Qualitative crop condition survey reveals spatiotemporal production patterns and allows early yield prediction

Santiago Beguería[a,1] and Marco P. Maneta[b,c,1]

[a]Department of Soil and Water, Estación Experimental de Aula Dei – Consejo Superior de Investigaciones Científicas (EEAD-CSIC), 50014 Zaragoza, Spain; [b]Geosciences Department, University of Montana, Missoula, MT 59812; and [c]Department of Ecosystem and Conservation Sciences, W.A. Franke College of Forestry and Conservation, University of Montana, Missoula, MT 59812

Edited by Ariel Ortiz-Bobea, Cornell University, Ithaca, NY, and accepted by Editorial Board Member Catherine L. Kling June 9, 2020 (received for review October 12, 2019)

Large-scale continuous crop monitoring systems (CMS) are key to detect and manage agricultural production anomalies. Current CMS exploit meteorological and crop growth models, and satellite imagery, but have underutilized legacy sources of information such as operational crop expert surveys with long and uninterrupted records. We argue that crop expert assessments, despite their subjective and categorical nature, capture the complexities of assessing the "status" of a crop better than any model or remote sensing retrieval. This is because crop rating data naturally encapsulates the broad expert knowledge of many individual surveyors spread throughout the country, constituting a sophisticated network of "people as sensors" that provide consistent and accurate information on crop progress. We analyze data from the US Department of Agriculture (USDA) Crop Progress and Condition (CPC) survey between 1987 and 2019 for four major crops across the US, and show how to transform the original qualitative data into a continuous, probabilistic variable better suited to quantitative analysis. Although the CPC reflects the subjective perception of many surveyors at different locations, the underlying models that describe the reported crop status are statistically robust and maintain similar characteristics across different crops, exhibit long-term stability, and have nation-wide validity. We discuss the origin and interpretation of existing spatial and temporal biases in the survey data. Finally, we propose a quantitative Crop Condition Index based on the CPC survey and demonstrate how this index can be used to monitor crop status and provide earlier and more precise predictions of crop yields than official USDA forecasts released midseason.

USDA | crop condition survey | crop condition index | crop monitoring | early yield prediction

The US Department of Agriculture (USDA) spends millions of dollars each year to collect, process, and disseminate to society information on agricultural production and markets. Farmers, agribusiness, and traders use this information for decision making and risk abatement, with demonstrated economic benefits and impacts on agricultural futures markets (1–8). Government agencies and research institutions also use this information for planning and research purposes, especially the world agricultural supply and demand estimates (WASDE), prospective planting and prospective acreage, crop production forecasts, or grain stocks reports. Very few studies, however, evaluate the effect of the Crop Progress and Condition (CPC) reports, despite these being one of the USDA products with the most significant number of subscribers and the one with the most frequent updates.

Issued weekly between early April and late November, the CPC report provides information on the progress (phenological states) and qualitative condition ratings of the most important crops grown in the United States. The report is based on an extensive voluntary survey conducted by agricultural extension agents, technicians, and other professionals in frequent contact with farmers. Thousands of surveyors in different states across the country are asked to assess the fraction of planted area that is at a specific stage of development (e.g., crop emergence, maturation, and various reproductive stages) and the fraction of the planted area that is in one of five qualitative condition categories. Raw data are gathered at the end of each week, and the official CPC report is issued at 4 PM Eastern Standard Time on the first business day of the following week.

With a fast turnaround time (the highest among the different USDA reports), the crop condition survey is a good and frequent indicator of the status of the crops at state and national levels and is an early indicator that can be used to anticipate positive or negative regional production anomalies. Fig. 1 plots the 2012 corn progress and condition report for the states leading the production of this crop to illustrate the value of this survey information for monitoring crop condition. Fig. 1 shows Likert-type plots with the weekly evolution of the survey during year 2012, the most adverse growing season for corn and soybeans in the United States in recent history. Likert plots show the proportion of responses over a finite set of specified categories, in this case the proportion of responses in each of the five categories assessing the status of a crop ("very poor," "poor," "fair," "good," and "excellent"). The plots assume equidistant categories and

ENVIRONMENTAL SCIENCES

AGRICULTURAL SCIENCES

Significance

We show how subjective information from qualitative crop rating surveys conducted weekly by the USDA can be transformed into a continuous crop condition index that integrates meteorological, agronomic, physiological, technological, and management factors. This index allows comparison of crop conditions between years and locations and provides superior information that enhances yield forecasting models. The proposed methodology can be used to develop better agricultural drought monitoring and early warning systems that can anticipate production anomalies and inform decision making.

Author contributions: S.B. designed research; S.B. and M.P.M. performed research; S.B. and M.P.M. analyzed data; and S.B. and M.P.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. A.O.-B. is a guest editor invited by the Editorial Board.

Data deposition: The CCI dataset and the R code of the yield prediction model may be downloaded from the institutional repository of Consejo Superior de Investigaciones Científicas (CSIC): http://hdl.handle.net/10261/201950.

[1]To whom correspondence may be addressed. Email: santiago.begueria@csic.es or Marco.Maneta@mso.umt.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1917774117/-/DCSupplemental.

First published July 16, 2020.

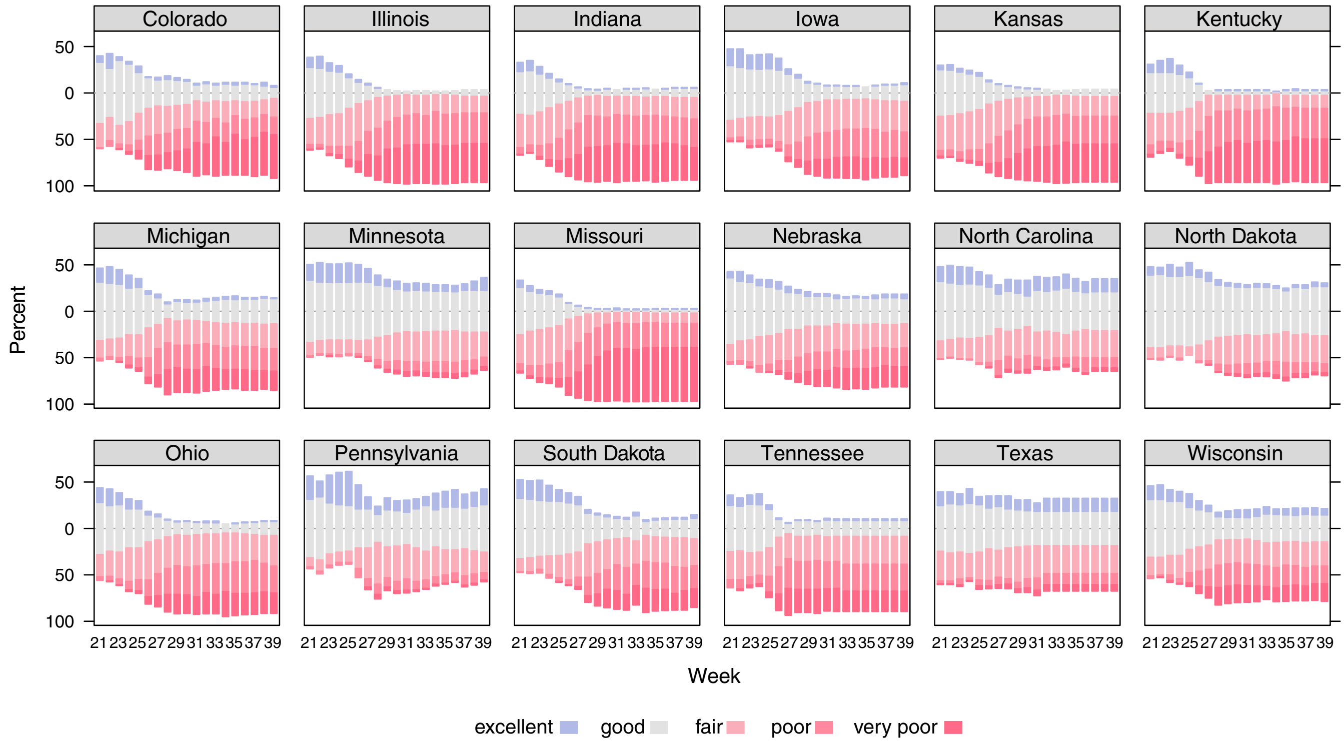

Fig. 1. Crop condition reports (Likert plots) for corn during 2012.

set the center of the scale in the middle of the good category, which is considered neutral according to the USDA instructions (*SI Appendix*, Table S1). After a relatively normal or even positive start of season at the end of May and early June, the survey responses become increasingly negative in most states as the prevailing drought conditions and unusually hot weather negatively affected corn crops.

Due to their capacity to inform about ongoing or developing crop anomalies that may affect production, the CPC reports influence futures markets of several traded commodities, as reflected by their rapid reaction upon the release of new CPC information (2, 4). Some studies have suggested that the sensitivity of markets to the CPC report has increased over recent years (9).

Despite its information content and long record and its demonstrated capacity to inform decision making, the USDA's CPC data are rarely used in scientific research. This may be reasonably attributed to the subjective nature of the data, which relies on the personal opinion of the surveyors. A few studies, however, show that CPC data provide valuable information for predicting crop yields (2, 10–15). The crop development data contained in the condition survey have also been used as a means to validate remote-sensing detection of crop progress and phenology at field scales (16–22). The difficulties of dealing with an ordinal variable in the context of quantitative models are certainly another reason for the limited use of this dataset. In the few studies that use CPC data, the researchers employ different transformations to convert ordinal data to another type more amenable to quantitative analysis. Some studies used a weighted sum of the five classes' percentage values using linear weights between 1 (for excellent condition) and 0 (very poor) to transform the ordinal variable to a continuous one (2, 12). The weights were chosen arbitrarily and were spaced at equal intervals, hence assuming an interval scale for the crop condition item. In other cases multivariate regression techniques were used to determine average yield corresponding to the different condition categories (11, 23). Other authors added the percentages of the two more positive classes (good and excellent) to generate a weekly index of crop goodness status (13–15). These approaches, however, did not solve the main problems with ordinal variables such as the magnitudes of intervals between categories or the precise location of the mean of the scale so positive and negative conditions can be effectively distinguished and quantified.

Evidence of Spatial and Temporal Effects in the Crop Condition Survey Data

To date, very few studies have analyzed the existence of systematic temporal and spatial biases in the CPC dataset. Such effects, however, can be reasonably expected in survey data based on the subjective perception of crop health and development. Spatial differences in the perception of crop status could arise from the very different environmental conditions in which crops are cultivated across the country or from technological differences such as the type of cultivars (varieties), cultivation practices, or cropping schedules. Also, the survey data span a relatively long time in which large technological and agronomic, market, and environmental changes have occurred. These effects, however, vary per state. An exploratory plot of the average scores of the survey by state, year, and week within the season reveals several sources of bias in the data (Fig. 2). For context, Fig. 2 also shows auxiliary information such as average yields per state and year and the most prevalent phenological state each week. In Fig. 2 we see noticeable crop status differences between states, and it could be hypothesized that these differences are related to the mean yields obtained in each case, with higher positive CPC scores in states reporting higher yields. There are also major differences between years, with low-yielding years obtaining a higher proportion of low CPC scores. Note that a long-term effect is apparent in the data: Due to technological advances, years that may be considered to have below-average yields at a point in time may have been considered above average during earlier years. The survey data seem to account for these known technology-driven long-term trends in crop productivity. Finally, and perhaps more surprisingly, there is evidence of biases within the growing season. Negative scores become more frequent as the season advances and the crop reaches critical phenological stages such as silking or doughing. The scores tend to recover slightly at the end of the season when the crop is mature in most fields. The existence of long-term effects as well as differences between start- and end-of-season scores has been documented for corn and soybeans at the national level (14, 24).

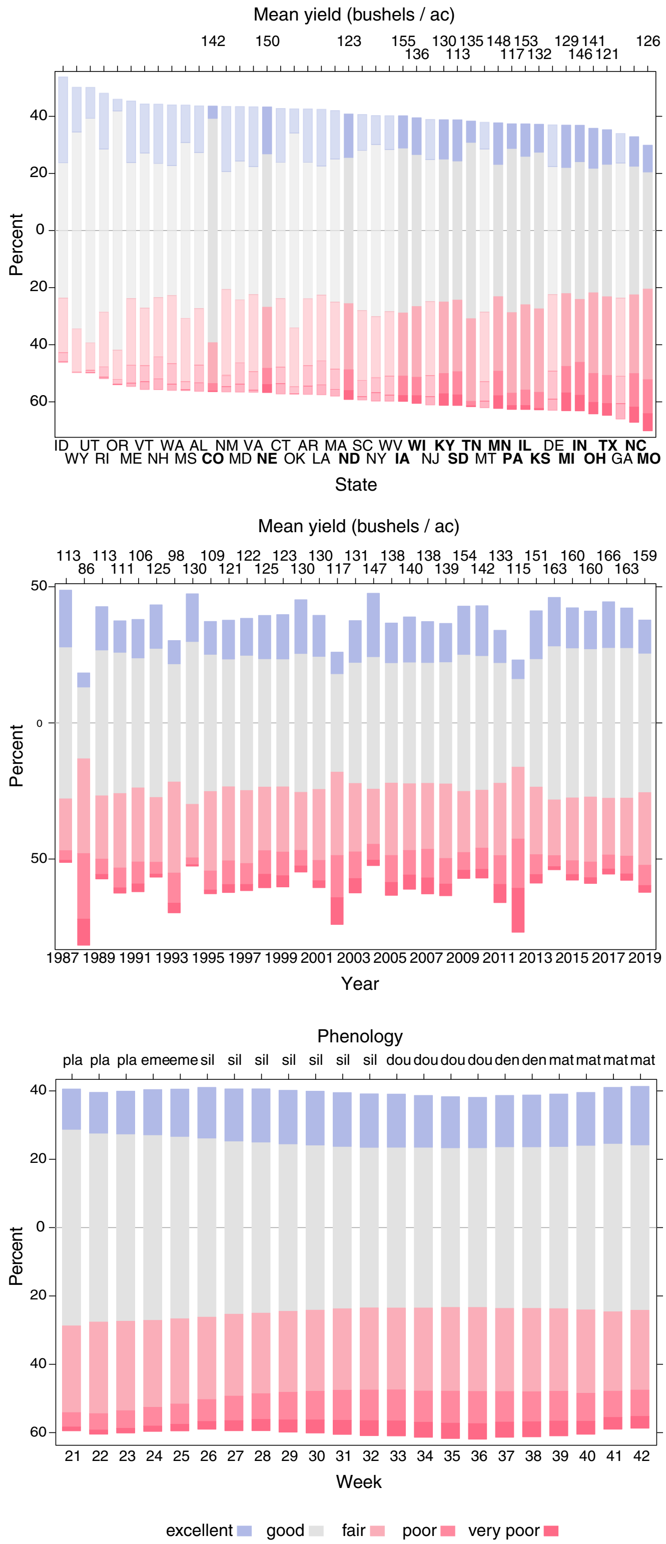

Fig. 2. Mean crop condition reports (Likert plots) for corn per state (*Top*), year (*Middle*), and week (*Bottom*). States highlighted in bold (*Top*) have available a complete and uninterrupted record from 1987 to 2018. Mean corn yields are shown in the upper axis for each year and for the states that have a complete record over the study period. The most frequent phenological state is also shown for each week, with the following codes: planted (pla), emerged (eme), silking (sil), dough (dou), dented (den), and mature (mat).

Formal Analysis

Our initial preliminary analysis of the crop condition data shows that further and more refined analysis is required to understand the nature of the dataset and maximize its information value. We conducted such analysis using a cumulative link mixed-effects model (25) with a form that allowed exploring the relevant spatial and temporal features of the dataset. The analysis revealed compelling human perception effects in the crop condition survey data not previously reported. However, and more importantly, the analysis resulted in the development of a homogenized, continuous crop condition index that can be used to compare relative crop development and health in space, in time, and within the growing season.

Similarities of the Underlying Models between Crops. Table 1 shows the linked mixed-effects model coefficients of the analysis of crop condition data. The θ coefficients are model intercepts that map the original ordinal variable (the crop condition survey classes) into a continuous variable and also provide a precise quantification of the metric distances between the survey categories. Interestingly, these coefficients were almost identical across crops, indicating the existence of an underlying perceptual model shared by all of the surveyors that does not change between crops. In other words, the survey categories (excellent, very poor, etc.) reflect the same degree of anomaly in the status of a crop, independent of the crop type. *SI Appendix*, Fig. S4 provides a graphic description of the models and shows the strikingly similar relative distance between categories in all crops. Note that the distribution of the perceptual model is left skewed: The mean of the distribution ($S = 0$) lies within the good category but toward its lower end close to the limit with the fair category. This is in agreement with the USDA definitions of the survey categories and implies that the survey reserves more categories (more granularity) to describe crops in worse-than-expected conditions.

Global Long-Term and Intraseasonal Effects. The β coefficients contain the fixed-effect terms of the analysis and have a global effect on the data. The long-term temporal effect (coefficient β_y) was very close to zero in all four crops and was statistically significant only for soybeans and cotton, indicating null or minimal long-term effects. The lack of a long-term trend in the mean of the CPC data is remarkable, considering the persistent increase in yields observed in the four crops during the study period (*SI Appendix*, Fig. S5). The analysis shows that surveyors account for this effect and adapt their scores to the expected crop performance associated with ever-improving technology and management practices.

Interestingly, the analysis showed the existence of a seasonal effect (coefficient β_w), which was significantly different from zero in all crops except winter wheat. Although the magnitude of this effect is low, as revealed by its SD it is interesting that the coefficients are negative, implying that condition scores tend to develop a low bias as the season develops. There is a logical explanation for this, since most growing seasons begin under normal conditions and more often than not develop normally throughout the season. Adverse events that negatively affect crops, on the other hand, occur less often but rapidly reduce the yield prospects from the normal expectation when the crop

Table 1. Model coefficients for the four crops: intercepts (θ), fixed effects (β), SD (σ), and correlations (ρ) of the random effects

	Corn	Soybeans	Cotton	Winter wheat
θ_1	−2.224*	−2.143*	−2.316*	−2.138*
θ_2	−1.504*	−1.357*	−1.487*	−1.360*
θ_3	−0.523*	−0.275*	−0.270*	−0.263*
θ_4	1.038*	1.318*	1.436*	1.367*
β_y	−0.040	0.078*	0.066*	0.007
β_w	−0.030*	−0.058*	−0.033*	−0.021
σ_s	0.614	0.141	0.346	0.347
σ_y	0.550	0.124	0.095	0.119
σ_w	0.044	0.034	0.063	0.059
ρ_{sy}	−0.948	−0.090	0.518	0.478
ρ_{sw}	0.354	0.198	0.585	0.361
ρ_{yw}	−0.197	0.163	0.169	0.363
σ_ϵ	0.477	0.433	0.496	0.252

*Significance at the confidence level $\alpha = 0.05$.

was planted. This generates the slightly high bias in the early season ratings and their apparent subsequent decline as the season progresses.

Differences across States. Since the CPC dataset is aggregated at the state level, it is possible to analyze the existence of spatial differences in the model parameters. This can be done by inspecting the random effect coefficients, which capture the variability of each state around the group-level intercept. The coefficient showing the highest variability (after the random term component, to be discussed later) was the state random coefficient (υ_s), which accounts for differences in the mean survey response across states (*SI Appendix*, Fig. S6). There is a clear relationship between the magnitude of the coefficient and the mean yields obtained in each state. States with high mean yields also had CPC responses with a higher positive bias. Similarly, states with low mean yields tended to have negatively biased CPC responses (*SI Appendix*, Fig. S7). The implication of this trend is that the perceptual model of the surveyors is homogeneous and spatially invariant: Surveyors could be randomly reassigned from one state to another, and their responses would still be representative of the local crop conditions.

The analysis also revealed differences in the long-term effects associated with each state, with some states showing positive values indicative of a temporal trend toward increasingly more positive survey scores and other states having negative values (*SI Appendix*, Fig. S8). These differences are related to the different yield trends experienced by each state. States where crop yields increased over time at rates higher than those of the group also tended to have positive coefficients and thus long-term trends in the crop condition survey response. The opposite is also true. This could be confirmed at least for corn and soybeans (*SI Appendix*, Fig. S9). Finally, our analysis showed that there is a seasonal effect that varies between states on a weekly basis (*SI Appendix*, Fig. S10) and that these differences are related to the timing of the main phenological stages of the crop, which also vary between states due to agro-climatic differences. This seasonal effect, however, is weaker than the long-term effect (*SI Appendix*, Fig. S11).

Random Effect. The residual coefficient (υ_r) explained a large fraction of the random variability of the CPC data, as shown by its variance. Once state, long-term, and seasonal effects are controlled by the corresponding terms of the model, this residual component represents the unbiased crop condition for each state and week. Since it has been formulated as a random effect per state, this component has a zero mean at the state level, with positive values indicating better-than-normal crop conditions (for a given state, year, and week), while negative values would indicate worse-than-normal conditions. For example, Fig. 3 shows the weekly change in the condition of corn, as represented by the random component of the model, in different corn-growing states during 2012. Fig. 3 illustrates how this component reflects the rapid worsening of the crop condition for the states that were affected by the severe drought that started between weeks 25 and 29. The histograms of the random component of the model adjusted for the four crops we consider in this study (Fig. 4) are left skewed, consistent with our previous discussion that the CPC has more categories defining worse-than-normal crop conditions than favorable crop conditions. The random component also displays differences across states (*SI Appendix*, Fig. S12), and it is interesting to confirm that there is a relationship between the variances of the random component and the yields recorded at each state (*SI Appendix*, Fig. S5).

Development of a Crop Condition Index

We have shown that, despite the subjective nature of the survey, the CPC data present highly robust characteristics across crops, states, and time. The main obstacle preventing a wider use of the CPC dataset is the ordinal character of the information. Our analysis framework, however, transforms the original data into a continuous variate, more suitable for mathematical analysis. The residual component of the model, once spatial and temporal biases and sources of variability have been eliminated, is therefore proposed as a quantitative crop condition index (CCI) (26). Because this CCI is a continuous variable, it can be used to monitor and assess the status of crops with a higher level of precision. Also, because the CCI is unbiased, it can be used to compare the status of crops between states, between years, and even between crops.

The CCI is, despite being negatively skewed, almost normally distributed; however, because it has different variances between states, it is not a fully standardized index. We have decided to leave the random component as it is and do not do any further transformation to standardize the CCI since these are intrinsic characteristics of the data that need to be preserved. In the next section, we provide an example of how the CCI can be used to provide early prediction of crop yields.

Early Prediction of Crop Yields Based on the Crop Condition Index

Early crop yield forecasts, along with crop acreage published by the USDA, are highly relevant for the agri-food sector. Accurate forecasts of yields are important to inform analysis and decision making. We develop a simple linear model of crop yields at the

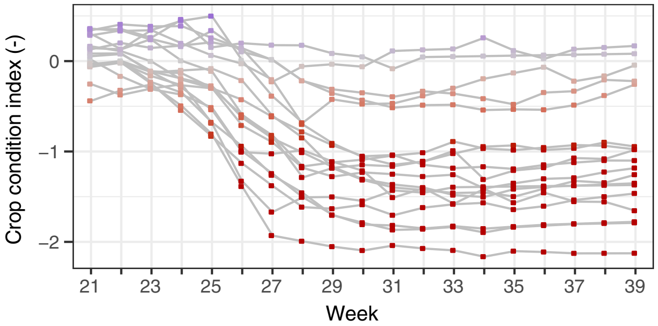

Fig. 3. Evolution of the crop condition index for corn during the 2012 growing season at different states. Positive values (in shades of blue) indicate better-than-normal conditions once spatial and temporal biases are accounted for, while negative values (red) indicate worse-than-normal conditions.

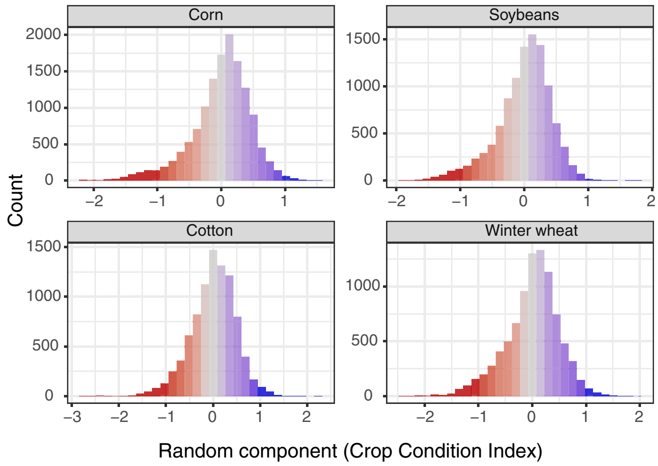

Fig. 4. Histograms of the random component (crop condition index) for the four crops, considering all states, years, and weeks. Color scheme is as in Fig. 3.

state level to demonstrate how the CCI can be used to monitor crop status and provide weekly predictions of yields. To evaluate the quality of these predictions we use a cross-validation approach to compare our weekly yield estimates with the end-of-season yield surveys conducted by the USDA. To further provide context, we compared the accuracy of our yield predictions based on the CCI with the USDA yield forecasts. The USDA forecasts are based on field surveys and farmer surveys and are usually reported three times during the time period when the CCI is available (27).

As an example, Fig. 5 depicts weekly predictions of corn yield in several states for the growing seasons 2015 to 2019. Predictions for the remaining crops and years are provided in *SI Appendix*, Figs. S14–S17. Fig. 5 includes instances of anomalous yields, like depressed corn yields in Missouri from the severe drought that affected the northern third of the state in the summer of 2018, or the anomalous conditions during the 2017 growing season in North and South Dakota associated with the flash drought that affected the Northern Plains. Similarly, it also contains instances of exceptionally good years like the record corn yields of 2018 in the Midwest. The plots also show the USDA forecasts, as well as end-of-season yield survey values. Yield predictions based on the CCI are typically very close to the USDA forecasts or occasionally are even closer to the target end-of-season yields. The plots illustrate the advantage of having weekly CCI data, which permits relatively accurate yield predictions many weeks before the first official USDA forecasts are issued. Fig. 6 shows the weekly evolution of goodness-of-fit, error, and bias statistics for CCI-based and USDA yield predictions. Additional model performance metrics at the national and state levels are provided in *SI Appendix*, Tables S4–S15. For comparison, in *SI Appendix*, Tables S4–S7 also include performance metrics for a null model used as control and from alternative models found in the literature. Both predictions achieve very high R^2 values, typically higher than 0.9 at the end of each crop season (0.75 in the case of cotton), and very low absolute errors. The R^2 values attained by the CCI-based predictions are similar to and in some cases higher than those of the USDA forecasts and achieve similar accuracy several weeks before the USDA releases the information. The CCI-based model also achieves better accuracy than other models that use raw crop condition data (11–13). Note that the model accuracy calculated through cross-validation, as used in this study, tends to produce more conservative goodness-of-fit values than the standard approach. The comparison with a null regression model shows that the model forecasting skill is provided by the CCI. Finally, it is also interesting to note that the yield predictions generated by the CCI-based model are virtually free of bias (mean error), while the USDA forecasts show a

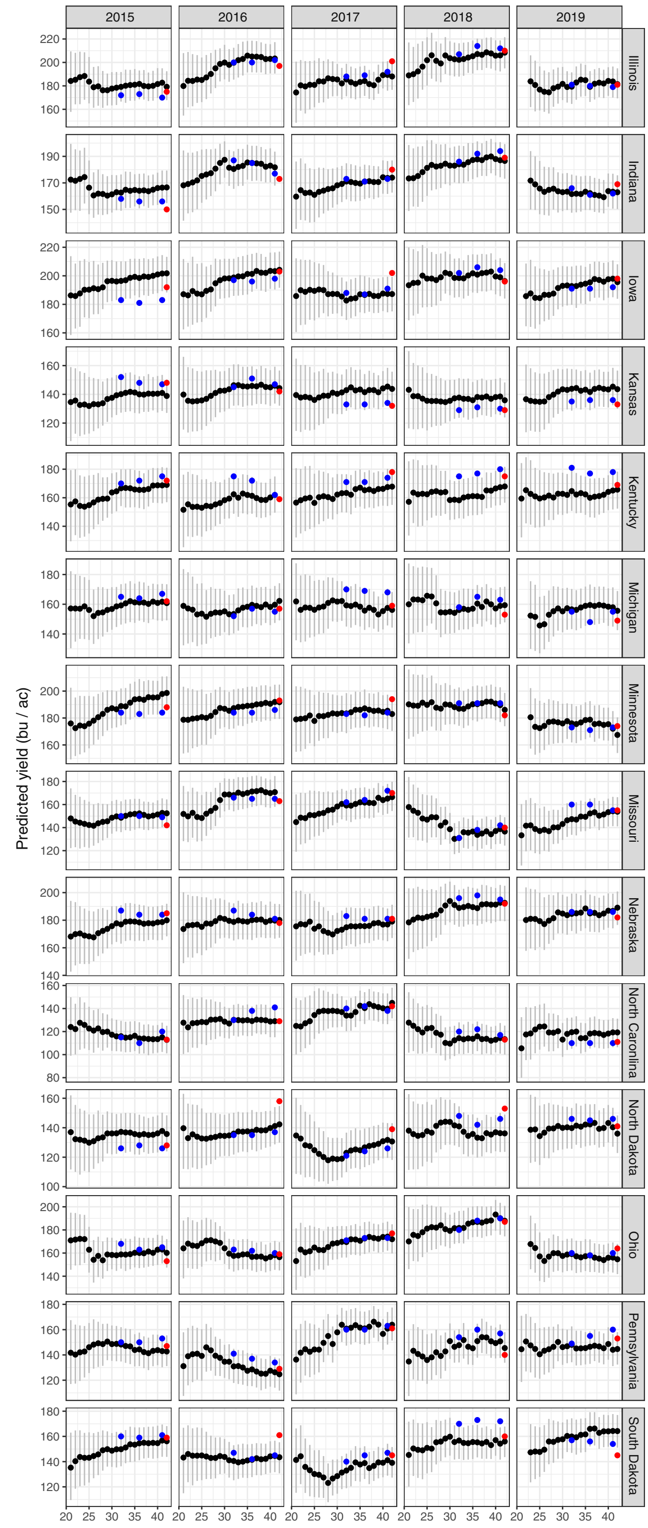

Fig. 5. Weekly prediction of corn yields per state, 2015 to 2019: linear regression model on crop condition index data (black dots), USDA forecasts (blue dots), and end-of-season USDA survey (red dot). The linear regression results have been obtained by excluding the year being predicted from the calibration set.

ENVIRONMENTAL SCIENCES

AGRICULTURAL SCIENCES

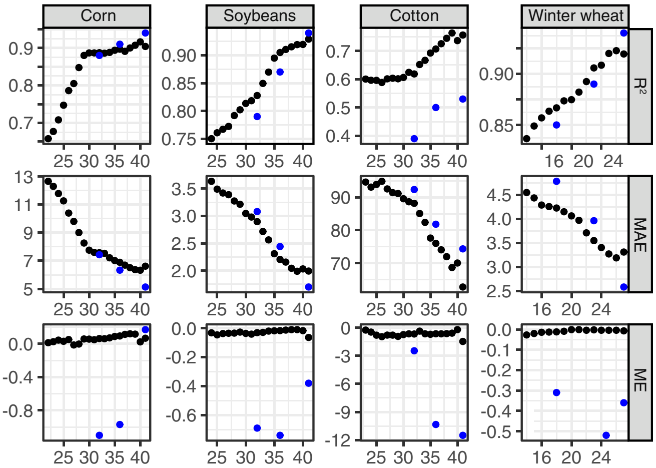

Fig. 6. Cross-validation statistics for corn yield predictions (mean values across states and years) based on a linear regression model upon crop condition index (black dots) and USDA yield forecasts (blue dots): coefficient of determination (R^2), mean absolute error (MAE), and mean error (ME).

slight negative bias. The spatial distribution of the goodness of fit is shown in Fig. 7.

Conclusion

Current crop-monitoring methodologies, operational yield forecasts, and early warning systems rely on remote-sensing imagery and crop models and use precipitation, temperature, or other meteorological data to detect anomalies, delineate their extent, and characterize their severity. Meteorological anomalies, such as drought, however, do not always affect agriculture because better management practices and technology often permit growers to maintain production under adverse climatic conditions. Without land surveys, many regions of the world can only rely on remote sensing and meteorological data as the basis for crop monitoring. In the United States, the weekly USDA crop condition survey provides an accurate assessment of the state of crops that integrates all relevant information and biophysical factors and considers specific local technological and management practices, as well as particular circumstances that may affect the timing and regular progress of a crop such as late planting dates. Our results demonstrate that a quantitative crop condition index can be developed based on the qualitative crop condition survey. This index permits the direct comparison of crop conditions across states and years, and its continuous nature makes it more amenable to be used in quantitative research. It also provides superior information that can be used to generate better operational crop monitoring and prediction systems at state scales in the United States.

Materials and Methods

Dataset. We downloaded crop condition data and other auxiliary variables (crop development, yields) from the Quick Stats database of the National Agricultural Statistics Service (28), for four major crops (corn, soybeans, winter wheat, and cotton), for the period between 1987 and 2018. Details on the data availability per state and crop are provided in *SI Appendix*, Table S2. Crop condition data consist of percentages for each of five condition classes (very poor, poor, fair, good, excellent), aggregated at the state scale, for each week during the growing season of each crop. It constitutes an example of an ordinal (or ordered categorical) variable, such as are widespread in scientific disciplines where humans are utilized as measurement devices, and Likert items are used to get information about a given problem. A Likert item is a simple question for which the response is codified on a discrete ordered scale ranging between two extreme values. The USDA crop condition survey can thus be considered a variety of a Likert item. Ordinal variables contain no metric information since the different levels of response do not indicate equal intervals between them. Therefore, a standard metric analysis is not feasible with ordinal data. There are techniques, however, suited to ordinal variables that allow answering relevant questions such as the distances between categories, the precise location of the mean category, or mean differences between different populations. One of such ways is the cumulative link model with a probit link, also known as the ordinal probit model. The cumulative link model allows transforming an original ordinal variable into a continuous, normal variate described by a mean and a SD. The process involves determining the threshold values that discriminate between the different classes of the variable, as is explained more formally in *SI Appendix*.

Preliminary Analysis. Likert plots (cumulative bar plots customarily used to portray ordinal variables) were used to explore crop condition survey data stratified per states, years, and weeks and help establish the model hypotheses. The reference line at 0 was set at the middle of the class fair, although further statistical analysis allowed us to determine more rigorously the mean of the distribution of the ordinal variable.

Statistical Analysis. A cumulative link mixed model (CLMM) was used to analyze the crop condition survey data. The model included a long-term linear trend component (variable year) and a seasonal component (variable week) as fixed effects and the state and interactions between state and year and state and week as random components. Another random component was included to represent the random variations that occur each week and on each state. This component reflects the crop condition anomalies, once the effects of state, year, and week have been accounted for. A probit link function was used to relate between these model components and the probabilities of each condition class,

$$\text{probit}\left(P\left(Y_i \le j \mid s, y, w\right)\right) = \theta_j + \beta_y y + \beta_w w + \upsilon_s + \upsilon_{y,s} y + \upsilon_{w,s} w + \epsilon_{si},$$

where $P(Y_i \le j)$ is the probability that condition of record i would correspond to class j or lower; s, y, and w are the state, year, and week corresponding to record i; θ_j is an intercept; β_y and β_w are model coefficients for the year and week fixed effects; $(\upsilon_s, \upsilon_y, \upsilon_w) \sim \mathcal{N}(\mathbf{0}, \Sigma)$ are multivariate normally distributed random intercept, year, and week effects; and $\epsilon_s \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is a random error. The latter term of the model (the random error) represents the crop condition index (CCI) for that particular state, year, and week, once all of the fixed and random effects have been accounted for. The model was fitted for each crop separately, and estimates of the model's parameters were obtained using a Laplace approximation to the maximum-likelihood function, as implemented in the ordinal R package (29).

Yield Prediction Model. We defined a hierarchical mixed-effects linear model of crop yields as

$$\mu_i(s) = \beta_0 + \beta_y\, y_i + \beta_c\, CCI_i + \upsilon(s) + \upsilon_y(s)\, y_i + \upsilon_c(s)\, CCI_i + \epsilon_i,$$

where $\mu_i(s)$ is the expected yield at state s and time i; β_0 is a global intercept; β_y and β_c are model coefficients for the long-term (year)

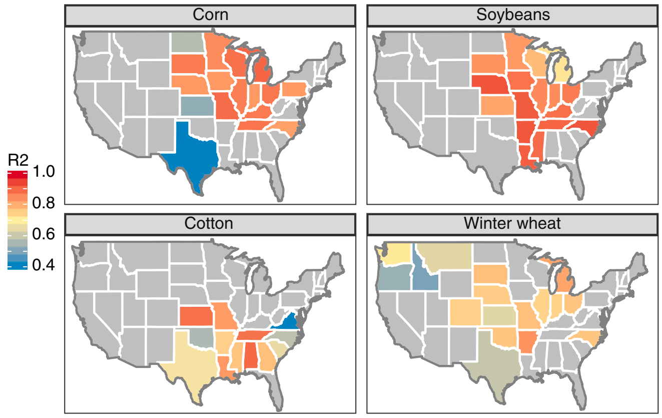

Fig. 7. Cross-validation (CV) R^2 values for end-of-season (weeks 41, 39, 41, and 25) yield predictions based on crop condition index, per state.

and CCI fixed effects; $(\upsilon, \upsilon_y, \upsilon_c) \sim \mathcal{N}(\mathbf{0}, \Sigma)$ are multivariate normally distributed random effects; and $\epsilon_s \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is a random error. The model was fitted for each crop and week during the season separately, and parameter estimates were obtained by the restricted maximum-likelihood (REML) method, as implemented in the lme4 R package (30, 31). *P* values for the fixed-effects coefficients were computed using the lmerTest package (32). Out-of-bag best linear unbiased predictions (BLUPs) were calculated for each crop and year in the dataset, which allowed for an unbiased assessment of the model's predictive power. The 95% prediction intervals around the BLUPs were estimated by drawing a sampling distribution for the random and the fixed effects and then sampling the fitted values across that distribution, as implemented in the merTools R package (33).

Data Availability. The CCI dataset and the R code of the yield prediction model may be downloaded from the institutional repository of CSIC: http://hdl.handle.net/10261/201950.

ACKNOWLEDGMENTS. S.B.'s research is supported by the Spanish Ministry of Science and Innovation (Grant CGL2017-83866-C3-3-R) and Fundación Biodiversidad, Ministry of Ecological Transition (Grant CA_CC_2018). M.P.M.'s support is from the USDA-National Institute of Food and Agriculture (NIFA) (Grant 2016-67026-25067) and the NASA Established Program to Stimulate Competitive Research (EPSCoR) program (Grant 80NSSC18M0025M). This research was made possible thanks to a mobility grant of the Spanish Ministry of Education, Culture, and Sports, within the framework of the "Programa Estatal de Promoción del Talento y su Empleabilidad en I+D+i," "Salvador de Madariaga" program.

1. B. Karali, Do USDA announcements affect comovements across commodity futures returns? *J. Agric. Resour. Econ.* **37**, 77–97 (2012).
2. R. Bain, T. R. Fortenbery, "Impacts of crop conditions reports on national and local wheat markets" in *NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management* (NCCC, St. Louis, MO, 2013), pp. 1–26.
3. G. V. Lehecka, "The reaction of corn and soybean futures markets to USDA crop progress and condition information," in *Southern Agricultural Economics Association Annual Meeting* (SAEA, Orlando, FL, 2013), pp. 1–24.
4. G. V. Lehecka, X. Wang, P. Garcia, Gone in ten minutes: Intraday evidence of announcement effects in the electronic corn futures market. *Appl. Econ. Perspect. Pol.* **36**, 504–526 (2014).
5. B. Karali, O. Isengildina-Massa, S. H. Irwin, M. K. Adjemian, "Changes in informational value and the market reaction to USDA reports in the big data era," in *Agricultural & Applied Economics Association Annual Meeting* (AAEA, Boston, MA, 2016), pp. 1–27.
6. O. Isengildina-Massa, B. Karali, S. H. Irwin, M. K. Adjemian, "The value of USDA information in a big data era" in *NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management* (NCC, Chicago, IL, 2016), pp. 1–28.
7. M. K. Adjemian, R. Johansson, A. McKenzie, M. Thomsen, Was the missing 2013 WASDE missed?*Appl. Econ. Perspect. Pol.* **40**, 653–671 (2017).
8. A. Fernandez-Perez, B. Frijns, I. Indriawan, A. Tourani-Rad, Surprise and dispersion: Informational impact of USDA announcements. *Agric. Econ.* **50**, 113–126 (2019).
9. G. V. Lehecka, The value of USDA crop progress and condition information: Reactions of corn and soybean futures markets. *J. Agric. Resour. Econ.* **39**, 88–105 (2014).
10. B. L. Dixon, S. E. Hollinger, P. Garcia, V. Tirupattur, Estimating corn yield response models to predict impacts of climate change. *J. Agric. Resour. Econ.* **19**, 58–68 (1994).
11. J. R. Kruse, D. B. Smith, "Yield estimation throughout the growing season" (Tech. Rep. 94-TR 29, Center for Agricultural and Rural Development, Iowa State University, Ames, IA, 1994).
12. N. Jorgensen, M. Diersen, Forecasting corn and soybean yields with crop conditions. Economics Commentator 547 (South Dakota State University, Brookings, SD, 2014).
13. S. Irwin, D. Good, When should we start paying attention to crop condition ratings for corn and soybeans? farmdoc daily 7, 24 May 2017.
14. S. Irwin, D. Good, How should we use within-season crop condition ratings for corn and soybeans? farmdoc daily 7, 1 June 2017.
15. S. Irwin, T. Hubbs, Measuring the accuracy of forecasting corn and soybean yield with good and excellent crop condition ratings. farmdoc daily 8, 27 June 2018.
16. B. D. Wardlow, J. H. Kastens, S. L. Egbert, Using USDA crop progress data for the evaluation of greenup onset date calculated from MODIS 250-meter data. *Photogramm. Eng. Rem. Sens.* **72**, 1225–1234 (2006).
17. H. Zhao, Z. Yang, L. Di, L. Li, H. Zhu, "Crop phenology date estimation based on NDVI derived from the reconstructed MODIS daily surface reflectance data" in *2009 17th International Conference on Geoinformatics* (IEEE, 2009), pp. 1–6.
18. T. Sakamoto, A. A. Gitelson, T. J. Arkebauer, MODIS-based corn grain yield estimation model incorporating crop phenology information. *Remote Sens. Environ.* **131**, 215–231 (2013).
19. T. Sakamoto, Refined shape model fitting methods for detecting various types of phenological information on major U.S. crops. *ISPRS J. Photogramm. Remote Sens.* **138**, 176–192 (2018).
20. L. Zeng *et al.*, A hybrid approach for detecting corn and soybean phenology with time-series MODIS data. *Remote Sens. Environ.* **181**, 237–250 (2016).
21. F. Gao, M. C. Anderson, D. Xie, "Spatial and temporal information fusion for crop condition monitoring" in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (IEEE, 2016), pp. 3579–3582.
22. F. Gao *et al.*, Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sens. Environ.* **188**, 9–25 (2017).
23. P. L. Fackler, B. Norwood, "Forecasting crop yields and condition indices" in *NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management* (NCR, Chicago, IL, 1999), pp. 1–18.
24. S. Irwin, T. Hubbs, Does the bias in early season crop condition ratings for corn and soybeans vary with the magnitude of the ratings? farmdoc daily 8, 20 June 2018.
25. T. M. Liddell, J. K. Kruschke, Analyzing ordinal data with metric models: What could possibly go wrong?*J. Exp. Soc. Psychol.* **79**, 328–348 (2018).
26. S. Beguería, M. P. Maneta, Qualitative crop condition survey reveals spatiotemporal production patterns [dataset], digital. CSIC. http://hdl.handle.net/10261/201950. Deposited 1 February 2020.
27. S. Irwin, D. Good, Opening up the black box: More on the USDA corn yield forecasting methodology. farmdoc daily 6, 26 August 2016.
28. USDA, National agricultural statistics service (NASS) quick stats database. https://quickstats.nass.usda.gov/. Accessed 15 July 2019.
29. R. H. B. Christensen, *ordinal: Regression Models for Ordinal data, R Package Version 2019.3-9*. http://www.cran.r-project.org/package=ordinal/. Accessed 15 July 2019.
30. D. Bates, M. Maechler, B. Bolker, S. Walker, *lme4: Linear Mixed-Effects Models Using 'eigen' and s4, R Package Version 1.1-21*. http://www.cran.r-project.org/package=lme4/. Accessed 15 July 2019.
31. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. *J. Stat. Software* **67**, 1–48 (2015).
32. A. Kuznetsova, P. B. Brockhoff, R. H. B. Christensen, lmerTest package: Tests in linear mixed effects models. *J. Stat. Software* **82**, 1–26 (2017).
33. J. E. Knowles, C. Frederick, *mertools: Tools for Analyzing Mixed Effect Regression models, R Package Version 0.3.0*. https://CRAN.R-project.org/package=merTools. Accessed 15 July 2016.