

Article

The Structural Identifiability of a Humidity-Driven Epidemiological Model of Influenza Transmission

Chunyang Zhang ^{1,2,†}, Xiao Zhang ^{3,†}, Yuan Bai ^{2,3,*}, Eric H. Y. Lau ^{2,3}  and Sen Pei ⁴¹ School of Mathematics, Changchun Normal University, Changchun 130032, China² WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China³ Laboratory of Data Discovery for Health Limited, Hong Kong Science and Technology Park, Hong Kong, China⁴ Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY 10032, USA

* Correspondence: yb424@hku.hk

† These authors contributed equally to this work.

Abstract: Influenza epidemics cause considerable morbidity and mortality every year worldwide. Climate-driven epidemiological models are mainstream tools to understand seasonal transmission dynamics and predict future trends of influenza activity, especially in temperate regions. Testing the structural identifiability of these models is a fundamental prerequisite for the model to be applied in practice, by assessing whether the unknown model parameters can be uniquely determined from epidemic data. In this study, we applied a scaling method to analyse the structural identifiability of four types of commonly used humidity-driven epidemiological models. Specifically, we investigated whether the key epidemiological parameters (i.e., infectious period, the average duration of immunity, the average latency period, and the maximum and minimum daily basic reproductive number) can be uniquely determined simultaneously when prevalence data is observable. We found that each model is identifiable when the prevalence of infection is observable. The structural identifiability of these models will lay the foundation for testing practical identifiability in the future using synthetic prevalence data when considering observation noise. In practice, epidemiological models should be examined with caution before using them to estimate model parameters from epidemic data.

Keywords: influenza; structural identifiability; climate-driven epidemiological model; scaling method



Citation: Zhang, C.; Zhang, X.; Bai, Y.; Lau, E.H.Y.; Pei, S. The Structural Identifiability of a Humidity-Driven Epidemiological Model of Influenza Transmission. *Viruses* **2022**, *14*, 2795. <https://doi.org/10.3390/v14122795>

Academic Editor: Ruian Ke

Received: 28 October 2022

Accepted: 10 December 2022

Published: 15 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Influenza epidemics cause considerable morbidity and mortality every year worldwide [1–6]. According to The World Health Organization, influenza leads to about 3 to 5 million cases of severe illness and about 290,000 to 650,000 deaths annually [7]. This burden might be alleviated by understanding historical and current transmission dynamics and predicting further transmission trends of influenza to assist public health authorities in designing effective interventions and vaccination strategies, especially in some special situations (for example, the emergence of influenza A (H1N1) viruses and the potential rebound of influenza in the post-COVID-19 pandemic period) [8–12].

Climate-driven epidemiological models, such as models driven by humidity, are mainstream tools for understanding seasonal transmission dynamics and predicting future trends in influenza activity. Recently, climate-driven epidemiological models have been successfully applied to recreate historical activity time series of influenza and to forecast the week with the highest influenza activity in temperate, tropical and subtropical regions (for example, in the United States and Hong Kong) [13–17]. Confident predictions using these models depend on testing various numerical optimization algorithms, such as filter-based data-assimilation algorithms, which fit the model to epidemic data when parameterizing

models to simulate the transmission dynamics of influenza [13]. However, the model structure identifiability needs to be tested to avoid the optimization algorithm falling into a set of the locally optimal solution.

Model identifiability includes structural and practical identifiability, involving investigation of whether unknown model parameters can be uniquely determined from noise-free epidemic data and accurately identified from noisy data, respectively [18–24]. The structural identifiability of the model is a fundamental prerequisite for practical identifiability and for the model to be used in practice [25]. It is necessary, but insufficient, to accurately identify model parameters from actual noisy data because a model that is structurally identifiable might be unidentifiable when noisy data are used. The scaling method, which is a structural identifiability analysis method that has been proposed in recent years, is based on the scale invariance of the equations [26]. Compared with existing structural identifiability methods (such as differential algebra), this method has the advantage of simple operation (no advanced computing skills are required) and low computational cost, particularly when analyzing high-dimensional non-linear models [27–30]. This method has been used to analyze the structural identifiability of mathematical modeling describing biological processes, such as the generalized mass-action model [31–33].

In this study, we apply the scaling method to analyse the structural identifiability of several types of commonly used humidity-driven epidemiological models. We investigate whether the key epidemiological parameters (infectious period, the average duration of immunity, the average latency period, the maximum and minimum daily basic reproductive number) can be determined simultaneously when the population prevalence of infected people is observable.

2. Methods

Here, we briefly introduce the process from building an epidemiological model to applying the model in real-world applications. After building an epidemiological model of influenza transmission, we test the structural identifiability of the model to investigate the properties of the model itself (the upper part of Figure 1). If the model is structurally identifiable, then we test the practical identifiability of the model, which determines whether the model is identifiable for noise data (e.g., reported noise), using synthetic data experiments. If the model is also practically identifiable, then this model can potentially be used in practice after evaluating its performance for specific functions (e.g., inference, forecasting, etc.). On the other hand, if a model is structurally unidentifiable, any parameter estimated by optimization algorithms might be unreliable; then, we need to consider modifying the model. For some complex models, for example, agent-based influenza transmission models [34], we can only test the practical identifiability directly after building the model (the lower part of Figure 1), as current mathematical methods may not be able to theoretically test the structural identifiability of these complex models. Here, we mainly focus on testing the structural identifiability of humidity-driven epidemiological models.

2.1. Humidity-Driven Epidemiological Model

We test the structural identifiability of several commonly used humidity-driven epidemiological models. The form of each model is as follows.

1. SIS model:

$$\begin{aligned} \frac{dS}{dt} &= \frac{I}{D} - \frac{\beta(t)IS}{N} - \alpha, \\ \frac{dI}{dt} &= \frac{\beta(t)IS}{N} - \frac{I}{D} + \alpha. \end{aligned} \quad (1)$$

2. SIRS model:

$$\begin{aligned}\frac{dS}{dt} &= \frac{N - S - I}{L} - \frac{\beta(t)IS}{N} - \alpha, \\ \frac{dI}{dt} &= \frac{\beta(t)IS}{N} - \frac{I}{D} + \alpha.\end{aligned}\quad (2)$$

3. SEIR model:

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta(t)IS}{N} - \alpha, \\ \frac{dE}{dt} &= \frac{\beta(t)IS}{N} - \frac{E}{W} + \alpha, \\ \frac{dI}{dt} &= \frac{E}{W} - \frac{I}{D}.\end{aligned}\quad (3)$$

4. SEIRS model:

$$\begin{aligned}\frac{dS}{dt} &= \frac{N - S - E - I}{L} - \frac{\beta(t)IS}{N} - \alpha, \\ \frac{dE}{dt} &= \frac{\beta(t)IS}{N} - \frac{E}{W} + \alpha, \\ \frac{dI}{dt} &= \frac{E}{W} - \frac{I}{D}.\end{aligned}\quad (4)$$

where N , S , E , and I represent the total number of people, the number of susceptible people, the number of exposed people, and the number of infectious individuals, respectively. $N = S + I$ in the SIS model, $N = S + I + R$ in the SIRS model, and $N = S + E + I + R$ in the SEIR model and the SEIRS model. α represents the rate at which influenza viruses are imported into the model due to travel. t represents time, such as the day, week or year. $\beta(t)$ represents the transmission rate at time t . D represents the mean infectious period in all four models. L represents the average duration of immunity in the SIRS model and the SEIRS model. W represents the average latency period in the SEIR model and the SEIRS model. The flow diagrams of these models are presented in Figure A1 in Appendix A.

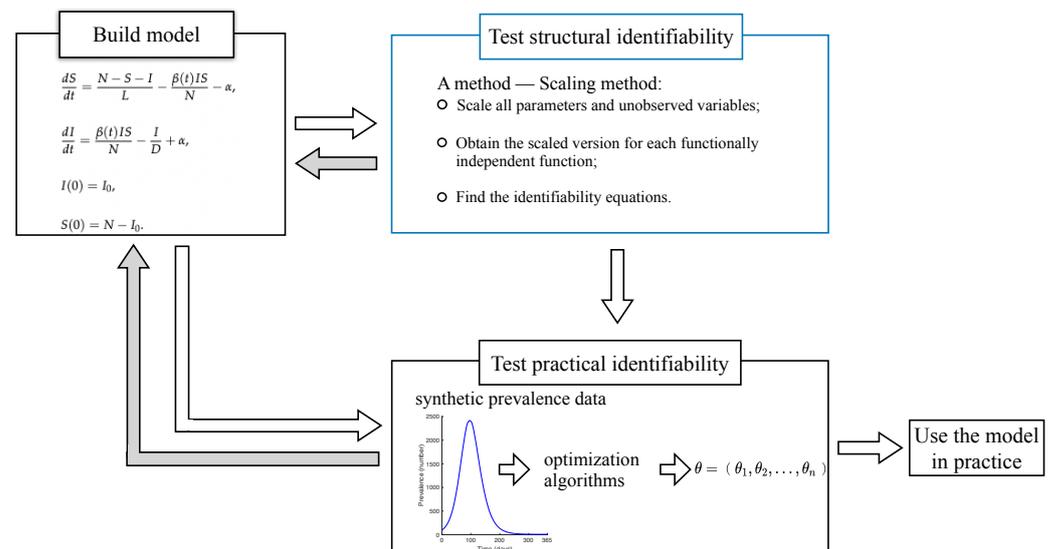


Figure 1. The process from building the model to putting it into use. The grey arrows show the feedback from testing the model identifiability results to modifying the model. The blue box illustrates the process of testing the model structural identifiability that our work focuses on.

The basic reproductive number, $R_0(t)$, represents the average number of secondary infections generated by a primary case in a fully susceptible population at time t , which is proportional to the transmission rate. The expression is as follows:

$$R_0(t) = \beta(t)D. \quad (5)$$

The influenza virus survival and transmission are relative to the absolute humidity (AH) shown from laboratory experiments [35]. The specific humidity (SH) is a measure of AH, in which $q(t)$ represents SH at time t . In this model, the humidity factor modulates $R_0(t)$ through an exponential relationship:

$$R_0(t) = e^{(a \cdot q(t) + b)} + R_{0\min}, \quad (6)$$

where $a = -180$ is estimated by fitting laboratory influenza virus survival to the value of AH using a regression model. $b = \log(R_{0\max} - R_{0\min})$, $R_{0\max}$ and $R_{0\min}$ are the maximum and minimum daily basic reproductive number, respectively. Parameter sets $\Theta_1 = \{D, R_{0\max}, R_{0\min}\}$ in the SIS model, $\Theta_2 = \{D, L, R_{0\max}, R_{0\min}\}$ in the SIRS model, $\Theta_3 = \{D, W, R_{0\max}, R_{0\min}\}$ in the SEIR model and $\Theta_4 = \{D, L, W, R_{0\max}, R_{0\min}\}$ in the SEIRS model may be estimated by fitting the model to the epidemic data. We analyse the structural identifiability of these models, which tests whether each parameter set can be uniquely determined simultaneously when the prevalence data is observable (using the observation record from the beginning to the end time, such as one influenza season in temperate regions).

2.2. The Frameworks for the Scaling Method

The scaling method is easy to use for identifying the structural identifiability of a non-linear model based on simple scaling transformations and the solution of simple sparse systems of equations [26]. The ordinary differential equations (ODE) model, which is applied to the frameworks of the scaling method, is as follows:

$$\begin{aligned} \frac{dx_i}{dt} &= f_i(x_1, \dots, x_s, x_{s+1}, \dots, x_n; \theta_1, \dots, \theta_m), \\ x_i(0) &= x_{i,0}, \quad i = 1, \dots, n, \end{aligned} \quad (7)$$

where $x_i(0)$ represents the initial conditions, $\frac{dx_i}{dt}$ represents the change of x_i over time, depending on m parameters θ_j , and the number of state variables is n . f_i is a function characterising the specific details of the change rate of x_i . The simplicity of this method depends on the ability to decompose functions f_i as a sum of P functional independent components, f_{ij} ,

$$\begin{aligned} \frac{dx_i}{dt} &= f_i(x_1, \dots, x_s, x_{s+1}, \dots, x_n; \theta_1, \dots, \theta_m) \\ &= \sum_{j=1}^P f_{ij}(\bar{x}_j, \bar{\theta}_j). \end{aligned} \quad (8)$$

A property of f_{ij} is that f_{ij} is functionally independent of f_{ik} for every $j \neq k$. Here, \bar{x}_j and $\bar{\theta}_j$ represent the subset of variables and parameters of function f_{ij} , respectively. In simple terms, if $f_i(x_1, x_2, \dots), \dots, f_n(x_1, x_2, \dots)$ are linearly independent functions, then the only solution of the equation

$$\sum_{i=1}^n a_i f_i(x_1, x_2, \dots) = 0 \quad (9)$$

is $a_1 = a_2 = \dots = a_n = 0$. The functional independence theorems used in this work are presented in Appendix A.

Based on decomposing functions, the steps of the scaling method are as follows:

Step 1. Scale all parameters and unobserved variables using unknown scaling factors, μ :

$$\begin{aligned} \theta_i &\rightarrow \mu_{\theta_i} \theta_i \quad i = 1, \dots, m, \\ x_j &\rightarrow \mu_{x_j} x_j \quad j = s + 1, \dots, n, \end{aligned} \tag{10}$$

and substitute them into Equation (8). The experiment measures variables x_1, \dots, x_s without modifying them (n is the total number of state variables, x_1, \dots, x_s is the observable state variable and x_{s+1}, \dots, x_n is the unobserved state variable).

Step 2. Obtain the scaled version for each functionally independent function. Namely,

$$f_{ij}(\bar{x}_j, \bar{\theta}_j) = f_{ij}(x_j, \mu_{\bar{\theta}_j} \bar{\theta}_j) \quad i = 1, \dots, s \tag{11}$$

and

$$f_{ij}(\bar{x}_j, \bar{\theta}_j) = \frac{1}{\mu_{x_i}} f_{ij}(\mu_{\bar{x}_j} \bar{x}_j, \mu_{\bar{\theta}_j} \bar{\theta}_j) \quad i = s + 1, \dots, n. \tag{12}$$

Step 3. Find scaling factor combinations that maintain the system invariant. Only the parameters $\bar{\theta}_j$ with a solution $\mu_{\bar{\theta}_j} = 1$ are identifiable. Only the variables, \bar{x}_j with $\mu_{\bar{x}_j} = 1$ are observable. Otherwise, parameters whose scaling factors are coupled form identifiable groups but cannot be identified independently.

3. Results

Here, we demonstrate how to use the simple scaling method to test the identifiability of the four humidity-driven epidemiological models introduced in the Methods section.

We consider a scenario where only I is observed, representing a kind of epidemic data that can be collected in practice. For example, in the UK, the COVID-19 Infection Survey identified those people testing positive for coronavirus (COVID-19) in private residential households (surveillance sensors) at a point in time to help the government make decisions on how to respond to the emerging epidemic and provide information to the public [36]. This infection survey can, in principle, be extended to survey influenza to identify new positive cases of influenza regularly around the influenza season. In this scenario, we can obtain I from the sentinel surveillance systems. We test whether the humidity-driven epidemiological models in Equations (1)–(4) are structurally identifiable, respectively.

3.1. SIS Model

For the SIS model, we test whether parameter set $\Theta_1 = \{D, R_{0\max}, R_{0\min}\}$ can be determined uniquely from the observable I . First, we investigate whether the differential equation in Equation (1) can be decomposed into a sum of linearly independent functions. This is a prerequisite for using the scaling method. For the differential equation associated with S , we have:

$$f_{S1} = \frac{I}{D}, \quad f_{S2} = -\frac{\beta(t)IS}{N}. \tag{13}$$

According to Theorem A1, the generalized Wronskian determinant is as follows:

$$W_S = \begin{vmatrix} \Delta_0 f_{S1} & \Delta_0 f_{S2} \\ \Delta_1 f_{S1} & \Delta_1 f_{S2} \end{vmatrix} = \begin{vmatrix} \frac{I}{D} & -\frac{\beta(t)IS}{N} \\ 1 & -\frac{\beta(t)(S+I)}{N} \end{vmatrix} = -\frac{\beta(t)I^2}{DN} \neq 0. \tag{14}$$

So, f_{S1} and f_{S2} are linearly independent functions. Similarly, for the differential equation associated with I , we have:

$$f_{I1} = \frac{\beta(t)IS}{N}, \quad f_{I2} = -\frac{I}{D}. \tag{15}$$

The corresponding generalized Wronskian determinant is as follows:

$$W_I = \begin{vmatrix} \Delta_0 f_{I1} & \Delta_0 f_{I2} \\ \Delta_1 f_{I1} & \Delta_1 f_{I2} \end{vmatrix} = \begin{vmatrix} \frac{\beta(t)IS}{N} & -\frac{I}{D} \\ \frac{\beta(t)(S+I)}{N} & -\frac{1}{D} \end{vmatrix} = \frac{\beta(t)I^2}{DN} \neq 0. \quad (16)$$

So, f_{I1} and f_{I2} are linearly independent functions. Next, we explore whether parameter set $\Theta_1 = \{D, R_{0\max}, R_{0\min}\}$ can uniquely be determined from the observable I using the scaling method. The steps of the scaling method are as follows:

Step 1. We scale the parameter set $\Theta_1 = \{D, R_{0\max}, R_{0\min}\}$ and the unobserved variable (S) by unknown scaling factors:

$$\begin{cases} D \rightarrow \mu_D D, \\ R_{0\max} \rightarrow \mu_{R_{0\max}} R_{0\max}, \\ R_{0\min} \rightarrow \mu_{R_{0\min}} R_{0\min}, \\ S \rightarrow \mu_S S. \end{cases} \quad (17)$$

Step 2. We obtain the scaled version for each functional linear independent function in Equations (13) and (15).

Step 3. We obtain the identifiability equations:

$$\frac{I}{\mu_S \mu_D D} = \frac{I}{D}, \quad (18)$$

$$\begin{aligned} \frac{I \mu_S S}{\mu_S N \mu_D D} & \left[e^{aq(t)} (\mu_{R_{0\max}} R_{0\max} - \mu_{R_{0\min}} R_{0\min}) + \mu_{R_{0\min}} R_{0\min} \right] \\ & = \frac{IS}{ND} \left[e^{aq(t)} (R_{0\max} - R_{0\min}) + R_{0\min} \right], \end{aligned} \quad (19)$$

$$\begin{aligned} \frac{I \mu_S S}{N \mu_D D} & \left[e^{aq(t)} (\mu_{R_{0\max}} R_{0\max} - \mu_{R_{0\min}} R_{0\min}) + \mu_{R_{0\min}} R_{0\min} \right] \\ & = \frac{IS}{ND} \left[e^{aq(t)} (R_{0\max} - R_{0\min}) + R_{0\min} \right], \end{aligned} \quad (20)$$

$$\frac{I}{\mu_D D} = \frac{I}{D}. \quad (21)$$

Manipulating the above formulas, the identifiability equations are:

$$\begin{cases} \mu_D = 1, \\ \mu_S = 1, \\ e^{aq(t)} (\mu_{R_{0\max}} R_{0\max} - \mu_{R_{0\min}} R_{0\min}) + \mu_{R_{0\min}} R_{0\min} \\ = e^{aq(t)} (R_{0\max} - R_{0\min}) + R_{0\min}. \end{cases} \quad (22)$$

From the last formula in Equation (22), we further manipulate this formula and obtain the following equation:

$$e^{aq(t)} [(\mu_{R_{0\max}} - 1)R_{0\max} - (\mu_{R_{0\min}} - 1)R_{0\min}] = (1 - \mu_{R_{0\min}})R_{0\min}. \quad (23)$$

$e^{aq(t)}$ is not equal to zero. When the left and right sides of the equation are equal, we have:

$$\begin{cases} (\mu_{R_{0\max}} - 1)R_{0\max} - (\mu_{R_{0\min}} - 1)R_{0\min} = 0, \\ (1 - \mu_{R_{0\min}})R_{0\min} = 0, R_{0\max} > 0, R_{0\min} > 0. \end{cases} \quad (24)$$

By solving the above equation, we can obtain:

$$\begin{cases} \mu_{R_{0\max}} = 1, \\ \mu_{R_{0\min}} = 1. \end{cases} \quad (25)$$

Therefore, the SIS model is identifiable. Namely, parameter set $\Theta_1 = \{D, R_{0\max}, R_{0\min}\}$ can be determined uniquely from the observable I .

3.2. SIRS Model

For the SIRS model, we test whether the parameter set $\Theta_2 = \{D, L, R_{0\max}, R_{0\min}\}$ can be determined uniquely from the observable I . First, we investigate whether the differential equation in Equation (2) can be decomposed into a sum of linearly independent functions. For the differential equation associated with S , we have: For the differential equation associated with S , we have:

$$f_{S1} = \frac{N - S - I}{L}, f_{S2} = -\frac{\beta(t)IS}{N}. \quad (26)$$

According to Theorem A1, the generalized Wronskian determinant is as follows:

$$W_S = \begin{vmatrix} \Delta_0 f_{S1} & \Delta_0 f_{S2} \\ \Delta_1 f_{S1} & \Delta_1 f_{S2} \end{vmatrix} = \begin{vmatrix} \frac{N-S-I}{L} & -\frac{\beta(t)IS}{N} \\ -\frac{2}{L} & -\frac{\beta(t)S}{N} - \frac{\beta(t)I}{N} \end{vmatrix} = \frac{\beta(t)[I(I-N) + S(S-N)]}{LN} \neq 0. \quad (27)$$

So, f_{S1} and f_{S2} are linearly independent functions. Similarly, for the differential equation associated with I , we have:

$$f_{I1} = \frac{\beta(t)IS}{N}, \quad f_{I2} = -\frac{I}{D}. \quad (28)$$

The corresponding generalized Wronskian determinant is as follows:

$$W_I = \begin{vmatrix} \Delta_0 f_{I1} & \Delta_0 f_{I2} \\ \Delta_1 f_{I1} & \Delta_1 f_{I2} \end{vmatrix} = \begin{vmatrix} \frac{\beta(t)IS}{N} & -\frac{I}{D} \\ \frac{\beta(t)(S+I)}{N} & -\frac{1}{D} \end{vmatrix} = \frac{\beta(t)I^2}{DN} \neq 0. \quad (29)$$

So, f_{I1} and f_{I2} are linearly independent functions. Next, we explore whether the parameter set $\Theta_2 = \{D, L, R_{0\max}, R_{0\min}\}$ can be determined uniquely from the observable I using the scaling method. The steps of the scaling method are as follows:

Step 1. We scale the parameters and unobserved variables by unknown scaling factors:

$$\begin{cases} D \rightarrow \mu_D D, \\ L \rightarrow \mu_L L, \\ R_{0\max} \rightarrow \mu_{R_{0\max}} R_{0\max}, \\ R_{0\min} \rightarrow \mu_{R_{0\min}} R_{0\min}, \\ S \rightarrow \mu_S S. \end{cases} \quad (30)$$

Step 2. We obtain the scaled version for each functional linear independent function in Equations (26) and (28).

Step 3. We obtain the identifiability equations:

$$\frac{N - \mu_S S - I}{\mu_S \mu_L L} = \frac{N - S - I}{L}, \quad (31)$$

$$\begin{aligned} & \frac{I \mu_S S}{\mu_S \mu_D D N} [e^{aq(t)} (\mu_{R_{0\max}} R_{0\max} - \mu_{R_{0\min}} R_{0\min}) + \mu_{R_{0\min}} R_{0\min}] \\ & = \frac{IS}{DN} [e^{aq(t)} (R_{0\max} - R_{0\min}) + R_{0\min}], \end{aligned} \quad (32)$$

$$\begin{aligned} & \frac{I\mu_S S}{\mu_D DN} [e^{aq(t)}(\mu_{R_{0max}} R_{0max} - \mu_{R_{0min}} R_{0min}) + \mu_{R_{0min}} R_{0min}] \\ & = \frac{IS}{DN} [e^{aq(t)}(R_{0max} - R_{0min}) + R_{0min}], \end{aligned} \tag{33}$$

$$\frac{I}{\mu_D D} = \frac{I}{D}. \tag{34}$$

Manipulating Equations (31)–(34), the identifiability equations are:

$$\left\{ \begin{aligned} & \mu_D = 1, \\ & \mu_S = 1, \\ & \mu_L = 1, \\ & (\mu_{R_{0max}} - 1)R_{0max} = (\mu_{R_{0min}} - 1)(1 - \frac{1}{e^{aq(t)}})R_{0min} \end{aligned} \right. \tag{35}$$

From the last formula in Equation (35), the left side of Equation (35) is constant, the right side of this equation is the function of t , and $e^{aq(t)}$ is not equal to zero. For the equation to be satisfied, $\mu_{R_{0max}} - 1 = 0$ and $\mu_{R_{0min}} - 1 = 0$. Hence, the system has a unique solution ($\mu_{R_{0max}} = \mu_{R_{0min}} = 1$). It follows that the SIRS model is identifiable.

3.3. SEIR Model

For the SEIR model, we test whether the parameter set $\Theta_3 = \{D, W, R_{0max}, R_{0min}\}$ can be determined uniquely from the observable I . First, we investigate whether the differential equation in Equation (3) can be decomposed into a sum of linearly independent functions. For the differential equation associated with S , we have:

$$f_{S1} = -\frac{\beta(t)IS}{N}. \tag{36}$$

According to Theorem A1, the generalized Wronskian determinant is as follows:

$$W_S = |\Delta_0 f_{S1}| = -\frac{\beta(t)IS}{N} \neq 0. \tag{37}$$

So, f_{S1} is a linearly independent function. Similarly, for the differential equation associated with E , we have:

$$f_{E1} = \frac{\beta(t)IS}{N}, \quad f_{E2} = -\frac{E}{W}. \tag{38}$$

The corresponding generalized Wronskian determinant is as follows:

$$W_E = \begin{vmatrix} \Delta_0 f_{E1} & \Delta_0 f_{E2} \\ \Delta_1 f_{E1} & \Delta_1 f_{E2} \end{vmatrix} = \begin{vmatrix} \frac{\beta(t)IS}{N} & -\frac{E}{W} \\ \frac{\beta(t)(S+I)}{N} & -\frac{1}{W} \end{vmatrix} = \frac{\beta(t)}{NW} [E(S+I) - IS] \neq 0. \tag{39}$$

So, f_{E1} and f_{E2} are linearly independent functions. For the differential equation associated with I , we have:

$$f_{I1} = \frac{E}{W}, \quad f_{I2} = -\frac{I}{D}. \tag{40}$$

The generalized Wronskian determinant is as follows:

$$W_I = \begin{vmatrix} \Delta_0 f_{I1} & \Delta_0 f_{I2} \\ \Delta_1 f_{I1} & \Delta_1 f_{I2} \end{vmatrix} = \begin{vmatrix} \frac{E}{W} & -\frac{I}{D} \\ \frac{1}{W} & -\frac{1}{D} \end{vmatrix} = \frac{I-E}{WD} \neq 0. \tag{41}$$

So, f_{I1} and f_{I2} are linearly independent functions. Next, we explore whether parameter set $\Theta_3 = \{D, W, R_{0max}, R_{0min}\}$ can be determined uniquely from observable I using the scaling method. The steps of the scaling method are as follows:

Step 1. We scale the parameter set $\Theta_3 = \{D, W, R_{0max}, R_{0min}\}$ and the unobserved variable (S) by unknown scaling factors:

$$\begin{cases} W \rightarrow \mu_W W, \\ D \rightarrow \mu_D D, \\ R_{0max} \rightarrow \mu_{R_{0max}} R_{0max}, \\ R_{0min} \rightarrow \mu_{R_{0min}} R_{0min}, \\ S \rightarrow \mu_S S, \\ E \rightarrow \mu_E E. \end{cases} \tag{42}$$

Step 2. We obtain the scaled version for each functional linear independent function in Equations (36), (38) and (40).

Step 3. We obtain the identifiability equations:

$$\begin{aligned} \frac{I\mu_S S}{\mu_S N \mu_D D} [e^{aq(t)} (\mu_{R_{0max}} R_{0max} - \mu_{R_{0min}} R_{0min}) + \mu_{R_{0min}} R_{0min}] \\ = \frac{IS}{ND} [e^{aq(t)} (R_{0max} - R_{0min}) + R_{0min}], \end{aligned} \tag{43}$$

$$\begin{aligned} \frac{I\mu_S S}{\mu_E N \mu_D D} [e^{aq(t)} (\mu_{R_{0max}} R_{0max} - \mu_{R_{0min}} R_{0min}) + \mu_{R_{0min}} R_{0min}] \\ = \frac{IS}{ND} [e^{aq(t)} (R_{0max} - R_{0min}) + R_{0min}], \end{aligned} \tag{44}$$

$$\frac{\mu_E E}{\mu_E \mu_W W} = \frac{E}{W}, \tag{45}$$

$$\frac{\mu_E E}{\mu_W W} = \frac{E}{W}, \tag{46}$$

$$\frac{I}{\mu_D D} = \frac{I}{D}. \tag{47}$$

According to formulas (45)–(47), it is easy to get:

$$\begin{cases} \mu_W = 1, \\ \mu_D = 1, \\ \mu_E = 1. \end{cases} \tag{48}$$

Pressing $\mu_D = 1$ into the formula (43), we can use Equation (23) in the SIS model. Then by Equation (24) in the SIS model, we can obtain:

$$\begin{cases} \mu_{R_{0max}} = 1, \\ \mu_{R_{0min}} = 1. \end{cases} \tag{49}$$

Next, pressing the above results into formula (44), we can get $\mu_S = 1$. Therefore, the SEIR model is identifiable. Namely, the parameter set $\Theta_3 = \{D, W, R_{0max}, R_{0min}\}$ can be uniquely determined from the observable I .

3.4. SEIRS Model

For the SEIRS model, we test whether the parameter set $\Theta_4 = \{D, L, W, R_{0max}, R_{0min}\}$ can be determined uniquely from observable I . First, we investigate whether the differential equation in Equation (4) can be decomposed into a sum of linearly independent functions. For the differential equation associated with S , we have:

$$f_{S1} = \frac{N - S - E - I}{L}, \quad f_{S2} = -\frac{\beta(t)IS}{N}. \tag{50}$$

According to Theorem A1, the generalized Wronskian determinant is as follows:

$$W_S = \begin{vmatrix} \Delta_0 f_{S1} & \Delta_0 f_{S2} \\ \Delta_1 f_{S1} & \Delta_1 f_{S2} \end{vmatrix} = \begin{vmatrix} \frac{N-S-E-I}{L} & -\frac{\beta(t)IS}{N} \\ \frac{1}{L} & -\frac{\beta(t)(S+I)}{N} \end{vmatrix} \quad (51)$$

$$= \frac{\beta(t)}{NL} [IS - (N - S - E - I)(S + I)] \neq 0. \quad (52)$$

So, f_{S1} and f_{S2} are linearly independent functions. Similarly, for the differential equation associated with E , we have:

$$f_{E1} = \frac{\beta(t)IS}{N}, \quad f_{E2} = -\frac{E}{W}. \quad (53)$$

The corresponding generalized Wronskian determinant is as follows:

$$W_E = \begin{vmatrix} \Delta_0 f_{E1} & \Delta_0 f_{E2} \\ \Delta_1 f_{E1} & \Delta_1 f_{E2} \end{vmatrix} = \begin{vmatrix} \frac{\beta(t)IS}{N} & -\frac{E}{W} \\ \frac{\beta(t)(S+I)}{N} & -\frac{1}{W} \end{vmatrix} = \frac{\beta(t)}{NW} [E(S + I) - IS] \neq 0. \quad (54)$$

So, f_{E1} and f_{E2} are linearly independent functions. For the differential equation associated with I , we have:

$$f_{I1} = \frac{E}{W}, \quad f_{I2} = -\frac{I}{D}. \quad (55)$$

The generalized Wronskian determinant is as follows:

$$W_I = \begin{vmatrix} \Delta_0 f_{I1} & \Delta_0 f_{I2} \\ \Delta_1 f_{I1} & \Delta_1 f_{I2} \end{vmatrix} = \begin{vmatrix} \frac{E}{W} & -\frac{I}{D} \\ \frac{1}{W} & -\frac{1}{D} \end{vmatrix} = \frac{I - E}{WD} \neq 0. \quad (56)$$

So, f_{I1} and f_{I2} are linearly independent functions. Next, we explore whether the parameter set $\Theta_4 = \{D, L, W, R_{0\max}, R_{0\min}\}$ can be determined uniquely from the observable I using the scaling method. The steps of the scaling method are as follows:

Step 1. We scale the parameter set $\Theta_4 = \{D, L, W, R_{0\max}, R_{0\min}\}$ and the unobserved variable (S) by unknown scaling factors:

$$\left\{ \begin{array}{l} L \rightarrow \mu_L L, \\ W \rightarrow \mu_W W, \\ D \rightarrow \mu_D D, \\ R_{0\max} \rightarrow \mu_{R_{0\max}} R_{0\max}, \\ R_{0\min} \rightarrow \mu_{R_{0\min}} R_{0\min}, \\ S \rightarrow \mu_S S, \\ E \rightarrow \mu_E E. \end{array} \right. \quad (57)$$

Step 2. We obtain the scaled version for each functional linear independent function in Equations (50), (53) and (55).

Step 3. We obtain the identifiability equations:

$$\frac{N - \mu_S S - \mu_E E - I}{\mu_S \mu_L L} = \frac{N - S - E - I}{L}. \quad (58)$$

$$\begin{aligned} & \frac{I \mu_S S}{\mu_S N \mu_D D} \left[e^{aq(t)} (\mu_{R_{0\max}} R_{0\max} - \mu_{R_{0\min}} R_{0\min}) + \mu_{R_{0\min}} R_{0\min} \right] \\ & = \frac{IS}{ND} \left[e^{aq(t)} (R_{0\max} - R_{0\min}) + R_{0\min} \right], \end{aligned} \quad (59)$$

$$\begin{aligned} & \frac{I\mu_S S}{\mu_E N \mu_D D} \left[e^{aq(t)} (\mu_{R_{0\max}} R_{0\max} - \mu_{R_{0\min}} R_{0\min}) + \mu_{R_{0\min}} R_{0\min} \right], \\ & = \frac{IS}{ND} \left[e^{aq(t)} (R_{0\max} - R_{0\min}) + R_{0\min} \right], \end{aligned} \quad (60)$$

$$\frac{\mu_E E}{\mu_E \mu_W W} = \frac{E}{W}, \quad (61)$$

$$\frac{\mu_E E}{\mu_W W} = \frac{E}{W}, \quad (62)$$

$$\frac{I}{\mu_D D} = \frac{I}{D}. \quad (63)$$

Similar to the previous derivation of the corresponding part of the SEIR model, it is easy to obtain:

$$\begin{cases} \mu_L = 1, \\ \mu_W = 1, \\ \mu_D = 1, \\ \mu_{R_{0\max}} = 1, \\ \mu_{R_{0\min}} = 1, \\ \mu_S = 1, \\ \mu_E = 1. \end{cases} \quad (64)$$

Therefore, the SEIRS model is identifiable. Namely, the parameter set $\Theta_4 = \{D, L, W, R_{0\max}, R_{0\min}\}$ can be determined uniquely from the observable I .

4. Discussion

In some existing studies, the authors explored the performance of different data assimilation algorithms when inferring parameters from observational epidemic data [13]. For example, ensemble filters are better at reproducing historical influenza incidence time series than particle Markov chain Monte Carlo. However, we still do not know whether the model is structurally unidentifiable, which affects the performance of the optimization algorithms. In this study, we applied the scaling method to analyse the structural identifiability of four types of commonly used humidity-driven epidemiological models when prevalence data is observable. Specifically, we investigated whether each parameter set $\Theta_1 = \{D, R_{0\max}, R_{0\min}\}$ in the SIS model, $\Theta_2 = \{D, L, R_{0\max}, R_{0\min}\}$ in the SIRS model, $\Theta_3 = \{D, W, R_{0\max}, R_{0\min}\}$ in the SEIR model and $\Theta_4 = \{D, L, W, R_{0\max}, R_{0\min}\}$ in the SEIRS model can be uniquely determined (D is the infectious period, L is the average duration of immunity, W is the latency period, and $R_{0\max}, R_{0\min}$ are the combination of the maximum and minimum daily basic reproductive number and the minimum daily basic reproductive number). We find each model is identifiable when the prevalence is observable. Aside from considering prevalence data as observational data, this study also considered the number of new cases as observational data. For example, when we introduced the differential equation of the new case changes over time in the SIRS model, the differential equation decomposition components (obtained based on Equation (8) from the frameworks of the scaling method part) are not linear-independent. So the scaling method cannot be used to determine whether the model is structurally identifiable when the new cases are observable.

Much work needs to be done to further test, validate, and improve the ability of humidity-driven epidemiological models to predict influenza activity. In the future, we will test more expanded humidity-driven epidemiological models (such as a model that includes more than one exposed state) to provide theoretical support for these models to be used in practice. This investigation involved testing of the structure identifiability of different humidity-driven epidemiological models. In the future, we will assess the practical identifiability of these models to provide synthetic experimental support to enable

these models to be used in practice. In addition, we will explore other model structure identifiability methods to test the structural identification of the humidity-driven epidemiological model when the number of new cases is observable. We will also consider more data types, such as the cumulative number of incidences.

In conclusion, our analysis suggests that the structural identifiability of these models can lay the foundation for testing practical identifiability in the future. In practice, epidemiological models should be examined with caution before using them to estimate model parameters from epidemic data.

Author Contributions: Conceptualization, Y.B., E.H.Y.L. and S.P.; methodology, C.Z., X.Z. and Y.B.; writing—original draft preparation, C.Z., X.Z., Y.B., E.H.Y.L. and S.P.; writing—review and editing, Y.B., E.H.Y.L. and S.P.; supervision, E.H.Y.L. and S.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the AIR@InnoHK programme of the Innovation and Technology Commission of the Hong Kong Special Administrative Region, the National Science Foundation of China (grant no. 72104208), the Health and Medical Research Fund, Food and Health Bureau, Government of the Hong Kong Special Administrative Region (grant no. 21200632), the National Institutes of Health (research funding to S.P. and grant no. R01AI163023), the Centers for Disease Control and Prevention (research funding to S.P. and grant no. U01CK000592 and 75D30122C14289) and the National Science Foundation (research funding to S.P. and grant no. DMS-2229605).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

We show the theorem to prove that the functions in the humidity-driven epidemiological model are linearly independent.

Theorem A1 ((Generalized Wronskian theorem) [37]). *Let $\Delta_0, \dots, \Delta_{n-1}$ be differential operators denoting by*

$$\Delta_s = \left(\frac{\partial}{\partial x_1}\right)^{j_1} \cdots \left(\frac{\partial}{\partial x_m}\right)^{j_m}, \quad j_1 + \cdots + j_m \leq s. \quad (\text{A1})$$

Then if the generalized Wronskian, W , over the functions $f_i(x_1, x_2, \dots, x_m)$, with $i = 1, \dots, n$,

$$W = \begin{vmatrix} \Delta_0 f_1 & \cdots & \Delta_0 f_n \\ \cdots & \cdots & \cdots \\ \Delta_{n-1} f_1 & \cdots & \Delta_{n-1} f_n \end{vmatrix} \quad (\text{A2})$$

is different from zero, then functions f_i are linearly independent. The only solution of the equation

$$\sum_{i=1}^n a_i f_i(x_1, x_2, \dots) = 0 \quad (\text{A3})$$

is $a_1 = \cdots = a_n = 0$.

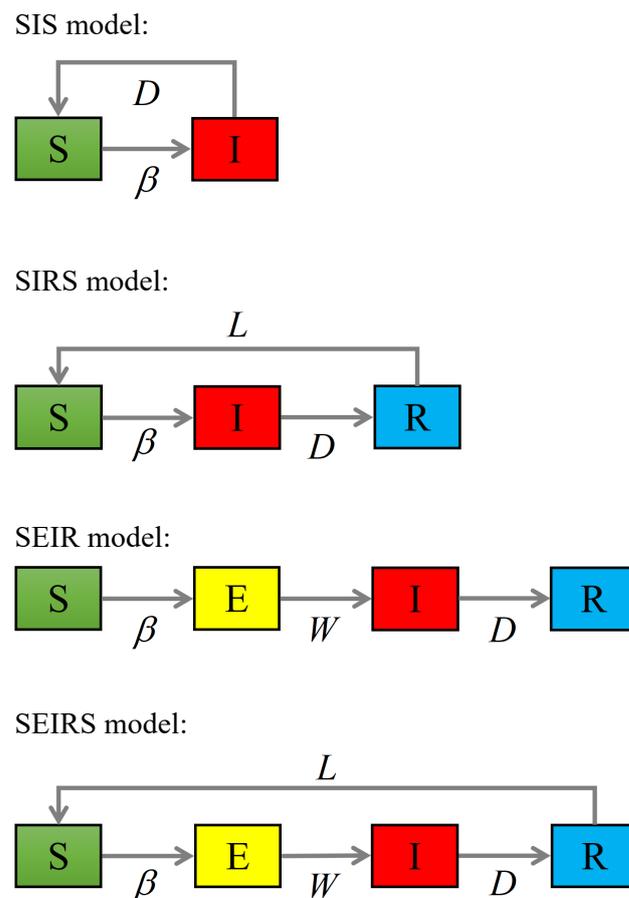


Figure A1. Flow diagram of the four types of humidity-driven epidemiological models.

References

- Iuliano, A.D.; Roguski, K.M.; Chang, H.H.; Muscatello, D.J.; Palekar, R.; Tempia, S.; Cohen, C.; Gran, J.M.; Schanzer, D.; Cowling, B.J.; et al. Estimates of global seasonal influenza-associated respiratory mortality: A modelling study. *Lancet* **2018**, *391*, 1285–1300. [[CrossRef](#)] [[PubMed](#)]
- Simonsen, L.; Spreeuwenberg, P.; Lustig, R.; Taylor, R.J.; Fleming, D.M.; Kroneman, M.; Van Kerkhove, M.D.; Mounts, A.W.; Paget, W.J.; Teams, G.C. Global mortality estimates for the 2009 Influenza Pandemic from the GLAMOR project: A modeling study. *PLoS Med.* **2013**, *10*, e1001558. [[CrossRef](#)] [[PubMed](#)]
- Paget, J.; Spreeuwenberg, P.; Charu, V.; Taylor, R.J.; Iuliano, A.D.; Bresee, J.; Simonsen, L.; Viboud, C. Global mortality associated with seasonal influenza epidemics: New burden estimates and predictors from the GLAMOR Project. *J. Glob. Health* **2019**, *9*, 020421. [[CrossRef](#)]
- Cox, N.J.; Subbarao, K. Global epidemiology of influenza: Past and present. *Annu. Rev. Med.* **2000**, *51*, 407–421. [[CrossRef](#)] [[PubMed](#)]
- Du, Z.; Nugent, C.; Galvani, A.P.; Krug, R.M.; Meyers, L.A. Modeling mitigation of influenza epidemics by baloxavir. *Nat. Commun.* **2020**, *11*, 2750. [[CrossRef](#)]
- Hayden, F.G.; Asher, J.; Cowling, B.J.; Hurt, A.C.; Ikematsu, H.; Kuhlbusch, K.; Lemenuel-Diot, A.; Du, Z.; Meyers, L.A.; Piedra, P.A.; et al. Reducing influenza virus transmission: The potential value of antiviral treatment. *Clin. Infect. Dis.* **2022**, *74*, 532–540. [[CrossRef](#)] [[PubMed](#)]
- Influenza (Seasonal). Available online: <https://www.who.int/health-topics/influenza-seasonal> (accessed on 6 October 2022).
- Du, Z.; Bai, Y.; Wang, L.; Herrera-Diestra, J.L.; Yuan, Z.; Guo, R.; Cowling, B.J.; Meyers, L.A.; Holme, P. Optimizing COVID-19 surveillance using historical electronic health records of influenza infections. *PNAS Nexus* **2022**, *1*, pgac038. [[CrossRef](#)] [[PubMed](#)]
- Du, Z.; Javan, E.; Nugent, C.; Cowling, B.J.; Meyers, L.A. Using the COVID-19 to influenza ratio to estimate early pandemic spread in Wuhan, China and Seattle, US. *EclinicalMedicine* **2020**, *26*, 100479. [[CrossRef](#)]
- Fineberg, H.V. Pandemic preparedness and response—lessons from the H1N1 influenza of 2009. *N. Engl. J. Med.* **2014**, *370*, 1335–1342. [[CrossRef](#)]
- Lee, S.S.; Viboud, C.; Petersen, E. Understanding the rebound of influenza in the post COVID-19 pandemic period holds important clues for epidemiology and control. *Int. J. Infect. Dis.* **2022**, *122*, 1002–1004. [[CrossRef](#)]

12. Baker, R.E.; Park, S.W.; Yang, W.; Vecchi, G.A.; Metcalf, C.J.E.; Grenfell, B.T. The impact of COVID-19 nonpharmaceutical interventions on the future dynamics of endemic infections. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 30547–30553. [[CrossRef](#)] [[PubMed](#)]
13. Yang, W.; Karspeck, A.; Shaman, J. Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics. *PLoS Comput. Biol.* **2014**, *10*, e1003583. [[CrossRef](#)] [[PubMed](#)]
14. Shaman, J.; Karspeck, A.; Yang, W.; Tamerius, J.; Lipsitch, M. Real-time influenza forecasts during the 2012–2013 season. *Nat. Commun.* **2013**, *4*, 2837. [[CrossRef](#)] [[PubMed](#)]
15. Yang, W.; Cowling, B.J.; Lau, E.H.; Shaman, J. Forecasting influenza epidemics in Hong Kong. *PLoS Comput. Biol.* **2015**, *11*, e1004383. [[CrossRef](#)] [[PubMed](#)]
16. Pei, S.; Kandula, S.; Yang, W.; Shaman, J. Forecasting the spatial transmission of influenza in the United States. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 2752–2757. [[CrossRef](#)]
17. Du, Z.; Fox, S.J.; Ingle, T.; Pignone, M.P.; Meyers, L.A. Projecting the Combined Health Care Burden of Seasonal Influenza and COVID-19 in the 2020–2021 Season. *MDM Policy Pract.* **2022**, *7*, 23814683221084631. [[CrossRef](#)]
18. Bellman, R.; Åström, K.J. On structural identifiability. *Math. Biosci.* **1970**, *7*, 329–339. [[CrossRef](#)]
19. Katafygiotis, L.S.; Beck, J.L. Updating models and their uncertainties. II: Model identifiability. *J. Eng. Mech.* **1998**, *124*, 463–467. [[CrossRef](#)]
20. Massonis, G.; Banga, J.R.; Villaverde, A.F. Structural identifiability and observability of compartmental models of the COVID-19 pandemic. *Annu. Rev. Control* **2021**, *51*, 441–459. [[CrossRef](#)]
21. Tuncer, N.; Le, T.T. Structural and practical identifiability analysis of outbreak models. *Math. Biosci.* **2018**, *299*, 1–18. [[CrossRef](#)]
22. Gallo, L.; Frasca, M.; Latora, V.; Russo, G. Lack of practical identifiability may hamper reliable predictions in COVID-19 epidemic models. *Sci. Adv.* **2022**, *8*, eabg5234. [[CrossRef](#)] [[PubMed](#)]
23. Wieland, F.G.; Hauber, A.L.; Rosenblatt, M.; Tönsing, C.; Timmer, J. On structural and practical identifiability. *Curr. Opin. Syst. Biol.* **2021**, *25*, 60–69. [[CrossRef](#)]
24. Villaverde, A.F.; Pathirana, D.; Fröhlich, F.; Hasenauer, J.; Banga, J.R. A protocol for dynamic model calibration. *Brief. Bioinform.* **2022**, *23*, bbab387. [[CrossRef](#)] [[PubMed](#)]
25. Sauer, T.; Berry, T.; Ebeigbe, D.; Norton, M.M.; Whalen, A.J.; Schiff, S.J. Identifiability of infection model parameters early in an epidemic. *SIAM J. Control Optim.* **2021**, *60*, S27–S48. [[CrossRef](#)] [[PubMed](#)]
26. Castro, M.; de Boer, R.J. Testing structural identifiability by a simple scaling method. *PLoS Comput. Biol.* **2020**, *16*, e1008248. [[CrossRef](#)] [[PubMed](#)]
27. Pohjanpalo, H. System identifiability based on the power series expansion of the solution. *Math. Biosci.* **1978**, *41*, 21–33. [[CrossRef](#)]
28. Chappell, M.J.; Godfrey, K.R. Structural identifiability of the parameters of a nonlinear batch reactor model. *Math. Biosci.* **1992**, *108*, 241–251. [[CrossRef](#)]
29. Denis-Vidal, L.; Joly-Blanchard, G. An easy to check criterion for (un) identifiability of uncontrolled systems and its applications. *IEEE Trans. Autom. Control* **2000**, *45*, 768–771. [[CrossRef](#)]
30. Xia, X.; Moog, C.H. Identifiability of nonlinear systems with application to HIV/AIDS models. *IEEE Trans. Autom. Control* **2003**, *48*, 330–336. [[CrossRef](#)]
31. Koelle, K.; Farrell, A.P.; Brooke, C.B.; Ke, R. Within-host infectious disease models accommodating cellular coinfection, with an application to influenza. *Virus Evol.* **2019**, *5*, vez018. [[CrossRef](#)]
32. Chis, O.T.; Banga, J.R.; Balsa-Canto, E. Structural identifiability of systems biology models: A critical comparison of methods. *PLoS ONE* **2011**, *6*, e27755. [[CrossRef](#)] [[PubMed](#)]
33. Miao, H.; Xia, X.; Perelson, A.S.; Wu, H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev.* **2011**, *53*, 3–39. [[CrossRef](#)] [[PubMed](#)]
34. Laskowski, M.; Demianyk, B.C.P.; Witt, J.; Mukhi, S.N.; Friesen, M.R.; McLeod, R.D. Agent-based modeling of the spread of influenza-like illness in an emergency department: A simulation study. *IEEE Trans. Inf. Technol. Biomed.* **2011**, *15*, 877–889. [[CrossRef](#)] [[PubMed](#)]
35. Shaman, J.; Pitzer, V.E.; Viboud, C.; Grenfell, B.T.; Lipsitch, M. Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biol.* **2010**, *8*, e1000316. [[CrossRef](#)]
36. Coronavirus (COVID-19) Infection Survey, UK Statistical Bulletins. Available online: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveys/pilot/previousReleases> (accessed on 21 October 2022).
37. Schmidt, W.M. *Diophantine Approximation*; Springer Science & Business Media: Berlin, Germany, 1996.