On Predictive Models of Forage Crops Productivity by Using Weather Variables: An Application in the Province of Ontario, Canada

by

Shuai Du

A thesis submitted to
The Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements
of the degree of

Master of Science

I.H. Asper School of Business
The University of Manitoba
Winnipeg, Manitoba, Canada
October 2017

Thesis co-advisors Author

Xuemiao Hao, Lysa Porth **Shuai Du**

On Predictive Models of Forage Crops Productivity by Using Weather Variables: An Application in the Province of Ontario, Canada

Abstract

The objective of this thesis is to establish predictive models for forage yield (productivity, tons per acre) using most relevant weather variables, such as precipitation and temperature from April to June. The outcome of this study is expected to be utilized to design indices and/or to set triggers for CAT bonds on forage crops in Ontario for the Government of Canada as it is exposed to tremendous agriculture risk exposures. We use forage crops data in Ontario, Canada, as an example. We propose to apply a single predictive model on a whole region, which is a vast area consisting of eight to ten counties which have a similar geographical environment. Seven models are tested for five regions with variables such as monthly rainfall, three months cumulative rainfall, average temperature, CDD (cooling degree days), etc. A new approach called weighted average temperature adjustment (WATA) is employed to deal with temperature data.

The results demonstrate that the selected predictive model(s) consistently and considerably better explain the relationship between forage yield and weather variables for regions.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to first express my gratitude to my two supervisors, Dr. Xuemiao Hao and Dr. Lysa Porth for countless time and energy they have devoted to me and my thesis. I firmly believe that this work could not have been done without encouragement and patient guide from them. Dr. Hao is an outstanding role model of my life; especially he has shown his rigorous attitude in scientific studying to me. Dr. Lysa Porth has the tremendous and extensive knowledge, and she always actively supports me and provides some precious suggestions when I need. Both of them made me have a deeper understanding of the actuarial industry and additionally, cultivated me to grow up.

I also want to say thank you to Dr. Milton Boyd and Dr. Jeffrey Pai as committee members and their invaluable feedback and comments that indeed helped my lot to improve and complete my thesis work.

I would also be very grateful for all my co–workers, Chunli Liang and Linghua Wei, with their powerful encouragement and I honestly have obtained beneficial experience from them.

Finally, I would like to say a big thank you to my parents and my wife, for all of the understanding, support and love over the last two years. Because of the long distance, I rarely spent the time in company with them, but I still can feel their unlimited love to me from kind words especially when I felt disappointed and disheartened. They are always my endless energy resource and could provide inspiration for me. I love them so much and be very thankful for what they have done to me.

Chapter 1

Introduction

Agriculture is major of social and economic significance (Porth and Tan, 2015). According to a report from Food and Agricultural Organization (FAO) of the United Nations in 2009, global food production needs to increase 70% by 2050 in order to feed continuing growing population. Therefore, promoting agricultural production is important and essential. However, agricultural risk exposure is tremendous. In particular, weather risk is dominant for the agriculture industry. Weather risk can be interpreted as agricultural loss due to adverse weather conditions, for example, drought, flood, hail, earthquake, hurricane, etc. As the previous literature and research point out (Zhu et al., 2015; USDA, 2014) adverse weather may contribute at least 70% of agricultural loss. So hedging adverse weather is the key procedure for agricultural producers to stabilize their income and willingness to continue being engaged in agriculture reproduction. However, hedging weather risk is challenging because weather risk is systematic and un-diversifiable, which means it can not be reduced or avoided by human management. Additionally, weather risk can be widespread and spatially correlated, which can have an effect on many farms for a given region (Porth et al., 2014).

1.1 Agricultural insurance in Canada

Agriculture insurance is a transferring risks technique, and it helps producers transfer risks to insurance companies. Barret et al. (2007) and Wang (2015) state that when farmers are faced with risks without adequate external risk protection or transferring mechanism, they may use fewer inputs, including fertilizer, water, etc., resulting in lower production in the end. Nowadays, agricultural insurance plays an important role in the agriculture sector. For instance, 90 % of the annual crop are insured in Ontario (Porth and Tan, 2015).

There are two major prevailing types insurance that is traditional indemnity-based insurance and index-based insurance.

Traditional indemnity-based insurance has three drawbacks, including moral hazard, adverse selection, and high administration cost. Moral hazard means farmers, as insured, may artificially or intentionally create damage on their crops resulting in reducing production in order to receive indemnities from insurance companies. For instance, they may carelessly regulate their land or irrigate land not on time and regularly. Adverse selection appears because policyholders are a group with the different level of risks, but insurance companies, as insurers, do not have the ability to distinguish these different risk levels so that insurer only could offer contracts to policyholders based on average risk level. Thus, premiums paid by farmers who are facing small risks may be subsidized to farmers who are exposed to high risks, but the premium paid is same for both sides. It may result in policyholders in low risks surrender from contracts. Under traditional insurance contract, on-field inspection is required, so the administration cost is high.

Index-based insurance as an alternative option can overcome above three disadvantages. Because whether indemnity being paid is solely depending on the generated value from underlying index so that this mechanism could ensure the index (such as precipitation, temperature, sunshine, area yield, satellite, etc.) is totally independent of policyholders' performance and meanwhile, insured difficultly manipulate the index to achieve more benefits. Obviously, on-filed inspection is no longer needed, which can reduce almost surely all administration cost, since actual losses on land do not affect indemnity. However, index-based insurance still has a shortcoming, which is basis risk. Designing the index and setting triggers of the insurance is also important.

1.1.1 Forage and forage insurance

In Canada, forage is planted 44% of the farming land as the second largest crop and specificity, the area of the plantation of forage is about 1,999,000 acres in the province of Ontario only (Wang, 2015). 80% of cows for beef production and 60% of dairy cows are fed forage (Porth and Tan, 2015). Thus, forage products are important for Canadian agriculture. Usually, forage harvest is from June to August with three cuts per year. As Simpson (2016) illustrates that there is a multi peril crop insurance program called as AgriInsurance in Canada, operated by each of ten provincial government crop insurance organizations and cooperated with the federal government as well. Farmers only need to pay 40% of the total premium and rest 60% is subsidized by the provincial and federal government.

1.1.2 Two examples of Ontario

In Ontario, there was a forage insurance named as simulated forage yield (SIMFOY) implemented in 1981, but it was terminated in 2004 (Wang, 2015). Since this insurance modelled forage yield with a complex combination of variables, such as rainfall, temperature, soil type, etc., farmers found that the structure of trading contract was too complicated to understand, especially for indemnities calculation process. In the end, it was abolished. Oppositely, there is a trading weather index insurance in Ontario called as Ontario forage rainfall plan (OFRP). Indemnities under OFRP only rely on cumulative precipitation, one variable. However, temperature and sunshine may also have a significant impact on forage yield. OFRP may not derive a high correlation between yield and rainfall.

1.2 Catastrophic risk management

In Canada, catastrophic risks faced by agricultural insurers are hedged by reinsurance. Nowadays, each province independently operates its crop insurance program, and the federal government offers a reinsurance program for provinces; see, for example, Ye et al. (2013). To date, five provinces, Alberta, Saskatchewan, Manitoba, Nova Scotia and New Brunswick, have joined into this reinsurance program. Some other provinces, including British Columbia, Ontario and Prince Edward Is-land, purchase reinsurance from private companies.

From the perspective of federal government, managing agricultural reinsurance is difficult and sometimes is very costly due to unpredictable extreme losses caused by catastrophic events, such as drought, flood, hail, earthquake, hurricane, etc. For instance, during the fiscal year 2002–2003, massive appropriations were allocated out to Alberta and Saskatchewan from the Federal Reinsurance Fund due to severe droughts (Ye et al., 2013). Additionally, as a report by the federal government demonstrated in 2011, there was a deficit in the balance of the Federal Reinsurance Fund. Thus it required the Minister of Finance to advance $497.5 million into the fund for each Reinsurance Fund agreement (Government of Canada, 2011; Ye et al., 2013). Thus, the federal government needs more techniques to manage agricultural risk exposures. Otherwise, it may experience substantial fiscal expenditures when severe catastrophes happen.

Catastrophe bond (CAT bond) may be a good alternative choice. CAT bond is a kind of event-linked bond with payoff depending on the occurrence of a specific event, such as earthquakes, drought, floods and hurricanes (Sun et al., 2015). Under a CAT bond contract, the interest or the principal of the bond may be terminated or delayed to issue to its investors if the CAT bond is triggered. For insurance companies or governments, CAT bond can help them transfer tremendous catastrophic risks to investors in the financial market. Therefore, if the federal government issues CAT bonds, it can use capitals from the financial market to cover potential extreme losses rather than increasing the budget of the Reinsurance Fund. The reduction of expenditures by issuing CAT bonds could provide the federal government opportunities to concentrate on improving its current agricultural risk management programs and to allocate spending on innovation and product development in relevant fields (Ye et al., 2013). So CAT bond may be a vital complement to the federal government's tool kit for catastrophic risk management.

1.3 Research objective

The objective of this thesis is to investigate the relationship between forage crops yield and weather variables (in this thesis, variables are precipitation and temperature). We search and select the best predictive model(s) of forage crops yield using weather variables for the province of Ontario. The outcome of this study is expected to be utilized to design indices and to set triggers for CAT bonds on forage crops in Ontario.

Chapter 2

Data

This chapter illustrates information about the data used in the analysis. The content of data consists of some basic weather information in county level, production of forage crops in farm level, areas of farming land, weather station locations, etc., in the province of Ontario, Canada. In addition, all data adopted in this thesis is provided by Agricorp, a provincial government crop insurance company in Ontario.

2.1 Data description

There are four datasets used in our analysis named Dataset 1, 2, 3 and 4. Dataset 1 mainly contains several types of weather information, such as precipitation in millimetre, maximum and minimum temperature of a day in degrees Fahrenheit, etc. The following list provides details for all fields in Dataset 1.

1. *id*: a unique seven-digit id number for each weather station.

2. *station name*: weather station's name that is based on the location of the station.

3. *county num*: means county number and same relation as between id and station name, each county has a specific number to be represented.

4. *county name*: county name can provide where this information is collected. In the other hand, it can show where these weather stations are placed.

5. *year*: the year, some weather data are reported.

6. *month*: the month, some weather information is recorded.

7. *day*: the day, some weather data are collected.

8. *rainfall*: it provides rainfall for a specific day measured in millimetres.

9. *sunshine*: it means average hours of sunshine received per day.

10. *maxtemp*: it shows the maximum temperature during the day measured in degrees Fahrenheit.

11. *mintemp*: same logic as Maxtemp, it means minimum temperature during the day measured in degrees Fahrenheit.

12. *three UTM labels*: UTM is short for Universal Transverse Mercator coordinate system, and it can provide locations on the surface of the earth.

13. *dist*: this label is abbreviated to distance, and it is expressed in kilometres and gives the shortest distance form the weather station to the closet Great Lake.

A sample of Dataset 1 is given in Table 2.1. Provided by weather stations from all counties of whole Ontario, these data are sorted by day, month and year. Months are from April to October and years included from 1967–2004. Although Dataset 1 contains weather information in the long period 1967–2004, the information is not statistical complete yet. For instance, there are some missing observations for rainfall and temperature. Particularly, missing observations for rainfall are significantly more than temperatures.

Table 2.1: Example of Dataset 1

id	station name	county num	county name	year	month	day	rainfall	sunshine	maxtemp	mintemp	UTM ZO NO	UTM North M	UTM East M	dist
6105760	OMPAH	10	FRONTENAC	1998	4	1	11.2		41.0	33.8	18	4937188	380119	36.9
6160473	BANCROFT/ OTTAWA	12	HASTINGS	1967	4	2	5.8	1.0	41.0	27.0	18	4993000	275000	114.3
6117684	SHANTY BAY	43	SIMCOE	2004	5	14	8.4		77.9	61.7	17	4918000	613000	36.2
6111859	COOKSTOWN	43	SIMCOE	2004	5	14	8.0		77.9	59.9	17	4894000	604000	43.2
6140954	BRANTFORD MOE	29	BRANT	1997	5	30		1.0	67.0	51.0	17	4777000	562000	40.3
6140954	BRANTFORD MOE	29	BRANT	1998	6	1	3.0				17	4777000	562000	40.3
6139149	DUMMY VINELAND	26	NIAGARA	2002	4	27	13.6	7.0	50.0	32.0	17	4781000	632000	1.7
6139149	DUMMY VINELAND	26	NIAGARA	2002	4	28	11.4	0.0	44.6	38.3	17	4781000	632000	1.7
6139149	DUMMY VINELAND	26	NIAGARA	2002	4	29	0.4	0.2	49.1	34.7	17	4781000	632000	1.7
6113490	HONEY HARBOUR	44	MUSKOKA	1997	4	25		7.3	59.0	38.0	17	4969000	593500	0.0
612GLEN	GLENCOE	39	MIDDLESEX	1997	5	29	9.0	0.0	56.0	50.0	17	4733000	442000	25.6
612GLEN	GLENCOE	39	MIDDLESEX	1997	5	30		1.0	64.0	53.0	17	4733000	442000	25.6
6105460	MORRISBURG	1	GLENGARRY	1999	7	17	3.0		89.6	62.6	18	4972000	485000	117.7
6160473	BANCROFT/ OTTAWA	12	HASTINGS	1982	5	8	21.8	0.0	54.0	48.0	18	4993000	275000	114.3
612CAST	CASTLETON	14	NORTHUM- BERLAND	1997	7	9	20.0	8.0	73.0	55.0	18	4886000	265000	12.7
6139145	VINELAND STATION/ TORONTO A	26	NIAGARA	1973	6	17	4.3	2.8	54.0	52.0	17	4781000	632000	1.7

Dataset 2 contains forage production information at farm level from the year 1981–2004. The following list provides details for all fields in Dataset 2.

1. *id*: id number for each farm consists of eight or nine numbers and each every farm has its id, which is to say, an id and a farm are one by one matched.
2. *county name*: same expression as mentioned in Dataset 1.
3. *township name*: it provides township names of farms locating.
4. *county num*: same expression as mentioned in Dataset 1 as well.
5. *yield*: it records forage crops production measured in tonnes per acre.
6. *acres*: it means total plantation areas of a farm expressed in acres.

A sample of Dataset 2 is given in Table 2.2. We want to point out that, majority data are ended by the year 2003, and there are many missing observations in three UTM categories data. Dataset 2 also includes species of forage crops that are alfalfa and timothy. Because of the severity of lacking information in Dataset 2, this part is not shown in Table 2.2.

Table 2.2: Example of Dataset 2

id	county name	township name	county num	year	yield	acres	UTM ZO NO	UTM North M	UTM East M
101113147	LAMBTON	WARWICK, TWP	38	2000	2.02	20.0	17	4769752	423303
101113147	LAMBTON	WARWICK, TWP	38	2001	1.69	20.0	17	4769752	423303
105038941	LAMBTON	ST CLAIR, TWP	38	2001	3.01	10.0	17	4738327	397767
101113147	LAMBTON	WARWICK, TWP	38	2003	2.02	10.0	17	4769752	423303
96312091	FRONTENAC	FRONTENAC ISLANDS, TWP	10	1981	1.83	10.8			
97393699	FRONTENAC	SOUTH FRONTENAC, TWP	10	1981	2.17	13.0			
104029690	FRONTENAC	FRONTENAC ISLANDS, TWP	10	1984	2.03	17.6			
100213708	FRONTENAC	SOUTH FRONTENAC, TWP	10	1981	2.26	17.5			
99129385	FRONTENAC	KINGSTON, CITY	10	1981	2.18	14.2			
97393699	FRONTENAC	SOUTH FRONTENAC, TWP	10	1982	2.03	8.3			
107797108	CHATHAM-KENT	CHATHAM-KENT,MUN	36	1989	3.03	5.0			
103385371	CHATHAM-KENT	CHATHAM-KENT,MUN	36	1994	3.36	5.0			

Since Datasets 3 and 4 are similar, we discuss and interpret them together. Datasets 3 and 4 provide location information of weather stations and farms respectively, via latitude, longitude and elevation coordinates as shown in Table 2.3.

The following list provides details for all fields in Datasets 3 and 4. We note that not all farms' locations are demonstrated in the Dataset 3, while there are only a few farms' locations are inclusive.

1. *id*: it has the same meaning in Dataset 1 for weather stations or farms.
2. *north east and zone*: these three labels are UTM coordinates.
3. *long and lat*: they are longitude and latitude coordinates. By using these coordinates, distances between weather stations and farms can be calculated.
4. *elevation*: it describes elevation of the weather station or the farm.
5. *fist year of data*: it expresses the first year that the weather station or the farm started to collect and report data.
6. *last year of data*: it means the last year that the weather station or the farm collected and reported data.
7. *ag div*: it shows which region the weather station belongs to.

Table 2.3: Example of Datasets 3 and 4 combined

id	north	east	zone	long	lat	elev	dist	fist year of data	last year of data	ag div
6073810	4966000	475000	18	-75.32	44.85	70.10	106.80	1997	2004	Eastern Ontario
6157831	4898000	379000	18	-76.51	44.23	91.21	0.85	1967	2004	Eastern Ontario
6110218	4882718	618482	17	-79.52	44.09	233.49	45.53	1998	2004	Central Ontario
6164432	4914000	680000	17	-78.74	44.36	257.63	52.07	1967	2004	Central Ontario
6169647	4918000	660000	17	-78.99	44.40	284.91	60.22	1997	2004	Central Ontario
6012198	5580671	490575	15	-93.13	50.38	359.30	352.36	1999	1999	Northern Ontario
6048261	5359429	328415	16	-89.32	48.36	200.74	4.35	1967	2004	Northern Ontario
6126499	4748000	406000	17	-82.15	42.88	200.99	18.48	1997	2004	Southern Ontario
6131415	4696000	401000	17	-82.20	42.41	179.86	16.80	1967	2004	Southern Ontario
6119500	4954000	489000	17	-81.14	44.74	187.87	0.43	1967	2004	Western Ontario
6122847	4800634	472449	17	-81.34	43.36	315.14	29.89	1998	2004	Western Ontario

2.2 Preparation

We perform the following preliminary operations on these datasets to facilitate our future analysis.

1. As mentioned in Chapter 1, forage crops often be harvest starting from June to August. Therefore, April, May and June's three months rainfall data are mainly considered in this analysis as the most contribution for forage growing. In other words, this season (i.e.,.second season, from April to June)can be called forage growing season.

2. We notice that in Dataset 2 forage production information is available only from County 9. Therefore Counties, 1 to 8, are excluded in this analysis.

3. Data without county information or weather station ID information are exclusive either.

4. There is an assumption that all missing observations in rainfall are treated as zeros precipitation in millimetre.

5. There are twenty–five weather stations that contain observations from 1967 to 2003, named as long term weather stations, and remaining three hundred and twelve weather stations collect short period observations from 1997 to 2003, called as short term weather stations.

6. The province is divided into five regions as following: Eastern, Western, Southern, Northern, and Central Regions shown in Figure 3.2. Each region is

consisting of seven to ten counties and the approach employed, uses OMAFRA (2017) for reference.

7. calculating distances from each farm, which has longitude and latitude, to its nearest long term weather station. The reason for this step is necessary and crucial, because weather data provided by long term weather stations may be represented for those counties that do not collect weather data from the year 1981 to 1996, which means some counties only contains short term weather stations.

8. Monthly rainfall, monthly temperature and CDD (cooling degree days, in this analysis, 50°F is adopted as threshold) for April, May and June are calculated out. Meanwhile, cumulative precipitation is obtained as well.

2.3 Regions, stations and farms

2.3.1 Regions

We follow the way of Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA, 2017) to divide the whole province of Ontario into five regions. There is a vital reason to do the partition. Since with the aim to achieve the objective of this thesis, a macro studying perspective is decent and using region level as analysis unit can provide such a macro perspective.

Figure 2.1 shows a map of UTM zones of Canada, and it is clearly demonstrating that province of Ontario locates in the middle-east of Canada and crosses through zone 15 to zone 18.

Figure 2.1: UTM

Note. Figure 2.1 UTM, retrieved June 28, 2017, from :http://www.cccmaps.com/gps.html. Copyright by Canadian Cartographics Corporation.

Figure 2.2 describes elevation conditions of the whole province of Ontario and are measured in meters. There is a legend consisted of several levels of colours, attached on the map and when the colour of a place on the map is closer to dark green, the lower elevation it has, vice versa.

Figure 2.3 shows the analysis units, five regions, in five different colours based on data provided by OMAFRA (2017). Blue is eastern region and yellow means central region. Red and light green are western and southern region respectively. The resting place is big green northern region. Based on Figures 2.2 and 2.3, we have the reason to believe that the method of dividing regions depends on elevation environment. From observation in Figures 2.2 and 2.3, we find that elevations within each region are similar. For instance, in the eastern region, elevation numbers are comparatively low, but in the central region, elevation numbers are significantly higher.

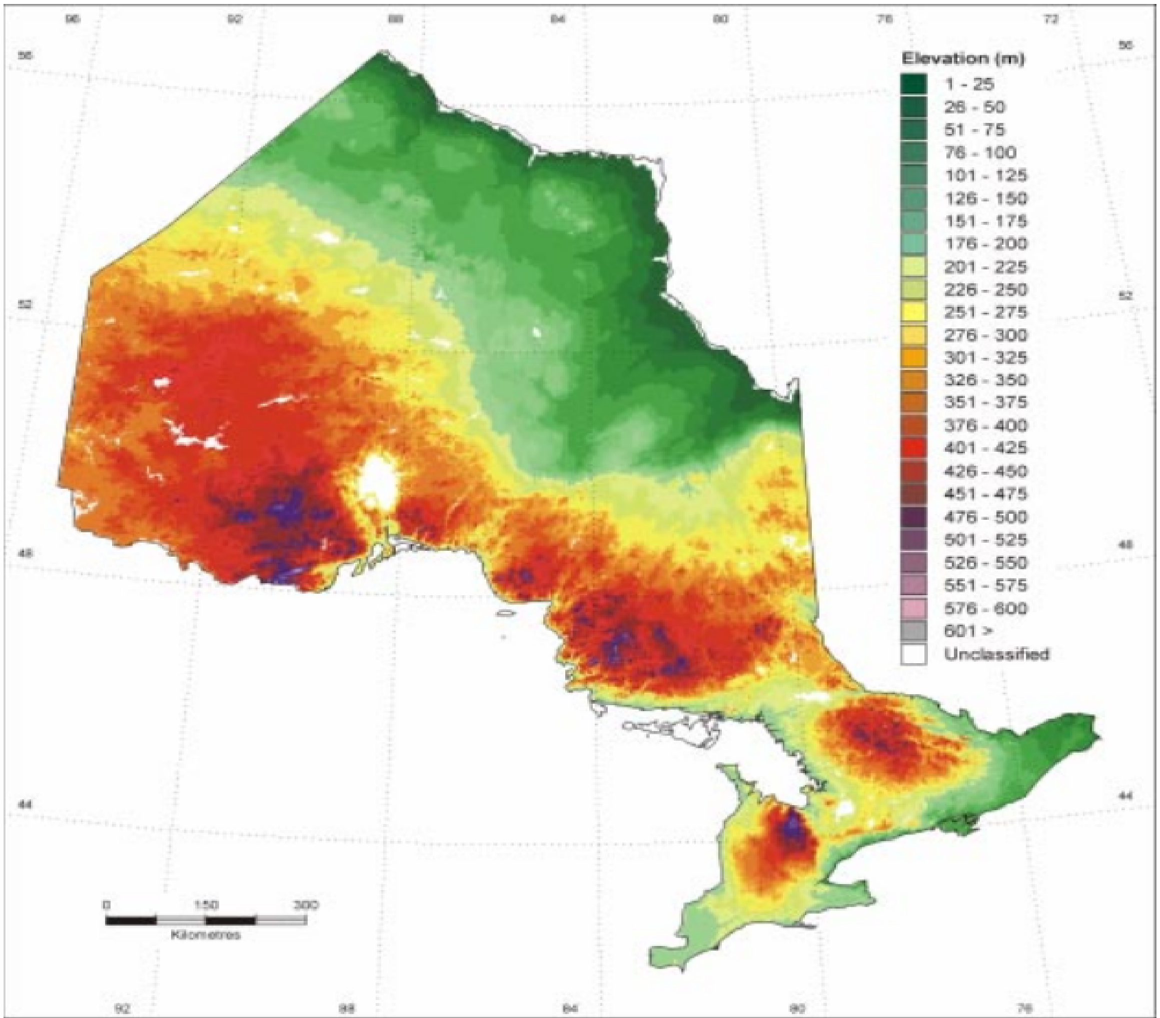

Figure 2.2: Elevation of Ontario

Note. Figure 2.2 Elevation of Ontario, retrieved June 28, 2017, from :https://oneclass.com/note/1513054-geography-2011ab-lecture-1. Copyright by Oneclass

2.3.2 Weather stations and farms

In order to clearly elucidate the information about weather stations and farms, a few figures are showing as following.

Figure 2.4 shows the map of locations of all farms, which have coordinates in terrain. It is quickly found that farms are mainly located in southern, western,

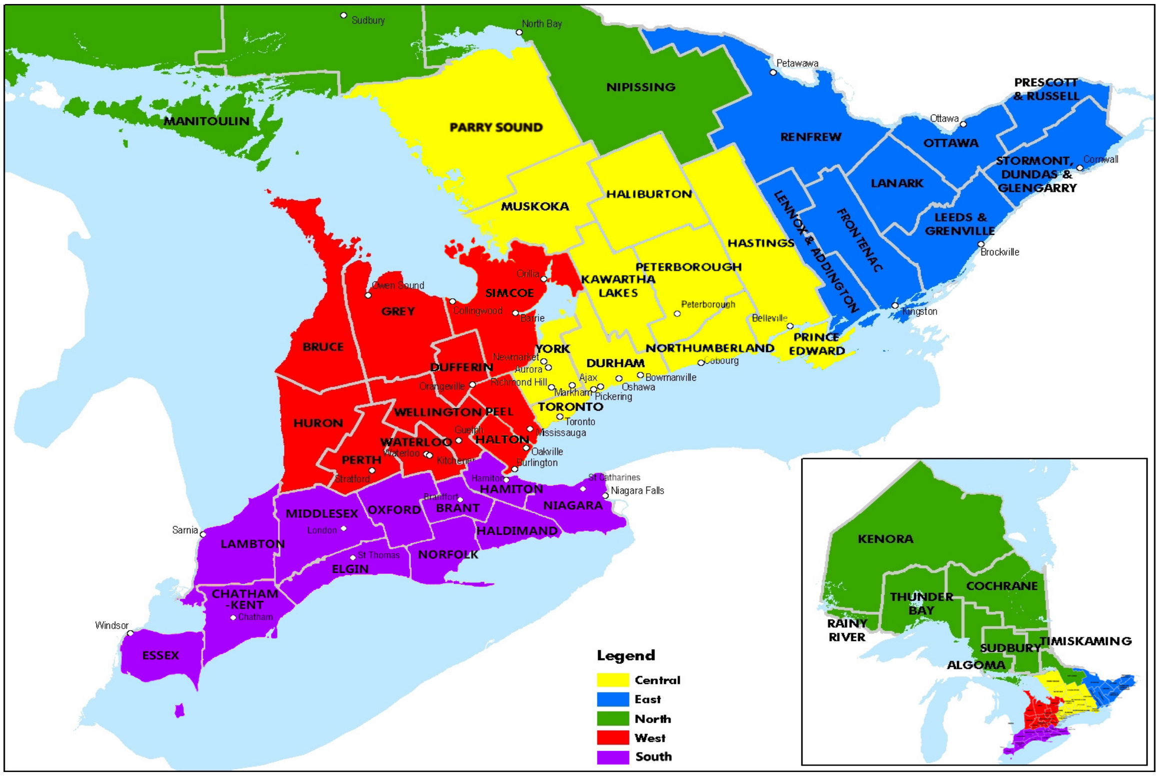

Figure 2.3: Five regions

Note. Figure 2.3 is generated by Photoshop

eastern and partly central regions. Coincidently, the majority of these farms are also locating in low elevation places, compared farms' locations and Figure 2.2.

Also, Figure 2.5 shows all locations of weather stations, and then these stations split into two sections: long term and short term. Figure 2.6 is the map of long term weather stations, and as mentioned before, long term weather stations are very few compared with a total number of stations. Figure 2.7 is the map for short ones. As shown in this figure, short term weather stations are comparatively much more than long term ones.

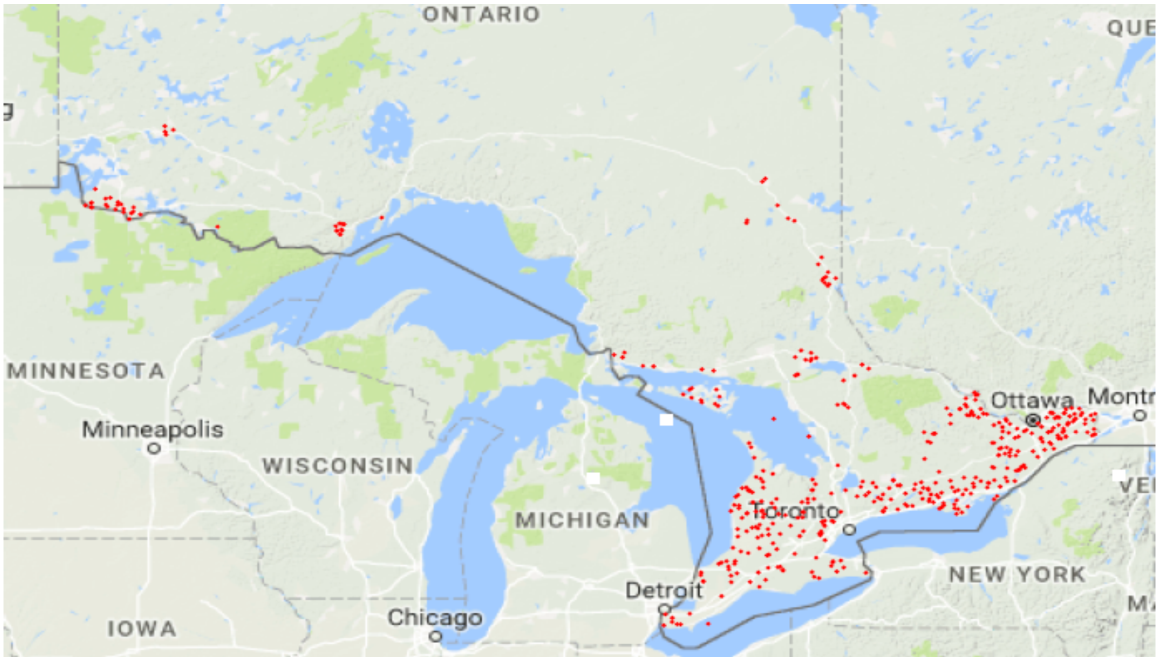

Figure 2.4: Farm locations

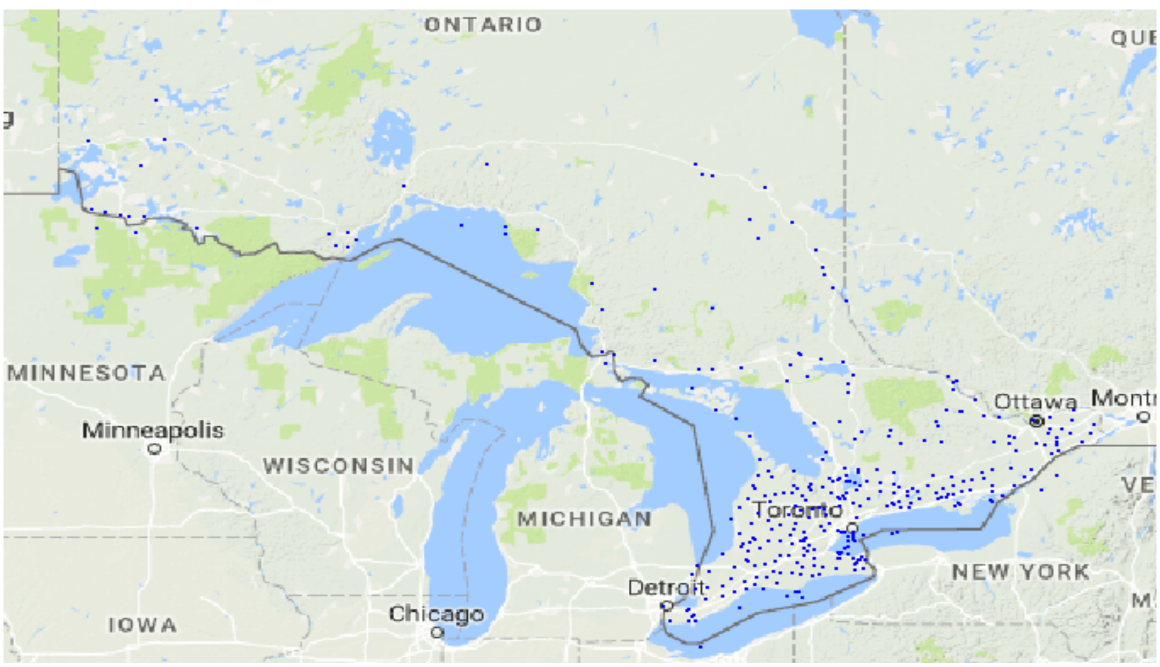

Figure 2.5: All weather stations

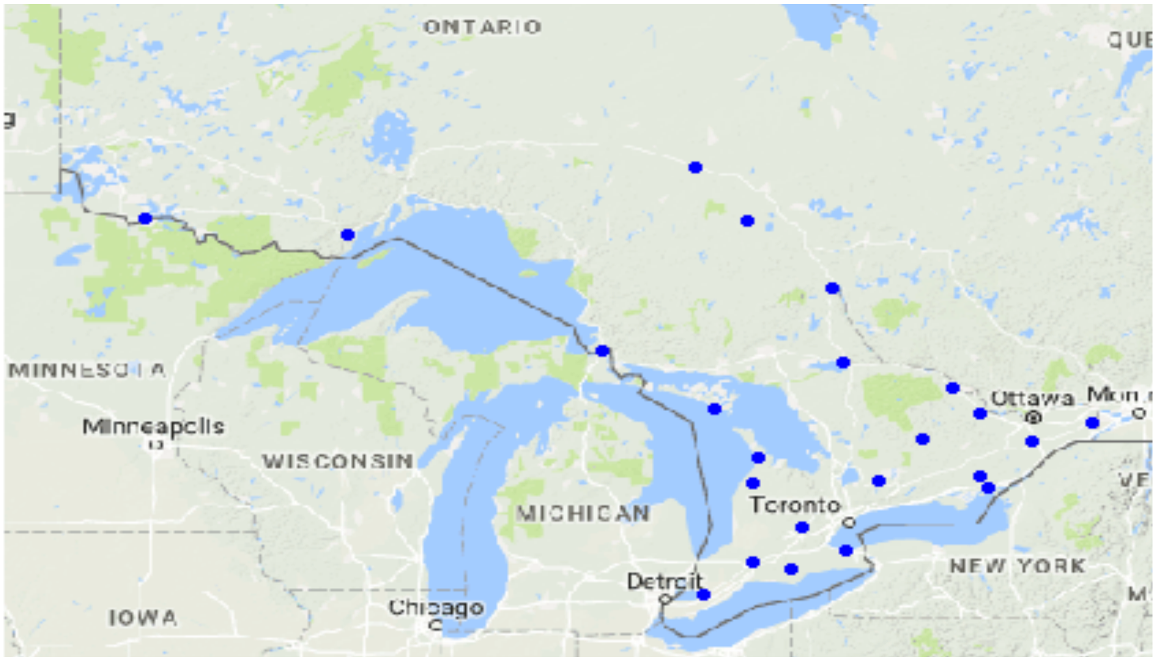

Figure 2.6: Long term weather stations

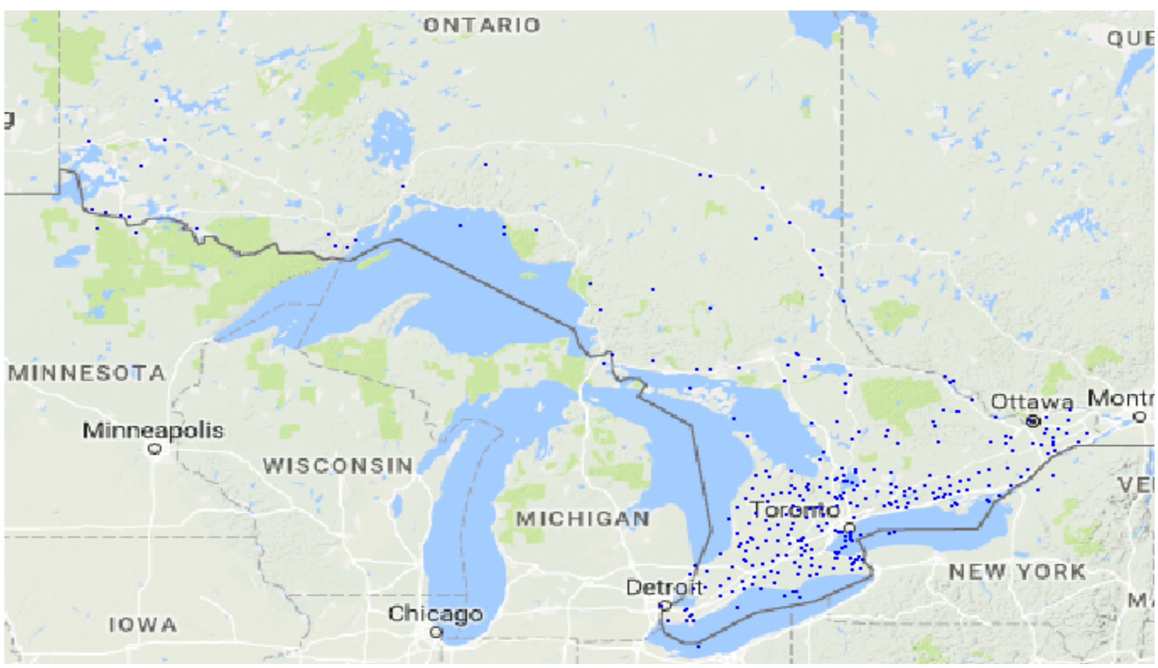

Figure 2.7: Short term weather stations

Chapter 3

Analysis

As Vedenov and Barnett (2004) address that temperature and precipitation are possibly contributing most to yield variability among all important factors. Therefore, in order to quantitatively analyse the potential relation between forage crops yield and climate variables, seven models are established and validated. In fact, the potential relation is not apparent and not effortlessly to derive. Each model is applied to be analysed from April 1st to June 30th in the years 1981 to 2003 and also in three levels of study unit: counties in Ontario, regions in Ontario and the whole province of Ontario. Besides, a new dealing temperature data approach called as weighted average temperature adjustment is introduced into this analysis (details are discussed in Chapter 4). Remarkably, this analysis shows insight that forage productivity is significantly impacted by weather variable as expected.

This chapter is constituted of four sections. Section 3.1 shows all seven models and corresponding explanation. Section 3.2 mainly discusses estimation of missing temperature observations from some counties without weather stations that collect weather data from the year 1981 to 1996. In Section 3.3 we demonstrate detail steps for dealing with levels of county, region, and province, respectively. Section 3.4 introduces weighted average temperature.

3.1 Predictive models

In Table 3.1, all seven models are revealed in order from Model 1 to Model 7. Among these models, Models 1, 2, 4, 5, 7 are used in previous literature (e.g., Teigen and Thomas, 1995; Turvey, 2001; Vedenov and Barnett, 2004) and the rest two, Models 3 and 6, are modified from Models 4 and 5 respectively. In particular, in Model 3 three months cumulative rainfall comes into use rather than monthly precipitation as in Model 4. Similarly, in Model 5, the variable of the number of cumulative cooling degree (CDD) days is employed to replace the number of cumulative cooling degrees above such predefined temperature threshold. The goodness of fit of these seven models is presented in Chapter 4 via R^2 and adjusted R^2.

Table 3.1: Model families

Model Type	Function Form
Model 1: Quadratic in deviations	$Y_t = \alpha_0 + \alpha_1 \cdot \Delta T_t^{June} + \alpha_2 \cdot (\Delta T_t^{April})^2 + \alpha_3 \cdot (\Delta R_{cum})^2 + \alpha_4 \cdot \Delta R_{cum} \cdot \Delta T_t^{April} + \alpha_5 \cdot \Delta R_{cum} \cdot \Delta T_t^{June} + \epsilon$
Model 2: Log-log in absolute values	$\log(Y_t) = \alpha_0 + \underline{\alpha_1} \cdot \log(\mathbf{T}_t) + \underline{\alpha_2} \cdot \log(\mathbf{R}_t) + \epsilon$
Model 3: Quadratic in absolute values	$Y_t = \alpha_0 + \alpha_1 R_{cum} + \underline{\alpha_2} \cdot \mathbf{T}_t + \alpha_3 {R_{cum}}^2 + \underline{\alpha_4} \cdot \mathbf{T}_t^2 + \underline{\alpha_5 R_{cum}} \cdot \mathbf{T}_t + \epsilon$
Model 4: Quadratic in absolute values	$Y_t = \alpha_0 + \underline{\alpha_1} \cdot \mathbf{R}_t + \underline{\alpha_2} \cdot \mathbf{T}_t + \underline{\alpha_3} \cdot \mathbf{R}_t^2 + \underline{\alpha_4} \cdot \mathbf{T}_t^2 + \underline{\alpha_5} \cdot \mathbf{R}_t \mathbf{T}_t + \epsilon$
Model 5: Quadratic in absolute values	$Y_t = \alpha_0 + \alpha_1 R_{cum} + \alpha_2 CDD_{65/50} + \alpha_3 R_{cum}^2 + \alpha_4 CDD_{65/50}^2 + \alpha_5 R_{cum} CDD_{65/50} + \epsilon$
Model 6: Quadratic in absolute values	$Y_t = \alpha_0 + \alpha_1 R_{cum} + \alpha_2 CDD_{65/50/days} + \alpha_3 R_{cum}^2 + \alpha_4 CDD_{65/50/days}^2 + \alpha_5 R_{cum} CDD_{65/50/days} + \epsilon$
Model 7: Log-log in absolute values	$\log(Y_t) = \alpha_0 + \alpha_1 \log(R_{cum}) + \alpha_2 \log(CDD_{65/50}) + \epsilon$

Notations used in above models are defined as follows:

1. Y_t is forage yield (production, tons per acre).

2. $\underline{\alpha_n}$ is a parameter vector.

3. R_{cum} is cumulative rainfall between April 1st and June 30th.

4. $\mathbf{R}_t$=($R_{April,t}$, $R_{May,t}$, $R_{June,t}$), they are monthly rainfall for relatively months from April to June in the year t.

5. $\mathbf{T}_t$=($T_{April,t}$, $T_{May,t}$, $T_{June,t}$), similarity, they are average monthly temperature for April, May, and June in the year t.

6. $CDD_{65/50/days}$ is the number of cumulative cooling degree days above either 65 or 50 degrees in Fahrenheit from April 1st to June 30th.

7. $CDD_{65/50}$ is the number of cumulative cooling degrees above either 65 or 50 degrees in F from April 1st to June 30th.

8. Δ means deviations from corresponding average values.

9. ϵ is a normally distributed error with mean 0 and variance σ^2.

3.2 Missing data and trending issue

As illustrated in Chapter 2, the number of long term weather stations is considerably rare, and it causes a series problem that some counties are not provided available data during the year 1981 to 1996. For solving this perplexity, the most straightforward method is removing these missing data or we can follow these steps to work out this drawback. First of all, find out how many farms in the county which do not collect any data before 1997 and then verify the closet long term weather stations for each of these farms. If the outcome of closet long term weather station is a unique one, the unique weather station will represent the whole county climate data for its missing years. Furthermore, if there more than one results derived, the size of these farms will be compared with and then the closet long term weather station for the largest farm will be adopted to represent the county. Under the second procedure, temperature data has been stationary and consistent, which is important for analysis, and facilitates us to educe a reliable result in the end.

For forage yield trending issue, some previous literature (e.g., Vedenov and Barnett, 2004; Turvey, 2001; Wang, 2015) detrended their yield data before employing into their models. Because the capacity of each acre for forage may have raised as agriculture technology innovated in the last three decades without the impact of weather. Therefore, yield data may not be directly utilized that is a reasonable assumption. There are various techniques can be used to detrend yield data, but usually, a linear trend equation is prevailing adopted to detrend yield data. For example, Vedenov and Barnett (2004) implemented a simple detrending procedure by fitting a log-linear model. In their model, a logarithmic trended yield is derived by:

$$\log(Y_t^{tr}) = \alpha_0 + \alpha_1(t - 1971),$$

where 1971 is the beginning year of yield data and t means every year between

1971 and 2001. Then, detrended yield can be calculated out followed by:

$$Y_t^{det} = Y_t \frac{Y_t^{tr}}{Y_{2001}^{tr}},$$

where 2001 is the ending year of yield data and Y_t is original yield data at year t. Alternatively, Wang (2015) discussed piecewise spline regression, and ARIMA approaches are also potential options for detrending.

Above presentation is discussed about trending issue in yield data and detrending methods for general studying purpose. However, whether detrending in this study, should depend on nature of our data. In order to identify this problem, annually average forage yield per acre for five regions and province are calculated.

Figure 3.1 presents the annual average yield per acre in units of tonnes per acre from the year 1981 to 2003 for five regions.

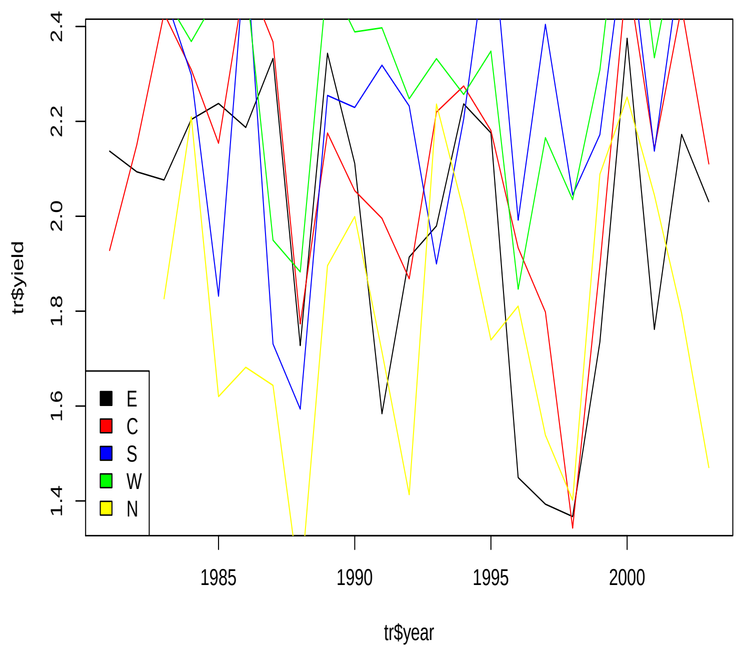

Figure 3.1: Average yield per acre from 1981 to 2003 for five regions

There is a legend on the down left corner of the figure and E means eastern region and so forth for rest letters. The interesting find from this figure is that there is no significant trend for every region during the period analysed.

Figure 3.2 also shows the annual average yield per acre in units of tonnes per acre from the year 1981 to 2003 but for the whole province of Ontario. Similarly, the outcomes fluctuate around 1.9 tonnes per acre dramatically, and it implies that there is no apparent trend for the province either.

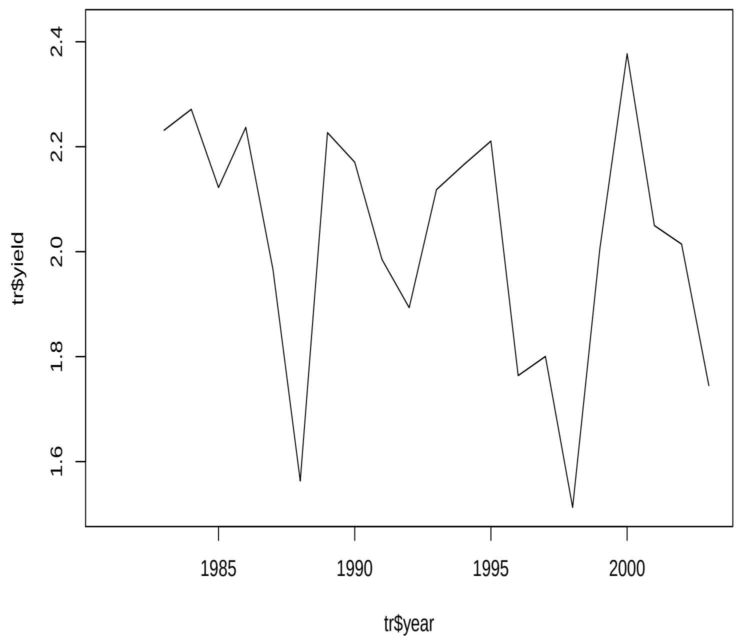

Figure 3.2: Average yield per acre from 1981 to 2003 for Province of Ontario

Consequently, detrending is needless in this study supported by sufficient evidence. Moreover, Ottman et al. (2013) also address that the trend of Alfalfa has not distinctly changed since 1980's in the western eleven states in the US. Coincidently, their research finding is same as the conclusion drawn in this analysis based on Ontario data. Therefore, the conclusion is reliable and analysis could proceed relying on this point.

3.3 Three levels of analysis unit

There are forty–seven counties, five regions and a province of data, which is to say, three levels of units can be analysed in this study. By applying the different level of analysis unit, different models are derived and then by comparing with these models, the best goodness fit one can be selected out. Later, results will show region level of analysis unit is the most appropriate level in this study.

3.3.1 County

If each individual county is chosen as the analysis unit, many R^2s and *Adjusted* R^2s will be achieved and more than that, calculating process is straightforward and simple. However, as discussed in previous chapter and sections, climate data for county level are not complete and consistent because of lacking sufficient long term weather stations. Thus, either only some counties with long term weather stations are utilized to analyse or data could be processed by the two approaches as presented in Section3.2 before introducing them into models.

The *Adjusted* R^2s' ranges of for the former method from 0% to 59.3349% which are extremely not stationary (numerical details is presented in Chapter4). It implies that choosing county level as analysis unit seems not to be decent. For later manner to process data, it is not suggested to be used because if those missing data of a given county replaced by data collected from its closet long term weather station(s)in another county and then these two counties will be homogeneous. Furthermore, the homogeneity may weaken accuracy of the goodness of models fit.

3.3.2 Five regions

We follow OMAFRA (2017) to divide the province of Ontario into five regions which are eastern, central, southern, western and northern region and detail information is in Table3.2. In the analysis unit of regions, all counties climate data within a region aggregate together, so that facilitates region level weather data to be complete and consistent. Region level data can be processed for models requirement. For instance, the eastern region contains eight single counties and temperature data are aggregated firstly and then for each day during April to June from 1981 to 2003, taking the average of all available eight counties' temperature as the temperature of eastern region for a given day. Monthly temperature can be obtained by summarizing these regions daily temperature from corresponding days and same logic for calculating cumulative temperature and even for regions precipitation.

Furthermore, quality data may output a better goodness fitting as expected. Subsequently, numerical results agree with this expectation. The ranges of $Adjusted$ R^2s before and after Akaike information criterion (AIC; Akaike, 1973) are narrow and some values are significantly high. It strongly implies region level of study unit is an appropriate selection. Moreover, Vedenov and Barnett (2004) also illustrate that a weather derivative contract should be applicable for a large geographic area. The two disadvantages of designed a contract at county level are available of climate data and a limited market for the contract. Hence, their research suggests a primary analysis unit is a level above a single county. Referring to this paper, the primary unit is region.

Table 3.2: Counties in five regions

East	Center	South	West	North
Frontenac county	Durham Regional municipality	Brant county	Bruce county	Algoma district
Leeds and Grenville United counties	Haliburton county	Chatham-Kent division	Dufferin county	Cochrane district
Lennox and Addington county	Hastings county	Elgin county	Grey county	Kenora district
Ottawa division	Kawartha Lakes division	Essex county	Halton regional municipality	Manitoulin district
Precott and Russell united counties	Muskoka district municipality	Haldimand-Norfolk regional municipality	Huron county	Nipssing district
Renfrew county	Northumberland county	Hamilton division	Peel regional municipality	Rainy River district
Stormont, Dundas and Glengarry counties	Parry Sound district	Lambton county	Perth county	Sudbury district
	Peterborough county	Middlesex county	Simcoe county	Sudbury regional municipality
	Prince Edward division	Niagara region municipality	Waterloo regional municipality	Thunder Bay
	York regional municipality	Oxford county	Wellington county	Timiskaming district

3.3.3 Province

Province level is not a suitable analysis unit as well because the whole province of Ontario is too large as a unit and it could not preferably represent and describe circumstance for counties level. Experience of a particular county may not be linked to the experience of a province.

3.4 Weighted average temperature adjustment

Weighted average temperature adjustment (WATA) is originality in this study and it means that another way to process temperature data in the case of region analysis unit. Not like computing arithmetic mean of a region inside counties as the region's corresponding temperature, WATA is defined by the following equation:

$$T_i = \sum_{j=1}^{n} \frac{c_{ji}}{C_i} t_{ji} \tag{3.1}$$

Variables are:

1. T_i: temperature of a region at time i.
2. t_{ji}: temperature of county j at time i.
3. c_{ji}: the number of acres of county j at time i.
4. C_i: the number of total acres of a region at time i.
5. n: is the total number of counties at time i.

The reason why we are introducing WATA into this paper is that when the forage yield (productivity) is calculated for a county or a region, it is a weighted average yield of all farms in fact, which is to say, large size farms have more impact on average yield than small size farms. Therefore, the more weights the farm assigned, the more impact it has in a model. Consequently, weather variables of large size farms may also have a considerable effect on models as well. The way of how to present the significant influence is assigning heavy weights on those counties that have larger forage plantation areas while calculating temperature. This is the basic idea why WATA is applied. Results from Chapter4 present when WATA is applied, for some models both R^2 and $Adjusted\ R^2$ values before and after AIC increase and meanwhile, give insight into the effectiveness of the new method WATA.

Chapter 4

Results

In this Chapter, the effectiveness of the weighted average temperature adjustment (WATA) is compared to the effectiveness of a straightforward approach. Meanwhile, the effectiveness of three analysis units is testified as well. Clearly, a higher number of R^2 or adjusted R^2, the more accurate description between forage yield and weather variables of a model has. Because R^2 is known as the coefficient of determination and it can be interpreted that how many percentages of response variables (i.e. forage yield in this thesis) can be explained by the model and parameters of variables are estimated by least squares method. However, R^2 has its drawbacks. For example, the more predictors added into the model, the higher R^2 value is obtained. Consequently, a model with more terms almost surely may achieve a better result, so that lessen the credibility of models forecasting. For this reason, R^2 is not the only measurement in our analysis and adjusted R^2 is employed as another useful statistical instrument. Adjusted R^2 only increase when a term indeed improves the model more than it would be expected which means adjusted R^2 penalises ineffective terms in the model.

Therefore, R^2 can be used to test whether an approach is effectual within different methods, and adjusted R^2 can be utilized to help us to analyse which model is optimized among all seven ones. The reason is that, when we compare two different two approaches to testify the better one, for example, straightforward way and WATA method, we need compare same models in same regions and R^2 value is such a statistical measurement satisfied this requirement. But when adjust R^2 values are computed, the underlying models are some variables truncated, which means we may achieve totally distinct models for same regions. There is no sense to compare different models employed different methods. As above discussed, we measure the model effectiveness by studying its R^2 and adjusted R^2 values. In addition, AIC has the same function as adjust R^2 and therefore, AIC is also used in this chapter to exam if these two ways provide same results in models selection.

Also, the way of processing missing temperature data discussed in Chapter 3 is solely used in WATA method, and for other approaches, those missing observations are excluded.

It is expected that region will be the best study unit among three units and WATA will substantially improve the effectiveness of all seven models forecasting and therefore obtain one or some new technique(s) to articulately reveal some details of the potential relation between forage productivity and climate factors.

4.1 Results of three analysis units

Table 4.1 demonstrates some adjusted R^2 results calculated from selected out counties which have long term weather data based on Models 1 to 7. For simplicity, only four counties are selected out and adjusted R^2 presenting as a sample here. Intuitively, the ranges of all seven models' results are apparent wide and significant not convergent. Majority values are low, including many zeros, except R^2 values of Models 3 and 4. In other words, Models 1 to 7 could not provide reasonable and convincing interpretation about forage yield and weather variables if county unit is adopted. However, from Table 4.1, a conclusion also can be drawn that Models 3 and 4 both have comparatively higher results than other models.

In Table 4.2, adjusted R^2 and R^2 values for five regions and province are presenting, and we can find in Table 4.2, adjusted R^2 values from Model 3 are more acceptable than results showing in Table 4.1, although improvement by applying region unit for Models 1 and 2 are insignificant. However, for some particular regions and other models, the improvement is sufficiently significant. For instance, referring to the central region in Model 3, the adjust R^2 is 0.5361 and moreover, its R^2 value is as high as 0.7680 which is rarely obtained from previous research in the literature. Same high results are also achieved by Model 4, and they are 0.4962 and 0.7252 respectively. Nevertheless, some insignificant results are also derived as displaying. For example, Models 3 and 4 in the northern region, Models 5 and 6 in the southern region are cases. Hence, for each region, a more accurate model can be selected out by comparing both values of adj R^2 and R^2 and therefore, the model can supply a more articulate relation to a larger area of land than county

Table 4.1: Results for counties

	M1		M2		M3		M4		M5		M6		M7	
County Number	adj R^2	R^2	adj R^2	R^2	adj R^2	R^2	adj R^2	R^2	adj R^2	R^2	adj R^2	R^2	adj R^2	R^2
26	0	0.0999	0	0.0893	0	0.4210	0	0.3665	0	0.1257	0	0.1626	0	0.0811
23	0.1043	0.3079	0	0.1728	0.0557	0.5278	0.1059	0.5123	0	0.1579	0	0.2078	0	0.06261
41	0	0.1578	0	0.1736	0.0939	0.5469	0	0.4372	0.3455	0.4942	0.4404	0.5676	0.0429	0.1299
25	0.0789	0.2882	0	0.1595	0.0292	0.5146	0.1068	0.502	0	0.1491	0	0.2277	0	0.0831

Table 4.2: Results of five regions and province

	M1		M2		M3		M4		M5		M6		M7	
Region	adj R^2	R^2	adj R^2	R^2	adj R^2	R^2	adj R^2	R^2	adj R^2	R^2	adj R^2	R^2	adj R^2	R^2
Western	0.2445	0.4162	0.0327	0.2086	0.2780	0.6390	0.0938	0.5057	0.3524	0.4996	0.3307	0.4828	0.0656	0.1506
Eastern	0.2359	0.4096	0.1818	0.3306	0.3653	0.6827	0.4111	0.6788	0.2007	0.3824	0.3430	0.4923	0.1777	0.2524
Central	0.0686	0.2803	0.1190	0.2792	0.5361	0.7680	0.4962	0.7252	0.2295	0.4046	0.2000	0.3819	0.1443	0.2220
Northern	-0.2173	0.0725	0.2085	0.3593	0.0594	0.5521	0.1416	0.5503	0.2020	0.3920	0.2108	0.3987	0.2431	0.3152
Southern	-0.0592	0.1815	-0.0124	0.1717	0.1177	0.5588	0.1069	0.5129	0.0578	0.2719	0.0562	0.2707	0.0519	0.1381
Whole Province	0.1227	0.3316	0.0433	0.2256	0.2686	0.6517	0.3346	0.6515	0.0153	0.2497	0.0352	0.2644	0.1019	0.1875

Table 4.3: The optimal model for each region and province

Region	Model	adj R^2	R^2
Western	M5	0.3524	0.4996
Eastern	M4	0.4111	0.6788
Central	M3	0.5361	0.7680
Northern	M7	0.2431	0.3152
Southern	M3	0.1177	0.5588
Whole Province	M4	0.3346	0.6515

unit.

Besides, one interesting finding is that results derived by the province analysis unit are always not the highest ones, which can support our argument in Chapter 3 that province is not an appropriate studying unit. Since province level could not provide a better interpretation of the relation between forage productivity and relative impact factors than treating the region as a research unit. For example, Model 5 in the northern region, both adjusted R^2 and R^2 of northern region are higher than values of the province.

The Table 4.3 shows the optimal model for each region and province based its adj R^2 value. Models 1 and 2 with insignificant values are not mainly studying subjects and therefore, we focus on analysing Models 3 to 7.

4.2 AIC selection

Table 4.4: AIC results of five regions and province

	M1	M2	M3	M4	M5	M6	M7
Region	R^2	R^2	R^2	R^2	R^2	R^2	R^2
Western	0.4099	0.1438	0.5420	0.4266	0.4819	0.4816	0.1438
Eastern	0.3563	0.3084	0.6786	0.6786	0.3824	0.4911	0.1942
Central	0.1947	0.1853	0.7477	0.6829	0.3508	0.3508	0.1853
Northern	0.0000	0.2753	0.4939	0.5463	0.3428	0.3492	0.2753
Southern	0.1594	0.1241	0.4672	0.4532	0.2219	0.2599	0.1241
Whole Province	0.3195	0.1552	0.6347	0.6234	0.2317	0.2423	0.1552

AIC selection can assist us to remove trivial factors in a model via assigning a penalty to extra parameters which should not be included into the model, so that meliorate forecasting credibility of the model. Usually, a model after AIC selection will achieve an AIC value and the smaller the AIC value is, the better goodness fit of the model has. However, in this thesis, an another way to interpret these AIC values is adopted. We again calculate R^2 values of each post-AIC model for each region as well. Because post-AIC models have already been removed insignificant variables and therefore, R^2 values of post-AIC models have same meaning as adjusted R^2 values, which is to say, the optimal model for each region can be obtained by comparing R^2 values. All R^2 values are presenting in Table 4.4. The optimal model for each region and province are selected out based on Table 4.4 showing and results are displaying in Table 4.5.

In Table 4.5, these optimal models are mainly concentrating on Models 3 and 4. However, in Table 4.3, the same conclusion could not be found.

Table 4.5: The optimal model for each region and province after AIC

Region	Model	R^2
Western	M3	0.5420
Eastern	M3& M4	0.6786
Central	M3	0.7477
Northern	M4	0.5463
Southern	M3	0.4672
Whole Province	M3	0.6347

4.3 Results of WATA

In order to exam whether WATA method can improve models forecasting credibility and effectiveness, results of WATA method will be compared to results of straightforward approach. However, there are two ways of selecting optimal models, which are depending on adjusted R^2 values and AIC, and therefore, they will be separately tested.

Firstly, optimal models of applying WATA and without WATA are presented in Table 4.6 by considering adjusted R^2 values. From this table, region of east, center and south have same optimal models and meanwhile, they have both higher adj R^2 and R^2 numbers. Referring to the western region, optimal model is same, Model 5, and results are exactly same as well. For region of north and whole province level, their optimal models altered but after applying WATA method, M6 and M3 can provide more accurate and credibility explanation between forage yield and climate variables for region of north and whole province respectively. Consequently, a conclusion can be drawn that adopting WATA method improves models predictive effectiveness.

Table 4.6: The results comparison without AIC

With WATA	Model	adj R^2	R^2	Without WATA	Model	adj R^2	R^2
Western	M5	0.3524	0.4996		M5	0.3524	0.4996
Eastern	M4	0.4599	0.7054		M4	0.4111	0.6788
Central	M3	0.5524	0.7762		M3	0.5361	0.7680
Northern	M6	0.3191	0.4893		M7	0.2431	0.3152
Southern	M3	0.3146	0.6573		M3	0.1177	0.5588
Whole Province	M3	0.3922	0.7106		M4	0.3346	0.6515

Similarly, second step is to compare optimal models of employing WATA and without WATA method. However, optimal models in this step are chosen according to each corresponding maximum R^2 value of every post-AIC model for each region. From Table 4.7, we can see that they have same selections of optimal models and also, eastern, southern and whole province results of using WATA all have significant higher R^2 values than results without adopting WATA. But the region of west, north and center, their R^2 values of WATA are slight lower than R^2 values without WATA. However, deviations of these R^2 values are insignificant and negligible. For detail results of adopting WATA method, they are showing in Table 4.8.

Table 4.7: The results comparison with AIC

With WATA	Model	R^2	Without WATA	Model	R^2
Western	M3	0.5342		M3	0.5420
Eastern	M3	0.7224		M3& M4	0.6786
Central	M3	0.7471		M3	0.7477
Northern	M4	0.5341		M4	0.5463
Southern	M3	0.6160		M3	0.4672
Whole Province	M3	0.7097		M3	0.6347

Notably, Model 3 in the southern region, its post-AIC R^2 values from 0.4672 tremendously raises to 0.6160 which increased 32%. But for other models and regions, this phenomenon does not appear. In order to interpret this, we propose an explanation.

Deviation in Table4.9 means the difference between the largest weight of a county and smallest weight of a county for a given year. Therefore, we find that the deviations of southern region are obliviously big than other regions and it may be the reason for the dramatical change in the southern region. It implies that WATA method might be extremely effective for places that have an unequal distribution of plantation land areas. However, the northern region also shows large deviation numbers from the year 1998 to 2003, and increments of R^2 values of major four models after applying WATA method are insignificant.

Table 4.8: Results of WATA

Region	M1		M2		M3		M4		M5		M6		M7	
	adj R^2	R^2	adj R^2	R^2	adj R^2	R^2	adj R^2	R^2	adj R^2	R^2	adjR^2	R^2	adj R^2	R^2
Western	-0.1296	0.1271	0.0283	0.2050	0.2322	0.6161	0.0216	0.4663	0.3524	0.4996	0.3307	0.4828	0.0656	0.1506
Western (after AIC)	0	0	0.1030	0.1438	0.4606	0.5342	0.2487	0.3511	0.4001	0.4819	0.3998	0.4816	0.1030	0.1438
Eastern	0.3459	0.4946	0.3970	0.5066	0.4926	0.7463	0.4599	0.7054	0.1924	0.3760	0.3348	0.4860	0.1748	0.2498
Eastern (after AIC)	0.4007	0.4551	0.4410	0.4918	0.5637	0.7224	0.5538	0.6755	0.1802	0.2174	0.3696	0.4842	0.1508	0.1894
Central	0.0741	0.2846	0.0916	0.2567	0.5524	0.7762	0.5370	0.7475	0.2221	0.3989	0.1297	0.3275	0.1326	0.2115
Central (after AIC)	0.0871	0.2531	0.1194	0.1594	0.6026	0.7471	0.5956	0.7427	0.2320	0.3018	0.2320	0.3018	0.1194	0.1594
Northern	-0.1103	0.1673	0.2095	0.3676	0.0968	0.5936	0.1465	0.5733	0.2787	0.4590	0.3191	0.4893	0.2863	0.3577
Northern(after AIC)	0	0	0.3059	0.3406	0.4206	0.5075	0.3344	0.5341	0.3825	0.4443	0.3825	0.4443	0.3059	0.3406
Southern	-0.0186	0.2129	0.0366	0.2117	0.3146	0.6573	0.1447	0.5335	0.1324	0.3296	0.0726	0.2834	0.0949	0.1772
Southern(after AIC)	0.0631	0.1908	0.1119	0.1523	0.4368	0.6160	0.2949	0.4551	0.1997	0.2725	0.1997	0.2725	0.1119	0.1523
Whole Province	0.1679	0.3660	0.0285	0.2136	0.3922	0.7106	0.3768	0.6736	0.0153	0.2497	0.0352	0.2649	0.1019	0.1875
Whole Province(after AIC)	0.2039	0.3555	0.1095	0.1519	0.4459	0.7097	0.4287	0.6735	0.1469	0.2282	0.1469	0.2282	0.1095	0.1519

Table 4.9: Deviations of regions

Deviation	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991
west	0.1610	0.1998	0.1557	0.1831	0.1525	0.1508	0.2001	0.1624	0.1430	0.3103	0.2502
east	0.3985	0.3183	0.4589	0.2653	0.2764	0.2595	0.2245	0.1702	0.3231	0.2099	0.2440
centre	0.3030	0.3420	0.1881	0.2234	0.2268	0.2249	0.2365	0.1887	0.1463	0.0890	0.1489
north	0.2892	0.2500	0.4179	0.2289	0.2240	0.2089	0.2337	0.2648	0.2166	0.2894	0.2283
south	0.4250	0.5284	0.4324	0.3309	0.4207	0.3827	0.4677	0.4296	0.2292	0.1853	0.1996
1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
0.2753	0.1838	0.2443	0.2561	0.2577	0.3149	0.2846	0.1901	0.2051	0.2288	0.2562	0.2755
0.2441	0.4416	0.2483	0.2365	0.2889	0.3056	0.3565	0.3157	0.3775	0.2562	0.4443	0.4611
0.0813	0.1240	0.1838	0.1311	0.1351	0.1484	0.3034	0.1602	0.2517	0.3384	0.2655	0.2102
0.2375	0.2354	0.2885	0.1775	0.1756	0.2775	0.6294	0.4153	0.5081	0.5470	0.5134	0.5486
0.1551	0.2233	0.1898	0.2671	0.2704	0.2362	0.2341	0.3954	0.2113	0.5349	0.2173	0.2486

Chapter 5

Summary and Future Research

5.1 Thesis summary

In this thesis, we investigate best predictive models explaining the relationship between forage crops yield and weather variables in the province of Ontario, Canada. Here, climate variables are for each month from April to June, including cumulative rainfall, monthly temperature in degrees, and monthly days with temperature in a particular range. The outcome of this study is expected to be utilized to design indices and to set triggers for CAT bonds on forage crops in Ontario.

Our main contributions are summarized as follows:

(i) We deal with missing weather observations. For example, lots of temperature observations are missing in the original datasets. We assume that for each county its temperature is as recorded by the closest weather station. Thus we are able to use the forage crop yield information from all counties fully.

(ii) We propose to search best predictive models based on five geographical regions. This is different from most research in the literature, in which predictive models are applied on one single farm or county.

(iii) We propose a new concept: weighted average temperature adjustment (WATA). The motivation is very straightforward. Actually, in each region, the cultivated area of forage crops varies a lot among counties every year. Thus the weighted average temperature with cultivated area being weight makes more sense in predicting forage crop yield than the arithmetic mean temperature does.

5.2 Future research

We would use the research output from this thesis in designing an appropriate CAT bond for agricultural insurance business in Ontario. There are several ways to design a CAT bond. For example, some CAT bonds use aggregate Loss Cost Ratio (LCR) as a trigger (Ye et al., 2013) and others use weather indices (Sun et al., 2015). We are interested in the index approach. In the following, we summarize the next steps and challenges that need to overcome to design a CAT bond eventually.

1. Predicting precipitation and temperature:
   (a) Based on historical weather data, we need to estimate the frequency and severity for extreme weather conditions. This involves using statistical models to estimate the occurrence of extreme weather conditions.
   (b) The weather variables in the five geographical regions of Ontario are likely to be correlated with each other. So we should investigate the

possibility that same severe weather condition may occur in more than one region in a single year.

2. Setting bond trigger:

   (a) The bond triggers may be set based on precipitation, temperature or both. Here we again need to take into account the spatial correlation between the five regions since the triggers represent the levels of catastrophic risk in Ontario as a whole.

3. Pricing the CAT bond:

   (a) Since CAT bonds linked to agriculture has not been issued yet, the market for CAT bonds is incomplete. So the pricing is a complicated and difficult work, and the martingale approach does not provide a unique price.

   (b) Jarrow (2010) proposed a closed for CAT bond pricing model, which may help us complete the pricing step.

   (c) Setting the return rate is also an important step. In order to attract investors to purchase and trade CAT bonds, CAT bonds have to provide a higher return rate than the risk-free rate. Usually, the recovery rate is assumed as LIBOR rate plus a fixed spread as shown in Sun et al. (2015).

4. Plausibility checking:

   (a) This step uses Monte Carlo simulation for precipitation and temperature. They, in turn, are used as inputs in our predictive models. The technique of importance sampling may help improve the accuracy and efficiency of the simulation.

(b) The plausibility of the CAT bond relies on if both stakeholders, the federal government, and the investors, are satisfied with the probability of default, hedging effectiveness, profitability, etc.

Bibliography

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle.*Proceedings of the Second International Symposium on Information Theory* (B. N. Petrov & F. Caski (Eds.)), Budapest: Akademiai Kiado.

Barnett, B. and Mahul, O. (2007). Weather index insurance for agriculture and rural areas in lower income countries.*American Journal of Agricultural Economics*, 89(5): 1241–1247.

Barret, C. B., Barnett, J. B., Carter, M. R., Chantarat, S. Hansen, J. W., Mude, A. G., Osgood, D. E., Skees, J. R., Turvey, C. G., and Ward, M. N. (2007). Poverty traps and climate risk: limitations and opportunities of index-based risk financing. *International Research Institute for Climate and Society Technical Report* 07-02.

FAO. (2009). Global Agriculture Towards 2050. *How to Feed the World 2050, High Level Expert Forum*. Accessed January 2015.

Fan, S. (2009). Setting Priorities for Public Spending for Agricultural and Rural Development in Africa. Policy Brief 12, IFPRI.

Government of Canada. (2011). Agriculture and Agri-Food Canada. *Evaluation of the Agricultural Disaster and Relief Program*. Office of Audit and Evaluation.

Hellmuth, M. E., Osgood, D. E., Hess, U., Moorhead, A., and Bhojwani, H. (2009). Index insurance and climate risk: Prospects for development and disaster management. Technical Report No. 2, Climate and Society, International Research Institute for Climate and Society, New York, USA.

Jarrow, R. A.(2010). A simple robust model for CAT bond valuation, *Finance Research Letters*, 7:72–79.

Ottman, M., Putnam, D., Barlow, V., Brummer, J., Bohle, M., Creech, E., Foster, S., Long, R., Marsalis, M., Norberg, S., Orloff, S., and Shewmaker, G. (2013). Long term trends and the future of the alfalfa and forage industry. *Western Alfalfa and Forage Symposium*,December:11-13.

OMAFRA. (2017). Area of census farms (acres) by county, Ontario: 1991, 1996, 2001, 2006, 2011 and 2016. Available at OMAFRA: http://www.omafra.gov.on.ca/english/stats/census/cty30a.htm.

Porth, L. and Tan, K., S. (2015). Agricultural Insurance More Room to Grow? *The Actuary, Society of Actuaries*, April/May 2015:36-41.

Porth, L., Pai, J., and Boyd, M. (2014). A portfolio optimization approach using combinatorics with a genetic algorithm for developing a reinsurance model. *Journal of Risk and Insurance*.

Simpson, A. (2016) An analysis of rainfall weather index insurance: the case of forage crop in Canada. Master thesis. Available at online: https://mspace.lib.umanitoba.ca/handle/1993/31247.

Teigen, L. D., and Thomas, M. (1995). Weather and yield, 1950–1994. *Economic Research Service*, USDA, Washington, DC, Staff paper:9527.

Turvey, C. G. (2001). Weather derivatives for specific event risks in agriculture. *Review of Agricultural Economics*, 23(2):333–351.

USDA (2014). World agricultural supply and demand estimates report (WASDE). Technical report, Office of the chief economist (OCE), United States Department of Agriculture.

Vedenov, D.V., and Barnett, B. J. (2004). Efficiency of weather derivatives as primary crop insurance instruments. *Journal of Agricultural and Resource Economics*, 29(3):387–403.

Woodard, J. D., Schnitkey, G. D., Sherrick, B. J., Lozano-Gracia, N., and Anselin, L. (2012). A spatial econometric analysis of loss experience in the U.S. crop insurance program. *Journal of Risk and Insurance*, 79(1):261286.

Wang, S. (2015) The design of weather index insurance for forage: the case of basis risk for the Canadian province of Ontario. Master thesis. Available at online: https://mspace.lib.umanitoba.ca/handle/1993/30835.

Sun, L., Turvey, C. G. and Jarrow, R. A. (2015). Designing catastrophic bonds for catastrophic risks in agriculture: Macro hedging long and short rains in Kenya. *Agricultural Finance Review*, 75(1):47–62.

Ye, W., Q., Poh, R., Wei, C., Zhou, S., Porth, L. and Weng, C., G. (2013). Securitization of Government of Canada risk exposure to agricultural disasters : A frame work for catastrophic bonds. *International Agricultural Risk, Finance and Insurance Conference*. Vancouver: June 16–18.

Zhu, W.,Tan, K. S., Porth, L. and Wang, C−W. (2015). Spatial Dependence and Aggregation in Weather Risk Hedging. Available at SSRN: https://ssrn.com/abstract=2669114 or http://dx.doi.org/10.2139/ssrn.2669114