Modelling Drought Option Contract Prices

A Thesis Submitted to the Committee on Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Masters of Science in the Faculty of Arts and Science

TRENT UNIVERSITY

Peterborough, Ontario, Canada

Applications of Modelling in the Natural and Social Sciences M.Sc. Program

September 2009

Your file Votre référence
ISBN: 978-0-494-53558-5
Our file Notre référence
ISBN: 978-0-494-53558-5

ABSTRACT

Modelling Drought Option Contract Prices

Jielin Zhu

To help farmers reduce their risk of income loss caused by drought, we introduce a new financial weather derivative called a drought option. We model and estimate drought option contract prices based on existing techniques used in temperature and precipitation option pricing.

The difference is, unlike the direct measure of temperature or precipitation, we need to find a index which can reflect the severity of drought in some long time period. Then we use historical burn analysis, index value simulation and daily value simulation to estimate the value of the contract based on data in dry regions in China.

Comparing the three evaluation methods, we attempt to determine which one is more reasonable for data with different lengths and different parameters in the contract.

Key Words: Drought option prices, reconnaissance drought index, potential evapotranspiration, goodness-of-fit test, mean-reverting process, daily temperature simulation, speed of monthly rainfall simulation.

Acknowledgements

I would like to record my gratitude and appreciation to my supervisors, Prof. M. Pollanen and K. Abdella. Their advice and guidance from the beginning of the research gave me great experience throught the work. They provided me encouragement and support in various ways. Their deep insights and passions in mathematics inspired and enriched my growth as a student and a researcher.

I gratefully thank the faculty and my fellow graduate students in the AMINSS program. They provided a friendly, helpful, cooperative environment for my research and study.

Finally, I express my appreciation to my family. I would not be where I am without their constant support and understanding.

Contents

List of Figures

List of Tables

Glossary

Actual evapotranspiration — The actual evaporation and transpiration for some specific crops under real climate conditions, 16

Aridity Index — Average of historical RDI which is used to reflect the dry condition in the specific place, 14

Crop Coefficient — A coefficient which can reflect the relationship between the actual evapotranspiration for a specific crop and the potential evapotranspiration, 16

Day Length The time each day from the moment the sun appears above the horizon during sunrise to the moment when the sun disappears below the horizon during sunset, 18

Potential evapotranspiration The evapotranspiration that would occur with an adequate water supply at all times, 16

Reconnaissance Drought Index A drought index which is used to measure the severity of drought by comparing precipitation and potential evapotranspiration, 13

Weather Derivative A financial contract in which people agree to make a deal in the future based on some specific weather conditions, 2

Chapter 1

Introduction

1.1 Motivation for the Research

Weather plays an important role in people's daily lives. It can also affect profits of companies directly and indirectly. For example, if it is a cold winter, a heating oil company may face increased demand. Or if there is too much rain in the summer, a zoo could lose many tourists resulting in decreased revenues. While weather can be predicted a few days in advance, it is difficult to predict for long periods and almost impossible to control. But that doesn't mean nothing can be done to reduce the risk of weather related financial loss.

In the past, the main financial instrument to reduce the risk of weather-related loss was insurance. However, an insurance can only be effective when a peak weather incident happens, such as a flood or tornado. It is hard to set an insurance contract

if there is only a little change in weather expected. Most of the time, when we want to control the financial effect of weather, we want to reduce the effect of cumulative weather measures, not disasters (Alaton, 2002).

In the mid-1990's, new derivative contracts called weather derivatives were introduced to reduce weather-related risks.

Financial Derivatives and Weather Derivatives

In the past few decades there was an explosive growth in the development of new financial assets called financial derivatives (Stampfli, 2001). They are not like other traded assets, such as commodities, stock, bonds and currencies, which have their own values. The value of a derivative depends on the values of some underlying assets. Actually, most financial derivatives can be seen as a contract between a buyer and a seller with some special terms, such as when to trade or at what price to trade. The most common types of derivatives are futures, options and swaps.

Weather derivatives are very similar to financial derivatives. The difference is that the value of a weather derivative depends on some weather indices which cannot be exchanged or traded (Alaton, 2002; Mraoua, 2005; Yoo, 2003; Cao, 2004). The details of these contracts will be given in next chapter.

Currently, weather derivatives are used to control the risks associated with temperature, precipitation or wind.

Drought

Since the birth of agriculture, farmers have been battling drought in different

regions of the world. Even though they try very hard to avoid its influence, drought is still one of the largest risks facing farmers. In developed countries, irrigation is common and electricity readily available, but both are a luxury for the farmers in many developing countries. To help those farmers, we will develop a financial tool called drought option contract to avoid the risk of drought. When the drought is serious, farmers can get some profits by exercising this contract and make up the loss from their yields. Actually, in current weather derivative markets, when people discuss using weather derivatives to control the risk of drought, they often choose to use rainfall derivatives. Specific drought option contract have not being discussed in the academic literature.

Of course when people mention the word "drought", the most common description is "the lack of water". But water is not the only factor which can affect drought conditions. Actually, it is hard to define drought to everyone's satisfaction (Redmond, 2002). Drought means different things to different people. While for one person drought might mean there's no rainfall during a 15 day period, for somebody else drought might mean in one month the amount of precipitation falls below some set level.

The American Meteorological Society divides drought into four groups: meteorological/climatological, agricultural, hydrological, and socioeconomic (Richard, 2002).

Meteorological drought is a long term phenomenon. It means the reduction or absence of precipitation caused by the change of climate conditions. If there is a lack

of water in the surface layers during an important time in the growing season, which would decrease the product of the crop, it can be seen as an agricultural drought. A hydrological drought focuses on a balance of water at a specific place. It means the water supply cannot meet the demand after a deficit of rainfall in a long period. Socioeconomic drought highlights the impact of meteorological, agricultural, and hydrological drought on some economic good.

In this thesis, we consider the problem as a combination of agricultural drought and socioeconomic drought. First of all, we want to help the farmers to hedge their risk when the agriculture drought results a reduction of the yield. Secondly, we use an option as a financial tool, whose price depends on the drought index which can measure the condition of the agricultural drought.

1.2 Objective of Thesis

Based on the above, to help farmers reduce the income risk caused by drought, precipitation derivatives may not be a good choice. But the main idea for all weather derivatives are the same. Therefore, the aim of this thesis is to propose a drought option contract and estimate a reasonable price for this contract.

It is almost the same with other kinds of weather options, except we need to find a proper index as the measure of drought. The measure of drought is not so obvious, like temperature or precipitation, because it is a combination of other weather factors.

The evaluation of the derivative contract is as important as the setting of contract.

We can consider the financial option contract first. The buyer of the option only holds the right to buy or sell something in the future. It is not so obvious that we could say how much this right should be worth. For weather options, it is more difficult because the underlying is even not a traded asset, but some kind of weather measure.

Currently, the equilibrium method is most commonly used to decide the price of a weather option contract. In other words, we choose the expectation of the present value of all possible profits as the price of contract. Therefore the profit or loss of the contract is directly affected by possible index values in the future. Thus our problem becomes: how to decide the possible index values. Considering the methods used to price other weather options, we decide to use three different models to evaluate the option contract (Cao, 2004; Muβhoff, 2006). Then, we will compare these three models to see which one is better for drought options.

1.3 Thesis Outline

Chapter 2

In this chapter, we want to introduce some background materials: the definition of option, description of drought indices and which kind of drought index we want to use, all the parameters which can affect the change of this drought index, and some mathematical algorithms which will be used in the subsequent chapters.

Chapter 3

In this chapter, we use a historical analysis model to calculate the drought option

price based on data from stations in China for the past 56 years. We analyze the basic characteristic of the drought index based on different periods. Then we compare the option prices based on different data lengths and different strike levels.

Chapter 4

In this chapter, we introduce the index value simulation model for the estimation of a drought index. From a statistical point of view, we want to describe the behavior of a drought index based on the value in the past. Then we simulate possible index values following the specific pattern, and get the option price. Based on three different goodness-of-fit tests, a non-negative test and a stability test, we attempt to find a better way to decide which distribution fits our data best. Then, we compare the results with those from the historical burn analysis.

Chapter 5

In this chapter, we try to use a daily value simulation model to simulate the index value. Unlike other kinds of weather indices, our drought index is influenced by at least two climate variables, which means we need to simulate daily processes for two different weather factors. In this chapter, we attempt to use a similar mean-reverting process to describe the behavior of both temperature and precipitation. Then we combine the results from those two models to calculate the possible values of a drought index and the option price. Finally, we compare them with those from the historical burn analysis and index value simulation.

Chapter 6

In this chapter, we discuss the work we have done, and indicate possible future work on how to improve the accuracy of the model.

Chapter 2

Background

In this chapter we will introduce some general terms which would exist with a high frequency in this thesis. First, we will show the basic knowledge about weather derivatives based on financial derivatives. Second, we describe how we can measure drought and what kind of factors we need. Then, we introduce the expectation principle as the rule of option pricing. Finally, we present some general algorithms used in this thesis.

2.1 Weather Derivatives

When we talk about weather derivatives, we can not avoid talking about financial derivatives, which are at the root of weather derivatives.

2.1.1 Options

The three most common kinds of derivative contracts are: Futures, Options and Swaps. In a future contract, the buyer and seller both agree that they will trade some stocks or commodities at a specific time in the future. A swap means the buyer and seller will sell each other something on a specific date in the future (Stampfli, 2001).

An option is a special contract. The buyer of the option has the right to choose whether he/she wants the deal to happen. With this contract, they will set a maturity date and an agreed price level, which is called the strike price. If the buyer has the right to sell something to the seller, this contract is called a put option, as the buyer will exercise this contract when the market price is lower than the strike price. If the buyer has the right to buy something from the seller, it is a call option. The buyer of a call option would exercise the contract when the market price is higher than the strike (Stampfli, 2001).

Actually, there are many standard types of option contracts existing which are traded. For example, a European option only allows the buyer to exercise the contract at the maturity, while, an American option allows the buyer to exercise the option at any time during the contract period (Stampfli, 2001). When we talk about the price of the option, we always use a European option or an American option as an example. We will introduce the details about pricing these in Section 2.6.

2.1.2 Drought Options

Options on financial assets and weather options are very similar at a contractual level. However, there is one fundamental difference, the underlying item in the contract, for example a temperature measure, cannot be traded directly like a financial asset such as oil. From a modeling point of view, this can make a huge difference. Another big difference is the weather phenomenon are often more localized, so the price of oil in New York and London will be highly correlated, but the temperature less so.

Therefore, for a drought option, as the underlying item is the measure of drought, the strike level should be also changed to some specific value of drought index, not a price. When the holder wants to buy a drought option contract, first of all the buyer and the seller should decide whether it is a call option and when the buyer could exercise the contract. Then they should set up the strike level, the period of this contract. Another item called tip should be brought in the contract, which reflects the relationship between the value of drought index and the loss of farmers' income.

This drought option can be used by the government to help the farmers. Right now under existing crop insurance policies, farmers are paid only after there has been a claim. That is, only after there has been a complete or partial crop failure. For small farmers in developing countries, such as subsistence farmers, this might occur only after they have lost their livelihood or are starving. However, the drought option contract is more sensitive than the insurance contract. For example, a put option contract is bought by the agriculture agency. The profit from this contract is kept

as a funding to assist farmers. As mentioned above, they need to choose a drought index as the underlying item, a specific value of the index, K, as the strike level, the tip S, and the period of the contract. The value of this asset will increase before the crops fail because it only depends on the severity of drought. Another difference with drought options is that, even though the drought is not so serious to destroy almost all the crops and trigger the insurance payout, the agency can still get a payback by exercising the deal as long as the real value of drought index I is smaller that the strike level K, which means the profit will be $S \cdot (K - I)$.

2.1.3 Weather Index

As the market for weather derivatives continues to develop, there is a need for more and more specific weather risks to be hedged, which means many kinds of weather indices exist. Therefore, for one kind of weather factor, there may be many kinds of indices.

For example, the most common types of weather derivatives are for the temperature and precipitation. For temperature, HDD (Heating Degree Day) and CDD (Cooling Degree Day) are used to measure how cold or hot it is outside. HDD is the difference when daily mean temperature is lower than 18° C, written as $HDD_i \equiv \max\{18 - T_i, 0\}$, where T_i is the temperature for day i. Similarly, CDD can be estimated by $CDD_i \equiv \max\{T_i - 18, 0\}$, which reflects the situation when more energy should be used to cool the house. They can be used to hedge against heating or air

conditioning costs (Alaton, 2002; Mraoua, 2005). For precipitation, we usually use the amount of the precipitation over a period as the index (Cao, 2004).

For drought, there is no single obvious measurement like degrees for the temperature or the amount of rainfall for the precipitation. But there are still some kinds of drought indices which we can consider to use in our problem. We will describe these in the next section.

2.2 Drought Indices

Different definitions of drought make it difficult to have a universal drought index. Drought is a very complex phenomenon. There are many factors which can affect the intensity and severity of drought. There is no simple index which can include all these factors.

Based on work of Richard in 2002, drought indices should meet the following four criteria (Richard, 2002):

1) the timescale should be proper to the problem we have at hand;

2) the index should be a quantitative measure which can be used for large-scale, long-continuing drought conditions;

3) the index should be appropriate to the problem being studied;

4) a long accurate past data set should be applicable to compute the index.

Since 1900's, scientists have designed many drought indices to satisfy their needs to measure drought (Richard, 2002; Keyantash, 2002). A lot of research has been

done to check the efficiency of those indices in Africa and Asia(Bhuiyan, 2004; Ntale, 2003).The most popular ones are Palmer's drought index (PDI) and the standardized precipitation index (SPI). Both indices focus on meteorological or hydrological drought (Alley, 1984; Guttman, 1997; Guttman, 1999).

With our problem, the drought index should satisfy the following conditions:

1) It can be used to measure agriculture drought;

2) It can be calculated in any appointed period;

3) It can include as many factors as possible, which can affect the severity of drought;

4) The data we have is enough for the calculation.

To measure agricultural drought, we can not only consider the supply of water in that region. Based on the work of Boken (2005) and Hanson (1991)'s work, agriculture drought is concerned with climate factors like evapotranspiration, temperature, precipitation planting dates and so on. What we find now is an index called a reconnaissance drought index (RDI), which we will show the details of in the following section. Based on this RDI, we can adjust this index by changing the potential evapotranspiration to other measurements.

2.3 Reconnaissance Drought Index (RDI)

The Reconnaissance Drought Index was first introduced by Tsakiris in 2004 (Tsakiris, 2005; Tsakiris, 2007). It was used to measure the severity of drought in the Mediter-

ranean region.

This index has three expressions. The first one is called the initial value of RDI, α_0. If we use α_0^i to stand for the initial RDI for the ith year, we can calculate it by the following formula:

$$\alpha_0^i = \frac{\sum_{j=1}^{12} P_{ij}}{\sum_{j=1}^{12} PET_{ij}}, i = 1 \ldots N \text{ and } j = 1 \ldots 12, \tag{2.1}$$

where P_{ij} and PET_{ij} are the precipitation and potential evapotranspiration of the jth month in the ith year with the unit $mm/(m^2 \cdot month)$, N is the number of years of the available data. Potential evapotranspiration reflects the upper limit of evapotranspiration in the region. Based on this formula, we know that lower the RDI is, more serious the drought will be.

The second one is called normalized RDI ($\mathrm{RDI_n}$), which is used to reflect the abnormality of the drought in this region. The normalized RDI in the ith year can be calculated by the following formula:

$$RDI_n^i = \frac{\alpha_0^i}{\bar{\alpha_0}} - 1, \tag{2.2}$$

where $\bar{\alpha_0}$ is called the aridity index, reflecting the arithmetic mean of all α_0 for N years. In this expression, when $\mathrm{RDI_n}$ is negative, the value of RDI is lower than the aridity index, which means the drought is serious in the place.

The third one, Standardized RDI ($\mathrm{RDI_{st}}$), is the standardization of the second one. If we use y^i to stand for $ln(\alpha_0^i)$, the $\mathrm{RDI_{st}}$ can be shown to be

$$RDI_{st}^i = \frac{y^i - \bar{y}}{\hat{\sigma_y}}, \tag{2.3}$$

where $\bar{y}$ is the arithmetic mean of all y^i, and $\hat{\sigma}_y$ is the standard deviation. The important assumption for this expression is that the α_0 follows the lognormal distribution.

From Tsakiris (2006), we can see that there are some advantages of the RDI:

1) It can be actually calculated for any period of time, even though Equation (2.1) is used for the whole year.

2) It can be used for measurement of the agricultural drought, as it considers the relationship of water supply as precipitation and water demand of the crop, which we use potential evapotranspiration to stand for.

3) This index actually includes more climate factors than PDI or SPI, such as day hour and crop coefficients. Both of them are included in potential evapotranspiration, which may lead to a more accurate measurement.

Actually, what we choose for the option underlying is the initial value of RDI. The normalized one can be seen as the the retained profit of the option if we use $\bar{\alpha_0}$ as the strike value. We will show the details in the following section. In Chapter 4 we will see that the assumption of the lognormal distribution cannot be satisfied most of the time. The value of RDI for any period in one year can be calculated by the following formula:

$$RDI^i = \frac{\sum_{j=1}^{m} P_{ij}}{\sum_{j=1}^{m} PET_{ij}}, i = 1 \ldots N \text{ and } j = 1 \ldots m, \tag{2.4}$$

where j=1 stands for the beginning day of the period we need, and m is the number of days for this period.

2.4 Parameters in RDI

From the Equation (2.1), we can see that the most important parts of this index are precipitation and potential evapotranspiration. Based on other drought indices, we could see that the precipitation plays the most important role in drought problems. As rain is a common weather phenomenon in people's lives and easily measured, here we only focus on the potential evapotranspiration part.

2.4.1 Potential Evapotranspiration (PET)

Evapotranspiration (ET) is used to describe the sum of evaporation and plant transpiration from the surface of the land to atmosphere. Evaporation focus on the movement of water to the air from sources such as the soil, and canopy interception. Transpiration measures the subsequent loss of water for a plant as vapor through stomata in leaves. Evapotranspiration plays an important part in the water cycle.

Potential evapotranspiration, PET, means evapotranspiration with no lack of water supply all the time. This term was used commonly in an earlier period. As the upper limit of ET, this term is hard to put into practice because it depends on the type of plants, soil and climatic conditions. This means scientists cannot use a universal definition of potential evapotranspiration.

Jensen first introduced the reference crop evapotranspiration (ET_r) to stand for the ET_p for alfalfa from 30 to 50 cm in height, many view those two terms as one thing (Burman, 1994). They are the basis for the measure for actual evapotranspiration.

Actual evapotranspiration (ET_a) is the actual loss of water from the land surface due to the processes of evaporation and transpiration. To get actual evapotranspiration, the calculation procedure is to calculate reference crop evapotranspiration first and then to multiply ET_r by a crop coefficient K_c (Burman, 1994), which is estimated by considering the growth stage of the vegetables and other conditions. We can see this actual evapotranspiration as the potential evapotranspiration for some specific crop, which makes more sense to use actual evapotranspiration in the estimation of RDI.

Therefore, in this thesis, we want to consider the actual evapotranspiration as a parameter of drought, which means we will have a new index. We call it Adjusted RDI (RDI_{ad}). It can be written as

$$RDI^i_{ad} = \frac{\sum_{j=1}^{m} P_{ij}}{\sum_{j=1}^{m} ET_{ij}}, i = 1 \ldots N \text{ and } j = 1 \ldots m \tag{2.5}$$

where ET_{ij} is the actual evapotranspiration in the jth day of this period in the ith year.

In the rest of the thesis, we use RDI to stand for Adjusted RDI (RDI_{ad}).

2.4.2 Actual Evapotranspiration

Since the 1950's, scientists have tried many methods to estimate evapotranspiration (Jensen, 1990). As there are many factors which can affect the value of evapotranspiration, the more parameters the formula includes, the more complicated the formula is. We can divide these methods into three categories (Xu, 2002):

1) mass-transfer based methods;

2) radiation based methods;

3) temperature based methods.

Because of data limitations, right now the applicable methods are the temperature based methods: Blaney-Criddle and Thornthwaite. From other research (McGuinness, 1972; USDA, 1964), the Blaney-Criddle method is considered to be more accurate.

SCS Blaney-Criddle Method

Scientists have developed many different versions of the Blaney-Criddle method since the original one appeared in the 1950's. Except for the FAO-24 version, the other methods can be directly used to estimate actual evapotranspiration with special crop coefficients (Burman, 1994). What we use in this thesis is called the SCS Blaney-Criddle method.

The climatic factors used in SCS Blaney-Criddle method are temperature and day hour length. In addition, there is another factor called the monthly consumptive use crop coefficients for the specific crop.

The formula used for the SCS Blaney-Criddle method is:

$$ET_a = k_c\, k_t\, f, \tag{2.6}$$

$$k_t = 0.0311T + 0.24, \tag{2.7}$$

$$f = \begin{cases} \frac{p(1.8T+32)}{100}, & \text{if } T > 1.67 \\ 0.003p, & \text{if } T \leq 1.67 \end{cases}, \tag{2.8}$$

where ET_a is the monthly actual evapotranspiration with unit $mm/(m^2 \cdot month)$, k_c is the consumptive crop coefficient for the SCS version, k_t is a climate coefficient, f is called monthly comsumptive-use factor which reflects the radiation variation based on temperature and day length, which is the length of the time from when the sun appears to when the sun disappears for every day, T is monthly mean temperature with unit °C and p is the percentage of day length cumulated in one month over a whole year. We should note that when air temperature is no more than 1.67 °C, the value of k_t is a constant number, 0.3.

While we hope that we can set our contract for any period, which means the ET_a should be applicable for any period, right now we don't have an efficient method for daily ET_a.

Even though based on work of Bordne in 2005, the SCS Blaney-Criddle method is used for daily evapotranspiration (Bordne, 2005), we need to test it by ourselves. If the actual error is not too big, we can use it directly.

In this thesis, to simplify the problem, we only consider the contract with full month period, like 3 months or 6 months.

2.5 Price of Drought Options

2.5.1 Profits of the Option

From the definition of the option and the drought index in Section 2.1 and 2.2, we can propose a put option contract for drought based on the characteristics of the RDI: smaller the RDI is, more serious drought it will be. We assume the strike price is K, the tip is S, and the period of this contract is from t=0 to t=T. We also assume this is a European option contract, which means it can only be exercised at the maturity date. Then at the maturity date T, the profit is known after they calculate the value of the drought index. The equation is given as follows:

$$f(I) = S \max(K - I, 0), \tag{2.9}$$

where f is the profit at the end of this period the holder will receive, and I is the value of the reconnaissance drought index as given in Equation (2.5). If I is smaller than K, then the holder will receive S(K-I), if I is bigger than K, the holder would choose not to exercise this contract and the profit would be 0.

2.5.2 Basic Method

For financial derivatives, the principle of the pricing method is called the no-arbitrage assumption. It means there's no arbitrage opportunity in the market. The contract price should be consistent with its future value. The representative method is the Black-Scholes equation (Stampfli, 2001). We create a replicated portfolio which has

the same payoff with the option, and estimate the price of this portfolio as the price of the option. However, for weather options, the underlying is not an asset, but a weather index, which means it cannot be traded. Thus this no-arbitrage assumption is not suitable for weather options.

The equilibrium method is always used to estimate the weather derivative contract. People use the present value of the expectation of the profits as the price of the option (Cao, 2004; Muβhoff, 2006). We use the drought option above as an example. The price is the discounted expectation of all possible profits when the contract is exercised.

This method can be shown to satisfy the following formula:

$$Pr = E[S\ e^{-\int_0^T r(u)\mathrm{d}u} f(I)|\mathbb{Q}], \tag{2.10}$$

where T is the contract period, r(u) is a function of the interest rate, I is the value of drought index, $\mathbb{Q}$ is the space which includes all possible value of I. Here S, the tip, reflects the payment of the contract, which depends on the relationship between the value of the drought index and the loss of farmers' income. For example, if a farmer buys this kind of drought option, he/she should choose a big S if he/she thinks the drought would damage his/her crops seriously. f(I) reflects the model for the profits, given by Equation (2.9).

To simplify the problem, we keep S constant as 1 in Equation (2.9) and assume the rate of interest is constant during the period T. Then Equation (2.10) can be written as

$$Pr = e^{-rT} E[max(K - I, 0)|\mathbb{Q}] \tag{2.11}$$

The value of Pr would be affected by the variation of drought index from the strike level (Stampfli, 2001).

Therefore, based on Equation (2.11), our problem is changed to setting the strike level and finding possible index value. The strike level depends on the RDI condition in the specific place. The analysis will be shown in Chapter 3.

2.5.3 Modeling Possible Index Value

Based on Equation (2.11), we should focus on finding possible values of RDI in $\mathbb{Q}$. In general, there are three kinds of methods for this part: Historical Burn Analysis; Index Value Simulation; Daily Value Simulation (Muβhoff, 2006).

The historical burn analysis method does not need to consider the future. What we need is only the information from the past. The index value simulation method needs one to think about the possible values of the index by considering the distribution of this index. The daily value simulation needs one to use stochastic methods to simulate the daily process of climate, for example, precipitation, to simulate the drought condition in the future. The details of these three models will be shown in Chapter 3, 4, 5 separately.

2.6 Mathematical Algorithms

2.6.1 Goodness-of-Fit Test

The goodness-of-fit test is used to determine if the samples follow a special distribution or which distribution is better to fit the samples. We compare the data with the assumptive density function for a parametric distribution under the null hypothesis H_0: if H_0 is true by testing many times with experimental data, we could generate data following the distribution in the hypothesis. In general, there are three kinds of tests: the Chi-Square test, the Kolmogorov-Simirnov test, and the Anderson-Darling test (James, 1971; Stephens, 1972).

In Chapter 4, the drought index is seen as a continuous variable in time. The proper distribution function should describe the probability that the value of index falls in some specific range. Different distributions have different density of probability functions. Normal distribution, lognormal distribution, Weibull distribution, gamma distribution are usually used in some engineering problems. Since many researchers have worked in this field, we won't discuss the characteristic of the distributions (Burley, 1999; Evans, 1993; Kotz, 2002; Milhesan, 2008).

In this thesis, to test the data lists and generate random numbers from a special distribution, we use a software package called BestFit. There are 20 kinds of distribution which can be considered. We can rank those fitting distributions by the 3 tests mentioned above. It can also generate 100000 random numbers following a specific

distribution.

2.6.2 Monte Carlo Simulation

If we assume that the RDI index follows some special distribution, and we want to use the basic method to calculate the option price based on Section 2.6, we need to calculate the expectation of all possible profits based on Equation (2.11).

We know that if the variable I follows a special distribution, the expectation of f(I) should satisfy the following equation:

$$E = \int_a^b p\ (I) f(I)\ \mathrm{d}I \tag{2.12}$$

where p(I) is the density function for the distribution, f(I) is the function of the option profits like the one in Equation (2.9) and (2.10), a and b are the range of possible value. For many distributions, even if we know the cumulative function and the value of parameters of this distribution, it's still hard to calculate the expectation by Equation (2.12) directly. Therefore, we consider the Monte Carlo method.

The basic idea of Monte Carlo method is that we randomly pick values which follow the chosen distribution, and estimate E by the following formula:

$$E = \frac{1}{N} \sum_{i=1}^{N} f(I_i) \tag{2.13}$$

where I_i follows the special distribution, and N is the number of random values we pick, f(I) is the profit function. The law of large numbers ensures that this estimate converges to the correct value as the number of draws, N, increases.

One disadvantage of Monte Carlo method is that the speed of convergence will become very slow when N increase (Glasserman, 2003). Based on the law of large numbers, we have that the Error is $O(\frac{1}{\sqrt{N}})$, which means if we want to get an extra decimal place of accuracy, we need 100 times as many samples. Therefore, even if we choose a very large size of sample, the results may not converge quickly enough. We will analyze this part in Chapter 4.

Random Number Generator based on the Transformation Method

To generate a random variable which follows a specific distribution, the most common way is to use an inverse-CDF method: An arbitrary random variable X can be generated from a uniform variable $U \sim \text{Uniform}(0,1)$ due to $X \sim F^{-1}(U)$ where F is the cumulative distribution of X.

In general, we can generate the random variable X from another random variable Z if we have $Z = g(X)$. If we know $X = g^{-1}(Z)$, and how to generate variable Z, we can use this formula to convert Z to X directly.

For example, consider the lognormal variable. We know the uniform variable and standard normal variable can be easily generated in every modern computer language. Also, we know X is a lognormal variable if X satisfies $X = e^Y$ where Y is a normal variable. Therefore, we can generate lognormal random variables directly based on the above.

But if $Z = g(X)$ has no unique solution, for example, we have solutions $x_0, x_1, ..., x_r$, we can not convert Z to X directly as we don't know which x_i we should choose.

Actually, we can approximate the probability of the conditional cumulative distribution by the following formula (Taraldsen, 2005)

$$p_i = P(X = x_i | Z = g(X)) = \frac{F(x_i/|g'(x_i)|)}{\sum(F(x_i)/|g(x_i)|)}, \tag{2.14}$$

where F is the cumulative distribution function of X.

If we choose $v_0 = 0, v_i = \sum_{j \leq i} p_j$, the variable v follows the uniform distribution in the interval (0,1). Then we have an algorithm when multiple roots exist:

1. Generate $V \sim$ Uniform$(0, 1)$, denoted by v;

2. Return $x = x_i$, where x_i is the unique root such that $v_{i-1} \leq v \leq v_i$.

This algorithm can be used to generate the inverse Gaussian variable in Chapter 4. We know the density function of the inverse Gaussian distribution is

$$IG(\mu, \lambda) = p(x) = (\frac{\lambda}{2\pi x^3})^{1/2} exp(\frac{-\lambda(x-\mu)^2}{2\mu^2 x}), \tag{2.15}$$

which can be obtained from the Wald distribution, whose density function is

$$Wald(\Phi) = p(x) = \sqrt{\frac{\Phi}{2\pi}} e^{\Phi} x^{-3/2} exp[\frac{-\Phi}{2}(x + x^{-1})]. \tag{2.16}$$

Therefore, if $X \sim Wald(\Phi)$, then $Y = \mu X$ would follow the inverse Gaussian distribution $IG(\mu, \mu\Phi)$.

We also have the relationship between the Wald variable and Chi-square variable:

$$g(X) = \Phi\frac{(X-1)^2}{X} = Z \sim \chi_1^2. \tag{2.17}$$

Solving Equation (2.17), we have two roots:

$$x_1 = 1 - \frac{1}{2\Phi}(\sqrt{z^2 + 4\Phi z} - z), \tag{2.18}$$

$$x_2 = 1/x_1. \tag{2.19}$$

We also have the conditional probability when $X = x_1$:

$$p_1 = P(X = x_1|Z = z) = 1/(1 + x_1). \tag{2.20}$$

Based on the above algorithm, with $x_1 < x_2$, if we want to choose x_1 as the value of the variable, we need to make sure $v < p_1 = 1/(1 + x_1)$.

This method can be used to generate random numbers following many other distributions too.

Van der Corput Sequence

By using a Quasi-random sequence, we can improve the efficiency of $\frac{1}{\sqrt{N}}$ in the error part for the Monte Carlo method. The Van der Corput sequence is an example of a one-dimension low-discrepancy sequence.

Any integer n has a unique expansion with base b. If $n = \sum_{i=0}^{\infty} a_i(n)b^i$, then the inverse expansion with base b is $\phi_b(n) = \sum_{i=0}^{\infty} a_i(n)b^{-i-1}$, which is the nth element of the Van der Corput sequence.

2.6.3 Gauss-Newton Method for Non-Linear Least-Square Problems

It is very hard to solve non-linear systems directly. Sometimes there are no exact solutions, or sometimes the form of the solution is too complicated to calculate. Therefore, we need to find another way to compute the "locally" optimal solution

(Bartholomew-Biggs, 2005).

One way is to convert the problem to a least-square problem. Assume we have a non-linear system

$$f_i(\vec{x}) = 0, i = 1, ..., m, \tag{2.21}$$

where $\vec{x} \in \mathbb{R}^n$. Instead of solving it directly, we turn to finding an optimal $\vec{x}$ which minimizes $\| f(\vec{x}) \|^2 = \sum_{i=1}^{m} f_i(\vec{x})^2$. The Gauss-Newton method is a common choice for the non-linear least-square problem.

The algorithm is:

1. Give a start guessing for $\vec{x}$, denoted by $\vec{x}^{(0)}$, and an iteration of convergence error (ER);

2. Linearize $\vec{f}$ near the kth estimation $\vec{x}^{(k)}$:

$$f_i(\vec{x}) \approx f_i(\vec{x}^{(k)}) + Dr(\vec{x}^{(k)})(\vec{x} - \vec{x}^{(k)}), i = 1, ..., m \tag{2.22}$$

where Dr is the Jacobian: $(Dr)_{ij} = \frac{\partial f_i}{\partial x_j}$;

3. Calculate

$$A^{(k)} = Dr(\vec{x}^{(k)}), \tag{2.23}$$

$$\vec{b}^{(k)} = Dr(\vec{x}^{(k)})(\vec{x}^{(k)}) - \vec{f}((\vec{x}^{(k)})), \tag{2.24}$$

which satisfy

$$\vec{f}((\vec{x}^{(k)})) + Dr(\vec{x}^{(k)})(\vec{x} - \vec{x}^{(k)}) = A^{(k)}\vec{x}^{(k)} - \vec{b}^{(k)}; \tag{2.25}$$

4. Solve the linear system:

$$\vec{x}^{(k+1)} = (A^{(k)T}A^{(k)})^{-1}A^{(k)T}\vec{b}^{(k)}; \tag{2.26}$$

5. Calculate $\| f(\vec{x}) \|^2$. If $\| f(\vec{x}) \|^2 < ER$, output $\vec{x}^{(k+1)}$ as the optimal approximation of the solution for this non-linear system; otherwise, repeat step 2, 3, 4.

2.6.4 Milstein Method to solve Stochastic Differential Equations

Assume a stochastic process has the form

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t, \tag{2.27}$$

on $t_0 \leq t \leq T$ with the initial value $X_{t_0} = X_0$, $a(t, X_t)$ and $b(t, X_t)$ stand for functions with variables t and X_t, and W is the Wiener process.

As $a(t, X_t)$ or $b(t, X_t)$ could be very complicated, it could be hard to solve Equation (2.27) directly. Therefore, it's important to find a way to simulate the approximation of the process as accurately as possible.

Here we choose the Milstein method (Kahl, 2004; Han, 2005). The explicit Milstein method is shown as:

$$X_{n+1} = X_n + a(t_n, X_n)\Delta_{t_n} + b(t_n, X_n)\Delta W + \frac{1}{2}b(t_n, X_n)b'_X(t_n, X_n)((\Delta W)^2 - \Delta_{t_n}), \tag{2.28}$$

where Δ_{t_n} is the time interval from t_n to t_{n+1}, X_{n+1} is the simulated value of X when the time is t_{n+1}, ΔW follows normal distribution $N(0, \sigma^2\Delta_t)$.

To make the approximation more accurate and convergent, we can use the implicit

Milstein method which is shown as:

$$X_{n+1} = X_n + a(t_n, X_{n+1})\Delta_{t_n} + b(t_n, X_n)\Delta W + \frac{1}{2}b(t_n, X_n)b'_X(t_n, X_n)((\Delta W)^2 - \Delta_{t_n}). \tag{2.29}$$

The difference between the Equation (2.28) and (2.29) is that there's an unknown X_{n+1} in the right hand side of (2.29). To make (2.29) work, we can use the result of Equation (2.28) as an approximation of X_{n+1}, and substitute it into the left side of Equation (2.29) to get the implicit approximation of X_{n+1}.

2.7 Java Code

In this thesis we choose to use Java code to calculate RDI and estimate the option price. Java is easy to write, compile, and debug because of the automatic memory allocation and garbage collection. Another important reason is, Java is independent on the platform. We can easily move Java program from one computer system to another, which means it's easy to make people in different fields cooperate on this program. Finally, Java is robust when we need to check for possible errors in the beginning.

2.8 Data

Based on Section 2.3 and 2.4, we need the data of temperature, precipitation, day hour length, and crop coefficient curve. In this thesis, the climate data are from the

China Meteorological Data Sharing Service System (http://cdc.cma.gov.cn/). They have daily and monthly data for more than 700 climate stations since the year 1951.

In order to minimize the error, we use observing data in those stations where the percentage of missing points is no more than 10. The gaps are filled by substituting random numbers following normal distribution. We choose data from Jinan station first, which have a history of drought. It is located in 40°N in the east part of China. Data included observations from January 1st, 1951 to December 31st, 2006.

The curve of consumptive crop coefficient is based on McGuinness's research (McGuinness, 1972).

Chapter 3

Historical Burn Analysis

The historical burn analysis method can be seen as the simplest method for the pricing of weather derivatives, including our drought option. By this method, people who want to calculate the price of the option need neither complicated mathematical knowledge nor specialized software.

3.1 The Description of the Model

The basic assumption of this method is that all the historical data can cover all the possibilities in the future, which means the price of the option only depends on the data we already have.

From the expectation principle in Chapter 2, based on Equation (2.11), we can

see that the price of a put option for drought is given by

$$Pr = \frac{1}{N}\, e^{-rT}(\sum_{i=1}^{N} \max(K - I_i, 0)), \tag{3.1}$$

where T is the maturity, r is the interest rate, I_i is the value of the drought index for the ith year during that specific period, N is the number of years of data, K is the strike value for drought indices.

As the threshold of drought indices, K should reflect the boundary between drought and no drought. Therefore, we should decide K based on every RDI list. From the definition of RDI in section 2.3, in the beginning, we suggest K could be n in (1, $\bar{\alpha}_0$), where $\bar{\alpha}_0$ is given by Equation (2.2).

K=1 can be used for those regions with a good irrigation system or no serious drought problems. For the holder of this put option, he/she can get a profit when the value of RDI is less than 1, which means the precipitation can not fit the need of actual evapotranspiration. Now, for dry regions, we suggest it is better to choose K=$\bar{\alpha}_0$, because $\bar{\alpha}_0$, as an aridity index, can reflect the dryness conditions in these regions.

3.2 Results of Actual Evapotranspiration

We use SCS Blaney-Criddle method to calculate actual evapotranspiration for our problem, and we check this formula to see if it is fitted to the daily cumulation.

From Chapter 2, we can see Equation (2.6) (2.7) and (2.8) are used for monthly

ET_a. The data we use now are those from Jinan station, and the daily values of the SCS Blaney-Criddle crop growth stage coefficients from alfalfa curve.

We calculate ET_a for every month over past 56 years by using monthly data and by summing up daily ET_a based on the same formula (2.6)-(2.8). Figure 3.1 shows the histogram of relative error between them. We can see that the percentage of those points whose relative error from the original monthly values is lower than 20% is 93.56. 0.45% of all points have a big difference from the original formula results (greater than 50% relative error). Based on this, we cannot say Equation (2.6)-(2.8)

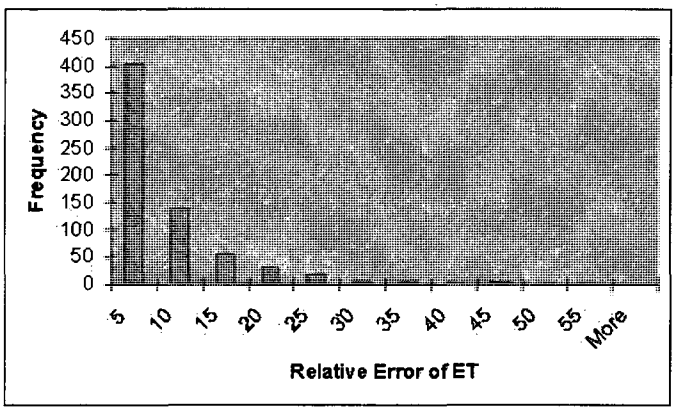

Figure 3.1: Histogram of Relative Error for ET

can be used for daily ET_a.

3.3 Results of RDI

Here we still use monthly and daily data separately to calculate RDI for different periods. If we assume that the SCS Blaney-Criddle method can be used to calculate

daily actual evapotranspiration, we can calculate adjusted RDI for any period by Equation (2.5). Right now we only consider the contracts for three months, half a year and one whole year, as the drought condition can always be reflected in a long period, to see if the drought in this place is serious.

Around Jinan station, the main crop is corn. Even though we can not get the consumptive crop coefficient of corn, we can use the curve of alfalfa instead of the curve of corn, because based on the definition of potential evapotranspiration and actual evapotranspiration in Section 2.4.1, the actual evapotranspiration of alfalfa can be seen as the potential evapotranspiration in that region. In Jinan, corn is planted in April or May, and they will be harvested in July or August. Based on the growth stage of the plants, we can set the period of the contract from April to August for corn. We can also choose the month when the crops grow or are yielded as a short period for the contract, for example, April to June.

RDI Lists with All Data

First we substitute all the monthly data we have in Equation (2.6)-(2.8). We calculate the value of RDI for different periods in a year. Table 3.1 and Table 3.2 show the results for the stations Jinan and Wuhan. Jinan is located in latitude 40°N and Wuhan is located around 30°N in China. Here we use the same crop coefficient for Wuhan station. This would give us a basic idea how different the RDI would be in dry and wet regions. In each table, we calculate the RDI lists for the whole year, half the year from spring to summer, the whole grown stage from April to August for

corn, the growth stage from April to June and harvest stage from July to August.

From Table 3.1, we can see that Jinan is always under a dry condition in the whole year, because the mean value of RDI from 1951 to 2006 is 0.51751. It means the supply of water from rainfall is only satisfying half of the crops' need for water. Similarly, in the period from March to August, or from April to August, the situation doesn't change. For the period from April to June, the situation is much worse. Even though we think the summer is a rainy season, the RDI list for the period from June to August shows that the water balance still doesn't look good. We believe here the day hour length and high temperature affects drought more than we thought before. If we compare the last three columns in Table 3.1, we could see that the aridity index for the period from June to August is higher that the one in the June-to-August period, which means the water balance in June is pretty good. On opposite, April and May would be very very dry because the aridity index for the period from April to June is much smaller than other periods. Then if we turn to Table 3.2, we could say that this region keeps a good water balance over the last 56 years. For the whole year or half a year, the aridity index which reflects the average of RDI is close to 1. And during the growth stage, the aridity index is 1.05126. If we look at the whole list, we can see that in this period, the value of RDI is often in the range [0.7,1.2].

Table 3.1: The RDI for Different Periods in Jinan

Y/P	1-12	3-8	4-8	6-8	4-6	7-8
1951	0.52539	0.49698	0.49394	0.74906	0.25372	0.56150
1952	0.45947	0.34171	0.31881	0.22496	0.40754	0.37211
1953	0.72129	0.84575	0.85277	1.36285	0.36318	1.00925
1954	0.62925	0.60145	0.61702	0.88782	0.36721	0.74114
1955	0.31937	0.23839	0.23546	0.45928	0.04016	0.30187
1956	0.58918	0.57828	0.57134	0.46755	0.67797	0.68849
1957	0.40815	0.46083	0.45147	0.57886	0.33011	0.57195
1958	0.43156	0.41641	0.41360	0.68651	0.16571	0.47289
1959	0.40306	0.40012	0.39720	0.46812	0.32490	0.40218
1960	0.34676	0.40058	0.41448	0.74956	0.12012	0.51026
1961	0.71698	0.61830	0.63525	1.08221	0.23504	0.83173
1962	0.89216	0.92069	0.95605	1.75325	0.21393	1.23500
1963	0.77145	0.94658	0.96642	1.30203	0.63318	1.00393
1964	0.86156	0.80175	0.81203	1.25986	0.39598	0.91661
1965	0.32007	0.33477	0.34290	0.56502	0.14238	0.39414
1966	0.36408	0.41804	0.40294	0.54224	0.26927	0.53493
1967	0.38331	0.34061	0.33489	0.53392	0.16060	0.43905
1968	0.23042	0.15355	0.15721	0.17900	0.13782	0.12846

Table 3.1: The RDI for Different Periods in Jinan

Y/P	1-12	3-8	4-8	6-8	4-6	7-8
1969	0.55575	0.55965	0.56025	0.78500	0.34187	0.57935
1970	0.39708	0.39695	0.40750	0.55947	0.26569	0.42121
1971	0.59660	0.65056	0.64426	0.89536	0.41236	0.83369
1972	0.51750	0.49685	0.50980	0.94469	0.13644	0.64193
1973	0.68649	0.71172	0.74190	1.20373	0.29997	0.90457
1974	0.55214	0.49764	0.49506	0.91555	0.12425	0.66312
1975	0.41561	0.34667	0.35514	0.51683	0.21006	0.40524
1976	0.61585	0.60566	0.62837	1.10904	0.19411	0.77268
1977	0.41617	0.41343	0.43228	0.63862	0.24236	0.47666
1978	0.64030	0.72264	0.73715	1.24524	0.29985	0.99702
1979	0.43087	0.39739	0.37906	0.50887	0.25677	0.44482
1980	0.54609	0.55748	0.55704	0.23350	0.85018	0.49288
1981	0.28168	0.31335	0.31858	0.40972	0.23562	0.41022
1982	0.47238	0.43768	0.45336	0.62411	0.30447	0.55831
1983	0.43264	0.40580	0.37802	0.50086	0.27073	0.33782
1984	0.54232	0.52012	0.52492	0.70196	0.36025	0.62623
1985	0.46727	0.38911	0.39839	0.53703	0.26638	0.40669
1986	0.25712	0.26831	0.26815	0.38617	0.16403	0.34724

Table 3.1: The RDI for Different Periods in Jinan

Y/P	1-12	3-8	4-8	6-8	4-6	7-8
1987	0.67826	0.78426	0.80867	1.19887	0.42700	0.98646
1988	0.41264	0.48626	0.49826	0.86989	0.17745	0.58466
1989	0.26779	0.26684	0.23445	0.32596	0.15262	0.29211
1990	0.57779	0.61528	0.62670	0.90594	0.35387	0.72953
1991	0.58795	0.62470	0.61343	1.00160	0.23260	0.70522
1992	0.40529	0.40277	0.41187	0.68204	0.15165	0.47233
1993	0.63949	0.61728	0.64322	0.77119	0.53205	0.68722
1994	0.60757	0.65709	0.67568	1.00754	0.37293	0.84799
1995	0.44468	0.45744	0.46642	0.76217	0.19444	0.59335
1996	0.63087	0.76458	0.78003	1.22875	0.37807	1.05254
1997	0.43525	0.37485	0.36587	0.49652	0.23388	0.34014
1998	0.54463	0.67185	0.68489	1.06975	0.32495	0.76478
1999	0.43089	0.46094	0.45921	0.51548	0.40996	0.49744
2000	0.53353	0.47735	0.50172	0.62470	0.38673	0.60076
2001	0.43923	0.40928	0.42041	0.45731	0.38961	0.55698
2002	0.33680	0.34548	0.35903	0.23924	0.47708	0.26090
2003	0.77899	0.68449	0.69684	1.06617	0.37362	0.76563
2004	0.84382	0.95134	1.00294	1.27369	0.76117	1.14499

Table 3.1: The RDI for Different Periods in Jinan

Y/P	1-12	3-8	4-8	6-8	4-6	7-8
2005	0.71103	0.58395	0.60535	0.87637	0.37833	0.71134
2006	0.47685	0.57705	0.60558	0.62870	0.58491	0.55179
Mean Value	0.51751	0.52176	0.52899	0.76017	0.31727	0.61752
Standard Deviation	0.15586	0.17965	0.18882	0.23880	0.15968	0.33582

Table 3.2: The RDI for Different Periods in Wuhan

Y/P	1-12	3-8	4-8	6-8	7-8	4-6
1951	0.98341	0.97449	0.95790	0.83686	0.97507	0.94099
1952	0.78822	0.71494	0.56905	0.41774	0.73573	0.39077
1953	0.81165	0.76553	0.68891	0.65526	0.88416	0.50361
1954	1.52036	1.71108	1.77183	1.58704	2.24640	1.30115
1955	0.87552	1.07079	0.95655	0.99640	1.49102	0.44726
1956	0.72402	0.86520	0.71537	0.48062	1.11758	0.31314
1957	0.99494	1.00139	0.98776	0.92267	1.14144	0.83814
1958	1.02657	0.94195	0.89608	0.59860	0.93551	0.85628
1959	1.08191	0.98887	0.96326	0.73659	1.97510	0.09089
1960	0.76131	0.73185	0.54887	0.48318	0.80689	0.30130
1961	0.72378	0.60142	0.47938	0.40343	0.83001	0.12491

Table 3.2: The RDI for Different Periods in Wuhan

Y/P	1-12	3-8	4-8	6-8	4-6	7-8
1962	1.20819	1.29084	1.33910	1.40062	1.04944	1.61208
1963	0.81985	0.92803	0.89483	0.71938	0.76058	1.03002
1964	0.97906	0.94982	0.90518	0.65274	1.68219	0.18161
1965	0.68349	0.59843	0.58776	0.47694	0.84379	0.33300
1966	0.52871	0.44467	0.42041	0.27699	0.69142	0.17509
1967	0.85381	0.63151	0.52433	0.33189	0.72513	0.32958
1968	0.72073	0.67565	0.58867	0.48563	0.62618	0.54938
1969	1.30144	1.55127	1.51377	1.78215	0.90036	2.13657
1970	0.93501	0.88256	0.82228	0.61581	1.33362	0.34813
1971	0.59391	0.50836	0.44887	0.31378	0.74068	0.18080
1972	0.79728	0.53239	0.41934	0.22503	0.66097	0.17273
1973	0.90239	0.83673	0.77902	0.48743	1.15333	0.41599
1974	0.70368	0.63157	0.61318	0.38397	0.85593	0.35105
1975	0.95261	0.94191	0.92488	0.74762	1.28512	0.58117
1976	0.66155	0.58892	0.54158	0.45685	0.98681	0.10012
1977	0.87390	0.92065	0.84225	0.50774	1.20314	0.49033
1978	0.56602	0.53628	0.47522	0.27542	0.89506	0.08380
1979	0.72222	0.81680	0.77325	0.70625	1.29695	0.26317

Table 3.2: The RDI for Different Periods in Wuhan

Y/P	1-12	3-8	4-8	6-8	4-6	7-8
1980	1.23403	1.43910	1.29820	1.52636	0.98519	1.64654
1981	0.82671	0.68175	0.61121	0.63148	0.74011	0.48598
1982	1.20782	1.21804	1.14444	1.30312	1.28732	0.99566
1983	1.38182	1.20418	1.23048	1.20517	1.52120	0.92696
1984	0.88633	0.86891	0.85914	0.91955	1.29922	0.43218
1985	0.74831	0.65841	0.58579	0.34949	0.82591	0.34535
1986	0.76623	0.68765	0.64519	0.59873	0.75276	0.53040
1987	1.05998	1.02016	0.95562	0.82567	1.07855	0.83377
1988	0.94674	0.95925	0.94893	0.84225	1.11826	0.78109
1989	1.23237	1.08174	1.04368	0.95586	1.45208	0.63449
1990	0.93467	0.84664	0.80120	0.62448	1.13100	0.48383
1991	1.31626	1.53332	1.47166	1.44733	1.21062	1.72865
1992	0.80337	0.90928	0.72113	0.66198	1.15977	0.27597
1993	1.17874	0.97162	0.88652	0.65749	1.06708	0.69047
1994	0.71587	0.63607	0.60278	0.60029	0.52294	0.68429
1995	0.89034	0.97691	0.99223	0.73395	1.37373	0.62232
1996	0.94231	1.00166	0.89264	1.02429	0.93536	0.84943
1997	0.64636	0.53690	0.54245	0.58308	0.44688	0.64409

Table 3.2: The RDI for Different Periods in Wuhan

Y/P	1-12	3-8	4-8	6-8	4-6	7-8
1998	1.14111	1.36168	1.31151	1.14886	1.15376	1.46981
1999	0.96213	1.14760	1.14106	0.99805	1.78149	0.46426
2000	0.78058	0.51610	0.52265	0.48313	0.67080	0.36747
2001	0.58602	0.44579	0.43319	0.27609	0.75293	0.11476
2002	1.02451	1.06775	0.98982	0.67357	1.25135	0.71334
2003	0.94193	0.95241	0.87988	0.77760	0.99451	0.76453
2004	1.04995	1.18029	1.21351	1.32581	1.16065	1.26992
2005	0.73358	0.58645	0.57359	0.49953	0.72476	0.39926
2006	0.67477	0.62962	0.64470	0.50348	0.64253	0.64702
Mean Value	0.90551	0.88845	0.83736	0.73467	1.05126	0.62937
Standard Deviation	0.21809	0.28980	0.30210	0.36138	0.35124	0.44868

The Strike Level

It seems like the use of a drought option is not so urgent in Wuhan. If we want to, we can set the strike as 1. For Jinan, we can say that there is a serious problem with the lack of rain if there is no irrigation system because from April to June corn needs water to grow. We cannot just use 1 as the strike in the contract, which is too far away from the real situation. From the RDI list for the period from April to

August, we can see that the aridity index is 0.5290. And for the last 56 years, only one RDI point is larger than 1. Therefore based on Section 2.5.2 and Equation (3.1), the aridity index would play a very important role in the option prices, not the diffusion of RDI any more. However, it's also not proper if we choose to use the aridity index as the strike. Unlike drought, the aridity reflects the usual dry condition in the past. In some cases, we want to consider some of the effects of the aridity in our contract. Then it's reasonable to choose some value between 1 and the aridity index. We should consider the specific situation when we put this contract into practice. Right now, we just assume strike is 0.5 for the period from April to June, 0.7 for the period from April to August and for the whole year.

Checking RDI with Cumulative Daily ET_a

If we assume Equation (2.6), (2.7), and (2.8) can be used for daily ET_a, we can calculate monthly ET_a by summing all daily values up. Table 3.3 shows the RDI lists for different periods based on daily ET_a in Jinan.

Table 3.3: The RDI based on Daily ET in Jinan

Y/P	1-12	3-8	4-8	6-8	4-6	7-8
1951	0.52296	0.33162	0.49923	0.75194	0.25909	0.56304
1952	0.42322	0.41222	0.33211	0.23923	0.43262	0.38678
1953	0.72610	0.39643	0.85678	1.37425	0.37360	1.02384
1954	0.63987	0.37052	0.63726	0.90316	0.38432	0.75399

Table 3.3: The RDI for Different Periods in Wuhan

Y/P	1-12	3-8	4-8	6-8	4-6	7-8
1955	0.32787	0.07402	0.24955	0.46579	0.05928	0.31283
1956	0.59650	0.67470	0.58278	0.47813	0.69466	0.69770
1957	0.39793	0.35707	0.45692	0.59286	0.33225	0.58252
1958	0.43453	0.19886	0.42278	0.69688	0.17851	0.48605
1959	0.40618	0.35183	0.40969	0.48135	0.33372	0.41509
1960	0.35931	0.13214	0.42714	0.76786	0.13204	0.53414
1961	0.73207	0.25037	0.64508	1.09916	0.25279	0.84632
1962	0.88113	0.26010	0.96642	1.76771	0.23065	1.24628
1963	0.77106	0.59614	0.98180	1.31786	0.64882	1.01310
1964	0.84786	0.36799	0.80486	1.27803	0.36822	0.92788
1965	0.33675	0.15506	0.34157	0.57250	0.14018	0.40472
1966	0.37496	0.30686	0.41349	0.55165	0.28354	0.54815
1967	0.39310	0.22876	0.34671	0.55189	0.18075	0.45779
1968	0.24521	0.15886	0.18088	0.19410	0.16115	0.14839
1969	0.55238	0.37364	0.56487	0.79583	0.36218	0.58566
1970	0.38941	0.23830	0.39801	0.56658	0.23177	0.43302
1971	0.57990	0.40373	0.64989	0.90212	0.42581	0.84338
1972	0.52253	0.16881	0.51993	0.95320	0.16346	0.65122

Table 3.3: The RDI for Different Periods in Wuhan

Y/P	1-12	3-8	4-8	6-8	4-6	7-8
1973	0.68606	0.31409	0.74129	1.20365	0.30757	0.90291
1974	0.54903	0.17153	0.50142	0.92920	0.13591	0.67322
1975	0.42325	0.20917	0.36913	0.52937	0.22067	0.41557
1976	0.60797	0.22008	0.63840	1.13283	0.20602	0.78370
1977	0.43036	0.24124	0.44671	0.65864	0.26156	0.49147
1978	0.63753	0.31201	0.74173	1.24651	0.31234	1.00244
1979	0.39409	0.27958	0.37341	0.51634	0.23483	0.46140
1980	0.56198	0.80869	0.56295	0.24765	0.84987	0.49971
1981	0.28973	0.26481	0.32655	0.41941	0.24480	0.42036
1982	0.47943	0.32558	0.46211	0.64001	0.31008	0.56547
1983	0.44632	0.34454	0.38410	0.51553	0.28044	0.35279
1984	0.55445	0.37722	0.53502	0.71308	0.36627	0.62949
1985	0.48328	0.28401	0.40857	0.54922	0.27283	0.41664
1986	0.27685	0.19432	0.28026	0.40565	0.17516	0.35937
1987	0.68493	0.43445	0.81829	1.20985	0.43360	0.99450
1988	0.42711	0.19750	0.51053	0.87855	0.19357	0.59981
1989	0.28857	0.24287	0.24524	0.34558	0.16892	0.30423
1990	0.58301	0.46236	0.63364	0.91461	0.36342	0.73736

Table 3.3: The RDI for Different Periods in Wuhan

Y/P	1-12	3-8	4-8	6-8	4-6	7-8
1991	0.59538	0.28722	0.62059	1.00807	0.24368	0.71141
1992	0.41276	0.18737	0.42138	0.68459	0.16313	0.48383
1993	0.65110	0.49575	0.65392	0.78123	0.53994	0.69670
1994	0.62397	0.36075	0.69421	1.02720	0.38540	0.86120
1995	0.45459	0.22213	0.47491	0.78031	0.20317	0.60340
1996	0.64804	0.37729	0.79181	1.24312	0.38934	1.06914
1997	0.44689	0.28656	0.37871	0.51262	0.25506	0.35140
1998	0.55142	0.35817	0.69250	1.07687	0.32661	0.77183
1999	0.43831	0.41510	0.46520	0.51917	0.42577	0.50629
2000	0.54572	0.41848	0.51191	0.63314	0.40160	0.61331
2001	0.45549	0.46173	0.43077	0.47810	0.39729	0.56832
2002	0.34684	0.42865	0.36867	0.24240	0.48207	0.26943
2003	0.78561	0.38874	0.70460	1.07073	0.38032	0.76921
2004	0.84386	0.69178	1.00035	1.28322	0.76170	1.15254
2005	0.71018	0.38956	0.61248	0.88260	0.38501	0.71375
2006	0.48360	0.52416	0.61507	0.63451	0.58943	0.55968
Mean Value	0.52247	0.33546	0.53757	0.77171	0.32673	0.62810

Most of values in Table 3.3 are close to those in Table 3.1. We can calculate the relative difference for RDI lists based on monthly data and RDI lists based on daily data for a whole year, half a year and 3 months. The results are shown in Table 3.4.

Table 3.4: The Relative Error of RDI based on Daily ET

Y/P	1-12	4-8	4-6
1951	0.47%	1.06%	2.07%
1952	8.57%	4.00%	5.80%
1953	0.66%	0.47%	2.79%
1954	1.66%	3.18%	4.45%
1955	2.59%	5.65%	32.25%
1956	1.23%	1.96%	2.40%
1957	2.57%	1.19%	0.65%
1958	0.68%	2.17%	7.17%
1959	0.77%	3.05%	2.64%
1960	3.49%	2.96%	9.02%
1961	2.06%	1.52%	7.02%
1962	1.25%	1.07%	7.25%
1963	0.05%	1.57%	2.41%
1964	1.62%	0.89%	7.54%
1965	4.95%	0.39%	1.57%

Table 3.4: The Relative Error of RDI based on Daily ET

Y/P	1-12	4-8	4-6
1966	2.90%	2.55%	5.03%
1967	2.49%	3.41%	11.15%
1968	6.03%	13.09%	14.48%
1969	0.61%	0.82%	5.61%
1970	1.97%	2.38%	14.63%
1971	2.88%	0.87%	3.16%
1972	0.96%	1.95%	16.53%
1973	0.06%	0.08%	2.47%
1974	0.57%	1.27%	8.58%
1975	1.81%	3.79%	4.81%
1976	1.30%	1.57%	5.78%
1977	3.30%	3.23%	7.34%
1978	0.43%	0.62%	4.00%
1979	9.33%	1.51%	9.34%
1980	2.83%	1.05%	0.04%
1981	2.78%	2.44%	3.75%
1982	1.47%	1.89%	1.81%
1983	3.06%	1.58%	3.46%

Table 3.4: The Relative Error of RDI based on Daily ET

Y/P	1-12	4-8	4-6
1984	2.19%	1.89%	1.64%
1985	3.31%	2.49%	2.36%
1986	7.13%	4.32%	6.35%
1987	0.97%	1.18%	1.52%
1988	3.39%	2.40%	8.32%
1989	7.20%	4.40%	9.65%
1990	0.90%	1.10%	2.63%
1991	1.25%	1.15%	4.55%
1992	1.81%	2.26%	7.04%
1993	1.78%	1.64%	1.46%
1994	2.63%	2.67%	3.24%
1995	2.18%	1.79%	4.30%
1996	2.65%	1.49%	2.89%
1997	2.60%	3.39%	8.30%
1998	1.23%	1.10%	0.51%
1999	1.69%	1.29%	3.71%
2000	2.23%	1.99%	3.70%
2001	3.57%	2.40%	1.93%

Table 3.4: The Relative Error of RDI based on Daily ET

Y/P	1-12	4-8	4-6
2002	2.89%	2.61%	1.04%
2003	0.84%	1.10%	1.76%
2004	0.01%	0.26%	0.07%
2005	0.12%	1.16%	1.74%
2006	1.40%	1.54%	0.77%

From this table, we can see that the relative difference between using monthly origin ET_a and daily ET_a for long periods is lower than 5%. But for short periods like 3 months, the relative difference is always around 10%. Here we should also notice that the absolute values of RDI from April to June is pretty small in Table 3.1 and 3.3. For example, the relative difference in 1955 is 32.25%, but the RDI from monthly data is 0.04016 and the RDI from cumulative daily ET_a is 0.05928. We could say they both stand for a very serious drought conditions.

Therefore, considering there is some missing points for daily data, and right now we use the average value of day length over a month in the daily ET_a calculations, some error cannot be avoided. If we really need to set the drought option contract between two specific days for a long period, it is still reasonable to use the same formula to calculate RDI. However in this thesis, we only consider to substitute monthly data in

Equation (2.6), (2.7), and (2.8).

RDI Lists with Only One or Two Variables

From Equations (2.5) and (2.6)-(2.8), we can see there are actually four variables in total: the amount of precipitation, the temperature, the day length, and the crop coefficient. As we don't know much about the crop coefficient right now, we can assume it as a constant because what we can use right now is the coefficient of alfalfa. Actually, the day length is only concerned with the latitude of the region. If we already decide where to set the contract, we can set the day hour length as a known curve. We can see the effect of different variables on drought by changing the historical records of the variable to the mean value over 56 years or replacing the day hour length curve by the one on another latitude. We still use data from Jinan station for example.

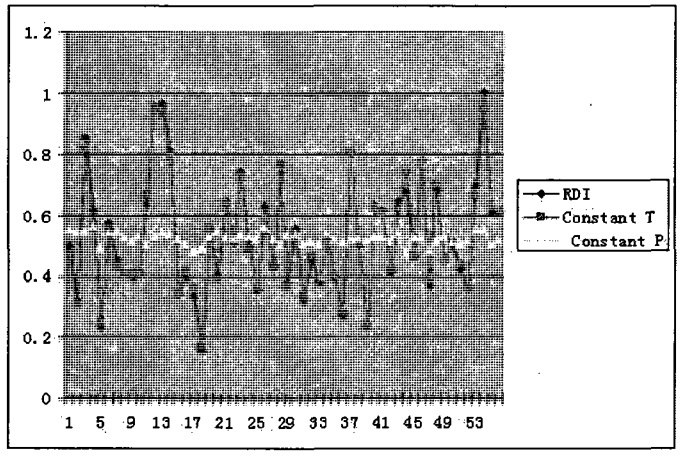

Figure 3.2: RDI with one variable in April-August

Therefore the latitude is settled as 40 °N. We calculate the RDI list for the period from April to August with one variable and use the mean value over 56 years to

replace another one.

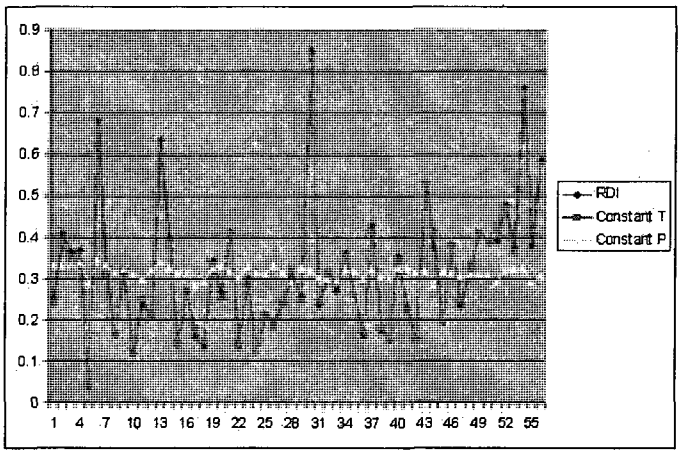

Figure 3.3: RDI with one variable in April-June

Figure 3.2 and Figure 3.3 show the results for different periods with only one variable: temperature or precipitation. In Figure 3.4, the curve of RDI shows the historical RDI in the past 56 years. The curve of Constant T means we use the mean value of monthly temperature over 56 years to replace the historical records of temperature. If we replace the historical precipitation by the average of monthly rainfall over 56 years, the results are shown as the curve of Constant P. The meanings of the three curves are similar in Figure 3.3. It is obvious that if we substitute monthly precipitation for each year and the mean values of monthly temperature, the curve is closer to the one with three variables than if we use the temperature as the only variable, which means that precipitation is the most important component for the measure of drought. This is the same as people always thought.

We can also change the parameter latitude to see the effect of day hour length.

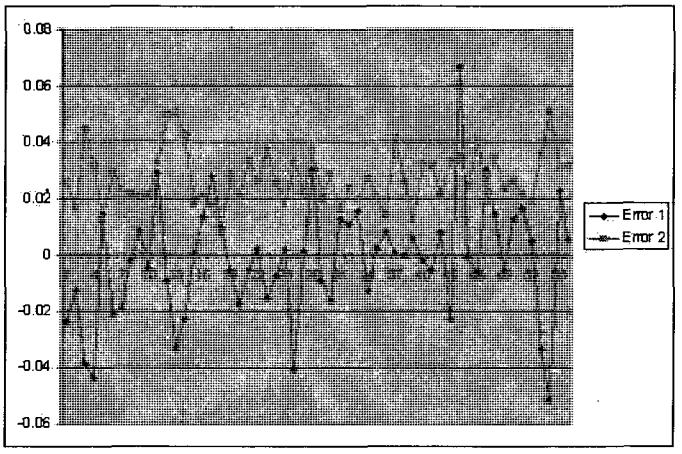

Figure 3.4: Difference between RDI lists for April-August

Figure 3.4 and Figure 3.5 compare the historical RDI, the RDI with constant temperature and the RDI with latitude 30 °N for the period from April to August and from April to June for the past 56 years. The curve of Error 1 in both figures shows the difference between RDI based on mean monthly temperature over 56 years and historical values of RDI. Similarly, the curve of Error 2 in Figure 3.4 and 3.5 shows the difference between RDI based on day hour length in latitude 30 °N and the historical values. It seems that the effect of the day hour length is similar to the monthly mean temperature. They only affect RDI in a small range. However, it's also obvious that the influence of day hour length is more stable than temperature. Based on Figure 3.4 and 3.5, the values of RDI in lower latitude is always bigger than those with latitude 40 °N. It's reasonable because for the period we choose, the day hour length is always longer in higher latitude.

This part can be seen as the previous analysis for Chapter 5. To simulate the

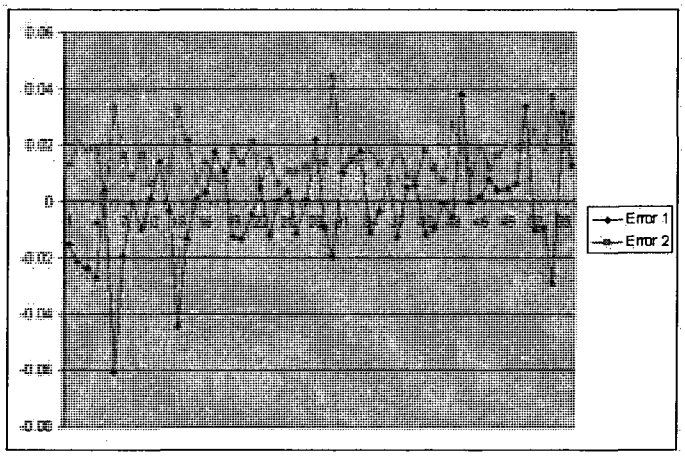

Figure 3.5: Difference between RDI lists for April-June

climate process in order to simulate the possible RDI value in the future, it is best if we can simulate the process for all the variables: the precipitation, the temperature, and the day hour length. If not, we should choose the most affective ones in our model. Because the day hour length curve for each latitude is available, in Chapter 5 we will simulate the process of both temperature and precipitation.

3.4 Results of the Drought Option Price

Now we have the basic value of I_i. We can use Equation (3.1) to calculate the price of the drought option. Here we can assume r=0.1. It is only an assumption of continuous rate of interest for the year.

We assume that this contract is signed to hedge the drought risk for the farmers in Jinan. If we set the contract for the period from April to August, we will use both

K=0.7 and K=$\bar{\alpha}_0$ as the strike level, which allow us to see the difference clearly. If we choose to use the contract to control the drought risk from April to June, we could choose the strike level as K=0.5 or K=$\bar{\alpha}_0$. We calculate the option price with all the K equals $\bar{\alpha}_0$ by using original monthly ET_a or the cumulative daily ET_a separately, and compare the difference by calculating the relative error of RDI based on daily ET_a from RDI based on monthly ET_a. The results are shown in Table 3.5. We also calculate the option price based on different strike levels for different periods. The results are shown in Table 3.6.

Table 3.5: Option Prices Based on Monthly and Daily ET with K=aridity index

	MonthlyET	DailyET	AbsError	Relative Error
Jan-Dec	0.05823105	0.05726732	0.00096373	1.68286%
Apr-Aug	0.07410387	0.07370554	0.00039833	0.54043%
Jun-Aug	0.09473529	0.09414235	0.00059294	0.62983%
Jul-Aug	0.13648905	0.13618975	0.0002993	0.21977%
Apr-Jun	0.05862795	0.05858515	4.28E-05	0.07306%

Here we could see that even though RDI from cumulative daily ET_a is 10% away from those based on the original monthly ET_a, the relative error of the option prices is only around 1% when the strike equals the aridity index. This result is reasonable because the fluctuation of RDI value affects option price the most based on Section 2.5.2. No matter whether we use monthly or daily ET_a, the fluctuation of RDI

Table 3.6: Option Prices Based on Monthly and Daily ET with Different Strike

	MonthlyET	DailyET	AbsError	Relative Error
Jan-Dec(K=0.7)	0.17638766	0.1717791	0.00460856	2.68284%
Apr-Aug(K=0.7)	0.18724644	0.17995118	0.00729526	4.05402%
Jun-Aug(K=0.7)	0.14405319	0.13683864	0.00721455	5.27230%
Jul-Aug(K=1)	0.29134917	0.28338004	0.00796913	2.81217%
Apr-Jun(K=0.5)	0.19633826	0.18850471	0.00783355	4.15563%

is similar to each other. For those with different strike levels, the relative error is around 5%. As we change the strike level for different periods, this could also affect the final results. Therefore, if we need to set the contract between some random days, this historical burn analysis method is still available.

Let us examine the columns of option prices based on monthly ET in both tables. If we use the aridity index as strike level, the option prices for most periods we chose are very low. In the rainy season July and August, the option price is higher than other contracts. It's reasonable because the standard deviation of the RDI for this period is twice as much as the one for other periods, and the option price is seriously affected by the fluctuation of RDI based on Section 2.5.2. However, if we set the strike a little bit closer to the water balance, the results are totally different. For most periods, the option prices are much higher, three times as much as the previous one, which means if we consider the dry condition in the contract, this would raise the

fluctuation of RDI under the strike level. And this shows us that it is really important how we decide which kind of strike level we choose.

Table 3.7: Option prices based on different data length with K=aridity index

	1951-2006	1957-2006	1967-2006	1977-2006	1987-2006	RE
Jan-Dec	K=0.5175	K=0.5147	K=0.5066	K=0.509	K=0.5392	
	0.05823105	0.05927542	0.05149668	0.05237679	0.05468512	13.077%
Mar-Aug	K=0.5218	K=0.5223	K=0.51	K=0.5213	K=0.5607	
	0.07012366	0.06956631	0.0637194	0.06451195	0.06488154	10.051%
Apr-Aug	K=0.529	K=0.5307	K=0.5185	K=0.5303	K=0.573	
	0.07410387	0.0736046	0.06720845	0.06854319	0.06904676	10.260%
Jun-Aug	K=0.6175	K=0.6181	K=0.6008	K=0.6081	K=0.6573	
	0.09473529	0.09441723	0.08826735	0.088455	0.08886422	7.328%
Jul-Aug	K=0.7602	K=0.7684	K=0.7358	K=0.7263	K=0.8	
	0.13648905	0.13492447	0.12660834	0.12570558	0.12277856	11.167%
Apr-Jun	K=0.3173	K=0.3131	K=0.3207	K=0.3516	K=0.3646	
	0.05862795	0.05676536	0.05713177	0.05735197	0.05343755	9.713%

Now we use some subsets of the RDI lists to calculate the option price. Now for a contract used in Jinan, we have climate records for 56 years, from 1951 to 2006. However, we should consider that there may not be enough data records to use if we

want to set the contract in some other places. Maybe they only have observations for 40 years, or much less, for example, 20 years. We want to see what would happen if we assume we only have data for the most recent 50, 40, 30 or 20 years.

Here we only calculate the option prices based on monthly ET_a. Table 3.7 shows the option prices with K=$\bar{\alpha}_0$ in different periods. Because the aridity index is concerned with data sets, if we choose to use different data sets, the strikes are different. Table 3.8 shows the option prices with some constant strike levels considering the dry condition in that period. The final column in both tables shows the relative error between the option price estimated based on data over 56 years and the most biased one based on the smaller data sets.

In Table 3.7, it seems that different data length doesn't make much difference for the final results. The relative error is around 10%. But if we turn to use some constant value as strike like Table 3.8, the conclusion would be different. If we only use the data in the most recent 50 or 40 years, it seems that the results of option price don't change a lot. However, when we only use the data in the most recent 20 years, the difference is obvious. For the period from April to August, the relative error between the original one and the 20-years-data one is 28.8%. Similarly, for the period from April to June, the relative difference is 30.6%.

This result is easy to understand. From Equation (3.1), when K equals 1, because n is not as big as possible, mean of K-I_i would change a lot if we add one big RDI in the RDI list, which means this option price is sensitive to the time length. But if we

Table 3.8: Option prices based on different data length with constant strike level

	1951-2006	1957-2006	1967-2006	1977-2006	1987-2006	RE
1-12(K=0.7)	0.17638766	0.17986768	0.1812595	0.17987126	0.15610641	12.992%
3-8(K=0.7)	0.18899318	0.18803025	0.1910371	0.18334147	0.1515263	24.726%
4-8(K=0.7)	0.18724644	0.18548385	0.18769	0.17968787	0.14533541	28.837%
6-8(K=0.7)	0.14405319	0.14425515	0.14829443	0.14501448	0.11030307	30.598%
7-8(K=1)	0.29134917	0.28282846	0.29430696	0.30490181	0.23814665	22.340%
4-6(K=0.5)	0.19633826	0.19906111	0.19267914	0.16857261	0.15046567	30.487%

use the strike K=$\bar{a}_0$, the value of $\bar{a}_0$ would change as the increase of the year number, which would always reflect the aridity condition for that region. This change would reduce the difference between each RDI value and $\bar{a}_0$. This would make the option price looks more stable than the other kind of contract.

Actually, we can say that the results of the historical burn analysis model depends on the time scale of records if we use a constant K. If there are not enough data records for the site, we should avoid this method to calculate the value of the option. This leads to the following questions: How many years of data records are enough? When can we say that we get a stable or good result?

Chapter 4

Index Value Simulation

In this chapter, we implement the index value simulation model to simulate the distribution of the index based on available data from Jinan station. The results obtained from this model is then compared to those obtained using the historical burn analysis as presented in the previous chapter.

4.1 The Description of the Model

In this analysis, the index is considered to be the only variable for determining the drought option. Unlike the historical burn analysis approach, the index simulation method considers not only the current and past values, but also the possible future values of the index.

In this model, the goodness-of-fit mentioned in Chapter 2 is used to select the best distribution that provides the simulation that is most consistent with the RDI

results. Then, a large number of random numbers exceeding 50,000 are generated to produce the new RDI distribution. Therefore, using Equation (2.11) and (2.13), the option price can be estimated by the following formula:

$$Pr = \frac{1}{N}\, e^{-rT}(\sum_{i=1}^{N} \max(K - I_i, 0)). \tag{4.1}$$

Equation (4.1) is similar to Equation (3.1). The difference is the definition of N. In historical burn analysis, N stands for how many years of data we have. Here N is the number of simulated RDI value following the specific distribution in index value simulation in Equation (4.1).

The most challenging aspect of this model is the determination of the best distribution function. Here we utilize the BestFit algorithm to determine the most appropriate distribution. The selected distributions are tested against the standard tests described in Section 2.6 including Chi-Square test, Kolmogorov-Simirnov test, and Anderson-Darling test. In addition to satisfying these tests, the random number generator based on the selected distribution must generate only non-negative random numbers to be physically meaningful as a measure of precipitation.

Therefore, the best fitting distribution must satisfy the following criteria:

1. The distribution must adequately satisfy the three goodness tests.

2. The random numbers generated from the distribution must be physically meaningful.

3. The distribution must be stable over the time scale considered in the RDI computation so that it will be valid for all relevant time intervals.

For a specific contract of drought put option the following procedure is followed. First, the value of the RDI are computed (see details in Section 3.3). Second, the RDI results are used in the BestFit algorithm to determine the distributions. Third, the rank of each distribution with respect to the various tests are estimated. Fourth, the best distribution is then selected to generate the random numbers and used to calculate the price of the option contract.

4.2 The Results of the Best Fitting Distribution

The BestFit algorithm is used to determine the most appropriate distribution from the RDI results. It should be noted that the lognormal distribution was first tested to determine whether it will provide the best fit. However, the results were not encouraging as the it did not provide consistently good fit.

4.2.1 The Ranking from the Goodness-of-Fit Tests

After substituting historical RDI in BestFit, we can have a ranking list based on specific goodness-of-fit test. The distributions on the top of the ranking is fitting data better than those at the bottom based on the value of statistics. For example, if we use Chi-square test, the test value for the triangle distribution is 5.714 which is the closest one to the critical value. On opposite, the test value for the Pareto distribution which is at the bottom of the ranking is 83.18, which means this distribution isn't good even though it pass the goodness-of-fit test.

As the ranking of the different distribution various according to the particular test, we initially chose those distributions with a rank of at least three for each of the different tests. The distributions are ranked using the weight 0.4, 0.3, and 0.3 for Chi-Square, Anderson-Darling and Kolmogorov-Simirnov test. After this initial trial iteration, other approaches can be considered. For example, when the RDI list for the period from April to August is used for the BestFit algorithm, the resulting distributions are ranked according to the various tests. According to the Chi-square test, the top five ranked distributions turned out to be triangle distribution, extreme value distribution, gamma distribution, inverse Gaussian distribution and beta general distribution. On the other hand according to the Anderson-Darling test, the best five distributions are ranked as extreme value distribution, inverse Gaussian distribution, lognormal distribution, Gamma distribution, and Pearson Type V distribution. Finally, according to the Kolmogorov-Simirnov test, the best rankings are Weibull distribution, beta general distribution, gamma distribution, inverse Gaussian distribution, and lognormal distribution. It is hard to say which distribution is the best. Therefore, after recalculating the ranking by the above iteration, we can see that the gamma distribution, inverse Gaussian distribution, and extreme value distribution are ranked as the top three by all the three tests and will be considered the best distributions. The summary of the ranking is presented in Table 4.1 for the various data periods we choose in Chapter 3. As the table demonstrates, no distribution is always the best one for all the RDI of different periods.

Table 4.1: BestFit Distributions for RDI Lists

	1	2	3
Jan-Dec	Triang	Weibull	ExtValue
Mar-Aug	Gamma	Pearson5	BetaGeneral
Apr-Aug	Gamma	InvGauss	ExtValue
Jun-Aug	InvGauss	Lognorm	Pearson5
Jul-Aug	ExtValue	Weibull	InvGauss
Apr-Jun	Log-logistic	ExtValue	Pearson5

4.2.2 Non-Negative Test

Once we have chosen the best distributions, we generate 50,000 random numbers from which we remove any distributions with negative values in order to obtain physically meaningful prices. Table 4.2 depicts the new ranking for each test and for each period. As the table demonstrates, many of the distribution functions in Table 4.1 fail to satisfy the non-negative test. For example all the three top-ranked distributions for the period from June to August fail the non-negative test.

Actually, after the non-negative check, we can see that the results in Table 4.2 change a lot. For example, for the period from March to August, we need to replace the inverse Gaussian and Pearson Type V distribution by beta general and extreme value distribution. And for the period from April to June and from June to August, we need to kick out all our first choices.

Table 4.2: BestFit Distributions for RDI Lists After Negative Check

	1	2	3
Jan-Dec	Triang	Weibull	ExtValue
Mar-Aug	Gamma	BetaGeneral	ExtValue
Apr-Aug	Gamma	InvGauss	ExtValue
Jun-Aug	ExtValue	Weibull	Triang
Jul-Aug	Weibull	Triang	BetaGeneral
Apr-Jun	Gamma	BetaGeneral	Weibull

In Table 4.3, we rank the distributions for the period from April to June and from July to August. It also shows the number of negative numbers for each distribution. From this table we could see that for both periods, after non-negative check, only three distributions can keep generating non-negative random numbers. Unfortunately, most of them are ranked sixth, seventh, or tenth, which are almost at the bottom of the ranking.

Since the BestFit has only 15 distributions in total, it is capable to consider the rank of all possible distributions. In general, if the number of distributions is large, considering all the distributions becomes impractical because the ranking of distributions cannot be copied directly, and one has to calculate the ranking in total by inputting the name of the distributions one by one which becomes tedious.

From Table 4.1 and 4.2, although the RDI lists can fit lognormal distribution, the

Table 4.3: Non-negative Test for Distributions for 4-6 and 7-8

Ranking	4-6	Negative Points #	7-8	Negative Points #
1	Log-logistic	26	ExtValue	7
2	ExtValue	26	Weibull	0
3	Pearson5	3	InvGauss	10
4	Logistic	1412	Lognorm	10
5	Lognorm	4	Log-logistic	89
6	Gamma	0	Normal	613
7	BetaGeneral	0	Pearson5	38
8	InvGauss	5	Triang	0
9	Normal	1188	Logistic	1058
10	Weibull	0	BetaGeneral	0

results are not satisfactory. For the six periods considered here, lognormal distribution can be only used in one of them. For other periods, lognormal distribution is neither the best one, nor passing the non-negative test. Therefore, one cannot assume that all the RDI lists fit lognormal distribution to make the problem easier to solve. Then it is reasonable that we don't choose the third expression of RDI in Chapter 2 as the index in our drought option contract.

In Table 4.2, some specific distributions are always ranked in the first three places. Actually, gamma, beta general, extreme value, triangle and Weibull distributions seem

to generate non-negative numbers all the time. However, most of the time they are not the best choices. While the other distributions seem to be a better fit according to the various tests, their performance with respect to the non-negative test seem to be dependent on the parameter values of the distributions which make the selection of the best distribution very challenging.

It seems difficult to find a distribution with good ranking and generating non-negative random numbers. Note however that, as depicted in Table 4.3, the frequency of the negative random numbers is extremely low. For most of distributions the number of negative random numbers never exceeded 50 out of 50,000 which is less than 0.1%. In order to test the validity of this assessment several sensitivity tests were carried out. It turns out that the frequency of the negative random numbers remains low and stable. Note that even when the size of the random numbers is doubled, the number of negative random numbers remains as low.

Therefore, it is reasonable to consider the removal of the negative random numbers for the highly ranked distributions. This can be done by removing the negative numbers completely and truncating the size of the random numbers accordingly. This method can give reasonable results as the frequency of the negative number is extremely small as shown above. The result of the option prices obtained from these approaches will be presented in Section 4.3.

4.2.3 Stability Test

In this section, the issue of stability of the distribution as the distribution of RDI index will be considered. Here 3 periods are chosen in order to determine the distribution ranking based on the subsets of RDI lists. The results are shown in Table 4.4, 4.5 and 4.6 for period 1-12, 4-8 and 4-6. In each table, we test the distributions in Table 4.1 with different data lengths and give the estimation of parameters.

Table 4.4: Stability Check for 1-12

T/L	56	50	40	30	20
Triang	1	4	8	5	3
	min=0.2005	min=0.19777	min=0.19429	min=0.2101	min=0.23135
	m.likely=0.4162	m.likely=0.4156	m.likely=0.4326	m.likely=0.4326	m.likely=0.4392
	max=0.94017	max=0.94399	max=0.88233	max=0.8906	max=0.91539
Weibull	2	3	2	4	1
	α=2.1779	α=2.0853	α=2.8229	α=2.3796	α=2.5099
	β=0.36357	β=0.35631	β=0.40348	β=0.35574	β=0.38147
	shift=0.1954	shift=0.1990	shift=0.1460	shift=0.1935	shift=0.2007
ExtValue	3	6	9	9	1
	a=0.44284	a=0.43932	a=0.43721	a=0.44052	a=0.46854
	b=0.13464	b=0.13484	b=0.12967	b=0.12698	b=0.13026

For example if one considers the period from April to August, the inverse Gaussian distribution, extreme value distribution and gamma distribution will be chosen as the

Table 4.5: Stability Check for 4-8

T/L	56	50	40	30	20
Gamma	1	Not Fit	Not Fit	3	Not Fit
	α=8.7073			α=8.6486	
	β=0.0642			β=0.067676	
	shift=-0.0304			shift=-0.0579	
InvGauss	2	3	3	4	3
	μ=0.7981	μ=0.82878	μ=0.69494	μ=0.83400	μ=0.66212
	λ=14.0822	λ=16.66057	λ=8.8002	λ=14.57738	λ=5.58731
	shift=-0.2691	shift=-0.3101	shift=-0.1867	shift=-0.3066	shift=-0.1450
ExtValue	3	1	1	1	2
	a=0.43985	a=0.43154	a=0.41778	a=0.43388	a=0.41336
	b=0.1599	b=0.15715	b=0.15974	b=0.16768	b=0.17879

most appropriating distributions. In Table 4.5, we could see that for inverse Gaussian distribution and extreme value distribution, the ranking doesn't change a lot with different data lengths. However, the data from recent 50, 40, or 20 years don't fit gamma distribution any longer. Therefore, even though gamma distribution is the best one from the first iteration, and it passes non-negative check, we should not use it as the distribution of RDI. Between the inverse Gaussian and extreme value distribution, we prefer the extreme value distribution, not only because the rank of this distribution is always better than the other, but the values of parameters for this

Table 4.6: Stability Check for 4-6

T/L	56	50	40	30	20
Log-logistic	1	6	3	2	3
	α=4.4404	α=2.6638	α=2.7866	α=3.0347	α=8.0236
	β=0.3596	β=0.20538	β=0.21652	β=0.23927	β=0.66199
	γ=-0.0708	γ=0.06946	γ=0.066827	γ=0.07509	γ=-0.31058
ExtValue	2	1	1	4	5
	a=0.24645	a=0.24602	a=0.25295	a=0.28166	a=0.29447
	b=0.12107	b=0.10992	b=0.11116	b=0.11446	b=0.1246
Pearson5	3	2	4	1	4
	α=13.204	α=5.9898	α=6.2842	α=6.7877	α=26.391
	β=6.5185	β=1.7065	β=1.8649	β=2.1802	β=18.923
	shift=-0.2169	shift=-0.0271	shift=-0.0308	shift=-0.02423	shift=-0.3805

distribution remain essentially the same. For the inverse Gaussian distribution, the value of λ changes a lot with different data lengths. In the next section we could see if this would make huge difference on option prices.

Similarly, for the period from January to December in Table 4.4, we would finally choose the Weibull distribution. The rank of this distribution is more stable, and the values of parameters don't change as much with different data lengths. For the period from April to June, in Table 4.6, the results lead to unexpected conclusions. All those

distributions which can generate negative random numbers still fit the subsets of the data lists. However, if we check the non-negative distributions in Table 4.2, gamma and Weibull distribution can fit the data list with 55 points but do not fit all other subsets, and beta general distribution is ranked around tenth. It seems that it's a bad idea to ignore all the distributions with negative numbers.

We hope the distribution we choose can always reflect the behavior of RDI no matter how many years of data we have. Therefore, the stability iteration can be used as the most important one to decide which distribution we should use.

For index value simulation method, we should consider the limit of the lack of data records too. If the data length is too short, this method doesn't have many advantages against historical burn analysis. However, if we check the stability of the distribution, it would make sense that this distribution can somehow reflect the frequency of RDI index, and the option price based on this distribution should be reasonable.

4.3 The Results of the Option Price

Once the random numbers are generated using the selected distributions, the price of the option contract can be computed using Equation (4.1). In these calculations it is assumed r=0.1 and we use the same strike level as those in Chapter 3.

4.3.1 Option Prices after Tests

Tables 4.7 depicts the computed option prices obtained from distributions selected using the ranking and the non-negative tests with different strike levels. Here the number of random numbers is 100000. For all the periods and the distributions considered, the option prices seem to converge and there is very little discrepancy from the results of the historical burn analysis. However, we should also see that the relative error for the last two periods is higher than others. After non-negative check, the distribution left is ranked low, which means it could not describe the RDI well.

This finding is reasonable as the results from historical burn analysis and index value simulation both reflect the expectation of the profits of the option, so that as long as we use sufficiently large enough data records, the results from the two methods are expected to be relatively close.

The remaining question from the previous section is whether it is necessary to do the non-negative check. For most distributions, no matter which period we choose, they could always generate some negative numbers. Table 4.8 depicts the case in which the negative numbers are discarded and the total random number size is truncated accordingly. We list the number of negative numbers in all 100000 numbers, option prices for the distributions in Table 4.1, and the biggest relative error compared to historical option prices.

Based on Table 4.2, for the period from March to August, during non-negative check with 50000 random numbers, beta general and extreme value distributions

Table 4.7: Option Price Compared with Historical Ones

	1	2	3	Historical	RE
Jan-Dec	Triang	Weibull	ExtValue		
K=mean	0.05754521	0.05721736	0.05822185	0.05823105	1.772%
K=0.7	0.17486741	0.17685126	0.1795586	0.17638766	1.766%
Mar-Aug	Gamma	BetaGeneral	ExtValue		
K=mean	0.06804662	0.06814147	0.06961309	0.07012366	3.052%
K=0.7	0.18814329	0.18745995	0.19107193	0.18899318	1.088%
Apr-Aug	Gamma	InvGauss	ExtValue		
K=mean	0.07216899	0.0719541	0.07367136	0.07410387	2.988%
K=0.7	0.18684759	0.18698024	0.18968462	0.18724644	1.285%
Jun-Aug	ExtValue	Weibull	Triang		
K=mean	0.09457096	0.09462239	0.09922325	0.09473529	4.523%
K=0.7	0.14479546	0.142197	0.14659589	0.14405319	1.734%
Jul-Aug	Weibull	Triang	BetaGeneral		
K=mean	0.13302472	0.12397935	0.13969621	0.13648905	10.090%
K=1	0.29032712	0.26737847	0.29147417	0.29134917	8.965%
Apr-Jun	Gamma	BetaGeneral	Weibull		
K=mean	0.05915839	0.05977087	0.06221157	0.05862795	5.760%
K=0.5	0.19173244	0.19175319	0.19252778	0.19633826	2.402%

Table 4.8: Option Prices Compared with Historical Ones including Negative Points

	1	2	3	Historical	RE
Jan-Dec	Triang	Weibull	ExtValue		
Negative #	0	0	0		
K=mean	0.05754521	0.05721736	0.05822185	0.05823105	1.772%
K=0.7	0.17486741	0.17685126	0.1795586	0.17638766	1.766%
Mar-Aug	Gamma	Pearson5	BetaGeneral		
Negative #	0	0	0		
K=Mean	0.06804662	0.06753036	0.06786055	0.07012366	3.840%
K=0.7	0.18814329	0.18783847	0.18868324	0.18899318	0.615%
Apr-Aug	Gamma	InvGauss	ExtValue		
Negative #	0	0	0		
K=mean	0.07216899	0.0719541	0.07367136	0.07410387	2.988%
K=0.7	0.18684759	0.18698024	0.18968462	0.18724644	1.285%
Jun-Aug	InvGauss	Lognorm	Pearson5		
Negative #	4	7	8		
K=Mean	0.0928481	0.09259562	0.09120931	0.09473529	3.866%
K=0.7	0.14207453	0.1417632	0.14032337	0.14405319	2.658%
Jul-Aug	ExtValue	Weibull	InvGauss		
Negative #	28	0	27		
K=Mean	0.13462023	0.13302472	0.13140339	0.13648905	3.870%
K=1	0.2961963	0.26737847	0.2908061	0.29134917	8.965%
Apr-Jun	Log-logistic	ExtValue	Pearson5		
Negative #	77	39	13		
K=Mean	0.05787283	0.05833659	0.0587987	0.05862795	1.305%
K=0.5	0.19406525	0.19280756	0.19321773	0.19633826	1.615%

cannot be used. However, when we try to generate 100000 random numbers in Table 4.8, both of them go through the non-negative check. For some distributions, it's hard to decide whether it can generate negative numbers or not. We should also notice

that for the last two periods, the relative errors become lower than those in Table 4.7. For the period from July to August in Table 4.8, Weibull distribution gives us the worst result with strike 1 even though only this distribution guarantees positivity.

Let's see the period from April to June in Table 4.3, the fourth distribution, logistic distribution gives us 1415 negative numbers with 50000 in total. If we calculate the option price with strike K=0.5 following this distribution, the price would be 0.19102675 with relative error 2.74%. We could say this result isn't good compared to the first three distributions, but not as bad as those based on non-negative distributions. Therefore, as expected, those distributions which generate fewer negative numbers yield reasonable option prices.

It seems that the ranking iteration is more important than the negative value check. The distribution with good ranking will lead to better results. Our best choice is the distribution with good ranking which can only generate positive points. However, negative values are not as bad as we thought as long as the negative points are below 1% in total.

We can choose one distribution as our best fitting one based on Table 4.4-4.6 for different periods by using the stability check. The results based on subsets of data records are presented in Table 4.9 for periods from January to December, April to August, and April to June separately, using constant strike value as before.

For the period from April to August, after the stability test, we can see that the inverse Gaussian distribution and the extreme value distribution both have good rank

Table 4.9: Option Prices with different length of data

1-12(0.7)	1951-2006	1957-2006	1967-2006	1977-2006	1987-2006	RE
Weibull	0.17685126	0.17879697	0.1815513	0.18046826	0.15584478	13.479%
Historical	0.17638766	0.17986768	0.1812595	0.17987126	0.15610641	12.992%
4-8(0.7)	1951-2006	1957-2006	1967-2006	1977-2006	1987-2006	RE
InvGauss	0.18698024	0.18524225	0.18849903	0.18220057	0.14724776	26.983%
ExtValue	0.18968462	0.18737687	0.19307224	0.18428832	0.15345643	23.608%
Historical	0.18724644	0.18548385	0.18769	0.17968787	0.14533541	28.837%
4-6(0.5)	1951-2006	1957-2006	1967-2006	1977-2006	1987-2006	RE
ExtValue	0.19280756	0.19559308	0.18973144	0.16469412	0.1521844	26.693%
Pearson5	0.19321773	0.19927843	0.19316501	0.16751694	0.15047973	28.401%
Historical	0.19633826	0.19906111	0.19267914	0.16857261	0.15046567	30.487%

with different lengths of data sets. However, results from the extreme value distribution seems a little better than those obtained from the inverse Gaussian distribution, with a smaller relative error when compared to the historical burn analysis. The results are similar for the period from April to June. Therefore, during the stability check, if more than one distributions have a good rank based on different data lengths, maybe we should choose the one whose value of parameters keep more stable than others.

Comparing the results with those from the historical burn analysis, we can see that the option prices are similar. For the results based on data in the most recent 20 years, those from the index value simulation is closer to the original ones than those from the historical burn analysis. Therefore, if we don't have sufficient data for the RDI, the index value simulation model is more reliable than the historical burn analysis, even though the results are not so good. Similarly, for the period from January to December, we think Weibull distribution is much better than other ones: its rank is good, it can only generate non-negative numbers and it's stable for our subsets of data lists. Even though the results with data in recent 40 or 30 years looks a little far away from the original values, here we should still use this result as the option price, which means the Weibull distribution can reflect the characteristic of RDI index.

When we don't have enough data for a stability check, it's difficult to decide which distribution we should choose only based on the ranking iteration and the non-negative check. In this case we should just choose the one with the best rank. Maybe this distribution would not fit the RDI list if we get more data, but right now we cannot avoid that possibility.

4.3.2 The Convergence of the Model

In this section, we consider the convergence of the Monte Carlo method. We consider the three periods and generate 100000 random numbers for each distribution to test

whether or not the results converge. As shown in Table 4.10 for the April to August period, the price remains 0.18. However, it should be noted that convergence is generally slow as shown on the table. This finding is consistent with the discussion provided in Section 2.6.2. Increasing the random number size by a hundred-fold will only improve the accuracy by one significant digit. In general, increasing the size of the random number beyond 100,000 is not feasible if one is using the BestFit algorithm and a more powerful algorithm must be considered in this case.

Table 4.10: Option Price with 100000 random numbers

	Type	Price 1	Price 2	Price 3	Price 4
1-12(0.7)	Weibull	0.17685126	0.17620741	0.17659686	0.17581592
4-8(0.7)	InvGauss	0.18698024	0.18711332	0.18686204	0.18592518
4-6(0.5)	ExtValue	0.19280756	0.19301155	0.19284184	0.1931022

Therefore, the multiple roots simulation algorithm which is described in Section 2.6.2 is considered as an alternative to the BestFit. The algorithm is written in Java code as shown in Appendix B. Here only the inverse Gaussian distribution is considered. For the April to August period, the results for larger random number sizes are presented in Table 4.11. Clearly the method converges. Moreover, using the low-discrepancy, Van der Corput sequence to replace the uniform random variables in Java gives a similar conclusion.

The option prices simulated from the inverse Gaussian distribution with the pa-

rameters: $\mu = 0.7981, \lambda = 14.0822, shift = -0.2691$ are shown in Table 4.11. As the size of the random number increases from 1000 to 1,000,000 the results based on uniform variable in Java are slowly converging to three digits. However, those from Van der Corput sequence seem to converge faster. The result from both method are consistent with that of the BestFit.

Table 4.11: Convergence Check for the Monte-Carlo Method

4-8(0.7)	1000	10000	100000	1000000	BestFit	Historical
Java	0.18723542	0.18961918	0.1868792	0.186384	0.18698024	0.18724644
Van der Corput	0.1871167	0.18651146	0.18665177	0.186636	0.18698024	0.18724644

In a word, if we want to get a more accurate option price based on index value simulation model, we could use Van der Corput sequence to replace the basic random sequence and generate much more random numbers following our best distribution. However, we should notice that it is considerable work to convert every distribution generator by ourselves.

4.4 The Procedure for the 2nd Model

As discussed in the previous sections, in order to use the index value simulation model one must consider a procedure to select the appropriate distribution and a procedure to compute the option prices.

In general, if the time scale is short, once the distributions are ranked with respect to the three goodness tests and the non-negative test is carried out, the distribution with good rank and relatively smaller number of negative random numbers is selected. However, if the time scale is relatively long then in addition one must consider the issue of stability. As discussed in the previous section, a subset of the RDI list must be considered at different years from which the option prices are computed.

As the size of the random number is limited in the BestFit algorithm, one needs to consider other algorithms if larger random number sizes are to be considered. In this thesis a couple of different approaches have been presented.

Chapter 5

Daily Value Simulation

In the previous two chapters two different methods were considered. While the first method was based only on the historical records therefore focused primarily on the effect of the past, the second method takes both the past and the future into account. However, both methods treat the RDI as the variable. As discussed in Chapter 2 the RDI is dependent on both precipitation and evapotranspiration which is determined by the monthly mean temperature for a specific location of the contract. Therefore, a third model called daily value simulation model is considered in this chapter.

5.1 The Description of the Model

In this model, the mean-reverting stochastic process is used to simulate the process of the daily mean temperature and monthly rainfall over the entire year. Then the period rainfall and evapotranspiration is obtained by summing up the related simulated

values. Finally, the option payoff is computed based on the repeatedly simulation of both precipitation and temperature where the average discounted payoff is used as the option price.

5.2 Simulation of Monthly Mean Temperature

Even though based on the Equation (2.5), (2.6), (2.7) and (2.8), only the mean monthly temperature is needed in order to calculate the actual evapotranspiration, it is relatively simple to simulate the daily mean temperature directly.

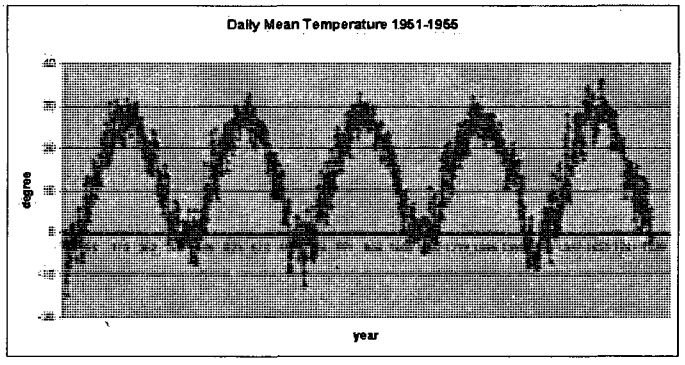

Figure 5.1: Historical Daily Mean Temperature

Figure 5.1 shows the historical records of daily mean temperature from 1951 to 1955 in Jinan station. We can see that the daily temperature is always changing following the season. It is always high in summer and low in winter. For one specific day, we can see that over 56 years, the temperature at that day is always changing around the mean value in a small range, for example, within 4 degree, depending on the climate condition for the specific region. Therefore, we can choose mean-reverting

stochastic process to simulate the behavior of daily temperature (Alaton, 2002; Benth, 2007).

We have the following Stochastic Differential Equation (SDE)

$$dT_t = dT_t^m + a(T_t^m - T_t)dt + \sigma_t dW_t, \tag{5.1}$$

whose solution is

$$T_t = (x - T_s^m)e^{-a(t-s)} + T_t^m + \int_s^t e^{-a(t-\tau)}\sigma_\tau dW_\tau, \tag{5.2}$$

where x is T_s, the temperature at time s when 0¡s¡t, T_t is the temperature at time t, T_t^m is the mean daily temperature at time t, a reflects the speed of the mean-reversion, σ_t reflects the variation from the mean value at time t and W_τ stands for the Wiener process.

Mean Daily Temperature

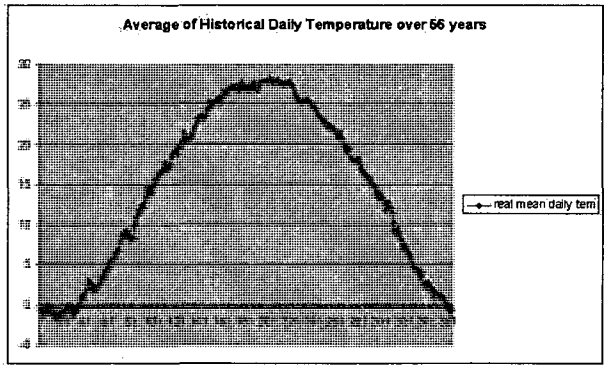

Figure 5.2: Historical Mean Monthly Temperature

T_t^m in Equation (5.1) stands for the mean daily temperature which is shown in Figure 5.2 where the average over 56 years is taken. If the mean temperature is

assumed to be a continuous variable then it is possible to simulate this curve by a sinusoidal function of the form:

$$sin(\omega t + \varphi), \tag{5.3}$$

where t denotes the time, measured in days. We let t=1,2,..., denote January 1, January 2 and so on. As we use one year (365 days) as the repetitative period, we have $\omega = 2\pi/365$. Since the yearly minimum and maximum mean temperature do not usually occur at January 1 and July 1 respectively, we shift the above curve by adding a phase angle φ. In order to account for global warming, the mean temperature is increased slightly each year through a linear variation with respect to time t. Therefore, the mean temperature is modeled as:

$$T_t^m = A + Bt + C\sin(\omega t + \varphi). \tag{5.4}$$

The Variation

The term $\sigma_t dW_t$ in Equation(5.1) reflects the discrepancy of real temperature from the simulated mean temperature. Actually, to make the problem easier, we choose the function σ_t to be a piecewise constant function. Here we assume the value of σ_t is a constant number for each month.

Parameter Estimation

Expanding Equation (5.4), we get

$$T_t^m = a_1 + a_2 t + a_3\sin(\omega t) + a_4\cos(\omega t), \tag{5.5}$$

where

$$A = a_1, \tag{5.6}$$

$$B = a_2, \tag{5.7}$$

$$C = \sqrt{a_3^2 + a_4^2}, \tag{5.8}$$

$$\varphi = \arctan(\frac{a_4}{a_3}) - \pi. \tag{5.9}$$

We can use the Gauss-Newton method to make the data fit this function, by solving

$$min_\xi [\![T_t - X]\!]^2, \tag{5.10}$$

where X is the data vector, ξ is the parameter vector (a_1, a_2, a_3, a_4). This formula means we want to reduce the discrepancy between the real temperature and the simulated ones as much as we can. Here the Gauss-Newton method is used as described in Chapter 2. The relative Java code is attached in Appendix B.

As σ_t reflects the variation of daily temperature, the first estimator is based on the quadratic variation of T_t:

$$\sigma_\mu = \frac{1}{N_\mu} \sum_{j=0}^{N_\mu - 1} (T_{j+1} - T_j)^2, \tag{5.11}$$

where N_μ is the number of days in month μ.

If we discretize Equation (5.1), for a given month μ, we have

$$T_j = T_j^m - T_{j-1}^m + aT_{j-1}^m + (1-a)T_{j-1} + \sigma_\mu \epsilon_{j-1}, j = 1, ..., N_\mu, \tag{5.12}$$

where $\{\epsilon_j\}_{j=1}^{N_\mu - 1}$ follow a standard normal distribution. We use $\bar{T}_j$ to stand for $T_j - (T_j^m - T_{j-1}^m)$. We can change (5.12) to

$$\bar{T}_j = aT_{j-1}^m + (1-a)T_{j-1} + \sigma_\mu \epsilon_{j-1}, \tag{5.13}$$

which we can see as a regression of today's temperature on yesterday's temperature. Therefore, the second estimator of σ_μ is

$$\hat{\sigma}_\mu^2 = \frac{1}{N_\mu - 2} \sum_{j=2}^{N_\mu} (\bar{T}_j - \hat{a}T_{j-1}^m - (1-\hat{a})T_{j-1})^2, \tag{5.14}$$

where $\hat{a}$ is an estimator of a, which is described as follows.

To estimate the mean-reverting speed parameter a, we can use the martingale estimation functions method. After we collect observations during n days, an efficient estimator $\hat{a}_n$ of a can be obtained by solving the equation

$$G_n(\hat{a}_n) = 0, \tag{5.15}$$

where

$$G_n(a) = \sum_{i=1}^{n} \frac{\dot{b}(T_{i-1}; a)}{\sigma_{i-1}^2} (T_i - E[T_i | T_{i-1}]), \tag{5.16}$$

and $\dot{b}(T_{i-1}; a)$ denotes the derivative w.r.t a of the term

$$b(T_t; a) = \frac{dT_t^m}{dt} + a(T_t^m - T_t). \tag{5.17}$$

Therefore, we have

$$\dot{b}(T_{i-1}; a) = T_t^m - T_t. \tag{5.18}$$

Then, to solve Equation (5.15), the only thing we need to know is the formula of $T_i - E[T_i|T_{i-1}]$. Since we know the solution of Equation (5.1) is (5.2), and that the expectation of the integration part is 0, when t=i and s=i-1, we have

$$E[T_i|T_{i-1}] = (T_{i-1} - T_{i-1}^m)e^{-a} + T_i^m. \tag{5.19}$$

Then, we have

$$G_n(a) = \sum_{i=1}^{n} \frac{T_{i-1}^m - T_{i-1}}{\sigma_{i-1}^2}[T_i - (T_{i-1} - T_{i-1}^m)e^{-a} - T_i^m]. \tag{5.20}$$

It's easy to estimate the parameter a from

$$\hat{a}_n = -ln(\frac{\sum_{i=1}^{n-1} Y_{i-1}(T_i - T_i^m)}{\sum_{i=1}^{n} Y_{i-1}(T_{i-1} - T_{i-1}^m)}), \tag{5.21}$$

where

$$Y_{i-1} \equiv \frac{T_{i-1}^m - T_{i-1}}{\sigma_{i-1}^2}, i = 1, 2, ...n. \tag{5.22}$$

Based on Equation (5.20), we know that $\hat{a}_n$ is the unique estimation.

A Correction for the Simulated Mean Daily Temperature

We use the data from Jinan station as an example. Based on the algorithm above, we get the estimation of the simulated mean daily temperature. We have a_1=13.8788, a_2=0.0035, a_3=-3.2112, a_4=-13.8329. Based on Equation (5.5), the simulated daily mean temperature should be

$$T_t^m = 13.8788 + 0.0035t - 3.2112\sin(\omega t) - 13.8329\cos(\omega t). \tag{5.23}$$

We use Equation (5.23) to simulate the mean daily temperature over the year, and compare them to the average daily temperature. In Figure 5.3, the curve called daily

tem by theta shows the simulation of mean daily temperature based on Equation (5.23) we can see that our simulated points are pretty close to the historical ones. It's also obvious that the real mean daily temperature cannot be as exact as a sine function like Equation (5.4). If we calculate the distance between the average of the temperature and the simulated one for each day, as shown in Figure 5.4, we can see that the trend of the distance is also following a sinusoidal function, with half a year as the full period. To make sure this characteristic is not a coincidence, we check the distance between the simulated and historical mean daily temperature based on data from other stations: Wuhan, Nenjiang, Xinyang, and Siping which are located in different latitude. It turns out that each curve has the same period half a year and a similar trend.

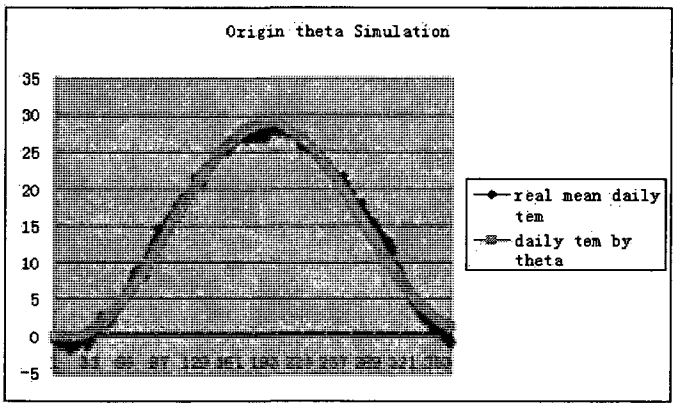

Figure 5.3: Mean Temperature Simulation

Therefore, we assume the simulation of the distance follows this sinusoidal func-

tion:

$$D(t) = d + \sum_{i=0}^{n} \beta_i sin(\frac{(2i+1)\pi(t-\nu)}{365/4}), \tag{5.24}$$

here we use n=1.

Based on Jinan's results for example, by Gauss-Newton method, we have the simulation of the distance:

$$D(t) = -0.0035 + 1.2143sin(\frac{\pi(t-35.99)}{365/4}) - 0.0615sin(\frac{3\pi(t-35.99)}{365/4}). \tag{5.25}$$

The simulated points of the distance are shown in Figure 5.4 as the red curve.

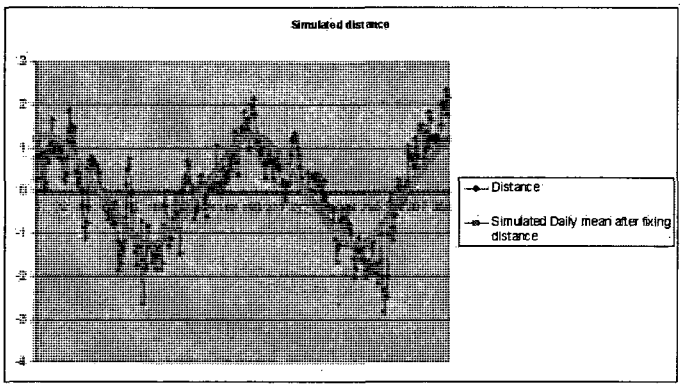

Figure 5.4: Temperature Distance Simulation

Now we can write the function of simulated mean daily temperature as the following form:

$$\bar{T}_t^m = T_t^m + D(t). \tag{5.26}$$

Therefore, we have $\bar{T}_t^m = 13.8788 + 0.0035t - 3.2112sin(\frac{2\pi t}{365}) - 13.8329sin(\frac{2\pi t}{365}) - 0.0035 + 1.2143sin(\frac{\pi(t-35.99)}{365/4}) - 0.0615sin(\frac{3\pi(t-35.99)}{365/4})$.

Then the whole curve of T_t^m for daily temperature over a year is shown in Figure 5.5 where it clearly shows that after the correction of T_t^m, the simulation is much closer to the historical one compared to the original T_t^m.

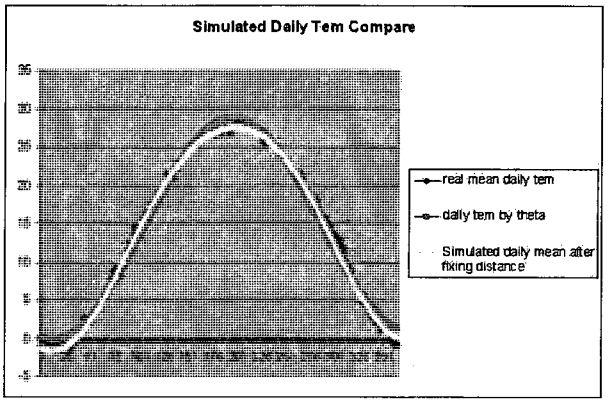

Figure 5.5: Simulated Mean Daily Temperature Compare

To confirm these findings, we calculated the mean simulated temperature for all the other stations which lead to the same conclusion.

Summary

In conclusion, to calculate the simulated list of daily mean temperature, we need to use Gauss-Newton method to estimate $a_1, ..., a_4$ in Equation (5.5) and other parameters in Equation (5.24) based on data in n years, and substitute them into Equation (5.5) and (5.24) to get the simulated daily mean temperature over a year. Then, we estimate the variation σ and the mean-reverting speed parameter a using Equation (5.11), (5.21) and (5.14) directly.

Note that it is difficult to use an integration solution like Equation (5.2) directly.

Therefore, we use a numerical method called the Implicit Milstein method which is described in Section 2.6.4.

The problem is reduced to the following iterative procedure:

$$T_{n+1}^{(0)} = T_n + (T_n^{m\prime} + a(T_n^m - T_n))\Delta t + \sigma_k \Delta W, \tag{5.27}$$

$$T_{n+1}^{(1)} = T_n + (T_{n+1}^{m\ \prime} + a(T_{n+1}^m - T_{n+1}^{(0)}))\Delta t + \sigma_k \Delta W, \tag{5.28}$$

where $T_{n+1}^{(1)}$ is the simulated point we get, σ_k is the value of diffusion parameter depending on the nth point located in month k.

Simulation of Monthly Mean Temperature

Based on data from Jinan station, we try to output the average of daily mean temperature, the simulated mean daily temperature, the average of simulated daily temperature over 10000 simulations, and the simulated daily mean temperature for one simulation for a whole year. These four curves are shown in Figure 5.6. The curve called daily tem by theta is the same one in Figure 5.3.

In the figure, the curve of the simulated mean daily temperature list is close to the average list of historical daily mean temperature. Therefore our estimate for T_t^m fits the historical data well.

5.3 The Simulation of Monthly Precipitation

Although the daily precipitation over many years does not show any specific trend, Figure 5.7 demonstrates that the monthly rainfall shows a trend that repeats every

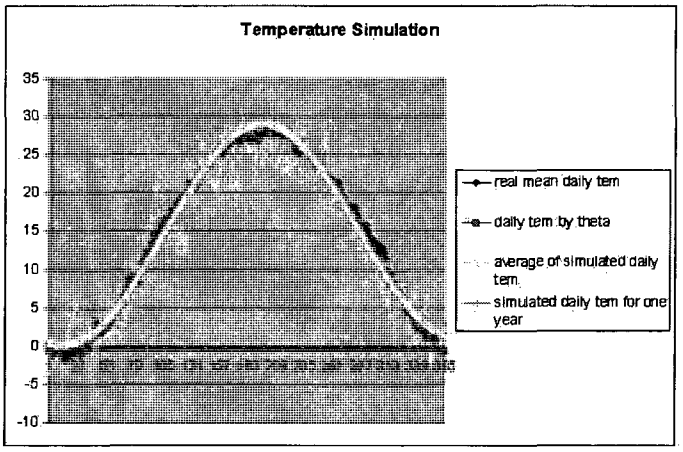

Figure 5.6: Simulated Daily Mean Temperature Compare

twelve points. Moreover, the monthly rainfall curve is shaped like the daily mean temperature. However, unlike temperature, precipitation hardly behaves in a continuous manner which makes it difficult to incorporate precipitation in the current continuous model.

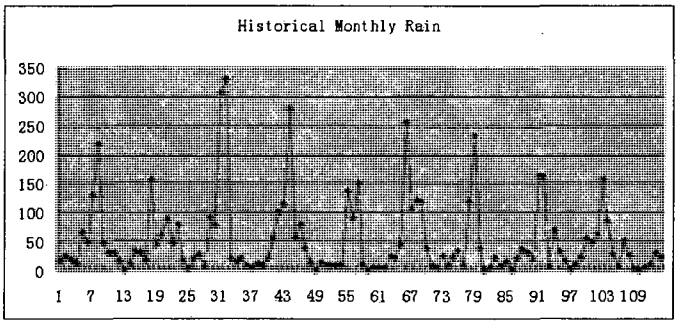

Figure 5.7: Monthly Rainfall over 10 years

5.3.1 Transforming Monthly Rain to Continuous Process

While precipitation does not behave continuously, the speed of rainfall, which is the differential of the rainfall w.r.t time, can be assumed to change continuously in time (Sachidananda, 1987). When rain starts the speed of precipitation increases continuously from zero to its peak and then decreases continuously to zero where it remains at zero until it rains again. In this formulation, the total amount of precipitation can be obtained by integrating the speed of precipitation over the period of interest as given by the formula:

$$Pre = \int_{t_2}^{t_1} P(t)\ \mathrm{d}t, \tag{5.29}$$

where $P(t)$ represents the speed of rainfall at time t, Pre represents the amount of precipitation during the period from time t_1 to time t_2.

Therefore, the monthly rainfall is essentially the monthly mean speed of precipitation with the unit $mm/(m^2 \cdot month)$. Recall that according to Equation (5.29), the amount of precipitation for each month can be expressed as $P_{ij} = \bar{p_{ij}} \cdot 1$, where P_{ij} and p_{ij} is the precipitation and the mean speed of monthly rain in the ith month and the jth year.

Therefore, in order to simulate monthly rainfall, it will be sufficient to simulate the speed of rainfall using the mean-reverting process.

5.3.2 Simulation of Monthly Precipitation Speed

As described above, the mean-reverting model will be used to simulate the speed of rain through the simulation of the monthly rainfall. Therefore the process that was used to simulate the daily temperature as described in Section 5.2 can be utilized except that the monthly values will be used instead. Thus we have the following stochastic process

$$dP_t = d\theta(t) + k(\theta(t) - P_t)dt + \sigma_t dW_t, \tag{5.30}$$

with $t \geq 0$, $k \geq 0$, and $\theta(t) > 0$, where P_t is the speed of rainfall at time t, $\theta(t)$ is the mean speed at time t, k reflects the speed of the mean reversion, σ_t reflects the variation from the mean value at time t.

After we have the simulated points, the following formulation can be used for the monthly rainfall:

$$Pre = \sum_{i=n_2}^{i=n_1} P_i \cdot \Delta t \tag{5.31}$$

where n_1 and n_2 respectively refer to for the beginning and end point in the month, P_i is the speed of rainfall at the point i, and Δt is the time interval in our simulation.

5.3.3 Mean-Reverting Process with Improvement

In the previous model, the parameter σ_t was modeled as a piece-wise function where σ_t was assumed constant for each month. In order to make the model more realistic, the diffusion part can be modeled as a function of P_t which leads to the following

SDE

$$dP_t = d\theta(t) + k(\theta(t) - P_t)dt + \sigma P_t^p dW_t, \tag{5.32}$$

with $t \geq 0$, $k \geq 0$, and $\theta(t) > 0$. As described in Equation (5.1), $\theta(t)$ is the simulated mean speed of monthly rain at time t over a year and a is the mean-reverting speed parameter. Here σ and p together reflect the volatility which is now dependent on previous time precipitation.

Simulated Mean Monthly Precipitation

Figure 5.8 depicts the mean monthly rain speed over 56 years using the daily data for the Jinan station. As in the temperature model, the real mean monthly precipitation will be assumed to fit a sinusoidal function of the form (Emmerich, 2005):

$$\theta(t) = m + \alpha sin(\frac{2\pi(t-\upsilon)}{12}), \tag{5.33}$$

where m is the mean of the sine curve, α determines the oscillation and υ is the horizontal shift.

To make the simulated curve closer to the historical mean one, the simulation can also be further improved by expanding θ in terms of the Fourier series(Emmerich, 2005):

$$\theta(t) = m + \sum_{i=0}^{n} \alpha_i sin(\frac{(2i+1)\pi(t-\upsilon)}{6}). \tag{5.34}$$

Of course, if the monthly rainfall data is directly substituted over the 56 period, it will be equivalent to using the historical mean monthly value rather than the real

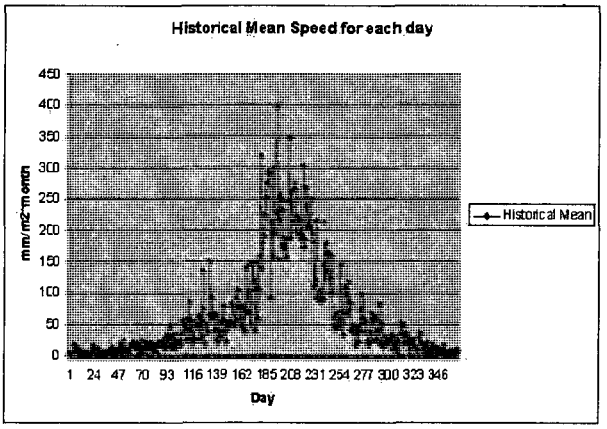

Figure 5.8: Mean Speed of Monthly Rain over 56 years

time speed which can lead to significant error in the simulation. In order to minimize this error, the amount of rain is substituted on a daily basis

Parameter Estimation

For the parameters in $\theta(t)$, one can still use the Gauss-Newton method to solve the least-square problem similar to Equation (5.6). However, according to Equation (5.34), first we need to decide how many α_i we should choose for the specific problem. The details of this analysis and the results will be shown in the next section.

Then, we want to find an unbiased estimation of mean-reverting parameter k. The original estimation of k is (Emmerich, 2005)

$$\hat{k} = \frac{1}{n}\sum_{i=0}^{n-1} \frac{P_{i+1} - P_i - \theta(i+1) + \theta(i)}{(\theta(i) - P_i)\Delta}. \tag{5.35}$$

Because this estimation will change with different length of data, we have the modification as:

$$\hat{k_b} = \frac{1}{\#I_b}\sum_{i \in I_b} \frac{P_{i+1} - P_i - \theta(i+1) + \theta(i)}{(\theta(i) - P_i)\Delta}, \tag{5.36}$$

with $I_b = \{i = 1, ..., n : |\theta(i) - P_i| > b\}$, b is a positive real number. With b getting bigger, $\hat{k}_b$ is convergent. Here we choose to use b=50 to get a proper estimation.

To estimate the diffusion parameters, first we square Equation (5.32) to get

$$(dP_t)^2 = [\theta'(t) + k(\theta(t) - P_t)dt + \sigma P_t^p dW_t]^2. \tag{5.37}$$

Then, based on Ito's integration rule, we have

$$\ln(dP_t)^2 = 2\ln(\sigma) + \ln(dt) + 2p\ \ln(P_t), \tag{5.38}$$

such that $ln(dP_t)^2$ and $ln(dP_t)$ have a linear relationship. The data list can now be substituted into Equation (5.38), from which one can use the linear regression method to get σ and p directly.

After the parameter estimation, based on (2.28), (2.29) and (5.32), the speed of monthly precipitation can be simulated by the following formula:

$$P_{n+1}^{(0)} = P_n + (\theta'(n) + k(\theta(n) - P_n^{(0)}))\Delta t_n + \sigma {P_n^{(1)}}^p \Delta W_n + \frac{1}{2}\sigma^2 p {P_n^{(1)}}^{2p-1}(\Delta W_n^2 - \Delta t_n), \tag{5.39}$$

$$P_{n+1}^{(1)} = P_n + (\theta'(n+1) + k(\theta(n+1) - P_{n+1}^{(0)}))\Delta t_n + \sigma {P_n^{(1)}}^p \Delta W_n + \frac{1}{2}\sigma^2 p {P_n^{(1)}}^{2p-1}(\Delta W_n^2 - \Delta t_n), \tag{5.40}$$

where $P_{n+1}^{(1)}$ is the simulated point at the n+1 step, $n = 0, ...N$, and N is the step size for one simulation process.

5.4 Correcting the Precipitation Model

Keeping the Positivity of the Simulation

Since P_t represents the speed function, it must be non-negative to have meaningful interpretation. Moreover, according to Equation (5.39) and (5.40), negative P_t lead to no value for the part P_t^{2p-1}. However, there is no guarantee that the simulation of Equation (5.32) yields positive values, especially when P_t is close to 0 at the beginning of the year (Kahl, 2004).

If we use X_t to stand for a stochastic process with

$$P(\{X_t > 0 \text{ for all } t\}) = 1, \tag{5.41}$$

the stochastic integration scheme possesses an eternal life time if

$$P(\{X_{n+1} > 0 | X_n > 0\}) = 1. \tag{5.42}$$

Otherwise, we say it has a finite life time.

If we can find a numerical method to solve this model which has an eternal life, as long as the initial value of historical precipitation is positive, we can make sure that all the following simulated points are positive. This is another reason to choose the implicit Milstein method to do the simulation (Emmerich, 2005).

For the mean-reverting process given by

$$dX_t = (\alpha - \beta X_t)dt + \sigma X_t^p dW, \tag{5.43}$$

with $\alpha, \beta, \sigma, p \in R^+$ and $p > \frac{1}{2}$, the Milstein method provides numerical positivity with the following restriction:

$$\Delta t < \frac{1}{\sigma^2}. \tag{5.44}$$

Compared with Equation (5.43), Equation (5.32) will guarantee positivity with

$$\alpha = k \cdot \theta(t) + \theta'(t) \in R^+. \tag{5.45}$$

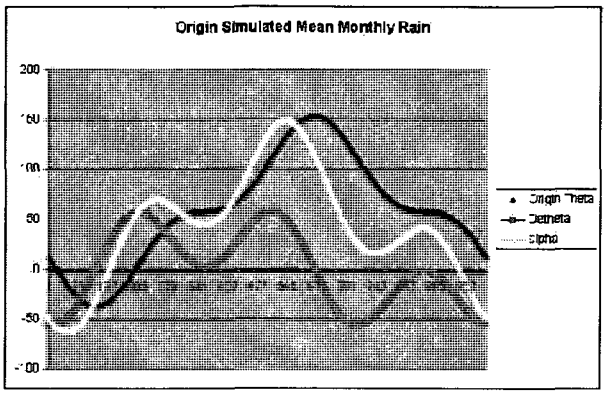

Figure 5.9: Simulated speed of monthly rain in Jinan

We now use the daily data from Jinan station with n=1, and get the parameter estimation as:

$$\theta(t) = 58.931 + 75.5712sin(\frac{\pi(t - 27.7482)}{6}) - 24.4185sin(\frac{3\pi(t - 27.7482)}{6}), \tag{5.46}$$

and k=2.2275, σ=4.3080, p=0.7762.

In Figure 5.9, the curve of Origin Theta shows the value of $\theta(t)$ based on Equation (5.46), the curve called Detheta shows the differential of $\theta(t)$, and the curve of alpha shows value of α based on Equation (5.45). As Figure 5.9 shows, the simulation of the mean speed of monthly precipitation does not stay positive due to the value α which is not always positive. Moreover, for our model whose diffusion part has term P_t, once a negative P_t value is obtained, according to Equations (5.39) and (5.40),

the simulated value for next step could never be calculated since $p < \frac{1}{2}$, which would lead to $2p - 1 < 0$ and no value for P_n^{2p-1}.

Therefore, in order to make sure positivity, $\theta(t)$ must be modeled appropriately. Since $\theta(t)$ is a continuous sinusoidal function and the speed of mean reversion k is assumed constant, there should be a boundary for the value of α in Equation (5.45) that ensures $\alpha > 0$. Thus we require,

$$\theta(t) > \frac{-\theta'(t)}{k}, \; t = t_0, ..., t_N. \tag{5.47}$$

If we only add a constant on the right hand of Equation (5.34) to make sure the new $\theta(t)$ satisfies Equation (5.45), we still have the same $\theta'(t)$. Therefore we must find $max_{t=t_0,...,t_N}(-\frac{\theta'(t)}{k})$ denoted by M so that the adjusted simulated mean speed of precipitation could be expressed as $\theta_{ad}(t) = \theta(t) + M + 1$, and $\theta'_{ad}(t) = \theta'(t)$.

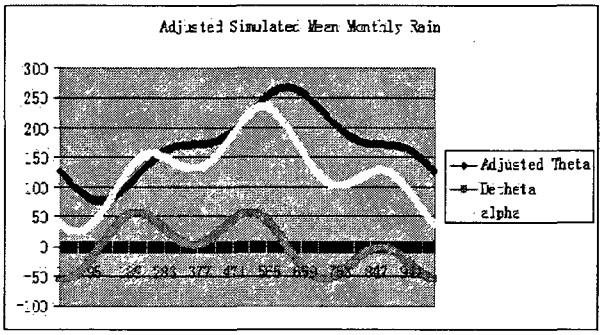

Figure 5.10: Adjusted Simulated speed of monthly rain in Jinan

Still using the Jinan data as an example, the results of $\theta_{ad}(t)$ are shown as the curve of Adjusted Theta in Figure 5.10 and the corrector for $\theta(t)$ is -40.977. The curve Detheta and alpha has similar meanings like those in Figure 5.9.

To use new θ in the algorithm, we need to notice that all the simulated points P_t here has mean value $\theta_{ad}(t)$. This change could also affect the diffusion part. In order to keep the process still similar to the original one, we have to modify the algorithm as follows:

1.$E = max(P_n^{(1)} - M - 1, v)$, where v is a uniform random variable in (0,0.1). If we use 0 to replace v, the diffusion part could be 0 if $2p-1 > 0$ and infinity if $2p-1 < 0$.

2.change (5.39) and (5.40) to

$$P_{n+1}^{(0)} = P_n + (\theta'(n) + k(\theta(n) - P_n^{(1)}))\Delta t_n + \sigma E^p \Delta W_n + \frac{1}{2}\sigma^2 p E^{2p-1}(\Delta W_n^2 - \Delta t_n), \quad (5.48)$$

$$P_{n+1}^{(1)} = P_n + (\theta'(n+1) + k(\theta(n+1) - P_{n+1}^{(0)}))\Delta t_n + \sigma E^p \Delta W_n + \frac{1}{2}\sigma^2 p E^{2p-1}(\Delta W_n^2 - \Delta t_n). \quad (5.49)$$

3. After calculating all the simulated points, we need to change the whole list back down to the original level:

$$P_n = P_n^{(1)} - M - 1. \quad (5.50)$$

Note that if $p < \frac{1}{2}$, then positivity is not yet guaranteed for large size simulation. For now we can run the program ignoring the negative values until an acceptable point is obtained.

A Correction for Simulated Mean Speed of Rain

Note that, since the number of steps simulated we choose to use over 12 months is 1000, Δt should equal 0.012. There are about 83 points in each month's simulation. Based on Equation (5.31) and (5.34), we can get the simulated mean monthly rain

over a year. The results are shown in Figure 5.11. The red curve is the average of monthly rainfall over 56 years. The yellow curve is our simulated value for mean monthly rainfall which is clearly inaccurate. Since the data used is for the last 56 years, the average monthly rain for a whole year is expected to be close to the real mean monthly rainfall.

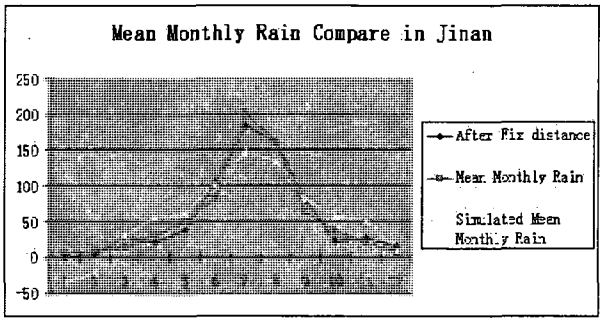

Figure 5.11: Simulated Mean Monthly Rain Compare

Note that, n=1 was used in Equation (5.46). However, in Figure 5.12, we have a similar simulation except when n=0. Thus n=1 is chosen for the model.

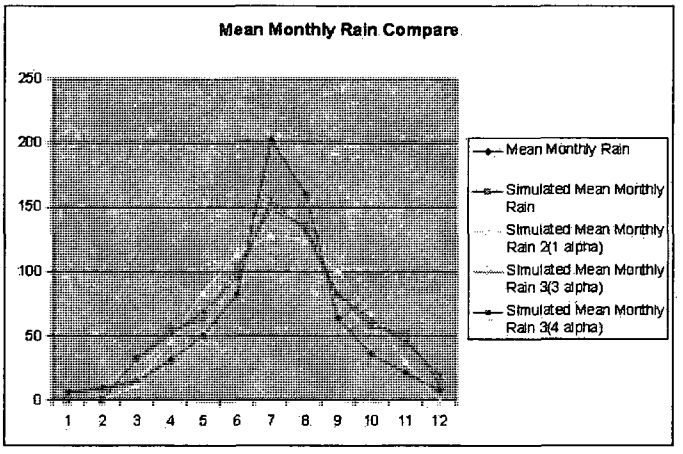

Figure 5.12: Simulated Mean Monthly Rain with Different Number of alpha

Similarly like the simulation of mean daily temperature, we also use a sinusoidal function to correct the discrepancy between the simulated and historical curve. Figure 5.13 shows the distance between the historical records and simulated points based on Equation (5.46). Based on Figure 5.13, we can still assume the period of the distance is 12 (months). This could change depending on the condition of different places.

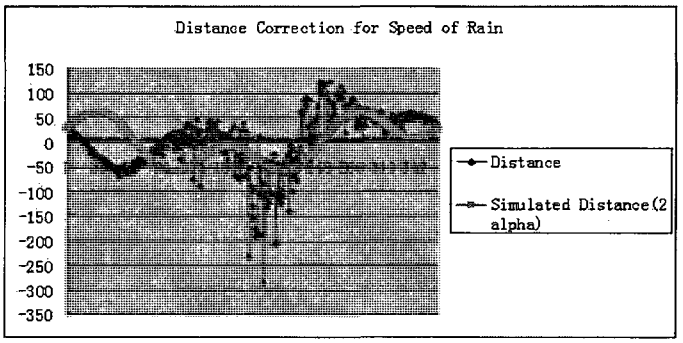

Figure 5.13: Distance between real and simulated speed of monthly rain

Therefore, we can assume the simulation of the distance to be of the form:

$$D(t) = d + \sum_{i=0}^{n} \beta_i sin(\frac{(2i+1)\pi(t-\nu)}{6}), \tag{5.51}$$

here we still use n=1.

Based on Jinan's results for example, by Gauss-Newton method, we have the simulation of the distance:

$$D(t) = 0 - 36.7251sin(\frac{\pi(t+0.7054)}{6}) - 0.9923sin(\frac{3\pi(t+0.7054)}{6}). \tag{5.52}$$

The simulated points of the distance are shown in Figure 5.13 as the red curve. Even though this new curve looks not so close to the original one at the beginning of the year, it's a continuous correction for our $\theta(t)$.

Now we can write the function of simulated mean monthly precipitation as the following form:

$$\bar{\theta}(t) = \theta(t) + D(t). \tag{5.53}$$

Therefore, we have $\bar{\theta}(t) = 58.931 + 75.5712sin(\frac{\pi(t-27.7482)}{6}) - 24.4185sin(\frac{3\pi(t-27.7482)}{6}) - 36.7251sin(\frac{\pi(t+0.7054)}{6}) - 0.9923sin(\frac{3\pi(t+0.7054)}{6})$.

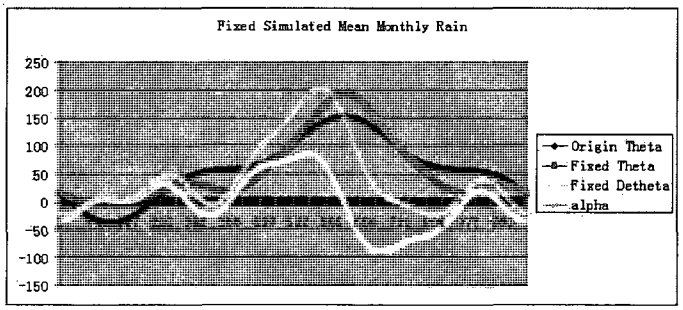

Figure 5.14: Fixed Mean Monthly Rain for a Whole Year

Then the whole curve of $\theta(t)$ with 1000 points called Fixed Theta is shown in Figure 5.14 where the newly simulated points for each months are shown as the blue curve in Figure 5.11. In Figure 5.14, the curve Origin Theta is the original simulation of mean speed of rainfall given by Equation (5.46). We also calculate the differential of $\bar{\theta}(t)$ and the value of α based on Equation (5.53) and (5.45), shown as the curve Fixed Detheta and alpha in Figure 5.14. Clearly, after correcting $\theta(t)$, the simulation is much closer to the historical one compared to the original $\theta(t)$ even though the simulation of the difference in Equation (5.52) isn't a good one.

To check our results, we can also calculate the mean simulated speed of rain for other stations. The results are similar to the above. We can say that the new

simulation of mean monthly precipitation is better than the original one.

Combination of the Modification

From the α curve in Figure 5.14, we can see that the correction function $D(t)$ cannot help avoiding the error from the limitation of positivity. Therefore, we need to combine the correction and the limitation of positivity for each simulation.

Then, the algorithm for the simulation of monthly precipitation become:

1. Decide the number of steps we should use for the simulation during a whole year based on Equation (5.44). Here we use 1000.

2. Do parameter estimation based on Equation (5.34)-(5.38).

3. Change $\theta(t)$ to $\bar{\theta}(t)$ based on (5.51) and (5.53).

4. At each step, the numerical simulation should follow the algorithm (5.48)-(5.49) in the first part of this section.

5. After the simulation, we substitute all simulated points in Equation (5.31) to get the simulated monthly rainfall and use them to evaluate the option price.

5.5 Analysis of the Results

In this section, we analyze the simulation of temperature and precipitation. Then we use them to estimate the value of drought option contract based on daily precipitation and speed of precipitation, and compare the results with those based on historical burn analysis and index value simulation.

5.5.1 The Results of Daily Temperature Simulation

Based on the model in Section 5.2, we can simulate the daily mean temperature with an initial value. To simplify the programming, here we set the initial value to be the real mean daily temperature on December 31st in the last year in our data. And we only simulate the daily mean temperature over a year, because in our contract we only predict the drought condition for one year.

Then, we calculate the mean monthly temperature and substitute them into Equation (2.6)-(2.8) to calculate the actual evapotranspiration.

We use data from Jinan and Siping station for example. In the previous section we already get the simulated mean daily temperature curve. Then, based on Equation (5.11) and (5.14), we get two series of σ. The results are shown in Table 5.1. From this table, we can see that the curve of σ for each month based on Equation (5.11) is smoother than those from Equation (5.14), which obviously varies between every month. Of course, we cannot conclude that the estimator of Equation (5.14) is better than (5.11) only based on these curves.

Therefore, we simulate the daily mean temperature over a year for 10000 times, and calculate monthly mean temperature for each simulation. Then, we calculate the mean and standard deviation of monthly mean temperature for historical data and simulated value for each month. The results are shown in Figure 5.15 and 5.16. We can see that the mean of historical and simulated value in Figure 5.15 are really close. The simulated mean monthly temperature is less than 2% away from the mean of

Table 5.1: Sigma Estimator for Daily Temperature

Month	Sigma Estimator 1	Sigma Estimator 2
Jan	2.473412771	2.535306941
Feb	2.726849447	2.939145292
Mar	2.89676842	3.116445658
Apr	3.19882791	3.220360937
May	3.086592084	2.888380112
Jun	3.062162518	2.431683392
Jul	2.712585873	1.966805113
Aug	2.660742508	1.751119392
Sep	2.645188804	1.901978641
Oct	2.720021982	2.412730342
Nov	2.698762944	2.831803427
Dec	2.667227284	2.617969293

the historical data. Even though for January and December, the simulated mean is around 40% higher than the historical ones, we should also notice that the absolute distance between the two values is around 0.5 °C, which is really small as a change of temperature.

Figure 5.16 shows that based on different σ series, the standard deviation for each month is also different. We can see that the standard deviation doesn't change a lot

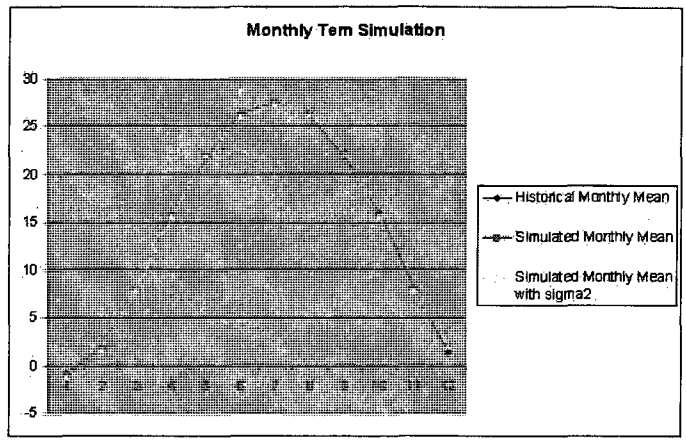

Figure 5.15: Monthly Temperature Simulation

for the first 4 months and last 2 months as the estimation of σ is pretty close to each other. But for other months, those from Equation (5.14) are closer to the standard deviation of historical data than those from Equation (5.11). Here we can say that Equation (5.14) can actually give us a better simulation of daily mean temperature.

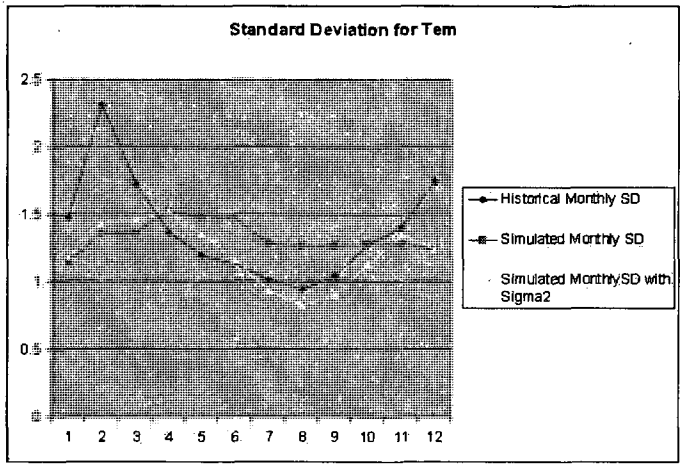

Figure 5.16: Simulated Standard Deviation Compare

By using data from Siping station, we can get the similar results. The mean of

simulated points are only around 5% away from historical results. Moreover, for most months, the relative error between monthly standard deviation based on Equation (5.14) and historical ones is lower than those based on Equation (5.11) and historical ones. They are still less than 10% away from the historical ones.

If we only simulate the process of daily mean temperature, and choose the average of historical monthly rain as the precipitation for the specific year in the contract, we can get another RDI list based on Equation (2.4), (2.6), (2.7), and (2.8). Compare them to the real RDI list over past 56 years, and the results are shown in Figure 5.17. We can see that the simulated RDI results change only over a narrow range.

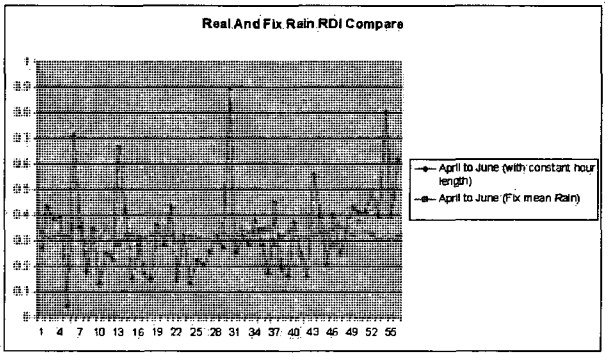

Figure 5.17: Simulated RDI with variable Temperature

Because the influence of mean temperature on RDI is relatively small it is not worth putting more effort to further improve the diffusion parameter.

Table 5.2: Simulated Mean Monthly Rain Compare

Month	Historical	Simulated	Average	Absolute error	RE of mean
Jan	6.2661	0.3840	1.5482	5.8820	93.87%
Feb	9.6607	3.8419	4.2409	5.8188	60.23%
Mar	14.3393	28.4677	27.7626	14.1284	98.53%
Apr	31.2286	20.9910	22.5753	10.2376	32.78%
May	49.4125	37.1675	39.7941	12.2450	24.78%
Jun	83.1179	106.1426	106.0027	23.0247	27.70%
Jul	201.9679	183.8513	186.5088	18.1165	8.97%
Aug	160.1911	160.5453	163.3399	0.3542	0.22%
Sep	63.0625	78.7743	91.5232	15.7118	24.91%
Oct	36.1964	22.8629	36.1383	13.3336	36.84%
Nov	20.9357	26.5345	31.3882	5.5988	26.74%
Dec	7.8893	14.7048	17.9086	6.8155	86.39%

5.5.2 The Results of Monthly Precipitation Simulation

Based on the model in Section 5.3, we can get the simulated monthly rainfall as long as we know the initial amount of precipitation of a specific month. Right now the initial value is the precipitation in the last December. Then we generate the simulated monthly precipitation for a whole year.

Based on the same model as the one we use for temperature simulation with

Table 5.3: Monthly Rain Standard Deviation Compare

Month	Historical	Simulated	Absolute error	RE of SD
Jan	8.0596	3.3362	4.7234	58.61%
Feb	8.7385	5.2941	3.4443	39.42%
Mar	11.1763	10.4675	0.7088	6.34%
Apr	26.3709	18.2921	8.0787	30.64%
May	39.4132	30.3748	9.0384	22.93%
Jun	58.8316	49.3789	9.4527	16.07%
Jul	105.5342	106.9998	1.4656	1.39%
Aug	95.1555	91.3618	3.7937	3.99%
Sep	51.4736	83.1080	31.6344	61.46%
Oct	30.5236	40.2103	9.6868	31.74%
Nov	22.4000	27.1438	4.7439	21.18%
Dec	8.4581	16.8490	8.3909	99.21%

monthly data, if we calculate the average of simulated value over all the simulation, the results are shown in Table 5.2 and 5.3. We use data from Jinan station, calculate the average of simulated monthly rain over 10000 simulations for the whole year, and compare them to the average value of the historical monthly rain. From the table we can see that the relative error of simulated mean monthly rain from April to November is between 0% to 37%. And the standard deviation in the middle of year is between

1% to 30% except the one in September. The biggest relative error in both tables is almost 100%.

Table 5.4: Simulated Mean Monthly Rain Compare by Model 2

Month	Historical	Simulated	Average	Absolute error	RE of mean
Jan	6.2661	0.0000	0.2362	6.0298	96.23%
Feb	9.6607	7.5444	7.8601	1.8006	18.64%
Mar	14.3393	25.5504	25.6453	11.3060	78.85%
Apr	31.2286	24.3370	24.1607	7.0679	22.63%
May	49.4125	39.8668	39.7016	9.7109	19.65%
Jun	83.1179	115.1314	115.0582	31.9403	38.43%
Jul	201.9679	183.5855	182.3929	19.5749	9.69%
Aug	160.1911	161.4174	160.2324	0.0414	0.03%
Sep	63.0625	76.2705	75.9524	12.8899	20.44%
Oct	36.1964	27.1644	27.2241	8.9723	24.79%
Nov	20.9357	23.8714	23.9580	3.0223	14.44%
Dec	7.8893	15.7477	16.1095	8.2203	104.20%

Then, we use the model whose diffusion part depends on the previously simulated point with daily data. From Table 5.4 and 5.5, we could see the comparison of mean monthly rainfall and standard deviation between historical and simulated results. Here the relative error of simulated mean monthly rain from April to November is

Table 5.5: Standard Deviation Compare by Model 2

Month	Historical	Simulated	Absolute error	RE
Jan	8.0596	0.5253	7.5343	93.48%
Feb	8.7385	4.8873	3.8512	44.07%
Mar	11.1763	19.0543	7.8780	70.49%
Apr	26.3709	23.0329	3.3379	12.66%
May	39.4132	26.2796	13.1336	33.32%
Jun	58.8316	54.3828	4.4487	7.56%
Jul	105.5342	91.6479	13.8863	13.16%
Aug	95.1555	93.5994	1.5561	1.64%
Sep	51.4736	65.6957	14.2221	27.63%
Oct	30.5236	30.9772	0.4537	1.49%
Nov	22.4000	21.3171	1.0829	4.83%
Dec	8.4581	19.3064	10.8483	128.26%

between 1% to 38%. And the standard deviation in the middle of year is between 1% to 34% without an odd in September. This means the second model may give us a better description of rain process.

For both methods, the simulated points at the beginning and the end of year is always far away from the historical ones. However, this trend is reasonable. We start the simulation at the beginning of year every time, if the initial value at December

31st is far away from the simulated mean value, it would take a while for the model to make it back to the mean curve. Generally, the simulation for each month looks good in the middle range of the year, which is also the growth period for most plants.

Compared with the simulation of mean monthly temperature, it's hard to say these two models show a good simulation. The relative error on the simulated mean monthly precipitation is always around 20% or 30%. However, this is consistent with other research results which are often based on the two-stage Markov chain model rather than the mean-reverting model (Cao, 2004). It is also similar to the results based on the same model based on data from Germany (Emmerich, 2005). Hence the result is not that unreasonable.

5.5.3 Option Price Comparison

Option Prices Comparison for Two Models

Here we still assume that we have the same drought option contract for Jinan as in Section 2.5.1. We set the same periods, constant strike level and data as those in the above two models. After we use daily value simulation model to estimate 100,000 possible values of monthly temperature and precipitation, we substitute them into Equation (4.1) to get the value of contract and compare it with those from historical analysis model. The results are shown in Table 5.6. Here Model 1 means the same mean-reverting model we use to simulate both daily temperature and monthly precipitation. In Model 2 we change the diffusion part and substitute daily data for

precipitation simulation.

Table 5.6: Option Price Compare for Jinan

	Jan-Dec(K=0.7)	Apr-Aug(K=0.7)	Apr-Jun(K=0.5)
Historical	0.17638766	0.18724644	0.19633826
Model1	0.14246321	0.17034594	0.0852701
Model2	0.16663164	0.18461022	0.1616114
RE1	23.813%	9.921%	130.255%
RE2	5.855%	1.428%	21.488%

From this table, we could see that the option prices from model 1 are not as good as those from model 2. For the period from January to December and from April to August, the relative error for both methods from historical results are within 20%. But for the period from April to June, the relative error of model 1 is larger than 100%, while it's only 21% for model 2. It seems like even though we cannot describe the behavior of rainfall well, the option price based on model 2 is much better than we thought. Of course we cannot make this conclusion only by this set of data but we must also check our model by substituting data from other stations.

Table 5.7 gives us the option price for the put option contract in Siping with the same period as above. We use the same strike level for each period. From each column in the two tables, we could see that even though the two places are located around 40°N, the option prices are totally different as long as the drought condition is

Table 5.7: Option Price Compare for Siping

	Jan-Dec(K=0.7)	Apr-Aug(K=0.7)	Apr-Jun(K=0.5)
Historical	0.05864985	0.08026772	0.06540103
Model1	0.05745684	0.08517523	0.01884279
Model2	0.07660143	0.10053972	0.06275855
RE1	2.076%	5.762%	247.088%
RE2	23.435%	20.163%	4.211%

different. In Table 5.7, for the first two periods, results from model 1 look closer to the historical ones. On the other hand the price for the period from April to June is too far away. Here we could say that model 2 gives us a much better estimation for the option price. The relative error from both models seems too high compared to those from the index value simulation model based on historical option prices. However, considering the significant error in modeling rainfall process this result is not that surprising.

Option Prices with Different Data Lengths

Just like Chapter 3 and 4, we can also use subsets of all the data to see the effect of time limitation on option prices. As we only use data in 40 or 20 years, all the parameter estimation should be different. Here we skip the parameter estimation part and consider the option prices.

Table 5.8-5.10 show the option prices based on different data lengths for different

Table 5.8: Option Price Compare for 1-12 in Jinan with Time Limitation

	K=0.7	1957-2006	1967-2006	1977-2006	1987-2006
Model1(H)	0.17638766	0.17986768	0.1812595	0.17987126	0.15610641
Model2(Weibull)	0.17685126	0.17879697	0.1815513	0.18046826	0.15584478
Model3(Monthly)	0.14246321	0.14311397	0.15055475	0.14541882	0.12260332
Model3(Daily)	0.16663164	0.16861026	0.1782004	0.17387247	0.15022846

Table 5.9: Option Price Compare for 4-8 in Jinan with Time Limitation

	K=0.7	1957-2006	1967-2006	1977-2006	1987-2006
Model1(H)	0.18724644	0.18548385	0.18769	0.17968787	0.14533541
Model2(ExtV)	0.18968462	0.18737687	0.19307224	0.18428832	0.15345643
Model3(Monthly)	0.17034594	0.16904443	0.17862279	0.1729103	0.14775786
Model3(Daily)	0.18461022	0.18671642	0.19884574	0.18730138	0.1583788

periods for all three models. The results for the three models are generally similar for each columns except for those in Table 5.10 which is based on monthly data. It seems the results based on daily value simulation model with daily data is closer to the results from the first two models than those with monthly data regardless of how many years of data is used.

In the last column in these three tables, the option prices change significantly

Table 5.10: Option Price Compare for 4-6 in Jinan with Time Limitation

	K=0.5	1957-2006	1967-2006	1977-2006	1987-2006
Model1(H)	0.19633826	0.19906111	0.19267914	0.16857261	0.15046567
Model2(ExtV)	0.19280756	0.19559308	0.18973144	0.16469412	0.1521844
Model3(Monthly)	0.0852701	0.083204	0.09117522	0.07966244	0.07025823
Model3(Daily)	0.1616114	0.16512109	0.16755921	0.15198266	0.14474896

regardless of the model used, which means the time limitation would always exist in our models, even though at the beginning of Section 2.5.3, we hope the index value simulation model and daily value simulation model can predict more possible values of RDI for Equation (2.11). Comparing the results from data over 20 years and those over 56 years for each model, we could see that the difference from historical burn analysis is always bigger than those from the other two. Comparing the prices in each row in Table 5.10, we could see that the prices based on mean-reverting process with daily data is more stable than the other two even though the original option prices based on data over 56 years is not accurate enough.

Finally, if sufficient data is not used then the estimated option prices will generally be inaccurate. However, even though the lack of data would always affect the accuracy of the estimation of option prices, index value simulation model and daily value process model can receive reasonable results compared to historical burn analysis model. Even

though our daily value simulation model cannot describe the behavior of precipitation well, it is still the most stable one.

Chapter 6

Conclusion

In this thesis we propose a new type of option contract that can be used to hedge against the financial risks of drought. We also develop three mathematical models to estimate the price of these options.

6.1 Conclusion

Unlike other weather derivatives, we need to consider more climate factors in drought option contracts. Such contracts potentially involve factors from biological and hydrological fields, and are dependent on not just location, but also the type of plants involved.

Right now the RDI can be seen as a good measure of agricultural drought with the limitation of the types of data, but other good indices might be developed in the future, which can include more climate factors and reflect the severity of drought.

In this thesis, we use three models to estimate the value of the drought contract for the period from January to December, April to August and April to June in Jinan with different strike levels. It turns out that the results from the three methods are very close to each other if we have enough historical data, which means that each of the three methods can reasonably be used.

However, if we wish to put this contract into practice, there may be other considerations. First of all, we should consider the relationship between the drought index and the influence of drought on farmers' income. Secondly, currently in developing countries with poor irrigation systems, it is difficult to obtain sufficient historical data sets, which means we cannot get an accurate description of drought in those places. In this situation we cannot use the historical analysis model. The index value simulation and daily index process can provide more information about future possible value of drought index. From our example with data in Jinan station, the index value simulation model is better. However, the possible disadvantage of this model is that it could be complicated to generate a large size of random numbers following different distributions to improve the efficiency of the price. The results from daily value simulation model is not as good as the second one because of the inaccurate precipitation model. However, it's simple to use only one model to describe the behavior of temperature and precipitation in different places.

6.2 Future Work

Improvements to the models presented in this thesis fall into two categories: improving the financial models or improving the weather models.

6.2.1 Improvements to the Contract

The focus of this thesis was to model drought conditions and so we ignored many financial factors to simplify the contracts. For example, as mentioned, we did not pay attention to the relationship of drought index and farmers' profits. We also ignored the change of the interest rate in the real market and, in general, we did not consider the potential market price of risk as the market is not complete (Cao, 2004).

6.2.2 Improvements to the Model

A number of improvements can also be made to improve the weather models used to price drought contracts. For instance, in this thesis we use the mean-reverting process to simulate monthly rain. Even though it is not so bad when we try to simulate the behavior in the long term, we can not say this model is efficient because the standard deviation from our simulations fluctuate too much from the real one.

Actually, we have another method which can simulate the daily rainfall. In this method, we can divide the process into two parts: the first one is the Markov chain which can approximate the probability of wet or dry days based on the whether the previous day is wet or dry; in the second part, we see the amount of precipitation

as the variable and try to find a goodness-of-fit distribution to fit the behavior of precipitation. This method is more often used when one wants to simulate the amount of rainfall in any period. This method was not implemented for this thesis.

Another problem is that we did not consider the location limitations of the climate models. The climate stations which collect precipitation data are usually located in large cities, which are far away from the farmland where the drought contract should be used. Even though the behavior of temperature may be similar between the neighborhood, the amount of precipitation may change a lot even though two locations are really close. Therefore, it is important to include the spatial basis risk in the model.

Bibliography

[1] Alley, W. M., *The Palmer Drought Severity Index: Limitations and Assumptions*, J. Clim. Appl. Meteor, **(1984)**, 23: 1271-1294.

[2] Alaton, P., Djehiche, B., and Stillberger, D, *On Modeling and Pricing Weather Derivatives.* Applied Mathematical Finance, Vol.9, **(2002)**, pp.1-20.

[3] Bartholomew-Biggs, M., *Nonlinear Optimization with Financial Applications*, Springer US, **(2005)**.

[4] Benth, F. E., and Saltyte-Benth, J., *The Volatility of Temperature and Pricing of Weather Derivatives*, Quantitative Finance, Vol.7, No.5, **(2007)**, pp.553-561.

[5] Bhuiyan, C., *Various Drought Indices for Monitoring Drought Condition In Aravalli Terrain of India*, In: Proceedings of the XXth ISPRS Conference. Int. Soc. Photogramm. Remote Sens., Istanbul, **(2004)**.

[6] Boken, V.K., Cracknell, A. P., and Heathcote, R. L., *Monitoring and Predicting Agricultural Drought*, Oxford University Press: Oxford, **(2005)**.

[7] Bordne, E. F., McGuinness, J. L., *Some Procedures for Calculating Potential Evapotranspiration.* The Professional Geographer, Vol.25, Iss.1, **(2005)**, pp.22-28.

[8] Bury, K., *Statistical Distributions in Engineering*, Cambridge, UK: Cambridge University Press, **(1999)**.

[9] Burman, R., and Pochop, L.O., *Evaporation, Evapotranspiration and Climatic Data*, ELSEVIER SCIENCE, SARA BURGERHARTSTRAAT 25, 1055 KV AMSTERDAM (NETHERLANDS), **(1994)**.

[10] Cao, M., Li, A., and Wei, J., *Weather Derivatives: A New Class of Financial Instruments*, York University, University of Toronto, and XL Weather & Energy Inc.,**(2004)**.

[11] Cao, M., Li, A., and Wei, J., *Precipitation Modeling and Contract Valuation: A Frontier in Weather Derivatives*,The Journal of Alternative Investments, Vol.7, **(2004)**, No.2, pp.93-99.

[12] Cao, M., and Wei, J., *Weather Derivatives Valuation and Market Price of Weather Risk*, The Journal of Futures Markets, Vlo.24, No.11, **(2004)**, pp.1065-1089.

[13] Emmerich, C.van, *Modelling and simulation of rain derivatives*, Master Thesis, University of Wuppertal, **(2005)**.

[14] Evans, M., N. Hastings, and B. Peacock, *Statistical Distributions*, 2nd ed., Hoboken, NJ: John Wiley & Sons, Inc., **(1993)**.

[15] Glasserman, P., *Monte Carlo Methods in Financial Engineering*, Stochastic Modelling and Applied Probability , Vol. 53, **(2003)**.

[16] Guttman, N. B., *Comparing the Palmer Drought Index and the Standardized Precipitation Index*, J. Amer. Water Resources Assn. Vol. 34, **(1997)**, pp. 113-121.

[17] Guttman, N. B., *Accepting the Standardized Precipitation Index: A Calculation Algorithm*, J. Amer. Water Resources Assn. Vol. 35, **(1999)**, pp. 311-322.

[18] Han, S., *Numerical Solution of Stochastical Differential Equations*, Master Thesis, University of Edinburgh and Heriot-Walt University, **(2005)**.

[19] Hanson, R. L., *Evapotranspiration and Droughts*, in Paulson, R.W., Chase, E.B., Roberts, R.S., and Moody, D.W., Compilers, National Water Summary 1988-89–Hydrologic Events and Floods and Droughts: U.S. Geological Survey Water-Supply Paper 2375, **(1991)**, pp.99-104.

[20] James, F., *Statistical Methods in Experimental Physics*, 2nd Edition, Amsterdam: North-Holland, **(1971)**.

[21] Jensen, M. E., Burman, R. D., and Allen, R. G., *Evapotranspiration and Irrigation Water Requirements*, American Society of Civil Engineers Manual and Reports on Engineering Practice, No.70, **(1990)**.

[22] Kahl, C., *Positive numerical integration of Stochastic Differential Equations*, Diploma Thesis, Wuppertal, **(2004)**.

[23] Keyantash, J. and Dracap, J. A., *The Quantification of Drought: An Evaluation of Drought Indices*, Bull. Amer. Meteor. Soc., 83, **(2002)**, pp.1167-1180.

[24] Kotz, S. and Nadarajah, S., *Extreme Value Distributions: Theory and Applications*, 1st Edition, World Scientific Publishing, **(2002)**.

[25] McGuinness, J.L., and Bordne, E.F., *A Comparison of Lysimeter-derived Potential Evapotranspiration with Computed Value* , ARS, USDA Tech. Bull. 1452, **(1972)**.

[26] Mihesan, V., *On a General Class of Beta Approximating Operators*, General Mathematics, Vol.16, No.4, **(2008)**, pp.115C126.

[27] Mraoua, M., and Bari, D., *Temperature Stochastic Modeling and Weather Derivative Pricing: Empirical Study with Moroccan Data*, Institute for Economic Research, OCP Group, Casablanca, **(2005)**.

[28] Muβhoff, O., Odening, M., and Xu, W., *Modeling and Hedging Rain Risk*, In Annual Meeting of the American Agricultural Economics Association (AAEA), Long Beach, USA, **(2006)**.

[29] Ntale, H. K., and Gan, T. Y., *Drought Indices and Their Application to East Africa*, International journal of climatology, vol.23, No.11, pp.1335-1357, **(2003)**.

[30] Redmond, K. T., *The Depiction of Drought: A Commentary*, Bull. Amer. Meteor. Soc., Vol.83, **(2002)**, pp.1143-1147.

[31] Richard R. Heim Jr., *A Review of Twentieth-Century Drought Indices Used in the United States* , Bull. Amer. Meteor. Soc., 83, **(2002)**, pp.1149-1165.

[32] Sachidananda, M., and D. Zrnić, *Rain Rate Estimates from Differential Polarization Measurements*, J. Atmos. Oceanic Technol., Vol.4, **(1987)**, pp.588C598.

[33] Stanpfli, J., and Goodman, V., *The Mathematics of Finance: Modeling and Hedging*, The Brooks/Cole Seires in Advanced Mathematics, **(2001)**.

[34] Stephens, M. A., *EDF Statistics for Goodness-of-Fit: Part I*, Technical Report, No.186, **(1972)**.

[35] Taraldsen, G., and Lindqvist, B. H., *The Multiple Roots Simulation Algorithm, the Inverse Gaussian Distribution, and the Sufficient Conditional Monte Carlo Method*, Statistics, No.4, **(2005)**, pp.2-8.

[36] Tsakiris, G., and Vangelis, H., *Establishing a Drought Index Incorporating Evapotranspiration*, Eur Water, 9-10, **(2005)**, pp.3-11.

[37] Tsakiris, G., Pangalou, D., and Vangelis, H., *Regional Drought Assessment Based on the Reconnaissance Drought Index*, Water Resources Management, Vol.21, No.5, **(2007)**, pp.821-833.

[38] United States Department of Agriculture, *Irrigation Water Requirements*, Technical Release No.21, **(1964)**.

[39] Xu, C. Y., and Singh, V. P., *Cross Comparison of Empirical Equations for Calculating Potential Evapotranspiration with Data from Switzerland*, Water Resources Management, Vol.16, **(2002)**, pp.197-219.

[40] Yoo, S, *Weather Derivatives and Seasonal Forecast*, Cornell University, Ithaca, New York, **(2003)**.

Appendix A

Java Code for Some Algorithms

Generate Van der Corput Sequence

```
public class VanderCorput {

    double Value(int n,int b)

    {

        int a = 0;

        int c = 0;

        do

        {

            a+=1;

            c = (int) (n/Math.pow(b,a));

        }

        while(c!=0);
```

```
int[] A = new int[a];

A[a-1] = (int)(n/Math.pow(b,a-1));


for (int i=a-2;i>=0;i--)

{

    double r = 0;

    for (int j=i+1;j<a;j++)

    {

        r+=A[j]*Math.pow(b,j);

    }

    double R = n-r;

    A[i] = (int)(R/Math.pow(b,i));

}


double x = 0;


for (int i=0;i<a;i++)

{

    x+=A[i]*Math.pow(b,-i-1);

}


return x;
```

```
    }
}

{\bf Random Number Generator Following Inverse Gaussian Distribution}

// Output Random Number List following
Wald Distribution with parameter mu, lambda

public class WaldRandom
{
    double[] List(double mu, double lamda, double[] X, double[] U)
    {
        double u,v,x,y,z;
        double fi = lamda/mu;
        int n = X.length;
        double[] L = new double[n];

        for (int i=0;i<n;i++)
        {
            z = Math.pow(X[i],2);
            y = 1-(Math.sqrt(Math.pow(z,2)+4*fi*z)-z)/(2*fi);
            v = (1+y)*U[i];
```

```
            if (v>1)
            {
                L[i] = 1/y;
            }
            else
            {
                L[i] = y;
            }
        }

        return L;
    }
}

// Convert Random Number List following Wald Distribution to one
following Inverse Gaussian Distribution with parameter mu, lambda, and
shift

public class IGRandom
{
    double[] List(double[] X, double mu, double lamda, double shift)
```

```
    {
        int n = X.length;
        double[] L = new double[n];

        for (int i=0;i<n;i++)
        {
            L[i] = mu*X[i]+shift;
        }

        return L;
    }
}

// Output m random number following Inverse Gaussian Distribution

import java.io.*;
import java.util.*;

public class InvGaussianList2
{
    public static double[] List(int m, double mu, double lamda, double shift)
    throws IOException
```

```
{
    IGRandom b = new IGRandom();
    Random r = new Random();
    VanderCorput d = new VanderCorput();
    WaldRandom e = new WaldRandom();
    MEAN g = new MEAN();
    PrintWriter out1 = new PrintWriter(new FileWriter("RandomInvGaussian.txt"));

    double[] L1 = new double[m+1];
    double[] L2 = new double[m+1];
    double[] L3 = new double[m+1];

    for (int i=1;i<m+1;i++)
    {
        L1[i] = r.nextGaussian();
        L3[i] = Math.random();
    }

    double[] W = e.List(mu,lamda,L1,L3);
    double[] L = b.List(W,mu,lamda,shift);
    double[] L4 = new double[m];
```

```
        for (int i=0;i<m;i++)
        {
            L4[i] = L[i];
            out1.println(L[i]);
        }
        out1.close();

        double x = g.mean(L4);
        double y = 0;

        for (int i=0;i<m;i++)
        {
            y+=Math.pow(L4[i],2);
        }

        double value2 = Math.sqrt((y-m*Math.pow(x,2))/(m));
        System.out.println("The mean value for the distribution is: "+x);
        System.out.println("The standard deviation for
        the distribution is: "+value2);

        return L4;
    }
```

```
}

{\bf Gauss-Newton Method to Estimate Mean Daily Temperature}

// M is the number of loops for Linear System;
X0 is the initial value for non-linear system;
X00 is the initial value for linear system;

public class GaussNewtonTem
{
    double[] List(int M, double Error, double[] X0, double[] X00, double[] S)
    {
        LinearSystem b = new LinearSystem();
        MatrixMultipler c = new MatrixMultipler();
        MatrixMultipler2 c2 = new MatrixMultipler2();
        MatrixMultipler3 c3 = new MatrixMultipler3();

        int m = 4;
        double[] X = X0;
        double[] D = S;

        double[] F = new double[365];
```

```
double[][] A = new double[365][m];


double[] B = new double[365];
double[] L = new double[m];


int i,j,k;
k = 0;


double x,y,z,s;
x = 0;
y = 0;


while(k<M)
{
    for (i=0;i<365;i++)
    {
        F[i] = X[0]+X[1]*(i+1)+X[2]*Math.sin(2*Math.PI*(i+1)/365)+
        X[3]*Math.cos(2*Math.PI*(i+1)/365)-D[i];
    }
    for (i=0;i<365;i++)
    {
        A[i][0] = 1;
```

```
        A[i][1] = i+1;


        A[i][2] = Math.sin(2*Math.PI*(i+1)/365);
        A[i][3] = Math.cos(2*Math.PI*(i+1)/365);
    }


    double[] C = c3.List(A,X,365,m);


    for (i=0;i<365;i++)
    {
        B[i] = C[i]-F[i];
    }


    double[][] A2 = c.List(A,365,m);
    double[] B2 = c2.List(A,B,365,m);
    L = b.List(A2,B2,X00,m,Error);


    for (i=1;i<m;i++)
    {
        z = Math.abs(L[i-1]-X[i-1]);
        y = Math.max(Math.abs(L[i]-X[i]),z);
    }
```

```
            for (i=0;i<m;i++)
            {
                X[i] = L[i];
            }

            k+=1;
        }

        if (y>Error)
        {
            System.out.println("The iteration exceeds");
        }

        return X;
    }
}

public class LinearSystem
{
    double[] List(double[][] A, double[] B, double[] X0, int n, double Error)
    {
        int i,j,M;
```

```
M = 0;

int N = 100000;

double[] R = new double[n];

double[] X = X0;

double[] Z = new double[n];

double a,x,y;

x = 0;

y = 0;


do

{

    y=0;

    for (j=0;j<n;j++)

    {

        x+=A[0][j]*X[j];

    }


    a = X[0];

    R[0] = B[0]-x;

    X[0] = X[0]+R[0]/A[0][0];

    Z[0] = Math.abs(X[0]-a);

    y = Math.max(Z[0],0);
```

```
x = 0;


for (i=1;i<n;i++)
{
    for (j=0;j<n;j++)
    {
        x+=A[i][j]*X[j];
    }


    R[i]=B[i]-x;
    a = X[i];
    X[i]=X[i]+R[i]/A[i][i];
    Z[i] = Math.abs(a-X[i]);
    y = Math.max(Z[i],Z[i-1]);
    x = 0;
}


M+=1;
if (M>N)
{
    System.out.println("The iteration exceeds");
    System.out.println("The error term right now is "+y);
```

```
                break;
            }
        }
        while (y>=Error);

        return X;
    }
}

public class MatrixMultipler
{
    double[][] List(double[][] A,int m,int n)
    {
        double[][] X = new double[n][n];
        int i,j,k;
        double x = 0;
        for (i=0;i<n;i++)
        {
            for (j=0;j<n;j++)
            {
                for (k=0;k<m;k++)
                {
```

```
                    x+=A[k][i]*A[k][j];
                }

                X[i][j] = x;
                x = 0;
            }
        }

        return X;
    }
}

public class MatrixMultipler2
{
    double[] List(double[][] A,double[] B,int m,int n)
    {
        double[] X = new double[n];
        int i,k;
        double x = 0;

        for (i=0;i<n;i++)
        {
```

```
            for (k=0;k<m;k++)
            {
                x+=A[k][i]*B[k];
            }
            X[i] = x;
            x = 0;
        }

        return X;
    }
}

public class MatrixMultipler3
{
    double[] List(double[][] A,double[] B,int m,int n)
    {
        double[] X = new double[m];
        int i,k;
        double x = 0;

        for (i=0;i<m;i++)
        {
```

```
            for (k=0;k<n;k++)
            {
                x+=A[i][k]*B[k];
            }
            X[i] = x;
            x = 0;
        }

        return X;
    }
}
```