MODIS-based Modeling of Corn and Soybean Yields in the US

David M. Johnson, Geographer
dave.johnson@nass.usdas.gov

United States Department of Agriculture
National Agricultural Statistics Service
www.nass.usda.gov

“... providing timely, accurate, and useful statistics in service to U.S. agriculture.”
United States Yield Year to Year Change %

- Corn
- Soybeans
NASS Crop Production reports

Corn Area Planted for All Purposes and Harvested for Grain, Yield, and Production – States and United States: 2009-2011 (continued)

<table>
<thead>
<tr>
<th>State</th>
<th>Yield per acre</th>
<th>Production</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2009 (bu)</td>
<td>2011 (bu)</td>
</tr>
<tr>
<td></td>
<td>(bushels)</td>
<td>(bushels)</td>
</tr>
<tr>
<td></td>
<td>(1,000 bushels)</td>
<td>(1,000 bushels)</td>
</tr>
<tr>
<td>Alabama</td>
<td>110.0</td>
<td>110.0</td>
</tr>
<tr>
<td>Arizona</td>
<td>170.0</td>
<td>180.0</td>
</tr>
<tr>
<td>Arkansas</td>
<td>140.0</td>
<td>140.0</td>
</tr>
<tr>
<td>California</td>
<td>180.0</td>
<td>180.0</td>
</tr>
<tr>
<td>Colorado</td>
<td>150.0</td>
<td>150.0</td>
</tr>
<tr>
<td>Connecticut</td>
<td>(NA)</td>
<td>(NA)</td>
</tr>
<tr>
<td>Delaware</td>
<td>135.0</td>
<td>135.0</td>
</tr>
<tr>
<td>Florida</td>
<td>180.0</td>
<td>180.0</td>
</tr>
<tr>
<td>Georgia</td>
<td>180.0</td>
<td>180.0</td>
</tr>
<tr>
<td>Hawaii</td>
<td>180.0</td>
<td>180.0</td>
</tr>
<tr>
<td>Illinois</td>
<td>170.0</td>
<td>170.0</td>
</tr>
<tr>
<td>Indiana</td>
<td>170.0</td>
<td>170.0</td>
</tr>
<tr>
<td>Iowa</td>
<td>160.0</td>
<td>160.0</td>
</tr>
<tr>
<td>Kansas</td>
<td>150.0</td>
<td>150.0</td>
</tr>
<tr>
<td>Kentucky</td>
<td>150.0</td>
<td>150.0</td>
</tr>
<tr>
<td>Louisiana</td>
<td>130.0</td>
<td>130.0</td>
</tr>
<tr>
<td>Maine</td>
<td>(NA)</td>
<td>(NA)</td>
</tr>
<tr>
<td>Maryland</td>
<td>140.0</td>
<td>140.0</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>140.0</td>
<td>140.0</td>
</tr>
<tr>
<td>Michigan</td>
<td>140.0</td>
<td>140.0</td>
</tr>
<tr>
<td>Minnesota</td>
<td>170.0</td>
<td>170.0</td>
</tr>
<tr>
<td>Mississippi</td>
<td>130.0</td>
<td>130.0</td>
</tr>
<tr>
<td>Missouri</td>
<td>120.0</td>
<td>120.0</td>
</tr>
<tr>
<td>Montana</td>
<td>150.0</td>
<td>150.0</td>
</tr>
<tr>
<td>Nebraska</td>
<td>170.0</td>
<td>170.0</td>
</tr>
<tr>
<td>Nevada</td>
<td>(NA)</td>
<td>(NA)</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>145.0</td>
<td>145.0</td>
</tr>
<tr>
<td>New Jersey</td>
<td>140.0</td>
<td>140.0</td>
</tr>
<tr>
<td>New York</td>
<td>130.0</td>
<td>130.0</td>
</tr>
<tr>
<td>North Carolina</td>
<td>110.0</td>
<td>110.0</td>
</tr>
<tr>
<td>North Dakota</td>
<td>130.0</td>
<td>130.0</td>
</tr>
<tr>
<td>Ohio</td>
<td>170.0</td>
<td>170.0</td>
</tr>
<tr>
<td>Oklahoma</td>
<td>150.0</td>
<td>150.0</td>
</tr>
<tr>
<td>Oregon</td>
<td>210.0</td>
<td>210.0</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>140.0</td>
<td>140.0</td>
</tr>
<tr>
<td>Rhode Island</td>
<td>140.0</td>
<td>140.0</td>
</tr>
<tr>
<td>South Carolina</td>
<td>110.0</td>
<td>110.0</td>
</tr>
<tr>
<td>South Dakota</td>
<td>145.0</td>
<td>145.0</td>
</tr>
<tr>
<td>Tennessee</td>
<td>140.0</td>
<td>140.0</td>
</tr>
<tr>
<td>Texas</td>
<td>130.0</td>
<td>130.0</td>
</tr>
<tr>
<td>Utah</td>
<td>150.0</td>
<td>150.0</td>
</tr>
<tr>
<td>Vermont</td>
<td>(NA)</td>
<td>(NA)</td>
</tr>
<tr>
<td>Virginia</td>
<td>130.0</td>
<td>130.0</td>
</tr>
<tr>
<td>Washington</td>
<td>130.0</td>
<td>130.0</td>
</tr>
<tr>
<td>West Virginia</td>
<td>130.0</td>
<td>130.0</td>
</tr>
<tr>
<td>Wyoming</td>
<td>140.0</td>
<td>140.0</td>
</tr>
<tr>
<td>United States</td>
<td>164.7</td>
<td>152.8</td>
</tr>
</tbody>
</table>

(*bu* is bushel. *NA* is not available; area harvested for grain not estimated.)

Crop Production 2011 Summary

**February 2012**

**Planted Acreage Update**

Survey respondents who reported acreage as not yet planted in Minnesota, Montana, North Dakota, and South Dakota during the survey conducted in preparation for the acreage report, released June 30, 2011, were re-estimated in July to determine how many of those acres were planted or still intended to be planted. Acreage estimates in this report reflect this updated information.

**Corn Production Up 4 Percent from 2010**

**Soybean Production Down 8 Percent from 2010**

**Cotton Production Down 9 Percent from 2010**

**All Wheat Production Down 1 Percent from July Forecast**

Corn production is forecast at 12.8 billion bushels, up 4 percent from 2010. If realized, this will be the third largest production total on record for the United States. Based on conditions as of August 1, yields are expected to average 153.0 bushels per acre, up 2 bushels from 2010, and the fourth highest yield on record. Acreage planted for all purposes is estimated at 92.3 million acres, unchanged from the June estimate. Area harvested for grain is forecast at 84.4 million acres, down less than 1 percent from June but up 4 percent from 2010.

Soybean production is forecast at 3.06 billion bushels, down 8 percent from last year. Based on August 1 conditions, yields are expected to average 41.4 bushels per acre, down 2 bushels from last year. Area for harvest in the United States is forecast at 73.8 million acres, down less than 1 percent from June and down 4 percent from 2010. Planted area for the Nation is estimated at 75.6 million acres, down fractionally from June.

All cotton production is forecast at 16.6 million 480-pound bales, down 9 percent from last year. Based on August 1 conditions, yields are expected to average 822 pounds per harvested acre, up 10 percent from last year. Upland cotton production is forecast at 15.8 million 480-pound bales, down 10 percent from 2010. American Pima production is forecast at 7.7 million bales, up 1 percent from last year. Producers expect to harvest 9.07 million bales of all cotton, down 10 percent from 2010. This harvested total includes 0.88 million bales of Upland cotton and 287.2 million bales of Pima cotton.

All wheat production, at 2.08 billion bushels, is down 1 percent from the July forecast and down 6 percent from 2010. Based on August 1 conditions, the United States yield is forecast at 45.2 bushels per acre, up 0.6 bushel from last month but down 1.2 bushels from last year.

Published no later than the 12th of each month.
Yields results primarily derived from two surveys

Agricultural Yield
• Farmer reported survey data of expected crop yields.
• Data obtained throughout the growing season.
• Conducted in all states except Alaska and Hawaii.
• Sample size in the 1000s per state.
• Farm operator contacts are selected from the March Crops/Stocks survey (small grains) and the June Crops/Stocks survey (late season crops and tobacco).
• Primarily telephone based.

Objective yield
• Corn, Cotton, Soybeans, Wheat, Potatoes.
• Only done in states where the commodities are primarily found.
• Samples selected from areas found in June Area Survey (“Acreage”).
• Performed at 100s of sample sites per state.
• Biophysical plant/seed measurements obtained.
• Each plot revisited a few times per season.
Estimating Yield from Remote Sensing

Third method for yield estimates

- There is a relationship between crop
  - Biomass, vigor, “greenness”, NDVI
  - and
  - Daytime land surface temperature
- And the resulting corn or soybean yield
- Utilize MODIS data to obtain biomass and temperature variables
- Do it for
  - National, State, “ASD”, and County levels
  - And map at pixel level
Moderate Resolution Imaging Spectroradiometer (MODIS)
Why MODIS?

- Global coverage
- Daily revisit rate
- 15 acre ground sample resolution
  - from red and near-infrared bands
- “Best of” image mosaics automatically generated
  - 8 and 16-day temporal windows
- Timely
  - data usually available within a couple of days
- Free distribution
  - downloaded via ftp
- Robust user group
  - nearly 20,000 citations so far
- Launched in 1999 and 2002
  - Two of them
  - 10-year plus history
- 6 year design life but still functioning fine
- Similar follow-on mission
  - VIIRS
MODIS NDVI data example
Calculation and use of NDVI

\[ \text{NDVI} = \frac{(\text{NIR} - \text{VIS})}{(\text{NIR} + \text{VIS})} \]

NIR = near-infrared
VIS = visible

Ranges from -1.0 to 1.0

NDVI is a related to:
- Plant health
- Chlophyll content
- “Greenness”
- Biomass
- Vegetation vigor
MODIS LST data example
Corn phenology fundamentals

Corn 5-year average 2006-2010

State
- Arkansas
- Illinois
- Indiana
- Iowa
- Kansas
- Louisiana
- Minnesota
- Missouri
- Mississippi
- North Dakota
- Nebraska
- Ohio
- Oklahoma
- South Dakota
- Wisconsin

Terra MODIS mean NDVI

16-day composite median date


8/1 9/1 10/1
Establishing the pixels that are only corn

FSA CLU/578 (early season)

- or -

NASS CDL (late season)

MODIS-scaled High probability sample of corn areas
Intersecting corn “mask” with MODIS data
County-level database developed

- Potential predictor variables (independent)
  - State (All major production Corn Belt states)
  - County (for each that had a published estimate, ~1000 of them)
  - Year (2006 – 2011)
  - 32 for each ranging every 8 days from February 18 – October 30
  - NDVI (derived from Terra “MOD09Q1” 250m)
  - Day LST (1:30 PM – from Aqua “MYD11A2” 1000m)
    - Thus 68 in total
- Forecast variable (dependent)
  - NASS published county level yield (available from NASS “Quickstats” webpage)
- Resulting database to evaluate ~5000 records
Corn yield dependence at county level
Corn Belt region, 2006-2011
Soybean yield dependence at county level
Corn Belt region, 2006-2011
Rulequest Cubist

CS_corn_5year

attribute definitions [CS_corn_5year.names]
training cases to be analyzed [CS_corn_5year.data]
test cases [CS_corn_5year.test]
rule-based model [CS_corn_5year.model]
output file [CS_corn_5year.out]

Model Construction Options

Form of Model:
- Rules alone
- Instances and rules
- Let Cubist decide

Use nearest

Committees of
- 5 members

Cross-validate

Use samples of
- % cases

Lock samples

Maximum rules
- 100

Extraction allowed
- 10 %

Unbiased rules

OK Defaults Cancel

Rules for CS_corn_5year

Rule 1/1: [47 cases, mean 64.59, range 18.3 to 135, est err 17.24]

if

MDR1.14 <= 6272.634
MDR1.22 <= 6741.595
EST_1330.02 <= 14405.63
EST_1330.04 <= 24906.01
then

yield = -90.46 + 0.0494 EST_1330.02 - 0.0046 EST_1330.04
+ 0.0499 LST_1330.04 + 0.8763 MDR1.31 + 0.0349 LST_1330.07
+ 0.0149 MDR1.03 + 0.0145 MDR1.25 + 0.015 LST_1330.23
+ 0.0083 LST_1330.03 + 0.0044 MDR1.20 - 0.0044 MDR1.12
+ 0.009 MDR1.22 + 0.6927 MDR1.29 + 0.0025 MDR1.11
+ 0.007 MDR1.21 + 0.6933 MDR1.24 + 0.0044 MDR1.25
+ 0.013 MDR1.29 + 0.004 LST_1330.30 + 0.004 LST_1330.32
+ 0.005 MDR1.34 + 0.8763 MDR1.10 + 0.8763 MDR1.32
+ 0.0027 MDR1.03 + 0.0036 MDR1.01 - 0.003 MDR1.19
+ 0.009 MDR1.04 + 0.0035 MDR1.02 + 0.0006 MDR1.20
+ 0.003 MDR1.23 + 0.6933 MDR1.26 + 0.003 MDR1.24 + 0.001 MDR1.32
+ 0.002 MDR1.33 + 0.8763 MDR1.31 - 0.002 MDR1.11 + 0.002 MDR1.33
+ 0.002 MDR1.34 + 0.006 MDR1.09 + 0.0066 MDR1.19
+ 0.004 MDR1.21 + 0.001 LST_1330.12

Rule 1/2: [44 cases, mean 100.00, range 49 to 162, est err 9.61]

if

MDR1.05 <= 5280.681
MDR1.14 <= 6272.634
MDR1.31 <= 4314.544
MDR1.14 <= 5280.681
LST_1330.04 <= 5280.681
EST_1330.05 <= 34902.5
LST_1330.12 <= 35486.29
LST_1330.32 <= 16195.42
then

yield = -1570.12 - 0.0972 EST_1330.06
+ 0.0749 LST_1330.01 - 0.038
+ 0.0373 MDR1.20 + 0.014 MDR1.11
+ 0.0383 LST_1330.04 + 0.011
+ 0.0211 MDR1.30.02 + 0.022
+ 0.0032 MDR1.27 + 0.0063 LST_1330.32

Rule 1/3: [37 cases, mean 104.21, range 37]

if

MDR1.04 <= 2563.265
MDR1.14 <= 6272.634
MDR1.14 <= 6272.634
MDR1.31 <= 4314.544
EST_1330.02 <= 14405.63
EST_1330.04 <= 24906.01
then

yield = -545.70 + 0.0189 MDR1.05 + 0.0136 MDR1.04 - 0.0109 MDR1.12
+ 0.0012 MDR1.16 + 0.0004 MDR1.10
Example county-level prediction output

<table>
<thead>
<tr>
<th>Case No.</th>
<th>Given Value</th>
<th>Predicted Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1701.0</td>
<td>92.49 + 24.78</td>
</tr>
<tr>
<td>2</td>
<td>1703.0</td>
<td>138.24 + 24.78</td>
</tr>
<tr>
<td>3</td>
<td>1705.0</td>
<td>64.58 + 24.78</td>
</tr>
<tr>
<td>4</td>
<td>1707.0</td>
<td>129.20 + 24.78</td>
</tr>
<tr>
<td>5</td>
<td>1709.0</td>
<td>106.46 + 24.78</td>
</tr>
<tr>
<td>6</td>
<td>17011.0</td>
<td>132.43 + 24.78</td>
</tr>
<tr>
<td>7</td>
<td>17013.0</td>
<td>104.15 + 24.78</td>
</tr>
<tr>
<td>8</td>
<td>17015.0</td>
<td>150.99 + 24.78</td>
</tr>
<tr>
<td>9</td>
<td>17017.0</td>
<td>127.98 + 24.78</td>
</tr>
<tr>
<td>10</td>
<td>17019.0</td>
<td>123.06 + 24.78</td>
</tr>
<tr>
<td>11</td>
<td>17021.0</td>
<td>122.79 + 24.78</td>
</tr>
<tr>
<td>12</td>
<td>17023.0</td>
<td>79.02 + 24.78</td>
</tr>
<tr>
<td>13</td>
<td>17025.0</td>
<td>45.38 + 24.78</td>
</tr>
<tr>
<td>14</td>
<td>17027.0</td>
<td>59.63 + 24.78</td>
</tr>
<tr>
<td>15</td>
<td>17029.0</td>
<td>94.61 + 24.78</td>
</tr>
<tr>
<td>16</td>
<td>17031.0</td>
<td>140.96 + 24.78</td>
</tr>
<tr>
<td>17</td>
<td>17033.0</td>
<td>65.67 + 24.78</td>
</tr>
<tr>
<td>18</td>
<td>17035.0</td>
<td>83.40 + 24.78</td>
</tr>
<tr>
<td>19</td>
<td>17037.0</td>
<td>143.90 + 24.78</td>
</tr>
<tr>
<td>20</td>
<td>17039.0</td>
<td>124.48 + 24.78</td>
</tr>
<tr>
<td>21</td>
<td>17041.0</td>
<td>125.98 + 24.78</td>
</tr>
<tr>
<td>22</td>
<td>17043.0</td>
<td>137.51 + 24.78</td>
</tr>
<tr>
<td>23</td>
<td>17045.0</td>
<td>104.61 + 24.78</td>
</tr>
<tr>
<td>24</td>
<td>17047.0</td>
<td>76.02 + 24.78</td>
</tr>
<tr>
<td>25</td>
<td>17049.0</td>
<td>66.14 + 24.78</td>
</tr>
<tr>
<td>26</td>
<td>17051.0</td>
<td>62.86 + 24.78</td>
</tr>
<tr>
<td>27</td>
<td>17053.0</td>
<td>101.90 + 24.78</td>
</tr>
<tr>
<td>28</td>
<td>17055.0</td>
<td>54.85 + 24.78</td>
</tr>
<tr>
<td>29</td>
<td>17057.0</td>
<td>109.45 + 24.78</td>
</tr>
<tr>
<td>30</td>
<td>17059.0</td>
<td>113.84 + 24.78</td>
</tr>
<tr>
<td>31</td>
<td>17061.0</td>
<td>104.00 + 24.78</td>
</tr>
<tr>
<td>32</td>
<td>17063.0</td>
<td>100.02 + 24.78</td>
</tr>
<tr>
<td>33</td>
<td>17065.0</td>
<td>64.23 + 24.78</td>
</tr>
<tr>
<td>34</td>
<td>17067.0</td>
<td>110.53 + 24.78</td>
</tr>
<tr>
<td>35</td>
<td>17069.0</td>
<td>69.21 + 24.78</td>
</tr>
<tr>
<td>36</td>
<td>17071.0</td>
<td>127.78 + 24.78</td>
</tr>
<tr>
<td>37</td>
<td>17073.0</td>
<td>121.29 + 24.78</td>
</tr>
<tr>
<td>38</td>
<td>17075.0</td>
<td>120.77 + 24.78</td>
</tr>
<tr>
<td>39</td>
<td>17077.0</td>
<td>63.20 + 24.78</td>
</tr>
<tr>
<td>40</td>
<td>17079.0</td>
<td>70.46 + 24.78</td>
</tr>
<tr>
<td>41</td>
<td>17081.0</td>
<td>66.27 + 24.78</td>
</tr>
<tr>
<td>42</td>
<td>17083.0</td>
<td>94.62 + 24.78</td>
</tr>
<tr>
<td>43</td>
<td>17085.0</td>
<td>123.96 + 24.78</td>
</tr>
<tr>
<td>44</td>
<td>17087.0</td>
<td>73.05 + 24.78</td>
</tr>
<tr>
<td>45</td>
<td>17089.0</td>
<td>131.46 + 24.78</td>
</tr>
<tr>
<td>46</td>
<td>17091.0</td>
<td>130.11 + 24.78</td>
</tr>
<tr>
<td>47</td>
<td>17093.0</td>
<td>120.26 + 24.78</td>
</tr>
<tr>
<td>48</td>
<td>17095.0</td>
<td>128.95 + 24.78</td>
</tr>
<tr>
<td>49</td>
<td>17097.0</td>
<td>120.66 + 24.78</td>
</tr>
<tr>
<td>50</td>
<td>17099.0</td>
<td>128.53 + 24.78</td>
</tr>
</tbody>
</table>

Corn

Weight by a 3-year average of harvested acres to derive ASD, state, and region estimates

Soybeans
“Voodoo Modeling”

Utilizing Rulequest Cubist software

– Learning tool to predict continuous (vs discrete) outcomes
– Allow for “composite” predictions using both
  • Instance-based
    – “Nearest neighbor”
    – Predicts the target value of a new case by finding the n most similar cases in the training data, and averaging their target values.
  • Model-based, via decision trees and piecewise linear regression
    – Divide and conquer strategy
    – Recursive splitting of training data to minimize intra-subset variation
  • Thus, for composite of instances and models:
    – Cubist finds the n training cases that are "nearest" (most similar) to the case in question. Then, rather than averaging their target values directly, Cubist first adjusts these values using the rule-based model.
– Also, does “Committee” models
  • made up of several rule-based models. Each member of the committee predicts the target value for a case and the members' predictions are averaged to give a final prediction
The relative error magnitude is the ratio of the average error magnitude to the error magnitude that would result from always predicting the mean value; for useful models, this should be less than 1!

The correlation coefficient measures the agreement between the cases' actual values of the target attribute and those values predicted by the model.
Soybean yield regression-tree model performance v. data timing county level, speculative region, 2006-2011

Similar story as for corn
Estimated Corn Yield
October 1, 2012
Reality check

Scene of a large hailstorm
2012 Results: Remote sensing vs NASS yield

State level average error

- corn = 5.8 bu./ac.
- soybeans = 3.1 bu./ac.
Models improvements for 2013

- **Corn**
  - relative |err|: 0.33 → 0.30
  - correl coeff: 0.93 → 0.95

- **Soybeans**
  - relative |err|: 0.31 → 0.30
  - correl coeff: 0.93 → 0.94

- **Absolute error unchanged**
  - ~8.0 bu/ac for corn, ~2.5 for soybeans
The end