VRB 83-09

REVIEW OF THE MICHAELS MODEL FOR PREDICTING CORN YIELDS IN VIRGINIA

by Jeanne L. Sebaugh, Ph.D.

The review of this model was undertaken using materials provided by Bill Arends as follows:

1. Progress Report, Statistical-Dynamic Models for Virginia Soybean and Corn Yields by Michaels and Smith, June 1, 1982,

2. Summary of Results to Date, September 27, 1982,

3. Summary, October 19, 1982,

4. Binder of SPSS computer runs, and

5. Box of punched cards containing data for the agricultural and meteorological variables from 1930 to 1979 by climatic division.

In addition, Dr. Michaels and his associate, Tom Smith, answered some of my initial questions over the telephone.

As a first step in the review process, an attempt was made to verify the data values provided in the punched cards. The Yield Evaluation Section (YES) in Columbia does not have any historic agricultural data for Virginia. Therefore, a computer printout of the data provided at the climatic division (CD) level and its aggregation to the state level has been included for verification with the official Virginia SSO records. The agricultural variables punched in the cards are acreage planted, production, and yield per planted acre. The area and production values are sums of county level values, and the yield value is the quotient of the two sums.

It is suggested that the first check be made of the values which have been calculated at the state level. Each year six climatic division area and production values were summed and then production was divided by area to obtain a yield per planted area. It is important that these state level values be confirmed by referring to the appropriate state publications. Another type of error in the climatic division values could occur if the proper counties were not summed within each climatic division. To check this, one needs to independently program the summing of the appropriate counties within each climatic division.

Another question regards the definition of corn yield used by the Virginia SSO over the period 1930-1979. Although corn for grain data is of current interest, in some states early corn yield and associated data are only available for corn for all purposes. It has only been since 1956 in Iowa and 1954 in Illinois that area, production, and yield have been reported separately for corn harvested for grain. Information on whether there has been any change in the reporting of corn yield, production or area estimates over the 1930 to 1979 time period in Virginia needs to be obtained. Also, it would be helpful in future analyses if the harvested area variable were also made available.

The climatic division monthly meteorological variables were checked by comparing the punched card values with values available to YES on the NOAA (National Oceanic and Atmospheric Administration, Department of Commerce) computer system in Suitland, Maryland. Their data were obtained from the National Climatic Center in Asheville, North Carolina. As shown in Table 1, most of the differences found are with the monthly precipitation variable. In the instances where the NOAA values are 10 or more, the University of Virginia precipitation error is due to an incorrect specification of the card columns for reading the precipitation data in the University of Virginia SPSS program. Whereas the card columns are specified as 11-15, 16-20 through 51-55, a change of the specification to 11-14, 15-19 through 50-54 would result in the values being read properly. (There are no occurrences of January precipitation greater than or equal to ten inches.) The reasons for the other precipitation differences are not known. The temperature differences in 1979 appear to be the result of the values being offset a month in the punched cards.

NOAA's data begin in 1931 rather than 1930, so the monthly weather data for 1930 could not be checked. Also, NOAA does not have a daily weather data base for Virginia available. Therefore, YES did not attempt to validate the weekly weather variables.

In deriving the weather variables, Michaels subtracts the mean over all years of the unadjusted variable value. Slight discrepancies were found with the mean values observed in the SPSS program. Table 2 shows the means from the University of Virginia SPSS program, from a SAS program run by YES using the weather data provided by Michaels and from a SAS program run by YES using the NOAA weather data.

The corn yield model reviewed here is the non-dynamic version. The list of weather variables supplied to the SPSS stepwise program included thirty-four of the thirty-eight variables in Table 1. The May and July temperature and precipitation variables were omitted from the list (more on this later). Also included in the list are six indicator variables and six trend variables, one of each type for each climatic division. Since a no-intercept model was specified, all six indicator variables (B1-B6) entered. Also, since there is a strong trend component over the 1930 to 1979 time period across the state, all six trend variables (TR1-TR6) entered. Twelve weather variables were selected by the University of Virginia SPSS program. Table 3 gives the coefficient estimates from the SPSS program and from two YES SAS runs, one using Michaels' weather data and one using NOAA's weather data. The differences are not great. However, YES was not able to duplicate the selection of the same set of twelve weather variables using the SAS stepwise procedure. Table 4 shows the variables selected by SAS with a significance level of entry equal to 0.45 and a significance level of staying equal to 0.10. (SAS uses significance level specifications rather than the specification of F values used by SPSS.) The YES runs were made using the means in Table 2. The weather variables in Table 4 were computed using the NOAA weather data.

YES has several questions about the way Michaels specified this model and defined and interpreted the weather variables. First of all, the specification of a no-intercept model only seems to be a mathematical requirement to avoid a singular design (X) matrix upon the simultaneous inclusion of six indicator variables. However, visual inspection of the coefficient estimates for the indicator variables reveals that the estimates for CDs 4 and 5 are similar to one another and higher than the estimates for the other CDs. These coefficient values should indicate relative yielding ability of the CDs after taking into account the effect of weather. One also notices in Tables 3 and 4 that some of the trend coefficient estimates, particularly in adjacent CDs, are similar. After performing some statistical analysis, YES found that an intercept model with one indicator variable, for CDs four and five, and four trend variables, one for CD1, one for CDs 2 and 3, one for CDs 4 and 5, and one for CD6, could be justified. Table 5 gives the results of the SAS stepwise procedure using these variables. This form of model specification seems desirable for several reasons. Fewer coefficients need to be estimated. Whereas, it might have been difficult to explain the small differences in B1 to B6, it should be much easier to find an explanation for the difference in yielding ability for CDs 4 and 5 versus 1, 2, 3, and 6. Also, if some explanation can not be found for the differences in trend coefficients, perhaps a single trend term should be employed.

The first report (June 1, 1982) did not include weekly weather variables. They were added in the September 27 summary. Although information about the crop calendar for corn in Virginia was not included in any of the material received, YES was told that particular weeks were chosen because of their importance in the growth and development of corn. The first week was May 1-7 in CDs 1 and 2 and May 8-15 in CDs 3 through 6. The second week was July 1-7 and the third week was July 8-15 for all CDs.

The list of variables in the University of Virginia SPSS program for the stepwise regression includes the temperature and precipitation variables for the above three weeks, both linear and quadratic. However, the monthly variables for May and July were excluded, except as they occurred in interaction terms. This disallows the direct impact of three weeks of weather in May and two weeks of weather in July from possibly being included in the model. YES added the May and July weather variables to the list and ran the SAS stepwise procedure. The results are shown in Table 6. May and July monthly precipitation variables did enter the model and the mean square residual was reduced. Two weekly weather variables, MT1 and JT2SQ, were still included in the model.

In the June 1, 1982 report, Michaels and Smith explained the use of two and three month interaction variables by saying that "These variables are intended to account for a nonlinear crop growth response to climatic events at time scales longer than one month duration. . .". YES would like to have a more thorough explanation and justification for these variables. For example, MJT for a given year is computed as [(this year's May temperature minus the long term average May temperature) times (this year's June temperature minus the long term average June temperature)] minus the long term average of quantity in brackets above. Without the last subtraction, the quantity would be positive if both May and June temperature were either above or below their long term average and would be negative if one month were below the long term average and the other month above its long term average. Mathematically, it would make no difference which month were below and which above average. With the last subtraction included, even if the temperatures for both months are average, MJT would equal -4.412 (- long term average of unadjusted MJT).

In a similar manner, the three month interaction terms, without the final subtraction, would be positive if the values for all three months were above average or if the values for any two out of the three months were below average. The term would be negative if the values for all three months or if the value for any one of the three months were below average. The impact on predicted yield would be the same regardless of the month.

Table 7 shows the results of a SAS stepwise procedure with all interaction terms omitted from the list of variables to be included. The mean square residual did not greatly increase and the results may be much easier to interpret. If the squared terms were simply defined as the product of the linear terms, rather than subtracting their long term averages, interpretation would be enhanced. However, the mathematical results would be the same except for the estimate of the intercept term.

One of the criteria that YES uses in model evaluation is consistency with scientific knowledge. In order to evaluate the model's consistency with scientific knowledge, it would be helpful to know the Virginia crop calendar and to be given a discussion by the model developers of the model produced by their computer program, such as why the coefficients of particular variables are positive or negative.

Although YES did not examine the dynamic version of the weekly model, the assumptions upon which that model is based were investigated. The dynamic model was designed to improve yield predictions in years whose yields are very good (much higher than trend, assuming normal weather) or very bad (much lower than trend, assuming normal weather). Weather variables are analyzed in a sequential fashion, and if an extreme value occurs (either 1.5 or 2.0 standard deviations from the statewide average, depending on the variable), then the values of the linear component of the corresponding type of weather variable (temperature or precipitation) in all the months following the extreme one are set to zero. For example, if MT1 is extreme (high or low), then the value of TJUN, JT1, JT2, TAUG, JAT, JJT, and JJAT are set equal to zero. The theory upon which Michaels and Smith base this process is that certain weather which occurs after the extreme occurrence has no further effect on corn yields.

Michaels performed the above process, obtained a new model using SPSS stepwise procedures, and concluded that yield estimation was not enhanced. Actually these results do not seem surprising, even if the underlying theory were correct. If weather variable extremes result in very good or very bad corn yields, then by setting variables in following months to zero, one is also equating normal weather with very good or very bad yields. Forcing extreme yields to correspond to both extreme and normal weather values should confuse the regression process.

Using the model in Table 7, YES computed predicted yields for each year from 1930 to 1979 in two ways: first, by using each year's trend and observed weather values and, second, by using each year's trend and normal weather values. The very good years and very bad years were defined as those whose reported yield was five bushels per acre higher or lower than the predicted yield with trend and normal weather. Next, years were chosen from this group whose predicted yield using trend and observed weather was either ten bushels per acre above or below the reported yield. These are the years for which improvement in yield prediction is most needed. Table 8 shows these year/CD combinations. The top nine lines are the very good years whose yields are not predicted accurately and the bottom thirteen years are the very bad years whose yields are not predicted accurately. Standardized Z-values (Z) and the contribution to the yield estimate (Cont.) are also given for the weather variables in the Table 7 model. The Z-value is for the linear variable only. For example the Z-value for PMAY is simply PMAY divided by the standard deviation of precipitation in May. The contribution value is the product of the coefficient and the weather variable for both linear and squared terms summed together where appropriate.

First of all, one notices that there are not that many occurrences of extreme values. Secondly, if the reason for the inability to predict the good or bad yields is related to contrary contributions to the predicted yield following a month with an extreme value, one would expect the following results. In the top nine lines, months with negative contributions pulling the prediction down would be observed following a month with a large positive contribution. In the bottom thirteen lines, months with positive contributions pushing the yield prediction up would be observed following a month with a large negative contribution. Although instances of this can be seen, not enough to make a difference of ten bushels per acre and not in a consistent, usable pattern.

Some other approach seems necessary. Of course, estimating yield per planted acre instead of yield per harvested acre compounds the problem, as one needs to attempt to model the extreme stress which can lead to very low yields. In addition, low yields per planted acre can also result from situations difficult to describe with monthly or weekly weather variables, like hail or freeze damage; or from factors like disease or insect damage. YES would suggest an investigation of these possibilities, particularly for some recent years in Table 8, like 1973, 1975, and 1977, before proceeding further with a dynamic model.

Table 1. Differences between U. of Virginia and NOAA monthly weather values

Year	Climatic Division	Monthly Number	Precipitation NOAA	Precipitation U. of V.	Temperature NOAA	Temperature U. of V.
1933	1	8	10.51	0.51		
1940	2	8	10.71	0.72		
	3	8	10.75	0.73		
	5	8	11.81	1.82		
1941	5	1	1.77	2.50		
	5	2	0.39	2.64		
	5	3	1.73	3.40		
	5	4	2.60	2.89		
	5	5	1.02	3.61		
	5	6	4.33	3.57		
	5	7	6.10	3.96		
	5	8	2.44	4.01		
	5	9	1.38	3.57		
	5	10	1.10	3.01		
	5	11	1.50	2.61		
	5	12	3.82	2.89		
1942	3	8	10.67	0.66		
	4	10	12.36	2.38		
1944	3	9	12.72	2.73		
1945	1	7	13.11	3.10		
	2	7	11.93	1.92		
1955	1	8	10.79	0.77		
	2	8	11.77	1.79		
	4	8	13.78	3.76		
1966	6	9	6.18	0.19		
1971	2	10	10.00	0.00		
1972	4	6	12.76	2.74		
1976	3	10	10.75	0.76		
	4	10	10.43	0.44		
1978	1	1	6.61	6.72		
	1	9	1.69	1.58		
	2	1	8.31	8.49		
	2	3	5.47	5.32		
	2	5	4.80	4.70		
	4	1	7.52	7.38		
	5	7	4.33	4.45		
	5	10	0.91	1.02		
	6	11	3.15	3.25		
1979	1	2			30.6	49.4
	1	3			49.5	57.3
	1	4			57.4	66.4
	1	5			66.4	70.1
	1	6			70.2	76.4
	1	7			76.5	77.3
	1	8			77.4	71.3
	1	9	10.91	1.09	71.2	58.8
	1	10			58.8	54.1
	1	11			54.1	43.2
	2	10	5.79	2.92		
	3	9	11.10	1.11		
	4	8	6.50	5.49		

Table 2. Means of monthly and weekly weather (WX) variables used for variable selection in the U. of Virginia corn model

Variable		Mean		
		U. of Virginia WX Data		NOAA WX Data
Name	Description	U. of V.-SPSS	YES-SAS	YES-SAS
PPRE	Sum of Jan.-April precip.	13.080	13.175	13.160
PMAY	May precipitation	3.880	3.873	3.865
PJUN	June precipitation	3.884	3.810	3.846
PJUL	July precipitation	4.5185	4.484	4.558
PAUG	August precipitation	4.167	4.150	4.415
PPRESQ	PPRE squared	10.0609	9.795	9.929
PJUNSQ	PJUN squared	2.0368	1.997	2.258
PAUGSQ	PAUG squared	3.5289	3.442	4.564
MJP	May * June precip.	0.1628	0.167	0.210
JJP	June * July precip.	0.2240	0.228	0.173
JAP	July * August precip.	0.2723	0.219	0.141
MJJP	May * June * July precip.	-0.0302	0.060	-0.089
JJAP	June * July * Aug. precip.	-0.4077	-0.279	-0.333
TMAY	May temperature	64.060	64.086	64.076
TJUN	June temperature	71.492	71.617	71.598
TJUL	July temperature	75.0287	75.118	75.113
TAUG	August temperature	73.974	73.987	74.005
TJUNSQ	TJUN squared	7.8688	7.915	7.839
TAUGSQ	TAUG squared	6.4546	6.304	6.319
TPJUN	June temp. * precip.	-0.5521	-0.663	-0.835
TPAUG	August temp. * precip.	0.0181	0.020	0.351
MJT	May * June temp.	4.6318	4.529	4.412
JJT	June * July temp.	4.7689	4.938	4.889
JAT	July * Aug. temp.	4.9289	4.915	4.953
MJJT	May * June * July temp.	-3.5459	-3.230	-2.974
JJAT	June * July * Aug. temp	-4.7072	-4.850	-4.746
MP1	May weekly precip.	0.808	0.814	
JP1	July 1-7 weekly precip.	0.93	0.931	
JP2	July 8-15 weekly precip.	1.11	1.109	
MP1SQ	MP1 squared	0.3979	0.381	
JP1SQ	JP1 squared	0.5571	0.537	
JP2SQ	JP2 squared	0.7091	0.683	
MT1	May weekly temp.	64.82	64.821	
JT1	July 1-7 weekly temp.	74.44	74.437	
JT2	July 8-15 weekly temp.	74.76	74.765	
MT1SQ	MT1 squared	12.6797	13.857	
JT1SQ	JT1 squared	12.3496	12.861	
JT2SQ	JT2 squared	10.8454	10.760	

n = 300 for monthly temperature and precipitation variables; n = 294 for weekly temperature variables; n = 293 for weekly precipitation variables

Table 3. Estimated coefficients for the U. of Virginia corn model

Variable	University of Virginia WX Data: U. of V.-SPSS	University of Virginia WX Data: YES-SAS	NOAA WX Data: YES-SAS
B1	10.662	10.734	10.210
B2	10.919	12.198	11.906
B3	12.421	13.270	12.784
B4	24.145	24.536	24.944
B5	22.802	24.170	24.345
B6	13.691	14.342	14.285
TR1	1.575	1.600	1.615
TR2	1.049	1.014	1.027
TR3	0.948	0.923	0.936
TR4	1.090	1.083	1.062
TR5	1.244	1.208	1.206
TR6	1.431	1.420	1.428
PPRE	0.562	0.515	0.508
PJUN	1.269	1.217	1.169
PAUG	1.035	1.090	1.093
PAUGSQ	-0.226	-0.249	-0.250
MJP	-0.548	-0.492	-0.524
TPJUN	-0.258	-0.248	-0.207
JP1	5.551	5.461	5.358
JP1SQ	-1.860	-1.883	-1.762
JP2	2.341	2.215	2.245
MT1	-0.489	-0.463	-0.435
MT1SQ	0.045	0.048	0.051
JT2SQ	-0.144	-0.142	-0.137
Residual			
Degrees of freedom	243	243	243
Sum of squares	11481.34	11818.35	11703.47
Mean square	47.25	48.64	48.16

Table 4. Variables and coefficients selected by SAS stepwise procedure using NOAA weather data and no intercept model

Variable	Coefficient
B1	14.009
B2	16.353
B3	15.485
B4	25.583
B5	23.793
B6	12.990
TR1	1.559
TR2	0.932
TR3	0.862
TR4	1.036
TR5	1.135
TR6	1.296
PPRE	0.663
PJUN	0.788
PAUG	1.214
PAUGSQ	-0.255
JAP	-0.265
TJUN	-0.535
TJUNSQ	-0.169
JJT	0.328
JP2	2.164
MT1	-0.498
JT1	-0.390
JT2	-0.310
JT2SQ	-0.188

Residual	
Degrees of freedom	243
Sum of squares	12591.996
Mean square	51.82

Table 5. Variables and coefficients selected by SAS stepwise procedure using NOAA weather data, an intercept model, and location and trend variables defined for both individual climatic divisions and sets of similar climatic divisions

Variable	Coefficient
Intercept	13.959
B45	11.252
TR1	1.556
TR23	0.960
TR45	1.070
TR6	1.266
PPRE	0.611
PJUN	0.927
PAUG	1.241
PAUGSQ	-0.233
JAP	-0.292
MJJP	-0.265
TJUN	-0.417
TJUNSQ	-0.170
JJT	0.374
MP1SQ	-0.750
JP2	2.171
MT1	-0.592
JT1	-0.405
JT2	-0.310
JT2SQ	-0.205

Residual	
Degrees of freedom	246
Sum of squares	12394.32
Mean square	50.38

Table 6. Variables and coefficients selected by SAS stepwise procedure using NOAA weather data, an intercept model, and location and trend variables defined for both individual climatic divisions and sets of similar climatic divisions. Monthly temperature and precipitation variables for May and July added to the list of weekly and monthly variables used by the University of Virginia.

Variable	Coefficient
Intercept	11.471
B45	10.772
TR1	1.640
TR23	1.037
TR45	1.210
TR6	1.426
FPRE	0.687
PMAY	0.695
PMAYSQ	-0.240
PJUN	1.281
PJUNSQ	-0.268
PJUL	3.283
PJULSQ	-0.581
MJJP	-0.273
TMAY	-0.470
TMAYSQ	0.152
MJT	-0.397
JJT	0.363
JAT	-0.179
TPJUN	-0.216
MT1	-0.395
JT2SQ	-0.109

Residual	
Degrees of freedom	247
Sum of squares	9185.87
Mean square	37.19

Table 7. Variables and coefficients selected by SAS stepwise procedure using NOAA weather data, an intercept model, and location and trend variables defined for both individual climatic divisions and sets of similar climatic divisions. Monthly temperature and precipitation variables for May and July added to the list of weekly and monthly variables used by the University of Virginia. Temperature by precipitation interaction terms and temperature or precipitation sequence terms of two or more months omitted from the list.

Variable	Coefficient
Intercept	11.900
B45	10.603
TR1	1.613
TR23	1.014
TR45	1.204
TR6	1.394
PPRE	0.655
PMAY	0.854
PMAYSQ	-0.219
PJUN	1.140
PJUNSQ	-0.189
PJUL	2.945
PJULSQ	-0.490
PAUG	0.498
PAUGSQ	-0.165
TJUN	-0.363
MT1	-0.437
JT2	-0.288
JT2SQ	-0.121

Residual	
Degrees of freedom	250
Sum of squares	10213.48
Mean square	40.85

. or very bad yields which were not predicted accurately

	PMAY		PJUN		TJUN		JT2		PJUL		PAUG	
nt.	Z	Cont.	Z	Cont.	Z	Cont.	Z	Cont.	Z	Cont.	Z	Cont.
1.0	-0.4	0.0	0.6	1.3	0.6	-0.6	-0.2	1.4	2.8	4.4	-0.5	0.0
-1.2	0.5	1.1	-0.5	-0.6	-0.8	0.8	-1.2	0.6	-0.1	0.8	-0.4	0.1
0.2	-0.2	0.3	0.5	1.2	1.5	-1.5	0.8	-0.2	-1.0	-4.8	0.8	1.1
-1.3	-1.7	-3.5	0.5	1.1	0.1	-0.1	1.0	-1.1	1.1	5.2	-0.9	-0.7
2.5	-0.2	0.2	-1.2	-2.2	-0.2	0.2	1.2	-1.6	0.1	1.7	1.5	0.6
1.0	-0.3	0.1	-0.8	-1.2	1.2	-1.2	0.7	0.0	-0.9	-4.1	0.3	1.0
-0.2	-1.3	-2.1	0.3	0.9	-0.1	0.1	-1.2	0.5	1.4	5.7	1.3	0.9
1.0	-0.4	-0.2	-0.1	0.3	1.0	-1.0	0.2	1.0	-0.1	0.9	-0.7	-0.3
-0.1	-1.3	-2.1	0.1	0.5	-0.5	0.5	-0.7	1.3	-0.3	0.1	-0.2	0.6
-1.1	-1.0	-1.4	-0.1	0.3	1.8	-1.8	1.7	-4.2	-1.5	-8.9	0.8	1.1
1.5	-1.5	-2.6	1.4	2.0	-0.5	0.5	-2.0	-1.9	-1.8	-11.7	-0.3	0.3
-0.6	-1.0	-1.3	-0.2	0.1	0.5	-0.5	0.5	0.5	-1.8	-12.4	-1.3	-2.0
-1.6	-1.5	-2.8	-0.8	-1.2	-1.4	1.4	1.3	-2.2	-1.5	-9.3	-0.4	0.3
0.3	0.0	0.6	-1.1	-2.0	0.3	-0.3	2.2	-7.1	-0.8	-3.2	-0.4	0.1
-0.5	-0.3	0.0	-1.6	-3.4	-1.0	1.0	0.6	0.1	-1.3	-7.2	-0.5	0.1
-0.5	0.2	0.8	-0.1	0.2	0.0	0.0	0.8	-0.3	-0.1	0.6	0.1	0.8
1.9	0.0	0.5	-0.3	-0.1	0.1	-0.1	1.7	-4.2	-0.6	-2.3	-1.0	-1.1
-1.0	0.0	0.6	0.0	0.5	-0.4	0.4	1.1	-1.2	-0.1	0.8	0.0	0.8
-1.4	-0.6	-0.4	-1.1	-1.9	0.5	-0.5	-0.2	1.4	0.8	4.6	0.3	1.0
1.8	1.2	1.4	-0.5	-0.6	1.0	-1.0	0.3	0.9	1.2	5.4	0.8	1.1
2.0	1.3	1.4	0.3	0.9	0.7	-0.7	0.1	1.2	-0.2	0.5	0.2	0.9
-1.4	-2.1	-4.8	0.0	0.5	0.6	-0.6	2.8	-11.4	-0.6	-2.1	0.0	0.7