Non-Parametric Analysis of ENSO Impacts on Yield Distributions: Implications for GRP Contract Design.

Paper prepared for presentation at the 2007 Annual Meetings
of the SAEA in Mobile, AL, February 4-7.

Authors:
Denis Nadolnyak, Agecon, Auburn University, nadolda@auburn.edu
James L. Novak, Agecon, Auburn University jnovak@acesag.auburn.edu
Dmitry V. Vedenov, Agecon, University of Georgia, dvedenov@agecon.uga.edu
Joel Paz, Bio. and Ag. Engineeering, University of Georgia, jpaz@uga.edu
Clyde W. Fraisse, Bio. and Ag. Engineeering, University of Florida, cfraisse@ifas.ufl.edu
Gerritt Hoogenboom, Bio. and Ag. Engineeering, University of Georgia, gerrit@uga.edu

ABSTRACT
The paper reports preliminary results of non-parametric analysis of historical and crop model generated peanut yield series in the Southwest Georgia. The results suggest ENSO phase dependent differences in yield distributions that are similar for both the simulated and actual series. The differences are magnified in GRP insurance premiums.

INTRODUCTION

The ElNino Southern Oscillation (ENSO) phases have distinct impacts on the climate in the Southeastern United States. It is reasonable to expect that they also affect crop yields (Hansen, Hodges, and Jones). The magnitude of these effects may be significant enough to accommodate them in agricultural decision making.

In this paper, we report preliminary results of non-parametric analysis of peanut yield distributions in the Southwest Georgia. The focus of the analysis is on establishing ENSO-dependent differences in the yield distributions and on evaluating their implications for area yield crop insurance the expected losses for which are calculated using county average yield series. The results of kernel density estimates of simulated yield data based on actual weather realizations in Colquitt county, GA, and actual county yield time series from the same geographical area suggest that there are distinct similarities between the simulated and actual peanut yield data and that the dissimilarities can be explained by known factors.

We also find that, while the non-parametric densities of the historical county average yield series may not always differ significantly in their means and higher moments between ENSO phases, the combined effect of the differences in densities seems to matter for area crop insurance calculations. In particular, the expected loss to coverage ratios are consistently the highest during the Neutral years and the lowest during the ElNino years for all 17 counties and for simulated yield series. At the same time, the ratios calculated for pooled data (all ENSO phases) using non-parametric densities seem to be only slightly higher than those calculated using normal distribution. These findings should be more relevant for the area yield insurance as opposed to the APH arrangements as the yield data used in designing contracts for the former reflects the systemic risk more dependent on climate than on the farm-level, basis risk factors accommodated in the APH plans.

Due to size limitations, the methodology and results presentation is fairly brief. The methodology section explains the crop simulation and density estimation techniques. In the results and discussion section, we briefly relay the preliminary findings, followed by suggestions for refining the methodology. But, first, a few words on the ENSO climate impacts in the U.S. Southeast.

ENSO Phases

The ElNino Southern Oscillation (ENSO) is an atmospheric phenomenon observed with irregular periodicity which is believed to affect global climate. The phase is determined by warming or cooling of the ocean surface in the western Pacific ocean, which changes the trade wind patterns and, subsequently, global weather for the duration of the phase and beyond (see http://meted.ucar.edu/climate/enso/index.htm for the basics of the process). The effects of ENSO phases on the climate in different geographical regions are complex and are not discussed here. In the Southeastern United States, more or less distinct ENSO dependent climate patterns have been observed. The table below provides a summary of the effects on temperature and precipitation across the Southeast.

El Nino/La Nina Impacts Across the Southeast U.S.					
Phase	Region	Seasons			
		Oct-Dec	Jan-Mar	Apr-Jun	Jul-Sep
El Nino	Peninsular Florida	Wet & cool	Very wet & cool	Slightly dry	Slightly dry to no impact
	Tri-State Region	Wet	Wet	Slightly wet	No impact
	Western Florida Panhandle	No impact	Wet	Slightly dry	No impact
	Central and North Ala. & Ga.	No impact	No impact	No impact	Slightly dry
La Nina	Peninsular Florida	Dry & slightly warm	Very dry & warm	Slightly wet	Slightly cool
	Tri-State Region	Slighly dry	Dry	Dry	No impact
	Western Florida Panhandle	Slightly dry	Dry	Dry	No impact
	Central and North Ala. & Ga.	Dry	Dry in the south, wet in NW Ala.	No impact	Wet in NW Ala.
Neutral	All Regions	No impact	No impact	No impact	No impact

Source: http://www.coastalclimate.org/climate/seimpacts.php

In addition to this, the likelihood of a severe freeze is much greater during a neutral phase than during either an El Nino or a La Nina event. As data come from the Southwestern Georgia, the impacts on the western Florida Panhandle seems to be the most appropriate. Certain effects of the ENSO cycles on simulated peanut yields in the Southeast have been observed (Garcia y Garcia et al; Cabrera et al; Fraisse et al). This paper extends the analysis by non-parametrically estimating yield data from a narrowly defined region.

METHODOLOGY

The analysis uses two data sets: a simulated set of peanut yields derived using daily weather information for the period of 1911 to 2003 and a set of historical county average yield series spanning the period from 1934 to 2005. A non-parametric technique of kernel density estimation is used for both datasets. This section briefly describes the crop simulation modeling and the kernel density approach.

Crop Simulation.

The Cropping System Model (CSM)-CROPGRO-Peanut model (Boote, Jones, and Hoogenboom, 1998; Jones et al., 2003) was used to simulate peanut yield responses to different climate, irrigation, and planting date scenarios. The CSM-CROPGRO-Peanut model, which is part of Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.0 (Hoogenboom et al., 2004), is a process-based model that simulates crop growth and development and the plant and soil water, and nitrogen balances. Long-term historical weather data (1900-2004) were obtained from the National Weather Service (NWS) Cooperative

Observer Program (COOP) network and compiled by the Center for Oceanic-Atmospheric Prediction Studies (COAPS), through the SECC. Weather variables include daily maximum and minimum temperatures and precipitation. A solar radiation generator, WGENR, with adjustment factors obtained for the southeastern USA was used to generate daily solar radiation data.

Georgia Green, a medium maturing runner-type peanut variety, was selected as the representative variety for all counties included in the simulation. It is currently one of the most commonly grown varieties in Georgia and other southeastern states. The soil profile data of three representative soils for each county were obtained from the soil characterization database of the USDA National Resource Conservation Service. Nine planting dates (April 16, 23; May 1, 8, 15, 22, 29; June 5 and 12) were considered in the simulation. These represent all possible planting dates at weekly intervals. The typical planting window for peanuts is between mid-April and mid-June. Peanut growth and development were simulated without irrigation to predict yield.

Regardless of their complexity and accommodation of biological and physical processes, the crop simulation models are deterministic. Therefore, whatever randomness in simulated yields we observe for same plots and management comes from random weather realizations. In this way, the simulated data is analogous to a controlled experiment. At the same time, it is nearly impossible to translate weather variability, expressed in so many ways, into yield variability through the model mechanics. For instance, cumulative measures of precipitation and solar radiation may not correlated with yields if the weather patterns are different, as evidenced by a comparison of the effect on plant growth of a week with four rainy days each followed by a sunny one with a week in which it rains four days in a row (the first one is likely to be more favorable). Thus, we do not try to deliberately draw parallels between climate research and our findings. Instead, we independently estimate the distributions of the simulated yields without forming any a priori expectations based the climate research.

The simulated annual data covers the period from 1911 to 2003 and assumes modern "best" management practices. This time period covers 14 ElNino, 17 LaNina, and 39 Neutral years. This is barely enough for distribution analysis, actual daily weather observation records do not go much further back. The actual weather observations used in the peanut yield simulation are from a weather station in Colquitt County in Southwestern Georgia, located in the heart of the Southeastern peanut producing region.

The nine simulated planting dates and three soil types, make for 2511 observations. The three soil types assumed are Tifton Loamy Sand, Cowarts Loamy Sand, and Troup Sand, the first being the most prevalent in the county (NRCS). As the differences in yields between the soil types were negligible and because peanuts are planted on all these soils, we did not distinguish between the soil types in most of the analysis.

Non-Parametric Density Estimation

Conventional parametric approaches to insurance analysis assume known functional forms for yield distributions. The most commonly assumed density is normal, which is justified by the Central Limit Theorem. However, empirical yield data do not always conform to theoretical priors due to a number of physical and biological attributes of plant growth. In particular, the bi-modality and skewness of yield distributions are often observed. Non-parametric density estimation accommodates these and other distributional idiosyncrasies.

The simplest way to estimate non-parametric density is to use a histogram. This, however, poses the problem of discontinuity and requires large samples. Smoothing the density between observations utilizes a kernel function, an estimator of local density around a datum. The localities overlap (to the degree of the kernel width), so that each kernel density depends not only on its own but also adjacent observations. The resulting density estimation is continuous, is supposed to be informationally richer than empirical rates, and improves the quality of comparisons with parametric densities. There is a range of possible kernel functions and kernel widths, the latter arguably being a more important specification. For the preliminary analysis, we chose Gaussian kernels as more commonly used in economics and the kernel width according to the Silverman's "rule of thumb" as optimal for the normal distribution family. Kernel density estimates can readily be compared to suggested parametric densities using Kolmogorov-Smirnov tests.

For a more thorough discussion of kernel density estimation, see the first chapter in Li and Racine.

RESULTS AND DISCUSSION:

The results reported in this section are preliminary and will benefit from refining the methodology used in the analysis. So far, the main findings are that

1) that there are some similarities between the simulated and actual peanut yield data and that the dissimilarities can be explained by known factors.
2) while the non-parametric densities of the historical county average yield series may not always differ significantly in their means and higher moments between ENSO phases, the "cumulative" difference in (non-parametric) densities seems to matter for area crop insurance calculations. In particular, the expected loss to coverage ratios are consistently the highest during the Neutral years and the lowest during the ElNino years for all 17 counties and for simulated yield series. At the same time, the ratios calculated for pooled data (all ENSO phases) using non-parametric densities seem to be only slightly higher than those calculated using normal distribution.

Simulated Yield Series

The table below provides some basic statistical parameters of the distribution of the simulated peanut yields by the ENSO phase.

Simulated peanut yield distribution, average of planting dates and soils.

enso	mean	sd	skewness	kurtosis	*min*	*max*
ElNino	2329.25	1003.81	0.43	2.92	*357*	*4902*
LaNina	2588.70	1282.87	0.31	2.03	*337*	*5519*
Neutral	2298.67	1179.43	0.85	3.29	*454*	*6297*
All years	2373.53	1177.45	0.66	2.85	*337*	*6297*

LaNina yields have the highest mean, which is associated with higher variance. The skewnesss or the simulated yield distribution during neutral years is fairly high (>0.5), while during the ENSO phases it is below moderate. Higher kurtosis is normally interpreted as greater "peakedness", which means more of the variance is due to infrequent extreme deviations, as opposed to frequent modestly-sized deviations. However, these moments convey more information when applied to parametric distributions.

The statistics for different planting dates and soil types show that the highest yields come from planting in between May 8-22 and Tifton Loamy Sand. However, the *relative* values of the distribution statistics are preserved throughout. As evaluating soil productivity and management practices is beyond the scope of this paper, we report such differences only in cases where they affect the distributional differences.

Assuming the simulated data represents a controlled experiment in that the yields do not depend on unknown stochastic influences, the important question is whether the yields depend on the ENSO phase. The results of t-tests of mean equality are reported in a table below.

H_0: mean(yield, EL) – mean(yield, LA) = 0		
Ha: diff < 0	Ha: diff != 0	Ha: diff > 0
Pr(T < t) = 0.0001	Pr(|T| > |t|) = 0.0002	Pr(T > t) = 0.9999
H_0: mean(yield, EL) – mean(yield, Neu) = 0		
Pr(T < t) = 0.6992	Pr(|T| > |t|) = 0.6016	Pr(T > t) = 0.3008
H_0: mean(yield, LA) – mean(yield, Neu) = 0		
Pr(T < t) = 1.0000	Pr(|T| > |t|) = 0.0000	Pr(T > t) = 0.0000

The tests confirm that LaNina yields are the highest in comparison to Neutral and ElNino, while there is no statistical difference between Neutral and ElNino. It has been suggested that the effects of the ENSO phases can carry over to the next year, but we did not find indications of such influences in the simulated data.

The table below shows the percentage differences in the simulated yields (with asterisk indicating significance).

Percentage Differences in Average Yields (row over column)

	EL	LA	Neu
EL	0	-10%*	1.3%
LA	11.1%*	0	12.6%*
Neu	-1.3%	-11.3%*	0

The significant difference in average yields depending on the ENSO phase is an interesting find, especially considering the fact that little ENSO climate impact has been found in South Georgia and western Florida. One reason for this discrepancy might lie in the details of the crop growth functions and their dependence not on the average temperature or precipitation, but on the finer details of climate, like the spacing of rainy and sunny days and other factors economists are ignorant of. Still, these findings are preliminary and must be taken with extreme caution.

The moments do not describe empirical distributions completely, and thus we proceed by reporting the differences between the distributions themselves. As common procedures for testing equality of variances rely on distributional assumptions which might not hold for the yield data, we use the non-parametric Kolmogorov-Smirnov test. The two sample test is based on the maximum absolute difference (D) between the CDFs for two continuous random variables. Unlike conventional statistical tests, this is a non-parametric test that does not require the variables to be normally distributed. The null hypothesis for the Kolmogorov-Smirnov test assumes there is no different in the CDFs associated with the two groups (i.e. school and home locations). The results of the distribution comparisons are reported in a table below. The largest observed difference between the two CDFs being examined was compared to the critical value of D at the 5 percent level of significance to determine if there is a statistically significant difference between the curves. The table below reports the results.

H_0: f(EL) = f(LA)			
Smaller group	D	P-value	Corrected
EL	0.1702	0.000	
LA	-0.0595	0.142	
Combined K-S	0.1702	0.000	0.000
H_0: f(EL) = f(Neutral)			
Smaller group	D	P-value	Corrected

EL	0.0471	0.189	
Neutral	-0.1188	0.000	
Combined K-S	0.1188	0.000	0.000
H_0: f(LA) = f(Neutral)			
Smaller group	D	P-value	Corrected
LA	0.0164	0.799	
Neutral	-0.1333	0.000	
Combined K-S	0.1333	0.000	0.000

The combined K-S statistics clearly show that the yield distributions in EL, La, and Neutral years are significantly different from each other (the differences are similar across planting dates and soil types). This means that it makes sense to proceed to integration of the pdf's in order to estimate the expected losses and their difference between ENSO phases.

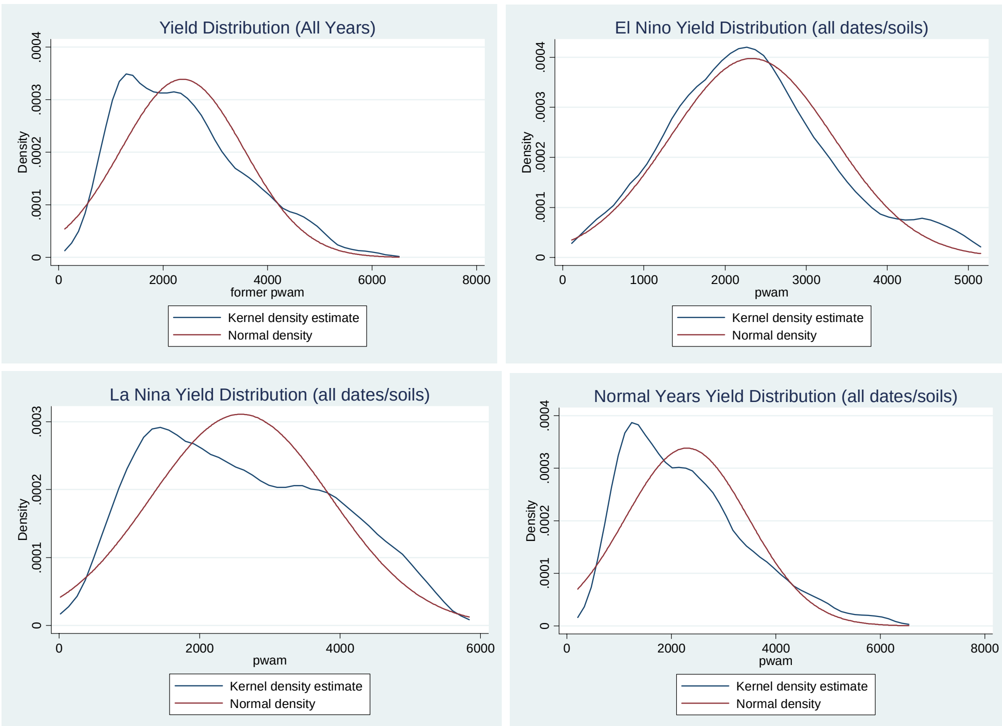
The observed differences in the distributions have immediate implications for insurance design, particularly for the group risk plans where rates for incomplete (less than 100%) coverage are calculated on the basis of assumed distributional parameters. Normally, such calculations assume normality defended using the Central Limit Theorem. However, even the simulated yield data shows that the simulated yield distributions are significantly different from their normal counterparts. The table below reports the results of one-sided Kolmogorov-Smirnov and Shapiro-Wilk tests for normality.

Kolmogorov-Smirnov test against normality (combined)			
	D	P-value	Corrected
All Years	0.0632	0.000	0.000
ElNino	0.0558	0.082	0.073
LaNina	0.0820	0.001	0.001
Neutral	0.0788	0.000	0.000
Shapiro-Wilk W test for normal data			
	W	z	Prob>z
All Years	0.95871	10.509	0.00000
ElNino	0.97961	4.690	0.00000
LaNina	0.96183	6.560	0.00000
Neutral	0.93921	9.929	0.00000

The results imply that the yields during neither of the three SO cycles, and even the pooled yields, are normally distributed with reasonable confidence levels. In light of the similar findings in the empirical literature, this is hardly surprising.

Finally, the plots of non-parametric yield distribution densities are shown in the figures below. Here report data from all soils and planting dates, as varying these parameters hardly shows any noticeable differences.

Kernel Densities of Simulated yields (Gaussian kernels and Silverman's kernel width).

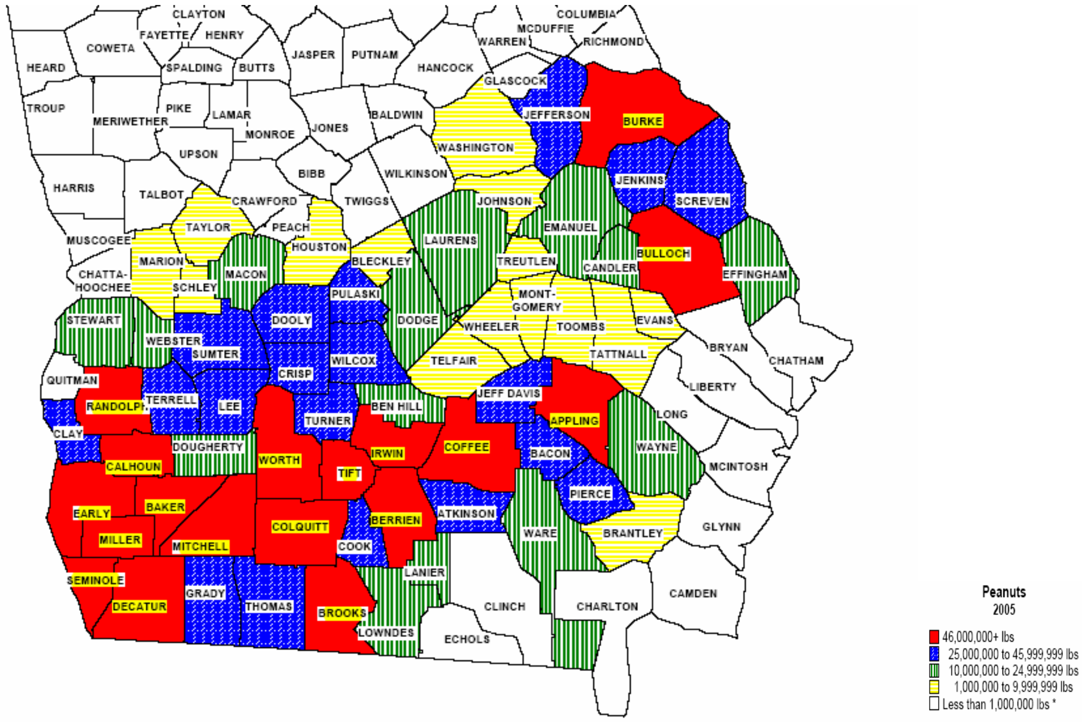


Visual inspection confirms that the differences among the SO phases. The ElNino yields seem to be more normally distributed than the even the pooled data. Perhaps this means that, indeed, the left skewness of yield data is the norm, and the ElNino yields deviate from this general pattern because ElNino years are wetter in the NW Florida and SW Georgia, contributing to plant growth (REF to Jones). The fact that the LaNina yields are higher on average and are least skewed (also confirmed by historical data) in the area believed to be dryer than average during LaNina years requires special consideration. We now proceed to the description of the historical (county average) data.

Historical Average County Yield Data.

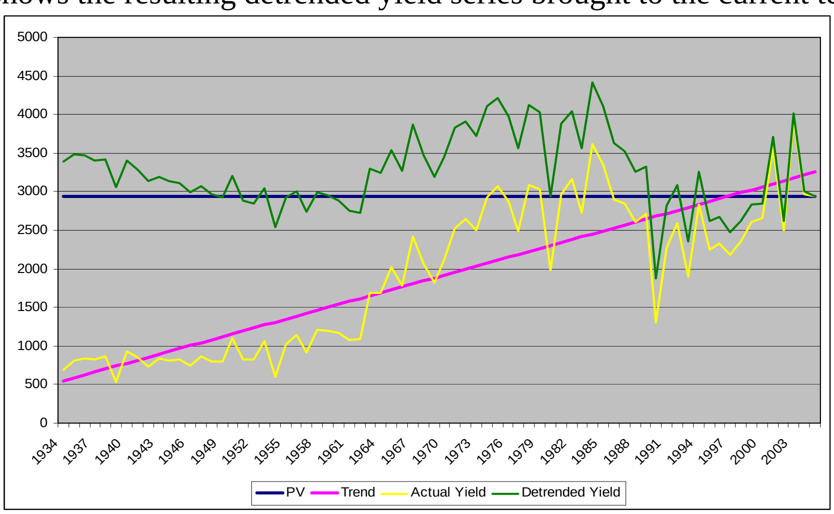
Brief Data Description

GA county average yield data for 1934-2005 was used for this analysis. 17 counties were chosen on the basis of the 2005 production volumes (more than 45,000,000lbs). While they might not have always been the top producers, the data shows that production leadership is more or less historical (this area did not experience peanut production shifts following the 2002 policy changes). Almost all the counties are located in Southwest Georgia. As the yield simulations were done for the Colquitt county weather, we are particularly interested in that county's time series.



The 1934-2005 series can not be used in its pristine form as there is an obvious upward trend and the deviations from it are likely to be heteroscedastic. Also, a mere visual inspection might also suggest a random walk. Detrending a time series can be a daunting task and a variety of methodologies are available. At this stage of the analysis, we used a linear trend and also tried to fit an ARIMA model.

A linear trend was generated by regressing the yields on the time variable and using robust variance matrix calculation. As the simulated data were generated assuming modern production practices, we assumed the last trend value to be the mean. This was done for purely illustrative purposes as this additive transformation does not affect the distributional properties of the series. Experimenting with non-linear trends showed no improvement in fits. The figure below shows the resulting detrended yield series brought to the current technology level.[1]



[1] Needless to say, this approach can be criticized on many grounds; detrending should be approached more carefully.

The looks of the detrended series suggest a random walk (confirmed by the tests reported later), which may also be a result of a few major and quite a few minor policy and technology shocks the industry experienced throughout the 20th century (for a review of the history of peanut production in the U.S., see REF). In other words, there are reasons to expect the data to be noisy and, because of this, peanut yield series may not be the best data to look for ENSO patterns, yet the crop is significant for the Southern agriculture. A possible refinement would be to use separate trends for several technological periods (as suggested in Garcia et al).

Distribution Analysis

In comparing the historical county-level yields with the simulated ones, we did not expect to get a perfect match. In particular, there is no reason to expect the average yields to be similar even when using the detrended series, as the NASS data used here includes both irrigated and non-irrigated yields. During the last 20 years, about 20% of the peanut acreage in the southwest Georgia has been irrigated, and the yields from irrigated production are typically ~30% higher. It is also well known that irrigated yields are less volatile due to independence of precipitation, which dampens any possible ENSO phase influences. The ENSO effects are further reduced because of the averaging of individual data in the county estimates. However, there is no reason not to expect similarities between the distributions of the real and simulated yields as those are shaped by a group of biological and physiological factors common to both practices. The table below lists the four moments of the Colquitt county yield distributions.

enso	**mean**	**sd**	**skewness**	**kurtosis**
ElNino	3316.9	362.5	-0.073	2.77
LaNina	3369.8	478.8	0.441	1.838
Neutral	3192.6	555.2	0.178	2.735
All years	3260.5	503.8	0.110	2.802

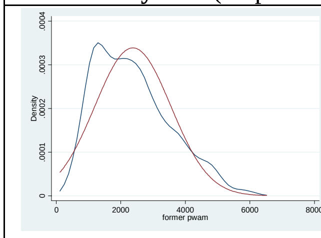
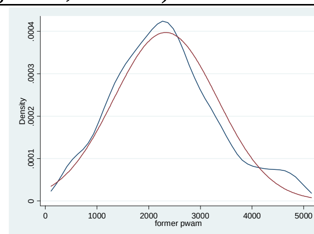
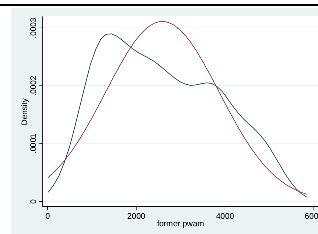
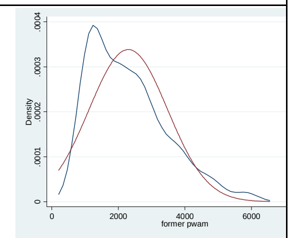
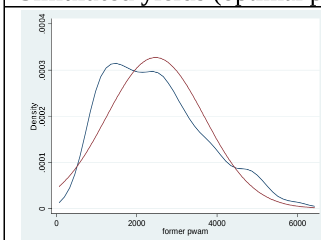
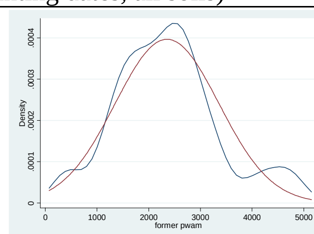
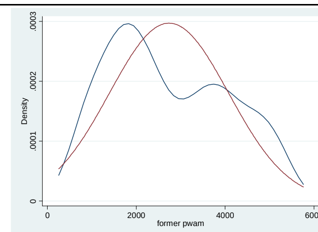
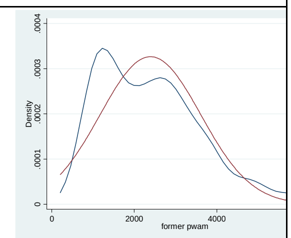
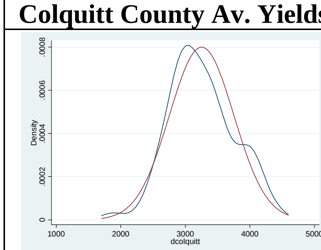
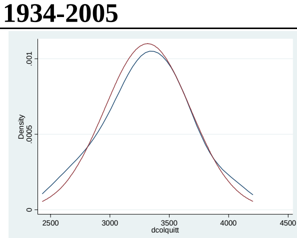
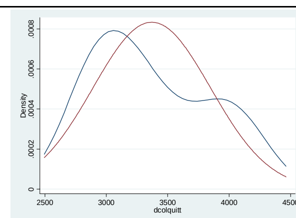
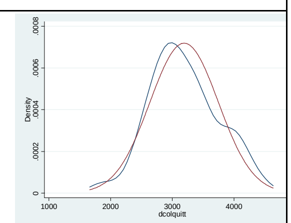
The average yields preserve the same relation as the simulated data: LaNina > ElNino > Neutral. The table below shows the percentage differences.

Differences in Average Yields between ENSO Phases, Colquitt county, GA (row/column)

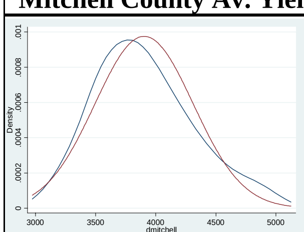
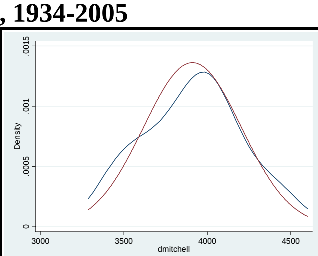
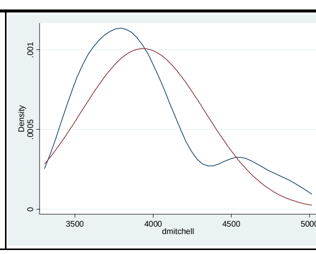
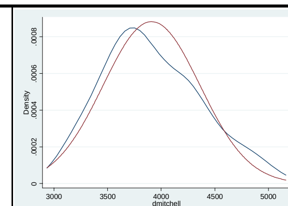
	EL	**LA**	**Neu**
EL	0	-1.6%	3.9%
LA	1.6%	0	5.6%
Neu	-3.9%	-5.3%	0

The average yields are statistically different only at levels greater than 23.4%. While the reasons for LaNina yields being the largest are still unclear, the Neutral year yields are probably the smallest because of the suggested higher freeze probability. The relative magnitudes of variance, skewness, and kurtosis seem to be different from the simulated yields. The significantly smaller variance of actual yields is explained by irrigation and by county-level averaging. Normally, higher differences between individual (simulated) and average yield variances should indicate greater heterogeneity among individual producers, which may account for the differences in the other distribution parameters (Barnett, Black, Hu, Skees, 2005).

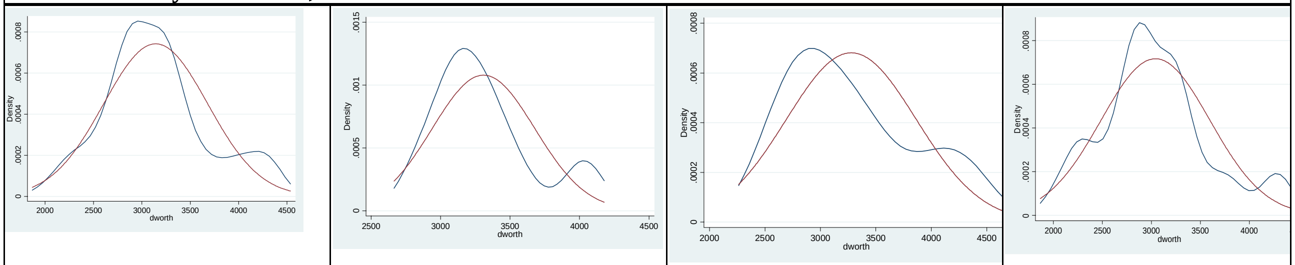
As was mentioned before, higher moments are less relevant when dealing with non-parametric distributions, when visual inspection is of greater importance. The table below presents comparisons of the simulated and actual yield distributions by ENSO phase with overlaid normal densities.

Simulated and Historical Yield Distributions, Colquitt, GA (normal overlay)			
All Years	**ElNino**	**LaNina**	**Neutral**
Simulated yields (all planting dates, all soils)			
			
Simulated yields (optimal planting dates, all soils)			
			
Colquitt County Av. Yields, 1934-2005			
			

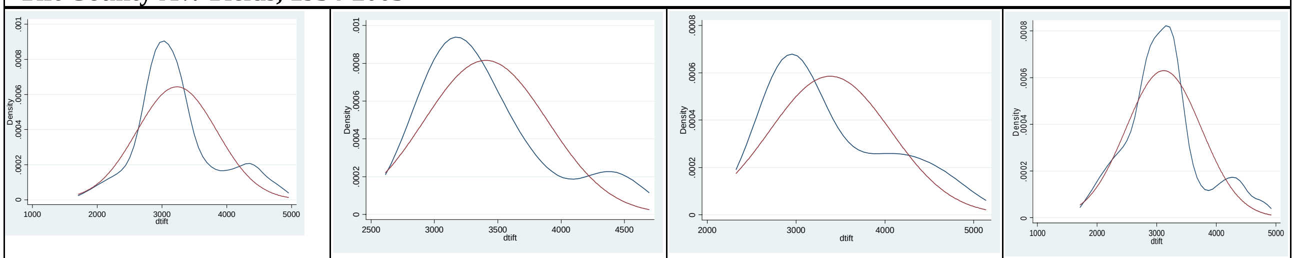
The ENSO-dependent similarity is largely preserved across counties. In both simulated and empirical series, the ElNino densities seem least skewed and closest to normal, while the LaNina densities show the greatest evidence of bi-modality. The Neutral year densities have the thinnest peaks, evidenced by the kurtosis value. In most cases, LaNina yields also show the strongest, and Neutral year the weakest, evidence of bi-modality which is also reflected in higher variance. In drawing these conclusions, it should be mentioned that the data span is sufficiently long making it unlikely that the differences in the ENSO dependent distribution patterns are due to chance. A similar table for counties adjacent to Colquitt provides an additional illustration.

Historical Yield Distributions for counties adjacent to Colquitt, GA (Gaussian kernel, normal overlay)			
All Years	**ElNino**	**LaNina**	**Neutral**
Mitchell County Av. Yields, 1934-2005			
			

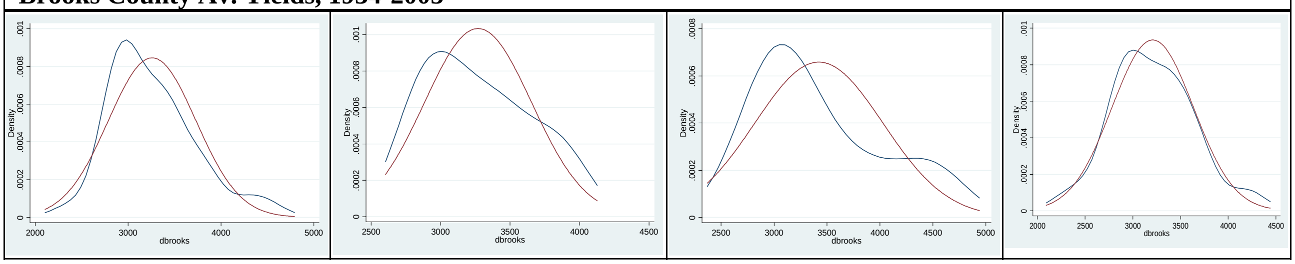
Worth County Av. Yields, 1934-2005



Tift County Av. Yields, 1934-2005



Brooks County Av. Yields, 1934-2005



As the analyzed distributions are quite idiosyncratic, we did not use parametric tests for comparing the moments of empirical and simulated distributions. Instead, we computed the frequencies of the incidences of yield distribution moments being the highest or the lowest during an ENSO phase for the 17 counties and compared them to the simulated data:

Frequencies of *Empirical* Distribution Moments being Highest and Lowest among the ENSO Phases, %

	mean			std			skewness			kurtosis		
	EN	LA	NE	EN	LA	NE	EN	LA	NE	EN	LA	NE
max	29	**71**	0	0	**82**	18	29	**65**	6	12	24	**65**
min	0	6	**94**	**100**	0	0	47	6	47	47	47	6

Frequencies of *Simulated* Distribution Moments being Highest and Lowest among the ENSO Phases, %

	mean			std			skewness			kurtosis		
	EN	LA	NE	EN	LA	NE	EN	LA	NE	EN	LA	NE
max	0	**100**	0	0	**100**	0	0	0	**100**	0	0	**100**
min	0	0	**100**	**100**	0	0	0	**100**	0	0	**100**	0

Clearly, there are similarities. The empirical data show that the majority of highest county average yields happen in the LaNina years and the overwhelming majority of lowest average yields happen in the Neutral years, which agrees with the simulated data. The empirical data also indicate that about 30% of the highest average county yields happen in the ElNino years. The high yields in LaNina years may be due to the observed higher cumulative solar radiation and the low yields during Neutral years may be due to the lack of it and also relative lack of rainfall. However, we cannot offer a compelling explanation for this right now. It is reasonable to expect that, *ceteris paribus*, higher yields during LaNina years should reduce expected insurance losses.

The standard deviations of the average county yield series are the highest during LaNina in 82 percent of the counties and they are always the lowest during ElNino, which also matches the simulated data. However, the skewness coefficient's relative frequency does not match the simulated data. Finally, the kurtosis is the highest during the neutral years in 65% of the counties analyzed and happen with equal frequencies during ElNino and LaNina years.

Apart from this, it is hard to describe these empirical distributions in much greater. The significance of distributional differences is determined by their effect on the methodology that uses these data. Our primary interest is the area yield insurance implications of using non-parametric distribution estimates.

Comparison of Expected Losses

Crop insurance contracts use forecasted (expected) yields and their (assumed) distributions to calculate actuarially fair premium rates based on the chosen coverage and scaling factors. The assumption of actuarial fairness is unrealistic as loading – accommodating (quasi)fixed costs in prates is a norm, but is relatively harmless for the purposes of this analysis.

In the simplest case, the rate is calculated as the ratio of the expected loss and the coverage. In this modeling exercise, we compute the expected loss for nonparametric densities using Gaussian kernels (the choice of width is described in the methodology section) and integrating using the trapezoid rule. We then calculate the loss to coverage ratios (effectively actuarially fair premiums) for an assumed expected yield and compare them among ENSO phases and also to the rates generated by assuming normal yield density. The rates are calculated for a hypothetical expected yield calculated using the series' trend from a robust regression. The omission of unusually low (catastrophic) yields in the non-parametric densities/distributions drives down the estimated premium rates. Without reliable information on how ENSO phases affect crop failures, there is no reason to try to accommodate them here. Thus, comparison with the actual GRP premiums is pointless at this point, and the term "premiums" is used interchangeably with "expected losses".

Below are two tables showing expected losses for Colquitt county data.

<u>Simulated</u> Yield Estimations of Expected Loss to Coverage Ratios by ENSO Phase

Legend: ___ - largest values; ___ - smallest values.

Actual Ranges			Expected Loss to Coverage Ratio				
MEAN	2509.886		All Years	ElNino	LaNina	Neutral	Normal
COVERAGE	70%	1756.92	8.51%	5.99%	8.30%	9.63%	6.59%
	80%	2007.909	11.39%	8.05%	10.71%	12.79%	8.89%
	90%	2258.897	14.49%	10.52%	13.29%	16.05%	11.51%

Colquitt County Estimations of Expected Loss to Coverage Ratios by ENSO Phase

Actual Ranges (smaller rates)			**Expected Loss to Coverage Ratio**				
Av. Yield	3260.508		All years	ElNino	LaNina	Neutral	Normal
COVERAGE	80%	2608.406	0.63%	0.01%	0.07%	1.21%	0.75%
	90%	2934.457	2.31%	0.92%	0.99%	3.51%	2.44%

The big difference between the simulated and actual data rates is due to (1) the averaged nature of the county data (only systemic risk), (2) the presence of much more stable irrigated yields in the county data, and (3) the broader range of simulated (individual) yields.

These differences notwithstanding, in this research, we are interested exclusively in the ordinal properties of the expected losses, i.e., in their differences between ENSO cycles. There is clearly a pattern in the differences among the expected losses that persists in both the simulated and the empirical data. In both cases, premiums (losses) are clearly the highest in the Neutral years and the lowest in the ElNino years for all coverage levels. This is a result of the generally lower mean and higher kurtosis in ElNino years noted in the previous section, meaning that the bulk of the mass is immediately to the left of the series average, hence the higher expected loss. Predictably, the expected losses estimated with pooled data fall in between the extremes. It is also worth noting that the expected losses based on the assumption of normality (using the series' mean and std) are relatively higher for the actual data and for the simulated data, which is probably a result of a smaller span of the former.

The premiums (losses) are also bigger in LaNina than in ElNino years. This is because the ElNino distribution is less skewed (although the variance seems counter-intuitive). Obviously, estimated loss probabilities and premium rates approach zero in spite of the fact that the true expected losses are likely to be positive, albeit small. Catastrophic loss probability can be constructed to address this issue (see Goodwin and Ker). However, investigation of the ENSO impact on catastrophic crop losses is beyond the scope of this paper.

The suggested explanation above also applies to the rest of the actual county data, which preserves the ordinal properties of the estimates. The table below contains analogous estimates for the rest of the counties.

Expected Loss to Coverage Ratios by ENSO Phases

			Non-Parametric Densities				Normal Parametric, All Years
Irwin	Av. Yield	3413.835	**All Yrs**	**ElNino**	**LaNina**	**Neutral**	
Coverage,%	70%	2389.685	0.36%	0.00%	0.00%	0.67%	1.19%
	80%	2731.068	1.08%	0.03%	0.08%	1.85%	2.87%
	90%	3072.452	2.94%	0.76%	1.18%	4.36%	5.24%
Worth	Av. Yield	3152.524					
Coverage,%	70%	2206.767	0.13%	0.00%	0.00%	0.24%	0.12%
	80%	2522.019	0.87%	0.00%	0.18%	1.48%	0.82%
	90%	2837.272	2.69%	0.09%	1.64%	3.88%	2.61%
Coffee	Av. Yield	3325.56					
Coverage,%	70%	2327.892	0.01%	0.00%	0.00%	0.03%	0.01%
	80%	2660.448	0.50%	0.02%	0.07%	0.81%	0.54%
	90%	2993.004	2.39%	0.81%	1.22%	3.17%	2.20%
Berrien	Av. Yield	3125.975					
Coverage,%	70%	2188.183	0.15%	0.00%	0.00%	0.29%	0.24%
	80%	2500.78	0.52%	0.00%	0.00%	0.96%	1.06%
	90%	2813.378	1.38%	0.09%	0.20%	2.31%	2.96%
Brooks	Av. Yield	3273.193					
Coverage,%	70%	2291.235	0.01%	0.00%	0.00%	0.03%	0.01%
	80%	2618.554	0.27%	0.00%	0.12%	0.45%	0.39%
	90%	2945.874	1.74%	0.85%	1.41%	2.10%	1.81%
Bulloch	Av. Yield	3147.958					
Coverage,%	70%	2203.571	0.39%	0.00%	0.32%	0.54%	0.56%
	80%	2518.366	1.37%	0.00%	1.48%	1.51%	1.68%
	90%	2833.162	3.28%	0.82%	3.68%	3.13%	3.81%
Tift	Av. Yield	3240.229					
Coverage,%	70%	2268.16	0.32%	0.00%	0.00%	0.58%	0.35%
	80%	2592.183	1.10%	0.00%	0.19%	1.96%	1.34%
	90%	2916.206	2.93%	0.38%	1.78%	4.46%	3.37%

Decatur	Av. Yield	4233.343					
Coverage,%	70%	2963.34	0.00%	0.00%	0.00%	0.00%	0.00%
	80%	3386.674	0.00%	0.00%	0.00%	0.00%	0.00%
	90%	3810.009	0.52%	0.06%	0.26%	0.70%	0.55%
Early	Av. Yield	3564.412					
Coverage,%	70%	2495.088	0.00%	0.00%	0.00%	0.00%	0.00%
	80%	2851.53	0.07%	0.00%	0.00%	0.14%	0.08%
	90%	3207.971	0.92%	0.03%	0.87%	1.28%	1.02%
Mitchell	Av. Yield	3916.57					
Coverage,%	70%	2741.599	0.00%	0.00%	0.00%	-0.29%	0.00%
	80%	3133.256	0.01%	0.00%	0.00%	0.04%	0.01%
	90%	3524.913	0.81%	0.23%	0.06%	1.08%	0.69%
Miller	Av. Yield	4038.246					
Coverage,%	70%	2826.772	0.00%	0.00%	0.00%	0.00%	0.00%
	80%	3230.597	0.08%	0.00%	0.00%	0.19%	0.12%
	90%	3634.421	0.92%	0.05%	0.44%	1.43%	1.11%
Baker	Av. Yield	4010.709					
Coverage,%	70%	2807.496	0.00%	0.00%	0.00%	0.00%	0.00%
	80%	3208.567	0.02%	0.00%	0.08%	0.01%	0.02%
	90%	3609.638	0.70%	0.06%	1.11%	0.73%	0.74%
Seminole	Av. Yield	3831.106					
Coverage,%	70%	2681.774	0.00%	0.00%	0.00%	0.00%	0.02%
	80%	3064.885	0.07%	0.00%	0.00%	0.14%	0.22%
	90%	3447.995	0.91%	0.42%	0.38%	1.34%	1.25%
Calhoun	Av. Yield	3735.677					
Coverage,%	70%	2614.974	0.00%	0.00%	0.00%	0.00%	0.00%
	80%	2988.542	0.05%	0.01%	0.00%	0.09%	0.07%
	90%	3362.109	0.87%	0.91%	0.48%	0.96%	1.07%
Burke	Av. Yield	3607.287					
Coverage,%	70%	2525.101	0.18%	0.00%	0.00%	0.39%	0.21%
	80%	2885.83	0.96%	0.00%	0.44%	1.52%	0.92%
	90%	3246.558	3.18%	0.67%	2.36%	3.92%	2.70%
Randolph	Av. Yield	3495.959					
Coverage,%	70%	2447.171	0.00%	0.00%	0.00%	0.00%	0.00%
	80%	2796.767	0.01%	0.00%	0.00%	0.02%	0.01%
	90%	3146.363	0.73%	0.08%	0.72%	0.97%	0.75%

The differences in the absolute values of expected losses among the counties are due to different variance of the county series (ranges from 387 to about 650), which is a universal factor for "inter-county" estimates (ENSO phases for the same county) but differs between the counties.

Again, it is important to note that these results are not readily comparable to the actual premiums because the latter are not likely to be actuarially fair and because our data do not include catastrophic losses (usually defined as yields more than two standard deviations below last four years' average). The most important finding is that they differ among the ENSO phases, and that can only be explained by the differences in their true non-parametric (as opposed to assumed theoretical) densities, even if the moments of the distributions seem to be similar.

In light of these preliminary findings, repeating the same exercise for a larger geographical area and using a more refined forecasting methodology seems to be an interesting venue to pursue.

Some work has been done using ARIMA models for intrapolation and forecasting. The integrated auto-regression and moving average models have the advantage that they let the data "speak for themselves" instead of imposing an a priori theoretical structure. This is done by regressing time series data on its own lagged differences and lagged errors (references to the methodology abound in econometrics textbooks). The process of selecting ARIMA model parameters - the order of auto-regression, moving average, and difference – involves a number of steps from examining correlograms and stationarity tests (Dickey-Fuller and Phillips-Perron) to choosing the best specification to explain the data. ARIMA models have been used in crop yield estimation for GRP/area yield insurance (Skees, Black, and Barnett). If the ARIMA models fit the data well, the effects of ENSO phases can be analyzed by (1) developing a "generic" model for the series (not accounting for the ENSO differences), (2) adding ENSO dummy variables and re-estimating the model. The coefficients of the dummies are then interpreted as measures of the ENSO effects (and as predictors as well).

Most of the series showed evidence of unit roots (random walk suggested earlier) but little sign on autoregression. The moving average appears to be of the second order. Thus, the ARIMA(0,1,2) specification seems to be the most appropriate (as in Goodwin and Ker).

Other methodologies worth exploring include spline regression analysis and (double) exponential smoothing.

CONCLUSION

The ElNino Southern Oscillation (ENSO) phases have distinct impacts on the climate in the Southeastern United States. It is reasonable to expect that they also affect crop yields. The magnitude of these effects may be significant enough to accommodate them in agricultural decision making.

In this paper, we report preliminary results of non-parametric analysis of peanut yield distributions in the Southwest Georgia. The focus of the analysis is on establishing ENSO-dependent differences in the yield distributions and on evaluating their implications for area yield crop insurance (GRP) the expected losses for which are calculated using county average yield series. The results of kernel density estimates of simulated yield data based on actual weather realizations in Colquitt county, GA, and actual county yield time series from the same geographical area suggest that there are distinct similarities between the simulated and actual peanut yield data and that the dissimilarities can be explained by known factors.

We also find that, while the non-parametric densities of the historical county average yield series may not always differ significantly in their means and higher moments between ENSO phases, the "cumulative" difference in (non-parametric) densities seems to matter for area crop insurance calculations. In particular, the expected loss to coverage ratios are consistently the highest during the Neutral years and the lowest during the ElNino years for all 17 counties and for simulated yield series. At the same time, the ratios calculated for pooled data (all ENSO phases) using non-parametric densities seem to be only slightly higher than those calculated using normal distribution. These findings should be more relevant for the area yield insurance as opposed to the APH arrangements as the yield data used in designing contracts for the former reflects the systemic risk more dependent on climate than on the farm-level, basis risk factors accommodated in the APH plans.

References:

Barnett, B.J., J.R. Black, Y. Hu, and J.R. Skees. 2005. "Is Area Yield Insurance Competitive With Farm Yield Insurance?" *Journal of Agricultural and Resource Economics*, 30(2): 285-301.

Boote, K.J., J.W. Jones, and G. Hoogenboom. 1998. “Simulation of Crop Growth: CROPGRO Model,” *Agricultural Systems Modeling and Simulation*, R.M. PEart and R.B. Curry, Eds., Marcel Dekker, New York, 113-133.

Cabrera, V.E., C.W.Fraisse, D. Letson, G. Podesta, J. Novak. “Impact of Climate Information on Reducing Farm Risk by Optimizing Crop Insurance Strategy,” Working Paper, University of Miami.

Fraisse, C.W., A. Garcia y Garcia, J.L. Novak, J.W. Jones, and G. Hoogenboom. “Using Crop Models and ENSO-Based Climate Forecast to Aid in Peanut Crop Insurance Decisions,” Working Paper, University of Florida.

Garcia y Garcia, A., G. Hoogenboom, L.C. Guerra, J.O. Paz, and C.W. Fraisse. “Analyzing Long-Term Historical Peanut Yield in Georgia with a Crop Simulation Model,” Working Paper, University of Georgia.

Goodwin, B.K., and A.P. Ker. 1998. “Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts,” *American Journal of Agricultural Economics*, 80:139-153.

Hansen, J.W., A.W. Hodges, and J.W. Jones. 1998. “ENSO Influences on Agriculture in the Southeastern United States,” *Journal of Climate*, 11:404-411.

Jones, J.W., G. Hoogenboom, C.H. Porter, K.J. Boote, W.D. Batchelor, L.A. Hunt, P.W. Wilkens, U.Singh, A.J. Gijsman, and J.T. Ritchie. 2003. “DSSAT Cropping System Model,” *European Journal of Agronomy*, 18:235-265.

Li, Q., J.S. Racine. *Nonparametric econometrics*. Princeton University Press, 2007.

Mavromatis, T., S.S. Jagtap, and J.W. Jones. 2002. “ElNino Southern Oscillation Effects on Peanut Yield and Nitrogen Leaching,” *Climate Research*, 22:129-140.

Mjelde, J.W., H.S.J. Hill, and J.F. Griffiths. 1998. “A Review of Current Evidence on Climate Forecasts and Their Economic Effects in Agriculture,” *American Journal of Agricultural Economics*, 80(5):1089-1085.

Podesta, G., D. Letson, C. Messina, F. Royce, R.A. Ferreyra, J.W. Jones, J.W. Hansen, I. Llovet., M. Grondona, and J. O’Brien. 2002. “Use of ENSO-related Climate Information in Agricultural Decision Making in Argentina: A Pilot Experience,” *Agricultural Systems*, 74(3):371-392.

Sherrick, B.J., F.C. Zanini, G.D. Schnitkey, and S.H. Irwin. 2004. “Crop Insurance Valuation under Alternative Yield Distributions,” *American Journal of Agricultural Economics*, 86(2):406-419.

Silverman, B.W. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York, 1986.

Skees, J.R., J.R. Black, and B.J. Barnett. 1997. “Designing and Rating an Area Yield Crop Insurance Contract,” *American Journal of Agricultural Economics*, 79(2):430-438.