Can Satellite Data Forecast Valuable Information from USDA Reports ? Evidences on Corn Yield Estimates

by

Pierrick Piette

Suggested citation format:

Piette, P. 2019. "Can Satellite Data Forecast Valuable Information from USDA Reports ? Evidences on Corn Yield Estimates." Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. Minneapolis, MN. [http://www.farmdoc.illinois.edu/nccc134].

Can Satellite Data Forecast Valuable Information from USDA Reports ? Evidences on Corn Yield Estimates.

Pierrick Piette[1]

Paper presented at the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management Minneapolis, Minnesota, April 15-16, 2019.

[1] Pierrick Piette is a PhD candidate in the Laboratoire de Science Actuarielle et Financière at the Université Claude Bernard Lyon 1, in partnership with Sinalys. Email: pierrick.piette@gmail.com. The author is greatful to Claudio Busarello, Kevin Dedieu, Tanguy Touffut and Madeleine Latapie as well as the NCCC-134 conference participants for their helpful insights and comments.

Can Satellite Data Forecast Valuable Information from USDA Reports ? Evidences on Corn Yield Estimates.

On the one hand, recent advances in satellite imagery and remote sensing allow one to easily follow in near-real time the crop conditions all around the world. On the other hand, it has been shown that governmental agricultural reports contain useful news for the commodities market, whose participants react to this valuable information. In this paper, we investigate wether one can forecast some of the newsworthy information contained in the USDA reports through satellite data. We focus on the corn futures market over the period 2000-2016. We first check the well-documented presence of market reactions to the release of the monthly WASDE reports through statistical tests. Then we investigate the informational value of early yield estimates published in these governmental reports. Finally, we propose an econometric model based on MODIS NDVI time series to forecast this valuable information. Results show that market rationally reacts to the NASS early yield forecasts. Moreover, the modeled NDVI-based information is significantly correlated with the market reactions. To conclude, we propose some ways of improvement to be considered for a practical implementation.

Keywords: Commodities market; Corn; Market information; NDVI; Satellite data; USDA reports.

1 Introduction

The value of public information in agricultural commodity markets has been a topic of great attention for many years in the literature (e.g., Summer and Mueller, 1989; Garcia et al., 1997; Isengildina-Massa et al., 2008; Dorfman and Karali, 2015; Gouel, 2018). More recently, with the improvement of data access and the emergence of the *Big Data* era, the interest in this subject has enhanced, especially on the possible declining value of USDA reports (Karali et al., 2019; Tack et al., 2017; Ying et al., 2017). Indeed, the recent advances in satellite data and remote sensing allow one to easily follow in near-real time the crop progress all over the world thanks to weather data or vegetation index (e.g., Prasad et al., 2006; Mkhabela et al., 2011). Nevertheless, untill now, studies show that USDA reports still have a significant impact on the commodities market. The purpose of this paper is to provide evidences that some valuable information contained in the governmental reports can be forecasted by using satellite data. Our regression analysis shows significant correlation between our satellite data-based forecasted information and the corn futures market reactions to the report releases. We also suggest some needed enhancements of our methodology for a practical application.

The monthly World Agricultural Supply and Demand Estimates (WASDE) report, provided by the United States Department of Agriculture (USDA), together with, for some specific months, the National Agricultural Statistics Service (NASS) Crop Production report, are the most valuable public sources of information for the U.S. commodities market. An extensive literature already examines the impact of these governmental report releases on the commodities market.

Summer and Mueller (1989) were the first to use an event-study framework to explore the impact of USDA reports on corn and soybean future prices over the period 1961-1982. They conclude that the information included in the governmental forecasts are considered as new and reliable by the market participants, since the changes in prices on days just following the publication are significantly higher than on other days. Isengildina-Massa et al. (2008) find consistent results by testing differences in variance of financial returns over the period 1985-2006 for the same commodities. Other similar studies have been conducted, and a large majority of them conclude on the fact that the commodities market is significantly impacted by the WASDE report releases (Garcia et al., 1997; Irwin et al., 2001; McKenzie, 2008).

On the contrary to the previous investigations, which rely on statistical tests such as F or Chi-squared test, other studies base their analysis on a regression model. Thus, Fortenbery and Sumner (1993) regress the changes in prices on, amongst other, a zero-one dummy variable for report dates. Lehecka (2014) quantifies the futures price reaction by using the crop condition information included in the weekly Crop Progress reports over the period 1986-2012. The author regresses the close-to-open return on the change in the percentages of the crop in excellent and good condition. The results highlight a significant and negative influence of the crop condition improvement on the prices change, which is a rational outcome in respect with the supply and demand theory.

More recently, other concerns have been raised with the significant increase of available information from multiple sources and the essor of *Big Data*. Intuitively, the augmentation of existing data sources should lower the impact of the USDA reports on the commodities market. However, Karali et al. (2019) and Ying et al. (2017) examine this hypothesis and conclude that the effects are not decreasing. Furthermore, it even seems that the informational value of some reports tends to enhance. Alternatively, Milacek and Brorsen (2017), based on the assumption that a private firm would have developed a WASDE reports forecasting model for trading purpose, determine the informational value of this latter. Abbott et al. (2016) also quantify the value of corn information contained in the WASDE reports to \$301 million, and, in particular, assess the corn yield information to \$188 million.

One of the data sources that private firms might use to forecast the USDA reports is satellite imagery. In this paper we focus on a specific vegetation index: the Normalized Difference Vegetation Index (NDVI), which is a reflectance index. Colwell (1956) was the first to explore the detection of crops healthiness by aerial infrared photographs. Kumar and Silva (1973) study into details the link between one crop reflectance and its chlorophyll activity. Indeed, they remark a specific signature in the near-infrared. Introduced by Rouse et al. (1974), the NDVI is nowadays one of the most used and studied vegetation index and is defined as follows

$$NDVI = \frac{R_{NIR} - R_R}{R_{NIR} + R_R}, \tag{1.1}$$

where R_{NIR} and R_R are the reflectance in the near-infrared and in red, respectively. Thus, a dense tropical forest NDVI value is positive from 0.6 to 0.8, while the bare soil leads to lower value around 0.1, and rock or snow have negative NDVI values.

The NDVI is widely used in agricultural remote sensing, and more specifically in crop yield forecasting. Many forecasting models have been developed based on simple linear regression, in particular for corn (Prasad et al., 2006), soybeans (Ma et al., 2001) and wheat (Mkhabela et al., 2011). Other more sophisticated algorithms have also been applied, such as Li et al. (2007) who use neural networks. Different variables, derived from the NDVI time series, can be used for yield forecasting. Thus, Mkhabela et al. (2011) estimate crop yield thanks to the mean NDVI value over the growth period, while Zhang et al. (2012) split the series into two distinct periods: from re-greening to heading and from heading to maturity. Numerous models exist due the specificities of each research project. Indeed, studies have been conducted in many countries such as, for examples, Zimbabwe (Svotwa et al., 2014), Hungary (Ferencz et al., 2004), China (Ren et al., 2008) or the U.S. (Prasad et al., 2006; Becker-Reshef et al., 2010a). Different data sources can be used, like the Advanced Very High Resolution Radiometer (AVHRR) (Rasmussen, 1997), the Moderate Resolution Imaging Spectroradiometer (MODIS) (Doraiswamy et al., 2005), or Sentinel-2 (Skakun et al., 2017a). Applications cover a large number of crop types (corn, soybeans, wheat, barley, rice, tobacco, potato, sugarcane, etc.). Furthermore, in some studies, authors associate NDVI with weather data, like rainfall estimates and humidity index (Prasad et al., 2006), to improve the forecasting model accuracy. In general, these studies conclude to a high and significant correlation between NDVI and crop yield, with a R^2 regularly up to 0.9.

More recently, with the improvement of satellite data resolution, more sophisticated remote sensing studies have emerged. Thus, Pervez and Brown (2010) manage to determine if a land is irrigated or not with an accuracy of 92% in California, while Peterson et al. (2011) detect the irrigation for different crops in Kansas with an accuracy of 88%. Beyond irrigation detection, crop mapping studies have gained even more interests (e.g., Wardlow and Egbert, 2008; Wardlow and Egbert, 2010; Skakun et al., 2017b; Gao et al., 2017; Zhong et al., 2019). The development of such maps is useful since it significantly improve the crop yield forecasting models (Maselli and Rembold, 2001; Kastens et al., 2005).

In this paper, we investigate wether one can forecast some of the valuable information from the USDA reports through satellite open-access data. We focus on the corn yield early estimates information from the NASS reports of August, September and October over the period 2000-2016, by observing the reaction of the corn futures with a maturity in December. The chosen satellite data is the NDVI values derived from MODIS abroad the NASA's satellites Terra and Aqua.

First, we test, on the data we base our study on, that the commodity market rationally reacts to the early yield estimates from the NASS reports. To do so, we follow the general econometric methodology of Lehecka (2014) to model the valuable information contained in the USDA announcements. Then, during the growing season, we use the MODIS NDVI data to forecast the corn yield through linear regression models, trained on the final crop yields. Finally, we focus on three specific NDVI-based estimates, which are the ones obtained around two weeks before the publication of the WASDE report, therefore forecasting the valuable information contained in the next governmental report. Following again the methodology of Lehecka (2014), we test the correlation between our NDVI-based forecasts of valuable information and the commodity market reactions.

Our findings show significant and rational correlations between our modeled information contained in the reports and the financial returns of corn futures. This result is in accordance with the exisiting literature (Irwin et al., 2001; Lehecka, 2014). Moreover, the correlation between the financial returns and our NDVI-based forecasted information is also significant and in line with the supply and demand theory. Since the NDVI data needed is available, at least, one week before USDA announcements, it highlights the possibility of knowing in advance some of the valuable information contained in the reports.

The remainder of this paper is organized as follows. In Section 2 we describe the data we base our study on. The methodology we apply is defined in Section 3. We present the results that we obtained in Section 4. Furthermore, Section 5 considers future possible improvements of the current work for a practical implementation. Finally, Section 6 concludes.

2 Data

2.1 NDVI Data

In this study, we use the MODIS NDVI data, available in open facilitated access thanks to the Global Agriculture Monitoring Project (Becker-Reshef et al., 2010b) initiated between the NASA, the University of Maryland and the USDA Foreign Agriculture Service (FAS). The geographic resolution is of 250 meters and the temporal one is 16 days (MOD44 16-days product). Thus, every 16 days, we obtain a mean image of the period. This methodology is due to the fact that NDVI is a reflectance index, and is therefore sensitive to weather conditions such as clouds when estimated from satellite (Whitcraft et al., 2015). The release of these images is done at a constant pace in the year (cf. Table 1). We also apply the standard water and crop mask (MOD 12) to focus on the crop conditions. MODIS data are available since February 2000.

2.2 Futures Data

On the contrary to Irwin et al. (2001) and Isengildina-Massa et al. (2008), we don't explore the market reaction through the futures whose maturity are the closest to the session of interest. We rather base our study on the corn futures with the December maturity of the very year. Since our results may be useful for the agricultural insurance companies' risk management, we focus on the price at risk for the revenue protection products, which is the new-crop harvest price, i.e., December for corn in the U.S.

Table 1: NDVI image periods and corresponding calendar dates.

Period	Starting Date	Ending Date
1	01 Jan.	16 Jan.
2	17 Jan.	01 Feb.
3	01 Feb.	17 Feb.
4	18 Feb.	05 Mar.
5	06 Mar.	21 Mar.
6	22 Mar.	06 Apr.
7	07 Apr.	22 Apr.
8	23 Apr.	08 May
9	09 May	24 May
10	25 May	09 Jun.
11	10 Jun.	25 Jun.
12	26 Jun.	11 Jul.
13	12 Jul.	27 Jul.
14	28 Jul.	12 Aug.
15	13 Aug.	28 Aug.
16	29 Aug.	13 Sep.
17	14 Sep.	29 Sep.
18	30 Sep.	15 Oct.
19	16 Oct.	31 Oct.
20	01 Nov.	16 Nov.
21	17 Nov.	02 Dec.
22	03 Dec.	18 Dec.
23	19 Dec.	03 Jan.

2.3 Crop yield estimates

Finally, to model the valuable information contained in the NASS reports, we use the early corn yield estimates of 10 states from the Corn Belt, namely Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Nebraska, Ohio, South Dakota and Wisconsin. Indeed, these states represent between 82 and 85% of the corn production in the United States. More particulary, we explore the market reaction to the release of August, September and October yield estimates over the period 2000-2016.

In order to train our NDVI-based yield forecasting model, we also use the final corn yield estimates, communicated by the NASS in the January following the harvest.

3 Methodology

3.1 Event study analysis

Following already existing event study methodologies (Irwin et al., 2001; Isengildina-Massa et al., 2008), we first check that the commodity market reacts to the release of the reports over the period of interest. We focus our research on the future returns to test if the variability is higher in the trading session just after the WASDE release than normal, i.e., sessions in a temporal window of 5 days before and after. Since we carry out our study over the period 2000-2016, we need to pay attention to the futures return definition. Indeed, before the year 2013, USDA reports were released at 8:30 a.m., while, since January 1, 2013, the statistical reports have now been published at 12:00 p.m.[2] Thus, if a market reaction exists, its timing might have changed over the years.

Therefore, we define the futures return of interest as

$$r_{t,i,N} = \ln\left(\frac{p^O_{t,i,N}}{p^C_{t-1,i,N}}\right) \times 100, \tag{3.1}$$

if the release happened strictly before 2013, and

$$r_{t,i,N} = \ln\left(\frac{p^C_{t,i,N}}{p^O_{t,i,N}}\right) \times 100, \tag{3.2}$$

if the report publication occured after January 1, 2013; and where $p^O_{t,i,N}$ and $p^C_{t,i,N}$ are respectively the opening and the closing price of corn futures with a maturity in December of year N for the session t of the event (i, N), i.e., the WASDE release of

[2] https://www.usda.gov/media/press-releases/2012/09/19/usda-announces-change-release-time-key-statistical-reports

month i of year N. Descriptive statistics of the defined financial returns are displayed in Table 2.

Then, we perform a F-test of equality of variances to compare the variability of the returns in the session just following the release of the report ($t = 0$) with the sessions of the event window ($t \in \{-5, \ldots, -1, 1, \ldots, 5\}$). Under market efficiency assumption, if the market participants consider that the reports contain new and reliable information, the futures price variability should be higher on the announcement date.

Table 2: Descriptive statistics of corn future returns over the period 2000-2016.

	r	$\lvert r \rvert$
Mean	0.009	0.772
Median	-0.044	0.456
1st Qu.	-0.468	0.189
3rd Qu.	0.442	0.442
Min	-17.477	0.000
Max	24.429	24.429
Variance	1.939	1.343

3.2 Regression analysis - USDA reports

The tests described in Section 3.1 are only designed to detect that new and reliable information is contained in the reports. However, this methodology doesn't allow us to know to which information the market exactly reacts to. Indeed, WASDE and NASS reports contain a large amount of statistics. In particular, Abbott et al. (2016) show that corn yield forecasts represent a significant value of \$188 million for the market. Inspired by Lehecka (2014) methodology, we test if the crop yield estimates are considered as valuable information.

Indeed, the early yield estimates are crucial data for the commodity market in the growing season. With the harvested area, it gives an early forecast of the next harvest production and, therefore, an outlook of the supply. All other factors being equal, a change in the crop yield forecasts should impact the futures price. More precisely, if the USDA scales up her expectation in terms of yield, the futures price should decrease, since the supply augments.

In the NASS report of month i, published in the year N, new early corn yield estimates, noted $Y_{k,i,N}$, are displayed for each state of interest k. We also note $Y_{k,N}$ the final yield estimates of state k for the year N. We focus on the reports released in August, September and October. In the current paper, we suppose that no other information is integrated by the market between two WASDE reports publication (this latter hypothesis will be discussed later in Section 5). Thus, for September and October, we model the new information provided by the report $i \in \{9, 10\}$ of year N as

$$X_{k,i,N} = \ln\left(\frac{Y_{k,i,N}}{Y_{k,i-1,N}}\right) \times 100. \tag{3.3}$$

However, this modeling cannot be applied to the August release since it is the first early NASS yield forecast of the year. We therefore define the information by

$$X_{k,8,N} = \ln\left(\frac{Y_{k,8,N}}{\overline{Y_{k,N}}}\right) \times 100, \tag{3.4}$$

where $\overline{Y_{k,N}}$ represents the Olympic mean of the final yields from the last 5 years before N, computed as follows

$$\overline{Y_{k,N}} = \frac{1}{3}\left(\sum_{n=1}^{5} Y_{k,N-n} - \max_{n\in\{1,\dots 5\}}\{Y_{k,N-n}\} - \min_{n\in\{1,\dots 5\}}\{Y_{k,N-n}\}\right) \tag{3.5}$$

The 5-years Olympic mean is a major index, and is notably used in the Agricultural Risk Coverage Program (ARC) (Kim et al., 2015).

Finally, and similarly to Lehecka (2014), we perform a regression analysis to determine the possible market reaction to the news by

$$r_{0,i,N} = \beta_0 + \beta_1 X_{k,i,N}. \tag{3.6}$$

With this equation, one can assess the significance of the linear correlation between the modeled USDA information contained in the reports and the future returns. However, the market reaction might not necessary be linear with the governmental news. Thus, we also estimate the Kendall rank correlation coefficient between $r_{0,i,N}$ and $X_{k,i,N}$ to release the linear relation hypothesis.

3.3 Regression analysis - NDVI forecats

In this section, we explore the possibility of forecasting the NASS reports information through the MODIS NDVI data. These images are highly correlated to the vegetation conditions, and therefore to corn conditions. The World Agricultural Outlook Board (WAOB) and the Foreign Agricultural Service (FAS) teams actually already use satellite imagery and weather analysis to monitor crop conditions in order to prepare the WASDE[3]. Hence, it is logical to use similar data when aiming to predict its information.

First, we develop corn yield forecasting models based on MODIS NDVI time series for each of the 10 states considered in the study. We recall the period notation used in the MODIS NDVI time series in Table 1. We note $V_{k,P,N}$ the mean value of NDVI over the

[3]https://www.usda.gov/oce/commodity/wasde/prepared.htm

state k during the period P of the year N. Then, for each period between 9 and 17, we define two variables derived from the MODIS NDVI time series: the growing phase total NDVI value, defined as follows,

$$G_{k,P,N} = \sum_{p=9}^{P} V_{k,p,N}, \tag{3.7}$$

and the maximum NDVI peak value reached during the season,

$$M_{k,P,N} = \max_{p \leqslant P} V_{k,p,N}. \tag{3.8}$$

Then, we model the NDVI-based yield forecasts by the following linear regressions for each state

$$Y_{k,N} = \beta_{k,P,0} + \beta_{k,P,1}.N + \beta_{k,P,2}.M_{k,P,N} + \beta_{k,P,3}.\left(G_{k,P,N} - M_{k,P,N}\right) + \epsilon_{k,P,N}, \tag{3.9}$$

where $Y_{k,N}$ is the final yield estimates of year N for the State k and $\epsilon_{k,P,N}$ is an error term. We estimate the coefficients through the ordinary least square methods, leading to $\hat{\beta}_{k,P,l}$ for each state k, $l \in \{0, 1, 2, 3\}$ and $P \in \{13, 15, 17\}$. Thus, we obtain NDVI-based early yield estimates

$$\hat{Y}^{NDVI}_{k,P,N} = \hat{\beta}_{k,P,0} + \hat{\beta}_{k,P,1}.N + \hat{\beta}_{k,P,2}.M_{k,P,N} + \hat{\beta}_{k,P,3}.\left(G_{k,P,N} - M_{k,P,N}\right). \tag{3.10}$$

In this methodology, we note two major points. Firstly, the NASS early yield estimates are never used to train the NDVI-based yield forecasting models: we only rely on the governmental final yield. We choosed to do so to avoid overfitting issue that training on the NASS early yield estimates may have raised. Secondly, the three NDVI periods we train our models on, are ending on the July 27, August 28 and September 29. The data needed is therefore available before the WASDE releases of August, September and October respectively.

Hence, following the same underlying idea developped in Section 3.2, we model the new information provided in the WASDE reports to the market participants by

$$\hat{X}^{NDVI}_{k,8,N} = \left(\ln\left(\hat{Y}^{NDVI}_{k,13,N}\right) - \ln\left(\overline{Y_{k,N}}\right)\right) \times 100, \tag{3.11}$$

$$\hat{X}^{NDVI}_{k,9,N} = \left(\ln\left(\hat{Y}^{NDVI}_{k,15,N}\right) - \ln\left(\hat{Y}^{NDVI}_{k,13,N}\right)\right) \times 100, \tag{3.12}$$

$$\hat{X}^{NDVI}_{k,10,N} = \left(\ln\left(\hat{Y}^{NDVI}_{k,17,N}\right) - \ln\left(\hat{Y}^{NDVI}_{k,15,N}\right)\right) \times 100, \tag{3.13}$$

respectively for the August, September and October releases.

Finally, we perform a regression analysis to determine if this modelled information is correlated enough with the information contained in the governmental reports. In other words, we test wether the NDVI-based information forecasts could have been statistically considered as information regarding the corn future price changes at the WASDE release dates. Similarly to Section 3.2, we test this hypothesis thanks to the following regression,

$$r_{0,i,N} = \beta_0 + \beta_1 \hat{X}^{NDVI}_{k,i,N}. \tag{3.14}$$

4 Results

4.1 Market reaction

In this section, we present the results obtained about the impact of the report releases on the commodities market. First, we display in Figures 1, 2 and 3 the mean absolute value of the financial returns according to session t relative to the report publication. For the months of August (Figure 1) and October (Figure 3), we note a higher value of the mean absolute return of the futures for the session just following the WASDE reports. However, for the September reports (Figure 2), results are less convincing: even though the reports session has one of the highest value, the difference is not clear when comparing to other sessions. Intuitively, we would conclude that the August and October reports contain new and reliable information for the commodity market, which therefore reacts, while the September ones have less impacts.

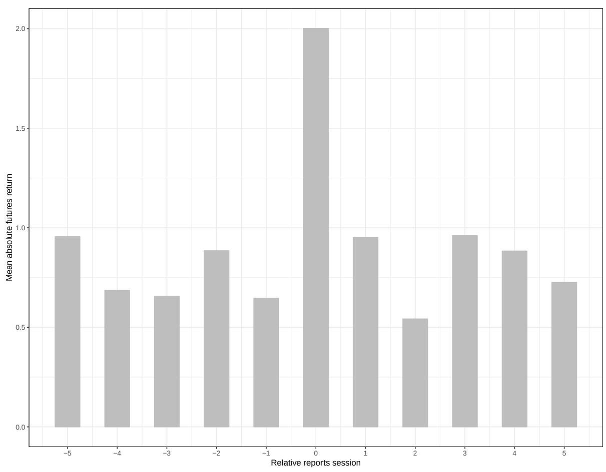

Figure 1: Mean absolute future returns relative to the August reports session during the period 2000-2016

To validate this graphical intuition, we perform a F-test, whose results are presented in

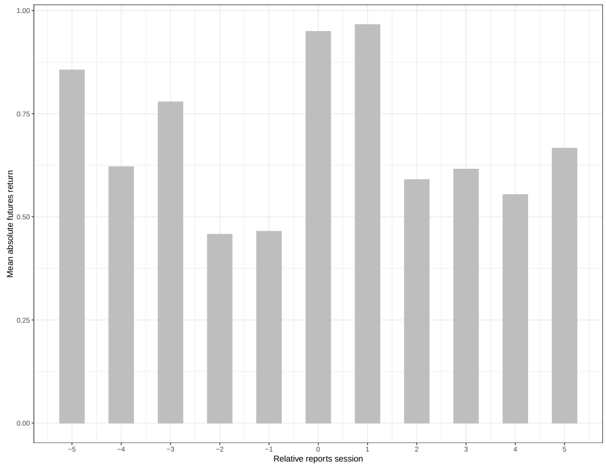

Figure 2: Mean absolute future returns relative to the September reports session during the period 2000-2016

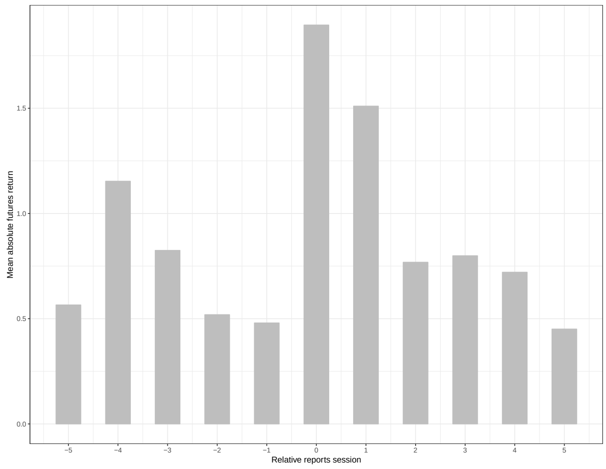

Figure 3: Mean absolute future returns relative to the October reports session during the period 2000-2016

Table 3. Return variance for all reports session considered is 4.2 times more important than pre and post reports return variance, and this difference is significant at a 0.001% level. However, we note that all the reports don't seem to have the same impact on the commodity market. Thus, the ratio of variances is around 5 for the releases of August and October, and only of 2.4 for September reports. This result is highlighted by the p-values, which indicate levels of confidence lower than 0.001% for both August and October, while the September one is around 1%.

We now test more specifically wether the information contained in the NASS early yield reestimation have an impact on the corn futures market. Results of the Equation 3.6 are displayed in Table 4, together with the Kendall rank correlation coefficient. First, we note that, for all the statistics, the estimates are negative with levels of confidence at least of 2%. This suggests that not only the future prices react to NASS early yield changes, but also that the reaction is rational. The significant negative sign of the estimated coefficient is consistent with the economic theory: a reduction (res. augmentation) of the crop yield expectations leads to an increase (res. decrease) of the

Table 3: Future return volatility test results for WASDE reports in Corn Markets, August-October, 2000-2016.

Reports	Reports Variance	Pre/Post Reports Variance	F-statistic	p-value
August	7.554	1.467	5.152	$< 1.10^{-5}$
September	2.073	0.879	2.362	0.0068
October	7.822	1.737	4.503	$< 1.10^{-5}$
All Reports	5.727	1.360	4.211	$< 1.10^{-5}$

commodity prices. Nevertheless, we remark that the significance levels differ depending of the publication month. As anticipated in the variances test, the reaction is less marked for September releases. Indeed, the p-values of Pearson's r and Kendall's τ are respectively only of 1.8% and 1.1%, while for other months the tests lead to a minimum of 0.03% level of confidence.

Table 4: Regression results of future prices impacts to early corn yield estimate changes, and their Kendall τ, August to October, 2000-2016

	$\hat{\beta}_0$		$\hat{\beta}_1$		Kendall τ	
Reports	Estimates	p-value	Estimates	p-value	Estimates	p-value
August	0.835	5.10^{-5}	-7.784	2.10^{-5}	-0.21	5.10^{-5}
September	-0.021	0.843	-8.001	0.018	-0.14	0.011
October	-0.273	0.196	-24.3	3.10^{-4}	-0.20	3.10^{-4}
All reports	0.162	0.122	-7.869	$< 1.10^{-5}$	-0.15	$< 1.10^{-5}$

The results obtained in this market reaction study are consistent with the existing literature, like for example Isengildina-Massa et al. (2008) who also detect a lower, while still statistically significant, impact of the September reports on the market. Thus, we can conclude on the fact that, during the period 2000-2016, the market participants considered that the WASDE reports of August, September and October, contain new and reliable information, and therefore react to it. Furthermore, we show evidences that, in particular, the NASS yield estimation changes provide valuable information to the commodity market, which seems to rationnaly react to it. This last result is line with the findings of Abbott et al. (2016) who estimate the corn yield informational value of WASDE reports to \$188 million.

4.2 Forecasting valuable information

In this section, we display the results from the NDVI-based analysis. First, we present in Table 5 estimates of the early yield forecasting models for each state based on the MODIS NDVI time series described in Equation 3.9. We note that the end of September models (period 17) achieve high adjusted-R^2 values, between 0.82 and 0.91, for almost every state. On the contrary, Kansas corn forecast model fail to fulfill such level of accuracy, even though it leads to a rather correct R^2 of 0.61. Similarly, the adjusted-R^2 increases throughout the growing season for all the states except Kansas and, to a lesser extent, Wisconsin. Thus, for example in Minnesota, at end of July the models lead to an accuracy of 0.54, which augments up to 0.77 around end of August and finally achieves an adjusted-R^2 of 0.85 at end of September. The special case of Kansas must be due to the fact that, contrary to other states where the two main crop types are corn and soybeans, Kansas is also a large winter wheat producer (cf. Table 6 for example). Hence, MODIS NDVI time series from wheat interfere with the corn ones when estimating the NDVI at a state level.

The NDVI-based early yield forecasting model results are in line with the WASDE's aim at providing accurate information about future yields, and perfecting it along the growing season. Thus, it is a logical candidate to model the information given by the WASDE reports. We therefore plot the modeled NDVI-based information (estimated with the Equations 3.11 to 3.13) to the NASS-based one (estimated with the Equation 3.3 and 3.4) in Figure 4. We graphically observe a postive correlation, which is statistically validated. The Pearson's r and Kendall's τ correlation coefficients are equal respectively to 0.68 and 0.35, both with a significance level of 0.1%.

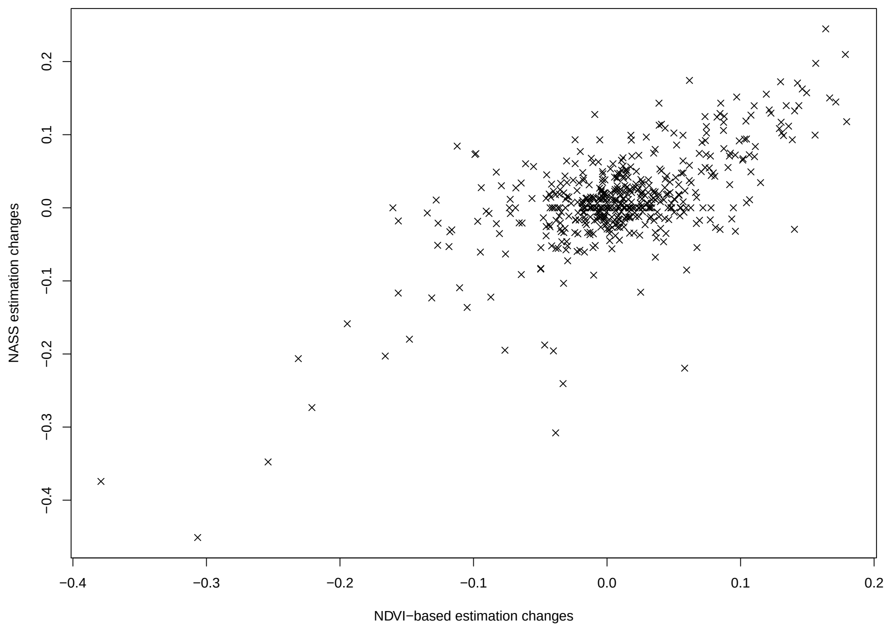

Figure 4: NASS yield estimation changes to NDVI-based yield estimation changes, August to October, 2000-2016

Although the NDVI-based information is significantly correlated to the early NASS yield estimation changes, the main goal of our study is to determine wether this correlation is strong enough for our NDVI-based yield estimation changes to be significantly correlated with the commodities market reactions. To assess this point we present the results of the Equation 3.14, together with the Kendall's τ, in Table 7. We

Table 5: Regression results of final corn yield estimate to NDVI time series, 2000-2016

States	Period	Dependant variable estimates: Intercept	Year	Maximum	$(G-M)$	Adjusted-R^2
Illinois	13	-3809^{*}	1.8^{*}	681^{***}	-81^{*}	0.60
	15	-2863^{**}	1.2^{*}	575^{***}	25	0.84
	17	$-2042.$	0.78	415^{**}	59^{*}	0.86
Indiana	13	-1784	0.84	458^{***}	-49^{*}	0.61
	15	-3219^{**}	1.4^{**}	623^{***}	33^{*}	0.85
	17	-2656^{**}	1.1^{*}	472^{***}	56^{*}	0.87
Iowa	13	-3085^{*}	1.5^{*}	407^{*}	-46	0.46
	15	-2903^{**}	1.2^{*}	693^{***}	1.9	0.77
	17	-2352^{**}	0.97^{*}	487^{**}	32.	0.82
Kansas	13	-3777	0.15	292^{*}	20	0.60
	15	303^{**}	-0.19	57	58^{*}	0.64
	17	964	-0.52	90	39^{*}	0.61
Michigan	13	-4284^{***}	2.2^{***}	224^{**}	-51	0.81
	15	-3478^{***}	1.6^{**}	407^{***}	40^{*}	0.87
	17	-3033^{**}	1.3^{**}	397^{**}	45^{*}	0.88
Minnesota	13	-3186^{*}	1.5^{*}	265	17	0.54
	15	$-1799.$	0.6	613^{**}	54^{**}	0.77
	17	$-1576.$	0.52	477^{**}	63^{***}	0.85
Nebraska	13	-3483^{**}	1.7^{**}	175^{*}	28	0.71
	15	-3103^{**}	1.5^{**}	108	39.	0.77
	17	-3072^{***}	1.5^{***}	0.7	47^{**}	0.89
Ohio	13	-3055	1.5	201.	64	0.53
	15	-1157	0.34	569^{***}	48^{**}	0.91
	17	$-1620.$	0.57	293.	79^{**}	0.91
South Dakota	13	-4122^{*}	2.0^{*}	224^{*}	5.8	0.64
	15	-3701^{**}	1.8^{*}	233^{*}	21	0.73
	17	-3517^{**}	1.7^{**}	98	49^{**}	0.84
Wisconsin	13	-3124^{***}	1.5^{**}	207^{**}	36	0.80
	15	-3001^{**}	1.4^{**}	310.	47^{*}	0.79
	17	-3167^{***}	1.4^{***}	215	50^{**}	0.83

Note: The levels of significance are noted as: $***$ for 0.1%, $**$ for 1%, $*$ for 5% and . for 10%.

Table 6: Area planted for corn, soybeans and winter wheat in Iowa and Kansas in 2016 (thousands of acres).

State	Crop	Planted area
Iowa	Corn	13,900
Iowa	Soybeans	9,500
Iowa	Wheat	25
Kansas	Corn	5,100
Kansas	Soybeans	4,050
Kansas	Wheat	8,500

remark that, when considering the three reports, the Pearson's correlation is significant at a 0.2% level. Furthermore, the $\hat{\beta}_1$ estimate is negative (-5), and therefore in line with the expectations. However, when focusing on specific month, we note that for the September one, no significant estimation emerge, the estimate even appear to be positive. This poor performance on the September reports is not completely a surprise. Indeed, these releases lead to the less important, but still statistically significant, market reactions quantified through the variance or regression analysis (cf. Table 3 and 4 and Figure 2). Kendall's τ study also indicates that the correlation is negative and significant for the August and October reports, while, for the September release, no significance arises. However, and on the contrary to the linear model, the rank correlation is not statistically significant when considering all reports together.

Table 7: Regression results of future prices impacts to early NDVI-based corn yield estimate changes, and their Kendall τ, August to October, 2000-2016

	$\hat{\beta}_0$		$\hat{\beta}_1$		Kendall τ	
Reports	Estimates	p-value	Estimates	p-value	Estimates	p-value
August	0.841	9.10^{-5}	-6.828	0.002	-0.12	0.023
September	-0.003	0.815	1.352	0.550	0.049	0.354
October	-0.400	0.054	-27.25	1.10^{-4}	-0.17	0.002
All reports	0.127	0.233	-5.001	0.002	-0.012	0.689

5 Perspectives

The results we present in the current paper are promising, however we believe that the methodology needs to be enhanced before a concrete and effective application on the commoditites market. Our statistical study aims at highlighting evidences that there is a possibility to forecast some of the valuable information contained in the USDA reports through the use of satellite data. These first findings should incite the private sector to develop similar methodologies for their risk management. We therefore propose some improvements that should be further investigate for a practical implementation.

First, we have only used NDVI to obtain our early yield estimates. To improve the accuracy of our model, a relative easy upgrade would be to also consider weather data (Mkhabela et al., 2011), or other vegetation index such as Enhanced Vegetation Index (Johnson et al., 2016). The use of a larger period to train the models could also be useful. In the current paper, we focus on the MODIS data, i.e. from 2000, while older data sources exist such as the AVHRR. However, the instruments and resolution being different, data first needs to be reprocessed (Pedelty et al., 2007). This issue will surely raise again in the future with the continuous enhancement of the satellite resolution (Skakun et al., 2018). More sophisticated data upgrades can be applied. Indeed, even though we apply a crop mask (MOD 12) to the MODIS NDVI data, we still conduct our study at a state level. Yet, many different crop types can be grown in a single state. It, therefore, leads to interferences between crops in the resulting NDVI time series. For example, Figure 5 displays the MODIS NDVI time series over the year 2016 in Kansas. We observe two peaks: one around the period 9 (mid May) and one around the period 15 (mid August). This results from the fact that Kansas is an important winter wheat producer (first peak), but also of corn and soybeans (second peak), as shown in Table 6. Recent studies highlight the possibility to elaborate a crop mapping thanks to remote sensing methods (Skakun et al., 2017b; Skakun et al., 2017a). Applying such a map should result to a better yield forecasting accuracy (Maselli and Rembold, 2001; Kastens et al., 2005; Zhang et al., 2019). Furthermore, if the map can be estimated early enough, it could also be used to forecast another valuable information from the USDA reports: the planted area. Indeed, the value to the market parcipants of such information is estimated to $145 million (Abbott et al., 2016).

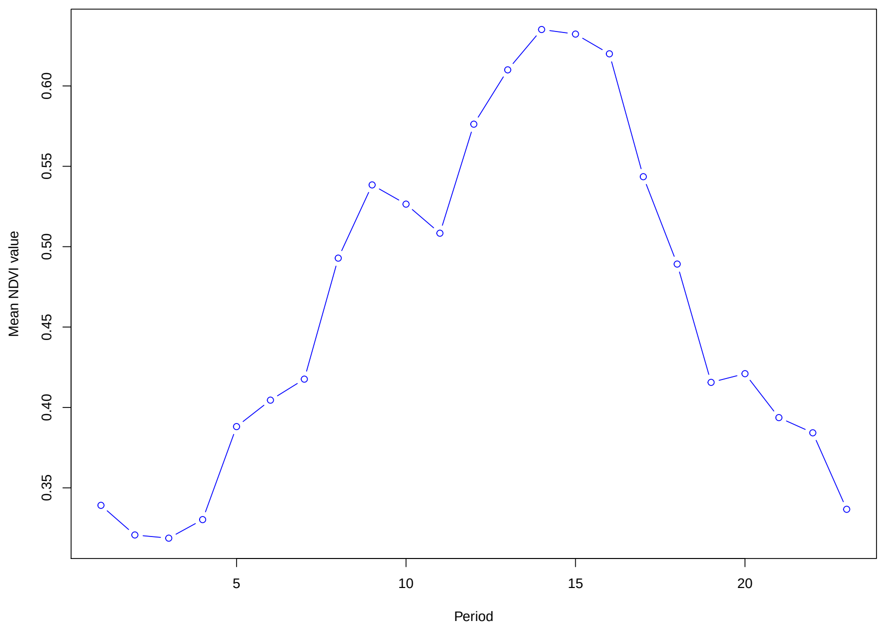

Figure 5: MODIS NDVI time series of Kansas over the year 2016.

The multiple crops problematic raises another limit of our paper: we focus on the corn future market only. The WASDE reports also contain valuable information about other agricultural commodities such as soybeans, wheat, barley, rice, etc. Moreover, as the

"W" suggests, the WASDE reports provide overviews of the supply and demand for these crops all around the world. For example, a careful attention is given to European (European Union, Ukraine and Russia) wheat market or Brazilian soybeans market. The agricultural commodities we focus on being storable, the different financial markets are linked one to another. Hence, to improve the forecasting performance of the market reactions to the WASDE releases, one would need to develop a more global model, in term of both crop type and localisation. Since the MODIS data cover the entire world, no major data problems should be expected.

Thirdly, in this research we focus on the public information contained in the USDA. More specifically, we evaluate the market reaction to the reestimation of yields as if no other information sources, public or private, were considered as newsworthy by the market participants. However, it is known that other governmental reports, published between two WASDE ones, contain valuable information on the crop conditions, notably the Crop Progress reports (Lehecka, 2014). In addition, many market participants have access to private forecasts on the future agricultural supply and demand (Karali et al., 2019). Therefore, it would be beneficial to take into account other sources of information in the modelling of the commoditites market reaction to the WASDE release.

Last but not least, we highlight significant and rational correlations in our results. For a practical implementation in the commodities market, the general idea of the current paper has to be converted into a trading strategy. Moreover, the performance should be assessed through a backtesting methodology before being put into practice. This method is an out-sample one, contrary to the study we conduct here. Thus, overfitting mentioned in Section 3.3 is no longer an issue. Hence, to increase the accuracy of the governmental valuable information forecasts, one should train the NDVI-based early yield model (Equation 3.9) not with the NASS final yield estimates $Y_{k,N}$, but rather with the corresponding NASS early yield estimates $Y_{k,i,N}$.

6 Conclusions

In this paper, we retrieve, over the period 2000-2016, the well-known result that corn futures market reacts to the WASDE report releases. Then, we evaluate the informational value contained in the early yield estimations from the NASS, published together with the WASDE. We find out that the changes in these estimations are significantly and rationally correlated to the financial returns of the corn futures. Then, we propose an econometric model to obtain early yield forecasts based on MODIS NDVI time series, available around two weeks before the actual publication of the WASDE reports. Finally, we show that changes between two of the NDVI-based early yield estimates are also significantly and rationally correlated to the corn future returns. Therefore, it seems possible to forecast some valuable information from the USDA announcements thanks to MODIS NDVI data.

Even though we do not pursue our work on this subject, our study can inspire commodity traders or agricultural insurance companies to optimize their financial and risk management strategies. Indeed, the value of the information is important for the market participants. Moreover, since the databases are provided in open access, the data acquisition cost can be considered as null. Hence, the incentives seem high enough for the private sector to invest in developping remote sensing tools for the commodities price risk management. Indeed, the approach presented in the current paper is only to be seen as a proof of concept rather than a directly useable methodology. In this perspective, we have proposed some improvements that one would need to focus on before putting the method into practice.

Nevertheless, giving that the data is provided by the NASA, i.e. a reliable governmental institution, the derived information can be considered as public. Hence, in the long term, if all market participants have developped such methodology, it may decrease the impact of the USDA reports.

References

Abbott, P., Boussios, D., and Lowenberg DeBoer, J. (2016). Valuing Public Information in Agricultural Commodity Markets: WASDE Corn Reports. *NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management.* St. Louis, MO.

Becker-Reshef, I., Vermote, E., Lindeman, M., and Justice, C. (2010a). A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sensing of Environment* 114.6, pp. 1312–1323.

Becker-Reshef, I., Justice, C., Sullivan, M., Vermote, E., and Tucker, C. (2010b). The Global Agriculture Monitoring (GLAM) project. *Remote Sensing* 2.6, pp. 1589–1609.

Colwell, R. (1956). Determining the prevalence of certain cereal crop diseases by means of aerial photography. *Hilgardia* 26.5, pp. 223–286. DOI: 10.3733/hilg.v26n05p223.

Doraiswamy, P. C., Sinclair, T. R., Hollinger, S., Akhmedov, B., Stern, A., and Prueger, J. (2005). Application of MODIS derived parameters for regional crop yield assessment. *Remote Sensing of Environment* 97.2, pp. 192–202. DOI: 10.1016/j.rse.2005.03.015.

Dorfman, J. H. and Karali, B. (2015). A Nonparametric Search for Information Effects from USDA Reports. *Journal of Agricultural and Resource Economics* 40.1, pp. 124–143.

Ferencz, C. et al. (2004). Crop yield estimation by satellite remote sensing. *International Journal of Remote Sensing* 25.20, pp. 4113–4149. DOI: 10.1080/01431160410001698870.

Fortenbery, T. and Sumner, D. A. (1993). The effects of USDA reports in futures and options markets. *Journal of Futures Markets* 13, pp. 157–173.

Gao, F., Anderson, M. C., Zhang, X., Yang, Z., Alfieri, J. G., Kustas, W. P., Mueller, R., Johnson, D. M., and Prueger, J. H. (2017). Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. *Remote Sensing of Environment* 188, pp. 9–25. DOI: 10.1016/j.rse.2016.11.004.

Garcia, P., Irwin, S. H., Leuthold, R. M., and Yang, L. (1997). The value of public information in commodity futures markets. *Journal of Economic Behavior & Organization* 32.4, pp. 559–570. DOI: 10.1016/S0167-2681(97)00013-9.

Gouel, C. (2018). The Value of Public Information in Storable Commodity Markets: Application to the Soybean Market. *NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*. Minneapolis, Minnesota.

Irwin, S. H., Good, D. L., and Gomez, J. K. (2001). The Value of USDA Outlook Information: An Investigation using event Study Analysis. *NCR Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*. St. Louis MO.

Isengildina-Massa, O., Irwin, S. H., Good, D. L., and Gomez, J. K. (2008). The Impact of Situation and Outlook Information in Corn and Soybean Futures Markets: Evidence from WASDE Reports. *Journal of Agricultural and Applied Economics* 40.1, pp. 89–103. DOI: 10.1017/S1074070800027991.

Johnson, M. D., Hsieh, W. W., Cannon, A. J., Davidson, A., and Bédard, F. (2016). Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agricultural and Forest Meteorology* 218-219, pp. 74–84. DOI: 10.1016/j.agrformet.2015.11.003.

Karali, B., Isengildina-Massa, O., Irwin, S. H., Adjemian, M. K., and Johansson, R. (2019). Are USDA reports still news to changing crop markets? *Food Policy*. DOI: 10.1016/j.foodpol.2019.02.005.

Kastens, J. H., Kastens, T. L., Kastens, D. L. A., Price, K. P., Martinko, E. A., and Lee, R.-Y. (2005). Image masking for crop yield forecasting using AVHRR NDVI time series imagery. *Remote Sensing of Environment* 99.3, pp. 341–356. DOI: 10.1016/j.rse.2005.09.010.

Kim, S., Zulauf, C., Roberts, M., and Cook, K. (2015). Performance of 5-Year Olympic Moving Average in Forecasting U.S. Crop Year Revenue for Program Crops. *NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*. St. Louis MO.

Kumar, R. and Silva, L. (1973). Light Ray Tracing Through a Leaf Cross Section. *Appl. Opt., AO* 12.12, pp. 2950–2954. DOI: 10.1364/AO.12.002950.

Lehecka, G. V. (2014). The Value of USDA Crop Progress and Condition Information: Reactions of Corn and Soybean Futures Markets. *Journal of Agricultural and Resource Economics* 39.1, pp. 88–105.

Li, A., Liang, S., Wang, A., and Qin, J. (2007). Estimating Crop Yield from Multi-temporal Satellite Data Using Multivariate Regression and Neural Network Techniques. *Photogrammetric Engineering & Remote Sensing* 73.10, pp. 1149–1157. DOI: 10.14358/PERS.73.10.1149.

Ma, B. L., Dwyer, L. M., Costa, C., Cober, E. R., and Morrison, M. J. (2001). Early Prediction of Soybean Yield from Canopy Reflectance Measurements. *Agronomy Journal* 93.6, pp. 1227–1234. DOI: 10.2134/agronj2001.1227.

Maselli, F. and Rembold, F. (2001). Analysis of GAC NDVI Data for Cropland Identification and Yield Forecasting in Mediterranean African Countries. *Photogrammetric Engineering & Remote Sensing* 67.5, pp. 593–602.

McKenzie, A. M. (2008). Pre-Harvest Price Expectations for Corn: The Information Content of USDA Reports and New Crop Futures. *Am J Agric Econ* 90.2, pp. 351–366. DOI: 10.1111/j.1467-8276.2007.01117.x.

Milacek, T. T. and Brorsen, B. W. (2017). Trading Based on Knowing the WASDE Report in Advance. *Journal of Agricultural and Applied Economics* 49.3, pp. 400–415. DOI: 10.1017/aae.2017.8.

Mkhabela, M. S., Bullock, P., Raj, S., Wang, S., and Yang, Y. (2011). Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. *Agricultural and Forest Meteorology* 151.3, pp. 385–393. DOI: 10.1016/j.agrformet.2010.11.012.

Pedelty, J. et al. (2007). Generating a long-term land data record from the AVHRR and MODIS Instruments. *2007 IEEE International Geoscience and Remote Sensing Symposium*. Barcelona, Spain: IEEE, pp. 1021–1025. DOI: 10.1109/IGARSS.2007.4422974.

Pervez, M. S. and Brown, J. F. (2010). Mapping Irrigated Lands at 250-m Scale by Merging MODIS Data and National Agricultural Statistics. *Remote Sensing* 2.10, pp. 2388–2412. DOI: 10.3390/rs2102388.

Peterson, D., Whistler, J., Egbert, S., and Brown, J. C. (2011). Mapping Irrigated Lands by Crop Type in Kansas. *Pecora 18-Forty Years of Earth Observation... Understanding a Changing World*, pp. 14–17.

Prasad, A. K., Chai, L., Singh, R. P., and Kafatos, M. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation* 8.1, pp. 26–33. DOI: 10.1016/j.jag.2005.06.002.

Rasmussen, M. S. (1997). Operational yield forecast using AVHRR NDVI data: Reduction of environmental and inter-annual variability. *International Journal of Remote Sensing* 18.5, pp. 1059–1077. DOI: 10.1080/014311697218575.

Ren, J., Chen, Z., Zhou, Q., and Tang, H. (2008). Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. *International Journal of Applied Earth Observation and Geoinformation*. Modern Methods in Crop Yield Forecasting and Crop Area Estimation 10.4, pp. 403–413. DOI: 10.1016/j.jag.2007.11.003.

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. NASA.

Skakun, S., Vermote, E., Roger, J.-C., and Belen, F. (2017a). Combined use of Landsat-8 and Sentinel-2A images for winter crop mapping and winter wheat yield assessment at regional scale. *AIMS Geoscience* 3.2, pp. 163–186. DOI: 10.3934/geosci.2017.2.163.

Skakun, S., Franch, B., Vermote, E., Roger, J.-C., Becker-Reshef, I., Justice, C., and Kussul, N. (2017b). Early season large-area winter crop mapping using MODIS NDVI data, growing degree days information and a Gaussian mixture model. *Remote Sensing of Environment* 195, pp. 244–258. DOI: 10.1016/j.rse.2017.04.026.

Skakun, S., Justice, C. O., Vermote, E., and Roger, J.-C. (2018). Transitioning from MODIS to VIIRS: an analysis of inter-consistency of NDVI data sets for agricultural monitoring. *International Journal of Remote Sensing* 39.4, pp. 971–992. DOI: 10.1080/01431161.2017.1395970.

Summer, D. A. and Mueller, R. A. E. (1989). Are Harvest Forecasts News? USDA Announcements and Futures Market Reactions. *Am J Agric Econ* 71.1, pp. 1–8. DOI: 10.2307/1241769.

Svotwa, E., Masuka, A. J., Maasdorp, B., Murwira, A., and Shamudzarira, M. (2014). Use of multi-spectral radiometer and MODIS satellite data correlations in developing models for tobacco yield forecasting. *International Journal of Agricultural Research* 3.1, pp. 147–154.

Tack, J., Coble, K., Johansson, R., Harri, A., and Barnett, B. (2017). The Potential Implications of 'Big Ag Data' for USDA Forecasts. SSRN Scholarly Paper ID 2909215. Rochester, NY: Social Science Research Network.

Wardlow, B. D. and Egbert, S. L. (2008). Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains. *Remote Sensing of Environment* 112.3, pp. 1096–1116. DOI: 10.1016/j.rse.2007.07.019.

— (2010). A comparison of MODIS 250-m EVI and NDVI data for crop mapping: a case study for southwest Kansas. *International Journal of Remote Sensing* 31.3, pp. 805–830. DOI: 10.1080/01431160902897858.

Whitcraft, A. K., Vermote, E., Becker-Reshef, I., and Justice, C. (2015). Cloud cover throughout the agricultural growing season: Impacts on passive optical earth observations. *Remote Sensing of Environment* 156, pp. 438–447. DOI: https://doi.org/10.1016/j.rse.2014.10.009.

Ying, J., Chen, Y., and Dorfman, J. H. (2017). Is the Value of USDA Annoucement Effects Declining over Time ? *NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*. St. Louis MO.

Zhang, H., Chen, H., and Zhou, G. (2012). The Model of Wheat Yield Forecast Based on Modis-Ndvi - A Case Study of Xinxiang. *ISPRES Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1-7.

Zhang, Y., Chipanshi, A., Daneshfar, B., Koiter, L., Champagne, C., Davidson, A., Reichert, G., and Bédard, F. (2019). Effect of using crop specific masks on earth observation based crop yield forecasting across Canada. *Remote Sensing Applications: Society and Environment* 13, pp. 121–137. DOI: 10.1016/j.rsase.2018.10.002.

Zhong, L., Hu, L., and Zhou, H. (2019). Deep learning based multi-temporal crop classification. *Remote Sensing of Environment* 221, pp. 430–443. DOI: 10.1016/j.rse.2018.11.032.