NCCC-134
APPLIED COMMODITY PRICE ANALYSIS, FORECASTING AND MARKET RISK MANAGEMENT

Agricultural Baselines Using Deep Learning

by

Siddhartha Bora and Ani Katchova

Suggested citation format:
Bora, S. and A. Katchova. 2022. "Agricultural Baselines Using Deep Learning." Proceedings of the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. [http://www.farmdoc.illinois.edu/nccc134].

Agricultural Baselines using Deep Learning

Siddhartha S. Bora and Ani L. Katchova*

Paper presented at the NCCC-134 Conference on Applied Commodity Analysis, Forecasting,and Market Risk Management, 2022.

*Professor, and Farm Income Enhancement Chair, Department of Agricultural, Environmental, and Development Economics at The Ohio State University. Address all correspondence to Ani L. Katchova by e-mail at katchova.1@osu.edu.

Agricultural Baselines using Deep Learning

Abstract

Long-term forecasts about commodity market indicators play an important role in informing policy and investment decisions by governments and market participants. The USDA's baseline reports are the primary source of long-term information for the US farm sector, but recent research shows that these projections do not stay informative for longer horizons. We examine whether the accuracy of long-term forecasts can be improved using deep learning methods. We first formulate a supervised learning problem and set benchmarks for forecast accuracy. We then train a set of deep neural networks on a training sample and measure their performance against the benchmark model on a test sample using a walk-forward validation strategy. We find that while the USDA baseline projections perform better for the shorter horizon, the performance of the deep neural networks improves for the longer forecast horizons. The findings may help future revisions of the forecasting process.

Keywords: commodity markets, forecasting, deep learning, USDA baselines

JEL Codes: C53, Q14

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 governments. Similarly, the forecasts of harvested acres and average yield provide information about the production of the commodities for the marketing year and help anticipate the ending stocks. The USDA's World Agricultural Supply and Demand Estimates (WASDE) provides forecasts about important commodities for the current marketing year. However, market participants may require information about the market trends beyond the current marketing year to inform their decisions. For example, forecasts for the next few years can facilitate comparisons of policy alternatives. 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 immensely help in planning the budget. Similarly, forecasts of long-term prices and crop yield 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 may be based on crop prices and yield forecasts for multiple marketing years. Given the importance of such information, the availability of reliable long-term forecasts may play an essential role in informing decisions by a range of participants in the farm sector.

The USDA's baseline projections, published every year in February, are one of the principal sources of long-term projections of the US farm sector. The baselines are produced by a team from 10 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 USDA, the Food and Agricultural Policy Research Institute (FAPRI) 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, Katchova, and Kuethe, 2021). As the evaluation of the baselines showed its limited predictive content, an investigation of alternative methods to improve the long-term projections becomes essential.

This study aims to forecast the harvested area and yield of three major field crops in the US using deep learning methods for the next five years. We use historical data on these variables to produce these forecasts. 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 twelve 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 and USDA baseline. Finally, we implement a suite of deep learning methods to forecast the harvested area and yield of corn, soybeans, and wheat and examine whether they offer any improvement over the traditional methods. We apply some recently developed deep learning methods to predict commodity market indicators, with particular emphasis on long short term memory (LSTM) recurrent neural networks (RNN), and convolutional neural networks (CNN) and their hybrids.

Our study contributes to the literature in several ways. We use some state-of-the-art deep learning methods to improve the baseline projections of commodity prices. 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, Bohte, and Oosterlee, 2018; Wan et al., 2019; Lara-Benítez et al., 2020), their use in predicting long-term agricultural statistics such as the USDA baselines have been limited. This study aims to bridge this gap. Our results suggest that deep learning networks may perform better than the official USDA projections at longer horizons. In particular, bias in the projections of harvested areas of crops may be reduced (Boussios, Skorbiansky, and MacLachlan, 2021). Efficient deep learning methods may have important implications for USDA baseline projection models and processes. The current baseline process is time-consuming, and the adoption of deep learning methods may help in reducing this timeline. In addition, the baseline projection process involves many agencies who work on specific components of the report and produce inputs for the composite model. Deep learning methods have the potential to make the baseline projection process more straightforward, faster, and more accurate. The importance of baseline projections became evident when the pandemic hit the economy, and policymakers needed information deep into the future to plan the recovery process. Deep learning methods with improved predictions may complement the existing baseline models for the baseline projections.

The remainder of this article is organized as follows. The next section describes our methodology. The third section describes the data, followed by results and discussions. The final section contains concluding remarks.

2 Data

Our dataset consists of historical values of harvested acres, farm price, and yield for corn, soybeans, and wheat since 1960. Together, these field crops constitute a significant share of the area under cultivation in the US. The values of these variables are obtained from the NASS Quickstats API (USDA National Agricultural Statistics Service, 2021). The values are averages for the marketing years, which differ by crop. The marketing year for corn begins on September 1 and comprises four quarters. For example, the marketing year 2020/21 for corn and soybeans starts on September 1, 2020, and ends on August 31, 2021. The 2020/21 marketing year for wheat begins on June 1, 2020, and ends on May 31, 2021. Our dataset of baseline consists of USDA agricultural baseline projections from 1997 to 2021. The baseline reports typically include estimates of the previous year(s) and projections for the next ten years. For example, the February 2020 USDA report contains realized estimates for 2018, provisional estimates for the year 2019, and projections for 2020–2029.

[FIGURE 1 ABOUT HERE]

Figure 1 shows the plots of harvested area, yield, and farm price of the three crops. The exact information set available to the committee producing the projections in the early years is difficult to retrieve due to the revisions made to the realized values over time. Often, the projections and estimates are revised long after they are first published. For example, there is no way of accessing the exact data used as the information set of the projections committee when the baseline projections were prepared for 1997. As the organizations involved with the projections process go through changes in personnel and information technology infrastructure over the years, the exact information used to produce the baselines are difficult to ascertain. Therefore, the best we have for the information set of the projections committee is the final estimates for the years 1960-2021 as they are currently available in NASS Quickstats. The committee may have had another version of the same estimates that were later revised to what is available to us today. In that sense, our methods use a revised information set than what the projections committee had back when the projections were made.

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 explain different traditional and deep learning methods used in this study.

3.1 The Prediction Problem

The prediction problem faced by the baseline projections committee is as follows. Let us denote the realized values of commodity indicators in year t by $\boldsymbol{y}_t$. The commodity indicators used in this study are harvested acres, yield, and farm price for corn, soybeans, and wheat. At year t, the projections committee makes projections $\hat{\boldsymbol{y}}_{t+h}$ for horizon $h \in \{0, 1, \ldots, 9\}$ for the next ten years, including year t using its information set I_t at year t. The information set I_t consists of realized values of R periods prior to year t, i.e. $I_t = \{y_{t-R}, y_{t-R+1}, \ldots, y_{t-1}\}$. Our goal is to use deep learning methods to produce projections using the same information set, and examine whether these projections have a superior performance over the USDA baseline projections.

We first transform the prediction problem to a supervised learning problem where a set of input features $\boldsymbol{X}$ are mapped to an output variable $\boldsymbol{y}$. Given our small sample size, we limit our attention to forecasts of upto five years, $h = \{0, \ldots, 4\}$. We also assume that upto five years lagged values are used to make the forecasts. Therefore, for a year t, our input $\boldsymbol{X}_t$ consists of vectors of upto lag five, and $\boldsymbol{y}_t$ consists of vectors of next five years of realized estimates.

3.2 Developing a Test Harness

A test harness ensures that all the deep learning methods used in the study is evaluated using a consistent approach for comparibility. 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 price, harvested area, and yield for corn, soybeans, and wheat between 1960 and 2021. Since we use up to five year lagged features in our deep learning algorithms to produce five year ahead forecasts, this yields a complete dataset of features (X) and output (y) between 1965 and 2017, a total of 53 years. We use the last 12 years of data as our test sample between 2006–2017, and the rest as a training 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. This validation strategy closely follows how the USDA produces baseline projections.

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:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(A_i - F_i)^2} \tag{1}$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{A_i - F_i}{A_i}\right| \tag{2}$$

where A_i and F_i are the realized values and forecasts of commodity farm prices, harvested area, and yield, and N is the sample size of the test or the training sample. For calculating in-sample forecast errors we use the sample size $N = 41$ for the training sample while for out-of-sample errors we use the test sample $N = 12$.

3.3 Benchmarking with Traditional Methods

We first develop a benchmark model to improve upon using deep learning methods. Our first choice is to use a naı"ve no-change forecast. This is a fairly naı"ve benchmark with high errors. Any econometric or deep learning method is expected to perform better than this naı"ve benchmark as the methods are sypposed to add some skill to forecasting. Our second choice to compare is the projections produced by the USDA-ERS in their baseline report. These projections are produced using a mixture of economic/econometric models, survey information, and expert opinions. As mentioned earlier, the exact information set used to produce these projections is difficult to ascertain, and therefore, the comparison with deep learning methods using the current training set may not be fully 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 in 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 consist of more than three layers.

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

3.4.1 Recurrent Neural Networks

Recurrent neural networks (RNN) 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) or solar radiation forecasting (Wang et al., 2019). Moreover, RNNs have done remarkably well at forecasting competitions, such as the recent M4 forecasting competition (Makridakis, Spiliotis, and Assimakopoulos, 2018).

In 1990s, Elman (1990) proposed an early RNN which generalizes feedforward neural networks 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 the Elmann 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, Carranza-García, and Riquelme, 2021).

Long Short-Term Memory (LSTM) networks were proposed in 1997 to address the vanishing and exploding gradients 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 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 (GRU) 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

Convolutional neural networks (CNN) are popular deep learning networks widely 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, Carranza-García, and Riquelme, 2021). A CNN consists of convolution layers, pooling layers, and fully connected layers. The pooling layers lower the spatial dimension of 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 process is more time-efficient than RNNs (Borovykh, Bohte, and Oosterlee, 2018). In addition, different convolutional layers can be stacked together to allow the transformation of raw data (Chen et al., 2020).

A recent trend in time series forecasting using deep learning has been the use of hybrid models. For example, LSTMs can be used with RNNs or CNNs depending upon the applications. Also, deep learning models can be used in conjunction with traditional econometric methods to achieve superior results. The winning entry of the M4 forecasting competition in 2018 used a hybrid ETS-LSTM model (Smyl, 2020). While the exponential smoothing component captured seasonality, the LSTM focused on non-linear trends and cross-learning from related series.

In this study, we use three deep learning models to forecast harvested area and yield of the field crops. The first one is a simple LSTM model. The second one is an encoder-decoder LSTM (LSTM-ED) 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. The third model is an LSTM with a CNN layer at the input. Each model is trained using either harvested area or yield of the three crops, resulting in three input features. The deep learning networks are trained using *keras* (Chollet et al., 2015) and *TensorFlow* (Abadi et al., 2015) libraries in Python. The hyperparameters of the models were chosen using trial and error. The models were trained for 500 training epochs with a batch size of 10. The average of 100 models were considered due to the stochastic nature of the deep learning models.

4 Results and Discussions

We present the forecast accuracy metrics for harvested area and yield of corn, soybeans and wheat for different methods in tables 1 and 2. The naïve benchmark is a low bar, and any model that yields smaller errors will be considered skillful. USDA baselines have smaller RMSE and MAPE than the naive benchmark for harvested area and yield for all three crops across all horizons. Any candidate algorithm to improve the baselines will need to have a couple of desirable properties. At a minimum, it must perform better than the naive benchmark. Second, it should improve the performance of the USDA baselines, at least for some horizons. In particular, smaller errors at longer horizons will be a good contribution, as USDA baselines tend to be less informative at longer horizons (Bora, Katchova, and Kuethe, 2021).

Apart from the naïve benchmark and USDA baseline, we present the forecast errors of the three deep learning algorithms used in our study: a vanilla LSTM (LSTM), an LSTM encoder-decoder (LSTM-ED), and an LSTM with a CNN layer (CNN-LSTM). The figures 2 and 3 shows the comparison of the forecast errors of all methods for harvested area and crop yield of three commodities respectively. All three deep learning methods generally show superior skill compared to the naive benchmark. Among the three deep learning methods, CNN-LSTM shows marginally better performance than the others, especially at longer horizons.

[**TABLE 1 ABOUT HERE**]

[**TABLE 2 ABOUT HERE**]

The USDA baselines perform better than the deep neural networks in shorter horizons. 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, deep neural networks rely solely on past patterns for predicting the values for the current year. The value of survey information and expert judgments diminishes as we move into the longer horizons. Our results show that some deep neural networks outperform the USDA baselines for forecasts of soybean harvested area from $h = 1$, and wheat harvested area at $h = 4$. The USDA baseline does better than the deep learning methods for crop yield forecasts. This is not surprising since USDA baselines for crop yield have been fairly accurate (Bora, Katchova, and Kuethe, 2021).

Another important observation is that the multivariate versions of the deep learning methods perform better than univariate versions, suggesting that they can learn the dependencies between different crops. One way to improve the predictions may be to add more input features to the problem, such as variables for additional crops, macroeconomic conditions, supply and use, and international trade. As in many high-dimensional small sample size applications of deep learning (Vabalas et al., 2019; Shen, Er, and Yin, 2022), incorporating additional features may help overcome challenges posed by limited training samples and facilitate better forecast performance.

[**FIGURE 2 ABOUT HERE**]

[**FIGURE 3 ABOUT HERE**]

5 Conclusions

Deep learning methods have shown great promise in forecasting in other fields, and their use in predicting long-term agricultural statistics such as the USDA baselines have been limited. This study aims to bridge this gap. Efficient deep learning methods can have important implications for USDA baseline projection models and processes. The current baseline process is time-consuming, as it takes more than eight months to produce the baseline report. The adoption of deep learning methods may help in reducing this timeline. In addition, the baseline projection process involves many agencies who work on specific components of the report and produce inputs for the composite model. Deep learning methods have the potential to make the process more straightforward and hence improve transparency. At the minimum, the deep learning methods can act as a complement to the existing models.

One of the limitations of the study is that the training sample is relatively small. Deep neural networks often perform better when the training sample is large, and at smaller samples they may lead to overfitting. However, the number of features for the baseline projections can be much higher than the currently implemented in this study. The problem of forecasting the commodities can a small sample size yet high-dimensional in terms of number of features. Future research may focus on incorporating more variables into the training sample. Another limitation about the small sample is that we have produced forecasts for only five years. Another limitation of the deep learning methods are their “black-box” nature, and when compared to economic modelling they are difficult to explain. However, advances in explainable deep learning methods may be able to address this issue in future.

References

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. 2015. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems." Software available from tensorflow.org.

Bora, S.S., A. Katchova, and T. Kuethe. 2021. "The Accuracy and Informativeness of Agricultural Baselines." Working paper.

Borovykh, A., S. Bohte, and C.W. Oosterlee. 2018. "Dilated convolutional neural networks for time series forecasting." *Journal of Computational Finance*, pp. .

Boussios, D., S.R. Skorbiansky, and M. MacLachlan. 2021. "Evaluating U.S. Department of Agriculture's Long-Term Forecasts for U.S. Harvested Area." Economic Research Report No. 285, Economic Research Service, USDA, Feb.

Chen, Y., Y. Kang, Y. Chen, and Z. Wang. 2020. "Probabilistic forecasting with temporal convolutional neural network." *Neurocomputing* 399:491–501.

Chollet, F., et al. 2015. "Keras."

Elman, J.L. 1990. "Finding structure in time." *Cognitive Science* 14:179–211.

Hochreiter, S., and J. Schmidhuber. 1997. "Long Short-Term Memory." *Neural Computation* 9:1735–1780.

Huang, Y., J. Li, W. Hou, B. Zhang, Y. Zhang, Y. Li, and L. Sun. 2020. "Improved clustering and deep learning based short-term wind energy forecasting in large-scale wind farms." *Journal of Renewable and Sustainable Energy* 12:066101.

Kim, H.Y., and C.H. Won. 2018. "Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models." *Expert Syst. Appl.* 103:25–37.

Lara-Benítez, P., M. Carranza-García, J.M. Luna-Romera, and J.C. Riquelme. 2020. "Temporal Convolutional Networks Applied to Energy-Related Time Series Forecasting." *Applied Sciences* 10.

Lara-Benítez, P., M. Carranza-García, and J.C. Riquelme. 2021. "An Experimental Review on Deep Learning Architectures for Time Series Forecasting." *International Journal of Neural Systems* 31:2130001, PMID: 33588711.

Makridakis, S., E. Spiliotis, and V. Assimakopoulos. 2018. "The M4 Competition: Results, findings, conclusion and way forward." *International Journal of Forecasting* 34:802–808.

Schmidhuber, J. 2015. "Deep learning in neural networks: An overview." *Neural Networks* 61:85–117.

Shen, L., M.J. Er, and Q. Yin. 2022. "Classification for high-dimension low-sample size data." *Pattern Recognition* 130:108828.

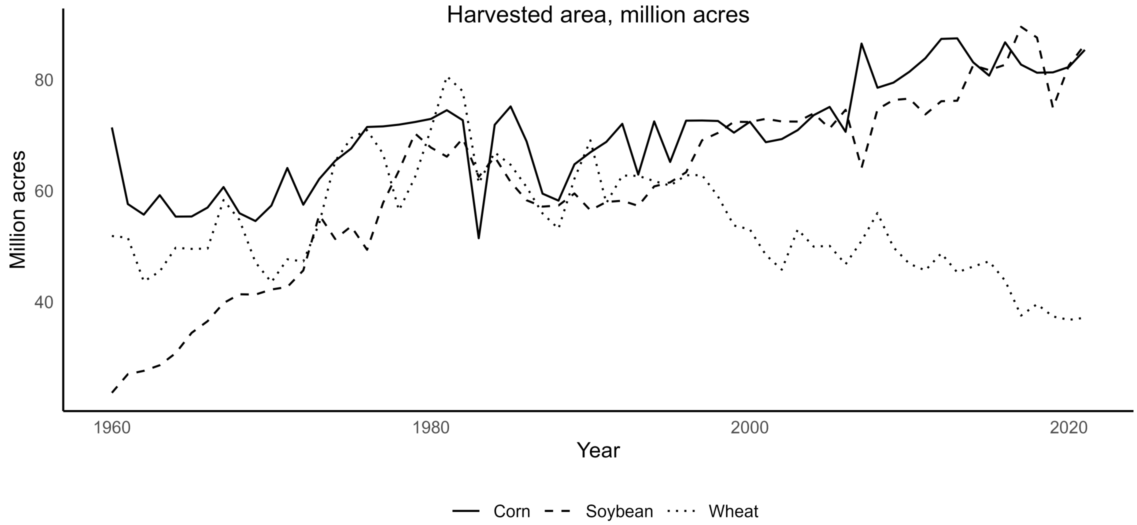
Smyl, S. 2020. "A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting." *International Journal of Forecasting* 36:75–85.

USDA National Agricultural Statistics Service. 2021. "NASS - Quick Stats: Ag Data Commons." Accessed on April 1, 2021.

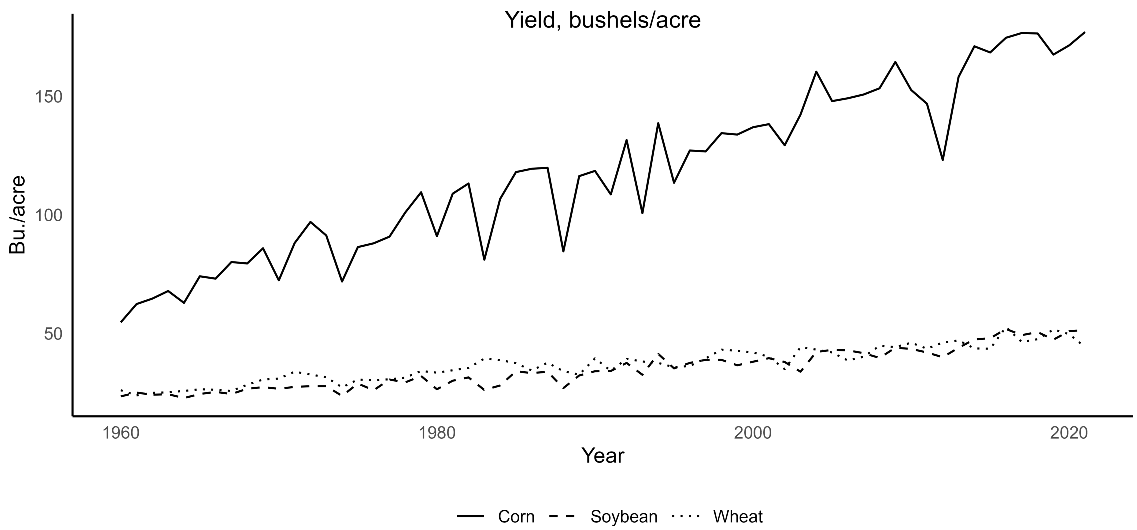
Vabalas, A., E. Gowen, E. Poliakoff, and A.J. Casson. 2019. "Machine learning algorithm validation with a limited sample size." *PLOS ONE* 14:1–20.

Wan, R., S. Mei, J. Wang, M. Liu, and F. Yang. 2019. "Multivariate Temporal Convolutional Network: A Deep Neural Networks Approach for Multivariate Time Series Forecasting." *Electronics* 8.

Wang, Y., Y. Shen, S. Mao, X. Chen, and H. Zou. 2019. "LASSO and LSTM Integrated Temporal Model for Short-Term Solar Intensity Forecasting." *IEEE Internet of Things Journal* 6:2933–2944.



(a) Historical harvested acres (in million acres), 1960-2021



(b) Historical yield (in bushels/acre), 1960-2021

Figure 1: Historical harvested acres and yield of corn, soybeans, and wheat between between 1960 and 2021

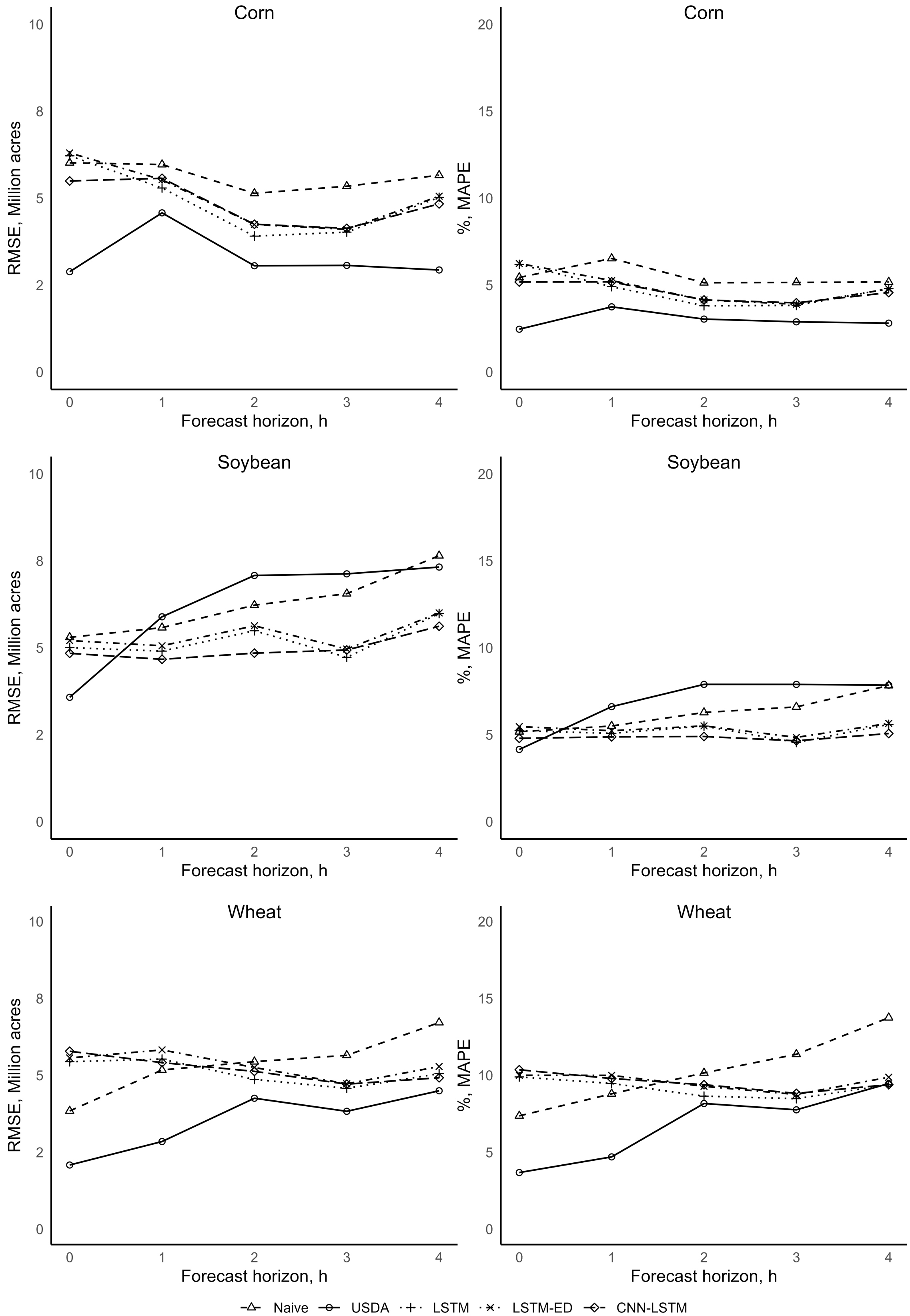


Figure 2: Root mean square errors (RMSE) and Mean Absolute Percent Errors (MAPE) for harvested area of corn, soybeans, and wheat

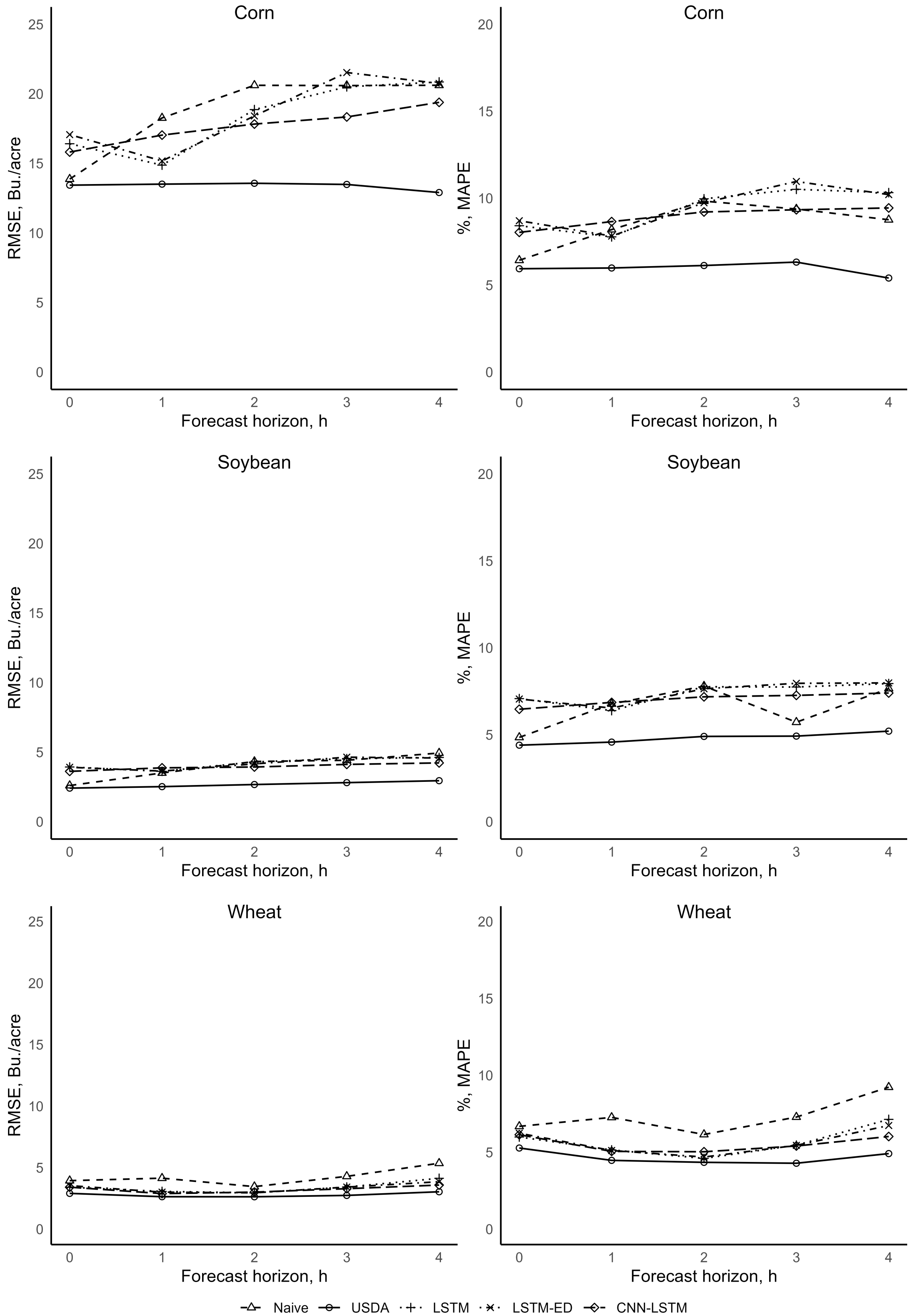


Figure 3: Root mean square errors (RMSE) and Mean Absolute Percent Errors (MAPE) for yield of corn, soybeans, and wheat

Table 1: Forecast accuracy for corn, soybeans and wheat harvested area

Horizon	Method	Corn		Soybeans		Wheat	
		RMSE	MAPE(%)	RMSE	MAPE(%)	RMSE	MAPE(%)
h=0							
	Naive	6.02	5.44	5.30	5.17	3.84	7.37
	USDA	2.89	2.46	3.58	4.15	2.09	3.68
	LSTM	6.22	6.19	5.01	5.23	5.44	9.90
	LSTM-ED	6.30	6.24	5.21	5.47	5.57	10.00
	CNN-LSTM	5.50	5.18	4.84	4.79	5.78	10.36
h=1							
	Naive	5.97	6.53	5.58	5.50	5.18	8.80
	USDA	4.58	3.75	5.89	6.61	2.85	4.70
	LSTM	5.29	4.91	4.90	5.08	5.53	9.47
	LSTM-ED	5.53	5.26	5.06	5.23	5.83	10.00
	CNN-LSTM	5.57	5.19	4.67	4.88	5.41	9.80
h=2							
	Naive	5.14	5.14	6.23	6.28	5.45	10.16
	USDA	3.06	3.05	7.08	7.90	4.26	8.17
	LSTM	3.91	3.81	5.49	5.50	4.87	8.65
	LSTM-ED	4.24	4.15	5.63	5.51	5.25	9.29
	CNN-LSTM	4.25	4.15	4.85	4.90	5.13	9.39
h=3							
	Naive	5.34	5.15	6.56	6.60	5.65	11.37
	USDA	3.07	2.89	7.13	7.89	3.83	7.76
	LSTM	4.02	3.83	4.73	4.56	4.58	8.48
	LSTM-ED	4.11	3.90	4.96	4.85	4.73	8.76
	CNN-LSTM	4.13	3.98	4.94	4.65	4.71	8.83
h=4							
	Naive	5.66	5.18	7.65	7.83	6.72	13.75
	USDA	2.94	2.81	7.32	7.85	4.50	9.47
	LSTM	5.02	4.80	5.99	5.59	5.05	9.39
	LSTM-ED	5.06	4.78	6.00	5.65	5.29	9.88
	CNN-LSTM	4.84	4.57	5.62	5.07	4.92	9.40

Table 2: Forecast accuracy for corn, soybeans and wheat yield

		Corn		Soybeans		Wheat	
Horizon	Method	RMSE	MAPE(%)	RMSE	MAPE(%)	RMSE	MAPE(%)
h=0							
	Naive	13.87	6.43	2.60	4.85	3.94	6.68
	USDA	13.44	5.94	2.41	4.40	2.91	5.28
	LSTM	16.42	8.42	3.93	7.08	3.41	5.99
	LSTM-ED	17.07	8.70	3.91	7.05	3.54	6.24
	CNN-LSTM	15.82	8.03	3.62	6.47	3.41	6.14
h=1							
	Naive	18.27	8.20	3.50	6.77	4.15	7.27
	USDA	13.51	5.99	2.52	4.58	2.64	4.47
	LSTM	14.88	7.78	3.62	6.39	3.07	5.15
	LSTM-ED	15.18	7.82	3.64	6.54	3.01	5.12
	CNN-LSTM	17.04	8.65	3.87	6.86	2.89	5.06
h=2							
	Naive	20.62	9.83	4.32	7.80	3.46	6.16
	USDA	13.57	6.13	2.67	4.90	2.63	4.35
	LSTM	18.87	9.97	4.30	7.76	2.96	4.60
	LSTM-ED	18.41	9.74	4.15	7.64	2.96	4.69
	CNN-LSTM	17.83	9.21	3.93	7.18	3.01	5.03
h=3							
	Naive	20.60	9.38	4.41	5.72	4.29	7.29
	USDA	13.49	6.32	2.81	4.92	2.74	4.29
	LSTM	20.51	10.51	4.53	7.74	3.43	5.46
	LSTM-ED	21.54	10.96	4.64	7.96	3.42	5.45
	CNN-LSTM	18.34	9.33	4.12	7.26	3.29	5.41
h=4							
	Naive	20.62	8.76	4.93	7.66	5.36	9.23
	USDA	12.91	5.41	2.95	5.21	3.04	4.91
	LSTM	20.86	10.32	4.62	7.94	4.13	7.14
	LSTM-ED	20.75	10.21	4.58	7.98	3.86	6.74
	CNN-LSTM	19.39	9.44	4.22	7.40	3.59	6.03