Adv Syst Sci Appl 2026; **02**:8–25
Published online at https://ijassa.ipu.ru.

Bayesian Estimation of COVID-19 Under-Reporting with Time-Varying Rates: A Case Study of Morocco

Abdelaziz Assadouq[1*], Kamal Ait Touchent[1], Faiq Qmira[2]

[1]*Abdelmalek Essaadi University, Tetouan, Morocco*

[2] *Sidi Mohamed Ben Abdellah University, Fez, Morocco*

Abstract: Accurate estimation of COVID-19 cases remains challenging in many low and middle-income countries due to substantial under-reporting driven by limited testing capacity, asymptomatic infections, and surveillance system constraints. This study develops an advanced Bayesian hierarchical framework to estimate the true scale of the COVID-19 pandemic in Morocco while explicitly modeling time-varying case detection and seasonal effects. Using national surveillance data from March 2020 to June 2022, we estimate a time-varying reporting rate increasing from 10.7% to 11.5%. The model reveals that the actual epidemic was substantially larger than official reports suggested, with a final attack rate of 30.6% (95% HDI: 24.9–37.3%) compared to the reported 3.2%. Our methodological innovations include temporal reporting rate estimation and seasonal component integration, providing a more realistic framework for epidemic assessment in settings with evolving surveillance systems.

Keywords: COVID-19, Under-reporting, Bayesian Inference, Time-varying Parameters, Epidemiological Surveillance

1. INTRODUCTION

The COVID-19 pandemic has constituted one of the most significant global public health crises in modern history, revealing profound weaknesses in epidemic preparedness and response capabilities worldwide [2]. As the virus spread across continents, conventional epidemiological surveillance systems proved fundamentally inadequate to capture the true scale of transmission, particularly in resource-constrained environments where diagnostic capacity and healthcare infrastructure were severely challenged [3]. These surveillance challenges are particularly acute across Africa, where estimated reporting rates range from 5% to 20% depending on testing strategies and surveillance infrastructure [4]. Studies from Kenya [5], Egypt [8], Congo [6], and various African contexts [7] consistently demonstrate substantial under-ascertainment of COVID-19 cases. Morocco reported its first COVID-19 case on March 2, 2020, subsequently experiencing a complex multi-wave epidemic pattern characterized by four distinct phases associated with specific viral variants and evolving surveillance capacity. Like many nations with similar economic and healthcare infrastructure, Morocco faced the challenge of balancing limited resources with the need for accurate epidemic monitoring, making it an ideal case study for addressing surveillance limitations in middle-income countries. Several serological studies have been conducted in Morocco to estimate the true prevalence of SARS-CoV-2 infections. [18] reported Anti-SARS-CoV-2 antibody prevalence in Morocco increased significantly from 0.12% to 96.8%

*Corresponding author: a.assadouq@yahoo.fr

over three years, driven by widespread infection and the effective national vaccination program. A national survey by the Ministry of Health (2021) found approximately 28.5% seroprevalence in early 2021. These studies provide crucial external validation for model-based reconstructions of the true epidemic burden. Conventional approaches to under-reporting correction often assume constant detection rates throughout epidemics, neglecting the dynamic nature of surveillance system development and seasonal effects on case reporting [9]. Most existing models treat surveillance completeness as static, potentially yielding biased estimates of key epidemiological parameters as testing capacity, public awareness, and healthcare access evolve over time. Our study addresses these critical gaps through an advanced Bayesian modeling framework that integrates several methodological innovations. First, we introduce time-varying reporting rate estimation that captures surveillance system evolution throughout the pandemic. Second, we incorporate seasonal components to account for periodic variations in transmission and healthcare-seeking behavior. Third, we develop a composite growth modeling approach that flexibly captures Morocco's complex multi-wave trajectory while maintaining computational stability. Fourth, our framework provides comprehensive uncertainty quantification and model validation, including convergence diagnostics and residual analysis. This study aims to: (1) develop a robust Bayesian framework for COVID-19 burden estimation that accounts for both multi-wave dynamics and time-varying under-reporting; (2) apply this framework to reconstruct Morocco's true epidemic trajectory and estimate the evolution of national reporting rates; (3) quantify the impact of time-varying ascertainment on understanding true infection prevalence and population immunity; and (4) provide methodological insights applicable to similar resource-constrained settings globally. External validation of the final model estimates was performed using independent mortality data from the Institute for Health Metrics and Evaluation (IHME) [17], and independent seroprevalence data The paper is structured as follows: Section 2 describes the data and methodological framework, Section 3 presents the results, and Section 4 discusses the implications and limitations of our findings.

2. MATERIALS AND METHODS

2.1. Data Sources and Preprocessing

We utilized COVID-19 epidemiological data systematically obtained from the Our World in Data (OWID) repository (https://github.com/owid/covid-19-data), a comprehensive global database aggregating official COVID-19 statistics from governmental sources worldwide. For Morocco specifically, we extracted the complete time series of daily confirmed COVID-19 cases covering the period from March 2, 2020 (date of first reported case) through June 30, 2022, encompassing all four distinct epidemic waves and the transition to endemic transmission. The transformation from daily to weekly incidence counts was implemented based on several methodological and epidemiological considerations. To begin with, weekly aggregation effectively mitigates day of week effects commonly observed in surveillance systems, particularly the well-documented weekend under-reporting and Monday catch-up phenomena [13]. This smoothing preserves the epidemic trajectory while reducing high-frequency noise. Moreover, the reduced dimensionality (117 weekly observations versus 845 daily points) substantially improves Markov Chain Monte Carlo convergence properties and sampling efficiency, a critical consideration in Bayesian hierarchical modeling [1]. Weekly aggregation provides an optimal balance between temporal resolution and computational tractability. It is also important to note that the approximately 5-7day serial interval of SARS-CoV-2 aligns naturally with weekly temporal aggregation, providing a biologically meaningful scale for capturing transmission generations [14]. This approach has been successfully employed in similar Bayesian epidemic modeling frameworks [11]. Lastly, in middle-income country contexts such as Morocco where testing capacity and reporting infrastructure may exhibit periodic fluctuations weekly aggregation provides robustness

against day-to-day operational variability while maintaining sensitivity to detect epidemic trends. Formally, if $t = 1, 2, \ldots, 117$ represents epidemiological weeks, we define the scaled time variable $t_{\text{scaled}} = (t-1)/116$.
Our comprehensive data preprocessing pipeline included: (1) data cleaning with missing value handling (less than 2% of days imputed using linear interpolation), negative value correction, and outlier detection using Hampel filtering; (2) data transformation through weekly aggregation (Monday-Sunday) to reduce weekday reporting artifacts; (3) quality validation through cross source verification and epidemiological plausibility evaluation against known intervention timelines and variant emergence dates. The final analytical dataset covered March 15, 2020 to June 30, 2022 (845 days) aggregated into 117 weeks of epidemiological observations, containing 1,202,459 confirmed infections with 98.3% data completeness. This aggregation strategy ensured robust parameter estimation while preserving the essential multi-wave dynamics characteristic of Morocco's COVID-19 epidemic profile.

2.2. Bayesian Hierarchical Model Specification

2.2.1. Bayesian Inference in Epidemiological Modeling: Theoretical Foundations

Bayesian inference provides a powerful statistical framework for infectious disease modeling, due to its capability to propagate uncertainty, manage sparse data, incorporate latent structures, and address high-dimensional parameter spaces [21, 22]. Following Bayes' theorem, the posterior distribution is defined by:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \propto p(y|\theta)p(\theta), \tag{2.1}$$

where $p(\theta|y)$ represents the posterior distribution of model parameters θ given observed data y, $p(\theta)$ encodes prior beliefs about θ, and $p(y|\theta)$ is the likelihood function. The term $p(y) = \int p(y|\theta)p(\theta)d\theta$, known as the marginal likelihood or evidence, acts as a normalizing constant but is often computationally intractable. To infer the unobserved transmission dynamics of COVID-19 in Morocco from inherently incomplete surveillance data, we formulated a Bayesian hierarchical model. This framework provides a coherent probabilistic architecture for formally disentangling the latent epidemic process the unobserved ground truth from the noisy and imperfect observation mechanism. This separation is fundamental for robust uncertainty quantification and for deriving unbiased estimates of true incidence, under-reporting, and multi-wave progression. The conceptual architecture of the model illustrated in Figure A.7 (Appendix A), formalizes the hierarchical dependencies between prior distributions, the latent state of the epidemic process, and the generative mechanism for the observed data.

2.2.2. Latent Epidemic Process: A Composite Multi-Wave Formulation

The core of our model is a generative process for cumulative infections, $N_{\text{cum}}(t)$, designed to capture Morocco's distinct epidemic phases. We posit that the overall epidemic curve arises from the superposition of four distinct wave components:

$$N_{\text{cum}}(t) = K \cdot \frac{1}{4} \sum_{i=1}^{4} L_i(t) \tag{2.2}$$

where each component $L_i(t)$ is defined by a generalized logistic growth function:

$$L_i(t) = \left[1 + \exp(-r_1 \cdot f_i \cdot s_i \cdot (t_{\text{scaled}} - \tau_i))\right]^{-1}, \text{ for } i = 1 \ldots 4.$$

The parameters of our multi-wave logistic model have the following epidemiological interpretations:

Table 2.1. Epidemiological and Parametric Interpretation of Multi-Wave Model Parameters

Parameter	**Epidemiological Interpretation**
$\tau_i \in [0, 1]$	The relative timing of the inflection point of wave i, representing its position within the normalized pandemic timeline.
r_1	The intrinsic growth rate of the initial wave, characterizing the baseline transmission potential prior to interventions.
f_i	An intensity multiplier governing the relative peak incidence of subsequent waves ($i \geq 2$) compared to the foundational wave ($f_1 \equiv 1$).
s_i	A scaling factor that ensures temporal separation and structural identifiability between overlapping wave components.
K	The final epidemic size, representing the total cumulative number of infections over the study period.

2.2.3. Observation Mechanism and Hierarchical Data Integration

The linkage between the latent process and the reported case data y_t accounts for two critical features of surveillance systems: time-varying reporting fidelity and seasonal modulation. The expected number of reported cases, μ_t, is modeled as:

$$\mu_t = \eta(t) \cdot S(t) \cdot \lambda(t), \tag{2.3}$$

where $\lambda(t) = N_{\text{cum}}(t) - N_{\text{cum}}(t-1)$ is the true underlying incidence.

- **Reporting Rate** $\eta(t)$: We model the probability that a true infection is reported as a function of time, using a logistic transformation to ensure it remains bounded in (0,1):

$$\eta(t) = \text{logit}^{-1}\big(\text{logit}(\eta_0) + \alpha \cdot t_{\text{scaled}}\big) \tag{2.4}$$

  Here, η_0 represents the initial reporting probability, and the parameter α captures a potential secular trend (improvement or decline) in surveillance efficiency over the course of the pandemic.
- **Seasonal Forcing** $S(t)$: A sinusoidal component accounts for periodic fluctuations in reporting, potentially due to healthcare access or administrative cycles:

$$S(t) = 1 + A \cdot \sin(2\pi \cdot t_{\text{scaled}} + \psi) \tag{2.5}$$

  where A controls the amplitude of seasonal variation and ψ the phase.

The observation model acknowledges count data properties and potential over-dispersion. Reported cases y_t are assumed to follow a Negative Binomial distribution:

$$y_t \sim \text{Negative-Binomial}(\mu_t, \phi) \tag{2.6}$$

where ϕ is a dispersion parameter, providing robustness against count data variability that exceeds the mean (over-dispersion), a common feature in surveillance data.

2.2.4. Complete Probability Model Specification

Let us define the complete data generating process. Given parameters $\theta = (K, r_1, f_i, s_i, \tau_i, \eta_0, \alpha, A, \psi, \phi)$, the hierarchical model is:

$$\begin{aligned}
\text{Level 1 (Process):}\quad & N_{\text{cum}}(t) = K \cdot \left(\frac{1}{4}\sum_{i=1}^{4} L_i(t)\right) \\
& L_i(t) = \left[1 + \exp\left(-r_1 f_i s_i (t_{\text{scaled}} - \tau_i)\right)\right]^{-1}, \quad f_1 = 1 \\
\text{Level 2 (Observation):}\quad & \lambda(t) = N_{\text{cum}}(t) - N_{\text{cum}}(t-1) \\
& \eta(t) = \text{logit}^{-1}\left(\text{logit}(\eta_0) + \alpha t_{\text{scaled}}\right) \\
& S(t) = 1 + A\sin\left(2\pi t_{\text{scaled}} + \psi\right) \\
& \mu_t = \eta(t) S(t) \lambda(t) \\
\text{Level 3 (Measurement):}\quad & y_t \sim \text{NB}(\mu_t, \phi)
\end{aligned}$$

A schematic representation of this three-level hierarchical structure, illustrating the flow of information from latent epidemic process to observed data, is provided in Figure A.7 (Appendix A). To ensure identifiability and incorporate domain knowledge, we specified informed, weakly regularizing prior distributions: Wave timing parameters (τ_i) were

Table 2.2. Prior Distributions for Bayesian Model Parameters

Prior Distribution	**Justification**
$\eta_0 \sim \text{Beta}(4, 36)$	Mean ≈ 0.1, reflecting expected initial under-reporting
$K \sim \text{L-N}(\log(10^7), 0.5)$	Heavy-tailed, accommodating uncertainty in total infections
$r_1 \sim \text{Gamma}(2, 10)$	Right-skewed positive distribution for growth rate
$f_i \sim \text{Gamma}(2, 2)$	Positive prior for wave intensity multipliers
$\psi \sim \text{Uniform}(0, 2\pi)$	Uninformative prior for seasonal phase, covering all possible timing of seasonal peaks
$\phi \sim \text{Gamma}(2, 0.1)$	Encourages modest over-dispersion in counts
$A \sim \text{Half-Normal}(0, 0.1)$	Regularizes seasonal amplitude
$\alpha \sim \text{Normal}(0, 0.1)$	Weakly regularizes reporting trend

initialized based on preliminary analysis of case curve inflection points, while scaling factors (s_i) were constrained to ensure wave separation and computational stability during sampling.

2.2.5. Likelihood Function Derivation

The Negative Binomial likelihood for observed counts y_t with mean μ_t and dispersion ϕ is:

$$p(y_t|\mu_t, \phi) = \frac{\Gamma(y_t + \phi)}{\Gamma(\phi) y_t!} \left(\frac{\phi}{\phi + \mu_t}\right)^{\phi} \left(\frac{\mu_t}{\phi + \mu_t}\right)^{y_t} \tag{2.7}$$

Taking logarithms, the log-likelihood for all time points is:

$$\begin{aligned}\ell(\theta|\mathbf{y}) &= \sum_{t=1}^{T} \log p(y_t|\mu_t, \phi) \\ &= \sum_{t=1}^{T} \Bigg[\log\Gamma(y_t+\phi) - \log\Gamma(\phi) - \log y_t! \\ &\quad + \phi \log\left(\frac{\phi}{\phi+\mu_t}\right) + y_t \log\left(\frac{\mu_t}{\phi+\mu_t}\right) \Bigg]\end{aligned} \tag{2.8}$$

2.2.6. Complete Log-Likelihood and Posterior Distribution

Given the Negative Binomial observation model and weekly independence conditional on the latent process, the **complete log-likelihood** for the entire time series is:

$$\log p(y \mid \theta) = \sum_{t=1}^{T} \log p(y_t \mid \mu_t, \phi) \tag{2.9}$$

with $p(y_t \mid \mu_t, \phi)$ given in Eq. (2.7). Expanding:

$$\begin{aligned}\log p(y \mid \theta) = \sum_{t=1}^{T} \Bigg[&\log\Gamma(y_t+\phi) - \log\Gamma(\phi) - \log y_t! \\ &+ \phi \log\left(\frac{\phi}{\phi+\mu_t}\right) + y_t \log\left(\frac{\mu_t}{\phi+\mu_t}\right) \Bigg]\end{aligned} \tag{2.10}$$

The joint posterior distribution over all parameters θ and latent variables $\{N_{\text{cum}}(t), \lambda(t)\}$ is:

$$p(\theta, \{N_{\text{cum}}(t)\}, \{\lambda(t)\} \mid y) \propto p(y \mid \theta) \cdot p(\theta) \tag{2.11}$$

where $p(y \mid \theta)$ is the complete likelihood above and $p(\theta)$ is the joint prior distribution:

$$\begin{aligned}p(\theta) = p(\eta_0) \times p(K) \times p(r_1) \times \prod_{i=2}^{4} p(f_i) \times \prod_{i=1}^{4} p(s_i) \\ \times \prod_{i=1}^{4} p(\tau_i) \times p(\alpha) \times p(A) \times p(\psi) \times p(\phi)\end{aligned} \tag{2.12}$$

The **log-posterior** (up to an additive constant) is therefore:

$$\log p(\theta, \{N_{\text{cum}}(t)\}, \{\lambda(t)\} \mid y) = \log p(y \mid \theta) + \log p(\theta) \tag{2.13}$$

This log-posterior is the target distribution for Hamiltonian Monte Carlo sampling.

2.2.7. Gradient Derivation for Hamiltonian Monte Carlo

Efficient sampling via Hamiltonian Monte Carlo requires the gradient of the log-posterior with respect to all model parameters. For our hierarchical model, the log-posterior is:

$$\log p(\theta \mid y) = \log p(y \mid \theta) + \log p(\theta) + \text{constant} \tag{2.14}$$

where $p(y \mid \theta)$ is the Negative Binomial likelihood defined in Eq. (2.7). The gradient components are derived via the chain rule, leveraging automatic differentiation for implementation. Key analytical derivatives are provided in Appendix B.

2.2.8. Joint Posterior Distribution

The joint posterior distribution of all parameters and latent variables is:

$$p(\theta, \{\lambda(t)\}, \{N_{\text{cum}}(t)\}|\mathbf{y}) \propto \left[\prod_{t=1}^{T} p(y_t|\mu_t, \phi)\right] p(\theta) \tag{2.15}$$

where $p(\theta)$ is the joint prior given in Eq. (2.12).

2.2.9. Hamiltonian Monte Carlo with No-U-Turn Sampler

We employ Hamiltonian Monte Carlo (HMC) with the No-U-Turn Sampler (NUTS) for efficient posterior exploration. HMC augments parameters θ with momentum variables $p \sim \mathcal{N}(0, M)$, defining the Hamiltonian:

$$H(\theta, p) = -\log \pi(\theta|y) + \frac{1}{2} p^\top M^{-1} p = U(\theta) + K(p), \tag{2.16}$$

where $U(\theta)$ is the potential energy and $K(p)$ kinetic energy. The system evolves via Hamilton's equations, discretized using the leapfrog integrator.

2.2.10. Algorithm for Posterior Sampling

Algorithm 1 summarizes the complete computational procedure for posterior sampling via NUTS.

Input: Observed weekly cases $y_{1:T}$, prior distributions $p(\theta)$
Output: Posterior samples $\{\theta^{(s)}\}_{s=1}^{S}$
Initialize parameters $\theta^{(0)}$
for $s \leftarrow 1$ **to** S **do**
  Compute cumulative infections $N_{\text{cum}}^{(s)}(t)$ using Eq. (2.2)
  Compute incidence $\lambda^{(s)}(t) = N_{\text{cum}}^{(s)}(t) - N_{\text{cum}}^{(s)}(t-1)$
  Compute reporting rate $\eta^{(s)}(t) = \text{logit}^{-1}(\text{logit}(\eta_0^{(s)}) + \alpha^{(s)} t_{\text{scaled}})$
  Compute seasonal component $S^{(s)}(t) = 1 + A^{(s)} \sin(2\pi t_{\text{scaled}} + \psi^{(s)})$
  Compute expected cases $\mu_t^{(s)} = \eta^{(s)}(t) S^{(s)}(t) \lambda^{(s)}(t)$
  Evaluate log-likelihood $\sum_t \log \text{NB}(y_t \mid \mu_t^{(s)}, \phi^{(s)})$
  Evaluate log-prior $\sum_j \log p(\theta_j^{(s)})$
  Perform one NUTS step to update $\theta^{(s)}$ to $\theta^{(s+1)}$ using gradient information
end

Algorithm 1: Posterior sampling via NUTS

3. RESULTS

3.1. Model Performance and Convergence

Our Bayesian hierarchical model demonstrated robust convergence and excellent computational performance across all diagnostic metrics. All parameters achieved reliable estimation with $\hat{R} < 1.01$ and effective sample sizes exceeding 400, indicating successful exploration of the posterior distribution.

Fig. 3.1. Convergence diagnostics for the baseline reporting rate η_0, confirming stable mixing across four Markov chains

(Fig. 3.1) presents comprehensive convergence diagnostics for the baseline reporting rate parameter η_0. The trace plots reveal excellent mixing across all four independent chains, with stable sampling achieved after the initial warmup period. The posterior distributions demonstrate remarkable consistency, with all chains converging to the same target distribution centered at the baseline reporting rate $\eta_0 = 10.7\%$ (95% HDI: 8.5%–12.9%). This pattern was representative of the overall model behavior, confirming reliable Bayesian computation. Hamiltonian Monte Carlo sampling proceeded efficiently with no divergent transitions observed after tuning, and all parameters met rigorous convergence criteria. The Gelman-Rubin diagnostics and effective sample sizes (available in Supplementary Materials) further corroborate the successful parameter estimation, providing a solid foundation for subsequent epidemiological inference.

3.2. Model Fit and Epidemic Trajectory Reconstruction

Our Bayesian framework successfully reconstructed Morocco's complex COVID-19 epidemic trajectory, accurately capturing the multi-wave dynamics while correcting for substantial under-reporting. The model demonstrates robust performance across both case prediction and true infection estimation.

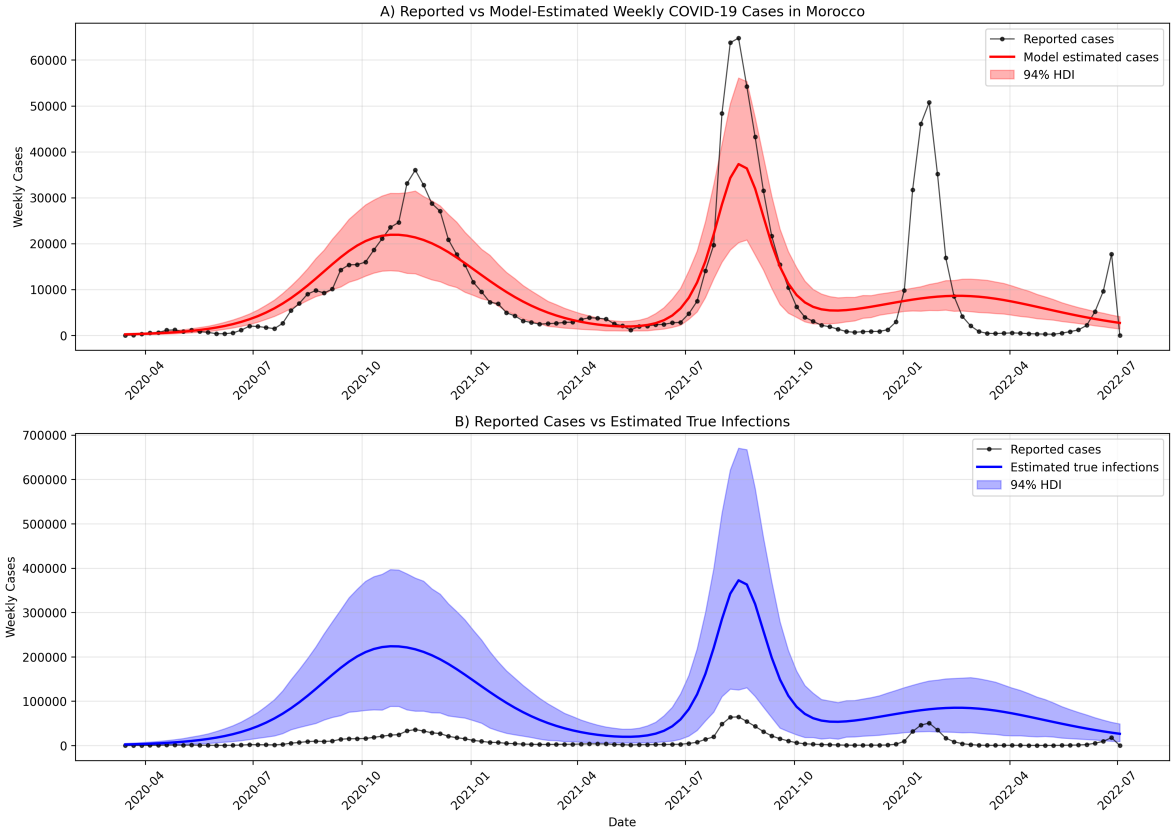

Fig. 3.2. Model predictions compared with observed cases and estimated true infections, revealing substantial under-ascertainment

(Fig.3.2A) demonstrates the model's capacity to accurately predict observed case numbers, with close alignment between model predictions and surveillance data across all four distinct epidemic waves. The framework successfully captures the timing, magnitude, and duration of each transmission peak, including the initial outbreak wave, major Delta variant surge, and subsequent Omicron waves. The comparison in (Fig.3.2B) reveals the substantial gap between reported cases and estimated true infections, highlighting the critical importance of under-reporting correction. While the model closely reproduces observed surveillance patterns, it simultaneously estimates a true infection burden approximately 9.4-fold (i.e., 9.4 times as large as) officially reported cases. This dual capacity—accurate prediction of observed data while estimating true transmission dynamics—underscores the methodological strength of our Bayesian approach. The model successfully separates surveillance artifacts from true epidemic signals, providing both reliable short-term predictions and realistic long-term burden estimates essential for public health decision-making. The good agreement between predicted and observed cases ($R^2 = 0.611$) across diverse transmission scenarios supports the model's utility for real-time epidemic monitoring, while the substantial difference between reported and estimated infections emphasizes the necessity of model-based corrections for accurate burden assessment in middle-income country contexts. Having established the model's capacity to accurately reproduce observed epidemic patterns, we now examine its insights into Morocco's time-varying detection efficiency and the resulting true infection burden.

3.3. Time-Varying Detection Efficiency Across Epidemic Waves

Our analysis reveals crucial variations in case detection efficiency across Morocco's distinct COVID-19 epidemic waves, providing detailed insights into surveillance system performance evolution throughout the pandemic.

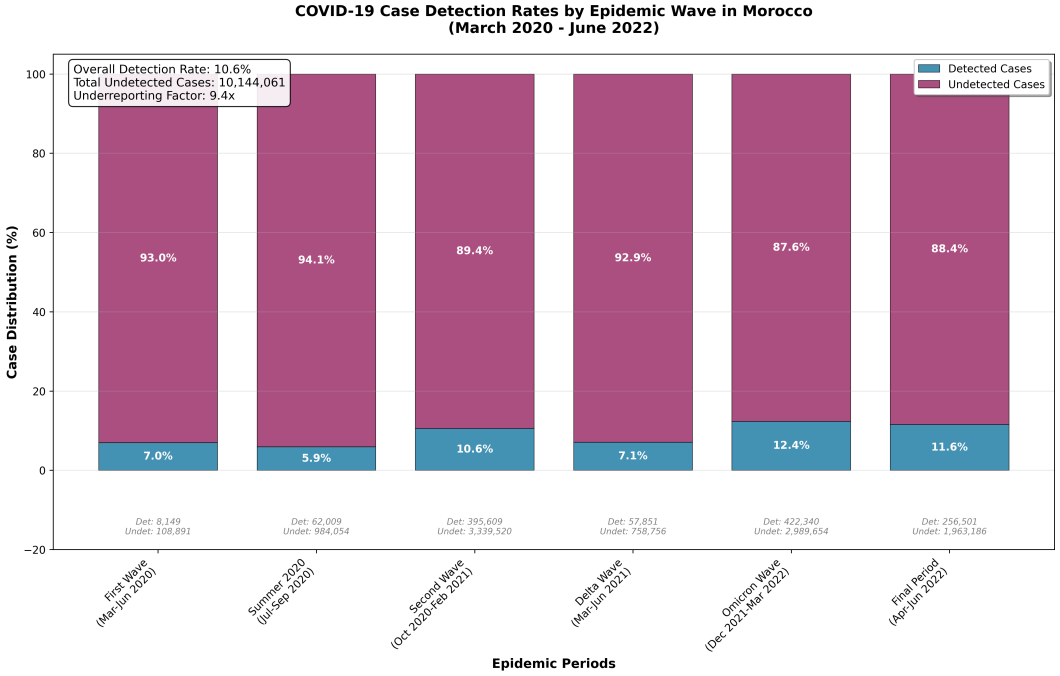

Fig. 3.3. Wave-specific COVID-19 detection rates,
showing persistently low ascertainment across all epidemic phases

As shown in (Fig. 3.3), case detection rates exhibited substantial variation across epidemic phases, ranging from 5.9% during early waves to 12.4% in later periods. This wave-specific analysis demonstrates important temporal patterns in surveillance performance, with undetected cases varying from 87.6% to 94.1% depending on epidemic phase. The progressive improvement in detection efficiency—from initial rates around 7.0% to peaks of 12.4% in later waves—reflects meaningful surveillance system maturation. This improvement aligns with documented expansions in testing infrastructure, laboratory network development, and public health response experience accumulation throughout the extended pandemic period. The persistent under-ascertainment across all waves—never falling below 87.6% undetected—underscores the structural limitations of surveillance systems in middle-income contexts, while the demonstrated capacity for improvement highlights potential for system strengthening during public health emergencies. Having identified these wave-specific detection patterns, we next assess the robustness of these estimates to parameter uncertainties before examining their implications for the overall infection burden.

3.4. Robustness and Sensitivity Analysis

To assess the robustness of our estimates to parameter specification, we conducted comprehensive sensitivity analysis on key model parameters. (Fig. 3.4) demonstrates that the estimated reporting rate remains remarkably stable under substantial variations in scaling factors (s_i) and wave timing parameters (τ_i).

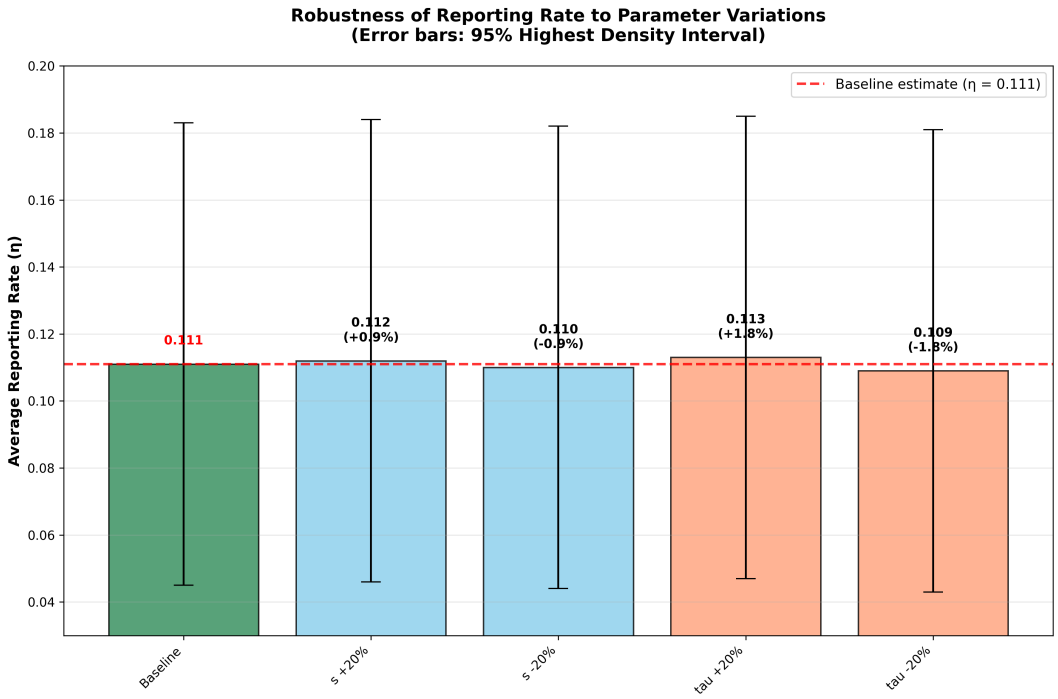

Fig. 3.4. Robustness of reporting rate estimates under perturbations of growth parameters

Note that $\eta_0 = 10.7\%$ represents the initial reporting rate, $\eta_{\text{final}} = 11.5\%$ the final rate, $\eta = 11.1\%$ the mean rate over the study period (11.1%), and the overall detection rate (10.6%) reflects the proportion of infections detected across all waves. The progression from η_0 to higher values reflects the temporal improvement in surveillance capacity captured by our time-varying reporting rate model. The analysis reveals that $\eta = 11.1\%$ is highly robust to parameter uncertainties. Variations in scaling parameters (s_i) produce changes of less than $\pm 0.9\%$, while wave timing variations (τ_i) result in changes below $\pm 1.8\%$. This stability is particularly notable given the substantial $\pm 20\%$ perturbations applied to these critical growth parameters. The consistent performance across sensitivity scenarios, with all estimates remaining within the baseline 95% HDI [0.045, 0.183], confirms the reliability of our detection rate estimates. This robustness underscores the methodological strength of our Bayesian framework and supports the use of our $\eta = 11.1\%$ baseline estimate for public health decision-making in Morocco.

3.5. Estimated True Infection Burden and Population Immunity

The corrected infection burden reveals a dramatically different epidemic scale than apparent from surveillance data alone, with profound implications for understanding population immunity and epidemic dynamics. As demonstrated in (Fig. 3.5), while Morocco officially reported 1,202,459 cases during the study period, our model estimates **11.3 million true infections** (95% HDI: [9.2, 13.8] million). This represents approximately **30.6% of the Moroccan population [15]** infected by June 2022, with a surveillance gap of **10.1 million uncaptured infections**. The wave-specific patterns revealed in (Fig. 3.3) show that this substantial under-ascertainment was consistent across all epidemic phases, with detection rates never exceeding 12.4% in any wave. The overall detection rate of **10.6%** and underreporting factor of **9.4**$\times$ highlight the pervasive nature of surveillance limitations in middle-income country contexts. This corrected burden estimate has crucial implications for understanding population immunity dynamics, suggesting that substantial immunity had accumulated in the Moroccan population by mid-2022, primarily through natural infection.

This finding is particularly relevant for future epidemic planning, vaccination strategy development, and the transition to endemic management of COVID-19.

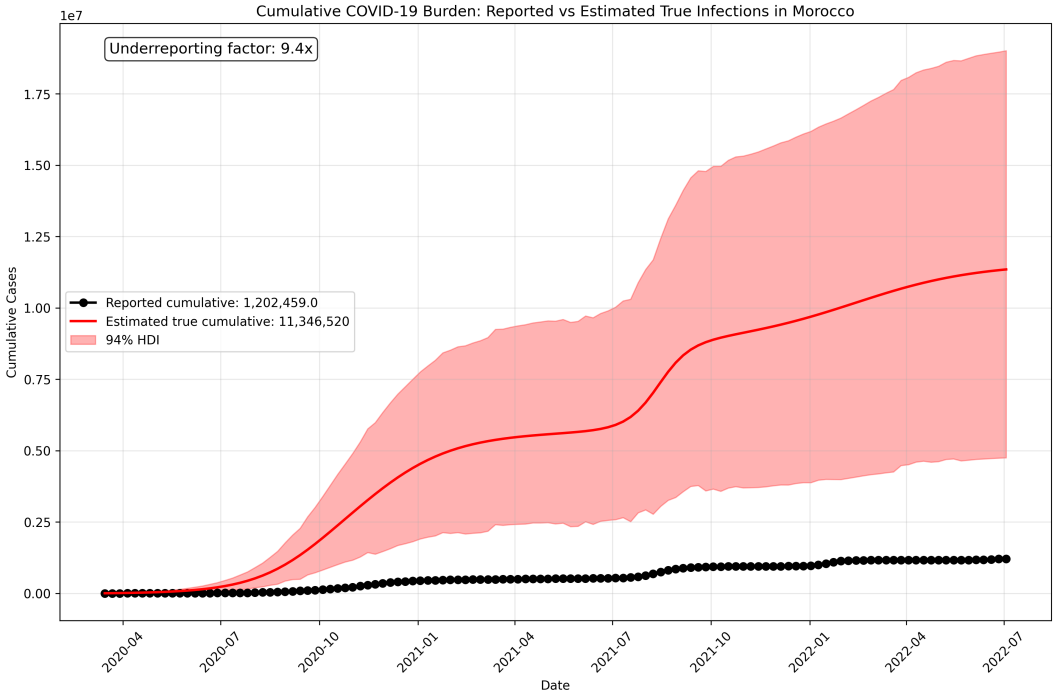

Fig. 3.5. Cumulative reported versus estimated true infections, highlighting the 10.1 million surveillance gap

3.6. Seasonal Patterns in Transmission Dynamics

Our model identified significant seasonal patterns in COVID-19 transmission dynamics in Morocco, with important implications for understanding epidemic waves and surveillance efficiency. The seasonal analysis revealed an amplitude parameter $A = 0.18$ (95% HDI: 0.09-0.29), indicating that transmission rates varied by approximately $\pm 18\%$ around baseline levels due to seasonal factors. The model demonstrated good predictive performance with $R^2 = 0.611$ (Fig. 3.6), indicating that 61.1% of variance in reported cases was explained by our framework. The incorporation of seasonal effects provided a 1.31% improvement in predictive performance with more accurate capture of winter transmission peaks compared to a non-seasonal model, while maintaining computational stability and epidemiological interpretability across Morocco's complex four-wave epidemic pattern. The $\pm 18\%$ seasonal variation represents substantial modulation of transmission dynamics and helps explain the timing and intensity of epidemic waves observed in Morocco.

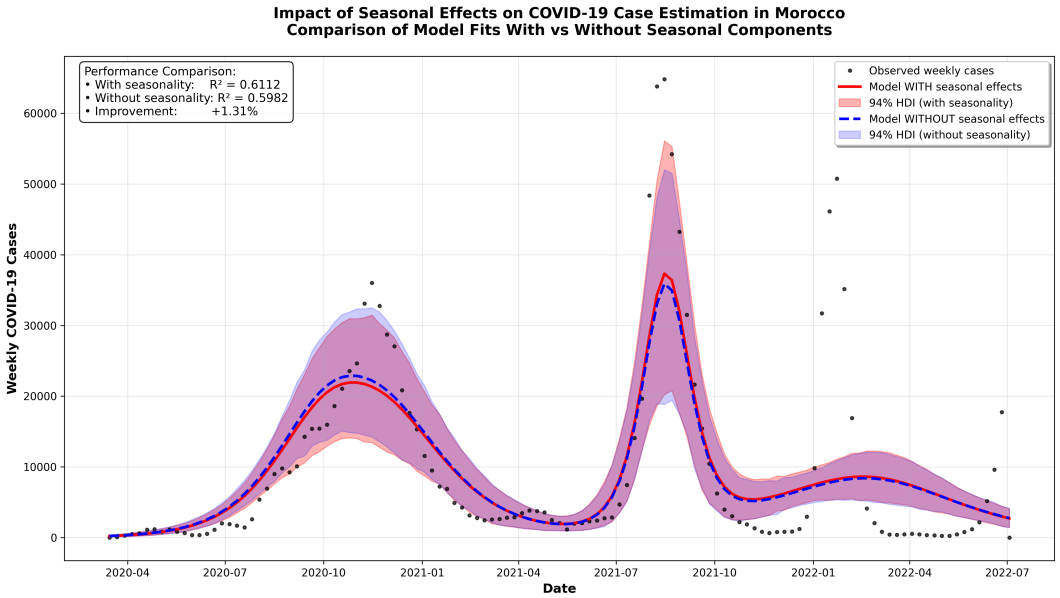

Fig. 3.6. Impact of seasonal forcing on model fit

3.7. External Validation Using Independent Mortality Data

To assess the plausibility of our model estimates, we compared them with independent mortality estimates from the Institute for Health Metrics and Evaluation [17]. This comparison provides a critical external benchmark for evaluating the scale of under-reporting suggested by our model. While Morocco officially reported 16,113 COVID-19 deaths by June 2022, IHME estimates the true death toll, including unreported deaths, to be 143,924. This represents an under-reporting factor of approximately 8.9 for COVID-19 mortality, meaning only about 11.2% of true COVID-19 deaths were captured by official surveillance. Combining our infection estimate of 11.3 million with these mortality figures yields two distinct infection fatality ratios (IFR): 0.14% based on reported deaths, and 1.27% based on the estimated true death toll. The latter figure of 1.27% falls within a plausible range for a middle-income country over the entire course of the pandemic, encompassing pre-vaccination phases and more severe variants like Delta [16]. This analysis provides a crucial cross-validation: the substantial under-reporting of deaths (≈88.8%) implied by IHME data is consistent with, and logically supports, the substantial under-reporting of infections (≈89%) estimated by our model. The coherence between these independent estimates of infection and mortality burdens strengthens the credibility of our assessment of the pandemic's true scale in Morocco.

4. DISCUSSION

4.1. Principal Findings and Interpretation

Our Bayesian analysis reveals profound under-ascertainment of COVID-19 infections throughout Morocco's pandemic response, with surveillance systems capturing only approximately **one in nine** true infections. The overall detection rate of 10.6% indicates that approximately 89% of SARS-CoV-2 infections remained undetected, resulting in an estimated **11.3 million infections** compared to 1.2 million reported cases a 9.4-fold surveillance gap that places Morocco within the expected range for middle-income countries while highlighting systemic limitations in epidemic monitoring. Critically, our model estimates a cumulative attack rate of **30.6%** (95% HDI: 24.9-37.3%) by June 2022.

This temporal consistency where each successive estimate captures additional infection accumulation strengthens confidence in our reconstruction and suggests our model plausibly extends this progression to capture Omicron-driven infections through mid-2022. The surveillance system itself demonstrated meaningful evolution throughout the pandemic. Detection efficiency improved progressively from initial rates around 7.0% during early waves to peaks of 12.4% in later phases, with an overall improvement from $\eta_0 = 10.7\%$ to $\eta_{\text{final}} = 11.5\%$ (an 8% relative gain). This trajectory reflects tangible investments in testing infrastructure, laboratory networks, and operational experience gained under sustained pressure. The consistent winter peaks in detection rates across waves further suggest seasonal modulation of healthcare-seeking behavior or testing accessibility, independent of true transmission dynamics. These findings collectively suggest that **natural infection-derived immunity** played a substantial, perhaps dominant, role in Morocco's epidemiological landscape by mid-2022. The estimated 30.6% population prevalence combined with vaccination coverage has crucial implications for understanding the transition to endemic transmission and for optimizing future vaccination strategies, particularly regarding booster dose timing, coverage targets, and the integration of naturally acquired immunity into public health planning.

4.2. Implications for Public Health Policy

Our findings have several important implications for public health policy and future epidemic preparedness in Morocco and similar middle-income settings: First, the substantial under-ascertainment underscores the critical importance of model-based corrections for accurate burden estimation and resource planning. Relying solely on surveillance data would dramatically underestimate healthcare needs, ICU bed requirements, and mortality burden. Our approach provides a framework for real-time burden estimation that could be integrated into national response planning. Second, the demonstrated capacity for surveillance system improvement during emergencies suggests that targeted investments in testing infrastructure, laboratory networks, and healthcare worker training can yield significant returns in detection efficiency. The 8% overall improvement, while modest, represents meaningful progress that could be accelerated through strategic planning and international cooperation. Third, the estimated 30.6% infection prevalence suggests that natural immunity likely played a substantial role in Morocco's epidemic trajectory by mid-2022. This has implications for vaccination strategies, particularly regarding the timing and targeting of booster doses, and suggests that future response plans should consider both vaccinated and naturally immune populations. Finally, the robust methodological framework developed here could be adapted for other infectious diseases and geographical contexts, providing a valuable tool for ongoing surveillance system strengthening and future epidemic preparedness.

4.3. Limitations and Methodological Considerations

Our study is subject to certain methodological constraints. First, as with all model-based approaches, our estimates depend on structural assumptions and prior distributions, though sensitivity analyses confirmed robustness to parameter variations. Second, weekly aggregation may obscure day-to-day detection variability, despite aligning with SARS-CoV-2's serial interval. Finally, we do not account for spatial heterogeneity in detection rates across Morocco's regions or seasonal variations in reporting behavior itself.

Notwithstanding these constraints, our core findings remain robust, as evidenced by comprehensive sensitivity analysis (Section 3.4) and strong coherence with independent mortality estimates (Section 3.7), lending credibility to our assessment of under-reporting in Morocco.

4.4. Future Research Directions

Our study suggests several promising directions for future research. First, the integration of seroprevalence data from national studies could provide external validation of our estimates and refine our understanding of infection-detection relationships. Second, extending the framework to incorporate spatial heterogeneity could reveal important subnational patterns in detection efficiency and inform targeted interventions. Third, applying similar approaches to other middle-income countries would enable comparative analysis of surveillance system performance and identification of best practices. Fourth, incorporating additional data streams such as wastewater surveillance, mortality data, or mobility patterns could enhance model accuracy and provide complementary perspectives on epidemic dynamics. Finally, the adaptation of this framework for other infectious diseases with substantial under-ascertainment, such as influenza or dengue, could provide valuable tools for comprehensive burden estimation and public health planning across multiple disease threats.

5. CONCLUSION

Our Bayesian analysis reveals that Morocco experienced substantial under-ascertainment of COVID-19 cases throughout the pandemic, with only 10.6% of true infections detected by surveillance systems. The estimated 11.3 million infections represent approximately 30.6% of the population infected by mid-2022, with detection efficiency showing progressive improvement from $\eta_0 = 10.7\%$ to $\eta_{\text{final}} = 11.5\%$ across the study period. The methodological innovations introduced—particularly time-varying reporting rates and seasonal components—provide a more realistic framework for epidemic assessment in settings with evolving surveillance capacity. Our approach demonstrates clear advantages over constant reporting rate models and offers valuable insights for public health decision-making in similar resource-constrained contexts globally. The substantial surveillance gap identified underscores the critical importance of model-based corrections for accurate burden estimation, while the demonstrated capacity for system improvement highlights the potential for strengthening public health infrastructure during emergencies. As Morocco and similar countries continue to navigate the transition to endemic COVID-19 management and prepare for future epidemic threats, the framework developed here provides a valuable tool for evidence-based decision-making and surveillance system strengthening. This study not only quantifies Morocco's pandemic experience but also provides an adaptable methodological framework that can enhance epidemic surveillance and response in resource-constrained settings worldwide.

REFERENCES

1. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., et al. (2013). *Bayesian Data Analysis* (3rd ed.). London, UK: Chapman and Hall.
2. Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet*, **395**(10228), 931–934. doi: 10.1016/S0140-6736(20)30567-5.
3. Walker, P. G. T., Whittaker, C., Watson, O. J., Baguelin, M., Winskill, P., et al. (2020). The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. *Science*, **369**(6502), 413–422.
4. Nkengasong, J. N., & Mankoula, W. (2020). Looming threat of COVID-19 infection in Africa: act collectively, and fast. *The Lancet*, **395**(10227), 841–842.
5. Uyoga, S., Adetifa, I. M. O., Karanja, H. K., Nyagwange, J., Tuju, J., Wanjiku, P., et al. (2021). Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Kenyan blood donors. *Science*, 371(6524), 79–82.

6. Niama, F. R., Koukouikila-Koussounda, F., Mayengue, P. I., Elguero, E., Ngoulou, T. B., et al. (2022). Prevalence of SARS-CoV-2 antibodies in the Republic of Congo in mid-2021. *IJID Regions*, **5**, 121–123. doi: 10.1016/j.ijregi.2022.09.012.
7. Thenon, N., Peyre, M., Huc, M., Touré, A., Roger, F., et al. (2022). COVID-19 in Africa: underreporting, demographic effect, chaotic dynamics, and mitigation strategy impact. *PLOS Neglected Tropical Diseases*, **16**(9), e0010735. doi: 10.1371/journal.pntd.0010735.
8. Medhat, M. A., & El Kassas, M. (2020). COVID-19 in Egypt: uncovered figures or a different situation? *Journal of Global Health*, 10(1), 010368. doi: 10.7189/jogh.10.010368.
9. de Oliveira, A. C. S., Morita, L. H. M., da Silva, E. B., Zardo, L. A. R., Fontes, C. J. F., et al. (2020). Bayesian modeling of COVID-19 cases with a correction to account for under-reported cases. *Infectious Disease Modelling*, **5**, 699–713.
10. Richards, F. J. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany*, **10**(2), 290–301.
11. Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., et al. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, **584**(7820), 257–261.
12. Bhatt, S., Ferguson, N., Flaxman, S., Gandy, A., Mishra, S., et al. (2023). Semi-mechanistic Bayesian modelling of COVID-19 with renewal processes. *Journal of the Royal Statistical Society Series A: Statistics in Society*, **186**(4), 601–615.
13. Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., et al. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine*, **172**(9), 577–582.
14. He, X., Lau, E. H. Y., Wu, P., Deng, X., Wang, J., et al. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, **26**(5), 672–675.
15. Haut-Commissariat au Plan. (2022). *Projections de la population totale du Maroc 2022–2050*. Rabat: Direction de la Statistique, [Online]. Available: https://www.hcp.ma/
16. World Health Organization. (2021). *Infection fatality rate of COVID-19 inferred from seroprevalence data*, [Online]. Available: https://iris.who.int/items/675a4c1e-e4f7-4296-9eb1-d4b5e22f6f0e
17. Institute for Health Metrics and Evaluation. (2022). *COVID-19 Mortality Estimates for Morocco*. Seattle: IHME, [Online]. Available: https://covid19.healthdata.org/
18. Majidi, H., El Omrani, S., Touil, A., Ezzikouri, S., & Benazzouz, B. (2026). Morocco 2020–2023: three-year serosurvey of anti-SARS-CoV-2 antibodies in outpatients and health care workers. *Egyptian Journal of Medical Microbiology*, **35**(2), 507–518.
19. Ministry of Health Morocco. (2021). *National COVID-19 Seroprevalence Survey*. Rabat: Internal report.
20. World Health Organization. (2021). *Global COVID-19 Seroprevalence Studies*. Geneva: WHO.
21. O'Neill, P. D., & Roberts, G. O. (1999). Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society: Series A*, 162(1), 121–129. doi: 10.1111/1467-985X.00125
22. Grinsztajn, L., Semenova, E., Margossian, C. C., & Riou, J. (2021). Bayesian workflow for disease transmission modeling. *PLOS Computational Biology*, **17**(1), e1008618.
23. Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, **15**, 1593–1623.
24. Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., et al. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, **76**(1), 1–32.
25. Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C. J., et al. (2023). PyMC: a modern probabilistic programming framework in Python. *PeerJ Computer Science*, **9**, e1516.

Appendices

A. MODEL STRUCTURE SCHEMATIC

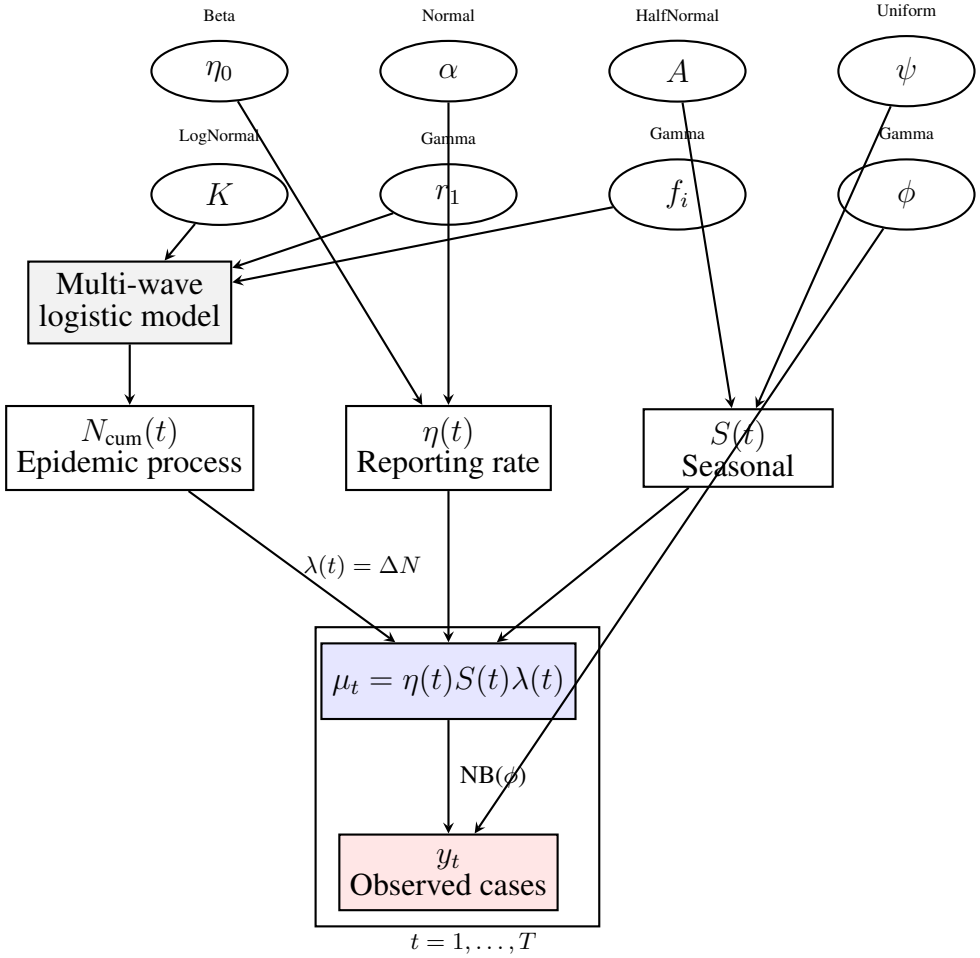

Fig. A.7. Schematic representation of the hierarchical Bayesian model structure.

B. HAMILTONIAN MONTE CARLO IMPLEMENTATION DETAILS

We use the HMC sampler with NUTS as implemented in PyMC. Gradients of the log-likelihood with respect to the model parameters are computed via automatic differentiation. The key derivatives needed for understanding the model are provided below. The detailed leapfrog and NUTS algorithms, along with the adaptation mechanisms, are standard in the literature and have been omitted here to keep the appendix concise.

B.1. Gradient Computations and Automatic Differentiation

For each time point t, the log-likelihood contribution is:

$$\ell_t = \log p(y_t \mid \mu_t, \phi) \tag{B.17}$$

The partial derivatives with respect to μ_t and ϕ are:

$$\frac{\partial \ell_t}{\partial \mu_t} = \frac{y_t}{\mu_t} - \frac{y_t + \phi}{\mu_t + \phi} \tag{B.18}$$

$$\frac{\partial \ell_t}{\partial \phi} = \psi(y_t + \phi) - \psi(\phi) + \log\left(\frac{\phi}{\phi + \mu_t}\right) + \frac{\mu_t - y_t}{\phi + \mu_t} \tag{B.19}$$

where $\psi(\cdot)$ is the digamma function.

B.1.1. Gradient of the Expected Mean μ_t

Recall $\mu_t = \eta(t)S(t)\lambda(t)$. The derivative with respect to a generic parameter θ_j is:

$$\frac{\partial \mu_t}{\partial \theta_j} = S(t)\lambda(t)\frac{\partial \eta(t)}{\partial \theta_j} + \eta(t)\lambda(t)\frac{\partial S(t)}{\partial \theta_j} + \eta(t)S(t)\frac{\partial \lambda(t)}{\partial \theta_j} \tag{B.20}$$

where:

$$\begin{aligned}
\lambda(t) &= N_{\text{cum}}(t) - N_{\text{cum}}(t-1) \\
\eta(t) &= \text{logit}^{-1}(\text{logit}(\eta_0) + \alpha t_{\text{scaled}}) \\
S(t) &= 1 + A\sin(2\pi t_{\text{scaled}} + \psi)
\end{aligned}$$

B.1.2. Gradients of the Logistic Growth Components

For $L_i(t) = [1 + \exp(-r_1 f_i s_i(t_{\text{scaled}} - \tau_i))]^{-1}$, let $z_i(t) = r_1 f_i s_i(t_{\text{scaled}} - \tau_i)$. Then:

$$\frac{\partial L_i(t)}{\partial z_i(t)} = L_i(t)[1 - L_i(t)] \tag{B.21}$$

Thus:

$$\frac{\partial L_i(t)}{\partial r_1} = L_i(t)[1 - L_i(t)] \cdot [-f_i s_i(t_{\text{scaled}} - \tau_i)] \tag{B.22}$$

$$\frac{\partial L_i(t)}{\partial f_i} = L_i(t)[1 - L_i(t)] \cdot [-r_1 s_i(t_{\text{scaled}} - \tau_i)] \tag{B.23}$$

$$\frac{\partial L_i(t)}{\partial s_i} = L_i(t)[1 - L_i(t)] \cdot [-r_1 f_i(t_{\text{scaled}} - \tau_i)] \tag{B.24}$$

$$\frac{\partial L_i(t)}{\partial \tau_i} = L_i(t)[1 - L_i(t)] \cdot [r_1 f_i s_i] \tag{B.25}$$

These gradients are computed efficiently via automatic differentiation in PyMC, ensuring numerical accuracy and scalability.

