# Weekly Agricultural Commodity Futures Forecasts Incorporating Corn Belt Weather Data

Patrick Travers
MSc. Computing (AI)
Dublin City University
Dublin, Ireland
patrick.travers3@mail.dcu.ie

Sheetal Kumar
MSc. Computing (Data Analytics)
Dublin City University
Dublin, Ireland
sheetal.kumar8@mail.dcu.ie

***Abstract*— Corn and soybean futures are heavily traded products on Chicago Board of Trade (CBOT). State-of-the-art next-day forecasts are produced by Wang et al. with artificial bee colony (ABC) forecast combination. This is despite weak base models in the ensemble. This study investigates whether similar results are achieved using only the strongest base models. Limitations of ABC are explored by comparison to a new combination technique: error-informed discretised search (EIDS). Wang et al.'s technique is extended to weekly forecast horizons. Weekly forecasts are further augmented by Corn Belt weather variables. Price series are smoothed using empirical mode decomposition (EMD) and variational mode decomposition (VMD). The series are modelled with ARIMA and RNN. Day-ahead and week-ahead forecasts are generated. Daily forecasts are less accurate than the original study. EIDS produces better forecast combinations in 100 milliseconds' computation than ABC does in 15 minutes. Week-ahead forecasting achieves MAPE 1.218%: a strong proof-of-concept. The effect of weather variables is mixed; the best weekly model ignores weather features.**

***Keywords—Agricultural commodities, CBOT, artificial bee colony, error-informed discretised search, corn belt, weather data***

All code relating to this study is available at https://gitlab.computing.dcu.ie/kumars8/2022-mcm-TERMPLATE

## I. Introduction

### A. Corn in USA

The United States is the largest producer of corn in the world by production volume with an approximate 31% share in 2020, followed by China (23%) and Brazil (9%). A typical annual nationwide corn yield is approximately 14 to 15 billion bushels [1]. Depending on price conditions, this volume of corn is approximately \$50 to \$80 billion dollars' worth in today's currency. Even in an advanced economy as USA, the scale of corn production and variability in its price mean it is a closely watched part of the financial system from many perspectives. There are just three markets consuming over 90% of US corn production. 10 to 20 percent of the US corn crop is sold abroad, making it the world's largest exporter [2]. Of domestically consumed corn, almost half goes towards feeding livestock. Most of the remaining produce is put towards ethanol production.

Price dynamics of primary economic outputs such as agricultural commodities, typically have knock-on effects throughout other economic activities. This is because commodities as raw materials are an input to various other industries. Livestock farmers using corn as feed are vulnerable to corn price fluctuations. So too are producers of ethanol where a starch-rich base material is essential. Considering these two primary areas of corn consumption, it is not difficult to imagine that fluctuations in corn prices have knock-on effects to meat and food prices, as well as gasoline prices where ethanol is a common additive. Furthermore, with USA representing the largest exporter of corn globally, the effects of price movements are felt well beyond its borders. The price and subsequent availability of US corn around the world is of interest to those influencing economic policies averting global food shortage risks.

### B. Futures Versus Spot Trading

Since corn price is prone to fluctuations, most trading is not carried out on-the-spot, rather via futures contracts mediated by a commodity exchange. Futures contracts allow a buyer and seller to agree a fixed price for a pre-specified quantity of a commodity to be exchanged at a specific future date. The benefit of trading via futures for both buyers and sellers is that it mitigates the risk of price swings, allowing for greater certainty around budgeting measures. Chicago Board of Trade (CBOT) represents the most significant commodity exchange in USA, where corn futures are one of the most heavily traded products. A CBOT contract's worth of corn represents a significant investment, entailing 5000 bushels, or just over 127 metric tons of raw produce.

A contract can be agreed from a couple of years in advance, up to the month before the contract's due date. While each contract is subject to its own dynamics, the more common approach to modelling futures is to examine continuous futures contracts, which represent a weighted aggregation of all currently available futures contracts on the exchange. This is more attractive from a modelling perspective because it affords a continuous series of sufficient length to allow patterns in price dynamics to emerge.

### C. The Corn Crop

Throughout the marketing year of corn there are many interacting factors influencing its market price. For corn producers, a considerable portion of their required inputs are variable in price. For example, fertiliser constitutes approximately 20% of the typical farm's costs, and is quite volatile in price. Pesticides can also vary considerably depending on market conditions, contributing normally over 10% of producers' costs. Oil-related inputs including fuel, lubricants and electricity constitute normally over 10% of costs [3]. Suffice to say, as is often stated in the literature, prices of agricultural commodities are non-linear. The health of a given year's crop plays a large part in determining future supply. USDA reports that temperature and moisture, as well as their interactions at critical moments during the corn cycle are critical to a healthy crop and high yields [4,5,6]. The present study supposes that of all nations producing a globally consequential portion of a crop, US corn is most suitable for analysing weather-impacts on price movements. As illustrated in Fig. 1, USA's corn output comes from a concentrated region where weather conditions during growing season have direct implications for crop yield, and consequently, futures prices as traded on CBOT.

### D. *The Scope of this Study*

The present study examines the cases of corn and soybean prices. In order to facilitate discussion of sufficient depth, corn is the primary focus of the study. However, soybean is a very similar crop, with similar geographical distribution, growth cycle, supply chain and price determinants. Indeed, such is their similarity that they exhibit the strongest co-movement of any two agricultural commodities in USA [7]. Therefore, the literature often analyses the cases of corn and soybeans together. Much of what has been stated above regarding corn applies to soybean just the same. In keeping with the literature, the present study will examine the case of soybean alongside the more economically consequential commodity of corn.

The following review gives a brief account of yield as a proxy for supply forecasting, followed by a more in-depth view of price forecasting literature. Section III outlines the current investigation's methodologies. Results of analyses are given in IV. The significance of findings are discussed in V. Section VI concludes.

## II. Literature Review

### A. *Weather and Climate*

Apart from the drier regions in its western reaches where irrigation systems are common, the US corn belt is largely a rain-fed crop system [5]. It is conventional wisdom among producers that weather can have an impact on a year's corn output. As part of a corn yield forecasting investigation, [8] has attempted to determine the extent to which this supposition holds. The study employed a heterogenous ensemble comprising linear regression, lasso regression, random forest and LightGBM models to predict county-level corn yields for the states of Iowa, Illinois and Indiana. Combining these models' averages was found to achieve 9.5% relative root mean squared error for annual corn yield by county. The dataset represented a host of weather and planting progress variables. In a post-hoc analysis of feature importance by time of year, it was found that daily precipitation during late May and early- to mid-October are the most useful features in this context. These dates coincide with sensitive planting and harvesting periods of the corn belt cycle.

Somewhat contradictory results of feature significance are reported in [9]. This work modelled county-

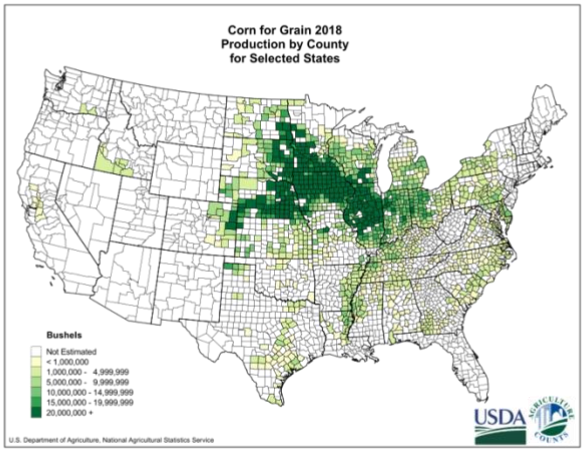


Fig. 1. Distribution of corn growth in USA in 2018 according to USDA.

average yields by state with normalised root mean squared error of 4.9% for corn and 5.5% for soybean. The predictions were achieved using a support vector machine (SVM), and the dataset used a modest selection of county-level weather-related variables: daily average temperature, temperature range, and total precipitation. That such results can be achieved with only these features indicates the significant impact of weather on corn and soybean growth. Although post-hoc analyses across different models must be interpreted with caution, the comparison between important variables here with those from [8] are interesting. The study finds that for both corn and soybean, temperature range for July and August are the most important features in general across all models. For corn, average temperature during that period was highly significant also. Conversely to [8] precipitation in June and July was a significant predictor for both crops – more so than precipitation during planting and harvesting. The study found that the weather variables generated the best predictions with the SVM when aggregated to weekly averages, rather than at daily granularity.

[10] takes the specific cases of corn, soybeans, and wheat, and attempts to determine the most salient variables for forecasting price movements at 1-month to 12-month horizons. Using ARCH and other autoregressive techniques, it is suggested that volatility in corn prices are best predicted by world stock-to-use ratio. This variable expresses the ratio between the year's ending stock (recall that for most industrial uses corn is virtually non-perishable and readily storable) as a proportion of the amount of corn used worldwide that year. Stock-to-use ratio exhibits an inverse relationship with corn price because it encapsulates current supply-demand circumstances. Similarly to [11], the paper finds that bond credit spreads and the stock market conditions are viable volatility predictors, with the added detail that this is best observed at longer forecast horizons. In the case of soybeans, these same financial variables also predict price volatility to a statistically significant degree, but function best at shorter forecast horizons. Also as in the case of [11], no place is found for climatic indicators such as El Nino/La Nina as viable predictors of price volatility.

The preceding observations are indicative of two trends within commodity forecasting literature. Supplyfocused yield forecasting studies have tended to leverage local weather conditions with some success. On the other hand, price forecasts have sided with slow-moving climatic variables such as El Niño/La Niña indicators, to little avail. Naturally this rule has its exceptions. An interesting area of commodity price prediction lately comes in the form of textually-informed models. [12] has leveraged news headlines to extract information relating to a range of a range of factor categories such as futures markets, international markets, macroeconomics, and weather forecasts. Much of the information obtained this way is similar to variables used in [8-11] above. However, the authors assert the distinct benefit of this approach is that it allows the model to gain access to weather forecasts as well as information on qualitative and administrative matters such as government subsidies, government reports and international trade policies. The study feeds textually-informed features into support vector regressor, random forest, and neural network models to predict Dalian Commodity Exchange (DCE) soybean futures prices. At 1-day and 5-day horizons, the text-based model fails to outperform ARIMA. However, at 60-day and 180-day horizons, the model does outperform the ARIMA baseline. Although the study is implemented on a static, scraped dataset, with little indication as to how the solution could be productionised, it is an example of weather conditions leveraged for price forecasting rather than relying on the lumbering influence of climatic indicators. Features with the potential to arrive to a forecasting model with this velocity are desirable, if not essential, for many applications in the practical setting.

### B. *Other Multivariate Approaches*

Price forecasting need not always draw features from disparate sources; some useful information can be found closer to home. [13] has achieved mean absolute percentage error (MAPE) of 3.49% for 1-month ahead forecasting of agricultural commodity spot prices. The study examines spot prices of 19 price series provided by China Animal Agriculture Association. Beginning with twenty-nine distinct features of time series such as linearity, curvature, lumpiness and seasonality, these features as applied to agricultural commodities are reduced to 7 categories of features using principal component analysis. The study's innovation is in the development of a model selection framework where the model for a given series is selected based on the constitution of that series, as described by the 7 feature categories. Forecast horizon is found to be one of the most broadly useful features across commodities, playing a part in the strongest 1-month-ahead model.

Another approach to extracting predictive power from commodity price series is to train and forecast more than one commodity simultaneously. [14] employs an LSTNet model to this end. On a set of 12 commodities from DCE, mean relative absolute error of 0.026 is achieved at a 3-day-ahead horizon. This low error is tempered by the fact that the study draws on some of the more statically priced DCE commodities; the same technique applied to CBOT corn and soybean, for example, seems unlikely to achieve similar results. The approach in general though may hold potential; a suitable agricultural commodity model, as LSTNet seems in this case, could be informed not by exogenous variables, but by training and forecasting price movements among a host of agricultural commodities concurrently.

### C. *Univariate forecasts*

There are studies which do not draw on exogenous variables in any way. Univariate time series forecasting is founded on the assumption that historical values hold sufficient information for predicting future values. As facile an assumption as this may seem, univariate time series prediction enjoys its share of successes, including in agricultural commodity price forecasting. Using a combination of vector error-correction model and multiple-output support vector regressor (VECM-MSVR), [15] has achieved MAPE of 2.47% in predicting the range (not value) of next-day price movements of corn traded on DCE. The study forecasts next-day minima and maxima, with final MAPE representing the mean error of upper and lower bound errors. It is an interesting approach and contribution in several ways. The general approach of forecasting minima and maxima is a worthwhile take on the problem, since it overcomes many of the limitations of ARCH-oriented volatility studies and point-estimate regression analyses; it combines the specificity of point-estimates while maintaining a measure of price variability. The drawback is that extrema are intrinsically more difficult to forecast than measures of central tendency, hence the somewhat sluggish next-day MAPE of 2.47%. MAPE errors for 3-day and 5-day horizons are 3.06% and 3.33% respectively. However, that these results were achieved without reference to exogenous variables stands as a reminder that univariate analyses have a place within the literature on agricultural commodity price movements.

An impressive result of 0.57% MAPE was achieved by [16] in forecasting next-day CBOT corn prices in a univariate fashion. The same error (0.57% MAPE) was also achieved for soybean. Using VMD for noise-reduction, a feed-forward neural network modelled the remaining signal. This approach was trialled for three other decomposition techniques: EMD, wavelet packet transform (WPT) and intrinsic time-scale decomposition (ITD). Of these techniques, VMD was found to furnish the lowest forecast error. In all cases, hyperparameters were found using particle swarm optimisation. This is a worthwhile result which perhaps could have been further extended. The various decomposition-NN combinations could have been combined into an ensemble to create a single forecasting mechanism leveraging the variance of different base models to make the most generalisable forecasts.

### D. *Wang et al.'s Contribution*

[17] has gone a distance towards filling some of the gaps by [16], and has produced a state-of-the-art (0.49% MAPE) forecast technique for next-day CBOT corn and soybean futures. There are three steps involved in achieving this result. First, price data for corn and soybean is denoised using VMD, EMD, and singular spectral analysis (SSA) techniques. Then, on each decomposed data series, ARIMA, SVR, RNN, GRU, LSTM models are fit. The resulting

models are combined with mean, median, trimmed mean and artificial bee colony (ABC) combination schemes. The best results are achieved when all fifteen base models (the product of five models fit to three different decompositions) are combined with the ABC scheme.

ABC optimisation is a genetic algorithm developed by the authors in a prior work [18], built upon the original inception by [19]. It is a means of searching a parameter space of arbitrarily large dimensions. It has been applied to agricultural commodity studies in various guises, including as a hyperparameter search strategy for modelling the dynamics of Colombian coffee prices [20]. The basic idea of ABC's functioning is that a 'swarm' of random solutions is generated, and each solution's fitness quantified (each solution analogous to a bee). A global variable keeps track of the best solution found to date. For a pre-specified number of iterations, each bee takes a random step within the search space. If the change increases the bee's fitness, it is kept. Bees converge around promising solutions by moving in the search space towards solutions already demonstrating high fitness. If a bee has failed to find any improvements after a specified number of failed attempts, the solution is abandoned in favour of a new, randomly generated solution. ABC in [17] was found at all times to outperform the naïve forecast combination schemes of mean, trimmed mean, and median.

Despite the successes of the study in generating highly accurate next-day forecasts, there exist a number of gaps in its contributions. Firstly, such is the ground covered by the study in the number of models built and decomposition techniques applied, there is insufficient detail on any particular model to allow reproducible results. For example, even though ARIMA is found to be the best-performing base model for both corn and soybeans, the paper does not mention the order of the models used. Indeed, the information given is that the order was chosen based on AIC and BIC. How AIC and BIC might have been combined to inform a decision, and what was the decided order remains unclear. The description of neural network models state that a single hidden layer of 32 nodes was used, with an Adam optimiser and a learning rate of 0.001. Nothing is said of key details such as input size. Details surrounding the ABC hyperparameters are lacking. Final combination weights decided by the ABC algorithm are not specified.

The study reports considerable differences in performance among different decompositions and models. Models fit to SSA perform nowhere near as well as EMD and VMD. The strongest base models are ARIMA and RNN. These observations hold for both corn and soybean. With the few details provided by the paper, it remains possible that the only strength of ABC above naïve forecast combinations lies in its ability to nullify the contributions of weaker models, applying a kind of 'pruning' effect to the ensemble. Without knowledge of the final ABC weights, it is impossible to know for certain. Furthermore, in relation to ABC, it is noted in [18] that ABC is often criticised for its slow convergence to global optima. If ABC converges (itself is no guarantee) it is said that it converges only very slowly, wasting time and memory resources exploring unfruitful avenues that are eventually abandoned. [17] offers no details regarding the resources used in its weight search.

All of the above considered, as well as the fact that the study has reported strong results, the present study finds its initial area of enquiry. This study wishes to examine whether the strong performance of ABC forecast combination is maintained when the ensemble consists of only the strongest decomposition-model combinations - namely, EMD-ARIMA, VMD-ARIMA, EMD-RNN and VMD-RNN. Since final model weights are not reported, it remains unknown why ABC-combined ensemble should perform so well. If ABC simply carries out a 'pruning' function, diminishing the contribution of weaker base models, perhaps a 'pruned' ensemble taking only the strongest individual base models would perform just as well as Wang's somewhat unwieldy 15-base-model ensemble.

**RQ1.** How does a 'pruned' ensemble of EMD-ARIMA, VMD-ARIMA, EMD-RNN, VMD-RNN perform, compared to Wang's 15-model ensemble under various combination schemes?

Within this investigation, examining ensembles with different numbers of base models will help determine whether ABC is converging to global optima. If an ensemble is outperformed by another ensemble using a subset of its base models, the larger ensemble has provably failed to converge to the global optimum. Noting the above criticisms of ABC inefficiency, the present study will investigate another related matter, left unanswered from the initial study:

**RQ2.** What are the time requirements for convergence of Wang's LL-ABC and can they be improved?

Another gap in the approach of [17] is that it's contribution ceases at next-day forecasting. For price forecasts to be of use for producers, consumers, and policymakers such a short horizon is perhaps impractical. Agricultural commodity futures contracts are a significant commitment. Practicalities around the delivery and storage of a contract's worth of produce requires careful management. In the case of policymakers, lengthier forecast horizons is even more important still. Government policymaking would be ill-advised to attempt to react to daily price fluctuations, if it were even possible. With this, forecasts at longer intervals should be better placed to fulfil the broader aims of agricultural commodity price forecasting.

A further natural extension is to investigate whether forecasts can be improved by leveraging extraneous variables. Given the pervasiveness in the literature of local weather rather than climatic features as useful contributory variables, the present study will investigate whether weather data can further improve [17]'s initial approach.

**RQ3.** How does the ensemble perform at weekly horizons incorporating corn belt weather data

## III. Methodology

### A. Hardware and Software

All code was written in Python 3.7.13 and executed in a Google Colaboratory environment. Processing for all code apart from neural network training was executed on 2-core Xeon 2.2GHz processor with 13GB RAM. Neural network training was carried out on Tesla K80 with 2496 CUDA cores and 12GB VRAM.

### B. *Datasets*

With the aim of the present paper to interrogate, reproduce, and extend the work of [17], the study employs the same dataset as used by that work. CBOT continuous futures contracts price data for corn and soybean from 1974 to 2017 is available in Appendix B of [17]. The authors report that the data is originally drawn from the website of CME.

It is worth making explicit that more data than 1974-2017 is available at source. Indeed, the relevant price series is available stretching back to 1960. The present study wishes to point out the motivation behind beginning the training set specifically in 1974. 1973 saw a distinct level change in corn futures prices, coinciding with the onset of the 1973-1976 recession in USA. Economic difficulties throughout USA and much of the world at that time reshuffled the landscape of agricultural commodity prices. This is why beginning training in 1974 is appropriate. More data is not necessarily always better when modelling in a complex non-linear context where even the units of the dependent variable – US dollars – are themselves subject to shifting in value over time.

Weather data for three weather stations across the state of Iowa (Cherokee, Mason, and Le Claire) is freely available from National Centers for Environmental Information [21]. Data from these stations was chosen due to their geographical spread across the state of Iowa and for their record completeness. Daily maximum temperature, daily minimum temperature (both degrees Celsius), and total daily precipitation (millimetres) are available with upwards of 99.8% completeness between 1974 and 2017. The one exceptional field is Le Claire's precipitation which is 97.8% complete between these dates.

### C. *Pre-processing*

*a) Data splits:* All data were split into training, validation, and test sets (80%, 10% and 10% respectively) to give the same sets as used in [17] (Fig. 2).

*b) Weekly Downsampling:* Since an aim of the current study is in extending predictive efforts from daily to weekly predictions, weekly versions of corn and soybean price series were made. Weekly series were created by calculating weekly mean of each daily price series. The following decomposition techniques outlined were applied to corn and soybean, daily and weekly price series.

*c) Empirical Mode Decomposition:* EMD was implemented using the Python emd package. EMD operates in a number of steps. First, all local maxima are identified and connected with a cubic spline. Then, the same is done with all local minima. The average of this 'upper envelope' and 'lower envelope' is then found to give an 'average envelope'. The 'noise' component of the series is identified by subtracting the average envelope from the original series. The original series is smoothed by subtracting this noise component. This process can be carried out recursively, removing more and more noise components until all that remains is the underlying trend. In this study only one noise component is removed to smooth the series.

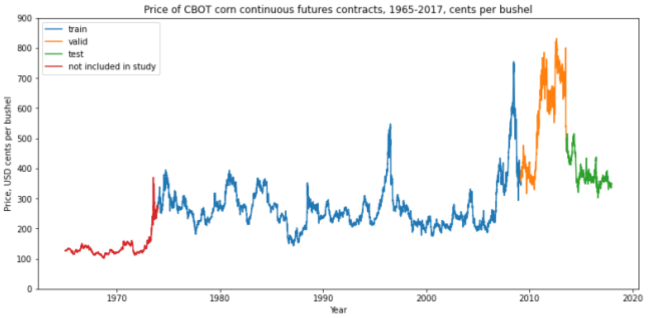


Fig. 2. Training, validation, and test sets for corn futures. These are identical splits to those used in [17].

Curiously, [17] reports that the daily price series can be decomposed to 11 different components. Using Python's emd package the present study could find only 9 decomposed components. The present study's daily EMD series is the sum of 8 components, with the highest frequency component discarded as noise. The weekly price series are the sum of 6 components, with the highest frequency series discarded as noise. A sample of results are displayed in Fig. 3.

*d) Variational mode decomposition:* VMD was implemented using the Python vmdpy package. The VMD technique was originally proposed by [22]. The process operates in a number of steps. First, the signal is decomposed into a specified number of modes. Wiener filtering is used to update the modes before centre frequencies are updated as the centre of each mode's power spectrum. The final step is in updating the process's Lagrangian multiplier in the style of Alternating Direction Method of Multipliers (ADMM).

Recommended VMD parameters [23] were found to work well in this case, along with the added specification borrowed from [17] that VMD should be decomposed using the same number of modes as were found with EMD. These specifications are as follows: alpha = 2000; tau = 0; DC = 0; init = 1; tol = 1e-7. K was set to 9 for daily series and 7 for weekly series.

*e) Weather pre-processing and feature engineering:* Missing values (0.32% of weather data) for all features for each weather station were imputed with the last valid entry in its series. This is an appropriate course given that missing values exhibit no signs of NMAR (not missing at random) criteria. Maximum and minimum daily temperatures were used to compute daily mid-range temperature, previously identified as a useful feature for corn yield modelling [9]. Data for all three stations was averaged to give what can be

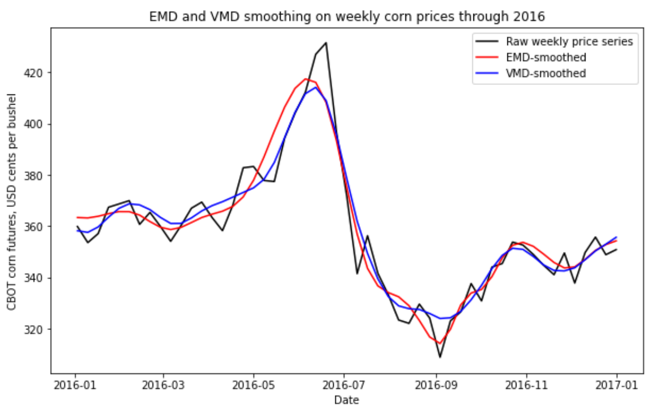


Fig. 3. The results of EMD and VMD smoothing for corn futures price series through 2016.

called mean corn belt figures. corn belt figures were then downsampled to weekly averages. Each datapoint was then subtracted by the average historical value for a week in that month. This results in 'normalising' weather figures across months, so that a 'dry June' and a 'dry October' will be of comparable value, even though those months typically experience different levels of rainfall. In order for the dataset to be appropriate for a recurrent neural network, each feature was scaled to have a mean of 0 and a standard deviation of 1. Given that different states of weather affect crops differently depending on the part of the crop cycle they occur, month number was added as a feature. All variables between November and March inclusive were set to zero, as this is outside the corn growth cycle, an idea borrowed from [8] where the technique proved effective. Month number was then transformed so that non-zero values were scaled equi-distantly between 0 and 1.

### D. Error Metrics Used

*a) Mean squared error:* Mean-squared error (MSE) is used as a loss function for neural network training.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

*b) Mean absolute percentage error:* Mean absolute percentage error (MAPE) is used as a means of forecast evaluation.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{|y_i|}$$

### E. Modelling

*a) ARIMA development:* ARIMA is a long-standing method of time series prediction. The model can be conceived oas a multiple linear regression, drawing on autoregressive and moving-average components at specific lags. In this study ARIMA is implemented using Python's pmdarima package.

A key assumption of ARIMA is stationarity of the time series data. The present study checks stationarity and chooses model order on a combined training and validation set, excluding only the test set. The Augmented Dickey-Fuller test available in Python's statsmodels package is used to test stationarity. In this study, AIC was used to choose the most appropriate model order. The results of this process are given in Table 1 below.

In the case of VMD weekly, both differenced autocorrelation and partial auto-correlation functions exhibit recurring patterns, indicating a seasonal component may be necessary. Various models were trialled, with emphasis on auto-regressive and moving-average components as appropriate. Ultimately, the model with the lowest AIC was found at lag 5, with model order (3,1,2)(2,0,2).

*b) RNN development:* In a regular feed-forward neural network, training data is fed forwards through the network before the back-propagation process updates weights back through the model. A recurrent neural network (RNN) differentiates itself from a feed-forward network by the inclusion of one or more hidden layers which not only output to the subsequent layer, but also serve as input to that layer for the next training instance. The present study implements RNN models using the Python's Keras deep learning framework.

For daily series, a number of hyperparameter assumptions were made where [17] has omitted detail. For one thing, input shape was decided at 60 units simply because this is mentioned as the window size for SSA decomposition. The single hidden layer was 32 nodes as specified in [17]. An Adam optimiser was used, and a learning rate of 0.001. It was arbitrarily assumed that tanh activation and mse loss functions were used. Furthermore, it is assumed the price series was scaled to between 0 and 1 for neural network training, as otherwise the process occurs only very slowly.

Hyperparameters and architecture for weekly models were chosen by experimentation, taking the model with lowest validation loss. Model architectures are given in Table 2. Relu activation and mse loss functions were used. Again, an Adam optimiser was used, with a learning rate of 0.001. As with weekly data, the price series was scaled to a range between 0 and 1 for training.

Daily models were trained for four epochs before the best model was chosen. Weekly models were given 50 epochs for training. With any longer training times, signs of overfitting began to appear with rising validation losses.

When it comes to RQ3, investigating performance at weekly horizons with weather data, RNN was used exclusively for this task. This is because for the purposes of bringing a weather-related features to the dataset, fitting the model to a smoothed dataset may be of little benefit, given that price fluctuations caused by weather events may be cancelled out in the smoothing process. Meanwhile, on examination of ACF and PACF plots, ARIMA does not seem a viable model for non-decomposed data. For this reason, weather is incorporated as supplementary data for the RNN fitted to non-decomposed data. It has been found elsewhere in the literature that models capable of accounting for non-linearities may be better placed to model weather-related

TABLE I. ARIMA MODEL ORDER

| Data series | ARIMA Model Order Trials | |
|---|---|---|
| | ***Maximum order evaluated*** | ***Order with minimal AIC*** |
| EMD-D | (3,1,5) | (3,1,5) |
| VMD-D | (3,1,3) | (2,1,3) |
| EMD-W | (3,1,3) | (3,1,2) |
| VMD-W | (3,1,2) | (3,1,2)(2,0,2) at lag 5 |

TABLE II. RNN MODEL ARCHITECTURE

| Data series | RNN Model Architectures | |
|---|---|---|
| | ***Input length*** | ***Hidden layer size(s)*** |
| EMD-D | 60 | 32 |
| VMD-D | 60 | 32 |
| EMD-W | 4 | 7 |
| VMD-W | 5 | 7, 4 |
| True-W | 5 | 7 |
| True-W +weather | 4 samples x 4 features | 8, 12, 8 |

impacts on crops, further reinforcing the decision to proceed with RNN ahead of ARIMA for this task [9].

*c) ARIMA forecasting:* Once model order for each series was decided, forecasts were generated by fitting an ARIMA to the entire dataset up to that point and generating a one-step-ahead forecast. This was done in both validation and test set forecasts.

*d) RNN forecasting:* Validation and test sets were arranged for forecasting to begin at the 1st datapoint of validation and test sets. For test set forecasts, RNN models were not retrained on a combined training and validation set. Once RNN forecasts were generated, it was necessary to rescale the forecasts to match the original prices series.

### F. Forecast Combination

The ABC technique employed is an implementation of LL-ABC as originally formulated in [18] and applied to agricultural commodities in [17]. This paper applies one slight tweak to LL-ABC in that the minimum number of bees in the elitist colony is not set as a proportion of colony size. Rather, it is set at a constant level of eight bees. Parameters working well in most cases were found as given in Table 3 below.

The drawbacks of ABC have been discussed. This study proposes a simple algorithm named error-informed discretised search (EIDS) which will stand as a fair comparison. Like ABC, EIDS is a means of searching the model weight state-space to empirically find the best solution. The steps involved in EIDS are the following:

- Calculate each base model's share of total MAPE.
- Calculate each base model's share of total inverse MAPE.
- Choose a step size for the search process (0.01 in this implementation) and the number of iterations after which the algorithm will halt if no improvements have been made.

Repeat until halting condition (no improvements after certain number of iterations: 1000 in this implementation):

- Randomly choose a weight to increase by step size, with probability of each element's share of inverse MAPE.
- Randomly choose a weight to decrease by step size, with probability of each element's share of MAPE.
- If the weight changes improve the ensemble's MAPE, the changes are kept. Otherwise, they are discarded.

TABLE III. ABC PARAMETERS

| ***Parameter*** | ***Value*** |
|---|---|
| Swarm size | 350 |
| Iterations through swarm | 5000 |
| Weight initialisation max | 2-model: 1; 4-model: 0.5; 5-model: 0.35 |
| Weight initialisation min | 2-model: 0; 4-model: 0.1; 5-model: 0.05 |
| Bees in elitist colony | 8 |
| Failed attempts before solution abandonment | 10 |

There are a couple of benefits to EIDS over and above ABC. Solutions where the sum of weights is not approximately equal to 1 are not entertained. This is an area where ABC wastes much time resources. EIDS also accounts for the fact that stronger base models should most often be weighted higher than weaker base models in an optimal ensemble. Furthermore, it should be noted that the memory footprint of EIDS represents only a tiny fraction of ABC's. Although memory resources are not a significant concern of either algorithm in this case, it is another advantage of EIDS.

## IV. RESULTS

Full validation and test results for both corn and soybean, as well as all combination weights for ABC and EIDS can be viewed on the project's associated GitLab (see link above). Performances of single models are given in Table 4. Tables 5 and 6 give daily and weekly ensemble MAPES respectively. Figure 4 visualises single-model weekly forecasts. Figure 5 displays forecasted values for the strongest weekly ensemble.

There are substantial differences in the performance of base models to that described by [17]. VMD-ARIMA performance at MAPE 0.543 is matched by 0.72 in [21]. The present study fails to replicate [17]'s EMD-ARIMA of 0.59, scoring 0.68.

Curiously, in all observed cases in the current study, validation error was greater than test error.

**RQ1)** 4-model daily corn forecasts with ABC combination scored MAPE 0.5588. With a four-model ensemble, median and trimmed mean combination schemes are equivalent, with MAPE 0.5964. A mean combination scheme scores MAPE 0.5927. ABC does still outperform mean, trimmed mean and median, so it cannot be said that its only benefit over and above these approaches is in its ability to prune weaker models.

At 0.56 MAPE, the four-strong ensemble performs comparably to Wang's MAPE but has not threatened to surpass the original fifteen-strong ensemble.

TABLE IV. SINGLE MODEL TEST SET MAPES FOR CORN

| **Model** | **Individual Model Test Set MAPE for Corn** | |
|---|---|---|
| | ***Daily forecast MAPE*** | ***Weekly forecast MAPE*** |
| EMD-ARIMA | 0.68 | 1.52 |
| VMD-ARIMA | 0.54 | 1.26 |
| EMD-RNN | 0.69 | 1.53 |
| VMD-RNN | 0.63 | 1.31 |
| True-RNN | 1.27 | 2 |
| True-RNN +weather | - | 2.27 |

TABLE V. DAILY ENSEMBLE TEST SET MAPES FOR CORN

| **Ensemble size** | **Daily Ensemble Test Set MAPEs** | |
|---|---|---|
| | ***ABC combination*** | ***EIDS combination*** |
| 2-Model | 0.5435 | 0.543 |
| 4-Model | 0.5588 | 0.5412 |
| 5-Model | 0.5553 | 0.5208 |

TABLE VI. WEEKLY ENSEMBLE TEST SET MAPES FOR CORN

| Ensemble size | Daily Ensemble Test Set MAPEs | |
|---|---|---|
| | *ABC combination* | *EIDS combination* |
| 2-Model | 1.263 | 1.242 |
| 4-Model | 1.266 | 1.218 |
| 5-Model | 1.287 | 1.266 |
| 5-Model+weather | 1.245 | 1.219 |

Table 4 shows that VMD-ARIMA performs at 0.543 MAPE. In this sense, this study finds that ABC as a combination technique has not even succeeded in beating the best single model.

**RQ2)** In addressing the matter of ABC convergence, the findings from RQ1 must be taken into account as a prerequisite. The fact that ABC has not even outperformed its own base model VMD-ARIMA, means that it certainly has not converged in this case. This is provably so because a better result could have been achieved by setting VMD-ARIMA's weight to 1, and all other weights to 0.

Timewise, in all cases ABC took 12 to 16 minutes to arrive at a solution.

The present research has outlined an alternative search-based approach to ABC, named EIDS, to overcome some of ABC's identified inefficiencies. Tables 5 and 6 compare EIDS and ABC combinations in terms of MAPE on various ensembles. In all cases EIDS took 75 to 125 milliseconds to complete – almost four orders of magnitude faster than ABC.

In all ensembles, EIDS has found lower test MAPE than ABC.

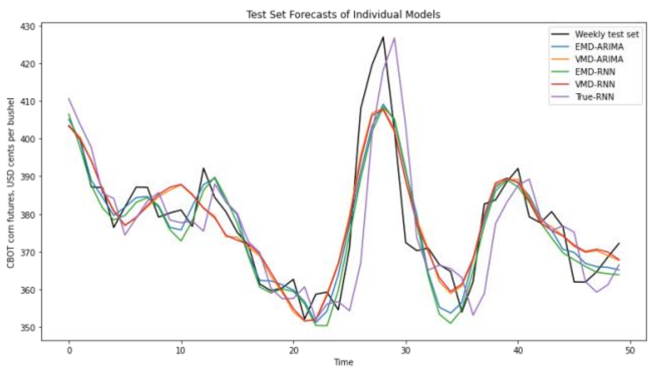


Fig. 4. Test set forecasts of selected base models on weekly data.

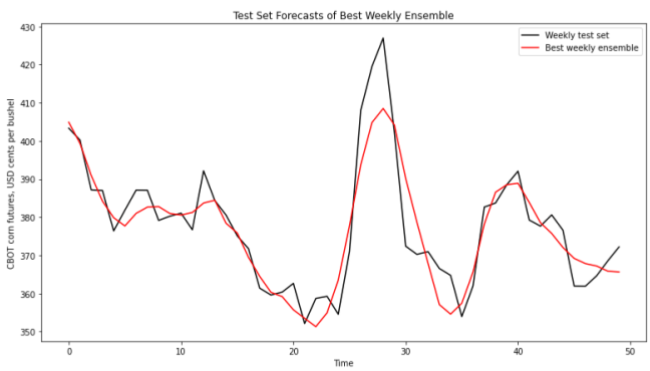


Fig. 5. Test set forecasts of the best weekly ensemble. The ensemble consists of five base models with combination weights decided by EIDS.

**RQ3)** An RNN model was found which made improvements to validation set forecasts on non-decomposed data. However, on test data this same model performed worse than the RNN without weather data. On the other hand, the weather-RNN, when put in a five-strong ensemble along with EMD-ARIMA, EMD-RNN, VMD-RNN, and VMD-ARIMA, gives better results than no-weather RNN in the same setup. Indeed, this ensemble with MAPE 1.219 comes very close to the best ensemble, which is the four-strong model combined using EIDS.

The strongest weekly model scored MAPE 1.218%.

## V. DISCUSSION

The consequences of each research question's results will be discussed presently, followed a general discussion of the present study, it's approach, strengths, and weaknesses.

**RQ1)** Insofar as RQ1 goes, if the 4-model ensemble had produced a comparable MAPE to [17]'s 0.49, it would naturally have followed that ABC had simply carried out a 'pruning' function, ridding the ensemble of contributions from weaker models. Apparently, the true picture is more complex. In the absence of the final weights of [17]'s ABC process, no further progress can be made in determining how models were combined in the original study.

Even with fewer base models, effective search-based forecast combination still exhibits pruning of weaker base models from the ensemble. However, the inclusion of weak base models can still reduce forecasting error. The present study finds that the strongest model for both daily and weekly series involve setting some weights to negative values, allowing the strongest base model to contribute even higher weights to the final ensemble. Where ensemble parsimony is harmed, forecast accuracy is still improved. On the other hand, for those more concerned with ensemble parsimony, the present study recommends VMD-ARIMA as the single strongest model. VMD-ARIMA is the dominant contributor to all strongest forecast combinations, performing comparably to these ensembles on its own.

**RQ2)** The implementation of ABC applied to the daily 4-model ensemble in this study has certainly not converged to a global optimum. This is proved by the fact that VMD-ARIMA alone achieves a lesser MAPE. The same point is reconfirmed by an even lower MAPE achieved using EIDS combination on the same base models. Even though the design of EIDS is less oriented towards global optimisation and more towards quickly achieving *approximately* optimal results, the question remains whether EIDS has indeed exhibited global convergence in this study. On this matter, we can establish with certainty that in 5-model weekly forecast combinations the global optimum has not been reached, because the ensemble's performance is surpassed by EIDS combination applied to a subset of its base models. On all other combinations it remains possible that a global optimum or near-global optimum has been reached. What is certain about EIDS in this study is that it has achieved lower test losses in all cases compared to ABC, in approximately 1e-4 of the time.

The reason EIDS outperforms ABC in terms of forecast accuracy is it simply allows the strongest base model(s) to hold sway. As discussed, VMD-ARIMA is the strongest single model with MAPE scores of 0.54 and 1.26 for daily and weekly forecasts respectively. While ABC does learn to rely on VMD, EIDS capitalises much more fully on its strength. EIDS often weights VMD-ARIMA greater than 1, setting other base model weights to negative values. On the other hand, ABC never set VMD-ARIMA's value greater than 1, and was less likely to find negative weights for weaker base models.

**RQ3)** MAPE of 1.218% was achieved in forecasting next-week corn futures. To the authors' knowledge, there is no contribution in the literature surpassing this figure.

The question of whether performance has been improved by weather data has brought more mixed results. True-RNN+weather exhibits stronger validation MAPE, but weaker test MAPE than True-RNN. When added to an ensemble with EMD-ARIMA, EMD-RNN, VMD-ARIMA and VMD-RNN, the ensemble produces better results than with True-RNN+weather. Since the contribution of True-RNN+weather to its ensemble is a negative weight, more discussion will be given to the former point. A plausible explanation for this validation-test set mismatch could be due to the validation set covering only four full growing seasons. Likewise for the test set. Perhaps this the model has overfit to validation weather conditions in some aspect, resulting in poor test-set performance. This potential flaw is not necessarily an isolated issue in the present study. There are a number of potential methodological holes in the present study. A general discussion around this topic is offered presently.

*a) Datasets*: The fact that only a few growing seasons are included in validation and test sets has been outlined. This is part of a broader criticism of the present research that forecasting models could have been helped by a greater volume of data. The motivation for beginning training in 1974 has been outlined. However, there are other avenues that could be explored for generating supplementary data rather than looking further into the past. For example, [24] has proposed a methodology for generating simulated price data for agricultural commodities. It may be worth investigating whether such a technique could benefit the present work.

This study has drawn on weather data to improve forecast accuracy, with mixed results. There are a host of other variables outlined at the outset which could have been trialled. Maybe incorporating more variables would have better captured the non-linear complexities of the wider picture. Even within the weather variables used (a similar selection to [9]), further weather-related variables such as wind intensity, soil temperature and soil moisture could have been applied.

This study has largely treated price data as a single continuous series. When the data is more closely considered, there are nuances which could have been better respected. Weekends and public holidays are completely ignored. Given that CBOT only operates on weekdays, the study's daily analyses violate an assumption of time-series forecasting that observations should be taken at regular temporal intervals.

Little subtlety is also shown to the price arcs of each year's own price dynamics. [8] has reported that a given model can give quite different results depending on a given year's price arc. This detail has been ignored.

*b) Pre-processing*: Another weakness of the study's approach can be found in the fact that the validation error achieved for all models is higher than test error. There are two plausible explanations for this: underfitting, or an unrepresentative or in some way unusual validation set. It seems more likely that the latter is the cause. ARIMA models were carefully selected and properly fit. Every measure was taken to find optimal RNN architectures and each model was selected for the lowest validation error. On casual observance (Fig. 1), the validation set (the same as in [17]) has much higher values than its adjacent years. Maybe further pre-processing could have mitigated this effect. For instance, log-transforming the data would have brought extreme values more into range.

Other approaches to stationarity for ARIMA pre-processing could have been attempted. The differenced corn price series exhibits high variance compared to the training set. Although the series passes the ADF stationarity test, it may have been more correct to manually homogenise series variance rather than relying on differencing alone to achieve stationarity. Furthermore, it was not investigated whether an RNN could have trained well on non-decomposed stationary data. This can sometimes be the case when an RNN has few datapoints for training, as with this study's weekly crop data.

The approach to weekly down-sampling has been to calculate weekly means. Another approach may have been to take weekly closing price. The present study says little about the point in week that the instantaneous price will equal the weekly mean. Relatedly, it should be explicated that the study has little to say about the price fluctuations of any one futures contract specifically, only of continuous futures contracts.

Two decomposition techniques have been used in this study: EMD and VMD. However, [16] has found that wavelet packet transform outperforms EMD for the purposes of price forecasting with a neural network. It may have been interesting to evaluate how this decomposition could have contributed to this study's ensemble.

*c) Modelling*: Criticisms have already been outlined around anomalous price values in the validation set. This may have had adverse knock-on effects at the modelling stage. Maybe the models could have been trained on data more selectively. Anomalous price values may be best left to the domain of time-to-event price spike prediction, while price forecasting may be best kept to periods of 'business as usual'. This separation raises a separate challenge of series classification to decide which time frames should qualify as 'normal', and which should be deemed anomalous.

*d) Forecast combinations*: This study has compared ABC combination with another search-based technique. However, it should be mentioned that there are other possibilities for calculating forecast combinations rather than exhaustive search, such as Bayesian model averaging. This was an unexplored avenue in the present study.

*e) General approach*: A drawback in the application of weather data to the problem may be that price swings may be more affected by weather forecasts rather than concurrent conditions. If this is the case, the current methodology would fail to pick up on this phenomenon. As already noted, [12] has made progress in assimilating weather forecast data into forecasts by drawing on news headlines. Maybe this approach would have been more fruitful in this problem instance. Furthermore, a text-based approach would open an entirely new prospect, in the form of US Department of Agriculture (USDA) reports. USDA regularly publishes reports regarding crop planting progress, plant health and projected yields. Estimating the impact of USDA reports on price fluctuations is an area unto itself (for example, [25]). Incorporating a text-based component to the solution would open up the enticing possibility of including this information into the price forecasting ensemble.

*e) Possible extensions:* Weekly forecasting with MAPE 1.218 stands as a worthwhile proof-of-concept. An interesting extension to the present study may be to extend forecasts from one-step-ahead to multi-step-ahead. EIDS or ABC could be used to find weights for effective combination at further intervals ahead distinct from one-step-ahead combinations, as different model performances may deteriorate at different rates at further horizons.

## VI. Conclusion

Corn is the most important agricultural commodity to the US economy. The present study has built on the work of [17], which has produced accurate forecasts of corn and soybean futures. ABC in [17] did not simply carry out a pruning function, ridding the ensemble of weak base models. On the other hand, it is an inefficient forecast combination technique, which was improved upon in this study by EIDS in both performance and time requirements. Week-ahead forecasts have been generated with 1.218% MAPE – a worthwhile contribution. Weekly forecasts have not been materially improved by incorporating corn belt weather data. Limitations of the study's approach have been outlined. It is suggested that future efforts could extend this study's efforts by developing combination schemes for multi-step-ahead agricultural commodity price forecasting.

## Acknowledgement

The authors would like to thank Dr. Juan Beltran, ECOM Trading, for discussion of available data sources, Prof. Jesus Crespo Cuaresma and Dr. Jaroslava Hlouskova for sharing data from previous studies, and our supervisor Prof. Martin Crane for his help, support, and guidance throughout the process.

## References

[1] FAO, *Faostat: FAO Statistical Databases*. Rome, Italy: Food & Agriculture Organization of the United Nations (FAO), 2000.

[2] C. Hutchins, "Feedgrains sector at a glance," *Usda.gov*. [Online]. Available: https://www.ers.usda.gov/topics/crops/corn-and-other-feedgrains/feedgrains-sector-at-a-glance/. [Accessed: 01-Aug-2022].

[3] L. F. Foreman, "Characteristics and production costs of U.s. corn farms," *Usda.gov*, 2001. [Online]. Available: https://www.ers.usda.gov/webdocs/publications/47146/15704_sb974-1_1_.pdf?v=0. [Accessed: 01-Aug-2022].

[4] "NCH-40," *Purdue.edu*. [Online]. Available: https://www.extension.purdue.edu/extmedia/nch/nch-40.html. [Accessed: 01-Aug-2022].

[5] "NCH-18," *Purdue.edu*. [Online]. Available: https://www.extension.purdue.edu/extmedia/nch/nch-18.html. [Accessed: 01-Aug-2022].

[6] "Approved by USDA's world agricultural outlook board," *Usda.gov*. [Online]. Available: https://www.ers.usda.gov/webdocs/outlooks/36651/39297_fds-13g-01.pdf?v=737. [Accessed: 01-Aug-2022].

[7] X. Yuan, J. Tang, W.-K. Wong, and S. Sriboonchitta, "Modeling co-movement among different agricultural commodity markets: A copula-GARCH approach," *Sustainability*, vol. 12, no. 1, p. 393, 2020.

[8] M. Shahhosseini, G. Hu, and S. V. Archontoulis, "Forecasting corn yield with machine learning ensembles," *Front. Plant Sci.*, vol. 11, p. 1120, 2020.

[9] V. R. Joshi, M. J. Kazula, J. A. Coulter, S. L. Naeve, and A. Garcia Y Garcia, "In-season weather data provide reliable yield estimates of maize and soybean in the US central Corn Belt," *Int. J. Biometeorol.*, vol. 65, no. 4, pp. 489–502, 2021.

[10] J. Crespo Cuaresma, J. Hlouskova, and M. Obersteiner, "Agricultural commodity price dynamics and their determinants: A comprehensive econometric approach," *J. Forecast.*, vol. 40, no. 7, pp. 1245–1273, 2021.

[11] B. Algieri and A. Leccadito, "Extreme price moves: an INGARCH approach to model coexceedances in commodity markets," *Eur. Rev. Agric. Econ.*, vol. 48, no. 4, pp. 878–914, 2021.

[12] J. Li, G. Li, M. Liu, X. Zhu, and L. Wei, "A novel text-based framework for forecasting agricultural futures using massive online news headlines," *Int. J. Forecast.*, vol. 38, no. 1, pp. 35–50, 2022.

[13] D. Zhang, S. Chen, L. Liwen, and Q. Xia, "Forecasting agricultural commodity prices using model selection framework with time series features and forecast horizons," *IEEE Access*, vol. 8, pp. 28197–28209, 2020.

[14] H. Ouyang, X. Wei, and Q. Wu, "Agricultural commodity futures prices prediction via long- and short-term time series network," *J. Appl. Econ.*, vol. 22, no. 1, pp. 468–483, 2019.

[15] T. Xiong, C. Li, Y. Bao, Z. Hu, and L. Zhang, "A combination method for interval forecasting of agricultural commodity futures prices," *Knowl. Based Syst.*, vol. 77, pp. 92–102, 2015.

[16] D. Wang, C. Yue, S. Wei, and J. Lv, "Performance analysis of four decomposition-ensemble models for one-day-ahead agricultural commodity futures price forecasting," *Algorithms*, vol. 10, no. 3, p. 108, 2017.

[17] J. Wang, Z. Wang, X. Li, and H. Zhou, "Artificial bee colony-based combination approach to forecasting agricultural commodity prices," *Int. J. Forecast.*, vol. 38, no. 1, pp. 21–34, 2022.

[18] W. Gao, H. Sheng, J. Wang, and S. Wang, "Artificial bee colony algorithm based on novel mechanism for fuzzy portfolio selection," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 5, pp. 966–978, 2019.

[19] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," *J. Glob. Optim.*, vol. 39, no. 3, pp. 459–471, 2007.

[20] A. Rodríguez and M. Melgarejo, "Identification of Colombian coffee price dynamics," *Chaos*, vol. 30, no. 1, p. 013145, 2020.

[21] National Centers for Environmental Information (NCEI), "Search," *Noaa.gov*. [Online]. Available: https://www.ncei.noaa.gov/cdo-web/search. [Accessed: 01-Aug-2022].

[22] K. Dragomiretskiy and D. Zosso, "Variational Mode Decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, 2014.

[23] "Vmdpy," *PyPI*. [Online]. Available: https://pypi.org/project/vmdpy/. [Accessed: 01-Aug-2022].

[24] P. Oktoviany, R. Knobloch, and R. Korn, "A machine learning-based price state prediction model for agricultural commodities using external factors," *Decis. Econ. Finance*, vol. 44, no. 2, pp. 1063–1085, 2021.

[25] A. M. McKenzie and Y. Ke, "How do USDA announcements affect international commodity prices?," *J. Commod. Mark.*, no. 100239, p. 100239, 2021.

## APPENDIX A – WORKLOAD SHARING

Patrick Travers, 21267100 – developed and coded the methodology including EIDS, obtained corn forecasts, and wrote the final draft.

Sheetal Kumar, 21262085 – Experimented with alternative base models, executed pre-processing, obtained soybean forecasts, completed technical documentation, and wrote the first draft.