SEASONAL KEY POINT BASED CROP PREDICTION; AN EVALUATION OF SPRING WHEAT IN NORTH DAKOTA AND THE CANADIAN PRAIRIE, AND COCOA IN IVORY COAST

A pilot study linking precipitation to price changes

Abstract

The main objective for the research is the development of a comprehensive and easily accessible method assessing the price changes of wheat and cocoa by predicting change in wheat and cocoa production using season key point in the growing season of wheat and cocoa and open precipitation data from ground based weather stations only. The found results are assessed towards their impact on the movement of wheat and cocoa prices.

J.A.A. Hogewoning

1384759

"Truth is ever to be found in simplicity, and not in the multiplicity and confusion of things"

- Sir Isaac Newton-

Preface

During a period of 6 months, this thesis was written as a graduation thesis for the master Water Management of the faculty of Civil Engineering of the TU Delft. The thesis was written during an internship at Ernst & Young (EY) Netherlands between the 1st of November 2016 and the 30th of April. The goal of the thesis was to find a common ground between engineering and finance through crops. Both finance and water management have a relation to crops, yet in a very different way. By looking for a common ground, maybe it is possible to bringing two different disciplines together.

I would like to thank my whole assessment committee and supervisors for their guidance, patience and inspiration. To Nick van de Giesen, Martine Rutten and Jos Timmermans, for sharpening my ideas and encouraging me to push harder as an engineer. To Floris van de Loo and Rob Balk, for helping me settle at EY and helping me concentrate not only on the theoretical, but also practical. I could not have created this thesis without your help and guidance. I would also like to thank Kasper Keizer and Dorien Lugt, for their contribution in the form of a review of the thesis.

I would also like to thank my parents and parents-in-law, brother and sister, girlfriend, friends, family, colleagues and anyone that was willing to listen to me going on about weather patterns. I have many people to give thanks to. To mom and dad, for never losing faith in that I could do it. To Jet and Erik, for providing comfort for doing something different. To Kees and Lies, to focus me on my studies, not everything around it. To Hans and Sylvie, for your support and encouragements. And of course, to my girlfriend Sophie, with whom I have been on this journey since the beginning. I couldn't have done it without out you. It has been a long journey. My last acknowledgement goes out to Fientje, who would have been the proudest of them all.

Index

Executive summary

In this research, a proof of concept is made to research the possibility of using publicly available rainfall data from weather stations in areas where wheat and cocoa is produced to make a predictive crop production model. The emphasis was put on using readily available online data to ensure the accessibility to the method to both people with an engineering and a financial back ground. The premise is that by using reports and crop specific requirements, seasonal key points of a crop season can be defined. Next, regression modelling between relative change in crop yield for wheat and crop production for cocoa, combined with the distinguished seasonal key points was performed. The process of distinguishing seasonal key points in a season using both qualitative and quantitative data, and then carrying out a regression using relative change in crop yield and production is introduced in this research as seasonal crop related key points prediction. The research provides a start in linking water management and finance together. From reports of the Food and Agricultural Organization (FAO) and the International Cocoa Organization (ICCO), wheat and cocoa prices were found to be dependent of supply and demand, with supply dependent on the production of the crops. The production of wheat and cocoa was found to be dominated by the effects of the weather. Therefore, to understand the effect of the weather on the production, is the first step in understanding the link from water management to finance.

For this research, wheat and cocoa production was assessed. As a first step, a qualitative review of wheat and cocoa was carried out to define what drives market prices of wheat and cocoa by using reports from the FAO and ICCO. While not the only factor, it was found that the weather plays an, if not the most important, role in the change of the prices of wheat and cocoa. The main drivers of wheat and cocoa are weather related production deficit, political and economic instability and expected surplus or deficit in production based on weather related news. The weather is found to be the main driver concerning the production of wheat and cocoa and thus the supply of wheat and cocoa according to the reports from the FAO and ICCO. To outset of this research is therefore to quantify the link between wheat and cocoa production versus the weather.

The price behaviour of wheat was assessed between 1960 and 2015 to distinguish the largest price increases and decreases. The largest price increases and decreases within one wheat year were assessed, being between June and May. All annual price changes were assessed, and as a selection for this thesis, the top 5 price increases and decreases were highlighted by using the "State of Food and Agriculture" reports from the FAO. Four of the top 5 price increases and decreases could be related to weather, confirming the strong influence the weather has on the prices via the impact the weather has on production. Price changes can also be used to indicate interesting years to research the cause of price increases from a wheat production perspective. By using reports concerning crop production from the FAO, and determining the wheat specific needed growing requirements from the FAO, the seasonal key points for wheat could be distinguished.

The price behaviour of cocoa was assessed between 1960 and 2015 to distinguish the largest price increases and decreases. The largest price increases and decreases within one cocoa year were assessed, being between October to September. All annual price changes were assessed, and as a selection for this thesis, the top 5 price increases and decreases are highlighted, as well as all the years between 2002 and 2015 by using annual reports from the ICCO and articles from the World Bank and World Cocoa Foundation (WCF). All of the top 5 price increases and decreases and production changes between 2002 and 2015 could be related to weather, confirming the strong influence the weather has on the prices via the impact the weather has on production. Price changes can also be used to indicate interesting years to research the cause of price increases from a cocoa production perspective. By using reports concerning crop production from the ICCO, World Bank and WCF, and determining the cocoa specific needed growing requirements from the ICCO and WCF, the seasonal key points for cocoa could be distinguished.

The research determined the dependency of the world wheat production on the wheat production of the USA and Canada. Over 64% of the variance in world wheat production was found to be explained by assessing the wheat production of the USA and Canada, confirming the important role of these countries concerning world wheat production.

The research also confirms the high dependence of the world cocoa production on the cocoa production of Ivory Coast and Ghana. Over 92% of the variance in world cocoa production was found to be explained by assessing the cocoa production of IC and Ghana, confirming the dominant role of these countries concerning world cocoa production.

For wheat, North Dakota and the Canadian Prairies were used as testing area to assess the dependency of the spring wheat yield to weather patterns in North Dakota and the Canadian Prairies. Data concerning yield was collected from the North Dakota State University Hettinger Extension Centre (NDSU HREC) and the ministry of Alberta, and data concerning rainfall was collected from the North Dakota Agricultural Weather Network (NDAWN) and the Koninklijk Nederlands Meteorologisch Instituut. Both the yield and rainfall data was collected over a period of 1997 to 2015 for North Dakota and over a period of 1961 to 2015 for the Canadian Prairies. An optimization process of first filtering and then using rainfall data of a specific month or parts of a crop year as defined by the reports from the FAO produced a method to define seasonal key points that affect wheat yield. Weather station data from the NDAWN and KNMI could subsequently be used to support the found key points in a season in a quantitative way. By using wheat production and seasonal key points, a regression based model could be formed to explain 54% of the variance of the spring wheat yield in North Dakota and 38% of the variance of the spring wheat yield in the Canadian Prairies.

For cocoa, Ivory Coast was used as testing area to assess the dependency of the world cocoa production to weather patterns in Ivory Coast. Data concerning production was collected from the ICCO and data concerning rainfall from the Koninklijk Nederlands Meteorologisch Instituut (KNMI). Both the production and rainfall data was collected over a period of 1960 to 2015. An optimization process of first filtering and then using rainfall data of a specific month or parts of a crop year as defined by the reports from the ICCO and WCF produced a method to define seasonal key points that affect cocoa production. Weather station data from the KNMI could subsequently be used to support the found key points in a season in a quantitative way. By using cocoa production and seasonal key points, a regression based model could be formed to explain 47% of the variance of the world cocoa production.

1. Introduction

In 1999, it was estimated that 1 trillion dollars of the 7 trillion-dollar economy of the United States alone was affected by weather (Hanley, 1999). Agriculture is one of the parts of the economy that is found to be dependent of the weather, as discussed throughout books such as "*Weather Economics*" by James Taylor (Taylor, 1968). While in the USA only 1% of the 2014 GDP is accounted for by agriculture, in other countries such as Nigeria and India the role of agriculture in the GDP can get as high as 15% (CIA, The World Factbook, 2014). Agriculture therefore plays an important role in both local and global economy. The agricultural sector is a producer of products to satisfy wants or needs, better known in finance as commodities. The weather is seen as a key driver concerning market prices of these agricultural or soft commodities (Trostle, 2008; Wright, 2009).

Agricultural commodities are commodities such as wheat, soybeans, cocoa, coffee and corn. Attempts have been made to isolate the main drivers of the prices and volatility of agricultural commodities using weather related causes. Especially wheat has been the subject of much research, due to the dominant status the crop has as one of the main diet staples worldwide. In these researches, adverse weather conditions and weather shocks are seen as being amongst the strongest drivers for commodity prices (Brockhaus, 2016; Algieri, 2016). The fact that weather affects crop yields and therefore crop prices is underlined. However, the exact meaning and implications of these so called adverse weather events is not described in the researches in a quantitative manner.

Climate change is expected to bring more weather extremes, which can have a severe impact on the agricultural industry worldwide (Rosenzweig, 2001) and therefore the agricultural commodity prices. These effects, such as droughts, and the reduction in crop harvest they cause, can be seen all around the word. It can be noted that Sub Sahara Africa (SSA), is hit particularly hard, with an expected reduced agricultural production of up to 25% compared to 2007 in the next 50 years (Cline, 2007).

In agricultural commodities, weather and finance come together. The weather aspect is responsible for the development and growth of the agricultural commodities, while the final production of the agricultural commodities affects the prices. Agricultural commodities can therefore be seen from two perspectives, being the weather related hydrological perspective and a financial perspective.

Financial perspective

From a financial perspective, the link between finance, crop yields and weather has been made. From France, the news company france24 reported "catastrophic" low wheat yields in 2016 due to heavy rains and little sunlight, resulting into higher wheat prices (France 24, 2016). In the Netherlands, Metro reported that a portion of French fries would become more expensive in 2016 due to bad harvest. The main reason behind this bad harvest was linked back to bad weather, with heavy rainfall in June as the main catalyst (Metro, 2016). Shortages in wheat production due to bad weather have even been linked as a driver for the Arabic Spring (Werrell & Femia, 2013).

In finance, attempts have been made to link seasonal agricultural commodities such as crops to financial predictive models. The focus is mainly on reducing risk in futures (Giot, 2003) or trying to predict the behaviour of prices in seasonal commodities by using models used in other commodities such as oil (Pirrong, 2011). In the latter, the new prices of for instance wheat are predicted using a seasonal production and seasonal storage of wheat. The model however is as of recent unsuccessful, with Pirrong stating that: *"The seasonal storage model fails rather spectacularly in characterizing the high frequency behaviour of periodically produced commodities"*, with the high frequency behaviour being the movement of the prices of agricultural commodities as a reaction to production.

Hydrological perspective
From a hydrological perspective, adverse weather and crop yields have been linked by multiple researchers (Antwi-Agyei, 2012; Changnon, 2012), with adverse weather such as floods, droughts and tornados affecting crop yields and subsequently crop prices. Research organizations such as the Food and Agricultural Organization (FAO) have performed research on the impact of weather on crop production. In a report from the FAO, links between water and crops yields are discussed, focusing on vital parts in the growth development of crops and water availability (FAO, 2012). The FAO states "*water has always been the main factor limiting crop production in much of the world"* (FAO, 2012).

Understanding and forecasting weather can therefore be interesting from both a financial and a hydrological perspective for agricultural commodities. However, even with modern technology, the forecasted weather has validity for about 10 days into the future, before accuracy drops. Although improvements are being made to push reliability past the 10-day mark, there is still room for improvement of accuracy beyond the 10-day mark and further (Bauer, 2015). Moreover, in the tropics, even less is known about the accuracy of weather prediction. Due to the chaotic nature of the weather in the tropics, predictions are hard to produce and have a shorter shelf life than in ex tropical areas, with weather in the tropics dominated by convective processes, as opposed to pressure and temperature differences in the mid-latitude areas of the world (Shukla, 1998).

For this research, two crops are researched concerning their link between weather, production and price changes. To keep the research diverse, a crop from inside the tropics and outside the tropics were chosen, due to the differences between tropical and extropical weather. As crop outside the tropics, wheat is selected, as crop within the tropics, cocoa is selected.

Wheat
As extropical crop, wheat was selected. As of 2013, the gross production value of wheat was estimated to be worth 198 billion USD (FAOSTAT, 2017). In the 1999 "Wheat: post-harvest operation" report, the FAO states that "*Wheat has been the staple food of the major civilisations in Europe, Western Asia, and North Africa for 8,000 years*" (FAO, 1999). The industry has benefited from the so called "Green revolution", which is defined by Cleaver as "*the rapid growth in Third World grain output associated with the introduction of a new package of tropic agricultural inputs. The package consists essentially of a combination of improved grain varieties, mainly rice and wheat, heavy fertilizer usage and carefully controlled irrigation. Without fertilizers or without controlled irrigation, the new varieties usually yield no more and sometimes less than traditional strains*" (Cleaver, 1972).

However, due to the effects of climate change, freshwater availability is expected to decrease in low latitude region that include heavily irrigated areas in India, China and Egypt (FAO, 2015). The FAO also quotes that "*Irrigation has been an important contributor to yield growth that underpinned much of the production increases over the past decades*" (FAO, 2012). Due to the expected effects of climate change, there is uncertainty and insecurity for wheat production towards 2050 (Rosegrant, 2009).

Due to the water dependency and subsequent weather dependency of wheat, research has been performed towards understanding the effect of water and weather on wheat (FAO, 2012). When assessing the "State of Food and Agriculture" (SOFA) reports from the FAO, weather events in Canada, the USA, Australia and Russia are related to lower crop yields and subsequent higher wheat prices. An example is for instance the price increase in wheat in 2010 due to "excess rain" in the USA and a heat wave in Russia. The heat wave in Russia triggered an export ban on Russian grain, stimulating market prices. (FOA, 2011; Oxfam, 2011). The reason for the importance of in particular Canada, the USA, Australia and Russia is due to the high contribution of the world wheat export these countries provide. In 2016, the USA was responsible of 14.8% of the world wheat export, with Canada second with 12.4%, Russia third with 11,6% and Australia fourth with 9.9% (CIA, 2017).
The areas of Canada, the USA, Australia and Russia are not the largest wheat producing countries in the world. As of 2017, China and India are the largest wheat producing countries in the world, but have a limited impact on the world wheat export and subsequent price of wheat due to high domestic consumption (FAOSTAT, 2017). Therefore, to research the impact of the weather on prices, the impact of the weather on the wheat production in the top exporting countries is evaluated.

From the SOFA reports of the FAO, the weather is mentioned in multiple reports as being an important factor concerning the price changes of wheat (FAO, 1977; FAO, 1995; FAO, 1992; FAO, 2008). In the reports, weather is described for specific areas with the focus on Canada, the USA, Australia and Russia, with descriptions such as drought or rainfall surplus. However, the weather is also indicated as "adverse" or "not optimal". The goal for this research is to determine the exact implication of the weather effects on the production of wheat. For this research, the top two wheat exporting countries, being the USA and Canada, are selected for further research concerning the link between weather and production.

Cocoa
As tropical crop, cocoa was selected. As of 2011, global sales of chocolate confectionary surpassed the 100-billion-dollar mark (WCF, 2012). While demand is constantly growing, the world cocoa production cannot always cope with the demand, leading to a higher frequency of deficits in cocoa production from 2001 onwards (ICCO, 2010). The world cocoa production is increasing, but is lagging behind demand. As a consequence, the cocoa market and subsequent cocoa prices can react volatile as a reaction to news from cocoa producing areas concerning cocoa production and distribution. Examples of this phenomenon are an increase of the cocoa price due to a coup d'état in Ivory Coast in 2001, an increase of the cocoa price due to dry weather in West Africa in 2004, an increase of 49% of the price of cocoa in 2007 due to news about and the effects of a *Harmattan*[1] in West Africa, but also a price decrease due to news about large amounts of rain in June in 2003 and a price decrease due to lower demand as a reaction a an economic crisis (ICCO, 2010).

From reports concerning cocoa from the International Cocoa Organization (ICCO), West Africa is mentioned in all reports as being an important factor concerning the changes in cocoa prices, with weather playing a key role (ICCO, 2008; ICCO, 2009; ICCO, 2010; ICCO, 2015). The indicated reason from the ICCO for the important role of West Africa concerning price changes, is the large contribution of West Africa to the world cocoa production. In 2014, over 70% of all of the cocoa in the world was produced in West Africa, with Ivory Coast (43%) and Ghana (21%) as largest producers (ICCO, 2016).

[1] A *Harmattan* is "*a dry dusty wind from the Sahara blowing towards the West African coast, especially from November to March*" (William Collins Sons & Co. Ltd., 2012).

In addition to in the reports from the ICCO, weather is also indicated in multiple other reports as main driver for the production of cocoa in West Africa (Bloomberg, 2017; WCF, 2012). In particular, the wet season, dry season or both seasons in West Africa are mentioned in all of the indicated reports as being the strongest indicators for a successful cocoa producing year. However, no quantitative link between cocoa production and the wet and dry season is defined.

As indicated, the cocoa market can be a volatile market. The problem with a volatile market, is the indicated fact that news concerning cocoa producing areas gains an important role. As of 2016, the market is highly concentrated, with the Stichting voor Economisch Onderzoek (SEO) stating that *"Due to recent mergers and acquisitions, the markets for chocolate manufacturing and processing are now dominated by a handful of companies"* and *"The market for cocoa processing is the most concentrated market segment of the value chain. UNCTAD (2015) estimates that in 2013 three very large agribusiness companies (ADM, Barry Callebaut, and Olam) controlled around 60 percent of the world's cocoa grindings"* (SEO , 2016). Risk protection services, such as production forecasts and price movement, are offered by companies such as Cargill to cope with the volatile prices of cocoa (Cargill, 2017). Due to the dominant position of the large companies, information services concerning risk protection and crop prediction of cocoa are however inaccessible to the public due to high prices. Next to the dominance of a number of companies, countries such as Ivory Coast are very secretive about the cocoa industry. Bloomberg stated in an interview that "*Ivory Coast's cocoa industry is notoriously secretive*" (Bloomberg, 2017).

The research

The main goal of the research was to link weather to price changes of agricultural commodities. The start of the research was to understand price change of wheat and cocoa, and to determine what drives prices of wheat and cocoa to either increase or decrease. The objective was to qualitatively understand price changes, determine the general dynamics of price changes and whether the weather had influence on the production and price change of wheat and cocoa, and what time period the weather was important and how strong the influence of weather was.

Second, the relative importance of the wheat production of the USA and Canada towards the world wheat production and the relative importance of West Africa, with the emphasis on Ivory Coast and Ghana, towards the world cocoa production was determined. Next, as an experiment, the wheat price via the production of wheat in North Dakota (ND) and the Canadian Prairies was linked to rainfall in specific time periods and the cocoa price via world cocoa production is linked to the rainfall in a specific time period in Ivory Coast. The reason for the choosing North Dakota, the Canadian Prairies and Ivory coast is elaborated on further in the introduction of wheat in chapter 2.3 and in the introduction of cocoa in chapter 5.3. The hypothesis is that due to the large dependency of world wheat export on ND and the Canadian Prairies and the world cocoa production on West Africa in a specific seasonal time period as indicated in the reports concerning wheat and cocoa, the possibility exists that the world export of wheat and the world production of cocoa can be predicted using rainfall data from the specified time period. The specific seasonal time periods in a wheat and cocoa growing season are defined as key points in the wheat and cocoa producing season. As a proof of concept, a linear regression between rainfall data from ND and the Canadian Prairies is used in combination with the relative change in yield in ND and the Canadian Prairies and to assess and quantify the link. For cocoa, a linear regression between rainfall data from Ivory Coast is used in combination with the relative change in world production of cocoa to assess and quantify the link.

The approach to crop production prediction in this research is an assessment technique, taking seasonal key points in a wheat and cocoa production year as indicated by wheat and cocoa production related reports qualitatively and supporting the key points quantitively with weather and production data. The technique uses open data from ground based weather station that are available online, to ensure full accessibility to the technique. The goal is to determine if it is possible to use reports concerning wheat and cocoa and rainfall data from weather stations to produce a method to predict wheat yield and cocoa production. By using readily available data, the problem of de secrecy and limited availability of news can be circumvented, while preserving the validity of the wheat and cocoa production prediction. As a final step, the found key points are validated using reports concerning the price movement of wheat and cocoa as provided by the FAO and ICCO.

As mention in this introduction, the dynamics of the wheat and cocoa market are assessed on the links from weather to price via supply. An assumption of this research is that the economy is demand and supply driven. If the supply is lower than the demand, prices increase due to scarcity of goods. By quantifying the link from weather to supply, the impact of the supply towards the price can be assessed.

The report is divided into three parts, being first the evaluation of wheat, then the evaluation of cocoa. After the evaluation of both crops, a common discussion and conclusion is made concerning the whole thesis.

1.1 The use of multivariate linear regression, P- and F-Values

In this research, multiple regressions were performed. In this section, the use of regressions is further explained. To correlate local production to world wheat and cocoa production and seasonal key points to spring wheat yield and world cocoa production, regressions were performed. For all the regressions, the Windows Excel tool "data analysis - regression" has been used. Al regressions in this research were conducted as linear regression functions as a first approximation.

The output of the regressions in Windows Excel delivers a set of values, of which the F- and P-values, R^2 values and coefficients related to the variables used in the regression were used in this research. For the production, the world wheat and cocoa production was used as dependent variable and the wheat production of the USA and Canada, and the cocoa production of Ivory Coast and Ghana as independent variables. For weather and change in production, the crop production is used as dependent variable and the weather as independent variables.

P- and F Values

P- and F-values are values representing significance of the found relations, with significance defined as "*of or relating to observations that are unlikely to occur by chance and that therefore indicate a systematic cause*" (William Collins Sons & Co. Ltd., 2012). The P-value is used to determine the significance of the independent variables, while the F-value is used to determine the significance of the regression. The P-value provides the significance level of a variable towards a regression with an independent variable. The null hypothesis stated for a P-value test is that the independent variable has no contribution towards a regression with the dependent variable. If the P-value is below a set significance level, the null hypothesis is rejected and the independent variable is perceived as being of importance towards the dependent variable (Minitab, 2017). For the F-value, the null hypothesis is that the model would also function without the independent variables, with a rejection of the null hypothesis being perceived as a dependency of the dependent variable on the independent variables (Minitab, 2017). The P and F-values are used to test the validity of the found results by implying a confidence interval to the P- and F-values. A full description of the multi variate regression and P and F-values are given in appendix VI.

For the regression to be accepted in this research, all the P and F values needed to be either equal to or smaller than 0.05 to guarantee a confidence interval of at least 0.95. This value was used to guarantee that the effect of the independent variable on the regression and the regression as a whole are off in less than 1 out of 20 cases. This standard is accepted in literature as an acceptable statistical value to test a hypothesis (Bland, 1995) and thus validity of an independent variable and the regression.

The coefficients of the independent variables found by Windows Excel represent the relative strength of the independent variable towards the tested dependent variable. If for instance coefficient α is a factor 100 smaller or larger than coefficient β, this could indicate that Windows Excel is trying to minimize the effect of α. The coefficients found by Windows Excel should be in the same order of magnitude as one another. Should one of the coefficients of a variable be a factor 100 or more smaller, the variable related to coefficient α is rejected and the regression is done again.

R^2 value

If both previous tests are satisfied, the R^2 value related to the regression as provided by Windows Excel was observed to determine the explained variance of the dependent variable (Lacey, 2017). The explained variance R is evaluated by using the Pearson coefficient to determine the strength of the correlation (Pearson, 1895). In appendix V, multivariate linear regression is explained further.

Data normality test
Before any of the regressions in this research could be validated, the variables used in the regression were tested on distribution. For a multivariate linear regression to be valid, both dependent and independent variables need to be distributed normally (Cambridge, 2017). The data from the variables are tested on skewness and kurtosis to determine if the variables are normally distributed. Both the skewness and kurtosis are measures to determine the shape of the probability distribution and if the distribution is a normal distribution. According to Pearson, a skewness between -2 and 2 is (Pearson, 1905) and a kurtosis between -2 and 2 are acceptable values for a normal distribution (Pearson, 1904). The skewness of a function is defined as "*a measure of the symmetry of a distribution around its mean*" while the kurtosis is defined as "*the state or quality of flatness or peakedness of the curve describing a frequency distribution in the region about its mode*" (William Collins Sons & Co. Ltd., 2012). The test for skewness and kurtosis were done using the function "data analysis – data information" in Windows Excel.

If the value of the skewness or the kurtosis are outside the set acceptance range, outliers as in extreme values are removed from the dataset conform common practice (Anscombe, 1960). In this research, extreme values are defined as values of the data set outside the range of the mean plus two standard deviations. The range of the mean plus two standard deviations is chosen in accordance with the empirical 68-95-99.7 rule, to define the interval in which 0.95 of all the values in the data set can be found (Pukelsheim, 1994). The removal of the outliers is done one extreme value at a time, to ensure that as much of the data set as possible is preserved. The extreme value removed is the value the furthest from the mean outside the 0.95 interval. The dataset is normalized again, before the kurtosis and skewness are determined. If the kurtosis and skewness were not in range after the removal of an extreme value, another extreme value was removed. If after the removal of all the extreme values the skewness and kurtosis were still outside the range of acceptance, the data was rejected from the research on the basis that the data was not normally distributed. To visualise the dataset, the data was normalised to provide a probability density function of the values within the dataset. The normalisation of the data was done using the following formula:

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ (eq. 1)$$

With:

x	=	Value from dataset
σ	=	the standard deviation of the data set
μ	=	the mean of the data set

Part I: Wheat

2. Continuation introduction wheat

2.1 Research Objective

Summarizing the introduction for wheat, the objective for this research was to link weather patterns to price changes of the agricultural commodity of wheat. First, an understanding of the dynamics and drivers of the prices of wheat is formed. Second, the relative importance of the wheat production of the USA and Canada towards the world wheat production is evaluated. Next, the hypothesis of the importance of rainfall in a specific time period towards the production and price change of wheat is tested both quantitatively and qualitatively. The hypothesis is tested for spring wheat production performance in North Dakota and the Canadian Prairie. The goal is to keep the research, the method and results fully accessible for both people with an engineering or financial background and provide a method to understand and predict wheat production subsequently assess the impact towards wheat prices.

2.2 Research Question

Following the set objective, the following research question concerning price change of wheat can be stipulated:

- *What are the main drivers behind the prices of wheat?*

For price change, the following sub questions can be stipulated:

- *In what years do the largest relative increases in price occur, taking inflation of the crop specific currency into account?*
- *Is it possible to determine the causes of the price increases and cluster these effects, such as shortage in supply due to natural effects?*
- *What is the relative strength of the different drivers on market prices?*

The following research questions can be stipulated concerning production:

- *How strong is the dependency of the world wheat production on spring wheat production in the USA and Canada?*

For production, the following sub questions can be stipulated:

- *What is the yearly world production of wheat?*
- *What is the yearly production of wheat of the USA and Canada?*
- *What part of the variance in the world production of wheat can be explained by the wheat production of the USA and Canada?*

For the proof of concept of weather dependency, the following research question can be stated

- *Is it possible to prove the importance of the weather in a specific time period in North Dakota and Canada towards the production performance of spring wheat?*

For the proof of concept, the following sub questions can be stipulated:

- *What are the crop's weather related specific requirements of spring wheat to successfully produce spring wheat?*
- *From reports, what are the indicated most important time periods for the production performance of spring wheat?*
- *What part of the variance of the local production performance of spring wheat can be explained by using the found important time periods?*
- *What part of the variance of the price of wheat can be explained by using the found important time periods?*

2.3 Geographical setting

The USA and Canada are the 2 largest wheat producing and exporting countries in the world as of 2016. The USA is responsible for 8.5% of the world wheat production and 14.8% of the world wheat export, Canada is a close second with 4.3% of the world wheat production and 12.4% of the world wheat export (CIA, 2017). For the USA, North Dakota is taken as area to test the hypothesis of seasonal key points for spring wheat production. For the wheat production of Canada, the spring wheat production of the Canadian Prairie is taken as area to test the hypothesis. The reason for choosing spring wheat is explained further on in this chapter.

North Dakota

In the USA, North Dakota (ND) is one of the largest wheat producing states with 17% in 2014 (USDA, 2016). In this research, only spring wheat production in ND is assessed for the hypothesis test of seasonal key points, because of the fact that the spring wheat production is covering 86% of the total wheat production in ND as of 2015 (NDWC, 2017). In Figure 1, the concentration of US spring wheat production is shown, were a strong concentration of production can be seen in ND. Next to being one of the largest wheat producing states in the USA, ND also has data available concerning rain and crop yields via the North Dakota State University Hettinger Research Extension Centre (NDSU HREC). An introduction to the NDSU HREC can be found in Appendix I.

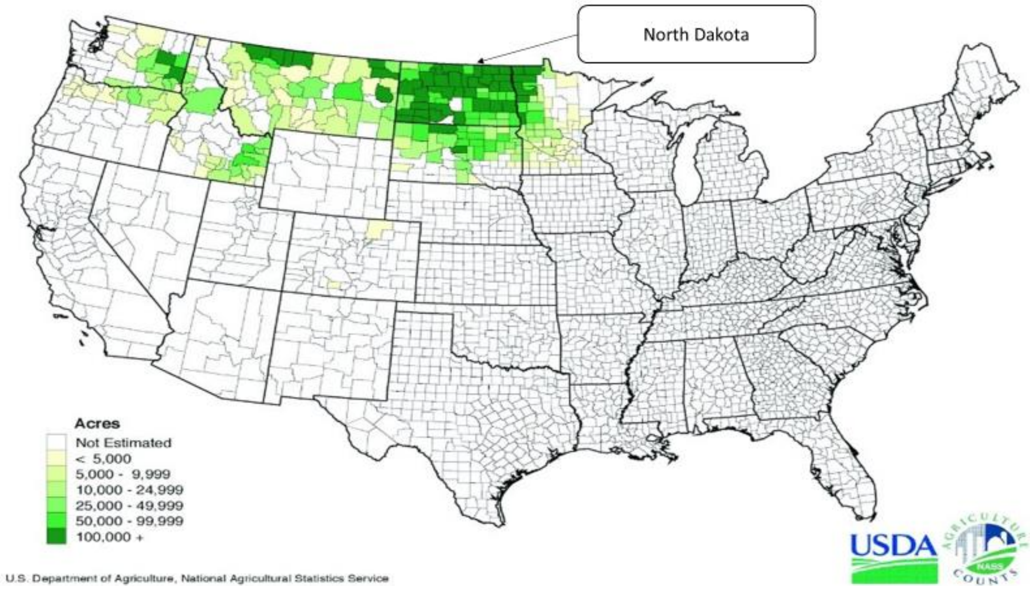


Figure 1: concentration of US spring wheat production with North Dakota (USDA, 2016)

North Dakota is situated in the North of the US on the USA Canada border. Situated in the Upper Midwest, ND has a humid continental climate with cold winters and hot summers. ND has a small part on the western border where the climate tends more towards a semi-arid climate before translating towards the continental climate. In general, ND has a period of warm and cold weather, with warmer temperatures arriving around April before the cold sets in around October. Snowfall is recorded from November up to March, being the main source of precipitation for those months. The wettest moths are May, June and July, and October is the driest month (NWSEND, 2002).

Canadian Prairie
The Prairie of Canada stretch over three provinces in the south of Canada, being Manitoba, Saskatchewan and Alberta, as can be seen in Figure 2. Together, the plains produce almost 91% of the total wheat production of Canada, of which spring wheat is the main contributor with 87% (Statistics Canada, 2016). A large amount of data is available from the Canadian Prairie both in meteorological and agricultural sense of for instance the ministry of Alberta. An introduction to the ministry of Alberta can be found in Appendix I.

Figure 2: Dispersion of wheat producing areas of Canada (CGC, 2017)

Concerning the climate of the Canadian prairies, the following quote from the book "The weather of the Canadian Prairies" can be found: "*The climatic regimes of the Prairie province are classified as either cold-temperate or sub-arctic and range from dry continental type conditions, in the southwest, to sub-arctic conditions in the northeast along the Hudson Bay coastline. The western mountain ranges have a pronounced effect on the precipitation patterns across the region and on winter temperatures. The summers are fairly short, warm and most of the time dry and stretch from May to August, while the winters stretch from November till late March, with large amounts of snow and general cold weather*" (Vickers, Buzza, Schmidt, & Mullock, 2000).

Relation climate to the production of spring wheat

Spring wheat is a type of wheat sown in the spring. The length of the growing season of spring wheat ranges from 100 to 170 days. The difference between spring wheat and winter wheat is that winter wheat is sown in the autumn and goes into dormancy throughout the winter. In areas with a severe winter, the production of spring wheat is preferred due to the destructive effect of the winter on the plants during winter dormancy (FAO, 2012). In ND, hard red spring wheat is cultivated, while in the Canadian Prairies, Canadian spring wheat is cultivated.

Translating the area specifics to the production of spring wheat in North Dakota and the Canadian Prairie, Peltonen-Sainio quotes the following concerning growing condition in northern climates: "*The most typical features of northern growing conditions are harsh winters, intensive, exceptionally rapid rate of development due to exposure of very early growth stages to long days and relatively rapidly increasing mean temperatures, generally cool mean temperatures during the growing season, risk of night frosts, early summer drought and risk of abundant precipitation close to harvest*". Peltonen-Saino further states the importance of the snowfall in the winter for growing wheat in northern countries. Snowfall provides a starting point in terms of moisture for the newly planted crops, but the melting snow also provides wet fields which need to dry before the wheat can be sown (Peltonen-Sainio, 2009).

Both North Dakota and Canada have harsh winters and small periods of crop growth. Due to the harsh winters in both Canada and North Dakota, spring wheat is planted between the second week of April towards the end of May and harvested mind August through October in North Dakota (NDSU, 2017) and in the month of May and harvested in October in Canada (Canadian Goverment, 2017).

3. Materials and Methodology

In this chapter, the material and methodology used in the research are discussed. In the materials section, all the used data for the research is discussed. Next, the methods of processing the data to achieve the results are discussed. In the method chapter, a general overview of the whole process of translating the data into results is given.

First, the price changes of wheat were assessed. For this research, agricultural and price reports from the FAO were used. For wheat, these were the yearly “The State of Food and Agriculture” (SOFA). From reports of the FAO, possible seasonal key points were defined qualitatively based on price changes, as well as the general dynamics of the wheat market. The used reports are introduced in section of materials, the method of using the reports is introduced in methods.

The impact of the rainfall is further researched in reports from the FAO, USDA, NDSU HREC and Ministry of Alberta. The goal is to qualitatively define seasonal key points as found from the reports on the wheat production performance from an agricultural perspective.

The wheat production of the USA and Canada were used to investigate how the countries influence the world wheat production. To determine the effect of weather on production, the crop performance is used. The materials and method of the wheat production assessment are introduced further on in this chapter.

Due to the high contribution of the USA and Canada to the world wheat production and export, the country of the USA and Canada were chosen to research the impact of local rainfall patterns on the crop performance of spring wheat. The rainfall data was collected from 8 stations in North Dakota 14 stations in within the Canadian Prairie, which are further discussed in the section of materials. The source and method of the spring wheat crop performance are introduced further on in this chapter.

Next, a multivariate regression analysis is performed. The goal is to assess rainfall data from the USA and Canada and research if seasonal key points found from the reports could be quantified concerning the contribution to the crop production. As part of the objective, the level of explained variance of the production of wheat as well as probabilistic values of the key points in a season were assessed to verify if the influence of the seasonal key points on the wheat production performance is not random. The method is elaborated on further on in this chapter.

As a final step, a binary Bayesian probabilistic model (BBPM) was created using the defined seasonal key points and performance data. Using the BBPM, the probabilities and the effect of the occurrence of the seasonal key points towards relative change in world wheat production could be defined. The full process is described further on in this chapter.

3.1 Materials

In this chapter, all the materials used for the research are discussed. All the details of the materials are discussed, such as length of used data series, limitations due to data availability and whether data is of a daily, monthly or yearly scale.

3.1.1 Market data wheat

The market data used to determine the price changes and the extreme increases and decreases the daily future prices from the National Agricultural Service of the USDA over a period of 56 years between 1960 to 2015 (USDA, 2016). The prices were provided in the form of future prices in US Dollars, meaning that the price were the quoted prices of future contracts. For wheat, future contracts are available for January, March, May, July and November (ICE, 2017). A future contract is a contract with an obligation to buy a commodity such as wheat in the future for a set price. The prices for the futures were provided on a monthly average scale. For this research, inflation is filtered out to produce the “real” or “clean” price changes (Malliaris, 2006), which will be further discussed in chapter 3.2.2. The future prices of wheat were in US dollars and therefore the inflation of the US dollar was taken to correct the future prices for inflation. For inflation, the inflation numbers of the US dollar from the Bureau of Labour and Statistics of the Department of Labour were used between 1914 to 2016 (BLS, 2016). The data concerning inflation was provided in monthly averages and in yearly averages.

3.1.2 Wheat reports

From the FAO, the annual SOFA reports concerning wheat production, prices and consumption were used. In the yearly reports, a complete overview of the wheat world price change, as well as production and consumption figures are provided. The annual SOFA reports used are from 1960 to 2013 and can be obtained via the website of the FAO (FAO, 2016). In appendix II, an example of a SOFA report is given.

Next to the SOFA reports, the FAO also provided crop specific information concerning the specifics of cultivating wheat (FAO, 2012). Next to reports from the FAO reports and articles from the NDSU HREC and Ministry of Alberta were used for crop specifics. For the NDSU HREC, the reports used were the “Western Dakota Crop Day” (WDCD) reports between 1984 and 2015. In the WDCD reports, the crop performance of wheat at two research locations in ND are discussed, with a particular focus on the weather, diseases and planting dates. As a proxy for the Canadian prairies, the reports from the ministry of Alberta were used, due to the central position of the province in the Canadian Prairies. From the ministry of Alberta, the annual crop season review reports between 2006 and 2015 were used. In the reports, the wheat growing conditions are discussed concerning the impact of the weather, insects and crop diseases. All reports can be found on the website of the ministry of Alberta (Ministry of Alberta, 2017).

3.1.3 Wheat production and yield per area

The crop production and performance data of wheat was retrieved from two sources, being the USDA and the North Dakota Wheat Commission (NDWC). The total wheat production data of the USA, Canada and the World was retrieved from the USDA Foreign Agricultural Services, department for Production Supply and Distribution (PS&D) website between 1960 and 2015. On the website, a custom query can be made concerning parameters such as production, distribution and export for a variety of crops (USDA, 2017). The data from the PS&D concerning the production of wheat was provided in thousands of megatons per crop year, where a crop year for wheat is defined between the start of June to the end of May the following year (OECD-FAO, 2007).

Next, the crop yield data was collected. Wheat production is dependent of a number of factors, of which weather is one. Another factor is the amount of area sown to produce wheat. The use of only production can give a wrong impression about the performance of a crop year concerning the production of wheat. For instance, two years can have the exact same amount of wheat production, but with for instance one area having twice as much area sown as the other area. The first area could then be categorized as either normal or underperforming, while the second area is normal of overperforming.

To filter out the effect of the sown areas, the performance of an area in the amount, or yield, of wheat produced per area is used for this research. For the yield data for Canada, the PS&D website was used as well. For the Canadian Prairies, the simplification of using the crop yield data of Canada as a whole from the PS&D was used for this research between 1960 and 2016. The wheat performance of the whole of Canada is a mix between spring wheat and winter. However, as a first approximation the simplification of taking the total yield for Canada was accepted due to the large contribution of spring wheat to the total wheat production in Canada. The unit of the yield data was provided in megaton per hectare. For ND, data concerning the spring wheat yield from the NDWC was used as part of the research (NDWC, 2016) between 1971 and 2015. The unit provided by the NDWC was in bushels per acre.

3.1.4 Weather station data

For the weather stations of ND and the Canadian Prairies, both snowfall and rainfall data was collected. Both snowfall and rainfall were found to be of importance for the cultivation of wheat when assessing northern countries (Peltonen-Sainio, 2009).

North Dakota

The weather stations of ND were collected from the North Dakota Agricultural Weather Network (NDAWN) website. The NDAWN is a network of weather stations in and around ND and is provided by the NDSU (NDAWN, 2017).In Figure 3, the selected weather stations for ND are presented. The selected weather stations are concentrated towards the south-western part of ND, because of the fact that the NDSU HREC is located in the south-western part of ND. Because of the use of reports of the HREC concerning the area and wheat production around the HREC, the stations were selected in close proximity to wheat testing site. The second reason for selecting the specified weather stations, is because these rain stations are the oldest weather stations in ND and therefore provide the largest amount of data.

In total, 8 weather stations were selected for rainfall data, being Bottineau 14W, Turtle Lake 4N, Hazen 2W, Dickinson 1NW, Jamestown 10W, Beach 9S, Bowman 4W and Hettinger NW[2]. Data was available from 1990 to 2015 for the stations Hettinger, Dickinson and Jamestown, for the other stations data was available between 1993 and 2015. Al data was available for the months of April to October. No missing data in the found in the dataset. Between November and March, the weather stations were not active due to the frost and snowfall in ND. Concerning snowfall, no direct measurements from stations was available. Therefore, overviews of the North Dakota State Climate Office (NDSCO) were used to visually determine the total snow depth between November and March around the stations of Jamestown, Bottineau and Dickinson. An example of the overview is presented in Appendix III.

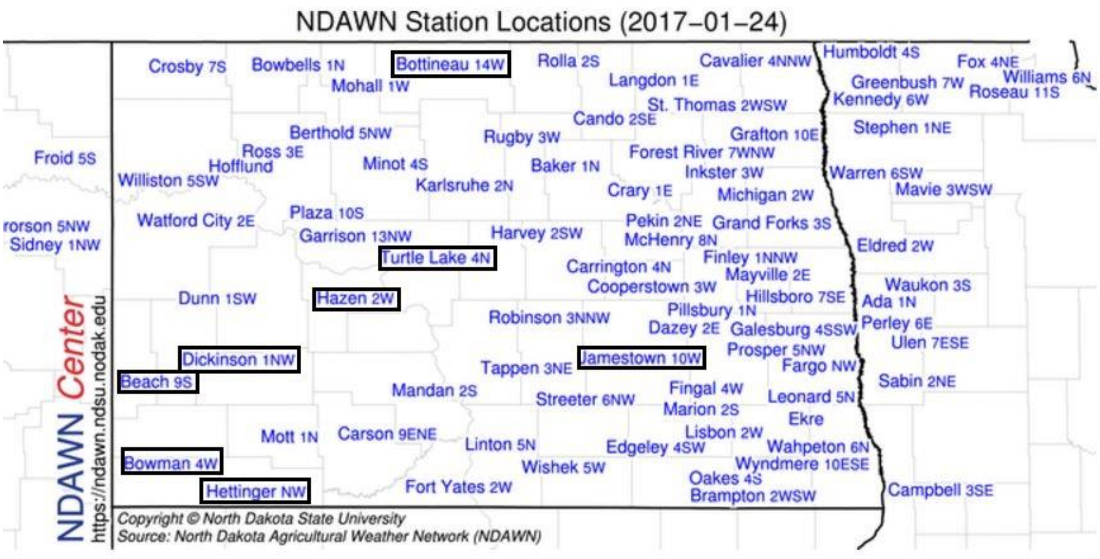


Figure 3: selected weather stations North Dakota (NDAWN, 2017)

[2] The name of the weather station is a combination between the name of the town closest to the weather station, and the distance and direction from the town. For instance, Hazen 2W means that the weather station is 2 miles West of Hazen.

Canadian Prairie
In Figure 4, the selected weather stations for the Canadian Prairie are presented. The data from the weather stations were obtained via the Koninklijk Nederlands Meteorologisch Instituut (KNMI) Climate Explorer web application. The KNMI Climate Explorer application (KNMI, 2017) is a collection of data from weather stations all over the world. Daily data from the KNMI was used as opposed to direct monthly data, due to the fact that no monthly data was available for Canada. As can be seen in Figure 4, 14 stations with daily data were selected for Canada. These stations were Medicine Head, Swift Current, Lethbridge, Estevan, Regina, Red Deer, Camrose, Saskatoon, Brandon, Altona, Climax, Calgary, Edmonton and Lloydminster. The reason for choosing the stations was to provide an even spread of stations over the whole Canadian Prairies. Data was available between 1883 and 2017, but not continuous for all stations. For instance, data for the weather station of Medicine Head was available between 1883 and 2007, while for Swift Current data was available between 1959 and 2017. Both the snow and precipitation data of Canada was collected from the KNMI Climate Explorer application. Rainfall data was provided in millimetres, while the snowfall data was provided in millimetres depth of snow.

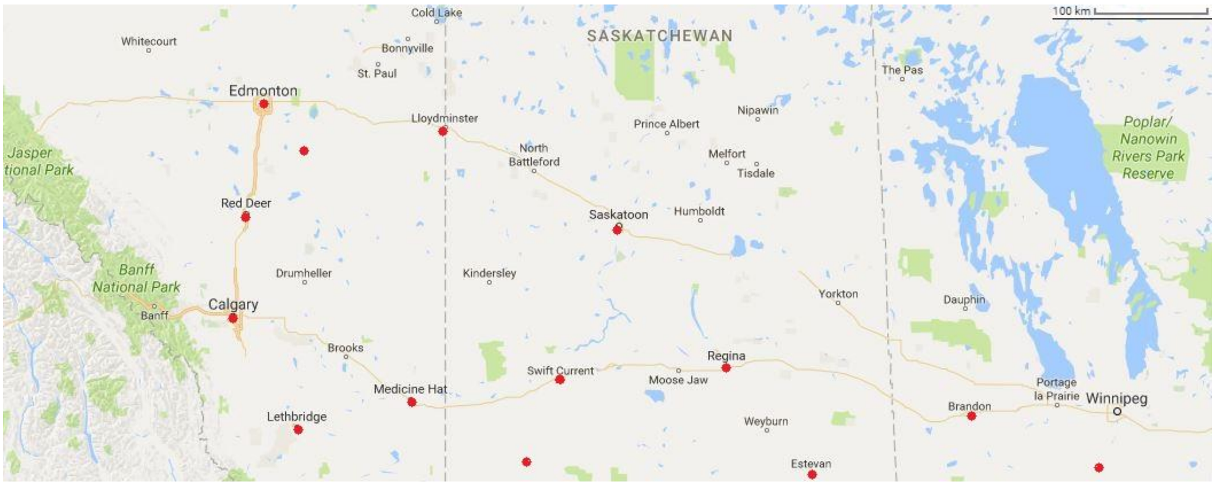


Figure 4: Selected weather stations in the Canadian Prairies, indicated with red dots (Google Maps, 2017)

3.2 Methods

In this chapter, the specifics concerning the used methods for the data processing are discussed. Theoretical background and the specifics of how a number of techniques are used in this research are also described. In Figure 5, the flowchart of the research is presented.

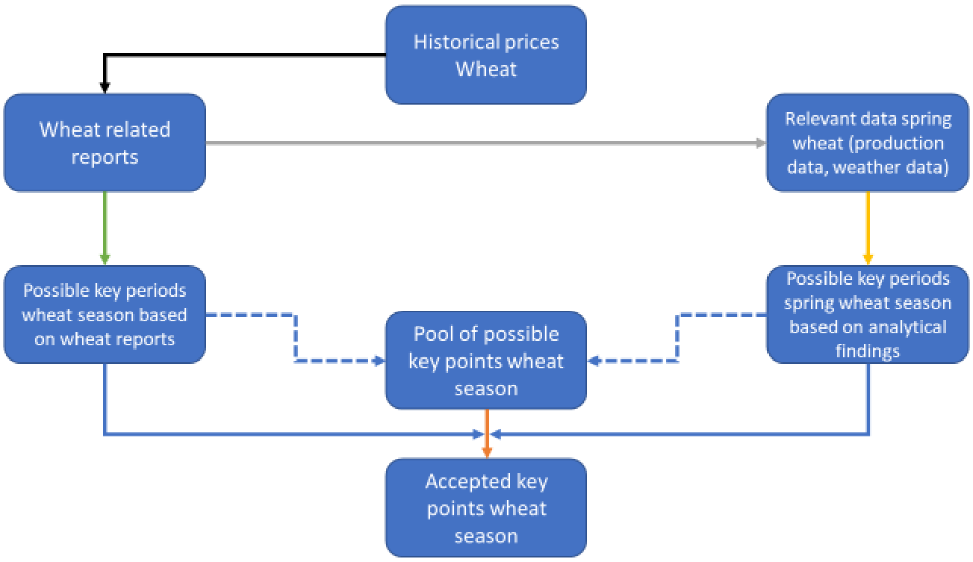


Figure 5: flowchart of the course of research for wheat, explanation given in text

To start of the research, the historical prices of wheat were assessed. In this step, the goal is to find the years in which the wheat price had the largest percental relative increase or decrease between 1960 and 2015. Also, monthly price changes were assessed to define the average price movement of wheat on a monthly scale within one crop year by taking the average monthly price change of all the crop years. All the prices were corrected to take inflation into account. This step is represented by the point "historical prices wheat" in Figure 5. The results from this step was used as a starting point for the next part of the research. The full process is described in 3.2.1.

Next, the found price increases and decreases were assessed using wheat related parts in reports from the FAO. The results from the previous step concerning the historical prices of wheat were used as a first indicator to determine in which years and months reports should be assessed in particular. The goal is to use the information of the largest price increases and decreases in combination with reports from the years and months in which the price changes occurred to identify the reason for the price change. This step is represented in Figure 5 by the solid black arrow between "Historical prices wheat" and "Wheat related reports".

In the same step of "Wheat related reports", crop specifics as defined by the FAO, USDA, NDSU HREC and Ministry of Alberta as well as climate specifics of ND and the Canadian Prairies were used to define key points in a spring wheat growing season. Crop specifics such as water demands, growing cycle and sowing and harvest times were researched. Area specifics for ND and the Canadian Prairies such as climate, seasonal changes in weather and temperatures were assessed at the same time. The combination of wheat related reports from the FAO, USDA, NDSU HREC and the Ministry of Alberta, as well as crop and climate specifics of ND and the Canadian Prairies were used to qualitatively define seasonal key points in the growing season.

The reports and historical prices were used to define the dynamics of the wheat market. This step is represented in Figure 5 as the solid green arrow between “wheat related reports” and “Possible key periods wheat season based on wheat reports”. The full process is described in 3.2.2.

The outcome of the wheat related reports was used as an input the evaluation of production and weather data. This step is represented by the grey arrow from “wheat related reports” to “relevant data spring wheat” in Figure 5. The goal was to use the information about the largest price changes in combination with the qualitative outcome of the reports to determine if the price changes can also be explained using spring wheat production data and rainfall data from ND and the Canadian Prairies. For the production data, the change in yield per area was used. Due to the fact that wheat yield is always increasing (FAO, 2012), the difference in yield as difference from the trend was taken. The process of using relative difference of spring wheat yield from the trend is described in 3.2.3.

Within the “relevant data spring wheat”, local production of the USA and Canada and world production was also assessed to determine the dependency of the world wheat production on the local production. A multivariate linear regression between world wheat production and wheat production from the USA and Canada was performed to determine whether a dependency between world wheat production and local wheat production as mentioned in the ICCO reports can be confirmed. The full process is described in 3.2.4.

In the “relevant data wheat” step, data from weather stations was also pre-processed for further use. First, the used weather station data was filtered for missing data. Second, the daily rainfall data per station is summed up to give a monthly and a yearly total. The total for a year was the total per wheat year as defined by the OECD-FAO between July and June. Third, the used weather stations are assessed on correlation to determine how strongly correlated the stations were to justify the use of an arithmetic mean of the stations. If a weather station was insufficiently correlated, the weather station would be rejected from the research until all weather stations left were sufficiently correlated to justify the use of an arithmetic mean. The correlation between weather stations was tested using the average rainfall on a monthly basis for a full year per weather station. Finally, the arithmetic mean of the weather stations was taken of the average monthly and yearly rainfall. The full process is described in 3.2.5.

To define key points in a spring wheat season in a quantitative sense, a linear regression between the relative change in spring wheat yield and the arithmetic mean of the weather stations was performed for both ND and the Canadian Prairies. This step is represented by the solid yellow arrow between “Relevant data spring” and “Possible key periods spring wheat season based on analytical findings” in Figure 5. To narrow down the variables used in the linear regression, the qualitative results from the spring wheat related reports were also used. This step is the results from the feedback from the reports as represented by the grey arrow in Figure 5. The outcome of the linear regression and the used variables are tested on probabilistic properties to determine if the finding were significant. The full process is described 3.2.6.

Next, both results from the reports and the analytical findings were combined in a pool of possible key points of a spring wheat growing season. This step is represented by the dotted blue arrow in Figure 5. Subsequently, to determine if the found key points were actual real key points in a spring wheat growing season, the possible key points were tested on both a qualitative basis by the indicated key points from the SOFA reports from the FAO and on a quantitative basis using the results from the linear regression. This step is represented by the solid orange arrow in Figure 5, while the qualitative and quantitative test is represented by the solid blue arrow in Figure 5. The full process is also described in 3.2.6.

As a final step, a binary probabilistic Bayesian model was applied to the confirmed seasonal key points to assess the conditional probability of a higher or lower wheat production by using the seasonal key points. In this step, the probability of a above or below the trend wheat production was made conditional based on the combination of a above or below average rainfall in one of the seasonal key points or a combination of seasonal key points. The full process is described in 3.2.7.

3.2.1 Historical prices of wheat

For the historical prices of wheat, future contract prices of wheat were collected as defined in 3.1.1. In appendix IV, a general introduction to prices and commodities is given as background information for the terms of commodities and future contracts. The future prices of wheat were quoted in dollars, but the future prices cannot be used directly in this research. Because of inflation, a US dollar in 1960 does not have the same value as a dollar in 2015. Inflation or deflation are defined as a general, continuous increase of decrease in prices, causing a reduction or an increase in the value of money (McIntosh, 2013). To make the price change of for instance 1960 usable in the same manner as in 2015, the future prices were first translated into relative change. In this research, wheat price changes were defined as relative price changes. To translate the crop price data to relative price change of wheat, the following formula was used:

$$RPC = {(X_n - X_{n-1})}/{X_{n-1}} * 100 \quad (eq. 2)$$

With:

RPC	=	Relative price change in %
X_n	=	the future price of month n in US dollars
X_{n-1}	=	the future price of month n-1 in US dollars

Relative price change was determined on both a monthly and yearly scale. For wheat, yearly relative price change is determined from June to May, to coincide with the wheat production year as defined by the OECD-FAO. The practice of taking the average of the year is in line with the practice of the average price determination of the USDA (USDA, 2016). By using the price change in a crop year, the effect of the crop production on the prices of the crop year can be obtained. For 2000 for instance, equation 2 takes the following shape:

$$RPC\ (2000/01) = {\Delta\left(\frac{2000}{01}\right)}/{X_{2000}} * 100\ (eq. 3)$$

With:

$RPC\ (2000/01)$	=	Relative price change for wheat year 2000/01 in %
$\Delta(2000/01)$	=	Difference between future price July 2000 and June 2001 in US dollars
X_{2000}	=	Future price of wheat in June 2000 in US dollars

By changing the prices to relative change, the prices were corrected with inflation, which is also a relative change. By subtracting the inflation, the goal was to define the “clean” price change in a month or year. The correction of the relative change of the prices of wheat by using inflation is an accepted method of correcting prices (Malliaris, 2006).

Inflation was therefore deducted from the relative price change to produce the "clean" value for the relative price change. If wheat prices increase with 4% in a year and inflation in the same year also increases with 4%, the "clean" relative price change is equal to 0%. Inflation is already given in a relative term of percentage on a yearly and monthly basis provided by the BLS (BLS, 2016), and could be directly deducted from the relative price change as found in equation 2. Inflation on a yearly scale was used according to the specific length of the crop year for wheat. The used formula is shown below:

$$RPC^{*} = RPC - I \; (eq. 4)$$

With:

RPC^{*}	=	the clean value of the relative price increase in %
RPC	=	the relative price change in % as defined in equation 2
I	=	the inflation for the assessed month/year in %

As a final step, the price increases and decreases were linked to their respective years and ranked from high to low.

3.2.2 Wheat related reports: price changes

In this section, the use of reports from the FAO, USDA, NDSU HREC and the ministry of Alberta is described. Reports from the FAO, USDA, NDSU HREC and the ministry of Alberta were used to define crop specifics key points for spring wheat. SOFA reports from the FAO are also used to determine if weather had an effect on the found price changes and if key points in a season related to weather can subsequently be defined using the reports.

3.2.2.1 Crop specific key points

To determine the crop specific needs for the growth of spring wheat, a review is conducted on available literature about the crop specifics for spring wheat. For spring wheat, the website of the FAO, NDSU HREC, USDA and ministry of Alberta were used to obtain specifics concerning the demand for growing spring wheat. The specifics provided information about the maximum and minimum amount of precipitation required, as well as length of the growing season and events during a growing season that could affect crop yield for spring wheat (FAO, 2012). The key periods of the season as defined by the crop specifics were later used to define the seasonal key points for the crop yield of wheat.

3.2.2.2 Wheat price change related reports

After the price changes for wheat are defined, a review of the SOFA reports of the FAO was performed to determine whether the weather had any influence on the relative price change of wheat. For this research, all price changes between 1960 and 2015 were assessed, with the top 5 price decreases and increases highlighted.

Production of wheat can be affected by different events, such as lower crop yields due to weather events, but also trade embargos, export stops and civil unrest. For all of the price changes, the reports that were available for wheat for that year were assessed to define the reason, or "driver", related to the price change. The SOFA reports were scanned on a specific link back to weather using words as "weather", "adverse", "negative effect", "unfavourable" specifically correlating weather to production decrease and price increase. For wheat, SOFA reports were used for weather related price changes.

3.2.3 Relevant data spring: production changes

As mentioned in the introduction, wheat yield is constantly increasing. Reasons for the constant yield increase are for instance more planting, but also crop improvement or better agricultural practices (Evans, 1980). As a consequence, wheat yield in different years could not be compared directly to each other. A specific amount of wheat production in for instance 1980 could be a very high performing crop year then, while the same production in 2000 would be medium or badly performing crop year.

A linear trendline of the spring wheat yield of ND and the Canadian Prairies was therefore used as a first approximation in this research. To determine whether a wheat year is underperforming, the difference from the linear trendline is used to determine how a wheat year was performing. As this research was a first trial, no other method than a linear trendline was used for the yield. The available data for the yield of ND and the Canadian Prairies was plotted using Windows Excel, after which a trendline is added to the plot.

For the linear regression and trendline, the dispersion of the residuals was used to visually determine the validity of the use of a linear regression (Stattrek, 2017). The residuals should appear to be dispersed randomly without a clear parabolic shape to enforce the use of a linear approximation. Crop yield under the trend will be defined as below expectation, while crop production above the trend will be defined as above expectation. The resulting outcome of the difference between the trend line is used in further steps of the research. For all steps, Windows Excel was used.

3.2.4 Relevant data spring wheat: influence local production wheat on world production wheat

To correlate the world wheat production to local wheat production, a multivariate linear regression of world wheat production as a function of wheat production in the USA and Canada was performed. The USA and Canada were here defined as local producing areas for wheat and independent variables, while the world wheat production was used as dependent variable. The regression for wheat was done using production data from 1960 to 2015 using the PS&D tool of the USDA as defined in the materials section. The outcome of the multivariate linear regression was tested on P-values, F-values and residuals to validate the linear regression. Finally, the R^2 value was used to define the explained variance of the regression and to determine the dependency of the world wheat production on the wheat production of the USA and the Canadian Prairies. For all steps, Windows Excel was used.

3.2.5 Relevant data spring wheat: pre-processing weather station data

To translate the weather stations data from ND and the Canadian Prairies into usable data for this research, a number of processing steps were taken. For this research, the weather stations were first checked for missing data. Missing data was given a zero as value concerning precipitation. If more than 5 days of a month were missing, the month was not used in further steps. For ND, no missing data was found. In the Canadian Prairie dataset, only the rainfall and snowfall data of the years between 1960 and 2015 were used, due to the fact that prices of wheat were available between 1960 and 2015. Between 1960 and 2015, no missing data was found for the Canadian Prairies. Next, rainfall per month and per wheat year were calculated by summing up all the recorded rainfall per month or wheat year.

Finally, the average rainfall in a specific month was determined by calculating the average rainfall of all specific months for a specific weather station. All calculations were done using Windows Excel. For instance, the average rainfall of the month of April for the Swift Current station was calculated by taking the average of all the recorded rainfall of all the months of April for the Swift Current station.

The correlation between stations was tested next to check if the average rainfall of weather stations in ND and the Canadian Prairies were correlated enough to justify the simplification of taking the arithmetic mean of all the weather stations. As a demand for the correlation index, correlation between all stations was required to be above 65% to ensure a “good” correlation conform the Pearson correlation coefficient (Pearson, 1895) to ensure that the rainfall in the assessed area was relatively equal at all assessed points and that therefore the area could be simplified to one area with an equally dispersed amount of rain. If the demand of 65% or above was not met, weather stations with the worst correlation to all other stations were rejected until all stations had a correlation above 65%.

3.2.6 Possible key periods spring wheat season based on analytical findings: linear regression

To quantify the crop specific seasonal key points from the reports from the FAO, USDA, NDSU HREC and ministry of Alberta, a regression of the difference from the trend for spring wheat yield as a function of the rainfall of all the months that have an effect on the wheat year was done. For ND and the Canadian Prairies, the arithmetic mean of the rainfall data per month of a spring wheat year were used as independent variables. From the regression, the month with the highest P-values was rejected and a new regression was done until all P-values of the used months were within a 0.95 confidence interval and the F value of the regression in general is also within a 0.95 confidence interval. As mentioned in 1.1, a confidence interval of 0.95 excludes randomness of the variables. Reports were also used to make the optimization process go faster by using months that are found to be of importance based on crop specific FAO, USDA, NDSU HREC and ministry of Alberta reports and climate and area specifics. The iterative process is repeated until a combination of variables is found that explain the highest amount of variance of the change in crop production while the significance of both the independent and dependent variables was preserved. However, before the results are accepted, the results are first checked with SOFA reports from the FAO to exclude spurious results (Burns, 2017)

The found seasonal key points are checked with the reports from the FAO, USDA, NDSU HREC and ministry of Alberta to check if the found time periods are both statistically accepted and in accordance with the reports as provided by the FAO, USDA, NDSU HREC and ministry of Alberta. The key of this step is the combination between reports and analytics. The goal is to ensure that the linear regression performed between difference from the trendline as a function of a key time period in the spring wheat growing season is both statistically sound while also being able to be explained by reports and crop specifics. By combining both qualitative and quantitative reasoning, the method should exclude spurious results of regression and only focus on variables that are related to the growing of spring wheat.

3.2.7 Binary Bayesian probabilistic model

By using both the data relative change in production from 3.2.3 and the results from the specified seasonal key points from 3.2.6, a binary Bayesian probabilistic model could be formed. The BBPM is a continuation of the results of the seasonal key point regression outcome of 3.2.6, connecting a probability to the seasonal key points of the regression. For the BBPM, the relative change in spring wheat yield was used in combination with the rainfall and snowfall data from ND and the Canadian Prairies of the seasonal key points. Both yield and seasonal key points were first translated into binary. For the spring wheat yield, yield above the trendline was defined as a 1, while the yield below the trend was defined as a 0. For the rainfall in the seasonal key points, rainfall above the average monthly arithmetic mean in a specific month was defined as a 1, rainfall below the arithmetic mean rainfall in a specific month was defined as a 0.

Because of the use of the seasonal key points as independent variables and the yield change as dependent variable in the linear regression, the probability of yield change was defined as conditional and dependent on the seasonal key points (Bernardo, 2000). To define the probability of an above or below average spring wheat yield, all the combinations of the independent variables were tested towards spring wheat yield. The probability and conditional probability were determined by using the historical occurrence of the events. For instance, if 5 of the 20 years assessed have an above trend spring wheat yield change, an unconditional probability of 0.25 was assigned to the probability that the relative spring wheat yield change will be above trend and an 0.75 change that the relative spring wheat yield change will be below trend.

4. Results

In this chapter, the results of the research are presented. First the results historical prices of wheat are presented. Both the yearly price changes as well as the average price change within one crop year are shown. Second, the price changes on both a monthly and yearly scale are explained using the information from the SOFA reports of the FAO, as well as the found seasonal key points from the reports and the crop specifics as found from the FAO, USDA, NDSU HREC and ministry of Alberta. Third, production is assessed, with an assessment of the trend of the spring wheat yield in ND and the Canadian Prairies as well as the dependency of the world wheat production on the wheat production of ND and the Canadian Prairies. Fourth, the results from the processed rainfall data of the used weather stations as well as the correlation index of the used weather stations is described. Finally, the outcome of the key point regression and subsequent BBPM are presented.

4.1 Historical prices of wheat

First, the average monthly price change of wheat as found between 1960 and 2015 is presented per crop year. The monthly price changes can be used as first indicators to determine the important months for wheat. Months with high price change could indicate important events concerning wheat and could therefore be important to assess. The average monthly price changes of wheat, corrected for inflation and in percentage are presented in Figure 6. As can be seen from Figure 6, price changes tend to increase in August, September, October and Nov, with the largest increase in August. The largest price decrease was noted in June, followed by July and May. In the review of the SOFA reports of the FAO in chapter 4.2, possible reasons for price changes are discussed.

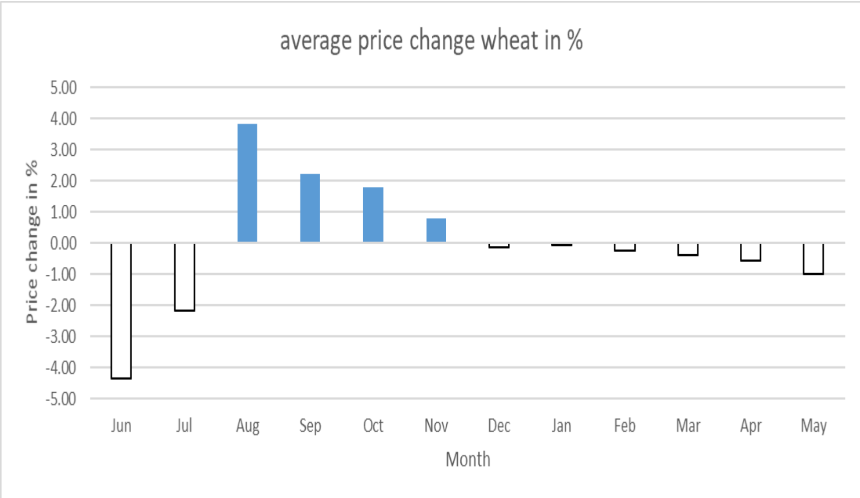


Figure 6: the average monthly price change of wheat in % of the wheat years between 1960 and 2015

When assessing the average standard deviation on a monthly basis, or volatility, different pattern can be found. In Figure 7, the average monthly volatility of wheat in percentage is given. It can be noted that throughout almost all the months, volatility is above 3%, with the largest volatility in August, June and July. It can also be noted that the in all the months the volatility of the wheat price is larger than the average price increase or decrease.

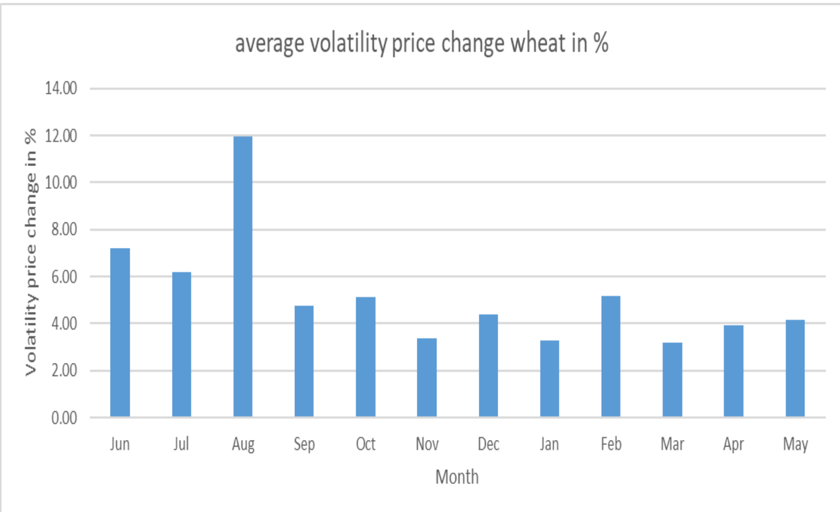


Figure 7: the average monthly volatility of the price change of wheat of the wheat years 1960 and 2015

Next, the price change on a yearly scale was assessed. In Table 1, the results of the price changes are shown, both ranked from high to low concerning percental price change, as well as chronologically. In chapter 4.2, the yearly price changes are assessed using the SOFA reports from the FAO.

Table 1: The price change in %, corrected for inflation, of wheat prices in a wheat year (June to May) from crop year 1960 to 2015, ranked according largest change (left) and chronological (right), with increase in green and decrease in orange.

year ranked	corrected increase %
2010	94.92
2007	73.62
1972	57.90
1995	46.90
1973	41.11
1991	37.26
1977	33.16
2005	23.55
2003	21.91
1987	19.14
2006	18.04
1993	17.65
2000	16.46
1988	14.88
1961	13.77
2002	12.10
1994	11.69
2012	11.52
1965	11.28
1970	10.40
1978	7.21
1975	6.23
1986	4.63
1969	2.62
1999	1.89
1962	1.52
1982	1.13
1960	0.96
1968	0.04
1983	-0.10
1963	-0.12
2016	-0.61
2001	-0.90
1966	-2.74
2013	-5.61
1980	-5.97
1964	-6.35
1985	-6.55
1984	-8.21
2004	-8.85
1979	-9.86
1971	-10.77
1974	-11.29
2011	-11.66
1967	-11.87
1998	-11.94
1992	-12.67
1981	-13.65
1997	-15.97
1989	-16.20
1990	-18.78
2015	-18.81
2014	-19.42
1996	-25.02
2009	-25.60
2008	-26.88
1976	-44.09

year	corrected increase %
1960	0.96
1961	13.77
1962	1.52
1963	-0.12
1964	-6.35
1965	11.28
1966	-2.74
1967	-11.87
1968	0.04
1969	2.62
1970	10.40
1971	-10.77
1972	57.90
1973	41.11
1974	-11.29
1975	6.23
1976	-44.09
1977	33.16
1978	7.21
1979	-9.86
1980	-5.97
1981	-13.65
1982	1.13
1983	-0.10
1984	-8.21
1985	-6.55
1986	4.63
1987	19.14
1988	14.88
1989	-16.20
1990	-18.78
1991	37.26
1992	-12.67
1993	17.65
1994	11.69
1995	46.90
1996	-25.02
1997	-15.97
1998	-11.94
1999	1.89
2000	16.46
2001	-0.90
2002	12.10
2003	21.91
2004	-8.85
2005	23.55
2006	18.04
2007	73.62
2008	-26.88
2009	-25.60
2010	94.92
2011	-11.66
2012	11.52
2013	-5.61
2014	-19.42
2015	-18.81
2016	-0.61

4.2 Wheat related reports

In this chapter, the results from assessing reports from the FAO, USDA, NDSU HREC and ministry of Alberta are described. The reports from the FAO, USDA, NDSU HREC and ministry of Alberta were used to define the key points in a season using crop specifics for wheat in general and spring wheat in particular, as well as attempting to explain the price changes on a monthly and yearly scale as found in chapter 4.1 by using the SOFA reports from the FAO. As partial input for the report assessment, the results from chapter 4.1 were used.

4.2.1 Crop specific key points wheat

When assessing the dependency of wheat production on the weather, the importance of the weather can be found in the form of rainfall, snowfall, temperatures and amount of sun (UK AHBD, 2017). In this research, only the dependency of wheat on water is assessed in the form of snowfall and rainfall. In Figure 8, the growing periods for winter and spring wheat are shown.

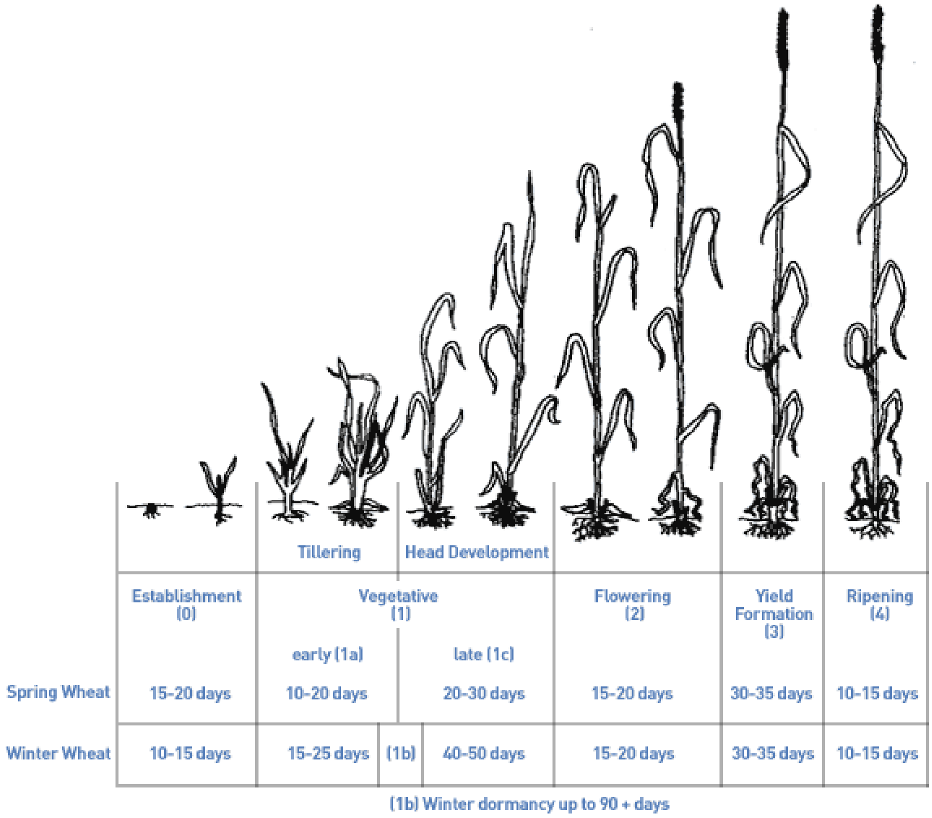


Figure 8: The growing periods of winter and spring wheat, number from 1 to 4 (FAO, 2012)

Concerning the water demand of wheat in general, the FAO states: "*In summary, provided there is adequate water during the establishment period (0) the critical periods for water deficit are: when the plants are some 15-cm tall, just completing tillering and just starting elongation; at this time, the total number of heads and number of potential seeds per head is being determined; at the end of head development to heading or the time that the flowering period (2) begins; water deficit will greatly reduce the number of seeds per head; At early yield formation period (early 3) when water deficits combined with hot, dry winds would result in an incomplete grain filling and a reduced yield of poor quality shrivelled grains*" (FAO, 2012).

From the reports of the NDSU HREC, the water demands for the spring wheat in ND can be found, being rainfall in the previous season, snow coverage, the amount of rainfall in April and May and the absence of droughts during the months of June and July (HREC, 1991; HREC, 1999; HREC, 1995; HREC, 2002; HREC, 2007; HREC, 2011). The rainfall in the previous season is seen as a proxy for stored moisture, providing soil moisture for the early growth stages of wheat. Snowfall in winter acts like a source for soil moisture, but can have a negative effect on growth too due to the flooding of fields, as happened in 2010 in ND (HREC, 2011). The importance of rainfall in the April, May, June and July can be related back to the crop specific demands as defined by the FAO and shown in Figure 8. The seasonal key points found from the crop specific literature from the HREC are therefore total amount of rain in the previous season, snowfall during the winter, rainfall in the early crop development stages between April and May, and the amount of rainfall in June and July.

From the website of the USDA, the usual sowing and harvesting time for ND could be found, as well as the average harvest season for the largest wheat exporting countries. According to the USDA, usual planting dates of spring wheat in ND commence in the middle of April and ends at the end of May, while sowing commences at the start of August and ends at the middle of September (USDA, 1997). In Table 2, the usual harvest times of the top wheat exporting countries is shown.

Table 2: dispersion of wheat harvest period, for 89% of the world wheat export during season 2014/15, with the harvest period per country indicated in green and the total contribution to the world export next to the name of the wheat exporting country (USDA, 2017).

harvest dates spring wheat	January	February	March	April	May	June	July	August	September	October	November	December
1. Canada (16.2%)												
2. United States (14.5%)												
3. Australia (11.4%)												
4. France(11.1%)												
5. Russia (10.1%)												
6. Germany (6.3%)												
7. Ukraine (3.8%)												
8. Kazakhstan (3.2%)												
9. Argentina (2.7%)												
10. Poland (2.3%)												
11. Romania (2%)												
12. Bulgaria(1.7%)												
13. Czech Republic (1.3%)												
14. Lithuania(1.3%)												
15. United Kingdom (1.1%)												

An interesting link can be noted from production schedule towards prices. From the overview of the harvest season, the average price decrease in June and July can be explained by the large inflow of wheat at the start of the harvest season. Following the principle of economy, a large supply equals a lower price if the demand is unchanged. Concerning price increases, one of the findings is that the average largest prices increases of wheat occurs at the end of the harvest season of most of the wheat producing countries. After a discussion with F. van de Loo, director at EY, a possible explanation could be the spread of the wheat production, with the true world production of wheat becoming apparent when in August, and prices subsequently reacting to the actual production (van de Loo, 2016).

From the “Alberta Crop Season in Review” reports of the ministry of Alberta, the importance of carryover moisture from the previous season, snowfall and rainfall in May, June and July (Ministry of Alberta, 2007; Ministry of Alberta, 2008; Ministry of Alberta, 2009; Ministry of Alberta, 2010; Ministry of Alberta, 2011; Ministry of Alberta, 2012; Ministry of Alberta, 2013; Minisrty of Alberta, 2014; Ministry of Alberta, 2015; Ministry of Alberta, 2016). As an example of the effect of the rainfall of the previous season and the snowfall in the winter before the sowing, the report of 2009 states: “*Also contributing to the dry conditions were the below average winter snowfall and a low carryover of moisture reserves from the previous crop season. The dry conditions, coupled with cool temperatures, caused significant delays in seeding operation and crop emergence*” (Ministry of Alberta, 2010). The importance of rainfall in the month of May can be found due to the early development phase of wheat taking place during the month of May. For instance, in the 2012 crop season review, the Ministry of Alberta states that “*The cool temperatures and several precipitation events in the May have resulted in crop development falling behind*” (Ministry of Alberta, 2013). Concerning the importance of the month of June and July, an example can be found in 2007, with the Ministry of Alberta stating that “*However, in some areas, heavy rainfall in June caused localized flooding and left water standing in low-lying fields. The persistent hot, dry weather across the province in July depleted soil moisture reserves and caused crops to abort flowering and podding, resulting in significant deterioration in crop conditions and yield potentials*” (Ministry of Alberta, 2008).

Seasonal key points for the growth of spring wheat in the Canadian Prairies as found from the reports of the ministry of Alberta are defined as rainfall previous season, total snowfall before the sowing of spring wheat, and rainfall in the months of May, June and July.

4.2.2 Wheat price change related reports

From the SOFA reports, a number of findings can be presented. First, the top 5 price decreases and increases as found in 4.1 are assessed. Next, the SOFA reports are assessed to define the general dynamics of the wheat industry. From the reports, wheat production related events are highlighted for the wheat producing areas of the world. The events are for instance an extremely wet or dry month or set of months, or extreme high or low temperatures for a specific year, periods of drought or water surplus. The reports are scanned for events that happen in North America and Canada in particular. The events related to North America and Canada in particular are collected and later tested to assess whether the events play a vital role for the crop yield in an area and can therefore be defined as a seasonal key point.

First, in Table 3 the top 5 price increases and decreases are presented with the reason for price change explained and the year in which the price change occurred.

Table 3: the top 5 of price increases and decreases for wheat, with the year in which the price change occurred, the magnitude of the price change and the reason for the price change according to SOFA reports form the FAO with a relation to weather

Year	Price change (%)	Reason
2010	94.92	Heat wave causing a drought in Russia, leading to lower wheat production and a subsequent export ban of Russian grain. Excess rainfall in the USA, drowning crops and causing lower wheat production, recovery of the world of the Global Food Crises (FOA, 2011; Oxfam, 2011)
2007	73.62	Third consecutive year of worldwide wheat production deficit due to "*weather-related production shortfalls in key exporting countries*" (FAO, 2008)
1972	57.90	Worldwide production of commodities declined due to adverse weather, shortages of grain in the USSR due to weather related lower production (drought), grain deal known as "the great grain robbery[3]" affecting wheat prices (FAO, 1973).
1995	46.90	Adverse weather worldwide causing reduced output of wheat and in the USA in particular, combined with lower production in 1993 and 1994 (FAO, 1996).
1973	41.11	Continued effect of "the great grain robbery", no weather-related production decline worldwide, record crop in the USA (FAO, 1974).
1976	-44.09	Second year of above expectation wheat production worldwide due to favourable weather, reaction to price increase in 1972 and 1973 (FAO, 1977).
2008	-26.88	Reaction to above expectation wheat production worldwide due to favourable weather, reaction to price increases during the Global Food Crises (FOA, 2011).
2009	-25.60	Reaction to above expectation wheat production worldwide due to favourable weather, reaction to price increases during the Global Food Crises (FOA, 2011).
1996	-25.02	Recovery of the USA concerning wheat production due to favourable weather, reaction to price increase of 1996 (FAO, 1997).
2014	-19.42	Recovery from the high wheat prices during the World Food Crisis, normal wheat production worldwide due to favourable weather (FAO, 2015).

[3] "The great grain robbery" is a name for a large grain deal between the USSR and the USA, in which the USA sold over 30% of the average production which they could not actually afford to sell due to domestic consumption. Due to a faulty system for subsidies, the USA lost over $300 million in subsidies on the Russian and other export sales (Luttrell, 1973).

As can be seen from Table 3, multiple factors can influence prices, with weather events highlighted as the main driving force behind four out of five of the price increases and all of the price decreases. In the SOFA reports, no specific months were highlighted, but rather areas and by what weather effect the area was hit.

The demand for wheat is steadily increasing throughout the years (USDA, 2017). Because of the steady increase in wheat demand, the prices of wheat are dependent on the production of wheat. From the reports of the FAO, negative weather effects such as a drought in summer or a surplus of rainfall in spring or summer can have a strong influence on the production of wheat and as a result on the price of wheat. Due to the spread-out character of the grain market, the total production of wheat is the main driver of the market prices, as can be confirmed by the price changes found in 4.1.

One of the factors influencing the price changes in particular in the years between 2006 to 2011, is the Global Food Crises (GFC). The GFC started with two years of lower production, being 2002 and 2003 (USDA, 2016). Two food crises occurred during 2006 to 2011 period, with one between 2007 to 2008 and one from 2010 to 2012. During the both the GFC, food prices rose "significantly", before collapsing again (EC, 2011). According to the UN, "*The food crisis is a result of a complex interplay of several factors. Some of these factors have recently emerged, such as excessive speculation in agricultural commodity futures markets, drought-induced crop failures in major grain- and cereal-producing regions and the surge in biofuel production in Europe and the United States. Other causes are longer-term, including reduced national and international investments in developing-country agriculture, distortions in the international trading system and changing consumption patterns. All these factors have adversely affected agricultural production*" (UN, 2011) .

The price increases are the results of a combination of production and market working according to the SOFA reports of the FAO. In particular, a serious of years with low wheat production in combination with increased speculation on the commodity market is proven to be an important combination concerning price increases between 2006 and 2011. Price increases due to only weather related production deficit occurred in 1993, 1994 and 1995 (FAO, 1996), and in 2002 and 2003 (FAO, 2005), where the price increases in 2005, 2006, 2007, 2010 and 2012 occurred due to the combination of market speculation and weather related production deficit (UN, 2011).

The price decreases in Table 3 were a reaction to a large price increase in the previous year followed by a recovery of the production of the next wheat year due to favourable weather or the absence of for instance periods of drought. The reason for the connection to the weather was the fact that due to crop specifics of wheat as discussed in 4.2.1, particular weather and climate requirements are needed for optimal wheat growth. When the crop specific demand of wheat was met concerning precipitation and the absence of long periods of drought, production is adequate and therefore supply should meet the demand. A reaction to adequate production and thus supply can be seen in 1989, 1990, 1996, 2004, 2008, 2009 and 2011 in Table 1 (USDA, 2017).

Next, the general dynamics of the wheat market as mentioned in the SOFA reports from the FAO were assessed. When assessing the SOFA report between 1960 to 2015 as defined in 3.1.2 concerning weather affecting production and price of wheat, in particular the weather leading up to the growing season and the weather during the growing season are discussed. The SOFA reports do not go into area specifics, so the results from the area specifics as found in 4.2.1 are used to better understand the importance of the pre-season precipitation and precipitation during the growing season.

The pre-season precipitation is of importance, due to the fact that the lack of precipitation in a previous season provides a low amount of soil moisture at the start of the growing season. Examples of the occurrence of a previous bad season can be seen in 1995, 2003 and 2007. However, an excess of water can also provide a negative effect by drowning the crops, as found in 4.2.1. Precipitation during a season is of importance for the growing of wheat. Droughts or a surplus of water can have a negative effect on crop growth, as found in 1972 and 2010. These findings are also in line with 4.2.1.

When combining all the SOFA reports of the FAO and the findings concerning the price change, the general dynamics of the wheat market can be defined. In Figure 9, the general dynamics are presented. In the figure, the green arrows represent quantifiable links, the red unquantifiable and the orange the result of the combination between green and red.

Other aspects such as policies, economy and currency can have an effect on the prices via supply and demand, but are unquantifiable from an engineering perspective. The resulting supply, demand, stock and finally prices are all result of the combination between the driving factors of the market of wheat. In this research, the link is specifically on the quantification of the weather, with the links represented by the green arrows. Because of the dependency of the production on the weather and the dependency of the price on the production, a continuation of the research towards the quantification of the link between the wheat production and weather with particular focus on the pre-season precipitation and the precipitation during the growing season.

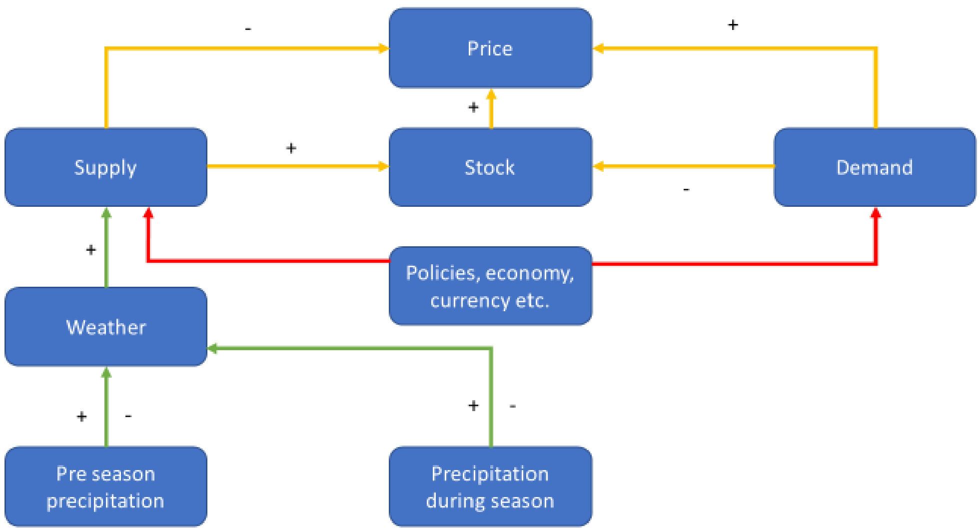


Figure 9: the dynamics of the wheat market, with green as quantifiable links, red as unquantifiable links and orange as links as a consequence of both quantifiable and unquantifiable links.

4.3 Relevant data wheat: yield changes

For wheat yield in ND and the Canadian Prairies, data from the NDWA and the USDA was used. To use the yield data provided by the NDWA and the USDA, processing steps were made to the data.

First, the spring wheat yield of ND is assessed. As can be seen from Figure 10, wheat yield in ND was constantly increasing throughout the years. For wheat, 52% of the variance of wheat yield can be explained by a trendline, which is a linear regression between wheat yield in ND and time. For the wheat yield in ND, a linear trendline was found with the formula:

$$EWY_{ND} = 0.42 * \mathrm{n} + 22.84 \; (eq. 5)$$

With:

EWY_{ND} = The Expected Wheat Yield North Dakota by using the linear trendline in bushels per acre

n = The assessed production year with start condition 1971

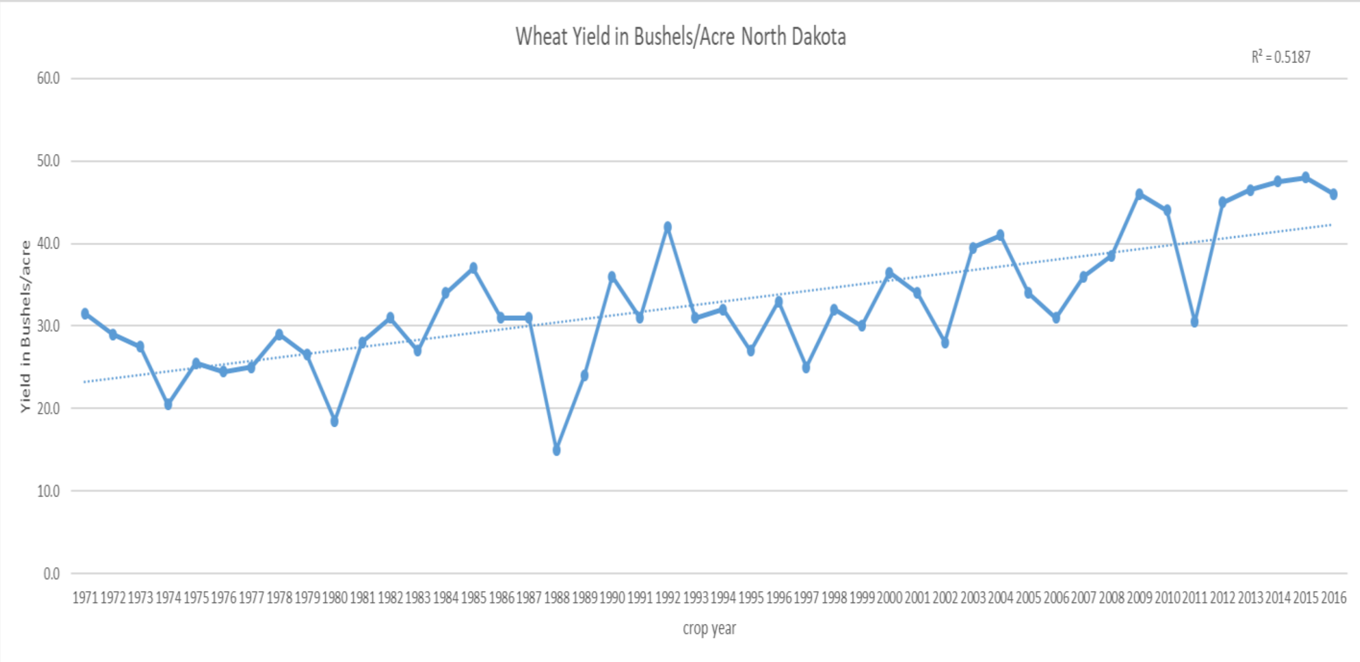


Figure 10: the wheat yield in ND as a function of time, with a linear trendline fitted to represent the trend in the data

From visual analysis of Figure 10, the yield per area of wheat in ND varies significantly per year, with for instance a difference of difference of 15 bushels per acre between 1987 and 1988. However, an increase in yield can be seen throughout the years.

In Figure 11, the residuals from the trend line can be seen. As can visually be deducted from the figure, the residuals do not appear to be randomly dispersed. In the figure, a slight parabolic form can be seen. The parabolic form of the residuals could indicate that a quadratic function could be a better fit for the trendline. However, for this research, a linear regression as first approximation was accepted.

To compensate of the increase in yield throughout wheat producing years, a trendline was accepted as a first assumption and approximation, and it was assumed that the wheat yield increases linearly as a function of time for this research. As a result, the relative difference from the linear trend line was accepted and used in further parts of this research.

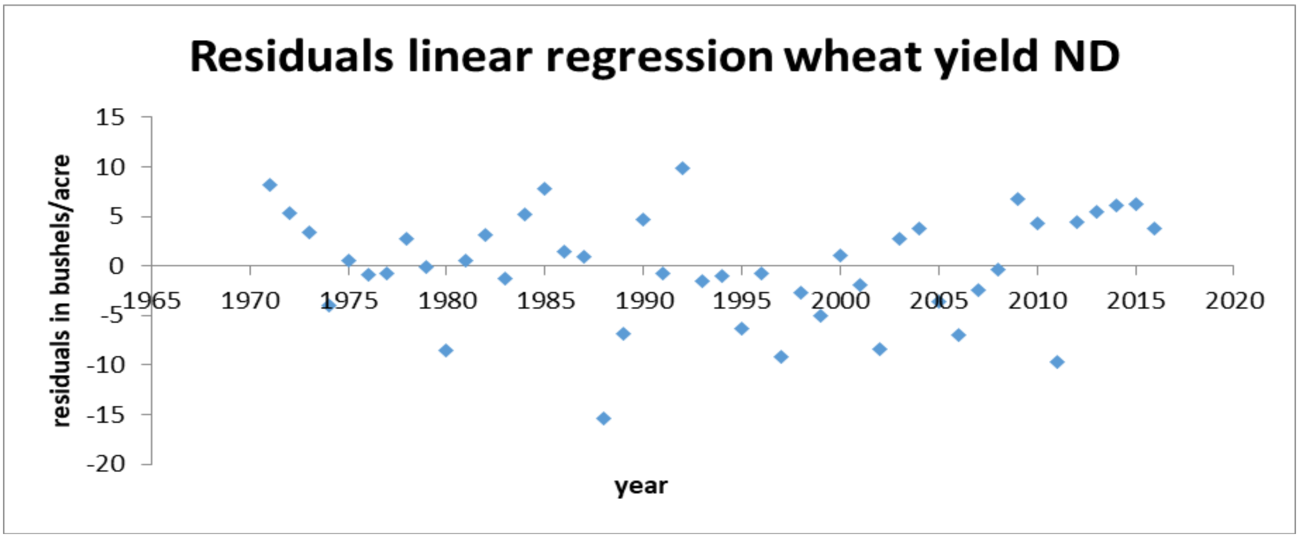


Figure 11: residuals for the linear regression trend line between wheat yield in ND as a function of time

Second, the wheat yield of the Canadian Prairies was evaluated. As mentioned in 2.1.3, the total yield of Canada was used as a first approximation. As can be seen from Figure 12, wheat yield in Canada was constantly increasing throughout the years. For wheat, 75% of the variance of wheat yield in Canada can be explained by a trendline, which is a linear regression between wheat yield and time. For the wheat yield in Canada, a linear trendline was found with the formula:

$$EWY_{CA} = 0.03 * \mathrm{n} + 1.25 \ (eq. 6)$$

With:
EWY_{CA} = The Expected Wheat Yield Canada by using the linear trendline in megaton per hectare
n = The assessed production year with start condition 1960/61

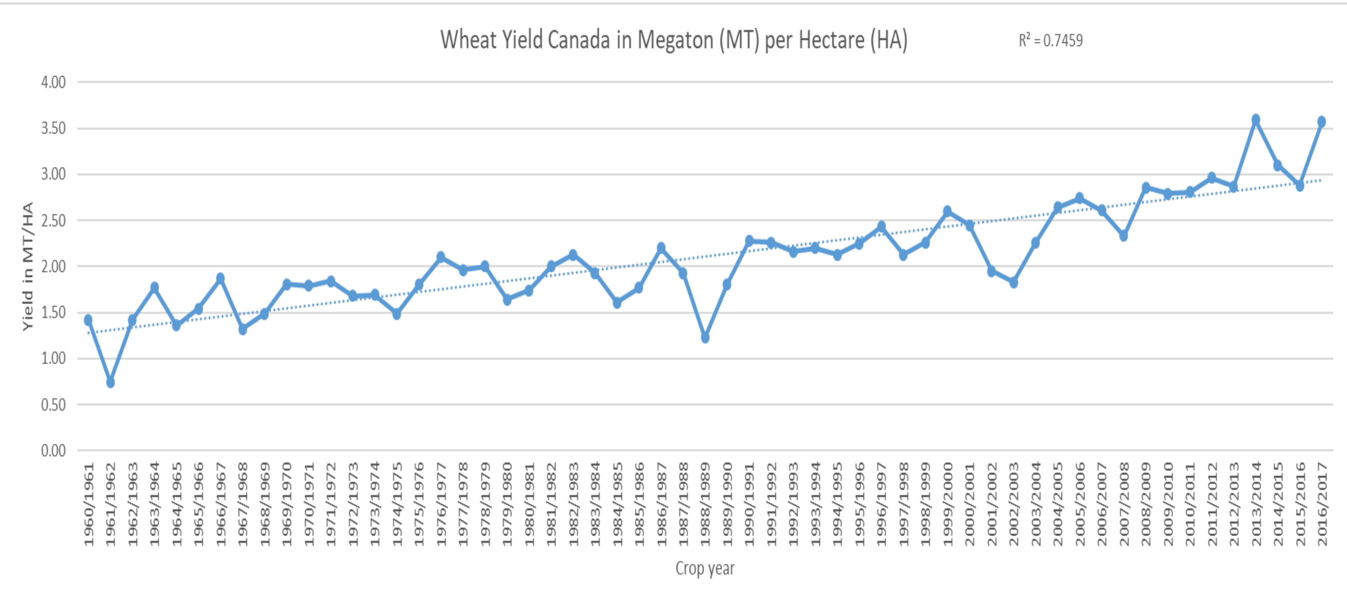


Figure 12: the wheat yield of Canada as a function of time, with a linear trendline fitted to represent the trend in the data

From visual analysis of, a steady increase in yield can be seen, with years such as 1988/89 and 2002/03 and 2003/2004 far under the trend and years such as 2013/2014 and 2016/2017 far above the trend.

In Figure 13, the residuals from the trend line can be seen. As can visually be deducted from the figure, the residuals again do not appear to be randomly dispersed, with a number of large negative outliers. In the figure, a slight parabolic form can be seen as well. The trendline was accepted as a first assumption and approximation, and it was assumed that wheat yield in Canada and subsequently the Canadian Prairies increases linearly as a function of time for this research. Consequently, the relative difference from the linear trend line was used in further parts of this research.

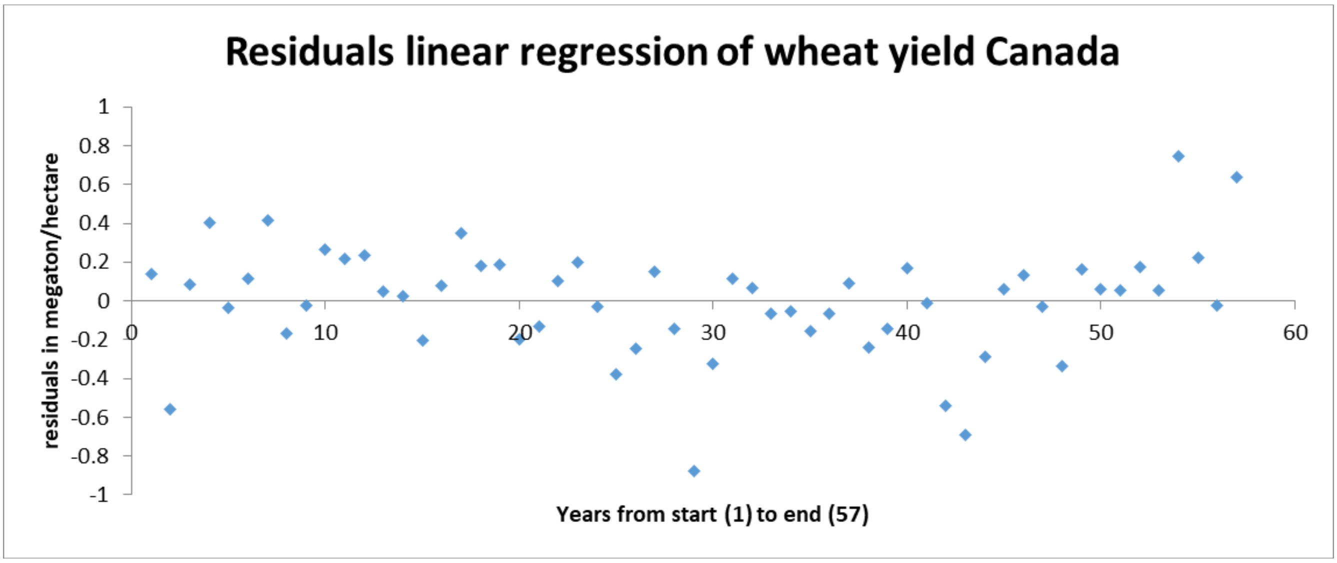


Figure 13: residuals for the linear regression trend line between world wheat production as a function of time

4.4 Relevant data wheat: influence local production wheat on world production wheat

A linear multivariate regression was used in this research as a first approximation to determine the strength of the wheat production of the USA and Canada towards the world wheat production.

The correlation of the USA and Canada was first tested using the Pearson coefficient, to determine if the variables were independent within the used dataset. The correlation between the USA and Canada was found to be 0.67, which indicates a moderate correlation according to the Pearson coefficient. Therefore, the use of the USA and Canada as independent variables in the linear multivariate regression is not fully justified due to the fact that the independent variables have a correlation towards each other. However, as a first approximation for this research, the use of the USA and Canada as independent variables was accepted.

A multivariate linear regression was performed with the world wheat production as dependent variable and the USA and Canada as independent variables. The regression was tested between 1960 and 2016 using data from the annual reports from the PS&D tool of the USDA as defined in 3.1.3. In Table 4, the results of the regression are shown. In the table, the R^2 value of the regression is shown, as well as the significant F value of the regression, the coefficients of the independent variables the USA and Canada and the P-values of the independent variables.

Table 4: the results of the regression of the world wheat production as a function the USA and Canada wheat production, with R^2, P- and F-value and coefficients

R^2	0.64
Significant F value regression	9.71×10^{-13}
Production USA P-value	0.0095
Production USA Coefficient	3.71
Production Canada P-value	2.93×10^{-6}
Production Canada Coefficient	13.93

When assessing the statistical significance of the variables of the regression models, the F- and P-values were within the requirement of being below 0.05. Therefore, the significance of the variables was proven in accordance with Bland (Bland, 1995). For the regression model, the coefficients of the variables were also acceptable, since both the coefficient of the USA (3.71) and Canada (13.93) in the regression were within the same order of magnitude. Therefore, both independent variables were of mutual importance for the regression model and none of the independent variables is repressed by Windows Excel.

Concerning the R^2 values, 0.64 of the variance of the wheat production could be explained by assessing the wheat production of the USA and Canada. By using the Pearson coefficient, the relation between the world wheat production and the wheat production predictor using the wheat production of the USA and Canada could be qualified as “moderate”. Therefore, the dependency of the world wheat production on the wheat production of USA and Canada is partly confirmed. While the USA and Canada provide 12.5% of the world production of wheat as of 2016, the effect of the wheat production of the USA and Canada on the world wheat production is still noticeable (USDA, 2017).

Based on the found results, a predictor for the world wheat production based on the wheat production of the USA and Canada can be formed. The results from the regression analysis can be summarized in equation 7. The resulting graph from equation 7 and Table 4 is shown in Figure 14.

$$PWPW = 3.71 * WP_{USA} + 13.93 * WP_{CA} \ (eq.7)$$

With

$PWPW$ =	Predicted World Production Wheat in thousands of tons
WP_{USA} =	Wheat Production USA in thousands of tons
WP_{CA} =	Wheat Production Canada in thousands of tons

As can be seen in Figure 14, the predictor for the wheat production is off in most of the years, but follows the same movement as the actual wheat production. The difference between the world wheat production predictor and the actual world wheat production can be explained. The USA and Canada contribute 12.5% to the world wheat production in 2016/17. However, the production of other wheat producing countries such as Russia, Australia, China and India also affect the world wheat production. The linear regression ignores the contribution of other regions, which in turn causes the regression to be off when the USA and Canada have a lower or higher wheat production, but other wheat producing countries do not. The actual wheat production reacts less volatile to the changes in wheat production in the USA and Canada, as can be seen in Figure 14. While the USA and Canada provide a large contribution to the world wheat production, the wheat production of the USA and Canada combined does not dominate the market. The finding is in line with the definition of a dominant market share, indicating a market share of over 60%, which is not obtained by the USA and Canada concerning wheat production (Athey & Schmutzler, 2001). The model is not stable, with large deviation from the actual world wheat production. In years where the wheat production of the USA and Canada is low as in for instance 1988/98 and 2002/03, the model assumes that the world production is low as well.

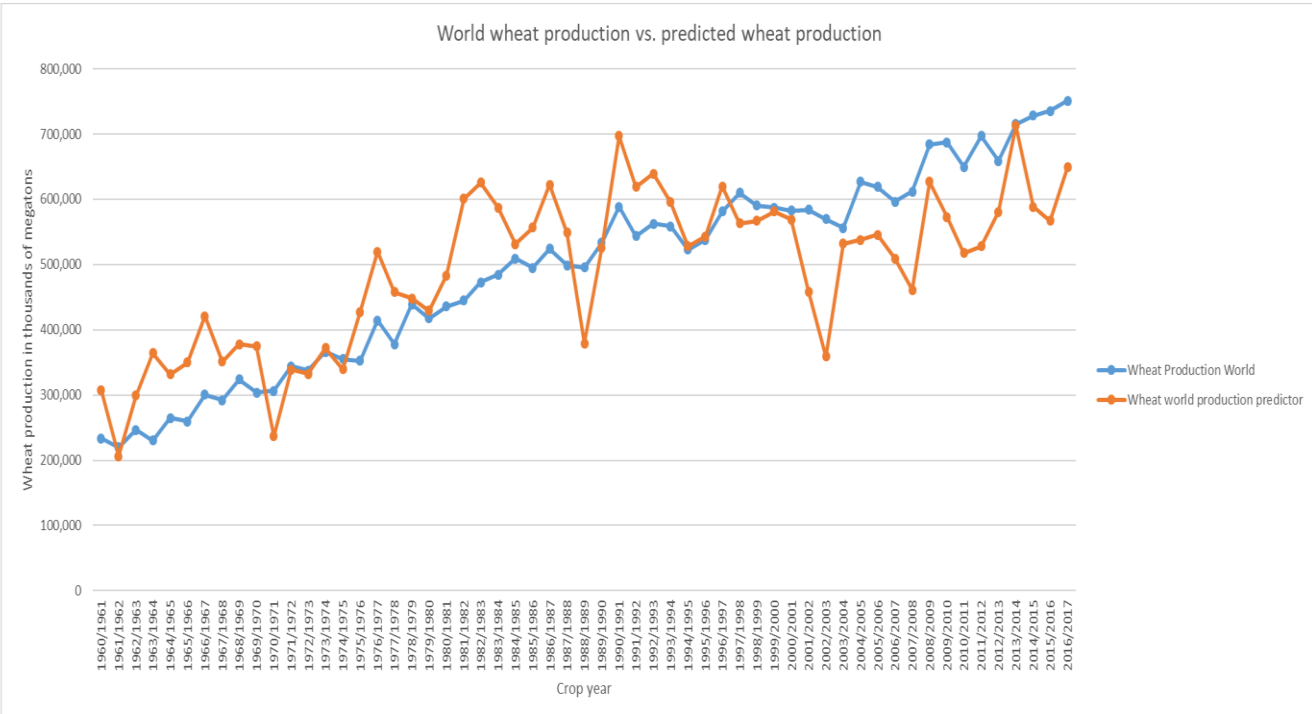


Figure 14: the actual world production of wheat (blue) versus the predicted world wheat production as a function of the USA and Canada (orange), given as a function of production in thousands of megatons versus the year of the harvest.

In Figure 15, the contribution to the residuals of the regression model for the independent variables USA and Canada are shown. In the figure, the residuals visually do not appear to be randomly dispersed. For the residuals of the USA, a slight mountain parabolic shape can be seen, indicating that a quadratic function for the production of the USA could be appropriate. For the residuals of Canada, the negative residuals visually appear to be more numerous than the positive residuals, and also concentrated between 15000 and 30000. Because of the limitation of this research to linear regressions only, the approximation was accepted.

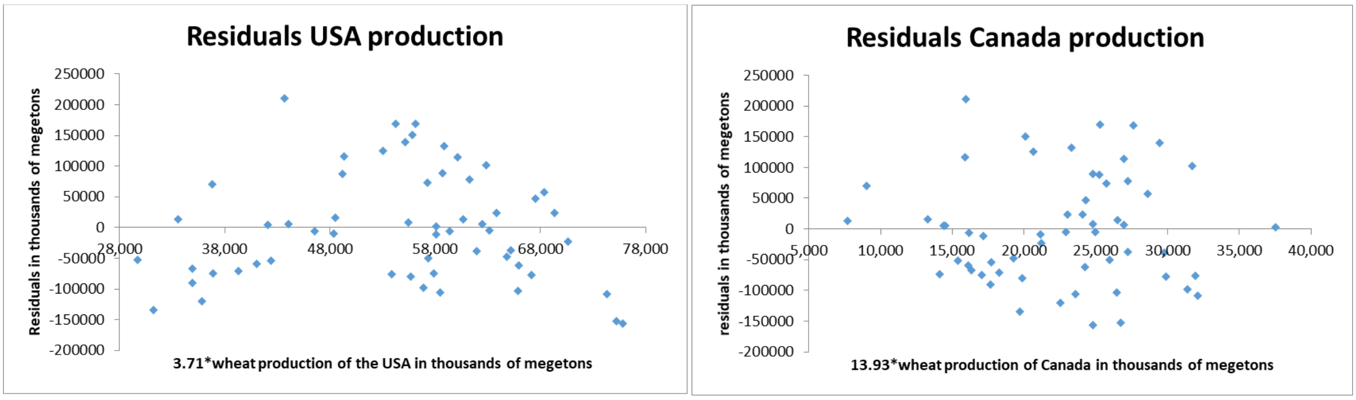


Figure 15: Residuals for the wheat production predictor of the USA and Canada in thousands of megatons. The graph shows the relative contribution of the predictor of the USA (left) and Canada (right) towards the world wheat predictor. On the vertical axis, the contribution towards the residual of the trendline in shown in thousands of megatons, while on the horizontal axis the relative wheat production contribution of the USA (left) and Canada (right) is presented in thousands of tons.

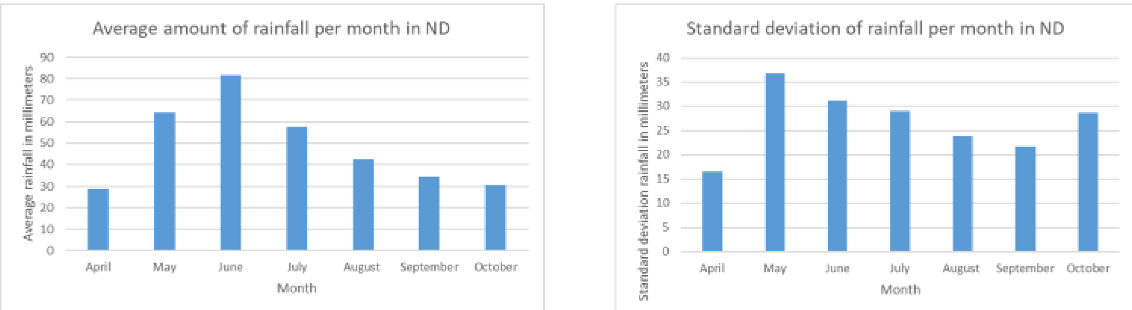
4.5 Relevant data wheat: weather station data

After cleaning up the data, a correlation matrix could be constructed to determine the correlation between the various weather stations used in ND and the Canadian Prairies. First, the weather stations of ND are assessed. The results of the correlation assessment are shown in Table 5. The correlations as shown in the correlation matrix are the Pearson correlations between the average monthly rainfall between April and October, consisting of 7 points in total. For ND, correlations reach a low of 0.8 between for Bowman and Jamestown. The correlation of 0.8 and higher can be qualified as “strong” according to the Pearson correlation. Therefore, the simplification to take the monthly arithmetic mean of all the available weather stations was accepted for this research.

Table 5: the correlation matrix of the Pearson correlation between average monthly rainfall between the used weather stations in ND

	Bottineau	**Turtle Lake**	**Beach**	**Dickinson**	**Jamestown**	**Bowman**	**Hazen**	**Hettinger**
Bottineau	1.00	0.96	0.90	0.94	0.92	0.84	0.93	0.95
Turtle Lake	0.96	1.00	0.96	0.98	0.98	0.86	0.99	0.96
Beach	0.90	0.96	1.00	0.99	0.93	0.94	0.94	0.97
Dickinson	0.94	0.98	0.99	1.00	0.95	0.92	0.96	0.99
Jamestown	0.92	0.98	0.93	0.95	1.00	0.80	0.99	0.91
Bowman	0.84	0.86	0.94	0.92	0.80	1.00	0.81	0.95
Hazen	0.93	0.99	0.94	0.96	0.99	0.81	1.00	0.93
Hettinger	0.95	0.96	0.97	0.99	0.91	0.95	0.93	1.00

After taking the arithmetic mean of the used weather stations, the average monthly rainfall and deviation corresponding to a specific month could be determined for ND. In Figure 16, the arithmetic mean of the weather stations per month and the corresponding deviation is shown.



	April	May	June	July	August	September	October
Average in mm	29	64	82	58	43	34	30
Standard deviation in mm	16	37	31	29	24	22	29
std dev/average	0.57	0.57	0.38	0.50	0.56	0.63	0.94

Figure 16: the average monthly rainfall of the dataset, using the arithmetic mean of the used weather stations in ND (top left), the standard deviation of the average monthly rainfall in the dataset, using the arithmetic mean of the used weather stations in ND (top right) and the numbers of the monthly average rainfall from the arithmetic mean, the standard deviation related to the monthly average and the ratio between the standard deviation relative to the average rainfall (central bottom)

The outcome of the analysis of the monthly arithmetic mean rainfall as shown in Figure 16 confirm the described weather as found in the description from NWSEND, with wet months in the summer period between May and July, and October the driest month (NWSEND, 2002). The standard deviations follow a different pattern as the rainfall, with the highest absolute deviation in September and October. In Figure 16, the averages, standard deviation and the ratio of the standard deviation to the average rainfall is given in the bottom central.

As can also be seen from Figure 16, the months with the highest mean to deviation ratio in relative sense are October with 0.94 and September with 0.63. The high ratio between mean rainfall and deviation translate into the observation that a high amount of variability could exists between the expected and actual rainfall in a specific month. The month with the highest standard deviation in absolute sense is the month of May. May is important due to the sowing and early development of spring wheat, as indicated in 4.2.1. With a ratio of the standard deviation to average rainfall of 0.57, the rainfall in May can divert significantly from the mean. High amounts of rainfall in June and July are needed for the further development of spring wheat, corresponding with the flowering period of wheat as defined in 4.2.1.

In Table 6, the correlation index of the snowfall around the three stations in ND are presented. The correlations as shown in the correlation matrix are the Pearson correlations between the average total snowfall from November to March for one year, consisting of 1 point in total. For ND, correlations reach a low of 0.7 between for Bottineau and Jamestown. The correlation of 0.7 and higher can be qualified as “strong” according to the Pearson correlation. Therefore, the simplification to take the arithmetic mean of all the available weather stations concerning snowfall was accepted for this research.

Table 6: the correlation matrix of the Pearson correlation between average total snowfall between the used weather stations in ND

	Jamestown	Bottineau	Dickinson
Jamestown	1.00	0.70	0.88
Bottineau	0.70	1.00	0.78
Dickinson	0.88	0.78	1.00

After taking the arithmetic mean of the used weather stations, the average yearly snowfall and deviation corresponding to the snowfall could be determined for ND. In Table 7, the arithmetic mean snowfall of the weather stations and the corresponding deviation is shown. It is not possible to relate the importance of the thickness of the snow coverage on itself, but as discussed in 4.2.1, the importance of the thickness of the snow coverage in combination with the rainfall of the previous season is of importance to the development of spring wheat in ND. The seasonal key points defined for ND based on the rainfall data were the rainfall of the previous year and snow coverage for soil moisture reserves, and rainfall during April, May, June and July for the sowing and developing for spring wheat.

Table 7: the average snowfall from the arithmetic mean, the standard deviation related to the average snowfall and the ratio between the standard deviation relative to the average rainfall

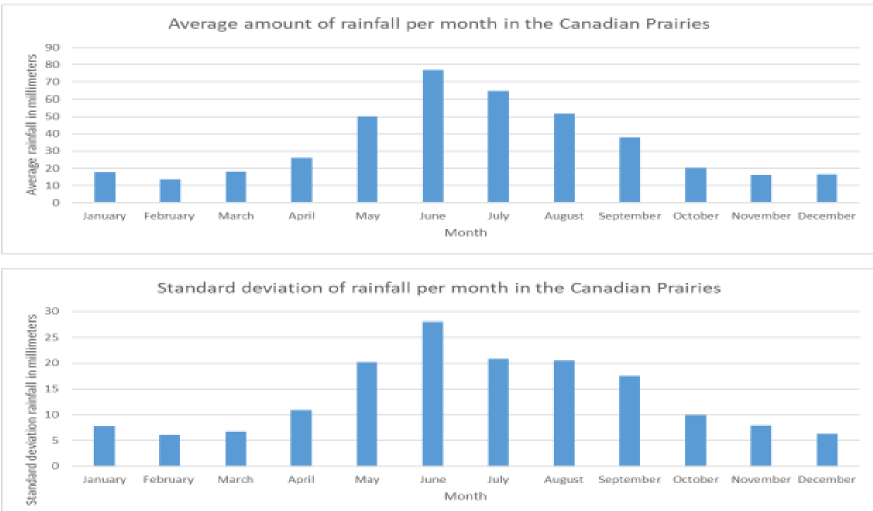
Average in mm	1102
Standard deviation in mm	435
Std dev/average	0.39

Second, the weather stations of the Canadian Prairie are assessed. The results of the correlation assessment are shown in Table 8. The correlations as shown in the correlation matrix are the Pearson correlations between the average monthly rainfall of one full year, consisting of 12 points in total. For the Canadian Prairies, correlations reach a low of 0.81 between for Climax and Edmonton. The correlation of 0.81 and higher can be qualified as “strong” according to the Pearson correlation. Therefore, the simplification to take the monthly arithmetic mean of all the available weather stations was accepted for this research.

Table 8: the correlation matrix of the Pearson correlation between average monthly rainfall between the used weather stations in the Canadian Prairies

	Medicine Head	Swift current	Lethbridge	Estevan	Regina	Red Deer	Camrose	Saskatoon	Brandon	Altona	Climax	Calgary	Edmonton	Lloydminster
Medicine Head	1.00	0.99	0.99	0.97	0.97	0.93	0.90	0.96	0.94	0.97	0.98	0.99	0.86	0.90
Swift current	0.99	1.00	0.98	0.97	0.98	0.94	0.92	0.97	0.94	0.97	0.97	0.99	0.89	0.92
Lethbridge	0.99	0.98	1.00	0.95	0.95	0.90	0.86	0.94	0.92	0.95	0.97	0.97	0.82	0.87
Estevan	0.97	0.97	0.95	1.00	0.99	0.96	0.93	0.98	0.98	0.99	0.94	0.99	0.92	0.95
Regina	0.97	0.98	0.95	0.99	1.00	0.97	0.95	0.99	0.98	0.99	0.95	0.99	0.93	0.96
Red Deer	0.93	0.94	0.90	0.96	0.97	1.00	0.99	0.99	0.98	0.98	0.88	0.97	0.99	0.99
Camrose	0.90	0.92	0.86	0.93	0.95	0.99	1.00	0.97	0.96	0.95	0.85	0.95	0.99	0.99
Saskatoon	0.96	0.97	0.94	0.98	0.99	0.99	0.97	1.00	0.97	0.99	0.94	0.99	0.96	0.97
Brandon	0.94	0.94	0.92	0.98	0.98	0.98	0.96	0.97	1.00	0.99	0.88	0.98	0.96	0.98
Altona	0.97	0.97	0.95	0.99	0.99	0.98	0.95	0.99	0.99	1.00	0.94	0.99	0.94	0.96
Climax	0.98	0.97	0.97	0.94	0.95	0.88	0.85	0.94	0.88	0.94	1.00	0.95	0.81	0.86
Calgary	0.99	0.99	0.97	0.99	0.99	0.97	0.95	0.99	0.98	0.99	0.95	1.00	0.93	0.96
Edmonton	0.86	0.89	0.82	0.92	0.93	0.99	0.99	0.96	0.96	0.94	0.81	0.93	1.00	0.99
Lloydminster	0.90	0.92	0.87	0.95	0.96	0.99	0.99	0.97	0.98	0.96	0.86	0.96	0.99	1.00

After taking the arithmetic mean of the used weather stations, the average monthly rainfall and deviation corresponding to a specific month could be determined for the Canadian Prairies. In Figure 17, the arithmetic mean of the weather stations per month and the corresponding deviation is shown.



	January	February	March	April	May	June	July	August	September	October	November	December
Average in mm	18	13	18	26	50	77	65	52	38	20	16	17
Standard deviation in mm	8	6	7	11	20	28	21	21	18	10	8	6
std dev/average	0.44	0.46	0.38	0.42	0.40	0.37	0.32	0.40	0.46	0.49	0.49	0.38

Figure 17: the average monthly rainfall of the dataset, using the arithmetic mean of the used weather stations in the Canadian Prairies (top), the standard deviation of the average monthly rainfall in the dataset, using the arithmetic mean of the used weather stations in the Canadian Prairies (middle) and the numbers of the monthly average rainfall from the arithmetic mean, the standard deviation related to the monthly average and the ratio between the standard deviation relative to the average rainfall (bottom)

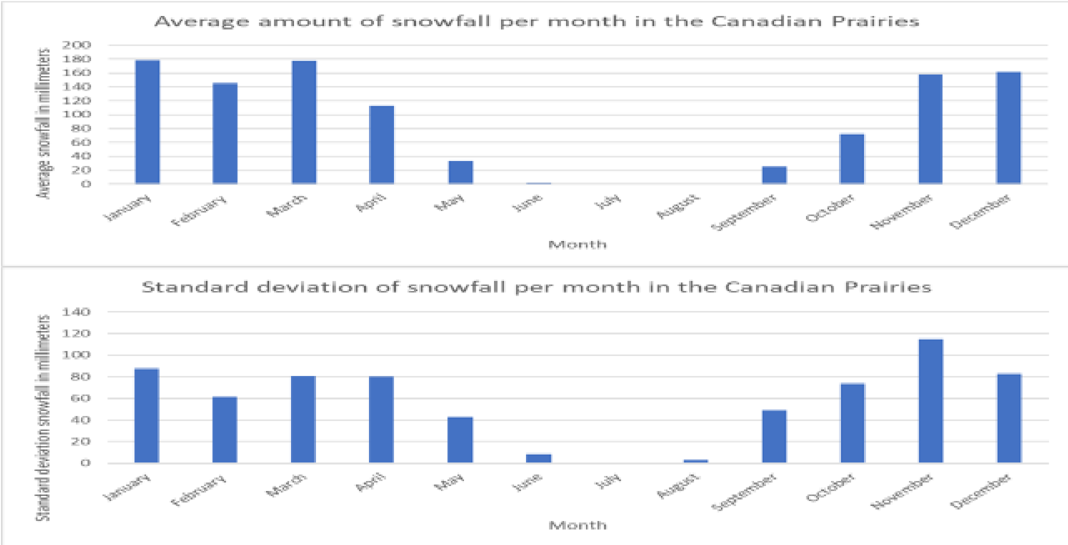
The outcome as shown in Figure 17 show the yearly pattern of the rainfall during a full year. As can be seen from the figure, the months of May, June, July and August are the wettest months of the year. The rainfall patterns diverge from the description as provided by Vickers, stating that the summers are "*generally dry*" (Vickers, Buzza, Schmidt, & Mullock, 2000). However, it could be noted that in Figure 17 only rainfall is accounted for, not snowfall.

Concerning snowfall, the results of the correlation assessment are shown in Table 9. The correlations as shown in the correlation matrix are the Pearson correlations between the average monthly snowfall of one full year, consisting of 12 points in total. For the Canadian Prairies, correlations reach a low of 0.81 between for Altona and Calgary. The correlation of 0.81 and higher can be qualified as "strong" according to the Pearson correlation. Therefore, the simplification to take the monthly arithmetic mean of all the available weather stations was accepted for this research.

Table 9: the correlation matrix of the Pearson correlation between average monthly rainfall between the used weather stations in the Canadian Prairies

	Medicine Head	Swift current	Lethbridge	Estevan	Regina	Red Deer	Camrose	Saskatoon	Brandon	Altona	Climax	Calgary	Edmonton	Lloydminster
Medicine Head	1.00	0.98	0.97	0.99	0.99	0.98	0.99	0.99	0.99	0.98	0.97	0.89	0.99	0.96
Swift current	0.98	1.00	0.98	0.98	0.99	0.98	0.96	0.98	0.98	0.93	0.93	0.92	0.97	0.96
Lethbridge	0.97	0.98	1.00	0.96	0.97	0.99	0.95	0.95	0.94	0.92	0.94	0.96	0.96	0.96
Estevan	0.99	0.98	0.96	1.00	0.98	0.98	0.99	0.99	0.99	0.97	0.97	0.89	1.00	0.97
Regina	0.99	0.99	0.97	0.98	1.00	0.97	0.98	0.99	0.99	0.96	0.94	0.88	0.97	0.95
Red Deer	0.98	0.98	0.99	0.98	0.97	1.00	0.96	0.96	0.95	0.93	0.96	0.96	0.97	0.97
Camrose	0.99	0.96	0.95	0.99	0.98	0.96	1.00	0.99	0.98	0.99	0.97	0.85	0.99	0.94
Saskatoon	0.99	0.98	0.95	0.99	0.99	0.96	0.99	1.00	0.99	0.98	0.95	0.85	0.99	0.94
Brandon	0.99	0.98	0.94	0.99	0.99	0.95	0.98	0.99	1.00	0.96	0.95	0.84	0.98	0.94
Altona	0.98	0.93	0.92	0.97	0.96	0.93	0.99	0.98	0.96	1.00	0.97	0.81	0.97	0.92
Climax	0.97	0.93	0.94	0.97	0.94	0.96	0.97	0.95	0.95	0.97	1.00	0.87	0.97	0.97
Calgary	0.89	0.92	0.96	0.89	0.88	0.96	0.85	0.85	0.84	0.81	0.87	1.00	0.87	0.93
Edmonton	0.99	0.97	0.96	1.00	0.97	0.97	0.99	0.99	0.98	0.97	0.97	0.87	1.00	0.96
Lloydminster	0.96	0.96	0.96	0.97	0.95	0.97	0.94	0.94	0.94	0.92	0.97	0.93	0.96	1.00

After taking the arithmetic mean of the used weather stations, the average monthly rainfall and deviation corresponding to a specific month could be determined for the Canadian Prairies. In Figure 18, the arithmetic mean of the weather stations per month and the corresponding deviation is shown.



	January	February	March	April	May	June	July	August	September	October	November	December
Average	179	145	177	113	33	2	0	0	25	72	158	161
Standard Deviation	88	61	81	80	43	8	0	3	49	74	115	83
Std dev/average	0.49	0.42	0.45	0.71	1.28	4.71	0.00	7.04	1.94	1.02	0.73	0.51

Figure 18: the average monthly rainfall of the dataset, using the arithmetic mean of the used weather stations in the Canadian Prairies (top), the standard deviation of the average monthly rainfall in the dataset, using the arithmetic mean of the used weather stations in the Canadian Prairies (middle) and the numbers of the monthly average rainfall from the arithmetic mean, the standard deviation related to the monthly average and the ratio between the standard deviation relative to the average rainfall (bottom)

When combining the outcome of both Figure 17 and Figure 18, the full spread of the precipitation in throughout the year can be seen. While rainfall from October to April is low, the snowfall is high and becomes the main source of precipitation. Also, the short summer can be seen, with only July being snow free. The findings confirm the finding from the reports of the ministry of Alberta as defined in 4.2.1. The snowfall throughout the winter provides a contribution to the soil moisture recharge of the area, but also dictates the sowing date due to the need for the snow to melt before spring wheat can be sown. The summer is short, with on average large amounts of rainfall to feed the crops. The rainfall during the growing season of spring wheat in the Canadian Prairies and after the harvesting of spring wheat contributes to the soil moisture reserve and are important for the production of spring wheat in the following season, as can be found in 4.2.1. The seasonal key points defined for the Canadian Prairies based on the rainfall data and crop specifics were therefore the rainfall of the previous year and snow coverage for soil moisture reserves, and rainfall during May, June and July for the sowing and developing for spring wheat.

4.6 Possible key periods wheat season based on analytical findings: linear regression

From the reports of the FAO, USDA, NDSU HREC and the ministry of Alberta, the vital factors for the growth of spring wheat were defined, as can be found from the reports as 4.2.1 and 4.2.2. and 4.5. As seasonal key points, the rainfall in the previous season, the snowfall in the winter before sowing, rainfall during the development stage of wheat and rainfall during the flowering period of wheat were selected, as confirmed by the findings chapter in 4.2.1 and 4.5. For ND, the months for development selected were April and May, while the flowering periods of wheat for ND June and July were selected. For the Canadian Prairies, the month for development selected was May, while the flowering periods of wheat for ND June and July were selected. For total rainfall in the previous season in ND, the rainfall of April to October of the previous year was taken, while for the Canadian Prairies the rainfall between September to August was taken. For ND, the rainfall between April to October was the only rainfall data available. For the Canadian Prairies, wheat harvesting starts in September, so the rainfall from September onwards does not contribute to the current wheat year anymore and contributes to the rainfall of the previous season up to August the following year. For snowfall, the total recorded snowfall was taken for ND while for the Canadian Prairies the sum of the snowfall from October to April was taken, to coincide with the Canadian winter as defined in 2.3.

First, the linear regression was optimized for ND. After testing and running the iterative regression model a number of times with filtering of the data using the key points as found in 4.2.1, the rainfall in the previous season, the total snowfall before the sowing and the rainfall in the month of May were found to be the key points concerning the relative change of spring wheat yield in ND. This was in line with the found reports and finding in chapter 4.2.1 and 4.5 due to the coincidence of the found seasonal key points with previously defined key points for the growth of spring wheat. The final regression was performed using the defined set of variables. The rainfall in the month of April, June and July were also tested, but the P- and F-values found were far above 0.05 and therefore not significant. Due to limited available data concerning snowfall, the regression was performed using 19 points between 1997 to 2015. From the regression, it can be stated that not all of the defined seasonal key points were found to be important towards the spring wheat yield in ND. In Table 10, the results of a regression of the relative change of spring wheat yield a function of the total rainfall of the previous season, the total snowfall before the sowing and the rainfall in the month of May in ND are shown.

Table 10: the results of the regression of the relative change of the spring wheat yield as a function of the total rainfall of the previous season, the total snowfall before the sowing and the rainfall in the month of May in ND, with R2, P- and F-value and coefficients

R^2	0.54
Significant F value regression	0.0073
Rainfall previous year P-value (mm)	0.0047
Rainfall previous year Coefficient (mm)	0.045
Snow P-value (mm)	0.051
Snow Coefficient (mm)	-0.0047
May P-value (mm)	0.089
May Coefficient (mm)	0.044

As can be seen from Table 10, the P-values of the chosen seasonal key points as independent variables are not all within the 0.95 confidence bound with values below 0.05, but values are relatively close to be interpreted as statistically significant. The F-value of the regression is below 0.05, indicating the significance of the regression as a whole. The regression has an R-squared of 0.54, which translates into the finding that more than half of the change in world wheat production could be explained by using a set of season key points in ND within a spring wheat year. All the coefficients of the independent variables found in the regression are in the same order of magnitude, confirming the relevance of the coefficients.

Among the coefficients, a number of interesting results were found. The amount of total rainfall in the previous season and the amount of rainfall in the month of May have an almost equal effect on the regression, with both coefficients around the 0.04. The importance of an adequate amount of rainfall both before the start of the season and at the development state of spring wheat is a reoccurring import factor already defined in 4.2.1. The results from the linear regression are interpreted that the linear regression agrees with the notion that the amount of rainfall in the previous season and during the development phase of wheat are of importance to a successful yield, in line with the findings in 4.2.1, 4.2.2 and 4.5.

A negative effect can be found for snowfall before the sowing season. The coefficient of the snowfall is a factor 10 smaller than the coefficient for both the rainfall variables, but this occurrence can be explained by the fact that on average snow has a density 10 times smaller than water (Judson, 2000). The negative effect can be explained by the peak loading of water large snow cover provides. A large cover of snow melting provides a large amount of water in a relative short period of time, as was the case during the 2010 growing season (HREC, 2011).

Based on the found results, a predictor for the spring wheat yield based on the arithmetic mean rainfall in the previous season, the snowfall before sowing and the month of May in ND could be made. The results from the regression analysis could be summarized in equation 8 as shown below:

$$PRCYSW_{ND} = 0.045 * R_{PS} - 0.0047 * S + 0.044 * R_M (eq.8)$$

With

$PRCYSW_{ND}$ = Predictor Relative Change Yield Spring Wheat North Dakota, relative to linear trendline in bushels per acre

R_{PS} = Rainfall previous season, sum of the arithmetic mean in millimetres

S = Snowfall before sowing, arithmetic mean depth in millimetres

R_M = Rainfall May, arithmetic mean in millimetres

Using the found equation, the comparison between the predictor and the actual difference of the spring wheat yield relative to the trend could be composed. In Figure 19, the comparison between the predictor and the actual change spring wheat yield relative to the trend is shown. As can be seen from Figure 19, the predictor for the change in spring wheat yield is not a perfect fit, but does follow the same general movement of the actual difference. Differences can be seen in particular in years where the combination of key points had limited influence on the spring wheat yield, such as 2006 and 2012. In 2006, high amounts of soil moisture were stored due to high precipitation in October. However, very high temperatures negatively affected crop maturing and lowered yield (HREC, 2006). In 2012, high amounts of stored soil moisture provided a good starting point for the growth of spring wheat, but a forced early harvest in July due to above average temperatures in June and July caused a loss in wheat yield (HREC, 2012). The parameter of temperature has not been included in this research, but is found through the linear regression as being of importance to the production of spring wheat in ND. A year in which the predictor mimicked the actual yield, was in 2011. From the Western Dakota Crop Day report in 2011, an “above normal” amount of snow was noted, causing fields to be “mudded”. Due to the large amount of moisture of the snowfall, fields could not be sown until late May (HREC, 2011).The findings confirm the negative effect of a surplus of snowfall.

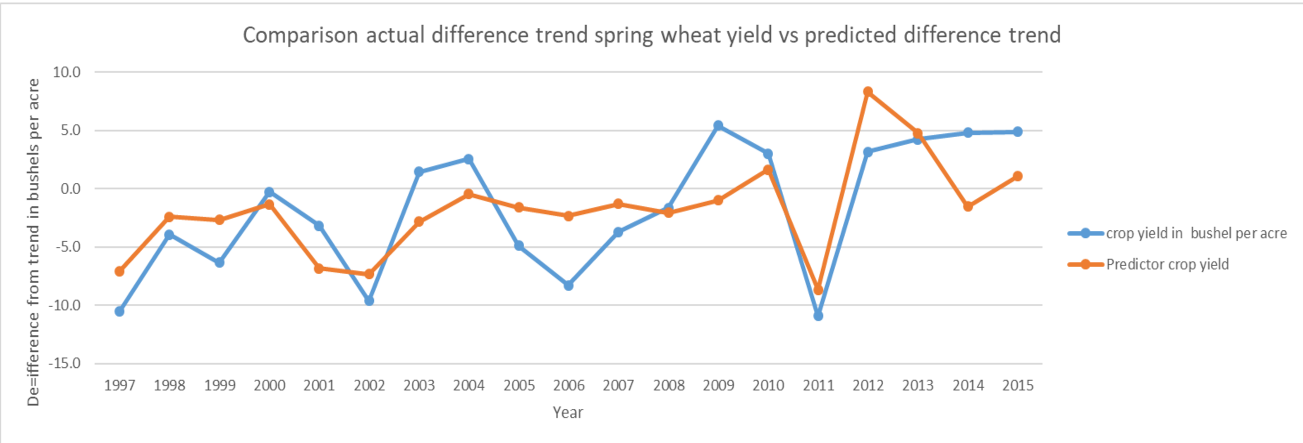
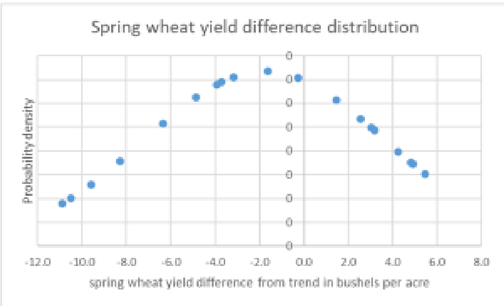
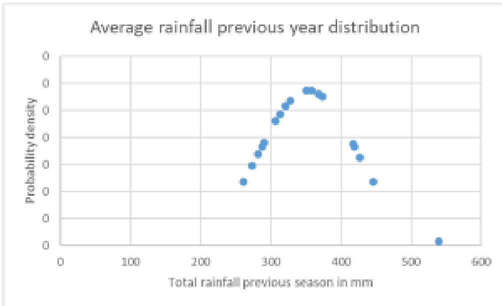


Figure 19: comparison between the actual difference of the actual spring wheat yield from the trend versus the predicted difference in spring wheat yield from the trend using the key point regression

As a final test, the rainfall and snowfall data, as well as yield differences are tested on skewness and kurtosis to determine if the data is normally distributed. In Figure 20, the probability distribution of the rainfall, snowfall and the yield difference from the trend of the spring wheat yield of ND is shown. When evaluating the data, the kurtosis of May was to be above the set boundary of 2. However, when evaluating the figure of the distribution of May, the main cause of the high kurtosis was the high above average rainfall points on the right of the peak. When assessing the results with the average rainfall as found in Figure 16, it can be stated that the even was outside the 0.95 confidence band of the mean plus two standard deviations. Therefore, the measurements can be seen as extreme events and rejected from the analysis. The skewness and kurtosis May with the extreme events filtered out and all of the other data are within the bounds as stated in 1.1. Therefore, the data was perceived as being normally distributed and the multivariate regression was subsequently accepted.



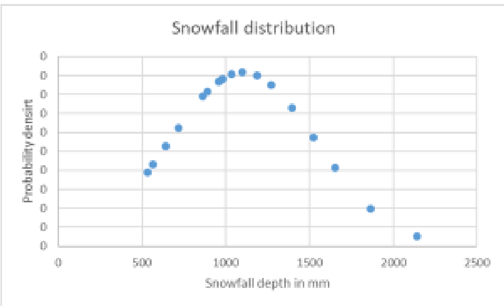


Spring wheat Yield	
Kurtosis	-1.29315
Skewness	-0.28902

Rainfall prev season	
Kurtosis	0.983187
Skewness	0.985059

Snowfall	
Kurtosis	0.193836
Skewness	0.848662

May	
Kurtosis	-1.08527
Skewness	0.183168



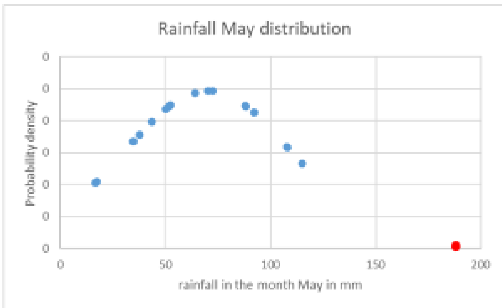


Figure 20: the probability distribution of the difference in spring wheat yield from trend, total rainfall previous season, snowfall and rainfall in May with tables containing the kurtosis and skewness of the yield, rainfall and snowfall data.

Second, the regression was optimized for the Canadian Prairies. After testing and running the iterative regression model a number of times with filtering of the data using the key points as found in 4.2.1, the rainfall of the previous season and the snowfall before sowing were found to be the key points concerning the relative change of spring wheat yield in the Canadian Prairies. This was in line with the found reports and finding in chapter 4.2.1 and 4.5 due to the coincidence of the found seasonal key points with previously defined key points for the growth of spring wheat. The final regression was performed using the defined set of variables. The rainfall in the month of May, June and July were also tested, but the P- and F-values found were far above 0.05 and therefore not significant. From the regression, it can be stated that not all of the defined seasonal key points were found to be important towards the spring wheat yield in the Canadian Prairies. In Table 11, the results of a regression of the relative change of spring wheat yield a function of the total rainfall of the previous season, the total snowfall before the sowing in the Canadian Prairies are shown.

Table 11: the results of the regression of the relative change of the spring wheat yield of the Canadian Prairies as a function the rainfall in the previous year and snowfall before sowing in the Canadian Prairies, with R2, P- and F-value and coefficients

R^2	0.38
Significant F value regression	3.9×10^{-6}
Snow P-value	0.007
Snow Coefficient	0.0003
Rainfall previous season P-value	4.8×10^{-6}
Rainfall previous season Coefficient	0.003

As can be seen from Table 11, the P-values of the chosen seasonal key points as independent variables are within the 0.95 confidence bound with values below 0.05, and can therefore be interpreted as statistically significant. The F-value of the regression is also below 0.05, indicating the significance of the regression as a whole. The regression has an R-squared of 0.38, which translates into the finding that more than a third of the change in spring wheat yield could be explained by using a set of season key points in the Canadian Prairies within a wheat year. All the coefficients of the independent variables found in the regression are in the same order of magnitude, confirming the relevance of the coefficients. The coefficient of the snowfall is again a factor 10 smaller than the coefficient of rainfall due to density differences.

Among the coefficients, a number of interesting results were found. The amount of total rainfall in the previous season and the snowfall before sowing both have a positive effect on the regression. The importance of an adequate amount of rainfall before the start of the season of spring wheat is a reoccurring import factor already defined in 4.2.1. The results from the linear regression are interpreted that the linear regression agrees with the notion that the amount of rainfall in the previous season and snowfall before sowing of importance to a successful yield, in line with the findings in 4.2.1, 4.2.2 and 4.5.

Based on the found results, a predictor for the spring wheat yield based on the arithmetic mean rainfall in the previous season and the snowfall before sowing in the Canadian Prairies could be made. The results from the regression analysis could be summarized in equation 9 as shown below:

$$PRCYSW_{CP} = 0.003 * R_{PS} + 0.003 * S \ (eq. 9)$$

With

$PRCYSW_{CP}$ = Predictor Relative Change Yield Spring Wheat Canadian Prairies, relative to linear trendline in megaton per hectare

R_{PS} = Rainfall previous season, sum of the arithmetic mean in millimetres

S = Snowfall before sowing, arithmetic mean depth in millimetres

Using the found equation, the comparison between the predictor and the actual difference of the spring wheat yield relative to the trend could be composed. In Figure 21, the comparison between the predictor and the actual change spring wheat yield relative to the trend is shown. As can be seen from Figure 21, the predictor for the change in spring wheat yield is again not a perfect fit, but does follow the same general movement of the actual difference. Differences can be seen in particular in years where the combination of key points had limited influence on the spring wheat yield, such as 1989/90, 2002/03 and 2013/14. In 1989 and 2002, North America and Canada were hit by an unusual drought, lowering wheat yield (FAO, 1990; FAO, 2003). In 2013, crop yields were helped by a long growing season due to mild temperatures and late winter weather (Minisrty of Alberta, 2014). The parameter of length of growing season as well as temperature has not been included in this research, but is found through the linear regression as being of importance to the production of spring wheat in the Canadian Prairies.

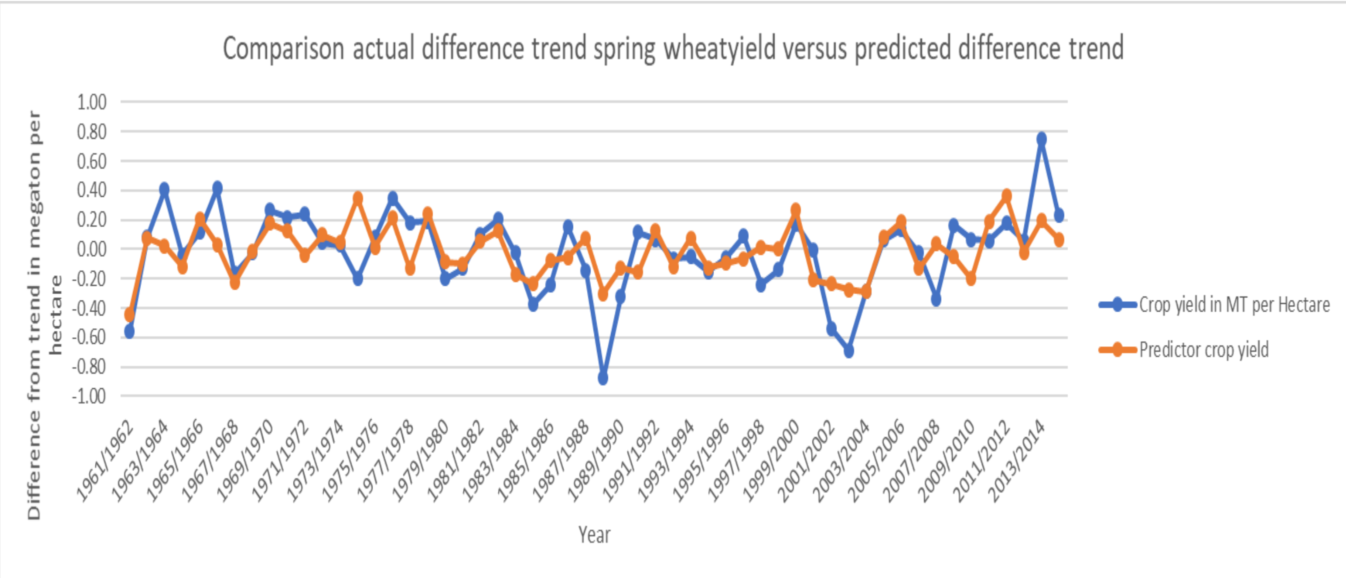
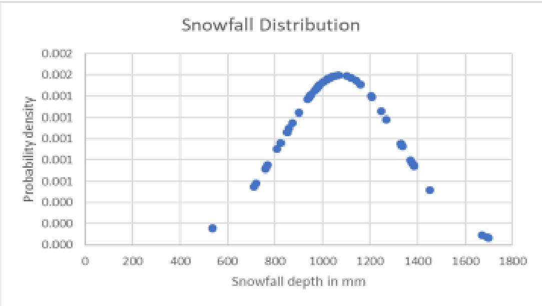
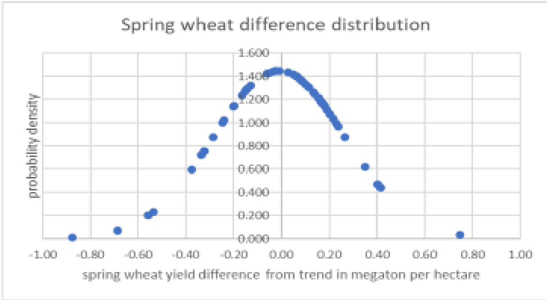
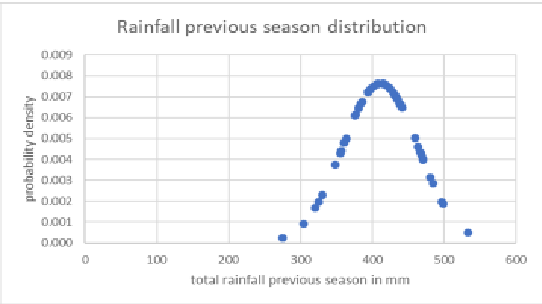


Figure 21: comparison between the actual difference of the actual spring wheat yield from the trend in the Canadian Prairies versus the predicted difference in spring wheat yield in the Canadian Prairies from the trend using the key point regression

As a final test for the regression, the rainfall and snowfall data, as well as yield differences are tested on skewness and kurtosis to determine if the data is normally distributed. In Figure 22, the probability distribution of the rainfall, snowfall and the yield difference from the trend of the spring wheat yield of the Canadian Prairie is shown. The skewness and kurtosis of all of the data were within the bounds as stated in 1.1. Therefore, the data was perceived as being normally distributed and the multivariate regression was subsequently accepted.







Snowfall	
Kurtosis	0.27
Skewness	0.61

Rainfall prev season	
Kurtosis	0.08
Skewness	-0.28

Spring wheat yield	
Kurtosis	1.77
Skewness	-0.63

Figure 22: the probability distribution of the difference in spring wheat yield from trend, total rainfall previous season and snowfall with tables containing the kurtosis and skewness of the yield, rainfall and snowfall data.

4.7 Binary Bayesian Predictive model

Following the results from chapter 4.6, a BBPM could be completed for ND and the Canadian Prairie. The BBPM was made to provide a probabilistic expectation following the regression formula as presented in 4.6 and using the same data sets. For the BBPM of ND, the relative change in spring wheat was assessed to have a conditional probability, while the rainfall in the previous season, snowfall before sowing and rainfall in the month of May have an unconditional probability. For the BBPM of the Canadian Prairies, the relative change in spring wheat was assessed to have a conditional probability, while the rainfall in the previous season, snowfall before sowing and rainfall in the month of May have an unconditional probability. To construct the BBPM, the seasonal key points and production were translated into binary. To translate the seasonal key points to binary, rainfall of snowfall above (1) or below (0) the arithmetic mean was used for both ND and the Canadian Prairies. For the yield, relative change above (1) or below (0) the trend was used to translate the yield to binary for both ND and the Canadian Prairies.

First, the BBPM for ND is presented. The formula for the probability for crop prediction is given below as equation 10 and the BBPM with the unconditional probability of the seasonal key points of the rainfall in the previous season, snowfall before sowing and rainfall in the month of May can be seen in Figure 23:

$$P\left(Y(ND)_{1,0}\right) = P\left(PS_{1,0}\right) * P\left(S_{1,0}\right) * P\left(M_{1,0}\right) * P\left(Y(ND)_{1,0} \middle| PS_{1,0}, S_{1,0}, M_{1,0}\right)\ (eq. 10)$$

With

$P\left(Y(ND)_{1,0}\right)$	= The probability of an above (1) or below (0) the trend change in spring wheat yield in ND occurring
$P\left(PS_{1,0}\right)$	= The probability of an above (1) or below (0) the mean amount of rainfall in the previous season occurring
$P\left(S_{1,0}\right)$	= The probability of an above (1) or below (0) the mean amount of Snowfall in January occurring
$P\left(M_{1,0}\right)$	= The probability of an above (1) or below (0) the mean amount of rainfall in May occurring
$P\left(Y(ND)_{1,0} \middle| PS_{1,0}, S_{1,0}, M_{1,0}\right)$	= The conditional probability of an above (1) or below (0) amount of the trend change in spring wheat yield in ND, conditional of the combination of above or below the mean rainfall in the previous season, snowfall before sowing and rainfall in the month of May occurring

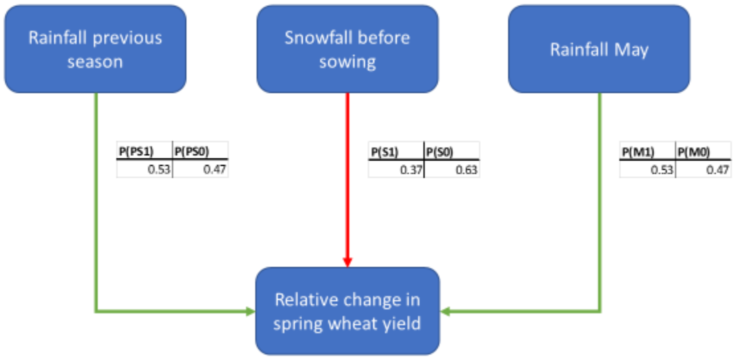


Figure 23: Binary Bayesian Probabilistic Model of the effect of the above or below the mean amount of rainfall of the seasonal key points of the previous season and the month of May, and the snowfall before sowing, with the unconditional probabilities corresponding to the seasonal key point based on historical occurrence between 1997 and 2015. The negative effect of the snowfall is indicated in red.

In Table 12, the conditional probability of the change in yield relative to the trend in spring wheat yield in ND is show. To determine all the conditional probabilities, all the combinations of above or below the mean amount of rainfall and snowfall for all the seasonal key points were calculated based on historical occurrence. The found provide an expectation based on historical data. Note that in one of the cases the conditional probability for yield was 0, indicating that the specific conditional probability did not occur between 1997 and 2015. An example is the conditional probability with Rainfall in the previous below mean (0), Snowfall above mean (1) and rainfall May below mean (0).

Table 12: The conditional probability of the of an above (1) or below (0) amount of the trend change in spring wheat yield (Y), conditional of the combination of an above or below the mean amount of rainfall in the previous season (PS), snowfall before sowing (S) and rainfall in the month of May (M).

PS	S	M	P(Y1)	P(Y0)
0	0	0	0.00	1.00
0	0	1	0.33	0.67
0	1	0	0.00	0.00
1	0	0	0.33	0.67
0	1	1	0.00	1.00
1	1	0	0.67	0.33
1	0	1	1.00	0.00
1	1	1	1.00	0.00

Next, the BBPM for the Canadian Prairies is presented. The formula for the probability for crop prediction is given below as equation 11 and the BBPM with the unconditional probability of the seasonal key points of the rainfall in the previous season and snowfall before sowing can be seen in Figure 24:

$$P\big(Y(CP)_{1,0}\big) = P\big(PS_{1,0}\big) * P\big(S_{1,0}\big) * P\big(Y(CP)_{1,0} \big| PS_{1,0}, S_{1,0}\big)\ (eq.11)$$

With

$P\big(Y(CP)_{1,0}\big)$	= The probability of an above (1) or below (0) the trend changes in spring wheat yield in the Canadian Prairies occurring
$P\big(PS_{1,0}\big)$	= The probability of an above (1) or below (0) the mean amount of rainfall in the previous season occurring
$P\big(S_{1,0}\big)$	= The probability of an above (1) or below (0) the mean amount of Snowfall in January occurring
$P\big(Y(CP)_{1,0} \big| PS_{1,0}, S_{1,0}\big)$	= The conditional probability of an above (1) or below (0) amount of the trend change in spring wheat yield in the Canadian Prairies, conditional of the combination of above or below the mean rainfall in the previous season and snowfall before sowing May occurring

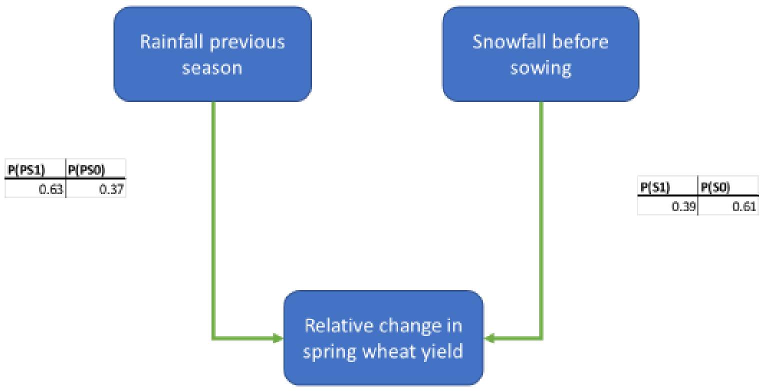


Figure 24: Binary Bayesian Probabilistic Model of the effect of the above or below the mean amount of rainfall of the seasonal key points of the previous season and the snowfall before sowing with the unconditional probabilities corresponding to the seasonal key point based on historical occurrence between 1961 and 2014.

In Table 13, the conditional probability of the change in yield relative to the trend in spring wheat yield in the Canadian Prairies is show. To determine all the conditional probabilities, all the combinations of above or below the mean amount of rainfall and snowfall for all the seasonal key points were calculated based on historical occurrence. The found provide an expectation based on historical data. Note that in none of the cases the conditional probability for yield was 0, indicating that all conditional probability did occur between 1961 and 2014.

Table 13: The conditional probability of the of an above (1) or below (0) amount of the trend change in spring wheat yield (Y), conditional of the combination of an above or below the mean amount of rainfall in the previous season (PS) and snowfall before sowing (S).

PS	S	P(Y1)	P(Y0)
0	0	0.57	0.43
0	1	0.47	0.53
1	0	0.33	0.67
1	1	0.67	0.33

As an example, to interpret the results, assume that the probability of an above average change in spring wheat yield for ND needs to be determined at the start of a wheat season with the assumption that the rainfall in all of the seasonal key points will be above the mean. Equation 8 takes the shape of:

$$P(Y(ND)_1) = P(PS_1) * P(S_1) * P(M_1) * P(Y(ND)_1|PS_1, S_1, M_1) \ (eq. 12)$$

The probability of the event actually occurring then becomes:

$$0.53 * 0.37 * 0.53 * 1 = 0.102$$

The probability of 0.102 is the multiplication of the probability of all the seasonal key points being above the mean multiplied by the conditional probability of the production change being above the trend.   All possible combinations of probabilities can be calculated to determine the probability of an event occurring. Note that this possibility is a blind possibility, as it is one of the options at the start of the season without knowing if the weather in the seasonal key points is above or below average. If for instance the probability of an above average spring wheat yield in ND is determined in April, with the knowledge that the rainfall of the previous season and snowfall before sowing were above average, equation 8 takes the shape of:

$$P(Y(ND)_1) = P(PS_1) * P(S_1) * P(M_1) * P(Y(ND)_1|PS_1, S_1, M_1) \ (eq. 13)$$

With $P(PS_1)$ and $P(S_1)$ being 1 due to the fact that the rainfall of the previous season and snowfall before sowing were above average. The probability of the event actually occurring then becomes

$$1 * 1 * 0.53 * 1 = 0.53$$

Part II: Cocoa

5. Continuation introduction cocoa

5.1 Research Objective

Summarizing the introduction for cocoa, the objective for this research was to link weather patterns to price changes of the agricultural commodity of cocoa. First, an understanding of the dynamics and drivers of the prices of cocoa is formed. Second, the relative importance of the cocoa production of Ghana and Ivory Coast towards the world cocoa production is assessed. Next, the hypothesis of the importance of rainfall in a specific time period towards the production and price change of cocoa is tested both quantitatively and qualitatively. The goal is to keep the research, the method and results as fully accessible for both people with an engineering or financial background and provide a method to understand and predict cocoa production subsequently assess the impact towards cocoa prices.

5.2 Research Question

Following the set objective, the following research question concerning price change of cocoa can be stipulated:

- *What are the main drivers behind the prices of cocoa?*

For price change, the following sub questions can be stipulated:

- *In what years do the largest relative increases in price occur, taking inflation of the crop specific currency into account?*
- *Is it possible to determine the causes of the price increases and cluster these effects, such as shortage in supply due to natural effects?*
- *What is the relative strength of the different drivers on market prices?*

The following research questions can be stipulated concerning production:

- *How strong is the dependency of the world cocoa production on the cocoa production in Ivory Coast and Ghana?*

For production, the following sub questions can be stipulated:

- *What is the yearly world production of cocoa?*
- *What is the yearly production of cocoa of Ivory Coast and Ghana?*
- *What part of the variance in the world production of cocoa can be explained by the cocoa production in Ivory Coast and Ghana?*

For the proof of concept of weather dependency, the following research question can be stated

- *Is it possible to prove the importance of the weather in a specific time period in Ivory Coast towards the world production and price of cocoa?*

For the proof of concept, the following sub questions can be stipulated:

- *What are the crop weather related specific requirements of cocoa to successfully produce cocoa?*
- *From reports, what are the indicated most important time periods for the production of cocoa?*
- *What part of the variance of the world production of cocoa can be explained by using the found important time periods?*
- *What part of the variance of the price of cocoa can be explained by using the found important time periods?*

5.3 Geographical setting

Ivory Coast

As mentioned in the introduction, Ivory Coast (IC) is the number one cocoa producing country of the world (ICCO, 2010). Figure 25 shows six of the world's largest cocoa producing areas of the world covering 80% of the world production of cocoa, with percentage of world production per area. While IC has significant importance for the cocoa market due to the large cocoa production, limited data is available from weather stations and production data. Political instability and civil war has had severe effects on IC on both an economic sense, as well as maintenance of the weather stations (UNDP, 2011).

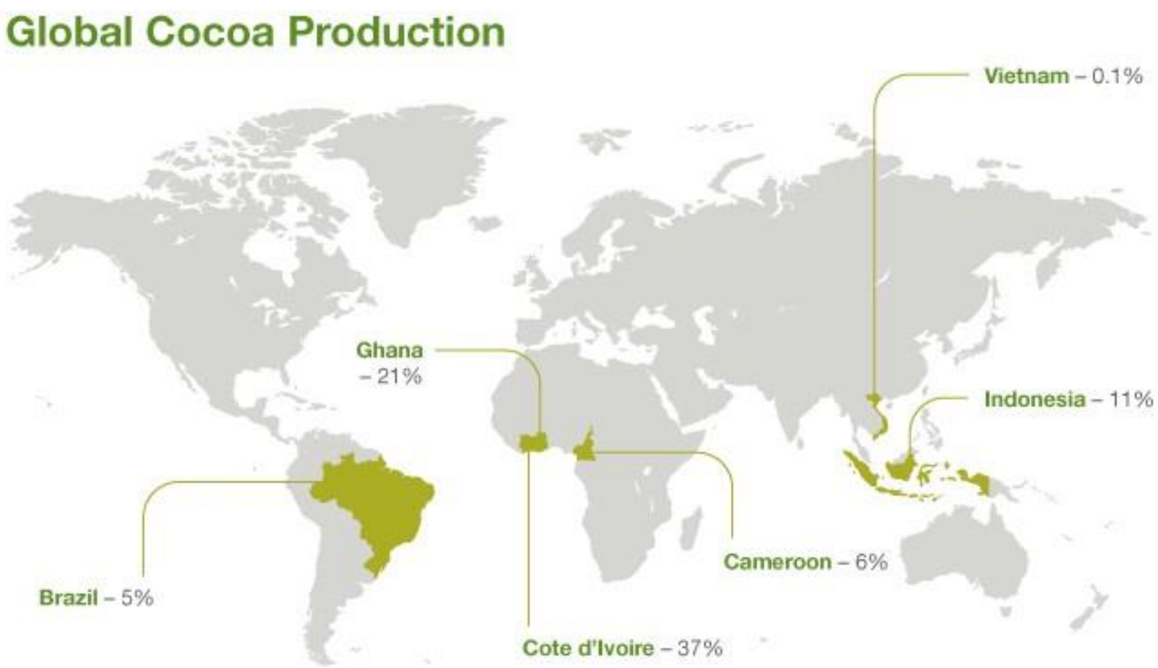


Figure 25: The largest cocoa producing areas of the world with percentages of world production (Cargill, 2016)

From the website of "Our Africa", the following quote can be made concerning the climate in IC: " *Near to the equator, Ivory Coast has a tropical climate with consistently high temperatures all year round. In the commercial capital of Abidjan on the coast, temperatures usually fall between 22°C and 32°C. The coastal region has the highest amount of rainfall, receiving 200-300cm on average. Much of the rain comes between May to July. Then there is a dry spell for a couple of months, followed by another shorter rainy season in October and November. The long dry season is from December to April. In the central forest region of Ivory Coast, it is hot, wet and humid for much of the year. Humidity is also high across the southern part of the savannah. Across the savannah in the north, a parching and dusty trade wind known as the Harmattan blows from the north-east from December to March*" (Our Africa, 2017).

Translating the area specifics to the production of cacao in IC, the following quote could be found concerning the weather on page 8 of the book "Cocoa: a guide to trade practices" from the International Trade Centre: "C*ountries with pronounced dry and wet seasons normally show two harvests a year, a main crop and a mid-crop. The relative sizes of these crops depend on how long the wet seasons last. A pronounced drought, or a long cool, rainy season, will have a major impact on the total tonnage produced – and on prices"* (ITC, 2001).

The cocoa production in IC has a main crop season and a mid-crop season, with the main crop season accounting for 90% of the cocoa production (ITC, 2001). The cocoa bean produced in IC is of the Forastero variety, which are "*basic in terms of flavour. Their colour is akin to the world standard (Ghana), but they are generally less well fermented and slightly higher in acidity. Prone to certain defects, including mould, germination and insect-damage, the beans also tend to be somewhat on the small side"* (ITC, 2001).

6. Materials and Methodology

In this chapter, the material and methodology used in the research are discussed. In the materials section, all the used data for the research is discussed. Next, the methods of processing the data to achieve the results are discussed. In the method chapter, a general overview of the whole process of translating the data into results is given.

First, the price changes of cocoa were assessed. From the reports, possible seasonal key points were defined qualitatively. For this research, cocoa related reports from the ICCO were used. For cocoa, these are the monthly and yearly reports from the ICCO obtained from the website of the ICCO (ICCO, 2017). The used reports are introduced in section of materials, the method of using the reports is introduced in methods, a general introduction to the ICCO can be found in appendix I.

The cocoa production of Ghana and IC were used to investigate how the countries influence the world cocoa production. The notion of a critical market share in cacao production was researched, to determine how much contribution to world production of cocoa is required to make an impact on world production noticeable. The materials and method of the cocoa production assessment are introduced further on in this chapter.

Due to the high contribution of IC to the world cocoa production, the country of IC was chosen to research the impact of local rainfall patterns on global cocoa yield. The rainfall data was collected from 8 stations in IC, which are further discussed in the section of materials. The source and method of the cocoa production are introduced further on in this chapter.

The impact of the rainfall in the qualitatively defined seasonal key points as found from the reports from the ICCO on the world cocoa production was researched. The goal is to assess rainfall data from IC and research if seasonal key points could be quantified concerning the contribution to the crop production. As part of the objective, the level of explained variance of the production of cocoa as well as probabilistic values of the key points in a season were assessed to verify if the influence of the seasonal key points on the cocoa production is not random. The method is elaborated on further on in this chapter.

As a final step, a binary Bayesian probabilistic model (BBPM) was created using the defined seasonal key points and production data. Using the BBPM, the probabilities and the effect of the occurrence of the seasonal key points towards relative change in world cocoa production could be defined. The full process is described further on in this chapter.

6.1 Materials

In this chapter, all the materials used for the research are discussed. All the details of the materials are discussed, such as length of used data series, limitations due to data availability and whether data is of a daily, monthly or yearly scale.

6.1.1 Market data cocoa

The market data used to determine the price changes and the extreme increases and decreases was the daily averaged cocoa futures prices of the ICCO via Index Mundi over a period of 31 years between 1986 to 2016 (ICCO, 2016). The prices were provided in the form of future prices in US Dollars, meaning that the price were the quoted prices of future contracts. A future contract is a contract with an obligation to buy a commodity such as cocoa in the future for a set price. For cocoa, future contracts are available for March, May, July, September and December (ICE, 2017). The prices for the futures were provided on a monthly average scale. For this research, inflation is filtered out to produce the "real" or "clean" price changes (Malliaris, 2006), which will be further discussed in chapter 6.2.1. The future prices of cocoa were in US dollars and therefore the inflation of the US dollar was taken to correct the future prices for inflation. For inflation, the inflation numbers of the US dollar from the Bureau of Labour and Statistics of the Department of Labour were used between 1914 to 2016 (BLS, 2016). The data concerning inflation was provided in monthly averages and in yearly averages.

6.1.2 Cocoa reports

From the ICCO, the annual and monthly reports concerning cocoa production, prices and consumption were used. The monthly reports give a detailed description of the monthly price movements for cocoa and also provide a suspected motive for the price change according to the ICCO. In the annual reports, a complete overview of the cocoa world price change, as well as production and consumption figures are provided. The annual reports used are from 2007 (ICCO, 2008), 2008 (ICCO, 2009), 2009 (ICCO, 2010), 2010 (ICCO, 2011), 2011 (ICCO, 2012), 2012 (ICCO, 2013), 2013 (ICCO, 2014), 2014 (ICCO, 2015), and 2015 (ICCO, 2016), as well as the market overview from 2010 (ICCO, 2010) and 2012 (ICCO, 2012). Monthly reports from the ICCO between 2008 and 2015 were also used. All the monthly reports were obtained from the website of the ICCO in January 2017 (ICCO, 2017). Next to reports from the ICCO, reports and articles from the World Bank and WCF were used too (WCF, 2012; The World Bank, 2001; The World Bank, 2007). In appendix II, an example of a monthly report of the ICCO is given. Next to the monthly and yearly reports, the ICCO also has reports with crop specific information concerning the specifics of cultivating cocoa (ICCO, 2016).

6.1.3 Cocoa production

The world production of cocoa was made available by courtesy of the ICCO. World production data for cocoa was available from 1960 to 2015, with no data missing. In the same document for the cocoa production, data concerning grindings, surplus and deficit for cocoa was also available (ICCO, 2016). The unit for the production data was in thousands of tons or megaton (MT). For this research, all cocoa is assumed to be fungible as a first approximation, not taking quality differences into account.

For Ivory Coast and Ghana, data was available for the cocoa production as a country on a crop year scale. From the annual reports of the ICCO as mentioned in 6.1.2, production numbers could be abstracted from IC and Ghana between 2002 and 2015. In the dataset, no missing data was found. The unit of the yield was in thousands of tons or megaton (MT).

6.1.4 Weather station data

The data from the weather stations were obtained via the Koninklijk Nederlands Meteorologisch Instituut (KNMI) Climate Explorer web application. The KNMI Climate Explorer application (KNMI, 2017) is a collection of data from weather stations all over the world. Daily data from the KNMI was used as opposed to direct monthly data, due to the fact that no monthly data was available for IC. The selected weather stations were dictated by two main conditions, being availability and the area where cocoa is produced. In Figure 26, all the available weather stations for IC are presented. As can be seen in Figure 26, only 14 stations with daily data were available for IC. Of the 14 stations, only 8 where in the zone where cocoa is produced. Therefore, 8 stations were selected for cocoa in IC.

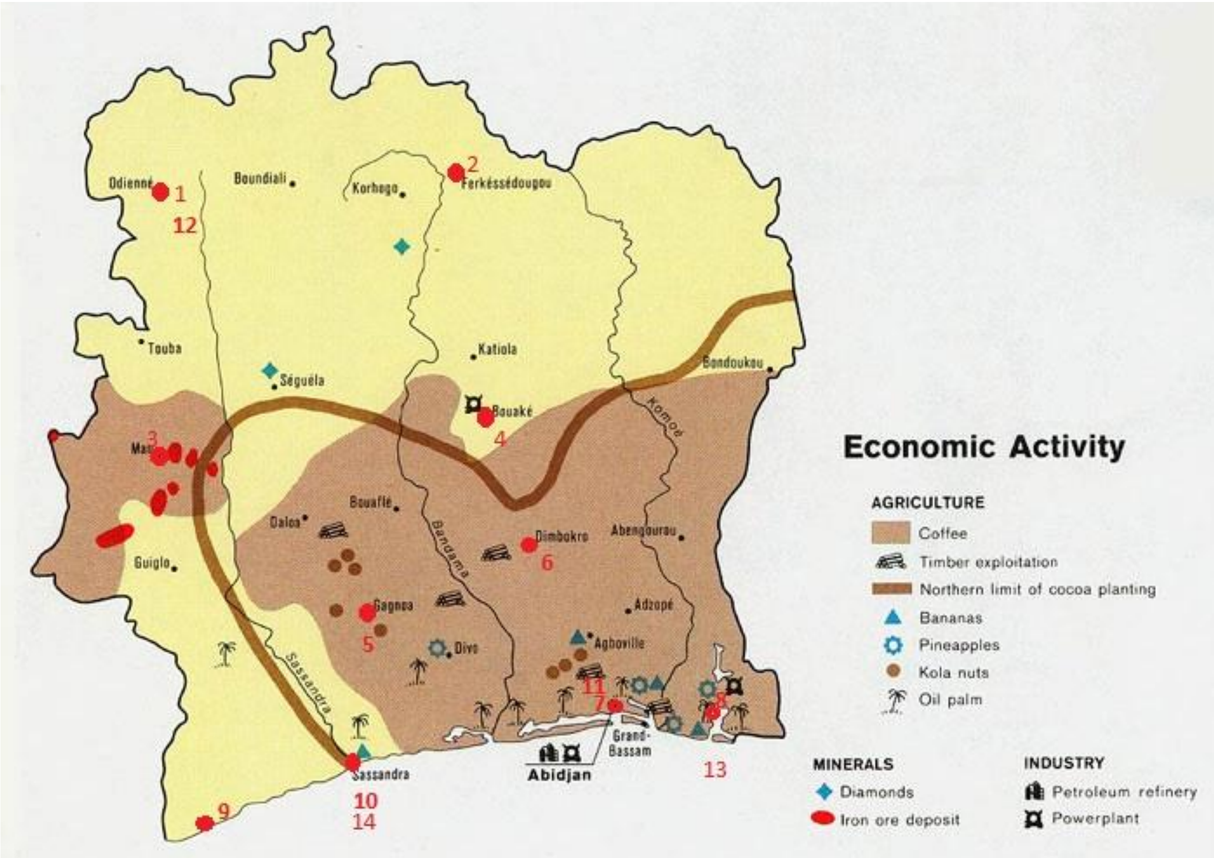


Figure 26: Available weather stations in Ivory Coast indicated with numbered red dots, with zone for the production of cocoa indicated (Aregheore, 2009)

The rainfall data was collected from the stations in Gagnoa, Abidjan, Adiake, Sassandra and Dimbokro. For Abidjan, Adiake and Sassandra, data was collected from 2 different weather stations. The double rainfall stations almost complemented each other, with data from one weather station ending where the recording of data started at another station. In months were both stations provided data, the average between the two stations was used.

Effectively, due to the double stations, data from 5 stations was collected. Data was available between 1945 and 2003, but with missing data especially prevalent from 1999 onward. This data coincides with the first Coup d'état in IC and the start of the civil unrest in IC (UNDP, 2011). 6 of the weather stations were active from 1945 up to 1980, while 2 stations (Adiake and Sassandra) were active from 1980 up to 2003. After 2003, no active weather stations were found via the KNMI Climate Explorer application. The unit of all data collected was in millimetres. All data was retrieved in January 2017 from the website of the KNMI Climate Explorer application (KNMI, 2017).

6.2 Methods

In this chapter, the specifics concerning the used methods for the data processing are discussed. Theoretical background and the specifics of which techniques are used in this research are also described. In Figure 27, the flowchart of the research is presented.

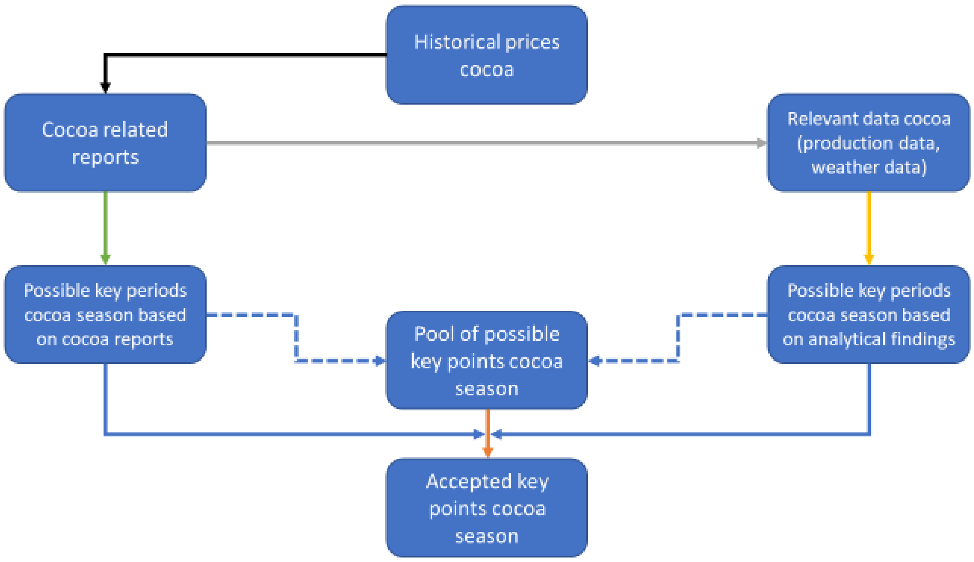


Figure 27: flowchart of the course of research, explanation given in text

To start of the research, the historical prices of cocoa were assessed. In this step, the goal is to find the years in which the cocoa price had the largest percental relative increase or decrease between 1986 and 2016. Monthly price changes were assessed to define the average price movement of cocoa on a monthly scale within one crop year by taking the average monthly price change of all the crop years. All the prices were corrected to take inflation into account. This step is represented by the point "historical prices cocoa" in Figure 27. The results from this step was used as a starting point for the next part of the research. The full process is described in 6.2.1.

Next, the found price increases and decreases were assessed using cocoa related reports from the ICCO, World bank and WCF. The results from the previous step concerning the historical prices of cocoa were used as a first indicator to determine in which years and months reports should be assessed in particular. The goal is to use the information of the largest price increases and decreases in combination with reports from the years and months in which the price changes occurred to identify the reason for the price change. This step is represented in Figure 27 by the solid black arrow between "Historical prices cocoa" and "Cocoa related reports".

In the same step of "Cocoa related reports", crop specifics as defined by the ICCO as well as climate specifics of IC were used to define key points in a cocoa growing season. Crop specifics such as water demands, growing cycle, harvest times and humidity were researched. Area specifics for IC such as climate, seasonal changes in weather and temperatures were assessed at the same time. The combination of cocoa related reports from the ICCO, as well as crop and climate specifics of IC were used to qualitatively define seasonal key points in the growing season. The reports and historical prices were used to define the dynamics of the cocoa market. This step is represented in Figure 27 as the solid green arrow between "cocoa related reports" and "Possible key periods cocoa season based on cocoa reports". The full process is described in 6.2.2.

The outcome of the cocoa related reports was also used as an input for the evaluation of production and weather data. This step is represented by the grey arrow from “cocoa related reports” to “relevant data cocoa” in Figure 5. The goal was to use the information about the largest price changes in combination with the qualitative outcome of the reports to determine if the price changes can also be explained using world cocoa production and rainfall data. For the production data, the world production of cocoa was used. Due to the fact that cocoa production is always increasing (Edwin & Masters, 2005), the difference in production as difference from the trend was taken. The process of using relative difference of cocoa production from the trend is described in 6.2.3.

Within the “relevant data cocoa”, local production of IC and Ghana and world production was also assessed to determine the dependency of the world cocoa production on the local production. A multivariate linear regression between world cocoa production and cocoa production from IC and Ghana was performed to determine whether a dependency between world cocoa production and local cocoa production as mentioned in the ICCO reports can be confirmed. The full process is described in 6.2.4.

In the “relevant data cocoa” step, data from weather stations was also pre-processed for further use. First, the used weather station data was filtered for missing data. Second, the daily rainfall data per station is summed up to give a monthly and a yearly total. The total for a year was the total per cocoa year as defined by the ICCO between October and September. Third, the used weather stations are assessed on correlation to determine how strongly correlated the stations were to justify the use of an arithmetic mean of the stations. If a weather station was insufficiently correlated, the weather station would be rejected from the research until all weather stations left were sufficiently correlated to justify the use of an arithmetic mean. The correlation between weather stations was tested using the average rainfall on a monthly basis for a full year per weather station. Finally, the arithmetic mean of the weather stations was taken of the average monthly and yearly rainfall. The full process is described in 6.2.5.

To define key points in a cocoa season in a quantitative sense, a linear regression between the relative change in world cocoa production and the arithmetic mean of the weather stations was performed. This step is represented by the solid yellow arrow between “Relevant data cocoa” and “Possible key periods cocoa season based on analytical findings” in Figure 27. To narrow down the variables used in the linear regression, the qualitative results from the cocoa related reports were also used. This step is the results from the feedback from the reports as represented by the grey arrow in Figure 27. The outcome of the linear regression and the used variables are tested on probabilistic properties to determine if the finding were significant. The full process is described 6.2.6.

Next, both results from the reports and the analytical findings were combined in a pool of possible key points of a cocoa growing season. This step is represented by the dotted blue arrow in Figure 27. Subsequently, the found key points were validated as actual real key points in a cocoa growing season by testing the possible key points on both a qualitative basis by the indicated key points from the reports from the ICCO and on a quantitative basis using the results from the linear regression. This step is represented by the solid orange arrow in Figure 27, while the qualitative and quantitative test is represented by the solid blue arrow in Figure 27. The full process is also described in 6.2.6.

As a final step, a binary probabilistic Bayesian model was applied to the confirmed seasonal key points to assess the conditional probability of a higher or lower cocoa production by using the seasonal key points. In this step, the probability of a above or below the trend cocoa production was made conditional based on the combination of a above or below average rainfall in one of the seasonal key points or a combination of seasonal key points. The full process is described in 6.2.7.

6.2.1 Historical prices of cocoa

For the historical prices of cocoa, future contract prices of cocoa were collected as defined in 6.1.1. In appendix IV, a general introduction to prices and commodities is given as background information for the terms of commodities and future contracts. The future prices of cocoa were quoted in dollars, but the future prices cannot be used directly in this research. Because of inflation, a US dollar in 1960 does not have the same value as a dollar in 2015. Inflation or deflation are defined as a general, continuous increase of decrease in prices, causing a reduction or an increase in the value of money (McIntosh, 2013). To make the price change of for instance 1960 usable in the same manner as in 2015, the future prices were first translated into relative change. In this research, cocoa price changes were defined as relative price changes. To translate the crop price data to relative price change of cocoa, the following formula was used:

$$RPC = {(X_n - X_{n-1})}/{X_{n-1}} * 100 \quad (eq. 14)$$

With:

RPC	=	Relative price change in %
X_n	=	the future price of month n in US dollars
X_{n-1}	=	the future price of month n-1 in US dollars

Relative price change was determined on both a monthly and yearly scale. For cocoa, yearly relative price change is determined from October to September, to coincide with the cocoa production year as defined by the ICCO (ICCO, 2016). The practice of taking the average of the year is in line with the practice of the average price determination of the ICCO to exclude price changes with a short time span such as price shocks and speculation, and determine the general trend of the price within one year (ICCO, 2016). By using the price change in a crop year, the effect of the crop production on the prices of the crop year can be obtained. For 2000 for instance, equation 11 takes the following shape:

$$RPC\ (2000/01) = {\Delta\left(\frac{2000}{01}\right)}/{X_{2000}} * 100\ (eq. 15)$$

With:

$RPC\ (2000/01)$	=	Relative price change for cocoa year 2000/01
$\Delta(2000/01)$	=	Difference between future price October 2000 and September 2001
X_{2000}	=	Future price of cocoa in October 2000

By changing the prices to relative change, the prices could be corrected with inflation, which is also a relative change. By subtracting the inflation, the goal was to define the “clean” price change in a month or year. The correction of the relative change of the prices of cocoa by using inflation is an accepted method of correcting prices (Malliaris, 2006).

Inflation was therefore deducted from the relative price change to produce the "clean" value for the relative price change. If the relative price change of is 4% in a year and inflation in the same year is also 4%, the "clean" relative price change is equal to 0%. Inflation is already given in a relative term of percentage on a yearly and monthly basis provided by the BLS (BLS, 2016), and could be directly deducted from the relative price change as found in equation 11. Inflation on a yearly scale was used according to the specific length of the crop year for cocoa. The used formula is shown below:

$$RPC^* = RPC - I \; (eq. 16)$$

With:

RPC^*	=	the clean value of the relative price increase in %
RPC	=	the relative price change in % as defined in (1)
I	=	the inflation for the assessed month/year in %

As a final step, the price increases and decreases were linked to their respective years and ranked from high to low.

6.2.2 Cocoa related reports: price changes

In this section, the use of reports from the ICCO is described. Reports from the ICCO were used to define crop specifics key points for cocoa. Reports from the ICCO are also used to determine if weather had an effect on the found price changes and if key points in a season related to weather can subsequently be defined using the reports.

6.2.2.1 Crop specific key points

To determine the crop specific needs for the growth of cocoa, a review is conducted on available literature about the crop specifics for cocoa. For cocoa, the website of the ICCO was used to obtain specifics concerning the demand for growing cocoa. The specifics provided information about the maximum and minimum amount of precipitation required, as well as length of the growing season and events during a growing season that could affect crop yield for cocoa (ICCO, 2016). The key periods of the season as defined by the crop specifics were later used to define the seasonal key points for the crop yield of both cocoa.

6.2.2.2 Cocoa price change related reports

After the price changes for cocoa are defined, a review of the reports of the ICCO was performed to determine whether the weather had any influence on the relative price change of cocoa. For this research, the top 5 price decreases and increases were highlighted, as well as the all the price changes between 2002 and 2015.

Production of cocoa can be affected by different events, such as lower crop yields due to weather events, but also trade embargos, export stops and civil unrest. For all of the price changes, the reports that were available for cocoa for that year were assessed to define the reason, or “driver”, related to the price change. The yearly and monthly reports were scanned on a specific link back to weather using words as “weather”, “adverse”, “negative effect”, “unfavourable” specifically correlating weather to production decrease and price increase. For cocoa, ICCO reports were used as well as reports from the world bank (The World Bank, 2007) (The World Bank, 2001), the Food and Agricultural Organization (FAO) (FAO, 2000; FAO, 1988; FAO, 1991) ,and the World Cocoa Foundation (WCF, 2012) for weather related price changes.

6.2.3 Relevant data cocoa: production changes

As mentioned in the introduction, cocoa production is constantly increasing. Reasons for the constant production increase are for instance more planting, but also crop improvement or better agricultural practices (Edwin & Masters, 2005). As a consequence, cacao production in different years could not be compared directly to each other. A specific amount of cocoa production in for instance 1980 could be a very high performing crop year then, while the same production in 2000 would be medium or badly performing crop year.

A linear trendline of the production was therefore used as a first approximation in this research. To determine whether a cocoa year is underperforming, the difference from the linear trendline is used to determine how a cocoa year was performing. As this research was a first trial, no other method than a linear trendline was used for the production. The available data for cocoa world production is plotted using Windows Excel, after which a trendline is added to the plot.

The dispersion of the residuals was used to visually determine the validity of the use of a linear regression (Stattrek, 2017). The residuals should appear to be dispersed randomly without a clear parabolic shape or heteroscedasticity to enforce the use of a linear approximation. Crop production under the trend will be defined as below expectation, while crop production above the trend will be defined as above expectation. The resulting outcome of the difference between the trend line is used in further steps of the research. For all steps, Windows Excel was used.

6.2.4 Relevant data cocoa: influence local production cocoa on world production cocoa

To correlate the world cocoa production to local cocoa production, a multivariate linear regression of yearly world cocoa production as a function of the yearly cocoa production in IC and Ghana was performed. IC and Ghana were here defined as local producing areas for cocoa and independent variables, while the world cocoa production was used as dependent variable. The regression for cocoa was done using annual production data from 2002 to 2015 using annual reports from the ICCO as defined in the materials section. The outcome of the multivariate linear regression was tested on P-values, F-values, residuals to validate the linear regression. Finally, the R^2 value was used to define the explained variance of the regression and to determine the dependency of the world cocoa production on the cocoa production of IC and Ghana. For all steps, Windows Excel was used.

6.2.5 Relevant data cocoa: pre-processing weather station data

To translate the weather stations data from IC into usable data for this research, a number of processing steps were taken. For this research, the weather stations were first checked for missing data. Missing data was given a zero as value concerning precipitation. If more than 5 days of a month were missing, the month was not used in further steps. In the dataset, the year of 1999 and 1966 were rejected due to missing data. Next, rainfall per month and per cocoa year were calculated by summing up all the recorded rainfall per month or cocoa year.

Finally, the average rainfall in a specific month was determined by calculating the average rainfall of all specific months for a specific weather station. All calculations were done using Windows Excel. For instance, the average rainfall of the month of April for the Abidjan station was calculated by taking the average of all the recorded rainfall of all the months of April for the Abidjan station.

The correlation between stations was tested next, to check if the average rainfall of weather stations in IC were correlated enough to justify the simplification of taking the arithmetic mean of all the weather stations. As a demand for the correlation index, correlation between all stations was required to be above 65% to ensure a “good” correlation conform the Pearson correlation coefficient (Pearson, 1895) to ensure that the rainfall in the assessed area was relatively equal at all assessed points and that therefore the area could be simplified to one area with an equally dispersed amount of rain. If the demand of 65% or above was not met, weather stations with the worst correlation to all other stations were rejected until all stations had a correlation above 65%.

6.2.6 Possible key periods cocoa season based on analytical findings: linear regression

To quantify the seasonal key points from the ICCO reports, a regression of the difference from the trend for cocoa production as a function of the rainfall of all the months that have an effect on the cocoa year was done. For IC, the arithmetic mean of the rainfall data per month of a cocoa year were used as independent variables. From the regression, the month with the highest P-values was rejected and a new regression was done until all P-values of the used months were within a 0.95 confidence interval and the F value of the regression in general is also within a 0.95 confidence interval. As mentioned in 1.1, a confidence interval of 0.95 excludes randomness of the variables. Reports were also used to make the optimization process go faster by using months that are found to be of importance based on crop specific ICCO reports and climate and area specifics. The iterative process is repeated until a combination of variables is found that explain the highest amount of variance of the change in crop production while the significance levels of both the independent and dependent variables were below the set level of 0.05. However, before the results are accepted, the results are first checked with reports from the ICCO to exclude spurious results (Burns, 2017)

The found seasonal key points are checked with the reports from the ICCO to check if the found time periods are both statistically accepted and in accordance with the reports as provided by the ICCO. The key of this step is the combination between reports and analytics. The goal is to ensure that the linear regression performed between difference from the trendline as a function of a key time period in the cocoa growing season is both statistically sound while also being able to be explained by reports and crop specifics. By combining both qualitative and quantitative reasoning, the method should exclude spurious results of a regression and only focus on variables that are related to the growing of cocoa.

6.2.7 Binary Bayesian probabilistic model

By using both the data relative change in production from 6.2.3 and the results from the specified seasonal key points from 6.2.6, a binary Bayesian probabilistic model could be formed. The BBPM is a continuation of the results of the seasonal key point regression outcome of 6.2.6, connecting a probability to the seasonal key points of the regression. For the BBPM, the relative change in cocoa production was used in combination with the rainfall data from IC of the seasonal key points. Both production and seasonal key points were first translated into binary. For the cocoa production, production above the trendline was defined as a 1, while the production below the trend was defined as a 0. For the rainfall in the seasonal key points, rainfall above the average monthly arithmetic mean in a specific month was defined as a 1, rainfall below the arithmetic mean rainfall in a specific month was defined as a 0.

Because of the use of the seasonal key points as independent variables and the production change as dependent variable in the linear regression, the probability of production change was defined as conditional and dependent on the seasonal key points (Bernardo, 2000). To define the probability of an above or below average cocoa production, all the combinations of the independent variables were tested towards crop production. The probability and conditional probability were determined by using the historical occurrence of the events. For instance, if 5 of the 20 years assessed have an above trend cocoa production change, an unconditional probability of 0.25 was assigned to the probability that the relative cocoa production change will be above trend and an 0.75 change that the relative cocoa production change will be below trend.

7. Results

In this chapter, the results of the research are presented. First the results historical prices of cocoa are presented. Both the yearly price changes as well as the average price change within one crop year are shown. Second, the price changes on both a monthly and yearly scale are explained using the information from the ICCO reports and other reports, as well as the found seasonal key points from the reports and the crop specifics as found from the ICCO. Third, production is assessed, with an assessment of the trend of the cocoa production as well as the dependency of the world cocoa production on the cocoa production of IC and Ghana. Fourth, the results from the processed rainfall data of the used weather stations as well as the correlation index of the used weather stations is described. Finally, the outcome of the key point regression and subsequent BBPM are presented.

7.1 Historical prices of cocoa

First, the average monthly price change of cocoa as found between 1985 and 2015 is presented. The monthly price changes can be used as first indicators to determine the important months for cocoa. Months with high price change could indicate important events concerning cocoa and should therefore important to assess. The average monthly price changes of cocoa, corrected for inflation and in percentage are presented in Figure 28. As can be seen from Figure 28, price changes tend to increase in December, February, June and July, with the largest increase in July. The largest price decrease was noted in October, followed by April, May, January, August and November. In the review of the reports of the ICCO in chapter 7.2, possible reasons for price changes are discussed.

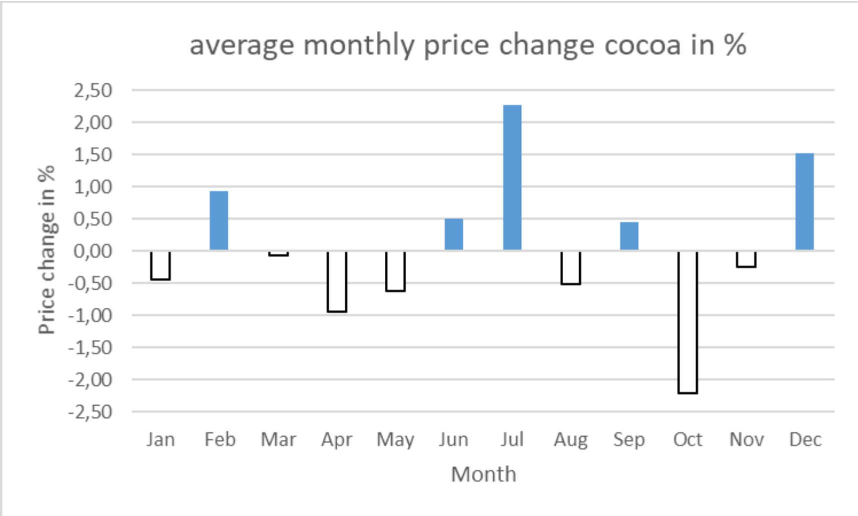


Figure 28: the average monthly price change of cocoa between cocoa year 1985 and 2015

When assessing the average standard deviation on a monthly basis, or volatility, different pattern can be found. In Figure 29, the average monthly volatility of cocoa in percentage is given. It can be noted that throughout almost all the months, volatility is above 5%, with the largest volatility in September. It can also be noted that the in all the months the volatility of the cocoa price is larger than the average price increase or decrease.

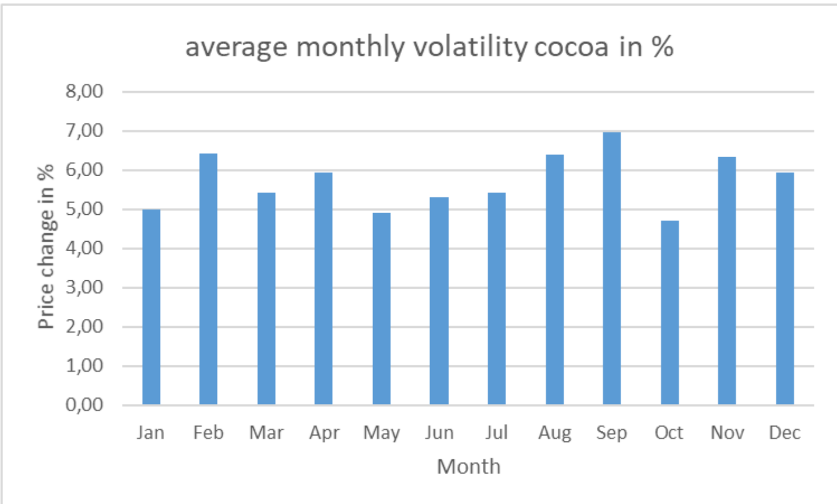


Figure 29: the average monthly volatility of the price change of cocoa between cocoa year 1985 and 2015

Next, the price change on a yearly scale was assessed. In Table 14, the results of the price changes are shown, both ranked from high to low concerning percental price change, as well as chronologically. From the price changes, it can be noted that the from the top 5 price increases, 4 years were after 2001. This finding is in line with the statement in the introduction that from 2001 onward, the cocoa industry is facing systematic shortcomings in supply and subsequently larger price volatility. In chapter 7.2, the yearly price changes are assessed using the reports from the ICCO.

Table 14: The price change in %, corrected for inflation, of cocoa in a cocoa year (Oct to Sept) from crop year 1985 to 2015, ranked according largest change (left) and chronological (right), with increase in green and decrease in orange.

year ranke	year ranked	change				change %
2001	**2002**	**98.19**		1985	1986	-6.35
2008	**2009**	**40.30**		1986	1987	-7.70
2007	**2008**	**35.61**		1987	1988	-39.33
2006	**2007**	**24.25**		1988	1989	-17.38
1989	**1990**	**23.45**		**1989**	**1990**	**23.45**
1996	**1997**	**17.19**		1990	1991	-6.93
2013	**2014**	**16.36**		1991	1992	-17.22
1992	**1993**	**14.90**		**1992**	**1993**	**14.90**
2000	**2001**	**14.12**		**1993**	**1994**	**12.57**
1993	**1994**	**12.57**		1994	1995	-6.44
2014	2015	5.63		1995	1996	2.83
2012	2013	4.69		**1996**	**1997**	**17.19**
2005	2006	4.54		1997	1998	-4.77
1995	1996	2.83		1998	1999	-37.83
2003	2004	1.31		1999	2000	-16.99
2004	2005	-2.03		**2000**	**2001**	**14.12**
2011	2012	-4.33		**2001**	**2002**	**98.19**
1997	1998	-4.77		2002	2003	-28.47
2010	2011	-5.03		2003	2004	1.31
2015	2016	-5.75		2004	2005	-2.03
1985	1986	-6.35		2005	2006	4.54
1994	1995	-6.44		**2006**	**2007**	**24.25**
1990	1991	-6.93		**2007**	**2008**	**35.61**
1986	1987	-7.70		**2008**	**2009**	**40.30**
2009	2010	-16.35		2009	2010	-16.35
1999	2000	-16.99		2010	2011	-5.03
1991	1992	-17.22		2011	2012	-4.33
1988	1989	-17.38		2012	2013	4.69
2002	2003	-28.47		**2013**	**2014**	**16.36**
1998	1999	-37.83		2014	2015	5.63
1987	1988	-39.33		2015	2016	-5.75

7.2 Cocoa related reports

In this chapter, the results from assessing reports from the ICCO is described. The reports from the ICCO were used to define the key points in a season using crop specifics for cocoa, as well as attempting to explain the price changes on a monthly and yearly scale as found in chapter 7.1. As partial input for the report assessment, the results from chapter 7.1 were used.

7.2.1 Crop specific key points

When assessing the dependency of cocoa production on the weather, the ICCO states that "*Cocoa production is highly sensitive to changes in weather conditions*" (ICCO, 2011). The ICCO also stated the following concerning crop specific needs of cocoa: "*Variations in the yield of cocoa trees from year to year are affected more by rainfall than by any other climatic factor. Trees are very sensitive to a soil water deficiency. Rainfall should be plentiful and well distributed through the year. An annual rainfall level of between 1,500mm and 2,000mm is generally preferred. Dry spells, where rainfall is less than 100mm per month, should not exceed three months. A hot and humid atmosphere is essential for the optimum development of cocoa trees. In cocoa producing countries, relative humidity is generally high: often as much as 100% during the day, falling to 70-80% during the night. The climate, soil, water supply, human actions and other environmental factors can also affect productivity*" (ICCO, 2017).

The seasonal key points found from the crop specific literature from the ICCO are therefore total amount of rain in a season, but also lack of rain in the dry season between December and April. Especially the dry season is of importance, as drought in an early stage of the growing season could have a strong negative effect on the harvest (ICCO, 2017). The wet season between May and July in IC is the main contributor of rainfall in IC as mentioned in 5.3, and can therefore be seen as an important contributor and seasonal key point concerning the total supply of rainfall in a crop year of cocoa. From the website of the ICCO, the average harvest season for the different cocoa producing countries could also be obtained. The findings are summarized in Table 15.

Table 15: dispersion of regular main crop of cocoa harvest period, for 89% of the world cocoa production during season 2014/15, with the harvest period per country indicated in green.

89% production 2014/2015	**Main crop**	January	February	March	April	May	June	July	August	September	October	November	December
Brazil 5%	**Oct-Mar**												
Cameroon 5%	**Sep-Feb**												
Côte d'Ivoire 42%	**Oct-Mar**												
Ecuador 6%	**Mar-Jun**												
Ghana 17%	**Sep-Mar**												
Indonesia 8%	**Sep-Dec**												
Nigeria 5%	**Sep-Mar**												

An interesting link can be noted from the key points as found from the reports towards prices. From the overview of the harvest season, the average price decrease in October can be explained by the large inflow of cocoa at the start of the growing season. The systematic decrease is a short-term effect of the market due to a temporary high availability of cocoa (ICCO, 2010). Following the principle of economy, a large supply equals a lower price if the demand is unchanged. Concerning price increases, one of the findings is that the average largest prices increases of cocoa occur at the start (December) and midpoint (February) of the dry season, as well as at the end of the wet and the main harvest season (July).

7.2.2 Cocoa price change related reports

Continuing with use of reports from the ICCO, WCF and World Bank to explain the price changes and understand the general dynamic of the cocoa industry, a number of interesting findings can be presented. First, the top 5 price decreases and increases as found in 7.1 are assessed. Next, the reports are assessed to define the general dynamics of the cocoa industry. From the reports, cocoa production related events are highlighted for the cocoa areas of Africa, Asia and the Americas. The events are for instance an extremely wet or dry month or set of months, or extreme high or low temperatures for a specific year. The reports are scanned for events that happen in West Africa and IC in particular. The events related to West Africa (WA) and IC in particular are collected and later tested to assess whether the events play a vital role for the crop yield in an area and can therefore be defined as a seasonal key point.

First, in Table 16 the top 5 price increases and decreases are presented with the reason for price change explained and the year in which the price change occurred.

Table 16: the top 5 of the relative price increases and decreases for a cocoa year, with the year in which the price change occurred, the magnitude of the price change and the reason for the price change according to reports form the ICCO, WCF or World Bank, all with a relation to weather

Year	Price change (%)	Reason
2001	98.19	Turning point cocoa prices, bad cocoa production expectation for 2002 due low rainfall in wet season in West Africa, lower production due to strong dry season 2001 in CI, coup d'état in CI (ICCO, 2010)
2008	40.30	Expected third consecutive year of production deficit of cocoa due to low rainfall in January in WA, but price increase suppressed by financial crisis (ICCO, 2010)
2007	35.61	Second sequential deficit, caused by weather effect of la Niña, causing too much rain in Indonesia with negative effect on production, no direct effects on the crop in WA (ICCO, 2009)
2006	24.25	Severe drought IC due to Harmattan, lower production due to el Niño causing lower rainfall in general in West Africa (ICCO, 2008)
1989	23.45	Lower production in IC due to human action by lower agricultural standards, lower cocoa prices for farmers due to constant surplus resulting in lower stimulus to produce cocoa in IC (Benjamin & Deaton, 1993)
1987	-39.33	Lower prices due to consecutive global production surplus of cocoa production, no specific reason for increase of production mentioned (FAO, 1988)
1998	-37.83	Lower prices due to production surplus following no abnormalities in the weather in cocoa producing areas worldwide, expected good harvest due to "favourable weather" in West Africa, enforced by a lower demand of cocoa globally (FAO, 2000)
2002	-28.47	Reaction to price increase 2001, expectation of good harvest 2003 due to good weather expectations in West Africa (ICCO, 2010)
1988	-17.38	Lower prices due to consecutive production surplus due to the rise of Indonesia and Brazil in cocoa production (FAO, 1988)
1991	-17.22	Reaction to strong price increases '89 and '90, more production in West Africa due to the absence of weather anomalies such as a long period of drought (FAO, 1991)

As can be seen from Table 16, multiple factors can influence prices, with weather events highlighted as the main driving force behind all the price increases and four out of five of the price decreases. The demand for cacao is steadily increasing throughout the years (ICCO, 2012). Because of the steady increase in cocoa demand, the prices of cocoa are dependent on the production of cocoa. From the reports of the ICCO, negative weather effects such as a strong dry season between December and April or a weak wet season between May and July can have a strong influence on the production of cocoa and as a result on the price of cocoa. The news about possible lower production also has a strong effect on the prices.

For instance, in 2001, prices increased as a second consecutive year of cocoa supply deficit was expected. Rather than being dependent on the actual production, the cocoa market reacted to the news of bad weather. However, the cocoa prices were also recovering from a long period of decline due to a period of structural cocoa production surplus (ICCO, 2010). This phenomenon can also be seen in Table 14, where prices leading up to the year 2000 were declining. The cocoa price was driven by expected production versus expected demand, but policies such as stimulation of agriculture (1989) or export stops (2010), reaction of the market to new about the weather affecting cocoa production, such as a Harmattan (2013), stronger or weaker currencies (1994) or an economic crisis also affected market prices due to the lower demand of luxury products such as chocolate (2008). In all of reports in the same year as the found price increases, a combination of causes with weather or weather alone was found to be the cause of the price increase. In all the reports, the effect of the weather on the price due to the effect the weather has on the production or expected effect on the production can be found. A number of years with consecutive supply deficits due to the weather can also be seen as a price stimulating event, such as the year 2006-2007-2008.

The price decreases in Table 16 were a reaction to a large price increase in the previous year or an expected surplus in production of the next cocoa year due to favourable weather or the absence of for instance periods of drought. The reason for the connection to the weather was the fact that due to crop specifics of cocoa as discussed in 7.2.1, particular weather and climate requirements are needed for optimal cocoa growth. When the crop specific demand of cocoa was met concerning precipitation and the absence of long periods of drought, production is adequate and therefore supply should meet the demand. A reaction to adequate production and thus supply can be seen in 1990, 1994, 2002 and 2009 in Table 14 (ICCO, 2016). Surplus or adequate supply for cocoa are also the main drivers for price decreases in the period of 1985 to 1988 and the period of 1997 to 1999 as can also be seen in Table 14 (ICCO, 2016).

The quote below was taken from the ICCO market review from 2012 and gave a good summary of all the aspects concerning the price of cocoa.

"This document covers the period from 2002/2003 to 2011/2012. The cocoa supply and demand situation has been generally characterized by wide fluctuations: while production has been increasing erratically with yearly growth rates of between minus 10% and plus 18%, grindings have been growing steadily at a slower pace of between two and seven per cent except for 2008/2009 when demand fell during the global economic crisis. Three seasons experienced a large production surplus, contrasting with two seasons with a significant production deficit. The ICCO Secretariat estimates that, during the current 2011/2012 campaign, demand will exceed supply.
During the period under review, the ICCO daily prices ranged between US$ 1,361 and US$ 3,730 per tonne. The minimum was reached in May 2004 when the market experienced a massive production surplus of nearly 290,000 tons. By contrast, prospects of a huge production surplus in 2010/2011 did not stop prices rocketing to 30-year highs in March 2011. However, the reason behind those high levels was mainly attributed to the export ban on Ivorian cocoa during the political crisis in Côte d'Ivoire, the largest cocoa producing country." (ICCO, 2012).

Next, the general dynamics of the cocoa market as mentioned in the annual reports and market reviews from the ICCO were assessed. When assessing the yearly report between 2006 and 2015 and the market reviews as defined in 6.1.2 concerning weather affecting production and price, two weather related seasonal key points could be defined, being the dry periods months of December, January and February and the wet season months of May, June and July. From the annual reports, the occurrence of strong dry season months was indicated as the driver with a negative effect on cocoa production in 2007, 2010, 2013 and 2014. The dry season was affected by the *Harmattan* phenomenon, but strong dry seasons were also noted in the assessed ICCO reports without a direct link to the *Harmattan* phenomenon. In 2008, dryer weather conditions in January were mentioned specifically as a cause for concern for cocoa production and reason for price increase, but the negative effect of a dry January on production was compensated with more rain later in the cocoa year. Concerning the wet season months, both positive and negative effects can be noted. From the annular reports, in 2008, 2011 and 2015 effects of the wet season had positive effects on the production. However, in 2009 and 2015, negative effects of the wet season were also noted, with a surplus in rain leading to diseases and fungi.

Another weather effect that has a strong influence on the cocoa production, according to the ICCO reports, is the El Niño Southern Oscillation (ENSO) effect. From the Climate Office of the North Carolina State University (NCSU), the following definition of the ENSO phenomenon can be quoted: "*ENSO stands for El Niño/ Southern Oscillation. The ENSO cycle refers to the coherent and sometimes very strong year-to-year variations in sea- surface temperatures, convective rainfall, surface air pressure, and atmospheric circulation that occur across the equatorial Pacific Ocean. El Niño and La Niña represent opposite extremes in the ENSO cycle. El Niño refers to the above-average sea-surface temperatures that periodically develop across the east-central equatorial Pacific. It represents the warm phase of the ENSO cycle, and is sometimes referred to as a Pacific warm episode. La Niña refers to the periodic cooling of sea-surface temperatures across the east-central equatorial Pacific. It represents the cold phase of the ENSO cycle, and is sometimes referred to as a Pacific cold episode*" (NDSU, 2017).

The ENSO effect has two direct effects on the cocoa production according to the annual reports from the ICCO. First, the El Niño phase of the ENSO effect reduces the amount of rainfall in the dry season in West Africa, Asia and South America, as specifically happened in 2006 (ICCO, 2008). The La Niña effect is not mentioned in the reports of the ICCO to affect West Africa, but affects the cocoa production in Indonesia due to a surplus of rainfall during the wet season, as specifically happened in 2007 (ICCO, 2009).

When combining all the reports of the ICCO, World Bank and WCF, and the findings concerning the price change, the general dynamics of the cocoa market can be defined. In Figure 30, the general dynamics are presented. In the figure, the green arrows represent quantifiable links, the red unquantifiable and the orange the result of the combination between green and red.

The prices of cocoa are dependent on the production of cocoa, but react strongly to news about negative weather effects in West Africa with a particular focus on either the wet or dry season. The reason for the strong reaction to the news about the weather is the expected impact the weather is going to have on the production. For the wet and dry season, a positive result is average to above average rainfall. A soft dry season with normal amounts of rain in December, January and February has positive effect, while a wet season with normal amounts of rain in May, June and July can have both a positive and negative effect.

A soft dry season with normal amounts of rain in December, January and February reduces the possibility of water stress in the cocoa trees. The months of May, June and July provide the largest contribution to the total amount of precipitation in IC. However, an above average amount of rain in one of the months can cause a surplus of water, which in turn increases the probability of a disease or fungi affecting the cocoa tree (ICCO, 2016). Concerning the effects of a surplus of rain, the Queensland Government of Australia states that "*Annual rainfall greater than 2500 mm may result in a higher incidence of fungal diseases*" (Queensland Government, 2017). The wet and dry season effect the weather, which in turn effects the supply via production.

Other aspects such as policies, economy and currency can have an effect on the prices via supply and demand, but are unquantifiable from an engineering perspective. The resulting supply, demand, stock and finally prices are all result of the combination between the driving factors of the market of cocoa. In this research, the link is specifically on the quantification of the weather, with the links represented by the green arrows. Because of the dependency of the production on the weather and the dependency of the price on both the production and the expected impact of the weather on the future production, a continuation of the research towards the quantification of the link between the cocoa production and weather with particular focus on the wet season months of May, June and July and the dry season months of December, January and February was taken.

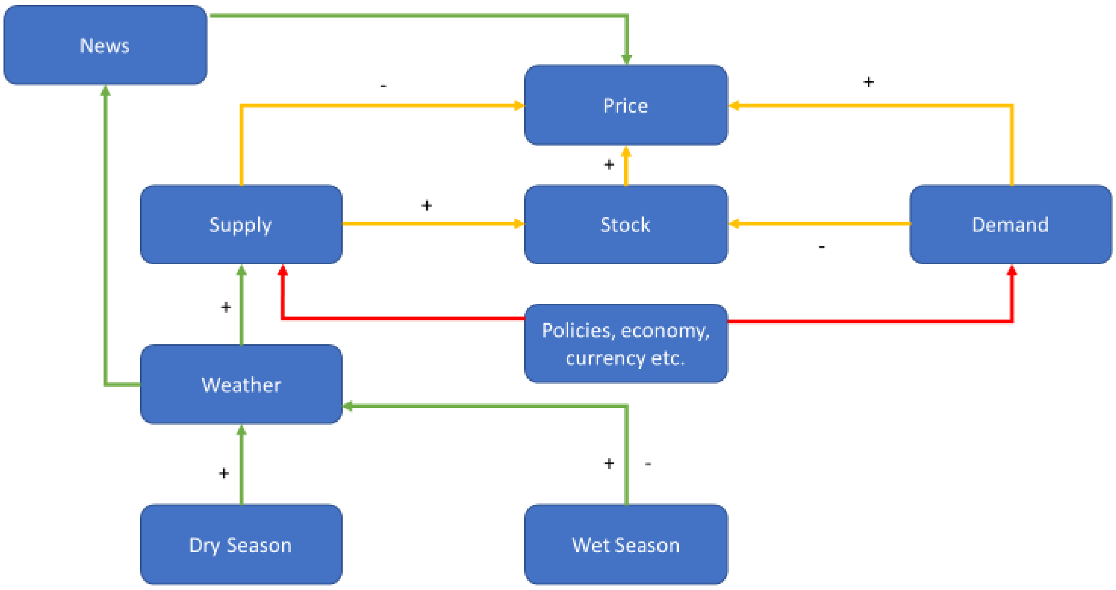


Figure 30: the dynamics of the cocoa market, with green as quantifiable links, red as unquantifiable links and orange as links as a consequence of both quantifiable and unquantifiable links.

7.3 Relevant data cocoa: production changes

For cocoa world production, data from the ICCO was used. To use the production data provided by the ICCO, a processing step was made to the data. As can be seen from Figure 31, cocoa world production was constantly increasing throughout the years. For cocoa, 92% of the variance of cocoa could be explained by a trendline, which is a linear regression between world production and time. For the world production of cocoa, a linear trendline was found with the formula:

$$EWCP = 57.1 * \mathrm{n} + 723.2 \ (eq.17)$$

With:

$EWCP$ = The Expected World Cocoa Production by using the linear trendline in thousands of tons

n = The assessed production year with start condition n=1 equal to cocoa year 1960/61

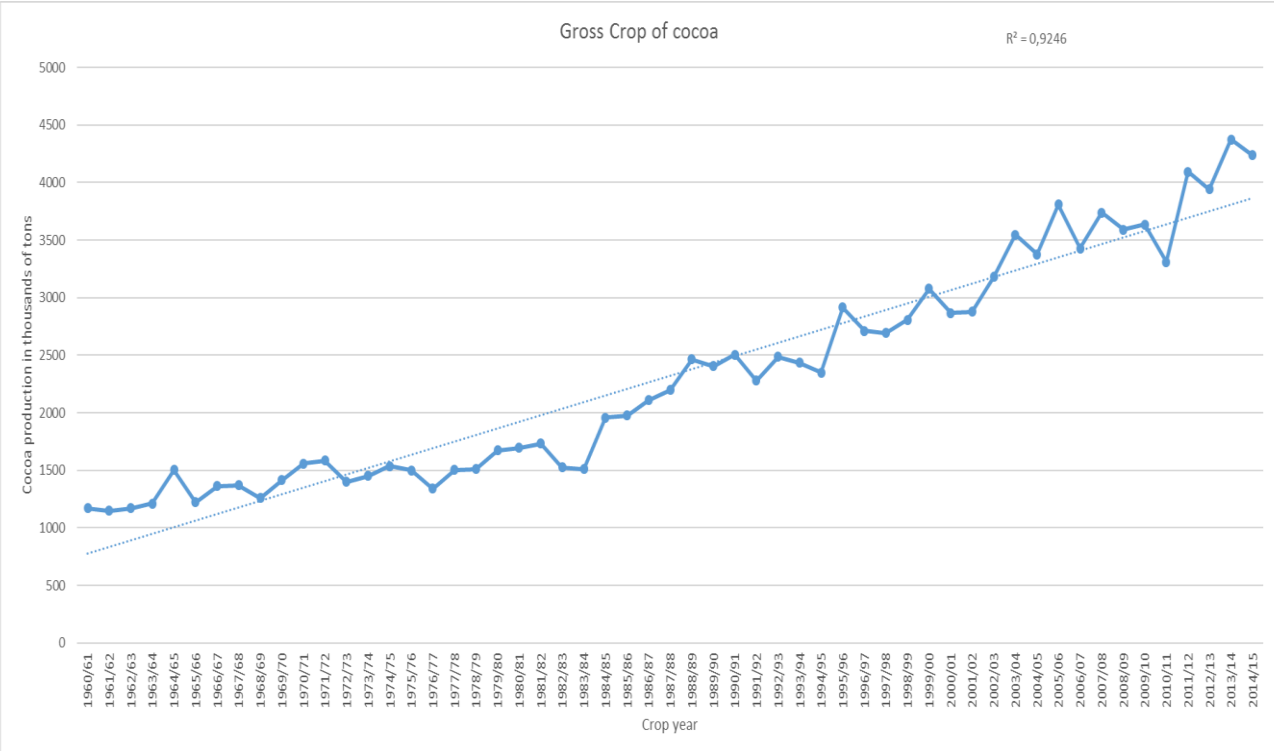


Figure 31: the world cocoa production as a function of time, with a linear trendline fitted to represent the trend in the data

From visual analysis of Figure 31, an increase in volatility from 2001 can be seen. This is in line again with the statement of the ICCO, indicating supply deficits and larger volatility in the world cocoa production. Although production is above the trendline, the "jumping" of the trendline is clearly visible from 2001 onwards. From 2001 onwards, large differences from year to year can be seen.

In Figure 32, the residuals from the trend line can be seen. As can visually be deducted from the figure, the residuals were not completely randomly dispersed. In the figure, a slight parabolic form can be seen. The parabolic form of the residuals could indicate that a quadratic function could be a better fit for the trendline. However, the trendline was accepted as a first assumption and approximation, and it was assumed that cocoa the production increases linearly as a function of time for this research. Therefore, the difference from the linear trend line was used in further parts of this research to indicate the change in production.

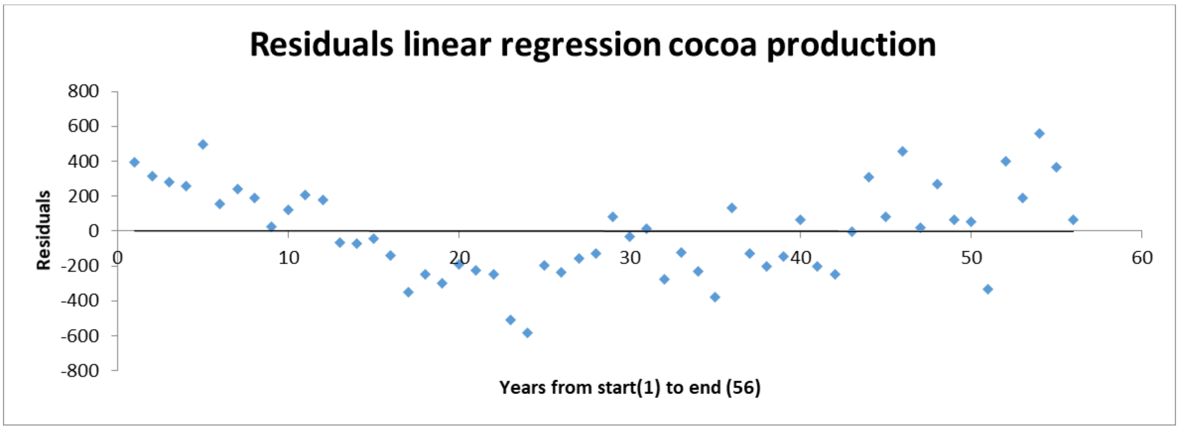


Figure 32: residuals for the linear regression trend line between world cocoa production as a function of time

7.4 Relevant data cocoa: influence local production cocoa on world production cocoa

Following the same reasoning as with the world cocoa production, a linear multivariate regression was used in this research as a first approximation to determine the strength of the cocoa production of IC and Ghana towards the world cocoa production.

The correlation of IC and Ghana was first tested using the Pearson coefficient, to determine if the variables were independent within the used dataset. The correlation between IC and Ghana was found to be 0.59, which indicates a weak correlation according to the Pearson coefficient. Therefore, IC and Ghana were used as independent variables in the linear multivariate regression.

A multivariate linear regression was performed with the world cocoa production as dependent variable and Ghana and IC as independent variables. The regression was tested between 2002 and 2014 using data from the annual reports from the ICCO as defined in 6.1.2. In Table 17, the results of the regression are shown. In the table, the R^2 value of the regression is shown, as well as the significant F value of the regression, the coefficients of the independent variables IC and Ghana and the P-values of the independent variables.

Table 17: the results of the regression of the world cocoa production as a function IC and Ghana cocoa production, with R^2, P- and F-value and coefficients

R^2	0.92
Regression Significant F-value	4.07×10^{-6}
Production Ivory Coast P-value	0.0035
Production Ivory Coast Coefficient	0.91
Production Ghana P-value	0.00022
Production Ghana Coefficient	1.69

When assessing the statistical significance of the variables of the regression models, the F- and P-values were within the requirement of being below 0.05. Therefore, the significance of the variables was proven in accordance with Bland (Bland, 1995). For the regression model, the coefficients of the variables were also acceptable, since both the coefficient of IC (0.91) and Ghana (1.69) in the regression were within the same order of magnitude. Therefore, both independent variables were of mutual importance for the regression model and none of the independent variables is repressed by Windows Excel.

Concerning the R^2 values, 0.92 of the variance of the cocoa production could be explained by assessing the cocoa production of IC and Ghana. By using the Pearson coefficient, the relation between the world cocoa production and the cocoa production predictor using the cocoa production of IC and Ghana could be qualified as “strong”. Therefore, the dependency of the world cocoa production on the cocoa production of IC and Ghana is confirmed. The finding is in line with the definition of a dominant market share, indicating a market share of over 60% (Athey & Schmutzler, 2001).

Based on the found results, a predictor for the world cocoa production based on the cocoa production of IC and Ghana can be formed. The results from the regression analysis can be summarized in equation 18. The resulting graph from equation 18 and Table 17 is shown in Figure 35.

$$PWPC = 0.91 * CP_{IC} + 1.69 * CP_{GH} \ (eq. 18)$$

With

$PWPC$ = Predicted World Cocoa production in thousands of tons

CP_{IC} = Cocoa Production Ivory Coast in thousands of tons

CP_{GH} = Cocoa Production Ghana in thousands of tons

As can be seen in Figure 33, the predictor for the cocoa production is off in the year 2008/09 and 2009/10. The difference between the world cocoa production predictor and the actual world cocoa production can be explained. While both IC and Ghana have a dominant role in the cocoa production with 61% of the world production in 2014/15, the production of other cocoa producing countries such as Indonesia with 11% in 2014/15 and Brazil with 5% in 2014/15 also affect the world cocoa production. The linear regression ignores the contribution of other regions, which in turn causes the regression to be off when IC and Ghana have a lower cocoa production, but Asian and South American cocoa producing countries do not. In 2008/09 and 2009/10, this situation was the case (ICCO, 2010) (ICCO, 2011). However, the predictor the follows the same movement as the actual production in the other years.

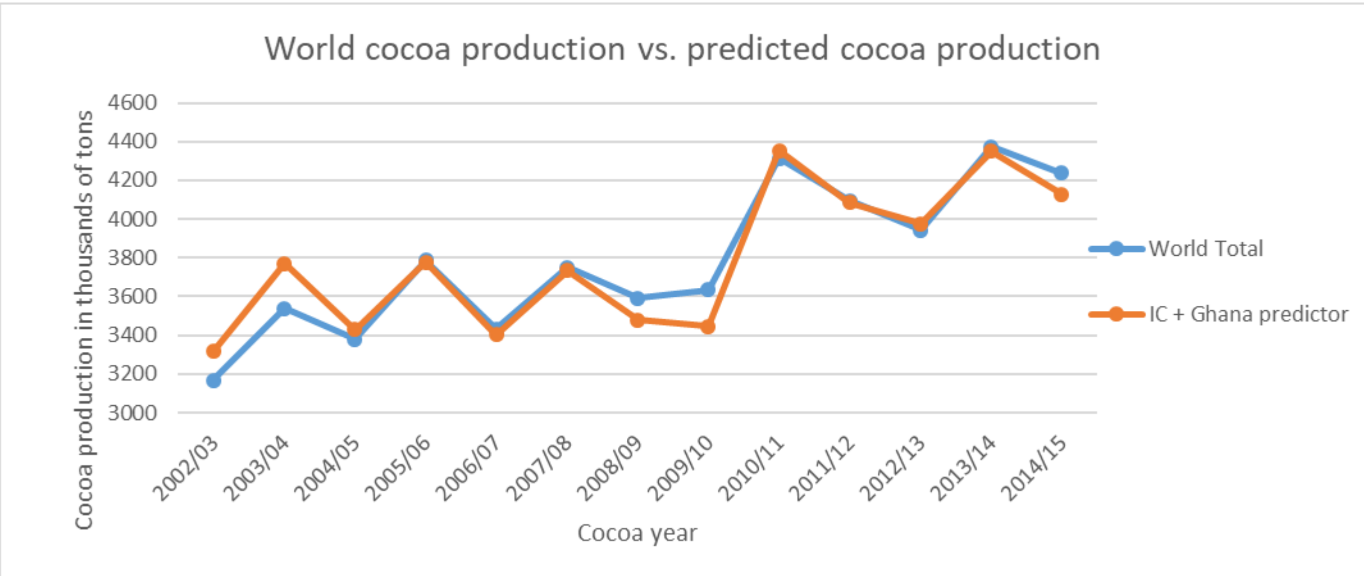
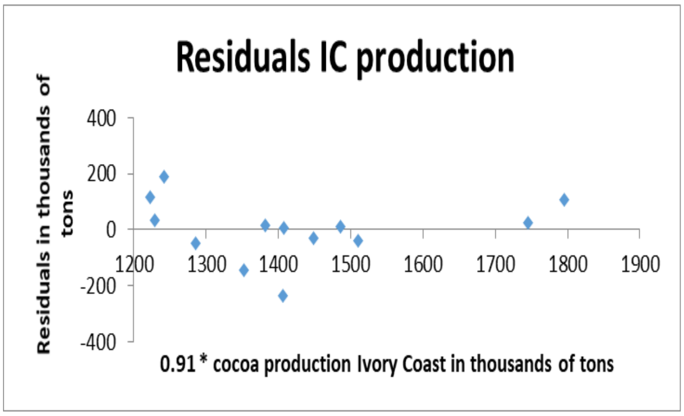


Figure 33: the actual world production of cocoa (blue) versus the predicted world cocoa production as a function of IC and Ghana (orange), given as a function of production in thousands of tons versus the year of the harvest.

In Figure 34, the contribution to the residuals of the regression model for IC and Ghana are shown. In the figure, the residuals visually appear to be not completely randomly dispersed. For IC, more positive than negative residuals can be noted, with most of the residuals below 1500. For Ghana, almost all of the residuals are positive, and slight heteroscedasticity can also be noted, with a larger distance from the zero axis for lower values and smaller distance from the zero axis for higher values. However, the linear model was accepted as a first approximation for this research.



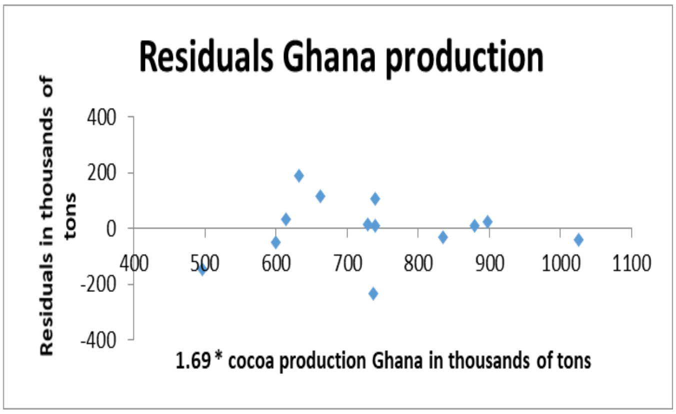


Figure 34: Residuals for the cocoa production predictor of IC and Ghana in thousands of tons. The graph shows the relative contribution of the predictor of IC (left) and Ghana (right) towards the world cocoa predictor. On the vertical axis, the contribution towards the residual of the trendline in shown in thousands of tons, while on the horizontal axis the relative cocoa production contribution of IC (left) and Ghana (right) is presented in thousands of tons.

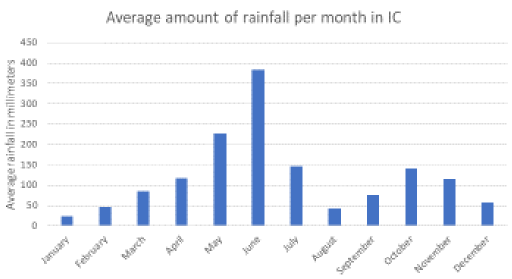
7.5 Relevant data cocoa: weather station data

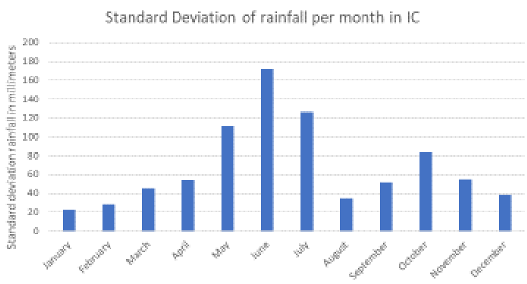
After cleaning up the data, a correlation matrix could be constructed to determine the correlation between the various weather stations used in IC. The results of the correlation assessment are shown in Table 18. The correlations as shown in the correlation matrix are the Pearson correlations between the average monthly rainfall for one year, consisting of 12 points in total. For IC, correlations reach a low of 0.65 between for Dimbokro and Sassandra. However, the data for WMO and other weather stations were from different years, so no direct correlation between the stations could be made for the same set of years. Therefore, the simplification to take the monthly arithmetic mean of all the available weather stations was accepted for this research.

Table 18: the correlation matrix of the Pearson correlation between average monthly rainfall between the used weather stations in IC

	Gagnoa	Abdijan	Adiake	Sassandra	Abidjan Ville	Adiake WMO	Sassandra WMO	Dimbokro WMO
Gagnoa	1.00	0.70	0.78	0.67	0.74	0.80	0.72	0.96
Abidjan	0.70	1.00	0.99	0.99	0.98	0.97	0.96	0.70
Adiake	0.78	0.99	1.00	0.98	0.99	0.98	0.97	0.76
Sassandra	0.67	0.99	0.98	1.00	0.99	0.95	0.97	0.65
Abidjan Ville	0.74	0.98	0.99	0.99	1.00	0.97	0.97	0.71
Adiake WMO	0.80	0.97	0.98	0.95	0.97	1.00	0.96	0.80
Sassandra WMO	0.72	0.96	0.97	0.97	0.97	0.96	1.00	0.71
Dimbokro WMO	0.96	0.70	0.76	0.65	0.71	0.80	0.71	1.00

After taking the arithmetic mean of the used weather stations, the average monthly rainfall and deviation corresponding to a specific month could be determined for IC. In Figure 35, the arithmetic mean of the weather stations per month and the corresponding deviation is shown.





	January	February	March	April	May	June	July	August	September	October	November	December
Average in mm	22	46	84	118	228	384	145	43	75	141	116	57
Standard Deviation in mm	23	28	46	54	111	172	126	35	52	84	55	38
Std dev/average	1.03	0.61	0.54	0.46	0.49	0.45	0.87	0.80	0.69	0.59	0.47	0.68

Figure 35: the average monthly rainfall of the dataset, using the arithmetic mean of the used weather stations in IC (top left), the standard deviation of the average monthly rainfall in the dataset, using the arithmetic mean of the used weather stations in IC (top right) and the numbers of the monthly average rainfall from the arithmetic mean, the standard deviation related to the monthly average and the ratio between the standard deviation relative to the average rainfall (central bottom)

The outcome of the analysis of the monthly arithmetic mean rainfall as shown in Figure 35 confirm the described weather as found in the description from Our Africa, with a wet period between May and July and a dry period from December to April (Our Africa, 2017). The standard deviations follow the same pattern as the rainfall, with the highest absolute deviation in the wet season. In Figure 35, the averages, standard deviation and the ratio of the standard deviation to the average rainfall is given in the bottom central.

As can also be seen from Figure 35, the months with the highest mean to deviation ratio are January, July and August, but December, September and February also show high a ratio of over 0.6. The high ratio between mean rainfall and deviation translate into the observation that a high amount of variability could exists between the expected and actual rainfall in a specific month.

Due to the high dependency of enough rainfall and the absence of long periods of drought, especially the dry season between December and January and the wet season between May and July are of importance for the production of cocoa (ICCO, 2016). From December to January, enough rainfall is needed to guarantee that the lower threshold of no more than 3 consecutive months with less than a 100 mm of rain is met. In the wet period between May and July, enough rain needs to fall to guarantee that the upper threshold of between 1500mm and 2000mm of rainfall within a crop year is met, but not too much rainfall to avoid diseases. However, the upper threshold of 2000mm was never breached in the assessed dataset. By assessing the rainfall data in combination with the findings from the ICCO reports as presented in 7.2.2, the importance of the dry period and wet period in IC towards the cocoa production of IC were confirmed.

7.6 Possible key periods cocoa season based on analytical findings: linear regression

From the reports of the ICCO, the importance of adequate amounts of precipitation in combination with the absence of long periods of drought were indicated as vital factors for the growth of cocoa, as can be found from the reports as 7.2.2. As seasonal key points, the rainy season from May to July and the dry season between December and March were suggested, as confirmed by the findings chapter 7.2.2 and 7.5. As a first approximation to define the correlation, a linear regression was used in this research. For the linear regression, the difference from the trendline concerning the world production of cocoa as found in 7.3 was used as dependent variable. For the independent variables, of the monthly arithmetic mean rainfall of the used weather stations in the same cocoa year as the relative change of production was used, as presented in chapter 7.5.

First, a regression was performed using all the month of the year as independent variables. A summary of the result of the regression is shown in Table 19, the full regression data can be found in appendix VII. Note that in Table 19, the months of May, June, December and January have the lowest P-values, indicating highest statistical significance.

Table 19: the results of the regression of the relative change of the world production of cocoa as a function the rainfall in all the months of the year in IC, with P-value

Variable	*P-Values*
November	0.89
March	0.63
April	0.61
September	0.51
August	0.38
July	0.28
October	0.24
February	0.14
January	0.13
December	0.08
June	0.08
May	0.004

After testing and running the iterative regression model a number of times with filtering of the data using the key points as found in 7.2.1, the months of December, January, May and June were found to be the key points concerning the relative change of world cocoa production. This was in line with the found reports and finding in chapter 7.2.2 and 7.5 due to the coincidence of the found seasonal key points with part of the dry season and wet season. The final regression was performed using the defined set of months. The total rainfall of the wet and dry season was also tested, but the P- and F-values found were far above 0.05 and therefore not significant. From the regression, it can be stated that not all of the wet and dry season were found to be important towards the cocoa production. In Table 20, the results of a regression of the relative change of world production of cocoa as a function of rainfall in the month of December, January, May and June in IC are shown.

Table 20: the results of the regression of the relative change of the world production of cocoa as a function the rainfall in the months of December, January, May and June in IC, with R2, P- and F-value and coefficients

R^2	0.47
Significant F value regression	$9.70x10^{-5}$
December P-value	0.035
December Coefficient	1.94
January P-value	0.013
January Coefficient	4.69
May P-value	0.006
May Coefficient	-0.88
June P-value	0.0053
June Coefficient	0.6

As can be seen from Table 20, the P-values of the chosen seasonal key points as independent variables are within the 0.95 confidence bound with values below 0.05, and can therefore be interpreted as statistically significant. The F-value of the regression is also below 0.05, indicating the significance of the regression as a whole. The regression has an R-squared of 0.47, which translates into the finding that almost half of the change in world cocoa production could be explained by using a set of season key points in IC within a cocoa year. All the coefficients of the independent variables found in the regression are in the same order of magnitude, confirming the relevance of the coefficients.

Among the coefficients, a number of interesting results were found. The amount of rainfall in the dry season months of December and January have a profound effect on the regression, with the coefficient of January being 5 times larger than the coefficient of the wet season months of May and June. The importance of an adequate amount of rainfall is a reoccurring import factor already defined in 7.2.1 and 7.2.2. The results from the linear regression are interpreted that the linear regression agrees with the notion that the amount of rainfall in the dry season is of large importance by assigning a large coefficient to months of the dry season. Concerning the effects droughts, the absence of a strong dry season or events such as a Harmattan or an El Niño and subsequently higher precipitation in the dry season is according to the ICCO of large importance for the development of cocoa. The relative importance of rainfall in the dry season is therefore underlined by the linear regression, in agreement with the reports as presented in 7.2.1 and 7.2.2.

A negative effect can be found for rainfall in the wet season month of May. May signifies the start of the wet season, with May on average not being the wettest month of the year. While adequate rainfall is important for the development for cocoa, a surplus of rainfall could have a negative effect on cocoa, promoting the susceptibility of cocoa trees to for instance fungi or rot. This effect occurred for instance in 2008 in IC (ICCO, 2009). June has a positive effect on the change in cocoa production, in line with the findings that June is the height of the wet season and therefore of importance due to the rainfall in June provides towards the total rainfall of the year. The results from the linear regression are interpreted that the linear regression agrees with the notion that the amount of rainfall in the wet season can have both a positive and negative influence on the cocoa production, by assigning both a positive and negative coefficient to months of the wet season. The findings are in line with the findings of the reports from the ICCO in 7.2.2.

Based on the found results, a predictor for the world cocoa production based on the arithmetic mean rainfall in the months of December, January, May and June in IC could be made. The results from the regression analysis could be summarized in equation 19 as shown below:

$$PRCWCP \; = 1.94 * R_D + \; 4.69 * R_{Ja} - 0.88 * R_M + 0.60 * R_{Ju} \; (eq.19)$$

With

$PRCWCP$ = Predictor Relative Change World cocoa production, relative to linear trendline in thousands of tons

R_D = Rainfall December, arithmetic mean in millimetres

R_{Ja} = Rainfall January, arithmetic mean in millimetres

R_M = Rainfall May, arithmetic mean in millimetres

R_{Ju} = Rainfall June, arithmetic mean in millimetres

Using the found equation, the comparison between the predictor and the actual difference of the cocoa production relative to the trend could be composed. In Figure 36, the comparison between the predictor and the actual change of cocoa production relative to the trend is shown. As can be seen from Figure 36, the predictor for the change in cocoa production is not a perfect fit, but does follow the same general movement of the actual difference. Differences can be seen in particular in years where the rainfall and subsequent cocoa production in IC was below the trend, but other areas such as Indonesia and Brazil did have a good harvest. The years 1988/89 is an example of this phenomenon, as also defined in 7.2.2. In the year 1993/94, the exact opposite is true, with a higher expected production based on weather in IC but with a lower actual production. However, the exact reason for the difference in 1993/94 is not known.

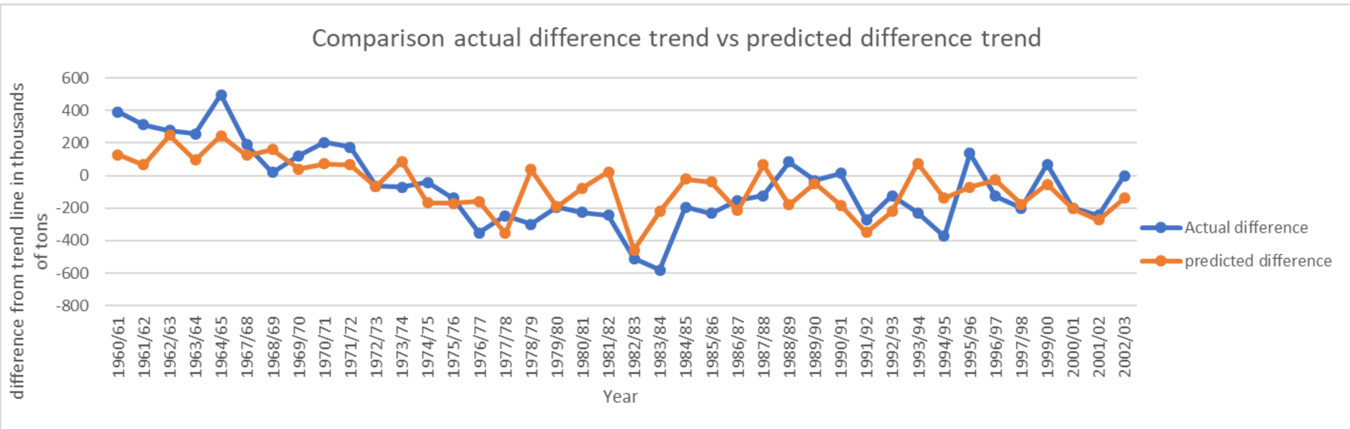
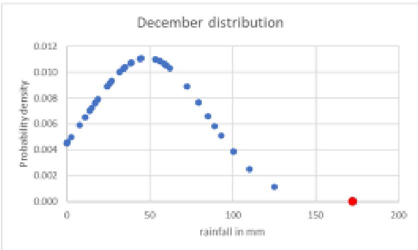


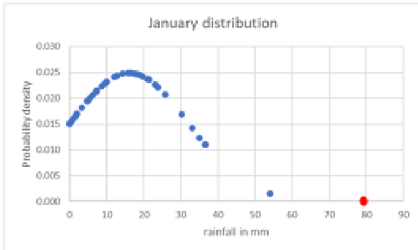
Figure 36: comparison between the actual difference of the actual cocoa production from the trend versus the predicted difference in cocoa production from the trend using the key point regression

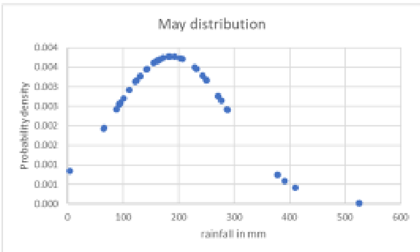
As a final test, the rainfall data and production differences are tested on skewness and kurtosis to determine if the data is normally distributed. In Figure 37 and Figure 38, the probability distribution of the rainfall and the production difference from the trend of the world cocoa production is shown. The kurtosis of December and January were above the set boundary of 2 without removal of the extreme values. However, when evaluating the figure of the distribution of both January and December, the main cause of the high kurtosis was a high above average rainfall points on the right of the peak in both months, indicated as the red dot in December and January in Figure 37. When assessing the results with the average rainfall as found in Figure 35, it can be stated that the events were outside the 0.95 confidence band of the mean plus two standard deviations. Therefore, the measurements can be seen as extreme events and rejected from the analysis. The skewness and kurtosis of December and January with the extreme events filtered out and all of the other data are within the bounds as stated in 1.1. Therefore, the data was perceived as being normally distributed and the multivariate regression was subsequently accepted.



December	
Kurtosis	-0.01
Skewness	0.70

January	
Kurtosis	0.81
Skewness	0.95





May	
Kurtosis	1.85
Skewness	1.16

June	
Kurtosis	-1.09
Skewness	0.15

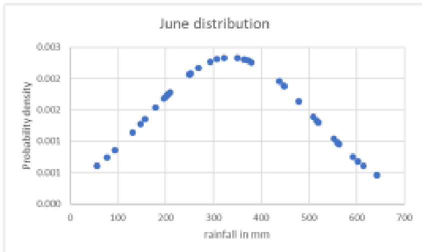
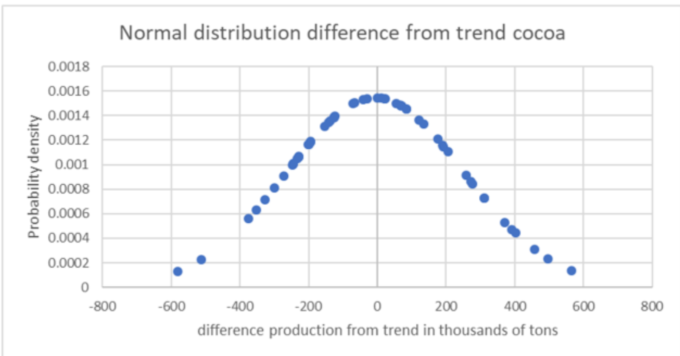


Figure 37: the probability distribution of the rainfall in December, January, May and June with tables containing the kurtosis and skewness of the rainfall data. For January and December, the extreme values removed are indicated in red.



Production difference from trend	
Kurtosis	-0.03
Skewness	0.32

Figure 38: the probability distribution of the difference from the trend of the world cocoa production with table containing the kurtosis and skewness of the difference from trend for the production of cocoa data

7.7 Binary Bayesian Predictive model

Following the results from chapter 7.6, a BBPM could be completed. The BBPM was made to provide a probabilistic expectation following the regression formula as presented in 7.6. For the BBPM, a conditional probability for relative change in production was defined, while the months of December, January, May and June an unconditional probability was defined. To construct the BBPM, the seasonal key points and production were translated into binary variables. To translate the seasonal key points to binary, rainfall above (1) or below (0) the arithmetic mean was used. For the production, relative change above (1) or below (0) the trend was used to translate the production to binary. The formula for the probability for crop prediction is given below as equation 20 and the BBPM with the unconditional probability of the seasonal key points of December, January, May and June can be seen in Figure 39:

$$P(P_{1,0}) = P(D_{1,0}) * P(Ja_{1,0}) * P(M_{1,0}) * P(Ju_{1,0}) * P(P_{1,0}|D_{1,0}, Ja_{1,0}, M_{1,0}, Ju_{1,0}) \ (eq. 20)$$

With

$P(P_{1,0})$ = The probability of an above (1) or below (0) the trend change in cocoa production occurring

$P(D_{1,0})$ = The probability of an above (1) or below (0) the mean amount of rainfall in December occurring

$P(Ja_{1,0})$ = The probability of an above (1) or below (0) the mean amount of rainfall in January occurring

$P(M_{1,0})$ = The probability of an above (1) or below (0) the mean amount of rainfall in May occurring

$P(Ju_{1,0})$ = The probability of an above (1) or below (0) the mean amount of rainfall in June occurring

$P(P_{1,0}|D_{1,0}, Ja_{1,0}, M_{1,0}, Ju_{1,0})$ = The conditional probability of an above (1) or below (0) amount of the trend change in cocoa production, conditional of the combination of above or below the mean rainfall in December, January, May and June occurring

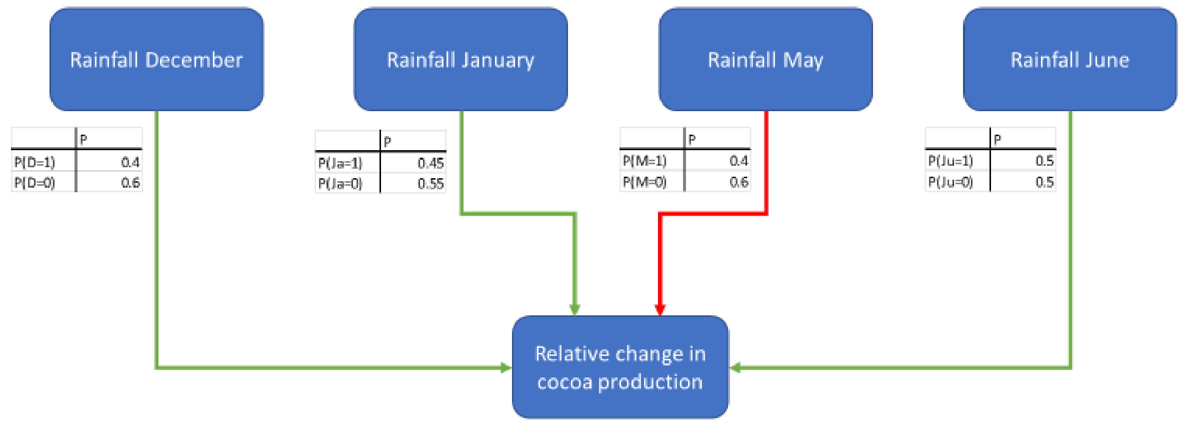


Figure 39: Binary Bayesian Probabilistic Model of the effect of the above or below the mean amount of rainfall of the seasonal key points of December, January, May and June, with the unconditional probabilities corresponding to the seasonal key point based on historical occurrence between 1960 and 2002.

In Table 21, the conditional probability of the change in production relative to the trend in cocoa production is show. To determine all the conditional probabilities, the combinations of above or below the mean amount of rainfall for all the seasonal key points were calculated based on historical occurrence. The found provide an expectation based on historical data. Note that some of the cases the conditional probability for production was 0, indicating that the specific conditional probability did not occur between 1960 and 2002. An example is the conditional probability with December above mean (1), January above mean (1), May below mean (0) and a June above mean (1).

Table 21: The conditional probability of the of an above (1) or below (0) amount of the trend change in cocoa production, conditional of the combination of an above or below the mean amount of rainfall in December, January, May and June.

D	Ja	M	Ju	P(P=1)	P(P=0)
0	0	0	0	0.17	0.83
0	0	0	1	0.00	1.00
0	0	1	0	0.00	1.00
0	1	0	0	0.17	0.83
1	0	0	0	0.50	0.50
0	0	1	1	0.00	1.00
0	1	0	1	1.00	0.00
1	0	0	1	0.67	0.33
0	1	1	0	0.33	0.67
1	0	1	0	0.50	0.50
1	1	0	0	0.00	1.00
1	1	1	0	0.00	0.00
1	1	0	1	0.00	0.00
1	0	1	1	0.00	1.00
0	1	1	1	1.00	0.00
1	1	1	1	0.80	0.20

As an example, to interpret the results, assume that the probability of an above average change in cocoa production needs to be determined at the start of a cocoa season with the assumption that the rainfall in all of the seasonal key points will be above the mean. Equation 20 takes the shape of:

$$P(P_1) = P(D_1) * P(Ja_1) * P(M_1) * P(Ju_1) * P(P_1|D_1, Ja_1, M_1, Ju_1)\ (eq.21)$$

The probability of the event actually occurring then becomes:

$$0.4 * 0.45 * 0.4 * 0.5 * 0.8 = 0.0288$$

The probability of 0.0288 is the multiplication of the probability of all the seasonal key points being above the mean multiplied by the conditional probability of the production change being above the trend. All possible combinations of probabilities can be calculated to determine the probability of an event occurring. Note that this possibility is a blind possibility, as it is one of the options at the start of the season without knowing if the weather in the seasonal key points is above or below average. If for instance the probability of an above average cocoa production is determined in March, with the knowledge that December and January were above average, equation 21 takes the shape of:

$$P(P_1) = P(D_1) * P(Ja_1) * P(M_1) * P(Ju_1) * P(P_1|D_1, Ja_1, M_1, Ju_1)\ (eq.22)$$

With $P(D_1)$ and $P(Ja_1)$ being 1 due to the fact that the months of January and December were above average. The probability of the event actually occurring then becomes

$$1 * 1 * 0.4 * 0.5 * 0.8 = 0.16$$

Part III: General Discussion and Conclusion

8. Discussion

This research focussed on using a combination of qualitative data in the form of wheat and cocoa reports and quantitative data in the form of rainfall and snowfall measurements of weather stations to determine the strength and impact of the rainfall in specific time periods in ND, the Canadian Prairies and IC towards the spring wheat production and world cocoa production. As a proof of concept, the research provides interesting results. The combination between quantitative and qualitative data provides a fast track method of isolating key points in a season. The key points can then be used to explain variance in the world production. One of the starting points of the research, being the preservation of accessibility to the method, is also met. The method provides an indication of the yield and production movement based on rainfall data readily available online. This could imply that the method could also be used offline and based on snowfall and rainfall data alone, if the seasonal key points for a specific crop can be defined. However, improvements to the technology can always be done. In this chapter, possible improvements on various part of the research as well as reflection on the performed research are done.

The inflation correction of market prices

For this research, the market data for wheat and cocoa in the form of future contracts as defined in 2.1.1 and 6.1.1 were translated into relative change of prices on both a yearly and monthly scale as shown 3.1 and 7.1, after which the data was corrected using inflation numbers for the dollar. The argument for using the technique of correcting for inflation, is to find the "clean" value of the price increase, as stated in the introduction 2.2 and 6.2. For cocoa, this could provide a problem. While prices of cocoa are quoted in dollars and the inflation of the dollar should be used to clean up the financial data, the simplification completely ignores the use of local currency and exchange rates. The inflation was taken for the US dollar, but are therefore also dependent of the US economy. From a financial perspective, the use of exchange rates could be researched. Large local currency drops such as in IC in 1991 can make cacao prices drop without an apparent reason when only evaluating the weather. There is still much discussion on what the best way is to incorporate inflation into price fluctuation of commodities (Browne & Cronin, 2010).However, this subject is outside the scope of this research.

The use of world production for cocoa

For the cocoa production, world data was used. While the goal of this research was to determine the effect of weather in IC towards the world production of cocoa, the step does introduce errors. The use of world production data versus local rainfall data from IC can be used as a first indicator, but ignores other effect that influence world production, such as the influence of cocoa production in other countries. As a logical continuation of this research, a follow up could be to use local production data of IC only in combination with seasonal key points. This step would clean up the production data by only assessing the area from where the rainfall data came. A further improvement could be the use of yield data, as was used with spring wheat. The use of yield data removes effects such as more area planted, and thus could be used to determine the effect of the weather on the cocoa production. However, the goal of this research was to test the effect of rainfall patterns in IC alone on the world production.

The method used in the research is as of yet applied to a world cocoa production, but could be focussed and narrowed down to assess smaller areas and thus local production. 42% of the variance of the world cocoa production could be explained by using rainfall data from IC alone. A logical next step is to look at cocoa production of Ghana and IC as standalone and use regression modelling to predict cocoa production on a local scale. When accuracy in the form of the variance explained is high for both IC and Ghana, the output can be used subsequently to predict the world production using the found correlation as presented in 7.4

Data source use
For this research, reports from the FAO and ICCO were used as main source of information. While OXFAM, Bloomberg, the WCF and World bank were also used, the information as provided from the FAO and ICCO were the main contributors. The FAO plays a dominant role in providing an insight into the agriculture of the world, the ICCO plays a dominant role in the world of cocoa, with most of the information concerning cocoa collected and subsequently distributed via the ICCO. While there is no direct way of circumventing these sources, diversification of the sources of data could lead to new findings. However, the data as provided by the FAO and ICCO were consistent with the findings and with other reports from OXFAM, Bloomberg, the world bank and the WCF, and the use of the FAO and ICCO reports can therefore be defined as valid.

Linear trendline production versus quadratic trendline cocoa
To use the world cocoa production, the difference from a linear trendline is taken. However, a positive quadratic trendline provides a better fit and a higher explanation of variance, as can be seen in Figure 40. The quadratic trendline explains almost 97% of the variance in world cocoa production versus 92% for the linear trendline. The better fitting positive quadratic trendline was rejected on the basis that the linear approximation would provide a good first insight into the possible use of the technology, as opposed to complicating the adjustment technique before functioning of the regression was proven. For further research, the adaptation of a quadratic trendline could be a first step in improving on the linear approximation.

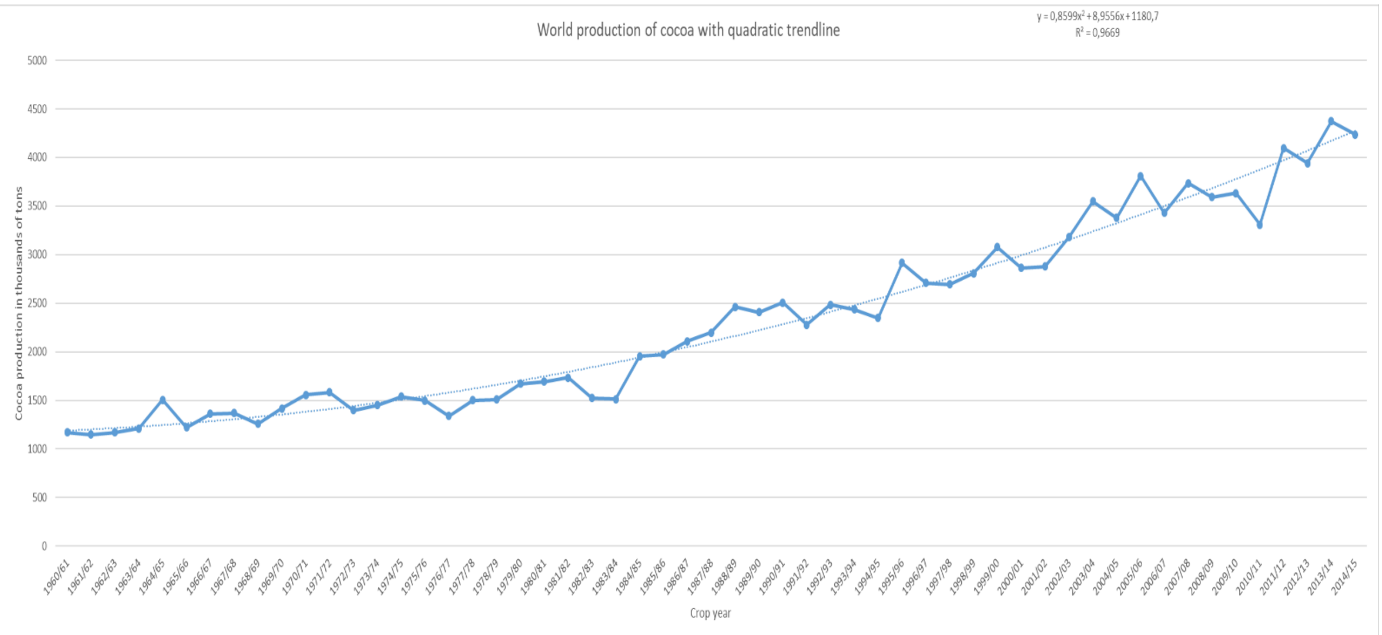


Figure 40: the world cocoa production as a function of time, with a quadratic trendline fitted to represent the trend in the data

Linear regression and residuals
As a first approximation and set boundary for this research, only linear regressions were used for production, yield and the multivariate regression. As can be seen from the residuals in Figure 11, Figure 13, Figure 15, Figure 32 and Figure 34, none of the residuals were fully randomly dispersed, indicating that a linear approximation might not be the best fit for the data. While the use of only linear regression is a valid first approximation, other forms of regression could be looked into. As a next step continuing on the results of this research, the use of for instance quadratic regression could be considered. The impact on the results by using for instance quadratic regression is however not known.

Processing of rainfall and snowfall data

The rainfall and snowfall data from weather stations are averaged out to be able to assess ND, the Canadian Prairies and IC as one area by using the arithmetic mean. While this step is defended by showing that the average monthly snowfall and rainfall of weather stations is strongly correlated, the step does ignore spatial variability of rainfall. A next step could be to assess the crop of an area within a specific range of a weather station. A Thiessen polygon could be used to mark the area of effect for the weather station in future research. This could also be a continuation of the research, focussing in on a local scale.

For IC, a closer look into the correlation between different weather stations is also advised. For instance, the weather stations of Dimbokro and Sassandra correlate for 65%. In this research, the fact that the data sets for Dimbokro and Sassandra are not available in the same time period. A direct comparison could not be made between the stations and is therefore the main reason for the just above average correlation.

Spread of the weather stations

The amount and spread of the weather stations is also a point of discussion. While more weather stations were available in the research areas, not all off the stations were used in the rainfall analyses. The main argument for not using all the available data, is the fact that this research is still a proof of concept in to the subject of key point predictive modelling. The goal of this research is to gain understanding in the larger dynamics of the effects of rainfall on crop yields and to determine whether key points in a season have in fact such a profound effect as stated in reports. In further research, a closer look into individual areas could be taken. A higher density of weather stations and possibly the use of privately owned weather stations to ensure a constant flow of data could be taken into account for further research. For instance, in IC, large differences between different climate zones in IC can be noted (Kouakau, 2012). These differences between climate zones were neglected for this research, but could be taken into account in a follow up study. The same reasoning for ND and the Canadian Prairies hold, with for instance in ND a climate difference between the south west part of ND as opposed to the rest of the area.

The use of production as price indicator

For this research, yield and production were used in the linear regression with seasonal key points. The argument for applying a linear regression to spring wheat yield and cocoa production, is because of the dominant role production has on the price changes of wheat and cocoa. When assessing the strongest driver behind the production of wheat and cocoa, the weather is found to be the strongest factor. Prices of wheat and cocoa change due to the change in production of wheat and cocoa, but cocoa prices also react strongly on the expected impact weather is going to have on the production of cocoa. By understanding the impact weather and in particular seasonal key points have on the production of wheat and cocoa, price changes of wheat and cocoa can be assessed to be valid price changes or invalid price changes. When for cocoa for instance there is limited rainfall in the dry period, a price increase of the cocoa market can be defined as valid when using the seasonal key point equation 17 and the BBPM, because a lower crop prediction combined with a high probability of a below the trend crop would be expected. To apply seasonal key points to price change expectations, validation by back testing with actual cocoa prices is needed. This subject is however outside the scope of this research.

Crop season key point predictor
For this research, an iterative process of performing linear regressions between wheat yield and cocoa production change, and snowfall and rainfall in specific months was performed. The risk of using regressions is that if enough data is used as input for the regression, the chance is high that a correlation will appear. In some cases, these correlations can be so called spurious results. These spurious results imply a correlation in the data that cannot be explained in real life. In this research, this element of randomly using variables and finding spurious results is countered by only accepting variables that can be explain a change in spring wheat yield and cocoa production both in a quantitative manner by using probabilistic significance and qualitative using reports and crop specifics of spring wheat and cocoa.

As an experiment for both wheat and cocoa, a linear step-wise regression was performed as a comparison to seasonal key point regression. The type of linear step wise regression used was backward elimination, eliminating independent variables with the highest P-values until all the variables were below the set 0.05 level of significance. For spring wheat, a regression between the spring wheat yield in ND and the Canadian Prairies versus the rainfall in the previous year, the snowfall before sowing and the rainfall during the wheat season was performed. For ND, the months within the wheat year used for the regression were April to October, covering all the found rainfall data as found in 4.5. For Canadian Prairies, the months within the wheat year used for the regression were October to September, to coincide with the end of the harvest season of the previous year as found in 4.2.1. For cocoa, the rainfall in the months from October to September were used, to coincide with one cocoa year.

For spring wheat yield, different results were found from the step wise regression than were found from seasonal key points. For ND, the month of September as opposed to May was found to have the strongest correlation with the spring wheat yield, with the results shown in Table 22. The explained variance of the step-wise regression is slightly higher than the seasonal key point, with 0.59 versus 0.54. The problem with the found results is that the importance of the month of September towards the spring wheat yield in ND cannot be defended using the qualitative data from the NDSU HREC. Spring wheat in ND is harvested from September to October, with a negative effect noted if high amounts of rainfall occur in September due to for instance the flooding of fields. Therefore, a negative coefficient rather than a positive coefficient for the month of September would have been expected. The question is if the month of September is a spurious result, or that there is another factor influencing the spring wheat yield in ND.

Table 22: the results of the step wise regression of the relative change of the spring wheat yield in ND as a function the rainfall in the previous year, snowfall before sowing and rainfall in the month September, with R2, P- and F-value and coefficients

R^2	0.59
Significant F value regression	0.0033
Rainfall previous year P-value (mm)	0.0065
Rainfall previous year Coefficient (mm)	0.041
Snow P-value (mm)	0.062
Snow Coefficient (mm)	-0.0042
September P-value (mm)	0.033
September Coefficient (mm)	0.10

For the Canadian Prairies, the month of November and December in combination with the rainfall of the previous year and the snowfall before sowing were found to have the strongest correlation with the spring wheat yield. The results are shown in Table 23. The explained variance was also found to be higher, with 0.38 for seasonal key points versus 0.5 for step wise regression. Similar to the findings of ND, the found importance of the months of November and December cannot be directly explained from literature. Both the coefficients have a positive coefficient, implying a positive contribution to high levels of rainfall in November and December. A problem with the findings is the fact that November and December are in the middle of the Canadian winter, with snowfall the dominant form of precipitation as found in 4.5. Also from 4.5, the months of November and December on average have a relative small contribution towards the total rainfall of a year with less than 10%. The months also have a relative small contribution towards soil moisture when compared to the rainfall of the previous year, with also less than 10% as found in 4.5. For Canada, the question also remains if the findings of the step wise regression are spurious results, or that other factors influence the crop yield in the Canadian prairies.

Table 23: the results of the step wise regression of the relative change of the spring wheat yield in the Canadian Prairies as a function the rainfall in the previous year, snowfall before sowing and rainfall in the month September, with R2, P- and F-value and coefficients

R^2	0.50
Significant F value regression	0.00000057
Snow P-value	0.0071
Snow Coefficient	0.00035
Rainfall previous season P-value	0.0071
Rainfall previous season Coefficient	0.00000074
November P-value	0.035
November Coefficient	0.0078
December P-value	0.026
December Coefficient	0.011

For cocoa, the linear step-wise provided the same results as found in 7.6, but took multiple steps more as opposed to the method of using seasonal key points. For the seasonal key points, the months outside the wet and dry season, being October, November, April, August and September, were also taken into account, but the P-values of these months were found to be above the 0.05 level of significance. By taking these months into account, 5 more steps of backward elimination had to be performed.

The new proposed method of seasonal key points appears to work better for cocoa than for spring wheat. For spring wheat, the question remains if the differences in the findings between seasonal key points and step wise regression are the result of spurious results or lurking variables that are not taken into account in this research. These findings could be taken into account for a continuation of this research.

Evaluation of the found seasonal key points

For ND, the Canadian Prairies and IC, a number of seasonal key points were found. For wheat, in particular the rainfall of the previous season and the total snowfall before the sowing of wheat were found to be of importance for both ND and the Canadian Prairies. However, as the linear regression model shows, the predictor is off in a number of years for both ND and the Canadian Prairies due to circumstances not related to precipitation. In particular, the length of the growing season and the temperatures during the growing were found to be of importance for the yield of wheat. In this research, the length of the growing season and the temperatures are not taken into account.

In an older research by J. Lee, a linear approximation of wheat yield in Ireland was made by using a combination of rainfall, temperatures and hours of sunshine (Lee, 1969). In recent research, a combination between precipitation, temperatures, crop water deficit, growing degree days, soil water available and normalized-difference vegetation index is used (Newlands, 2014). For this research, only rainfall and snowfall data was assessed towards the yield of spring wheat. From the results, it could be stated that the found seasonal key points have an important contribution towards the yield of spring wheat, but the predictor is not completed yet. For further research, the length of the growing season due to temperatures at the start and end of the spring wheat growing seasons as well as temperatures during growing season could be taken into account. The addition could still make use of the weather stations, but a higher degree of communication with for instance farmers in ND and the Canadian Prairies is needed to keep track of the sowing and harvest dates.

After the linear regression and validation with the reports from the ICCO, the months of December, January, May and June were accepted to be seasonal key points. From the reports concerning crop specifics of cocoa from the ICCO and the area specifics of IC, the wet and dry season were indicated as the most important parts of the cocoa year. However, the months of February and June, while part of the dry and wet season, were rejected as seasonal key points due to significance above 0.05. For the dry season, an argument could be made that after December and January, most of the dry season has already passed. When assessing Figure 35, February is not capable of compensating low rainfall in both December and January on average. December and January are also found not to be correlated when the average rainfall of both months was tested, with a correlation of -0.006. A possible reason therefore is that when December and January have passed, the largest contribution to the dry season has also passed. When assessing the wet season, a similar discussion could be made. May and June are the start of the wet season, with June being the wettest month of the year. The standard deviation of the rainfall in June is more than the average rainfall in the month of July. A dry June can therefore not be compensated by the rainfall in July on average. As with the dry season, the first two month of the wet season, being May and June, seem to define the wet season of a cocoa year.

The use of key point predictor for market prices

One of the interesting possibilities for using this research to assess market prices of for instance cocoa, can be found when combining the average movement of cocoa market prices and the seasonal key points in the cocoa growing season. As can be seen from Figure 28, prices on average tend to increase the most in July and February. As discussed in chapter 7.2.2, the price increases in July and February are related to the expected impact of the weather during the wet and dry season. However, when assessing the results of the seasonal key point regression as presented in Table 19, the weather in parts of a cocoa producing year with the largest impact on the production are not found in July and February, but rather in December, January, May and June. The cocoa production tends to depend, or “pivot”, around a set of months throughout a cocoa growing season. For Ernst & Young, the months could be used to assess the possible price change of cocoa futures in the coming months based on expected production change, and can subsequently use the prediction to advice clients on an appropriate course of action. However, for the model to be used for active risk mitigation, a higher level of accuracy and thus a higher level of explained variance is needed.

For wheat, the same principal could hold towards the total production of wheat. Before the same principal can be applied though, the explained variance of wheat production should be towards 90%, as was the case for cocoa. As a continuation of the research for wheat, a critical market share of the production of over 60% is advised to provide a predictor model for the production of wheat.

The binary Bayesian Probabilistic Model
The BBPM is a continuation and visualization of the regression model. While the linear regression did offer an equation linking snowfall and rainfall to difference in spring wheat yield and cocoa production from the trend, the model did not provide a direct visual insight of the probability of certain event actually happening. Therefore, the BBPM is a logical continuation of the research, providing an insight in the probabilities of combinations of variables occurring. Concerning the BBPM, a discussion point can be made about the use of a binary system. A drawback of using only two categories is the simplification of the data, ignoring slight differences in the data and can therefore give a wrong impression concerning the probability of an above or below trend production. One of the problems however of a BPM with more categories is the higher amount of combinations for conditional probabilities and the subsequent need for a larger data set. Some of the combination of the seasonal key points as found in the research were not found to occur within the assessed data set using a binary model.

As a first approximation with a relatively small data set of 40 points, the use of a binary model is therefore seen as appropriate for this research and the scale of the data set. With for instance 4 categories for cocoa, 256 possible combinations between seasonal key points are possible, overstretching the data set.

Concerning the probabilities as found for wheat in ND, three of the combinations provided a 100% chance of a combination of seasonal key points resulting in a above or below average spring wheat yield. While the outcome of the analysis based on historic events is correct, a 100% chance does not occur in real life. Therefore, the 100% as found for the BBPM for ND can be seen as improbable. The main reason found was the limited dataset used for ND, consisting of 19 points. Due to the three defined key points and the binary system, eight combinations of conditional probabilities for spring wheat yield are possible. The BBPM does provides a start to translate seasonal key points into probabilities concerning spring wheat yield. For the active use of the BBPM of ND, the dataset should be expanded in future research.

The use of weather stations as an alternative for cocoa production
During the research, the high level of seclusion of the cocoa producing industry was noted. Most data concerning cocoa production was closely kept by the ICCO and was not readily available for research purposes. A reason could be that due to the high value of the cocoa industry, developments in the field of making predictions of cocoa production are not made readily available. While it is highly unlikely that no research is done, the availability of the research is limited. The development of a simple yet accessible method using weather stations that could also predict cocoa production change could therefore be an interesting alternative to oppose the secluded nature of the cocoa industry.

9. Conclusion

As a main objective of this research, the goal was to link price changes of wheat and cocoa to weather. As part of the main objective, the general dynamics of the wheat and cocoa market, the role of the USA and Canada towards the world wheat production and the role of Ivory Coast and Ghana towards the world cocoa production were defined. Next, the hypothesis of the role of the weather in North Dakota, the Canadian Prairies and Ivory Coast in particular time periods, or seasonal key points, towards the yield of spring wheat and the world production of cocoa was tested.

To define the general dynamics of the wheat and cocoa market, relative price changes of wheat and cocoa on a yearly basis were used. The relative price changes were corrected for inflation were used to define years with the largest price increases and decreases. The results of the price evaluation were subsequently used to determine the cause of the price increases and decreases based on qualitative date from reports of the ICCO, FAO, World Bank and WCF. Simultaneously, the area specifics of North Dakota, the Canadian Prairies and Ivory Coast and the crop specifics of spring wheat and cocoa were used to determine seasonal key points. For wheat, the world wheat production and the wheat production of the USA and Canada were used to determine the relative importance of the wheat production of the USA and Canada towards the world wheat production. For cocoa, the world cocoa production and cocoa production of IC and Ghana were used to determine the relative importance of the cocoa production of IC and Ghana towards the world cocoa production. The assessment was performed by using a linear regression between world production and the production of the selected countries.

The difference from the trend of spring wheat yield in ND and the Canadian Prairies and world cocoa production in IC were defined as dependent variable to test the hypothesis of weather dependency of the spring wheat production in ND and the Canadian Prairies and the world cocoa production on weather in IC. Rainfall and snowfall data from weather stations in ND, the Canadian Prairies and IC was collected and the monthly and total arithmetic mean of stations within the growing area of spring wheat and cocoa were used as independent variables for the testing of the hypothesis. A linear regression was performed using seasonal key points based on crop and area specifics, and subsequently validated by using the qualitative data from the NDSU HREC, ICCO, WCF, FAO, Ministry of Alberta and World Bank reports. Finally, a Bayesian model was made based on the occurrence of combinations of seasonal key points within the assessed data set.

Concerning the set research questions, the following answers can be provided:

- The main drivers of wheat and cocoa are actual production deficit, political and economic instability and expected changes in production based on news. Four out of five of the price increases of wheat and all of the price increases of cocoa can be related to weather affecting production, confirming the strong influence the weather has on the prices via the impact the weather has on production. Price changes can also be used to indicate interesting years to research concerning the possible weather-related cause of price increases and decreases.
- Following the regression analysis between the world wheat production and wheat production of the USA and Canada, it can be stated that the world wheat production is dependent of the wheat production of the USA and Canada. With 64% of the variance of the world wheat production explained, the important role of the wheat production of USA and Canada is confirmed. However, the linear regression is off in a number of years due to the influence of production in other countries, such as Russia and Australia. Following the regression analysis between world cocoa production and the cocoa production of IC and Ghana, it can be stated that the world cocoa production is highly dependent on the cocoa production of IC and Ghana.
- Over 92% of the variance in world cocoa production can be explained by assessing the cocoa production of IC and Ghana, confirming the dominant role of these countries in concerning world cocoa production.
- By using the reports from the FAO and ICCO, a visualisation of the dynamics of the wheat and cocoa market could be created. From the reports of the FAO and ICCO, the impact of seasonal key points on the prices via the production could be defined for both cocoa and wheat.
- By using qualitative data in the form of rapports of the ICCO, FAO, NDSU HREC, the Ministry of Alberta, WCF and World Bank, key periods and crop specifics for spring wheat and cocoa could be defined further. Weather station data from the NDAWN, NDSCO and the KNMI could subsequently be used to support the found key points in a season in a quantitative way. By using a combination of the two inputs, a regression based model could be formed to explain 54% of the variance of the spring wheat yield in ND using the rainfall of the previous year, the total snowfall before sowing and the rainfall in the month of May, 38% of the variance of the spring wheat yield in the Canadian Prairies using the rainfall of the previous year and the total snowfall before sowing and finally 47% of the variance of the world cocoa production using the months of December, January, May and June. The key points and crop specifics could be used to exclude spurious results and speed up the linear regression process.
- The Bayesian model continues where the linear regression model stops by providing an overview of the probabilities of a combination of seasonal key points occurring. The Bayesian model provides a binary probability, with the categories of above and below the mean, and a conditional probability of relative change in spring wheat yield and cocoa production.

On a whole, the research provided an interesting view into the world of wheat and cocoa, but also provides a starting point for further research. Leonardo da Vinci once stated that "*Water is the driving force of all nature*". It is in the view of this research that, even 500 years later, that statements still hold very true.

Bibliography

Algieri, B. (2016). A Roller Coaster Ride: An Empirical Investigation of the Main Drivers of Wheat Price. In M. Kalhul, & J. v. Braun, *Food Price Volatility and Its Implications for Food Security and Policy* (pp. 207- 238). Bonn: Springer Open.

Andah, W. E. (2003). Water, Climate, Food, and Environment.

Anscombe, F. (1960). Rejection of outliers. *Technometrics 2*, 123-147.

Antwi-Agyei, P. (2012). Mapping the vulnerability of crop production to drought in Ghana using rainfall, yield and socioeconomic data. *Applied Geography* , Volume 32 324-334.

Aregheore, E. M. (2009). *Country Pasture/Forage Resource Profiles CÔTE D'IVOIRE.* Rome: FAO.

Athey, S., & Schmutzler, A. (2001). Investment and market dominance. *RAND Journal of Economics*, 1–26.

Barron, J. (2003). Dry spell analysis and maize yields for two semi-arid locations in east Africa. *Agricultural and Forest Meteorology*, 23-37.

Barry. (2006). *Comparative study of river basin development and management.*

Bauer, P. (2015). The quiet revolution of numerical weather prediction . *Nature 525* , 47-55.

Benjamin, D., & Deaton, A. (1993). Household Welfare and the Pricing of Cocoa and Coffee in Cote d'Ivoire: Lessons from the Living Standards Surveys. *The World Bank Economic Review, vol 7 no 3*, 293-318.

Bernardo, J. (2000). *Bayesian Theory.* West Sussex: Wiley.

Bland, M. J. (1995). Multiple significance tests: the Bonferroni method. *BMJ*, 310.

Bloomberg. (2017, January 15). *Bloomberg*. Retrieved from Chocolate Eaters Drive Record Cocoa-Output Deficit: Commodities: https://www.bloomberg.com/news/articles/2013-12-17/chocolate-eaters-drive-record-cocoa-output-deficit-commodities

Bloomberg. (2017, January 20). *Bloomberg.* Retrieved from news: https://www.bloomberg.com/news/articles/2017-03-22/cocoa-slump-threatening-top-producer-is-deja-vu-for-80s-traders

BLS. (2016, December 12). *Bureau of Labor Statistics: Consumber Price Index*. Retrieved from Consumer Price INdex: http://www.bls.gov/cpi/

Brockhaus, J. (2016). When Do Prices Matter Most? Rice,Wheat and Corn Supply Response in China. In M. Kalkuhl, J. v. Braun, & M. Torero, *Food Price Volatility and Its Implications for Food Security and Policy* (pp. 435-456). Bonn: Springer Open.

Browne, F., & Cronin, D. (2010). Commodity prices, money and inflation. *Journal of Economics and Business*, 331-345.

Burns, W. C. (2017, April 20). *Burns*. Retrieved from Spurious Correlations: http://www.burns.com/wcbspurcorl.htm

Cambridge. (2017). *Cambridge*. Retrieved from Cambridge university press: http://assets.cambridge.org/97805217/60119/excerpt/9780521760119_excerpt.pdf

Canadian Goverment. (2017, May 1). *Agriculture*. Retrieved from Crop Production spring wheat: http://www.gov.mb.ca/agriculture/crops/production/print,spring-wheat.html#seeding

Cargill. (2016, December 6). *Cargill: Cocoa supply chains* . Retrieved from Cocoa supply chains : http://www.cargill.com/corporate-responsibility/responsible-supply-chains/cocoa/index.jsp

Cargill. (2017, April 15). *Cargill*. Retrieved from cocoa price risk services: https://www.cargill.com/food-beverage/cocoa-chocolate/price-risk-services

Cesteletti, A. (2008). Water reservoir control under economic, social and enviromental constrains. *Elsevier*.

CGC. (2017, April 6). *Canadian Grain Commission*. Retrieved from Canadian Grain Commission, quality of western Canadian Wheat 2011 : https://www.grainscanada.gc.ca/wheat-ble/harvest-recolte/2011/hqww/hqww11-qrbo11-2-eng.htm

Changnon, S. A. (2012). *Increasing Losses from Weather Extremes in Illinois: Their Causes and Future Implications.* Champaign: University of Illinois Urbana-Champaign.

CIA. (2013). *The World Fact book.* Washington DC.

CIA. (2014). *The World Factbook.* Washington D.C,: CIA.

CIA. (2015). *The World Factbook.* Washington D.C.: CIA.

CIA. (2016). *The World Factbook.* Washington D.C.: CIA.

CIA. (2017). *The World Fact Book.* Washington: CIA.

CLAAS. (2016, December 12). *CLAAS: calendrier de recolte*. Retrieved from Recolte: http://www.claas.fr/fascination-claas/actualites/calendrier-de-recolte

Cleaver, H. M. (1972). The Contradictions of the Green Revolution. *The american Economic review vol 62*, 177-186.

Cline, W. (2007). *Global Warming and Agriculture: Impact Estimates by Country.* Peterson Institute.

Cogneau, D., & Jedwab, R. (2010). Commodity Price Shocks and Child Outcomes: The 1990 Cocoa Crisis in Cote d'Ivoire. *G-MonD working paper no. 9*.

Davies, D. B. (2000). *Urban Drainage.* London: Spon Press.

Desanker. (2002). *Cimate Change Vulenrability in Africa: Impact of Climate Change.* Center for African Development Solutions.

Dircke, P. (2015). Climate change adaptation; innovative tools and strategies in Delta City Rotterdam. *Water Practice and Technology*.

Dixon, A. (1999). Water saving potential of domestic water reuse systems using greywater and rainwater in combination. *Water Science and Technology*.

EC. (2011). *Causes of the 2007-2008 global food crisis identified.* Brussel: European Commission.

Edwin, J., & Masters, W. (2005). GENETIC IMPROVEMENT AND COCOA YIELDS IN GHANA. *Experimental Agriculture*, 491-503.

Evans. (1980). The Natural History of Crop Yield: A combination of improved varieties of crop plants and technological innovations continues to increase productivity, but the highest yields are approaching limits set by biological constraints. *American Scientist*, 388-397.

FAO. (1962). *The state of food and agriculture 1962.* Rome: FAO.

FAO. (1966). *The state of food and agriculture 1966.* Rome: FAO.

FAO. (1968). *The state of food and agriculture 1968.* Rome: FAO.

FAO. (1971). *The state of food and agriculture 1971.* Rome: FAO.

FAO. (1972). *The state of food and agriculture 1972.* Rome: FAO.

FAO. (1973). *The state of food and agriculture 1973.* Rome: FAO.

FAO. (1974). *The state of food and agriculture 1974.* Rome: FAO.

FAO. (1976). *The state of food and agriculture 1975.* Rome: FAO.

FAO. (1977). *The state of food and agriculture 1977.* Rome: FAO.

FAO. (1982). *The state of food and agriculture 1982.* Rome: FAO.

FAO. (1988). *Commodity Review and Outlook 1988-89.* Rome: FAO.

FAO. (1988). *The state of food and agriculture 1988.* Rome: FAO.

FAO. (1989). *The state of food and agriculture 1989.* Rome: FAO.

FAO. (1990). *The State of Food and Agriculture.* Rome: FAO.

FAO. (1990). *The state of food and agriculture 1990.* Rome: FAO.

FAO. (1991). *Commodity review and outlook 1990-1991.* Rome: FAO.

FAO. (1991). *The state of food and agriculture 1991.* Rome: FAO.

FAO. (1992). *The state of food and agriculture 1992.* Rome: FAO.

FAO. (1993). *The state of food and agriculture 1993.* Rome: FAO.

FAO. (1994). *The state of food and agriculture 1994.* Rome: FAO.

FAO. (1995). *The state of food and agriculture 1995.* Rome: FAO.

FAO. (1996). *The state of food and agriculture 1996.* Rome: FAO.

FAO. (1997). *The state of food and agriculture 1997.* Rome: FAO.

FAO. (1998). *The state of food and agriculture 1998.* Rome: FAO.

FAO. (1999). *WHEAT: Post-harvest Operations.* Rome: FAO.

FAO. (2000). *Commodity Market Review 1999-2000.* Rome: FAO.

FAO. (2001). *The state of food and agriculture 2001.* Rome: FAO.

FAO. (2002). *The state of food and agriculture 2002.* Rome: FAO.

FAO. (2003). *The state of food and agriculture 2003.* Rome: FAO.

FAO. (2005). *The state of food and agriculture 2005.* Rome: FAO.

FAO. (2006). *The state of food and agriculture 2006.* Rome: FAO.

FAO. (2007). *The state of food and agriculture 2007.* Rome: FAO.

FAO. (2008). *The state of food and agriculture 2008.* Rome: FAO.

FAO. (2012). *Crop yield response to water.* Rome: FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS.

FAO. (2012). *Crop Yield response to water.* Rome: FAO.

FAO. (2012). *The state of food and agriculture 2012.* Rome: FAO.

FAO. (2012). *WORLD AGRICULTURE TOWARDS 2030/2050.* Rome: FAO.

FAO. (2014). *http://www.fao.org/nr/water/aquastat/countries_regions/GHA/GHA-CC_eng.pdf*. Retrieved from AQUASTAT.

FAO. (2015). *Climate change and food systems: global assessments and implications for food security and trade.* Rome: FAO.

FAO. (2015). *The State of food and Agriculture 2015.* Rome: FAO.

FAO. (2016, December 20). *FAO*. Retrieved from SOFA Reports: http://www.fao.org/publications/sofa/2016/en/

FAO. (2017, April 5). *FAO*. Retrieved from About FAO: http://www.fao.org/about/en/

FAO. (2017). *FOA, Terreset Eaux*. Retrieved from Wheat: http://www.fao.org/land-water/databases-and-software/crop-information/wheat/fr/

FAOSTAT. (2017, May 20). *FAO*. Retrieved from FAOSTAT: http://www.fao.org/faostat/en/#data/QV

Farooq, M. (2009). Plant Drought Stress: Effects, Mechanisms and Management. In M. Farooq, A. Wahid, N. Kobayashi, D. Fujita, & S. Basra, *Sustainable Agriculture* (pp. 153-188). Pakistan: Springer Netherlands.

FOA. (2010). *Water use by sector and source.* Retrieved from AQUATAT online database.

FOA. (2011). *The state of food and agriculture 2010-2011.* Rome: FAO.

FOA. (2017). *FAO*. Retrieved from The state of food and Agriculture: http://www.fao.org/publications/sofa/2016/en/

France 24. (2016, 12 12). *France 24: Agriculture*. Retrieved from Agriculture: http://www.france24.com/en/20160809-french-wheat-harvest-catastrophic-30-year-low-economy-agriculture

Giot, P. (2003). Market risk in commodity markets: a VaR approach. *Energy Economics* , volume 25, 435-457.

Google Maps. (2017). *Google Maps*. Retrieved from Google Maps: https://www.google.nl/maps

Goverment of Alberta. (2017, April 5). *agriculture alberta*. Retrieved from agriculture alberta about the ministry: http://www.agriculture.alberta.ca/app21/ministrypage

Hanley, M. (1999). Hedging the force of nature. *Risk Professional*, 1:21-25.

HREC. (1991). *Western Dakota Crops Day.* Hettinger: NDSU.

HREC. (1995). *Western Dakota Crops Day.* Hettinger: NDSU.

HREC. (1999). *Western Dakota Crops Day.* Hettinger: NDSU.

HREC. (2002). *Western Dakota Crop Day.* Hettinger: NDSU .

HREC. (2006). *Western Dakota Crop Day.* Hettinger: NDSU.

HREC. (2007). *Western Dakota Crops Day.* Hettinger: NDSU.

HREC. (2011). *Western Dakota Crops Day.* Hettinger: NDSU.

HREC. (2012). *Western Dakota Crop Day.* Hettinger: NDSU.

HREC. (2017, January 5). *West Dakota Crop Day*. Retrieved from annual reports: https://www.ag.ndsu.edu/hettingerrec/agronomy

ICCO. (2008). *Annual Report 2006/07.* London: International Cocoa Organization.

ICCO. (2009). *Annual Report 2007/08.* London: ICCO.

ICCO. (2010). *Annual Report 2008/09.* London: ICCO.

ICCO. (2010). *THE WORLD COCOA ECONOMY: PAST AND PRESENT.* London: ICCO.

ICCO. (2011). *Annual report 2009/2010.* London: ICCO.

ICCO. (2012). *annual report 2010/11.* London: ICCO.

ICCO. (2012). *THE WORLD COCOA ECONOMY: PAST AND PRESENT.* London: ICCO.

ICCO. (2013). *annual report 2011/2012.* london: ICCO.

ICCO. (2014). *Annual report 2012/2013.* London: ICCO.

ICCO. (2014). *COCOA MARKET REVIEW FEBRUARY 2014.* London: ICCO.

ICCO. (2014). *COCOA MARKET REVIEW JUNE 2014.* 2014: ICCO.

ICCO. (2015). *Annual report 2013/2014.* London: ICCO.

ICCO. (2016). *Annual report 2014/2015.* London: ICCO.

ICCO. (2016, December 6). *ICCO Annual reports*. Retrieved from Annual reports: http://www.icco.org/about-us/icco-annual-report.html

ICCO. (2016, December 6). *ICCO monthly averages of daily prices*. Retrieved from ICCO : https://www.icco.org/statistics/cocoa-prices/monthly-averages.html

ICCO. (2016, December 6). *ICCO: Growing Cocoa*. Retrieved from Growing Cocoa: https://www.icco.org/about-cocoa/growing-cocoa.html

ICCO. (2016, September 10). *Index Mundi*. Retrieved September 2016, from Index Mundi: http://www.indexmundi.com/commodities/?commodity=cocoa-beans&months=360

ICCO. (2016). *QBCS, Vol. XLII No. 3.* London: ICCO.

ICCO. (2017, March 29). *ICCO*. Retrieved from FAQ ICCO: https://www.icco.org/faq.html

ICCO. (2017, April 29). *ICCO*. Retrieved from documentation: https://www.icco.org/home/icco-documentation.html

ICE. (2017, May 30). *London Cocoa Futures*. Retrieved from ICE: https://www.theice.com/products/37089076/London-Cocoa-Futures

ICE. (2017, May 30). *Wheat futures*. Retrieved from ICE: https://www.theice.com/products/37089081/UK-Feed-Wheat-Futures/specs

Investopedia. (2016, December 12). *Investopedia Terms: Derivative Definition*. Retrieved from Derivative Definition: http://www.investopedia.com/terms/d/derivative.asp?ad=dirN&qo=serpSearchTopBox&qsrc=1&o=40186

Investopedia. (2016, December 6). *Investopedia Terms: Hedge Definition*. Retrieved from Hedge Definition: http://www.investopedia.com/terms/h/hedge.asp?ad=dirN&qo=serpSearchTopBox&qsrc=1&o=40186

Investopedia. (2016, December 6). *Investopedia Terms: Market definition*. Retrieved from Market definition: http://www.investopedia.com/terms/m/market.asp?ad=dirN&qo=investopediaSiteSearch&qsrc=0&o=40186

ITC. (2001). *Cocoa: A guide to trade practices.* Geneve: INTERNATIONAL TRADE CENTRE UNCTAD/WTO.

Judson, A. (2000). Density of Freshly Fallen Snow . *Bulletin of the American Meteorological Society in the Central Rocky Mountains*, 1577-1587.

Jung. (2006). Regional climate change and the impact on hydrology in the volta basin of West Africa. *Wissenschaftliche Berichte FZKA*.

KIM, H. H. (1992). Urban heat island. *International Journal of Remote Sensing*.

KNMI. (2017, April 6). *KNMI*. Retrieved from KNMI Climate Explorer: https://climexp.knmi.nl/about.cgi?id=someone@somewhere

KNMI. (2017, January 12). *KNMI*. Retrieved from Climate Explorer application: https://climexp.knmi.nl/start.cgi?id=someone@somewhere

Kouakau, k. E. (2012). Analyze of climate variability and change impacts on hydro-climate parameters: case study of Côte d'Ivoire. *International Journl of Scientific & Engineering Research, Volume 3, Issue 2,*.

Lacey, M. (2017, April 22). *Yale University*. Retrieved from Statistical topics: http://www.stat.yale.edu/Courses/1997-98/101/stat101.htm

Lautze. (2005). Driving forces behind African transboundary water law: Internal, external, and implications.

Lee, J. (1969). Effects of Weather on Wheat Yields. *Irish Journal of Agricultural Research*, 349-357.

Live Moya. (2013). *Marketing to the low-income consumer*. Retrieved from Live Moya.

Luttrell, C. B. (1973). *The Russian Wheat Deal — Hindsight vs. Foresight.* St. Louis: Federal Reserve Bank of St. Louis.

Malliaris. (2006). US inflation and commodity prices: Analytical and empirical issues. *Journal of Macroeconomics*, 267-271.

Martins, J. (2012). *Market segmentation of the .*

McIntosh, C. (2013). *Cambridge Advanced Learner's Dictionary.* Cambridge: Cambridge University Press.

Mentens, J. (2006). Green roofs as a tool for solving the rainwater runoff. *Landscape and Urban Planning*.

Metro. (2016, December 12). *Metro: Binnenland*. Retrieved from Binnenland: http://www.metronieuws.nl/nieuws/binnenland/2016/11/je-portie-friet-wordt-een-stukje-duurder

Minisrty of Alberta. (2014). *Alberta 2013 Crop Season in Review.* Edmonton: Agriculture and Forestry.

Ministry of Alberta. (2007). *Alberta 2006 Crop Season in Review.* Edmonton: Agriculture and Forestry.

Ministry of Alberta. (2008). *Alberta 2007 Crop Season in Review.* Edmonton: Agriculture and Forestry.

Ministry of Alberta. (2009). *Alberta 2008 Crop Season in Review.* Edmonton: Agriculture and Forestry.

Ministry of Alberta. (2010). *Alberta 2009 Crop Season in Review.* Edmonton: Agriculture and Forestry.

Ministry of Alberta. (2011). *Alberta 2010 Crop Season in Review.* Edmonton: Agriculture and Forestry.

Ministry of Alberta. (2012). *Alberta 2011 Crop Season in Review.* Edmonton: Agriculture and Forestry.

Ministry of Alberta. (2013). *Alberta 2012 Crop Season in Review.* Edmonton: Agriculture and Forestry.

Ministry of Alberta. (2015). *Alberta 2014 Crop Season in Review.* Edmonton: Agriculture and Forestry.

Ministry of Alberta. (2016). *Alberta 2015 Crop Season in Review.* Edmonton: Agriculture and Forestry.

Ministry of Alberta. (2017, February 10). *Ministry of Alberta*. Retrieved from Agriculture and Forestry: http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/sdd4191

Minitab. (2017, May 30). *F-test*. Retrieved from Minitab: http://blog.minitab.com/blog/adventures-in-statistics-2/what-is-the-f-test-of-overall-significance-in-regression-analysis

Minitab. (2017, May 30). *P-values and coefficients*. Retrieved from Minitab: http://blog.minitab.com/blog/adventures-in-statistics-2/how-to-interpret-regression-analysis-results-p-values-and-coefficients

Monnier, O. (2017, May 20). *Bloomberg*. Retrieved from Bloomberg news: https://www.bloomberg.com/news/articles/2017-03-13/cocoa-plunge-mutinous-soldiers-may-widen-ivory-coast-deficit

NDAWN. (2017, April 6). *NDAWN*. Retrieved from NDAWN Stations : https://ndawn.ndsu.nodak.edu/

NDSCO. (2017, January 5). *NDSCO*. Retrieved from NDSCO snow: https://www.ndsu.edu/ndsco/data/snow/

NDSU. (2017, 5 April). *NDSU*. Retrieved from NDSU HettingerREC: https://www.ag.ndsu.edu/HettingerREC/about-us

NDSU. (2017, April 20). *NDSU*. Retrieved from Climate Office: http://climate.ncsu.edu/climate/patterns/ENSO.html

NDSU. (2017, May 1). *North Dakota State University*. Retrieved from Spring wheat articles: https://www.ag.ndsu.edu/crops/spring-wheat-articles/optimum-and-last-planting-date-minimum-stands-crops-loss-from-frost-or-hail

NDSU HREC. (2016, December 20). *NDSU HREC*. Retrieved from annual reports: https://www.ag.ndsu.edu/hettingerrec/agronomy

NDSU HREC. (2017, April 30). *HREC*. Retrieved from HREC annual reports: https://www.ag.ndsu.edu/hettingerrec/agronomy

NDWC. (2016, December 6). *Buyers and processors: prodcution data*. Retrieved from prodcution data: http://www.ndwheat.com/buyers/default.asp?ID=301

NDWC. (2017, April 6). *North Dakota Wheat Commission*. Retrieved from NDWC Charts and Statistics: http://www.ndwheat.com/buyers/default.asp?ID=539

Newlands, N. (2014). An integrated, probabilistic model for improved seasonal forecasting of agricultural crop yield under environmental uncertainty. *Frontiers in eviromental science*, 1-21.

NWSEND. (2002). *Fargo, North Dakota Climate.* Kansas City: National Weather Service Eastern North Dakota.

Nyatuame, M. (2014). Statistical Analysis of Rainfall Trend for Volta Region in Ghana. *International Journal of Atmospheric Sciences*, 11 pages.

Obuobie. (2008). Estimation of groundwater recharge in the context of future climate change in the White Volta River Basin, West Africa.

OECD-FAO. (2007). *OECD-FAO Agricultural Outlook 2007-2016.* Rome: FAO.

Our Africa. (2017, April 6). *Our Africa, Ivory Coast*. Retrieved from Climate and Agriculture: http://www.our-africa.org/ivory-coast/climate-agriculture

Oxfam. (2011). *The Impact of Russia's 2010 Grain Export Ban.* Oxford: Oxfam GB.

Pearson, K. (1895). Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London 58*, 240-242.

Pearson, K. (1904). *On the theory of contingenxy and its relation to association and normal correlation.* London: Dulau and Co.

Pearson, K. (1905). *On the general theory of skew correlation and non-linear regression.* London: Dulau and Co.

Peltonen-Sainio, P. (2009). Improving Farming Systems in Northern European Conditions. In O. Victor, *Crop Physiology* (pp. 71-97). Oxford: Elsevier .

Pirrong, C. (2011). *Commodity Price Dynamics; A Structural Approach.* Cambridge: Cambridge University Press.

Pukelsheim. (1994). The Three Sigma Rule. *American Statistician 4*, 88-91.

Queensland Government. (2017, April 30). *Queensland goverment*. Retrieved from Department of Agriculture and Fisheries: Growing Cocoa: https://www.daf.qld.gov.au/plants/fruit-and-vegetables/fruit-and-nuts/other-fruit-crops/growing-cocoa#climate

Ritchison, D. (2016). Information concerning NDAWN. (J. Hogewoning, Interviewer)

Rosegrant, M. (2009). *Exploring alternative futures for agricultural knowledge, science and technology.* Canberra: Australia Centre for International Agricultural Research.

Rosenzweig, C. (2001). Climate change and extreme weather events, Implications for food production, plant diseases, and pests. *GLOBAL CHANGE & HUMAN HEALTH*, Volume 2, 90-104.

Schnittker, J. A. (1973). The 1972-1973 food price spiral. *Brookings Papers on Economic Activity*, Vol. 4, issue 2, pages 498-507.

Semadeni-Davies. (2008). The Impact of climate change and urbanisation on drainage in Helsingborg, Sweden, Combined sewer system. *Journal of Hydrology*.

SEO . (2016). *Market Concentration and Price Formation in the Global Cocoa Value Chain.* Amsterdam: SEO.

Shukla, J. (1998). Predictability in the Midst of Chaos: A Scientific Basis for Climate Forecasting. *Science vol. 282 no. 5389*, 728-731.

Statistics Canada. (2016). *Production of principal field crops, July 2016.* Otawwa: Statistics Canada.

Statisticshow. (2017, April 20). *Statisticshow*. Retrieved from f-statistic: http://www.statisticshowto.com/f-statistic/

Statsdirect. (2017, March 20). *Basics*. Retrieved from P values: http://www.statsdirect.com/help/basics/p_values.htm

Stattrek. (2017, April 20). *Stattrek*. Retrieved from regressions: http://stattrek.com/regression/residual-analysis.aspx

Taylor, J. A. (1968). *Weather Economics .* Aberwystwyth: Peragom Press.

The World Bank. (2001). *Aid and Reform in Africa Lessons from Ten Case Studies.* Washington D.C. : The World Bank.

The World Bank. (2007). *Cote d'Ivoire: Volatility, Shocks and Growth Policy Research Working Paper.* Washington: The World Bank.

Thompson. (1975). Weather variability, climatic change, and grain production. *Science 188*, 535-541.

Trostle, R. (2008, May 5). *"Global Agricultural Supply and Demand: Factors Contributing to the Recent Increase in Food.* Retrieved from Economic Research Service, U.S. Department: http://www.ers.usda.gov/Publications/WRS0801/WRS0801.pdf

UK AHBD. (2017, May 19). *UK AHBD.* Retrieved from Wheat Growth Guide: http://webdoc.agsci.colostate.edu/wheat/linksfiles/UKAHDBgrowthstageguide.pdf

UN. (2011). *The Global Social Crisis.* New York: UN.

UNDP. (2006). *Human development report.*

UNDP. (2011). *The Conflict in Côte d'Ivoire and its Effect on West African Countries: A Perspective from the Ground* . New York: UN.

USDA. (1997). *Usual Planting and Harvesting Dates.* Washington: USDA NASS.

USDA. (2016, December 6). *Agricultural Prices National Agricultural Statistics Service (NASS)*. Retrieved from NSDA Economics, Statisctics and Market Information System: http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do;jsessionid=F154BA78C7C50C021C8CA924EDB72FD5?documentID=1002

USDA. (2016). *Crop Production Historical Track Records.* Washington D.C. : United States Department of Agriculture & National Agricultural Statistics Service.

USDA. (2016, December 20). *USDA ERS*. Retrieved from Wheat Data: https://www.ers.usda.gov/data-products/wheat-data/

USDA. (2017). *Grain: World Markets and Trade.* Washington: USDA.

USDA. (2017, April 5). *USDA*. Retrieved from USDA About USDA: https://www.usda.gov/our-agency/about-usda

USDA. (2017, January 10). *USDA: PS&D*. Retrieved from PS&D: https://apps.fas.usda.gov/psdonline/app/index.html#/app/home

Vaes, G. (1999). The impact of rainwater reuse on cso emissions. *Water Science and Technology*.

van de Giesen, N. (2001). Competition for water of the Volta basin.

van de Loo, F. (2016, December 20). Discussion concerning the price increase of wheat in August. (J. Hogewoning, Interviewer)

Vickers, G., Buzza, S., Schmidt, D., & Mullock, J. (2000). *The Weather of The Canadian Prairies.* Ottawa: NAV Canada.

WCF. (2012). *Cocoa Market Update.* Washington: WORLD COCOA FOUNDATION.

Werrell, C. E., & Femia, F. (2013). *The Arab Spring and Climate Change: A Climate and Security Correlations Series.* Washington D.C : Center for American Progress.

William Collins Sons & Co. Ltd. (2012). *Collins English Dictionary.* Glasgow: HarperCollins.

Wright, B. (2009). *"International Grain Reserves and Other Instruments to Address Volatility in Grain Markets."* . Retrieved from Policy Research Working Paper 5028: http://econ.worldbank.org/

Wung, T.-C. (2006). Rainwater reuse supply and demand response in urban elementary school of different districts in Taipei. *Resources, Conservation and Recycling*.

Yi, L. (2011). An overview of reclaimed water reuse in China. *Journal of Environmental Sciences*.

Appendix

Appendix I

Description of FAO, USDA, NDSU HREC, Ministry of Alberta and ICCO

Food and Agricultural Organization

The Food and Agricultural Organization (FAO) is an agency of the United Nations (UN). Founded in 1945 and with the slogan "Fiat Panis", Latin for "let there be bread", the FAO strives to help eliminate hunger, malnutrition and food insecurity, make agriculture more sustainable and productive and create inclusive and efficient agricultural and food systems. As a central knowledge hub for agriculture for the UN, the FAO releases a number of reports every year and provides readily available knowledge about a variety of crops (FAO, 2017).

United States Department of Agriculture

The United States Department of Agriculture (USDA), is the department of the United States government that focusses on food, agriculture, natural resources, rural development amongst other subjects. The USDA is made up of 29 agencies and covers a large array of subjects, such as farm and foreign agriculture services, food safety and marketing and regulation programs. The USDA collects data locally and globally, and monitors the global production and developments concerning agriculture (USDA, 2017).

North Dakota State University Hettinger Research Extension Centre

The North Dakota State University (NDSU) provides an important agricultural service in the form of the Hettinger Research Extension Centre (HREC). The HREC produces important data about the agricultural activities in the United States largest grain producing state. As part of the NDSU, the HREC continues research and development on an academic level (NDSU, 2017).

Ministry of Alberta, department Agriculture and Forestry

From the official website of the ministry, "*The ministry is responsible for the policies, legislation, regulations and services necessary for Alberta's agriculture, food and forest sectors to grow, prosper and diversify; inspires public confidence in wildfire and forest management and the quality and safety of food; supports environmentally sustainable resource management practices; and leads collaboration that enables safe and resilient rural communities*" (Goverment of Alberta, 2017).

International Cocoa Organization

The international Cocoa Organization (ICCO), according to their official website, "*a global organization, composed of both cocoa producing and cocoa consuming member countries. Now located in Abidjan, Côte d'Ivoire, the ICCO was established in 1973 to put into effect the first International Cocoa Agreement which was negotiated in Geneva at a United Nations International Cocoa Conference*". The ICCO is a central hub for all the cocoa production in the world and has been of vital importance concerning market transparency, sustainable cocoa production and consumption and as a centre for knowledge and information concerning production, consumption and other statistics (ICCO, 2017).

Appendix II

An example of a SOFA report from the FAO (1972) and a monthly report of the ICCO

only slightly because of a decrease in U.S.S.R. production.

In North America an easing of controls on production and policy attitudes in general, plus the farm legislation of 1970, appear to have encouraged production increases. A number of factors probably contributed to the larger crops. The farm programme of the United States allowed more freedom of choice in planting, there were higher feed prices and a tendency to overplant against a return of the 1970 corn leaf blight.

A modest 1971 recovery in Australia indicates an adjustment to the reduction of wheat areas in the recent past, and an increase of noncereal crops. In New Zealand, farm production and grazing improved in 1971, generally favoured by better weather than in the drier 1970 growing season.

Every year food and other agricultural production is affected by a plethora of recurring factors and phenomena. During 1971, droughts and excessive dry periods occurred in Afghanistan, Cuba, Haiti, Somalia and the Yemen Arab Republic. Floods, typhoons, heavy storms, volcanic eruptions and earthquakes caused damage in many countries, including Brazil, Congo, Ethiopia, Guyana, the Khmer Republic, Nepal, the Philippines and the United States.

Pests and disease exacted their usual tribute from all countries, but were responsible for particularly heavy losses in Congo and Zaire, where manioc disease struck and spread; in Ethiopia, where food crops were lost to the army worm; in Guyana, which suffered blast damage to the rice crop; in Lesotho, where there were losses in maize and sorghum to the boll worm; and in the Philippines, where the tungro virus damaged part of the rice crop.

Production of main commodities[4]

World aggregates for individual crops showed modest increases for the major food items. Cereals, with the exception of rice, made notable gains, the developed market economies providing most of the total world increase, although the developing countries generally reported larger harvests than in 1970.

World production of cereals rose 8 percent over the previous year, making this commodity group the principal contributor to agriculture's performance in 1971. Some countries reported sizable gains in total cereal production with increases in certain crops of a third or more over 1970.

World wheat output has been estimated at some 353 million tons. This is about 11 percent larger than the 1970 crop and more than 6 percent above the previous record of 1968. It exceeded the 1970 harvest in all regions (including China) except the U.S.S.R., and there the crop surpassed early forecasts. In part, this rise was a recovery from deliberate cutbacks by the large producing countries of North America in previous years, and in part the result of the use of high-yielding varieties and more inputs, and in some cases better weather.

Western European wheat production, at 56.6 million tons, was about 19 percent over 1970. France, Italy, Portugal, the United Kingdom and Yugoslavia all had record crops, with the Federal Republic of Germany and Spain reporting large increases over the previous year. In eastern Europe, Albania, Czechoslovakia, Hungary and Poland all had bumper crops. The U.S.S.R. was an important exception, with a harvest somewhat below the 1970 level.

In North America, the Canadian crop was almost 60 percent higher than in 1970, and the United States crop was a record. In South America, the large wheat-producing countries had better crops than in 1970, as did many countries in Africa and Asia. India and Turkey had record wheat harvests, escaping the very dry weather that plagued some of western Asia, reducing wheat and other crops in Afghanistan, Iran, Iraq and Pakistan. Oceania's wheat crop, down in 1970, rose by about 10 percent in 1971.

Barley followed the pattern of wheat, making a good gain in percentage terms, while maize made still greater gains. The year 1971 saw a complete reversal of the world's coarse grains situation to one of abundant supplies. World production rose by 9 percent as output in the United States and western Europe more than recovered from the previous year's setback, and other regions also showed increases.

Rice was the one cereal crop that lagged in 1971, at less than 1 percent below the 1970 world level, mainly reflecting a cutback of 14 percent in Japan (which has allowed some reduction in the country's surplus stocks), of 5 percent in Pakistan, and the lack of outstanding growth in the other principal producing countries. Many did report small increases, however.

The world output of pulses dropped a little from 1970, many producers reporting approximately the same harvest. Root crops followed a comparable pattern.

The 1971 world output of centrifugal raw sugar was slightly lower than in 1970. Although some large producers reported increases, three out of the four largest producers reported sizable decreases. Cuba, which had a record crop in 1970, suffered an almost 30 percent drop in its cane sugar output in 1971, and India's cane crop dropped almost 12

[4] For a more detailed account of the commodity situation, see *FAO commodity review and outlook 1971-1972.*

COCOA MARKET REVIEW

JANUARY 2010

The current review reports on cocoa price movements on international markets during the month of January 2010. **Chart I** illustrates price movements on the London (NYSE Liffe Futures and Options) and New York (ICE Futures US) markets in January. **Chart II** shows the evolution of the ICCO daily price, quoted in US dollars and in SDRs, from November 2009 to January 2010. **Chart III** depicts the change in the ICCO daily price Index, the Dow Jones-UBS Commodity Index and the US Dollar Index. **Chart IV** presents recent changes in daily price volatility of cocoa compared to coffee.

Chart I: Cocoa bean prices on the London (NYSE Liffe) and New York (ICE) futures markets
January 2010

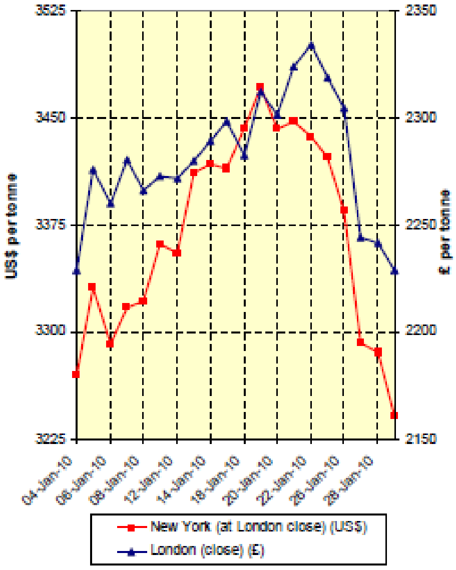
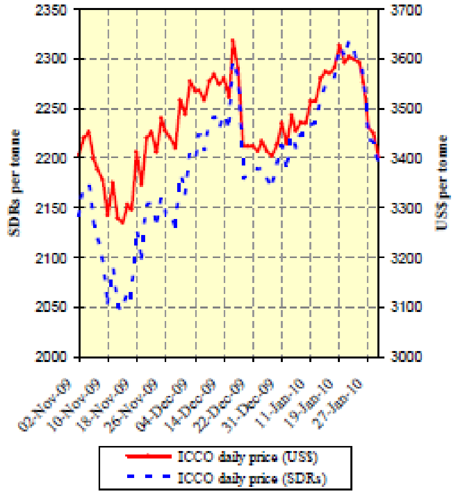


Chart II: ICCO daily prices
November 2009 – January 2010



Note: The ICCO daily price for cocoa beans is the average of the quotations of the nearest three active futures trading months on NYSE Liffe and ICE Futures U.S. at the time of London close, converted into US$ and SDRs using the appropriate exchange rates.

Price movements

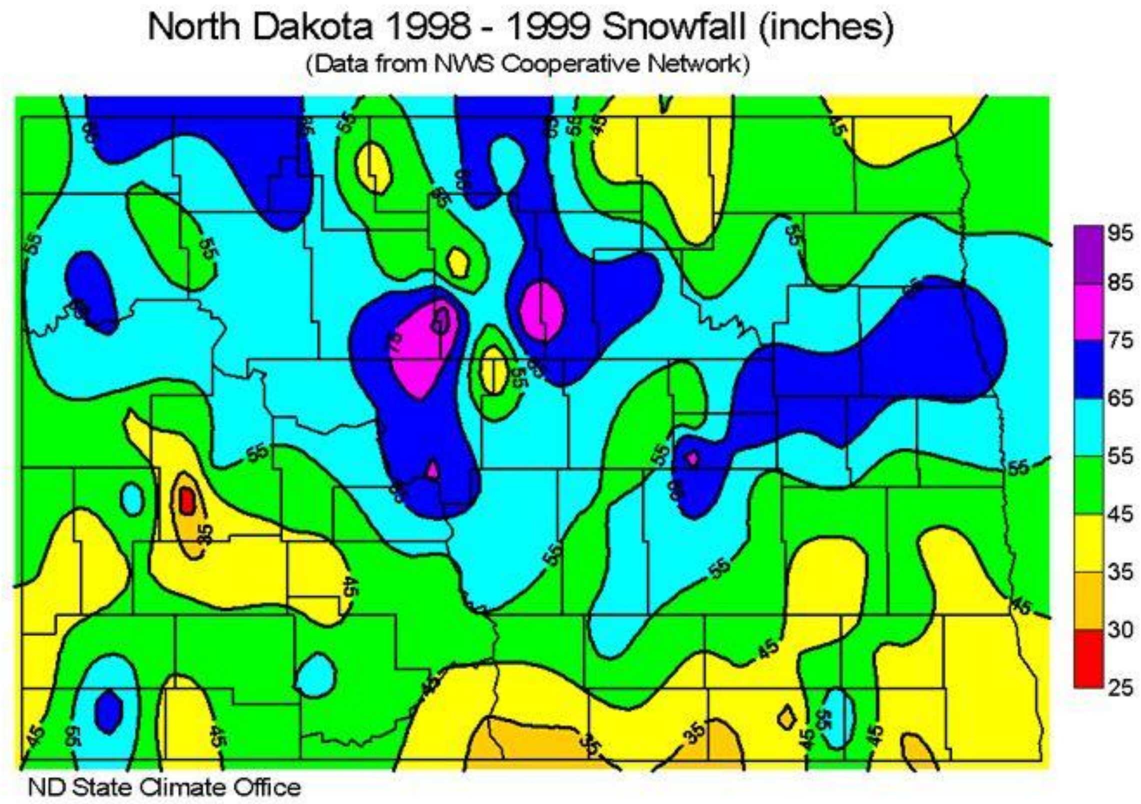
In January, the ICCO daily price averaged US$3,525 per tonne, up by US$27 compared to the average price recorded in the previous month (US$3,498), and ranged between US$3,402 and US$3,626.

The cocoa futures markets followed an upward trend in January right through to the third week of the month when the London futures market experienced a new record price, surging to its highest level for over 32 years to £2,334 per tonne while in New York, prices reached US$3,472 per tonne.

The major bullish factor arose from waning weekly cocoa arrivals in Côte d'Ivoire. Contrary to the previous season, the main crop (October – March) in Côte d'Ivoire witnessed an early and strong start with port arrivals figures indicating that production in the first three months was 17% higher than during the same period of the previous year, reaching 595,000 tonnes as at the end of December 2009. However, since the beginning of January, despite the fact that Côte d'Ivoire has been spared a strong *harmattan* (a dry wind that usually sweeps down from the Sahara between December and March, negatively affecting cocoa output), weekly arrivals have declined significantly, a sign that the main crop has been tailing off. The decrease would have been more acute without the smuggling of cocoa beans originating from Ghana, due to the significantly lower fixed price paid to Ghanaian farmers for their cocoa of Gh¢2,400 (US$1,655) per tonne, compared to the fluctuating price offered to farmers in Côte d'Ivoire of about US$2,200 per tonne on average in January.

Appendix III

Example of snowfall depth measurements for North Dakota



Appendix IV

An introduction to market prices

To buy goods for everyday use, people go to places that sell these goods. In a broad term, we can call these places markets. Hence also the term supermarket for instance. By definition, a market is an open or covered place where people meet to buy or sell something (McIntosh, 2013). These places can however be either physical or virtual. In finance, the definition of a market does not stray far from this definition as well. In finance, a market is a medium where buyers and sellers can exchange specific goods (Investopedia, 2016). These goods can be diverse, such as stocks (the stock market), houses (the housing market) or commodities (the commodities market). It is the later that is of interest to this thesis.

A commodity is a basic good, such as oil, iron ore or grains (McIntosh, 2013). These goods are in their basic form and are interchangeable within one type of commodity. Commodities can be divided in two categories, being hard and soft commodities. Hard commodities are commodities that are mined ore subtracted from the earth, while soft commodities are commodities such as grain and livestock. The commodities market is a place for buying and selling these commodities.

Historically, people went to the market to procure commodities as materials for businesses. For instance, a baker going to the market to buy flower from the miller, who in his turn bought wheat from the farmer. The other way around, the farmer sells to the miller who sells to the baker again. But times have changed and with it the businesses. The insecurities connected to simply going to the market are no longer acceptable in a time like today. To compensate for the risk, financial products were devised to minimize risk and comply with the growth in demand. These financial products are called derivatives and consist out of options, futures and swaps. In this thesis, only options and futures are addressed.

The term derivatives are used for a financial product that derives value from an underlying asset rather than having any value itself (Investopedia, 2016). A future for instance, is a contract that obligates a buyer to buy a commodity for instance. A contract is an obligation to buy, but the buying date is set in the future. Hence the name future of future contract in used. At the end of the future contract, the seller needs to deliver the sold product, while the buyer needs to buy. For a seller, this means security that his or her product gets sold. For a buyer, this means security and assurance that the materials he or she needs for production are available at the time the buyer wants them to be available. This principle is called hedging (Investopedia, 2016).

An option is the same as a future, but without the obligation to buy. Within the time when the option is active, the option can be executed, meaning that the contract can be set in motion. One difference between futures and options is that usually for an option a premium is charged, due to the flexibility in the contract.

The value of commodities that are most common today, are the prices from the futures market. There are many markets, all with different specialties. Chicago for instance has the CME group, while Wall street has the stock exchange. The prices that can be seen online or on the newspaper, are actual prices of trades that have been done on the one of these markets. The problem with futures and options, is the fact that these contracts are not only available for companies using the commodities for their business, but also for buyers that do not use them.

This is what is called speculation. Buyer buy a contract and try to sell the contract again when there is a profit in it. This principle makes understanding the market difficult, because the relation is not purely supply and demand driven anymore.

Appendix V

Multiple variate linear regression

The following description is found from the website of Yale concerning Multiple variate linear regression:

"Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. Every value of the independent variable x is associated with a value of the dependent variable y. The population regression line for p explanatory variables x1, x2, ... , xp is defined to be $\mu_y = \beta_0 + \beta_{1x1} + \beta_{2x2} + \dots + \beta_{pxp}$. This line describes how the mean response μ_y changes with the explanatory variables. The observed values for y vary about their means μ_y and are assumed to have the same standard deviation σ. The fitted values b0, b1, ..., bp estimate the parameters $\beta_0, \beta_1, \dots, \beta_p$ of the population regression line.

Since the observed values for y vary about their means μ_y, the multiple regression model includes a term for this variation. In words, the model is expressed as DATA = FIT + RESIDUAL, where the "FIT" term represents the expression $\beta_0 + \beta_{1x1} + \beta_{2x2} + \dots \beta_{pxp}$. The "RESIDUAL" term represents the deviations of the observed values y from their means μ_y, which are normally distributed with mean 0 and variance σ. The notation for the model deviations is ε.

Formally, the model for multiple linear regression, given n observations, is

$yi = \beta_0 + \beta_{1xi1} + \beta_{2xi2} + \dots \beta_{pxip} + \varepsilon_i$ for i = 1,2, ... n.

In the least-squares model, the best-fitting line for the observed data is calculated by minimizing the sum of the squares of the vertical deviations from each data point to the line (if a point lies on the fitted line exactly, then its vertical deviation is 0). Because the deviations are first squared, then summed, there are no cancellations between positive and negative values. The least-squares estimates b0, b1, ... bp are usually computed by statistical software.

The values fit by the equation b0 + b1xi1 + ... + bpxip are denoted $\hat{y}_i$, and the residuals ei are equal to yi - $\hat{y}_i$, the difference between the observed and fitted values. The sum of the residuals is equal to zero.

The variance σ^2 may be estimated by $s^2 = \frac{\sum e_i^2}{n - p - 1}$, also known as the mean-squared error (or MSE).

The estimate of the standard error s is the square root of the MSE" (Lacey, 2017).

Appendix VI

P-values and F-values

Concerning the P and values, the following description can be found from Stats direct:
"The P value, or calculated probability, is the probability of finding the observed, or more extreme, results when the null hypothesis (H0) of a study question is true – the definition of 'extreme' depends on how the hypothesis is being tested. P is also described in terms of rejecting H0 when it is actually true, however, it is not a direct probability of this state.

The null hypothesis is usually a hypothesis of "no difference" e.g. no difference between blood pressures in group A and group B. Define a null hypothesis for each study question clearly before the start of your study.

The only situation in which you should use a one-sided P value is when a large change in an unexpected direction would have absolutely no relevance to your study. This situation is unusual; if you are in any doubt then use a two-sided P value.

The term significance level (alpha) is used to refer to a pre-chosen probability and the term "P value" is used to indicate a probability that you calculate after a given study.

The alternative hypothesis (H1) is the opposite of the null hypothesis; in plain language terms this is usually the hypothesis you set out to investigate. For example, question is "is there a significant (not due to chance) difference in blood pressures between groups A and B if we give group A the test drug and group B a sugar pill?" and alternative hypothesis is " there is a difference in blood pressures between groups A and B if we give group A the test drug and group B a sugar pill".

If your P value is less than the chosen significance level then you reject the null hypothesis i.e. accept that your sample gives reasonable evidence to support the alternative hypothesis. It does NOT imply a "meaningful" or "important" difference; that is for you to decide when considering the real-world relevance of your result.

The choice of significance level at which you reject H0 is arbitrary. Conventionally the 5% (less than 1 in 20 chance of being wrong), 1% and 0.1% (P < 0.05, 0.01 and 0.001) levels have been used. These numbers can give a false sense of security.

In the ideal world, we would be able to define a "perfectly" random sample, the most appropriate test and one definitive conclusion. We simply cannot. What we can do is try to optimise all stages of our research to minimise sources of uncertainty. When presenting P values some groups find it helpful to use the asterisk rating system as well as quoting the P value:

P < 0.05

P < 0.01

P < 0.001

Most authors refer to statistically significant as P < 0.05 and statistically highly significant as P < 0.001 (less than one in a thousand chance of being wrong).

The asterisk system avoids the woolly term "significant". Please note, however, that many statisticians do not like the asterisk rating system when it is used without showing P values. As a rule of thumb, if you can quote an exact P value then do. You might also want to refer to a quoted exact P value as an asterisk in text narrative or tables of contrasts elsewhere in a report." (Statsdirect, 2017)

The F value can be used in a same way as the P value as a test of significance, but has an application for the Analysis of Variance (ANOVA) for a multivariate linear regression. Concerning the F values, the following description can be found:

‘An F statistic is a value you get when you run an ANOVA test or a regression analysis to find out if the means between two populations are significantly different. It’s similar to a T statistic from a T-Test; A-T test will tell you if a single variable is statistically significant and an F test will tell you if a group of variables are jointly significant.

What is “Statistically Significant“? Simply put, if you have significant result, it means that your results likely did not happen by chance. If you don’t have statistically significant results, you throw your test data out (as it doesn’t show anything!); in other words, you can’t reject the null hypothesis” (Statisticshow, 2017).

Appendix VII

First outcome of the regression model based on the arithmetic mean of the monthly rainfall in IC, as a function of the relative change in cocoa production.

	Coëfficiënten	Standaardfou	T-statistische gegever	P-waard	Laagste 95%	Hoogste 95	Laagste 95,0	Hoogste 95,0
November	-0,094368806	0,692653013	-0,13624254	0,892604398	-1,513204184	1,324466573	-1,513204184	1,324466573
March	0,459872306	0,935061409	0,491809738	0,626687537	-1,455514162	2,375258775	-1,455514162	2,375258775
April	-0,477640047	0,929548233	-0,513841058	0,611394646	-2,381733285	1,426453191	-2,381733285	1,426453191
September	-0,531279048	0,787602544	-0,674552224	0,505492185	-2,144609725	1,082051628	-2,144609725	1,082051628
August	-1,193918305	1,337838828	-0,892423123	0,379772032	-3,934356914	1,546520305	-3,934356914	1,546520305
July	0,494572928	0,4510375	1,096522859	0,282191593	-0,429335507	1,418481364	-0,429335507	1,418481364
October	0,63301287	0,53052557	1,1931807	0,242811235	-0,453719495	1,719745236	-0,453719495	1,719745236
February	1,969946966	1,287817718	1,529678415	0,137316649	-0,668028045	4,607921977	-0,668028045	4,607921977
January	3,498538993	2,218693206	1,57684667	0,126063373	-1,046248015	8,043326	-1,046248015	8,043326
December	1,960135018	1,091079582	1,796509668	0,083210781	-0,27484019	4,195110225	-0,27484019	4,195110225
June	0,471165046	0,258189793	1,824878668	0,078706344	-0,05771277	1,000042862	-0,05771277	1,000042862
Snijpunt	-273,9854218	98,50886773	-2,78132749	0,009575802	-475,77169	-72,19915366	-475,77169	-72,19915366
May	-1,089692353	0,345699025	-3,152141813	0,003842353	-1,797824705	-0,38156	-1,797824705	-0,38156