



A hybrid approach to study and forecast climate-sensitive norovirus infections in the USA

Juping Ji ^{a,b}, Shohel Ahmed ^c, Hao Wang ^{c,*}

^a School of Mathematics and Information Sciences, Guangzhou University, Guangzhou, Guangdong, 510006, PR China

^b Guangzhou Center for Applied Mathematics, Guangzhou University, Guangzhou, Guangdong, 510006, PR China

^c Department of Mathematical and Statistical Sciences & Interdisciplinary Lab for Mathematical Ecology and Epidemiology, University of Alberta, Edmonton, AB T6G 2R3, Canada

ARTICLE INFO

MSC:

34D20

49M05

92B05

Keywords:

Norovirus

Transmission rate

Inverse method

Environmental conditions

Machine learning

Generalized boosting model

ABSTRACT

Norovirus, responsible for acute gastroenteritis and foodborne diseases in the United States, is influenced significantly by environmental factors. This study employs a hybrid approach to develop a foodborne disease model that incorporates indirect incidence to examine the correlation between norovirus outbreaks and environmental conditions, specifically focusing on the impact of temperature and humidity on virus transmission. By analyzing weekly average climate data and confirmed case data from four United States regions (Southern, Northeastern, Midwestern, and Western), we assess the mortality rates and estimate transmission rates using the inverse method. Our numerical results confirm that norovirus outbreaks predominantly occur in colder months. However, higher temperatures or increased humidity during warmer months appear to mitigate the spread of the virus. Utilizing climate data, this study also forecasts transmission rates and infection cases up to eight weeks in advance using a generalized boosting machine learning model.

1. Introduction

Norovirus (NoV) is a highly infectious agent responsible for causing acute gastroenteritis and foodborne illness. Symptoms such as nausea, vomiting, diarrhea, and abdominal cramps typically manifest within 12 to 48 h following exposure to the virus (Anon, 2023a). Most individuals recover from norovirus symptoms within one to three days. Due to its highly contagious nature, norovirus can rapidly infect individuals across all age groups (Morioka et al., 2006; Siebenga et al., 2009). In the United States, norovirus leads to millions of intestinal illness cases annually, affecting approximately one in every fifteen residents (Hall et al., 2014). Currently, there are no vaccines or specific treatments available for norovirus, making prevention reliant on strict personal hygiene practices and isolation during outbreaks.

Norovirus is recognized as a seasonal virus, with its outbreaks being influenced by environmental factors like temperature and humidity (Adler and Zickl, 1969; Chenar and Deng, 2017a,b; Colas de la Noue et al., 2014; Kim et al., 2012; Lopman et al., 2009; Rohayem, 2009; Verhoef et al., 2008). Research conducted by Chenar and

Deng highlights how these environmental elements impact norovirus transmissions; specifically, they discovered that low temperatures ranging from -6.6 °C to 20 °C, relative humidity levels between 10% and 66%, and rainfall occurring from one day to three months before an outbreak are significant contributors to the prevalence of norovirus (Chenar and Deng, 2017a). Similarly, Lopman et al. observed that in every region of England and Wales, temperature and humidity consistently affect norovirus outbreaks (Lopman et al., 2009). Their findings suggest that factors such as lower temperatures, reduced relative humidity, diminished population immunity, and the emergence of new norovirus strains are all independently associated with increased norovirus activity. Furthermore, numerous studies have indicated that norovirus outbreaks typically occur in the colder months, with a notable peak during winter (Adler and Zickl, 1969; Mounts et al., 2000).

Norovirus, known for its rapid and versatile transmission, imposes a significant economic burden due to its numerous transmission routes. According to the Centers for Disease Control and Prevention (CDC) (Anon, 2023a), norovirus can infect humans through various

* Corresponding author.

E-mail address: hao8@ualberta.ca (H. Wang).

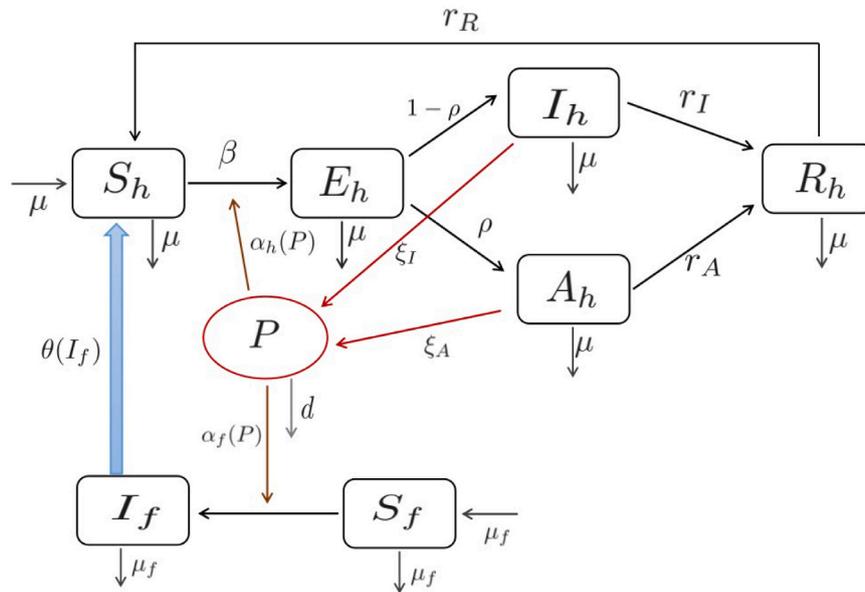


Fig. 1. Flow of disease transmission within the System (2.1) is depicted through various colored lines representing different processes: Black lines indicate the transitions between different health statuses. Red lines illustrate the virus shedding from infected individuals. Orange lines denote indirect infections resulting from environmental factors. The blue line represents transmission due to consuming contaminated shellfish.

methods including inhalation of particles from vomit, fecal-oral spread, consumption of contaminated food and water, and contact with contaminated environmental surfaces. The virus primarily spreads through direct person-to-person contact and foodborne transmission, making it a leading cause of foodborne diseases in the United States. Often, outbreaks are associated with food contaminated during preparation by infected restaurant workers.

Shellfish, particularly those harvested from waters contaminated with untreated sewage, are frequently linked to norovirus outbreaks (Campos and Lees, 2014; Chenar and Deng, 2018; Prato et al., 2004; Sugieda et al., 1996; Westrell et al., 2010). Consuming raw or undercooked shellfish, especially oysters that have been exposed to fecal contamination, poses a significant risk for infection (Chenar and Deng, 2018). Research, including studies by Prato et al. has highlighted raw shellfish consumption as a major risk factor (Prato et al., 2004). Additionally, Westrell et al. noted that individuals with compromised immune systems or chronic diseases are particularly vulnerable to norovirus when consuming contaminated oysters, and incidents of norovirus outbreaks linked to oyster consumption are on the rise (Westrell et al., 2010).

This paper is structured as follows: Section 2 provides an in-depth exploration of the Norovirus compartmental model, specifically focusing on the integration of environmental variables that influence the transmission dynamics. Section 3 assesses the mortality rates of norovirus across four distinct regions in the USA, each with varying environmental conditions. In Section 4, we analyze a time-dependent transmission rate and examine the impact of weather conditions on this rate. Section 5 describes how we use a machine learning approach combined with climate data to predict weekly infection cases. The limitations of our study are discussed in Section 6. Finally, our findings are presented and discussed in the concluding Section 7.

2. Model formulation

Norovirus infection can occur through various transmission routes (Bitler et al., 2013; Campos and Lees, 2014; Mathijs et al., 2012; Matsuyama et al., 2017; Rushton et al., 2019; Ushijima et al., 2014; S. Xiao et al., 2017), including close contact with an infected individual, touching surfaces contaminated with the virus and then touching the

face, and consuming contaminated food or beverages. Direct person-to-person transmission is generally considered the primary pathway for spreading norovirus. It is important to note that norovirus infections can be asymptomatic (CDC, 2011). However, in many outbreaks, contaminated food and environmental factors also play crucial roles (Scallan et al., 2011; Smith et al., 2012; Ushijima et al., 2014; Widdowson et al., 2005). The persistence of the virus in the environment significantly influences its transmission. This study examines these three main routes to understand the transmission dynamics of norovirus.

To model the dynamics of norovirus transmission, we utilize the Susceptible–Exposed–Infectious–Recovered framework. In our model, the total human population size (N_h) is segmented into five categories: susceptible (S_h), exposed (E_h), symptomatic infected (I_h), asymptomatic infected (A_h), and recovered (R_h) individuals, with $N_h = S_h + E_h + I_h + A_h + R_h$. Additionally, we denote the density of norovirus on environmental surfaces as P . Considering the consumption of raw shellfish as a risk factor for norovirus, we define S_f and I_f as the susceptible and infected shellfish populations, respectively. The compartmental model’s schematic diagram is outlined below.

Cold and dry environmental conditions, characterized by low temperatures and relative humidity, influence the spread of norovirus. Given the connection between environmental factors and norovirus outbreaks, we integrate temperature and humidity into our compartmental model. We propose the following infectious disease model aims to investigate how these environmental variables impact the transmission dynamics of norovirus:

$$\begin{cases} \frac{dS_h}{dt} = \mu N_h - \frac{\beta(T,H)S_h(I_h+E_h+\theta_A A_h)}{N_h} - \alpha_h(P)S_h - c\theta(I_f)S_h - \mu S_h + r_R R_h, \\ \frac{dE_h}{dt} = \frac{\beta(T,H)S_h(I_h+E_h+\theta_A A_h)}{N_h} + \alpha_h(P)S_h + c\theta(I_f)S_h - (\delta + \mu)E_h, \\ \frac{dI_h}{dt} = (1 - \rho)\delta E_h - (r_I + \mu)I_h, \\ \frac{dA_h}{dt} = \rho\delta E_h - (r_A + \mu)A_h, \\ \frac{dR_h}{dt} = r_I I_h + r_A A_h - (r_R + \mu)R_h, \\ \frac{dP}{dt} = \xi_I I_h + \xi_A A_h - d(T, H)P, \\ \frac{dS_f}{dt} = \mu_f N_f - \alpha_f(P)S_f - \mu_f S_f, \\ \frac{dI_f}{dt} = \alpha_f(P)S_f - \theta(I_f) - \mu_f I_f \end{cases} \quad (2.1)$$

Table 1
The parameter description of System (2.1).

Parameter	Description	Unit	Value	Source
β	Transmission rate for direct transmission	week ⁻¹		
θ_E	Relative transmissibility of Exposed individuals		0.55	Assumed
θ_A	Relative transmissibility of asymptomatic infected individuals		0.55	Assumed
$1/\delta$	The mean length of latent period	week	0.57	Assumed
ρ	Proportion of asymptomatic infected individuals		0.6	Assumed
μ	Natural death rate	week ⁻¹	0.00025	Lane et al. (2019)
r_I	Symptomatic infected individuals recovery rate	week ⁻¹	0.583	Assumed
r_A	Asymptomatic infected individuals recovery rate	week ⁻¹	1	Assumed
r_R	Rate at which recovered individuals lose immunities	week ⁻¹	0.038	Lane et al. (2019)
ξ_I	Virus shedding rate from symptomatic infected individuals	cells ind ⁻¹ week ⁻¹	94.5	Assumed
ξ_A	Virus shedding rate from asymptomatic infected individuals	cells ind ⁻¹ week ⁻¹	23.8	Assumed
d	Mortality rate of Norovirus	week ⁻¹		Estimated
a_1	Maximum rate of infection	week ⁻¹	0.00015	Assumed
a_2	Maximum harvesting rate	Number week ⁻¹	0.00015	Assumed
a_3	Maximum rate of infection	week ⁻¹	0.0001	Assumed
b_1	Half-saturation of the virus density on environmental surfaces	cells	3 000 000	Assumed
b_2	Half-saturation of the harvesting infected shellfish	Number	500 000 000	Assumed
b_3	Half-saturation of the virus density in the sea	cells	2 000 000	Assumed
c	Infection due to contact with shellfish	1/Number	0.0001	Assumed
μ_f	Mortality rate of shellfish	week ⁻¹	0.0006	Assumed

with nonnegative initial conditions $S_h(0), E_h(0), I_h(0), A_h(0), R_h(0), P(0), S_f(0), I_f(0)$. A schematic diagram for the system (2.1) is presented in Fig. 1.

Here, $\beta(T, H)$ denotes the transmission parameter for direct transmission, which is influenced by temperature and humidity, where T represents the temperature in degrees Celsius and H signifies the percent relative humidity. θ_E and θ_A indicate the relative transmissibility of exposed and asymptomatic infected individuals, respectively (Ozawa et al., 2007). The term $1/\delta$ represents the incubation period. Recovery rates for symptomatic and asymptomatic infected individuals are denoted by r_I and r_A , respectively. Asymptomatic infections comprise a proportion ρ , while symptomatic infections make up $1 - \rho$. The natural death rate of individuals is represented by μ . Since recovered individuals may lose immunity and become susceptible to reinfection with norovirus (CDC, 2011), r_R is used to denote the rate at which immunity is lost. The virus shedding rates for symptomatic and asymptomatic individuals are indicated by ξ_I and ξ_A , respectively (Atmar et al., 2008). The parameter $d(T, H)$ represents the death rate of norovirus on environmental surfaces, also dependent on temperature and humidity. The indirect components of the incident term are $\alpha_h(P)$ and $\alpha_f(P)$, defined as $\alpha_h(P) = \frac{a_1 P}{P + b_1}$ and $\alpha_f(P) = \frac{a_3 P}{P + b_3}$. μ_f stands for the natural death rate of shellfish, and N_f represents the total population size of shellfish. The infection rate from consuming infected shellfish is given by $\theta(I_f) = \frac{a_2 I_f}{I_f + b_2}$. Detailed descriptions of these parameters are provided in Table 1 for the system (2.1).

3. Evaluation of $d(T, H)$

In 2014, Colas de la Noue et al. (2014) carried out experiments to investigate how relative humidity and temperature affect the survival of norovirus. Their study measured the persistence of the virus over a 20-h period under varying humidity levels, ranging from low (10% RH) to saturated (100% RH), at temperatures of 9 °C and 25 °C. The findings, detailed in Table 2, indicate that both temperature and relative humidity significantly influence norovirus survival. Specifically, the data shows that higher temperatures tend to reduce the survival time of the virus under consistent humidity conditions.

Based on the data from Table 2, we applied linear regression to model the relationship among temperature, relative humidity, and the survival of norovirus on surfaces. We fit a linear equation to estimate the mean survival time of norovirus ($Survival_{NoV}$) as a function of temperature in degrees Celsius (T) and percent relative humidity (H),

Table 2
Relationship between relative humidity and norovirus survival at different temperatures. Source: Adapted from Colas de la Noue et al. (2014).

Temperature (°C)	Relative humidity (%)	Mean survival (%)
9	10	63.9
	35	55.6
	55	24.3
	85	56.4
	100	59.5
25	10	51
	35	14.9
	55	Not detected
	85	Not detected
	100	3.7

measured in weeks. The resulting model is represented by the equation:

$$Survival_{NoV} = 0.815986 - 0.019067T - 0.002193H.$$

This equation suggests that increases in both temperature and humidity are associated with a decrease in the survival time of norovirus on surfaces.

Given that e^{-dt} represents the survival probability of the virus at time t , and using the function for the survival of norovirus on surfaces, we relate it to the survival probability with $e^{-dt} = Survival_{NoV}$. Setting $t = \frac{5}{42}$ weeks, we aim to calculate the mortality rate of norovirus under various weather conditions weekly. The formula for calculating the mortality rate $d(T, H)$ becomes:

$$d(T, H) = -8.4 \ln(Survival_{NoV}) = -8.4 \ln(0.815986 - 0.019067T - 0.002193H). \tag{3.1}$$

To assess the impact of environmental conditions on the mortality rate of norovirus, we selected four regions within the USA: Southern, Northeastern, Midwestern, and Western. We gathered weekly average climate data spanning from January 19, 2022, to December 31, 2023, to calculate the mortality rates of norovirus (Anon, 2023b). To mitigate regular weekly fluctuations, we utilized a centered 3-week moving average approach. The weekly average temperatures and relative humidity for these regions are depicted in Fig. 2. Notably, the Southern region exhibits higher temperatures compared to the other regions, whereas temperatures in the Midwestern region are generally lower. In terms of humidity, the Southern and Western regions report lower average relative humidity levels than the Northeastern and Midwestern regions.

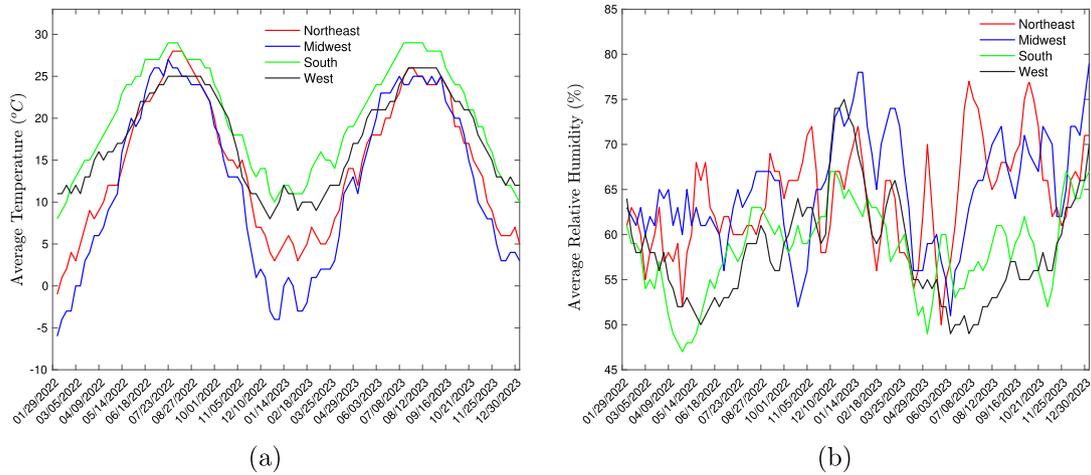


Fig. 2. Weekly average temperature and relative humidity across four regions in the USA.

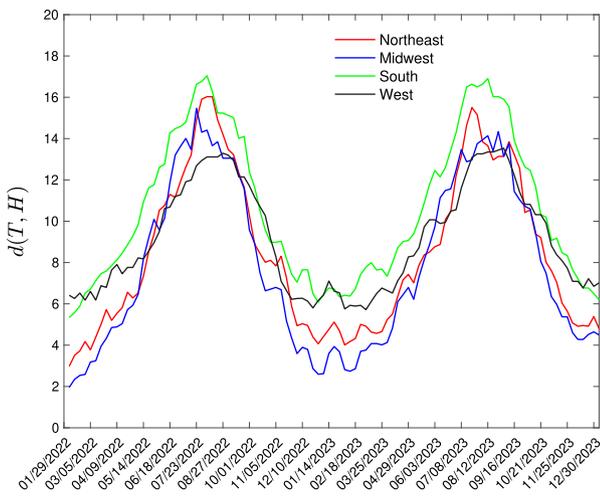


Fig. 3. Weekly mortality rates of norovirus on surfaces across four regions in the USA.

After substituting the weekly temperature and relative humidity data into our mortality rate function $d(T, H)$ as outlined in Eq. (3.1), we calculated the weekly mortality rates of norovirus for each of the four regions: Southern, Northeastern, Midwestern, and Western. These rates are visually represented in Fig. 3. From this figure, it is evident that the mortality rate of the virus on surfaces is consistently higher in the Southern region compared to the other regions. Additionally, we observe that the mortality rates across all regions are lower during the colder months (November to April) and higher during the warmer months (June to October). This pattern supports the conclusion that higher temperatures accelerate the decline in norovirus survival, resulting in increased viral mortality rates.

4. A time-varying transmission rate

In this section, we investigate how environmental factors influence the transmission of norovirus by examining the time-varying transmission rate, $\beta(t)$, across four regions in the USA. Weekly, laboratories report the total number of norovirus tests conducted and the positives identified to the CDC, which then calculates and publishes the average percentage of positive tests (Anon, 2023c). Annually, the USA reports between 19 to 21 million cases of norovirus. For our study, we assume that approximately 1.5% of the population undergoes testing each week. Using this CDC data (Anon, 2023c), we have generated a plot

(Fig. 4) that displays the weekly confirmed cases of norovirus across the four regions. The plot indicates that the number of confirmed cases typically peaks during the late fall, winter, and early spring.

Inverse method (Ji et al., 2023; Kong et al., 2015; Pollicott et al., 2012; Wang and Wang, 2023; Wang et al., 2022a,b), a discrete method for estimating the time-varying transmission rate inversely, is an effective approach for determining the transmission rates of infectious diseases. By utilizing data from the weekly confirmed cases in four regions of the USA, we will estimate the transmission rates in each of these regions separately using the inverse method.

Now, let $S_h[i]$, $E_h[i]$, $I_h[i]$, $A_h[i]$, $R_h[i]$, $P[i]$, $S_f[i]$, and $I_f[i]$ represent the state variables of the compartmental model system (denoted as System (2.1)) for week i . We use the term $(1-p)\delta E_h(t)$ to approximate the notification data, where $M[i] = (1-p)\delta E_h(t)$ represents the weekly notification data for the i th week. This allows us to express $E_h[i]$ as $\frac{M[i]}{(1-p)\delta}$ for each week i , where $i = 1, 2, \dots, K$, and K is the total number of weeks for which notification data are available. With the relationship $N_h = S_h(t) + E_h(t) + I_h(t) + A_h(t) + R_h(t)$ defining the total population in the human compartment, we can determine the total population at the beginning of our observational period as $N_h = S_h[1] + E_h[1] + I_h[1] + A_h[1] + R_h[1]$. This calculation helps track the progression of the infection through different stages within the model framework (see Figs. 6 and 7). From System (2.1), it follows that $I_h[i] = I_h[i-1] + (1-p)\delta E_h[i-1] - (r_I + \mu)I_h[i-1]$, $A_h[i] = A_h[i-1] + p\delta E_h[i-1] - (r_A + \mu)A_h[i-1]$, $R_h[i] = R_h[i-1] + r_I I_h[i-1] + r_A A_h[i-1] - (r_R + \mu)R_h[i-1]$, $P[i] = P[i-1] + \xi_I I_h[i-1] + \xi_A A_h[i-1] - d(T, H)P[i-1]$, $S_f[i] = S_f[i-1] + \mu_f N_f - \alpha_f(P[i-1])S_f[i-1] - \mu_f S_f[i-1]$, $I_f[i] = I_f[i-1] + \alpha_f(P[i-1])S_f[i-1] - \theta(I_f[i-1]) - \mu_f S_f[i-1]$, $S_h[i] = N_h - E_h[i] - I_h[i] - A_h[i] - R_h[i]$, for $i = 2, 3, \dots, K$.

$$\beta[i-1] = \frac{[E_h[i] - E_h[i-1] + (\delta + \mu)E_h[i-1] - \alpha_h(P[i-1])S_h[i-1] - c\theta(I_f[i-1])S_h[i-1]]N_h}{S_h[i-1](I_h[i-1] + \theta_E E_h[i] + \theta_A A_h[i-1])}$$

for $i = 2, 3, \dots, K-1$, and $\beta[K] \approx \beta[K-1]$.

According to the United States Census regions (Anon, 2023d), the total populations of the northeastern, midwestern, southern, and western regions in 2023 are 56,983,517, 68,909,283, 130,125,290, and 78,896,805, respectively. The parameter values are detailed in Table 1. Figs. 5(a) through 8(a) display the transmission rates $\beta(t)$ of norovirus in these four regions, derived using the inverse method. The data reveal that transmission rates are higher during the cold months compared to the warm months, confirming that norovirus transmission is lower during the warm summer months, while the colder, drier winter months tend to facilitate its spread. Subsequently, we use the derived transmission rates $\beta(t)$ to model the weekly notification data in each region.

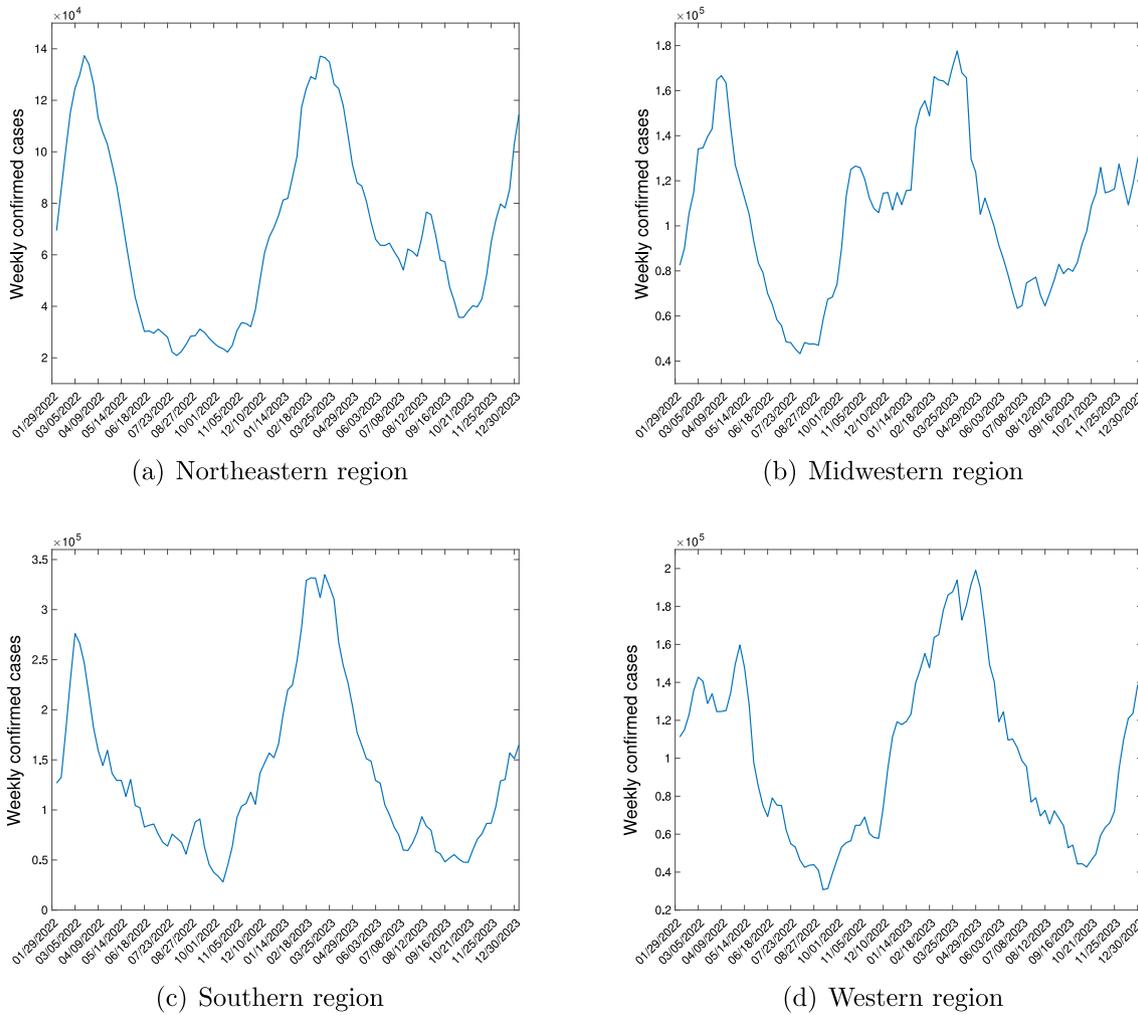


Fig. 4. Weekly confirmed cases of norovirus across four regions in the USA.

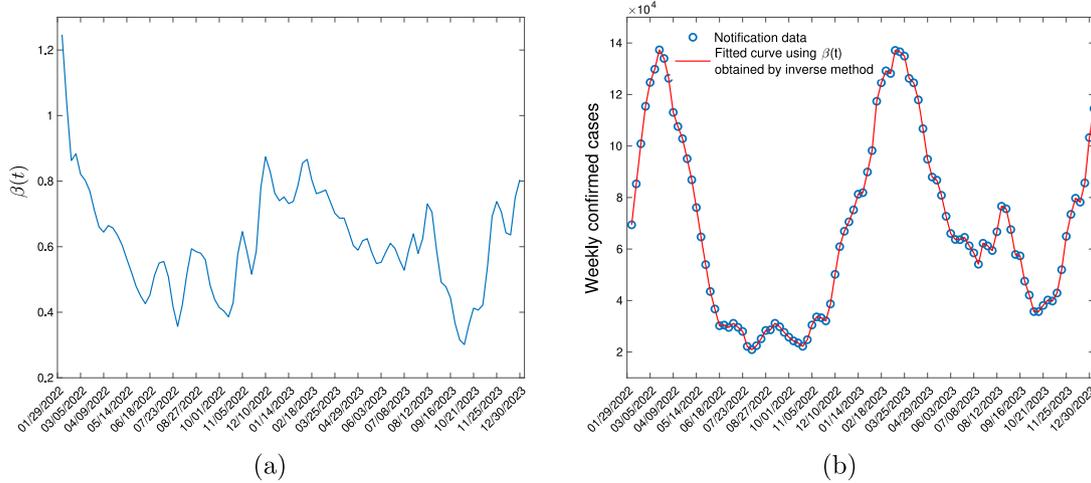


Fig. 5. Transmission rate in Northeastern region obtained by the inverse method and the fitting with notification data. Initial condition is $(S_h[1], E_h[1], I_h[1], A_h[1], R_h[1], P[1], S_f[1], I_f[1]) = (55334378, 99139, 250000, 500000, 800000, 200000, 1000000, 10000)$.

By incorporating the weekly transmission rate $\beta(t)$ into System (2.1), we calculate $E_h(t)$ and then plot the curve of $(1 - p)\delta E_h(t)$ to compare it with the notification data. The fitting results, shown in Figs. 5(b) through 8(b), demonstrate that the transmission rates obtained through

the inverse method align nearly perfectly with the notification data across all four regions.

In Fig. 9, we present the weekly confirmed cases and transmission rates of norovirus, derived using the inverse method, to examine the

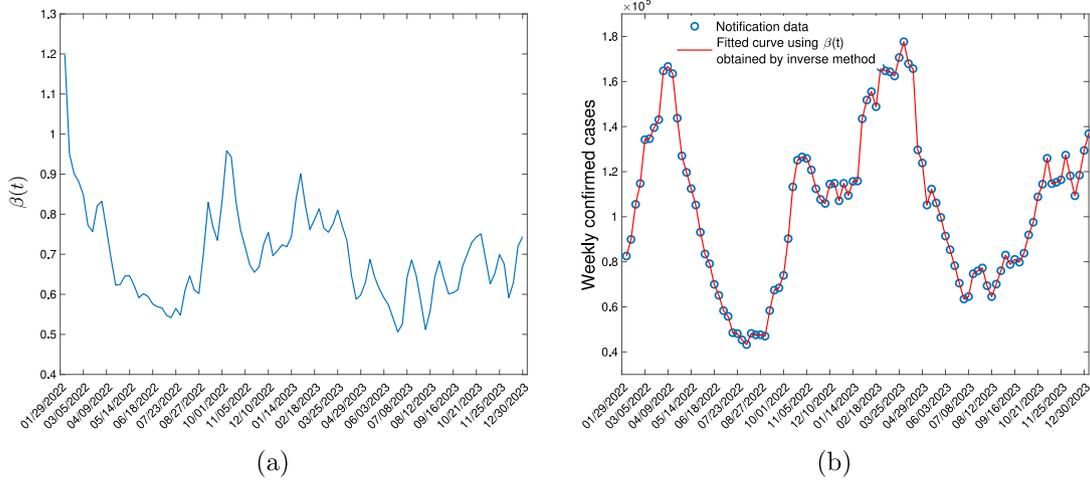


Fig. 6. Transmission rate in Midwestern region obtained by the inverse method and the fitting with notification data. Initial condition is $(S_h[1], E_h[1], I_h[1], A_h[1], R_h[1], P[1], S_f[1], I_f[1]) = (66591360, 117923, 400000, 800000, 1000000, 200000, 1000000, 10000)$.

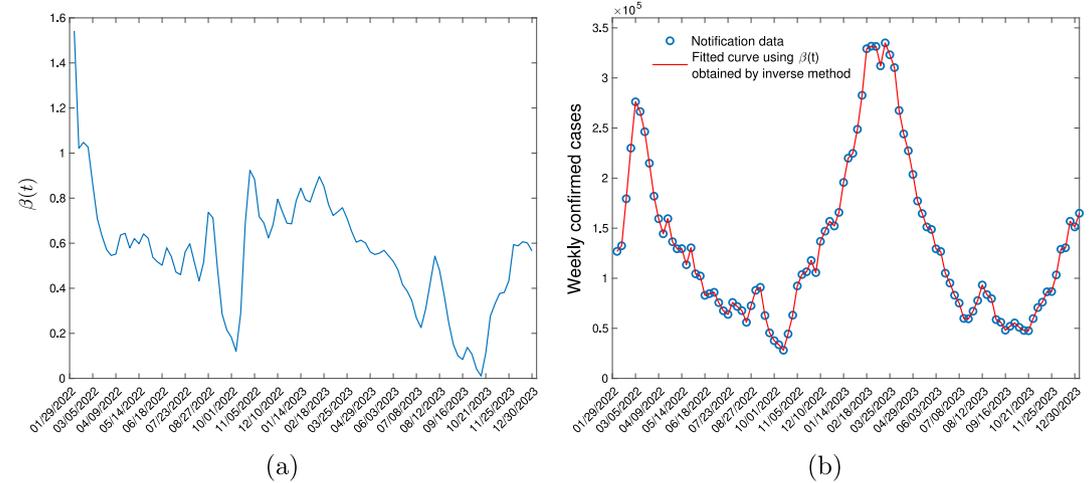


Fig. 7. Transmission rate in Southern region obtained by the inverse method and the fitting with notification data. Initial condition is $(S_h[1], E_h[1], I_h[1], A_h[1], R_h[1], P[1], S_f[1], I_f[1]) = (115944128, 181162, 4000000, 8000000, 2000000, 200000, 1000000, 10000)$.

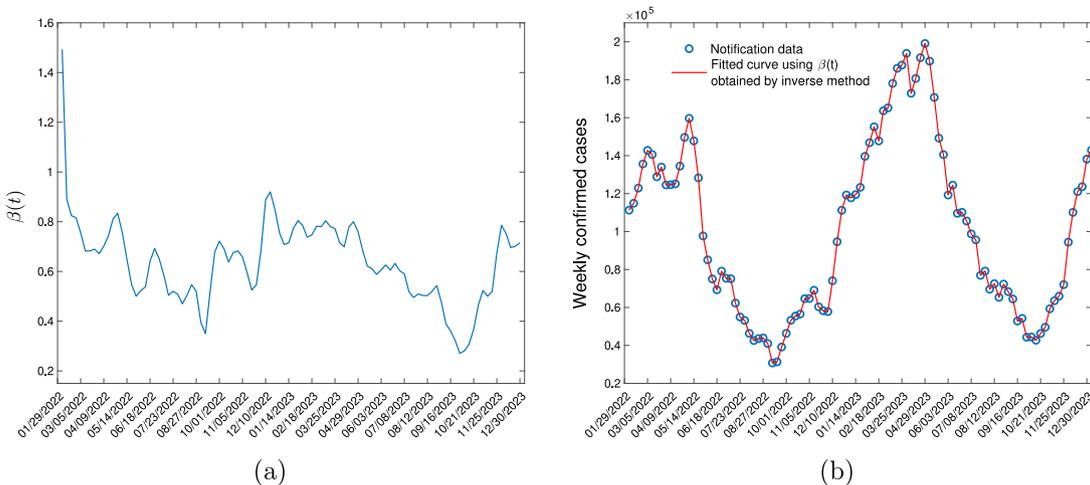


Fig. 8. Transmission rate in Western region obtained by the inverse method and the fitting with notification data. Initial condition is $(S_h[1], E_h[1], I_h[1], A_h[1], R_h[1], P[1], S_f[1], I_f[1]) = (76837866, 158939, 300000, 600000, 1000000, 200000, 1000000, 10000)$.

impact of environmental conditions on the virus's transmission. As depicted in the figure, the confirmed cases during the colder months

are notably higher in the Southern regions compared to other areas. Interestingly, the transmission rates in the Southern region closely

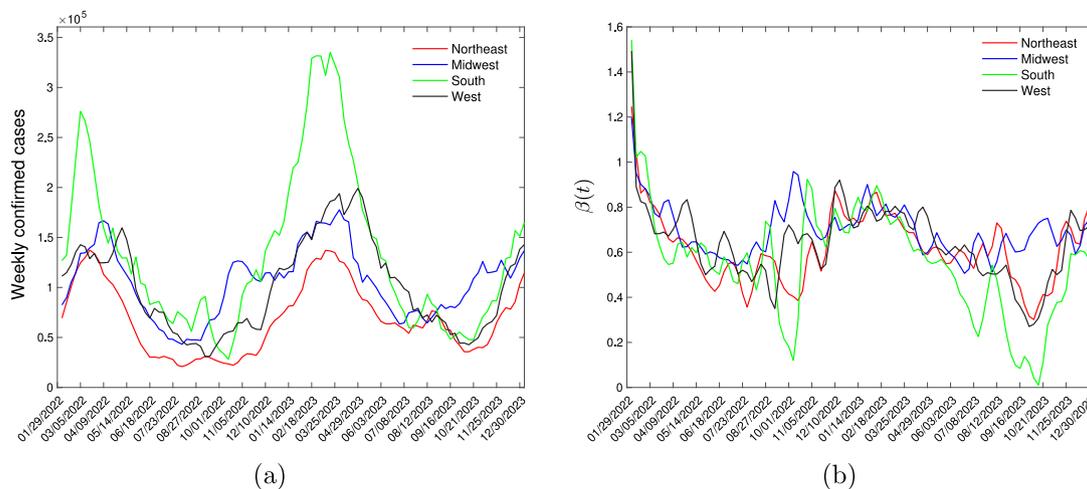


Fig. 9. Weekly confirmed cases and transmission rates of norovirus across four regions in the USA.

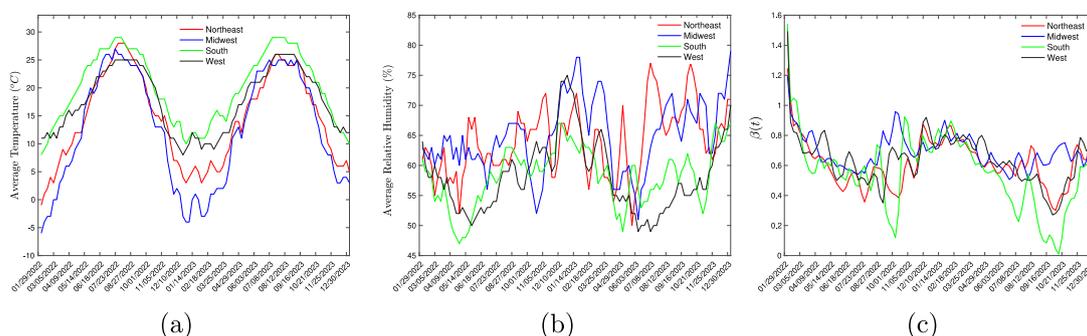


Fig. 10. Weekly average temperature, relative humidity and transmission rate across four regions in the USA.

mirror those in the other three regions. This observation coincides with the fact that norovirus mortality rates are higher in the Southern region during colder months and that the virus does not survive well in higher temperatures.

In Fig. 10, we show the relationship among temperature, relative humidity and the transmission rate obtained by the inverse method. We observe that when the temperature or humidity is increasing, the transmission rate will decrease in the overall trend. Norovirus outbreaks are more frequent in colder months, whereas higher temperatures and increased humidity during warmer months seem to mitigate virus spread. These findings align with the hypothesis that norovirus transmission is lower in warmer climates and that such conditions help reduce the spread of the virus.

5. Machine learning and prediction

In this section, we aim to investigate the relationship between the transmission rate of norovirus and climatic factors using machine learning techniques. Specifically, we will utilize the gradient boosting machine (GBM), a robust predictive tool (Friedman, 2001; Natekin and Knoll, 2013). Our approach involves applying GBM to model how the average temperature and relative humidity influence the transmission rate. We will conduct this analysis using the gbm package and the prediction function within the R statistical software environment. This methodology will help us accurately predict transmission rates based on environmental conditions.

In this analysis, we focus on the Western region, where we note a notably low mortality rate and an uptick in weekly confirmed cases beginning in December 2022. We set the start date for our training data

on January 19, 2022, and progressively extend the training duration starting from 40 weeks. As detailed in Table 3, we train the Gradient Boosting Machine (GBM) on various durations, increasing incrementally from 40 to 80 weeks in 4-week intervals, and then test the models over the subsequent 8 weeks for each training period. The training dataset consists of the weekly transmission rate determined via the inverse method, alongside relevant climatic factors. With the climatic data provided for the testing periods, we employ the trained GBM models to predict future transmission rates. These predictions allow us to estimate weekly infection cases by applying the series of trained and tested weekly transmission rates to the equation $(1-p)\delta E_h(t)$ within System (2.1). We then plot this curve and compare the predicted infection rates against the actual notification data of confirmed cases, enabling a comprehensive evaluation of the model’s accuracy and effectiveness in predicting norovirus transmission under varying climatic conditions.

In statistics, there are several methods to evaluate the accuracy of forecast models against actual outcomes. Two commonly used measures are Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). MAE quantifies the average magnitude of errors in a set of predictions, without considering their direction, while MAPE provides a measure of prediction accuracy as a percentage, which is particularly useful for comparison across different datasets (Khair et al., 2017; Willmott and Matsuura, 2005). We plan to employ both MAE and MAPE to assess the discrepancies between the predicted infection cases and the actual confirmed cases, as well as the differences between the transmission rates predicted by GBM models and those calculated using the inverse method. The formulas for MAPE and MAE are as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - x_i}{x_i} \right| \quad \text{and} \quad MAE = \frac{1}{n} \sum_{i=1}^n |y_i - x_i|$$

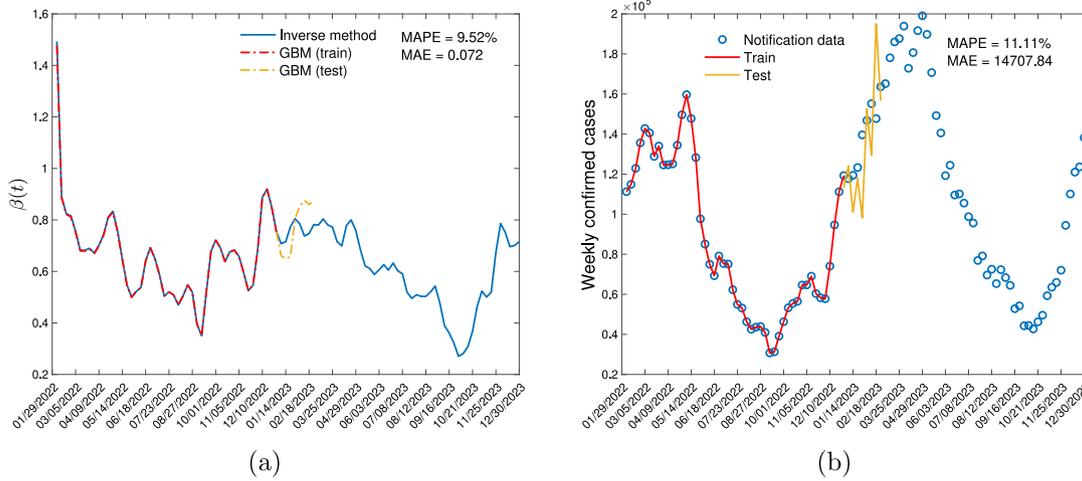


Fig. 11. Using climate data in the western region, train 48 weeks from January 19 to December 3, 2022, and test 8 weeks from December 4, 2022 to January 28, 2023. (a) Transmission rates and predicted transmission rates. (b) Weekly confirmed cases and predicted infection cases.

Table 3
Training and testing durations.

Train length (weeks)	Train duration	Test duration
40	Jan 19, 2022–Oct 8, 2022	Oct 9, 2022–Dec 3, 2022
44	Jan 19, 2022–Nov 5, 2022	Nov 6, 2022–Dec 31, 2022
48	Jan 19, 2022–Dec 3, 2022	Dec 4, 2022–Jan 28, 2023
52	Jan 19, 2022–Dec 31, 2022	Jan 1, 2023–Feb 25, 2023
56	Jan 19, 2022–Jan 28, 2023	Jan 29, 2023–Mar 25, 2023
60	Jan 19, 2022–Feb 25, 2023	Feb 26, 2023–Apr 22, 2023
64	Jan 19, 2022–Mar 25, 2023	Mar 26, 2023–May 20, 2023
68	Jan 19, 2022–Apr 22, 2023	Apr 23, 2023–Jun 17, 2023
72	Jan 19, 2022–May 20, 2023	May 21, 2023–Jul 15, 2023
76	Jan 19, 2022–Jun 17, 2023	Jun 18, 2023–Aug 12, 2023
80	Jan 19, 2022–Jul 15, 2023	Jul 16, 2023–Sept 9, 2023

Table 4
MAPE and MAE for predicted infection cases versus notified confirmed cases across various training durations.

Train length (Weeks)	MAPE (%)	MAE
40	14.57	13 808.11
44	27.42	22 771.7
48	11.11	14 707.84
52	16.76	23 189.95
56	17.05	26 159.73
60	16.34	29 768.08
64	23.71	46 473.74
68	25.41	44 599.62
72	29.69	46 390.44
76	36.45	53 244.59
80	40.78	57 225.57

where x_i denotes the actual value of the i th data point, y_i represents the predicted value for the i th data point, and n is the total number of observations.

In this study, we employ GBM with a learning rate of 0.01, utilizing 1000 trees that assume a Gaussian distribution for the response variable. The model requires a minimum of 10 observations in the terminal nodes of the trees and utilizes 10-fold cross-validation. The default tree depth is set at 30. According to Table 4, training the GBM for 48 weeks results in relatively low MAPE and MAE values. As depicted in Fig. 11, the MAPE for the prediction result based on the GBM is 11.11%. Although the trained transmission rates align closely with those derived from the inverse method, the predicted transmission rates show some discrepancies. Our numerical analysis further reveals that the transmission rates modeled fit nearly perfectly with the actual notified confirmed data, and the predicted infection

cases also align well with the real data, indicating the effectiveness of the GBM in capturing the dynamics of norovirus transmission under the study conditions.

Relative influence indicates the relative importance of each variable in training the model (Friedman, 2001). Here, we compute the relative influence of temperature and relative humidity in the GBM. The relative influence of temperature and humidity when trained for 48 weeks from January 19 to December 3, 2022 is shown in Fig. 12. We can observe that temperature has the higher weight of relative influence, which is 55%. In this work, temperature is the more important factor in training the model.

6. Limitations

This paper acknowledges several limitations in its findings. Firstly, the analysis is limited to only two environmental factors: temperature and relative humidity. Other factors, such as rainfall, and sunlight may have an impact on the transmission of norovirus (Lopman et al., 2009). Extreme rainfall events may lead to flooding, hence promoting the outbreaks of norovirus.

Secondly, while our results numerically demonstrate that higher temperatures and humidity levels can mitigate the transmission of norovirus during warmer months, deriving an explicit formula for the transmission rate β as a function of temperature (T) and humidity (H) remains a complex and unresolved challenge within the scope of this analysis.

Thirdly, there are significant limitations concerning the data used. Since linear regression is a simple data analysis technique and commonly used, we chose it to obtain a preliminary model to explore the relationships among temperature, relative humidity, and the survival of norovirus on surfaces. Due to the lack of additional data, we are

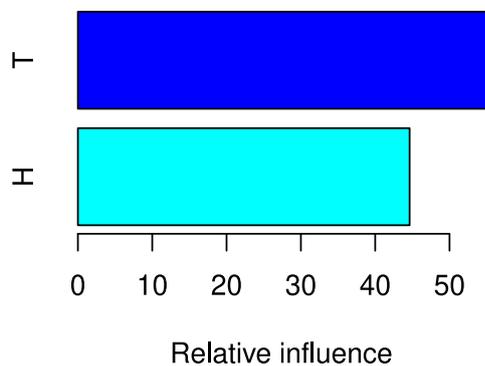


Fig. 12. Relative influence of temperature (T) and humidity (H) when trained for 48 weeks from January 19 to December 3, 2022.

currently unable to determine their nonlinear relationship. Many infected individuals may be unaware of their condition and consequently do not seek testing. We have assumed that only 1.5% of the population undergo testing each week, which likely results in an underestimation of the actual number of infections, potentially skewing the predictive accuracy of our models. The values of some parameters are poorly studied in the literature. Therefore, we based our assumptions on existing research results and logical reasoning.

Lastly, the epidemiology of norovirus is complicated by its diverse modes of transmission and high viral diversity. Our study focuses only on three primary routes of transmission, which may not fully encapsulate the complete dynamics of how norovirus spreads.

7. Discussion

In this study, we employed a hybrid modeling approach to understand the dynamics of norovirus transmission. We developed a foodborne disease model with an indirect incidence mechanism to investigate the link between norovirus outbreaks and environmental conditions. Utilizing regression techniques, we modeled the relationship between climatic factors – specifically temperature and relative humidity – and the survival rate of norovirus. We applied a linear regression model to estimate the survival rate of norovirus under various combinations of temperature and humidity conditions. We then calculated the mortality rates of norovirus in four distinct regions of the United States (Southern, Northeastern, Midwestern, and Western) by analyzing climate data from January 19, 2022, to December 31, 2023, for each region. By examining the weekly confirmed cases and corresponding weather data, we derived the time-varying transmission rates using the inverse method for each region independently. The transmission rates determined through the inverse method are closely aligned with the observed notification data. Our findings corroborate the notion that the transmission rate of norovirus diminishes in warmer months, while colder and dryer conditions favor the spread of the virus. These results highlight the significant impact of environmental factors on the transmission dynamics of norovirus.

To explore the relationship between the transmission rate of norovirus and environmental factors, we utilized a gradient-boosting machine (GBM) to estimate the transmission rates based on climate data from four regions in the USA. We also predicted the weekly number of infection cases using this model. The predictive performance of the GBMs was assessed using mean absolute error (MAE) and mean absolute percentage error (MAPE), both of which measure the accuracy of the transmission rate estimations and the alignment of predicted with confirmed case numbers. Our model demonstrated exceptional data fitting capabilities, accurately modeling the transmission dynamics of norovirus. Leveraging climate data, our model is capable of forecasting the transmission rate and subsequent infection cases for up to eight

weeks into the future. The accuracy and foresight provided by our model are invaluable for public health and healthcare facilities, particularly in preparing for the heightened burden of norovirus infections during the winter months.

There are more than 200 diseases caused by consuming contaminated foods or beverages by pathogenic bacteria, viruses, or parasites, and fungi. Norovirus, Salmonella (non-typhoidal), Clostridium perfringens, Campylobacter and Staphylococcus aureus are five major foodborne diseases, where the outbreak of norovirus, Salmonella, and Campylobacter are associated with the environmental conditions (Kovats et al., 2004; Tam et al., 2006). While Salmonella and Campylobacter are bacterial infections, it would be insightful to explore and compare the transmission dynamics of diseases caused by viruses and those caused by bacteria, highlighting their similarities and differences.

CRediT authorship contribution statement

Juping Ji: Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Shohel Ahmed:** Writing – review & editing, Data curation. **Hao Wang:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Funding

This project was supported by National Natural Science Foundation of China (No. 12401635), MfPH (Mathematics for Public Health) and OMNI (One Health Modelling Network for Emerging Infections) funded by Natural Sciences and Engineering Research Council of Canada - Emerging Infectious Diseases Modelling Initiative, as well as IUSEP funding (Department of Mathematical and Statistical Sciences at the University of Alberta).

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.

Data availability

The datasets used and/or analyzed during the current study are available publicly (Anon, 2023a,b,c,d).

References

- Adler, J.L., Zickl, R., 1969. Winter vomiting disease. *J. Infect. Dis.* 119, 668–673.
- Anon, 2023a. Centers for Disease Control and Prevention. Available from <https://www.cdc.gov/norovirus>.
- Anon, 2023b. Weather data. Available from <https://www.wunderground.com/history>.
- Anon, 2023c. Norovirus Regional Trends. Available from <https://www.cdc.gov/surveillance/nrevss/norovirus/region.html>.
- Anon, 2023d. United States Population Growth by Region. Available from https://www.census.gov/popclock/data_tables.php?component=growth.
- Atmar, R.L., Opekun, A.R., Gilger, M.A., et al., 2008. Norwalk virus shedding after experimental human infection. *Emerg. Infect. Dis.* 14, 1553–1557.
- Bitler, E.J., Matthews, J.E., et al., 2013. Norovirus outbreaks: a systematic review of commonly implicated transmission routes and vehicles. *Emerg. Infect. Dis.* 141 (8), 1563–1571.
- Campos, C.J., Lees, D.N., 2014. Environmental transmission of human noroviruses in shellfish waters. *Appl. Environ. Microbiol.* 80 (12), 3552–3561.
- CDC, 2011. Updated norovirus outbreak management and disease prevention guidelines. *MMWR Morb. Mortal. Wkly. Rep.* 60, 1–15.
- Chenar, S.S., Deng, Z., 2017a. Environmental indicators for human norovirus outbreaks. *Int. J. Environ. Health Res.* 27 (1), 40–51.
- Chenar, S.S., Deng, Z., 2017b. Environmental indicators of oyster norovirus outbreaks in coastal waters. *Mar. Environ. Res.* 130, 275–281.
- Chenar, S.S., Deng, Z., 2018. Development of artificial intelligence approach to forecasting oyster norovirus outbreaks along Gulf of Mexico coast. *Environ. Int.* 111, 212–223.

- Colas de la Noue, A., Estienney, M., Aho, S., et al., 2014. Absolute humidity influences the seasonal persistence and infectivity of human norovirus. *Appl. Environ. Microbiol.* 80, 7196–7205.
- Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Ann. Statist.* 29, 1189–1232.
- Hall, A.J., Wikswold, M.E., Pringle, K., et al., 2014. Vital signs: foodborne norovirus outbreaks—United States, 2009–2012. *MMWR Morb. Mortal. Wkly. Rep.* 63 (22), 491–495.
- Ji, J., Wang, H., Wang, L., Ramazi, P., Kong, J.D., Watmough, J., 2023. Climate-dependent effectiveness of nonpharmaceutical interventions on COVID-19 mitigation. *Math. Biosci.* 366, 109087.
- Khair, U., Fahmi, H., Hakim, S., Rahim, R., 2017. Forecasting error calculation with mean absolute deviation and mean absolute percentage error. *J. Phys. Conf. Ser.* 930, 012002.
- Kim, S.J., Si, J., Lee, J.E., Ko, G., 2012. Temperature and humidity influences on inactivation kinetics of enteric viruses on surfaces. *Environ. Sci. Technol.* 46, 13303–13310.
- Kong, J.D., Jin, C., Wang, H., 2015. The inverse method for a childhood infectious disease model with its application to pre-vaccination and post-vaccination measles data. *Bull. Math. Biol.* 77, 2231–2263.
- Kovats, R.S., Edwards, S.J., Hajat, S., et al., 2004. The effect of temperature on food poisoning: a time-series analysis of salmonellosis in ten European countries. *Emerg. Infect. Dis.* 10, 443–453.
- Lane, D., Husemann, E., Holl, D., Khaled, A., 2019. Understanding foodborne transmission mechanisms for Norovirus: A study for the UK's Food Standards Agency. *European J. Oper. Res.* 275, 721–736.
- Lopman, B., Armstrong, B., Atchison, C., Gray, J.J., 2009. Host, weather and virological factors drive norovirus epidemiology: time-series analysis of laboratory surveillance data in England and Wales. *PLoS One* 4 (8), e6671.
- Mathijs, E., Stals, A., et al., 2012. A review of known and hypothetical transmission routes for noroviruses. *Food Environ. Virol.* 4, 131–152.
- Matsuyama, R., Miura, F., Nishiura, H., 2017. The transmissibility of noroviruses: Statistical modeling of outbreak events with known route of transmission in Japan. *PLoS One* 12 (3), e0173996.
- Morioka, S., Sakata, T., et al., 2006. A food-borne norovirus outbreak at a primary school in Wakayama prefecture. *Jpn. J. Infect. Dis.* 59 (3), 205–207.
- Mounds, A.W., Ando, T., Koopmans, M., et al., 2000. Cold weather seasonality of gastroenteritis associated with Norwalk-like viruses. *J. Infect. Dis.* 181, 284–287.
- Natekin, A., Knoll, A., 2013. Gradient boosting machines, a tutorial. *Front. Neuroinformatics* 7.
- Ozawa, K., Oka, T., Takeda, N., Hansman, G.S., 2007. Norovirus infections in symptomatic and asymptomatic food handlers in Japan. *J. Clin. Microbiol.* 45 (12), 3996–4005.
- Pollicott, M., Wang, H., Weiss, H.H., 2012. Extracting the time-dependent transmission rate from infection data via solution of an inverse ODE problem. *J. Biol. Dyn.* 6, 509–523.
- Prato, R., Lopalco, P.L., et al., 2004. Norovirus gastroenteritis general outbreak associated with raw shellfish consumption in South Italy. *BMC Infect. Dis.* 4, 37.
- Rohayem, J., 2009. Norovirus seasonality and the potential impact of climate change. *Clin. Microbiol. Infect.* 15 (6), 524–527.
- Rushton, S.P., Sanderson, R.A., et al., 2019. Transmission routes of rare seasonal diseases: the case of norovirus infections. *Phil. Trans. R. Soc. B* 374, 20180267.
- S. Xiao, S., Tang, J.W., Li, Y., 2017. Airborne or fomite transmission for norovirus? A case study revisited. *Int. J. Environ. Res. Public Health* 14 (12), 1571.
- Scallan, E., Hoekstra, R.M., et al., 2011. Foodborne illness acquired in the United States—major pathogens. *Emerg. Infect. Diseases* 17 (1), 7–15.
- Siebenga, J.J., Vennema, H., et al., 2009. Norovirus illness is a global problem: Emergence and spread of norovirus GII.4 variants, 2001–2007. *J. Infect. Dis.* 200 (5), 802–812.
- Smith, A.J., McCarthy, N., et al., 2012. A large foodborne outbreak of norovirus in diners at a restaurant in England between January and 2009. *Emerg. Infect. Dis.* 18, 1695–1701.
- Sugieda, M., Nakajima, K., Nakajima, S., 1996. Outbreaks of norwalk-like virus-associated gastroenteritis traced to shellfish: coexistence of two genotypes in one specimen. *Emerg. Infect. Dis.* 116, 339–346.
- Tam, C.C., Rodrigues, L.C., O'Brien, S.J., Hajat, S., 2006. Temperature dependence of reported *Campylobacter* infection in England, 1989–1999. *Emerg. Infect. Dis.* 12, 119–125.
- Ushijima, H., Fujimoto, T., et al., 2014. Norovirus and foodborne disease: a review. *Food Saf.* 2 (3), 37–54.
- Verhoef, L., Depoortere, E., Boxman, I., et al., 2008. Emergence of new norovirus variants on spring cruise ships and prediction of winter epidemics. *Emerg. Infect. Dis.* 14, 238–243.
- Wang, X., Wang, H., 2023. Discrete inverse method for extracting disease transmission rates from accessible infection data. *SIAM J. Appl. Math. Special Section Preview*.
- Wang, X., Wang, H., Ramazi, P., Nah, K., Lewis, M., 2022a. A hypothesis-free bridging of disease dynamics and non-pharmaceutical policies. *Bull. Math. Biol.* 84 (5), 57.
- Wang, X., Wang, H., Ramazi, P., Nah, K., Lewis, M., 2022b. From policy to prediction: Forecasting COVID-19 dynamics under imperfect vaccination. *Bull. Math. Biol.* 84 (9), 90.
- Westrell, T., Dusch, V., et al., 2010. Norovirus outbreaks linked to oyster consumption in the United Kingdom, Norway, France, Sweden and Denmark, 2010. *Eurosurveillance* 15 (12), 19524.
- Widdowson, M.A., Sulka, A., et al., 2005. Norovirus and foodborne disease, United States, 1991–2000. *Emerg. Infect. Dis.* 11 (1), 95.
- Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* 30, 79–82.