Data to Donations: Towards In-Kind Food Donation Prediction across Two Coasts

Esha Sharma[1‡‖], Lauren Davis[2§†], Julie Ivy[3§‖], Min Chi[4‡‖]
Dept. of Computer Science[‡], Dept. of Industrial & Systems Engg.[§],
North Carolina State University[‖], North Carolina A&T State University [†],
Raleigh, NC [‖], Greensboro, NC[†]
{esharma2[1], jsivy[3], mchi[4]}@ncsu.edu, lbdavis@ncat.edu[2]

***Abstract*—Our goal in this work is to build effective yet robust models to predict unreliable and inconsistent in-kind donations at both weekly and monthly levels for two food banks across coasts: the Food Bank of Central Eastern North Carolina in North Carolina and Los Angeles Regional Food Bank in California. We explore three factors: *model*, *data length*, and *window type*. For the *model*, we evaluate a series of classic time-series forecasting models against the state-of-the-art approaches such as Bayesian Structural Time Series modeling (BSTS) and deep learning models; for the *data length*, we vary training data from 2 weeks to 13 years; for the *window type*, we compare sliding vs. expanding. Our results show the effectiveness of different models heavily depends on the *data length* and the *window type* as well as characteristics of the food bank. Motivated by these findings, we investigate the effectiveness of employing an average of all predictions formed *by considering all three factors* at both monthly and weekly levels for both food banks. Our results show that this average of predictions significantly and consistently outperforms all classical models, deep learning, and BSTS for the donation prediction at both monthly and weekly levels for both food banks.**

***Index Terms*—Food Insecurity, Humanitarian Supply Chain, Bayesian Structural Time Series, Long Short Term Memory, Training Length, Expanding and Sliding Window**

I. Introduction

Food insecurity in a household refers to the inability to obtain nutritious food to maintain a healthy and active lifestyle for the entire family [1]. More than 37 million Americans, including over 11 million children, experienced some sort of food insecurity in 2018 [1, 2]. Due to COVID-19, these numbers *more than doubled* by April 2020 [3]. Hunger relief organizations such as food banks attempt to meet the needs of the food insecure [4]. These organizations achieve these goals by recovering surplus food from a variety of sources and distributing it to charitable agencies that in turn, distribute this food to food insecure populations. Now more than ever before, there is an urgent need for building effective, efficient, and equitable food distribution systems that provide food *equitably* to those in need, *efficiently* maximize the yield of the donated supply minimizing waste, and distribute food in a cost *effective* manner. However, the distribution networks of food banks are dynamic and consist of multiple configurations (e.g., hub and spoke) with many charitable autonomous agency partners that receive donated food. Many food banks depend on both in-continuous and non-obligatory donations from benefactors [5, 6]. As a result, food aid distribution is challenged by the *unpredictable timing* and *quantity* of the donated food supply which in turn, makes transportation needs, usage, and routes difficult to anticipate [6, 7].

Our goal in this work is to build an effective yet robust model to predict *in-kind donations* made to food banks. Generally speaking, donations made to food banks can be classified as *monetary* vs. *in-kind*. The former refers to *cash* donations from the federal and state governments and offers flexibility in management, while the latter refers to *non-cash* donations made by individuals and retail donors that vary from day to day [5, 6]. *In-kind donations* require the immediate assignment of manpower to manage them. In one of the food banks in this work, for instance, over half of the food distributed is perishable [8]. Food banks are unable to provide adequate refrigeration and storage due to uncertainty about in-kind donations, which can lead to food wastage [6]. Predicting these donations in advance not only helps food banks plan storage facilities for fresh and nutritious food better, but also helps them plan distributions that are equally distributed [7].

While in-kind donations are important, they can be very challenging to predict as there are so many variables such as income, social status, age, and gender that impact non-cash donations from retailers and individuals [9, 10, 11]. In addition, unobserved factors such as empathy can be a driving force in individual donors' donations [12] while retail donations are susceptible to changes in the economy, which is volatile, and to customer consumption patterns [5].

In this work, we investigate in-kind donations for two food banks on opposite coasts of the United States: *The Food Bank of Central and Eastern North Carolina (FBCENC) in North Carolina (NC)* and *Los Angeles Regional Food Bank (LARFB) in Los Angeles, California*. At both the monthly and weekly levels, we investigate the impact of three factors: the *model*, the *data length*, and the *window type*.

For the *model*, we explore classical time-series models including Autoregressive Integrated Moving Average Model (ARIMA), Simple Moving Average (MA), Exponentially Weighted Moving Average Model (ETS) as well as advanced models such as deep learning models such as Long Short Term Memory (LSTM) and Bayesian Structural Time Series Models (BSTS). For the *data length*, we vary data length from

2 weeks to 13 years since the length of training data can greatly influence donation predictions and different tasks often have different "optimal" data lengths [13, 14, 15, 16, 17, 18, 19]. When it comes to the *training windows*, we compare expanding vs. sliding windows. In the former, the *origin of the data* remains fixed so that the training data expands with each time-step, whereas in the latter, the origin of the data slides so that the *length of training data* remains fixed with each time-step. Studies have shown that expanding windows are more effective for tasks involving long-term dependencies [14, 16, 20, 21, 22] while sliding windows are more suitable for tasks involving short-lived trends in data [13, 15, 23, 24]. According to our preliminary findings, the effectiveness of the different models depends largely on the length of the training data and the type of window used. In light of this, we explore averaging predictions because past studies have shown that simple averages of predictions often outperform complex methods as well as combinations of forecasts [25, 26]. Overall, our results show that across both food banks, averaging these predictions consistently beats the individual methods at both the weekly and monthly levels. In summary, our main contributions are:

- To our knowledge, this is the first work that predicts donations to two different food banks on both a monthly and weekly basis.
- This is the first work to apply BSTS and LSTM to predict food bank donations and compare their performance with classical time-series forecasting models such as ARIMA, MA, and so on, by systematically considering the impact of different data lengths and window types on the effectiveness of various models.
- Our results show that a simple method that takes the average of predictions by combining different models, data lengths, and window types can outperform all basic models at both monthly and weekly levels for the two food banks.

II. Background and Related Work

Classical time-series models: Previous studies have explored the use of classical time-series models, which include ARIMA, MA, ETS, and Support Vector Regression (SVR), for predicting food bank donations [6, 27, 28, 29]. For example, Davis et al. compared ARIMA, MA, and ETS for monthly donation prediction using six years' donation data from one food bank and showed that ETS performed the best in terms of the least Mean Average Percentage Error (MAPE) [6]. Pugh and Davis compared SVR against ETS on the same task with nine years of donation data and found the former outperformed the latter [27]. A classical time-series model employs the frequentist method of making predictions and assumes that the parameters in the model remain fixed for the entire time series of data. Given the high degree of unpredictability in donations made to the food bank, a more suitable model should be flexible and be able to accommodate the dynamics of donation behaviors [30, 31].

BSTS Models: BSTS [32] relies on Bayesian statistics for making predictions and assumes that the state of the system is continually changing. Therefore, BSTS is highly adaptive and is capable of capturing complex, uncertain, and dynamic patterns in a time-series such as that of the in-kind donations. BSTS originated from a long line of research involving applying *Bayesian averaging (BA)* over multiple models to avoid inferior predictions from a single model [32]. BA predicts by constructing an ensemble model that combines multiple predictions weighted by their similitude to the empirical distribution of the predicted variable [33]. For example, Wright et al. applied BA to a simple linear regression to predict inflation in the United States; they found that BA gave superior forecasts as compared to a simple averaging technique over multiple forecasts generated by the linear regressions and more importantly, they found that BA was robust to outliers [34]. Scott and Varian later proposed BSTS and used it to predict consumer sentiment and showed that BSTS is more accurate than a baseline naive autoregressive model [32]. Since then, BSTS has been widely applied in humanitarian-related applications such as predicting unemployment claims and retail sales [35], alcohol-related harms [36], and positive COVID-19 cases [37].
LSTM has been widely applied to handle large-scale sequential data and has demonstrated superior performance in time-series forecasting as compared to classic time-series models. LSTM can capture short term and long term dependencies in temporal sequential data [38, 39, 40, 41, 42, 43]. There have been several studies comparing LSTM to classic time-series forecasting. For example, Simai et al. found that LSTM outperformed ARIMA on a wide range of forecasting tasks such as stock indices prediction, Housing index prediction, Food and Beverage index prediction and so on [20]. Similarly, LSTM outperformed ARIMA on the task of bitcoin price predictions in [44].
Data Lengths Explored: Past research has shown that for making accurate forecasts, choosing an appropriate data length is critical [13, 14, 15, 16, 17, 18]. Skabar et al. examined the effectiveness of different training data lengths (from 10 days to 250 days) on DJIA trading strategies. They suggested using the short-term and medium-term periodicity in data to determine the training data length [45]. A study by Li et al. explored how training data lengths affected detection of heart rhythm disorders and found that an "optimally" short training data length led to faster detection of faulty heart rhythms while an extremely short length caused incorrect predictions [46]. In a study closely related to ours, Davis and Pugh examined three options for training data length: 96, 48, and 24 months and found that SVR models trained using 24 months of data delivered the best performance (lowest MAPE for food bank donations) [27].
Window Types Explored: Expanding and sliding windows have both been extensively explored in past research. Generally speaking, the expanding window has been used to make predictions in time-series data with long term dependencies and has been used for a wide range of forecasting tasks, such as prediction of stock indexes, housing indexes, food and

beverage indexes [20], and electric load forecasts [22]. On the other hand, the sliding window has been used to forecast time series data with short-term to medium-term dependencies, or to make more rapid forecasts. It has been used, for instance, for the classification of electroencephalography in [23], for the prediction of cloud computing resource demand in [24], and for stock market forecasts in [15]. In summary, prior research suggests that the expanding window functions better when the frequency of data is greater than a few weeks or months [14], while the sliding window comes in handy when dealing with time-series data of high frequency (e.g. in hours) [14].

Prior work on Average of Predictions: Past work has found that combining predictions from multiple predictors by employing the average value of individual predictors is effective in combining these predictions [25, 26, 47]. Note that this simple averaging differs from BSTS in that BSTS makes predictions using a structural time-series and assumes that all the components in this equation are normally distributed. It utilizes Bayesian statistics to determine how the predicted component changes with changes in the remaining components of the equation. Based on these distributions learned and the new values seen, BSTS simulates a distribution of predictions and then ensembles these predictions by weighting them by their probability of being close to the empirical distribution. On the other hand, a simple average of the predictions gives all the predictors the same weight and does not make use of any prior knowledge of the system while constructing this average.

III. Methods

Problem Definition: Our dataset is a uni-variate time-series data which can be represented as $Y = \{y_1, y_2, ..., y_N\}$, where N is the total number of time-stamps. For example, we have $N = 167$ monthly events and $N = 728$ weekly events in FBCENC dataset. The goal of this work is to predict y_t, the donation at t using $y_1, y_2, ..., y_{t-1}$. In the following, we use $\hat{y}_t$ for *predicted* donation while y_t is the actual donation.

Three Types of Predictive Models

Four Classical Models

1) ARIMA employs the autocorrelations between successive points on a time-series and the past errors in prediction to predict the next value on a time-series [48]. We have:

$$y_{\text{t}} = \sum_{i=t-p}^{t-1} \alpha_{\text{i}} y_{\text{i}} + e_{\text{t}}$$

where p is a hyperparameter referring to the number of the most recent terms to be considered; $\alpha_{\text{t-p}}$, $\alpha_{\text{t-p+1}}$, .. $\alpha_{\text{t-1}}$ are their corresponding weights. Additionally, e_t is defined as:

$$e_{\text{t}} = \sum_{j=t-q}^{t-1} \beta_{\text{j}} < \hat{y}_{\text{j}} - y_{\text{j}} >$$

where q is the number of the most recent prediction errors to be considered and $\beta_{\text{t-q}}$, $\beta_{\text{t-q+1}}$,... $\beta_{\text{t-1}}$ are corresponding weights. Note that we need at least two data points to construct an ARIMA model given $p < t$.

2) Simple (equally-weighted) MA models the value of y_t in terms of the average of the most recent *m* values where $m \geq 1$.

$$y_{\text{t}} = 1/m \sum_{i=t-m}^{t-1} y_{\text{i}}$$

Note that the same weight is given to the past *m* terms. MA needs least two data points to construct a model since $m < t$.

3) ETS models the value of y_t in terms of *all* the past values [49]. We generated two variations of the ETS model: the one without trend and seasonality is referred as ***ETS-Plain***, in which the weight given to the past value decreases as the distance from the value being predicted increases.

$$y_{\text{t}} = \sum_{i=1}^{t-1} \alpha^{t-i} y_{\text{i}} \text{ where } 0 < \alpha < 1$$

4) ETS-Plus integrates additive trend and seasonality into consideration. We have

$$y_{\text{t}} = \sum_{i=1}^{t-1} \alpha^{t-i} y_{\text{i}} + \varrho_{\text{t-1}} + \kappa_{\text{t-p}}$$

where $0 < \alpha < 1$, $\varrho_{\text{t-1}}$ is the trend at time $t-1$ and κ is the seasonality at time $t-p$ where p in the number of periods in the time-series. For example, p is 12 when there is a seasonality of 12 months in the data

Note that for both variations of ETS, to learn α we need at least two data points to produce a prediction at monthly level and six data points for weekly level prediction. We replace the missing predictions for ETS at monthly and weekly level with donation data from the previous time-step for uniform comparison across all models.

Bayesian Structural Time Series Modeling (BSTS) employs Bayesian averaging over multiple models to build a single model while preserving the intricate temporal nature of a time-series. Generally speaking, a structural time-series model can be directly decomposed into its time-varying components [50, 51, 52]. The BSTS models used in this work can be decomposed into its *trend*, *seasonality*, and *regression components*. BSTS leverages Bayesian statistics to capture the trend and seasonality in the data and the parameters in BSTS come from a distribution instead of being single values.

BSTS models predict y_t using the state of the system S_t, noise η_t, and a set of weights α_t. S_t is determined by y_{t-1}, ϱ_{t-1} which is the *trend* at $t-1$, κ_{t-1} which is the *seasonality* at $t-1$, and a noise component ω_t.

$$y_{\text{t}} = \alpha_{\text{t}} S_{\text{t}} + \eta_{\text{t}} \tag{1a}$$

$$S_{\text{t}} = y_{\text{t-1}} + \varrho_{\text{t-1}} + \kappa_{\text{t-1}} + \omega_{\text{t}} \tag{1b}$$

BSTS assumes all the components in Eq. 1b follow normal distributions. More specifically, it is assumed that η_t and ω_t are independent and identically distributed and they follow a normal distribution with mean μ = 0 and a corresponding standard deviation of ${\sigma_\eta}^2$ and ${\sigma_\omega}^2$ respectively. BSTS uses the Monte Carlo Markov Chains (MCMC) and the Kalman filter to predict the marginal likelihoods for parameter α_{t} which is then uses to predict y_t. For this, BSTS observes the past values of the state $S_1, S_2, ..S_{t-1}$, and the values of the dependent variable $y_1, y_2, ..y_{t-1}$ to determine ${\sigma_\eta}^2$ and ${\sigma_\omega}^2$ for the noise η_t, ω_t and the marginal likelihoods for α_t. It then uses the information on marginal likelihoods of α_{t} learnt along with the next observed value for S_{t} to predict y_{t}. Note that we need at least two data points to construct an BSTS model.

Long Short Term Memory (LSTM) primarily passes information through the LSTM cells [53]. In the standard LSTM cell unit, the cell state $\boldsymbol{c}^t$ serves as an *internal memory* and controls the information flow. It is generated by forgetting

information through a forget gate $\boldsymbol{f}^t$, the most recent cell state $\boldsymbol{c}^{t-1}$, adding new information through an input gate $\boldsymbol{i}^t$, and a candidate cell state $\tilde{\boldsymbol{c}}^t$. To apply LSTM to our uni-variate time-series data, we have the following:

$$\begin{aligned}\boldsymbol{i}^t &= \text{sigmoid}(\boldsymbol{W}_h^i \boldsymbol{h}^{t-1} + \boldsymbol{W}_y^i y^{t-1}),\\ \tilde{\boldsymbol{c}}^t &= \tanh(\boldsymbol{W}_h^c \boldsymbol{h}^{t-1} + \boldsymbol{W}_y^c y^{t-1}),\\ \boldsymbol{f}^t &= \text{sigmoid}(\boldsymbol{W}_h^f \boldsymbol{h}^{t-1} + \boldsymbol{W}_y^f y^{t-1}),\end{aligned} \tag{2}$$

where $\boldsymbol{h}^{t-1}$ is a hidden state output by $\boldsymbol{c}^{t-1}$, $\{\boldsymbol{W}_h \in \mathbb{R}^{H\times H}$, $\boldsymbol{W}_y \in \mathbb{R}^H\}$ denote network parameters to be trained and H is the number of hidden nodes. The new cell state can be obtained as follows:

$$\boldsymbol{c}^t = \boldsymbol{f}^t \otimes \boldsymbol{c}^{t-1} + \boldsymbol{i}^t \otimes \tilde{\boldsymbol{c}}^t, \tag{3}$$

where $\otimes$ denotes entry-wise product. Finally, we generate the hidden states by filtering the new cell state through an output gate layer $\boldsymbol{o}^t$, and produce the prediction of donation at each event using a sigmoid function with parameter $\boldsymbol{U}$:

$$\begin{aligned}\boldsymbol{o}^t &= \text{sigmoid}(\boldsymbol{W}_h^o \boldsymbol{h}^{t-1} + \boldsymbol{W}_y^o y^{t-1}),\\ \boldsymbol{h}^t &= \boldsymbol{o}^t \otimes \tanh(\boldsymbol{c}^t).\end{aligned} \tag{4}$$

Training Window: Expanding vs. Sliding

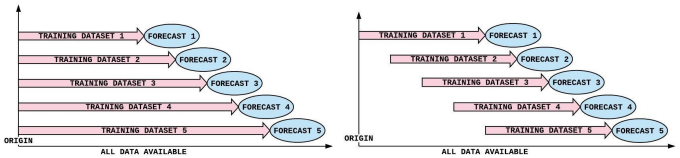


Fig. 1: Expanding (Left) vs. Sliding (Right)

In this work, we used Temporal cross-validation (Temporal-CV) because it was not feasible to use traditional cross-validation (CV) to evaluate the effectiveness of time-series models due to the temporal dependence in the data [18, 54]. In each round, we performed *one single step ahead* using the training data length and evaluated the predicted value against the ground truth; this data was then added to the training data for the next round. Figure 1 shows that when carrying out Temporal-CV, *Expanding Window* keeps the *original training data fixed* while *Sliding Window* keeps *the training data length fixed*.

Data Length For FBCENC, we had 14 years' donation data from 2006 to 2020. For LARFB, we had 6 years' donation data from 2013 to 2019. For the monthly and weekly predictions, different lengths of training data and levels of granularity were used.

Averaging Predictions To calculate the Average value of predictions at a time-step t, we take the simple average of all the predictions available at that time step.

IV. Two Data-Sets

We focused on the gross weight of the in-kind donations made to the FBCENC (from July 2006 to May 2020) and LARFB (from November 2014 to October 2019) by non-government sources, such as individuals and retailers such as Walmart. The weights are in unit of *pounds (LB)*. These two food banks serve two drastically different areas. FBCENC [8] has six branches and serves 34 counties in North Carolina while LARFB [55] caters to Los Angeles county in California. As a result, there are significant differences in the trend, seasonality, and distribution of donations made to the two food banks. For example, using aggregated data we found that the weekly data differs significantly from monthly data in that the former fluctuates much more than the latter. For FBCENC, the monthly donations vary from $1,883K$ to $8,036K$ LB while the weekly donations vary from $3K$ to $8,037K$ LB. For LARFB, the monthly donations vary from $1,419K$ to $4,025K$ LB and the weekly donations vary from $84K$ to $1,930K$ LB.

We also explored the impacts of various economic elements such as stock indices like DJIA and NASDAQ as well as big donors such as Walmart and Food Lion stocks on our foodbank donations. For the FBCENC monthly data, the correlation coefficient r between FBCENC monthly donation and the DJIA/NASDAQ stock one month prior can be as high as $0.79 \sim 0.82$. This correlation remained more or less unchanged even using stock indices up to six months prior. Similarly, the month-old walmart stock prices showed a noticeable correlation of 0.75 with the monthly donation data. For Walmart stock prices, the correlation ranged from 0.73 to 0.75 from two to six months prior. However, no such correlation was found between features of the economy and donations made for LARFB.

V. Experimental Setup

1) Experimental Settings: Table I displays the list of the settings for each of the three factors we considered (Models, Data Length, and Window Type). For the Model, we explored *four* classical time-series models that were explored on the task of food bank donations: *ARIMA*, *MA*, *ETS-Plain*, and *ETS-Plus*; *three* more advanced models that have not been applied on this task: a deep learning model *LSTM* and two BSTS-based models: *BSTS-Plain* and *BSTS-Plus*. BSTS-Plain is the original BSTS model while BSTS-Plus incorporates economic factors discussed in section IV above into the original BSTS model. In addition to the settings explored, all models were tuned using a grid search to find the optimal hyper-parameters. The ETS, ARIMA, and MA models were tuned with the R forecast package [49] using a state-space search technique. The two BSTS models searched the optimal number of iterations for Monte Carlo Markov Chains (MCMC), and the seed for the algorithm. The BSTS models were implemented in R with BSTS package. For the seed, we experimented with 1, 3, 5 ... 130. For the the number of iterations, we experimented with 10, 20 ... 6000. We chose a seed of 30 and 3000 iterations for this problem. The LSTM models were implemented in Keras with Tensorflow as the backend engine. For the number of layers and neurons, we experimented with 1, 2, 4 layers and 2, 4, 6, 8 neurons. For the number of epochs, we experimented with 20, 40 ... 1000 epochs. We chose 100 epochs with four neurons and one intermediate layer for this problem.

2) Experimental Setup: We employed Temporal-CV to evaluate our different methods and took fiscal years into

TABLE I: Experimental Settings

Experimental Setting	Variation Explored	
Seven Models	ARIMA, MA, ETS-Plain, ETS-Plus; BSTS-Plain, BSTS-Plus, LSTM	
Data Length	Monthly Model (Up to 18)	2, 4, 6, 8, 10 months; 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 years
	Weekly Model (Up to 21)	2, 4, 6, 8, 10, 12, 24, 36 weeks; 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 years
Two Window Types	Expanding, Sliding	

consideration. Note that a fiscal year is determined by the corresponding organization and thus, we have July-June for FBCENC and November-October for LARFB. For the rest of the paper, $year_t$ denotes the fiscal year starting in year t and ending in year $t+1$. For each $year_t$, we built and evaluated our models using different data lengths *(up to the maximal training data available)* and window types settings shown in Table I. More specifically, to implement Temporal-CV for a year $year_t$, our training data consisted of the most recent data with length l if available before $year_t$. We fit the model on the training data and made a one step ahead prediction for the next step in the time-series. For Temporal-CV at weekly level, we employed the same process and made weekly predictions.

For both monthly and weekly predictions, we applied 13 rounds of Temporal-CV for each of fiscal year from year 2007 to 2019 for FBCENC and 5 rounds from 2014 to 2018 for LARFB. For a given model and corresponding window type, we explored different training data length settings based on the corresponding setting in Table I and the amount of training data available. For example, for $year_{2007}$, an ARIMA with Expanding window would be able to explore *six* settings (five monthly setting + 1 year) and the number would increment 1 for each subsequent fiscal year. Therefore, a total of 156 possible settings can be explored for ARIMA with Expanding window. By combining the two window types, for each model our experiment explored $156 = 312$ settings for monthly prediction for FBCENC. Similarly, for each model, our experiment explored 40 settings for monthly prediction for LARFB. The same general procedure was used for weekly predictions: for each model we explored 195 settings for FBCENC and 55 settings for LARFB respectively.

For each evaluation metric, we compared different models *overall* performance first and then compared their performance by < 24 months vs. ≥ 24 months. This is because consistent with prior research, our results also showed that the length of training data available has a significant impact on models performance. By utilizing results from [27] which showed that 24 months of data delivered the best performance, we further categorized our results by training data length: 1) with less than 24 months of training data and 2) with at least 24 months of training data. For example, to calculate each model's monthly performance using < 24 months for FBCENC, we included a total of 156 setting which is based on 6 data length settings that are less than 24 months ($2, 4, 6, 8, 10$ months and one year) $\times$ two window types $\times$ 13 fiscal years. To calculate the each model's performance using ≥ 24 months data, we used the remaining $312 - 156 = 156$ settings.

3) Evaluation of Prediction Performance: To evaluate and compare the performance of different models with their various settings, we employed two evaluation metrics that are commonly used in the field: *Mean Average Error (MAE)* [56] which is defined in equation 5a and *Mean Average Percentage Error (MAPE)* [24, 44] which is defined in equation 5b.

$$MAE = (\sum_{i=1}^{t} abs(\hat{y}_i - y_i))/t \tag{5a}$$

$$MAPE = (\sum_{i=1}^{t} (abs(\hat{y}_i - y_i)/y_i) * 100)/t \tag{5b}$$

VI. Results and Discussion

Tables II and III show the overall performance of different models at a monthly and weekly level respectively. For each table, the performance for FBCENC is in the upper part of the table and for LARFB is in the lower part. For each foodbank, we compare the *four* classical time-series forecasting models against the *two* BSTS-based and LSTM models and our simple Averaging prediction. In each of these tables, the best model in each category of models is in **bold**; the *best* and the *second-best* models across ALL are labeled with ** and ‡.

Both tables compare different models across all the corresponding settings on the two evaluation metrics: MAE and MAPE. In each metric, we report an average value and a standard deviation for multiple settings of the model. For both monthly and weekly, we first compare all the models' "overall" performance by averaging the model across all the data length and window type settings and then further compare their performance with less than 24 months of training data (columns 4 and 7) and with at least 24 months of training data (columns 5 and 8) respectively. In the following, we will first compare the performance of the seven models against our simple Averaging and then we will shed some lights on the impact of the data length and window type on the task of in-kind donation predictions.

A. Model Performance

1) Monthly: For FBCENC (Upper), Table II shows that amongst the four classical models, ETS-Plain has the best overall MAE performance and also for both with < 24 months and with ≥ 24 months training data; in terms MAPE, ETS-Plain also have the best overall MAPE and with ≥ 24 months training data; ARIMA has the lowest MAPE with < 24 months training data. Among the three advanced methods, BSTS-Plain is the best for both evaluation metrics overall and for ≥ 24 months training data. However, when we have < 24 months training

TABLE II: Performance (±standard deviation) of Monthly Models

Food Bank	Model	Mean Average Error (MAE) (10^3 lb)			Mean Percentage Error (MAPE)		
		Overall	$<$ 24 mon	$\geq$ 24 mon	Overall	$<$ 24 mon	$\geq$ 24 mon
FBCENC	ARIMA	459.00 (±436.66)	473.99 (±458.35)	455.23 (±447.93)	11.69 (±9.07)	**11.96 (±9.21)**‡	11.14 (±9.07)
	MA	464.25 (±432.92)	482.15 (±443.74)	452.88 (±448.66)	11.72 (±8.68)	12.11 (±8.70)	10.99 (±8.74)
	ETS-Plain	**452.28 (±426.55)**‡	**468.99 (±437.61)**‡	**435.58 (±437.60)**	**11.60 (±9.03)**‡	12.00(±9.22)	**10.57 (±8.51)**
	ETS-Plus	455.68 (±425.13)	473.00 (±438.54)	437.62 (±435.11)	11.68 (±9.01)	12.09 (±9.24)	10.61 (±8.45)
	BSTS-Plain	**466.55(±371.12)**	527.76 (±446.38)	**408.59 (±365.27)**‡	**12.31 (±8.97)**	13.67 (±10.02)	**10.28 (±7.80)**‡
	BSTS-Plus	479.48 (±361.76)	529.44 (±448.53)	429.54 (±349.75)	12.61 (±8.96)	13.72 (±10.14)	10.77 (±7.67)
	LSTM	491.57 (±458.40)	**489.40 (±457.29)**	508.21 (±501.95)	12.18 (±8.44)	**12.24 (±8.98)**	11.88 (±8.68)
	Average	**415.97 (±423.68)****	**434.56 (±449.20)****	**406.84 (±419.44)****	**10.69 (±9.13)****	**11.11 (±9.59)****	**9.94 (±8.15)****
LARFB	ARIMA	346.62 (±267.64)	343.34 (±260.87)	367.15 (±347.79)	15.05 (±9.02)	14.91 (±8.77)	15.91 (±11.59)
	MA	**316.60 (±251.90)**‡	**306.64 (±249.87)****	374.83 (±335.08)	**13.75 (±8.85)**‡	**13.32 (±8.78)****	16.25 (±11.51)
	ETS-Plain	330.90 (±273.50)	326.25 (±269.37)	**358.83 (±326.12)**	14.16 (±9.37)	13.99 (±9.28)	**15.31 (±10.84)**‡
	ETS-Plus	336.85 (±273.02)	333.59 (±268.25)	362.04 (±330.42)	14.41 (±9.32)	14.28 (±9.21)	15.44 (±10.80)
	BSTS-Plain	410.53 (±248.83)	406.04 (±256.46)	424.08 (±291.33)	18.08 (±9.32)	17.81 (±9.66)	18.99 (±11.33)
	BSTS-Plus	411.38 (±247.70)	407.68 (±255.73)	418.98 (±275.56)	18.18 (±9.41)	17.92 (±9.70)	18.87 (±10.97)
	LSTM	**326.06 (±256.06)**	**321.86 (±249.04)**	**352.58 (±337.86)****	**14.25 (±8.56)**	**13.99 (±8.15)**	**15.62 (±11.73)**
	Average	**313.84 (±262.80)****	312.11 (±258.60)‡	352.68(±323.19)‡	**13.53 (±9.45)****	13.45 (±9.23)‡	**15.28 (±11.25)****

For each category, the best model is in **bold**; The *best* and the *second-best* models across ALL are labeled with ** and ‡.

TABLE III: Performance (± standard deviation) of Weekly Models

Food Bank	Model	Mean Average Error (MAE) (10^3 lb)			Mean Percentage Error (MAPE)		
		Overall	$<$ 24 mon	$\geq$ 24 mon	Overall	$<$ 24 mon	$\geq$ 24 mon
FBCENC	ARIMA	**311.19 (±263.22)**‡	350.37 (±285.47)	**264.59 (±275.33)****	71.30 (±677.06)	74.55 (±669.57)	41.33 (±257.08)
	MA	336.15 (±302.24)	**347.03 (±307.71)**‡	330.20 (±309.25)	73.81 (±701.37)	75.01 (±700.71)	46.46 (±202.66)
	ETS-Plain	344.36 (±317.85)	370.98 (±333.77)	316.87 (±334.98)	**69.71 (±627.90)**	**73.39 (±629.12**)	**41.23 (±176.25**)
	ETS-Plus	353.63 (±311.58)	376.45 (±340.12)	328.99 (±317.30)	73.40 (±671.00)	76.02 (±671.72)	44.19 (±184.98)
	BSTS-Plain	**338.37 (±269.55**)	376.72 (±306.36)	**285.32 (±302.26)**	75.51 (±748.43)	79.35 (±745.61)	**40.45 (±191.85)**‡
	BSTS-Plus	349.07 (±271.03)	393.17 (±315.68)	286.70 (±301.72)	76.11 (±732.51)	80.51 (±729.70)	40.64 (±191.53)
	LSTM	347.90 (±335.30)	**355.58 (±331.98)**	347.15 (±352.59)	**67.35 (±619.24)****	**68.88 (±619.10)****	42.29 (±163.18)
	Average	**297.54 (±294.84)****	**329.89 (±322.14)****	273.85 (±286.00)‡	67.39 (±682.24)‡	70.64 (±680.55)‡	**38.57 (±195.07)****
LARFB	ARIMA	121.05 (±117.09)	122.70 (±117.63)	115.87 (±127.02)	26.79 (±39.44)	27.11 (±39.70)	26.57 (±41.85)
	MA	**113.57 (±111.53)**‡	**113.86 (±112.31)**‡	115.55 (±123.95)	**25.33 (±38.52)**‡	**25.38 (±38.65)**‡	26.43 (±41.90)
	ETS-Plain	117.00 (±112.98)	118.99 (±114.33)	111.22 (±122.52)‡	25.88 (±36.41)	26.28 (±36.25)	**25.55 (±42.05)**‡
	ETS-Plus	117.35 (±113.04)	119.34 (±114.42)	**111.21 (±122.12)**	25.94 (±36.41)	26.34 (±36.25)	**25.55 (±41.97)**‡
	BSTS-Plain	128.24 (±107.67)	128.57 (±111.22)	129.42 (±120.47)	27.57 (±31.25)	27.94 (±35.01)	27.28 (±28.14)
	BSTS-Plus	128.63 (±107.24)	129.56 (±110.74)	126.00 (±120.73)	27.83 (±32.13)	28.33 (±35.82)	**26.54 (±26.85)**
	LSTM	**116.14 (±110.94)**	**115.52 (±111.41)**	**122.23 (±128.06)**	**25.95 (±37.86)**	**25.49 (±36.82)**	29.28 (±46.43)
	Average	**107.54 (±114.02)****	**108.33 (±115.72)****	**110.82 (±118.62)****	**23.93 (±36.26)****	**24.13 (±37.24)****	**25.07 (±36.10)****

For each category, the best model is in **bold**; The *best* and the *second-best* models across ALL are labeled with ** and ‡.

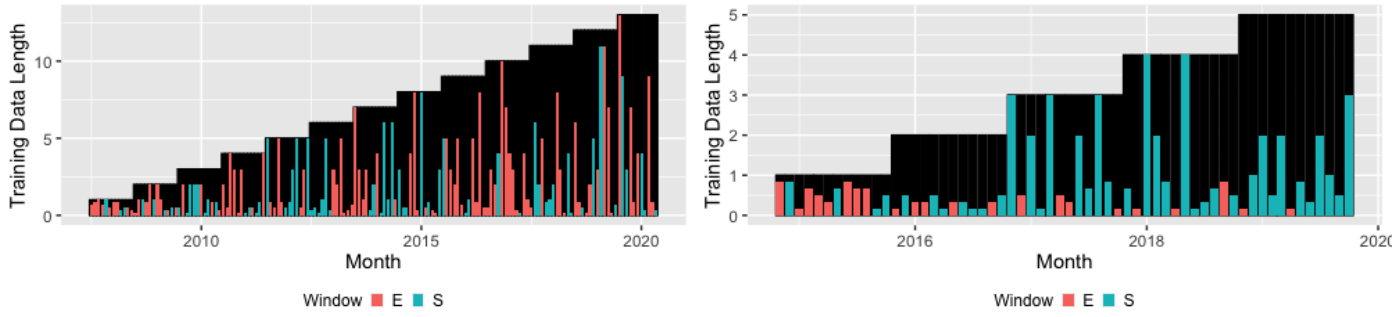
data, LSTM performs the best. Finally, our simple Averaging prediction performs the best across all the models and across the two training data length categories.

For LARFB (Lower), Table II shows that amongst the four classical models MA performs the best overall and for $<$ 24 months training data while ETS-Plain perform the best for $\geq$ 24 months training data. Among the three advanced methods, LSTM is the best across the board. Finally, our simple Averaging prediction performs the best across all the models on MAE and MAPE for overall (columns 3 and 6) and are the best or the second best with very close performance to the best model for either $<$ 24 months data or with $\geq$ 24 months data.

In short, across the two food banks and three categories of models, our simple Averaging prediction performed the best overall.

2) Weekly: Table III compares different models' performance at a weekly level. For FBCENC (Upper), Table III shows that in terms of MAE, amongst the four classical models ARIMA has the best overall performance and also with $\geq$ 24 months training data. However, with $<$ 24 months training data, MA performs the best. In terms of MAPE, ETS-Plain has overall best performance and also for both with $<$ 24 months and with $\geq$ 24 months' training data. Amongst the three advanced methods, BSTS-Plain and LSTM split the best performance. The former has the best overall MAE, MAE with $\geq$ 24 months training data and MAPE with $\geq$ 24 months training data while LSTM has the best MAE with $<$ 24 months training data, the best overall MAPE and MAPE with $<$ 24 months training data. Finally, our simple Averaging prediction performed the best or the second best with very close performance to the best model across the six columns.

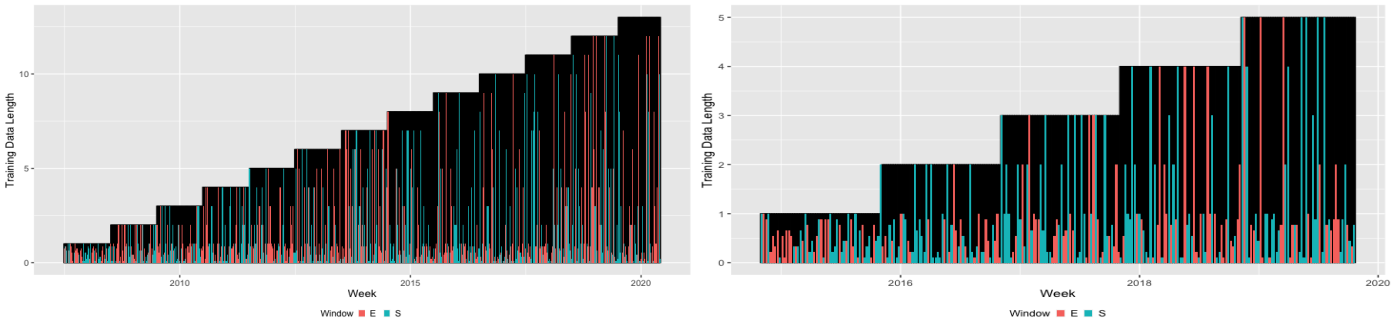
For LARFB (Lower), Table III shows that amongst the four classical models MA performs the best overall and for $<$ 24 months training data while ETS-Plus perform the best for $\geq$ 24 months training data. Among the three advanced methods, LSTM performs the best except on MAPE with $\geq$ 24 months training data. In short, our simple Averaging prediction generally performed the best across all the models on both MAE and MAPE for weekly predictions.



(a) FBCENC (b) LARFB

Fig. 2: Data Length & Window Type Comparison - Monthly

S



(a) FBCENC (b) LARFB

Fig. 3: Data Length & Window Type Comparison - Weekly

B. Data Length & Window Type

The Figure 2 and 3 show the effect of Data Length and Window Type on the weekly and monthly levels respectively. In each figure, the x-axis refers to the month and year, while the y-axis refers to the available training data length (in black) which increases over time. For each month, we pick the best model (with the smallest MAE) and describe the corresponding training data length and the window type involved (colored bars with *red bars* bars for expanding windows while *blue bars* for sliding window). These graphs show that as more training data available, the best model often involves much shorter training data. Furthermore, we can observe that no clear patterns on Data Length or Window Type that would give the optimal performance. In short, this observation suggests that there is a clear need to combine all three factors: Model, Data Length, and Window Type for prediction.

VII. Conclusion and Future Work

The prediction of food donations is necessary to prevent wastage and to provide food-insecure households with equal and timely access to donations. With data on potential donations, food banks can more efficiently allocate manpower and plan storage and distribution of donations. In this work, examined methods of predicting food bank donations involving two food banks with radically different characteristics and at a monthly and weekly level. For this, we did a detailed analysis of the problem involved looking at three aspects: the best predictive model, the optimal amount of historical data, and the type of window that provided the best predictions. We found that no one method gives consistently good predictions of food donations due to the changing nature of the data and the presence of concept drift in the data. Thus, we use an Average of the predictions derived from each combination. Although this approach is simple, it outperformed the base methods for both food banks consistently both on a monthly and weekly basis. From this, we can conclude that for this problem, we must combine predictions generated based on the observed data. Our future work will focus on combining these predictions with more robust meta-learning models. Furthermore, since these in-kind donations are not made by a single donor, but rather by many donors, we will investigate a bottom-up learning approach for predicting these donations in our future work.

References

[1] A. Coleman Jensen, C. Gregory, and A. Singh, "Household food security in the united states in 2018," *USDA-ERS Economic Research Report*, no. 270, 2019.

[2] F. A. Website, "Facts about poverty and hunger in america," *https://www.feedingamerica.org*.

[3] L. Bauer, "The covid-19 crisis has already left too many children hungry in america," *https://www.brookings.edu/blog/*, 2020.

[4] M. Berner and K. O'Brien, "The shifting pattern of food security support: food stamp and food bank usage in north carolina," *Nonprofit and voluntary sector quarterly*, vol. 33, no. 4, pp. 655–672, 2004.

[5] I. S. Orgut *et al.*, "Achieving equity, effectiveness, and efficiency in food bank operations: Strategies for feeding america with implications for global hunger relief," in *Advances in managing humanitarian operations*. Springer, 2016, pp. 229–256.

[6] L. B. Davis *et al.*, "Analysis and prediction of food donation behavior for a domestic hunger relief organization," *International Journal of Production Economics*, vol. 182, pp. 26–37, 2016.

[7] C. Bazerghi, F. H. McKay, and M. Dunn, "The role of food banks in addressing food insecurity: a systematic review," *Journal of community health*, vol. 41, no. 4, pp. 732–740, 2016.

[8] FBCENC, "http://www.foodbankcenc.org."

[9] R. Bennett, "Factors underlying the inclination to donate to particular types of charity," *International Journal of Nonprofit and Voluntary Sector Marketing*, vol. 8, no. 1, pp. 12–29, 2003.

[10] A. Sargeant, "Charitable giving: Towards a model of donor behaviour," *Journal of marketing management*, vol. 15, no. 4, pp. 215–238, 1999.

[11] B. B. Schlegelmilch, A. Diamantopoulos, and A. Love, "Characteristics affecting charitable donations: empirical evidence from britain," *Journal of Marketing Practice: Applied Marketing Science*, vol. 3, no. 1, pp. 14–28, 1997.

[12] G. A. Verhaert and D. Van den Poel, "Empathy as added value in predicting donation behavior," *Journal of Business Research*, vol. 64, no. 12, pp. 1288–1295, 2011.

[13] N. Laptev, J. Yosinski, L. E. Li, and S. Smyl, "Time-series extreme event forecasting with neural networks at uber," in *International Conference on Machine Learning*, vol. 34, 2017, pp. 1–5.

[14] R. Yang. Omphalos, uber's parallel and language-extensible time series backtesting tool.

[15] S. Selvin, R. Vinayakumar, E. Gopalakrishnan, V. K. Menon, and K. Soman, "Stock price prediction using lstm, rnn and cnn-sliding window model," in *2017 international conference on advances in computing, communications and informatics (icacci)*. IEEE, 2017, pp. 1643–1647.

[16] M. G. Elfeky, W. G. Aref, and A. K. Elmagarmid, "Stagger: Periodicity mining of data streams using expanding sliding windows," in *Sixth International Conference on Data Mining (ICDM'06)*. IEEE, 2006, pp. 188–199.

[17] A. Noureldin, A. El-Shafie, and M. R. Taha, "Optimizing neuro-fuzzy modules for data fusion of vehicular navigation systems using temporal cross-validation," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 1, pp. 49–61, 2007.

[18] L. J. Tashman, "Out-of-sample tests of forecasting accuracy: an analysis and review," *International journal of forecasting*, vol. 16, no. 4, pp. 437–450, 2000.

[19] S. P. Schnaars, "A comparison of extrapolation models on yearly sales forecasts," *International Journal of Forecasting*, vol. 2, no. 1, pp. 71–85, 1986.

[20] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A comparison of arima and lstm in forecasting time series," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018, pp. 1394–1401.

[21] B. M. Althouse, Y. Y. Ng, and D. A. Cummings, "Prediction of dengue incidence using search query surveillance," *PLoS Negl Trop Dis*, vol. 5, no. 8, p. e1258, 2011.

[22] H. Al-Hamadi and S. Soliman, "Short-term electric load forecasting based on kalman filtering algorithm with moving window weather and load model," *Electric power systems research*, vol. 68, no. 1, pp. 47–59, 2004.

[23] A. R. Marathe, A. J. Ries, and K. McDowell, "Sliding hdca: single-trial eeg classification to overcome and quantify temporal variability," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 2, pp. 201–211, 2014.

[24] S. Islam, J. Keung, K. Lee, and A. Liu, "Empirical prediction models for adaptive resource provisioning in the cloud," *Future Generation Computer Systems*, vol. 28, no. 1, pp. 155–162, 2012.

[25] R. T. Clemen, "Combining forecasts: A review and annotated bibliography," *International journal of forecasting*, vol. 5, no. 4, pp. 559–583, 1989.

[26] V. R. R. Jose and R. L. Winkler, "Simple robust averages of forecasts: Some empirical results," *International journal of forecasting*, vol. 24, no. 1, pp. 163–169, 2008.

[27] N. Pugh and L. B. Davis, "Forecast and analysis of food donations using support vector regression," in *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 2017, pp. 3261–3267.

[28] I. A. Nuamah, L. Davis, S. Jiang, and N. Lane, "Predicting donations using a forecasting-simulation model," in *2015 Winter Simulation Conference (WSC)*. IEEE, 2015, pp. 1880–1891.

[29] L. G. Brock III and L. B. Davis, "An approach to approximating contributions received from supermarkets by food banks," in *IIE Annual Conference. Proceedings*. Institute of Industrial and Systems Engineers (IISE), 2012, p. 1.

[30] R. Oloruntoba and R. Gray, "Humanitarian aid: an agile supply chain?" *Supply Chain Management: an international journal*, vol. 11, no. 2, pp. 115–120, 2006.

[31] L. N. Van Wassenhove, "Humanitarian aid logistics: supply chain management in high gear," *Journal of the Operational research Society*, vol. 57, no. 5, pp. 475–489, 2006.

[32] S. L. Scott and H. R. Varian, "Bayesian variable selection for nowcasting economic time series," National Bureau of Economic Research, Tech. Rep., 2013.

[33] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, "Bayesian model averaging: a tutorial," *Statistical science*, pp. 382–401, 1999.

[34] J. H. Wright, "Forecasting us inflation by bayesian model averaging," *Journal of Forecasting*, vol. 28, no. 2, pp. 131–144, 2009.

[35] S. L. Scott and H. R. Varian, "Predicting the present with bayesian structural time series," *International Journal of Mathematical Modelling and Numerical Optimisation*, vol. 5, no. 1-2, pp. 4–23, 2014.

[36] C. McQuire, K. Tilling, M. Hickman, and F. de Vocht, "Forecasting the 2021 local burden of population alcohol-related harms using bayesian structural time–series," *Addiction*, vol. 114, no. 6, pp. 994–1003, 2019.

[37] N. Feroze, "Forecasting the patterns of covid-19 and causal impacts of lockdown in top ten affected countries using bayesian structural time series models," *Chaos, Solitons & Fractals*, p. 110196, 2020.

[38] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on lstm recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, 2017.

[39] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by lstm," *Energy*, vol. 148, pp. 461–468, 2018.

[40] Y. Liu, C. Yang, K. Huang, and W. Gui, "Non-ferrous metals price forecasting based on variational mode decomposition and lstm network," *Knowledge-Based Systems*, vol. 188, p. 105006, 2020.

[41] Z. Zhao, W. Chen, X. Wu, P. C. Chen, and J. Liu, "Lstm network: a deep learning approach for short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.

[42] M. Schultz and S. Reitmann, "Prediction of aircraft boarding time using lstm network," in *2018 Winter Simulation Conference (WSC)*. IEEE, 2018, pp. 2330–2341.

[43] S. Yao, L. Luo, and H. Peng, "High-frequency stock trend forecast using lstm model," in *2018 13th International Conference on Computer Science & Education (ICCSE)*. IEEE, 2018, pp. 1–4.

[44] E. Karakoyun and A. Cibikdiken, "Comparison of arima time series model and lstm deep learning algorithm for bitcoin price forecasting," in *The 13th multidisciplinary academic conference in prague 2018 (the 13th mac 2018)*, 2018, pp. 171–180.

[45] A. Skabar and I. Cloete, "Investigation of the effect of training and prediction window sizes on neural financial prediction models," in *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics*, vol. 4. IEEE, 2003, pp. 2133–2137.

[46] Q. Li, C. Rajagopalan, and G. D. Clifford, "Ventricular fibrillation and tachycardia classification using a machine learning approach," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1607–1613, 2013.

[47] A. Bouchachia and S. Bouchachia, "Ensemble learning for time series prediction," 2008.

[48] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[49] R. J. Hyndman *et al.*, "Package 'forecast'," *https://cran. r-project.org/*.

[50] S. L. Scott and H. R. Varian, "Predicting the present with bayesian structural time series," *Available at SSRN 2304426*, 2013.

[51] S. R. Jammalamadaka, J. Qiu, and N. Ning, "Multivariate bayesian structural time series model," *The Journal of Machine Learning Research*, vol. 19, no. 1, pp. 2744–2776, 2018.

[52] A. Harvey and S. Koopman, "Structural time series models," *Encyclopedia of Biostatistics*, vol. 7, 2005.

[53] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[54] R. J. Hyndman, "Measuring forecast accuracy," *Business forecasting: Practical problems and solutions*, pp. 177–183, 2014.

[55] LARFB, "https://www.lafoodbank.org."

[56] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error over the root mean square error in assessing average model performance," *Climate research*, vol. 30, no. 1, pp. 79–82, 2005.