Evaluating the Utility of Passive Microwave-Derived Soil Moisture Estimates for Forecasting Canola Yields across the Canadian Prairies

by

Jenelle White

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements for the degree of
Master of Science
in
Geography

Guelph, Ontario, Canada

© Jenelle White, August, 2018
ABSTRACT

EVALUATING THE UTILITY OF PASSIVE MICROWAVE-DERIVED SOIL MOISTURE ESTIMATES FOR FORECASTING CANOLA YIELDS ACROSS THE CANADIAN PRAIRIES

Jenelle White
University of Guelph, 2018

Advisor:
Dr. Aaron Berg
Committee Member:
Dr. Catherine Champagne

Soil moisture is a key variable in the determination of crop yields in arid regions around the world. While in-situ soil moisture measurements are sparsely-distributed, passive microwave remote sensing offers an efficient and accurate means for acquiring large-scale observations of surface soil moisture. This research evaluated the utility of passive microwave-derived soil moisture for forecasting canola yields across the Canadian Prairies using soil moisture observations obtained by the Soil Moisture Ocean Salinity Mission (SMOS) satellite. Initial work explored the relationship between soil moisture and canola yields and determined that canola yields are strongly associated ($p < 0.01$, df = 1) with excess soil moisture conditions throughout the growing season, and in particular, during the stand establishment stage (SM $\geq 26.6\%$), in low-yielding years. The Integrated Canadian Crop Yield Forecaster (ICCYF) was then employed to assess the added-value of utilizing SMOS soil moisture observations for forecasting canola yields. Improved model fit ($R^2_{diff} > 0$) was observed across most of the Canadian Prairies when SMOS soil moisture indices were included in the ICCYF, however $R^2$ values revealed that model performance overall was relatively low. These findings suggest that while passive microwave-derived soil moisture observations provide an effective indicator of canola yields, forecast skill is limited by the short temporal record of these datasets.
ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Aaron Berg, for his continual guidance, support and enthusiasm throughout this research over the last two years. I would also like to thank Dr. Catherine Champagne for serving as my committee member; this research would not have been possible without your invaluable expertise. Additionally, I would like to give thanks to Agriculture and Agri-Food Canada, the National Sciences and Engineering Research Council of Canada (NSERC), the Canadian Space Agency, the Queen Elizabeth II Graduate Scholarship in Science and Technology, and the Canada First Research Excellence Fund: Food From Thought Initiative for providing funding for the research completed in this thesis.

I would like to extend thanks to the many people who assisted me throughout this research. Thank-you to Dr. Jaison Thomas and Adam Bonnycastle for your assistance with the MATLAB and ArcGIS scripts used. Thank-you to Dr. Jon Warland and Dr. Yinsuo Zhang for your support and expertise regarding the iterative chi-square technique and Integrated Canadian Crop Yield Forecaster (ICCYF) model. Finally, thank-you to my parents – my biggest cheerleaders – and my partner Will for always providing me with endless support and encouragement. I could not have done this without all of you by my side.
TABLE OF CONTENTS

ABSTRACT ................................................................................................................................. ii
ACKNOWLEDGEMENTS ........................................................................................................... iii
LIST OF FIGURES ...................................................................................................................... vi
LIST OF TABLES ....................................................................................................................... viii
Chapter 1.0: Introduction ........................................................................................................... 1
  1.1 Background ......................................................................................................................... 1
  1.2 Research Aims and Objectives ........................................................................................... 3
  1.3 Thesis Outline .................................................................................................................... 3
Chapter 2.0: Canola Yield Sensitivity to Passive Microwave-Derived Soil Moisture Estimates in Saskatchewan, Canada ................................................................. 5
  Abstract ................................................................................................................................. 5
  2.1 Introduction ....................................................................................................................... 6
  2.2 Methods ............................................................................................................................. 9
    2.2.1 Study Area .................................................................................................................... 9
    2.2.2 Yield and Climate Data ............................................................................................... 11
    2.2.3 Soil Moisture Data ..................................................................................................... 12
    2.2.4 Iterative Chi-Square Analysis .................................................................................... 13
  2.3 Results and Discussion ...................................................................................................... 15
    2.3.1 Temperature ................................................................................................................ 15
    2.3.2 Precipitation ................................................................................................................. 19
    2.3.3 Soil Moisture ................................................................................................................ 21
  2.4 Conclusions ....................................................................................................................... 24
Chapter 3.0: Assessing the Performance of SMOS Soil Moisture Observations for Forecasting Canola Yields across the Canadian Prairies ......................................................... 26
  Abstract ................................................................................................................................. 26
  3.1 Introduction ....................................................................................................................... 27
  3.2 Methods ............................................................................................................................. 31
    3.2.1 Study Region ................................................................................................................ 31
    3.2.2 Input Data .................................................................................................................... 32
      3.2.2.1 Yield Data ............................................................................................................ 32
      3.2.2.2 Agroclimate Data ............................................................................................... 33
      3.2.2.3 AVHRR NDVI Data ............................................................................................ 35
LIST OF FIGURES

**Figure 2.1:** Census Agricultural Regions (CARs) of the province of Saskatchewan. Canola growing regions across Canada are depicted in grey. .............................................................. 10

**Figure 2.2:** Three-week running average daily chi-square values for maximum temperature in years of high (solid) and low (dashed) canola yields. Dotted grey lines indicate significance thresholds (±7, p < 0.01, df = 1). ........................................................................................................ 16

**Figure 2.3:** Three-week running average daily chi-square values for minimum (nocturnal) temperature in years of high (solid) and low (dashed) canola yields. Dotted grey lines indicate significance thresholds (±7, p < 0.01, df = 1). ........................................................................................................ 17

**Figure 2.4:** Three-week running average daily chi-square values for rainfall in years of high (solid) and low (dashed) canola yields. Dotted grey lines indicate significance thresholds (±7, p < 0.01, df = 1) ........................................................................................................ 19

**Figure 2.5:** Three-week running average daily chi-square values for SMOS percent volumetric surface soil moisture in years of high (solid) and low (dashed) canola yields. Dotted grey lines indicate significance thresholds (±7, p < 0.01, df = 1). ........................................................................................................ 21

**Figure 3.1:** Census Agricultural Regions (CARs), ecodistricts, townships and the major canola producing regions across the Canadian Prairies. Selected climate stations are represented by the black triangles. ........................................................................................................ 31

**Figure 3.2:** Data and modelling workflow of the Integrated Canadian Crop Yield Forecaster (ICCYF). Adapted from Chipanshi et al. (2015). ........................................................................................................ 40

**Figure 3.3:** Ecodistricts with SMOS soil moisture indices as predictors of canola yields. A) Ecodistricts that included SMOS soil moisture indices as a predictor at least once over the 7-year period under study; and B) Ecodistricts that consistently selected SMOS soil moisture indices as a predictor every year ........................................................................................................ 47

**Figure 3.4:** Percent occurrence of SMOS soil moisture indices in developed ecodistrict yield models. A total of 1,382 ecodistrict models were developed between 2010 and 2016. ........................................................................................................ 49

**Figure 3.5:** Bravais and Pearson coefficient of determination (R²) across the Canadian Prairies for A) developed ecodistrict-level yield models including agroclimate and NDVI indices; and B) developed ecodistrict-level yield models including agroclimate, NDVI and soil moisture indices. Black hatches denote statistically significant R² values at the 95% confidence level (p < 0.05). ........................................................................................................ 51
**Figure 3.6:** \( R^2_{\text{diff}} \) based on the correlations between predicted and surveyed canola yields when SMOS soil moisture indices are included and excluded from the ICCYF. A positive \( R^2_{\text{diff}} \) (blue) indicates model improvements with the inclusion of SMOS soil moisture observations, while a negative \( R^2_{\text{diff}} \) (red) indicates a decline in the model’s performance. ........................................... 52

**Figure 3.7:** A) Coefficient of variation (CV) of surveyed township-level canola yields across the Canadian Prairies from 2010 to 2016, and B) Spatial density of canola crop land area based on the temporal and spatial frequency of the crop between 2009 and 2016 (AAFC, 2017b). .......... 54
LIST OF TABLES

Table 3.1: Top ten predictors selected by ecodistrict regression models based on input datasets including agroclimate and NDVI indices alone (Agroclimate + NDVI), and agroclimate, NDVI and soil moisture indices (Agroclimate + NDVI + SM). Agroclimate variables include the effective growing degree days (EGDD), the crop water stress index (SI) and precipitation (P). Numbers following these indices (i.e. 5,6,7 and 8) represent May, June, July and August, respectively. SD and Sum prior to the indices represent whether it is the standard deviation or temporal sum. For three-week average NDVI indices, the numbers following refer to the weeks of the year used for the averaging calculation. Similarly, for soil moisture indices, the number following refers to the week of the year the observation was made.
Chapter 1.0: Introduction

1.1 Background

Crop yield forecasting is essential for determining potential and actual food losses that can aid in the development of export-import policies, food security policies, and efficient land management practices. In the Canadian Prairies, accurate yield forecasts are extremely important as the three Prairie provinces (i.e. Manitoba, Saskatchewan and Alberta) account for roughly 80% of Canada’s total cropland area and play a vital role in the country’s economy and in national and global food security (Statistics Canada, 2018b). However, accurate crop yield forecasts are highly complex as agricultural systems are extremely susceptible to climate variability (Basso et al., 2013; Ren et al., 2012). Improved understanding of the relationships between climatic variables and crop yields is therefore required as climate variability intensifies under global climate change (IPCC, 2013).

While numerous factors exist that both directly and indirectly impact crop yields (e.g. soil type, land management practices, seed varieties, meteorological factors), soil moisture is widely regarded as one of the main determinants of agricultural production as it directly controls the amount of water and nutrients available to support crop growth and development (Champagne et al., 2012; Dobriyal et al., 2012; Holzman et al., 2014; Judge, 2007; Maybank et al., 1995; McGinn and Shepherd, 2003; Nadler, 2007). In the Canadian Prairies, soil moisture is a key factor limiting agricultural yields as crop water demands often exceed precipitation amounts (Champagne et al., 2012, 2011; McGinn and Shepherd, 2003; Nadler, 2007). However, the relationship between soil moisture and yield has remained poorly understood due to the challenges historically associated with soil moisture quantification.

Soil moisture is highly variable, both spatially and temporally, due to varying soil types
and textures, vegetation, topography, meteorological conditions and land management practices (Barrett and Petropoulos, 2014; Dobriyal et al., 2012; Engman and Chauhan, 1995; Holzman et al., 2014). Traditional, ground-based soil moisture measurements (e.g. gravimetric and in-situ methods) are therefore largely unrepresentative of soil moisture conditions as they lack the spatial density required to capture soil moisture variability present across the landscape (Champagne et al., 2012, 2011; Dobriyal et al., 2012; Holzman et al., 2014). As a result, current crop yield forecasting systems mainly rely on soil moisture estimates derived from water balance models using temperature and precipitation inputs (Baier and Robertson, 1996; Bornn and Zidek, 2012; Robock et al., 2000). However, recent advancements in remote sensing platforms offer a promising solution for large-scale soil moisture measurement, with several satellites launched in the past decade dedicated to global soil moisture retrievals. Microwave L-band (1.41 GHz) platforms in particular, such as the Soil Moisture Ocean Salinity mission (SMOS; Kerr et al., 2010) and Soil Moisture Active Passive mission (SMAP; Entekhabi et al., 2010) satellites, are able to acquire accurate, global measurements of surface soil moisture (depth < 5 cm) at frequent temporal resolutions (1-3 days) and at a greater spatial density than ground-based soil moisture measurements (Barrett and Petropoulos, 2014; Champagne et al., 2012; Kerr et al., 2012). The short temporal record of most operating satellites (< 10 years) however, has largely limited the use of remotely-sensed soil moisture data in crop yield forecasting as a 30-year data record is typically required in order to define baseline soil moisture conditions and extreme events (Sheffield et al., 2004). Some recent research has examined the use of shorter temporal record satellite data and has found that in most cases, records of five years or more can quantify most of the variability in soil moisture with little bias for analyzing drought and excess moisture conditions (Champagne et al., 2011). This research will evaluate if soil moisture observations
obtained by the SMOS satellite can adequately capture this relationship for forecasting crop yields. This research will not only improve our understanding of the soil moisture-yield relationship but will also facilitate the development of more reliable crop yield forecasting systems, incorporating soil moisture as a key variable.

1.2 Research Aims and Objectives

The aim of this research is to evaluate the use of remotely-sensed soil moisture observations obtained by the SMOS satellite as a predictor of canola yields across the Canadian Prairies. This aim will be addressed through two key research objectives:

1. Determine the critical time periods and threshold conditions in which canola exhibits the greatest sensitivity to soil moisture stress using passive microwave observations obtained by the SMOS satellite and compare the significance of the observed associations to other more traditionally used climate indicators (e.g. precipitation, temperature).

2. Assess the added-value of utilizing remotely-sensed soil moisture observations from the SMOS satellite for forecasting canola yields across the Canadian Prairies.

1.3 Thesis Outline

This thesis is structured into four chapters with two distinct manuscripts addressing each of the research objectives listed above. In Chapter 2, a manuscript addressing the first research objective and submitted for peer-review entitled, “Canola yield sensitivity to passive microwave-derived soil moisture estimates in Saskatchewan, Canada,” is presented. This manuscript authored by Jenelle White, Aaron Berg, Catherine Champagne and Jon Warland discusses the motivations behind the study, the methods used, and the results, along with their implications. Chapter 3 addresses the second research objective: Assessing the added-value of utilizing remotely-sensed soil moisture observations obtained by the SMOS satellite for forecasting
canola yields across theCanadian Prairies. This manuscript chapter includes a description of the problem context and the methods and datasets used, as well as a presentation of the research findings. Lastly, Chapter 4 provides a summary of the overall thesis findings and recommendations for future research surrounding the inclusion of passive microwave soil moisture observations within crop yield forecasting systems.
Chapter 2.0: Canola Yield Sensitivity to Passive Microwave-Derived Soil Moisture Estimates in Saskatchewan, Canada

Abstract

Canola production is primarily attributed to variations in precipitation and temperature, however the direct relationship between soil moisture and canola yield has yet to be fully evaluated. Recent advancements in microwave remote sensing offer the potential for timely and accurate acquisitions of large-scale soil moisture data; however, the short temporal record of these datasets presents a challenge for defining extreme soil moisture events, and satellite limitations surrounding the detection of surface conditions prevents a direct agronomic link between satellite soil moisture and plant available water. This study examines the potential of passive microwave surface soil moisture observations from the Soil Moisture Ocean Salinity (SMOS) mission satellite as an indicator of canola yields across Saskatchewan over a 6-year period from 2010-2015. An iterative chi-square analysis was employed to investigate the association between SMOS soil moisture and canola yield data from Saskatchewan’s 20 Census Agricultural Regions (CARs). Associations between canola yields and traditionally used climate indicators (temperature and rainfall) were compared to assess the relative effectiveness of SMOS soil moisture as an indicator of yield. Cooler-than-average temperatures ($T_{\text{max}} \leq 26^\circ\text{C}$ and $T_{\text{min}} \leq 12^\circ\text{C}$) at the beginning of August were shown to have a positive impact on canola yield, while high nocturnal temperatures ($T_{\text{min}} \geq 15^\circ\text{C}$) in mid-July had a negative effect. Precipitation was found to have a minimal impact on canola production, showing little deviation from daily rainfall conditions in normal-yielding years in both high- and low-yielding years, likely due to greater-than-average rainfall that was observed during the period of study. The strongest associations were observed between soil moisture and canola yield, particularly in low-yielding years with excess soil moisture exhibiting significant associations ($p < 0.01, df = 1$) throughout the growing season. The beginning of June,
coinciding with the stand establishment stage, was identified as a critical time period in which excess soil moisture (≥ 26.6%) resulted in yield losses. The findings of this study suggest that soil moisture observations obtained by passive microwave satellites, such as SMOS, may be effectively utilized as predictors of crop yields, particularly in areas experiencing excess soil moisture conditions.

2.1 Introduction

Canola, also known as oilseed rape, is the most valuable field crop in Canada, contributing approximately $26.7 billion to the Canadian economy each year (LMC International, 2016; Statistics Canada, 2009). Canola however, is extremely susceptible to environmental stresses as it is a cool-season crop (Bérard et al., 2001; Brandt and McGregor, 1997; Morrison, 1993), growing best at temperatures between 10 and 30°C (Ontario Ministry of Agriculture, Food and Rural Affairs, 2016), and can require up to 480 mm of water during the growing season (Alberta Agriculture and Forestry, 2016). Numerous studies have investigated the adverse effects of temperature and precipitation extremes on canola crops (e.g. Aksouh-Harradj et al., 2006; Angadi et al., 2000; Bérard et al., 2001; Gan et al., 2004; Kutcher et al., 2010; Meng et al., 2017; Morrison, 1993; Morrison and Stewart, 2002), with high temperature and water stress having shown to result in infertility and yield losses for various Brassica species. Improved understanding of the relationship between climatic variables and canola production across the Canadian Prairies however, is required as climate extremes intensify and become more frequent under global climate change.

The timing and magnitude of climatic stresses are key determinants of the severity of impacts on crop yields, as a plant’s vulnerability to stress varies at different stages of growth (Aksouh-Harradj et al., 2006; Angadi et al., 2000; Champagne et al., 2011; McGinn and
Shepherd, 2003; Mkhabela et al., 2011). Knowledge of these critical time periods and threshold conditions would therefore allow for more informed yield prediction methodologies, policy decisions and land management planning, and would better equip farmers and society to respond to climate variability. In a study analyzing the effects of temperature and precipitation on canola yields across the province of Saskatchewan, Kutcher et al. (2010) was able to identify the timing of temperature and rainfall thresholds indicative of canola yield variations using the iterative chi-square analysis developed by Joseph M. Caprio (1966). Long-term climate and yield data from 1967 to 2001 revealed that yield losses across Saskatchewan’s 20 Census Agricultural Regions (CARs) over the 35-year period were associated with high temperatures and low precipitation at the beginning of July, coinciding with the flowering stage of development in Saskatchewan (Kutcher et al., 2010). Low-yielding years were found to have a greater number of days with daily minimum and maximum temperatures in excess of 16 and 31°C, respectively, and fewer days with more than 2 mm of rainfall during this critical time period (Kutcher et al., 2010). High-yielding years on the other hand, exhibited associations with greater-than-average precipitation and cooler nocturnal temperatures (Kutcher et al., 2010).

Few studies however, have aimed to identify the critical time periods and threshold conditions in which canola, and many other crops, exhibit the greatest sensitivity to soil moisture stress, despite its recognized importance for crop growth and development. Soil moisture is widely regarded throughout literature as one of the main determinants of yield in many arid regions around the world (Champagne et al., 2012; Dobriyal et al., 2012; Holzman et al., 2014; Maybank et al., 1995; McGinn and Shepherd, 2003; Nadler, 2007). In the Canadian Prairies, soil moisture is a key factor limiting agricultural yields as crop water demands often exceed precipitation amounts, and thus soil available water (Champagne et al., 2012, 2011; McGinn and
Shepherd, 2003; Nadler, 2007). Excess soil moisture conditions produced by major rainfall events and/or high snowmelt runoff however, also play a significant role in limiting yields across the Prairies through the production of anoxic soil conditions and increased salinity and sodicity (Bedard-Haughn, 2009; Holzman et al., 2014). These soil moisture extremes (i.e. too much or too little water) can have profound impacts on crop development and annual yields when they occur during critical stages of growth. The identification of these critical time periods and threshold conditions however, have been largely limited due to the challenges and limitations historically associated with soil moisture quantification.

Soil moisture is highly variable, spatially and temporally, due to varying soil types, vegetation cover, topography, climate conditions, and land management practices, which poses a significant challenge for large-scale quantification (Barrett and Petropoulos, 2014; Dobriyal et al., 2012; Engman and Chauhan, 1995; Holzman et al., 2014). Therefore, while traditional, in-situ soil moisture measurements provide highly accurate, point-estimates of soil moisture extremes, they lack the spatial density required to capture soil moisture variability present on a larger scale (Champagne et al., 2012, 2011; Dobriyal et al., 2012; Holzman et al., 2014). Recent advancements in remote sensing methods however, such as the microwave L-band (1.41 GHz) Soil Moisture Ocean Salinity mission (SMOS; Kerr et al., 2010) and Soil Moisture Active Passive mission (SMAP; Entekhabi et al., 2010) satellites, can provide global measurements of surface soil moisture (depth < 5 cm) at a greater spatial density than ground-based measurements (Barrett and Petropoulos, 2014; Champagne et al., 2012; Kerr et al., 2012). Launched in 2010, several studies in recent years have shown that the SMOS satellite can be utilized to effectively capture extreme events, such as drought or flooding, across agricultural regions (e.g. Chakrabarti et al., 2014; Champagne et al., 2015), including the Canadian Prairies, with root mean squared
errors (RMSE) less than 0.10 m³m⁻³ (Champagne et al., 2016). However, most operating satellites have a short temporal record which can pose a significant challenge to assessing soil moisture-yield relationships. Therefore, the aim of the current study is to investigate the potential of SMOS soil moisture as an indicator of canola production in the Canadian Prairies using a 6-year period from 2010-2015. Following the approach of Kutcher et al. (2010), this will be achieved through two objectives: 1) Determine the critical time periods and threshold conditions in which canola exhibits the greatest sensitivity to soil moisture stress using passive microwave observations obtained by the SMOS satellite; and 2) Compare the significance of the association between SMOS soil moisture variations and canola yields to other more traditionally used climate indicators (e.g. temperature and precipitation) in order to assess the utility of the SMOS satellite for yield prediction purposes.

2.2 Methods

2.2.1 Study Area

All analyses were conducted over the province of Saskatchewan within the Canadian Prairies (Figure 2.1). Saskatchewan is the largest producer of Canola in Canada, accounting for roughly 52.0% of the total canola seeded area and 49.4% of the total canola production in the country each year (Statistics Canada, 2017a). Located in the most northerly portion of the North American Great Plains, Saskatchewan is largely characterized by its rich, fertile soils that support the extensive agricultural production occurring across the province (Campbell et al., 2002). Saskatchewan’s soils mainly consist of black/dark-gray, dark-brown and brown Chernozemic soils, ranging in texture from loamy sands to clays, which vary across three agro-climatic zones: sub-humid, semi-arid and arid (Mkhabela et al., 2011; Nadler, 2007). The most arid regions are found in southwestern Saskatchewan and are significantly drier than the semi-
arid and sub-humid regions to the northeast, which experience cooler summers and colder winters (Government of Canada, 2006). Average winter and summer temperatures are -10°C and 15°C, respectively, with temperatures varying more than 65°C common in any given year (Environment and Climate Change Canada, 2017a). Mean daily temperatures in July, the warmest month, vary from an average of 18.5°C in Swift Current in the southwest to 15.8°C near Cigar Lake in the northeast (Environment and Climate Change Canada, 2017a).

Saskatchewan receives some of the lowest precipitation amounts in the country due to its location on the leeward side of the Rocky Mountains and distance from the moderating influence of large water bodies. Annual precipitation amounts vary from 357 to 497 mm in Swift Current and Cigar Lake, respectively (Environment and Climate Change Canada, 2017a), with a provincial average of just 395 mm (McGinn and Shepherd, 2003). As a result, water availability

Figure 2.1: Census Agricultural Regions (CARs) of the province of Saskatchewan. Canola growing regions across Canada are depicted in grey.
is often a major factor limiting crop yields across Saskatchewan and the Canadian Prairies, making this region extremely prone to agricultural drought (Government of Canada, 2006; Maybank et al., 1995; McGinn and Shepherd, 2003; Meng et al., 2017; Nadler, 2007).

2.2.2 Yield and Climate Data

Saskatchewan is divided administratively by Statistics Canada (2015) into 20 CARs (Figure 2.1) composed of adjacent consolidated subdivisions that range in size from 9,421 to 357,553 km². These CARs are utilized for disseminating agricultural statistics by the Census of Agriculture, including the number of farms, crop area, livestock inventories, farm capital, operating costs and yield (Statistics Canada, 2015). Canola yields by Small Area Data Region, analogous to CARs, were acquired from Statistics Canada (2017b) in kilograms per hectare for each year from 2010 to 2015. Over the period of study, canola production across Saskatchewan varied from 5,692,600 tonnes in 2010 to 9,536,800 tonnes in 2015 (Statistics Canada, 2017b).

Daily temperature and precipitation data collected across Canada at meteorological stations is available from Environment and Climate Change Canada (2017b) by station. Approximately 40 climate stations across Saskatchewan, Manitoba and Alberta were used to represent the climate of Saskatchewan’s 20 CARs. In Saskatchewan, the number of stations per CAR varies from 1 to 5. Station-based daily maximum ($T_{\text{max}}$) and minimum ($T_{\text{min}}$) temperature and daily rainfall data spanning the growing season (1 May to 30 September) were acquired from Environment and Climate Change Canada and quality controlled, gap-filled and aggregated to the CAR-level by Agriculture and Agri-Food Canada (AAFC) for each of the six years under study (2010-2015). Data outside 1 May to 30 September were not considered as canola is seeded during the month of May and combined in late September in the province of Saskatchewan (Kutcher et al., 2010).
2.2.3 Soil Moisture Data

The SMOS satellite was launched in November 2009 by the European Space Agency (ESA) with the aim to provide global measurements of volumetric surface (depth < 5 cm) soil moisture at a spatial resolution of 35-50 km every 1-3 days and with an accuracy of 0.04 m$^3$m$^{-3}$ (Kerr et al., 2012). SMOS features a 2-D interferometric Microwave Imaging Radiometer using Aperture Synthesis (MIRAS) that operates in the L-band frequency (1.41 GHz) to passively collect naturally emitted microwave energy from the Earth’s surface across an ~1000 km swath (Kerr et al., 2010, 2001). This emittance, referred to as the microwave brightness temperature ($T_B$), is related largely to the attenuating effects of the soil dielectric properties, and thus soil moisture, at L-band frequencies (Engman and Chauhan, 1995; Njoku and Entekhabi, 1996).

Several studies have found that SMOS exhibits an overall dry bias, and therefore, tends to underestimate soil moisture (e.g. Al Bitar et al., 2012; Champagne et al., 2016; Djamai et al., 2015; Jackson et al., 2012). A recent study by Champagne et al. (2016) for example, found a root mean squared error (RMSE) of 0.05 m$^3$m$^{-3}$ between volumetric soil moisture estimates obtained by the SMOS satellite and in situ networks across Saskatchewan from 2010 to 2014. This underestimation can be problematic for identifying soil moisture thresholds, which rely heavily on the absolute accuracy of the soil moisture estimates; however, is less important for determining the overall relationship between soil moisture and yield as this depends more on the relative soil moisture trends (Champagne et al., 2016, 2015).

Level 2 SMOS soil moisture data was obtained from the European Space Agency and re-gridded to a 0.25-degree resolution using version 6.20 of the SMOS soil moisture processor by AAFC (2016). The dataset is created daily and is quality flagged to remove areas where rainfall and/or snow cover is present and where the Radio Frequency Interference (RFI) is above an
acceptable threshold (AAFC, 2016). Daily percent volumetric surface (depth < 5 cm) soil moisture data (AAFC, 2016) was collected for 1 May to 30 September between 2010 and 2015 as a series of GeoTIFF rasters and aggregated to the CAR-level using a weighted average in ArcGIS 10.2©. Dates with missing SMOS retrievals for a given CAR were gap-filled as the average of the two surrounding days. Approximately 0.8% (i.e. 155 out of 18,360 samples) of the daily CAR-level SMOS data required gap-filling. Temporally, the first year of SMOS soil moisture acquisition (i.e. 2010) had the greatest number of missing SMOS observations (i.e. 78 of the 155 missing samples), while spatially, CARs located in southwestern Saskatchewan (e.g. 4730, 4733, 4740, 4741), or Palliser’s Triangle, exhibited a greater number of days with missing SMOS retrievals over other CARs.

2.2.4 Iterative Chi-Square Analysis

The iterative chi-square analysis, developed by Caprio (1966), is a statistical procedure designed to investigate the association between climate observations and discrete biological data. The technique was originally developed as a means to compare climate records and crop yields (Caprio, 1966) and has since been applied to relate temperature and precipitation observations to wheat yield records in Montana (Kalma et al., 1992), apple (Caprio and Quamme, 1999) and grape yields (Caprio and Quamme, 2002) in British Columbia, canola yields in Saskatchewan (Kutcher et al., 2010), and cabbage, onion and rutabaga yields in Ontario (McKeown et al., 2005; Warland et al., 2006). The technique has also seen applications relating temperature and precipitation variables to tree-ring growth in southern Arizona (Caprio et al., 2003) and in comparing temperature, precipitation and modelled soil moisture estimates to grasshopper pest populations in Alberta (Powell et al., 2007).
The iterative chi-square analysis identifies the timing, magnitude and direction (positive or negative) of the relationship between individual daily climate observations and crop production throughout the growing season and determines the climate threshold values above or below which the most significant associations occur (Caprio, 1966; Kutcher et al., 2010; McKeown et al., 2005; Powell et al., 2007). The technique iteratively compares the number of days that meet a threshold condition in high- or low-yield years (observed) to normal-yielding years (expected) within a 3-week moving window, generating an overall chi-square statistic:

\[ \chi^2 = (O - E)^2 / E \]  

where \( O \) is the observed number of days that meet the condition and \( E \) is the expected, or “theoretical”, number of days meeting the condition. If the number of days meeting the threshold condition in each yield class is not proportionally distributed, the returned chi-square value departs from zero. A positive chi-square value is then assigned if there are an excess number of days meeting the threshold condition (i.e. observed > expected), while a negative value is assigned if there are a deficit of days (i.e. observed < expected; Caprio, 1966).

Traditionally, the iterative chi-square technique has been applied to datasets spanning 20-100+ years (e.g. Caprio et al., 2003; Caprio and Quamme, 2002, 1999; Kalma et al., 1992; Kutcher et al., 2010; McKeown et al., 2005; Powell et al., 2007; Warland et al., 2006); however, due to the temporal limitations associated with utilizing remotely-sensed data (i.e. most satellites have an operating life < 10 years), in this study, the iterative chi-square technique was applied for only a 6-year period spanning the SMOS data record (i.e. 2010-2015). Using MATLAB™ R2015a software, the iterative chi-square technique was utilized to compare daily temperature, rainfall and soil moisture observations from the SMOS satellite in high- or low-yielding years to years of normal canola yields between 2010 and 2015 across the 20 CARs of Saskatchewan.
Daily temperature, precipitation and soil moisture data were averaged for each CAR over the 6-year period, giving a total of 120 samples; significantly less than the 700 samples used by Kutcher et al. (2010). Low-, normal- and high-yield classes were determined based on quartiles, where low-yield years were defined as the bottom 25% of samples, normal-yield years between 25% and 75%, and high-yield years the top 25%. Thresholds were searched in steps of 1°C from -10 to 35°C for both daily maximum and minimum temperatures and in 0.2 mm increments from 0 to 50 mm for daily rainfall; the same as those used by Kutcher et al. (2010). Additionally, thresholds were searched in steps of 0.2% from 0 to 52.4% for daily percent volumetric soil moisture either above (high-to-low scan) or below (low-to-high scan) each value. For this analysis however, only high-to-low scans will be considered, following the approach of Kutcher et al. (2010).

2.3 Results and Discussion

The results of the iterative chi-square analysis for daily maximum and minimum temperature, daily rainfall and SMOS daily percent volumetric surface soil moisture are presented in Figures 2.2 to 2.5 and discussed in further detail below. Three-week running averages of daily chi-square values summarized across the 20 CARs of the province of Saskatchewan are plotted and the threshold values in which the most significant associations occur identified. Associations in years of low and high canola yields are represented by the dashed and solid lines, respectively. A significant difference between yield classes exists if the chi-square statistic exceeds ±7 (p < 0.01, degrees of freedom = 1; Caprio, 1966).

2.3.1 Temperature

Associations between canola yields and daily maximum and minimum temperatures over the study period are presented in Figures 2.2 and 2.3, respectively. High-yield years exhibited a
deficit of days with $T_{\text{max}} \geq 26^\circ$C (Figure 2.2) and $T_{\text{min}} \geq 12^\circ$C (Figure 2.3) at the beginning of August, while low-yielding years had an excess number of days with nocturnal temperatures ($T_{\text{min}}) \geq 15^\circ$C in July (Figure 2.3). Low-yielding years however, showed little to no deviation in daily maximum temperatures from normal-yielding years, with some association observed in July, with a greater number of days with warm maximum temperatures (Figure 2.2).

These results indicate that canola is extremely susceptible to heat stress, similar to the findings of previous studies, and indicate that the 6-year period examined provided sufficient data to accurately assess the temperature-yield association when compared to studies conducted over a longer period of record (e.g. Angadi et al., 2000; Brandt and McGregor, 1997; Gan et al., 2004; Kutcher et al., 2010; Meng et al., 2017; Morrison, 1993; Morrison and Stewart, 2002; Polowick and Sawhney, 1988). High temperatures have been shown to have adverse effects on canola development and can lead to substantial yield losses through increased flower and pod abortions and reductions in the number and size of seeds per flower (Aksouh-Harradj et al., 2006; Angadi et al., 2000; Gan et al., 2004; Morrison, 1993; Morrison and Stewart, 2002). Gan

![Figure 2.2](image)

**Figure 2.2:** Three-week running average daily chi-square values for maximum temperature in years of high (solid) and low (dashed) canola yields. Dotted grey lines indicate significance thresholds ($\pm 7$, $p < 0.01$, df = 1).
et al. (2004) for example, reported average yield reductions of 26 and 59% for various *Brassica* species when stresses of 28/18°C (i.e. day/night temperatures) (moderate stress) and 35/18°C (high stress) were imposed for 10 days, respectively. However, the greatest reductions in yield were observed when high temperature stress (35/18°C) was applied during the flowering and pod formation stages of development, with average yield losses of 58 and 77%, respectively, compared to a 15% average yield reduction when the same stress was applied during the bud formation stage (Gan et al., 2004).

Numerous studies have found that canola exhibits the greatest sensitivity to temperature stress during the flowering and pod formation stages of development (e.g. Aksouh-Harradj et al., 2006; Angadi et al., 2000; Brandt and McGregor, 1997; Gan et al., 2004; Kutcher et al., 2010; Morrison, 1993; Morrison and Stewart, 2002). In analyzing the effects of temperature and precipitation extremes on canola yields in Saskatchewan from 1967 to 2001, Kutcher et al. (2010) found that canola yields were most adversely affected by temperature stress at the

![Figure 2.3: Three-week running average daily chi-square values for minimum (nocturnal) temperature in years of high (solid) and low (dashed) canola yields. Dotted grey lines indicate significance thresholds (±7, p < 0.01, df = 1).](image)
beginning of July, coinciding with the flowering stage of development. In this study, the most significant associations between high nocturnal temperatures ($T_{\text{min}} \geq 15^\circ\text{C}$) and reduced canola yields were observed in mid- to end-July; approximately 2 weeks later than the critical time period identified by Kutcher et al. (2010). This is expected as excess moisture and cooler-than-average temperatures unrepresentative of the historical climate norms across Saskatchewan delayed seeding, and therefore, crop development and growth stage, in several years (i.e. 2010, 2011, 2013, 2014) over the 6-year period between 2010 and 2015 (Barthet, 2014, 2013, 2011, 2010). Additionally, the present study identified the beginning of August, coinciding with the end of the flowering stage and beginning of the pod development stage, as a crucial time period in which cooler temperatures ($T_{\text{max}} \leq 26^\circ\text{C}$ and $T_{\text{min}} \leq 12^\circ\text{C}$) allowed for improved canola yields.

Temperatures ranging from 25 to 35°C have been reported throughout the literature as critical temperatures for heat stress in canola (Angadi et al., 2000; Gan et al., 2004; Morrison, 1993; Morrison et al., 1989; Morrison and Stewart, 2002; Polowick and Sawhney, 1988). Kutcher et al. (2010) found daily maximum temperatures in excess of 31°C and nocturnal temperatures in excess of 16°C to be associated with yield losses between 1967 and 2001. In addition, the positive effects of cooler-than-average nocturnal temperatures on canola production were also observed (Kutcher et al., 2010). In the current study, canola yield reductions across Saskatchewan from 2010 to 2015 were associated with nocturnal temperatures in excess of 15°C, similar to the threshold identified by Kutcher et al. (2010). However, associations between temperature and canola production in high-yielding years were much more significant. Yield improvements were found to have significant associations with daily maximum temperatures less than 26°C and nocturnal temperatures less than 12°C at the beginning of August. Cooler-than-average temperatures were a defining characteristic of the 2010-2015 period and played a
significant role in improving canola production. Significant portions of each growing season during the 6-year period experienced below-average temperatures (Barthet, 2015, 2014, 2013, 2012, 2011, 2010), with mean temperature differences from normal up to -4°C (AAFC, 2017a); therefore, minimizing the adverse effects associated with heat stress observed in previous long-term studies.

2.3.2 Precipitation

Figure 2.4 shows the associations between daily rainfall and canola yields across Saskatchewan from 2010 to 2015. Both high- and low-yield years showed little deviation in daily rainfall from normal-yielding years, with some significance seen in late August and early September. Low-yield years had a fewer number of days with more than 0.4 mm of rain at the start of September, while high-yield years had a fewer number of days with little to no rain at the end of August.

Previous studies have found that water stress can have a profound effect on seed yield, oil concentration and seed weight for various canola cultivars when applied before or after flowering (Mailer and Cornish, 1987; Meng et al., 2017; Neilsen, 1997). Kutcher et al. (2010) for

![Figure 2.4](image)

**Figure 2.4**: Three-week running average daily chi-square values for rainfall in years of high (solid) and low (dashed) canola yields. Dotted grey lines indicate significance thresholds ($\pm 7, p < 0.01, df = 1$).
example, found that water stress (daily rainfall < 2 mm) during the flowering stage of
development had a negative effect on canola yields in Saskatchewan between 1967 and 2001. However, the results of the present study indicate water stress played a much smaller role in canola production across Saskatchewan from 2010 to 2015. The 6-year period under study experienced extreme rainfall conditions; growing season departure from average precipitation varied from surpluses in excess of 120 mm across southern Saskatchewan in 2010 and 2014, to deficits in excess of 120 mm in 2013 (AAFC, 2017a). Abnormally wet conditions were present across significant portions of Saskatchewan’s canola growing regions during the flowering stage of development for four of the six growing seasons under study (i.e. 2010, 2011, 2012, 2013) (AAFC, 2017a). As a result, the current study was unable to capture the negative effects of water stress on canola yields during this critical growth stage, as observed in previous studies, due to the 6-year data record being unrepresentative of historical climate norms in the region.

Increased precipitation however, does not necessarily translate into improved canola yields, as shown in the present study. Water stress is considered an important abiotic factor controlling crop production (Campbell et al., 1992) and numerous studies have utilized precipitation as an indictor of crop water stress (e.g. Caprio and Quamme, 1999; Kalma et al., 1992; Kutcher et al., 2010; McKeown et al., 2005; Warland et al., 2006). However, the effects of precipitation on crop development are not direct; rather, precipitation indirectly affects crop growth through its impacts on soil available water (Haferkamp, 1988; Neenu et al., 2013). A significant rainfall event for example, can have a prolonged effect on crop growth and development as the wetness may be retained in the soil long after the rainfall event itself (Koster and Suarez, 2001). If water-logging occurs, this can have detrimental impacts on crop development (Kanwar et al., 1988). As such, rainfall may not always be the best indicator of
water stress, as seen in the current study; rather soil moisture may be a better descriptor of water stress as it is able to more effectively capture the “memory” of precipitation events (Koster and Suarez, 2001).

2.3.3 Soil Moisture

The associations between canola yields and daily percent volumetric surface soil moisture observations obtained by the SMOS satellite are shown in Figure 2.5. Low-yield years exhibited significant deviations in daily soil moisture from normal-yielding years, with chi-square values exceeding the significance threshold (+7, p < 0.01, df = 1) throughout the entire growing season. The strongest association was observed at the beginning of June with a greater number of days in low-yield years with daily soil moisture values exceeding 26.6%. High-yield years however, showed little deviation in daily soil moisture compared to normal-yielding years, with some association observed at the beginning of July and in late August-early September with a fewer number of days experiencing excess soil moisture conditions.

These results indicate that excess soil moisture had a profound, adverse effect on canola

![Figure 2.5: Three-week running average daily chi-square values for SMOS percent volumetric surface soil moisture in years of high (solid) and low (dashed) canola yields. Dotted grey lines indicate significance thresholds (±7, p < 0.01, df = 1).]
production across Saskatchewan during the period of study and are in general agreement with previous findings. Numerous studies have shown that both excess and deficit soil moisture conditions can have a significant impact on the development of most crops (Haferkamp, 1988; Holzman et al., 2014; Kanwar et al., 1988; Vough and Marten, 1971). Excessive soil moisture in particular, can create numerous problems for crop production. Wet soils can negatively impact initial crop development through delayed seeding, if the soil is too wet to seed, and can result in seeds failing to germinate (Champagne et al., 2012; Kanwar et al., 1988). Excess soil moisture conditions throughout development can have a significant impact on the exchange of important gases, nutrient cycling and root development, and can lead to the production of toxic substances due to the lack of oxygen in the soil (Canola Council of Canada, 2013; Holzman et al., 2014; Kanwar et al., 1988). Detrimental impacts can arise if water-logging occurs after seed germination (Canola Council of Canada, 2013; Kanwar et al., 1988); excess soil moisture at this critical time period can be fatal, while wet conditions at a more well-developed stage of crop development can result in relatively little damage (Kanwar et al., 1988). A study by Kanwar et al. (1988) for example, found that corn yields were most heavily impacted by excessive soil moisture conditions during the plant establishment stage and were least affected when excess soil moisture was present during the yield formation stage. The results of the current study also show that the impact of excess soil moisture on canola production across Saskatchewan is greatest during stand establishment at the beginning of June. Past this stage, excessive soil moisture conditions were found to play a significant, but smaller role, in canola production, similar to the findings of Kanwar et al. (1988). These results are expected as canola is extremely susceptible to water-logging and can experience yield reductions in as little as three days under excess water conditions (Canola Council of Canada, 2013). The results of the present study also demonstrated
the positive effects of having fewer days with excessive soil moisture conditions, with associations observed in July and around the end of August and beginning of September in high-yielding years.

Previous studies however, have also shown that soil moisture deficits can adversely affect crop production. Water stress has been found to have a negative impact on leaf development, decreasing the crop’s photosynthetic area and subsequently stunting growth (Slayter, 1974; Vough and Marten, 1971). In a study of alfalfa growth under severe soil moisture stress, Vough and Marten (1971) found that soil moisture deficits had a significant impact on dry matter accumulation, resulting in decreased yields. Similarly, Denmead and Shaw (1960) reported yield reductions in corn of 25, 50 and 21% as a result of soil moisture stress applied prior to, during and after the silking stage, respectively. Yields of tall wheatgrass, timothy, alfalfa and aslike clover have also shown significant reductions as a result of decreased soil moisture (Gifford and Jensen, 1967), likely through the mechanisms discussed above. The negative impacts of soil moisture deficits on canola production however, were not observed in the current study due to the abnormally wet conditions that persisted throughout much of the 6-year period between 2010 and 2015, as discussed previously.

Numerous studies investigating climate impacts on crop yields have cited temperature and precipitation as key factors controlling crop growth and development (e.g. Angadi et al., 2000; Brandt and McGregor, 1997; Campbell et al., 1992; Kutcher et al., 2010; Meng et al., 2017). Recent studies at the University of Manitoba for example, concluded that temperature was the most important environmental factor controlling canola production across western Canada (Canola Council of Canada, 2017). However, the current study reveals that soil moisture may provide more information on the growth and development of canola across Saskatchewan than
temperature and precipitation alone. Chi-square values exceeding the significance threshold throughout the growing season in low-yielding years, and in excess of 100 ($p=0.01$, $df=1$) from mid-May to mid-June, indicate a stronger association between soil moisture and yield across Saskatchewan from 2010 to 2015 compared to the other three indicators assessed ($T_{\text{max}}$, $T_{\text{min}}$, rainfall).

2.4 Conclusions

This research has shown that there is a strong relationship between satellite soil moisture and canola yields across Saskatchewan during this period of greater-than-average precipitation. As climate variability intensifies under global climate change, there is an ever-growing need to develop a greater understanding of the effects of climatic variables on crop production. This knowledge is particularly important across the Canadian Prairies, which contain approximately 80% of Canada’s total cropland area (Kittson et al., 2007), including 99% of the total canola seeded area (Statistics Canada, 2017a). While soil moisture is considered one of the main determinants of crop yields in arid regions, such as the Canadian Prairies (Champagne et al., 2012; Dobriyal et al., 2012; Holzman et al., 2014; Maybank et al., 1995; McGinn and Shepherd, 2003; Nadler, 2007), the quantitative relationship between soil moisture and yield has remained poorly understood due to the challenges and limitations historically associated with soil moisture quantification. Thus, the aim of the current study was to investigate the potential of utilizing soil moisture observations obtained from passive microwave satellites, such as SMOS, to assess the effects of soil moisture on canola production across the province of Saskatchewan.

The identification of the critical time periods and threshold satellite soil moisture conditions affecting canola yields in Saskatchewan provides a unique insight into the relationship between yield and soil moisture that has not been investigated before. The iterative chi-square
analysis revealed that the beginning of August was a critical time period in which cool temperatures ($T_{\text{max}} \leq 26^\circ\text{C}$ and $T_{\text{min}} \leq 12^\circ\text{C}$) had a positive impact on canola yields, while low-yielding years exhibited associations with high nocturnal temperatures ($T_{\text{min}} \geq 15^\circ\text{C}$) in mid-July. Precipitation was found to have a minimal impact on canola production, showing little to no deviation in daily rainfall in both high- and low-yielding years, compared to normal-yielding years. Significant associations however, were observed between SMOS soil moisture and canola yields in low-yielding years. Excess soil moisture was revealed to have a profound, negative effect on canola production across Saskatchewan between 2010 and 2015, with the strongest associations observed at the beginning of June (SM $\geq 26.6\%$). To a lesser extent, the results of the chi-square analysis also demonstrated the positive effects of having fewer days with excess soil moisture at the beginning of July and end of August.

The association between excess soil moisture and reduced canola yields was by far the strongest association observed, followed by nocturnal temperatures. Therefore, the findings of this study suggest that soil moisture observations obtained by the SMOS satellite may be effectively utilized as a predictor of crop yields, particularly in areas experiencing excess soil moisture conditions. However, due to the abnormally wet 6-year data record, the potential of SMOS soil moisture as an indicator of canola yields in dry years remains unknown. Previous studies have argued that the surface soil moisture observations obtained by microwave satellites, such as SMOS, are unrepresentative of root zone soil moisture deficits (Capehart and Carlson, 1997; Wilson et al., 2003). Further work is therefore required to assess whether SMOS, or other satellite soil moisture datasets, can effectively capture the relationship between soil moisture and yield under dry soil moisture conditions in order to truly determine the potential of SMOS soil moisture for yield prediction purposes.
Chapter 3.0: Assessing the Performance of SMOS Soil Moisture Observations for Forecasting Canola Yields across the Canadian Prairies

Abstract

Remote sensing offers several benefits over traditional, ground-based methods for forecasting crop yields, including greater spatial and temporal coverage and reduced costs. Satellite-derived vegetation indices are widely utilized in yield forecasting models, however they can be heavily impacted by atmospheric conditions due to their reliance on visible and near-infrared portions of the electromagnetic spectrum. Given the importance of soil moisture for crop development, satellite soil moisture platforms, such as the Soil Moisture Ocean Salinity Mission (SMOS) satellite, present a unique opportunity for yield forecasting as they offer an efficient and accurate means for acquiring large-scale observations of surface soil moisture at frequent temporal resolutions (1-3 days) using microwave radiation. The objective of this study was to investigate the use of remotely-sensed soil moisture observations obtained by the SMOS satellite for forecasting canola yields across the Canadian Prairies within Agriculture and Agri-Food Canada’s (AAFC) Integrated Canadian Crop Yield Forecaster (ICCYF) model. Weekly SMOS soil moisture data was combined with agroclimate variables and normalized difference vegetation index (NDVI) data derived from the Advanced Very High Resolution Radiometer (AVHRR) platform and used as an input for forecasting canola yields at the township-scale across the Canadian Prairies from 2010 to 2016. Top predictors were identified and regression models were built using a robust least angle regression (RLARS) and leave-one-out cross-validation (LOOCV) scheme. Selected predictors revealed that canola yields across the Canadian Prairies are highly sensitive to soil moisture conditions throughout the growing season and in particular, during week 24. In addition, soil moisture was found to provide a better descriptor of canola stresses than the more widely utilized NDVI, being selected as a predictor in 74.2% of
developed ecodistrict models over the 7-year period, compared to just 41.2% for NDVI. The contribution of SMOS soil moisture data to the model’s forecast skill was determined by calculating the difference between model $R^2$ values (i.e. $R^2_{\text{diff}}$) when SMOS soil moisture predictors were included and excluded from the forecast, respectively. The results showed varying degrees of model improvements when SMOS soil moisture observations were included as potential predictors within the ICCYF; $R^2_{\text{diff}}$ values ranged from -0.42 to +0.43 across the study area. However, the majority of ecodistricts under study (53.3%) showed improved model fit (i.e. $R^2_{\text{diff}} > 0$) with observed canola yields when SMOS soil moisture indices were included in the model. Overall, greater improvements in the ICCYF performance were observed in Manitoba and Saskatchewan where meteorological stations are more sparsely distributed. However, performance for both sets of model inputs was relatively low with $R^2$ values ranging from 0 to 0.74 (mean = 0.13) and from 0 to 0.52 (mean = 0.12) across the study area when soil moisture was included and excluded from the model, respectively. These findings suggest that while SMOS soil moisture observations may provide a more effective indicator of canola yields, the ICCYF’s performance at the township-scale, where interannual yield variability is often quite high, is limited by the short temporal satellite record.

### 3.1 Introduction

As climate variability intensifies and the global demand for food increases, there exists a growing need to improve the accuracy and robustness of crop yield forecasting systems (IPCC, 2013; Basso et al., 2013; Holzman et al., 2014; Kogan et al., 2013). Crop yield forecasts are extremely important for determining potential and actual food losses that can aid government, private and producer groups in the development of export-import policies, food security policies, and efficient land management practices. However, traditional crop yield forecasts, conducted
through farm surveys (e.g. Statistics Canada, 2018a) or by experts in the field based on their
evaluation of crop statuses, are extremely subjective and time-consuming and often
unrepresentative due to small sample sizes (Basso et al., 2013; Statistics Canada, 2018a).

Satellite remote sensing offers several benefits over traditional, ground-based methods
for forecasting crop yields, including greater spatial and temporal coverage (Barrett and
Petropoulos, 2014; Judge, 2007; Patel et al., 2006; Ren et al., 2012), reduced costs (Basso et al.,
2013; Boken and Shaykewich, 2002; Maas, 1988; Newlands et al., 2014), and the elimination of
human-related biases and errors (Chipanshi et al., 2015; Mkhabela et al., 2011). Leaf area index
(LAI), green area index (GAI), meteorological parameters (e.g. temperature, rainfall,
evapotranspiration) and various vegetation indices (VIs) (e.g. normalized difference vegetation
index (NDVI), normalized difference water index (NDWI), enhanced vegetation index (EVI))
are commonly derived from remote-sensing platforms and utilized in yield forecasting models
(Kogan et al., 2013; Maas, 1988). VIs in particular, are widely utilized for forecasting crop yields
as they can provide valuable information regarding vegetation vigour and drought conditions that
can be utilized in the evaluation of crop health throughout the growing season (Peng et al.,
2014). The most common VI, the NDVI, is implemented in several yield forecasting models
around the world including China’s Crop Watch (Wu et al., 2014), the European Commission
Joint Research Centre’s (JRC) Monitoring Agriculture with Remote Sensing (MARS) Crop
Yield Forecasting System (MCYFS; van Diepen et al., 2004), the Crop Condition Assessment
Program (CCAP) of Statistics Canada (Reichert and Caissy, 2002), and Agriculture and Agri-
Food Canada’s (AAFC) Integrated Canadian Crop Yield Forecaster (ICCYF; Chipanshi et al.,
2012).
The ICCYF is a geospatial modelling tool integrating agroclimatic variables and near-real
time NDVI values derived from the National Oceanographic and Atmospheric Administration’s
(NOAA) Advanced Very High Resolution Radiometer (AVHRR) platform to produce regional
yield forecasts for major Canadian crops across the agricultural landscape of Canada (AAFC,
2015). Historical and near-real time climate, NDVI and crop field survey data are utilized to
generate in-season probabilistic yield forecasts which are updated throughout the growing season
as more data becomes available (Chipanshi et al., 2015). Past studies have shown significant
success with the ICCYF model generating skillful forecasts for spring wheat, barley and canola
approximately 1 month before harvest, and 3-4 months before the official release of Statistics
Canada’s survey results (Chipanshi et al., 2015; Newlands et al., 2014). However, the wide-
spread use of VIIs in yield forecasting models has inherent limitations due to the reliance on
visible and near-infrared portions of the electromagnetic spectrum; these portions of the
electromagnetic spectrum are unable to penetrate through cloud cover, heavy rain, or fog, and
therefore, the availability of cloud-free, clear-sky data during optimal growing periods is a
significant limitation to their use in crop yield forecasting (Peng et al., 2014).

Several studies in recent years have investigated the potential of utilizing microwave
satellite observations of soil moisture for monitoring agricultural systems (e.g. Chakrabarti et al.,
2014; Champagne et al., 2016, 2015, 2012, 2011; Djamai et al., 2015; Ines et al., 2013; Liu et al.,
2016; Nearing et al., 2012). Unlike optical or thermal imaging satellites, microwave platforms
are able to monitor the Earth’s surface under all-weather conditions, day or night, and are less
heavily impacted by dense vegetation canopies and surface roughness, making them well-suited
for agricultural applications (Engman and Chauhan, 1995; Judge, 2007; Kerr et al., 2010).
Furthermore, soil moisture is a key variable in the determination of crop yields in many rain-fed
regions around the world, including the Canadian Prairies, as it directly controls the amount of water and nutrients available to support crop growth and development (Champagne et al., 2012; Dobriyal et al., 2012; Holzman et al., 2014; Judge, 2007; Maybank et al., 1995; McGinn and Shepherd, 2003; Nadler, 2007). However, direct measurements of soil moisture have been largely excluded from crop yield forecasting due to the challenges historically associated with soil moisture quantification (Bornn and Zidek, 2012; Holzman et al., 2014).

Soil moisture is extremely heterogeneous, spatially and temporally, due to varying soil types, vegetation cover, topography, climate conditions, and land management practices (Barrett and Petropoulos, 2014; Dobriyal et al., 2012; Engman and Chauhan, 1995; Holzman et al., 2014). Traditional, in-situ point-measurements of soil moisture therefore, lack the spatial density required to capture soil moisture variability present on a larger scale (Champagne et al., 2012, 2011; Dobriyal et al., 2012; Holzman et al., 2014). As a result, indirect estimates of soil moisture are often derived from water balance models using temperature and precipitation inputs for use in yield forecasting models (Baier and Robertson, 1996; Bornn and Zidek, 2012; Robock et al., 2000). However, recent developments in microwave remote sensing platforms, such as the microwave L-band (1.41 GHz) Soil Moisture Ocean Salinity mission (SMOS; Kerr et al., 2010) satellite, now offer an efficient and accurate means for acquiring large-scale surface (depth < 5 cm) soil moisture data at frequent enough temporal resolutions (1-3 days) for near-real time monitoring (Barrett and Petropoulos, 2014; Champagne et al., 2012, 2011; Kerr et al., 2012). Therefore, the objective of this study is to investigate the use of remotely-sensed soil moisture observations obtained by the SMOS satellite for forecasting canola yields across the Canadian Prairies within the ICCYF model.
3.2 Methods

3.2.1 Study Region

This study encompasses the canola growing regions of the Canadian Prairies (Figure 3.1) which are defined using AAFC’s (2017b) spatial density of major crops dataset. This dataset shows the areas where major crops can be expected across the agricultural regions of Canada based on the temporal and spatial frequency of the crop between 2009 and 2016 (AAFC, 2017b). Restricted to the most southerly portions of the three Prairie provinces, this area extends westward from south-central Manitoba (97°W) to western Alberta (120°W) and northward from the U.S border (49°N) to the Peace River region in Alberta (58°N).

The three Prairie provinces account for 99% of the total canola seeded area and production in Canada each year (Statistics Canada, 2017a). Saskatchewan is the largest producer of canola in the country, accounting for roughly 49.4% of the total canola production, followed

![Figure 3.1: Census Agricultural Regions (CARs), ecodistricts, townships and the major canola producing regions across the Canadian Prairies. Selected climate stations are represented by the black triangles.](image)
by Alberta (34.6%) and Manitoba (15%), while the remaining 1% is produced in British Columbia (0.4%), Ontario (0.3%), Quebec (0.2%) and New Brunswick (0.1%) (Statistics Canada, 2017a). The Canadian Prairies are largely characterized by rich, fertile soils that support the highly-diverse and extensive agricultural production occurring across the region. Soils consist mainly of black/dark-gray, dark-brown and brown Chernozemic soils corresponding to three agro-climatic zones: Sub-humid, semi-arid and arid, respectively (Mkhabela et al., 2011; Nadler, 2007). The most arid regions are found in southern Alberta and south-western Saskatchewan, transitioning to more semi-arid and sub-humid climates moving to the north and east (Government of Canada, 2006). The climate of the Canadian Prairies is predominately described as continental with long, cold winters, short, hot summers and relatively low precipitation amounts (Government of Canada, 2006; Mcginn, 2010). Average winter and summer temperatures are -10°C and 15°C, respectively (Environment and Climate Change Canada, 2017a), while temperatures in July, the warmest month, vary from an average of 15.9°C in Edmonton, 18.8°C in Regina, and 19.5°C in Winnipeg (Environment and Climate Change Canada, 2017a). Annual precipitation amounts range from just 250 mm in the arid grasslands of southern Alberta and south-western Saskatchewan to a maximum of 500 mm in the sub-humid Manitoba lowlands, making this region extremely prone to agricultural drought (Government of Canada, 2006; Maybank et al., 1995; McGinn and Shepherd, 2003; Nadler, 2007).

3.2.2 Input Data

3.2.2.1 Yield Data

The Canadian Prairies are administratively divided into a 36-square mile (~10x10 km) township grid system (Figure 3.1) (Government of Canada, 2018). Each township is composed of 36 1-square mile sections, which are further sub-divided into quarter sections (i.e. field-level,
Canola yields in kilograms per hectare across the Canadian Prairies were acquired from AAFC at the township-scale (~10x10 km). The township yield data is interpolated and aggregated from quarter section yields of major crops in kilograms per acre that are provided to AAFC by private crop insurance agencies in Manitoba, Saskatchewan, and Alberta and is available from 2010 onwards. Canola yields were obtained from 2010 to 2016 for each township within the canola growing regions of the Canadian Prairies in order to coincide with the SMOS data record (i.e. 2010 to present). A total of 6,017 reporting townships were identified within the canola growing regions of Manitoba, Saskatchewan and Alberta.

3.2.2.2 Agroclimate Data

Agroclimate variables were derived from station-based daily temperature and precipitation data collected by Environment and Climate Change Canada and other partner institutions through the Drought Watch program operated by the National Agroclimate Information Service (NAIS) of AAFC (2018a). A total of 324 climate stations across Manitoba, Saskatchewan and Alberta were used to represent the climate of the 6,017 townships under study (Figure 3.1).

Daily maximum and minimum temperature and precipitation data for each of the 324 climate stations was obtained from Environment and Climate Change Canada from 2010 to 2016 and was quality controlled and gap-filled by AAFC. The data was then input into the Versatile Soil Moisture Budget (VSMB) model (Baier et al., 2000) at a daily time step along with soil physical parameters (e.g. available soil water holding capacity, AWHC) obtained from the Canadian Soil Information Service (CanSIS) database (2018) in order to generate the agroclimatic indices to be used in the ICCYF (Figure 3.2). A biometeorological time scale
(Robertson, 1968) using spring wheat crop parameters and water extraction coefficients (Akinremi et al., 1996) was applied to quantify generalized canola phenological development stages within the VSMB model. For the purposes of this study, the response of canola to soil and atmospheric environments was assumed to be similar to that of spring wheat. This assumption was proven valid in a previous study conducted by Chipanshi et al. (2015) in which the ICCYF was used to forecast spring wheat, barley and canola yields across the agricultural landscapes of Canada. The effective growing-degree days (EGDD) above a base temperature of 5°C, soil water availability (SWA), crop water stress index (SI) and crop seeding date were generated for each climate station by the VSMB model at a daily time step to be included as potential predictors within the ICCYF. The crop water stress index was defined within the VSMB model as:

\[
SI = 1 - \frac{AET}{PET}
\]  
[3.2]

where AET and PET are actual and potential evapotranspiration, respectively (Moran et al., 1994), and computed using the Baier and Robertson (1996) algorithm. Soil water availability was defined as the percentage of the AWHC at any given time (Newlands et al., 2014).

Other agroclimate variables included in the ICCYF model as potential predictors were: Precipitation (P), the number of frost days (FrostD; daily minimum temperature below 2°C) and the number of heat stress days (HeatD; daily maximum temp above 30°C). All daily agroclimatic indices from May to August were then temporally-summed (EGDD, P, FrostD, HeatD) or -averaged (SWA, SI) by month across each of the 7 growing seasons under study. The standard deviations of daily precipitation, EGDD and SI over each month were also calculated and used as potential predictors. The resulting indices were then assigned to the 6,017 townships across the study area. For townships with no climate stations, the nearest climate station was used.
3.2.2.3 AVHRR NDVI Data

Weekly NDVI composites derived from the AVHRR platform are available from the NOAA’s (2016) Global Vegetation Index (GVI) dataset. The NDVI is based on the inherent absorption of visible red wavelengths ($\rho_{RED}$; 620-670 nm) by chlorophyll pigments in the palisade layer of a healthy green plant leaf and the reflection of near-infrared ($\rho_{NIR}$; 841-876 nm) wavelengths within the spongy mesophyll (Campbell and Wynne, 2011). It is defined as:

$$\text{NDVI} = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$$

[3.3]

with values ranging from -1 for non-vegetated surfaces (e.g. water, snow, ice or cloud cover) to +1 for healthy, green vegetation, while stressed vegetation will display positive NDVI values approaching zero (Campbell and Wynne, 2011; NOAA, 2016).

AVHRR-NDVI weekly composites spanning the growing season were acquired for Julian Weeks 18 to 40 (i.e. May to the end of September) from 2010 to 2016 with a spatial resolution of 1 km across the agricultural regions of the Canadian Prairies. The raw data is quality controlled and cloud-affected pixels are identified and removed prior to generating the weekly composites (Reichert and Caissy, 2002). AAFC’s (2017b) spatial density of major crops dataset for canola was applied in order to restrict the AVHRR-NDVI data to the canola growing regions of the Canadian Prairies, removing any non-agricultural and non-canola pixels from the analysis (Figure 3.2). The resulting weekly NDVI composites were aggregated to the township-level and three-week running means were calculated and included as potential predictors in the ICCYF. The maximum weekly NDVI value over the growing season for each township was also extracted and included as a potential predictor.
3.2.2.4 Satellite Soil Moisture Data

Launched in November 2009 by the European Space Agency (ESA; Kerr et al., 2012), the SMOS satellite is the first satellite designed to monitor soil moisture and sea surface salinity on a global scale (Mecklenburg et al., 2012). The satellite features a 2-D interferometric Microwave L-band (1.41 GHz) Imaging Radiometer using Aperture Synthesis (MIRAS) that passively captures naturally emitted microwave energy (i.e. microwave brightness temperature, $T_B$) from the Earth’s surface across a 1000 km swath every 1-3 days with a spatial resolution of 35-50 km (Kerr et al., 2010). The measured emittance can be used to estimate surface soil moisture content due to the large difference in the dielectric constant of dry soil (~3.5) and water (~80); as the soil moisture content increases, the dielectric constant of the soil can increase from roughly 3.5 (i.e. dry soil) to a value of 20 or greater when saturated, therefore increasing the attenuation of microwave energy (Barrett and Petropoulos, 2014; Schmugge, 1983).

Level 2 weekly SMOS soil moisture data for North America, available from AAFC (2017c), was produced by the ESA using version 6.20 of the SMOS soil moisture processor and re-gridded to a 0.25-degree resolution. The SMOS soil moisture processor uses an iterative approach to quantify soil moisture and vegetation opacity through minimizing a cost function, given by the sum of the squared weighted differences between measured and modelled brightness temperature data, by iteratively adjusting values for the soil dielectric constant (Kerr et al., 2012). The dataset is created on a daily timestep and is averaged to 7-day periods using the International Organization for Standardization (ISO) week (AAFC, 2017c). Areas where rainfall and/or snow cover is present and where the Radio Frequency Interference (RFI) is above an acceptable threshold are removed through quality flags (AAFC, 2017c). Weekly percent volumetric surface (depth < 5 cm) soil moisture data (AAFC, 2017c) for North America was
obtained for Weeks 18 to 40 between 2010 and 2016 as a series of GeoTIFF rasters and aggregated to the township-level (~10x10 km) using a weighted averaging technique adopted from Koster et al. (2000). Average soil moisture, $S_i$, in township $i$ was defined in ArcGIS 10.2© as:

$$S_i = \frac{\sum_n S_n A_{ni}}{\sum_n A_{ni}} \tag{3.4}$$

where $n$ corresponds to the number of SMOS pixels intersecting township $i$, $S_n$ is the weekly percent volumetric surface soil moisture for SMOS pixel $n$ and $A_{ni}$ is the fractional area of township $i$ occupied by SMOS pixel $n$ (Koster et al., 2000).

The aggregated weekly SMOS data was then masked to the canola growing regions using AAFC’s (2017b) spatial density of major crops dataset for canola and included in the ICCYF model (Figure 3.2). Three-week running means were also calculated and included as potential predictors. Samples (i.e. Township-year combinations) missing weekly SMOS soil moisture data for any week over the growing season were removed from the analysis; of the 41,611 samples (i.e. Township-year combinations) considered in this study, 1,922 were missing soil moisture data for at least one week during the growing season, of which 83 townships (out of 6,017 townships) were missing soil moisture data for at least one week across all reporting years, leaving a total of 5,934 townships and 39,689 township-year samples to be used in the analysis.

3.2.3 The ICCYF Model

The ICCYF of Agriculture and Agri-Food Canada provides regional yield forecasts for six major Canadian crops: Spring wheat, durum wheat, barley, corn, soybeans and canola (AAFC, 2015). The model integrates historical agroclimate variables, crop yields, and near-real time remote sensing data (i.e. NDVI) within a geographic information system (GIS) in order to
generate probabilistic yield forecasts throughout the growing season. Currently, forecasts within the ICCYF are performed at the Census Agricultural Region (CAR)-level. CARs are utilized for disseminating agricultural statistics (e.g. number of farms, farm capital, yield) by the Census of Agriculture (Statistics Canada, 2015) and can range in size from 4,726 to 1,509,850 km², spanning multiple climate and soil zones (Kouadio et al., 2014; Mkhabela et al., 2011). As a result, crop yield responses to environmental factors can vary considerably within a single CAR (Chipanshi et al., 2015; Kouadio et al., 2014; Zhang et al., 2017).

In this study, the ICCYF was calibrated to forecast canola yields across the Canadian Prairies at the township-scale (~10x10 km). While still relying on administrative boundaries, previous studies have shown that the smaller township-scale is able to capture most of the yield variability with a forecast skill comparable or better than that of the CAR-level (Zhang et al., 2017, 2015). However, modelling at the township-scale is limited by the availability of yield data as township-level yields are only available from 2010 onwards, while modelling at the CAR-scale is made over a period of approximately 30 years using historical yield data to drive the model. Using a more spatially dense but temporally short data record allows for grouping of townships based on similarities in the key predictors of crop yield, thus increasing the number of samples that are available to build a statistically robust model. Therefore, townships were grouped by ecodistricts in order to increase the number of observations in the short data record and generate yield forecasts at the township-scale. An ecodistrict is defined as a subdivision of an ecoregion characterized by relatively homogenous climate, soil, landscape and ecological characteristics (AAFC, 2013). A total of 216 ecodistricts were used to group the 5,934 townships across the study area and a single yield model representing the yield-environment response relationships of the townships was developed for each ecodistrict.
The ICCFY crop yield modelling process is depicted in Figure 3.2 and was coded using R Statistical Language (Version 3.0.3, R Development Core Team) (R Foundation, 2014). Two input datasets were utilized to generate yield forecasts for canola across the Canadian Prairies and evaluate the potential of SMOS soil moisture observations as a potential predictor: 1) agroclimate plus NDVI indices (i.e. excluding SMOS soil moisture observations); and 2) agroclimate plus NDVI and SMOS soil moisture indices. For both model inputs, all potential predictors (70 for agroclimate and NDVI input and 114 for agroclimate, NDVI and soil moisture input) were subjected to a robust least angle regression scheme (RLARS; Efron et al., 2004; Khan et al., 2007) in order to evaluate and rank the correlations between canola yields and each of the predictor variables. A maximum of 5 predictors were selected for each ecodistrict. A robust leave-one-out cross-validation (LOOCV) scheme (Khan et al., 2010) was then applied to finalize the predictors and coefficients of each yield forecasting model by removing any false predictors selected from contaminated data (i.e. outliers; Chipanshi et al., 2015; Khan et al., 2010). A spatial correlation analysis among neighbouring ecodistricts was then conducted using the Bayesian statistical approach, as described by Bornn and Zidek (2012), in order to improve and stabilize the model’s prediction performance. Historical data for the statistically-selected neighbouring ecodistricts were used along with the forecasting ecodistrict’s data in order to generate the prior distribution of the predictors. The posterior distribution of the predictors was obtained using a Markov-Chain Monte Carlo (MCMC) scheme (Dowd, 2006) fed by the prior distributions and near-real time data obtained at the time of forecast. Further information on the ICCYF modelling methodology can be found in Newlands et al. (2014).

A defining characteristic of the ICCYF is the ability to generate forecasts at any point throughout the growing season. Thus, for in-season forecasting, random forest machine learning
Figure 3.2: Data and modelling workflow of the Integrated Canadian Crop Yield Forecaster (ICCYF). Adapted from Chipanshi et al. (2015).
(Liaw and Wiener, 2002) is applied to estimate the unobserved variables required to make a yield forecast by a specified date. The estimated variables are then input into the yield forecasting model along with the variables observed in near-real time in order to forecast the yield probability distribution for each township. For this analysis however, end-of-season forecasts (i.e. September) were performed and thus, no variable estimation was required. End-of-season forecasts were chosen for this study as they incorporate all available information and therefore, due to the nature of the ICCYF model, represent the best forecast results achieved (Chipanshi et al., 2015). Typically, in crop yield modelling, forecasted yields will get closer to the actual yields later in the growing season; however, due to the selection algorithm incorporated into the ICCYF, the developed yield model may only contain predictors during critical stages (e.g. during stand establishment or flowering) and therefore, predictors obtained later in the growing season will have no effect on the forecasted yields or errors (Kouadio et al., 2014).

Overall, the final yield model is a multivariate linear regression equation, defined as:

\[ Y_t = \alpha_0 + \alpha_1 t + \sum_{i=2}^{n} \alpha_i X_i + \epsilon_t \]  

[3.5]

where \( Y_t \) is the crop yield for year \( t \), \( \alpha_0 \) is the regression intercept, \( \alpha_1 t \) represents the technology trend over time, \( \epsilon_t \) is the error term and \( X_i \) is the predictor \( i \) in year \( t \), where \( i \) could be any of \( n \) potential predictors (Chipanshi et al., 2015). Yield models were generated for all ecodistricts with a minimum of 12 samples (i.e. township-year combinations) available for model building. A total of 1,382 ecodistrict models were developed for 200 ecodistricts over the 7-year period.
3.2.4 Model Evaluation and Assessment of SMOS Soil Moisture Performance

Given the limited availability of data used in this study (i.e. 7-year SMOS data record), a leave-one-year-out cross validation (LOOCV), as performed by Mkhabela et al. (2011), Kouadio et al. (2014), and Chipanshi et al. (2015), was applied in order to evaluate the ICCYF’s performance. For each of the seven years under study, data for a single year (i.e. forecast year) was set aside while the remaining six years of data (i.e. training data) were used to calibrate the ecodistrict yield models. Forecasts were then generated at the township-scale for the forecast year using the observed variables and the developed one-year-removed ecodistrict models.

Model forecast skill was evaluated for each ecodistrict using the Bravais and Pearson coefficient of determination ($R^2$). $R^2$ describes the proportion of the of the observed yield variance that can be explained by the model predictors (Chipanshi et al., 2015) and is defined as:

$$ R^2 = \left( \frac{\sum_{i=1}^{n} (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n} (O_i - \bar{O})^2 \sum_{i=1}^{n} (P_i - \bar{P})^2}} \right)^2 \tag{3.6} $$

where $O_i$ and $P_i$ are the observed and forecasted yields for township $n$ in forecast year $i$, respectively, while $\bar{O}$ and $\bar{P}$ represent the mean observed and forecasted yields across the ecodistrict over the 7 years of study. Values for $R^2$ lie between 0 and 1, where 0 indicates that the model is unable to explain any of the observed yield variability and 1 indicates that the model accounts for all the observed yield variation. Higher $R^2$ values therefore indicate a greater model fit. $R^2$ values were only calculated for ecodistricts in which forecasts were generated for more than one year in order to eliminate any erroneous zero $R^2$ values. Three ecodistricts (612, 827 and 833) were found to have only one yield forecast generated and thus, removed from the
model evaluation, leaving 197 ecodistricts with which to evaluate the model’s performance.

In order to quantify the contribution of SMOS soil moisture observations to the model’s forecast skill, the difference in $R^2$ values ($R^2_{\text{diff}}$) between the two inputs was calculated for each ecodistrict. Following the approach of Koster et al. (2011, 2010) and Thomas et al. (2016), the $R^2_{\text{diff}}$ was calculated by subtracting the $R^2$ value associated with the agroclimate and NDVI input (i.e. without SMOS, $R^2_{\text{WO SMOS}}$) from the $R^2$ value of the agroclimate, NDVI and SMOS input ($R^2_{\text{SMOS}}$):

$$R^2_{\text{diff}} = R^2_{\text{SMOS}} - R^2_{\text{WO SMOS}}$$

where a positive $R^2_{\text{diff}}$ indicates an improvement in the ICCYF model performance when SMOS soil moisture indices are included as predictors and a negative $R^2_{\text{diff}}$ indicates a decline in the model’s performance.

3.3 Results and Discussion

3.3.1 Selected Predictors

The predictors selected by each ecodistrict to model the township-level canola yields varied across the study area and according to the input dataset used. When only agroclimate and NDVI indices were included as potential predictors, the standard deviation and/or sum of the effective growing degree days (EGDD) from May to August were among the most frequently selected predictors (Table 3.1). The standard deviation of the crop water stress index in May (SDSI_5), average NDVI over weeks of the year 28 to 30 (NDVI28_30), the maximum NDVI value observed during the growing season (NDVI_MAX) and the standard deviation of precipitation in July (SDPcpn_7) were also among the predominant predictors. However, when SMOS soil moisture observations were included in the ICCYF model, soil moisture during week 24 (SM24) was the top predictor (Table 3.1). Other leading predictors included soil moisture
Table 3.1: Top ten predictors selected by ecodistrict regression models based on input datasets including agroclimate and NDVI indices alone (Agroclimate + NDVI), and agroclimate, NDVI and soil moisture indices (Agroclimate + NDVI + SM). Agroclimate variables include the effective growing degree days (EGDD), the crop water stress index (SI) and precipitation (P). Numbers following these indices (i.e. 5,6,7 and 8) represent May, June, July and August, respectively. SD and Sum prior to the indices represent whether it is the standard deviation or temporal sum. For three-week average NDVI indices, the numbers following refer to the weeks of the year used for the averaging calculation. Similarly, for soil moisture indices, the number following refers to the week of the year the observation was made.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Agroclimate + NDVI</th>
<th>Agroclimate + NDVI + SM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SDEGDD_7</td>
<td>SM24</td>
</tr>
<tr>
<td>2</td>
<td>SDEGDD_5</td>
<td>SDEGDD_7</td>
</tr>
<tr>
<td>3</td>
<td>SumEGDD_5</td>
<td>SumEGDD_5</td>
</tr>
<tr>
<td>4</td>
<td>SDEGDD_8</td>
<td>NDVI28_30</td>
</tr>
<tr>
<td>5</td>
<td>SDEGDD_6</td>
<td>NDVI_MAX</td>
</tr>
<tr>
<td>6</td>
<td>SDSI_5</td>
<td>SDEGDD_5</td>
</tr>
<tr>
<td>7</td>
<td>SumEGDD_7</td>
<td>SM40</td>
</tr>
<tr>
<td>8</td>
<td>NDVI28_30</td>
<td>SDSI_5</td>
</tr>
<tr>
<td>9</td>
<td>NDVI_MAX</td>
<td>SM29</td>
</tr>
<tr>
<td>10</td>
<td>SDPcpn_7</td>
<td>SumEGDD_7</td>
</tr>
</tbody>
</table>

during weeks 29 and 40 (SM29, SM40), as well as the standard deviation and sum of the EGDD in May and July, average NDVI over weeks of the year 28 to 30 (NDVI28_30), the maximum weekly NDVI value observed over the growing season (NDVI_MAX) and the standard deviation of the crop water stress index in May (SDSI_5).

The predictors selected when SMOS soil moisture observations were not included in the model (i.e. agroclimate and NDVI indices alone) are in general agreement with those identified in a previous study by Chipanshi et al. (2015). In their study, the standard deviation and sum of the EGDD in July and August, the crop water stress index and precipitation in July, and NDVI from July through August were identified as the top predictors of canola yields across the Canadian Prairies at the CAR-level using the ICCYF and historical data from 1985 to 2012.
These results therefore suggest that the 7-year period under study provided sufficient data to accurately identify the variables accounting for the majority of the variance observed in the canola yield data.

The recurrence of temperature-related indices (e.g. SumEGDD, SDEGDD), both when SMOS soil moisture observations were included and excluded from the model, reveals that temperature plays a significant role in the determination of canola yields across the Canadian Prairies, similar to the findings of previous studies. Canola is a cool-season crop, growing best at temperatures between 10 and 30°C (Ontario Ministry of Agriculture, Food and Rural Affairs, 2016); therefore, extreme temperatures have been found to have a profound effect on canola production. Kutcher et al. (2010) for example, reported a 12% reduction in canola yields across Saskatchewan for every 7 days of maximum temperatures greater than 30°C. In addition, yield losses between 1967 and 2001 were found to be strongly associated with daily minimum and maximum temperatures in excess of 16 and 31°C, respectively, at the beginning of July, coinciding with the flowering stage of development (Kutcher et al., 2010). Similar findings were reported by White et al. (2018), in which nocturnal temperatures (i.e. Tmin) greater than 15°C in mid-July were found to negatively impact canola yields across Saskatchewan between 2010 and 2015. In this study, temperatures (i.e. EGDD) in July were also identified as important meteorological factors controlling canola yields, being selected amongst the top predictors for both model inputs (Table 3.1). However, temperatures in May were also found to be crucial in the determination of canola yields (Table 3.1). This is likely due to cooler-than-average temperatures that were present across the Canadian Prairies at the beginning of the growing season in four of the seven years under study (i.e. 2010, 2011, 2013, 2014), which resulted in delayed seeding and thus, crop development (Barthet, 2014, 2013, 2011, 2010).
The selected predictors indicate that soil moisture was the number one factor controlling canola production across the Canadian Prairies between 2010 and 2016 (Table 3.1). Soil moisture is widely regarded as one of the main determinants of crop yields in arid regions around the world due to its impacts on water availability, nutrient cycling and field accessibility for seeding, harvest and field management (Champagne et al., 2016, 2015; Dobriyal et al., 2012). Numerous studies have shown that both excess and deficit soil moisture conditions can have a significant impact on the development of most crops (Haferkamp, 1988; Holzman et al., 2014; Kanwar et al., 1988; Vough and Marten, 1971). Water stress for example, has been found to limit leaf development, therefore decreasing the crop’s photosynthetic area and stunting crop growth (Slayter, 1974; Vough and Marten, 1971), while wet soils throughout the growing season can hinder the exchange of important gases between the roots and soil, impacting nutrient cycling and root development (Canola Council of Canada, 2013; Holzman et al., 2014; Kanwar et al., 1988).

In the current study, excess soil moisture conditions played a significant role in the determination of canola yields as the 7-year period under study experienced abnormally wet conditions unrepresentative of historical climate norms (AAFC, 2017a); departure from average growing season precipitation (AAFC, 2017a) revealed surpluses in excess of 120 mm spanning significant portions of the canola growing regions across the Canadian Prairies in four of the seven growing seasons under study (i.e. 2010, 2012, 2014 and 2016). Of the 200 ecodistricts considered in this study, 193 selected SMOS soil moisture indices as a predictor of canola yields at least once over the 7-year period, while 72 ecodistricts consistently selected SMOS soil moisture indices every year of the study (Figure 3.3). The 7 ecodistricts that did not select SMOS soil moisture indices were primarily located along the boundaries of the canola growing regions.
(Figure 3.3, top) and therefore, had lower percentages of canola crop land area (Figure 3.7, bottom). The 72 ecodistricts that consistently selected SMOS soil moisture indices however, showed distinct spatial clustering in eastern Saskatchewan and on the leeward side of the Rocky

Figure 3.3: Ecodistricts with SMOS soil moisture indices as predictors of canola yields. A) Ecodistricts that included SMOS soil moisture indices as a predictor at least once over the 7-year period under study; and B) Ecodistricts that consistently selected SMOS soil moisture indices as a predictor every year.
Mountains in Alberta (Figure 3.3, bottom). Eastern Saskatchewan in particular, was heavily impacted by excess moisture over the 7-year period under study with ponding early in the growing season resulting in delayed seeding in several years, with some areas being designated as too wet to seed (Barthet, 2014, 2012, 2011, 2010; Champagne et al., 2015). In Alberta, increased moisture likely had a positive impact on canola production during this period of greater-than-average precipitation as soil moisture is often a key factor limiting crop yields across the province where precipitation amounts during the growing season are, on average, significantly less than crop water demands (Alberta Agriculture and Forestry, 2015).

While NDVI is widely utilized in yield forecasting models, the results of this study indicate that SMOS soil moisture observations provide a more accurate descriptor of canola stress across the Canadian Prairies than NDVI. Over the 7-year period under study, the number of ecodistrict models that selected NDVI as a predictor of canola yields varied from just 41.2% to 43.1% when SMOS soil moisture observations were included and excluded from the model, respectively, while SMOS soil moisture indices were selected as a predictor in 74.2% of developed ecodistrict models. This relatively low occurrence of NDVI predictors suggests that NDVI may not be an effective indicator of canola yields across the Canadian Prairies, similar to the findings of Chipanshi et al. (2015). Using the ICCYF at the CAR-level, Chipanshi et al. (2015) found that only 6 of the 28 CARs in Alberta and Saskatchewan selected NDVI as a predictor of canola yields. This is likely due to the fact that canola, and other Brassica oilseeds, have a distinct “yellow-up” that occurs with the appearance of bright yellow flowers (Sulik and Long, 2016, 2015) which contributes green and red light to the canopy-level signal (Behrens et al., 2006; Shen et al., 2009). As a result, the ability of NDVI to adequately capture canola biomass for use in yield forecasting models is impacted by the confounding spectral signals of
the reproductive flowers during the flowering stage of development (Shen et al., 2010; Sulik and Long, 2015). Only when the flowers begin to wilt and fall-off (i.e. yellow-down) during maturation (mid- to end-July) is NDVI able to provide an effective indication of canola yields (Sulik and Long, 2016), as seen in the current study. Therefore, other remote sensing-derived indices, such as soil moisture, may be more relevant for forecasting canola yields.

3.3.2 Soil Moisture Timing

The frequency of the selected soil moisture predictors throughout the growing season is illustrated in Figure 3.4. The percent occurrence corresponds to the percentage of ecodistrict models that selected a specific soil moisture index as a predictor of canola yields. A total of 1,382 ecodistrict models were developed over the 7-year period under study.

The selected predictors reveal that canola yields across the Canadian Prairies are highly sensitive to soil moisture conditions throughout the growing season. Soil moisture during week 24 in particular, was found to play a significant role in the determination of canola yields being selected as a predictor in 15.1% of the developed ecodistrict models (Figure 3.4). Soil moisture during weeks 40, 29 and 25 were also identified as crucial time periods for canola development, being selected as a predictor of canola yields in 9.7, 8.2, and 7.7% of the developed ecodistrict models.

![Figure 3.4: Percent occurrence of SMOS soil moisture indices in developed ecodistrict yield models. A total of 1,382 ecodistrict models were developed between 2010 and 2016.](image-url)
models, respectively. These findings are in general agreement with those of White et al. (2018) in which canola yields across Saskatchewan between 2010 and 2015 were found to exhibit significant associations with excess soil moisture conditions throughout the growing season in low-yielding years. The beginning of June however, coinciding with the stand establishment stage, was identified as a critical time period in which canola yields were most adversely affected by excess soil moisture (White et al., 2018); outside this time period, excessive soil moisture conditions were found to play a significant, but smaller role, in canola production (White et al., 2018), similar to the findings of the current study. The weakest associations between soil moisture and canola yields were observed from mid- to end-August with chi-square values ranging from just 12.3 to 19.6 (White et al., 2018). The current study also reveals that soil moisture conditions during this time period play a much smaller role in the determination of canola yields; week 34 soil moisture was only selected as a predictor in 0.6% of the developed ecodistrict models, while soil moisture in weeks 33 and 35 were selected in 1.8 and 2.2%, respectively (Figure 3.4).

3.3.3 Comparison of Yield Model Performance

A comparison of the Bravais and Pearson coefficient of determination ($R^2$) reveals improved model performance across the majority of the Canadian Prairies when SMOS soil moisture observations are included as potential predictors within AFFC’s ICCYF (Figure 3.5, Figure 3.6). When only agroclimate and NDVI indices (i.e. excluding soil moisture) were used as inputs in the ICCYF, $R^2$ values ranged from 0 to 0.52 across the 197 ecodistricts considered in the model evaluation (Figure 3.5, top). However, when SMOS soil moisture indices were included in the ICCYF model, $R^2$ values across the study area ranged from 0 to 0.74 (Figure 3.5, bottom). An $R^2_{\text{diff}}$ test revealed that approximately 53.3% of ecodistricts exhibited improved
model fit (i.e. $R^2_{\text{diff}} > 0$) with observed canola yields when SMOS soil moisture observations were included as potential predictors within the ICCYF, while correlation improvements (i.e. $R^2_{\text{diff}}$) greater than 0.10 were observed in 39 (19.8%) ecodistricts (Figure 3.6). In contrast,

**Figure 3.5:** Bravais and Pearson coefficient of determination ($R^2$) across the Canadian Prairies for A) developed ecodistrict-level yield models including agroclimate and NDVI indices; and B) developed ecodistrict-level yield models including agroclimate, NDVI and soil moisture indices. Black hatches denote statistically significant $R^2$ values at the 95% confidence level ($p < 0.05$).
declines in model performance (i.e. $R^2_{\text{diff}} < 0$) were observed in 46.7% of ecodistricts, of which 24 (12.2%) exhibited correlation declines greater than 0.10 (Figure 3.6).

Spatially, higher $R^2$ values were observed across Saskatchewan and Manitoba for both sets of model inputs, while the majority of ecodistricts in Alberta exhibited poorer model fit (Figure 3.5). These regions with higher $R^2$ values generally aligned with townships with high yield variations ($r = 0.19$) and greater canola crop density ($r = 0.23$) (Figure 3.7). The average coefficient of variation (CV) of yield for example, expressed as the ratio of the standard deviation over the mean, across townships in Manitoba, Saskatchewan and Alberta was 26.9, 23.1 and 20.9% (Figure 3.7, top), respectively, while mean $R^2$ values were 0.22, 0.15 and 0.08, respectively, when soil moisture was included in the model (Figure 3.5, bottom).

![Figure 3.6: $R^2_{\text{diff}}$ based on the correlations between predicted and surveyed canola yields when SMOS soil moisture indices are included and excluded from the ICCYF. A positive $R^2_{\text{diff}}$ (blue) indicates model improvements with the inclusion of SMOS soil moisture observations, while a negative $R^2_{\text{diff}}$ (red) indicates a decline in the model’s performance.](image-url)
R\textsuperscript{2}_\text{diff} values across the canola growing regions of the Canadian Prairies however, revealed varying degrees of model improvement when soil moisture observations obtained by the SMOS satellite were included as potential predictors within the ICCYF. R\textsuperscript{2}_\text{diff} values ranged from -0.42 to +0.43 across the study area (Figure 3.6). The largest positive R\textsuperscript{2}_\text{diff} (+0.43) was observed in the Whitesand Plain ecodistrict in east-central Saskatchewan, while the largest negative R\textsuperscript{2}_\text{diff} (-0.42) was observed in the Blueberry Upland ecodistrict in the Peace River region of Alberta. This variability is likely due to the large variability present in the township-level canola yields over the 7-year period under study. The CV of yield ranged from 0% for townships with only one reporting year to 121.6% across the study area, with an average CV of 22.8% (Figure 3.7, top), indicating that canola production across the Canadian Prairies is highly dependent upon environmental conditions, similar to the findings of previous studies (e.g. Chipanshi et al., 2015). This high dependence presents a significant challenge for the development of accurate yield forecasts, especially when aggregating to the larger ecodistrict-scale; while ecodistricts are considered relatively homogenous (AAFC, 2013), yields can vary considerably across an ecodistrict due to small-scale variations in climate, soil, and landscape characteristics, as well as non-environmental controls such as seed varieties and land management practices (Canola Council of Canada, 2017). Developing one yield model that accurately captures the yield-environment response relationships of all townships within an ecodistrict is therefore, an incredibly challenging task.

Despite the variability observed in the ICCYF’s performance, several distinct spatial patterns can be observed. For example, model fit in east-central Saskatchewan was largely improved by the inclusion of SMOS soil moisture observations; R\textsuperscript{2}_\text{diff} values ranged from -0.15 to +0.43, with an average R\textsuperscript{2}_\text{diff} of +0.12 (Figure 3.6). This is expected as canola yields in eastern
Saskatchewan between 2010 and 2016 were found to be highly dependent on soil moisture conditions, as discussed previously, selecting SMOS soil moisture indices as a predictor of yield for all 7 years under study (Figure 3.3). Declines in model performance were observed in north-

Figure 3.7: A) Coefficient of variation (CV) of surveyed township-level canola yields across the Canadian Prairies from 2010 to 2016, and B) Spatial density of canola crop land area based on the temporal and spatial frequency of the crop between 2009 and 2016 (AAFC, 2017b).
central Saskatchewan directly to the west (Figure 3.6) where soil moisture played a much smaller role in the determination of canola yields over the 7-year period (Figure 3.3). High degrees of model improvement were also observed in some areas of Alberta where soil moisture is a key factor controlling crop yields, as discussed in Section 3.3.1 (Alberta Agriculture and Forestry, 2015). Most ecodistricts in the Peace River region of Alberta exhibited declines in model performance when SMOS soil moisture observations were included in the ICCYF, with 14 of the 19 ecodistricts (i.e. 73.7%) in the region having $R^2_{\text{diff}}$ values less than zero (Figure 3.6). Soil moisture conditions in the Peace River region are generally higher and less variable than in the arid grasslands of southern Alberta, therefore playing a smaller role in crop development (Alberta Agriculture and Forestry, 2015). Soil moisture in the central Peace River however, tends to be more limiting than the surrounding areas; average plant available water between 1988 and 2002 showed just 60 to 75 mm in 120 cm depth of soil, while surrounding regions were found to have 75 to 125 mm of plant available water on average (Alberta Agriculture and Forestry, 2015). This is reflected in the current study, in which the Fahler Plain, Dunvegan Plain and Debolt Plain ecodistricts in central Peace River showed correlation improvements of 0.08, 0.09 and 0.11, respectively, when SMOS soil moisture observations were included as potential predictors (Figure 3.6).

Overall, greater improvements to the ICCYF performance were observed across Manitoba and Saskatchewan when SMOS soil moisture observations were included as potential predictors (Figure 3.6). Of the 33 ecodistricts coinciding with the province of Manitoba, 66.7% had a positive $R^2_{\text{diff}}$, compared to 50.6% of ecodistricts in Saskatchewan and Alberta, while mean $R^2_{\text{diff}}$ were 0.02, 0.03 and 0.003 across Manitoba, Saskatchewan and Alberta, respectively. Greater model improvements in Manitoba and Saskatchewan with the inclusion of SMOS soil
moisture indices is likely due to the lower density of climate stations across these provinces compared to Alberta, indicating that even with its relatively coarse spatial resolution, the SMOS satellite provides greater spatial coverage over these areas than current climate networks can provide. However, despite the improvements observed in the ICCYF performance, model performance with both inputs was relatively low. Mean $R^2$ values were 0.13 and 0.12 when soil moisture was included and excluded from the model, respectively, while approximately 97.5 (i.e. with SMOS) and 99.5% (i.e. without SMOS) of ecodistricts under study exhibited $R^2$ values less than 0.50 (Figure 3.5). These results suggest that the ICCYF’s performance at the township-scale, regardless of the data inputs, may be significantly limited by the availability of data, as determined in previous studies (e.g. Zhang et al., 2015). Higher correlations are likely to be observed between the developed models and survey yields as the record of the model building data increases over time.

3.4 Conclusions

Remote sensing data sources are increasingly being used in crop yield forecasting systems due to the greater spatial and temporal coverage and lower costs they provide over traditional, field survey methods (Barrett and Petropoulos, 2014; Basso et al., 2013; Newlands et al., 2014). The wide-spread use of optical remote sensing VIs (e.g. NDVI) in yield forecasting models however, presents several challenges due to the reliance on visible and near-infrared portions of the electromagnetic spectrum, which are largely limited to cloud-free, clear-sky conditions (Peng et al., 2014). Therefore, this study aimed to assess the potential of passive microwave-derived soil moisture estimates for forecasting canola yields across the Canadian Prairies within AAFC’s ICCYF model.

The majority of ecodistricts under study (53.3%) showed improved model fit ($R^2_{\text{diff}} > 0$)
when soil moisture observations obtained by the SMOS satellite were included as potential predictors within the ICCYF. Selected predictors revealed that canola yields across the Canadian Prairies were highly sensitive to soil moisture conditions throughout the growing season during this period of greater-than-average precipitation. Soil moisture during week 24 in particular, was found to play a significant role in the determination of canola yields over the 7-year period under study, being selected as the top predictor. Overall, greater improvements in model performance were observed in Manitoba and Saskatchewan, indicating passive microwave-derived soil moisture observations can provide valuable information in regions where climate stations are sparsely distributed. Furthermore, soil moisture was found to provide a better descriptor of canola stresses than the more commonly utilized NDVI, with SMOS soil moisture indices being selected as a predictor in 74.2% of developed ecodistrict models, while NDVI indices were selected in only 41.2%. These findings further highlight that NDVI may be a less effective indicator of canola yields, as determined in previous studies (e.g. Behrens et al., 2006; Chipanshi et al., 2015; Shen et al., 2010, 2009, Sulik and Long, 2016, 2015), supporting the need for other, more relevant remote sensing-derived indices, such as soil moisture, in yield forecasting models.

Despite the improvements in model performance observed across the majority of ecodistricts, a large amount of variability was observed spatially across the study area in the degree of model improvement when SMOS soil moisture was included as a potential predictor. Furthermore, $R^2$ values revealed that performance with both model inputs was relatively low, suggesting that the ICCYF’s performance at the township-scale, and when using passive microwave-derived soil moisture observations, is limited by the availability of these datasets. Future work should therefore focus on assessing the potential of utilizing SMOS soil moisture observations for forecasting yields at the township-scale as the data record of these two datasets
increases. Incorporating soil moisture as a key variable in crop yield forecasting models will not only provide a more accurate descriptor of crop stresses but allow for improved yield forecasting which will have significant implications for the development of export-import policies, food security policies, and efficient land management practices.
Chapter 4.0: Conclusions

Crop yields are inherently variable, often varying dramatically from year-to-year (Ozaki et al., 2008). Accurately forecasting these variable yields is incredibly complex, however it has become increasingly important as climate variability intensifies under global climate change (Kogan et al., 2013; Ozkan and Akcaoz, 2002). While soil moisture is a key variable in the determination of crop yields (Champagne et al., 2012; Dobriyal et al., 2012; Holzman et al., 2014; Maybank et al., 1995; McGinn and Shepherd, 2003; Nadler, 2007), the inclusion of direct soil moisture measurements within crop yield forecasting systems has been largely limited by the nature of soil moisture data; traditional ground-based measurements of soil moisture are sparsely-distributed and therefore, are highly unrepresentative of the soil moisture variability present across the landscape (Champagne et al., 2012, 2011; Holzman et al., 2014; Peng et al., 2014). This research aimed to improve both the accuracy and robustness of crop yield forecasting systems by evaluating the utility of passive microwave soil moisture observations obtained by the SMOS satellite for forecasting canola yields across the Canadian Prairies.

This thesis addressed two key research objectives in order to evaluate the potential of SMOS soil moisture observations for forecasting canola yields. The first was to determine the critical time periods and threshold conditions in which canola exhibits the greatest sensitivity to soil moisture stress, as captured by the SMOS satellite, and compare the significance of the observed associations to other more traditionally used climate indicators (e.g. precipitation, temperature). Of the four indicators assessed (rainfall, maximum and minimum temperature, and soil moisture), the strongest associations were observed between SMOS soil moisture observations and canola yields. Low-yielding years in particular, exhibited significant associations ($p < 0.01$, degrees of freedom = 1) with excess soil moisture conditions throughout
the growing season, while the strongest associations were observed in the beginning of June, coinciding with the stand establishment stage. Soil moisture in excess of 26.6% at this critical time period was found to have a profound, negative effect on canola production across Saskatchewan between 2010 and 2015. Overall, the results of this study demonstrated that soil moisture observations obtained by passive microwave satellites, such as SMOS, may be effectively utilized as an indicator of crop yields, particularly in areas experiencing excess soil moisture conditions.

The second objective of this thesis was to assess the added-value of utilizing remotely-sensed soil moisture observations from the SMOS satellite for forecasting canola yields across the Canadian Prairies. SMOS weekly percent volumetric soil moisture estimates were included as potential predictors within Agriculture and Agri-Food Canada’s (AAFC) Integrated Canadian Crop Yield Forecaster (ICCYF) to generate regional yield forecasts for canola at the township-scale across the Canadian Prairies. The contribution of SMOS soil moisture data to the model’s forecast skill was determined by calculating the difference between model $R^2$ values ($R^2_{\text{diff}}$) when SMOS soil moisture predictors were included and excluded from the model, respectively. The majority of ecodistricts under study (53.3%) showed improved model fit (i.e. $R^2_{\text{diff}} > 0$) when soil moisture observations obtained by the SMOS satellite were included in the ICCYF model, while selected model predictors further revealed that canola yields across the Canadian Prairies are highly sensitive to soil moisture conditions throughout the growing season and in particular, during week 24. However, despite these improvements, a large amount of variability was observed spatially across the study area in the degree of model improvement when SMOS soil moisture was included as a potential predictor with $R^2_{\text{diff}}$ values ranging from -0.42 to +0.43. Furthermore, $R^2$ values revealed that performance with both model inputs was relatively low,
with mean $R^2$ values of 0.13 and 0.12 when soil moisture was included and excluded from the model, respectively. These findings suggest that while SMOS soil moisture observations may provide a more effective indicator of canola yields, the use of passive microwave-derived soil moisture observations in yield forecasting models is limited by the short temporal record of these datasets.

It is anticipated that this research will facilitate the development of more reliable crop yield forecasting methodologies, incorporating remotely-sensed soil moisture as a key variable, therefore, allowing for more informed policy decisions and land management planning. Recent advancements in passive microwave remote sensing platforms offer an efficient and accurate means for acquiring large-scale surface (depth < 5 cm) soil moisture data at a greater spatial density than ground-based soil moisture measurements and at frequent enough temporal resolutions (1-3 days) for near-real time monitoring (Barrett and Petropoulos, 2014; Champagne et al., 2012, 2011; Kerr et al., 2012). However, the short temporal record of most operating satellites (< 10 years), including SMOS, has largely limited the use of these datasets as a 30-year data record is typically required in order to establish baseline soil moisture conditions and define extreme events (Sheffield et al., 2004).

The results of this research suggest that the 7-year SMOS data record provided sufficient data to accurately capture periods of extreme soil moisture stress at both the CAR- and township-scale. However, the short temporal record did prove to be problematic in the development of accurate linear regression yield models. Future research would therefore benefit from a longer soil moisture time series in order to improve forecasting skill. In addition, the 7-year period between 2010 and 2016 was abnormally wet. Previous studies have suggested that microwave surface soil moisture observations are unrepresentative of root zone soil moisture deficits.
(Capehart and Carlson, 1997; Wilson et al., 2003); therefore, a longer data record is required to assess whether these datasets can capture soil moisture-related yield variations under dry conditions in order to truly determine the potential of passive microwave soil moisture observations for crop yield forecasting. Additionally, the coarse spatial resolution of these datasets is still a significant limitation as soil moisture varies considerably over small areas due to varying soil types and textures, vegetation, topography, meteorological conditions and land management practices (Barrett and Petropoulos, 2014; Dobriyal et al., 2012; Engman and Chauhan, 1995; Holzman et al., 2014). Therefore, crop yield forecasting, especially at the smaller township-scale, would benefit from future missions with higher spatial resolution soil moisture products.

Absolute percent saturated surface soil moisture products (AAFC, 2016) derived from the SMOS satellite were utilized in this research in order to quantify the threshold soil moisture conditions indicative of canola yield variations. However, several studies have found that SMOS exhibits an overall dry bias, and therefore, tends to underestimate soil moisture, due to errors inherent in the SMOS retrieval algorithm (Al Bitar et al., 2012; Champagne et al., 2016; Djamai et al., 2015; Jackson et al., 2012). Future research therefore, may want to investigate the use of percent saturated surface soil moisture anomaly products (AAFC, 2018b) for forecasting crop yields as this relies less on the absolute accuracy of the soil moisture estimates and more on the relative soil moisture trends (Champagne et al., 2016, 2015).

Overall, the thesis findings reveal that soil moisture observations obtained by the SMOS satellite provide an effective indicator of canola yields, particularly in areas experiencing excess soil moisture conditions, and that forecast accuracy is improved when SMOS soil moisture is included as a potential predictor. While this research investigated the potential of passive
microwave-derived soil moisture estimates for forecasting canola yields, the potential of these datasets to forecast yields of additional crops should be explored in future studies. Improved understanding of the soil moisture-yield relationship for various crops is crucial for informing more robust yield forecasting methodologies across the Canadian Prairies in order to better equip farmers and society to respond to climate variability.
References


68


70


Statistics Canada, 2018b. Table 32-10-0153-01 - Total area of farms and use of farm land, historical data, CANSIM (database). https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3210015301&pickMembers%5B0%5D=1.1&pickMembers%5B1%5D=3.2 (accessed 7.13.18).


