THE INCREDIBLE SHRINKING DOLLAR?

page 24
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 soybeans, 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 90-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,
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.

<table>
<thead>
<tr>
<th>Basalt Yield Index (Planted yields from normalized data)</th>
<th>NASS Planted Yields (from harvested production and planted acreage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IOWA CORN</td>
<td>IOWA CORN</td>
</tr>
<tr>
<td>Report date:</td>
<td>Report date:</td>
</tr>
<tr>
<td></td>
<td>Aug</td>
</tr>
<tr>
<td>1990</td>
<td>75.6</td>
</tr>
<tr>
<td>1991</td>
<td>116.9</td>
</tr>
<tr>
<td>1992</td>
<td>135.6</td>
</tr>
<tr>
<td>1993</td>
<td>140.0</td>
</tr>
<tr>
<td>1994</td>
<td>151.1</td>
</tr>
<tr>
<td>1995</td>
<td>153.0</td>
</tr>
<tr>
<td>1996</td>
<td>156.0</td>
</tr>
<tr>
<td>1997</td>
<td>159.2</td>
</tr>
<tr>
<td>1998</td>
<td>161.1</td>
</tr>
<tr>
<td>1999</td>
<td>163.6</td>
</tr>
<tr>
<td>2000</td>
<td>166.0</td>
</tr>
<tr>
<td>2001</td>
<td>168.0</td>
</tr>
<tr>
<td>2002</td>
<td>170.1</td>
</tr>
<tr>
<td>2003</td>
<td>172.6</td>
</tr>
<tr>
<td>2004</td>
<td>175.0</td>
</tr>
</tbody>
</table>

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

Source: Central Valley Project

...continued

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 dry surface conditions will impact the crop’s development much different from unusually cold and dry weather. Therefore the models contain nonlinear interactions between the two sets of variables. A primary reason why yields vary at a location relate to the interaction of moisture at temperature at the site.

Because locations can vary in the productivity, the NASS yield value for each county are normalized. Specifically, we calculate the median yield for each county and use the average of the monthly departures. For example if the mean is 100 bushels of soybeans in county A, and in a particular year received a yield of 50 bushels, the value used in the model is 50. That permits the anomalies to efficient translate changes in growing conditions with expected yield.

RESULTS & VALIDATION
To test the hypothesis that the data can lead the NASS reports, predict values can be compared to the yields values in the NASS August report based upon the agency’s July survey.

52 FUTURES | September 2006
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 demon-
strate high correlation to final NASS yield estimates.

"Inside Iowa" (left) shows results from our model for Iowa, during the
period 1988 through 2004. The SSMI 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 anom-
lies in temperature and wetness. The yield data were not trended, although
we are well aware that all conditions being equal, we should expect a sub-
stantially larger yield in 2004 than 1988. This is one reason why SSMI
predictions for 1988 are higher.

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 sea-
non (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,
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.

TRACING VALUES

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

<table>
<thead>
<tr>
<th>NASS Track Yields regressed on Basist Yield Index</th>
<th>Annual Obs.</th>
<th>Annual Correlation</th>
<th>Monthly Obs.</th>
<th>Monthly Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND Wheat</td>
<td>13</td>
<td>96.8%</td>
<td>52</td>
<td>94.7%</td>
</tr>
<tr>
<td>KS Wheat</td>
<td>12</td>
<td>97.5%</td>
<td>48</td>
<td>90.0%</td>
</tr>
<tr>
<td>NE Soy Bean</td>
<td>13</td>
<td>96.1%</td>
<td>52</td>
<td>90.2%</td>
</tr>
<tr>
<td>IL Soy Bean</td>
<td>13</td>
<td>97.6%</td>
<td>52</td>
<td>92.1%</td>
</tr>
<tr>
<td>No. IL Soy bean&quot;</td>
<td>13</td>
<td>96.8%</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>IA Corn</td>
<td>13</td>
<td>98.7%</td>
<td>52</td>
<td>95.0%</td>
</tr>
<tr>
<td>OH Corn</td>
<td>13</td>
<td>96.4%</td>
<td>52</td>
<td>94.2%</td>
</tr>
<tr>
<td>Average</td>
<td>12.9</td>
<td>97.7%</td>
<td>61.3</td>
<td>92.7%</td>
</tr>
</tbody>
</table>

* 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.