Food Control 121 (2021) 107630

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

# A surveillance of food borne disease outbreaks in India: 2009–2018

Akshay Bisht [a,b], Manoj P. Kamble [a], Pritesh Choudhary [a], Kartikey Chaturvedi [a], Gautam Kohli [a,c], Vijay K. Juneja [d], Shalini Sehgal [e], Neetu Kumra Taneja [a,*]

[a] *Department of Basic and Applied Science, National Institute of Food Technology Entrepreneurship and Management (NIFTEM), Kundli, India*
[b] *School of Food and Advanced Technology, Massey University, Auckland, New Zealand*
[c] *European Joint Master's Program-FIPDES, AgroParis Tech, Paris, France*
[d] *USDA-Agricultural Research Service, USA*
[e] *Bhaskaracharya College of Applied Sciences, University of Delhi, Delhi, India*

ARTICLE INFO

*Keywords:*
India
Food-borne disease
Outbreaks
Surveillance
IDSP

ABSTRACT

Knowledge about distribution of food-borne outbreaks and implicated food-vehicles helps in mitigating the risk of food-borne diseases and is critical for designing strategies to control them. In this study data from Integrated Disease Surveillance Program (IDSP) and Open Government Data Platform India (OGDPI) on food-borne outbreaks for the period 2008–2018 was consolidated and analysed. The modelling methods of Gaussian distribution model (GAM) and Autoregressive Integrated Moving Average (ARIMA) were used to probe influence of climatic factors (temperature and rainfall) on food-borne outbreaks. Data analysis showed that states of West Bengal (31.22), Karnataka (29.11) and Gujarat (22.67) reported maximum average outbreaks and contributed to 31.5% illnesses and 8.7% deaths. Amongst 19.6% of outbreaks, grains and beans were found to be food-vehicle causing maximum outbreaks (32.7%), while chemically contaminated food caused maximal deaths (70%). Weak correlations of climatic factors with outbreaks resulted in poor performance of ARIMA models. GAM model was validated and predicted 356 outbreaks for the year 2020, late-April to mid-July being most prevalent months. The analysis also revealed inclination of current surveillance program towards chemically contaminated food that resulted in maximal deaths (70%), while biological agents were observed to be under-reported. Despite the limitations, available data shows that food-borne disease outbreaks remain a public health concern in India. Therefore, it is imperative for India to strengthen its disease surveillance program by undertaking capacity-building initiatives at state/local health-care levels and connecting causative agents of outbreaks. This would help in efficient implementation of risk assessment and risk management strategies.

## 1. Introduction

The emergence of new diseases and resurgence of existing infectious diseases is amongst the major public health challenges faced by nations across the world. Amongst all other ailments, food-borne diseases continues to be the major cause of morbidity and mortality, resulting in substantial socio-economic losses globally (Devleesschauwer, Haagsma, Mangen, Lake, & Havelaar, 2018; World Health Organisation, 2015). This necessitates the need for prompt monitoring, reporting and dissemination of timely warnings of food borne disease outbreaks. National level organizations such as the Centers for Disease Control and Prevention (CDC) in USA and many others across other nations have developed various disease surveillance systems to improve public health. In USA, food-borne diseases are responsible for approximately 48 million illnesses, 128,000 hospitalizations, and 3000 deaths each year (Scharff et al., 2016). While in Australia, food gastroenteritis accounts for 5.4 million illnesses, 14,700 hospitalizations, and 76 deaths each year (Hall et al., 2005). In addition to the aforementioned nations, several reports monitoring and analysing the impact of consuming food contaminated with microbes, toxins or chemicals, have been published for other developed nations such as England and Wales (Adams et al., 2019), Canada (Bélanger, Tanguay, Hamel, & Phypers, 2015), Netherlands (Pijnacker et al., 2019), France (Vaillant et al., 2005; Van Cauteren et al., 2017), and New Zealand (Pattis, Lopez, Cressey, Horn, & Roos, 2017). However, there are limited studies highlighting the burden of food-borne diseases in low to medium-income countries, including India.

The food produced in developing and under-developed nations is

* Corresponding author.
*E-mail address:* neetu.taneja@niftem.ac.in (N.K. Taneja).

https://doi.org/10.1016/j.foodcont.2020.107630
Received 13 August 2020; Received in revised form 10 September 2020; Accepted 11 September 2020
Available online 12 September 2020
0956-7135/

 

more prone to contamination because of manifold reasons such as lack of access to clean water for food preparation, poor transportation, inadequate storage infrastructure, poor personal hygiene and improper handling thus making the population more susceptible to diseases (Dandage, Badia-Melis, & Ruiz-García, 2017; Jahan, 2012; Odeyemi, 2016). If food safety standards in India remain at current state, over 100 million annual cases of food-borne diseases are estimated, which would increase to 150–177 million cases by 2030 (Kristkova, Grace, & Kuiper, 2017). The current estimate however represents only tip of the iceberg, and the exact health burden due to contaminated food is vastly under-reported and unknown. Food-borne diseases are self-limiting therefore most of the patients with short-term mild symptoms do not seek medical care, thus making the cases/incidences go unreported (Gibbons et al., 2014). Further, many pathogens that are commonly associated with food can also be transmitted indirectly *via* the environment, animals, or infected/carrier person, thereby complicating the estimation of proportion of diseases originating from food sources (Hald et al., 2016). To fulfil this gap, Integrated Disease Surveillance Program (IDSP) was started by the Ministry of Health and Family Welfare, Government of India, in 2004 for effective management and minimization of food-borne diseases. Food-borne diseases are preventable and their timely investigation and reporting may help in reducing the risk of spread to a larger population. Therefore, surveillance of food-borne diseases aims at reducing the transmission and occurrence of similar outbreaks by monitoring outbreak trends, identifying and eliminating the implicated food, identifying specific pathogens that cause illness, determining the transmission pathway for specific pathogens, identifying vulnerable groups and providing information to the policymakers (Ford, Miller, Cawthorne, Fearnley, & Kirk, 2015; Sudershan, Naveen, Kashinath, Bhaskar, & Polasa, 2014; Wu et al., 2018).

National Centre for Disease Control (NCDC) conducts surveillance for several communicable and non-communicable diseases, including food-borne disease outbreaks in India through IDSP. IDSP is a decentralized program that identifies and investigates outbreaks at district, state, and central surveillance units. The key elements of IDSP include identification of an outbreak, its investigation, and control, followed by analysis and dissemination of results to develop a response strategy (Appendix: Fig. S1). At the current system, district and state surveillance units identify outbreaks and report them to a central surveillance unit for a weekly publication. The first weekly report of IDSP, i.e., 'Disease alerts/outbreaks reported and responded to by states/UTs,' was published for 25 weeks, ending on June 21, 2009. Although much of the information reported is collected from the suspect, paramedical, and medical officers, it cannot be directly translated into policies. Thus, in order to make the information useful to understand the disease epidemiology and risk factors, and to devise necessary preventive and control measures, it is pertinent to analyse the available data and predict the probable outbreaks. This report summarizes the food-borne disease outbreaks reported by IDSP between 2009 and 2018 and proposes a model for predicting annual food-borne illnesses in India in the years to come. The findings are intended to be used by health departments and policymakers for developing better policies and implementing appropriate healthcare measures to prevent and control food-borne outbreaks.

## 2. Methods

### 2.1. Data collection

A food-borne disease outbreak is defined as the occurrence of $\geq 2$ cases of a similar illness resulting from the ingestion of a common food (Bennett et al., 2018). This report summarizes the food-borne disease outbreaks reported before July 2019, in which the first illness onset occurred between the 25th week of June 2009 to the 52nd week of December 2018. The weekly surveillance reports were collected from the IDSP website (https://idsp.nic.in/), and the data related to food poisoning outbreaks were segregated. Information collected from each report includes the date of illness onset, state reporting the outbreak, number of illnesses and deaths, and suspected food vehicles. The outbreaks affecting 30 or fewer individuals (illness + death) were defined as small outbreaks, otherwise referred to as large outbreaks. Outbreaks with other modes of transmission, such as person-to-person contact and animal contact, were excluded.

### 2.2. Classification of food vehicle

Foods and ingredients that were suspected or confirmed as the source of outbreaks were classified in 7 categories, as given in Table 1. Outbreaks that were due to the implicated food containing one single contaminated ingredient or different ingredients belonging to a single commodity were assigned to the same category. Outbreaks associated with food ingredients from multiples categories such as prasad in the temple, mid-day meal, marriage function, community festival, death ceremony, chicken and cauliflower curries, and a mixture of vegetables and milk items were not included in food commodity group analysis. Also, the outbreaks with missing or incomplete information on food vehicles were not attributed to any commodity.

### 2.3. Data analysis

#### 2.3.1. Statistical analysis

Descriptive statistics was performed to analyse the outbreaks, illnesses, and deaths in different states over ten years. Rate of incidence reported per 10,000,000 individuals was estimated using the latest published population data from the 2011 Indian census (http://censusindia.gov.in/). For broader understanding of the outbreaks, analysis of the data, its impact and significance on population, the outbreaks were categorised into small and large outbreaks based upon the number of persons falling sick and/or deaths due to foodborne illnesses. The threshold for defining a small outbreak is number of incidences 30 or less, which is the median number of cases per outbreak. Differences between small and large outbreaks and dependency of outbreaks on a food commodity were assessed using the Mann Whitney *U* test and chi-square test ($P < 0.05$), respectively.

**Table 1**
Classification of food commodities and the local food product that may serve as a vehicle for foodborne outbreak in India.

| Food category | Food products | Number of outbreaks | Reference |
|---|---|---|---|
| Fish and shellfish | Raw fish, fish curry, boiled crabs | 25 | https://idsp.nic.in/ |
| Dairy | Milk and its products, curd, butter milk, kulfi, ice-cream | 65 | |
| Meat-poultry and egg | Chicken curry, pilav (beef), chicken biryani, toddy (pork), chicken and mutton curry, egg curry, boiled eggs | 58 | |
| Grains and beans | Fried rice, roti, kollu rasam, saboodana, khichdi, wild herb jatropha, moong, keshiri, bengal gram, cowpeas (lobhiya), ratanjyot seeds, gram, peas, lemon rice, bhatura seeds | 172 | |
| Sweets | Shirni, sweet buniya, uppita, gulab-jamun, kaju-katali, kaddu kheer, rasmalai, paravannam, malida, laddu, rasgulla | 94 | |
| Fruits and vegetables | Mushrooms, leafy vegetables, sprout, vine-stalk, sambhar, potato curry, spinach curry, sause pala, tomato curry, fruits and nuts juice | 94 | |
| Chemical in food | Contamination of food with chemicals like lead, mercury or addition of organic phosphorous and alum in cold drink and alcohol | 18 | |

#### 2.3.2. GAM models for predicting food borne outbreaks

The outbreak data was also used to find out possible modelling solution for prediction. The amplitude version of the Gaussian distribution (GAM) model (equation (1)) was used for the fitness evaluation of data trends and to develop the predictive model (Bhatt et al., 2017; Shoukri, Asyali, Van Dorp, & Kelton, 2004). The simulation graph explaining the progression of the model in idealistic conditions is shown as Appendix (Fig. S2). Levenberg Marquardt algorithm was used for iteration at 1000 epochs and error rate $e^{11}$. The predictive values were regressed with actual data set to determine the robustness of the model in the form of $r^2$ values ($P < 0.05$).

$$y = y_0 + Ae^{-\frac{(x-xc)^2}{2w^2}} \quad (1)$$

where, $y_0$: initial value (offset); x: time (months); xc: maximum value of the curve (centre); w: width of the curve ($w > 0.0$); A: intensity of progression (amplitude). The model was fitted using OriginPro 2020 SR1 9.7.0.188.

#### 2.3.3. ARIMA models for timeseries analysis of foodborne outbreaks and climatic factors

An autoregressive integrated moving average (ARIMA) model is next generation model best at dealing with nonstationary time series data, used for predicting and forecasting in which values of time series are transformed and expressed linearly in terms of previous and current state along with residual series (Zhang, Zhang, Young, & Li, 2014). The transformation is achieved through the following equation:

$$X(t) = \Phi_1 X(t-1) + \dots + \Phi_p X(t-p) + \theta_1 \varepsilon(t-1) + \dots + \theta_q \varepsilon(t-q) \quad (2)$$

where $(p,d,q)x(P,D,Q)_s$, $P$ autoregressive of seasonal order, $p$ autoregressive of non-seasonal order, and so on. The subscripted letter 's' indicates length of seasonal period, which in present study is 12 (corresponding to 12 months in a year). The model was developed using open source language Python (version 3.0, 2020) involving steps of data treatment, identification, estimation and diagnostic following the protocols by Zhang et al. (2014). The identification was performed on the basis of autocorrelation functions (ACF). Model performance was evaluated based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) (Fabozzi, Focardi, Rachev, & Arshanapalli, 2014).

## 3. Results

### 3.1. Outbreak, illness and death

During the ten-year period ranging from 2009 to 2018, a total of 2688 food-borne disease outbreaks, resulting in 153,745 illnesses, and 572 deaths were reported to IDSP. An average of 269 (range: 67–383) outbreaks, 15,375 (range: 5147–23,425) illnesses and 57 (range: 26–109) deaths were reported each year. The average annual rate of food-borne disease outbreaks was 2.2 outbreaks per 10,000,000 individuals with a maximum of 3.2 in 2016. Maximum cases of illness were reported in 2013 and 2016, affecting 22,177 and 23,425 individuals, respectively, contributing to 30% of all reported food-borne

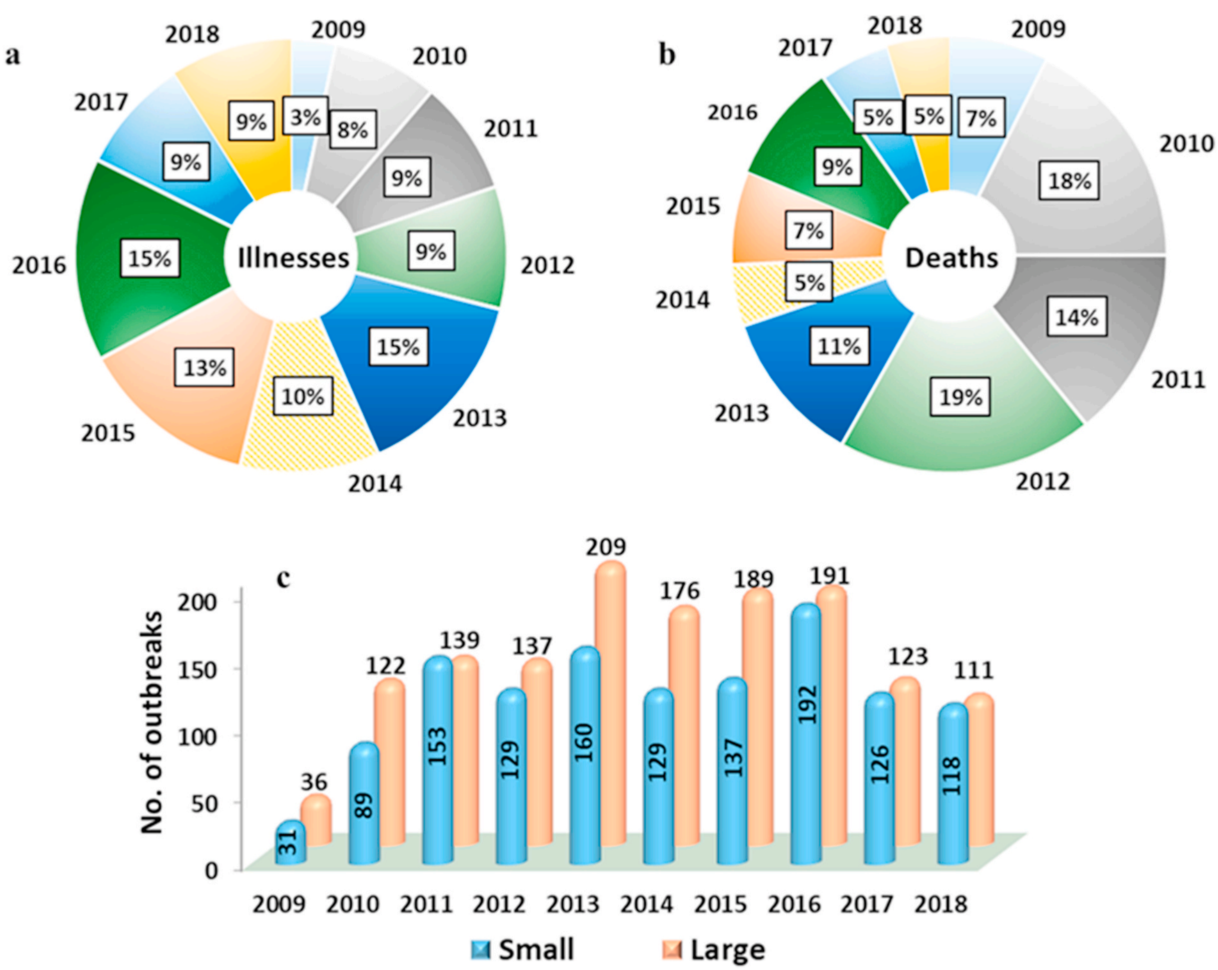


**Fig. 1.** Distribution of percentage of food-borne (a) illnesses and (b) deaths over different years in India. (c) Double bar graph showing the number of small and large outbreaks reported during 2009–2018 in India.

illnesses (Fig. 1a). Up till December 2018, only 0.37% of cases have resulted in deaths, out of which approximately 69% of deaths occurred between 2009 and 2013, with the majority of deaths in 2012 accounting for 19% of all reported deaths (Fig. 1b). In the latter 5 years, the fraction of deaths decreased to 31%; however, the illness increased to 56%.

The reported outbreaks were classified as small and large based on the number of affected people (Fig. 1c). 53% of the outbreaks were identified as large, with the highest number of 209 large outbreaks in 2013. A maximum of small outbreaks was reported in 2016 (192). However, on applying the Mann Whitney *U* test, no significant difference ($P < 0.05$) was found between the occurrence of small and large outbreaks, indicating that the food-borne outbreaks equally affect small (≤30) and large (>30) population. Fig. 2a shows the distribution of outbreaks across different months and reveals the occurrence of outbreaks was dependent on the month of the year of the reported incidence/outbreak. About 54% of outbreaks were reported between March to July of each year, with a maximum average outbreak of 36 in May. On the contrary, only 31% of outbreaks were reported between August to December, with a minimum average outbreak of 13 in December. A similar trend was observed for the number of affected individuals (Fig. 2b). 54% of the individuals suffer from food-borne diseases between March to July, while only 28.5% are affected between August to December.

### 3.2. State-wise distribution of outbreaks

Fig. 3 categorizes the different states and union territories (UT) of India into 5 groups based on the average annual outbreaks. West Bengal, Karnataka, and Gujarat are the states with maximum reported average outbreaks accounting for 31.22, 29.11, and 22.67, respectively. These three states together contributed to 31.5% illnesses and 8.7% deaths over ten years. However, on comparing the rate of outbreaks per 10,

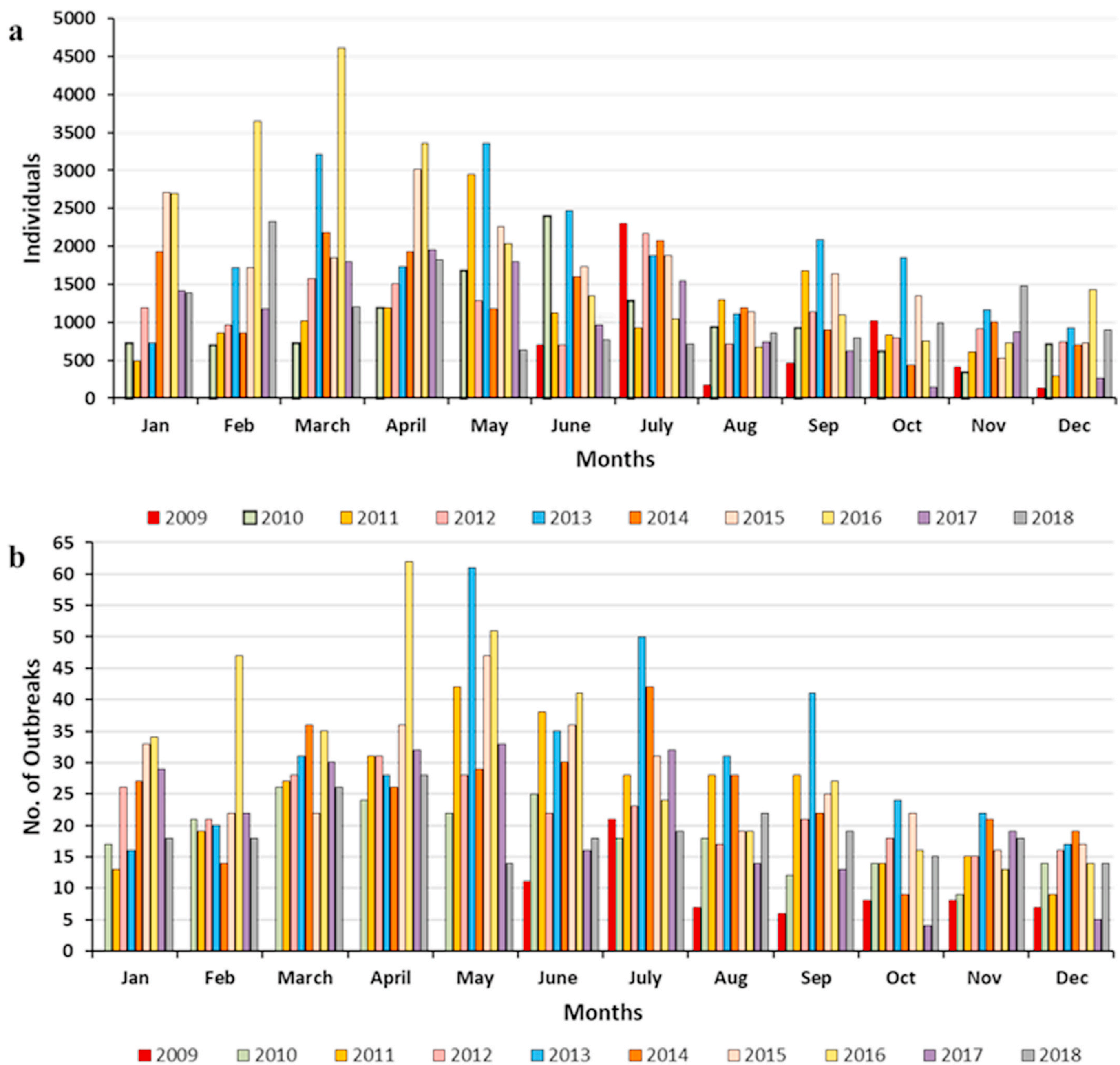


**Fig. 2.** Monthly distribution of (a) food-borne disease outbreaks and (b) affected individuals during 2009–2018 in India.

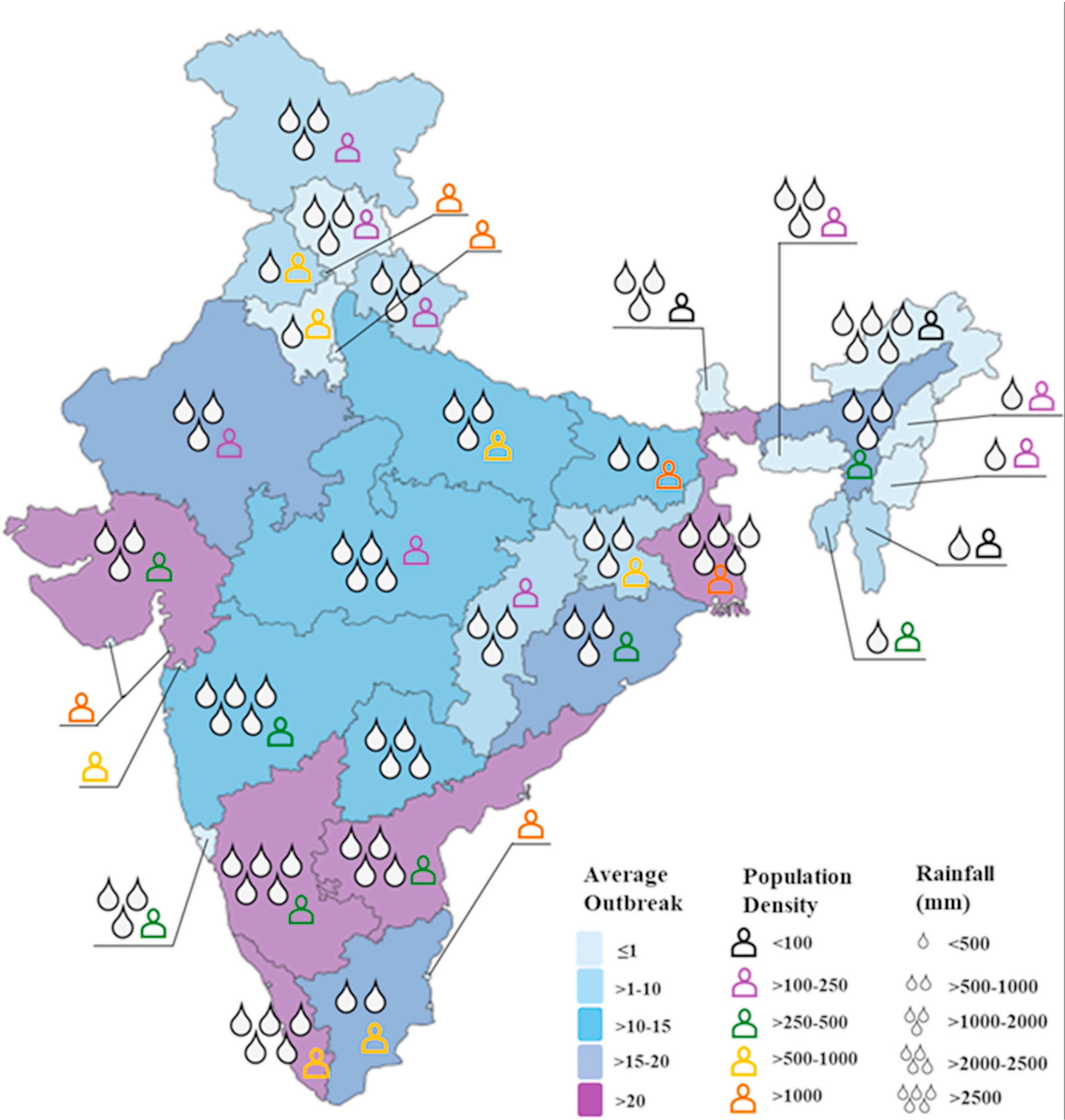


**Fig. 3.** Average annual food-borne outbreaks and average annual rainfall (mm) reported by different states/union territories of India during 2009–2018. Population density as per 2011 census.

000, 000 individuals, Kerala (6.42) tops the list closely followed by Assam (6.02). On the contrary, the rate of the outbreak for West Bengal, Karnataka, and Gujarat was 3.42, 4.77, and 3.75, respectively, though it was still higher than the national average (2.2).

For the first year of the surveillance programme in India, 17 states/ union territories (UT) reported food-borne disease outbreak to IDSP. In 2009, a total of 67 outbreaks were reported of which 12 outbreaks occurred in Kerala (illnesses: 1303; deaths: 0), 10 in Rajasthan (illnesses: 944; deaths: 7) and 9 in Uttar Pradesh (illnesses: 130; deaths:14), while 1 outbreak was reported from each Himachal Pradesh (illnesses: 40; deaths: 0), Karnataka (illnesses: 53 deaths: 0), Maharashtra (illnesses: 42; deaths: 0), Punjab (illnesses: 19; deaths: 0), Tamil Nadu (illnesses: 134; deaths: 0) and West Bengal (illnesses: 79; deaths: 1) (Appendix: Fig. S3). Noticeably, by 2018, 34 states/UT participated in a surveillance programme and reported food-borne disease outbreaks to IDSP.

The total number of outbreaks reported by each state/UT during 2009–2018 varied (range: 1–282), but no pattern was observed. Only one outbreak was reported by Arunachal Pradesh (median: 1; range: 1-1; IQR: 0), Daman and Diu (median: 1; range: 1-1; IQR: 0), Dadar and Nagar Haveli (median: 1; range: 1-1; IQR: 0), and Delhi (median: 1;

range: 1-1; IQR: 0); contrary to Andhra Pradesh (median: 17; range: 3–40; IQR: 10.5), Assam (median: 18.5; range: 4–27; IQR: 13), Gujarat (median: 21.5; range: 5–33; IQR: 11.75), Karnataka (median: 27.5; range: 1–44; IQR: 4.5), Kerala (median: 22; range: 11–30; IQR: 3.75), Madhya Pradesh (median: 11; range: 5–18; IQR: 4.5), Maharashtra (median: 16; range: 1–18; IQR: 4), Orissa (median: 14.5; range: 2–40; IQR: 16.75), Rajasthan (median: 14.5; range: 10–34; IQR: 6.25), Tamil Nadu (median: 14.5; range: 1–31; IQR: 16.5), Uttar Pradesh (median: 13.5; range: 3–34; IQR: 10.25) and West Bengal (median: 34.5; range: 1–49; IQR: 23.25) from where at least one outbreak was reported every year (Fig. 4a and b).

### 3.3. Food as a vehicle

A single suspected or confirmed food or ingredient as a vehicle was reported for limited outbreaks (19.6%). Grains and beans (32.7%) were associated with maximum outbreaks while chemical contaminated food (3.4%) with minimum. Further, grains and beans were the leading cause of both small (27.4%) and large (37.4%) outbreaks, whereas the least percent of small (2.5%) and large (4.3%) outbreaks were associated with chemically contaminated food (Fig. 5a). On comparing the effect of implicated food commodity on the size of the outbreak, a significant difference was found for all food categories except chemical in food, sweets, and fish and shellfish. Fruits and vegetables (24.1% vs 12.5%), and meat-poultry and egg (12.3% vs 8.2%) made up a significantly larger proportion of small outbreaks than large outbreaks. On the other hand, grains and beans (37.4% vs 27.4%) and dairy (15.3% vs 9%) made up a significantly larger proportion of large outbreaks than small outbreaks. During 2009–2018, the implicated food was reported to be responsible for 16.3% of illnesses and 23.3% of deaths, of which, the implicated food commodities commonly associated with illnesses were grains and beans (42%), sweets (17%) and dairy (16%); while fish and shellfish (5%) and food contaminated with chemicals (3%) contributed the least (Fig. 5b). However, consumption of food contaminated with chemicals was the leading cause of deaths (70%) followed by fruits and vegetables (11%), grains and beans (5%), and meat-poultry and egg (5%) (Fig. 5c). Interestingly, sweets were associated with only 1% of deaths.

### 3.4. Modelling food-borne outbreaks

Table 2 summarizes the calculated values for GAM. For most years, the suggested model reasonably predicted the distribution. However, for the outbreak data of the year 2009 and 2018, poor predictions were observed with $r^2$ value of 0.63 and 0.54 values, respectively. For the year 2009, perhaps, it may be due to the fact that the data was available for only 7 months whilst in the latter case data did not show convergence. The fitness of the model was comparatively higher, with 95% confidence bounds regressed with observational data. The parameter 'A' provided the progression 'rate' of the outbreak whereas the parameter 'area' correlates with the 'number' of outbreaks which were highest for the year 2013 and 2016 and matched the observation data set. In a normal scenario, the predictive model suggests that the outbreak distribution around the respective year would depend on two major factors: the outbreak case start point (cases observed in the first month of the year) along with the rate of progression. Parameter 'xc' provided useful information on the probability of months observing a greater number of food-borne outbreaks in India, which are March to July. This observation is concurrent with the data set used for model development. Using the estimated parameters GAM distribution, a generalized predictive distribution model for food-borne outbreaks in India may be proposed as in equation (2). GAM predictions and observed food-borne outbreaks are illustrated in Fig. 6.

$$y = y_0 + 23.47 \pm 16.33 e^{-\frac{(x-5.41)^2}{6.40 \pm 4.21}} \tag{3}$$

Thereafter, Equation (2) was used, with parameter constraints, for prediction and validation of food-borne outbreaks for the year 2019, as depicted in Fig. 7. Statistical significance for the performance of the GAM model studentized *t*-test was performed while linear regression was performed. The model fitted well with 95% confidence constraints for the outliers, with $r^2 = 0.84$ (Fig. 7b). However, perfect fitting could not be observed.

The outbreak data was further modelled with abiotic/climatic (temperature and rainfall) stresses to investigate relationship between them (Fig. 3). The seasonality data is a time series data having high randomness and noise resulting in poor fitting and correlation with conventional tools. ARIMA successfully resolves such data sets by reducing noise and linearizing the data (Zhang, Zheng, & Feng, 2019; Zhang et al., 2014). In this study three different ARIMA models were tried, first ones with bi-factors (temperature/rainfall with outbreaks) and later with tri-factors. Prior to model development data was checked for correlation using ACF, the results suggested weak correlation coefficient $r^2$ values of 0.26 for temperature, 0.29 for rainfall and 0.13 for population density in terms of outbreaks, same was evinced in cases of residuals. This owed to high noise and randomness in the data set. The results along with model structures are appended in Table 3. Since, higher the AIC and BIC values more distant predicted values are from the actual values (Fabozzi et al., 2014). All the models observed very high AIC and BIC values, signifying lack of predictive capabilities of model. However, in comparison to the previously developed ARIMA epidemiological models both AIC and BIC values were 50% lower (Du et al., 2017; Fabozzi et al., 2014; Lindström, Tildesley, & Webb, 2015).

## 4. Discussion

Despite increasing awareness about food safety, the incidences of food-borne disease have increased to an epidemic level across the globe resulting in 600 million cases of illness and 420,000 deaths in 2010 (Havelaar et al., 2015; World Health Organisation, 2015). Therefore, understanding the prevailing risk situation is critical for developing policies for reducing food-borne diseases in the future. In India, a surveillance program has been put into action to collect information about outbreaks that may be used for designing preventive actions. Herein is presented the first study summarizing the food-borne disease outbreaks in India reported between 2009 and 2018.

During the ten years of study, a total of 2688 outbreaks were reported in India. A significantly low number of outbreaks in 2009 is due to reporting for less than seven months of the year, limiting to 17 states/UT. The effect of less data was also clearly observed while fitting the gaussian amplified distribution model. An increase in the following years could be primarily due to surveillance artifact because of the increase in infrastructure, investment, the participation of more states/UT, and the development of referral lab networks rather than an actual increase in the number of outbreaks. Further, recruitment of 395 epidemiologists, 61 microbiologists, and 21 entomologists at the state/district level by December 2013 and improvement in information technology system in 2013 (IDSP, 2019b) could have contributed towards more outbreaks being identified and reported. Clearly, the surveillance system has continuously improved in the country as the number of reporting states/UT increased to 34 by 2018 from 17 in 2009. However, only 47% of reported outbreaks were categorized as small, indicating that small outbreaks are most likely underestimated because of less number of cases seeking health care interventions. Even if a small outbreak is identified, it is mostly not prioritized due to the limited impact on the affected population. For instance, the Indian surveillance program monitors several diseases, including Dengue, Chikungunya, Typhoid Fever, Cholera, Malaria, Diphtheria, and Pneumonia that have greater and severe complexities in individuals than food-borne diseases (IDSP, 2019a). The underestimation of outbreaks, however, would adversely impact the planning, kind of strategies, and measures implemented to combat and control the food-borne outbreaks.

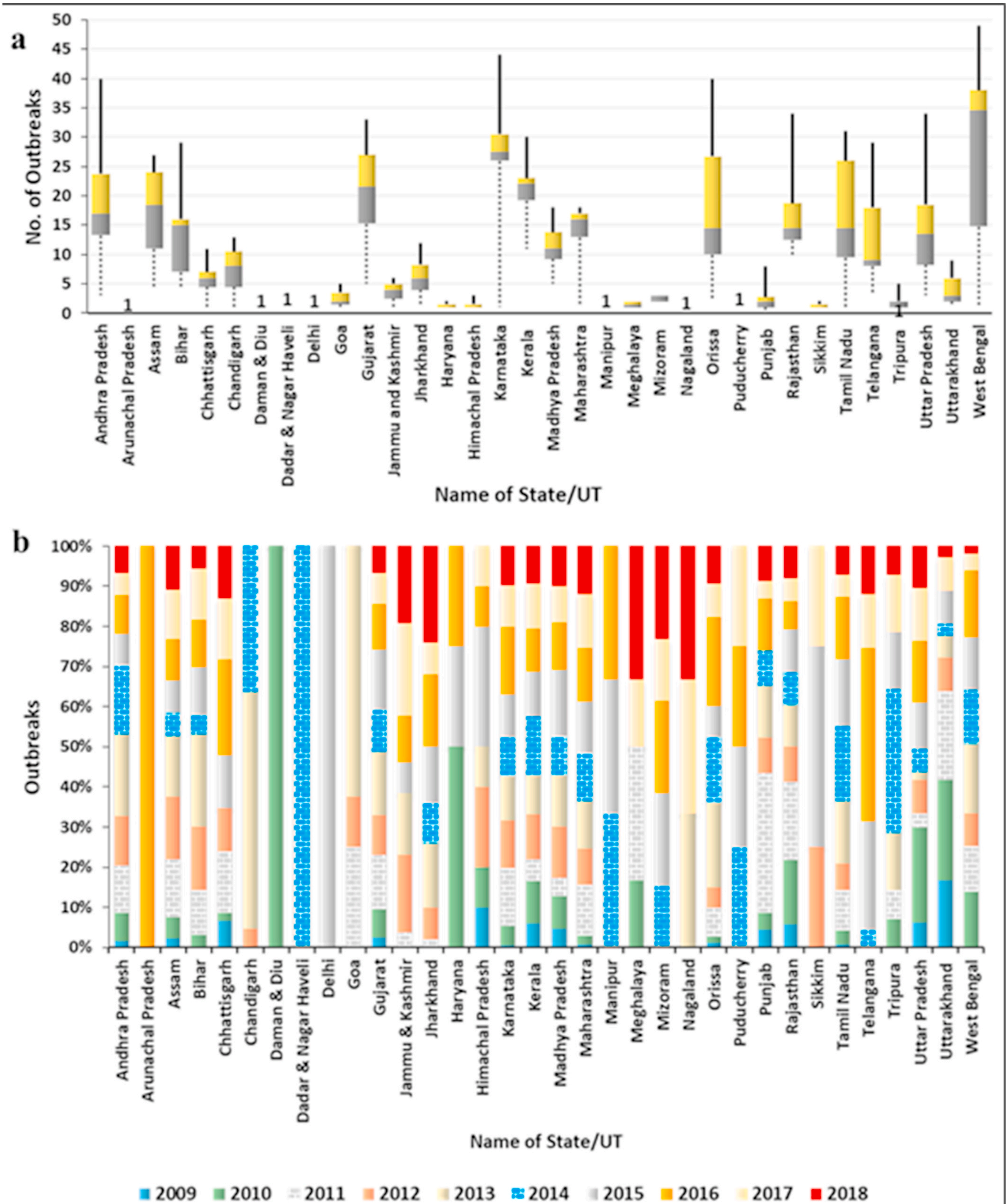


**Fig. 4.** (a) Box plot showing the median of food-borne disease outbreaks and estimated inter-quartile range and (b) Stacked column showing percentage outbreak in 34 states/union territories (UT) of India during 2009–2018.

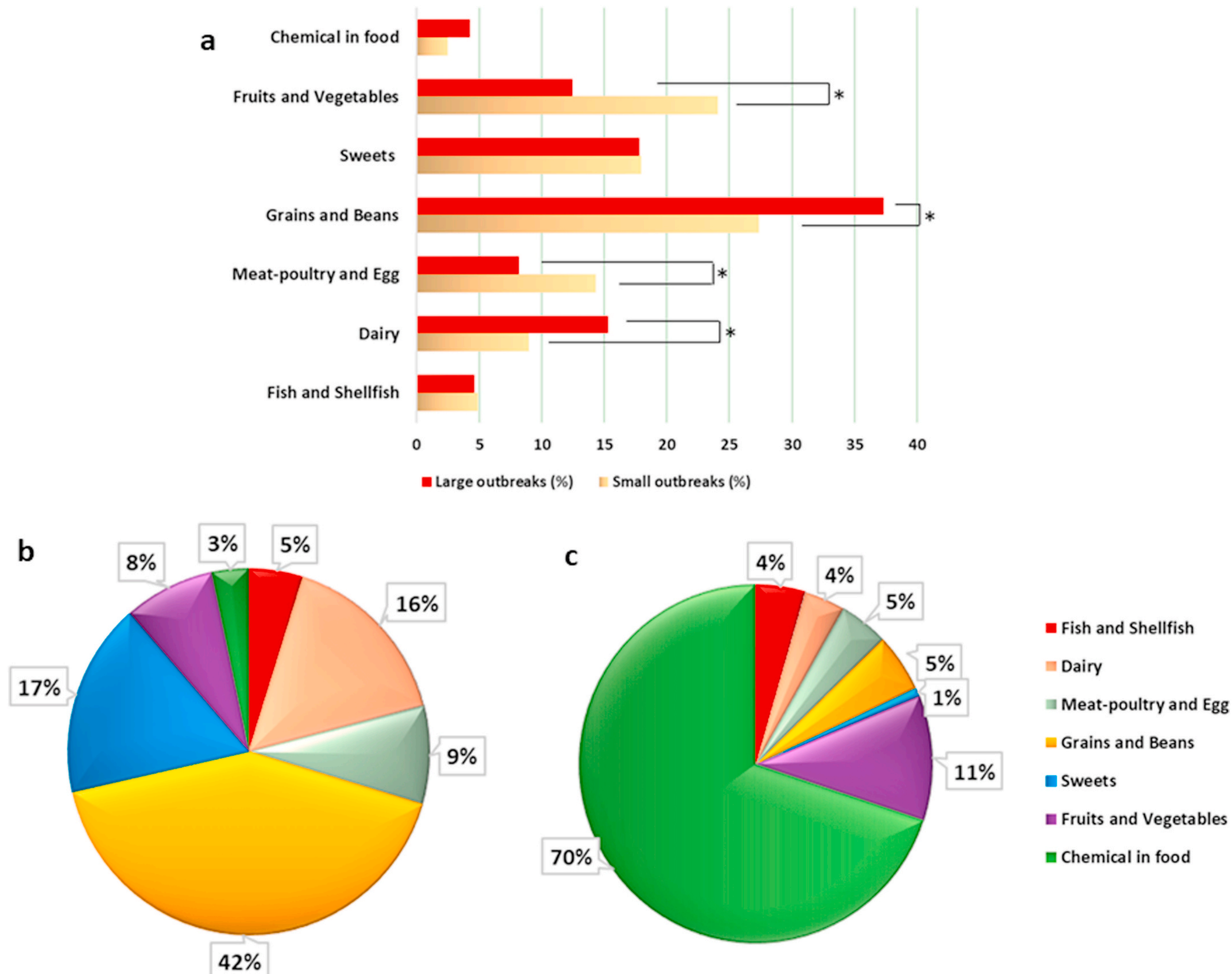


**Fig. 5.** (a) Percentage distribution of small and large outbreaks for different implicated food categories. Percentage of (b) illnesses and (c) deaths associated with different food categories during 2009–2018 in India. * Statistically significant difference between small and large outbreaks at $P < 0.05$.

**Table 2**
Parameter estimation and fitness evaluation data of foodborne outbreaks in India using gaussian amplified distribution model.

| Year | $y_0$ (outbreaks) | Xc (month) | W[b] | A[b] | FWHM (outbreakXmonth) | Area (outbreakXmonth) | Statistics[a] |
|---|---|---|---|---|---|---|---|
| 2009 | 9.1 | 7.00 | 0.19 | 11.90 | 0.35 | 4.4 | 0.63 |
| 2010 | 11.44 | 4.21 | 2.62 | 13.65 | 6.18 | 59.76 | 0.86 |
| 2011 | 7.28 | 5.69 | 2.80 | 30.32 | 6.59 | 212.82 | 0.89 |
| 2012 | 16.66 | 3.84 | 2.40 | 11.97 | 5.63 | 71.78 | 0.87 |
| 2013 | 11.02 | 6.29 | 2.85 | 35.33 | 6.72 | 252.66 | 0.84 |
| 2014 | 22.31 | 6.90 | 0.73 | 19.28 | 1.72 | 35.24 | 0.78 |
| 2015 | 22.00 | 5.12 | 1.03 | 24.13 | 2.42 | 62.05 | 0.75 |
| 2016 | 15.35 | 3.92 | 2.23 | 37.85 | 5.26 | 211.84 | 0.79 |
| 2017 | 28.77 | 10.95 | 2.61 | 20.27 | 6.16 | 132.84 | 0.74 |
| 2018 | 17.5 | 3.51 | 0.25 | 68.50 | 0.59 | 43.14 | 0.54 |

[a] Adjusted $r^2$ values at 95% confidence.
[b] Unit less constants. FWHM; Full width at half maximum of peak.

Geographical and seasonal variations were observed in IDSP data. Observations were consistent with the results of GAM predicting the distribution of food-borne outbreaks. West Bengal has constantly reported maximum outbreaks accounting for 10.5% outbreaks causing 16.8% cases of illness and 2.3% deaths. However, the maximum deaths over ten years were reported in Assam (14.5%), with only 6.4% outbreaks. Moreover, several states/UT, such as Arunachal Pradesh, Delhi, and Daman and Diu have only reported one outbreak in ten years. Similar non-uniform distribution of food-borne disease outbreaks was observed in Australia (OzFoodNet Working Group, 2018), China (Wu et al., 2018) and USA (Gruber, Bailey, & Kowalcyk, 2015). The discrepancy in outbreaks could possibly be because of differences in lifestyle, available resources, education of people, the effectiveness of surveillance teams, and the management of an outbreak rather than the

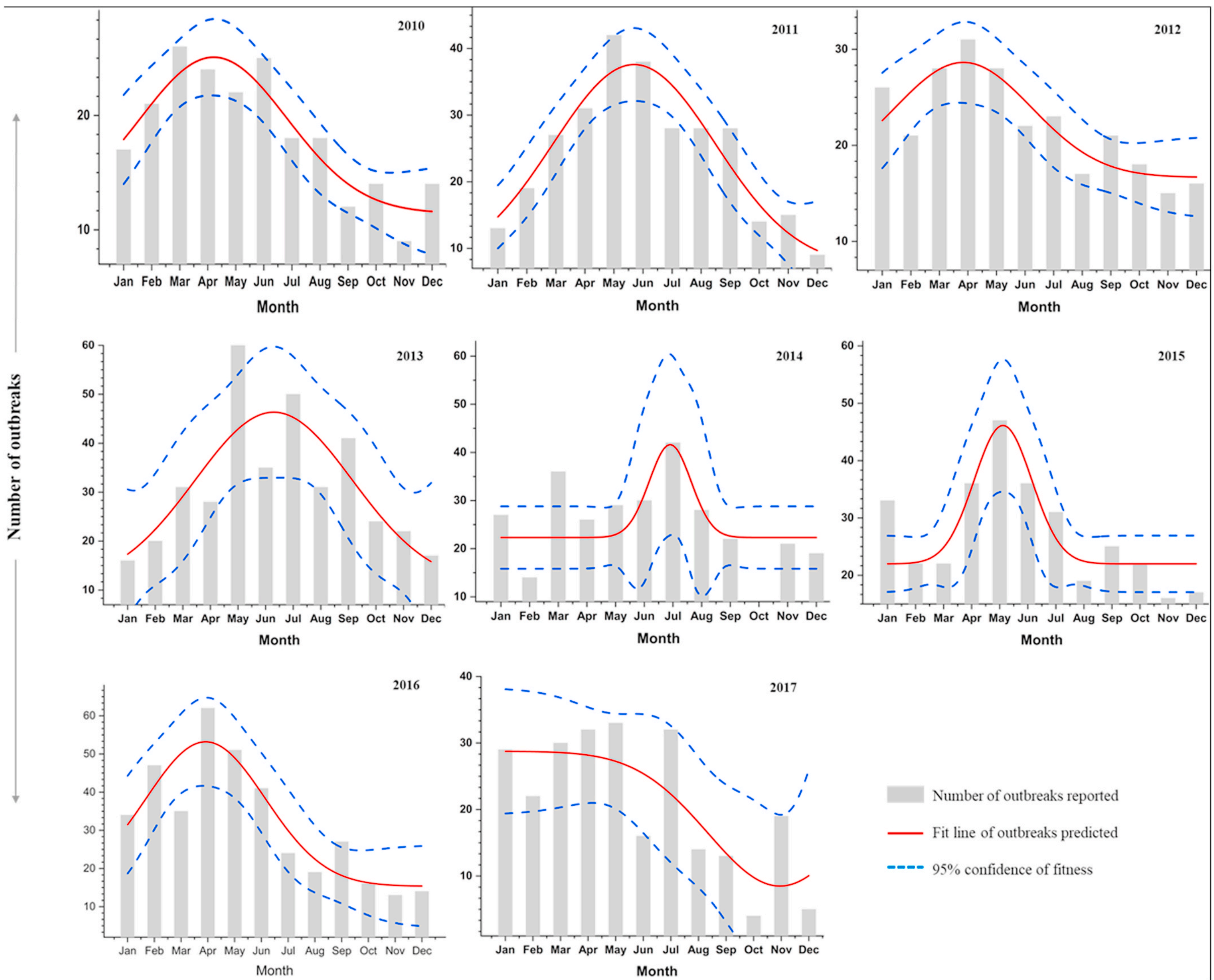

Fig. 6. Plots of gaussian amplified distribution models of outbreak data, arranged year wise.

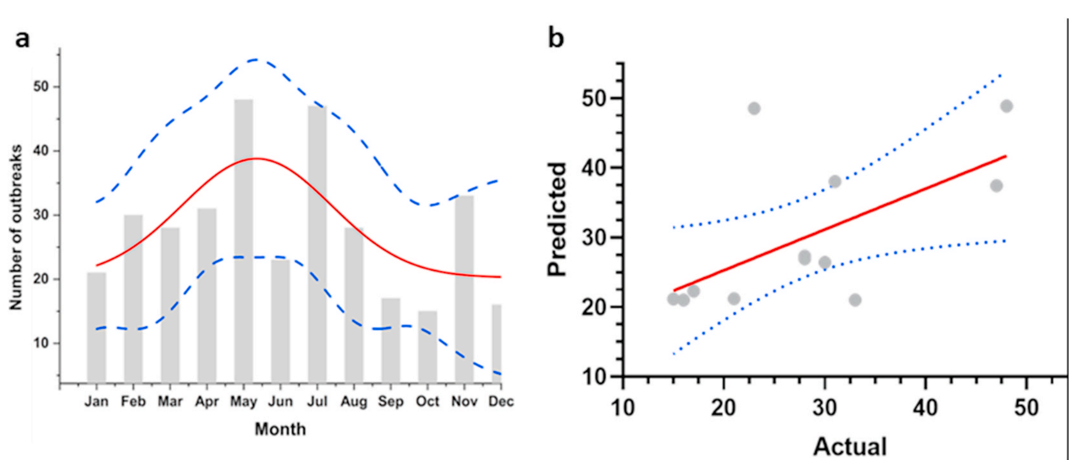

Fig. 7. (a) Plot of gaussian amplified distribution model of outbreak and (b) validation of the proposed model using 2019 data.

actual disparity in the occurrence of an outbreak. The climatic conditions in the country further affect the distribution of outbreaks. High numbers of outbreaks are reported in the summer and early monsoon months. Similarly, in China, food-borne disease outbreaks peak during warmer months (May to September) (Wu et al., 2018). In addition, as reported by World Health Organisation (2015), the occurrence of

**Table 3**
Estimation of ARIMA models for each factor.

| Environmental parameter | Identifier | Constant | Lag | AIC | BIC |
|---|---|---|---|---|---|
| Outbreak x Rainfall | ARIMA (5,2,0) | −0.122 ± 0.876 | −1.742 ± 0.240 | 670.209 | 690.296 |
| Outbreak x Temperature | ARIMA (5,2,0) | 4.328 ± 50.714 | 13.472 ± 2.945 | 1042.524 | 1113.788 |
| Outbreak x Temperature x Rainfall | ARIMA (5,2,1)x (5,2,1) | 7.704 ± 100.373 | 87.522 ± 27.058 | 1638.676 | 1658.495 |

food-borne diseases is distributed amongst people from all age groups, particularly children below 5 years of age, so reporting demographic data is crucial to identifying vulnerable groups. Thus, studying the geographic and demographic distribution of food-borne outbreaks is essential in designing a need-based mitigation program and resource allocation.

Further, plant-based food is associated with the vast majority of outbreaks; however, the distribution amongst different food categories varied. For example, grain and beans are the leading cause of outbreaks, followed by fruits, vegetables and sweets. On the contrary, in the USA (Gould et al., 2013), Australia (OzFoodNet Working Group, 2018), Netherlands (Pijnacker et al., 2019), and Hong Kong (Chan & Chan, 2008) food-borne disease outbreaks are associated mainly with animal origin food. One possible reason for the higher association of plant food to food-borne outbreaks could be the large vegetarian population in India that consumes cereal grain, vegetables, and dairy as their staple diet. In USA, between 2010 and 2014, plant-basses food such as fruits, vegetable raw crops, sprouts and seeded vegetables were the most common implicated foods for multistate foodborne outbreaks (Crowe, Mahon, Vieira, & Gould, 2015). Different percentages are estimated when illness and death are used as the attribution unit. For example, although grains and beans were the leading cause of illnesses attributed to a single food commodity, grains and beans contribute to only 5% of all deaths. In contrast, chemically contaminated food is implicated to the least percent of illnesses but is associated with the maximum number of deaths caused by a single food commodity. Likewise, in China, chemicals added to food are responsible for 14% cases and 55% deaths (Wu et al., 2018). Chemical additives are added to increase the storage life (nitrates, sodium hydroxide, and sodium benzoate) and appearance (carmoisine, sunset yellow, and fast green) of food, however, the use of chemicals above the permissible limit may be toxic. Additionally, food may get accidentally contaminated with chemicals at any point in its supply chain (by the pre-harvest to post-harvest practices) and may not always be a result of adulteration. The use of fertilizers and pesticides in agriculture and inappropriate discard of industrial waste introduces metals like mercury, copper, cadmium, and lead, and phosphates in the food chain that can cause food-borne diseases if consumed by individuals.

Similarly, GAM and ARIMA modelling method used in this study for analysing and predicting the distribution pattern of food-borne outbreaks were effective for such types of data sets. The integration of environmental factors as covariates in the form of marginal means further enhance the predictive capabilities of the model (He, Li, Edmondson, Rader, & Li, 2012). This was achieved for multivariate model development using ARIMA. Another solution for such complex time-series data trends is Support Vector Machine (SVM), which employs machine learning algorithms for linear legalization of variance classifiers (Mittag et al., 2012; Tapak, Hamidi, Fathian, & Karami, 2019; Yu, Liu, Valdez, Gwinn, & Khoury, 2010). While, the GAM model offered a more consistent prediction of highly variant univariate outbreak data, despite that the model failed to incorporate extreme outlier single value in the year 2018 of month May, and low data size of the year 2009, outbreak cases reported. Same was evinced in multivariate ARIMA models where AIC and BIC values suggested distant predictions compared to the observed values (Ceylan, 2020). With more consistent monitoring and a couple of more data sources, the model can be expected to have more robustness and reliability in terms of prediction with lesser variance in the constants necessary to develop generalized predictive model. A major constraint observed which was inherent due to the nature of the outbreak dataset itself is lack of traceability to the food source, vectoring the illness. Perhaps, this is something which policy makers and monitoring agencies may look into. This limitation may have compromised the predictive capabilities of ARIMA model, were trends of seasonality variations are supposed to be correlated with the outbreaks observed. Lack of source traceability also resulted in poor robustness of ARIMA with population density data. Other limitations observed was differences of geo-spatial distribution of dataset, this was attempted to overcome by interpolation of data set, however high noise lead to high variance in interpolated values. Similarly, while validating the GAM model for the year 2019 and comparing with the actual reported food-borne outbreaks, both under and over-estimation of predicted values were observed. GAM model underestimated the peak values of food-borne outbreak for the year 2019 while overestimating the end values of 'y' as the year progressed. This perhaps is a limitation rooted in the large spatial distribution of data source (geographical diversity, food types, etc.), resulting in non-normal distribution of outbreak data. Therefore, the model performance was evaluated using a 95% confidence region, which accommodated most outliers of native predicted values. However, the data set of 2019 food-borne outbreaks still had two extremist values/outliers for the months of June and November, thereby reducing the model's performance. GAM's failure in compensating extreme outlier data is well observed by other researchers (Bhatt et al., 2017; Shoukri et al., 2004). However, considering the closest possible predictions for the major part of the year, we believe it could possibly be adopted for fairly predicting a relatively close number of outbreaks in India.

As stated previously, food-borne disease outbreaks are preventable. Identification of vulnerable groups can help in developing target specific food safety programs. The majorities of outbreaks reported in India are large and take place in public gatherings like temples, marriage functions, canteens, mid-day-meal at schools, and community festival celebrations, contrary to retail and home setting where most of the small outbreaks occur. Knowledge about the food preparation setting as a possible source of an outbreak is essential to understand the cause and risk factors of an outbreak. For instance, in USA (Dewey-Mattia, Manikonda, Hall, Wise, & Crowe, 2018) the majority of outbreaks are attributed to the food prepared in restaurants (61%) followed by catering and banquet facilities (14%) and home (12%), indicating the regulations related to food handling, preparation and storage practices in restaurant needs to be scrutinized for reducing food-borne outbreaks. Thus, IDSP should add more information about the food preparation setting in their database. Further information about the type of pathogen that infects the food is vital in knowing the food transmission pathway for specific pathogens, limit the spread of the pathogen, and decide the preventive treatment. Monitoring etiological agents will also help in the timely identification of antibiotic-resistant strains and emerging or re-emerging species. Different microbial species may have dominance in specific food categories. For instance, in the USA, outbreaks caused by *Salmonella* were primarily linked with poultry and eggs, whereas Shiga-toxin producing *E. coli* and norovirus were mostly linked with beef and leafy vegetable outbreaks, respectively (Gould et al., 2013). However, data about etiological agents is missing from IDSP reports.

Additionally, the growth of microbes is influenced by environmental conditions, which further reflect seasonal variations in the occurrence of outbreaks. Summers are more appropriate for the growth of bacteria, while viruses such as norovirus and rotavirus grow more rapidly during winters, thus dominate food-borne disease outbreaks differently throughout the year (Gruber et al., 2015; Wu et al., 2018). Most of the food-borne diseases reported in India are bacterial caused by bacteria such as *Staphylococcus aureus*, *Bacillus cereus*, *Escherichia coli*, *Salmonella,* and *Vibrio parahaemolyticus* (Vasanthi & Bhat, 2018); however identification of other microbial species in food may be limited due to lack of facilities. During 1998–2008, viruses replaced bacteria as the leading cause (43%) of food-borne outbreaks in the USA (Gould et al., 2013). A possible reason for this shift may be due to the employment of molecular techniques at local health centers that made the identification of norovirus possible. Further in Netherlands, in 2009, *Toxoplasma gondii*, thermophilic *Campylobacter* spp., rotaviruses, noroviruses and *Salmonella* spp. were the leading cause of foodborne disease burden (Havelaar et al., 2012). So, laboratory facilities with advanced molecular subtyping methods should be developed at the district level to detect a wide variety of pathogens, as most of the outbreaks are limited locally. With the advancement in food processing and the development of long supply chain networks, a pathogen may disseminate to a broader population and result in more severe multistate outbreaks. Food safety needs to be ensured at all levels of farm to fork; therefore, collaborative efforts between policymakers, food business owners, researchers, and consumers are imperative to reduce food-borne disease outbreaks.

However, there are three fundamental limitations to the findings in this report. Firstly, the information on specific food commodities responsible for an outbreak is missing or incomplete in many IDSP reports; the conclusion drawn by assigning the food to one of the 7 categories might not be a true representative of outbreaks with unknown food vehicles. Further, the lack of laboratory confirmation of the implicated food may be misleading. Secondly, only a small fraction of food-borne illnesses that occur each year are identified as outbreaks. For example, most of the outbreaks reported by IDSP happen at a public gathering, and there may be a large sum of outbreaks caused by similar pathogen or food but are unidentified. The true number of outbreaks is underestimated because some outbreaks may be undetected as many people do not seek medical care or are not investigated or not reported to the surveillance department. This also results in poor GAM predictions as a consequence of extreme outliers resulting in under/over estimation of predicted values for food-borne outbreaks. Further, food-borne illnesses have nonspecific symptoms such as nausea, vomiting, and diarrhoea, which can easily be linked to other types of outbreaks. Consequently, the impact of the assumption that the reported outbreaks and food commodities are a random sample of all outbreaks occurring in the population is uncertain. Thus, interpretation of statistical differences in outbreak reporting and the food commodities contributing to outbreak illnesses over time were made based on limited available data. Finally, IDSP is a dynamic surveillance database dependent upon the reporting from state and district health departments; the previously submitted reports may be updated or deleted by the agency whenever new information becomes available. Therefore, the results of the analysis represented in this report are based on the data available at a single point of time on the IDSP website and might differ slightly from future reports.

To conclude, 2,688 food-borne outbreaks were reported in India between 2009 and 2018. Though this number of outbreaks is much lower than the true outbreaks as most outbreaks are not reported; thus, a clear conclusion about vulnerable groups cannot be made. To improve the surveillance, it is recommended to include the laboratory analysis of all outbreaks for the etiological agent and implicated food so that meaningful information may be inferred from the data. Further, most of the outbreaks occur in public settings by consuming freshly prepared food, so the emphasis on making regulations and educating people about hygienic and sanitary practices should be made to improve the food safety across the country.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## CRediT authorship contribution statement

**Akshay Bisht:** collected and analyzed the data, wrote the manuscript. **Manoj P. Kamble:** conceived and designed the study, collected and analyzed the data. **Pritesh Choudhary:** conceived and designed the study, collected and analyzed the data. **Kartikey Chaturvedi:** collected and analyzed the data, wrote the manuscript. **Gautam Kohli:** made critical revisions to the manuscript. **Vijay K. Juneja:** made critical revisions to the manuscript. **Shalini Sehgal:** made critical revisions to the manuscript. **Neetu Kumra Taneja:** conceived and designed the study, wrote the manuscript, All authors reviewed and approved the final manuscript.

## 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.

## Acknowledgment

Authors are grateful to NIFTEM for infrastructural support and encouragement.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2020.107630.

## References

Adams, N., Byrne, L., Edge, J., Hoban, A., Jenkins, C., & Larkin, L. (2019). Gastrointestinal infections caused by consumption of raw drinking milk in England & Wales, 1992–2017. *Epidemiology and Infection, 147*, E281.

Bélanger, P., Tanguay, F., Hamel, M., & Phypers, M. (2015). Food-borne illness: An overview of food-borne outbreaks in Canada reported through outbreak summaries: 2008-2014. *Canada Communicable Disease Report, 41*(11), 254–262.

Bennett, S., Sodha, S., Ayers, T., Lynch, M., Gould, L., & Tauxe, R. (2018). Produce-associated food-borne disease outbreaks, USA, 1998–2013. *Epidemiology and Infection, 146*(11), 1397–1406.

Bhatt, S., Cameron, E., Flaxman, S., Weiss, D., Smith, D., & Gething, P. (2017). Improved prediction accuracy for disease risk mapping using Gaussian process stacked generalization. *Journal of The Royal Society Interface, 14*(134), 20170520.

Ceylan, Z. (2020). Estimation of COVID-19 prevalence in Italy, Spain, and France. *The Science of the Total Environment*, 138817.

Chan, S., & Chan, Z. (2008). A review of food-borne disease outbreaks from 1996 to 2005 in Hong Kong and its implications on food safety promotion. *Journal of Food Safety, 28*(2), 276–299.

Crowe, S. J., Mahon, B. E., Vieira, A. R., & Gould, L. H. (2015). Vital signs: Multistate foodborne outbreaks—United States, 2010–2014. *Morbidity and Mortality Weekly Report, 64*(43), 1221–1225.

Dandage, K., Badia-Melis, R., & Ruiz-García, L. (2017). Indian perspective in food traceability: A review. *Food Control, 71*, 217–227.

Devleesschauwer, B., Haagsma, J., Mangen, M. J., Lake, R., & Havelaar, A. (2018). The global burden of food-borne disease. In R. T (Ed.), *Food safety economics* (pp. 107–122). Cham: Springer.

Dewey-Mattia, D., Manikonda, K., Hall, A., Wise, M., & Crowe, S. (2018). Surveillance for food-borne disease outbreaks—United States, 2009–2015. *MMWR Surveill Summ, 67* (10).

Du, P., Zheng, H., Zhou, J., Lan, R., Ye, C., Jing, H., et al. (2017). Detection of multiple parallel transmission outbreak of Streptococcus suis human infection by use of genome epidemiology, China, 2005. *Emerging Infectious Diseases, 23*(2), 204.

Fabozzi, F. J., Focardi, S. M., Rachev, S. T., & Arshanapalli, B. G. (2014). *The basics of financial econometrics: Tools, concepts, and asset management applications*. John Wiley & Sons.

Ford, L., Miller, M., Cawthorne, A., Fearnley, E., & Kirk, M. (2015). Approaches to the surveillance of food-borne disease: A review of the evidence. *Food-borne Pathogenic Disease, 12*(12), 927–936.

Gibbons, C., Mangen, M., Plass, D., Havelaar, A., Brooke, R., Kramarz, P., et al. (2014). Measuring underreporting and under-ascertainment in infectious disease datasets: A comparison of methods. *BMC Public Health, 14*(1), 147.

Gould, L., Walsh, K., Vieira, A., Herman, K., Williams, I., Hall, A., et al. (2013). Surveillance for food-borne disease outbreaks—United States, 1998–2008. *MMWR Surveill Summ, 62*(2), 1–34.

Gruber, J., Bailey, J., & Kowalcyk, B. (2015). Evaluation of US poison center data for surveillance of food-borne disease. *Food-borne Pathogenic Disease, 12*(6), 467–478.

Hald, T., Aspinall, W., Devleesschauwer, B., Cooke, R., Corrigan, T., Havelaar, A., et al. (2016). World health organization estimates of the relative contributions of food to the burden of disease due to selected food-borne hazards: A structured expert elicitation. *PloS One, 11*(1), Article e0145839.

Hall, G., Kirk, M., Becker, N., Gregory, J., Unicomb, L., Millard, G., et al. (2005). Estimating foodborne gastroenteritis, Australia. *Emerging Infectious Diseases, 11*(8), 1257.

Havelaar, A. H., Haagsma, J. A., Mangen, M.-J. J., Kemmeren, J. M., Verhoef, L. P., Vijgen, S. M., et al. (2012). Disease burden of foodborne pathogens in The Netherlands, 2009. *International Journal of Food Microbiology, 156*(3), 231–238.

Havelaar, A., Kirk, M., Torgerson, P., Gibb, H., Hald, T., Lake, R., et al. (2015). World Health Organization global estimates and regional comparisons of the burden of food-borne disease in 2010. *PLoS Medicine, 12*(12), Article e1001923.

He, J., Li, H., Edmondson, A. C., Rader, D. J., & Li, M. (2012). A Gaussian copula approach for the analysis of secondary phenotypes in case–control genetic association studies. *Biostatistics, 13*(3), 497–508.

IDSP. (2019a). Diseases under surveillance. In *About IDSP*.

IDSP. (2019b). IDSP achievements. In *About IDSP*.

Jahan, S. (2012). Epidemiology of food-borne illness. In B. Valdez (Ed.), *Scientific, health and social aspects of the food industry* (pp. 321–342). InTech.

Kristkova, Z., Grace, D., & Kuiper, M. (2017). *The economics of food safety in India: A rapid assessment*. Netherlands: Wageningen University & Research.

Lindström, T., Tildesley, M., & Webb, C. (2015). A bayesian ensemble approach for epidemiological projections. *PLoS Computational Biology, 11*(4), Article e1004187.

Mittag, F., Büchel, F., Saad, M., Jahn, A., Schulte, C., Bochdanovits, Z., et al. (2012). Use of support vector machines for disease risk prediction in genome-wide association studies: Concerns and opportunities. *Human Mutation, 33*(12), 1708–1718.

Odeyemi, O. (2016). Public health implications of microbial food safety and food-borne diseases in developing countries. *Food & Nutrition Research, 60*.

OzFoodNet Working Group. (2018). Monitoring the incidence and causes of diseases potentially transmitted by food in Australia: Annual report of the OzFoodNet network 2012. *Communicable Diseases Intelligence, 42*, S2209–S6051.

Pattis, I., Lopez, L., Cressey, P., Horn, B., & Roos, R. (2017). *Food-borne disease in New Zealand 2016*. Wellington: Ministry for Primary Industries.

Pijnacker, R., Friesema, I., Mughini Gras, L., Lagerweij, G., van Pelt, W., & Franz, E. (2019). *Disease burden of food-related pathogens in The Netherlands*. Bilthoven: National Institute for Public Health and the Environment, 2018.

Scharff, R., Besser, J., Sharp, D., Jones, T., Peter, G., & Hedberg, C. (2016). An economic evaluation of PulseNet: A network for food-borne disease surveillance. *American Journal of Preventive Medicine, 50*, S66–S73.

Shoukri, M., Asyali, M., Van Dorp, R., & Kelton, D. (2004). The Poisson inverse Gaussian regression model in the analysis of clustered counts data. *Journal of Data Science, 2*(1), 17–32.

Sudershan, R., Naveen, K., Kashinath, L., Bhaskar, V., & Polasa, K. (2014). Food-borne infections and intoxications in Hyderabad India. *Epidemiology Research International*, 942961.

Tapak, L., Hamidi, O., Fathian, M., & Karami, M. (2019). Comparative evaluation of time series models for predicting influenza outbreaks: Application of influenza-like illness data from sentinel sites of healthcare centers in Iran. *BMC Research Notes, 12*(1), 353.

Vaillant, V., Valk, H. D., Baron, E., Ancelle, T., Colin, P., Delmas, M.-C., et al. (2005). Foodborne infections in France. *Foodbourne Pathogens & Disease, 2*(3), 221–232.

Van Cauteren, D., Le Strat, Y., Sommen, C., Bruyand, M., Tourdjman, M., Da Silva, N., et al. (2017). Estimated annual numbers of food-borne pathogen–associated illnesses, hospitalizations, and deaths, France, 2008–2013. *Emerging Infectious Diseases, 23*(9), 1486–1492.

Vasanthi, S., & Bhat, R. (2018). Management of food safety risks in India. *Proceedings of the Indian National Science Academy, 84*(4), 937–943.

World Health Organisation. (2015). *WHO estimates of the global burden of food-borne diseases. Food-borne diseases burden epidemiology reference group 2007-2015*. Geneva: World Health Organisation.

Wu, Y., Liu, X., Chen, Q., Liu, H., Dai, Y., Zhou, Y., et al. (2018). Surveillance for food-borne disease outbreaks in China, 2003 to 2008. *Food Control, 84*, 382–388.

Yu, W., Liu, T., Valdez, R., Gwinn, M., & Khoury, M. J. (2010). Application of support vector machine modeling for prediction of common diseases: The case of diabetes and pre-diabetes. *BMC Medical Informatics and Decision Making, 10*(1), 16.

Zhang, X., Zhang, T., Young, A. A., & Li, X. (2014). Applications and comparisons of four time series models in epidemiological surveillance data. *PloS One, 9*(2), Article e88075.

Zhang, H., Zheng, L., & Feng, L. (2019). Epidemiology, diagnosis and treatment of moyamoya disease. *Experimental and Therapeutic Medicine, 17*(3), 1977–1984.