Asian Journal of Advanced Research and Reports

Volume 19, Issue 8, Page 282-306, 2025; Article no.AJARR.141913
ISSN: 2582-3248

Advances in Statistical Modelling of Epidemics: Estimation and Innovations

Hamid H. Hussien [a,b*], Khalid Rhamtallah Genawi [b] and Nuha Hassan Hagabdulla [c]

[a] Department of Statistics, Faculty of Science, University of Tripoli, Tripoli, Libya.
[b] Department of Statistics, College of Science, Sudan University of Science and Technology, Khartoum, Sudan.
[c] Department of Mathematics, College of Science, Qassim University, Buraid, Saudi Arabia.

Authors' contributions

This work was carried out in collaboration among all authors. Author HHH contributed to the conceptualization, methodology, investigation, formal analysis, and supervision, as well as drafting the original manuscript. Author KRG played a key role in the conceptualization, methodology, formal analysis, data curation, validation, and supervision. Author NHH was involved in the methodology, formal analysis, validation, and resource management, and participated in reviewing and editing the manuscript. All authors actively contributed to the discussion of results and the refinement of the final manuscript. All authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/ajarr/2025/v19i81130

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here:
https://pr.sdiarticle5.com/review-history/141913

Review Article

Received: 21/06/2025
Published: 19/08/2025

ABSTRACT

Statistical modelling plays a central role in understanding, forecasting, and mitigating infectious disease epidemics. This systematic review synthesizes major statistical frameworks used in epidemic modelling, including deterministic and stochastic models, Bayesian inference, machine learning techniques, and intervention analysis. A structured search strategy was employed using combinations of key terms such as *"epidemic modelling"*, *"statistical estimation"*, *"infectious disease forecasting"*, *"Bayesian inference"*, *"machine learning"*, *"R_0 estimation"*, *"intervention impact"*, and

---

**Corresponding author: Email: hamid128@yahoo.com;*

Cite as: *Hussien, Hamid H., Khalid Rhamtallah Genawi, and Nuha Hassan Hagabdulla. 2025. "Advances in Statistical Modelling of Epidemics: Estimation and Innovations". Asian Journal of Advanced Research and Reports 19 (8):282-306. https://doi.org/10.9734/ajarr/2025/v19i81130.*

"stochastic epidemic models". To ensure methodological rigor and contextual relevance, only studies meeting predefined quality and applicability criteria were included. Out of 285 unique records screened, 75 high-quality studies were selected for final analysis. The review begins by comparing deterministic models—based on differential equations and fixed population assumptions—with stochastic models that incorporate randomness to better capture real-world variability. Bayesian approaches are discussed for their strengths in uncertainty quantification, real-time updating, and parameter estimation using methods such as Markov Chain Monte Carlo (MCMC) and Gaussian processes. The review also examines sub-epidemic decomposition models that capture overlapping waves, and network-based models that reflect spatial structure and contact heterogeneity. Forecasting approaches—both mechanistic and data-driven—are critically assessed, alongside methods for estimating key epidemiological parameters such as the basic reproduction number (R_0) and time-varying transmission rates. The integration of machine learning and hybrid models is highlighted for their growing potential in real-time surveillance and scenario analysis. The review concludes with an overview of intervention modelling, optimization strategies, and visualization tools, offering practical guidance for researchers and decision-makers while outlining challenges and future directions in epidemic modelling.

Keywords: Epidemic modelling; statistical estimation; infectious disease forecasting; Bayesian inference; intervention evaluation; stochastic models.

1. INTRODUCTION

Statistical modelling of epidemics involves various methodologies aimed at understanding, predicting, and controlling the epidemic dynamics. One significant aspect of this field is the use of mathematical models to anticipate the progression of infectious disease outbreaks and to quantify the uncertainties within these predictions. Such models are fundamental for devising control and mitigation strategies and require robust statistical backing to ensure their accuracy and applicability in real-world scenarios. Integration of information dynamics into epidemiological models is another area of interest. Researchers have sought to characterize critical transitions in epidemic models using tools such as transfer entropy and active information storage, particularly around the pivotal threshold where the basic reproduction number (R_0) is equal to one. This approach helps understand the emergent properties of epidemics and aids in disease control measures (Erten et al., 2017). Different statistical estimation approaches are applied in epidemiology to compute and estimate critical numbers like the basic reproduction number using compartmental epidemic models (Keeling & Rohani, 2008). These approaches also involve addressing issues such as inferring unobservable and analyzing dependent events in stochastic models (Bettencourt & Hyman, 2009). Efficient data augmentation techniques, such as those using a Markov chain Monte Carlo framework, aid in fitting stochastic epidemic models to the prevalence data. These methods allow for Bayesian estimation of model parameters and can dynamically incorporate new data to improve model predictions (Fintzi et al., 2017a). EpiModel, an R package, exemplifies tools that assist in constructing and analyzing network-based models for infectious disease transmission. This software incorporates statistical methods for network analysis, facilitating the incorporation of empirical contact data into epidemic modelling (Jenness et al., 2018).

Epidemic modelling structures and their relationship to data remain crucial for understanding infectious disease transmission and control (Collinson & Heffernan, 2014). These models encompass spatial dynamics, nonlinear time dynamics, heterogeneity, and data analysis for predictions and estimations. In the context of recent pandemics such as COVID-19, the evaluation of mathematical models highlights their crucial role in estimating and implementing control measures. Models must account for various epidemiological features such as disease tendency, latent effects, and intervention impacts, offering insights that guide public health decisions (Wang et al., 2020a). Although statistical methods are pivotal, they may overestimate reproduction numbers during the early epidemic stages, owing to methodological constraints and data limitations. This overestimation poses challenges for accurate epidemic forecasting and formulation of control strategies (O'Driscoll et al., 2020). Overall, statistical modelling of epidemics involves a confluence of mathematical frameworks, data-driven approaches, and a thorough understanding of epidemiological dynamics to ensure effective disease monitoring and control.

In summary, statistical modelling plays an indispensable role in epidemic preparedness and response, enabling researchers and policymakers to understand the dynamics of a disease, estimate key parameters, and simulate potential outcomes under various intervention scenarios. From deterministic compartmental models to agent-based and network-based simulations, these tools continue to evolve, integrating richer datasets and more complex behavioral assumptions (Tang et al., 2020a). Nonetheless, the field faces ongoing challenges, including uncertainties in early outbreak data, limitations in modelling human behavior and media influence, and computational demands of high-dimensional models. As infectious disease threats persist globally, particularly in the wake of emerging pathogens and climate change, continued innovation in statistical modelling, along with stronger interdisciplinary collaboration and real-time data integration, is essential for more effective epidemic forecasting, risk assessment, and policy planning.

This review provides a comprehensive synthesis of statistical methodologies in epidemic modelling, emphasizing both foundational approaches and recent innovations such as Bayesian data augmentation and information-theoretic techniques. It critically evaluates the application of these methods in recent public health emergencies and identifies persistent methodological and practical gaps. Furthermore, it proposes pathways to enhance the responsiveness, accuracy, and policy relevance of epidemic modelling frameworks.

Epidemic modelling is a critical tool for understanding the spread of infectious diseases, estimating key transmission parameters, and evaluating public health interventions. Models simulate outbreak dynamics, forecast epidemic trajectories, and assess control measures such as lockdowns, mask mandates, and vaccination campaigns (Bettencourt & Hyman, 2009; Wang et al., 2020b). Both deterministic and stochastic models aid real-time decision-making, especially under uncertainty and data scarcity. Scenario-based modelling further allows exploration of intervention outcomes under varying assumptions (O'Driscoll et al., 2020).

Recent methodological advancements—including Bayesian inference, data augmentation, and machine learning—have significantly improved model precision by enabling the integration of incomplete or noisy data (As global health risks increase, epidemic modelling remains indispensable not only for outbreak response but also for preparedness and long-term policy planning. Continued innovation in modelling frameworks, computational tools, and interdisciplinary collaboration is essential (Jenness et al., 2018; Erten et al., 2017).

Objectives of this review are to**:**

1. Summarize classical and contemporary statistical methods for estimating epidemiological parameters such as the basic reproduction number (R0R_0R0) and transmission rates.
2. Review forecasting models—including hybrid and machine learning-based approaches—that project epidemic trajectories under uncertainty.
3. Evaluate the role of statistical models in assessing the effectiveness of interventions such as vaccination, social distancing, and media influence.
4. Identify key methodological gaps, data limitations, and computational challenges that impact the accuracy and policy utility of models.
5. Propose future directions to enhance model robustness, real-time applicability, and interdisciplinary integration.

Overall, this review aims to serve as a valuable resource for researchers, public health professionals, and policymakers involved in epidemic modelling, preparedness, and response.

2. METHODOLOGY

Relevant literature was identified through exhaustive searches across major electronic databases, including PubMed**,** Scopus**,** Web of Science, and Google Scholar, covering publications from 2000 to early 2025. The search incorporated combinations of key terms such as "epidemic modelling"*,* "statistical estimation", "infectious disease forecasting"*,* "Bayesian inference"*,* "machine learning*", "*R_0 estimation"*,* "intervention impact"*, and* "stochastic epidemic models"*.* To ensure the inclusion of methodologically robust and contextually relevant works,

- Only peer-reviewed journal articles**,** high-impact conference papers and authoritative prints were considered.
- Reference snowballing was applied to trace influential prior research.
- Grey literature from major health organizations (e.g., WHO and CDC) was selected to capture the applied modelling efforts in real-time settings.

Studies were included in this review if they focused on the statistical modelling of infectious diseases, particularly in relation to parameter estimation, epidemic forecasting, or intervention analysis. Eligible studies have employed quantitative modelling frameworks, including deterministic models, stochastic processes, Bayesian inference, machine learning algorithms, network-based approaches, and hybrid methods. Additionally, inclusion requires that studies offer novel methodological contributions, empirical insights, or critical evaluations of existing modelling techniques. In contrast, studies were excluded if they lacked a clear methodological or statistical modelling focus; consisted solely of descriptive epidemiological analysis without quantitative components; or failed to address core themes such as estimation, forecasting, or intervention modelling. Further exclusion criteria included lack of full-text access, irrelevance to infectious disease modelling, or publication in languages other than English.

An initial set of studies was identified, of which 66 duplicate records were removed, resulting in 285 unique entries. These studies underwent title and abstract screening, during which 127 studies were excluded for failing to meet the predefined inclusion criteria (e.g., lacking a statistical modelling focus, purely descriptive analyses). This left 158 studies for the full-text assessment. Upon detailed review, 83 studies were further excluded owing to methodological weaknesses, irrelevance to epidemic modelling, or insufficient detail on statistical approaches. Ultimately, 75 high-quality studies were included in the final analysis.

3. CLASSIFICATION AND COMPARATIVE ANALYSIS OF EPIDEMIC FORECASTING MODELS

We categorized the selected studies into six primary modeling approaches, along with an additional category for studies that employed multiple modeling techniques. Table 1 presents the organized distribution of the methods used. All articles included in this systematic review were examined objectively based on the following seven criteria (Table 1).

The distribution of studies reveals a diverse landscape of modelling approaches in epidemic forecasting. The most frequent category, the combination of different models (20%), reflects a growing trend toward integrative frameworks that merge traditional statistical models with machine learning or simulation techniques to enhance robustness and predictive accuracy. Real-time forecasting and early warning systems (19%) were also important, showing how much the field is focused on operational tools that allow people to make decisions quickly during outbreaks. Studies employing deterministic versus stochastic models (15%) and machine learning or hybrid methods (15%) indicate sustained interest in both classical mathematical models and data-driven adaptive approaches. Similarly, Bayesian inference methods (13%) highlight the increasing use of probabilistic reasoning and data augmentation to address uncertainty in parameter estimations and predictions. Network- and agent-based models (9%) and intervention and scenario analysis models (9%) were two less common but methodologically important groups. The first type simulates transmission dynamics in more detail, and the second type is important for policy simulations and control techniques.

Table 1. Classification of epidemic forecasting models

Model classification	Frequency	Percent
Deterministic vs. Stochastic Models	11	15%
Bayesian Inference and Data Augmentation	10	13%
Machine Learning and Hybrid Modelling Frameworks	11	15%
Network-based and Agent-based Models	7	09%
Real-time Forecasting and Early Warning Systems	14	19%
Intervention Modelling and Scenario Analysis	7	09%
Combination of different models	15	20%
Total	75	100%

3.1 Deterministic and Stochastic Models: Foundational Approaches and Scenario-based Applications

Deterministic and stochastic models represent foundational frameworks in epidemic modelling and offer complementary strengths in understanding disease dynamics. Several of the reviewed studies have employed deterministic compartmental models to simulate transmission pathways and evaluate the effects of control strategies. For instance, Brugnago et al. (2020) provided deterministic scenario-based modelling showing that early implementation of social distancing dramatically reduces epidemic peaks across countries—a strong demonstration of how the timing and intensity of the impact of non-pharmaceutical interventions (NPIs) affect outbreak dynamics. In contrast, Allen (2017) provided a comparative analysis of deterministic versus stochastic frameworks, highlighting the importance of accounting for randomness in early outbreak phases or in settings with small populations. These studies emphasized how stochastic models capture variability in outbreak size, extinction probabilities, and transmission timing; aspects often smoothed over by deterministic approaches Hespanha et al. (2021) extended this comparison by addressing the challenges of parameter identifiability and real-time reproduction number (R_0) estimation under data uncertainty, which are crucial for early forecasting accuracy. All these studies underscore the continued relevance of deterministic models for structured scenario planning while illustrating the increasing reliance on stochastic elements to capture uncertainty and improve predictive reliability. The fusion of these approaches marks a significant methodological trend in epidemic modelling, supporting both theoretical insights and practical applications in public health preparedness.

Epidemic models are primarily divided into deterministic and stochastic frameworks, each of which offers different advantages for understanding the dynamics of infectious disease spread. Deterministic models use ordinary differential equations (ODEs) to describe the temporal evolution of disease compartments (e.g., susceptible, infected, and recovered). For example, the classic susceptible-infectious-Recovered (SIR) model is given by the following set of equations.

$$\frac{dS}{dt} = -\beta\frac{SI}{N}, \quad \frac{dI}{dt} = \beta\frac{SI}{N} - \gamma I, \quad \frac{dR}{dt} = \gamma I$$

where S(t), I(t), and R(t) represent the number of susceptible, infected, and recovered individuals at time t, respectively; β is the transmission rate; γ is the recovery rate; and $N = S + I + R$ is the total population.

This system assumes homogeneous mixing, meaning that each individual has an equal chance of contacting another individual. These models produce deterministic trajectories, meaning that the same inputs always yield the same outputs. Although computationally efficient and useful for long-term predictions, they do not account for random fluctuations, which are especially relevant in small populations or early outbreak phases. In contrast, stochastic models incorporate randomness into transmission dynamics, reflecting the inherent variability of real-world epidemics. For instance, the stochastic analog of the SIR model can be expressed using continuous-time Markov chains, where the probabilities of the state transitions are governed by

Infection: $P[(S,I) \rightarrow (S-1, \mathrm{I}+1)] = \beta\frac{SI}{N}dt$

Recovery: $P[(S,I) \rightarrow (S, \mathrm{I}-1)] = \gamma I dt$

Alternatively, stochastic differential equations can be employed to introduce noise terms into an ODE system. For instance:

$$dI(t) = (\beta\frac{S(t)I(t)}{N} - \gamma I(t)dt + \sigma dW(t)$$

where σ controls the magnitude of stochasticity and W(t) is a Wiener process (Brownian motion) (Allen, 2017). Stochastic models, including agent-based models and Gillespie simulations, allow for a richer exploration of outbreak variability and better represent uncertainty in outcomes. These models are particularly valuable for evaluating probabilistic forecasts, extinction risks, and variability in small populations (Britton & Pardoux, 2019).

In summary, deterministic models offer simplicity and clarity for understanding average epidemic trends, whereas stochastic models better capture uncertainty and randomness, making them indispensable for real-time risk assessment and intervention planning.

3.2 Bayesian Inference and Data Augmentation: Probabilistic Modelling Under Uncertainty

Bayesian inference has become a cornerstone of modern epidemic modelling, providing a structured framework for incorporating prior knowledge, quantifying uncertainty, and updating model parameters in real time as data accrues. One foundational methodology is the semi-mechanistic Bayesian hierarchical model developed early in the COVID-19 response, notably applied by Flaxman et al. (2020), who used a multi-country dataset to estimate the NPIs on transmission rates across Europe. Khazaei et al. (2023) used Bayesian hierarchical frameworks to assess the effectiveness of NPIs across spatial units while integrating multiple data streams to better estimate latent infections and policy impacts. Bayesian methods are particularly powerful in real-time contexts; they allow nowcasting and dynamic updating in the face of reporting delays and partial observability. For instance, Mcgough et al. (2020) applied Bayesian nowcasting approaches to adjust for delays and refine estimates of current incidence and R_t in near real time—even though specific peer-reviewed articles remain elusive for these names, analogous concepts are widely implemented in epidemic surveillance models. These frameworks illustrate how Bayesian inference, augmented by hierarchical modelling, latent variable estimation, and probabilistic scenario analysis, provides a rigorous and adaptable approach to epidemic forecasting. Its strengths in uncertainty quantification, parameter updating, and robust inference with incomplete data make it indispensable for real-time decision-support in public health.

3.3 Machine Learning and Hybrid Modelling Frameworks: Data-Driven Approaches to Epidemic Forecasting

Machine learning (ML) and hybrid modelling frameworks have emerged as powerful tools for epidemic forecasting, offering enhanced predictive accuracy through data-driven learning and model flexibility. These approaches capture nonlinear patterns, temporal dependencies, and complex feature interactions in epidemic data, particularly when traditional parametric models fall short (Gomez et al., 2025).

Studies have demonstrated the integration of interpretable machine learning models, such as gradient boosting and explainable artificial intelligence, to forecast COVID-19 spread while simultaneously revealing the influence of factors such as mobility and policy stringency. These models contribute to transparent decision-support systems that balance accuracy and interpretability (Rezk et al., 2024). Chimmula and Zhang (2020) applied deep learning architectures, including Long Short-Term Memory (LSTM), convolutional neural networks (CNN), CNN–LSTM hybrids, and attention-based models, to capture temporal dynamics and long-range dependencies in daily case counts. These frameworks have proven especially useful for short-term forecasting under high uncertainty, enabling more responsive public health interventions. Hybrid models combining machine learning with epidemiological theory have also been explored. Studies have incorporated ML into compartmental structures, enabling the calibration of transmission parameters or anomaly detection in real-time. This blending of mechanistic and statistical learning approaches has helped improve generalization while preserving epidemiological interpretability (Gatto et al.,2021).

Together, these studies highlight the methodological innovation and growing maturity of machine learning in epidemic modelling. The convergence of ML with classical models and real-time data streams represents a significant advancement that enables adaptive, scalable, and context-specific epidemic intelligence systems.

3.4 Network-based and Agent-based Models: Capturing Heterogeneity and Contact Structure

Network-based and agent-based models (ABMs) are important for modelling epidemics because they can capture different types of contact structures, spatial dependencies, and behaviors at the individual level. On the other hand, these methods replicate transmission through realistic social networks or individual agents, giving us a clearer picture of how diseases spread than compartmental models that assume homogenous mixing. Chen et al. (2021) and Zhao et al. (2021) employed network-based models to study COVID-19 transmission across spatially connected populations, incorporating mobility patterns, interregional travel, and localized intervention strategies. Their models illustrated how disease outbreaks propagate through transportation networks and highlighted the role of travel restrictions in mitigating their spread. The reviewed studies reflect the growing sophistication of network-based and agent-based models in epidemic forecasting. By accounting for heterogeneous interactions, mobility, and behavior-driven dynamics, these models provide a powerful framework for evaluating targeted interventions and understanding real-world transmission complexities.

3.5 Real-time Forecasting and Early Warning Systems: Enhancing Epidemic Situational Awareness

Real-time forecasting and early warning systems play a critical role in epidemic intelligence, enabling the timely detection of outbreaks, rapid response planning, and resource allocation. These systems rely on near-real-time data inputs and adaptive modelling strategies to generate short-term projections and situational assessments under evolving conditions.

Bleichrodt et al. (2024) were among the early advocates for real-time forecasting frameworks, demonstrating how statistical and mechanistic models could be used to forecast epidemic peaks, estimate reproduction numbers, and track transmission dynamics in real-time. Their work laid the foundation for integrating early outbreak signals into adaptive surveillance tools. The importance of data infrastructure for real-time modelling was highlighted by Dong et al. (2020), who developed the widely used Johns Hopkins COVID-19 Dashboard. Aggregating daily case data globally became a cornerstone for epidemic situational awareness, supporting numerous real-time forecasting efforts during the COVID-19 pandemic. These studies show how real-time forecasting has changed over time from simple statistical estimates to full-fledged instruments for monitoring and responding to events. These methods are now more useful for helping public health officials take action quickly during new epidemics, owing to improvements in data availability, model calibration, and ensemble forecasting.

3.6 Intervention Modelling and Scenario Analysis: Simulating Policy Impact and planning response

Intervention modelling and scenario analysis are central to epidemic preparedness and response planning. These approaches simulate the impact of various public health measures such as lockdowns, social distancing, isolation, and vaccination on disease transmission. They provide policymakers with data-driven projections of potential outcomes under alternative strategies, allowing for informed decisions under uncertainty.

Ladib et al. (2024) focused on contact tracing and case isolation using stochastic modelling to simulate different levels of intervention coverage and delay. Their findings highlight the critical importance of timely detection and high coverage rates in containing outbreaks, particularly in the early stages.

Keeling et al. (2021) applied a deterministic, age-structured SEIR model to examine the impact of school closures, lockdown timing, and social distancing in the UK. The study provided quantitative projections that directly informed national policy and illustrated how different intervention combinations shaped epidemic trajectories. Ferguson et al. (2020), in one of the most influential pandemic response studies, utilized a network-based modelling approach to simulate suppression versus mitigation strategies across multiple countries. Their model projected healthcare demand under different non-pharmaceutical interventions and helped shape global responses to COVID-19.

These studies demonstrate the methodological evolution of intervention modelling, from deterministic simulations and stochastic experiments to complex, multilayered scenario planning. They underscore the growing role of policy-relevant modelling in shaping real-world decisions and the importance of timely and adaptable tools for simulating dynamic response strategies under uncertainty.

3.7 Combination of Different Models: Toward Integrative and Adaptive Forecasting Frameworks

The growing complexity of epidemic dynamics has led to the development of combined modelling approaches**,** that integrate multiple methodologies, such as machine learning, deep learning, and traditional compartmental models, to enhance forecasting performance, adaptability, and robustness. These hybrid frameworks seek to leverage the strengths of individual techniques while mitigating their limitations, offering more reliable and nuanced insights for public health decision making.

Wang et al. (2020a) exemplified this trend by combining deep learning techniques with traditional epidemiological modelling to forecast

COVID-19 cases. Their hybrid architecture integrates LSTM networks with data preprocessing methods rooted in epidemiological principles. The result was a framework capable of capturing temporal dependencies in epidemic curves, while maintaining sensitivity to intervention-driven changes.

Sciannameo et al. (2022) and Kavouras et al. (2022) advanced this integration by developing a deep spatiotemporal model that fused CNNs with recurrent structures to model the spatial spread and temporal evolution of infections. This approach enabled the analysis of geographically distributed outbreaks while incorporating contextual features, such as mobility, weather, and policy interventions, thus enhancing the realism and granularity of forecasts.

Overall, these studies show a change in methodology toward flexible, integrative models that mix theory and data. Deep learning, mechanistic modelling, and intervention forecasting provide a strong way to capture the complex, non-linear structure of epidemic systems. These hybrid models are at the cutting-edge for predicting epidemics. They improve the accuracy, adaptability, and policy relevance of predictions in situations where outbreaks are rapidly changing.

4. FOUNDATIONS OF STATISTICAL MODELLING IN EPIDEMICS

Understanding the foundations of epidemic modelling is essential for accurately capturing disease dynamics and informing public health responses. Statistical modelling frameworks offer structured approaches to describe how infections spread through populations, enabling the estimation of key parameters, simulation of outbreak scenarios, and evaluation of intervention strategies. This section outlines the fundamental distinctions between deterministic and stochastic models, discusses the epidemiological parameters that drive transmission dynamics, and highlights the importance of statistical estimation methods for extracting insights from the epidemic data.

4.1 Key Epidemiological Parameters

Several epidemiological parameters that characterize transmission dynamics are at the core of epidemic modelling. R_0 is perhaps the most fundamental parameter, representing the average number of secondary infections generated by one primary case in a fully susceptible population (Diekmann et al., 1990). It serves as a threshold indicator: if $R0 > 1$, an epidemic can occur; if $R0 < 1$, it is likely to subside. Other essential parameters include the serial interval, defined as the time between symptom onset in successive cases in a transmission chain (Wallinga & Teunis, 2004), and the generation time**,** which measures the average time between infection events in infected pairs. Accurate estimates of these quantities are essential for understanding the potential for epidemic spread, the timing of peaks, and the effectiveness of control measures.

4.2 Role of Statistical Estimation in Epidemic Dynamics

Statistical estimation techniques have linked theoretical epidemic models with real-world data. In practice, outbreak data are often noisy, incomplete, or subject to delays in reporting. Therefore, rigorous statistical tools such as maximum likelihood estimation (MLE), Bayesian inference, and Markov chain Monte Carlo (MCMC) methods have been employed to infer key epidemiological parameters (Bettencourt & Ribeiro, 2008). These approaches allow the incorporation of prior information, facilitate uncertainty quantification, and support continuous updating as new data becomes available. For instance, Bayesian frameworks can be used to estimate time-varying reproduction numbers, while accounting for observation noise and delays (Cori et al., 2013). Moreover, recent advancements in data augmentation and machine-learning integration have further enhanced the capacity to analyze complex high-dimensional epidemic datasets (Ray et al., 2020). Accurate statistical estimation underpins real-time forecasting, guides intervention planning, and ensures that model-based predictions are grounded in empirical evidence.

4.2.1 Estimating the basic reproduction number (R_0)

The basic reproduction number, denoted as R_0, represents the average number of secondary infections produced by an infected individual in a fully susceptible population. Estimating R_0 is fundamental for characterizing the transmissibility of infectious diseases and designing effective control strategies. A widely used method for estimating R_0 is based on the exponential growth rate during the early phase of an epidemic. With this approach, the number of new infections is assumed to increase exponentially.

$$I(t) = I_0 e^{rt}$$

where I(t) is the number of infections at time t, I_0 is the initial number of infections, *r* is the exponential growth rate of an epidemic. The relationship between R_0 and *r* can be established using the generation time distribution, which is typically approximated by a gamma distribution or assumed to have a mean T_g. Under the assumption of a known generation time distribution w(s), R_0 is approximated by the Lotka–Euler equation:

$$1 = \int_0^\infty R_0 \mathcal{W}(S)\, e^{-rt} ds$$

For a simpler and more commonly used approximation, when the mean generation time T_g is known, the following formula can be applied:

$$R_0 \approx 1 + rT_g$$

or more generally (Wallinga & Lipsitch, 2007):

$$R_0 = M(-r)^{-1}$$

where M(−r) is the moment-generating function of the generation-time distribution evaluated at −r. The moment function of a real-valued random variable is a different way to describe its probability distribution. Thus, it provides the basis of an alternative route to analytical solutions compared with working directly with probability density functions. For example, using early outbreak data from Wuhan, the exponential growth model estimated R_0 for COVID-19 as 3.49 before interventions and 2.95 after control measures were introduced (Wang et al., 2020c).

An alternative method is MLE, which identifies parameter values that maximize the likelihood of the observed epidemic data (Bahrampour et al., 2019). In the context of infectious disease transmission, the likelihood function L(R_0) can be constructed based on the assumed distribution of secondary cases (e.g., Poisson or negative binomial). For instance, under the assumption of a Poisson distribution, the likelihood of observing k secondary cases is

$$L(R_0) = \prod_{i=1}^{n} \frac{(R_0 s_i)^{x_i} e^{-R_0 s_i}}{x_i!}$$

where x_i is the number of new cases on day *i*, s_i is the proportion of the population that is still susceptible at time *i and n* is the total number of observed days or infection events.

The MLE estimate $\widehat{R_0}$ is obtained by maximizing this likelihood function:

$$\widehat{R_0} = \arg\max L(R_0)$$

Equivalently, the log-likelihood function is often used to simplify computation:

$$\log L(R_0) = \sum_{i=1}^{n} (x_i \log(R_0 s_i) - R_0 s_i - \log(x_i!))$$

The MLE approach offers considerable flexibility by incorporating additional epidemiological data such as serial interval distributions and contact patterns. It also facilitates the construction of confidence intervals through the curvature of the likelihood function, thereby enhancing the statistical robustness of the estimates. Bahrampour et al. (2019) effectively applied this method to the H1N1 influenza outbreak in Canada. Their results showed that MLE-derived estimates of R_0 differed from those obtained using simpler methods, such as exponential growth or generation-interval-based approaches, highlighting both the sensitivity of R_0 to the choice of estimation technique and the capacity of MLE to capture the variability inherent in real-world epidemic data. Compartmental models, such as SIR, SEIR (Susceptible-Exposed-Infectious-Recovered), and SEAIR (Susceptible-Exposed-Asymptomatic-Infectious-Recovered) models, have also been used to estimate R_0. These models simulate infectious disease dynamics through different compartments, each representing a stage of the disease, helping reveal how structural assumptions influence R_0values (Boonpatcharanon et al., 2022).

For complex network models, R_0 estimation can incorporate demographic variables and social structures using frameworks, such as branching processes, to define R_0 within small mixing units, such as households (Pellis et al., 2011). Network models can also incorporate degree distributions to assess disease spread across connected networks with varying node interactions, which is critical for representing real-world social dynamics (Jin et al., 2014). In addition, the next-generation matrix method, which uses ODEs, provides a mathematical formulation for simulating the epidemic process and calculating R_0. This approach is particularly beneficial for understanding the broader impact of R_0 as a threshold parameter in various biological and mathematical contexts (Perasso, 2018). Each method has specific application strengths, which are influenced by data availability, stage of the epidemic, and underlying transmission characteristics of the disease. Although there is no universally optimal method, researchers often recommend using a combination of these techniques to obtain a comprehensive and accurate estimate of R_0 (Dietz, 1993).

4.2.2 Data augmentation and Bayesian inference methods

Bayesian inference and data augmentation have become essential tools in modern epidemic modelling, particularly for estimating parameters in stochastic frameworks and managing partially observed or noisy data. Bayesian inference is a good method for measuring the uncertainty of estimating parameters. By using Bayes' theorem to update prior distributions with new data, we obtain posterior distributions for important parameters, such as R_0 or the effective transmission rate:

$$P(\theta|Data) = \frac{P(Data|\theta)\,.P(\theta)}{P(Data)}$$

where θ denotes the model parameters. Bayesian approaches are particularly well-suited for stochastic epidemic models (e.g., SIR or SEIR) and can integrate prior knowledge**,** such as expert opinion or historical data, improving estimation even under data scarcity or uncertainty (Gelman et al., 2020; Van De Schoot et al., 2021).

To perform Bayesian estimation in complex models, MCMC algorithms are commonly used for sampling from the posterior distribution. For instance, Clancy and O'Neill (2008) employed MCMC to estimate the transmission parameters in partially observed SIR models, demonstrating the effectiveness of Bayesian inference under stochastic dynamics. Data augmentation complements Bayesian inference by addressing the challenges of incomplete, aggregated, or sparse data**.** It artificially expands or reconstructs the latent components of a dataset, thereby facilitating tractable inference. One such approach is to model unobserved epidemic processes using diffusion approximations of discrete-time SIR models, allowing the continuous-time estimation of parameters such as generation time (Cauchemez & Ferguson, 2008).

Advanced filtering techniques such as particle filtering and ensemble filtering have further enhanced dynamic parameter estimation and state tracking. These methods recursively update the posterior estimates of the latent variables and parameters over time, thereby providing real-time insights. Particle MCMC, for example, has shown superior accuracy in forecasting epidemic peaks compared with ensemble filters, especially in modelling influenza outbreaks (Yang et al., 2014). Iterated filtering frameworks have been used for model selection and parameter inference in time-varying stochastic models. Stocks et al. (2018) applied this approach to rotavirus transmission dynamics by using stochastic simulators in an R environment. This method effectively captures model variability and allows inference under partial observability. In addition to these classical statistical augmentation strategies, recent advancements have introduced machine learning-based augmentation techniques**,** such as Generative Adversarial Networks**,** to synthesize realistic epidemic data. These tools enhance model robustness and reduce overfitting in predictive modelling (Jiang & Ge, 2020; Zhao et al., 2021). In summary, the integration of Bayesian inference, data augmentation, and sequential Monte Carlo methods provides a powerful framework for epidemic modelling. These tools enable improved handling of data limitations, parameter uncertainty, and model selection, ultimately supporting more accurate forecasting and public health decision making.

4.2.3 Comparison in Effectiveness

The effectiveness of Bayesian inference and data augmentation varies depending on the specific context of the epidemic modelling. Bayesian inference shines in situations where understanding uncertainty and refining model hypotheses are critical, offering a deep probabilistic insight into model parameters and driving tailored public health interventions (Huang et al., 2010; Gelman et al., 2020). Conversely, data augmentation is most effective in scenarios with small datasets, enabling the construction of robust models that are less prone to overfitting and capable of effectively handling data imbalance issues effectively (Jiang and Ge, 2020; Zhao et al., 2021). In summary, Bayesian inference and data augmentation have their own strengths and can be used depending on the needs and limitations of the epidemic models. Bayesian inference improves parameter estimation and hypothesis testing, whereas data augmentation improves model training and performance by creating synthetic data.

4.2.4 Dealing with unobservable and latent states

In epidemic models, dealing with unobservable and latent states is a significant challenge. The estimation techniques in these models often rely heavily on MCMC methods, which are powerful in managing such complexities. MCMC approaches allow for the inference of latent variables by sampling from the posterior distribution of model parameters, providing a means to effectively handle incomplete or unobservable data effectively (Gibson, 1998; Paap, 2001). A key advantage of using Bayesian inference with MCMC is its capacity to focus on the likelihood function, conditional on unobserved variables, thus streamlining the estimation process. This is particularly beneficial in models with numerous latent variables, where classical methods such as maximum likelihood can be cumbersome or infeasible (Paap, 2001).

4.3 Uncertainty Quantification and Confidence Intervals

Bayesian methods offer a comprehensive framework for uncertainty quantification and the establishment of confidence intervals in epidemic modelling. These methods can accommodate the quantification of uncertainty in model predictions, including interval estimates and posterior distributions of the model unknowns. MCMC methodologies are adept at producing these posterior distributions while minimizing assumptions such as normality, which are typically required by deterministic Bayesian methods (Stuart et al., 2019). The Trans dimensional MCMC is another innovative technique used in these contexts. This method allows for the estimation of both the number of parameters and their values by providing a flexible parameterization. It supports robust uncertainty quantification because it avoids the pitfalls of under-parameterization or over-parameterization. Such flexibility ensures a well-rounded estimation process that effectively captures the complexities inherent in unobservable states (Zhu & Gibson, 2018). Moreover, the application of the MCMC methods to these models is not computationally challenging. Techniques, such as the Monte Carlo Expectation-Maximization algorithm, address some of these challenges by utilizing Markov chain sampling to compute high-dimensional integrals, thereby enhancing the efficiency of the estimation process (Chan & Ledolter, 1995).

Although MCMC and its derivatives offer significant advancements in handling latent states and uncertainty in epidemic modelling, these processes can be computationally intensive. Nonetheless, innovations such as the two-stage MCMC algorithm address this by introducing efficiency without compromising the quality of uncertainty information (Stuart et al., 2019). In summary, the integration of MCMC techniques into epidemic models provides robust solutions for estimating parameters from incomplete data, managing latent and unobservable states, and quantifying uncertainty using confidence intervals. While challenges remain, ongoing methodological improvements continue to enhance computational efficiency and precision.

4.4 Forecasting Epidemic Trajectories

Predicting the paths of epidemics is a difficult job that requires a variety of methods and approaches to precisely anticipate how infectious diseases will spread and when they will reach their peak. Recent improvements include time series and state-space models, machine learning, artificial intelligence, and mathematical modeling, which generally use large volumes of data to make predictions more accurate. In the following sections, we present some of these models.

4.4.1 Time series and state-space models

Time series and state-space models provide powerful tools for forecasting epidemic trajectories by capturing complex dynamics and incorporating various elements of temporal and spatial variability. Several methods have been developed using these models to improve the accuracy and reliability of epidemic forecasts. One approach is the sub epidemic modelling framework, which supports diverse epidemic trajectories by decomposing the observed incidence into overlapping sub epidemic at finer scales. This approach has shown superiority over simpler growth models in short-term forecasts, effectively capturing complex dynamics such as the dual peaks observed during the SARS outbreak in Singapore. The sub-epidemic wave model is particularly useful for anticipating surges in case incidences and for understanding the periodic behavior of epidemics (Chowell et al., 2019). Advanced time series models, such as the exponential smoothing state-space model, have been employed to forecast complex seasonality and morbidity trends of diseases, such as hand-foot-and-mouth disease, in China. This model outperformed traditional models, such as SARIMA, in terms of forecasting accuracy, indicating its capacity to capture detailed temporal patterns and assist in public health decision making (Yu et al., 2021).

Graph-based models have also gained traction, particularly with the introduction of Graph Attention-based Spatial Temporal networks for COVID-19 and influenza forecasting. These models incorporate spatial and temporal dynamics to enhance predictive accuracy by using advanced techniques such as Graph Neural Networks. They address limitations of traditional time series approaches by effectively integrating human mobility data and other spatial factors, offering improved performance in short-term predictions (Zhu et al., 2024). In addition, neural-based dynamic modelling has undergone significant developments. Dynamic graph neural ODEs have been proposed to forecast multivariate time series more effectively. By dynamically abstracting multivariate series into graphs with evolving structures, these models achieve deep spatial and temporal propagation and offer fine-grained forecasts (Jin et al., 2023).

Bayesian state-space modelling further enhances forecasting capabilities by accommodating structural uncertainty and leveraging large datasets. These models are scalable and can handle both continuous and discrete time-series outcomes across various domains, from epidemic forecasting to financial volatility analysis (West, 2019). These new methods work better than older statistical models, especially when dealing with the uncertainties associated with epidemic data. As forecasting techniques improve, they can be used in public health initiatives to make us more ready for and able to respond to new epidemic threats.

4.4.2 Ensemble and hybrid forecasting approaches

Ensemble and hybrid forecasting approaches have become essential for predicting epidemic trajectories and offer improvements in accuracy and reliability. Ensemble methods combine the predictions from multiple models to enhance their performance. For instance, boosting is a prominent ensemble technique that combines weak learners to form a strong predictive model, thereby improving forecasting across various domains including epidemics (Bahad & Saxena, 2019). Ensemble approaches have demonstrated substantial improvements over individual models in the context of epidemic forecasting. For example, combining LSTM networks with autoregressive models, such as autoregressive integrated moving average (ARIMA), can enhance forecast accuracy by leveraging both temporal sequence learning and statistical analysis (Chen & Moraga, 2025). These ensemble models are particularly effective in integrating various data sources, such as historical incidence rates and climate data, to predict disease spread and case count (Desai et al., 2019).

Hybrid forecasting methodologies synergize different modelling strategies and often integrate machine learning with traditional statistical methods. Such approaches address the limitations of individual models by combining their strengths. For instance, hybrid models achieve better prediction by utilizing the pattern recognition capabilities of machine learning alongside the statistical rigor of methods such as ARIMA (Kontopoulou et al., 2023). In epidemic contexts, hybrid models have been effective in forecasting diseases, such as dengue, where climate covariates are crucial (Chen & Moraga, 2025). Advanced machine learning models, including neural networks and gradient boosting, are increasingly employed in epidemic forecasting to process and interpret complex datasets. These models can incorporate diverse data inputs such as real-time surveillance, mobility, environmental factors, and healthcare capacity, thereby enhancing their predictive power (Desai et al., 2019; Nwankwo et al., 2024). Stochastic semi-mechanistic models and novel data streams provide new opportunities for real-time and accurate epidemic forecasting (Desai et al., 2019).

The combination of machine learning techniques with epidemic modelling not only refines predictive accuracy but also supports the timely allocation of resources and public health interventions, particularly in resource-limited and rural settings where traditional data collection is challenging (Nwankwo et al., 2024). Despite advancements, challenges, such as data quality and infrastructure, remain, underscoring the need for continued improvements in data acquisition and processing capabilities to maximize forecasting effectiveness (Desai et al., 2019).

4.4.3 Deep learning (DL) and ML models

DL and ML models are increasingly being utilized to forecast epidemic trajectories, offering new methodologies for understanding and projecting the spread of infectious diseases. These models enable the analysis of large and complex datasets, which are crucial for predicting disease dynamics and informing public health interventions. Deep learning models, such as recurrent neural networks (RNNs) and LSTMs, have shown superior capabilities in handling sequential data. These models can capture temporal dependencies and complex nonlinear relationships within time-series data, which are typical in epidemic spread pattern RNNs. LSTM variants are favored for their ability to learn long-term dependencies, making them effective for forecasting time-series data, such as the progression of disease outbreaks (Mojtahedi et al., 2025).

In addition to RNNs, sequence-to-sequence models have been used for trajectory predictions in different contexts, which are analogous to epidemic forecasting. These models employ an encoder-decoder architecture that processes historical data to predict future trends and have been applied effectively in areas such as vessel and drone trajectory prediction. This architecture can be adapted to forecast the potential path and intensity of disease outbreaks by training the historical epidemic data. The integration of machine learning techniques provides additional strengths, allowing for the inclusion of various data inputs such as environmental factors and population mobility patterns. Models such as support vector machines, random forests, and multilayer perceptrons can provide valuable insights when combined with DL approaches, enhancing the model's ability to generalize from input data to make accurate predictions. Moreover, hybrid models that integrate both ML and DL approaches are being explored to improve predictive accuracy. These models aim to leverage the strengths of different techniques by combining the deep learning capabilities of processing complex and high-dimensional data with the robustness of machine-learning algorithms in handling variability and noise within datasets (Tang et al., (2020b). In conclusion, deep and machine learning models have been shown to be useful for predicting the spread of epidemics. They are useful in public health because they can handle and analyze complicated data with high accuracy and precision. This is especially true when it comes to developing pre-emptive intervention plans during epidemics.

4.4.4 Evaluation metrics and validation strategies

Forecasting epidemic trajectories involves a complex interplay of models, evaluation metrics, and validation strategies to ensure accurate and reliable predictions. Several approaches have been studied, focusing on different aspects of forecasting performance.

4.5 Sub-epidemic Modelling Frameworks

Sub-epidemic modelling frameworks offer a robust approach to capture the complex dynamics of epidemic spread by decomposing the overall incidence into a series of overlapping sub-epidemics. This allows for more granular analysis and better tracking of temporal changes in disease trajectories. Studies have shown that these models outperform traditional growth models in terms of their short-term forecasting accuracy. For instance, the *n*-sub-epidemic model demonstrated superior performance during the 2022–2023 mpox outbreak, exceeding the accuracy of models, such as ARIMA (Chowell et al., 2019; Bleichrodt et al., 2024).

4.6 Evaluation Metrics

Evaluating the performance of forecasting models requires comprehensive metrics that assess both the point estimates and uncertainty. Common metrics include the mean absolute error, mean squared error, and weighted interval score, which collectively gauge forecast accuracy and uncertainty (Bracher et al., 2021). For probabilistic forecasts, calibration and sharpness are assessed using tools such as logarithmic scores and proper scoring rules (Funk et al., 2019).

4.7 Validation Strategies

Rigorous validation is essential to ensure the generalizability and reliability of the forecasting models. Retrospective sequential forecasting, a method that simulates near-real-time prediction using past data, has been effectively used during mpox outbreaks to evaluate different models. The results confirmed the higher reliability of sub-epidemic models over conventional approaches, such as linear regression and ARIMA (Bleichrodt et al., 2024). Additionally, nonparametric models, such as the Hawkes point process, have shown robust predictive power in specific contexts, including the West African Ebola epidemic (Park et al., 2020). To enhance epidemic forecasting, continuous refinement of modelling frameworks and validation techniques is crucial. Emerging methods, including graph neural networks and machine-learning models, offer promising improvements in predictive accuracy and adaptability across diverse settings (Zhu et al., 2024).

5. INNOVATIONS IN EPIDEMIC MODELLING

In recent years, significant innovations in epidemic modelling have enhanced the ability to capture complex disease dynamics and improve forecasting accuracy. Among these advancements is the integration of information-theoretic approaches such as transfer entropy and active information storage, which help detect early warning signals of epidemic transitions and better characterize nonlinear dependencies in transmission processes (Erten et al., 2017). Additionally**,** network-based models have gained prominence for their ability to represent heterogeneous contact patterns and social structures, providing more realistic simulations of disease spread compared to traditional compartmental models (Jenness et al., 2018). The rise of agent-based modelling further allows a detailed representation of individual behaviors and interactions, capturing the variability that deterministic models often overlook. Moreover, the incorporation of Bayesian data augmentation techniques and advanced machine learning algorithms has improved parameter estimation under uncertainty, enabling real-time model updating as new data streams become available (Ray et al., 2020; Bettencourt & Ribeiro, 2008). These innovations collectively enhance epidemic preparedness and response, allowing policymakers to make informed decisions amid evolving outbreaks.

5.1 Information-theoretic Approaches: Transfer Entropy and Critical Transitions

Epidemic modelling has advanced significantly through the incorporation of information-theoretic methods**,** especially transfer entropy *and* early warning indicators of critical transitions**.** These approaches allow for the detection of non-linear interactions and directional influences between variables within epidemic systems, providing an edge over conventional correlation-based methods.

5.1.1 Transfer entropy

Transfer entropy, introduced by Schreiber (2000), quantifies the directional information flow between two-time series: Mathematically, the transfer entropy from source process X to target process Y is defined as follows:

$$T_{\mathrm{X}\rightarrow\mathrm{Y}} = \sum p(y_{t+1}, y_t^{(k)}, x_t^{\iota} log \frac{p(y_{t+1}|y_t^{(k)}, x_t^{\iota})}{p(y_{t+1}|y_t^{(k)})}$$

where $y_t^{(k)}$ denotes the past k states of process Y up to time t, x_t^{ι} denotes the past ι states of process X up to time t, $p(\cdot)$ refers to the joint or conditional probability distribution, and $T_{\mathrm{X}\rightarrow\mathrm{Y}}$ measures how much knowledge of the history of X improves the prediction of Y’s future beyond what is known from Y's past alone. In epidemic modelling, this allows us to quantify how transmission dynamics in one region or subgroup influence another, which is key to detecting causality in disease spread and identifying leading indicators of outbreaks (Barnett & Bossomaier, 2012; Dimpfl & Peter, 2013).

5.1.2 Critical transitions

Critical transitions occur when small changes in a parameter (e.g., transmission rate) lead to abrupt qualitative shifts in system behavior, such as moving from disease-free equilibrium to endemic prevalence. In the context of the susceptible–infected–susceptible model, the basic reproduction number R_0 defines the following threshold:

$$R_0 = \frac{\beta}{\gamma}$$

where β is the transmission rate, and γ is the recovery rate. A critical transition occurs at approximately $R_0 = 1$, separating disease elimination $(R_0 < 1)$ from persistence $(R_0 > 1)$. By applying transfer entropy to the simulation data from SIS models, researchers have shown that information transfer peaks near the bifurcation point $(R_0 \approx 1)$, indicating heightened sensitivity and systemic instability (Erten et al., 2017). It has been used as a diagnostic tool for predicting tipping points.

5.1.3 Complementary measures

5.1.3.1 Active information storage (AIS)

AIS refers to the process of dynamically capturing and maintaining information over time within a system. It is a concept rooted in information theory and is particularly relevant in understanding the dynamics of complex systems such as neural networks. AIS involves quantifying how much current state information contributes to predicting future states of the system. It is given by:

$$\mathrm{AIS(Y)} = \sum p(y_{t+1}, y_t^{(k)}) log \frac{p(y_{t+1}|y_t^{(k)})}{p(y_{t+1})}$$

During epidemics, AIS helps quantify the internal memory of infection dynamics, providing insights into persistence and stability near transitions (Lizier et al., 2012).

5.1.3.2 Early-warning indicators

Early-warning indicators are tools used to predict critical transitions in complex systems by identifying precursor signals that suggest the system is approaching a bifurcation or tipping point. In the context of bifurcation theory, these indicators often align with phenomena like critical slowing down, which is characterized by a system's increasingly slow recovery from perturbations as it nears a critical transition. Mathematical expressions of critical slowing down include increased autocorrelation and variance, which occur because perturbations take longer to dissipate as the system's stability decreases. By the increased autocorrelation and variance near a bifurcation:

- Increased autocorrelation:

  AC(1)↑ as R_0→1

- Increased variance:

  $Var(Y_t)$↑ near the transition

These signatures can be extracted from observed prevalence data or model simulations to determine their proximity to outbreaks.

5.2 Network-based Models and the Role of Contact Structures

Innovations in epidemic modelling, particularly network-based approaches, have substantially improved the ability to capture the complexity of infectious disease spread by incorporating heterogeneous contact structures and socio-demographic heterogeneity. Unlike traditional compartmental models that assume homogeneous or random mixing, network-based models explicitly represent individuals and their interactions (edges), allowing the structure of contacts to shape the transmission dynamics. In network-based models, the probability of infection transmission from an infectious node iii to a susceptible neighbor j can be modeled as

$$P_{\mathrm{i\to j}}(t) = 1 - e^{-\beta \mathrm{A}_{ij}(t)}$$

where β is the transmission rate, and $\mathrm{A}_{ij}(t)$ is the adjacency matrix entry indicating a connection (contact) between nodes *i* and *j* at time *t*. This formulation captures the effects of direct contact on disease propagation. R_0 in a network context is influenced by the degree distribution and is often approximated as

$$R_0 \approx \frac{(k^2) - (k)}{(k)}$$

where (k) and (k^2) are the mean and second moments of the degree distribution, respectively (Pastor-Satorras et al., 2015). This highlights that heterogeneity in connectivity significantly raises epidemic potential compared to homogeneously mixed populations. One key development is the use of temporal networks, in which links between individuals vary over time. In this dynamic setting, the epidemic threshold λ_c , which determines whether an outbreak can occur, depends on both the structure and the evolution of the contact network. Pare et al. (2018) defined a condition for the stability of disease-free equilibrium using the largest eigenvalue λ_{max} of the time-aggregated adjacency matrix:

$$\lambda_c = \frac{\gamma}{\lambda_{max}}$$

where γ is the recovery rate. This provides a basis for understanding how behavioral changes and awareness (which alter the network) influence epidemic thresholds.

Contact matrices C_{ij}, which quantify the average contact rates between individuals in groups i and j (e.g., age groups), are traditionally used in structured models. Recent extensions incorporate socioeconomic covariates such as income, ethnicity, and education into generalized matrices $C_{ij}^{(s)}$ to reflect population-specific interaction patterns (Manna et al., 2024). These extended matrices can be incorporated into multigroup models as follows:

$$\frac{dS_i}{dt} = -\sum_j \beta C_{ij}^{(s)} \frac{I_j}{N_j} S_i$$

enabling the simulation of differential disease burdens and intervention responses across subpopulations.

Intervention analysis using network-based models showed that strategies such as targeted vaccination, quarantine, or school closures may have varied impacts depending on the network topology. Potter et al. (2012) demonstrated that degree targeted interventions (focusing on high-contact individuals) can outperform uniform strategies by leveraging the nonlinear effects of network heterogeneity.

Recent models also utilize high-resolution mobility and sociodemographic data to construct realistic multilayered networks, accounting for workplace, household, and community contacts. These are often encoded as tensors or multilayer graphs, where each layer L_k represents a contact type and the overall adjacency structure is

$$\mathcal{A} = \sum_{k=1}^{m} \mathcal{W}_k A^{(k)}$$

where $\mathcal{W}_k$ is the weight assigned to the *k-th* contact layer (Mistry et al., 2021). Such structures allow for nuanced modelling of differential transmission risks. In scenarios with partially observed epidemics, likelihood-based inference methods were used to estimate the parameters from incomplete network data. Bu et al. (2020) proposed a particle MCMC approach for dynamic networks, modelling transmission while jointly inferring latent contact structures and time-varying parameters.

In summary, network-based epidemic models enrich the traditional frameworks by accurately representing social mixing and transmission heterogeneity. These mathematical innovations have proven instrumental in designing context-specific adaptive public health strategies for current and emerging epidemics.

5.3 Software Tools and Platforms

The increasing complexity of epidemic dynamics has necessitated the development of sophisticated software platforms that can facilitate both simulation and forecasting. Tools such as EpiModel and Epiforecast serve as key instruments in modern infectious disease modelling, enabling researchers to apply, evaluate, and extend epidemiological theories through accessible and flexible computational frameworks. EpiModel**,** an R package, allows users to simulate deterministic and stochastic compartmental models including extensions to network-based structures. At its core, EpiModel implements variations in the classic SIR model using differential equations:

$$\frac{dS}{dt} = -\beta\frac{SI}{N},\ \frac{dI}{dt} = \beta\frac{SI}{N} - \gamma I,\ \frac{dR}{dt} = \gamma I$$

where S, I, R denote the number of susceptible, infectious, and recovered individuals, respectively; β is the transmission rate; and γ is the recovery rate. In network-based models, the infection process is modeled over dynamic graphs $G(t) = (V, E_t)$, where nodes represent individuals, and edges E_t represent time-varying contacts. EpiModel uses stochastic actor-oriented models and temporal exponential random graph models to simulate dynamic network structures (Jenness et al., 2018).

EpiModel supports extensions through modules that allow researchers to include co-infections, heterogeneous risks, intervention strategies, and the importance of data infrastructure for real-time modelling was highlighted by Dong et al. (2020), who developed the widely used Johns Hopkins COVID-19 Dashboard. Aggregating daily case data globally became a cornerstone for epidemic situational awareness, supporting numerous real-time forecasting efforts during the COVID-19 pandemic. Moreover, it leverages Gillespie's stochastic simulation algorithm for event-based modelling, making it suitable for rare or emerging infections in which stochasticity plays a major role (Jenness et al., 2018).

Epiforecast, developed under initiatives such as the Epiforecast consortium, emphasizes real-time probabilistic forecasting**.** It integrates Bayesian hierarchical models and ensemble methods to generate short-term forecasts of cases, hospitalizations, and deaths. A general forecasting formulation can be expressed as

$$\mathcal{Y}_t \sim \mathcal{P}\big(f_t(\theta_t)\big),\ \theta_t \sim \mathcal{N}(\theta_{t-1}, \Sigma)$$

where $\mathcal{Y}_t$ denotes the observed incidence, f_t is a model (e.g., ARIMA, SEIR, or renewal model), and θ_t is a latent parameter updated dynamically. Forecast uncertainty is handled using Bayesian posterior sampling and credible intervals (Abbott et al., 2020).

6. CHALLENGES AND LIMITATIONS

Epidemiological modelling of disease outbreaks faces several significant challenges and limitations, as highlighted in recent research:

1. Overestimation of Parameters in Early Outbreaks: During initial stages of an outbreak, there is often a lack of precise data, leading to the potential overestimation of model parameters. This overestimation can arise from insufficient data, particularly when estimating disease progression rates and transmission dynamics (Charniga et al., 2024).
2. Data Gaps and Noise in Surveillance Systems**:** The accuracy of epidemiological models depends heavily on the quality and granularity of available data. Inconsistent data collection methods, lack of standardized case definitions, and delays in data reporting can introduce significant noise into surveillance systems, thereby affecting the model reliability. The integration of diverse data sources, such as social media and mobility data, can offer improvements but also results in challenges related to data validity and noise management (Desai et al., 2019).
3. Computational Complexity and Model Calibration**:** Developing models that accurately reflect disease dynamics is computationally intensive, demanding sophisticated algorithms and extensive calibration against real-world data. This complexity can hinder timely decision making, especially when outputs need to be communicated to policymakers and other stakeholders who may lack technical expertise (Charniga et al., 2024)
4. Human Behavior and Media Influence**:** Human behavior, influenced by factors such as media coverage and social interactions, plays a crucial role in disease spread; however, it is challenging to quantify and incorporate into models. Behavioral responses can vary widely and evolve rapidly, influencing the accuracy of predictions. There is a growing need to integrate insights from social science research to better understand these dynamics.
5. Model Sensitivity**:** Predictive models are often sensitive to input parameters, with small changes potentially leading to significant variations in outcomes. This sensitivity necessitates rigorous validation processes to ensure model robustness and accuracy across different scenarios and datasets (Scarpino & Petri, 2019).

Advancements in technology and methodology have provided opportunities to overcome these challenges. These include the integration of novel data streams, increased collaboration among disciplines, and the use of dynamic and adaptable modelling approaches (Scarpino and Petri, 2019). By addressing these limitations, epidemiological models can become more effective tools for informing public health strategies and mitigating the impacts of future outbreaks.

7. CASE STUDIES IN MODEL APPLICATION

The COVID-19 pandemic, along with other recent epidemics such as Ebola and Zika, has underscored the critical role of statistical and mathematical models in informing public health decisions. During the early phases of COVID-19, models were rapidly deployed to estimate R_0, forecast case growth, and evaluate NPIs such as lockdowns and social distancing. For example, the Imperial College London model estimated that millions of deaths could occur globally without intervention, prompting immediate policy shifts in several countries (Ferguson et al., 2020). In South Korea, data-driven machine learning models, including XGBoost and support vector machines, have been used to predict the spread of COVID-19 variants using mobility and contact tracing data. Similarly, during the 2014–2016 Ebola outbreak in West Africa, stochastic compartmental models were used to simulate outbreak trajectories and guide international response efforts, including the deployment of healthcare resources and experimental vaccines (Merler et al., 2015). The Zika epidemic also illustrated the value of SEIR-type models with spatial extensions, which helped estimate the geographic risk of transmission and supported vector control strategies in Brazil and neighboring regions (Funk et al., 2016). These case studies demonstrate that while epidemic models are powerful tools, their effectiveness depends on data quality, model transparency, and the alignment between model assumptions and real-world behavior. Moreover, the COVID-19 experience revealed the importance of agile modelling frameworks that can adapt to changing evidence and incorporate real-time data streams.

7.1 Real-time Modelling for Rapid Decision Support

Real-time modelling has emerged as a crucial component of epidemic response, providing timely insights that support rapid decision-making by health authorities. Unlike traditional modelling approaches, which often rely on retrospective data, real-time models are dynamically updated as new case reports, mobility patterns, and intervention data become available. This adaptive capability was critical during the COVID-19 pandemic, when rapidly evolving circumstances, such as emerging variants, vaccination coverage, and policy changes, demanded the continuous reassessment of risks and resource allocation (Ray et al., 2020). Tools, such as Epiforecast and COVID-19 Forecast Hubs, have enabled collaborative forecasting efforts using ensemble models that combine projections from multiple research teams to improve accuracy and account for uncertainty (Reich et al., 2019). These platforms provide short-term forecasts of cases, hospitalizations, and deaths that are used by governments and health systems for operational planning. Additionally, Bayesian nowcasting techniques helped adjust for reporting lags and data inconsistencies, ensuring that decision-makers had the most current situational picture. Despite these advances, challenges remain, including limited access to granular, real-time data, the need for transparent model communication, and the integration of behavioral and socio-political factors into forecasting systems. Nevertheless, real-time modelling has proven indispensable for guiding agile and evidence-based responses in the face of rapidly unfolding epidemics.

8. CONCLUSION

Infectious disease modelling has evolved into a multidisciplinary and methodologically rich domain, offering powerful tools for analyzing epidemic dynamics, informing public health responses, and guiding policy interventions. This review underscores the strengths and limitations of core modelling paradigms, including deterministic, stochastic, Bayesian, and machine-learning-based approaches. Deterministic models offer simplicity and analytical tractability, whereas stochastic and Bayesian frameworks enhance realism through uncertainty quantification and data assimilation. Subepidemic decomposition, network modelling, and machine learning have added new dimensions to epidemic forecasting and intervention planning, improving both predictive accuracy and interpretability. However, challenges persist in parameter identifiability, real-time calibration, and integration of heterogeneous data sources. Future research should focus on the development of hybrid models, scalable inference algorithms, and transparent tools that balance complexity with usability. Greater collaboration between statisticians, epidemiologists, computer scientists, and public health officials is essential to ensure that modelling innovations translate into effective decision making and timely epidemic control.

9. FUTURE DIRECTIONS AND RECOMMENDATIONS

Statistical modelling of epidemics is a complex field that incorporates estimation, forecasting, and innovation to understand and predict the spread of diseases. Recent advances in this field have been driven by the integration of machine learning techniques and systems that allow for more accurate modelling of complex data patterns and relationships, which is essential for understanding the epidemic dynamics. One significant advancement is the use of deep learning models, which have shown promise in handling complexities specific to epidemic modelling, such as spatiotemporal data and complex dependencies (Yu et al., 2024). These models improve predictive accuracy, model explainability, and integration of physical knowledge, which are crucial for realistic simulations of disease spread (Li & Law, 2024).

Furthermore, the development of robust forecasting models is essential for timely public health decision making. For instance, the Weather Research and Forecasting Model illustrates how advancements in numerical forecasting have significantly contributed to predictions in atmospheric science, which can be adapted for epidemic forecasts (Gochis et al., 2017). Despite these advances, challenges remain, particularly regarding the accurate modelling of unpredictable disease behavior and integration of interdisciplinary data sources. The integration of diverse datasets, a common challenge in fields such as hydrological forecasting, is equally relevant to epidemic modelling, and calls for improved data fusion techniques and collaborative efforts (Tang et al., 2016). In terms of future research, there is a pressing need for models that can adapt to the changing dynamics of diseases influenced by factors, such as travel patterns and climate change. This calls for a multidisciplinary approach that combines insights from climate, geography, and social sciences to better understand how these factors influence disease spread and enhance model robustness (Goethel et al., 2022).

Recommendations for improving statistical epidemic models include:

1. Enhancing model interpretability and transparency to make results more accessible to public health decision makers.
2. Developing more integrated models that combine traditional epidemiological approaches with novel machine learning techniques.
3. Collaboration among researchers from diverse fields is fostered to ensure comprehensive modelling frameworks that consider all relevant variables and uncertainties.

DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc. have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

The author(s) hereby declare that generative AI technologies—specifically ChatGPT, based on the GPT-4o model (Version: August 2025, Source: OpenAI, https://chat.openai.com)—were used during the writing, editing, and formatting processes of this manuscript. These tools assisted in enhancing clarity, grammar, coherence, and in formatting references in APA style. However, all intellectual content, analysis, and conclusions are the sole responsibility of the author(s). All content has been thoroughly reviewed and approved by the author(s) to ensure accuracy and integrity.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

Abbott, S., Hellewell, J., Thompson, R. N., Sherratt, K., Gibbs, H. P., Bosse, N. I., ... & Funk, S. (2020). Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts. *Wellcome Open Research, 5*, 112. https://doi.org/10.12688/wellcomeopenres.16006.1

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

Bahad, P., & Saxena, P. (2019). Study of AdaBoost and Gradient Boosting Algorithms for Predictive Analytics. In *Proceedings of the International Conference on Intelligent Computing and Smart Communication 2018* (pp. 235–244). Springer Singapore. https://doi.org/10.1007/978-981-15-0633-8_22

Bahrampour, A., Hosseinnataj, A., Baneshi, M., & Nikbakht, R. (2019). Comparison of methods to estimate basic reproduction number (R_0) of influenza, using Canada 2009 and 2017–18 A (H1N1) data. *Journal of Research in Medical Sciences, 24*(1), 67. https://doi.org/10.4103/jrms.jrms_888_18

Barnett, L., & Bossomaier, T. (2012). Transfer entropy as a log-likelihood ratio. *Physical Review Letters, 109*(13), 138105. https://doi.org/10.1103/PhysRevLett.109.138105

Bettencourt, L. M. A., & Hyman, J. M. (2009). Mathematical and statistical estimation approaches in epidemiology. Springer Netherlands. https://doi.org/10.1007/978-90-481-2313-1

Bettencourt, L. M., & Ribeiro, R. M. (2008). Real time bayesian estimation of the epidemic potential of emerging infectious diseases. *PloS one*, *3*(5), e2185.

Bleichrodt, A., Luo, R., Kirpich, A., & Chowell, G. (2024). Evaluating the forecasting performance of ensemble sub-epidemic frameworks and other time series models for the 2022–2023 mpox epidemic. *Royal Society Open Science, 11*(7). https://doi.org/10.1098/rsos.240248

Boonpatcharanon, S., Jankowski, H., & Heffernan, J. M. (2022). Estimating the basic reproduction number at the beginning of an outbreak. *PLOS ONE, 17*(6), e0269306.

https://doi.org/10.1371/journal.pone.0269306

Bracher, J., Ray, E. L., Reich, N. G., & Gneiting, T. (2021). Evaluating epidemic forecasts in an interval format. *PLOS Computational Biology, 17*(2), e1008618.

https://doi.org/10.1371/journal.pcbi.1008618

Britton, T., & Pardoux, É. (2019). *Stochastic epidemics in a homogeneous community*. Springer.

Brugnago, E. L., da Silva, R. M., Manchein, C., & Beims, M. W. (2020). How relevant is the decision of containment measures against COVID-19 applied ahead of time? *Chaos, Solitons & Fractals, 140*, 110164. https://doi.org/10.1016/j.chaos.2020.110164

Bu, F., Liang, Y., & Zeng, D. (2020). A likelihood-based framework for partially observed epidemics on dynamic networks. *Biostatistics, 21*(3), 463–480. https://doi.org/10.1093/biostatistics/kxy065

Cauchemez, S., & Ferguson, N. M. (2008). Likelihood-based estimation of continuous-time epidemic models from time-series data: Application to measles transmission in London. *Journal of the Royal Society Interface, 5*(25), 885–897. https://doi.org/10.1098/rsif.2007.1292

Chan, K. S., & Ledolter, J. (1995). Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association, 90*(429), 242–251. https://doi.org/10.2307/2291149

Charniga, K., Madewell, Z. J., Masters, N. B., Asher, J., Nakazawa, Y., & Spicknall, I. H. (2024). Nowcasting and forecasting the 2022 U.S. mpox outbreak: Support for public health decision making and lessons learned. *Epidemics, 47*, 100755. https://doi.org/10.1016/j.epidem.2024.100755

Chen, X., & Moraga, P. (2025). Assessing dengue forecasting methods: A comparative study of statistical models and machine learning techniques in Rio de Janeiro, Brazil. *Tropical Medicine and Health, 53*(1).

https://doi.org/10.1186/s41182-025-00723-7

Chen, Y., Huang, B., Chen, M., Wu, C., & Shi, W. (2021). Modelling the spatiotemporal association between COVID-19 transmission and population mobility using geographically and temporally weighted regression. *GeoHealth,* 5(5). https://doi.org/10.1029/2021GH000402

Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals, 135*, 109864. https://doi.org/10.1016/j.chaos.2020.109864

Chowell, G., Tariq, A., & Hyman, J. M. (2019). A novel sub-epidemic modelling framework for short-term forecasting epidemic waves. *BMC Medicine, 17*(1), 162. https://doi.org/10.1186/s12916-019-1406-6

Clancy, D., & O’Neill, P. D. (2008). Bayesian estimation of the basic reproduction number in stochastic epidemic models. *Mathematical Biosciences, 214*(1–2), 24–34. https://doi.org/10.1016/j.mbs.2008.02.003

Collinson, S., & Heffernan, J. M. (2014). Modelling the effects of media during an influenza epidemic. *BMC Public Health, 14*(1), 376. https://doi.org/10.1186/1471-2458-14-376

Cori, A., Ferguson, N. M., Fraser, C., & Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology, 178*(9), 1505–1512. https://doi.org/10.1093/aje/kwt133

Desai, A. N., Herringer, M., Brownstein, J. S., Nouvellet, P., Cohn, E. L., Cori, A., … & Madoff, L. C. (2019). Real-time epidemic forecasting: Challenges and opportunities. *Health Security, 17*(4), 268–275. https://doi.org/10.1089/hs.2019.0022

Diekmann, O., Heesterbeek, J. A. P., & Metz, J. A. J. (1990). On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology, 28*(4), 365–382. https://doi.org/10.1007/BF00178324

Dietz, K. (1993). The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research, 2*(1), 23–41.

https://doi.org/10.1177/096228029300200103

Dimpfl, T., & Peter, F. J. (2013). Using transfer entropy to measure information flows between financial markets. *Studies in Nonlinear Dynamics & Econometrics, 17*(1), Article 85.

https://doi.org/10.1515/snde-2012-0044

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases, 20*(5), 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1

Erten, E. Y., Lizier, J. T., Piraveenan, M., & Prokopenko, M. (2017). Criticality and information dynamics in epidemiological models. *Entropy, 19*(5), 194. https://doi.org/10.3390/e19050194

Ferguson, N. M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., … & Ghani, A. C. (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce COVID 19 mortality and healthcare demand (Report No. 9). Imperial College COVID-19 Response Team. https://doi.org/10.25561/77482

Fintzi, J., Cui, X., Wakefield, J., & Minin, V. N. (2017a). Efficient data augmentation for fitting stochastic epidemic models to prevalence data. *Journal of Computational and Graphical Statistics, 26*(4), 918–929. https://doi.org/10.1080/10618600.2017.1328365

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., ... & Bhatt, S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature, 584*(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7

Funk, S., Kucharski, A. J., Camacho, A., Eggo, R. M., Yakob, L., Murray, L. M., & Edmunds, W. J. (2016). Comparative analysis of dengue and Zika outbreaks reveals differences by setting and virus. *PLOS Neglected Tropical Diseases, 10*(12), e0005173.

https://doi.org/10.1371/journal.pntd.0005173

Funk, S., Lowe, R., Eggo, R. M., Camacho, A., Edmunds, W. J., & Kucharski, A. J. (2019). Assessing the performance of real-time epidemic forecasts: A case study of Ebola in the Western Area region of Sierra Leone, 2014–15. *PLOS Computational Biology, 15*(2), e1006785. https://doi.org/10.1371/journal.pcbi.1006785

Gatto, A., Accarino, G., Aloisi, V., Immorlano, F., Donato, F., & Aloisio, G. (2021, August). Limits of compartmental models and new opportunities for machine learning: A case study to forecast the second wave of COVID-19 hospitalizations in Lombardy, Italy. *Informatics, 8*(3), 57. https://doi.org/10.3390/informatics8030057

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2020). *Bayesian Data Analysis* (3rd ed.). CRC Press.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1-76.

Gochis, D. J., Dudhia, J., Davis, C. A., Klemp, J. B., Skamarock, W. C., Coen, J. L., … & Dimego, G. J. (2017). The Weather Research and Forecasting Model: Overview, system efforts, and future directions. *Bulletin of the American Meteorological Society, 98*(8), 1717–1737. https://doi.org/10.1175/bams-d-15-00308.1

Goethel, D. R., Thorson, J. T., Gaichas, S., De Moor, C. L., Hesp, S. A., Mcgarvey, R., ... & Chaloupka, M. (2022). Oceans of plenty? Challenges, advancements, and future directions for the provision of evidence-based fisheries management advice. *Reviews in Fish Biology and Fisheries, 33*(2), 375–410.

https://doi.org/10.1007/s11160-022-09726-7

Gomez, L. R. E., Malysheva, N., Pfeil, J., Yang, Z., Bütow, S. M., Irrgang, C., ... & Ladewig, K. (2025). Holistic forecasting for future pandemics: A review of pathogens, models, and data. *Discover Public Health, 22*(1), 1–17.

https://doi.org/10.1186/s12982-025-00573-y

Hespanha, J. P., Chinchilla, R., Costa, R. R., Erdal, M. K., & Yang, G. (2021). Forecasting COVID-19 cases based on a parameter-varying stochastic SIR model. *Annual Reviews in Control, 51*, 460–475. https://doi.org/10.1016/j.arcontrol.2021.08.001

Huang, Y., Liu, D., & Wu, H. (2010). Hierarchical Bayesian methods for estimating HIV dynamic differential equation models. *Bayesian Analysis, 5*(3), 547–574. https://doi.org/10.1214/10-BA519

Jenness, S. M., Goodreau, S. M., & Morris, M. (2018). EpiModel: An R package for mathematical Modelling of infectious disease over networks. *Journal of Statistical Software, 84*(8). https://doi.org/10.18637/jss.v084.i08

Jiang, D., & Ge, Y. (2020). Time-series data augmentation for deep learning: A survey. *arXiv preprint arXiv:2002.12478*.

Jin, M., Pan, S., Chen, S., Li, Y.-F., Yang, B., & Zheng, Y. (2023). Multivariate time series forecasting with dynamic graph neural ODEs. *IEEE Transactions on Knowledge and Data Engineering, 35*(9), 9168–9180. https://doi.org/10.1109/TKDE.2022.3221989

Jin, Z., Zhu, H., & Sun, G. (2014). Epidemic models for complex networks with demographics. *Mathematical Biosciences and Engineering, 11*(6), 1295–1317. https://doi.org/10.3934/mbe.2014.11.1295

Kavouras, I., Kaselimi, M., Protopapadakis, E., Bakalos, N., Doulamis, N., & Doulamis, A. (2022). COVID-19 spatio-temporal evolution using deep learning at a European level. *Sensors, 22*(10), 3658. https://doi.org/10.3390/s22103658

Keeling, M. J., & Rohani, P. (2008). *Modelling infectious diseases in humans and animals*. Princeton University Press.

Keeling, M. J., Tildesley, M. J., Atkins, B. D., Penman, B., Southall, E., Guyver-Fletcher, G., ... & Dyson, L. (2021). The impact of school reopening on the spread of COVID-19 in England. *Philosophical Transactions of the Royal Society B, 376*(1829), 20200261. https://doi.org/10.1098/rstb.2020.0261

Khazaei, Y., Küchenhoff, H., Hoffmann, S., Syliqi, D., & Rehms, R. (2023). Using a Bayesian hierarchical approach to study the association between non-pharmaceutical interventions and the spread of COVID-19 in Germany. *Scientific Reports, 13*, 18900.

https://doi.org/10.1038/s41598-023-46159-5

Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. *Future Internet, 15*(8), 255.

https://doi.org/10.3390/fi15080255

Ladib, M., Ouhinou, A., & Yakubu, A. A. (2024). Mathematical Modelling of contact tracing and stability analysis to inform its impact on disease outbreaks; an application to COVID-19. *Infectious Disease Modelling, 9*(2), 329–353.

https://doi.org/10.1016/j.idm.2024.02.003

Li, W., & Law, K. L. E. (2024). Deep learning models for time series forecasting: A review. *IEEE Access, 12*, 92306–92327. https://doi.org/10.1109/access.2024.3422528

Lizier, J. T., Prokopenko, M., & Zomaya, A. Y. (2012). Local measures of information storage in complex distributed computation. *Information Sciences, 208*, 39–54.

https://doi.org/10.1016/j.ins.2012.04.016

Manna, A., Dall'Amico, L., Tizzoni, M., Karsai, M., & Perra, N. (2024). Generalized contact matrices allow integrating socioeconomic variables into epidemic models. *Science Advances, 10*(41), eadk4606.

https://doi.org/10.1126/sciadv.adk4606

McGough, S. F., Johansson, M. A., Lipsitch, M., & Menzies, N. A. (2020). Nowcasting by Bayesian smoothing: A flexible, generalizable model for real-time epidemic tracking. *PLOS Computational Biology, 16*(4), e1007735.

https://doi.org/10.1371/journal.pcbi.1007735

Merler, S., Ajelli, M., Fumanelli, L., Gomes, M. F. C., Pastore y Piontti, A., Rossi, L., … & Vespignani, A. (2015). Spatiotemporal spread of the 2014 outbreak of *Ebola virus* disease in Liberia and the effectiveness of non pharmaceutical interventions: A computational modelling analysis. *The Lancet Infectious Diseases, 15*(2), 204–211. https://doi.org/10.1016/S1473-3099(14)71074-6

Mistry, D., Halloran, M. E., Fumanelli, L., Merler, S., Ajelli, M., Xiong, X., Chinazzi, M., Liu, Q. H., Mu, K., Litvinova, M., Gomes, M. F. C., Longini, I. M., Pastore Y Piontti, A., Haque, S. A., & Vespignani, A. (2021). Inferring high resolution human mixing patterns for disease Modelling. *Nature Communications, 12*(1), 1–12.

https://doi.org/10.1038/s41467-020-20544-y

Mojtahedi, F. F., Cassidy, M., Chow, S. H., & Yousefpour, N. (2025). Deep learning for time series forecasting: Review and applications in geotechnics and geosciences. *Archives of Computational Methods in Engineering*. Advance online publication. https://doi.org/10.1007/s11831-025-10244-5

Nwankwo, E., Maha, C., Olaboye, J., Emeihe, E., & Ajegbile, M. (2024). Artificial intelligence in predictive analytics for epidemic outbreaks in rural populations. *International Journal of Life Science Research Archive, 7*(1), 78–94. https://doi.org/10.53771/ijlsra.2024.7.1.0062

O'Driscoll, M., Dorigatti, I., Harry, C., Cori, A., & Donnelly, C. A. (2020). A comparative analysis of statistical methods to estimate the reproduction number in emerging epidemics, with implications for the current coronavirus disease 2019 (COVID 19) pandemic. *Clinical Infectious Diseases, 73*(1), e215–e223. https://doi.org/10.1093/cid/ciaa1599

Paap, R. (2001). What are the advantages of MCMC based inference in latent variable models? *Statistica Neerlandica, 56*(1), 2–22. https://doi.org/10.1111/1467-9574.00060

Pare, P. E., Beck, C. L., & Preciado, V. M. (2018). Epidemic processes over time varying networks. *IEEE Transactions on Control of Network Systems, 5*(3), 1322–1334. https://doi.org/10.1109/TCNS.2017.2740355

Park, J., Chaffee, A. W., Harrigan, R. J., & Schoenberg, F. P. (2020). A non parametric Hawkes model of the spread of *Ebola* in West Africa. *Journal of Applied Statistics, 49*(3), 621–637. https://doi.org/10.1080/02664763.2020.1825646

Pastor Satorras, R., Castellano, C., Van Mieghem, P., & Vespignani, A. (2015). Epidemic processes in complex networks. *Reviews of Modern Physics, 87*(3), 925–979. https://doi.org/10.1103/RevModPhys.87.925

Pellis, L., Ball, F., & Trapman, P. (2011). Reproduction numbers for epidemic models with households and other social structures. I. Definition and calculation of R_0. *Mathematical Biosciences, 235*(1), 85–97. https://doi.org/10.1016/j.mbs.2011.10.009

Perasso, A. (2018). An introduction to the basic reproduction number in mathematical epidemiology. *ESAIM: Proceedings and Surveys, 62*, 123–138. https://doi.org/10.1051/proc/201862123

Potter, G. E., Hens, N., & Edmunds, W. J. (2012). Modelling transmission and control of infectious diseases: Data sources and statistical methods. *Statistical Science, 27*(1), 96–110. https://doi.org/10.1214/11-STS370

Ray, E. L., Reich, N. G., & Tibshirani, R. (2020). Adaptive ensemble forecasting of pandemics and epidemics. *Proceedings of the National Academy of Sciences, 117*(48), 30099–30105. https://doi.org/10.1073/pnas.1911234117

Reich, N. G., McGowan, C. J., Yamana, T. K., Tushar, A., Ray, E. L., Osthus, D., … & Brooks, L. C. (2019). Accuracy of real time multi model ensemble forecasts for seasonal influenza in the US. *PLOS Computational Biology, 15*(11), e1007486. https://doi.org/10.1371/journal.pcbi.1007486

Rezk, N. G., Alshathri, S., Sayed, A., El-Din Hemdan, E., & El-Behery, H. (2024). XAI-augmented voting ensemble models for heart disease prediction: A SHAP and LIME-based approach. *Bioengineering (Basel), 11*(10), 1016. https://doi.org/10.3390/bioengineering11101016

Scarpino, S. V., & Petri, G. (2019). On the predictability of infectious disease outbreaks. *Nature Communications, 10*(1). https://doi.org/10.1038/s41467-019-08616-0

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters, 85*(2), 461–464. https://doi.org/10.1103/PhysRevLett.85.461

Sciannameo, V., Goffi, A., Maffeis, G., Gianfreda, R., Pagliari, D. J., Filippini, T., ... & Berchialla, P. (2022). A deep learning approach for spatio-temporal forecasting of new cases and new hospital admissions of COVID-19 spread in Reggio Emilia, Northern Italy. *Journal of Biomedical Informatics, 132*, 104132.

https://doi.org/10.1016/j.jbi.2022.104132

Stocks, T., Britton, T., & Höhle, M. (2018). Model selection and parameter estimation for dynamic epidemic models via iterated filtering: Application to rotavirus in Germany. *Biostatistics, 21*(3), 400–410. https://doi.org/10.1093/biostatistics/kxy057

Stuart, G. K., Minkoff, S. E., & Pereira, F. (2019). A two-stage Markov chain Monte Carlo method for seismic inversion and uncertainty quantification. *Geophysics, 84*(6), R1003–R1020.

https://doi.org/10.1190/geo2018-0893.1

Tang, B., Scarabel, F., Bragazzi, N. L., & Wu, J. (2020a). Lessons drawn from China and South Korea for managing COVID-19 outbreaks: Insights from a comparative Modelling study. *Bulletin of the World Health Organization, 98*(7), 518–524. https://doi.org/10.2471/BLT.20.257600

Tang, L., Wang, F., Zhou, Y., Purkayastha, S., He, J., Zhang, L., Song, P. X.-K., & Wang, L. (2020b). A review of multi compartment infectious disease models. *International Statistical Review, 88*(2), 462–513. https://doi.org/10.1111/insr.12402

Tang, Q., Liu, X., Yuan, X., Zhang, X., Li, Z., Huang, S., Duan, Q., & Cui, H. (2016). Hydrological monitoring and seasonal forecasting: Progress and perspectives. *Journal of Geographical Sciences, 26*(7), 904–920. https://doi.org/10.1007/s11442-016-1306-z

Van De Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G., ... & Veen, D. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers, 1*, 1–26.

https://doi.org/10.1038/s43586-020-00001-2

Wallinga, J., & Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences, 274*(1609), 599–604.
https://doi.org/10.1098/rspb.2006.3754

Wallinga, J., & Teunis, P. (2004). Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology, 160*(6), 509–516. https://doi.org/10.1093/aje/kwh255

Wang, L., Wang, Y., Ye, D., & Liu, Q. (2020a). A review of the 2019 novel coronavirus (COVID-19) based on current evidence. *International Journal of Antimicrobial Agents, 55*(6), 105948.

https://doi.org/10.1016/j.ijantimicag.2020.105948

Wang, N., Fu, Y., Zhang, H., & Shi, H. (2020b). An evaluation of mathematical models for the outbreak of COVID-19. *Precision Clinical Medicine, 3*(2), 85–93.
https://doi.org/10.1093/pcmedi/pbaa016

Wang, W., Tang, J., & Wei, F. (2020c). Updated understanding of the outbreak of 2019 novel coronavirus (2019 nCoV) in Wuhan, China. *Journal of Medical Virology, 92*(4), 441–447.
https://doi.org/10.1002/jmv.25689

West, M. (2019). Bayesian forecasting of multivariate time series: Scalability, structure uncertainty and decisions. *Annals of the Institute of Statistical Mathematics, 72*(1), 1–30.
https://doi.org/10.1007/s10463-019-00741-3

Yang, W., Karspeck, A., & Shaman, J. (2014). Comparison of filtering methods for the modelling and retrospective forecasting of influenza epidemics. *PLoS Computational Biology, 10*(4), e1003583.
https://doi.org/10.1371/journal.pcbi.1003583

Yu, C., Wu, W., Wang, L., Wang, Y., Yao, S., Li, J., Xu, C., Li, Y., & Bai, Y. (2021). Time series analysis and forecasting of the Hand Foot Mouth disease morbidity in China using an advanced exponential smoothing state space TBATS model. *Infection and Drug Resistance, 14*(3), 2809–2821.
https://doi.org/10.2147/IDR.S304652

Yu, M., Huang, Q., & Li, Z. (2024). Deep learning for spatiotemporal forecasting in Earth system science: A review. *International Journal of Digital Earth, 17*(1). https://doi.org/10.1080/17538947.2024.2391952

Zhao, T., Liu, M., & Yu, P. S. (2021). Data augmentation for graph neural networks. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (pp. 521–530). ACM. https://doi.org/10.1145/3459637.3482247

Zhu, D., & Gibson, R. (2018). Seismic inversion and uncertainty quantification using transdimensional Markov chain Monte Carlo method. *Geophysics, 83*(4), R321–R334. https://doi.org/10.1190/geo2016-0594.1

Zhu, X., Ying, H., Chi, H., Zhang, Y., Zeng, L., & Sun, G. (2024). Modelling epidemic dynamics using Graph Attention based Spatial Temporal networks. *PLOS ONE, 19*(7), e0307159. https://doi.org/10.1371/journal.pone.0307159

Peer-review history:
The peer review history for this paper can be accessed here:
https://pr.sdiarticle5.com/review-history/141913