*Journal of Agricultural and Applied Economics* (2019), **51**, 402–416
doi:10.1017/aae.2019.5

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

# Predicting Soybean Yield with NDVI Using a Flexible Fourier Transform Model

Chang Xu and Ani L. Katchova*

Department of Agricultural, Environmental, and Development Economics, The Ohio State University, Columbus, Ohio, USA
*Corresponding author. Email: katchova.1@osu.edu

**Abstract**

We use models incorporating the normalized difference vegetation index (NDVI) derived from remote sensing satellites to improve soybean yield predictions in 10 major producing states in the United States. Unlike traditional methods that assume an ordinary least squares model applies to all observations, we allow for global flexibility in the relationship between NDVI and soybean yield by using the flexible Fourier transform (FFT) model. FFT results confirm that there is a nonlinear response of soybean yield to NDVI over the growing season. Out-of-sample predictions indicate that allowing for global flexibility with the FFT improves the predictions in time-series prediction and forecasting.

**Keywords:** Flexible Fourier transform model; forecasting; NDVI; remote sensing; soybean yield

**JEL Classifications:** C14; C53; Q16

## 1. Introduction

Many agencies, both public and private, exert significant efforts to make crop yield forecasts (Irwin, Sanders, and Good, 2014). Accurate and timely crop yield forecasts are valuable in many ways for market participants. At the aggregate level, crop yield forecasts help the price discovery process and improve market efficiency; they also aid decision makers in formulating rapid decisions to accommodate humanitarian actions and provide disaster assistance. At the individual level, crop yield forecasts are used to set crop insurance premiums by insurance companies, and they provide critical information for producers to make adjustments to improve their farm profitability.

In recent years, there has been an increasing interest in using remote sensing data to help improve crop yield forecasting. Remote sensing collects, archives, processes, and distributes satellite-derived data (Senay, 2016). For example, the normalized difference vegetation index (NDVI) contains helpful information generated by remote sensing procedures that can be used to predict crop yields. NDVI is a measure of biomass density on the surface of the earth, usually produced by a space platform. NDVI is defined as follows:

$$\mathrm{NDVI} = (\mathrm{NIR} - \mathrm{RED})/(\mathrm{NIR} + \mathrm{RED}),$$

where NIR stands for the reflectance of the near-infrared bands and RED stands for the reflectance of the visible bands of the electromagnetic spectrum. According to electromagnetic theory, live vegetation absorbs the blue and red bands of sunlight and reflects most of the green band of sunlight. Dying vegetation, to the contrary, absorbs mostly the green band of sunlight and reflects mostly the blue and red bands of sunlight. Barren soil reflects moderately both the visible and near-infrared bands of the electromagnetic spectrum. Generally, the higher the NDVI, the more

NIR light is reflected and the less RED light is reflected, and therefore, the target area includes more vegetation.

Because remote sensing provides information with a similar level of accuracy and accessibility regardless of the location and economic development of the country, using remote sensing data to predict crop yield has the potential to be applied in less developed countries in a cost-effective manner. In comparison, traditional, survey-based forecasts are relatively expensive and labor intensive.

Previous NDVI-based forecasting studies (Lv, 2014) utilized ordinary least squares (OLS) regression, which assumes that a global coefficient applies to each location invariantly. However, a global coefficient may hide location variation. Because of differences in local climate, soil conditions, and farm practices, the correlation between NDVI and crop yields may be highly localized. Using a global coefficient to forecast site-specific crop yield may be biased and thus may cause less informed decisions by market participants.

We use a flexible Fourier transform (FFT) model to allow for global flexibility in crop yield forecasts based on NDVI. This is the first study to our knowledge to examine how the correlation between NDVI and soybean yield varies by location and to use this global flexibility of the FFT model to improve the forecast performance of soybean yields. We then compare FFT with OLS in terms of out-of-sample forecast performance. Two hypotheses are tested: (1) the relationship between NDVI and crop yield is nonlinear using the FFT model; and (2) the proposed FFT model outperforms OLS in terms of ex ante forecasting accuracy, because FFT introduces flexibility in modeling the soybean yield–NDVI relationship and allows the soybean yield–NDVI elasticity estimates to vary across observations.

This article is organized as follows: Section 2 presents some background information on current practices used for crop yield forecasting and remote sensing for crop yield forecasting; Section 3 introduces the data sources and the FFT model we use; Section 4 presents a descriptive analysis and regression and forecasting results, comparing the FFT method and the traditional OLS method; and Section 5 concludes the article.

## 2. Background and related literature

### *2.1. Overview of current crop yield forecast methods*

There are two types of crop forecasts: survey-based forecasts and regression-based forecasts. Survey-based forecasts tend to be more accurate, especially when the harvest date is approaching, usually available shortly before or around harvest time, but they are also more expensive and labor intensive. Regression-based forecasts are more cost effective and can be available largely ahead of harvest; however, their accuracy may be compromised.

Survey-based forecasts that are used by the U.S. Department of Agriculture, National Agricultural Statistics Service (USDA-NASS) are made by conducting annually an agricultural yield survey (AYS) and an objective yield survey (OYS), the details of which can be found in USDA-NASS (2012). In the AYS, farmers are asked to self-report their anticipated yields, which may become the actual yields if harvest has begun. In the OYS, NASS sends technical personnel to the field to take objective measurements and counts of the plants. Both AYS and OYS are conducted monthly from May to November, but soybean yield data are collected and soybean yield forecasts are published from August to November. The final forecast is released in January of the next year. The typical cycle of soybean production in the major producing states in the United States is as follows: planting is in May and June, flowering is in July (which is its moisture/temperature-sensitive stage), filling is in August, maturation is in September, and harvesting is between October and November.

The second type of crop forecast is the regression-based forecast. This type of forecast is used mostly by private agencies and occasionally as a supplementary forecast by public agencies. For

https://doi.org/10.1017/aae.2019.5 Published online by Cambridge University Press

example, the World Agricultural Outlook Board (WAOB) releases the World Agricultural Supply and Demand Estimates regression-based forecasts, which use trend analysis and crop weather regression models. Unlike the forecasts released by NASS at the end of the year, the WAOB releases early forecasts throughout the growing season, from May to August (Irwin, Sanders, and Good, 2014). The comparison between NASS and WAOB yield forecasts and an evaluation of WAOB forecast accuracy can be found in Irwin, Good, and Sanders (2015). The crop weather model (also known as the modified Thompson model) utilizes a year trend variable, monthly weather variables, and an indicator if the crop is planted late. The crop condition model utilizes a year trend variable, the proportion of the crop planted after a certain date (e.g., May 30 for soybeans; Irwin, Good, and Tannura, 2009), and the proportion of the crop rated as good or excellent by USDA (Crop Progress Report). The model we propose in this study is based on the crop weather model but also adds NDVI variables. According to the literature, the modified Thompson model produces a good fit but performs poorly when events (such as insects and diseases) that cannot be captured by a weather variable negatively affect crop yields. We hypothesize that using NDVI can also monitor for insects and diseases because NDVI is a direct indicator of the greenness/health of the vegetation, with the additional benefit that NDVI data are immediately available at a low cost compared with the methods that rate crop conditions. Because regression-based forecasts typically rely on aggregate-level information, such as climatological variables at the county or regional level, a limitation of the regression-based forecasts is their inability to incorporate farm-level characteristics such as managerial skills or soil characteristics. However, regression-based forecasts can become useful when farm-level data are lacking, which is prevalent in many cases, especially in yield forecasting in developing countries.

### *2.2. Crop yield forecasting using remote sensing*

There have been numerous studies documenting the correlation between NDVI and crop yield forecasts, at the national (Maselli and Rembold, 2002), regional, county (Bolton and Friedl, 2013), and field level (Ferencz et al., 2004). Tucker (1979) determined that a time-integrated NDVI is largely correlated with crop yields when the vegetation is at the maximum level of greenness. Some studies focus on intra-annual variability showing how the correlation between the vegetation index and crop yields varies by the crop cycle and planting date (Basnyat et al., 2004). These studies suggest choosing NDVI data over a specific period for each type of crop in order to produce better forecasts. The weekly availability of NDVI data makes this crop-specific specification achievable. Lv (2014) suggests using earlier May NDVI and the change in NDVI over the crop planting and harvesting season for the most accurate yield forecasting. Johnson (2014) finds that crop yields are highly correlated with NDVI and daytime land surface temperature. The author conducts a regression of crop yields on NDVI for every week of the growing season and finds that the week in which the correlation is at its peak is at the beginning of August.

In addition to NDVI derived from the National Aeronautics and Space Administration's (NASA) Earth Observing System (EOS) Moderate Resolution Imaging Spectoradiometer, called eMODIS, other indexes and images have been used. For example, Doraiswamy and Cook (1995) is one of the earliest studies that used Advanced Very High Resolution Radiometer (AVHRR) imagery. AVHRR data are coarser, whereas eMODIS data are finer; AVHRR data are available for an extended period, whereas eMODIS data are only available after 2000. Later, Ferencz et al. (2004) also used AVHRR and a vegetation index called general yield unified reference index. Bolton and Friedl (2013) suggest to incorporate crop phenology and use a combination of the EVI2 (two-end enhanced vegetation index), NDVI, and normalized differenced water index (NDWI) for crop yield forecasting. They distinguish between semiarid and non-semiarid areas. They find that vegetation indexes are the best type of indexes for predicting in non-semiarid areas, whereas the NDWI is the best index for prediction in semiarid areas, because the water index is sensitive to irrigation in these semiarid areas.

https://doi.org/10.1017/aae.2019.5 Published online by Cambridge University Press

Instead of using traditional statistical models, Bose et al. (2016) utilize spiking neural networks from machine learning to analyze a remote sensing spatiotemporal relationship. Their work focuses on finding the optimum number of variables (or "features" in machine learning) to be included in regression analysis using machine learning techniques. They find that this type of prediction can be made 6 weeks before harvest with an average accuracy of 95.64%. They find that the year 2002 had the largest forecast error because of the 2002 drought. Adrian (2012) applies the Bayesian hierarchical model. This model is suitable for modeling data with clusters. It produces unique estimates for each state while requiring the estimates from each state to also follow a prior distribution. Johnson et al. (2016) focus on comparing forecast performance using linear versus nonlinear machine learning techniques and find that nonlinear models are not necessarily advantageous compared with linear models. Li et al. (2007) find that neural network techniques improve corn predictions compared with multivariate analysis. Kaul, Hill, and Walthall (2005) find that a nonlinear model only outperforms the linear model for barley. Mkhabela et al. (2011) categorize the Census Agricultural Regions (CARs) into three distinct agroclimatic zones; however, even within CARs, there might be multiple soil types. Bolton and Friedl (2013) emphasize the importance of delineating the boundary between farmland and nonfarmland, such as grassland and forests, because nonfarmland may contaminate the NDVI–crop yield relationship. Delineation can be done by using a land cover map such as Landsat Thematic Mapper (TM) data (Bolton and Friedl, 2013). Another method of delineation is to identify single pixels as agricultural or nonagricultural vegetation using statistical correction analysis (Maselli and Rembold, 2002). Among those studies, there are soybean forecasts in the United States using remote sensing (Lobell and Asner, 2003; Prasad et al., 2006). Chang et al. (2007) focus on using NDVI to map corn and soybean farmland.

Fieuzal, Sicre, and Baup (2017) make corn yield forecasts using both a real-time approach and a diagnostic approach. The real-time approach updates the estimates dynamically after the newest image is acquired, whereas the diagnostic approach utilizes all the image data throughout the season. The authors find the two best estimates perform comparably. Burke and Lobell (2017) regress the agreement between satellite-based yields and field-reported yields as a function of farm size and find the vegetation index can most accurately predict crop yield when the field size is large.

All of the abovementioned studies employ a global model to produce the regression results that fit all observations, with the major difference among the studies being the specific model they use. To the best of our knowledge, this study is the first one to employ models that produce site-specific regression results, allowing heterogeneous responses of soybean yields across counties. This is also the first study to our knowledge that applies the FFT model to examine the yield-NDVI relationship.

## 3. Data and methods

### *3.1. Data*

We use data for 797 counties from 10 major soybean-producing states in the United States from 2000 to 2016. According to NASS, the soybean production from these 10 states accounted for 78.5% (in 2016) and 79.8% (2000–2016 average) of the total soybean production in the United States (see Table 1 for soybean production and yield by state). Mkhabela et al. (2011) state that if a crop is not the dominant crop in the region, NDVI would give a poor prediction of crop yield because it cannot distinguish between different crops. The soybean yield data are obtained from the USDA-NASS QuickStats (https://quickstats.nass.usda.gov [accessed December 1, 2017]). This database provides official published aggregate statistics on U.S. soybean yields and the value of soybean production. Soybean yield is measured in bushels per acre. The NDVI data we use are from eMODIS onboard NASA's EOS Terra satellite. Landsat TM and eMODIS are two mainstream imagery sources. Though Landsat TM has a better spatial resolution (30 m) than eMODIS (250 m), the latter provides a better temporal resolution (daily) than the

https://doi.org/10.1017/aae.2019.5 Published online by Cambridge University Press

**Table 1.** Soybean production and yield in 10 major producing states

| State | Soybean Production[a] | Soybean Yield | Soybean Production | Soybean Yield | Soybean Production | Soybean Yield |
|---|---|---|---|---|---|---|
| | 2015 | | 2016 | | 2000–2016 Average | |
| Illinois | 544,320 | 56 | 592,950 | 59 | 461,082 | 48 |
| Iowa | 553,700 | 56.5 | 571,725 | 60.5 | 478,456 | 49 |
| Minnesota | 377,500 | 50 | 393,750 | 52.5 | 296,659 | 42 |
| Indiana | 275,000 | 50 | 324,300 | 57.5 | 260,298 | 48 |
| Nebraska | 305,660 | 58 | 314,150 | 61 | 237,676 | 49 |
| Missouri | 181,035 | 40.5 | 271,460 | 49 | 196,832 | 39 |
| Ohio | 237,000 | 50 | 263,780 | 54.5 | 206,655 | 45 |
| South Dakota | 235,520 | 46 | 255,915 | 49.5 | 161,933 | 37 |
| North Dakota | 185,900 | 32.5 | 249,000 | 41.5 | 125,860 | 32 |
| Arkansas | 155,330 | 49 | 145,700 | 47 | 119,596 | 39 |
| Ten states | 3,050,965 | 49 | 3,382,730 | 53 | 2,545,047 | 43 |
| U.S. total | 3,926,339 | 48 | 4,306,671 | 52.1 | 3,190,025 | 42 |

[a]Soybean production is measured in 1,000 bushels. Soybean yield is measured in bushels/acre.

former (16-day cycle). For monitoring purposes, we chose the eMODIS data. The eMODIS instrument onboard the Terra satellite achieves global coverage on a daily basis and provides 7-day composited data sets for its suite of products. Each data set provides NDVI information in GeoTIFF format that contains the reflective indices captured by Terra satellite at the resolution of 250 m from 2000 onward. Ag-Analytics converts the 250-m-resolution raw images to county-level NDVI. Ag-Analytics is an open-source, open-access database that provides data on agricultural finance, environmental finance, insurance, and risks (Woodard, 2016). We calculate county-level monthly NDVI values by taking a monthly average of the weekly NDVI values provided by Ag-Analytics. Climatological data are obtained from PRISM (parameter-elevation regressions on independent slopes model) Climate Data from Oregon State University and Ag-Analytics. We include two weather variables: maximum temperature over a month and average monthly precipitation. County boundary shapefiles are obtained from the U.S. Census Bureau. We obtain a sample of 12,027 county-year observations for the FFT analysis.

### *3.2. Flexible Fourier transform model*

When estimating crop yield response to input variables, traditional models use regional and temporal dummies to capture spatial and intertemporal heterogeneity. Adding dummy variables can only capture the difference in the value of the dependent variable across locations and time; it does not take into account how the relationship varies according to site-specific and time-specific characteristics. Another type of model uses a quadratic functional form to estimate the relationship between crop yield and weather variables, assuming that crop yield is nonlinearly related to the weather variable. However, these models may suffer from model misspecification, especially if there is a threshold effect, driven by environmental risks such as drought and flooding (Cooper, Nam Tran, and Wallander, 2017).

Gallant (1984) first proposed flexible Fourier functional transform to generate unbiased production function approximation and proved its mathematical validity. Cooper, Nam Tran, and Wallander (2017) applied an FFT function to estimate the relationship between crop yield

https://doi.org/10.1017/aae.2019.5 Published online by Cambridge University Press

and temperature. We follow the approach and modeling in Cooper, Nam Tran, and Wallander (2017) for the flexible Fourier function, which can be presented as follows:

$$\begin{aligned}\text{Soybean yield} = \beta_0 &+ \sum_{m=April}^{August} (\beta_{1m} MaxTemp_m + \beta_{2m} MaxTempSquare_m) \\ &+ \sum_{m=April}^{August} (\beta_{3m} Precipitation_m + \beta_{4m} PrecipitationSquare_m) \\ &+ \sum_{m=April}^{September} (\beta_{5m} NDVI_m) + \delta_0 TimeTrend + \sum_{s=1}^{9} \delta_s StateDummy_s \\ &+ 2\sum_{\alpha=1}^{A}\sum_{j=1}^{J}\left\{v_{j\alpha}\cos\left[jk_\alpha' s(NDVI)\right] - w_{j\alpha}\sin\left[jk_\alpha' s(NDVI)\right]\right\} + error \qquad (1)\end{aligned}$$

In this model, the dependent variable is soybean yield in a county for a given year. $\beta_0$ is the constant term. $MaxTemp_m$, $Precipitation_m$, and $NDVI_m$ are the maximum temperature, the average precipitation, and the average NDVI in month $m$, respectively. We include the weather variables from April to August, following the standard specification in the literature (Cooper, Nam Tran, and Wallander, 2017). We include NDVI variables through September, following the remote sensing literature (Li et al., 2007). The advantage of the FFT function is that it not only allows for model flexibility but also incorporates multivariate estimation, which is difficult to achieve through other nonparametric models such as kernel regression.

$PrecipitationSquare_m$ and $MaxTempSquare_m$ are the squared terms of $MaxTemp_m$ and $Precipitation_m$. *TimeTrend* equals the year minus 1999. $StateDummy_s$ is the state dummy variable. *NDVI* is a vector with each element being $NDVI_m$. *s*(*NDVI*) is the scaled version of *NDVI* such that each element of s(*NDVI*) is in the range of $[0, 2\pi]$. In our case, only NDVI variables are transformed.

The $\beta_0 + \sum_{m=April}^{August} (\beta_{1m} MaxTemp_m + \beta_{2m} MaxTempSquare_m) + \sum_{m=April}^{August} (\beta_{3m} Precipitation_m + \beta_{4m} PrecipitationSquare_m) + \sum_{m=April}^{September} (\beta_{5m} NDVI_m)$ terms represent the quadratic regression part. $\beta_{1m}$, $\beta_{2m}$, $\beta_{3m}$, $\beta_{4m}$, and $\beta_{5m}$ are parameters to be estimated. The $2\sum_{\alpha=1}^{A}\sum_{j=1}^{J}\{v_{j\alpha}\cos[jk_\alpha' s(NDVI)] - w_{j\alpha}\sin[jk_\alpha' s(NDVI)]\}$ term models the functional flexibility using FFT. Similar to the Taylor expansion, which uses a series of polynomial terms to approximate the true function, the Fourier function uses a series of trigonometric terms to approximate the true function. The Fourier functional form is believed to be the only known functional form that satisfies the Sobolev condition, meaning that the difference between the approximated function and the true function approaches zero as the sample size becomes arbitrarily large. For a proof that the Fourier function satisfies the Sobolev condition, refer to Gallant (1994). In the model, $k_\alpha$ ($\alpha = 1, 2, \ldots, A$) is the elementary multi-index vector, whose dimension equals the dimension of $x_{FFT}$, whereas $A$ is the total number of elementary multi-indexes. The vector $k_\alpha$ can be obtained in the following way: first, exhaust the list of $k_\alpha$, such that $k_\alpha$ has only integer elements and the sum of the absolute value of each element in $k_\alpha$ is no greater than $K$, where $K$ is predetermined; second, delete any $k_\alpha$ whose first nonzero element is negative; and third, delete any $k_\alpha$ whose elements have a common integer divisor. Monahan (1981) introduced a Fortran code to produce the set of elementary multi-index vectors. Also in the model, $J$ is the order of the Fourier transformation, whereas $v_{j\alpha}$ and $w_{j\alpha}$ are parameters to be estimated. We use the following parametrization: $K = 2$, $J = 2$, which are chosen such that the rule of thumb—the number of variables after transformation is roughly the square root of the number of observations (Fenton and Gallant, 1996)—is satisfied. Because there are 12,027 observations in the data we use, we include a total of 120 variables after the adding the transformed NDVI variables.

https://doi.org/10.1017/aae.2019.5 Published online by Cambridge University Press

The model degenerates to the traditional OLS model when $v_{j\alpha} = 0$ and $w_{j\alpha} = 0$. In the following discussion, the OLS model refers to equation (1), with $v_{j\alpha} = 0$ and $w_{j\alpha} = 0$ imposed. By testing the statistical significance of variable $v_{j\alpha}$ and $w_{j\alpha}$, we can decide whether the traditional quadratic model should be rejected in favor of the more flexible FFT model.

A review of the relevant literature reveals that the FFT model has been used/tested by scholars in different studies, fields, and situations. Chang et al. (2016) used the FFT to model the nonlinear effect of temperature on electricity demand. Becker, Enders, and Lee (2006) proposed a unit root test with a Fourier functional transform. Enders and Li (2015) approximated structural breaks in U.S. GDP trends using Fourier forms. Jones and Enders (2014) provided a summary on using Fourier forms to model structural breaks.

### 3.3. Prediction and forecast

We compare the prediction performance of the FFT model versus the OLS model. We conduct out-of-sample predictions and evaluate prediction performance by comparing prediction errors measured by the root-mean-square error (RMSE) and the mean absolute error (MAE), between FFT and OLS, for three schemes: time-series prediction, cross-sectional prediction, and panel prediction. RMSE and MAE are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{\text{N}}\sum_{i=1}^{\text{N}}(y_i - \hat{y}_i)^2}. \tag{2}$$

$$\text{MAE} = \frac{1}{\text{N}}\sum_{i=1}^{\text{N}}|y_i - \hat{y}_i|. \tag{3}$$

Both RMSE and MAE are commonly used measures to evaluate prediction performance. They measure the difference between true and fitted values for soybean yield. The unit for both RMSE and MAE is bushels per acre. In a time-series prediction, we first select a year for prediction, then we use observations from all other years to generate the model, and after that we predict the soybean yield for the selected year using the fitted model, weather data, and NDVI data from the selected year. In cross-sectional prediction, similarly, we select a state for prediction, then we use observations from all other states to generate the model, and after that we predict the soybean yield for the selected state using the fitted model, weather data, and NDVI data from the selected state; in panel prediction, similarly, we make the prediction for a selected year and state. Though commonly used, a shortcoming of using RMSE or MAE to measure prediction performance is that we do not know whether the predicted yield overestimates or underestimates the final actual yield.

We make predictions and forecasts using the regression results from the models. In this study, prediction refers to cases where we may use data afterward to predict for a specific time; forecast refers to cases where we only use data up to a certain year to make predictions for that year.

## 4. Results

### 4.1. Descriptive analysis

The descriptive statistics for the main variables are reported in Table 2. The average soybean yield across all states and years is 43.11 bushels per acre. From April to July, the average maximum temperature and average NDVI increase steadily and reach their peak levels in August. The average precipitation is highest in the months of May and June. These variables are included as suggested by the modified Thompson model (Thompson, 1963) to account for weather effects.

https://doi.org/10.1017/aae.2019.5 Published online by Cambridge University Press

**Table 2.** Descriptive statistics

| Variable | Number of Observations | Minimum | Median | Maximum | Mean | Standard Deviation |
|---|---|---|---|---|---|---|
| Soybean yield | 12,027 | 2.9 | 44 | 73.1 | 43.11 | 10.05 |
| Maximum temperature,[a] April | 12,027 | 0.37 | 17.32 | 27.34 | 17.06 | 3.44 |
| Maximum temperature, May | 12,027 | 13.65 | 22.33 | 30.67 | 22.39 | 2.56 |
| Maximum temperature, June | 12,027 | 19.7 | 27.37 | 36.06 | 27.37 | 2.28 |
| Maximum temperature, July | 12,027 | 22.82 | 29.41 | 38.91 | 29.58 | 2.4 |
| Maximum temperature, August | 12,027 | 20.13 | 28.76 | 39.47 | 28.92 | 2.35 |
| Precipitation, April | 12,027 | 4.17 | 85.14 | 424.08 | 91.26 | 48.98 |
| Precipitation, May | 12,027 | 5.27 | 108.28 | 355.26 | 113.08 | 52.23 |
| Precipitation, June | 12,027 | 7.64 | 105.01 | 376.5 | 115.8 | 58.07 |
| Precipitation, July | 12,027 | 0.89 | 87.5 | 354.27 | 94.43 | 49.66 |
| Precipitation, August | 12,027 | 0 | 82.04 | 438 | 90.01 | 51.54 |
| NDVI, April | 12,027 | −0.01[b] | 0.33 | 0.79 | 0.35 | 0.12 |
| NDVI, May | 12,027 | 0.13 | 0.42 | 0.85 | 0.45 | 0.13 |
| NDVI, June | 12,027 | 0.24 | 0.59 | 0.87 | 0.58 | 0.1 |
| NDVI, July | 12,027 | 0.27 | 0.74 | 0.89 | 0.73 | 0.08 |
| NDVI, August | 12,027 | 0.27 | 0.75 | 0.88 | 0.72 | 0.1 |
| NDVI, September | 12,027 | 0.24 | 0.6 | 0.87 | 0.6 | 0.1 |

[a]Temperatures are measured in degrees Celsius; precipitation is measured in inches.
[b]Negative normalized difference vegetation index (NDVI) denotes snow cover.

### *4.2. Flexible Fourier transform regression results*

All FFT models were developed using Matlab R2017a (The MathWorks Inc.), following the methodology in Cooper, Nam Tran, and Wallander (2017). Figures showing FFT results were made using the ArcMap 10.3 software. The estimation results from the model incorporating FFT terms are reported in Table 3. Because of the substantial number of variables (including 84 transformed NDVI variables), we only report the results for the main variables, including the untransformed weather variables and NDVI variables. However, the rest of the transformed variables are also included in the model-fitting process. We calculate elasticities by applying the mean value theorem to get the numerical approximation of the derivatives and fixing the values of independent variables at the median value for each variable for each county. Thus, we obtain an elasticity estimate for each county. We present the minimum, median, and maximum of FFT elasticity estimates across counties in columns 2 through 4 in Table 3. For comparison purposes, we also use the OLS regression results to calculate elasticity estimates for each county and report the elasticity summary from the OLS regression in columns 5 through 7 in Table 3. The OLS model refers to equation (1) with $v_{j\alpha} = 0$ and $w_{j\alpha} = 0$ imposed. For the weather variables, except for the July maximum temperature and the April average precipitation, the median of elasticity estimates derived from OLS and the median of elasticity estimates from FFT have the same sign. On average, higher temperatures from April to June and higher precipitation levels from June to August lead to higher soybean yields. On the other hand, higher temperatures in August and higher precipitation levels in May are associated with lower soybean yields.

Although the median of elasticity estimates for weather variables across counties is very similar between the FFT and OLS results, the median elasticity estimates of NDVI variables differ

https://doi.org/10.1017/aae.2019.5 Published online by Cambridge University Press

**Table 3.** Elasticity estimates from flexible Fourier transform (FFT) and quadratic ordinary least squares (OLS) models

| | FFT | | | Quadratic OLS | | |
|---|---|---|---|---|---|---|
| | Minimum | Median | Maximum | Minimum | Median | Maximum |
| Maximum temperature, April | 0.02 | 0.08 | 0.29 | 0.04 | 0.07 | 0.21 |
| Maximum temperature, May | −0.75 | 0.22*** | 3.74 | −0.82 | 0.27*** | 1.38 |
| Maximum temperature, June | −0.11 | 0.42*** | 5.99 | −0.29 | 0.39*** | 1.62 |
| Maximum temperature, July | −0.94 | −0.04*** | 1.14 | −1.1 | 0.04*** | 0.86 |
| Maximum temperature, August | −3.82 | −0.48** | −0.3 | −1.46 | −0.53 | −0.37 |
| Precipitation, April | −0.03 | 0.0013*** | 0.04 | −0.04 | −0.0047** | 0.0048 |
| Precipitation, May | −0.15 | −0.01* | 0.01 | −0.18 | −0.01* | 0.004 |
| Precipitation, June | −0.05 | 0.03*** | 0.3 | −0.07 | 0.04*** | 0.08 |
| Precipitation, July | −0.29 | 0.04*** | 0.35 | −0.44 | 0.05*** | 0.11 |
| Precipitation, August | 0.01 | 0.09*** | 0.53 | 0.03 | 0.09*** | 0.16 |
| NDVI, April | −3.27 | −0.03*** | 2.08 | −0.22 | −0.07*** | −0.04 |
| NDVI, May | −1.04 | −0.06* | 1.09 | −0.22 | −0.09*** | −0.04 |
| NDVI, June | −8.62 | −0.15*** | 1.14 | −0.01 | −0.01 | −0.0032 |
| NDVI, July | −2.27 | 0.45*** | 8.36 | 0.08 | 0.14*** | 0.26 |
| NDVI. August | −3.74 | 0.34 | 7.46 | 0.13 | 0.22*** | 0.32 |
| NDVI, September | −5.11 | 0.09 | 2.48 | −0.11 | −0.06*** | −0.04 |
| Number of observations | | 12,027 | | | 12,027 | |
| State fixed effects | | Yes | | | Yes | |
| Year trend effects | | Yes | | | Yes | |
| Adjusted $R^2$ | | 0.721 | | | 0.701 | |
| Rank test between Fourier and OLS | | | $F$ (84,11906) = 11.132 | | | |

Notes: Because of the nonlinearity of the FFT regression, we report the elasticity estimates rather than the coefficient estimates of the main variables. Significance here indicated by asterisks corresponds to the significance of the untransformed variables. Asterisks (*, **, and ***) denote significance level of 10%, 5%, and 1%, respectively. In addition to these variables, an additional 84 Fourier transformed variables of normalized difference vegetation index (NDVI) are included in the analysis—their coefficient estimates are not reported here, but they are included in the elasticity calculations.

significantly between the FFT and OLS results, in terms of both sign (September NDVI) and magnitude (April–August NDVI). NDVI elasticities estimated from the FFT model have a wider range than those generated by OLS, because of the inclusion of the transformed NDVI variables. The OLS results suggest that the August NDVI has a greater impact on soybean yields than the July NDVI, whereas the FFT results suggest the opposite. According to Table 3, when the July NDVI increases by 10%, the median soybean yield significantly increases by 4.5% or 1.94 bushels per acre. The median effect of August NDVI is also positive, though not significant.

By testing the significance of the coefficient estimates for the Fourier terms, we can test whether the FFT specification is overfitting the data. In Table 3, we present an $F$ test of the FFT regression versus the OLS regression; we find that the coefficients on the transformed Fourier terms are jointly significantly different from zero, and thus the OLS is rejected in favor of the FFT regression.

The geographic distribution of coefficient estimates from FFT is presented in Figure 1. In each panel, we present the geographic distribution of the median of the elasticity estimates of NDVI for

https://doi.org/10.1017/aae.2019.5 Published online by Cambridge University Press

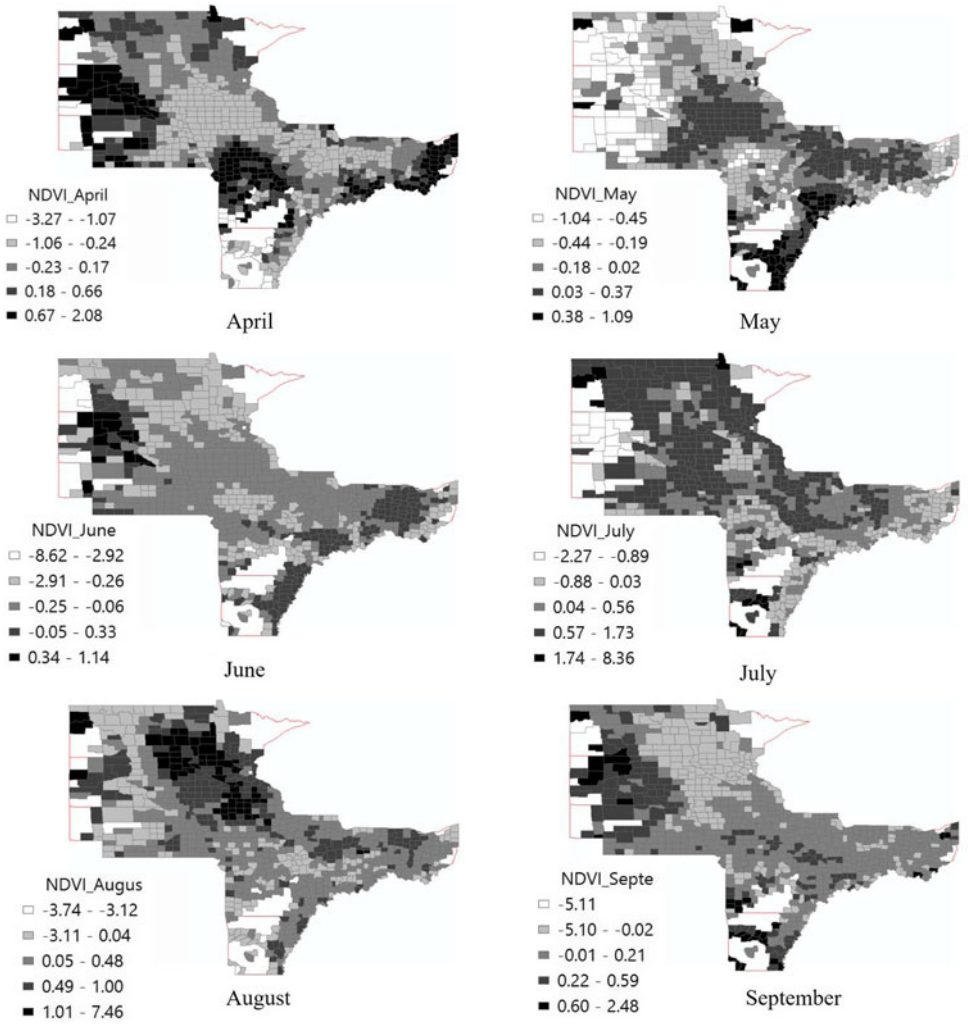


**Figure 1.** Geographic distribution by state of elasticity estimates from flexible Fourier transform, April–September. Note: NDVI, normalized difference vegetation index.

each month (April, May, June, July, August, and September, respectively) across different counties. For some counties in the north of North Dakota, central Minnesota, central Indiana, western Arkansas, and southwestern Missouri, soybean yields are highly responsive to July NDVI, but less responsive to August NDVI. For most counties in Ohio and in eastern Arkansas, in contrast, the soybean yield is responsive to August NDVI, whereas it is less responsive to July NDVI. For some counties in the western parts of North Dakota and South Dakota, soybean yields are responsive to April NDVI, whereas they are less responsive to August NDVI. These geographic differences in soybean yield responsiveness to NDVI show that global flexibility needs to be considered when making yield predictions.

### *4.3. Prediction and forecast results*

The results of the time-series prediction and cross-sectional prediction performance for FFT versus OLS are shown in Table 4. The bolded numbers show cases where the FFT error is lower than the OLS error. On average, FFT performs better than OLS in time-series predictions because both

https://doi.org/10.1017/aae.2019.5 Published online by Cambridge University Press

**Table 4.** Out-of-sample prediction performance: time-series and cross-sectional prediction

| | MAE | | RMSE | | | MAE | | RMSE | |
|---|---|---|---|---|---|---|---|---|---|
| Year | OLS | FFT | OLS | FFT | State | OLS | FFT | OLS | FFT |
| 2000 | 4.811 | 4.981 | 6.071 | 6.214 | North Dakota | **5.421** | **4.926** | **6.510** | **6.003** |
| 2001 | 3.879 | 3.936 | 4.921 | 5.022 | South Dakota | 5.358 | 6.273 | 6.883 | 8.729 |
| 2002 | **4.828** | **4.808** | **6.239** | **6.196** | Iowa | **4.801** | **4.078** | **5.851** | **5.118** |
| 2003 | **5.321** | **5.025** | **6.696** | **6.385** | Ohio | 4.318 | 4.717 | 5.335 | 5.771 |
| 2004 | 4.240 | 4.352 | 5.427 | 6.021 | Illinois | **6.200** | **5.578** | **7.582** | **6.947** |
| 2005 | **4.269** | **4.137** | **5.427** | **5.234** | Indiana | **4.617** | **4.468** | **5.546** | **5.457** |
| 2006 | **4.378** | **4.148** | **5.555** | **5.342** | Nebraska | 10.052 | 10.354 | 12.769 | 13.220 |
| 2007 | 4.620 | 4.741 | 6.203 | 6.338 | Minnesota | 4.979 | 4.981 | 6.343 | 6.532 |
| 2008 | 4.073 | 4.251 | 5.227 | 5.450 | Missouri | 4.685 | 4.723 | 5.919 | 5.925 |
| 2009 | **4.397** | **4.192** | **5.753** | **5.415** | Arkansas | **7.613** | **7.262** | **9.570** | **9.224** |
| 2010 | **3.857** | **3.757** | **4.925** | **4.875** | Average | **5.804** | **5.736** | 7.231 | 7.293 |
| 2011 | 4.374 | 4.531 | 5.573 | 5.671 | | | | | |
| 2012 | **5.929** | **5.739** | **7.467** | **7.328** | | | | | |
| 2013 | **4.945** | **4.894** | **6.189** | **6.174** | | | | | |
| 2014 | **4.257** | **4.238** | **5.402** | **5.372** | | | | | |
| 2015 | **4.492** | **4.464** | **5.839** | **5.798** | | | | | |
| 2016 | **5.358** | **5.336** | 6.584 | 6.602 | | | | | |
| Average | **4.590** | **4.561** | **5.853** | **5.849** | | | | | |

Notes: Bolded numbers indicate that flexible Fourier transform (FFT) has lower prediction errors and therefore outperforms ordinary least squares (OLS). MAE, mean absolute error; RMSE, root-mean-square error.

MAE and RMSE for FFT are lower than those for the OLS model. For cross-sectional predictions, FFT has a higher RMSE on average, but a lower MAE than OLS does.

Our results show that time-series predictions on average are more accurate than cross-sectional predictions in terms of smaller predicting error. RMSE and MAE from time-series predictions are consistently lower than cross-sectional predictions.

We also conduct out-of-sample panel predictions. We randomly select 1,000 observations from all years and states, and predict the soybean yields for these 1,000 observations by OLS and FFT, using all other observations excluding these 1,000 observations. We then compare the predicted soybean yields with the actual yields and calculate the RMSE and MAE. We then repeat this sampling process 200 times. The histogram shown in Figure 2 is of the distribution of RMSE and MAE. Two findings are interesting. First, panel prediction has much lower prediction error than both time-series and cross-sectional predictions in Table 4. This suggests that when predicting soybean yield for a certain location, it is useful to include the already publicized yield data from other locations into the training sample. Second, FFT has a consistently lower prediction error than the OLS model. FFT can improve the prediction performance by a modest 0.3% according to MAE, or 0.4% according to RMSE. This percentage is obtained by dividing the prediction error by the mean of crop yield (average MAE is 0.138, average RMSE is 0.1684, and mean soybean yield is 43.11).

The predictions so far may have used data from future periods to predict current soybean yields. Therefore, we now include forecasts where soybean yield predictions are only based on data from previous periods (Table 5). For RMSE, there are 10 years out of 16 years where

https://doi.org/10.1017/aae.2019.5 Published online by Cambridge University Press

**Table 5.** Out-of-sample forecast performance

| | MAE | | RMSE | |
|---|---|---|---|---|
| Year | OLS | FFT | OLS | FFT |
| 2001 | **11.136** | **10.636** | **12.882** | **12.678** |
| 2002 | 6.303 | 6.502 | 7.956 | 8.235 |
| 2003 | **4.696** | **4.327** | **6.188** | **5.639** |
| 2004 | **5.956** | **5.395** | 7.117 | 7.394 |
| 2005 | **7.201** | **6.979** | **8.448** | **8.239** |
| 2006 | **4.758** | **4.546** | **6.319** | **6.051** |
| 2007 | 5.297 | 5.559 | 6.865 | 7.157 |
| 2008 | 4.242 | 4.758 | 5.475 | 6.143 |
| 2009 | **4.240** | **4.074** | **5.495** | **5.203** |
| 2010 | **4.353** | **4.277** | **5.545** | **5.489** |
| 2011 | **5.132** | **4.911** | **6.765** | **6.410** |
| 2012 | **6.239** | **6.124** | **7.835** | **7.730** |
| 2013 | 4.845 | 4.994 | 6.080 | 6.247 |
| 2014 | **4.240** | **4.218** | **5.388** | **5.335** |
| 2015 | **4.488** | **4.383** | **5.854** | **5.748** |
| 2016 | **5.358** | **5.336** | 6.584 | 6.602 |
| Average | **5.530** | **5.439** | **6.925** | **6.894** |

Notes: Bolded numbers indicate that flexible Fourier transform (FFT) has lower forecast errors and therefore outperforms ordinary least squares (OLS). MAE, mean absolute error; RMSE, root-mean-square error.

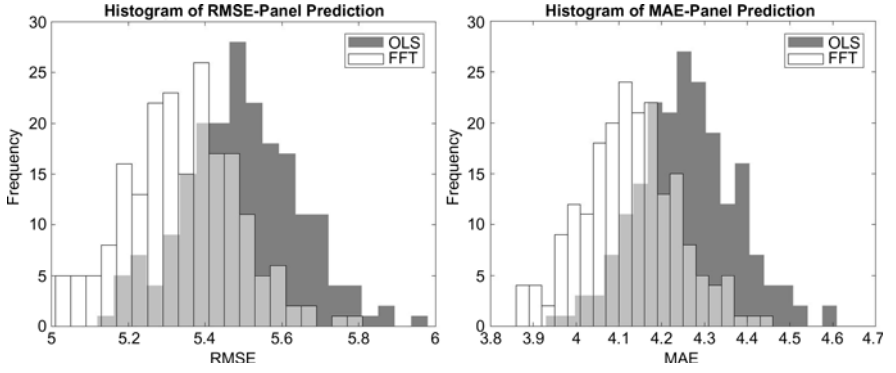


**Figure 2.** Histogram of root-mean-square error (RMSE) and mean absolute error (MAE) between ordinary least squares (OLS) and flexible Fourier transform (FFT).

FFT outperforms OLS. For MAE, there are 12 years out of 16 years in which FFT outperforms OLS. In terms of average error, FFT has smaller RMSE and MAE than OLS does. Although the forecasts are more realistic in terms of being based only on data from previous periods, the average prediction errors are unsurprisingly higher than those for the predictions using all data including from future periods in Table 4.

https://doi.org/10.1017/aae.2019.5 Published online by Cambridge University Press

**Table 6.** Out-of-sample prediction performance with county fixed effects: time-series and cross-sectional prediction

| | MAE | | RMSE | | | MAE | | RMSE | |
|---|---|---|---|---|---|---|---|---|---|
| Year | OLS | FFT | OLS | FFT | State | OLS | FFT | OLS | FFT |
| 2000 | 4.1246 | 4.1621 | 5.2858 | 5.2991 | North Dakota | 4.1488 | 5.2305 | 4.8985 | 6.0372 |
| 2001 | 3.2586 | 3.4405 | 4.1627 | 4.3735 | South Dakota | 4.9816 | 6.0233 | 5.9978 | 6.9592 |
| 2002 | 3.5647 | 3.6346 | 4.5595 | 4.7035 | Iowa | 7.5167 | 8.6121 | 8.2265 | 9.2306 |
| 2003 | **4.8895** | **4.8078** | **6.0952** | **6.0159** | Ohio | 5.6129 | 6.7518 | 6.6215 | 7.6794 |
| 2004 | **3.764** | **3.7491** | **5.0481** | **4.9968** | Illinois | 6.8525 | 7.5782 | 7.6378 | 8.3345 |
| 2005 | **3.8178** | **3.8177** | **4.7016** | **4.6617** | Indiana | 6.2513 | 7.2419 | 6.9124 | 7.8397 |
| 2006 | **3.2359** | **3.148** | **4.219** | **4.1832** | Nebraska | 14.0015 | 16.1125 | 14.7071 | 16.7112 |
| 2007 | 3.9388 | 4.1172 | 5.0193 | 5.2704 | Minnesota | 6.3847 | 7.5013 | 7.1637 | 8.2296 |
| 2008 | **3.7597** | **3.6614** | **4.8143** | **4.7068** | Missouri | **5.6902** | **4.9128** | **6.6107** | **5.7695** |
| 2009 | **3.898** | **3.7821** | **5.0509** | **4.8781** | Arkansas | **6.6991** | **6.3144** | **7.6860** | **7.3413** |
| 2010 | **2.7168** | **2.7111** | 3.4873 | 3.4901 | Average | 6.8139 | 7.6279 | 7.6462 | 8.4132 |
| 2011 | **3.6341** | **3.6156** | **4.7376** | **4.6817** | | | | | |
| 2012 | 4.7872 | 4.8085 | 6.0166 | 6.0447 | | | | | |
| 2013 | 3.4953 | 3.8161 | 4.464 | 4.8339 | | | | | |
| 2014 | 2.9774 | 3.0689 | 3.8074 | 3.9477 | | | | | |
| 2015 | **3.3629** | **3.3447** | **4.4685** | **4.4134** | | | | | |
| 2016 | **3.9657** | **3.9004** | 4.8404 | 4.8539 | | | | | |
| Average | 3.7016 | 3.7303 | 4.7461 | 4.7813 | | | | | |

Notes: Bolded numbers indicate that flexible Fourier transform (FFT) has lower prediction errors and therefore outperforms ordinary least squares (OLS). MAE, mean absolute error; RMSE, root-mean-square error.

We also conducted a panel model regression, which included county fixed effects, to explore the within variation of the data. The results show lower prediction errors in a time-series prediction and higher errors in cross-sectional prediction for models with county fixed effects (Table 6) when compared with the models without county fixed effects (Table 4). The models with county fixed effects show that OLS has smaller prediction errors than the FFT model when the prediction is cross-sectional. However, for time-series prediction with county fixed effects, out of 17 years, there are 10 (8) times when FFT outperforms OLS in terms of smaller mean MAE (RMSE). Overall, the use of county fixed effects explored the within variation and improved prediction over time, but worsened cross-sectional prediction.

According to the Crop Production report (USDA-NASS, 2016), the root-mean-square percentage error (RMSPE) of AYS/OYS forecasts for soybeans was 6.6% in 2016. In comparison, the RMSPE of our FFT model forecasts in 2016 using time-series prediction was 9.29%. Though the RMSPE from our FFT model is greater than that from USDA survey forecasts, FFT model forecasts can substantially save labor and survey costs.

## 5. Conclusions

In this study, we used FFT to account for global flexibility in the relationship between NDVI throughout the growing season and soybean yield. We produced county-specific coefficients

https://doi.org/10.1017/aae.2019.5 Published online by Cambridge University Press

and elasticities of NDVI on soybean yield. We found that the response of soybean yield to NDVI is different across locations. For some counties located in the northern states, soybean yield is highly positively related with the July NDVI, whereas for other counties located in the south, the August NDVI is a better indicator of the soybean yield. Traditional OLS models seem to underestimate the response of soybean yield to July and August NDVI.

Furthermore, we conducted out-of-sample predictions/forecasts and compared their performances for the OLS and FFT models. On average, predictions in time-series and forecasts from the FFT model outperform those from the OLS models in terms of lower prediction errors. We found that FFT models generally result in better out-of-sample predictions and forecasts than OLS models.

A limitation of this work is that it does not distinguish pixels of soybean crops from those of other crops or vegetation types. Nevertheless, incorporating NDVI in the model still results in significant coefficients and an improved fit. Future work can use filters to select pixels that are highly likely to be soybean crops. However, the use of globally flexible models may capture the heterogeneous soybean to total land ratios across counties by allowing a flexible and nonlinear relationship between NDVI and yield, compared with OLS, thus alleviating the contamination caused by other crops. Future work that applies land cover filters may improve the results even further.

This study uses data from the 10 major soybean-producing states in the United States for which data are readily available. Our results show that using the FFT model helps improve the prediction accuracy (lowers the prediction error) especially in panel predictions. The goal is to improve on the forecast accuracy of soybean yield in order to allow market participants to make more informed decisions with respect to anticipated crop yield and possible resulting prices. The FFT model also has the potential to forecast crop yields in less developed countries where ground fieldwork is too expensive to conduct or where the meteorological network is sparse—making this an alternative feasible solution in making crop yield predictions.

**Author ORCIDs.** Ani L. Katchova 0000-0002-7307-4073

**Acknowledgements.** We thank Joseph Cooper for sharing code with us and Joshua Woodard for sharing NDVI data with us. We thank two anonymous referees and the editor. All errors are our own.

## References

**Adrian, D.** "A Model-Based Approach to Forecasting Corn and Soybean Yields." Paper presented at the Fourth International Conference on Establishment Surveys, Montréal, Québec, Canada, June 11–14, 2012.

**Basnyat, P., B. McConkey, B. Meinert, C. Gatkze, and G. Noble**. "Agriculture Field Characterization Using Aerial Photograph and Satellite Imagery." *IEEE Geoscience and Remote Sensing Letters* **1**, 1(2004):7–10.

**Becker, R., W. Enders, and J. Lee**. "A Stationarity Test in the Presence of an Unknown Number of Smooth Breaks." *Journal of Time Series Analysis* **27**, 3(2006):381–409.

**Bolton, D.K., and M.A. Friedl**. "Forecasting Crop Yield Using Remotely Sensed Vegetation Indices and Crop Phenology Metrics." *Agricultural and Forest Meteorology* **173**, (May 2013):74–84.

**Bose, P., N.K. Kasabov, L. Bruzzone, and R.N. Hartono**. "Spiking Neural Networks for Crop Yield Estimation Based on Spatiotemporal Analysis of Image Time Series." *IEEE Transactions on Geoscience and Remote Sensing* **54**, 11(2016): 6563–73.

**Burke, M., and D.B. Lobell**. "Satellite-Based Assessment of Yield Variation and Its Determinants in Smallholder African Systems." *Proceedings of the National Academy of Sciences of the United States of America* **114**, 9(2017):2189–94.

**Chang, J., M.C. Hansen, K. Pittman, M. Carroll, and C. DiMiceli**. "Corn and Soybean Mapping in the United States Using Modis Time-Series Data Sets." *Agronomy Journal* **99**, 6(2007):1654–64.

**Chang, Y., C.S. Kim, J.I. Miller, J.Y. Park, and S. Park**. "A New Approach to Modeling the Effects of Temperature Fluctuations on Monthly Electricity Demand." *Energy Economics* **60**, SC(2016):206–16.

**Cooper, J., A. Nam Tran, and S. Wallander**. "Testing for Specification Bias with a Flexible Fourier Transform Model for Crop Yields." *American Journal of Agricultural Economics* **99**, 3(2017):800–17.

**Doraiswamy, P.C., and P.W. Cook**. "Spring Wheat Yield Assessment Using NOAA AVHRR Data." *Canadian Journal of Remote Sensing* **21**, 1(1995):43–51.

https://doi.org/10.1017/aae.2019.5 Published online by Cambridge University Press

**Enders, W., and J. Li**. "Trend-Cycle Decomposition Allowing for Multiple Smooth Structural Changes in the Trend of US Real GDP." *Journal of Macroeconomics* **44**, SC(2015):71–81.

**Fenton, V.M., and A.R. Gallant**. "Qualitative and Asymptotic Performance of SNP Density Estimators." *Journal of Econometrics* **74**, 1(1996):77–118.

**Ferencz, C., P. Bognar, J. Lichtenberger, D. Hamar, G. Tarcsai, G. Timar, and B. Szekely**. "Crop Yield Estimation by Satellite Remote Sensing." *International Journal of Remote Sensing* **25**, 20(2004):4113–49.

**Fieuzal, R., C.M. Sicre, and F. Baup**. "Estimation of Corn Yield Using Multi-Temporal Optical and Radar Satellite Data and Artificial Neural Networks." *International Journal of Applied Earth Observation and Geoinformation* **57**, (May 2017):14–23.

**Gallant, A.R.** "The Fourier Flexible Form." *American Journal of Agricultural Economics* **66**, 2(1984):204–8.

**Gallant, A.R.** "Identification and Consistency in Semi-Nonparametric Regression." *Advances in Econometrics* 1(1994):145–69.

**Irwin, S., D. Good, and D. Sanders**. "Understanding and Evaluating WAOB/USDA Soybean Yield Forecasts." *Farmdoc Daily* 5(May 7, 2015):84.

**Irwin, S., D. Good, and M. Tannura**. *Early Prospects for 2009 Corn Yields in Illinois, Indiana, and Iowa*. Urbana: Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, Marketing and Outlook Brief 09-01, 2009.

**Irwin, S.H., D.R. Sanders, and D.L. Good**. *Evaluation of Selected USDA WOAB and NASS Forecasts and Estimates in Corn and Soybeans*. Urbana: Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, Marketing and Outlook Research Report 2014-01, 2014.

**Johnson, D.M.** "An Assessment of Pre- and Within-Season Remotely Sensed Variables for Forecasting Corn and Soybean Yields in the United States." *Remote Sensing of Environment* **141**, (February 2014):116–28.

**Johnson, M.D., W.W. Hsieh, A.J. Cannon, A. Davidson, and F. Bédard**. "Crop Yield Forecasting on the Canadian Prairies by Remotely Sensed Vegetation Indices and Machine Learning Methods." *Agricultural and Forest Meteorology* **218–219**, (March 2016):74–84.

**Jones, P.M., and W. Enders.** "On the Use of the Flexible Fourier Form in Unit Root Tests, Endogenous Breaks, and Parameter Instability." *Recent Advances in Estimating Nonlinear Models*. J. Ma, and M. Wohar, eds. New York: Springer, 2014, pp. 59–83.

**Kaul, M., R.L. Hill, and C. Walthall**. "Artificial Neural Networks for Corn and Soybean Yield Prediction." *Agricultural Systems* **85**, 1(2005):1–18.

**Li, A., S. Liang, A. Wang, and J. Qin**. "Estimating Crop Yield from Multi-Temporal Satellite Data Using Multivariate Regression and Neural Network Techniques." *Photogrammetric Engineering and Remote Sensing* **73**, 10(2007):1149–57.

**Lobell, D.B., and G.P. Asner**. "Climate and Management Contributions to Recent Trends in U.S. Agricultural Yields." *Science* **299**, 5609(2003):1032.

**Lv, X.** "Remote Sensing, Normalized Difference Vegetation Index (NDVI), and Crop Yield Forecasting." Master's thesis, University of Illinois at Urbana-Champaign, Urbana, 2014.

**Maselli, F., and F. Rembold**. "Integration of LAC and GAC NDVI Data to Improve Vegetation Monitoring in Semi-Arid Environments." *International Journal of Remote Sensing* **23**, 12(2002):2475–88.

**Mkhabela, M., P. Bullock, S. Raj, S. Wang, and Y. Yang**. "Crop Yield Forecasting on the Canadian Prairies Using MODIS NDVI Data." *Agricultural and Forest Meteorology* **151**, 3(2011):385–93.

**Monahan, J.H.** *Enumeration of Elementary Multi-Indices for Multivariate Fourier Series*. Raleigh: North Carolina State University, Institute of Statistics Mimeo Series No. 1338, 1981.

**Prasad, A.K., L. Chai, R.P. Singh, and M. Kafatos**. "Crop Yield Estimation Model for Iowa Using Remote Sensing and Surface Parameters." *International Journal of Applied Earth Observation and Geoinformation* **8**, 1(2006):26–33.

**Senay, G.** "The Power of Remote Sensing: Global Monitoring of Weather, Water, and Crops with Satellites and Data Integration." *Resource Magazine* **23**, 2(2016):6–9.

**Thompson, L.M.** *Weather and Technology in the Production of Corn and Soybeans*. Ames: Iowa State University, CARD Reports 17, 1963. Internet site: https://lib.dr.iastate.edu/card_reports/17/ (Accessed May 8, 2017).

**Tucker, C.J.** "Red and Photographic Infrared Linear Combinations for Monitoring Vegetation." *Remote Sensing of Environment* **8**, 2(1979):127–50.

**U.S. Department of Agriculture, National Agricultural Statistics Service (USDA-NASS)**. *The Yield Forecasting and Estimating Program of NASS*. Washington, DC: USDA-NASS, Statistical Methods Branch, Staff Report No. SMB 12-01, 2012.

**U.S. Department of Agriculture, National Agricultural Statistics Service (USDA-NASS)**. *Crop Production*. Washington, DC: USDA-NASS, 2016. Internet site: https://nass.usda.gov/Publications/Todays_Reports/reports/crop0818.pdf (Accessed May 8, 2017).

**Woodard, J.** "Big Data and Ag-Analytics: An Open Source, Open Data Platform for Agricultural and Environmental Finance, Insurance, and Risk." *Agricultural Finance Review* **76**, 1(2016):15–26.

**Cite this article:** Xu C and Katchova AL (2019). Predicting Soybean Yield with NDVI Using a Flexible Fourier Transform Model. *Journal of Agricultural and Applied Economics* **51**, 402–416. https://doi.org/10.1017/aae.2019.5