Yield Model Development

BRAZIL WHEAT YIELD COVARIANCE MODEL

(B84-10103) BRAZIL WHEAT YIELD COVARIANCE MODEL (National Oceanic and Atmospheric Administration) 22 p HC A02/MF A01 CSCL 02C

SUSAN L. CALLIS
CLARENCE SAKAMOTO

USDA/NOAA
CIAD/MODELS BRANCH
ROOM 200, FEDERAL BLDG
608 CHERRY STREET
COLUMBIA, MO 65201

Lyndon B. Johnson Space Center
Houston, Texas 77058
A model based on multiple regression was developed to estimate wheat yields for the wheat-growing states of Rio Grande do Sul, Parana, and Santa Catarina in Brazil. The meteorological data of these three states were "pooled" and the years 1972 to 1979 were used to develop the model since there was no technological trend in the yields during these years. Predictor variables were derived from monthly total precipitation, average monthly mean temperature, and average monthly maximum temperature.
BRAZIL WHEAT YIELD COVARIANCE MODEL

by

Susan L. Callis

and

Clarence Sakamoto

AISC Models Branch

January 16, 1984
INTRODUCTION

The purpose of this study was to select monthly weather variables that could be used to estimate wheat yields for the wheat growing areas of Brazil. Wheat is grown in seven states in Brazil: Rio Grande do Sul, Santa Catarina, Parana, Sao Paulo, Mato Grosso, Goias, and Minas Gerais. Rio Grande do Sul is Brazil's original wheat-producing state; until 1972 it was the country's most important production area. Increasingly, however, Parana, Mato Grosso and Sao Paulo have grown in importance. In 1977, Parana took over the number one spot. Figure 1 shows the wheat-growing areas of Brazil.

Although wheat has been grown since the sixteenth century, Brazil has yet to develop a high-quality wheat variety that produces well under the country's widely varied climatic conditions. Brazil's wheat crops have continuously been plagued with problems resulting in consistently low yields. Frosts and plant disease occasionally reduce yields. The high cost discourages use of fungicides. Expansion of acreage sown to wheat was met with cultivating problems and high costs. New land areas cultivated in wheat are highly acidic and low in fertility. Fertilizer is costly. Late-season rains frequently delay harvest and reduce yield.

The climate of the southern wheat-producing states is "subtropical humid"; rainfall is relatively abundant and well-distributed throughout the year, with slightly more rainfall in the warm months. There is usually no season of drought. The northern states are "semiarid" with a winter dry season and less total annual rainfall. Summers are hot and winters are mild. Parana and southern Sao Paulo are the northern limit for frost occurrence.

Wheat is planted in the months of April through June and is harvested in November and December.
Figure 1. Wheat-growing areas of Brazil. (J. McQuigg, R. Willis, 1982, personal communication)
METHOD

Multiple regression analysis of yield with selected agroclimatic indices was used to derive a suitable model. The index P-PET (precipitation minus potential evapotranspiration) was used in the regression equations to represent available soil moisture, monthly precipitation, and monthly maximum temperature.

The regression equation is:
\[ \hat{Y} = \alpha + B_1T_X_i + B_2R_i + B_3(P-PET)_i + E \]

where
- \( \hat{Y} \) = Estimated yield,
- \( \alpha \) = Constant,
- \( B_j \) = Coefficients of variables \( j = 1-3 \),
- \( T_X_i \) = Maximum temperature for month \( i \),
- \( R_i \) = Total precipitation for month \( i \),
- \( (P-PET)_i \) = Precipitation minus PET for month \( i \), and
- \( E \) = Unexplained error.

In developing the model, various procedures of the Statistical Analysis System (SAS Institute, Inc., 1979) were used. The procedures used and the operations performed with each are summarized in the Appendix. The selected model had the highest \( R^2 \) and included variables that were significant at the 10 per cent level and agronomically meaningful.

DATA

The Brazil crop data were obtained from the Foreign Agricultural Service (Sam Ruff, Personal Communication, 1982). The data was recorded with year of yield as year of harvest, so the weather influencing the crop occurred during year - 1.
Meteorological data from 1972 through 1977 were used to model because there is no apparent trend in the yield data during this period. Furthermore, 1977 represented the latest available data. Table 1 lists the stations used to derive the meteorological data sets for each state. Figure 2 shows the location of each station.

PROCEDURES

Since Rio Grande do Sul's meteorological data has the longest period of record, initial models were developed for this state alone. Various weather variables were tried in regression equations in many different combinations. The results were not good. Several different trend were tried with different regression models, but none were significant at the 10 per cent level. From historical accounts, it is believed that some sort of trend of increased yield began in 1962 because of increased use of fertilizer and more adaptable varieties of wheat. Yet the yield data did not indicate this. It was decided to model for years 1972-1979 for which it was believed no trend existed. Eliminating years of data created the problem of fewer degrees of freedom. However, by combining data for the states of Rio Grande do Sul, Parana, and Santa Catarina, a covariance model could be developed.

Variables used in the regression equations for the covariance model, included "dummy variables" for Parana and Santa Catarina. The "dummy variables" adjust the contributions to yield of both states to a base yield which, in this case, is Rio Grande do Sul's yield. The "dummy variable" for Parana was not significant at the 10 per cent level in the final model. The coefficient for Santa Catarina's "dummy variable" is negative, indicating that its yield is below that of the norm set by Rio Grande do Sul.
<table>
<thead>
<tr>
<th>STATE</th>
<th>METEOROLOGICAL STATION</th>
<th>WMO NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mato Grosso</td>
<td>Campo Grande</td>
<td>83611</td>
</tr>
<tr>
<td></td>
<td>Ponta Pora</td>
<td>83702</td>
</tr>
<tr>
<td></td>
<td>Pres Prudente</td>
<td>83716</td>
</tr>
<tr>
<td>Parana</td>
<td>Londrina</td>
<td>83766</td>
</tr>
<tr>
<td></td>
<td>Santa Branca</td>
<td>83810</td>
</tr>
<tr>
<td></td>
<td>Curitiba</td>
<td>83842</td>
</tr>
<tr>
<td></td>
<td>Porto Uniao</td>
<td>83864</td>
</tr>
<tr>
<td>Rio Grande do Sul</td>
<td>Ijuí</td>
<td>83805</td>
</tr>
<tr>
<td></td>
<td>Sao Borja</td>
<td>83901</td>
</tr>
<tr>
<td></td>
<td>Passo Fundo</td>
<td>83914</td>
</tr>
<tr>
<td></td>
<td>Julio de Castilhas</td>
<td>83935</td>
</tr>
<tr>
<td></td>
<td>Santa Maria</td>
<td>83936</td>
</tr>
<tr>
<td></td>
<td>Sao Gabriel</td>
<td>83957</td>
</tr>
<tr>
<td></td>
<td>Bage</td>
<td>83980</td>
</tr>
<tr>
<td>Santa Catarina</td>
<td>Porto Uniao</td>
<td>83864</td>
</tr>
<tr>
<td></td>
<td>Vaceria</td>
<td>83918</td>
</tr>
<tr>
<td></td>
<td>Sao Joaquin</td>
<td>83920</td>
</tr>
<tr>
<td>Sao Paulo</td>
<td>Pres Prudente</td>
<td>83716</td>
</tr>
<tr>
<td></td>
<td>Bauru</td>
<td>83722</td>
</tr>
<tr>
<td></td>
<td>Jau</td>
<td>83723</td>
</tr>
<tr>
<td></td>
<td>Tiete</td>
<td>83777</td>
</tr>
<tr>
<td></td>
<td>Ataliba Leonel</td>
<td>83803</td>
</tr>
<tr>
<td></td>
<td>Tatui</td>
<td>83816</td>
</tr>
</tbody>
</table>

Table 1. Meteorological Stations Used to Derive Data Sets for the Brazil Wheat Model.
Figure 2. Location of Meteorological Stations Used to Derive Data Sets for the Brazil Wheat Model.
The following is the selected model:

- **DUM**: Dummy variable to adjust Santa Catarina's yield
- **TM7**: July Mean Temperature
- **P-PET8**: August P-PET
- **SP-PET8**: Squared August P-PET
- **P-PET9**: September P-PET
- **RDFN11**: Deviation from normal of November precipitation

Too high temperatures in July and too much rainfall in November both reduce yield. The negative coefficient for the linear term P-PET8 suggests that in August excess precipitation above demand, PET8, is damaging to yield. This is reasonable in Brazil when in August the crop is in the tillering stage. The quadratic term indicates that increased yield is favorable at some level when PET is higher than precipitation. However, one is cautioned not to extend this interpretation beyond the limits of the data base used. The statistics of the selected model are summarized in Table 2.

The same variables were used in regression equations for only the two states of Rio Grande do Sul and Santa Catarina. The problem encountered was that models with agronomically reasonable variables had too low an R²; models with acceptable R² had too many variables for the number of degrees of freedom.

Finally, modeling was attempted for the northern states of Mato Grosso and Sao Paulo. Overlapping plots of weather variables versus yield were made for both states. From these plots, it was determined that modeling for a combination of data for these two states would not be acceptable; their climates are too different. Next, reasonable weather variables were tried in regression for both states separately. No suitable models were derived. For information, plots of yields for both states are shown in Figure 3.
TEST RESULTS

A jackknife test was run on the selected model. In this test, a year was eliminated from the crop data and the model was used to predict that year’s yield. This process was done for each successive year beginning with 1972. The test had to be run separately on each state. The results are printed on Tables 3 through 5 and plotted on Figures 4 through 6.
APPENDIX

Definition of Variables

P-PET, precipitation minus potential evapotranspiration, is used a measure of the amount of moisture available for plant growth. Potential evapotranspiration is determined by the procedure developed by Thornthwaite (1948). It requires only temperature:

\[ \text{PET} = \left( \frac{10T}{I} \right)^{a} \]

where \( I \) = heat index, which is the sum of the 12 monthly indices \( i \),

\[ i = \left( \frac{I}{5} \right)^{1.517} \]

\( T \) = monthly temperature in °C, and

\( a \) = an empirical exponent \( 6.75 \times 10^{-7}I^{3} - 7.71 \times 10^{-5}I^{2} + 1.79 \times 10^{-2}I + 0.49 \).

The duration of daylight is used to adjust potential evapotranspiration as a portion of 12 hours.

Statistical Analysis System Procedures Used

PROC CORR Computes correlation coefficients between variables, including Pearson product-moment and weighted product-moment correlation.

PROC PLOT Graphs one variable against another, producing a printer plot.

PROC STEPWISE Provides five methods for stepwise regression. Stepwise is useful when selecting variables to be included in a regression model from a collection of independent variables.

PROC STEPWISE FORWARD Begins by finding the one-variable model that produces the highest \( R^{2} \). For each of the other independent variables, FORWARD calculates F-statistics reflecting the contribution to the model if the variable were to be included.
PROC STEPWISE BACKWARD

Begin by calculating statistics for a model including all the independent variables. The variables are deleted from the model one by one until all the remaining variables produce F-statistics significant at the .10 level.

PROC STEPWISE STEPWISE

The stepwise method is a modification of the forward selection technique, differing in that variables already in the model do not necessarily stay there. After a variable is added (as in the forward selection method) the stepwise method looks at all the variables already included in the model and deletes any variable that does not produce an F-statistic significant at the .10 level. Only after this check is made and the necessary deletions accomplished can another variable be added to the model.

PROC STEPWISE MAXR

(Maximum R^2 improvement) Unlike the three techniques above, this method does not settle on a single method. Instead it looks for the "best" two-variable model, the "best" three variable model, and so forth.

PROC PETM

Uses latitude and mean monthly temperature to calculate Thornthwaite's potential evapotranspiration for each month.

PROC ZINDEX

Uses monthly PET's, precipitation, SS (beginning moisture in surface layer), AMCS (available water capacity in surface layer), SU (beginning moisture in the underlying layer), and AMCZ (available water capacity in the underlying layer) to calculate Palmer's soil moisture budget, drought index Z, ET, and ETi.
## Table 2. Statistics of Brazil Wheat Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>B Value (Unstandardized)</th>
<th>Std Error</th>
<th>Type II SS</th>
<th>F</th>
<th>Prob&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>27.91e71003</td>
<td>1.4541e184</td>
<td>43.54320431</td>
<td>11.39</td>
<td>0.0036</td>
</tr>
<tr>
<td>DUM2 (Santa Catarina)</td>
<td>-4.70e22434</td>
<td>0.4313e438</td>
<td>22.28065116</td>
<td>5.89</td>
<td>0.0273</td>
</tr>
<tr>
<td>T+7</td>
<td>-1.04e13789</td>
<td>0.0195e438</td>
<td>21.36220762</td>
<td>5.59</td>
<td>0.0302</td>
</tr>
<tr>
<td>P PH3</td>
<td>0.00e016214</td>
<td>0.0000e426</td>
<td>17.27858241</td>
<td>4.72</td>
<td>0.0484</td>
</tr>
<tr>
<td>S CECA</td>
<td>0.02e43603</td>
<td>0.0094e298</td>
<td>29.33485172</td>
<td>7.68</td>
<td>0.0131</td>
</tr>
<tr>
<td>D FEITO</td>
<td>-0.04206852</td>
<td>0.0098e494</td>
<td>70.35548490</td>
<td>18.41</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

Variables in the model are significant at the 0.1000 level.
<table>
<thead>
<tr>
<th>ETA3</th>
<th>ETA4</th>
<th>ETA5</th>
<th>ETA6</th>
<th>ETA7</th>
<th>CONTRI1</th>
<th>CONTRI2</th>
<th>CONTRI3</th>
<th>CONTRI4</th>
<th>CONTRI5</th>
<th>CONTRI6</th>
<th>CONTRI7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.255569</td>
<td>0.2746</td>
<td>0.2997</td>
<td>0.3451</td>
<td>0.3595</td>
<td>-0.0337</td>
<td>-0.0394</td>
<td>0.4309</td>
<td>0.5904</td>
<td>0.9104</td>
<td>-7.5438</td>
<td>1.6201</td>
</tr>
<tr>
<td>0.5904</td>
<td>0.6188</td>
<td>0.6451</td>
<td>0.6717</td>
<td>0.7004</td>
<td>-0.0371</td>
<td>-0.0427</td>
<td>0.4309</td>
<td>0.5904</td>
<td>0.9104</td>
<td>-7.5438</td>
<td>1.6201</td>
</tr>
<tr>
<td>0.3924</td>
<td>0.4179</td>
<td>0.4434</td>
<td>0.4693</td>
<td>0.4953</td>
<td>-0.0352</td>
<td>-0.0408</td>
<td>0.4309</td>
<td>0.5904</td>
<td>0.9104</td>
<td>-7.5438</td>
<td>1.6201</td>
</tr>
</tbody>
</table>

Table 3. Results of Jackknife Test for State of Parana Wheat Model
Figure 4. State of Parana Wheat Model.
Table 4. Results of Jackknife Test for State of Rio Grande do Sul Wheat Model.
Brazil Wheat Jackknife Results—Test of Model 1975-1980

_0_ = Observed Yield
_p_ = Models Predicted Yield

Plot of Yield vs Yieldyr
Symbol used is _p_
Plot of Observed Yield vs Yieldyr
Symbol used is _0_

Figure 5. State of Rio Grande do Sul Wheat Model
Figure 6. State of Santa Catarina Wheat Model