RandPunkt
Journal

ISSN 00009172 (Online)
APPLIED MATHEMATICS SERIES

A hybrid stochastic-deterministic operator framework for early detection of severe respiratory diseases with pandemic potential

Martha Rios Castro[1,*]

[1] *Department of Scientific Innovation, Instituto Nacional de Cardiología Ignacio Chávez, Mexico City, 14080, Mexico*

*Correspondence to: **M. Rios** ▪ marthar8@hotmail.com

Volume 1, Issue 1, 2026
Published on: 05 November 2026

Pages: [1–14] ▪ 14 pages
Article DOI: 10.5281/zenodo.24149964

Abstract. We formulate the early detection of severe respiratory diseases with pandemic potential as a hybrid stochastic-deterministic operator problem. The central object is a transition operator $T = D + S$, where D is a deterministic component driven by confirmed clinical data and S is a stochastic component driven by probabilistic propagation patterns.

We classify twelve early-warning algorithms as discrete realizations of T and evaluate them under the alarm-characterization framework of Daluwatte et al. Using a single sample of sixteen early warnings, we introduce a weighted alert functional $f(x_k) = (k-1) + x_k$ that maps the alert distribution into a scalar score. The results show that only the hybrid algorithms H1 and H3, together with the deterministic algorithm D4, attain high scores in the "early" category, while no purely stochastic algorithm produces early warnings.

We then construct a hybrid operator for a real-time implementation using a risk-weight matrix P derived from sociodemographic risk and symptom severity, alongside a cluster estimator. The work provides a unified operator formalism and a constructive hybrid operator consistent with theoretical predictions. The system is implemented in Python and deployed on mobile platforms, offering a reproducible tool for resource-limited settings.

Keywords: hybrid stochastic-deterministic operators • early-warning algorithms • epidemiological risk-weight matrices • weighted alert functional • mathematical epidemiology • pandemic preparedness

MSC (2020): 92D30 • 60J20 • 47A10 • 62P10 • 65C20

Received: 05 September 2026 ▪ Accepted: 05 October 2026 ▪ Published: 05 November 2026 © 2026 CC by ▪ 4.0v

1 Introduction

The growing threat of epidemic outbreaks, driven by globalization and high human mobility, has highlighted the urgent need for mathematical tools capable of predicting the emergence of severe respiratory diseases with pandemic potential, of viral origin [1]. Mathematical modeling of infectious diseases has a long history, with deterministic compartmental models based on ordinary differential equations providing the foundational framework [2, 3, 4]. The problem admits a natural operator formulation: given a state space $\mathcal{X}$ of epidemiological configurations (susceptible, infected, recovered, and their sociodemographic covariates), an early-warning algorithm is a map

$$T : \mathcal{X} \longrightarrow \mathcal{A}, \qquad \mathcal{A} = \{\text{early}, \text{on time}, \text{late}, \text{false}, \text{missed}\}, \tag{1}$$

that assigns to each configuration an alert category. The central question of this work is which class of operators T minimizes the detection delay while preserving reliability under incomplete data.

In this context, computational algorithms have proven essential for processing large volumes of data in real time, allowing the early identification of unusual patterns that could signal the onset of an outbreak [5]. These systems not only slow down the spread of the disease but also provide public health authorities with a valuable window of time to implement preventive measures, such as vaccination, travel control, or social distancing [6]. The ability to foresee outbreaks with pandemic potential through algorithms based on predictive analysis would radically transform the global response to health emergencies [7]. Global early-warning models have evolved to incorporate time-series analysis, spatial scan statistics, and Bayesian spatio-temporal frameworks [8]. However, in developing countries, where healthcare infrastructure is fragile and resources are limited, the implementation of these systems is particularly crucial [9]. The lack of effective early warning systems in these nations puts the population at serious risk, as any delay in response could significantly worsen the impact on health and the economy [10].

In this work, we present a system designed under a stochastic-deterministic profile, which can be implemented for the early detection of epidemics. We validated this system by categorizing twelve early-warning algorithms, with the aim of assessing whether a stochastic-deterministic approach is suitable for the early prediction of epidemic outbreaks, given its potential to combine accuracy with flexibility [5]. Mathematically, this hybrid approach corresponds to a decomposition

$$T = D + S, \tag{2}$$

where D is the deterministic component (driven by confirmed clinical data) and S is the stochastic component (driven by propagation patterns and random factors). Hybrid stochastic-deterministic frameworks have been developed for compartmental models, enabling simulation-based inference while preserving conservation principles [11]. This decomposition allows systems to better adapt to complex situations and imperfect data, which is particularly relevant in developing countries, where early detection systems often provide limited data [9].

2 Materials and Methods

This section is divided into two parts: first, we apply the alarm-characterization methodology of [12] to the twelve algorithms described below; second, we propose a computational implementation based on the best algorithms evaluated here.

2.1 Types of Algorithms

The first part of this analysis considers the classification of the twelve algorithms into deterministic, stochastic, and hybrid. We describe each class below in operator-theoretic terms.

2.1.1 Deterministic algorithms.

Deterministic systems are based on a deterministic function D and require confirmed data to generate predictions and issue alerts. These models are accurate in their analysis and predictions, as long as they have precise and up-to-date information. The classical SIR model of Kermack and McKendrick [2] and its extensions, such as the MSEIR model [13], are canonical examples. However, their main limitation lies in the need for confirmed clinical data, which means that the alert may be generated too late, when the spread of the virus has already reached a significant level. Formally, if D depends only on confirmed cases $C(t)$, the detection time satisfies $t_{\text{alert}} \geq t_{\text{confirm}}$, and since t_{confirm} is itself delayed by the asymptomatic period and underreporting, deterministic operators anchored on confirmed cases are structurally unable to produce early warnings. Consequently, these systems may not be effective in the early detection of pandemics [14].

2.1.2 Stochastic algorithms.

Stochastic systems primarily use a stochastic function S or employ probabilistic models that do not necessarily rely on confirmed clinical data, but rather on propagation patterns and random factors to predict the spread of a disease [15, 16]. Although these systems can be useful for identifying potential outbreaks before cases are clinically confirmed, their lack of support from medical evidence makes them less reliable in the eyes of the health sector. The inherent uncertainty of these models can lead to false alarms, which can result in widespread distrust and poor decision-making [17]. In operator terms, S acts on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and produces alerts with a high variance, which is why purely stochastic operators tend to over-trigger.

2.1.3 Hybrid algorithms.

The integration of deterministic and stochastic components in an early-warning system results in a hybrid operator $T = D + S$ that combines the strengths of both models, overcoming their respective limitations. Hybrid approaches to compartmental models have been formalized as piecewise deterministic Markov processes (PDMPs), in which compartments can switch between stochastic and deterministic regimes based on population thresholds [11]. These hybrid systems are capable of using confirmed data to ensure the accuracy and reliability of their predictions, while also incorporating probabilistic analysis to anticipate potential outbreaks before they are clinically confirmed. In this way, they offer a greater capacity to generate early and reliable alerts, allowing health authorities to implement preventive measures in a timely and efficient manner. By merging the certainty of confirmed clinical data with the predictive capability of stochastic models, hybrid systems represent a significant evolution in the fight against pandemics. They are particularly useful for monitoring emerging diseases with high potential for spread, as they allow for a quicker response based on a comprehensive analysis of the situation. While pure deterministic and pure stochastic systems have their own advantages and limitations, it is the integration of both approaches that offers the most robust and efficient solution for early pandemic detection. Hybrid systems are not only capable of warning about the spread of an emerging virus with greater anticipation, but they also provide a reliable framework for mitigating the impact of potential pandemics. Currently, due to high global connectivity and the emergence of unpredictable diseases, early warning systems need to incorporate components from both hybrid and stochastic systems to achieve a more useful response to respiratory epidemics and avoid significant human and economic losses.

2.2 Evaluation of algorithms

A systematic review of the literature was conducted on PubMed, resulting in the collection of twelve primary articles (Tables 1–3). These articles present mathematical algorithms designed to generate early

epidemiological alerts for severe respiratory infections. Four algorithms with a stochastic-deterministic (hybrid) profile were selected (Table 1), four with a stochastic profile (Table 2), and four with a deterministic profile (Table 3).

Table 1. Revised hybrid-type algorithms. Each row corresponds to a hybrid operator $T = D + S$ whose deterministic part D is anchored in confirmed clinical data and whose stochastic part S captures propagation patterns.

Discipline	Subdiscipline	Algorithm	Reference
Probability	Discrete Markov Chains, Sequences and Series	H1: StochCum Method	[18]
Probability	Partial Differential Equations, Sequences and Series	H2: SLIAR	[19]
Probability	Neural Networks	H3: Stochastic metapopulation epidemic model	[20]
Probability	Stochastic–deterministic coupling	H4: Hybrid stochastic–deterministic model	[11]

Table 2. Revised stochastic-type algorithms. Each row corresponds to a purely stochastic operator S acting on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Discipline	Subdiscipline	Algorithm	Reference
Probability	Probabilistic Distributions	E1: SIDR	[21]
Probability	Heuristic Programming	E2: SHR	[22]
Probability	Hidden Markov models	E3: EM algorithm	[23]
Probability	Stochastic processes	E4: Stochastic epidemic model	[17]

The algorithms (Tables 1–3) were subjected to the methodology of [12], and to a single sample composed of sixteen early warnings. The aforementioned methodology classifies each alert as: early, timely, late, false, or lost. To understand this evaluation, we show the following example.

2.3 Example of the Procedure

The H3 algorithm is evaluated by the methodology [12], and the sixteen early warnings it receives are rated as:

This same distribution with respect to 16 is equivalent to:

Then we apply the functional $f(x_k) = (k-1) + x_k$ to those values, according to the rule: lost, $k = 1$; false, $k = 2$; late, $k = 3$; on time, $k = 4$; early, $k = 5$. Particularly for the H3 algorithm, if we select the maximum value $x_k = 3 = 0.88$ (proportion of the number of alerts), then $(k-1) + x_k = (5-1) + 0.88 = 4 + 0.88 = 4.88$. This functional is the discrete analogue of a weighted norm on the alert space: each category is assigned an integer level k and the fraction x_k of alerts in that category modulates the score within the interval $[k-1, k]$. The choice of $k = 1$ for missed and $k = 5$ for early reflects the epidemiological priority: an early alert is worth five times a missed one in terms of the response window it provides to public-health authorities. The functional is monotone in x_k and bounded in $[0, 5]$, which makes it suitable for ranking heterogeneous algorithms.

Table 3. Revised deterministic-type algorithms. Each row corresponds to a purely deterministic operator D driven by confirmed clinical data.

Discipline	Subdiscipline	Algorithm	Reference
Mathematical Analysis	Sequences and Series	D1: SIR model	[2]
Mathematical Analysis	Differential Equations	D2: MSEIR	[13]
Mathematical Analysis	Differential Equations	D3: Early warning signals	[24]
Mathematical Analysis	Sequences and Series	D4: ISM	[25]

Table 4. Raw alert counts for algorithm H3 over a single sample of sixteen early warnings.

Algorithm	False	Early	On Time	Late	Missed
H3	0	**14**	2	0	0

Table 5. Normalized alert proportions for algorithm H3.

Algorithm	False	Early	On Time	Late	Missed
H3	0	14/16 = 0.88	2/16 = 0.13	0	0

Table 6. Weighted alert categories for algorithm H3. The value 4.88 in the "Early" column is obtained from the functional $f(x_k) = (k-1) + x_k$ with $k = 5$ and $x_k = 0.88$.

Algorithm	False	Early	On Time	Late	Missed
H3	0	**4.88**	0.13	0	0

This last value of 4.88 corresponding to “early alert” is the indicator of the H3 algorithm.

2.4 Evaluation procedure

Analogously to this example, the remaining eleven algorithms were evaluated, and their results can be seen in Table 7.

Table 7. Raw alert counts for the twelve algorithms over a single sample of sixteen early warnings. The maximum value in each row is shown in bold.

Algorithm	False	Early	On Time	Late	Missed
H1	0	**8**	8	0	0
H2	0	0	0	**14**	2
H3	0	**14**	2	0	0
H4	0	0	3	**13**	0
E1	0	0	0	**12**	4
E2	0	0	0	**14**	2
E3	0	0	0	**10**	6
E4	0	0	0	**13**	3
D1	0	0	0	**15**	1
D2	0	0	0	**14**	2
D3	0	0	0	**15**	1
D4	0	**12**	4	0	0

According to the procedure [12], each distribution of the alerts presented in Table 7 was evaluated, resulting in Table 8.

Table 8. Normalized alert proportions for the twelve algorithms.

Algorithm	False	Early	On Time	Late	Missed
H1	0	0.5	0.5	0	0
H2	0	0	0	0.88	0.13
H3	0	0.88	0.13	0	0
H4	0	0	0.19	0.81	0
E1	0	0	0	0.75	0.25
E2	0	0	0	0.88	0.13
E3	0	0	0	0.63	0.38
E4	0	0	0	0.81	0.19
D1	0	0	0	0.94	0.06
D2	0	0	0	0.88	0.13
D3	0	0	0	0.94	0.06
D4	0	0.75	0.25	0	0

The graphical interpretation of Table 9 as a trend curve with a fourth-degree polynomial function shows the proportion of alerts per record (Figure 1). The choice of a fourth-degree polynomial requires clarification, because the curve serves two distinct purposes that must not be conflated. First, as a pedagogical device, the curve is required to pass through one representative point per alert category, and a polynomial of degree d can interpolate at most $d + 1$ points; since there are five categories, degree 4 is the minimal degree that allows the curve to visit all five. Each representative point is defined as the category centroid, i.e., the mean of the weighted scores $f(x_k)$ across all algorithms that place at least one alert in that category. Second, the curve is not intended as a statistical fit to the data in Table 9; it is a visual aid to compare the relative positions of the five categories. If a reader wishes to interpret the curve as a trend estimator rather than as an interpolant, a lower-degree regression (degree 1 or 2) with

Table 9. Weighted alert categories for the twelve algorithms, obtained by applying the functional $f(x) = (k-1)+x$ to the normalized proportions of Table 8.

Algorithm	False	Early	On Time	Late	Missed
H1	0	**4.50**	3.50	0	0
H2	0	0	0	2.88	1.13
H3	0	**4.88**	3.13	0	0
H4	0	0	3.19	2.81	0
E1	0	0	0	2.75	1.25
E2	0	0	0	2.88	1.13
E3	0	0	0	2.63	1.38
E4	0	0	0	2.81	1.19
D1	0	0	0	2.94	1.06
D2	0	0	0	2.88	1.13
D3	0	0	0	2.94	1.06
D4	0	**4.75**	3.25	0	0

a reported coefficient of determination would be more appropriate, and we note this as an alternative reading of the same figure. The interpolation is therefore a deliberate display choice, not a claim about the underlying data- generating process.

Figure 1. Distribution of weighted values from Table 9 based on the function $f(x_k) = (k-1) + x_k$ for each type of alert: lost, $k = 1$; false, $k = 2$; late, $k = 3$; on time, $k = 4$; early, $k = 5$. The curve is a fourth-degree polynomial that interpolates the five category centroids; it is a display aid, not a statistical fit. Only H1, D4, and H3 receive a high score for Weighted Alert.

2.5 Description of the proposed system

The proposed system is a hybrid operator that takes into account clinical symptoms characteristic of severe airway diseases, and non-clinical factors of population vulnerability. The system is designed to operate in hospital emergency services and health clinics in small towns and villages. Particularly, the latter characterize healthcare in developing countries or those experiencing disordered growth. The system is written in Python and can be downloaded for free from GitHub, and run on any mobile phone with an Android or iOS operating system.

Mathematically, the operator is decomposed as in (2): the deterministic part D encodes the clinical symptoms of the most severe viral diseases that have affected the global population in the last 60 years (Table 10), while the stochastic part S encodes the sociodemographic risk factors (Table 11).

2.5.1 Step 1: Assign numerical values to the risks

Sociodemographic factors (transmission risk). We assign a numerical scale for the sociodemographic factors, where 1 is the lowest risk and 3 is the highest:

- High sociodemographic risk: 3 (e.g., high population density, limited access to healthcare)
- Moderate sociodemographic risk: 2 (e.g., moderate mobility, low educational level)
- Low sociodemographic risk: 1 (e.g., low overcrowding, adequate access to health care)

Symptoms (danger level of the virus). We assign another scale for the symptoms of the viruses, also from 1 to 3:

Table 10. Classification of the hazard levels of the main respiratory viruses and their severe symptoms. The "Dangerousness" column encodes the deterministic component D of the operator.

Virus	Dangerousness	Exclusive or Mild Symptoms	Severe Symptoms
SARS-CoV-2 (COVID-19)	High	Loss of smell (anosmia) and taste (ageusia). Skin rashes and "COVID toes".	Severe respiratory distress. Extreme fatigue with rapid onset. Intense chills.
Influenza (H1N1) A	High	Intense muscle and joint pain. Extreme fatigue with rapid onset. Intense chills.	Wheezing and rapid breathing in children. Apnea in babies. Pneumonia and bronchiolitis. "Barking" cough (croup). Laryngotracheobronchitis.
Respiratory Syncytial Virus (RSV)	Moderate and high	Wheezing and rapid breathing in children. Apnea in babies. Pneumonia and bronchiolitis in infants and the elderly.	"Barking" cough (croup). Laryngotracheobronchitis.
Parainfluenza	Moderate	"Barking" cough (croup). Laryngotracheobronchitis.	Pneumonia in immunocompromised individuals.
Adenovirus	Moderate	Conjunctivitis. Diarrhea (children). Urinary tract infections.	It usually does not cause severe symptoms.
Seasonal coronaviruses	Low	Nasal congestion. Sore throat. Mild cough. Mild headache.	Exacerbation of asthma or COPD.
Rhinovirus (common cold)	Low	Frequent sneezing. Nasal drip. Sore throat. Very low or absent fever.	Severe respiratory distress. Skin rashes and "COVID toes".

Table 11. Key factors that exacerbate the transmission of respiratory viruses. The "Dangerousness" column encodes the stochastic component S of the operator.

Factor	Dangerousness	Description
Population density	High	Densely populated areas facilitate close contact, increasing the risk of contagion.
Limited access to health services	High	Lack of access to medical services delays diagnosis and treatment.
Housing conditions (overcrowding)	High	Overcrowding and lack of ventilation in households increase virus transmission.
Working conditions (close contact)	High	Jobs requiring close physical contact and the impossibility of teleworking increase the risk.
Population mobility	Moderate	High mobility facilitates the spread of viruses to other geographic areas.
Low educational level	Moderate	People with less access to education may have less knowledge about preventive measures.
Social and community habits	Moderate	Customs such as hugs, handshakes, and large gatherings favor virus transmission.
Reluctance to seek medical care	Moderate	Cultural beliefs or distrust in medicine delay seeking treatment.
Air pollution	Moderate	Air pollution weakens the respiratory system.
Poor hygiene standards	Moderate	Lack of adequate hygiene practices increases virus transmission.
Cold and dry climate	Low	Cold and dry climate facilitates virus spread because people spend more time indoors.
Misinformation about viruses	Moderate	The spread of myths and misinformation reduces vaccination rates.

Table 12. Classification of the 32 states of Mexico according to their level of risk from respiratory viruses (Table 10) and sociodemographic factors (Table 11).

State	Virus	Sociodemographic
Estado de México	High	High
Aguascalientes	High	High
Baja California	Moderate	High
Baja California Sur	Moderate	High
Campeche	Moderate	High
Chiapas	Moderate	Moderate
Chihuahua	Low	Low
Ciudad de México	High	High
Coahuila	Moderate	Moderate
Colima	Moderate	Moderate
Durango	Low	Low
Guanajuato	Moderate	Moderate
Guerrero	Moderate	Moderate
Hidalgo	Moderate	Moderate
Jalisco	Moderate	Moderate
Michoacán	Moderate	Moderate
Morelos	High	High
Nayarit	Low	Low
Nuevo León	Moderate	Moderate
Oaxaca	Moderate	Moderate
Puebla	High	High
Querétaro	Moderate	Moderate
Quintana Roo	Moderate	Moderate
San Luis Potosí	Moderate	Moderate
Sinaloa	Moderate	Moderate
Sonora	Low	Low
Tabasco	Moderate	Moderate
Tamaulipas	Moderate	Moderate
Tlaxcala	High	High
Veracruz	Moderate	Moderate
Yucatán	Moderate	Moderate
Zacatecas	Moderate	Moderate

- High risk due to symptoms: 3 (e.g., severe respiratory difficulty, pneumonia)
- Moderate danger due to symptoms: 2 (e.g., croup cough, wheezing)
- Low risk due to symptoms: 1 (e.g., nasal congestion, low fever)

2.5.2 Step 2: Compute a normalized risk score

To combine the two risk dimensions, we use a normalized score rather than a probability. The base formula is:

$$S(\text{risk}) = \frac{v_{\text{socio}} + v_{\text{symptom}}}{6}. \tag{3}$$

We divide by 6 because it is the maximum possible sum (3 from sociodemographics + 3 from symptoms = 6). This formula generates values between 0 and 1, which we use as normalized scores in the risk-weight matrix described below. The arithmetic mean of the two ordinal scores is a first-order modeling choice: it treats the sociodemographic and symptomatic dimensions as equally important and additive, which is a simplifying assumption rather than a probabilistic derivation. The resulting score $S(\text{risk})$ must not be interpreted as a transition probability; it is a relative risk indicator that ranks the nine combinations of the two dimensions.

2.5.3 Step 3: Compute the entries of the matrix

Example of computing an entry of the matrix. Let us consider a state with high sociodemographic danger and high symptom danger: $v_{\text{socio}} = 3$, $v_{\text{symptom}} = 3$. We use the formula:

$$S(\text{high}, \text{high}) = \frac{3+3}{6} = 1. \tag{4}$$

This value of 1 is the maximum normalized score, indicating that both sociodemographic and symptom factors are at their highest danger level. In the risk-weight matrix, this score is used as a heuristic weight, not as an empirically estimated probability of remaining in the state.

Example: Moderate sociodemographic and low by symptoms:

$$S(\text{moderate}, \text{low}) = \frac{2+1}{6} = 0.5. \tag{5}$$

This indicates a relative risk level of 0.5 for a state with moderate sociodemographic danger and low symptom danger. As above, the value is a normalized score, not a transition probability.

2.5.4 Step 4: Complete the risk-weight matrix

We can now construct the risk-weight matrix $P = (p_{ij})$ using these calculations for each combination:

Table 13. Risk-weight matrix $P = (p_{ij})$ for the nine risk combinations. The entries are normalized scores in $[0, 1]$, not transition probabilities; consequently, the rows do not sum to 1 and P is not a stochastic matrix in the Markov-chain sense.

Risk factors / Symptoms	High symptom	Moderate symptom	Low symptom
High sociodemographic	1	0.83	0.67
Moderate sociodemographic	0.83	0.67	0.5
Low sociodemographic	0.67	0.5	0.33

2.5.5 Step 5: Interpretation

If a state has high sociodemographic danger and high symptom danger, the normalized score attains its maximum value of 1, reflecting the highest combined danger from both dimensions. If a state has moderate sociodemographic danger and moderate symptom danger, the score is 0.67. If a state has low sociodemographic risk and low symptom risk, the score is 0.33. These values are relative risk indicators, not literal probabilities of remaining in a state.

With this information, the system, based on the postal code of the person completing the questionnaire, retrieves their geographical position and calculates the local incidence proportion within a 3-kilometer radius, based on the average of incidents. Formally, if $\{x_1, \ldots, x_n\}$ are the geolocated cases, the local incidence proportion is

$$\hat{C} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{|x_i - x_0| \le 3\,\mathrm{km}\}}, \tag{6}$$

where x_0 is the centroid of the postal-code region. The choice of a 3 km radius is a compromise between two competing requirements: a smaller radius (e.g., 1 km) would fragment urban clusters into statistically unstable units, while a larger radius (e.g., 10 km) would merge epidemiologically distinct neighborhoods. Three kilometers corresponds approximately to the typical catchment area of a primary-care clinic in Mexico, which makes the estimator operationally meaningful. Spatial scan statistics provide a complementary framework for cluster detection by identifying geographic areas where monitoring indicators deviate significantly from the expected level [8].

Subsequent to that, update each cluster taking into account the average length of hospital stay (Table 14), in intensive care and in intermediate care.

Table 14. Detailed analysis of the average days that patients with severe respiratory viral infections spend in intensive care (ICU) and intermediate care (IMC).

Infection	ICU	IMC	Observations
COVID-19 (SARS-CoV-2)	10–15 days	7–10 days	Severe cases may require prolonged mechanical ventilation
Influenza H1N1 (2009)	7–14 days	5–10 days	Complications such as pneumonia may prolong ICU stay
SARS (2002–2003)	10–20 days	5–8 days	Patients with severe pneumonia may require longer ICU time
MERS (2012–present)	10–14 days	7–10 days	High mortality, commonly associated with severe respiratory failure
RSV	7–14 days	5–8 days	Common in children and older adults with severe respiratory complications
General Average	**9–15 days**	**6–9 days**	Average ICU and IMC stay based on the infections analyzed

Once these calculations are performed, a graph of the daily distribution of the accumulation, the daily threshold, and a linear approximation using the least squares method is generated (Figure 2). Mathematically, if $\{y_t\}_{t=1}^{T}$ are the daily accumulated cases, the least-squares estimator of the slope is

$$\hat{\beta} = \frac{\sum_{t=1}^{T} (t - \bar{t})(y_t - \bar{y})}{\sum_{t=1}^{T} (t - \bar{t})^2}, \tag{7}$$

and the alert is triggered when y_t exceeds the threshold $\tau = 10$. The threshold $\tau = 10$ was calibrated empirically: it corresponds to the smallest daily count that, in the Mexican states analyzed, preceded a sustained upward trend in the least-squares slope $\hat{\beta}$ during the early phase of the COVID-19 pandemic. A

lower threshold (e.g., $\tau = 5$) would trigger alerts on normal seasonal fluctuations, while a higher threshold (e.g., $\tau = 20$) would delay the alert by several days, defeating the purpose of early detection. The calibration procedure and the underlying daily series are documented in Figure 2 and in the supplementary material available at the project repository. The choice of a first-degree (linear) least-squares fit, rather than a higher-degree polynomial, is deliberate: in the early phase of an outbreak the signal is dominated by exponential growth, but over short time windows (5–7 days) a linear approximation is sufficient and more robust to noise.

Figure 2. Daily evolution of grouped cases in Mexico. The black dots represent the actual daily incident data. The dashed burgundy line shows the trend obtained through a first-degree least squares fit; the slope $\hat{\beta}$ is the best linear unbiased estimator of the local trend. The dark green horizontal line indicates the alert threshold set at 10 cases per day.

2.6 Software availability

The source programs, as well as the supplementary files, are available on GitHub: `https://github.com/polancodf2024/WARNINGS` and the executable program for both cell phones and computers on the website.

3 Results

Table 7 shows that most algorithms issue a late alert, which implies that the spread of the disease will increase before the health authority considers issuing a warning to the population. Table 9 and Figure 1 show that only three algorithms register a high score for early warning: H1 (4.50), D4 (4.75), and H3 (4.88). The ordering H3 > D4 > H1 is significant: H3 is a hybrid stochastic meta-population model, D4 is a deterministic index (the modified overcrowding index), and H1 is the StochCum hybrid method. The fact that H3, a hybrid stochastic meta-population model, achieves the highest score is consistent with the possibility that the stochastic component S is not merely a corrective term but may contribute significantly to early-detection performance when coupled with a deterministic anchor; however, the data presented do not establish a causal relationship. D4's high score, despite being purely deterministic, is explained by its dependence on emergency-room saturation rather than on confirmed cases: it is a deterministic operator acting on a proxy variable that leads the confirmed-case curve. H1, being a balanced hybrid, scores slightly lower because its deterministic component is anchored on confirmed cases and therefore inherits part of their delay. No stochastic algorithm was found that issued early warnings (Table 8), which confirms that S alone is insufficient: the best stochastic algorithm (E4) reaches only 2.81 in the "Late" category and 1.19 in "Missed", with no early alerts at all.

In operator terms, this means that the stochastic component S alone is insufficient to produce early alerts; the deterministic component D must be present to anchor the operator in confirmed data. Conversely, D alone is also insufficient when it is anchored on confirmed cases, since it then inherits their delay. Only the composition $T = D + S$ achieves the required balance. The numerical ranking H3 > D4 > H1 is consistent with this theoretical prediction: H3 has the strongest S component, D4 has a deterministic component acting on a leading proxy, and H1 has a balanced but confirmed-case anchored D.

4 Discussion

Recent pandemic events have caused health crises worldwide, primarily due to the high virulence and transmissibility of emerging viruses like SARS-CoV-2, as well as their adaptability in humans, all of

which have played a crucial role in its global spread [26]. Understanding these mechanisms at the time allowed for the design of effective prevention strategies, such as the use of masks or hand hygiene, but these strategies proved ineffective due to the magnitude of epidemic outbreaks.

In this study, we propose an early-warning operator that, by its nature (probabilistic and stochastic components), can help mitigate the damage of these severe health events. Given that most upper respiratory tract diseases spread through air or close contact, the probabilistic components (Table 11) include essential factors that accelerate disease transmission. This makes them more useful today, as nowadays, an infected person can travel to another continent and, in less than 5 hours, carry a highly infectious disease with them, making this component very valuable.

Regarding the deterministic components, the clinical factors hold significant importance. While it is clinically known that a certain percentage of the infected population can trigger an early alert, this alert typically arrives late due to the fact that the majority of the population either lacks the habit of visiting a doctor or, worse yet, lacks the necessary financial resources to receive medical attention. The asymptomatic period, during which most people present, is another important factor that facilitates the spread of the disease before its detection, underscoring the importance of early diagnosis and epidemiological surveillance. Therefore, the percentage of microbiologically confirmed infections is always likely to be biased. Moreover, studies show that for every infected and confirmed person, at least 8 other people are infected, so if only deterministic components are considered, the alert will always be late. Being a hybrid system, ours could prevent a high burden of infections, mitigating the impact of these pandemics primarily on the most vulnerable population.

We use stochastic characteristics to account for other important variables in our mixed system that cannot predict an early warning. One of them is the spread of the disease, which generally depends on the immune response of the hosts; this, in turn, depends on multiple biological, environmental, and genetic variables, making it impossible to predict the outcomes with precision due to its random nature [15]. Each organism develops a unique immune system as a result of its genetics, which generates a unique response to infections and vaccines, creating variability in the onset of a disease and its spread. Depending on the study's approach and the model's objective, mathematical models can represent vaccinated individuals in both deterministic and probabilistic ways.

We can determine a fixed proportion of the population that has received vaccinations in a deterministic model (such as the classic SIR or SEIR models). This approach assumes that the dynamics of the population, whether vaccinated or not, follow differential equations, with predictable outcomes provided the initial conditions and parameters are known. A probabilistic model, primarily focused on probability, treats vaccination as an event with a corresponding percentage of probability. This approach encapsulates the inherent uncertainty of a vaccine's impact on each individual or its distribution within the population. It is crucial to acknowledge that all organisms, to varying degrees, possess the capacity to mutate or develop resistance to treatments or vaccines, thereby intensifying the impact of a disease. Moreover, individual factors like age, health status, and vaccine efficacy influence the likelihood of an infection. Consider the SARS-CoV-2 variants that pose a challenge to control strategies. The operator is intended to incorporate immune response variability, which we expect will contribute to precise and robust early alerts, though comparative performance remains to be established.

The risk-weight matrix (Table 13) deserves a separate comment. Its diagonal entries (1, 0.67, 0.33) indicate that the score is largest when both dimensions coincide at the same level, which is consistent with the intended semantics: the score ranks the combined risk of a (sociodemographic, symptom) pair, and the highest combined risk occurs when both dimensions are high. Because the entries are normalized scores rather than transition probabilities, the matrix is symmetric and its rows do not sum to 1; it should not be interpreted as a Markov transition kernel. The symmetry of the matrix reflects the equal weighting of sociodemographic and clinical risks, which is a modeling choice rather than a biological necessity. A

natural extension would be to introduce a parameter $\alpha \in [0, 1]$ such that

$$p_{ij} = \alpha \frac{v_{\text{socio}}}{3} + (1 - \alpha) \frac{v_{\text{symptom}}}{3}, \tag{8}$$

which would allow the model to be calibrated to regions where one of the two components dominates. This is left for future work.

The spatial discretization (Table 12) reveals a heterogeneous risk landscape: Estado de México, Ciudad de México, Aguascalientes, Baja California, Baja California Sur, Campeche, Morelos, Puebla, and Tlaxcala appear as high/high, while Chihuahua, Durango, Nayarit, and Sonora appear as low/low. This heterogeneity justifies the cluster estimator: a single national threshold would be inappropriate, whereas a local cluster-based alert can adapt to the regional risk level. The 3 km radius is consistent with the typical catchment area of a primary-care clinic in Mexico, and the average length of stay (Table 14) determines the time window over which cases are aggregated.

Finally, the threshold $\tau = 10$ (Figure 2) was calibrated empirically. A lower threshold would trigger alerts on normal seasonal fluctuations, while a higher threshold would delay the alert by several days. The choice of a first-degree least-squares fit, rather than a higher-degree polynomial, is deliberate: over short time windows (5–7 days) a linear approximation is sufficient and more robust to noise. The intersection of the fitted line with the threshold marks the estimated alert time, which in the Mexican data preceded the observed peak by several days. The World Health Organization's Epidemic Intelligence from Open Sources (EIOS) system exemplifies the integration of diverse data sources for early detection, and our operator is complementary to such global surveillance initiatives [7].

5 Conclusions

Hybrid behavioral operators for generating early warnings in airborne respiratory diseases that classify the population into Susceptible, Infected, and Recovered (SIR), and some of their variants (H1, D4, H3), are the ones that best interpret the imminence of an epidemic event with a minimal number of infections, demonstrating to be a comprehensive and effective strategy to address current challenges in epidemiological surveillance. Mathematically, the decomposition $T = D + S$ provides a principled explanation of why hybrid operators are expected to be preferable to purely deterministic and purely stochastic ones: D anchors the prediction in confirmed data, while S extends the detection horizon beyond confirmed cases. The numerical ranking H3 > D4 > H1 is consistent with this prediction, with H3's strong stochastic component giving it the highest early-warning score. Therefore, its use is recommended to generate early warnings in resource-limited settings, where imperfect data is the norm rather than the exception.

Supplementary information

Supplementary information is available at the journal's website and at

`https://github.com/polancodf2024/WARNINGS`.

Acknowledgements

The author thank the anonymous referees for their careful reading of the manuscript and for their constructive comments, which substantially improved the clarity and rigor of the mathematical framework and the presentation of the epidemiological results.

Funding

This research received no external funding.

Authors' contributions

M.R.C conceived the study, developed the mathematical framework, and wrote the manuscript.

References

[1] Lemey, P., Rambaut, A., Bedford, T., Faria, N., Bielejec, F., Baele, G., Russell, C. A., Smith, D. J., Pybus, O. G., Brockmann, D., and Suchard, M. A., *Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza H3N2*, PLoS Pathogens 10.**2** (Feb. 2014), e1003932, doi:10.1371/journal.ppat.1003932.

[2] Kermack, W. O. and McKendrick, A. G., *Contributions to the mathematical theory of epidemics–I. 1927*, Bulletin of Mathematical Biology 53.**1-2** (1991), pp. 33–55, doi:10.1007/BF02464423.

[3] Heesterbeek, J. A. P., *Epidemic Models, Deterministic*, Encyclopedia of Biostatistics, Wiley, 2005, doi:10.1002/0470011815.b2a07015.

[4] Martcheva, M., *An Introduction to Mathematical Epidemiology*, vol. 61, Texts in Applied Mathematics, Springer, 2015, doi:10.1007/978-1-4899-7612-3.

[5] Pramod, A., Abhishek, J. S., and Suganthi, K., *Epidemic outbreak prediction using machine learning models*, arXiv:2310.19760, Oct. 2023, https://arxiv.org/abs/2310.19760.

[6] Chouit, E. M., Rachdi, M., Bellafkih, M., and Raouyane, B., *Harnessing the Power of Artificial Intelligence and the Internet of Things for Improved Epidemic Forecasting*, Advances in Environmental Engineering and Green Technologies Book Series, IGI Global, 2024, doi:10.4018/979-8-3693-3807-0.ch005.

[7] World Health Organization, *WHO upgrades its public health intelligence system to boost global health security*, WHO News Release, Oct. 2025, https://www.who.int/news/item/13-10-2025-who-upgrades-its-public-health-intelligence-system-to-boost-global-health-security.

[8] Chen, Y., Chu, C. W., Chen, M. I.-C., and Cook, A. R., *The utility of LASSO-based models for real time forecasts of endemic infectious diseases: A cross country comparison*, Journal of Biomedical Informatics 81 (May 2018), pp. 16–30, doi:10.1016/j.jbi.2018.02.014.

[9] Gomez Piraban, P. R., *Early warning for emerging infectious disease outbreaks: Digital disease surveillance for public health preparedness and response*, https://typeset.io/papers/early-warning-for-emerging-infectious-disease-outbreaks-13fribzx, 2023.

[10] Isiaka, A. B., Anakwenze, V. N., Ilodinso, C. R., Anaukwu, C. G., Ezeokoli, C. M.-V., Noi, S. M., and Adonu, R. M., *Harnessing Artificial Intelligence for Early Detection and Management of Infectious Disease Outbreaks*, International Journal of Innovative Research and Development 13.**2** (Feb. 2024), doi:10.24940/ijird/2024/v13/i2/feb24016.

[11] Germano, D. P. J., Zarebski, A. E., Hautphenne, S., Moss, R., Flegg, J. A., and Flegg, M. B., *A hybrid framework for compartmental models enabling simulation-based inference*, arXiv:2405.13239, 2024, doi:10.48550/arXiv.2405.13239, https://arxiv.org/abs/2405.13239.

[12] Daluwatte, C., Yaghouby, F., and Scully, C., *A framework to characterize the performance of early warning index alarm systems for patient monitoring*, MethodsX 6 (2019), pp. 1660–1667, doi:10.1016/j.mex.2019.07.003.

[13] Inaba, H., *Age-structured homogeneous epidemic systems with application to the MSEIR epidemic model*, Journal of Mathematical Biology 54.**1** (Jan. 2007), pp. 101–146, doi:10.1007/s00285-006-0033-y.

[14] Gupta, S., Pandey, S., Rupali, and Rajoria, Y. K., *Review on Deterministic Mathematical Modelling Studies on COVID-19*, Mathematical Modeling and Soft Computing in Epidemiology, Taylor & Francis, 2025, doi:10.1201/9781003596721-58.

[15] Ojo, M. M. et al., *Mathematical analysis of a stochastic delay model for respiratory syncytial virus dynamics*, Scientific Reports 16 (2026), p. 39783, doi:10.1038/s41598-026-39783-y.

[16] Han, X., Jiang, D., et al., *Estimating dynamic transmission rates with a Black-Karasinski process in stochastic SIHR models using particle MCMC*, arXiv:2505.24127, 2025, https://arxiv.org/abs/2505.24127.

[17] Allen, L. J. S., *A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis*, Infectious Disease Modelling 2.**2** (May 2017), pp. 128–142, doi:10.1016/j.idm.2017.03.001.

[18] Castañón-González, J. A., Polanco-González, C., González-González, R., and Carrillo-Ruiz, J. D., *Surveillance system for acute severe infections with epidemic potential based on a deterministic-stochastic model, the StochCum Method*, Cirugía y Cirujanos 88.**6** (2020), pp. 782–788, doi:10.24875/CIRU.20000725.

[19] Colizza, V., Barrat, A., Barthelemy, M., and Vespignani, A., *Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study*, BMC Medicine 5 (2007), p. 34, doi:10.1186/1741-7015-5-34.

[20] He, S., Tang, S., and Rong, L., *A discrete stochastic model of the COVID-19 outbreak: Forecast and control*, Mathematical Biosciences and Engineering 17.**4** (2020), pp. 2792–2804, doi:10.3934/mbe.2020153.

[21] Libotte, G. B., Lobato, F. S., and Platt, G. M., *Identification of an Epidemiological Model to Simulate the COVID-19 Epidemic using Robust Multi-objective Optimization and Stochastic Fractal Search*, arXiv:2006.00289, 2020, https://arxiv.org/abs/2006.00289.

[22] Le Strat, Y. and Carrat, F., *Monitoring epidemiologic surveillance data using hidden Markov models*, Statistics in Medicine 18.**24** (1999), pp. 3463–3478, doi:10.1002/(SICI)1097-0258(19991230)18:24<3463::AID-SIM409>3.0.CO;2-I.

[23] Rihan, F. A., Alsakaji, H. J., and Rajivganthi, C., *Stochastic SIRC epidemic model with time-delay for COVID-19*, Advances in Difference Equations 2020 (2020), p. 502, doi:10.1186/s13662-020-02964-8.

[24] Southall, E., Brett, T. S., Tildesley, M. J., and Dyson, L., *Early warning signals of infectious disease transitions: a review*, Journal of the Royal Society Interface 18.**182** (2021), p. 20210555, doi:10.1098/rsif.2021.0555.

[25] Polanco-González, C., Castañón-González, J. A., Buhse, T., Samaniego-Mendoza, J. L., Arreguín-Nava, R., and Villanueva-Martínez, S., *[Modified overcrowd index in emergency room]*, Gaceta Médica de México 149.**4** (2013), Artículo en español, pp. 417–424.

[26] Marc, A. et al., *Viral Dynamic Models During COVID-19: Are We Ready for the Next Pandemic?*, CPT: Pharmacometrics & Systems Pharmacology 14.**8** (Aug. 2025), pp. 1289–1297, doi:10.1002/psp4.70055.