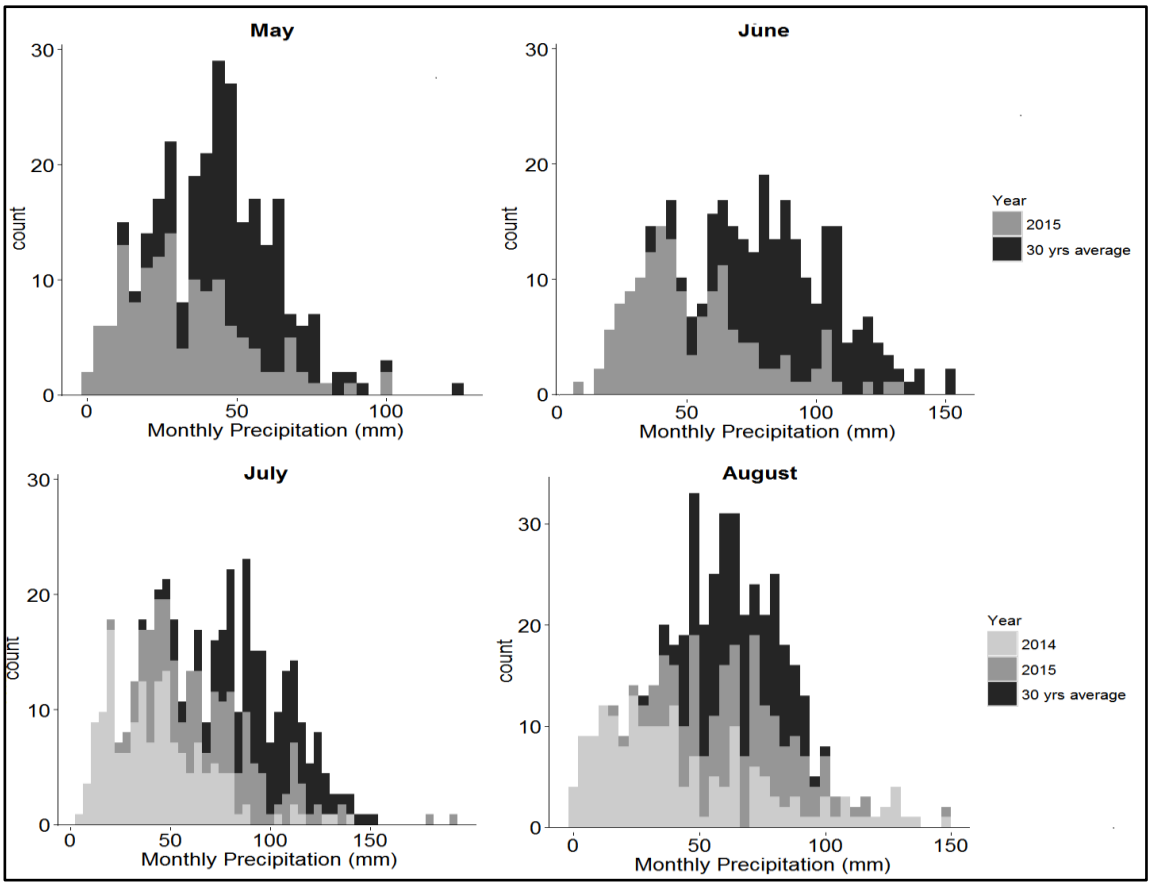


Figure 2.5 Daily precipitation frequency in 2014, 2015 and 30-years average (1984 -2013)
Amount of station (n) in 2014 is n=179, 2015 is n=136, and 30-years average is n=151

2.2.2 CaPA System precipitation estimates

CaPA System estimates are near real-time 6h accumulated precipitation at 10km resolution produced at 00, 06, 12, and 18 UTC. CaPA System estimates were produced by Environment and Climate Change Canada (ECCC). CaPA System estimates combined with AAF station observations of WS, Temp, and RH were used to calculate FWI System indexes.

CaPA System (Mahfouf *et al.,* 2007) merges station observations and radar quantitative precipitation estimates (QPEs) with a background precipitation grid field provided by the Global Environment Multiscale (GEM) model using an optimal interpolation method (Daley 1991) through a series of processes. First, differences between the station observations and 10km GEM forecasts are calculated at locations where precipitation is observed. GEM forecasts are used as

the background field due to the low station density in northern Canada (Figure 2.7a). Second, these differences are interpolated using simple kriging through experimental variogram analysis, and the resulting increment is added to the background field. A cubic root transformation is applied to both the observations and GEM forecasts before the interpolation to reduce the skewness. A back-transformation is applied to the background field. Finally, the radar QPEs are up-scaled to 10km and are assimilated into the analysis with consideration of horizontal correlations in the rain gauges and radar composite (Fortin *et al.,* 2015). Doppler range of 120km is used to calculate radar QPEs because there is greater confidence within this range. To avoid the errors associated with the model spin-up, each 6h analysis of CaPA uses GEM forecasts produced 6h before the start of the accumulation period. Since GEM forecasts are modelled outputs, CaPA System has the greatest confidence in areas close to ground observations and radar (Lespinas *et al.,* 2015).

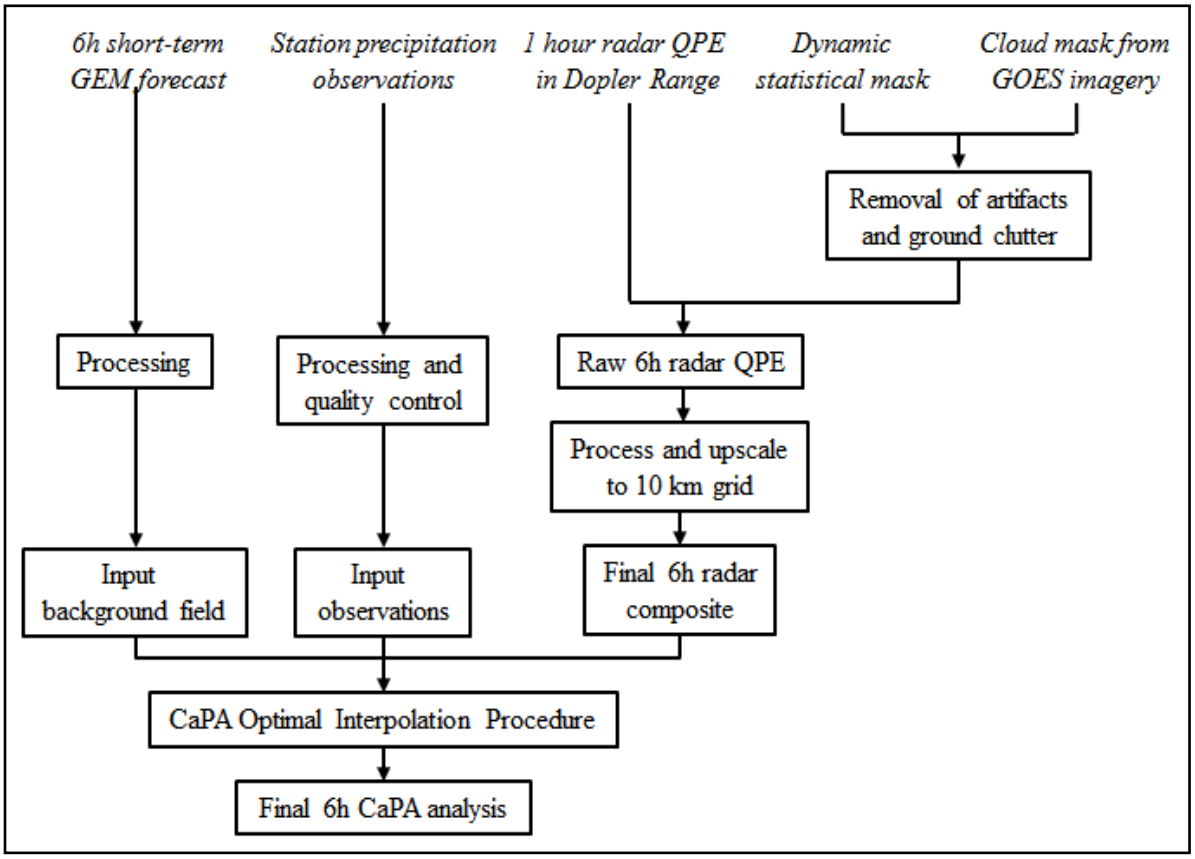


Figure 2.6 Data assimilation workflow of CaPA System
Information is adapted from Lespinas *et al.,* (2015) and Fortin *et al.,* (2015)

ECCC operates ~80 weather stations (1.58 stations / 10 000 km^2) in the study area; these stations are concentrated on the edge of the study area (Figure 2.7b). The ECCC station network and AAF station network (Figure 2.4) are unique station networks operated by different agencies. Our study area is covered by four radar stations using a 120km Doppler range (Figure 2.7b). The 6h CaPA System outputs were post-processed in two steps. First, the four 6hrs analysis (00, 06, 12, and 18 UTC) were added together to generate the 24h precipitation at 18 UTC (12:00 LST). Second, the 24h CaPA accumulations were extracted to the AAF stations located in the validation area (Figure 2.4) using the nearest neighbour search; these extractions were used as CaPA estimated precipitation at each weather station and were used for further analysis.

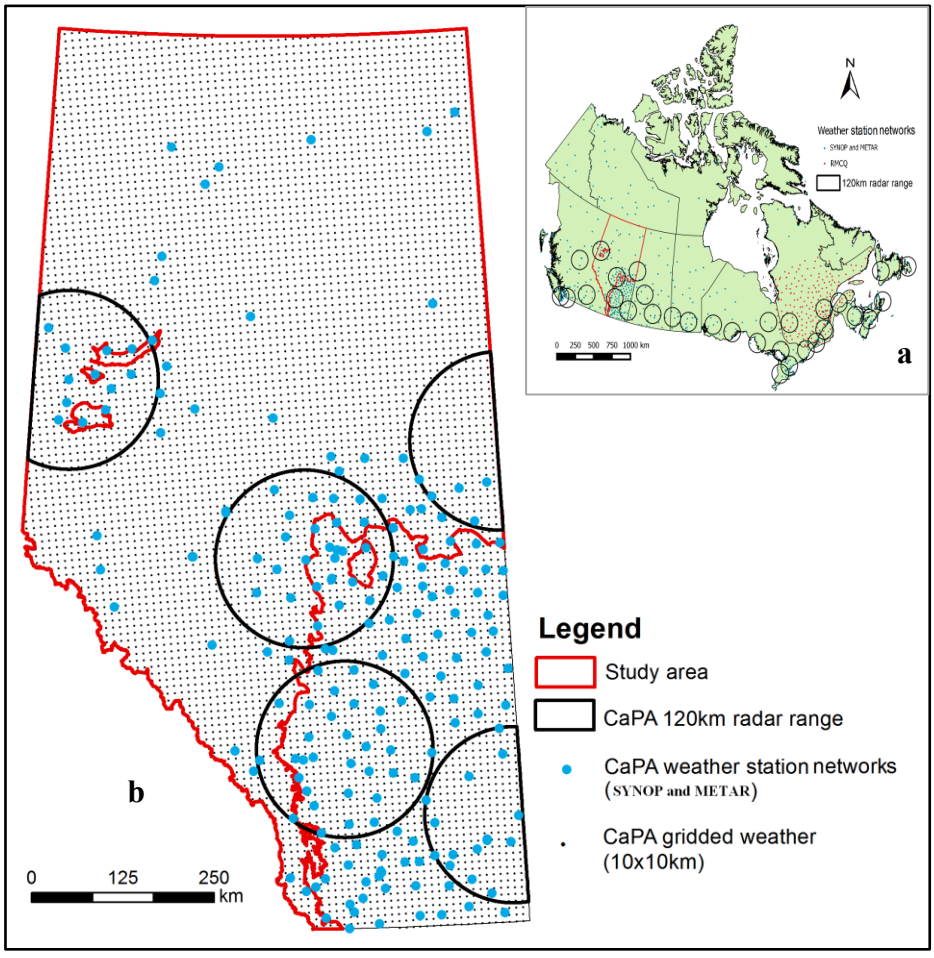


Figure 2.7 Stations networks and radar composite of CaPA System (a) in Canada and (b) in study areas

The station networks include manual and automatic synoptic station (SYNOP); aviation routine weather report (METAR); and two cooperative networks, the American Meteorological cooperative (SHEF) and the Réseau Météorologique Coopératif du Québec (RMCQ). There are 31 Canadian radars are integrated into CaPA System and 4 of them cover the study area. The radar composite is constructed using the Doppler range (120km).

2.3 Interpolation algorithms

Canadian fire management agencies interpolate weather observations from surrounding stations to estimate fire danger in areas between weather stations. In this study, I selected five types of interpolation algorithms to compare with the CaPA System in estimating daily precipitation in data scarce area (a review of fire weather interpolation is presented in Section 1.2.4). These five interpolation algorithms were classified into three groups: 1) non-geostatistical methods – inverse distance weighting, smoothed thin-plate spline, and non-smoothed thin-plate spline; 2) univariate geostatistical method – ordinary kriging; and 3) hybrid method –regression kriging (i.e., a regression of precipitation observations with CaPA System estimates and kriging on the regression residuals).

Smoothing the observed data is often implemented in interpolation algorithms to reduce the skewness of climate variables. There is a debate regarding which smoothing technique – square root (Hutchinson 1995), cubic root (Stidd 1973), or natural log – is the best for precipitation data. We implemented all three smoothing to find an appropriate smoothing technique for this study. Overall, we evaluated 18 algorithms of 6 methods (i.e., CaPA System and 5 interpolation methods as listed in Table 2.3). There was no transformation technique applied to inverse distance weighting because inverse distance weighting was used as a control group in the analysis as it is the interpolation method used by the Canadian Wildland Fire Information System (CWFIS) (Natural Resource Canada 2008).

For each interpolation algorithm, precipitation estimates were generated using a leave-one-out cross-validation (LOOCV) procedure (Stone 1974). The LOOCV removed one station from the AAF station network in the validation area (Figure 2.4) at a time. Then, precipitation of the removed station was estimated using the rest of the AAF stations in the study area for every interpolation algorithm. After the interpolation, the removed station was put back into the dataset and the next station was removed in sequence until all the stations in the validation area had been removed once. The LOOCV procedure was repeated every day in the study period to generate consecutive estimates of precipitation.

Table 2.3 Candidate algorithms used to estimate daily precipitation in this study

Candidates	Transformation of PCP[1]	Acronym
Inverse distance weighting	n/a	IDW
Canadian Precipitation Analysis System[2]	n/a	CaPA
Smoothed Thin-plate spline	n/a	tps_s
Smoothed Thin-plate spline	square root	tps_s_ sqrt
Smoothed Thin-plate spline	cubic root	tps_s_ cbrt
Smoothed Thin-plate spline	natural log	tps_s_ln
Non-smoothed Thin-plate spline	n/a	tps_ns
Non-smoothed Thin-plate spline	square root	tps_ns_sqrt
Non-smoothed Thin-plate spline	cubic root	tps_ns_cbrt
Non-smoothed Thin-plate spline	natural log	tps_ns_ln
Ordinary kriging	n/a	ok
Ordinary kriging	square root	ok_sqrt
Ordinary kriging	cubic root	ok_cbrt
Ordinary kriging	natural log	ok_ln
Regression kriging with CaPA	n/a	rk(capa)
Regression kriging with CaPA	square root	rk(capa)_ sqrt
Regression kriging with CaPA	cubic root	rk(capa)_cbrt
Regression kriging with CaPA	natural log	rk(capa)_ln

[1]Transformation of PCP was applied to PCP observations prior to the interpolation, and a back-transformation is applied to the interpolated PCP estimates. Among the three transformations, we will choose one technique that produces the best PCP estimates for the rest of the analysis.
[2]The CaPA system estimates were extracted from CaPA System grids using the nearest neighbour search algorithm.

The five interpolation methods (Table 2.3) share the same general estimation formula:

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \quad [1]$$

Where $\hat{Z}$ is the estimated precipitation value at the point of interest x_0, Z is the observed precipitation at the sampled point x_i, n is the sampled points used for the interpolation, and λ_i is the weight assigned to the sampled point (Webster and Oliver 2007). The difference among these interpolation algorithms lies in the way of calculating the weights (λ_i). Detailed descriptions of these algorithms are documented in Li and Heap (2011). Descriptions for each method were provided as follows and the R codes of each algorithm are presented in Appendix 2.

2.3.1 Non-geostatistical algorithms

Currently, fire management agencies use interpolation algorithms ranging from simple weighted averages to more sophisticated cubic splines to assess fire danger across landscapes; all of these interpolation algorithms fall into the category of non-geostatistical methods.

Inverse Distance Weighting

Inverse Distance Weighting (IDW) is a simple interpolation algorithm based on the function of inverse distance (Shepard 1968), where the weight, λ_i, decreases with the distance to the point of interest increases. The weight is expressed as $\lambda_i = 1/d_i^p / \sum_{i=1}^{n} 1/d_i^p$, where $d_i$is the distance between x_0 and x_i, and p is the power parameter that determines how much the weights diminish with distance. As p increases, more weight is placed on the nearby observations, and a p of zero means the interpolated value is the average of sampled observations. I used a p of 2 according to the current interpolation algorithm used by the CWFIS. Calculations of the precipitation estimates for the IDW algorithm were completed in the R programming language (R Development Core Team 2016) using the `gstat` package (Pebesma 2004).

Smoothed Thin-Plate Spline

Smoothed Thin-plate spline (TPS) uses the sampled observations to fit a spline surface with the constraint of smoothing (Wahba 1990). TPS can be viewed as a generalization of linear regression where the parametric model is replaced by a smooth non-parametric function (Haylock *et al.,* 2008). Smoothed TPS optimizes the amount of the data smoothing to minimize the predictive error measured by generalized cross validation (GCV). TPS surface is fitted by univariate spline function of location variables (latitude and longitude) (Hutchinson 1998a), or multivariate spline function with an additional variable of elevation (also called cubic spline) (Hutchinson 1998b). The inclusion of elevation in TPS surface is beneficial for climate fields (Tait *et al.,* 2006), but preliminary analysis of the data showed that precipitation in validation area was not well correlated with elevation, which is a result of relatively small variations in elevation (Figure 2.3). Therefore, I only used location variables for the spline function in this study; equations of the smoothed TPS are given in Wahba (1990) and Hutchinson (1993).

Smoothed TPS is robust to the interpolation of many climate variables across large domains (Price *et al.,* 2000). The assumption of smoothed TPS is that the interpolated surface is

perfectly smooth, which is often not true for many climate fields (e.g., summer precipitation). In addition, the smoothing effect may violate the sampled observations and produce unrealistically smooth results (Wahba 1990).

Non-smoothed Thin-Plate Spline

Non-smoothed TPS fits the spline surface without the constraint of smoothing, that is, the spline surface is fitted exacted with the sampled observations. Compared to the smoothed TPS, the exact spline may produce local artifacts of excessively high or low values, but non-smoothed TPS may be superior to smoothed TPS when extreme values are of critical concerns. Precipitation estimates for smoothed and non-smoothed TPS were calculated using the `fields` packages (Douglas *et al.,* 2016) in R.

2.3.2 Geostatistical algorithm - ordinary kriging

Ordinary Kriging (OK) is a commonly used geostatistical method that provides the best linear unbiased estimates (BLUE), where the estimated value for the point of interest is a weighted linear combination of sampled observations (i.e., the sum of weights is 1) (Matheron1963). OK is similar but more advanced than IDW, as the weight λ_i of OK is estimated by minimizing the variance of the prediction errors (Isaaks and Srivastava 1989). This is achieved by constructing a semi-variogram that models the difference between neighboring values. Details about the semi-variogram can be found in Li and Heap (2008). Semi-variograms can be fitted with simple models such as Nugget, Exponential, Spherical, Gaussian; and the nested sum of one or more simple models (Burrough and McDonnell 1998). In this study, I fitted the semi-variogram using a spherical model that is commonly used in interpolation studies. Additionally, a different semi-variogram was built for each point of interest.

Compared to non-geostatistical algorithms, the strength of ordinary kriging is its ability to model the spatial structure (variance) of the sampled observations. An assumption of ordinary kriging is data stationarity, that is, the mean of the interpolated variable is constant within the search window, which is often not true. This makes OK unsuitable for interpolation over large domains and often requires data transformation (Li and Heap 2014).

2.3.3. Hybrid algorithm – regression kriging

Regression kriging (RK) is a hybrid method that combines a regression of the sampled observations on secondary variables with the simple kriging of the regression residuals (Odeh *et al.,*1995). Equations of regression kriging can be found in Li and Heap (2008). In this study, I implemented the regression kriging algorithm in three steps. For each point of interest, I first fitted a linear regression of sampled precipitation observations and CaPA System estimates. Then, a simple kriging of the regression residual (predicted value minus observation) was performed by constructing semi-variogram with a spherical model. Finally, I summed the predicted value of linear regression and the kriged value to generate the final estimate. Precipitation estimates for ordinary kriging and regression kriging were calculated using the `gstat` package (Pebesma 2004) built in R.

Compared to ordinary kriging, regression kriging does not require the interpolated data to be stationary within the search window. Regression also can extend its algorithm to a broader range of regression techniques such as GLM by allowing separate interpretation of the two interpolated components (Hengl *et al.,*2007).

2.4 Canadian Forest Fire Weather Index (FWI) System

The FWI System (Van Wagner 1987) uses daily noon weather station records of ground temperature, relative humidity, 10m open wind speed, and 24h precipitation to produce six indicators of the relative fire danger in a local region (Figure 2.8). I used precipitation estimates (i.e., CaPA System and interpolation algorithms) and AAF station observations (i.e., Temp, WS, and RH) to calculate the indexes of FWI System. The FWI System calculated using all the AAF station observations were treated as true values.

FWI System moisture codes (i.e., FFMC, DMC, and DC, see Figure 2.8) are modelled using a dynamic bookkeeping system that tracks the drying and wetting of fuel layers. The three fuel moisture codes are strongly influenced by precipitation; with the occurrence of rain events increasing the fuel moisture content of the fuel layers, thus decreasing the values of the fuel moisture codes (Van Wagner 1987). However, these fuel moisture codes are influenced by precipitation of different magnitudes and a minimal precipitation amount (effective rain) of

0.5mm, 1.5mm, and 2.8mm is required to reduce the code values of FFMC, DMC, and DC respectively (Van Wagner 1987). The effectiveness of precipitation at reducing the values of the fuel moisture codes is negatively correlated to the amount of precipitation and is positively correlated with the code values prior to the rain, all of which is determined by the absorption rate of fuel layers (Lawson and Armitage 2008). The FFMC, DMC, and DC have a rain capacity of 0.62mm, 15mm, and 100 mm respectively, that is, depending on the previous day's code values, precipitation over that capacity becomes run-off and has no effect on reducing the code values. Fuel layers also start to dry in the absence of rain, leading to the increase of fuel moisture code values. The drying time lags (the time it requires for fuel to lost ~2/3 of its free moisture toward equilibrium moisture content) are 2/3 day, 12 days, and 52 days respectively for the FFMC, DMC, DC under standard condition (Temp 20°C, RH 45%, WS 13km/h) (Van Wagner 1987).

Initial analysis showed that errors of the precipitation estimates were accumulated (error propagation) in DC and DMC due to their long drying time lag. I did not adjust this error propagation and used the estimated previous day's fuel moisture codes to calculate today's fuel moisture codes because this is the approach used by the CWFIS for fire danger mapping (personal communication with Bo Lu). The calculation of FWI System was conducted in R using the `"cffdrs"` package (Wang *et al.,* 2017) which follows the equations and calculations described in Van Wagner (1987).

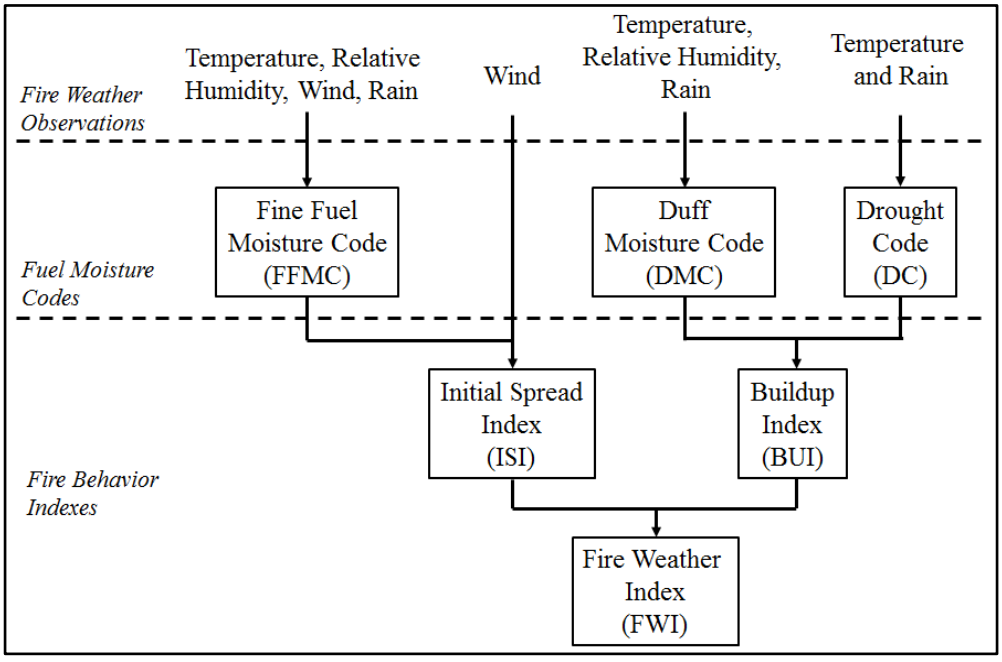


Figure 2.8 Structure of the Canadian Forest Fire Weather Index System (Van Wagner 1987)

2.5 Model evaluation

2.5.1 Skill scores

I used continuous and categorical scores to evaluate the candidates in estimating precipitation of any amounts and the effective precipitation for the FWI System fuel moisture codes. The formulas of skill scores used in this study are summarized in Table 2.4.

Table 2.4 Definition of skill scores used in this study.

Skill Score	Equations[1]
Mean Absolute Error (MAE)	$\mathrm{MAE} = \frac{1}{n \times k}\sum_{i=1}^{n}\sum_{j=1}^{k}\left\lvert P_{i,j} - O_{i,j}\right\rvert$
Mean error (Bias)	$\mathrm{Bias} = \frac{1}{n \times k}\sum_{i=1}^{n}\sum_{j=1}^{k}\left(P_{i,j} - O_{i,j}\right)$
Equitable Threat Score (ETS)	$\mathrm{ETS} = \frac{a-R(a)}{a+b+c-R(a)}$; $R(a) = \frac{(a+b)(a+c)}{a+b+c+d}$
Frequency Bias Index (FBI)	$\mathrm{FBI} = \frac{a+b}{a+c} - 1$

[1]Explanation of the variables: $P_{i,j}$ is the predicted PCP value at weather station i on day j, $O_{i,j}$ is the observed PCP value at the same weather station i on the same day j. n is the number of validation stations and k is the number of days in the study. *a, b, c, d* are defined using the contingency table as described by Wilks (2011). Specifically, *a* is the fraction of hits (e.g., PCP>0.5mm in the estimates and PCP>0.5mm in the observations); *b* is the fraction of false alarms (e.g., PCP>0.5mm in the estimates and PCP<0.5mm in observations); *c* is the fraction of misses (e.g., PCP<0.5mm in the estimates and PCP>0.5mm in the observations; *d* is the fraction of correct rejections (e.g., PCP<0.5mm in the estimates and PCP<0.5 in the observations).

Mean Absolute Error (MAE) is a natural, unambiguous measure of average error (Willmott and Matsuura, 2005). MAE describes the errors in the same unit as the precipitation (mm) with the best score of zero. I did not choose the commonly used Root Mean Square Error (RMSE) because its sensitivity to outliers (here, extremely large PCP events) could inflate the skill, especially when tested dataset does not follow a normal distribution (Willmott and Matsuura 2005). Chai (2014) also suggested MAE is a better choice to evaluate precipitation errors than RMSE because RMSE tends to overemphasize outliers.

Mean error (Bias) is a continuous measure of the direction of the average errors between estimates and observations (Stanski HR 1989)**.** The best score of Bias is zero; a positive/negative score means over-prediction /under-prediction.

Equitable threat score (ETS) is a categorical measure of the fraction of estimates/observations that are correctly predicted, adjusting by R(a), which is the expected amount of estimates that would be correctly classified if the estimate were completely random (Wilks 2011). R(a) makes ETS insensitive to the climatology of the event (e.g., dryness and wetness) and it is often used in the verification of rainfall in NWP models (Wilks 2011). ETS ranges from -1/3 to 1, with a perfect score of 1 and values less than 0 indicating no skill.

Frequency bias index (FBI) is a categorical measure of the ratio between the frequency of estimates and the frequency of observations (Wilks 2011). FBI indicates whether the forecast system has a tendency to under-prediction (FBI < 0) or over-prediction (FBI > 0) events. FBI does not measure how well the estimate corresponds to the observations. Note that I have subtracted 1 from the conventional definition of FBI, so the perfect score of FBI is 0.

The categorical indexes ETS and FBI were calculated by grouping the minimal precipitation thresholds for fuel moisture codes, i.e., 0.5 mm, 1.5 mm, and 2.8 mm for FFMC, DMC, and DC respectively. Therefore, the following bins were used for the calculation: 0< precipitation ≤0.5 mm; 0.6 mm< precipitation ≤ 1.5 mm; 1.6 mm < precipitation ≤ 2.8 mm; precipitation>2.9 mm.

2.5.2 Statistical tests

In this study, I also performed a hypothesis test to identify statistical significance in the precipitation and FWI System estimates with algorithms listed in Table 2.3. Because the precipitation data are often spatially and temporally correlated (Livezey and Chen 1983), I used a resampling hypothesis test procedure (Hamill 1999) that simultaneously accounts for these correlations by building a distribution through repeated random sampling of the tested data, which satisfies the independence requirement in the tests (Wilks 1997).

The resampling hypothesis test includes two major steps. First, a resampling one-way ANOVA is applied to precipitation / FWI System estimates from all the algorithms to determine if there were statistical differences between these estimates. Then, a resampling posthoc test was applied to identify which pair of algorithms produced significantly different estimates. The resampling posthoc test combines the resampling paired t-tests (Hamill 1999) and Holm-Bonferroni p-values adjustment (Holm 1979) to accommodate for multiple comparisons. Details of the resampling posthoc test were summarized in Appendix 3. I provided a description of the resampling procedure as follows.

Stationary Block Bootstrapping

The resampling is achieved through stationary block bootstrapping (SBB) (Politis and Romano 1994), which was used by Lespinas *et al.,* (2015) to evaluate the performance of the CaPA system in Canada. SBB is based on resampling blocks of observations with random length to form a stationary pseudo-time series. The length of each block follows a geometric distribution that is defined as $\Pr(X = m) = (1-p)^{m-1} \cdot p$; where p is a fixed number in [0, 1]; search algorithm to a time-space data series through the following processes: 1) an initial date was randomly selected for each bootstrap sample; 2) a block contains b days of observations from the initial date was enforced, where b follows the geometric distribution with a mean block length of m. A m of 3days was chosen for precipitation due to its averaged temporal autocorrelation (Figure 2.9); 3) to preserve the spatial correlation, observations from all the locations within the block were collected; 4) the resampling of the time-space blocks is repeated with replacement until the reconstructed sample contains as many observations as the original sample; and 5) steps 1 to 4 were repeated 1000 times to generate 1000 resamples. The mean block length for precipitation and FWI System were chosen according to their average temporal autocorrelation as plotted in Figure 2.9.

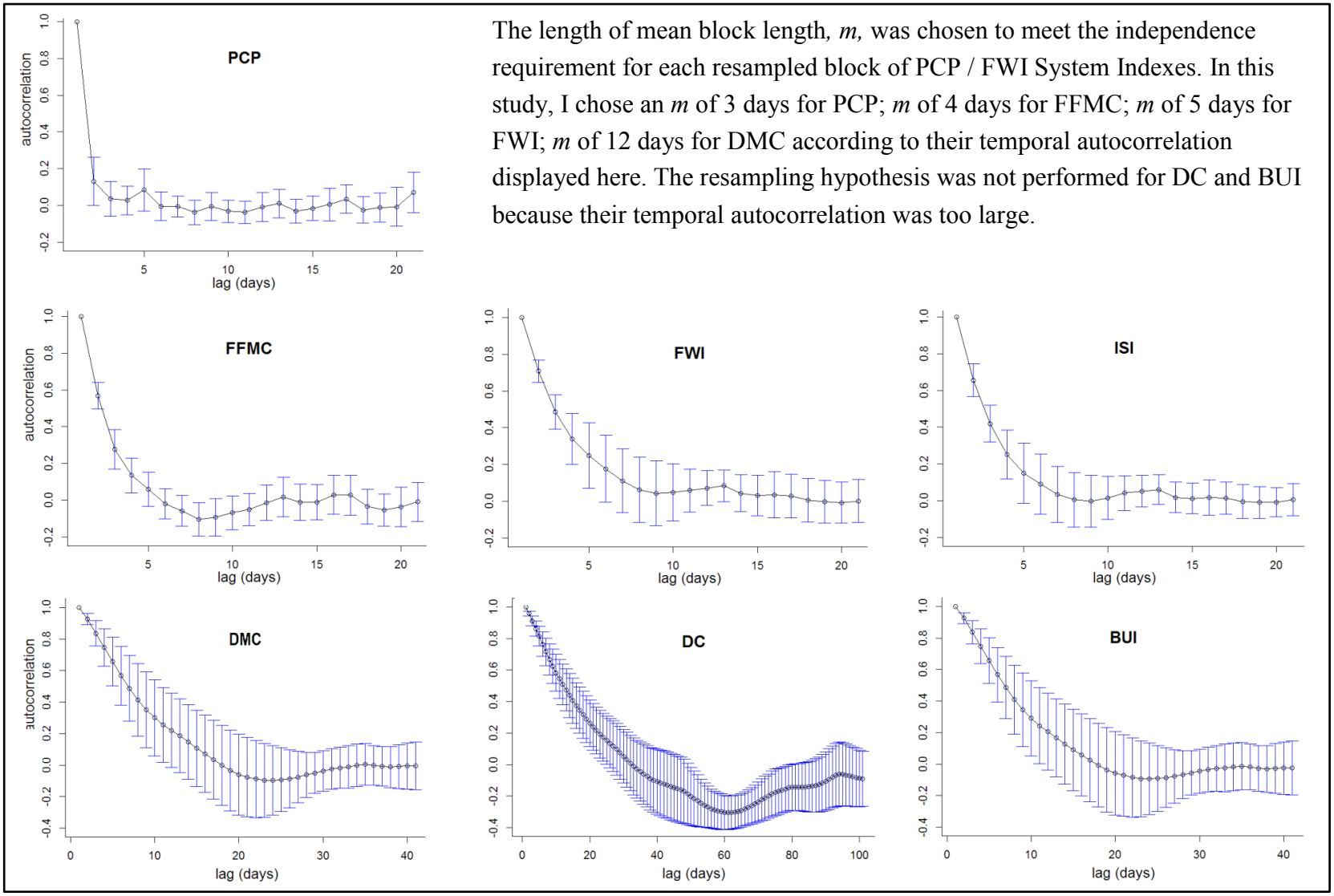


Figure 2.9 Temporal autocorrelation of daily precipitation and FWI System
Autocorrelation are averaged for all stations (n=81) in 2015.Error bars are calculated as the averaged temporal autocorrelation plus or minus the standard deviation of the temporal autocorrelation for all stations

2.6 Sensitivity analysis of weather station density

Because the performance of the interpolation methods is highly related to weather station density (Daly 2006; Hofstra *et al.,* 2010), I performed a sensitivity analysis of weather station density. This analysis involved random sampling of fire weather stations at densities defined in Table 2.5. The sensitivity analysis applied all the aforementioned analysis steps and its process can be summarized as the following: 1) 10 % of the fire weather stations in the study area were randomly selected (includes stations in the validation area and buffer area); 2) LOOCV procedure was performed to generate precipitation estimates in the selected stations using interpolation method listed in Table 2.3 (estimates of CaPA System were extracted from the closest grid cell); 3) FWI System were calculated using the precipitation estimates with observed Temp, RH, and WS; 4) precipitation and FWI System estimates were evaluated using the skill scores listed in Table 2.5 (with emphasis on MAE); 5) step 1 to step 4 were repeated 100 times and the mean of the resulted skill scores were calculated; 6) steps 1 to 5 were repeated with random sampling 25%, 50%, 75% and 90% of available stations.

Table 2.5 Weather station density used in the sensitive analysis

Year	Scenario	Area (km^2)	No. of stations	No. of stations per 10 000km^2
2014	10% selected	505 964	18	0.36
	25% selected		45	0.89
	50% selected		90	1.78
	75% selected		135	2.67
	90% selected		162	3.20
2015	10% selected	505 964	14	0.28
	25% selected		34	0.67
	50% selected		68	1.34
	75% selected		102	2.02
	90% selected		123	2.43

Chapter 3 Results

3.1 Overall model performance

3.1.1 Continuous scores and the choice of smoothing

In 2014 and 2015, the kriging related algorithms ok_cbrt, rk(capa)_cbrt, and ok_sqrt, produced precipitation estimates with the smallest MAE values (Figure 3.1a, 3.1c). Compared to IDW, these algorithms had ~23.5%, 21.3%, and 21.7% lower MAE, respectively (Table 3.1). CaPA System had lower mid-tiered performance (13th of 18) and was 7% lower in MAE than that of IDW (Table 3.1). Smoothed TPS related algorithms had upper mid-tiered performance (7th -9th of 18) when performed with smoothed precipitation and were 17.5% to 20.2 % lower in MAE than that of IDW (Table 3.1). Non-smoothed TPS was always higher in MAE than that of smoothed TPS, but non-smoothed TPS had lower MAE than that of CaPA System (Figure 3.1). Ordinary Kriging (OK) was the worst overall algorithm and was the only algorithm that performed worse than IDW (2.2% greater MAE comparing to IDW).

The Bias ranking showed that IDW, regression kriging, and smoothed TPS had the smallest bias when performed with non-smoothed precipitation (Figure 3.1b, 3.1d). Ordinary kriging with smoothed precipitation had the largest negative bias (underestimation), while ordinary kriging with non-smoothed precipitation had the largest positive bias (overestimation). Additionally, CaPA System produced a relatively large negative bias amongst all the algorithms (Figure 3.1b and Figure 3.1b). Although IDW produced a very low bias, it does not necessarily indicate that IDW is a good algorithm. Because IDW also produced the largest MAE (Figure 3.1); the low bias of IDW may result from the cancelling-out of overestimates and underestimates.

Smoothing precipitation observations using square root, cubic root, and natural log greatly reduced the prediction errors (i.e., MAE values, Figure 3.1a and 3.1c). However, interpolation algorithms performed with smoothed precipitation also tended to underestimate precipitation values, which is problematic for FWI System that are sensitive to seasonal precipitation (e.g., DMC and DC, as explained in Section 3.3). Among the three smoothing techniques, I selected the square root smoothing technique for the rest of the analysis and reduced the candidate algorithms from 18 to 10 as listed in Table 3.2. I chose square root

because all three smoothing techniques performed similarly in reducing the prediction errors, but square root smoothing technique resulted in a slightly smaller bias compared with the other two smoothing techniques.

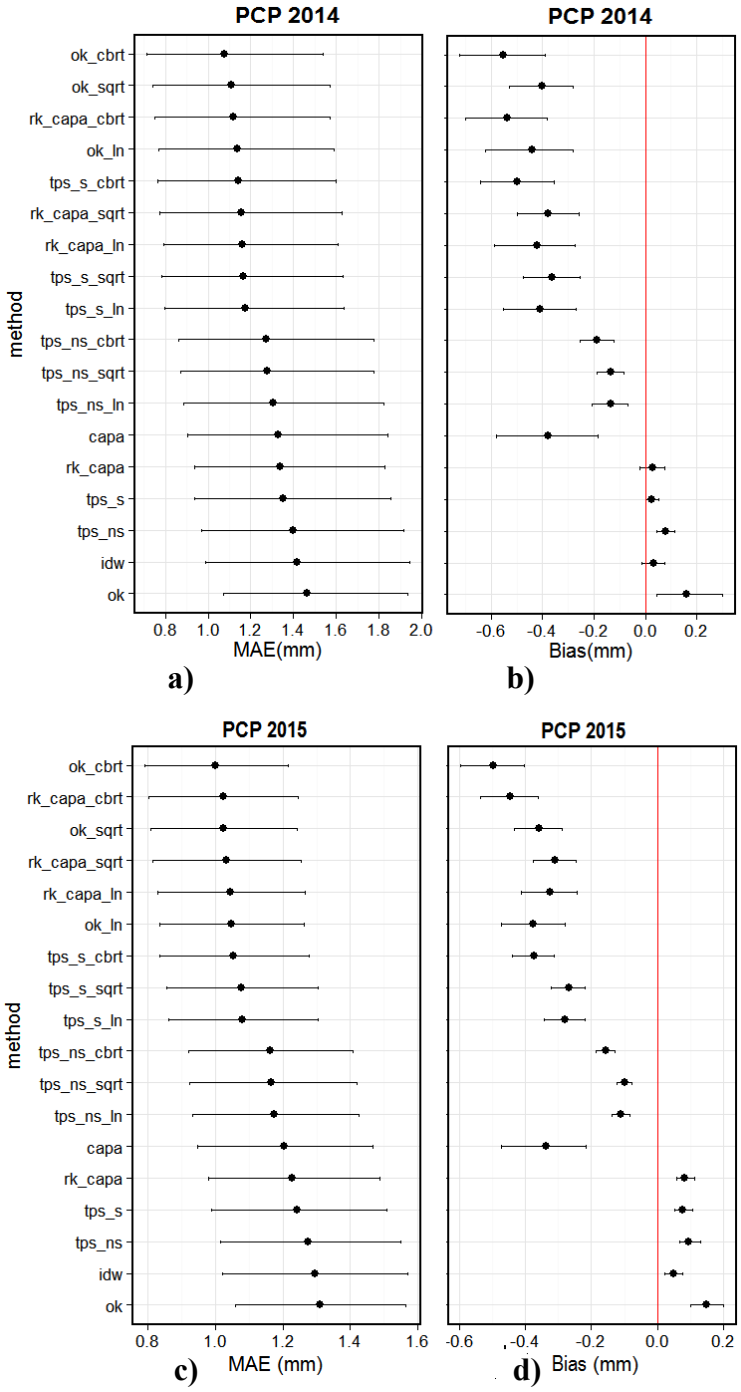


Figure 3.1 95% Confidence interval of candidate algorithms in validation area using resampled MAE and Bias

Algorithms are ranked from the best (top) to the worst (bottom) according to the mean of 1000 replicates (dark dots).

Table 3.1 MAE and Bias of candidate algorithms in validation area

MAE and Bias were calculated by averaging all the stations.

Candidate Algorithms	Transformation of PCP	MAE(mm) 2014	MAE(mm) 2015	MAE(mm) Percent change vs IDW (%)[1]	Bias(mm) 2014	Bias(mm) 2015
IDW	n/a	1.42	1.30	n/a	0.04	**0.05**
CaPA	n/a	1.33	1.20	(7.0)	(0.38)	(0.34)
Smoothed TPS	n/a	1.34	1.24	(5.5)	**0.03**	0.08
	square root	1.14	1.08	(18.3)	(0.36)	(0.27)
	cubic root	1.12	1.05	(20.2)	(0.49)	(0.37)
	natural log	1.15	1.08	(17.5)	(0.4)	(0.28)
Non-smoothed TPS	n/a	1.37	1.27	(2.9)	0.08	0.11
	square root	1.26	1.17	(10.6)	(0.13)	(0.10)
	cubic root	1.25	1.16	(11.4)	(0.18)	(0.16)
	natural log	1.28	1.17	(10.0)	(0.13)	(0.11)
Ordinary kriging	n/a	1.47	1.31	2.2	0.16	0.15
	square root	1.11	1.02	(21.7)	(0.4)	(0.36)
	cubic root	**1.08**	**1.0**	**(23.5)**	(0.55)	(0.5)
	natural log	1.14	1.05	(19.5)	(0.44)	(0.38)
Regression kriging with CaPA	n/a	1.33	1.23	(5.9)	**0.03**	0.08
	square root	1.12	1.03	(19.6)	(0.38)	(0.31)
	cubic root	1.12	1.02	(21.3)	(0.52)	(0.45)
	natural log	1.16	1.04	(19.2)	(0.42)	(0.32)

[1] Percentage change (%) equals to the difference between the MAE values of candidate algorithms and IDW divided by the MAE value of IDW; the values were averaged for 2014 and 2015.

Bold: the best performing algorithm. Underline: the worst performing algorithm.

Table 3.2 Resampling paired t-test of candidate algorithms in validation areas using MAE

	MAE (mm) averaged over all stations in 2014 and 2015 (test statistics)									
Algorithm[1]	ok	idw	tps _ns	tps _s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk (capa)_s	ok _s
2014	*1.47*	*1.42*	*1.37*	*1.34*	*1.33*	*1.33*	*1.25*	*1.14*	*1.16*	***1.12***
2015	*1.31*	*1.30*	*1.27*	*1.24*	*1.23*	*1.20*	*1.16*	*1.08*	*1.03*	***1.02***
ok	–							▲	▲	▲
idw		–				▲		●▲	▲	▲
tps_ns			–				▲	▲	▲	▲
tps _s				–			▲	▲	●▲	▲
rk(capa)					–			▲	▲	▲
CaPA						–		▲	▲	▲
tps_ns_s							–	▲	▲	▲
tps_s_s								–		
rk(capa)_s									–	
ok_s										–

The better / worse performance than the algorithms on its left / right side.

Bold: the best performing algorithm. Underline: the worst performing algorithm.

●: comparisons achieved a significance level of α=0.05 in year 2014

▲: comparisons achieved a significance level of α=0.05 in year 2015

The output of the resampling post-hoc tests are presented in **Appendix 3**, Table A3-1.

Statistical tests (Table 3.2) showed that in 2015, the top three algorithms - ok_sqrt, rk(capa)_sqrt, and tps_s_sqrt produced significantly better estimates of precipitation than other algorithms regarding MAE, but these three algorithms were not significantly different from each other. CaPA had significantly lower MAE than IDW in 2015. Algorithms with square root smoothing of precipitation had significantly lower MAE compared to non-smoothed algorithms. In 2014, there were fewer comparisons that showed statistical significance (Table 3.2); this may be caused by the smaller dataset in 2014 (49 days versus 123 days).

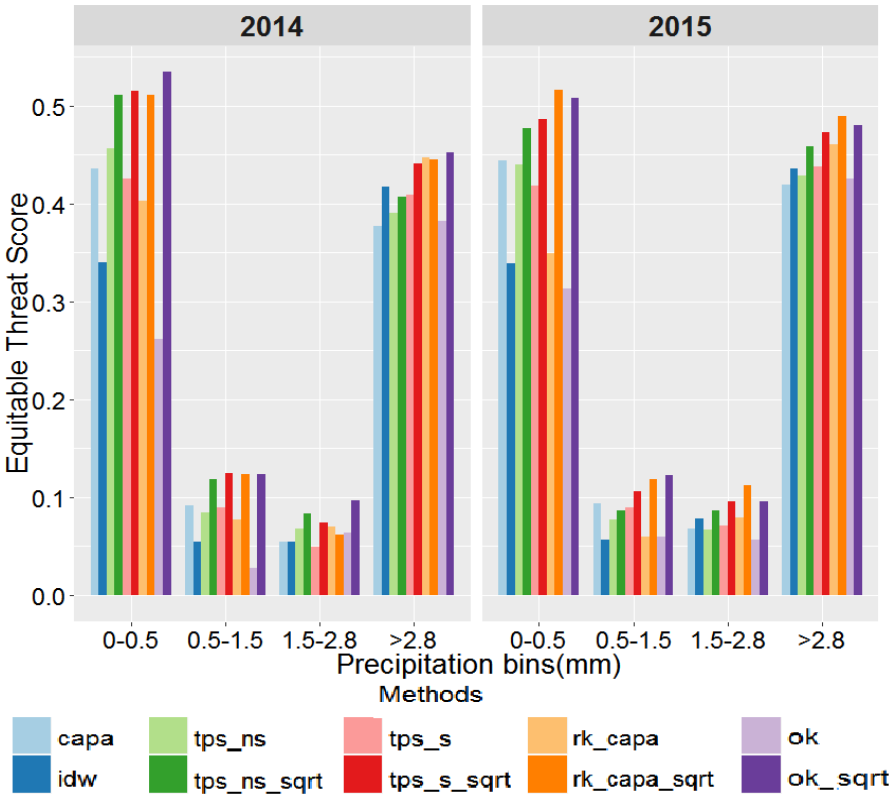
Considering both MAE and Bias, regression kriging with CaPA estimates as covariate emerged as the best overall algorithm in estimating precipitation. Examples of 24 hour precipitation estimated using the 10 candidate methods are provided in **Appendix 4,** all of which showed the advantages regression kriging. I also evaluated these algorithms using the unit-less categorical scores by grouping precipitation observations based on the minimal precipitation values that are required to reduce the fuel moisture codes values.

3.1.2 Categorical scores

In 2014 and 2015, every algorithm was better at predicting precipitation <0.5 mm and precipitation >2.8mm than precipitation between 0.5-2.8 mm (ETS, Figure 3.2a). The poor skill in predicting precipitation between 0.5-2.8 mm was a result of large positive bias (overestimation) as indicated by FBI (Figure 3.2b). Additionally, every algorithm underestimated precipitation <0.5 mm; CaPA and interpolation algorithms with smoothing underestimated precipitation >2.8 mm, while algorithms without smoothing overestimated precipitation > 2.8 mm (Figure 3.2b). In particular, CaPA had the second worst ETS and FBI scores in predicting precipitation >2.8 mm (Figure 3.2), which is due to Capac's tendency of underestimating precipitation in that range (Table 3.3b).This underestimation of precipitation will result in large errors for DMC and DC as explained in Section 3.4.

In 2015, rk(capa)_sqrt had the best ETS scores for each precipitation category and was significantly better than OK, IDW, CaPA, and rk(capa) (Table 3.3a). There was no significant difference between the top few algorithms regarding ETS. Rk(capa) was significantly less biased than other algorithms in predicting precipitation >2.8 mm, while tps_ns_sqrt was significantly less biased in predicting precipitation between 0.5 – 2.8 mm (Table 3.3b).

a)



b)

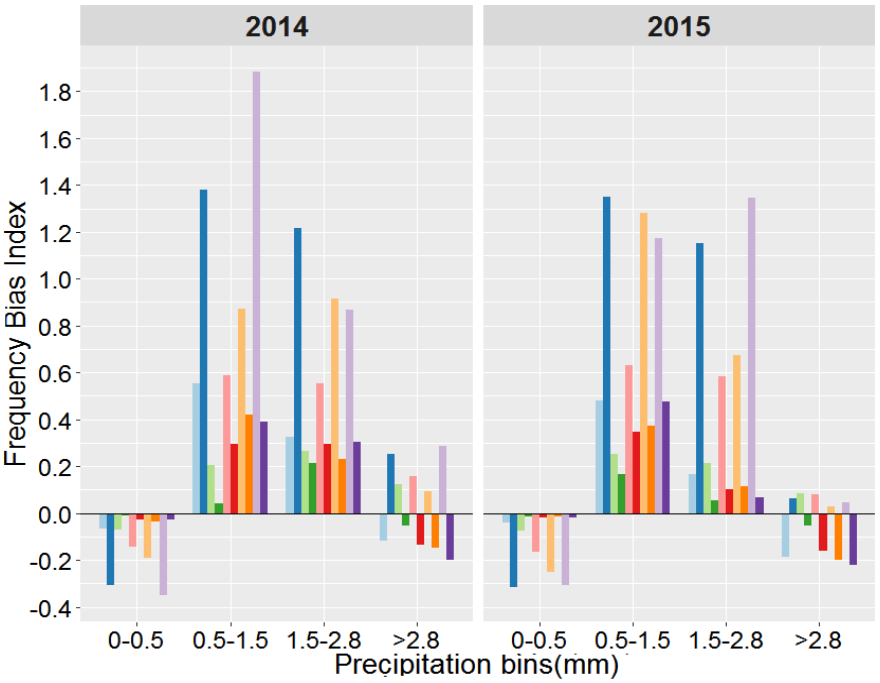


Figure 3.2 Bar plots of (a) ETS and (b) FBI of precipitation estimates in 2014 and 2015
PCP ≤0.5 mm has no effect to fuel moisture codes; 0.5 mm <PCP ≤1.5 mm only affect FFMC; 1.5 mm <PCP ≤ 2.8 mm only affect FFMC and DMC; PCP >2.8 mm affect FFMC, DMC, and DC. ETS and FBI values were calculated by pooling all the station estimates together. The perfect ETS score is 1 and value < 0 indicates no skill; the perfect FBI score is 0 and a positive/negative value indicates a tendency of overestimation/underestimation.

Table 3.3 Resampling paired t-test of candidate algorithms in validation areas using (a)ETS and(b)FBI

a) **ETS averaged over all stations in 2015 (test statistics)**

PCP (mm)	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
0.5 – 1.5	*0.06*	*0.06*	*0.08*	*0.09*	*0.06*	*0.09*	*0.09*	*0.11*	***0.12***	***0.12***
1.5 – 2.8	*0.06*	*0.08*	*0.07*	*0.07*	*0.08*	*0.07*	*0.09*	*0.10*	***0.11***	*0.10*
> 2.8	*0.42*	*0.44*	*0.43*	*0.44*	*0.46*	*0.42*	*0.46*	*0.47*	***0.49***	*0.48*

Algorithm	ok	idw	tps _ns	tps _s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
ok	–								●▲■	●
idw		–						●	●	●
tps_ns			–						■	●■
tps_s				–						
rk (capa)					–			●	●	●
CaPA						–			■	
tps_ns_s							–			
tps_s_s								–		
rk(capa)_s									–	
ok_s										–

Bold: the best performing algorithm. Underline: the worst performing algorithm. A higher ETS score indicates greater skill.

● comparisons achieved a significance level of α=0.05 for PCP range from 0-0.5 mm

▲ comparisons achieved a significance level of α=0.05 for PCP range from 1.5 to 2.8 mm

■ comparisons achieved a significance level of α=0.05 for PCP > 2.8 mm

The output of the resampling post-hoc tests is presented in **Appendix 3**, Table A3-2.

b) **FBI averaged over all stations in 2015 (test statistics)**

PCP (mm)	ok	idw	tps _ns	tps _s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
0.5 – 1.5	*1.17*	*1.35*	*0.25*	*0.63*	*1.28*	*0.48*	***0.17***	*0.35*	*0.37*	*0.47*
1.5 – 2.8	*1.34*	*1.15*	*0.21*	*0.59*	*0.67*	*0.17*	***0.05***	*0.10*	*0.11*	*0.07*
> 2.8	*0.05*	*0.06*	*0.08*	*0.08*	***0.03***	*(0.20)*	*(0.06)*	*(0.16)*	*(0.19)*	*(0.20)*

Algorithm	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
ok	–		▲			●▲	▲	▲	▲■	▲■
idw		–	●▲	▲	▲	●▲	●▲	●▲	●▲■	●▲■
tps_ns			–	●	▲	■	▲■	■	■	■
tps_s				–		■	●▲	▲	▲■	▲■
rk (capa)					–	▲■	●▲	●▲■	●▲■	▲■
CaPA						–	●	●	●	
tps_ns_s							–	●	●■	■
tps_s_s								–		
rk(capa)_s									–	
ok_s										–

Similar to Table 3.3a, but showing the results for FBI. A FBI score of zero indicates no bias, while a positive/negative score indicates over/underestimate.

The output of the resampling post-hoc tests are presented in **Appendix 3**, Table A3-3.

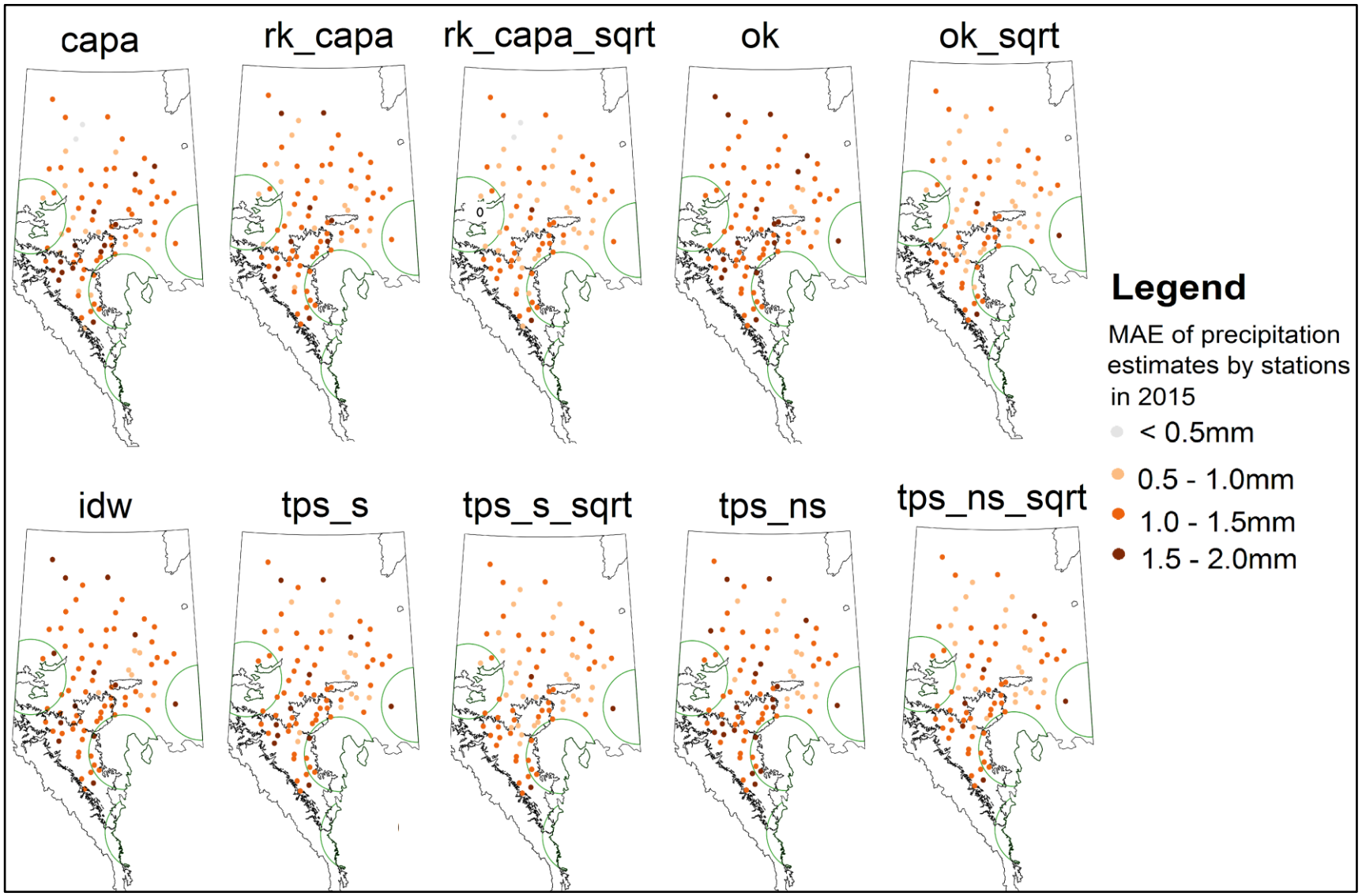


Figure 3.3 MAE of precipitation estimates by stations in 2015

MAE was averaged over the 2015 study period (123days). Stations with lighter colour performed better than stations with darker colour.

3.2 Spatial variations

By calculating MAE values on a station-by-station basis, we found that algorithms performed differently in radar and non-radar areas, as well as in Boreal and Foothills areas (Figure 3.3).

3.2.1 Radar versus non-radar

In radar areas, CaPA related methods were the top performing algorithms (Figure 3.4a) and rk(capa)_sqrt performed the best; these algorithms were the only algorithms that had reduced MAE values in radar areas (Table 3.4). Statistical tests in radar areas were tempered by the fewer stations (n=8), but rk(capa)_sqrt still showed significantly lower MAE than 5 out of 9 other algorithms (Table 3.4). CaPA System, however, was only significantly better than IDW and rk(capa) was significantly better than tps_ns (Table 3.4). In non-radar areas (Figure 3.4b and Table 3.4), all the algorithms had very similar MAE values compared to that of overall validation area; this may due to 90% of AAF weather stations are located in non-radar covered areas.

ETS (FBI) scores in radar and non-radar areas (Figure 3.5) followed a similar pattern of ETS (FBI) scores in validation area (Figure3.2). In radar areas, rk(capa)_sqrt had the best ETS scores in each precipitation category and had significantly greater ETS scores than that of OK and IDW when predicting precipitation between 0.5-1.5 mm (Table 3.5). In radar areas, CaPA had improved ETS scores in every precipitation category compared to non-radar areas (Figure 3.5a), but the improvement was not significant (Table 3.5). In radar areas, CaPA had the least improvement of ETS scores in predicting precipitation >2.8mm; because CaPA had a larger tendency to underestimate precipitation >2.8mm compared to non-radar areas (e.g., FBI changed from -0.17 to -0.33, Figure 3.5b). Both CaPA and rk(capa)_sqrt were significantly more biased than other algorithms in predicting precipitation >2.8mm. In radar areas, rk(capa) was the least biased algorithm in predicting precipitation>2.8mm and showed significant improvement over CaPA, rk(capa)_sqrt, and ok_sqrt (Table 3.6).

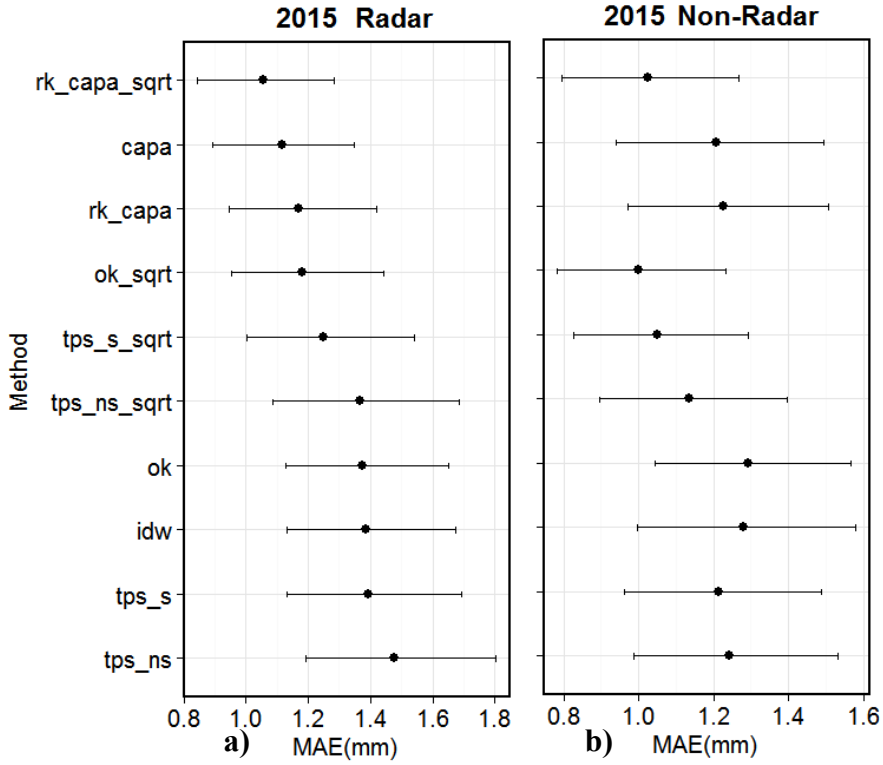


Figure 3.4 95% Confidence interval of candidate algorithms in (a) radar and (b) non-radar using resampled MAE.

Algorithms are ranked from the best (top) to the worst (bottom) according to the mean of 1000 replicates (dark dots) in radar covered areas.

Table 3.4 Resampling paired t-test of candidate algorithms in radar and non-radar areas using MAE

	MAE (mm) averaged over stations in radar / non-radar for 2015 (test statistics)									
Algorithm	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
Non-radar	*1.30*	*1.29*	*1.25*	*1.22*	*1.23*	*1.21*	*1.14*	*1.06*	*1.03*	***1.01***
Radar	*1.37*	*1.39*	*1.47*	*1.39*	*1.17*	*1.12*	*1.36*	*1.24*	***1.01***	*1.18*
Percent change (%)	*(5.1)*	*(7.2)*	*(15.0)*	*(12.2)*	*5.1*	*8.0*	*(16.2)*	*(14.5)*	*2.0*	*(14.4)*

Algorithm	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
ok	–						▲	▲	●▲	●▲
idw		–				●		▲	●▲	●▲
tps_ns			–		●		●▲	▲	●▲	●▲
tps_s				–			▲	●▲	●▲	●▲
rk (capa)					–			▲	●▲	▲
CaPA						–		▲	▲	▲
tps_ns_s							–	▲	▲	▲
tps_s_s								–		▲
rk(capa)_s									–	
ok_s										

Bold: the best performing algorithm in radar/non-radar regions. Underline: the worst performing algorithm in radar/non-radar regions.

▲: comparisons achieved a significance level of α=0.05 in non-radar region

●: comparisons achieved a significance level of α=0.05 in radar region.

The output of the resampling post-hoc tests are presented in **Appendix 3**-Table A3-4

a)

b)

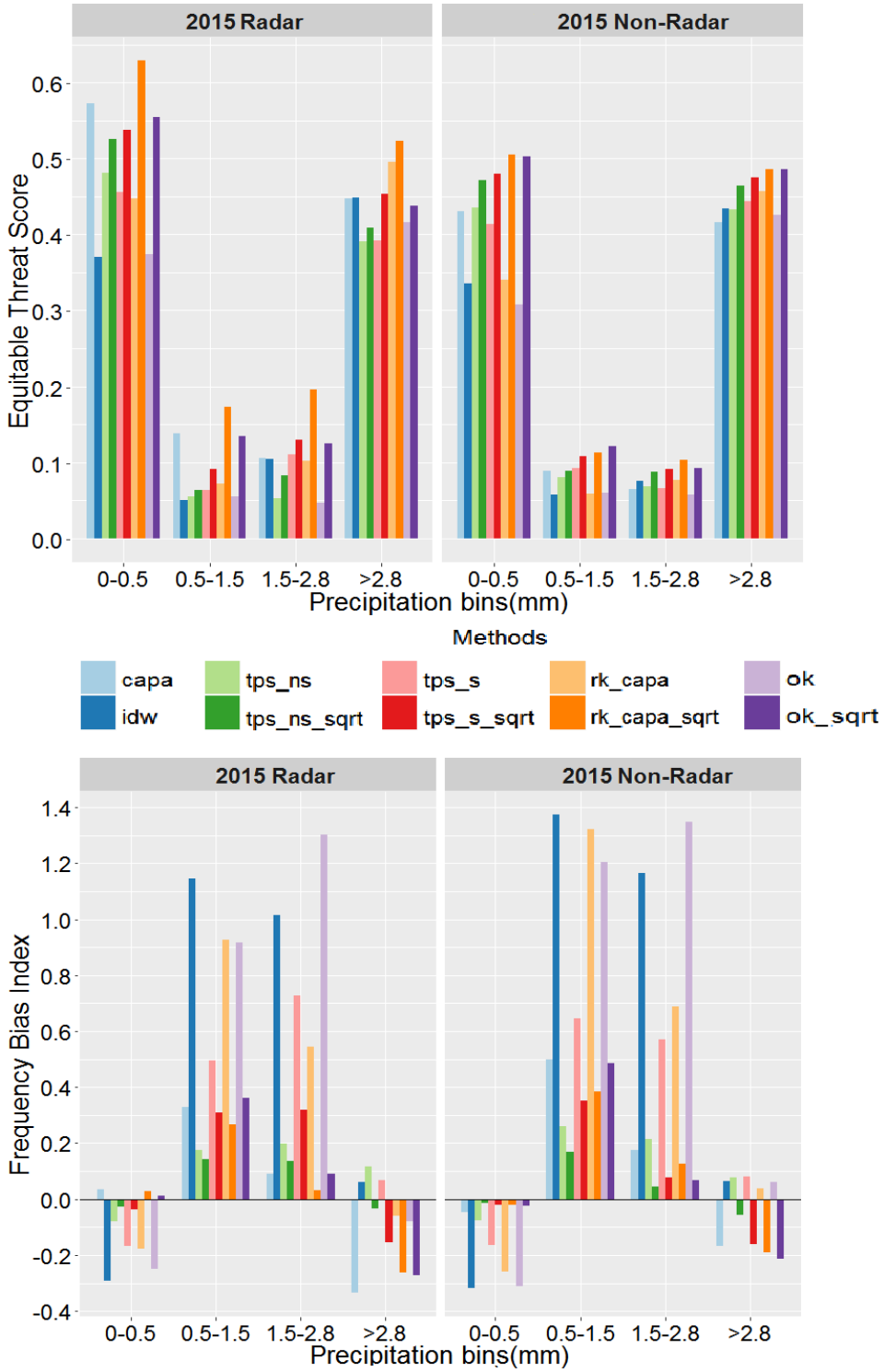


Figure 3.5 Plots of (a) ETS and (b) FBI in radar and non-radar areas

ETS and FBI values were calculated by polling all station estimates within the radar / non-radar areas in 2015.

Table 3.5 Resampling paired t-test of candidate algorithms in radar and non-radar areas using ETS

		ETS averaged over stations in non-radar / radar in 2015 (test statistics)									
Regions	PCP (mm)	ok	idw	tps_ns	tps_s	rk (capa)	CaPA	tps_ns_s	tps_s_s	rk(capa)_s	ok_s
Non-radar	0.5 – 1.5	*0.06*	*0.06*	*0.08*	*0.09*	*0.60*	*0.09*	*0.09*	*0.10*	*0.11*	***0.12***
	1.5 – 2.8	*0.06*	*0.08*	*0.07*	*0.07*	*0.08*	*0.06*	*0.09*	*0.09*	***0.10***	***0.10***
	> 2.8	*0.43*	*0.43*	*0.43*	*0.43*	*0.46*	*0.42*	*0.46*	*0.48*	***0.49***	***0.49***
Radar	0.5 – 1.5	*0.05*	*0.05*	*0.06*	*0.06*	*0.07*	*0.14*	*0.06*	*0.09*	***0.17***	*0.13*
	1.5 – 2.8	*0.05*	*0.10*	*0.05*	*0.11*	*0.10*	*0.11*	*0.08*	*0.13*	***0.20***	*0.12*
	> 2.8	*0.42*	*0.45*	*0.39*	*0.39*	*0.51*	*0.45*	*0.42*	*0.45*	***0.52***	*0.43*

	Algorithm	ok	idw	tps_ns	tps_s	rk (capa)	CaPA	tps_ns_s	tps_s_s	rk(capa)_s	ok_s
Radar	idw		–						●	●■	●
	tps_ns			–							●■
	tps_s				–						
	rk (capa)					–			●		●
	CaPA						–			■	
	tps_ns_s							–			
	tps_s_s								–		
	rk(capa)_s									–	
	ok_s										–
Radar	ok	–								●	
	idw		–							●	
	tps_ns			–							
	tps_s				–				■		
	rk (capa)					–					
	CaPA						–				
	tps_ns_s							–			
	tps_s_s								–		
	rk(capa)_s									–	
	ok_s										–

Bold: the best performing algorithm. Underline the worst performing algorithm. A higher ETS score indicates greater skill.

The outputs of the resampling posthoc tests are presented in **Appendix 3**, Table A3-5 and Table A3-6.

● Comparisons achieved a significance level of α=0.05 for PCP range from 0-0.5 mm

▲ comparisons achieved a significance level of α=0.05 for PCP range from 1.5 to 2.8 mm

■ comparisons achieved a significance level of α=0.05 for PCP > 2.8 mm

Table 3.6 Resampling paired t-test of candidate algorithms in radar and non-radar areas using FBI

		FBI averaged over stations in non-radar / radar in 2015 (test statistics)									
Regions	PCP (mm)	ok	idw	tps_ns	tps_s	rk (capa)	CaPA	tps_ns_s	tps_s_s	rk(capa)_s	ok_s
Non-radar	*0.5 – 1.5*	*1.20*	*1.3*	*0.26*	*0.65*	*1.32*	*0.50*	***0.17***	*0.35*	*0.38*	*0.49*
	1.5 – 2.8	*1.35*	*1.17*	*0.21*	*0.57*	*0.69*	*0.17*	***0.04***	*0.08*	*0.12*	*0.07*
	> 2.8	*0.06*	*0.06*	*0.08*	*0.08*	***0.04***	*(0.17)*	*(0.06)*	*(0.16)*	*(0.19)*	*(0.21)*
Radar	*0.5 – 1.5*	*0.92*	*1.37*	*0.18*	*0.49*	*0.93*	*0.33*	***0.14***	*0.31*	*0.27*	*0.36*
	1.5 – 2.8	*1.30*	*1.02*	*0.20*	*0.73*	*0.55*	*0.09*	*0.14*	*0.32*	***0.03***	*0.09*
	> 2.8	*(0.08)*	*0.06*	*0.12*	*0.07*	*(0.06)*	*(0.33)*	***(0.03)***	*(0.16)*	*(0.26)*	*(0.27)*

	Algorithm	ok	idw	tps_ns	tps_s	rk (capa)	CaPA	tps_ns_s	tps_s_s	rk(capa)_s	ok_s
Non-radar	ok	–		▲			▲	▲	▲	●▲	▲
	idw		–	▲	▲	▲	●▲	●▲	●▲■	●▲	●▲■
	tps_ns			–		▲	●▲■	▲■	■	■	■
	tps_s				–		■	●■	▲■	▲■	▲■
	rk (capa)					–	▲	●▲	●▲■	●▲	●▲■
	CaPA							●			
	tps_ns_s								●		■
	tps_s_s										
	rk(capa)_s										
	ok_s										
Radar	ok	–									
	idw		–					●	●	●▲■	▲■
	tps_ns			–			■	■		■	■
	tps_s				–		■			■	
	rk (capa)					–	■			■	■
	CaPA						–	■			
	tps_ns_s							–			■
	tps_s_s								–		
	rk(capa)_s									–	
	ok_s										–

Bold: the best performing algorithm. Underline: the worst performing algorithm. A FBI score of zero indicates no bias, while a positive / negative score indicates over / underestimate.
The outputs of the resampling post-hoc tests are presented in **Appendix 3**, Table A3-7 and Table A3-8.
● comparisons achieved a significance level of α=0.05 for PCP range from 0-0.5 mm
▲ comparisons achieved a significance level of α=0.05 for PCP range from 1.5 to 2.8 mm
■ comparisons achieved a significance level of α=0.05 for PCP > 2.8 mm.

3.2.2 Foothills versus Boreal

The top performing algorithms in Foothills and Boreal were identical (Figure 3.6), but all the algorithms produced larger MAE in Foothills compared to Boreal (Figure 3.6). This suggests that all the algorithms performed worse in Foothills, which mat due to the climatology between the two areas is very different (e.g., Foothills is much wetter than the Boreal, see Table 2.2).

Analysis based on the 2015 dataset showed that every algorithm was more accurate (i.e., higher ETS score) and was less biased (i.e., smaller absolute FBI score) in Foothills than in Boreal, especially in predicting precipitation <2.8 mm (Figure 3.7). This result is contrary to the result indicated by MAE and can be explained by 1) Foothills was wetter than Boreal (a 31% larger daily precipitation, see Table 2.2); and 2) the nature of MAE and ETS (FBI) is different. MAE is a continuous measure of average error and is dependent on the climatology of the event (e.g., magnitude of precipitation event). Therefore, MAE may be greater in wet areas than dry areas. However, ETS (FBI) is dependent on a unit-less contingency table and is insensitive to the climatology of the event. It is, therefore, more appropriate to use ETS (FBI) to compare the precipitation methods in Foothills and Boreal. Additionally, the greater skill of algorithms in Foothills may be due to a higher weather station density in this region, that is, Foothills had a 146% higher station density compared to Boreal regions (Table 2.1). The outputs of statistical tests were presented in **Appendix 3**.

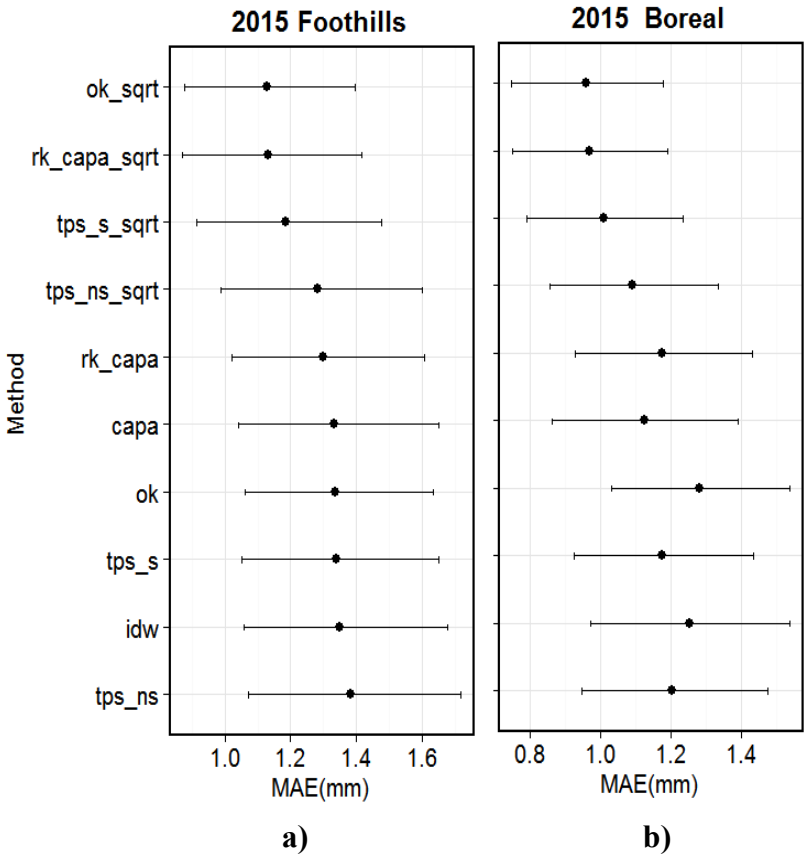


Figure 3.6 95% Confidence interval of candidate algorithms in (a) Foothills and (b) Boreal using resampled MAE

Algorithms are ranked from the best (top) to the worst (bottom) according to the mean of 1000 replicates (dark dots) in Foothills areas.

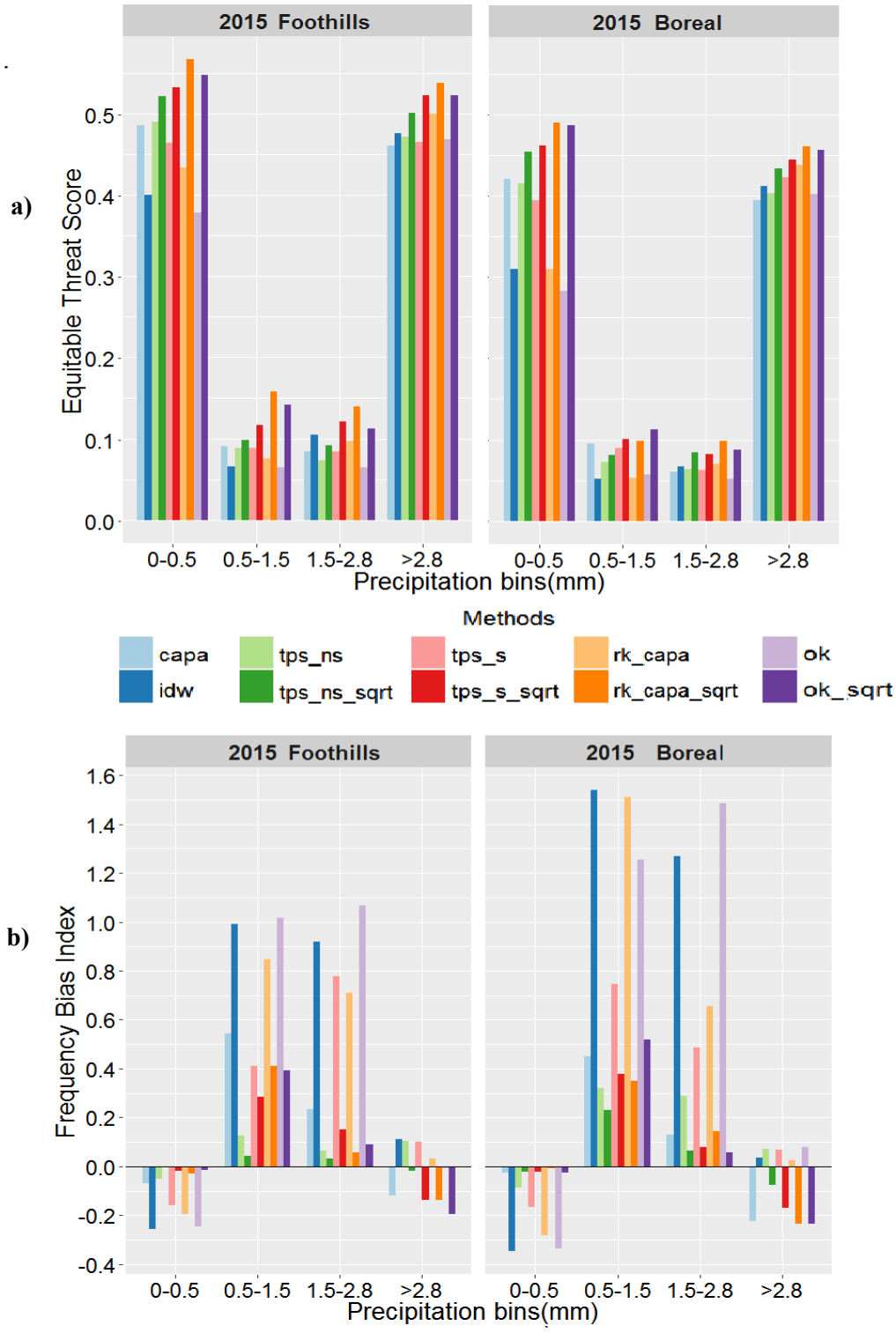


Figure 3.7 Plots of (a) ETS and (b) FBI in Foothills and Boreal
ETS and FBI values were calculated by pooling all station estimates in 2015.

3.3 FWI System

The rk(capa)_sqrt had the overall lowest MAE values for FFMC, ISI, and FWI estimates that were calculated using estimated precipitation and observed Temp, RH, and WS (Table 3.7). Compared to IDW, rk(capa)_sqrt had 16%, 29%, 22% lower MAE for FFMC, ISI, and FWI estimates, respectively. CaPA System showed mid-tired MAE values for FFMC, ISI, and FWI estimates, while Ordinary kriging showed the largest MAE values for the same three FWI System (Table 3.7). Statistical tests showed that FFMC estimated with rk(capa)_sqrt had significantly lower MAE than the other algorithms and FWI estimated with rk(capa)_sqrt had significantly lower MAE than OK, IDW, CaPA, and smoothed TPS (Table 3.8).

Rk(capa) had the overall lowest MAE values for DMC, DC, and BUI estimates, while CaPA had the largest MAE values for the same three FWI System (Table 3.7). The poor performance of CaPA in estimating DMC, DC, and BUI may be due to its large tendency to underestimate precipitation>2.8mm as showed in Figure 3.2b.. Statistical tests of DMC showed that only a single comparison archived significance, i.e., rk(capa) had significantly lower MAE than CaPA (Table 3.8). No statistical tests were performed on DC as explained in Section 2.5.2 (Figure 2.9).

In non-radar areas, the performance of algorithms were similar to that for the overall study areas, i.e., rk(capa)_sqrt / ordinary kriging was the best / worst performing method for FFMC, ISI, and ISI, while rk(capa) / CaPA was the best / worst performing method for DMC, DC, and BUI (Table 3.7). In radar areas, the performance of methods was not as consistent across the FWI System Indexes. For example, the best / worst performing method for FWI changed to rk(capa) / ok_sqrt, while the best performing method for DC changed to tps_s (Table 3.7). Although CaPA related methods had smaller MAE values for FFMC, ISI, and FWI under radar, CaPA had increased MAE values for DMC, DC, and BUI estimates. This may be due to CaPA had increased tendency to underestimate precipitation>2.8mm in radar areas compared to non-radar areas (Figure 3.5). In Foothills and Boreal, the performance of all the algorithms were consistent with the overall study areas, but all the algorithms produced lower MAE values of FWI System estimates in Foothills (Table 3.7), which is in agreement with the ETS values of precipitation estimates presented in Figure 3.7a.

Table 3.7 MAE of FWI System calculated using estimated PCP and observed RH, Temp, WS

Calculation of fuel moisture codes using estimated previous day values. MAE values are calculated by pooling all the stations in validation areas, radar areas, non-radar areas, Boreal regions, and Foothills regions.

Region	Method	PCP(mm)	FFMC	ISI	FWI	DMC	DC	BUI
Validation area (n = 81)	capa	1.20	4.23	0.69	2.74	9.68	66.52	13.49
	idw	1.30	4.58	0.82	2.80	7.98	49.78	10.74
	tps_ns	1.27	4.20	0.71	2.52	7.79	49.95	10.76
	tps_ns_s	1.16	3.92	0.64	2.41	7.82	49.27	10.80
	tps_s	1.24	4.29	0.76	2.64	7.98	46.99	10.86
	tps_s_s	1.08	3.79	0.63	2.47	8.68	53.75	11.89
	rk(capa)	1.23	4.06	0.75	2.61	**7.75**	**45.54**	**10.65**
	rk(capa)_s	1.03	**3.57**	**0.58**	**2.36**	8.48	55.22	11.76
	ok	1.31	4.65	0.90	3.05	8.46	48.31	11.38
	ok_s	**1.02**	3.62	0.59	2.46	9.22	60.24	12.76
Non-radar (n = 73)	capa	1.21	4.23	0.71	2.76	9.53	64.27	13.26
	idw	1.29	4.55	0.84	2.87	8.20	49.43	10.99
	tps_ns	1.25	4.13	0.71	2.55	7.96	48.81	10.92
	tps_ns_s	1.14	3.84	0.65	2.43	7.96	47.64	10.88
	tps_s	1.22	4.23	0.77	2.69	8.18	47.06	11.10
	tps_s_s	1.06	3.72	0.63	2.49	8.78	53.41	11.97
	rk(capa)	1.23	4.08	0.77	2.67	**7.91**	**44.11**	**10.82**
	rk(capa)_s	1.03	**3.54**	**0.59**	**2.38**	8.53	53.39	11.73
	ok	1.30	4.61	0.92	3.13	8.71	47.34	11.67
	ok_s	**1.01**	3.55	0.60	2.44	9.22	58.91	12.69
Radar (n = 8)	capa	1.12	4.18	0.51	2.58	11.07	87.10	15.59
	idw	1.39	4.89	0.67	2.17	5.99	53.03	8.46
	tps_ns	1.47	4.88	0.67	2.25	6.28	60.34	9.31
	tps_ns_s	1.36	4.58	0.61	2.28	7.02	64.09	10.55
	tps_s	1.39	4.82	0.65	2.16	6.10	**46.40**	8.64
	tps_s_s	1.24	4.37	0.57	2.26	7.70	56.87	11.19
	rk(capa)	1.17	3.91	0.57	**2.01**	**5.75**	58.52	**8.45**
	rk(capa)_s	**1.05**	**3.71**	**0.49**	2.19	8.07	71.93	11.97
	ok	1.37	4.99	0.73	2.35	6.12	57.19	8.71
	ok_s	1.18	4.34	0.61	2.62	9.21	72.43	13.43
Boreal (n = 53)	capa	1.13	4.27	0.73	3.00	10.65	68.21	14.90
	idw	1.26	4.74	0.90	3.15	9.20	50.82	12.42
	tps_ns	1.21	4.32	0.77	2.84	9.07	51.92	12.45
	tps_ns_s	1.10	4.02	0.70	2.71	9.04	49.50	12.38
	tps_s	1.18	4.32	0.80	2.89	9.08	49.76	12.41
	tps_s_s	1.01	3.84	0.67	2.71	9.92	51.98	13.45
	rk(capa)	1.18	4.14	0.81	2.91	**8.96**	**48.20**	**12.34**
	rk(capa)_s	0.97	**3.61**	**0.63**	**2.62**	9.82	57.51	13.54
	ok	1.29	4.84	1.00	3.49	9.97	49.08	13.44
	ok_s	**0.96**	3.64	0.65	2.69	10.68	59.32	14.61
Foothills (n = 28)	capa	1.35	4.15	0.60	2.21	7.85	63.32	10.81
	idw	1.36	4.28	0.68	2.14	5.69	47.82	7.56
	tps_ns	1.39	3.99	0.59	1.92	5.45	46.20	7.56
	tps_ns_s	1.29	3.73	0.53	1.88	5.45	48.83	7.79
	tps_s	1.35	4.24	0.68	2.16	5.97	41.75	7.92
	tps_s_s	1.20	3.69	0.54	2.00	6.32	57.11	8.95
	rk(capa)	1.31	3.91	0.63	2.02	**5.36**	**40.49**	**7.43**
	rk(capa)_s	**1.14**	**3.49**	**0.51**	**1.85**	5.96	50.90	8.38
	ok	1.35	4.29	0.73	2.26	5.59	46.87	7.47
	ok_s	**1.14**	3.58	0.52	2.02	6.46	61.99	9.28

Bold: algorithm had the lowest MAE for the indexes of FWI system.
Underline: algorithm had the highest MAE for the indexes of FWI indexes.
n: number of stations.

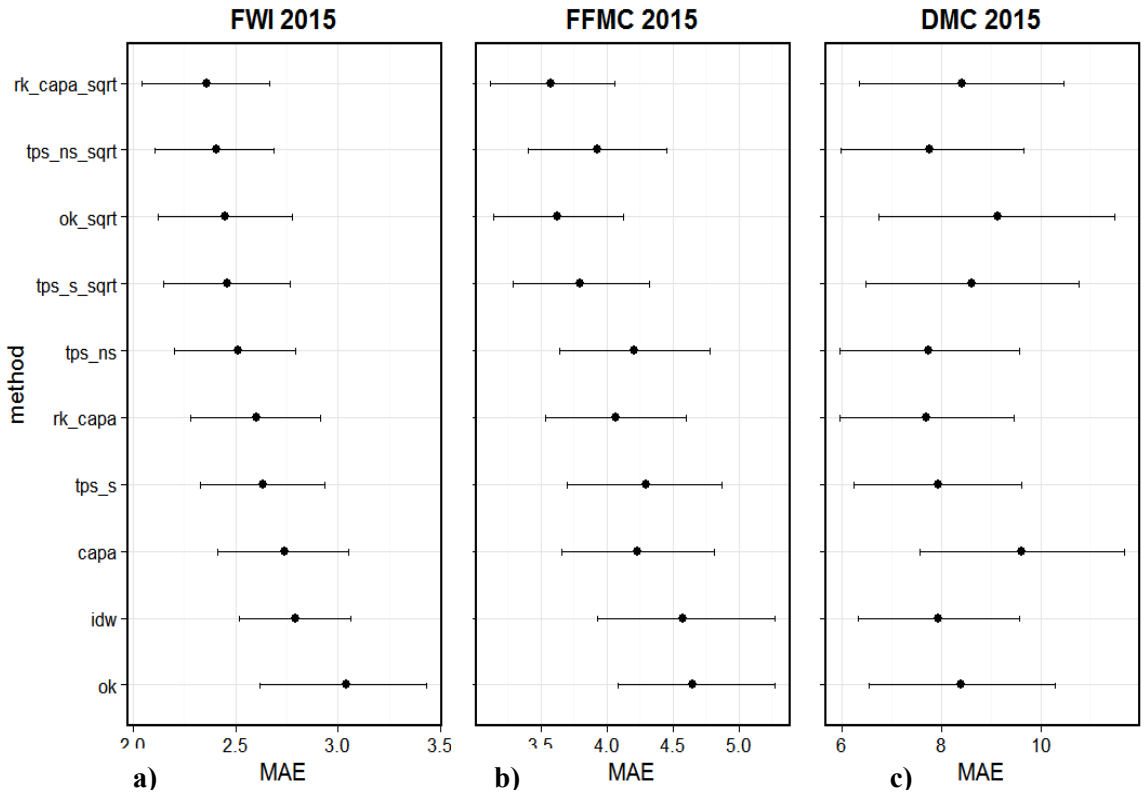


Figure 3.8 95% Confidence interval of resampled MAE of precipitation estimates for (a) FWI, (b) FFMC, and (c) DMC in 2015

Algorithms are ranked from the best (top) to the worst (bottom) according to the mean of 1000 replicates (dark dots) of FWI index.

Table 3.8 Resampling paired t-test of FWI, FFMC, and DMC estimates in validation areas using MAE

	MAE in overall study region for 2015 (test statistics)									
Algorithm[1]	ok	idw	CaPA	tps_s	rk (capa)	tps _ns	tps _s_s	ok_s	tps _ns_s	rk(capa) _s
FWI	*3.05*	*2.80*	*2.74*	*2.64*	*2.61*	*2.52*	*2.47*	*2.46*	*2.41*	***2.36***
FFMC	*4.65*	*4.58*	*4.23*	*4.29*	*4.06*	*4.20*	*3.79*	*3.62*	*3.92*	***3.57***
DMC	*8.46*	*7.98*	*9.68*	*7.98*	***7.75***	*7.79*	*8.68*	*9.22*	*7.82*	*8.48*

Method	ok	idw	CaPA	tps_s	rk (capa)	tps _ns	tps _s_s	ok_s	tps _ns_s	rk(capa) _s
ok	–				▲		▲	▲	▲	●▲
idw		–					●▲	▲	●▲	●▲
CaPA			–		■			●▲	●	●▲
tps_s				–			▲	▲	▲	●▲
rk (capa)					–			▲		▲
tps_s_s						–	▲	▲	▲	▲
tps_ns							–	▲	▲	▲
ok_s								–	▲	
tps_ns_s									–	▲
rk(capa)_s										–

●:significance level of α=0.05 for FWI; ▲:sig. level of α=0.05 for FFMC; ■: sig. level of α=0.05 for DMC

Bold: best performing algorithm. Underline: worst performing algorithm.

The output of the resampling post-hoc tests are presented in **Appendix 3**-Table A3-14.

The estimates of FWI System showed two groups of responses to precipitation estimates (Table 3.7). The performance of algorithms in estimating FFMC, ISI, and FWI was consistent with their performance in estimating precipitation, while the performance of these algorithms in estimating DMC, DC, and BUI was not as consistent. This is because long drying time lags in DMC and DC (12 and 52 days) resulted in an error propagation in precipitation estimates, making DMC and DC estimates be largely influenced by underestimation of true precipitation values.

The error propagation of DMC and DC can be seen in Figure 3.9a, which shows that underestimation of large rainfall events can make a dramatic change to all FWI System Indexes, but DMC and DC need more time to recovery compared to FFMC and FWI. For example, CaPA System greatly underestimated a large precipitation event (~40mm) at the White Court Auto station at the end of May (Figure 3.9a). This underestimation of actual precipitation values made FFMC, FWI, DMC, and DC estimated with CaPA System being all higher than observed values. As FFMC and FWI estimated with CaPA System recovered quickly after a few dry days, DMC and DC estimated with CaPA System required more time to recover and remained higher than observations during the rest of the study period (Figure 3.9a). This error propagation of DMC and DC can be corrected by using observed previous day's fuel moisture values for the FWI System calculation (Figure 3.9b).

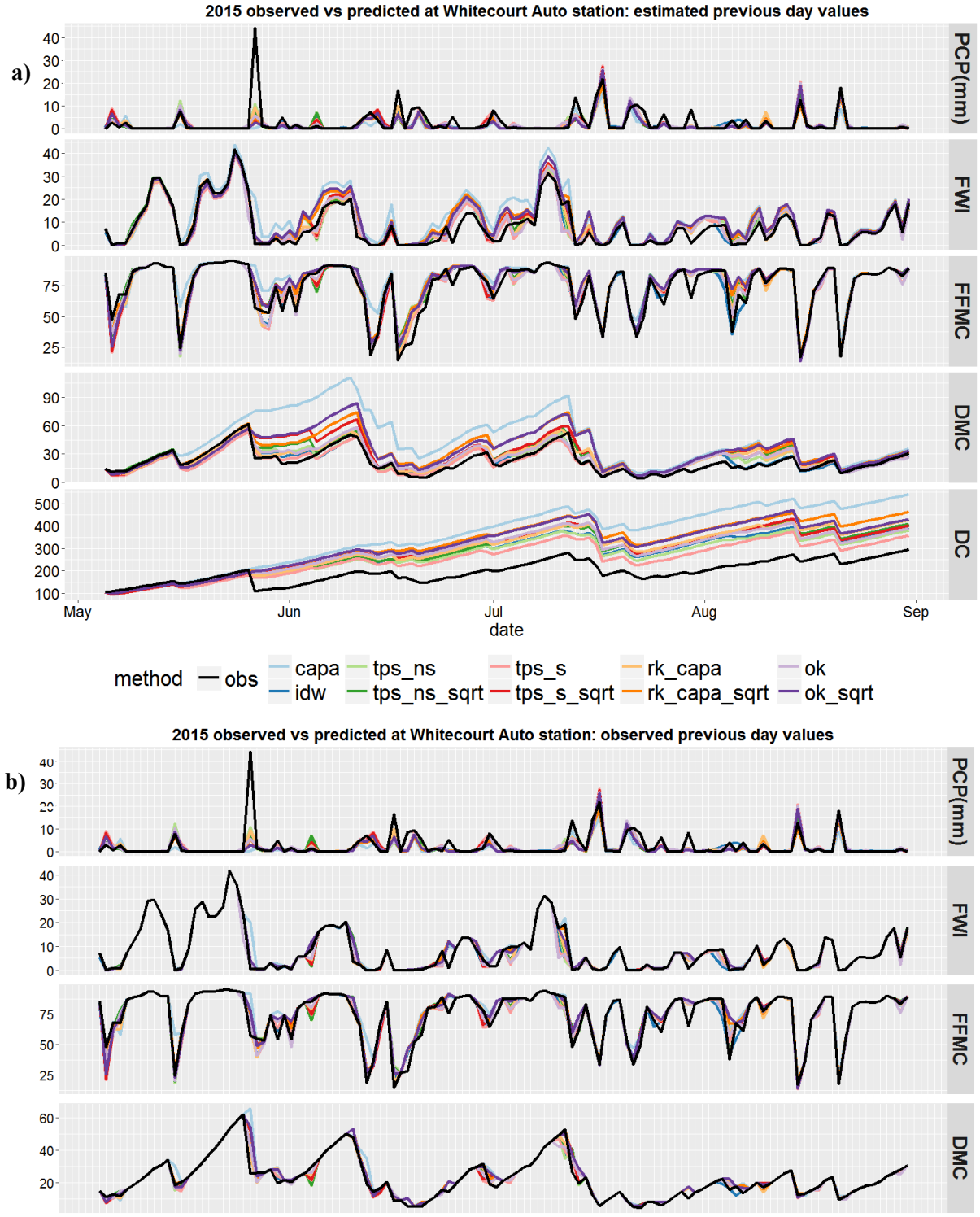


Figure 3.9 Time series plots of observations versus predictions of precipitation and FWI System calculated using (a) estimated previous day values and (b) observed previous day values at Whitecourt Auto station in 2015 (123 days)

Whitecourt Auto is located in the validation Foothills region and is covered by the 120km Doppler radar range; and received the largest PCP accumulations in 2015. FWI System codes were calculated using estimated PCP and observed RH, Temp, WS.

3.4 Weather station density

Thus far, all the results presented above using all weather stations in the study area (an average of 2.69 weather stations / 10 000km^2 in 2015). In this section, we examined the performance of interpolation algorithms under various station density scenarios. This was achieved by a random selection of 10%, 25%, 50%, 75%, and 90% of all the stations in our study areas. MAE values of precipitation and FWI System estimates at every station density scenario were summarized in Table 3.9.

For precipitation, every interpolation algorithm had reduced MAE values as the station density increased (Table 3.9, 3.10). The MAE values of CaPA System remained the same because the station network integrated into the CaPA system was constant (~1.58 weather stations / 10 000km^2) during the analysis. The best performing method shifted frequently as the station densities increased: CaPA was the best performing method when there were less than 0.6 weather stations /10 000km^2; rk(capa)_sqrt was the best performing method when there were 0.6 -2.25 weather stations / 10 000km^2; and ok_sqrt was the best performing method when there were more than 2.25 weather stations / 10 000km^2. The worst performing method also shifted between tps_ns, IDW, and OK as the station densities increased.

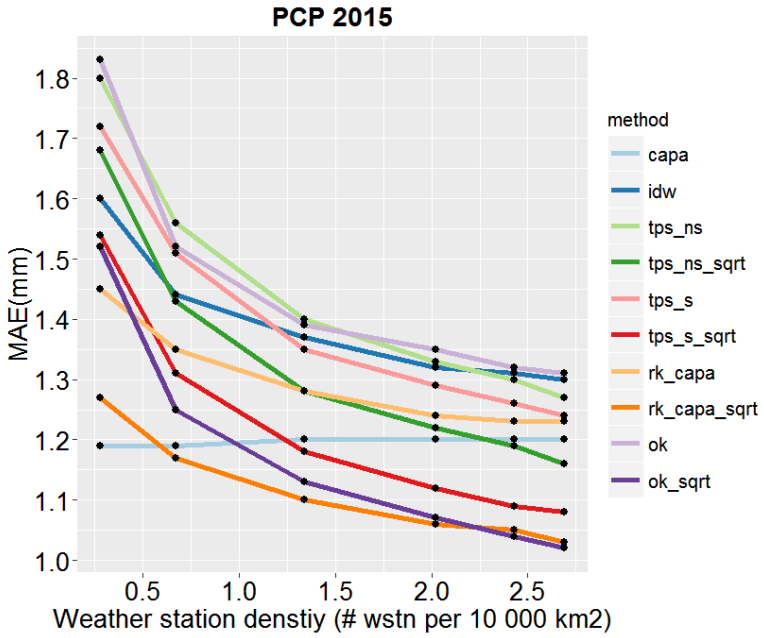


Figure 3.10 MAE of candidate algorithms with changing weather station densities in 2015
MAE values are averaged over the 100 resampling replicates

The sensitivity of FWI System to station densities was divided into two groups according to the dying time lags of the indexes (Figure 3.11 and Table 3.9).

The quick drying indexes FFMC, ISI, and FWI had a similar trend to the precipitation estimates at every station density scenario (Figure 3.11). For example, CaPA was the best performing method for FFMC, ISI, and FWI indexes when station density was lower than 0.6 weather stations / 10 000km^2; rk(capa)_sqrt was the best performing method for FFMC, ISI, and FWI indexes when station density was higher than 0.6 weather stations / 10 000km^2. The difference between the top few interpolation algorithms became smaller as the station density increased. For example, the MAE values of FWI System estimated with tps_ns_sqrt was similar to the MAE value of FWI System estimated with rk(capa)_sqrt when station density was greater than 2.5 weather stations / 10 000km^2 (Figure 3.11).

For slow drying indexes DMC, DC, and BUI, rk(capa) was the best algorithm regardless of station densities (Figure 3.11). The worst performing algorithm for DMC shifted from ok_sqrt to CaPA when station density exceeded 2.0 weather stations / 10 000km^2. The worst performing algorithm for DC shifted from ok_sqrt to CaPA when station density exceeded 1.7 weather stations / 10 000km^2.

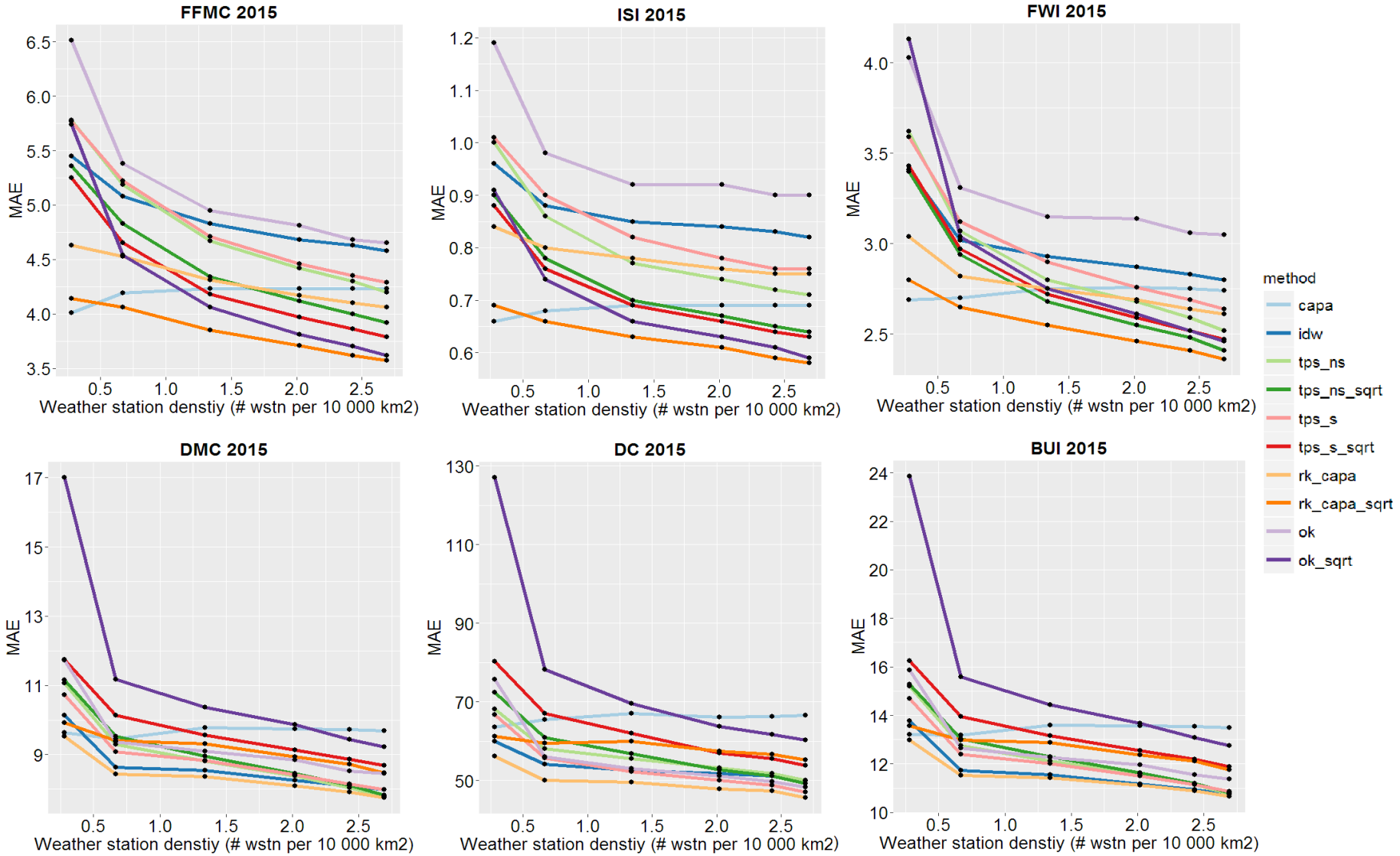


Figure 3.11 MAE of FWI System with changing weather station densities in 2015

MAE values are averaged over the 100 resampling replicates. FWI System indices were calculated using estimated PCP and observed RH, Temp, WS; calculation of fuel moisture codes used values estimated from the previous day.

Table 3.9 Effect of station density on MAE of precipitation and FWI System in 2015

FWI System codes were calculated using estimated PCP and observed RH, Temp, WS; calculation of fuel moisture codes using estimated previous day values. MAE values are averaged over the 100 resampling replicates.

Index	Density[1]	% [2]	capa	idw	tps_ns	tps_ns_s	tps_s	tps_s_s	rk (capa)	rk (capa)_s	ok	ok_s
PCP	0.28	10%	**1.19**	1.60	1.80	1.68	1.72	1.54	1.45	1.27	1.83	1.52
	0.67	25%	1.19	1.44	1.56	1.43	1.51	1.31	1.35	**1.17**	1.52	1.25
	1.34	50%	1.20	1.37	1.40	1.28	1.35	1.18	1.28	**1.10**	1.39	1.13
	2.02	75%	1.20	1.32	1.33	1.22	1.29	1.12	1.24	**1.06**	1.35	1.07
	2.43	90%	1.20	1.31	1.30	1.19	1.26	1.09	1.23	1.05	1.32	**1.04**
	2.69	100%	1.20	1.30	1.27	1.16	1.24	1.08	1.23	1.03	1.31	**1.02**
FFMC	0.28	10%	**4.01**	5.45	5.78	5.36	5.77	5.25	4.63	4.14	6.51	5.74
	0.67	25%	4.19	5.08	5.19	4.83	5.22	4.65	4.53	**4.06**	5.38	4.54
	1.34	50%	4.23	4.83	4.67	4.34	4.71	4.18	4.31	**3.85**	4.95	4.06
	2.02	75%	4.23	4.68	4.42	4.12	4.46	3.97	4.17	**3.71**	4.81	3.81
	2.43	90%	4.23	4.63	4.30	4.00	4.35	3.86	4.10	**3.62**	4.68	3.70
	2.69	100%	4.23	4.58	4.20	3.92	4.29	3.79	4.06	**3.57**	4.65	3.62
ISI	0.28	10%	**0.66**	0.96	1.00	0.90	1.01	0.88	0.84	0.69	1.19	0.91
	0.67	25%	0.68	0.88	0.86	0.78	0.90	0.76	0.80	**0.66**	0.98	0.74
	1.34	50%	0.69	0.85	0.77	0.70	0.82	0.69	0.78	**0.63**	0.92	0.66
	2.02	75%	0.69	0.84	0.74	0.67	0.78	0.66	0.76	**0.61**	0.92	0.63
	2.43	90%	0.69	0.83	0.72	0.65	0.76	0.64	0.75	**0.59**	0.90	0.61
	2.69	100%	0.69	0.82	0.71	0.64	0.76	0.63	0.75	**0.58**	0.90	0.59
FWI	0.28	10%	**2.69**	3.41	3.62	3.40	3.59	3.43	3.04	2.80	4.03	4.13
	0.67	25%	2.70	3.02	3.07	2.94	3.12	2.97	2.82	**2.65**	3.31	3.04
	1.34	50%	2.75	2.93	2.80	2.68	2.90	2.72	2.75	**2.55**	3.15	2.75
	2.02	75%	2.76	2.87	2.68	2.55	2.76	2.59	2.69	**2.46**	3.14	2.61
	2.43	90%	2.75	2.83	2.59	2.48	2.69	2.52	2.64	**2.41**	3.06	2.52
	2.69	100%	2.74	2.80	2.52	2.41	2.64	2.47	2.61	**2.36**	3.05	2.46
DMC	0.28	10%	9.63	10.13	11.07	11.16	10.73	11.74	**9.52**	9.93	11.72	17.00
	0.67	25%	9.46	8.64	9.29	9.53	9.08	10.13	**8.44**	9.40	9.36	11.18
	1.34	50%	9.77	8.55	8.81	8.96	8.83	9.57	**8.37**	9.32	9.10	10.37
	2.02	75%	9.74	8.25	8.39	8.45	8.42	9.13	**8.10**	8.94	8.85	9.87
	2.43	90%	9.72	8.10	8.05	8.11	8.15	8.87	**7.92**	8.73	8.53	9.44
	2.69	100%	9.68	7.98	7.79	7.82	7.98	8.68	**7.75**	8.48	8.46	9.22
DC	0.28	10%	63.56	59.97	68.06	72.37	66.69	80.27	**56.16**	61.19	75.63	127.0
	0.67	25%	65.48	54.09	58.06	60.85	55.69	66.98	**50.02**	59.46	55.94	78.18
	1.34	50%	67.02	52.58	55.60	56.74	52.16	62.04	**49.48**	59.97	52.93	69.60
	2.02	75%	66.13	51.73	53.13	52.70	49.96	56.93	**47.82**	57.35	51.05	63.73
	2.43	90%	66.25	51.15	51.79	51.16	48.69	55.50	**47.28**	56.56	49.74	61.74
	2.69	100%	66.52	49.78	49.95	49.27	46.99	53.75	**45.54**	55.22	48.31	60.24
BUI	0.28	10%	13.21	13.77	15.19	15.29	14.70	16.25	**12.98**	13.57	15.86	23.84
	0.67	25%	13.19	11.74	12.75	13.04	12.39	13.95	**11.53**	12.98	12.65	15.59
	1.34	50%	13.60	11.56	12.12	12.29	12.00	13.17	**11.43**	12.88	12.26	14.43
	2.02	75%	13.56	11.16	11.58	11.62	11.50	12.54	**11.12**	12.37	11.95	13.67
	2.43	90%	13.54	10.95	11.15	11.20	11.14	12.19	**10.90**	12.11	11.55	13.09
	2.69	100%	13.49	10.74	10.76	10.80	10.86	11.89	**10.65**	11.76	11.38	12.76

[1] Density is the amount of stations per 10 000km^2. [2] Number of selected weather stations was rounded up.

Bold: best performing method (smallest MAE). Underline: worst performing method (biggest MAE)

Chapter 4 Discussion

In this chapter, a discussion of the following a few issues was provided, including (1) the mid--tiered performance of CaPA System; (2) impacts of weather station density on model performance in Alberta; (3) advantages and disadvantages of using regression kriging with CaPA as covariate for fire weather interpolation; (4) impacts of error propagation in FWI System, and (5) possible sources of errors and future research directions.

4.1 Evaluation of the CaPA System

Our results showed that the CaPA System was a mid-tiered method in estimating precipitation (13^{th} of 18) in the forested area of Alberta, except in Doppler radar covered area, where CaPA System produced the second best precipitation estimates. There are three reasons for the mid-tiered performance of the CaPA System.

Firstly, the density of ECCC weather stations incorporated into CaPA System was much lower than the density of AAF weather stations used by the other interpolation methods (1.58 versus 2.56 weather stations / 10 000 km^2). Weather station observations are one of the three inputs for CaPA System. As suggested by Mahfouf *et al.,* (2007), CaPA System is more accurate in areas close to weather station observations; otherwise, CaPA System relies on the less accurate GEM forecasts. In general, a density of 1.17 stations / 10 000 km^2 was required by CaPA System to outperform GEM forecasts (Lespinas *et al.,* 2015). In our case, ECCC station density exceeded this threshold, but AAF station density was even higher and enabled some interpolation methods to outperform CaPA System. Studies have suggested that the accuracy of interpolation methods decrease with station density (Flannigan *et al.,* 1998; Hofstra *et al.,* 2010). This study found that when AAF station density dropped below 0.6 weather stations / 10 000km^2, CaPA System would be superior to all interpolation methods considered in this study. This finding is important because it proves the validity of using CaPA System in data scarce areas (e.g., northern Alberta). The threshold of 0.6 weather stations / 10 000km^2 can be used by fire management agencies to determine whether the use of CaPA System will be a valuable asset for fire danger rating in areas between weather stations. Additionally, ECCC and AAF operate two

separate weather station networks; we feel that it would be beneficial to both agencies by integrating the two data sources.

Secondly, only 10% of our study area was covered by 120km Doppler radar range. This is unfortunate because our study showed that the accuracy of CaPA System was greatly improved in areas under the radar (~8% improvement regarding MAE). Our finding agrees with the work by Fortin *et al.,* (2015), who indicated that the inclusion of radar composite significantly improves the CaPA System. We also found that the largest improvement of CaPA System under radar lies in predicting precipitation less than 1.5 mm, leading to greater improvement of FFMC, which is more sensitive to small precipitation events. Flannigan *et al.,* (1998) also suggested that radar would be used best to identify areas with and without precipitation compared to interpolation methods. The usefulness of CaPA System will increase if there was an additional radar station in the center of Boreal forest in Alberta.

Thirdly, CaPA System resulted in a larger negative Bias comparing to interpolation methods, especially when predicting precipitation greater than 2.8mm. This tendency of underestimation was true for CaPA System regardless of radar coverage and can be attributed to the cubic root smoothing applied to CaPA System inputs (Lespinas *et al.,* 2015). Our study showed that when performing square root, cubic root, and natural log on precipitation observations, all the interpolation methods yielded a significant reduction in prediction errors. However, smoothing the observed precipitation inevitably resulted in an underestimation of precipitation; with cubic root smoothing technique having the biggest negative bias compared to the other two smoothing techniques. Underestimating larger amounts of precipitation is problematic when estimating FWI System that is sensitive to seasonally accumulating precipitation (e.g., Drought Code, as discussed in Section 4.3).

Overall, we recommended that the CaPA System could be used to estimate fire danger in areas with low fire weather station density and areas under Doppler radar coverage. However, CaPA System should be used with care when large precipitation events are critical concerns.

4.2 Selection of the best method

Regression kriging produced the best estimates of precipitation regardless of the radar coverage. This is expected, as many interpolation studies have suggested that multivariate geostatistical algorithms are superior to univariate geostatistical algorithms (e.g., ordinary kriging) and simple distance-only-based algorithms (Seo *et al.,* 1990a; Seo *et a.,* 1990b; Haberlandt 2007; Li and Heap 2014). By treating CaPA System as a covariate, regression kriging produced better precipitation estimates than using CaPA System alone, especially when estimating large precipitation events (>2.8mm). The findings of our study suggest that regression kriging that uses gridded precipitation analysis as covariate may also address problems found by Flannigan *et al.,* (1998) and Horel *et al.,* (2014). In particular, regression kriging has resolved the systemic underestimation from radar estimates (Flannigan *et al*., 1998) and gridded precipitation from Real-Time Mesoscale Analysis (RTMA) (Horel *et al.,* 2014). Additionally, computational processing time is critical to fire operation agencies. In this study, we found that regression kriging algorithm did not necessarily require more processing time than that of IDW (e.g., 3mins versus 1min to interpolate precipitation over 10,000 points). We, therefore, recommend the use of regression kriging with CaPA as a covariate to estimate precipitation and fire danger in areas between weather stations.

Smoothed TPS was an upper mid-tiered algorithm. This finding proves the validity of using smoothed TPS to estimate fire danger across landscapes as suggested by Flannigan and Wotton (1989). We also found that smoothed TPS was superior to non-smoothed TPS, but smoothed TPS may overly smoothed the precipitation field and resulted in negative bias. Both smoothed and non-smoothed TPS could include elevation adjustment, making them a good fit to interpolation in mountainous areas (Hutchinson *et al.,* 2009).

On the other hand, IDW, which is currently used by SFMS (Englefield *et al.,* 2000), was found to be the second worst algorithm. This finding is in agreement with Flannigan and Wotton (1989), who suggested that IDW is a crude method and could not address the spatial variability of summer precipitation, leading to poor estimates of FWI System Indexes. We recommend that IDW should be replaced by a more sophisticated method.

4.3 Impacts of error propagation in FWI System

In this study, we observed that the model performance was affected by error propagation in the DMC and DC. The error propagation is attributed to their long drying time lags (the time it requires for fuel to lose ~2/3 of its free moisture toward equilibrium moisture content) of 12 and 52 days (Van Wagner 1987). Our study showed that algorithms that frequently underestimated large precipitation (>2.8mm) experienced the most error propagation in DMC and DC. This is because DMC and DC track moisture content in compacted organic soil layers, where are affected by large precipitation events and are sensitive to seasonal precipitation (Lawson and Armitage 2008).

When an extreme precipitation event (e.g., 40mm) is missed or underestimated at the beginning of fire season, all the FWI System will be overestimated. As FFMC will recover quickly to the code value before the precipitation (1-2days), DC and DMC may propagate this error until the end of fire season and may result in a twice larger estimate compared to the true value (see Figure 3.9a).Since DMC and DC are used for long-term fire management planning, overestimating of these indexes will result in false alarms and cause unnecessary pre-suppression resource allocation. In addition, overestimating DMC values would also cause problems in predicting the occurrence of lightning-caused fires (Wotton and Martell 2005). In our case, both CaPA System and smoothing effect resulted in substantial underestimation of precipitation and produced the worst estimates for DMC, DC. Therefore, CaPA System and smoothing effect should be used with care when DMC and DC are of critical importance.

Our study showed that using previous day's *observed* fuel moisture codes to calculate today's fuel moisture codes would eliminate the error propagation (see Figure 3.9b). When error propagation was removed, the performance of all the algorithms in estimating FWI System was stable. This approach may provide a solution to adjust the error propagation of DMC and DC when previous day's observations are available. Our study suggests that regression kriging without smoothing resulted in the best estimates for DMC and DC, but more options on this topic should be investigated in the future.

4.4 Limitations and future research

Data quality is one of the major constraints in this study. We assumed that AAF station observations were representative of true values and we used AAF station data to validate the CaPA System estimates. Although we performed quality control procedure on the AAF station data, our quality control procedures may be different from those implemented by ECCC. AAF and ECCC station observations may all be impacted by the following factors: 1) differences in the types of rain gauge may results in various levels of accuracy; 2) rain gauge measurements may be biased on windy days; 3) animals and woody debris may also alter the observations; 4) some small precipitation events may evaporate before been measured. In this study, we evaluated the candidate algorithms for one and a half fire seasons that were drier than normal. Although we are concerned with the dry years for fires and results of this study may apply to wet years, it would be better to validate all the candidate algorithms for fire seasons with different climatology (e.g., a wet year).

As discussed previously, one of the many feasible future research directions is to test the performance of regression kriging with other gridded analysis products. Gridded dataset such as the Global Precipitation Measurement (GPM) (Smith *et al.,* 2007) and PERSIANN (Sorooshian *et al.,* 2005) could be tested because both of them calibrate satellite data. Using remote sensing for fire danger rating should also be evaluated for future research, especially for the slow drying indexes DC and DMC (Oldford *et al.,* 2007). Additionally, alternative interpolation algorithms should also be evaluated. These interpolation algorithms include random forest which combines IDW and ordinary kriging with concepts from machine learning (Sanabria *et al.,*2013).

Chapter 5 Conclusions

In this study, we addressed the challenges of estimating fire danger rating between weather stations through improved precipitation estimates. We also quantified the performance of interpolation methods with different station densities. To achieve these objectives, we compared precipitation estimates from eighteen algorithms of six methods and assessed the impacts of these precipitation estimates on the Fire Weather Indexes (FWI) System. The 6 methods include one gridded precipitation analysis (CaPA System) and 5 interpolation methods (IDW, smoothed TPS, non-smoothed TPS, ordinary kriging; and regression kriging). Estimates of these algorithms were generated using station observations through leave-one-out cross-validation and were evaluated using a range of skill scores (Bias, MAE, ETS, and FBI) with the resampling hypothesis test.

Regression kriging was the best overall algorithm in estimating precipitation between weather stations in the study area, closely followed by ordinary kriging and smoothed TPS, all of which were performed on square root smoothed precipitation. Regression kriging with smoothing had 19.6% less MAE than that of IDW. CaPA System had mid-tiered performance (13^{th} of 18), but it produced significantly better precipitation estimates than that of IDW. In areas within 120km Doppler radar, CaPA System had improved performance and ranked second. However, CaPA System tended to largely underestimate precipitation values greater than 2.8 mm regardless of the radar coverage. This could be problematic for indexes of the FWI System that are sensitive to seasonal precipitation (e.g., DC, as discussed in Section 4.3).The tendency of underestimating large precipitation values was also found when applying square root smoothing to observed precipitation.

Weather station density significantly affected the performance of interpolation algorithms (Figure 3.11). When weather station density dropped below 0.6 stations / 10 000km^2, CaPA System became the best algorithm; whereas when station density was above this threshold, regression kriging with smoothing was the best algorithm. Although all the interpolation algorithms performed better when station density increased, we found that these algorithms did not respond equally to the increasing station densities. For example, when weather station density was above 2.25 weather stations / 10 000km^2, the top three performing algorithms had

very similar performance; while when weather station density was below this threshold, regression kriging with smoothing was superior to ordinary kriging and smoothed TPS(Figure 3.11).

FWI System were significantly better estimated using improved precipitation estimates, e.g., FWI values estimated from regression kriging with smoothing had 22% less MAE than that of IDW. Our study showed that the FWI System had two groups of responses to precipitation estimates, depending on 1) the drying time lags of the indexes (Figure 3.9) and 2) weather station density (Figure 3.11). For quick drying indexes (i.e., FFMC, ISI, and FWI), regression kriging with smoothing was the best performing algorithm when weather station density was above 0.6 weather stations / 10 000km^2. Otherwise, CaPA System was the best performing algorithm. For slow drying indexes (i.e., DMC, DC, and BUI), regression kriging without smoothing was the best performing algorithm regardless of the weather station density. This is because slow drying indexes are more sensitive to seasonal precipitation, therefore these indexes preferred methods with lower Bias (e.g., regression kriging without smoothing). Overall, we recommend the use of regression kriging with CaPA as a covariate to estimate fire danger across landscapes.

References

Balshi, M. S., McGuire, A. D., Duffy, P., Flannigan, M., Walsh, J., & Melillo, J. (2009). Assessing the response of area burned to changing climate in western boreal North America using a Multivariate Adaptive Regression Splines (MARS) approach. *Global Change Biology*, 5(3), 578-600.

Benfield, A. (2016 May). *Global Catastrophe Recap. Aon Benfield. 15 pgs.* Retrieved from http://thoughtleadership.aonbenfield.com/Documents/20160608-ab-analytics-if-may-global-recap.pdf

Beverly, J., & Martell, D. (2005). Charaterizing extreme fire and weather events in the Boreal Shield ecozone of Ontario. *Agriculture and Forest Meteorology*, 133:5-16. http://doi.org/10.1016/j.agrformet.2005.07.015

Beverly, J., & Wotton, M. (2007). Modelling the probability of sustained flaming: Predictive value of fire weather index components compared with observations of site weather and fuel moisture conditions. *International Journal of Wildland Fire*, 16, 161-173. DOI: 10.1071/WF06072.

Beverly, J., Flannigan, M., Stocks, B., & Bothwell, P. (2011). The association between Northern Hemisphere climate patterns and interannual variability in Canadian wildfire activity. *Canadian Journal of Forest Research*, 41: 2193–2201. doi:10.1139/x11-131.

Burrough, P., & McDonnell, R. (1998). The principles of geographical information systems.*Oxford University Press.*

Cary, G., Keane, R., Gardner, R., Lavorel, S., Flannigan, M., Davies, I.,Mouillot, F, et al. (2006). Comparison of the sensitivity of landscape-fire-succession models to variation in terrain, fuel pattern, climate and weather. *Landscape Ecology*, 21:121–137. doi:10.1007/s10980-005-7302-9.

CIFFC. (2013). CIFFC Canada Report . http://www.ciffc.ca/images/stories/pdf/2013_canada_report.pdf: Canadian Interagency Forest Fire Center Inc.

Countryman, C. (1972). The fire environment concept. Berkeley, CA: USDA Forest Service, Pacific Southwest Forest and Range Experiment Station.

Cressie, N. (1993). Statistics for Spatial Data. *John Wiley and Sons*, 900 pp.

Daly, C. (2006). Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology*, 26,707-721.

Daley, R. (1991). Atmospheric Data Analysis. *Cambridge,UK: Cambridge University Press*, 457pp.

de Groot W.J, Flannigan, M., & Cantin, A. (2013). Climate change impacts on future boreal fire regimes. *Forest Ecology and Management*, 294:35-44 http://dx.doi.org/10.1016/j.foreco.2012.09.027.

Douglas, N., R. F., John, P., & Stephan, S. (2016). *fields: Tools for Spatial Data. R package version 8.3-6.* Retrieved from http://CRAN.R-project.org/package=fields

Englefield, P., Lee, B., & Suddaby, R. (2000). Spatial fire management system . *ESRI International User Conference* (p. 489). San Diego, California: ESRI, San Diego, California, USA. 10 p.

Evans, A. (2013). Investigation of enhancements to two fundamental components of the statistical interpolation method used by the Canadian Precipitation Analysis (CaPA) (Master's thesis). *Retrieved from the Department of Civil Engineering, University of Manitoba.*, Available online at http://hdl.handle.net/1993/22276.

Flannigan, M., & Harrington, J. (1988). A study of the relation of meteorological variables to monthly provincial area burned. *Journal of Applied Meteorology*, 4: 441-452.

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

Flannigan, M., Krawchuk, K., de Groot, W., Wotton, B., & Gowman, L. (2009). Implications of changing climate for global wildland fire. *International Journal of Wildland Fire*, 18:483-507.

Flannigan, M., Logan, K., Skinner, W., & Stocks, B. (2005). Future Area burned in Canada. *Climatic Change*, 72: 1–16.doi:10.1007/s10584-005-5935-y.

Flannigan, M., Wotton, B., & Ziga, S. (1998). A study on the interpolation of fire danger using radar precipitation estimates. *International Journal of Wildland Fire*, 8(4):217-225.

Flannigan, M., Wotton, B., Marshall, G., de Groot, W., Johnston, J., Jurko, N., & Cantin, A. (2016). Fuel moisture sensitivity to temperature and precipitation: climate change implications. *Climatic Change*, 134: 59-71. doi:10.1007/s10584-015-1521-0.

Flat Top Complex Wildfire Review Committee. (2012, May 18). *Flat Top Complex Wildfire Review Committee*. Retrieved from http://wildfire.alberta.ca/wildfire-prevention-enforcement/wildfire-reviews/documents/FlatTopComplex-WildfireReviewCommittee-A-May18-2012.pdf

Forestry Canada Fire Danger Group. (1992). *Development and Structure of Canadian Forest Fire Behavior Prediction System.* Ottawa: Foresty Canada Science and Sustainable Development Directorate.

Fortin, V., Roy, G., Donaldson, N., & Mahidjiba, A. (2015). Assimilation of radar quantitative precipitation estimations in the Canadian Precipitation Analysis (CaPA). *Journal of Hydrology*, 531: 296-307.

Government of Alberta. (2015, January 16). Retrieved from http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/sag6299

Gillett, N., Weaver, A., Zwiers, F., & Flannigan, M. (2004). Detecting the effect of climate change on Canadian forest fires. *Geophysical Reseach Letters*, 31, L18211, doi:10.1029/2004GL020876.

Haberlandt, U. (2007). Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *Journal of Hydrology*, 332, 144-157.

Hamill, T. (1999). Hypothesis test for evaluating numerical precipitation forecasts. *America Meteorological Society*, 155-167.

Haylock, M. R., Hofstra, N., Klein Tank, V., Klok, E., Jones, P., & New, M. (2008). A European daily high-resolution gridded dataset of surface temperature and precipitation for 1950–2006. *Jurnal of Geophysical Research*, 113, D20119. DOI: 10.1029/2008JD010201.

Hengl, T., Heuvelink, G., & Rossiter, D. (2007). About regression-kriging:

From equations to case studies. *Computers&Geosciences*, 33(10):1201-1315. doi.org/10.1016/j.cageo.2007.05.001.

Heyerdahl, E. K., B. L., & Agee, J. K. (2011). Spatial controls of historical fire regimes: a multiscale example from the Interior West, USA. *Ecology*, 82: 660–678. DOI:10.1890/0012-9658(2001)082[0660: SCOHFR]2.0.CO;2.

Holden, Z. A., Morgan, P., Crimmins, M. A., Steinhorst, R. K., & Smith, A. M. (2007). Fire season precipitation variability influences fire extent and severity in a large southwestern wilderness area, United States. *Geophysical Research Letters*, 34(16), [L16708]. DOI: 10.1029/2007GL030804.

Hofstra, N., Haylock, M., New, M., Jones, P., & Frei, C. (2008). Comparison of six methods for the interpolation of daily European climate data. *Journal of Geophysical Research*, 113: D21110.

Hofstra, N., New, M., & McSweeney, C. (2010). The influence of interpolation and station network density on the distributions and trends of climate variables in gridded daily data. *Climate Dynamics*, 35:841-858.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.

Horel, J. D., Ziel, R., Galli, C., Pechmann, J., & Dong, X. (2014). An evaluation of fire danger and behaviour indices in the Great Lakes Region calculated from station and gridded weather information. *International Journal of Wildland Fire*, 23(2), 202-214.

Hutchinson, M. (1993). On the thin plate splines and krigging. *Computer Science and Statistics*, 25: 55-62.

Hutchinson, M. (1995). Interpolating mean rainfall using thin plate smoothing splines. *International Journal of Geographical Information System*, 9(4): 385-403,DOI: 10.1080/02693799508902045.

Hutchinson, M. (1998a). Interpolation of rainfall data with thin plate smoothing splines–Part I:two-dimensional smoothing of data with short range correlation. *Journal of Geographic Information and Decision Analysis*, 2(2): 139-151.

Hutchinson, M. (1998b). Interpolation of rainfall data with thin plate smoothing splines–Part II: analysis of topographic dependence. *Journal of Geographic Information and Decision Analysis*, 2(2):152-167.

Hutchinson, M., McKenney, D., Lawrence, K., Pedlar, J., Hopkinson, R., Milewska, E., & Papadopol, P. (2009). Development and Testing of Canada-Wide Interpolated Spatial Models of Daily Minimum–Maximum Temperature and Precipitation for 1961–2003. *Journal of Applied Meteorology and Climatoloy*, 48: 725-741.

Isaaks, E., & Srivastava, R. (1989). An Introduction to Applied Geostatistics. *Oxford University Press*, 561 pages.

Johnson, E. (1972). Fire and vegetation dynamics: Studies from the North America boreal forest. *Cambridge: Cambridge University Press.*

Johnston, L. (2016). Mapping Canadian Wildland Fire Interface Areas. Retrieved from https://era.library.ualberta.ca/files/cnv935306n#.WKSHzlUrJtQ.

Krezek-Hanes, C., Ahern, F., Cantin, A., & Flannigan, M. (2011). Trends in large fires in Canada, 1959-2007. Canadian Biodiversity: Ecosystem Status and Trends 2010, Technical Thematic Report. *No. 6.* Ottawa, ON 48p http://www.biodivcanada.ca/default.asp?lang=En&n=137E1147-0: Canadian Councils of Resource Ministers., AB: Natural Resources of Canada, Canadian Forest Service, Northern Forestry Centre.

Lawson,B.D & Armitage,O.B. (2008). Weather guide for canadian forest fire danger rating system. Edmonton: Canadian Forest Service, Northern Forestry Centre.

Lee, B., Alexander, M., Hawkes, B., Lynham, T., Stocks, B., & Englefield, P. (2002). Information systems in support of wildland fire management decision making in Canada. *Computers and Electronics in Agriculture*, 37, 185-198. Retrieved from Canadian Wildland Fire Information System: http://cwfis.cfs.nrcan.gc.ca/background/dsm/fwi

Lespians, F., Fortin, V., Roy, G., Rasmussen, P., & Stadnyk, T. (2015). Performance evaluation of the canadian precipitation analysis (CaPA). *Journal of Hydrometeorology*, 16: 2045-2064.

Li, J., & Heap, A. (2008). A review of spatial interpolation methods for environmental scientists. *Geoscience Australia*, p.137. Record 2008/23.

Li, J., & Heap, A. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factor. *Ecological Informatics* , 6, 228-241 http://dx.doi.org/10.1016/j.ecoinf.2010.12.003.

Li, J., & Heap, A. (2014). Spatial interpolation methods applied in the environmental science. *Environmental Modelling&Software*, 53:173-189.

Livezey and Chen. (1983). Statistical filed significance and its determine by Monte Carlo technique. *Monthly Weather Review*, 111:46-49.

Ly, S., Charles, C., & Degre, A. (2011). Geostatistical interpolation of daily rainfall at catchment scale:the use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrology Earth System*, 15, 2259–2274.

Mahfouf, J., Brasnett, B., & Gagnon, S. (2007). A Canadian precipitation analysis (CaPA) project: Description and preliminary results. *Atmosphere Ocean*, 45,1-17.

Martell, D. (2001). Forest fire management. Chapter 15. In E. A. Johnson, & K. E. Miyanishi, *Forest Fires: Behaviour and Ecological effects* (pp. pp. 527-583). San Diego: Academic Press.

Matheron, G. (1963). Principle of geostatistics. *Economic Geology*, 58:1246-1266.

McArthus, A. (1968). Fire behaviour in eucalypt forests. (p. Leaflet No. 107). India: Ninth Commonwealth Forestry Conference.

McKenney, D., Pedlar, J., Papadopol, P., & Hutchinson, M. (2006). The development of 1901–2000 historical monthly climate models for Canada and the United States. *Agricultural and Forest Meteorology*, 138, 69–81.

Merrill, D., & Alexander, M. (1987). Glossary of forest fire management terms.*Fourth Edition.* Ottawa, ON: Canadian Committee on Forest Fire Management, National Research Council of Canada.No 26516.

Moritz, M., Parisien, M., B. E., Krawchuk, M., Van Dorn, J., Ganz, D., & Hayhoe, K. (2012). Climate change and disruptions to global fire activity. *Ecosphere*, 3: 1-22. doi:10.1890/ES11-00345.1.

Natural Regions Committee. (2006). Natural regions and subregions of Alberta. Edmonton, Alberta,Canada: Complied by DJ Downing and WW Pettapoece.

Natural Resource Canada. (2008, 07 11). Canadian wildland fire information system - Canadian Forest Fire Weather Indexes (FWI) System. Retrieved from http://cwfis.cfs.nrcan.gc.ca/background/dsm/fwi

Odeh, I., McBratney, A., & Chittleborough, D. (1995). Further results on the prediction of soil properties from terrain attribute: heterotopic cokriging and regression-kriging. *Geoderma*, 67 (3-4): 215–226. doi.org/10.1016/0016-7061(95)00007-B.

Oldford, S., Leblon, B. M., & Flannigan, M. (2007). Predicting slow-drying fire weather index fuel moisture codes with NOAA-AVHRR images in Canada's northern boreal forests. *International Journal of Remote Sensing*, 27(18): 3881-3902.

Parisien, M., Kafka, V., Hirsch, K., Todd, J., Lavoie, S., & Maczek, P. (2005). *Mapping wildfire susceptibility with the BURN-P3 simulation model.* Edmonton, Alberta. 36pp http://cfs.nrcan.gc.ca/publications?id=25627: Nature Resource Canada, Canadian Forest Service, Northern Forestry Centre.

Parisien, M.-A., Parks, S., Krawchun, M., Flannigan, M., Bowman, L., & Morite, M. (2011). Scale-dependent controls on the area burned in the boreal forest of Canada, 1980–2005. *Ecological Applications*, 21(3):789-805. DOI: 10.1890/10-0326.1.

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

Politis, N., & Romano, J. (1994). The Stationary Bootstrap. *Journal of the American Statistical Association*, 89:1303-1313.

Price, D., McKenney, D., Nalderc, I., Hutchinson, M., & Kesteven, J. (2000). A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data. *Agricultural and Forest Meteorology*, 101 (2000) 81–94.

R Core Team. (2015). *foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase. R package version 0.8-66.* Retrieved from http://CRAN.R-project.org/package=foreign

R Development Core Team. (2016). *R*: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org.

Rowe, J. (1983). The role of fire in northern circumpolar ecosystems-Concepts of fire effects on plant individuals and species. *New York: John Wiley and Sons.*

Sanabria, L., Qin, X., Cechet, R., & Lucas, C. (2013). Spatial interpolation of McArthur's Forest Fire Danger Index across Australia: Observational study. *Journal of Environmental Modelling & Software*, 50,37-50. doi>10.1016/j.envsoft.2013.08.012.

Seo, D., Krajewski, W., & Bowles, D. (1990a). Stochastic Interpolation of Rainfall Data From Rain Gages and Radar Using Cokriging 1. The design of Experimetns. *Water Resources Research*, 26, NO. 3, 469-477,

Seo, D., Krajewski, W., & Bowles, D. (1990b). Stochastic Interpolation of Rainfall Data From Rain Gages and Radar Using Cokriging 2. results. *Water Resources Research*, 25, NO.5, 915-924.

Shen, S., Dzikowski, P., Li, G., & Griffith, D. (2001). Interpolation of 1961–97 daily temperature and precipitation data onto Alberta polygons of eco district and soil landscapes of Canada. *Journal of Applied Meteorology*, 40: 2162–2177.

Shepard, D. (1968). A Two-Dimensional Interpolation Function for Irregularly Spaced Data. *Proceedings of 23rd ACM national conference* (pp. Pages 517-524). New York: ACM Digital Library. doi>10.1145/800186.810616.

Smith, E., & Coauthors. (2007). *International Global Precipitation Measurement (GPM) Program and Mission: An Overview.* Measuring Precipitation from Space: URAINSAT and the Future, V. Levizzani and F. J. Turk, Eds., Springer, 611–653.

Sorooshian, S., Hsu, K., Imam, B., & Hong, Y. (2005). Global Precipitation Estimation from Satellite Image Using Artificial Neural Networkds. *Journal of Applied Meteorology*, 36,1176-1190. Retrieved from http://gwadi.org/sites/gwadi.org/files/Sorooshian_L2.pdf.

Stanski HR, W. L., Wilsom, L.J., Burrows, W.R. (1989). Survey of common verification methods in meteorology. WMO World Weather Watch Technical Report No.8

Stidd, C. (1973). Estimating the precipitation climate. *Water Resources Research*, 9:1235-1241.

Stock, B., & Martell, D. (2016). Forest fire management expenditures in Canada: 1970-2013. *The Forestry Chronicle*, 92, 298-306.

Stocks, B., Lawson, B., Alexander, M., Van Wagner, C., McAlpine, R., Lynham, T., & Dubé, D. (1989). The Canadian Forest Fire Danger Rating System: an overview. *The Forestry Chronicle*, 65(6): 450-457.

Stocks, B., Mason, J., Todd, J., Bosch, E., Wotton, B., Amiro, B., . . . Logan, K. (2003). Large forest fires in Canada, 1959–1997. *Journal of Geophysical Research*, 107(D1):. Retrieved from Canadian National Fire Database: http://cwfis.cfs.nrcan.gc.ca/ha/nfdb

Stone. (1974). Cross-validation choice and assessment of statistical predictions. *Journal of the Royal Statistical Society.Series B (Methodological)*,36(2): 111-147.

Strong, W., & Leggar, K. (1992). Ecoregions of Alberta. Alberta Forestry, Lands and Wildlife, Pub. No. T/245. Map at 1:1,000,000.

Tait, A., Henderson, R., Turner, R., & Zheng, X. (2006). Thin plate smoothing spline interpolation of daily rainfall for nNew Zealand suing a climatological rainfall surface. *International Journal of Climatology*, 26: 2097–2115.

Taylor, S., & Alexander, M. (2006). Science, technology, and human factors in fire danger rating: the Canadian experience. *International Journal of Wildland Fire*, 15, 121-135.

Turner, J., & Lawson, B. (1978). Weather in the Canadian forest fire danger rating system. A user guide to national standards and practices. Victoria, BC. Information Report BC-X-177. 40 p.: Fisheries and Environment Canada, Canadian Forest Service, Pacific Forest Research Centre.

Tymstra, C., Bryce, R., Wotton, B., Taylor, S., & Armitage, O. (2010). Development and Structure of Prometheus: the Canadian Wildland Fire Growth Simulation Model. Edmonton, Alberta. 88pp: Nature Resource Canada, Canadian Forestry Service, Northern Forestry Centre.

Van Wagner CE. (1987). Development and Structure of the Canadian Forest Fire Weather Index System. Forestry Technical Report 35. Canadian Forestry Service, Ottawa, ON.

Wahba, G. (1990). Spline Models for Observational Data. *CBMSNSF Regional Conference Series in Applied Mathematics* (Vol. 59, 169pp). Society for Industrial and Applied Mathematics.

Wang, X., Parisien, M.-A., Flannigan, M., Parks, S., Anderson, K., Little, J., & Tylot, S. (2014). The potential and realised spread of wildfires across Canada. *Global Change Biology*, 20(8):2518-2530.DOI: 10.1111/gcb.12590.

Wang, X., Thompson, D., Marshall, G., Tymstra, C., Carr, R., & Flannigan, M. (2015). Increasing frequency of extreme fire weather in Canada with climate change. *Climatic Change*, 130: 573. doi:10.1007/s10584-015-1375-5.

Wang, X., Wotton, B., Cantin, A., Parisien, M., Anderson, K., Moore, B., & Flannigan, M. (2017). cffdrs: an R package for the Canadian Forest Fire Danger Rating System. *Ecological Processes* , 6:5 DOI 10.1186/s13717-017-0070-z.

Webster, R., & Oliver, M. A. (2007). Geostatistics for Environmental Scientists, 2nd Edition. Chichester: John Wiley&Sons, Ltd.

Wilks, D.S. (1997). Resampling hypothesis tests for autocorrealted fields. *Amerian Meteorological Society*, 65-78.

Wilks, D. (2011). Statistical methods in the atmospheric sciences, 3rd Edition,. *Academic Press (*Vol.100, 704pp*).*

Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error over root mean square error in assessing average model performance. *Climate Research,* 30:79-82.

Wotton, B. (2009). Interpreting and using outputs from the Canadian Forest Fire Danger Rating System in research applications. *Environmental Ecology Stat*, 16: 107-131.

Wotton, B. M., Martell, D., & Logan, K. (2003). Climate change and people caused. *Climatic Change*, 60, 275–295. doi:10.1023/A:1026075919710.

Wotton, B., & Flannigan, M. (1993). Length of fire season in a changing climate. *The Forestry Chronicle*, 69(2):187-192.

Wotton, M., & Martell, D. (2005). A lightning fire occurrence model for Ontario. *Canadian Journal of Forest Research*, 35(6):1389-1401. DOI: 10.1139/x05-071.

Appendix 1: Fire weather stations used in the study

Table A1-1. Summary of names, years of operation, locations, types, elevation, and located regions for weather station used in this study.

Natural region / Station name	Weather station					Regions	
	Years	Location		Station type[1]	Elevation (m)	Validateiton area[2] (75km buffer)	Radar coverage[3] (120km)
		Lat (^{o}N)	Lon (^{o}W)				
Boreal							
Adair	14-15	59.16	118.33	LO	394	yes	no
Algar	14-15	56.16	111.72	LO	751	yes	no
Battle River	14-15	57.17	117.66	LO	674	yes	no
Bison Lake	14-15	57.11	116.51	LO	716	yes	no
Chinchaga	14-15	57.12	118.33	LO	881	yes	no
Clear Hills	14-15	56.60	119.41	LO	1074	no	yes
Conklin	14-15	55.62	111.17	LO	677	no	yes
Calling Lake	14-15	55.21	113.20	RS	610	yes	no
Chisholm	14-15	54.94	114.04	LO	686	yes	no
Cowpar Lake	14-15	55.95	110.50	LO	553	no	no
Cadotte	14-15	56.30	116.43	LO	733	yes	no
Deadwood	14-15	56.64	117.39	LO	648	yes	no
Economy Creek	14-15	54.79	118.23	LO	858	yes	yes
Enilda	14-15	55.32	116.20	LO	717	yes	no
Ells River	14-15	57.21	112.31	LO	573	yes	no
Picadelly Auto	14-15	57.65	114.53	PAW	367	yes	no
Lambert Auto	14-15	58.03	114.14	PAW	368	yes	no
Whitesands Auto	14-15	59.43	114.90	PAW	962	no	no
Foggy Mountain	14-15	58.69	114.99	LO	899	yes	no
Grovedale	14-15	54.98	118.94	RS	671	no	yes
Grande	14-15	56.30	112.23	LO	559	yes	no
Gordon Lake	14-15	56.62	110.49	LO	515	no	no
Gift Lake	14-15	55.90	115.76	LO	691	yes	no
Hines Creek	14-15	56.25	118.61	RS	671	yes	yes
Hawk Hills	14-15	57.66	117.42	LO	730	yes	no
Heart Lake	14-15	54.91	111.34	LO	866	yes	yes
Kimiwan	14-15	55.99	116.61	LO	770	yes	no
Bovine Creek Auto	14-15	56.11	112.55	PAW	518	yes	no
Ground Zero Auto	14-15	55.01	110.61	PAW	737	no	yes
Dunkirk Auto	14-15	56.77	112.49	PAW	479	yes	no
Keane Auto	14-15	58.31	110.28	PAW	337	no	no
Lacorey	14-15	54.45	110.76	RS	579	no	yes
Legend	14-15	57.45	112.88	LO	855	yes	no
Mclennan	14-15	55.71	116.89	RS	625	yes	no
Muskeg Mountain	14-15	57.14	110.89	LO	615	no	no
May	14-15	55.56	112.40	LO	866	yes	no
Otter Lakes	14-15	56.70	115.92	LO	784	yes	no
Basnet Auto	14-15	57.35	119.76	PAW	714	no	no
Puskwaskau	14-15	55.22	117.49	LO	916	yes	no
Red Earth	14-15	56.66	115.11	LO	627	yes	no
Rock Island Lake	14-15	55.33	113.46	LO	722	yes	no
Salteaux Auto	14-15	54.92	114.78	PAW	736	yes	no
Kinuso	14-15	55.30	115.49	PAW	611	yes	no
Loon River	14-15	57.14	115.08	PAW	455	yes	no
Wabasca Auto	14-15	55.96	113.83	PAW	555	yes	no
Chip Alpac	14-15	56.53	113.47	PAW	573	yes	no
Sneddon Creek	14-15	56.20	119.40	RZ	648	no	yes
Salt Prairie	14-15	55.66	115.83	LO	763	yes	no
Sandy Lake	14-15	55.82	113.67	LO	632	yes	no

(Continued)

Natural region / Station name	Weather station					Regions	
	Years	Location		Station type[1]	Elevation (m)	Validation area[2] (75km buffer)	Radar coverage[3] (120km)
		Lat (°N)	Lon (°W)				
Sand River	14-15	54.65	110.98	LO	756	no	yes
Stony Mountain	14-15	56.39	111.28	LO	762	yes	no
Smith	14-15	55.16	114.03	RS	610	yes	no
Vega Auto	14-15	54.43	114.43	PAW	713	no	yes
Valleyview2	14-15	55.35	117.13	RS	700	yes	no
Meekwap Auto	14-15	54.62	116.66	PAW	836	yes	no
Fox Creek Auto	14-15	54.40	116.80	PAW	850	yes	no
Timeu Creek Auto	14-15	54.66	114.60	PAW	730	yes	no
Whitemud	14-15	56.42	118.00	LO	835	yes	yes
Wandering River	14-15	55.20	112.50	RS	579	yes	no
Watt Mountain	14-15	58.65	117.51	LO	793	yes	no
Zama	14-15	58.60	119.17	LO	653	no	no
Buffalo	14	57.95	116.20	LO	801	yes	no
Birch Mountain	14	57.69	111.83	LO	798	yes	no
Doig	14	56.96	119.53	LO	1059	no	no
Emend 05	14	56.81	118.36	TAW	663	yes	no
Ponderosa Auto	14	59.21	117.06	PAW	388	yes	no
Rainbow Lake	14	58.50	119.41	TAW	540	no	no
Fontas	14	57.78	119.50	LO	1055	no	no
Hotchkiss	14	57.33	118.96	LO	990	no	no
Jean Lake	14	57.50	113.88	LO	745	yes	no
Kirby Lake	14	55.36	110.65	LO	691	no	yes
Keg	14	57.64	118.35	LO	950	yes	no
Notikewin	14	56.87	118.60	LO	937	yes	no
Jackpine Auto	14	56.88	116.56	PAW	663	yes	no
Round Hill	14	55.30	112.00	LO	694	yes	no
Trout Mountain	14	56.80	114.42	LO	796	yes	no
Teepee Lake	14	56.46	114.12	LO	782	yes	no
Chipewyan Lakes	15	56.99	113.42	LO	615	yes	no
Bistcho Lake Auto	15	59.91	118.96	PAW	576	no	no
Livock	15	56.46	113.19	LO	650	yes	no
Panny	15	57.18	114.61	LO	625	yes	no
Rocky Lane Agrd	15	58.45	116.48	PAW	301	yes	no
Smoky Lake Agdm	15	54.26	112.50	PAW	675	no	no
Tompkins Landing Agcm	15	58.02	116.85	PAW	343	yes	no
Whitefish	15	56.18	115.47	LO	705	yes	no
Foothills							
Ansell	14-15	53.55	116.50	LO	968	yes	no
Athabasca	14-15	53.41	117.79	LO	1582	no	no
Aurora	14-15	52.65	115.72	LO	1331	yes	no
Yaha Tinda Auto	14-15	51.65	115.36	PAW	1486	no	no
North Ghost	14-15	51.57	114.86	PAW	1477	no	yes
Blackstone	14-15	52.77	116.33	LO	1576	yes	no
Brazeau	14-15	53.04	115.44	LO	1101	yes	yes
Carrot Creek	14-15	53.45	115.87	LO	1056	yes	yes
Deer Mountain	14-15	54.91	115.16	LO	1161	yes	no

(Continued)

Natural region / Station name	Weather station					Regions	
	Years	Location		Station type[1]	Elevation (m)	Validated area[2] (75km buffer)	Radar coverage[3] (120km)
		Lat (^{o}N)	Lon (^{o}W)				
Wildhay Auto	14-15	53.86	117.56	PAW	1041	yes	no
Schwartz Creek Auto	14-15	53.42	116.52	PAW	986	yes	no
Grande Cache	14-15	53.92	118.87	PAW	1250	no	no
Flat Top	14-15	55.15	114.82	LO	1039	yes	no
Kakwa	14-15	54.42	118.98	LO	1230	no	no
Marten Mountain	14-15	55.47	114.79	LO	1005	yes	no
Pinto	14-15	54.78	119.40	LO	1044	no	yes
Clearwater Auto	14-15	51.99	115.24	PAW	1280	no	no
Ram Falls Auto	14-15	52.09	115.84	PAW	1641	no	no
Elk River Auto	14-15	52.97	115.86	PAW	1067	yes	no
Rocky	14-15	52.52	114.88	LO	1106	no	no
Marten Hills Auto	14-15	55.53	114.56	PAW	1000	yes	no
Saddle Hills	14-15	55.62	119.72	LO	967	no	yes
Smoky	14-15	54.40	118.30	LO	1201	yes	no
Snuff Mountain	14-15	54.68	117.54	LO	943	yes	no
Sweathouse	14-15	54.90	116.75	LO	892	yes	no
Tom Hill	14-15	53.93	116.33	LO	1289	yes	no
Freeman Auto	14-15	54.57	115.38	PAW	821	yes	no
Eta Lake Auto	14-15	53.18	115.69	PAW	927	yes	yes
Windfall Auto	14-15	54.19	116.24	PAW	830	yes	no
Berland Hills Auto	14-15	54.15	117.89	PAW	1170	yes	no
White Mountain	14-15	55.69	119.24	LO	1021	no	yes
Berland	14	54.09	117.40	LO	1231	yes	no
Eagle	14	54.46	116.44	LO	1048	yes	no
Goose Mountain	14	54.75	116.03	LO	1385	yes	no
House Mountain	14	55.05	115.60	LO	1176	yes	no
Horse Creek Pb	14	54.02	117.84	TAW	1136	yes	no
Imperial	14	54.47	115.57	LO	1202	yes	no
Lovett	14	53.03	116.66	LO	1477	yes	no
Mayberne	14	53.86	116.67	LO	1453	yes	no
Meridian	14	55.55	114.18	LO	940	yes	no
Pass Creek	14	54.23	116.84	LO	1081	yes	no
Swan Dive	14	54.73	115.36	LO	1241	yes	no
Simonette	14	54.23	118.41	LO	1379	yes	no
Tony	14	54.41	117.49	LO	1007	yes	no
Whitecourt	14	54.03	115.72	LO	1172	yes	yes
Yellowhead	14	53.24	117.14	LO	1477	yes	no
Yellowhead	14	54.09	117.40	LO	1231	yes	no
Bald Mountain	15	54.81	118.92	LO	955	no	yes
Berland Auto	15	54.09	117.40	PAW	1230	yes	no
Eagle Auto	15	54.46	116.44	PAW	1048	yes	no
Imperial Auto	15	54.47	115.57	PAW	1202	yes	no
Pass Creek Auto	15	54.23	116.84	PAW	1081	yes	no
Swan Dive Auto	15	54.73	115.36	PAW	1241	yes	no
Tony Auto	15	54.41	117.49	PAW	1007	yes	no
Whitecourt Auto	15	54.03	115.72	PAW	1172	yes	yes
Rocky Mountain							
Highwood Auto	14-15	50.41	114.73	PAW	1576	no	no
Willow Creek Auto	14-15	50.24	114.35	PAW	1446	no	no
Kananaski Boundary Auto	14-15	50.93	115.12	PAW	1464	no	no
Peter Lougheed Park	14-15	50.71	115.12	PAW	1622	no	no

(Continued)

Natural region / Station name	Weather station					Regions	
	Years	Location		Station type[1]	Elevation (m)	Validation area[2] (75km buffer)	Radar coverage[3] (120km)
		Lat (^{o}N)	Lon (^{o}W)				
Elbow Auto	14-15	50.91	114.69	PAW	1413	no	yes
Blue Hill	14-15	51.71	115.22	LO	1987	no	no
Baseline	14-15	52.13	115.43	LO	1908	no	no
Livingston Gap Auto	14-15	49.88	114.38	PAW	1403	no	no
Castle Auto	14-15	49.39	114.34	PAW	1352	no	no
Poll Haven Auto	14-15	49.03	113.60	PAW	1611	no	no
Blairmore Ranger Station	14-15	49.61	114.45	PAW	1311	no	no
Entrance Auto	14-15	53.38	117.70	PAW	1045	yes	no
Kakwa Auto	14-15	54.18	119.06	PAW	1344	no	no
Ironstone	14-15	49.57	114.50	LO	2079	no	no
Mockingbird Hill	14-15	51.42	115.07	LO	1907	no	yes
Moose Mountain	14-15	50.94	114.84	LO	2431	no	yes
Porcupine	14-15	49.89	114.01	LO	1820	no	no
Kootenay Plains Auto	14-15	52.06	116.41	PAW	1294	no	no
Adams Creek	14	53.73	118.57	LO	2178	no	no
Blackstone Burn	14	52.60	116.63	TAW	1669	no	no
Baldy	14	52.53	116.13	LO	2082	yes	no
Carbondale	14	49.43	114.36	LO	1807	no	no
Cline	14	52.18	116.41	LO	2050	no	no
Copton	14	54.18	119.42	LO	1860	no	no
Grave Flats	14	52.89	116.99	LO	2076	no	no
Hailstone Butte	14	50.21	114.46	LO	2370	no	no
Junction Mountain	14	50.57	114.65	LO	2241	no	yes
Kananaskis	14	50.61	115.07	LO	2129	no	no
Nose Mountain	14	54.56	119.63	LO	1487	no	no
Obed	14	53.57	117.50	LO	1605	yes	no
Raspberry Ridge	14	50.30	114.64	LO	2360	no	no
Sugarloaf	14	49.95	114.54	LO	2516	no	no
Torrens	14	54.31	119.68	LO	1815	no	no
Upper Sask Pb	14	52.01	116.53	TAW	1631	no	no
Barrier Lake	15	51.05	115.08	LO	2021	no	yes

[1] LO: lookout towers, RS: ranger weather stations, PAW: permanent remote automatic weather Stations, TAW: temporary remote automatic weather Stations.

[2] yes: weather station is within the validated area and used as a validation point.

[3] yes: weather station is within the 120km radar effective range.

Note: all the weather stations are operated by Alberta Environment and Sustainable Resource Development

Appendix 2: Interpolation algorithms R codes

```
# Title: Leave-one-out cross-validation of interpolation methods
# Date: Mach 21, 2017
# Author: Xinli Cai
```

```
# Load required libraries
library(fields)
library(gstat)
library(sp)
library(rgdal)
library(rgeos)
library(automap)
library(foreign)

# Clean workspace and set working directory
rm(list=ls())
path <- c("E:/1 Data/AB ws/2014&2015_weather_data")
setwd(path)

# Import weather data (Here I used 2015 data as an example)
all_wstn<- read.csv("all_weather_data_2015.csv")
val_wstn<- read.csv("validated_weather_data_2015.csv")
head(val_wstn)

##   id      lat       lon  jd     date temp    rh ws  wd prec capa
## 1 AD 59.16063 -118.3295 124 5/4/2015  5.0 47.25  0   0    0 0.01
## 2 AD 59.16063 -118.3295 125 5/5/2015  8.0 52.96  6 315    0 0.00
## 3 AD 59.16063 -118.3295 126 5/6/2015 11.0 29.32  5 270    0 0.00
## 4 AD 59.16063 -118.3295 127 5/7/2015 11.5 35.06  8 360    0 0.00
## 5 AD 59.16063 -118.3295 128 5/8/2015 14.0 35.80  9 360    0 0.53
## 6 AD 59.16063 -118.3295 129 5/9/2015 15.0 45.72 10 180    0 0.08
```

id and jd columns uniquely specify each row. Here, prec is the observed daily precipitation at each AAF station while capa is the precipitation estimated by CaPA system. Data have been cleaned using procedures described in Section 2.2.1.

Thin-plate spline

Calculations of TPS-related methods were using fields package following Douglas et al., 2016. Here, I displayed the leave-one out cross-validation (LOOCV) procedure for smoothed / non-smoothed TPS, and square root transformed TPS algorithms. The following function performs the TPS related methods for multiple weather stations over a period.

```
tps_loocv <- function(df_valid, df_fit, method){
    #df_valid <- val_wstn (weather stations in the validation area, see Figure 2.4)
    #df_fit <- all_wstn (all the weather stations in the study area, see Figure 2.4)

    days<- unique(df_valid$jd) # n=129
    id <- unique(df_valid$id) # n=81

    # Create a new dataframe to store the interpolated values
```

```
    foo <- as.data.frame(matrix(ncol=length(days)+1,nrow=length(id)))
    names(foo)<-c("id",paste("jd",days,sep="_"))
    foo$id<-id

    # Loop through all the days and weather stations
      for(i in 1:length(days)){
       for(j in 1:length(id)){

        # LOOCV:remove one weather station at a time
        test<- df_fit[df_fit$jd == days[i] & ! df_fit$id == id[j], ]

        # TPS interpolation algorithms

        # (1) Smoothed TPS: Smoothing parameter lambda is determined by GCV

        if(method=="tps_s"){
        spl <- Tps(as.matrix(test[,c("lon","lat")]), test[,"prec"])
        intpltd <-  predict(spl, df_valid[df_valid$jd==days[i]&df_valid$id==id[j], c
("lon","lat")])
          if(intpltd<0|is.na(intpltd)){intpltd <- 0}
            foo[j,i+1] <- intpltd
             }

        # (2) Non-smoothed TPS: lambda =0 (no smoothness constraints)

        if(method=="tps_ns"){
        spl <- Tps(as.matrix(test[,c("lon","lat")]), test[,"prec"], lambda=0)
        intpltd <- predict(spl, df_valid[df_valid$jd==days[i]&df_valid$id==id[j], c("
lon","lat")])
           if(intpltd<0|is.na(intpltd)){intpltd <- 0}
              foo[j,i+1] <- intpltd
             }
          }

        # (3)Smoothed TPS with square root transforming of precipitation observations
        # Transformation is applied before the interpolation and a back-transformatio
n is applied to the estimates

        if(method=="tps_s_sqrt"){
        spl <- Tps(as.matrix(test[,c("lon","lat")]), sqrt(test[,"prec"]))
        intpltd <-  predict(spl, df_valid[df_valid$jd==days[i]&df_valid$id==id[j], c
("lon","lat")])
           if(intpltd<0|is.na(intpltd)){intpltd <- 0}
             foo[j,i+1] <- intpltd^2
              }

    # Reshape the outputs

    output <- reshape(foo, idvar="id", varying = list(2:ncol(foo)), v.names=method, d
irection="long")
    output$time <- rep(days, each=length(id))
     } # i
    return(output)
}
```

Kriging

Calculations of Kriging related methods were using gstat package following Pebesma 2004. I display the algorithms of ordinary kriging and regression kriging.The function below performs LOOCV procedure using kriging related algorithms for multiple weather stations over a period of time.

```
Kriging_loocv<- function(df_fit, df_valid, method){
    #df_valid <- val_wstn (weather stations in the validation area, see Figure 2.4)
    #df_fit <- all_wstn (all the weather stations in the study area, see Figure 2.4)

    days<- unique(df_valid$jd)
    id <- unique(df_valid$id)
    id <-  factor(id, levels=levels(df_fit$id))

   # Create a new dataframe to store the interpolated values
    foo <- as.data.frame(matrix(ncol=(length(days)+1),nrow=length(id)))
    names(foo)<-c("id",paste("jd",days,sep="_"))
    foo$id <- id

    # Loop through all the days
    for(i in 1:length(days)){

    # For each day,converting the dataframe to spatialpointsdataframe
    # Spatial reference system used is EPSG:3402,which is centred in Alberta.

      test <- df_fit[df_fit$jd==days[i], ]
      coordinates(test) = ~lon+lat
      proj4string(test) <- "+proj=longlat +ellps=GRS80"
      test<- spTransform(test, CRS("+init=epsg:3402"))

    # Perform LOOCV for all the weather stations

     for(j in 1:length(id)){
     # Removed weather station at a time
     valid <- test[test@data$id==id[j], ]
     # Use the reminder stations to built the semi-variogram
     fit <- test[!test@data$id==id[j], ]

     # For dry days (no rain events were observed at all the stations), setting inter
polated values to zero. Because variogram can not be fitted when observed precipitati
on is unique for all the stations.

     if(length(unique(fit$prec))==1){
          foo[j,i+1] <- unique(fit$prec)
                  }
     # For locations where there is no CaPA estimated value, do not perform the inter
polation.Because regression kriging needs valid CaPA estimates for each removed stati
on.
      else if(unique(is.na((fit@data$capa)))){
                                foo[, i+1] <- NA
                }
      else{
```

```
       # Ordinary kriging: semi-variogram is fitted with a spherical model

       if(method=="ok"){
        m <- autofitVariogram(prec~1, fit, model="Sph")
        plot(m)
        ok <- krige(prec~1, fit, valid, model = m$var_model)
         if(ok@data$var1.pred<0|is.na(ok@data$var1.pred)){ok@data$var1.pred <- 0}
            foo[j,i+1] <- ok@data$var1.pred
            }

       # Regression kriging using CaPA estimates as covariate: semi-variogram is fitt
ed with a spherical model
        if(method=="rk"){
         m <- autofitVariogram(prec~capa, fit, model="Sph")
         plot(m)
         rk <- krige(prec~capa, fit, valid, model = m$var_model)
          if(rk@data$var1.pred<0|is.na(rk@data$var1.pred)){rk@data$var1.pred <- 0}
            foo[j,i+1] <- rk@data$var1.pred
                }
             }
         } # j: days
    } # i:stations

    # Reshape the output
    output <- reshape(foo, idvar="id", varying = list(2:ncol(foo)), v.names=method, d
irection="long")
    output$time <- rep(days, each=length(id))
    return(output)
}
```

Call the functions

```
# Smoothed TPS
tps_s <- tps_loocv(df_valid = val_wstn, df_fit = all_wstn, method="tps_s")
# Non-smoothed TPS
tps_ns <- tps_loocv(df_valid = val_wstn, df_fit = all_wstn, method="tps_ns")
# Smoothed TPS with sqaure root transforming
tps_s_sqrt <- tps_loocv(df_valid = val_wstn, df_fit = all_wstn, method="tps_s_sqrt")
# Ordinary kriging
ok <- Kriging_loocv(df_valid = val_wstn, df_fit = all_wstn, method="ok")
# Regression kriging
rk <- Kriging_loocv(df_valid = val_wstn, df_fit = all_wstn, method="rk")
```

Appendix 3: Resampling post-hoc tests outputs

In this study, a resampling post-hoc test procedure that combines a resampling paired t-test (Hamill 1999) and Holm-Bonferroni p-value adjustment (Holm 1979), was applied to identify statistical differences between estimates of precipitation and FWI System indexes. In Appendix 3, we presented the outputs of resampling post-hoc tests applied to precipitation and FWI System estimates were also presented.

Table A3-1. The resampling post-hoc test outputs of PCP estimates in 2014 and 2015 over validation area. The PCP estimates are evaluated using MAE (mm).

		MAE (mm) averaged over all stations in 2014 and 2015 (test statistics)									
	Method	ok	idw	tps _ns	tps _s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk (capa)_s	ok_s
	2014	*1.47*	*1.42*	*1.40*	*1.35*	*1.34*	*1.33*	*1.28*	*1.16*	*1.16*	*1.11*
	2015	*1.30*	*1.29*	*1.27*	*1.23*	*1.22*	*1.20*	*1.16*	*1.07*	*1.03*	*1.02*
Year	Method	ok	idw	tps _ns	tps _s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk (capa)_s	ok_s
2014	ok	–	1.00	1.000	0.858	0.767	0.858	0.264	0.099	0.130	0.076
	idw		–	1.000	0.414	0.414	0.767	0.130	**0.000**	0.076	0.076
	tps_ns			–	0.416	0.560	1.000	0.076	0.099	0.112	0.043
	tps _s				–	1.000	1.000	0.076	0.043	**0.000**	0.099
	rk(capa)					–	1.000	0.285	0.043	0.043	0.043
	CaPA						–	1.000	0.280	0.240	0.240
	tps_ns_s							–	0.112	0.099	0.099
	tps_s_s								–	1.000	0.264
	rk(capa)_s									–	0.480
	ok_s										–
Year	Method	ok	idw	tps _ns	tps _s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk (capa)_s	ok_s
2015	ok	–	1.00	1.000	0.247	0.144	0.247	0.060	**0.000**	**0.000**	**0.000**
	idw		–	1.000	0.387	0.300	**0.022**	0.085	**0.000**	**0.000**	**0.000**
	tps_ns			–	0.144	0.154	0.247	**0.000**	**0.000**	**0.000**	**0.000**
	tps _s				–	1.000	0.680	**0.022**	**0.000**	**0.000**	**0.000**
	rk(capa)					–	1.000	0.060	**0.000**	**0.000**	**0.000**
	CaPA						–	0.680	**0.000**	**0.000**	**0.000**
	tps_ns_s							–	**0.000**	**0.000**	**0.000**
	tps_s_s								–	0.060	**0.000**
	rk(capa)_s									–	1.000
	ok_s										–

Bold: comparison achieved a statistical significance at α level of 0.05. The method had better / worse performance than the methods on its left / right side.

Table A3-2. The resampling post-hoc test outputs of PCP estimates in 2015 over validation area. The PCP estimates are evaluated using ETS.

		ETS averaged over all stations in 2015 (test statistics)									
	PCP (mm)	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
	0.5 – 1.5	*0.06*	*0.06*	*0.08*	*0.09*	*0.06*	*0.09*	*0.09*	*0.11*	***0.12***	***0.12***
	1.5 – 2.8	*0.06*	*0.08*	*0.07*	*0.07*	*0.08*	*0.07*	*0.09*	*0.10*	***0.11***	*0.10*
	> 2.8	*0.42*	*0.44*	*0.43*	*0.44*	*0.46*	*0.42*	*0.46*	*0.47*	***0.49***	*0.48*
PCP	Method	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
0.5 – 1.5	ok	–	1.000	0.826	0.155	1.000	0.264	0.450	0.037	**0.017**	**0.000**
	idw		–	0.380	0.070	1.000	0.070	0.378	**0.000**	**0.000**	**0.000**
	tps_ns			–	1.000	0.858	1.000	1.000	0.099	0.099	**0.000**
	tps_s				–	0.174	1.000	1.000	0.735	0.322	0.155
	rk (capa)					–	0.189	0.418	**0.000**	**0.000**	**0.000**
	CaPA						–	1.000	1.000	0.374	0.174
	tps_ns_s							–	0.450	0.208	0.250
	tps_s_s								–	1.000	0.528
	rk(capa)_s									–	1.000
	ok_s										–
1.5 – 2.8	ok	–	0.899	1.000	1.000	0.780	1.000	0.234	0.123	**0.000**	0.084
	idw		–	1.000	1.000	1.000	1.000	1.000	1.000	0.544	1.000
	tps_ns			–	1.000	1.000	1.000	0.660	0.407	0.044	0.704
	tps_s				–	1.000	1.000	0.988	0.432	0.234	0.980
	rk (capa)					–	1.000	1.000	1.000	0.160	1.000
	CaPA						–	1.000	0.490	0.044	0.704
	tps_ns_s							–	1.000	0.988	1.000
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	0.980
	ok_s										–
> 2.8	ok	–	1.000	1.000	1.000	0.696	1.000	1.000	0.513	**0.000**	0.105
	idw		–	1.000	1.000	1.000	1.000	1.000	0.076	0.041	0.041
	tps_ns			–	1.000	0.736	1.000	0.105	0.076	**0.000**	**0.000**
	tps_s				–	1.000	1.000	1.000	0.128	0.076	0.041
	rk (capa)					–	0.174	1.000	1.000	0.150	1.000
	CaPA						–	0.448	0.128	**0.000**	0.105
	tps_ns_s							–	1.000	0.513	0.650
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	1.000
	ok_s										–

Equitable threat score (ETS): values range between -0.333 to 1, where 1 is perfect skill.
Bold: comparison achieved a statistical significance at α level of 0.05.

Table A3-3. The resampling post-hoc test outputs of PCP estimates in 2015 over validation area. The PCP estimates are evaluated using FBI.

		FBI averaged over all stations in 2015 (test statistics)									
	PCP (mm)	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
	0.5 – 1.5	*1.17*	*1.35*	*0.25*	*0.63*	*1.28*	*0.48*	***0.17***	*0.35*	*0.37*	*0.47*
	1.5 – 2.8	*1.34*	*1.15*	*0.21*	*0.59*	*0.67*	*0.17*	***0.05***	*0.10*	*0.11*	*0.07*
	> 2.8	*0.05*	*0.06*	*0.08*	*0.08*	***0.03***	*(0.20)*	*(0.06)*	*(0.16)*	*(0.19)*	*(0.20)*
PCP	Method	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
0.5 – 1.5	ok	–	0.888	0.048	0.114	1.000	**0.000**	0.030	0.030	0.030	0.153
	idw		–	**0.000**	0.030	1.000	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	tps_ns			–	**0.000**	0.030	0.048	0.476	0.594	0.533	0.533
	tps_s				–	0.060	0.580	**0.000**	0.030	0.144	0.888
	rk (capa)					–	0.048	**0.000**	**0.000**	**0.000**	0.048
	CaPA						–	**0.000**	0.240	0.533	1.000
	tps_ns_s							–	**0.000**	**0.000**	0.160
	tps_s_s								–	1.000	0.752
	rk(capa)_s									–	0.819
	ok_s										–
1.5 – 2.8	ok	–	1.000	**0.000**	0.060	0.170	**0.000**	**0.024**	**0.024**	**0.000**	**0.000**
	idw		–	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	tps_ns			–	0.090	**0.024**	1.000	**0.000**	0.988	0.645	0.496
	tps_s				–	1.000	0.060	**0.000**	**0.000**	**0.000**	**0.000**
	rk (capa)					–	**0.024**	**0.000**	**0.000**	**0.000**	**0.000**
	CaPA						–	0.645	1.000	1.000	1.000
	tps_ns_s							–	1.000	1.000	1.000
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	1.000
	ok_s										–
> 2.8	ok	–	1.000	1.000	1.000	1.000	0.066	0.480	0.066	**0.000**	**0.000**
	idw		–	1.000	1.000	1.000	0.046	0.368	0.028	**0.000**	**0.000**
	tps_ns			–	1.000	0.624	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	tps_s				–	0.162	**0.000**	0.028	0.028	**0.000**	**0.000**
	rk (capa)					–	**0.024**	0.076	**0.000**	**0.000**	**0.000**
	CaPA						–	0.028	1.000	1.000	1.000
	tps_ns_s							–	0.066	**0.000**	**0.000**
	tps_s_s								–	0.480	0.170
	rk(capa)_s									–	0.780
	ok_s										–

Frequency bias index (FBI): adjusted from -1 to infinity, where 0 is perfect skill. Positive/negative FBI means overestimate/underestimate **Bold:** comparison achieved a statistical significance at α level of 0.05.

Table A3-4. The resampling post-hoc test outputs of PCP estimates 2015 validation radar and validation non-radar area. The PCP estimates are evaluated using MAE (mm).

		MAE averaged over stations in validation radar / non-radar area (test statistics)									
	Method	ok	idw	tps_ns	tps_s	rk (capa)	CaPA	tps_ns_s	tps_s_s	rk(capa)_s	ok_s
	Non-radar	*1.30*	*1.29*	*1.25*	*1.25*	*1.24*	*1.22*	*1.15*	*1.06*	*1.03*	*1.01*
	Radar	*1.37*	*1.39*	*1.47*	*1.39*	*1.17*	*1.12*	*1.36*	*1.25*	*1.05*	*1.18*

Region	Method	ok	idw	tps_ns	tps_s	rk (capa)	CaPA	tps_ns_s	tps_s_s	rk(capa)_s	ok_s
Non-radar	ok	–	0.945	0.864	0.320	0.320	0.330	**0.000**	**0.000**	**0.000**	**0.000**
	idw		–	0.864	0.320	0.513	0.198	0.040	**0.000**	**0.000**	**0.000**
	tps_ns			–	0.320	0.945	0.864	**0.000**	**0.000**	**0.000**	**0.000**
	tps_s				–	0.945	0.945	**0.024**	**0.000**	**0.000**	**0.000**
	rk (capa)					–	0.945	0.040	**0.024**	**0.000**	**0.000**
	CaPA						–	0.289	**0.024**	**0.000**	**0.000**
	tps_ns_s							–	**0.000**	**0.000**	**0.000**
	tps_s_s								–	0.320	**0.024**
	rk(capa)_s									–	0.350
	ok_s										–

Region	Method	ok	idw	tps_ns	tps_s	rk (capa)	CaPA	tps_ns_s	tps_s_s	rk(capa)_s	ok_s
Radar	ok	–	1.000	0.798	1.000	0.032	0.060	1.000	0.060	**0.000**	**0.000**
	idw		–	1.000	1.000	0.075	**0.000**	1.000	0.240	**0.000**	**0.000**
	tps_ns			–	0.798	**0.000**	0.110	**0.000**	0.060	**0.000**	**0.000**
	tps_s				–	0.032	0.075	1.000	**0.000**	**0.000**	**0.000**
	rk (capa)					–	1.000	0.306	0.968	**0.000**	1.000
	CaPA						–	0.270	0.798	0.640	1.000
	tps_ns_s							–	0.240	0.060	0.110
	tps_s_s								–	0.075	0.640
	rk(capa)_s									–	0.060
	ok_s										–

Bold: comparison achieved a statistical significance at α level of 0.05. The method had better / worse performance than the methods on its left / right side for validation non-radar.

Table A3-5. The resampling post-hoc test outputs of PCP estimates in 2015 **validation non-radar** area. The PCP estimates are evaluated using **ETS**.

ETS averaged over stations in validation r non-radar area (test statistics)

PCP (mm)	ok	idw	tps_ns	tps_s	rk (capa)	CaPA	tps_ns_s	tps_s_s	rk(capa)_s	ok_s
0.5 – 1.5	*0.06*	*0.06*	*0.08*	*0.09*	*0.60*	*0.09*	*0.09*	*0.10*	*0.11*	***0.12***
1.5 – 2.8	*0.06*	*0.08*	*0.07*	*0.07*	*0.08*	*0.06*	*0.09*	*0.09*	***0.10***	***0.10***
> 2.8	*0.43*	*0.43*	*0.43*	*0.43*	*0.46*	*0.42*	*0.46*	*0.48*	***0.49***	***0.49***

PCP	Method	ok	idw	tps_ns	tps_s	rk (capa)	CaPA	tps_ns_s	tps_s_s	rk(capa)_s	ok_s
0.5 – 1.5	ok	–	1.000	0.784	0.232	1.000	0.462	0.608	0.038	0.102	**0.000**
	idw		–	0.462	0.038	1.000	0.232	0.416	**0.000**	**0.000**	**0.000**
	tps_ns			–	1.000	0.738	1.000	1.000	0.102	0.150	**0.000**
	tps_s				–	0.102	1.000	1.000	0.949	0.738	0.124
	rk (capa)					–	0.416	0.416	**0.000**	0.038	**0.000**
	CaPA						–	1.000	0.784	0.416	0.070
	tps_ns_s							–	0.784	0.500	0.351
	tps_s_s								–	1.000	0.949
	rk(capa)_s									–	1.000
	ok_s										–
1.5 – 2.8	ok	–	1.000	1.000	1.000	1.000	1.000	0.735	0.440	0.045	0.264
	idw		–	1.000	1.000	1.000	1.000	1.000	1.000	0.585	1.000
	tps_ns			–	1.000	1.000	1.000	0.837	0.648	0.369	0.832
	tps_s				–	1.000	1.000	0.585	0.816	0.301	0.930
	rk (capa)					–	1.000	1.000	1.000	0.816	1.000
	CaPA						–	1.000	0.957	0.301	0.592
	tps_ns_s							–	1.000	1.000	1.000
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	1.000
	ok_s										–
> 2.8	ok	–	1.000	1.000	1.000	1.000	1.000	1.000	0.702	**0.000**	0.132
	idw		–	1.000	1.000	1.000	1.000	0.702	0.132	**0.000**	0.041
	tps_ns			–	1.000	1.000	1.000	0.076	0.041	0.041	**0.000**
	tps_s				–	1.000	1.000	1.000	0.132	0.076	0.076
	rk (capa)					–	0.180	1.000	1.000	0.203	1.000
	CaPA						–	0.280	0.105	**0.000**	0.105
	tps_ns_s							–	1.000	1.000	0.702
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	1.000
	ok_s										–

Equitable threat score (ETS): values range between -0.333 to 1, where 1 is perfect skill.
Bold: comparison achieved a statistical significance at α level of 0.05.

Table A3-6. The resampling post-hoc test outputs of PCP estimates in 2015 **validation radar area.** The PCP estimates are evaluated using **ETS**.

		ETS averaged over stations in validation radar area (test statistics)									
PCP (mm)		ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
0.5 – 1.5		*0.05*	*0.05*	*0.06*	*0.06*	*0.07*	*0.14*	*0.06*	*0.09*	***0.17***	*0.13*
1.5 – 2.8		*0.05*	*0.10*	*0.05*	*0.11*	*0.10*	*0.11*	*0.08*	*0.13*	***0.20***	*0.12*
> 2.8		*0.42*	*0.45*	*0.39*	*0.39*	*0.50*	*0.45*	*0.42*	*0.45*	***0.52***	*0.43*
PCP	Method	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
0.5 – 1.5	ok	–	1.000	1.000	1.000	1.000	0.288	1.000	1.000	**0.000**	0.363
	idw		–	1.000	1.000	1.000	0.084	1.000	0.891	**0.000**	0.350
	tps_ns			–	1.000	1.000	0.185	1.000	1.000	0.152	0.608
	tps_s				–	1.000	0.350	1.000	1.000	0.123	0.754
	rk (capa)					–	0.682	1.000	1.000	0.043	0.754
	CaPA						–	0.682	1.000	1.000	1.000
	tps_ns_s							–	1.000	0.123	0.891
	tps_s_s								–	0.123	1.000
	rk(capa)_s									–	1.000
	ok_s										–
1.5 – 2.8	ok	–	1.000	1.000	0.646	1.000	1.000	1.000	0.344	0.090	1.000
	idw		–	1.000	1.000	1.000	1.000	1.000	1.000	0.378	1.000
	tps_ns			–	1.000	1.000	1.000	1.000	0.624	0.264	1.000
	tps_s				–	1.000	1.000	1.000	1.000	0.520	1.000
	rk (capa)					–	1.000	1.000	1.000	0.851	1.000
	CaPA						–	1.000	1.000	1.000	1.000
	tps_ns_s							–	1.000	0.864	1.000
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	0.492
	ok_s										–
> 2.8	ok	–	1.000	1.000	1.000	0.672	1.000	1.000	1.000	0.164	1.000
	idw		–	0.810	0.444	1.000	1.000	1.000	1.000	0.627	1.000
	tps_ns			–	1.000	0.273	1.000	1.000	0.544	0.088	1.000
	tps_s				–	0.342	1.000	1.000	**0.000**	0.088	0.928
	rk (capa)					–	0.972	0.468	1.000	1.000	1.000
	CaPA						–	1.000	1.000	0.490	1.000
	tps_ns_s							–	0.928	0.088	1.000
	tps_s_s								–	0.806	1.000
	rk(capa)_s									–	0.240
	ok_s										–

Equitable threat score (ETS): values range between -0.333 to 1, where 1 is perfect skill.
Bold: comparison achieved a statistical significance at α level of 0.05.

Table A3-7. The resampling post-hoc test outputs of PCP estimates in 2015 **validation non-radar area**. The PCP estimates are evaluated using **FBI.**

	FBI averaged over stations in validation non-radar area (test statistics)									
PCP (mm)	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
0.5 – 1.5	*1.20*	*1.3*	*0.26*	*0.65*	*1.32*	*0.50*	***0.17***	*0.35*	*0.38*	*0.49*
1.5 – 2.8	*1.35*	*1.17*	*0.21*	*0.57*	*0.69*	*0.17*	***0.04***	*0.08*	*0.12*	*0.07*
> 2.8	*0.06*	*0.06*	*0.08*	*0.08*	***0.04***	*(0.17)*	*(0.06)*	*(0.16)*	*(0.19)*	*(0.21)*

PCP	Method	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
0.5 – 1.5	ok	–	0.995	0.028	0.028	0.995	0.063	0.028	0.028	**0.000**	0.095
	idw		–	**0.000**	**0.000**	0.995	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	tps_ns			–	**0.000**	0.028	**0.000**	0.240	0.832	0.516	0.507
	tps_s				–	0.063	0.760	**0.000**	0.095	0.112	0.948
	rk (capa)					–	0.046	**0.000**	**0.000**	**0.000**	**0.000**
	CaPA						–	**0.000**	0.252	0.583	0.995
	tps_ns_s							–	**0.000**	0.046	0.102
	tps_s_s								–	0.995	0.760
	rk(capa)_s									–	0.847
	ok_s										–
1.5 – 2.8	ok	–	1.000	**0.000**	0.076	0.136	**0.024**	**0.000**	**0.000**	**0.000**	**0.000**
	idw		–	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	tps_ns			–	0.060	**0.000**	1.000	**0.024**	0.504	0.888	0.480
	tps_s				–	1.000	0.076	0.042	**0.000**	**0.024**	**0.000**
	rk (capa)					–	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	CaPA						–	0.495	1.000	1.000	0.990
	tps_ns_s							–	1.000	0.871	1.000
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	1.000
	ok_s										–
> 2.8	ok	–	1.000	1.000	1.000	1.000	0.140	0.324	0.092	0.054	0.054
	idw		–	1.000	1.000	1.000	0.140	0.432	**0.000**	0.030	**0.000**
	tps_ns			–	1.000	1.000	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	tps_s				–	0.480	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	rk (capa)					–	0.054	0.105	**0.000**	0.030	**0.000**
	CaPA						–	0.092	1.000	1.000	1.000
	tps_ns_s							–	0.072	0.030	**0.000**
	tps_s_s								–	1.000	0.324
	rk(capa)_s									–	0.854
	ok_s										–

Frequency bias index (FBI): adjusted from -1 to infinity, where 0 is perfect skill. Positive/negative FBI means overestimate/underestimate

Bold: comparison achieved a statistical significance at α level of 0.05.

Table A3-8. The resampling post-hoc test outputs of PCP estimates in 2015 **validation radar area.** The PCP estimates are evaluated using **FBI.**

		FBI averaged over stations in validation radar area (test statistics)									
	PCP (mm)	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
	0.5 – 1.5	*0.92*	*1.37*	*0.18*	*0.49*	*0.93*	*0.33*	***0.14***	*0.31*	*0.27*	*0.36*
	1.5 – 2.8	*1.30*	*1.02*	*0.20*	*0.73*	*0.55*	*0.09*	*0.14*	*0.32*	***0.03***	*0.09*
	> 2.8	*(0.08)*	*0.06*	*0.12*	*0.07*	*(0.06)*	*(0.33)*	***(0.03)***	*(0.16)*	*(0.26)*	*(0.27)*
PCP	Method	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
0.5 – 1.5	ok	–	1.000	0.272	0.672	1.000	0.319	0.310	0.272	0.405	0.782
	idw		–	**0.000**	0.114	1.000	0.041	**0.000**	**0.000**	**0.000**	0.114
	tps_ns			–	0.310	0.041	1.000	1.000	1.000	1.000	1.000
	tps_s				–	0.625	1.000	0.336	1.000	1.000	1.000
	rk (capa)					–	0.210	0.041	0.144	0.272	0.405
	CaPA						–	1.000	1.000	1.000	1.000
	tps_ns_s							–	1.000	1.000	1.000
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	1.000
	ok_s										–
1.5 – 2.8	ok	–	1.000	0.224	1.000	0.300	0.180	0.084	0.180	0.084	0.043
	idw		–	0.120	1.000	0.425	0.120	0.120	0.204	**0.000**	**0.000**
	tps_ns			–	0.364	1.000	1.000	1.000	1.000	1.000	1.000
	tps_s				–	1.000	0.378	0.390	0.480	0.204	0.120
	rk (capa)					–	0.798	0.644	1.000	0.319	0.248
	CaPA						–	1.000	1.000	1.000	1.000
	tps_ns_s							–	1.000	1.000	1.000
	tps_s_s								–	1.000	0.770
	rk(capa)_s									–	1.000
	ok_s										–
> 2.8	ok	–	0.408	0.189	0.280	1.000	0.096	1.000	0.825	0.081	0.081
	idw		–	1.000	1.000	0.847	0.081	0.900	0.115	**0.000**	**0.000**
	tps_ns			–	1.000	0.306	**0.000**	**0.000**	0.032	**0.000**	**0.000**
	tps_s				–	0.825	**0.000**	0.825	0.032	**0.000**	0.032
	rk (capa)					–	**0.000**	1.000	0.825	**0.000**	**0.000**
	CaPA						–	**0.000**	0.058	0.847	1.000
	tps_ns_s							–	0.285	0.058	**0.000**
	tps_s_s								–	0.432	0.132
	rk(capa)_s									–	1.000
	ok_s										–

Frequency bias index (FBI): adjusted from -1 to infinity, where 0 is perfect skill. Positive/negative FBI means overestimate/underestimate

Bold: comparison achieved a statistical significance at α level of 0.05.

Table A3-9 The resampling post-hoc test outputs of PCP estimates 2015 **validation Boreal and validation Foothills area**. The PCP estimates are evaluated using MAE (mm).

	MAE averaged over stations in validation Boreal / Foothills area (test statistics)									
Method	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
Boreal	*1.29*	*1.26*	*1.21*	*1.18*	*1.18*	*1.13*	*1.10*	*1.01*	*0.97*	*0.96*
Foothills	*1.34*	*1.35*	*1.38*	*1.34*	*1.30*	*1.34*	*1.28*	*1.19*	*1.13*	*1.13*
Percent change (%)	*(4.4)*	*(7.4)*	*(12.9)*	*(8.5)*	*(9.9)*	*(16.3)*	*(16.2)*	*(14.7)*	*(14.9)*	*(15.8)*

Region	Method	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
Boreal	ok	–	0.554	0.384	0.120	0.068	**0.020**	**0.000**	**0.000**	**0.000**	**0.000**
	idw		–	0.534	0.190	0.156	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	tps_ns			–	0.165	0.534	0.126	**0.000**	**0.000**	**0.000**	**0.000**
	tps_s				–	0.554	0.384	**0.000**	**0.000**	**0.000**	**0.000**
	rk (capa)					–	0.378	0.054	**0.000**	**0.000**	**0.000**
	CaPA						–	0.534	0.038	**0.000**	**0.000**
	tps_ns_s							–	**0.000**	**0.000**	**0.000**
	tps_s_s								–	0.143	0.080
	rk(capa)_s									–	0.534
	ok_s										–

Region	Method	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
Foot-hills	ok	–	1.000	1.000	1.000	1.000	1.000	1.000	**0.000**	**0.000**	**0.000**
	idw		–	1.000	1.000	1.000	1.000	0.950	0.029	**0.000**	**0.000**
	tps_ns			–	1.000	0.528	1.000	**0.000**	**0.000**	**0.000**	**0.000**
	tps_s				–	1.000	1.000	0.756	**0.000**	**0.000**	**0.000**
	rk (capa)					–	1.000	1.000	0.052	**0.000**	**0.000**
	CaPA						–	1.000	0.144	0.029	**0.000**
	tps_ns_s							–	0.075	0.029	**0.000**
	tps_s_s								–	0.880	0.299
	rk(capa)_s									–	1.000
	ok_s										–

Bold: comparison achieved a statistical significance at α level of 0.05. The method had better / worse performance than the methods on its left / right side for validation Boreal area.

Table A3-10. The resampling post-hoc test outputs of PCP estimates in 2015 **validation Boreal.** The PCP estimates are evaluated using **ETS.**

		ETS averaged over stations in validation Boreal (test statistics)									
PCP (mm)		ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
0.5 – 1.5		*0.06*	*0.05*	*0.07*	*0.09*	*0.05*	*0.10*	*0.08*	*0.10*	*0.10*	***0.11***
1.5 – 2.8		*0.05*	*0.07*	*0.06*	*0.06*	*0.07*	*0.06*	*0.08*	*0.08*	***0.10***	*0.09*
> 2.8		*0.40*	*0.41*	*0.40*	*0.42*	*0.44*	*0.39*	*0.43*	*0.45*	***0.46***	***0.46***
PCP	Method	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
0.5 – 1.5	ok	–	1.000	1.000	0.330	1.000	0.160	1.000	0.160	0.140	0.044
	idw		–	1.000	0.044	1.000	0.140	0.520	0.080	0.111	**0.000**
	tps_ns			–	1.000	1.000	0.805	1.000	0.080	0.392	0.330
	tps_s				–	0.111	1.000	1.000	1.000	1.000	1.000
	rk (capa)					–	0.140	0.550	0.080	0.044	0.044
	CaPA						–	1.000	1.000	1.000	1.000
	tps_ns_s							–	0.552	1.000	0.459
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	1.000
	ok_s										–
1.5 – 2.8	ok	–	1.000	1.000	1.000	1.000	1.000	0.851	0.546	0.180	0.640
	idw		–	1.000	1.000	1.000	1.000	1.000	1.000	0.430	1.000
	tps_ns			–	1.000	1.000	1.000	1.000	1.000	0.615	1.000
	tps_s				–	1.000	1.000	1.000	1.000	0.722	0.900
	rk (capa)					–	1.000	1.000	1.000	0.640	1.000
	CaPA						–	1.000	1.000	0.396	1.000
	tps_ns_s							–	1.000	1.000	1.000
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	1.000
	ok_s										–
> 2.8	ok	–	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.660	0.504
	idw		–	1.000	1.000	0.840	1.000	1.000	0.544	0.160	0.129
	tps_ns			–	0.962	0.682	1.000	0.296	0.190	**0.000**	0.129
	tps_s				–	1.000	1.000	1.000	0.720	0.918	0.525
	rk (capa)					–	0.129	1.000	1.000	1.000	1.000
	CaPA						–	0.975	0.672	0.044	0.160
	tps_ns_s							–	1.000	1.000	0.720
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	1.000
	ok_s										–

Equitable threat score (ETS): values range between -0.333 to 1, where 1 is perfect skill.
Bold: comparison achieved a statistical significance at α level of 0.05.

Table A3-11. The resampling post-hoc test outputs of PCP estimates in 2015 **validation Foothills.** The PCP estimates are evaluated using **ETS**

		ETS averaged over stations in validation Foothills (test statistics)									
PCP (mm)		ok	idw	tps_ns	tps_s	rk (capa)	CaPA	tps_ns_s	tps_s_s	rk(capa)_s	ok_s
0.5 – 1.5		*0.06*	*0.07*	*0.09*	*0.09*	*0.08*	*0.09*	*0.10*	*0.12*	***0.16***	*0.14*
1.5 – 2.8		*0.07*	*0.10*	*0.07*	*0.08*	*0.10*	*0.08*	*0.09*	*0.12*	***0.14***	*0.11*
> 2.8		*0.47*	*0.48*	*0.47*	*0.46*	*0.50*	*0.46*	*0.50*	*0.52*	***0.54***	*0.52*
PCP	Method	ok	idw	tps_ns	tps_s	rk (capa)	CaPA	tps_ns_s	tps_s_s	rk(capa)_s	ok_s
0.5 – 1.5	ok	–	1.000	1.000	0.840	1.000	1.000	0.840	0.152	0.000	0.000
	idw		–	1.000	0.840	1.000	1.000	0.840	0.170	0.042	0.080
	tps_ns			–	1.000	1.000	1.000	1.000	1.000	0.152	0.152
	tps_s				–	1.000	1.000	1.000	1.000	0.152	0.264
	rk (capa)					–	1.000	1.000	0.540	0.000	0.042
	CaPA						–	1.000	1.000	0.117	0.638
	tps_ns_s							–	1.000	0.352	0.352
	tps_s_s								–	0.840	0.840
	rk(capa)_s									–	1.000
	ok_s										–
1.5 – 2.8	ok	–	0.492	1.000	1.000	0.798	1.000	1.000	0.387	0.264	1.000
	idw		–	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	tps_ns			–	1.000	1.000	1.000	1.000	0.520	0.387	1.000
	tps_s				–	1.000	1.000	1.000	1.000	0.780	1.000
	rk (capa)					–	1.000	1.000	1.000	1.000	1.000
	CaPA						–	1.000	1.000	0.180	1.000
	tps_ns_s							–	1.000	1.000	1.000
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	1.000
	ok_s										–
> 2.8	ok	–	1.000	1.000	1.000	1.000	1.000	1.000	0.123	0.000	0.000
	idw		–	1.000	1.000	1.000	1.000	1.000	0.340	0.195	0.288
	tps_ns			–	1.000	1.000	1.000	0.340	0.288	0.259	0.372
	tps_s				–	1.000	1.000	1.000	0.123	0.000	0.042
	rk (capa)					–	1.000	1.000	1.000	0.340	1.000
	CaPA						–	1.000	0.630	0.228	0.667
	tps_ns_s							–	1.000	1.000	1.000
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	1.000
	ok_s										–

Frequency bias index (FBI): adjusted from -1 to infinity, where 0 is perfect skill. Positive/negative FBI means overestimate/underestimate

Bold: comparison achieved a statistical significance at α level of 0.05.

Table A3-12. The resampling post-hoc test outputs of PCP estimates in 2015 **Validation Boreal.** The PCP estimates are evaluated using **FBI.**

		FBI averaged over stations in validation Boreal (test statistics)									
PCP (mm)		ok	idw	tps_ns	tps_s	rk (capa)	CaPA	tps_ns_s	tps_s_s	rk(capa)_s	ok_s
0.5 – 1.5		*1.26*	*1.54*	*0.32*	*0.75*	*1.51*	*0.45*	***0.23***	*0.38*	*0.35*	*0.52*
1.5 – 2.8		*1.49*	*1.27*	*0.29*	*0.49*	*0.66*	*0.13*	*0.07*	*0.08*	*0.14*	***0.06***
> 2.8		*0.08*	*0.04*	*0.07*	*0.07*	***0.02***	*(0.22)*	*(0.08)*	*(0.17)*	*(0.24)*	*(0.24)*
PCP	Method	ok	idw	tps_ns	tps_s	rk (capa)	CaPA	tps_ns_s	tps_s_s	rk(capa)_s	ok_s
0.5 – 1.5	ok	–	0.963	0.050	0.171	1.000	0.034	0.050	**0.000**	**0.000**	0.234
	idw		–	**0.000**	**0.000**	1.000	**0.000**	**0.000**	**0.000**	0.034	0.034
	tps_ns			–	**0.000**	0.034	0.578	0.858	1.000	1.000	0.936
	tps_s				–	0.088	0.105	**0.000**	0.034	0.034	0.936
	rk (capa)					–	0.034	**0.000**	**0.000**	0.034	0.050
	CaPA						–	0.034	1.000	0.936	1.000
	tps_ns_s							–	0.120	0.735	0.578
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	0.812
	ok_s										–
1.5 – 2.8	ok	–	1.000	0.030	0.030	0.133	**0.000**	**0.000**	**0.000**	0.048	**0.000**
	idw		–	0.030	**0.000**	0.030	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	tps_ns			–	0.812	0.105	0.819	**0.000**	0.240	0.375	0.133
	tps_s				–	0.819	0.170	0.105	0.048	0.088	0.030
	rk (capa)					–	0.030	**0.000**	**0.000**	**0.000**	**0.000**
	CaPA						–	1.000	1.000	1.000	1.000
	tps_ns_s							–	1.000	1.000	1.000
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	1.000
	ok_s										–
> 2.8	ok	–	1.000	1.000	1.000	1.000	0.050	0.289	0.050	0.030	0.030
	idw		–	1.000	1.000	1.000	0.050	0.960	0.100	0.050	0.030
	tps_ns			–	1.000	1.000	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	tps_s				–	0.960	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	rk (capa)					–	**0.000**	0.152	0.030	**0.000**	**0.000**
	CaPA						–	0.030	1.000	1.000	1.000
	tps_ns_s							–	0.063	**0.000**	**0.000**
	tps_s_s								–	0.352	0.252
	rk(capa)_s									–	1.000
	ok_s										–

Frequency bias index (FBI): adjusted from -1 to infinity, where 0 is perfect skill. Positive/negative FBI means overestimate/underestimate

Bold: comparison achieved a statistical significance at α level of 0.05.

Table A3-13. The resampling post-hoc test outputs of PCP estimates in 2015 **Validation Foothills.** The PCP estimates are evaluated using **FBI.**

		FBI averaged over stations in validation Foothills (test statistics)									
	PCP (mm)	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
	0.5 – 1.5	1.031	1.000	0.131	0.413	0.855	0.545	0.044	0.287	0.413	0.391
	1.5 – 2.8	1.076	0.931	0.067	0.784	0.716	0.248	0.033	0.147	0.062	0.094
	> 2.8	-0.011	0.109	0.103	0.099	0.032	-0.122	-0.022	-0.141	-0.140	-0.201
PCP	Method	ok	idw	tps _ns	tps_s	rk (capa)	CaPA	tps _ns_s	tps _s_s	rk(capa) _s	ok_s
0.5 – 1.5	ok	–	0.963	0.050	0.171	1.000	0.034	0.050	**0.000**	**0.000**	0.234
	idw		–	**0.000**	**0.000**	1.000	**0.000**	**0.000**	**0.000**	0.034	0.034
	tps_ns			–	**0.000**	0.034	0.578	0.858	1.000	1.000	0.936
	tps_s				–	0.088	0.105	**0.000**	0.034	0.034	0.936
	rk (capa)					–	0.034	**0.000**	**0.000**	0.034	0.050
	CaPA						–	0.034	1.000	0.936	1.000
	tps_ns_s							–	0.120	0.735	0.578
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	0.812
	ok_s										–
1.5 – 2.8	ok	–	1.000	0.031	1.000	1.000	0.115	**0.000**	0.052	0.031	0.031
	idw		–	**0.000**	1.000	1.000	0.031	**0.000**	**0.000**	0.031	**0.000**
	tps_ns			–	**0.000**	**0.000**	1.000	1.000	1.000	1.000	1.000
	tps_s				–	1.000	0.286	**0.000**	0.052	**0.000**	**0.000**
	rk (capa)					–	0.096	**0.000**	**0.000**	**0.000**	**0.000**
	CaPA						–	1.000	1.000	0.798	1.000
	tps_ns_s							–	1.000	1.000	1.000
	tps_s_s								–	1.000	1.000
	rk(capa)_s									–	1.000
	ok_s										–
> 2.8	ok	–	0.336	0.432	0.230	1.000	0.825	1.000	0.230	0.336	0.064
	idw		–	1.000	1.000	0.825	0.175	0.432	**0.000**	**0.000**	0.034
	tps_ns			–	1.000	0.825	0.112	**0.000**	**0.000**	**0.000**	**0.000**
	tps_s				–	0.825	0.192	0.336	**0.000**	0.093	**0.000**
	rk (capa)					–	0.093	0.825	**0.000**	0.034	**0.000**
	CaPA						–	0.432	1.000	1.000	0.825
	tps_ns_s							–	0.093	0.112	**0.000**
	tps_s_s								–	1.000	0.825
	rk(capa)_s									–	0.156
	ok_s										–

Frequency bias index (FBI): adjusted from -1 to infinity, where 0 is perfect skill. Positive/negative FBI means overestimate/underestimate

Bold: comparison achieved a statistical significance at α level of 0.05.

Table A3-14. Resampling post-hoc tests of 1000-bootstrapped MAE for FWI System calculated by interpolating precipitation, and observed RH, WS, Temp in 2015 over validation area.

Index	ok	idw	tps_ns	tps_s	rk (capa)	CaPA	tps_ns_s	tps_s_s	rk(capa)_s	ok_s
FFMC	4.65	4.58	4.21	4.29	4.07	4.23	3.93	3.8	3.58	3.63
DMC	8.40	7.93	7.74	7.93	7.69	9.62	7.77	8.61	8.42	9.15
FWI	3.04	2.79	2.51	2.64	2.60	2.74	2.41	2.46	2.36	2.45

Index	Method	ok	idw	tps_ns	tps_s	rk (capa)	CaPA	tps_ns_s	tps_s_s	rk(capa)_s	ok_s
FFMC	ok	–	0.855	0.112	0.144	**0.000**	0.112	**0.000**	**0.000**	**0.000**	**0.025**
	idw		–	0.112	0.112	0.112	0.112	**0.025**	**0.000**	**0.000**	**0.000**
	tps_ns			–	0.534	0.534	0.855	**0.000**	**0.000**	**0.000**	**0.025**
	tps_s				–	0.102	0.855	**0.000**	**0.000**	**0.000**	**0.000**
	rk (capa)					–	0.328	0.399	0.090	**0.000**	**0.025**
	CaPA						–	0.112	**0.025**	**0.000**	**0.000**
	tps_ns_s							–	**0.000**	**0.000**	**0.025**
	tps_s_s								–	**0.025**	**0.025**
	rk(capa)_s									–	0.534
	ok_s										–
DMC	ok	–	0.805	0.540	0.960	0.510	0.504	0.504	1.000	1.000	1.000
	idw		–	1.000	1.000	1.000	0.126	1.000	1.000	1.000	0.861
	tps_ns			–	1.000	1.000	**0.000**	1.000	0.510	1.000	0.675
	tps_s				–	0.728	0.164	1.000	0.744	1.000	0.728
	rk (capa)					–	0.086	1.000	0.527	0.510	0.456
	CaPA						–	0.164	0.195	0.044	1.000
	tps_ns_s							–	0.456	1.000	0.638
	tps_s_s								–	1.000	0.858
	rk(capa)_s									–	0.638
	ok_s										–
FWI	ok	–	0.416	0.104	0.189	0.189	0.440	0.064	0.064	**0.000**	0.104
	idw		–	0.039	0.288	0.138	0.680	**0.000**	**0.000**	**0.019**	0.039
	tps_ns			–	0.064	0.632	0.084	0.039	0.632	0.084	0.632
	tps_s				–	0.680	0.416	0.039	0.120	**0.000**	0.198
	rk (capa)					–	0.375	0.198	0.520	0.189	0.520
	CaPA						–	**0.000**	0.039	**0.019**	**0.000**
	tps_ns_s							–	0.392	0.680	0.680
	tps_s_s								–	0.138	0.680
	rk(capa)_s									–	0.064
	ok_s										–

Bold: comparison achieved a statistical significance at α level = 0.05.
Note: The resampling post-hoc test is a combination of resampling paired t-test with Holm-Bonferroni p-values adjustment and is documented in Section 2.4.3

Appendix 4: Examples of 24-h precipitation estimated using the candidate methods

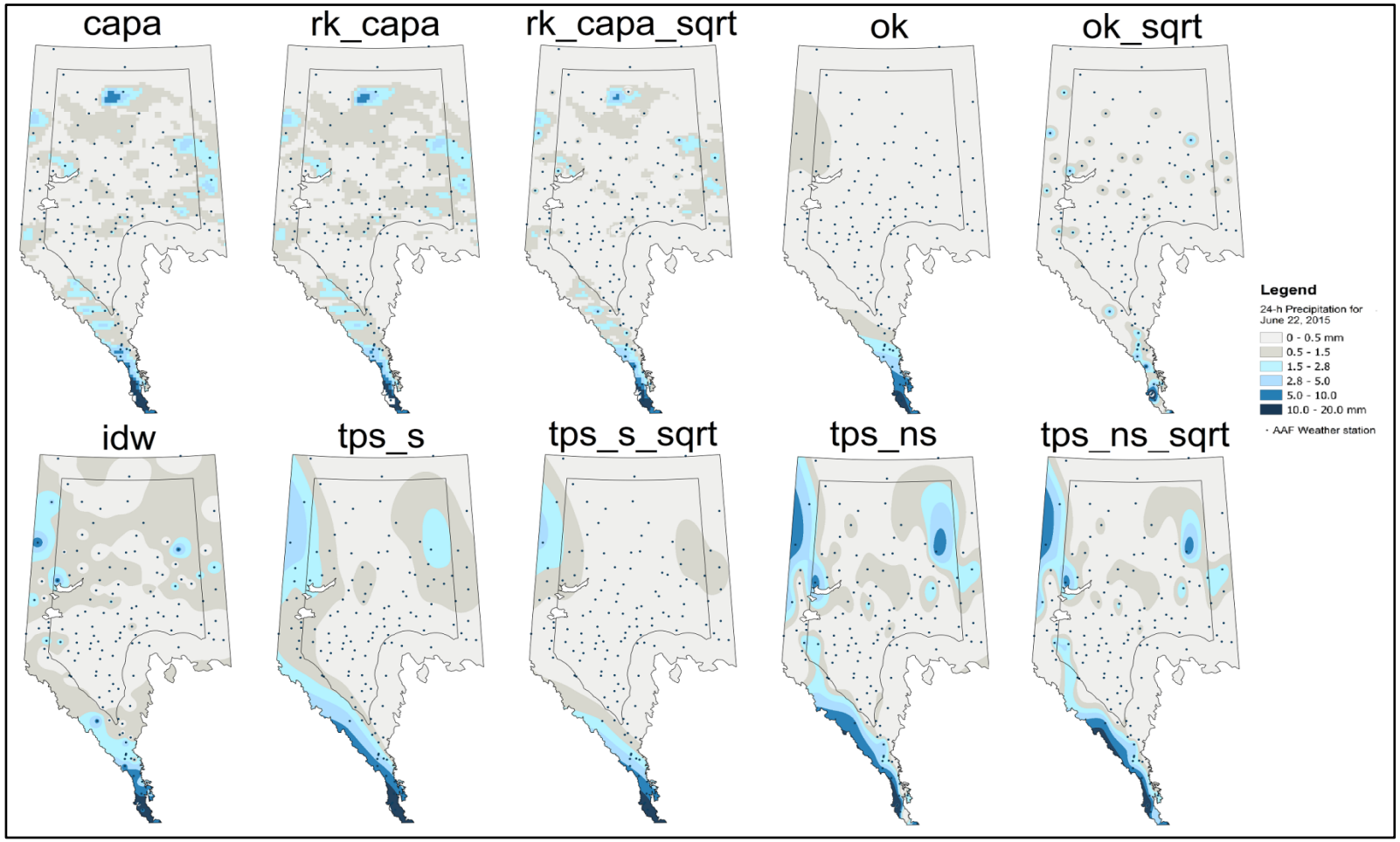


Figure A4-1 24-h precipitation estimated using the candidate methods on May 26, 2015 (a light precipitation day). Each dot represents a weather station. This map shows the advantage of using regression kriging with CaPA System in estimating precipitation in our study area. For example, there was a small precipitation event on the north of the study area that was captured by CaPA System and regression kriging related methods, while the rest of the interpolation methods could not capture this precipitation event due to the lack of weather stations in that area.

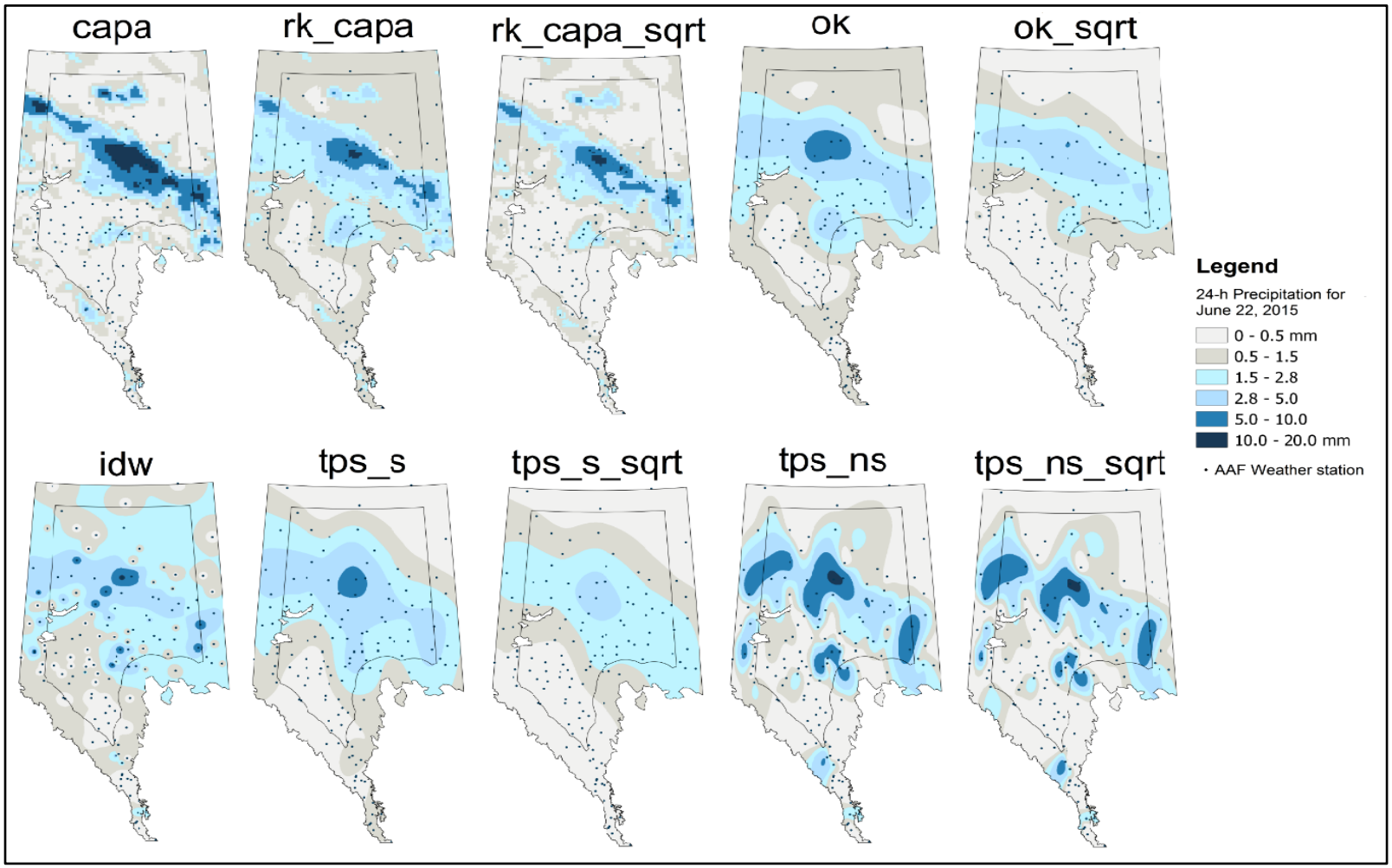


Figure A4-2 24-h precipitation estimated using the candidate methods on June 22, 2015 (a rainy day).

In a rainy day, precipitation estimated with the 10 methods varied greatly. Regression kriging with CaPA performed the best because (1) regression kriging could capture precipitation events identified by both the CaPA System and the AAF fire weather stations; (2) regression kriging also produced a gridded precipitation map with more details compared with the other interpolation methods. IDW produced a precipitation map with the signature "bullseye", which were not realistic compared to the variability of precipitation. Additionally, these maps also showed that the square root of the observed precipitation could result in underestimation.