CROP YIELD PREDICTION USING SATELLITE INDICES[1]

William T. Liu[2] and Felix Kogan[3]

ABSTRACT

This study is to explore the potential application of NOAA/AVHRR based satellite indices to soybeans yield estimate for Brazil. Two soybeans yield models, including Paraná State and country were constructed using observed yield data and NOAA/ AVHRR GVI visible, near infrared and thermal channels weekly data for the period of 1985 to 1995 and were evaluated using data period of 1996 to 1998. The results of model validation showed that the percentage errors of 1996, 1997 and 1998 were –8.74, 0.12 and 2.81 respectively for the Paraná model and 4.05, -3.87 and 5.51 respectively for the country model. It is concluded that the satellite indices data are useful for crop production monitoring.

INTRODUCTION

Remote sensing data provide much more information about land surface features, including agricultural crops behavior. It has been shown that AVHRR data obtained from NOAA polar orbiting satellite platforms can be used as a sole and also complimentary to weather data source of information for the purposes of monitoring crop conditions and productivity on a large scale area (Tucker et al 1986; Quarmby, et. al., Rasmunsen; 1997, Unganai et al 1998). This research, in addition to focus on vitally important agricultural region, has a goal of testing the new AVHRR-based method in areas with sufficient moisture supply. The objectives were to investigate the nature of the relationship between satellite-derived and agricultural indicators and to simulate productivity of soybeans in Brazil using new AVHRR-based indices as predictors.

---

[1] Este pesquisa é financiadopela FAPESP, Proc. N° : 1997/11568-3.
[2] Department of Atmospheric Sciences, IAG/USP, Rua do Matao 1226, IAG/USP, Cidade Universitaria, Butantan, 05508-960-SP, Sao Paulo, Brazil. Emial: will@model.iag.usp.br
[3] Physical Scientist, Climatic Research and Application Division, NESDIS/NOAA, NOAA Science Center 5200, Auth Rd, Camp Springs, MD 27026, USA. Email: fkogan@orbit1s.nesdis.noaa.gov

DATA

Two data sets were used in this study, soybeans yield and NOAA/AVHRR, during the period of 1985 through 1998. Two areas, including the Paraná state and Brazil were selected for constructing soybeans prediction models. Official yield statistics of soybeans gathered by government agencies in Brazil, IBGE, were used. NOAA AVHRR GVI data for the period of 1985 to 1998 provided by NOAA/NESDIS were used. The equations 1 and 2 describe the calculation of satellite indices: Vegetation Condition Indices (VCI) and Temperature Condition Indices (TCI), used in this study.

$$VCI = (NDVI - NDVI_{min}) / (NDVI_{max} - NDVI_{min})*100 \quad (1)$$

$$TCI = (BT_{max} - BT) / (BT_{max} - BT_{min}) * 100 \quad (2)$$

Where NDVI, $NDVI_{max}$, and $NDVI_{min}$ are the smoothed weekly NDVI, its multi-year absolute maximum, and minimum, respectively; BT, BT_{max}, and BT_{min} are similar values for brightness temperature.

RESULTS AND DISCUSSION

The crop yield is generally affected by the technological improvements through time and by the annual weather fluctuation. The contribution of technological improvements was approximated by trend term and weather-related fluctuations of yield around the trend were estimated through AVHRR-based indices. Annual yield increase rate due to technological improvements of each region was obtained by applying the linear regression technique using historical yield data against year. Table 1 estimates technological trend during 1986 to 1995. The results showed that the technological trend had significant contribution with slope of 0.5364 with R^2 of 0.3364 for the Paraná State and slope of 0.6593 with R^2 of 0.5819 for country.

The dynamics of soybeans yield and TCI/VCI for the week of the highest correction is shown in Figure 1. There is clearly identified critical period when yield is sensitive to TCI variations. Highest correlation coefficient between yield variation from trend and TCI occurred at week 28 in Paraná model and at week 31 in country model.

Yield Models

The model construction consists of two steps: modeling of yield trend and deviation from trend. For the deviation, the yield ratio data were used as dependent variable and TCI/VCI during one to 4 week periods as dependent variables. In order to select better candidate variables, correlation coefficient lower than 0.3 were excluded. One to two candidate models were then selected through the "STEPWISE" regression process available in the software SAS ("Statistical Analysis System"). Each model contains only one or two variables using data period from 1986 to 1994. Data from 1996 to 1998 were used for model validation. The 1995 satellite data were not available.

Table 2 summarizes the yield departure from trend (ratio) models. For the Paraná model, the determination coefficient (R^2) was 0.5876 and the root mean square error (RMSE) was 0.952%. For the country model, the R^2 was 0.3544 and RMSE was 9.4%. It was observed that TCI /VCI from January (week 27) to mid February (week32) were good predictors of yield deviation from trend during flowing and grain filling stages of soybeans..

Model Validation

Models in Tables 1 and 2 were used to estimate actual soybeans yield during 1986 to 1998 (except 1995). Figures 1 and 2 showed simulated and observed yield for the Paraná and Brazil respectively. Within the period used for model construction (1986 to 1994), The models had RMSE lower than 10%. Both models had RMSE lower than 10% that was considered acceptable. The simulated corresponded quite well with low yield in 1991 for PR. But in year 1994, the effect of low yield was not detected by the model due to excess water at stages of maturity and harvest.

Models were then evaluated independently by comparing simulated to observed yields during 1996 to 1998. These yield data were not included in the model construction. Table 6 and Figures 8 and 9 show the results of the comparison. The results of model validation showed that the percentage errors of 1996, 1997 and 1998 were –8.74, 0.12 and 2.81 respectively for the Paraná model and 4.05, -3.87 and 5.51 respectively for the country model. The result of both dependent and independent model evaluations showed that the percentage errors of both models were lower than 10%.

CONCLUSION

It is concluded that the AVHRR-based indices explored in this research showed to be useful for crop production monitoring. In the major areas of Brazilian soybeans cultivation, the temperature-based index is more informative about possible fluctuation of soybeans yield and production in Brazil. In some regions, a combination of satellite and in situ data may likely improve yield estimate. The future plans of this research will include application of the same techniques for simulation of other crops more sensitive to weather fluctuations.

REFERENCES

Kidwell, K. B., 1994, Global Vegetation Index User's Guide. US Dept. of Commerce, NOAA.

Kogan F.N. (1997), Global dought watch from space. *Bulletin of American Meteorological Society*, 78:621-636.

Rao, C.R.N., J. Chen, 1995: Inter-satellite calibration linkages for the visible and near-infrared channels of the Advanced Very High Resolution Radiometer on the NOAA-7, -9,and -11 spacecrafts. *International Journal of Remote Sensing.* 16, 1931-1942.

Quarmby, N. A., M. Milnes, T. L. Hindle and N. Silleos, 1993, The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction. *International Journal of Remote Sensing*, **14**, 199-210.

Rasmussen M.S. (1997). Operational yield forecasting using AVHRR NDVI data: prediction of environmental and inter-annual variability. *International Journal of Remote Sensing*. 18:1059-1077.

Unganai L.S. and Kogan, F.N., 1998, Drought Monitoring and corn yield estimation in Southern Africa from AVHRR data, *Remote Sensing of Environment*, **63**, 219-232.

Weinreb, M. P., G. Hamilton and S. Brown, 1990, Nonlinearity correction in calibration of the Advanced Very High Resolution Radiometer infrared channels. *Journal of Geophysical Research*, **95**, 7381-88.

Table 1. Results of the soybeans yield trend for the Paraná state and Brazil studied. (yield data of 1985 to 1995 reported by IBGE)

Region ID	Region Name	Regression Equation Intercepted	Slope	R^2	rmse (%)	range (%)
4	Paraná	-27.252	0.5364	0.3364	14.90	-12.5 ~ 26.7
9	Brazil	-40.858	0.6593	0.5819	10.70	-11.8 ~ 53.9

Table 2. Yield estimation models, R^2 and root mean square error (RMSE) for the Paraná state and Brazil.

Region N$^{o.}$	State/Brazil	Alternative Model*	R^2	RMSE
4	Paraná	$69.1393+0.9628T4-0.00656(T4)^2$	0.5876	9.52
9	Brazil	$76.1875+0.284V3+0.00096(T3)^2$	0.3544	9.40

- obs: T represents TCI; V represents VCI; the subscript number indicates the week number; the normal number indicates the period summed through the number of weeks which varies from region to region.
- For Paraná state: $T4= (T_{26}+ T_{27} + T_{28} + T_{29})/4$;
- For Brazil: $T3= (T_{31}+ T_{32} + T_{33})/3$; $V3= (V_{31}+ V_{32} + V_{33})/3$;

Table 3. Comparison of observed and model estimated soybeans for the Paraná state and Brazil.

State/Brazil	Year	Yield (kg/ha)		% error
		Observed	Estimated	
PR	1996	2677	2443	-8.74
	1997	2580	2583	0.12
	1998	2558	2630	2.81
BRAZIL	1996	2195	2284	4.05
	1997	2298	2209	- 3.87
	1998	2360	2490	5.51

Fig. 1 – Comparison of observed and model estimated soybeans yield for the Paraná State.