COPULA BASED STOCHASTIC WEATHER GENERATOR AS AN APPLICATION FOR CROP GROWTH MODELS AND CROP INSURANCE

A Dissertation

by

MIRIAM JUAREZ TORRES

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	James W. Richardson
	Dmitry Vedenov
Committee Members,	Ximing Wu
	Michael Longnecker
Head of Department,	C. Parr Rosson III

December 2012

Major Subject: Agricultural Economics

ABSTRACT

Stochastic Weather Generators (SWG) try to reproduce the stochastic patterns of climatological variables characterized by high dimensionality, non-normal probability density functions and non-linear dependence relationships. However, conventional SWGs usually typify weather variables with unjustified probability distributions assuming linear dependence between variables. This research proposes an alternative SWG that introduces the advantages of the Copula modeling into the reproduction of stochastic weather patterns. The Copula based SWG introduces more flexibility allowing researcher to model non-linear dependence structures independently of the marginals involved, also it is able to model tail dependence, which results in a more accurate reproduction of extreme weather events.

Statistical tests on weather series simulated by the Copula based SWG show its capacity to replicate the statistical properties of the observed weather variables, along with a good performance in the reproduction of the extreme weather events.

In terms of its use in crop growth models for the ratemaking process of new insurance schemes with no available historical yield data, the Copula based SWG allows one to more accurately evaluate the risk. The use of the Copula based SWG for the simulation of yields results in higher crop insurance premiums from more frequent extreme weather events, while the use of the conventional SWG for the yield estimation could lead to an underestimation of risks.

DEDICATION

I thank my God, JEHOVA. The LORD has given me one of the greatest desires of my heart. He is my rock in the middle of the storm, my high tower and my deliverer; He trained my hands for war and my fingers for my everyday battles. The LORD is gracious and full of compassion; He conceded me one of the best gifts, rejoice in my labor.

I dedicate this dissertation to my family: Jesus Antonio Lopez Cabrera, my husband; Emilio and Lilian Helena, our children; Marbella Cabrera, Raul Juarez and Guadalupe Torres, our parents. This project became a reality because of your love, sacrifice, confidence, commitment, understanding and patience. No phrase is able to completely describe my gratitude and the love that my heart feels for you.

Moreover the profit of the land is for all;
Even the king is served from the field.
(Ecclesiastes 5:9)

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support of Dr. James W. Richardson and the wise direction of Dr. Dmitry Vedenov, thank you. I also would like to thank my committee members Drs. Ximing Wu and Longnecker for their guidance.

I thank people of the Department of Agricultural Economics at Texas A&M University for contributing to this enriching experience, in particular to my colleague Leonardo Sanchez for being a friend and an excellent workmate.

I also extend my gratitude to the Mexican government, Consejo Nacional de Ciencia y Tecnología (CONACYT) and the Comexus Fulbright-Garcia Robles for providing me the financial support for attaining this goal. I am also grateful for the support and guidance of the Drs. Antonio Yunez Naude and Juan Carlos Moreno Brid.

A special acknowledgement for Jesus Antonio Lopez Cabrera, my husband, who believed in me from beginning to end. His love, sacrifice and support have been some of the most amazing things of my life.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Stochastic Weather Generators (SWG) are numerical models that try to reproduce the statistical properties from observed historical climate series. Climatological variables are complex systems, characterized by high dimensionality, non-normal probability density functions and non-linear dependence relationships.

In the last decade copula methods and their applications have experienced a significant progress. In particular, the desire of reproducing more accurately stochastic patterns has conducted the application of copula procedures in the modeling of natural hazards.

In this research, copula approach is used to develop a SWG where the modeling of the joint distribution of weather variables satisfies two objectives. First, to model the non-linear dependence structures within weather variables. Second, to reproduce more accurately extreme weather patterns through the use of copula families like Gumbel and Clayton, or even through copula mixtures which introduce more flexibility. In addition, Copula based SWG considers Brownian motion process to emulate the daily time stochastic dynamics of the weather variables.

An evaluation on the performance of the Copula based SWG is carried out in terms of its use in crop growth simulation models and the modeling of insurance policies where no historical yield data is available.

For the sake of simplicity, this research is structured as follows:

- Chapter II discusses the foundations, functioning and characteristics of the SWGs. Also, this chapter describes in detail the motivation, the methodology and the structure of the Copula based SWG.
- Chapter III presents a statistical validation on the Copula based SWG against the observed data and against Richardson's (1981) SWG, one of the most broadly used SWG. Basic statistics, quantile analysis and non-parametric tests of goodness of fit are estimated for both simulated weather series.
- Chapter IV depicts the parametric calibration process for the Camelina crop in the Environmental/Policy Integrated Climate Model (WinEPIC). Next, yields are simulated using both SWGs and the resulting series are comparatively evaluated in terms of their distributions. The estimation of insurance premiums for both yield-series is carried out.
- Chapter V summarizes the most important findings, discuss some results and suggest some opportunity areas for future research.

CHAPTER II

THE MODELING OF WEATHER VARIABLES WITH COPULA APPROACH

Introduction

SWGs are a fundamental input for crop simulation models. These statistical models are able to produce long synthetic weather series, while they offer a solution for missing data by simulating key properties of the observed weather records. However, because climatological and meteorological phenomena are complex, characterized by non-normal probability density functions, such models have not attained a satisfactory quantification of uncertainty.

This study proposes an alternative SWG, based on copula methodology, to simulate the climate variables: precipitation, maximum temperature and minimum temperature, required by the WinEPIC to simulate crop yields. The main objective of this research is to apply the copula technique for the simulation of multivariate climatological variables capturing more accurately their nonlinear dependence structure and the occurrence of extreme events.

The Copula approach has several advantages; however, the most important for this research is its flexibility that allows researchers to model dependence structures between random variables independently of the marginal involved. Also, the copula technique offers different treatments on dependence structures for extreme events, common in weather variables. In fact, the initial hypothesis of this study is that multivariate probability distribution resulting from the copula approach might capture more accurately long-term changes in the hydrologic cycle and weather patterns of a specific region because it can model different patterns of dependency and extreme events.

The proposed SWG's specification is a hybrid copula, which incorporates different families of multivariate distributions using the conditional mixtures approach. This technique allows additional flexibility because it can interpolate from perfect conditional negative dependence to perfect conditional positive dependence, with conditional independence in between (Salvadori et al. 2007). In some cases, marginal distributions are specified under nonparametric specifications because parametric distributions are a poor description of the climatological process under these stochastic variables.

This research proposes the selection of particular observations for solving the dimensionality problem. The selected dates consider the observations with the highest average anomalies per month. Copula estimation is applied for the selected observations and simulated, while the daily dynamic of weather series is emulated by a random walk described by a geometric Brownian motion.

Model estimation is carried out by Maximum Likelihood (ML) methods. Also, ML is used to determine the best specification for copula family because multivariate goodness of fit tests (GOF) do not lead to an unique, or even conclusive criterion (Genest and Favre 2007a). Data from three weather stations located in Montana, Washington and Texas are used for this research.

Background and Motivation

Climatological and meteorological phenomena are complex and characterized by non-normal probability density functions. Temperature belongs to bounded and skewed distributions usually parameterized by the gamma distribution. Precipitation is a central issue in the climate modeling because it is ruled by nonlinear physical processes which are highly variable in space and time (Schölzel and Friederich 2008). However, it is known that traditional models have not attained an adequate quantification of uncertainty for climate (Wilks and Wilby 1999).

SWGs are numerical models that try to reproduce the statistical properties from observed historical climate series, mainly maximum temperature, minimum temperature and precipitation. In theory, these models are able to generate long synthetic weather series that preserve the statistical properties of the original data. SWGs satisfy the random number generator conditions: efficiency (fast and small use of memory), repeatability (exactly reproduction of sequences) and portability (L'ecuyer 2004). With these characteristics, SWGs are able to reproduce sequences of weather data whose behavior is very close to the "truly random" observed weather patterns in a broad variety of climates and regions. SWGs are not weather forecasting algorithms, which are deterministic weather models that typically operate by numerically integrating partial differential equations. SWGs behave statistically as weather data, which no weather sequence can be duplicated at a given time in either the past or future (Wilks and Wilby 1999).

SWG parameters comprise a concise summary of climate behavior and use Monte Carlo methods as a random number generator for simulation whose output statistically resembles daily weather data at a location, where any particular simulated weather sequence will be duplicated (Wilks and Wilby 1999).[1]

Typically, SWGs award a determinant role to the precipitation process. Precipitation is a complex stochastic process whose occurrence gives rise to numerous physical processes in nature affecting the statistics of many non-precipitation variables (Wilks and Wilby 1999). Precipitation has a high zeros rate which introduce a discontinuity in its probability distribution between zero and nonzero observations. In terms of temporal correlations, precipitation owns a mixed character of discrete and continuous variable.

Because of these features, precipitation is modeled in two treatments, occurrence and intensity processes. Some authors combine the use of first-order Markov models to characterize occurrence and pseudo-random number generators to describe intensity by the generation of independent and identically distributed (i.i.d.) draws.

SWGs have numerous applications; their parameters can be interpolated to generate synthetic daily data for unobserved locations. Also, they are frequently used in climate change studies for impact evaluation by using modifications in climatic means and variabilities, predicted by the Global Climate Models (GCM) (Semenov et al. 1998). In particular, SWGs have been widely used as input in crop simulation models because

[1]In computational statistics, random variate generation includes two steps. First, the generation of i.i.d. random variables with Uniform(0, 1); second, the application of transformation on these random i.i.d. U(0, 1) to imitate random variates and random vectors from arbitrary distributions.

of their ability to generate missing data with only monthly or seasonal statistics. Additionally, they are able to produce long enough series to allow good estimates of the probability of extreme events that affect crop yield.

This research will focus on the construction of a SWG, based on copula methodology, as a component in WinEPIC, originally developed in the early 1980's, to assess the effect of erosion on productivity (Williams et al. 1984). WinEPIC is a comprehensive crop simulation model to analyze cultural practices and cropping systems on production, soil quality, water quality, water and wind erosion, and profits.[2] WinEPIC's components include weather simulation, hydrology, erosion, sedimentation, nutrient cycling, pesticide fate, crop growth, soil temperature, tillage, economics, and plant environment control (Williams et al. 1984). Crop growth is one of the most important simulated processes because soil productivity is expressed in terms of crop yield. Thus, the model is sensitive to crop characteristics (weather, soil and fertility) and to other properties. Potential plant growth is simulated on daily basis and constrained by three stress factors: soil, strength, temperature and aeration (Williams et al. 1989).

WinEPIC includes the SWG developed for Richardson (1981) for simulating precipitation, radiation, maximum and minimum temperature. The model configuration, awards a primary role to precipitation, preserving the dependence in time, the correlation between variables, and the seasonal characteristics in actual weather data for the location. Precipitation is characterized in two stages. First, a Markov chain exponential model – with two states, wet and dry– describes the occurrence process where the

[2] WinEPIC considers homogeneous weather, soils, and management systems in field-size area up to 100 ha and operates on a daily basis.

probability of rain is conditioned on the wet or dry status of the previous day, which exhibit persistence or positive serial autocorrelation.[3] So, wet and dry runs tend to clump together in volume more strongly than could be expected by chance. Second, daily precipitation amount, given a wet day, is supposed to be independently determined by an exponential distribution (Richardson 1981). The inputs for the model are monthly means and coefficients of variation for each variable.

Richardson's SWG considers that for each variable the dependence structure (serial correlation) is characterized by a first order linear autoregressive model. Although this model operates on a daily time step, their simulations do not show longer-term variations (Wilks and Wilby 1999). The random values for the current SWG show a lower monthly mean temperature and monthly accumulated precipitation with respect to the observed weather data. The SWG currently cannot capture the variability year to year, as their statistics vary only through a fixed annual cycle.

Authors such as Semenov et al. (1998) and Wilks (1990) have pointed out that sensivity analysis in crop simulation models have shown that stochastic simulations for weather variables based on mean temperatures values produce overestimated crop yields.

Semenov et al. (1998) showed that changes in climate variability and extreme weather events can have a major effect on crop growth simulation and the associated agricultural risk. In particular, because the occurrence of extreme weather events are better correlated with changes in the variability of climate variables than with changes in

[3] The behavior of the Markov chain is ruled by the transition probabilities that specify the conditional probabilities for the system to be in each of its possible states during the next time period. In a first order Markov chain, the transition probabilities controlling for the next stage of the system depend only on the current state of the system (Wilks, 2011).

the mean values. As a consequence, a SWG modeled by copulas with more accurate properties could have multiple implications in WinEPIC use because copulas are able to model tail dependence, which relates to dependencies of extreme events. Copula estimation of anomalies on climatic variables could produce more accurate yields simulation because crop simulation models incorporate a mixture of non-linear responses of the crop to its environment (Semenov et al. 1998). Thus, more precise simulations could result in the estimation of more accurate yields generated by the crop simulation model because crop simulation models incorporate a mixture of nonlinear responses of the crop to its environment components.

A Copula Based Stochastic Weather Generator

Traditional modeling of climate variables relies on a multivariate distribution, which is usually characterized jointly under the same parametric family and their pattern of dependence is assumed to be linear. Any kind of high dimensional multivariate distribution is either limited in covariance structure or comes with a high number of parameters (Schölzel and Friederichs 2008).

According to Genest and Favre (2007a), the traditional multivariate approach has disadvantages because it dismisses additional information from their individual behavior. The implicit rigidity in the dependence pattern might avoid the incorporation of variability and, also it would prevent capturing the long term changes in the climatological process.

In contrast, the copula approach might incorporate additional information into the climate simulation providing important insights to give a more accurate representation of the underlying processes of climatological variables. Besides, this methodology allows a multivariate dimension analysis for climatological variables, while the modeling of dependence structures between random variables can be far from linearity and independent of the marginal laws involved. In addition, the copula technique offers different treatments on dependence structure for extreme events, common in climatic variables. These properties jointly might help to detect long-term changes in the hydrologic cycle of a specific location (Genest et al. 2007b).

In recent years, the applications of copulas in simulation of multivariate data, extreme value analysis and modeling dependence structure has increased in climatological phenomena analysis. Favre et al. (2004) used copulas to analyze the multivariate hydrological frequency. Schölzel and Friederichs (2008) studied the appropriateness of the application of copulas to meteorological and climatological phenomena summarizing the problem of goodness of fit for copulas and analyzing the connection with multivariate extremes. More recently, in the planning and management of water resources Wong et al. (2010) modeled droughts using copulas to simulate duration, peak, intensity and average intensity.

However, no application in climate analysis that involves copula methods in the design of a SWG has been developed, and this is precisely the objective of this research: the use of copula methodology in the development of a SWG for precipitation, maximum temperature and minimum temperature.

Basically the idea of modeling climatic variables using copula methods relies on the dynamics of these variables. Every year weather observations follow a determined cycle: high temperature realizations in summers and low temperature realizations during winters. Although weather realizations are stochastic, their differences between one day and the next one are not far. For example, usually the temperature on June 1st is at most two or three degrees different from June 2nd, or even on June 5th. There exist evident dependence patterns which might not be adequately modeled assuming linear dependence, for example: maximum temperature realizations have a strong connection with minimum temperature realizations, lower realizations in minimum temperature are associated with rainfall occurrence.

Furthermore, the copula approach provides additional flexibility because the true probability distribution of the weather variables is unknown. Copula methods allow researcher to fit individually the probability distribution for every variable and then to model the dependence relationship between the variables with different copula family specifications.

However, the copula estimation in daily basis implies a dimensionality problem. In a high-dimensional distribution model, where the whole surface is estimated using a set of observations, the more points are considered in the estimation, more accurate will be the surface estimations; but the cost for on higher dimensionality is the low reliability of the estimation. The solution that this research proposes is the selection of specific dates to perform the copula estimation. For such purpose 12 dates per year were selected

– one per month – where the dates are determined by the maximum monthly average historical anomalies.[4]

The parameters of the probability distribution for the marginals are individually estimated, while the trivariate copula parameters are jointly estimated for the selected dates. The simulation of the weather variables implies the description of their joint probability distribution at the selected time which is the bordering conditions of the climate stochastic dynamic simulation, while the Brownian Bridge reflects the intertemporal dynamic of weather variables evolving on a path forward through time.

Basically the SWG structure would impose anything but a joint dependence structure and the Brownian motion process between the simulated structures to emulate their daily time stochastic dynamics.[5]

The strong connection between random weather variables and their daily sequence validate the assumption of the Brownian Bridge stochastic process to interpolate the copula realizations.[6] Turvey (2005) used a similar idea to daily pricing of weather insurance for ice wine in the Niagara Peninsula of southern Ontario. The Brownian Bridge results from conditioning of the Brownian motion on its endpoints and its behavior depends on its parametric space. The potential advantage of Brownian

---

[4] Anomalies measure deviations over a certain period of time (month, season or year) from the long-term climate statistics, relating to their calendar period.

[5] The Brownian motion is the most known form of the Wiener-Levy process, which has been adopted as the probabilistic model for numerous natural phenomena. Brownian motion describes the random movement of particles in multidimensional space, such as the stochastic process in weather and hydrology.

[6] The Brownian Bridge conserves the Brownian motion properties: Gaussian, centered, with independent increments and diffusion Markovian or Martingale property, that is:

$$E[W(t_n)|W(t_{n-1}), W(t_{n-2}), \dots, W(t_1)] = W(t_{n-1}), \quad t_1 < t_2 < \cdots < t_n$$

Bridge relays in its use with variance reduction techniques and low discrepancy methods (Glasserman 2010).

Dependence Patterns of Weather Variables

Although weather generation was performed for three locations – one in Montana, one in Texas and one in Washington – figure 1, in appendix A, just describe climate patterns on one weather station, Conrad in Pondera County, Montana.

Figure 1, in appendix A, depicts the relations between the multiple pairs of intersections for weather variables and provide some insight about the form of the dependence between weather variables. The scalar plots show asymmetric distributions, the key is the difference in the ends of the distributions which describe tail dependence and, in turn, extreme weather events. Maximum temperature corresponds to lower bound precipitation while minimum temperature corresponds to upper bound precipitation. On the other hand, maximum temperature and minimum temperature show a positive dependence in the right corner suggesting a positive concordance between these two variables.[7]

Figure 2, in appendix A, shows positive concordance between minimum temperature and precipitation with more weight in just one tail of these multivariate

---

[7] Concordance refers to the probability of having large (or small) values of both variables X and Y is high, while the probability of having "large" values of X together with small values of Y (or vice versa) is low (Cherubini et al. 2004).

distributions. Asymmetric copulas – such as Clayton, Gumbel or Frank – are able to model multidimensional movement only in one corner.

Methods and General Theory About Copulas

Copulas are joint cumulative distribution functions that describe dependencies among variables independent of their marginals (Joe 1997). In other words, the copula is a multivariate distribution with all univariate margins being U(0,1) that represent dependencies between variables (Cherubini et al. 2004).

According to Nelsen (2006) copulas satisfy mainly four conditions:

i)$\forall$ u Є [0,1], C(1,…,1,u,1,…,1) = u;

ii)$\forall$ u_i Є [0,1], $C(u_1,\ldots,u_n) = 0$ if at least one of the u_i equals zero;

iii)C^m is grounded and m-increasing.[8]

In terms of an m-dimensional distribution F with marginal cumulative distribution functions $(F_1,\ldots,F_m)$, and a j^{th} univariate margin F_j, the copula associated with F is a distribution function $C:[0,1]^m \rightarrow [0,1]$ that satisfies

$$F(x) = C\ (F_1(X_1), \ldots, F_m(X_m)), \qquad x \in R^m \tag{1}$$

Besides, if F is a continuous m-variate distribution function with univariate margins $F_1,\ldots,F_m$, and quantile functions $F_1^{-1},\ldots,F_m^{-1}$, then

$$C(u) = F(F_1^{-1}(u_1), \ldots, F_m^{-1}(u_m)) \tag{2}$$

This is because if X~F and F is continuous then $(F_1(x_1),\ldots,F_m(x_m))$~C and if U ~ C, then $(F_1^{-1}(U_1),\ldots,F_m^{-1}(U_m))$ ~ F. Copula is the distribution of a random vector,

[8] Let A_1 and A_2 non-empty subsets of R and a function G: $A_1xA_2 \rightarrow R$ and denote a_i the least element of A_i, i=1,2. The function G is grounded if for every (v,z) of A_1xA_2, $G(a_1,z) = 0 = G(v,a_2)$, (Cherubini et al. 2004).

$U = (U_1, \ldots, U_m)$, where each $U_j \sim U(0,1)$, C is a continuous function and increasing, which guarantees that right derivatives exist.

In terms of multivariate weather data simulation, copula representation is more than convenient because of their probabilistic interpretation. The Sklar theorem states that if all $F_1, \ldots, F_m$ are continuous, then copula C^m is uniquely determined on the range of $F_1, \ldots, F_m$. As a consequence, the joint probability density of multivariate distributions can be presented as the product of the marginal probability densities and the copula density, which is the canonical representation (Cherubini et al. 2004).

$$f_x(\boldsymbol{x}) = \quad c_x\big(F_1(x_1), F_2(x_2), \ldots, F_m(x_m)\big) \cdot \prod_{j=1}^{m} f_j\big(x_j\big) \tag{3}$$

where

$$c_x\big(F_1(x_1), F_2(x_2), \ldots, F_m(x_m)\big) = \frac{\delta(C(F_1(x_1), F_2(x_2), \ldots, F_m(x_m)))}{\delta F_1(x_1), \delta F_2(x_2), \ldots, \delta F_m(x_m)} \tag{4}$$

Two important implications are derived from Sklar theorem. First, the independent representation from marginals of the copula defines the dependence structure in the multivariate structure (Nelsen 2006). This separation between marginal distributions and dependence creates the flexibility to use marginals from different types of distributions that describe better the multivariate phenomena. The second implication is the possibility for simulating random variables with the same probability distributions as original data and preserving the dependence structure through the copula.

Copula Families

Each copula family or class is represented by its density and conditional distribution function and the parameter or a vector of parameters. Families characterize dependence functional forms related to properties that include reflection symmetry, extreme value copula, multivariate extendibility, as well as dependence properties (Joe 1997).

Some families such as Gaussian and t-student (copulas of elliptical distributions) are frequently used in all areas of study because of their advantage in extension to arbitrary dimensions. However, they are restricted to radial symmetry and they do not have a close form expression, which could imply a high cost in high-dimensional estimation. An exhaustive list of copula families can be found in Nelsen (2006).

Archimedean copula is another class, particularly used in the modeling of climate and hydrological phenomena.[9] These copulas specifications are easy to construct and they allow a broader variety in dependence structures, such as tail dependence (Nelsen 2006). Archimedian copulas can be constructed by a generator function φ: I→R^+, which defines a subcall or family of Archimedean copulas. Generator function (φ) is continuous, decreasing and convex with φ(1)=0 (Cherubini et al. 2004). Given a generator and its pseudo-inverse, the next equation states the generation of an Archimedean copula C^A.[10]

$$C^A(x_1, \dots, x_m) = \varphi^{-1}(\varphi(x_1) + \cdots + \varphi(x_m)) \quad (5)$$

The generator function must be strict (strictly monotonic, continuous and strictly increasing) to allow multivariate extension of the copula. Archimedean copula properties

[9] Salvadori et al. 2007; Schölzel and Friederichs 2008.

[10] The pseudo-inverse function is ϕ^{-1}: [0,inf] →I, continuous and non-increasing on [0,inf] and strictly decreasing on [0, ϕ(0)] and by composition with the generator gives the identity, $\phi^{-1}(\phi(x))=x$.

$$\phi^{-1}(x) = \begin{cases} \phi^{-1}(x) & 0 \le x \le \phi(0) \\ 0 & \phi(0) \le x \le +\infty \end{cases}$$

are: symmetry, associativity and easy identification of their level curves by the following condition:[11]

$$\{(x_1, \dots, x_m) \in I^m : C(x_1, \dots, x_m) = K\} \tag{6}$$

The most frequent source of generators for m-dimensional Archimedean copulas are the Laplace inverse transformations for distribution functions, which existence is guaranteed only when function φ is completely monotonic (Cherubini et al. 2004). Copulas describe naturally the dependence between multivariate extremes. The tail dependence concept in a bivariate distribution rates the amount of dependence in the upper-quadrant tail or lower-quadrant (Salvadori et al. 2007).

The three copulas used in this research are Gumbel, Clayton and Frank.[12] Gumbel m-copula belongs to the Gumbel-Hougaard family, which is the only extreme value of the Archimedean family. Gumbel is completely monotonic, has upper tail dependence, extreme value copula and partial multivariate extension.[13]

$$Generator \quad \varphi(u) = (-\ln(u))^{\alpha} \tag{7}$$

$$\varphi^{-1}(t) = \exp\left(-t^{\frac{1}{\alpha}}\right) \tag{8}$$

$$C(u_1, u_2, \dots, u_m) = \exp\left\{-\left[\sum_{i=1}^{m}(-\ln u_i)^{\alpha}\right]^{\frac{1}{\alpha}}\right\} \quad for \quad \alpha > 1 \tag{9}$$

Clayton m-copula, belongs to Clayton family, is completely monotonic and owns lower and upper tail dependence.

---

[11] Generator is strict if it satisfies ϕ(0)=+inf, (Cherubini et al. 2004)

[12] An exhaustive list of Archimedean copulas or other classes, see Nelsen (2006); Salvadori et al. (2007).

[13] Extreme value distributions and their three types (Gumbel, Frétchet and the Weibull) provide the only non-degenerated limit laws for adequate transformed maxima of identical and independently distributed random variables. For a detailed reference in this issue, consult Embrechts et al. (2001).

$$Generator \quad \varphi(u) = \mathrm{u}^{-\alpha} - 1 \tag{10}$$

$$\varphi^{-1}(t) = (t+1)^{-\frac{1}{\alpha}} \tag{11}$$

$$C(u_1, u_2, \ldots, u_m) = \left[\sum_{i=1}^{m} (\mathrm{u}_\mathrm{i}^{-\alpha} - n + 1)\right]^{-\frac{1}{\alpha}} \quad for \;\; \alpha > 0 \tag{12}$$

Frank m-copula, belongs to Frank family, is completely monotonic, it owns reflection symmetry, partial multivariate extension and extension to negative dependence.

$$Generator \quad \varphi(u) = ln\left(\frac{\exp(-\alpha u) - 1}{\exp(-\alpha) - 1}\right) \tag{13}$$

$$\varphi^{-1}(t) = -\frac{1}{\alpha} \ln(1 + e^{t}(e^{-\alpha} - 1)) \tag{14}$$

$$C(u_1, u_2, \ldots, u_m) = -\frac{1}{\alpha} \ln\left\{1 + \frac{\prod_{i=1}^{m}(e^{-\alpha u_i} - 1)}{(e^{-\alpha} - 1)^{m-1}}\right\} \quad for \;\; 0 \le \alpha \le \infty \; when \; n \ge 3 \tag{15}$$

Mixtures of Conditional Distributions

The conditional mixture method allows extending bivariate copulas to an arbitrary dimension, at the same time that introduces additional flexibility in the model. By this technique, the construction of an m-multivariate family starts from two dimensional marginals. Salvadori et al. (2007) mentions that "these families of multivariate distributions can be made to interpolate from perfect conditional positive dependence with conditional independence in between."

Copula family can depict a range of dependence structure. Furthermore, copula mixture is able to model different dependence patterns in multivariate distributions. Thus, this method allows one to model the dependence pattern by pairs of variables capturing the best dependence structure in each pair of variables using the conditional sampling method. For example, the trivariate Frank copula is symmetric and cannot admit tail dependence, while the mixture Frank-Clayton-Gumbel admits asymptotic tail dependence and asymmetry.

The conditional approach, used here, is a unifying method for constructing multivariate distributions with a given family copula for each bivariate margin. However, the conditional mixture effectively enhances flexibility when there is a gain constructed in a common base measure for all component mixtures: the likelihood function. This model is especially effective with large sample sizes (Salvadori et al. 2007).

M-variate distributions can be constructed based on (m-1) dimensional margins, which must have m-2 variables in common. If one is given, F_{12}, F_{23},…,$F_{m-1,m}$, $m \geq 3$, it is possible to build a m-variate distribution starting with the trivariate distribution $F_{i,i+1,i+2} \in \mathcal{F}(F_{i,i+1}, F_{i+1,i+2})$, next the four-variate distributions from $F,\ldots,F_{i+3} \in F(F_{i,i+1,i+2}, F_{i+1,i+2,i+3})$, and so on. There exist a bivariate copula C_{ij} associated with the (i, j) bivariate margin of the m-variate distribution. For (i, j) with $|j - i| > 1$, C_{ij} measures the amount of conditional dependence in the i^{th} and j^{th} variables, given those variables with indices in between (Joe 1997). Following Joe (1997), the next equation shows the construction of a trivariate copula family.

$$F_{123}(\boldsymbol{x}) = \int_{-\infty}^{x_2} C_{13}(F_{1|2}(x_1 | x_2) F_{3|2}(x_3 | x_2) F_2(dx_2), \tag{16}$$

The arguments of the integrand are conditional cumulative distribution functions ($F_{1|2}$ and $F_{3|2}$) obtained from F_{12} and F_{23}. They can be written in terms of copulas because by construction, equation (16) is a trivariate distribution with univariate margins F_1, F_2, F_3 and bivariate margins F_{12} and F_{23}. C_{13} can be interpreted as a copula representing the amount of conditional dependence between the first and third univariate margins given the behavior of the second (Joe 1997). This method can be extended recursively to an m-dimensional copula.

$$F_{1,..,m}(\boldsymbol{x}) = \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_{m-1}} C_{1m}(F_{1|2\ldots m-1}(x_1 \mid x_2, \ldots, x_{m-1}). F_{2|2\ldots m-1}(x_m \mid x_2, \ldots, x_{m-1}). F_{2\ldots m-1}(dx_2, \ldots, dx_{m-1}) \tag{17}$$

Where $F_{1|2\ldots m-1}$ and $F_{m|2\ldots m-1}$ are conditional distributions functions obtained from $F_{1\ldots m-1}$ and $F_{2\ldots m-1}$ (Joe 1997). Copulas can be derived directly by using the integral representation in equation (17) and Sklar's theorem.

This research applies the conditional mixture method for the estimation of a trivariate copula. This approach includes not only diverse marginal distributions, but also different copula family specifications. The trivariate copula is expressed in equation (18).

$$C_{123}(u_1, u_2, u_3) = \int_0^{u_2} C_{13}\left(\frac{\partial C_{12}(u_1, x)}{\partial u_2}, \frac{\partial C_{23}(x, u_3)}{\partial u_2}\right) dx \tag{18}$$

In general terms, the estimation and simulation of copulas is possible by the calculation of partial derivatives, as the following equations show.

$$c_{123}(u_1, u_2, u_3) = \frac{\partial^3 C(u_1, u_2, u_3)}{\partial u_1 \partial u_2 \partial u_3}$$

$$= \frac{\partial^2 C_{13}\left(\frac{\partial C_{12}(u_1, x)}{\partial u_2}, \frac{\partial C_{23}(x, u_3)}{\partial u_2}\right)}{\partial u_1 \partial u_3} \times \frac{\partial^2 C_{12}(u_1, u_2)}{\partial u_1 \partial u_2} \times \frac{\partial^2 C_{23}(u_2, u_3)}{\partial u_2 \partial u_3} \quad (19)$$

$$= c_{13}\left(\frac{\partial C_{12}(u_1, u_2)}{\partial u_2}, \frac{\partial C_{23}(u_2, u_3)}{\partial u_2}\right) \times c_{12}(u_1, u_2) \times c_{23}(u_2, u_3)$$

with

$$\frac{\partial C_{ij}\left(u_i, u_j, \theta_{ij}\right)}{\partial u_i} \quad (20)$$

$$c_{ij}\left(u_i, u_j\right) = \frac{\partial^2 C_{ij}\left(u_i, u_j, \theta_{ij}\right)}{\partial u_i \partial u_j} \quad (21)$$

Thus, different specification families can be used to give more flexibility to the specification. Three different parameters substituting equation (20), and (21) into (19) result in a three-variable-three parameter copula density $c_{123}\left(u_1, u_2, u_3;\ \theta_1, \theta_{2,}\theta_{3,}\right)$. This expression can be used to estimate the parameter values by Maximum Likelihood. As Cherubini et al. (2004) mentions, this approach is elegant but the calculation of the inverse function analytically could be challenging.

Archimedean copulas provide advantages in estimation because they can be rewritten in their Laplace transformation representation and then estimated more easily. If $C(u_1, u_2, \dots, u_m) = \varphi^{-1}[\varphi(u_1) + \varphi(u_2) + \dots + \varphi(u_m)]$ is an Archimedian m-variate copula with generator $\varphi(.)$. For a k=2,…,m,

$$C_k(u_k \,|\, u_1, \dots, u_{k-1}) = \frac{\varphi^{-1(k-1)}[\varphi(u_1) + \varphi(u_2) + \dots + \varphi(u_k)]}{\varphi^{-1(k-1)}[\varphi(u_1) + \varphi(u_2) + \dots + \varphi(u_{k-1})]} \tag{22}$$

$$\frac{\partial^{k-1} C_{k-1}(u_1, \dots u_k)}{\partial u_{1,} \dots, \partial u_{k-1}} = \varphi^{-1(k-1)}[\varphi(u_1) + \varphi(u_2) + \dots + \varphi(u_{k-1})].\prod_{j=1}^{k-1} \varphi^{(1)}(u_j) \tag{23}$$

Thus, the parameter estimation in the case of the copulas would result more straightforward.

Copula Based Simulation

Simulation of multivariate vectors with given m-distributions can be carried out by calculating partial derivatives of the copulas because the integral operator is removed from equation 18 when the copulas of interest are differentiated. Eventually only composite functions of partial derivatives for bivariate copulas are evaluated (Salvadori et al. 2007).

The algorithm to simulate multivariate copula distribution has a nested structure. Salvadori et al. (2007) provides a straightforward method based on Sklar Theorem. Assume that **F** is a multivariate distribution with continuous marginals $F_1,\dots,F_m$ that can be represented by a m-copula, $\mathbf{C}^m$. Then, the generation of a vector $(X_1,\dots,X_m) \sim F$ can be done by simulating a vector $(U_1,\dots,U_m) \sim \mathbf{C}$, where the random variables U_i's are Uniform [0,1].

Because copulas are invariant to transformations, the simulated random vector **X** has the same dependence structure as vector U. The following equation shows the joint application of the Sklar theorem and the Probability Integral Transform.

$$U_i = F_i(X_i) \Leftrightarrow X_i = F_i^{[-1]}(U_i) \tag{24}$$

Where i=1,…,m, the X_i's random variables have marginal distributions of F_i's, (which do not necessarily belong to the same distribution) and a joint distribution F. The whole simulation process for k variables is described by the following steps.

First, let u_1 to be the random realization of the random variable U_1' uniform on [0,1]. The simulated variable is u_1. For the sake of the simulation of X, set

$$X_1 = F_{x1}^{-1}(U_i) \tag{26}$$

The next step is to simulate u_1 and u_2 based on the joint distribution function F. For this purpose, u_2 sampled from U_2, must be conditioned on the event $\{U_1=u_1\}$

$$u_2 = F_2^{-1}\left(u_2 \middle| u_1\right) = P(U_2 \le u_2 \mid U_1 = u_1) \tag{27}$$

Where the conditional functions can be expressed as

$$\frac{\delta_{u_1} C(u_1, u_2\ )}{\delta_{u_1} C(u_1)} = \delta_{u_1} C(u_1, u_2\ ) \tag{28}$$

Where u_2' is the realization of a random variable U_2' uniform on [0,1] and independent of U_1'

Thus, successively for example to simulate u_k, sampled from U_k and consistent with the joint distribution function F or previously sampled $u_1,\dots,u_{k-1}$; U_k must be conditioned to the events $\{U_1=u_1, U_2=u_2,\dots,U_{k-1}=u_{k-1}\}$

$$u_k = F_k^{-1}\left(u_k \middle| u_1, \dots, u_{k-1}\right) = \boldsymbol{P}\{U_k \le u_k \mid U_1 = u_1, \dots, U_{k-1} = u_{k-1}\} \tag{29}$$

$$F_k^{-1}\left(u_k \middle| u_1, \dots, u_{k-1}\right) = \frac{\delta_{u_1,\dots,u_{k-1}} C(u_1, \dots, u_k)}{\delta_{u_1,\dots,u_{k-1}} C(u_1, \dots, u_{k-1})} \tag{30}$$

Where u_k' is the realization of the random variable U_k' uniform on [0,1], and independent of U_1',…, U_{k-1}. Finally, by the probability integral transform it is possible to generate the sample (x_1,…,x_m).

$$(x_1, \dots, x_m) = (F_1^{-1}(u_1), \dots, F_m^{-1}(u_m)) \tag{31}$$

Estimation Methods

The canonical representation for the multivariate density function in equation (23), allows decomposing the statistical modeling of copulas in two steps: first the identification and modeling of the marginal distributions; and second, the estimation of the suitable copula function. This procedure can be generalized to mainly three methods: the Exact Maximum Likelihood (EML) method, the inference for the marginal (IFM) method and the canonical maximum likelihood (CML) method.

EML method assumes a parametric family of the copula and parametric marginal distribution, and it simultaneously estimates both sets of parameters (Cherubini et al. 2004). However, its efficiency depends on the numerical complexity of the optimization problem which increases with the dimensionality of the random vectors (Schölzel and Friederichs 2008).

IFM is a two stage estimation process based on maximum likelihood. The first stage consists of the estimation for the univariate marginal distribution parameters and the second stage estimates the copula parameters. This procedure is consistent and asymptotically normal under regularity conditions (Cherubini et al. 2004).

Finally, CML is also a two-step method based on maximum likelihood. The first step involves the estimation of the marginals using the empirical distributions and second, the estimation of the copula density by using maximum likelihood estimation method. This method produces consistent estimates of the copula parameters and their standard errors.

This research carries out the estimation in two steps. First, the estimation of the univariate marginal distributions parameters, assuming parametric specifications was carried out. Second, the kernel smoothing technique was applied to compare, or even, to attain a better fit of the distributions, with emphasis on precipitation.

Dependence Measures

Pearson correlation and linear dependence concepts do not capture the complete dependence dimensions of non-normal distributions. Linear correlation is not preserved when nonlinear transformations are applied to random variables. Multivariate models, such as copulas, require dependence measures that can be able to capture and identify their dependence properties.

Kendall's tau (τ) is used for compatibility conditions. Thus, Kendall's tau and Spearman's rho(ρ) are multidimensional measures of monotone dependence for continuous variables that are invariant respect to strictly increasing transformations, which is mainly the characteristic of copulas and non-normal distributions. Other important property of τ and ρ is that are increasing with respect to the concordance ordering (Joe 1997).[14]

---

[14]Concordance means the degree to which large or small values of one random variable associate with large or small values of another and as a rank correlation. Concordance measure satisfies properties: completeness; normalized measure, symmetry, continuity and concordance zero when variables are independent. These properties imply invariance respect to increasing transformations (Cherubini et al. 2004).

Other measures such as tail dependence, positive quadrant dependence and the concordance ordering are also basic in the analysis of multivariate extreme value distributions and copulas.

F be a continuous multivariate cumulative distribution function (c.d.f.) and let $(X_1,X_2,\dots,X_n)$ and $(X'_1,X'_2,\dots,X'_n)$ be independent random vectors with distribution F (Joe 1997). Kendall's tau is[15]

$$\tau = Pr[(X_1 - X'_1)(X_2 - X'_2) \dots (X_n - X'_n) > 0] \\ - Pr[(X_1 - X'_1)(X_2 - X'_2) \dots (X_n - X'_n) < 0] \quad (32)$$

$$\tau = 2\Pr(Pr[(X_1 - X'_1)(X_2 - X'_2) \dots (X_n - X'_n) > 0] - 1 = 4 \int F\, dF - 1 \quad (33)$$

F be a continuous multivariate c.d.f. with univariate margins $F_1,F_2,\dots,F_n$ and let $(X_1,X_2,\dots,X_n)\sim F$; then the Spearman's rho is the correlation of $F_1(X_1)$, $F_2(X_2),\dots,F_n(X_n)$. As F_1 and F_2 are U(0,1) random variables under the continuity assumption, their expectations are ½, their variances are ½ and Spearman's rho is in the following equation.

$$\rho = 12 \iint F_1(X_1)\, F_2(X_2) \dots F_n(X_n) dF(X_1, X_2, \dots, X_n) - 3 \quad (34)$$

$$= 12 \iint \overline{F}\, dF_1 dF_2 \dots dF_n - 3$$

[15] The condition $(X_1 - X'_1)(X_2 - X'_2) \dots (X_n - X'_n) > 0$ denotes $(X_1, X_2, \dots X_n), \left(X'_1, X'_2, \dots, X'_n\right)$ are two concordant vectors where one of the vectors has the larger value for both components. The condition $(X_1, X_2, \dots X_n), \left(X'_1, X'_2, \dots, X'_n\right) < 0$ refers $(X_1, X_2, \dots, X_n), (X'_1, X'_2, \dots, X'_n)$ are two discordant pairs where for each pair one component is larger than the corresponding to the other component and one is smaller (Joe 1997).

Because Kendall's tau and Spearman rho are invariant to strictly increasing transformations, they can be expressed in terms of a copula with the *C* associated with F (Joe 1997).

$$\tau = \int 4\, C\, dC - 1$$

$$\rho = 12 \int \dots \int u_1\, u_2 \dots u_n dC(u_1, \dots u_n) - 3 \tag{35}$$

$$= 12 \iint \overline{C}(u_1 \dots u_n) du_1\, du_2 \dots du_n - 3$$

Tail dependence captures the dependence in extreme values measured in the upper-quadrant or lower quadrant and it is also invariant to increasing transformations. Tail dependence is defined for a multivariate copula C in the following equations.

$$\lim_{u \to 1} \frac{\overline{C}(u, u, \dots, u)}{u} = \lambda_u \tag{36}$$

$$\lim_{u \to 0} \frac{\underline{C}(u, u, \dots, u)}{u} = \lambda_L \tag{37}$$

There exist upper tail dependence if $\lambda_u \epsilon (0,1]$ and no upper tail dependence if $\lambda_u = 0$. On the other hand, C has lower tail dependence if $\lambda_L \epsilon (0,1]$ and no lower tail dependence if $\lambda_L = 0$.

Copula Methods Applied to a Stochastic Weather Generator

The application of the copula technique for modeling climatological variables implies overcoming some challenges. First, no general criterion for selecting the copula family has been established because there is not a generalized GOF test methodology for multivariate copula. Second, the marginal distributions are unknown; however, marginal distributions determine the copula of a distribution and the rate of convergence in the tail dependence. Third, copula does not solve the problem of dimensionality, but allows several kinds of dependence structures redirecting the problem toward finding parametric distributions for high dimensional random vectors (Schölzel and Friederichs 2008).

Data

The SWG methodology is applied to simulate climate for three weather stations: one in Montana, one in Washington and one in Texas. The weather stations are Pondera County, Conrad-MT1974, Spokane, Spokane County, Washington and Temple, Bell County, Texas.

All of these climatological stations provide daily information about maximum temperature, minimum temperature and precipitation. The information was obtained from the National Oceanic and Atmospheric Administration (NOAA) website, from January 1st 1960 to December 31st 2010.[16]

Based on daily historical average temperatures for 50 years (1960-2010), the date selection criterion for the copula estimation is focused on the dates with the highest absolute deviations from mean with respect to the average monthly observation.

Variable Detrending

The modeling of the weather variables requires the decomposition of the series when some sequential or cyclical patterns are observed. Temperature cannot be well modeled using random walks because these variables include seasonal variations,

[16] http://gis.ncdc.noaa.gov/map/cdo/?thm=themeDaily

cyclical patterns and high autocorrelation, which explains that its short-term behavior differs from the long-term behavior (London 2007).

In general terms the standard methodology consists of decomposing the series in long-term trend, seasonal behavior and white noise. In particular, for the case of temperature, given the seasonal and cyclical nature of the temperature the model for detrending the series could incorporate mean reversion in the process because the temperature seems to vary between (London 2007).

Harmonic analysis is useful to extract the fluctuations and variations in the series, using sin and cosine functions.[17] Application of harmonic series requires three adjustments (Wilks 2011). First, the fundamental frequency term $w_1 = 2\Pi/n$ rescales proportionally time to angular measure, i.e. specifies the fraction of the full cycle over the whole data series (given n, the length of the data is considered as a full cycle of 360^o or 2π radians in angular measure). Second, the amplitude (C_1) is the determination of the stretching or compressing of the cosine or sine into the range of the data. Third, the phase angle or phase shift (ϕ) that makes the lateral adjustment of the harmonic function.

$$Y_t = \bar{y} + C_1 \cos\left(\frac{2\Pi t}{n} - \phi\right) + C_1 \sin\left(\frac{2\Pi t}{n} - \phi\right) \quad (38)$$

$$C_1 \cos\left(\frac{2\Pi t}{n} - \phi\right) = A_1 \cos\left(\frac{2\Pi t}{n}\right) + B_1 \sin\left(\frac{2\Pi t}{n}\right) \quad (39)$$

[17] These periodic functions have repetitive patterns every 2π radians or 360^o and they oscillate around their average value of zero and attain maximum values of +1 and minimum of -1. The cosine function is maximized at 0^o, 360^o and so on, the sine function is maximized at 90^o, 450^o and so on.

Where $A_1 = C_1 \cos(\phi)$ and $B_1 = C_1 \sin(\phi)$ are the amplitudes of an upshifted cosine and sine waves. The parameters A1 and B1 are calculated by using standard regression methods.

This detrending technique was applied to daily observations of the three variables for all of the weather stations, considering a cycle of 365 days. The coefficients of the regressions were significant for minimum and maximum temperatures. Once the trend was removed from these data series, from the detrended data some dates were selected for the copula parameter estimation. However, any trend specification was significant for precipitation. Figure 3, in appendix A, shows the application of this method for the maximum temperature monthly anomalies with data from Conrad, Montana weather station. Trend was highly significant: in Conrad weather station, trend explains 66% in maximum and minimum temperature behavior; in Spokane weather station, trend explains 80% and 70% respectively, and in Temple weather station, the trend explains 69% and 75% respectively (see tables 1.A 1.B, and 1.C, in appendix B).

Thus, the estimated trend was taken away from the original daily observations and hence, all estimation processes were carried out using detrended data series.

Selection Process for Marginal Distributions

Parametric distributions have been widely used to model climate variables. In parametric density fitting, the criterion of selection for the best fit distribution relay on Maximum Likelihood as a competitive indicator of goodness of fit, especially if the parametric densities have the same number of parameters. However, when the number of parameters differs, the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are able to derive a conclusion about fitting the distributions.

Even when the assumption is not justified, frequently Gaussian distributions are assumed for modeling temperature using the Box-Cox transformation (Wilks 2011). In contrast, although it is mathematically possible to fit precipitation into a Gaussian distribution, the adjustment is not good because of its asymmetry and right skeweness (Wilks 2011), see figure 4.C. in appendix A. Also, its mixed character (discrete and continuous) and its discontinuity in probability distribution between zero and non-zero observations increases the difficulty of its estimation. On the other hand, the versatility of the gamma distribution for modeling precipitation is suitable, but the estimation of the two parameters for a gamma is complex because they do not exactly correspond to the moments of the distribution (Wilks 2011).[18]

The pitfall of the parametric approach is the a priori assumption of the parametric functional form of the variable to be estimated. Misspecification often occurs because restrictive assumptions can result in a misrepresentative characterization of the true density, thus producing erroneous estimates that lead to unsound inference. Nonparametric characterization of the marginal distribution is a potential option because of its flexibility. Instead of assuming a functional form, nonparametric representation requires some regularity conditions such as smoothness and differentiability. However, non-parametric approach requires more data to achieve the same grade of precision as a

[18] The gamma probability distribution function can take a broad range of shapes depending on its shape parameter α and the scale parameter β which stretch or squeeze the function to the right or the left depending on data. The mean is the product of the two parameters ($\alpha\beta$) and variance in $\alpha\beta^2$, and draws simulated from estimators show that the median is below the real median because the distribution is positively skewed (Wilks 2011).

parametric model, which is not a problem for the case of weather data (Wand and Jones 1995).

The nonparametric density estimator assumes no pre-specified functional form Kernel.

$$\hat{f}(x;h) = \frac{1}{nh}\sum_{i=}^{n} K\left\{\frac{(x - X_i)}{h}\right\} \quad (40)$$

Where K is a function that satisfies $\int K(x)dx = 1$, which is the kernel and h is bandwidth or window width and is a positive number. K is chosen to be a unimodal probability density function that is symmetric about zero ensuring that $\hat{f}(x;h)$ is a density (Wand and Jones 1995). For a given sample size n, if h is small, the resulting estimator will have a small bias but a large variance. Conversely, if h is large, the resulting estimator will have a small variance but large bias. Minimization of the Mean Square Error (MSE) – which is the error measure of the estimation of the density at a single point of the density kernel function – is a consequence of the bandwidth optimal selection, which requires the balance of the bias squared and the variance terms.

There are numerous kernel functions: uniform, triangular, biweight, triweight, Epanechnikov, normal. However, as Wand and Jones (1995) pointed out, the choice of the shape of the kernel function is not a particular important, but the choice of the bandwidth value is the big issue.

Table 2 shows, in appendix B, the results of the parametric estimation for the marginals. The AIC and the BIC show that for maximum temperature and minimum temperature the best parametric specification is the normal distribution.[19] In contrast, the AIC and the BIC pointed out that the best specification for the precipitation is the extreme value distribution; however, this distribution hardly provides a good description of the data distribution. Thus, parametric specification shows a poor fit for precipitation because of the high rate of zero rainfall.

In general terms, the graphs of the adjusted probability distribution function (p.d.f.) show a good kernel fit and also a good parametric fit under the normal distribution of the maximum temperature. In the case of precipitation and minimum temperature, kernel clearly attains a better fit. Although the parametric distribution could result more efficient, the large volume of weather data provides reliability on non-parametric estimations that usually captures more accurately the probability in the tails of the distribution.

For these reasons, this research adopts the kernel specification of the probability distribution for the three weather variables originated in all weather stations (see figures 4, 5, and 6, in appendix A).

Copula Estimation

The goodness of fit test (GOF) helps to determine if the observed data are well modeled by the specified dependence structure of the multivariate distribution for the specific family of parametric copulas.[20] However, the development of a GOF test for the mixture copula exceeds the primary objective of this research.

---

[19] The most negative AIC and BIC indicates the best adjustment.

[20] There are three groups of methodologies. The simplest approach assumes dependence structures. The second kind uses statistical tests of the arbitrary parameters such as the rank-based statistics, kernels,

Although important advances have been attained in GOF test, the formal methodology to test the GOF for a copula is just recently emerging (Genest et al. 2009). Most of the progress has been done for a one-dimensional test, while in the multivariate case, there is no consensus. Although recent advances in copulas GOF have centered in "blanket test type", in the multivariate case the advances are not robust enough because the value of the statistics depends on the order in which the variables are conditioned. So, different conditioning decisions could lead to different results (Genest et al. 2009).

Because of the inconclusive information that GOF can provide in multivariate analysis, the selection of the appropriate copula family was based on a ranking copula criterion that measures the likeness of that sample coming from a given distribution. Although maximum likelihood criterion cannot be properly the criterion for the selection of the copula family because parametric distributions are unknown, it is possible to use maximum likelihood as a common base measure for all component mixtures that indicates if the conditional mixture effectively enhances flexibility.

It is impossible to prove all copula mixtures; however, in this context the maximum likelihood provides some discernment about the applicability of a particular distribution to every sample. Table 3, in appendix B, shows AIC and the BIC statistics for the considered weather stations, where the three best specifications are attained by the one-parameter-Gumbel copula family.

---

weight functions and associated smoothing ad hoc categorization of the data. Finally, the "Blanket tests" can be applied to any specification and do not require selection for kernel and optimal bandwidth (Genest et al. 2009).

Brownian Bridge Treatment and Construction

In general terms, Brownian motion describes the random movement of particles in multidimensional space. By definition, the Brownian Motion on [0,T] is the stochastic process $\{W(t),\ 0 \le t \le T\}$ which satisfy the following properties (Glasserman 2010):

i) Centered, W(0)=0;

ii) The Mapping t → W(t) is, with probability 1, a continuous function on [0,T];

iii) The increments $\{W(t_1)-W(t_0), W(t_2)-W(t_1),\ldots,W(t_k)-W(t_{k-1})\}$ are independent for any k and any $0 \le t_0 < t_1 < \cdots < t_k \le T$;

iv) W(t)-W(s) ~N(0, t-s) for any $0 \le s < t \le T$

v) As a consequence of (i) and (iv) it can be inferred that

W(t) ~ N(0, t) for $0 < t \le T$.

As a stochastic process, Brownian motion has the property of scaling invariance property, which identifies a transformation on the space of functions which changes the individual Brownian random functions but leaves their distribution unchanged (Glasserman 2010).

The Brownian Bridge has stationary increments but non-independent, in contrast with Brownian motion that has independent increments. Dependent increments in Brownian Bridge are the result of conditioning the final value to be canceled in the considered interval. Let $Z_1,\ldots,Z_n$ be independent standard normal random variables. For a standard Brownian motion set t_0=0 and W(0)=0. The subsequent values can be generated as

$$W(t_{i+1}) = W(t_i) + \sqrt{t_{i+1} - t_i}\, Z_{i+1}, \quad i = 0, \dots, n-1 \tag{41}$$

For X~ BM(μ, σ^2) with constant m and s, given X(0), set

$$X(t_{i+1}) = X(t_i) + \mu(t_{i+1} - t_i) + \sqrt{t_{i+1} - t_i}\, Z_{i+1}, \quad i = 0, \dots, n-1 \tag{42}$$

These methods are exact in the sense that the joint distribution of the simulated values $[W(t_1),\dots,W(t_n)]$ or $[X(t_1),\dots,X(t_n)]$ are the same for the joint distribution of the corresponding Brownian motion at $[t_1,\dots,t_n]$. The vector $[W(t_1),\dots,W(t_n)]$ is a linear transformation of the vector of increments $[W(t_1), W(t_2) – W(t_1), \dots, W(t_n) – W(t_{n-1})]$ because these increments are independent and normally distributed, then $[W(t_1),\dots,W(t_n)]$ has a multivariate normal distribution (Glasserman 2010).

For a standard Brownian Motion, the mean $E[W(t_i)] = 0$, so for the covariate matrix and $0 < s < t < T$; using the independence of the increments,

$$Cov[W(s), W(t)] = Cov\big[W(s), W(s) + \big(W(t) - W(s)\big)\big]$$

$$= Cov[W(s), W(s)] + Cov\big[W(s), \big(W(t) - W(s)\big)\big] = s + 0 = s \tag{43}$$

If Cov denotes the covariance matrix of $[W(t_1),\dots,W(t_n)]$, then

$$Cov_{ij} = \min(t_i, t_j) \tag{44}$$

Given the Brownian motion $\{W(t), t\geq 0\}$, $T>0$, then

$$B(t) = W(t) - \frac{t}{T} W(T), \qquad t \in [0, T] \tag{45}$$

Is a stochastic process of Brownian Bridge independent of W(T). However, when the Brownian bridge realizations satisfy $B(t) = x$ and $B(T) = y$ they are the initial and final points, respectively, and the Brownian Bridge can be expressed as

$$B_{0,x}^{T,y}(t) = x + W(t) - \frac{t}{T}(W(T) - y + x) \tag{46}$$

In fact, the Brownian Bridge Matlab program generates the underlying Brownian motion process by successive increments. The Brownian Bridge construction involves a process that begins with the generation of the final value $W(t_n)$, then filling in the intermediate values amounts to simulating a Brownian Bridge from $0=W(0)$ to $W(t_n)$. Next, $W(t_{[n/2]})$ is sampled, and valued between times $t_{[n/2]}$ and t_n are filled in to simulate the Brownian Bridge from $W(t_{[n/2]})$ to $W(t_n)$ and so on.

A Brownian bridge constructed from a Brownian motion with drift μ, is the same as the one constructed from a standard Brownian motion, only the first step (sampling the rightmost point) would change. Instead of sampling $W(t_n)$ from $N(0, t_n)$, it would be sampled from $N(\mu t_n, t_n)$. The conditional distribution of $W(t_1),\ldots, W(t_{n-1})$ given $W(t_n)$ is the same for all values of μ (Glasserman 2010).

The d-dimensional Brownian Bridge construction implies the application of independent one-dimensional constructions to each one of the coordinates. To include a drift vector for $BM(\mu,I)$ process, it must be added $\mu_i t_n$ to $W_i(t_n)$ at the first step of the construction of the i^{th} coordinate, the remaining parts of the construction are the same (Glasserman 2010).

To construct $X \sim BM(\mu,\Sigma)$, X can be can be represented as $X(t) = \mu\, t + BW(t)$ with B as a dxk matrix, $k \leq d$, satisfying $BB^T=\Sigma$ and W a standard k-dimensional Brownian motion. Then a Brownian Bridge construction can be applied to $W(t_1),\ldots,W(t_n)$ and recover $X(t_1),\ldots,X(t_n)$ through a linear transformation.

X~ BM(μ,Σ) means that the process X is a Brownian motion with drift μ – with μ being a vector in R^n and Σ a nxn matrix, positive definite or semidefinite –. X is a continuous sample paths, with initial value X(0) = 0 and independent increments with

$$X(t) - X(s) \sim N\big((t-s)\mu, (t-s)\Sigma\big) \tag{47}$$

Let B a dxk matrix satisfying $BB^T = \Sigma$ and W is a standard Brownian motion on R^k, then the process BM(μ,Σ) is defined by

$$X(t) = \mu\, t + B\, W(t) \tag{48}$$

In particular the law of X depends on B only through BB^T, then the process in equation 48 solves the stochastic differential equation

$$dX(t) = \mu\, dt + B\, dW(t) \tag{49}$$

So, extending the definition to a d-dimensional Brownian motion to deterministic, time varying μ(t), and Σ(t) through the solution to

$$dX(t) = \mu(t)\, dt + B(t)\, dW(t) \tag{50}$$

Where $B(t)B(t)^T = \Sigma(t)$. this process has continuous sample paths, independent increments and

$$X(t) - X(s) \sim N\left(\int_s^t \mu(u)du\, , \int_s^t \Sigma(u)du\right) \tag{51}$$

In this terms, if X~ BM(μ,Σ) then

$$Cov\left(X_i(\mathrm{s}), X_j(\mathrm{t})\right) = \min(s,t)\Sigma_{ij} \tag{52}$$

Let Z_1,Z_2,…be independent N(0,1) random vectors in R^d. The standard d-dimensional Brownian motion at times $0 = t_0 < t_1 < \ldots < t_n$ by setting W(0) = 0 and

$$W(t_{i+1}) = W(t_i) + \sqrt{t_{i+1} - t_i}\, Z_{i+1}, \quad i = 0, \dots, n-1 \tag{53}$$

Equivalent to applying the one-dimensional random walk construction separately to each coordinate of W. To simulate X~ BM(μ,Σ), first matrix B is found for which $BB^T = \Sigma$. If B is dxk, let Z_1, Z_2,… be independent standard normal random vectors in R^k. Set X(0)=0 and

$$X(t_{i+1}) = X(t_i) + \mu(t_{+1} - t_i) + \sqrt{t_{i+1} - t_i}\, B\, Z_{i+1}, \quad i = 0, \dots, n-1 \tag{54}$$

The simulation of BM(μ,Σ) is simple once Σ has been factored (Glasserman 2010).

However, in the particular case of the Brownian Bridge construction for the SWG, in the construction of the tridimensional Brownian motion one of the variates (Z_1, Z_2…) is truncated to emulate the precipitation behavior. As a consequence μ and Σ have to reflect such circumstance. The coefficients of the multivariate normal are determined by fitting the historical weather variables (maximum temperature, minimum temperature and precipitation) using the maximum likelihood estimation method. The parameters estimation is from a population with single truncated sample, normal p.d.f. and the truncation point of zero. Cohen (1991) shows the analytical solutions for $\bar{x}$ and σ, derived using maximum likelihood estimation.

When restriction occurs only in one of the variates of the multivariate distribution, such as in the case of precipitation; say X= (x_1,x_2,x_3), is the trivariate distribution with the following p.d.f. equation.

$$f(\boldsymbol{X}) = 2\pi^{-3/2} \left| \Sigma^{ij} \right|^{-1/2} \exp^{(-1/2)(\boldsymbol{x}-\mu)'\Sigma^{-1}(\boldsymbol{x}-\mu)} \tag{55}$$

For left truncated samples, the analytical solutions for the estimates μ_1 and σ_1 have close form solutions. As Cohen (1991) shows solutions for x_1 (truncated variate) can be calculated only from marginal data of x_1, without consider for any of the other variates.

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} \tag{56}$$

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n} \tag{57}$$

$$\theta(\xi) = \frac{Q(\xi)}{Q(\xi) - \xi'} \tag{58}$$

$$Q(\xi) = \frac{\phi(\xi)}{1 - \Phi(\xi)} \tag{59}$$

$$\mu = \bar{x} - \theta(\xi)(\bar{x} - T) \tag{60}$$

$$\sigma^2 = s^2 + \theta(\xi)(\bar{x} - T) \tag{61}$$

Where n is the number of truncated rain-rate samples, $\theta(\xi)$ is the auxiliary estimation function, and $\phi(\xi)$ and $\Phi(\xi)$ are probability distribution function and cumulative distribution function of the standard normal distribution, respectively.

Estimation of parameters of the remaining two variates and their correlation coefficients show the following pattern.

$$\hat{\mu}_j = \overline{x_j} - \overline{r_j} \frac{s_j}{s_1} (\overline{x_1} - \hat{\mu}_1), \tag{62}$$

$$\hat{\sigma}_j = \overline{s_j} \sqrt{\frac{1 - \hat{\lambda} \left(1 - \overline{r_{ij}^2}\right)}{1 - \hat{\lambda}}} \tag{63}$$

$$\hat{\rho}_{ij} = \frac{\overline{r_{ij}} - \hat{\lambda}\left(\overline{r_{ij}} - \overline{r_{1i}}\,\overline{r_{1j}}\right)}{\sqrt{\left[1 - \hat{\lambda}\left(1 - \overline{r_{1i}^2}\right)\right]\left[1 - \hat{\lambda}\left(1 - \overline{r_{1j}^2}\right)\right]}} \tag{64}$$

For i=1,2,…,p-1, j=2,3,…,p,i<j.

$$\hat{\lambda} = 1 - \frac{\overline{s_1^2}}{\overline{\overline{\sigma_1^2}}} \tag{65}$$

Since by definition r_{ii}=1, the last equation for i=1 becomes

$$\hat{\rho}_{1j} = \frac{\overline{r_{ij}}}{\sqrt{\left[1 - \hat{\lambda}\left(1 - \overline{r_{1j}^2}\right)\right]}} \tag{66}$$

For more details on this issue, consult Cohen (1991).

Brownian Bridge is useful in the context of this application because it is able to generate high quality sequences to outline the paths of the Brownian motion process, by sampling the weather generated by copulas acting as milestones (or borderline conditions), the sequence can be filled using Monte Carlo methods. The Brownian Bridge could represent an advantage when it is used with variance reduction techniques and low-discrepancy methods.

Summary of the Applied Methods for the Copula Based SWG

The applied methods that compose the Copula Based SWG are briefly detailed. First, the selection of the dates for the estimation of the Copula parameters is carried out based on the criterion of the highest average monthly anomaly. Second, the original daily weather observations are detrended to extract the cyclical patterns in the series. Third, the detrended selected dates are used to estimate the parameters from different copula mixture specifications and the best specification is determined. Four, parameters on the trivariate normal with one truncated variate are obtained from the detrended weather variables using daily observations, one set of parameters per month. Five, simulation of the weather variables is carried out using the Copula parameters previously obtained. Six, Brownian Bridge generation is carried out to emulate daily dynamics of the weather variables using the Copula simulations as the borderline of the Copula based SWG and the trivariate normal simulations. Seven, trend is incorporated into the daily simulated variables.

Summary

The proposed technique considers the application of the copula methods for the stochastic generation of daily values for the weather variables: maximum temperature, minimum temperature and precipitation. This methodology solves three important issues. First, the selection of the marginal distributions which was determined individually based on the best fit for every variable including parametric and non-parametric approach. Second, the solution of the dimensionality problem that relays in the reduction of the sample for the copula estimation. This principle selects the dates with the highest average monthly anomalies in the sample. Third, the copula family selection criterion for the final representation of the multivariate model, which was established on a common base measure for all component mixtures: the likelihood function. Under this measure the best specification for the three models was the one-parameter Gumbel.

CHAPTER III

STATISTICAL VALIDATION FOR THE COPULA BASED STOCHASTIC WEATHER GENERATOR

Introduction

The statistical validation of the Copula based SWG represents a key issue in the generation of weather series for crop simulation models such as the WinEPIC. The copula based SWG was tested in nine locations with two different climatic conditions. Weather stations are located in Conrad, Montana; Spokane, Washington and Temple, Texas.

An evaluation of the Copula based SWG performance versus the Richardson's SWG performance, currently used in the WinEPIC, is carried out to learn about their strengths and limitations. Copula based SWG does not assume parametric specifications; instead, Copula based SWG was designed on non-parametric modeling using kernel smoothing and copula methods to capture jointly the adjacent weather patterns in the series. Also, Copula based SWG relies on Brownian Motion to emulate the daily behavior of the weather series and used Monte Carlo methods to replicate the behavior of the observed weather series.

First, because any parametric functional form is being used, a non-parametric two sample Kolmogorov-Smirnov test can be used to evaluate the performance of the Copula method to replicate the distribution for the weather series.

Next, a deeper analysis by periods is applied to get more detailed information on the Copula based SWG's performance in contrast to the Richardson's SWG performance.

In general terms, although the Copula based SWG provides a good representation and an acceptable replication of the observed weather patterns from historical data, there is no a conclusive evidence on which SWG has the best performance. However, one remarkable characteristic of the Copula based SWG is that I t provides accurate representations on magnitudes of extreme weather events in temperatures.

Monte Carlo Methods

Weather models are stochastic representations that replicate daily variations on weather. The parameters of such models represent specific characteristics of the local climate that the Monte Carlo simulation technique reproduces by random number generators which resembles daily weather. Thus, the weather series generated cannot be duplicated at any time. Monte Carlo methods are a fundamental component of the SWG and the laws that govern the samples generated by this method are also applicable to those daily weather series generated by SWG.

Monte Carlo simulation method is a numerical calculation method that performs numerical computations of random variables. Basically Monte Carlo is a method which simulates independent realizations of the stochastic event z as an estimate for the probability or expectation of the phenomenon via an appropriate estimator obtained from independent samples (Asmussen and Glynn 2007). The probability $z = P(W_n > x)$ can be calculated as the sample proportion of the W_{rn} that is greater than x. The estimator $\hat{z}$

of $z = \mathbb{E}Z$ is developed by an algorithm that generates independent and identical distributed (i.i.d.) random variables $Z_1,...,Z_R$ and estimates z from the sample by the expectation estimator.

$$\hat{z} = \widehat{z_R} = \frac{1}{R}\sum_{r=1}^{R} 1\{W_{rn} > x\} \tag{67}$$

Where $\mathbb{1}$ is the indicator function and the Law of Large Numbers (LLN) guarantees that the algorithm converges to as the number of independent replications goes to infinite (Asmussen and Glynn 2007). Monte Carlo method is able to generate independent sequences under the distributional assumptions defined.

Precision or the number of simulation required for attaining convergence under the LLN can be improved using the central Limit Theorem (CLT). Thus, assuming $\sigma^2 \stackrel{\text{def}}{=} Var\ Z < \infty$

$$\sqrt{R}(\hat{z} - z) \xrightarrow{D} N(0, \sigma^2) \ \ as\ R \to \infty \tag{68}$$

This can be expressed as

$$\hat{z} \approx z + \frac{\sigma V}{\sqrt{R}}\ where\ V \sim N(0,1) \tag{69}$$

When R is large $\hat{z}$ converges in distribution with a convergence rate of the order $R^{-1/2}$. However, because the error for large R is asymptotically normally distributed, the error for large R depends on the standard deviation σ and it is possible to assess accuracy by the confidence intervals derived from the normal distribution (Asmussen and Glynn 2007). Because of the CLT, z_α denotes the α-quantile of the normal distribution $\Phi\ (z_\alpha) = \alpha$ with the asymptotic probability of the event

$$\left(z_{\frac{\alpha}{2}}\sigma\sqrt{R} < \hat{z} - z < z_{1-\frac{\alpha}{2}}\sigma\sqrt{R}\right) \tag{70}$$

So, the interval is

$$\left(\hat{z} - z_{1-\frac{\alpha}{2}}\sigma/\sqrt{R}\,, \hat{z} - z_{\frac{\alpha}{2}}\sigma/\sqrt{R}\right) \tag{71}$$

Because in practice ${\sigma_z}^2$ is unknown, it can be estimated by its sample estimator

$$\widehat{s_z}^2 = \frac{1}{R-1}\sum_{r=1}^{R}(z_r - \hat{z})^2 \tag{72}$$

In general terms, precision on $\widehat{s_z}^2$ is complex to obtain; however, because z is Gaussian then $\widehat{s_z}^2$ tends toward σ_z^2 for every large N (Huynh et al. 2008).

$$\left(\hat{z} - z_{1-\frac{\alpha}{2}}s/\sqrt{R}\,, \hat{z} - z_{\frac{\alpha}{2}}s/\sqrt{R}\right) \stackrel{\text{def}}{=} \hat{z} \pm z_{1-\frac{\alpha}{2}}s/\sqrt{R} \tag{73}$$

Where 1-α is the asymptotic confidence interval for z. The speed of convergence is measured by the size of the confidence interval. The standard choice for 1-α= 95% =1.96, so the confidence interval is $\hat{z} \pm 1.96\ s/\sqrt{R}$.

There are several methods for increasing the efficiency of Monte Carlo simulation by reducing the variance of simulation estimates. However, the implementation of a reduced-variance estimator with a valid confidence interval requires sacrificing some potential variance reduction.

Methods such as antithetic sampling, control variates, conditional sampling, stratified sampling or importance samplings are common. However, they vary in effectiveness and complexity (Asmussen and Glynn 2007). While the antithetic sampling is easier to be implemented because it does not require specific information about a simulated model, their efficiency is minor. In contrast, importance sampling is the most complex method because it has the capacity to exploit detailed knowledge about a model (often in the form of asymptotic approximations) to produce orders of magnitude variance reduction (Glasserman 2010). The adequate application of importance sampling method can attain an effective reduction in variance.

Monte Carlo Method in the Copula Based Stochastic Weather Generator

Monte Carlo methods are a fundamental component of the Copula based SWG. They get involved in two different stages of the climate generation. First, the conditional Monte Carlo method provides variance reduction in the simulation of Copula draws. This three-step process for the multivariate copula simulation is broadly described in Chapter II, which basically follows Cherubini et al. (2004) and Salvadori et al. (2007).

Second, Monte Carlo method is involved in the simulation of the Brownian Bridge to emulate the daily dynamic of the weather series. Brownian Bridge uses high quality sequences to outline the paths of the Wiener process, by sampling points acting as the milestones; they can fill the trajectory Monte Carlo sampling or even better quasi-Monte Carlo Methods (Brandimonte 2006). The property of stationary independent increments of the Brownian Bridge makes the simulation process equivalent to the random variable generation from a specific infinitely divisible distribution (Glasserman 2010). Because Brownian Bridge relies on Brownian Motion, it exhibits Markovian property that aggregates more persistence in the simulated weather series.

Statistical Tests on Simulated Weather Data

Daily weather data for three locations with highly differentiated weather patterns across the United States were generated by the Copula based SWG. Following Richardson's (1981) research, observations from three Weather stations in Conrad, Montana; Spokane, Washington and Temple, Texas were used to test their accuracy properties.

Parameters were estimated with data from Conrad and Spokane weather stations using 50 years of daily observations (1960-2010); for Temple the estimation was carried out with 42 years of daily data. Year data were partitioned into 12 observations per year (one per month) according to the highest average anomalies recorded and whose distributions are replicated by Copula methods. The complete methodology of the SWG based in copulas is accurately described in the previous chapter.

Two-Sample Kolmogorov-Smirnov Test

The two-sample Kolmogorov-Smirnov test (KS) is applied to Copula simulations to compare the c.d.f. of the generated weather series vs. the c.d.f. of original observed weather data at each one of the three locations in Montana, Washington and Texas. In this context, this non-parametric test compares two unknown c.d.f.s: F for the observed data and G for the simulation, quantifying the distance between the empirical distribution functions of the two samples through the test statistic in the following expression

$$D_n = \sup |F_{n1} - G_{n2}| \tag{74}$$

Where F_n_1 is the empirical c.d.f. from a sample of n_1 data values (observed weather data) and G_n_2 is the empirical cdf from a sample of n_2 data values (simulated data), being F_{n1}, G_{n2} continuous distributions.

The null hypothesis is H_0: $F_{n1} = G_{n2}$. The fit is measured by the statistic $D_n$with its asymptotic distribution and the limiting distribution $\sqrt{n}D_n$ is distribution free, in consequence, the reasonable criterion is to reject H_0 if D_n is large. (Mood et al. 1974).

A sample of 120-year draws for every one of the three locations was generated using both SWGs and the two-sample Kolmogorov-Smirnov test was applied.

Table 4.A and 4.B, in appendix B, show the p-values for the selected dates generated. Simulated dates that reject the H_0 are marked with asterisks, in these cases the probability distribution of the simulated weather data does not correspond to the probability distribution of the observed data. In the case of the Copula based SWG simulations the H_0 is rejected in 37% of the cases, while in the case of Richardson SWG is 27%; however the cases of rejection are concentrated in the simulation for the Spokane weather station.

However, this rate of rejection in the case of the Copula based SWG can be attributed to the fact that the KS test is more sensitive to median values than to extreme values of the distribution, and Copulas precisely tend to capture more information from the tails of the observed distributions (or extremes of the distribution). In particular, this result makes sense because the Gumbel Copula family used to model weather series is characterized by upper tail dependence.

Quantile Analysis

Quantiles of the distributions are calculated to analyze in detail the differences in the distributions for the simulated weather series versus the observed weather series. A 120-year simulation was performed to carry out the quantile analysis.

The quantiles of a distribution are points taken at regular intervals c.d.f. function that provides nonparametric estimators of their population counterparts based on a set of independent observations $\{X_1,X_2,..,X_n\}$ from the distribution F. Quantile of the distribution F is define by the following expression.

$$Q(p) = F^{-1}(p) = \inf\{x: F(x) \geq p\}, \qquad 0 < p < 1 \tag{75}$$

Let $\{X_{(1)},X_{(2)},..,X_{(n)}\}$ denote the order statistics of $\{X_1,X_2,..,X_n\}$ and let $\widehat{Q}_\iota(p)$ denotes the ith sample quantile.

Table 5, in appendix B, shows values of the weather variables for different quantiles of the distribution. The Copula based SWG generates weather series significantly closer to the original observed data. Although the reproduction of the weather patterns is consistent, the replication of the climate is comparatively better for the station of Spokane, Washington and Temple, Texas than for Conrad, Montana. The values of the lower percentiles are more accurate in the case of the simulations generated by the Copula based SWG. This result could be attributed to the property of copulas to capture more information in the tails of the distribution.

Statistical Analysis of the Simulated Weather Series

The validation of a weather generator based only on the analysis of their moments distribution (mean, standard deviation, skewness and kurtosis) is insufficient. A more accurate description of the occurrence of precipitation by season provides key information to evaluate the performance of the Copula based SWG. For such purpose 28-day period indicators were calculated for both, the generated and the observed weather data series. Next, mean values of accumulated precipitation amounts (cm), mean number of rainy days, mean minimum temperature and mean maximum temperature per period were calculated.

Table 6, in appendix B, shows that the simulated mean precipitation amounts do not differ significantly from the values obtained from the observed data. However, the replication of the amount of water from precipitation is more accurate for locations with higher amounts of water such as Temple, Texas than in locations with low levels of rainfall during the year. The average number of days per period generated by the Copula based SWG was in general terms close to the observed data. However, the Copula based SWG shows certain inflexibility in replicating the amounts of water and the recurrence of rain periods in highly variable precipitation patterns.

The same analysis is applied for the daily simulated temperatures. Table 7, in appendix B, shows the mean maximum temperature and the mean minimum temperature for 120-years of generated series and for the observed weather series. The means for the maximum and minimum temperature in the three weather stations are close to the observed data. The differences in averages can be mainly attributed to the detrending technique by harmonic analysis described in detail, in Chapter II.

Both SWGs reproduce significantly close weather patterns in the three weather stations. However, there is no conclusive evidence about how to rank the accurateness of these models.

Table 8, in appendix B, summarizes the capacity of the Copula based SWG to reproduce the distribution of the annual extreme temperatures in minimum temperature and maximum temperature series. The comparative analysis of the generated and the observed data in Table 8, in appendix B, confirms that Copula based SWG reproduces much closer the patterns of extreme events in weather series. Both, Copula based SWG and Richardson's SWG, show about the same number of days with extreme temperatures; however, the Copula based SWG shows a better replication in magnitude of the temperature extreme events of the observed data for the three weather stations.

Summary

Tests on copula based SWG showed that the model is able to represent the main features for the distributions of the observed weather variables. For the three weather stations, the Gumbel with a single parameter was the best specification. Although there is no a clear insight about which SWG has the best performance, the copula based SWG shows a better performance in the reproduction of the extreme weather events.

CHAPTER IV

COMPARATIVE EVALUATION OF THE COPULA BASED STOCHASTIC WEATHER GENERATOR: AN APPLICATION FOR CROP GROWTH MODELS AND CROP INSURANCE

Introduction

The performance of the SWG is evaluated in terms of the Camelina yields produced by two different weather generators – Copula based SWG versus Richardson's SWG, currently used in WinEPIC – and in terms of Average Production History (APH) insurance schemes for Camelina.

Camelina is an oilseed crop recently growed in some North areas in the United States. Because no historical data on Camelina yields are available for rating of the new insurance scheme, the alternative solution is to obtain these data from crop growth models like the Environmental Policy Integrated Climate model (WinEPIC). In such models, weather is one of the main determinants of crop yields. Given that the Copula based SWG more accurately reproduces the observed extreme weather events, it is expected that the yields generated using this weather generator will reflect more accurately the effect of extreme weather events on insurance premiums. This exercise is applied in the specific location of Conrad, Pondera County, Montana. For the sake of this research, some issues and results from the Risk Management Agency study (RMA 2011) are considered.

Energy Crops

The global crisis triggered by the rise of the world food prices during 2007 and 2008 caused a renewed interest for oilseed crops as a feedstock for renewable fuels. In particular, part of increase in food prices was attributed to the diversion of food crops (maize in particular) for producing first-generation biofuels in coincidence with weather-related cereal production shortfalls in Australia, U.S., EU, Canada, Russia and Ukraine.[21]

The Energy Independence and Security Act of 2007 had a considerable impact on U.S. energy policy, making the production of corn more profitable than other crops which lead to significant increasing of corn acreage and reduction in soybean and wheat acreages. In this circumstance, non-food crops like Camelina, Jatropha, Crambe, Castor bean, safflower, switch grass, seashore mallow and mustard are being considered for biofuel production. These crops can prosper on marginal agricultural land where edible crops do not.

This research will focus the analysis on the crop of *Camelina sativa* (Camelina). This is a spring annual oilseed plant of the mustard (Brassicaceae) family (genus Cruciferae), a distant relative to canola. This crop, originally from Central Asia and traditionally cultivated in Europe, has shown some advantages over other oilseed plants: it matures earlier than other oilseed crops, it is more drought tolerant, greater spring freezing tolerant, and more resistant than canola to flea beetles. All these features imply greater economic advantages from reduced production costs in some climates compared to other oil crops such as canola or oilseed rape (Johnson 2007).

[21] From 2004 to 2007 the use of maize for ethanol grew exponentially, using about 70% of the increase in global maize production for such purposes (Donald, 2008).

Camelina is adequate for the same growing areas as canola, flax, and mustard such as Idaho, Montana, Minnesota, Oregon, North Dakota, South Dakota, and Washington. However, agronomical trials on Camelina in Texas have not shown a favorable experience on yield and quality.[22] Specialists pointed out that until additional equipment and/or genetic improvements take place, Camelina yields will not be competitive with other cool-season, oil-seed crops. The two major limitations to Camelina are establishment of very small seed and shattering prior to harvest. However, this result could be inconclusive because of limited experience with Camelina in Texas.[23]

At present, contract farming is used as the predominant method for producing Camelina in states with acceptable performance (Montana, North Dakota, Oregon, and Washington). Processors and first handlers contract growers and set up production conditions under fixed price terms (RMA 2011).

Mainly the demand for Camelina comes from the U.S. Navy for its biodiesel jet fuel and other companies such as Great Plains Oil & Exploration-The Camelina Company, Sustainable Oils, and Willamette Biomass Processors (RMA 2011).

In Montana Camelina has been grown since 2004. In 2010 crop year 9,900 acres of Camelina were planted, of which 9,400 acres were harvested. In 2009 20,800 acres were planted of which 19,500 acres harvested (NASS website, last accessed 5/5/2012: http://quickstats.nass.usda.gov). Montana farmers consider that Camelina is a low input,

---

[22] Camelina trials were conducted in the Agricultural Experimental Stations of Texas A&M University for between 5-7 locations across Texas for 3 years (2008-2010).

[23] Personal Communication with Dr. Gaylon D. Morgan, Associate Professor and Cotton Specialist, Texas AgriLife Extension Service - Texas A&M University.

low risk, and high efficiency crop (RMA 2011). Table 9, in appendix B, shows the economic advantages of growing Camelina in Montana instead of other crops such as canola or spring wheat. Camelina production requires 33% less fertilization than canola and has additional properties that implicitly reduce its production risk.

RMA (2011) evaluated energy crops that are commercially grown and dedicated to energy production in terms of their insurance feasibility. RMA (2011) found that some of these energy crops like Camelina in Montana, Oregon and Washington are feasible because its characteristics are similar to other insured crops in these locations. In addition, higher premium estimates in Montana point out a greater demand for insurance in this State.[24]

The purpose of this research is not to design of a new insurance scheme per se, but the comparison of the performance of two weather generators in terms of their use in crop growth models as an instrument for the rating of new insurance schemes where no historical data on yields is available.

It is a priori known that the copula based SWG reproduces more accurately the extreme weather event patterns than Richardson's SWG, currently used in WinEPIC as the weather generator. For this reason, some of the issues and results of the RMA (2011) feasibility study are considered.

Furthermore, this research incorporates new dimensions to the problem initially formulated by RMA (2011). The proposed analysis evaluates the implications of using the copula based SWG in terms of fair premiums in insurance schemes for Camelina.

[24] There exists a precedent of Camelina insurance in Saskatchewan, Canada (Canada, Saskatchewan Government, 2012).

Physiology and Cultural Practices for Camelina

Camelina can be grown in semi-arid regions on dryland or with minimum rainfall, where other crops cannot be grown. In Montana, Camelina's advantage lies in better management of moisture and cold tolerance (Johnson 2007). Camelina grows up to 90 centimeters tall and has branched smooth or hairy woody stems. Camelina seed contains 29-41% oil compared to 20% in soybeans and the remaining 60-70% germplasm can be used as livestock feeding as a potential soybean meal replacement in finishing beef cattle. (Ehrensing and Guy 2008).

To simulate Camelina with WinEPIC is necessary to specify the biological parameters for the crop. Camelina is on 85-100 day lifetime-crop with a physiological cycle of six stages: seedling (0-14 days), leafing (15-45 days), blossom (46-58 days), green boll (59-77 days), boll ripening (78-100 days) and maturity (100th day) (RMA 2011). Planting dates are variable across regions. In North Central Montana, spring planting dates are from late March to late April; in Eastern Montana dates range from late April to early June.[25]

In Montana, Camelina is planted no deeper than ¼ to ½ inch using 3 to 5 lbs of seed per acre. Seedbed preparation is done by drilling the seed very shallow using packer wheels to ensure good seed to soil contact and a firm seedbed. Other cultural practices suggest the distribution of the seed onto a clean seed bed followed by a harrow or rollers.

[25] A difference of one month between planting and harvesting occurs between western Montana and Southeastern Montana (Billings), USDA website.

Seed to soil contact and soil compaction is vital and planting too deep will cause poor or no establishment (McVay and Lamb 2008).[26]

Different rotation patterns are possible for Camelina. In Montana, producers replace their fallow land with Camelina between wheat crops in a wheat/Camelina/wheat rotation pattern (RMA 2011). Better results for Camelina yields have been observed in Montana and Oregon when Camelina was planted in fields with a previously fallow or growing wheat, barley, peas or lentils. Poor performance is shown when it is planted consecutively or following canola or another Brassica such as brown mustard, canola or rapeseed. Farmers use Camelina to replace fallow in their crop rotation system because Camelina can stabilize exposed soils for erosion control.

Camelina responds to nitrogen, sulphur and phosphorus fertilizer application. Sustainable production suggests an application of nitrogen lower than 90 pounds per acre and no than less than 32 pounds of phosphorous. Areas with higher yield potential (more available moisture) may experience response to increased fertilizer rates (Ehrensing and Guy 2008).

Camelina should be planted in fields with limited weed pressure to reduce competition. A burn down of broadleaves and grassy weeds utilizing Round Up (glyphosate) is recommended prior to planting to lessen weed competition during establishment. Like canola, no herbicides are necessary because Camelina produces a natural exudate from its roots (alleopathy) and it is highly sensitive to long-term herbicide residuals (McVay and Lamb 2008).

[26]Authors suggest a minimum tillage or no-tilled (Mc Vay 2008; Ehrending and Guy 2008).

Camelina is alternaria blackspot resistant and is highly resistant to a wide range of blackleg isolates (Leptosphaeria maculans), which are major diseases of canola or Brassica crops. It exhibits variation for resistance to sclerotinia stem rot, brown girdling root rot, and downy mildew, suggesting that disease resistant cultivars can be developed (Ehnrensing and Guy 2008). Other diseases such as clubroot, white rust, and aster yellows limit its adaptation. Camelina is also susceptible to viral diseases like turnip crinkle virus and turnip rosette virus that are transmitted by flea beetles (RMA 2011).

Camelina must be harvested within a few days of maturity because pods mature, the seed easily falls from the pod. The seed moisture content must be less than eight percent to ensure proper storage quality (RMA 2011).

Several agronomical trials have been performed on Camelina by the Montana Agricultural Experiment Stations of the Montana State University. Since 2004 Dr. K. A. McVay has led a broad research program at seven agronomical experiment stations of the Montana State University to determine the best management practices. In particular, McVay and Khan (2011) conducted a two-year study to evaluate the effects of stand reduction on Camelina at different growth stages; his results suggest that Camelina exhibits plasticity to maintained grain yield across a wide range of stand reductions under dryland conditions. Yield plasticity is higher at the rosette compared to when the stand was reduced at bolting stage (McVay and Khan 2011).

Although agronomic trials provide useful information about productive practices, their results in terms of yields depend on the particular production practices that agronomists are evaluating. For this reason, the yields on agricultural experimental stations could be different from actual farmers. As a consequence, information regarding cultural practices for the calibration of the Camelina yields was obtained from farmers in Pondera County, Montana. RMA gathered information on management practices and yields from producers consistently growing Camelina for 4 years under the same management practices (RMA 2011).

WinEPIC Crop Calibration

Farmers cultivating new crops experience higher uncertainty about the response of the plant to cultural practices, weather, and natural disasters. For this reason, yield simulation can be effective for the risk analysis and the development of instruments that allow coping with such risks, such as the insurance.

Plant growth simulation models represent a feasible option for this purpose. In particular the WinEPIC evaluates production strategies considering sustainability, erosion (wind, sheet, and channel), economics, water supply and quality, soil quality, plant competition, weather and pests.[27] Also, WinEPIC is able to simulate hundreds of years of daily potential plant grow constrained by the minimum of five stress factors (water, nitrogen, phosphorus, temperature and aluminium toxicity). WinEPIC models the phenological development of the crop based on a wide set of equations that capture the processes of daily growth from emergence to harvest (Williams et al. 1989).

Farming practices are set up for WinEPIC in the form of crop production schedules. Such plans specify application rates, dates of operations prior to and during the growing season for tillage, planting, pesticides, irrigations, fertilizers, and harvesting.

[27]WinEPIC was developed in the early 1980's by Dr. J.R.Williams, Blacklands Research and Extension Center, Texas AgriLife Research, Texas A&M University, Temple, Texas,

In addition, management information such as rates and dates of crop production inputs facilitates the simulations for tillage, irrigation, and fertilizer procedures (RMA 2011). Based on the cultural practices documented in Montana region where Camelina is an important crop, parameters on leaf area development, temperature responses, development-rate, radiation-use efficiency, nitrogren and phosphorous concentrations in plant biomass and plant growth process of the crop are adjusted to calibrate the Camelina growth model.

This research considers the information and the calibration parameters from the RMA's (2011) research as a base to simulate Camelina yields. According to RMA (2011), a local Conrad producer provided his best estimates of 2007-2010 non-irrigated yields. The 2008 yield was reduced significantly by shattering and the 2009 yields suffered from harvesting losses. The farmer's rotation was wheat-fallow-Camelina, so Camelina production followed a summer fallow period for each specific field. Thus, a rotation of minimum-till winter wheat/no-till fallow/direct-seeded spring seeded Camelina was utilized for calibrating the WinEPIC model and to simulate the yields as closely as possible.

The farming practices, in table 10 in appendix B, detail the production schedules about the cultural practices, schedules for tillage, planting, fertilization, pesticide applications, and harvesting operations along with management decisions and dates for each operation, seeding rates, and application rates of fertilizers, and pesticides. The daily weather information used for the simulation was obtained from the closest weather station to the farm, Conrad MT1974.

The usual way to evaluate the robustness of the crop growth model calibration is a graph with the simulated output by the model on the “x” axis versus yield observations on the “y” axis. This graph highlights the comparison of model predictions and system measurements, a 45° line or 1:1 line would indicate a perfect adjustment of simulated data respect to the observed data. In general terms every generation from the model is accompanied by an error, even for robust models data are scattered around the 1:1 line.[28]

Divergence lines, usually of ±15%, could be determined by the observed coefficient of variation (CV) for the variable under consideration. If the model is robust, data should be located between these divergence lines, 80 or 90% of the points (Soltani and Sinclair 2012).

As table 11, in appendix B, shows the yield series simulated for a 300-acres Camelina farm that has an average of 1,576.5 lbs/acre versus 1,489.5 lbs/acre for the observed data. Figure 7, in appendix A, depicts the relationship of simulated to producer yields year-by-year in table 11, in appendix B. The regression line with a 1.001 slope and R-squared of 0.81 indicate a satisfactory adjustment.

An APH Insurance Scheme for Camelina

This research uses the Average Production History (APH) insurance scheme to carry out the comparative insurance analysis on yields generated by the Copula based Stochastic Weather Generator vs. yields generated by Richardson (1981) SWG.

This insurance scheme was evaluated by the RMA (2011) as a feasible application for energy crops. In particular for the case of Camelina, RMA concluded that

[28] This graph is common for output variables with a single value for each situation, e.g. days to maturity, crop yield, crop mass at maturity, and so forth (Soltani and Sinclair 2012).

the reasons for the pertinence of the APH insurance are mainly two. First, Camelina does not trade on a commodity exchange. Instead, in regular basis the prices are established a priori in the farming contract as a result of a private negotiation between the farmer and the processor or first handler. Second, although the lack of clear price mechanisms prevents the design of a revenue protection insurance program, in the other hand the specific contract conditions on production conditions and delivery process reduces the risk of revenue variability (RMA 2011).

APH is based on historical loss experience data and it has been broadly applied by the RMA for insuring crops in the United States. The APH provides a yield risk protection guarantee for the producer against shortfalls in yield as a function of a proven yields and its selected level, which trigger the indemnity payment (Coble et al. 2010).

The ratemaking procedure deals strictly with the derivation of the expected loss component, represented by the Loss Cost Ratio (LCR).[29] RMA establishes rates for every crop separately and at any level of coverage, expected losses are aggregated geographically for a group of similar risks, typically by county. Furthermore, there are other tailoring criteria for adjusting the rate to an individual producer. Basically the ratemaking procedure has five steps: one, adjusting the loss (indemnity) and exposure (liability) to a common coverage level; two, derivation of county unloaded rates; third, base rate loading; fourth, capping rate changes; five, updating of practices and group factors (Milliman and Robertson Inc. 2000).

[29] LCR measures the loss per unit of exposure, which is obtained by dividing indemnity by liability.

When historical losses are available, rating methodologies are based on historical loss costs ratios. Historical indemnity and liability are used to evaluate the premium rate through the LCR. The incorporation of this information guarantees that moral hazard or changes in production practices as a famer reacts to insurance are reflected in the expected value of future costs.

When historical data is not available like in the case of a new insurance scheme like Camelina, the ratemaking procedure relies on the yield data available and on simulated losses (Goodwin and Mahul 2004). In this context simulation can be helpful for rating this scheme because the observed experience could not reflect either the full range of potential outcomes or the current distribution of exposures (Coble et al. 2010).

This rating method could have the required flexibility to reflect the heterogeneous risk characteristics, such as Coble et al. (2010) recommends. Simulation based rating is able to consider the effects on yield of: different variety on soils, weather, topography and cultural practices within a country.

Unfortunately there are two weaknesses in the yield simulation based approach for insurance ratemaking purposes. First, insurance parameters such as premium rates are sensitive to the assumptions made in modeling yield distributions. Second, this procedure does not capture the impact of insurance on farming practices. The yields used to build the insurance parameters are not from insured farmers, so it could reflect the behavior of the uninsured farmer, which naturally affects the perception and how he copes with risk. In contrast historical indemnity data reflect the impacts of farming practices on yields.

Coble et al. (2010) affirm that loss experience base rating has an advantage with respect to a yield simulation based approach because crop insurance indemnifies losses that are not normally reflected in planted acre yields mainly from three sources: preventing planting provisions, replant payment provisions, and quantifying quality loss. A valid insurance rating system requires a procedure to evaluate the convergence of the rates and the observed experience (Coble et al. 2010).

Yield Modeling and Rating of a New Insurance Contract

The Camelina APH insurance schemes developed in this research use as a base the simulation base rating and, by construction, embraces mainly three implicit assumptions. First, yields stochasticity is entirely originated in the WinEPIC model by weather. Second, yields simulation is produced under the same technological conditions, no technological improvement is considered. Third, patterns of physiological development in the crop, erosion in soils and the general conditions of the environment are considered without alteration and climate change issues are not considered.

By its nature, crop yield risk is mainly driven by climate. Thus, crop yield loss events and consequently yield risk are determined by the extreme but infrequent events. So, the analysis of the APH insurance schemes for Camelina capture the effect of having a more accurate replication of the extreme weather patterns observed.

Crop yields are the result of the interactions among several factors mostly related to environmental conditions, which implies that the evaluation of their risk probabilities is determined by the accurate assessment of the probability of these events (Coble et al. 2010). Thus, simulated yields provide details on the frequency of occurrence of extreme events such as floods, droughts, etc. at the same time that reflect the characteristics of the present production systems and cultural practices. In contrast, the observed historical yield series could reflect loss experiences from crop production systems and technologies no longer in use, which might create distortions in the effective risk valuation.

In a new insurance scheme the rating procedure heavily relies on yield data available and on simulated losses. Thus, Camelina indemnity payments can be simulated from the yield data (Goodwin and Mahul 2004). In this context, the weather generators have important implications for yield probability distribution functions; in particular regarding the tails of the probability distribution. Thus, weather generator yield distribution will directly impact expected insurance payouts and the premium rates derived from the estimated yield densities.

For the sake of simplicity, assume a yield insurance contract at a predetermined fixed price that pays indemnities if the actual yields fall below some threshold defined by the guarantee (liability). In this scheme, the two parameters are influenced by the underlying yield distribution. Yield guarantee determines the total liability or the maximum possible indemnity paid in the event of total loss and it reflects the expected yield and establishes the conditions in which the indemnity disbursements are paid. The premium or price of the insurance reflects the likelihood and the expected level of loss that corresponds to the coverage level specified in the contract. An actuarially fair premium equals the expected insurance loss (expected indemnities). The premium rate is expressed as the ratio of expected loss to total liability, it means the dollars paid in premium for each dollar of liability.

Assuming that there is an adequate representation of the yield density $f(y)$, the contract with level of coverage λ, and the expected insurance yield μ.

$$Expected\ Yield = \mu = E(y) = y \int_{-\infty}^{\infty} f(y)\, dy \tag{76}$$

The Expected Insured Loss (EIL) is the product of the probability of a loss times the expected loss, given that a loss occurs.

$$EIL(y) = E\{ \max(\lambda\mu - y, 0) \} = \tag{77}$$

$$= Prob(y < \lambda\mu)[\lambda\mu - E(y|\, y < \lambda\mu)] \tag{78}$$

$$= \int_{-\infty}^{\lambda\mu} f(y)\, dy \left\{ \lambda\mu - \frac{\int_{-\infty}^{\lambda\mu} y\, f(y) dy}{\int_{-\infty}^{\lambda\mu} f(y) dy} \right\} \tag{79}$$

$$Actuarially\ Fair\ Premium = Expected\ Indemnity = EIL \tag{80}$$

In this simple scheme, the insurance premium is the expected LCR. The reserve load is the cost of reserves the insurer must set to pay unexpected losses with a determined degree of confidence (Goodwin and Mahul 2004).

In general terms, crop yields are negatively skewed because they show more frequent yields near the maximum than yields near the minimum. Because insurance issues occur toward the tail of the distribution rather than near its median, parameter and model error in the estimation of the yield distribution are compounded when the yield distribution is translated into insured losses (Goodwin and Mahul 2004). Thus, the selection of the appropriate specification of the yield distribution is primary and no consensus exists because observed yield distributions tend to be inconsistent across region, production potential and soils.

Modeling of the yield probability distribution includes basically two approaches: parametric and non-parametric. Parametric methods use available data and maximum likelihood method or moment estimation procedures to obtain the parameters of the distribution. Under this approach the fundamental assumption is that the true distribution of the data is a priori known (Goodwin and Mahul 2004).

Usually beta, Weibull, gamma log-normal and normal distributions are considered for modeling crop yield distributions. However, a normal distribution could not be suitable because the systemic risk (covariate component) violates the assumptions of the central limit theorem about i.i.d. Yields reflect agricultural risk which is composed of systemic and non-systematic risk. Systemic risk is caused by weather, pest, or natural phenomena that uniformly affect entire geographical areas (Goodwin and Mahul 2004).

The alternative approach to model the yield distribution is the non-parametric distribution; however, the weakness of this approach is the bin width and placements of bins. The non-parametric kernel approach provides additional flexibility because it imposes a minimal structure on the estimated distribution. The only requirement is enough observations to estimate reliable probabilistic estimates.

There exists a tradeoff between efficiency and bias in the selection between parametric versus non-parametric probability distributions. The incorrect selection of a parametric distribution can create bias in estimates of the distributions which results in inaccurate insurance premiums rates. However, when the parametric distribution is known the resulting estimates are the most efficient because they attain the Crámer-Rao lower bound.

So, the best strategy consists in estimating the yield probability distribution of the parametric type when the distribution is a priory known. As a consequence, non-parametric distribution only could represent an improvement in distribution modeling under two conditions: when there is enough data to attain a reliable estimate and the functional form of the distribution of the observed data is unknown.

Camelina Yield Distribution

By construction Camelina simulated yields are not subject to technological change or differentiated cultural practices. Instead, they are generated by the same data-generating process in the WinEPIC. As expected, trend is not significant for both Camelina yield series, see table 12, in appendix B.

Camelina simulated yields from both SWGs have important differences in terms of standard deviation. Table 13, in appendix B, shows higher standard deviations on simulated data by the copula based SWG which could reflect the effect of a more accurate description of the observed data.

Yields were modeled using mainly three parametric distributions beta, gamma and Weibull. The selection criterion for the probability distribution relays on the best representation of the left tail of the yield probability distribution. The probabilities associated with the left tail of the distribution mostly determine the expected insured loss. Thus, based on this criterion and the AIC and BIC criterion showed in table 14, in appendix B, the Weibull probability distribution attained the best fit for the Camelina yields in both cases, when Copula based SWG is used and also when Richardson SWG is applied.

Figures 8 and 9, in appendix A, show that Weibull distribution attains a better fit for the Camelina yield simulated using Richardson SWG and for the yield generated using the Copula Based SWG.

Also, the yield simulated using the Copula based SWG has fatter tails (kurtosis 3.85), and it is more positive skewed (1.26). In contrast, distribution generated by the yields from Richardson SWG are less positive skewness (0.90) and has a more peakedness distribution with kurtosis of 4.37, bigger than yields generated by the Copula based SWG.

APH Insurance Scheme for Camelina

In the insurance scheme the guarantee considered equals the APH multiplied by the selected coverage level. Following RMA (2011), additional considerations related to loss adjustment procedures or insured causes of loss for Camelina are those applicable to small grains like canola and rapeseed because growth stages and losses are similar.

The comparison of the yields series generated by the two SWG is made in terms of the insurance results. Seven yield insurance policies entail the different percentages of the coverage of the APH (50%, 60%, 65%, 70%, 75%, 80% and 85%). The unloaded premiums were calculated under three different parametric probability distributions and the non-parametric kernel smoothing. Insurance analysis is done under the assumption that the APH yield equals the average simulated yield and the consideration that all risk contained in these series is entirely generated by the stochastic weather.

The examination of the unloaded fair premiums for both Camelina yields shows substantial differences.[30] Table 15, in appendix B, shows that the premiums from the yields generated using Copula based SWG are significantly higher than premiums from almost two fold when the coverage increases at 70% and higher. The premium for the yield simulated with yieldcop is $7.92 per acre in comparison with $2.59 per acre for the yieldRich. At 70% coverage level the unloaded premium for the yieldcop is $20.3 per acre, while for the yieldRich the unloaded premium is $9.06 per acre. At the 85% of coverage the unloaded premium is $34.55 per acre for yieldcop, while for yieldRich is $18.19 per acre. The fully loaded base premium at 85% APH coverage is $56.71 per acre for the yieldcop, while for the yieldRich is $29.86 per acre.

Following RMA (2011), the fully loaded premium is calculated by dividing the unloaded fair premium by 0.90 that corresponds to the unit division load factor and then again dividing by the 0.88 that corresponds to the Federal Crop Insurance Corporation (FCIC) disaster reserve factor and finally multiplying by 1.3 which is the qualitative load factor that adjust for taking in account the additional risk different from climate on the regression equation for physical relationships and production functions in the WinEPIC model (RMA 2011). The loss cost ratio for all the insurance policies is the ratio of the expected loss and the liability.

The difference between the APH insurance schemes generated by the two weather simulators is significant. The loss cost for the yield generate using copula

[30] The amounts mentioned here consider the Weibull probability distributions for both Camelina yield series.

approach at an 85% level of coverage attains the 28.02%, while the loss cost for the yield generated by Richardson approach at the same coverage level is 15%.

These differences indicate the underestimation of the agricultural risks in the ratemaking process of a new insurance scheme based on the use of SWG that does not accurately reproduce extreme weather event patterns. In particular, differences in insurance premiums can be attributed to probability distributions with fatter tails, where the extreme weather events are reflected.

Summary

The results of this analysis strength the evidence of the RMA (2011) feasibility study for the development of an APH insurance scheme for Camelina in conditions of farming contract with prices pre-established.

A new insurance scheme with no historical data available requires, for the ratemaking process, the yield generation from a crop growth model such as the WinEPIC. The generated yields are proxies of the real Camelina yields for specific locations and under particular production practices; in consequence it is possible to obtain tailored approximations of unsubsidiazed unloaded and loaded fair premium estimates.

Under this approach, weather is the only source of uncertainty in the crop growth model. This condition allowed a deeper analysis of the use of the SWG in crop growth models for yield simulation. In particular, the use of the Copula based SWG, which reproduces more accurately the extreme weather events, showed important differences in the generated yields and the APH insurance schemes.

A comparative analysis found evidence of the underestimation of risks on Camelina yields when Richardson's SWG is used. The copula based SWG generated more positively skewed and fatter tails in Camelina yield distribution than the yields distribution generated by the Richardson SWG. As a consequence, when Copula based SWG is used in the WinEPIC, the generated yields reflect higher premiums as a result of more risk from more frequent extreme weather.

Furthermore, another advantage in the use of crop growth models in the ratemaking process of new insurance schemes reside is the possibility of include additional heterogeneity through the incorporation of differentiated site characteristics involved in production, such as soils, topography, cultural practices and weather. Also, the simulation of yields could be an effective tool to incorporate the analysis of the effects of climate change in insurance policies.

CHAPTER V

SUMMARY, CONCLUSIONS AND FUTURE RESEARCH

Stochastic Weather Generators represent a key issue in the generation of weather series for multiple applications in agriculture, such as in crop growth simulation models like the WinEPIC. However, conventional SWGs have many shortcomings. Climate variables are complex and characterized by non-normal probability density functions, while usually SWG assumes parametric probability distributions that are not close enough to the observed data. Also, SWGs assume linear dependence between variables which lead to the simulation of inaccurate climatic variables. Likewise, SWGs tend to under-estimate the occurrence of extreme weather events from observed data.

The main objective of this research is to design a SWG based on copula methodology that more accurately models the nonlinear dependence structure and the occurrence of extreme events between precipitation, maximum temperature, and minimum temperature. An additional objective of this research is to provide a clear insight of the advantages of the use of this SWG in the crop growth models for its applicability in insurance.

The idea of modeling climatic variables using copula methods relies on the behavior and structure of these variables. The copula modeling of the weather variables depicts their joint probability distribution considering their dependence patterns, which are far from linear dependence. The dimensionality problem in copula estimation was solved by the selection of 12 dates with the highest average monthly anomalies. Thus, the weather series simulated by copula methods are the bordering conditions of the weather stochastic simulator, while the Brownian Bridge uses Monte Carlo methods to replicate the daily dynamic of weather variables evolving on a path forward through time.

Copula methods provide the flexibility to model dependence structures between random variables independent of the marginal distributions involved. The selection of the marginal distributions was between the normal distribution and the non-parametric kernel smoothing specification. Although the copula based SWG can incorporate a hybrid specification on copula families – that describe different dependence patterns – and numerous specifications were tested, the final specification was the one-parameter Gumbel family.

Statistical tests on simulated weather showed that Copula based SWG is able to represent the main features for the distributions of the observed weather variables. Although the comparative analysis of the copula based SWG versus Richardson SWG did not provide a clear insight about which SWG has the best performance in terms of their simulations, there is evidence that indicates the copula based SWG has better performance in the reproduction of the extreme weather events.

For a comprehensive evaluation on the SWG, this research considers a comparative analysis of two SWGs in the generation of yields using the WinEPIC.[31] The simulated yields are proxies of real Camelina yields for specific locations and under particular production practices. Assuming that the only source of uncertainty in yield

[31] It refers a comparison of the yield generated by the copula based SWG versus the yields generated by Richardson SWG.

simulation is the weather, insurance estimation provides some insights of the advantages for the use of a copula based SWG in the generation of yields and the risk modeling of insurance policies where no historical yield data is available.

Non-parametric statistical tests were applied to the simulated series of climate variables because the true probability distribution for these weather variables is not known. These tests showed that copula based SWG had an acceptable replication of the observed weather patterns. In particular, the SWG simulated series showed an accurate reproduction of the extreme weather event patterns. Although in general there is no a conclusive evidence about if the copula based SWG has a better performance than Richardson's SWG, in terms of extreme weather events the reproduction of the simulations derived from the copula based SWG showed to be more accurate.

In terms of the yields generated and insurance analysis, there is evidence of significant differences in the yields generated by the two SWGs. The copula based SWG generated Camelina yields with a distribution that was more positively skewed and with fatter tails. Furthermore, there is evidence of underestimation of risks on Camelina yields derived from Richardson's SWG. As a consequence, when Copula based SWG is used in the WinEPIC, the yields generated reflect higher crop insurance premiums as a result of greater risk from more frequent extreme weather.

This research can be extended in multiple ways:

- More advances can be attained in terms of the copula specification to more accurately capture the dependence patterns between weather variables.

- Multiple criteria for the selection of the dates estimated by copula methods can be development for the SWG.
- Higher climate variability could be incorporated into the SWG by the incorporation of changes in means and variances in Brownian Motion that resembles the daily changes in the climate variables.
- Climate change analysis could be included by adding changes in parameters for the Brownian Motion by decades, emulating recent research which has focused on the study of the patterns of decadal variability in precipitation and temperature.
- There are additional advantages in the use of crop growth models in the ratemaking process for new insurance schemes.
    - The possibility of adding heterogeneity into the yields simulation through the incorporation of differentiated site characteristics involved in production, such as soils, topography, cultural practices and weather.
    - Crop growth models for the yield simulation could be an effective tool to incorporate in the analysis the effects of climate change on crop insurance policy.

REFERENCES

Asmussen, S. and P. W. Glynn. 2007. *Stochastic Simulation: Algorithms and Analysis*. Series: Stochastic Modeling and Applied Probability. New York: Springer.

Brandimonte, P. 2006. *Numerical Methods in Finance and Economics. A MATLAB-Based Introduction*. 2nd ed. New York: John Wiley and Sons.

Canada, Saskatchewan Government. 2012. *Crop insurance. 2012. Terms and Conditions: Camelina.* Crop Insurance Corporation. http://www.saskcropinsurance.com/cropinsurance/terms/camelina.pdf (last accessed: 5/5/2012)

Cherubini, U., E. Luciano and W. Vecchiato. 2004. *Copula Methods in Finance*. West Sussex: John Wiley & Sons.

Coble, K. H., T. O.Knight, B. K. Goodwin, M. F. Miller and R.M. Rejesus. 2010. *A Comprehensive Review of the RMA APH and COMBO Rating Methodology*. Final Report. U.S. Risk Management Agency, March.

Cohen, A. C. 1991. *Truncated and Censored Samples: Theory and Applications*. Statistics, Textbooks and Monographs. New York: John Wiley and Sons.

Donald, M. 2008. *A Note on Rising Food Prices*. Policy Research Working Paper 4682, World Bank. Washington D.C.

Ehrensing, D.T. and S. O. Guy. 2008. "Camelina." Department of Agriculture, Oregon State University, Extension Service, Informative Bulletin Number EM-8953-E. http://extension.oregonstate.edu/catalog/pdf/em/em8953-e.pdf (last accessed: 5/6/2012)

Embrechts, P., F. Lindskog, A. McNeil. 2001. *Modeling Dependence with Copulas and Applications to Risk Management. Handbook of Heavy Tailed Distributions in Finance*, ed. S. Rachev, San Diego: Elsevier.

Favre, A.-C., S. El Adlouni, L. Perreault, N. Thi´emonge and B. Bobee. 2004. “Multivariate Hydrological Frequency Analysis Using Copulas.” *Water Resources Research* (40): W01101.

Genest, C. and A.-C Favre. 2007a. “Everything you Always Wanted to Know About Copula Modeling but Were Afraid to Ask.” *Journal of Hydrologic Engineering 1*2 (4): 347–368.

Genest, C., A. – C. Favre, J. B´eliveau and C. Jacques. 2007b. “Metaelliptical Copulas and their Use in Frequency Analysis of Multivariate Hydrological Data.” *Water Resources Research* 43(9): W09401.

Genest, C., B. R´emillardb and D. Beaudoinc. 2009. “Goodness-of-Fit Tests for Copulas: A Review and a Power Study”. *Insurance: Mathematics and Economics* (44): 199–213

Glasserman, P. 2010. *Monte Carlo Methods in Financial Engineering.* Applications of Mathematics. Stochastic Modeling and Applied Probability 53. New York: Springer.

Goodwin, B. K. and O. Mahul. 2004. *Risk Modeling Concepts Relating to the Design and Rating of Agricultural Insurance Contracts*. Policy Research Working Paper 3392. World Bank. Washington, D.C. August.

Huynh, H. T., V. S. Lai and I. Soumaré. 2008. *Stochastic Simulation and Applications in Finance with MATLAB® Programs*. West Sussex, England: The Wiley Finance Series.

Joe, H. 1997. *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Johnson, D. 2007. *Camelina.* Montana State University, Northwestern Agricultural Research Center. Harvesting Clean Energy: Conference VII, Creston, Montana. http://harvestcleanenergy.org/2010-conference-resources/2007-harvesting-clean-energy-conference/2007-conference-presentations/johnson-pdf (last accessed: 5/5/2012)

L'ecuyer, P. 2004. *Random Number Generation.* Working paper, D´epartement d'Informatique et de Recherche Op´erationnelle, University of Montreal. http://www.iro.umontreal.ca/~lecuyer (last accessed: 5/11/2011)

London, J. 2007. *Modeling Derivatives Applications in Matlab, C++, and Excel*. Sussex: Pearson Education Editors.

McVay, K. A. and P. Lamb. 2008. *Camelina Production in Montana. Report*. Montguide MT200701AG. Montana State University, Agricultural Experimental Station, Bulletin, May. http://msuextension.org/publications/AgandNaturalResources/MT200701AG.pdf (last accessed: 5/5/2012)

McVay, K. A. and Q. A. Khan. 2011. "Camelina Yield Response to Different Plant Populations under Dryland Conditions." *Agronomy Journal* 103(4):1265-1269.

Milliman and Robertson Inc. 2000. *Actuarial Documentation of Multiple Peril Crop Insurance Ratemaking Procedures*. Document prepared for USDA, Risk Management Agency.

Mood, A., F. A. Graybill and D.C. Boes. 1974. *Introduction to the Theory of Statistics.* 3rd ed. Singapore: McGraw Hill International Editions.

National Agriculture Statistics Service, (NASS), USDA.

http://quickstats.nass.usda.gov (last accessed: 5/5/2012)

Nelsen, R. 2006. *An Introduction to Copulas.* Series in Statistics. New York: Springer Verlag.

Richardson, C. W. 1981. "Stochastic Simulation of Daily Precipitation, Temperature and Solar Radiation." *Water Resources Research* (17): 182-190.

Risk Management Agency (RMA). 2011. *Feasibility Study for Insuring Dedicated Energy Crops*. *Final Submission*. U.S. RMA, Solicitation No. N10PC18199, May. http://www.rma.usda.gov/pubs/2011/energyfeasibility.pdf (last accessed: 5/6/2012)

Salvadori, G., C. Michele, N. T. Kottegoda and R. Rosso. 2007. *Extremes in Nature*. *An Approach Using Copulas*. New York: Springer.

Schölzel, C., and P. Friederichs. 2008. "Multivariate Non-normally Distributed Random Variables in Climate Research Introduction to the Copula Approach." *Nonlinear Processes in Geophysics* 15 (5): 761–772.

Semenov, M.A., R.J. Brooks, E.M. Barrow and C.W. Richardson. 1998. "Comparison of WGEN and LARS-WG Stochastic Weather Generators for Diverse Climates." *Climate Research* (10): 95-107.

Soltani, A. and T. R. Sinclair. 2012. *Modeling Physiology of Crop Development, Growth and Yield*. Massachusetts: CAB International Eds.

Turvey, C.G. 2005. "The Pricing of Degree-Day Weather Options." *Agricultural Finance Review* (65): 59-86.

Wand, M. C. and M. P. Jones. 1995. *Kernel Smoothing*. 1st ed. New York: Chapman and Hall Editors.

Wilks, D. 1990. “Maximum Likelihood Estimation for the Gamma Distribution Using Data Containing Zeros.” *Journal of Climate* (3): 1495-1501.

Wilks, D. 2011. *Statistical Methods in the Atmospherical Sciences*. International Geophysic Series; v. 100. 3rd ed. San Diego: Elsevier.

Wilks, D. and R. Wilby. 1999. “The Weather Generation Game: A Review of Stochastic Weather Models.” *Progress in Physical Geography* 23(3): 329-357.

Williams, J.R., C.A. Jones, and P.T. Dyke. 1984. “A Modeling Approach to Determining the Relationship between Erosion and Soil Productivity.” *Transactions of the ASAE* (27):129-144.

Williams, J. R. C.A. Jones, J. R. Kiniry and D.A. Spanel. 1989. “The EPIC Growth Model.” *Transactions of the ASAE*, 32(2):129-144.

Wong, G., M. Leonard, A. Metcalfe and M. Lambert. 2010. “Modeling Dependence Structure in Drought Simulations.” *Geophysical Research Abstracts* (12): 40-68. EGU2010-4068, 2010, EGU General Assembly. Vienna, Austria,

APPENDIX A

FIGURES

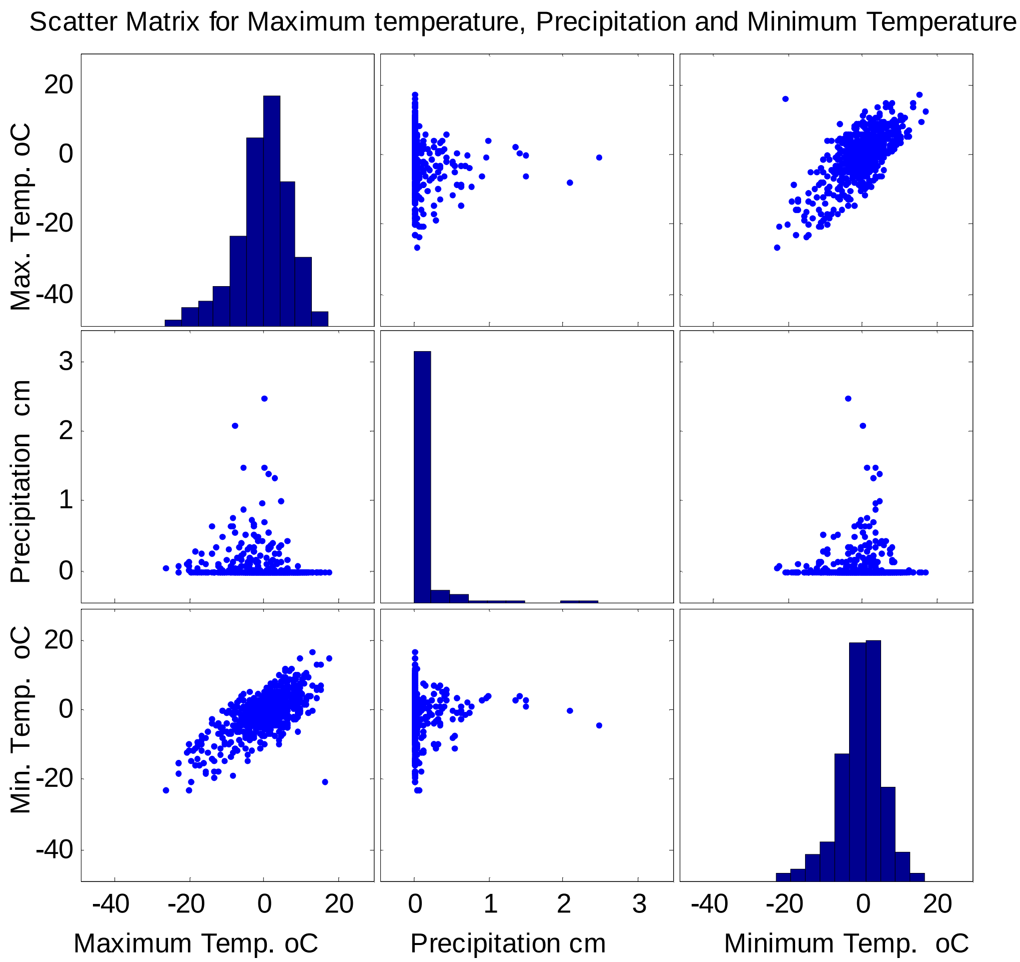


Figure 1. Bidimensional Scatter Matrix for Weather Variables, Conrad, Pondera County, Montana: Selected Dates for Estimation

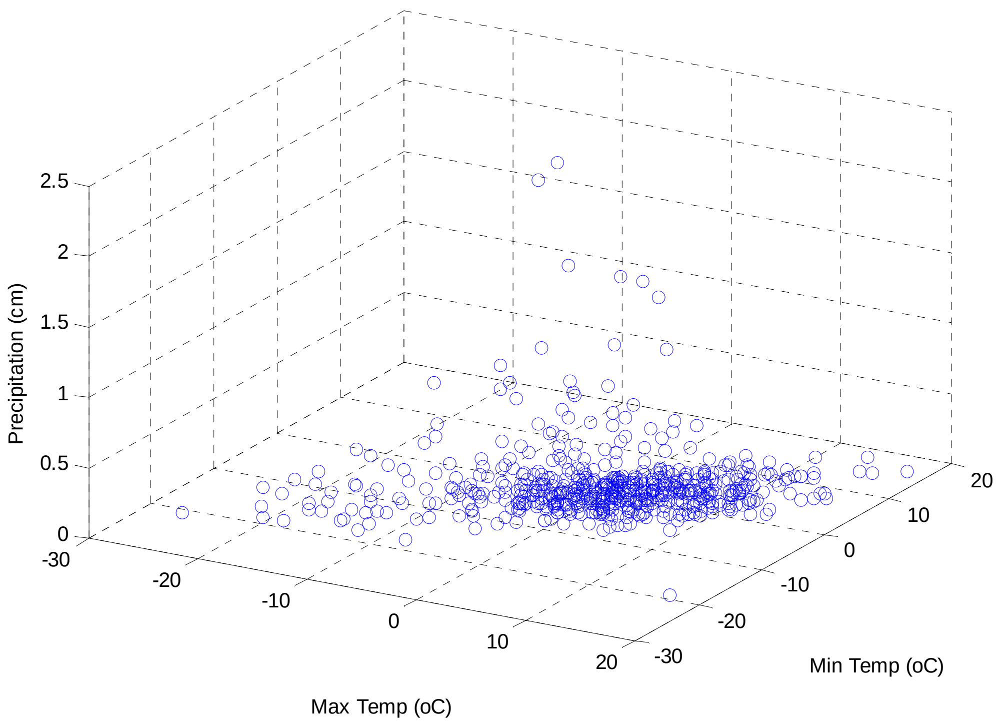


Figure 2. Tridimensional Scatter Plot for Climate Variables: Precipitation, Minimum Temperature and Maximum Temperature

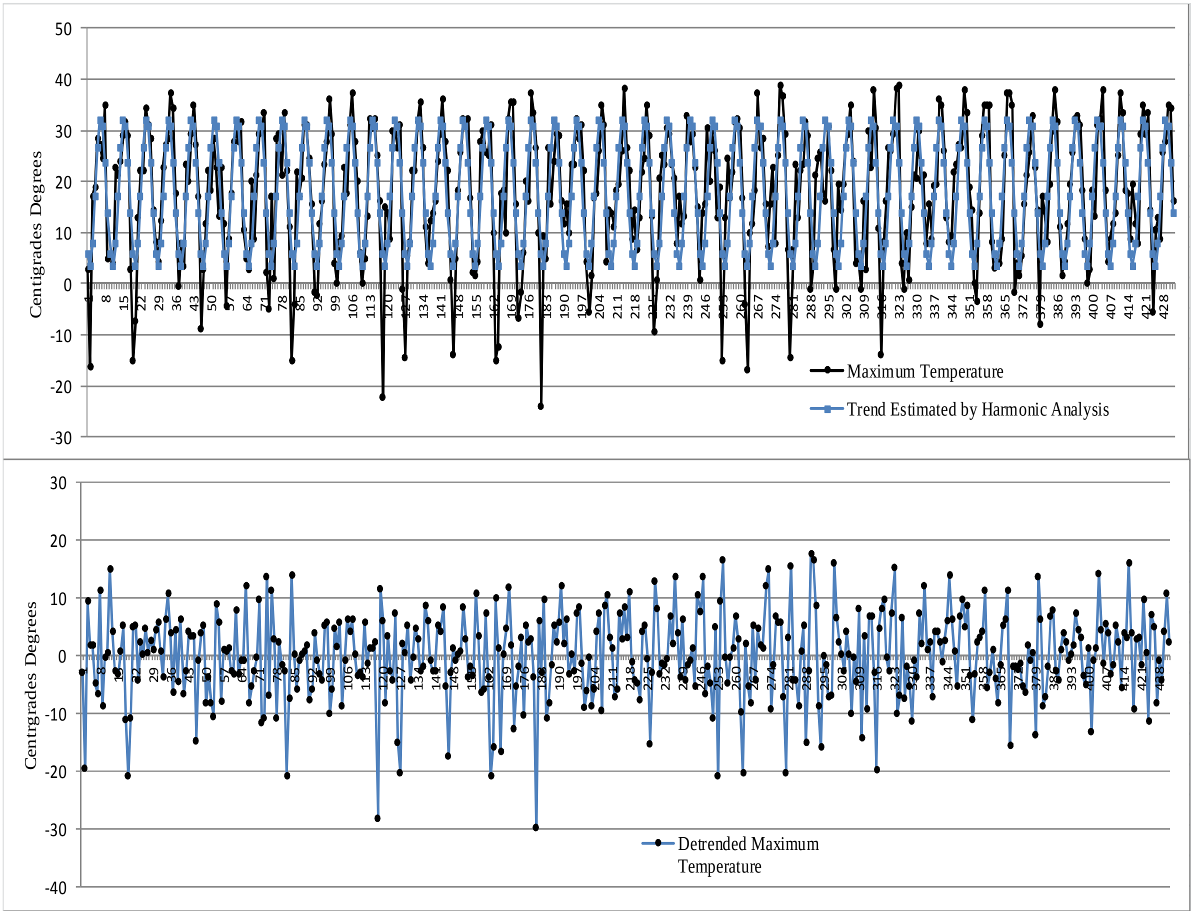


Figure 3. Detrend Technique Based in Harmonic Analysis Applied to Maximum Temperature Anomalies, Conrad-MT1974, Pondera County, Montana (1960-2010)

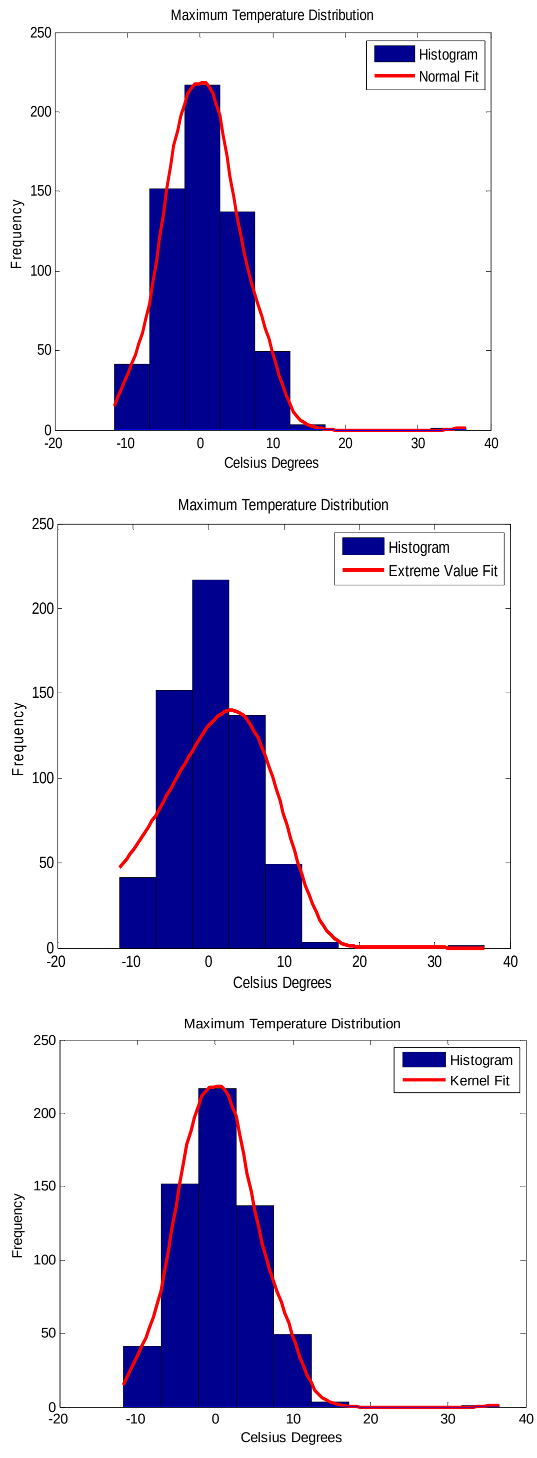


Figure 4.A. Probability Distribution Fit: Conrad, Pondera County, Montana: Maximum Temperature

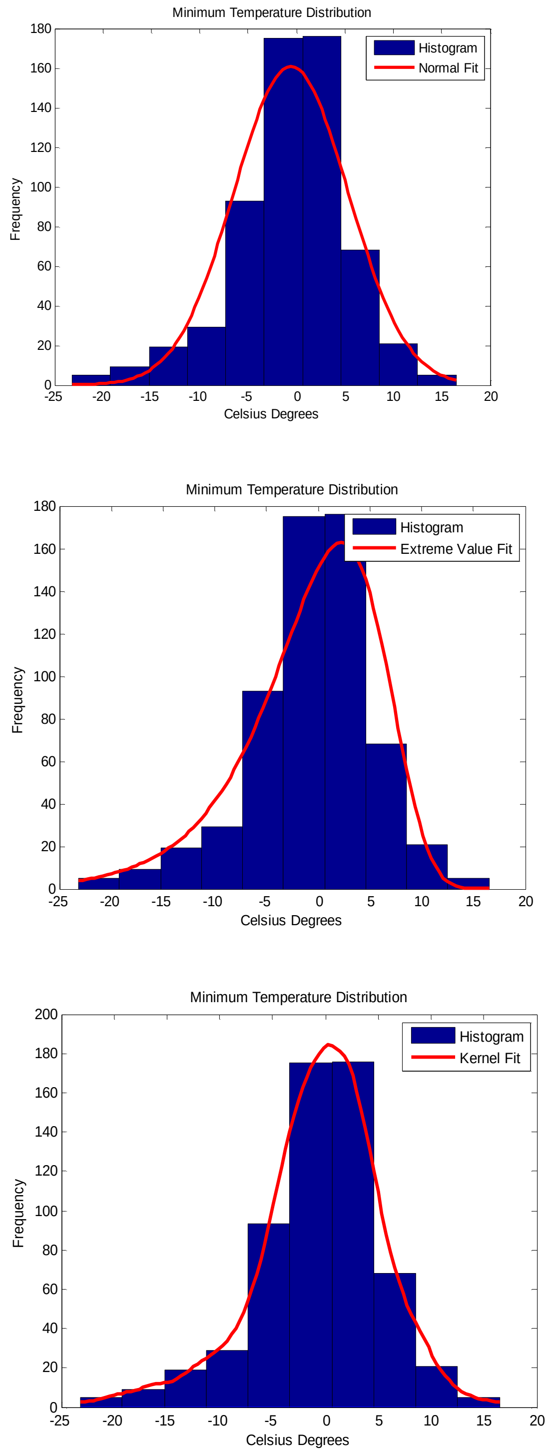


Figure 4.B. Probability Distribution Fit: Conrad, Pondera County, Montana: Minimum Temperature

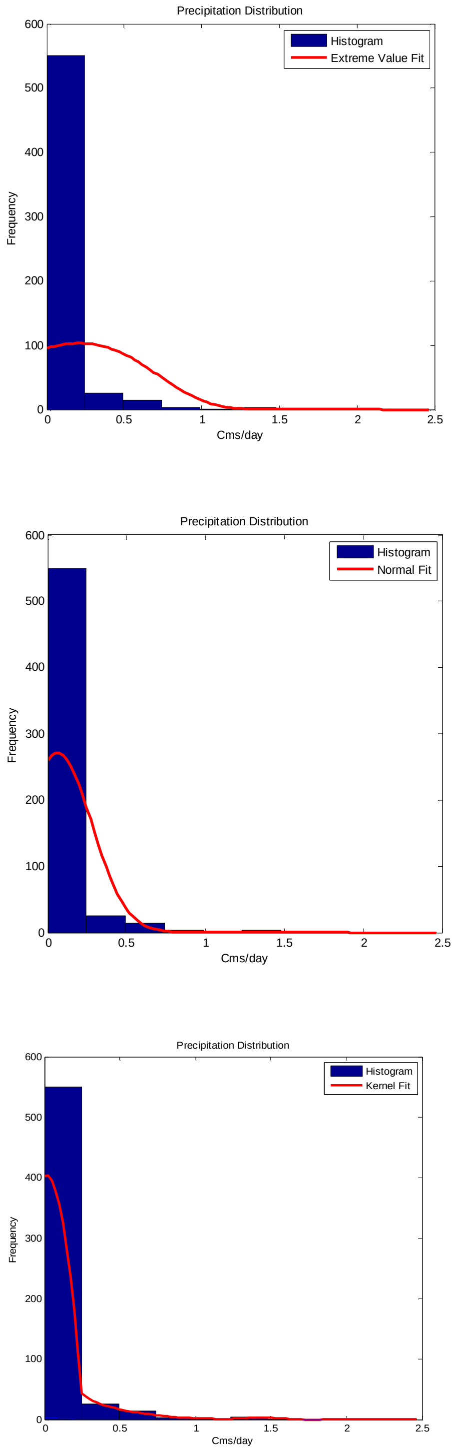


Figure 4.C. Probability Distribution Fit: Conrad, Pondera County, Montana, Precipitation

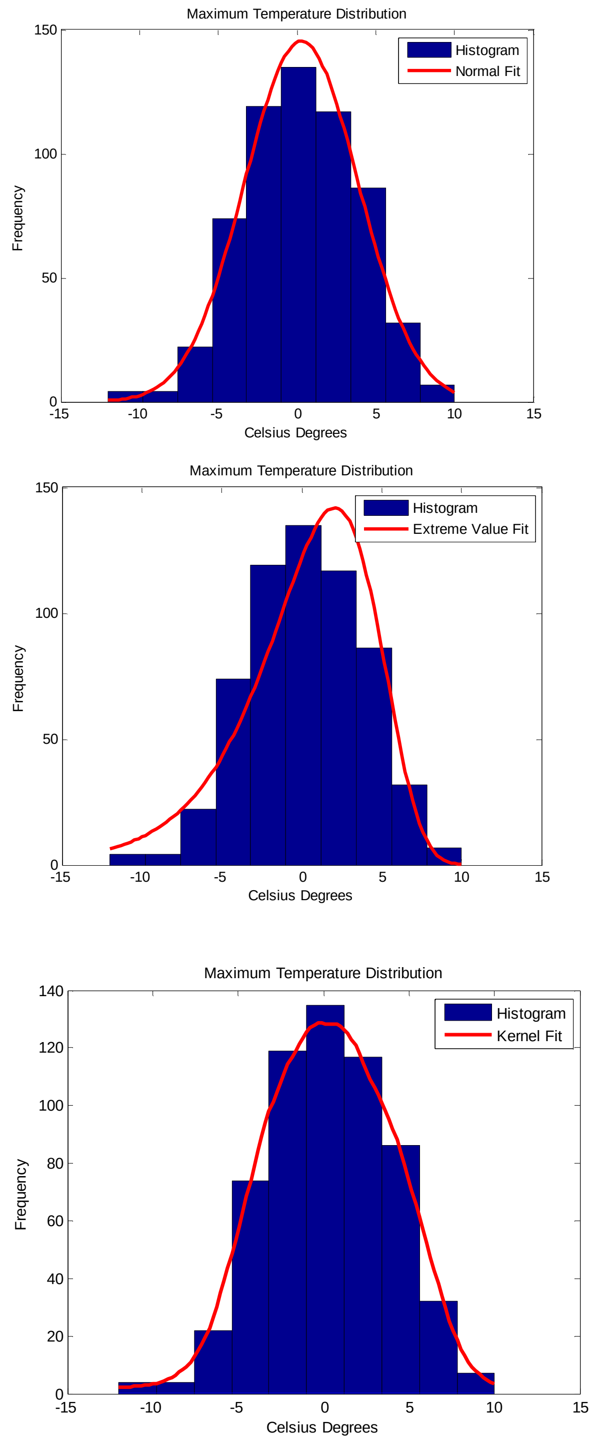
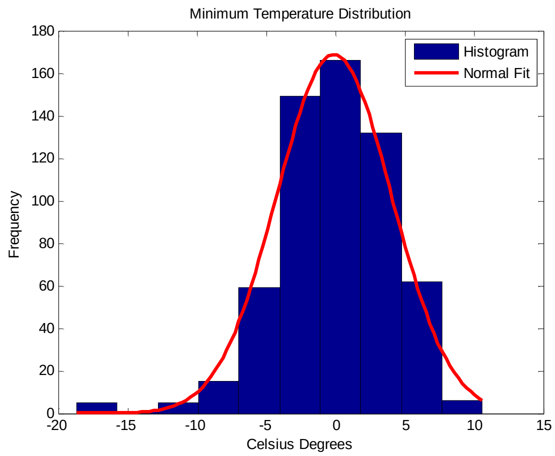
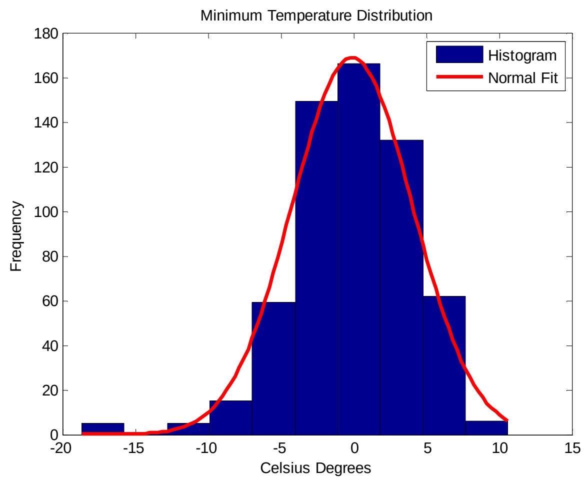


Figure 5.A. Probability Distribution Fit: Spokane, Spokane County, Washington Maximum Temperature





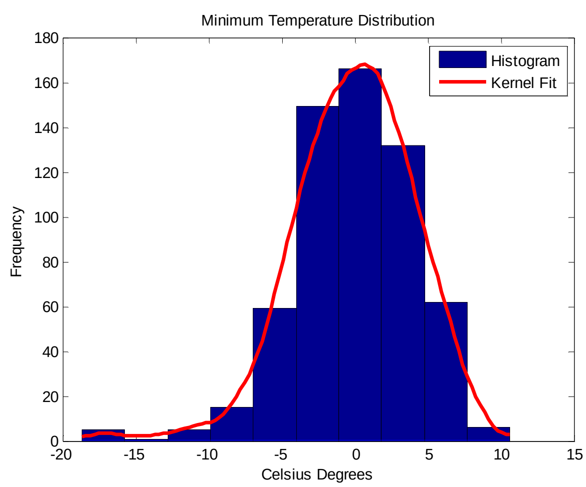


Figure 5.B. Probability Distribution Fit: Spokane, Spokane County, Washington Minimum Temperature

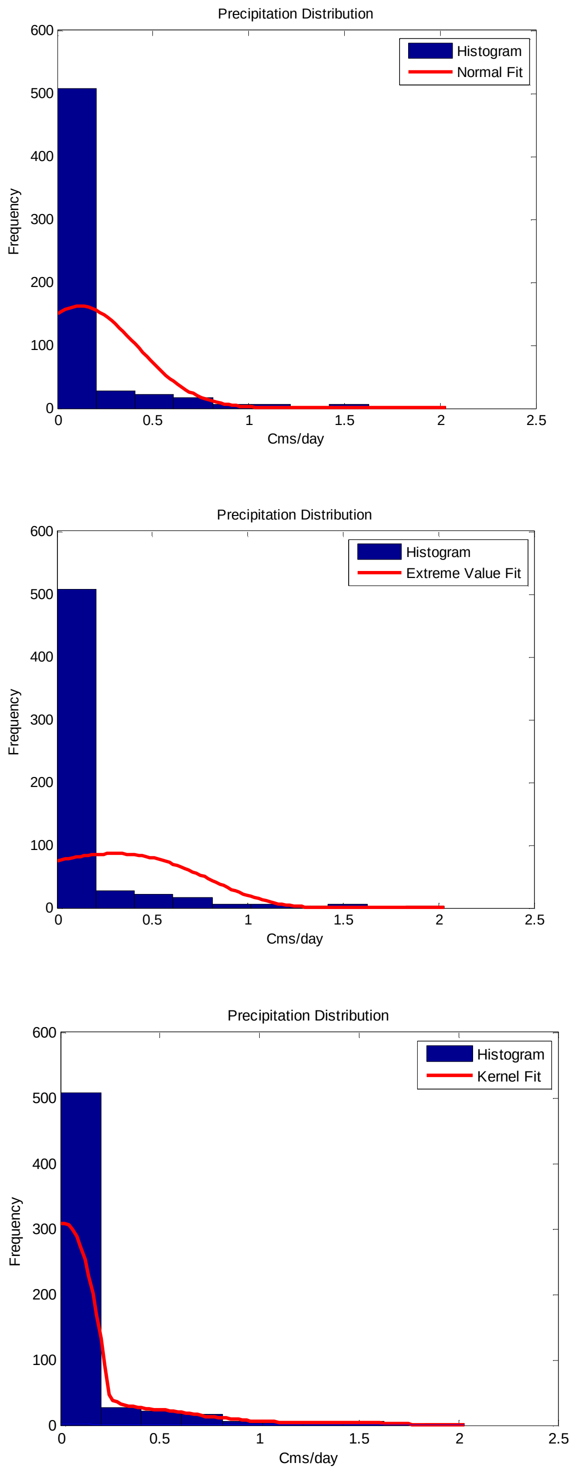
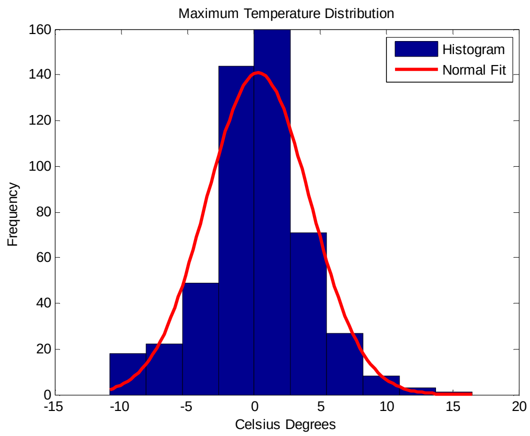
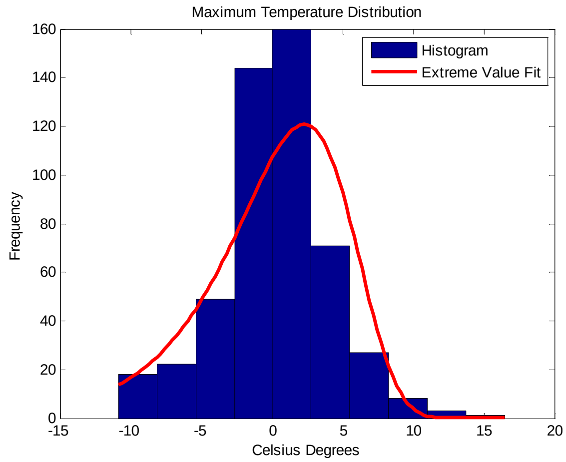


Figure 5.C. Probability Distribution Fit: Spokane, Spokane County, Washington Precipitation





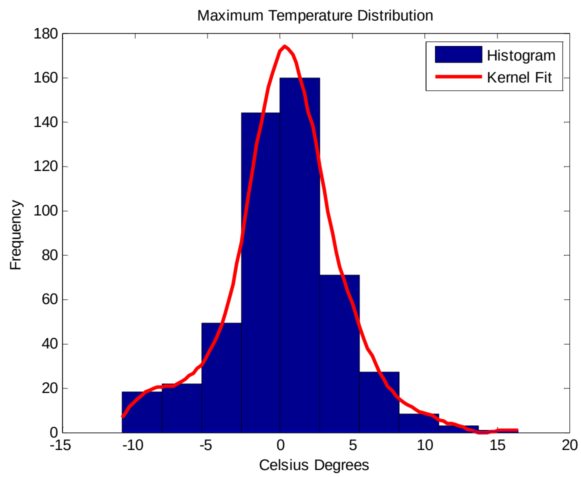
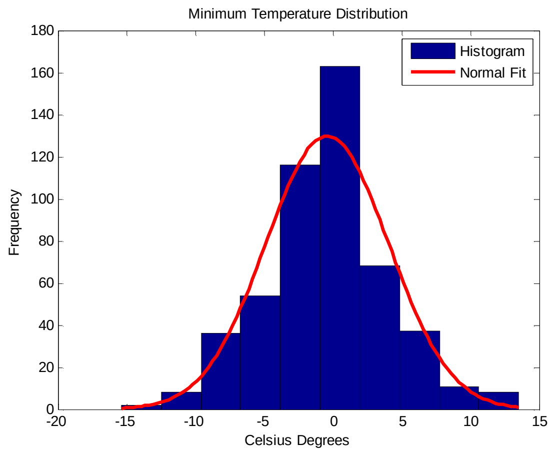
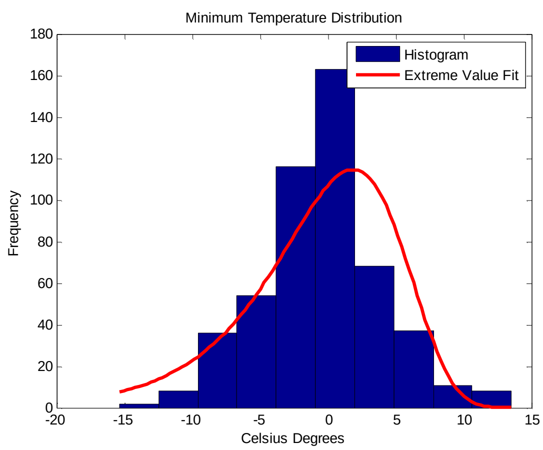


Figure 6.A. Probability Distribution Fit: Temple, Bell County, Texas Maximum Temperature





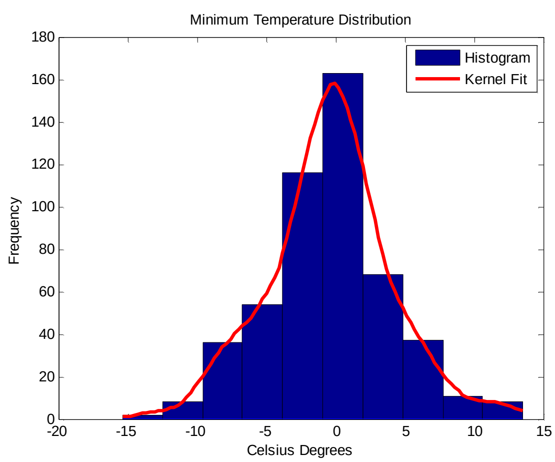


Figure 6.B. Probability Distribution Fit: Temple, Bell County, Texas Minimum Temperature

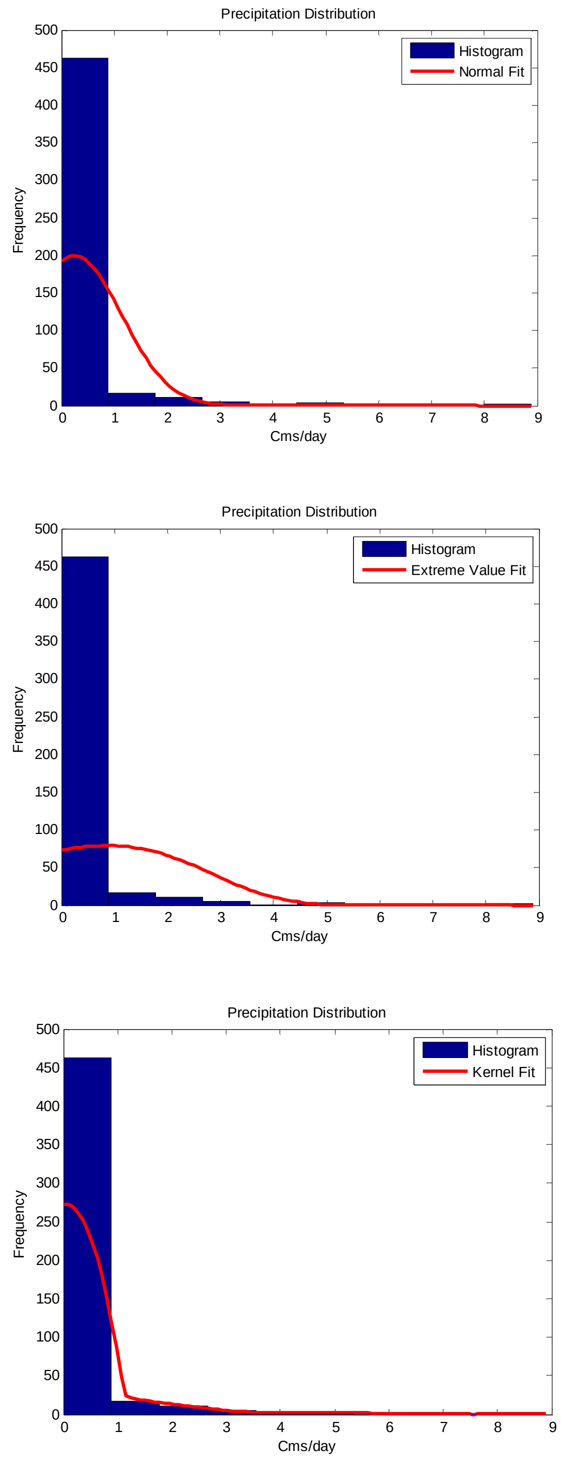


Figure 6.C. Probability Distribution Fit: Temple, Bell County, Texas Precipitation

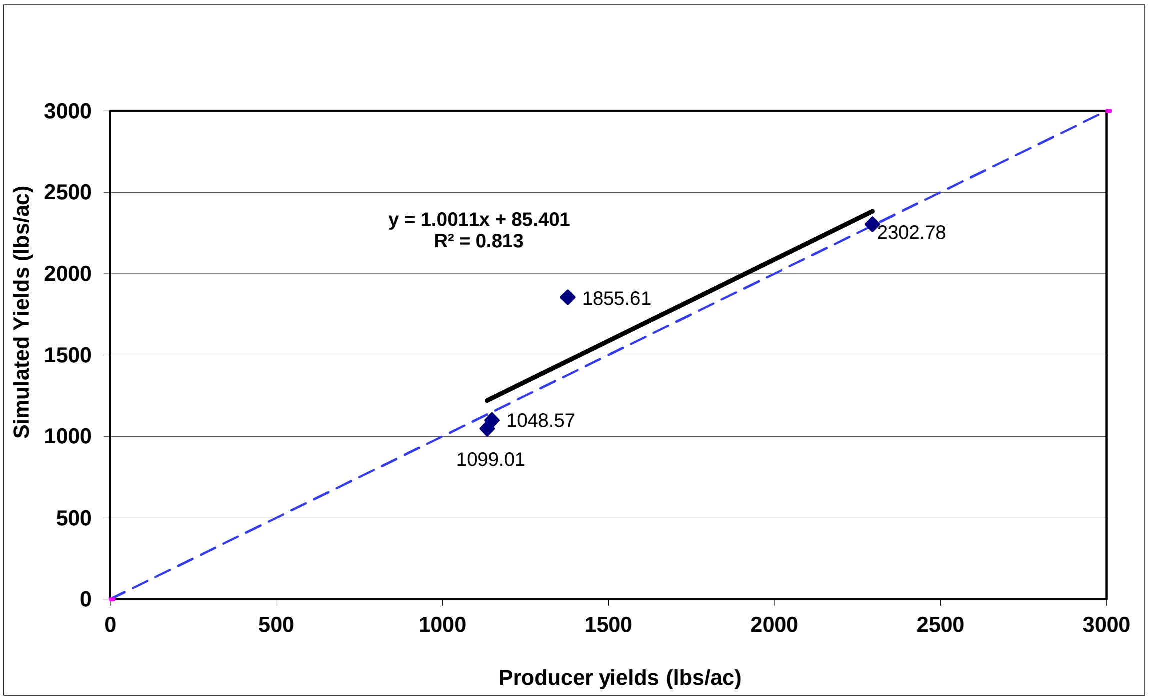


Figure 7. Camelina Yields Calibration: Comparison Producer vs. Simulated, Conrad, Pondera County, Montana

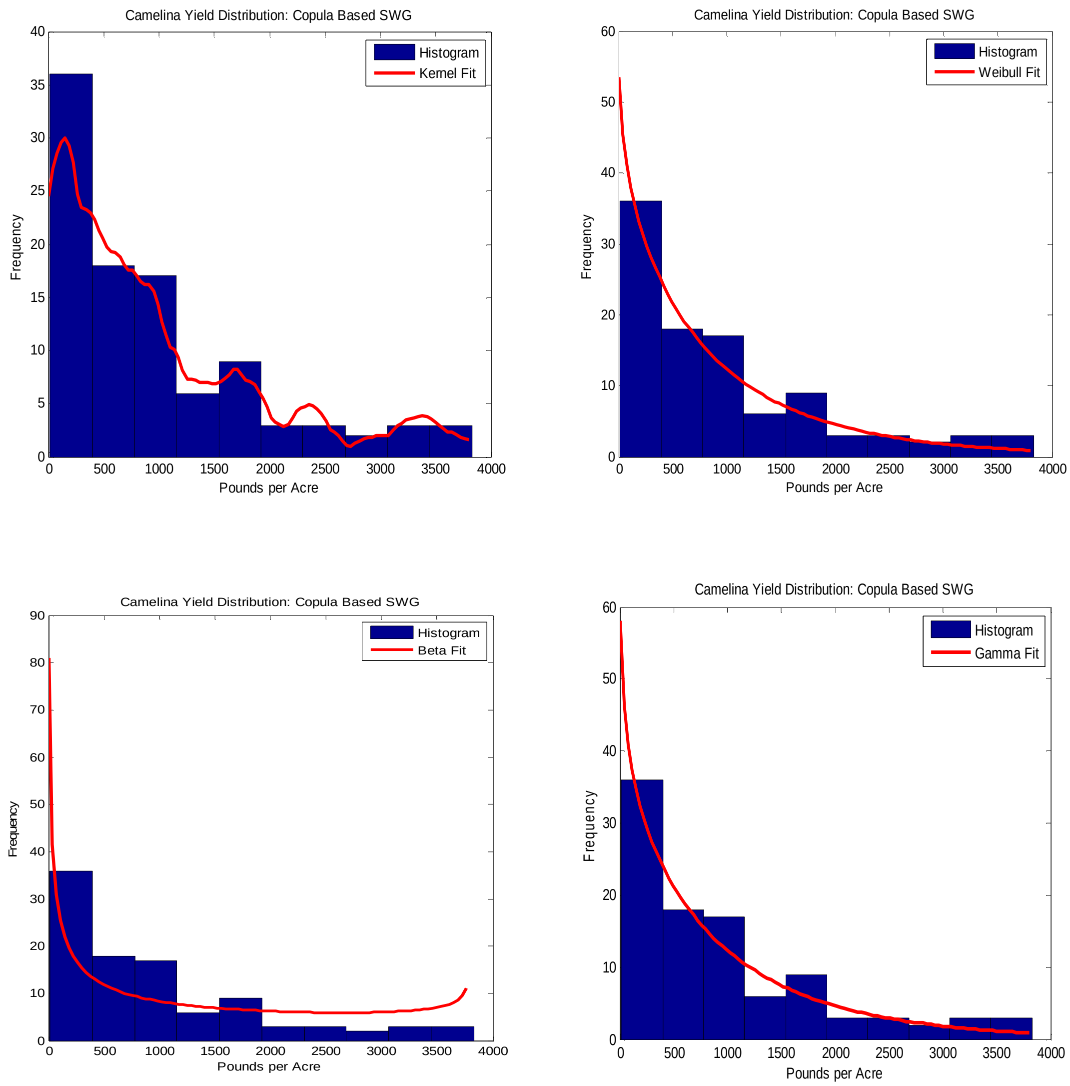


Figure 8. Camelina Simulated Yields Under Different Probability Distributions Copula Based SWG

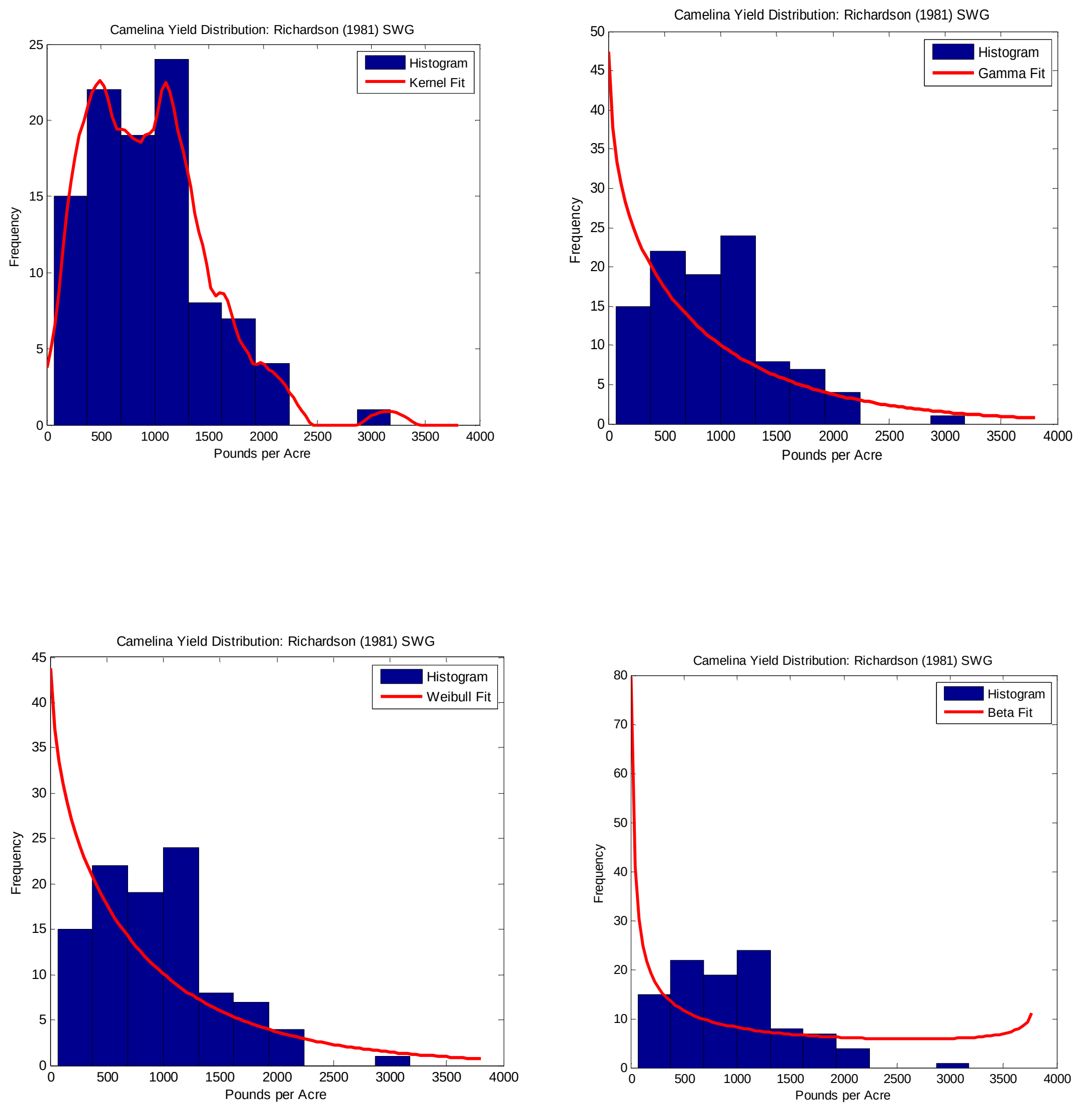


Figure 9. Camelina Simulated Yields Under Different Probability Distributions Richardson SWG

APPENDIX B

TABLES

Table 1.A. Detrending Regression for Conrad, Pondera County, Montana

Maximum Temperature				
Observations	18250.00			
F(2, 18250)	17992.50			
Prob > F	0.00			
R-squared	0.66			
Adj R-squared	0.66			
Variable	Coefficient	Std Error	t- Statistic	P>t
Cost	-13.04	0.07	-181.25	0.00
Sint	-4.02	0.07	-55.97	0.00
Const	14.02	0.05	275.55	0.00

Minimum Temperature				
Observations	18250.00			
F(2, 18250)	17788.95			
Prob > F	0.00			
R-squared	0.66			
Adj R-squared	0.66			
Variable	Coefficient	Std Error	t- Statistic	P>t
Cost	-11.10	0.06	-181.68	0.00
Sint	-3.10	0.06	-50.70	0.00
Const	-1.49	0.04	-34.60	0.00

Table 1.B. Detrending Regression for Spokane, Spokane County, Washington

Maximum Temperature				
Observations	18250.00			
F(2, 18250)	36107.00			
Prob > F	0.00			
R-squared	0.80			
Adj R-squared	0.80			
Variable	Coefficient	Std Error	t- Statistic	P>t
Cost	-13.37	0.05	-258.81	0.00
Sint	-3.79	0.05	-72.34	0.00
Const	14.23	0.04	389.23	0.00

Minimum Temperature				
Observations	18250.00			
F(2, 18250)	21437.75			
Prob > F	0.00			
R-squared	0.70			
Adj R-squared	0.70			
Variable	Coefficient	Std Error	t- Statistic	P>t
Cost	-8.66	0.47	-198.06	0.00
Sint	-2.64	0.47	-60.39	0.00
Const	3.05	0.03	98.59	0.00

Table 1.C. Detrending Regression for Temple, Bell County, Texas

Maximum Temperature				
Observations	15330.00			
F(2, 15330)	16884.71			
Prob > F	0.00			
R-squared	0.69			
Adj R-squared	0.69			
Variable	Coefficient	Std Error	t- Statistic	P>t
Cost	-9.91	0.06	-174.56	0.00
Sint	-3.26	0.06	-57.43	0.00
Const	25.18	0.04	627.19	0.00

Minimum Temperature				
Observations	15330.00			
F(2, 15330)	22778.32			
Prob > F	0.00			
R-squared	0.75			
Adj R-squared	0.75			
Variable	Coefficient	Std Error	t- Statistic	P>t
Cost	-9.95	0.05	-203.89	0.00
Sint	-3.08	0.05	-63.12	0.00
Const	12.82	0.34	371.54	0.00

Table 2. Parametric Distributions Fit for Weather Variables, Three Weather Stations

Distribution	α	β	- Σ log L	AIC	BIC
	Conrad MT1974, Pondera County, Montana				
	Normal				
Maximum Temperature	0.40	5.16	1,835.79	-3,667.58	-3,658.79
Precipitation	0.06	0.22	-64.48	132.95	141.75
Minimum Temperature	-0.57	5.90	1,916.06	-3,828.12	-3,819.33
	Extremevalue (Gumbel)				
Maximum Temperature	3.05	7.64	2,028.13	99,999.00	99,999.00
Precipitation	0.21	0.53	387.32	-770.64	-761.84
Minimum Temperature	2.22	5.38	1,920.69	3,986.14	3,994.93
	Exponential				
Precipitation	0.06		-1,071.89	2,145.79	2,150.18
	Spokane WA, Spokane County, Washington				
	Normal				
Maximum Temperature	0.29	3.62	1,622.86	-3,241.72	-3,232.93
Precipitation	0.12	0.30	128.04	-252.08	-243.29
Minimum Temperature	-0.09	4.16	1,705.92	-3,407.85	-3,399.05
	Extremevalue (Gumbel)				
Maximum Temperature	2.08	3.42	1,651.93	3,467.29	3,476.08
Precipitation	0.30	0.52	421.92	-839.85	-831.05
Minimum Temperature	1.88	3.68	1,703.00	3,551.31	3.56E+03
	Exponential				
Precipitation	0.12		-685.35	1,372.70	1,377.10
	Temple TX, Bell County, Texas				
	Normal				
Maximum Temperature	0.36	3.90	1,398.35	-2,792.71	-2,784.27
Precipitation	0.24	0.89	656.70	-2,928.01	-1,300.95
Minimum Temperature	-0.40	4.47	1,466.00	-2,928.01	-2,919.56
	Extremevalue (Gumbel)				
Maximum Temperature	2.30	4.20	1,456.83	3,102.95	3,111.39
Precipitation	0.86	2.08	1,019.89	-2,035.78	-2,027.34
Minimum Temperature	1.84	4.66	1,519.48	3,213.06	3,221.50
	Exponential				
Precipitation	0.24		-205.38	412.76	416.98

Table 3. Copula Mixture Estimation for Climatic Variables by Weather Stations

Copula Mixture	Param. 1	Param. 2	Param. 3	- Σ log L	AIC	BIC
	Conrad MT1974, Pondera County, Montana					
Clayton	0.001000			0.041399	1.917202	6.314132
Frank	0.001000			0.025487	1.949025	6.345955
Gumbel	**1.100000**			**5.732520**	**-5.244449**	**-9.465039**
Clayton, Clayton,Clayton	0.001000	0.077741	-0.048485 *	-1.036122	8.072244	21.263033
Clayton, Clayton,Gumbel	1.022122	-0.164136	0.846158 *	-19.985389	45.970779	59.161568
Frank, Frank, Frank	0.001000	-1.099259	-1.418806	-13.851869	33.703739	46.894528
Frank, Frank, Clayton	0.001000	0.862911	-0.044400 *	-1.105269	8.210538	21.401327
Frank, Frank, Gumbel	1.000204	-1.297100	0.832434	-20.887645	47.775289	60.966078
Gumbel,Gumbel, Gumbel	1.513458	0.952485 *	0.751978 *	-25.867954	57.735909	70.926698
Gumbel, Frank, Clayton	0.813972 *	0.925946	-0.043871	-4.338217	14.676433	27.867222
Gumbel, Frank, Gumbel	0.995191	1.000036	0.842727 *	-25.061201	56.122402	69.313191
	Temple TX, Bell County, Texas					
Clayton	0.00100			0.05886	6.10288	1.88229
Frank	0.00100			0.02635	6.16790	1.94731
Gumbel	**1.10000**			**-41.69724**	**-29.63757**	**-25.41698**
Clayton, Clayton,Clayton	0.00100	0.14990	-0.07667 *	-3.51428	25.69033	13.02856
Clayton, Clayton,Gumbel	0.00100	0.04713	0.77590 *	-32.56102	83.78380	71.12203
Frank, Frank, Frank	0.00100	1.21032	-2.14424 *	-29.12381	76.90940	64.24763
Frank, Frank, Clayton	0.00100	1.73605	-0.07553 *	-5.27117	29.20411	16.54234
Frank, Frank, Gumbel	0.00100	1.61078	0.79008 *	-34.17259	87.00695	74.34518
Gumbel,Gumbel, Gumbel	0.90662 *	1.11905	0.77290 *	-41.69724	102.05625	89.39448
Gumbel, Frank, Clayton	0.87410	1.75801	-0.07423 *	-7.30074	33.26324	20.60147
Gumbel, Frank, Gumbel	0.90662 *	1.11897	0.77243 *	-40.54667	99.75512	87.09335
	Spokane WA, Spokane County, Washington					
Clayton	0.00100			0.11772	1.76457	6.16150
Frank	0.00100			0.02776	1.94448	6.34141
Gumbel	**1.10000**			**-32.98232**	**-37.37925**	**-32.98232**
Clayton, Clayton,Clayton	0.00100	0.21398	-0.12842	-11.13180	28.26361	41.45439
Clayton, Clayton,Gumbel	0.00100	1.13314	0.89037 *	-12.96967	31.93935	45.13014
Frank, Frank, Frank	0.00100	0.19995	0.87737	-16.54238	39.08476	52.27555
Frank, Frank, Clayton	0.00100	1.77027	-0.10693 *	-13.77605	33.55209	46.74288
Frank, Frank, Gumbel	0.00100	1.72906	-1.00565 *	-16.30425	38.60849	51.79928
Gumbel,Gumbel, Gumbel	0.00100 *	1.85597	0.89791 *	-34.14157	74.28315	87.47394
Gumbel, Frank, Clayton	0.73401 *	1.83158	-0.13861	-34.72642	75.45285	88.64364
Gumbel, Frank, Gumbel	0.81940 *	2.74101	0.86815 *	-33.25106	72.50213	85.69292

Note: * Parameters that do not satisfy monotonicity conditions.

Table 4.A. Two-Sample Kolmogorov-Smirnov Test for Selected Dates Copula Based Stochastic Weather Generator (P-Values)

Selected Dates	Max Temp	Precipitation	Min Temp
	Conrad, Montana		
January, 9th	0.028 *	0.267	0.002 **
February, 1st	0.001 **	0.998	0.000 **
March, 3rd	0.453	0.999	0.525
April, 3rd	0.737	0.995	0.242
May, 1st	0.372	0.999	0.095
June, 1st	0.000 **	0.072	0.019 *
July, 1st	0.001 **	0.071	0.046 *
August, 7th	0.011 *	0.009 **	0.112
September, 5th	0.571	0.886	0.225
October, 1st	0.744	0.995	0.306
November, 4th	0.819	0.969	0.063
December, 1st	0.763	0.669	0.489
	Spokane, Washington		
January, 9th	0.069	0.001 **	0.107
February, 1st	0.159	0.050 *	0.001 **
March, 3rd	0.113	0.024 *	0.304
April, 3rd	0.012 *	0.804	0.017 *
May, 1st	0.077	0.362	0.058
June, 1st	0.089	0.765	0.360
July, 1st	0.082	0.999	0.011 *
August, 7th	0.007 **	0.530	0.000 **
September, 5th	0.691	0.701	0.730
October, 1st	0.022 *	0.739	0.172
November, 4th	0.023 *	0.004 **	0.123
December, 1st	0.000 **	0.000 **	0.371
	Temple, Texas		
January, 9th	0.035 *	0.393	0.015 *
February, 1st	0.013 *	0.049 *	0.040 *
March, 3rd	0.057	0.232	0.074
April, 3rd	0.007 **	0.959	0.530
May, 1st	0.113	0.001 **	0.203
June, 1st	0.000 **	0.114	0.013 *
July, 1st	0.000 **	0.989	0.000 **
August, 7th	0.013 *	0.980	0.003 **
September, 5th	0.002 **	0.985	0.000 **
October, 1st	0.111	0.965	0.343
November, 4th	0.046 *	0.866	0.199
December, 1st	0.967	0.994	0.142

Note: * Reject H_0 at 5% significance level

** Reject H_0 at 1% significance level

Table 4.B. Two-Sample Kolmogorov-Smirnov Test for Selected Dates Richardson Stochastic Weather Generator (P-Values)

Selected Dates	Max Temp	Precipitation	Min Temp
	Conrad, Montana		
January, 9th	0.226	0.560	0.022 *
February, 1st	0.167	0.989	0.062
March, 3rd	0.433	0.610	0.085
April, 3rd	0.448	0.662	0.794
May, 1st	0.028 *	0.081	0.073
June, 1st	0.019 *	0.053	0.267
July, 1st	0.628	0.404	0.794
August, 7th	0.190	0.404	0.069
September, 5th	0.062	0.696	0.139
October, 1st	0.017 *	0.880	0.145
November, 4th	0.104	0.960	0.056
December, 1st	0.867	0.960	0.464
	Spokane, Washington		
January, 9th	0.074	0.008 **	0.000 **
February, 1st	0.005 **	0.008 **	0.000 **
March, 3rd	0.001 **	0.326	0.000 **
April, 3rd	0.055	0.100	0.000 **
May, 1st	0.673	0.010 **	0.000 **
June, 1st	0.718	0.000	0.056
July, 1st	0.673	0.050	0.098
August, 7th	0.000 **	0.999	0.000 **
September, 5th	0.039	0.368	0.000 **
October, 1st	0.026 *	0.308	0.000 **
November, 4th	0.001 **	0.001	0.000 **
December, 1st	0.006 **	0.000	0.000 **
	Temple, Texas		
January, 9th	0.082	0.166	0.860
February, 1st	0.407	0.211	0.945
March, 3rd	0.377	0.673	0.339
April, 3rd	0.709	0.860	0.860
May, 1st	0.356	0.231	0.186
June, 1st	0.252	0.403	0.087
July, 1st	0.015 *	0.915	0.108
August, 7th	0.200	0.999	0.067
September, 5th	0.356	0.915	0.209
October, 1st	0.915	0.761	0.938
November, 4th	0.274	0.403	0.615
December, 1st	0.399	0.575	0.549

Note: * Reject H_0 at 5% significance level

** Reject H_0 at 1% significance level

Table 5. Comparative Quantile Analysis for Three Locations[32]

	Precipitation Amount, cm			Maximum Temperature, °C			Minimum Temperature, °C		
Quantile	**Observed**	**Copula**	**Richardson**	**Observed**	**Copula**	**Richardson**	**Observed**	**Copula**	**Richardson**
				Conrad, Montana					
0.025	0.000	0.000	0.000	-13.300	-8.430	-10.509	-25.000	-19.554	-23.416
0.05	0.000	0.000	0.000	-7.800	-5.078	-6.691	-21.100	-16.894	-19.987
0.1	0.000	0.000	0.000	-1.100	-1.570	-2.270	-15.600	-13.919	-15.950
0.2	0.000	0.000	0.000	4.400	2.882	3.499	-9.400	-10.824	-10.628
0.3	0.000	0.000	0.000	7.800	6.621	7.845	-5.600	-8.173	-6.548
0.4	0.000	0.000	0.000	11.100	10.101	11.656	-2.800	-5.286	-3.130
0.5	0.000	0.001	0.000	14.400	13.668	15.179	-0.600	-2.231	-0.109
0.6	0.000	0.002	0.000	18.300	17.021	18.574	2.200	0.850	2.653
0.7	0.000	0.004	0.000	21.700	20.708	21.926	5.000	4.014	5.170
0.8	0.000	0.012	0.000	25.000	24.521	25.360	7.800	7.213	7.532
0.9	0.203	0.227	0.189	28.900	29.184	29.218	10.000	10.269	10.171
0.975	0.864	0.700	0.862	32.800	35.305	34.067	13.300	14.168	13.460
				Spokane, Washington					
0.025	0.000	0.000	0.000	-5.000	-4.310	-9.620	-13.300	-9.946	-21.603
0.05	0.000	0.000	0.000	-2.200	-2.193	-6.219	-10.000	-8.247	-18.306
0.1	0.000	0.000	0.000	0.600	0.513	-2.214	-6.100	-6.355	-14.364
0.2	0.000	0.000	0.000	3.900	4.117	3.000	-2.800	-3.920	-9.242
0.3	0.000	0.000	0.000	7.200	7.296	6.939	-1.100	-1.713	-5.409
0.4	0.000	0.000	0.000	10.000	11.101	10.356	0.600	0.707	-2.121
0.5	0.000	0.002	0.000	13.300	14.906	13.668	2.800	3.323	0.757
0.6	0.000	0.004	0.000	17.200	18.782	16.967	5.000	5.676	3.400
0.7	0.025	0.013	0.000	21.100	22.232	20.440	7.800	7.952	5.899
0.8	0.127	0.142	0.057	25.000	25.227	24.114	10.000	10.098	8.255
0.9	0.406	0.385	0.310	29.400	28.778	28.416	12.800	12.460	10.881
0.975	1.036	0.847	1.051	33.900	33.279	33.821	16.700	15.418	14.170
				Temple, Texas					
0.025	0.000	0.000	0.000	5.600	8.628	5.834	-3.900	-1.949	-3.887
0.05	0.000	0.000	0.000	8.300	10.808	8.879	-1.700	-0.089	-1.718
0.1	0.000	0.000	0.000	12.800	13.258	12.514	1.100	1.579	0.916
0.2	0.000	0.000	0.000	17.200	16.658	17.173	4.400	4.265	4.598
0.3	0.000	0.000	0.000	21.100	19.400	20.764	7.700	6.622	7.802
0.4	0.000	0.000	0.000	23.900	22.198	23.771	10.600	9.397	10.868
0.5	0.000	0.000	0.000	26.700	25.397	26.549	13.900	12.355	13.943
0.6	0.000	0.000	0.000	28.900	28.454	29.058	17.200	15.529	16.841
0.7	0.000	0.000	0.000	31.700	30.986	31.440	19.900	18.297	19.272
0.8	0.025	0.025	0.000	33.300	33.538	33.684	21.700	20.588	21.130
0.9	0.533	0.802	0.618	35.600	36.810	36.122	22.800	23.421	22.878
0.975	2.769	2.901	2.767	37.800	41.595	39.152	23.900	26.973	25.023

[32] In weather stations the minimum reported amount of precipitation is 0.0254cm (0.01 inches).

Table 6. Average Rainfall Amount and Average Number of Rainy Days by 28-Day Period

	Precipitation Amount in Cms.			Number of Rainy Days		
Period	**Observed**	**Copula**	**Richardson**	**Observed**	**Copula**	**Richardson**
			Conrad, Montana			
1	0.976	1.590	0.858	5.020	4.625	3.067
2	0.798	1.502	0.801	4.120	4.708	2.720
3	1.193	1.573	1.148	4.580	5.125	3.367
4	2.111	2.002	1.853	5.040	5.525	3.780
5	3.335	1.903	3.722	6.620	5.767	5.187
6	6.672	1.272	5.227	9.800	4.542	6.787
7	4.086	2.323	3.960	7.420	5.958	5.293
8	2.483	2.535	2.681	5.440	6.542	4.507
9	2.925	1.950	2.565	5.900	5.058	4.573
10	2.156	1.357	1.897	4.960	4.367	3.513
11	1.037	1.699	1.137	3.660	4.558	2.533
12	1.127	1.703	1.108	4.420	4.242	3.007
13	1.036	2.210	0.901	4.700	5.925	2.947
			Spokane, Washington			
1	4.390	3.053	1.233	12.314	8.958	4.253
2	3.807	3.196	1.149	10.824	9.575	4.020
3	3.477	2.946	1.555	10.275	6.233	5.093
4	3.035	2.978	2.760	9.118	4.967	6.267
5	3.123	3.106	5.468	8.627	5.408	8.187
6	3.872	2.295	6.699	8.765	4.042	9.080
7	1.805	2.287	5.888	5.118	3.533	8.300
8	0.977	1.531	4.010	3.216	3.350	6.153
9	1.710	1.594	3.651	4.647	2.625	6.080
10	1.659	3.588	2.715	5.314	5.258	5.433
11	2.839	4.311	1.596	7.353	6.642	3.860
12	5.558	3.675	1.159	12.667	6.683	3.573
13	5.443	3.111	1.223	12.824	8.717	4.060
			Temple, Texas			
1	4.999	8.728	5.098	6.561	6.608	5.380
2	6.792	7.112	5.753	6.951	5.208	5.540
3	5.552	5.798	5.480	6.512	4.408	5.340
4	5.297	4.364	5.947	5.927	4.383	5.030
5	10.555	6.271	9.093	6.976	5.125	5.450
6	9.318	10.320	9.153	6.171	6.017	5.000
7	5.472	8.680	6.428	4.634	5.442	3.840
8	4.265	6.061	4.283	3.220	5.700	2.950
9	6.122	6.591	5.720	4.439	5.833	3.890
10	8.556	4.532	9.753	5.902	5.250	5.110
11	9.223	5.943	7.331	6.024	5.117	4.360
12	6.554	9.549	6.826	5.634	6.283	4.820
13	6.209	10.246	7.352	5.976	7.333	5.240

Table 7. Average Maximum Temperature and Average Minimum Temperature by 28-Day Period

	Maximum Temperature °C			Minimum Temperature °C		
Period	**Observed**	**Copula**	**EPIC**	**Observed**	**Copula**	**EPIC**
			Conrad, Montana			
1	0.03	0.48	1.16	-13.47	-13.06	-13.04
2	3.38	1.23	3.26	-10.59	-12.20	-10.89
3	6.73	5.35	6.99	-7.85	-9.15	-7.70
4	12.62	11.15	12.61	-2.92	-4.32	-3.15
5	17.37	17.12	17.79	1.60	1.28	1.69
6	21.56	22.44	21.86	6.19	5.72	5.88
7	25.90	26.45	25.88	9.07	8.41	8.87
8	28.48	27.05	27.57	10.24	8.66	9.45
9	25.86	24.62	25.50	7.92	7.04	7.42
10	20.04	19.89	20.19	2.78	2.77	2.80
11	14.10	13.20	14.09	-1.93	-2.85	-2.20
12	5.74	6.81	6.86	-8.12	-7.93	-7.59
13	1.03	2.74	1.83	-11.91	-11.39	-12.05
			Spokane, Washington			
1	0.45	0.66	1.88	-5.60	-6.12	-11.35
2	3.77	1.43	2.89	-3.65	-5.74	-9.63
3	8.16	5.37	5.69	-1.42	-3.23	-6.87
4	12.64	11.23	10.35	1.06	0.47	-2.57
5	17.31	17.91	15.24	4.38	4.85	1.95
6	21.51	23.70	19.89	8.24	8.91	6.02
7	26.21	27.31	24.46	11.32	11.25	9.30
8	29.90	27.81	26.95	13.77	11.45	10.31
9	26.38	25.51	24.73	11.14	10.25	8.55
10	20.96	21.55	19.00	6.61	7.95	3.99
11	12.91	15.47	13.21	1.44	4.04	-0.88
12	4.71	9.58	7.06	-1.99	-0.05	-5.88
13	0.57	4.30	2.69	-5.44	-3.69	-10.35
			Temple, Texas			
1	13.83	15.45	14.57	1.85	2.44	2.29
2	16.31	15.23	16.14	3.70	2.55	3.66
3	19.99	17.15	19.84	7.13	4.66	7.16
4	24.13	21.25	24.24	11.60	8.96	11.54
5	27.75	26.57	27.88	15.90	14.02	15.93
6	31.05	31.54	31.02	19.51	19.01	19.33
7	33.80	34.69	33.57	21.83	22.08	21.53
8	35.40	35.65	35.03	22.54	22.88	22.34
9	34.81	34.03	34.37	21.98	21.42	21.48
10	30.44	30.75	30.51	17.75	17.73	17.87
11	25.29	26.36	25.71	12.59	13.41	12.82
12	19.70	21.48	20.74	7.35	8.72	7.98
13	15.26	17.54	15.95	3.26	4.58	3.76

Table 8. Annual Average Temperature and Number of Days of Extreme Events by, Weather Station

Weather Variable	Observed	Copula	Richardson
Conrad			
Maximum Temperature, °C	40.60	47.50	49.99
Minimum Temperature, °C	-27.20	-28.53	-45.58
Days ≥ 35 °C	1.10	5.08	6.38
Days ≤ 0°C	183.84	205.13	183.93
Spokane			
Maximum Temperature, °C	42.20	42.02	51.42
Minimum Temperature, °C	-24.40	-24.56	-43.37
Days ≥ 35 °C	6.41	3.27	6.17
Days ≤ 0°C	138.27	113.95	172.73
Temple			
Maximum Temperature, °C	43.30	55.30	48.32
Minimum Temperature, °C	-14.40	-9.68	-16.23
Days ≥ 35 °C	54.61	55.43	52.11
Days ≤ 0°C	31.49	18.92	29.14

Table 9. Summary on Camelina Production Characteristics in Montana

Features	Description
Varieties	18 varieties of Camelina in dryland
Yeld	Average yield ranges from 500 to 1,850 lbs/acre depending on variety and geographical characteristics related soils, climate, inclination, etc.
Oil content	29 to 41% equals 60-80 gallons/acre
Price	18-16 cents/lb (2008), Great Plains 9-12 cents/lb (2009), Great Northem Growers Coop 16 cents/lb (2009), Bill Schillinger, WSU
Return	With yield of 1,585 lbs./acre and a price of 16 cents/pound, gross return is $142.61/acre and net return of $105.61/acre.
Costs	-With yield of 1,500 lbs/acre, total operating cost is $46.67/acre, breakeven production cost is $1.56/bushel (1 bushel= 50 lbs.), smaller than Canola ($4.33/acre) and spring wheat ($1.81/acre). -Fertility requirements 33.3% less than Canola production.

Note: This data corresponds to several Agronomic Experimental Stations of Montana State University located across Montana.

Sources: Johnson, D. (2007). Presentation Prepared for the Harvesting Clean Energy Conference VII, MSU Northwestern Agricultural Research Center, Montana State University.

Table 10. Physiology and Cultural Practices Scheduling for Camelina Calibration in Conrad Farm, Pondera County, Montana

Stage	Approximate dates
Planting	Early April-Early June
emerging	Mid April-mid June
Blooming	Mid June-Early July
Turning	Late June-Late July
Harvesting	Mid July-Late August
Activity	**Exact Dates**
Pesticide	March 15
Fertilize	March 20
Pesticide	April 1
Fertilize	April 15
Planting	April 15
Fertilizer	June 12
Hauling	July 1
Harvest	July 30

Source: Adapted from RMA (2011)

Table 11. Producer Vs. Simulated Yield

Grain Yield in lbs/acre

Year	Producer	Simulated
2007	1,150.0	1,099.0
2008	2,295.0	2,302.8
2009	1,135.0	1,048.6
2010	1,378.0	1,855.6
Average	**1,489.5**	**1,576.5**

Table 12. Detrending Regressions for Camelina Yields

Yields Simulated Using Richardson (1981) SWG

Observations	100.00			
F(1, 98)	1.40			
Prob > F	0.24			
R-squared	0.01			
Adj R-squared	0.00			
Root MSE	555.23			
Variable	Coefficient	Std Error	t- Statistic	P>t
Trend	-2.28	1.92	-1.18	0.24
Const	1065.86	111.88	9.53	0.00

Yields Simulated Using Copula Based SWG

Observations	100.00			
F(1, 98)	3.67			
Prob > F	0.06			
R-squared	0.04			
Adj R-squared	0.03			
Root MSE	992.44			
Variable	Coefficient	Std Error	t- Statistic	P>t
Trend	8.41	4.39	1.92	0.06
Const	539.32	276.29	1.95	0.05

Table 13. Camelina Yields Basic Statistics: Conrad, Pondera County, Montana

SWG	Mean	Std Dev.	Minimum	Maximum
Copula based	967.07	963.28	17.46	3,827.62
Richardson (1981)	950.95	556.35	64.99	3,178.69

Table 14. Probability Parametric Distributions for Camelina Yields from Conrad, Pondera County, Montana

Distribution	Range & Params.	α	β	Max Likelihood	AIC	BIC
				Copula Based SWG		
Beta	0≤x≤1, α>0, β>0	0.42	0.75	-29.26	62.51	67.72
Weibull	0≤x<∞, α>0, β>0	922.58	0.90	792.86	-1,581.73	-1,576.52
Gamma	0≤x<∞, α>0, β>0	0.83	1,164.29	768.50	-1,533.00	-1,527.79
				Richardson (1981) SWG		
Beta	0≤x≤1, α>0, β>0	0.75	1.09	-5.39	14.77	19.99
Weibull	0≤x<∞, α>0, β>0	1,069.58	1.78	756.88	-1,509.75	-1,504.54
Gamma	0≤x<∞, α>0, β>0	2.62	363.20	775.45	-1,546.90	-1,541.69

Table 15. APH Insurance for Camelina Yields for Different Coverage Levels

Coverage	50%	60%	65%	70%	75%	80%	85%
Copula Based SWG, APH Yield 967.07 Lbs/Acre and Guarantee Price of $0.15 per Pound							
Liability	**72.53**	**87.04**	**94.29**	**101.54**	**108.80**	**116.05**	**123.30**
Unloaded Fair Premium (Dollars per Acre)							
Weibull	7.92	13.22	16.53	20.30	24.55	29.30	34.55
Beta	6.07	9.80	12.09	14.66	17.55	20.75	24.28
Gamma	7.90	13.15	16.43	20.15	24.36	29.05	34.24
Kernel	5.83	9.98	12.60	15.63	19.08	22.97	27.31
Fully Loaded Premium (Dollars per Acre)							
Weibull	12.99	21.70	27.13	33.31	40.29	48.09	56.71
Beta	9.96	16.09	19.84	24.07	28.80	34.06	39.85
Gamma	12.96	21.59	26.96	33.08	39.98	47.68	56.20
Kernel	9.57	16.37	20.69	25.65	31.31	37.70	44.82
Loss Cost (Percentage)							
Weibull	10.92	15.19	17.53	19.99	22.56	25.24	28.02
Beta	8.37	11.26	12.82	14.44	16.13	17.88	19.69
Gamma	10.89	15.11	17.42	19.85	22.39	25.03	27.77
Kernel	8.04	11.46	13.37	15.39	17.54	19.79	22.15
Fully Loaded Base Premium (Percentage)							
Weibull	17.92	24.93	28.78	32.81	37.03	41.44	45.99
Beta	13.74	18.48	21.04	23.70	26.48	29.35	32.32
Gamma	17.87	24.80	28.60	32.58	36.75	41.09	45.58
Kernel	13.19	18.81	21.94	25.26	28.78	32.48	36.35
Richardson (1981) SWG, APH Yield 950.95 Lbs/Acre and Guarantee Price of $0.15 per Pound							
Liability	**71.32**	**85.59**	**92.72**	**99.85**	**106.98**	**114.11**	**121.25**
Unloaded Fair Premium (Dollars per Acre)							
Weibull	2.59	5.14	6.91	9.06	11.62	14.67	18.19
Beta	3.3	5.9	7.5	9.4	11.6	14.0	16.8
Gamma	2.51	5.17	7.04	9.31	12.04	15.27	19.02
Kernel	2.58	5.17	6.95	9.10	11.63	14.60	18.00
Fully Loaded Premium (Dollars per Acre)							
Weibull	4.25	8.44	11.34	14.86	19.08	24.07	29.86
Beta	5.45	9.62	12.30	15.40	18.97	23.04	27.62
Gamma	4.13	8.49	11.55	15.28	19.76	25.07	31.22
Kernel	4.24	8.49	11.41	14.93	19.09	23.96	29.55
Loss Cost (Percentage)							
Weibull	3.63	6.01	7.45	9.07	10.87	12.85	15.00
Beta	4.66	6.85	8.08	9.40	10.80	12.30	13.88
Gamma	3.52	6.04	7.59	9.32	11.25	13.38	15.69
Kernel	3.62	6.04	7.50	9.11	10.87	12.79	14.85
Fully Loaded Base Premium (Percentage)							
Weibull	5.96	9.86	12.23	14.89	17.84	21.10	24.63
Beta	7.64	11.24	13.26	15.43	17.74	20.19	22.78
Gamma	5.79	9.92	12.46	15.31	18.47	21.97	25.75
Kernel	5.94	9.92	12.31	14.95	17.84	21.00	24.37