Predicting crop yields 50 | Where is an exchange? 68

www.futuresmag.com September 2006

Futures®

News, analysis and strategies for futures, options and stock traders

THE INCREDIBLE *SHRINKING* DOLLAR?

page 24

US $6.95 CAN $8.95

HIGHLINEMEDIA

New applications of satellite data can better predict growing conditions worldwide. This can be used to forecast crop production that leads the widely followed government reports.

Microwave imaging that predicts yields

BY ALAN BASIST, ROBIN HULT, SAMUEL SHEN, NEIL THOMAS & MARC BASIST

Imagine if you could forecast crop conditions better and faster than the U.S. Department of Agriculture (USDA). Well, it's possible using technology known as special sensing microwave imaging (SSMI).

This technology can objectively calculate changes in growing conditions and yields for major crops at the county-equivalent level throughout the world. This permits independent and objective assessment of yield where limited data previously existed.

Crop models that exploit this data use the statistical relationships between temperature and wetness variations and yield figures at the county level. Running on near real-time SSMI data, the output is highly correlated with yield values supplied by the National Agricultural Statistical Service (NASS), which are followed world-wide as the definitive source of crop data. Moreover, the SSMI derived yield index provides an excellent technique to objectively assess yields without extensive, expensive and subjective field surveys.

The benefit for the trader is clear: faster, accurate, more affordable crop assessments result in better models. These models result in satellite derived accurate forecasts, and ostensibly, more profitable trades.

THE TECHNOLOGY

This technique uses the microwave spectrum to identify changes in surface wetness and temperature. It then incorporates these changes, measured as anomalies, into crop models, which explain variations in yields for soybean, corn, wheat and cotton in the United States. Alternative methods, such as traditional field surveys, are based on few and frequently unrepresentative spot observations and these findings tend to be subjective in nature.

SSMI technology was initially developed to monitor surface temperature and wetness from microwave energy naturally emitted from the land surface. The SSMI can observe, monitor and measure the land surface under almost all sky conditions. Thus, SSMI provides better risk coverage than optical-based satellite methods because clouds can cover much of the earth's surface at any time.

The temperature measurement tool was calibrated on an extensive network of surface stations. The wetness measurement tool is a composite of any source of moisture near the surface. These developed models have been combined and integrated as two inputs to create yield indexes for corn, soybeans, wheat and cotton.

The data come from a satellite platform flown by the Defense Meteorological Satellite Program (DMSP) that orbits the globe 14 times a day, and has been doing so since 1987. The DMSP satellites have sun-synchronized overpasses at 6 a.m. and 6 p.m. These satellite overpasses occur twice daily and are processed into 1/3 x 1/3 degree "pixels" by the National Environmental Satellite and Data Information and Satellite (NESDIS). These data are archived at NOAA's Satellite Active Archive (SAA) in near real time.

The data received from these satellite observations are processed into three classes of values: the actual, climatology and anomaly. Both the temperature, measured in Celsius, and wetness measurements are available as morning and afternoon observations.

Anomalies are departures from the expected value for that location and time of year. The surface wetness index is derived as the percentage of the radiating surface that is in any form of moisture (liquid water). Anomalies for the wetness product are defined by a cumulative probability function, where low values are extremely dry and high values are extremely wet for that location and time of year.

Using techniques that measure the true spatial structure of the temperature is elusive in most areas of the world because isolated point measurements are smeared across the region, hiding the true spatial structure and gradients. This is particularly true in mountainous areas or regions where steep and irregular gradients in temperature and precipitation occur. The satellite sensor's ability to monitor the true surface wetness and temperature patterns, and departures from normal in near-real time provides a great utility to an array of applications.

The satellite observations are averaged throughout a base period from 1988 to 2005 for each month at every 30-kilometer pixel across the land surface. The mean values are compared against the observed temperatures for a particular time and location. The departure from the mean defines the temperature anomaly, which identify whether a location has above, below or average temperatures during that time of year. Anomalies range from severely dry to severely wet using a cumulative probability scale.

"World assessment" (right) shows the full global structure of land surface temperatures for a week in July 2005 (top chart). It identifies the areas where temperatures are above average (in much of the United States, Brazil and eastern Russia) as well as areas where temperatures are below normal (such as Canada, Argentina and parts of China). The bottom chart shows the corresponding surface wetness anomalies for the same week. It shows much of China slightly wetter than normal, while the largest positive wetness anomalies are in northeastern Australia and across much of India.

By transferring the magnitudes of these anomalies to our crop model, we can explain variations in yield by measuring the combined impact of surface wetness and temperature anomalies during planting, emergence, filling, maturation and harvest.

METHODOLOGY

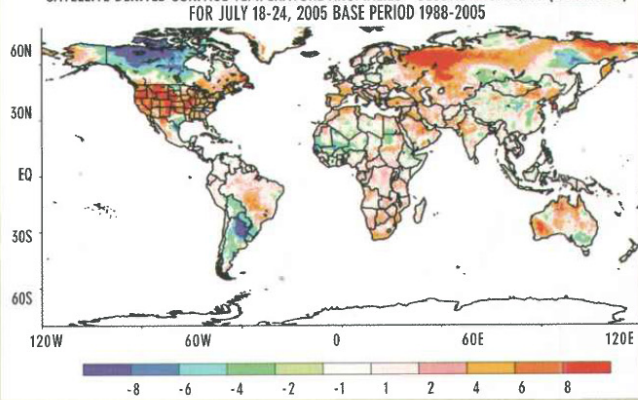
Numerous approaches have tried to model how changes in growing conditions impact variability in yield. Some of the techniques integrate remotely sensed data into the models. However, the inter-annual variations of soil moisture can be quite different from actual observations. Soil moisture is an important variable in assessing growing conditions, and if a certain technique cannot measure it accurately, then a new technique is needed.

Using only two sets of predictors,

WORLD ASSESSMENT

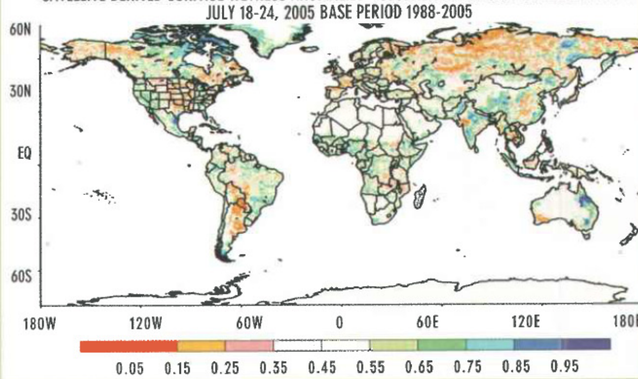
This shows how temperatures around the world vary from their norm. New technology allows this data collection to be unaffected by cloud cover.

SATELLITE DERIVED SURFACE TEMPERATURE ANOMALIES - GLOBAL ANOMALIES (DEGRESS C) FOR JULY 18-24, 2005 BASE PERIOD 1988-2005



We also can measure various moisture levels around the world. By inputting the temperature and moisture data into our model, we can improve on crop forecasts.

SATELLITE DERIVED SURFACE WETNESS ANOMALIES - GLOBAL STANDARDIZED ANOMALIES FOR JULY 18-24, 2005 BASE PERIOD 1988-2005



Source: Commodity Hedgers

INSIDE IOWA

As expected, partly because it takes time to perform field surveys and analyze the findings, the SSMI data generally lead the NASS reports. For instance, in 1992 the models consistently increased yields from the end of June, and the NASS yields approached the SSMI model results later.

	Basist Yield Index (Planted yields from normalized data) IOWA CORN				Differenc		NASS Planted Yields (from harvested production and planted acreage) IOWA CORN				
Data ending:	30-Jun	31-Jul	31-Aug	30-Sep	Sept 30 - --- Final	Report date:	Aug	Sep	Oct	Nov	Final
1988	91 3	84.9	84.3	84 6	6.7	1988	75.8	73.9	77.6	78.6	79.5
1989	116.0	116.1	116.1	113.5	2.7	1989	106.9	105.0	110.8	115 7	114 7
1992	132.5	145.9	147.3	147.2	14.8	1992	128.5	128.5	132.4	142.3	144.2
1995	122.6	121 3	118.2	119.6	1 7	1995	131 6	121 8	117 9	117.9	119.9
1996	128.8	135.9	136.0	134.5	2.4	1996	126.0	127.9	131.8	136.7	134.7
1997	125.3	125.8	129.2	132.2	-4 4	1997	136.6	136.6	136.6	136 6	134 6
1998	139.8	138.6	137.1	138.4	1.2	1998	139.6	139.6	139.6	141.5	141.5
1999	147.0	139.3	143.2	142.3	-4.0	1999	147 3	147.3	146.3	145.3	145.3
2000	154.0	154 0	148.0	146 3	1 9	2000	151.2	151.2	144.4	142 4	141.5
2001	147.8	147.0	144.7	144.6	7.2	2001	137.4	134.5	137.4	143.2	142 3
2002	158.9	153 1	154.2	154.7	3	2002	141.8	144.7	154.4	155 4	160 3
2003	151 1	151.0	142.8	143.1	-7.8	2003	152.9	149.0	150.9	153.8	151.9
2004	161 1	163 1	175.2	177.0	1.3	2004	158.2	159.1	175.7	178.7	176.7

* Archived satellite data was incomplete in 1990,1992,1993 and 1994

Source: Commodity Hedgers & USDA

monthly wetness and temperature anomalies, we can relate changes in the growing conditions to fluctuations in yield potential. This is done by correlating the SSMI monthly anomalies to yield and their explanation in yield changes as the growing season progresses. The final county-level statistics provided by the NASS can be used to judge the accuracy and stability of the models.

Monthly anomalies for the months influencing growing conditions of various crops are correlated to final yield values. Because the models are calibrated against final yield, they should correctly converge on the final yield as the growing season unfolds. We can test this hypothesis using the monthly NASS estimated yield values, which are known as track yields. The period of study begins with the August report (July survey) and ends with the final report, which is the January annual crop production.

The main inputs to the models are the anomalies during the main plant growth cycle: vegetative, reproduction, seed-pod filling and maturation. The model parameters and correlation coefficients are generated using nonlinear regression analysis. The independent, or exploratory, variables are monthly anomalies of wetness and temperature, and the statistical procedure corresponds a beta coefficient to each of these independent variables.

The data set used to generate the weights for the independent variables is huge. Data are drawn from about 100 counties and 10 years of validation data. This effectively creates 1,000 growing years for testing the accuracy and value of the model.

But that doesn't mean that what works well in one area works well in another. In one particular area, anomalous hot and wet surface conditions may be optimum, while in another area it may be better if conditions are cool and wet. So, developing a model on statewide statistics allows it to determine the best relationships to accommodate regional differences.

Clearly, factors other than temperature and wetness affect crop production. However, because we are using variations from the norm, then the effect of other important variables, such as soil types and climatic conditions, can be ignored because they typically don't change for a specific area. After all, the reason particular crops are grown at a location is because that location provides some stability growing conditions.

Obviously, unusually hot and d surface conditions will impact tl crop's development much different from unusually cold and dry weathe Therefore the models contain son nonlinear interactions between the two sets of variables. A primary reas why yields vary at a location relat to the interaction of moisture ar temperature at the site.

Because locations can vary in the productivity, the NASS yield valu for each county are normalize Specifically, we calculate the mea yield for each county and use the ave age to derive annual departures. F instance if the mean is 100 bushels f county A, and in a particular year received a yield of 50 bushels, tl value used in the model is 50%. Th permits the anomalies to efficient translate changes in growing cond tions with expected yield.

RESULTS & VALIDATION

To test the hypothesis that the SS data lead the NASS reports, predicte values can be compared to the yiel values in the NASS August repor based upon the agency's July survey.

second model correlates the SSMI data through July with NASS's analyses ending in August (reported in September). Sequentially, the models are correlated using SSMI anomalies ending in August with the NASS September survey (reported in October). Because NASS releases its end-of-month report in the middle of the following month, the SSMI predictions therefore precede NASS by more than five weeks.

The models must prove the validity of the assumption that changes in growing conditions can identify changes in potential yield, and more important, provide leading indications of NASS reported yields and demonstrate high correlation to final NASS yield estimates.

"Inside Iowa" (left) shows results from our model for Iowa, during the period 1988 through 2004. The SMMI data set is complete for the last 10 years, and the near-real-time SSMI data are operationally downloaded each day. All NASS anomalies are calculated in terms of a 17-year (1988 to 2004) base period. The models are calibrated on these NASS anomalies, which are normalized by counties. The normalization procedure describes an average year as 100%. A result of less than 100 is the amount yields are below average, and more than 100 is the percent it is above average.

One hypothesis in this study is that changes in yield are highly correlated to changes in field conditions during the growing season, and that the SSMI-based models converge on the final yield in a trend similar to NASS. Indeed, the SSMI data lead the NASS reports. The models actually explain variation in yield as early as the end of June that NASS does not report until mid-August.

During most years, the models converged on the final yield in a consistent pattern, indicating that growing conditions can be determined to be anomalies in temperature and wetness. The yield data were not trended, although we are well aware that all conditions being equal, we should expect a substantially larger yield in 2004 than 1988. This is one reason why SSMI predictions for 1988 are higher.

PREDICTING YIELDS

On average the accuracy of the SSMI was 1% better than NASS, and when 1988 (an atypical drought year) is excluded, it was almost 2% higher. That's with more than a month lead time.

Verification of Estimation Error
Absolute Error Relative to Final NASS Yield

	Basist end of Jun	NASS middle of Aug
1988	14.8%	4.8%
1989	1.1%	6.8%
1992	8.1%	10.9%
1995	2.2%	9.8%
1996	4.4%	6.5%
1997	6.9%	1.4%
1998	1.2%	1.4%
1999	1.2%	1.3%
2000	8.9%	6.9%
2001	3.9%	3.4%
2002	0.9%	11.5%
2003	0.5%	0.6%
2004	8.8%	10.5%
Average	4.8%	5.8%
w/o 1988	4.0%	5.9%

Source: Commodity Hedgers & USDA

We can test this hypothesis further by comparing the expected yields for Iowa corn early in the growing season. The earliest available NASS data is the July survey reported in August, which we can compare with the SSMI-based yields at the end of June. A 100 means a perfect first relationship (SSMI at the end of June and NASS in the middle of August) of the final yield value. See "Predicting yields," above. For the vast majority of the years, the June models identified final yield better than NASS July numbers.

The next question is whether the SSMI-based predictions lead the NASS change in yield as the growing season develops. To test this, we can use yield values from three crops in high-production states: corn, soybeans and wheat. The study used spring wheat from North Dakota and winter wheat from Kansas. It used soybeans from Illinois and Nebraska, and corn from Iowa and Ohio. Initial findings were only based on the relationship between SSMI-based final yield values (end of September) with the NASS final yield values from the USDA Annual Crop Production report released in January the following year. The correlation was more than 96% for each crop in the various states.

Next, to increase the amount of data used for this test, we can expand it to include four periods of the growing season for each crop and state. This better tests the ability of the SSMI data to lead NASS as the growing season advances. As mentioned above, the comparison is the end of June value for the SSMI data to end of July values for NASS, and then advanced the models in a monthly time step throughout the growing season (see "Tracking values," page 54).

The results of all these tests reveal three findings. First, NASS is generally conservative, moving yields slowly in the proper direction. Second, it takes time to perform field surveys and analyze the results. Third, the accurate, near real-time and objective data can provide valuable information and shorten the analysis period that relates field condition to final yield.

Because the SSMI data set is objective, global and scientific, it is being used by numerous governmental and commercial organizations that have a need to make accurate assessments of future agricultural production. It has rapidly become the scientific tool of choice in predicting and more thoroughly understanding global yields on a near real time basis. Now sophisticated traders may join these esteemed organizations by taking advantage of this superior technology. FM

TRACKING VALUES

Generated a month before the NASS reports, the SSMI data are surprisingly accurate in forecasting the government figures across several crops.

NASS Track Yields regressed on Basist Yield Index

	Annual Obs.	Annual Correlation	Monthly Obs.	Monthly Correlation
ND Wheat	13	98.8%	52	94.7%
KS Wheat	12	97.5%	48	90.0%
NE Soy Bean	13	96.1%	52	90.2%
IL Soy Bean	13	97.6%	52	92.1%
No. IL Soy bean*	13	98.8%	N.A.	N.A.
IA Corn	13	98.7%	52	95.0%
OH Corn	13	96.4%	52	94.2%
Average	12.9	97.7%	51.3	92.7%

* Monthly county estimates are not available from NASS

Source: Commodity Hedgers & USDA

Note: For additional references and suggestions for further reading on this subject, and for the online article "A better model for crop forecasts," please go to www.futuresmag.com. The online article provides more specific technical information on the model.

Alan Basist worked as a research meteorologist in the federal government for 15 years. He has an expertise in satellites and agricultural monitoring. Robin Hult has a history of software development and is a CPA. Sam Shen is the chairman of the mathematics department at San Diego State University with a research interest in climate data analysis. Neil Thomas is a GIS expert. Marc Basist is responsible for sales and marketing and can be reached at Marc@commodityhedgers.com.