

# Multi-step commodity forecasts using deep learning

Agricultural  
Finance  
Review

Siddhartha S. Bora

*Davis College of Agriculture and Natural Resources, West Virginia University,  
Morgantown, West Virginia, USA, and*

Ani L. Katchova

*Department of Agricultural, Environmental, and Development Economics,  
The Ohio State University, Columbus, Ohio, USA*

269

Received 31 August 2023  
Revised 27 April 2024  
Accepted 28 April 2024

## Abstract

**Purpose** – Long-term forecasts about commodity market indicators play an important role in informing policy and investment decisions by governments and market participants. Our study examines whether the accuracy of the multi-step forecasts can be improved using deep learning methods.

**Design/methodology/approach** – We first formulate a supervised learning problem and set benchmarks for forecast accuracy using traditional econometric models. We then train a set of deep neural networks and measure their performance against the benchmark.

**Findings** – We find that while the United States Department of Agriculture (USDA) baseline projections perform better for shorter forecast horizons, the performance of the deep neural networks improves for longer horizons. The findings may inform future revisions of the forecasting process.

**Originality/value** – This study demonstrates an application of deep learning methods to multi-horizon forecasts of agri-cultural commodities, which is a departure from the current methods used in producing these types of forecasts.

**Keywords** Agricultural baselines, USDA baselines, Commodity projections, Forecast evaluation, Deep learning

**Paper type** Research paper

## 1. Introduction

The availability of long-term information about commodity markets plays a vital role in policy and investment decisions by market participants. The forecasts of season-average farm prices of major field crops such as corn, soybeans, and wheat are widely used to inform decisions by farmers, agricultural businesses and the government. Similarly, the forecasts of harvested area and yield provide information about production of the commodities for the marketing year and help anticipate ending stocks. The USDA's World Agricultural Supply and Demand Estimates (WASDE) provide forecasts about commodities for the current marketing year. However, market participants may need information about market trends beyond the current marketing year to

### JEL Classification — C53, E37, Q14, Q18

© Siddhartha S. Bora and Ani L. Katchova. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

Work on this article was performed while Ani L. Katchova was on a sabbatical at USDA-ERS. We thank Utpal Vasavada and Saleem Shaik for their helpful comments and suggestions. This study was supported in part by Cooperative Agreement #58-3000-1-0068 from the USDA-ERS.

**Funding:** Scientific Article No: 3473 of the West Virginia Agricultural and Forestry Experiment Station. This study is funded in part by Cooperative Agreement #58-3000-1-0068 from the USDA, ERS. This material is based upon work that is partially supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, [Hatch Project], under [WVA00779].



Agricultural Finance Review  
Vol. 84 No. 4/5, 2024  
pp. 269-296  
Emerald Publishing Limited  
0002-1466  
DOI 10.1108/AFR-08-2023-0105

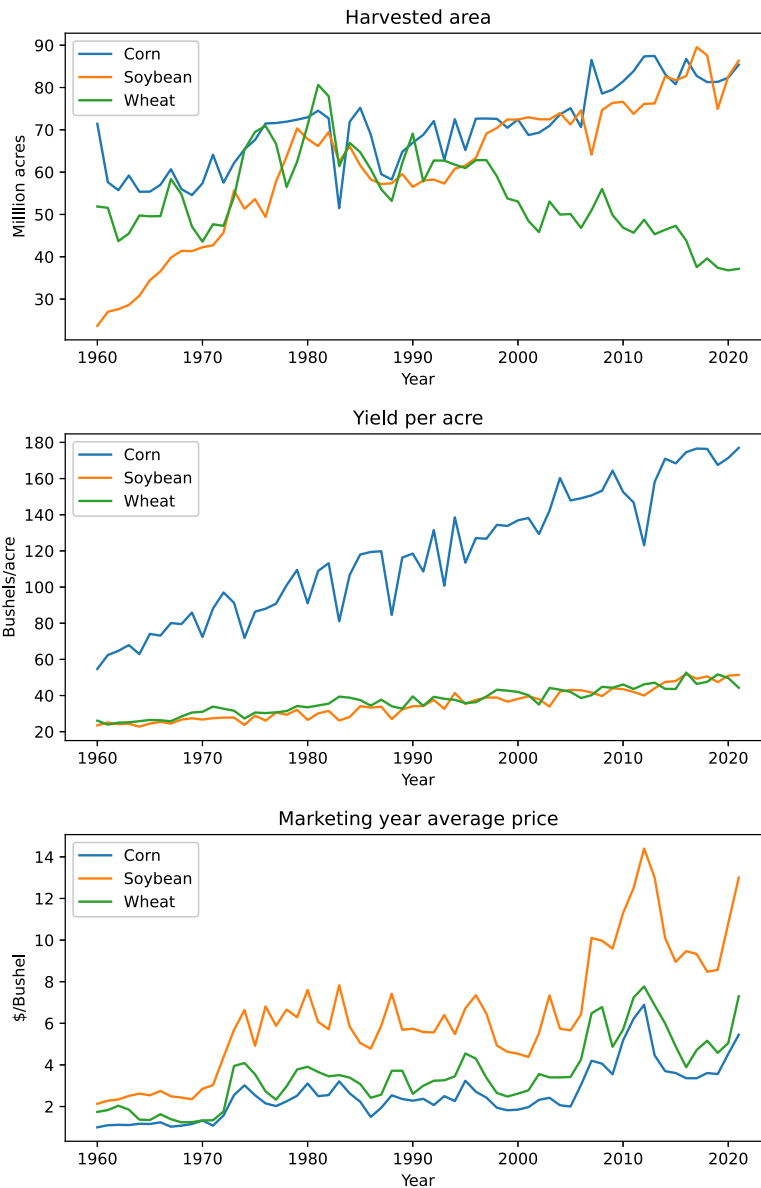
DOI 10.1108/AFR-08-2023-0105

inform their decisions. For example, forecasts for the next few years can facilitate comparisons of policy alternatives by government agencies. Similarly, long-term forecasts can help estimate the outlays of various farm program costs under the federal budget. The Farm Bill programs are typically implemented in five-year cycles, and having information for the next five years will help immensely in planning the budget. Similarly, long-term prices and crop yield forecasts may help farmers inform their long-term decisions about planting, crop choice, and land use. For example, the decision to enroll farmland in federal programs like conservation reserve programs (CRP) may be informed by crop prices and yield forecasts for multiple years into the future. The importance of reliable long-term forecasts became evident when the pandemic hit the economy, and policymakers required information far into the future to plan the recovery process.

The USDA's baseline projections, published every year in February, are one of the principal sources of long-term information about the USA farm sector. The baselines are produced by a team from ten USDA agencies, including the Economic Research Service (ERS), and contain annual projections of key measures of agricultural market conditions for the next decade. These projections facilitate comparisons of policy alternatives by providing a conditional "baseline" scenario based on specific macroeconomic, weather, policy and trade assumptions. Over the years, the baseline projections have been used for a variety of purposes, including estimating farm program costs and preparing the president's budget. In addition to the USDA, the Food and Agricultural Policy Research Institute (FAPRI), University of Missouri, produces similar ten-year projections of key agricultural variables. The baseline projections are produced through a mixture of the output of quantitative models and expert opinions. Previous studies show that many variables in the USDA baseline projections are biased and that the predictive content of the baselines diminishes after a few years (Bora *et al.*, 2023; Katchova, 2024; Fang and Katchova, 2023; Chandio and Katchova, 2024). As the evaluation of the baselines has shown their limited predictive content, an investigation of alternative methods to improve the long-term projections becomes essential.

This study aims to forecast the harvested area, yield and farm price of three major field crops in the USA for the next five years using deep learning models. Our investigation is performed in three steps. First, we formulate a supervised learning problem for the forecasting process and develop a test harness to compare the performance of various methods based on a train-test split of the sample. The last ten years were used as a test sample using a walk-forward validation approach. Second, we benchmark the performance of traditional methods such as a naïve no-change forecast, exponential smoothing and USDA baseline reports. Finally, we implement a suite of deep learning models to predict commodity market indicators, with particular emphasis on long short-term memory (LSTM) recurrent neural networks (RNN), convolutional neural networks (CNN) and their hybrids. We train the deep learning models using a large number of input features reflecting macroeconomic indicators, demographic trends, weather variability, global trade and demand and supply of key commodities.

Previous studies have looked at the potential of using machine learning to improve commodity forecasts. Using satellite and weather data, Roznik *et al.* (2023) show that XGBoost-based machine learning models can produce reasonably accurate crop yield forecasts comparable to those produced by WASDE reports. We investigate the potential of employing deep learning techniques in forecasting multi-year forecasts of crop harvested acres, yield and farm price. The closest available information to such forecasts are the USDA baselines, which are the result of a complex process involving econometric analysis and expert inputs from various agencies within the USDA. Compared to traditional time-series forecasting models, deep neural networks, particularly LSTMs, excel at capturing nonlinear dependencies within sequential data (Panigrahi and Behera, 2017) and can model complex relationships between various factors such as macroeconomic conditions, weather and crop variables. With careful feature selection for economic interpretation, deep neural networks can also process diverse types of inputs.



Source(s): Created by the authors using data sourced from NASS Quickstats

**Figure 1.**  
Historical harvested  
area, yield and farm  
price of corn, soybeans  
and wheat, 1961–2021

Another benefit is that the forecasts can be generated using publicly accessible data by any interested market participant.

Our study contributes to the literature in several ways. We use state-of-the-art deep learning methods to improve the long-term forecasts of commodity market indicators.

While deep learning methods have shown great promise in forecasting in other fields (Kim and Won, 2018; Huang *et al.*, 2020; Wang *et al.*, 2019; Borovykh *et al.*, 2019; Wan *et al.*, 2019; Lara-Benítez *et al.*, 2020), their application in predicting long-term agricultural statistics such as the USDA baselines has been limited. This study aims to bridge this gap. Our results suggest that deep learning networks may perform better than the official USDA baselines at longer forecast horizons. In particular, when the USDA baselines perform well, deep learning models match the accuracy, but if the USDA baselines do not perform well, deep learning models perform better. These findings may have important implications for future revisions of the USDA baseline models and processes. Deep learning models with improved accuracy may offer insights for the existing USDA baseline models by determining where improvements to the accuracy and performance of these models can occur. The existing process of producing the baseline reports involves many agencies, which work on specific components of the report and create inputs for the composite model. Deep learning methods have the potential to contribute to their work by identifying where the original USDA baseline models can be improved so they become more accurate and perform better.

The remainder of this article is organized as follows: The next section describes the various datasets used in this study. The third section describes the methodology, followed by results and discussion. The final section contains concluding remarks.

## 2. Data

Our dataset of the target variables consists of historical values of harvested area, yield and farm price of corn, soybeans and wheat in the USA since 1961. Together, these three field crops constitute a significant share of the area under cultivation in the USA. The values are averages for the marketing years, which differ by crop. The marketing year for corn and soybeans begins on September 1 and comprises four quarters. For example, the marketing year 2021–2022 for corn and soybeans starts on September 1, 2021, and ends on August 31, 2022. The 2021–2022 marketing year for wheat begins on June 1, 2021, and ends on May 31, 2022. All this information was obtained using the National Agriculture Statistics Services (NASS) Quickstats API (USDA National Agricultural Statistics Service, 2024). Figure 1 shows the plots of harvested area, yield and farm price of the three crops for the period 1961–2021. The figures suggest that many of these indicators are highly correlated, and they may be related to each other or to other macroeconomic, weather or trade indicators. For example, the loss of wheat harvested area over the years is accompanied by a contemporaneous increase in soybean harvested area.

An archive of the USDA agricultural baseline projections since 1997 is available at the Albert R. Mann Library at Cornell University (USDA ERS, 2024). The baseline reports typically include estimates of the previous year(s) and projections for the next ten years. For example, the February 2022 USDA report contains realized estimates for 2020, provisional estimates for 2021 and projections for 2022–2031 (USDA Office of Chief Economist, 2022). The exact information set, which was available to the committee producing the projections in the early years, is difficult to retrieve due to a lack of information on the variables that were in the information set of the committee and the revisions made to the realized values over time. As the organizations involved with the projection process go through personnel and information technology infrastructure changes over the years, the exact information used to produce the baselines is challenging to ascertain. The projections and estimates are often revised long after they are first published. For example, there is no way to access the exact data used as the information set by the committee when the baseline projections were prepared for 1997. We can assume that the committee made the best use of the information they had at that time. To mimic the forecasting process of the committee, we

try to provide many features such as macroeconomic, population, trade and weather information input to train our deep learning models. The committee may have had a different set of variables and/or different values for these variables that were later revised to what is available today. Our goal is to use deep learning methods to produce the forecasts using a similar information set and examine whether these forecasts have superior performance over the USDA baselines.

We use data from several sources as input features to train the deep learning models. First, we use the lagged values of commodity indicators to forecast their future values. We also include several macroeconomic, population, trade, and weather variables for the world and the USA as input features to our models. These include growth rates for gross domestic product (GDP) and population. For the USA economy, we also include inflation, unemployment, labor market participation and interest rates. We also include features that represent changes in weather in the world and the USA over time. To account for temperature changes all over the world, we include global annual average temperature

Feature	Descriptive statistics				
	Mean	Median	SD	Min	Max
US CPI inflation	3.73	3.02	2.76	-0.36	13.55
US population growth	1.00	0.96	0.26	0.12	1.66
US GDP growth	2.97	3.10	2.20	-3.40	7.24
US unemployment rate	6.00	5.65	1.60	3.50	9.70
US labor force participation rate	63.64	63.90	2.70	58.70	67.20
US merchandise trade	15.49	15.57	5.19	6.40	24.03
US CPI index 2010	61.63	62.46	35.25	13.71	124.27
US Real interest rate	3.64	3.07	2.29	-1.28	8.59
US GDP per capita growth	1.95	2.03	2.16	-4.33	6.31
US lending interest rate	7.07	6.82	3.28	3.25	18.87
World annual population growth	1.57	1.57	0.34	0.94	2.13
World GDP growth	3.47	3.62	1.74	-3.27	6.56
World GDP per capita growth	1.87	2.06	1.62	-4.27	4.82
World merchandise trade	32.84	30.60	10.50	16.04	51.07
World urban growth	2.44	2.44	0.32	1.69	3.23
Global temperature anomalies	0.38	0.36	0.31	-0.14	0.99
Average temperature	52.65	52.54	1.07	50.88	55.28
Maximum temperature	64.52	64.34	1.13	62.68	67.69
Minimum temperature	40.78	40.70	1.08	38.87	43.13
Heating degree days	4485.52	4516.00	290.17	3780.00	5030.00
Cooling degree days	1258.90	1235.00	125.40	1016.00	1547.00
Precipitation	30.81	30.62	2.16	25.70	34.96
Corn ending stocks	1571.36	1435.91	855.04	425.92	4881.68
Corn exports	1628.84	1821.32	622.53	396.36	2752.62
Corn use, feed and residual	4628.84	4659.45	889.28	2966.15	6131.62
Corn use, food seed industrial	2428.12	1533.50	2354.49	314.94	7056.77
Corn imports	14.23	7.09	23.74	0.47	159.95
Soybean crush	1257.64	1253.99	546.07	0.00	2190.00
Soybean ending stocks	237.39	208.08	155.67	0.00	909.04
Soybean exports	908.27	804.65	554.86	0.00	2265.43
Soybean imports	7.12	3.38	12.12	0.00	71.76
Wheat domestic	1018.68	1113.70	231.45	581.10	1389.28
Wheat ending stocks	852.60	845.25	347.27	305.82	1904.98
Wheat exports	1055.19	1040.40	255.55	543.99	1770.72
Wheat imports	56.55	40.71	53.61	0.99	172.47

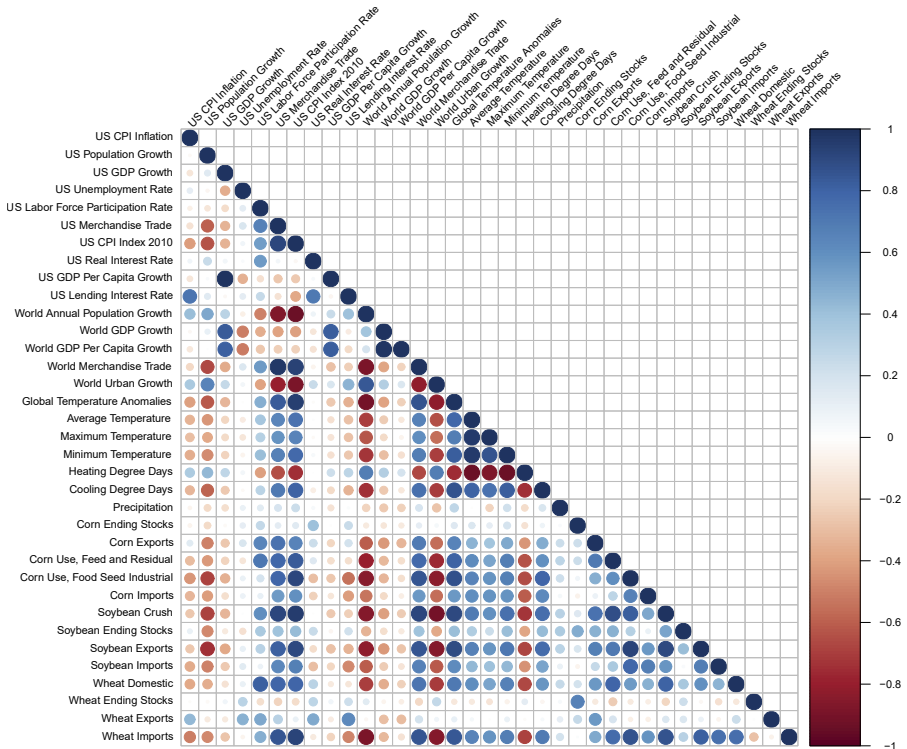
Source(s): Created by the authors

**Table 1.**  
Descriptive statistics  
of input features

anomalies, measured as deviations from the 20th century average. The macroeconomic data are taken from the World Bank Open Data Catalog. Similarly, we include the USA's annual average temperature, maximum temperature, minimum temperature, precipitation and heating and cooling degree days. All weather information was obtained from the National Oceanic and Atmospheric Administration (NOAA) (National Centers for Environmental Information, 2022). Finally, we add commodity balance sheet variables representing domestic use, imports, exports and ending stocks of corn, soybeans and wheat as input features. The commodity balance sheet information is extracted from the Production, Supply and Distribution (PSD) Database published by the USDA Foreign Agricultural Service (USDA Foreign Agricultural Service, 2024). We have provided the descriptive statistics of the input features in Table 1 and their correlation plot in Figure 2.

**3. Methodology**

In this section, we define our prediction problem and proceed to develop a test harness for comparing the performance of the methods used in this study. We then describe the different traditional and deep learning methods used in this study.



**Figure 2.**  
Correlation plot of  
input features

**Note(s):** The features were first scaled using MinMax Scaler before calculating correlation indices

**Source(s):** Created by the authors

### 3.1 The prediction problem

We denote the realized or actual values of commodity indicators of harvested acres, yield and farm price for corn, soybeans and wheat in year  $t$  by  $y_t$ . At year  $t$ , the forecaster makes a forecast  $\hat{y}_{t+h|t}$  for horizon  $h \in \{0, 1, \dots, H-1\}$  for  $H$  future years, including year  $t$  using lagged values of the commodity indicators and a set of other covariates such as macroeconomic, population and weather variables. Although the baselines are for  $H = 10$  years, we limit our attention to forecasts of up to five years due to the small length of the time period, i.e.  $h = \{0, \dots, 4\}$ . Similarly, we assume that up to five years of lagged values of input features are used to produce the forecasts.

We first transform the prediction problem into a supervised learning problem where a set of input features  $X$  is mapped to an output variable  $y$ . For year  $t$ , our input  $X_t$  consists of vectors of all input features up to lag five and  $y_t$  consists of vectors of the next five years of values of the target variables (harvested acres, yield and farm price of corn, soybeans and wheat). From our dataset for the time period 1961–2021, we construct  $\{X_t, y_t\}$  pairs for 52 years between 1966 and 2017. This yields a three-dimensional array of input features  $X$  with dimensions  $(52, 5, n\_features)$ , where  $n\_features$  is the total number of input features. This is important since the deep learning models used in this study accept three-dimensional input. We use a total of 44 input features in this study; however, this number can be augmented by including additional features.

### 3.2 Developing a test harness

A test harness ensures that all deep learning methods used in this study are evaluated using a consistent approach for comparability. The important components of our test harness are the train-test split validation strategy and the evaluation criteria.

**3.2.1 Train-test split.** Our dataset contains commodity market variables of harvested area, yield and farm price for corn, soybeans and wheat between 1961 and 2021. Since we use up to five-year-lagged features in our deep learning algorithms to produce five-year-ahead forecasts, this results in a complete dataset of features ( $X$ ) and output ( $y$ ) between 1966 and 2017, for a total of 52 years. We used the last ten years of the data as our test sample between 2008 and 2017, representing close to 20% of the entire sample. As preferred in time-series applications, we use a walk-forward validation strategy, allowing updated information to train the model as we progress through the years in the test sample. We use an expanding training window approach, which means the training sample increases as we walk through the test sample. For example, we train a model using 42 samples between 1966 and 2007 to produce forecasts for 2008. We then add the sample for 2008 back to the training sample to produce forecasts for 2009 and so on. This validation strategy closely follows how the USDA produces the baseline reports as forecasters make use of new information as it becomes available. Another choice is to use a sliding window, where the oldest training sample is dropped as we add a new sample, keeping the length of the training sample constant. However, we prefer an expanding window as we would like to make use of all the information available, and our sample size is small.

**3.2.2 Evaluation criteria.** We will use two widely adopted error metrics to measure the performance of the proposed methods: root mean squared error (RMSE) and mean absolute percent error (MAPE). The RMSE is calculated at the level of the variables, while the MAPE is calculated relative to the actual level of the variables according to the following formulas:

$$RMSE_h = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_{t+h} - \hat{y}_{t+h|t})^2} \quad (1)$$

$$MAPE_h = \frac{100}{T} \sum_{t=1}^T \left| \frac{y_{t+h} - \hat{y}_{t+h|t}}{y_{t+h}} \right| \quad (2)$$

where  $y_{t+h}$  are the realized values,  $\hat{y}_{t+h|t}$  are forecasts of the target variable at horizon  $h$  and  $T$  is the sample size of the test or the training sample. For calculating in-sample forecast errors, we use the sample size  $T = 42$  for the training sample, while for out-of-sample errors, we use the test sample  $T = 10$ .

### 3.3 Benchmarking with traditional methods

**3.3.1 Naïve benchmark.** We first develop a benchmark model to improve upon using deep learning methods. A natural choice is to use a naïve no-change forecast, where we consider the most recent year's value as the forecast for the next five years. This is a fairly naïve benchmark that would result in high forecast errors. Any econometric or deep learning method is expected to perform better than this naïve benchmark, as the methods are supposed to add some skill to forecasting.

**3.3.2 Simple exponential smoothing (ETS).** We also use the simple exponential smoothing (ETS) method, which is useful for forecasting when the time series have no clear trend or seasonal pattern. The ETS forecast is a weighted average of past observations, where the weights decay exponentially for older observations. The ETS method can be expressed in terms of the following equations (Hyndman and Athanasopoulos, 2021),

$$\hat{y}_{t+h|t} = \ell_t \quad (3)$$

$$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1} \quad (4)$$

where  $\ell_t$  is the level of the variable at time  $t$ . The smoothing parameter  $\alpha$  represents the rate at which the weight placed on past observations decreases.

**3.3.3 Exponential smoothing (ETS) with trend.** We then use an extension of the simple ETS method, which allows a trend (Holt, 2004). Some of our data series, such as crop yield, shows a clear time trend, and farm price may also be trending upward over the years. The ETS method with trend can be expressed as Hyndman and Athanasopoulos (2021),

$$\hat{y}_{t+h|t} = \ell_t + hb_t \quad (5)$$

$$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad (6)$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1} \quad (7)$$

where  $\beta$  is an additional smoothing parameter for the trend. We use the implementations of ETS and ETS with trend methods in Python *statsmodels* library to produce the forecasts (Seabold and Perktold, 2010).

**3.3.4 Auto-regressive integrated moving average (ARIMA).** Auto-regressive integrated moving average (ARIMA) models are useful in forecasting when the time series can be made stationary by differencing. An ARIMA ( $p$ ,  $d$  and  $q$ ) model consists of  $p$  autoregressive terms and  $q$  lagged forecast errors in the prediction equation and needs  $d$  times differencing to achieve stationarity. While ARIMA models are traditionally univariate, ARIMA including exogenous features (ARIMAX) is also available. We employ an implementation of ARIMA model from the *pmdarima* library in Python, which automates the calibration of the ARIMA models in their *auto\_arima* function (Smith, 2017).

*3.3.5 USDA baseline report.* Our final choice for comparison is the projections produced by the USDA in their baseline report. These projections are produced using a mixture of economic/econometric models, survey information and expert opinions. We calculate the error metrics for baseline projections up to five years for the test period 2008–2017 for comparison with the other methods used in our study. As mentioned earlier, the exact information set used to produce these projections is challenging to ascertain. Therefore, the comparison with deep learning methods using the current training set may not be entirely justifiable.

### *3.4 Deep learning methods*

The methods discussed in the previous section are traditional time-series forecast models. However, in recent years, deep neural networks have become popular for forecasting time series (Schmidhuber, 2015). Neural networks are a collection of algorithms used in pattern recognition. Deep learning refers to a subset of neural networks, which consists of more than three layers.

The most basic deep learning networks are feed-forward neural networks (FNNs) that do not allow recursive feedback, such as the multi-layer perceptron (MLP). The computational architecture of FNNs consists of three layers: an input layer, hidden layer(s) and an output layer. Since two consecutive layers have only direct forward connections, FNNs ignore the temporal nature of the data and treat each input independently. Therefore, they are of limited use in dealing with our data, which are inherently temporal and sequential. We consider two main families of deep learning methods that account for temporal dependence in sequences, namely RNN and CNN. We also explore hybrid deep learning models, which have become more popular in recent years.

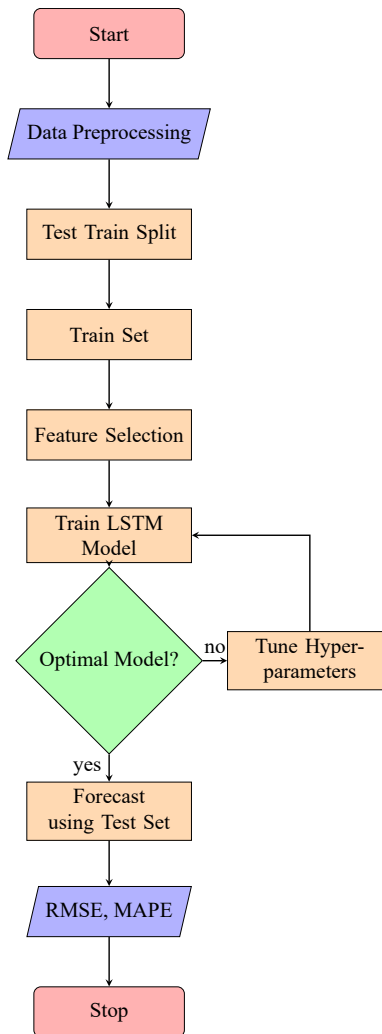
*3.4.1 Recurrent neural networks.* RNNs are popular in time series prediction applications. An RNN allows recursive feedback, and each RNN unit can take the current and previous input simultaneously. They are widely used for prediction in different fields, including stock price forecasting (Kim and Won, 2018), wind speed forecasting (Huang *et al.*, 2020) and solar radiation forecasting (Wang *et al.*, 2019). Moreover, RNNs have done remarkably well at forecasting competitions, such as the recent M4 forecasting competition (Makridakis *et al.*, 2018). In a recent study, Medvedev and Wang (2022) used RNNs to predict the volatility of the S&P 500 Index (SPX) for pricing options, with good success. However, we are not aware of any studies applying RNNs to forecast long-term information about agricultural markets.

Elman (1990) proposed an early RNN, which generalizes FNN by using recurrent links in order to provide networks with dynamic memory. This type of network is more suitable for handling ordered data sequences like financial time series. While Elman's RNN model is simple, training these models is difficult due to inefficient gradient propagation. In particular, the problem of vanishing and exploding gradients makes it challenging to learn long-term dependencies. Due to vanishing gradients, it may take a long time to train the model, while the exploding gradients may cause the model's weights to oscillate (Lara-Benítez *et al.*, 2021).

LSTM networks were proposed to address the vanishing and exploding gradient problems faced by standard RNNs (Hochreiter and Schmidhuber, 1997). LSTMs can model long-term temporal dependencies without compromising short-term patterns. LSTM networks have a similar structure to Elman's RNN but differ in the composition of the hidden layer, known as the LSTM memory cell. Each LSTM cell has three gates: a multiplicative input that controls memory units, a multiplicative output that protects other cells from noise and a forget gate. Gated recurrence units (GRUs) are simplified versions of LSTMs that replace the forget and input gates with a single update gate to

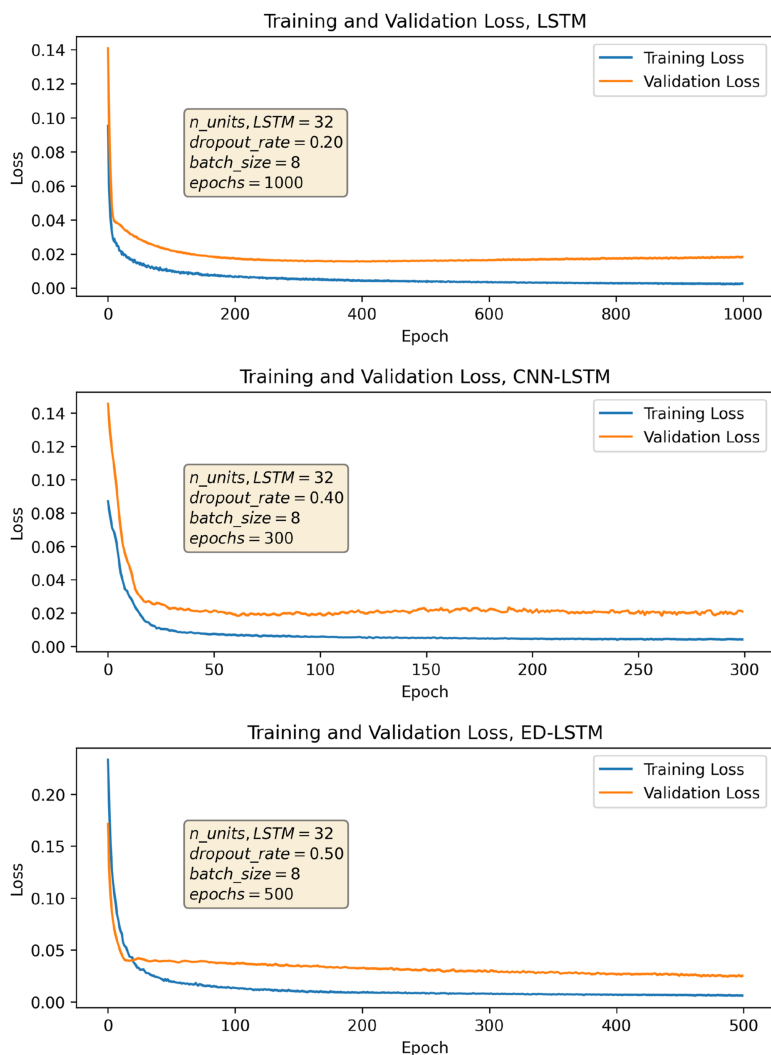
reduce trainable parameters. An RNN can also have stacked recurrent layers to form a deep RNN.

*3.4.2 Convolutional neural networks.* CNNs are mainly used in classification applications such as speech recognition, object recognition and natural language processing (NLP). However, with some adjustments, they can be used for time-series predictions as well. A CNN uses the convolutional operation to extract meaningful features from raw data and create feature maps (Lara-Benítez *et al.*, 2021). A CNN consists of convolution layers, pooling layers and fully connected layers. The pooling layers lower the spatial dimension of the feature maps, while the fully connected layers combine the local features to form global features. As CNNs have a smaller number of trainable parameters, the learning



**Figure 3.**  
Flowchart  
representing the  
forecasting process

**Source(s):** Created by the authors



Source(s): Created by the authors

Figure 4.  
Learning curves for the  
LSTM models

process is more time-efficient than RNNs (Borovykh *et al.*, 2019). In addition, different convolutional layers can be stacked together to allow the transformation of raw data (Chen *et al.*, 2020).

Hybrid models are a recent trend in time-series forecasting using deep learning. For example, depending on the application, LSTMs can be used with RNNs or CNNs. Also, deep learning models can be used with traditional econometric methods to achieve superior results. The winning entry in the M4 forecasting competition in 2018 used a hybrid ETS-LSTM model (Smyl, 2020). While the ETS component captures seasonality, the LSTM focuses on non-linear trends and cross-learning from related series.

3.4.3 *Training the deep neural networks.* In this study, we use three deep learning architectures to forecast commodity market indicators.

- (1) *Vanilla LSTM:* The first architecture that we use is a simple LSTM model with one LSTM layer.
- (2) *Encoder-decoder LSTM (ED-LSTM):* The second architecture that we use is an ED-LSTM with two layers. The first layer reads the input sequence and encodes it into a fixed-length vector, and the second layer decodes the fixed-length vector and outputs the predicted sequence.
- (3) *CNN-LSTM:* The last architecture that we use consists of an LSTM preceded by a convolution layer at the input.

Each architecture has a multiple-input multiple-output configuration, with three-dimensional tensors as inputs and output. The input consists of lagged values of the commodity variables, and additional features representing macro-economic

Horizon	Method	Corn		Soybeans		Wheat	
		RMSE	MAPE(%)	RMSE	MAPE(%)	RMSE	MAPE(%)
<i>h = 0</i>							
	Naive	4.03	4.06	4.65	4.13	3.86	7.31
	USDA	1.84	1.70	3.40	3.79	2.02	3.69
	ETS	3.33	3.61	4.60	4.17	3.86	7.31
	ETS Trend	3.16	3.40	3.93	3.58	3.85	7.31
	ARIMA	3.12	3.31	4.63	4.30	4.67	8.67
<i>h = 1</i>							
	Naive	4.84	5.51	5.68	5.49	4.85	8.74
	USDA	2.77	2.61	5.60	5.86	2.55	4.34
	ETS	4.18	4.22	5.65	5.63	4.85	8.74
	ETS Trend	3.99	4.17	4.31	4.39	4.83	8.61
	ARIMA	3.95	4.09	4.64	4.62	7.70	16.13
<i>h = 2</i>							
	Naive	4.76	4.61	6.71	6.84	5.58	10.52
	USDA	3.03	2.87	7.20	7.93	4.01	8.39
	ETS	4.63	4.58	6.66	6.76	5.58	10.52
	ETS Trend	4.50	4.76	5.45	5.97	5.49	10.20
	ARIMA	4.45	4.71	5.65	5.94	10.13	22.74
<i>h = 3</i>							
	Naive	4.55	4.31	6.97	6.99	6.19	13.58
	USDA	3.21	2.95	7.10	7.62	4.08	8.44
	ETS	4.99	4.84	7.00	7.18	6.19	13.58
	ETS Trend	4.95	5.31	5.42	5.53	6.00	13.11
	ARIMA	4.89	5.24	5.63	5.68	11.99	27.98
<i>h = 4</i>							
	Naive	4.11	3.86	8.20	8.58	7.28	15.57
	USDA	2.90	2.73	7.55	7.81	4.72	10.05
	ETS	5.14	4.72	8.14	8.63	7.28	15.57
	ETS Trend	5.10	5.44	5.80	5.67	7.07	14.94
	ARIMA	5.05	5.40	5.98	5.69	13.54	32.27

**Table 2.** Forecast accuracy for corn, soybeans, and wheat harvested area using traditional/ econometric methods

**Source(s):** Created by the authors

Horizon	Method	Corn		Soybeans		Wheat	
		RMSE	MAPE(%)	RMSE	MAPE(%)	RMSE	MAPE(%)
<i>h = 0</i>							
	Naive	15.18	7.53	2.82	5.48	4.15	6.74
	USDA	14.69	6.88	2.55	4.72	2.82	4.70
	ETS	15.52	8.80	3.20	5.90	3.55	6.08
	ETS Trend	13.73	6.96	3.16	5.69	2.60	4.41
	ARIMA	17.03	9.01	3.18	5.72	3.52	5.95
<i>h = 1</i>							
	Naive	19.95	9.37	3.67	6.98	4.06	6.89
	USDA	14.79	6.99	2.65	4.77	2.67	4.20
	ETS	18.54	10.24	4.02	7.13	3.41	5.72
	ETS Trend	14.31	7.44	3.48	6.24	2.49	3.99
	ARIMA	17.63	9.07	3.58	6.35	3.51	6.25
<i>h = 2</i>							
	Naive	21.99	10.51	4.60	8.25	3.21	5.48
	USDA	14.66	6.77	2.82	5.07	2.83	4.65
	ETS	19.22	9.85	4.67	8.05	3.37	5.85
	ETS Trend	13.75	6.80	3.66	6.21	2.77	4.67
	ARIMA	16.23	7.61	3.69	6.06	2.50	4.25
<i>h = 3</i>							
	Naive	21.92	10.02	4.82	6.52	4.00	6.60
	USDA	14.20	6.50	3.00	5.29	2.88	4.47
	ETS	19.53	9.88	5.12	8.34	3.98	6.41
	ETS Trend	13.38	6.48	3.75	6.48	2.72	4.31
	ARIMA	14.82	7.08	3.93	6.53	3.57	5.58
<i>h = 4</i>							
	Naive	22.53	10.05	5.40	8.88	5.50	9.04
	USDA	13.42	5.39	3.18	5.66	3.28	5.48
	ETS	20.39	11.14	5.76	10.55	4.78	8.12
	ETS Trend	12.81	5.73	4.01	7.32	3.16	5.52
	ARIMA	13.35	5.69	4.00	7.29	5.06	9.38

**Source(s):** Created by the authors

**Table 3.**  
Forecast accuracy for  
corn, soybeans and  
wheat yield using  
traditional/  
econometric methods

environment, weather and commodity balance sheet information. The outputs are five future time steps of the nine target variables: acres, *yields* and *farm prices* of corn, soybeans and wheat. As the input features are correlated, we use a principal component analysis (PCA) of the scaled input features and retain 11 orthogonal features, representing 95% variance in the features. This allows us to reduce the dimensionality of the feature vector.

The forecasting process is outlined in [Figure 3](#). We take a number of steps to ensure that we do not overfit the LSTM models. We train the deep learning networks using the *Keras* ([Chollet, 2015](#)) and *TensorFlow* ([Abadi et al., 2015](#)) libraries in Python. To reduce the chances of overfitting, we carefully choose the model hyperparameters using a grid search methodology over a validation set. In particular, we choose over a hyperparameter space all combinations of select values of the number of LSTM units, batch size, the number of training epochs and the dropout rate. The number of LSTM units refers to the complexity of the LSTM layers. The number of epochs indicates the number of passes

Horizon	Method	Corn		Soybeans		Wheat	
		RMSE	MAPE(%)	RMSE	MAPE(%)	RMSE	MAPE(%)
<i>h = 0</i>							
	Naive	1.05	16.07	1.42	10.25	1.08	17.99
	USDA	0.92	13.25	1.55	10.34	0.93	13.58
	ETS	1.05	16.07	1.55	11.72	1.08	17.99
	ETS Trend	1.07	16.86	1.62	12.32	1.10	18.25
	ARIMA	1.15	20.84	1.44	10.43	1.02	16.86
<i>h = 1</i>							
	Naive	1.61	28.14	2.38	17.60	1.64	26.92
	USDA	1.22	15.20	2.07	13.43	1.11	13.89
	ETS	1.61	28.14	2.44	18.52	1.64	26.92
	ETS Trend	1.66	29.64	2.53	18.93	1.71	28.92
	ARIMA	1.82	37.41	2.47	18.07	1.72	28.06
<i>h = 2</i>							
	Naive	1.89	32.57	2.89	22.36	1.77	27.80
	USDA	1.40	17.47	2.44	16.02	1.29	14.43
	ETS	1.89	32.57	2.96	23.76	1.77	27.80
	ETS Trend	1.97	35.04	3.10	26.10	1.89	29.23
	ARIMA	2.14	42.23	3.03	23.15	1.81	28.55
<i>h = 3</i>							
	Naive	1.92	37.38	3.06	25.50	1.82	29.64
	USDA	1.46	20.56	2.60	16.03	1.52	19.82
	ETS	1.92	37.38	3.15	26.19	1.82	29.64
	ETS Trend	2.04	40.98	3.38	29.47	1.98	31.44
	ARIMA	2.04	41.17	3.22	26.50	1.76	27.92
<i>h = 4</i>							
	Naive	1.91	37.19	3.27	28.63	2.02	31.77
	USDA	1.40	20.35	2.71	18.55	1.66	23.17
	ETS	1.91	37.19	3.42	29.14	2.02	31.77
	ETS Trend	2.08	40.93	3.73	32.59	2.23	36.00
	ARIMA	1.97	38.38	3.47	29.90	1.86	31.47

**Table 4.** Forecast accuracy for corn, soybeans, and wheat farm price using traditional/ econometric methods

**Source(s):** Created by the authors

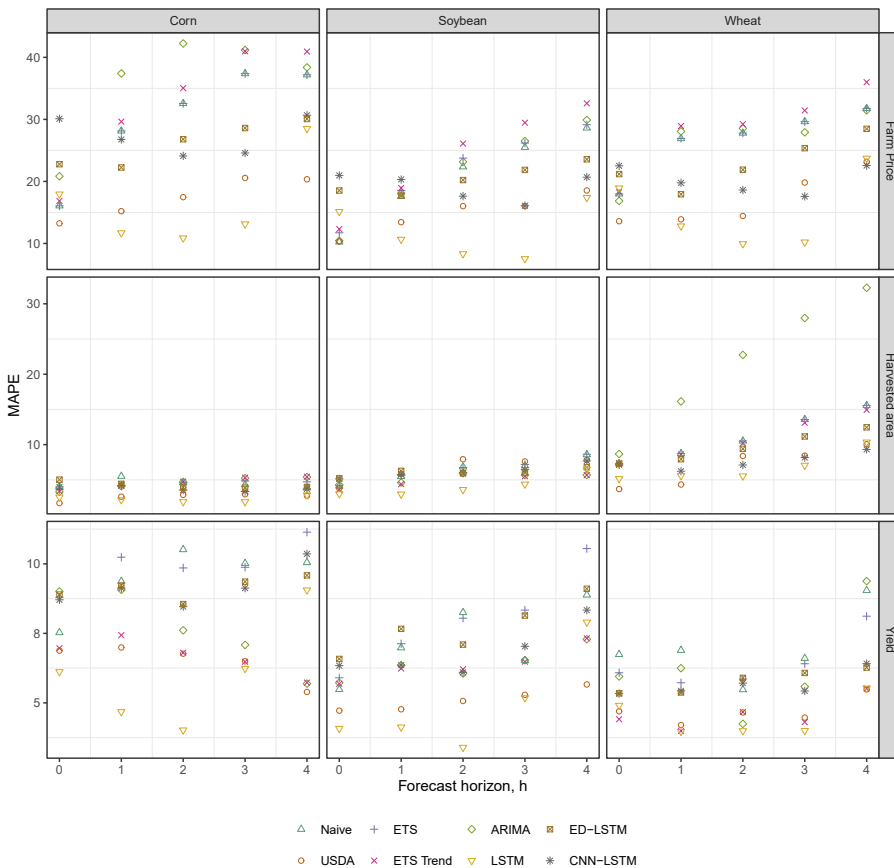
through the training dataset, while the batch size hyperparameter indicates the number of training samples used before the model parameters are updated again.

In Figure 4, we plot the training and validation losses against training epochs for the tuned configurations for each of our models using a 20% validation split. While there is an expected gap between the learning curves of training and validation sets, there is no sign of severe overfitting. The literature shows that while sample size does contribute to the efficiency of the LSTM model predictability, a model is inefficient with small sample sizes. Boulmaiz *et al.* (2020) show these results for rainfall-runoff modeling and streamflow predictions. To predict streamflow, Boulmaiz *et al.* (2020) train an LSTM model with different lengths of data ranging from 3 to 15 years. The results of this study indicate the efficiency of the LSTM model using a small, three year sample size compared to a benchmark model that requires nine years of data to yield similar results. The study tested varying sizes of data from hundreds of watersheds to train an optimal LSTM model. Where it is traditionally expected that more data would result in higher model accuracy, this study finds that, beyond an optimal value, model accuracy does not improve significantly with the addition of data.

Due to the stochastic nature of the deep learning models, we consider the average of 100 models and report the mean and standard deviation of the errors in these models. As a standard practice, we normalize the input features using a MinMaxScaler, which transforms all feature values to be in the range [0,1]. However, we also train our models using a standard scaler, and the results are comparable. We compile the models using the Adam optimizer (Kingma and Ba, 2014) and a Huber loss function, which is less susceptible to outliers (Huber, 1964). We present our main results for a sliding training window where older training data is replaced with newer data during walk-forward validation and include an expanding training window as additional results.

#### 4. Results and discussions

We present the forecast accuracy metrics for the harvested area, yield and farm price of the three commodities using traditional econometric models in Tables 2-4. The naïve benchmark is a low bar, and any model that yields smaller errors than this naïve benchmark will be considered skillful. The USDA baselines have smaller RMSE and MAPE

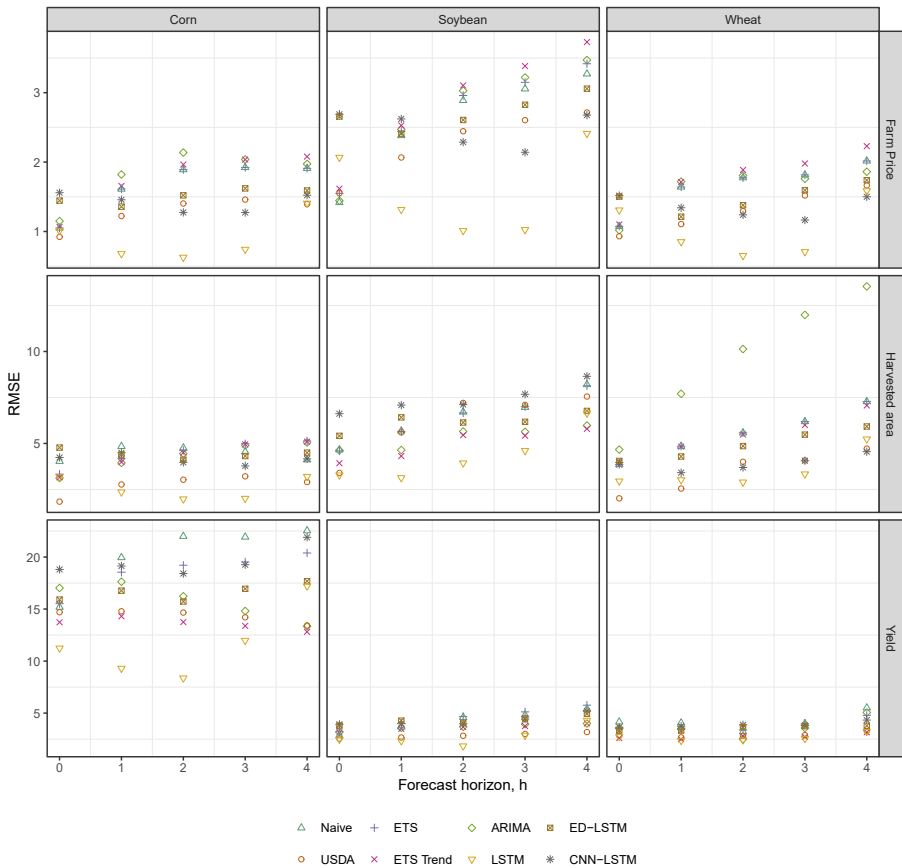


**Figure 5.** Mean absolute percent errors (MAPE) for farm price, harvested area and yield of corn, soybeans and wheat

Source(s): Created by the authors

than those of the naïve benchmark for harvested area, yield and farm price for all three crops across all horizons. Any candidate algorithm to improve the baselines would need to have a couple of desirable properties. At a minimum, it must perform better than the naïve benchmark. Second, it should improve the performance of the USDA baselines, at least for some horizons. In particular, smaller forecast errors at longer horizons would be a good contribution, as the USDA baselines tend to be less informative at longer horizons (Bora *et al.*, 2023). Figures 5 and 6 show the comparison of the forecast errors of all methods for harvested area, yield and farm price, respectively, of the three commodities.

As expected, the RMSEs and MAPEs of the naïve benchmark are very high for most indicators across all horizons. The ETS methods, with or without trend, do not result in a considerable improvement in accuracy and have errors that are comparable to the naïve benchmark for forecasts of harvested area and farm price. For crop yield forecasts, the ETS with trend model performs well. The USDA baseline and the deep neural networks generally show superior skill compared to the naïve benchmarks. We focus the rest of our discussion on the performance of the USDA baseline and the three deep learning models in Tables 5–7.



**Figure 6.** Root mean square errors (RMSE) for farm price, harvested area and yield of corn, soybeans and wheat

Source(s): Created by the authors

Horizon	Method	Corn		Soybeans		Wheat	
		RMSE	MAPE(%)	RMSE	MAPE(%)	RMSE	MAPE(%)
$h = 0$	LSTM	3.176 (0.014)	2.543 (0.019)	3.282 (0.021)	3.011 (0.022)	2.952 (0.022)	5.164 (0.044)
	CNN-LSTM	4.239 (0.238)	3.997 (0.133)	6.618 (0.571)	5.142 (0.3)	3.962 (0.072)	7.131 (0.111)
	ED-LSTM	4.776 (0.029)	5.027 (0.033)	5.416 (0.046)	5.207 (0.055)	4.051 (0.03)	7.306 (0.064)
$h = 1$	LSTM	2.358 (0.016)	2.17 (0.019)	3.137 (0.021)	2.963 (0.023)	3.023 (0.019)	5.595 (0.04)
	CNN-LSTM	4.494 (0.231)	4.18 (0.129)	7.085 (0.583)	5.781 (0.299)	3.412 (0.072)	6.23 (0.117)
	ED-LSTM	4.396 (0.034)	4.434 (0.032)	6.418 (0.046)	6.275 (0.052)	4.292 (0.03)	7.929 (0.059)
$h = 2$	LSTM	1.98 (0.018)	1.887 (0.019)	3.939 (0.02)	3.615 (0.027)	2.898 (0.022)	5.548 (0.046)
	CNN-LSTM	3.972 (0.25)	3.551 (0.144)	7.114 (0.573)	5.996 (0.3)	3.694 (0.068)	7.114 (0.13)
	ED-LSTM	4.115 (0.041)	3.902 (0.044)	6.14 (0.051)	5.928 (0.056)	4.859 (0.032)	9.395 (0.065)
$h = 3$	LSTM	2.011 (0.018)	1.881 (0.018)	4.611 (0.024)	4.39 (0.03)	3.339 (0.023)	7.058 (0.06)
	CNN-LSTM	3.779 (0.242)	3.347 (0.131)	7.676 (0.57)	6.451 (0.301)	4.06 (0.064)	8.168 (0.143)
	ED-LSTM	4.317 (0.037)	3.889 (0.034)	6.176 (0.052)	6.03 (0.058)	5.48 (0.037)	11.16 (0.083)
$h = 4$	LSTM	3.204 (0.025)	2.842 (0.026)	6.645 (0.033)	6.508 (0.036)	5.239 (0.039)	10.355 (0.081)
	CNN-LSTM	4.161 (0.247)	3.703 (0.133)	8.651 (0.555)	7.691 (0.296)	4.552 (0.071)	9.326 (0.169)
	ED-LSTM	4.501 (0.042)	3.992 (0.042)	6.769 (0.056)	6.791 (0.064)	5.923 (0.035)	12.462 (0.08)

Source(s): Created by the authors

**Table 5.**  
Forecast accuracy for  
corn, soybeans, and  
wheat harvested area  
using LSTM

The USDA predicts more accurately the harvested area of crops for the current year compared to the other methods (Tables 2 and 5). At  $h = 0$ , the MAPEs of the USDA baselines for corn, soybean and wheat harvested areas are 1.70, 3.79 and 3.69%, respectively. The MAPEs of USDA baselines of corn harvested area remain low at longer horizons, with LSTM forecasts matching its performance closely for  $h = \{1, 2, 3, 4\}$ . The vanilla LSTM shows better accuracy than all other models for  $h = \{1, 2, 3\}$ , but its MAPEs are high for  $h = \{0, 4\}$ . The USDA baselines do not perform well in predicting the harvested area of corn and soybeans for longer horizons, with large increases in MAPEs between  $h = 0$  and  $h = 4$  for both crops. For  $h = \{2, 3, 4\}$ , both ED-LSTM and CNN-LSTM forecasts have comparable accuracy for soybean and wheat harvested areas, with the CNN-LSTM model performing slightly better.

Horizon	Method	Corn		Soybeans		Wheat	
		RMSE	MAPE(%)	RMSE	MAPE(%)	RMSE	MAPE(%)
<i>h = 0</i>							
	LSTM	11.243 (0.055)	6.118 (0.035)	2.482 (0.012)	4.077 (0.024)	2.742 (0.009)	4.907 (0.021)
	CNN-LSTM	18.808 (0.951)	8.708 (0.248)	3.922 (0.179)	6.339 (0.176)	3.646 (0.181)	5.334 (0.18)
	ED-LSTM	15.911 (0.1)	8.896 (0.065)	3.803 (0.022)	6.58 (0.053)	3.335 (0.021)	5.343 (0.044)
<i>h = 1</i>							
	LSTM	9.296 (0.051)	4.679 (0.031)	2.33 (0.012)	4.128 (0.025)	2.37 (0.01)	3.971 (0.022)
	CNN-LSTM	19.137 (0.961)	9.127 (0.262)	3.985 (0.185)	6.359 (0.185)	3.739 (0.194)	5.433 (0.178)
	ED-LSTM	16.769 (0.108)	9.221 (0.062)	4.271 (0.027)	7.662 (0.051)	3.321 (0.022)	5.373 (0.044)
<i>h = 2</i>							
	LSTM	8.371 (0.055)	4.021 (0.031)	1.839 (0.011)	3.397 (0.023)	2.419 (0.01)	3.993 (0.021)
	CNN-LSTM	18.413 (0.96)	8.462 (0.248)	3.982 (0.2)	6.098 (0.187)	3.872 (0.199)	5.71 (0.195)
	ED-LSTM	15.726 (0.113)	8.554 (0.064)	4.073 (0.028)	7.095 (0.059)	3.662 (0.024)	5.898 (0.043)
<i>h = 3</i>							
	LSTM	11.982 (0.067)	6.236 (0.035)	2.86 (0.012)	5.182 (0.025)	2.559 (0.012)	4.003 (0.025)
	CNN-LSTM	19.263 (0.931)	9.127 (0.247)	4.426 (0.194)	7.034 (0.181)	3.833 (0.197)	5.427 (0.194)
	ED-LSTM	16.948 (0.112)	9.359 (0.074)	4.493 (0.031)	8.14 (0.061)	3.748 (0.025)	6.076 (0.054)
<i>h = 4</i>							
	LSTM	17.211 (0.071)	9.059 (0.046)	4.303 (0.017)	7.902 (0.037)	3.227 (0.015)	5.525 (0.033)
	CNN-LSTM	21.896 (0.909)	10.359 (0.26)	5.197 (0.192)	8.335 (0.183)	4.331 (0.187)	6.41 (0.173)
	ED-LSTM	17.677 (0.147)	9.585 (0.084)	4.934 (0.035)	9.105 (0.068)	3.78 (0.03)	6.27 (0.052)

**Table 6.** Forecast accuracy for corn, soybeans, and wheat yield using LSTM

**Source(s):** Created by the authors

The USDA projections of crop yields are fairly accurate across horizons, with MAPEs around 5% (Table 3). As observed in Figure 1, crop yield has a strong time trend for all crops, making it easier to predict if the trend is correctly identified. The ETS trend model closely matches the performance of the USDA model for all three crops, suggesting the USDA might be using a similar model that includes trends to predict crop yield. The LSTM model has lower MAPEs than those of the USDA baselines for horizons  $h = \{1, 2, 3\}$ , but its accuracy drops sharply at  $h = 4$ .

The deep learning models show noticeable improvement in accuracy while predicting farm prices, which is an indicator that the USDA baselines have much lower accuracy at longer horizons (Tables 4 and 7). At  $h = 0$ , the MAPEs of the USDA baselines are the lowest among all models; however, they increase for longer horizons. Between  $h = 0$  and  $h = 4$ , the

Horizon	Method	Corn		Soybeans		Wheat	
		RMSE	MAPE(%)	RMSE	MAPE(%)	RMSE	MAPE(%)
$h = 0$	LSTM	1.012 (0.003)	17.954 (0.073)	2.068 (0.006)	15.131 (0.054)	1.309 (0.004)	18.934 (0.069)
	CNN-LSTM	1.559 (0.019)	30.1 (0.335)	2.689 (0.062)	20.972 (0.346)	1.517 (0.029)	22.52 (0.344)
	ED-LSTM	1.445 (0.005)	22.78 (0.108)	2.656 (0.01)	18.534 (0.081)	1.503 (0.005)	21.185 (0.101)
$h = 1$	LSTM	0.679 (0.003)	11.708 (0.077)	1.316 (0.007)	10.665 (0.065)	0.852 (0.005)	12.812 (0.077)
	CNN-LSTM	1.457 (0.017)	26.762 (0.35)	2.622 (0.055)	20.312 (0.327)	1.342 (0.028)	19.767 (0.336)
	ED-LSTM	1.358 (0.006)	22.244 (0.137)	2.409 (0.012)	17.739 (0.104)	1.213 (0.006)	17.935 (0.106)
$h = 2$	LSTM	0.628 (0.003)	10.888 (0.079)	1.012 (0.007)	8.334 (0.065)	0.653 (0.004)	9.946 (0.079)
	CNN-LSTM	1.275 (0.017)	24.114 (0.379)	2.288 (0.06)	17.627 (0.341)	1.24 (0.029)	18.616 (0.364)
	ED-LSTM	1.521 (0.007)	26.802 (0.149)	2.607 (0.015)	20.219 (0.117)	1.379 (0.007)	21.889 (0.13)
$h = 3$	LSTM	0.742 (0.004)	13.138 (0.081)	1.026 (0.008)	7.541 (0.064)	0.708 (0.004)	10.222 (0.072)
	CNN-LSTM	1.273 (0.019)	24.583 (0.374)	2.142 (0.061)	16.089 (0.34)	1.165 (0.03)	17.584 (0.366)
	ED-LSTM	1.621 (0.007)	28.608 (0.165)	2.826 (0.016)	21.857 (0.152)	1.594 (0.008)	25.364 (0.146)
$h = 4$	LSTM	1.407 (0.005)	28.528 (0.13)	2.412 (0.011)	17.393 (0.097)	1.586 (0.006)	23.735 (0.107)
	CNN-LSTM	1.523 (0.018)	30.664 (0.366)	2.677 (0.052)	20.688 (0.336)	1.5 (0.027)	22.561 (0.37)
	ED-LSTM	1.591 (0.006)	30.059 (0.181)	3.058 (0.014)	23.567 (0.141)	1.739 (0.008)	28.479 (0.155)

Source(s): Created by the authors

**Table 7.**  
Forecast accuracy for  
corn, soybeans, and  
wheat farm price  
using LSTM

MAPEs of USDA corn price baselines increase from 13.25 to 20.35%. For the same horizons, the MAPEs of the soybean and wheat price baselines increase from 10.34 to 18.55% and from 13.58 to 23.17%, respectively. The price forecasts from the LSTM model show very low MAPEs at horizons  $h = \{1, 2, 3\}$ , but their performance decreases drastically at  $h = 4$ . Given that farmers frequently choose between various crops when planting, being able to reliably predict long-term commodity prices has implications for estimating outlays for federal programs.

The USDA baselines generally perform better than all other methods for the current-year forecasts ( $h = 0$ ). For example, the current-year USDA baselines for harvested areas of corn, soybeans and wheat have MAPEs of 1.7, 3.79 and 3.69%, respectively, which are among the

Horizon	Method	Corn		Soybeans		Wheat	
		RMSE	MAPE(%)	RMSE	MAPE(%)	RMSE	MAPE(%)
<i>h = 0</i>							
	LSTM	2.972 (0.015)	2.404 (0.018)	3.239 (0.026)	2.939 (0.029)	3.1 (0.024)	5.566 (0.043)
	CNN-LSTM	3.473 (0.173)	3.339 (0.095)	4.996 (0.35)	4.278 (0.17)	3.755 (0.065)	6.795 (0.11)
	ED-LSTM	5.302 (0.028)	5.687 (0.032)	6.255 (0.06)	6.097 (0.067)	5.022 (0.043)	9.413 (0.086)
<i>h = 1</i>							
	LSTM	2.348 (0.014)	2.213 (0.016)	3.019 (0.021)	2.818 (0.023)	2.977 (0.019)	5.348 (0.037)
	CNN-LSTM	4.083 (0.156)	3.947 (0.086)	5.705 (0.339)	5.224 (0.175)	3.526 (0.068)	6.388 (0.127)
	ED-LSTM	4.849 (0.036)	4.921 (0.038)	7.297 (0.066)	7.318 (0.074)	4.996 (0.046)	9.234 (0.086)
<i>h = 2</i>							
	LSTM	1.916 (0.015)	1.774 (0.018)	3.809 (0.021)	3.561 (0.024)	2.945 (0.02)	5.559 (0.049)
	CNN-LSTM	3.578 (0.166)	3.308 (0.092)	5.705 (0.335)	5.366 (0.17)	3.655 (0.061)	7.101 (0.124)
	ED-LSTM	4.417 (0.04)	4.208 (0.04)	6.842 (0.069)	6.633 (0.072)	5.459 (0.051)	10.415 (0.099)
<i>h = 3</i>							
	LSTM	1.931 (0.018)	1.74 (0.018)	4.426 (0.026)	4.271 (0.031)	3.378 (0.025)	7.057 (0.058)
	CNN-LSTM	3.43 (0.184)	3.168 (0.098)	6.196 (0.347)	5.819 (0.179)	3.881 (0.066)	7.844 (0.146)
	ED-LSTM	4.368 (0.041)	3.925 (0.042)	6.964 (0.087)	6.773 (0.085)	6.045 (0.055)	12.201 (0.11)
<i>h = 4</i>							
	LSTM	3.119 (0.026)	2.842 (0.031)	6.502 (0.034)	6.492 (0.038)	5.575 (0.047)	10.82 (0.107)
	CNN-LSTM	3.84 (0.163)	3.571 (0.087)	7.097 (0.32)	6.927 (0.171)	4.248 (0.071)	8.614 (0.166)
	ED-LSTM	4.516 (0.041)	3.95 (0.043)	7.559 (0.084)	7.586 (0.08)	6.488 (0.053)	13.639 (0.115)

**Table 8.** Forecast accuracy for corn, soybeans, and wheat harvested area using LSTM (using expanding training window)

**Source(s):** Created by the authors

lowest of all models. The current-year USDA crop yield baselines have low MAPEs as well, though the deep learning methods have comparable performance at  $h = 0$ . Similarly, the MAPEs of the USDA baselines are the lowest for the current year forecasts of farm prices of the three crops (13.25% for corn price, 10.34% for soybean price and 13.58% for wheat price). These findings show that for indicators like yield, for which the USDA baselines are relatively accurate, the deep learning methods do not show much improvement in their accuracy. However, for indicators like farm prices and, to some extent, harvested areas that are more difficult to predict and have high errors, the deep learning methods can be used to improve the accuracy of the USDA baselines. This is not surprising since the USDA enjoys rich market and survey information and expert judgments for making predictions for the current year. On the other hand, all other methods rely solely on past patterns for predicting

Horizon	Method	Corn		Soybeans		Wheat	
		RMSE	MAPE(%)	RMSE	MAPE(%)	RMSE	MAPE(%)
$h = 0$	LSTM	10.894 (0.049)	5.817 (0.032)	2.481 (0.011)	3.966 (0.023)	2.701 (0.011)	4.774 (0.023)
	CNN-LSTM	14.647 (0.479)	7.625 (0.158)	3.228 (0.091)	5.671 (0.093)	3.242 (0.14)	4.922 (0.125)
	ED-LSTM	17.412 (0.136)	9.672 (0.083)	4.305 (0.03)	7.38 (0.059)	3.654 (0.023)	5.903 (0.045)
$h = 1$	LSTM	8.769 (0.047)	4.217 (0.029)	2.251 (0.013)	3.887 (0.024)	2.343 (0.01)	3.986 (0.02)
	CNN-LSTM	16.05 (0.439)	8.385 (0.143)	3.518 (0.093)	6.039 (0.102)	3.302 (0.131)	5.082 (0.129)
	ED-LSTM	18.145 (0.155)	10.047 (0.086)	4.733 (0.039)	8.447 (0.069)	3.677 (0.03)	6.035 (0.052)
$h = 2$	LSTM	7.563 (0.048)	3.504 (0.029)	1.708 (0.012)	3.153 (0.026)	2.42 (0.011)	4.017 (0.021)
	CNN-LSTM	14.96 (0.467)	7.506 (0.144)	3.28 (0.095)	5.513 (0.108)	3.428 (0.14)	5.359 (0.128)
	ED-LSTM	17.023 (0.173)	9.278 (0.091)	4.493 (0.041)	7.773 (0.072)	4.01 (0.034)	6.455 (0.056)
$h = 3$	LSTM	11.159 (0.06)	5.664 (0.037)	2.704 (0.014)	4.88 (0.029)	2.58 (0.013)	4.183 (0.027)
	CNN-LSTM	15.536 (0.489)	7.903 (0.142)	3.604 (0.089)	6.173 (0.106)	3.419 (0.15)	5.079 (0.128)
	ED-LSTM	18.254 (0.177)	10.236 (0.098)	4.899 (0.047)	8.795 (0.084)	4.048 (0.033)	6.519 (0.066)
$h = 4$	LSTM	16.964 (0.079)	8.933 (0.051)	4.282 (0.017)	7.841 (0.04)	3.331 (0.015)	5.705 (0.035)
	CNN-LSTM	18.46 (0.463)	9.313 (0.153)	4.384 (0.083)	7.508 (0.109)	3.79 (0.125)	5.91 (0.135)
	ED-LSTM	19.389 (0.203)	10.666 (0.108)	5.356 (0.049)	9.854 (0.087)	4.047 (0.034)	6.599 (0.064)

Source(s): Created by the authors

**Table 9.**  
Forecast accuracy for  
corn, soybeans, and  
wheat yield using  
LSTM (using  
expanding training  
window)

the values for the current year. The value of survey information and expert judgments diminishes as we move into longer horizons. At longer horizons, the LSTM model performs better than the USDA baselines, or at least matches them. Among the three deep learning methods, the LSTM models show the most accurate performance across indicators over the horizon.

Our study provides a working example to demonstrate that deep learning methods may produce more accurate multi-step commodity forecasts. One way to improve the predictions may be to add more input features to the problem, such as variables for additional crops. As in many high-dimensional, small-sample applications of deep learning (Vabalas *et al.*, 2019; Shen *et al.*, 2022), incorporating additional features may help overcome challenges posed by limited training samples and facilitate better forecast performance. Such

Horizon	Method	Corn		Soybeans		Wheat	
		RMSE	MAPE(%)	RMSE	MAPE(%)	RMSE	MAPE(%)
<i>h = 0</i>							
	LSTM	0.986 (0.003)	16.766 (0.068)	1.972 (0.006)	14.244 (0.047)	1.277 (0.003)	18.28 (0.063)
	CNN-LSTM	1.472 (0.02)	28.238 (0.339)	2.293 (0.044)	18.344 (0.271)	1.457 (0.03)	21.326 (0.334)
	ED-LSTM	1.453 (0.005)	21.602 (0.106)	2.708 (0.01)	18.514 (0.084)	1.513 (0.005)	20.741 (0.095)
<i>h = 1</i>							
	LSTM	0.609 (0.003)	10.407 (0.062)	1.235 (0.006)	9.835 (0.06)	0.777 (0.004)	11.801 (0.064)
	CNN-LSTM	1.483 (0.019)	27.543 (0.382)	2.42 (0.044)	18.985 (0.302)	1.361 (0.032)	19.977 (0.372)
	ED-LSTM	1.365 (0.006)	21.094 (0.132)	2.45 (0.012)	16.944 (0.102)	1.218 (0.007)	17.251 (0.123)
<i>h = 2</i>							
	LSTM	0.575 (0.003)	10.027 (0.078)	0.952 (0.006)	7.822 (0.062)	0.602 (0.003)	9.131 (0.06)
	CNN-LSTM	1.251 (0.02)	23.466 (0.371)	2.061 (0.042)	16.422 (0.332)	1.222 (0.033)	18.121 (0.373)
	ED-LSTM	1.494 (0.007)	24.961 (0.152)	2.564 (0.013)	18.769 (0.118)	1.338 (0.007)	20.601 (0.127)
<i>h = 3</i>							
	LSTM	0.703 (0.003)	11.865 (0.074)	0.957 (0.007)	6.982 (0.055)	0.659 (0.004)	9.615 (0.059)
	CNN-LSTM	1.186 (0.02)	22.463 (0.333)	1.823 (0.047)	14.072 (0.335)	1.08 (0.034)	16.304 (0.401)
	ED-LSTM	1.586 (0.007)	26.881 (0.191)	2.748 (0.015)	20.372 (0.155)	1.534 (0.008)	23.545 (0.137)
<i>h = 4</i>							
	LSTM	1.315 (0.005)	25.846 (0.133)	2.257 (0.011)	15.853 (0.106)	1.479 (0.006)	22.097 (0.115)
	CNN-LSTM	1.433 (0.018)	28.119 (0.373)	2.414 (0.048)	18.307 (0.349)	1.372 (0.031)	20.285 (0.383)
	ED-LSTM	1.549 (0.008)	28.713 (0.2)	2.978 (0.016)	22.298 (0.157)	1.68 (0.01)	26.776 (0.169)

**Table 10.** Forecast accuracy for corn, soybeans, and wheat farm price using LSTM (using expanding training window)

**Source(s):** Created by the authors

high-dimensional networks might need a more complex architecture than the ones used in this study. The three deep learning models used in this study are still relatively simple compared to what a production-ready model with more input features and additional target variables to cover the entire baseline report would entail. For robustness purposes, we provide additional results using an expanding training window in Tables 8–10, with similar forecast errors to the sliding window. In addition, we provide the results using a standard scaler in Tables 11–13, with similar forecast errors to the MinMaxScaler.

### 5. Conclusions

In this study, we developed three deep learning models for predicting harvested area, yield and farm price of three major field crops for five years into the future and compared their

Horizon	Method	Corn		Soybeans		Wheat	
		RMSE	MAPE(%)	RMSE	MAPE(%)	RMSE	MAPE(%)
<i>h = 0</i>							
	LSTM	2.602 (0.039)	2.537 (0.036)	3.036 (0.031)	2.785 (0.031)	3.485 (0.029)	6.343 (0.061)
	CNN-LSTM	4.222 (0.139)	3.816 (0.1)	5.636 (0.139)	5.687 (0.116)	8.806 (0)	17.944 (0.001)
	ED-LSTM	4.809 (0.021)	5.023 (0.024)	6.213 (0.024)	6.425 (0.029)	4.389 (0.019)	8.163 (0.041)
<i>h = 1</i>							
	LSTM	1.827 (0.022)	1.721 (0.023)	2.895 (0.022)	2.786 (0.028)	2.022 (0.019)	3.912 (0.042)
	CNN-LSTM	4.261 (0.134)	3.832 (0.089)	5.568 (0.165)	5.346 (0.123)	9.952 (0)	21.373 (0)
	ED-LSTM	3.797 (0.022)	3.698 (0.028)	7.065 (0.025)	7.165 (0.03)	4.838 (0.02)	9.225 (0.043)
<i>h = 2</i>							
	LSTM	1.435 (0.019)	1.41 (0.02)	3.13 (0.026)	3.082 (0.032)	1.672 (0.021)	3.29 (0.044)
	CNN-LSTM	4.218 (0.143)	3.641 (0.1)	5.527 (0.161)	5.334 (0.12)	11.217 (0)	25.004 (0)
	ED-LSTM	3.282 (0.025)	3.072 (0.026)	6.918 (0.026)	6.809 (0.035)	5.687 (0.02)	11.471 (0.045)
<i>h = 3</i>							
	LSTM	1.802 (0.022)	1.786 (0.025)	3.479 (0.029)	3.315 (0.038)	2.209 (0.025)	4.398 (0.051)
	CNN-LSTM	4.266 (0.153)	3.501 (0.1)	5.789 (0.174)	5.501 (0.13)	12.294 (0)	28.186 (0)
	ED-LSTM	3.209 (0.03)	2.952 (0.03)	7.161 (0.029)	7.104 (0.037)	6.558 (0.021)	13.756 (0.049)
<i>h = 4</i>							
	LSTM	3.484 (0.057)	3.464 (0.057)	5.147 (0.051)	4.827 (0.057)	5.403 (0.066)	10.092 (0.112)
	CNN-LSTM	4.688 (0.149)	4.025 (0.1)	6.702 (0.168)	6.529 (0.112)	13.161 (0)	30.916 (0)
	ED-LSTM	3.471 (0.029)	3.158 (0.031)	7.978 (0.031)	8.304 (0.039)	7.305 (0.023)	15.896 (0.054)

**Table 11.**  
Forecast accuracy for  
corn, soybeans, and  
wheat harvested area  
using LSTM (using  
sliding training  
window and standard  
scaler)

**Source(s):** Created by the authors

performance against a naïve benchmark, ETS with and without trend, ARIMA and USDA baselines. Except for ETS with a trend model for crop yields, the ETS methods do not significantly improve forecast accuracy over the naïve benchmark. The USDA baselines perform well in forecasting crop yield but do not perform as well in forecasting harvested area and farm price, especially at longer horizons. The deep learning models show better accuracy than the USDA baselines in forecasting at longer horizons, most notably in predicting farm prices, where the USDA baselines show poor accuracy. The results suggest that deep learning methods can, at the very least, match the accuracy of USDA baselines for most indicators while offering significant improvement in accuracy for indicators that the USDA baselines do not predict well.

Deep learning methods have shown great promise in forecasting in other fields, but their use in predicting long-term agricultural statistics such as the USDA baselines has been

Horizon	Method	Corn		Soybeans		Wheat	
		RMSE	MAPE(%)	RMSE	MAPE(%)	RMSE	MAPE(%)
<i>h = 0</i>							
	LSTM	13.099 (0.094)	6.968 (0.049)	3.026 (0.021)	5.225 (0.042)	2.331 (0.019)	3.518 (0.041)
	CNN-LSTM	17.209 (0.321)	9.08 (0.146)	4.031 (0.103)	6.84 (0.14)	3.724 (0.085)	5.922 (0.108)
	ED-LSTM	16.524 (0.06)	9.313 (0.039)	4.077 (0.014)	7.142 (0.03)	3.569 (0.012)	5.841 (0.026)
<i>h = 1</i>							
	LSTM	7.122 (0.06)	3.792 (0.034)	2.148 (0.017)	3.695 (0.033)	1.764 (0.014)	2.639 (0.024)
	CNN-LSTM	17.096 (0.409)	8.663 (0.176)	3.959 (0.132)	6.579 (0.164)	3.509 (0.086)	5.413 (0.132)
	ED-LSTM	17.601 (0.067)	9.831 (0.038)	4.576 (0.016)	8.164 (0.032)	3.446 (0.012)	5.48 (0.024)
<i>h = 2</i>							
	LSTM	5.181 (0.054)	2.499 (0.031)	1.538 (0.013)	2.573 (0.023)	1.936 (0.015)	3.002 (0.026)
	CNN-LSTM	14.954 (0.426)	7.068 (0.171)	3.703 (0.143)	5.826 (0.171)	3.7 (0.106)	5.636 (0.144)
	ED-LSTM	17.466 (0.067)	9.591 (0.042)	4.62 (0.017)	8.113 (0.034)	3.93 (0.013)	6.187 (0.026)
<i>h = 3</i>							
	LSTM	6.208 (0.065)	3.057 (0.039)	1.999 (0.018)	3.232 (0.033)	2.139 (0.017)	3.298 (0.034)
	CNN-LSTM	16.024 (0.417)	7.697 (0.167)	4.018 (0.142)	6.445 (0.17)	3.743 (0.104)	5.458 (0.138)
	ED-LSTM	18.795 (0.082)	10.804 (0.051)	5.179 (0.019)	9.265 (0.037)	4.09 (0.015)	6.552 (0.031)
<i>h = 4</i>							
	LSTM	18.227 (0.146)	9.144 (0.076)	4.727 (0.031)	8.376 (0.061)	3.731 (0.026)	6.384 (0.057)
	CNN-LSTM	18.972 (0.393)	9.11 (0.17)	4.781 (0.122)	7.953 (0.162)	4.298 (0.094)	6.726 (0.133)
	ED-LSTM	20.401 (0.083)	11.785 (0.048)	5.789 (0.018)	10.828 (0.039)	4.237 (0.015)	6.958 (0.033)

Source(s): Created by the authors

**Table 12.** Forecast accuracy for corn, soybeans and wheat yield using LSTM (using sliding training window and standard scaler)

limited. Our study shows that efficient deep learning methods can have important implications for baseline models and processes. Since different commodity or country experts work on specific parts of the baseline report and produce inputs for the composite model, deep learning methods can provide insights on which specific parts of the baselines can be improved. The USDA can use the findings and insights from deep learning models to improve the baseline models that produce the initial baselines or the expert adjustment process that produces the final baselines, with the goal to improve the overall accuracy and performance of the baselines. Therefore, the deep learning methods can provide useful insights for the existing baseline projection models.

One of the limitations of this study is that the training sample is relatively small in the terms of years. Deep neural networks often perform better when the training sample is large,

Horizon	Method	Corn		Soybeans		Wheat	
		RMSE	MAPE(%)	RMSE	MAPE(%)	RMSE	MAPE(%)
<i>h</i> = 0	LSTM	0.783 (0.004)	14.415 (0.088)	1.339 (0.007)	9.689 (0.063)	0.944 (0.005)	12.883 (0.096)
	CNN-LSTM	1.497 (0.023)	28.687 (0.361)	2.579 (0.056)	19.695 (0.356)	1.515 (0.034)	22.957 (0.43)
	ED-LSTM	1.37 (0.003)	21.555 (0.093)	2.488 (0.008)	17.772 (0.081)	1.45 (0.004)	20.43 (0.08)
<i>h</i> = 1	LSTM	0.413 (0.002)	6.336 (0.051)	0.806 (0.006)	5.917 (0.05)	0.408 (0.003)	6.386 (0.06)
	CNN-LSTM	1.383 (0.025)	24.689 (0.353)	2.459 (0.061)	17.355 (0.318)	1.35 (0.031)	19.81 (0.376)
	ED-LSTM	1.155 (0.004)	19.077 (0.09)	1.956 (0.009)	14.863 (0.08)	1.077 (0.005)	15.615 (0.09)
<i>h</i> = 2	LSTM	0.364 (0.003)	6.166 (0.062)	0.627 (0.005)	5.04 (0.044)	0.338 (0.003)	5.169 (0.053)
	CNN-LSTM	1.26 (0.028)	20.179 (0.376)	2.334 (0.06)	15.648 (0.325)	1.237 (0.034)	16.685 (0.387)
	ED-LSTM	1.216 (0.005)	23.728 (0.115)	1.971 (0.01)	16.764 (0.096)	1.136 (0.005)	18.52 (0.094)
<i>h</i> = 3	LSTM	0.387 (0.003)	6.845 (0.074)	0.689 (0.007)	5.677 (0.061)	0.41 (0.004)	6.09 (0.061)
	CNN-LSTM	1.329 (0.03)	21.854 (0.373)	2.301 (0.061)	15.553 (0.322)	1.243 (0.033)	16.584 (0.326)
	ED-LSTM	1.329 (0.005)	26.649 (0.126)	2.129 (0.011)	18.316 (0.115)	1.279 (0.005)	21.685 (0.103)
<i>h</i> = 4	LSTM	1.182 (0.006)	23.034 (0.152)	1.97 (0.014)	15.253 (0.102)	1.241 (0.008)	19.306 (0.14)
	CNN-LSTM	1.572 (0.022)	28.339 (0.379)	2.777 (0.048)	19.75 (0.29)	1.548 (0.025)	21.183 (0.35)
	ED-LSTM	1.424 (0.006)	29.435 (0.136)	2.494 (0.011)	21.182 (0.112)	1.492 (0.005)	25.403 (0.109)

Source(s): Created by the authors

**Table 13.**  
Forecast accuracy for  
corn, soybeans and  
wheat farm price using  
LSTM (using sliding  
training window and  
standard scaler)

and at smaller samples, they may lead to overfitting. While our time period is limited, the number of features can be made much larger than in the current study. The USDA baseline report publishes hundreds of indicators, representing a high-dimensional prediction problem where the sample size is much smaller than the number of features. Future research may incorporate more input features and produce forecasts for additional target variables. However, it may require careful feature selection and dimensionality reduction strategies to overcome the challenge of high-dimensionality. Another limitation of the small sample is that we were able to produce forecasts for only five years. A limitation of the deep learning methods is their “black box” nature, making them difficult to explain when compared to economic modeling. However, advances in explainable deep learning methods may be able to address this issue in the future.

## References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. and Zheng, X. (2015), “TensorFlow: large-scale machine learning on heterogeneous systems”, Software, available at: <https://www.tensorflow.org/>
- Bora, S.S., Katchova, A.L. and Kuethe, T.H. (2023), “The accuracy and informativeness of agricultural baselines”, *American Journal of Agricultural Economics*, Vol. 105 No. 4, pp. 1116-1148, doi: [10.1111/ajae.12350](https://doi.org/10.1111/ajae.12350).
- Borovykh, A., Bohte, S. and Oosterlee, C.W. (2019), “Dilated convolutional neural networks for time series forecasting”, *Journal of Computational Finance*, Vol. 22, pp. 73-101, doi: [10.21314/jcf.2019.358](https://doi.org/10.21314/jcf.2019.358).
- Boulmaiz, T., Guermoui, M. and Hamouda, B. (2020), “Impact of training data size on the lstm performances for rainfall-runoff modeling”, *Modeling Earth Systems and Environment*, Vol. 6 No. 4, pp. 2153-2164, doi: [10.1007/s40808-020-00830-w](https://doi.org/10.1007/s40808-020-00830-w).
- Chandio, R. and Katchova, A.L. (2024), “Similarities in the usda international baseline projections”, *Journal of the Agricultural and Applied Economics Association*, doi: [10.1002/jaa2.129](https://doi.org/10.1002/jaa2.129).
- Chen, Y., Kang, Y., Chen, Y. and Wang, Z. (2020), “Probabilistic forecasting with temporal convolutional neural network”, *Neurocomputing*, Vol. 399, pp. 491-501, doi: [10.1016/j.neucom.2020.03.011](https://doi.org/10.1016/j.neucom.2020.03.011), available at: <https://www.sciencedirect.com/science/article/pii/S0925231220303441>
- Chollet, F. (2015), “Keras”, available at: <https://github.com/fchollet/keras>
- Elman, J.L. (1990), “Finding structure in time”, *Cognitive Science*, Vol. 14 No. 2, pp. 179-211, doi: [10.1016/0364-0213\(90\)90002-e](https://doi.org/10.1016/0364-0213(90)90002-e), available at: <https://www.sciencedirect.com/science/article/pii/036402139090002E>
- Fang, X. and Katchova, A.L. (2023), “Evaluating the OECD-FAO and USDA agricultural baseline projections”, *Q Open*, Vol. 3 No. 2, pp. 1-29, doi: [10.1093/qopen/qoad029](https://doi.org/10.1093/qopen/qoad029).
- Hochreiter, S. and Schmidhuber, J. (1997), “Long short-term memory”, *Neural Computation*, Vol. 9 No. 8, pp. 1735-1780, doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- Holt, C.C. (2004), “Forecasting seasonals and trends by exponentially weighted moving averages”, *International Journal of Forecasting*, Vol. 20 No. 1, pp. 5-10, doi: [10.1016/j.ijforecast.2003.09.015](https://doi.org/10.1016/j.ijforecast.2003.09.015), available at: <https://www.sciencedirect.com/science/article/pii/S0169207003001134>
- Huang, Y., Li, J., Hou, W., Zhang, B., Zhang, Y., Li, Y. and Sun, L. (2020), “Improved clustering and deep learning based short-term wind energy forecasting in large-scale wind farms”, *Journal of Renewable and Sustainable Energy*, Vol. 12 No. 6, 066101, doi: [10.1063/5.0016226](https://doi.org/10.1063/5.0016226).
- Huber, P.J. (1964), “Robust estimation of a location parameter”, *The Annals of Mathematical Statistics*, Vol. 35 No. 1, pp. 73-101, doi: [10.1214/aoms/1177703732](https://doi.org/10.1214/aoms/1177703732).
- Hyndman, R.J. and Athanasopoulos, G. (2021), *Forecasting: Principles and Practice*, 3rd ed., available at: <https://otexts.com/fpp3/> (accessed 13 May 2024).
- Katchova, A.L. (2024), “Do revisions improve agricultural baselines?”, *Journal of the Agricultural and Applied Economics Association*, Vol. 3 No. 1, pp. 78-99, doi: [10.1002/jaa2.100](https://doi.org/10.1002/jaa2.100).
- Kim, H.Y. and Won, C.H. (2018), “Forecasting the volatility of stock price index: a hybrid model integrating lstm with multiple garch-type models”, *Expert Systems with Applications*, Vol. 103, pp. 25-37, doi: [10.1016/j.eswa.2018.03.002](https://doi.org/10.1016/j.eswa.2018.03.002).
- Kingma, D.P. and Ba, J. (2014), “Adam: a method for stochastic optimization”, available at: <https://arxiv.org/abs/1412.6980>

- Lara-Benítez, P., Carranza-García, M., Luna-Romera, J.M. and Riquelme, J.C. (2020), “Temporal convolutional networks applied to energy-related time series forecasting”, *Applied Sciences*, Vol. 10 No. 7, p. 2322, doi: [10.3390/app10072322](https://doi.org/10.3390/app10072322), available at: <https://www.mdpi.com/2076-3417/10/7/2322>
- Lara-Benítez, P., Carranza-García, M. and Riquelme, J.C. (2021), “An experimental review on deep learning architectures for time series forecasting”, *International Journal of Neural Systems*, Vol. 31 No. 03, 2130001, PMID: 33588711, doi: [10.1142/S0129065721300011](https://doi.org/10.1142/S0129065721300011).
- Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2018), “The m4 competition: results, findings, conclusion and way forward”, *International Journal of Forecasting*, Vol. 34 No. 4, pp. 802-808, doi: [10.1016/j.ijforecast.2018.06.001](https://doi.org/10.1016/j.ijforecast.2018.06.001).
- Medvedev, N. and Wang, Z. (2022), “Multistep forecast of the implied volatility surface using deep learning”, *Journal of Futures Markets*, Vol. 42 No. 4, pp. 645-667, doi: [10.1002/fut.22302](https://doi.org/10.1002/fut.22302).
- NOAA National Centers for Environmental information (2022), “Climate at a glance: National time series”, available at: <https://www.ncei.noaa.gov/cag/> (accessed 18 July 2022).
- Panigrahi, S. and Behera, H. (2017), “A hybrid ets-ann model for time series forecasting”, *Engineering Applications of Artificial Intelligence*, Vol. 66, pp. 49-59, doi: [10.1016/j.engappai.2017.07.007](https://doi.org/10.1016/j.engappai.2017.07.007), available at: <https://www.sciencedirect.com/science/article/pii/S0952197617301550>
- Roznik, M., Mishra, A.K. and Boyd, M.S. (2023), “Using a machine learning approach and big data to augment waste forecasts: empirical evidence from us corn yield”, *Journal of Forecasting*, Vol. 42 No. 6, pp. 1370-1384, doi: [10.1002/for.2956](https://doi.org/10.1002/for.2956).
- Schmidhuber, J. (2015), “Deep learning in neural networks: an overview”, *Neural Networks*, Vol. 61, pp. 85-117, doi: [10.1016/j.neunet.2014.09.003](https://doi.org/10.1016/j.neunet.2014.09.003), available at: <https://www.sciencedirect.com/science/article/pii/S0893608014002135>
- Seabold, S. and Perktold, J. (2010), “statsmodels: econometric and statistical modeling with python”, *9th Python in Science Conference*.
- Shen, L., Er, M.J. and Yin, Q. (2022), “Classification for high-dimension low-sample size data”, *Pattern Recognition*, Vol. 130, 108828, doi: [10.1016/j.patcog.2022.108828](https://doi.org/10.1016/j.patcog.2022.108828), available at: <https://www.sciencedirect.com/science/article/pii/S0031320322003090>
- Smith, T.G. (2017), “Pmdarima: arima estimators for Python”, available at: <http://www.alkaline-ml.com/pmdarima>
- Smyl, S. (2020), “A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting”, *International Journal of Forecasting*, Vol. 36 No. 1, pp. 75-85, doi: [10.1016/j.ijforecast.2019.03.017](https://doi.org/10.1016/j.ijforecast.2019.03.017), available at: <https://www.sciencedirect.com/science/article/pii/S0169207019301153>
- USDA ERS (2024), “USDA agricultural projections”, available at: <https://usda.library.cornell.edu/concern/publications/qn59q396v?locale=en> (accessed 13 May 2024).
- USDA Foreign Agricultural Service (2024), “Production, supply and distribution database”, Foreign Agricultural Service, Department of Agriculture, available at: <https://data.nal.usda.gov/dataset/production-supply-and-distribution-database>
- USDA National Agricultural Statistics Service (2024), “Nass – quick stats: Ag data commons”, available at: <https://data.nal.usda.gov/dataset/nass-quick-stats> (accessed 13 May 2024).
- USDA Office of Chief Economist (2022), “USDA Agricultural Projections to 2031”, Prepared by the Interagency Agricultural Projections Committee Long-Term Projections Report Oce-2022-1, United states department of agriculture.
- Vabalas, A., Gowen, E., Poliakoff, E. and Casson, A.J. (2019), “Machine learning algorithm validation with a limited sample size”, *PLOS ONE*, Vol. 14 No. 11, pp. 1-20, doi: [10.1371/journal.pone.0224365](https://doi.org/10.1371/journal.pone.0224365).
- Wan, R., Mei, S., Wang, J., Liu, M. and Yang, F. (2019), “Multivariate temporal convolutional network: a deep neural networks approach for multivariate time series forecasting”, *Electronics*, Vol. 8

Wang, Y., Shen, Y., Mao, S., Chen, X. and Zou, H. (2019), "Lasso and lstm integrated temporal model for short-term solar intensity forecasting", *IEEE Internet of Things Journal*, Vol. 6 No. 2, pp. 2933-2944, doi: [10.1109/jiot.2018.2877510](https://doi.org/10.1109/jiot.2018.2877510).

---

**Corresponding author**

Siddhartha S. Bora can be contacted at: [siddhartha.bora@mail.wvu.edu](mailto:siddhartha.bora@mail.wvu.edu)

---

For instructions on how to order reprints of this article, please visit our website:

[www.emeraldgroupublishing.com/licensing/reprints.htm](http://www.emeraldgroupublishing.com/licensing/reprints.htm)

Or contact us for further details: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)