Socio-Economic Planning Sciences 82 (2022) 101041

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

journal homepage: http://www.elsevier.com/locate/seps

Food waste reduction and economic savings in times of crisis: The potential of machine learning methods to plan guest attendance in Swedish public catering during the Covid-19 pandemic[★]

Christopher Malefors[a,*], Luca Secondi[b], Stefano Marchetti[c], Mattias Eriksson[a]

[a] Department of Energy and Technology, Swedish University of Agricultural Science, Box 7032, 75007, Uppsala, Sweden
[b] Department for Innovation in Biological, Agro-Food and Forest Systems (DIBAF), University of Tuscia, Via S.Camillo De Lellis, Snc, Viterbo (Vt), 01100, Italy
[c] Department of Economics and Management, University of Pisa, Via Ridolfi 10, Pisa, 56124, Italy

ARTICLE INFO

Keywords:
Food waste school kitchens forecasting
random-forest system optimization

ABSTRACT

Food waste is a significant problem within public catering establishments in any normal situation. During spring 2020 the Covid-19 pandemic placed the public catering system under greater pressure, revealing weaknesses within the system and generation of food waste due to rapidly changing consumption patterns. In times of crisis, it is especially important to conserve resources and allocate existing resources to areas where they can be of most use, but this poses significant challenges. This study evaluated the potential of a forecasting model to predict guest attendance during the start and throughout the pandemic. This was done by collecting data on guest attendance in Swedish school and preschool catering establishments before and during the pandemic, and using a machine learning approach to predict future guest attendance based on historical data. Comparison of various learning methods revealed that random forest produced more accurate forecasts than a simple artificial neural network, with conditional mean absolute prediction error of <0.15 for the trained dataset. Economic savings were obtained by forecasting compared with a no-plan scenario, supporting selection of the random forest approach for effective forecasting of meal planning. Overall, the results obtained using forecasting models for meal planning in times of crisis confirmed their usefulness. Continuous use can improve estimates for the test period, due to the agile and flexible nature of these models. This is particularly important when guest attendance is unpredictable, so that production planning can be optimized to reduce food waste and contribute to a more sustainable and resilient food system.

1. Introduction

The Covid-19 pandemic revealed an urgent need to investigate how the food supply chain reacts during a crisis, when it is necessary to save resources and allocate them to where they are most needed [1]. The food system is currently under stress on all levels, from food service provider to farms. This requires the system to be quite robust and sufficiently flexible to allow swift changes due to rapidly changing consumption patterns, in order to avoid building large imbalances into the system that eventually generate shortage of food in some places and food waste in others. Only by reducing these imbalances can the food system develop in a sustainable way, ensuring food security without unnecessary resource consumption.

The global food system is a major driver of land use change [2,3], depletion of freshwater resources [4,5], climate change [6,7], and pollution of aquatic and terrestrial ecosystems through excessive nitrogen and phosphorus inputs [8–10]. The current steep trajectories of population and consumption growth further increase the importance of finding solutions that can meet food demand in a sustainable fashion [11]. Five major areas have been identified as possible ways of achieving a sustainable food system by 2050 [12]. However, predicted climate change will alter the conditions for food production and future crises of different types that impact food supply and food consumption in one way or another are likely, so it is highly important to understand effects arising from disruptive events, like the Covid-19 pandemic. Unlike many other countries, Sweden did not close down preschools and primary

★ The article was handled by Editor in Chief, Vedat Verter

* Corresponding author.
E-mail address: christopher.malefors@slu.se (C. Malefors).

https://doi.org/10.1016/j.seps.2021.101041
Received 17 December 2020; Received in revised form 17 February 2021; Accepted 23 February 2021
Available online 2 March 2021

schools in the initial phase of the Covid-19 pandemic. This created a somewhat unique situation where schools were open but had to apply restrictions, e.g., forcing pupils with the slightest symptoms to remain at home for at least two days after becoming symptom-free. This created a situation with much lower rates of guest attendance in school canteens than usual, potentially leading to large volumes of overproduction unless the schools were agile enough to adjust their production swiftly. As the pandemic also led to illness among kitchen staff, the work environment in canteens was likely to be more stressed than usual, making it unlikely that reducing food waste by saving and re-using leftovers was prioritized. However, in times of crisis the food system needs to be resilient and not waste resources, but allocate them to where they are most needed. Better forecasting of guest attendance has been suggested as a way to reduce food waste in normal situations [13], but knowledge is lacking on how feasible such forecasting is in abnormal situations.

The aim of this study was therefore to assess the accuracy with which a machine learning model could predict guest attendance in Swedish primary school canteens during the initial phases of the Covid-19 pandemic. The goal was to contribute to a more resilient and sustainable food system by developing knowledge on how to make production planning more resource-efficient in times of crisis.

2. Consumption pattern and waste during crisis

Previous food waste studies of the hospitality sector report that roughly 20% of food produced ends up as waste [14–20]. This should be considered the normal level, and could be expected to rise when consumption patterns and guest attendance shift rapidly. Little is known of the situation during the initial phase of the Covid-19 pandemic, as statistics have not yet been compiled. However, in a normal situation public catering is a significant actor in Sweden, with approximately 50% of all midday meals in the food service sector served in the public sector [21], by catering establishments in hospitals, preschools, schools, and elderly care units. School kitchens are the most significant actor in the public catering segment, with almost 1.3 million meals served to Swedish pupils every day throughout the school year. School meals are provided free of charge, funded by taxes [22]. Similar approaches to school meals are used in Finland [23] and Estonia [24]. Other countries serve also school meals, although at a cost to the guest, so the concept of school meals is widespread. In Sweden, planning for the number of meals to serve is generally based on the number of pupils registered in the school, but despite these optimal conditions, fluctuations in canteen guest numbers still occur. This is one of the risk factors for food waste generation identified previously [25], since kitchen staff do not get timely information regarding the number of guests.

Previous studies have also shown that kitchen staff and food service providers generally add an extra margin in meal production [25,26] in order to avoid running out of food, which is seen as a negative outcome in the eyes of the guests and a source of shame for kitchen staff. The problem is therefore two-fold; there is little knowledge about how many guests will turn up to a specific meal, and the kitchen needs to prepare an acceptable margin of food to avoid shortages and loss of goodwill. Reduced food waste is important to decrease the environmental impacts from the food system, as it reduces the overall amount of food that needs to be produced [27–29]. United Nations Sustainable Development Goal (UN-SDG) 12.3 is to halve per-capita consumer and retail food waste by 2030, and to reduce food losses along production and supply chains. Several studies (e.g. Refs. [12,30]) cite significantly reduced food waste levels as a prerequisite for a sustainable food system in the future. However, such studies provide few suggestions on how exactly to achieve the transition to a more resource-efficient society.

In times of crisis and associated changes in food consumption patterns, the risk of food waste increases, since the food system is not agile enough to adapt to rapid changes. The current pandemic may not occur again, but society is likely to face future crises that impose stress on the food system. It is therefore important to gain as much knowledge as possible on consumption patterns during the current crisis, in order to be prepared for the next crisis.

2.1. Food planning, management and forecasting: the integration with a statistical perspective

The challenge of sustainable production and consumption can also act as a stimulus for enhancing the linkage between rational food planning and the use of statistical methods for obtaining accurate predictions of food production, therefore improving organization and management, and ultimately reducing food waste generation.

Food production is a complex process where uncertainty is relevant, mainly due to stochastic supply and demand, and variability in raw materials and ingredients [31]. This results in differences between planned production and actual output, with important economic, social, and environmental impacts.

Several statistical methods and approaches for forecasting food production have been proposed in the literature, some specifically referring to meal requirements in restaurants, catering services [32,33] and the overall food service sector [34]. The importance of accurately predicting expected meal consumption in the overall hospitality sector has been highlighted by the Covid-19 pandemic [35]. Use of known past data for estimating and forecasting one or more future values in the data series would allow private managers and public bodies to plan the right amount of food to buy and produce, and to adjust staff levels so that food can be prepared and served efficiently [36]. Time series forecasting through regression models [37], operation research and machine learning methods [31], and the Prophet algorithm [36,38] have proven to be useful for this purpose.

The machine learning statistical framework of analysis has already been used for food waste management in the context of driver assessment for household food waste [39]. In this study, we specifically focused on the potential of the random forest approach [40] for predicting guest attendance and meal planning in Swedish public catering, using an estimation strategy that first considered the overall available data and then data at kitchen level, as detailed in section 3.

3. Material and methods

The work was carried out in three steps: i) Collecting data on guest attendance before and during the initial phase of the Covid-19 pandemic; ii) modeling number of guests using a machine learning approach; and iii) evaluating the model in relation to actual guest attendance.

3.1. Data

The complete dataset available for the analysis covered 18 primary school kitchens and 16 preschool kitchens in Sweden. These were selected as suitable test subjects for forecasting models because they were willing to share their data upon request and because most had data available for several years prior to the pandemic. Sixteen of the primary schools and all of the preschools belonged to one municipality, while the remaining two schools belonged to two other municipalities. Thus, no random selection of units was performed. All of the selected kitchens serve meals to pupils ranging in age from around 1 to 15 years. The analysis focused on lunch, as this is the most commonly served meal in the selected kitchens, although breakfast and snacks may also be served. A Swedish school year consists of at least 178 days between late August and early June [41]. The autumn semester includes one week of holiday in late October/early November. A winter holiday of around three weeks covers the Christmas and new year period. The spring semester has one holiday week in February and one week around Easter, plus scattered national holidays from early May to June. Schools do not provide teaching during holidays but often remain open to provide childcare for younger pupils, and therefore serve meals during these periods. The number of meals served during holidays is therefore lower than during ordinary school time. Preschools are kept open during holidays, but guest attendance can still be influenced if families with older siblings use the school holidays for vacation. A typical school kitchen offers pupils two to three lunch options. The menu follows a five-to seven week cycle of menus agreed at meetings between the heads of catering and public catering managers. This study focused on the initial phase of the Covid-19 pandemic in Sweden, which was defined as starting on March 18, 2020, when the Swedish government introduced a strategy for social distancing, including restrictions on school attendance, and ending on June 9, 2020, which marked the official end of the spring semester.

Data on the number of guests eating school meals from 2010 to the beginning of June 2020 were collected by the municipalities themselves, by counting plates after every lunch. The reason for counting plates is that this number is used for internal accounting, where the public catering organization charges the school organization for the lunches. There are several methods for counting the number of plates. One involves drawing tally marks on the dishwasher for each full tray, and then multiplying the total by the capacity of each tray (usually 18 plates). Another approach is to collect one plate from each full tray and all plates from the last incomplete tray and calculate the total number of plates. In this study, it was assumed that one plate was equal to one portion, which was equivalent to one guest. However, this might not necessarily be accurate, since guests are allowed to re-fill their plate or take several plates. Plate data are thus an approximation of the number of guests served, since no point-of-sale data are available to extract information regarding guest flow as the meals are free of charge to pupils.

3.2. Machine learning-based models

A strategy and methodology for forecasting guest attendance (portions) for a given kitchen in Sweden were developed using a sub-set of the data covering the period August 2019–June 2020. The data collected by the kitchens were transformed into a standardized format similar to that proposed by Ref. [42]. Since all field trips and other extracurricular activities were canceled during the pandemic, all holidays and extreme values defined as falling outside the inter-quartile range per school and school semester were removed before entering the modeling step. This is because kitchens already have a good understanding of guest seasonality and of when guests will be missing due to upcoming holidays, but struggle with the variability arising on normal weekdays. The intention was for the filter to remove known features (such as holidays and known study visits in the period prior to the pandemic) from the dataset, in order to focus on the modeling aspects.

With the filter in place, a strategy for estimation going from the pooled data set to the single kitchen data was applied. First, the models were calculated by considering all kitchens and an attempt was made to forecast number of portions using only school and kitchen features. Second, some features relating to the pandemic crisis in Sweden were added. Two kitchens were selected as examples, both serving food in primary schools, the first to children aged 6–12 years (Kitchen 1) and the second to children aged 6–9 years (Kitchen 2). The set of features can be denoted as inputs (predictors or independent variables), which are measured or pre-set and have some influence on one or more outputs (responses or dependent variables). The goal is generally to use the inputs to predict the values of the outputs. This type of approach is called supervised learning. The opposite is "unsupervised learning", where there are only features and no measurement of outcome. The target in this step of the present work was to forecast guest attendance through number of portions, and therefore a machine learning approach was used. Machine learning approaches are designed to make the most accurate predictions possible, whereas statistical approaches are designed to allow inferences about the relationships between variables to be drawn.

3.2.1. Machine learning-based models: the random forest approach

The classical machine learning approach is based on training and testing data. The training data are used to tune a prediction model that is able to correctly predict the test data. Here, the times series of data on portions served in school canteens over time were split into two parts. The first part of the time series was used as training data to tune a machine learning model that could effectively predict the "last" part of the series as accurately as possible. If the last part of the series was well fitted, this was taken as an indication of the capability of the model to forecast portions in future unobserved time.

There are many different machine learning methods for forecasting time series. Those pre-tested in this study were artificial neural networks, Poisson auto-regressive models, and random forests. The best results in terms of accuracy of predictions were obtained with the random forest approach.

Random forest is a tree-based method involving stratification or segmentation of the space of inputs into a number of simple regions. With the aim of making a prediction for a given observation, the mean of the training observations in the region to which it belongs is used. The set of splitting rules used to segment the predictor space can be summarized in a tree, so these types of approaches are known as tree-based methods. Since simple tree-based methods have proven not to be completely effective for prediction purposes, bagging, boosting, and random forest have been introduced. Each of these approaches involves producing multiple trees, which are then combined to yield a single consensus prediction that is very effective [43]. A tree consists of a series of splitting rules, starting at the top of the tree. The necessary steps for building trees can be summarized as follows:

1. Use recursive binary splitting to grow a large tree on the training data, stopping only when each terminal node has fewer than some minimum number of observations.
2. Apply cost complexity pruning to the large tree in order to obtain a sequence of best sub-trees, as a function of α.
3. Use K-fold cross-validation to choose α. That is, divide the training observations into K folds. For each k in 1 to K:
   (a) Repeat Steps 1 and 2 on all but the kth fold of the training data.
   (b) Evaluate the mean squared prediction error on the data in the left-out kth fold, as a function of α.
   (c) Average the results for each value of α, and pick to minimize the average error.
4. Return the sub-tree from Step 2 that corresponds to the chosen value of α.

The recursive binary splitting step (point 1) consists of selecting the input X_j and the cut-point s such that splitting the input space into the regions $\{X||X_j < s\}$ and $\{X||X_j \geq s\}$ leads to the greatest possible reduction in the residual sum of squares (RSS) of the output. That is, all inputs $X_1, \dots, X_p$ and all possible values of the cut-point s for each of the inputs are considered, and then the input and cut-point giving the tree with the lowest RSS are chosen. Without loss of generality, for any j and s, the following pair of half-planes are defined:

$$R_1(j,s) = \{X||X_j < s\} \text{ and } R_2(j,s) = \{X||X_j \geq s\} \tag{1}$$

by searching the value of j and s that minimize the following quantity:

$$\sum_{i:x_i \subseteq R_1(j,s)} \left(y_i - \widehat{y}_{R_1}\right)^2 + \sum_{i:x_i \subseteq R_2(j,s)} \left(y_i - \widehat{y}_{R_2}\right)^2 \tag{2}$$

where $\widehat{y}_{Ri}$ is the mean output for the training observations in $R_i(j,s)$. The last split regions are called terminal nodes. Let the large tree built as described above (point 1) be T_0. This tree can over-fit the data, obtaining poor performance predicting in predicting the test data. Therefore, it is pruned back to obtain a sub-tree in such a way that minimizes the error in the test data. This procedure is carried out through the complexity pruning step (point 2) described above. For a sequence of trees indexed by a non-negative tuning parameter α, each α has a sub-tree $T \subset T_0$ to minimize:

$$\sum_{m=1}^{|T|} \sum_{x_i \subseteq R_m} \left(y_i - \widehat{y}_{R_m} \right)^2 + \alpha |T| \tag{3}$$

where $|T|$ represents the absolute number of terminal nodes of the tree T and R_m is the region corresponding to the m^{th} terminal node.

Construction of trees following the procedure described above suffers from a high level of variability. From the perspective of reducing dispersion, and therefore the uncertainty of the estimates, the use of bootstrap aggregation (also called bagging) leads to a consistent reduction in the variance through the basic rule that averaging a set of observations leads to a reduction in variability. As a result, a natural way to reduce the variance, and therefore increase the prediction accuracy of a statistical learning method, is to take a large number of samples from the training set, build separate prediction models using each sample, and average the resulting predictions.

In addition to bagged trees, random forests provide an improvement by de-correlating the trees. Similarly to bagging, a number of decision trees is built on bootstrapped training samples, by considering each time a split in a tree, i.e., a random sample of l inputs chosen as split candidates from the full set of p inputs. The split can be used only on those l inputs. A fresh sample of l predictors is taken at each split, typically setting $l \approx \sqrt{p}$. This method, even though it sounds counter-intuitive, leads to construction of uncorrelated trees, thus greatly reducing the variance by averaging them. In contrast, the bagging method may build many correlated trees, which does not lead to a reduction in variance. Variance of the average of B identically distributed random variables is $\rho\sigma^2 + B^{-1}(1-\rho)\sigma^2$, where ρ is the pairwise correlation coefficient and σ^2 is the variance of a variable.

The algorithm for the random forest approach can be summarized in the following steps:

1. For $b = 1$ to B:
   (a) Draw a bootstrap sample of size N from the training data.
   (b) Grow a random-forest tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size is reached. Then:
      i. Select l inputs at random from the p inputs.
      ii. Pick the best input/split-point among the l.
      iii. Split the node into two daughter nodes.
2. Output the ensemble of trees T_b, $b = 1, \ldots, B$.

As a result, a prediction for a given point x is obtained as $\widehat{f}_{rf}^{B}(x) = B^{-1}\Sigma_{b=1}^{B} T_b(x)$.

3.3. Selected measures and tools for forecasting statistical and economic validity

To assess how well the test data were predicted through the trained random forest, we calculated mean squared error (MSE), defined as:

$$MSE = T^{-1} \sum_{t=1}^{T} \left(y_t - \widehat{y}_t \right)^2 \tag{4}$$

where y_t is the number of portions served the tth day and $\widehat{y}_t$ its prediction, and T is the number of days in the test dataset. The lower the MSE, the better the model. We also computed mean absolute prediction error for the test and training dataset as:

$$MAPE = T^{-1} \sum_{t=1}^{T} \left| y_t - \widehat{y}_t \right| / y_t \tag{5}$$

For series which includes $y_t = 0$ this measure cannot be used. An alternative approach is to compute conditional MAPE (cMAPE) for $y_t \neq 0$,

$$cMAPE = T^{-1} \sum_{t=1}^{T} \left(\left| y_t - \widehat{y}_t \right| \Big/ y_t \right) I(y_t \neq 0) \tag{6}$$

where $I(y \neq 0)$ is an indicator function equal to one when $y_t \neq 0$, and 0 otherwise. To compare the prediction power of the random forest between the kitchens, the coefficient of variation (CV) was calculated as the ratio between the square root of MSE and mean number of portions served in the relevant period.

With the aim of evaluating the practical implications and impact of our models and their usefulness in reducing economic, environmental (and social) impacts related to food waste, we estimated the economic costs (and the associated savings) for the individual kitchen in three different scenarios:

i. A "no-plan" scenario, representing a situation in which each kitchen prepares every day a number of meals equal to the total number of pupils enrolled.
ii. An "actual" scenario, representing the actual number of meals served, with the costs computed based on the daily information available in the dataset.
iii. A "forecast" scenario, representing the number of meals forecast by our random forest model.

By assigning to each meal a cost of 2 Euros (corresponding to 22 Swedish krona, the average cost per meal in the municipalities in 2019), we computed the progressive cumulative sum of costs for preparation of meals on each of the days in the test period. This value represented a first measure of the total costs incurred by kitchens in the three different scenarios. We then compared the three daily series of costs both by analyzing the differences between the actual scenario and the forest scenario, and by comparing each of these two series with the no-plan scenario. The median Absolute Relative Bias (ARB) was used for this purpose and was calculated as the relative difference between the daily actual and predicted economic values.

4. Results

4.1. Guest attendance during the pandemic

Fig. 1 displays changes over time in the number of guests in all school kitchens studied, and the deviation in the number of guests during the pandemic in contrast to before the pandemic. Three periods are shown in the diagram, the time before the pandemic, the initial stage of the pandemic between 18 March (indicated by the dashed vertical line in Fig. 1) and the Easter holidays, and the steady state of the first wave from the Easter holidays to the summer holidays. The number of pupils enrolled remained constant during the period, as this number normally does not change during the school year. However, the average attending school lunch was on average 181 pupils/day before the pandemic, 144 pupils/day during the initial phase, and 185 pupils/day during the steady state. There was also a clear drop in guest attendance during the first day of the initial phase (i.e., 19 March), where the guest attendance was only 50% of that in previous weeks. However, this drop in guest attendance was short-lived; aggregated to weekly average, the drop in guest attendance in week 12 was only 25% compared with the previous week. However, this was still noticeable in comparison with the average during the years prior to 2020, where a drop in guest attendance during week 12 as not visible at all.

In the school canteens studied, the guest attendance during the steady state period was almost back to normal, although fewer pupils than normal attended daycare in the Easter holiday (Fig. 1). This is in contrast to the preschools, where there was a clear drop in guest

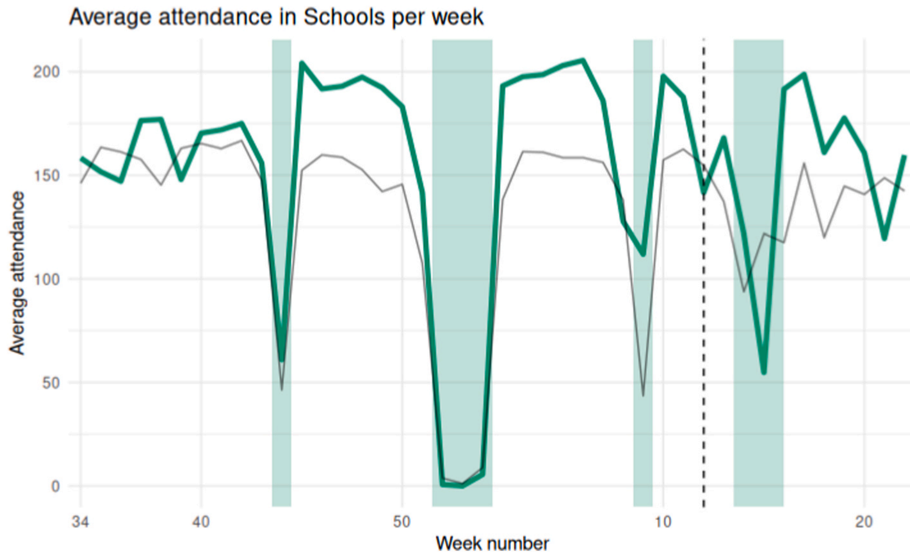

Fig. 1. Average number of primary school meal guests over time on a weekly basis before and during the pandemic. ( ) indicates data for the school year 2019/2020 and ( ) indicates data from previous school years (2010–2018). The dashed line is the day (18 March) when Sweden introduced guidelines for establishments operating in the public catering sector. The shaded area ( ) indicates school breaks, with less activity.

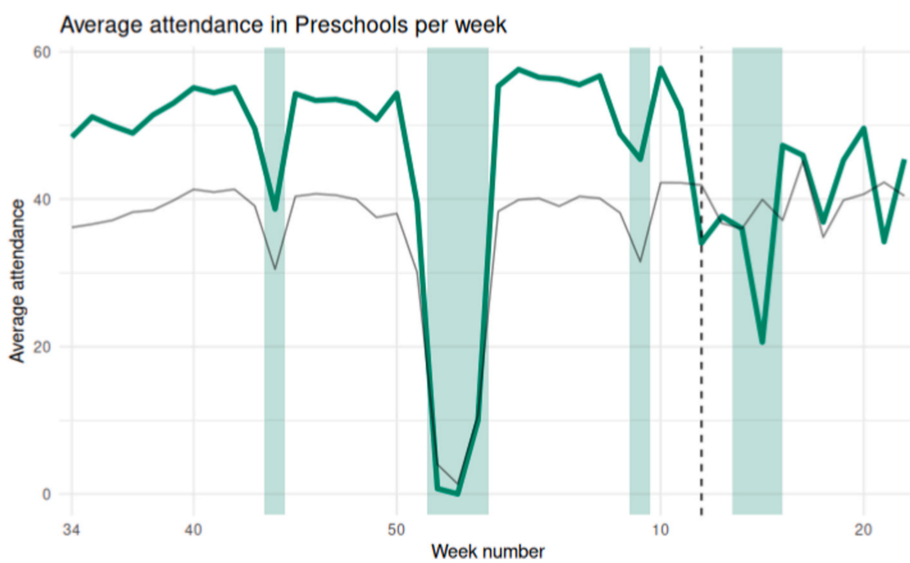

Fig. 2. Average number of preschool meal guests over time on a weekly basis before and during the pandemic. ( ) indicates data for the school year 2019/2020 and ( ) indicates data from the previous school years (2010–2018). The dashed line is the day (18 March) when Sweden introduced guidelines for establishments operating in the public catering sector. The shaded area ( ) indicates school breaks, with less activity.

attendance during both the initial phase and the steady state (Fig. 2). There were also fewer children attending preschool during the week before Easter holiday (even though this is not an actual Easter holiday in preschools, which is clear in the previous year).

4.2. Model specification, estimation, and validation

With the aim of showing the adequacy of the selected method and in order to achieve the objective of the present analysis, we first analyzed the "pooled" dataset (i.e., the dataset summarizing, for each day, the total number of meals served in the 18 kitchens considered in the analysis) and then focused on single kitchens, carrying out separate analyses, estimates, and forecasting.

We started the estimation process by fitting a random forest with all the available inputs: Holiday (categorical with level Bank holiday, Vacation, Planning day, and None), Year (categorical, 2019, 2020), Year day (numerical, number of days in the year, e.g., 4 January is day 4, 28 of December is day 362), Quarter (categorical), Month (categorical), Trend (numerical), Weekday (categorical, from Monday to Sunday), Weekend (categorical, yes, no), and Week (numerical, number of weeks in the year). The random forest model was fitted in the R environment [44], using the package randomForest [45]. Once we had estimated the model for the entire dataset, we concentrated on the two selected single kitchens with the aim of obtaining more accurate predictions. For the two individual kitchens, we selected the inputs Holiday, Year day, Month, Weekday, Weekend, and Week and assessed which were the most important in the final model. We drew a 500 bootstrap sample on which we grew random forest trees, selecting at random two inputs from among the six in the model. We then used the final random forest model to try to predict the test data.

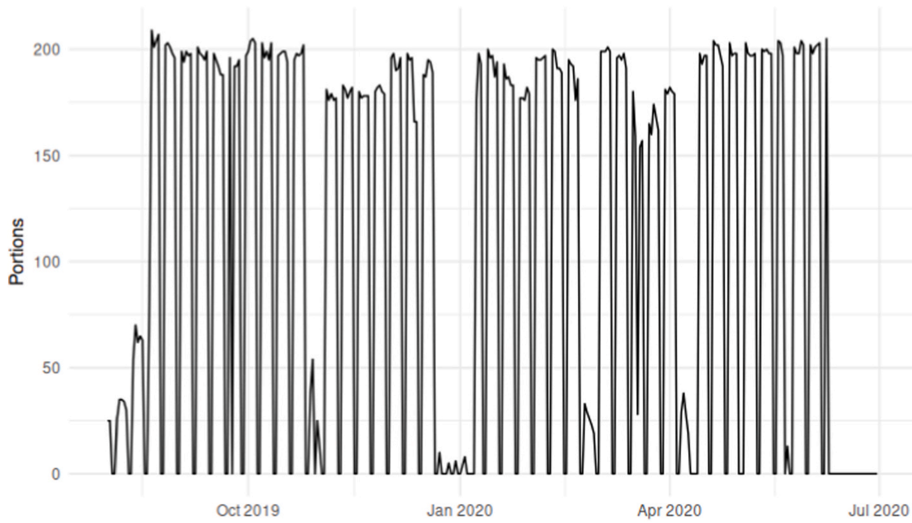

Fig. 3. Time series of portions served by Kitchen 1 in the period 1 August 2019–30 June 2020.

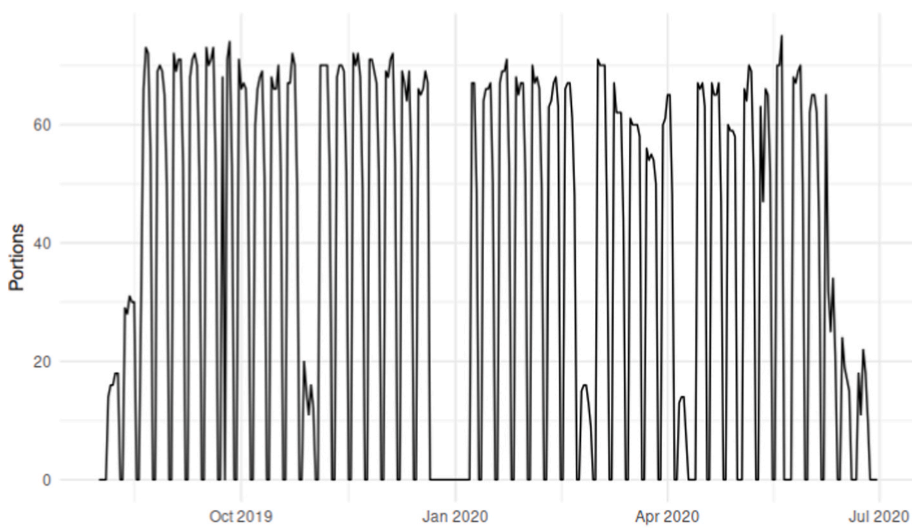

Fig. 4. Time series of school meal portions served by Kitchen 2 in the period 1 August 2019–30 June 2020.

4.2.1. *Definition of the training and testing datasets*

The results for the two individual kitchens, Kitchen 1 and Kitchen 2, are displayed in Figs. 3 and 4, respectively. During the period covered by the analysis, the schools were closed to some extent during 73 vacation days (where the school is open to provide childcare for some pupils), two planning days (where only teachers are present in the schools), 48 weekends, and four bank holidays (where the school is completely closed). The dataset had 125 missing days (mainly weekends) for Kitchen 1 and 119 for Kitchen 2, which were assigned a value of zero school meal portions served.

From Fig. 3, it can be seen that Kitchen 1 served zero portions during weekends and few portions in August, late October/early November, Christmas holidays, end of February, and middle of April. At the beginning of June, the number of portions served fell to zero (due to the kitchen closed down for summer holidays). On other days, the number of portions served varied between 154 and 209.

The series of portions served by Kitchen 2 was similar to that for Kitchen 1. The main difference was zero portions served during Christmas holidays and some portions served in June (Fig. 4) due to day care activities. In addition, Kitchen 2 served a lower number of daily portions, ranging between 44 and 75 during the normal period.

We divided the series into two parts (training set and test set). We tuned the random forest model for each kitchen using the training data and then used the model to predict the data in the test set. The test dataset for Kitchen 1 was set as 21 May-8 June 2020, since days after 8 June had zero portions served. The test dataset for the Kitchen 2 covered the period 9–28 June 2020. The training dataset for Kitchen 1 covered the period 1 August 2019–20 May 2020, while that for Kitchen 2 covered the period 1 August 2019–8 June 2020.

4.2.2. *Random forest estimation for the pooled data set and at kitchen level: statistical and economic validation*

For the pooled dataset for all kitchens studied, MSE was equal to 59322 for the training data and 3036 for the test data. Conditional MAPE (cMAPE) was equal to 0.39 and 1.28 for training and test data, respectively. The random forest approach was applied at kitchen level to capture the characteristics and peculiarities of each series. This approach made it possible to analyze in depth the statistical and economic validity of the estimated models and to compare model performance for different daily series and kitchens.

For the datasets at kitchen level, MSE for Kitchen 1 was around 609 for the test data and 175 for the training data, and cMAPE was around 0.448 and 0.137 and for the training and test data, respectively. Using a very simple neural network with one hidden layer, the MSE of the test

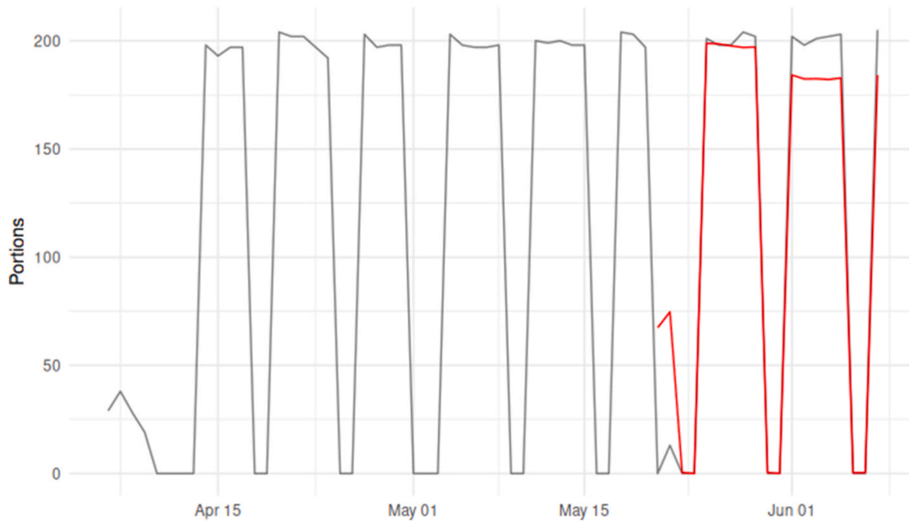

Fig. 5. Number of school meal portions served by Kitchen 1 from April 2020 to June 2020. Black line: observed data, red line: predicted data. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

data was around 11600 and cMAPE was around 4.104.

Fig. 5 shows the time series from April 2020 to June 8, 2020 (black line) with the superimposed red line of the random forest predictions for the test data (21 May-8 June). As can be seen, the last 19 days of the series were rather well predicted by the model. Using more inputs related to socioeconomic contextual variables (such as characteristics of the pupils attending the school) could probably improve the prediction ability of the model.

To assess whether older data improved the prediction power of the random forest model, we tried to predict the same test data using training data on the number of portions served from August 1, 2017. For Kitchen 1, the inputs Weekday and Weekend were initially the most important, followed by Holiday. Based on the relative importance of the metrics, we decided to build a final random forest model using the inputs Holiday, Year, Day of the year, Month, Weekday, and Weekend. Using a longer run (three years) changed the importance of the inputs, with Holiday being the most important input, followed by Weekdays and Weekend, while Month, Year and Day of the year were least important. The MSE of the test set was around 555, slightly better than obtained using the model trained on a shorter run (one year, MSE = 609).

We also checked whether including information on the pandemic stages improved the prediction power. To do this, we added two inputs: number of daily new Covid-19 cases in the municipality where the kitchen was located and a dummy identifying days before or after March 19, 2020, which corresponded to the beginning of the restrictions in Sweden. However, neither of these inputs improved the prediction power of the random forest model.

In an economic assessment, we computed the cumulative sum of kitchen costs for serving meals in the three different scenarios presented in section 3.3. Fig. 6 shows the distribution of meal costs over the test period for Kitchen 1 (May 21-June 8) on comparing the daily progressive meal costs in the no-plan scenario (dashed red line in Fig. 6), the actual scenario (blue bars), and the forecast scenario (orange bars). Two different types of comparisons were carried between the three scenarios. The results confirmed the effectiveness of the forecasting method from an economic perspective. The expected cumulative costs for serving meals in the entire test period were found to be 4454 Euros for the actual (real) data and 4461 Euros for the forecast data. This demonstrated good accuracy of prediction and high levels of economic savings (and distance) compared with the total costs that the kitchen would have incurred if no planning (i.e. no-plan scenario) had been implemented (costs = 5382 Euros).

The alignments between actual and forecast data also showed the economic value of avoiding waste. In Fig. 6, the distance for each day

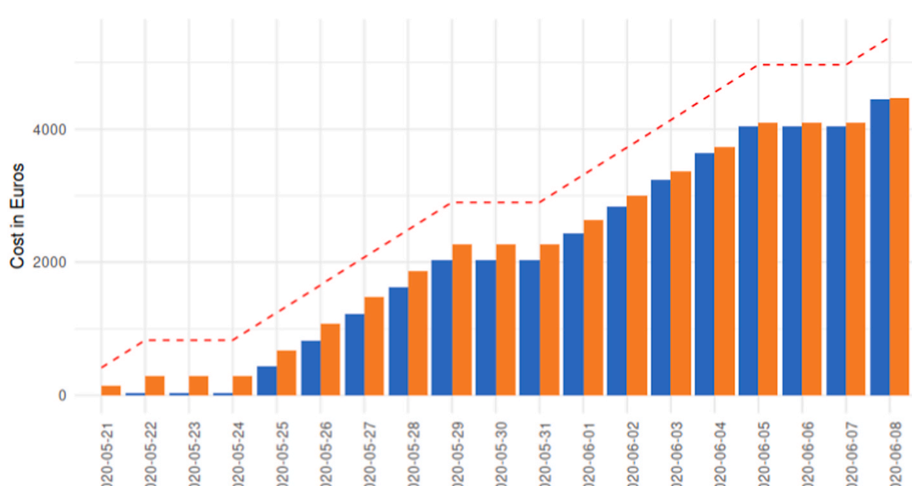

Fig. 6. Results of economic scenario analysis for Kitchen 1. (■) Sum of meal costs based on actual data. (■) Sum of meal costs based on forecasted data. (- - -) Sum of meal costs based on total enrolled students.

between the dashed red line (no-plan scenario) and the forecast scenario (orange bars) represents the costs avoided by adopting a machine-learning approach to forecasting meals. The total saving for Kitchen 1 was approximately 921 Euros, compared with approximately 928 Euros saved from using the series of actual (real) data (blue bars). Median ARB was equal to 0.0884%, thus confirming the potential of the machine learning model for food planning and management.

Prediction using random-forest models was also performed for Kitchen 2 (Fig. 7). Based on availability of data, the test data period was 9–28 June 2020,[1] after which there was a sharp decrease in the actual number of meals served (see Fig. 4). The training set covered the period 1 August 2019–8 June 2020. Other inputs in the initial random forest model were the same used for Kitchen 1, as were the inputs of the refined model (selected using the importance metric).

The MSE was around 19 for the training data and 107 for the test data, while cMAPE was 0.095 for the training data and 0.487 for the test data.

Economic validation of meal forecasting for Kitchen 2 (Fig. 8) also revealed the overall discrete adherence between the cumulative sum of kitchen costs incurred during the test period, which was approximately 532 Euros based on the actual data (therefore the real costs of served meals) and approximately 774 Euros based on the estimated random-forest results. In contrast, the theoretical cost of meals in the no-plan scenario was 2072 Euros.

Thus the total saving from predicting meals through the machine learning approach during the test data period (9–28 June 2020) was approximately 1298 Euros, compared with 1540 Euros based on observed data. It is worth noting that although we found greater over-estimation of number of meals served on the last days of the test data period, the average daily deviation from the costs of the no-plan scenario (i.e., cooking meals for all pupils enrolled) amounted to 64 Euros for the forecast values and 77 Euros for the actual values. The median ARB value (0.296) confirmed the high level of variability in the estimates, with the first quarter of the differences having relative bias <0.15.

The coefficient of variation (CV) for Kitchen 1 was 0.1186 and 0.2106 for the training and test data, respectively, while for Kitchen 2 it was 0.1165 and 0.7773, respectively. Thus, the CV of the training set was quite similar and small for both kitchens, indicating that the random forest models were well trained. The CV of the test set was higher, particularly for Kitchen 2. The cMAPE values showed similar trends. As can be seen in Fig. 7, the prediction line (red) was lagged with respect to the actual data (black line), so the prediction was not so good. It was unclear whether this was related to pandemic-specific circumstances. Unfortunately, adding inputs related to the pandemic (number of daily new cases and the dummy) did not improve the prediction power of the model.

5. Discussion

Based on data on guest attendance for meals in school and preschool canteens, it is clear that the pandemic period started exactly after the press conference announcing restrictions in Sweden. The day after this press conference saw the lowest attendance of children during a normal week day in the schools and preschools studied. However, by the second day after the start of the restrictions guest attendance started rising again and in the weeks just after the start of the restrictions the attendance in schools was almost back to normal, given that the time just before a holiday normally has a low level of guest attendance [13]. During the pandemic steady state phase between Easter holiday and summer holiday guest attendance went down, but it normally does so especially during the weeks including national holidays in May. Thus during this period the pattern of guest attendance was close to normal for the period, as if there were no pandemic going on at all.

In preschools, it was more apparent that there was a pandemic going on, with guest attendance after the restrictions were announced down to around 70% of what would be expected based on the regular pattern during April–June. Thus preschools seemed to suffer more long-lasting effects of the pandemic or the restrictions, probably due to three factors. First, school attendance is compulsory and it is actually illegal for parents to keep pupils away from school if they are not sick, while there is no such legal obligation for preschool children. Second, younger children are well-known for catching other corona viruses causing regular colds and runny noses, which are normally not considered serious diseases, but affected children were restricted from attending preschool during this period. Third, many parents were asked to simply stay at home with their children if possible (e.g., if they were on parental leave with younger siblings), in order to prioritize preschool staff time for the children of key workers.

It is probably impossible to predict changes in guest attendance during the initial phase of a pandemic, but guest attendance after just a few weeks of restrictions appeared to be fairly steady and predictable. In both the schools and preschools studied, guest attendance followed the pattern in the previous year, but with a slight shift downwards, especially in preschools. However, since this downshift seemed quite constant over the period, it must be considered predictable given that the previous days set the level of the forecast.

Introduction of a machine learning approach to predict guest attendance in Swedish public catering proved to be an effective tool for avoiding overcatering, and could therefore be used to reduce economic, environmental, and social impacts of wasting food. Machine learning techniques have been shown previously to improve demand estimation compared with conventional multiple linear regression methods [46]. Specifically, the random forest learning method was found to be able to predict number of portions that kitchens had to serve in schools, setting a “safety margin” to avoid insufficient portions. However, some methodological consideration must be resolved in future research to enable deeper analysis. We found a good level of prediction power of the random forest models from a statistical perspective and considerable amounts of economic savings from a strict economic perspective. We also found quite a high level of uncertainty characterizing the test period, particularly for Kitchen 2, which was characterized by a less standard daily series. This suggests a need not only to assess various types of models, but also to collect “auxiliary” information on e.g., the school, pupils, and kitchen characteristics, to improve the accuracy of forecasting.

The results could be viewed as very case-dependent, given that a future crisis would most likely not follow exactly the same pattern as the COVID-19 pandemic and that the Swedish public catering system is rare in a global perspective. However, certain aspects are general and should therefore make the results more generalizable. First, even though the next pandemic will probably take a different shape, societies will have experiences from the Covid-19 pandemic, and base restrictions and measures on these experiences. This happened already during the second wave of the Covid-19 pandemic (autumn 2020), when countries all over Europe were trying to restrict the spread of the virus in society, but without closing down schools, a similar approach to that applied in Sweden during spring 2020. Second, even though the concept of free school meals is rather rare, meals are served in school canteens in many countries, and the forecasting techniques tested in this paper should be applicable to other settings to increase the precision of predicting future guest attendance. An interesting finding by Ref. [47] for the household sector was an observed reduction in levels of food waste during the pandemic in an Italian context, due to careful food planning and management. However, those authors stress the important role of contextual variables in reducing food waste during a lockdown scenario, and their findings are not easily transferable to other sectors of the food supply chain.

[1] Kitchen 2 served portions up to the end of June, so we attempted to predict the last 20 days of available data. Kitchen 1 served portions until 8 June, according to our data.

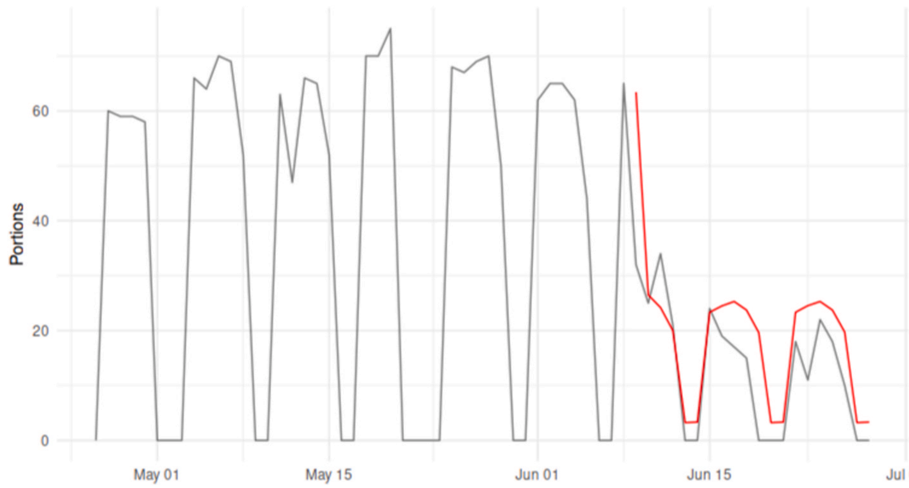

Fig. 7. Number of school meal portions served by Kitchen 2 from April 2020 to June 2020. Black line: observed data, red line: predicted data. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

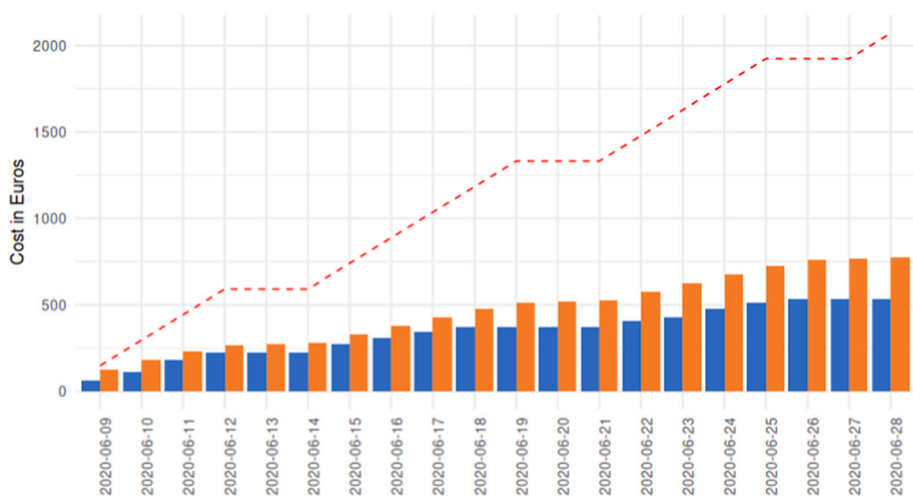

Fig. 8. Results of economic scenario analysis for Kitchen 2. ( ) Sum of meal costs based on actual data. ( ) Sum of meal costs based on forecasted data. ( ) Sum of meal costs based on total enrolled students.

Based on the results in this paper, policy recommendations can be formulated for a future situation where there are restrictions in society to contain a pandemic, but where schools and preschools are kept open. The main recommendation is use a forecasting tool before a crisis to avoid overproduction, but also to train the forecast on a normal situation. The situation in the first days of restrictions is difficult to predict, since it is highly dependent on people's reactions and precautions, rather than the actual number of infected individuals. During this period, the best strategy would be to accept that too much food will be purchased and therefore shift the focus to handling this overproduction by re-using leftovers or preparing food in smaller batches, so as to adjust production more precisely, as recommended by the Swedish Food Agency in normal situations [48]. During this period, it would also be good to adjust the forecast so that recent days weigh higher in the predictions, or simply shift to a simple model using moving average or similar models, as exemplified in Refs. [13,49]. In the case of underprediction, shortages will occur and a sufficient margin needs to be in place. Optimization of this margin is something that is described by Ref. [13] who also proposes a system where a backup stock is used to handle days when the forecast delivers an underprediction to meet demand. Such a system would need to be evaluated to see if it has the desired food waste reduction effect.

For the remainder of the pandemic period, it is likely that the normal forecast could be applied with good precision. There might be a new slightly lower level, e.g., if some guests are prevented from attending the lunch servings, but since this would be a known factor it should be possible to adjust the forecast based on these numbers. In the random tree approach, pandemic or non-pandemic could be used as a parameter to better train the machine learning algorithm. In summary, a good forecasting tool is likely to predict guest attendance during a pandemic with good accuracy, but not during the initial phase, where production adjustments are also needed. One of the key aspects in implementing forecasting in kitchens is to achieve trust in the forecasts among the staff and have an acceptable margin in place. One major advantage that random forest or any other tree-based method provides is that they are easy to explain [50], which can be one way of establishing trust in the forecasts that they provide. Without trust in the information provided by the forecast, it is unlikely to be used by the kitchen staff and therefore potential benefits with a forecasting system are in vain. However, using a method that is explainable and not of a black box type might lower the trust threshold and increase the likelihood of implementation and usage among end-users. Accurate and trusted forecasting is therefore a tool that can be useful to prevent overproduction and increase sustainability both before and during a pandemic.

6. Conclusions

It is impossible to forecast demand early during extreme events. Instead, the focus should be on dealing with surplus food or preparing smaller batches during the initial phase of a pandemic or other extreme event. However, after a short period into the Covid-19 pandemic a forecasting tool was able to predict guest attendance at preschool and primary school meals with accuracy. A random forest approach was found to predict guest attendance during the test period with a conditional absolute mean error of 0.448–0.487 on kitchen level. The most important factors for the accuracy of the forecast were found to be weekday and weekend, while number of daily new Covid-19 cases had no prediction power in the model. The potential savings from use of random tree forecasting was found to be 921–1298 Euros compared with a situation where a meal is cooked for every pupil enrolled in schools. Random forest forecasting appears therefore to be a suitable tool for implementing in public catering in order to reduce food waste and contribute to a more sustainable food system.

CRediT authorship contribution statement

Christopher Malefors: Conceptualization, Methodology, Software. Resources, Data Curation, Formal analysis, Writing - review & editing. **Luca Secondi:** Conceptualization, Methodology, Software, Data Curation, Formal analysis, Writing - review & editing. **Stefano Marchetti:** Methodology, Software, Data Curation, Formal analysis, Writing - review & editing. **Mattias Eriksson:** Conceptualization, Methodology, Software, Resources, Data Curation, Formal analysis, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by H2020 ERA-net Cofund on Sustainable Food Production and Consumption (SUSFOOD2) by the Swedish Research Council for Sustainable Development (Formas), grant number FR-2018/0001.

References

[1] Ocampo L, Yamagishi K. Modeling the lockdown relaxation protocols of the philippine government in response to the COVID-19 pandemic: an intuitionistic fuzzy DEMATEL analysis. Soc Econ Plann Sci 2020;72:100911. https://doi.org/10.1016/j.seps.2020.100911.
[2] Foley JA. Global consequences of land use. Science 2005;309(5734):570–4. https://doi.org/10.1126/science.1111772.
[3] Newbold T, Hudson LN, Hill SLL, Contu S, Lysenko I, Senior RA, Börger L, Bennett DJ, Choimes A, Collen B, Day J, Palma AD, Díaz S, Echeverria-Londoño S, Edgar MJ, Feldman A, Garon M, Harrison MLK, Alhusseini T, Ingram DJ, Itescu Y, Kattge J, Kemp V, Kirkpatrick L, Kleyer M, Correia DLP, Martin CD, Meiri S, Novosolov M, Pan Y, Phillips HRP, Purves DW, Robinson A, Simpson J, Tuck SL, Weiher E, White HJ, Ewers RM, Mace GM, Scharlemann JPW, Purvis A. Global effects of land use on local terrestrial biodiversity. Nature 2015;520(7545):45–50. https://doi.org/10.1038/nature14324.
[4] Shiklomanov IA. World water resources at the beginning of the twenty-first century. Cambridge, UK New York: Cambridge University Press; 2003.
[5] Wada Y, van Beek LPH, van Kempen CM, Reckman JWTM, Vasak S, Bierkens MFP. Global depletion of groundwater resources. Geophys Res Lett 2010;37(20). n/a–n/a, 10.1029/2010gl044571.
[6] Edenhofer O, editor. Climate change 2014 : mitigation of climate change : working group III contribution to the fifth assessment report of the intergovernmental panel on climate change. New York, NY: Cambridge University Press; 2014.
[7] Vermeulen SJ, Campbell BM, Ingram JS. Climate change and food systems. Annu Rev Environ Resour 2012;37(1):195–222. https://doi.org/10.1146/annurev-environ-020411-130608.
[8] Cordell D, White S. Lifes bottleneck: sustaining the world's phosphorus for a food secure future. Annu Rev Environ Resour 2014;39(1):161–88. https://doi.org/10.1146/annurev-environ-010213-113300.
[9] Robertson GP, Vitousek PM. Nitrogen in agriculture: balancing the cost of an essential resource. Annu Rev Environ Resour 2009;34(1):97–125. https://doi.org/10.1146/annurev.environ.032108.105046.
[10] Diaz RJ, Rosenberg R. Spreading dead zones and consequences for marine ecosystems. Science 2008;321(5891):926–9. https://doi.org/10.1126/science.1156401.
[11] Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, Pretty J, Robinson S, Thomas SM, Toulmin C. Food security: the challenge of feeding 9 billion people. Science 2010;327(5967):812–8. https://doi.org/10.1126/science.1185383.
[12] Springmann M, Clark M, Mason-D'Croz D, Wiebe K, Bodirsky BL, Lassaletta L, de Vries W, Vermeulen SJ, Herrero M, Carlson KM, Jonell M, Troell M, DeClerck F, Gordon LJ, Zurayk R, Scarborough P, Rayner M, Loken B, Fanzo J, Godfray HCJ, Tilman D, Rockström J, Willett W. Options for keeping the food system within environmental limits. Nature 2018;562(7728):519–25. https://doi.org/10.1038/s41586-018-0594-0.
[13] Malefors C, Strid I, Hansson P-A, Eriksson M. Potential for using guest attendance forecasting in Swedish public catering to reduce overcatering. Sustainable Production and Consumption 2021;25:162–72. https://doi.org/10.1016/j.spc.2020.08.008.
[14] Boschini M, Falasconi L, Giordano C, Alboni F. Food waste in school canteens: a reference methodology for large-scale studies. J Clean Prod 2018;182:1024–32. https://doi.org/10.1016/j.jclepro.2018.02.040.
[15] Byker CJ, Farris AR, Marcenelle M, Davis GC, Serrano EL. Food waste in a school nutrition program after implementation of new lunch program guidelines. J Nutr Educ Behav 2014;46(5):406–11. https://doi.org/10.1016/j.jneb.2014.03.009.
[16] Engström R, Carlsson-Kanyama A. Food losses in food service institutions examples from Sweden. Food Pol 2004;29(3):203–13. https://doi.org/10.1016/j.foodpol.2004.03.004.
[17] Falasconi L, Vittuari M, Politano A, Segrè A. Food waste in school catering: an Italian case study. Sustainability 2015;7(11):14745–60. https://doi.org/10.3390/su71114745.
[18] Malefors C, Callewaert P, Hansson P-A, Hartikainen H, Pietiläinen O, Strid I, Strotmann C, Eriksson M. Towards a baseline for food-waste quantification in the hospitality sector—quantities and data processing criteria. Sustainability 2019;11(13):3541. https://doi.org/10.3390/su11133541.
[19] Eriksson M, Osowski CP, Malefors C, Björkman J, Eriksson E. Quantification of food waste in public catering services – a case study from a Swedish municipality. Waste Manag 2017;61:415–22. https://doi.org/10.1016/j.wasman.2017.01.035.
[20] Eriksson M, Malefors C, Bergström P, Eriksson E, Osowski CP. Quantities and quantification methodologies of food waste in Swedish hospitals. Sustainability 2020;12(8):3116. https://doi.org/10.3390/su12083116.
[21] Swedish National Food Agency, Fakta om offentliga måltider 2019 - kartläggning av matsvinn i kommunalt drivna förskolor, skolor och äldreboenden., Livsmedelsverkets rapportserie, Uppsala.
[22] Swedish Parliament, Svensk författningssamling, Skollag (school law) 201 0:800.
[23] Tikkanen I, Urho U-M. Free school meals, the plate model and food choices in Finland. Br Food J 2009;111(2):102–19. https://doi.org/10.1108/00070700910931940.
[24] Ministry of Education and Research of the Republic of Estonia. Pre-school, basic and secondary education. Estonian Ministry of Education and Research 2019. https://www.hm.ee/en/activities/pre-school-basic-and-secondary-education. accessed 2020-01-24.
[25] Steen H, Malefors C, Röös E, Eriksson M. Identification and modelling of risk factors for food waste generation in school and pre-school catering units. Waste Manag 2018;77:172–84. https://doi.org/10.1016/j.wasman.2018.05.024.
[26] Boschini M, Falasconi L, Cicatiello C, Franco S. Why the waste? a large-scale study on the causes of food waste at school canteens. J Clean Prod 2020;246:118994. https://doi.org/10.1016/j.jclepro.2019.118994.
[27] Bajželj B, Richards KS, Allwood JM, Smith P, Dennis JS, Curmi E, Gilligan CA. Importance of food-demand management for climate mitigation. Nat Clim Change 2014;4(10):924–9. https://doi.org/10.1038/nclimate2353.
[28] Foley JA, Ramankutty N, Brauman KA, Cassidy ES, Gerber JS, Johnston M, Mueller ND, O'Connell C, Ray DK, West PC, Balzer C, Bennett EM, Carpenter SR, Hill J, Monfreda C, Polasky S, Rockström J, Sheehan J, Siebert S, Tilman D, Zaks DPM. Solutions for a cultivated planet. Nature 2011;478(7369):337–42. https://doi.org/10.1038/nature10452.
[29] Garnett T. Where are the best opportunities for reducing greenhouse gas emissions in the food system (including the food chain)? Food Pol 2011;36:S23–32. https://doi.org/10.1016/j.foodpol.2010.10.010.
[30] J. O. Karlsson, G. Carlsson, M. Lindberg, T. Sjunnestrand, E. Röös, Designing a future food vision for the nordics through a participatory modeling approach, Agron Sustain Dev 38 (6). doi:10.1007/s13593-018-0528-0.
[31] Garre A, Ruiz MC, Hontoria E. Application of machine learning to support production planning of a food industry in the context of waste generation under uncertainty. Operations Research Perspectives 2020;7:100147. https://doi.org/10.1016/j.orp.2020.100147.
[32] Lasek A, Cercone N, Saunders J. Restaurant sales and customer demand forecasting: literature survey and categorization of methods. In: Smart city 360°. Springer International Publishing; 2016. p. 479–91. https://doi.org/10.1007/978-3-319-33681-7_40.

[33] Blecher L, Yeh RJ. Forecasting meal participation in university residential dining facilities. J Foodserv Bus Res 2008;11(4):352–62. https://doi.org/10.1080/15378020802519637.
[34] Miller JL, A Thompson P, Orabella M. Forecasting in foodservice: model development, testing, and evaluation. J Am Diet Assoc 1991;91(5):569–74.
[35] Filimonau V. The prospects of waste management in the hospitality sector post COVID-19, Resources. Conserv Recycl 2020:105272doi. https://doi.org/10.1016/j.resconrec.2020.105272.
[36] Yurtsever M, Tecim V. Forecasting meal requirements using time series methods in organization. In: Economic and financial challenges for balkan and eastern European countries. Springer International Publishing; 2020. p. 243–54. https://doi.org/10.1007/978-3-030-39927-6_15.
[37] Ryu K, Sanchez A. The evaluation of forecasting methods at an institutional foodservice dining facility. J Hospit Financ Manag 2003;11(1):27–45. https://doi.org/10.1080/10913211.2003.10653769.
[38] S. J. Taylor, B. Letham, Forecasting at scaledoi:10.7287/peerj.preprints.3190v2.
[39] Grainger MJ, Aramyan L, Piras S, Quested TE, Righi S, Setti M, Vittuari M, Stewart GB. Model selection and averaging in the assessment of the drivers of household food waste to reduce the probability of false positives. PloS One 2018;13(2):e0192075. https://doi.org/10.1371/journal.pone.0192075.
[40] Amit Y, Geman D. Shape quantization and recognition with randomized trees. Neural Comput 1997;9(7):1545–88. https://doi.org/10.1162/neco.1997.9.7.1545.
[41] Swedish Parliament. Svensk författningssamling. Skolförordningen 2011:185.
[42] Eriksson M, Osowski CP, Björkman J, Hansson E, Malefors C, Eriksson E, Ghosh R. The tree structure — a general framework for food waste quantification in food services. Resour Conserv Recycl 2018;130:140–51. https://doi.org/10.1016/j.resconrec.2017.11.030.
[43] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. Springer-Verlag New York Inc.; 2009. URL, https://www.ebook.de/de/product/8023140/trevor_hastie_robert_tibshirani_jerome_friedman_the_elements_of_statistical_learning.html.
[44] R Core Team R. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2020. URL, https://www.R-project.org/.
[45] Liaw A, Wiener M. Classification and regression by randomforest. R News 2002;2(3):18–22. URL, https://CRAN.R-project.org/doc/Rnews/.
[46] Nair DJ, Rashidi TH, Dixit VV. Estimating surplus food supply for food rescue and delivery operations. Soc Econ Plann Sci 2017;57:73–83. https://doi.org/10.1016/j.seps.2016.09.004.
[47] Principato L, Secondi L, Cicatiello C, Mattia G. Caring more about food: the unexpected positive effect of the covid-19 lockdown on household food management and waste. Soc Econ Plann Sci 2020:100953doi. https://doi.org/10.1016/j.seps.2020.100953.
[48] Swedish National Food Agency. Handbok för minskat matsvinn - för verksamheter inom vård. skola och omsorg.; 2020.
[49] Malefors C. Food waste in the food service sector - quantities, risk factors and reduction strategies, Lic thesis. Swedish University of Agricultural Sciences; 2021, ISBN 978-9 1-576-9828-5.
[50] James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. Springer-Verlag GmbH; 2017. URL, https://www.ebook.de/de/product/29292548/gareth_james_daniela_witten_trevor_hastie_robert_tibshirani_an_introduction_to_statistical_learning.html.

Christopher Malefors is a PhD-candidate in the Department of Energy and Technology at the Swedish University of Agricultural Sciences. He received his M. Sc in Sociotechnical Systems Engineering from Uppsala University. His research interest include system optimization and understanding of establishments working in the food service sector with focus on food waste reduction measures.

Luca Secondi is Associate Professor of Economic Statistics at the University of Tuscia (Viterbo– Italy). Since November 2019 he has been appointed as Rector's Delegate for national and international ranking. He obtained the PhD in Applied Statistics at the University of Florence and was visiting researcher at the American University (Washington-DC), the Shanxi and Jiangxi universities (China) and the University of Castilla-La Mancha (Spain). His main research interests cover the area of consumer behaviour especially regarding food waste and circular economy issues. Starting from November 2020 he participates as task leader and supports the coordination of the HORIZON-2020 LOWINFOOD project.

Stefano Marchetti is Associate Professor at the University of Pisa, he holds a PhD in Applied Statistics from the University of Florence (2009). He spent a visiting period at the Centre for Census and Survey Research, University of Manchester. He is member of the International Association of Survey Statisticians, the Italian Statistical Society. and the Dagum Centre on Advanced Statistics for Equitable and Sustainable Development. He is Associate Editor for PLOS One and Journal of Statistics: Advances in Theory and Applications. His research focus mainly on small area estimation. He participated to H220 and FP7 projects.

Mattias Eriksson works as an associate professor at the Department of Energy and Technology at the Swedish University of Agricultural Science in Uppsala. He received a PhD in technology 2015 for his thesis titled *Supermarket food waste - Prevention and management with the focus on reduced waste for reduced carbon footprint*. The focus of Erikssons research, teaching and outreach has been on sustainable food systems and especially food waste and life cycle assessment of agricultural products and waste management systems.