Crop Progress and Condition Layers

United States Department of Agriculture

Arthur Rosales

National Agricultural Statistics Service

Research and Development Division

Washington DC 20250

Data Description

January 2021

Crop Progress and Condition Layers

January 2021

Arthur Rosales

Summary

Crop Progress and Condition Layers are gridded geospatial datasets which are fully synthetic representations of confidential, county level data. These new data are available for U.S. corn, soybeans, cotton, and winter wheat, at a weekly cadence during the growing season. The current archive of these datasets span growing-season weeks for all years from 2015 to present. This document serves as an introduction of these data and is intended to give a useful overview of their origins, as well as their characteristics and limitations.

This document is divided into four sections. Section 1 describes United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) procedures for creating the confidential, county-level data. Section 2 describes the method for creating the synthetic, gridded datasets which represent the county-level data. Section 3 briefly describes the file format, and section 4 provides a few notes that may be of interest to the data user, as well as contact information.

1. Overview of Raw County Data

NASS provides crop progress and condition estimates for selected crops on a weekly basis during the crop specific growing season.

1.1. Survey Procedures

Crop progress and condition estimates are obtained from data provided by the non-probability crop progress and condition survey. NASS reports typically cover the full U.S. growing season and are based on survey data collected each week from early April through the end of November. The non-probability crop progress and condition surveys include input from approximately 3,600 respondents whose occupations provide them opportunities to make visual observations and frequently bring them in contact with farmers in their counties. Based on standard definitions, these respondents subjectively estimate the progress of crops through various stages of development, as well as the progress of producer activities. They also provide subjective evaluations of crop conditions. Weekly reports are reviewed for reasonableness and consistency. Aggregation of the data to state and national levels relies on weights derived from historical NASS acreage estimates, the original county reports are unweighted.

1.2. Reporting Procedure

Most respondents complete their questionnaires on Friday or early Monday morning and submit them to their local state NASS Field Office by mail, telephone, fax, e-mail, or through a secured internet website. A small number of reports are completed on Thursday, Saturday, and Sunday. Regardless of when questionnaires are completed, respondents are asked to report for the entire week ending on Sunday. For reports submitted prior to the Sunday reference date, respondents must make projections for weekend changes in progress and condition. The majority of reports are submitted on Monday morning, via the secure website.

Respondents are sent written reporting instructions at the beginning of each season and are contacted periodically to ensure proper reporting. Terms and definitions of crop stages and condition categories used as reporting guidelines are available on the NASS website at www.nass.usda.gov/Publications/National_Crop_Progress.

1.3. Progress and Condition Definitions

Progress reporting is crop-specific and based on standard definitions of phenological stages. As a rule, any crop can progress from 0% planted to 100% harvested, see Table 1. Crop condition is based on subjective evaluations made by responding local experts, see Table 2. The condition of any crop for any week is represented in 5 exhaustive categories, ranging at the extremes from 100% very poor to 100% excellent.

Table 1. Example of Weekly Corn Crop Progress Data at the County Level

County	CCRNHVPG	CCRNMAPG	CCRNDEPG	CCRNDOPG	CCRNSIPG	CCRNEMPG	CCRNPLPG
47001	0	7	51	78	100	100	100
47003	0	55	100	100	100	100	100
47005	0	15	72	100	100	100	100
47007	0	30	69	94	100	100	100
47009	0	7	51	78	100	100	100
47011	0	7	51	78	100	100	100
47013	0	7	51	78	100	100	100
47015	0	19	100	100	100	100	100

CCRNHVPG, CCRNMAPG, CCRNDEPG, CCRNDOPG, CCRNSIPG, CCRNEMPG, CCRNPLPG correspond to "Harvested", "Matured", "Dented", "Doughed", "Silked", "Emerged", and "Planted" for corn progress.

Table 2. Example of Weekly Corn Crop Condition Data at the County Level

COUNTY	CCRNVPCD	CCRNPOCD	CCRNFACD	CCRNGOCD	CCRNEXCD
01001	2.68	0.67	36.6	49.33	10.72
01003	0	5	60	35	0
01005	60.75	24.75	8.75	5.75	0
01007	0	3	16	77	4
01009	0	6.36	22.09	57.18	14.37
01011	18.2	3.8	25	40.2	12.8
01013	0	8	64.5	26.5	1
01015	0	7.2	31.3	50.1	11.4

CCRNVPCD, CCRNPOCD, CCRNFACD, CCRNGOCD, CCRNEXCD correspond to "Very Poor", "Poor", "Fair", "Good", and "Excellent" for corn condition.

1.4. Restriction of Raw County Data

Although data are collected at the county level, the survey is designed to provide national and state estimates. This is done in part to protect the confidentiality of growers whose operations may comprise a vast majority of the production in a county. As a function of this design, it is common for only one or two respondents to complete a questionnaire for any given county for any given week. As such, estimates about county-level progress and condition based on raw county data alone are unstable compared to the current state estimates based on the aggregated data. However, this aggregation results in loss of information about spatial trends at the county level.

2. Dataset Creation

In response to increasing demand for higher resolution crop progress and condition data which reveal spatiotemporal trends within states, NASS has begun the creation of geospatially referenced, gridded datasets which represent the raw county data in a way that protects farmer confidentiality. This protection is achieved by taking several steps which abstract and obscure the original data, while preserving the valuable spatiotemporal trends that are of interest to researchers. These datasets are created in two main steps: (1) reinterpret the original, percentage-in-category variables into single numerical indices at the county level, (2) mapping and interpolating this new county data to cover relevant areas of the conterminous US. The following sections describe these steps.

2.1. Creating Crop Progress and Condition Indices

The process begins by creating weekly datasets of numeric indices which represent county-level crop progress or condition. While the calculation is slightly different for progress and condition,

in both cases there is a simple, arithmetic formulation based on original reported data. At this stage, all reported data has already been averaged at the county level across multiple respondents (assuming there was more than one respondent).

2.1.1. Crop Condition

Formulation for creating the Crop Condition Index is consistent for all crops. For any given week, the index is calculated as follows:

$$Condition = (5 * excellent + 4 * good + 3 * fair + 2 * poor + very\_poor)/100$$

Where *excellent*, *good*, *fair*, *poor*, and *very_poor* are each the percentage of crop reported to be in that quality category. This results in a numeric index with values ranging from 1 to 5 for each week in each county, where 1 corresponds with a crop that is 100% "Very Poor", and 5 corresponds with a crop that is 100% "Excellent". An example of how the original data corresponds with the resulting index is given in Table 3.

Table 3. Example of Weekly Crop Condition Index Data at the County Level

County	Excellent	Good	Fair	Poor	Very Poor	Condition Index
04029	100%	0%	0%	0%	0%	**5**
04031	80%	20%	0%	0%	0%	**4.8**
04033	25%	20%	50%	5%	0%	**3.65**
04035	0%	50%	50%	0%	0%	**3.5**
04037	20%	20%	30%	20%	10%	**3.2**
04039	0%	60%	25%	15%	0%	**3.45**
04041	20%	75%	5%	0%	0%	**4.15**
04043	40%	50%	10%	0%	0%	**4.3**

2.1.2. Crop Progress

The Crop Progress Index is more complicated than the Crop Condition Index, because the phenological stages are different for different crops. For the purpose of this document, the wheat progress formula will be showcased, but the process is similar for other crops.

The NASS recorded phenological stages for wheat are "planted", "emerged", "headed", and "harvested". The formula used to create the wheat Crop Progress Index is as follows:

$$Wheat\ Progress = (planted + emerged + headed + harvested)/400$$

This results in an index with values ranging from 0 to 1, where 0 represents 0% planted, and 1 represents 100% harvested. All crops have the same value range. Those who wish to tie this index value to the original phenological categories should be aware that this formula equally

spaces the phenological states across the range of the index, even though phenological stages are not equally spaced across time.

2.2. Creating Weekly Layers

Going forward, the process for creating the data layers is the same for both condition and progress, regardless of crop. For each crop, week, and product (crop progress or condition), a gridded layer is interpolated from the county level index data that was created using steps in Section 2.1.

2.2.1. Creating Point Datasets

In order to run the interpolation algorithm, the county-level data first needs to be expressed geographically as a set of points in space which correspond with latitude/longitude coordinates. It was decided to create a point dataset which corresponds with the general land-cover extent of the crop of interest. This was done by creating a set of polygons corresponding to counties, where each polygon represents the approximate crop extent of each county. Within each of these polygons, a set of random points was created, where the count of points is proportional to the county acreage for the crop of interest. Known cropland extent is based on recent years of the NASS Cropland Data Layers (CDL) and county crop acreage is based on NASS estimates. There are many counties with negligible cropland for the crops of interest but were provided with a point at the county centroid, in case crop progress and/or condition was reported for those counties. A mockup example of the corn points dataset for one year are given in Figure 1.

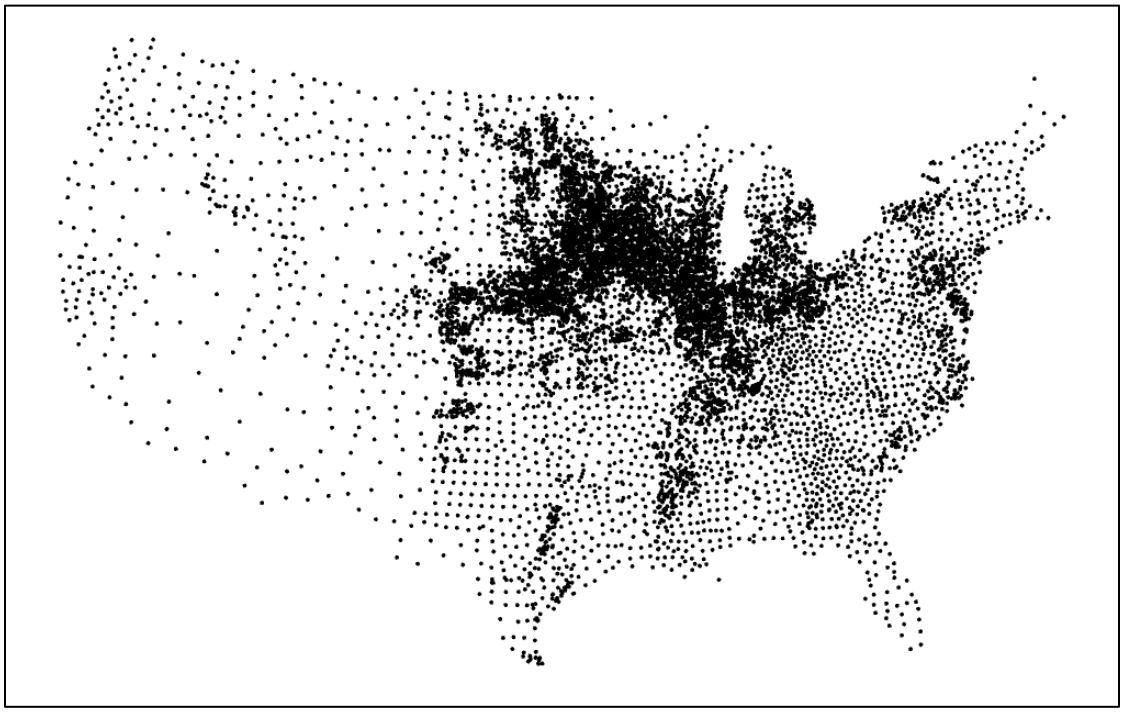

Figure 1. Example of acreage/location-based point dataset for use in kriging interpolation.

2.2.2. Kriging

Kriging is a spatial interpolation method that predicts variable values in unobserved locations based on variable values in observed locations. Kriging relies on spatial autocorrelation to fit a model which is then applied to the existing data to create a prediction surface. For these datasets, ArcGIS PRO software is used to link the numeric index data discussed in Section 2.1 to the point datasets described in 2.2.1. for each crop, week, and product. All points within a county boundary are given the same county index value. The ArcGIS PRO Simple Kriging algorithm is then used to create the prediction surface covering the lower 48 states based on the point data. Since the purpose of generating these datasets includes obscuring the original data, a "focal statistics" filter is applied to the resulting gridded datasets. A focal statistics algorithm averages gridded values across a local neighborhood of cells. This results in a smoothing effect, which diminishes outliers and rough artifacts in the kriging-based gridded layers. Interpolation can result in values outside the range of the original input data. Therefore, progress and condition index values outside the theoretical range were truncated to match the expected maximums and minimums. An example of a smoothed, kriging-based surface is given in Figure 2.

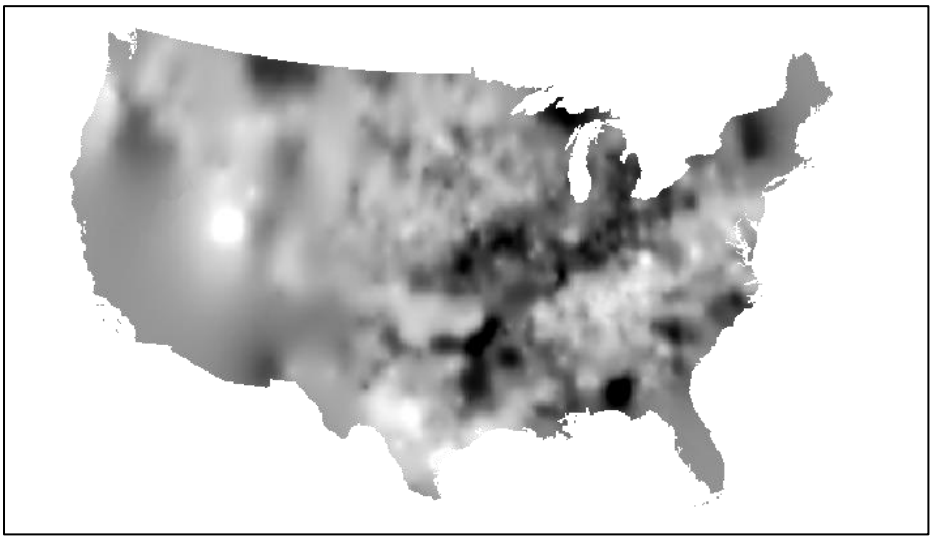

Figure 2. Example of weekly, smoothed gridded layers representing crop condition, masked to conterminous US.

2.2.3. State Masking

Finally, the smoothed gridded surfaces are masked only to states which had reported data. Of note, if even a single county within a state had reported data, the entire state will be visible. The entire process described in Section 2 of this paper is performed for every week, crop, and product for every year going back to 2015, resulting in the final Crop Progress and Condition Layers. Going forward, layers will be created within one or two days of the state-level reports currently made available by NASS.

3. File Format

All gridded datasets prepared are in the widely used geoTIFF format. The resolution is 9km and uses the NAD 1983 Contiguous USA Albers Projected Coordinate System. The data is stored in 32-bit floating format. “No data” is represented by the value -9999. The data layers are snapped to the NASA Soil Moisture Active Passive Satellite data, to facilitate ease of coordinated use between datasets. Each file follows the following naming convention:

*CropProductyy*w*ww*,

Where *Crop* indicates crop, *Product* indicates if the layer is for progress or condition, *yy* indicates year, and *ww* indicates NASS week. For example, *CornProg19w23* pertains to corn progress for 2019, NASS week 23.

4. Data Notes and Caveats

4.1. Different Experts Report Differently

The local experts who volunteer to report on crop progress and condition have varying reporting burdens and styles. A single expert may report for either a single county or multiple counties. Additionally, this results in varying levels of reporting precision depending on the week and region. For example, some experts may round their crop condition percent-in-category reports to even multiples of ten, whereas others will report with one or two decimal places. These irregularities may vary across space and time, with the number of experts and level of reporting precision usually increasing in areas of high importance to the crop, during the peak weeks of the growing season.

4.2. Crop Progress Layers Not Strictly Monotonically Increasing

For *in situ* crop progress, it is clearly impossible for the maturity to revert. True crop progress should be monotonically increasing. However, each of the weekly crop progress layers are created independently of data reported in previous weeks. This mostly does not create issues, but there is possibility for certain pixels in some weeks to show reduced progress when compared to the same location in a previous week. This is because the amount of data available changes throughout the growing season, and pixel values result from interpolation. Therefore, if new county data becomes available in a week which covers a location which previously was unreported and is less progressed than surrounding areas, the interpolated pixel values near that location may be lower than they were previously. Data users who are concerned by this should be aware of this possibility and be prepared to perform calculations to correct it if necessary.

4.3. Contact Information

For questions or comments, contact sm.nass.rdd.gib@usda.gov in the Research and Development Division (RDD) of USDA NASS.