A Study on the Effectiveness of Seasonal Forecasting of Precipitation and Temperatures for Soybean Yields in Central Iowa

Brian Folkers, Chase Van Valkingburg, Elizabeth Lichtenberg

Department of Meteorology, Iowa State University

Introduction

Soybean yields contribute over $3.5 billion annually to the agriculture economy in Iowa (www.usda.gov). As a result, the state of Iowa plants more acres of soybeans than any other state. This leads to the fact that soybean yields per acre are greater in Iowa than anywhere else in the world (www.usda.gov). In order to produce these maximum yields, seasonal forecasts are used to predict in advance what types of weather events could play a role in soybean development. Weather events such as extreme flooding, droughts, El Nino and La Nina play a substantial role as to how bountiful soybean yields will be across Central Iowa. In this study, we will explore how the seasonal forecast for precipitation, temperature, and growing degree days (GDD) compare to the actual reported values of these three variables, looking at data from 1979 –1999, for the months of April, May, June, July, August and September. Next, we will compare this data to yearly soybean yields to see if there is a correlation between the seasonal forecast and how abundant soybean yields were. Finally, we will look for a correlation to see how accurate the seasonal forecasts are compared to actual measured values of precipitation, temperatures and GDD's and their affect on soybean yields.

Data and Methods

Observed Data

In order to collect data for crop yields, precipitation, temperature, and GDD for Central Iowa, we selected four spatially adjacent counties. The counties we selected were: Story, Dallas, Polk and Boone. Next, we collected annual crop yield data from the United States Department of Agriculture for these four counties. We used monthly data from April through September, for the years 1979-1999. In order to collect observed climate data (precipitation, temperature, and GDD), we selected one city that was located in each of the four counties. The cities we selected were: Ames, from Story county, Perry, from Dallas county, Ankeny, from Polk county, and Jefferson from the Boone-county area. We chose Jefferson, which is the closest city to Boone County, since it did not have a city we could use to collect desired data from. Our observed climate data was taken from the Iowa Environmental Mesonet from April through September for the same 21- year period. The observed temperatures we used were monthly averaged temperatures taken from April through September. As for the observed precipitation data, it was taken as the total growing season precipitation. To calculate the observed GGD, we used the equation GDD=(Monthly avg. temp – 50^{o}F)*30, where 50^{o}F was used as the minimum baseline temperature, for all six months, then added together and spatially averaged.

Hindcast Data

The hindcast data was taken from the seasonal forecasting web page. In order to collect precipitation and temperature data, we first found two grid points in Central Iowa representative of our four county area. The spatially averaged grid points that we used were: 42N 93.5W and 41N 94.5W. The data was taken from the same 21-year period, 1979-1999, and for the same six months, April through September, as in the observed data process. The hindcast temperatures we used were monthly averaged temperatures taken from April through September. The hindcast precipitation was spatially averaged and converted to inches over the entire growing season. To calculate the hindcast GDD, we used the equation GDD=(Monthly avg. temp – 50^{o}F)*30, where 50^{o}F was used as the minimum baseline temperature, for all six months, then added together and spatially averaged.

Analysis

In Figure 1, the estimated GDD is compared to the observed GDD to see if it was acceptable to use. Looking at the graph, it is shown that the estimated GDD was acceptable to use so we could now use GDD as part of the hindcast output since it was not previously calculated for us. The graph in Figure 2 compares hindcast precipitation to the observed precipitation. As shown on the first line graph, GDD are greatly overestimated. The reason for comparing these data points was to see how accurately the model predicted temperature, and therefore growing degree days could be interpreted from that model. Figure 2 shows that the models greatly underestimate precipitation. In most cases the models only effectively predict half of the amount of actual precipitation. Figures 3a and 3b compare observed GDD to soybean yield to see if the GDD affected the yields and if they did by how much. Figures 4a and 4b show observed precipitation versus soybean yield. We plotted these graphs to see if extreme weather events caused bias in the linear regression line. In Figures 5 and 6 we implemented the hindcast data.

Starting with these two graphs, we studied the plot to see what kind of correlation is presented. This was done to see how the predicted graphs compared to the actual graphs. Figure 7 shows simply the actual yield each year. There seemed to be a bias present because our trendline increased per year. We assumed that there was a bias present because increases in agricultural technology will give better yields in general each year. We then assumed that this bias that was present would cause bias in our linear regression charts, so we eliminated that bias by subtracting the trendline. Figure 8 shows this unbiased trendline, with an R^2 of zero. With this new unbiased distribution, we re-graphed yield vs. GDD and yield vs. precipitation to see if our linear regressions looked different. This analysis is shown in figures 9a and 9b, and figures 10a and 10b, respectively.

Results

We first decided to compare the hindcast data to actual observed data. In Figure 1 actual GDD is compared to monthly average estimated GDD and forecasted GDD. There is a correlation between the actual GDD and the estimated GDD. The difference between the two is minimal and our estimated GDD was a little lower than the actual. There is a definite trend between hindcast GDD and actual GDD, but the hindcast overestimates the GDD by a large margin. After deciding that there is visible bias between the hindcasts and the observed data, we looked at correlations between GDD and precipitation and yields

Figures 3 and 4 show little or no correlation for precipitation and GDD vs. yield. Their R^2 value is very small. However, when looking at changes in precipitation from year to year, there seemed to be a negative correlation on the line graph. This correlation was not seen when yield vs. GDD was plotted.

After analyzing yields vs. precipitation and GDD, we decided to see if the models did a comparable job. We assumed that they would not, because of the bias present in both the temperature models and the precipitation models. Figures 5 and 6 show this analysis, and the graphs look much different than the observed data graphs did. This lead us to believe that the model forecasts would not do that great of a job of predicting yields, unless the bias could be found and corrected.

While looking at bias, we realized that when we plotted the yield for each year, the trendline increased with an R^2 value of 0.2641. This would potentially indicate a bias in our original trendline, which was plotted against GDD and precipitation. This bias is shown in figure 7. Since agricultural technology has increased each year, we assumed that overall yields would increase with time. We decided to take this bias out of our plot by subtracting the trendline from our graph. The R^2 value is not too low, but we wanted a flat trend line. To correct this bias we subtracted the trend line equation:

$$y = 285007x - 5 \cdot 10^8$$

and found a new and unbiased trend line. Figure 8 shows a new unbiased distribution of yields each year, with an R^2 value of zero. We then plotted this unbiased distribution against GDD and precipitation to see if the graphs looked different. Figures 9a and 9b show GDD vs. the new distribution. Figure 9a shows a nice bell looking distribution, which would indicate that too few GDD is bad, but so is too many GDD. The line graph on figure 9b looks similar to the line graph on figure 3b.

Figures 10a and 10b show the new distribution against precipitation. The bell shaped curve is even more evident in this plot, which would indicate that too little precipitation is bad, as in the drought of 1988. Too much precipitation is also bad for yields, as in the flood of 1993. Most data points fall in the middle of the graph. Also in figure 10b, the negative correlation is more pronounced than it was in figure 3b. When precipitation changes from year to year, the yield changes in the opposite direction.
In most of the years, when precipitation decreases compared to the previous year, the yield increases compared to the previous year. It also works when the precipitation increases compared to the previous year, the yield decreases compared to the previous year. This would possibly indicate a negative correlation that would not necessarily be as easily seen on a linear regression chart. Overall with this new unbiased trendline linear regressed again GDD and precipitation, most of our data points fell with in the center of the bell curve, which is good for higher values for R^2 on a trendline of the second order. We used the second order trendline because of the bell shaped curve we obtained. This trendline makes sense because you would not want an over abundance of precipitation, but you would not want too little precipitation either. The same results were shown for GDD but not as convincingly. The R^2 value was a little lower than it was for the precipitation graph. Once again though, GDD is an indication of overall average temperatures, and for a good yield too little heat would be bad for crop growth, but too much heat would also be bad for crop growth.

Conclusion

Overall the index seemed to work best when the precipitation fluctuated more from year to year. When the precipitation decreased a significant amount from the previous year, the yield increased a significant amount from the previous year. This, however, did not occur in all situations. This can be easily seen in figure 3 and figure 10. The Growing Degree Days did not appear to have any conclusive data. Our R^2 value is much better for precipitation than it is for growing degree days. We believe from our data that a yearly fluctuation in precipitation is the most important factor in having good crop yields. This is most evident from our data in 1993 and 1994. In 1993, Iowa received the most precipitation in our period and the second lowest yield. In 1994, Iowa received a below average amount of precipitation and recorded the highest yield in our period.

We also decided that it is possible that the bias with the trend line and yields was caused by the improvement in technology over the past 20 years. However, errors in our analysis might not only depend on the biased yield distribution. Crop yields can depend on many more things than GDD and precipitation. Other important factors are soil moisture content at the beginning and end of the growing season, heat waves that occurred for short amounts of time, floods, droughts, and killing frosts at the end of the season. Most of this data can slip through the coarse data provided by monthly averaged data, which is also spatially averaged.

Monthly average seasonal forecasts have very good potential to be accurate at predicting soybean yields. More work can be done to find statistical correlations between not only GDD and rainfall, but also other factors such as the ones listed above. Biases are present in forecast data, but the trends are easily seen in our graphs. If these biases can be determined through future studies, crop yields could be determined from seasonal forecasts.

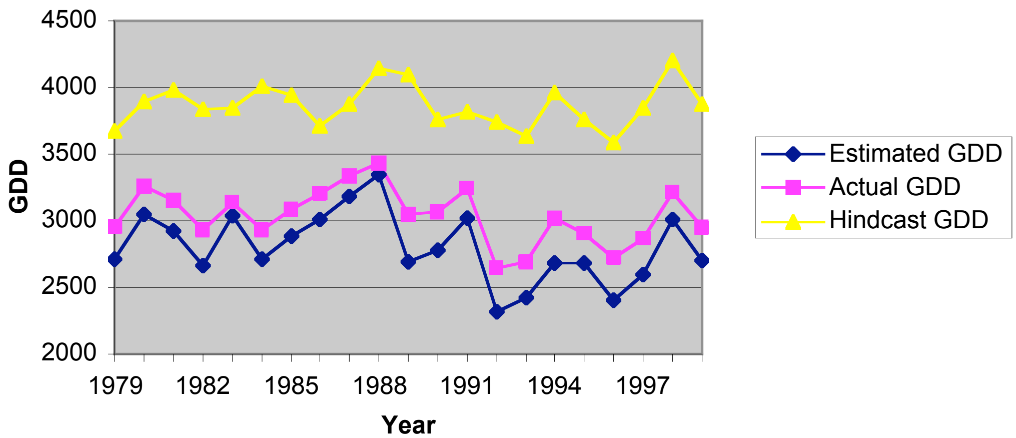
References

Dept. of Geological and Atmospheric Sciences, cited 2003: Iowa State University Meteorology 455/555 Course Seasonal web site. [Available online at http://www.meteor.iastate.edu/classes/mt455.]

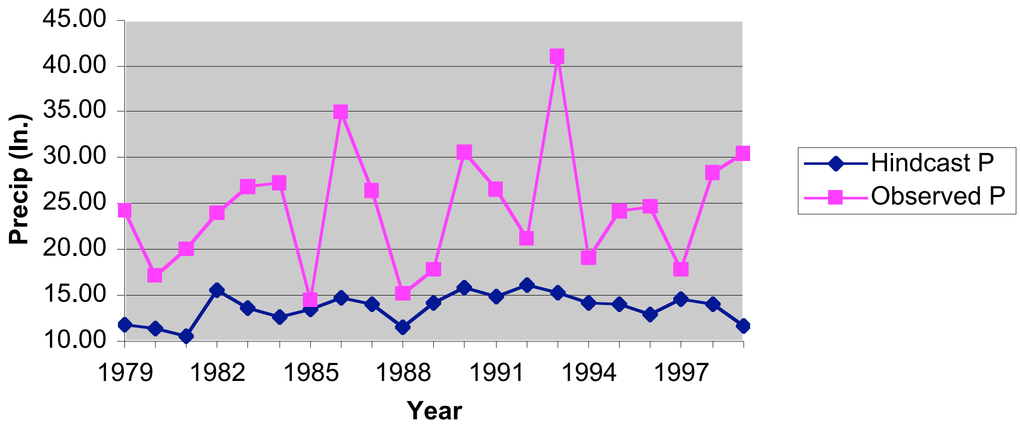
Todey, Dennis, cited 2003: Iowa Environmental Mesonet. [Available online at http://mesonet.agron.iastate.edu/climodat/table.html.]

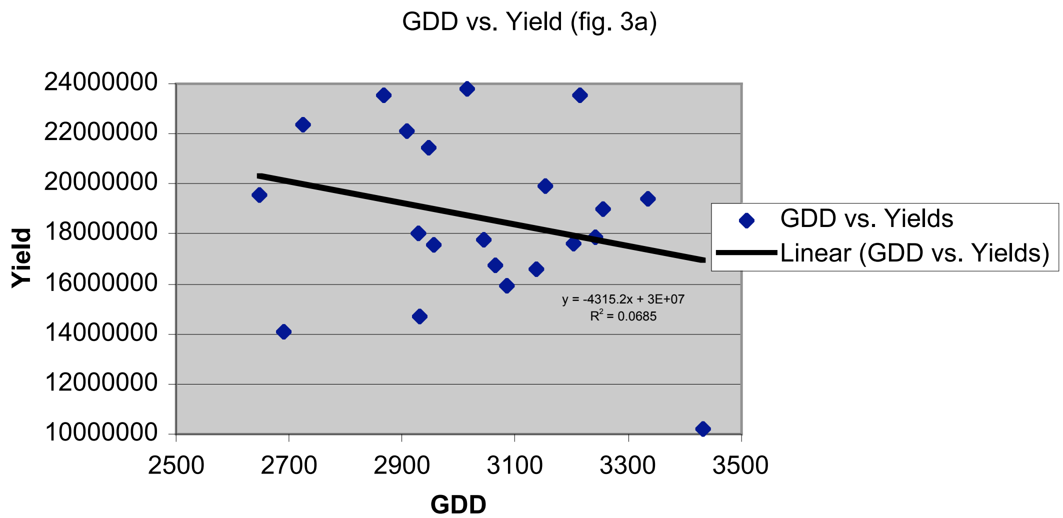
United States Department of Agriculture, cited 2003. [Available online at http://www.usda.gov.]

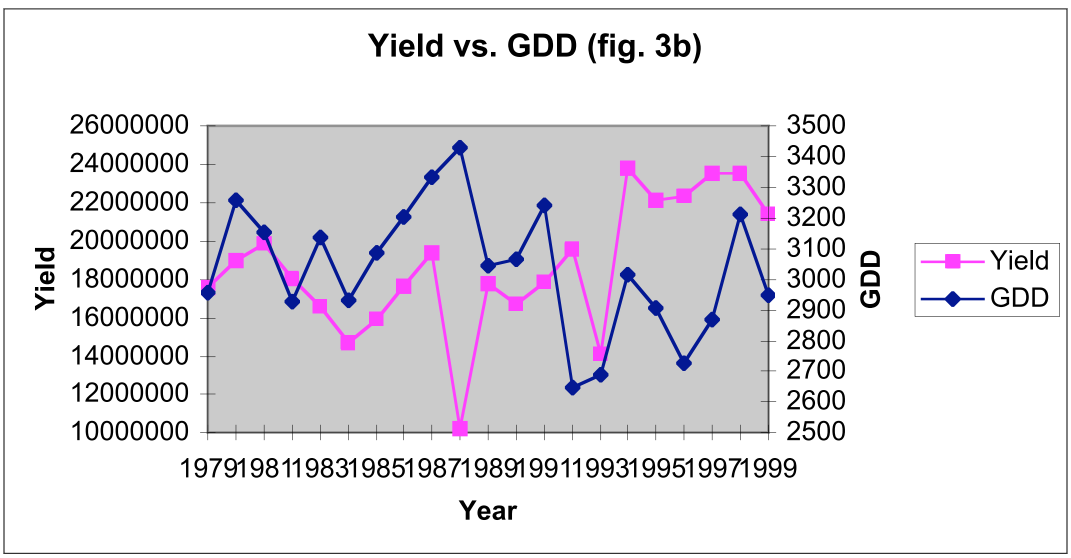
Comparison of GDD Sources (fig. 1)

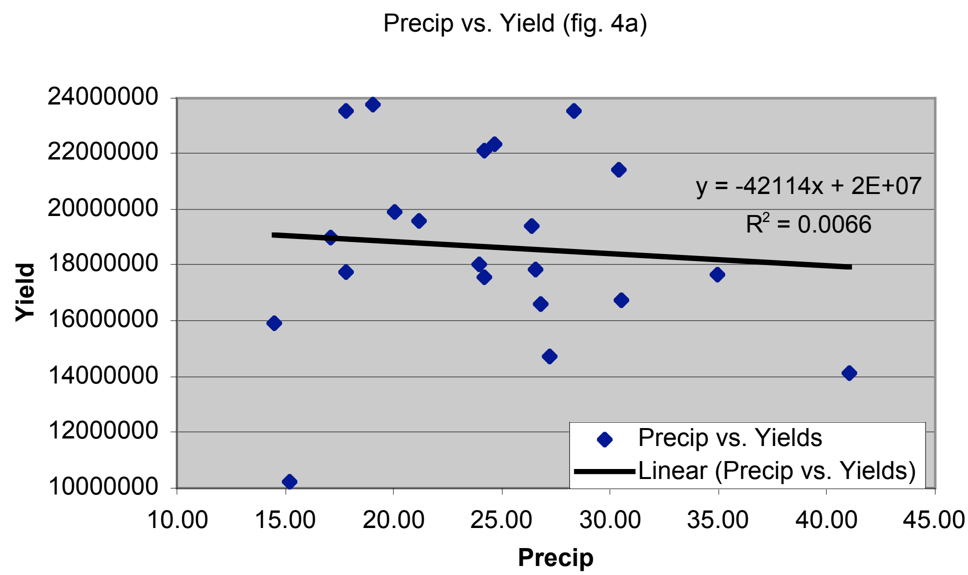


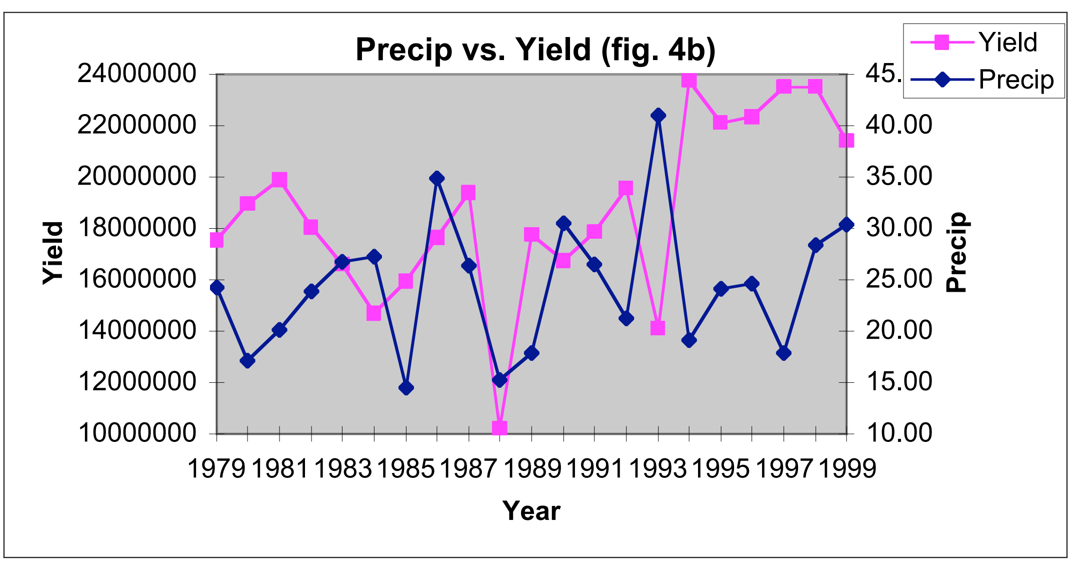
Hindcast P vs. Observed P (fig. 2)

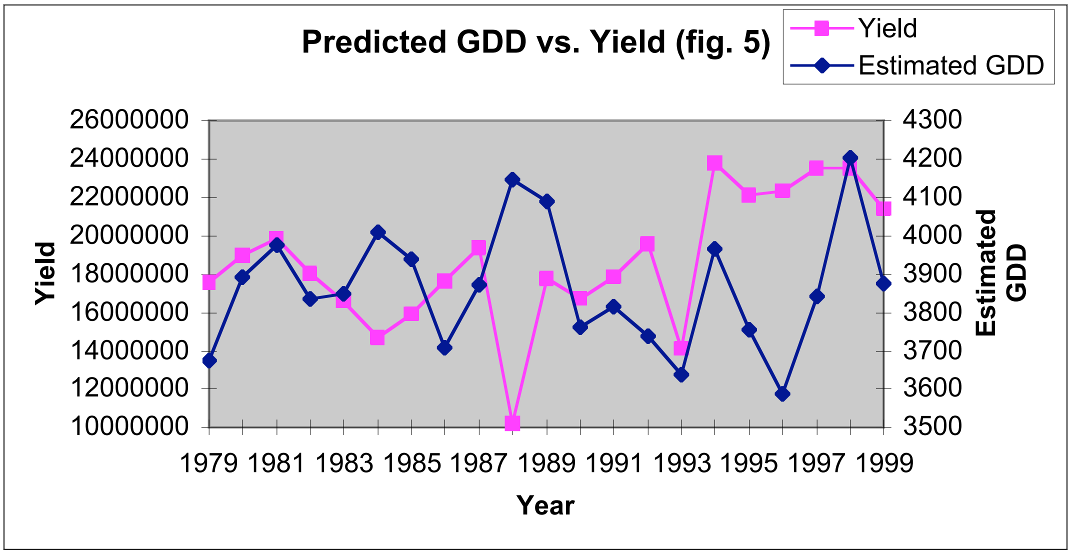


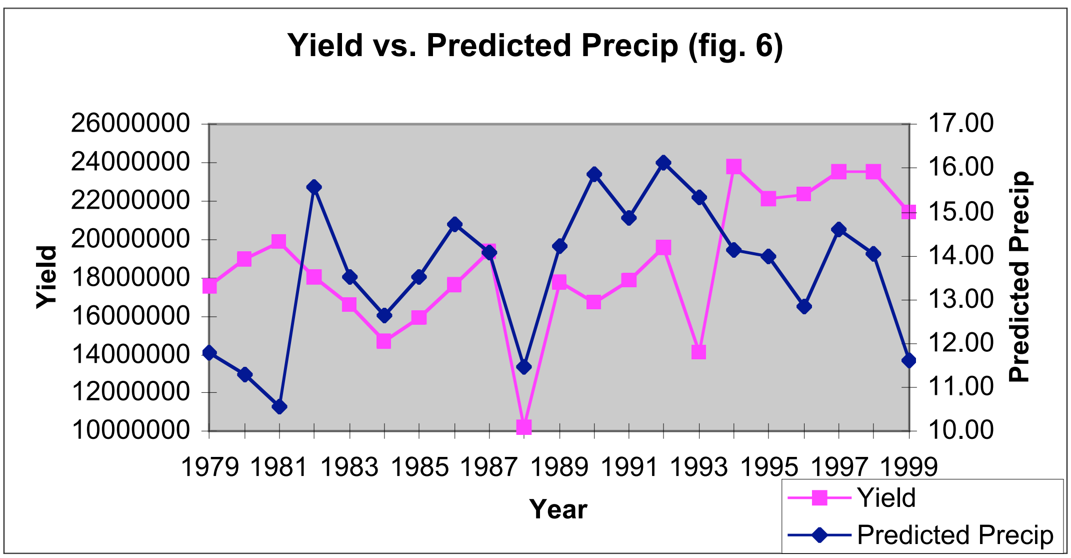

GDD vs. Yield (fig. 3a)
Yield
24000000
22000000
20000000
18000000
16000000
14000000
12000000
10000000
2500
2700
2900
3100
3300
3500
GDD
y = -4315.2x + 3E+07
R^2 = 0.0685
GDD vs. Yields
Linear (GDD vs. Yields)


Yield vs. GDD (fig. 3b)
Yield
26000000
24000000
22000000
20000000
18000000
16000000
14000000
12000000
10000000
GDD
3500
3400
3300
3200
3100
3000
2900
2800
2700
2600
2500
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
Year
Yield
GDD

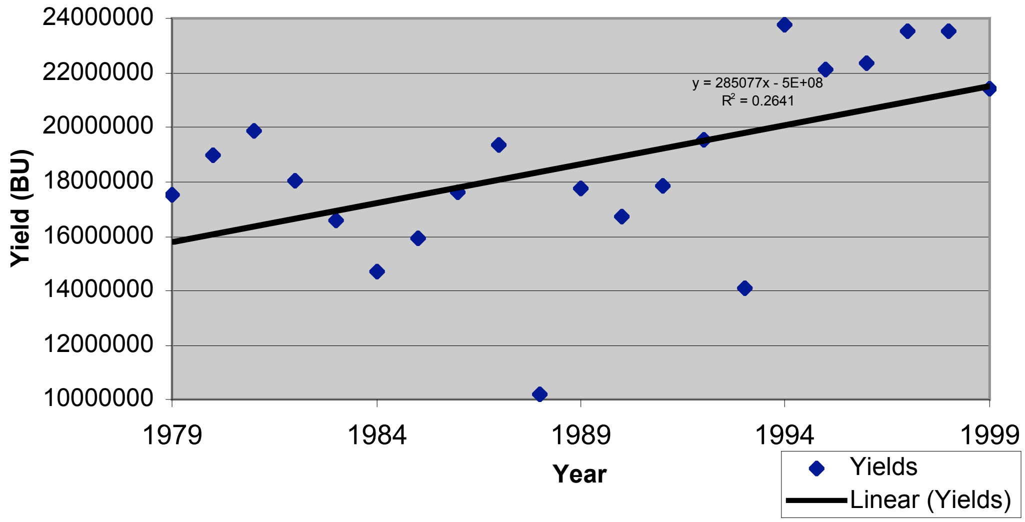

Precip vs. Yield (fig. 4a)
y = -42114x + 2E+07
R² = 0.0066
Precip vs. Yields
Linear (Precip vs. Yields)
Yield
Precip
24000000
22000000
20000000
18000000
16000000
14000000
12000000
10000000
10.00
15.00
20.00
25.00
30.00
35.00
40.00
45.00

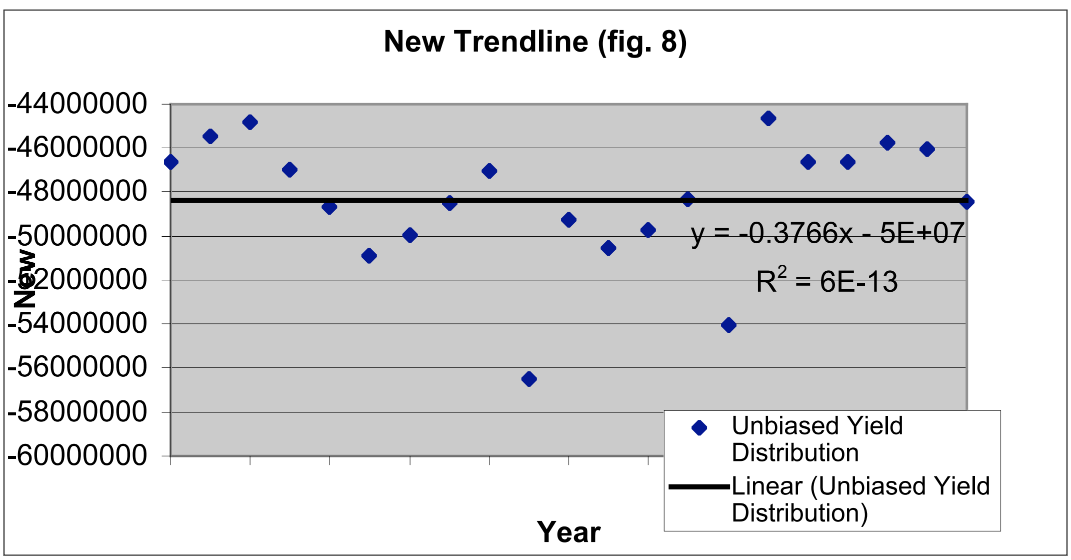

Precip vs. Yield (fig. 4b)
Yield
Precip
Year
24000000
22000000
20000000
18000000
16000000
14000000
12000000
10000000
45.
40.00
35.00
30.00
25.00
20.00
15.00
10.00
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999

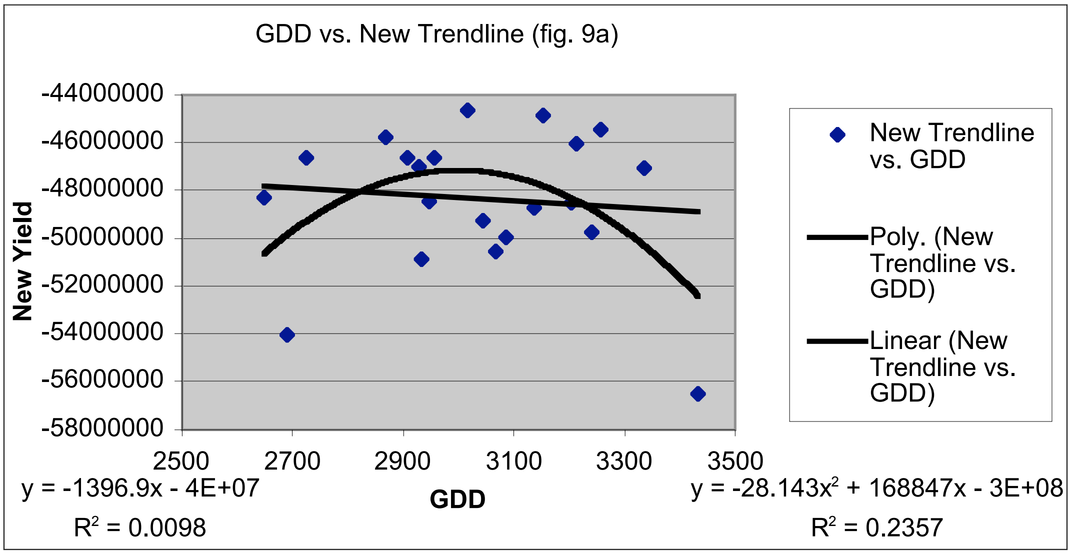

Predicted GDD vs. Yield (fig. 5)
Yield
Estimated GDD
26000000
24000000
22000000
20000000
18000000
16000000
14000000
12000000
10000000
4300
4200
4100
4000
3900
3800
3700
3600
3500
Yield
Estimated GDD
1979 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999
Year

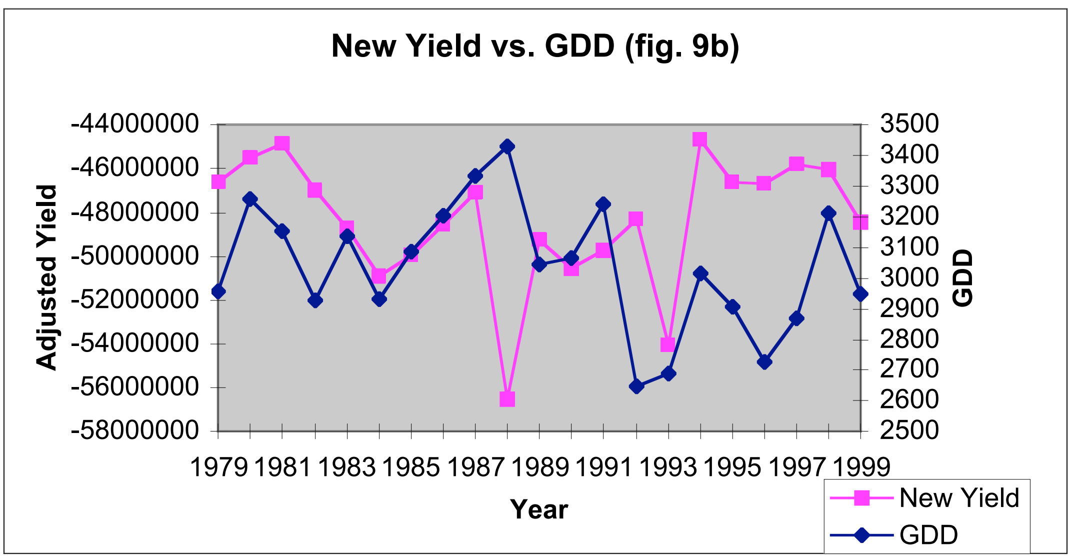

Yield vs. Predicted Precip (fig. 6)
26000000
24000000
22000000
20000000
18000000
16000000
14000000
12000000
10000000
17.00
16.00
15.00
14.00
13.00
12.00
11.00
10.00
Yield
Predicted Precip
1979 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999
Year
Yield
Predicted Precip

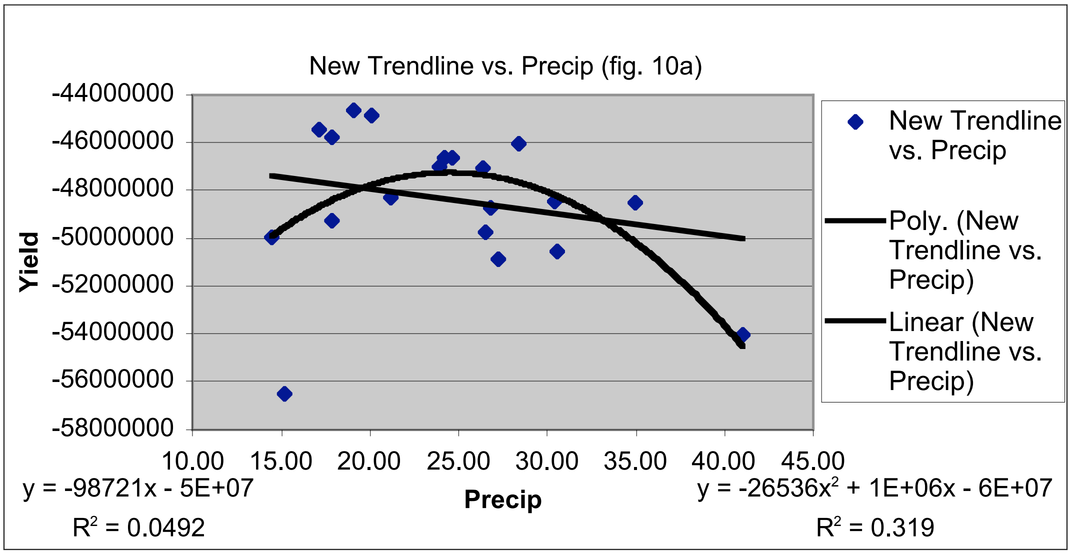
Yields (fig. 7)

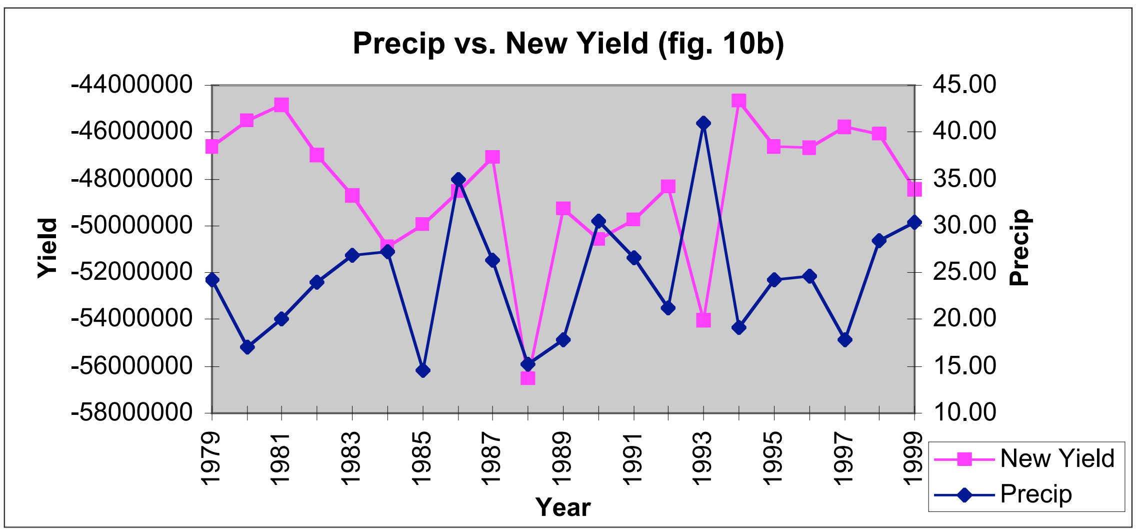

y = 285077x - 5E+08
R^2 = 0.2641
24000000
22000000
20000000
18000000
16000000
14000000
12000000
10000000
Yield (BU)
1979
1984
1989
1994
1999
Year
Yields
Linear (Yields)


New Trendline (fig. 8)
-44000000
-46000000
-48000000
-50000000
-52000000
-54000000
-56000000
-58000000
-60000000
New
y = -0.3766x - 5E+07
R^2 = 6E-13
Unbiased Yield Distribution
Linear (Unbiased Yield Distribution)
Year


GDD vs. New Trendline (fig. 9a)
-44000000
-46000000
-48000000
-50000000
-52000000
-54000000
-56000000
-58000000
New Yield
2500
2700
2900
3100
3300
3500
GDD
New Trendline vs. GDD
Poly. (New Trendline vs. GDD)
Linear (New Trendline vs. GDD)
y = -1396.9x - 4E+07
R² = 0.0098
y = -28.143x² + 168847x - 3E+08
R² = 0.2357


New Yield vs. GDD (fig. 9b)
-44000000
-46000000
-48000000
-50000000
-52000000
-54000000
-56000000
-58000000
Adjusted Yield
3500
3400
3300
3200
3100
3000
2900
2800
2700
2600
2500
GDD
1979 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999
Year
New Yield
GDD


New Trendline vs. Precip (fig. 10a)
-44000000
-46000000
-48000000
-50000000
-52000000
-54000000
-56000000
-58000000
Yield
10.00
15.00
20.00
25.00
30.00
35.00
40.00
45.00
Precip
New Trendline vs. Precip
Poly. (New Trendline vs. Precip)
Linear (New Trendline vs. Precip)
y = -98721x - 5E+07
R2 = 0.0492
y = -26536x2 + 1E+06x - 6E+07
R2 = 0.319


Precip vs. New Yield (fig. 10b)
-44000000
-46000000
-48000000
-50000000
-52000000
-54000000
-56000000
-58000000
Yield
45.00
40.00
35.00
30.00
25.00
20.00
15.00
10.00
Precip
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
Year
New Yield
Precip